13

Publishing Your Results

McKinney et al. [1989] report that more than half the published articles that apply Fisher's exact test do so improperly. Our own survey of some 50 biological and medical journals supports their findings. This chapter provides you with a positive prescription for the successful publication of the results of testing procedures. First, we consider the rules you must follow to ensure that your data can be analyzed by statistical methods. Then, we provide you with a number of simple rules to prepare your report for publication.

13.1 Design Methodology

It's never too late to recheck your design methodology. Recheck it now in the privacy of your office rather than before a large and critical audience. All our testing methods rely on the independence and/or the exchangeability of the observations. Were your observations. independent of one another? What was the experimental unit? Were your subjects/plots assigned at random to treatment? If not, how was randomization restricted? With complex multi-factor experiments, you need to list the blocking variables and describe your randomization scheme.

13.1.1 Randomization in Assignment

Are we ever really justified in exchanging labels among observations? Consider an experiment in which we give six different animals exactly the same treatment. Because of inherent differences among the animals, we end up with six different measurements, some large, some small, some in between. Suppose we arbitrarily label the first three measurements as "controls" and the last three as "treatment." These arbitrary labels are exchangeable and thus the probability is 1 in 20, that the three "control" observations will all be smaller than the three "treatment." Now suppose we repeat the experiment, only this time

we give three of the animals an experimental drug and three a saline solution. To be sure of getting a positive result, we give the experimental drug to those animals who got the three highest scores in the first experiment. Not fair, you say. Illegal! Illegitimate! No one would ever do this in practice.

In the very first set of clinical data I received for statistical analysis was brought by a young surgeon. He described the problems he was having with his chief of surgery. "I've developed a new method for giving arteriograms, which I feel can cut down on the necessity for repeated amputations. But my chief will only let me try out the technique on patients that he feels are hopeless. Will this affect my results?" It would and it did. Patients examined by the new method had a very poor recovery rate. But, of course, the only patients who'd been examined by the new method were those with a poor prognosis. The young surgeon realized that he would not be able to test his theory until he was able to assign patients to treatment at random.

Not incidentally, it took us three more tries until we got this particular experiment right. In our next attempt, the chief of surgery—Mark Craig of St. Eligius in Boston—announced that he would do the "random" assignments. He finally was persuaded to let me make the assignment using a table of random numbers. But then he announced that he, and not the younger surgeon, would perform the operations on the patients examined by the traditional method to make sure "they were done right." Of course, this turned a comparison of methods into a comparison of surgeons and intent.

In the end, we were able to create the ideal "double blind" study: The young surgeon performed all the operations, but the incision points were determined by his chief after examining one or the other of the two types of arteriogram.

13.1.2 Choosing the Experimental Unit

The exchangeability of the observations is a sufficient condition for a permutation test to be exact. It is also a necessary condition for the application of any statistical test. Suppose you were to study several pregnant animals that had been inadvertently exposed to radiation (or acid rain or some other undesirable pollutant) and examine their offspring for birth defects. Let $X_{ij}, i = 1, \ldots, I; j = 1, \ldots, J$, denote the number of defects in the jth offspring of the ith parent; let $Y_i = \sum_j X_{ij}, i = 1, \ldots, I$ denote the number of defects in the ith litter. The $\{Y_i\}$ *may* be exchangeable (we would have to know more about how the data were collected to determine this). The $\{X_{ij}\}$ are not: The observations within a litter are interdependent; what affects a parent affects all her offspring. In this experiment, the litter is the correct experimental unit.

The viewpoints of the observer and the statistician can be quite different. If we wear two hats—serving both as observer and statistician, recognition of this distinction can be painful. For example, in a typical toxicology study a