

December 20, 1938

Dear Betty,

I am very glad to hear your news of the Statistical Tables and of the reviews which you and Cornish are doing in Australian Journals. It will be immensely helpful to have them reviewed by people who understand statistical methods. Many thanks also for the note on middle 6. I must have this checked up.

I have frequently been confronted with data such as you describe, in which there is no genuine basis for a test of significance at all. One can usually find some basis which is probably, though not certainly, not misleading. For example, it may appear that, in fitting a growth curve to the means - or, if suitable, to their logarithms - the mean square residual ceases to fall after a quadratic or cubic has been fitted, so that one has an estimated error variance for the mean of a number of animals at any one date. There is certainly no validity in applying this to compare the means of two different groups of animals, since, conceivably, for example, each animal might follow a cubic as closely as one pleased, ^{and} yet different animals follow
^

different cubics. Still, using data in which individual weights have been recorded, one can see whether the ~~one~~ ^{first} estimate is, or is not, appreciably greater than the true one. I do think, however, that you will have to insist on individual weights being ordinarily provided.

On the question of soil temperature, as you know there is quite an effective theory connecting the sequence of surface temperatures with the sequence at any depth. On reasonably homogeneous soils the formula

$$e^{-\frac{px}{v}} \cos p\left(t - \frac{x}{v}\right)$$

gives the amplitude and phase of a harmonic wave with period $\frac{2}{p}$. v is the velocity of penetration into the soil, and x the depth. For waves of different periods v is proportional to the square root of p , the constant proportionality depending on the conductivity and heat capacity of the soil. One should be able, therefore, to express the temperature at any depth and time in terms of the surface temperature up to that time. This theory seems to bear a little on the regression procedure you mention, in the sense that predictions will be systematically too high or too low according to the time of year, and also ^{to} the annual variation at different places, both these disturbances being due to the annual wave of temperature. For example, in England the temperature at

at any depth, following a given sequence of a months surface temperature, will be higher in Autumn than in Spring. Again, given the preceding series of surface temperatures during, let us say, March, the temperature at any depth would be higher in England than in Russia, because the annual heat wave is more important in Russia than in England.

I should, therefore, if I were tackling the problem, use the previous readings, grouped in such a way as your man has done, not to predict the absolute temperature, but to predict deviations from the normal temperature at that depth and at that time of year, for I have no doubt that the amplitude of the annual thermal oscillation is greater at continental than at coastal stations.

I have no doubt, however, that you have already compared the data with a view to detecting systematic errors. If these are not giving trouble, I should certainly use the residual mean square to measure the precision of the formula.

With respect to Bliss, I suppose the point you have in mind is that, when χ^2 is satisfactory, he prefers to use the theoretical precision based on the number of observations to the empirical estimate based on the mutual agreement of the different points on his curve. In this I think he is certainly right. On the other hand, there are many causes which may disturb the value of χ^2 , such as heterogeneity in the material tested, or irregularity in

the true form of the response curve. Hence, when X^2 is excessive, one cannot safely use the precision appropriate to a given number of observations. It is easy to understand why the test of significance based on the empirical estimate often fails to show significance, when it is indicated by the simpler formula, for there are usually only a few degrees of freedom available for estimating the precision, and an appropriate allowance for the error of estimation makes one require considerably ^{higher} error ratios before judging them to be significant.

With best wishes for Christmas and the New Year,

Yours sincerely,