be encouraged to begin the study of statistics with the course having calculus as a prerequisite, and should not be put into the necessarily slower group not having the calculus. I believe that these elementary courses should begin with the theory of probability, but should go on to the chief distribution functions used in practice, and should include applied problems and work on calculating machines.

Putting a sound program of statistical teaching into effect will take time, partly because of the scarcity of suitable teachers of statistics. Nevertheless, the process is well under way, and the prospects are good for substantial improvements in the teaching of statistics. A body of able young research men possessing the requisite knowledge of statistical fundamentals is now in existence and is growing. Some of the recent textbooks represent striking improvements. The Institute of Mathematical Statistics itself, with the *Annals of Mathematical Statistics*, is perhaps the best evidence of a changed view making for better things.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

DISCUSSION OF PROFESSOR HOTELLING'S PAPER

BY W. EDWARDS DEMING

It is a pleasure to endorse Professor Hotelling's recommendations; in fact we have been following them pretty closely in the courses in the Graduate School of the Department of Agriculture. As a matter of fact, he has indirectly played an influential part in building up this set of courses, because some of our best instructors are his former students.

Listening to Professor Hotelling's paper, I was thinking of the possibility that some of his recommendations might be misunderstood. I take it that they are not supposed to embody all that there is in the teaching of statistics, because there are many other neglected phases that ought to be stressed. In the Bureau of the Census the population division alone has augmented its force by approximately 3500 statistical clerks during the past six months. They come from diverse schools and it has been interesting to observe how many of them have the idea that all the problems of sampling and inference from data can be solved by what are commonly known as modern statistical techniques—correlation coefficients, rank correlation coefficients, chi-square, analysis of variance, confidence limits, and the like. Most of them are shocked to learn that many of the so-called modern "theories of estimation" are not theories of estimation at all, but are rather theories of distribution and are a disappointment to one who is faced with the necessity of making a prediction from his data, i.e., of basing some critical course of action on them. The conviction that such devices as confidence limits and Student's t provide a basis for action regardless of the size of the sample whence they were computed, even under conditions of statistical control, is too common a fallacy. On the other hand, many simple but worthy devices are neglected. A histogram, for instance, can be a genuine tool of prediction if it is built up layer by layer in different legends so as to distinguish the different sources whence the data are derived. The modern student, and too often his teacher, overlook the fact that such a simple thing as a scatter diagram is a more important tool of prediction than the correlation coefficient, especially if the points are labeled so as to distinguish the different sources of the data. Most students do not realize that for purposes of prediction the consistency or lack of it between many small samples may be much more valuable than any probability calculations that can be made from them or from the entire lot. Students are not usually admonished against grouping data from heterogeneous sources. Of those that are not guilty of indiscriminate grouping, many are inclined to rely on statistical tests for distinguishing heterogeneity, rather than on a careful consideration of the sources of the data. Too little attention is given to the need for statistical control, or to put it more pertinently, since statistical control (randomness) is so rarely found, too little attention is given to the interpretation of data that arise from conditions not in statistical control.

Nevertheless, the fundamentals of probability and sampling theory, and the mathematics of the distribution functions, though by themselves they do not qualify anyone for high-grade statistical work, are ultimately essential for proficiency in statistics. Since they are seldom learned away from the university they are properly made the main theme of teaching. The university is the place to learn the studies that are so difficult to get outside of it.

Above all, a statistician must be a scientist. The skepticism of many first class scientists of today for modern statistical methods should be a challenge to statistical teaching. A scientist does not neglect any pertinent information, yet students of statistics are often taught to do just the opposite of this, and are accused of being old-fashioned for daring to think of combining experience with the new information provided by a sample, even if it is a pitifully small one. Statisticians must be trained to do more than to feed numbers into the mill and grind out probabilities; they must look carefully at the data, and take account of the conditions under which each observation arises. It is my feeling that the chief duty of a statistician is to help design experiments in such a way that they provide the maximum knowledge for purposes of prediction; another is to compile data with the same object in view; and still a third function is to help bring about some changes in the source of the data. Scientific data are not taken merely for inventory purposes. There is no use taking data if you don't intend to do something about the sources whence they arise.

Bureau of the Census,
Washington