C. Tolmie

REFIGURING ANTHROPOLOGY

First Principles Of Probability & Statistics

David Hurst Thomas
American Museum of Natural History

Waveland Press, Inc.
Prospect Heights, Illinois

For information about this book, write or call:

Waveland Press, Inc.
P.O. Box 400
Prospect Heights, Illinois 60070
(312) 634-0081

For permission to use copyrighted material, the author is indebted to the following:

FIG. 2.1. By permission of the Trustees of the British Museum (Natural History).
TABLE 2.3. (p. 24) *From Physical Anthropology: An Introduction* by A. J. Kelso. Reprinted by permission of the publisher, J. B. Lippincott Company. Copyright © 1974. (p. 25) Reproduced by permission of the Society for American Archaeology from *Memoirs of the Society for American Archaeology*, Vol. 11, 1956.
FIG. 3.1. From Hulse, Frederick S. *The Human Species: An Introduction to Physical Anthropology*. Copyright © 1963 by Random House, Inc.
FIG. 3.2. From Dozier, Edward P., *The Pueblo Indians of North America*. Copyright © 1970 by Holt, Rinehart and Winston, Inc. Reproduced by permission of Holt, Rinehart and Winston. [This book reissued 1983 by Waveland Press, Inc.]
FIG. 3.4. Reproduced by permission of the Society for American Archaeology from *American Antiquity*, Vol. 35 (4), 1970.
FIG. 3.5. Reproduced by permission of the American Anthropological Association from the *American Anthropologist*, Vol. 73 (3), 1971.
FIG. 13.14. From *Biometry* by Robert R. Sokal and F. James Rohlf. W. H. Freeman and Company. Copyright © 1969.

Second Printing
The 1976 version of this book was entitled *Figuring Anthropology*.

ISBN 0-88133-223-2

Printed in the United States of America.

16 A Parting Word of Caution

- *Ritual or ceremonial is a fixed set of solemn observances. It need not be tied up with religion but develops whenever behavior is taken seriously*—R. Lowie

16.1 STATISTICAL PRACTICES AS RITUAL BEHAVIOR

Some skeptics suggest that the application of statistical methods by social scientists amounts to little more than conventionalized ritual behavior. Taken in its proper anthropological perspective, this statement does indeed contain a kernel of truth.

Item: Emile Durkheim viewed ritual as "rules of conduct" prescribing appropriate behavior in the presence of "sacred objects."

Item: Ecclesiastical rituals generally employ a professional clergy or priesthood, often under the control of the central temple. Members of this select priesthood have acquired intricate ritual knowledge which is dispensed to the proletariat only under tightly specified circumstances.

Item: The magician believes that he *controls* supernatural power under restricted ritualized conditions. That is, he is confident that he possesses a tested formula. If he carries out the ritual exactly as prescribed by the formula, he will obtain the guaranteed results. The supernatural power has no volition of its own. It must respond because the enlightened magician has power over power.

Item: A. R. Radcliffe-Browne suggested that the performance of significant rituals, such as the Trobriand Islanders' *kula* ring, raises sentiments in the actors which are beneficial to the society in general. These "rites of solidarity" tend to enhance the sense of group identity by coordinating the action of individual members for group benefit.

Some limited parallels between ritual behavior and statistical behavior seem unavoidable. In fact, the very terms "ritual" and "rite" derive directly from the Latin *ritus* and recall the Greek *arithmos* meaning "number." The common phrase "do it by the numbers" further links the concepts of rite and number. The six steps of hypothesis testing introduced in Chapter 9 appear distressingly similar to Durkheim's "rules of conduct," and anybody who has ever witnessed an orientation tour to a local computer center should have little difficulty recognizing the "sacred objects" of statistics. Furthermore, most computer centers, and even some Anthropology Departments, employ "professional statistical consultants" who dispense information about the steps which must be followed if the statistical formulas are to respond in the guaranteed fashion. A "cult of the computer" has even arisen recently within anthropology, complete with sacred texts, omnipotent deities, and a host of ritual paraphernalia. But like the Trobriand *kula* ring, this cult has ramifications far transcending the immediate details of ritual. The current statistical awareness in the social sciences has fostered a far-reaching reexamination of the nature of scientific explanation, the quality of acceptable evidence, and the degree of bias introduced during anthropological fieldwork.

The point here is that statistical procedures can indeed function as mere ritual if we so desire. We could easily develop a false sense of security toward our conclusions and a static attitude of self-righteousness, content merely to utter prescribed phrases at prescribed times in prescribed places. The following guidelines are designed to keep anthropological statistics from slipping into the mystical realm of myth and ritual.

● *ART, SCIENCE—you seem to have paid a fairly high price for your happiness said the Savage.*—A. Huxley

16.2 THE TEN COMMANDMENTS OF STATISTICS

Some years ago, a discontented tribe of Social Scientists known as the Anthropologists left their homeland on the bleak Plateau of Impressionistic Groping, settling first here, then there, yet never finding true contentment. They tarried for a time amidst the Temples of Anthropometry. But they soon moved on, dissatisfied with weaving tapestries from meaningless cranial measurements. Their journeys were fraught with hazards: First they tottered above the precipice of Racial Determinism and Social Darwinism; then they navigated through the circular eddies of nineteenth-century Cultural Evolutionism, only to be trapped for decades in the box canyons of Historical Particularism.

But their wandering at last appeared to end when they came upon the new land—the lush Valley of Quantification. The Valley appeared to be a land of plenty: plenty of numbers, plenty of formulas, and plenty of computers. The Anthropologists surged forth brandishing their computer programs, their Munsell Color Charts, their random digit tables, and all the other implements necessary to exploit this rich new environment. Then amidst their revelries, they heard a voice. At first a murmur, the sound grew louder and louder, until they recognized the voice of The Science: "Ye must proceed with care. I can offer potent weapons to smite the antagonistic hordes of Small Samples and Probabilistic Uncertainty, but understand and accept the adherent responsibilities. Learn and keep the Covenant of Statistics. If ye steadfastly keep this Covenant, then ye shall be a special treasure unto me, for only you the Anthropologists, of all the tribes of Social Science, do indeed embrace all Peoples of the Earth. But if ye choose to disobey the Commandments of the Covenant and rape the resources of the valley of Quantification, I shall surely banish ye every one to the vapid hinterlands of pseudo-Science. Ye shall be derided and ridiculed by the Tribes of Biometrists, Econometrists, Psychometrists, and—yes indeed—even your close neighbors, the Sociologists."

And it came to pass on the third day, that there were thunders and lightnings. A dense cloud arose surrounding the Machine of Computing, and all the Anthropologists trembled.

And when the voice of the trumpet sounded long, and waxed louder and louder, the Science belched twice from deep within the 64K memory of the Machine of Computing, and the CalComp plotter slowly scribed the following lines:

> ● *If you fall in love with a machine there is something wrong with your lovelife. If you worship a machine there is something wrong with your religion.*—L. Mumford

COMMANDMENT I. Thou shalt not worship the 0.05 level.

Few cows seem more sacred in the pantheon of statistics than the conventional 0.05, 0.01, and 0.001 levels of statistical significance. Of these, surely the 0.05 level is the most revered. The popularity of these particular levels results no doubt from simple convenience and tradition, but anthropologists have come to rely heavily upon them to the virtual exclusion of other levels of probability. The question must arise as to whether most people really understand what levels of significance are all about. The alpha level deals only with one's willingness to commit a Type I error (that is, to reject a true H_0). The smaller the alpha, the larger becomes the probability of committing a Type II error (unless the sample size is increased). It is a sin to read more into the significance level than this.

Alpha should be chosen, whenever possible, after the sample size has been established because significant results are known to emerge more readily in large samples. Meehl (1967) once tested 55,000 Minnesota high school students for such diverse factors as sex, order of birth, religion, club participation, hobbies, and so on. Significant relationships emerged in fully 90 percent of all cases. "Meehl's paradox" illustrates that the research hypothesis is bound to be confirmed in very large samples in over half the cases by chance alone, whether or not the hypothesis is actually true. Pelto (1970) and Benfer (1968) have also urged caution in evaluating statistical results from large anthropological samples. A rigorous level of significance (say, 0.01 or less) should generally accompany such large samples, while a greater rate of Type I error ($\alpha = 0.05$ or even larger) is often permissible when n is small.

The best advice regarding the levels of statistical significance is simply to use common sense. Weigh the consequences of committing both Type I and Type II errors. Would one error exact a higher price than the other? Search for ancillary sources of information—previous studies, existing theory, similar conditions in allied disciplines—to provide clues (see for example, Labovitz 1968; Skipper, Guenther, and Nass 1967). But above all, do not select the 0.05 level as a comfortable ritual.

● *This teacher went into her classroom about fifteen minutes before the class was supposed to begin work and caught a bunch of her boys down in a huddle on their knees in the corner of the room. She demanded of them what they were doing, and one of them hollered back and said, "We were shooting craps." She said, "That's all right, I was afraid you were praying."*—S. Ervin

COMMANDMENT II. Thou shalt not infer a causal relationship from statistical significance.

Science attempts to isolate "universals" or "laws" from the seemingly incoherent tangle of reality; the successful quest for laws ultimately leads to clarification of specific causes known to produce predictable effects. The biological and physical sciences generally design laboratory or field experiments in which all causal factors save one are held constant; in this manner, the complex causal nexus is reduced to manageable proportions. But anthropology is rarely blessed with tightly designed experiments employing built-in controls. Anthropologists must accept circumstances largely as they exist in the real world, usually without the luxury of holding selected factors constant. Because social phenomena generally involve multiple causality, anthropologists often must rely upon analysis of extant associations. The judicious use of statistical techniques can be of great help in sorting out the patterned cultural responses from chance occurrences, but anthropologists must explicitly recognize that some phenomena are better explained by statistics than others. Sometimes statistics don't fit.

Anthropologists employ two classes of statistics: (1) *descriptive statistics* characterize central tendency, dispersion, correlation, and regression; and (2) *inferential statistics* infer unknown population parameters from incomplete sample data.

Descriptive generalizations of social phenomena are commonplace in anthropology.

Ninety percent of societies with Crow/Omaha kin terms have some form of unilineal descent.

Horticulturalists have a significantly higher degree of matrilocal residence than do hunter-gatherers.

The prehistoric Basketmaker II people hunted with the *atlatl* rather than bow and arrow.

The potlatch of the prehistoric Kwakiutl closely resembled the redistributive feasts of New Guinea and Melanesia.

Anthropology is also committed to unravelling relationships at a higher level of analysis. If *X*, then *Y*.

Unilineal descent *causes* Crow/Omaha kinship terminology.
Increased population pressure *caused* Mesolithic peoples to turn to an alternative subsistence base.
The principal *cause* of warfare among slash-and-burn horticulturalists is the limited availability of secondary forest growth.

Three important steps are necessary to establish the validity of such causal statements: (1) identify the important factors; (2) determine the predictor-predicted relationship among these factors; (3) establish and test the causal mechanism. Statistical methods greatly facilitate the first two steps, but difficulties are encountered on the third level. Associational statements (step 2) illustrate only that a significant relationship exists between variables *X* and *Y*. *Nothing in the framework of statistical inference justifies the automatic assertion that Y is causally related to X.* Statistics only establish that *X* serves as a predictor of *Y*. The causal arrow must be supplied by considerations extending beyond commonplace statements of statistical significance and association. Marvin Harris (1968: 621) has underscored this warning in *The Rise of Anthropological Theory:*

> *The difference between causal factors and mere predictive ones is not to be taken lightly. It is the difference in knowing whether wounds cause gunshots or gunshots cause wounds. Bullet holes are excellent predictors of gunshots. As all devotees of Hercule Poirot know, there is a high correlation between gunshots and bullet holes, but no murder has yet admitted of the possibility that it was the fatal wound which caused the gun to discharge its contents. Along the same lines, one might note how reliably rain predicts clouds, or how frequently fire engines are found near burning buildings. If any additional examples are needed, they are available to anyone who has a movie projector which can be run in reverse.*

This is not, of course, to assert that deft anthropologists should avoid causal inference. Quite to the contrary. Statistical methods can even be appropriate to such studies, particularly the judicious use of advanced correlation and regression techniques as discussed by Blalock (1964) and Boudon (1965). But causal statements never directly follow from statistical association.

A comprehensive discussion of causal analysis of cultural phenomena is beyond the present scope. The interested reader is referred to Pelto (1970: chapter 9), Harris (1968: chapter 21), and Köbben (1970). Other discussions can be found in Cohen and Naroll (1970: 5); Murdock (1949: chapter 7); Watson, LeBlanc, and Redman (1971: 140–150); and Clarke (1968: chapter 12).

COMMANDMENT III. Thou shalt not confuse statistical significance with substantive significance.

The initial step in hypothesis testing translates a research hypothesis into operational statistical statements. The next four steps are directed toward testing these statistical hypotheses: H_0 either survives the test or is rejected. The final step translates the numerical findings back into substantive, anthropologically relevant statements. The translation and subsequent decoding from the numerical jargon of statistics to the everyday discourse of anthropology creates a potential chasm of confusion for the unwary.

A bridge between the substantive and the statistical can be seen in the familiar "If *A*, then *B*" perspective of the logician. *B* is a logical consequent of *A*. Let us call theory *A* the anthropological hypothesis and *B* the logical consequent having certain statistical implications. The statement that the Mousterian culture was practiced by European Neanderthals is a plausible theory. Certain consequences must logically follow if this theory is correct. We suspect, for instance, that only Neanderthal skeletons should appear in Mousterian contexts, and Neanderthals should appear *only* in Mousterian sites. *If* the Mousterian culture was practiced only by Neanderthals, *then* only Neanderthal skeletons should appear in Mousterian sites: If *A*, then *B*. The *B* proposition is a substantive proposal easily translated into a statistical contingency table.

	Neanderthal Skeletons	
Mousterian artifacts	+	−
+	*a*	*b*
−	*c*	*d*

The operational statistical hypotheses which follow are:

$$H_0: \quad ad \leq bc \qquad H_1: \quad ad > bc$$

A decision can be made regarding the tenability of the null hypothesis by tabulating inventories of major Middle Paleolithic sites. Suppose that the evidence weighs strongly against H_0—cells *a* and *d* are heavily loaded—and these results are declared significant at 0.01. The statistical decision rejects the null hypothesis in favor of H_1. The substantive conclusion is that Neanderthal skeletons are indeed associated with artifacts of the Mousterian culture.

Several cautions must be entered. First of all, the level of statistical probability applies only to the statistical conclusion, *not to the substantive decision*. The testing procedure *does not* warrant the statement that "there's less than a 0.01 chance that Neanderthals were not Mousterians." This conclusion grossly confuses the statistical with the substantive (see also Commandment VIII).

Second, an appropriate measure of association such as phi or gamma should accompany the level of significance. It is an error to weigh the level of significance more heavily than the actual size of a difference or the magnitude of association. Anthropological literature often considers only the *p* value as a "measure of association" without noting the actual size of the observed difference. The general goal of high predictability in social science is a laudable one, but this goal must not be confused with a high level of statistical significance. A 1 percent difference will prove statistically significant with a large enough sample, but in a practical sense such small differences are meaningless in modern social science (Selvin 1957).

Finally, one must avoid committing the *fallacy of affirming the consequent* (Blalock 1972: chapter 8). Proposition *B* appears to be correct in our example because Neanderthal skulls generally co-occur with Mousterian artifacts. But we cannot conclude from this finding that *A must* also be true, that Neanderthals in fact made those tools or that Neanderthals necessarily carried the Mousterian cognitive set about in their heads. We can only be certain that *A* is true by absolutely establishing that *no other statements* could predict *B*, such as: If *C*, then *B*, or if *D*, then *B*. This task is generally impossible because there seem to always be alternative explanations for the *B* phenomena. If Neanderthal were a relic collector, for example, then the Mousterian artifacts associated with Neanderthal burials could be heirlooms or antiques rather than the actual cultural remains of Neanderthals. Or perhaps Neanderthal's European successor, Cro-Magnon, was the pothunter. The Mousterian sites could conceivably represent prehistoric "museums," set up by a history-conscious Cro-Magnon. You can probably think of equally *possible*, although implausible, explanations for the fact that Mousterian tools and Neanderthal bones co-occur at archaeological sites. Because this is so, one can never prove that theory *A* must be true. In fact, infinitely many theories exist to explain any set of facts, as long as one is willing to discard enough other theories (see Kemeny 1959: chapter 5 on this point).

The term *significance test* can be a culprit creating problems with statistical results. As mentioned earlier, Naroll urges us to call our statistical tools *insignificance tests*, emphasizing that only irrelevant relationships can be established with statistical authority. Kish (1959: 139) suggests scrapping the term "significance" altogether and proposes that we speak of "tests against the null hypothesis." Either H_o survives these tests or not. Hays (1973: 384) regards the value of statistical results in terms of their "surprisal value." When the results appear likely under H_o, then their "surprisal value" is quite low. But the surprisal value is high when results appear unlikely under H_o, and we can direct actions accordingly. Keep in mind that many cases exist when statistical tests are inappropriate in the first place. Some results should simply be presented with appropriate confidence intervals, and the matter of "significant or not" be left to the reader.

This brief discussion considers only some of the severe problems arising when statistical significance is confused with substantive importance. Suitably alerted, the reader is urged to consider the growing body of literature on the misuse of statistical machinery in the social sciences. Some of the more relevant articles have been compiled by Morrison and Henkel (1970) and Steger (1971); see also Naroll's (1971) review of the problem.

● *If fifty million people say a foolish thing, it is still a foolish thing.*—B. Russell

COMMANDMENT IV. Thou shalt not confuse statistical significance with strength of association.

An alarming example from psychology illustrates the prevalence of this fourth sin in the social sciences. The results of a hypothetical statistical analysis were presented to a department of psychologists at a well-known American university. Nine faculty members, each with a Ph.D., and ten graduate students were asked to determine the credibility of findings in two sets of experiments. The first experiment used sample sizes of $n = 10$ and the second involved $n = 100$. Both samples produced significant results. The alarming upshot was that the psychologists felt more confident with the *large* sample results than with significant results derived from the small samples *at the same alpha level*! Not only did they confuse the relationship of n and p, but they were also guilty of sins VIII (using p as a measure of significance) and IV (confusing statistical significance with strength of association). This is particularly poignant, since the average psychologist is surely more statistically aware than the average anthropologist. Rosenthal and Gaito (1963) and Bakan (1967: chapter 1) consider further implications of the above study.

The point is that knowledge of *only* the level of statistical significance tells us precious little about the true magnitude of relationship under study. Several other factors, particularly sample size and the power of the given test, are known to influence probability levels in a manner quite independent from the magnitude of association. A linear correlation coefficient of $r = 0.11$, for example, can be the basis for rejecting H_o in one study, while $r = 0.25$ will fail to reject the null hypothesis in a second study, provided the sample sizes are large enough (300) and small enough (50), respectively (Morrison and Henkel 1969). A complete description of one's statistical analysis should include a statement of statistical significance (if relevant), an appropriate measure of associational strength, and the absolute magnitude of the observed differences.

COMMANDMENT V. Thou shalt not modify a priori hypotheses or the level of significance in light of specific sample data.

It is considered *cheating* to modify one's hypothesis to conform with a previously examined set of data. Once a hypothesis has been developed from a set of data, it becomes farcical to "test" for statistical independence within those same data. Of course they're related! That's what made you suspect the relationship in the first place! When hypotheses are "tested" upon the data that originally suggested them, and statistical significance levels are computed, a spurious sense of validity results. The computed levels of significance may have almost no relation to the true level (Selvin 1957 provides a quantitative example of this sin).

Alpha is only a statement of willingness to commit a Type I error. This willingness should be based upon a general familiarity with the empirical situation, but *not* upon the data contained in a specific sample. Like the statistical hypotheses, the level of significance is an a priori value set in advance of data manipulation and remaining intact throughout the testing episode.

● *I begin to smell a rat.*—M. de Cervantes

But there is the argument that "logic is timeless . . .," and it matters little whether a theory is conceived before, after, or during collection of data. In fact, very few investigators actually set significance levels or define all statistical hypotheses prior to beginning actual fieldwork. This is acceptable because of the special definition of *data*, discussed in Chapter 2. Data are not objects, but rather are empirical observations *made on* those objects. Thus, collection of data is more an analytical process than something one does "in the field." The important concept here is that formation of hypotheses be kept *logically* independent from the data used to test them; temporal order is only a side issue (see LeBlanc 1973, and Williams, Thomas, and Bettinger 1973 for conflicting sentiments on this point).

● *Some people, unfortunately, are lightning bugs; they carry their illumination behind them.*—S. Ervin

COMMANDMENT VI. Thou shalt not collapse contingency tables to generate significant results.

This sin was discussed earlier in Section 11.6. An attribute or ordinal scale, such as postmarital residence pattern or mode of settlement, can generally be dichotomized (or trichotomized) in a number of different ways. Obviously, one's "pet" hypothesis can be favored by the surreptitious manipulating of the cutting points on an ordinal scale. In fact, a judicious adjustment of categories can even change the directionality of a contingency table.

In a manner of speaking, the after-the-fact collapsing of contingency tables is actually a specialized example of the sin warned against in Commandment V (modifying hypotheses after inspecting the data). Contingency tables can legitimately be collapsed only for reasons which are clearly independent of the relations under investigation, such as inadequate data within certain cells (see Yule and Kendall 1937).

● *Don't pull the crime if you can't pull the time.*—Anonymous

COMMANDMENT VII. Thou shalt not test hypotheses with a shotgun.

This book has discussed how to test statistical hypotheses. It must be pointed out, however, that the actual *testing* of hypotheses is but one component in the overall scheme of science. It is equally important to realize just how hypotheses are derived in the first place.

There is no magic formula for generating good ideas. In fact, more than one brilliant scientific notion has arisen from what many people consider most "unscientific" circumstances. Take the case of the chemist Kekule, who had been wrestling with devising an appropriate structural formula to account for the behavior of the benzene molecule (as described by Hempel 1966: 16). One evening in 1865, Kekule was dozing before his fireplace, gazing into the fire. As the flames danced about, one flame seemed to whirl and catch its own tail. Kekule awoke with a start, suddenly realizing he just had solved the problem of benzene structure: The molecule is a hexagonal ring, originally suggested by the flame's dancing shape. Flashes of insight such as this are rare, and it has been said that the true role of genius in science is to suggest new hypotheses, to create order out of seeming chaos. It matters little how this insight occurs.

Once the creative hunches have been generated, deciding which of them best explains the data becomes a purely mechanical process. Research hypotheses are translated into operational statistical hypotheses which are then probed for their ability to explain phenomena of the real world. The null hypothesis either survives this test or perishes, depending upon its performance relative to the data.

These well-known canons of science are mentioned here to reinforce one very simple point. Significance testing is relevant only to scientific validation, not to scientific exploration. Good ideas can come from anywhere and dredging through a mass of unsorted data can be an illuminating procedure. In fact, computing preliminary statistics of description can be helpful in looking for insights. But tests of statistical significance can be misleading when applied in such preliminary studies. The *fishing expedition approach*—applying tests of significance to huge bodies of data in the hope of coming up with meaningful comparisons—seriously undermines the assumptions of significance testing. The results are much too lenient with respect to Type I errors (Kish 1959). It is possible only to test statistically an existing hypothesis. Such tests cannot explain *why* the hypothesis is true nor can they ferret out that hypothesis in the first place. An anthropological example should underscore this difficulty.

In 1966, Robert Textor published a volume entitled simply *A Cross-Cultural Summary*. Textor's "book" actually consists of 20,000 statistical intercorrelations of cultural traits. The 536 pages consist largely of computer printout, weighing a whopping eight (!) pounds. The objective behind tabulating this mass of data is twofold: (1) to allow the anthropologists to test a wide variety of already existing hypotheses, and (2) to generate new hypotheses and hunches. Of the literally tens of thousands of intercorrelations calculated by the computer, only the "best" 20,000 were included in the volume. The "weak" or "less supportable" correlations were winnowed out, using a cutoff point of $\alpha = 0.10$.

Textor's motives are surely worthy, but as Marvin Harris (1968: 621) has suggested, the 20,000 computer correlations "may actually be . . . a source of ignorance and confusion as well as of correlation." There is always a danger that the uninitiated will confuse correlation with causality (Commandment II) and judge a particular relationship as *confirmed* by its mere presence in the printout. One might argue, for example, that:

> *Cultures where male genital mutilation is present tend to be those where metal working is present because the computer showed that results will occur by chance fewer than one time in 10,000 cases.*

Or one could assert that:

> *Cultures where the plow is present tilt more toward being those where a high god is present, with $p = 0.012$.*

This statement is a direct quote from the Textor printout. One's argument could be further buttressed by the rationalization that the probabilities must be right—aren't they computed using Fisher's Exact Test?

Assertions of this sort impress only the statistically naïve. These "laws" are merely the catch of a massive fishing trip. Lacking the interconnecting causal nexus, the "statistical significance" becomes irrelevant because the alpha level has been undermined by repetitive testing. *By chance alone*, we expect about 10 percent (for $\alpha = 0.10$) of the correlations to be spurious.

The point is not to discredit or make light of Textor's massive contribution. Obviously, the effort summarizes a mountain of quantitative cultural data. If simplistic errors in interpretation arise from *A Cross-Cultural Summary*, the fault will certainly *not* lie with the author, Robert Textor. I simply wish to caution the reader, yet again, against any unwarranted reverence for quantitative data in general and the levels of statistical significance in particular.

● *Anthropologists as a group do not know what they know; they do not know the questions to which they have accumulated the answers.*—D. French

COMMANDMENT VIII. Thou shalt not take *p* as a measure of significance.

A common error in the anthropological literature is to take the *p*-value (or alpha level) as a quantitative "measure" of significance for a total population. One encounters statements to the effect that a probability of $p < 0.05$ means "there's less than 5 percent probability that my field results could be due to chance," or "we can be 95 percent confident in our findings," or "we can predict the future outcomes with 95 percent accuracy." Unfortunately, none of these statements is consistent with the inference model employed in statistical hypothesis testing. "*p*" is merely an *inference* about a specific *sample* rather than a concrete measure of a population characteristic. The probability value is just an a priori condition allowing the researcher to determine whether or not H_o should be rejected. That decision will be incorrect a projected number of times per hundred such trials (assuming the assumptions hold). Conditions inferred from large samples—such as differences in means, correlations, associations, and so forth—may well hold true for the total population, but they nevertheless remain inferences.

COMMANDMENT IX. Thou shalt not take the assumptions of statistical models in vain.

Honor these assumptions and hold them inviolate. Better to use a less powerful test or a nonparametric alternative than to rape the underlying model.

COMMANDMENT X. It is a sin to place mathematical elegance before anthropological relevance.

Do not covet the mathematical rigor of neighboring disciplines, nor their infinite replicability, nor their level of predictability, nor anything that is of physical science. The objectives of social science are worthy in themselves and need not be subverted by a seductive technological or statistical reductionism.

● *Beware fanciful desires; you may get lucky.*—P. Roth

16.3 ANTHROPOLOGY, STATISTICS, AND COMMON SENSE

● *Next to knowing when to seize an opportunity, the most important thing in life is to know when to forego an advantage.*—B. Disraeli

When asked if he could read music, folk musician Pete Seeger once replied, "Not enough to hurt my playing." This impious attitude can profitably be applied to statistical thinking on the social sciences: Anthropologists should practice just enough statistics not to hurt their anthropology.

This caveat is not a pitch for ignorance. If I didn't think statistics were important for the anthropologist, I wouldn't have written this book. I am merely pointing out that the euphoric refrain of statistical rigor can lead the unwary down a path eventually ending in superficiality and fruitless inquiry. As Bakan has warned (1967),

> *We must overcome the myth that if our treatment of our subject matter is mathematical it is therefore precise and valid. We need to overcome the handicap associated with limited competence in mathematics, a competence that makes it possible for us to run tests of significance while it intimidates us with a vision of greater mathematical competence if only one could reach up to it.*

I have argued elsewhere in this book that a knowledge of statistics is essential for modern anthropologists so that they will be able to judge when such methods must not be used, as well as when such techniques will be invaluable. It is now time to consider one final statistical test: *Berkson's Interocular Traumatic Test* (Edwards, Lindman, and Savage 1963). The method is quite simple and straightforward, but an elusive test to teach students. The test involves no numbers, no formulas, no tables of probabilities. What could be easier? Berkson's Interocular Traumatic Test states simply: *You know what the data mean when the conclusion hits you between the eyes.* No further statistical methods are involved whenever the Berkson test applies. Learn to apply it well.

● *It requires a very unusual mind to make an analysis of the obvious.*—A. Whitehead