Othmar W. Winkler

Interpreting Economic and Social Data

A Foundation of Descriptive Statistics

Springer

Interpreting Economic and Social Data

Othmar W. Winkler

Interpreting Economic and Social Data

A Foundation of Descriptive Statistics

Springer

Professor Othmar W. Winkler
Georgetown University
The McDonough School of Business
Washington DC 20057
USA

ISBN 978-3-540-68720-7 e-ISBN 978-3-540-68721-4
DOI 10.1007/978-3-540-68721-4
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009920218

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

It is the responsibility of scientists never to suppress knowledge, no matter how awkward that knowledge is, no matter how it may bother those in power. We are not smart enough, to decide which pieces of knowledge are permissible and which are not. . .

Carl Sagan*

Introduction[1]

On a snowy winter morning I boarded a crowded city bus, unable to use my bicycle for the usual 10 km commute to Georgetown University. In preparation for teaching that morning, I began perusing the textbook on business and economic statistics that I had adopted for this course. At the next stop, a young woman took the seat next to me, the only one remaining in the full bus. Shortly after settling in, she turned to me:: 'Excuse me sir, is this statistics?' she motioned to the textbook. 'Yes', I responded, surprised, 'Business- and Economic Statistics.' At this, a look of revulsion overcame her "Ugh. . . Statistics was the only subject I could never handle in college. . ." She trembled at a memory that still haunted and upset her. At this reaction a series of similar, though less dramatic, occurrences came to mind.[2] Few other academic subject seem to evoke the distaste that the mention of statistics seems to elicit. Does it have to be that way?[3] I grappled with possible explanations for a long time This book is my response that evolved gradually over decades of teaching a variety of business, economic and general statistics courses, using the newest textbooks available, and being involved in survey work and statistical consulting. I wondered why these textbooks on business and economic statistics presented the subject matter as a watered-down version of mathematical statistics, which itself evolved from problems of measurement and observation in the natural sciences. These textbooks treat socio-economic data like the measurements in the natural sciences and present the subject as an application of probability, grouped around the Gaussian, Poisson, F, χ^2 and other statistical distributions, sampling and statistical inference. There was no interest or concern about how to interpret the messages about society contained in the wealth of published economic and social data.[4] They fail to see that this is the raîson d'ètre of the entire statistical enterprise which also should be the main purpose of statistics courses for social scientists. These courses fail to present statistics

* Quoted from "Be careful what you pray for. . .You just might get it" Larry Dossey, M.D. Harper, San Francisco, 1997, p. 165.

as the instrument for scanning the economic and social environment and to monitor important aspects of social reality.

It is the aim of this book to re-orient statistics towards making sense of economic and social data. It is an attempt to rehabilitate 'descriptive statistics' as a respectable part of statistics, re-orienting it toward the description of society which in fact was its original purpose and still is the ultimate goal of all statistical endeavors. This book is addressed to the literate and numerate public, trying to open their eyes to various basic facts that are commonly overlooked, in short, to lead them to a fuller awareness of simple basics, to encourage asking questions and to look for answers in the fine print that accompanies tabulations of socio-economic data.

It is also the aim of this book to draw attention to the neglected twilight zone, the no-man's land between the partisan efforts of statisticians who, inspired by applications in the natural sciences, turned to probability, controlled experiments, model building, etc. on one hand, and the applied fields of social sciences on the other. Statistical theorists feel that they are the guardians of a true science, concerned with the purity of its theoretical core with little regard for interpreting economic and social data. On the other side are social scientists and economists concerned about discovering timeless laws of economics, intent on condensing them into mathematical models. They too are less concerned about using statistical data to monitor and also influence events in society. And last, but not least, there are the dedicated statistical foot soldiers who take censuses and surveys, and prepare tabulations. They too have no time for making sense of their data about society.

As you may notice from the 'Outline,' this book departs from the usual structure, but instead follows the steps of the statistical process in a rather abstract, theoretical manner, from the very start of conceptualizing the socio-economic phenomenon to be investigated to the final tabulation of the data. Standard topics, like the Gaussian curve, probability theory and symmetrical, well-behaved frequency distributions are treated at the end of the book, if at all. The initial chapters deal at length with topics that are usually missing in textbooks[5] such as aggregation, statistical aggregates and ratios. They form the backbone for the interpretation of socio-economic data. Then follow three chapters on time series as the most frequent form in which data are published. These chapters are given priority over frequency distributions in one or more dimensions, treated toward the end.

This book, by the way, is not meant as an introduction to statistics, nor as a "how-to-do, hands-on" manual. Its concern is to make sense of socio-economic data. and to shine new light on various misconceptions the reader may have acquired in previous statistics courses. Only a minimum of mathematics will be required. Calculations are relegated to the five optional appendices. Although mathematical statisticians may find this book pedestrian and simplistic, some abstract thinking is involved and the reader is asked to be patient with unfamiliar ideas.

This book is intended for everyone who has to deal with data about society: students and teachers in business, economics and social science courses, economists, social scientists, financial analysts, market researchers, business and economic forecasters, sociologists, managers, demographers, even geographers. It is my hope

that the chapters of this book will open up a new understanding of socio-economic data for their meaningful interpretation, allay bad feelings toward our field, and stimulate further developments in the indicated direction.

Description of Chapters

Chapter 1 provides a short view of the developments that led to the present situation in socio-economic statistics. The powerful influence and the band wagon effect of the developments in statistics in biology, agriculture came to dominate all fields of statistical application. This chapter points out that socio-economic statistical data are quite different from the measurements in the sciences.

Chapter 2 traces the statistical process, from the conception and formulation of a socio-economic phenomenon, such as unemployment, poverty, productivity or crime; to the identification and recording of the relevant 'real-life-objects' which portray that social or economic phenomenon: human beings, entities such as corporations, or events, such as births, work accidents or business mergers. The simplified records of these 'real-life-objects' then become the 'statistical-counting-units'.

In Chap. 3 the subsequent grouping of these 'statistical-counting-units' into suitable aggregates is discussed. These new entities, the statistical aggregates, are defined by their three 'dimensions': the subject matter, the time period, and the extent of the geographic area covered. As to the subject-matter "dimension", the qualitative characteristics of the statistical counting units are important for the formation of a hierarchy of sub-aggregates. The magnitude of each of the three 'dimensions' of an aggregate determines how to interpret the gains and losses from aggregating the 'statistical-counting-units'. These statistical aggregates represent the bulk of the data in socio-economic statistics. They are quite distinct from the data in the natural sciences, an important matter that has not received due attention.

In Chap. 4 a variety of ratios is discussed as simple and effective analytical tools. These ratios allow us to perceive and make sense of the underlying economic and social reality conveyed by these aggregates. Despite their pervasiveness and importance, ratios have rarely been discussed.

Chapters 5, 6 and 7 study the development, over time, of economic and social phenomena through time-series of socio-economic data.

Chapter 5 presents a critical view of the customary decomposition of time series into trend, seasonal pattern, business cycle and randomness. Instead of the mathematical decomposition into the standard components, time- series should be understood as quantitative economic and social history that can be interpreted meaningfully through a hierarchy of simple ratios between aggregates. These figures are not to be understood as abstract algebraic numbers.

Chapter 6 explores the fact that statistical data lose their relevance over time and become obsolete and less relevant for anticipating the future of a situation in society. Good forecasting requires acquaintance with the historic development of the underlying economic or social forces. Much depends on the speed with which

the data become obsolete. The level of aggregation also affects obsolescence. All this requires judicious decisions regarding the weight older data should be given in a forecasting model, and the point in the past from which on the data of a time series should be disregarded.

Chapter 7 has two parts. In the first part, Sect. 7.1, Price-Index-Numbers are discussed as an important type of time series. A simpler, ratio-based approach is presented[6] that is more transparent and easier to interpret than the historic Price-Index-Number formulations currently in use, allowing for understanding and interpreting the actual changes in price levels.

In the second part, Sect. 7.2, Index-Numbers of Production are critically reviewed. Different production concepts are discussed and simpler ways of measuring production and productivity are developed.

Chapter 8 deals with the interpretation of highly asymmetric frequency distributions that predominate in economic and social data. Simple measures are presented to deal appropriately with these highly asymmetrical data, to assess and interpret centrality, asymmetry and dispersion.

Chapter 9 discusses the puzzling case of one particular regression analysis that changed my views on cross-sectional data in general. Without going into the algebra of their calculation, specific problems in Regression and Correlation with aggregate data are discussed

Chapter 10 explores the relationship between statistics and the calculus of probability. Although socio-economic statistics is numeric, using mathematical symbols, algebra, geometry and graphs, it must not be considered as a branch of mathematics.[7] Socio-economic statistical data have an important conceptual non-numeric component that defies a numbers-only approach. One must keep in mind that its purpose is the perception of very real economic and social happenings in historic time, and in geographic and subject-matter space. Misuses of probability, foremost the mis-interpretation and misuse of "Statistical Significance," are critically reviewed[8].

Chapters 11 and 12, explore areas that social, business, and economic statistics has in common with subjects that do not readily come to mind as linked with statistics. While exploring these areas in these two final chapters the nature of socio-economic statistics is further clarified.

Chapter 11 has more in common than is usually acknowledged.[9] When statistics is not considered as a branch of mathematics, however, it is easier to see that macro economics, really National Accounting – which is essentially economic statistics – keeps track of the economy like financial accounting keeps track of a business corporation. The discussion reveals surprising affinities between socio-economic statistics and financial accounting.

Chapter 12 discusses the importance of geographic-spatial distributions, a matter that has been absent from the theory of statistics, though not from statistical field-work. Although specialized quantitative-statistical research abounds in geography, the geographic-spatial dimension has not been recognized as belonging to statistics[10] and ought to be included in its theory.

Acknowledgments I gratefully acknowledges help received from colleagues, Michael Czinkota, International Marketing, Jackie Hoell, Business Statistics; Harvey Iglarsh, Management Science; Denis F. Johnston Sociology, Ursula Kueck, Economic Statistics, Alan Mayer-Sommer, Accounting; Stanley Nollen, Managerial Economics; Kimberly Sellers, Mathematical Statistics; Annette Shelby, Business Communication; Robert Thomas, Marketing; Wilfried Ver Eecke, Philosophy, Daniel Westbrook, Econometrics, Elizabeth Winkler, English, Deepak Sahay, my graduate student assistant. Each has contributed in some way with critical observations, knowledge and encouragement to shape the ideas in this manuscript. And last but not least, Dr. Niels Peter Thomas of Springer Verlag without whose initiative this book would not have been written.

Notes

1. This book is the result of my six decades of teaching all kinds of statistics on graduate and undergraduate levels, as well as the result of my involvement in many interesting statistical studies and surveys. I gradually arrived at the conviction that much of current statistical theory does not apply to economic and social data, a situation that can be imagined as two only partially overlapping circles or ellipses, one representing the general theory of statistics and the other, socio-economic statistics. The small area of intersection where these two geometric figures overlap represents the methods of the general theory of statistics that cover the needs of socio-economic statistics.
2. This is the consequence of the fact that statistics has been identified with mathematics in general and with the calculus of probability in particular. The reaction to mathematics, described in the following address to the German mathematical society, is a familiar story in which the word 'mathematics' can readily be substituted with 'statistics': "It's the same old refrain: . . .I hate math . . . Pure torture from the start of school. It's a total mystery how I ever managed to graduate. . . A nightmare for me. . .Mathematical formulas are. . . pure poison. They just turn me off . . . Complaints such as these are heard all the time . . . educated people express them routinely with a blend of. . .defiance and pride. They assume their listeners sympathy. . .." p. 9 "Drawbridge Up", translated from "Zugbruecke Ausser Betrieb" Hans Magnus Enzensberger, translated from the German by Tom Astin. Published by A.K. Petus, LTD, Natuck, Ma, 1999.
3. The recent trend toward reducing the required course offerings of statistics in business schools and economic departments in European and also in American universities, seems to be rooted in similar experiences. It appears as a reaction to a sense of frustration with the kind of statistics taught. Though intellectually challenging, it seems to be of insufficient relevance for the social sciences, at least at the undergraduate level, to justify the effort needed to master it. The amount of time dedicated to this kind of statistics is to be allotted to other courses, believed to be more relevant.

   Business and economic statistics courses in academic curricula are dominated by mathematical statistics that evolved from the physical-natural sciences. These courses further assume that the problems and data in economics and related fields are like those in the natural sciences, and that all students who take a statistics course want to become professional statisticians. This situation has become entrenched because those, believed to be qualified to teach statistics in economic and business curricula, are expected to have a background in mathematical statistics, taught in mathematics departments. These same academicians also act as reviewers of statistical journals, favoring mathematically oriented manuscripts for publication. That situation has continued for some time, but efforts are beginning to be made to change that.

   The recent discussion in statistical journals, e.g. in the 2003 issues of the "Allgemeines Statistisches Archiv", revealed a concerned questioning by academic statisticians, why statistics, as a core course in economic curricula, is losing ground, and why the required hours of teaching statistics are being curtailed. In short, why the interest in statistics is waning. Peter von der Lippe, Sibylle Schmerbach "Mehr Wirtschaftsstatistik in der Statistikausbildung

fuer Volks- und Betriebswirte" Allgemeines Statistisches Archiv 87, pp. 335–344, 2003 – Hans Peter Litz, Curriculare und fachsystematische Aspekte einer univrsitaeren Wirtschafts-und Sozialstatistik, Allgemeines Statistisches Archiv, 88, pp. 347–361, 2004; – Werner Gruenewald, Hans-Joachim Mittag, Michael Mueller, Reiner Staeglin, Peter Lorscheid und Roland Gnoss, "Diskussionsbeitraege zu "Mehr Wirtschaftsstatistik in der Statistikausbildung fuer Volks- und Betriebswirte", AStA 88, pp. 100–117, 2004- Peter von der Lippe, Sibylle Schmerbach "Antwort zur Discussion um 'Mehr Wirtschaftstatistik in der Statistikausbildung fuer Volks- und Betriebswirte'" AStA 88, pp. 362–367. Evidently colleagues in economics departments and business school have lost patience with this kind of statistics as it has been taught. Yet, the most they feel able to do is to support the reduction of the amount of statistics in the curriculum. In analogy to Lester C. Thurow's highly critical assessment of the (then) current state of economics, the situation of statistics could be named appropriately "Dangerous Currents: the State of Business-, Economic- and Social Statistics." Lester C. Thurow, Dangerous Currents: The State of Economics, Random House, New York, 1984 (esp. Chap. 4 "Econometrics: the icebreaker caught in the ice").

4. In the following chapters the term 'data' will be used in the plural when referring to tabulated statistical numbers, (. . .data are. . .) and in the singular when referring to general, not necessarily quantitative, information (. . .data is. . . .).
5. The following quotation is applicable to socio-economic statistics, expressing what such a required change in mental outlook entails:

   "This experiential horizon is socially and historically, that is, conjecturally conditioned; it shares in the historical character of the whole life of man. With Thomas Kuhn we can register a dual form of progress in our human cogitation: (a) a homogeneous, mostly continuous development within one and the same intellective model. . .this is an evolutive progress; (b) progress through a fundamental change in the (conjunctural) horizon of experience or the 'intellective model,' whereby the meaning already attained has to be 'translated' anew; this process entails something of a 'revolution' (. . .prior to the 'revolution' there is always a prerevolutionary situation in which for some time past the model had actually ceased to work.) Every sharp change in an intellective and empirical horizon still has its own history! Besides long periods of quiet, homogeneous progress, every so often history exhibits more fundamental jerks: a transition from one historical or conjunctural intellective horizon to another. When a new model has been found (p. 580) it takes time to be accepted by everybody as new evidence. . . for a while old and new culture models will co-exist; the respective champions of the two models often come into conflict; there is even polarization at times: two groups of people, though contemporaries, live in mutually 'alien' worlds, they cease to understand each other. For the result. . .goes deep and reaches wide. As Wittgenstein says: . . .'What were ducks before the revolution are rabbits afterwards' What in the old model of physics was a solid. . .chair appears in the new atom-model as a kind of empty space with atoms and molecules whirling. . .about inside it. An 'outsider' hearing about this for the first time, will either shake his head in disbelief – or angry protest- over such a new-fangled aberration, since the chair's solidity seems perfectly obvious. . .(p. 581) Schillebeeckx, Edward, *Jesus, an Experiment in Christology*, A Crossroad Book. The Seabury Press, New York, 1979 (A translation of "Jezus").
6. In American textbooks of business- and economic statistics the measurement of price levels usually is presented as a historic peculiarity, separate from the rest of the material that is based on the theory of probability. The straightforward description of the reality of prices is eclipsed by the pseudo-problems of Index number theory. Because socio-economic statistics, in general, has shown little interest in the simple description of reality, no effort was made in price statistics e.g. to clarify, what the things or objects to be reported, the 'statistical counting units', ought to be. See: Winkler, O.W. "Measuring the Price Reality or Measuring a Price Illusion?" *Proceedings of the 11th Tagung fuer Preismessung,* Lutherstadt-Wittenberg, Germany, July 2006.
7. The concept of a 'sample space,' for example, must remain limited to those instances in which data were actually selected as samples. There is a tendency to consider all data as samples,

even those that include the entire population. Similarly, the concept of a 'super-population' though useful for sampling theory, is not to be considered for this approach. Another example is the mathematical formulation of a regression surface. It is of interest only within the domain of the actually reported data. The facility with which analysts 'transform' and distort their data, is symptomatic for this attitude, e.g. Mosteller, F., Tukey, J. *Data Analysis and Regression*, Addison Wesley Publ. Co, Reading, Mass. 1977. Such manipulations can distort the reality underlying the data, and are to be avoided.

8. The following quotation is but one of the innumerable examples: "The statistical methodology for analyzing data. . . is called statistical inference. . . The logical foundation of statistical inference is the mathematical theory of probability" (p. 75) Neter, John et al. (Boston 1978) Or another example: "The data may be obtained from published sources, through survey research or by designed experiment. However obtained, the data are the observed outcomes or responses of some phenomenon of interest or underlying *random variable.*" (p. 3). (italics added for emphasis). Berenson, et al. (New Jersey, 1983).
9. This chapter follows closely Winkler, Othmar W. "Statistics in Accounting other than Sampling" *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1976, pp. 654–659 as well as a modified version of this article entitled "Secret Allies?" *Management Accounting*, National Association of Accountants, Montvale, NJ, June 1985, pp. 48–53.
10. Such topics are described in publications with titles such as "*Statistical Concepts in Geography*" John Silk, London, Allen & Unwin 1979, esp. pp. 22–26 and 206–249 A list of other pertinent literature in Winkler, O.W. "The Interface of Geography and Economic Statistics" *Proceedings of the Business and Economic Statistics Section, AS*A, Washington, DC, 1981.

Contents

Chapter 1
Developments in Socio-Economic Statistics

1.1 Stating the Problem

Statisticians accept as a self evident principle that there is one general theory of statistics that applies equally to all fields,[1] biology, economics, engineering, demography, environmental sciences, sociology, etc. (Fig. 1.1).

Yet, important applications in economics and the social sciences in general are not covered by what today is considered 'the theory of statistics.'

This calls for a review of the situation, of methods that do not apply, and important aspects of socio-economic applications that are not supported by statistical theory. The peculiar nature of the data in socio-economic statistics requires a different basis than is available at present[2] and makes it unlikely that a general 'Theory of Statistics' can satisfy the needs of this scientific field. Historically, the turn toward inference came from the discovery of random sampling, from experimentation in agriculture and other applications in the natural sciences. We proceed as if socio-economic statistical data are like those in the sciences, ignoring that they differ in important ways. Because of this, the applications of social, business and economic statistics are not adequately supported by today's statistical theory (Fig. 1.2).

1.2 The Anglo-American Influence

The influence of the Anglo-Saxon bio-mathematicians came to dominate the development of statistical theory. The ideas of K. and E. Pearson, R. Fisher, F. Yates, Wm. S. Gossett, M.M. Bartlett, J. Neyman, and other biometricians from the British school of thought found a fertile ground in the USA, partly due to the accessibility of their publications through the common language, and partly due to their common interest in the bio-sciences and engineering. The resulting development could be called the Anglo-American theory of statistics having entered business and economic statistics as 'decision-making under uncertainty' of value for business corporations and government. The Anglo-American statistical theory moved probability into a prominent position about which more is to be said in Chap. 10. Yet, the bulk of actual statistical work in the social sciences is directed primarily at the realistic perception of socio-economic phenomena such as price level movements,

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_1, © Springer-Verlag Berlin Heidelberg 2009

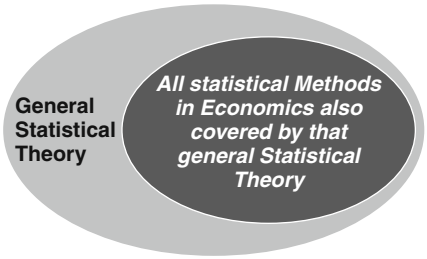

Fig. 1.1 Belief that statistical Theory covers the needs of the social sciences

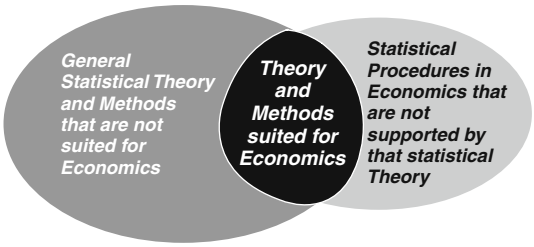

Fig. 1.2 Partial Availability of Statistical Methods in Economics and the Social Sciences – Absence of an adequate Theory of Socio-Economic Statistics

demographic developments, industrial production, foreign trade or labor problems. The subsequent evaluation and interpretation of the data is the important aim of all statistical efforts. The present theory of Anglo-American statistics, however, is not directed at the interpretation of the economic and social situations described by these data, yet insisting that the available theory is appropriate and sufficient.

Authors of textbooks on business and economic statistics acknowledge their debt to the mathematicians and biologists R. Fisher, K. Pearson and 'Student', but do not acknowledge a greater debt to W. Leontief, R. Stone, S. Kuznets, J. Tinbergen, E. Laspeyres and others for their contributions to socio-economic statistics proper. The roots of this obvious mis-orientation go back to Adolphe Quetelet's physique sociale, his idea of physical laws governing society like the laws in the physical sciences that were recent discoveries of his time. This idea, typical of his 'Zeitgeist' had a long-lasting influence. Quetelet popularized the idea that society could be treated as if it were a branch of the natural sciences. This idea was also accepted and developed by mathematical economists like Walras and Marshall, later leading into econometrics. All this consolidated the influence of these positivist ideas,[3] particularly by econometricians like R. Frisch and T. Haavelmo.[4]

The other, related source of this mis-direction is the mistaken assumption, that socio-economic statistical data are point-like and objective like individual measurements in the natural sciences. The present theory, based on this, ignores the subjective and aggregative nature of our data.

The contributions of the Anglo-American statistical theory to economic and social statistics are essentially the probability-based methods of inference and the calculus of probability.

Strictly speaking, probability calculus only allows insight into the effects of the randomization of experiments and the random selection of items in samples. Even in legitimate applications of sampling, statisticians, in actual practice, often prefer samples that are representative rather than random. They prefer samples by stratification or by purposive selection over samples selected according to the principles of unrestricted, pure random procedures. For example, in price statistics, typically consumed goods, a market basket, is chosen to be priced, rather than a random selection of goods. The calculus of probability allows for inferences into the corresponding parameters in the population, and by extension, the testing of hypotheses about the numeric value of such parameters. The concentration on these topics has prevented a balanced development of a general theory of business and economic statistics.[5] It simply goes too far to view economic processes as random experiments,[6] and all economic and social data as random samples, even when it is obvious that the data in question are populations, in the statistical sense. Business statisticians and econometricians routinely refer to deterministic economic processes as random variables and random deviations.[7] Even a population completely surveyed by a census is treated as a random sample from an imaginary super-population.[8] Similar is the widespread practice of treating the time series of yearly, quarterly, monthly, etc. economic data[9] as samples of[10] sampling-units that are assumed to be randomly selected from some population[11] of such units. Statistical procedures based on such assumptions have only a tenuous relation to the underlying economic reality. One might say that the Anglo-American theory of statistics perceives economic reality through the conceptual framework of probability, as if viewing reality through a probabilistic lens.[12] Yet, neither social nor economic development occur 'by chance' or 'at random.' On the contrary, it is precisely through interpreting the data about those economic and social economic situations that the underlying forces can be studied.

To these significant problems one must add the problems of measurement[13] of economic and social data, which are quite different from the measurements in the natural sciences. This will be discussed in greater depth in the following chapters.

One ought not to be surprised that such development took place predominantly in countries with free market economies, in which the individual economic actors link their perception of daily uncertainty to the assumption that the world actually may be stochastic. In socialist countries, e.g. in the former eastern block, on the other hand, government control and economic planning may have caused resignation, but not a feeling that the economy consisted of random processes. The theory of business, management and economic statistics in those countries did not assign a comparable place[14] to probability theory.[15] The question whether the economy, and consequently also socio-economic statistics, should be understood as an actual network of stochastic processes, is a political and philosophic issue, not a mathematical one.[16]

It is also worth noting that Statistics appears to be developing in three steps. Initially its theory developed around 'parametric statistics' – mostly in the sciences. Later it was extended to include 'non-parametric statistics.' This book initiates a further extension into 'non-quantitative statistics' a development in widening circles, beginning with symmetrical, well behaved frequency distributions, progressing from these more manageable continuous quantitative variables e.g. in the engineering data of industrial quality control, to the less manageable discrete variables, and in this book to the vast area of the characteristics that cannot even be measured or expressed in numbers but are of fundamental importance: the qualitative variables or attributes. Sampling, randomness, probability and inference are removed from their exalted place as the center of attention, while extending statistics in new directions by exploring the interface with other fields, accounting, human geography and philosophy – epistemology.

1.3 Socio-Economic Statistics and Decision Theory

(I have made up my mind. Don't confuse me with the facts)
(Anonymous)

In the late sixties, many universities in the USA began consolidating the courses on Business and Economic Statistics with courses on Decision Theories and Decision Making. The administrative convenience was evident. The real reason, however, was the obvious affinity between these two groups of courses: both were presented as based on a stochastic view of society and probability theory. Statistics was presented as an extension of making decisions under uncertainty. Such consolidation seemed only a question of time. Nevertheless, some serious objections had to be raised against it.

First, the conditions under which probability calculus, particularly the frequentist kind of probability that prevailed in courses of statistics, can predict the results of games of chance differ from those of actual business decisions. Their risk is of a different nature than that evident in games of chance. In the latter the rules of the game are fixed and known to the players in advance (the decision makers). All possible outcomes are known beforehand. Once the game begins, the rules cannot be changed. The outcomes can be predicted only for the long run, that is, when such a game is continued for many rounds. There are indeed few economic decisions of this invariant and repetitive nature[17] in which the probability rules of games of chance can be applied meaningfully.[18] Most business or economic decisions are made either as a compromise between the divergent views of the situation by the voting members of an executive committee, or by a corporate executive officer, without the tensions and benefits of a multidimensional perception of the situation. Economic decisions are judged by their success in the marketplace, and are based on a multiplicity of short and long range considerations, the most important of which often cannot even be quantified. Rarely can such decisions be made according to the rules of games of chance.[19] The study of such decisions is of great interest but

really belong in courses of management, finance or marketing, rather than in one of socio-economic statistics.

Second, it is important to understand how statistical input is brought to bear on business decisions. It provides the economic panorama for the decision, together with other non-statistical information. Typical were the weekly sessions of the directorate of the Du Pont de Namour corporation at which the updated, pertinent economic data series were presented and discussed.[20] No immediate, concrete decisions followed from the knowledge of these data. Its high-level participants kept this statistical panorama, as it were, in the back of their minds, for the appropriate moment when a decision would be made. This is akin to a situation after a college examination when the instructor publishes the distribution of grades, and each student can assess his position among his peers. Those who ought to make changes in their study habits[21] will not necessarily act[22] based on such available information.[23] If, however, they do decide to act,[24] then they will use the given information as a guide[25] in that decision-making process, but will not allow themselves to be forced to act in a specific way, like a cogwheel in a mechanical gear box.[26] Nobody can object to a course in decision-making, but it should not take the place of business, economic and social statistics properly speaking.

1.4 Misconceptions in Socio-Economic Statistics

Statistics is often popularly characterized as measuring and counting. This uncritical transfer of concepts from the natural sciences to the social sciences is misleading. It is important to note that the data* in socio-economic statistics are of a different nature,[27] to be further discussed in the next chapter. We are not helping matters by referring to the determination of a characteristic as a measurement. Whatever name we assign this process, measuring e.g. the weight of a piece of zinc oxide on an electronic scale, clearly is a different proposition than e.g. determining, through financial accounting, the net value of a business firm for a given time period. Both are referred to as measurements. Yet, the data in the social sciences are not the result of direct observations done by an objective, specially trained outside observer, like for example in microbiology. Most socio-economic statistical data are, in contrast, self-observed, intended to inform about facts that are on a questionnaire or verbally reported to a survey taker, who usually does not do the observation of facts him/herself. These observations properly speaking, are usually carried out by those who are to be observed, as self reporting.[28] Very few statistical observations relating to economic facts or other aspects of society are made directly by an objective outside observer, like in the daily work of scientists. Instead, the accuracy and veracity of the information depends on the level of education, good will, disposition to cooperate, and the honesty and unfailing memory of the interviewed. This provides an

*In the following the word 'data' will be used in the plural (Data are...) when referring to statistical numbers. 'Data' will be used in the singular (data is...) when referring to any kind of information, not just numeric ones.

important difference with the data in the natural sciences. As the subsequent discussion will show, socio-economic data are aggregates that transmit the economic and social reality differently than is generally assumed. One must realize that economic facts, such as e.g. the employment and labor force participation status of persons is not determined or measured with a precision gauge or with an electronic scale.

Nor is counting what it appears to be. The economic entities, e.g. business firms, report their characteristics via questionnaires. It is these questionnaires, not the persons, business firms, etc. that are the things to be counted, as will be discussed in Chap. 2. These questionnaires will be aggregated and provide a stripped down, abstract picture of socio-economic reality, as discussed in Chap. 3. Statistics' role is that of a reduction lens which condenses phenomena that are too far dispersed to be perceptible.[29] These economic and social phenomena are too widely scattered, geographically, subject-mater-wise and over time, to be perceptible without the help of statistics. It acts as a macroscope,[30] an instrument that allows for the perception of things that are too big or too widely scattered to be seen, the opposite of a microscope, which amplifies phenomena that are too small for the unaided eye. The individual cases themselves, represented by questionnaires, are of little interest. It is their distribution over regions, time and subject-matter categories that is the key to interpreting socio-economic phenomena.

The Anglo-American statistical theory is aimed at phenomena of the natural sciences in which the regional-geographic aspect is of far less importance than in the socio-economic phenomena. This unrecognized difference in the phenomena is also the cause of the present disregard for the distinctive nature of socio-economic statistical data, most of which are aggregates.[31] Because of this false orientation, statistical numbers are treated as if they were un-historic, un-geographic, direct measurements that in subsequent analyses could be manipulated mathematically (transformed), or be condensed into a single measure – e.g. of central tendency such as the arithmetic mean or a least squares trend line of these numbers – as the best approximation of what is believed to be their' true value'. This view, treating socio-economic data as if they were deviations and residuals from some true value,[32] afflicted by a random disturbance,[33] is at the heart of today's thinking. Analysts of socio-economic data seem to forget that the calculated numbers e.g. of a fitted trend curve, yield hypothetical, not real values, and that the 'raw' socio-economic statistical data, made to appear as mere random deviations,[34] are the ones to be taken seriously. They are not deviations from a calculated fiction that is removed from economic and social reality, but are the values that represent actual reality.

Aggregation has been occasionally referred to as 'data reduction'. This is another misleading expression: it is not the number of cases that is being reduced,[35] but the detail with which the counting-elements in higher-level aggregates are described. In larger aggregates there is less detail available, but the number of cases in the aggregate remains the same or is larger. Instead of referring to aggregation as data reduction, I proposed 'phenomenon enhancement'[36] stressing the positive effects of aggregation. Data are to be grouped in such a manner that the socio-economic phenomenon of interest is best highlighted or enhanced. For example, the salaries

of the employees of a large business firm were grouped in such a way that they created the erroneous impression that male and female employees were paid the same for comparable work, when in reality this was not the case. Thus the social phenomenon sex-discrimination can be made to disappear through the clever re-grouping of employees' salaries. Obviously, the responsible interpretation of such statistical data requires re-grouping the salaries of the employed men and women in a manner that reveals, rather than obscures or hides, the phenomenon in question.

1.5 Symptomatic Omissions

The gaps and omissions found in the textbooks of business and economic statistics reveal the areas that are similarly missing from the theory of socio-economic statistics.

Statistical aggregates are neither discussed nor recognized in their actual geographical historical-institutional context. Population and other economic censuses are hardly ever mentioned in textbooks. Statistical theory rarely contributes to the understanding of categorical or qualitative characteristics.[37] Yet, these categorical variables that cannot be determined with precision, prevail in socio-economic data, and are more important than the quantitative characteristics in describing socio-economic reality. Because of its orientation toward measurable, quantitative characteristics of the natural science data, the theory in textbooks of business and economic statistics fails to discuss the important classifications of economic activities SIC and NAICS, of occupations, and of products. Similarly ignored are the important geographic or spatial distributions. Separate specialized treatments of these topics do exist[38] but are not part of the theoretical foundation of statistics as applied to the social sciences. Nor is there a place for considerations of an international kind at a time when globalization requires the attention of leaders in business, politics and the economy.[39]

Most frequency distributions in socio-economic statistics are decidedly asymmetric. Yet, the orientation toward data in the natural sciences, where symmetrical distributions prevail, has not recognized this. As previously stated, the phenomena in the social sciences differ from those in the natural sciences and these typically highly asymmetric frequency distributions require special treatment with classes of unequal widths, a matter that is rarely mentioned, and whose interpretation, though important, is not on their agenda.

Related is the fact that the statistical perception of reality – disparagingly referred to as only descriptive statistics – is least valued. Publishers of textbooks have advertised, as an improvement in a new edition, that the space allotted to descriptive statistics has been further reduced, in favor of more statistical inference. That misses the point, however, that the original purposes for which business economic and social statistics are produced, is to scan society and its changes, and to report its findings. Statistical methods, most of them transferred from statistics in biology and other natural sciences, hardly take note of economic and social factors and do not present

methods to study phenomena such as the extent and intensity of unemployment in different parts of the country, by age, gender, race, occupation, industrial activity, etc. By failing to acknowledge the subject matter-time-space dimensions of social phenomena, statistical theory has turned its back on socio-economic reality, limiting its concerns to concepts of random sample selection, random variable, inference from samples, least squares. and related sample-theoretical considerations.[40] It is in vogue to construct and study models of reality, rather than to study that reality itself. It is questionable that much can be learned about a situation[41] through simulation exercises[42] and testing of hypothetical models.

Much effort is expended on testing probabilistically, whether a chosen statistical model fits the given data. This kind of research rarely allows insight into and understanding of real socio-economic phenomena.[43] The substantive areas of statistical applications to society which e.g. filled the bigger second volume of W. Winkler's *Grundriss der Statistik*[44] are simply not dealt with in the Anglo-American textbook literature. Obviously there is no statistical theory available to aid those who must deal with real socio-economic phenomena.

In short, the practical application of statistics to business and economics is not properly supported by an appropriate theory. Theory and practice are like the two intersecting, only partly overlapping, circles or ellipses of a Venn diagram (see Fig. 1.2). Those topics to which statistical theory is correctly applied are represented by the relatively small common area of these ellipses. This happens in data that belong to the sciences, such as industrial quality control and related engineering data, as well as data of aptitude and skill tests of employees. These belong to the sciences towards which today's statistical theory is oriented, while statistical theory fails to recognize that economic and social data are either aggregates or derived from such aggregates. At any rate, they are quite different from data in the natural sciences.

The amount of attention devoted to random draws from urns and decks of cards, to throws of dice and coins, and to spins of roulette wheels, are out of proportion to the minor importance probability concepts play in the interpretation of actual socio-economic data. In general, textbook authors are concerned with computing, rather than with explaining and interpreting the results.[45] Thus, generations of future leaders in business and society are instructed in theoretical knowledge that is relevant to socio-economic statistics only to a very limited extent.[46]

1.6 Recommendations for the Future

1.6.1 Beyond Sampling and Inference

What should a future theory of business, economic and social statistics contain? Although sampling techniques and the inference from samples are important, socio-economic statistics literally has been trapped for decades in it as its near-exclusive theory. The situation has not changed with the emergence of non-parametric methods of inference and multivariate analyses. Despite their limited scope, sampling,

inference and decisions based on it are treated as if they were *The Theory of Statistics.* It was precisely these limited concerns that have kept statistical theorists from returning to the interpretation of the situations described by socio-economic data, which really is the ultimate purpose of statistics. Historically there were similar episodes of the exclusive and limited concern with certain topics. At the turn of the 20th century, for example, discussion centered on the measures of location, dispersion, and index numbers. Neither one of these developments contributed significantly to interpreting socio-economic data

The time has come to break out of the confinement of many decades of exclusive concern with sampling and inference[47] and to re-orient statistics to interpret the phenomena of society through all kinds of data, not only those from samples. The entire process, from the early draft of the concept of what exactly is to be investigated, to the final presentation and the appropriate storage of results, must be part of a theoretical framework of data interpretation.[48]

As statistical aggregates are the instruments through which reality is perceived, these aggregates, the data, ought to be the starting point of all statistical theorizing. Aggregation must be recognized as centrally important. Instead, statisticians have turned to probability to look for answers and by doing so, have further put off the real task of interpreting the situations in society as they are reflected in the data.

1.6.2 Shifts in Emphasis

A shift ought to take place, from the frequency distribution approach with the tempting mathematical treatment of numeric characteristics, that prevails in the data of the natural sciences, to the less tractable qualitative and geographic characteristics, the typical determinants of socio-economic data. These, though not as readily convertible to numbers, are the basic features of the data about economic and social phenomena. Returning to its two original functions of capturing and interpreting reality, statistics must deal with distributions by attributes and geographic regions.

The importance of formulating and testing hypotheses about situations in society for managers, business analysts, politicians and lawmakers must be questioned, despite its great interest for research in the natural sciences. Most of the hypotheses formulated in econometrics cannot be legitimately tested in the same way as e.g. hypotheses in the engineering problems of statistical quality control.

The discussion of price measurement needs to be expanded beyond the customary formalistic treatment. Basic issues need to be discussed such as, ‘What is price?’, ‘What is its nature?’, and ‘What is production?’ Price level changes should be discussed as part of time series, not as a separate oddity. The recent, more inclusive social indicators should become part of the wider discussion of economic indicators. All this should become part of a foundation for descriptive socio-economic statistics.[49]

1.6.3 Filling Voids

The classification systems which underlie the aggregates of socio-economic data are rarely discussed in textbooks on socio-economic statistics. They should also become part of statistical theory. The relation between the socio-economic phenomena and the statistical data aggregates will have to be clarified. In the interpretation of time series and in forecasting, such a comprehensive statistical theory must allow for the combination of the quantitative description of these unique, historic and geographic socio-economic situations with the tools of historiography, sociology, philosophy, management, and economics, not with probability theory except in those instances where it is truly warranted.

National accounting, as part of macroeconomics, also belongs in socio-economic statistics, but is not mentioned in textbooks, even though it is the descriptive framework that integrates all statistical efforts regarding the economy. W. Leontief's input-output scheme, which captures the dynamics of the economy, also belongs in a course on socio-economic statistics. These two separate areas belong and ought to be discussed in courses and textbooks of statistics. The interpretation and prediction of regional, mostly non-experimental socio-economic data requires the re-thinking of their foundation. Just as economic and social phenomena are the point of departure and the final destination of any statistical enterprise, so also must the theory cover that entire process from beginning to end. This much broader theoretical basis should cover both statistical description and statistical inference, keeping in mind, however, that every statistical effort requires interpretation, but not necessarily inference. Such a broadened theoretical foundation should be capable of sharing its concerns with epistemology, sociology, geography, economic history, the science of management, accounting, social ethics, and of course, with economics. The calculus of probability, though, will be less prominent. Only little of what Leonard J. Savage had to say will be of use as a foundation for the theory of socio-economic statistics.[50]

Electronic computers, with their ever-increasing capacity for storing numbers, text and formulas, free statisticians from burdensome sorting and computing, indeed from the drudgery and tedium of what constituted the bulk of their work. This was reflected in the expression 'Tabellenknechte' (slaves of tabulations), coined to describe statisticians' work before the arrival of computers. These should allow statisticians more time to think about the meaning of their results unless they allow the complexities of computer technology to take the place of the drudgery from which they have been recently liberated.

There is also another danger rooted in the ease with which readily available canned statistical procedures and models can be accessed. The F, t, chi- square, and other statistical tests, often routinely and inappropriately applied, can create the illusion that useful, even scientific analysis has been accomplished. Yet, too often the appropriate conditions for using these tests are not given, and fail to help to understand the socio-economic situation. Computers, however, can be very useful in the meaningful interpretation of socio-economic data by aggregation/dis-aggregation, which is discussed in greater depth in subsequent chapters.

1.6.4 Toward a De-centralized Understanding of Data

The envisioned foundation of descriptive statistics requires a different attitude toward data about business, the economy and society: neither as the highly accurate measurements of natural science phenomena, in which the historic time and geographic place of the measurement is of minor importance, nor as random variables and random samples. On the contrary, in socio-economic data, their location, place in a historic context, and geographic region are of major interest, in realistically portraying these spatial-historical-institutional socio-economic phenomena (to be discussed in the next Chapter). This requires a very different approach to socio-economic statistical data[51] than the present understanding that treats them as abstract mathematical quantities. As a consequence of this mis-understanding, essential areas have been excluded that really belong to socio-economic statistics.

The assumption that data are only random deviations from some 'true value' is a carry-over from the thinking developed in the natural sciences. For example, the scatter of data in a regression diagram is typically considered a deviation from that center represented by the mathematically-determined regression line. The least-squares regression or trend line is held to be a valid approximation of the natural laws presumably underlying the behavior of chemical or physical processes. When dis-aggregating a socio-economic data set, however, the data in the sub-aggregates usually have regression lines with different parameters than the data in their aggregate. This indicates that there is no counterpart in society to the laws that govern physical phenomena, a matter that is further discussed in Chap. 9.

American and other societies experience the pull toward greater economic and political autonomy and decentralization,[52] while at the same time different forces work in the opposite direction, toward greater concentration. The present reduction in the functions and powers of Federal Agencies in the United States are a testimony to this trend toward decentralization The principle of subsidiarity recognizes the greater importance to citizens of what goes on in their immediate neighborhood and in the local district vis-à-vis matters affecting the country or the world as a whole. In statistical data about society an analogous situation should be expected. Averages and other values of centrality and trend values, representing those central values in society, lose their present preponderance that statistics has adopted from the natural sciences. In short, socio-economic data should be recognized as pieces of statistical evidence in their own right, not as deviations from some central value or trend.

This view of socio-economic data as not having a natural, necessary center from which they randomly deviate, is an important feature to be taken into account when interpreting data. This matter is followed-up in the next chapters.[53] The thinking about socio-economic data ought to shift away from its present belief that they have a center relying on means, trends and the dispersion around them, toward an understanding of socio-economic data as amorphous structures that can be aggregated or de-aggregated by subject categories, regions and time periods, without having such a center.

In conclusion, it appears that there were two different approaches to socio-economic statistics, one characterized as logisch-sachlich (logical-factual, subject-matter-oriented), the other as mathematical (probabilistic). This distinction overlaps to a great extent with another distinction between descriptive versus inferential statistics as well as, to some extent, with the dichotomy in social-science-statistics and natural-science-statistics.

Notes

1. The remarks in 'The Images that have shaped Accounting Theory' are quite pertinent also for statistics. This becomes more evident when one substitutes in the following passage the words 'accountants,' 'accounting' and 'accounting theory' with 'statisticians,' 'socio-economic statistics' and 'theory of economic and social statistics.'

   "The way accountants shape and understand the world of organization and management is influenced by the images which they bring to their subject of investigation. . .social scientists . . .developmental constructs within which they. . .make sense of the. . .ambiguous experiences. The image of the subject. . .shapes the way in which reality is understood. . .(and). . .generates insights. . . and thereby provides a basis for. . .research. . .we explore the images which have shaped accounting theory. . .how contemporary debates. . . revolve around competing images. . .the image usually offers no more than one particular. . .. insight. (p. 307). The idea that reality can be defined through numbers defines the basic paradigm of accounting and provides constraints upon accounting. It generates a way of seeing that takes precedence over other ways of seeing. . .**Only what is quantifiable**. . .**in numbers is objective and real**. . .Four principal images have shaped. . . financial accounting. They are those which treat accounting as a historical record, as descriptor of current economic reality, as an information system, and as a good. (p. 308). An appreciation of the nature of these images allows us to see how many controversial issues in accounting surface and resurface. . .and how in appreciating that they involve the advocacy of competing images we might use the insight provided. . .to gain . . .understanding of the issues being discussed. (p. 309) Accounting theorists and practitioners. . .cannot. . .fix attention on phenomena without any prior suppositions. These. . .are powerful in shaping our understanding of reality, and. . .have influenced. . ..the direction of accounting theory and practice. . .The implicit . . .assumptions. . .resulted in controversies. . .which are in fact the conflicts between competing images."

   Davis, Stanley W., Krishnagopal Menon, Gareth Morgan, in *Accounting Organizations and Society,* Vol. 7, Dec. 1982 pp. 307–318.
2. The direction of recent developments was expressed symptomatically, when the Institute of Mathematical Statistics decided to change the name of its Journal from *The Annals of Mathematical Statistics* to *Annals of Statistics*' Such implicit equating of mathematical statistical theory with general statistical theory is tantamount to declaring the latter as a branch of mathematics, largely at the exclusion of descriptive socio-economic data and methods.
3. A contemporary observer of the scene remarks sarcastically: "The material about business behavior that students read about in economics textbooks and almost all of the new theoretical material developed by mainstream professionals and published in the professions leading journals **was composed by economists who sat** down **in some comfortable chair and**. . . **simply made it up**. . ." p. 1.

   Barbara R. Bergmann, Prof. Emerita, U Md. In "Needed: A New Empiricism" *The Berkeley Electronic Press*, Joseph Stiglitz and Aaron Edlin editors, mm-9788-12157875@bepress.com, 2006
4. Haavelmo, Trygve, "The Probability Approach in Econometrics" *Supplement to Econometrica*, Vol. 12, July 1944, pp. 1–118 and pp. I–VI, University of Chicago.

5. The following quotation is but one of the innumerable examples: "The statistical methodology for analyzing data. . .is called statistical inference. . . The logical foundation of statistical inference is the mathematical theory of probability" (p. 75).
   Neter, John, Wasserman, William, and Whitmore, G.A. *Applied Statistics*, Allyn & Bacon, Boston, 1978.
6. "The data may be obtained from published sources, through survey research or by designed experiment. However obtained, the **data are the observed outcomes** or responses **of some phenomenon of interest or underlying random variable**." (p. 3).
   Berenson, Mark, Levine, David M, Goldstein, Matthew, *Intermediate Statistical Methods and Applications*, Englewood Cliffs, New Jersey, Prentice-Hall, Inc. 1983.
7. ". . .the disturbance term u(i) may be used as a substitute for all the excluded or omitted variables from the model. . . the joint influence of these variables may be. . . non-systematic or random. **Hopefully, their combined effect can be treated as a random variable** u(i). . ." p. 27.
   Gujarati, Damodar, *Basic Econometrics*, McGraw-Hill, New York, 1978, p. 27. (letters bolded for emphasis)
8. "The justification for the inference derives almost entirely from the assumption of a certain model, sometimes narrowly specified as a super-population with known shape. . ." (p. 80, and pp. 108–111, 121)
   Cassel, C.M., Särndal, C.E., Wretman, J.H., *Foundations of Inference in Survey Sampling*, New York, John Wiley & Sons, Inc. 1977.
9. Wallis uses large sample theory for tests of significance – as do numerous other econometricians – for the time series data of hog-corn price ratios and the annual series of commercially slaughtered hogs, in million heads, 1935–1971 as if these yearly totals were selected by a random sampling procedure from a population of such yearly totals. Those values selected – the actual time series before the analyst – are assumed to constitute a random selection from among a very large number of such yearly time series values, an absurd rationalization.
   Wallis, Kenneth F., "Multiple time series analysis and the final form of econometric models," *Econometrica,* Vol. 45, No.6, Sept. 1977, pp. 1487–1490.
10. Parzen, Emmanuel, "An approach to empirical time series analysis", *Time Series Analysis Papers*, San Francisco, Holden Day, 1967, pp. 551–565.
11. Hatanaka, Mitsuo Suzuki "A Theory of the Pseudo spectrum and its Application to non-stationary Dynamic Econometric Models", Chapter 26 in *Essays in Mathematical Economics in Honor of Oskar Morgenstern* edited by Martin Shubik, Princeton, N.J., Princeton University Press, 1967, esp. p. 444.
12. This impression has been confirmed independently by the Austrian mathematical statistician Adolf Adam:

    "Die derzeit dominierende Schule der mathematischen Statistic umgibt die statistischen Basisinformationen mit einem stochastischen Schleier, indem sie behauptet, dass jedes statistische Kollektive eine 'Zufalls-Stichprobe' aus einer fiktiven verteilungsstabilen und unbeschränkten Grundgesamtheit sei (ein Paradigma, "das formalwissenschaftlich sehr ergiebig, sachwissenschaftlich aber höchst problematisch ist)."

    Translated: "The presently prevailing school of mathematical statistics surrounds all basic statistical information with a stochastic veil, by insisting, that every statistical collective is a 'random sample' from a fictitious, unlimited base collective, the distribution of which is stable (a paradigm which is most fruitful from a formal-scientific point of view, but most problematical from the point of view of a subject matter science.)" Adolf Adam, "Grundriss einer Statistischen Systemtheorie" in: *Festschrift für Wilhelm Winkler* Wirtschaftsverlag Dr. Anton Orac, Wien 1984, p. 38.
13. Morgenstern, Oskar, *On the Accuracy of Economic Observations*, Princeton, N.J., Princeton University Press, Second ed., 1963. Also: Winkler, Othmar W. "Determining Classes in Frequency Distributions of Economic Data", *Proceedings of the Business and Economic Statistics Section of ASA*, Washington, D.C. 1983, pp. 487–492. esp. p. 488.

14. *Statistische Praxis – Zeitschrift für Rechnungsführung und Statistik* (– journal for industrial record keeping, accounting, and statistics) Zentralamt für Statistik, Berlin, Deutsche Demokratische Republik (DDR).
15. See e.g. the textbooks by G. Forbrig against which none of the objections hold which must be made against the textbooks which follow the Anglo-American understanding of the theory of statistics as applied to business and economics. Forbrig, Gotthard, *Grundriss der Industrieestatistik*, Vol. I, Verlag die Wirtschaft, Berlin (DDR) 1965. Vol. II, with Rumen Janakieff, Verlag die Wirtschaft, Berlin (DDR), 1967 and Forbrig, Gotthard, Brosch, Otmar, Wolff, Ursula *Betriebsstatistik* Verlag die Wirtschaft, Berlin (DDR), 1983 Also: Winkler, O.W.: "Unterschiedliche Ansätze zur Wirtschafts- und Sozialstatistik in Ost und West" *Jahrbücher für Nationalökonomie und Statistik,* Band (Vol.) 208/5, pp. 459–492, Gustav Fischer Verlag, Stuttgart, Germany, Sept. 1991.
16. The following passage is revealing, confirming the earlier observation:
    "Modern statisticians equalized probability and random event as if dividing objective occurrence into two types of phenomena, one being certainty, the other being uncertainty. . .. In so far as the view of Soviet school of "Statistical Theory" which divides sciences relating to the objective phenomena of the universe, according to whether or not it has any class nature, into natural sciences (mechanics, physics, chemistry, biology) and social sciences (sociology, economics, political science); and put the two opposing each other, holding **that random events can only appear in natural phenomena, there being no or very little random event in social phenomenon**." (p. 7)
    Dai Shiguang, Professor of Statistics, Chinese People's University, Beijing, China, in: *Dialectical Materialism, The Guiding Ideology of Applied Statistics*, Oikon Publishers, Hong Kong, 1984.
17. Without intending to do so, the author of a well known American textbook, presents an example of the inadequacy of stochastic decision criteria. It deals with a unique decision concerning the switch to a new package for the principal product of a firm, not with a routine, repeatable situation. This did not prevent the author from treating this historically unique situation in the life of this firm as a continuing game of chance, despite the fact that the only two alternative outcomes were the threat of a heavy loss, versus the lure of a great gain. With this one example the author demonstrated Bayesian and non-Bayesian decision techniques through two entire chapters (Chapters 15 and 16) of his textbook, with expected long run monetary outcomes and utilities as the decision criteria.
    Mansfield, Edwin, *Statistics for Business and Economics, Methods and Applications*, second ed., New York, W.W. Norton & Co. 1983.
18. "There was almost invariably a confusing identification of mathematical probability with probability in real life. . . (p. 203) a considerable part of Annals of Mathematical Statistics has been mathematics that has had only a tenuous connection with applied statistics. . ." (p. 204).
    Doob, J.L. "Foundations of probability theory and its influence on the theory of statistics" in: D.B. Owen, *On the History of Statistics and Probability*, Statistical Textbooks and Monographs, Vol. 17, Marcel Dekker, Inc. New York, 1976.
19. "The theory of decision functions helps very little in the decisions of ordinary life and it is a mistake to claim that it does except in the relatively few cases where the pay-off matrix has a genuine reality. The theory of games has proved to be disappointing and it is a fair question to ask whether the player of any non-trivial game has ever been able to improve his play by working through it." (p. 52).
    Kendall, Maurice, "Statisticians – Production and Consumption", Lecture invited by the President of the American Statistical Association, *The American Statistician*, Vol. 30–2, Washington, D.C., May 1976, p. 52.
20. Villers, Raymond, *Dynamic Management in Industry*, Prentice Hall, Englewood Cliffs, N.J. 1960, pp. 139–147.
21. Sarepta Paper Co. (C) Abridged Case, #9-678-166, Harvard University Business School, case studies, Soldiers Field, Boston, 1978.

22. Simon, Herbert A., *Models of Thought*, especially Chapters 4.2 "Trial and Error Search in Solving difficult Problems", and 5.5 "Problem Solving and Rule Induction" New Haven, Yale University Press, 1979.
23. *Mead Corporation Abridged Case, # 9-678-165*, Harvard University Business School, case studies, Soldiers Field, Boston, 1978.
24. Schonberger, Richard, *Japanese Manufacturing Techniques – Nine hidden Lessons in Simplicity*, New York, The Free Press, 1982.
25. "While it is true that the good companies have superb analytical skills, we believe that their major decisions are shaped more by their values than by their dexterity with numbers." (p. 51) Peters, Thomas J., Waterman, Jr. Robert H., 'In Search of Excellence,' New York, Harper & Row Publ., 1982.
26. Williamson, O. Chapter 2 "The Organizational Failures Framework", in: *Markets and Hierarchies-Analysis and Antitrust Implications-29.6 A Study in the Economics of Internal Organization*. Free Press, New York, 1975.
27. Winkler, Othmar, W., "On the Nature of Statistical Information in Business and Economics", *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C., 1964, pp. 64–74.
28. In the "Bochumer Betriebsräte Befragung" the questionnaire, directed at 2,171 mid-sized metalworking enterprises was to 'measure', or more accurately, to explore the relationship between its Betriebsrat (the labor representation) and the firm's management. The researchers used the following questionnaire with five spelled-out questions, to be selected by management to describe the kind of relationship that management believes exists in that enterprise. The following five pre-printed questions to be answered by management, selecting that answer that best represents the situation in managements opinion. This is a typical example of 'measurement' in the social sciences, a far cry from measurement in the natural sciences.

    Q1.: "Die meisten technischen oder organisatorischen Veränderungen muessen gegen den Betriebsrat durchgesetzt werden"
    Q2.: "Manchmal ist es schwierig dem Betriebsrat die gemeinsamen Betriebs-und Belegschaftsin teressen zu vermitteln"
    Q3. "Technische oder organisatorische Veränderungen werden vom Betriebsrat uneingeschränkt unterstuetzt"
    Q4.: "Der Betriebsrat betrachtet technische oder organisatorische Veränderungen nicht als sein Aufgabenfeld und beteiligt sich nicht"
    Q5.: "Der Betriebsrat wird an solchen Veränderungen nicht beteiligt" From "Kooperation zwischen Betriebsrat und Management" Alexander Dilger, *Jahrbücher fuer Nationalökonomie und Statistik*, Band 276, Heft 5, Sept.2006 p. 565/566.

29. A stunning example of 'a phenomenon being too large to be perceived' impressed the author during a performance of Charles Dickens' 'A Christmas Carol' at the Ford Theater in 1983, in Washington, D.C. The phenomenon in question was the Ghost of X-masses to come who appeared on the dimly lit stage as a huge, shadowy figure, obviously on stilts, enveloped in a flowing, enormous black cape. That shadow of a figure seemed to fill the entire stage, dwarfing Scrooge who seemed oblivious of that ghost's presence. But so did most of the audience who did not notice the ghost's presence until later during this act, as could be noted by the public's ooohs and aaahs, and eventual reaction to its discovery, dramatically illustrating that there can exist things which are too big or too dispersed to be readily perceived.
30. I proposed this expression to visualize the descriptive function of statistics. A colleague subsequently informed me that this expression had been invented earlier by the French biologist, Joel De Rosnay' who applied it to a different area but with the same intent, published as *Le Macroscope – Vers une vision globale* Editions du Seuil 1975, rue Jacob, Paris 6. In this remarkable book that author expressed, as 'macroscope', the concept of an instrument that permits to view jointly ecology and all related physical and economic processes. He does not mention socio-economic statistics, though, for which his concept is perfectly suited.

31. Winkler, Othmar W., op.cit. 1964, pp. 68,69.
32. Firebaugh, Glenn, "Assessing Group Effects – A Comparison of two Methods" in: Edgar F. Borgatta, David J. Jackson, editors *Aggregate Data – Analysis and Interpretation*, Sage Publications, Beverly Hills, 1980.
33. "...we need to specify the probability distribution of...u(i)... which is random by assumption... since the probability distribution of these estimators are necessary to draw inferences about their population values...the void can be filled if we are willing to assume that the u(i)'s follow some probability distribution...in the regression context it is usually assumed that the u(i)'s follow the normal distribution..."
Gujarati, Damodar, op. cit. p. 71, see footnote #10.
34. Winkler, Othmar W. "A Critical View of Time Series Analysis" *Proceedings of the Business and Economics Statistics Section of ASA*, American Statistical Association, Washington, D.C. 1966, pp. 352–370.
35. Ehrenberg, A.C.S. *Data Reduction*, London, John Wiley & Sons, 1975.
36. Winkler, Othmar W., "Determining Classes in Frequency Distributions of Economic Data" *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1983, p. 489.
37. Winkler, Othmar W. op. cit. 1983, p. 487.
38. See e.g. Silk, John, *Statistical Concepts in Geography* London, Allen & Unwin 1979, esp. pp. 22–26 and 206–249.
39. "At the undergraduate level, the image is that statistics is difficult and irrelevant.... math was rated as more pleasant, more relevant, and less difficult than statistics. A distressing aspect...was that the course...substantially reduced students' fear of the subject, but did little to change their notions of its relevance"
Jordan, Eleanor W., Stroup, Donna F. "The Image of Statistics", *Collegiate News and Views*, Vol. XXXVII, No. 3, Spring 1984, pp. 11–13.
40. American sociologists claim that new research and insights rarely come about through the contact of the professor with his students, but that research topics are determined by a professional elite, whose consent or rejection determines the development of the field. Jencks, Christopher, Riesman, David, "The Art of Teaching," in: Anderson, Charles H., Murray, John D., editors *The Professor*s Cambridge, Massachusetts, Schenkman Publ. Co., Inc. 1971.
41. Briefs, Henry W. *Three Views of Method in Economics*, especially Chapter 5, "Econometrics" pp. 60–72 Washington, D.C., Georgetown University Press, 1960.
42. Arrow, Kenneth J. "Classificatory Notes on the Production and Transmission of Technological Knowledge" in *American Economic Review*, May 1969, pp. 29–35.
43. "When we turn from physiology to culture and ask why America, Mexico, and the Caribbean have more crime than most of Europe and Asia, the obstacles to intellectual progress look to me even more formidable than when we try to understand the effects of genes. The record of the past generation is also less encouraging. We are not, I think, any closer to understanding why cultures differ from one another, or why they change over time, than we were thirty years ago. Worse yet, young social scientists are seldom interested in such questions. Most of them seem to prefer mathematical games, which are absorbing and fun to play but have almost no chance of telling us anything useful about anything..."
Christopher Jencks, "Genes and Crime," *NY Review of Books*, XXXIV, 2 (12 Feb '87), pp. 33–41.
44. Wilhelm Winkler, Grundriss der Statistik II Gesellschaftsstatistik, 2nd revised edition, Wien 1948, Manzsche Verlags und Universitaets Buchhandlung.
45. Lawrence Klein, reproducing another author's monthly time series data in his textbook treats these as if these monthly data were the elements of a randomly selected sample!
Klein, Lawrence *Introduction to Econometrics* Englewood Cliffs, Prentice Hall, 1962 pp. 33–48, esp. 33, 34.
46. "It would be interesting to know how much of what we teach as statistics is really needed by statisticians.... more importantly it would be interesting to know what we do not teach that a

statistician ought to know . . . one cannot help feeling that we are turning lose on the world a number of insufficiently rounded individuals."
Kendall, Maurice "Statisticians – Production and Consumption" op.cit.1976, p. 53

47. Typical for the prevailing attitude are the remarks Richard Savage's as the discussant at a meeting in which statisticians of renown expressed their conviction that non-sampling errors were far more important for survey results than sampling errors. "Survey operations. . .call. . .for the measurement of the differences between what is done and what is ideal. . .the evidence is not strong that a systematic, comprehensive effort is needed. . . a casual reading of the papers has not given me a good idea of the nature of the camel, although Tore Dalenius assures us the gnat is sampling error. . ." p. 45.
Savage, Richard (I.R.), Discussion of paper "Sampling and Non-sampling Errors" by Tore Dalenius, C.A.Brooks, Barbara Bailar, L. Madow, in: *Proceedings of the Social Statistics Section of ASA* Part I, Washington, D.C. 1977, p. 45

Professor Bruckmann, in a different part of the world, in a different context, elaborates on this attitude:

"So ist etwa vom Standpunkt der Anwendung aus der systematische Fehler ungleich bedeutsamer als der Zufallsfehler. Spricht man aber einen mathematischen Statistiker auf den systematischen Fehler an, so reagiert er mit Verständnislosigkeit und Gleichgültigkeit; da der Zufallsfehler berechenbar ist, der systematische Fehler jedoch nicht, kommt dieser in seinem Gedankengebäude gar nicht vor."

Translation of Professor Bruckmann's statement:

"From the point of view of application, the systematic or non-sampling error is incomparably more important than the sampling error. If, however, one addresses a mathematical statistician about the systematic error, he will react with incomprehension and indifference (!); because the sampling (random) error can be computed, while the systematic error cannot be computed, he does not even give any thought to the latter (. . . has no place in his mental construct.)"

Gerhart Bruckmann, "Statistik und Prognostik" in *'Festschrift für Wilhelm Winkler*, *Schriftenreihe der Osterreichischen Statistischen Gesellschaft*, Band 1, Adolf Adam ed., Wirtschaftsverlag Dr.Anton Orac, Wien 1984, p. 54.

Attention should be drawn also to the growing awareness among statisticians of non-sampling errors. Witness to this e.g. the International Symposium on Panel Surveys, scheduled by ASA in Washington, D.C., November 1986 which dealt explicitly with "Non-sampling Error Issues" in nine sessions, each lasting half a day. Reported in: Amstat News, Washington, D.C. May 1986, p. 11, 12.

48. It is interesting to note that in the sciences and in engineering the important role of statistics as an instrument of perception of its phenomena is just being rediscovered under the name of 'exploratory data analysis' (EDA). It has been contrasted with the now conventional, inductive role of statistics, called 'confirmatory analysis.' Since the appearance of J.Tukey's book a considerable literature has come into existence under the name of EDA. Tukey, John *Exploratory Data Analysis (EDA)*, Reading, Massachusetts, Addison & Wesley, 1976.
49. Johnston, Denis F. *Social Indicators and Social Reporting in the United States*, Unesco, Paris, SS/C/44/82/2, 1982, Analytical and Methodological Studies.
50. Especially his statement that: "Statistics proper can perhaps be defined as the art of dealing with vagueness and with interpersonal differences in decision situations. . ." (Chapter 8) Savage, Leonard J., op. cit.
51. "Universities certainly ought to be critical toward everything new. Yet they are not just "conservatories" – institutions in which existing knowledge is preserved for posterity like in a museum – which are allowed to act as if we already knew everything that is essential, and one has only to fill in minor details that have been missing. That scholars exclude themselves from new "essentialities," that is exactly what is unbearable in today's state of affairs. Nobody wants to Re-learn, to revamp ones idea of the world (Weltbild)."
Driesch, Hans. 'Parapsychologie', Rascher Verlag, Zurich, 1943, p. 6 (this quotation is a translation from the German text).

52. Naisbitt, John Chapter 5, "From Centralization to Decentralization" in: 'Megatrends', New York, Warner Books, Inc. 1982, pp. 97–129.
53. To many scientists. . .facts are things that simply are the case: they are discovered through properly passive observation of natural reality. To such views Fleck replies that facts are invented, not discovered. . .the appearance of scientific facts as discovered things is itself a social construction: a made thing. . ..at the cognitive level this is manifested in the notion of the 'thought style' (Denkstil) and at the social level through the entity that expresses and maintains it, the 'thought collective' (Denkkollektiv). . .Fleck asserts the essentially social nature of thought. . .. cognition 'is the most socially-conditioned activity of man, and knowledge is the paramount social creation.' All cognition and perception occur within a particular thought style. . .as 'a readiness for direct perception'. . .The thought style constrains the ways in which it is possible to perceive reality. One is inducted into a thought style through an "initiation rite;" within it the shape of answers is preformed in the nature of permissible questions. . .The development of any scientific fact is an artifact of the social processes of communication. . .The passive component of knowledge is what we call "existence" or "reality"; it seems independent of us and compels us to accept it. . .This passive feature will itself be a function of the assumptions and goals adopted by a particular thought collective and is.explicable in terms of cultural history. What appears. . .a function of "reality" in one set of circumstances may appear actively constructed in another. . .Whatever is known has always seemed indubitable to the knower; equally, whatever is alien has seemed fanciful and arbitrary.

    Steven Shapin, a review of the book Genesis and Development of a Scientific Fact written under the title, *A View of Scientific Thought* , by Ludwik Fleck, translated from the German edition by F. Bradley and T.J. Trenn, R.K. Merton, T.J. Trenn editors. University of Chicago Press, 1979. Science, Vol. 207, March 7 1980 pp. 1065–6, T.J. Trenn, R.K. Merton, and T.J. Trenn editors. University of Chicago Press, 1979.

Chapter 2
From the Facts in Society to Socio-Economic Data

2.1 Socio-Economic Phenomena – The Starting and End Point of Statistics

The intent of this chapter is to clarify the nature of socio-economic statistical data, and the role statistics is playing in capturing socio-economic phenomena. This role has been seldom discussed but is a fundamental issue concerning the nature of socio-economic statistical data,[1] and the manner in which they convey socio-economic reality. The following discourse may strike some readers as unnecessary, perhaps as not even belonging to statistics. Yet, a good understanding of this preliminary phase should provide the user of statistical data with an understanding of the data-creation process as an important first step of interpretation.

2.1.1 Flaws in the Perception of Socio-Economic Reality

To properly interpret data, an understanding of the nature of the elementary building blocks,[2] the 'statistical-counting-units' and their role in portraying economic phenomena, is needed. A comparison suggests itself with the role that atoms and molecules are believed to play in the physical world. The 'statistical-counting-units' could be thought of as the equivalents of the atoms in physics. The summation of these statistical-counting-units in statistical aggregates could be compared to molecules that are made up of such atoms. These molecules then make up the substance of objects, which then are somehow comparable to phenomena in the social sciences. Despite the appearance of simplicity and mathematical precision of statistical data presenting socio-economic phenomena, like 'price level,' 'unemployment,' or the GDP, these phenomena and the data portraying them, are more ambivalent and elusive than is commonly realized.

2.1.1.1 The Socio-Economic Phenomena

Let me start with the beginning of any statistical investigation: defining the phenomenon to be studied, what it is, where and when it can be found, and how

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_2, © Springer-Verlag Berlin Heidelberg 2009

it should be captured statistically. To repeat the obvious, the phenomena in society are quite different from phenomena in the natural sciences. They also differ in the manner in which 'real-life-objects' project the socio-economic phenomena.[3] The temperature at which water reaches the boiling point, for example, should be expected to be the same in socialist China as in capitalist USA, in a stone age community in Australia's outback as in a futuristic community in California. Aside from the influence of the barometric pressure – depending on the altitude above sea level – the boiling point of water was probably the same during the time of the French Revolution as during the Punic Wars of ancient Rome. Minor changes may have occurred in reaction to changes in our solar system and in the galaxy to which it belongs. It seems unlikely that a research grant would be available for studying differences in the boiling points of water between cultures, in different continents, or in different historical epochs. Compare this with research in the social sciences where the opposite assumption applies: nothing should be expected to remain the same from one social stratum to another, from one country or culture to another, or even from one month to the next. Social phenomena are known for their rapid change, their unpredictable evolution and their great variety. Statistical data must keep up with this dynamism, and statistical theory ought to be prepared to interpret the phenomena that underlie those data. It should not be a surprise that statisticians have been uncomfortable approaching this topic. They seem to consider a discussion of economic and social phenomena as lying outside the purview of statistics.[4] Yet, a foothold in this foreign area must be obtained.

It appears that socio-economic phenomena can be abstracted from actual situations of society on at least three levels.

1. At the most abstract level, one might consider phenomena such as The Business Firm, The Production Plant or New Venture Creation.[5] At such a high level of abstraction a general theory of the firm might be derived from the study of existing firms, regardless of country, culture, stage of economic development, type of product, state of technology, phase in the business cycle, etc. An analogue abstraction in the natural sciences could be the abstract phenomenon 'Tree' derived from the most diverse forms of life without regard to species, type of wood, fruits, leaves, height, shape, climate, location and ecosystem in which it exists.
2. At a less abstract level one can view the same business firms, production plants or New Venture Creation as parts of an economic system, focusing on their interaction with other entities of their kind as sellers and purchasers, still largely disregarding regional and period-specific circumstances, except for the implicit assumption of a free, western-style market society. This less abstracted phenomenon might roughly be identified as an 'Industry,' and may be as different from the socio-economic phenomenon 'Business Firm' as the natural science phenomenon 'Pine Forest' is from 'Tree'.
3. At an even less abstract level these same Business Firms represent even more concrete phenomena in socio-economic situations in which the location, state

of development at a given time, type of products it deals with, and many other particulars that define such a business firm are not abstracted and assumed away to the same extent. Possible examples might be the Steel industry in Sweden at the turn of the century, or the British leather industry in the decade following World War II.[6] These phenomena are as different from the previous phenomena as the phenomenon 'Washington National Forest of West Virginia in the 1980s' is from the more abstract, general phenomenon 'Pine Forest.'

Each one of these three levels of phenomena is embodied or projected by the same factories, to stay with the example of a production plant, but at different levels of abstraction. One might say that the natural sciences, and imitating them, (micro) economic theory, mostly deals with phenomena at levels of abstraction 1 and 2, econometrics with levels 2 and 3, but socio-economic statistics mostly with levels 3 or higher numbered levels not listed here, of even less abstract, more concrete phenomena.

A typical case, illustrating the need for statistics to clarify a social and cultural phenomenon, could be poverty, before one can even consider collecting data and planning future tabulation of results.[7] It should also be noted that the need to clarify such social phenomena as 'business firm' 'unemployment' or 'work accident' has led to the creation of national and international agencies like the Bureau of Labor Statistics of the US Government's department of Labor (e.g. for NAICS, the North American Industrial Classification) (ILO) the International Labor Organization and (ISI) the International Statistical Institute.

2.2 The 'Projecting Agents' of Socio-Economic Phenomena

In sociology, economics, management, and other business areas, specific socio-economic phenomena are portrayed or projected by specific items, events, buildings and all kinds of things such as e.g. cars and in general, 'durable consumer-goods.' These 'projecting agents' can also be contractual documents that seem to exist only as a piece of paper but are anchored in the laws and customs of society. All of these will be referred to in the following as 'real-life-objects.'

Socio-economic phenomena, at all levels of abstraction, are projected by appropriate 'real-life-objects' as the 'projecting agents', somewhat like the invisible field of a magnet is projected by iron filings scattered on a sheet of paper placed on top of that magnet. The iron particles become projecting agents of the phenomenon 'magnetism' by their reaction to these polarizing forces that exert an effect on these particles. Quetelet's example of a circle drawn with chalk on a blackboard comes to mind although he intended to illustrate with it the 'Law of Large Numbers.' When looking through a magnifying glass, he relates, the individual chalk particles can be seen, spread randomly over the rough surface of the blackboard. When looking at all those particles together, however, the shape of their array in a circle, which in this instance is the phenomenon, becomes evident.[8]

After the appropriate branches of the social sciences have defined a social or economic phenomenon to be investigated, it is the task of statistics to identify, locate and record those 'real-life-objects' that portray that phenomenon.

2.2.1 Different Types of 'Real-Life-Objects'

Understanding those 'real-life-objects' is a first step of data interpretation. A great variety of such 'real-life-objects'[9] exists, that act as projecting agents[10] for socio-economic phenomena. Human beings are the most important of the great variety of 'real-life-objects' that are of interest to society – no offense is meant when referring to human beings as 'real-life-objects' as a technical-statistical term. It can be an individual person, or a group of persons, like a 'family', a 'household' or other groups of people, e.g. in a mental institution, in hospitals, jails, or retirement homes.

These 'real-life-objects' can also be things related to socio-economic activities, such as mines, farms, retail establishments, production plants, railroad companies (with their rail network), corporations, but also machines, farm animals, and produced goods. Political-administrative districts can become 'real-life-objects', such as counties, metropolitan areas, census tracts, even plots of land cultivated with certain field crops. Other, quite different kinds of 'real-life-objects' can be legal documents like shares, mortgages, vehicle registrations, birth certificates, building permits and bonds.

The most frequent kind of 'real-life-objects', however, are neither people nor buildings or things. They are occurrences of social relevance, such as sales, strikes, accidents. Into this category of 'real-life-objects' belong events that are *beginnings* e.g. the birth of a person, foundation of a firm, issuance of a share, the issue of a construction permit or the creation of a new job, *changes* e.g. in the occupation of a person or in the line of production of a firm, and *terminations* e.g. the withdrawal of a person from the labor force or the conclusion of a debt through full payment, the completion of the construction of a dwelling unit or the bankruptcy filed by a business firm. These occurrences can become the 'real-life-objects' of interest, independently of the persons, things or events in which they occur. These beginnings, changes and endings are of interest independently of the 'real-life-object' in which they occur, though always in relation to it, whereby the description of the 'real-life-object' in (or on) which an occurrence takes place becomes one of its characteristics. An example would be the opening of a new supermarket, where the 'real-life-object,' the beginning of a firm, is characterized by the size and kind of business in which it occurs.

2.2.2 Substance and Individuality of 'Real-Life-Objects'

These 'real-life-objects' differ widely regarding their physical substance. On one extreme are those that consist predominantly of a physical mass like lumber, coal,

gasoline, cement, fuels and raw materials. These are needed to project socio-economic phenomena such as importation, exportation, or as the input of certain raw materials in a production process. The problem with them is that they lack natural units that can be counted and measured.

On the other extreme are 'real-life-objects' that have only a symbolic substance: a mortgage, the piece of paper that represents that financial contract and is part of the important phenomenon 'long-term investment.' Occurrences usually have only a minimal physical substance: a birth certificate or a marriage license. Some occurrences have no physical substance at all such as a business transaction in which merchandise and money is exchanged informally, without a written record – the substance of the traded merchandise must not be confounded with the substance of the transaction itself, which is the 'real-life-object' properly speaking from a statistical point of view. Such lack of a physical substance in 'real-life-objects' causes the problem of under-reporting because of the difficulty in locating and recording them.

A different, though related matter, is the individuality of these 'real-life-objects'. It refers to their appearance as something clearly distinct from their environment and from other 'real-life-objects'. A 'real-life-object' may consist of one single piece or unit, such as a car. At times a 'real-life-object' may consist of various individual pieces, each of which could become a 'real-life-object' in its own right. A 'Corporation,' for example is a 'real-life-object' of one kind. Its various retail establishments or production plants can become separate 'real-life-objects' in which case they represent a different kind of economic phenomenon.

The delimitation of the individuality of an object often suggests itself naturally, such as in a motor vehicle, farm animals, or fruit trees.[11] This is not the case in a variety of socio-economic 'real-life-objects' whose individuality must be defined by the social scientist, such as e.g. a business firm, an I.O.U., a work-accident or a strike. Raw materials, many semi-finished products, and fuels present problems in this regard. Bulk products like cement, cotton, chemicals, lumber, oil, coal, electricity or gas do not have naturally individualized pieces that one might use as 'real-life-objects.'

Other materials do have individualized pieces, but the exact determination of their number and characteristics is not worth the trouble, such as metal screws, nails, apples, bricks, pencils or cigarettes to give a few examples. In such instances the weight, length, surface or volume of their physical bulk is substituted, such as tons, bushels, board feet, KWH, or certain forms of packaging, such as barrels (oil), sacks (potatoes), crates, bales, or even the 'production of the day.' These are not truly individualized objects but pseudo-objects. The number representing the measure of their weight or volume are scale units of measurement, not, as is sometimes mistakenly believed, individual objects. Such units-of-measurement, as stand-ins, are pseudo 'real-life-objects' that are treated as homogeneous, in contrast to individualized 'real-life-objects' that can be quite heterogeneous and require a correspondingly more sophisticated statistical approach.

2.2.3 Life Span and Timing of 'Real-life-Objects'

Every 'real-life-object' has a duration or life-span, no matter how short it may be. That life-span has a beginning, various phases of development, and an end. (e.g. see Fig. 7.1) No object really exists as just a point in time, even if for practical purposes it may be treated as such. Beginnings, changes and terminations themselves usually are complex occurrences. The establishment of a new business firm, for instance, may take months. It is a lengthy process which itself has a beginning, duration, and a termination. The onset of the beginning may be considered in even finer detail and further phases might be distinguished about it, such as a beginning e.g. the moment at which this beginning phase actually is initiated, a development of this early stage, and an ending, which is the point in time when this beginning stage is terminated. The possibility of such refinements has a certain importance for the precision with which real-life-objects can be recorded statistically, and to clarify some old problems in statistics like 'the index-number-problem'.[12]

The issue of when exactly a 'real-life-object' is captured statistically can be important. It allows to link-up each object with other 'real-life-objects' in a 'historic landscape'. This matter is important because statistical survey procedures tend to isolate 'real-life-objects' from their actual surroundings, thereby tending to ignore potentially important information about their socio-economic context. More about this will be discussed in Chap. 5, Longitudinal Analysis-Part 1 – Looking to the Past.

2.2.4 Location, Extension and Mobility of 'Real-Life-Objects'

Every 'real-life-object' has a definite relation to its location. Reference to it as the 'geographic characteristic' treats location as an intrinsic quality of an object, at a par with other characteristics. This assessment is inaccurate, however, and prevented statistical theory from dealing with the geographic dimension of socio-economic phenomena. Regional phenomena differ due to the special economic and environmental characteristics of each area, which are implied and summarily stated through a 'real-life-objects' geographic location. Even 'real-life-objects' with only a symbolic, minimal physical substance like the *sale* of a car or the *issuance* of a mortgage happen in a place on the map. The geographical location on which a sale takes place, though not an attribute of the 'real-life-object' 'sale' is, like the time at which it happened, important for grouping these objects into meaningful aggregates (more in Chap. 3).

Every object also has a geographic extension. A farm occupies a certain amount of land with certain surface and soil characteristics. So does a strike which takes place in some production plant. The plant's physical and geographic extension is usually also the geographic extension of that 'strike.'

Objects can be fixed or mobile with regard to their location. Most 'real-life-objects' are neither absolutely fixed, nor completely mobile. Even houses and large firms have been moved to different locations. It is the high mobility

of some 'real-life-objects' that creates problems for statistics. Examples are the whereabouts of the rolling stock of a trucking firm or of a railroad company. These problems create uncertainty, not unlike the measuring problems in atomic physics.

2.2.5 Attributes and Variables

These 'real-life-objects' project an economic phenomenon through their properties. The attributes – qualitative characteristics or non-measurable variables – of these real-life-objects describe pervasive, essential aspects of an object, through non-numeric, nominal description. They cannot be determined with accuracy or measured on an interval or ratio scale. Quantitative characteristics, on the other hand, expressing intensity or the magnitude of some feature, can be determined accurately, but contribute little to characterize the object.[13] Both kinds of determining the characteristics of a 'real-life-object' are needed as mutual complements.[14]

Every property which characterizes a 'real-life-object' may be understood as a partial description of its nature. Behind the customary distinction in qualitative characteristics (attributes) and quantitative characteristics (variables) really is another distinction, according to the width of the segment of the integral nature of the 'real-life-object' which is provided by a given characteristic. Qualitative characteristics capture in literary form essential and pervasive aspects of the 'real-life-object,' but cannot be determined succinctly. The wider that slice out of the nature of a 'real-life-object', a specific attribute, the less precisely can it be determined. The so-called quantitative characteristics, on the other hand, refer to narrow segments of the nature of the 'real-life-object' which can be determined more accurately. The narrower this segment, the **more precisely** it can be captured (measured), but **the less information** is obtained concerning that 'real-life-object'.

As a first approximation, a wide part of the nature of a 'real-life-object' is described through a qualitative characteristic. In consecutive, progressively finer determinations (descriptions) the nature of that initial segment of the 'real-life-object' is then further defined. At the end of such a wedge-like penetration into the nature of the 'real-life-object', quantitative, measurable characteristics can add the sharpness that was missing in the initial description by the attributes. The same holds for the tabulations made of such characteristics of the 'real-life-objects.'

When the 'real-life-object' is an occurrence, it is also characterized by the 'real-life-object' to which it belongs, or on which it is happening. The characteristics of non-individualized 'real-life-objects,' e.g. raw materials, are summarily estimated. From the socio-economic point of view they usually are of little interest – although they may be of interest e.g. from a quality-control, that is, engineering point-of-view.

To summarize, the *qualitative* description alone is imprecise, e.g. a firm described only by the nature of its products. The *quantitative* description alone has little meaning, e.g. a firm described only by the number of its employees, or the size of last month' sales, without an indication of its qualitative characteristics like the industry to which it belongs, the kind of products, form of ownership, capital structure, etc. The description of a 'real-life-object' by attributes does not need to be supplemented by quantitative characteristics – measurements – in order to be comprehensible.

The analysis by attributes is Basic, but the description by one or more quantitative variables alone is not meaningful. Quantitative Variables are only complementary.

These observations should alert the user of data to first clarify these issues by asking questions, using the answers as the first tool of a meaningful interpretation of data. This understanding also underlies the structure of this book, where Chaps. 3 and 4 discuss the qualitative nature of data, followed by a discussion of their development through time, in Chaps. 5, 6 and 7. The quantitative characteristics, usually treated at the outset, are discussed in this book in Chaps. 8 and 9, only after the statistical issues with qualitative characteristics.

2.3 From 'Real-Life-Object' to 'Statistical-Counting-Unit'

'Quod non est in acta, non est in mundo'
(What is not on record, does not exist – A basic tenet of Roman law.)

The printed socio-economic data do not directly deal with the 'real-life-objects' that were discussed, but with simplified statistical sketches of these, that I would like to call the 'statistical-counting-units.' It is these that are tabulated, not the 'real-life-objects' themselves. The user of statistical data knows only about those 'real-life-objects' of which questionnaires or computer accessible evidence – the 'statistical-counting-units' – exist. A clear distinction must be made between the 'real-life-objects' out there in reality, and the 'statistical-counting-units,' the sketches of these 'real-life-objects' in electronic or in other storable form. That seemingly subtle distinction, however, is important and must be kept in mind when interpreting socio-economic data (Fig. 2.1).

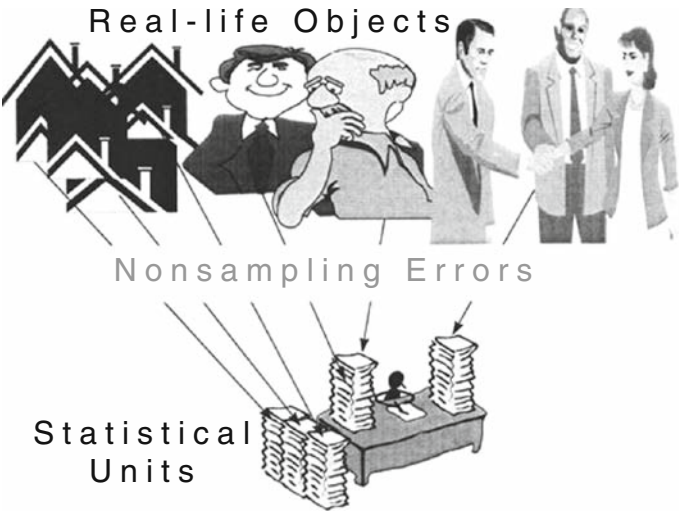

Fig. 2.1 From the real-life object to the 'statistical-counting-unit'

2.3.1 Surveying the 'Real-Life-Objects'

The process which transforms the 'real-life-objects' into 'statistical-counting-units' usually is the statistical survey. It can be a census, a sample, or some administrative listing that exists for other purposes but is made available to statistics.

Known is the population census. There are other, less known economic census operations: census of agriculture, of mining, manufacturing, whole-sale-retail establishments, and service industries. Even less known is the US census of governments, in which the local governments in the US are the real-life-objects. Because a census is a costly, major operation that requires a legal basis, a professional staff and big budget allocations, it is carried out only at 5 or 10 year intervals, and the different censuses are scheduled at different times because of the limited administrative capacity of census bureaus.

Another matter are the abundant sample surveys. Unless they are undertaken by a public or private professional sampling organization, they seldom serve a serious statistical purpose, but are used as a pretext to draw attention to a new product or some political cause.

Statistical theory has spent much thought and effort on improving the sample design in selecting the real-life-objects and managing the inevitable (mathematical) sampling error. As already mentioned, sampling theory and inference has dominated the discussion of statistics at the expense of nearly everything else.

This statistical process extracts from the rich reality of the existing 'real-life-objects a simplified – and often distorted – sketch of it on a questionnaire or other means of recording. It is a reduction process that is not reversible: The real-life object, e.g. a human person, cannot be reconstructed from a questionnaire, regardless of how much detail it contains and how conscientiously it has been filled out. Furthermore, once recorded, each 'statistical-counting-unit' starts its own existence, separate from, and independent of that of the real-life object. Even if the latter should disappear completely, the 'statistical-counting-unit' remains, as a lasting testimony to the former's existence. When tabulated, it survives even the destruction of the original record, on a questionnaire, punch-card, magnetic tape, CD or other device.

Statistical surveys record the real-life-objects in isolation from their socio-economic context. Usually real-life-objects of one kind are enumerated together, such as the dairy farms located in a country in a census of agriculture. Different types of real-life-objects are surveyed at different times, by different agencies, usually according to different criteria and definitions. No integral census has yet been accomplished that would report together human beings, factories, farms, mines, wholesale and retail establishments, banks and other service establishments, with their relevant characteristics. This inability to survey the entire society and its activities together, at the same time, results in discrepancies and variations in the data that have nothing to do with chance occurrences in the economy, but result from the truncation of socio-economic phenomena through the statistical process

2.3.2 The 'Statistical-Counting-Units'

It is interesting to consider the differences between "measurement" in the natural sciences and the corresponding statistical activity in the social sciences. In the natural sciences these measurements are the result of observations by objective especially trained observers, like in the bio sciences, so to speak from the outside of the thing to be measured. In the socio-economic setting the person providing the information e.g. in a population survey, really is the "object" to be observed. That self-reported information from many different informants of varying competence and intelligence is collected by survey takers, who themselves often are insufficiently prepared for that task, acting mostly as mail carriers, not like the observers in the natural sciences. The truthfulness and accuracy of such information depends on the cooperation of these interviewees, a matter that cannot be guaranteed, despite existing laws that require it. Neither their honesty nor the accuracy of their memory can be guaranteed. That is a fundamental, important difference between socio-economic statistical data and the measurement data in the natural sciences.

Statistical data have been variously classified. The distinction in 'Punkt- and Streckenmassen'[15] (point- and line masses), for example, is based on the length of life of the real-life-objects: some real-life-objects are perceived as being points in time, of short duration. Others last long, occupying a 'Strecke' that is, a considerable stretch of time. But every real-life-object has a certain duration. Considering its life span as point-like and short, or as long lasting, is a relative matter. Moreover, this distinction ignores the fact, that we do not deal with the real-life-objects themselves but with the 'statistical-counting-units' which are, by their nature, points in time and space, regardless of the length of life of the real-life object.

Another distinction in 'Bestands- and Bewegungsmassen' – inventories of a mass of stationary real-life-objects and masses of moving real-life-objects that are not stationary – is based on the spurious distinction between existence-units which are real-life objects that remain in their location without moving, and motion-units, that is, real-life objects that are on the move, without a fixed relation to a place in a geographic region. That obscures the fact, that every 'statistical-counting-unit' is a static record, fixed in a certain time and location, regardless of whether a real-life-object is static or dynamic.[16]

A distinction could be made between different types of 'statistical-counting-units' according to the occasion of their registration:

1. Real-life-objects are contacted by mail, telephone or personal visit by a concerted effort to record them, and approached at a certain point in time as in a census or sample survey, or
2. A government or private institution records the real-life object on the occasion of some event that triggers a registration, such as a beginning of something, a change of its characteristics, or its termination, carried out for other than statistical purposes. Typical is the registration of the birth of a child, the issue of a building permit for an addition to an existing building or for a new building, the registration of the bankruptcy of a firm (death), or the periodic re-registration of motor vehicles. In most of these instances the registration is requested by law,

is carried out as a continuing operation, often for the purpose of taxation, not originally for statistical purposes.

The first type leads to the statistical registration of all real-life-objects of a kind, as a (more or less) simultaneous cross section. On such occasions, events connected to these real-life-objects are also recorded, such as sales and costs during the past year in a census of enterprises.

The second type leads to the formation of ‘statistical-counting-units’ at uneven time intervals although the point in time at which the real-life-objects are registered can be important. The recording agency acts as a point at which an occurrences is registered, related to the issuing of a license or permit, or acting as a checkpoint for the flow of real-life-objects, like in studies of road traffic. The real-life-object on which the occurrence happens is often also registered. This distinction in cross sectional and longitudinal registration is roughly identical to another more familiar distinction: data collected for statistical purposes, and data collected as a by-product of administrative activities.

Both procedures yield still-pictures of the real-life-objects, somewhat like a photographic snapshot – except that less detail is retained. The purpose of such a statistical registration is not really to describe in detail the individual real-life-objects but to capture some socio-economic phenomenon in which that real-life-object is involved. In all these instances, statistical surveys yield still-pictures of the phenomenon. Its dynamism can be approximated through arranging these static still-pictures in sequence over time, such as e.g. yearly inventory figures for business units that are recorded by their accounting departments in a census-type operation, or monthly production totals for factories as the real-life-objects recorded in a continuous registration procedures.

At times various real-life-objects are registered collectively as one figure, such as the cattle on a farm in an agricultural census. No separate ‘statistical-counting-units’ are recorded on that occasion for each animal in the herd. Similarly, no ‘statistical-counting-units’ are produced in the case of the production of substances and materials that do not form individualized ‘real-life-objects’. In these instances, statistical information bypasses the formation of statistical units, in many instances even omitting to record the number of ‘pseudo-real-life-objects’ like barrels of crude oil produced. Instead of their number, only their total weight, volume or value is recorded. Similarly, statistics records the Kilo-watt-hours of electricity produced or consumed.

Statistical materials which are prepared from individually recorded ‘statistical-counting-units’ call for a detailed analysis of all the characteristics of the units which were investigated. Additional computations may help assess the respective socio-economic phenomenon. The more details of the real-life-objects were recorded in the ‘statistical-counting-units’, the more of the features of the phenomenon can be studied.

Statistical materials which were not prepared from individual ‘statistical-counting-units’ cannot be interpreted in much detail but become useful in the form of ratios and index numbers.

The schema in Fig. 2.1 visualizes this transition from the real-life object to become a 'statistical-counting-unit'. For empirical socio-economic studies, the 'statistical-counting-units' are the de-facto projecting agents of the socio-economic phenomena, not the 'real-life-objects' themselves that exist out there in society. These 'statistical-counting-units', then, are the actual building elements of the data in our field. It should be stressed again that one individual 'statistical- counting-unit' does not correspond to, nor is it comparable with an individual observation or measurement in the natural sciences. The differences in their respective roles will be further discussed in the next chapter on aggregation.

These 'statistical-counting-units' are only of transitory importance. It is their aggregation that yields the data that are of interest and are to be interpreted. A large number of 'statistical-counting-units' in an aggregate does not imply a greater validity of a statistical statement; nor does it establish the socio-economic phenomenon with greater certainty. The 'Law of Large Numbers' – the Central Limit Theorem – simply does not apply to socio-economic statistical data, except when actual samples are analyzed inferentially. These statistical elements link socio-economic reality 'out there' with the tabulated data, the aggregates, 'in here'. In the next chapter these aggregates will be explored into which the 'statistical-counting-units' are assembled. It is precisely through these aggregates that socio-economic data and the underlying phenomena can be interpreted.

Notes

1. See also: Winkler, Othmar W. "Statistical Flaws in Econometricians' Perception of Economic Reality" Roy C. Brown ed., *Quantity and Quality of Economic Research*, Vol. I, University Press of America, Lanham, MD, 1985, pp. 295–354.
2. Kendall, M.G. and Buckland W.R. *A Dictionary of Statistical Terms.* Hafner Publishing Corp., New York, 1957. United Nations, General Principles for a Housing Census
   UN Statistical Papers Series M, No. 28, New York 1958. United Nations, Seleccion de una Unidad Estadistica Adecuada para las Encuestas Economicas, UN Document E/CN.3/244, New York, 1957.
   United Nations, *Statistics of Enterprises*, UN Doc. E/CN.3/169,and E/CN.3/169 Add.1, New York, 1954.
   United Nations, *Statistics of Individual Industries*, UN Doc.E/CN.3/176, New York, 1954.
   United Nations, *The Statistical Unit in Economic Inquiries*, UN Doc. E/CN.3/259.
3. This section follows: Winkler, O. W. "The Units (Elements) of Economic Statistics," *Bulletin de L' Institute International de Statistique,* Ottawa, August 1963, Vol. XL, Book 1, pp. 305–306.
4. Statisticians are not the only ones who have difficulty with what one would call socio-economic phenomena. In a (yet unpublished) paper "Issues Management: A New Direction for Public Relations Professionals" Professor Annette Shelby states: "T. Campbell has called identifying the issues the most neglected step in issues management. One of the major difficulties is the vast number of possible issues in. . .the internal and external environments. A further problem is the "mushiness" of the signals perceived which. . . may identify emerging trends. The ambiguity is so great that F.J.Connor, President and CEO of American Can Co. doubts that issues identification can be formalized. 'Someone has to identify in the landscape of distinguished elements those that are significant, . . . He has to read meaning into what others, if they see the elements at all, simply dismiss. He has to see a new pattern emerging from an old. . .'

Despite reservations, organizations continue to search for a systematic approach to identifying issues. . . S. Goodman provides a useful summary: 'monitoring published information, using company volunteers to identify and track issues, establishing issues committees as a part of the board of directors. . .' Additionally, many organizations provide data from polling, content analysis, and general scanning. . ." Annette Shelby, p. 8,9, paper presented at the Conference on Organizational Policy and Development, Louisville, April 1984. Although the expression "socio-economic phenomenon" is not mentioned in this paper, it gives an excellent description of such phenomena.

5. Gartner, William. B. "A Framework for Describing and Classifying the Phenomenon of New Venture Creation," *Academy of Management Review* 10 (4), 1985, pp. 696–706.
6. The reader very likely has encountered statements like the following without being aware that it was what here is called a socio-economic phenomenon. 'Unemployment in Morocco in 1984' would be such an example. "The IASS, jointly with the INSEE and the 'Direction de la Statistique Maroc', is organizing a seminar on the statistical approach of the non structured sector and its effects on . . .. employment. . .in Rabat (Morocco). . . October 1984 to analyze the employment situation in a country, through household statistics alone. Those sources . . .allow a fair. . . knowledge of the active, employed and unemployed population. . . The main interest of these sources lies in grasping the whole of the phenomenon. . . and. to serve as a basis for the projections of planners. . . (p. 10). . .a systematic comparative analysis of the. . . information sources. . .is. . . .allowing. . . a new light on the employment phenomenons, (sic) underemployment and unemployment. . .differences in concepts, definitions, observation methods and fields of inquiry make such comparisons difficult. . .the ways to apprehend those phenomenons can. . .be modified . . .and improved" (p. 11).
   The Survey Statistician, Jl. of the International Association of Survey Statisticians, ISI, June 11, 1984.
7. The critical observation by a clinical psychologist of the practice to transfer results from 'individual differences research' to the study of 'individual behavior' may further illustrate the need to clarify the phenomena to be studied while also showing the distinction between socio-economic phenomena and science phenomena. He states that: (. . .kind of knowledge that individual differences research can legitimately . . .yield is neither "idiographic" nor "nomothetic" in any sense that a personality theorist would be compelled to take seriously. . .(p. 143). . .The task is to make salient the errors of reasoning by which the empirical findings generated by individual differences research are made to seem interpretable at the level of the individual. . .these errors are. . .just different versions of. . . the psychologist's fallacy. According to (William) James. . . whenever the empirical properties of data are uncritically assumed to reflect psychological properties of the persons observations on whom occasioned. . . the analysis of those data. (pp. 143–145).

   (Prior to a political election. . .pollsters. survey representative samples of the voting population in an attempt to forecast the outcome of the election. . ..in one . . .poll it has been found that 51% of the sampled voters favor candidate A, while 49% favor candidate B. . . .within the sample as a whole, there is a slight preference for candidate A, . . . no one would . . . interpret such findings as evidence that within each voting person there is (p. 144) a "mild preference" for candidate A over candidate B. Indeed, if this were the case, then an accurate poll would have revealed that candidate A was favored by 100% of the voters, and candidate B by no one. A 51–49% outcome might reflect any one of a very large number of underlying dynamics, one of which is that 51% of the sampled voters adamantly prefer candidate A while 49% just as adamantly prefer candidate B. . . then no one's preference is "mild". . . By the. . . same token,. . . a 99–1% outcome could be pointed to as evidence that there is an (overwhelming" preference for candidate A within the sample. . .such data would not constitute evidence that any voting person has an overwhelming preference one way or the other. . .it might just as well be true that each of 99% of the voters has but a mild preference for A, and each of 1% of the voters but a mild preference for B. . .no one in the sample could . . . be said to have an "overwhelming" preference at all. The point. . .is. . .whatever trends might be revealed by the. . .election poll. . . regarded as empirical trends that are. . ..in the overall pattern

revealed by the votes. Whether or not there are any voting persons who have psychological inclinations corresponding to the identified trends is a matter on which the available data are altogether mute... election polls fail to ... inform us about voters... A pollster can ... ignore epistemological problems...because... What is central is not ...knowledge about the...voters, but knowledge about the pattern of the ... votes (145) ... The point is that if the central theoretical assertion of generic structuralism is valid, the empirical findings generated by individual differences research can not ... ever establish that fact. (155).
James T. Lamiell, *The Psychology of Personality – An Epistemological Inquiry*, Columbia University Press, New York, 1987.

This psychologist's concern, perhaps without being aware, illustrates the difference between a socio-economic and a science phenomenon. (This topic is further discussed in section 3.3 "The Interpretation of Aggregates.").

8. quoted in: Winkler, Wilhelm, op.cit. 1948, Vol. II, p. 304.
9. Weinhandl, Ferdinand, *Die Gestaltanalyse*, Verlag Kurt Stenger, Erfurt, 1927.
10. This section follows essentially Osgniach, Augustine J., OSB, *The Analysis of Objects*, J.F. Wagner, New York, 1938, and Brugger, S.J. Walter, *Philosophisches Wörterbuch*, sechste Auflage, Verlag Herder, Freiburg, 1957.
11. There are of course directly opposing views of this kind of 'reality'. Mysticism is essentially a belief that reality is oneness...mystics believe that our common perception of the universe as containing multitudes of discrete 'real-life-objects'...trees, birds, horses, ourselves – all separated from one another by boundaries is a mis-perception, an illusion. To this...mis-perception...that most of us mistakenly believe to be real, Hindus and Buddhists apply the word "Maya." They...hold that true reality can be known only by experiencing the oneness through a giving up of ego boundaries. It is impossible to really see the unity of the universe as long as one continues to see oneself as a discrete object, separate and distinguishable from the rest of the universe.
M. Scott Peck, *The Road less Traveled* Simon & Schuster, N.Y. 1978, p. 96
12. Also see exhibit #1 in: Winkler, O. W. "A New Approach to Measuring Agricultural Production", *Proceedings of the Business and Economic Statistics Section of ASA*, Washington, D.C., 1963, p. 31
13. "The subjects which are most interesting in themselves do not lend themselves best to accurate observation and systematic study. But the two kinds of gradings can compensate for each other over a wide range of disciplines, in which they combine in variable proportions, and thus uphold throughout a steady level of scientific value. The supreme exactitude and scientific coherence of physics compensate for the comparative dullness of its inanimate subject matter, while the scientific value of biology is maintained at the same level as that of physics by the greater intrinsic interest of the living things studied, though the treatment is much less exact and coherent...science must accept to an important extent the pre-scientific conception of these subject matters. The existence of animals was not discovered by zoologists, nor that of plants by botanists, and the scientific value of zoology and botany is but an extension of man's pre-scientific interest in animals and plants. Psychologists must know from ordinary experience what human intelligence is, before they can devise tests for measuring it scientifically...If the scientific virtues of exact observation and strict correlation of data are given absolute preference for the treatment of a subject matter which disintegrates when represented in such terms, the result will be irrelevant to the subject matter and probably of no interest at all (p. 139) It... requires...hat we should explain all kinds of experience in terms of atomic data. This is of course the program of a mechanistic world view ... conjured up by Laplace's imagination has diverted attention from the decisive sleight of hand by which he substitutes a knowledge of all experience for a knowledge of all atomic data. Once you refuse this deceptive substitution, you see that the Laplacean mind understands precisely nothing and that whatever it knows means precisely nothing. Yet the spell of Laplacean delusion remains unbroken to this day. The ideal of strictly objective knowledge, paradigmatically formulated by Laplace, continues to sustain a universal tendency to enhance the observational accuracy and systematic precision of science, at the expense of its bearing on its subject matter....Scientific stringency,

inflexibly resolved to denature the vital facts of our existence, continues to sustain this conflict, which may yet issue in a sweeping reaction against science as a perversion of truth. . ..a complex historical movement has since then led, along a number of mutually related lines, to the establishment in our time of the scientific method as the supreme interpreter of human affairs. (p. 141).

The thing to realize is that a knowledge of physics and chemistry would in itself not enable us to recognize a machine. Suppose you are faced with a problematic real-life-object and try to explore its nature by a meticulous physical or chemical analysis of all its parts (added observation: obviously this is done in a precise, quantitative manner). You may thus obtain a complete physico-chemical map of it. At what point would you discover that it is a machine. . . and. . . how it operates? Never. . .you cannot even put this question, let alone answer it, though you have all physics and chemistry at your finger-tips, unless you already know how machines work. . .The physico-chemical topography of the real-life-object may in some cases serve as a clue to its technical interpretation, but by itself it would leave us completely in the dark in this respect. The complete knowledge of a machine as a 'real-life-object tells us nothing about it as a machine. (p. 330). . .the observation of the. . . real-life-object in terms of physics and chemistry will spell complete ignorance of what it is. . .the more detailed knowledge we acquire of such a thing, the more our attention is distracted from seeing what it is. . .Some physical and chemical characteristics of a machine, such as its weight, size and shape, or its fragility,. . .will be of interest in themselves on certain occasions, for example to a carter undertaking the transport of the machine. But this is about as much as the scientific study of a machine can achieve when pursued in itself, without reference to the principles by which the machine performs its purpose (p. 331) Physical and chemical knowledge can form part of biology only in its bearing on previously established biological shapes and functions: a complete physical and chemical topography of a frog would tell us nothing about it as a frog, unless we knew it previously as a frog. . .(p. 342) This is true in respect of inanimate things like . . . machines. But its major importance emerges only when we turn to living beings, where an important additional feature is added to it: . . .our recognition of individuals." (p. 343).
Polanyi, Michael, *Personal Knowledge – Towards a Post-Critical Philosophy*, The University of Chicago Press, Chicago, 1958.

14. The following excerpt from the weekly *Newsletter DH+S REVIEW* published by Deloitte Haskins + Sells, of August 18, 1986 unexpectedly confirmed my contention: "Boards of directors of the largest U.S. companies favor a stronger link between pay and performance. This finding comes from a survey conducted by Sibson & Company, Inc., which includes responses from 120 board members and 80 senior human resources executives of the 345 largest U.S. companies. The survey found that 60% of all directors responding thought that "more effectively relating pay to performance" is the most important issue facing large companies. **Qualitative aspects** of performance **were deemed** by directors **to be more important than quantitative measures**:" (highlighting added for emphasis). Directors Responses to Selected Measures of Evaluating CEO Performance Relative importance Assigned by

Directors			
Quantitative measures	%	Qualitative measures	%
EPS over 2–5 years	61	Establishing strategic direction.....	89
Return to shareholders.....	57	Building management team.....	80
Return on inv. Capita.....	37	Leadership qualities.....	73
Return on stockholders' equity.....	35	Providing for succession.....	71
Return measure trends.....	33	Implementing strategy.....	57

15. Winkler, Wilhelm, op.cit. Vol. I, p. 25.
16. Kellerer, Hans, *Übertragung einiger in der Bevölkerungswissenschaft gebräuchlichen Begriffe und Methoden auf das Wirtschaftsleben,* Muenchen 1951, p. 5.

Chapter 3
Structure and Nature of Socio-Economic Data: The Aggregates

Divide et Impera
Julius Caesar
[Subdivide (de-aggregate) and interpret]

3.1 The Tri-Dimensional Frame of an Aggregate

Socio-economic phenomena deal not only with a subject-matter aspect but also with a time and a regional-geographic aspect. The real-life-objects, and their corresponding 'statistical-counting-units' that portray those phenomena, partake in those three aspects that can be conveniently visualized as the three perpendicular vectors or dimensions of a coordinate system. This means that every aggregate[1] that deals with socio-economic phenomena can be understood as occupying a tri-dimensional space like in a Cartesian coordinate system (Fig. 3.1).

The subject-matter dimension can be presented on the vertical vector of a statistical aggregate, or on any other of the vectors, if so preferred. The sub-divisions of the subject-matter, e.g. the major groupings of the classification of economic activities, can be indicated in linear form by corresponding tick-marks.[2]

For the social sciences the development of phenomena over time is of great interest. Time, therefore, should be marked on the second vector of that – still empty – Cartesian space, facing the observer, using the customary subdivisions of the calendar (months, quarters, semesters).

On the third, geographic vector, the administrative regions are plotted as a one-dimensional sequence. Regional districts, reduced to linear form, are projected on the geographic vector. Figure 3.2 shows the tri-dimensional frame of a statistical aggregate before the 'statistical-counting-units' are placed into it. One could imagine this empty space framed by the three vectors to look like an empty fish tank with its three dimensions.

It is important to recognize that these three dimensions are present in all statistical data. This is easily overlooked, because data published in tabular form usually present only two of these three dimensions, either the subject-matter and time, or the subject-matter and geographic-territorial-dimension. This is true as much for aggregates as it is for other data-materials that are derived from aggregates.

When the time dimension is small, like in a census or inventory, the tri-dimensional character of a statistical aggregate shrinks to a seemingly

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_3,

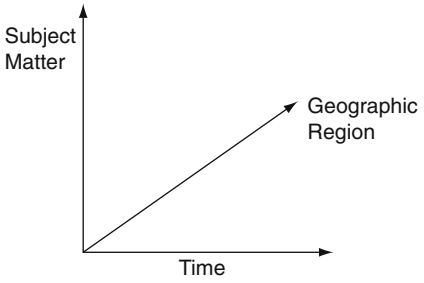

Fig. 3.1 The tri-dimensional coordinate system

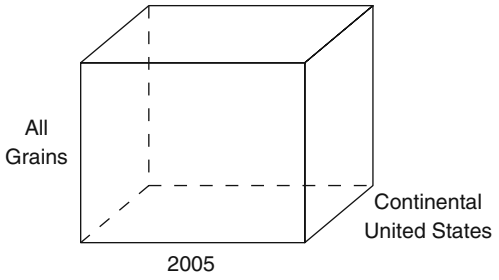

Fig. 3.2 The conceptual frame of an aggregate

two-dimensional sheet and is easily overlooked. Yet, like a sheet of paper that, regardless how thin it may be, still has a thickness that becomes evident when e.g. 500 sheets of such thin papers are packaged as a ream. In the case of a survey, the time dimension of the resulting data consists of those few hours or days – in a census of a big country that may be many months – needed to accomplish the field work, capturing a specific socio-economic phenomenon at that particular point in time. It may take that long to locate the respective 'real-life-objects,' to interview or canvass them and forward the result to a central location, an office, to produce the 'statistical-counting-units'. The placement of the resulting aggregates on the time vector, Fig. 3.3, is important, because it allows to connect them to other statistical and non-statistical materials.[3]

The aggregate-space can be subdivided in a variety of different ways. Sub-aggregates can be formed by introducing additional sub-set[4] forming criteria in any

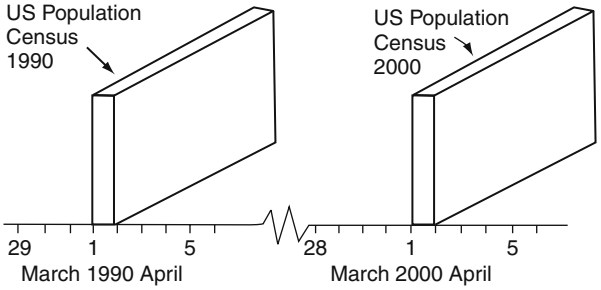

Fig. 3.3 Placement of aggregates along the time vector

one of the three dimensional vectors, or simultaneously, in any **two** of the vectors or in all **three** of these vectors at the same time. Consider, for example, the conceptual box for the statistical aggregate, 'Total U.S. Grain Production, 2005.'[5]

A breakdown along the vertical subject-matter vector, e.g. by economic categories, can be created by slicing the space of the aggregate horizontally into an array of thinner horizontal spaces, each representing another one of the grain species as separate sub-aggregates. Note also that each of these new, thinner conceptual spaces extends over the entire geographic region of the entire USA, for the full length of the entire year. Each of these horizontal – still empty – conceptual spaces, represented by boxes, represents the different categories or species of grain that are harvested, whose separate information is of interest, and of which sufficient quantities can be expected to be produced during the entire year, in the entire territory of the United States, such as wheat, barley, oats, corn, sorghum, etc. Fig. 3.4. The original total space of the aggregate is taken apart into a set of sub-aggregates along the vertical vector. Each of these sub-aggregate spaces can then become a separate aggregate, as a piece of potentially meaningful statistical information.

If one were to study the timing when grain was produced, one would form sub-aggregates along the time vector according to trimesters, or shorter subdivisions of time, breaking up, so to speak, the entire space of the original aggregate, in the example of the 'Total Grain Production, 2005,' into vertical slices (Fig. 3.5).

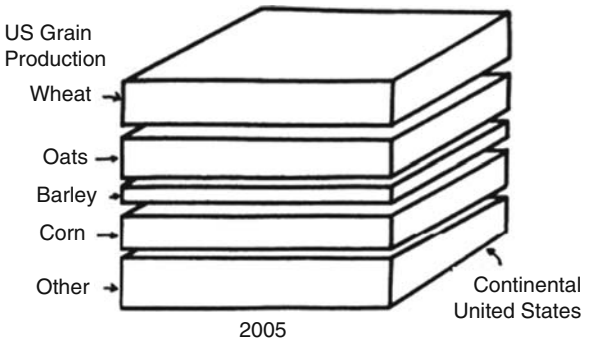

Fig. 3.4 Forming sub-aggregates by subject-matter

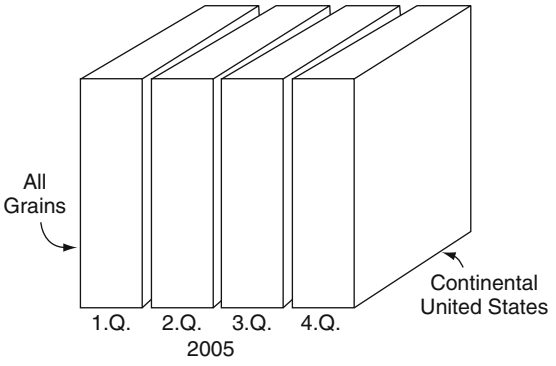

Fig. 3.5 Forming quarterly subdivisions

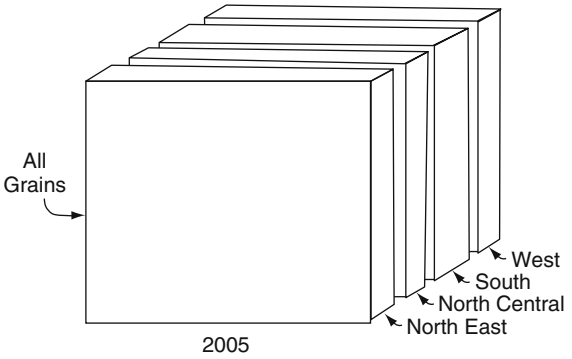

Fig. 3.6 Forming geographic subdivisions of the aggregate

The geographic territory can also be subdivided along the geographic vector, Fig. 3.6 to study that grain production for the major regions of the country.

Sub-aggregate spaces can also be formed simultaneously along any two dimensions. Figure 3.7 shows a breakdown by subject matter and geographic subdivisions for the entire year.

It can also be subdivided according to geographic regions and shorter time intervals, Fig. 3.8. The total of All grain produced in 2005 can also be subdivided by subject matter and time, for the entire country, Fig. 3.9.

Finally, sub-aggregates can be formed along all three dimensions at the same time. In this example, the production of the different grain species would be subdivided by region and by shorter time interval – e.g. quarters – Fig. 3.10. It should be noted that qualitative, non-measurable characteristics or attributes of the 'statistical-counting-units' are the important subset-forming criteria, although the quantitative or measurable characteristics – often incorrectly referred to as random variables – have a similar, though much less important role in forming sub-sets.

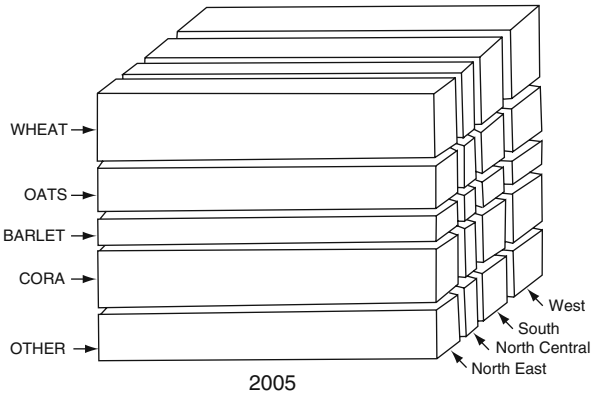

Fig. 3.7 Breakdown of grain production by species and region for the entire year

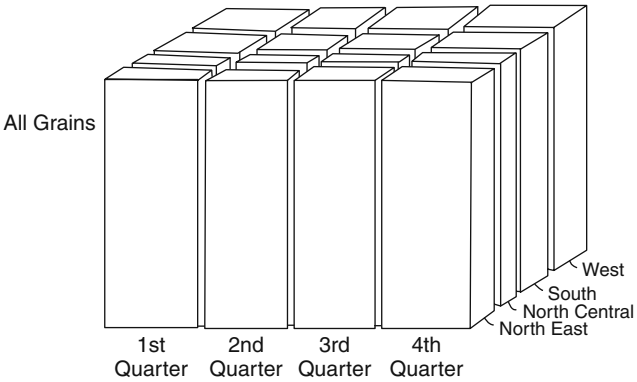

Fig. 3.8 Breakdowns by region and time interval for the entire grain production

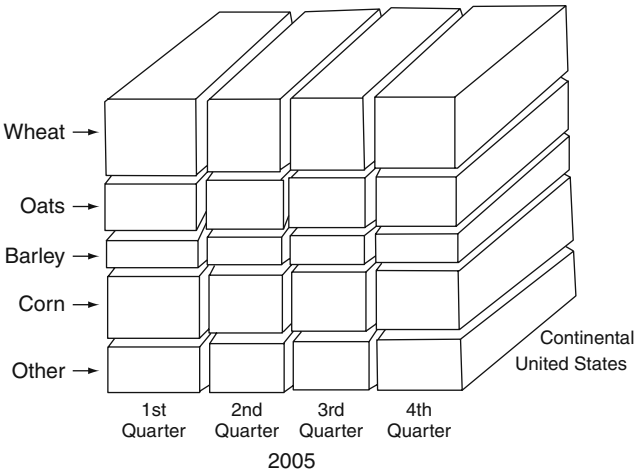

Fig. 3.9 Breakdown of grain production by species and time interval (*quarters*) for the entire country

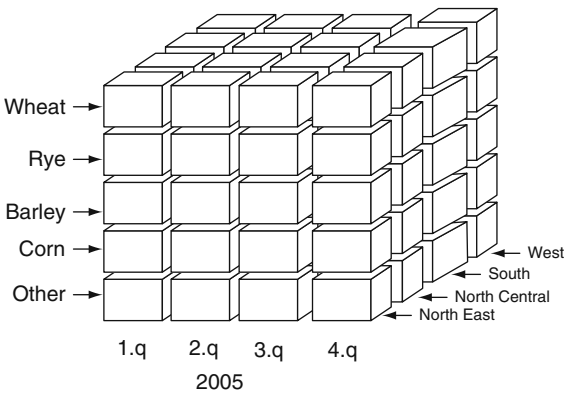

Fig. 3.10 Breakdown by subject-matter, time and region

3.2 The Nature of Socio-Economic Statistical Aggregates

The empty conceptual boxes represent the defined frame of a statistical aggregate. The elementary building blocks of statistical data are then formed by placing the 'statistical-counting-units' into these empty conceptual spaces. It is the number and location of the 'statistical-counting-units' within the aggregate **together** with the definitions of subject matter, region and time period, that yield the statistical aggregate as the primary statistical data-material (Fig. 3.11).

The internal structure of an aggregate is of no interest, and is treated as if it were unknown. It can be revealed only by decomposing the aggregate. This is the key to a better understanding of the data in statistical tabulations, because nearly all these data are aggregates or based on aggregates.

Through the comparison with other, similarly defined aggregates the structural features of a socio-economic phenomenon can be revealed. That overview, however, is achieved at the expense of detail in the 'statistical-counting-units', resulting in a loss of information, a notable fact about which statistical theory so far had nothing to say.

The interrelation of this frame of an aggregate and the 'statistical-counting-units' within it, as well as the relation between aggregates merits additional discussion. A colleague, expert in market research, referred to the de-aggregation of census material as 'peeling the onion – you continue breaking the aggregate down until you arrive at the smallest group.' This was an astute observation but not close enough to describe the actual nature of statistical aggregates. If one continues to break a given set into subsets, one never reaches a core comparable to the cen-

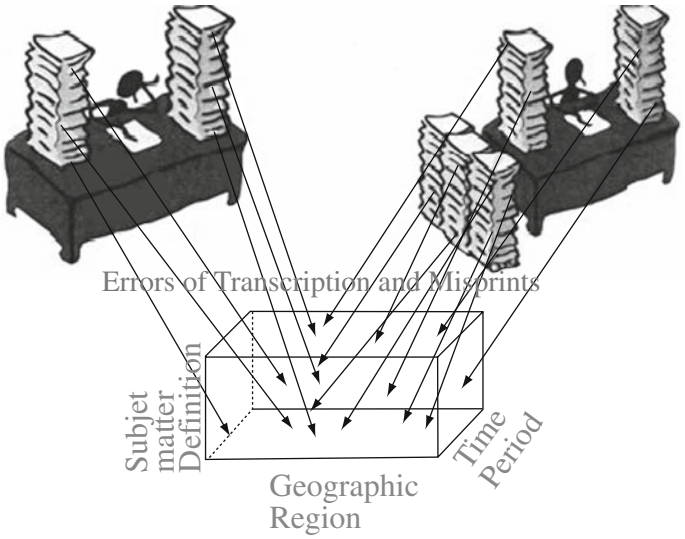

Fig. 3.11 Creating a statistical figures filling the concept-space of the Aggregate with 'statistical-counting-units'

ter of an onion. To the contrary, the aggregate can be taken apart into separate, smaller aggregates which can become independent statistical materials when taken out of their original context. In contrast to living beings whose parts are functionally related to each other, the parts of a statistical aggregate do not have such a functional inter-relationship.

I would like to propose a different image: Statistical aggregates are like a trick toy, similar to 'Chinese boxes.' A Chinese box consists of a cubic box that is completely filled with identical, smaller cubic boxes that fit exactly, without leaving any empty space. Each one of these smaller cubic boxes in turn contains a set of tightly fitting, even smaller, solid cubes. To add to the surprise, when opening these boxes, each box has another color. Statistical aggregates are like these Chinese boxes. There is a difference, though: aggregates are not as limited; because they can always be further disassembled into even smaller sub-aggregates. Moreover, there is always a possible, larger statistical aggregate of which the given aggregate can be considered a sub-aggregate. Although extremely large aggregates can be imagined – e.g. the GDP for the decade 1990–1999 for all countries of North, Central and South America – or an extremely small one – e.g. the Sales in the men's clothing department of Macy's store #13 during the Saturday-sale between 8:00AM and noon on Nov. 7. In reality the realm of meaningful aggregates, of course, is much narrower, limited by considerations of usefulness and practicality.[6] The definitive lowest limit for possible smallest aggregates is the detail with which the 'statistical-counting-units' are recorded. The pun regarding excessive specialization also holds regarding the usefulness of de-aggregation with regard to learning about a socio-economic phenomenon: 'One learns more and more about less and less, until one knows everything about nothing.'

Given that one has no interest in the internal structure of the aggregate, viz. about the distribution of the 'statistical-counting-units' within the three-dimensional space of the aggregate, these 'statistical-counting-units' could be concentrated at the beginning of the time dimension of the aggregate, e.g. during the weeks of January, Fig. 3.12.

As another possibility, the 'statistical-counting-units' could be concentrated in the middle of the time interval, e.g. July–August, Fig. 3.13.

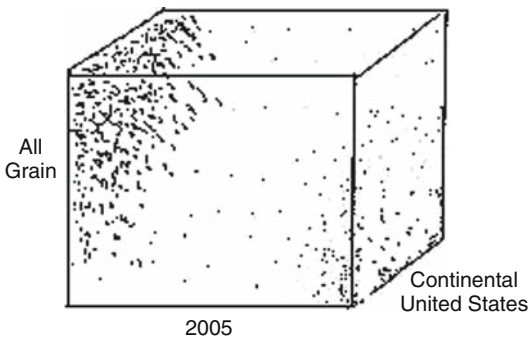

Fig. 3.12 Units concentrated at beginning

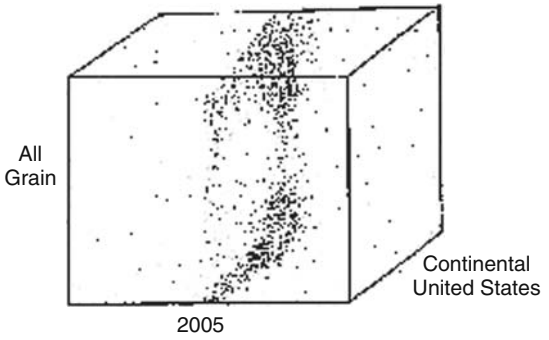

Fig. 3.13 Units concentrated in center of the time dimension

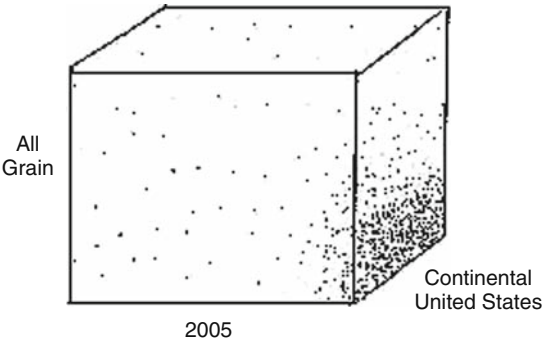

Fig. 3.14 Units concentrated at the end of an aggregate

Also most of the 'statistical-counting-units' could happen at the end of the time interval, e.g. in December, Fig. 3.14.

Something analogous, of course, could happen with regard to the other dimensions of the aggregate: the 'statistical-counting-units' bunched together at the beginning, in the middle or at the end of the subject matter- or the geographic dimensions (not shown). Although one can make reasonable assumptions about the likely distribution of 'statistical-counting-units' within an aggregate, one cannot know how they are really distributed, until one actually disassembles the aggregate into sub-aggregates. The distribution remains undetermined below the given level of aggregation. That means that one can only perceive those aspects of a socio-economic phenomenon that are projected by the 'real-life-objects', and only to the extent to which they could be located in society, and their characteristics were correctly recorded in the 'statistical-counting-units'. Depending on the 'size' of an aggregate – that is, according to the number of smaller subdivisions, with respect to subject matter, geographic and time extension of the definitions which are included, not according to the number of statistical-counting-units – one focuses on different aspects of that phenomenon.[7] This statistical perception of a socio-economic phenomenon differs by its inherent, inevitable fuzziness from the manner in which phenomena in the sciences are perceived.

Summing up, in a statistical aggregate all features in the 'statistical-counting-units', which are not explicitly stated in the definition of the aggregate, are ignored

and disappear. In other words, the larger a statistical aggregate, the broader and more inclusive its definition with regard to subject matter, region and time interval, the fewer of the features of the individual 'statistical-counting-units' in that aggregate remain recognizable. This leaves correspondingly fewer details available for the interpretation of a socio-economic phenomenon. There are usually more 'statistical-counting-units' in the larger aggregates, but this is not always the case and one cannot count on it as a given.

3.3 The Interpretation of Aggregates

The larger a statistical aggregate, viz. the broader and more inclusive its definition is with regard to subject matter, time interval, or region, the fewer of the features of the individual 'statistical-counting-units' included in that aggregate, are available to reveal features of the phenomenon. Features of the 'statistical-counting-units' not specifically stated in the definition of an aggregate, become de-activated and leave fewer features available for the interpretation of a socio-economic phenomenon.

Statistical aggregates, akin to the arithmetic mean, level out the content of the aggregate. Like the arithmetic mean, aggregates are an abstract, artificial creation of statistics that has no counterpart in perceptible reality. Such an aggregate, in isolation, can hardly be interpreted. In linking it to other, similarly-defined aggregates, however, it can be compared meaningfully with regard to other time periods or other geographic regions. The intuition was correct that led to the pronouncement 'comparison is the soul of statistics.'

The value of such comparisons lies in the overview, the big picture of a socio-economic phenomenon that otherwise is not available. These 'statistical-counting-units' are the actual building elements of the data in our field but are only of transitory interest. They become valuable through their aggregation that yields the data that are to be interpreted. But a large number of 'statistical-counting-units' in an aggregate neither implies a greater validity of a statement, nor establishes the socio-economic phenomenon with greater certainty. It is not clear either how the Law of Large Numbers – the Central-Limit Theorem – is supposed to apply to aggregates.

The specific location of the 'statistical-counting-units' within this three-dimensional hierarchical structure of an aggregate is what matters. The social situation is revealed through the frequencies with which 'statistical-counting-units' with certain characteristics occur within the system of statistical aggregates and sub-aggregates. One's perception of social reality might be comparable to the visual perception through the multifaceted eyes of a housefly. It presumably sees in its surroundings only the movement of things and changes in their brightness, but does not recognize the specific shapes of these things. Similarly, statistical aggregates allow one to notice only the changes in the phenomenon through an increase or decrease over time and over regions, in the number and location of the 'statistical-counting-units' in these tri-dimensional spaces of aggregates. If one changes the width of these definitions of the aggregate, e.g. by defining more

narrowly the limits of one or more of these three dimensions, then one observes the same socio-economic phenomenon, so to speak, with another, differently-structured pair of insect eyes. One perceives then the local and short-lived aspects of the same socio-economic phenomenon, through the fluctuation in the frequencies with which the 'statistical-counting-units' occur in the sub-aggregates.

In the natural sciences, each individual measurement represents an observation that gives account of the phenomenon. Although it may be distorted by mistakes in measuring and other errors of perception, there is empirical evidence that these measurement-errors tend to cancel each other out, the more of such individual repeat-measurements can be obtained. The scientific phenomenon then can reasonably be expected to emerge with less distortion. Hence, the importance of the Gaussian distribution, also tellingly referred to as the error distribution or normal curve.

In the socio-economic field, however, neither the 'statistical-counting-unit' nor the aggregates are the counterparts of measurements in the sciences. Rather, the 'statistical-counting-units' portray socio-economic phenomena such as unemployment, production, foreign trade, inflation, or sex discrimination, like the small, colored mosaic stones that compose a mural. Each individual stone represents something quite different than the picture which these stones compose collectively. Not one of these mosaic stones, individually, gives the slightest idea of that picture. Yet together, they create it. Socio-economic phenomena are like such murals: the 'statistical-counting-units' function like the small mosaic stones, and their aggregates represent clusters of similar mosaic stones in that mural.[8]

Consider 'unemployment,' a socio-economic phenomenon that can be perceived only when all the persons in search of a job – but lacking one at the time of a statistical survey is taken – are viewed together. The intensity and structure of that socio-economic phenomenon is given by the clustering of those who cannot find a job in certain geographic locations, industries, occupations, gender, racial and age groups. The socio-economic phenomenon, 'unemployment' represents something that is different from that which each one of those individuals represents. A person without a job is not the economically and politically important phenomenon 'unemployment.'

Small, medium and large aggregates are needed to penetrate into the different aspects of a given socio-economic phenomenon. These are different, complementary ways of perceiving the same phenomenon.[9] The actual distribution of the units inside the aggregate remains unspecified, and to that extent the aggregate is fuzzy. When formerly separate aggregates are joined together into larger aggregates, the detail of the formerly separate categories, regions and time periods is submerged and disappears in the wider category, region and time period of the new aggregate. The 'mix' within any aggregate is not known until it is revealed by dis-aggregation. The analyst should be prepared to find substantial differences from one sub-region and time-sub-period to another. Obviously knowledge of the subject area is more valuable to interpret statistical aggregates than proficiency in probability calculus.

The concreteness of meaning vanishes to the extent to which the definitions of the aggregate become more inclusive. The loss of 'concreteness' in the process of forming larger aggregates is unpredictable. The size of the aggregate – size with

regard to the inclusiveness of the definitions of any of its three dimensions, – and the number of 'statistical-counting-units' in it, are not necessarily related, although larger, that is, more widely or inclusively defined, aggregates can be expected to contain a larger number of 'statistical-counting-units' than smaller aggregates.[10] At any rate, fewer characteristics of the 'statistical-counting-units' in large aggregates are defined than in smaller aggregates. This also accounts for the reason that the 'statistical-counting-units' have fewer features in common in large aggregates.

When dealing with the analytical value of a large aggregate, the substance of a cloud comes to mind. At a distance, that cloud can be perceived as a distinct entity. Closer up, however, its distinctness vanishes and the cloud does not seem to have much tangible substance left; one might even fly through such a cloud – e.g. traveling by airplane – without noticing much of its presence, except when flying into an electrically charged cumulus cloud.

To explore this important point further, consider the movements over time of the value of a market basket of goods. The prices recorded for specific quantities of each sold/purchased item in a base period is important for price measurement. Using that concept seems to imply that the products in that basket and their prices remain separate entities, such as a quart of milk, a package of carrots, a bunch of celery, a bar of butter, half a dozen eggs, a loaf of bread, etc. in the belief that the statistical information about these items also remains individualized after aggregation. But this takes that romantic notion of going shopping with a woven wicker-basket too literally. Conceptually statistical aggregation does to these separate products what a food blender would do to them physically; namely, converting them into a smooth, uniform pulp in which eggs, carrots and the other items have disappeared. The process of statistical aggregation transforms the characteristics of the items in the group in an analogous manner. The resulting aggregate is a new something with new characteristics that are a composite of all ingredients, their formerly separate characteristics, yet is different from those of the original ingredients. The more varied the characteristics of the items that were included in that aggregate, the less clearly defined the resulting pulp – the statistical aggregate.[11]

The view of an economy through the highly aggregated data of the GDP is qualitatively different from, and deals with different aspects of the economic situation, than the view conveyed by less aggregated statistical data. This is not a subtle point of minor importance, but potentially a big issue that has to be confronted. All this is to be kept in mind when interpreting such aggregate data.[12] Although aggregation is at the heart of statistics, no attention has been paid to the 'loss of meaning through aggregation'. Statisticians, social scientists and economists in particular, even lack an awareness of its existence as indicated by the fact that they do not even have an appropriate terminology for it.

3.4 Tabular Presentation of Aggregates

What has been said about aggregates ought to find its expression in an appropriate tabular presentation. In fact, properly presented tables could be the first tool of statistical interpretation.

There is no awareness among analysts of socio-economic data that important differences exist in the interpretation of statistical aggregates of different size or magnitude. It would help if such differences would be indicated by the size of the printed characters and the intensity of the printing ink. The headings, text and numbers of large aggregates – large with regard to their subject-matter grouping, their region and/or the length of their time period – ought to be printed in large, characters. Headings, text and numbers of smaller aggregates ought to be printed in correspondingly smaller print-characters.

The degree of vagueness of the aggregates should be indicated by the corresponding color-intensity of the printing ink. The large totals, corresponding to large, vague aggregates should be printed in faded hues of gray. Smaller aggregates are to be printed in correspondingly more intensive, darker inks to indicate symbolically their greater closeness to reality. Solid black ink should be reserved for the data of the smallest possible aggregates. The same would hold if other than black printing colors were used. Let a simple set of data serve as an example. Ordinarily, such data would appear printed as shown in Table 3.1.

The proposed presentation of those data in contrast would appear as in Table 3.2a and b. Statistical aggregates can only be interpreted properly if they are broken down in sub-aggregates along its three vectors, since small, medium and large aggregates allow one to perceive the equally important complementary features of a given socio-economic phenomenon.[13]

More appropriate tabulations than the current way of printing socio-economic data would remind the user of the differences that exist in the meaning of these data. This awareness should help to interpret data appropriately reducing the need for some of the subsequent data manipulations that are believed to be necessary today.

In conclusion, it bears repeating, that **each aggregate** – regardless of whether it has many or few 'statistical-counting-units's – **is just one piece of statistical information**.

Table 3.1 The usual way of presenting data

US. Federal Government Direct Expenditures, Fiscal Year 2003

Total expenditures	**719,031,000,000**
Education	70,685,000,000
Public Welfare	265,105,000,000
Health/Hospitals	68,374,000,000
Highways	72,455,000,000
Police protection	9,860,000,000
Correctional Facilities	36,938,000,000
Natural Resources	17,110,000,000
Parks/Recreation	4,636,000,000
Governmental administration	42,846,000,000
Interest on Government Debt	31,295,000,000

Table 3.2 **a** Data appropriately presented; **b** The same data even more appropriately presented

a

US. Federal Government Direct Expenditures, Fiscal Year 2003

Total Expenditures...719,031,000,000	
Education	170,685,000,000
Public Welfare	265,105,000,000
Health/Hospitals	68,374,000,000
Highways	72,455,000,000
Police Protection	9,860,000,000
Correctional Facilities	36,938,000,000
Natural Resources	17,110,000,000
Parks/Recreation	4,636,000,000
Governmental Administration	42,846,000,000
Interest on Government Debt	31,295,000,000

Or

US. Federal Government Direct Expenditures, Fiscal Year 2003

Education	170,685,000,000
Public welfare	265,105,000,000
Health/Hospitals	68,374,000,000
Highways	72,455,000,000
Police protection	9,860,000,000
Correctional facilities	36,938,000,000
Natural resources	17,110,000,000
Parks/recreation	4,636,000,000
Governmental administration	42,846,000,000
Interest on Government Debt	31,295,000,000
Total Expenditures...719,031,000,000	

b

US. Federal Government Direct Expenditures, Fiscal Year 2003

Education		170,685,000,000
Public Welfare		265,105,000,000
Health/Hospitals		68,374,000,000
Highways		72,455,000,000
Police Protection		9,860,000,000
Correctional Facilities		36,938,000,000
Natural Resources		17,110,000,000
Parks/Recreation		4,636,000,000
Governmental Administration		42,846,000,000
Interest on Government Debt		31,295,000,000
	Total Expenditures	719,031,000,000

Notes

1. The terms Aggregates and Aggregation in this book are not used in the sense in which economists and econometricians have used them, and as treated in H. Theil's book. These terms are used here in the simple and literal sense of adding together groups of such 'statistical-counting-units' that have certain characteristics in common, are located in a specified geographic area, and occur during a specific time interval. It was a pity that econometricians did not decide to explore the nature of socio-economic statistical data from the elementary building blocks up. They could have avoided the man-made Aggregation-Problem or the problems of 'consistent aggregation.'
   Theil, Henry, *Linear Aggregation of Economic Relations*, North Holland Publ. Co. Amsterdam, 1954, Green, H.A. John, *Aggregation in Economic Analysis-An Introductory Survey*, Princeton University Press, Princeton, NJ, 1964.
2. See: Winkler, Othmar, op. cit. 1964, pp. 67, 68, footnote #29.
3. By a remarkable coincidence the geographer Brian Berry published a Matrix, an identical three-dimensional scheme, to expound his ideas about future quantitative research in human geography. Only the time and geographic vector in his model are reversed, the geographic instead of the time dimension, occupying the more prominent position facing the observer. Brian J.L. Berry, "Approaches to regional Analysis: A Synthesis" in: Brian J.L. Berry and Duane F. Marble, editors,*Spatial Analysis: A Reader in Statistical Geography*, Prentice Hall, Inc. Englewood Cliffs, NJ, 1968, pp. 24–34.
4. The term 'set' in many a reader's mind has become identified with the new math in which the concept of algebraic numbers has been re-formulated in set-theoretical terms. It is unfortunate that the sets and subsets of abstract algebraic numbers are essentially one-dimensional. There is no reason, however, why these set concepts should not be applied to statistical aggregates which are tri-dimensional.
5. Winkler, Othmar W. "A New Approach to 'Measuring' Agricultural Production" *Proceedings of the Business and Economic Statistics Section of ASA* Washington, D.C., 1962, pp. 30–36. See "Appendix" pp. 34–35.
6. Discussed in Winkler, Othmar W. "The Nature of Statistical Information in Business and Economics" op. cit., pp. 69, 70.
7. Further examples in: Winkler, Othmar "A Critical View of Time Series Analysis in Business and Economics" *Proceedings of the Business and Economic Statistics Section of ASA*, Washington, D.C., 1966, pp. 352–370.
8. The statistical perception of socio-economic phenomena is also similar to the photographic process (before the time of digital cameras and digital photography). When taking a black and-white picture e.g. of a group of persons, the light reflected by them caused the light-sensitive chrome-silver particles of the film emulsion to become dark or light. All of these individual chrome-silver particles, each one in its place on the super-amplified 'blowup' of that photo, became visible as separate dots of gray or black on the blank background of the photographic paper. Taken out of their context, these chrome-silver particles would be meaningless dots of various shades of gray or black. Nothing is left of the picture if these emulsion elements were arrayed in a frequency distribution according to their degree of color intensity, from black at the beginning of that scale, to shades of gray and white on the other end of that scale. In socio-economic statistics the objects – really their images, the 'statistical-counting-units' – play the role of the chrome-silver elements. Analogously, the phenomenon is materialized by the objects (through the 'statistical-counting-units) in their actual location with respect to subject-matter category, geographic location and time. Only then can the socio-economic phenomenon be perceived.

   Let me further illustrate with other examples the relationship between socio-economic phenomena, the 'real-life objects' and the 'statistical-counting-units' in which these phenomena materialize. Each one of us exemplifies the phenomenon 'population' through our personal characteristics. Yet, 'population' is something quite different than the individuals that compose

it. Some of these individuals partake in the more concrete phenomenon (level 3, see above) 'population of Washington D.C., June 2007.' Similarly related are the 'individuals who have no job' to the phenomenon 'unemployed person,' 'unemployment' and 'unemployment in Washington D.C., June 2007.'

Or consider the phenomenon 'road traffic.' It materializes in all sorts of 'vehicles,' yet 'traffic' is something quite different from these vehicles, having its own flow, rhythm and existence. Traffic continues to exist, even when all of these vehicles have been taken out of circulation and were replaced by different, new ones. The phenomenon 'traffic' always transcends and survives, so to speak, its individual projecting elements. It should become obvious that e.g. the phenomena 'vehicle,' 'road traffic' and 'traffic on Roosevelt bridge, Washington D.C., March 5, 2005 between 8:00 and 8:30 AM' are different aspects of the same engineering, economic and social phenomenon.

9. Suppose it were possible to get the weekly figures of the agricultural production for every subspecies of all grain products, for every county of the USA. That crop situation would be presented through a confusing number of many very small aggregates with production figures that can differ widely from one subspecies, region and week to another. It would be like viewing a countryside from a low-flying airplane: much detail moves by very rapidly, but the observer does not perceive the larger context of that countryside. Returning to the example of the grain harvest, the statistical figures about that same crop situation, presented only in totals for each grain species, by four major regions of the country, and e.g. in yearly totals, again show the same phenomenon 'grain production' but offer a different view, like looking at the same countryside from a land-surveying satellite in space that circles the earth. The higher that vantage point, the better the overall picture can be appreciated, though much detail has vanished form sight and seems lost. Winkler, Othmar, op.cit. 1964, p. 74.
10. The *number* of statistical units is only loosely related to the 'size' of the aggregate, that is, the manner in which it is defined. Consider, for example, the aggregate 'total wheat production in the USA during 2005'. The definition includes the harvest of all the species of wheat, during the twelve months of that year and in all regions of the country, including those that cannot produce wheat. If we reduce this aggregate to only those few summer months in 2005 during which the bulk of wheat was harvested, then the number of units (e.g. bushels) in that reduced aggregate hardly will differ from the number of units in the aggregate that referred to the wheat production of the twelve months of that year. Similarly, the number of units in the aggregate will hardly be affected if a different aggregate of wheat production were formed with a geographic dimension that is reduced to the wheat producing regions only. Even if the geographic dimension of that aggregate were reduced to only 10% of the territory of the original aggregate, the total number of statistical units in it would remain about the same.
Winkler, Othmar, footnote #12 in: "The Nature of Statistical Information in Business and Economics" op. cit., 1964, p. 74
11. O. Winkler, op. cit. 1983.
12. Pioneers like R. Frisch and T. Haavelmo, to give an example of such developments, under the spell of the natural sciences, have ventured to transpose the paradigms of 'measurement' and 'measurement errors' from the natural to the social sciences, with the intent of converting the 'soft' social science of economics into the more respectable, 'hard science' of econometrics.
13. Winkler, Othmar, "The Meaning of Statistical Data in Business and Economics" an abstract, *Bulletin of the 35th Session of ISI*, Vol. XLI – Book 2, Belgrade, 1965, pp. 978–980.

Chapter 4
Ratios in the Social Sciences

4.1 Why Discuss Ratios?

Besides the proper tabulation of aggregates, ratios are simple yet powerful tools of interpretation.[1] The algebraic operation of division establishes a binary relationship between two numbers, to subdivide, but more importantly for statistics, to compare. The importance of ratios resides in the fact that a statistical aggregate is an artificial creation that has no counterpart in the perceptible world of human experience. Its informational content, the big picture of a socio-economic phenomenon can only be revealed through connecting a given statistical aggregate with other, similarly abstract creations, namely other statistical aggregates.

It should not surprise, then, that a great variety of ratio applications can be found in the Social Sciences. Many important ratios known as 'rates,' 'percentages,' prices, 'index numbers,' 'GDP- per-capita,' arithmetic, geometric or harmonic means, and probabilities, are ratios but are seldom recognized as such. Data users need guidance to interpret these ubiquitous statistical tools, because a statistical aggregate alone, by itself, cannot be interpreted – remember: 'comparison is the soul of statistics.' This important topic has received little attention[2] because statistical theory essentially deals with data that are the measurements in the natural sciences that do not need ratios for their interpretation. Ignoring the centrality of aggregates, and the fact, that events in society happen in historic time and actual geographic regions[3] has made statistics as an academic discipline less relevant for the social sciences,[4] gradually marginalizing it in academic curricula.

4.2 Classifications of Ratios

Ratios can be classified according to (1) the purpose for which a ratio is computed, and (2) the conceptual closeness between numerator and denominator.

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_4,

4.2.1 Ratios Classified by Their Purpose

Every ratio is computed for a purpose. Even though there are always various reasons why a given ratio is computed, one usually stands out and prevails in any given application. The following purpose-categories are listed in the order of frequency of their use.

4.2.1.1 Reference Ratios use the figure in the denominator, (D) as a reference or standard for the figure in the numerator, (N). Though every ratio, regardless of its purpose, also serves as a reference, most ratios do so explicitly, like index numbers, demographic rates, and turnover rates.

4.2.1.2 Adjustment Ratios 'adjust' the (N) figure through division by the (D) figure. These are ratios that 'deflate' value aggregates, that 'de-seasonalize' or 'de-trend' a time series, purporting to eliminate the influence of inflation, of the recurring effects of the seasons of the year or the general trend of a series. Although computed for other reasons, density ratios and 'per capita' figures in a sense also implicitly intend to 'adjust' the (N) aggregate.

4.2.1.3 Causation Ratios, computed to reveal suspected cause-effect relationships between the (N) and (D) aggregates, are e.g. 'input-output ratios', the 'productivity coefficient of labor', the 'productivity coefficient of capital', the 'birth rate' of a human population.

4.2.1.4 Estimation Ratios serve to project on a 'population' the structures found in a sample, to interpolate missing data in a time series and to extrapolate – forecast – a time series into the future.

4.2.2 Ratios Classified by Closeness of (N) and (D)

Of particular importance for the interpretation of a ratio is the closeness of the data in (N) and (D). On one end of the spectrum are ratios in which two aggregates who's subject-matter, geographic area and time period refer to the same subject-matter categories, time-period and geographic area, and were produced by the same survey methods, such a ratio obviously has the same definitions as the (N) or (D) aggregates.

On the other end of that spectrum is a ratio between aggregates whose definitions of subject-matter, time-period and geographic area are different, having none of these determinants in common. Such a ratio resists meaningful interpretation. Obviously, the more alike the definitions of (N) and (D) with regard to their three dimensions, and the methods by which these (N) and (D) figures were created, the more meaningfully can such a ratio be interpreted.

The following six categories classify ratios according to the closeness of the definitions of their (N) and (D) aggregates, beginning with ratios in which (N) and (D) have the greatest affinity, progressing towards those with the least affinity.

4.2.2.1 Density Ratios. The, the (N) aggregate is divided

(a) by its geographic extension in square miles, km^2 or acres, e.g. the 'Number of Persons per Square Mile' or 'bushels per acre'. In these ratios the figure in the numerator is defined for the same geographic area as the figure in the denominator, the number of km^2 or acres of the numerator aggregate.
(b) by its time extension, hour, day or other subdivision of the calendar, such as 'Number of fire alarms in the metropolitan area of a city, per day', for a given month.

4.2.2.2 Distribution Ratios establish the relation between an aggregate or total and its sub-aggregates or sub-totals. This kind of ratio, often presented as a percentage, can be computed for each one of the three dimensions of the aggregate,

(a) by subject matter subcategories, (1) by attribute, e.g. total spending by types of expenditure; all demographic participation rates, (2) Also by continuous or by discrete variables, such as a frequency distribution in which the class frequencies are expressed as percentages of the total.
(b) by regions, such as 'Population of each state of the USA' as a percent of 'Total US population.'
(c) by time-periods, such as monthly figures as a percentage of the year's total.

4.2.2.3 Ratios Between Sub-aggregates

(a) by subject-matter categories e.g. the gender-ratio in a population, e.g. males (females) as a percentage of females (males).
(b) by regions e.g. 'Murders in the metropolitan area of Chicago's district A compared to murders in district B' for a given year.
(c) by time periods e.g. 'work accidents during the second semester as a percentage of work accidents in the first semester' – considering the total number of work accidents for the entire year as the pertinent aggregate of which these two figures are sub-aggregates.

4.2.2.4 Ratios in Which the 'Statistical-Counting-Units' in (N) Are Different from those in (D) but Obtained from the Same 'Real-Life-Objects'. In this category belong most averages e.g. the 'average hourly wage' of a group of workers. The wages paid, in (N) are divided by the 'total number of hours worked', by that same group of workers in (D). The (N) aggregate 'wages paid' is different and obtained separately from the same group of workers as the (D) aggregate 'hours worked' Yet both, (N) and (D) are obtained from the same workers as the 'real-life-objects'.

4.2.2.5 Flow-, Rotation- and Turnover Ratios relate flow or activity data with data about the stock on which that flow or activity occurs, such as the 'labor-turnover rate', the 'inventory turnover', the 'birth-rate'. The (N) and (D) aggregates are less closely related than those in the previous categories of this list

4.2.2.6 Ratios Between Aggregates of Different Real-Life Objects. An example for such a difficult-to-interpret ratio would be the ratio in which at least two of

the three dimensions in the (N) aggregate are different than the definitions of the (D) aggregate. To this type of ratio belong various 'per-capita' ratios, such as 'New books published per-capita' for the population of a country. The aggregates in these ratios are not directly, if at all related to each other.

The foregoing classifications should assist in judging the potential for the meaningful interpretation of a ratio. The inverse of many ratios is also meaningful and should be interpreted in the same way. For example, a sub-aggregate can be expressed as a fraction of the corresponding total aggregate, but the total can also be expressed as a multiple of its part. Similarly 'bushels per acre' and 'acres per bushel' are both meaningful statements about the same situation. So is 'liters-per-kilometer' (or gallons-per-mile) used, and/or 'kilometers-per-liter' or 'miles-per-gallon' to describe the efficiency of a gasoline motor. The form in which a ratio is used is a matter of convenience. The multiplication by 100 or other multiples of 10 does not affect the information contained in a ratio, which makes the distinction between ratios and percentages irrelevant.

SUMMARY OF RATIOS

	Same aggregate total and subtotal	Between Sub-aggregates	Same real-life object but different S.C.U. s	Flow data to stock data; turnover rates	Different real-life Objects
Reference	% Frequency distribution, monthly sales as % of yearly sales, region as % of country	Male-female ratio, car accidents in city A to car accidents in city B		Demographic ratios, inventory turnover	Price index numbers, density Ratios e.g. Bushels per Acre
Adjustment			Most averages hourly wages		Seasonal adjustment, Value time series in $ of 1980
Causation				Productivity ratios of labor and capital	
Estimation	From sample to population	Time series forecast interpolation			

4.3 The Interpretation of Ratios

When calculating a ratio one does not just connect the algebraic numbers of two statistical figures, but one also links the definitions of the (N) with that of the (D). This is the key to an appropriate interpretation of the underlying socio-economic

situation. Let us look at a simple density ratio. Its interpretation depends on size and nature of the aggregates in (N) and (D), and on the purpose for which a ratio is computed. The ratio of the aggregate 'Total US Population', in (N), and the aggregate 'Total surface of the USA, in square miles' in (D), yields the ratio 'Population per square mile in the USA' a density measure that is interpreted exactly like the (N) aggregate 'Total population of the USA.' This ratio, therefore is as undefined with respect to location within the country – whether and to what extent these are agricultural or urban, mountainous or flat, deserts or areas of water – and as undefined with regard to people's characteristics: their nationality, gender, race, age, occupation, income, state of health, etc. The ratio 'USA's population per square mile' is silent about the kind of people, the land and how that population is distributed within that land. Despite their vagueness, the (N) and (D) figures together convey a new statement of some interest about both, the population and the geographic area. The inverse of this ratio 'square miles per person', the amount of 'elbow room' available to each person, expressed in (fractions of) square miles, is just as meaningful – or devoid of meaning . The same holds for the interpretation of ratios of type 4.2.2.2 and 4.2.2.3 of this classification. Although intuitively appealing to characterize this and other ratios, by adding the statement 'on average', the reference to an 'average' really does not explain a ratio, because averages themselves are ratios. Such a statement tries to elucidate one ratio by reference to another, equally unexplained ratio.

Then consider the distribution ratio of the unemployed in a geographic sub-region, in (N), and the unemployed in the entire country, in (D). The 'statistical-counting-units' in (N) and (D) not only differ with regard to the detail with which their geographic location is defined, but implicitly also with regard to the economic and social characteristics of that particular region, that are not present in other parts of that country. Such a 'distribution ratio' reveals something new, that neither of these two aggregates alone reveals. But that 'something', the numeric value of this distribution ratio, is being determined only to the extent of the **less** detailed definition of the involved aggregates. In other words, the information contained in this ratio is dominated by the less detailed, vaguer definition in the (D) aggregate, Fig. 4.1. The larger (D) circle represents the wider, more inclusive, less sharply defined geographic characteristics of the unemployed of the entire country, used as the denominator. The smaller circle, representing the numerator (N), completely contained in the (D) circle, identifies it as a subset with regard to its regional definition and implicitly, also of other special socio-economic features of the unemployed in this sub-region. Subject-matter and time-period are the same for (N) and (D). Even though this ratio appeared to be aimed at the smaller (N) aggregate, it also says something about the bigger (D) aggregate. Yet the information contained in that ratio is as vague as the more general, less detailed descriptive definition of unemployment in the entire country.

Another example is the 'crude labor participation rate'. The (N) aggregate contains the 'economically active' population above a certain age; the (D) aggregate refers to the entire population. That ratio is also dominated by the more inclusive, less sharply defined, and vaguer definition of the (D) aggregate, although their geographic area and time period are identical.

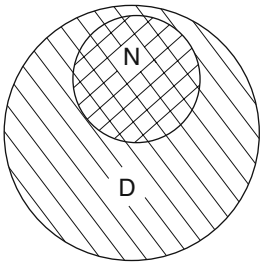

Fig. 4.1 Geographic location (and socio-economic facts) in (N) more narrowly defined than in (D)

Consider one more example of this type, 'number of accidents per thousand hours worked'. Suppose the accidents happen in a large manufacturing firm, are of a specific kind of injury, and happen only to men who work in one specific occupational category, in only one of its plants. The numbers of these accidents that happen during each year in this plant are the aggregate in (N) of this ratio. The hours worked by all workers, used as (D), also include the hours of the workers who are not in this occupation and were not exposed to the risk of this type of accident. It is a more inclusive, much wider statistical aggregate that is undefined with regard to the occupation of those workers in that firm, their gender, position, age, wage, department in the plant, etc. When interpreting the resulting accident rate, the degree of vagueness of the larger, not of the narrower aggregate, dominates the interpretation of this ratio. It refers then to all the workers, in all plant locations even if they cannot possibly have that kind of accident which was included in the numerator figure. The result could appear as 0.0002 accidents per hour worked, or more popularly, as 2 accidents per 10,000 hours worked by all workers. It would obviously make more sense, to use as (D) only the hours worked by the workers exposed to that kind of accident.

In ratios of types 4.2.2.4–4.2.2.6, definitions in (N) and (D) do not coincide, either with regard to subject-matter, with regard to geographic region, or with regard to the time-period of these aggregates. Consider a birth rate (4.2.2.5). The (N) aggregate represents the number of birth-events during a given year, the (D) aggregate is the numerically much larger mid-year inventory of that population. Even though the (D) aggregate is limited to a precisely specified, point in time, this ratio is considered to be valid for the entire year, Fig. 4.2. Here the definition of the time interval in the (D) aggregate is dominated by the wider definition of the numerator aggregate (N) with regard to time. The geographic area is the same for (N) and (D) but the subject-matter is not. In (N) are the birth-events, – not the persons born properly speaking – in (D) an estimate of all persons. The choice of the point in time for that population count, within that year can make some difference in the numeric value of that ratio. Although at first blush it seems to make sense to relate births to the human population, but on careful thought, the diversity in age, marital status and other pertinent population characteristics makes the total human population a less than ideal reference, and a 'cause' only in a vague, unspecified sense. That vagueness is also the setting in which to interpret that birthrate.

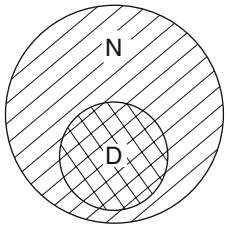

Fig. 4.2 Time in (D) defined more narrowly than in (N)

Then consider a time series of the sales of a business firm. Assume that the aggregate in year t is used as the base period, in (D), The sales figure contained sales of products A, B, C and E. In the later period, t+i product A is no longer carried and is not in the sales aggregate, (N). Included, however, is the new product F, which was not yet sold during the base period. That aggregate of the later period now contains products B, C, E and F. When computing the ratio between these two sales aggregates, its interpretation corresponds to the union of these partly overlapping subject-matter definitions. That ratio refers equally to all five products A, B, C, E and F, even though only sales of products B, C and E are contained in both, (N) and (D) – a situation common in time series. Graphically the sales figures in that ratio correspond to the entire shaded area, Fig. 4.3. The resulting number of this ratio is to be interpreted as less clearly defined than either (N) or (D), viz. as the union of (N) and (D).

No detail with regard to the product-mix can be distinguished in such a ratio. The conceptual situation shown in Fig. 4.3 is typical of time series, particularly of the chained price-index- numbers. The less the subject-matter definitions of (N) and (D) overlap, the less clear is the interpretation of the corresponding ratio. More about this in Chap. 7, on index numbers.

Finally, consider a ratio in which none of the three dimensions in the (N) aggregate is co-extensive with the definition of the (D) aggregate. An example would be a ratio between the GDP of the USA in 1985 and the GDP-equivalent for the USSR for 1980. That ratio, regardless, which is the (N) or the (D), attempts to compare two distant countries, with different climatic conditions, quite different economic systems, for different years, produced by different national accounting systems and by differences in the confidence these data deserve shown in Fig. 4.4.

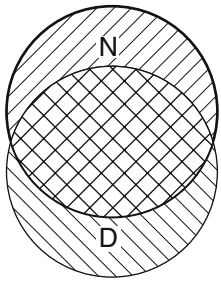

Fig. 4.3 Overlapping but not identical subject-matter definitions of (N) and (D)

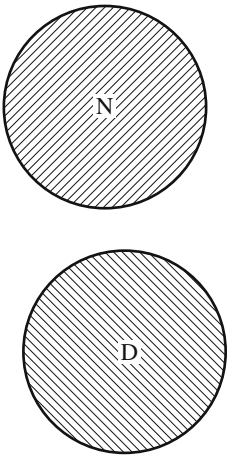

Fig. 4.4 All definitions of subject-matter, time and area are different

In general, the more alike the definitions of (N) and (D) with regard to its three dimensions, the more meaningfully can such a ratio be interpreted.

Contrary to widespread belief, the algebraic operation of division does not change the real-life objects in (N) nor those in the (D) aggregates. That is as true for any 'adjustment' of data, be that the 'deflation' of a financial statement expressed in the currency of a base period, or a time series, the data of which are expressed in units of the currency, 'deflated' with a consumer-price index-number, or its 'de-seasonalization' with a seasonal index. Ratios **do not alter** the reality of the aggregates in the (N) or (D). Its original structure, including its 'real-life-objects' and the corresponding 'statistical-counting-units', remain unaffected by the division.[5] The intuitive impression, though, is correct, that something new has been created. That 'something,' however, is not the 'deflated,' 'de-seasonalized,' or otherwise 'adjusted' figure in the numerator, but that ratio itself, as something new, with features of its own.

4.4 Ratios Between Ratios

So far, we have only dealt with ratios between aggregates, that one may call 'first-order ratios'. If the numbers in the (N) and in the (D) themselves are ratios, the result would be a 'second-order ratio'. If (N) and (D) each are a ratio between ratios, viz such 'second-order ratios', such a ratio would be a 'third-order ratio'. Evidently, the guidelines for the interpretation of 'first-order ratios' – the ones discussed so far – are also valid for higher-order ratios with the added complication that the definitions of more than two aggregates must then be considered. The 'deflated' and 'adjusted' aggregates in reality are ratios of a higher order. They are not usually recognized as ratios because they appear expressed in dollars, tons, etc. in the guise of regular statistical data properly speaking.

When their structure is not symmetrical, when the figures in (N) above and(D) below the main fraction-line are different, e.g. the (N) of such a ratio is a simple income aggregate, but the (D) of that ratio is itself a 'first order ratio' e.g. between the price-aggregates of a price-index number, then one would refer to it as a 'partial second order ratio' which would be a 'partially higher-order ratio.' Another example of a partial second-order ratio is 'real net income per capita' in which the (N) but not the (D) is a 'first-order ratio'.

An example of a complete 'second-order ratio' is 'per capita income of region A as a percentage of per capita income of region B'. In this case both, (N) and (D) each are themselves first-order ratios. The resulting second-order ratio is conceptually – but not necessarily also numerically – identical to a first-order ratio between the pooled incomes of both regions in (N), and the pooled populations of both regions in the (D). Both, the first- and second-order ratios refer to the same situation and their interpretation should be analogous. Also all Log-Linear and Logit models are really 'second-order ratios'.[6]

The 'index of unit-labor-costs' is computed as a third-order ratio.[7] A ratio of the fourth order is the 'true net migration rate' $Rbar(i) = Mbar(i)/Pbar(i)^2$. This ratio – and many others as well – looks deceptively simple because of the custom of introducing new symbols at each successive stage, which leads to a simpler, uncluttered expression.[8] It is not uncommon for economists and sociologists to work with ratios of the fourth and even higher orders. The highest level ratio I have encountered was the partial eighth-order ratio of the average change in the irregular to the 'average change in the trend cycle'.[9] The true contribution of such ratios to the understanding of a socio-economic situation really is difficult to assess.

To give one more illustration of the potential dangers and the problematic of interpreting ratios of a higher order, consider the ratio 'real average wages of miners in city X,' which had actually been computed in a country of South America. The wage totals had been averaged and 'deflated,' by a consumer-price-index. The latter was constructed with the family-spending-pattern of one single working class family, whose wage earners were not miners, in the distant capital of that country. The kinds of goods included, as well as their price levels differed substantially from those in that mining district. The price deflator was also from an earlier time period. The ratio 'real wages of miners in city X in year t' then contained four differently defined aggregates.[10] The resulting 'real wages' of this seemingly compact social group intermixed their nominal, average wages with the price and consumption situation of a very different socio-ethnic group that was far removed in time and geographic region. The resulting ratio therefore reflected the situation to be interpreted in a hazy, very diffuse and unspecified way. Worst of all, it was not even recognized as a ratio, let alone, as a ratio of a higher order. These 'real average wages' were much less real and useful for policy analysis than the un-deflated wage averages of those mine-workers would have been. In an apparent effort to avoid this pitfall, labor productivity ratios are now computed in the USA as ratios between a production index number and an index number of hours worked.[11] Unfortunately analysts are deceiving themselves about the value of that complex and vague 'second-order ratio.'

4.5 Other Kinds of Statistical Materials

In business, economics and the social sciences, there is one other type of statistical material which is neither a ratio nor an aggregate, but based on the latter. It is the difference between an aggregate of beginnings, and an aggregate of terminations, representing a class of flow data that, because of their important uses, deserve to be mentioned separately. It would be very difficult, for example, to ascertain directly the amount of currency in circulation for a country, at a given point in time. It is more promising and less troublesome to estimate the phenomenon 'money in circulation' from the stock at the beginning of the time interval, adding the amount of 'new money issued' and subtracting 'money withdrawn from circulation' during that period. Each is a genuine statistical aggregate, yet the end result, their difference, is not. Accounting figures about flow phenomena in a business firm also are based on differences between aggregates, such as data on inventories and investment. The interpretation of this kind of material hardly poses new problems.

Notes

1. This Chapter follows: Winkler, Othmar "The Meaning of Ratios in the Social Sciences – a Sketch of a Foundation of Descriptive Statistics" *Proceedings of the 39th Session of ISI*, Vol. XLV – Book 2, Vienna 1973, pp. 580–586 See also the strong endorsement by: Walter Kraemer "Statistik in den Wirtschaftwissenschaften" *Allgemeines Statistisches Archiv*, 80, 2 2001 *Translated:* "most of the influential empirical studies in sociology and in the economic sciences do not derive their influence from sophisticated statistical procedures but from an unexpected approach, from the juxtaposition of series of data so far not seen as related, a new way of looking at the subject ... most of the great contributions of statistics to progress in the economic sciences were not achieved through advanced (statistical) methods but through rather descriptive, mathematically undemanding, simple tools." :(original text:) ". . .schoepfen die meisten einflussreichen empirischen Studien in Soziologie und Wirtschaftswissenschaften ihren Einfluss nicht aus raffinierten statistischen Verfahren, sondern aus einem ungewohnten Sichtwinkel, aus der Gegenueberstellung bislang getrennter Datenreihen, aus einer frischen Betrachtungsweise ihres Gegenstandes. . .sind die meisten grossen Beitraege der Statistik zum Fortschritt in den Wirtschaftswissenschaften nicht durch fortgeschrittene Verfahren, sondern durch eher deskriptive, formal anspruchslose Werkzeuge geschaffen worden. In den uebrigen Sozialwissenschaften treten die statistischen Methoden sogar noch ausgepraegter hinter den Daten selbst zurueck. . .". p. 196.
2. Winkler, Wilhelm, *Die Statistischen Verhaeltniszahlen, Eine Methodologische Untersuchung*, Franz Deuticke, Wien (1923). This unusual and thorough treatment of ratios has a different focus than the discussion in this chapter.
3. See e.g. Walker, Helen M., "Degrees of Freedom" viewing data as mathematical structures in an n-dimensional, abstract sample space. in: A. Haber, R. Runyon, P.Badia, editors, *Readings in Statistics*, Addison Wesley Publ. Co., Reading, Mass. 1970, pp. 130–146.
4. Kendall, M.G., "On the future of statistics – A second look", *Journal of the Royal Statistical Society,* Series A, Vol. 131, Part 2 (1968), p. 184.
5. Winkler, Othmar, op. cit., 1966, pp. 356–357.
6. Knoke, David, Peter J. Burke, *Log-Linear Models,* Sage University Papers #20. Series: Quantitative Applications in the Social Sciences. Ed. John L. Sullivan. Sage Publications, Inc. Beverly Hills, 1980.

7. U.S. Dept. of Labor, "Unit Labor Costs in the US and Ten Other Nations 1960–71", *Monthly Labor Review*, Washington, D.C. (July 1972), p. 6.
8. Zachariah, K.C. "A Note on the Census Survival Ratio Method of Estimating Net Migration" *JASA, Journal of the American Statistical Association,* Washington, D.C. (March 1962), pp. 175–183.
9. Early, John, "Measures of Variability for Seasonally Adjusted Series", *Monthly Labor Review*, Washington, D.C. February 1972, pp. 66–67.
10. This account was a personal communication by an employee of the Dirección General de Estadística of Ecuador, who at that time was my student at the postgraduate center CIEF, of the Organization of American States, in Santiago, Chile, when I was teaching a course on Estadística de Producción, Precios y Trabajo', 1954–1959
11. U.S. Dept. of Labor, "Output per Man Hour Measures: Industries," Chapter 26 in *BLS Handbook of Method*s, Bulletin 1711, Washington, D.C. (1971), pp. 219–226. and the 1997 version of *BLS Handbook of Methods*, Ch. 10 "Productivity Measures: Business Sector and Major Sub-sectors", pp. 89–109.

Chapter 5
Interpreting Longitudinal Data, Part 1 – Looking to the Past

Thirty years ago, before old monk had studied Zen, he saw the mountains as mountains, waters as waters. Later when he came to know a good master and was first initiated into Zen, he no longer saw mountains as mountains or waters as waters. Now that he had got a resting place, he again sees that mountains are only mountains and waters only waters[1]

5.1 Nature and Purpose of Socio-Economic Time-Series

Let me start this chapter with a piece of Chinese wisdom, not because of its superficial similarity to the 'mountains' in the graph of a time-series, but because it expresses the evolution of my own thinking on this matter. The present state of time-series analysis seems captive in the second phase mentioned by Chin-Yuan Wei-Hsin, where things as vital as waters and mountains, under the guidance of learned teachers, are understood to be something else. In analogy, statistical concepts and ideas that have ignored the true nature of socio-economic data dominate our approach to their longitudinal analysis. Chapters 2, 3 and 4 laid the foundation for a more realistic approach to time-series, arriving finally, like the old monk, at the unromantic realization about the true nature of 'mountains and waters'.

Time-series, because of their importance in understanding the evolution and development of today's business, economic, and social situations,[2] are treated in this book ahead of frequency distributions which are customarily given preferred treatment in textbooks and presented ahead of time-series but are less important for the interpretation of economic and social situations and the corresponding data.

It is natural albeit unfortunate that statisticians have paid attention only to the numerical aspect of time-series, making the search for patterns of variation the main objective. They seem to forget that socio-economic time-series are the quantitative history of a situation. The interpretation of time-series therefore, though expressed in numbers, ought to be above all historiographical. This has two aspects:

1. The nature of the phenomena and their definition, the corresponding 'real-life-objects', and the detail of the survey and computing procedures have evolved and changed over time. These are matters that are to be considered when attempting

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_5, © Springer-Verlag Berlin Heidelberg 2009

the interpretation of a socio-economic time-series, particularly if that series reaches far into a past of which the historical conditions must be expected to have been different.
2. The data in a time-series are embedded in a wider historical landscape of other events.

Time-series data, therefore, are not simply a sequence of algebraic numbers, but are part of the general history, presented in statistical terms. Pertinent other historical information must also be considered, e.g. a strike that affected the industry in certain regions or the incipient globalization. The present numbers-only statistical approach to time-series leaves no room, and is ill prepared, to take into consideration important other historical non-quantitative facts. This matter will be further discussed in the next chapter on extending a time series into the future.

5.2 Types of Time-Series

The differences in statistical materials that lead to different kinds of time-series are generally ignored. On the other hand, the distinction between 'aggregative' (cumulative) and 'non-aggregative' (non-cumulative) series[3] is not helpful because all socio-economic data are aggregates at least in one of their three dimensions, or are based on such tri-dimensional aggregates. There are then at least three different types of time-series:

1. Time-series of data in which the time-dimension is aggregated (such as monthly sales data)
2. Time-series of data in which the time-dimension is not aggregated (such as inventories and censuses), although these data are aggregates with regard to category and geographic region, and
3. Time-series of derived types of statistical materials,

   a. data that are the differences between aggregates, such as net-earnings, net investment or money in circulation.
   b. data that are ratios, such as real wages, sales in dollars of 2000, per-capita income, index-numbers of prices and quantities, and social indicators.

The treatment of these types of time-series differs. The following discussion will be limited to the prevalent first two types of time-series. Time-series of type 3 b – ratios and index-numbers – will be discussed in Chap. 7.

5.2.1 Time-Series of Data with the Time-Dimension Aggregated

Most time-series of socio-economic data are of this type. These data consist of accumulations of 'statistical-counting-units' during time-intervals, e.g. the monthly data of births. These data cannot be treated like 'measurements' in the natural sciences,

as points in time, in subject-matter and in geographic place. An understanding of aggregation and of statistical aggregates is essential to the following discussion. Each of the three-dimensional larger and smaller 'conceptual spaces' of an aggregate can be used to form a time-series. As an example, the conceptual space 'all new construction,' as yearly aggregates, Fig. 5.1, or as monthly aggregates Fig. 5.2.

Then the geographically smaller sub-aggregate 'All yearly construction activity in the Northeast Region of the US' can be established separately as a time-series. Time-series of other subsets can also be formed, e.g. by more narrowly defined construction activities, such as 'private residential construction.' The latter can be presented as a time-series for the entire country, or only for some of its regions, as a time-series for construction activity during the entire year as yearly data, or as a time-series for data covering shorter time-intervals. There is a hierarchy of possible time-series of such three-dimensional sets, super-sets, and sub-sets concerning an industry, like the construction activities. Each of these possible subsets can be presented as a separate time-series. The corresponding 'statistical-counting-units' – building permits issued, or value added to construction in progress, or completions of construction projects, etc. – must be considered to be scattered unevenly throughout the inside of these conceptual spaces. Further information can be obtained by forming sub-aggregates, Fig. 5.2.

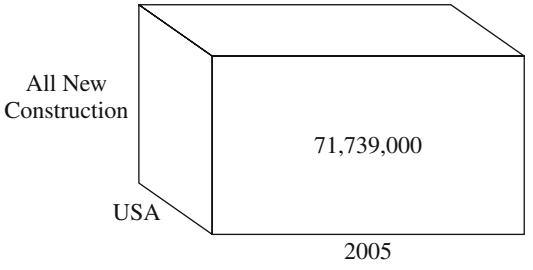

Fig. 5.1 All yearly construction activities, USA

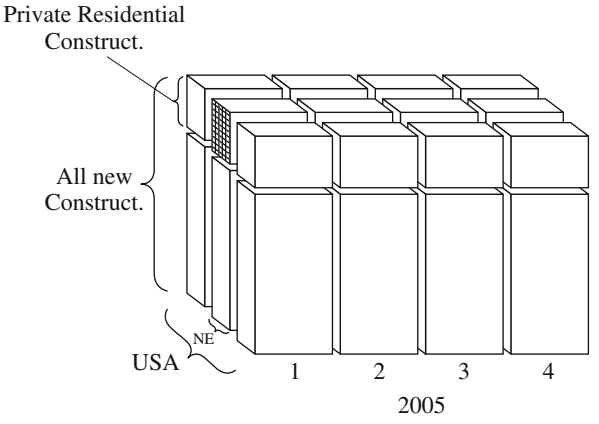

Fig. 5.2 Types of yearly construction, USA

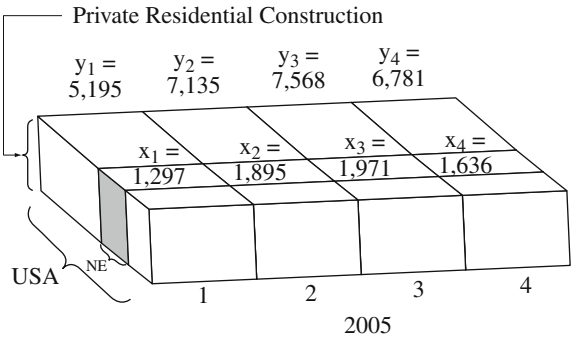

Fig. 5.3 Quarterly private residential constructions, USA

Quarterly time-series of the sub-aggregates of private construction activity can also be formed, Fig. 5.3. These figures are totals that result from the progressive accumulation of the statistical counting units over the chosen time-interval, reaching their full count at the end of each time-period. That is where the resulting totals really belong, not in the center, where these totals are usually placed. The detailed information is usually not available that would be necessary to construct a line-graph of the accumulation over time inside an aggregate. It would show an irregular increasing trace because the statistical counting units accrue at irregular intervals, forming a kind of ogive for each of the sub-time-periods, Fig. 5.5. A similar picture would emerge in aggregates of any length of time, such as a month, quarter or a year. The fact that the total belongs at the end of each time-interval could be used to advantage when applying moving averages.

This can further be detailed, as quarterly private construction by major geographic regions, Fig. 5.4.

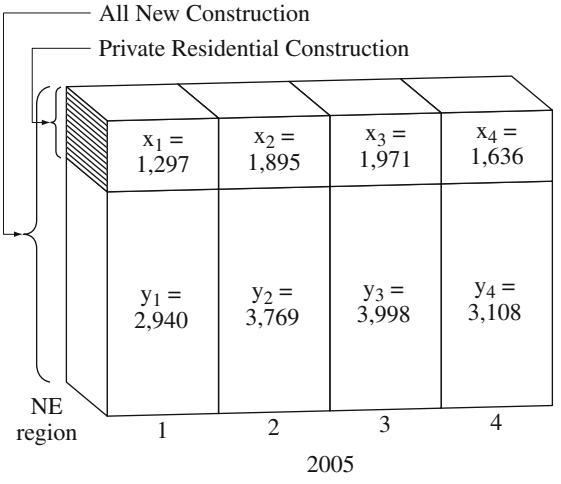

Fig. 5.4 Quarterly private constructions for the North East region

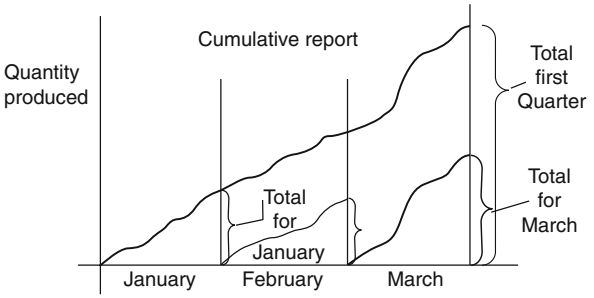

Fig. 5.5 Aggregation as accumulation over time

5.2.2 Time-Series of Inventories and Censuses

The misleading term 'non-aggregative' is used for time-series of business inventories and censuses. In this category also belong series of the yearly financial reports – on balance sheets of corporations – on monthly reports of unemployment, or on quality control samples taken at regular intervals. The label 'non-aggregative' is misleading because, as already pointed out, aggregation is always involved, be that by subject-matter, by geographic area, and often by both. Data of this type of statistical material should be used as benchmarks for related data that are occurring and collected continuously.

The arrangement of such inventory-type figures into a time-series is based on the implicit belief that change in the underlying socio-economic situation between inventory dates, is gradual and predictable, Fig. 5.6.

Yet, in time-series of this kind one inevitably faces problems of unexpected turns, disruptions and discontinuities in the underlying socio-economic situation. The direction, abruptness, or the cyclical nature and intensity of change usually are not reported. Census and inventory data, therefore, should not be used as time-series! But if such inventory-type data are assembled as a time-series, they must be interpreted with great caution.[4] The hypothetical inventories of Fig. 5.6 seem to indicate a declining trend, drawn as a dotted line, when in reality – the development between each such inventory figures is usually unknown and not part of the graph – the actual development could have an upward trend, the opposite if what that time-series of inventory figures seems to convey. An exception to this occurs when such inventory-type data are combined with the continuous information on

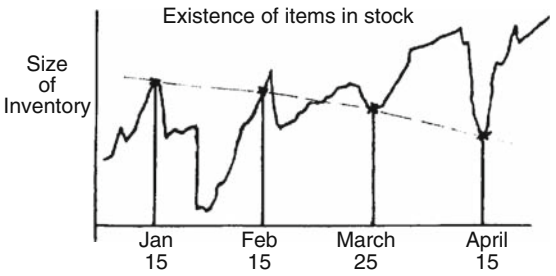

Fig. 5.6 Inventories with unreported developments in between

new entries and exits from the stock of the inventory, like in the constant updating of the inventory of merchandise with the electronic check-out information of items passing through the cash register of retail businesses. In all other situations the best one can do is to estimate the speed and direction with which the underlying socio-economic phenomenon may have evolved between inventories or censuses. That information can then be used to interpolate a value of such a series for points of time between these inventories or censuses, based on pertinent available statistical and other non-numeric information.

5.3 The Problem of (Dis-)Continuity

In this fast moving age, continuity in the socio-economic phenomena can never be taken for granted. The lack of continuity in a time-series, however, can be a serious matter. When interpreting any time-series one must always be prepared to encounter change that could be drastic, e.g. when using the yearly household-expenditure surveys taken in October, but especially when the data are taken at larger intervals, such as the decennial censuses. Discontinuity not only depends on the length of time between each two data (aggregates) but also on the speed and magnitude of the unreported development in inventory-type time-series. A shorter time interval does not necessarily mean fewer changes. The monthly data of business inventories of products with a high turnover could be more discontinuous than e.g. the information of a time-series of the quinquennial census of an agriculture that changes only very gradually.

Discontinuity can also be present in less obvious ways, when the 'statistical-counting-units' in a time-cumulative series are collected only during a short part of each period e.g. during the second week in each month, for a monthly series on hourly wages. In that case the de facto statistical coverage of this socio-economic phenomenon 'cost of labor' is only partial and discontinuous. The monthly data in retail price indexes are usually the listed price quotations taken in a few types of stores, in selected locations, only once each month, e.g. on the Thursday in the third week of each month. The obvious discontinuities in such series may not be noticed when the data are appearing in aggregate form, e.g. as monthly figures Given the importance of discovering the dynamics of a socio-economic phenomenon, the time-dimension should be covered as completely as possible.

Interpreting a discontinuous time-series is like attending a concert with sound-proof earplugs, but removing these briefly, say, every thirty seconds. The less changes there are in the music, the closer will that which one perceives under such circumstances, resemble the music being played. During more rapid passages, however, the music one would perceive would differ – and make less sense – than that which is actually being performed. In contrast, analysts often have been unaware that they too have perceived a socio-economic development in such a discontinuous, erratic manner, with 'statistical earplugs'.

Discontinuity in a time-series is not always obvious and its magnitude usually unknown. Inventory-type aggregates in particular report the evolution of socio-

economic phenomena in a manner that can easily be misinterpreted. If there is reason to believe that the omissions are substantial, then such a sequence of inventory-type data should not even be attempted as a time-series. But discontinuity can be embedded in all kinds of data and the interpreter of time-series must always be on the alert for hints to the possible presence and seriousness of hidden discontinuities.

5.4 A Critical Look at the Basic Model of a Time-Series

One should expect that a neophyte, who is learning to interpret a time-series, would be taught to understand the data as statistical reflections of the social or economic events that shaped it. The current approach to socio-economic time-series, however, insists on 'isolating' the supposed 'components' and also 'eliminating' their effects, regardless of whether this really contributes to the understanding of the economic or social situation. There have been refinements to this model but no real changes to this approach. The meaning and relevance of these components of a time-series have rarely been explored.[5]

5.4.1 The Secular Trend

Statisticians and social scientists seem agreed that a trend is the 'component' that is most important, most meaningful and easiest to determine. A trend line is usually started at some arbitrarily selected point in the time-series. Although there is always a justification why one particular point in a time-series, and not another, has been chosen as the starting point, it is only one among many other possible alternative starting points. It could be imposed by circumstances when the data source omits an earlier part of the series, or when statistical coverage started later in the history of that development and earlier data are not available. More likely, it will be arbitrarily chosen. Whichever the case, it is important to realize that for every different starting point the computations yield a different trend-line.[6] In addition, the trend value changes with each new figure that accrues to the series. If one considers this, it is amazing that people believe that the one trend-line which they computed is the only possible line, **the trend** of that time-series. But there is more to be said about trend.

Consider the optimal circumstance: some economic activity started in 1975, ended in 2005, with the statistical record complete and available. Even in such a situation, the trend cannot be determined as a unique, clear line for the following reason: Moving averages are based on the same smoothing principle[7] as the corresponding moving totals that are aggregates. The placement of their numeric values, on the y-axis of the graph of that time-series, is a single unique value. Its placement on the x-axis of that graph, however, is undetermined within the time-interval of each moving total or moving average. The intuitively preferred placement on the horizontal axis is the midpoint of each time-interval. But that ignores the basic indeterminateness of an aggregate for the reasons outlined in Chap. 3. The placement on the horizontal x-axis of the numeric value of the moving average is not uniquely

determined within the time interval of the corresponding aggregate (Figs. 3.12, 3.13 and 3.14). despite the fact that most accumulation would be completed by the end of each period, Fig. 5.5. The numeric values of the moving averages could be placed anywhere along the entire length of the time-interval of a newly computed moving average. The numeric value y e.g. of a three-year moving average, could be placed anywhere on x, from January 1 of the first year to December 31 of the third year.

The centered moving average is generally preferred. But other off-center moving averages could equally be justified.[8] The most likely point on the x-axis, though, on which to plot the corresponding y-value would be at the endpoint of each time-interval. When applying this understanding of the uncertainty on the x-axis of the corresponding y-value on the vertical y-axis for the computation of the trend-line, one would get a sheaf of different, parallel trend-lines that are bounded by the lines resulting from the placement of the lowest and highest y-averages on the vertical axis for a specific x-value. These limiting y-values would determine a trend-zone.[9]

A mathematically fitted linear or non-linear least-squares trend-line can be considered the 'limiting' value of a moving average of aggregates whose time-dimension has been extended to an extreme. The points of a graph of moving averages forming a trend-line, therefore, can be considered an approximation to a 'least-square' line. The closeness of its approximation is determined by the number n of the time-series data in the n-term moving average. A trend-line fitted by least-squares, in principle, should be considered to be as undetermined as a moving average trend-line of comparable smoothness. Notice that this indeterminateness of the trend is due to aggregation, it is unrelated to the uncertainty margin due to sampling error.

How wide is such a trend zone? No general statements can be made, except that for any given value on x there will be a number of different $\bar{y}$ values, e.g. four for a quarterly moving average, twelve for a 12-month moving average, ideally to be placed at the end of each original period. This is determined by computing the n different trends for an n-term moving average. For a larger number of time-periods included in a moving average, experimentation with 2, 3, 5 and 10-term moving averages has not shown that such a trend-zone approaches a limit. One can conclude that the trend-zone for an algebraically fitted least-squares trend-line will not only be very wide but also drawn on the graph as a barely perceptible zone or canal to indicate symbolically its vague meaning of being 'a broad sweeping motion through the data'.

What is the socio-economic meaning of a 'trend'? Each aggregate in the time-series about e.g. 'construction activity', is the result of the building activities in many regions with different economic characteristics, and within each region, of many different kinds of construction, each behaving differently in each one of these regions. The over-all development of these many, rather independent parts, becomes an amorphous something as the net effect of these separate developments. In the case of construction activities, the overall trend is the outcome of a multitude of separate and independent activities that are the result of regional economic and climatic conditions. There are, of course, some factors common to all building activities throughout the entire large country. In general though, the higher the level

of aggregation with regard to time, region and type of construction activity, and the smoother the resulting trend zone, the fewer are the common forces that drive the different building activities in different regions and the less influence they should be expected to exert. The trend does not represent a force that has actively shaped the historic development, nor does it indicate the direction which the series would have taken in the absence of differences in the development of regional building activities. The statistical trend is the passive residue of the aggregation process, the net-effect of divergent, decentralized developments at the local and regional levels and evidently not the result of **one** single, pervasive economic force. It is difficult to say what exactly this remainder, the trend, means in terms of socio-economic reality.

It would be important to express also visually this lack of concrete meaning of the trend by tracing it as a not very visible, hazy, gray zone. Regardless of how it is computed, the trend should be plotted into the time-series graph as a hazy, barely visible and noticeable zone, sweeping through the clearly marked original data.[10] This would more realistically indicate the scant meaning that can be attributed to these general, vague forces and would be a reminder of the negligible contribution trend analysis really makes to the study of the development over time of socio-economic phenomena. Nobody can tell what the series would have looked like if certain facts in the economy had been different. The re-arrangement of class boundaries of the aggregates in a time-series does not change the economic and social reality that had happened, and nothing has been eliminated e.g. by 'de-trending' a series. All statistical counting units in the time-series remained in their places within these aggregates. Only the focus has been changed, away from the monthly or quarterly detail which is then no longer visible.

5.4.2 Seasonal Fluctuations

Next, the isolation and elimination of seasonal fluctuations, that are superficially associated with the calendar, is considered useful. 'Seasonal fluctuations' are caused by physiological, climatic, psychological, and institutional factors, not the formal divisions of the calendar. Each one of these contributing factors has a development of its own. This 'seasonal component', then, is an aggregate of the seasonal components of the various pertinent factors, each with its own, different development. These individual sub-patterns of seasonality themselves are subject to change at different speed. The seasonal pattern of steel production, for example, is a weighted average of the seasonal patterns of all the uses of steel. In times of prosperity the seasonal pattern of steel for automobiles is weighted much heavier relative to that of tin plate for canning food, than is the case in times of depression, because sales of canned food fluctuate less during a business cycle than do sales of new automobiles. Consequently the seasonal pattern shifts with the ups and downs of the economy because of shifts in the proportions of steel going into uses with different seasonal patterns. Since ordinarily one of the first steps in studying business cycles is to make a seasonal adjustment of the data, there is danger of either eliminating part of the cyclical fluctuation or of interpreting as cyclical some fluctuation that

is really seasonal.[11] Statisticians treat this heterogeneous mix as one single source of fluctuations, assuming either a stable pattern or one that changes according to a discernible regularity.[12] This kind of analysis in monthly and quarterly time-series reveals little about the underlying socio-economic situation.

It is customary to 'remove' this seasonal component with a twelve-month, or four-quarter moving average by dividing each value by the monthly or quarterly seasonal pattern that has been distilled from the monthly or quarterly fluctuations of the series. What is left in a 'seasonally adjusted' time-series is essentially a short-term trend that is ambiguous, for the reasons discussed under trend analysis. Eliminating these seasonal fluctuations does not reveal the values which this series would have had in the absence of those seasonal forces. It cannot be known what the figures of a time-series would have been if certain facts in the climate, shopping customs, or the civic and religious holidays in that society had been different.[13] The re-arrangement of statistical numbers does not change or 'eliminate' the underlying social and economic facts, although it allows highlighting those effects in the series that reach beyond a moving twelve-month, or four-quarter period. Every 'statistical-counting-unit' that was reported and became part of the time-series remained in the same spot within the three-dimensional structure of each aggregate. What is changing, though, is the focus, by ignoring the monthly or quarterly detail.[14]

When 'isolating' these fluctuations, one attempts to bring two differently focused pictures of the same socio-economic reality into a simultaneous focus. One takes the broader-focused, de-seasonalized time-series as the reference-horizon and plots on it the monthly figures that still contain the seasonal effects. Because these two pictures lie at different focal planes, one cannot see both in focus; one of them always perceived as blurred[15] Yet data users believe that they can see both simultaneously in focus, the more detailed, 'close-up' nearer picture of the original data, and the more general, more 'distant picture' of the de-seasonalized or de-trended series. Figure 5.7 illustrates this. A hazy canal of seasonally adjusted figures, really a kind of trend, sweeps through the clearly marked original data. If, however, the de-seasonalized data are perceived clearly as the line of reference, then the original data that contain the effects of seasonality can only be perceived as a hazy canal,

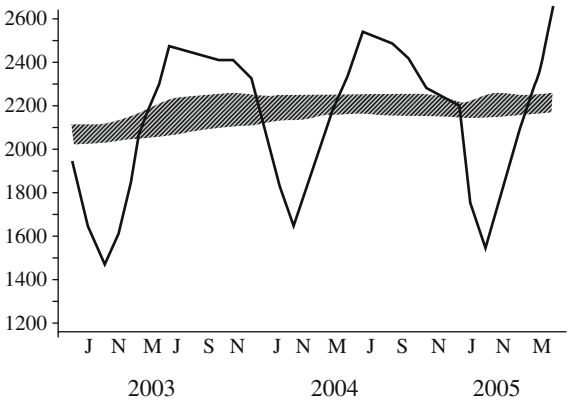

Fig. 5.7 Clear seasonal data against fuzzy de-seasoalized data

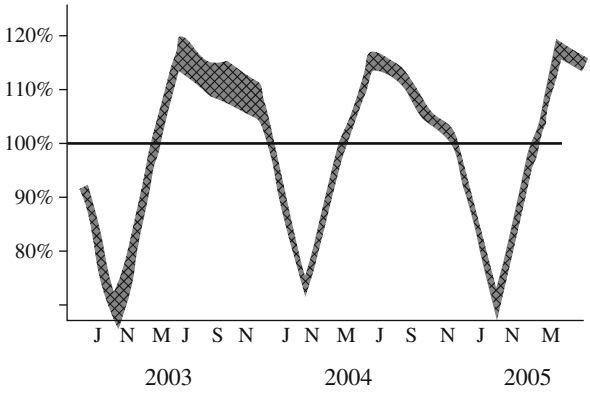

Fig. 5.8 Fuzzy seasonal data against clear de-seasonalized trend

as if out-of-focus, Fig. 5.8. One can see only one of them in focus, not both. Just try to look at a pencil you are holding and at the same time also at a picture further away on the wall behind. One can either see the pencil clearly while the picture at a distance appears blurred and out of focus. Or one focuses on the picture in the distance behind the pencil, then the image of the pencil close-up will be out of focus. We are hardly aware of this limitation because our brain is accustomed not to attempt to focus simultaneously on objects at different distances. The data and computations for the two graphs are shown in Tables B.1, B.3, B.4 and B.5 of the Appendix to chapter 5.

Rather than attempting to interpret the seasonality of highly aggregated time-series, it is more rewarding to study separately, for each region and socio-economic activity, the possible relationship between a given time-series and various factors that are known to have an impact on the investigated phenomenon, like the habits of consuming, buying, investing, etc. In the case of construction activity such factors to consider would be meteorological facts of temperature, humidity, rainfall, sunshine and the corresponding uneven hours actually worked. Their effect on the construction data and their possible relation to the divisions of the calendar should be used to trace their influence in the seasonal short-term fluctuations of the series.

5.4.3 Random Fluctuations

Everyone seems to know what 'random fluctuations' in a time-series are, yet an explanation is hard to come by. In reality, these irregular, fluctuating movements in a time-series are the leftovers after having 'removed' trend and seasonality from a time-series. Size and shape of these unexplained residuals depends on the type of trend-line and the seasonal pattern that were chosen. Besides unintended flaws that can happen during the statistical process, errors of copying, errors of arithmetic and misprints, other fluctuations are the results of events that on occasion affect society, like a strike, or a natural disaster like a disruptive hurricane that can affect the situation reflected in a time-series.[16] The irregular fluctuations due to these occurrences

are assumed to behave randomly, approximately following a Gaussian distribution. But these so-called random fluctuations in a time-series are determined in shape and size by the choice of trend-line and seasonal pattern.[17] Obviously, these 'random fluctuations' are not the result of specific 'random' events in society but the result of the adopted time-series model. 'Random fluctuations' are the waste-basket for the unexplained remainder variations, after the 'components' have been 'isolated' and 'removed' from the data. Random fluctuations generally defy interpretation and have no practical value for the interpretation of a time-series.

This model of socio-economic time-series with 'components' that can be disassembled, as one would unscrew and remove parts of a machine, dominates statisticians' thinking. It seems to obviate the need to delve into cumbersome detail of the data and the need to search for the regional and subject-area-specific significant socio-economic events underlying the data. The real contribution of trend and seasonal analysis is vastly overrated and is not as obvious as is believed. Random fluctuations will be further discussed in Sect. 10.5.

5.4.4 The Business Cycle

Until a few decades ago much of the time-series literature dealt with Business Cycles. Although they never were really 'cycles' in the physicist's sense, they are becoming even less so in view of a growing depression-consciousness in government and the private sector of the economy. During the last decades economists have learned much about economic development, have acquired more experience and are less reluctant to steer the economy in order to minimize the threat of inflation or recessions and if possible, to eliminate downswings altogether. The recent growing integration of the economies of all countries is creating new situations that have never before existed. Obviously the idealized stereotype of the 'business cycle' is rapidly becoming a thing of the past. How to handle statistically the internationally connected irregularities of economic development is beyond our present knowledge.

As early as 1955, G. Colm anticipated that: "I do believe that the changes which have taken place make it likely that the business-cycles of the future will have little similarity to those of the past. It is most likely that both the duration and character of the cycle have basically changed. Therefore, I doubt that direct analogies with the past can give us much help in identifying our present position in the cycle or analyzing the economic outlook."[18] What is left is an awareness that economies grow and behave in hard-to-predict, uneven spurts that can not even in a figurative sense be considered as 'cyclical' and 'inevitable'.

5.5 A More Germane Approach to Socio-Economic Time-Series

The time-series model described in the previous section forces socio-economic history into a conceptual straight-jacket. The mathematical algorithms used to decide

which parts of a time-series are essential, and which are not, are insensitive to the facts in society, allowing interpretation only in the broadest terms.

Instead of such a mechanistic approach to the analysis of a time-series I propose a historically more sensitive approach, taking seriously every wiggle and fluctuation in a time-series. A simpler way of studying a time-series that does not exceed the analytical potential of the data should be based on comparisons of sub-series, formed along the geographic, the subject-matter and the time-dimension, to reveal aspects of the original time-series that were hidden within its aggregates. This can be achieved by establishing second-order ratios between the first-order ratio 'relative change in one of the sub-series,' $x_t/ x_{(t-1)}$, and the first-order ratio 'relative change in the more broadly defined original series,' $y_t/ y_{(t-1)}$ assuming the role of a 'trend-equivalent.' Then

$$k_{(t,t-1)} = [\{x_t/x_{(t-1)}\}/\{y_t/y_{(t-1)}\}]^*(100)$$

or with a fixed base period

$$k_{t,o} = [\{x_t/x_o\}/\{y_t/y_o\}]^*(100)$$

Such a comparison of the changes at different levels of aggregation pinpoints the impulses that are responsible for a given fluctuation, by time of occurrence, by geographic location, and by socio-economic activity, like a tourist on a cruise-ship watching the wavelets in its swimming pool, while oblivious of the waves in the ocean along which the cruise-ship is moving.

One also could construct time-series of aggregates that are larger than the one being studied, given that such data are available, forming what one might call time-series of super-aggregates in relation to which the given time-series becomes a sub-series. A variety of such super-series can be formed for a given time-series, as trend-equivalents, against which the fluctuations of the given series could be meaningfully assessed.

This relies on the use of series with different degrees of smoothness. Smoothness in a time-series is the result of the 'width' of the definitions of its aggregates: those more broadly defined tend to give a time-series of smoother appearance than those whose data are defined more narrowly with respect to any of its three dimensions. Smoothness is the result of the internal compensation of the countervailing effects of forces that are active in certain parts of the tri-dimensional conceptual space of the aggregates. Smoothness as such, achieved at the expense of detail, is neutral, neither desirable nor undesirable.

The ratios k reveal features of the development which otherwise may be overlooked. By doing this in three dimensions, by regions, by socio-economic activities, and by time-intervals, the origin and development of each influence can be traced, identified and its intensity measured. This kind of analysis differs from the usual time-series analysis in the attention given to the economic and social detail of the situation. Changes in direction and intensity of the coefficients $k_{(t,t-1)}$ or $k_{(t,0)}$ could be due to changes in the statistical procedures or a technical-statistical glitch.

Beyond that, however, and more importantly, they can be interpreted as indications of changes in the socio-economic phenomenon, due to impulses from specific, traceable events at a certain time, in a specific socio-economic activity, located in a specific geographic region. The effects of local and shorter impulses can be disentangled from the effects of more pervasive general forces, revealing fluctuations that could not be noticed because they were hidden inside the aggregates of the wider, more inclusive definitions of the larger, smoother time-series. The interpretation of the discrepancies between the two series, measured by the ratio k, will lead to a search for actual happenings that should be identifiable with each revealed up- or downswing. In this way, the share of each region and each economic activity in producing effects in the aggregate can be assessed.

5.5.1 Editing as a First Step of Preparation

When attempting to interpret a time-series, one should at the outset assume that every change and fluctuation in the economy or social situation that is meaningful and of interest can be accounted for. Not even minor wiggles in the series should be disregarded as too insignificant. Nonetheless, before interpreting, some editing must be done to adjust the data for some formal effects that are not part of the underlying socio-economic phenomenon.

Here belongs the formal difference in the length of the time-interval. Months of 30, 31 or 28 (29) days should be adjusted to time-periods of equal length. A more discriminating 'adjustment' would take into account the number of working days in each time-period. If possible, the number of days actually worked should be used or even better, the number of man-hours actually worked. This may be cumbersome or impossible, since the necessary information for such adjustments may vary for each region and socio-economic activity[19] and may not be available.

Let $x_{(t-1)}$ be the quantity produced (work done) in periods $t-1$, and $x_{(t)}$ the quantity produced in the following period t. The simple first-order ratio $\{x_t/x_{(t-1)}\}$ represents the relative change in that production series. If $M_{(t-1)}$ is the number of **M**an-hours worked in time-period $t-1$, and M_t the number of **M**an-hours worked in the following period, the ratio $M_t/M_{(t-1)}$ is then the relative change in Man-hours worked either as a link-relative or fixed-based ratio. The second order ratio $k_{(t,t-1)}$ indicates by how much (in percent) a change in the produced quantity x from period $(t-1)$ to (t) is due to a change in all causes except the formal number of hours worked.

$$k_{(t,t-1)} = [\{x_t/x_{(t-1)}\}/\{M_t/M_{(t-1)}\}]^*100 \text{ or } k_{(t,o)} = [\{x_t/x_o\}/\{M_t/M_o\}]^*100$$

Then the procedures of data collection and the definitions and measures used – e.g. a change to a different concept of production – to be discussed in Sect. 7.2 – the change from local currencies to the Euro, or a change in the design of the sample should be reviewed, inspected for changes during the time-period under consideration, and if necessary, the data adjusted. The analyst must then assess whether the information for such adjustments is available and worth the effort and cost.

5.5.2 *Relating Data of Regrouped Geographic Areas*

Simpler to interpret than the standard trend and seasonal analysis, especially the moving averages, are the geographic-regional aspects of time-series that highlight events in smaller geographic areas. It amounts to comparing time-series of data of differently-sized geographic areas Let the data of a time-series of the larger geographic area be 'y' and the data of the time-series of its smaller geographic sub-areas 'x'. Both series are for the same time-intervals and economic activities. For example, the statistical figures for 'private residential construction put in place in the northeast region' are quarterly, just as the figures for the private residential construction for the entire territory of the USA.

The latter, the larger conceptual frame, larger with respect to its three dimensions, is the more inclusive and smoother series that is used as a 'trend-equivalent' for the series of its geographic subtotals. Figures 5.9 and 5.10, show the individually divergent construction activities in the North East sub-region (for data see the Appendix B). Analogous computations can be made for the other sub-regions.

The time-series data of the larger region are the result of the net effect of the ups and downs in the data of the time-series of each sub- region by combining the data of their activities in its sub-regions. The fluctuations visible in the aggregates of that geographically larger series 'Private Residential Construction in the entire USA' show what is left after internally compensating the fluctuations caused by the forces that affect that construction activity in the sub-regions.

These local building and construction activities differ with regard to regional detail and can behave more erratically, reflecting the ups and downs of that indus-

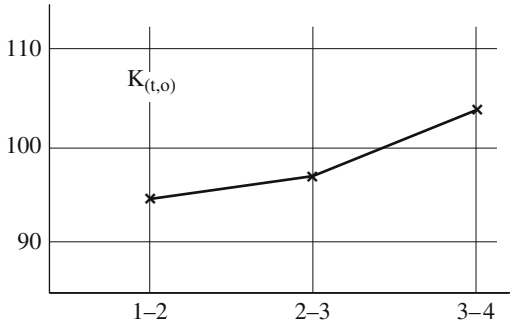

Fig. 5.9 Private residential construction in the North East region i = 1, as a % of changes in the private residential construction industry in the entire USA between each two consecutive quarters of that year

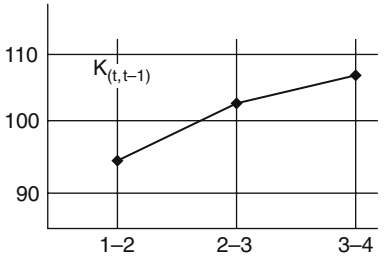

Fig. 5.10 Same data as in Fig. 5.9, as a link-relative or chain index for the same region

try in those narrower regions in relation to that activity in the entire country. The series of the large region, the entire USA, is smoother as the developments in its sub-regions offset each other. The data of that large region can be used as the reference horizon, or the 'trend-equivalent' that, unlike a customary trend-line, can be interpreted meaningfully in terms of the regional socio-economic situations of that industry.

The arithmetic of this statistical technique is simple. Take $x_{i=1,t}$, $x_{i=2,t}$, $x_{i=3,t}$ $x_{i=4,t}$ as the data of quantity or value of the construction activity in the smaller geographic areas, labeled 1, 2, 3, 4, ... in period (t), and y_t as the quantity or value of construction activities in the larger geographic area for the same economic activity and same time-period (t). Then the second-order ratio $k_{i=I,(t,t-1)}$ gives the percent change of the geographically smaller series for region i = 1 relative to the geographically larger, presumably smoother series for the entire country.

$$k_{i=1,(t,t-1)} = [\{x_{i=1,t}/x_{i=1,t-1}\}/\{y_t/y_{t-1}\}]^*100 \text{ or } k_{i=1,(t,0)} = [\{x_{i=1,t}/x_{i=1,0}\}/\{y_t/y_0\}]^*100$$
$$k_{i=2,(t,t-1)} = [\{x_{i=2,t}/x_{i=2,t-1}\}/\{y_t/y_{t-1}\}]^*100 \text{ or } k_{i=2,(t,0)} = [\{x_{i=2,t}/x_{i=2,0}\}/\{y_t/y_0\}]^*100$$
$$k_{i=3,(t,t-1)} = [\{x_{i=3,t}/x_{i=3,t-1}\}/\{y_t/y_{t-1}\}]^*100 \text{ or } k_{i=3,(t,0)} = [\{x_{i=3,t}/x_{i=3,0}\}/\{y_t/y_0\}]^*100, \text{ etc.}$$

The number k−100 indicates the percentage by which the change in the smaller geographic area exceeds or falls short of the change in the larger region. The graphs of Figs. 5.9 and 5.10 show the changes in 'Private Residential Construction in Region i = 1' as a percentage of the changes in the same economic activity and same time period for the entire USA. Figure 5.10 presents the link-based second-order ratios $k_{i,(t,t-1)}$ that connect the first-order ratios of each two consecutive t values of the geographic sub-region i = 1. A similar calculation accomplishes this for the other sub-regions (graphs not shown).

The interpretation is straight-forward: it is the amount of construction activity in that particular smaller geographic region, above or below the general development in the construction industry throughout the country.

That same technique further allows studying the inner structure of each of the smaller regions which were investigated. In this case, the figures of each one of the sub-series can be further broken down by their own sub-regions. What were originally sub-regions become the new y's and their component series are the corresponding ${}_{j=1}x_{i=1,t}$; ${}_{j=2}x_{i=1,t}$; ${}_{j=3}x_{i=1,t}$ etc. the left-subscripts 1, 2, 3, etc. indicating the new sub-(sub-regions) of what were originally the right-subscripted small areas i = 1, 2, 3 etc. The new ${}_j k_{(i,t,t-1)}$ and ${}_j k_{(i,t,0)}$ have an analogous interpretation.

There is no need to fit a least-squares trend line. It should be recalled, that the least-squares criterion was derived from and applies to situations in which 'normally' distributed errors of measurement have affected the point-like, individualized measurement-observations in the natural sciences. The situation in social-science data, however, is different, as already discussed. The ratio k neither reveals 'residuals' or 'random fluctuation,' nor does it attempt to de-trend the series, but focuses on the regional detail against the background of the broader picture of that economic activity in the larger geographic area.

When bringing together the large and the narrow time-series through the k-ratio, though dealing with the same economic forces, these series lie, as it were, on dif-

ferent focal planes and therefore cannot both be seen 'in focus.' When focusing on the data of the small region, the larger region is not in focus, like the distant object discussed in Sect. 5.4.2 The meaning of the ratio k, though obviously focusing on the narrower local series, has the larger series as the trend-equivalent horizon of reference. That ratio, therefore, is defined like the broader, less detailed of the two series of aggregates, viz. the series with the wider geographic region (see Chap. 4).

5.5.3 Relating Data of Re-grouped Socio-Economic Activities

Something analogous can be achieved by comparing a time-series of a subset of an economic activity with the time-series of the broader grouping of that industry or economic activity. It amounts to relating time-series which include different types of socio-economic activities, different products, or different types of crimes, etc. in the same geographic region In the larger, more broadly defined time-series, which can include either all economic activities in that geographic region or only some that are specifically related to construction activities (see Table B.2 of Appendix to chapter 5), the smoothing effect again is achieved through the partial compensation of the divergent movements of its sub-series. Those fluctuations that remain in the time-series of the broader grouping of economic activities in that sub-area, are shaped by economic forces that affect that broader grouping of economic activities in that region. The interpretation of these ratios 'k' is analogous to that of geographic sub-regions, described in the previous section.

Figure 5.12 presents the same situation as Fig. 5.11 but as a fixed-based ratio $k_{i,(t,o)}$ for the same sub-regions

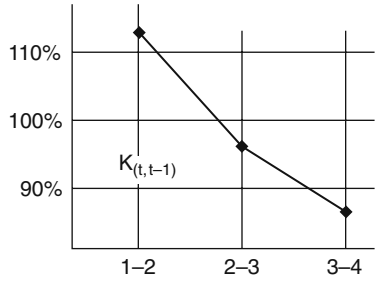

Fig. 5.11 Link based ratios $k_{i,(t,t-1)}$ of residential construction as a percent of all economic activity in region i = 1

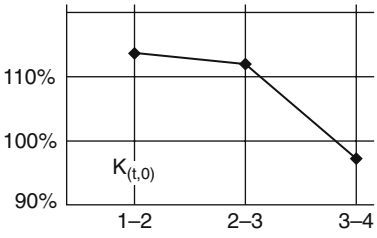

Fig. 5.12 Fixed base ratios $k_{i,(t,0)}$ of residential construction as a % of all economic activity in region i = 1

5.5.4 *Relating Data Grouped in Time-Intervals of Different Lengths*

The quarterly or monthly data of a time-series can also be related by the ratio k to each corresponding year's average quarterly values – or monthly data of a monthly time-series – used as the trend-equivalent for the same geographic region and economic activity. These ratios, however, are somewhat more difficult to interpret. Most of what is to be said here has already been stated. The partial second-order ratios e.g. for a quarterly series $k_{(i,t,t-1)}$ or $k_{(i,t,0)}$ for the i quarterly data, would be computed as follows. Instead of x for the smaller region/activity as in the formulas above, now the shorter time interval, formerly x e.g. of a **q**uarter, in the following presented by q, and the former, larger aggregate, y, can in the following be read as the '**y**early' aggregate.

$$k_{1,2004} = [q_{(1,2004)}/\{y_{2004}/4\}]^{*}100, \quad k_{2,2004} = [q_{(2,2004)}/\{y_{2004}/4\}]^{*}100,$$
$$k_{3,2004} = [q_{(3,2004)}/\{y_{2004}/4\}]^{*}100, \quad k_{4,2004} = [q_{(4,2004)}/(y_{2004}/4)]^{*}100$$
$$k_{1,2005} = [q_{(1,2005)}/\{y_{2005}/4\}]^{*}100, \quad k_{2,2005} = [q_{(2,2005)}/\{y_{2005}/4\}]^{*}100,$$
$$k_{3,2005} = [q_{(3,2005)}/\{y_{2005}/4\}]^{*}100, \quad k_{4,2005} = [q_{(4,2005)}/\{y_{2005}/4\}]^{*}100, \text{ etc}$$

Their values indicate the relative excess or shortfall of the smaller, more narrowly defined quarterly series above or below the series with the larger, yearly aggregates used as the trend-equivalent. The same, of course, holds with respect to monthly series.

A more sophisticated and complex calculation that would better reveal the inherent indeterminateness of the larger time aggregates, would use a time series of four quarter moving averages – instead of simply using the yearly total as in the example above – of the time-aggregates for the same geographic region and economic activity. It would better show the implied uncertainty in the contrast of the quarterly series relative to the development of the smoother yearly moving aggregates. The resulting k-ratios use a non-centered moving total of four quarters, placed at the end of each time period, as the horizon of reference, allowing to see the within-year developments. The exact location of each smaller quarterly (or monthly) time-period on the horizontal time-axis is undetermined with respect to the larger time-interval. For reasons discussed earlier, the figure of, e.g. the fourth quarter $q_{(4,2004)}$ of 2004 should be linked to the four-quarter total $Y_4 = q_{(1,2004)} + q_{(2,2004)} + q_{(3,2004)} + q_{(4,2004)}$ comprising the quarters 1 to 4 of 2004 so that both totals end in December 31.

Placing the aggregate value at the endpoint of each period has the practical advantages of not losing the last, most recent terms of the time-series, when using moving averages, and of ensuring that the values of the ratios k will not be affected by the type of developmental growth in the series. The argument against the use of arithmetic moving averages in situations of geometric growth loses its urgency *

* The subscripts 4, 3, 2004 in the first k-ratio indicate that quarter 4 of 2004 is divided by quarter 3, of 2004; the subscripts 4 and 3 refer to the moving or progressively advancing artificial totals of four quarters each.

when the Y_t-values are moving totals

$$k_{(4,3,2004;4,3)} = [\{q_{4,2004}/q_{3,2004}\}/\{Y_4/Y_3\}]^*100, \text{ or}$$
$$k_{(1,4,2005;5,4)} = [\{q_{1,2005}/q_{4,2004}\}/\{Y_5/Y_4\}]^*100, \text{ or}$$
$$k_{(2,1,2005;6,5)} = [\{q_{2,2005}/q_{1,2005}\}/\{Y_6/Y_5\}]^*100, \text{ or}$$
$$k_{(3,2,2005;7,6)} = [\{q_{3,2005}/q_{2,2005}\}/\{Y_7/Y_6\}]^*100$$

One must keep in mind that the fourth quarter of 2004, q_4, is also part of the moving totals Y_5,Y_6 and Y_7

$$Y_3 = q_{(4,2003)} + q_{(1,2004)} + q_{(2,2004)} + q_{(3,2004)}$$
$$Y_4 = q_{(1,2004)} + q_{(2,2004)} + q_{(3,2004)} + \mathbf{q}_{(4,2004)}$$
$$Y_5 = q_{(2,2004)} + q_{(3,2004)} + \mathbf{q}_{(4,2004)} + q_{(1,2005)}$$
$$Y_6 = q_{(3,2004)} + \mathbf{q}_{(4,2004)} + q_{(1,2005)} + q_{(2,2005)}$$
$$Y_7 = q_{(4,2004)} + q_{(1,2005)} + q_{(2,2005)} + q_{(3,2005)}$$

One could relate $q_{(4,2004)}$ with nearly equal justification to any of the totals Y_t, for $t = 4, 5, 6$ or 7, in which $q_{(4,2004)}$ is contained, and compute the coefficients:

$$k_{(4,3,2004;4,3)} = [\{q_{4,2004}/q_{3,2004}\}/\{Y_4/Y_3\}]^*100, \text{ or}$$
$$k_{(4,3,2004;5,4)} = [\{q_{4,2004}/q_{3,2004}\}/\{Y_5/Y_4\}]^*100, \text{ or}$$
$$k_{(4,3,2004;6,5)} = [\{q_{4,2004}/q_{3,2004}\}/\{Y_6/Y_5\}]^*100, \text{ or}$$
$$k_{(4,3,2004;7,6)} = [\{q_{4,2004}/q_{3,2004}\}/\{Y_7/Y_6\}]^*100$$

The values of these k-ratios and the computations are shown in columns 9, 10, 11, 12 and 13 of Table B.3 of the Appendix to chapter 5. This should give an idea of the uncertainty that results from the four possible trend lines when the overlapping, larger totals of n periods are used. Here again one must face the fact that there is a **trend-zone,** not a trend-line when using moving totals or moving averages. Statisticians' attention to minor issues, such as whether to 'center' the moving averages has deflected attention form the bigger issue of the indeterminateness of the results.

Though seemingly absent from the discussion in the two preceding sections, this vagueness always exists when a narrowly defined aggregate is compared with one that is more broadly defined. In fact the percent changes in aggregates of different size comparing geographic regions and subject-matter groupings should be understood as a hazy zone with fuzzy borders, not a clearly drawn trend-line. The comparison between a given q value with each of the four Y-aggregates containing that q value, has a somewhat different meaning. The conditions or economic forces that have shaped both aggregates are only in part the same, and each of these possible comparisons will reveal a somewhat different aspect of the underlying economic and social forces.

Summing up, the interpretation of seasonal changes seen against the backdrop of changes in the same time-series of aggregates with wider time-intervals is less

revealing than the interpretation of regional differences, or the different development of a specific economic activity, observed against the backdrop of the time-series that represents the bigger geographical or industrial context.

5.5.5 Concluding Observations on Time-Series

The suggested fact- and location-oriented interpretation of time-series is based on a different approach to statistical materials. It is a change from the prevailing natural-science mathematical-mechanical approach that treats the aggregates of socio-economic data as if they were point-like measurements, aiming to discover laws in nature, with randomly distributed measurement-errors, to a socio-economic statistical approach that recognizes and respects the historic and geographic uniqueness of each event. The origin of each fluctuation cannot be pinpointed or explained by relying on statistical information alone. A meaningful interpretation of time-series must be supplemented and combined with an inquiry into the social, economic, regional and historic developments of the situation presented in the data. All knowable circumstances must be blended into a fact-oriented interpretation. The aim is not to discover general economic 'laws' that are removed from historic time and actual geographic region, but to provide decision makers with useful background-information about quantitative aspects of the broader economic and social panorama that are not otherwise obtainable.

The proposed procedures require access to a wealth of statistical detail. Results are achieved through comparisons among time-series of a different degree of detail regarding geographic areas, subject-matter groupings and time-intervals. It relies on the smoothing effects of larger totals that ignore detail of development in time, or of geographic region or of subject matter. The meaningful interpretation of time-series requires acquaintance with the nature of the data as never before, as well as with the general economic and social situation reported in the time-series. The proposed simple statistical procedures should reveal socio-economic phenomena through 'insight' and 'overview' – see the motto 'Divide et Impera' under the heading of Chap. 3 which is the essence of the suggested methods here – to be carried out as far as the available data permit. The results will always be relevant and simple to interpret.

Notes

1. John Wu, *Beyond East and West* Sheed & Ward, 1951. Author Wu, quoting Chin-Yuan Wei-Hsin elaborates further:
   "The second view of this saying is more interesting than the first, but the third is the most interesting of all. You see mountains as mountains and waters as waters. There is nothing extraordinary about it and that is why it is so extraordinary. . ."
2. Professor Parzen insists that time-series of engineering data are of the same kind as time-series of business-, economic or social data. He believes that ". . .it is of vital importance that the present tendency toward a schism between the statistical literature, the engineering literature

and the economic literature be arrested with the aim of developing a true inter- disciplinary field of time-series analysis. . .”
Parzen, Emmanuel, book review of Hannan, E.J., “Time-Series Analysis” in *Journal of the American Statistical Association JASA*, 57:505, June 1962, p. 505.

3. Frederick A. Ekeblad, op.cit. on p. 589 distinguishes between ‘cumulative data’ and non-cumulative data’. “data which appear on the income- or profit and loss statement of a company are generally cumulative data. Data which appear on a balance sheet are generally non-cumulative data. . .” Ekeblad considers this distinction important. Richmond distinguishes between” ‘aggregative type series’, such as production, sales, income and expenditures; while the ‘non- aggregative series’ include population, prices, number of employees and interest rates”
Samuel B. Richmond, *Statistical Analysis,* 2nd ed., The Ronald Press Co. N.Y., 1964, p. 375. This distinction is made as an afterthought to the treatment of time-series, a marginal remark on the occasion of discussing how to change the scale of a regression equation. Most of the data declared to be non-cumulative are so only with regard to time-dimension, not, however, with respect to subject- matter categories and geographic regions.
4. “It must not be overlooked that our method of determining the values of the series at fixed equal intervals of time may suppress evidence of oscillatory movements which have a period equal to those intervals or to some sub-multiple of them. Suppose for instance, that there was a systematic oscillation in the English population expressible by a harmonic component with period of exactly 10 years, by observing the series at 10-yearly intervals we should never find any evidence of this effect, . . . In the population case, of course we have collateral evidence to indicate that no such oscillation exists, but where nothing is known of the series otherwise we can never exclude the possibility of a period exactly equivalent to our time interval. Sometimes in fact, we know that it is there, and choose our interval so as to exclude the oscillation from consideration. For instance, in our sheep population we know that there is a seasonal effect within the year, which is not brought out in Table 29.2 because the sheep census is taken on June 4th each year.”
Maurice G. Kendall, The Advanced Theory of Statistics Vol. II, 3rd edition, Hafner Publishing Co. N.Y. 1951 p. 368, 369
5. “The logician or economist who wants to be difficult can always maintain that, although any series can be separated out into three specified components as a matter of mathematical or statistical analysis, **the results throw little or no light on the causal influences at work that produced the series**. To such a critic we have to concede, I think, that in carrying out the analysis we have at the back of our minds the strong possibility that the three elements are due to independent causal systems. If he refuses to accept this view – and some economists do – we can only invite him to produce a better statistical method. . .” Maurice G. Kendall, op. cit., p. 371.
6. A good discussion of this phenomenon can be found in William Neiswanger, *Elementary Statistical Methods,* The McMillan Co., New York, 1956, p. 515, 516. The trend value for five least squares trend lines, fitted to the production of steel ingots and steel for castings (in millions of tons):

Period	Equation	x-units	Origin
1930–1953	Y=66.583+1.79x	6 mo.	Jan. 1,1942
1917–1953	Y=59.514+1.77x	1 yr.	Jul. 1,1935
1900–1929	Y=34.433+0.73x	6 mo.	Jan. 1,1915
1900–1953	Y=48.722+0.72x	6 mo.	Jan. 1,1927
1917–1938	Y=42.140-0.15x	6 mo.	Jan. 1,1928

If these equations were unified as to their origin, the y-intercept values would differ even more. Note the negative slope in the last equation.

7. Arthur F. Burns and Wesley C. Mitchell, *Measuring Business Cycles*, NBER, New York 1947, Note 14 on p. 46 "in practice we use ratio-to-twelve-month moving totals instead of to moving averages. The two yield the same final results, but the former is a more economical method of calculation."
8. Note: The averaging effect refers to its position among the n numeric values of the moving total on the vertical axis of a typical time-series graph with time marked on the horizontal axis, the corresponding numeric values on the vertical y-axis. Although this average value can be clearly placed on the vertical axis, its placement on the horizontal time-axis is undefined within the limits of the time-interval of the corresponding moving total. The custom of placing that average value on the midpoint of the time-interval on the horizontal axis of the corresponding moving total is an unwarranted extension of the averaging principle. Notice that the calendar values of 'time,' marked on the x-axis, have not been averaged!
9. "The concept of trend is more difficult to define. Generally one thinks of it as a smooth broad motion of the system over a long term of years. But 'long' in this connection is a relative term and what is long for one purpose may be short for another . . . the rise over a particular century might appear as part of a slow oscillatory movement, so that any inference from the 'trend' in a particular century to the effect . . . would be quite false. . ." Maurice G. Kendall, op. cit. p. 370.
10. Winkler, Othmar, op.cit. 1966, Figure 5, p. 357 cit., figures 7 and 10, and Tables 1 and 2 on p. 368.
11. W. Allen Wallis, Harry V. Roberts, *Statistics – A New Approach,* The Free Press, Glencoe, Ill. 8th Print. 1960, pp. 586–587
12. Compare this with the erratic development of seasonal patterns of egg prices. Waugh, Frederick V., *Graphic Analysis in Economic Research*, Agricultural Handbook No. 84, US. Dept. of Agriculture, Agricultural Marketing Service, Washington 1955, p. 24, 25 (The development of egg prices in December between 1924 and 1952, as ratio-to-moving averages).
13. See Winkler, Othmar W., "Unrecognized Possibilities for Simplifying Production Index Numbers," *Proceedings of the Business and Economic Statistics Section ASA*, 1963, p. 5, Footnote #5.
14. What happens when a seasonal pattern is "isolated" from a time-series? First, the time-series is de-seasonalized as described above. The ratio for every observed month is divided by the smoothed value for each month, then averaging all January ratios to get the seasonal pattern for January, and so on, proceeding thus with each one of the twelve months. This seasonal pattern is then used to de-seasonalize all values of the time-series. Every figure in the original series is then divided by the values of the seasonal pattern. The twelve- month moving average, so to speak, is used as a reference horizon. But the "seasonal pattern" will be different according to where the moving average is placed: at the customary midpoint or elsewhere, using all the possible off-center moving averages. This brings the uncertainty to light that is inherent in the comparison between aggregates of different width. The meaning of the ratio of two statistical materials is dominated by the conceptually wider of the two, and is as vague as the wider of these aggregates. The de-seasonalized series consists of ratio to-moving average figures and has become a sequence of 'partial third order ratios.' Its values are basically as undetermined as the monthly detail is within a yearly total, as was explained in Sect. 4.3. To each one of the twelve, or in the case of quarterly series, of four different trend values for a given month, correspond twelve, or four different ratio-to-moving average values, for a quarterly time-series, forming a zone. The seasonal pattern then becomes a 'gray' not clearly determined seasonal zone of values, not a clear, single line.
15. Anderson, Oscar, Jr. "Eine Neue Variante der Saisonbereinigung von Statistischen Zeitreichen" *Mitteilungsblatt für Mathematische Statistik,* Heft 1" 1950, p. 50. Anderson determines a "Saison-corridor" $k \pm 2\sigma$ which also conveys this idea of uncertainty, for quite a different reason, though, as he assumes a normal distribution of all ratio-to-moving average figures.
16. An excellent discussion of errors, though not in the context of random fluctuations in time-series can be found in Morgenstern, Oskar, *On the Accuracy of Economic Observations,* Princeton University Press, Princeton, N.J. 2nd rev.ed. 1963.
17. See e.g. Samuel Richmond, op. cit. Fig. 17-5 on p. 412 and Fig. 18-2 on p. 418.

18. Colm, Gerhard, “Economic Barometers and Economic Models” in, The Review of Economics and Statistics Vol. XXXVII, No. 1, Feb. 1955, Harvard Univ. Press, p. 56.
19. Números Indices de la Producción Industrial, United Nations, Estudios de Métodos No. 1, ST/STAT/SER.F/1, New York 1950.

Chapter 6
Longitudinal Analysis, Part 2 – Looking to the Future

Forecasting: The Predictive Value of Statistical Data

*We must start out with the premise that forecasting is not a respectable human activity and not worthwhile beyond the shortest periods. Strategic planning is necessary precisely because we cannot forecast.**

Bluntly stated, it does not seem possible for anybody to predict the future, be that of an individual person, of an industry or of an entire economy. We are occasionally reminded of this, as in the sudden emergence of the world oil crisis of 1973 which hit the western economies unprepared, an episode which apparently nobody had anticipated. Anyone trying to determine the future is faced with this conundrum: it really cannot be done, yet it is necessary to know and must be attempted nonetheless.

6.1 The Forecaster and the Past

In this book on interpreting socio-economic statistical data, the stress is more on data than on methods. No forecasting method is better than the data and how it uses them. Therefore attention will be centered on the statistical data-inputs on which such forecasts are based.[1] This chapter will be concerned with the potential for prediction, that is, the predictive value of the data in a socio-economic time-series. When forecasting models are evaluated, their lack of success is often apologetically referred to as the lack of precision of the data or as the influence of a human factor in forecasting.[2]

The term predictive value is sometimes understood to mean the ability of one time series to give advance notice of changes in another series which lags behind[3] In the following, predictive value will be used in the broader sense, as allowing for the anticipation of future socio-economic developments in the data of every time-series.

Every statistical figure has a fixed relationship to a certain point in the development of a socio-economic situation. As that situation develops and unfolds, the analyst must keep abreast of the changes in the social or economic situation with

*Drucker, Peter "*Management, Tasks, Responsibilities*, Practices" Harper & Row, New York, 1974, p. 124.

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_6, © Springer-Verlag Berlin Heidelberg 2009

new data. When the availability of a statistical figure is delayed, the situation it describes corresponds to the state of affairs at the moment of analysis only to the extent to which that socio-economic situation has not changed. Unfortunately, in this rapidly-developing society, something of the relevance for describing the current situation, even of the most recent data, has already been lost by the time those data become available. The assumption of continuity in patterns and relationships, that underlies every attempt at forecasting, must be understood in light of this basic fact.[4]

Imagine, for example, how useful the time series of the index of industrial production would be to a forecaster who, at the end of the year receives the August figure as the latest available datum. How well would he or she be informed about the production situation of December? How useful would that time series be in a forecast for the first, or even later semesters of the following year? Or how useful can the 2000 census of retail establishments be to a forecaster who must rely on it as the only available information in 2006, years after the census has been taken? There is even a point in time beyond which a given statistical figure ceases to be of value in assessing the current, let alone the future state of a socio-economic situation. This important fact of 'data obsolescence' is of much less importance in the data of the physical/natural sciences, which are, for the most part, accurate measurements of facts that hardly change over time. It is because of statistical theory's reliance on methods developed in the natural sciences that economists and social scientists have not paid attention to the fact of data obsolescence and its consequences.

A good part of forecasting consists of understanding the past and tracing the historic roots of the economic and social forces that are responsible for the present state of the situation in order to estimate how these forces may affect the future. Such an understanding of the history of a situation lies at the heart of the matter. In the best of cases, one should not expect to be able to predict the future further than one can trace the development of the present situation back to its past.[5] The forecaster should treat the different parts of a time series as being of different value, paying more attention to the newer figures than to the older ones. In fact, attention can be confined to a certain, limited time-span that keeps advancing as time moves on. One must not conceive of a socio-economic time series as an ordinary climbing vine that continues to grow at the tip of its runners while remaining fully alive in all its parts. One must instead conceive of it as one of those rare creepers, the older parts of which die off gradually while continuing to sprout new leaves and roots at the tip of its runners. With these, it clings to the new ground and feeds on it. In short, one must not burden a forecast with data that have become obsolete, and therefore irrelevant for anticipating the future developments of a socio-economic situation to be forecast.

6.2 Statistical Obsolescence, Its Causes and Measurement

It appears that it has not yet been fully understood that socio-economic data expire like perishable food products or medicines, and eventually become irrelevant for extending the time series into the future, regardless of the original cost and value of

the data. The process of obsolescence in the data, the fading-out of their descriptive value through the loss of timeliness continues unrelentingly. After some time every statistical figure will eventually lose its value in understanding the present situation, let alone its future. Expired socio-economic data have become irrelevant for anticipating the future course of the time series and ought to be kept out of the forecasting process. Forecasters, without being fully conscious of this fact, have recognized the need to obtain the latest data as quickly as possible and are willing to trade off their timeliness for completeness of coverage and accuracy of the most recently published monthly data.[6]

Statistical obsolescence stems from changes in the underlying causal system, and is due to factors that are internal and external to the socio-economic situation to be forecast.

1. Internal factors are at work whenever older workers, older equipment, buildings and long-serving fixed capital assets are replaced by newer ones that perform at higher levels of speed and quality. Changes are caused through the superior training of young entrants into the labor force, new installation of robotic systems, continuous upgrading of existing electronic equipment and computers. All this is getting managed with the increasing knowledge and efficiency of organizations under the influence of young graduates of business schools and economic departments. The combined effects are usually not directly reflected and recognized in the performance of the economy, such as in the data on production, exports, and the quality of life in society, such as the data on health, crime, family life, etc.

Many small changes happen when workers are retrained, new accounting concepts e.g. of depreciation are more widely adopted, business transactions are made faster and cheaper, more business firms allow flextime for their employees, employees working from their home on a computer via the internet, to name just a few of these changes. There is also a growing awareness of the responsibilities of business corporations to the environment beyond their immediate stake-holders, the availability of new synthetic materials that replace, e.g. steel with plastic, improvements in the design of established products, the growing acceptance of the use of addictive drugs and stimulants in society, to mention only some of the many facts that alter the responses of the socio-economic actors. are at work. These ubiquitous internal changes, in all segments of life, lead to corresponding losses of continuity with their corroding effects on all data that deal with society.

2. External factors are those that refer to changes in the general socio-economic setting, such as the switch from war to peace production, racial integration, and measures dealing with sex and race discrimination in employment (e.g. Title IV in US labor legislation) ordered by law, changes in the interest rate determined by the Central Bank of the country – in the USA the Federal Reserve Board – in short, every change in existing laws and government regulation that affects the industry or society in the region for which the forecast is to be made.

Every indication of an internal or external change is also an indication of additional obsolescence in the data before that change. Shifts in the combination of socio-economic forces are gradual and usually remain unnoticed in the data of a time series. Obsolescence operates as an unspectacular erosion that usually does

not leave traces that could be noticed in the data. Only occasionally changes in the environment also leave a perceptible mark in the data of a time series. This process of becoming irrelevant is unpredictable and takes place with uneven, changing speed within the same series. The loss of continuity and the corresponding increase in obsolescence are more pronounced in the longer time intervals between surveys, such as the quinquennial economic censuses or the decennial population census in the US, and the data farther back in time of a continuous series. It is here that the forecaster must proceed with well-informed, perceptive judgment. That obsolescence, eroding all statistical data, also affects the relationship between different time series, especially when a time series has been used as a proxy series for other series. A different source of data obsolescence, but of similar importance, are the occasional changes in the definition and in the methodology of gathering the data.[7]

Despite numerous hints at the need for staff and upkeep of the forecasting models in the description of actual statistical forecasts, statistical obsolescence has not been explicitly considered. Although forecasters have been intuitively aware of data obsolescence, nothing has been written or done about it.

How far back should the data of a time series be used as inputs into a forecasting model? Obviously, there is no simple unique answer that would apply to all situations. The forecaster will have to study each situation for all internal and external facts that may have affected the continuity of the causal system which underlies the socio-economic situation to be forecast. But the task to investigate and judge the impact of these factors on the data at hand pertains to the manager, engineer, sociologist, the applied economist or demographer, in short, the expert in the subject matter, not the mathematical statistician! Those experts should appraise the importance of the loss in continuity, assessing that corrosion as objectively as possible and how much of the continuity of each figure may have been lost from one time period to the next. The investigation into the relevance of the data is to be extended to the entire series, also to determine the cutoff point, beyond which the data of that time-series are not to be used at all. In forecasts performed continuously, the relevance of the data of each period must be reassessed and the assigned weights be adjusted for each new forecast to be made. It is important that this be done by informed expert judgment, not mechanically by a mathematical formula, like in ‘exponential forecasting’.

Obsolescence of data is to be estimated as the amount of lack of continuity in the underlying technical, economic and social conditions that connect the present situation to the earlier periods. The result of such an assessment of the degree of continuity-obsolescence is to be indicated by weights. These weights can be expressed by decimal fractions between 1.0 and 0.0 and should be used as multipliers of the corresponding data of the series to be forecast. A continuity rating of, e.g. 1.0 would indicate that between the time period under investigation and the moment it is to be used in the forecast, no changes in the internal and external factors were found. The corresponding obsolescence would be 0.0.

If the underlying socio-economic situation should have changed completely, the weight expressing continuity would be zero, which would also be assigned to all

the data before the one with the rating of a 0.0 continuity. The corresponding multiplicative weight for obsolescence will be 1.0.

These individually assessed, continuity/obsolescence ratings between each two consecutive periods will have to be concatenated into joint ratings. A joint rating of say, 0.1 – a 10% continuity or 90% obsolescence – means that the information about the figure of that time period should be used in a forecast with only 1/10 of the importance given to the newest, completely up-to-date figure at the time when the forecast is to be made.

As a hypothetical example, consider a time series in which the pair-wise continuity ratings were established by subject-matter experts, back to 1990. The degree of continuity and obsolescence of each year's figure, expressed as decimal fractions between 0 and 1, are to be used as multiplicative weights – not to be mistaken for probabilities. A rating of 1 indicates that the conditions in that industry have not changed. A rating of 0 would indicate that the conditions have changed to such an extent that there was no point of using the information of that time period as well as of all data before it.

Suppose that the degree of continuity in the data, or the lack of it, as obsolescence – the data of that series themselves to which these weights apply are not shown – with regard to the previous year's figure, was determined to be as in Table 6.1.

The figure '.94' for the data of 2005 would indicate that between 2005 and 2006 the situation underwent only minor changes. The high stability in the situation was assessed as .94, with a corresponding loss of continuity of only about 6%. The 2005 figure of the time-series (not shown) is to be used in a forecast made in 2006 with 94% of its value. Between the years 1995/96, in contrast, major discontinuities in that industry were observed, leaving only 40% of the conditions to carry over into the following year. This low continuity assessment corresponds to a complementary loss through obsolescence of about 60%. On the other hand, there were no changes observed in the underlying socio-economic situation between 2003 and 2004.

Table 6.2 shows the joint continuity ratings of that series (but not the data of that time-series themselves) beginning with the most recent time period. The corresponding joint obsolescence is the complement to 1.0 but is not to be used in analogy to the joint continuity figures.

These ratings indicate the cumulative continuity and obsolescence of the earlier data of that time series (data not shown) to forecast the scenario of the socio-economic setting for this series beyond 2006. These computations indicate that e.g. the figures for the year 2000 are to be used only with a weight of .739 to indicate their reduced importance in comparison with the importance, appraised as the weight of 1.0 of the latest available figure – the one for 2006 – of that time-series. To be extrapolated into the future, beyond 2006. Although the weight of .739 is the result of expert assessment of the situation, really an educated guesswork, and the decimals in that rating need not be taken too literally, it should serve as a guide to the forecaster, regardless of the forecasting method used. Stated differently, the supposed changes in the underlying economic and social situation of that hypothetical time-series have caused a loss of actuality through obsolescence of about

Table 6.1 Year-to-year continuity/obsolescence ratings 1990–2006

Year	1990–91	1991–92	1992–93	1993–94	1994–95	1995–96	1996–97	1997–98	1998–99	1999–00	2000–01	2001–02	2002–03	2003–04	2004–05	2005–06
Continuity	.94	.98	.80	.75	.79	.40	.35	.60	.80	.86	.91	.97	.99	1.0	.90	.94
Obsolescnc	.06	.02	.20	.25	.21	.60	.65	.40	.20	.16	.09	.03	.01	0.0	.10	.06

Data of that time-series itself are not shown.

Table 6.2 Joint continuity/obsolescence ratings in 2006 of a hypothetical time-series (figures not shown) for the years 1990–2006

Year	Cumulative ratings	Continuity	Obsolesc.
2005–06	.94	=.940	.060
2004–05	.94∗.90	=.846	.154
2003–04	.94∗.90∗1	=.846	.154
2002–03	.94∗.90∗1∗.99	=.838	.162
2001–02	.94∗.90∗1∗99∗.97	=.812	.188
2000–01	.94∗.90∗1∗.99∗.97∗.91	=.739	.261
1999–00	.94∗.90∗1∗.99∗.97∗.91∗.86	=.636	.364
1998–99	94∗.90∗1∗.99∗.97∗.91∗.86∗.80	=.509	.491
1997–98	94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60	=.305	.695
1996–97	94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35	=.107	.893
1995–96	.94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40	=.043	.957
1994–95	.94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40 ∗.79	=.034	.966
1993–94	.94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40 ∗.79∗.75	=.025	.975
1992–93	94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40 ∗.79∗.75∗.80	=.020	.980
1991–92	.94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40 ∗.79∗.75∗.80∗.98	=.020	.980
1990–91	.94∗.90∗1∗.99∗.97∗.91∗.86∗.80∗.60∗.35∗.40 ∗.79∗.75∗.80∗.98∗.94	=.019	.981

26%. These older data are not to be used on par with the latest figures, but with the indicated 'discount due to obsolescence.' The 1997 figure of that time-series should be used in a forecast with only approximately 30.5% of its value compared with the full 100% value of the latest 2006 figure.

Inevitably, different experts will assign different weights to the data of a given time-series. Part of the gift a forecaster needs, is to seek and gain insight from expert advice. Forecasting is indeed an art regarding the use of data. The intuitive understanding of this fact accounts for the growing popularity of exponential forecasting and its variants, that discount older data, although at an arbitrarily chosen, fixed rate.

Changes in the underlying causal system have been measured before.[8] For purposes of forecasting, however, a more alert perception of historical changes is required. It is obvious that all data before 1996 of this hypothetical time-series (Table 6.2) should no longer be used in 2006 for forecasts beyond that year.

Figures 6.1 and 6.2 show the impact of Government legislation on Bovine Spongiform Encephalopathy in the UK.[9] Although not really a social phenomenon, it did have vast economic implications and provides a dramatic example of a complete break in a time-series that is clearly perceptible on all levels. The first, most drastic government intervention took effect on June 21, 1988, making BSE

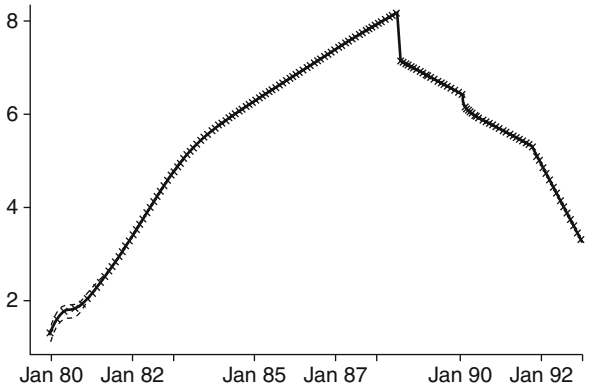

Fig. 6.1 Three government interventions in the UK 'level component' of the time-series of incidents of mad cow disease

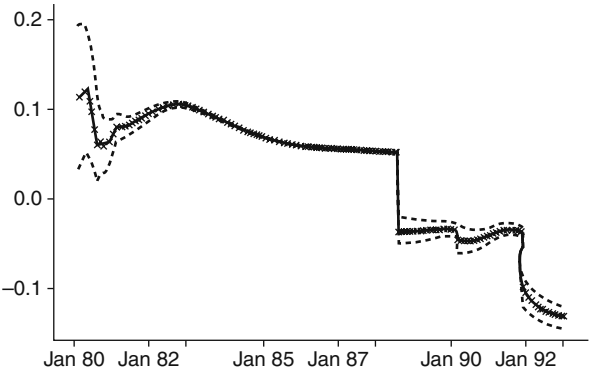

Fig. 6.2 Three government interventions in the mad-cow disease in the UK growth component of the time-series

notifiable and provided for isolation of the BSE suspected animals. Also banned was the use of ruminant-derived protein in ruminant feedstuffs. A second government intervention, effective as of February 14, 1990, introduced 100% compensation for all animals slaughtered under the compulsory slaughter scheme. A third government intervention took effect on November 6, 1991. The first of these three disruptions made the series before June 1988 useless for forecasting. All the data before that break would receive a continuity rating of 0. It is one of the rare examples of a complete break in the underlying conditions that also is visible in the data.

This proposed limitation of the data in a time series contrasts with the belief that 'there is strength in numbers,' implying that the longer a time series the better its forecast. This kind of reasoning is not a legitimate application of sampling theory, according to which larger samples allow more reliable conclusions about a population. The data of economic time series, however, are not sample observations drawn at random from an immutable, timeless population. They are instead a sequence of statistical still-pictures, describing numerically successive stages in the historic development of a socio-economic situation. In many time series, each figure, that is, all the data are statistically speaking, populations, and the regression coefficients computed between such time series are those between populations. If the data of a time series are samples, then interval estimates for each time period could be determined and the upper and lower confidence limits for the population parameter be forecast separately.

Our thinking is dominated by statistical inference to such an extent that we seem unable to conceive of statistical data, including the data of most economic and social time series, as anything but random samples from some hypothetical, timeless population. It should have become clear by now that the data of a socio-economic time-series cannot be treated like a set of simultaneously existing sample units. The 'strength-in-numbers' argument fortime-series, therefore, has no strength at all and is out of place. The failure to understand the true, descriptive nature of socio-economic time-series data has kept forecasting methods and forecasters from limiting their time-series to only the relevant, more recent and often short parts of a time-series.

6.3 Obsolescence and Size of the Aggregate

The question has been raised by practitioners whether the forecast of a time series can be improved by forecasting its sub-series and then combining these partial forecasts of the sub-time-series. Is such a combined forecast really superior to a direct forecast of the larger aggregates in a time-series?

Remember that a time-series that consists of large aggregates describes a broader, less detailed picture. Such a time-series shows only those major net-changes in the socio-economic situation that reach beyond the aggregation limits with regard to time-interval, subject-matter and geographic territory. Everything else has disappeared for practical purposes and is hidden from view. The result is a time-series

that not only fluctuates less but is also more resistant to obsolescence. Its data do not become obsolete as quickly, because the broad picture it describes is not as strongly affected by the innumerable day-to-day changes that occur in smaller regions and more narrowly defined subject-matter categories.

Time-series of more narrowly defined, de-aggregated data fluctuate more frequently and more strongly and are sensitive to minor changes in the business scene. Consequently these more narrowly defined sub-time-series are more strongly affected by obsolescence that increases more rapidly. The forecasting span can reach at most, and not further, into the future, gradually fading out, than the rate of a similarly fading effect of obsolescence allows to trace the present situation back into the past. Time series based on wider aggregates – wide with regard to any one, two, or three of its dimensions as discussed before† – have a lower rate of obsolescence and permit longer-range forecasts than those based on more narrow aggregates. Time series of smaller aggregates have a more rapid rate of obsolescence, allowing only for shorter-ranging forecasts. If various short-ranging forecasts are combined into the semblance of a forecast of the total series, that forecasting range will extend no further into the future, and will be as limited as that of its component series. It will not reach as far as the forecasting span of the not de-aggregated time series. In light of the foregoing, the belief that combining the forecasts of the sub-time-series to extend the reach and quality of a long-range forecasts, ignores the characteristics of aggregates and cannot be upheld.[10]

The obsolescence ratings should also be considered when exploring the relationship between various time series using *n-dimensiona*l multi-variate analysis. The rates of obsolescence of these *n* time series are likely to differ. In that case, the joint obsolescence for each data point of a given time period would be the geometric mean of the obsolescence ratings of each individual time-series, for each time period. When calculating regression parameters, these weights should be used like frequencies with which each of the multidimensional points on the regression surface is to be weighted. In all this it must be kept in mind that the proposed measure of continuity/obsolescence is an attempt at quantifying the estimates of changes in the historic developments. Algebraic refinements will hardly improve that situation.

6.4 Some Conclusions

Few forecasting models have performed well consistently. The reason for this, one may suspect, has not been so much due to faults in the economic logic on which they rest, but to the indiscriminate input of data, disregarding their obsolescence. These forecasting models will improve their performance if their parameters are computed with proper regard for statistical obsolescence, the backward-oriented counterpart of the forward-oriented predictive value. But no matter how clever a forecasting model is conceived, a forecast that relies only on the data of the past, proceeds like a person

† For this discussion refer to Sects. 3.3 and 5.5.

who walking backwards, advances with his back toward the direction in which he intends to move, while carefully scrutinizing the surrounding where he came from.

When adjusting seasonal fluctuations by electronic computers, the earlier models had a limited data storage capacity, accommodating data of a time-series for at most 15 years, which was deplored as a drawback.[11] In reality, such a limitation was a blessing because a span of the most recent fifteen years in a time-series is probably more than should be used anyway for most forecasting purposes in these rapidly changing times.

One could also expect some benefit from this discussion for the practical aspect of obsolescence for the storage capacity of electronic data banks. Obsolescence, rightly understood, should lead to a continuous turnover within the storage area of data banks. As soon as data begin to fade beyond the point of high usefulness, they ought to be transferred from the more costly, interactive storage area into a cheaper, less readily accessible archive, and eventually into inactive storage areas for their occasional use for some historic research. Such constant rotation would alleviate storage situations and lead to a more economical use of electronic memories. Compromises, though, will have to be made between different forecasting uses of the same series, and of data, whose component series have different expiration ranges.[12]

There is also another practical side to this: Insurance companies determine the ratios of insurable events from time series data, ratios that occasionally are incorrectly referred to as probabilities. These ratios, used for insurance purposes, are not changed often enough. But, insurance rates should not only be re-calculated to adjust for major changes in society, but more regularly from revolving sets of data, that include the newest data, while eliminating older data that no longer represent the social, demographic and economic reality.[13] That holds for all kinds of insurance policies, life, car, accident, home-owner insurance and many more. This is another area where the concept of data-obsolescence has potentially serious economic consequences.

Another conclusion is of a more academic nature. When relative frequency distributions are computed from time series, e.g. to determine insurance rates, to be used as approximations to the 'true' probability distributions, statisticians who lean toward the 'objective' interpretation of probabilities will include as many data of a time series as possible. Yet, to make these ratios more useful, statistical obsolescence will have to be taken into consideration. Such pseudo-probability distributions should be based on a revolving set of data in which the newest data are continuously added while the expired data are discarded. Their obsolescence ratings must also be continuously updated. In a situation of rapid obsolescence, these probabilities may be based on a fairly short part of the time series, approaching the situation of 'subjective' probabilities. In fact, the latter can be understood as the limiting case of a probability distribution derived from time-series data with extreme obsolescence. Thus the notion of obsolescence can open up an unexpected perspective on the possibility of bridging the opposing views held by proponents of subjective and objective probability.

Finally let me propose a plausible defense, for the frequent case of forecasts that missed the mark: assume you developed the perfect forecasting model that is expected to give results that are excellent and very close. But two things are bound to happen. (1) As forecasting is not a spectator sport, but made to guide action, those who ordered the forecast will take advantage of that predicted boon or act to ward off the predicted threat. And (2) Other forecasters will also have made forecasts. Even if their forecasts are not as good as yours, pro-active action will probably also be taken based on their forecasts. By the time the future you predicted so well arrives, it has been tampered with to such an extent, thanks to the effects of all these other forecasts, that it becomes something quite different from the future that had existed at the time when you made your forecast. So you can defend your own forecast by explaining that it would have been perfect if everybody else just had left that future alone!

Notes

1. This chapter follows closely Winkler, Othmar W., "Data Obsolescence and Forecasting" *The 11th Federal Forecasters Conference, Papers and Proceedings,* pp. 119–126, U.S. Department of Education, Washington, D.C. 2000 and, "Business Forecasting: The Predictive Value of Statistical Data," *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1967, pp. 381–385.
2. James B. Wong, *Business Trends and Forecasting, an annotated guide to theoretical and technical publications and to sources of data,* Gale Research Co. Detroit, Mich, 1966, p. 31 and Walter E. Hoadley Jr. "The Importance and problems of Business Forecasting" in: Herbert Prochnow, ed. *Determining the Business Outlook*, Harper Brothers, New York, 1954, p. 23.
3. Milton H. Spencer, Colin G. Clark, Peter W. Hoguet, *Business and Economic Forecasting*, pp. 202, 203.
4. ". . .for this philosopher of process (Whitehead) the primary metaphysical question is that of how to hold together the sense of permanence with that of perishing (p. 110). . .why is it that "my having existed" is indestructible? . . . there is after all an aspect of permanence to the universe. . .If these experiences perished in an absolute sense we would not even be able to talk about them. They would be nothing and therefore could not be referred to (p. 111) . . . our duration is not merely one instant replacing another; if it were, there would never be anything but the present – no prolonging of the past into the actual, no evolution, and no concrete duration. Duration is the continuous progress of the past which gnaws into the future and which swells as it advances. . .the past is preserved by itself, automatically. In its entirety, probably it follows us at every instant; all that we have felt, thought and willed from our earliest infancy is there, leaning over the present which is about to join it, pressing against the portals of consciousness that would fain leave it outside. . .Our past, then, as a whole, is made manifest to us in its impulse; it is felt in the form of tendency, although a small part of it only is known in the form of idea" (quoted from Bergson, Creative Evolution, p. 5).

   Whitehead's philosophy expands upon Bergson's insight that all of our past experiences remain present to us. . .It shows even more explicitly than Bergson's that the preservation of the past in the present applies to the whole of cosmic reality and not only to our own human memory. It maintains that it is of the very essence of physical reality. . .that whatever has happened in the past still abides in the present. . .We cannot overemphasize that every occasion is at root experiential. . . . the data that it experiences is the past. This past includes not only the immediate one, but also, in a vague sense . . . that of the entire universe. . .nothing is ever totally lost. In perishing, the occasions of experience are not relegated to absolute nothingness but instead are assigned an "objective immortality" in the experience of subsequent occasions. The

transition of things does not entail loss but preservation. In their perishing, cosmic events hand themselves over to assimilation by the present and thus are allowed to persist "immortally" as causally influential moment after moment. (p. 113) The incursion of novelty into the world means that the present has to give way, has to perish. . .. this fading of the present into the past, and then the fading of the past itself is the "ultimate evil in the temporal world."

Paul Tillich interprets this anxiety about the loss of the present into an irretrievable past in terms of our own mortality. What makes us anxious about our having to die. . . is not simply the possibility of our ceasing to be. . . it is the "anxiety of being eternally forgotten." The possible fading of our memory into complete oblivion where no traces of our having existed remain is what terrorizes us. Tillich maintains that humans were never able to bear the thought of having their experience thrust into a past where it would be totally lost. And this is the reason why they have always sought . . .to erect obstacles to the diminishment of their memory. (p. 114) Unless perishing is not absolute. . .it is extremely difficult to imagine how anything could be imbued with lasting significance. . . (p. 115)

John F. Haught, *The Cosmic Adventure – Science, Religion and the Quest for Purpose*, Paulist Press, New York, 1984."

5. The forecaster is like an observer standing on a bridge over a big, flooded river, looking downstream. From the configuration of objects floating on that river – uprooted trees, housetops, etc. – he may try to anticipate what the next scene on the river underneath, floating down from behind him, will look like. He will rely more heavily on the recent changes in the picture than on the more distant scene further downstream in anticipating what the river will bring next.
6. The US Bureau of Labor Statistics (BLS) publishes its price and productivity data in reverse time sequence, and limited to a short time span of older data, which makes sense in view of the foregoing discussion.
7. Spencer, Milton et al. op. cit. p. 91.
8. e.g. Gregory C. Chow "Tests of Equality between Sets of Coefficients in two Linear Regressions" Econometrica, Vol. 28, 3, July 1960, Also: "Das Lexis'sche Dispersionsverfahren," in: Wilhelm Winkler *Grundriss der Statistik I*, Wien 1947, Manzsche Verlagsbuchhandlung, pp. 73–79.
9. Sandy D. Balkin, Assessing the Impact of Government Legislation on BSE in the U.K., p. 104.

   "Bayesian dynamic linear models (DLM). . .**operate according to the principle of Management by Exception where an exception is relevant expert information from an external source** or a monitoring signal indicating that the performance of the current model is inadequate. . .**This** sequential **mode**l development **allows the incorporation of external subjective information** concerning future beliefs. . ..the DLM method includes a tool called **Retrospective Assessment.** . . .**it is useful in determining 'what happened' given all current information**. . . .will use this type of analysis to assess the impact of government legislation . . ." in

   Papers and Proceedings of the 11th Federal Forecasters Conference 2000, Washington D.C. pp. 99–107.
10. e.g. David C. Melnikoff, "Long Term Projections and Business Decisions", *Proceedings of the American Statistical Association, Business and Economic Statistics Section,* 1957, p. 337 upper right.
11. Julius Shiskin, Harry Eisenpress "Seasonal Adjustments by Electronic Computer Methods" *NBER, Technical paper No. 12,* New York, National Bureau of Economic Research, 1958, p. 427, especially his reference to Method I. The fact that the capacity of computers since then has been extended to 50 and more years does not change my point.
12. Georgetown University Library has begun in 1999 to remove books for which there was only minimal demand, from its 'active' shelves at the Main Campus and Medical libraries, transferring them to a geographically remote off-campus storage location that is cheaper but less rapidly accessible.

13. An example of this: “Motorcycle riders have long had to pay far more for insurance than automobile drivers. Geico. . ..now insures higher-risk divers. That includes older motorcycle riders, the fastest growing group of owners. In 1985, the median age of US motorcyclists was 27. By 2004, . . .the median age had risen to 41, as moneyed middle-agers entered the market. Motorcycle accidents have also risen with the age of riders. In 2005. . .reported that the average age of riders killed in accidents has risen in the past 10 years. The data show large increases in crash deaths of riders ages 40 and older. Riders 50 and older accounted for the steepest climb in motorcycle fatalities in 2005 compared with the previous year, especially those riding large bikes. . .Motorcycle deaths in 2006 were predicted to increase for the ninth straight year, . . . according to preliminary data from NHTSA. The agency projected nearly 5,000 such death for the year. Geico goes cruising for motorcyclists. . .”, *The Washington Post*, July 2, 2007, part D, pp. 1, 2.

Chapter 7
Longitudinal Analysis, Part 3 – Index Numbers

Grau, teurer Freund, ist alle Theorie, doch grün des Lebens goldner Baum *

7.1 The Measurement of Prices

7.1.1 Introductory Observations

Index numbers are unique to socio-economic statistics and are well-established. The need for index numbers has never been seriously questioned. They also pursue the longitudinal study of socio-economic phenomena, but are not recognized as belonging to the analysis of time series. Index numbers seem to be different, with theories and problems of their own. The discussion has not really gone much beyond the thinking of Laspeyres and Paasche some 125 years ago. It reflects the science-inspired ideal of that epoch concerning the measurement of 'prices' p, 'quantities' q, and the role of weights.[1] There was a time when the discussion of Price-index-numbers dominated the scene. By simply reversing prices p and quantities q, a price-index-number formula seemed readily convertible to a quantity-index-number with price weights. As will be discussed in the following, the classic 'index-number-problem' is really a pseudo problem. Price measurement can be greatly simplified by recognizing the misconceptions about the nature of 'price', about the complex nature of 'quantities,' and about the unrecognized features of socio-economic data. All this has been 'off the radar screen' of statistical theory.

As mentioned in Chap. 1, statistics in the social sciences, to which the measurement of price-levels and their changes belongs, has a dual role to play: Firstly, it must define the price phenomenon and report correctly on it.[2] In this descriptive role, price statistics functions as a "macroscope", (see note 30 in Chap. 1) which makes visible phenomena that are too large and too widely dispersed geographically and conceptually. Secondly, statistics then must interpret these phenomena, that are then reported in statistical tables, to guide decisions that impact the economy and the society.

---

* Johann Wolfgang von Göthe, Faust I Studierzimmer, Mephisto 2038; translated: "Gray, my dear friend, is all theory, but green is life's golden tree".

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_7,

These two functions of statistics, the capture and reporting of a phenomenon, and its subsequent interpretation, are to be carried out separately. The official measurement of prices, however, has been under pressure not only to report prices with price-index-numbers, but also to 'adjust' the recorded prices for changes in the priced items to obtain the theoretical ideal of the 'pure price change'. To achieve this, it has been taken as an undisputed necessity that price observation ought to be made from a set of carefully defined products and services that does not change over time. It has also been accepted, without questioning, to use the seller's price expectation as the "price-per-unit" (or "price-per pound," "price per gallon," etc.) price-tags on the merchandise instead of the a ctually paid prices. These per-unit price-quotations further required appropriate weights that had to be determined from an occasional household consumer survey.

In addition to intending to measure the abstract concept of changes in the 'pure price,' price-index-numbers are also expected to inform about changes in the 'cost of living,' a related but different matter. To achieve these separate goals with a single price-measure, statistical agencies make efforts to manipulate the 'raw' data, e.g. through "hedonic adjustments," thereby distorting the reported price reality. No need was felt to clarify the concept "price", since calling it 'p' did not seem to require further explanations.

7.1.2 A (Very) Brief History of Price Statistics

In 1764, when G.B. Carli investigated the development of retail prices in a small city-state of today's Italy, he could restrict his observations to a few basic products that were the popular staples of that time, such as wheat, wine and olive oil. Markets and consumer habits remained unchanged for decades. A century later, the economists Laspeyres (1864)[3] and Paasche (1866)[4] published their proposals for the measurement of prices in the economy of one of the 39 small kingdoms and principalities in what today is Germany, each with its own currency, postal stamps and customs duties. Though their economies were more developed than the city-state economies at the time of Carli,[5] they were hardly comparable with today's economies. At that time, Laspeyres' proposal, to follow the prices of a fixed list of typically-consumed products over the course of time, may then have reflected more closely the reality of price developments than today.

These simple historic conditions for the measurement of prices no longer exist. Compared to the limited number of products likely to have been traded at Laspeyre's time, the number of products and services has grown exponentially. The dynamism of today's markets must have been unimaginable at that time for those concerned with measuring price changes. Today, products that seem to have existed forever disappear from the markets, new products surge incessantly, and the quality and packaging of existing products change frequently.[6] Because of the overabundance of products in the modern economy, marketing, market research and advertising in newspapers, radio and television have become steady features of our lives.[7] The academic discussion of price statistics, however, does not seem to acknowledge these

historic changes. The theoretically demanded fixed 'shopping basket' of goods and services, on which price measurement is based, is becoming frequently revised[8] under pressure from the market dynamism. These pressures explain the growing popularity of the chain-price-indexes as a compromise between the strict comparability of the products to be priced, demanded by theory, and today's rapidly changing market reality.[9]

7.1.3 Statistical Price Collection and Market Reality

As mentioned earlier (Chap. 1) statistics has become dominated by the problems and methods developed in the natural sciences, despite the fact that statistics had its origin in the social sciences, as the description of the human population.[10] The new, differently-oriented statistical theory showed little interest in the problems of the social, economic and demographic sciences. Statisticians also accepted, somewhat uncritically, economists' abstract assumptions about prices.[11] that for every merchandise on the free market economy the price is formed at the point of intersection between the demand and supply curves as a unique monetary value $p_{i\ t}$, that price is the amount of currency per-unit, per kilo, per liter, etc., that price can be obtained from the price-tag posted by the seller on the merchandise, and that price indeed was a characteristic of the merchandise.

These ideas are at the root of the felt need for price-index-numbers. The attendant complexities and complications have created a comfortable academic cottage industry of index-number-research to solve the problems created by assumptions that had never been critically investigated or challenged. The problems of how to aggregate these unit-price quotations lead to the creation of a variety of different index-numbers. The concern about these formulas, however, distracted attention from the basic problems of how best to measure price-levels and their changes. The search for the "pure price" of an individual merchandise and its changes, the holy grail of economists, is only of limited practical interest to guide business decisions and economic policy.

A personal shopping experience for computer appliances some 30 years ago was 'the last straw' needed to change my view about price. I purchased a package of ten 5¼ inch Kodak diskettes in the discount house W. Bell for $12.95. Then I went to the bookstore Dalton, next door, to retrieve a book I had ordered. There I noticed, in a display of computer paraphernalia, the identical set of diskettes in the same packaging, for $6.50. I also bought that package, returned to the discount house and showed the manager that identical package with its much lower price tag. The manager, visibly embarrassed, paid me the price difference.

I was puzzled by the fact that vastly different prices existed side by side, in the same place, for the exact same product at the same time, and that 'price' could be changed with ease. Obviously 'price' was not, as I had believed like everybody else, a quality of the product at par with its other characteristics, such as its weight, size, or packaging. That minor episode convinced me that "price" really belonged

to the transaction, the business deal between the seller and the buyer, not the thing or service that was sold or bought. This also proved to be the key to the answer of what the 'real-life-objects' in price statistics should be, whose characteristics are to be recorded as the 'statistical-counting-units', aggregated and published in tabulations.

In actuality, there are usually different "prices-per-unit" for the same consumer product[12] at the same point in time, in the same market region, through all kinds of discounts, e.g. for buying a larger quantity of the same product,[13] as well as for a variety of pretexts[14] and through different local sales taxes and fees. There are also differences in price for the same merchandise in sales outlets with different trade volumes.[15]

That personal shopping experience made it clear that in price statistics the 'real-life-object' ought to be the business transaction[16] not the merchandise. The salient characteristics of a transaction are the 'Price' as the total amount of money paid, as well as the type of product involved in the transaction. When this is recognized, the merchandise, its physical characteristics and other detail, become of secondary importance for price statistics. It is the commercial transaction and its characteristics that ought to be recorded and reported, whereby the type of merchandise or service and their physical, technical and legal features are also valuable pieces of information. Yet the salient feature of the transaction is the actual amount of money paid, including all kinds of discounts and surcharges. As a related matter, for reporting the price-reality on the markets it is not important to identify which portion of a price goes to the private sector as cost and profit, and which goes to the public sector as sales taxes. Statistics has to include all this in reporting the prices actually paid in business transactions.

Summing up, price therefore is not to be understood as a characteristic of the merchandise or service, but as the principal feature of the commercial transaction, in which seller and buyer are motivated by economic and psychological considerations.[17] The transaction-prices depend to a greater extent on the perception of the social and financial pressures and constraints under which both sides approach the transaction, the power of persuasive advertising, and the negotiating skills of the participants, than on the detailed specifications of the products.[18]

7.1.4 What Really Is 'Price'?

The per-unit price-tags affixed on the products, for the indicated reasons, are too often not the actual amounts of money paid, the actual prices. To appropriately describe the price reality, the statistic of prices has to be, in principle, a statistic of the universe of all transactions with their actually paid amounts of currency. It has to include and capture the prices of the fast growing number of new or changed products, and the multiplicity of simultaneously existing, different prices for the same products.[19]

As already mentioned, the actual transactions have to be the 'real-life-objects' of which the corresponding 'statistical-counting-units' are recorded with the

effectively-paid amounts of money and the type of product as their important characteristics. In price statistics, it is the occurrence of individual transactions, and the corresponding agreed sums of all payments, expressed in the official currency, of the transacted – sold/purchased – goods and services that ought to be recorded as the 'statistical-counting-units' during a certain time span, in each geographic region,

"Price" originates only when a seller and a buyer have agreed on all conditions of the transaction, above all the amount and form of payment. The "price-per-unit" of the price-tag is useful for shoppers to compare seller's price-suggestions of competing alternative products, but is more often than not, a poor substitute for the prices actually paid. These price-tags in stores are really the "price bid" or the "seller's price proposal" but are not yet prices in an economic and statistical sense. It would be more accurate to talk of the "price of a purchase" or "the price of one specific transaction of the purchase of a merchandise" instead of talking of the "price of a merchandise."

When interpreting published price data it is important to find out what really was recorded as "price." Substitutes for the actual transaction-prices usually tend to misrepresent the reality of changes in price-levels.

Statistically Price, the main characteristic of the transaction, exists only during the short duration of a transaction,. 'Price' does not exist before or after a transaction. It is, in principle, a point-like event that takes place at a specific moment in time and a specific geographic location. This short-lived, point-like nature of price is easily overlooked because the suggested price-offers, on price-tags attached to the merchandise, displayed in store windows or listed in sales catalogues, can remain unchanged for an extended time. But those are only price-offers, price-suggestions to potential buyers, not actual prices.

Occasionally a seller publicizes a price-offer in the form of higher price-tags than he expects to sell. Typical are the prices-tags in car dealerships that offer buyers the illusory satisfaction of having been able to successfully negotiate a lower price. It is another instance where the price-tag can misrepresent the actual price situation.

If prices could be made visible when and where they occur, they would light up, at irregular intervals, like the brief glow of lightning bugs (fire flies, glow worms) on a warm summer night, everywhere and at any time.

The academic discussion, and also the statistical praxis, has proceeded under mistaken assumptions about price that led to the insistence that price-changes can only be obtained from a set of products, a 'market-basket of goods and services' that remains unchanged during the time-period for which price-level changes are to be observed. Under pressure from the rapid expansion of goods and services in the economy, the need for change is reluctantly beginning to be recognized. The difficulty to make a change stems from the ingrained historic understanding of price as a feature of the product or service instead of the transaction. Despite strong objections from purists, there is no doubt that the popularity of chain-price-indexes is growing as an underhanded form of relaxing the demand of price theory for strict comparability over time of the items in the 'market-basket.'

7.1.5 The Price Aggregates[20]

To produce a measure of price-level changes, the recorded, individual sales-transactions and their prices must be aggregated. Price statistics comes to the public in the form of price-aggregates. Changes in price levels are revealed by ratios of such price-aggregates.

Part of the problem has been uncritically extending the requirements for comparing the prices of an individual product over time to comparing statistical aggregates of prices. The conditions to compare price-aggregates over time or between regions are different than when comparing the prices-tags of individual, in great detail defined products. Because price-aggregates deal with transactions, not individual products, they retain fewer details, are more abstract, less sharply defined, which allows comparing price-aggregates that do not necessarily contain prices of the same merchandises, and justifies forming ratios of such price-aggregates

As was discussed in Chap. 3, all detail of the transaction-prices in a price-aggregate that is not specifically spelled out as part of the definition of such an aggregate, becomes invisible, disappears and becomes inexistent for all purposes. Consequently it is not necessary that only the price-aggregates of the transactions of exactly identical products can be used to follow the price development over time. Instead, the transaction-prices of similar, even different, products that belong to a given group of products, even if they differ in technical detail, can be included in the price-aggregate to report on changes in price levels over time, or to compare price levels in different geographic regions. In these price-aggregates, no single individual transaction can be identified or remains perceptible. All 'statistical-counting-units' that are included in a price-aggregate are treated as anonymous, as identical and as equally unimportant. Consequently, it is not necessary to follow only the prices of a 'market-basket' of an unchanging list of products and services. This holds for the price-aggregates of regional groups, for the price aggregates of product-groups, service-groups, industries as economic activity groupings, and certainly for the price- aggregates of the economy as a whole.

These insights justify the inclusion of the transactions of the enormous number of new products together with the transactions of older products and services of today's world-wide interconnected economies to measure changes in the level of prices on local, regional and national markets. In other words, price-measurement ought to include the great variety of transaction-prices that happen during a given time span on regional markets. Insisting on an unchanging 'shopping-basket' to measure price levels may have been acceptable in the early days of price measurement, it may have been less acceptable at the time of Laspeyres,[21] but cannot be justified any longer for today's economies.

When interpreting published price-data, it is important to find out what is used as 'price' and how it was recorded. This information can be obtained from instruction manuals for the persons charged with the collection of prices. 'Prices' taken from the price-tags of a list of products that have been kept without change for five or ten years, that once were believed to be typical, may have become sidelined, even irrelevant by the flood of new products, shopping outlets and the discounts offered,

that probably are ignored in such a price report. In addition, the type of stores where these price-readings were collected, may not represent the variety of prices to be found in larger stores and discount houses that offer quantity discounts that small local 'mom-and-pop' neighborhood stores cannot offer.

Ideally, price reporting should canvas that huge universe of all bar-code or otherwise recorded, effectively paid amounts of money of actual transactions. That random sample should be stratified by product-groups and regions, cover all types of stores and be as large as possible.[22] That sample need not be of the same size in the compared time periods or regions.[23]

It should also be noted that transactions are different in various groups of products and services. Transactions, that is, sales, e.g. in the housing market, are different from those in grocery stores or in specialized shoe stores. These different types of transactions are approximately matched by the usual product groups for which partial or sub-price-index-numbers are published. In order to interpret these partial price-index-numbers, attention must also be paid to the possibility that different procedures may have been used to produce them.

7.1.6 How Changes in Price Levels Ought to Be Measured

The preceding discussion was intended to make it clear that today's price-index-numbers are based on assumptions that can no longer be justified. The customary price-index-numbers rely on a fixed selection of a 'market-basket' of goods and services that is a non-random sample of the universe of products and services of a past time-period, of the price-tags, not of their actual sales and actually paid prices, collected only on a certain day of each month, from a few selected, supposedly typical, stores. That 'market-basket' excludes the numerous new products and services, as well as those previously existing products that in the intervening time have been improved or otherwise changed. That fixed 'market-basket' approach ignores price discounts and 'specials', and the growing availability to consumers of wholesale prices through price clubs and warehouse sales. It also ignores changes in consumer tastes and shopping habits, all matters that do affect the general price levels.

The 'market-basket' relies on fixed quantity weights from a separate household survey while ignoring the actual frequency of the sales of the included products and services which could have changed since that household survey was taken. It is hard to know in what direction and by how much the present kind of reporting distorts the reality of prices and their changes.[24] The formal criteria to judge the quality of an index number, proposed by I. Fisher, do not include the most important one, namely how well a given index-formula captures and represents the actual changes in the level of prices, which is the real purpose of price-measurement.

The following formulation uses the largest possible random sample of all recorded transactions and their actually-paid prices. Such a sample is to be selected from the bar-code readings or other written evidence of the sales of goods and services of specific groups of products during a given time span, in a geographic area, e.g. the month of October and the city of Berlin.[25] Ideally these random

samples taken from actual transactions in each time period for the same region – or for different regions in the same time period – will be of the same size. These aggregates have to include the same hierarchy of subgroups of products and services, although not necessarily including the exact same individual items. This formula can simply be written as:

$$\mathbf{PLI} = \left[\left(\sum \mathbf{p_{i,t}}\right)\left(/\right)\left(\sum \mathbf{p_{i,0}}\right)\right] \tag{7.1}$$

where p_i , transaction-price i, is the total amount of currency, paid in cash or by other means, in the ith transaction during the time segment t. The sum of these individual monetary payments, $\sum p_{i,t}$ is the transaction-price total for the intended group or class. That could be the sum for each category of products/services, or the sum for each industry, also that sum for each region, and of course, the sum of transaction prices for the entire economy.

If, as is possible, these samples for consecutive time periods or different regions are not of the same size, that is, do not include the same number of transactions, then these transaction-price-aggregates need to be converted to average values to make them formally comparable. Averaging can be done in two different ways: First, it can be computed meaningfully as "average price per unit sold (transacted)" :

$$\left[\left(\sum \mathbf{p_{i,t}}\right)/\sum \mathbf{u_{i,t}}\right]_{\mathbf{U}} \text{ or } \left[\left(\sum \mathbf{p_{i,t}}\right)/\mathbf{U_t}\right]_{\mathbf{U}} \tag{7.2a}$$

And, second, as "average price per transaction" :

$$\left[\left(\sum \mathbf{p_{i,t}}\right)/\sum \boldsymbol{\tau}_{\mathbf{t}}\right]_{\mathbf{T}} \text{ or } \left[\left(\sum \mathbf{p_{i,t}}\right)/\mathbf{T_t}\right]_{\mathbf{T}} \tag{7.2b}$$

In (7.2a) and (7.2b) p_i (for price) is the **total amount** of currency – cash or other means of payment – that was paid in the transaction τ_i, during the time segment t. in the given region or country. The sum of these individual monetary payments, $\sum \mathbf{p_{i,t}}$ yields the price total for the intended group of products, region and, eventually, the price total for the entire economy. In (7.2a) $\mathbf{U_t} = \sum \mathbf{u_{i\,t}}$ is the number of units i, sold in all the transactions during period t. Dividing by U_t determines the average price per-unit of that aggregate of all goods and services sold during period t. That 'unit' could be an individual piece, a unit of measurement like one Kg, a Liter, Meter, or any other type of measurement that is customary for a given type of product.

In (7.2b) the division of the price total by $\sum \boldsymbol{\tau}_{\mathbf{t}}$, – or $\mathbf{T_t}$, – the total number of transactions, yields the average price per transaction for each one of the groups of products, regions or the entire economy.

The double ratio[26] in (7.3a), a ratio of the second-order, becomes the **P**rice **L**evel **I**ndicator PLI_U, per unit transacted, the more important of the two price level indicators,

$$\mathbf{PLI_U} = \left[\left(\sum \mathbf{p_{i,t}}\right)/\sum \mathbf{u_{i,t}}\right]/\left[\left(\sum \mathbf{p_{i,0}}\right)/\sum \mathbf{u_{i,0}}\right] \tag{7.3a}$$

This 'Price-per-Unit' PLI_U in (7.3a) would be the principal indicator of changes in the level of prices of each group of transactions of goods and services, as well as for the entire economy.

In (7.3b) the PLI_T would show changes in the average amount of currency paid per transaction, calculated for the same groups of transactions as in (7.3a).

$$\mathbf{PLI_T} = \left[\left(\sum \mathbf{p_{i,t}}\right) / \sum \boldsymbol{\tau}_{\mathbf{i,t}}\right] / \left[\left(\sum \mathbf{P_{i,o}}\right) / \sum \boldsymbol{\tau}_{\mathbf{i,o}}\right] \tag{7.3b}$$

These two parallel price-level indicators are complementary, revealing, at the corresponding level of aggregation, two different aspects of the broader price developments that can be made for each one of the customary groups of transactions of goods and services. This would allow to study the price development in each region and for the country as a whole. Such comparisons can also be made between regions, for the same time periods.[27] Both measures consider all new products, discounts and departures from the listed price tags. Products that are not available any longer, that are likely to be included in the conventional official price reports, would be automatically excluded from this kind of measure. These PLI price measures are to be computed from the largest possible random samples that are to be selected from among the millions of daily transactions in all kinds of stores and situations, usually through the method of bar-code readings. As explained before, these transactions are like points in time and will not necessarily be of the same kind of product and service.[28] These PLI indicate changes in price levels through changes in the amount of money paid for merchandise as well as changes in the choice of products by consumers, – e.g. shifting from buying cheaper goods to more expensive luxury goods or changing in the opposite direction from more expensive to cheaper products – describing in a realistic manner, what is really going on in the economy.[29] These proposed measures are easier to interpret than the current price index numbers.

Neither these PLI price-measures – nor the current retail price index numbers – are to be interpreted as indicators of changes in the 'cost of living.' Though related, this requires a different kind of information, such as the description of the social group of households to which that 'cost of living' refers, and how and when their consumption habits were explored using only the prices of those goods that the selected population group is consuming. Retail-price-indexes are often presumed to measure both, the general price level – which they do not do well – and the 'cost of living.' Interpreting such data one must ask 'Whose cost of living? How large a sample? Are the transactions of the goods those that the targeted social class is purchasing? In the kind of stores where they actually do their shopping? What about changes in their consumption patterns?' Different population groups have different choices of products and patterns of consumption. A general "cost of living" of an entire population can neither be produced nor meaningfully be interpreted. When a regular consumer-price-index-number is called a 'cost-of-living-index' or COLI, the authority responsible for such a price measure, although it may honestly believe it, simply is in error.

7.1.7 Summing Up

The monumental changes in the historic development of modern economies require a careful re-thinking of the concepts and assumptions traditionally employed in the measurement of prices and changes in price-levels

The current approach takes for granted that "price" is a characteristic of the merchandise, that prices can only be compared over time or regions for exactly identical products/services, that at any time there exists only one single market price of each merchandise, that such a unique price exists for a prolonged time and is the same as the 'per-unit' price-tags, really the sellers price proposals, weighted with the household-expense structure of a small sample of an earlier period. This approach to price level measurement, based on these deeply-rooted assumptions, also overlooks that changes in price levels also occur because of product substitution, when buyers change their shopping habits to adjust to changes in their economic situation.

When interpreting published price-index-numbers it is important to know which products and services are included in its 'shopping-basket' supposed to represent the vast majority of products available on the markets but actually not included in the price index. To assess the real price situation, it also would be necessary to get an estimate of size and frequency of the many departures from listed prices through discounts, rebates and sales taxes. These precautions are recommended even in those instances in which a price measurement may be based on the records of actual transactions. On the other hand, the fact that some price-index-series today are chain indexes rather than fixed-based price indexes, an attempt to update the list of priced items to keep that information closer to the market reality, does not eliminate the caveats. Objections by theoreticians to this growing praxis are based on a two-fold misunderstanding: considering price as part of the merchandise instead as part of the transaction, and also failing to understand the nature of aggregation. As discussed in Chap. 3, comparing large price aggregates calls for a different approach, with less rigorous, more relaxed conditions for comparability, than the strict conditions for the comparability of individual products over time and between regions.

Some official price series have begun to experiment with adjusting the price-index-numbers of durable goods, particularly cars and large appliances like washing machines, for improved engineering features that supposedly change the nature of the product by improving its performance. This so-called "hedonic price adjustment," is based on the belief that a close relationship exists between price and quality that can be formulated mathematically, and that the prices of these products ought to be adjusted correspondingly. Such adjustment of the list-price of an improved product is aimed at re-establishing the strict comparability of its price before and after such a change happened in that product. This approach overlooks the fact that the price of actual transactions is determined to a greater extent by the competitive market situation, the bargaining power of seller and buyer, and clever commercial advertising, than by changes in the technical specifications of the car or appliance. It is preferable that the data-input into price-index-numbers not be adjusted. Prices ought to be reported without such theory-based adjustments, but as they actually occur. Instead of interfering with the reporting function of price statistics the possible effects on price of changes in the technical features of merchandise, e.g. of cars,

washing machines, computers or hospital care, should be left to separate academic research, as part of the interpretation of the price reports.

This request for keeping the price measurement free of added adjustments must also be expressed with regard to using a measurement of the general price level as an indicator of the cost-of-living. It is not possible to achieve these two goals with one and the same price measure. Either the general price level is presented objectively, regardless of who the buyers and sellers are, or a typical household budget of a specific segment of the population is priced over time, without regard for the general price situation. One index number cannot be at the same time a COGI – a cost-of-goods index and a COLI – cost-of-living index. These two possible measures require conditions that are not compatible and that one single index cannot fulfill.[30] These are interesting issues that have to be explored separately. When interpreting price index numbers it is important to determine clearly what a given price index really represents, regardless of the official claims that can be misleading.[31].

This chapter was meant to alert the reader to shortcomings of current price measurements, and provide a conceptual model for comparison, as a guide for their proper interpretation.

7.2 Index Numbers of Quantities

To get a time series of quantities,[32] which appear to be the simpler component of any index number, apparently all that is needed is to exchange the roles of p and q in a price-index-number. Quantity-index-numbers are used to report changes in agricultural production, and productivity changes in the manufacturing and service industries. Price- and other weights are used to aggregate data about heterogeneous products – the proverbial apples and oranges – reported in crates, units, barrels, tons, or in other measures, into a common time-series of produced or exported quantities.[33] The 'need' to convert heterogeneous data into a common expression of value or of man-hours makes all index number formulas rigid.[34] They become too complex for a meaningful interpretation.[35] There is no justification or explanation in the literature why such a quantity index numbers should be computed at all. The stated purpose is to summarize the divergent developments of different products or industries.[36] This goal, however, can better be achieved by forming separate time series of statistical aggregates that show the value, volume or weight of the produced or exported quantities. These time series can be expressed as simple ratios to highlight changes over time.

What really is 'production'? Let us look at some of the finer points that are relevant for statistical inquiry.

7.2.1 Four Complementary Production Concepts

The produced output of any industry, including agriculture, is a flow phenomenon, projected, so to speak, by the harvested crops and fruits, the produced men's suits, cars, kitchen appliances, etc. This flow phenomenon happens in different areas of

the country, during the seasons of the year, in all economic activities, including every kind of products.

Along the production flow, many different observation points can be chosen where the phenomenon 'production' can be monitored. Assuming the data are adequate,[37] there are important neglected problems of a conceptual nature in production statistic that need to be clarified. Let me discuss these for the situation in agricultural production.[38] Some field crops are reported as 'produced' in the first phases of the harvesting process, some crops are reported in later stages of elaboration on the farms, some during the final sale of the products, or even in other industries which acquire the agricultural product as raw material. Hardly two products are reported in the same stage of processing, with the same deductions for loss, waste or consumption within the producing farm.[39]

The production of most products is a continuous process that begins with the setting up of a production run, or with the preparation of the soil in agriculture, and ends with the sale of the product. It can be imagined as a straight line on which the beginning and ending are not points but short segments. In fact, the beginning as well as the end itself is a process with a start, a certain duration, and an end. When magnified, each of these stages itself also has a pattern of beginning, development and ending. There are at least three different and distinct points in the end phase of the production process at which the items in production can be considered as 'produced.' Some of the physical characteristics, quantities, place and time differ in which the products become produced (Fig. 7.1).Each of these possible endpoints along the production flow corresponds to a different concept of production. The first two are technical concepts, based on technical considerations, be that engineering or agronomical in nature. The third concept is both a technical and a monetary concept. The fourth concept is a monetary-economical concept only, net of costs and other deductions. These different production concepts usually are associated with different quantities, locations and times of occurrence. The amount of labor and capital investments for each of these production concepts also differs. It is, therefore, not indifferent for the interpretation of production data, where along this flow through the production process statistics has observed the reported output.

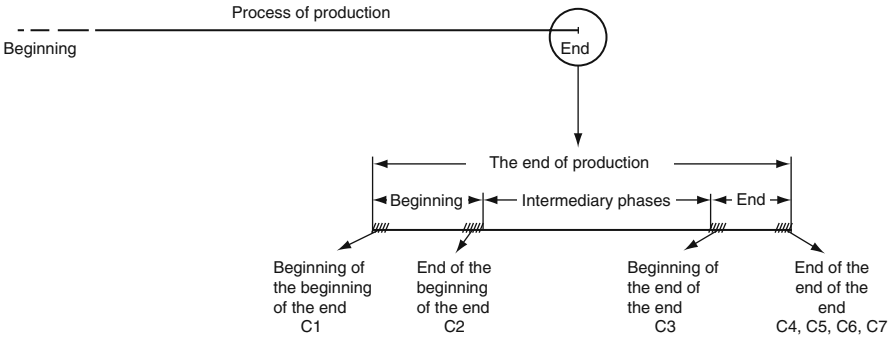

Fig. 7.1 Production concepts in the flow of production

C1 (Production Concept 1) is an engineering or technical-agronomical concept of production. It refers to the moment when the finished car or washing machine passes the last inspection and becomes ready to be moved off the production line or agricultural field to some storage location. In the case of the harvest of potatoes it is the moment in which the potatoes are dug from the ground, passing from nature into the disposition of the farmer. The production process has been completed and the product is finished in a technical sense. Industrial products have been examined or tested and presumably function as intended. The produced item is on the factory floor, or on the fields of the farm, ready to be stored or to be sold. The corresponding quantities, particularly in agricultural production, are 'gross' in a technical sense; as leaves, shells, etc. may have to be removed later. The produced quantities correspond to the 'beginning of the beginning of the end' of Fig. 7.1 Excluded are those quantities that had to be discarded or got destroyed during the production run, or that were lost during the harvesting process. On the other hand, the figures of this concept still include those produced items which will be lost, damaged, stolen, or destroyed later. The data of this concept may coincide with the last pre-harvest crop estimates, and meaningful yields-per-acre ratios can be established only with these figures of C1. They are also essential to determine the need for labor, transportation and storage facilities, to guide economic and management policies on production. The product has not yet been sold and consequently, these quantities are not yet 'produced' in a monetary and economic sense. Most of the labor input up to this point is direct production labor. Only an insignificant part is non-production labor, for supervision, time keeping, security, etc.

C2 (Production Concept 2) is another technical concept. The same finished product now has been packaged or crated, has been safely stored, e.g. in a centralized location on the farm, e.g. in a barn, and is ready to be shipped. Losses may reduce the number, quality or size of the produced items. Although not yet sold, these items may be transported from the production site to points of sale. They are not yet 'produced' in a monetary-economic sense. Combined with the inventories at the beginning and the end of the same period and the figures of C3 for the same period, shrinkage, theft, waste, accidental damage and other losses during the production process can be evaluated. There can be substantial differences between the numbers of a class of technically finished items being 'produced' according to the production concept C1, and concept C2 because more labor and capital have been added through packing, shipping and storing. In agricultural production, the figures of C2 are 'net' in an agronomical sense[40] and allow to estimate the availability of nutritional values and raw materials for the industries that process these crops.

C3 (Production Concept 3) refers to the moment of actual sale of such technically 'produced' items. Money is promised or paid. Only when the sale has been finalized – such a sale itself can be a complex procedure that takes some time, (see Figs. 7.1 and 7.2) – can a given product be considered also 'produced' in a monetary-economic, marketing sense. Production concept C3 refers to the sold quantities of produced goods in a given time interval, and to the actual receipts in current market values. The time and place on which these quantities are reported may also differ from those of C1 and C2. In agriculture, for example, the figures of C3 refer to the

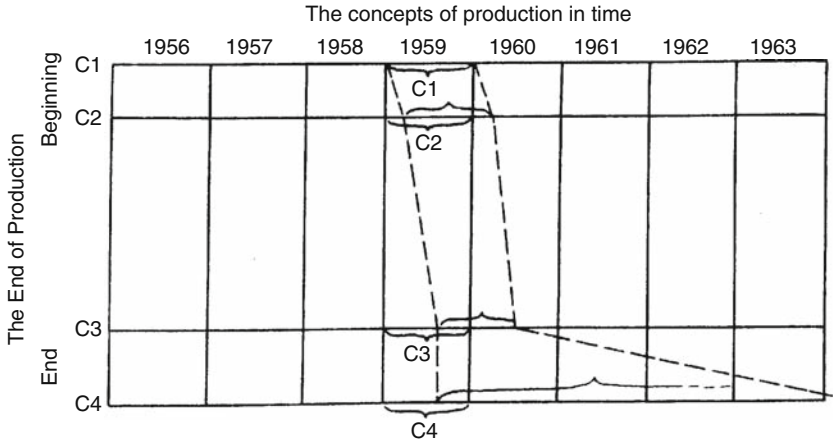

Fig. 7.2 Time relationship of the production concepts

moment of the sale of the crops, describing the final (gross) economic outcome, are given in physical measures of quantity (e.g. in bushels, tons, etc.) and in the actually received monetary values. The C3 figures are available separately for each product, as part of the economic gross product of the firm. Costs are not considered for the values of C3.

Not included in the production concept C3 are the parts of the goods that were produced but not sold in this time period, those parts that were consumed in the own production process, such as retaining parts of the crop for seed, those lost during transportation or storage, and those given away without sale. On the other hand, included in a time interval's C3 production figures are crops and products sold from the inventory of items that had been produced during a previous time period. The produced quantities of C3 for a certain time interval, therefore, may differ substantially from the quantities of the production concepts C2 and C1.Only when the items are also sold at the moment in which they become technically finished – 'produced' – do the production figures of C1, C2 and C3 coincide. The comparative analysis of the figures of C1, C2 and those of C3 gives a picture of the consumption- storage- or inventory situation in an industry, as well as the degree to which an industry (e.g. animal husbandry) of a region is associated with the market.

C4 This production concept is a purely monetary-economic concept. It is a refinement of C3, also taking costs into consideration, as well as all deductions, such as returned merchandise and sales that later were cancelled. Depending on the degree of sophistication and on the industry, C4 really can be considered the 'value added,' or the net profit of production. C4 covers an entire group of net production concepts. It is no longer a product-specific technical and engineering production concept, but refers to the net sales of the output of the farm, business firm or industrial organization.[41]

The produced quantities of concepts C1 and C2 should be used when the technical efficiency or productivity of the production factors is to be studied. The goods or

services produced according to production concepts C1 and C2 can be reported as the number of items produced, their weight, and/or their volume, and in the case of services, the hours of service rendered. That production can not yet be reported as a monetary value. In many industries, particularly in agricultural production, it also must be clarified which economic activities really belong to it. Data on agricultural production are often taken at later stages of further industrial processing in derived non-agricultural industries which may be vertically integrated with farms.

In those cases, the produced quantities reported for C3 are smaller because certain parts have been removed, recorded in different time periods and sometimes in different places. They also report a bigger C4 or 'value added'. If aggregated with the production figures of other farms or factories that are not vertically integrated to the same extent, the interpretation of such mixed quantity aggregates becomes questionable. Obviously the aggregates of the mix of quantities of different production concepts are less meaningful.[42]

Figure 7.2 shows the timing of a given year's production according to the four production concepts C1 to C4. It shows that the output produced during a given time interval can extend over subsequent time intervals, but also the quantities produced according to concepts C2, C3 and C4 of that same time interval can include quantities produced in previous periods e.g. the production figures given on a calendar-year basis do not necessarily show only the results of the production efforts of that year. A similar graph (not shown) could be constructed for the geographic origin of the production figures that correspond to each production concept.

Although four different production concepts seem to be excessive – in some industries e.g. agriculture, even more than four production concepts could be distinguished – they simply recognize the reality of a complex situation, providing a new perspective for index numbers of quantities.

7.2.2 What Is Quantity?

The produced quantities of concepts C1 to C3 can be reported in different ways, a matter that further complicates the situation. These alternatives ways are:

1. the number of produced items or harvested crop units,
2. the physical weight of the produced items, e.g. in tons,
3. their physical volume, e.g. in bushels, crates, or,[43] in some instances, barrels
4. their length e.g. board feet, or their surface area
5. in other ways such as the protein content of a crop.

The counted number of the produced items, the unit count, is the basic quantity.[44] All other measures are derived and complementary, and are not as important as the unit count. It does not matter whether the weight, volume, length or surface area of the produced items is ascertained directly, or, as in the case of the protein content of a crop, derived by applying sample estimated 'average protein content per unit ' to the reported units. The monetary value, though, of these produced

quantities exists only for C3 and C4. The rules for grouping the produced items into classes by product or industry are the same as those discussed in earlier chapters. Everything stated there – especially in Chap. 3 – about aggregation of 'statistical-counting-units' also applies to the quantities presented in production data, including index-numbers.

Let me refer to the lowest, first level of grouping at which the 'statistical counting units' are added, 'groups of the first order' *or* 'groups of the lowest order'. Reading Fig. 7.3 from the bottom up, the process of aggregation is demonstrated for the single product 'Wheat.' On closer look one discovers a surprising variety of 'wheat'. All crops of e.g. the variety 'hard red winter wheat' are added to form a homogeneous group, despite the fact that there are some regional variations in the length of their spikes, in the size of the grains, their protein content, weight, color, etc. forming subspecies of 'hard red winter wheat'. Fig. 7.3 shows eight other, similarly homogeneous first-order groups of North American wheat varieties. The totals of each one of these different varieties of wheat can be considered as sub-totals, and can be grouped together, forming the category 'Wheat,' which is a *group of the second-order*. In a similar way, corn, rye, etc. form crop totals, each of which is composed of many lowest- or first order groupings. At the next higher level of forming conceptual groups, the groups of the next-higher order: 'Wheat,' 'Rye,' 'Barley,' etc. can be added to a group of an even higher level, 'All Grains.' Note that the data of e.g. the subgroup 'Wheat' could not directly be added to the data of the sub-group 'Tomatoes,' or any other higher-order grouping. In general, statistical groups of the nth order must include all available lower groups of order n-i (i = 1, 2, 3, . . ., n−1). This aggregation process must be conceptually complete at each level of aggregation before it can be joined with other, similarly complete aggregates to

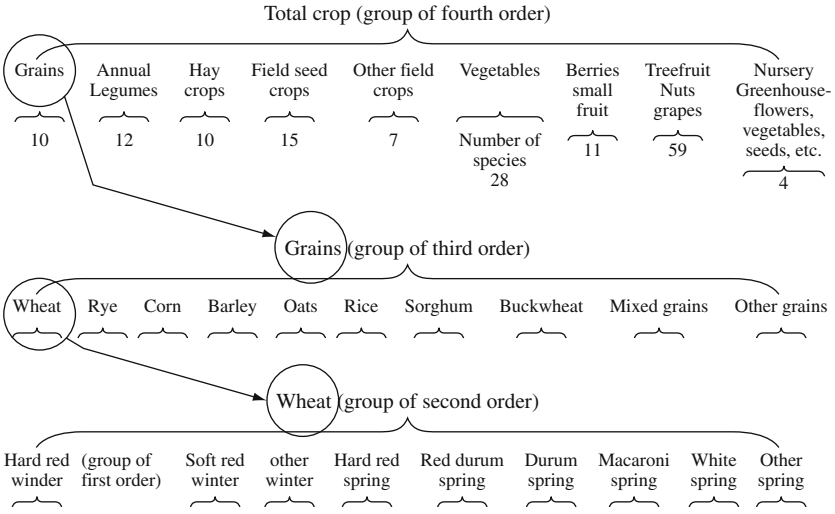

Fig. 7.3 Hierarchy of statistical groups in statistical aggregates

an aggregate of the next higher level. Aggregation must be conceptually exhaustive, proceeding in layers.

The detailed knowledge of the harvested species is lost when merging or aggregating these different species of products in an aggregate or group of the next-higher order. Only those few features of the 'statistical- counting-units' remain that are common to all items in the aggregate, as discussed in Chap. 3. When interpreting the comparison of the statistical aggregate 'All Varieties of Grain' for different regions of the country at a given time, or for one region over successive time periods, one should expect to find in these aggregates different crop varieties, even different kinds of grain species. Obviously, the comparison of such inclusive groupings has only a limited meaning. Though possible, such comparisons are made frequently and must be interpreted with the awareness that these aggregates do not have a simple, concrete meaning. Few data users will want to add up the produced physical quantities of the group 'grains-legumes-vegetables-roots-berries-fruit-Nuts' into the highest level-aggregate 'Total (physical) Quantity of Agricultural Production'. Obviously, no practical purpose would be served by such an aggregate of the physically produced items, as 'Total count of all the products harvested (produced) in agriculture, anywhere in the USA at any time during the year 2008'. Besides the few common features, nothing is known about the distribution within. Such an aggregate definitely should not even be attempted.

While the nebulous nature and doubtful usefulness of such a huge aggregate is obvious for aggregates of unit-count figures, the situation is far less obvious if those same heterogeneous items are converted – with the help of conversion tables – to volume or weight. The transformation of all items into their physical weight, e.g. into tons, according to conversion tables of period (t), w_t, of course, does not change this lack of conceptual affinity of the heterogeneous items in such an aggregate. It is to be interpreted with the same reluctance as an aggregate of the unit-count of these same items. What is different, though, in such an aggregate, is the role played by the physical weight of the items. The heavier items are given a greater importance in the aggregate. If unit prices were used to weigh the different items, the distortion in the aggregate would be even more pronounced, giving a disproportional weighting to expensive items. In all these instances the question about such an aggregate would be: 'the total weight of what?' or 'the total value of what?' Although statisticians assure us that heterogeneous items cannot be added, prone to mention that 'you can't add apples and oranges', they do not seem to feel that way after 'converting' these items e.g. into their market price, weight, or volume.[45] In reality the glue that holds the items together in a statistical aggregate is their conceptual affinity, regardless of how tenuous that may be, which in turn affects the interpretation of such an aggregate, as discussed in Chap. 3. At any rate, no transformation into their physical weight, volume, value or whatever is necessary to be able to aggregate items.[46] Yet, that is done in quantity-index-numbers of industrial and agricultural production. In fact, the index-numbers of quantities are a confused mix of different conversion rates, in addition to mixing different production concepts for each produced type of products, whereby each product is reported in another modality of the produced quantities.

7.2.3 The Role of Weights in Production-Index-Numbers

At the heart of the construction of price index numbers was the unquestioned conviction, that the 'price' p of a product is the per-unit monetary amount on the price-tag, and that quantity weights q were needed to aggregate these prices per unit. By simple inverting the roles of p and q, these price index formulas could also be used to report produced quantities. The following six arguments are used to justify the need for weights:

Argument 1 'Weighing quantities with prices allows to transform quantity figures that are reported in different kinds of measurements e.g. in board feet, barrels, crates, KWH, sacks, etc, into a common expression of monetary value $\Sigma Q_t P_0$.'

This seemingly ingenuous solution not only ignores that there are differences in the produced quantities according to the different production concepts, but it also ignores that the market price, the physical weight or volume of a given product change independently of each other. Consider the production of motors of a certain kind. Through changes in materials or changes in design, e.g. replacing steel parts with parts made of hard plastic, these motors may become heavier or lighter, larger or smaller and the value added by the producer will vary according to the number of parts purchased from outside suppliers, in addition to the fluctuations in the prices at which they are sold. A time series of the **number** of motors produced reports about a different feature of that production than a time series of their **physical weight**, or their **volume**, their g**ross sales valu**e, or the **value-added**. Each of such time series would describe a different aspect of that same production results. The relationship between these different aspects of the motors also changes over time.

To interpret such data meaningfully**,** one must find out if all the produced items are reported in the same modality and measuring unit. If the physical weight of the produced items is of concern, e.g. for the purpose of determining appropriate means for their transportation, then the physical weight for each item in the time-series should be presented. Usually the production reports for different products in such a quantity-index-number are given in different modalities, e.g. some are directly given as the number of items produced, others in measures of their length e.g. board feet; others may be reported in measures of their volume, such as the number of crates, barrels or sacks, while others are reported in measures of their weight, e.g. tons. Such heterogeneous reporting could be standardized with the help of up-to-date conversion tables with the necessary transformation coefficients e.g. expressing all products in terms of their physical weight: the reported board feet of one product into tons, the barrels of another product into tons, the reported sacks of another item into their weight in tons, etc. Based on samples, these conversion coefficients ought to be kept up to date for every product, in each region. Although this means additional work, interpreting the resulting series will be more meaningful because the produced items are all expressed in the same type of measure. Multiplying the differently measured quantities by their 'prices' – the discussion in Sect. 7.1 showed that 'price' itself is not a simple phenomenon – of these items is irrelevant for transforming the sacks of one product and the unit count of another into an appropriate common reporting standard of these products.[47]

Argument 2. 'Weighing quantities with unit prices or with other weights allows to cover gaps in statistical reporting.' Whenever production data for a given product cannot be obtained, the weight corresponding to this product can be 'imputed' to the weight of another, related product, for which data are available. The weight assigned to each product indicates its relative importance in the total structure of production and looks like an inexpensive device to fill gaps in reporting.

Imputation is based on the assumption that the product, whose information is missing, develops parallel to the production of another product for which data are available. Such an assumption cannot usually be proved for the same reason for which imputation was believed necessary, because its data were unavailable. Even if it developed parallel in the past, there is no guarantee that it will continue to do so. When production-index-numbers were first established, statistical information was scarcer, and sampling not yet widely used. Imputation, then, was an inconspicuous and convenient stop-gap procedure. Today statistical reporting is better equipped to cover the important aspects of production, and sampling has become a sophisticated, practical technique. Imputation, then, though cheap, is an inferior means to estimate the otherwise unobtainable information.

Argument 3. '*Transforming* dissimilar objects, such as computers and shoes, *into* a common expression of *their economic value* seems to be a simple way out of the logical impossibility of aggregating dissimilar objects, like adding apples and oranges' This reasoning follows the same misguided logic about statistical aggregation as discussed in Chap. 3. It also ignores the existence of different production concepts. At any rate, there really should be no need for price or other weights.

Argument 4. 'Data *on physical quantities* must be "*weighted" by their prices* or sales values *so that each product* in the aggregate *will be represented according to its economic importance*.'

This most oft-heard argument in support of price-weights is another consequence of ignoring the different concepts of production, the internal process of aggregation and how to interpret such aggregates.

Argument 5. 'Prices must be held constant so that the changes in the produced quantities can be observed free of price changes.' This argument implies, that to observe the 'true' development of production, the prices of these products must be held constant, as if multiplying the produced quantities with a base-period price was the same as actually freezing prices at the base-period level, like in a controlled lab experiment. This reasoning is inspired by the success of controlled experiments in the sciences.

Suppose one really could control the economy, forcing the prices to remain the same as in the base-period. The quantities produced under such controlled circumstances would in subsequent periods certainly differ from the quantities that were actually produced without such price controls. It is important to understand that the algebraic manipulation of the produced quantities q_t in period (t) with the base period t=0 prices p_0 of these products cannot have the same effects as actually being in a position to control the markets. by freezing the respective prices of these products and being able to enforce such a 'freeze'. The obvious futility of this argument has been the ironic point of one of W. Somerset- Maugham's short stories.[48] It is

basically that bank director's kind of reasoning that is used to justify price weights in quantity indexes and, analogously, insisting on using quantity weights in price indexes.

Argument 6. '*Price* and other weights really are of subordinate importance in quantity-index-numbers.' In fact the opposite is true. Whatever the weights – gross values, value-added, hours worked, yields per acre, etc. – it is these weights that a 'quantity- index-number' extends over time. In reality the **weights determine the meaning** of a 'quantity- index-number'.

Consider a monthly production index number that combines individual series of the numeric count of the produced items, while the weights are the values-added by these products in a base period. Although such a series seems to be an index number of the produced physical quantities, weighted by the value-added of each of these products, it really became a 'value-added-index-number' in which the produced units were proxies for the unavailable monthly or yearly value-added figures. They serve to extend forward over time, the values-added of a base period used as index weights.

Both, the series and the index weights ought to be of the same kind. In practice the choice is limited by the availability of data for the series to update the weight structure of the base-period. The quality of a monthly index-number series depends on how well the selected products represent the production of the entire economic sector that the index represents, and how closely these selected series are related to the choice of weights. If measuring the GNP (Gross National Product) concept is the purpose of an index-number, but GNP figures are not available, the next best indicators would be the value-added figures, both for the industry weights and for the monthly or quarterly series updating these weights. Other types of information will yield weights and series that are even less suited for that purpose. Depending on the degree to which the chosen series are related to e.g. GNP values, a quantity index number can be judged by the proportion of weights and series that are excellent, good, fair or poor substitutes.

Each cell in the following Fig. 7.4 stands for a combination of base-period weights and representative monthly, quarterly or yearly production series. Dotted lines cutting midway across these cells show frequent combinations between types of series used, in the head of the table, and types of weights, in the stub of Fig. 7.4. The corresponding 'quantity-index-numbers' result in some cases in a 'Good' index number of production, in most cases only in a 'fair' index number, in some cases a 'poor,' or 'very poor' index number. If, on the other hand, gross-value-series are used in conjunction with GNP weights, sometimes an 'excellent', more often a 'good enough,' only 'fair,' or poor index number may result. These ratings refer to the adequacy of the conceptual framework of 'production-index-numbers', not to the quality of actual coverage of this socio-economic situation. A similar graph could be drawn for other kinds of index numbers, such as indices of land usage or of labor requirements.

To interpret a quantity index number the content of the individual production series, as well as their weights, are to be carefully scrutinized. In light of the previous discussion, an index-number of e.g. physical production with value-added weights

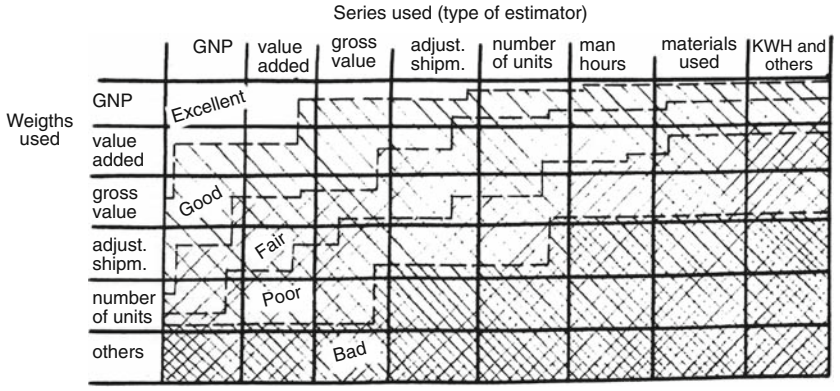

Fig. 7.4 The quality of different time series carrying forward different types of weights

should be recognized as an indicator of the values-added in a base period that are extended over time and carried forward by different quantity series as estimators.[49]

All in all, weighing with prices is problematic for a variety of reasons. Under the pretext of practical convenience, or theoretical necessity, price or other weights complicate and confuse the issues that 'quantity-index-numbers' were meant to report. To interpret production data meaningfully, attempts should be made to obtain conceptually pure time series, based on the quantities of any one of the production concepts C1, C2 or C3, and of one series of monetary values of production.[50]

A careful rethinking of the concepts of 'output,' 'price' and other components that 'measure' output and productivity in mining, agriculture, construction, manufacturing and services, should lead to statistical measures that show the actual socio-economic situation more clearly and are easier to interpret.

7.3 Measures of Factor Productivity

Productivity measures often are computed as ratios between a 'quantity-index-number' of produced output and an index-number of labor input. These 'ratios-between-ratios' – see Sect. 4.5, or higher-order ratios – are to measure the effectiveness of the factors of production, usually of labor. In constructing 'output-index-numbers', no distinction is made between the different production concepts. Only C1 and C2 series allow meaningful comparisons with the factors of production employed. Here also, the use of weights confuses the issues. Aggregation, properly understood, has no need for any kind of weights. The customary index-number formulas of Laspeyre or Paasche should not be used as measures of produced quantities. The same holds for the labor inputs, which do not need weights in order to be aggregated. In fact, weights make it more difficult to make sense of these ratios. The productivity-index-numbers, as computed today, do not measure correctly changes in the effectiveness or 'productivity' of any one production factor, in particular of the production factor 'labor'.

7.3.1 The Production Factor 'Labor'

Of all inputs into the 'production' of a good or service, labor has been the most important, the simplest to obtain, and easiest to interpret. Variations in the number of hours worked seem more directly linked to variations in the quantity and quality of the end product than variations in any other factor of production. Long-run improvements in output and productivity, however, are not usually due to corresponding changes in labor. When fewer hours are needed to produce the same amount of output, automation, improved product design, introduction of more efficient machinery and tools, improved work-layout, and improvements in management usually are the reasons. If key personnel in the production process receives additional training thereby upgrading the efficiency of 'labor directly involved in the production process', that ought to be credited to 'management' as a separate production factor, rather than to the catch-all category 'labor'. One might even conclude that improvements in the labor-output ratio usually are due to everything else except for that labor, that is directly involved in producing the goods or services properly speaking. In the aggregate at least, it is difficult to attribute specific improvements in productivity to any one of the production factors, such as 'long term capital investment,' 'land and natural resources,' 'management' or to 'labor on the factory floor'. In lieu of multivariate studies, it has been customary to use the hours worked by 'production labor' as the next-best available and seemingly reasonable indicator of changes in the work situation. It must be kept in mind, though, that there is a narrow range within which a given worker can improve his speed, skill and accuracy. Most of the older, experienced workers already function on the high end of their learning curve. The margins are narrow within which the efficiency of individual workers, and that of the work force as a whole, can be improved.

Productivity measures aim at technical efficiency, not at the cost effectiveness of a production process. Productivity measures, therefore, ought to refer to physical output as reported with C1 or C2. The monetary concepts C3 or C4 introduce issues that have little to do with the technical efficiency of the production process, such as special discounts for package deals or large-quantity purchases, clearance sales, the inclusion of services, such as 'free' delivery, insurance, parts- and repair warranties, in addition to the vagaries of the market. Whenever the monetary production concepts are used, the **cost** of labor, not the number of hours worked or the size of the work force ought to be used. When C3 or C4 figures are used, the issue shifts from measuring the effectiveness – or productivity – of labor, to measuring its cost effectiveness vis a vis the other production factors.

7.3.2 Other Production Factors

Human skills, unaided by tools or machines, are relatively stable because workers quickly reach the ceiling on their learning curve. That makes it doubtful that

increases in output are due to a corresponding increase in skill levels, speed and accuracy, in short, in the efficiency of the workers. An improved 'labor productivity' is usually due to more and better capital investment and better management, two other important but usually invisible factors of productivity. By taking into consideration only the amount (hours) of labor, and the amount of product produced, and not the improvements through a simpler engineering design for faster, less prone to produce defective products that have to be reworked or discarded, changes in the 'labor-input / product-output' ratio are hardly related to that which human labor properly speaking, as a separate production factor by itself, accomplished. When the data on labor productivity show improvement, that result may be due to everything else except to improvements in the production factor 'labor' itself. Not included in that statistical measure are important matters such as improved skill levels and schooling of the production workers, their motivation, inventiveness, and cooperation with management,. The gains in productivity over time for individual products, groups of products or an entire industry, reveal less improvements in the productivity of 'labor,' as changes in the skills of management and in the availability and proper use of short and long term capital investment.

Summing up, measures that report changes in productivity are formed by ratios between a 'quantity-index-number' of produced output and an index-number of labor input. These higher-order 'ratios-between-ratios' (see Sect. 4.5) are affected by every problem concerning the interpretation of 'price-index- numbers', discussed in Sect. 7.1, 'quantity-index-numbers' and index-numbers of labor input. These index numbers, ill suited for measuring actual price movements, are even less suited to measure the true changes in the productivity (efficiency) of 'labor'.[51] The link between hours of human effort spent in the production of an item, and a particular final product is tenuous, except in the special case of a small-scale production of e.g. 'orthopedic shoes made to measure' by one shoemaker. That and other kind of handcraft productions may be the only cases where one really can determine the number of hours of manual effort to make one product, e.g. a pair of shoes, and changes in the productivity of human labor.

When dealing with aggregates of products, the timing of input and output further blurs the relationship between the produced output and the input of labor because only part of the labor applied during a given time period is incorporated in that period's output. The longer and more complex the production process of a product, the less of the labor input during a given time period is actually applied to the physical output produced during that same time period (see Fig. 7.2). How then should one interpret data on productivity? Keep these critical considerations in mind and get as much information about the productivity data at hand as possible. Don't be shy to ask questions about the concepts used, the reliability and scope of the actual sources of the collected data and quite generally, get information about the production processes in that particular industry. At any rate, be on guard when interpreting the data of the supposed 'gains in productivity' over time.

Notes

1. The following sections draw heavily on thoughts published in Winkler, Othmar W., (1962) "A New Approach to 'Measuring' Agricultural Production" *Proceedings of the Business and Economic Statistics Section of ASA*, Washington, D.C., pp. 30–36, and (1985) "The Actual Information Content in Statistical Productivity Measurements", presented at the 45th Congress of the International Statistical Institute, *Book 2, Contributed Papers* pp. 391–392, ISI, Amsterdam 1985.
2. In American textbooks on business and economic statistics the measurement of prices is treated as a historic peculiarity and usually placed as one of the last chapters, disconnected from the general probability based theory of statistics. Part of this special treatment is the consequence of the convoluted manner to capture the reality of prices in complicated index-numbers creating pseudo-problems that overshadowed the task of reporting the movements of prices. Given the lack of interest in the descriptive function of statistics in economics, no need was felt to clarify what exactly the real-life-objects in price statistics are, or ought to be, from which prices, as the statistical-counting-units, are to be recorded.
3. Richard Maul, "Laspeyres als statistisch-nationalökonomischer Forscher" Inaugural Dissertation zur Erlangung der Doktorwürde der Wirtschafts- und Sozialwissenschaftlichen Fakultät, J.W. Göthe Universität, Frankfut a.M. Rockhausen 1930. Also: "Laspeyres index, proposed by German Economist Etienne Laspeyres for measuring current price or quantity levels relative to those of a selected base period. . .*Encyclopedia Britannica, Micropedia*, Vol. VI, p. 61, 1976.
4. "Studien über die Natur der Geldentwertung und ihre praktische Bedeutung in den letzten Jahrzehnten. Auf Grund Statistischen Detailmaterials". Dr. Herman Paasche, Privatdozent zu Halle A/S. Jena, Verlag von Gustav Fischer. 1878 Ferner: 'Paasche Index' index developed by German economist Hermann Paasche for measuring current prices or quantity levels relative to those of a selected base period. . . *Encyclopedia Britannica, Micropedia* Vol. VII, p. 661, 1976.
5. David Blackbourn. *The Long Nineteenth Century – A History of Germany 1780–1918*, Oxford University Press, New York, Oxford, 1998.
6. To give just one example of the multitude of new products, the "*The Washington Post*" of January 22., 2005, Section F 1 and F 12 "A New Twist on Everyday Stuff – the latest in Technology Design and Gadgets for Homes" newly invented and available products such as e.g. "An outdoor Ring of Fire combines Flames and Fountain", or "Steam vac-duo Vacuum Cleaner", "Efficient, longer lived LED Floodlights" a "Fabric Freshener Clothing Steamer", "Garage Doors of Steel and Glass" to mention just a few of the new products described in that news article.

   Another hint of this phenomenon comes from the "Harvard Business Review." The article "Innovation from the Outside in" proposes "To trump rivals, you must drive top-line growth. And that means **generating a constant stream of new products and services**. But if you're depending solely on your R&D team to innovate, you're putting your company at risk. Why? In many firms, R&D productivity is flattening while innovation budgets are climbing faster than revenues. A solution? Gather promising ideas outside your company – from other industries and from vendors, customers, even competitors. Then develop those ideas into new or refined offerings-quickly and cheaply. . ." *HBR ONPOINT ALERT* – March 2006. Obtained via e-mail from <Harvard_Business_Online@hbsp.ed10.net>.
7. "Advertising messages are one of the building blocks that constitute the massive commercializing trend in our society. But the start is really with the availability of products everywhere. These in-your-face strategies include intensive distribution with roll-outs of products into all major distribution points (Levi's, Nike) and the new development of dense networks of distribution outlets (Starbucks, McDonald's). Products are now available in more locations and stores are open longer hours than ever. Not only are there more shopping malls with more stores and amenities than ever, the various brands are available in a wider variety of stores, from their own boutiques, to department stores, to discount stores and factory outlets, to ware-

house outlets – not to mention gray goods, pirated copies and counterfeits. . . . in the U.S. You can buy almost anything at any time of the day if your really want to. Online shopping means that you can buy a used car or a new camera, survey the new fashions, do your banking and reserve an airplane ticket any time of the day (and night). To use a marketing phrase: "You've come a long way, baby!" (p. 27) "The sheer volume of products and services, coupled with their ever shorter life cycles and the accompanying promotional noise, would lead one to suspect that consumers in the U.S. are less in control of things than before. . . the most recent research suggests that things are even worse than before" p. 50

"Today there is a great deal of technological innovation and a high rate of new product introduction in categories as diverse as detergents, toothbrushes, shampoos, water, wines, sneakers, retailing, computers, even financial products. . ..we choose where to buy on price and what to buy on brand." (p. 32)

Johny K. Johansson, *In Your Face – How American Marketing Excess Fuels Anti-Americanism* FT (financial Times) Prentice Hall, Upper Saddle River, NJ, 2004.

8. See also "Executive Summary" the Boskin Commission: "The American economy is flexible and dynamic. New products are being introduced all the time and existing ones improved, while others leave the market. The relative prices of different goods and services change frequently, in response to changes in income and technological and other factors affecting costs and quality. This makes constructing an accurate cost of living index more difficult than in a static economy. . ." pp. i–iii, *Toward a More Accurate Measure of the Cost of Living* Final Report to the Senate Finance Committee, Michael J. Boskin, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, Dale Jorgenson, December 4, 1996.

   The numerous profound changes in the markets and in the American society that happened within a short time can be noticed convincingly in the statistical tables of the studies of the BLS (US Bureau of Labor Statistics). Charles Mason, Clifford Butler, "New basket of goods and Services being priced in revised CPI" Monthly Labor Review, January 1987, pp. 3–22.

   ". . .children themselves are under fierce advertising and peer group pressures to want everything material and everything prestigious that is available in the society and to be deeply convinced that this is their right, which their parents are failing to satisfy. Advertising, as we know it in our culture, has gone far beyond information and invitation. It works quite insidiously to create discontent, especially in young people, to persuade them that they will not be respected by their peers and cannot be happy unless and until they possess whatever item is being advertised. . ." p. 101, Monika K. Hellwig *Public Dimensions of a Believer's Life*, Rowman & Littlefield Publishers, Inc. Lanham-Boulder-New York-Toronto-Oxford, 2005.

9. Peter von der Lippe defends the theoretical necessity of the strict comparability of the products whose prices are to be compared, against the overwhelming pressure of the changing actuality. In his carefully thought out discourse he insists on the need for the purity of comparisons of strictly identical products, defending it against the theoretically unsatisfactory compromise of the yearly changing shopping basket of the chain price index numbers.

   Peter von der Lippe, *Chain Indices, A Study in Price Index Theory* Volume 16 of the Publication Series Spectrum of Federal Statistics, Verlag Metzler- Pöschel, Stuttgart, 2001

   The Boskin Commission, though, arrives at the opposite conclusion: ". . .BLS should move away from the assumption that consumers do not respond at all to price changes in close substitutes . . .these moves would alleviate the problem of **the growing irrelevancy of "baskets" based on decade-old consumption patterns**, reduce significantly the substitution and formula bias, and facilitate the speedier introduction of new goods and services into the index. . ..the BLS **should organize itself for "permanent" rather than decadal revisions of the CPI.** Both the weights and the priced commodity and services assortment need more frequent updating. . ." p. 79, 80 ". . .**moving to a notion of a new "basket" each year** will allow for a faster introduction of new items and new outlets. . .":p. 82

   Toward a More Accurate Measure of the Cost of Living, Final Report to the Senate Finance Committee, Michael J. Boskin, etc. op.cit.

10. "Almost all of what is today the philosophy and practice of applied statistics derives from ideas originated by (Ronald A.) Fisher" John C. Gower, Fisher Memorial Lecture at SMPQ, *Amstat News*, American Statistical Association, Washington, D.C. February 1995.
11. An exception is Oskar Morgenstern's description of the complexity of transactions of wholesale deals, and how their prices are formed, which comes very close to the concept of "price" developed in the following by this author.
    Morgenstern, *Oscar (1963) On the Accuracy of Economic Observations*, Princeton University Press, 2d ed. , notably p. 185.
12. Price differentiating has become an important tool of marketing and is used with great success but has complicated the situation in the markets. Newly available information on the internet seems to clarify the situation.

    Another example taken from a typical weekly shopping trip at a Safeway store (a chain of grocery retailer-supermarket, selling groceries, paper products, pharmaceutical and cosmetic products). On the print-out of one typical bill was the price of one purchased item shown as **"3/22/06 Turnover apple, 4 CT – Total Price $3.39, Price with Safeway Club Card $2.50. – you save $0.89 with Card"**. Such 'club cards' can be easily obtained, intending to foster customer loyalty in the competition with other, similar grocery chains such as "Giant", "Food Lyon" '"Price club," or "Costco". It is another example of the importance not to rely on price tags for a realistic statistic of prices, but to rely on the actually paid prices.
    "An Index of Prices has more to report than the price once the Myth of the Unique Price is abandoned . . .." p. 39 George Stigler, James F. Kindahl, *The Behavior or Industrial Prices*, NBER, New York 1970
13. A memo sent to the faculty of Georgetown University's McDonough School of Business, dated June 23, 2003, offered the (at that time) latest version of the "Software for econometric analysis, Stata 8" at the following prices: US $ 469 for the purchase of one Program, US $ 347 for each program with the purchase of two programs, for the purchase of 3–6 Programs, US $256 per Program, and for the purchase of 6–10 Programs $215 pro Program. The "price" for this item clearly depended on the prices paid in actual transactions. Reliance on the price-tag for one unit is bound to misrepresent the situation.
14. Such discounts are granted for all sorts of pretexts e.g. for senior citizens, or because it is Saturday – the infamous Saturday-night specials in the arms trade, or when the buyer is a member of a price club, or can present coupons that can be obtained free from newspapers and from the internet. Well-known are also discounts for the purchase of larger quantities of a given merchandise, or because the seller wants to get rid of less popular items, or so-called 'fire sales'. Add to it discounts for special customers. Discount coupons usually are limited to special products, on special days, or to be redeemed during a limited time only. Some discounts are 'open-ended' e.g. "$5.00 back for a purchase of at least $15.00" these can be proportionally subtracted from the money paid in proportion to the list price of each product that is included in such a transaction. As most of those products belong to the same price aggregate it is indifferent how an open-ended discount is distributed.

    Also see Susanne Wied-Nebbeling ". . .Da die meisten der Befragten Firmen ihre Kaufpreise differenzierten und/oder Rabatte gewährten, wichen die tatsächlichen Verkaufspreise häufig von den kalkulierten Preisen ab" (As most of the business firms that were interviewed , differentiated their selling prices, or granted discounts, departed the actual transaction sales prices from the prices that had been calculated) p. 4 *Das Preisverhalten in der Industrie* J.C.B.Mohr, Tübingen 1985.

    In a reverse of that situation, actually paid prices can be higher than the price tag or advertised price. A colleague reported about his trip to a scientific conference. He purchased a flight return ticket with the discounter *Travelocity* on the internet for $186, but in the end his credit card was charged with $560 because of airport and other added taxes that were not mentioned in the initial offer. The "pure price" usually differs from the actually paid price, even if seldom to such an extreme extent.
15. The *Washington Pos*t of May 31, 2004 reports, in its Business Section pages E1 and E 9, under the title "Wal-Mart Triggers Tumult in Toyland" the case of drastic price differences for

toys. The specialty store "Tree Top Toys, Inc," an independent retail chain store with outlets in Washington, D.C. and the suburb McLean, in Virginia, reports that the toy "Chunky Farm" is very popular with brisk sales at $32 each. Eighteen months later that phenomenal demand ended, although not the supply of that toy, at the old price, while in the nearby Wal-Mart – the most powerful retail discount store in the USA with hundreds of stores throughout the country – sells that same toy "Chunky Farm" for $14. The producer of "Chunky Farm," "Shelcore Inc" reports that he has no influence on the price at which it sells in the two competing stores that sell his product in retail trade.

Another toy, "Leap Pad Learning System" sold for $54.95 at 'Tree Top Toys, Inc' and was sold at the same time in the nearby Wal-Mart store for $33.24.

16. "The majority of business transactions are simple, routine, often automated, like the home delivery of a newspaper, or the purchase of a cup of coffee from a vending machine. Yet, even in the most routine or mechanized transaction has the same basic elements: the contracting partners, an agreement as to the product or service, its quantity and quality, the form and timing of payment, as well as place and time. In the case of the coffee vending machine it is tacitly taken for granted that the coin or credit card operating machine is functioning properly, that a non-leaking cup will be supplied that it will be filled with hot coffee of a certain quality, with sugar and milk optionally available. If the buyer disagrees with the conditions, he does not have to use that machine; he is free not to carry out this purchase. In all these instances 'Price' is a characteristic of the transaction concerning this product, not of the merchandise or service itself that was exchanged; Price, therefore, must not be considered as removed from human influence – which marketing experts and advertisers have long recognized. Prices do not just happen as a result of impersonal, anonymous, sweeping market forces of demand and supply. The 'scientific' outlook of econometrics would have us believe this. Instead, every price is the result of an agreement, often arrived at after hard-fought battles and final compromises between individual buyers and sellers. It is one of our modern-day myths that prices exist out there' as objective, impersonal facts, removed from human influence. Prices, to the contrary, are determined – raised or lowered – by people's' decisions. Price, that complex object, can be obtained from a written document, electronically (bar code readings) or just verbally from one of the participants in the transaction. Every transaction is characterized by the type of products, and its quantity and quality, the form of payment – cash or some form of bank credit – in addition to the form of packaging and service e.g. insurance or delivery to buyer's location. In addition, every transaction is also characterized by the kind of buyer e.g. a hospital or individual household, and the seller, e.g. a chain store. Usually the simplest form of transaction is taken as the prototype, which suggested the identification of price with the merchandise. But the socio-economic character of price becomes clear when one considers the most complicated transaction as the model, such as the purchase of one firm by another in a business merger. The numerous legal conditions and notarized agreements regarding the shareholders, the time frame and localities such a transaction is to be carried out make it evident that the agreed upon price belongs to that transaction more properly than to the object that was sold." Winkler, O. (1987) "A New Approach to the Measurement of Prices" *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, Washington, DC. pp. 608–613, esp. p. 609
17. "The individual price is an event that is not nearly as much governed by the 'economic laws of supply and demand' as it is by the judgement and negotiating prowess of the parties involved" .Winkler, O. (1991) "Unlocking the Secrets of Price Aggregation – Recording Prices without Weights" *Proceedings of the Business and Economic Statistics Section of ASA*, Alexandria, VA. USA.
18. The price of the textbook "International Marketing" 7th. edition by Michael Czinkota and Ilkka Ronkkainen, Thomson South-Western Publishing, London, Cincinnati, 2005 is priced at US $ 156.25 in the bookstore of Georgetown University, and was sold to the students, with a local sales tax of 7.5%, for US $167.97. That same textbook was sold in London for the

equivalent of US $70.00 Some of the students of that course imported that textbook from England for the additional cost for shipment of US $5.00.

Professor Czinkota. reported another personal experience with L.L.Bean, one of the largest mail order houses in the USA. He attempted to purchase a pair of shoes from their newest catalogue. He inquired by telephone, asking for various details regarding the leather, the shoe sole, the last, etc. The price was listed as US $160 At the end of his questioning the sales lady tried to close the deal. When Dr. Czinkota declined, she inquired for his reasons, why he did not wish to make that purchase. He responded: "too expensive" .to which she reacted with a counter question "What is your price limit?" whereupon he responded "$100". At that point she made a counter offer "Then $99.95 is that OK?" He accepted and bought that pair of shoes for US $ 99.95 instead of $160. Not every vendor has that margin and flexibility in price deals, nor does every customer have experience in negotiating. Yet this is a typical instant of the discrepancy between list prices and real transaction prices as they happen in the market place.

As a further example, why a statistic of prices that relies on list prices or on price-tags, misrepresents the market situation is given by Lent und Dorfman who report on the discrepancy between the price measure for air travel in the official CPI (consumer price index) that relies on the published list prices for airline tickets, and their experimental price index of actually paid prices for those same flight tickets. The CPI price index of the list prices for airline tickets, between Q4 1998 and Q2 2003 indicates a rise in airline ticket prices of 15,4 % compared to a rise in the actually paid prices for airline tickets for the same flight routes, during that same time span of only 6.6% (Table 2, p. 23 op. cit. "This difference is probably due mainly to (1) the different target formulas used (Fisher or Jevons /Modified Laspeyres) and (2) the ATPI's (Actual Airline Ticket Price Index) inclusion of special discount fares that involve differential pricing, combined with consumers' increasing use of special discount tickets during the period. (e.g. frequent-flier awards and internet specials). . ." p. 23, Janice Lent, and Alan H. Dorfman, "A transaction Price Index for Air Travel", *Monthly Labor Review* June 2005 pp. 16–31.

19. To measure the development of prices, sales and other taxes on transactions, usually reported separately from price, are to be included in a price -index. These add-ons do affect the purchasing decisions, and as co-determinants of prices must be included in the measurement of prices, to get a realistic picture of the situation. It is indifferent to the buyer which part of his price payment will go to the private or the public sector.

    "In addition to unit valuation. . . pricing issues include treatment of taxes and comparability between private-sector scanner data and census Bureau/BLS data. The CPI collects prices without sales taxes; then a calculated tax is applied separately using secondary data. Scanner data also do not include taxes. . .since A.C. Nielsen does not disclose the exact location of outlets, it is not always clear what tax rate should be added to item prices . . .One possibility would be to calculate a population-weighted average sales tax each month for each item based on the outlet usage patterns of consumers in each geostrata. . ." p 270
    Charles L. Schultze, "*At What Price?" Conceptualizing and Measuring Cost-of-Living and Price Indexes*, National Academy Press, Washington, D.C. 2002
20. Winkler, O. (2005) "Aggregates and Aggregation in Economic- and Social Statistics" 55th Session of ISI, Sydney, Australia, April 2005, *record #1159 on CD*.
21. In addition, the weighting of the price-per-unit price-tags with fixed quantities, has the following implication that has been ignored. Assuming that the quantities of the products contained in the shopping basket of the price-index could effectively be frozen at their level by government decree, the prices of these products would be different from those actually observed and used in the price index, prices that happened without such government interference. The fixed quantities of a base – or any other – period in the index create the illusion of having achieved something analogous to an experiment in the natural sciences, of actually having "been able to hold constant" the important factor q of price formation. The illusion arises from the subliminally implied suggestion, that the market situation, reflected in the price-index-numbers, was the result of those fixed quantities. Without a doubt the prices of the items in the current

price-index-number would be different if those base-period quantities q really could have been imposed on the markets.

22. Non-matched random samples of the prices of actually purchased airline tickets in successive time periods are used by the US Department of Transportation as a valid method. of price research. – see Lent, Dorfman *MLR*, op.cit.
23. Lowe Robin, (1998) Televisions: Quality Changes and Scanner data" *Proceedings of the Fourth Meeting of the International Working Group on Price Indices,* US Department of Labor, Washington, DC. "Calculations based On Scanner Data. A large retailer, with many stores across Canada, provided the data used for these calculations. . ..The price is the actual transaction price before taxes. . . (p. 10) . . .the most noticeable result is that all (scanner based) indexes show a substantial decline while the orthodox (CPI = Consumer Price Index) indexes rose slightly . . . How do we account for these differences? 1. Items. . . listed as "all other", included in this (scanner) database, are not included in the CPI indexes. They have fallen most. . . .the indexes for 20 inch and 27 inch (TV sets) separately are also significantly lower than in the CPI regular survey. . ..2. There was a shift in 1997 towards the higher priced products. . .. The monthly-chained (scanner data based) index reflects the substitution between representative commodities that the CPI index does not. (p. 11) the reason for this has to do with how representative the scanner sample is (p. 12) . . .Conclusions: . . .The differences between these indexes and ones obtained from scanner data are striking and difficult to explain. (p. 15) . . .the few comparisons that have been made using prices of constant quality suggest that the range of impact between doing quality adjustments well or badly is still small compared to the difference between the scanner data and the regular (official) survey results (for the CPI). (p. 16)"
24. Feenstra, Robert C. and Shapiro, Matthew D. (2003) *Scanner Data and Price Indexes" Studies in Income and wealth*, Vol. 6 NBER The University of Chicago Press

    "When data are based on actual transactions, as opposed to survey samples of price quotations, revenues and expenditures, there is the potential for measurement to closely reflect the underlying variable being measured. Basing data on actual transactions also creates opportunities for modeling the behavior underlying the transactions. . . . Scanner data are electronic records of transactions that establishments collect as part of the operation of their businesses. The most familiar and now ubiquitous form of scanner data is the scanning of bar codes at checkout lines of retail stores. . ..(p1). . .Increased computerization . . .provide enormous scope for measuring consumer activity transaction by transaction. . ..Scanner data provide a census of all transactions rather than a statistical sample. Scanner data are collected continuously. . .Scanner data provide simultaneous and consistent observations on both price and quantity. Electronic transmission of scanner data from point of collection to point of analysis can provide for substantial increases in the timeliness and accuracy of observations. They allow conceptual as well as functional changes in price measurement. . .. The most obvious (benefit) . . .is the elimination of sampling error inherent in estimating the average price based on a relatively small sample of prices for an item (p. 2) . . .scanner data can reduce the need to measure prices for a limited number of items at a limited number of outlets . . ..The slow incorporation of changes in goods and shopping patterns is an important source of overstatement of inflation by the Consumer Price Index (p. 3)"

    Neubauer, Werner von, (1998) "Preisindex versus Lebenshaltungskostenindex: Substitionseffekte und ihre Messung" *Jahrbücher für Nationalökonomie und Statististik*, Vol. 217/1, p. 51 (Translated):: Considering that the conventional theoretical model of the ‚Cost-of -Living-Index' is not known to stand out as being operational, and if one adds the serious departures of the world of theoretical models from reality, there can be only one conclusion: The fiction of a true ‚Cost-of-Living-Index' and its approximation by a „Superlative Price-Index" is untenable and misleading. A 'Price-Index' with a constant , fixed 'market-basket' that is updated in reasonable intervals is in the end yet a more reliable tool. . ..whether one aims at a 'Price-Index' or a 'Cost-of-living-Index' the phenomenon of substitution, as a consequence of changes in the structure of prices, is a reality .. . ."

(Original text) "Bedenkt man, dass schon das überkommene theoretische Modell des Lebenshaltungskostenindex sich nicht durch Operationalität auszeichnet und nimmt man die gravierenden Abweichungen der Wirklichkeit von der Modellwelt hinzu, so kann es nur eine Konsequenz geben: Die Fiktion des "wahren Lebenshaltungskosteneindex" und seiner Approximation durch 'superlative indizes' ist unhaltbar und irreführend. Ein Preisindex mit konstantem 'Warenkorb', der in angemessenen. . . Abständen aktualisiert wird, ist am Ende doch ein verlässlicheres Instrument. . . .ob man nun einen 'Preisindex' oder einen. . . 'Lebenshaltungskostenindex' anstrebt, das Phänomen der Substitutionen im Gefolge von Preisstrukturänderungen existiert. . ."

25. "Scanner Data have a number of potential advantages over price measurements based on survey sampling. Scanner data include the universe of products sold, whereas sampling techniques capture only a small fraction of the population. Scanner data are available at very high frequency, whereas the cost of survey sampling typically limits data to monthly or lower frequency. Finally, scanner data provide simultaneous information on quantities sold in addition to prices, whereas survey techniques typically collect separate data on price and quantity . . . at different frequencies and for different samples . . . (p. 123) ". . .it has been quite common to use the high-frequency variation in prices and sales available from scanner data. In the marketing literature it is well recognized that a great deal of substitution occurs across weeks in response to changes in prices and advertising. . ..store-level data for tuna and toilet tissue contain a dip in sales in the weeks following a promotion. . .there is also a high substitution between different varieties of tuna, depending on whether they are on sale or not. . ..it would be highly desirable to construct weekly price indexes in a way that takes this behavior into account. . ..to construct "true" or "exact" price indexes, we need to have a well-specified model of consumer demand, which includes the response to sales and promotions' (p. 124)
Robert C. Feenstra and Matthew D. Shapiro, "High-Frequency Substitution and the Measurement of Price Indexes" *Studies in Income and Wealth*, Vol. 64, National Bureau of Economic Research, published in Ch. 5 of *Scanner Data and Price Indexes,* The University of Chicago Press, 2004

'. . .Scanner price data may replace or reduce the need to visit stores to price items. . .could generate a more representative selection of items for pricing. Scanner data include the universe of products sold . . .. whereas the current quota sampling method only records prices for a small fraction of items on store shelves. Scanner quotes are available if the item has been sold during the pricing periods. . .the BLS collects prices for selected items whether or not they have been sold at the . . .identified outlets . . . since scanner data can include the universe of transacted prices at covered outlets, samples are refreshed continuously and new items appear in the data much more frequently. . .scanner record actual transaction prices, not shelf prices at which transaction may or may not have taken place for the relevant period. . ." p. 267 ". . .scanner data could expand geographic coverage . . . data from non-metropolitan area outlets are also available. In contrast, the CPI uses data from only 87 metropolitan areas." p. 268
Schultze, Charles L. op. cit.

26. This PLI appears structurally to be like the index number formula by Drobisch, and a value index but is based on a different concept of price, and can be used also for the transaction prices of products that are not directly comparable, for any level of aggregation. That is also true of the early efforts by Dutot who had the right intuition but not the right concept of price. Hoffmann und Kurz made the following remark about the Dutot Index:

". . .The Distance between the Dutot indices and the geometric means increased as well for the chained attached samples. . . but remained within reasonable limits. . .Even with the differences being relatively small, it is remarkable that without exception the Dutot index gives lower estimates of price changes than the Jevons index. This indicates that the rate of price increase is lower for expensive dwellings than for cheap dwellings. . . **the Dutot index can be described as a weighted arithmetical mean of price changes**, with the weights being proportional to the relative level of rents. . . **the Dutot index corresponds to an exact Laspeyres-index**. . ." p. 29, 31

Hoffmann, Johannes, Kurz, Claudia (2002) "*Rent Indices for housing in West Germany 1985 to 1998*" discussion paper x/02 Economic research Centre of the Deutsche Bundesbank.

"Scanner Data allow transaction prices to be averaged over the relevant period. . . scanner data are typically produced using aggregated unit values- a quantity weighted average price of an item. The simplest version is calculated as sales revenue divided by number of units sold. . .. the main criticism of unit pricing is that it produces a price at which no single item may actually have been sold . . . (Footnote 10: This is the case because stores sell the same item a t different prices, which then are averaged. Unit values may be the average of prices over a time period, across some set of outlets. . . or even across product codes that have only minor differences in characteristics. Multi store unit pricing implicitly accepts the assumption that consumers switch easily between outlets in response to price changes. . .) . . ." the unit value index more accurately p. 268 reflects the preferences of the shopper who searches out the lowest prices each week and also the consumer who stockpiles during a particularly good special but then purchases nothing until the next special. . ..in some instances, few consumers purchase at the shelf price that the BLS agent happens to observe. How many people buy Chicken-of-the-sea tuna fish when the 'Bumblebee' next to it is on sale for half price? . . . there is substantial consumer substitution across weeks in response to price changes and advertising. . . Using shelf prices assumes rigidity in consumer shopping behavior, since items in each week of pricing are treated independently and that elasticity of substitution among them is zero. . . at some level, price averaging must take place to construct any price index. . ." p. 269 Schultze, Charles F. op. cit. . . .

27. ". . .purchasing power parities (ppp's) are . . .interspatial price indexes (by analogy with the inter-temporal price indexes such as consumer price indexes) . . .the methodology and theory underlying their calculation are identical to those of more familiar index numbers. . .there are differences of emphasis, however, between inter-temporal and interspatial price indexes. An important difference is the choice of the goods and services making up the basket. . ..It is more difficult to choose a basket of goods and services equally representative and characteristic in two or more countries. Even in neighboring countries . . . one may encounter different preferences for a variety of reasons (tastes, climate, size and type of packaging, etc.) . . .although ppp's . . .can be calculated. . .with consumer price index theory, the usual coverage of ppp's is that of the goods and services which make up gross domestic product. . .. for a product to be included in the list it must be available in at least two of the countries concerned . . .the list must be representative of the expenditure category (basic heading) and characteristic of at least one country. (p. 8) Dryden, John, Reut Katrina, Slater Barbara (1987) *Monthly Labor Review*, December, US Bureau of Labor Statistics, Washington, D.C. The proposed PLI is much less limited for such regional comparisons than the method for regional comparison described in this article.
28. A special position takes rents of apartments and the costs of owning an apartment or a house. Monthly rents are to be included in the price level measurements as the transaction prices of services. In the ownership of apartments and houses all current costs, that have not already been registered as purchases in stores, such as e.g. the materials for the addition of a wooden patio, are to be included in the 'transaction price of services." To be added, however, in these "prices of service" are the costs of maintenance, cleaning and repairs in the house/apartment and the garden that belongs to it. Open for discussion remains the issue of whether the costs of new construction as additions to the existing structures, should be included.
29. The two versions of the PLI yield complementary but different information about changes in price levels. The PLI_T, "price level change per transaction" informs about changes in the amount of money spent, on average, in each transaction. When buyers increase their spending from one dozen eggs, on average, in the month before, this will appear as a doubling of the amount spent in the Sum $\sum \mathbf{p_{i,t}}$ of the PLI_T "price pro transaction". The PLI_U "price level change-per-unit of merchandise sold," will not indicate a change in the average price-per-unit"

An increase in the PLI_T can have a variety of causes. e.g. customers buy a larger amount of units of the same merchandise at the same price per unit. It can also be because customers

buy the same amount of units but at higher prices, which could be because the prices rose, or customers switch to buy in stores that sell the same item at that time for a higher price-per-unit. It could also be because customers prefer to buy other, more expensive goods. In each case more money is spent, on average, per unit for different reasons but with the same end-effect. It could even be the result of a change in the income situation of the customers.

Another issue that may be objected is the issue of adding vastly different products into these price aggregates. How can e.g. the prices of eggs be added together with the prices of cars? These prices of individual transactions of products are not added directly, but are added to transaction prices of similar transactions of goods in the same category. Then, in a subsequent step, these different transaction groups are aggregated to larger totals in which the relative importance of each product group is weighted automatically with the amount of money spent for each group. (For more detail concerning proper aggregation see Figure 7.2.3, in the next chapter on measuring quantities) Because each of these groups of products will be purchased with regularity, only changes in purchasing habits will cause the aggregate to indicate a change e.g. purchases of a larger number of expensive items, or purchases of the same number of costly items but at higher or lower prices. The regular and systematic aggregation of these price-groups of different products and services results in a balanced over-all picture of the real price situation.

Both versions of the PLI are transparent and easy to understand and are calculated with the same raw data input. Neither version of the PLI corresponds to the fiction of a "pure price change," or of a "cost of living indicators," but simply record that which is going on "out there in the economy."

30. Personal information by John Astin, Organizer and Discussant of Session No 52 "Use of Hedonic Methods for Quality-adjusted Prices", 54th Session of ISI, Berlin 2003
31. This chapter treats the measurement of prices and price levels from a realistic, statistical survey-oriented approach, not the usual economic-theory-approach to prices as the 'Index-Number-Problem'. Problems of reporting the prices of services of all kinds, particularly health, housing and transportation services, not covered here, deserve to be treated separately.
32. The following sections draw heavily on thoughts published in Winkler, Othmar, "A New Approach to 'Measuring' Agricultural Production" *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1962, pp. 30–36, and Winkler, O.W. "The Actual Information Content in Statistical Productivity Measurements", presented at the 45th Congress of the *International Statistical Institute, Book 2, Contributed Papers*, pp. 391–392, ISI, Amsterdam 1985.
33. "A series of articles describing "National and International Indexnumbers of Agricultural Production." as well as discussions with persons who compute or use them have raised doubts about the necessity for and the value of index numbers of production". John Black, Bruce D. Mudgett, "Research in Agricultural Indexnumbers – Scope and Methods", *Social Science Research Council,* Bull No. 10, New York, March 1978, pp. 97–100; M.I. Klayman, "Note on National Indice of Agricultural Production", *Monthly Bulletin of Agricultural Economics and Statistics*, IV-2, FAO, Rome, Jan. 1955, pp. 4–7; M.I. Klayman, "Indices" M.I. Klayman, "Indices Internacionales de Producción Alimenticia y Agrícola", *Boletín Mensual de Economia y Estadística Agrícolas*, II-1, FAO, Rome, March 1953, pp. 1–6, P.V.Sukhatme, P.L.Sherman, "Indices Internacionales de Producción Agrícola", *Boletín Mensual de Economia y Estadística Agricoles,* V-3, FAO, Rome, March 1956, pp. 1–4. For further References see note '6', p. 53 and 54.

Utility considerations are conspicuously absent from the literature. Obviously the consumer preferences and utilities of any one individual are out of place: the same agricultural or industrial products may be purchased on the same market by the purchasing agent of a factory (as raw material), by wholesalers and retailers, but also by ultimate consumers. Their utilities, if they exist at all, are hardly comparable: Donald Davidson, Jacob Marschak, "Experimental Tests of a Stochastic Decision Theory", pp. 233–269 and S.S.Stevens, "Measurement, Psychophysics, and Utility", pp. 18–63, both in: *Measurement: Definition and Theory*, ed. W. Churchman, P. Ratoosh, John Wiley & Sons, New York, 1959.

34. Dudley J. Cowden, R.W. Pfouts, "Index-numbers and Weighted Means", *The Southern Economic Journal,* XIX-1, July 1952, p. 90: ". . .Statistical Price and Quantity Index-numbers are essentially unrealistic, . . .problems involving construction or use of index-numbers are always difficult and must be handled with apprehensive care. . .".
35. The simple looking formula of Laspeyres, for example, becomes a monstrosity if all the changes that actually are incorporated in the course of time, in such an index-number, were expressed in symbolic notation. The complications which arise have been traced in some detail in Winkler, O. "Unrecognized Possibilities for Simplifying Production Index Numbers" *Proceedings of the Business and Economics Statistics Section, ASA*, Washington, D.C. 1963, pp. 2–8.
36. ". . .Index-numbers are devices for measuring differences in the magnitude of a group of related variables. These differences may have to do with . . .the physical quantity of goods produced. . .", Frederick E. Croxton, Dudley J. Cowden, *Applied General Statistics,* 2nd ed. 5th printing Prentice Hall, Englewood Cliffs, N.J., 1960.
37. In the case of agriculture, production records are incomplete or missing on smaller farms, the cooperation by farmers is uneven, and the representativeness of the sample often not adequate. These production figures are among the least accurate economic data. How soon and to what extent these socio-technical problems can be overcome is anybody's guess. See especially: Morgenstern, Oskar, op. cit
38. Approximately 450 other pertinent publications which were used for this study are listed in: Winkler, Othmar W. Bibliografia – Estadística de Producción, Precios y Trabajo, *Centro Interamericano de Enseñanza Estadística Económica y Financiera,* Santiago, Chile 1959, CIEF 2523, 'pp. 17, 18, 25, 28–33, 48–54, 69, 70, 75, 76.
39. The Agricultural Estimation and Reporting Services of the US Department of Agriculture, *Miscellaneous Publication No. 703*, USDA, Washington D.C., Dec. 1949
    Agricultural Production and Efficiency, *Major Statistical Series of the US Dept. of Agriculture,* Vol. 2, Agricultural *Handbook No. 118*, Sept. 1957
    Hernán Montoya, "Metodologia en Estadísticas Agrícolas de las Américas", *Estadística, Journal of the IASI* Vol.IV-13, 14 and 15, March, June and Sept. 1946, Washington D.C.
40. Compare with this the conventional concept as presented in: Felix Rosenfeld, *Course of Lectures in Agricultural Statistics,* FAO, Rome 1957, pp. 47, 123, 124.
41. In my paper "A New Approach to 'Measuring' Agricultural Production" I had to distinguish four, instead of one net production concept for the special case of 'Agriculture.'

    C4 (corresponding to the concept C3 in chapter 7.2 of this book) is given in physical measures of quantity and in sales values, and is available on the farm for each product separately. The monetary values are part of the economic gross product of the farm.

    C5 (part of the concept C4 in chapter 7.2) refers to the total sales of the farm, and is given in monetary values only. It includes all gross monetary incomes from agricultural activities at the current actually paid prices. It is a total, comprising the monetary values of C4, for each of the more important crops for which separate figures were gathered, the money income from sales of minor crops for which no such separate figures were collected, sales of agricultural sub-products such as straw, but also of the sales of animals and products of non-agricultural activities on the farm as long as these are not important in the farm total. The figures of C5 correspond to the amount of money really received by agriculture, that is, the money flow into that sector of the economy. Combined with cost information the agricultural "value added" can be estimated.

    C6 (another concept for the concept C4 of chapter 7.2) is another monetary concept of interest. It is a value estimate, preferably at the current market prices, of all agricultural products produced in the given year. The quantities of this chapter's C3 that were consumed or processed in an industrial activity on the farm, bartered with neighbors or put into storage, will be assessed and added to the sales figures of C5. Considering the instability of prices and the size of the unsold production, this estimate may give an unrealistic picture for agriculture as a whole.

See also: Paul Studenski, *The Income of Nations*, New York University Press, New York, 1958, pp. 176–177, 268.

C7 is a more refined version of C4 in chapter 7.2. The figures of this concept, a form of net production, deducting costs from the figures of C5 or C6, can be determined only for the farm as a whole, not for specific crops. The aggregate for all farms gives the "value added" by agriculture, the share in the GNP or NNP depending on the concept of gross production used, and the kind of costs deducted. It is the most important of all the concepts for economists, and really contains itself a group of concepts.

Summarizing, these seven production concepts complement but do not replace one another. The figures of these concepts should not be adjusted for seeds or fodder as is customary now, but taken for what they are. See M.I.Klayman, "Concept of Production", p. 2 and "Summary Table of Principal Features of National Indices of Agricultural Production", pp. 4–7, in "*Note on National Indices of Agricultural Production*", see endnote 1. In some instances the time, place and quantity of the produced goods is the same for all concepts of production. Generally, however, the figures of the different concepts will not coincide in amount and time period of occurrence. Between the successive stages, time elapses, more labor and capital is added, and certain quantities get lost – by shrinkage and waste in the process of cleaning, by insects, rodents, decay, or theft. The figures become 'net' to an increasing extent, first technically, then economically. Meaningful aggregates can be formed only with the figures of the same concepts. Concepts C5, C6 and C7 may be adjusted for change in the value of the monetary unit by an adequate deflator.

42. This problem has not been recognized in the literature. A good understanding in: Bjorn Koch, *Metodologia de la Estadística de la Indústria Minera en las Naciones Americanas,* IASI, Washington D.C., Sept. 1947, Document ISC/252/IASI-P-S.

    Lacking at present a more dependable criterion, at least those economic activities should be excluded from agricultural production which are also found in a country as independent industries and surpass small-scale operations for own use in the farms. Without going here into more detail it should be mentioned that the figures of production concepts C1, C2 and C3 must report the crops before their industrial transformation. Instead of dividing agricultural production into "primary" and "secondary" production as does Heinz Kraus, in "Measuring the Volume of Agricultural Production", *Jl. of Farm Economics,* XXXII-4/1, Nov. 1950, p. 609, in order to compute an index number of agricultural production on the basis of protein values, it should be recognized that animal breeding in all forms is really another 'industry' which uses as inputs the agricultural products on the farms. The figures of C4 should include only the quantities and values of the sales of agricultural products without further processing.
43. E.g. *Anuario Estadístico do Brasil,* IGBE-CNE, Rio de Janeiro, 1954. The crop statistics of the following fruits are given in 1000 units: Abacaxi p. 91, Banana p. 95, Coco do Bahia p. 100 and Oranges p. 103.
44. ". . .The numerosity of collections of objects (number in the layman's sense) constitutes the oldest and one of the most basic scales of measurement. It belongs to the class that I have called the ratio scales. . .", S.S. Stevens, "*Measurement, Phsychophysics, and Utility*", p. 20, note 2. op. cit. in note 33.
45. A clear statement of this position can be found in: A.F.Burns, "The Measurement of the Physical Volume of Production", *The Quaterly Journal of Economics,* XLIV, Feb. 1930, especially pp. 251, 253, 258–260.
46. Another aspect of Laspeyre's and Paasche's index formulae is the lack of comparability over time. Their insistence on rigid, strictly comparable structures in the aggregates paradoxically causes a lack of comparability. It is precisely this insistence on formal comparability which makes it necessary to perform constant underhanded adjustments, sometimes shown in small print as footnotes, for new products that were added, and for those products that are no longer on the market and cannot be priced anymore. The relative stability over time of higher-order aggregates makes weights of any kind – unit prices, hours worked, protein content per unit, etc. – unnecessary. Besides, prices, as discussed earlier, have important problems of their own

which make them undesirable as weights, especially when prices are set by governmental decree. (see e.g. M.I. Klayman, "El Sistema de Ponderación" in: "*Números Indices Internacionales de la Producción Alimentaria y Agrícola*" etc. note #1). Quantity measures are distorted by price (and other) weights, and become value estimates about which more is to be said in the following.

If a characteristic of the units cannot be easily determined, it can be estimated using a conversion-table, e.g. of unit-protein content for each fruit variety, which was established some time ago. In the case e.g. of physical weight the expression

$$\Sigma N_t w_0 \tag{1a}$$

is acceptable as an estimate of

$$\Sigma N_t w_t \tag{1b}$$

only if there is little or no difference between the w(t) and the w(o). The protein conversion tables of 1928, however, will not reflect the protein situation of 2006 and will not give a good estimate for the protein content of the crops of that year. It is also untenable, to rationalize such base-weighted aggregates from the need to isolate the changes in the number of crop-units, eliminating the per- unit changes. The development of the protein-per-unit of the different crops is a separate sub-question, not to be confused with the more general problem of "how does the weight-aspect of the produced quantities compare over time or between regions.

Estimating the weight, volume, proteins, etc. for a given crop really means estimating different aspects of agricultural production. These index numbers – really crop relatives – are given by:

$$Qw_n = \Sigma N_t w_t / \Sigma N_o w_o \tag{2a}$$

or approximated by

$$Qw_o = \Sigma N_t w_o / \Sigma N_o w_o \tag{2b}$$

$$Qvol_n = \Sigma N_t Vol_t / \Sigma N_o Vol_o \tag{3a}$$

or approximated by

$$Qvol_o = \Sigma N_t Vol_o / \Sigma N_o Vol_o \tag{3b}$$

$$Qprot_n = \Sigma N_t prot_t / \Sigma N_o prot_o \tag{4a}$$

or approximated by

$$Qprot_o = \Sigma N_t prot_o / \Sigma N_o prot_o \tag{4b}$$

They would measure complementary quantity aspects of the same phenomenon. It seems reasonable to assume that index numbers of quantities (2), (3) and (4) – and others not shown – will show divergent developments.

The numerator of a price-weighted quantity aggregate of a Laspeyre quantity index really is:

$$\Sigma Q_t P_0 = \Sigma(N_{w,t})P_{w,0} + \Sigma(N_{v,t})P_{v,0} + \Sigma(N_{u,t})P_{u,0}$$

in which P_0 – at the left of the equation – is the base-period price per whatever type of unit (e.g. Kg, Liter, etc.) is applied to 'quantity Q' at time t. On the right side of the equation N_w is the reported **n**umber of **w**eight units of those produced products that are reported in that manner, and $P_{w,0}$ is the **p**rice per unit of **w**eight for each of those products in the base-period, e.g. dollars per Kg., per ton, etc. – N_v stands for those products that are given in measures of volume, e.g. dollars per bushel, and $P_{v,0}$ is the price per **v**olume-unit in the bas-period. – and $N_{u,t}$ is the production figure of those products that are reported in units produced . $P_{u,0}$ is the base-period **p**rice per produced **u**nit of those products that are reported in units, e.g. dollars per umbrella or per thousand oranges. One should keep in mind that the produced quantities

N which correspond to the four production concepts (C1), . . .(C4) in period (t). and discussed in the following section are:

$$(C1)N_t \geq (C2)N_t \geq (C3)N_t \geq (4)N_t.$$

The differences may be due to shrinkage, theft or loss during storage, elimination of shells in agricultural products, etc. . Differences in the corresponding per-unit weights for each type of measure and product category also can happen. Then

$$(C1)w_t \geq (C2)w_t \geq (C3)w_t \geq (C4)w_t$$

Therefore the numerator of a Laspeyres quantity index number would be as follows:

$$(C1)N_t w_{w,t} \geq (C2)N_t w_t \geq (C3)N_t w_t \geq (C4)N_t w_t.$$

These aggregates are formed, freely mixing the different kinds of measurements in which the produced quantities are reported, not to mention the different concepts of production in which they are expressed.

47. To illustrate this point, let us assume a quantity- index, computed from three products which basically remain the same during both periods of observation: a gasoline motor, a type of electric motor and a gas turbine. Assume further, that for each of these the number of units produced, their total weight in tons and their total volume in cubic feet are available.

Product	Measure	Time period 0	Time period 1
Electric motor	Number of units	7,000	9,000
	weight/unit (kg)	175	165
	Motor Capacity (cft)	280	290
	Price/unit $	$ 200	–
Gasoline motor	Number of units	10,000	17,000
	weight/unit (lbs)	260	250
	Motor Capacity (cft)	50	65
	Price/cft	$ 5	–
Gas turbine	Number of units	1,000	1,000
	Weight/unit (tons)	0. 5	0.5
	Volume (cft)	600	620
	Price/ton	$ 1000	–
	Total number of units	18,000	27,000
	Percent increase of number units	**100.0**%	**150.0**%
	Total weight (tons)	0.7998082	0.78467934
	Percent increase of weight	**100.0**%	**98.1%**
	Total volume (cft)	930	975
	Percent increase	**100.0**%	**104.8**%
	Laspeyre's Quantity Index	**100.0**%	**128.6**%

$$\Sigma Q_t P_o / \Sigma Q_0 P_0 = [9,000{*}200 + 65{*}5 + 0.5{*}1,000]/[7,000{*}200 + 50{*}5 + 0.5{*}1,000]$$
$$= [1,800825]/[1,400,750] = 1.2856 \text{ or } 128.56\%$$

Laspeyres' price weighted quantity index number	100.0%	128.56%

If one is interested in a production index, say of the tonnage produced, e.g. for purposes of transportation, the way to proceed is to get the produced tonnage figures for each, add these and compute a time series of tonnage.

Assume then that, as usually happens, the products are reported in different ways: the electric motors in the number of units produced, the gasoline motors in volume of their cylinders (in cft.) and the gas turbines in weight (tons). The proper way to proceed would be to convert

these different measures into their weight, in tons, with the help of conversion factors established by sampling. Instead most often price is used as the glue to hold together the information about these differently expressed products. But the use of unit prices introduces a very different aspect, namely their value on the market, instead of the relative importance of each product according to their weight in the aggregate. A current-priced weight-index-number (Paasche index) would further aggravate the difference between a time series of the weight of produced items and a production-index-number in which the available information is converted into values through price-weights.

48. When the old parson retired his young successor, an efficient administrator, reviewed the duties of the church's employees . During that review the new pastor noticed that the elderly sexton had not been in charge of the parish register. When asked to do so, the elderly man declined explaining that he neither could read or write. At this the sexton's employment was terminated. The dismissed sexton, walking away in a somber mood, tried to light a cigarette, discovers that he has no matches. Looking for a tobacco store but finding none in that neighborhood, the idea struck him, to open such a tobacco store in this street. His wife, the cook of that parish, had also been dismissed. So he decided to open a tobacco store together with his wife, right there, across from a school, he selling newspapers and tobacco wares, his wife selling candy to schoolchildren. Business is brisk from the start. For the first time in his life he is in a position to save money in the nearby bank. Later, on the occasion of one of his regular visits to that bank, the bank director summoned him to his office, proposing an investment portfolio, more lucrative than his savings account. "We watched your growing savings account. We thought you would be much better served with this portfolio of high-yielding shares. All you need to do is to sign here" pointing to the proper place in the prepared document. This successful client hesitated, then declined the offer. The bank director, puzzled, by his successful clients refusal, inquired why this excellent deal did not seem acceptable. Hesitatingly the client came forward with his real reason: "I can't sign my name, Sir, I never learned to read or write!" The bank director, shaking his head in disbelief, and as if speaking to himself, stated: "this man is a genius!" Then, turning to his customer, exclaimed with pathos: Where would you be today if you had learned to read and write! Ah, responded the client I would still be the sexton at St. Andrew's Church and would not have a penny to save – W. Somerset-Maugham, "Three Short Stories," filmed ca. 1955
49. The makers of the FRB (**F**ederal **R**eserve **B**oard, the central bank of the US) index of industrial production seemed well aware of this fact as can be seen from remarks in the 'technical notes on weights,' FRB bulletin Dec. 1953, pp. 1076–1078.
50. The proposed approach to interpreting production data is in line with the clamor for more detailed information, in such a way as to allow to study the dynamics e.g. of regional development. To mention only two out of the field: Philip M. Raup, "Structural Changes in Agriculture and Research Data Needs", *Jl. of Farm Economics,* XLI-5, Dec. 1959, pp. 1480–1491; Theodore W. Schultz, "Reflections on Agricultural Production, Output and Supply", *Jl. of Farm Economics,* XXXVIII-3, Aug. 1956, p748–762.
51. *BLS Handbook of Methods, Bull 2490,* Chapter 10 "Productivity Measures-Business Sector and Major Subsectors, pp. 89–109, US Department of Labor, Bureau of Labor Statistics, Washington DC 1997.

Chapter 8
Cross Sectional Analysis in One Dimension

"Not everything that can be measured, counts.
Not everything that counts can be measured"
A. Einstein

8.1 Frequency Distributions in Perspective*

One cannot interpret the data of a frequency distribution without first paying attention to the fact, that the 'statistical-counting-units' had to be sorted and grouped according to some of their qualitative, non-numeric or categorical, and regional characteristics to create the necessary framework. One should keep in mind, as was discussed in Chap. 2, that these quantitative, measurable characteristics, are only complementary to those more essential, qualitative characteristics,[1] although in a few instances frequency distributions can become important in their own right.[2]

Arranging the 'statistical-counting-units' according to the magnitude of one of their measurable characteristics removes them from their historic and geographic context, and treats them as if they existed simultaneously, as it were, side-by-side. This provides a cross sectional view of the situation against the backdrop of the qualitative characteristic of the aggregates in which these 'statistical-counting-units' are embedded. When more than one quantitative variable is involved, then the 'statistical-counting-units' are arranged in a space of two or more dimensions, leading to simple and multiple regression and correlation analysis, also presenting a static cross-section, a still-picture of a dynamic social or economic situation, discussed in the next chapter.

The values of the selected quantitative characteristic of the 'statistical-counting-units', e.g. the weekly salaries of all construction workers, paid in the last week of May 2007, are irregularly spread out and tend to pile up around certain points on the horizontal scale. This picture of the irregularly spread out, ungrouped data, may not give a good idea of the likely shape of their distribution. By grouping these dispersed values into classes along the horizontal scale, forming sub-aggregates as it were, the individual detail of each measurement is sacrificed to yield a simplified distribution that shows the general contour more clearly, the well-known frequency

*This section follows: Winkler, Othmar, W. "Forming Classes in Frequency Distributions of Data in Business and Economics" *Proceedings of the Business and Economic Statistics Section, ASA* Washington, D.C. 1983, pp. 487–492.

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_8,

distribution. Everything about aggregates that was discussed in Chap. 3 also applies to the data grouped in the classes of a frequency distribution.

The usual graphic presentation of the frequencies of a distribution are bar graphs (histograms) and line graphs. The line graph of cumulative frequencies is the 'ogive'. Although bar graphs are well known, few users are aware that the class frequency is represented by the area of the bar, not the height, and indicates the density of the measurements along the horizontal axis. . If all classes in a frequency distribution are of the same width, then the height of each bar is proportional to the area and can be interpreted as if it were the class frequency (more in Sect. 8.5). If the class intervals are not equal, as is typical in socio-economic frequency distributions, the height of each bar of the distribution must be adjusted by the size of the interval of each class. Such an adjustment for differences in class widths is needed in bar as well as in line graphs for the appropriate interpretation of the distribution. In line graphs the adjusted height of each class is marked over the midpoint of the interval. In a cumulative frequency distribution, (ogive) the frequencies, accumulated up to and including the frequency of that class, are not adjusted and plotted at the end of each class interval, and are connected by straight lines. These classes, through the choice of their width and their placement along the horizontal scale, are meant to allow an overview of the big picture of the distribution of a particular quantitative characteristic. When one has access to the 'raw' ungrouped data, the person or the computer program in charge of that task should take the specifics of the ungrouped data into consideration – accumulation of data on certain values, and stretches on the horizontal scale with no data at all – to minimize distortion when establishing frequency classes.

Statistical methods dealing with quantitative variables occupy a position in textbooks of business-, economic- and social statistics, that is out of proportion to their limited importance for interpreting the underlying socio-economic phenomena reflected in the data. The numeric values of these quantitative variables resemble measurements in the sciences, invite mathematical manipulation, and have popularized the impression, that socio-economic statistics is a subset of mathematical statistics. Yet the real contribution of quantitative variables to the understanding of a socio-economic situation is only complementary to that of the more fundamental but less tractable qualitative, (categorical) characteristics, as discussed in Chap. 2.[3]

8.2 Linking Cross-Sectional and Longitudinal Analysis

Imagine the weekly paychecks of the workers of a factory. They are the 'real-life-object' from which a 'statistical-counting-unit', their weekly wage, is recorded. As discussed in Chap. 2 these workers as the 'real-life-objects' continued to exist in that factory, but their wages may change from one week to the next. Each of these workers has a history of earlier payments, so to speak, an extension in time that cannot be seen in the distribution of the wages of this particular week. In the weeks before, each wage could have been higher, or lower than the one that is captured in

this one week's frequency distribution, taken in one specific instant of the continuing existence of each worker as the 'real-life-objects'. The value of each 'statistical-counting-unit' at that moment is fixed, as if frozen in time.

Or suppose that in a study of business firms, the number of their employees is one of their quantitative characteristics. One of these firms employed 2,000 workers in June. A month earlier it had 2,100 and in the following month of July it employed only 1,800 workers. Other firms also experience changes in their employment situation. A frequency distribution of these firms by number of employees in June would capture those business firms in a specific moment of their dynamic development over time. It is important, then, that something be known about the usual magnitude of fluctuations in that quantitative characteristic of these 'real-life-objects' and their 'statistical-counting-units' in that specific frequency distribution. This information should be brought to bear on the analysis by forming classes that are cognizant of that kind of fluctuation of each value – the quantitative characteristic 'number of employees' of each business firm – that is to be plotted in a frequency graph.

The analyst of cross-sectional data does not know whether the reported value of the quantitative characteristic 'number of employees' at an earlier point in time was higher and has moved downward to its present value on the scale, or whether it was lower and has moved up to its present position, perhaps continuing its upward trend. These dynamic developments have to be taken into consideration when forming classes so as to preserve and enhance the socio-economic content of the distribution, as well as when assessing the classes of a published frequency distribution.

The number of employees of small firms should be expected to fluctuate less from one point in time to the next, than the number of employees of larger firms. Numerous studies in a variety of fields have confirmed this phenomenon.[4] The weekly wages of cotton mill workers, that I studied also showed this: the wages of the low paid workers tended to fluctuate less from one week to the next than the wages of the higher paid workers. These fluctuations, by the way, were not symmetrical: When a worker received a higher pay than usual in some week, the amount of additional income was small. When in another week that same worker received a lower than his usual pay, that shortfall in income was larger. Figure 8.1 which shows such wages fluctuations. These ups and downs that one might call 'Eigenschwankung' – self-fluctuation of this quantitative characteristic of these workers over time – in a frequency distribution should be taken into consideration when determining appropriate class intervals. These 'statistical-counting-units' that are grouped in a given frequency class at the point in time when they were registered, may be recorded in a different class if that survey of these same workers were taken at another point in time. Because of the uncertainties in their development, they will very likely be replaced by other 'statistical-counting-units' that before were on the high side of the next-lower class, or at the low end of the next-higher class. The width and placement of appropriately calibrated classes not only collates and simplify the scattered distribution of individual 'measurements' into a frequency distribution but in a sense also stabilizes it.

This calls for narrower class intervals along the smaller values on the horizontal scale, and wider class intervals for the larger values, with a gradual transition of

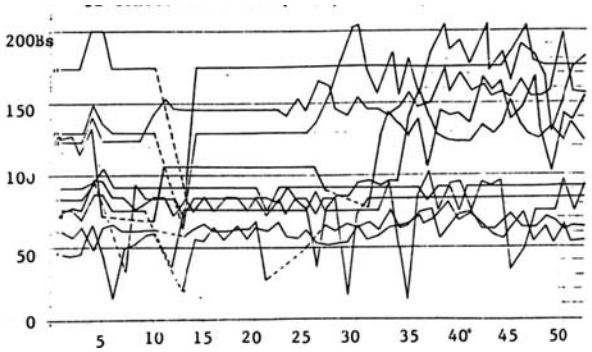

Fig. 8.1 Wage fluctuations of workers' wages. Ten randomly selected cotton mill workers, Caracas, Venezuela

these class intervals from narrowest to widest. This heteroscedasticity in frequency distributions of socio-economic data is the rule, not the exception, even though statistical theory treats it as if it were an occasional aberration.

8.3 'Measurement' in Socio-Economic Statistics

Although it has already been broached in Chaps. 1 and 2, this matter bears repeating here as it affects the formation of classes in frequency distribution. The term 'measurement' has the connotation of scientific, precise and objective determination of the value of a characteristic. The implication, of course, is that it is a quantitative characteristic. But no precision instruments are used in the social sciences to 'measure' quantities. These values are determined and reported by each informant in a statistical survey according to his or her judgment and convenience. Few quantities in socio-economic statistics are determined and reported objectively by qualified outside observers.One exception is the appraisal of the condition of a building by a professional real estate appraiser. In most other instances statistics must rely on what the respondents report in questionnaires, face-to-face or telephone interviews, with an occasional checking-up on the reasonableness or likely reliability of the reported information. The veracity and accuracy of the responses can vary considerably for different topics and from one informant to the next.[5] That includes the omission of some information that was inadvertently overlooked, intentional mis-information, even the outright refusal to cooperate. Information about business corporations is also rendered by an accountant or a contact person, who for much of the requested information, must rely on accounting reports'.[6]

Now consider e.g. the classification of farms by the number of cattle. Farmers with only ten cows have no problem knowing the exact number of their stock, and cannot easily misreport this fact without calling attention to it. The owner of a ranch with 1,000 cattle, on the other hand, may really not know the size of his herd with the same accuracy. His animals may roam freely over large areas, can get lost, stolen,

or killed by predators, and new calves may have been added to the herd. If such a rancher wants to under- or over represent his livestock he could get away with reporting as few as 600 or as many as 1,500 cattle without arousing suspicion.[7] The 'statistical-counting-unit' of that same 'real-life-object' could end up in a different class than the one into which it really belonged because of that inherent uncertainty.

Similar is the accuracy of reported detail. Consider e.g. the statistical report on the value of monthly car sales. A small dealer can report his sales rounding his sales figures by $100 with no material loss of information. For a big dealer, the margin of irrelevancy may be as large as $1,000 or more.

Many self-reported figures come from accounting records that often are based on concepts and interpretations that are still under debate.[8] The resulting uncertainty is akin to the uncertainty one can experience when determining one's federal income tax. Depending on how one views certain expense items and deductions, the reported taxable income for the same person and period can differ quite a bit.

All this is a far cry from the precision and reliability with which the measurements in the natural sciences can be reported. Most of what applies there to measurements and measurement errors is hardly relevant for socio-economic data, whose errors and uncertainties are of a different kind and magnitude. Expressions such as 'measuring the price level,' or 'measuring unemployment' obviously are euphemistic figures of speech. All socio-economic data, including frequency distributions, are closer to historiography, describing quantitative aspects of society, than to the sciences that proceed with ever more sophisticated and precise measuring instruments. But what does this have to do with interpreting frequency distributions?

8.4 Determining Class Intervals

The socio-economic situation cannot be perceived more succinctly than these substantial inaccuracies permit. The range of these inaccuracies increases with the magnitude of the reported 'measures' which has implications for the interpretation of these cross-sectional data.

The class intervals of frequency distributions, therefore, should be as wide or wider, but not narrower, than the likely margin of inaccuracy in the data themselves. That margin should be the guideline when determining appropriate class sizes. As these inaccuracies are not uniform across the range of a variable, using class intervals of equal width for small and large values in a frequency distribution would suppress unequal amounts of detail. This determination of classes in frequency distributions, a form of aggregation, should not indiscriminately hide detail, but to the contrary, should reveal the features that are material, – a term used in accounting for 'essential' – to the analysis of socio-economic situations. Care must be exercised when condensing ungrouped data into classes, to assure that important details are not lost in this step. The situation is akin to the loss of evidence at the scene of a crime through careless police work. No amount of sophisticated forensic arguments later, in a court of law, can make up for that loss of evidence at the crime scene.

The 'raw' data should be grouped in such a manner that important detail is preserved for subsequent interpretation. Although users of published material have no influence on how those data have been grouped, this discussion is meant to provide a perspective on what to watch for when interpreting frequency distributions.

The widths and location of these classes should take into consideration both, the likely margins of uncertainty of these 'measurements', as well as the estimated amplitude of the 'Eigenschwankung' of the 'real-life-objects', using the wider of these two as the criterion for grouping the data. The lack of precision in the 'measurements', and the uncertainty caused by the 'Eigenschwankung' of the 'real-life-objects', increases in both kinds of uncertainty with their magnitude, but need not parallel each other. The grouping of 'raw' data in classes should be accomplished with an estimate of both, the 'measurement error' – really the reporting errors properly speaking – and the 'Eigenschwankung' of the data. Frequency distributions classified in this manner will better reveal the true shape of the phenomenon than frequency distributions with classes that ignore these criteria.[9]

The link between the implicit time aspects and the cross sectional character of the data in frequency distributions has been ignored. Appropriate class-sizes link longitudinal with cross sectional aspects, while reducing detail along the range of values in approximately the same proportion. By contrast, class intervals of same size reduce unevenly the originally available detail, too much at the beginning of he scale, not enough toward its end. Before deciding on the width and location of classes, the plot of the raw data on an unabridged, full-length scale[10] should be studied for existing patterns. The ungrouped 'measurements' can accumulate at certain scale values, while leaving empty stretches in-between (Fig. 8.2, the original data available on request). These empty stretches are as important as the clusters themselves[11] (see Fig. 8.3). When clear patterns emerge, the clusters and empty stretches should guide the establishment of classes, placing the centers of the classes where the clusters are centered, to preserve the pattern. Then, the class limits are to be determined around the chosen class centers. If these centers are awkward numerical values, they should not be changed to more convenient scale values of e.g. multiples of five or ten, because those usually are unrelated to the socio-economic phenomenon in question and would distort the information contained in the data. One should 'let the chips fall where they may' and, if possible, establish class centers at whatever irregular intervals become apparent.

If necessary, one may insert empty 'filler classes' with a frequency of zero. If, on the other hand, only few observations cover a long stretch of the tally sheet, then one single, wide class interval could take care of such an uninteresting stretch.

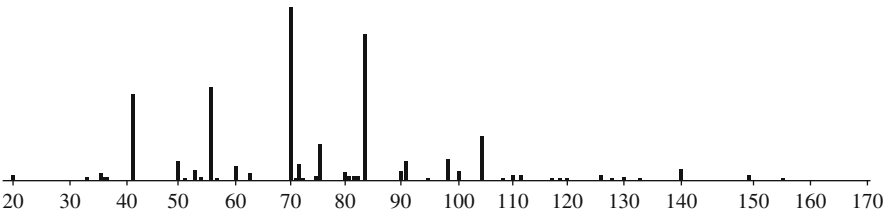

Fig. 8.2 The unabridged Tally of the 482 Wages. Construction workers, Caracas Venezuela

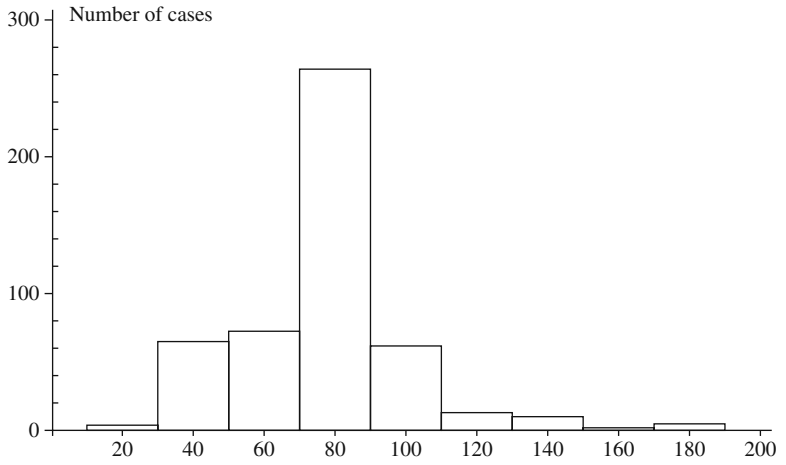

Fig. 8.3 Conventional grouping, wages of 482 construction workers. Caracas, Venezuela – uniform classes of 20 Bs (Bolivares)

Genuine clusters of data on the tally sheet, however, must not be grouped together in one large class together with stretches of little 'activity.' Figure 8.3 shows the consequences of improperly grouped data.

Figure 8.4 shows another careless grouping of the same data, differing only in the location of the class centers. Stretches with an increased density of data should be preserved in appropriately smaller classes. Information that was preserved in a first grouping can be discarded later if deemed appropriate. But the analysis must not start by disregarding, from the outset, a potentially interesting information in the tally sheet (Fig. 8.5).

Sometimes no clear pattern emerges on the tally sheet. If there are grounds to believe that the 'Eigenschwankung' of the counting elements is about the same along the entire range of values, as one might expect in quality control measurements and other engineering and science data, then size and location of the classes

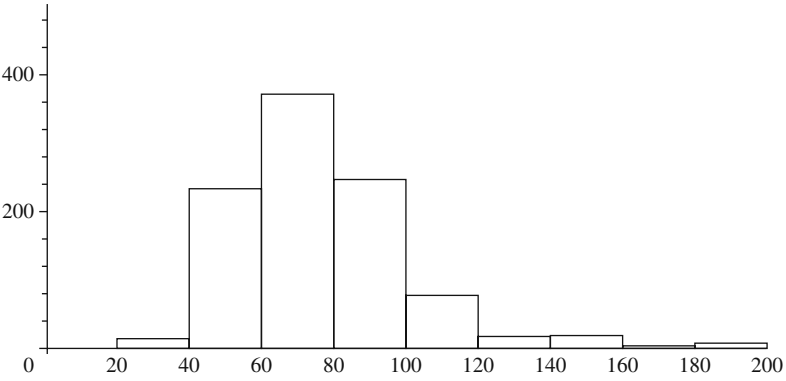

Fig. 8.4 An alternative conventional grouping of 482 wages

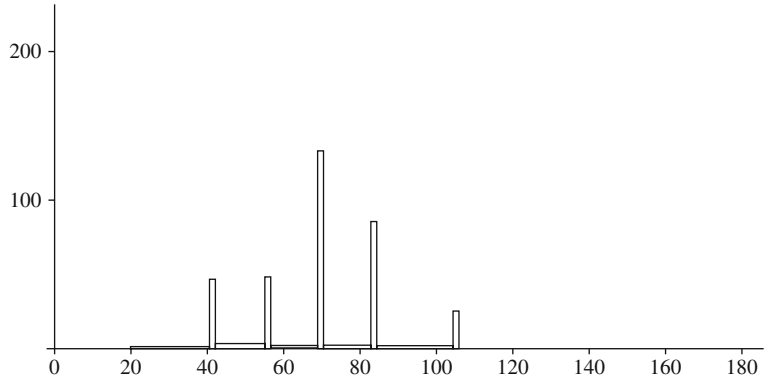

Fig. 8.5 Classes that preserve the shape of the original distribution of the 482 wages of construction workers

may be indifferent and need not be treated in the described manner. Theoretically, of course, the distribution of quality control, other engineering and psychological testing data should cluster around one prominent value in a symmetrical, bell-shaped pattern. In that case, typical in the sciences, it is not important how many classes, how wide these should be, and where they are to be placed on the horizontal scale.

Social science data, more likely than science and engineering data, are a mixture of different social or economic 'species.' A closer look at the data, disaggregating them by one of their more detailed other attributes that may be related to this variable, can reveal a mix of different types of socio-economic 'real-life-objects' in a frequency distribution, that could clarify the absence of a clear pattern in the data.

Suppose that the yearly net incomes of the construction firms of a region are distributed without a clear pattern. In such a case, one should separate the net incomes of those construction firms that specialize in building roads from those that build town houses, or bridges, or high-rise apartments. The tallies and frequency distributions of these subsets, formed by their qualitative characteristic 'Type of construction' usually shows clearer patterns.

Before deciding on a final grouping, the frequency graphs (e.g. bar graphs – histograms) of different arrangements of those same tallies should be explored using different class widths, and different sets of class midpoints.[12] Then **that** arrangement is to be adopted, that comes closest to the pattern of the ungrouped data on the tally sheet. This is a subject-matter oriented approach to frequency distributions, not a mathematical one.

8.5 Interpreting Frequency Distributions of Unequal Class-Intervals

Socio-economic frequency distributions have always been presented with unequal class-intervals. Without exception, the publications of all countries use narrow class intervals for the smaller values, and larger class intervals, that increase at an

accelerated rate, for higher values. The apparent consensus about these empirical progressions is remarkable, characterized by what one may call the 'gearshift phenomenon' – in analogy to the shifting of gears in a car, in response to different demands of the terrain on the motor.[13]

Statistical praxis has followed the right intuition concerning classes of unequal width in frequency distributions, but without theoretical guidance. The class sizes so established progress at an excessive rate. A better balance must be found between theory's lack of understanding for the need of unequal classes, and praxis' untutored, unsupervised zeal that leads to an excessive progression of class intervals.

Converting the 'raw data' of any socio-economic frequency distribution to logarithms transforms such data into a distribution that resembles more a bell-shaped distribution, in hopes to make socio-economic data resemble those in the sciences. Such a logarithmic transformation of the data may satisfy a mathematically oriented statistician, but distorts the reality of the social or economic situation and becomes difficult to interpret. It is insensitive to the uneven spread of the data values, clustering around certain scale values, is indifferent to the varying size of 'measurement errors' and oblivious of the effects of the 'Eigenschwankung' in the data. Mathematical transformations of the data, instead of adding value to the interpretation of the social or economic reality reflected in those data, tend to gloss over or even obliterate that evidence.

Better for the interpretation of quantitative variables than a mathematical transformation of the data, hoping to make them approach a Gaussian distribution, with equal class sizes, are the following two principles.

(1) The **classes should enhance and preserve**, not obliterate, the pattern of distribution in the ungrouped data.
(2) **The loss of detail** in the data, from lowest to highest value, **in all classes** should be approximately **the same.** That can accommodate both, 'measurement errors' of unequal size and 'Eigenschwankung', both getting larger as the scale values advance. In socio-economic data that leads to classes of unequal width, narrower classes on the low end of the scale eliminating about the same relative amount of detail in the data as the classes in the middle and on the high end of the scale. The exact location of those non-overlapping, contiguous classes is not at issue.

These two criteria that should guide the formation of classes can contradict each other in some distributions. In those instances a compromise can lead to wide classes alternating with an occasional class of a narrow width. There are no rules, demanding that the intervals of classes in frequency distributions must be equal, or have to increase at a steady rate. Before deciding on a choice of classes for the distribution of a quantitative variable, a few possible, alternative groupings of the data should be tried. The one which best complies with the two criteria for grouping should then be selected as the final version from among those tentative groupings.

To give a correct view of a graph of unevenly grouped values, **one must adjust the heights of** the frequency **bars** in a histogram or the heights in a line graph for differences in class width. Without such adjustments, the graph will mislead,

overstating the height of the larger classes. In a histogram **the frequency** of a class is **represented by the area** of the bar. As f = h*c (f = frequency of class, h = height of a bar in a histogram or in a line graph, c = class interval) the adjusted height of the bar or of a line graph to be plotted at the class midpoint, is h = f / c. The vertical scale of such a histogram or line graph then marks the 'frequency per unit of measurement on the horizontal scale' of each class. This has not received attention because the discussion in textbooks is limited to fairly symmetrical distributions usually with equal class intervals. As the height in those histograms is proportional to the area, its frequency can be plotted as the height without the need of adjustment.

The standard patterns of increasing class sizes found in official publications of economic and social frequency distributions ignore both criteria for proper class formation, particularly the second, 'relative equal loss of detail in each class along the scale'.[14] Neither the use of uneven class intervals nor the criteria to determine them has been explained or justified. The routinely published frequency distributions of socio-economic facts, using uneven class intervals, preserve too much detail in the first two or three classes, and not enough in the classes past the lower third of the distribution. The desire to give disproportionately more detail at the beginning of the frequency distributions resulted from the need for detailed information about smaller producers, farms, income recipients, etc. while also trying to display in the same distribution the large values, avoiding too many empty, narrow classes. Today's frequency distributions with uneven class intervals mix two schemes that should be presented separately as two parallel frequency distributions with different sets of classes. In either set, the sizes of the class intervals will not advance as rapidly, yet in accord with the uncertainty of measurements and the 'Eigenschwankung' of each data set. The first of these should have classes at the beginning of the distribution that are wider than those presently in use, but their widths should not advance as rapidly in size. This would lead to a larger number of classes, whereby the widest (last) class of this arrangement will be much narrower than the last classes used today in publications. The other set of classes would start with a class width that is much wider than customary in the beginning of the frequency distribution. The subsequent classes are also progressing in width, but less rapidly than those of today. The class interval at the end of the frequency distribution of that second arrangement will be approximately as wide as those last classes in use now (Fig. 8.6). Although the amount of detail suppressed in each class through aggregation will differ in these two groupings, yet the relative amount of loss, from class to class, should be similar in both distributions. In either case, the progression in width from one class to the next ought to be justified by some knowledge of the uncertainties of measurement and the 'Eigenschwankung' of the data in question, and, if possible, be ascertained empirically.

Related is the seemingly inevitable problem of open ended classes. This last undetermined but presumably very wide class does not follow the 'gearshift pattern' of the other classes. The occasional exposure to the raw data of such frequency distributions convinced me that the last, open-ended class really consists of at least two, usually more tan two classes that would increase at the rate used in the rest of the distribution. In some cases the numeric values of the few outliers in the open

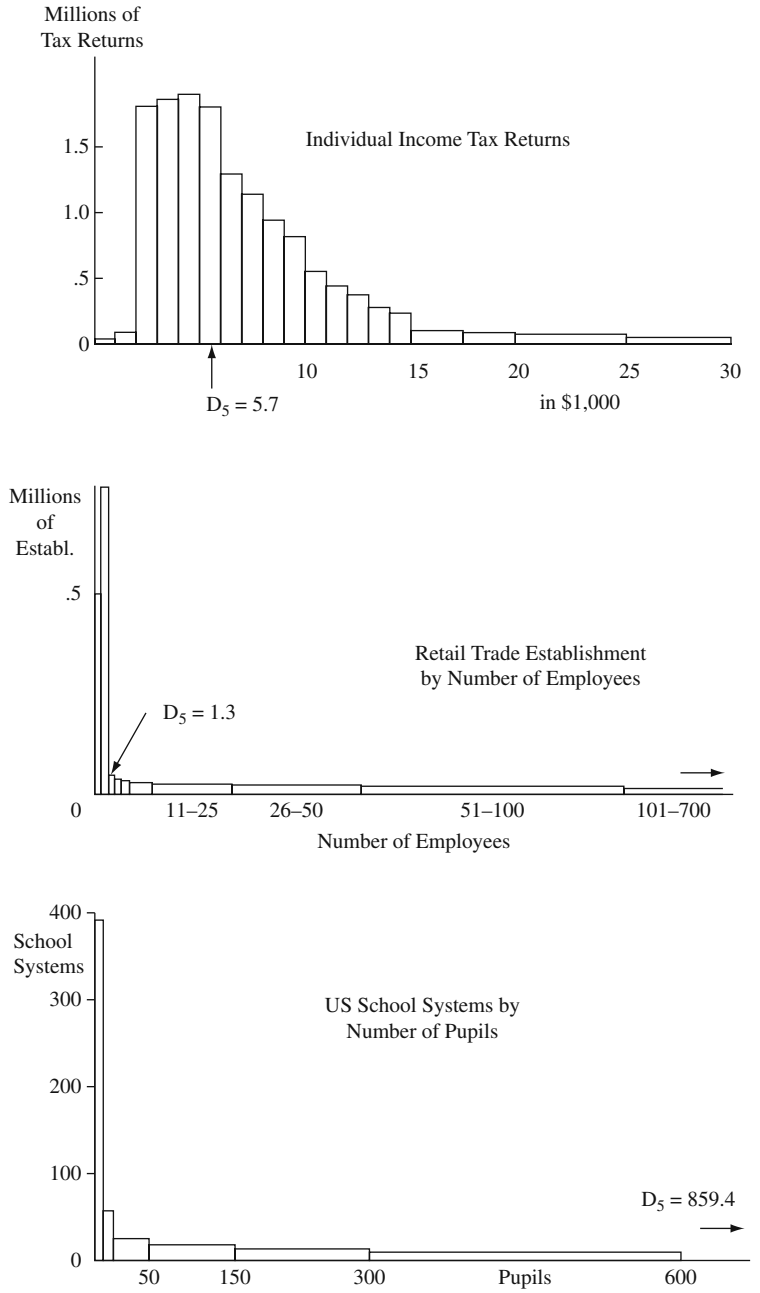

Fig. 8.6 Typical frequency distributions in socio-economic data

ended class are extreme, devaluing the results of computations with the data of such a frequency distribution to guesswork. Instead of leaving the last class open-ended, to shelter the anonymity of the few extreme outliers, some information about them should be conveyed to the reader by listing their approximate values in a source note. If there are too many cases in the open ended class, it should be closed by extrapolating at least two classes according to the same 'gearshift pattern' that was used in that distribution.[15]

Sizes and placement of uneven classes and properly adjusted graphs are important tools to interpret frequency distributions. Even when the user of such data has no influence on the structure of published frequency distributions, this discussion offers conceptual tools that should be helpful to interpret such data. This has not received the attention it deserves because statistical theory has been fixated on frequency distributions that are patterned after the Gaussian distribution, usually referred to as 'the normal distribution' – as if that well-known and readily tabulated distribution with tables of probabilities, de rigeur to be included in every statistics textbook, were the norm or ought to be the norm – is not the type of distribution that prevails in social science data.

8.6 Interpreting the Cumulative Frequency Distribution

Although textbooks show how to construct a cumulative frequency distribution, no attempts are made to interpret it[16]. Yet a cumulative distribution of frequencies reveals interesting features about the socio-economic situation beyond what the frequency distribution reveals.

The cumulative frequency distribution is shaped both by the strength or efforts of each individual case, and by the general social, political and economic environment under which these individual cases have achieved the values that were statistically registered and became 'statistical-counting-units'. The strength of each individual case – the 'real-life-objects' of Chap. 2 – is compared to that of all other cases, and also pitted against the obstacles and resistant forces of the environment in which they function. The up-sloping, S-shape of any cumulative frequency distribution, the ogive, can be imagined as the contour of a sea shore, a rocky cliff, whose profile is shaped like such an ogive. Each individual value in this frequency distribution – and presumably also in those values that have not yet occurred, but are bound to occur later – can be imagined as an ocean wave coming in against that shore. Each wave, washing up against that steep, S-shaped profile of the shore, will break at some point, ending there its advance. Most waves will not reach very high up, ending their run somewhere against the resistance of the steep portion of that shore. Only the few, occasional powerful waves may wash over most of that shore profile, reaching high up before having spent their momentum and their energy, coming to a halt. Once they have reached the flattening portion on the upper end of that coast, however, the resistance to their further advancement decreases, making it easier for them to advance even further.

This image of a coastline, that initially rises steeply, then levels off toward the higher end, invites comparing it to the usual, right-skewed frequency distribution, with the typical long, thin tail-end to the right, with the bulk of the data occurring and gathered at the beginning, lower part of the horizontal scale. Consider an income distribution; the initially steep shore profile of its ogive indicates how difficult it is in that society for its members to advance economically, from one income level to the next higher level in the lower income ranges (on the horizontal scale). Even a big effort does not allow them to advance by much in that steep part of the cumulative distribution. Relatively few individuals, perhaps with the support of their families, or a particularly gifted and driven, perhaps also criminal personality, on the other hand, can come up against this 'income profile' like that occasional powerful wave that seems to flood with apparent ease over the initial, steep part of that coast, overcoming all institutional and societal resistance. Once over that steep portion, the flattening-out shore profile places a decreasing resistance to the oncoming wave that has made it up to there. In terms of the distribution of personal incomes, to stay with that analogue profile of a sea coast, once reaching the upper, flattened portion of the coast, even a minor, additional effort on the part of such a 'wave' can produce an effect equivalent to an added income of a second million dollars, referring to the cumulative distribution of actual – one would hope, honestly reported – personal incomes. In some sense all frequency distributions in their cumulative form, reflect the operation of the societal situation, individual cases, facing a kind of selection process, shaped by the institutional constraints and resistance to advancement of a society (Fig. 8.7).

Rarely are socio-economic frequency distributions skewed to the left, gradually rising at the beginning of the horizontal scale, toward its peak on the higher end of that scale. The corresponding ogive looks like a shore profile that initially rises gently, like a gradually up-sloping sandy beach, becoming steep at the end. Such an ogive can be interpreted as a shore situation in which most of the 'waves' have no difficulty rolling over that initial, flat part of the shore profile, breaking further

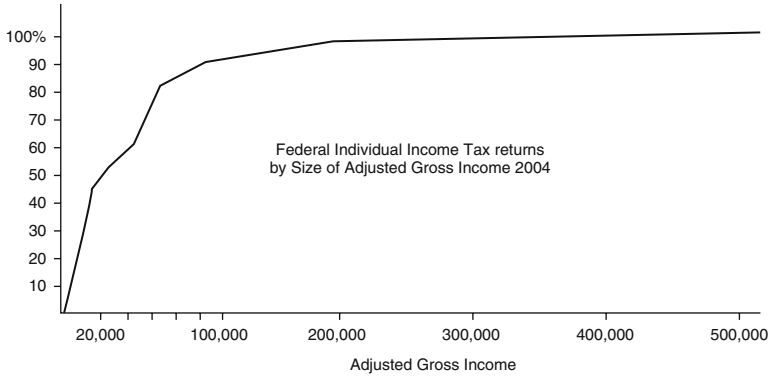

Fig. 8.7 Ogive of the grouped income distribution in Fig. 8.6

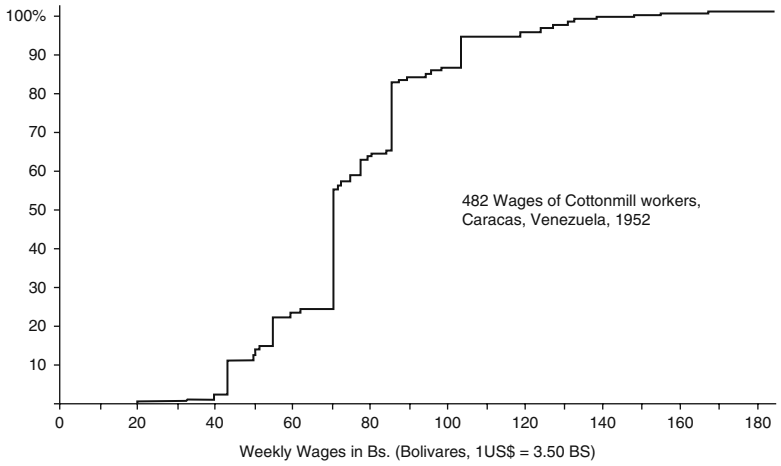

Fig. 8.8 Ogive of the Ungrouped wages of Fig. 8.2

back, higher up where it becomes steep, putting up a growing, increasingly stronger resistance to its further advancement. It could be the ogive of a society with a highly progressive income tax structure, making it increasingly difficult for persons or corporations that in an environment with a less punishing tax structure would earn large after-tax incomes.

When plotting the percentages instead of the absolute values of the cumulative frequencies, and also converting the horizontal scale into percentages, one can readily compare the ogives of different frequency distributions.

Ogives can also be established from ungrouped data. Such ogives present a serrated profile of many small steps. Each additional 'statistical-counting-unit' as the 1/N th case, becomes a tiny 'riser' of the many small steps that form a cumulative frequency curve (Fig. 8.8). From here it is only a small, further step, to extract additional information from standardized ogives by using the cumulative total – e.g. personal income – value on the horizontal axis to arrive at one of the variants of a 'measure of concentration'[17]

Summing up, it is important that something be known about the 'Eigenschwankung' of the 'real-life-objects', and the uncertainties involved in the process of recording the 'statistical-counting-units', the data that form the frequency distribution. This knowledge is to be used when interpreting a frequency distribution. Finally, it should be kept in mind that socio-economic phenomena are portrayed mainly through qualitative variables and the geographic location of the 'statistical-counting-units'. Quantitative variables, presented in frequency distributions, though having historically drawn attention early on and are well suited for numeric manipulation, are less important, acquiring meaning only in conjunction with some qualitative characteristics as their complement. (see Endnote 2)

8.7 Interpreting Measures of Location, Dispersion and Asymmetry

Socio-economic frequency distributions, as a rule, are asymmetrical – usually skewed to the right. This fact compromises the use of parametric inference, which is predicated on the normality of population distributions. Asymmetry also places in jeopardy the value of methods of location or central tendency, and measures of dispersion, all of which really make sense only in fairly symmetrical frequency distributions. Logarithmic and other empirical data transformations, make the distributions resemble more closely the data model in the natural sciences but tend to distort the underlying social or economic facts. The widely used measures of location, dispersion and asymmetry need to be re-evaluated in light of the approach to socio-economic data taken in this book.

8.7.1 Measuring 'Central Tendency' and Location

In the early days of statistics the number of measures of location or central tendency proliferated. Of those only the arithmetic mean survived, with the median and mode taking distant second and third places.

8.7.1.1 The Arithmetic Mean

Although most computer programs automatically calculate the arithmetic mean, one should keep in mind, that it is essentially a per-unit ratio in a socio-economic statistical aggregate, and is as vague as the latter (Chap. 3 and Sect. 4.3). If one were to plot the mean of the ungrouped values of a numeric characteristic of the 'statistical-counting-units' it would be a barely visible, gray mark amidst the clearly marked original values. As a further indication of the belief that socio-economic statistics is like statistics in the sciences, the arithmetic mean was seen as the point at which the lower and the upper portion of a distribution are in balance, like the equilibrium of a physical body suspended at its center of gravity. In social-science data, however, to consider the arithmetic mean as such a center of gravity is an unrealistic fiction.[18]

The arithmetic mean has a different role and interpretation in measurements of engineering and the natural sciences. In those distributions the arithmetic mean is meant to represent the 'true value' from which the values in the frequency distribution can be assumed to deviate due to random disturbances in the production run or in the measuring procedure. Although data in the social sciences are beset by all kinds of 'errors' (see e.g. Chap. 2) they are not like those in the 'hard sciences.' The transfer of that paradigm to the data in the social sciences is misguided, because our data lack a comparable center. The arithmetic mean, instead, is to be interpreted as one of the ratios, discussed in Chap. 4.

8.7.1.2 Other Measures of Centrality

The **Median** is the numeric value of 'the "statistical-counting-unit" in the center,' in the typically right-skewed frequency distributions of socio-economic data, below the value of the arithmetic mean. The median considers only the position in the center of the rank-ordered data, from smallest to largest, not their numeric values. The lack of sensitivity of this measure to outliers makes it more desirable for socio-economic frequency distributions than the arithmetic mean. In the early days of statistics it was involved in a measure of asymmetry. The median value is easy to interpret, giving an idea of the location of the center. It can be of interest e.g. when comparing different frequency distributions.

The **Mode** is a deceptively simple looking measure. It is not really a measure of centrality in a distribution but the location of the value on the horizontal scale of the highest density in a given neighborhood. Most socio-economic distributions have more than one such modal points, none of which is necessarily in the middle of the distribution. These various **modal values** usually are hidden within the progressively wider class intervals. The classes containing one or more **modes** – in addition to the obvious one in the class containing the highest frequency, the modal class – may not be noticed among their neighboring classes that may have higher frequencies. When interpreting a frequency distribution of socio-economic data this fact should be kept in mind, and not too much is to be made of the one obvious modal value. Each one of those additional modal values in a frequency distribution is a hint that a subcategory, with its own shape of a distribution, had been lumped together with other frequency distributions of different sub-aggregates to form the frequency distribution under consideration.

Only few analysts rely on the mode(s). Karl Pearson even preferred to completely bypass the mode in his well-known measure of asymmetry (For an example of various modes in a distribution see Fig. 8.2, the tally sheet of weekly, ungrouped wages).

8.7.1.3 The Fractile Values

Median, Quartiles, Sextiles, Octiles, Deciles, Percentiles[19] are values of individual 'statistical-counting-units'. 'Fractile' is the name for any of these values on the horizontal scale, marking the points where 50, 25, 16.67, 12.5, 10 and 1% of the distribution are located. They are, like the median, non-parametric, rank-order determined. When a **fractile** value falls between the values of two 'statistical-counting-units', its value is indeterminate between these lower and upper neighboring values. In grouped data fractiles have to be estimated (interpolated), thereby losing some of their immediacy and simplicity of interpretation. The **median** is one of the fractiles that can provide, in conjunction with other fractile values, a measure of dispersion (see the next section), and allows to assess the degree of asymmetry of a frequency distribution (Sect. 8.7.3).

8.7.2 Interpreting Measures of Dispersion

The **standard deviation** is the complement to the arithmetic mean and considered to be **the** undisputed **measure of dispersion**. The dominant role of the standard deviation – and its value without the square root, the variance – was never questioned as the true measure of dispersion because of its clearly defined role in the normal distribution which is central to statistical theory. Through its role as the paradigm of measurement errors in the sciences, and its role in sampling theory, the 'Gaussian' distribution believed to be 'normal' in sciences, morphed into the idealized prototype of all frequency distributions. The normal distribution, however, **is not** the prototype of the phenomena in the social sciences and consequently the standard deviation does not have a theoretically backed[20] privileged standing like in the natural sciences. Given the tri-dimensional structure of socio-economic data, the meaning of the standard deviation as an 'ideal measure of distance' from its 'ideal center' (the arithmetic mean), is not clear. One can agree with the statement of a well-known textbook author that: 'The variances of skewed distributions do not have any descriptive meaning at all'.[21]

The standard deviation, as well as the variance, and consequently also, to a certain degree the arithmetic- and other means in frequency distributions of socio-economic data, defy a meaningful interpretation[22] That is also true, by and large, of the mean deviation, computed either from the arithmetic mean, as $\Sigma(|x_i - \mu|)/N$, or from the median as $\Sigma(|x_i - M|)/N$. Both formulas assume that all data are in essence repeated measurements of the same, single true value from which they deviated randomly, due to errors of measurement. Neither the standard deviation nor the mean deviation, and implicitly, the arithmetic or any other mean, can be recommended for socio-economic distributions. Data in the social sciences do not have a center comparable to the assumed center of the normal distribution in the 'hard sciences'.

To express the spread of the data in a frequency distribution, measures should be used that do not consider the data as 'deviations' from some expected, supposedly true value, but take into consideration the decentralized nature of socio-economic data. Corrado Gini's measure of the average inter-unit spread comes to mind. It certainly is the sort of non-central measure of dispersion like the one discussed here. The difficulty of actually computing Gini's measure has prevented its application in praxis. Simpler measures are needed that do not depend on the assumption of a center, from which the data are considered to deviate randomly. Measures based on inter-fractile distances fit this requirement, particularly the 'average inter-quartile distance,' determining the average spread of each 25% of the distribution as

$$\overline{Q} = \Sigma(Q_i - Q_{i-1})/4 \text{ for } 1 \leq i \leq 4$$

where Q_0 is the value of the first "statistical-counting-unit" at the beginning of the sorted, ungrouped array of the data, or in grouped data, the numeric value of the lower end of the first class. Q_2 the median, and Q_4. the value of the last "statistical-counting-unit" or the upper end of the last class. The interpretation of this measure is

clear and simple.$\overline{Q}$ then is the average spread or dispersion of 25% of the 'statistical-counting-units', regardless of the shape of their frequency distribution.

The distance between each two **deciles**, the average inter-decile distances gives a finer account of the spread of the 'statistical-counting-units'

$$\overline{D} = \Sigma(D_i - D_{i-1})/10 \text{ for } 1 \leq i \leq 10$$

In which each $(D_i - D_{i-1})$ for $1 \leq i \leq 10$ contains 10% of the data in a frequency distribution.

Measures of centrality, together with a measure of dispersion, however, do little to clarify the underlying socio-economic situation and are not really helpful to interpret a socio-economic frequency distribution.

8.7.3 Interpreting Measures of Asymmetry

Though regularly taught, 'measures of skewnes' are rarely used[23] and even less interpreted because they are difficult to make sense of. Fig. 8.6 is to remind the reader of the common shape of distributions in the social sciences (see also Table C.1 of the Appendix to this chapter for various measures of asymmetry computed for this and other distributions e.g. note in Fig. 8.9).

The same inter-fractile intervals, used in measures of dispersion, suggested in the last section, can also be used to measure asymmetry Fig. 8.9 for the distribution of

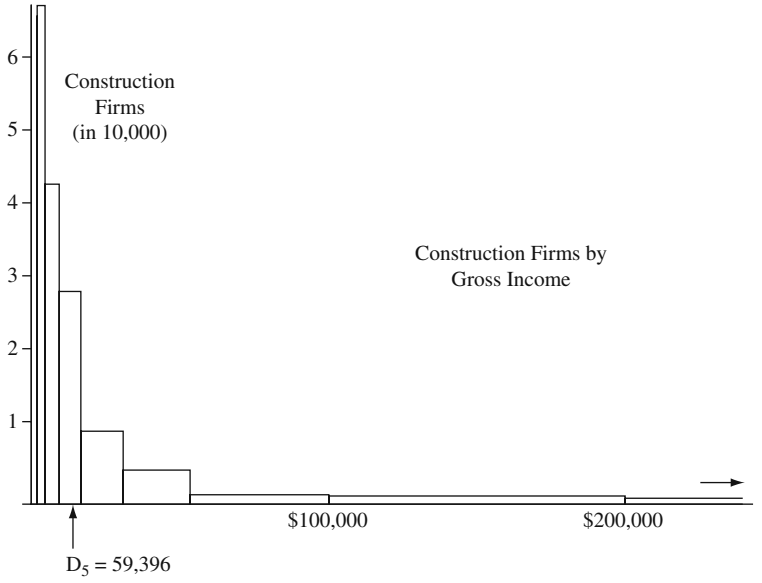

Fig. 8.9 Construction firms by gross income

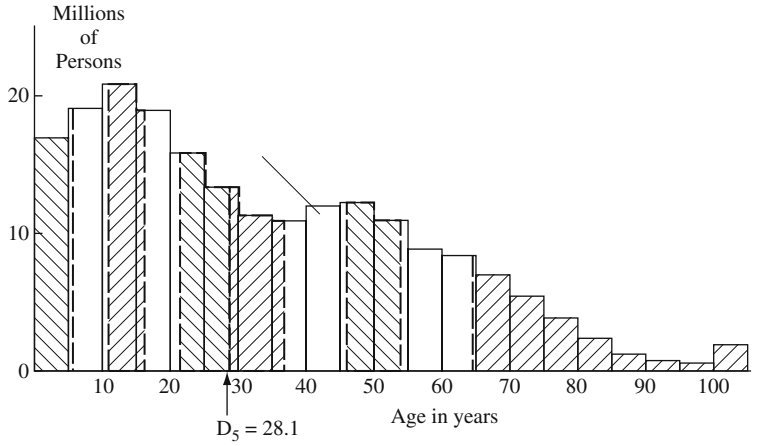

Fig. 8.10 US populations by age, 1970

age in Fig. 8.10. The resulting measures do not involve the 'moments' of a distribution about the mean, and are easy to interpret. Inter-decile distances are sensitive to departures from symmetry, and can be determined in ungrouped data and estimated reasonably well in grouped data (frequency distributions). This measure of asymmetry $\overline{A}$ consists of the geometric mean of the five ratios A_i between concentrically placed inter-decile distances (Fig. 8.11). The numerator of each of these ratios contains the inter-decile distance, the difference between two contiguous deciles to the right of the median of a frequency distribution. The denominator contains the inter-decile distances symmetrically placed to the left of the median, (which is the fifth decile, D_5).

The differences between each two of the eleven decile points D_i, for $1 \leq i \leq 10$, are the inter-decile distances $D_i - D_{i-1}$. There are ten such inter-decile distances[24] corresponding to the subdivision of the entire frequency distribution into ten segments, each containing 10% of the data. This measure uses the interval on the horizontal scale that each segment occupies, to determine the degree of asymmetry in each concentric segment of the frequency distribution.

The median is the value around which the lack of symmetry is measured. The inter-decile distances in the lower half of the distribution, to the left of the median, are compared with the inter-decile distances to the right of the median. Then A_1 – 'A' for asymmetry – is the ratio between the inter-decile distance $D_6 - D_5$ to the immediate right of the median, in the numerator, with the inter-decile difference to the immediate left of the median, $D_5 - D_4$ in the denominator. Each inter-decile distance encompasses 10% of the persons in the age distribution of the U.S. population of 1970.[25] The ratio of these two differences

$$A_1 = (D_6 - D_5)(/)(D_5 - D_4) = (36.65 - 28.07)(/)(28.07 - 21.32)$$
$$= 8.58/6.75 = 1.27,$$

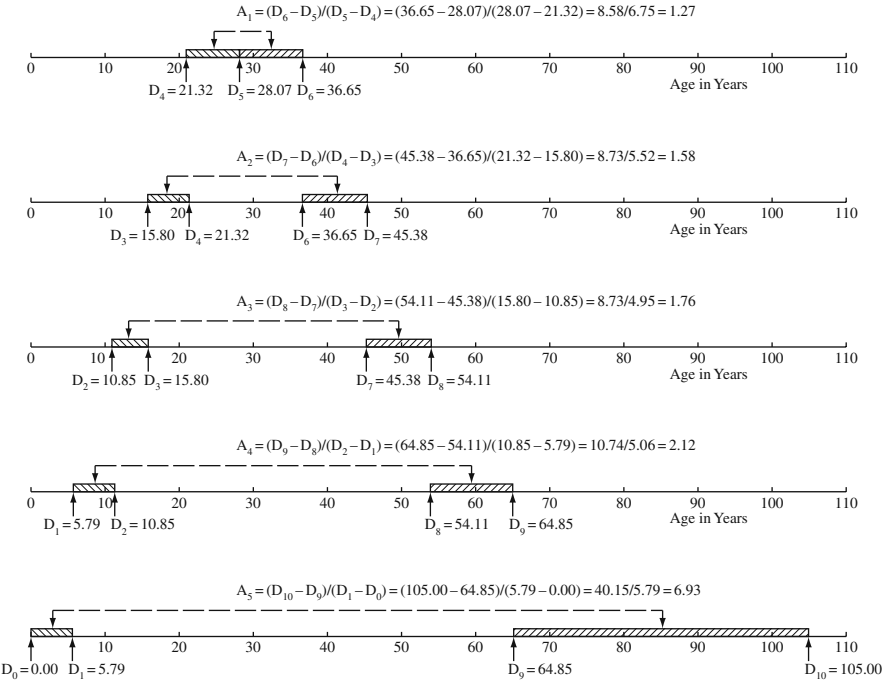

Fig. 8.11 Schematic presentation of Inter-decile Ratios

as can be seen in Fig. 8.11. The ratio A_1 informs about the central 20% of that distribution indicating that the inter-decile distance to the right of the center (the median or fifth decile) ($D_6 - D_5$), is 27% more extended than the inter-decile distance ($D_5 - D_4$) immediately to the left of the center. A_1 contains no information about any other part of the distribution. A_2 explores in an analogous manner the symmetry or lack of it in the next concentric zone, covering the next 20% of the distribution (see Fig. 8.11).

$$A_2 = (D_7 - D_6)/(D_4 - D_3) = 1.58$$

It tells us that the 10% of the population between the ages D_6 = 36.65 years and D_7 = 45.38 years, with an inter-decile distance of ($D_7 - D_6$) = 8.73 years, to the right of the median, is 1.58 times more spread out than the 10% of the population between D_3 = 15.80 years and D_4 = 21.32 years, ($D_4 - D_3$) = 5.52 years. The ages of 10% of the persons to the right of the middle of the age distribution is 58% more spread out than the younger persons of the corresponding concentric 10% to the left of the median age. A_1 and A_2 measure the asymmetry in the central 40% of this frequency distribution, between D_3 and D_7. Each concentric pair of inter-decile distances is assessed separately. For a detailed analysis of this measure, as well as comparisons with other measuring methods of asymmetry, see Appendix C. The

row vector of these six ratios for this distribution is (A_1, A_2, A_3, A_4, A_5; $\overline{A}$) = (1.3, 1.6, 1.8, 2.1, 6.9; 2.2) completely describing the nature of its asymmetry, is easy to determine, (relatively) reliable (because of the need to interpolate, to estimate these A_i values inside each class) and easy to understand. Interpreting this numeric result indicates that e.g. the oldest 10% of the population is 6.9 $\cong$ 7 times more dispersed than the youngest 10%. Stated differently, the density at the beginning of the frequency distribution of ages is seven times higher per year for the persons of ages 0–5.79 years than for each year in the last group of 10% of the highest-aged segment of the US population, between ages 64.85 to 105 years. In short, there are seven times more young people per year of age in the age group of those six years or younger folks, than for each year of elderly person of age 65 or older. Such comparisons, by the way, can also be made for any two inter-decile age groups, not only for those symmetrically placed in a mirroring position as in the A_i measure see the Appendix C to this chapter for more interpretations.

Notes

1. Suppose one is informed about the size of the assets of organizations – a quantitative variable – without also being informed about the kind of economic activity they pursue, the kind of products they produce, their legal status as a publicly or privately owned corporation, their geographic location, type of investment and a number of other non-quantitative pieces of information about those organizations. The information about the asset sizes alone, without any other qualitative information about these business firms would serve no purpose and be meaningless, unless and until some of their qualitative characteristics have been specified.
2. The following are a few examples:

A. Product standardization, e.g. the garment and shoe industry rely on size classes of body measurements for the production and marketing of a reduced number of shoe and garment sizes. Forming the proper classes greatly affects inventory costs. (J. Sittig, "Standardization of Garment Sizing by Statistical Methods" *Bulletin de L'Institute International de Statistique*, Tome XXXIV, 1e Livraison, 28th Session of ISI, Rome, 1960, pp. 317–22).

B. The frequency data on business firms, available through the IRS (Internal Revenue Service, to collect the federal income tax) are presented in unequal classes that are determined by federal law. This has direct consequences for research on the financing, marketing and management of those business firms. Researchers must pursue their goals within the constraints of these uneven class intervals as given by the corporate income tax law. (M. Gupta "The Effect of Size, Growth and Industry on the Financial Structure of Manufacturing Corporations" *Journal of Finance*, Sept. 1969, pp. 517–529.)

C. There is a growing need for coordination of statistics among countries. In particular the international comparability of frequency distributions with unequal classes has concerned the statistical offices of the UN, EEOC, OAS and others. These offices have issued many recommendations, among which the coordination of uneven class intervals in economic distributions is of special interest. "In countries where the bulk of industrial activity takes place in small establishments, there should be many classes with small spans for the lower size groups, and few classes with wide spans for the upper size groups … common lower or upper limits to class intervals among countries are 10, 40, 100, 500, and 1,000; and for countries with many small establishments, five. It is suggested that countries use at least these points …" *Studies in Methods: Industrial Censuses and Related Enquiries,* UN Publ. Series F, No. 4 Vol. I, pp. 200–204

D. In the audit of inventories and customer accounts, book values are usually grouped into strata which are really the classes of a population frequency distribution of book values. The issue of forming classes of unequal widths is of substantial interest to auditors given the importance of stratified sampling. T. W. McRae, *Statistical Sampling for Audits and Control*, John Wiley & Sons, N. Y. Chichester, 1974 esp. "The Problem of Stratification" pp. 65–73. The author suggests among other methods the use of a Lorenz Curve to determine the most suitable strata which is just another name for what this chapter is concerned with, viz. classes.

3. "The matter of clarity vs. obscurity may also be approached from the point of view of Whitehead's philosophy of perception . . . . One of the most important axioms that I have found in Whitehead's thought is that those things which are most clear and distinct are not necessarily the most real. "Those elements of our experience which stand out clearly and distinctly in our consciousness are not its basic facts." (Alfred North Whitehead, *Process and Reality*, corrected ed. ed. by David Ray Griffin and Donald W. Sherburne, New York, The Free Press, 1978, p. 162) . . . And if mathematics is at hand we can easily slip our abstractions into the niche of the purely quantitative. Mathematics deals quite easily with the quantitatively clear and distinct, but it has trouble with the qualitatively opaque and important. In order to make things clear it has to dispense with most of what is relevant in a phenomenon, whether the latter be an atom or the universe. Whitehead's advice is to mistrust our abstractions since they are not identifiable with concrete reality. (p. 140) The problem of science and religion arose . . . because of the modern bias that the clear and distinct are also the most fundamental and that lack of clarity means absence of realism . . . It is one of the. . .aspects of Whitehead's thought that it challenges both Descartes and the empiricists. Whitehead questions Descartes' assumption that 'clear and distinct' necessarily means concretely real. . . .His own 'radical empiricism' goes deeper than the abstractions that are always the result of our attempts to clarify. (p. 142). . .radical empiricism. . .recognizes. that our senses bite off only a tiny contemporary cross-section of reality and that our abstractive intellects may remove us even further from the intrinsic reality, depth and importance of things. . . (p. 143)" John Haught, op. cit.
4. Particularly a study of the 12 daily measurements of the fluctuations of the height and weight of 40 orphaned school children during a period of 30 consecutive days, that I carried out with a measuring instrument especially designed and built for this research, for the Instituto Nacional de Nutrición, in Caracas, Venezuela, in 1953. It was published as . "Medición de Talla y Peso en Niños: Cuanta Confianza Merecen?" *Anales Venezolanos de Nutrición*, Fundación Cavendes (2), pp 127–138, 1997, Caracas, Venezuela I found a very similar result in another, unpublished study of the weekly fluctuations of the wages of a group of 50 cotton mill workers in Caracas, Venezuela 1951.
5. Oskar Morgenstern, *On the Accuracy of Economic Observation*, Princeton University Press, Princeton, NJ, 2nd ed. 1963.
6. Anecdotal rumors have it that there are three kinds of accounting data: 'accounting for income tax purposes' ' accounting for a statistical report' and 'accounting for management decsions'. There is also further discussion about accounting in Chap. 11. See also the subsequent Endnote 8.
7. "The Botswana Agricultural Census of 1982 highlighted . . . response errors in the data supplied on cattle numbers . . . a tendency of respondents to understate the size of the herd . . . particularly when the respondent is not the holder, and a tendency of respondents to exclude oxen from their herd count – some holdings reported owning and using oxen for cultivation but declared nil oxen as part of their cattle herd. Pilot post enumeration checks of holdings using but not declaring ownership of oxen as part of their cattle herd suggested that the overall estimate of the cattle population would need to be raised by about 17%" Robin Rothfield, Australian Bureau of Statistics, *The Survey Statistician* IASS, #15 – June 1986.
8. "Sunrise Senior Living yesterday reported that it plans to reduce recorded profits by $98 million to $107 million for 1999 though 2005 as it corrects the way it accounted for certain

joint ventures. The McLean company, which operates hundreds of communities for senior citizens in the United States, Canada and Europe, said those profits would likely be moved to show up as income from 2006 to 2008. The company offered no specific timetable for when it would formally restate its results, saying only a filing was near.

Brad Rush, Sunrise's chief financial officer, said the company was going through the 'last stages' of its accounting review and was close to submitting its recast results for 2003 through 2005" "Sunrise to cut Profits by About $100 Million" *The Washington Post*, Washington, DC, Jan 28, 2007 p. D4.

9. For a long time Sturges' formula – H. A. Sturges "The Choice of a Class Interval" *JASA 21, 1926* pp. 65–66 – was the only contribution of statistical theory to the formation of classes in frequency distributions. The number of classes, not class intervals or locations of those classes, are determined by this formula through analogy with a binomial probability distribution. The actual frequency distribution is assumed to consist of pure algebraic numbers that result from Bernoulli trials, removed from historic time and geographic space. The resulting classes are of equal width. Sturges' algorithm was neither meant for skewed distributions, nor does it throw light on the socio-economic situation underlying the data.
   See: Robert Parsons, *Statistical Analysis: A Decision Making Approach*, Harper & Row, N. Y. 1974, p. 13, Note 1.
   Neyman's criterion is concerned with allocation of sampling units over various strata so as to minimize total sampling variation. It deals with establishing classes in frequency distributions without specifically saying so. The purpose of this grouping procedure is to minimize the variability of the counting elements in the strata, not to highlight the economic content in the frequency distribution. Although this is an improvement over Sturges' approach, it also considers the data as abstract numbers, not as descriptors of some socio-economic reality. Other proposals for optimum stratification are similar, e.g. Tore Dalenius "The Problem of Optimum Stratification – Fundamentals of Stratified Sampling" in *Selected Publications*, Vol. 5 University Institute of Statistics, Uppsala, Sweden, reprinted from *Skandinavisk Aktuarietidskrif*t, 1950, pp. 203–213.
   Another "method" which explicitly addresses the formation of classes in frequency distributions is the empirical rule found in most textbooks of Business and Economic Statistics: "divide the range of the frequency distribution into at least five but no more than 25 equal classes". Such a procedure does not take into consideration the phenomenon to be studied. It is indifferent about how much and which detail can be tolerated to be lost during the formation of frequency classes. A wrong paradigm for socio-economic data underlies this empirical rule.
   Clustering algorithms – empirical procedures guided only by the proximity between individual measurements – developed independently of statistics, so far have not been used yet to determine classes in frequency distributions. These procedures identify clusters of counting elements and the existence of gaps between clusters. They are not directly concerned with forming classes for frequency distributions, but appear to be a promising tool which would be suited to gain insight into the social forces that act on the data. Clustering is also part of the procedure suggested in the following. J. A. Hartigan, *Clustering Algorithms*, N. Y. John Wiley & Sons l975.
10. As a first approach, use a tally sheet with an amplified, complete scale, drawn on paper that is as large as needed, e.g. a roll of continuous wrapping paper, to plot all the occurring values along an unabridged horizontal scale. The resulting tally is more revealing than the procedures available on computer packages, e.g. all programs called 'FREQUENCY' in SAS and SPSS, which abridge the "empty," unused distances between the occurring cases.
11. This set of 482 wages of construction workers in Caracas came to my attention in 1950, through the director of the Seguro Social de Venezuela, in Caracas who at that time was my student. It appeared to be just another, uneventful data set. When I analyzed these data in the described manner, however, I discovered oddly spaced clusters. Unfortunately, by then I had lost contact with that student to get this unexpected phenomenon explained. Therefore those clusters could not be explained. It was also too late to request additional information about these workers that would allow to determine the 'Eigenschwankung' of this material. That

insight, that the fluctuations over time of each 'statistical-counting-unit' – the reported wages of each worker – should be used for appropriate class intervals, occurred to me much later.

In this context John W. Tukey's well-known Stem and Leaf procedure should be mentioned. His approach, however, is only a streamlined version of equal class sizes (stems) with no regard for differences in precision and 'Eigenschwankung'. *Exploratory Data Analysis,* Addison Wesley Publishing Co. Reading, 1977, pp. 1–25.

12. The ungrouped data can also be grouped in overlapping, gradually advancing classes of the same width, e.g. of 5. Then, instead of classes $0 \leq x < 5,\ 5 \leq x < 10$, etc. classes like $0 \leq x < 5,\ 1 \leq x < 6,\ 2 \leq x < 7,\ 3 \leq x < 8,\ 4 \leq x < 9,\ 5 \leq x < 10$, etc. When plotted, these advancing classes will give the approximate contours of a line graph of that distribution – assuming that a class-width of 5 is called for.
13. While the driving force, e.g. the motor of a car, operates within a small range of the number of revolutions of its drive shaft, the effort demanded from that motor can vary considerably with the terrain. With the help of gears, however, the relatively same effort of the motor can be transformed into slower speed on uphill travel, or higher speeds on level ground posing only moderate resistance to the car's advancement, yet all along, using roughly the same motor force. By shifting into the next higher gear the car speed can be raised to a new level in steps that parallel the widening class intervals in these frequency distributions of official publications. The analogy reflects the intuitive understanding of statistical practitioners who "shift" to the next wider class like a driver of a car shifting the gear to a larger gear in a terrain that places less resistance to achieve higher speeds with the approximately same effort of the motor.
14. e.g. under 1, 1 to under 2, 2 to under 5, 5 to under 10, 10 to under 25, 25 to under 50, 50 to under 100, 100 to under 250, etc., e.g. of farms by size of farmland in ha., or of business firms by number of employees, or its long term assets, in million $, of politico-administrative districts by size of population, etc. This kind of pattern originated in some administrative activity in a distant past long ago and has been accepted by statistics such as e.g. the legally established groupings of incomes for tax purposes. From there it has been adopted worldwide, for most socio-economic frequency distributions. Although the user of these published data has no influence on how these frequency distributions were made, nonetheless, the discussion of this chapter should prove pertinent and helpful when interpreting such frequency distributions.
15. Occasionally the total value of the characteristic in that last class is provided – e.g. total assets of the firms in each class in a frequency distribution of assets. In that situation the average value of assets can be determined for the open ended class, and used as its midpoint. Then an upper limit for that class can be estimated. There is no assurance, though, that such an estimated upper limit really will include all of these very large data-values in the open end of the distribution.
16. Mathematical statistics had the correct intuition, giving priority to the cumulative distribution function F(x). The probability density function f(x) is then considered as secondary, as the probability density function, the first derivative of the cumulative density function. But no explanation of the reasons for preferring the cumulative over the simple frequency distribution. were given.
17. See e.g. Bruckmann, Gerhart "Konzentrationsmessung" as Chap. 26, The Measurement of Concentration, in: Bleymüller, Josef, Gehlert, Günther and Gülicher, Herbert, *Statistik für Wirtschaftswissenschaftler* München, Verlag Franz Vahlen, 1981, zweite überarbeitete und erweiterte Auflage, pp. 185–190.
18. Take, for instance, the work force of a factory where workers earn different weekly wages because of different occupations, skill levels, diligence and experience. Assuming that their earnings, instead of being paid out on Friday, were pooled and re-distributed, so that each receives the same average amount as the wage for that week.

    In the next pay period the former high earners will either leave for another job or adjust to this situation by working less diligently and earning less, while the lower paid workers,

receiving the same average amount as their wage for that week which is higher, without any greater effort on their part, would not serve as an incentive to work faster and better to justify that unexpected increase in their wages. The average weekly wage of that next 'round', showing these effects, would be lower, initiating a downward spiral. The arithmetic mean, consequently, does not answer questions such as: "what would these workers have earned if everyone had received the same pay," or "what would they have earned if everyone had the same professional qualifications and motivation." This illusory idea of the arithmetic mean can be taken as an indicator of the effects of a re-distribution of wealth that may have been at the root of socialist economic thinking. One may have to find answers to such questions through experimentation and research, not through theoretical "armchair solutions," like computing an arithmetic mean. If that were really carried out, such a re-distributed arithmetic mean wage would neither be the ideal of distributive justice, nor a prudent move from management's point of view. It would discourage the skillful and diligent high paid workers, while not serving as incentive to increase the output of the less proficient, less qualified lower paid workers.

19. In the literature the terms "Fractiles," "Quantiles," and "Percentiles" are used interchangeably. "Deciles and "Quartiles are known best. Less familiar are Sextiles.and Octiles.
20. In the formula of the standard deviation $(x_i - \mu)$ is the distance of a value x from a 'center of the distribution' that rarely is of particular interest for the social situation under study. $(x_i - \mu)^2$ is the area of a square – a geometric figure – which has the indicated distance from μ as the length of the side of a geometric figure, the square. Adding together the areas of all these squares (as geometric figures), that have as the length of its side the $(x_i - \mu)$, results in a large geometric figure, the area of which is the sum of all individual deviations squared. The division by N, the number of such squares of deviations from the mean, gives the area of the geometric figure, the 'average square'. Its square root, σ, is the length of the side of that average square-shaped geometric figure. The value of σ indicates that distance from the fictitious center μ by which not quite 1/3 of the total value in the normal distribution would be concentrated. That is not true for other than the normal distribution. Where the number of actual cases fall within $\mu - \sigma$ and within $\mu + \sigma$ is not as predictable. The standard deviation expresses the actual diversity of individual values in the frequency distribution for the two equally distant points $\mu \pm \sigma$. The individual values of the frequency distribution are believed to differ, on average, by the amount σ from μ. Although universally accepted the standard deviation does not give an intuitively understandable sense of dispersion.
21. A.S.C. Ehrenberg, *Data Reduction*: *Analysis and Interpretation of Statistical Data* John Wiley & Sons, London-New York, 1975, p. 203.
22. Chebychev's theorem,

$$P(|x - \mu| \geq k\sigma) \leq 1/k^2$$

is the best general statement about sigma for other than the normal distribution. But that theorem gives reasonable answers only for values of $k > 1$. For $k = 1$ the theorem states that the probability of a given value x falling beyond one standard deviation can be as high as 1, which indicates that, based on the knowledge of sigma, nothing can be said about the distribution in its main part. For distances of $k<1$ standard deviations, Chebychev's expression cannot be determined at all.

23. See Rutledge Vining, "Regional Variation in Cyclical Fluctuation," *Econometrica,* Vol. 13–3, especially p. 198;
James P. George, "Negative Skewness and its Significance in Relation to Distribution of Performance Ratings of Civil Service Employees," '1958 Proceedings of the Social Statistics Section of ASA', Chicago, pp. 259–260; William Droms, Charles W. Miller, and Grant A. La Cert, "Financial Characteristics of Small Retailers," paper presented at the Annual Meeting of the Financial Management Association, Oct. 13–14 1977 (unpublished, courtesy of Dr. Wm Droms).
24. The expression "inter-decile distance" is clearer than "inter-decile deviation" which might be mistaken to mean an average of the ten individual inter-decile differences. It is also clearer than the term "inter-decile range" which could be misunderstood to mean the difference between

two deciles that are far apart like e.g. $D_7 - D_3$ in analogy to the "inter-quartile range". The beginning of the distribution, the value of the first "statistical-counting-unit" or the beginning of the first class, is the decile point $D_{0,}$ The tenth, last decile point, D_{10} is the value of the highest "statistical-counting-unit" , or the end of the last class.

25. "Age by sex, for the United States and Outlying Areas 1970 and 1960," Table 49 for ages 0 to 75 years, p. 1–263, and Table 50 for ages above 75 years, pp. 1–265, Vol. I, Pt. 1, Sec. 1, *1970 US Population Census, US Summary,* Bureau of the Census, Department of Commerce, Washington, D.C., 1973. The D_I values are interpolated in that grouped data material.

Chapter 9
Cross Sectional Analysis in More Than One Dimension

9.1 The Case that Provoked the Re-interpretation of Regression

The following personal experience represented a milestone in my thinking about regression, and in fact, about socio-economic data in general. The data were from the sex-discrimination lawsuit,[1] in which I was a statistical expert-witness for the plaintiff. The group of female employees claimed to be paid less than male librarians at entry level, to have received smaller pay raises and fewer promotions, and yet to have the same qualifications and identical work requirements as their male counterparts. The plaintiff's lawyer approached me to explore and statistically establish the facts of their claim. The District Court ordered the unrestricted access to the employment records of the professional men and women in that large agency of the Federal Government.

For the discovery of such alleged differential treatment of equally qualified female employees the group of 32 librarians, 18 male and 18 female professionals, seemed ideally suited to serve as a pilot study for the alleged discriminatory treatment of female employees. All of these professional men and women held comparable academic degrees and were charged with the same duties. I started with a simple linear regressions between salary and length of service for men and an analogous, separate study for women. This first approach provided unexpected, puzzling information. Given the salary structure of the federal government, I expected a positive association between length of service and salary, regardless of gender. The regression equation, computed separately for the male librarians[2] was SAL(M) = \$16,900 + \$1,380*ERDAEMPL Obviously it meant: 'The average beginning (at entering) salary for men was \$16,900 with a yearly raise of approximately \$1,380 for each additional year employed as a librarian at ERDA' which seemed to be a sensible interpretation (Fig. 9.1a). I expected to discover a similar situation among the equally qualified female librarians, but probably on a lower salary level. I was unprepared, however, for what I discovered. The linear regression equation for women was SAL(W) = \$26,500 – 1,020*ERDAEMPL (Fig. 9.1b). An analogous interpretation of the regression line of the female librarians in that government agency would state that these female librarians were paid an average entry salary of \$26,500. making those with the least time on the job the highest paid employees in that department. The slope of the linear regression line for female librarians,

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_9, © Springer-Verlag Berlin Heidelberg 2009

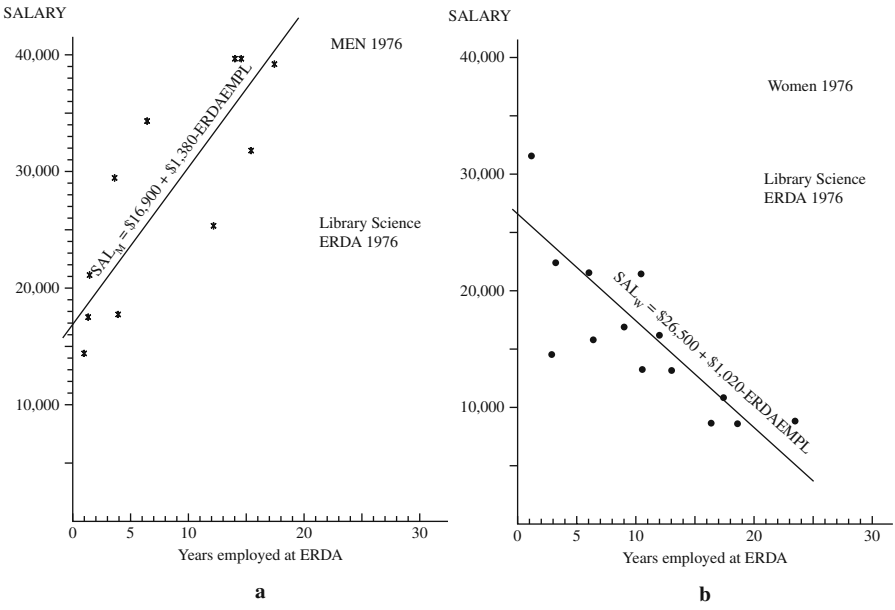

Fig. 9.1

$\beta_1 = -\$1,020$ indicated that the longer these female librarians had served at ERDA, the less they got paid, reducing their high entrance salary by $1,020 for each additional year at ERDA. – the Greek symbol β to indicate that these 18 women librarians were treated as a 'population,' not as a sample. The salary of these women apparently decreased, on average, by $ 1,020 per year. According to these data, they apparently got paid less the longer they worked. That obviously made neither sense nor was a credible description of the situation. It defied what is known about the US Federal Government's pay scale, and failed to conform to the general experience of the business world. What then could have caused this unexpected, puzzling statistical result? The plotted data for the male librarians revealed a sensible picture: lower salary for those who had been a shorter time with ERDA, and correspondingly higher salaries for the male librarians who had served longer on the job. The slope of the linear regression line was $\beta_M = +\$1,380$ per year of service The data for women librarians, however, did not make sense. Who would continue to work for an employer who seemed to penalize seniority on the job? This obviously had to be some error or data mix-up. I contacted the person in the government department responsible for these figures and got the promise of a careful review made credible by the court-ordered nature of this investigation. By the next day I was assured that no errors were involved and that the data represented the situation correctly. This excluded the possible explanation of an error in the data and made the situation even more puzzling.

The slope b or β. Then it dawned on me that the issue was not an error in the data or in the calculations. The error was in the interpretation! A change in

the interpretation, then, resolved this conundrum and began to make sense of the situation. It was also a ευρεκα moment of enlightenment, understanding that all regression analyses, not only this one, treat the data as cross-sections, regardless of the facts, as if the 'statistical-counting-unit's existed simultaneously, side-by-side. β_1 is not a dynamic factor of 'change – growth or decline – in Y for a one-unit increase in X'. Instead, **the slope** is **the difference** between simultaneously existing 'statistical-counting-units', assumed to be contemporaries at the point in time when these data were recorded. It was only a small further step to realize, that every 'statistical-counting-unit' is really situated at the intersection of the present with the past, usable for both, a cross sectional, static view of the situation, as well as a longitudinal, developmental perspective of that same situation.

A close look at individual librarians' employment histories revealed that the female librarians who entered this government agency after the promulgation of the law 'Title VII', – one of the laws against sex-discrimination was promulgated in 1969 and expanded in 1974 – were hired at starting salaries that were substantially higher than the starting salaries of the newly hired male librarians and of course, also much higher than the salaries at which women librarians had been hired years before the promulgation of these anti-sex-discrimination laws. These old-timers were still with that agency. This imbalance was the result of management's reaction to these anti-sexdiscriminaiton laws and the threat of law suits. Such a law suit was actually filed as a class-action sex-discrimination law-suit against that government agency, by women from all departments, not just the librarians. The situation reflected management's inadequate and inappropriate reaction to that legislation that was intended to correct discriminatory treatment of women. Management upgraded only the **average salaries** of these female librarians but made no attempt to adjust the salaries of its older, aggrieved employees. Management obviously assumed that only salary averages for each department would be used to check on compliance with that law, that required the elimination of discriminatory treatment of women.

Scatter-diagrams and regression equations always represent a cross-sectional view of a situation, not its longitudinal aspects, even if they deal with characteristics involving time, such as the 'employee length of service'. The dynamism of the salary situation, though, is indirectly revealed because all regression analyses give a cross-sectional view of the situation. The coefficient β_1 therefore **is not to be interpreted as** the amount by which the salary of an employee **increased** for each additional year of service. Instead the slope β_1 is to be interpreted as the amount by which the salaries of any two female librarians employed by that agency, existing side by side in 1976, differed overall by negative $1,020 for each year of difference of their affiliation with that agency; the longer ago they had joined ERDA as librarians, the lower was their salary at the time of this report, on average by $1,020. It was a static cross section view of the income situation in that agency in 1976.

In any regression case the slope b_1 or β_1 **must not** be interpreted as an 'increase – or decrease', or as 'the change in Y for a unit change in X.' It must be interpreted as a neutral 'the slope is the average difference in the Y-values that is associated with a difference in the X-values by one scale unit'.

This regression shows the result of the development of these salaries over time and also hints at gross differences in the development of these salaries. Female librarians who had joined the agency decades earlier, were paid much lower entry salaries. They also had received raises over the years, that were probably the same percent-increase but that were smaller dollar-amounts than the salary increases of male librarians of comparable qualifications and service. They were also substantially below the salaries of more recently hired women. The apparent paradox of female employees who seemed to be paid less the longer they worked at that agency was resolved when I recognized the **cross sectional nature** of every regression analysis.

In another statistical study of the real estate market in the suburbs of Washington, DC I found in multiple-regression analyses that houses with 4 bedrooms sold for $400 more than houses with 3 bedrooms. Does this mean that by adding a 4th bedroom, a house with 3 bedrooms would have been sold for $400 more? It cannot mean this, because in none of these sold houses the number of bedrooms had been changed. It indicates, though, that in that location at that time a house with four bedrooms, ceteris paribus, sold, by and large, for $400 more than another with 3 bedrooms. It does **not** indicate, however, that a given three-bedroom house in that location and at that time would have sold for more by that statistically determined amount if it had actually been expanded by adding one additional bedroom. It can only be taken as a stationary indication of differences that existed in that housing market.

If one actually had tried to investigate the effect of adding a fourth bedroom to a three bedroom house, one would have had to place various three-bedroom houses in that region on the market, sell them to the best offer, then oblige the new owners to enlarge these 3-bedroom houses by an additional bedroom. These enlarged but otherwise unchanged houses would have to be put back on the market and sold. Assuming that in the meantime the market situation has not changed, the difference of these converted 4-bedroom house-prices with their former 3-bedroom house-price would represent the changes in price due to the actual increase of the property by one bedroom. Evidently, if it were feasible at all, such an experiment would not be worth the trouble and cost. Yet, this is how one would have to proceed if one really wanted to ascertain a dynamic picture of the market value of the change in one of its characteristics. Given the practical impossibility of actually conducting such an experiment, one is inclined to accept, incorrectly though, the value of the stationary β_1 as the best available estimate of such a possible 'change in X – here number of bedrooms – on Y, the price of the property'.

In conclusion, the slope(s) of any regression line or surface, $ß_i$ must be interpreted as a static, cross-sectional statement. The interpretation must avoid dynamic expressions like 'change, increase or decrease' but use instead stationary expressions like 'differ by' or, 'on average is greater or smaller by the amount of b' to make users aware of the cross-sectional nature of every regression analysis.

Y-intercept, β_0, b_0 or 'a'. A few words about the mathematical Y-intercept. In most instances that number has no meaningfully interpretation. Although the regression function continues to negative infinity on the left side, and to positive

infinity on the right side, the statistical interpretation of the algebraic equation of the regression function can be made only within the limited range of those values of X for which values of Y are available. Only that portion of the mathematical regression function can be interpreted as statistically meaningful, that lies within the range of the data. Outside of that range, the mathematical function of the regression line does not make statistical sense and has no meaningful interpretation in terms of the problem. In all those cases the Y-intercept cannot be interpreted.

Usually the origin, at $X = 0$, falls outside the X-range of the data. The Y-intercept β_0, b_0 or 'a' of the regression line can be interpreted meaningfully only in those infrequent instances in which the data actually begin at, or extend to both sides of the Y-axis, beyond $X = 0$. Only then can the Y-intercept also have a socio-economic meaning. This is one of the many instances where statistics parts company with mathematics. The two fields overlap with regard to the use of numbers, but are different in their orientation (more in Chap. 10).

Regression analysis in socio-economic statistics deals with the real world that exists 'out there' functioning as a 'macroscope,' (Chap. 1 and Sect. 7.1) but is not concerned with what might be, should be or what is thinkable as expressed in econometric models, statistical forecasting being an exception.

When interpreting a regression line or surface everything that has been stated concerning the one-dimensional arithmetic mean (Sect. 4.1), can also be affirmed by expanding it to two or even more dimensions, the regression line or regression surface. In the social sciences regression analysis usually does not encounter clear-cut, one-on-one cause-consequence relationships. Besides, relationships that appear at one level of aggregation change, even vanish when aggregating or de-aggregating the same data. The current regression model, adopted from the natural sciences, considers the data that deviate from the chosen mathematical regression function – the vast majority – as the randomly distributed errors of measurement that follow a Gaussian distribution. That model is not appropriate for most socio-economic phenomena that are represented by data that are aggregates.

9.2 An Irreverent View of Least Squares

The least-squares algorithm of regression/correlation does not specify which type of mathematical function to use. This decision is left to the researcher. The least-squares algorithm only assures that the chosen line or surface will be fitted to the data in such a manner that the sum of the squares of the vertical distances to the Y-values of the data will be the smallest of all possible other positions of the chosen regression function. 'Least squares' guarantees only the 'best fit' of a pre-selected algebraic function to a set of data, regardless of how well-suited the chosen function is for the data and for the relationship.

Linear regression proceeds like a person who wants to place a straight bamboo rod with the help of rubber bands of equal size that are to be fastened to nails that were not fully hammered in, that are spread out like data points, or like the

cross-sected cable endings, sticking out from the surface of the scatter-diagram – an image proposed in Chap. 8 – to which the rubber bands can be fastened. To carry this out, two persons would be needed: one to steady the bamboo rod holding it in place, the other to fasten the rubber bands to the rod and vertically to the nail heads that are at various distances from the rod, like statistical data from the regression line. Those nails or 'cable endings' that are farther away from the rod will cause the rubber bands to be correspondingly more stretched, exerting a stronger pull on the rod than those connected to nearer points. Rubber bands connected to more distant points will be taut, those connected to nearer points will be slack. When the rubber bands have been connected to all points above and below the rod, its sudden release will cause the rod to 'snap' out of its initial, hand-held position, and after some corrective oscillations, will settle in a position of least tension of all the rubber bands. If one assumes that the tension in the rubber bands increases with the square of their being extended, this rod then would have been fitted by the least squares principle. A procedure like this one determines algebraically the values of $ß_o$, for $ß_1$, r, and if more independent variables are involved, of all the β_i.

The real issue, however, is whether and to what extent, the mathematics of regression and correlation contributes to a better understanding of a given socio-economic situation. What then is the contribution of such a least squares regression line to the understanding of a socio-economic situation? Measures that indicate how well the regression line fits the data, such as the coefficients of correlation and determination, say little about the underlying complex socio-economic relationship. The 'double precision' with which the calculations can be accomplished by a computer program is spurious, not like in the hard sciences, as recognized by statements like: The formal association between the values of X and Y, expressed by the regression function $Y = \beta_0 + \beta_1 X$, or as $y = a + bx$, or with a more complex formulation' should not be taken to mean that there exists a causal, or any relationship at all between them.'

When regressing the salaries of professional men and women against, among other variables, length of service in ERDA's employment situation, the scatter of the data-points was not due to 'random deviations' from their 'true salaries' as given by the values of the regression equation, but due to concrete, traceable reasons that could be established in each case.[3] Despite the existence of an official federal pay scale for government employees, every individual departure from this pay scale could be traced and explained, not the least of which was 'sex-discrimination.' The values of the regression line, by the way, reflected like an average, the numerous individual circumstances of each employee rather than the official pay scale. Further de-aggregation by department within each agency led to ever different regression equations at each new level of de-aggregation.

As previously mentioned, the phenomena in society, its demography, economy and sociology should be expected to change, often substantially, from period to period, and from one region to another. This is an important characteristic by which the phenomena and data in the social sciences differ from those in the natural sciences. This scenario is very different from that implicit in regression analysis that was introduced to the social sciences from the bio-sciences.[4]

The least-squares regression function, linear or non-linear, arithmetic or logarithmic, simple or multiple, in one or two-steps, regular or step-wise, is assumed to reveal the underlying relationship between variables. Every individual observation that depart from this regression line or regression surface, which is believed to represent the underlying, common relationship or force, are considered deviations, errors due to influences that are estimated to behave like random deviations. When it is believed that differences in the time of year or differences in the regional districts account for some of the dispersion, helping to explain those deviations, instead of using dummy or binary variables for the time periods, e.g. trimesters and regions. I found it advantageous to recognize the short-lived and regional nature of the phenomena of economic and social statistics by carrying out regression analyses separately for each regional and time subgroup. In this manner interesting information can be discovered about the relationship of the investigated variables at different times, and in different districts. That allows a factual interpretation of the respective economic and social situation. Instead of using dummy variables for regions or time periods, the regression analysis should be carried out separately for each of these regions and time period.

9.3 On Precision in Regression with Aggregate Data

In socio-economic data these parameters cannot be determined as succinctly as the computer generated results might indicate. The reason is aggregation, apart from and independent of sampling. The blurring effects of aggregation can best be illustrated in regression between the time-aggregated data of time series although this effect is also present in regression between data that are aggregated by subject-matter and geographic area. These latter forms of aggregation have the same effect,[5] even though they cannot be demonstrated as readily.

Although aggregation is at the heart of statistics, and regression analysis is a most widely used statistical method, no attention has been paid to the 'loss of meaning through aggregation'. Statistics does not even have a proper terminology for the kind of loss caused by aggregation, although the effect of grouped data on correlation analysis has received some attention.[6] Correlation coefficients computed from aggregated data were higher than in the ungrouped data. Grouping obviously reduced the scatter around the regression line but was not believed to affect the regression parameters.[7]

As was discussed in Chap. 3, the of loss detail of the original 'statistical-counting-units' in aggregation leads to an indeterminateness of the results, causing socio-economic reality to be perceived only in a hazy manner. In regression analysis monthly, quarterly or yearly data are treated as if they existed simultaneously, regardless of the fact that they occurred in ordered sequence over time. Even when in a scatter-diagram the points representing the aggregates of a time series are marked with their date of occurrence, the regression algorithm ignores the time element, treating them as simultaneously existing.

Regressions between the aggregates of socio-economic time series are not exempted from the blurring effect of aggregation. When e.g. monthly data are totaled into yearly aggregates they lose the distinctness of having occurred during a specific part of the year. Shipments made during February can no longer be identified in a yearly total of shipments. They could as well have occurred in October. When the original statistical 'monthly containers' are discarded, their numerical content becomes merged with that of other months into a wider 'yearly container'. The only thing that is certain, made clear in the definition, is that these data can neither predate January 1 nor postdate December 31 of the yearly interval. This determines the time frame for the numerical values of both series that are to be used in a regression of time series of yearly aggregates Examples of such socio-economic aggregates are monthly, quarterly or yearly produced or exported quantities, prices and index numbers whose exact location within each time period is indeterminate except for the fact that they lie between the beginning and ending of each aggregation period. The longer that period, the farther apart are these limits, and the less can be known about the moment at which each 'statistical-counting-unit' actually occurred within the time aggregate. Regression analysis, on the other hand, customarily calculates its coefficients with any number of decimals and proceeds as if the value of each X and Y data pair were a sharply profiled, point-like numeric value, akin to an individual, single measurement in the natural sciences that would be represented as a point on a regression chart.

Consider the two series 'Capital Expenditures of all **US** **M**ulti **Na**tional **C**orporations (USMNCs), including the Parents and all **M**ajority-**O**wned **F**oreign-**A**ffiliates (MOFAs),' referred to as S1 (Series 1), and 'Value Added by all USMNCs, Parents and all MOFAs' referred to as S2 (Series 2), both in Table 9.1. Then consider the data pair S1, $X_{10} = \$436{,}405$ and S2, $Y_{10} = \$2{,}688{,}123$ in the linear regression equation. $\hat{S}2 = \$501,661.44 + \$4.21(S1)$. Both, S1 and S2 cover the time-interval of the year 2003. There is an unspecified large number of time-points in both years that could be paired up. These points cannot be further apart than 365 days. In such instances it is implicitly and tacitly assumed that the aggregated values of both series have occurred in the center of the year, July 1 (Fig. 9.2, points e, but also points a and c). Although the regression equation between S1 and S2 can be determined with any desired level of precision, the indeterminateness of these yearly aggregates, however, precludes a precise determination of that relationship. With the help of two assumptions the possible, limiting extreme boundaries for a regression relationship can be determined. First, assume that the yearly values of S1 have occurred at the end of each year (on December 31 instead of July 1), while the values of the other series S2, are assumed to have occurred at the beginning of each year (point d). The formerly synchronous aggregates S1 and S2 of any period (year) are now treated as if they were displaced against each other by 365 days. The values of S2 are lagged with regard to the values of S1, point d in Fig. 9.2, and panel B of Table 9.2. This amounts to regressing the numerical values of S1 on the numerical values of the previous year for S2 The other assumption reverses the order, considering the figures of S1 as having occurred at the beginning of each year, and those of S2 at the end. (point b, Fig. 9.2). S2 for each year is paired up

Table 9.1 US multinational corporations (USMNCs). Parents and all majority-owned foreign-affiliates (MOFAs)

#	Yea	Capital expenditure S1	Value added S2
1	1994	303,364	1,717,488
2	1995	323,616	1,831,046
3	1996	340,510	1,978,948
4	1997	398,037	2,094,318
5	1998	411,155	2,100,773
6	1999	453,032	2,480,739
7	2000	506,950	2,748,106
8	2001	524,215	2,478,056
9	2002	443,388	2,460,411
10	2003	436,405	2,688,123

Source: Survey of Current Business, July 2005, Vol. 85 / #7 p. 10, in millions of current

with the previous year's numerical value of S1, as LagS1 viz. the aggregate of S2 for 2003 is paired up with the value of S1 for 2002. These non-lagged X-values and lagged Y-values give the other limiting regression equation (Table 9.2 B). The resulting bordering regression lines $Y_{(LagS1,S2)}$ and $Y_{(S1,LagS2)}$ enclose the zone which contains the regression relationship. Figure 9.3 shows the uncertainty of the regression relationship caused by using yearly rather than monthly or weekly figures. Graphically it should be represented by a zone of varying intensities of gray, barely visible and fading out at the lower border, the 'floor' and the upper border, the 'ceiling' of that zone. The most intensive color is in the middle of that zone.

This margin of uncertainty is always present when aggregate data are involved. It is relatively simple to visualize the effects of aggregation in data aggregated according to their time dimension. Something analogous is happening with aggrega-

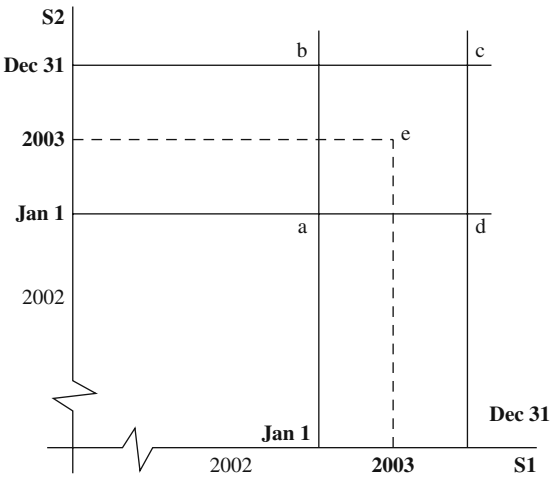

Fig. 9.2 The possible time limits for S1 and S2

Table 9.2 Lagged regressions of S1 and S2 of Table 9.1

A		**B**	
Lag S1	S2	S1	Lag S2
303,364	1,831,046	323,616	1,717,488
323,616	1,978,948	340,510	1,831,046
340,510	2,094,318	398,037	1,978,948
398,037	2,100,773	411,155	2,094,318
411,155	2,480,739	483,032	2,100,773
483,032	2,748,106	506,956	2,480,739
506,950	2,478,056	524,215	2,748,106
524,215	2,460,411	443,388	2,478,056
443,388	2,688,123	436,406	2,460,411
$\text{Lag } \hat{S}2 = 349,861.7 + 4.328876628(S1)$		$\hat{S}2 = 932,886.0254 + 3.33768833(\text{LagS1})$	

tion by geographic area, and by aggregating subject-mater categories. Aggregation of data usually happens in all three dimensions at the same time. This blurring effect, consistently ignored by economists and econometricians, exists in addition to the sample-related uncertainties, recognized as the 'confidence interval of the regression line,'[8] which has a different meaning. When regression is used as a forecasting tool, the aggregation margin of the regression line must be recognized in addition to the uncertainties discussed in Chap. 6 on forecasting.

The regression line **ought to be** understood as a **barely visible, hazy zone** with contours that gradually fade out on its upper and lower edges. The regression must **not** be drawn as a heavy line passing through the barely visible data points, but to the contrary, these data-points are to be plotted as strongly marked points in

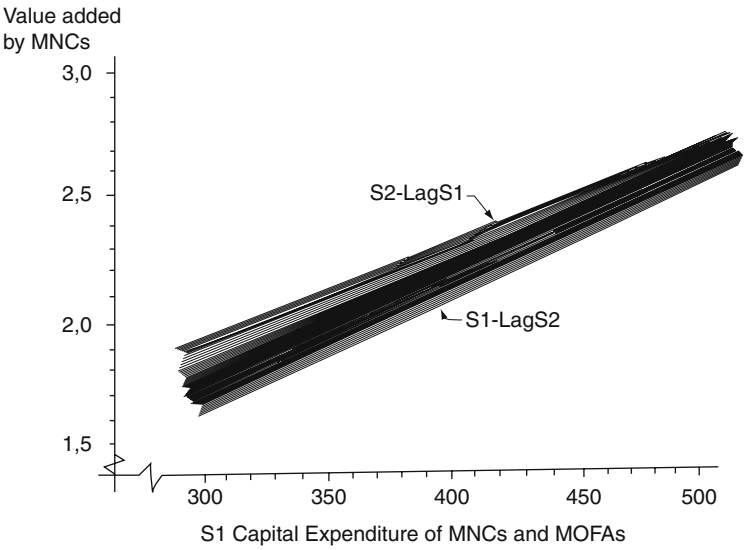

Fig. 9.3 Regression of aggregate data S1 and S2 and the Margin of uncertainty

the graph because they are closer to reality than the regression line (not shown in Fig. 9.3). This is an application of what was pointed out in Sect. 3.4 about visualizing aggregates of different 'sizes.' Inference from regression in the sample to the corresponding population regression function is not the dominant aspect when interpreting multivariate cross-sectional relationships of aggregates, particularly when the data represent all cases in a district or of an industry.

9.4 Different Forms of Data Association in the Social Sciences

The assumption that every phenomenon is caused by some natural laws comes from the bio-sciences. That law is reflected in a clear relationships, that is inevitably affected by random measurement errors that follow a Gaussian distribution. In the social sciences such clear-cut relationships of the linear or non-linear type are the rare exception. The different nature of the phenomena in the social sciences and the aggregate nature of most data in economics and business are causing violations of the standard model of measuring relationships. Good alternatives do not seem to be available. Statistics, of course, is not alone in this dilemma. Other social sciences face a similar situation.[9]

It would be reasonable to develop methods that are more appropriate to the concrete situations as they actually exist in our field. We should keep in mind that today's general least squares linear regression model originated in Galton's and Pearson's solutions to specific problems in human biology. We now approach every situation with this procedure, uncritically accepting those assumptions as if they were the realities of our data. Here are vast unmet statistical needs to be explored by social scientists.

The uncertainties surrounding the frequent regression analyses between time series data, discussed in the previous section, are a special case of the general relationship between aggregated socio-economic variables. Typically, these regressions are characterized by a wide and often widening scatter of the data points (heteroscedasticity), and low R-values. The departures of social science data from the basic model that underlies today's accepted regression analysis are not to be taken lightly.

Instead of considering the regression line (or regression surface) as of primary importance, while paying little attention to the individual values, one ought to proceed in the opposite direction. The individual, scattered data points are to be taken as the primarily given, concrete facts from which one ought to proceed, encasing between boundary lines the area which appears to contain the relationship. The original data points are the primary facts not to be downgraded to mere 'deviations' from the computed least squares regression values $\hat{Y}_x$. These, based on all data points, have the same lack of 'concreteness' as the corresponding aggregate of these data.

When investigating the relationship between the grades received by my students in two of my first economic statistics courses and the percentage of classes they missed, I noticed that the scatter-diagram of these data roughly resembled a rectangular triangle, formed by the X and Y axes, and by a line sloping down

from upper left – high grades and low absence – to the lower right – more frequent absences and lower grades – creating the impression of a triangle[10] (Fig. 9.4). Despite an evident relationship between absence from class and semester grade, the coefficient of correlation, using linear regression, was $r = -.18024$. Interpreting this $r^2 = .03249$ seemed to indicate that knowing the % of absence from classes barely helps to explain – 3% of the scatter of grades received in those courses. Although the linear regression line $\hat{Y} = 18.625 - 0.1854x$ is in the right direction, the strong reverse heteroscedasticity at the low values of X, tapering off toward the higher values of x, – heteroscedasticity usually unfolds in a reverse manner: in the dependent variable that corresponds to the higher values of the independent variable X. Although this grade distribution seemed to make sense, it did not conform to a trend around which the data would be scattered evenly in the shape of a normal distribution. Teachers of quantitative subjects are well aware, that students' regular class attendance is strongly related to their success in the course, yet good attendance does not guarantee a good grade. Missing many classes, however, makes it increasingly more difficult, even for bright students, to succeed in such a course. Depending on how such a course is conducted, – in those courses class attendance and taking notes were indispensable conditions because textbooks were not avail-

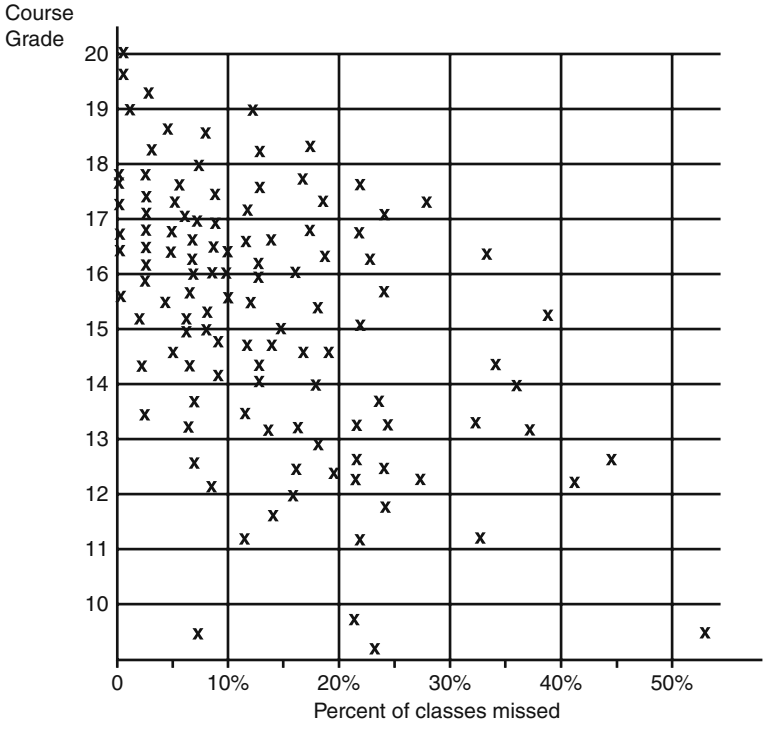

Fig. 9.4 (*Down*) Student course grades, (*across*) % Absence from classes, Universidad Central de Venezuela, Facultad Economica

able. Yet there was more involved to achieve academic success. Good attendance alone did not guarantee a good grade. Participation in those classes was a 'conditio sine qua non.' Neither linear nor non-linear regression models do justice to such a situation.

Later, while exploring the earlier mentioned manpower data of ERDA, I repeatedly encountered scatter-diagrams that resembled triangles. The computer-printed scatter-diagram of the 'salary and length of employment at the agency,' shows a different triangular shape, Fig. 9.5.[11] Figure 9.6 shows a different triangular shape in the scatter-plot of length of time of employees in that large workforce at ERDA and time worked before joining ERDA for another agency of the Federal Government.

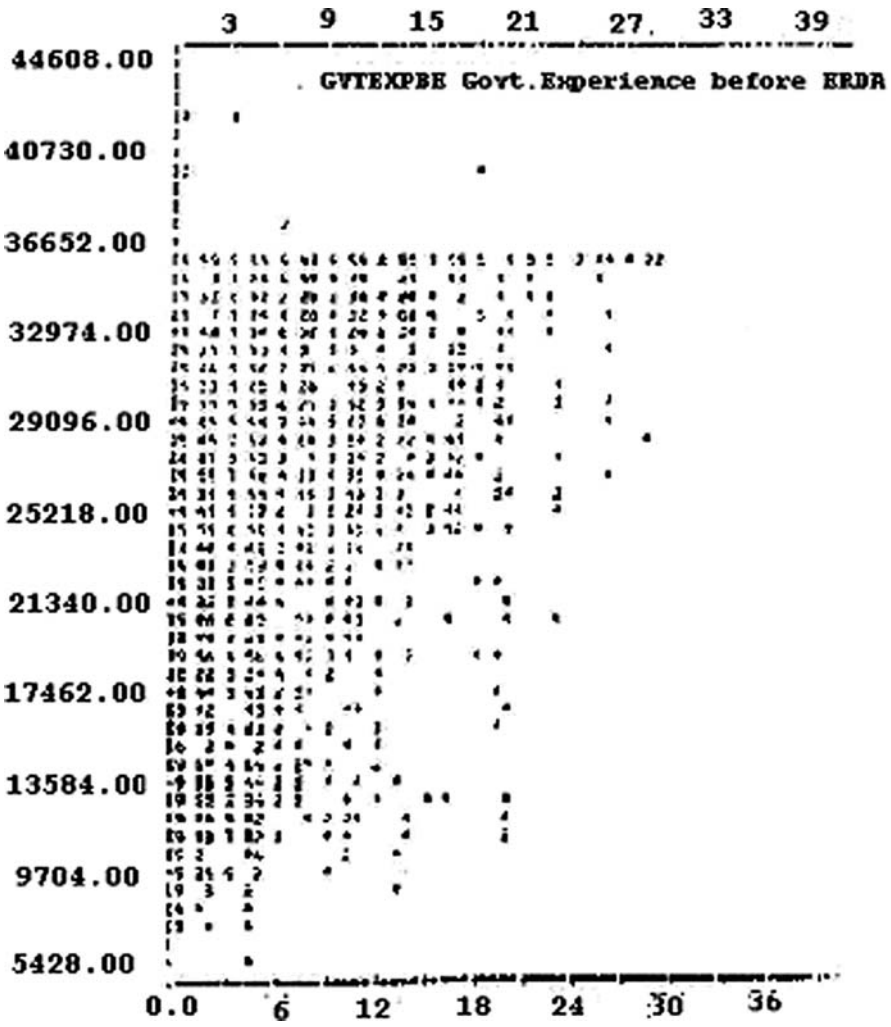

Fig. 9.5 (*Down*) Salary at ERDA, (*across*) Length of employment at ERDA

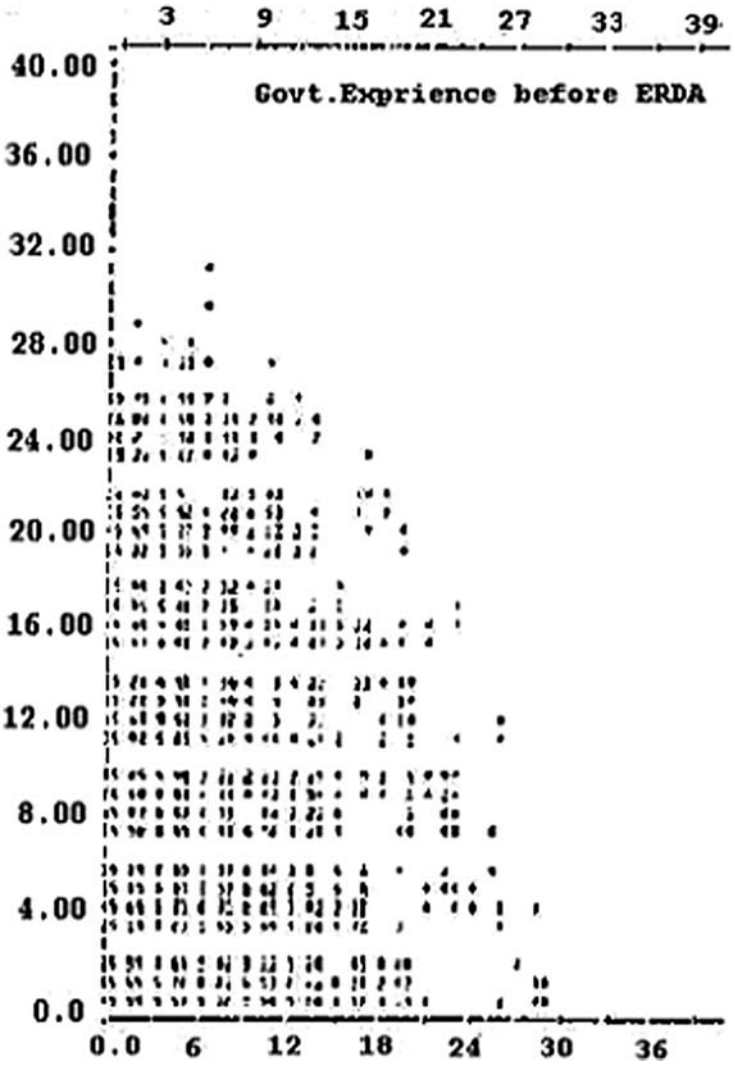

Fig. 9.6 (*Down*) Length of employment with ERDA; (*across*) Length of previous employment with the Federal Government

A different view of this situation is given in Fig. 9.7 in which the standardized residuals are plotted (using the SPSS program 'Regression') against the standardized predicted regression values of such a unique, triangle-shaped regression relationship. Figure 9.8 gives evidence of the existence of such triangular relationships also in the sciences.[12] This type of regression relationship turned out to be more frequent than expected. Transformations are of no help because the underlying functional relationship is of a different kind. The independent variables do not evoke a unique response in the dependent variable, but serve as the necessary condition for an array

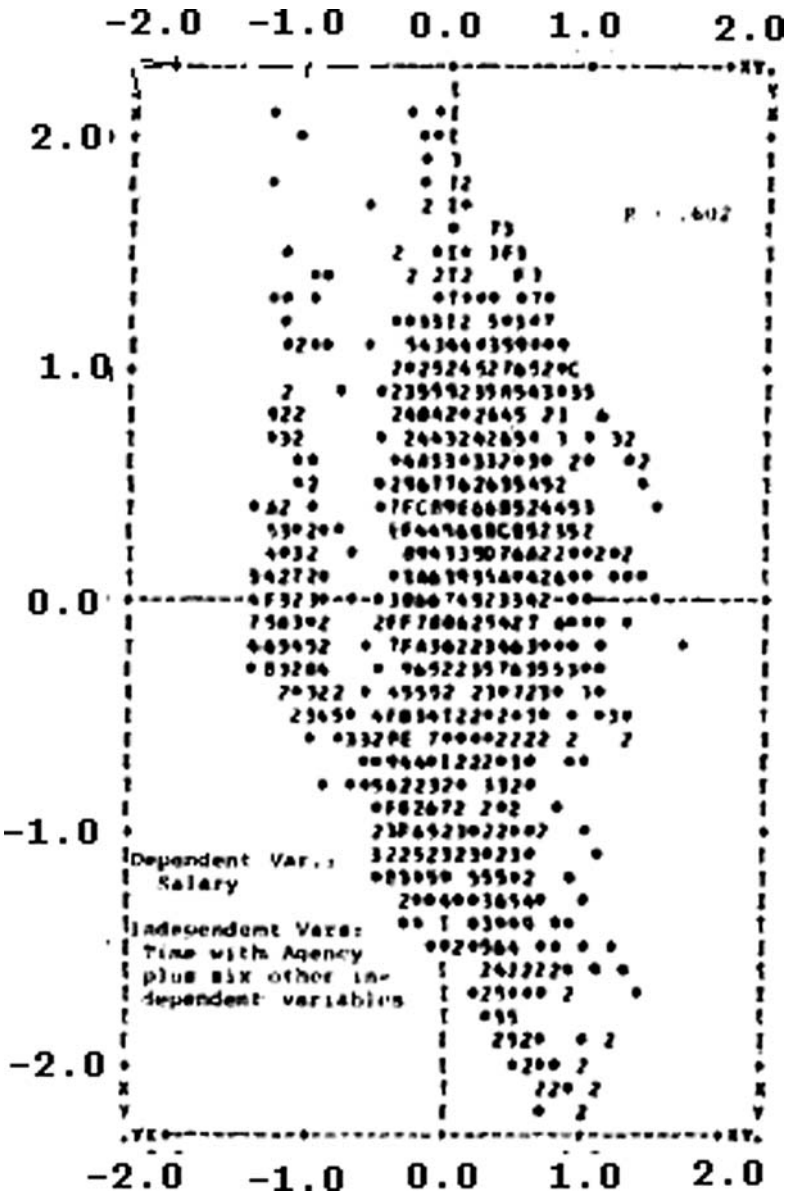

Fig. 9.7 (*Down*) Standardized deviations of salary at ERDA, (*across*) standardized length of employment and six other independent variables

of possible responses within certain boundaries that can be determined. The literature on regression and correlation[13] does not seem to have paid attention to this phenomenon. What looked like exceptional and unique data materials that defied the standard regression approach turned out to be more frequent than expected.

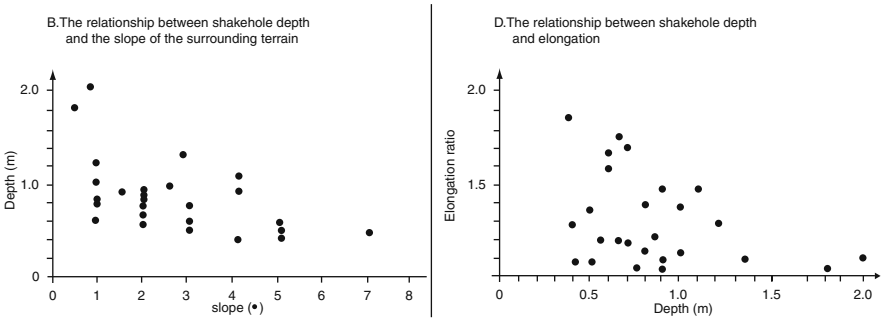

Fig. 9.8 Size, depth and slope of shake-holes

A different approach was called for that does not insist on discovering single-valued natural laws among the fuzzy phenomena of the social sciences. That approach can be generalized and extended to a large number of applications. Even though regular least squares regression analysis may be tried in these instances, the data clearly call for a different, more appropriate model for more complex relationships between variables than the model of the unique one-on-one type of the standard regression and correlation analysis. Obviously some kind of relationship does exist, although it is not of the single-valued variety, and a way should be found to assess this kind of association. The customary data transformations[14] such as taking the logarithms of the Y-values, the inverse, square or cube of the X-values does not remedy the situation[15] nor make it more suitable for an application of the standard least squares regression.[16]

9.5 Suggestions for an Alternative Model

The variables in the data in Figs. 9.3, 9.4, 9.5, 9.6 and 9.7 obviously did not display a one-to-one relation. In an earlier publication I discussed a different conception of measuring relationships in social science data, titled 'triangular correlation'[17] Data are not to be considered as deviating from the central values of a regression line, but without having such a center, they are loosely held together by limiting conditions which may lead to a distribution of the data in the shape of triangle or other, similar geometric shapes. If the data are not enclosed by the two axes and a third line but are more of the kind of a heteroscedastic sprawling distribution, the two bordering lines that enclose such data can be viewed as a regression line split in two, the lower half and the upper half. These two lines of enclosure are then to be fitted, e.g. by the least squares procedure, to all the bordering data, the ceiling line forming a 'ceiling' to be fitted to every largest data-point corresponding to a given X. The other bordering line, as a 'floor' is to be fitted to the lowest values for each X. In the situations presented in the figures of this chapter the customary regression line or regression surface is of little value and cannot be meaningfully interpreted.

Let me therefore propose a model based on a less sharply defined relationship between the X and Y variables, that can be stated with a certain latitude, between a lower limit of Y for a given X, $Y'_{L,x}$ and an upper limit $Y'_{u,x.}$ Both must make sense for the topic under study, and be meaningful in terms of the socio-economic situation. Neither of these limits need be linear. Most Y-values are assumed to be indifferently located inside those limits. Ideally they are evenly placed between the lower and upper bounds. Assuming the bivariate case, of a dependent variable Y and only one independent variable X, a 'floor' is determined, e.g. by least-squares lines fitted to the lowest Y-values and a 'ceiling' to the highest Y-values for each X-value respectively. All values could be treated as deviations from equally, that is, indifferently spaced computed $Y'_{i,x}$-values that are to be determined for each X-value. The sum of the square of the nearest distances of all Y_x-values from these $Y'_{i,x}$ values is determined for each X-value, and a measure of closeness of the data to this less stringent model developed in analogy to the coefficient of simple correlation. The proposed procedure is one possible alternative for the numerous instances in the social sciences where the clear, unique and homoscedastic trend of a relationship, postulated by the current correlation model, is not appropriate.

To repeat, the proposed model of an X – Y relationship has a 'floor' and a 'ceiling' and equally likely, evenly spaced points $Y'_{i,x}$ in between. If e.g. there are five Y-values that belong to a specific X-value, then two of the equidistant computed $Y'_{i,x}$-values will be located on the bordering equations, and three will be equally distant between these points on the borders. The measure determines how close the data points $Y_{i,x}$ come to this model of an equal distribution between the two staked-out limits. In this model the data points $Y_{i,x}$ are not considered as 'errors' or 'deviations.' Neither normality of distribution, nor homoscedasticity of the data is postulated. For each X as many $Y'_{i,x}$-values are expected as there are actual $Y_{i,x}$ data points corresponding to a given X, instead of only one such $\hat{Y}_x$ as in the usual regression model. The discrepancy of the observed $Y_{i,x}$ values from the computed equally spaced points, $Y'_{i,x}$ can be determined, squared and used for a measure of association – or lack or it. Appendix D will show how the locations of these hypothesized equally distant points between floor and ceiling are determined.

Then the distance of each observed $Y_{i,x}$ from its hypothesized, nearest $Y'_{i,x}$ is determined for a given X, and squared. Like in regular regression, the double $\Sigma\Sigma$ $[Y_{i,x} - Y'_{i,x}]^2$ is an indication of irregular distribution inside the area of the 'triangle.' The first summation is for all $[Y - Y'_{i,x}]^2$ of a given X, the second summation adds these discrepancies of observed from computed values over all the X values. There are n_x points $Y'_{i,x}$ for each one of the k X-values, and nk different $Y'_{i,x}$ values. This measure of discrepancy is not directly comparable to ordinary least squares curve fitting. In regular regression the parameters of the regression line and the values of $\hat{Y}_{i,x}$ are determined from all the observed X and Y values. In the proposed measure of this one-to-many relationship the two equivalents of a single regression line, the 'floor' and 'ceiling' are determined only from the lowest and highest Y-values as shown in Appendix D. Y-values falling outside the boundaries, above or below, are treated like the other Y-values that fall within these boundaries. Then this coefficient

of triangular correlation, $R_{\blacktriangledown}$ proceeds like the regular coefficient of correlation 'r' and can be interpreted analogously. In the case of complete conformity of the data with the stipulations of this model, the double sum of the discrepancies, connecting vertically each $Y_{i,x}$ with its nearest $Y'_{i,x}$ becomes zero and $R^2_{\blacktriangledown} = 1$. Every deviation of an observed Y-value from its expected $Y'_{i,x}$-value reduces the credibility of the model assumptions for the given set of data. The difference with regular correlation is that the residual sum of squares is computed from many $Y'_{i,x}$-values instead of from only one $\hat{Y}_{i,x}$for each X. The same holds for the coefficient of triangular determination, $R^2_{\blacktriangledown}$. The proposed method, by the way, is not limited to pronounced triangular situations.

Interpreting the results of the sample used to demonstrate the calculations, in Appendix D, one may question whether it is preferable to conclude that only 9% of the salaries in that agency are related to the length of service of its employees – an admission of 'no relationship' of the kind that one would expect to find in the natural sciences – or to conclude that 84% of those salaries are related to length of service in the described not very specific relationship between between a lower and an upper limit. The first conclusion obviously does not take into account the known fact of the yearly salary increases that accompany within-grade (within-step) and between-grade promotions and adjustments. The second recognizes that such salary differentials exist at all salary levels but can be perceived at the given high level of aggregation with certainty only at the lower fringe and at the federal pay ceiling as the other fact in this relationship. The vagueness of the interaction between these two variables can only be reduced by bringing additional information to the analysis. This can be done by creating smaller, more defined sub-aggregates, e.g. by type of occupation, position, attained education, geographic location of the sub-agencies, and in this case, not least of all, the gender of employees. At this high level of aggregation the proposed less strictly formulated 'triangular regression' model highlights more correctly the existence of some relationship, than does the customary correlation analysis, that was designed for other kids of phenomena like those in the natural sciences. The latter leads to the conclusion, that no relationship exists between salary and length of service, certainly not a relationship comparable to a natural law in the bio-sciences.

This alternative model of 'one-to-many' relationships can be helpful in all social science data with indications of heteroscedasticity, or simply wide scatter. The lower and upper regression boundaries should be determined in the manner proposed here, splitting in two, as it were, the single regression line, and thereby extending the concept of an 'aggregation zone' that was found to exist in most linear and non linear regressions and correlations. The proposed $R^2_{\blacktriangledown}$ can be computed as a meaningful alternative in addition to the usual R^2 of the standard approach.

On further thought, the proposed approach to cross-section analysis, marking off the space within which a relationship appears to exist, can be considered to be the general case. When 'floor' and 'ceiling' of that space converge, collapsing into only one single line, then the well known regression and correlation formulas become the limiting case of the more general approach outlined in this section. In the socio-economic field the case of the 'collapsed floor and ceiling line' type relation

will be the exception rather than the rule. This outline of a 'one-to-many' relation between variables could be one of the new tools statistical theory needs to develop to do justice to the particularities of economic and social data.

Notes

1. Chewning vs. Seamans, U.S. District Court, District of Maryland, case # 760334, 1976. The agency of the Federal Government, ERDA (ERDA = Energy Resources Development Agency of the federal government, the forerunner of DOE = US Department Of Energy) was sued by a group of 173 professional women employees. Although not part of this group of plaintiffs, a large number of women belonged to the class that would be affected by the outcome of that lawsuit. That study explored 8,164 employee records with 65 characteristics each. Some of the highlights of this exploration are discussed in my paper : "Discrimination Against Women, A Multivariate Analysis" *Proceedings of the Social Statistics Section of ASA,* Washington, D.C. 1979, pp. 354–359, especially the tables of regression coefficients, pp. 357–359.
2. Of the18 records of the male librarians only 12 were used in this regression analysis. In six records one or more of its variables were missing. Of the 18 women librarians only 14 records were used omitting the four records with variables missing.
3. This discovery was made during the detailed study of the employment records of professionals in ERDA (**E**nergy **R**esource **D**evelopment **A**gency), of the Federal government in 1976, described in: Winkler, Othmar, "Discrimination against Women – A Case Study in Multivariate Analysis" *Proceedings of the Social Statistics Section of the American Statistical Association (ASA),* Washington, D.C. 1978, pp. 354–359 and Chewning vs. Seamans, U.S. District Court, District of Maryland, case # 760334, 1976.

   The class of female professional employees, including these librarians, alleged that they had been discriminated that is, disadvantaged in regard to promotions and salaries advances compared to their male counterparts in that agency. I singled out this particular set of professionals because it was one of the few groups in that large federal agency in which the number of women equaled the number of men, with men and women having comparable qualifications. It seemed best suited to explore if women really were treated differently.

   This study was also a perfect example of the role of statistics in the social sciences. Statistics makes it possible to perceive phenomena in our society, such as inflation, discrimination, poverty, etc. This perception or description of a social or economic phenomenon is its basic function. Inference from samples is often necessary. When the inference-phase has been concluded, however, when confidence and significance statements have been made to account for the effects of randomness in the selection of the sampling-units, then the research turns to the basic issues of the field that it had set out to study in the first place. Description , as it is condescendingly referred to, is the raison d'etre of data collection and analysis, regardless of how sophisticated the former. The current dichotomy into 'descriptive statistics' and 'statistical theory' properly speaking, fails to recognize the descriptive nature of socio-economic statistics. This study of sex-discrimination at ERDA is typical of the contribution statistics can make to the social sciences: it served to reveal the social reality of that elusive phenomenon 'Discrimination by Gender' in a government agency. It was crucial in deciding this litigation, yet this study did not rely in any way on statistical inference.
4. One should remember that the word "regression" itself has its origin in human biology. It was coined by Pearson when he discovered that the body sizes of sons "regressed" to those of their fathers.
5. Winkler, Othmar W. "Aggregation and Aggregates In Economic- and Social Statistics" 55th Session of ISI, Sydney, Australia – In this paper I explored the numeric effects of changes in the proportion of its five sub-price indexes on the value of the national price index. The results are reproduced in Appendix D of this book.

6. C.E. Gehl, Katherine Biehl, "Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material," *JASA* Vol. 29–185A, New Series No. 185A, pp. 169–170. Jan K. Wisniewski, "Pitfalls in the Computation of the Correlation Ratio" *JASA* 29-188, December 1934, pp. 416–417.
7. J.S. Cramer "Efficient Grouping, Regression and Correlation in Engel Curve Analysis" *JASA* 59–305, March 1964, pp. 233–250.
8. Regression analysis between time series usually should be treated as regression between populations, even if there are only a few yearly, quarterly or monthly values in that time series available.
9. Consider Moral Theology. This social science has developed an abstract model of man from which a theory of proper moral behavior has been derived, which often does not correspond to the real behavior of human beings. Pastoral Theology had to be developed as another discipline to bridge this gap. Quoting from a recent study: ". . .homosexual acts are considered intrinsically immoral according to Moral Theology. When ministering pastorally to homosexual persons, however,. . .primary concern is to help them live as stable a Christian life as possible in their particular situations. . . one can recommend them to seek such a homosexual partnership and one accepts this relationship as the best they can do in their present situation. . ." D.Doherty, ed. Dimensions of Human Sexuality, Doubleday, 1979, pp. 218–219.
10. The situation was aggravated by the fact that no textbook in Spanish for economic statistics was available and I had not yet written my own notes to be distributed in that course. Students had to take notes on their own of the lectures during class. The grading system was between 1 and 20 points. Students who received less than 11 points failed the course. . . Othmar W. Winkler "La Vida de la Facultad en Cifras" *Boletín Informativo No. 4 de la Facultad de Sciencias Económicas y Sociales,* Universidad Central de Venezuela, Caracas, Venezuela, 1950, pp. 123–152, esp. p. 140.
11. Individual values are marked on this scatter plot produced by SPSS with *. More than one person coinciding on that same point – receiving that salary with that length of service – are marked by the numbers 2 up to 9. When more than 9 persons coincid on one value on the scatter- diagram, letters are used. A is used for 10 persons, B for 11, etc.
12. Both panels of this figure are taken from "Shake holes, A morphometric Field Project for Sixth-Form Geographers," *Geography,* Vol. 65 pt. 3, July 1980, No. 288. and are included here to show the ubiquity of such a 'Triangular Correlation' even in the natural sciences where one would expect to find the customary one-on-one relationships.
13. I explored this topic under the narrower title:. "On Triangular Correlation" *Proceedings of the Business and Economics Statistics Section of ASA*, Washington, D.C. 1980, pp. 606–611.
14. Norman Draper, William G. Hunter "Transformations: Some Examples Revisited" *Technometrics* Vol. 11, No. 1, Feb. 1969, pp. 23–40.
15. See e.g. John Neter, William Wasserman Applied Linear Statistical Models, Richard D. Irwin, Inc. 1974 Homewood, Ill. pp. 121–139.
16. See op.cit. Winkler, O.W. (1980), Figs. 7, 8 and 9 on p. 611.
17. Winkler, Othmar W. "On Triangular Correlation" 'Proceedings of the Business and Economic Statistics Section of ASA', Washington, D.C. 1980, pp. 606–611.

Chapter 10
Socio-Economic Statistics and Probability

*Leicht bei einander wohnen die Gedanken doch hart im Raume stossen sich die Sachen**

10.1 Introduction – The Big Picture[1]

So, you are a statistician? Then you could help me win at blackjack (a popular card game involving probability). Occasionally I find myself asked such a question at social gatherings, a reaction to the mention of statistics that takes a distant second to the reaction described at the beginning of this book. It reveals the identification of 'statistics' with the calculus of probability and mathematics in the minds of educated people. This should not come as a surprise, as it stems from the growing persuasion that statistical theory is identical with inference and the mathematics of probability. Social data, besides being treated as if they were 'measurements' in the natural sciences, are treated as the outcome of random experiments or as random samples even when they are clearly populations or unique, historic events that were not a planned experiment.[2] Testing the statistical significance of characteristics of data that obviously are populations, or to analyze them with statistical methods based on inference and on the concept of random sampling, is pseudoscience. Uncritical adoption of the stochastic view of socio-economic reality and the indiscriminate use of terms like 'random variable,' 'random deviation,' and 'random error' are compounding these misconceptions and misuses. The 'what, when and where' of probability in the data relevant to a particular situation, ought to be analyzed much more carefully before the high-powered tools of statistical inference are employed.

Such misuses of probability came to my attention during various sex-discrimination class-action lawsuits, other than the one described in Chap. 9, in which I was also involved as an expert witness. There the correct understanding of probability was crucial, as the outcomes of these lawsuit depended on probabilistic arguments. In these cases, the perpetrators of misuses were the statistical consultants, and the judges who accepted their opinions. The ultimate responsibility for these misuses, however, must be traced to the academic enterprise. First in line

*freely translated: 'It is easy to theorize, but reality is something else, it is another matter' Wallensteins Tod, 2nd Act, 2nd Scene, by Friedrich Schiller This quotation could as well have been placed under the very first title of this book.

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_10, © Springer-Verlag Berlin Heidelberg 2009

are the authors of textbooks on business, economic and social statistics, and the publishers of these textbooks who encourage these authors to write mainstream textbooks to assure acceptance in the college market. They all have contributed to reinforce the trend, over the last decades, of moving probability into an ever-more prominent position. These textbooks, that disseminate and perpetuate these misconceptions, contribute to shape the mindset of future business managers, economists, social scientists, in short, all who are going to use statistical data. Responsibility also falls on the editors of statistical journals, who influence the direction of the field by their selective choice of manuscripts. This trend is reinforced and complemented by the hiring practices of business schools, social science and economics departments that require a background in mathematical statistics as proof of qualification as a teachers of their statistics courses. Given the central position of probabilistic reasoning in today's statistics textbooks and literature, a careful, critical scrutiny of its proper and improper use is overdue.

10.2 A Historic Perspective on Probability in the Social Sciences

A broad-brush historic overview may be helpful to place into proper perspective the way in which probability theory came to dominate statistics in the social sciences.

The early statistical congresses and literature dealt mainly with dialogues between the producers of official economic and social statistics about defining economic and social phenomena. Later, economists and policy makers began to show interest in these data as users. Early in the 20th century, the literature became dominated by discussions on 'price-index-numbers', 'quantity-index-numbers' and time series. Probability was not yet an issue. In the 1930s and 1940s the focus turned to sampling, which coincided and supported the emergence of econometrics, opening the field to probability considerations[3] Economists hoped to study their field with more rigorous, mathematically and statistically sophisticated procedures, particularly multivariate analysis. The rise of econometrics also opened new vistas for marketing, business and economic statistics. The idealized model of statistical data in these fields was that of measurements in the natural sciences with the attendant errors of measurement for which the calculus of probability seemed appropriate.

Another impulse to the application of probability[4] came when K. Pearson and especially R. Fisher introduced the mathematics of designing agricultural and biological experiments. Swedish statisticians introduced the concepts of the infinite population and of the super-population, with regard to which every set of data could be considered a random sample. Thanks to these scientists, statisticians, and the American penchant for formal algorithms, general statistical theory and also business- and economic statistics worldwide, have become ever more closely associated with the calculus of probability.[5] In today's theory of business and economic statistics hardly any other theoretical considerations seem to exist.[6] The presence of a random error component in all socio-economic models is taken for granted,

with little concern for the exact meaning of such error terms. In fact, statisticians today hardly consider the non-probabilistic, subject-matter-oriented-analysis of socio-economic data as their concern.[7]

These developments in business and economic statistics happened mostly in countries of the western hemisphere, whose economies depend on the free market system with the concomitant uncertainty for individual participants. In the USA, these developments have led to the current state where business and economic statisticians have come to view every situation as a random process or a random experiment, regardless of whether randomization really was involved. Some authors approach statistical data as if they were samples regardless of the manner in which the data were collected. Some textbooks exclusively use the terms 'random variables' and 'random deviations,' without justifying the term 'random'.[8] Obviously there is a connection between the perceived uncertainty in these market economies and the readiness of those who study it, to proceed as if the world around them really were such a random world.[9]

In the former East-block countries, on the other hand, governmental planning and control may have fostered a sense of frustration, but not a comparable sense of uncertainty. Statisticians in those countries have not witnessed a similar fusion of applied economic and administrative statistics with the theory of probability.[10]

10.3 What Really Is Probability in the Social Sciences?

The calculus of probability is believed to be the foundation of statistical theory. In fact, statisticians have come to consider the non-probabilistic analysis of socio-economic data as lying outside their discipline.[11] Once the true nature of socio-economic phenomena and of the statistical numbers dealing with them, are understood as unique, historic facts, it becomes obvious that probability – and methods based on them – do not play the central role attributed to them by mathematically-oriented statisticians. Many routine applications of probability-based inference in economic and social data are neither warranted nor contribute to the interpretation of the socio-economic forces reflected in the data. To clarify the circumstances under which the use of probability is justified also helps to clarify the nature of probability.

The statement 'a probability is the ratio between the number of expected outcomes and the total number of possible outcomes' begs the issue. It does not indicate what distinguishes that ratio, called a probability, from the countless other ratios in demography, finance, economics or sociology that are not considered to be probabilities. What is that special quality that makes the ratio between two numbers a probability? The answer lies in the stochastic, random or chance nature of the process that creates the data from which such a ratio, the probability, is calculated. But what is random or chance? Where in management, in the economy, in a society do genuinely stochastic processes occur? Where does random in the sense of the oft-invoked games of chance take place in the life of modern society? When does it occur?

The concepts of stochastic, random or chance are different from a privately held 'Weltanschauung,' e.g. that events in a person's life do not seem to follow any discernible pattern nor have discernible causes, in other words, that things in his or her life just happen at 'random.' Even if one may privately hold such a view of one's own life, it is not legitimate to transfer it unexamined to socio-economic processes, and to statistical data about these. The relationship of the life of such an individual to the macro phenomena of society can be thought of somewhat like the relationship of a macro molecule – itself following its own rules of behavior, seemingly at random, like its atoms – to large combinations of such molecules in living matter, which are neither controlled by the same rules as its atoms, nor by the non-stochastic laws that guide its molecules.

Is an earthquake such a chance event? Or a fire in a factory? Or the discovery of oil on a location on land or at sea? Such real-life events are given as examples that are allegedly governed by chance.[12] Behind each one of these events, there are processes or causal systems at work, which, in principle, can be known.[13] This is true even if at this time these processes may not yet be fully understood, or may not even have been identified as such. Hence these processes ought to be considered as not yet adequately predictable at this time.[14] For if they should be discovered to be truly probabilistic, researchers should no longer waste time and effort with such projects, nor should private, public or academic research facilities invest resources in them.

Man's empirical understanding of his environment has changed over the millennia like that of a child as it grows up. At first, a child has not grasped the relationship between a cause and a consequence. It pulls on the lamp cord and the lamp falls off the table. For the child this accident was 'chance,' a consequence with no awareness of a cause. The more highly developed that child's understanding of its surroundings, the more it will relate happenings with preceding actions, not necessarily his own. The child will push back, as it were, the frontier of such 'chance' events.[15] Analogously, there are domains in our adult world that formerly were not well understood but now are becoming clarified[16] and removed from the realm of fortuitous or chance events.[17]

Let another example illustrate what might be considered the essence of 'chance.' Imagine a person returning to his home. His house key is on a key ring with four different keys that are similarly designed. Under normal circumstances it is clear to that person which key belongs to which lock. Minor differences in the jagged ridges, length grooves and heads of those keys make it easy for the owner to distinguish between them. In daylight that person will select the appropriate key for the lock of his door as a perfectly deterministic process.

Imagine that same person returning after dark with only dim street-light illuminating the scene. Even in such a situation most people would have little difficulty selecting the proper key on first try.

Assume further a cold night, as an added aggravation, too cold to rely on tactile identification of the ridges of these keys. The selection of the appropriate key may begin to resemble a random selection process. The owner's chances of identifying the proper key on first try are roughly given by the number of available keys on that

key ring. They resemble now equally likely possibilities. The lack of light and the numbing temperature, depending on the severity of such aggravations, could convert this key-selection process, setting aside a tried but failed key, into a situation that resembles a process of 'sampling without replacement.'

As an additional aggravation, imagine that the owner also was under the influence of alcohol. That person may no longer keep a tried but not proper key from being tried again. Perhaps he may even be lucky if he can manage to insert the key in the keyhole. That situation has become a game of chance with equal probabilities for each key, as 'sampling with replacement,' one of those situations which statistical theory readily summons as if this were the prevailing case in individual economic or managerial situations, in fact, in the free market society at large.

Both examples were attempts to illustrate the relative nature of 'random' or 'chance' events.[18] It should be added that the socio-economic domain is neither subject to the strict laws of the physical world of nature, nor to the indeterminacy of subatomic particles. It is the domain of man-made laws and regulations of society, of social customs, the particularities of a specific culture, and the spontaneous actions and foibles of its citizens, that shape socio-economic statistical data but do not determine them as the laws of nature that underlie the processes and data in the sciences. As the reader may notice, I am siding with those who hold the first of professor J. Haught's list of different views of 'chance.' As to the mathematical view of 'chance,' I consider it confined to genuine 'random sampling' and inference based on such samples. In fact, probability is of little use for the interpretation of socio-economic statistical data, while another kind of uncertainty, that is created by aggregation, ought to be taken seriously.

Chance does not appear to be a characteristic that is objectively inherent in a process. Rather it resides in the observer's difficulty, inability or lack of interest to understand the deterministic causal system that actually is at work.[19] 'Randomness' or chance of a process – excluding behavior at the atomic or even molecular levels in the sciences – is not an ontological category that exists outside of and separate from causality, so to speak in parallel and in competition with it, but is one possible way to look at reality.[20] A stochastic approach to a deterministic situation may be convenient, but it must be kept in mind that true randomness in socio-economic statistics can be guaranteed only when it is specifically created by man with the help of gambling devices, tables of random numbers, or the computer-generated random numbers of some programs. Only then does the selection process of such a carefully-designed random sample, or the explicit randomization of a social experiment[21] deserve to be treated as a random process and random event. But true randomness, as probability theory seems to take for granted, does not occur of its own accord in the processes involved in social science data.[22] Probabilities are the quantified expressions of the randomness of a situation. They can be assessed either as 'objective' or 'subjective' probabilities, though ultimately both are determined by human judgment. Most important, however, is the assessment of the generating process itself as either truly random, or as not truly random. In the latter case, the ratios formed between their data cannot be considered to be genuine probabilities.[23] As

mentioned before, true probabilities have the external trappings of ordinary ratios. But ratios are true probabilities only when they are computed from data that result from a genuine random process. If this is not the case, then such ratios are falsely considered to be probabilities.

10.4 Typical Misuses of Probability

It is useful to recall the stringent rules that are required to guarantee that the sample selection is truly random. Such assurance of randomness is required, supported by the threat that every detail of that procedure is liable to be cross-examined in a court of law for an audit sample to certify the financial conditions of a business firm.[24] This means that the difficulties involved in making a process truly random, such as the selection of n elements from a population, cast doubt on the widely held belief that true randomness is easy to achieve, and can occur everywhere and anytime in the economy. Randomness is assumed to be an unquestioned fact. Proof of randomness is neither requested nor expected.

In the following, the numerous misuses of probability in socio-economic statistics are listed according to the degree to which they are obvious.

1. The most obvious misuse occurs when statistical inference is applied to data that without a doubt are populations,[25] a situation that obviously is not the result of a random process.[26] Those who perform hypothesis tests and inferential statements on populations, intending to confer scientific legitimacy on an investigation, seem to be unaware that such tests are a pretentious illusion, pseudo-science, and unaware that this exposes their lack of understanding of statistical inference.[27] This kind of misuse of probability is facilitated by the fact that computer programs provide the values of the t, z, F, chi-square and other probability distributions along with the requested statistical procedures, regardless of whether the data are random samples, non-random samples or populations.[28] This includes the practice of applying inferential reasoning to socio-economic time series of statistical populations.[29] It should not be an excuse that such practices have a long history reaching back to the early econometricians.[30] The proper way to redeem such statements would be to declare.[31] '**if these data were a random sample,**' thus distancing the person responsible for such a misuse of probability from being accused of ignorance or of intentionally misleading those who may use such a study.[32]

The concept of a super-population, an imaginary, hypothetical population from which any data at hand can be considered to have been selected by an unspecified random procedure, was invented to justify the abuses of statistical inference.[33] The super-population concept also implies a direct misuse of probability because the (population) data at hand are not selected by any valid random procedure from an existing data pool.[34]

2. Inferential procedures are applied to sample data that were selected by a process that does not produce valid random results.[35] Consider the 'purposive sample' selection of six regional offices of a Federal Government agency, out of

over 30 such regional offices, to 'discover' the extent of discriminatory practices. Plaintiff and defendant agreed upon this deliberate, particular, non-random selection of six workstations (sub-agencies) as representative clusters of the work force of that Federal Agency. Clearly, the conditions for the valid application of probabilistic inference were not given. Yet the statistician for the defendant – the US Government – insisted on using F, t and chi-square tests.[36] Statements like ". . .there is a 1% probability that the difference between salaries of men and women are due to chance. . ."[37] cannot have the same meaning as such a statement would have in the situation of a genuine random sample. In general, data are assumed to be random without proof of randomness in the data generating process. If that process was not random, like in the case of the selection of six regional government offices, then such a situation is a misuse of probability. That is apart, however, from the fact that this selection did provide a clear and convincing picture of the sex-discrimination that went on in general in that government agency.

3. Often, statistical ratios are considered to be legitimate probabilities. This confuses ratios that describe a concrete historic situation, with genuine probabilities. The accident, mortality, and other rates in actuarial tables on car insurance, life insurance, house insurance, etc. belong to this category. People with certain characteristics had car accidents, got sick, died, etc. leading to specific, observed frequency distributions. Ratios computed from these aggregates describe what came to pass in the social reality 'out there' with regard to certain groups of people (e.g. ages 13 to 19), in certain places (e.g. eastern USA), and in certain time periods (e.g. during the decade 1970–1979).[38] When the proneness to traffic accidents changes, e.g. through improved driver's education courses, stricter enforcement of drunk-driving laws, traffic monitoring by cameras, or when changes in mortality through medical breakthroughs take place, new ratios have to be computed from the new situation, replacing the existing insurance rates that date from an earlier era. These statistical ratios show the effects of complex social forces as they existed at a certain point in historic time and geographic region.

Mortality and related ratios have the external trappings of probabilities, but can be considered to be such only when the individuals to whom these ratios are applied have been selected at random from the very same population from which such tables were computed. Only then could one consider those ratios in actuarial tables to be true probabilities. When an insurance policy is written for an individual at a later point in time, or for a person who lives under conditions that differ from the living conditions of those in the original population from whom these ratios were determined, and when such a new client was not selected by a random process, then these ratios are not probabilities properly speaking[39] although they are useful as estimation or forecasting ratios. The misuse here consists in attributing to descriptive historic ratios the quality of being true, timeless probabilities. This kind of confusion between ratios and probabilities may not have practical consequences because, in every-day applications, these ratios are understood to be just that. Nonetheless, it is important to distinguish between ratios that are true probabilities, and ratios that describe historic, regional and social conditions. All (objective) probabilities are ratios, but few ratios are probabilities.

4. An even less direct form of misuse, mostly in American and British higher education and textbooks, consists of the fact, that important issues have been displaced by the study of probability theory, issues that should be part of statistics applied to business, economics and the social sciences in general.[40] To list just a few of those displaced statistical issues: data collection, – the term 'census' is mentioned only in very few textbooks, and seldom on more than two lines – the structure and interpretation of aggregates, economic classifications – even fewer textbooks mention SIC or NAICS and the principles underlying their classification – price and other socio-economic indicators, input-output analysis and much that has been relegated to macroeconomics, are topics that have become excluded, displaced by the teaching of probability. American textbooks of business and economic statistics give the impression that these topics do not even belong into socio-economic statistics. The attention given to probability theory, probabilistic sampling, inference and hypothesis testing is out of proportion to their actual importance in the social sciences,[41] giving the appearance that socio-economic statistics is a branch of mathematics, with disregard of the epistemology of economic and social phenomena that ought to receive preferred attention.[42] Although this is only an indirect misuse of probability, it has the most serious, long-term consequences.

10.5 Random Fluctuations

The 'random fluctuations' in time series and in regression analysis of socio-economic data are not usually included in the discussion of probability. Statisticians and econometricians act as if they knew what random fluctuations are, yet do not explain them, e.g. by classifying events in society as 'random' or 'non-random,' nor identifying those parts of the fluctuations in a time series and regression analysis that supposedly are due to such events specifically declared to be random. The 'random fluctuations' to which they refer are not determined in this manner.[43] Instead, a least squares trend-line is fitted to the data in such a way that the original data in the series, usually aggregates, remain above and below that line. That estimated trend-line is presumed to represent the true, long-lasting values underlying a time series that were hidden beneath the veneer of the shorter seasonal and random fluctuations, like the random errors of measurements in astronomy.

The values of the trend in this model are treated as if they were closer to reality and more real than the actually occurring data. Depending on the analytical context, these data are called 'residuals,' 'random deviations' or simply 'errors'. They appear in a time series only after the trend or trend- cycle – and in many instances also the seasonal pattern – has been 'eliminated.' These so called 'random fluctuations' which remain in such a 'de-trended' series, however, differ in shape and size depending on the choice of trend-line and the seasonal 'filter' that was used to screen out certain 'wavelengths' in the series. These fluctuations are declared to be 'random' or white noise, from the preconceived conviction that random fluctuations are supposed to exist, using as the prototype, the scientific measurement in experimentation in the natural sciences.

British researchers such as K. Pearson and R. Fisher gave a tremendous impulse to the application of probability.[44] The influence of these bio-mathematicians reached far beyond their immediate fields. Due to their influence statistical theory, particularly in the USA, has become closely identified with the calculus of probability.[45] The theoretical treatment of socio-economic time series[46] was inspired by the 'natural-science' model, which basically considers the aggregate data of socio-economic time series as randomly selected measurements from an infinite population. The academic community takes the presence of a random error component in socio-economic models for granted, neither demanding nor providing proof of its actual existence. Because everybody seems to know what they are, no efforts are deemed necessary to explain the meaning of such 'error terms'.[47]

10.6 Some Other Misuses of Probability

In order to classify a ratio as a probability, the evidence must be convincing that the process that created the data was really 'stochastic'. The attitude among statisticians and econometricians is the opposite: assume any situation to be stochastic, and the data as produced by chance unless there is strong evidence to the contrary.[48] This pro-stochastic bias in socio-economic statistics and econometrics has been at the heart of the misuses listed here. It is a consequence of the historic misconception that statistical aggregates are 'measurements' like those in the natural sciences. When scrutinized with the suggested criteria few socio-economic situations and processes will qualify as 'stochastic' and few data to be considered as 'random variables.'

In order to generate a process that is truly stochastic, substantial efforts are required (see H. Arkin, Endnote 24). Most of the processes in our field are essentially deterministic, even though the causes may not yet be fully understood. This should be kept in mind when data are conveniently treated as stochastic. Although economic theorists and social scientists are apparently aware of this, it is convenient and respectable to consider the economy – and society for that matter – as a system of random processes and random experiments. This view of society confines statistical theory to probabilistic reasoning. It is time to draw attention to spurious and outright false uses of probability, to clarify the limits of what probability theory can contribute, and to re-establish a better balance with the interpretation of data about concrete economic situations. These misuses reinforced the exaggerated importance attributed to games of chance and to probabilistic reasoning for the study of socio-economic data. The mistaken assumption of natural science models of measurement for socio-economic statistical data was the root cause, and their use in sampling reinforced our reliance on probability as the theoretical foundation of socio-economic statistics.

Leading statisticians have begun to take issue with that misuse in business and economic statistics.[49] The mathematics of statistical inference is only complementary and ancillary to the statistical description of facts in society. That, not inference, is the ultimate purpose of statistics and also the aim of this book. Stated differently,

every statistical investigation in society is descriptive of socio-economic reality, but not every statistical investigation requires the probability-based reasoning of inference.

10.7 Probability and Sampling

Probability was welcomed and integrated into statistical theory and practice only after the discovery and popularization of random sampling. In the excitement and novelty of this discovery, probability then entered into other phases of economic and social statistics through inference. The possibility to formalize inferential conclusions about a population from sample, and the possibility to assess the risk of that conclusion being in error has led to a general acceptance of probability into the theory of statistics. Yet survey practitioners have always tried to use as much information about the populations to be sampled as possible, in this way limiting the possible distorting effects of pure or 'simple' randomness of selection. Stratification, was one way to reduce the risks associated with 'simple (unrestricted) random sampling'. Selection of sampling elements – the 'real-life-objects' and their corresponding 'statistical-counting-units' – by probability is used only after all available information about the population to be sampled has been applied. As shown in this schema (see Fig. 10.1) simple random sampling, that is, the pure form of randomness, is only one of many ways actual samples are taken and often is reduced to a limited application.

The aim of a sample is to obtain a representative picture of the population, faithfully replicating its structure, but not that the sample be random. One could imagine an ideal sample obtained by the reverse selection process: removing from the population all those 'statistical- counting-units' that are not needed for a good replica of the population, Such an ideal sample would consist of only those 'statistical-counting-units' that had to be left in place after removing all those that were not needed.[50]

The actual process of selecting and taking the sample includes getting to know the population, preparing detailed lists of units to be selected into the sample, determining strata and clusters in the population, which are all aimed at reducing the need to leave selection of the sampling elements to chance. Although it is impersonal and objective, the probability-based selection ignores possible improvements in a sample's representativeness through some knowledge of the internal structure of that population. Theorists prefer an impersonal, probabilistic selection of all sample elements. Practitioners, on the other hand, without openly acknowledging that fact, try to limit random selection as much as possible. In stratified sampling, samples are taken from each stratum. Although the numbers in each sub-sample are smaller, the selected sampling elements are more homogeneous, reducing variability and uncertainty. Most samples end up as a compromise by using all available knowledge through stratification and clustering, and relying on some form of probability selection as little as possible, which is more practical than 'simple random sampling'. 'Systematic sampling with one or more random starts' would be such a random

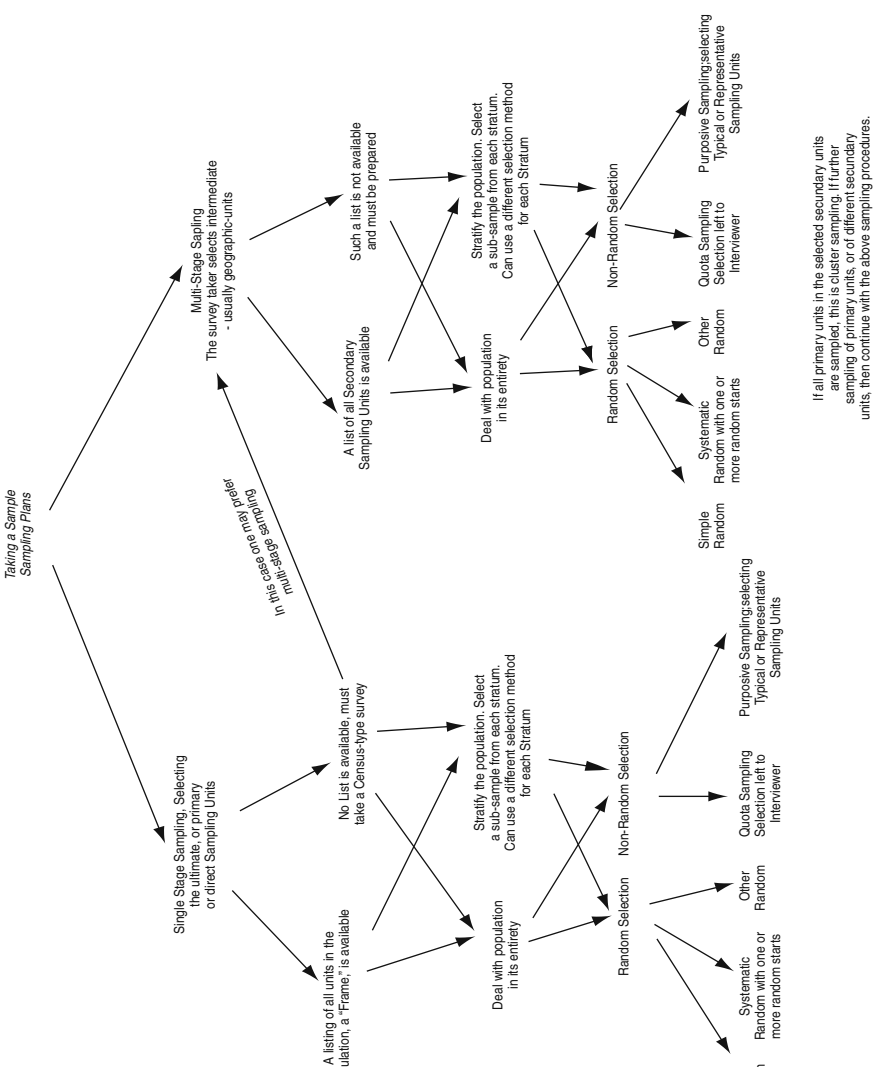

Fig. 10.1 Schema of the ways a sample can be taken

procedure that may be simpler than 'simple random sampling'. The worst case is the selection of the sampling elements left to the judgment of the interviewer or survey taker which is prone to be 'haphazard', not random properly speaking in a mathematical sense and likely to be biased. The name 'simple random sample', by the way, is misleading, because it usually is not as simple to execute as that label suggests, requiring detailed advance information about the population to be sampled.

The foregoing discussion and the examples given in Appendix E may clarify, why randomness is considered important. The practical problems of a pure random selection of the 'statistical-counting-units' has lead to the numerous compromises in the praxis of sampling. The schema in Fig. 10.1 shows the many ways samples can be selected. 'Simple random' selection, obviously the theoretical ideal, is only one among various more practical and preferred sampling procedures. Most samples take advantage of whatever is known about the population.

10.8 Misuses of 'Statistical Significance'

The interpretation of statistical results includes invariably, often as the main argument, a statement of the 'statistical significance' of the result. It is usually given at the 5% level and if possible, also at more stringent, smaller percentage levels. From published and oral presentations by researchers it is obvious that the expression 'statistically significant' is commonly understood to mean that 'this result is important, because it is validated by an accepted scientific-mathematical test.'

A numeric example in Appendix E should help refresh the basics of this matter, which also shows why theorists – but less the practitioners of sampling – prefer 'simple random sampling'. The 'sampling distributions' of any statistical measure computed from a random sample is the foundation of statistical inference, usually by testing the 'Null hypothesis' H_0. Let me recapitulate the meaning of inference with the slope of the regression analysis of sample data

Assume the sample-slope b of the linear least-squares equation representing the relationship between the two characteristics x and y of the n 'statistical-counting-units' indicates a positive relationship selected randomly form a population of such x and y values that are not related. Suppose the sample slope $b > 0$. A positive sample slope would mean that the y-values that correspond to small x-values are low and the y-values that correspond to bigger x-values also are bigger. Then a null-hypothesis H_0: 'there is no relationship between the x and y pairs in the population' is set up as a straw-man. In other words, one assumes that there is no relationship between x and y in the population and the slope of the x and y pairs $\beta = 0$. How likely could one, by coincidence, in a random selection process, have selected as a random sample only those 'statistical-counting-units' that together show a positive (or negative) value of the slope b, thereby erroneously leading to the conclusion that such a relationship between x and y does exist in the population. 'Rejection of H_0 at the 5% level of significance', which happens when the sample slope b exceeds the value(s) that marks the point(s) in the tail(s) of the sampling distribution that leaves

5%. (or 2.5% depending on how the Null-Hypothesis is formulated) in the tail-end. If that happens, one is lead to the conclusion that the Null-Hypothesis 'there is no relationship between x and y in the population' is not true and can be rejected. Although such a conclusion may be 95% likely to be correct, there is a 5% *chance that the null-hypothesis H_0 actually is true*, because there is a 5% chance of such a sample to stem from a tail-end of the sampling distribution (for more about this concept see Appendix E) of all b-values from a population that indeed **does not** have any x–y relationship.

These tests usually are conducted at the '5%, 1% or 0.1% levels of significance.' One asterisk * is used to indicate that the sample result is 'significant' at the 5% 'level of significance': In a sample regression it would mean that 'the sample slope b' (supposing a positive numeric result $b>0$) is at most 5% likely to have been selected from a population with no relationship between x and y . That could happen because of the remote possibility, to have selected from that population n 'statistical-counting-units' that by pure chance of selection, happened to yield a sample regression slope of the given magnitude. It says that if one took 100 repetitions of sample size n, selected randomly from that population with no x–y relationship and computed for each of these 100 random samples of size n the slope b, only 5 of these sample slopes would have a sample b-value as large or larger than the one that was found in this case. That kind of analysis, though, says nothing about the true value of the population-slope β, nor whether the fact that the sample $b>0$ is of any material importance for the problem at hand.

Two asterisks, ** are customary to indicate 'significance at the 1% level', in the case of computing the sample slope b, that it is at most 1% likely to have been selected from a population that is hypothesized to have no x–y relationship. This result is referred to as 'highly significant' or very unlikely to have captured in the sample only those 'statistical-counting-units' that, by happenstance, resulted in a slope $b>0$, even though the population slope was zero.

The symbol *** indicates that the sample slope b is even less likely, that is, with a probability of at most only 0.1% or 1 in 1,000 of such repeated random samples, selected from a population in which the Null-hypothesis, 'no relationship between x and y in the sampled population' is true. This is referred to as 'a most highly significant result.'

These tests are automatically performed by every computer program, regardless of whether the data were a sample, were taken truly at random, or if they are actually a population. At any rate, the scholarly community takes them very seriously. When the result is not 'statistically significant' and H_0 cannot be rejected this test, however, **does not confirm that** the assumption of $\mathbf{H_0}$ 'no relationship in the population' **is true**. There is always the possibility that some other than the assumed $\beta = 0$ relationship exists between x and y in the population. More importantly, it does not say anything about the underlying social or economic situation. The use of statistical significance does lend an air of scientific seriousness, rigor and competence of the researcher, and implies that something of value was achieved. If the situation should warrant it, such a test may be a desirable, yet only preliminary, first step in the interpretation of a socio-economic situation when sampling is involved.

10.9 Toward a Stochastic Worldview?

While in countries of the western hemisphere a trend in economic and social statistics has gained ground to consider data as the results of random processes, developments outside of the profession point in the opposite direction. The awareness appears to gain support that most events in the social sphere are the result of identifiable causal chains. Various phenomena of our time seem to be symptomatic of such a turning away from chance and probability when problems of daily living are concerned.

The existence and importance of R&D in corporations is a strong indication that management believes that the causes of everything concerning their products are, in principle knowable, should be explored and whenever possible, brought under control. In areas as far apart as industrial quality control and cancer research, subtle deterministic leads often have successfully been followed up to bring under human control what formerly was believed to be just chance (see e.g. Endnotes 13, 14 and 16).

Another symptom is the growing number of lawsuits against manufacturers. Accidents, damage or loss caused by faulty products are no longer accepted by the public as unavoidable chance events. Manufacturers are held responsible for product quality more than ever before, beyond customary warranties, and are even taken to court. These manufacturers are not only seen as the source of such problems but are also expected to exert full control over their production processes. Chance is no longer accepted by the public as a valid excuse for defective product quality. The courts recognize the existence of such responsibilities, and the reasonableness of the public's expectations.

The increasing number of preventive recalls by manufacturers when flaws in a product are detected and a reaction by the public is feared, also belongs here. Management rightly perceives that the public no longer considers chance and probability as practical working principles in their lives.

The more recent focus in business school curricula, according to AACSB† standards, is on the ethics of institutional decision making, on the regulatory environment of business and on communication,[51] not on probability. A view of future attitudes is emerging that will be intercultural, multinational, and global – but not stochastic.[52]

Notes

1. (Translation) "... In contrast to statistics as supplier of data, its role as the inductive interpreter of data, the importance of mathematical statistics in the economic and social sciences is declining. ... The reason for this declining appreciation of mathematical statistics in the

†AACSB = American Association of Collegiate Schools of Business, a non-governmental, privately organized supervisory body that accredit all business schools that accept their suggested standards and, once every decade, submit themselves to their outside review and scrutiny.

economic and social sciences is a newly emerging distaste, for mathematical models and methods in general. . ..'Mathematics is a first-rate game, serving more to entertain than to serving to clarify economic phenomena. (quote from Nobel price winner J.R. Hicks) p. 193 . . . Statistics ought to re-think its role in the economic sciences . In teaching and in research classical mathematical statistics dominates. . .. but that classical mathematical statistics with its stress on the most efficient exploitation of a given set of data is bypassing and failing to meet the needs of the economic and social sciences. It is time to recognize the enormous mistake to transfer and impose that model-based mathematical statistics, that has been so successful in the natural- engineering- and biosciences, on the economic and social sciences. Instead, statisticians in the economic and social sciences ought to recognize again that at the root, statistics is to mirror and reproduce reality. (p. 197) Walter Krämer, „Statistik in den Wirtschafts- und Sozialwissenschaften" *Allgemeines statistische Archiv* (AstA) 85.2, 2001.

"Anders als die Statistik als Datenlieferant nimmt die Statistik als Dateninterpretierer, d.h die inductive, mathematische Statistik, in den Wirtschaft- und Sozialwissenschaften heute an Bedeutung ab . . . .Der Grund fuer den abnehmenden Stellenwert der Mathematischen Statistik in den Wirtschaftswissenschaften ist ein neues und allgemeines Unbehagen gegen mathematische Modelle und Methoden ueberhaupt . . . .'Die Mathematik ist ein prima Spiel. Es dient mehr dem Spass als der Erhellung wirtschtlicher Phänomene (quote from Nobelpreisträger John R. Hicks)' p. 193 . . .Die Statistik sollte ihre Rolle in den Wirtschaftwisenschaften überdenken. In der Lehre und in der Forschung dominiert die klassische Mathematische Statistik. . .Insofern geht die klassische mathematische Statistik mit ihrer Betonung auf der möglichst effizienten Auswertung gegebener Datensätze an den Bedürfnissen der Wirtschafts- und Sozialwissenschaften vorbei. . .Es ist an der Zeit, den grossen Fehler einzusehen, die in den Natur- Ingenieur- und Biowissenschaften so erfolgreiche modellgestützte mathematische Statistik auch den Wirtschafts- und Sozialwissenschftn überzustülpen; stattdessn sollten sich die Wirtschafts- und Sozialstatistiker wieder mehr auf die Wurzeln der Statistik als der Abbildung der Wirklichkeit besinnen (p. 197)

2. "We know that academic contributions from developed countries (especially the USA) tend to crowd out practical contributions (especially from the LDC's) in the programs and in our journals. Our ISI should not become merely another in a crowded field of **journals dominated by sterile, academic mathematics, mostly in English. I am always fighting that harmful momentum**. . ." (highlighted for emphasis). Leslie Kish, President IASS (International Association of Survey Statisticians – Part of ISI) 1983–1985, 'The Survey Statistician,' Journal of the International Association of Survey Statisticians, International Statistical Institute, No. 12, Dec. 1984, p. 16
3. Morris H. Hansen, William G. Madow, "Some Important Events in the Historical Development of Sample Surveys," pp. 73–102, in D.B. Owen, ed. *On the History of Statistics and Probability*, Statistics Textbooks and Monographs, Vol. 17, Marcel Dekker, Inc. N.Y., 1976.
4. William G. Cochran, "Early Development of Techniques in Comparative Experimentation," pp. 1–26, in D. B. Owen, op. cit.
5. Jerzy Neyman, "The Emergence of Mathematical Statistics: A Historical Sketch with Particular Reference to the United States," pp. 147–194, in D.B. Owen, op.cit.
6. Boyd Harshbarger, "History of the Early Developments of Modern Statistics in America" (1920–1944) pp. 131–146, in D.B. Owen, op.cit.
7. The statement "When inference is finished there is nothing else left to do for a statistician" was made by Dr. Charles A. Mann during his deposition as statistical consultant for the defendant – the US Federal Government – in the class-action law-suit Chewning v. Seamans (ERDA) June 1978.
8. (Translation): "The idea to conceive of situations in society as caused by random processes must be seen not only as a blind alley from the point of view of the theory of cognition but worse, it closes access to the realistic imagining of processes in society (p. 357). . .and in addition, the sobering conclusions that the authors draw about the apperception and meaning of the approach of a "probabilistic social statistics" will please only very few readers. For

those who begin to question in earnest what the real meaning of the customary probabilistic judgment of a t-value of the consumption-function for the entire economy might be, **there remains in the end, in addition to the careful description of the data, only the field of a narrowly conceived theory of sampling** (p. 357)"

„Die Idee soziale Sachverhalte als durch Zufallsgeneratoren zustande gekommen zu betrachten muss nicht nur als Erkenntnistheoretische Sackgasse betrachtet werden, sondern sie verstellt den Weg zur Gewinnung adäquater Vorstellungen von sozialen Prozessen (p. 357) . . .zum andern werden die ernüchternden Schluesse die die Autoren aus ihrer Diskussion über die Sinnhaftigkeit des Ansatzes der 'probabilistischen Sozialstatistik' ziehen, nur wenige Leser erfreuen. **Letztlich bleibt neben der sorgfältigen Description von Daten lediglich der Bereich der engeren Stichprobentheorie** wer ernsthaft begonnen hat zu fragen, was mit einem so üblichen Wahrscheinlichkeitsurteil wie etwa mit dem t-Wert einer gesamtwirtschaftlichen Konsumfunktion implizierten, tatsächlich gemeint sein könnte. . ..(p358)"

Bookreview of „Wahrscheinlichkeit:Begriff und Rethorik in der Sozialforschung", Götz Rohwer, Ulrich Pötter, Juventa, 2002 in: *Allgemeines Statistisches Archiv* , 89.3, 2005 Literatur/Books, reviewed by Andreas Behr, pp. 356–358.

9. ". . .statements for very large enterprises cannot be 100% free from error; **this would contradict the probabilistic nature of the world**. . ." p. 28 Oskar Morgenstern, *On the Accuracy of Economic Observations,* Princeton University Press, Princeton, N.J., 1950, (highlighted for ephasis).
10. See any issue of the now discontinued journal ‚Statistische Praxis – *Zeitschrift fuer Rechnungsfuehrung und Statistik*, Zentral Bureau fuer Statistik, Deutsche Demokratische Republik (East Germany). Also Winkler, O.W. "Contrasting Approaches to Socio-Economic Statistics in East and West" *Proceedings of the Business and Economic Statistics Section, ASA*, Alexandria, VA 1989, pp. 345–350. and "Unterschiedliche Ansätze zur Wirtschafts- und Sozialstatistik in Ost und West" *Jahrbücher für Nationalökonomie und Statistik,* Band (Vol.) 208/5, pp. 459–492, Gustav Fischer Verlag, Stuttgart, Germany, Sept. 1991.
11. Snedecor, George W. ". . .The only object of description is to facilitate inference. The purpose of examining data is to estimate probabilities that enter into the making of decisions. In practically all statistical investigations . . .inferences are extended beyond the data in hand, otherwise one would be wasting his time sorting dead bones. . . . Random sampling is the appropriate subject matter of elementary statistics. The graver problems of non-random sampling should be left to professionals in the specialized disciplines involved. . . His restraint from any sample-to-population inference removed the first five chapters from the domain of statistics**.** Mere arithmetical and graphical descriptions of data have no more than historical value". This remarkable statement by George W. Snedecor in his book review of A Primer of Statistics for Political Scientists V.O. Kay Ju. Thomas Y. Crowell Comp., New York 1954. Reviewed in: *Journal of the American Statistical Association JASA,* Volume 50, No. 270, Washington, D.C. June 1955, pp. 608, 609 early on stated a view of statistical theory and method that still seems to prevail today.
12. Doob, J. L. "Foundations of Probability Theory and its Influence on the Theory of Statistics" in D.B. Owen, *On the History of Statistics and Probability,* Statistical Textbooks and Monographs, Vol. 17, Marcel Dekker, Inc. New York, 1976.
13. Nothing in society happens by chance or at 'random', even though the causes may not become public knowledge, such as decisions made out of the public view by business management and the political establishment. A good example was the sudden, strong decline in gasoline prices before the American Congressional elections in November 2006. As reported later in a news analysis at radio station 90.9 FM, the American oil companies lowered the price of gasoline ten times the amount by which their cost of crude oil declined, with the intent to strengthen the fortune of the Republican party candidates that favored their business. The connection of the sudden plunge of gasoline prices at the tank stations (gasoline or benzene pumps) with the hard-fought, political fight before the elections, was too obvious, for the general public to

accept the official version that this was "only due to the vagaries of the market", presenting this politically sensitive matter as an 'innocent' random event.

14. "Experience to date indicates that difficulties in identifying problems have delayed statistics far more, than difficulties in solving problems" John W. Tukey, "Unsolved Problems of Experimental Statistics" Summary of paper, *JASA* 49-266, June 1954, p. 343.
15. The following passage seems to be completely in agreement with this image:
    ". . .we are compelled to accept the view that for something even to be actual at all it must possess at least a minimum of order. Total absence of internal patterning . . .amount(s) to non-actuality. . .To be actual is to be something definite. . .being ordered (p. 84) . . .our (Alfred North Whitehead's) hypothesized principle of order. . .is non-interfering and unobtrusive. . .it is causal in the deepest sense. . .not in a mechanical. . .way. Because of its unobtrusive, formatively causal, rather than mechanically coercive, mode of influencing the universe it is inevitable that there would be deviations from the intelligibility inherent in order. These deviations are what we call chance occurrences. . .the principle of order. . .is also. . .the source of novelty. . .whenever novelty invades a situation of order the result is at least momentary deviation from the fixed arrangements of the past. . .the momentary breakdown of harmonious patterning may give rise to occurrences for which the word "chance" is appropriate. . . these deviations, while unintelligible from the point of view of one frame of order, might not be without intelligibility from within a wider angle of vision . . .it is legitimate to state that at least some things which appear without intelligibility from an earlier perspective may in principle become intelligible within a later and wider perspective. If this is the case, then, it may be simply impossible for us ever to have a controlling and objectively comprehensive understanding of what chance really is. (p. 85) John F. Haught, The Cosmic Adventure, Science, Religion and the Quest for Purpose*, Paulist Press,* New York/Ramsey, 1984, pp. 84–85.
16. "An abnormality in part of the brain that controls breathing, arousal and other reflexes may be what causes **s**udden **i**nfant **d**eath **s**yndrome (SIDS), a study said. The discovery could explain why babies lying face down are more likely to die because, in that position an infant's reflexes, including head turning and arousal, are harder to trigger when breathing is challenged, the report from Children's Hospital Boston and Harvard Medical school said. These findings provide evidence that SIDS is not a mystery but a disorder that we can investigate with scientific methods and someday may be able to identify and treat', said Hannah C. Kinney of the Boston hospital, an author of the paper. The study, published in this week's Journal of the American Medical Association, was based on autopsy data from 31 infants who had died from SIDS and 10 who had died from other causes between 1997 and 2005 in California. In the SIDS victims, a look at the lowest part of the brainstem, the medulla oblongata, found abnormalities in nerve cells that make and use serotonin, one of the chemicals in the brain that help coordinate breathing, blood pressure, sensitivity to carbon dioxide and temperature, the report said" (The Washington Post, November 1, 2006.)
17. Jung, Carl "It is our . . .positive conviction, that everything has causes which we call natural and which we at least suppose to be perceptible. Primitive man, on the other hand, assumes that everything is brought about by invisible, arbitrary powers. . . Only he does not call it chance, but intention. Natural causation is to him. . .not worthy of mention. If three women go to the river to draw water, and a crocodile seizes the one in the center and pulls her under, our view of things leads us to the verdict that it was pure chance that particular woman was seized. The fact. . .seems to us natural enough, for these beasts occasionally do eat human beings. For primitive man such an explanation completely obliterates the facts, and accounts for no aspect of the whole exciting story. Archaic man is right in holding our view of the matter to be superficial or even absurd, for the accident might not have happened and still the same interpretation would fit the case. The prejudice of the European does not allow him to see how little he really explains things in such a way. Primitive man expects more of an explanation. What we call chance is to him arbitrary power. It was. . .the intention of the crocodile to seize the woman who stood between the other two. If it had not had this intention it would have

taken one of the others. But why did the crocodile have this intention? These animals do not ordinarily eat human beings. Crocodiles are really timid animals, and. easily frightened. . .it is an unexpected and unnatural event when they devour a man. Such an event calls for explanation. Of his own accord the crocodile would not take a human life. By whom, then, was he ordered to do so? (p. 132). . . To this extent (archaic man) behaves exactly as we do. But he goes further than we. He has. . .theories about the arbitrary power of chance. We say: Nothing but chance. He says: Calculating intention. He lays the chief stress upon the. . .occurrences that fail to show the causal connections which science expects. . .that constitute the other half of happenings in general. He has long ago adapted himself to nature in so far as it conforms to general laws; what he fears is unpredictable chance whose power makes him see in it an arbitrary and incalculable agent. Here again he is right. (p. 133) Chap. 7, "Archaic Man" in: *Modern Man in Search of a Soul*, A Harvest/HBJ Book, Harcourt Brace Jovanovich, New York, 1933.

18. John Haught gave an interesting overview of this matter: "chance is used in at least the following five ways:

    1. ". . .there is. . .the epistemological usage of the term. . .to thinkers as diverse as Laplace, Einstein and Russell a chance event is one whose cause is unknown. This usage . . .makes chance. . . a kind of cover-up for our own ignorance. Chance is a blind spot in our understanding rather than an objective fact resident in nature. . . .since all events must have causes, according to this classical framework, there really are no such things as chance occurrences. "God does not play at dice with the universe," as Einstein put it. . .in this view, chance does not exist. It is merely an expression of the limitedness of our knowing. . .For some theists chance is actually. . .a hidden way God's omnipotence determines all things. . .for the atheist chance is . . . a confused expression cloaking our own ignorance of the iron-clad, impersonal laws of a deterministic universe. In either case chance does not really exist.
    2. Another way of understanding "chance" is the mathematical. . .we ask what are the "chances" that a flipped coin will land tail-up. While mathematics cannot decide the answer in any single case, it can formulate laws of probability according to which we can make fairly accurate predictions (78) regarding the outcome of a large number of coin tosses. In this context "chance" occurrences are deviations from statistical regularities. In themselves they are surds, lacking any systematical intelligibility. A common question posed by science today is whether the origin of life and the mutations involved in evolution are such irrational, unplanned and disorderly deviations. It is in this connection especially that the question of purpose in evolution arises. Could life and evolution possibly be the implementation of a divine purposiveness if they are carried along so prominently on a stream of chance happenings?
    3. A third context. . .is. . .existential. Here "chance" refers to any occurrence which, without interrupting the known laws of natural causation, shows up as an absurdity disturbing the order of our human existence. Existential chance appears when two independent physically causal series intersect in such a way as to make us ask fervently: "Why did that have to happen to me. . .? For example pigeon droppings. . . invariably make their way earthward because of the deterministic laws of gravitational attraction. If I on my bicycle, following another independent trajectory, just "chance" to pass underneath such a natural occurrence at the relevant moment. . .the point is that we have here two independent causal series, both blindly following the laws of physics. . . the fact that a human being is involved gives their intersection a dimension that would otherwise be absent. One can. . .think of many more tragic examples of existential chance. . .some modern writers. . .interpret our very birth and existence on this planet as such an absurd crossing of incongruous paths.
    4. A fourth denotation, often given to the term "chance" is a physical one. A number of modern physicists hold that events at the sub-atomic level are not only (79) indeterminable or unpredictable by scientific observation, but. . .are also unpredictable even in principle. Contrary to the determinists, who see all events as the predictable result of antecedent

causes, physical indeterminists insist that at the sub-atomic level there are happenings which are "uncaused," arising spontaneously and unpredictably out of a mysterious depth to which our science of causes cannot penetrate. This speculation of recent physics has encountered a great deal of resistance, even from scientists of the stature of Einstein. . .the hypothesis of physical chance posits an indeterminacy at the base of cosmic reality, and. . .forces us to ask whether the natural world is influenced by any sort of ordering principle. . .an important qualification. . . needs to be made with respect to this hypothesis of physical chance. Physics can allow for indeterminacy in. . . microcosmic occurrences without rejecting the predictability that occurs when large numbers of these occurrences coalesce to make up macroscopic entities. . . Our world. . .appears to be a composite of indeterminacy and order.

5. Another way of using the idea of chance is the metaphysical. . .chance . . . providing the definitive answer to ultimate questions such as, Why am I here? Why did life appear? Why is there anything at all? . . .In this application "Chance" often takes the same place that "God" takes in classical theology. . .it is almighty (though not all- good); it lies beyond the scope of scientific method (since science can deal only with the recurrent, the orderly and the predictable) . . .Chance comes close to being the object of worship and devotion since it is the metaphysical source of all things. . .It is an hypothesis brought onto the scene when human ingenuity and resourcefulness are lacking. It is a deus-ex- machina that puts the lid on further inquiry and delivers us from the need to unravel the story of nature with further careful, patient, rational inquiry. We can see. . .that . . .there is an imaginative component associated with employment of the term "chance" that explains its psychological attractiveness to its devotees. . . (p. 81) John F. Haught, op. cit. pp. 78–81.

19. "There is a growing consensus among cancer researchers that some people are more susceptible than others, that cancer does not always strike at random . . . Investigators are beginning to devise tests that may predict who is most susceptible to cancer . . ."*Science*, Vol. 207, 29, February 1980 p. 967 in: "Research News"
20. Although we are talking here of socio-economic macro phenomena it may be of interest to hear what a biochemist has to say about the laws of nature and probability. The author, Pierre Lecomte du Nouy "distinguishes two views about 'Laws of chance.'

    1. The tolerant view today. . .can be understood as the orthodox position to be defined in Henri Poincare's own words:

       "The expression 'laws of chance' does not necessarily mean the absence of regularity, but a regularity the effects of which are so complex that its detailed analysis is removed from our ability. At best we can determine general tendencies which result from a large number of partial effects, which in part compensate each other"

       Emile Borel on his part declares:

       "The hallmark of phenomena which we designate as 'random' or 'caused by chance' is their dependence from causes which are too complex to be known and investigated by us" In this view of things the blame is laid on the weakness of our senses and of our mind, that we take refuge in probability calculus. There are countless 'laws' which we do not know yet; yet that is irrelevant as on our (atomic and subatomic) level everything happens as if those laws did not exist and chance was the only decisive factor. One can ask whether these fundamental laws that we do not know yet, themselves are laws of chance, or, if they express to the contrary a pre-ordained order. . . (p. 157) Either there are individual laws that suffice to explain many phenomena, relegating the laws of chance into the role of artificial interpretation and are valid only on our level of understanding, or
    2. There are only and exclusively the laws of chance, acting at all levels. This is the intolerant view of probability, which, philosophically speaking, is in accord with the monistic view. . ."

       From the foregoing one can see clearly the two ways in which one can understand the laws of nature. . . (p. 158) the rules by which man tries to explain and foretell the

interrelation of the phenomena of nature. . . .The second attitude denies the existence of phenomena that are controlled by laws other than those of chance, regardless of the level on which they occur. This mode of viewing appears to be the credo of the pure materialists or mechanists. . .”(p. 159) in: Pierre Lecomte du Nouy, *Der Mensch vor den Grenzen der Wissenschaft* Gustav Kilpper Verlag, Stuttgart, 1952, p. 157, my translation from the German text.

21. Mansfield, Edwin *Statistics for Business and Economics – Methods and Applications*, W.W. Norton Co., New York 1980, p. 61 and p. 100.

    Also Wm. Mendenhall, J.E. Reinmuth, “Since. . .will rarely include all the variables affecting the response in the experiment, random variation in the response is observed. . .” p. 466.

    Statistics for Management and Economics, Duxbury Press, N. Scituate, Ma. 1978.
22. See also. Winkler, Othmar “Minimum Wages and Employment,”*Proceedings of the Bus. & Econ. Statistics Section of ASA,* Washington, D.C. 1973, pp. 634–639, also Harlow D. Osborne, O. Winkler, “Measuring the Indirect Employment Effect of Strikes,” *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1970, pp. 581–586.
23. Zellner, Arnold *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Inc. New York, 1971

    “The Bayesian approach. . .involves a quantification of such phrases as “probably true” or “probably false” by utilizing numerical probabilities to represent degrees of confidence or belief that individuals have in propositions. . . (p. 7). It is of the utmost importance to realize that. . .there is a unified and operational approach to problems of inference in econometrics. . .whether we analyze. . .time series, regression models, the approach and principles will be the same” (p. 11). Obviously ‘randomness’ – a situation akin to that in games of chance – is here implicitly assumed to exist.
24. Arkin, Herbert *Handbook of Sampling for Auditing and Accounting,* Ch. 2 “Selecting the Sample” and Ch. 3 “The Mechanics of Sampling,” pp. 19–48. 2nd ed. McGraw-Hill Book Co., New York, 1974.
25. “the justification for the inferences derives almost entirely from the assumption of a certain model, sometimes narrowly specified as a super-population with known shape. . .”
    Cassel, C.M., Särndal, C.E., Wretman, J.H. *Foundations of Inference in Survey Sampling*, John Wiley & Sons, Inc., New York, 1977, p. 80, also pp. 81–83, 108–111, 121.
26. Kendall, Maurice G., Alan Stuart, *The Advanced Theory of Statistics Vol.* I, 2nd ed., Hafner publ. Co., New York, 1963, esp. p. 206.

    Gujarati, Damodar “the data of Table 2.1 represent the population, we can. . .compute the conditional probabilities of Y, p(Y,X), . . .for X = $80. . .there are five Y values. . .the probability of obtaining any one of these is 1/5. . .”(p. 21) *Basic Econometrics*, New York, 1978.

    Fisher v. Dillard University, “Defendant attacks the validity of plaintiff’s survey on the ground that it represents only a very small statistical sample. . .” referring to the entire faculty of 74 members of Dillard University. 499 F. Supp. 525 (1980), p. 535

    EEOC v. Local Union No. 38 “plaintiff’s expert. . .arrive at the opinion that the percentage shortfall of minority indentures for the years 1977, 78 and 79 was not the result of chance. . .one of the statistical tests he employed. . .was a binomial test. . .the shortfall of 7 1/2 minorities in the year 1977. . .could have resulted from chance only in a probability of 1–50 thus inferring the disparity was due to bias. . .the tell-tale data in Exh. 12 leads. . .to the conclusion that chance had no part in the indenture selection process.” 24 EPD P31, 553, Feb. 2, 1981, pp. 19280–19281. These are statements that treat populations as random samples without specifying ‘form which population it was drawn by which procedure’.
27. Kruskal, William H., Tanur, Judith M. ed.

    “Another difficulty arises in the use of significance tests (or any other procedures of probabilistic inference) when data consist of a complete census for the relevant population. . .it would be highly questionable to use sampling theory. . .to test the null hypothesis that the population correlation coefficient is zero. . . it is hard to see how the numbers can reasonably be regarded as a sample of any kind. . .” (p. 951)

"Significance, Tests of" *International Encyclopedia of Statistics*, The Free Press, a Division of Macmillan Publ. Co., Inc. N.Y., Vol. 2, 1978, also: Louis J. Braun "Statistics and the Law: Hypothesis Testing and Its Applications to Title VII Cases" That author uses an entire work force – a population – to demonstrate the calculation of the z-test for the difference between proportions, "Illustration 3" p. 77, the chi-square test for a population of 10,000 "Illustration 4" p. 79, p. 83, and a z-test for the differences between means, "Illustration 6" p. 87 for another hypothesized population.
Hastings Law Journal, Vol. 32, Sept. 1980.

28. Siskin, Bernard, *Statistics in Employment Discrimination Cases*, Course Manual, Federal Publications, Inc. Washington, D.C., 1980, pp. A-4 and A-5. Chi-square tests are applied to a population of 264 applicants.
29. Gujarati, D, op. cit., pp. 125–127 treating the multiple regression of personal consumption expenditures and personal disposable income in the US, 1956–1970, in billions of 1958 dollars as a sample of n=15, testing for the significance of the beta-hat with 12 degrees of freedom! – which is strictly reserved for randomly taken samples, Wallis, Kenneth F. uses large sample theory for the time series of the hog-corn price ratios of 'statistical populations,' and the annual series of commercially slaughtered hogs, in million heads, 1935–1971 with tests of significance, in "Multiple Time Series Analysis and the Final Form of Econometric Models"*Econometrica*, Vol. 45, No. 6, Sept. 1977, pp. 1487–1490.

    Mendenhall, W. and Reinmuth, J.E. request in problems 11.50 and 11.51 regression analysis between the Fed. Reserve discount rate and bank debits in trillions of dollars, 1962–1968, and a test of the significance for slope, with an alpha=.01 op. cit. pp. 398, 399, see Endnote in Chap. 2.
30. Haavelmo, Trygve, "The Probability Approach in Econometrics" *Supplement to Econometrica*, Vol. 12, July 1944, pp. 1–118 and pp. I–VI, University of Chicago.
31. D. Gujarati, op. cit. "The disturbance term u(i) may be used as a substitute for all the excluded or omitted variables from the model. . .the joint influence of these variables may be. . .nonsystematic or random. Hopefully, their combined effect can be treated as a random variable u(i) . . ." (p. 27).
32. A statement like this would have benefited the otherwise very interesting study "Democracy and National Economic Performance: The Preference for Stability" Dennis Quinn, John T. Woolley, *American Journal of Polkitical Science,* vol. 45, No. 3, July 2001 pp. 634–657. This study uses all available data of 84, 92 and 106 countries for its different models. It is a descriptive study of entire populations, and although no random or other sample selection from a fictitious population was performed and no inference in the sense of sampling theory was attempted, the high levels of significance of some results of this purely descriptive study was stressed.
33. Cassel, M. Särndal, C.E. Wretman, H.J. op. cit.

    "The role of sample design. . .in classical survey sampling. . .is an all-important element, whereas adherents of newer, super-population based inference claim that the way in which the actual sample was selected is unimportant. . ." (p. 9).
34. Gertrude Cox, "There are more general frontiers I wish to mention. . .how far are we justified in using statistical methods based on probability theory for the analysis of non-experimental data? Much of the data used in. . .descriptive methods. . .are non-experimental. . ." (p. 10). "Statistical Frontiers" *JASA*, Vol. 52, No. 277, 1957, pp. 1–12
35. Cassel, C.M. Särndal, C.E. Wretman, J.H. op. cit.

    "Traditional inference theory. . .is preoccupied with independent and identically distributed observations (**even though it is often evident that data have not been gathered at random**) . . ." (p. 10).Emphasis added.
36. Neter, John, Wasserman, William, Whitmore, G.A. *Applied Statistics*, Allyn & Bacon, Boston, 1978 "such a judgement sample. . .cannot, however, be utilized to evaluate. . .the margin of error due to sampling" (p. 182).

37. Source is proprietary, relating to unpublished material gathered for a sex-discrimination law suit. Information supplied on request.
38. "Motorcycle riders have long had to pay far more for insurance than automobile drivers. Geico (one of the largest insurance corporations) built its business on insuring safe drivers but now insures higher-risk divers. That includes older motorcycle riders, the fastest-growing group of owners. In 1985 the median age of US motorcyclists was 27. By 2004 . . . the median age had risen to 41. . .Motorcycle accidents have risen with the age of riders. In 2005 the. . . the average age of riders killed in accidents has risen in the past 10 years. . . large increases in crash deaths of riders age 40 and older. .Riders 50 and older accounted for the steepest climb in motorcycle fatalities in 2005 . . .motorcycle death in 2006 were predicted to increase for the ninth straight year. . . none of this has discouraged Geico. . ." *The Washington Post*, July 2, 2007 part D p. 1, 2
39. Ben-Horim, Moshe, Levy, Haim distinguish clearly between relative frequency and probability, pp. 188, 189. *Statistics Decisions and Applications in Business and Economics*, Random House, New York 1981.
40. Neter, John, Wasserman, W., Whitmore, G.A. op. cit.
"The statistical methodology **for analyzing data**. . .is called statistical inference. . .The logical foundation of statistical inference **is the mathematical theory of probability"** (p. 75)
41. Mendenhall, W. Reinmuth, J.E. op. cit. "The objective of statistics is to make inferences about populations based on information contained in a sample" (p. 222).
42. I have expressed these ideas in my review of the textbook *Statistical Analysis for Business: A Conceptual Approach*, Smith, L. and Williams, D. The Wadsworth Publishing Co. Belmont, CA, 1971,Winkler, Othmar in: *JASA* June 1974, Vol. 69, No. 346, pp. 576–577.
43. Burns, Arthur F. Wesley C. Mitchell,
". . .For strictly speaking, smoothing does not eliminate erratic movements; it merely redistributes the values of the original data in a manner predetermined by the particular formula used. . .p. 318. . .In historical series the effects of cyclical and random forces cannot be separated even over the course of a full cycle. Random factors constantly play on business at large and on each of its many branches, and their effects register in different ways under different circumstances. . . (p. 320) . . . **erratic movements therefore cannot be identified with effects of 'random forces' as in our experimental model**. The 'erratic movements' that stand out clearly in time series and are obliterated by smoothing are simply the short-term oscillations, other than seasonal variations, that play, so to speak, on the back of specific cycles. Not clearly revealed by the data and not smoothed out by the usual formulas are other effects of random forces. Indeed the general contours of cyclical movements themselves may have been shaped, in part by the same forces that produced the short term oscillations. These observations instill caution in the use of smoothing devices. . ."
Measuring Business Cycles New York, 1947, p. 322 (the highlighting for emphasis)
44. William G. Cochran, "Early Development of Techniques in Comparative Experimentation", pp. 1–26, in D. B. Owen, *On the History of Statistics and Probability* Marcel Dekker, Inc. New York, 1976
45. Jerzy Neyman, "The Emergence of Mathematical Statistics: A Historical Sketch with Particular Reference to the United States", pp. 147–194, in D.B. Owen, op. cit.
46. Boyd Harshbarger, "History of the Early Developments of Modern Statistics in America (1920–1944) pp. 131–146, in D.B. Owen, op. cit.
47. Lawrence Klein, reproducing another author's monthly time series data in his textbook treats these monthly data as the elements of a randomly selected sample. Klein, Lawrence *Introduction to Econometrics* Englewood Cliffs, Prentice Hall, 1962 pp. 33–48.
48. Typical is the following statement in Gujarati, D. op. cit. "we need to specify the probability distribution of . . .u(i) . . .**which is random by assumption**. . .since the probability distribution of these estimators are necessary to draw inferences about their population values. . .the void can be filled **if we are willing to assume that the u(i)'s follow some probability distribution**

. . .in the regression context it is **usually assumed that the u's follow the normal distribution"** (p. 71). (highlighted for emphasis).

49. "A. . .danger is the proliferation of mathematical acrobatics in the world of abstraction. . .How many different 'estimators' have been formulated and their mathematical properties studied in detail? . . . perhaps several hundreds. . . How many have been used or are likely to be used, in practice in the world of reality? . . . only a very few. . ."

    Prof. Mahalanobis then added the following remarks, off the record, which did not appear in the Bulletin of that meeting: "does it help in the world of reality to assume a population to be taken from an infinite population? Observations have two limits: Heisenberg's Minimum, and Einstein's Speed of light. That bounds reality, makes it finite. The rest are mathematical acrobatics. . . If you mix the world of reality with the world of abstractions, terrible confusion results. Every number must have a space and time coordinate" Mahalanobis, P.C. "Contributions to Sampling" *Proceedings of the 38th Session of ISI,* Washington, DC, volume XLI-Book 1, 1971, p. 259.

    "There has been some discussion lately about the future of theoretical statistics, and concern about its alienation from applications. . .the electronic computer has to a considerable extent already deemphasized the usefulness of pure mathematical analysis in reaching important statistical research results. . . the nature and amount of recent separation of theory from applications seems far from optimal." Särndal, Carl-Erik, "Contributions to Sampling" *Proceedings of the 38th Session of ISI* Washington, D.C. Volume XLI-Book 1, p. 264

    "In statistics we use logarithms in two ways: for graphic representation and for the building of formulae. The first use is fully legitimate. . . the other use of logarithms consists in going in ever higher abstraction mathematical formulae which have lost any reasonable contact with the matter to be studied, and which are in the truest sense of the word 'transcendent', that is, transcending the capacity of the human intelligence to find in them any connection with reality. That is not statistics any more, that is mathematical gymnastics, which I have, most strongly, rejected in my book "Grundfragen der Oekonometrie" and which I, not less strongly, reject for the whole field of statistics!" Winkler, Wilhelm, *Bulletin of the International Statistical Institute*, Vol. LXI-1, p. 263, Belgrad 1965, Diskussion to Yuzo Morita's "A new Method of delimitating Urbanized Areas in Population Census Statistics".

    Thurow, Lester C. *Dangerous currents: The State of Economics*, New York, Random House 1983 (especially Ch. 4 "Econometrics: An Icebreaker Caught in the Ice")

50. Imagine a large mural composed of N mosaic stones, small colored ceramic tiles. Then tiles are removed from that mosaic in such a way, that the picture remains recognizable. The removal of these small ceramic tiles would not be limited in number, taking away all the tiles that are not essential for the picture. That would mean that many tiles can be removed from large, contiguous areas, on the other hand, few if any, could be removed from critical areas where every tile represents an important part of that mural. One would end this process, when removing one more tile would begin to make the mural unrecognizable. The tiles that are left in place, then, would be the ideal sample of tiles that still provides a fair idea of the picture presented in that mural. This idealized sampling process relies on deciding not only which tiles to remove from the picture, but also how many. Such a procedure would require the full knowledge of the structure of the population – which, of course, would make the need for sampling unnecessary.
51. Rogene Buchholz, Business Environment-Public Policy: *A Study of Teaching and Research in Schools of Business and Management*, Center for the Study of American Business, Working paper no. 41, AACSB, St. Louis, Mo. 1979. Dan Bertozzi, Jr., "The Legal Environment of Business as an Appropriate Minor for the Business School Curriculum" *Collegiate News and Views*, Fall 1981, Vol. XXXV, No. 1, pp. 27–31, Southwestern Publ. Co.
52. Also remember this book's Sect. 1.3 'Statistics and Decision Theory'

Chapter 11
The Interface Between Statistics and Accounting

11.1 Socio-Economic Statistics and Accounting

Various quantitative fields border on socio-economic statistics. Some come to mind immediately, such as econometrics. Others may not even be thought of as having anything in common with statistics. Accounting is one of those areas[1] that apparently has little in common with statistics. Exploring what these two fields may have or may not have in common[2] is meant to clarify the nature of socio-economic statistics.

An initial clue can be found in the name of an activity related to socio-economic statistics, viz 'National Accounting.' Both business accounting and national accounting aim at summarizing the financial situation of an organization during a specific time period. Their concepts and principles, coping with similar problems, are analogous. The main difference lies in the labels of these activities. At the level of the business firm they are called financial or cost 'accounting.' At the national level they are called macroeconomics, and occasionally as well as 'economic statistics.' This indicates that the differences between accounting and statistics cannot be as great as is generally assumed.[3] The reason why this is not evident is the development of statistics, increasingly becoming a branch of mathematics, dominated by probability theory, and mathematical statistics. These fields, beginning in a supportive role as subordinate contributors to statistics applied to the social sciences, now seem to have taken it over.

Despite the belief that business, economic and social statistics is mostly about sampling and inference, it is worth repeating that the real task and purpose of socio-economic statistics is the perception and objective reporting of economic and social phenomena like unemployment or price level changes. This task of describing and reporting is the reason why the inferential detour is undertaken. Socio-economic statistics can do without inference, but it can never do without description. When one becomes aware of this, it becomes easier to see that statistics has much in common with accounting, more than with other quantitative disciplines despite the fact that accounting does not come to mind when thinking of related fields.

It may come as a surprise that the definitions of, e.g. business statistics and (business) accounting do not suggest great differences between them. One definition of financial accounting reads as follows:

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_11, © Springer-Verlag Berlin Heidelberg 2009

"Its function is to provide quantitative information primarily financial in nature, about economic entities that is intended to be useful in making economic decisions. . .reasoned choices among alternative courses of action. . ."[4] Another definition states: "A primary objective of the accounting process is the measurement of changes in assets and liabilities. . ."[5] On the other hand, business statistics is defined as 'a body of methods and theory applied to numerical evidence in making decisions in the face of uncertainty.'[6] If the words 'accounting' and 'statistics' were omitted and one were not told the source of the definition, it would be difficult to tell which definition describes business or economic statistics and which accounting. From this, it becomes evident that the disciplines of statistics and accounting, in principle, must have a lot in common. On the practical working level this has been less evident. There are numerous differences in terminology that mask existing similarities. Yet, there are also terms that seem identical but have a different meaning in each field.

1. Accounting and statistics use different terms to cover the same substance. Accountants 'post' an entry in the journal while statisticians enter a 'tally' about that same event in a tally sheet, or mark a response in a questionnaire or make an electronic entry in an excel worksheet. Then the accountant 'debits' or 'credits' that transaction in two complementary accounts while a statistician might classify such an event simultaneously in the body of a contingency table, that is, in the intersection of two classification criteria. Accountants have 'ledgers' and 'auxiliary ledgers,' accounts payable, accounts receivable and control accounts, while statisticians talk of 'groups' or 'classes' and subgroups and subclasses. The 'cost centers' become 'departments,' and a sample check of accuracy of an inventory count becomes a post enumeration survey (PES).
2. The accountant confines his attention mainly to exchange transactions, even though he must use his judgment to determine the impact on his recorded values of other events, inside and outside the business firm. The statistician's task even in the same business environment is much broader, going beyond the tasks of accountants, also recording and evaluating data in the areas of personnel, quality control, sales and sales forecasting, as well as the corresponding data for the entire industry in which a given firm is operating. In short, statistics gets involved in anything that happens within and without a given firm. In principle, though, both face the same problems of locating, identifying, recording, aggregating, reporting and interpreting their quantitative results. Overall, statistics, like accounting, is also concerned but not limited to monetary values.
3. Occasionally the opposite happens when both disciplines use the same term with different connotations. 'Independent variables' in accounting means assets and liabilities; 'dependent variable' means the owner's equity, a residual amount that can be calculated only when the assets and liabilities have been determined. In statistics these terms have a different, well-known literal meaning in regression analysis.
4. There are further differences in their philosophical outlook that has kept accountants and statisticians apart. The accountant is supposed to be factual in the basic reporting of the events within his purview. He leans heavily on 'generally

accepted principles' to achieve this. He also has direct access to the facts – documents, that in the preceding chapters were referred to as the 'statistical-counting-units' – that become his data. Occasionally, he also may have to resort to sampling, e.g. when faced with a recalcitrant management.

The statistician on the other hand, does not usually have his facts as readily available as the accountant, and often must devise ways to overcome this handicap within given financial constraints. Nor does he view these facts as straightforwardly as the accountant, for he was taught to view socio-economic reality as if through a veil of random distortions.

In addition, statistics is not as closed a system like accounting, but rather a sequence of loosely related methods that were developed in response to needs in different disciplines, at different times. Even the terminology for each of these sub-areas of statistics is different, reflecting this lack of inner cohesion. For instance, consider the arithmetic mean. Depending on the context in which it is used, it is referred to as 'an average', as the 'arithmetic mean', as $\bar{x}$, μ, λ, E(x), or even as just a ratio.

The isolation and separate development of statistics and accounting was not a coincidence but the result of a long history of separate pursuits of their quantitative tasks, of different basic philosophies that developed out of different problem settings, and of different demands from the users of their output. Both fields have rallied around the consolidation of their separate professional ways.

11.2 Areas Common to Statistics and Accounting

Statistical studies in which financial data were gleaned from accounting records, balance sheets and income statements, are not what we are concerned with here. We are looking at applications of genuine statistical procedures in routine accounting work, of statistical reasoning in accounting, or simply of finding areas that statistics and accounting have in common, even though both sides may not have paid attention to that fact. In a number of instances, accountants acknowledge statistical methods other than sampling as belonging to statistics, such as ratio analysis and forecasting But instances in which the commonness has not been properly recognized are the areas of special interest in this chapter, particularly where accountants are treading on statistical territory, or if you prefer, statisticians have proceeded like accountants.

11.2.1 Common Areas that Are Recognized

11.2.1.1 Ratios

Accountants operate with entire networks of ratios,[7] either calculating them explicitly, or only mentally estimating percentages to make rough comparisons. Accountants refer to many of these as 'statistical ratios.' It is unfortunate that newer textbooks in business, economic and social statistics hardly discuss ratios, which is a serious omission in view of their great importance in applied work.

11.2.1.2 Time Series

Accountants rely in their daily work on simple procedures for the analysis of time series. They apparently have no use for the abstract, sophisticated time series models developed by statisticians. The reason is not a lack of sophistication on the part of accountants, but differences in perception of the facts they deal with in their respective environments. More dialogue in this area between statisticians and accountants would be constructive. The hard-nosed approach accountants take to their economic reality is closer than that of statisticians who perceive it as if through a veil of probabilities. Statisticians in the social sciences would benefit from adopting some of accountants' sober sense of reality. Then, time series analysis and other statistical methods will become of greater interest to accountants.[8]

11.2.1.3 Forecasting and Other Estimation Procedures

These are other areas where accountants routinely perform statistical tasks. The development of the annual budget requires estimating in advance the expected receipts and payments of cash, the estimates of cash flow are naturally based to some extent on the ledger accounts showing past cash receipts and payments. The 'aging of the accounts receivable' is another example of routine forecasting done by accountants. The 'balance sheet approach' and the simpler 'income statement approach' are really standard statistical estimation procedures.[9] The basics of regression analysis used in accounting should be mentioned here. In all forecasting applications, the 'going concern' concept of accounting also underlies all work done in business and economic statistics.

Both fields, accounting and statistics, use estimation procedures to fill gaps in their reported information. Accountants prefer simple ratio-based estimates, using analogue and past experience estimation procedures. Here also, statistical achievements could be put to good use in accounting such as e.g. the Bureau of Census' computerized 'hot deck matrix' estimation approach.

11.2.1.4 Grouping, Aggregation and Classification

Problems of this kind are as important in socio-economic statistics as in accounting. 'Accounting relies on classification as an indispensable part of its analysis ... accounting separates the mass of raw data into categories. This is a significant analytical process in itself and facilitates further grouping, associating and interrelating of the classified data.'[10] and: "First we record business events as they occur, second we classify these events into groups so that the mass of detailed information will be in compact usable form, and third we summarize the classified information into ... financial statements ... the ultimate objective of accounting is the use of this information, its analysis and interpretation. The accountant is always concerned with the significance of the figures he has produced ..."[11]

The problems of interpreting aggregates that result from grouping individual journal entries into accounts, or from grouping 'statistical-counting-elements' into

statistical classes are basically the same. Such aggregates, though, are more difficult to analyze in socio-economic statistics because the counting elements often cannot be expressed in a common monetary unit,[12] as is customary in accounting. In the area of group formation, statistics, unfortunately, does not have much to offer. Accountants seem to have developed the art of grouping – aggregation and de-aggregation – to a higher degree than statisticians. The accounting theory that is concerned with clarifying whether a transaction is to be considered as a liability or an asset, an income or an expense is concerned with classification. So is the theory that recommends that accounts receivables with a credit balance (overpayment by customer) should not be netted against the usual debit balances of accounts receivables, but be treated as a liability. Accounting theory exhibits good sense from the point of view of aggregation.

11.2.1.5 Index Numbers and the Adjustment of Changes in Purchasing Power

Despite a steady erosion in the purchasing power of the dollar in the US over various decades, accountants prefer to assume that the value of the dollar is stable.[13] Accountants, though, are familiar with statistical 'price-index-numbers' and with procedures to adjust for changes in purchasing power. Their reluctance to use this statistical tool more freely – despite official endorsements[14] is an indication of their good sense to mistrust this inadequate tool. Stating that the use of 'price-index-numbers' in accounting has a greater potential than is presently recognized does not properly account for the unsolved problems of interpreting the 'adjusted' values. There is still much clarifying to be done by statisticians before accountants' rightful skepticism will have been assuaged.

The use of these five method-areas by accountants could be expanded in each area if statisticians would explore the analytical needs of that constituency and also take a fresh look at the basics of socio-economic statistics.

11.2.1.6 Sampling

Any form of recognized random sampling that has legal standing in a court of law. It may be the only statistical method that accounting recognizes as not a method of its own. It is used in many applications and often is left to be performed by a statistician.

11.2.2 Common Areas that Are Not Recognized

11.2.2.1 Bookkeeping

This part of accounting activities is essentially a statistical activity. The vast system of controls – e.g. the subsidiary ledger system for control over plant and equipment to prevent losses by theft or waste – over every change in the status of a firm's property and performance is basically a statistical recording system. The resulting accounting records – often the source for statistical studies – are not just another source outside of statistics, but are statistical data themselves.

All who take statistics as part of their education in the social, economic and business fields hardly learn about this important surveying and recording phase of statistics. This phase is omitted as theoretically unrewarding, apparently also for lack of a conceptual foundation. The newer textbooks of statistics have little to say about the nature of our data, of data collection and of aggregation. As an academic discipline statistics has become divorced from the socio-economic phenomena that are to be studied with the help of data. Accounting, in contrast, has preserved this relationship between the data and the phenomena to be described, especially with the concepts of 'net gain' and 'net earnings.' That is why the recording activities of accounting have not been recognized by statistics as essentially one of its own tasks.

11.2.2.2 Double Entry Bookkeeping as (m)∗(n) Statistical Tables

The similarities with statistics become more evident if all bookkeeping entries, which are made pair wise in separate ledger accounts, are visualized as statistical entries in one of various two-dimensional contingency tables. The columns of such two-way classifications represent e.g. all the customer accounts while the rows represent the operational accounts indicating the mode of payment. Every accounting entry would be matched by a tally in the respective cell of such a contingency table. The accountant is usually interested in the dollar value of an event although on certain occasions he also may be interested in the unit count. The statistician keeps count on both, unit count of such events, and value for each cell and account group. Events which reduce an account could be marked by an entry with a minus sign in the same statistical table, or another identical table could be set up in which only reductions of a ledger account are entered without regard to whether this was an accounting debit or credit. At the end of the accounting period, the entries in every cell in both tables are then consolidated (netted) and the net figures for every subtotal determined – row and column wise – indicating the state of affairs at the end of that reporting period. The point here is to show that statistics, in principle, has coped with the general problem of double classification for which accounting has developed a highly specialized, more efficient solution.

11.2.2.3 Agreement on Basic Working Principles

The following concepts or assumptions including, 'consistency,' 'objectivity,' 'conservatism,' 'obsolescence,' 'materiality,' 'disclosure,' and 'accounting entity,' that have been formulated as basic accounting principles, are also important working principles in statistics. Following is a brief review of these concepts.

a. **Consistency** implies that a particular accounting method once adopted, should not be changed from one period to the next. This requirement is important because it enables users of financial statements to intelligently interpret changes in financial position.[15] Such methodological continuity is just as important and desirable in statistical work. Survey procedures and methods of analysis must not be changed inadvertently to maintain comparability over time. This is the reason

why e.g. in 'price-index-numbers' the 'shopping basket' of goods that are to be priced monthly, is maintained as long as possible, to the detriment, though, of a more realistic representation of changes in the market situations.

b. **Objectivity** is essentially a term which allows for some reasonable latitude in the quality of the evidence. Although the cost of a depreciable asset may be objectively determined, judgment is needed to determine the depreciation method and the period depreciation expense that should be used.[16] Statisticians have, in principle, always followed this tenet in their work, as far as humanly possible. The rule of not publishing statistical information of less than three business firms in any region or aggregate, as well as all other rules to guarantee the privacy of statistical information is to safeguard the objectivity of results.
c. **Conservatism** in accounting refers to the reasonable expectation by creditors and stockholders that the reporting should not be unduly optimistic.[17] There are many areas of statistical reporting – e.g. price index numbers or employment data – where de facto conservatism has evolved. The external political pressures on statisticians to be cautious are quite comparable to those pressures that advise accountants to be conservative.
d. **Obsolescence** relates to the capacity of a plant asset to render services to a particular company for a particular purpose. . . it is probably a more significant factor than physical deterioration in putting an end to the usefulness of most depreciable assets.[18] Although statisticians seldom deal with this kind of obsolescence, there is a strong analogy with regard to statistical data which also become obsolete. Statistical theory has not paid attention to this problem, nonetheless, statisticians have dealt with it in a few ways on a practical level by speeding up the publication of provisional socio-economic data that later are revised. The custom to limit the published time series data to only a few recent periods implies an awareness of the obsolescence of data. This valuable accounting concept has been applied in statistical practice without perhaps an awareness that it was dealing with obsolescence. It is a matter that ought to be recognized by statistical theory, particularly in time series and forecasting.[19]
e. **Materiality** is the reasonable expectation that knowledge of the item would influence the decision of prudent users of financial statements.[20] This is what statisticians have been doing in weighing the cost of surveys against sample size and the inclusion of additional features in a questionnaire. There is also a difference between a 'statistically significance' of a sample result, and the actual importance or 'materiality' of that result. The difference between two arithmetic means, for example, could be statistically 'highly significant' if the samples are very large, but of no practical consequence, that is 'not material' in the social or economic context.
f. **Disclosure of relevant information** includes non quantitative, narrative type information that is to accompany the figures such as terms of borrowing arrangements, contractual provisions, accounting methods used in preparing financial statements, changes in method and other significant events such as a strike, new legislation pending, problems of procuring raw materials, etc.[21] Disclosure is an important principle which at present is not sufficiently heeded

by statistical theory. Head-notes and footnotes in statistical tables in general, and of tables of time series in particular, are considered of little interest to statistical theory. Yet statistics has for a long time coped with the legal and political problems of 'disclosure' – really the opposite of it, maintaining the privacy of information. Apart from laws that aim at guaranteeing the privacy of detailed information in statistical surveys of persons and businesses, statistics also limits the published detail to at least three business firms in a region or industry grouping.

g. **Legal Entity** is the 'Accounting Entity' and 'Cost Centers' that are the organizational levels at which reporting takes place. Each of these concepts has an exact counterpart in statistics. Occasionally data are gathered and published within a business firm at the level of the 'corporation,' at the level of each one of the 'establishment' within a corporation, and at the level of each one of its 'producing units' or 'departments' within an establishment.

All of these accounting principles are also valid in statistics as they guide the quantitative description of these socio-economic phenomena.

11.3 Concluding Observations on Statistics and Accounting

While the common ground between accounting and e.g. economics or tax law, as well as the common ground between statistics and e.g. operations research or econometrics has been explored, little had been done concerning the relationship between accounting and statistics. This chapter pointed to areas where accounting is using statistical sampling as well as other statistical methods and concepts, a matter that is not obvious because of the separate historic developments of these two quantitative disciplines. Yet these two quantitative disciplines have a great deal in common, more than either has been aware of. A dialogue between statisticians and accountants could benefit the perception of socio-economic reality. Both fields can learn from each other. Statisticians, in particular, would be reminded that interpretation is the ultimate purpose for economic and social data, a matter that ought to be part of statistical theory. At present statistical theory for the social sciences has developed and is geared to the data in the natural and engineering sciences, that deal with 'measurement data' and phenomena of a different kind. Not much of the present theory applies to the bulk of socio-economic data. The analyst trained in mathematical statistics may feel uncomfortable since his training has not prepared him to deal with our kind of data. There is a real need for a change of attitude.

In contrast, accounting is fully aware of its fact-recording function. It has developed systems to register every purchase, sale and other relevant fact, their 'real-life-objects', then classifying and grouping them as the corresponding book entries, the equivalent of the 'statistical-counting-units', then drawing up a comprehensive picture of the firm's status and internal flow phenomena. For statisticians it is wholesome to keep in mind that description by counting and valuing – thus

accounting for certain economic and social phenomena – has started statistics on its own development, and still makes up the bulk of its practical work today. The comparison with accounting also highlights the extent to which statistical theory has been caught up in sampling and inference, and lost sight of its original descriptive tasks.

It appears that statistical theory has not developed the appropriate concepts for those areas where statistical methods have been used in accounting. Although the notions of probability do not necessarily seem appropriate for reporting financial statements, accountants are moving away from the earlier belief that their valuation procedure yielded a unique, true picture of the firm's business performance during a period. It is not only probability that accountants may find useful, but also the concept of uncertainty in aggregated data, discussed in earlier chapters.

Although statisticians have valuable professional associations, such as ASA, these are not functioning like their accounting counterparts, the AICPA, the APB (now FASB), or the SEC, that codify and standardize the principles of accounting. The statistical organizations have not had the effect on the statistical profession as these professional accounting organizations have had on their field. A closer contact between statisticians' and accountants' organizations in the future, without posing a threat to either professional group, could prove beneficial for the advancement of both disciplines. Government agencies may act as a catalysts, but the work will have to be done by academicians.

Notes

1. This chapter follows closely Winkler, Othmar W. "Statistics in Accounting other than Sampling" *Proceedings of the Business and Economic Statistics Section of ASA,* Washington, D.C. 1976, pp. 654–659 as well as a modified version of this article entitled "Secret Allies?" *Management Accounting,* National Association of Accountants, Montvale, NJ, June 1985, pp. 48–53.
2. When asked to teach the graduate course "Statistics for Accounting" in 1975. It soon became clear that this was to be a course in audit sampling with a smattering of general statistical inference. I used the textbooks by Herbert Arkin, *Handbook of Sampling for Auditing and Accounting,* 2nd ed. McGraw-Hill Book Co. New York, 1974, and Robert W. Vanasse, *Statistical Sampling for Auditing and Accounting,* 2nd ed. McGraw-Hill Book Co. New York, 1968, supplemented by Daniel, Wayne D., Terrell, James C. *Business Statistics: Basic Concepts and Methodology*, Houghton Mifflin Co. Boston, 1975. Other books dealing with statistics in accounting, conceive of statistics narrowly as a tool for audit sampling only.
3. Winkler, Othmar, Unterschiedliche Ansaetze zur Wirtschafts- und Sozialstatistik in Ost und West *Jahrbuehcer fuer Nationaloekonomie und Statistik*, Band 208/5, pp. 459–492 Gustav Fisher Verlag, Stutgart, Gemany, Sept. 1991.

   It is of interest to draw attention to the fact that in the former socialist countries of the east bloc, statistics and accounting in the industrial enterprise were considered inseparable parts of management planning and control. This difference in approach is expressed in the titles of some of their publications. *Statistische Praxis* to name one example, formerly the official statistical journal of the DDR (Deutsche Demokratische Republik – East Germany) was subtitled, *Zeitschrift für Rechnungsführung und Statistik* (Journal of Accountancy and Statistics.)
4. AICPA Professional Standards, ACS Section 1023.01, Bulletin No. 3, p. 7201, Oct. 1970.

5. Meigs, Walter B., Mosich, A.N., Johnson, Charles E. *Accounting the Basis for Business Decisions*, 3rd ed. McGraw-Hill Book Co., New York, 1972, p. 2.
6. Lapin, Lawrence L. *Statistics for Modern Business Decisions,* The Harbrace Series in Business and Economics, Harcourt, Brace, Jovanovich, Inc., New York, 1973, p. 2.
7. Evidence of this is the popularity of publications such as Dun & Bradstreet's annual *Key Business Ratios* or Robert Morris Associates' *Annual Statement Studies*.
8. For an extended discussion of this topic see Chap. 5 of this book, "Longitudinal Analysis, Part 1 – Looking to the Past".
9. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 33.
10. DeMaris, E.J., Fess, P.E., Moyer, C.A., Perry, K.W., Wyatt, A.R., Zimmerman, V.K., Mautz, R.K. *A Statement of Basic Accounting Postulates and Principles*, Center for International. Education. and Research in Accounting., Univ. of Ill., 1964, p. 28.
11. Meigs, Walter B., Mosich, A.N., Johnson, Ch. E. op. cit. p. 2.
12. See Chap. 7 of this book "Longitudinal Analyses, Part 3 – Index Numbers" and Chap. 3 on Aggregates.
13. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 404.
14. "Financial Statements Restated for General Price-Level Changes," Statement No. 3, AICPA 1969. Here also belongs a new SEC requirement that certain large publicly held companies publish such data as a supplement, starting in 1977.
15. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 404.
16. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 403.
17. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. pp. 406–407.
18. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 335
19. See Chap. 6 of this book on Forecasting.
20. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. p. 406.
21. Meigs, Walter B., Mosich A.N., Johnson, Charles E. op. cit. pp. 405–406.

Chapter 12
Socio-Economic Statistics and Geography

12.1 A First Assessment

Location is not only a prime determinant of market value for real estate property, but more generally is important in all socio-economic phenomena. Every individual case, that is, every 'real-life-object' and its corresponding 'statistical-counting-unit' in a statistical data base must contain detailed information about its geographic location. Although location is usually not one of the properties of the 'statistical-counting-units', their geographic location becomes important when incorporating them into an aggregate.

The practical need for reliable geographic work in statistical surveys has long been recognized. Statistical theory, however, has ignored this geographic feature of its phenomena, proceeding as if human activity were not place-bound.[1] This attitude dates back to the time when the sciences, which inspired the paradigms of present day statistical theory, were not yet concerned with location-related ecological phenomena.

On the other hand, research in geography has developed in a quantitative, particularly statistical direction. Geographers, city planners and regional economists are increasingly relying on statistical data and methods for their work. Statistics and geography overlap, although geography as a related discipline does not come to mind when considering subjects that border on or even overlap with statistics. Although the geographic component of statistical data was discussed in earlier chapters, the relationship between business, economic and social statistics with economic, social and political geography needs further clarification.[2] The prominent geographic component in all socio-economic statistical work requires that a theory of this field of statistics at least must recognize this geographic component. At present, textbooks of business and economic statistics completely ignore it. This, unfortunately, has resulted in the neglect of the international aspects of statistics which at present is treated as a timeless and place-less mathematical discipline. Statisticians are also missing a great opportunity to make a valid and unique contribution to the internationalization of business and economic curricula.

Of the two major fields of geography – physical geography and human geography – mostly the latter is of concern here.[3] It includes economic and social geography, political geography, (e.g. the geography of federal spending in a county),

O.W. Winkler, *Interpreting Economic and Social Data*,
DOI 10.1007/978-3-540-68721-4_12, © Springer-Verlag Berlin Heidelberg 2009

historical geography (the 'human impact'), behavioral geography and political studies from a geographer's perspective.

Books such as 'People in Durham: A Census Atlas'[4] and statements by geographers indicate that no clear boundaries between statistics and geography seem to exist. Let geographers speak about this matter: "the Negro question is taken up mainly through the voice of individuals and the testimony of statistics . . . the clash of cultures . . . by way of the regional novel supported by impersonal surveys . . ."[5] "This 'Synthesis of information collected from a wide variety of sources' adopts a systematic approach to the human geography of Western Europe . . . stress is laid on both temporal and spatial variation in human activity and on the processes responsible for the variations . . ."[6]

"Three questions are central . . . where is the economic activity located? what are the characteristics of the activity? . . . to what other phenomena is the economic activity related? . . . the student is introduced to various techniques . . . location quotients and the index of diversification . . . It is a pity that so many of the maps are drawn using state, rather than county data, as in the first edition. . ."[7]

Various statistical methods have been devised to deal with distributions of points, lines and areas on the terrestrial surface. It would be important to incorporate them into business and economic and social statistics. Maps as models of the earth's surface are the language common to statisticians and geographers. Yet, their interests differ. Statisticians construct choropleth maps as an added means of displaying regional data. Geographers use statistical data to add to their maps a stronger sense of socio-economic realism. Unfortunately, geographic maps are seldom found in textbooks of business and economic statistics – as if location were irrelevant for statistics, and hence, for its theory. Geographers on the other hand, use census results and areal distributions. They show concern for the non-sampling characteristics of socio-economic data, and for the meaning of statistical aggregates.

12.2 Statistics in Geography

12.2.1 Using Statistical Data

For the statistician, unfamiliar with the work of geographers, the use of statistical data in geography is amazing. One can find for example, geographers discussing at length the taking of a population census.[8] In one study, the author used data of the population censuses of 1938, 1951, 1964 and 1973, to compute detailed rural-urban growth differentials for each area[9] in a study of population changes in Colombia. Another geographic study ". . . focuses upon the data sources . . . with facsimile reproductions of some of the data sources. . . It is refreshing for the critical comment on the reliability of individual data sources. . . The book demonstrates the wealth of readily available material for project and practical work in urban geography. . ."[10] ". . .details abstracted from the population register can prove invaluable in tracing the turnover of populations and provide a basis for population density, migration and

community structures . . .”[11] “. . .the author prefers the telling phrase or quotation and the hand specimen to the statistical overview. . .”[12] Or “there is little reference to recent work in industrial geography and hence questions concerning patterns of job loss, or locational change within large multi-plant firms, are largely ignored.”[13] “. . .It is impossible to state with accuracy the number of industrial estates in Britain, mainly because of the variety of definitions which are used. For the same reason, it is difficult to obtain time-series data on industrial estate development. . .the estimate of 500 industrial estates nationwide is a much too conservative figure . . . a figure nearer 900 seems more likely”.[14] Concerning a Census Atlas: “. . . This . . . is a pioneering case study of how to cope with huge amounts of spatial data . . . It pioneers not only the computer techniques but also the ways in which distributions are mapped. A signed chi-square statistic overcomes the statistical skewness of many census variables, so highlighting high or low distributions set against national norms . . . using 1 km grid squares comparative studies using a consistent spatial framework . . . is the innovatory and exciting first of a series of national population atlases that will at last have a substantive research use.”[15]

These quotations from the work of geographers concerning their use of statistical data are instructive for those who were unaware of the widespread use of statistical data and methods in geography. In another study, 370 large industrial firms in south-western Germany, Switzerland and northern Italy were surveyed and the results analyzed.[16] Or: “It pays. . .only limited attention to various key questions. . .including the question as to whether the geographical pattern of regional economic development bears any relationship to the . . . hypothesis of a cumulative concentration of economic activity . . . relative to its periphery. A . . . source of frustration is the fact that in a book of over 400 pages, the actual text accounts for less than half, the remainder comprising statistical and other appendices . . . The chief strength of the book lies in its compilation and analysis of an original data set covering population, employment and GDP (gross domestic product) in 76 regions . . . for the years 1950, 1960 and 1970 . . . From these are derived ten and twenty year growth rates for the absolute figures and for ratios between them leading to a two-level shift and share analysis. Maps and dispersion graphs illustrate the main features of the interpretation. The analysis . . . reveals significant findings . . . that regional differences in relative wealth, measured by GDP per head of the population, decreased . . . over the study period . . . in each of the larger member countries. . .Molle’s book presents statistical findings of considerable value. . .”[17]

Another review states: “(this) reviewer . . . was attracted . . . to the account of the economic infrastructure and the . . . illuminating chapter on tertiary industries and tourism. The human geography . . . is based upon statistics up to 1968 or 1969 . . .”[18]

Another book: “examines selected human aspects only . . . data for the 70s . . . the new (1969–1976) administrative structure, the demographic reversal of the 70s, the effects of the oil crisis, the social-residential effects of affluence, the effects on regional development of state intervention . . . the changing distribution of regional and urban populations . . . amply supported by maps and tables . . .”[19]

“An international text dealing with structure and organization of retailing on a global scale. . .a. . . .major contribution to the growing literature on marketing

geography and the emergence of this subject as a distinctive sub-branch of the discipline. . .Many of the statistics taken for the US are outdated but the comparative theme is not unduly affected by this and the active researcher will shortly be blessed with a new Census of Business . . . little reference . . . is given to recent models of retail location. . .the role of changes in consumer behavior, perception research and the impact of planning policies, particularly through redevelopment programs. . ."[20]

This cross section of recent geographers' work and other geographers' critical opinions about it conveys an idea of the great importance they attribute to statistical data and procedures.

12.2.2 Using Statistical Methods

In the following, geographers tell us how they view statistical methods as they apply them to their work:

"Scattergrams, correlation and regression have become part of the stock in trade of geographers in the past twenty years . . . trend lines are usually assumed to be linear . . . often . . . variables have 'curvilinear' rather than linear relationships and . . . the nature of these relationships is at least as interesting as anything a correlation coefficient may reveal. (p. 126) . . . it should be understood that the methods of running medians for small data sets with the horizontal scale ranked, and column medians for large data sets . . . are entirely pragmatic and do not rest on any foundation of mathematical theory. The lines . . . have been called trend-lines rather than regression lines, . . . they are not used to derive correlation coefficients and their corresponding levels of significance; nor can they be used to estimate the statistical significance of residuals . . . measures of statistical significance lose some of their value (sic!) when applied to whole populations. . ."[21]

Another geographer used the Mann-Whitney test (improperly) to test statistically for the presence of 'contagion diffusion'.[22] In the study of urban geography "techniques available for testing hypotheses and theories concerning urban structure . . . provides an integrated treatment of the . . . important aspects of the subject in which statistical techniques are learnt in the context of the search for solutions to real problems . . . and worked examples to illustrate the statistical techniques . . ."[23]

The (mis)use of probability theory by geographers reflects the careless handling of this topic on the part of statisticians. "The chapters on probability theory and spatial probability models stray from model-building per se to treat analysis via classical statistical testing . . . the chi-squared and t distributions are not fully explained . . . the difference between model-building and model testing is unclear. . ."[24]

The geographer P. Lewis interprets probabilistic (statistical) statements for geographic work as follows: "When we speak of a map being a realization of a random procedure, we are not implying that there is something that allocates factories or people to sites in some region. We are suggesting that we use some simple model to establish what sort of distribution would arise if such a random procedure were used experimentally. This then provides us with some tangible sense of random without supposing that human behavior is manipulated by some procedure in the same way

in which we operate the experimental procedure. Nor are we saying that judgment and decision are irrelevant to behavior . . . There is no prototype of random. Random variables can take on values in accordance with a variety of constraints"[25]

L.J. King is more outspoken: ". . . the more important consideration is whether or not a random distribution of settlements has any real geographic significance. In the North Dakota area. . ..the critical fact appears to be that the settlements are located along three major transportation axes which have an approximate east- west orientation rather than that the distributions statistically random . . . *the concept of randomness with respect to settlement patterns might well be disregarded*, *except* for the fact *that the value of* $R = 1$ ***is a convenient*** . . . ***origin*** from which ***to measure*** the tendencies toward an ***aggregation or uniform spacing of settlements*** . . ."[26] (highlighting added for emphasis)

King's observations make an important statement about probability models in geography, denying them legitimacy for geographic research on human settlements. His remarks deserve further comment.

Assume a geographer wishes to determine whether a given regional distribution of store locations is random. It is an unrealistic assumption, yet deep in their heart, researchers will not accept as the final conclusion of their efforts, that some distribution on the earth's surface has indeed occurred for no particular rhyme or reason, just 'by chance.'

Statistical tests for randomness allow for two possible outcomes: either one can reject the hypothesis that this distribution is random (e.g. when these stores are highly concentrated in a few clusters) or not reject it.

Suppose, first, that the distribution of stores in a region clearly is not random. Then the null-hypothesis 'the distribution on the map follows a chance pattern' is rejected, say, at the 5% level of significance. This outcome suggests to the researcher, that he is now encouraged to search for the real causes for this particular distribution, for causes other than 'chance.' He is also warned at the same time to remain alert to the small probability – smaller than the chosen 5% 'level of significance' – that even this seemingly non-random geographic distribution of store units, although not very likely, could have come about by some chance mechanism.

Then consider the case where the given distribution of store locations does not show a clearly discernible pattern, and the statistical test indicates that the hypothesis, 'these stores are distributed at random' cannot be rejected. Does that mean that these entrepreneurs have drawn random numbers to determine the best location for their stores? Not likely, in view of what is known from the ample marketing research on store location. What, then, does such a finding mean?

It means, that a geographic distribution, like this one, could be imagined as having been selected by a chance mechanism 'at random' from among the innumerable, theoretically possible alternative placements of these dots (stores) on the map. It does not affirm that the actual geographic distribution of these stores was really produced by some known chance process. Nor does it say anything about the reasons why a given dot, or group of dots, landed on that particular place on the map.

Next, consider that much geographic and other detail of the 'statistical-counting-units' has been eliminated at this level of regional (and other) aggregation. Little

remains, therefore, that would allow one to trace the finer detail of the actual circumstances that contributed to a particular placement pattern. Superficially speaking, the situation is not very different from the case where actual location of the points on the map was determined by random draws for the values of each (x, y) coordinate on the map. The contribution of this part of statistical theory to geographic analysis, other than alerting the geographer to such possibilities – is negative, not just 'not positive.' In addition to facing the fact that no regional social forces are revealed as acting on his distribution, statistics implicitly advised the geographer to stop searching for the reasons of a specific regional distribution. Statistics, with its 'scientific' mathematical apparatus, advised that the placement of the stores, studied by the geographer, could as well have been produced by a chance mechanism. From then on, there is no encouragement to do further research that would be worth his effort. Statisticians should make it clear to geographers that speculations in probability do not contribute anything of social or economic relevance. Yet, if statistics is not narrowly identified with the calculus of probability, then statistics can make valuable contributions to the discipline of geography.

12.3 Geography in Statistics

Brazil is the one country that officially acknowledges the need for a close working relation with geography in statistical work. The name of its central statistical office is 'Fundaçao IBGE' (. . .**I**nstituto **B**rasileiro de **G**eografia e **E**stadistica). Geography and statistics both deal with facts located on the earth's surface. Statistical practitioners know that the socio-economic phenomena which they are capturing statistically, have a location in 'real geographic space.' It is this important, non-probabilistic feature which links geography with socio-economic statistics. Statistical theorists, on the other hand, have remained by and large, oblivious to the geographic dimension of socio-economic statistics. While much has been written by statisticians with regard to the 'time dimension' of its data, very little has been written by statisticians about the ever-present geographic dimension.[27]

Statistical theory tends to present its distributions abstracting from the regional component. Geographic (dis)aggregation, however, places socio-economic phenomena into their realistic context, maintaining and restoring a measure of realism to the data. The contributions made by regional breakdowns, and measures of proximity and distance between these geographic groupings, which human geography can provide, contrast favorably with the dubious contributions of the probabilistic analysis of statistics to academic geography. Statistical theory completely disregards geographic location, a dimension needed to restore some realism to the probability orientation of its theoretical models.

This situation is reflected in the fact, and not by coincidence, that more has been written about statistics by geographers, than about geography by statisticians. Statistical theorizing has obviously lost sight of the fact that statistics in the social sciences, consists for the most part of conceptual definition and classification of its socio-economic phenomena and objects in economic and geographic categories.

12.4 Differences Between Statistics and Geography

According to geographers, statistics does not apply to certain parts of geography:

"...The standard economic atlas of the distribution of population, minerals, agricultural and manufacturing activity encourages the impression that economic activity is merely a technical affair: in this country, there is copper, ... in the next manufacturing industry. The 'State of the World Atlas' makes it clear that the social organization of economic activity is crucial. Its maps show not just how much output is produced where, but who profits from it and whose labor produces it; not just oil and food output, but oil and food power ... particularly striking are the maps of tax havens and free production zones ... Geographers have a...record of attributing to nature or technology what is attributable to political and social organization ... this (atlas) prompts the reader to think about ... the political disputes in and against states, military governments, foreign military presence, refugees...things which appear in the news but not apparently in the minds of geographers."[28]

Another geographer comments: "... The book takes its origin in 'the widespread reassessment of the positivist movement in science'... it reaches the conclusion that 'the age of positivism is dead'...traces the empirical background of the antecedents of geography and how the subject fits into the course of scientific and philosophical development from the time of Bacon and Newton to those of Humboldt and Ritter. It outlines the swings in intellectual fashion..."[29]

Finally: "American urban geographers have ... with the aid of quantitative methods searched primarily for universal laws, regularities, and order in the spatial structure of urban distribution and patterns. Tempted by the common euphoria of the 1950s and 1960s which was the era of hard science in America, geographers believed that urban phenomena could be studied in basically the same way as phenomena in the natural sciences and that the scientific approach and methodological procedures of the natural sciences could be adapted directly to urban geography. It was also hoped that this approach would help to overcome the lack of respectability ... of human geography ... and of urban geography in particular ... It has ... made urban geography too abstract and general, because it has taken much of the 'geography' out of the discipline, since most ... of the cultural-genetic variations of cities throughout the world had to be 'purposely sacrificed' in order to arrive at general similarities, regularities, and order in urban spatial structure ... These cultural-genetic linkages are rather difficult to understand and measure ... intangible elements such as values, perceptions, traditions, preferences or dislikes of society, none of which can easily or not at all be transferred into numbers and dealt with by the computer. The workings. are manifested in the form of laws or regulations, societal choice patterns, and document-able historical records including the works of poets, philosophers, painters, and other artists...a growing number of urban scholars has begun to ... use such records to explain the geographically relevant regional differences of urban phenomena ... a more sophisticated education is needed to engage in such projects than has been displayed by most of the quantifying urban students ..."[30]

A different problem is "In geography ... quantitative efforts have focused on statistical testing, but in many cases the underpinnings of those tests are weak..."[31]

Finally there are topics that interest geographers but not statisticians: 'The chapters overall deal mainly with a review of central place theory, regional hierarchies of centers, trade areas and the role of markets and fairs, and contrasts between the structure of retailing in urban, rural and suburban areas.'[32]

12.5 A Revised Assessment

It is evident from the quotations that statisticians and geographers are interested in the same socio-economic facts. Unfortunately, this common interest has not been translated into corresponding adjustments of statistical theory. What today is taught as statistics for the students of business, economics and the social sciences lacks an awareness of geographic location and ignores geography as a closely related academic discipline. In contrast, geographers seem fully aware of the 'descriptive' potential of socio-economic statistical data and cope reasonably well with statistical reasoning by introducing inference and models of geographic random distributions, testing the hypothesis that an actual geographic distribution of stores or cities is random.

Geographers describe and analyze regional distributions of society's activities, use statistical methods, and have contributed statistical methods of their own. Statisticians, on the other hand, are charged with the collection and presentation of these data, use maps for display and descriptive analysis of socio-economic facts. Statisticians face the regional aspects of their data reluctantly preferring by far to deal with the more tractable quantitative variables.

Neither statistics nor geography is clearly defined against each other. Probability-based methods and the concept of 'random' – except for random sampling – are irrelevant for geographers, contributing nothing to the understanding of real situations that are of interest to them. Geographers are more directly concerned with the nexus of their data with the socio-economic reality of regionally dispersed phenomena than statisticians, who do not seem to care about that nexus. The latter are absorbed in technical questions of data collection and inference, but show little concern for the connection of their numbers with reality, and the validity of their assumptions.[33]

Statisticians and geographers also differ in their educational backgrounds, and in their professional lingo. These differences are more a matter of emphasis than of substance. Both have to continue toward a fuller understanding of the simultaneous area-time-subject matter dynamism of society.

Notes

1. The term 'spatial' which is customary among geographers is avoided in the following because it now refers to the growing number of activities in interplanetary or outer space. The location or 'rootedness' of objects on the earth will be referred to as 'locational', 'areal', 'regional', or simply as 'geographic.' Although the earth's surface is not flat, for many statistical and

geographical purposes it can be treated as if it consisted of level areas on which the location of socio-economic facts can be determined.

2. An interesting attempt by a geographer was made concerning the boundaries between geography and psychology, somewhat along the lines of this chapter: Golledge, R.G. "Substantive and Methodological Aspects of the Interface between Geography and Psychology" *Proximity and Preference-Problems in the Multidimensional Analysis of large Data Sets*, R.G. Golledge, John N. Rayner, Univ. of Minnesota Press, Minneapolis, 1982, pp. xix–xxxix.
3. Although questions of land conservation, climate and other topics of physical geography may also concern social scientists. Air photography and satellite reconnaissance of surface cover, the Landsat series in particular, has brought the work of geographers and statisticians closer together. Important seasonal changes in planting can now be sensed directly and transmitted through space-born multi spectral scanners. See e.g. Allen, J.A., "Remote Sensing in Land, and Land-use Studies," in: *Geography*, Vol. 65, pt. 1, 1980, pp. 35–42.
4. Review by B.E. Coates, *Geography*, Vol. 62, pt. 1, Jan 1977, p. 58, of J.C. Dowdney, and D.W. Rhind, editors, Durham: University of Durham Census Research Unit, 1976.
5. Review by J.A. Edwards, *Geography*, Vol. 67, pt. 1, Jan 1982, p. 86. of J.W. Watson, *Social Geography of the United States*, London: Longmann, 1979.
6. Review by S. Hanslip, *Geography*, Vol. 67, pt 2, April l982, p. 167, of B.W. Ilbery, *Western Europe: A systematic Human Geography*, Oxford University Press, 1981.
7. Review by Davis, J.F. of J.W. Alexander, L.J. Gibson, *Economic Geography*, London, Prentice Hall, 2nd ed. 1979, in *Geography*, Vol. 64, pt. 4, 1979, pp. 352–3.
8. P.T.H. Unwin, "The census of India 1981", *Geography*, Vol. 66, pt. 3 1981, pp. 221–2.
9. Williams, Lynden S., Griffin, Ernst C. "Rural and Small-Town Depopulation in Colombia", *The Geographical Review*, Vol. LXVIII, 1978, N.Y. pp. 13–30.
10. Review by M.B. Gleave, *Geography*, Vol. 66, pt. 1 Jan 1981, p. 73, of J.R.Short, *Urban Data Sources (Sources and Methods in Geography)* London: Butterworth, 1980.
11. R.L. Gant, J.A. Edwards, "Electoral Registers as a Resource for Geographical Inquiry", *Geography*, Vol. 62, pt. 1, Jan. 1977, pp. 17–24.
12. Review by J.T. Coppock, *The Geographical Journal*, Vol. 148, pt. 2, July, 1982, p. 238, of W.R. Mead, *An historical Geography of Scandinavia*, Academic Press, London, 1981.
13. Review by D.J. Smallbone, *Geography*, Vol. 67, pt. 4, October 1982, p. 364 of D. Horsfall, *Manufacturing Industry*, Oxford, Blackwell, 1982.
14. J.R. Bale, "Industrial Estate Development and Location in Post-war Britain", *Geography*, Vol. 62, pt. 2, April 1977.
15. Review by M. Blakemore, *Geography,* Vol. 66, pt. 4, 1981, p. 325, *People in Britain: A Census Atlas Census Research Unit*, Dept. of Geography, Office of Population Censuses and Surveys and General Register Office (Scotland), London, (HMSO), 1980.
16. Review by Wolfgang Brücher, *Die Erde,* 112. Jahrg. Heft 1–2. pp. 138–139, of W. Mikus, G. Kost, G. Lamche, H. Musall, *Industrielle Verbundsysteme*, Geographisches Institut der Universität Heidelberg, 1979.
17. Review by A.J. Hunt, *Geography*, Vol. 66, Pt. 3, 1981, p. 249, of Willem Molle, B.van Holst and H. Smit,*Regional Disparity and Economic Development in the European Community* Farnborough: Saxon House 1980, Studies in Spatial Analysis.
18. Review by S.J. Baker, *Geography*, Vol. 60, pt. 2, 1975, pp. 162–163, of F.F. Ojany, R.B. Ogendo, Human Geography Longman, Kenya, 1973.
19. Review by A.J. Hunt, *Geography*, Vol. 66, pt. 3 July 1981, p. 249, of M.T. Wild, *West Germany: A Geography of its People*, Folkestone: Wm. Dawson, 1979.
20. Review by R.L. Davies, *Geography*, Vol. 66, pt. 1, Jan 1981, p. 75, Beaujeu-Garnier and Annie Delobez *Geography of Marketing*, London: Longman,1979.
21. Hammond, R., "Linear Regression and Curvilinear Trend lines", *Geography*, Vol. 67, pt. 2, April 1982, pp. 126–133, esp. 132.
22. J.R. Bale, "Geographical Diffusion and the Adoption of Professionalism in Football in England and Wales", *Geography*, Vol. 63, pt. 3, 1978, pp. 188–197, esp. p. 192.
23. Review by M.B. Gleave, *Geography*, Vol. 66, pt. 1, Jan 1981, p. 73, of J.R. Short, *Urban Data Sources* London: Butterworth, 19.

24. Review by R.J. Johnston, *Geography*, Vol. 66, pt. 3, July 1981, pp. 252–253, of R.W. Thomas and R.J. Huggett, Modeling in *Geography: a Mathematical Approach*, London: Harper and Row, 1980.
25. Peter Lewis, *Maps and Statistics* London, Halstead Press 1977, p. 59
26. King, Leslie J. "A Quantitative Expression of the Pattern of Urban Settlements in Selected Areas of the US." in: *Spatial Analysis –A Reader in Statistical Geography* ed. Brian J.L. Berry, Duane F. Marble, University of Chicago Press, 1968, p. 166.
27. See e.g. Roberto Bacchi, "Graphical Methods for Statistical Analysis", *Bulletin of ISI, Proceedings of the 43rd Session,* Buenos Aires, 1981, pp. 1003–1026, and the bibliography of eight of his publications.
28. Review by A. Sayer, *Geography*, Vol. 66, pt. 3 July 1981, of M. Kidron and R. Segal, *The State of the World Atlas: A Pluto Press Project*, London, Pan Books and Heinemann Educational, 1981.
29. Review by W.R. Mead, *Geography*, Vol. 67, pt. 3, July 1982, of: Margarita Bowen *Empiricism and Geographical Thought from Francis Bacon to Alexander von Humboldt* Cambridge: Cambridge University Press, 1981.
30. Prof. Dr. Lutz Holzner (Milwaukee). This English language summary is entitled "Cultural-genetic responses to processes of urbanization – an anti-positivist view" (Die Kultur-genetische Forschungsrichtung in der Stadtgeographie (eine nicht-positivistische Auffassung). Dept. of Geography, U of Wisconsin, Milwaukee, Wis., in: *Die Erde*, Jahrgang 112, 1981 pp. 173–184, esp. p. 174.
31. see: R.W. Thomas et al. op. cit.
32. Review by R.L. Davies, *Geography of Marketing*, J.Beaujeu-Garnier, Annie Delobez, London: Longman, 1979, the English translation of *Geographie du Commerce* Paris, 1977, Geography, Vol. 66, pt. 1, Jan 1981.
33. Typical is the following example, although not from human geography: "In all applications on e.g. the probability of 20 of the next 30 years experiencing above-average rainfall, assuming p=.5 and independence of events over time, is given by P(x=20) = (30 over 20)(.5) 2(.5)0=.0280 a rather tedious calculation. However. . .a normal distribution with the same mean and standard deviation provides a close approximation to the binomial distribution." John Silk *Statistical Concepts in Geography*, London, Allen & Unwin, 1979, p. 90.

Afterthoughts

*Yesterday's heresy has become today's Orthodoxy**

The ideas discussed in this book evolved during six decades of teaching and doing applied statistical research. These thoughts, some as sudden, unexpected insights that defied statistical orthodoxy, others with long gestation periods, are reflected in content and sequence of the chapters. All this turned into a statistical autobiography.

It all began at the University of Vienna, home of the Austrian (marginalist) school of economic thought, as a student of a well-known statistician, my father, Professor Wilhelm Winkler. Studying under him laid a solid foundation of social, economic and demographic statistics. It was, however, also the beginning of my uneasy feelings about the role and nature of probability in socio-economic statistics.

Later, in the library of the Banco Central de Venezuela in 1948, it struck me that behind all social and economic statistical work is some phenomenon of society that must be defined, and the 'objects' in which it is materialized – people, things, or a variety of events – must be located and recorded.

Then, on my way to teaching economic statistics at the Universidad Central de Venezuela, in Caracas, the oversized mosaic on the wall of a big building struck me as the paradigm of statistics. The role that the small pieces of colored mosaic stones played in portraying the picture of a patriotic theme reminded me of the role the individual 'real-life-objects' play to reveal some phenomenon of society, but that this fact translates to actual data only to the extent that these objects de facto are also recorded as 'statistical counting-units'.

Later, attending with my family the performance of Charles Dickens' 'A Christmas Carol,' in Washington, DC, an actor on stilts appeared on stage as the ghost of Christmases to come. Covered with a huge cape, his oversized presence seemed to fill that entire stage. It was so big that it took some time before the attending audience noticed and recognized it (more in Endnote 29, Chap. 1). That episode at that moment struck me as the paradigm of social phenomena: they are too large and too widely dispersed to be readily perceived. It is precisely the task of descriptive statistics to scan these elusive, widely spread-out social and economic

*The origin of this provocative statement is hard to trace. The prime 'suspects' are Karl Rahner, S.J., but also Hans Küng, Bernard Häring and Edward Schillebeeckx.

phenomena, and make them 'visible.' Statistics, essentially functions as a macroscope, the appropriate instrument to accomplish this task (Endnote 31, Chap. 1).

While working with a medical researcher on his study of leucocytes, and in my own research of the daily changes in the measurements of Venezuelan schoolchildren's height and weight,[1] I became aware of the fundamental difference between 'measurements' in the natural sciences, and 'measurements' in the social sciences, discussed in Sect. 1.4.

I remember the moment in 1955, when driving on the Avenida Bernardo O'Higgins in Santiago, Chile, returning from teaching a class on statistics of agricultural and industrial production at the CIEF,[†] it suddenly struck me that every statistical "measurement" has three separate determinants: its subject-matter definition, its geographic location, and the moment in time when it is recorded. The totals of such individual statistical facts, their aggregates, consequently also have these three 'dimensions,' as discussed in Chaps. 2 and 3.

While giving a seminar on price-index-numbers to a group of labor union representatives from the Caribbean, in Front Royal, Virginia, it dawned on me that price-index-numbers were essentially ratios of price aggregates as were many other economic data: de-seasonalized time series, the GDP and all monetary data 'at constant prices,' even arithmetic means, other averages and measures of dispersion. I was surprised about how ubiquitous and important ratios really are in the social sciences, which prompted me to dedicate an entire chapter to ratios. That seminar was also the beginning of my slowly evolving different understanding of 'price', 'price level' and its changes, and how it ought to be measured.

I also became aware then that in the social sciences time-series, the longitudinal studies of a situation, were more frequent and more important than cross-sectional studies, like the frequency distributions. This is reflected in the sequencing and number of chapters: three chapters on time series precede the single chapter on frequency distributions, and one on regression. This is contrary to current thinking that is dominated by statistics in the sciences, in which the importance and sequence of these topics is reversed.

When viewing the film "Trilogy" based on a short-story by Somerset Maugham, the scene hit me head-on where the surprised banker arrives at a logical but patently wrong conclusion. It was the same kind of logic implicit in many econometric assumptions. The film's story (see a more detailed account, Endnote 17, Sect. 7.2) dealt with an illiterate church sexton who was fired by the new pastor because he could not take care of the birth, marriage and death registers. Relieved of his lifelong job, he reluctantly opened a tobacconist shop and in time became a successful business man. His growing savings-account prompted the director of that bank to prepare for him a portfolio of high-earning stock investments. When that client was asked to sign that proposal, he refused. A probing question revealed the reason for his refusal, namely that this client could neither read nor write, the banker amazed

[†]Centro Interamericano de Enseñanza Estadística Económica y Financiera, Project #10 of the Organiation of American States.

exclaimed in a dramatic scene: "you are a genius! where would you be if you could read and write" to which the client responded "Ah, I would still be the sexton at St. Andrew's church and would not have a penny to save."

At this startling finale with that banker's sensible but wrong conclusion, suddenly Laspeyre's price-index-number came to my mind. Its implicit assumption: that by holding constant the quantities at the level of the base period, the variable q, like in a science experiment, one seemed to be able to isolate the variable p. But if one actually could have enforced that the quantities sold during the base period on all markets remained unchanged, the ensuing prices would have developed differently from those actually used in the price-index, without such a control of the quantities. This story also reminded me of the kind of logic that appears to underlie the assumptions of econometric models.[2] (Endnote 22 in Sect. 7.1 and Endnote 17 in Sect. 7.2.)

The understanding of aggregates as vague structures, comparable to the vaporous substance of clouds, occurred to me on a flight to an ASA[‡] meeting in Los Angeles, when the airplane flew into a dense cloud. This lead to the realization that trends, in time-series and regression, based on aggregated data, are equally vague. In a graph these trends ought to be represented as hazy, barely visible trend-zones or trend-canals whose upper and lower borders gradually fade out, instead of the customary presentation of trend as a single, clearly drawn line.

Another ευρηκα moment happened when it became clear to me, during a routine shopping experience, that 'price' belongs to the **transaction** of a merchandise, not the merchandise itself, discussed in Sect. 7.1.

During a data-mining experience (Chapter 9.1) as the statistical expert-witness in a law-suite, I discovered that really all frequency distributions in one or more dimensions, Chaps. 8 and 9, are cross-sections and must be interpreted as static pictures of the situation, not as is customary, as dynamic ones, e.g. when interpreting the slope of regression lines (Sect. 9.1). These are a few personal highlights that were involved in the evolution of my thoughts, laid-out in this book.

Then one day a large publisher of statistical material approached me to write a book about the ideas that I had presented at annual meetings of ASA. After initial encouragement by the editor, the verdict, about the completed manuscript "A Foundation of Statistics" was as follows "This is not even statistics, forget about publishing it to avoid the embarrassment for the publisher and the author." For these reviewers the foundation of statistics was the mathematics of probability and inference, not the data and their interpretation.

Eventually this book evolved from that original manuscript as "Interpreting Social and Economic Data," a title suggested by a colleague who had reviewed that manuscript. A version of the original title was retained as the subtitle "A Foundation of Descriptive Statistics." Implicit in "Foundation," of course, is the notion that the book is about basic principles, but not the usual kind of textbook.

Times are changing. Doubts are beginning to arise about the mathematically oriented statistics courses as an appropriate academic discipline for the study of the

[‡]American Statistical Association, Alexandria, VA.

social sciences. Of all students of business, economics or sociology who had to endure statistics courses, few are interested in a career in statistics. Yet most textbooks are written as if to prepare professional statisticians, and are taught by professors with a pedigree in mathematical statistics with scant interest in the descriptive part of social and economic statistics. There is growing pressure in business schools, departments of economics and sociology to reduce the hours of teaching this kind of statistics to free up time for other subjects that are believed to be more relevant.

With this book I hope to counter that tendency, by proposing a different direction for these courses of statistics. It is written for everybody who has to make sense of statistical data and use them. I also hope to rehabilitate descriptive statistics, the neglected Cinderella of the profession, nudging statistical theory in the indicated, different direction. Description, after all, is the reason why statistical efforts are undertaken in the first place.

Those who feared or expected mathematical formulas may be relieved, or disappointed, to discover that social and economic statistics, though numeric, is essentially quantified history of society, not a branch of mathematics. I certainly hope that it will change the kind of reaction to statistics as the incident that I witnessed, reported in the Preface. The redundancy of basic statements, repeated in many chapters, is intended for those readers who are selectively interested only in some of the chapters.

Let me conclude with a call for caution. Social and economic statisticians have to tread a delicate line between recognizing the limitations of statistical data on one hand, while acknowledging their practical value and usefulness on the other hand, to ward off attempts to cut the budget for statistics and to reduce academic credit-hours of courses in business, economic or social statistics. The threat comes from the bureaucracies of the departments in universities and government to reduce and shift budget allocations for statistics to other projects.

Finally I would like to acknowledge once more that this book owes its existence to Springer Verlag's Dr. Niels P. Thomas. It would not have been written without his initiative.

Notes

1. Winkler, Othmar (1997) "Medición de Talla y Peso en Niños: Cuanta Confianza Merecen?" *Anales Venezolanos de Nutrición* Fundación Cavendes, (2), pp. 127–138, Caracas, Venezuela. ("Measurements of Height and weight of Children: how much Confidence do they deserve?").
2. Winkler, Othmar (1985) "Statistical Flaws in Econometricians' Perception of Economic Reality" *Quantity and Quality in Economic Research, Vol. I*, papers of the first International Conference on Economic Research, Roy C. Brown ed. University Press of America, Lanham, MD, 1985, pp. 295–354. Winkler, Othmar (1995) "Rethinking Statistical Theory for Economics and the Social Sciences" paper presented at the 50th Congress of ISI, Beijing, China, August 1995. *Bulletin de l'Institute International de Statistique, Livraison 2,* pp. 1286–1287.

Appendix A
Appendix to Chapter 3

Data users ought to be warned about that vagueness by correspondingly pale printing of totals. Yearly totals e.g. ought to be printed in barely visible tones of gray.

To quantify that vagueness of a very large aggregate I experimented with different internal structures of the CPI for the US economy, which is such a very large aggregate. I explored systematically how such an internal change might affect the numeric value of such an aggregate, even though its definition remains exactly the same. In the following you see what might happen if the relative importance of these seven price sub-aggregates (the usual five product and two service groups) and their Index Numbers were changed by 10%. The CPI-U – Consumer Price Index for Urban Consumers – in 1985 was 323.9 (1967 = 100). Under that rather extreme assumption the value of that CPI could be as low as 314.1 or as high as 333.9, with any outcome in-between possible. (See Table A.1 and Fig. A.1).

These possible outcomes are the quantitative assessment of the vagueness of this large aggregate, not probabilities. This aggregate is the CPI value for the entire year 1986, for all prices of all goods and services, for the entire economy of that vast country, the USA. At that level of aggregation nothing is known about its inner structure and when using this figure, the internal structure of this aggregate generally is of no interest.

The relative importance of the sub-aggregates within an aggregate could be different from that actually observed. One would not know this until de-aggregating the total CPI into its major and minor component groups.

The chosen re-shuffling of the relative importance of the weights of the seven sub-aggregates by 10% is based on a rather extreme assumption, unlikely to happen in the short run. Under this assumption of the re-weighting these seven sub-aggregates, about 80% of all possible values for that year's CPI would be located between 318.3 and 329.3.

This quantitative exploration of the possible range of uncertainty of this large aggregate is an attempt to give a quantitative expression to the uncertainty that is implicit in the definition of that aggregate. This CPI figure is a highly abstract entity that cannot really be visualized or imagined.

Table A.1 Systematic exploration of the effect of a 10% re-allocation of weights (Consumer spending Survey Data, 1982–1984, in % of family budget)*

F&B	H	A&U	T	MC	E	OG&S	CPI	
+5%	−5%	0	0	0	0	0	321.6	
+5%	0	−5%	0	0	0	0	328.8	
+5%	0	0	−5%	0	0	0	323.1	
+5%	0	0	0	−5%	0	0	318.9	
+5%	0	0	0	0	−5%	0	325.8	
+5%	0	0	0	0	0	−5%	322.6	
0	+5%	−5%	0	0	0	0	331.2	
0	+5%	0	−5%	0	0	0	321.3	
0	+5%	0	0	−5%	0	0	328.2	
0	+5%	0	0	0	−5%	0	325.2	
0	+5%	0	0	0	0	−5%	325.5	
0	0	+5+	−5%	0	0	0	318.3	
0	0	+5%	0	−5%	0	0	**314.1**	Lowest value
0	0	+5%	0	0	−5%	0	321.0	
0	0	+5%	0	0	0	−5%	318.0	
0	0	0	+5%	−5%	0	0	319.8	
0	0	0	+5%	0	−5%	0	326.7	
0	0	0	+5%	0	0	−5%	323.7	
0	0	0	0	+5%	−5%	0	330.9	
0	0	0	0	+5%	0	−5%	327.8	
0	0	0	0	0	+5%	−5%	320.9	
17.840	42.637	6.524	18.696	4.796	4.380	5.128		
−5%	+5%	0	0	0	0	0	326.2	
−5%	0	+5%	0	0	0	0	319.2	
−5%	0	0	+5%	0	0	0	324.9	
−5%	0	0	0	+5%	0	0	328.9	
−5%	0	0	0	0	+5%	0	322.1	
−5%	0	0	0	0	0	+5%	325.2	
0	−5%	+5%	0	0	0	0	316.8	
0	−5%	0	+5%	0	0	0	322.5	
0	−5%	0	0	+5%	0	0	326.7	
0	−5%	0	0	0	+5%	0	319.8	
0	−5%	0	0	0	0	+5%	322.8	
0	0	−5%	+5%	0	0	0	329.7	
0	0	−5%	0	+5%	0	0	**333.9**	Highest value
0	0	−5%	0	0	+5%	0	326.9	
0	0	−5%	0	0	0	+5%	330.0	
0	0	0	−5%	+5%	0	0	328.2	
0	0	0	−5%	0	+5%	0	321.3	
0	0	0	−5%	0	0	+5%	324.3	
0	0	0	0	−5%	+5%	0	317.1	
0	0	0	0	−5%	0	+5%	320.2	
0	0	0	0	0	−5%	+5%	327.1	

* F&B = Food and Beverage; H = Housing; T = Transportation; MC = Medical Care; E = Education and communication; OG&S = Other Goods and Services

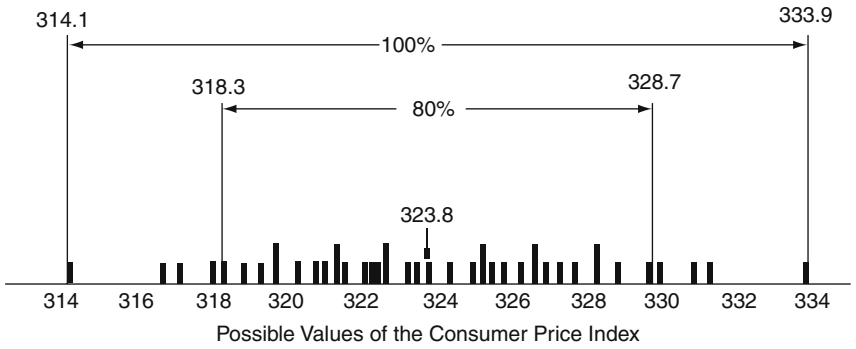

Fig. A.1 Exploring the effects on the large price aggregate "Consumer Price Index USA 1985" of a 10% (−5% and +5%) change in the relative importance of its seven sub-aggregates

Appendix B
Appendix to Chapter 5

Table B.1 New Construction put in place, US Quarterly (not actual) Fiugures, in million $

	2004				2005			
	1	2	3	4	1	2	3	4
Private Residential Nonfarm, USA	5,347	7,106	7,469	6,589	5,195	7,135	7,568	6,781
North East Region only	1,260	1,710	1,820	1,650	1,297	1,895	1,971	1,636
Private total const. USA	9,604	11,753	12,664	11,870	9,751	12,914	13,817	13,517
Total New Constr. USA	13,399	16,766	18,559	17,174	13,692	18,622	20,381	19,044
North East Region only	3,070	3,860	4,080	3,730	2,940	3,769	3,998	3,818

Table B.2 New construction put in place, 2005 totals in million $

Private Residential Nonfarm Construction (26,679)	Total Private Construction (53,817)	New Constr. put in Place Total (71,739)
Private Nonresidential Construction Buildings, except farm and public Utilities (20,679)		
Farm Construction (1,195)		
Public Utilities (5,178)		
Buildings, excl. Military (7,443)	Public Total Constr. (17,922)	
Military Facilities (883)		
Highways and Streets (9,596)		

Table B.3 Schema of the ratios and their uncertainty

			Private Residental Constr. N. E.						Private Residental Constr. N. E.					
	Year	Quarter	Total (3)	$\frac{y_t}{5{,}440}$ (4)	$\frac{y_t}{6{,}477}$ (5)	$\frac{y_t}{6{,}662}$ (6)	$\frac{y_t}{6{,}813}$ (7)	x_t (8)	$x_t = 1{,}650$ (9)	9/4 (10)	9/5 (11)	9/6 (12)	9/7 (13)	Low-High (14)
2004	1.	1+2+3+4	6,440	100.00	–	–	–	1,260						
	2.	2+3+4+1	6,477	100.57	100.00	–	–	1,710						
	3.	3+4+1+2	6,662	100.45	102.86	100.00	–	1,820						
	4.	4+1+2+3	6,813	105.80	105.19	102.27	100.00	1,650						
2005	1.	1+2+3+4	6,799	105.57	104.97	102.05	99.80	1,297	78.61	78.2	76.4	76.5	78.8	76.4–78.8
	2.	2+3+4+1	6,822	–	105.33	102.40	100.13	1,895	114.85	111.0	109.2	112.5	114.7	119.2–114.7
	3.	3+4+1+2	7,027	–	–	105.48	103.14	1,971	119.45	112.9	113.8	116.7	115.8	112.9–116.7
	4.	4+1+2+3	7,170	–	–	–	105.24	1,636	99.15	93.9	94.1	94.0	94.2	93.9–94.2
	1.	1+2+3+4	7,144	–	–	–	–	1,320 2,100 2,114 1,610						

Table B.4 Summary of results

	$k_{(t,0)}$			$k_{t,t-1}$		
	1–2	1–3	1–4	1–2	2–3	3–4
Private Res. Constr. N.E. to Total new Constr. N.E.	114.5% +14.5%	111.8 +11.8%	97.1 –2.9%	114.5 +14.5%	97.6 –2.4%	86.9 –13.1%
Priv. Res. Constr. N.E. to Priv. Res. Constr. USA	94.0 –6.0%	95.9 –4.1%	103.5 +3.5%	94.0 –6.0%	102.0 +2.0%	107.9 +7.9%
Priv. Res. Constr. N.E. to Total New Constr. USA	107.4 7.4%	102.1 +2.1%	90.7 –9.3%	107.4 +7.4%	95.1 –4.9%	88.8 –11.2%
Priv. Res. Constr. N.E. to 4-Quarter moving Totals	109.2% to 114.7%	112.9 to 116.7%	93.9 to 94.2%	142.9% to 145.6%	101.7% to 103.4%	80.7% to 83.2%

Table B.5 Results of calculations demonstrated in Tables B3 and B4

Year	Month	Total Private Residential Construction	12-month		Trend		Ratio-to-mov. average	
			Moving Total (placed at end of period)	Moving Aver.	High	Low	High	Low
2002	Jan	1,543						
	F	1,368						
	M	1,605						
	A	1,906						
	M	2,141						
	June	2,373						
	J	2,357						
	A	2,377						
	S	2,330						
	O	2,210						
	N	2,113						
	D	1,965	24,292	2,024	2,138	2,024	97.0	91.9
	Jan	1,669	24,418	2,035	2,154	2,035	82.0	77.2
2003	F	1,456	24,506	2,042	2,166	2,042	71.2	67.2
	M	1,698	24,599	2,050	2,180	2,050	82.8	77.9
	A	2,018	24,711	2,059	2,197	2,059	98.1	91.9
	M	2,254	24,824	2,069	2,210	2,069	109.1	102.1
	June	2,495	24,942	2,078	2,219	2,078	120.1	114.0
	J	2,470	25,055	2,088	2,225	2,088	118.4	111.0
	A	2,446	25,124	2,094	2,231	2,094	116.6	109.5
	S	2,419	25,213	2,101	2,235	2,101	115.0	108.2
	O	2,408	25,411	2,118	2,235	2,118	113.7	107.7
	N	2,357	25,655	2,138	2,235	2,138	110.3	105.4
	D	2,153	25,843	2,154	2,235	2,154	99.9	96.4
	Jan	1,813	25,987	2,166	2,235	2,166	83.7	81.1
2004	F	1,626	26,157	2,180	2,235	2,180	74.6	72.8
	M	1,906	26,365	2,197	2,235	2,197	86.9	85.3
	A	2,188	26,535	2,210	2,235	2,196	99.5	97.8

Table B.5 (continued)

Year	Month	Total Private Residential Construction	12-month		Trend		Ratio-to-mov.	
			Moving Total (placed at end of period)	Moving Aver. (placed at end of period)	High	Low	average High	average Low
	M	2,345	26,626	2,219	2,235	2,192	107.0	105.0
	June	2,570	26,701	2,225	2,235	2,192	117.2	115.0
	J	2,546	26,777	2,231	2,235	2,192	116.1	113.9
	A	2,492	26,823	2,235	2,235	2,192	113.7	111.5
	S	2,405	26,809	2,234	2,234	2,192	109.7	107.6
	O	2,311	26,712	2,226	2,226	2,192	105.5	104.0
	N	2,229	26,584	2,215	2,215	2,192	101.6	100.5
	D	2,076	26,507	2,209	2,219	2,192	94.6	93.6
	Jan	1,788	26,482	2,207	2,224	2,192	81.6	79.8
2005	F	1,580	26,436	2,203	2,229	2,192	72.0	69.0
	M	1,827	26,357	2,196	2,233	2,192	83.4	81.5
	A	2,134	26,303	2,192	2,236	2,192	97.4	95.5
	M	2,371	26,329	2,194	2,241	2,194	108.2	105.8
	June	2,630	26,389	2,199	2,241	2,199	119.6	117.2
	J	2,591	26,434	2,203	2,233	2,203	117.6	115.6
	A	2,527	26,469	2,206				
	S	2,450	26,514	2,209				
	O	2,370	26,573	2,214				
	N	2,283	26,627	2,219				
	D	2,138	26,689	2,224				
	Jan	1,843	26,744	2,229				
2006	F	1,627	26,791	2,233				
	M	1,873	26,837	2,236				
	A	2,191	26,894	2,241				
	M	2,367	26,890	2,241				
	June	2,534	26,794	2,233				

Appendix C
Appendix to Chapter 8

Asymmetry and its measurement is not considered to be of great importance. In deference to this wrongly held attitude, continuing the discussion of the US Population age-distribution has been relegated to this appendix.

$$A_3 = (D_8 - D_7)/(D_3 - D_2) = 1.76.$$

The right side of the age distribution in the ages between D_7 and D_8 is 76% more extended than the left side inter-decile distance between the ages that mark the 20 and 30% points D_2 and D_3.

$$A_4 = (D_9 - D_8)/(D_2 - D_1) = 2.12$$

The 10% of the population between the ages D_8 and D_9 is more than twice as widely dispersed as the 10% of the population in the younger ages between D_1 and D_2.

$$A_5 = (D_{10} - D_9)/(D_1 - D_0) = 6.93.$$

One would not expect such a high asymmetry in that outer part of the distribution, the right tail-end spread out seven times as much as the left tail end, a surprising fact that neither a look at the tabulated data nor the histogram reveals. Overall, except for the bi-modality which creates a moderate U-shape in the center, this appears to be a reasonably well behaved frequency distribution with only moderate asymmetry.

$\overline{A} = (A_1^* A_2^* A_3^* A_4^* A_5)^{1/5}$ = fifth root of $\{(1.27)(1.58)(1.76)(2.12)(6.93)\} = 2.2$ for the entire age distribution. It can be interpreted as follows:

The inter-decile distance between any two deciles above the median, on average, covers more than twice the number of years (of age) than the corresponding inter-decile distance below the median. The geometric mean $\overline{A}$ becomes more meaningful after having taken cognizance of the detailed A_i ratios, as given in Table C.1. The row vector of these six ratios

$$(A_1, A_2, A_3, A_4, A_5; \overline{A}) = (1.3, 1.6, 1.8, 2.1, 6.9; 2.2)$$

Table C.1 Comparative measures of asymmetry

Part A

Name of distribution	Name of measure						
	$\alpha 3$	S	Sk(P)	φ (p=.25)	sk(Q)	sk(M)	$\overline{A}$
Any symmetric distr.	0	0	0	0	0	1	1.0
Population by age	.5	.2	.6	.3	.2	1.5	2.2
Individual income tax returns	1.3	.3	.7	.3	.2	1.5	2.5
Establishments, by No. of employees	3.6	.3	.8	0*	−.1	.9	7.6
Establishments, by size of sales	4.2	.2	.7	0	0	1.0	15.4
School systems, by No. of pupils	7.9	.2	.7	.1	.5	2.6	18.5
Construction firms, by gross income	23.5	.2	.6	.1	.6	3.6	54.6

*The exact value is −.0009.

Part B

	A_1	A_2	A_3	A_4	A_5	$\overline{A}$
Any symmetric distribution	1	1	1	1	1	1.0
Population by age	1.3	1.6	1.8	2.1	6.9	2.2
Individual income tax returns	1.3	1.6	2.0	3.2	6.7	2.5
Establishments, by No. of employees	1.0	.8	.7	18.6	2,361.1	7.6
Establishments, by size of sales	1.0	1.0	1.0	54.5	15,860.0	15.4
School Systems, by No. of pupils	1.3	3.2	7.1	24.3	3,069.1	18.5
Construction firms, by gross income	2.3	3.2	10.0	36.4	178,212.0	54.6

completely describes the nature of the asymmetry of this distribution, is easy to use, easy to understand, and performs reliably in all situations.

In general terms,

$$A_i = \{D_{(5+i)} - D_{(5+i-1)}\}/\{D_{(5-i+1)} - D_{(5-i)}\} \text{ for } 1 \leq i \leq 5$$

in which $\overline{A}$ = nth root of the products of the n A_i values. A ratio of 1 means that the inter-decile distance on the left side of a distribution is equally as dispersed as the corresponding concentric inter-decile distances on the right side, regardless of the shape of the distribution, e.g. irregular U shaped. In left-skewed distributions the left side of the distribution is more spread out than the right side.[1] For the routine determination of the A_i-values existing computer programs can be adapted.

Instead of deciles frequently other fractiles are used, especially quartiles. That measure compares interquartile distances that refer to each 25% of the distribution, measuring the asymmetry in less detail.[2] Ungrouped data, of course, are ideal for computing these same measures of asymmetry.

Although skewness and kurtosis are customarily treated together, most socio-economic distributions are so strongly asymmetrically shaped that there is no point in considering kurtosis, that is, the degree of "flatness of their hump."

C.1 Comparing A_i with Other Measures of Asymmetry

The best known measure of asymmetry is based on the third moment around the mean. It is given either as

$$\alpha 3 = [\Sigma\, f_j\{X_j - \mu\}^3/N]/\sigma^3, \text{ or } \Sigma f_j[(X_j - \mu)/\sigma]^3/N.$$

Occasionally the variant $(\alpha 3)^2$ can be found.

Pearson's well known measure sk (P) $= 3(\overline{x} - M)/\sigma$ although it uses the median M, it really intends to use the Mode in conjunction with the arithmetic mean.[3] This measure is not sensitive to more pronounced forms of asymmetry.

Hotelling discussed another measure of asymmetry[4] that he called 's' but will be used here as 'sk(H)', skewness according to Hotelling.

$$\text{sk(H)} = (\text{mean} - \text{median})/\sigma.$$

This measure is essentially like Pearson's formula except that it does not exceed the values of -1 and $+1$, assuming the value 0 in the case of symmetry.

Another general version of a wide variety of measures of skewness, φ is[5]

$$\varphi = [X(p) + X(1 - p) - 2^*X(p = .5)]/\sigma$$

depending on which quantile is used for p.

A similar, better known measure for skewness, based on the quartile-points Q_i is

$$\text{sk(Q)} = (Q_1 + Q_3 - 2\text{Med})/(Q3 - Q1)$$

which can also be written as

$$[(Q_3 - Q_2) - (Q_2 - Q_1)]/(Q_3 - Q_1).$$

It is like phi in the numerator if $p = .25$. In the denominator the interquartile range $Q_3 - Q_1$ replaces sigma as a measure of dispersion. The only measure that resembles the proposed A_i is

Miller's formula[6]

$$\text{sk(M)} = (Q_3 - Q_2)/(Q_2 - Q_1).$$

These measures of asymmetry are ratios between some expression of the lack of symmetry in the numerator, and some measure of dispersion in the denominator. The numerator is expressed as a multiple of that dispersion, hopefully indicating direction and intensity of asymmetry. The comprehension of that denominator is important to properly interpret each measure. (For the source of the information of

Table C.1 see the Endnote 7). A rapid scanning, by columns, of Part A of Table C.1 reveals that only $\alpha 3$ is sensitive to strong asymmetry. All other measures fail to reflect the increase of asymmetry from the first to the last distribution. Pearson's measure sk(P) is no exception. sk(M) does better but cannot fully do justice to the real situation because it is confined to the central 50% of the distribution. It indicates, e.g. that for the last distribution in Table C.1, Part A, 25% of the units to the immediate right of the median are 3.6 times – or 260% – more widely spread out than the 25% to the immediate left of the median. This indicates a substantial departure from $\alpha 3$ symmetry in the central portion of that distribution but ignores the more extreme asymmetrical situation in the outlying portions of this distribution. $\alpha 3$ is more sensitive. Its high value for the last distribution states that the weighted, averaged and cubed deviations of the class centers from the mean of this distribution are 23.45 times as large as the cubed standard deviation, whatever that may mean, and asymmetric to the right. On the other hand $\overline{A} = 54.59$ states that, on average, a 10% segment of the population, between two decile values to the right of the middle (the median), occupies 54 1/2 times as much distance on the horizontal scale – business receipts during 1973 in $100,000's – as the corresponding concentric 10% segment of those construction firms, to the left of the median. The five individual A_i values give a step-by-step account of the peculiar asymmetric structure of this distribution, especially the enormous asymmetry in its outlying parts. Thus A_5 indicates that the last 10% of that distribution occupies a scale length 178,212 times that of the inter-decile distance of the very first 10% of these construction firms' receipts. The gross receipts of the top 10% of construction firms are 178,212 times more spread out form the smallest in that top class – who is a large firm nonetheless – to the biggest in the top class, than the gross receipts of the 10% between the smallest and the biggest among the very small construction firms, that is, among the smallest 10% of all the entrepreneurs in the construction field. It is easy to understand what is going on in this distribution. Indeed, the A_i vectors for the other distributions reveal great differences in regard to asymmetry which even a careful visual inspection of the histograms is unlikely to reveal. $\alpha 3$, in contrast, defies a meaningful interpretation, even though it is sensitive. All other measures of asymmetry are surprisingly ineffective, obviously were not meant for our kind of frequency distributions and are not suited for socio-economic data.

The finer breakdown given in the A_i vector allows some interesting insight, e.g., into the peculiar distribution "Retail Trade Establishments by Number of Employees" (Table C.1, Part B, row 4). There is an increasing asymmetry toward the left in its center area, but strong asymmetry in the opposite direction, to the right, in the outer 40% of the data, a fact that remains unnoticed by other measures. Φ and sk(Q) indicate "perfect symmetry" because of the simultaneous presence of a left and a right skewness within the limited range covered by both measures, with asymmetries that completely cancel each other. In all instances φ has been computed for a $\pi = .25$. Even sk(M) is low as compared to its values for the other distributions listed in the table. In contrast, A_1, A_2, and A_3 indicate growing if moderate left asymmetry within the central 60% portion of this distribution. The overwhelming asymmetry in the opposite direction in the outer 40% dominates and completely reverses this picture.

Similar comparisons, row wise, of the different measures in Table C.1 for the other frequency distributions show the greater sensitivity of the A_i measure.

Hotelling's measure computed for the age distribution of the U.S. population in 1970 gives sk(H) = .24 stating that the difference between mean and median is 1/4 of the standard deviation. It would not only be important to know exactly what sigma means in this distribution, but also what a discrepancy in the numerator signifies that is 1/4 of sigma. Neither figure can be clearly understood. $\alpha 3 = 1.26$ for the age distribution, means that the cubed, weighted and averaged deviations of the class midpoints from the arithmetic mean of the distribution are 1.26 times as great as (or 26% greater than) the cube of the standard deviation. The standard deviation has too long been taken for granted, with no questions asked about its meaning in other than the Gaussian distribution. A statement like "if the value of $\alpha 3 > .5$ there is considerable skewness present"[8] gives no insight into the nature of the asymmetry of such a distribution nor is intuitively understandable.

All measures of asymmetry that use σ cannot be interpreted in practical, meaningful terms. sk(p) = .58 simply says that the distance of the arithmetic mean from the median, multiplied[9] by 3, is a little more than half the standard deviation, another piece of statistical information that defies interpretation. .My proposed $\overline{A} = 2.20$ on the other hand does not pose similar difficulties of interpretation.

Notes

1. For more detail see: Winkler, Othmar W. "A new Measure of Asymmetry for Data in Business and Economics" *Proceedings of the Business and Economic Statistics Section, ASA* Washington, D.C., 1977, pp. 723–728.
2. To appreciate this compare the A_1, A_2 and A_3 values with the values of sk(M) in Table C.1. sk(M) is quartile-based, covering essentially the same parts of a distribution.
3. K. Pearson believed that for biological data Med - Mode $= 2(\hat{x}$ - Med), hence $(\overline{x}$ - Mode) $= 3(\overline{x}$ - Med), preferring the latter expression because of the uncertainty in determining the Mode. This 2:1 relationship evidently does not hold for most social and economic frequency distributions.
4. Harold Hotelling and Leonard M. Salomons, "The Limits of a Measure of Skewness," *The Annals of Mathematical Statistics* Vol. III, pp. 141–142, 1932, and Raymond Garver, "Concerning the Limits of a Measure of Skewness," ibid., pp. 358–360.
5. Robert W. Resek, "Alternative Tests of Skewness: Efficiency Comparison Under Realistic Alternative Hypotheses," *Proceedings of the Business and Economic Statistics Section, ASA,* 1974, St. Louis Meeting, Washington, D.C., pp. 546–551. Resek evidently was aware of occasional complexities in the structure of asymmetry. He states: "F(Xsub-pi) = pi defines the Xsub-pi, the (100 pi) percentile of distribution. It appears that phi may indicate zero skewness for one value of pi and non zero skewness for another, or even positive for one, and negative for another." (p. 546).
6. Herman P. Miller, "Elements of Symmetry in the Skewed Income Curve," *JASA*, Vol. 50–269, March 1955, pp. 55–71. I am ascribing this formula to Miller even though he quoted it as two separate measures, viz., (1 - Q1/Q2) for the left side and (Q3/Q2 - 1) for the right side. He then compared these two separate expressions informally, assessing the degree of asymmetry of various income distributions in this manner. If this ratio is formalized sk(M) appears much like the A(i) measures.

7. "Taxable Returns, Form 1040A Returns: Sources of Income and Tax Items by size of Adjusted Gross Income," US Internal Revenue Service, *Statistics of Income 1973,* Table 1.9, p. 34, Publication 79 (11–76), Washington, D.C., 1976.

   "Establishments by Number of Paid Employees." Due to a mistake, the data for "Single and Multi Units - United States, by Kind of Business: 1972, Retail Trade, Total, Establishments by Number of Units: (Table 2a, p. 1–66) was used for the actual computations. In addition, the entries in rows 1 to 4, 7 and 8 of the column "Establishments" were omitted for specific reasons from the calculations, making it a somewhat unrealistic distribution. Apart from this, the comparisons between the different measures of asymmetry remain valid. *US Census of Retail Trade 1972,* Bureau of the Census, Department of Commerce, Washington, D.C., 1976.

   "Sales Size of Establishments, United States, by Kind of Business: 1972 Retail Trade, Total," Table 2b, p. 1–95, *US Census of Retail Trade 1972,* ibid. "Public School Systems, Schools, and Pupils Enrolled, by States: 1972," Table 13, p. 40, "All Systems, by Enrollment Size (pupils enrolled, Oct. 1971), Total," *Census of Governments 1972,* Bureau of the Census, Department of Commerce, Washington, D.C., 1973.

   "Receipts, Selected Deductions and Net Profit, by Selected Industries and Size of Business Receipts - Total Revenue Service," *Business Income Tax Returns Publication 438* (3–76), Washington, D.C., 1976.
8. Robert Parsons, *Statistical Analysis: A Decision Making Approach* Harper & Row, New York, 1974, p. 93.
9. It is seldom made clear that this is based on Pearson's early "discovery," that arithmetic mean to median, and median to mode relate in the proportion of 1:2, and that consequently the distance between arithmetic mean and mode is three times the distance between arithmetic mean and median. He gave no reasons why the mode of a distribution is to be avoided even though it seems easy to understand and interpret. The supposed relation of 1:2, by the way, could never be confirmed for any distribution!, let alone, for any socio-economic frequency distribution.

Appendix D
Appendix to Chapter 9

To demonstrate how to calculate the proposed decentralized model of data association the data of Table D.1 are used. These 90 cases were randomly selected from the data of salaries $Y_{i,x}$ and years of employment X of the sex-discrimination study of ERDA The $Y'_{i,x}$ are the decentralized equivalent of the (centralized) values of the $\hat{Y}_{i,x}$-values of the customary unique regression line. Instead of only one computed regression-line value for a given X in the case of the data in Table D.1, there are two delimiting lines, a ceiling-line and a floor-line and more than one $Y'_{i,x}$ computed value at equal distances. In the case of a distribution of the data in the shape of a triangle, these $Y'_{i,x}$ -values will be between one ceiling line and the horizontal axis as the floor line.

The locations of these hypothesized points $Y'_{i,x}$ between floor and ceiling are determined as equally distant points by simple interpolation between floor and ceiling. Then, for a given X, the distance of every observed $Y_{i,x}$ from their hypothesized, nearest $Y'_{i,x}$ is determined and squared. Like in regular regression, the double summation $\Sigma\Sigma[Y_{i,x} - Y'_{i,x}]^2$ is an indication of the irregular distribution inside the 'triangle' or within the heteroscedastic swarm of data, in the shape of a fan. The first summation is for all deviations of the data-points of a given X, $[Y_{i,x} - Y'_{i,x}]^2$. The second summation adds these squared discrepancies of observed from expected values over all the X-values. There are n_x computed points $Y'_{i,x}$ for each one of the k X-values, and nk different $Y'_{i,x}$-values in total. This measure of discrepancy is not directly comparable to ordinary least squares curve fitting. In regular regression the parameters of the straight or non-linear regression-line and the values of $\hat{Y}_{i,x}$ on that line are determined from all the observed X and Y-values. In the proposed measure of this one-to-many relationship the two equivalents of a regression line, the 'floor'-line is determined only from the lowest Y-values, and the 'ceiling'-line is determined only from the highest Y-values. After having established the boundary lines, Y-values falling outside the boundaries, below the 'floor' and above the 'ceiling', are treated like the other deviating Y-values inside these boundaries.

This coefficient of triangular correlation, really a 'one-to-many' correlation $R_{\blacktriangledown}$, proceeds like the regular coefficient of correlation 'r' and can be interpreted analogously.[1] In the case of complete conformity of the $Y_{i,x}$-data with the corresponding expected $Y'_{i,x}$ -values of this model, the double sum of the discrepancies,

Table D.1 Calculations for the rank-ordered values of Length of Employment at ERDA, X and their Salary, Y of 90 randomly selected employees*

No	X	$Y_{i,x}$	$Y'_{i,x}$	$[Y_{i,x} - Y'_{i,x}]$	$[Y_{i,x} - Y'_{i,x}]^2$
1	0	10.0	9.72	.28	.08
2	0	10.2	15.12	−4.92	24.21
3	0	11.9	20.53	−8.63	74.48
4	0	17.5	25.93	−8.43	71.06
5	0	22.1	31.33	−9.23	85.19
6	0	32.5	36.73	−4.23	17.89
7	0	42.5	42.13	0.37	0.14
8	1	9.9	10.08	−0.18	0.03
9	1	10.3	14.66	−4.36	19.01
10	1	11.2	19.23	−8.03	64.48
11	1	12.5	23.81	−11.31	127.92
12	1	15.2	28.39	−13.19	173.98
13	1	20.8	32.96	−12.16	147.87
14	1	31.5	37.54	−6.04	36.48
15	1	38.0	42.11	−4.11	16.89
16	2	9.2	10.44	−1.24	1.54
17	2	12.5	14.97	−2.47	6.10
18	2	17.5	19.49	−1.99	3.96
19	2	23.5	24.01	−0.51	0.26
20	2	31.0	28.53	2.47	6.10
21	2	37.0	33.05	3.95	15.6022
22	2	38.2	37.57	0.63	0.40
23	2	42.0	42.09	−0.09	0.01
24	3	11.5	10.81	0.69	0.48
25	3	15.0	17.32	−2.32	5.38
26	3	16.2	23.31	−7.11	50.55
27	3	19.0	29.57	−10.57	111.72
28	3	33.5	35.82	−2.32	5.38
29	3	42.0	42.07	−0.07	0.01
30	4	11.3	11.17	−0.13	0.02
31	4	12.5	18.89	−6.39	40.83
32	4	18.5	26.61	−8.11	65.77
33	4	28.3	34.33	−6.03	36.36
34	4	42.5	42.05	0.45	0.20
35	5	11.8	11.53	0.27	0.07
36	5	22.5	19.15	3.35	11.22
37	5	31.2	26.78	4.42	19.54
38	5	38.6	34.41	4.19	17.56
39	5	41.5	42.03	−0.53	0.28
40	6	12.2	11.89	0.31	0.10
41	6	18.0	18.17	−0.17	0.03
42	6	26.3	23.94	2.36	5.57
43	6	30.6	29.96	0.64	0.41
44	6	36.0	35.99	0.01	0.00
45	6	42.0	42.01	−0.01	0.00
46	7	12.6	12.25	0.35	0.12
47	7	39.0	41.99	−2.99	8.94
48	8	12.5	12.61	−0.11	0.01
49	8	16.8	16.81	−0.011	0.00
50	8	22.0	21.00	1.00	1.00

* These 90 employee records were selected at random from the 2,172 records on which the sex-discrimination study of this lawsuit against the management at ERDA was based.

Table D.1 (continued)

No	X	$Y_{i,x}$	$Y'_{i,x}$	$[Y_{i,x} - Y'_{i,x}]$	$[Y_{i,x} - Y'_{i,x}]^2$
51	8	28.1	25.19	2.91	8.47
52	8	30.2	29.39	0.81	0.66
53	8	33.3	32.87	0.43	0.18
54	8	39.5	37.78	1.72	2.96
55	8	43.0	41.97	1.03	1.06
56	9	13.0	12.97	0.03	0.00
57	9	37.5	41.95	−4.45	19.80
58	10	12.8	13.33	−0.53	0.28
59	10	18.0	22.87	−4.87	23.72
60	10	30.6	32.40	−1.80	3.24
61	10	40.0	41.93	−1.93	3.72
62	11	13.0	13.70	−0.70	0.49
63	11	23.1	18.40	4.70	22.09
64	11	27.5	23.10	4.40	19.36
65	11	32.5	27.80	4.70	22.09
66	11	35.5	32.51	2.99	8.94
67	11	39.6	37.21	2.39	5.71
68	11	42.1	41.91	0.19	0.04
69	12	15.0	14.06	0.94	0.88
70	12	20.0	23.33	−3.33	11.06
71	12	30.2	32.61	−2.41	5.81
72	12	37.6	41.89	−4.29	18.40
73	13	14.0	14.42	−0.42	0.18
74	13	25.5	21.28	4.22	17.81
75	13	32.5	28.14	4.36	19.01
76	13	39.8	35.01	4.79	22.94
77	13	41.5	41.87	−0.37	0.14
78	14	15.3	14.78	0.52	0.27
79	14	20.0	19.29	0.71	0.50
80	14	30.1	23.80	6.30	39.69
81	14	32.6	28.31	4.29	18.40
82	14	37.5	32.83	4.67	21.81
83	14	41.1	37.34	3.76	14.14
84	14	42.5	41.85	0.65	0.42
85	15	14.8	15.14	−0.34	0.12
86	15	27.0	20.74	6.62	39.19
87	15	35.2	25.82	9.38	87.98
88	15	40.0	31.15	8.85	78.32
89	15	40.6	36.49	4.11	16.89
90	15	42.0	41.83	0.17	0.03
					1, 832.06

connecting vertically each $Y_{i,x}$ with its nearest $Y'_{i,x}$,-value, becomes zero and $R^2_{\blacktriangledown} = 1$. Floor and ceiling are the estimates of the border-values for a given X. Every deviation of observed $Y_{i,x}$-values from their expected $Y'_{i,x}$-value reduces the credibility of the model assumptions for the given set of data. It is an indication that there are other forces at work in addition to those represented by the X-values used in that $R_{\blacktriangledown}$, whose effects are 'hidden' in the given level of aggregation of the data. The difference with regular correlation is that the residual sum of squares is

computed from many $Y'_{i,x}$ values instead of from only one $\hat{Y}_{i,x}$ for each X. The same, of course, holds for the coefficient of triangular determination, $R^2_{\blacktriangledown}$.

In the calculations the lowest and highest $Y'_{i,x}$ are placed on the two boundary lines, while the remaining $n_i - 2$ $Y'_{i,x}$-values for a given X are assumed equally spaced in-between. Although there is no reason why one would exclude these boundary values for placement of the n $Y'_{i,x}$-values, one could also determine the n+1 intervals excluding the boundary values. Such difference in computing, however, had only a minimal effect on the final sum of squared residuals in this simulation.

These computed reference values of Y are assumed to be equally distant between a "floor" value and a computed "ceiling" value of Y, as shown[2] in Table D.1.

The general equation for these computed $Y'_{i,x}$ points is:

$$Y'_{i,x} = Y'_{L,x} + [(i-1)(/)(n-1)][Y'_{U,x} - Y'_{L,x}]$$

where $Y'_{L,x}$ is the lower boundary value or 'floor" of this regression relationship, preferably determined by the least-squares procedure applied to to all the lowest $Y_{i,x}$-data – the subscript L stands for 'lowest value for a given X-value of the floor equation. $Y'_{U,x}$ is the upper boundary value or "ceiling," determined in analogy to the 'floor' from the highest $Y_{i,x}$-data. It represents the highest $Y'_{n,x}$-value. The equation for the ith partition point, for the case of n Y- observations that correspond to a given x, can be restated as

$$Y'_{i,x} = Y'_{L,x} + Y'_{u,x}[(i-1)/(n-1)]$$

If b_o is the intercept and b_1 the slope of the 'floor' or lower boundary equation, these lines can be stated as:

$$Y'_{L,x} = b_{L,o} + b_{L,1}\,X$$

and the 'ceiling' or upper boundary equation as:

$$Y'_{Ux} = b_{u,o} + b_{U,1}X$$

The *i*th partition point is the proportional, linearly interpolated point between these two boundary values for a given X.

In the following this simple partitioning by interpolating the boundaries is illustrated with the data of the cases #30 to #34, all having the same X=4, (see Table D.1). The same procedure interpolating for five $Y'_{i,x}$ –values would also be applied to the five cases #35 to #39, all having the value X=5, as well as the five cases #73 to #77 that have the same value of X=13. The following formula for the partition of the five observed $Y_{x=4}$ can be used for any X-value with five Y-observations. Here follow the computations for X = 4:

$$Y'_{i=1,x=4} = Y'_{L,x=4} = 9.7213 + 0.3631\ X = 11.1737$$
$$Y'_{i=2,x=4} = .75^{*}Y'_{L,x=4} + .25^{*}Y'_{U,x=4} = 17.8246 + 0.2659\ X = 18.8882$$
$$Y'_{i=3,x=4} = .50^{*}Y'_{L,x=4} + .50^{*}Y'_{U,x=4} = 25.9279 + 0.1704\ X = 26.6239$$
$$Y'_{i=4,x=4} = .25^{*}Y'_{L,x=4} + .75^{*}Y'_{U,x=4} = 34.0312 + 0.0750\ X = 34.3312$$
$$Y'_{i=5,x=4} = Y'_{u,x=4} = 42.1346 - 0.0204\ X = 42.0530$$

For all X-values who have five Y- values the corresponding $Y'_{i,x}$ -values can be determined by those five equations for the respective X-value. It should be noted that the boundary equations $Y'_{L,x}$ and $Y'_{U,x}$ are the same for the entire set, regardless of how many points correspond to a given X.

The sum of residual squares – values of Y "unexplained by or not associated with the variable X " – is 1,832.0628. Comparison of this residual with the total sum of squares 11,531.9693 provides the triangular coefficient of determination $R^2_{\blacktriangledown}$ in this case it is (11, 531.9693−1, 832.0628)(/)11, 531.9693 = 0.8411318. The suggested interpretation of this $R^2_{\blacktriangledown} = 84\%$ would indicate that, within the boundaries established by the floor and ceiling, about 84% of employee salaries Y at that agency are associated with the variable X, their years of employment at that agency, in an approximate, somewhat fuzzy manner, due to other variables that have not yet been considered.

Additional information about these employees would allow to reduce the distance between the floor and ceiling, the lower and upper limits, and show more clearly a stronger association between the same X and Y-values, length of service and salary, reducing that fuzziness. In regular multiple regression one would have to add additional explanatory variables to improve its R^2. In the case of $R^2_{\blacktriangledown}$ this would be achieved by forming sub-aggregates and pursuing this analysis on a more detailed level of aggregation, e.g. by gender of the employees, and/or by the level of their education. Then the floor and ceiling equations would be closer together and the scatter of the conditional Y-values within would give a clearer, less fuzzy picture.

Compared with the coefficient of determination, the regular correlation model of a one-to-one relationship between length of service X and Salary Y of these same data would indicate $r^2 = 0.09489$. In other words, only 9.5% of these salaries appeared to be associated with length of service. That is approximately one ninth of $R^2_{\blacktriangledown}$, barely a relationship of the single-linear equation kind in these in the face of compelling reasons to expect a closer association, given the federal government regulations for employment and promotion. It seems logical to expect some rather evident relationship but not of the standard regression type. The square root gives the coefficient of triangular correlation $R_{\blacktriangledown} = 0.91713$ which compares to r = 0.30803 which indicates a weak relationship of the least-squares linear kind that is customary in the sciences. The necessary calculations for the triangular correlation/regression are simple, including the algebraic determination of the two boundary equations.

Let me briefly sketch a possible test which might help decide whether to prefer the suggested fuzzy correlation or the regular succinct but often less meaningful regression model. In the latter, the Y-values are assumed to be normally distributed

and centered on the $\hat{Y}'$ value of the regression line. In the $R_{\blacktriangledown}$ measure the individual Yi,x values are treated as residuals from their $Y'_{i,x}$ vertically equally distant values, which correspond to a rectangular-type distribution model. Such parallel computations would allow a direct comparison of these two different ways to measure association between variables for the same data set. First treated as normally distributed about the regression line, assumed to be homoscedastic along the entire range of X-values, then also treated in the manner proposed for the less stringent assumptions of this model that would be better suited for usually heteroscedastic social-science data. If there is a substantial difference in the "Sum of Squares of Errors", the SSE values, the triangular regression model should be considered to be better suited.

The simulated example intentionally modeled a situation of little promise to show that even there a meaningful statement of a non-central relationship between variables is possible.

Notes

1. (Eplained Variation) (/) (Total Variation) = (Total Variation – Unexplained Variation) (/) (Total Variation)
2. These $\hat{Y}_{i,x}$ points are spaced differently depending on the number of Y values for a given X. Two $\hat{Y}_{i,x}$ values are always placed on the boundary values, the others are equally spaced between them. If there are only two Y values, like in the cases of X = 7 and X = 9, these $\hat{Y}_{i,x}$ points have the values of their boundary equations, This is a convenient but not the only possible way to proceed. If there had been only one Y value to a given X the $\hat{Y}_x$ would be computed as the exact mid-value between the corresponding values of the floor and the ceiling Equations.

Appendix E
Appendix to Chapter 10

E.1 Random Sampling and Sampling Distributions

The following is to refresh the reader's memory. Let a simple numeric example show what is involved. Assume a population of 10 persons, identified by letters A through J and consider an associated characteristic x for each person. Then a sample of 2 is to be selected, the sample arithmetic mean of the characteristic x determined to estimate the population mean μ_x of that characteristic x of each person. Any one of the 10 could be chosen to occupy the first place in the sample, and anyone from the 9 remaining to be the other person selected into that sample. There are $10 \times 9 = 90$ possible selections of 2. As a sample consisting of A and B and another consisting of B and A would be the same for determining the arithmetic mean, it should be disregarded as redundant. There are then 45 really different possible selections of a samples of two. In mathematical terms, [10 * 9](/) 2 = or more generally,[1] N!(/) $[(N - n)!(n!)] = (^{N}_{n})$ As only the n places of that sample are to be filled , the complete factorial of N is to be limited to the first n places by division with the factorial of $(N - n)!$ The redundant, n possible re-arrangements within the selected sample are also to be eliminated by division with n!

Then the $\bar{x}$-values for each one of the 45 different samples are computed and arranged in a frequency distribution,[2] called the 'sampling distribution of $\bar{x}$'. It is centered on the arithmetic mean of the population μ (more precisely μ_x) and also equal to the arithmetic mean of those 45 $\bar{x}$-values, $\mu_x = \mu_{\bar{x}}$.

Now consider a somewhat larger population (say N = 5000) again taking a sample of only n = 2. There are (5000)(4999)(/) 2 = 12,497,500 different possible selections of a sample of 2. A sample of size 10 would generate $(5000 \times 4999 \times 4998 \times 4997 \times 4996 \times 4995 \times 4994 \times 4993 \times 4992 \times 4991)(/)(10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1) = (9.678073483)^{36}/(3, 628, 800) = (2.667017604)^{30}$ possible, different selections of n = 10. To write down this gigantic number, one would have to move the decimal point 30 positions to the right, and fill the newly opened, many empty spaces with 0. This is the astronomic number of possible, different samples of size n = 10.

Assume again that the purpose of taking a sample is to estimate the arithmetic mean of that population of N = 5,000. Then assume that the arithmetic mean $\bar{x}$

is to be computed for the quantitative characteristic x in each one of these possible samples of size n = 10. Then these zillions of possible sample $\overline{x}$-values are to be arranged as a frequency distribution – let a computer do that. This 'sampling distribution of $\overline{x}$' will approximate a symmetrical, bell-shaped curve, closely resembling a mathematically perfect Gaussian or normal distribution. The center of that essentially symmetrical distribution coincides with the arithmetic mean of these zillions sample $\overline{x}$-values, $\mu_{\overline{x}}$, and also coincides with the arithmetic mean of the population, μ_x. The dispersion or spread of this 'sampling distribution' of all possible $\overline{x}$-values depends on the dispersion of the individual X-values of the population, N = 5000, and on the sample size. The standard deviation of this 'sampling distribution' of $\overline{x}$, called the 'sampling error of $\overline{x}$' is $\sigma_{\overline{x}}$. In mathematical terms, $\sigma_{\overline{x}} = \sigma(/)\{(\sqrt{n})[(N-n)(/)(N-1)]^{1/2}\}$. When the sample size n increases, this $\sigma_{\overline{x}}$ becomes smaller, but only by the square root of n. A tenfold increase in sample size e.g. from n =10 to n = 100 only causes a reduction of the spread of that 'sampling distribution of $\overline{x}$' by a factor of 3, not 10. To get that 'sampling distribution of $\overline{x}$' more closely concentrated around the population arithmetic mean,[3] μ_x, the sample would have to be quite a bit larger than n = 10. In some situations the dispersion, – σ_x in the numerator – can be reduced by judiciously stratifying the population and sampling these more homogeneous sub-populations, then combining these separate results with weights proportional to the sub-populations.

Although familiar to most users of socio-economic statistical data, this is reviewed here to clarify what 'statistically significant' really means, to avoid misunderstanding these basics of statistical inference, and to remind data users why theorists' insistence on random sampling particularly 'simple random sampling' is important for valid inferences.

E.2 Statistical Significance

What was discussed concerning the 'sampling distribution of $\overline{x}$' also holds generally for 'sampling distributions' of most other statistical measures, e.g. the sample slope 'b'. This understanding of 'sampling distributions' for most all measures computed from a sample and their shape and connection to the size n of the sample is the basis for the application of statistical inference in particular, the concept of 'statistical significance' tested as a 'Null hypothesis' H_0. Consider the study of a possible relationship in a population between two characteristics, represented as the variables X and Y. H_0 would be formulated as 'No relationship exists between the characteristics x and y in the population', which means that the slope of the regression line would be $\beta = 0$. When H_0 is true, there is a chance that a particular sample was randomly selected from the low end or from the high end of the sampling distribution of b-values. If that happens, the slope of such a sample b is substantially bigger or smaller than 0. A large b-value could happen e.g. when some of the elements or 'statistical-counting-units' selected from such a population into that sample had small values of x and y, while other randomly selected 'statistical-counting-units' happened to have large values for x and y, resulting in $b > 0$. For such a sample

to occur, showing a strong relationship and a slope b $\neq$ 0, there is a small chance, referred to as the 'Level of Significance' to be selected from a population that supposedly – according to the null-Hypothesis – has no relationship between x and y. The probability of that happening is called alpha, α, and can be set at 5%, or at lower percentages such as 1 or 0.1%, the situation in which the statistical counting units in such a sample would show a strong relationship between its x and y values, leading to the erroneous conclusion, that such a positive or negative relationship between x and y apparently also exists in the population.

The computer-produced probability statements usually would give the probability of a b-value as large or larger – or as small or smaller – than that obtained in the sample, from the tail end of the sampling distribution of b-values centered on $\beta = 0$ ('no relationship between X and Y in the population'). A computer generated P-value (e.g. of 0.08) would indicate that 8% of the sampling distribution of b-values can depart from a hypothesized $\beta = 0$ as far as that obtained in the sample, or farther. That P = .08 would be the probability of a b-value to come from that tail-end of the sampling distribution of b-values that begins at the obtained b-value. Such a result, in fact all P-values higher than 5% would be considered 'statistically significant at the 5% level'. More is to be said in the following sections. It does not insist that the sample slope actually belongs to a population of 'no-slope', but is a warning not to be too ready to believe to have received the final answer in that sample result. The P-value, e.g. 8% is the likelihood of drawing a premature, wrong conclusion.

Computer programs – such as the popular statistical packages SAS, SPSS or MINITAB – treat all data as if they were random samples, taken from some large, unspecified population. Computer programs do not request proof that the data have been selected by a valid random procedure from a properly identified population. Nonetheless, citing computer-generated probabilities of 'levels of significance' has become a requirement for research and a prerequisite for getting it published.

But few published studies in the social science literature use data that are true random samples. The data used in political-, social science and economic research are more likely to be whatever data became available to the researcher, perhaps entire 'populations' or part of a 'population', but not true random samples. In those instances the computer-supplied probability figures concerning 'statistical significance' should not be taken for what they appear to be.

It may be worth repeating, how difficult it really is to create a sample that is truly random.[4] The 'haphazard' or casual selection of 'statistical-counting-units' at the discretion of the interviewer or researcher certainly does not satisfy the concept of 'simple or other random sampling.'

The conclusion: in most real-life applications, the computer produced 'statistical significance' figure does not have the meaning statistical theory attributes to it. When a sample is not truly random, these probabilities do not lend legitimacy to the concept of 'statistical significance.'

Now consider another simplified example to clarify the meaning of 'statistical significance.' Assume the overly simple situation of a 'population' N = 100 in which each element or unit has two variables, an X-value and a Y-value. Twenty five cases (elements or 'statistical-counting-units') in that simulated population have

generally small X-values and small Y-values, another 25 cases have large X-values and large Y-values, a third group of 25 has small X-values and large Y-values, the last 25 cases have large X-values and small Y-values. In each one of these four groups the X and Y values are scattered, not identical. Yet, these $N = 100$ cases taken together, indicate no relationship between its two variables X and Y. The slope of that population regression line, $\beta = 0$ is a horizontal line expressing this supposedly non-existing relationship between X and Y.

Next, assume that a sample of $n = 10$ units is selected from that population. Various things can happen: the sample regression coefficient b of the variables x and y may indicate 'no relationship,' $b \approx 0$. It could also happen that the selected ten units may indicate a negative relationship, $b < 0$ when some of the selected units have small values of x and large y-values, while other units with large values of x have small y-values. It could also happen that $b > 0$ when some of the 10 randomly selected units have small values of x and small y-values, while other selected units in that sample have large values of x and large y-values. You can see that, depending on which units were selected into the sample, its sample slope b may or may not reflect the slope $\beta = 0$ of this population.

Then assume that all possible samples, or groupings of 10 elements, are assembled in a systematic manner and the slope b of the regression line for each sample is computed The result of '100 chose 10' are 17,310,309,460,000 different groupings or 'combinations' of 10 elements – at least one element being different between any two groupings – a procedure referred to as 'sampling without replacement'. For each one of these selected samples of $n = 10$ the slope b of the regression line would be computed, yielding the hard-to-imagine number of 17,310,309,460,000 different sample b-values ! When plotted, these 17 zillion sample b-values would form a huge, bell-shaped, symmetric frequency distribution, the **'sampling distribution of b,'** following closely the Gaussian or normal distribution. At the center of this hypothesized distribution, on the horizontal axis, the arithmetic mean of all the b-values, $\mu_b = 0$, that corresponds to the slope $\beta = 0$ of this hypothetical population. Even in this unrealistically small setting of $N = 100$ and $n = 10$, the numbers in the 'sampling distribution of b' are staggering.

Most of these sample b-values, plotted on the horizontal b-axis, roughly 68%, are within $\pm 1\sigma_b$ of the center, $\mu_b = \beta = 0$. The spread or dispersion of that distribution of b-values depends on how dispersed the X-values of the 100 'statistical-counting-units' are in $N = 100$. The wider these values are spread, the wider and flatter will also be its bell-shaped sampling distribution of b-values. The standard deviation of that sampling distribution, expressing a measure of its dispersion, is called the '**standard error of b**'[5] (denoted σ_b) and is always much smaller than the standard deviation σ_x of the original population.

E.3 Levels of Significance

Ideally, every sample should reflect the situation in the population. In most samples, however, that is not the case, due to the **'sampling error.'** It has become

standard to accept 95% of the samples in the sampling distribution as not rebutting H_0. It is also standard to reject H_0 at the 5% level, (alpha) $\alpha = 5\%$ of all samples from both tail ends of the sampling distribution taken together, as the '**5% level of significance.'** To determine those values on the horizontal b-axis (of the sampling distribution of b) e.g. for a two-sided test of H_0, that leaves 2.5% of all the b-values of the sampling distribution in either of its two tail ends, $(0.025) \times 17{,}310{,}309{,}460{,}000 = 432{,}757{,}736{,}400$. In other words, there will be 432,758 million b-values in the lower tail end, beyond the point, called $b_{0.025}$, and another 432,758 million b-values in the upper tail-end, beyond the point $b_{0.975}$. These two points, symmetrically placed, mark the **five percent** of all the possible sample b-values from that sampling distribution of b-values, from a population of $N = 100$ in which there is no relationship between the X and Y variables, a slope of the regression line $\beta = 0$. From the beginning of this kind of statistical analysis the 5% and 1% 'levels of significance' were introduced as probabilities small enough to justify rejecting H_0 – namely rejecting that there is 'no relationship between the variables X and Y in that hypothetical (unknown) population of $N = 100$'. These 432,758 million b-values that are at or below the critical $b_{0.025}$ have larger negative b-values, and as many above $b_{0.975}$ have larger b-values with a positive sign. Some of these negative or positive sample 'outliers' deceptively may even have very large b-values.

Analogously another set of symmetrically placed b-values can be determined that excludes **one percent** of these 17 zillion b-values. On the low end of that sampling distribution, the value $b_{0.005}$ leaves $(0.005) \times 7{,}310{,}309{,}460{,}000 = 86{,}551{,}547{,}280$ that is 86,552 million b-values at or below this point. The other, mirror-image critical value on the horizontal b-axis, $b_{0.995,}$ leaves another 86,552 million b-values, that is half of one percent of these 17 zillion possible sample b-values, at or above it in the upper tail-end of that bell shaped sampling distribution of b-values.

When one takes $n = 10$ individual elements from that population of $N = 100$ units, by an approved random procedure – which is the equivalent of selecting at random one sample result from the 17 zillion possible samples – and computes its b-value for the regression between the variables x and y, such a sample b-value should fall anywhere between $b_{0.025}$ and $b_{0.975}$ with a 95% probability. Analogously, the probability of such a sample b-value to fall between the critical values $b_{0.005}$ and $b_{0.995}$, should be 99%.

Assuming one obtains a sample's b-value that falls outside of $b_{0.025}$ or $b_{0.975}$, the two-and-a- half percent points, one would be inclined to conclude that this sample was taken from a population in which the variables X and Y were correlated (or 'not uncorrelated'). One would reasonably conclude that there is a negative relationship in that unknown population if $b < b_{0.025}$, or a positive correlation if $b > b_{0.975}$. It would be only 5% likely that such a sample stemmed from the hypothesized population in which there was no relationship between X and Y. In other words, this would not be a very likely event, and the risk of misjudging is only 0.05 of **falsely rejecting the hypothesis $H_0 : \beta = 0$**, called a Type I error. In that case one would conclude that the sample b-value is **statistically significant at the 5% level**, usually indicated by a single asterisk *.

If the b-value of one's sample exceeds the $b_{0.005}$ or the $b_{0.995}$ values, one would be justified with even more reason to reject the null-hypothesis 'no relationship between x and y in the population.' The chance of drawing the wrong inference has an even smaller probability. Such a sample b-value would be declared '**statistically significant at the 1% level**', often indicated by **. In this, and the previous instance, one would reject the null-Hypothesis, H_0: $\beta = 0$ substituting it with a point estimate of the population $\beta \cong b$. Three and more asterisks are to assure the analyst with an even greater probability that one can count on the population having a similar slope as the sample, indicated by ***

E.4 Interpreting Statistical Significance

The application of the foregoing to a real-life situation, in which the true population parameter is not known but is to be estimated by a random sample, requires a small leap of faith. In real life usually much larger sample sizes are involved. The computer will use a mathematical approximation to estimate the Sampling Distribution of the parameter in question, using the required information from the sample data as estimates to determine the sampling error – in this case the standard deviation of the sampling distribution of b. The computer print-out usually does not show the critical values $b_{0.005,}$ $b_{0.025}$ and $b_{0.975}$, $b_{0.995}$. Instead, most statistical packages print the probability of 'a result as large or larger than the one obtained, given that the Null Hypothesis is true'. The smaller that probability, the less likely is the Null Hypothesis true, and can be rejected with correspondingly greater assurance. *The* unspoken *condition for that process of statistical inference* is that the *sample* in question *was selected by a true random procedure,* as if selecting that sample from such a sampling distribution of the measure in question instead of selecting the sampling units (elements) directly from the population. In this example it would be the 'statistic' b, the slope of the sample regression line. These probabilities, and the implied assurances about the inference, would be less reliable or even invalid if that were not a true random sample.

Some afterthoughts: The larger the sample, the more likely will the Null Hypothesis be rejected. (That is so because the sample size 'n' enters into the denominator of the 'sampling error'- the standard deviation of the sampling distribution. The larger n, the smaller that 'sampling error' and the narrower that 'sampling distribution'). Most samples in political- or social science studies are large enough to be able to often reject $\mathbf{H_0}$. If n is large enough one may reject $\mathbf{H_0}$ at any of the customary levels of significance, but the result may be meaningless and irrelevant for the situation being studied, or as accountants have it, may not be material. In other words, statistical significance of a result does not necessarily mean that it is of value in an investigation. That question has to be judged from a subject-matter point of view, beyond the statistical- probabilistic point-of-view.

If it is obvious that the data used in a research are not a random sample, but e.g. just 'all the data that one could get hold of' one might use the probabilities of the

'levels of significance' to make a declaration that shows the researcher's understanding of these statistical concepts while at the same time distancing him/herself from an illegitimate use of the procedure, and also satisfying the demands of reviewers of professional journals, as follows: "...these data are not a random sample but e.g. a population. If, however, these data had been a randomly selected sample, then the results of this research would have allowed to reject the Null-hypothesis H_0 at such and such a 'level of significance,' which supports the authors suggested hypothesis ..."

What if one cannot reject the Null Hypothesis? Check if the fault lies with the wide dispersion among the sample elements. That can happen if in the data different, relatively homogeneous subsets were intermixed. If that were the case, try to separate these subsets and compute the intended measures separately for each group. Otherwise go back to the 'drawing board' and rethink whether the underlying working Hypothesis was appropriate. Perhaps a re-wording of that thesis and some re-calculating of the data can save the effort already expended for a research project. In the end the probabilities of 'levels of significance' can be informative, even if they should not be taken at face value because the data were not a randomly selected sample. In any case, not being able to reject H_0 **does not mean** that H_0 is true. Alternative hypotheses could be formulated and considered. In most instances, the matter of statistical significance should be treated as an interesting but not a determining feature in the interpretation of socio-economic data, regardless of whether one can or cannot reject H_0.

Notes

1. The exclamation sign !, "factorial of" is a mathematical symbol, an operator, that orders the multiplication of that number by all ordinal numbers that are smaller than itself, in descending order down to the number 1. In this case, $(10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)/[(8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)^*(2 \times 1)] = 10!/[(10-2)!^*2!]$ or more generally , $[(N)(N-1)(N-2)\ldots 3^*2^*1]^*[N-2)(N-3)\ldots 3^*2^*1]^*[2^*1]$.
2. Depending on the x-values of the 10 persons in that miniature 'population' the distribution of the $\overline{x}$-values will be less spread-out than the 10 different x-values in that population, and tend to be somewhat more symmetrical than the distribution of the 10 population x-values. At any rate, the average of the 45 $\overline{x}$-values, $\mu_{\overline{x}}$ (mu with subscript $\overline{x}$) will be the same as the arithmetic mean of the population μ_x. There is also a relationship between the spread or dispersion in the population data and the dispersion among the 45 $\overline{x}$-values. Due to the very small sizes of population and sample that relationship is more complicated than when population and sample size are much bigger.
3. Because $\sqrt{10} = 3.16$ and $\sqrt{100} = 10$. The reduction in $\sigma_{\overline{x}}$ is 10 / 3.16 = 3.16 not 100 / 10 = 10.
4. When I taught Statistics for Auditing in the Program for the degree of a Master in Accounting, at Georgetown University, it surprised me and became clear to me, how difficult it really is in an audit situation to create the necessary conditions that would qualify, and be accepted in a court of law, as a valid statistical random sample that guarantees the required impartiality of an audit. It made me aware of the stringent conditions, required to make the results of a statistical audit sample stand up in a court of law The amount of effort and circumspection necessary to select a data-set as a sample that is truly random – and hopefully also representative

of the situation – may surprise everyone who has not dealt with audit sampling. O. Winkler "Secret Allies- Accountants and Statisticians come from different worlds but have a common Methodology and Mission" *Management Accounting,* June 1985. See also H. Arkin, Endnote 22, Chap. 10.

5. $\sigma_b = \sigma_{y \cdot x}(/)\{\Sigma X^2 - n\,\overline{X}^2)^{1/2}$ The farther apart the X-values, the bigger that part of the formula. But on the other hand, the larger the sample size n, the more that portion is reduced.

Index