Johnson & Johnson

file Battelle Deposition Study

New Brunswick, N.J.
January 12, 1977
File #157

Subject: Dr. Wehner's Manuscripts

TO: Dr. B. Semple

In my earlier letter to Mr. Buschbom, I pointed out several suggestions to improve the presentation of the original paper. For the paper on the "Leaching of Radionuclides" I offer some additional suggestions which would standardize the statistical terminology closer to its current usage in the statistical literature. Some of these could also apply to the original paper.

The suggestions which follow are based largely on a distinction between the unknown constant parameters of a mathematical distribution and the estimates of these parameters which are calculated from the data and which vary because of sampling variability and possibly other causes.

1. Current statistical practice usually restricts the use of Greek characters such as μ or σ to represent the unknown parameters of a distribution and Roman characters such as $\bar{x}$ and s to represent the estimates of the parameters which are computed from experimental data. Many writers also use modified Greek characters such as $\hat{\mu}$ and $\hat{\sigma}$ (mu-hat or sigma-hat) to represent estimates of parameters.

2. In the past the terms "mean" and "average" were frequently used interchangeably and to some extent this confusion still persists in the literature. The best current usage, however, restricts the term "mean" to represent the unknown parameter of a distribution (actually the first moment of a distribution) and the term, "average" to the estimate of the parameter calculated from data.

3. In the case of the parameter, "standard deviation", (actually the square root of the second moment about the mean) there is no recognized simple word to represent the estimate calculated from data; therefore, the most acceptable terms are the awkward expressions, "estimate of the standard deviation" or the "sample standard deviation". By consistently using s or $\hat{\sigma}$ to represent the estimate, one can get around the awkward expressions.

4. In the paper, averages are frequently shown along with "$\pm\sigma$" or "$\pm 2\sigma$" to demonstrate the variability in the data. Using the notation described above, these should be "$\pm$s" or "$\pm$2s" if the s represents the "estimate of the standard deviation" of individual observations. On the other hand, if the authors have computed the "standard error of the average" then

#4 cont.

this should be so stated or the correct notation should be employed, as follows:

$$s_{\bar{x}_n} = \frac{s}{\sqrt{n}} \quad , \text{ where}$$

s = the estimate of the standard deivation of individual observations.

$s_{\bar{x}_n}$ = the standard error of the average or the estimate of the standard deviation of averages of size n.

n = number of observations used to compute the average.

Some authors use $s_{\bar{x}_n}$ instead of s because the former is much smaller and the presentation thereby "looks better". Whichever term is used, it should be so stated.

J. N. Sivertson

/jdm

cc Mr. G. E. Heinze
Dr. G. Hildick-Smith
Mr. G. Lee
Dr. G. Lord
Dr. R. Rolle
Dr. W. T. Sherman