SWITCHING REGRESSIONS
TOBIT MODEL)

D. C. GILLILAND
J. HANNAN

FALLACIES, STATISTICAL

The rapid growth in the development and application of statistical methodology in this century has been accompanied by a corresponding increase in fallacious statistical reasoning and misuses of statistics. The potential for statistical fallacies has been enhanced by the development of statistical computer packages (creating greater access to complex procedures) and the increasing need for statistical analyses in government agencies, industries, and diverse academic disciplines. The fallacies included in this article were chosen because of their frequent occurrence and because they merit mention in introductory courses in statistical methods. We have not attempted to provide an exhaustive catalog. An indexing of types of fallacies and a discussion of a greater variety of them are given by Good [12, 13]. Other sources with several good examples of statistical fallacies and misuses of statistics include Campbell [8], Cohen [10], Freedman et al. [11], Huff [14, 15], Moran [18], and Wallis and Roberts [23].

We first describe three major types of errors, each of which seems to be responsible for a variety of fallacious arguments, and then we describe briefly several other fallacies which occur in other settings.

FAILURE TO INCLUDE RELEVANT COVARIATES

Many bivariate associations disappear or diminish when appropriate control variables are introduced. Failure to recognize this fact often leads to wrong conclusions about relationships between variables.

For categorical data*, it is misleading to restrict attention to a two-dimensional contingency table when that table is more properly viewed as a marginal distribution of a higher-dimensional array. Bickel et al. [3] illustrate this in their discussion of data concerning admission into graduate school at Berkeley. Investigation of the 2×2 table of admissions decision (admit, do not admit) by sex revealed that relatively fewer women than men were admitted into the graduate school. When admissions were considered separately by academic department, however, the apparent bias against women disappeared. The explanation for the shift in results is that the proportion of women applicants tended to be highest in disciplines that were most competitive for admissions in having the highest rejection rates. Bishop et al. [4, pp. 41–42] present a similar example. An apparent association between amount of prenatal care and infant survival disappears when the data are considered separately for each clinic participating in the study. A pooling of the data from the clinics ignores the dependence of both infant survival and amount of prenatal care on clinic.

The fact that partial associations may be very different in nature from unconditional bivariate associations has been expressed through conditional probabilities as *Simpson's paradox* (see Simpson [20] and Blyth [5, 7]). This paradox states that even if $\Pr(A \mid BC) > \Pr(A \mid B^cC)$ and $\Pr(A \mid BC^c) > \Pr(A \mid B^cC^c)$, it is possible that $\Pr(A \mid B) < \Pr(A \mid B^c)$. In the context of a $2 \times 2 \times 2$ cross-classification of three dichotomous variables X, Y, and Z, Simpson's paradox implies that it is possible to have a positive partial association between X and Y at each level of Z, yet a negative unconditional association between X and Y, due to the nature of the association of Z with both X and Y. For a numerical example, see Blyth [5, p. 264]. Bishop et al. [4, p. 39] give conditions under which partial associations are the same as unconditional associations, so that multidimensional contingency tables* can be meaningfully collapsed.

The same remarks regarding covariates apply to quantitative variables. For example, the mean salary for men may exceed the mean salary for women for the faculty at a particular university. When factors such as department, rank, and number of years in rank are controlled, however, it is possible that the difference in the means may change appreciably. In a related example, Cochran [9] discusses how failure to use age standardization in comparing two populations can lead to fallacious conclusions.

FAILURE TO ADJUST THE ERROR RATE FOR MULTIPLE INFERENCES

The importance of using multiple comparison procedures for making pairwise inferences about several means is emphasized in many statistical methods textbooks (e.g., Snedecor and Cochran [21, p. 272]; *see also* MULTIPLE COMPARISONS). The authors of these books note that the use of a standard error rate (such as 0.05) for each of a large number of inferences may result in an unacceptably large probability of at least one error occuring (e.g., at least one type I error or at least one confidence interval not containing the parameter it is designed to enclose). Fallacious arguments can easily occur in many other contexts from using a single-inference error rate when several inferences have actually been conducted.

This type of error frequently occurs when the need for a multiple-inference approach is not obvious. For example, a researcher analyzing a large data set on several variables may screen it, using computer packages to compute correlations, chi-squares, analyses of variance, regression analyses, etc., on various combinations of the variables. In some cases several competing tests may be conducted to test the same hypothesis. The researcher may select from the computer printout everything achieving significance at the 0.05 level and report those results as if the corresponding hypotheses and analyses were the only ones considered.

A more subtle failure to adjust for multiplicity of inferences results from the tendency of research journals in many fields to publish only those studies that obtain statistical significance at a certain level. If a large number of researchers independently test the same true null hypothesis, there is a good chance that a type I error will be published. Researchers who do not obtain significant results may be discouraged from submitting their findings or feel pressured to find ways of achieving significance (e.g., other tests, more data). Walster and Cleary [24] give a good discussion of this problem. They also emphasize the importance of replication of previously published research so that type I errors are exposed.

Fallacious arguments of a similar nature can occur from treating the maximum or minimum of a set of random variables as if it had the same distribution as an arbitrary one of the variables. This error occurs when a researcher selects the two most distant sample means (out of a collection of several means) as the most interesting finding of a survey and compares those means using standard two-sample tests or confidence intervals. The error also occurs in variable selection for a regression model when at each stage, one tests the significance of the partial effect of X_i on Y *after* having selected X_i because it had the largest such effect out of some set of variables. Many events that seem to be very unusual occurrences when viewed in isolation may seem rather common when considered in proper context. Suppose that a coin shows heads in each of 10 tosses. We would probably suspect that the coin is unbalanced in favor of heads, since the probability of such a rare event if the coin were balanced is only $(\frac{1}{2})^{10} = 0.00098$. However, if we were told that this coin had had the greatest number of heads out of 1000 coins that had been tossed 10 times each, we would be less likely to believe that it was biased (see Wallis and Roberts [23, p. 116]).

Another misuse that occurs when multiple inferences are not recognized as such is the application of fixed sample-size methods after obtaining each new observation in a sequential sampling scheme (see Moran [18] and Armitage et al. [1]). The result we predict may occur if we wait long enough.

CONFUSION OF CORRELATION WITH CAUSATION

Fallacious arguments often result from the belief that correlation* implies causation*. The everyday usage of the word "correlation" in the English language probably contributes to the confusion. The fallacy can often be shown by illustrating the lack of partial association when certain control variables are introduced, as in the sex–graduate admissions study by Bickel et al. [3]. Yule and Kendall [26, Chaps. 4, 15, 16] discuss the problems in detail.

A special case of the correlation–causation fallacy is the "post hoc" fallacy that if A precedes B, it must be a cause of B. Campbell [8, p. 172] illustrates the fallacy by reference to the plane traveler who requests that the captain not turn on the "fasten seat belts" sign, since it always seems to result in a bumpy ride. For other examples of the post hoc fallacy, see Huff [14, Chap. 8].

Other types of misapplications of correlation coefficients abound. Barnard [2] explains why "astonishingly high correlations" need not be especially noteworthy. For example, a correlation arbitrarily close to 1 may be produced by a single outlying observation. Another common error occurs in the generalization of inferences to a different sampling unit. For example, "ecological correlations" computed from rates or totals for units such as counties or states may be very different in magnitude from correlations obtained using data on individuals. Freedman et al. [11, pp. 141–142] state that the correlation between average income and average education computed for nine regions in the United States is about 0.7, whereas it is approximately 0.4 when computed for individuals from census data.

REGRESSION FALLACY

The phenomenon of regression* toward the mean for the bivariate normal regression* model was first noticed by Sir Francis Galton* in his studies of X = father's height and Y = son's height. Based on this phenomenon he made the fallacious conclusion that the variability in heights must decrease with time. A counterexample is given by Good [13], who notes that a reversal of the labeling of Y and X would necessarily force the conclusion that variability in heights is increasing with time. The fact that "highly unusual" observations tend to be followed by more regular observations, when not recognized or not understood, has resulted in various types of fallacies and superstitions (e.g., that a professional athlete having an outstanding first year will have a "sophomore jinx"). For further discussion, see Freedman et al. [11, pp. 158–159].

NEGLECTING ASSUMPTIONS

Misuses of statistical procedures commonly occur from severe violations of basic assumptions concerning method of sampling, required sample size (for use of asymptotically based formulas), measurement scale of variables, and distribution of variables. For example, fallacious conclusions could result from using formulas based on simple random sampling for data collected as a cluster sample, from treating time-series data as an independent identically distributed sequence, from applying the chi-square test* of independence to a contingency table having a small total frequency or having ordered rows or columns, and from blindly applying techniques such as regression, analysis of variance*, and factor analysis* to dichotomous variables. A sociology Master's thesis is rumored to exist which contains about 100 applications of the Kolmogorov–Smirnov* two-sample test, *none* of which attains significance at the 0.05 level. The author apparently applied the standard form of the test designed for continuous variables to highly discrete data, for which that test is highly conservative.

NEGLECTING VARIATION

Fallacies or misleading statements often occur from a failure to consider variation*.

Examples include reporting a percentage without listing its base sample size (and hence its standard error) and attributing importance to a difference between two means which could be explained by sampling error.

MISINTERPRETATION OF STATISTICAL TESTS OF HYPOTHESES

The results of statistical tests are often misinterpreted due to factors such as confusion of statistical significance with practical significance and acceptance of the null hypothesis without consideration of the power function. For good discussions, see Kruskal [17, p. 456] and Kish [16].

FALLACIOUS PROBABILITY REASONING

Many advances in the historical development of probability occurred because of fallacious arguments in gambling situations which led to seemingly contradictory results in the pocketbook (see Freedman et al. [11, pp. 223–225], Huff [15, pp. 63–69], and Todhunter's [22, Chap. XIII] discussion about D'Alembert's fallacies). Among the most common errors are the following: treating sample points as equally likely when they are not; misinterpreting the law of large numbers, as when arguing that in 1,000,000 flips of a fair coin, the number of heads is bound to be within a few units of 500,000; not understanding independent trials or conditional probability, as in the argument that a sequence of 10 consecutive heads in coin flipping is almost sure to be followed by a tail; misuse of the additive law, as in the argument that for n independent trials with probability p of success on each, the probability is np of at least one success. Many fallacious probabilistic arguments result from an unawareness of certain paradoxes. For example, it is tempting to argue that $P(Y > X) > \frac{1}{2}$ and $P(Z > Y) > \frac{1}{2}$ implies that $P(Z > X) > \frac{1}{2}$. The transitivity paradox shows that this need not be the case even if X, Y, and Z are independent (see Blyth [6]).

FALLACIES WITH TIME-SERIES DATA

Failure to recognize the special problems occurring in the analysis of time-series data often leads to fallacious conclusions. For example, strong spurious associations may result from correlating variables which are measured over time and have similar trends (see Huff [14, p. 97]). Errors in statistical analysis often result from applying formulas based on independent observations, such as by using standard regression procedures and disregarding effects of serially correlated error terms (see Wonnacott and Wonnacott [25, pp. 136–147]). Richman and Richman [19] show how fallacious conclusions concerning changes in the level of heroin addiction result from improper analyses of time-series data.

IMPROPER BASE

A common statistical error in the news media is the comparison of frequencies based on different totals. Examples include the statement during the Vietnam war that it was safer to be in the army than driving on the nation's highways due to the lower yearly death total in the war (rates of death should be compared, preferably within age groups).

MAKING AN INFERENCE WITHOUT THE NECESSARY COMPARISON

This error commonly appears through the reporting of only one row of a contingency table. For example, a criminal rehabilitation program might be criticized because participants in it have a recidivism rate of 50%. Without being given the corresponding rate for nonparticipants or for other programs, we would have a difficult time making a judgment.

BIASED DATA THROUGH MEASUREMENT ERROR OR INTERVIEWER EFFECT

Wallis and Roberts [23, p. 96] quote a survey in which the percentage of blacks interviewed who felt the army to be unfair to their race was 35% for those people having a black interviewer and 11% for those having a white interviewer (see also Huff [14, p. 24]).

BIASED DATA DUE TO IMPROPER SAMPLING FRAME

The average number of children in families having students in a particular school would tend to be overestimated by sampling children from that school, since a large family is more likely to be represented in the sample than a small family.

UNCRITICAL RELIANCE ON COMPUTERS

Among the likely consequences of the development of computer-based statistical packages have been a greater relative frequency in the use of statistical procedures that are inappropriate to a problem or which the researcher fails to understand, errors associated with unrecognized multiple inferences due to searching for significant results, and the attributation of greater accuracy to the results than the data warrant (*see* COMPUTERS AND STATISTICS).

References

[1] Armitage, P., McPherson, C. K., and Rowe, B. C. (1969). *J. R. Statist. Soc. A*, **132**, 235–244.

[2] Barnard, G. A. (1977). *J. R. Statist. Soc. A*., **140**, 200–202.

[3] Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975). *Science*, **187**, 398–404.

[4] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis*. MIT Press, Cambridge, Mass.

[5] Blyth, C. R. (1972). *J. Amer. Statist. Ass*., **67**, 364–366.

[6] Blyth, C. R. (1972). *J. Amer. Statist. Ass*., **67**, 366–373.

[7] Blyth, C. R. (1973). *J. Amer. Statist. Ass*., **68**, 746.

[8] Campbell, S. K. (1974). *Flaws and Fallacies in Statistical Thinking*. Prentice-Hall, Englewood Cliffs, N.J. (Good source of elementary examples; similar in level to Huff's books.)

[9] Cochran, W. G. (1968). *Biometrics*, **24**, 295–313.

[10] Cohen, J. B. (1938). *J. Amer. Statist. Ass*., **33**, 657–674.

[11] Freedman, D., Pisani, R., and Purves, R. (1978). *Statistics*. W. W. Norton, New York. (Excellent introductory statistics text which highlights several statistical fallacies.)

[12] Good, I. J. (1962). *Technometrics*, **4**, 125–132.

[13] Good, I. J. (1978). Fallacies, statistical. In *International Encyclopedia of Statistics*, Vol. 1, W. H. Kruskal and J. M. Tanur, eds. Free Press, New York, p. 344. (Contains large bibliography of literature on statistical fallacies.)

[14] Huff, D. (1954). *How to Lie with Statistics*. Norton, New York. (Contains discussions of several basic types of statistical fallacies; written for the layman.)

[15] Huff, D. (1959). *How to Take a Chance*. Norton, New York. (Good examples of fallacious probabilistic arguments.)

[16] Kish, L. (1959). *Amer. Sociol. Rev*., **24**, 328–338.

[17] Kruskal, W. H. (1978). Significance, tests of. In *International Encyclopedia of Statistics*, Vol. 2, W. H. Kruskal and J. M. Tanur, eds., 944–958. Free Press, New York.

[18] Moran, P. A. P. (1973). *Commun. Statist*., **2**, 245–257. (Discusses 15 settings in which fallacious statistical analyses commonly occur.)

[19] Richman, A. and Richman, V. (1975). *Amer. Statist. Ass., 1975 Proc. Social. Statist. Sec.*, 611–616.

[20] Simpson, E. H. (1951). *J. R. Statist. Soc. B*, **13**, 238–241.

[21] Snedecor, G. W. and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. Iowa State University Press, Ames, Iowa.

[22] Todhunter, I. (1865). *A History of the Mathematical Theory of Probability*. Macmillan, London.

[23] Wallis, W. A. and Roberts, H. V. (1962). *The Nature of Statistics*. Free Press, New York.

[24] Walster, G. W. and Cleary, T. A. (1970). *Amer. Statist*., **24**, 16–19.

[25] Wonnacott, R. J. and Wonnacott, T. H. (1970). *Econometrics*. Wiley, New York.

[26] Yule, G. U. and Kendall, M. G. (1937). *An Introduction to the Theory of Statistics*. Charles Griffin, London.

(CAUSATION
LOGIC IN STATISTICAL REASONING)

ALAN AGRESTI

FARLIE–GUMBEL–MORGENSTERN DISTRIBUTIONS

The Farlie–Gumbel–Morgenstern (FGM) system of bivariate distributions includes all