HOW TO DETERMINE WHICH OF TWO VARIABLES IS BETTER FOR PREDICTING A THIRD VARIABLE

WILLIAM DOWELL BATEN

A VERY important problem arising in certain agricultural experiments is the determination of which one of two variables is better for predicting a third variable. For example, measurements are made on the lengths of heads of wheat and counts on the number of spikelets and the number of kernels per head, and it is desired to know whether the number of kernels per head can better be predicted on the average from the length of the heads than from the number of spikelets per head. Can a steer's weight be better predicted on the average from his heart girth than from some other body measure? Can the area of a bean leaflet be better predicted from its length than from its width?

The natural way to go at solving this problem is to test for significance between the standard errors of estimate arising from the errors from the two predicting values. Assume that predicting linear equations have been found for predicting the number of kernels per head of wheat from the length of the head and from the number of spikelets per head and that these equations together with the standard errors of estimate are

\[
\hat{y} = \bar{y} + b(x - \bar{x}) \quad ; \quad \sigma_e = \sigma_y \sqrt{1 - r^2_{yx}}
\]

and

\[
\hat{y} = \bar{y} + d(z - \bar{z}) \quad ; \quad \sigma_e = \sigma_y \sqrt{1 - r^2_{yz}}
\]

where \(\hat{y}, \bar{y}, x,\) and \(z\) represent, respectively, the predicted number of kernels per head, mean number of kernels, length of the head, and the number of spikelets per head, and \(r_{yx}\) and \(r_{yz}\) the respective correlation coefficients. The difficulty in testing the two standard errors of estimate arose from the fact that they were correlated in a peculiar manner. Hotelling solved the problem for the first time and gave a test for testing significance. Since this test is so recent (November 1940) and might be overlooked for several years by agriculturists not interested in derivations of formulas, and since it has been desired by many research workers, it seems fitting to present several applications of this helpful test and bring it to the attention of those interested.

APPLICATIONS

The first application pertains to determining whether or not the length of a head of wheat is better for predicting on the average the number of kernels per head than the number of spikelets with grain

---

1Contribution from the Department of Mathematics, Michigan State College, East Lansing, Mich. Journal Article No. 512 (n.s.), Michigan Agricultural Experiment Station. Received for publication March 8, 1941.

2Research Associate in Statistics.