The Philosophical Relevance of Statistics
Author(s): Deborah G. Mayo
Source: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1980, Volume One: Contributed Papers (1980), pp. 97-109
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/192556
Accessed: 15/02/2010 23:02

The Philosophical Relevance of Statistics

Deborah G. Mayo

Virginia Polytechnic Institute
and State University

1. Introduction

The foundations of a number of scientific theories are more or less relevant to philosophy. The better a theory is at elucidating the structure of science and scientific method the more a study of its foundations is relevant for philosophers, particularly philosophers of science. But while philosophers have studied probability and induction, statistics has not received the kind of philosophical attention mathematics and physics have. Although the terms "philosophy of induction" and "confirmation theory" are common, the term "philosophy of statistics" is rarely used; and despite the fact that modern statistical methods have been used increasingly in a number of sciences, specific developments of statistics have been little noted in the philosophy of science literature.

Statistics has been fundamental for determining such things as whether a substance causes cancer, which method of psychotherapy gives the most recoveries, whether censorship causes the instability of nations, whether a fossil belongs to a man or a chimpanzee, whether there is regularity in the oscillations of the velocity of gas in the sun's atmosphere, whether an individual authored a piece of writing, whether the economy will improve, and numerous other issues involving uncertainties. More and more sciences have incorporated statistical techniques leading to a number of new fields, such as biometrics, econometrics and psychometrics. With such widespread use of statistics and so many public policy decisions based upon it, getting clear on the nature of statistical reasoning is imperative. But what is the logic involved in making such statistical inferences, and when are statistical inferences valid? Practitioners have been far more concerned with applying statistics and absorbing statistical techniques into their respective sciences than with answering questions concerning the foundations of these tools. This is not surprising, since it is the business of the philosopher and not the scientist to scrutinize

PSA 1980, Volume 1, pp. 97-109

the foundations of inference methods. However, philosophers have so far been unsuccessful in adequately answering these foundational questions, and controversy and confusion about which methods to use and how to interpret them are greater than ever, both among philosophers and those who apply statistics.

In this paper I want to show that it is important for philosophers to work on foundational questions in statistics because of their relevance both to theoretical and to what might be termed applied problems of philosophy. I begin by considering the philosophical problem of induction and then go on to discuss the reasoning that leads to causal generalizations in science and how statistics elucidates the structure of science as it is actually practiced. I hope to show that in addition to being relevant for building an adequate theory of scientific inference, statistics provides a link between philosophy, science, and public policy.[1] As philosophy of science moves out of the positivistic framework, and becomes more interested in a framework that is relevant to actual problems in science, a tool for bridging the gap between philosophy and science is required. I claim that statistics can provide such a tool, and if I am right, philosophy of statistics has a significant role to play in developing a new framework for philosophy of science.

2. Problems of Induction

What is sometimes referred to as the "old" problem of induction arose from the empiricist principle that the acceptance and rejection of theories is to be solely the result of observation and experiment, and the recognition by Hume that no finite amount of experimentation can justify general theories. Unable to show that induction could be relied on to produce conclusions that were even highly probable, scientific method appeared unjustifiable. This lead some philosophers to attempt a less rigorous means of justifying induction; namely to show that the basic rules for inductive inference in science correspond to what common sense and science deem intuitively reasonable. But what are the basic rules for inductive inference in science? Attempts to answer this question gave rise to the "new" problem of induction. (See Skyrms 1975.)

Many attempts to set out a theory of inductive inference have proceeded by setting up a quantitative relation M which is to measure the evidential strength between data and a hypothesis, and thus the "new" problem of induction involves the problem of finding relation M. While the hypothesis could not be logically certain given the data, the possibility of it having some degree of probability suggested itself, and hence, the use of probability functions in setting up relation M. Typically, M is defined as measuring degrees of belief, confirmation, support, credibility, plausibility and the like. Setting up a purely formal theory of inductive inference by specifying an evidential measure M was very attractive to Positivist philosophers, and the most complete inference theory of this sort is that of Carnap and his successors. Here, M is to hold between two propositions:

one expressing a hypothesis, the other data; and it measures the degree of confirmation the data afford the hypothesis. A number of difficulties were found to stand in the way of their attempt to reconstruct scientific method by means of a formal system (e.g., the "grue" paradox as set out by Goodman (1965)). The simple languages for which Carnap's theories are designed are incapable of expressing many kinds of hypotheses and data needed in actual scientific inferences. And there is also the problem of choosing which of infinitely many functions to use for M.

In general, the inductive logics proposed by philosophers have been too oversimplified to do the actual work of scientific inference.[2] A more promising approach would be to build a theory of induction upon the already well-worked out techniques of statistical theories, methods of hypothesis testing, and estimation. Such a theory of inference will be far more complicated than one composed simply of an evidential relation M. As such, statistics will not provide a "logic" of inference in the ordinary sense, but I think it can provide systematic methods which can cope with the complexities of actual scientific inferences involving probabilities. While theories based on statistics are not going to resolve the "old" problem of induction better than other theories of inductive inference, they can do a better job with the "new" problem, which involves setting up an adequate theory of inductive inference. Given theories which are adequate for the tasks of scientific inference, as I believe those based on statistics are, it is possible to assess how well they accomplish these tasks and so justify preferring some theories of statistics to others. This would provide a much needed basis for resolving controversies between rival theories of statistics.

3. Controversies in the Foundations of Statistics

The major controversies in the foundations of statistics are between two rival inference philosophies: the <u>orthodox</u> or <u>standard</u> philosophy (e.g., Neyman-Pearson and Fisherian) and the non-standard philosophies (e.g., Bayesian and Likelihood). Despite the fact that most of the statistical methods used in the sciences come from the standard theory of statistics, the standard methods have increasingly come under attack both by philosophers and practitioners. The great majority of philosophers who have taken an interest in this controversy have come to support non-standard methods of inference. Among the most philosophically relevant of the criticisms raised against the standard methods are that they fail to adequately represent scientific inference, that they are unable to perform the tasks to which they are put in science, and that they are not objective. Such criticisms are of relevance to those philosophers interested in the logic of scientific inference. If the criticisms against the standard method are correct, then much of what passes as science is misconceived and one would not want to build a logic of inference upon it. In general, the problems concerning the foundations of statistics are relevant for philosophy of science because any adequate account of inductive inference and scientific method will have to cope with precisely these problems.

Hence philosophy of statistics provides an important source of problems for theories of scientific inference, and any theory of inference which is constructed without an awareness of such problems is likely to be irrelevant to actual scientific practice.

Most of the philosophical criticisms of statistical methods are based upon their formulation in statistical texts, and because these formulations have little relation to how statistics is actually practiced, the resulting criticisms have little bearing on much of the use of statistics in science. This point was illustrated at a recent conference on statistical foundations, where J. Neyman (1977, p. 97) told of the following exchange between an "authority" in non-standard statistics and a practicing statistician. "The Authority: 'You must not use [standard inference methods]; they are discredited!' Practicing Statistician: 'I use [them] because they correspond exactly to certain needs of applied work.'" A major task for philosophers of statistics, as I see it, is to uncover and make explicit that which enables certain statistical tools to "correspond exactly" to the needs of the practicing statistician. Understanding what enables these methods to perform valid and informative inferences when they do, is an important step towards forming a theory of inference adequate to science as it is actually practiced. This would be a significant contribution even for those practitioners who do correctly use statistics, for even they may not correctly comprehend why the methods work. At most they follow "rules of thumb" which, while not presently included in the formal expositions of statistics, might be made precise enough to be incorporated in an eventual philosophical theory of inference. To accomplish this task, a philosophical examination of the aims of science and the role that statistics actually plays in science is needed. Constructing an adequate theory of inference would shed light not only on certain controversies in statistics but on certain problems in philosophy as well.

Many of the problems of statistics are not very different from those with which philosophers of science wrangle. For example, statistics is concerned with the following questions: What should be observed and what may be inferred from the resulting observations? How well does data confirm or fit a model? What effects do the processes of observation and measurement have upon the results observed? How can spurious relationships be distinguished from genuine lawlike ones? How can a causal hypothesis be tested? Although the language in which statistical questions are expressed differs from the terminology of philosophers of science, the concerns of one often parallel the other. For example, where philosophers want to know what evidence is needed to reject a theory, a statistician wants to know what is the best <u>rejection region</u> for a test hypothesis. Similarly, the philosopher's concern about objectivity of observation is parallel to the statistician's concern to obtain unbiased observations by such means as <u>randomization</u>. Statistical problems are relevant to philosophy in two ways: the solutions to the problems statistics has solved may suggest solutions to parallel problems in philosophy (e.g., obtaining objectivity of observation via randomization), and those statistical problems which remain unsolved may suggest analogous philosophical problems not recognized by philosophers. Many of the statistician's concerns are, in a sense, applied versions of some of the concerns of philosophers of science, and as Kempthorne (1976) suggests, statistics may be seen as "applied philosophy of science."

4. Statistics and Causality: Testing a Causal Claim

Few problems have generated as much philosophical interest as the problem of causality and the reasoning behind causal inferences in science. Statistical tests play an important role in establishing causal (and other) generalizations in science, but precisely how they do so is not something revealed in statistical theory itself. Consideration of the relationship between statistical inferences and substantive scientific inferences, such as those involving causality, is typically considered by statisticians to fall outside their proper domain. A philosophical theory of scientific inference which did include such considerations could provide a link between statistical and scientific inferences, and in doing so elucidate the logic of causal inferences. An examination of the role of statistical tests in evaluating causal claims is necessary for such a theory of inference, as well as for assessing the criticism that the (standard) tests are irrelevant for making causal inferences.

Consider for example the claim that saccharin causes cancer in rats. If this claim is true it does not mean that all rats fed saccharin will get cancer. Hence, in testing this claim, an observable prediction will be an assertion about the average or mean number of saccharin-fed rats which get cancer. (The average is an example of a statistic.) However, the average number of saccharin-fed rats who get cancer may be high even if saccharin is causally irrelevant to cancer, since it may be the result of some other cause. What one wants to know is whether on the average the number of cancers in saccharin-fed rats (i.e., treated rats) is significantly greater than the number of cancers in rats not fed saccharin (i.e., control rats). When possible,[3] this information may be obtained by carrying out a comparative random experiment.

Without going into the details of the experimental design,[4] the idea is roughly to take a random sample of rats who do not have cancer, treat half of them with saccharin (for a suitable period of time) while leaving the other half untreated. At the end of the experiment the average number of rats having cancer in each group is recorded. Let M_t and M_c be the mean number of rats observed to have cancer in the saccharin-treated and control groups respectively. In this case, one wants to know whether M_t is sufficiently greater than M_c to consider that a genuine causal relationship between saccharin and cancer exists. (In other cases one may not be interested in the specific direction of a difference.) To answer this, a statistical test of the significance of differences may be carried out.

In a nutshell, what the test consists of is a rule which designates, before the experiment, which observations are going to be taken to reject the statistical hypothesis under test; that is, the test or null hypothesis h_0. These observations make up the rejection region, and it is specified so that there is a given probability, called the size or the significance level of the test, of obtaining such observations given that the null hypothesis is true. Conventionally, significance levels are chosen to be .05 or .01. One obtains the observation and either rejects or fails to reject h_0 according to whether or not it falls in the rejection region. (In Neyman-Pearson tests, failure to reject h_0 leads to accepting the alternative hypothesis.)

The logic of testing basically follows the pattern of modus tollens, and in the saccharin example it takes the following form. Since, if saccharin is causally irrelevant to cancer (in rats), then there will be no difference in the mean cancer rates in the populations of saccharin-treated and non-saccharin-treated rats, by rejecting the consequent (that the means are equal), it can be concluded that saccharin is not causally irrelevant. Letting μ_t and μ_c be the mean cancer rate(s) in the populations of saccharin-treated and control rats respectively, the null hypothesis can be formally expressed by h_0: $\mu_t - \mu_c = 0$. The alternative may be that $\mu_t - \mu_c$ is either not equal to zero (two-sided test), or is less than or greater than zero (one-sided tests); and in this case one is likely to be interested in the last of these. As the entire population of treated and control rats are unavailable for inspection, a (random) sample of each is taken and the difference between the mean number of cancers is observed. Thus, the data for this test may be modeled by the statistic S: $M_t - M_c$. The following four levels summarize the way the test statistically models the actual observations and the scientific claim:

Scientific Claim: Saccharin is causally irrelevant to cancer in rats.
Statistical Hypothesis: h_0: $\mu_t - \mu_c = 0$

Statistical Data: S: $M_t - M_c$ (test statistic)

Scientific Data: Observations on a random experiment of saccharin-treated and control rats.

The test consists of the rule: reject h_0 just in case the probability of S given h_0 is less than .01 or .05 depending on the signficance level (i.e., just in case S falls in the rejection region).

However, when tests are carried out by mechanically setting up a .05 or .01 significance level and then rejecting or accepting the null hypothesis, the resulting inferences are often misinterpreted, and many of the criticisms of tests are the result of such misinterpretations. To avoid misuses and misunderstandings, testing logic must be better understood. Tests, as I see them, function primarily to detect certain types of discrepancies between observations and hypotheses and between hypotheses. By a suitable choice of test statistic, larger discrepancies correspond to smaller probabilities. One wants to reject h_0 in our example not directly because the observed difference S is very improbable under the assumption that h_0 is true, but rather because S is sufficiently large. While it is often not realized, with enough observations an improbable difference may actually be of trivial magnitude and then the resulting rejection of h_0 is misconstrued. Such misuses may largely be avoided by formulating the test so that it rejects h_0 just in case an observed discrepancy is large enough to indicate that h_0 is false. Precisely how I think such a reformulation of tests can be made to do this is a matter to be taken up in a separate paper. (See Mayo 1980).

5. Lawlikeness and Causation by Chance

A related problem of philosophical significance (particularly for resolving the "new" problem of induction) is distinguishing between genuine or lawlike (or projectible) regularities and spurious ones. Although the standard theory of statistical testing plays a major role in rooting out lawlike relationships in science, philosophers have rarely made direct use of these tests in dealing with this problem. This is unfortunate, as I think an examination of the logic of statistical tests and their role in science is quite relevant to the philosophical problem of lawlikeness.

Statistical tests provide a systematic way of determining whether an observed difference is large enough to indicate a lawlike relationship or small enough to be attributed to non-lawlike or <u>chance</u> factors. The null hypothesis typically takes one of the standard forms for asserting that the difference is "due to chance factors". In our example, the null hypothesis models the claim that the relationship between saccharin and cancer is not lawlike by asserting that the difference in means is due to chance. Accepting (or failing to reject) the null is tantamount to concluding that the observed difference is not indicative of a systematic or lawlike relationship. In testing a non-lawlike relationship (e.g., between saccharin and eye-color in rats) one would not want to reject the null hypothesis since it would be true. However if the null is true, there will still be discrepancies between the hypothesized difference in means (i.e., 0) and the observed difference in means. Hence, what is needed is a standard for determining whether or not a discrepancy can be plausibly attributed to chance, and probability models provide such a standard. The basic probability models are derived from observations generated by probabilistic mechanisms such as tossing a fair coin (i.e., probability of "heads" is .5). The type of discrepancy between .5 and the proportion of "heads" observed in tossing a fair coin is a standard measure for a discrepancy that is due to chance.

The notion "due to chance" is in need of careful philosophical scrutiny, but here a few remarks must suffice. It is often wrongly thought that attributing a discrepancy to chance means that the observations stem from a probabilistic mechanism. As I see it, it may simply mean that the type of discrepancy is similar to the type known to arise from a probabilistic mechanism, and hence may be represented by a probability model. It may merely serve as a way of saying that the discrepancy is not large enough to be of interest at a certain stage of research. As the practice of attributing certain discrepancies to chance is frequent in science, any theory of scientific inference must take account of it, and examining statistical tests is relevant for doing so.

If the null hypothesis in our example is rejected, it can be concluded that saccharin is not causally irrelevant to cancer in rats. However, this does not imply the conclusion that saccharin causes cancer in rats. The reason is that even with the randomization techniques of experimental design, not all of the relevant factors can be controlled (or even known), and the difference may be due not to saccharin but to these uncontrolled factors. Still, the statistical hypothesis test may be seen as a step in the reasoning processes that lead to inferring causality. Rejecting the null hypothesis (in an appropriately designed test) at least suggests that further experimentation is warranted; accepting it permits the causal claim to be denied.

Getting clear on the reasoning behind causal inferences is not only relevant to the theoretical problems of causality and lawlikeness, but to applications of philosophy to problems concerning public policy involving statistics. Questions about the justification of such policy (e.g., banning substances claimed to "have been determined to cause cancer in rats") are very much dependent upon being able to analyze the validity of the statistical reasoning on which they rest.

6. The Structure of Scientific Activity

One very general concern shared by both statisticians and philosophers of science is to provide a structure for scientific activity, and our saccharin example suggests how statistics may be seen to provide such a structure. Although the statistician is interested in providing a structure at the "working level", the structure it provides is relevant for the philosopher's interest in the structure of science. That statistics functions to provide a structure for science at all is typically not recognized. I attribute this to the overly narrow conception of the role of a theory of inference both in philosophical theories of inference and in presentations in statistical texts. Philosophical theories of inference typically begin with the assumption that the data and hypotheses are given, and the only job that remains is to construct a measure of relationship between them. In presenting statistical inference, statistical texts also typically start out with the data and hypotheses given; and in addition it is assumed that the data follows a known distribution. In fact, inference does not begin only after data and hypotheses are given, nor is the data known to follow a given distribution. Statistical reasoning enters at a number of different stages of scientific research, and several statistical inferences may be involved in a single research effort.

Statistical reasoning, construed broadly, is involved in the initial planning of experiments, the gathering and modeling of data, and in the construction and evaluation of hypotheses and theories. In other words, statistical reasoning is involved in constructing and linking the four levels sketched in our saccharin example. On the level of the data, statistics (or more exactly, the statistical theory of experimental design) is involved in specifying what should be observed, and how to observe it, and then in modeling the observation by means of a statistic. Statistical inference links this data to the statistical hypothesis, and lastly, the statistical hypothesis is linked to the substantive scientific claim which it models. Statistics may first be used to detect potentially interesting relationships in the data, to indicate which quantities are likely to be relevant in an eventual theory, and hence should be studied further. For example, one would first check whether there was any relationship between saccharin and cancer before going on to test a specific relationship.

As different inference principles may be relevant at different stages of inquiry, an adequate theory of inference is going to have to be more complex than generally thought. While the standard theory of statistics provides a conglomeration of inference principles, there has not been an attempt to unify them, and doing so is an important task for a philosophical theory of inference built upon (the standard theory of) statistics. (It may also be possible to unify some of the standard principles of inference with non-standard principles.) Each stage of inquiry involves a different kind of model, and it is by linking these models that statistics provides a structure for scientific research. Hence a philosophical analysis of the statistical inference principles involved at different stages of research will offer insight into the structure of sciences which employ statistics. This is particularly true for the biological and social sciences which are faced with much less controlled experiments than the physical sciences, and hence have a greater need for statistics.

Experiments in the physical sciences typically enjoy relatively little error in measurement; items to be measured are nearly identical with respect to the property of interest and they are relatively stable over time. As such, in measuring the effect of a change in things like temperature or pressure, one is not faced with the same variability faced by the scientist of sociology, psychology and biology in measuring effects on animals and humans. Understanding the statistical methods for accumulating knowledge in the so-called inexact sciences is relevant for answering a number of philosophical questions concerning these sciences. The following are examples of such questions: Are the social sciences methodologically or logically distinct from natural science? Are the social sciences inferior? Can they be objective? Building a theory of scientific activity on the model of the physical sciences, and failing to recognize the roles of statistics in science, has often lead to narrow-minded answers to these questions.

Statistical considerations are not absent even in those sciences considered exact. The reason is that only a finite sample of data is available, and the accuracy, precision, and reliability of the data are limited by distortions introduced by the processes of measurement and observation. Scientific theories contain theoretical concepts which do not exactly match up with things that can be observed; data are finite and discrete while theories may refer to an infinite number of cases, and continuous variables such as weight and temperature. As such, the data can rarely be expected to agree exactly with theoretical predictions, and because of this, the observable prediction is often a statistical claim. Hence, whether theories are considered to be probabilistic or deterministic, studying them often gives rise to very similar statistical situations. By examining the role of statistics in both social and natural sciences, it might be possible to see the differences between them as simply being a matter of degree. In any case a clearer picture of their relationship may emerge.

7. Conclusion

To summarize, a study of statistics, its problems and its roles in science, is far more relevant to philosophy than is typically recognized. Though here I have only been able to sketch a few of the major reasons for thinking this, hopefully enough points have been raised to suggest a number of possible ways of putting the foundations of statistics to philosophical work. I have indicated the relevance of statistics for the (new) problem of induction and for elucidating the structure of actual scientific practice and the reasoning behind causal inferences. As such, statistics may be seen to be relevant to general problems of philosophy of knowledge. The following remark of Oscar Kempthorne is of interest in this regard: "To resolve the obscurities about probability and inductive inference is equivalent in my opinion to laying out a philosophy of knowledge. Workers in statistics and particularly those working on theories of statistical inference are on the boundaries of philosophy often, and usually, I fear, without being aware of the fact." (1971, pp. 482-483).

In addition to being useful for theoretical problems of philosophy, statistics provides a bridge linking philosophy to applications of science in government and society, where statistics is extensively used. If the positivistic framework of science is to be replaced by a framework that is relevant to actual scientific problems, philosophy of statistics has an important role to play in its construction. Ronald Giere has put it this way: "One may hope that in time philosophers will appreciate the significance of the [statistical] significance test controversy for inductive logic and it will begin to appear in the philosophical literature. Eventually we may have something relevant to say to sociologists and psychologists who, it seems, very much want and need to hear something relevant from someone who understands the full dimensions of the problem." (1972, p. 180).

Notes

[1] I will not here discuss the philosophical relevance of decision theory, which is a branch of statistics, and which has found applications to moral and political philosophy and is relevant to public policy. For a discussion of this see Suppes (1961).

[2] For a good discussion of the inductive theories of a number of philosophers see Giere (1979a).

[3] When such a randomized experiment is not available, as in historical surveys, more complicated statistical arguments are required.

[4] For a detailed discussion of some of the experiments used in determining the relationship between saccharin and cancer in rats, as well as an excellent explanation of statistical methods in science generally, see Giere (1979b).

References

Carnap, R. (1962). Logical Foundations of Probability. 2nd Edition. Chicago: University of Chicago Press.

---------- (1971). "A Basic System of Inductive Logic, Part I." In Studies in Inductive Logic and Probability, Vol. I. Edited by R. Carnap and R. Jeffrey. Berkeley: University of California Press. Pages 33-165.

Finch, P. D. (1976). "The Poverty of Statisticism." In Harper and Hooker (1976). Pages 1-44.

Fisher, R. A. (1959). Statistical Methods and Scientific Inference. 2nd Edition. Edinburgh: Oliver and Boyd.

Giere, R. N. (1972). "The Significance Test Controversy." British Journal for the Philosophy of Science 23: 170-181.

------------ (1976). "Empirical Probability, Objective Statistical Methods, and Scientific Inquiry." In Harper and Hooker (1976). Pages 63-93.

------------ (1979a). "Foundations of Probability and Statistical Inference." Current Research in Philosophy of Science. Edited by Peter D. Asquith and Henry Kyburg. E. Lansing: Philosophy of Science Association. Pages 503-533.

------------ (1979b). Understanding Scientific Reasoning. New York: Holt Rinehart and Winston.

------------ (1980). "Causal Systems and Statistical Hypotheses." In Applications of Inductive Logic. Edited by L. J. Cohen. New York: Oxford University Press. In Press.

Godambe, V.P. and Sprott, D. A. (eds.). (1971). Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston of Canada.

Good, I. J. (1971). "The Probabilistic Explication of Information, Evidence, Surprise, Causality, Explanation, and Utility." In Godambe and Sprott (1971). Pages 108-122.

Goodman, N. (1965). Fact, Fiction and Forecast. New York: Bobbs Merrill Company, Inc.

Harper, W. L. and Hooker, C. A. (eds.). (1976). Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Vol. II. (University of Western Ontario Series in Philosophy of Science, Vol. 6.) Dordrecht: D. Reidel

Kempthorne, O. (1971). "Probability, Statistics, and the Knowledge Business." In Godambe and Sprott (1971). Pages 470-499.

------------- (1976). "Statistics and the Philosophers." In Harper and Hooker (1976). Pages 273-309.

Lieberman, B. (ed.). (1971). Contemporary Problems in Statistics. New York: Oxford University Press.

Mayo, D. (1980). "Testing Statistical Testing." In Philosophy and Economics. (Western Ontario Series in Philosophy of Science.) Edited by Joseph C. Pitt. Dordrecht: D. Reidel. In Press.

Morrison, D. E. and Henkel, R. E. (eds.). (1970). The Significance Test Controversy-A Reader. Chicago: Aldine Publishing Co.

Neyman, J. (1977). "Frequentist Probability and Frequentist Statistics." Synthese 36: 97-131.

Rosenkrantz, R. (1973). "The Significance Test Controversy." Synthese 26: 304-321.

Skyrms, B. (1975). Choice and Chance. 2nd Edition. California: Dickenson Publishing Co., Inc.

Spielman, S. (1978). "Statistical Dogma and the Logic of Significance Testing." Philosophy of Science 45: 120-135.

Suppes, P. (1961). "The Philosophical Relevance of Decision Theory." The Journal of Philosophy 58: 605-614.