Book Reviews

Book Reviews

ASSOCIATE EDITORS

Burrows computes correlations (which are typically more than .7) between characters in all of Jane Austen's novels for their usage rates of the 30 most common words. He performs a principal-component analysis of the characters who have at least 2,000 words of dialogue. Because of the uniformly high and positive correlations, the first principal component, which accounts for 80% of the variance, is essentially the average usage rate for the 30 most common words. Of his second principal component (only 4.4% of the variance), Burrows says that the change from one loading extreme (Harriet Smith, whose prospects and character Emma tries to improve) to the other (Collins, the pompous, apologetic clergyman in *Pride and Prejudice*) is a movement from "garrulousness and intellectual indiscipline, through . . . civil and articulate speech-habits, to formality and dignity, and onward to pomposity" (p. 132).

Burrows finds a satisfying overall differentiation between the Bennet ladies (Elizabeth's mother and sisters) and those of Darcy's circle. To explain the high correlations of .926 and .945 between Elizabeth and Lady Catherine (Darcy's aunt) and between Elizabeth's father and Lady Catherine, Burrows observes that these "characters are united in a firmness of manner and control of ideas that readily distinguish them from Mrs. Bennet [Elizabeth's mother] and Lydia [Elizabeth's wild sister]" (p. 87). Firmness suggests, for example, low usage of weak forms of emphasis such as *quite* and *very*. A "firm control of ideas" relates to near-zero correlations with Burrows's second principal component.

Burrows continues, "Mr. Bennet and Lady Catherine are brought still closer by . . . expressions of authority and habits of command . . . which they share with Sir Thomas Bertram [a character in *Mansfield Park*]" (p. 88). Command is related to high usage of second-person pronouns, as when Sir Thomas, attempting to intimidate Fanny (the heroine) says "*You* are not to be judged by the same rule. *You* do not owe me the duty of a child. But, Fanny, if *your* heart can acquit *you* of ingratitude . . . ."

For heroines and heroes, Burrows computes usage rates for consecutive blocks of 1,000 words of their dialogue. Growth and increasing self-knowledge are satisfyingly reflected by changing patterns for the 30 most common words, and some pairs of lovers (e.g., Emma and Mr. Knightly) converge.

Burrows's statistical inexperience shows when he refuses to perform chi-squared tests (p. 40) that would give highly significant results for several characters whose expected values for *us*, although below the magic number of 5 that produces a warning on most statistical computing packages, are still above 4 (p. 40). Burrows (p. 38) makes the dubious claim that *of*, whose observed usage rates by characters vary from 10.6 per thousand to 37.6 per thousand, is more variable than *we*, whose rates vary from .85 per thousand to 16.2 per thousand, because larger chi-squared values are associated with the commoner word. At times he overinterprets differences in correlations that appear only in the second and third decimal places.

Burrows claims that "The evidence of the very common words bears at times on questions hotly contested in the critical and scholarly controversies of recent years . . . [and] increases our understanding in areas where literary criticism has seldom found firm ground" (p. 2). For example, the fact that Collins correlates most highly with Darcy tends to support the suggestion (made by numerous critics) that Collins is a parody of Darcy. In showing that some aspects of character and development have a basis in simple word counts, Burrows has countered the extreme deconstructionist view that literary analysis is completely subjective.

Burrows succeeds in his central objective of demonstrating that "*From no other evidence* than a statistical analysis of the relative frequencies of the very common words, it is possible to differentiate . . . appropriately among the idiolects of Jane Austen's characters and even to trace the ways in which an idiolect can develop in the course of a novel" (p. 2).

MALCOLM J. SHERMAN
State University of New York at Albany

REFERENCES

Lord, R. D. (1958), "De Morgan and the Statistical Study of Literary Style," *Biometrika*, 45, 282.

Mendenhall, T. C. (1887), "The Characteristic Curves of Composition," *Science*, 9 (supp.), 237–249.

——— (1901), "A Mechanical Solution to a Literary Problem," *Popular Science Monthly*, 60, 97–105.

Mosteller, Frederick, and Wallace, David L. (1984), *Applied Bayesian and Classical Inference: The Case of the Federalist Papers* (2nd ed. of *Inference and Disputed Authorship: The Federalist*), New York: Springer-Verlag.

Yule, G. Udny (1938), "On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship," *Biometrika*, 30, 363–390.

Statistics Explained: Basic Concepts and Methods.
R. Kapadia and G. Andersson. Chichester, U.K.: Ellis Horwood, 1987. 234 pp. $36.95.

This slim volume advertises itself on its front cover as a "serious and entertaining introduction to statistics." The book definitely keeps its promise to be entertaining. Potential readers, however, need to be made aware of the limited and very idiosyncratic coverage of statistics provided. Far from being an introductory statistics textbook, the book's scope and mathematical sophistication is closer to that of Huff's (1954) book.

With the exception of one brief chapter on inference toward the end, the word *statistics* is used throughout as synonymous with *descriptive statistics*. Thus the book has the obligatory chapters on measures of central tendency and variability and tabular, as well as pictorial, data summaries (Chap. 4–6). Correlation and regression (which occupy a brief Chap. 10) are the only other substantive statistical topics in the book. The book goes well beyond the usual coverage of descriptive statistics, however, by expanding on such topics as "common fallacies in [descriptive] statistics" (Chap. 2) à la Huff (1954) and the "power and ubiquity of [descriptive] statistics" (Chap. 1) and providing a listing and critique of various sources of official statistics (Chap. 3). Using newspaper headlines, clippings, and figures, the book's goal is to give its readers a "systematic training in analysing statistical information" (p. 11).

For the American market, one serious disadvantage is the almost exclusive use of British examples. (The exceptions are Swedish examples, leftovers from a Swedish version of the book first published by Andersson.) Although many of these examples serve perfectly well to illustrate the authors' points, I suspect that many Americans will be only marginally interested in unemployment rates in Wales or the percentages of adults in various British social classes reading particular British newspapers. In a book that makes extensive use of such examples, lack of interest in the content of these examples could be a serious handicap in keeping the readers' attention. Chapters on various sources of official statistics or British economic indicators and their interpretation lose much of their interest and utility this side of the Atlantic.

A brief and superficial chapter on chance-and-probability theory (Chap. 7) is used to introduce a fairly technical chapter on "polls and questionnaires" (Chap. 8). There is a nice discussion of common sampling methods and the effect of nonresponse, with useful tips and caveats for consumers of survey research. There is little in the book, however, to prepare readers for a disproportionately detailed and technical treatment of confidence intervals or the margin of error, a problem acknowledged by the authors, who encourage their readers in a footnote to "skip this section on first reading" (p. 148). It is far from clear that a second reading of the book would improve comprehension. The topic of interval estimation is continued in its short penultimate chapter on statistical inference, but it lacks the clarity and coherence of the authors' treatment of descriptive statistics. Without any theoretical or conceptual motivation, the chapter appears to be a half-hearted afterthought.

Being targeted mainly at the "non-specialist user of statistical methodology" with "technical details kept to a minimum" (front cover), the book's intermittent instruction in the use of Minitab seems strangely out of place. A future edition should take care of some uneven editing and organization of topics. Chapter 2, for example, uses concepts from probability theory (e.g., the notion of a sample space) that are not introduced until Chapter 7. The exercises in Chapter 3 use terms like "pie chart," "bar diagram," "line of best fit," and "Spearman's rank correlation coefficient" that remain undefined until Chapter 6 or later. Chapter 9, on "measuring inflation," appears completely isolated and unmotivated.

This book is definitely not a substitute for a conventional introductory statistics textbook. Like other books, such as those by Moore (1985), Agresti and Finlay (1986), and Hildebrand (1986), the book is aimed at students with little mathematical background who will most likely be mainly consumers of statistical information. Thus these books place greater emphasis on statistical ideas and applications over statistical methods. Unlike them, however, the present book lacks the breadth of statistical topics that one would expect to find in a textbook for even a nontechnical introductory course. Nevertheless, as a book on "statistical civics," it could make a nice supplement. Most chapters provide a set of useful integrative exercises and projects, with solutions to the exercises in the back.

ELKE U. WEBER
University of Chicago

REFERENCES

Agresti, A., and Finlay, B. (1986), *Statistical Methods for the Social Sciences* (2nd ed.), San Francisco: Dellen.

Hildebrand, D. K. (1986), *Statistical Thinking for Behavioral Scientists*, Boston: Duxbury Press.
Huff, D. (1954), *How to Lie With Statistics*, New York: W. W. Norton.
Moore, D. S. (1985), *Statistics: Concepts and Controversies* (2nd ed.), San Francisco: W. H. Freeman.

NEW EDITIONS

An Introduction to Statistical Methods and Data Analysis.

Lyman Ott. Boston: PWS-Kent, 1988. xiii + 945 pp.

The first edition of this work was published in 1977 and favorably reviewed in JASA (Abeles 1983). The new edition seems to retain the good features of previous ones, and reorganization seems to meet some of the objections raised by Abeles. The content of the newest edition seems to include more of the good things—more exercises, more work on residual analysis, more illustrative computer output, more emphasis on diagnosing problems with assumptions, and so forth.

REFERENCE

Abeles, F. (1983), Review of *An Introduction to Statistical Methods and Data Analysis*, by L. Ott, *Journal of the American Statistical Association*, 78, 739–740.

PUBLICATIONS RECEIVED

Collections

Cuttance, Peter, and Ecob, Russell (eds.). *Structural Modeling by Example*. Cambridge, U.K.: Cambridge University Press, 1987. x + 319 pp. $47.50.
Den Hollander, F., and Maasen, H. (eds.). *Mark Kac Seminar on Probability and Physics Syllabus 1985–1987*. Amsterdam: Centre for Mathematics and Computer Science, 1988. vi + 162 pp. Dfl. 25.30 (paperback).
Ermoliev, Y., and Wets, R. J.-B. (eds.). *Numerical Techniques for Stochastic Optimization*. Berlin: Springer-Verlag, 1988. xiii + 571 pp. $74.
Ghosh, J. K. (ed.). *Statistical Information and Likelihood*. New York: Springer-Verlag, 1988. xviii + 365 pp. $39.70 (paperback).
Harper, William L., and Skyrms, Brian (eds.). *Causation in Decision, Belief Change, and Statistics*. Dordrecht, Holland: D. Reidel, 1988. xix + 252 pp. $69.
Iri, M., and Yojima, K. (eds.). *System Modelling and Optimization*. Berlin: Springer-Verlag, 1988. x + 787 pp. $120 (paperback).
Kacprzk, Janusz, and Fedrizzi, Mario (eds.). *Combining Fuzzy Imprecision With Probabilistic Uncertainty in Decision Making*. Berlin: Springer-Verlag, 1988. ix + 399 pp. $42.90 (paperback).
Mendez, R. H., and Orszag, S. A. (eds.). *Japanese Supercomputing*. New York: Springer-Verlag, 1988. iv + 161 pp. $27.40 (paperback).
Philippou, A. N., Horadam, A. F., and Bergum, G. E. (eds.). *Applications of Fibonacci Numbers*. Dordrecht, Holland: D. Reidel, 1988. xx + 213 pp. $79.
Ricci, Gianni, and Velupillai, Kumaraswamy (eds.). *Growth Cycles and Multisectoral Economics: The Goodwin Tradition*. Berlin: Springer-Verlag, 1988. iii + 126 pp. $18.30 (paperback).
Ross, John A., and Mauldin, W. Parker (eds.). *Berelson on Population*. New York: Springer-Verlag, 1988. xii + 275 pp. $28.95.
Sargan, John Denis (Esfandiar Maasoumi, ed.). *Contributions to Econometrics* (Vols. 1 & 2). Cambridge, U.K.: Cambridge University Press, 1988. Vol. 1, viii + 304 pp., $54.50; Vol. 2, viii + 296 pp., $54.50.
Schiehlen, W., and Wedig, W. (eds.). *Analysis and Estimation of Stochastic Mechanical Systems*. New York: Springer-Verlag, 1988. vi + 350 pp. $41.10 (paperback).
Simcox, David E. (ed.). *U.S. Immigration in the 1980s: Reappraisal and Reform*. Boulder, CO: Westview Press, 1988. vii + 308 pp. $30 (paperback).
Stephens, N. M., and Thorne, M. P. (eds.). *Computers in Mathematical Research*. Oxford, U.K.: Clarendon Press, 1988. xii + 251 pp. $57.50.
Tarozzi, Gino, and Van der Merwe, Alwyn (eds.). *The Nature of Quantum Paradoxes*. Dordrecht, Holland: D. Reidel, 1988. vii + 443 pp. $115.
Visek, Jan Amos (ed.). *Information Theory, Statistical Decision Functions, Random Process* (Vol. B). Dordrecht, Holland: D. Reidel, 1988. 439 pp. $99.

Books

Abrahams, D. Mark, and Rizzardi, Fran. *BLSS: The Berkeley Interactive Statistical System*. New York: W. W. Norton, 1988. xx + 412 pp. $21.95 (paperback).
Adams, John S. *Housing America in the 1980s*. New York: Russell Sage Foundation, 1988. xviii + 328 pp. $65.
Bean, Frank D., and Tienda, Marta. *The Hispanic Population of the United States*. New York: Russell Sage Foundation, 1987. xxiv + 456 pp. $42.50.
Beltzer, A. I. *Acoustics of Solids*. Berlin: Springer-Verlag, 1988. xi + 235 pp. $59.
Bianchi, Suzanne M., and Spain, Daphne. *American Women in Transition*. New York: Russell Sage Foundation, 1986. xxii + 286 pp. $32.50 (cloth); $14.95 (paperback).
Bond, Kenneth M., and Scott, James P. *Essential Business Statistics: A Minitab Framework*. Boston: PWS-Kent, 1988. xiii + 332 pp. $42.75.
Bremaud, Pierre. *An Introduction to Probabilistic Modeling*. New York: Springer-Verlag, 1988. xii + 207 pp. $34.
Buchanan, James. *The Political Economy of the Welfare State*. Stockholm: The Industrial Institute for Economic and Social Research, 1988. 31 pp. SEK 40 (paperback).
Budinski, Kenneth G. *Surface Engineering for Wear Resistance*. Englewood Cliffs, NJ: Prentice-Hall, 1988. xii + 429 pp. $52.
Castillo, Enrique. *Extreme Value Theory in Engineering*. San Diego, CA: Academic Press, 1988. xv + 389 pp. $59.95.
Chiswick, Barry R. *Illegal Aliens*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1988. ix + 160 pp. $18.95 (cloth); $11.95 (paperback).
Chow, Yuan Shih, and Teicher, Henry. *Probability Theory* (2nd ed.). New York: Springer-Verlag, 1988. xviii + 467 pp. $42.
Dharmadhikari, Sudharkar, and Kumar, Joag-Dev. *Unimodality, Convexity, and Applications*. San Diego, CA: Academic Press, 1988. xiii + 278 pp. $64.50.
Drost, F. C. *Asymptotics for Generalized Chi-Square Goodness of Fit Tests*. Amsterdam: Centre for Mathematics and Computer Science, 1988. ix + 104 pp. Dfl. 16.50 (paperback).
Duffie, Darrell. *Security Markets, Stochastic Models*. San Diego, CA: Academic Press, 1988. xx + 358 pp. $44.95.
Fishburn, Peter C. *Nonlinear Preference and Utility Theory*. Baltimore: Johns Hopkins Press, 1988. xiv + 259 pp. $47.50.
Flury, Bernhard, and Riedwyl, Hans. *Multivariate Statistics: A Practical Approach*. London: Chapman & Hall, 1988. xi + 296 pp. $69.50 (cloth); $32.50 (paperback).
Galambos, Janos. *Advanced Probability Theory*. New York: Marcel Dekker, 1988. vii + 405 pp. $99.75 (U.S.A. and Canada); $119.50 (elsewhere).
Gudder, Stanley. *Quantum Probability*. San Diego, CA: Academic Press, 1988. xii + 316 pp. $49.50.
Healy, M. J. R. *GLIM: An Introduction*. Oxford, U.K.: Clarendon Press, 1988. ix + 130 pp. $39.95.
Hutchinson, T. P. *Road Accident Statistics*. Adelaide, South Australia: Rumsby Scientific Publishing, 1987. xxv + 292 pp. $45.