112

APPENDIX TO A PAPER BY H.G.THORNTON AND P.H.H.GRAY ON THE NUMBERS OF BACTERIAL CELLS IN FIELD SOILS*

* Thornton, H.G. and Gray, P.H.H. (1934) The numbers of bacterial cells in field soils, as estimated by the ratio method. *Proceedings of the Royal Society of London*, B, 115: 522-540.

Appendix

By Professor R. A. Fisher, Sc.D., F.R.S.

It is important to avoid confusion between the two types of test for which χ^2 is used in this paper. If bacteria only are counted we have a single series of numbers representing the counts obtained from n' different fields, *e.g.*,

$$x_1, x_2, \ldots, x_{n'}.$$

If the different fields contained equal quantities of soil, in which after thorough mixture the bacteria were dispersed at random, it is known (Fisher, 1932, sect. 16) that the number x should constitute a sample of n' from Poisson series, such that the probability of counting just x organisms is

$$e^{-m}\frac{m^x}{x!};$$

where m is the average number in all fields. When m is not unreasonably small, *i.e.*, if it exceeds about 5 units, the homogeneity of the different fields may be easily tested by calculating what is known as the index of dispersion, defined as

$$\chi^2 = \frac{s\,(x-\bar{x})^2}{\bar{x}},$$

with $n'-1$ degrees of freedom; for, on the hypothesis of homogeneity, this index will be distributed in a well-known distribution, independent of the value of m. Clearly when the values of x vary greatly χ^2 will be high, and since its distribution is known and readily available it is easy to see if its value is too high to admit of the hypothesis of homogeneity.

When both bacteria and indigo particles are counted, each field will provide two numbers, and we are concerned with the stability not of either number in itself, but of the ratio between them. This is equivalent to testing the proportionality of the entries in a $2 \times n'$ table of frequencies, for which the table of χ^2 may equally be used. If b and i stand for the numbers of bacteria and indigo particles counted in any field, and if B and I are corresponding totals for all fields, χ^2 may be written in the form (Fisher, 1932, sect. 21)

$$\frac{1}{BI}\,S\left\{\frac{(bI-iB)^2}{b+i}\right\},$$

where the degrees of freedom are, as before, one less than the number of fields. Thus the same distribution may be used for testing the homogeneity of the ratios as for testing that of the absolute numbers.

Where homogeneity exists the importance of obtaining clear evidence of it is twofold. In the first place, the fact that the actual discrepancies between different fields, drops, or slides are of the same magnitude as those which would arise by pure chance in sampling homogeneous material affords a guarantee that the technique employed has attained the greatest possible precision subject only to the limitation imposed by the number of fields counted. In the second place, when the precision has been thus raised to the highest possible level it depends only on the extent to which the material is counted, and can therefore be inferred from the total numbers. This is a considerable convenience in estimates of standard error, intended for use in tests of significance.

If p is the ratio $B/(I+B)$, then it is well known that, with homogeneous material, the variance of p is given by

$$V(p) = \frac{BI}{(I+B)^3};$$

from this, since I and B are both large, we may readily calculate the variance of the ratio B/I, and what is more important, that of its logarithm. For if $r = B/I$, then for small variations of p and r, dp and dr

$$\frac{dp}{p^2} = \frac{dr}{r^2},$$

whence

$$V(r) = \frac{r^4}{p^4} V(p) = \frac{(I+B)^4}{I^4} \cdot \frac{BI}{(I+B)^3} = \frac{B(I+B)}{I^3}.$$

Again if $Z = \log r$

$$dZ = \frac{1}{r} dr,$$

whence

$$V(Z) = \frac{1}{r^2} V(r) = \frac{I^2}{B^2} \cdot \frac{B(I+B)}{I^3} = \frac{I+B}{IB} = \frac{1}{I} + \frac{1}{B}.$$

To find the sampling variance of the natural logarithm of the ratio B/I, it is therefore only necessary to take the sum of the reciprocals of the total numbers of bacteria and indigo particles counted. This rapid and convenient test could not be relied on unless, as is shown by the χ^2 tests, practically all the existing variation is ascribable to purely random errors.

When this is not so, the values of Z derived from different drops will not agree so closely as is theoretically possible, and an empirical standard error based on actual deviations between values of Z for parallel drops should be used in tests of significance.

REFERENCE

Fisher, R. A. (1932). "Statistical Methods for Research Workers" (Oliver and Boyd).