

Science, Philosophy of Science and Statistics

Introduction

Webster's (second edition) defines statistics in two ways. In the singular, statistics is "The science of the collection and classification of facts on the basis of relative number or occurrence as a ground for induction; systematic compilation of instances for the inference of general truths; the doctrine of frequency distributions." In the plural, statistics are "Classified facts respecting the condition in various respects of the people in a state, or representing any particular class or interest; especially those facts that can be stated in numbers, or in tables of numbers." Note that the singular definition of statistics is that of a scientific discipline dedicated to developing theories and procedures to provide a "ground for induction" and a means "... for the inference of general truths." In other words, a discipline dedicated to evaluating data as a basis for making decisions. The plural definition of statistics could be adequately condensed to "numerical facts" or "numerical observations." On the other hand, the word "statistic" without a terminal "s" connotes an "... estimate of a number or numbers describing a numerical property of an assemblage or population..." So the plural definition of statistics along with the word "statistic", provide a second purpose for the discipline of statistics: to summarize and describe data or information and to estimate properties of assemblages or populations of instances or individuals.

This semester our intention is to examine each of these facets of the discipline of statistics. For now I will focus on the relationship between statistics as a means "... for the inference of general truths", and another discipline that also endeavors to develop a methodology to achieve the same end: the philosophy of science.

Philosophy of Science: In a Nutshell

Have you ever asked yourself; why does science appear to progress? If you have, you have probably been led to ponder a few possible answers such as:

- 1) "scientific method" (whatever it may be) allows or insures that the right questions are asked and answers are rapidly forthcoming.
- 2) something inherent in "natural science" questions make them more amenable to solution than social science, political, or ethical matters.
- 3) scientists have been attacking their questions for a longer time, with greater funding (brute force).
- 4) non-scientists aren't very clever.

You may as well have questioned the question. Does science really progress? Or at least, any faster than fields of endeavor generally considered non-scientific?

If you were either adventurous or foolish enough (as the case maybe) to think about the initial question this much, you probably either concluded: gee, "I don't know," and went to the basketball game. Or inquired further; "How can we answer these questions, and then went to the basketball game. Of course, after celebrating a victory for the home team with a six pack, you would wake up groggy and shaky the next morning, but the questions would still be gnawing at your mind.

Having gone this far one has at least five alternatives:

- A) Go about your business, don't worry about the logical structure of science, it works! "Scientists build bridges that don't fall down."
- B) Drop what you are doing and pursue these questions.
- C) Do nothing, just worry about these questions occasionally.
- D) Check the literature to see if some answers have been suggested.
- E) Check to see if it is a degree requirement.

Of course, since you are all biologists at SFSU, you tend not to treat serious questions cavalierly, and since you consider yourselves scholars as well as scientists you have a command of the literature, and a facility with library research that allows you to make quick work of this matter.

But, there's a problem. Everybody seems to be interested in these questions. Philosophers, scientists, sociologists and historians alike. And this means that there are masses of books and papers that all purport to deal with these problem in some form.

Now for the average student, who is overworked and underpaid, there just aren't enough hours in the day to deal with these "peripheral questions". So you are at the mercy of your mentors (dementors) or your task masters. If they think it important enough maybe you'll sit through a couple of boring lectures on this topic and read a few papers on it, or maybe they'll find something else unreasonable for you to do in your spare time. So without further adieu, one oversimplified, and slightly warped view of science and how it progresses.

The popular view of "scientific method" that existed and to a great extent is rampant today is attributed to Francis Bacon (1561-1626). One could describe it thusly:

A scientist carries out carefully measured observations and experiments in the area in which he hopes to make discoveries, and gradually amasses a lot of hard data. Sooner or later this data begins to exhibit certain general features, and these in turn suggest a law like hypothesis. The scientist seeks to verify this hypothesis by finding evidence that conclusively proves it. If successful, it becomes a permanent addition to the body of our certain knowledge, and so the scientist will have discovered a regularity of nature, a law that can be put to use wherever it applies.

In other words, science is seen to proceed from specific instances to "law like" general statements by finding specific instances that verify the general statement (hypothesis). This method of proceeding from observed instances to general statements is the process known as induction. In fact, an element of this "inductive method" served as (far as I know) the first criterion suggested to demarcate science and non science:

Demarcation Criterion I - Verifiability: Statements of an intuitive nature, arguments from authority, or on any other ground were considered not to be scientific. Only statements of observable fact, and statements validly induced from them were scientific, and only these statements were genuine certainties.

Scientific progress therefore was the accretion of new certainties, new inductive statements, and observable facts. This was the ruling view at least until the 1920's, and actually is still quite popular. However, in the 18th century David Hume (1739-1790) asked some embarrassing questions about induction. Hume contended:

- no number of observation statements could logically entail a general statement. In other words, the "weight of the evidence," the number of instances observed to be consistent with the hypothesis was insufficient to logically demand that an hypothesis be true.

The classic puerile example that Hume used to illustrate his point had appeared in logic texts since the middle ages and continues to appear to this day. It goes like this:

"All swans are white."

- By induction, millions of swans had been observed by millions of observers and all were white.

- But Hume contended, that it was still possible that somewhere, as yet un-observed, there were swans that were not white.

Yet, all general statements (scientific theories, laws or hypotheses) are essentially of this sort. They purport to explain all instances of the phenomena in question, even those that have not yet been observed. Because of this every induced general

statement can have its black swan. For example Newton's law of gravitation is taken to hold anywhere at any time. Yet, if as Hume contends, it cannot be verified by observation, how has it been proved?

From the time of Hume until the 1920's this famous problem of induction, was treated by most of the world's great philosophers. In the 1920's, events had or were occurring that led Karl Popper to propose a solution to the problem of induction.

First of all, philosophers are as steeped in "physics envy" as are ecologists so that Einstein's work on relativity, and the growth of quantum mechanics did not go without notice. Second, the avant-garde of philosophers were centered in Vienna, Austria and were known as the "Vienna circle". Because of their envy of these counter-intuitive advances in physics, they endeavored to apply scientific method to philosophy and founded the philosophical school known as logical positivism. This is basically the old inductive method applied to determine what sorts of statements were meaningful. Only statements of observable fact and general statements that could be induced from them are considered meaningful by logical positivists. Only general statements that could be proved were meaningful. Everything else was metaphysical. The verifiability of a statement rendered it scientific, hence the logical positivists and inductivist demarcation criterion between science and non science is a verification criterion. Only verifiable statements are scientific.

How does Popper fit in? He was a school teacher in Vienna at the same time and was interested in philosophy. He had occasion to meet and debate a few of the members of the Vienna Circle at intellectual gatherings held in individual homes. But he wasn't a member of the Vienna circle, more a thorn in their collective side.

In 1934 Popper published his book (in German) *Logik der Forschung* in this he presented his solution to "the problem of induction", and suggested a new criterion by which he could demarcate science and non science. His solution to the problem of induction goes like this:

General statements (universal statements, theories, or hypotheses) cannot be proved, but they can be disproved. For example, "All swans are white." - No number of observations of white swans will prove this statement. But, a single observation of a black swan can disprove it.

So, Popper contends that a search for conclusive verification is irrational, but that attempted refutation is rational. Popper's contention amounts to a claim that there is a logical asymmetry between verification and falsification. But, this implies something that most find discomfoting. We never actually know a scientific statement to be true, only for it to be false or not false.

Demarcation Criterion II - Falsifiability: According to Popper science is demarcated from non science since general scientific statements are at best in

principle - falsifiable - rather than verifiable as the inductivists and logical positivists contend.

To Popper the *modus operandi* of science is to pose a "bold conjecture" and to attempt to critically refute it. Failure to refute the conjecture does not imply that it is "proved", but continued failure to refute in the face of the most clever, rigorous and critical test imaginable does increase the "relative truth content" or "verisimilitude" of the conjecture. Remember Newtonian mechanics was indeed verisimilitudinous prior to the advent of relativistic physics.

To summarize Popper's view:

1. We do not start from observations.
2. We do not generalize from these to form hypotheses.
3. We do not verify our hypothesis.
4. We do not arrive at final certain knowledge.
5. We pose a "bold conjecture" (a general statement or hypothesis).
6. We attempt to critically refute this hypothesis by performing clever experiments or collecting the crucial contradictory observations.
7. Repeated failure to refute our conjecture increases its verisimilitude.

Science is then seen as a series of superseded theories that did not arise out of our observation and experiments, but rather the theories precede and are tested by observation and experiment.

This is not to say that observations play no role in theory development, only that different sets of observations must serve in theory development than those used to test a theory. Also, Popper does not suggest that observations and experiments are the neutral arbitrators of science. For our minds are not "empty buckets." Rather we see and filter all "data" with a plethora of implicit theory. All knowledge is theory laden - never take an observation for granted.

Now if you were to read Popper, you would find not only a formal logical discussion laced with symbolic logic and logical "proofs", but also examples and discussion from the history of science (particularly physics). Are we then to interpret Popper as describing how he thinks science works and progresses, or as prescribing how scientists should proceed?

As I mentioned Popper suggests that all knowledge is theory laden. He also contends that language, as well, is theory laden. If so, maybe an examination of our language will shed some light on how descriptive or prescriptive Popper is.

Consider:

Have you ever heard the old adage - "The disproof is in the pudding"?

The point that I am trying to make is that our language, and probably most of our thought is steeped in the inductivist - logical positivist tradition not in the falsificationist mode. So is Popper talking and writing about what he think has happened or what he wants to happen? I believe more of the latter. Note also that he suggests only one method and his arguments are all based on formal logic.

Another view, that of T.S. Kuhn (1970), as to how science progresses can be summarized as follows. Kuhn sees the activities of scientists as being dictated by paradigms - major theories or ruling world views. Paradigms are presented as textbook wisdom and are no longer questioned by scientists. Rather, they serve to structure and orient scientific endeavors because they suggest what questions can be legitimately asked, what assumptions should be made, and conceivably what approaches should be applied. Periods when scientists work under such paradigms he terms "normal science." Controversies that exist during these times usually arise because individual scientists may be laboring under different paradigms. Kuhn suggests that scientists laboring under different paradigms, although they may attempt to argue, do not communicate because the language, ideas, background assumptions, and methodology associated with different paradigms are incommensurable. That is, it is as if one scientist was talking about apples and the other about oranges, or as I have heard said "... one guy was talking television and the other was listening on radio."

How then do ideas or paradigms change and how does science progress? Kuhn suggests that a change in paradigm is precipitated by anomalous observations or data, or phenomena that cannot be explained by the existing paradigm. These anomalies foster a growing sense of crisis that may lead to critical scrutiny of the existing paradigm. However, Kuhn believes that if reasonable alternative paradigms are not available then these anomalies would be just set aside, ignored, or explained away as special cases with no bearing on the existing paradigm. When change does come Kuhn believes that it is usually revolutionary - not gradual. Two examples that may fit Kuhn's dialectic are:

- Darwin's proposition of the theory of evolution by natural selection to replace the concept of divine creation and the fixity of species. The anomalies Darwin could explain included what fossils represented, how plants and animals could be modified under domestication and breed true, and disjunct patterns of geographical distributions of plants and animals.

- The theory that crustal plates are in continuous motion resulting in continental drift. This view was first proposed in the 1840's, but a reasonable geophysical mechanism was not proposed until the 1960's. Continental drift explained the "fit of continents", the presence of coal in Australia, and the occurrence of closely related fossil and extant species with widely disjunct geographical distributions.

When simplified the debates surrounding the proposition of these theories may appear to mimic Kuhn's notions of a paradigm change. However, Kuhn's dialectical view of how ideas change in science may represent a false historiography imposed on these examples after the fact rather than a process necessary to scientific progress.

Kuhn also describes what he calls an "immature science" as a discipline without paradigms. Usually such a discipline is characterized by a lack of textbooks, little hierarchical accrual of knowledge and researchers who work independently on different questions.

Again how much of this view of change in science is description and how much prescription? Kuhn is both a historian and a philosopher of science, and his writings are liberally laced with examples from the history of science. Note also that under Kuhn's view much of the tension between opposing or alternative paradigms and much of the reason for their change is psychological! When do the anomalies and the sense of crisis cease to stimulate "willful suspension of experimental results", and instead lead to a paradigm change? Presumably, a Popperian falsificationist approach may be applied to "normal science" questions. But in Kuhn's view it is not a single critical experiment that falsifies a paradigm and therefore forces a search for another. Kuhn even suggests that paradigm changes are generational changes. Senior scientists that have labored their life under one paradigm never relinquish it. Rather, the next generation of young scientists are responsive to the existing anomalies and adhere to the new paradigm.

So far I have presented two views concerning how scientists should or do do science. These are aptly described by the title of Kuhn's paper in the book "Criticism and the Growth of Knowledge," "Logic of Discovery or Psychology of Research." One view, Popper's, being based purely on formal logic denying or at least minimizing the role of the researchers' psychological state, social milieu, or other economic or political conditions. The other, Kuhn's view emphasizing a psychosocial decision to abandon one paradigm in favor of another.

Another young philosopher of science, Imre Lakatos, who died in a motorcycle accident in the early 1970's, presents a more comprehensive view of science and its methods (Lakatos and Musgrave 1970). Lakatos' philosophy which he calls the "Methodology of Scientific Research Programmes" subsumes both Kuhn and Popper's views, as well as including more about the sociology of science.

Lakatos refines Popper's falsificationist doctrine by distinguishing "naive falsificationism" from "sophisticated falsificationism." Naive falsificationism is the contention that a theory or perhaps an entire research program can be falsified by a single observation statement. The black swan or critical refutation of which Hume and Popper speak. On the other hand, sophisticated falsification denies the possibility that a single critical experiment or refutation can or even should be sufficient to lead a single researcher or a discipline to abandon a particular theory or research program. In sophisticated falsificationism, an existing theory or research programme (read paradigm) can only be falsified by the proposition of a new theory that: 1) predicts novel facts, 2) explains existing observations that are inconsistent with previous theory or are considered anomalous in light of existing theory, and 3) allows rapid corroboration of novel facts.

"... no experiment, experimental report, observation statement or well corroborated low level falsifying hypothesis alone can lead to falsification. There is no falsification before the emergence of a better theory."

So in Lakatos' view, sophisticated falsificationism leads to a "progressive problem shift" since the new theory has some excess empirical content over the old theory. Lakatos also describes degenerating problem shifts as instances where new theories or research programmes that lack this excess empirical content gain favor for whatever reason.

A final aspect of Lakatos' view concerns the sociology of problem shifts. He defines what he calls the negative and positive heuristic of a research programme. First I will return to Lakatos' definition of a research programme: a line of investigation that has a certain amount of continuity among its progressive and degenerative problem shifts and consists of methodological rules, some of which tell us what paths of research to avoid (negative heuristic) and others tell us what paths to pursue (positive heuristic). The negative heuristic is the hard core of the research programme and the discipline forbids one to question it, protecting it from attack by inventing *ad hoc* auxiliary hypotheses to explain anomalous observations. The positive heuristic of a research program specifies how to handle anomalies through modifying the auxiliary protective hypotheses. So these heuristics dictate the behavior of researchers in defending or challenging some aspect of the hypothetical substance of a research program. Maybe Lakatos is describing how to do normal science in Kuhnian terms?

Once again Lakatos' view is a very specific model as to how science progresses. His writings are again laced with citations from the history of physics, so we don't know if he is being merely descriptive or prescriptive. The key point to remember about Lakatos' view is that it gives the Popperian notion of the logical structure of scientific inquiry - conjecture and refutation - a central role in scientific progress, yet places science in the context of the researchers psychology, the sociology of the discipline, and the broader social economic and political milieu.

The final notion of scientific method and scientific progress that I will present is that of Paul Feyerabend (1975). Feyerabend's view of scientific method is best described by the title of a Cole Porter song, "Anything Goes." Feyerabend believes that individual scientists should lie, steal, cheat or do whatever is necessary to insure that their own ideas hold sway. If this can be accomplished by reasoned argument, by aping Popperian, positivist or, Kuhnian notions of method, or by blatant dis-ingenuities it makes no difference. The strongest most persuasive individual's ideas will win out until they are replaced by ideas proposed by someone more persuasive. Presumably to be persuasive an individual scientist must advocate more useful and verisimilitudinous ideas, adduce more facts or observations to support one's ideas, and propose theories that explain novel facts.

Feyerabend's view has been interpreted as an argument for methodological pluralism when it comes to doing science. As such, it has been used to attack Popper's notions of science as a process of conjecture and refutation, and to legitimize science by advocacy.

The Relationship Between Scientific and Statistical Hypothesis Tests

Concern about method in science is motivated by the hope that understanding how science works will enable us to somehow do science better. But, what do we mean by "do science better," and how can we evaluate whether one or another methodological or philosophical approach to science is really better? As a practicing scientist rather than a philosopher of science, my purpose is not to attempt to show that one scientific method or philosophy of science is logically better or worse than another. Rather, I will examine the practical impact of laboring under one or another philosophy of science or one another view of scientific method.

If scientists are motivated by a desire to understand how nature works, and we view scientific method as a process of attempting to answer questions to this end, then we can specify three characteristics that one might demand of a "good" scientific method. First, scientific method should tend to lead to a "true" picture of nature and how it works. Second, scientific method should minimize the risk of making errors which could lead to an incorrect view of nature or at least divert effort into fruitless dead ends. Lastly, the application of scientific method should lead most rapidly to the correct solutions to the question posed. In other words, we would desire a method that would most often lead us down the most direct path through the maze of possible research endeavors to the correct solution to the problem posed.

Unfortunately, it is considerably easier to catalogue the characteristics one would desire in a good scientific method than it is to actually assess the relative merits of one or another method in regard to these characteristics. Such a comparison would fall in the domain of an historian of science. What data could one collect? Could a comparison be made of the accuracy, facility and rapidity with which scientists using

one or another method solve research questions? Should such a comparison only be attempted between scientists using different methods to answer the same questions, or between progress in different scientific disciplines that use different approaches? I believe that all such comparisons are doomed to failure because of inequities in the difficulty of different research questions, and of resources committed by different research groups toward solving the same question. These endeavors would also be plagued by the lack of information chronicling the methods used by different scientists, and because of the large element of creativity involved in generating possible answers to a particular question. They also would be confounded because scientists seldom strictly adhere to a single methodological approach. Instances of multiple independent discovery present some hope for actually performing such a comparison, but accurately evaluating the efficiency of a particular research program is still problematic, much less attributing efficiencies or inefficiencies to specific scientific philosophies or scientific methods.

Scientific Method and Error Elimination

I have painted a pessimistic picture of our ability to empirically evaluate the relative merits of different scientific methods. However, I still believe that we can compare scientific methods by examining the propensity of each method to lead to certain kinds of errors in judging the merits of a theory or hypothesis. In other words, we may be able to use the second criterion I mentioned above to establish a basis for preferring one particular scientific method over another.

The process of evaluating a theory or hypothesis can ultimately lead to two distinct correct conclusions, and two distinct kinds of errors. We can correctly reject an hypothesis when it is false, or fail to reject it when it is true, and we can incorrectly reject an hypothesis when it is true or fail to reject it when it is false. In statistical hypothesis testing these two kinds of errors are termed Type I and Type II errors, respectively, and procedures to control the chance of making these errors are well developed. If we view theory evaluation or hypothesis testing in general as analogous to statistical hypothesis testing, then we would like to adopt a scientific method that minimizes the chance of making both of these kinds of errors, or that at least is more effective at minimizing the more costly of these kinds of errors. Which kind of error might be more costly depends on the hypothesis being examined and the motivation for the inquiry.

In statistical hypothesis testing, the tested hypothesis is often in the form of a "null" hypothesis. A null hypothesis is simply a conjecture that a particular causal process or relationship is not evidenced in the observations under study. While the tested hypothesis could also posit a process or relationship, most often the presence of some individual process or relationship or class of processes or relationships is proposed as an alternative hypothesis. Strong et al (1979) have incorrectly suggested that null hypothesis have logical primacy relative to other hypotheses.

However, both kinds of hypotheses can serve as the objects of examination and testing. Rather, the widespread use of "null" hypotheses in statistical hypothesis testing arises because of an asymmetry in the costs of Type I and Type II errors and the ease with which the probability of making these errors can be controlled. In tests of statistical hypotheses, the probability of rejecting the tested hypothesis when it is actually true (Type I error) is set by the investigator and is equal to the significance level (α) of the stated test. The probability of a Type II error (β), failing to reject the tested hypothesis when false is indirectly controlled by the experimenter by specifying μ , the sample size, and the alternative hypothesis, but also depends on the variance of the sampling distribution of the test statistic. The chance of making a Type I error is therefore known prior to performing a test, but unless prior information on the expected variance of the phenomenon of interest is available, the probability of a Type II error cannot be specified in advance of a test. Because of this, the risk of making a Type II error is often greater than of making a Type I error. If the tested hypothesis is a "null" hypothesis, then by setting α we can closely control the likelihood of concluding that some process or relationship is manifested by the evidence when actually it is not. However, it is more difficult to know, if we fail to reject the tested null hypothesis, the chances that we do so incorrectly. While an effort can be made to minimize this risk, the risk of failing to find a process or relationship to be important when it truly is, will generally be greater than of finding a process or relationship to be important when it is not. So given a tested hypothesis that is a null hypothesis, one tends to err on the side of ignorance. That is, one will tend to make Type II errors more often than Type I errors.

If the tested hypothesis is not null and posits the operation of some process or the existence of some relationship, then the implications of Type I and Type II errors are quite different. Since the risk of rejecting the tested hypothesis when true is lower than the risk of failing to reject it when false, if the tested hypothesis is not null we will tend more often to err by find a process or relationship to be important when it is not truly so.

But which of these errors is more costly? Would we rather err on the side of ignorance or by concluding that a process or relationship is important when it is not truly so? As I said above this depends on the motivation for a particular inquiry. In general for the evaluation of scientific theories, I believe one would rather err on the side of ignorance. My reasons for this are twofold. First, if we set more stringent standards for concluding that a theory is true, we are less apt based on incomplete or weak evidence to champion a pet theory that may actually be false. In this sense, it is a means to guard against advocating theories by exposing all theories to close scrutiny. Second, and as a result of the first point, it will tend to prevent us from proceeding to elaborate more theories based on the incorrect acceptance of a prior theory. In other words, it requires a more rigorous demonstration of the reasonableness of a theory before that theory can be used to justify more elaborate, comprehensive, or ancillary theories. It is of interest to note that the basic tenet of American jurisprudence, that the accused is innocent until proven guilty, parallels my contention that decision rules in scientific hypothesis testing should be designed to

err on the side of ignorance, if to err at all. On the other hand, had our inquiry involved an applied problem such as screening drugs or chemicals for toxicity to humans, the cost of a Type II error, failing to conclude that a substance is toxic when it truly is, will generally be greater than of concluding that a substance is toxic when it is not. In this instance, using a non-null tested hypothesis would allow tighter control of the chances of making the more costly error.

Another important aspect of error elimination in hypothesis testing involves the specification of the alternative hypothesis. Seldom can we construct an experiment in the environmental sciences where the negation of a null tested hypothesis logically entails the acceptance of a specific alternative hypothesis. Our inability to use such "strong inference" in the environmental sciences (Platt 1964) should be viewed as both a mandate for more carefully planned experiments and as a warning that stating the full class of reasonable alternative hypotheses is as important as stating the tested null hypothesis (Chamberlin 1890). This class of alternative hypotheses should contain hypotheses of multiple causation and of complex causation, that is, factor interaction (Hilborn and Stearns 1983, Quinn and Dunham 1983), in addition to single factor theories. Failure to identify the full class of alternative hypotheses can lead to serious delays and errors if the correct alternative is not among the class of stated alternative hypotheses.

The importance of both of these aspects of error elimination in hypothesis testing; erring on the side of ignorance and delineating the full scope of alternative hypotheses was underscored by Chamberlin (1890) in his delightful essay "The Method of Multiple Working Hypotheses." But, how do these ideas relate to more formal prescriptions about how to do science? To me the formalism embodied in statistical hypothesis testing, in the vein I have outlined above, is most similar to the deductive falsificationist approach to theory evaluation outlined by Popper (1959) and Lakatos (1970). An hypothesis, usually a null hypothesis, is posed and evidence is gathered in an effort to reject the tested hypothesis. Rejection of the tested hypothesis leads to testing alternative hypotheses, rather than the immediate acceptance of an alternative. Failure to reject the tested hypothesis leads to a more concerted effort to perform a rigorous test.

Quinn and Dunham (1983) find greater similarity between statistical hypothesis testing and an inductive scientific method. In an inductive scientific method, specific instances provide a basis for adducing the truth of a general theory. The perceived regularity displayed by data is used to adduce the truth of a particular theory. The perception of the regularity in the evidence constitutes the rejection of an implicit null hypothesis that such regularity does not exist, in favor of the adduced theory. As such the probability of rejecting the implicit null hypothesis of no regularity when it is true, a Type I error, is very high. This is the essence of Hume's "problem of induction" outlined in terms of statistical hypothesis testing. Quinn and Dunham (1983) contend that, in ecology, data usually precede the statement of a specific hypothesis and hypothesis test, and that a statistical examination of the data is used to adduce support for some *a posteriori* hypothesis. While I agree that the *modus operandi* of

many ecologists has traditionally been of this form, and that this procedure can be useful in generating hypotheses, it is certainly unsuited to testing an *a priori* hypothesis and can lead to a higher Type I error rate than nominally specified. If an hypothesis is stated after the data have been collected and examined it is still possible to apply the formalisms of statistical hypothesis testing the control for Type I and Type II errors. However, determining the exact probability of making these kinds of errors is very difficult (Selvin and Stuart 1966).

All this is not to say that there is only one way to do science. But rather that the costs and risks involved in each approach must be evaluated in light of the motivation for the inquiry to which they are to be applied. I suggest that in basic science, posing a null tested hypothesis, attempting to critically refute it, and delineating the full class of reasonable alternative explanations will tend to minimize the risks of making the most costly errors in hypothesis testing.

References

- Barber, B. 1961. Resistance by Scientists to Scientific Discovery. *Science* 134: 596-602.
- Broad, W.J. 1979. Paul Feyerabend: Science and the Anarchist. *Science* 206: 534-537.
- Caplan, A.L. 1977. Tautology, circularity, and biological theory. *American Naturalist* 111: 390-393.
- Chamberlin, T.C. 1890. The method of multiple working hypotheses. *Science* 15: 92.
- Connor, E.F. and D. Simberloff. 1986. Competition, scientific method, and null models in ecology. *American Scientist* 74: 155-162.
- Dolby, G.R. 1982. The role of statistics in the methodology of the life sciences. *Biometrics* 38: 1069-1083.
- Ferguson, A. 1976. Can evolutionary theory predict? *American Naturalist* 110: 1101-1104.
- Feyerabend, P. 1975. *Against Method*. (Verso: London).
- Finch, P.D. 1979. Description and analogy in the practice of statistics. *Biometrika* 66: 195-208.

- Hilborn, R. and S. Stearns. 1982. On inference in ecology and evolutionary biology; the problem of multiple causes. *Acta Biotheoretica* 31: 145-164.
- Holton, G. 1975. On the role of themata in scientific thought. *Science* 188: 328-338.
- Holton, G. 1978. *The Scientific Imagination: Case Studies*. (Cambridge Univ. Press: Cambridge).
- James, F.C. and C.E. McCulloch. 1985. Data analysis and the design of experiments in ornithology. In, *Current Ornithology*, R.F. Johnston (ed.) Vol. 2, pp. 1-63. Plenum Press.
- Kruskal, W.H. 1978. Significance, tests, of p. 944-958. In, W.H. Kruskal and J.M Tanuar (eds.) *International Encyclopedia of Statistics*. MacMillian: New York.
- Kuhn, T.S. 1970. *The Structure of Scientific Revolutions*. (Univ. Chicago Press: Chicago), 2nd. ed.
- Kuhn, T.S. 1977. *The Essential Tension*. (Univ. Chicago Press: Chicago).
- Lakatos, I. and A. Musgrave (eds.) 1970. *Criticism and the Growth of Knowledge*. (Cambridge Univ. Press: Cambridge).
- Magee, B. 1974. Karl Popper: The Useful Philosopher. *Horizon* 16(5): 52-57.
- Merton, R.K. 1973. *The Sociology of Science*. (Univ. Chicago Press: Chicago).
- Platt, J.R. 1964. Strong Inference. *Science* 146: 347-353.
- Popper, K.R. 1959. *The Logic of Scientific Discovery*. (Harper and Row: New York).
- Popper, K.R. 1963. *Conjectures and Refutations*. (Harper and Row: New York).
- Popper, K.R. 1972. *Objective Knowledge*. (Clarendon Press: Oxford).
- Quinn, J.F. and A.E. Dunham. 1983. On hypothesis testing in ecology and evolution. *American Naturalist* 122: 602-617.
- Selvin, H.C. and A. Stuart. 1966. Data-dredging procedures in survey analysis. *American Statistician* 320: 20-23.
- Strong, D.R., L.A. Syszka, and D.S. Simberloff. 1979. Tests of community wide character displacement against null hypotheses. *Evolution* 33: 897-913.
- Strong, D.R. 1980. Null hypotheses in ecology. *Synthese* 43: 271-285.