THE COMPARISON OF DIFFERENT SCALES OF MEASUREMENT FOR EXPERIMENTAL RESULTS[1,2]

By W. G. Cochran

Iowa State College

1. Introduction. In some fields of research, the development of a satisfactory method for measuring the effects of experimental treatments constitutes a difficult problem. The estimation of the vitamin content of preparations of foods furnishes a good example; for most of the vitamins several years of work were required to construct a reliable method of assay. In other cases, where the ideal method for measuring treatment responses is costly or troublesome, a search may be made for a more convenient substitute. Thus in pasture or forage-crop experiments the species composition of a plot may be estimated by eye inspection as a substitute for a complete botanical separation. As a third example we may quote experiments in cookery, where the flavor and quality of the dishes are subject to the whims of human taste. Frequently a panel of judges is employed, each of whom scores the dishes independently. It is not easy to determine how the panel should be chosen, nor how representative its verdicts are of consumer preferences in general.

When such problems are investigated, experiments may be carried out specifically for the purpose of comparing two or more methods or *scales* of measurement. Where the process of measurement affects only the final stages of the experiment, as in the last two examples quoted above, all that is necessary is to score the *same* experiment by the various scales under consideration. In comparing two different methods of assaying vitamins, on the other hand, independent experiments are frequently required, the only common feature being that the same set of treatments is tested in both experiments.

In the interpretation of the results of such experiments, two types of comparison are of general interest. One concerns the *relations* between the scales. It may be summed up rather loosely in the question: Are the effects of the treatments the same in all scales? For a more exact formulation, consider the case of two scales, which is probably the most frequent in practice. Let ξ_{1t}, ξ_{2t} be the true means of the tth treatment as measured on the two scales. We may wish to examine the following hypotheses:

(i) *Scales equivalent:*

$$(1) \qquad \xi_{1t} = \xi_{2t}, \qquad \text{(all } t\text{)};$$

(ii) *Scales equivalent, apart from a constant difference:*

$$(2) \qquad \xi_{1t} = \xi_{2t} + \epsilon, \qquad \text{(all } t\text{)};$$

---

[1] Paper presented at a meeting of the Institute of Mathematical Statistics, Washington, D. C., June 18, 1943.

[2] Journal Paper No. J-1136 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project 514.