23

STATISTICAL APPENDIX TO A PAPER BY J. DAVIDSON ON BIOLOGICAL STUDIES OF *APHIS RUMICIS**

* Davidson, J. (1922) Biological studies of *Aphis rumicis*: reproduction on varieties of *Vicia faba*. *Annals of Applied Biology*, 9: 135-142.

(Tables I and II from Davidson's paper are reprinted here at the end of the appendix by Fisher.)

APPENDIX

STATISTICAL CONSIDERATIONS INVOLVED IN TABLES I AND II OF THE ABOVE PAPER.

BY R. A. FISHER, M.A.

Fellow of Gonville and Caius College; Statistician, Rothamsted Experimental Station, Harpenden.

The discussion of the probable error to be attached to Dr Davidson's aphis infestation numbers, involves points of statistical interest, which have hitherto not, in print at least, received adequate treatment.

 Annals of Applied Biology, 9: 142-145, (1922).

Since only five infections were made on each of the varieties tested it is clear that only the roughest estimate of the standard deviation can be based upon the data for a single variety. For a standard deviation estimated from a sample of five is not only subject to very large errors of sampling, but is distributed in a markedly skew manner; on the other hand, the causes of variation, whatever they may be, must be closely analogous in all the varieties tested; and this fact should enable us to make use of the information supplied by the whole of the material, to estimate with some accuracy the probable error to be ascribed to each value.

The process of obtaining a single probable error from the deviations of a number of distinct groups has been applied successfully in cases where it may be assumed that the groups are equally variable; as is the case when a correlation ratio is determined on the assumption of the equal variability of the arrays. In the case of the infestation numbers no such assumption is *a priori* plausible, for setting aside the lowest mean which is evidently exceptional, the 18 means range from 286 to 1037. Moreover, an inspection of the figures for the individual plants shows that the higher numbers are, as is to be expected, actually the more variable. There exists, however, a class of distribution, of which the Poisson Series is the classical example, in which the variance is proportional to the mean.

To test whether this is the case with the infestation numbers, the quantities

$$\frac{n\mu_2}{\bar{x}}$$

were calculated for each variety, where $\bar{x}$ is the mean and μ_2 the second moment of each sample; $n = 5$ in every case but two, for which only four counts were available. In these cases the quantity was increased by one-third, and used with the others to obtain the aggregate. Adding the 18 quantities obtained from the several varieties, and dividing by $18(n-1)$, we find that on the average the variance is nearly 28·5 times the mean. If now the variance is proportional to the mean we have for each variety

$$\sigma^2 = 28{\cdot}5\bar{x}$$

and the probable error is thence calculated with 18 times the weight with which it would be calculated from a single variety.

A precise test of the accuracy of the above assumption is afforded by the distribution of

$$\frac{n\mu_2}{\sigma^2}$$

for, if the standard deviations have been correctly calculated, this must be distributed as in χ^2 when $n' = 5$ in Elderton's table of Goodness of Fit. For the 17 values we have the following comparison.

	$\frac{n n_2}{\bar{x}}$			
	Expected m	Observed	e	$\frac{\epsilon^2}{m}$
0–2	4·49	4	− ·49	·05
2–4	5·61	8	+ 2·39	1·02
4–6	3·52	1	− 2·35	1·03
6–8	1·83	2		
8–10	·87	1	+ ·44	·12
10–	·69	1		
				$x^2 = 2{\cdot}22$

Since we have introduced one empirical value we must take $n' = 3$, and obtain $P = {\cdot}336$. Thus in 33 cases out of 100 we should expect a worse fit to occur by chance, and this shows that the totality of the observations shows no significant deviation from the set of standard deviations which we have assigned to them.

With a knowledge of the probable errors of the infestation numbers it is possible to test to what extent the several varieties may be grouped together as possibly identical in respect of aphid infestation. From the genetic standpoint it is of the highest importance to determine the continuity or discontinuity of susceptibility; and it is only too frequently that statisticians infer the continuity of a variable quantity without testing to what extent the apparent continuity is due to genuine continuity, due to a multitude of underlying genetic factors, and to what extent a real discontinuity has been obscured by chance variation. In the present case the homogeneity of certain groups is suggested by the mean values. Without being able to test this delicate point with rigour, it is worth while to note that the numbers obtained from certain groups are consistent with the supposition of identical susceptibility.

The group of three varieties (Nos. I, II, XIX) giving mean infestation numbers 1007 to 1037, may be regarded as samples from a single group with mean infestation number at 1020. The standard deviation of the individual plants from this value is 167, and that calculated as explained above is 171.

The variety No. III with infestation number 737 appears to be isolated.

The four varieties (Nos. IV, V, VI, VII) with infestation values 513 to 616 are distributed consistently with the supposition that they are samples of a group with mean value at 570, the observed standard deviation is 105 while the calculated one is 127.

The nine varieties (Nos. VIII, IX, X, XI, XII, XIII, XIV, XV, XVI) with infestation numbers 370 to 451 are distributed about a mean of 407, the observed standard deviation is 103 as against a calculated value 108. The two remaining varieties (Nos. XVII and XVIII) do not seem to be assignable to any other group, and may be regarded at least provisionally as each representing a separate grade of immunity; the very low susceptibility and the high variability of the last variety (No. XVIII) especially makes it worth a more extensive study.

(*Received Oct. 25th*, 1921.)