

Book Reviews

Books for review are to be sent to the Book Review Editor Jan Wretman, Statistical Research Unit, Statistics Sweden, S-11581 Stockholm, Sweden.

ANDREWS, D.F. and HERZBERG, A.M., Data: A Collection of Problems from Many Fields for the Student and Research Worker <i>Norman R. Draper</i>	343
BENGTSOON, T., FRIDLIZIUS, G., and OHLSSON, R. (Eds.), Pre-Industrial Population Change: The Mortality Decline and Short-Term Population Movements <i>Erland Hofsten</i>	344
FINGLETON, B., Models of Category Counts <i>Herdis Thorén Amundsen</i>	345
FOWLER, Jr., F. J., Survey Research Methods <i>Sven Eriksson</i>	346
PURI, M. L. and SEN, P.K., Nonparametric Methods in General Linear Models <i>Marianne Frisen</i>	347
KIEFER, J.C., Collected Papers III: Design of Experiments <i>Yadolah Dodge</i>	347
TREWIN, D., (Ed.), Statistics in the Labor Market: The Role in Planning and Policy Formulation <i>Berndt Öhman</i>	349

Andrews, D. F. and Herzberg, A. M., Data: A Collection of Problems from Many Fields for the Student and Research Worker. Springer-Verlag, New York, 1985, ISBN 0-387-96125-9 and ISBN 3-540-96125-9, 442 pp., DM 138.

The subtitle tells it all. Seventy-one sets of data are given and there really is something for everybody. We have all heard of “The Number of Deaths by Horsekicks in the Prussian Army” data. See them on p. 18. How about “Yearly Yields of Grain and Straw for Eighteen Plots in Broadbalk, Rothamsted 1852–1925”? See pp. 22–29.

Perhaps you would like to read how the Lord Brounker measured the errors due to the recoiling of guns round about 1734? Start on p. 57. Do you know how thick the ozone column was at Arosa, Switzerland, 1926–71, measured in Dobson units? You will if you read p. 76. How much rain fell in Adelaide, Australia, 1839–1977? Find it on pp. 106–117.

Four hundred mice received various doses of Red Dye No. 40. How long did they live? Study pp. 240–243. Would you like to know the physical characteristics of 79 urine specimens? All right, better skip pp. 250–251, then! Do not miss the map of the Galapagos Islands

on p. 293, however. Are famous people less likely to die in the month preceding their birthday than at any other time? Study the data on pp. 431–433 and decide for yourself!

The data (also available in machine readable form from D. F. Andrews, Statistics Department University of Toronto, Ontario, Canada M5S 1A1, for a nominal charge) were contributed to the authors as a result of data requests made to 200 individuals. No analyses are given; one looks up the references oneself if one wants to see what analyses have been done in the past. There are “sets of data which have been used frequently to demonstrate statistical techniques... [and sets of data]... showing the complexity and variety of issues facing statisticians” (p. 1). Each set is preceded by its source, the name of the contributor (if there is one) and a description which, as one gathers from the preface, is either written by the contributor, or written by the authors and checked by the contributor, if one exists. Which is which is never clear, except of course when there is no contributor!

A few other minor mysteries remain. What is aetiology? (p. 199). What sort of mammal is the *Perognathus formosus*? (p. 287). What does “a large homogeneous pool of serum” look like? (p. 151). The “oral glucose toler-

ance test" has an ugly ring to it on p. 211. What does "a diet . . . of rum and damper" consist of besides rum? (p. 307). Perhaps we do not need to know any of these things to use the data but it does make the reader doubt that "All the problems can be readily understood by non-specialists" as is claimed on p. 1.

Let us put these minor quibbles aside, however. This is a magnificent, fascinating book with 71 intriguing sets of data, ready and available to test your latest statistical theory, or to strain your new computer program, or maybe just to amuse or sadden you during a quiet read in the evening. The typesetting is excellent, the layout clear, the quality high and the typos few and far between. You will undoubtedly need this book at some time. Therefore, make sure your library has it on hand. Better, buy one yourself!

Norman R. Draper
University of Wisconsin, Madison
U.S.A.

Bengtsson, T., Fridlitzius, G., and Ohlsson, R. (Eds.), *Pre-Industrial Population Change: The Mortality Decline and Short-Term Population Movements*. Almqvist & Wiksell International, Stockholm, 1984, ISBN 91-22-00741-5, 419 pp., SEK 340.

This volume consists of revised versions of sixteen papers, which were presented at the International Symposium, organized by the Department of Economic History, University of Lund, Sweden in 1981. The symposium was devoted to two related topics, namely 1) the secular decline in mortality and the factors which might have initiated this process, and 2) short-term variations in a population and their relation to fluctuations in the pre-industrial economy. Five of the authors are from Sweden (Bengtsson, Fridlitzius, Jonsson, Ohlsson, Söderberg), two from Great Britain (Lee, Schofield) two from Canada (Norrie, Percy), and one each from Finland (Kaukiainen), France (Goubert), Italy (Corsini), Switzerland (Perrenoud), and USA (Richards). This review concentrates on only a few of the essays.

In the introductory essay, R. Schofield con-

siders the decline in mortality after 1750. He refers to the conventional view that improved food supply was the dominating factor behind the decline in mortality and, in particular, behind the disappearance of mortality peaks, thus resulting in an accelerated growth of population. Against this view Schofield maintains that many factors were at play and, in particular, that immunological aspects must be taken into account.

Schofield has based his argument on data for England, France and Sweden. In all three countries the decline in mortality was considerable. However, there were significant differences with regard to the ensuing population growth rates. Schofield argues that Sweden confirms the classical view that the increased rate of population growth was primarily caused by declining mortality, whereas in England increasing fertility also played an important role. Both the Swedish and the English pattern can be seen in the French data.

Schofield seems to have overlooked the fact that in Sweden the increase in fertility played a more important role than the decline in mortality for the rapid population growth which followed after 1810.

Schofield's essay is followed by five essays that all consider the decline in mortality and its causes. The conclusions in these essays are similar to Schofield's, i.e., they reject the view that the mortality decline was primarily caused by an improved food supply.

In section two of the book, one finds an essay by R. Ohlsson and T. Bengtsson based on an analysis of the data on crude birth rates, crude death rates and real wages. These data have been recorded since the beginning of Swedish vital statistics, i.e., from 1749. Utilizing spectral analysis to analyze the annual fluctuations, the authors come to the conclusion that there is a relation between real wages (and harvests) and crude birth and death rates. They maintain that "the results clearly indicate that the 18th century Swedish economy was mainly Malthusian in character."

The essay does not include any discussion about the meaning of the word "Malthusian"; it would seem that the authors are of the opinion that it is appropriate to use this term as soon as a short-term relationship can be established between the level of living and the crude vital rates.

To my mind the word "Malthusian" can only be used in a situation when there is a tendency for the population to grow more rapidly than the means of subsistence. It may be argued that such a tendency, if it exists, will be accompanied by short-term variations in population related to variations in the economic conditions. However, the opposite does not hold true; from the fact that there are short-term variations does not follow that a Malthusian rule applies. One may very well imagine a population that in the long run is constant because of an ecological equilibrium with the environment, but in which there are nevertheless short-term variations in the vital rates, e.g., caused by climatic variations.

As an overall opinion it may be said that the volume is ambitious and will no doubt prove very useful to research workers in the field. It might perhaps be added that future research in historical demography would benefit from closer contacts with research in other, closely related fields, such as demography proper, ecology and immunology.

Erland Hofsten
University of Stockholm

Fingleton, B., *Models of Category Counts.* Cambridge University Press, Cambridge, 1984, ISBN 0-521-25297-0 (hard cover), ISBN 0-521-27283-1 (paperback), 187 pp. £ 17.50 (hard cover), £ 6.95 (paperback).

The author's aim is to give an account of recently developed statistical methods for the analysis of data which can be organized in tables of category counts (contingency tables). The book is intended for social scientists. The emphasis is on application, and numerous examples of analyses of real data sets are given.

The construction of contingency tables is well explained in the first chapter, with a step by step introduction of models and methods for two-way tables. Nevertheless, the book favours readers with some knowledge of probability theory and statistical methods, and some familiarity with the analysis of contingency table data.

Log-linear models are the main models for the analysis in the book. It is pointed out that there are several different choices of parameter restrictions in use for log-linear models and that care must be taken in interpreting the parameters.

Multidimensional tables and saturated log-linear models are treated in Chapter 2. Fingleton warns against starting one's analysis by looking at two-way tables, and against placing too much confidence in the interpretation of collapsed tables from multidimensional data.

In Chapter 3, unsaturated log-linear models and their uses are treated. Model selection strategies, with the aim of choosing a parsimonious model to fit the data, are explained. After mentioning the selection of parameters based on saturated models, stepwise methods, and screening, the author treats more extensively the simultaneous test procedures (STP) as introduced by Aitkin. STP provides adjusted significance levels when several parameters are tested; see Aitkin (1979).

Rather late in the book, in Chapter 4, we find the theoretical assumptions about the distribution of the data underlying the methods already treated. Next, the author discusses problems that are introduced when the data are sampled by various designs from finite populations. It is not easy to give a clear account of this subject. The numerous references to recent literature may be of use to the reader.

Chapter 5 treats problems in analyzing data from large, unwieldy tables. A number of special models are introduced: uniform association, proportional odds, extreme ends, etc.

The shortcomings of the usual large sample methods when used for small samples, are mentioned, but "exact" test methods are not given.

Chapter 6 is named "Counts from Temporal Models." The main subject is panel data models and analyses, which are quite extensively treated.

The Appendix contains 13 listings of GENSTAT directives which have been used for examples in the text. Other well-known program packages are mentioned and used in some of the examples.

There is a large list of references, the most recent ones dating from the first part of 1983.

Some of the flaws common to statistical texts on this level occur also here: "Expected frequency" and "probability" sometimes mean the theoretical values but more often their estimates (a passage on p. 38 partly explains the usage). The same symbol is used for several (but not all) different statistics with an asymptotic chi-square distribution.

The book concentrates on Data Analysis. Occasionally, however, the author has some remarks about the connection between the interpretation of the statistical results and the real problem which the data are supposed to throw a light upon. This feeds the hope that more importance will gradually be given to the inclusion of the *a priori* theory and knowledge about the problem in the formulation of models and analyses for categorical data. In many cases it should be possible to bridge the gap between traditional statistical analysis and pure data analysis with the tools that are now at our disposal.

Reference

Aitkin, M. (1979): A Simultaneous Test Procedure for Contingency Table Models. *Applied Statistics*, 28, pp. 233–242.

Herdis Thorén Amundsen
University of Oslo
Norway

Fowler, Jr., F. J., *Survey Research Methods*. (Applied Social Research Methods Series, Vol. 1) Sage Publications, Beverly Hills, 1984, ISBN 0-8039-2347-3 (hardback), 0-8039-2348-1 (paperback), 159 pp., £ 17.75 (hardback), £ 8.95 (paperback).

Fowler is a social psychologist who has devoted many years to survey methodology and applied survey research within the social sciences. In the author's own words, the purpose of the book is "to give a sound basis for evaluating data collection procedures to those who would collect, analyze or read about survey data."

Most books with this ambition are completely dominated by the sampling aspects of a survey. However, contrary to what Fowler believes, there are other books that concentrate

mainly on the measurement process of surveys (e.g., Warwick and Lininger (1975)).

The outline of Fowler's book is as follows:

1. Introduction (8 pages)
2. Sampling (26)
3. Nonresponse (16)
4. Methods of Data Collection (13)
5. Designing Questions (26)
6. Design of a Questionnaire (8)
7. Survey Interviewing (20)
8. Preparing Survey Data for Analysis (8)
9. Ethical Issues in Survey Research (6)
10. Providing Information About Survey Methods (4)
11. Errors in Perspective (6)

In a methodological textbook on survey research, one expects a layout with figures, tables, and examples elucidating problems and methods with authentic data sets. There is not a single example of this kind in the book; there is only one figure and a few tables. The examples that are employed are in the text and are often of the type "Suppose we have the following data . . .," or consist of relatively general references to specific elements of certain studies.

The best and major part of the book is the section on measurement methodology. Here the reader has a lot to learn about choice of measurement method, formulation of questions and design of questionnaires.

What is missing is a chapter on "Study objectives and specification of the statistical problems." Under this heading, I would like to have found an exposition on target population and sampling frame (these two points are discussed in the book), elementary units, definitions of various study concepts, classifications to be used in the study, definitions of study variables, which parameters to estimate, domains of study (population subgroups for which separate results are desired), accuracy, precision and number of units in the sample. Drawing up a time schedule and estimating the total study cost are two other important questions that should be addressed in a book of this nature.

It is obvious that the author has come to applied survey research via the "social science route" and not the "statistical route." There are a number of unforgivable mistakes in the chapter on sampling. To give a few examples of elementary errors: The intention with stratified sampling is not always to produce a sam-

ple that is likely to reflect the distribution of the whole population. This is very inefficient in many cases. It is also incorrect that "stratification never hurts the precision of the sample." The author uses "odds" and "probability" as synonyms.

This is the first handbook I have ever seen which gives an erroneous formula for the standard error of the mean of a simple random sample. It is defined as the square root of V^2/n , where " V^2 " = the variance, the sum of the squared deviations from the sample mean." Notice that apart from the very strange notation V^2 , the term $n-1$ is missing in the denominator!

In summary: There are books which give a better overview of the whole survey process, e.g., the previously mentioned one by Warwick and Lininger (1975). The sampling part of Fowler's book does not meet acceptable standards. The chapters on measurements in surveys are, however, competently written. The potential reader, perusing this part of Fowler's book, should also complement it with another text on the planning and sampling aspects of a survey. A suitable complement, written for social scientists, is Kalton (1983).

References

- Kalton, G. (1983): *Introduction to Survey Sampling*. Sage Publications, Beverly Hills.
 Warwick, D. P. and Lininger, C. A. (1975): *The Sample Survey: Theory and Practice*. McGraw-Hill, New York.

Sven Eriksson
 Hässle Pharmaceutical Company
 Mölndal, Sweden

- Puri, M. L. and Sen, P. K.**, *Nonparametric Methods in General Linear Models*. Wiley, New York, 1985, ISBN 0-471-70227-7, ix+399 p., £ 57.80.

The book gives theoretical results for the case with a parametric linear model without any distributional assumptions. The methods treated are based on ranks. The book is mainly concerned with proofs of asymptotic normality (or chi-squaredness) and the associ-

ated results for tests and estimators.

This is by no means a handbook for the practitioner who wants robust techniques. The combination of the theory of ranks with the theory of linear models obviously requires heavy mathematics. Familiarity with advanced calculus, real analysis and matrix algebra as well as a good knowledge of probability theory, nonparametric statistics, linear models and multivariate analysis is assumed.

There is quite a number of exercises but most of them are of the type: "Provide a formal proof of Lemma X.X.X." The book is closely related to *Nonparametric Methods in Multivariate Analysis* (1971) by the same authors. The results of Chapters 4, 5, and 6 of the earlier book are generalized in the new book.

Reference

- Puri, M. L. and Sen, P. K. (1971): *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Marianne Frisén
 University of Göteborg
 Sweden

- Kiefer, J. C.**, *Collected Papers III: Design of Experiments*. (L. D. Brown, I. Olkin, J. Sacks, and H. P. Wynn, Eds.) Springer-Verlag, New York, 1985, (with the co-operation of the Institute of Mathematical Statistics), ISBN 3-540-96004-X, xxv+718 pp., DM 128.

Jack C. Kiefer, one of the foremost mathematical statisticians of this century, was born in Cincinnati in 1924 and died untimely at Berkeley in 1981 at the age of 57 leaving behind a monument of work that forms the foundation of many areas of statistics. He was a man of tremendous talent who contributed profoundly to many different areas of statistics, including sequential analysis, multivariate decision theory, and, in particular, the optimal design of experiments.

Inspired by A. Wald's (1943) paper on the "efficient design of statistical investigation," Jack Kiefer together with Jacob Wolfowitz revolutionized the science of experimentation by introducing decision theoretic ideas in the

field of design of experiments. He alone brought the theory of optimal design of experiments to its perfection.

This volume contains 45 of Kiefer's papers on design of experiments from 1958 to 1984, reprinted in chronological order. (The last paper was jointly written with H.P. Wynn and appeared three years after Kiefer's death.) The volume also contains a preface by A. S. Hedayat, a paper by H. P. Wynn entitled "Jack Kiefer's Contributions to Experimental Design" (first published in the *Annals of Statistics*, 1984), and seven commentaries on 28 of Kiefer's papers by C. S. Cheng, Z. Galil, T. Mitchell, F. Pukelsheim, J. Sacks, W. J. Studden, and H. P. Wynn.

This volume can be considered the bible on optimal design of experiments. As we know, the Bible is very hard to read and understand in its original version, which is also true with Kiefer's papers. Even specialists have a hard time reading Kiefer's papers. Many read a few pages and then trust Kiefer's professional dignity and reliability when quoting his results.

Having a collection of Kiefer's papers on optimal designs along with reviews from other statisticians who are active in the field in one volume is of great value. This remarkable volume contains many unsolved and exciting problems. Kiefer was extremely generous to share his research problems with others. If a young researcher is looking for unsolved problems in the area of optimal designs, he should absolutely read Kiefer's papers. This volume draws together the most important results of nearly three decades of research. It is indeed a book that no one can do any new work without referring to.

Kiefer did solid work but did not cover all areas. For example, he did no work on designs for comparing treatments with one or more controls, nor did he work on optimal designs for sequential experimentations. Repeated measurements designs and nonlinear models are other examples of areas where Kiefer did no research. It appears that Kiefer was a perfectionist. He did not want to solve anything easily for the sake of applications. Someone should rewrite Kiefer's work for practitioners. After all, he was a mathematical statistician and not a statistician. To date, Fedorov's (1972) book is the only book in the field of optimal design and is an important basic reference.

The comments on Kiefer's papers, written by former students and colleagues, give utterance to deep respect, admiration and affection for Kiefer. Just like Kiefer's papers, these comments also contain many exciting unsolved problems. For example, Pukelsheim at the end of his comments writes: "A question, which in the work of Kiefer (and of other authors) is not always treated adequately, is whether there exists an optimal design. Whereas for D-optimality existence poses no problem, there are other criteria for which an optimal design fails to exist; . . . To date it is an open problem whether the Equivalence Theorem of the general theory continues to hold when no optimal design exists; . . ."

Wynn reviews seven of Kiefer's papers and discusses in particular Kiefer's (1959) paper on "optimum experimental designs." Kiefer read this paper before a research methods meeting of the Royal Statistical Society on May 27th, 1959, with professor G. A. Barnard in the chair. A discussion followed by Tocher, Quenouille, Mallows, Curnow, Yates, Beale, Cox, Barnard, and Stone and a reply by Kiefer. With regard to this paper, Wynn writes:

"In this early sequence of papers the paper to the Royal Statistical Society stands out partly for the almost orchestral attack on it by the British School of experimental design stemming from the work of R. A. Fisher and F. Yates. This is one of the few published clashes between the Fisherian British school and the Waldian decision school for which J. Kiefer was ambassador on that occasion. Needless to say, the British had met their match; his reply is masterly.

The paper itself is a discursive thesis on optimum design with particularly useful sections on invariance and the support size of a design."

As Hedayat puts it in his preface to this volume:

"There are numerous examples of expansive writing in his papers which give his wide perception and share his ideas and methodologies with the reader. A reading of his papers on design is replete with examples of his scholarship, his innovativeness, ingenuity and strength as a researcher."

I fully agree with the preface and heartily recommend this volume to everyone who

wants to find out how useful mathematics is in the theory of experimental design.

References

- Fedorov, V. V. (1972): *Theory of Optimal Experiments*. Academic Press, New York.
 Wald, A. (1943): *On the Efficient Design of Statistical Investigation*. *Annals of Mathematical Statistics*, Vol. 14, pp. 134–140.

*Yadolah Dodge
 University of Neuchâtel
 Switzerland*

Trewin, D. (Ed.), *Statistics in the Labor Market: The Role in Planning and Policy Formulation*. (Lecture Notes in Statistics, Vol. 5) Marcel Dekker, Inc., New York, 1983, ISBN 0-8247-1912-3, vii+163 pp., SEK 561.

This book is a result of a symposium on labour statistics in Australia. It is a collection of papers, reflecting several topics and problems that were treated at the conference.

The first paper (Chapter 1) treats the impact of unemployment benefits on the level of unemployment. An increase in benefits seems to have some impact on total unemployment, but the impact is very different for different groups. It is also shown that a greater part of the unemployed gets benefit payments when payments increase, although the total number of unemployed does not increase considerably. A closer analysis shows that increased benefit payments have no impact on inflow into unemployment but some impact on the duration of unemployment.

The second paper argues for better statistical data concerning technological change. The third and fourth papers deal with the impact of immigration on the labour market. They show a complex picture. It is stressed that immigration affects not only the supply of labour but also the demand for labour and not only total supply of labour but also its characteristics (structure). The importance of immigration is judged very differently. According to Chapter 3, the proportion of overseas-born persons in the Australian labour force is very

small (p. 72), but they are a very substantial group (26%) according to Chapter 4 (p. 75).

One of the papers (Chapter 6) discusses the growing demand for statistical data on regional and local levels. Sample survey data are often not reliable below the state level. Census data are not timely and Australian register data provide little information on small domains. The paper analyzes different methods for estimating small domain employment statistics but concludes that we should perhaps “refrain from producing certain small domain statistics because [we] feel that the errors are too large or that they are not measurable” (p. 124). The last paper in the book gives an account of the role of labour statistics in the IMPACT project in Australia. This project seems to be a macroeconomic model for long-time planning and policy analysis.

There is no discussion of the concept and content of labour statistics. I think labour statistics – or statistics on the labour market – consist of information on:

- supply of labour: labour force, employment, unemployment,
- demand for labour: jobs, employment (or filled posts), vacancies,
- the function of the labour market: gross flows, mobility and so on,
- the price of labour: earnings, wages and labour costs,
- working conditions in a broader sense (than wages),
- labour market policy.

Chapter 5 is a paper on labour statistics in Australia. Evidently, the Australians define labour statistics in about the same way as I do above. One difference is that they include what they call other related statistics (above all a consumer price index). As is pointed out, nearly all economic statistics could be considered to be of some relevance (p. 108), but I think we agree that labour statistics are first and foremost concerned with the labour market. Our concept of labour statistics is therefore essentially the same.

The Australian paper also gives an account of the different types of statistics in Australia:

1. *Surveys*

- household surveys, e.g., the monthly population survey,
- employer surveys, to obtain labour force data and information on income, earnings and hours.

2. *Censuses*

- population censuses, every five years,
- censuses of business enterprises and establishments.

3. *Administrative data*

- often as by-products of administrative processes,
- a monthly civilian employees series, based on the state payroll tax, is discontinued, however, and will perhaps be replaced by a survey of employers.

The main source of statistics on the labour supply is the monthly population survey, comprising about 0.67 % of the Australian population. The sampling errors are too high, however, to allow regional estimates to be published. A survey of employers is proposed, as indicated, and it will provide estimates of employees and wages with an industry dissection.

Data on the demand side of the labour market are not as available as data on the supply side, the Australian paper asserts (p. 103). With regard to wages, there are annual and quarterly statistics on earnings. Moreover, the monthly labour force surveys collect information on the weekly earnings of wage and salary earners, once a year (in August). There is little information on costs of labour, apart

from wages and salaries. Industrial accident statistics are produced annually as are statistics on trade union membership, while statistics on industrial disputes are produced monthly.

A comparison with Swedish labour statistics shows many similarities. There are about the same types of statistics and about the same problems. The Australians have more statistics on the supply of labour than on the demand for labour, more on earnings than on labour costs. One difference should be mentioned concerning labour force surveys: the Swedish surveys collect data on individuals, not households, and earnings data are not collected.

This book contains papers of interest, from a statistical as well as from a labour economics point of view. Some of the papers are examples of how you can use labour statistics, some of them are primarily concerned with labour statistics as such. Together, they make a valuable volume on labour statistics, but the book is not intended to be used as a textbook on labour statistics. Rather it is written for specialists in the labour market field.

Berndt Öhman
Statistics Sweden