25 April 1933.

H.J. Buchanan Wollaston, Esq.,
Fisheries Laboratory,
LOWESTOFT.

My dear Buchanan Wollaston:

Thanks for your letter of 21 April 1933, which I have just received on my return from Plymouth.

If there were data for two previous years, and this kind of thing fairly often happens, one can either (1) combine the two independent variates into one as by adding them together if this combination seems undoubtedly more appropriate than any other; or, (2) better, but more laborious, treat the two or more sets of data for what are going to be used as independent variates, quite independently, doing an analysis of variance of each and an analysis of covariance of each pair, and of each with the dependent variate; and then use the Sum of Squares and products from the error lines of each table to form equations for the multiple regression of the dependent on the independent variates.

No, the covariance as estimated from any given number of degrees of freedom is not distributed in the same way as is a
variance. In fact, in different samples it may sometimes be positive and sometimes negative, so that one never takes its log to form a z from.

On the question of the reality of plankton differences at different stations I think I should try in the first instance whether the mean square differences between samples taken at 24 hours apart at different stations is or is not significantly different from the mean square between samples taken 24 hours apart at the same place. I cannot myself doubt that such an effect exists but if it is called in question and needs demonstrating, I think this would demonstrate it.

If the probability integral of a distribution has been tabulated so that you can read off the value of a variate from the percentage frequency above or below it, then a handy method of sampling is to take Tippett's Random Sampling Numbers as giving these percentages and reading the variate value of the sample individual from them. With respect to the distinction of modes one might say that practically nothing is known because "observed modes" really only appear when the data are grouped and what modes you can get will depend on the size and precision of the grouping intervals. As a working test of bimodality it might I suppose be worth trying to take the three highest group frequencies and use twice the middle one minus the
other two as a sort of measure of convexity. With some material perhaps the five highest frequencies with a similar measure would do better. Such measures of convexity might, I think, be expected to be distributed rather differently in unimodal and bimodal distributions and might perhaps provide a criterion. To establish one would require, I suppose, rather ample experimental samples from distributions of these two types similar in other respects to the material you are studying.

Yours sincerely,