STUDIES IN THE VARIABILITY OF POCK COUNTS

By P. ARMITAGE

Statistical Research Unit of the Medical Research Council, London School of Hygiene and Tropical Medicine

(With 2 Figures in the Text)

1. INTRODUCTION

The titration of viruses of the pox group is frequently performed by making counts of focal lesions, or 'pocks', produced on the chorio-allantoic membranes of chick embryos (Beveridge & Burnet, 1946). Workers using this technique commonly observe great variation between pock counts caused by samples of the same virus suspension in different eggs. McCarthy, Downie & Armitage (1958) have investigated the variation observed in neutralization tests with variola virus, and have found a particularly simple relationship between the variance of a set of replicate counts, and their mean (the variance being, on the average, about thirteen times the mean). It seemed worth while to continue this investigation, using counts made by other workers and with other viruses. I have analysed some published series, but much of the data is unpublished and was kindly put at my disposal by the workers concerned.

The purposes of the analysis were:

(a) To see whether the simple type of relationship between variance and mean, found by McCarthy *et al.* (1958), was generally valid.

(b) If the relationship varied from one set of data to another, to see whether it was characteristic of the individual laboratory, or of the virus used.

(c) To shed light, if possible, on the mechanism causing the variability in the counts.

(d) To formulate proposals for the statistical treatment of pock counts.

2. DATA

I have analysed eight different sets of data, details of which are given below. *A–D* seemed to be among the most extensive sets of data in the literature; *E–H* are unpublished.

A. Experiments by Burnet & Lush (1939) with herpes virus

(Neutralization experiments on the chorio-allantoic membrane, using virus suspensions with or without the addition of immune serum.) These have been divided into two series:

(i) Four groups of replicate counts from the stock suspension of virus (from table I of Burnet & Lush, 1939).

(ii) Twenty-seven groups of counts from various serum-virus mixtures (from tables II, III, V and VI, Burnet & Lush, 1939).

I have included the low counts bracketed by the authors, and where a group contains a membrane for which a count was 'impossible owing to secondary foci' the remaining counts have been used. Groups with only two counts have not been used, since they provide very little information for present purposes. The remaining groups consisted of between 3 and 16 counts, and the mean count per group varied from 0·3 to 61·8.

B. Experiments by Fenner, Marshall & Woodroofe (1953, table 1) *with myxoma virus*

(Neutralization tests with sera taken from three rabbits, 1 and 15 months after recovery from myxomatosis, in the absence of re-infections.) These have been divided into two series:

(i) Twenty-three groups of counts using either immune sera taken in March 1951 (1 month after recovery) or non-immune sera (for control tests).

(ii) Nineteen groups of counts using immune sera taken in May 1952, 15 months after recovery.

The groups consisted of between 3 and 6 counts, and the mean count per group varied from 4·5 to 87·0.

C. Experiments by French & Reeves (1954) *with five viruses isolated from mosquitoes, and fowl-pox*

(Neutralization tests with fowl serum immune to virus CF-2 (table 1 of French & Reeves), CF-3 (their table 2) and fowl-pox (their table 3)). The data have been divided into six series:

(i) Twelve groups of counts from mixtures of virus CF-2 and the three immune sera.

(ii) Eleven groups of counts from CF-3 and immune sera.

(iii) Twelve groups of counts from CF-4 and immune sera.

(iv) Twelve groups of counts from CF-6 and immune sera.

(v) Thirteen groups of counts from CA-1 and immune sera.

(vi) Ten groups of counts from fowl-pox and immune sera.

The groups consisted of between 3 and 6 counts, and the mean count per group varied from 0·2 to 62·8.

D. Experiments by McCarthy et al. (1958) *with variola virus*

(Neutralization tests with mixtures of virus and immune serum or virus and non-immune serum.) There were six series, with altogether 92 groups of counts. Each group contained 3, 4 or 5 counts, and the mean count per group varied from 11·0 to 200·2.

E. Experiments with cowpox virus

Sixty-six groups of counts, made in the course of titrations of sixteen suspensions of cowpox virus. Each suspension was prepared from a single lesion from the previous titration. The groups consisted of between 3 and 10 counts, and the mean count per group varied from 0·1 to 413·3.

F. Experiments with vaccinia virus

Eighty-one groups of counts, made over a period of about a year. They consisted of between 4 and 39 counts per group, and the mean count per group varied from 1·0 to 96·2.

G. Experiments with viruses of cowpox, alastrim, vaccinia and variola

(Mostly inoculations of supernates of lightly centrifuged preparations from infected chorio-allantoic membranes):

(i) Eleven groups of counts with cowpox (the mean count per group varying from 2·0 to 56·8).

(ii) Twelve groups of counts with alastrim (mean count varying from 1·6 to 191·2).

(iii) Ten groups of counts with vaccinia (mean count varying from 1·6 to 91·2).

(iv) Sixteen groups of counts with variola (mean count varying from 0·8 to 150·0).

The number of counts per group was either 4 or 5.

H. Experiments with vaccinia virus

(i) 115 groups of counts on the Lister Institute strain, passed alternately in rabbits and sheep.

(ii) 162 groups of counts on the Lister Institute strain, adapted to rabbits.

The experiments took place over a period of almost 2 years. The groups consisted of between 3 and 6 counts, and the mean count per group varied from 0·4 to 93·4.

3. STATISTICAL ANALYSIS

The analysis for each set of data followed the method used by McCarthy *et al.* (1958), which is described in detail in their paper. The first step is to calculate, for each group of counts, the mean, $\bar{x}$, and the estimate of variance, s^2 (the sum of squares about the mean divided by one less than the group size). The object of the analysis is to estimate a supposedly linear relationship between the logarithm of σ^2 (the expected value of s^2 for a given $\bar{x}$) and the logarithm of $\bar{x}$. Log σ^2 is estimated by

$$y = \log s^2 + c_r,$$

where c_r is a small correction term depending on the number, r, of counts in the group. The regression line of y on log $\bar{x}$ is then calculated, each value of y being weighted by a factor, w_r, which depends on r and is inversely proportional to the variance of y. Differences between the positions or slopes of the regression lines for different series are tested by standard methods of regression analysis.

If, for any set of data, log σ^2 is uniquely determined by a linear relationship in terms of log $\bar{x}$, the values of y should vary about the fitted line to a predictable extent. In fact, if the counts are normally distributed, the residual weighted mean square about the regression line should be unity, apart from sampling fluctuations. In practice the counts are far from normally distributed, and a correction for the skewness and kurtosis of their distribution may be applied (see §4 below).

As explained in §4, for data E it was necessary to fit a quadratic regression line relating y to log $\bar{x}$. This was carried out by standard methods, using the same weights as before.

4. RESULTS

The sets of data A, B, C, D, G and H were each subdivided into a number of series as indicated in §2. In no instance were there significant, or even suggestive, differences in the slopes or positions of regression lines of y on log $\bar{x}$, fitted separately to the different series within a set, and consequently a single regression line has been fitted for each set of data.

For most sets of data it was clear that the regression was effectively linear, but in a few doubtful cases a quadratic term was included and tested for significance. In only one instance, for set E, was the quadratic regression coefficient significant at the 5 % level, and for all the other sets the linear regression has been taken. As an illustration of the scatter diagrams of y and log $\bar{x}$, Fig. 1 shows those for sets A and E, with the fitted regression lines.

The equations of the regression lines are given in Table 1 together with the expected coefficient of variation (standard deviation expressed as a percentage of the mean), at mean counts of 5, 20 and 100. These values are obtained by calcula-

Table 1. *Regression lines relating log variance to log mean, and coefficient of variation for mean counts of* 5, 20 *and* 100, *for eight sets of data*

Data	Regression line	Coefficient of variation, from regression line, at mean count of		
		5	20	100
A	$Y = \quad 0\cdot148 + 1\cdot264 \log \bar{x}$	65·6	39·4	—
B	$Y = -0\cdot142 + 1\cdot657 \log \bar{x}$	64·4	50·8	—
C	$Y = \quad 0\cdot090 + 1\cdot209 \log \bar{x}$	58·7	33·9	—
D	$Y = \quad 0\cdot852 + 1\cdot148 \log \bar{x}$	—	74·5	37·5
E	$Y = \quad 0\cdot044 + 1\cdot242 \log \bar{x} + 0\cdot197 (\log \bar{x})^2$	63·9	49·6	45·5
F	$Y = \quad 0\cdot011 + 1\cdot518 \log \bar{x}$	68·7	49·2	—
G	$Y = \quad 0\cdot349 + 1\cdot303 \log \bar{x}$	85·4	57·7	30·1
H	$Y = \quad 0\cdot257 + 1\cdot560 \log \bar{x}$	94·3	69·5	—
	Poisson distribution with same mean (coefficient of variance $= 100/\sqrt{\bar{x}}$)	(44·7)	(22·4)	(10·0)

tion of 100 (antilog $\frac{1}{2}Y)/\bar{x}$, where Y is the value of y predicted by the regression line for the value of $\bar{x}$ considered. Coefficients of variation have been calculated only for values of $\bar{x}$ covered by the ranges of mean counts given in §2. In Fig. 2 the regression lines are drawn, over the observed ranges of log $\bar{x}$.

Table 2 gives the regression coefficients of y on log $\bar{x}$, arranged according to the type of virus used. For sets C and G, slopes are given separately for the different viruses used by the same worker. The slopes for each set as a whole are given in the last column. For set E, for which a quadratic line was required, the slope of the best fitting straight line is given, for comparison with the other values. There is no obvious evidence from Table 2 that the slope depends on the type of virus used,

a perhaps surprising observation in view of the differences in size of lesions caused by different viruses. Nor is there any suggestion in the results of the analysis that the slope is affected by the presence in the inoculum of immune serum. On the
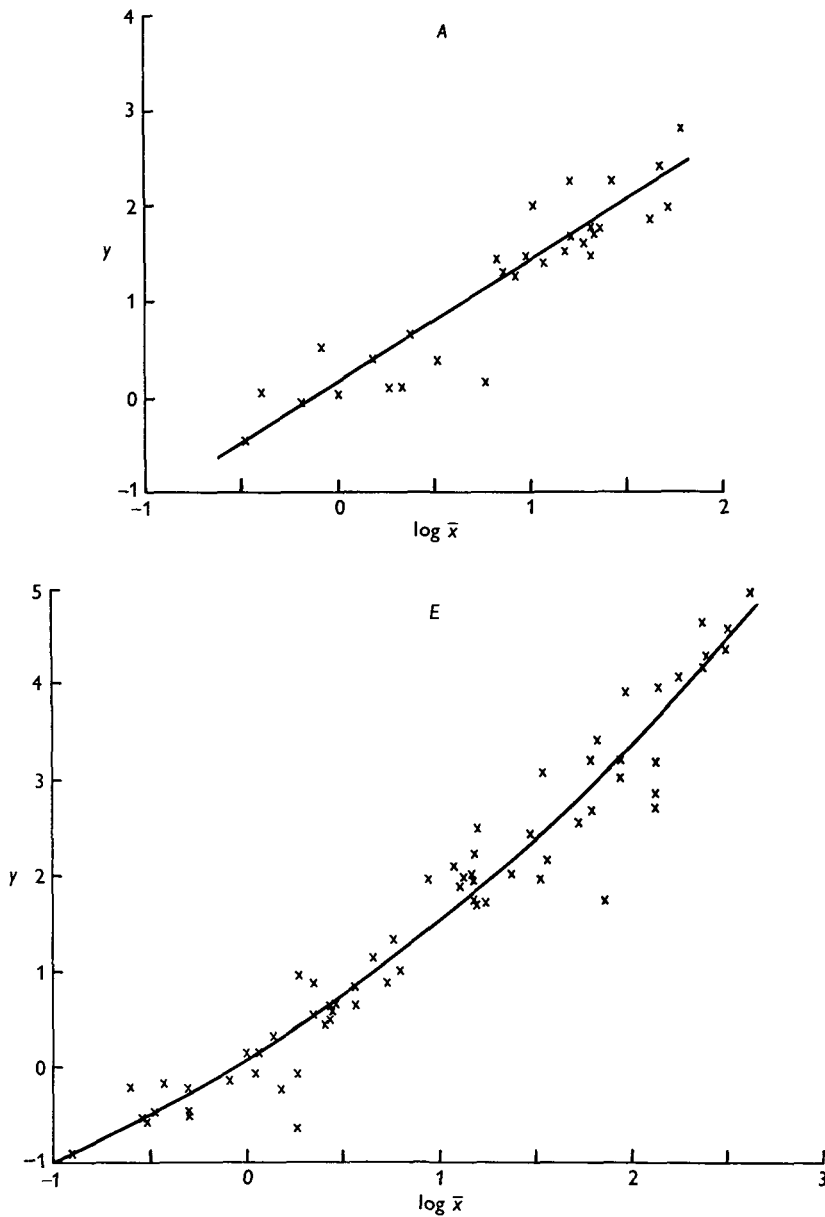

Fig. 1. Scatter diagrams showing relationships between y (the estimated log variance) and $\log \bar{x}$ (the log of the mean count), for sets A and E.

other hand, there are very definite differences in slope between the different sets of data; excluding E, a test of heterogeneity of the slopes gives $\chi^2 = 41\cdot2$ on 6 D.F.; $P < 10^{-5}$. It seems, then, that the variability of pock counts, for a given mean

count, is characteristic of a particular laboratory, and may not vary appreciably from one occasion to another in the same laboratory, even when different viruses are being used.
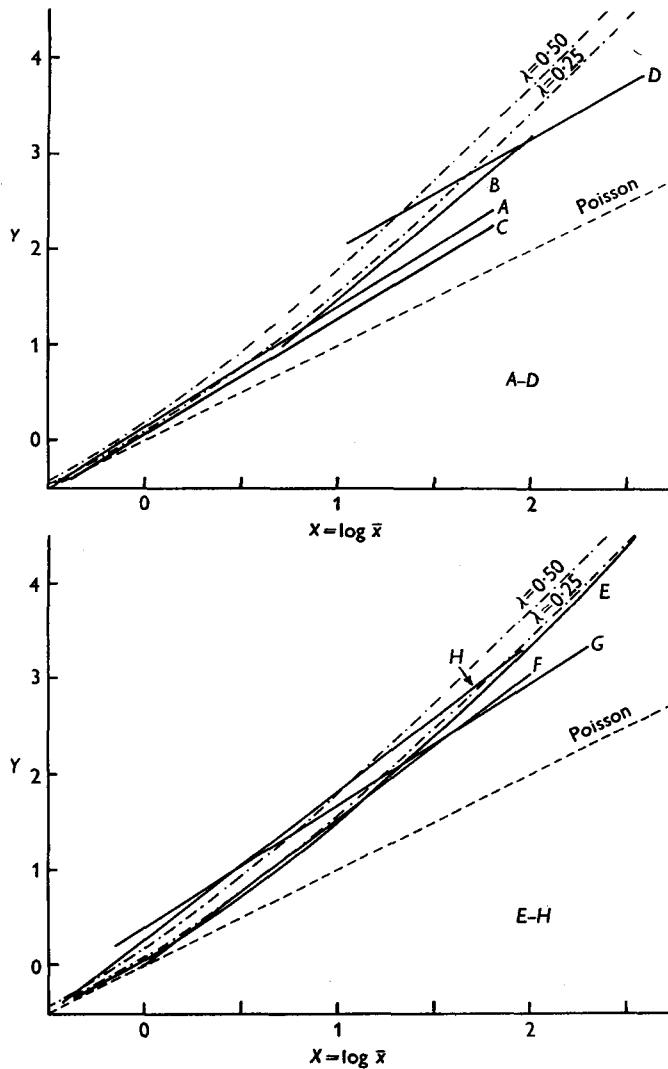

Fig. 2. Relationship between Y (the predicted value of y, the estimated log variance) and $X(= \log \bar{x}$, the log of the mean count). —— Regression lines for sets of data $A–H$. – – – – Theoretical relationship for Poisson distribution. —·—·— Theoretical relationship for model (b), with $\lambda = 0.25$ and 0.50.

McCarthy *et al.* (1958) pointed out that the slope for their data was not significantly different from unity, a value which would imply that the variance was proportional to the mean. Table 2 shows, however, that all the calculated slopes are greater than unity, and all except one of the pooled values significantly so.

The residual mean square about the regression line fitted to each set of data is shown in Table 3. The point of interest raised by these residual mean squares is

36

that a value significantly above its expectation suggests that the variability, in that set of data, cannot be uniquely related to the mean count, as it apparently changes from one occasion to another. On the other hand, a residual mean square not significantly above its expectation provides no evidence of such changes in the variability. As stated in §3, these mean squares would have an expectation of unity if replicate counts were normally distributed, and if, for any set of data,

Table 2. *Regression coefficients of log variance on log mean ($\pm$ standard errors)*

	Herpes	Vaccinia	Variola	Alastrim	Myxoma	Cowpox	Fowlpox	Mosquito viruses	Total
A	1·26±0·10	—	—	—	—	—	—	—	1·26±0·10
B	—	—	—	—	1·66±0·11	—	—	—	1·66±0·11
C	—	—	—	—	—	—	1·31±0·14	1·19±0·07	1·21±0·06
D	—	—	1·15±0·12	—	—	—	—	—	1·15±0·12
E	—	—	—	—	—	(1·59)*	—	—	(1·59)*
F	—	1·52±0·05	—	—	—	—	—	—	1·52±0·05
G	—	1·10±0·37	1·48±0·22	1·25±0·23	—	1·46±0·31	—	—	1·30±0·12
H	—	1·56±0·04	—	—	—	—	—	—	1·56±0·04

* The slope of the best fitting straight line. The coefficients in the quadratic regression equation $Y = a + b_1 \log \bar{x} + b_2 (\log \bar{x})^2$ are $b_1 = 1·24 \pm 0·10$ and $b_2 = 0·20 \pm 0·05$.

Table 3. *Residual mean squares in the weighted analyses of variance of y*

Data	Residual mean square (D.F.)
A	1·159 (29)
B	0·697 (40)
C	0·746 (68)
D	0·719 (90)
E	1·449 (63)
F	1·801 (79)
G	2·044 (47)
H	0·989 (274)

y were distributed about an expected value $\log \sigma^2$, uniquely determined by a linear relationship with $\log \bar{x}$. In general, however, distributions of pock counts are appreciably skew (positive skewness leads to a positive correlation between y and $\log \bar{x}$ in samples from a given population, and tends to *decrease* the residual variation of y), and have excessive kurtosis (which tends to *increase* the residual variance). The effect of non-normality was examined in some detail by McCarthy *et al.* (1958, appendix). In the notation of their paper, the corrected expectation of the residual mean square of y is $(1 - \hat{\rho}^2)(1 + \frac{1}{2}g^2)$, which (for the four series considered there) takes values of 0·99, 1·20, 0·86 and 1·09. These values are, on average, sufficiently near unity to suggest that the correction for non-normality is not very large, and the matter has not been pursued in the present investigation. Of the values shown in Table 3, it seems likely that only those for sets F and G, and possibly that for E, would be significant.

5. MODELS FOR VARIABILITY

It is of some importance to discover why pock counts vary as much as they do, for a knowledge of the mechanism of variability might suggest some way of reducing it. I shall first discuss the relevance of the present results to two particular models purporting to explain variability.

(a) *Poisson theory*

Suppose that

(i) Each lesion is caused by infection of the membrane by only one particle.

(ii) The inocula received by a group of membranes are samples of equal size from a suspension of randomly spaced virus particles.

(iii) All membranes are equally susceptible to infection, in the sense that the probability that a randomly chosen particle will initiate a lesion is independent of which membrane it falls on. (Note that it is not specified here whether the failure of some particles, and the success of others, in initiating infection is due to inherent differences between particles, or to variations in susceptibility of different cells on a given membrane, or to some random feature of the interaction of particle and cell.)

(iv) Particles act independently, so that the chance that a particular particle initiates infection does not depend on how many other particles do so.

Under these assumptions, the numbers of lesions formed on replicate membranes will follow a Poisson distribution. Not only are the mean and variance of the distribution of counts equal, but the expectation of s^2 for a given $\bar{x}$ (which we have denoted by σ^2) is equal to $\bar{x}$. Hence, the theoretical regression of y on $\log \bar{x}$ is

$$Y = \log \bar{x}. \tag{1}$$

The expected value of the slope, b, of the regression line is therefore unity.

As is pointed out above, there is a general tendency for b to exceed unity, and Fig. 2 shows that all the regression lines lie well above the dotted line representing (1). Table 1 shows the extent to which the coefficients of variation (calculated from the regression line) exceed that expected on Poisson theory.

The fact that pock counts usually show variation in excess of that expected on a Poisson theory is, of course, well known. The only authors consistently able to avoid this excessive variability appear to be Reid, Crawley & Rhodes (1949), who report pock counts with fowlpox which show *smaller* variation than that expected on Poisson theory. This effect might in general be due to some sort of overcrowding phenomenon, preventing the formation of high numbers of lesions, but this explanation seems unlikely to be true for this particular set of data. Westwood, Phipps & Boulter (1957) were able to approach the lower limit of Poisson variability by meticulous care in the technique of egg handling. These authors maintain that excessive variability is due primarily to a number of abnormally low counts, the occurrence of which may be related to the presence of non-specific lesions or albumin-sac encroachment. The removal of these technical difficulties would therefore be expected to increase the mean count. Burnet & Faris (1942) similarly maintain that under optimum conditions most of the virus particles present in the

36-2

inoculum will give rise to lesions, and that low counts are obtained on a few 'non-responsive' membranes. They further say that 'abnormally high counts beyond the range of random sampling will only be obtained as a result of technical error or a failure to recognize secondary foci as such'. This view seems to imply that, unless secondary lesions are counted, the distributions of counts will be negatively skew, except perhaps for groups of counts with low means (when the reduction of a few counts may have very little effect on the skewness).

Table 4. *Distribution of the sign of (mean–median) for 10 groups of counts selected at random from each set of data*

	Direction of skewness		
Data	−	0	+
A	4	1	5
B	5	2	3
C	6	0	4
D	7	0	3
E	5	0	5
F	4	0	6
G	3	0	7
H	4	2	4
	38	5	37

One has the impression, on looking through the data analysed in the present paper, that negative skewness is not markedly more frequent than positive skewness even when the counts are fairly high. A full analysis of the whole data would be rather lengthy, but to check the first impression I have taken a random sample of ten groups of replicate counts from each set of data, omitting groups with means less than five. As a simple indication of the direction of skewness I have recorded the sign of the difference between the mean and the median. Positive and negative signs indicate, respectively, positive and negative skewness. The distribution of signs shown in Table 4 confirms the impression recorded above, that there is no strong tendency for counts to show negative skewness. It is difficult to investigate precisely the shape of the distribution of counts when, as in most of the data examined here, the number of replicate counts is small. However, in four of the six series of data *D*, there were no significant differences between the means of different groups. (The groups of counts in Series III were made with samples of the same virus suspension added to different normal sera; those in each of the other three series were made with samples from the same virus-serum mixture.) On pooling the groups within each series the distributions shown in Table 5 are obtained. Three of the four series show marked positive skewness.

Departure from a Poisson distribution implies that at least one of the assumptions on which the Poisson theory is based is incorrect. The assumptions which might be invalid are:

(1) *A lesion is caused by one virus particle.* Excess variability would occur for instance if secondary lesions were counted, even though the primary lesions might follow a Poisson distribution. Secondary lesions are normally not counted.

(2) *Particles act independently.* The most likely way in which this assumption might be invalidated seems to be the possibility of some sort of overcrowding effect, preventing the full development of high numbers of lesions.

(3) *All membranes receive equal inocula.* Variations in size of the value of fluid inoculated into the egg are probably too small to explain the high variability of counts, but the amount reaching the chorio-allantoic membrane might vary considerably.

Table 5. *Distributions of counts in four series of experiments from set D*

Count	Series...	II	III	IV	V*a*
			Frequency		
0–		1	1	17	4
10–		6	0	25	17
20–		14	0	15	16
30–		14	5	10	9
40–		17	3	5	7
50–		8	3	4	8
60–		9	7	1	4
70–		3	4	1	5
80–		6	5	—	2
90–		1	9	—	1
100–		0	14	—	2
110–		1	9	—	0
120–		—	6	—	0
130–		—	4	—	0
140–		—	6	—	1
150–		—	4	—	—
160–		—	0	—	—
170–		—	1	—	—
180–		—	2	—	—
190–		—	2	—	—
200–		—	0	—	—
210–		—	1	—	—
220–		—	0	—	—
230–		—	0	—	—
240–		—	1	—	—
		80	87	78	76

(4) *The membranes are equally susceptible.* Susceptibility might vary for genetic reasons, in which case one might hope to improve matters by decreasing genetic variation in the hens. It might be influenced by the environment to which the egg has been exposed, and this also might be made more uniform. Or it might vary as a result of technical accidents, such as damage to the membrane during inoculation. The ability of Westwood *et al.* (1957) to reduce variability by technical modifications suggests that this may be so.

Paragraphs (3) and (4) above may be summarized by saying that the Poisson model assumes that each membrane has the same *expected* number of lesions, the additional variation being caused by the random distribution of infective particles, whereas it may be true that the expected number varies from one membrane to another. A model allowing for such variation may now be considered.

36-3

(b) Variation in expected counts

Suppose that, for any population of egg membranes, the coefficient of variation (i.e. standard deviation expressed as a percentage of the mean) of the expected counts is constant, irrespective of the density of virus particles in the inoculum. If the variation were caused by differences in susceptibility, for instance, this condition would be met if the expected counts on any two membranes bore the same ratio to each other, for all viral densities. The observed count on any one membrane follows a Poisson distribution about its expected value.

This model has been proposed by Polissar & Shimkin (1954) in a study of the distribution of induced pulmonary tumours in mice, and by Fenner & McIntyre (1956) in a study of myxomavirus.

Let the distribution of expected counts have mean m and variance λm^2 (the coefficient of variation thus being $100\sqrt{\lambda}$ %). Then it may be shown that the variance of the observed counts is

$$V = m + \lambda m^2.$$

(This is equivalent to equation 9 of Polissar & Shimkin (1954) and that given by Fenner & McIntyre (1956, p. 251), and was obtained by Newbold (1927, p. 521) in a study of accident proneness.)

Defining σ^2 in terms of s^2 and $\bar{x}$, as in §3, we shall have approximately,

$$\sigma^2 = \bar{x} + \lambda\bar{x}^2. \tag{2}$$

The accuracy of this approximation would presumably depend on the form of the distribution of expected counts, but we may note that when $\lambda = 0$ we have the Poisson model, and then, as pointed out above, $V = m$ and $\sigma^2 = \bar{x}$ exactly.

For small vules of $\bar{x}$, (2) approaches the equation $\sigma^2 = \bar{x}$ appropriate for the Poisson distribution. This means that the variation in expected counts becomes increasingly unimportant, in comparison with the Poisson contribution, as the mean decreases. For large values of $\bar{x}$, (2) approaches the equation $\sigma^2 = \lambda\bar{x}^2$, showing that in this extreme the Poisson contribution becomes negligibly important.

Polissar & Shimkin (1954) found that (2) provided a good fit to the observed relationship between s^2 and $\bar{x}$ for tumour counts, with an estimated value of λ of 0·265. For comparisons with our previous regression analyses it is convenient to transform (2) to provide the relationship between $Y = \log \sigma^2$ and $X = \log \bar{x}$. We have

$$Y = \log (\bar{x} + \lambda\bar{x}^2)$$
$$= X + \log (1 + \lambda \text{ antilog } X). \tag{3}$$

For small values of $\bar{x}$, (3) approaches the equation $Y = X$, and for large values of $\bar{x}$, the equation $Y = \log \lambda + 2X$. The curve relating Y to X thus has as a lower asymptote the dotted line shown in Fig. 2 corresponding to the Poisson distribution, which has a slope of 1, and as an upper asymptote a line with a slope of 2. As an illustration, the curves for $\lambda = 0·25$ and 0·50 are shown by partly dotted lines in Fig. 2. The curve will be non-linear, with a positive curvature. Positive curvature was detected in only one of my sets of data, but it might well have been obscured

by sampling fluctuations. A straight line fitted to a series of points with a true regression curve given by (3) would be expected to have a slope somewhere near the value of dY/dX evaluated at the observed means of X and y. From (3) it is readily shown that

$$dY/dX = 2 - \text{antilog } (X - Y),$$

and the expected regression slope would therefore be approximately $\hat{b} = 2 - \text{antilog}$ $(\overline{X} - \overline{y})$. Taking $\overline{X}$ as fixed, the sampling variance of $\hat{b}$ would be

$$\text{var } (\hat{b}) = \{2 \cdot 3026 \text{ antilog } (\overline{X} - \overline{y})\}^2 \text{ var } (\overline{y}),$$

where var $(\overline{y})$ is the variance of $\overline{y}$, calculated in the usual manner. To compare the observed slope, b, with the expected value $\hat{b}$ we could take, approximately, $t = (b - \hat{b})/\{\text{var } (b) + \text{var } (\hat{b})\}^{\frac{1}{2}}$ on the number of degrees of freedom shown in Table 3. The comparisons, excluding set E, are made in Table 6. In each case b is lower than $\hat{b}$, and in all but one instance the difference is highly significant.

It is clear from Fig. 2, however, that the quadratic line fitted to data E is quite close to the theoretical curve for $\lambda = 0 \cdot 25$. This impression is confirmed by calculation of the observed and expected gradients at the point on the quadratic regression line where X is equal to the mean value for this set of data, $0 \cdot 77780$. Substitution of this value of X in the regression equation gives $Y = 1 \cdot 12962$, and $\hat{b} = 2 - \text{antilog}$ $(X - Y) = 1 \cdot 56$. The actual gradient of the regression line at this point is $1 \cdot 55$.

The general conclusion is, then, that in at least six of the eight sets of data, the variability in counts fails to increase with the mean as rapidly as would be expected by the model.

Table 6. *Regression coefficients of log variance on log mean, expected on the hypothesis (b) of proportionate expected counts, and their differences from the observed regression coefficient*

Data	Expected slope	
	$\hat{b} \pm \text{s.e. } (\hat{b})$	$(b - \hat{b}) \pm \text{s.e. } (b - \hat{b})$
A	$1 \cdot 61 \pm 0 \cdot 06$	$-0 \cdot 34 \pm 0 \cdot 11$
B	$1 \cdot 84 \pm 0 \cdot 02$	$-0 \cdot 18 \pm 0 \cdot 11$
C	$1 \cdot 42 \pm 0 \cdot 04$	$-0 \cdot 21 \pm 0 \cdot 07$
D	$1 \cdot 92 \pm 0 \cdot 01$	$-0 \cdot 77 \pm 0 \cdot 12$
E	---	---
F	$1 \cdot 72 \pm 0 \cdot 01$	$-0 \cdot 21 \pm 0 \cdot 06$
G	$1 \cdot 83 \pm 0 \cdot 03$	$-0 \cdot 52 \pm 0 \cdot 12$
H	$1 \cdot 84 \pm 0 \cdot 01$	$-0 \cdot 28 \pm 0 \cdot 04$

Whether excess variability is due to differences in susceptibility of membranes or to technical hazards, it seems that these factors exert proportionately less influence at high than at low counts. This would be expected if, for instance, high counts tend to be inhibited by some sort of overcrowding effect. An alternative explanation may be the tendency not to record some high counts when the pocks become confluent. In this connexion it may be relevant that when counts are high, the mean counts, at successive dilutions, may differ by a smaller ratio than that of the dilution step (Burnet & Faris, 1942).

6. ANALYSIS OF TITRATION EXPERIMENTS

I shall consider here the statistical analysis of an experiment in which counts are made at different dilutions. The main object is to express the activity of the preparation by estimating the mean number of lesions, Z, at a particular dilution. Multiplication by the reciprocal of the dilution factor gives an estimate of the density of 'effective' particles in the inoculum, and the reciprocal of this figure may be used as an estimate of the dilution required to produce, on the average, one lesion. A secondary object may be to check whether the mean count decreases in proportion to the dilution.

The choice of a method of analysis will depend on the type of variability observed. If model (a) of §5 were valid, a very simple method of analysis based on the theory of Poisson and binomial distributions could be applied. However, this model seems to be quite inappropriate, at least for most experiments.

If model (b) of §5 were valid, the use of the logarithm of the count would present two advantages. First, the variance of log count would be approximately constant, a desirable feature in most statistical analyses. Secondly, the mean log count would be approximately linearly related (with a slope of unity) to the log dilution. Unfortunately, the frequent presence of zero counts would necessitate some modification (such as the calculation of $\log (x + \frac{1}{2})$), since the logarithm of zero is minus infinity, and this might appreciably affect the linearity at high dilutions. In any case model (b) of §5 appears to be generally inappropriate, and if the logarithmic transformation were applied to any of the data examined in this paper the variance of the log count would decrease with its mean.

Kleczkowski (1949), examining the distribution of counts of lesions on half leaves, caused by plant viruses, found that the standard deviation was related linearly to the mean, and that the transformed variate $\log (\text{count} + c)$ had a variance virtually independent of the mean for any value of c between about 15 and 80.

An alternative procedure for pock counts is to analyse the untransformed counts, making empirical use of the observed variability. As an example, we consider the following counts of fowlpox lesions, taken from table III of French & Reeves (1954):

	Dilution	
	10^{-6}	10^{-5}
	5	18
	6	32
	2	14
	5	37
	5	42
	0	38
Mean	3·8333	30·1667

Simple estimates of the mean count at 10^{-6}, with 95 % confidence limits, can be made either from the observed counts at 10^{-6}, or from those at 10^{-5} (if we assume for the moment that the expected means at the two dilutions will differ by tenfold).

If $\bar{x}$ is the mean of a group of n counts, and $s^2 = \Sigma(x - \bar{x})^2/(n-1)$, the sample estimate of variance, 95 % confidence limits for the mean are given by

$$\bar{x} \pm t_{0.05}s/\sqrt{n},$$

where $t_{0.05}$ is the tabulated value of the t-distribution corresponding to $P = 0.05$ and $n-1$ D.F. For both groups in our example, $n = 6$ and $t_{0.05} = 2.571$, and the calculations give

Dilution	Sample variance s^2	95 % confidence limits for mean $\bar{x} \pm t_{0.05}s/\sqrt{n}$
10^{-6}	5·3667	1·40 and 6·26
10^{-5}	132·1667	18·1 and 42·2

Dividing the values obtained at 10^{-5} by 10, we obtain two estimates of (and limits for) the mean count at 10^{-6}: 3·83 (1·40–6·26) and 3·02 (1·81–4·22).

For a given n, more precise estimates will usually be obtained from the lower dilutions, since the coefficient of variation usually decreases as the mean increases. A more efficient way of analysing the data, though, is to obtain a combined estimate by using all the available results. There is no standard statistical solution to this problem. The method recommended below is based on a number of approximations. In particular, counts are assumed to be normally distributed, but this should not be a serious drawback.

Let u_i be the ratio of the concentration of virus, at the ith dilution, to that at the 'standard dilution' to which the count is to refer. Let $\bar{x}_i$ and s_i^2 be the mean and sample variance (defined as above), and n_i the number of counts. Let $z_i = \bar{x}_i/u_i$ be the estimate of the expected count, Z, at the standard dilution. Let $W_i = n_i u_i^2/s_i^2$. W_i is the reciprocal of the estimated sampling variance of z_i and is the appropriate weight to attach to z_i. The combined estimate of Z is

$$\bar{z} = \Sigma W_i z_i / \Sigma W_i.$$

The variance of $\bar{z}$ cannot be determined exactly because the weights are defined only in terms of *estimated* variances. The best available estimate of the variance of $\bar{z}$ appears to be that given by Cochran & Carroll's (1953) modification of a result due to Meier (1953):

$$\text{var}(\bar{z}) = \frac{1}{\Sigma W_i}\left\{1 + 4\sum_{i=1}^{k}\frac{1}{v_i'}\theta_i(1-\theta_i)\right\},$$

where k is the number of groups,

$$v_i' = n_i - 1 - \frac{4(k-2)}{k-1},$$

and

$$\theta_i = W_i/\Sigma W_i.$$

Note that for $k = 2$, as in the present example, $v_i' = n_i - 1$, the number of degrees of freedom for s_i^2.

Approximate 95 % confidence limits for Z are given by

$$\bar{z} \pm t_{0.05}\sqrt{\{\text{var}(\bar{z})\}},$$

where $t_{0.05}$ is the tabulated value of t for $P = 0.05$ and $f = 1/\left\{\sum_{i=1}^{k}\theta_i^2/(n_i-1)\right\}$ D.F.

The full calculations for the present example (in which $k = 2$) are as follows:

| | Dilution | | |
	10^{-6}	10^{-5}	Total
u_i	1	10	—
n_i	6	6	—
$\bar{x}_i$	3·8333	30·1667	—
z_i	3·8333	3·0167	—
s_i^2	5·3667	132·1667	—
W_i	1·1181	4·5397	5·6578
$W_i z_i$	4·2860	13·6949	17·9809
θ_i	0·1976	0·8024	1·0000
v_i'	5	5	—
$\theta_i(1-\theta_i)/v_i'$	0·03171	0·03171	0·06342
$\theta_i^2/(n_i-1)$	0·00781	0·12877	0·13658

$$\bar{z} = 17\cdot9809/5\cdot6578 = 3\cdot1781,$$

$$\mathrm{var}\,(\bar{z}) = \frac{1}{5\cdot6578}\{1 + 4(0\cdot06342)\} = 0\cdot2216,$$

$$f = 1/0\cdot13658 = 7\cdot32.$$

95 % confidence limits for Z: $3\cdot178 \pm (2\cdot35)\sqrt{(0\cdot2216)} = 2\cdot07$ and $4\cdot28$. (The factor $2\cdot35$ is obtained by interpolation between the entries for 7 and 8 D.F.)

The confidence range for Z, $2\cdot07$–$4\cdot28$, is, as we should expect, slightly narrower than that of $1\cdot81$–$4\cdot22$ given above, which is based on the data for the lower dilution. An alternative method is given by James (1956), but this involves the first few terms of an asymptotic series which is only accurate for reasonably large degrees of freedom. It should be noted that Meier's method also may be inaccurate for small samples.

It is rather more difficult to provide a satisfactory statistical solution to the second problem—that of testing whether the observed mean counts are proportional to the virus concentration (i.e. whether the different values of z_i are consistent with each other). As a first approximation we could ignore the sampling error in the estimates of variance, s_i^2. If these variances were known precisely, for $k = 2$ we could regard

$$v = \frac{z_1 - z_2}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}$$

as a normal variate, with a critical value of $1\cdot960$ at the 5 % significance level. To allow for the sampling errors of the s_i^2 two somewhat controversial solutions have been proposed. One, due to W. V. Behrens and R. A. Fisher, is the basis of a table by Sukhatme (Fisher & Yates, 1953, table VI) for $P = 0\cdot05$ and $0\cdot01$. The other, due to B. L. Welch, is the basis of a table by A. A. Aspin (Pearson & Hartley, 1954, table II) for $P = 0\cdot10$ and $0\cdot02$, and of that by Trickett, Welch & James (1956) for $P = 0\cdot05$ and $0\cdot01$. However, none of these tables allows degrees of freedom less than 6,* and in that by Trickett et al. (1956) the lowest value for

* Tables for using the Behrens test with small degrees of freedom are to appear in the fifth edition of Fisher & Yates's tables (cf. Fisher & Healy, 1956).

degrees of freedom is 8. They will consequently be of limited use for pock count data where groups of five readings or less are common. In the present example, $v = 0.7734$, which is undoubtedly non-significant.

For $k > 2$, as a first approximation we could regard

$$Q = \Sigma W_i (z_i - \bar{z})^2 = \Sigma W_i z_i^2 - (\Sigma W_i z_i)^2 / \Sigma W_i$$

as a χ^2 variate on $k - 1$ degrees of freedom. For equal n_i, Cochran (1937) gives a second approximation, but the use of his formula is inadvisable for values of n_i below about 7. An alternative approximation is given by James (1951) and Welch (1951). According to their result, the critical value of Q exceeded with a probability P is

$$Q_P = \chi_P^2 \left[1 + \frac{3\chi_P^2 + (k+1)}{2(k^2 - 1)} \sum_{i=1}^{k} \left(\frac{1 - \theta_i^2}{n_i - 1} \right) \right],$$

where χ_P^2 is the tabulated value of χ^2 of $k - 1$ D.F., corresponding to the same probability, P. This approximation will get worse as the n_i become smaller, and it is not at all certain how accurate it is for small values.

The calculations for $k = 3$ are illustrated on the following data, which are counts of cowpox lesions from set E. Note that, since $k = 3$, $v_i' = n_i - 1 - (4/2) = n_i - 3$.

	Dilution			
	10^{-5}	10^{-4}	10^{-3}	Total
u_i	1	10	100	—
	0	9	64	—
	0	13	69	—
	1	13	81	—
	1	14	104	—
	1	20	130	—
	2	26	—	—
	2	—	—	—
	2	—	—	—
n_i	8	6	5	—
$\bar{x}_i$	1·1250	15·8333	89·6000	—
z_i	1·1250	1·5833	0·8960	—
s_i^2	0·69643	37·3667	748·300	—
W_i	11·487	16·057	66·818	94·362
$W_i z_i$	12·923	25·423	59·869	98·215
$W_i z_i^2$	14·51	40·25	53·64	108·43
θ_i	0·1217	0·1702	0·7081	1·0000
$1 - \theta_i$	0·8783	0·8298	0·2919	—
v_i'	5	3	2	—
$\theta_i(1 - \theta_i)/v_i'$	0·02138	0·04708	0·10335	0·17181
$\theta_i^2/(n_i - 1)$	0·00212	0·00579	0·12535	0·13326
$(1 - \theta_i^2)/(n_i - 1)$	0·11020	0·13771	0·02130	0·26921

Logically, the first step is to test the consistency of the three groups. For a test at the 5 % significance level, the tabulated value of χ^2 on 2 D.F. is $\chi_P^2 = 5.991$. Hence,

$$Q_P = 5.991 \left[1 + \tfrac{1}{16}(17.973 + 4)(0.26921) \right]$$
$$= (5.991)(1.3697) = 8.21.$$

The value of Q is

$$108\cdot43 - (98\cdot215)^2/(94\cdot362) = 6\cdot20,$$

which falls below the critical value, Q_P. It is concluded that there is no strong evidence of inconsistency between the three groups of counts. Proceeding with the calculations:

$$\bar{z} = 98\cdot215/94\cdot362 = 1\cdot0408,$$

$$\mathrm{var}\,(\bar{z}) = \frac{1}{94\cdot362}\{1 + 4(0\cdot17181)\} = 0\cdot01788,$$

$$f = 1/0\cdot13326 = 7\cdot50.$$

95 % confidence limits for Z: $1\cdot041 \pm (2\cdot33)\sqrt{(0\cdot01788)} = 0\cdot729$ and $1\cdot353$.

SUMMARY

1. The relationship has been examined between the variance and the mean of pock counts made by a number of different workers.

2. In all cases the variance was considerably greater than that expected if the distributions were of the Poisson form. Recent work of Westwood *et al.* (1957) suggests that variation approaching the Poisson expectation can be achieved by appropriate technical measures.

3. With one exception, there appeared to be a linear relationship between the logarithm of the variance and the logarithm of the mean. The slope of this line varied from one laboratory to another, but not, apparently, from one occasion to another in any one laboratory.

4. The hypothesis that for any set of egg membranes the expectations of the pock counts bear constant ratios to each other, irrespective of the virus density, was untenable: the observed variances increase less rapidly with high counts than would have been expected on this theory.

5. Statistical methods are proposed for comparing and combining sets of counts at different dilutions. If Poisson variability were achieved, simpler methods would be available.

REFERENCES

BEVERIDGE, W. I. B. & BURNET, F. M. (1946). The cultivation of viruses and rickettsiae in the chick embryo. *Spec. Rep. Ser. Med. Res. Coun. Lond.* no. 256.

BURNET, F. M. & FARIS, D. D. (1942). The technique of quantitative chorioallantoic virus titration. *J. Bact.* **44**, 241–8.

BURNET, F. M. & LUSH, D. (1939). The inactivation of herpes virus by immune sera: experiments using the chorio-allantoic membrane technique. *J. Path. Bact.* **48**, 275–86.

COCHRAN, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Suppl. J. R. statist. Soc.* **4**, 102–18.

COCHRAN, W. G. & CARROLL, S. P. (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. *Biometrics*, **9**, 447–459.

FENNER, F., MARSHALL, I. D. & WOODROOFE, G. M. (1953). Studies in the epidemiology of infectious myxomatosis of rabbits. I. Recovery of Australian wild rabbits (*Oryctolagus cuniculus*) from myxomatosis under field conditions. *J. Hyg., Camb.*, **51**, 225–44.

FENNER, F. & McINTYRE, G. A. (1956). Infectivity titrations of myxoma virus in the rabbit and the developing chick embryo. *J. Hyg., Camb.*, **54**, 246–57.

FISHER, R. A. & HEALY, M. J. R. (1956). New tables of Behrens' test of significance. *J. R. statist. Soc.*, B, **18**, 212–16.

FISHER, R. A. & YATES, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*, 4th ed. London: Oliver and Boyd.

FRENCH, E. L. & REEVES, W. C. (1954). A group of viruses isolated from naturally infected mosquitoes collected in the Murray Valley area of Victoria and New South Wales. *J. Hyg., Camb.*, **52**, 551–62.

JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, **38**, 324–9.

JAMES, G. S. (1956). On the accuracy of weighted means and ratios. *Biometrika*, **43**, 304–21.

KLECZKOWSKI, A. (1949). The transformation of local lesion counts for statistical analysis. *Ann. appl. Biol.* **36**, 139–52.

McCARTHY, K., DOWNIE, A. W. & ARMITAGE, P. (1958). The antibody response in man following infection with viruses of the Pox group. I. An evaluation of the Pock counting method for measuring neutralizing antibody. *J. Hyg., Camb.* (In the press).

MEIER, P. (1953). Variance of a weighted mean. *Biometrics*, **9**, 59–73.

NEWBOLD, E. M. (1927). Practical applications of the statistics of repeated events, particularly to industrial accidents. *J. R. statist. Soc.* **90**, 487–535.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, vol. i. Cambridge University Press.

POLISSAR, M. T. & SHIMKIN, M. B. (1954). A quantitative interpretation of the distribution of induced pulmonary tumours in mice. *J. nat. Cancer. Inst.* **15**, 377–403.

REID, D. B. W., CRAWLEY, J. F. & RHODES, A. J. (1949). A study of fowl pox virus titration on the chorio-allantois by the pock counting technique. *J. Immunol.* **63**, 165–71.

TRICKETT, W. H., WELCH, B. L. & JAMES, G. S. (1956). Further critical values for the two-means problem. *Biometrika*, **43**, 203–5.

WELCH, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, **38**, 330–36.

WESTWOOD, J. C. N., PHIPPS, P. H. & BOULTER, E. A. (1957). The titration of vaccinia virus on the chorio-allantoic membrane of the developing chick-embryo. *J. Hyg., Camb.*, **55**, 123–39.