January 24,
1941

My dear Fox,

I handed Dr Johnson's paper and Peierl's Appendix for criticism to W.L.Stevens of this Department, and he promptly ~~replied~~ reached with the enclosed. As what he says in this note may not give Dr Johnson all the help she needs, I enclose also a detailed statement of the calculations necessary, which she will, I think, have no difficulty in carrying out. If she is in any doubt about this, I shall, of course, be glad to give her a hand later.

Yours sincerely,

The Respiratory Function of the Haemoglobin of the Earthworm

Criticism of Statistical Methods Employed.

The method proposed is extremely clumsy and does not make the best use of the data. It is, therefore, unnecessary to consider in any detail whether or not the formulae derived are ~~theretect~~ theoretically correct. An idea of the quality of these formulae may however be judged from the author's approximation that a and c are uncorrelated, though the fact that c = b/a shows that not only are they correlated, but also that the degree of correlation is deducible.

We note also that the author allows for weight of worms by simply dividing by the weight, although he himself shows elsewhere that this is not the correct allowance (footnote p.5).

Similarly he allows for initial oxygen consumption by assuming that the regression of log oxygen consumption second hour on first hour is unity, in spite of the facts that there is no reason to believe that it is, and that the data would provide evidence, which he entirely ignores, as to the real value of this regression coefficient.

Furthermore he apparently estimates separately the error for each experimental series, although a much more accurate estimate is obtained by pooling the squared residuals.

It is accordingly recommended that the present statistical treatment be scrapped, and instead the data be treated, by the method of the Analysis of Covariance, taking as ~~independent variate the logarithm consumption first hour, and log weight~~.

~~P.T.O.~~

dependent variate the logarithm of the oxygen consumption in the second hour and as independent variates the logarithm consumption first hour, and log weight.

The regression equation would then be formed from the sums of squares and products for the 83 degrees of freedom for error

y stands for log oxygen consumption in second hour } for one worm-pair
x_1 " " log weight } for one worm-pair
x_2 " " log oxygen consumption in the first hour } for one worm-pair

The expected value of y for given values of x_1 and x_2 is then expressed in the form

$$Y = a + b_1 x_1 + b_2 x_2$$

wherein $\underline{a}$ is expected to depend on the partial pressure of oxygen, and on C.O. treatment.

For each set of n' experiments (there being ten such sets) agreeing in these respects calculate

Totals	$S(x_1)$	$S(x_2)$	$S(y)$
Mean	$\bar{x}$	$\bar{x}_2$	$\bar{y}$
S. ~~Mean~~ Squares	$(p) = S(x_1^2) - \frac{1}{n'}\{S(x_1)\}^2$	$(q) = S(x_2^2) - \frac{1}{n'}\{S(x_2)\}^2$	$(w) = S(y^2) - \frac{1}{n'}\{S(y)\}^2$
S. ~~Mean~~ products	$(r) = S(x_1 x_2) - \frac{1}{n'}S(x_1)S(x_2)$	$(l) = S(x_1 y) - \frac{1}{n'}S(x_1)S(y)$	$(m) = S(x_2 y) - \frac{1}{n'}S(x_2)S(y)$

Add the values of p from the 10 sets of experiments to a total P

Do the same with q, r, l, m, w

Solve the equation

$$Pb_1 + Rb_2 = L$$
$$Rb_1 + Qb_2 = M$$

The values of b_1 and b_2 are thus derived from the data as a whole. The sum of squares of y not accounted for by these two factors is

$$W - b_1 L - b_2 M$$

To this corresponds 81 degrees of freedom, i.e., this is the sum of the squares of all quantities $(y - b_1 x_1 - b_2 x_2\cdot)$ Dividing by the d.f. gives the mean square

$$s^2 = \frac{W - b_1L - b_2M}{81}$$

In applying the correction x_1 and x_2 are measured from any conveniently chosen central or standard values, which must be the same for all ten series. For the mean $\bar{y}$ of any set, subtract

$$b_1\bar{x}_1 + b_2\bar{x}_2$$

where $\bar{x}_1$ and $\bar{x}_2$ are the corresponding means in this set.

Then the sampling variance (squared standard error) of the mean of 9 corrected sets of the form $y - b_1x_1 - b_2x_2$ is $\frac{1}{9}s^2$, that of a different set of ~~xxxx~~ 8 is $\frac{1}{8}s^2$, and that of the difference between the two corrected means is $(\frac{1}{8} + \frac{1}{9})s^2$. Significance can be judged by comparing differe ces with the standard error thus calculated.

I think this will give you the tests of significance you want. Section 49.1 of my book gives the test of significance with a further refinement which I imagine your data will not need. It might be as well for Miss Johnson to look up this chapter.