Dear Sir,

I have been looking at the examination papers in Statistics for 1925, primarily with an idea of gaining a knowledge of the degree of difficulty which is customary. My first impression is that quite a large proportion of questions involve very great difficulties, and I am at some loss to know in what manner the student was expected to attack them. Perhaps I may make myself more clear by a concrete instance.

Questions 2 and 5 in the Honours School for Commerce (paper I) refer to the age distribution of deaths from cancer; at the end of question 5 is the supplementary question:

"Do the results differ significantly from those calculated directly?"

The first answer that suggests itself is:

"It is impossible that they should for both sets of results are derived from the same sample, and to differ significantly would be to indicate that they were derived from samples of different populations."

If this is to misunderstand the question, I suppose we should understand it thus:

Two estimates (e.g., of the mean) have been made by different methods from the same sample, if this sample were normal, as is assumed in 5, the discrepancy between the two estimates will be distributed about zero in calculable manner, use the actual discrepancies observed to test this hypothesis of normality.

If this is right the question seems to be one of great
difficulty. The students will doubtless be able to calculate the standard error of the estimate in question 2, once they have disposed of the question of the two deaths at specified age (which from question 5 are evidently to be cut among the deaths over 70), and of the distribution of deaths below 15; perhaps, and this is a point I should like to know about, they could handle the sampling errors in the method of 5. If I am not mistaken we have for the mean

\[
\sigma^2_m = \frac{\sigma^2}{m} \left\{ x^2 - \frac{\beta^2}{x^2} - 2x' x \frac{\beta'}{x} + x^2 \frac{\beta'(-\beta)}{x^2} \right\} \div (x - x')^2
\]

where \( \beta \) and \( \beta' \) are the proportions below 45 and above 70, \( x \) and \( x' \) the corresponding normal deviates, one of them negative, and \( x \) and \( x' \) the corresponding ordinates. This seems pretty tough, though they may be familiar by practice with this line of work. How are they to allow for the correlation in random samples between the two estimates? Of course, this can be done, for as a matter of fact the mean, even from grouped data is so near to being efficient that we can in practice subtract the variances of the two estimates to obtain the variance of their difference. To know this, however, the student would have to have gone a good way beyond any text book.

I enclose an attempt at the problem, so that you can see more exactly what I mean. I take it that the students will have been trained in a number of correct methods appropriate to certain types of statistical data, and that the purpose of the examination is to test

(a) primarily, if a particular student can recognise which method is appropriate and carry it through correctly, and

(b) secondarily, if he can recognise and discuss, but not necessarily solve, difficulties other than those for which he has been taught the correct procedure.

(c) perhaps, I ought to add in the primary section that he should understand the meanings attached to statistical terms, and be able to give a reasoned justification of the processes employed.
The main trouble, I suppose, is that there are so many definitely incorrect methods in the literature that it is difficult to keep them out of (a), and (c) becomes exceedingly difficult, being involved in disputed points of statistical theory.

Perhaps my best plan is to ask you for a syllabus, somewhat detailed in respect to the type of numerical exercises example, and of the text book material which the students should have covered. (I do not suppose that my own book has been available for this year's students). With this in hand I feel I shall be more able to cooperate usefully in the examination.

(Sgd) R.A. Fisher.