Discussion of the paper, Teaching Hypothesis Testing, Can It Still Be Useful by Henrik Dahl

Michael F. Capobianco

St. John's University, New York

300 Howard Ave.

Staten Island, NY 10301 USA

capobiam@stjohns.edu

The point is made in Section 3 that "... we should concentrate not just on the mathematical formulae, but making clear the philosophy behind the methods."

I couldn't agree more. This is one reason why I have always advocated beginning the teaching of statistical testing with nonparametric tests. In these tests, the mathematics is generally so simple that it doesn't obscure the students' ability to see and understand the underlying principles and concepts such as those mentioned in this paper namely, model, parameter, test statistic, sampling distribution, null and alternative hypotheses, and critical region.

The example chosen: "How many excellent grades should be tolerated", is a good one, I believe, for two reasons. First, students easily relate to it, and second it makes use of a relatively simple sampling distribution, the binomial. I was happy to see that the concept of the power of a test was used here in the process of comparing different critical regions and sample sizes. I much prefer the use of *power* rather than type II error since it is based on the critical region as is type I error.

The question raised regarding the equivalence (or non-equivalence) of the two decisions, "ridiculing" and "freeing a culprit" is a most interesting one, and really should be handled by the use of a Loss Function. Prof. Dahl mentions the application to medical diagnosis, an area I personally studied some time ago. One might consider the situation in the following way: A patient is to be tested for a serious disease. Either she has the disease, or she doesn't. These are the "states of nature", only two. There are also only two decisons possible upon observation of the patient's symptons namely, "she has the disease" , "she doesn't have the disease". Hence a Loss Function could be set up as a 2 by 2 matrix.

The entries would be the losses incurred under the various combinations of states and decisions.

The function should reflect the fact that deciding that the patient doesn't have the disease when in reality she does is much more serious than deciding that she has the disease when in reality she doesn't. An expected loss can now be computed using the probabilities of Type I and Type II errors, and the merits of the test judged on the basis of this value.

In Section 5, Prof. Dahl states: "If the data set is small, non-rejection may be due to insufficient power." This is, of course, correct. However, I should like to point out that by the same token, <u>rejection</u> with a small sample yields a certain assurance of a correct result.

I should like to conclude with my main objection to hypothesis testing which is not generally taught in classes, and ought to be. Let's suppose that we reject the null hypothesis. Hooray! We feel we know something, or at least we know something up to some small probability of being wrong. But what is it we know? Suppose we had been testing that the population mean was zero. We get a rejection; it is not 0. But how much different from 0 is it. The difference is "statistically significant", but is it "practically significant"? Is the difference so small that in practical applicaion in the field in question, it really makes no difference? Of course, we can compute the size of the difference by means of a confience interval, and I much prefer interval estimation to hypothesis testing. However, we are then faced with the dificulty of deciding just what the confidence interval means. Suppose that we compute that the mean is between 1 and 3 with 95% confidence. What does this mean? It means one and only one thing namely, that in order to obtain this interval we have used a procedure which, in the long run, will yield an interval which contains the true value of the parameter (the mean) 95% of the time. What does that tell us about whether or not our interval, 1 - 3 , contains the true mean? NOTHING!