



Are the Data Worth Owning?

Author(s): J. Ross Macdonald

Source: *Science*, New Series, Vol. 176, No. 4042 (Jun. 30, 1972), p. 1377

Published by: American Association for the Advancement of Science

Stable URL: <http://www.jstor.org/stable/1734564>

Accessed: 01/05/2009 11:05

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=aaas>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science*.

<http://www.jstor.org>

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Science serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by the AAAS or the institutions with which the authors are affiliated.

Editorial Board

1972

ALFRED BROWN
JAMES F. CROW
THOMAS KUHN
ELLIOTT W. MONTELL

FRANK PRESS
FRANK W. PUTNAM
WALTER O. ROBERTS

1973

H. S. GUTOWSKY
ARTHUR D. HASLER
RUDOLF KOMPFFNER
DANIEL E. KOSHLAND, JR.

GARDNER LINDZEY
RAYMOND H. THOMPSON
EDWARD O. WILSON

Editorial Staff

Editor

PHILIP H. ABELSON

Publisher
WILLIAM BEVAN

Business Manager
HANS NUSSBAUM

Managing Editor: ROBERT V. ORMES

Assistant Editors: ELLEN E. MURPHY, JOHN E. RINGE

Assistant to the Editor: NANCY TEIMOURIAN

News and Comment: JOHN WALSH, DEBORAH SHAPLEY, ROBERT GILLETTE, NICHOLAS WADE, CONSTANCE HOLDEN, BARBARA J. CULLITON, SCHERRAINE MACK

Research News: ALLEN L. HAMMOND, WILLIAM D. METZ, THOMAS H. MAUGH II

Book Reviews: SYLVIA EBERHART, KATHERINE LIVINGSTON, KATHRYN MOUTON

Cover Editor: GRAYCE FINGER

Editorial Assistants: MARGARET ALLEN, ISABELLA BOULDIN, BLAIR BURNS, ELEANORE BUTZ, RONNA CLINE, ANNETTE DIAMANTE, MARY DORFMAN, JUDITH GIVELBER, MARLENE GLASER, CORRINE HARRIS, OLIVER HEATWOLE, CHRISTINE KARLIK, MARSHALL KATHAN, MARGARET LLOYD, JANE MINOR, DANIEL RABOVSKY, PATRICIA ROWE, LEAH RYAN, LOIS SCHMITT, YA LI SWIGART, ALICE THEILE

Guide to Scientific Instruments: RICHARD SOMMER

Membership Recruitment: LEONARD WRAY; **Subscriptions:** BETTE SEEMUND; **Addressing:** THOMAS BAZAN

Advertising Staff

Director
EARL J. SCHERAGO

Production Manager
BONNIE SEMEL

Advertising Sales Manager: RICHARD L. CHARLES

Sales: NEW YORK, N.Y. 10036: Herbert L. Burkland, 11 W. 42 St. (212-PE-6-1858); SCOTCH PLAINS, N.J. 07076: C. Richard Callis, 12 Unami Lane (201-889-4873); CHICAGO, ILL. 60611: John P. Cahill, Room 2107, 919 N. Michigan Ave. (312-DE-7-4973); BEVERLY HILLS, CALIF. 90211: Winn Nance, 111 N. La Cienega Blvd. (213-657-2772)

EDITORIAL CORRESPONDENCE: 1515 Massachusetts Ave., NW, Washington, D.C. 20005. Phones: (Area code 202) Central office: 467-4350; Book Reviews: 467-4367; Business Office: 467-4411; Circulation: 467-4417; Guide to Scientific Instruments: 467-4480; News and Comment: 467-4430; Reprints and Permissions: 467-4483; Research News: 467-4321, Reviewing: 467-4440. Cable: Advancesci, Washington. Copies of "Instructions for Contributors" can be obtained from the editorial office. See also page xv, *Science*, 24 December 1971. **ADVERTISING CORRESPONDENCE:** Room 1740, 11 W. 42 St., New York, N.Y. 10036. Phone: 212-PE-6-1858.

Are the Data Worth Owning?

Amitai Etzioni has recently (14 April 1972, page 121) raised the question of who should ultimately own the data. He points out that, since data is (sic) often lost or becomes inaccessible, agencies that finance data collection or preparation should require that it be made available to others by deposition in a data bank or library.

A more fundamental question than who should own the data is, Are the data even worth owning? Unfortunately, the answer is usually an embarrassing and costly "No" across the entire spectrum of research. The problem usually lies in lack of knowledge about the trustworthiness of the data. Measures of uncertainty are usually not given at all; even when they are, they are themselves generally untrustworthy. Lancelot Hogben has stated that "less than one percent of research workers clearly apprehend the rationale of statistical techniques they commonly invoke."

Further, David Lide, head of the National Standard Reference Data System of the National Bureau of Standards, estimates that from 50 to over 90 percent of the published raw data available for producing trustworthy, evaluated results for the physical properties of scientific materials cannot, in fact, be used for this important purpose. A good illustration of the basic difficulty has been given by the late W. J. Youden of NBS. He states that, of 15 observations of the mean distance to the sun published from 1895 to 1961, each worker's estimated value is outside the uncertainty limits set by his immediate predecessor.

Both systematic and random errors occur in all experimental situations. They should be estimated, discussed, and cited separately, as Churchill Eisenhart has pointed out. Ideally, systematic errors should be estimated by independently measuring the quantity in question with a different apparatus, preferably one that operates on a different principle from that of the original apparatus. One should strive to make the estimated maximum systematic error comparable to or smaller than the estimated root-mean-square random error of the experiment. When it is impractical to obtain independent estimates of the systematic error, a good rule of thumb is to multiply one's best estimate of it by a factor of 3.

Even when estimates of individual errors (deviations) are calculated, it is unusual for the experimenter to check the deviations for stochastic independence and to state the result of such a check. Rarely indeed does one find mention of the statistical distribution that the deviations appear to follow. Without such knowledge, however, one cannot assess the meaningfulness of such important derived quantities as ordinary confidence limits and standard deviations. When individual measurements are to be analyzed by such techniques as least squares, one seldom finds experimenters replicating the individual measurements closely enough to obtain trustworthy estimates of uncertainties for use in weighted least squares. Yet only thus can one verify the assumptions implicit in even unweighted least squares. Nonlinear least squares analysis is becoming much more common these days, but one never finds it shown that the bias in parameter estimates introduced by this estimation technique is safely smaller than the sampling error. Finally, there are almost always random errors present in the values of all variables measured (except in whole number cases), not just in the "dependent" variable, as is usually assumed in ordinary least squares analysis. Although a generalized least squares technique is necessary and available to handle such situations, it is hardly ever used—nor is the need for it usually recognized.

Clearly, much further education in data analysis, presentation of results, and the need to call in a statistician is necessary before a high proportion of published data can be properly used for more than qualitative purposes.—J. ROSS MACDONALD, *Chairman, Numerical Data Advisory Board, National Research Council.*