J. Hyg., Camb. (1974), **73**, 27
Printed in Great Britain

Statistical analysis of data bearing on the number of particles required to form a plaque

By J. G. KALBFLEISCH and D. A. SPROTT
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(*Received* 26 *October* 1973)

SUMMARY

Methods of statistical analysis are presented for one or more dilution series experiments where the quantity of interest is the number of virus particles required to infect a cell. These methods are illustrated on several data sets drawn from the literature. Data from seven series, which have been used to support a two-particle model in the literature, are here shown to reject such a model decisively, whereas fifteen other experiments are found to be in excellent agreement with a one-particle model.

INTRODUCTION

A fundamental question that has been of interest in virology for some time is the number of virus particles required to infect a cell. A straight line relationship between concentration of virus innoculated and number of plaques observed is usually taken to be evidence for a single particle (Dulbecco, 1952; Dulbecco & Vogt, 1954; Khera & Maurin, 1958; Kjellén, 1961; Cooper, 1961; Boeyé, Melnick & Rapp, 1966). If two particles are required, the relationship between plaque count and dose is quadratic. In general, if h particles are required to form a plaque, the number of plaques is proportional to the hth power of the concentration (Dulbecco & Vogt, 1954; Cooper, 1961; Boeyé *et al.* 1965). Where evidence concerning linearity or non-linearity has been presented, it appears to have been assessed by examining visually graphs of concentration versus observed plaque count, or by presenting estimates of h. No quantitative or statistical assessment of the strength of the evidence in this regard appears to have been made.

The effort and expense involved in a thorough statistical analysis will usually be negligible in comparison with that required to obtain reliable data. It therefore seems only reasonable that, when observations have been carefully made, a complete analysis should be undertaken in order to extract the maximum amount of information from them.

It is the purpose of this paper to present appropriate methods of statistical analysis. The necessity of a statistical analysis is demonstrated on data which have been taken to support a two-particle model in the literature but which are here shown to reject such a model decisively. Hence these methods can sometimes detect departures from the predicted relationship that are not apparent from visual inspection alone.

THE MATHEMATICAL MODEL

Suppose that there are k successive dilutions of an initial viral concentration by a dilution factor d, giving $k+1$ concentrations in proportion to $1:d^{-1}:d^{-2}:\ldots:d^{-k}$. The assumptions usually made are that it requires h particles to infect a cell; that the initial concentration of virus particles is not too great; and that the virus particles are randomly distributed among the cells on the plate. It can then be shown that the theoretical or expected plaque count at dilution level j (concentration d^{-j}) will be

$$\nu(d^{-j})^h = \nu\theta^j,$$

where ν is the expected number of plaques arising from the undiluted suspension, and $\theta = d^{-h}$ (Dulbecco & Vogt, 1954; Cooper, 1961; Boeyé *et al.* 1966). Furthermore, it follows *from the same assumptions* that the actual plaque count, which is subject to random fluctuations about this expected value, will follow a Poisson distribution (Reid, Crawley & Rhodes, 1949; Kjellén, 1961; Cooper, 1961; Boeyé *et al.* 1966; Fisher, 1970, 54–63). The probability of observing y_j plaques at level j is then

$$(\nu\theta^j)^{y_j} \exp(-\nu\theta^j)/y_j!.$$

This model was used by Alling (1971). If there are n_j separate flasks or plates at dilution level j, then y_j will be used to denote the total number of plaques on all n_j plates, and $\nu\theta^j$ will be replaced by $n_j\nu\theta^j$.

The statistical analysis presented in the following sections flows from the above model. The quantity of interest is h, the number of particles required to form a plaque. The quantity ν is not usually of interest, and is eliminated from the analysis by the technical device of conditioning on the total plaque count over all dilutions. The analysis is then based on the fact that the maximum likelihood estimate of h has approximately a normal distribution. For justification and mathematical details see Kalbfleisch & Sprott (1974).

STATISTICAL ANALYSIS OF A SINGLE SERIES

Define y_j, n_j, d, h, and $\theta = d^{-h}$ as above, and let $\log d$ denote the natural logarithm of d. The following quantities enter into the analysis:

$$\left.\begin{aligned} X &= y_0+y_1+y_2+\ldots+y_k; \\ T &= \quad y_1+2y_2+\ldots+ky_k; \end{aligned}\right\} \tag{1}$$

$$\left.\begin{aligned} B &= n_0+n_1\theta+n_2\theta^2+\ldots+n_k\theta^k; \\ A &= \quad n_1\theta+2n_2\theta^2+\ldots+kn_k\theta^k; \\ D &= \quad n_1\theta+2^2n_2\theta^2+\ldots+k^2n_k\theta^k; \end{aligned}\right\} \tag{2}$$

$$S = -\left(T - X\frac{A}{B}\right)(\log d); \tag{3}$$

$$I = X(DB-A^2)(\log d)^2/B^2. \tag{4}$$

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

Table 1. *Analysis of plaque counts from Boeyé et al.* (1966)

Isolate	Dilution level 0	1	2	$\hat{h}$	$\hat{I}$	u ($h = 2$)	u ($h = 1{\cdot}8199$)
1	122	10	2	2·0280	19·63	0·124	0·922
2	176	10	4	2·1085	25·18	0·544	1·448
3	170	19	2	1·9116	32·34	−0·503	0·521
4	266	40	5	1·6755	70·54	−2·725	−1·213
5	264	38	4	1·7303	64·88	−2·172	−0·722
6	306	42	3	1·8076	67·63	−1·582	−0·101
7	186	22	2	1·8889	36·58	−0·672	0·417

The first step in the analysis is the computation of $\hat{h}$, the maximum likelihood estimate (M.L.E.) of h. This is the value of h which is best supported by the data in the sense that, when $h = \hat{h}$, the probability of the observed plaque counts is as great as it possibly can be under the model. The estimate $\hat{h}$ may be obtained by solving the equation $S = 0$ using the procedure described in the Appendix. For discussions of maximum likelihood estimation, see Finney (1964, p. 80) and Kempthorne (1969, p. 167).

If X, the total plaque count over all dilutions, is not too small, then the maximum likelihood estimate $\hat{h}$ will have approximately a normal distribution with mean h and variance $\hat{I}^{-1}$, where $\hat{I}$ is the value of I computed from (4) using $h = \hat{h}$. Hence the quantity u defined by

$$u = (\hat{h} - h)\sqrt{\hat{I}} \tag{5}$$

will have approximately a standardized normal distribution (a normal distribution with mean 0 and variance 1), for which tables are readily available. In order to determine whether a proposed value of h is consistent with the data, the corresponding u-value is computed from (5) and is compared with the tables. An improbably large or small (negative) value of u provides evidence against the proposed value of h. An approximate 95 % confidence interval for h is given by $\hat{h} \pm 1{\cdot}96/\sqrt{\hat{I}}$, this being the set of h-values for which u lies within the central 95 % of the standardized normal distribution.

Example. Columns 2, 3 and 4 of Table 1 give plaque counts from Boeyé *et al.* (1966, Table 4), which were taken as supporting a two-particle model ($h = 2$). There are three dilution levels ($k = 2$), the dilution factor d is $\sqrt{10}$, and each plaque count is the total over $n_j = 2$ plates. The fifth column gives $\hat{h}$ for each isolate, computed as in the Appendix. Column 6 gives $\hat{I}$, which is computed from (4) with $h = \hat{h}$. Column 7 gives the value of u for each isolate as computed from (5) with $h = 2$. Isolates 1, 2, 3, 6, and 7 are now seen to be consistent with $h = 2$ because the corresponding values of u are reasonable ones (within the central 95 % of a standardized normal distribution). However, isolates 4 and 5 yield u-values which differ from zero by more than 1·96, and therefore contradict the assumption that h is 2.

The last column of Table 1 gives the value of u for each isolate as computed from (5) using $h = 1{\cdot}8199$. These values will be used in the next section.

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

COMBINATION OF DATA FROM SEVERAL SERIES

Given the results of r dilution series experiments, such as the seven isolates of Table 1, two questions will be of interest. First, one will wish to know whether the data are homogeneous; that is, whether there exists a single value of h which is compatible with the data from all r experiments. Secondly, assuming a common value of h, one will wish to combine the data from the r experiments to give a single estimate of h, or to test some theoretical value such as $h = 2$ in the preceding example.

Combination of the data

The overall maximum likelihood estimate of the common value of h in r experiments can be obtained easily on an electronic computer. However, the following analysis is computationally much simpler, and will give almost identical results provided that the total plaque count X is fairly large in each experiment.

Let the M.L.E. of h in the ith experiment be denoted by $\hat{h}_i$, with approximate variance $\hat{I}_i^{-1}$ as calculated in the last section. An overall estimate of h may be obtained as a weighted average of the individual estimates:

$$\bar{h} = \Sigma \hat{I}_i \hat{h}_i / \Sigma \hat{I}_i. \tag{6}$$

This has approximately a normal distribution with mean h and variance $(\Sigma \hat{I}_i)^{-1}$. Hence the quantity

$$z = (\bar{h} - h) \sqrt{\Sigma \hat{I}_i} \tag{7}$$

has approximately a standardized normal distribution, and may be used to assess an hypothesized value of h on the basis of the combined data from all of the experiments.

Heterogeneity of the data

To test for heterogeneity in the data, compute

$$\chi^2 = \Sigma u_i^2, \quad \text{where} \quad u_i = (\hat{h}_i - \bar{h}) \sqrt{\hat{I}_i}. \tag{8}$$

Under the assumption of homogeneity, χ^2 has approximately a chi-square distribution with $r-1$ degrees of freedom, and this distribution is extensively tabulated. An improbably large value of χ^2 would show that there was no single value of h which was compatible with the data from all r experiments.

Example (*continued*). The values from Columns 5 and 6 of Table 1 may be substituted in (6) to give the weighted average $\bar{h} = 1{\cdot}8199$, with variance

$$(\Sigma \hat{I}_i)^{-1} = (316{\cdot}77)^{-1}.$$

The values $u_1, u_2, \ldots, u_7$, computed from (5) with $h = 1{\cdot}8199$, are given in the last column of Table 1, and their sum of squares is $\chi^2 = 5{\cdot}395$. The value of χ^2 is a very probable one, lying close to the 50 % point of a chi-square distribution with 6 degrees of freedom, and hence there is no evidence of heterogeneity among the seven isolates.

To determine whether the combined data of all seven isolates are consistent

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

with the two-particle model, set h equal to the theoretical value 2. Then (7) gives $z = -3{\cdot}205$. The chance of obtaining a value so far from zero in a standardized normal distribution is less than 0·003, and hence *the combined data from all seven isolates are incompatible with the two-particle model.*

It is perhaps not obvious from a visual inspection of the estimates of h in Table 1 that the combined experiment provides such strong evidence against the theoretical value $h = 2$. In fact, *it has been concluded elsewhere in the literature that these data support the two-particle theory.* This example shows the need for quantitative methods, without which the strength of the evidence cannot adequately be assessed.

Further analysis may be undertaken in an attempt to determine the reason for the departures from the two-particle model. The assumption of a Poisson distribution can be checked statistically and appears to be satisfactory. The possibility of a one-particle model is easily ruled out because $h = 1$ is even more decisively rejected by the data than $h = 2$. Indeed, no integer value of h is compatible with the data. A possible explanation is that generally two particles are required to form a plaque, but there is a small probability that one particle will suffice.

DATA CONFORMING TO A ONE-PARTICLE MODEL

In the preceding example, the dilution factors and numbers of dilution levels were the same in all seven isolates. For a more complex example, data from $r = 15$ dilution series experiments were taken from the following four easily accessible sources:

(1) Dulbecco (1952); (2) Dulbecco & Vogt (1954);

(3) Khera & Maurin (1958); (4) De Maeyer (1960).

The plaque frequencies and dilution factors for these 15 experiments are recorded in Table 2 according to their source. The numbers of plates used are given in parentheses. For instance, in experiment $(2d)$ the dilution factor d was 3, and there were three dilution levels ($k = 2$). At dilution level 0 there were 2 plates with a total of 46 plaques; at level 1 there were 6 plates with a total of 61 plaques; and at level 2 there were 10 plates with a total of 36 plaques. The remaining dilution levels 3, 4, 5, and 6 were not used in experiment $(2d)$.

The second column of Table 3 gives $\hat{h}$ for each of the 15 experiments computed as in the Appendix. The calculations may be performed by desk calculator in a few hours, or by electronic computer in a few seconds. For the latter, all that is required is a routine to evaluate S and I as defined by (3) and (4). The calculations outlined in the Appendix can then be carried out by repeatedly applying this programme.

The third column of Table 3 gives the value of I when $h = \hat{h}$ for each of the fifteen experiments, and column four gives the value of u for each experiment, computed from (5) using the theoretical value $h = 1$. All fifteen values of u lie within the central 95 % of a standardized normal distribution. Hence each individual experiment is compatible with a one-particle model, and this agrees with *the conclusions reached in the literature.*

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

Table 2. *Plaque counts and numbers of plates used in fifteen experiments*

Experiment	Dilution level 0	1	2	3	4	5	6	Dilution factor d
(1) *a*	297 (2)	152 (2)	—	—	—	—	—	2
b	112 (2)	124 (7)	—	—	—	—	—	3
c	79 (1)	23 (1)	—	—	—	—	—	3
d	50 (1)	—	12 (1)	2 (1)	—	—	—	2
e	26 (1)	10 (1)	—	—	—	—	—	3
(2) *a*	305 (3)	238 (4)	—	—	—	—	—	2
b	47 (1)	46 (2)	—	—	—	—	—	2
c	82 (2)	84 (6)	—	—	—	—	—	3
d	46 (2)	61 (6)	36 (10)	—	—	—	—	3
e	102 (4)	99 (8)	92 (16)	—	—	—	—	2
(3) *a*	66 (2)	44 (2)	27 (2)	17 (2)	11 (2)	4 (2)	4 (2)	$\sqrt[5]{10}$
b	178 (2)	63 (2)	—	6 (2)	0 (2)	—	—	$\sqrt{10}$
c	180 (4)	27 (2)	6 (2)	2 (2)	—	—	—	$\sqrt{10}$
(4) *a*	264 (2)	25 (2)	—	—	—	—	—	10
b	476 (2)	39 (2)	—	—	—	—	—	10

Table 3. *Analysis of plaque counts from Table 2*

Experiment	$\hat{h}$	$\hat{I}$	u ($h = 1$)	u ($h = 1{\cdot}0164$)
(1) *a*	0·9664	48·31	−0·234	−0·348
b	1·0477	71·03	0·402	0·264
c	1·1232	21·50	0·571	0·495
d	1·2372	27·68	1·248	1·162
e	0·8697	8·72	−0·385	−0·433
(2) *a*	0·7729	64·23	−1·820	−1·951
b	1·0310	11·17	0·104	0·049
c	0·9781	50·08	−0·155	−0·271
d	0·8390	102·74	−1·632	−1·798
e	1·0739	93·73	0·715	0·557
(3) *a*	1·0435	89·11	0·411	0·256
b	1·0083	144·68	0·100	−0·097
c	1·1113	81·20	1·003	0·855
(4) *a*	1·0237	121·08	0·261	0·080
b	1·0865	191·12	1·196	0·969

It is still possible that the combined data from all fifteen experiments might contradict the one-particle model, and hence further analysis is desirable. The values from columns 2 and 3 of Table 3 may be substituted into (6) to give the weighted average $\bar{h} = 1{\cdot}0164$, with variance $(1126{\cdot}35)^{-1}$. The values $u_1, u_2, \ldots, u_{15}$, computed from (5) with $h = 1{\cdot}0164$, are given in the last column of Table 3, and their sum of squares is 11·15. The 50 % point of a chi-square distribution with 14 degrees of freedom is 13·34, so that the value obtained is not an unusually large one. Hence there is no evidence of heterogeneity among the fifteen experiments.

If h is set equal to the theoretical value 1, then (7) gives $z = 0{\cdot}550$. From tables

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

of the standardized normal distribution, the chance of a more extreme value is greater than 50 %. Hence the one-particle model is in accord with the combined data from all fifteen experiments.

The preceding example shows that a statistical analysis applied to diverse experiments performed over a wide period of time and in different places can exhibit a convincing compatibility with the hypothesis in question ($h = 1$). The results of such an analysis would seem to be more compelling than solely examining point estimates $\hat{h}$ or graphs of plaque count *vs.* dose.

We should like to thank Dr W. S. Rickert and Dr W. F. Forbes for helpful suggestions.

APPENDIX

Computation of $\hat{h}$ for a single series

Let y_i and y_j be the two largest plaque totals, and define

$$\tilde{h} = (\log [n_i y_j / n_j y_i])/(i-j) \log d.$$

If the series has only two dilution levels ($k = 1$), then $\hat{h} = \tilde{h}$. However if $k > 1$, $\tilde{h}$ gives only a first approximation to $\hat{h}$, and additional calculations are required. Let h_1 denote some initial guess at the value $\hat{h}$. (For instance, one might choose the proposed theoretical value, or else take $h_1 = \tilde{h}$). Putting $h = h_1$, one computes θ, B, A, D, S, and I using the formulas (1) to (4). A closer approximation to $\hat{h}$ will then be given by $h_2 = h_1 + \Delta$, where $\Delta = S/I$. This procedure may now be repeated with h_2 as the new initial value to obtain yet a better approximation, $h_3 = h_2 + \Delta$. One continues in this fashion until the correction factor Δ becomes sufficiently small.

For example, consider the data of isolate one in Table 1. Here

$$d = \sqrt{10}, \quad k = 2 = n_0 = n_1 = n_2, \quad X = 134, \quad \text{and} \quad T = 14.$$

For an initial guess at $\hat{h}$ one may select the theoretical value, $h_1 = 2$, and the following results are then obtained:

$$\theta = 0{\cdot}1 \qquad B = 2{\cdot}22 \qquad A = 0{\cdot}24 \quad D = 0{\cdot}28$$

$$S = 0{\cdot}5601 \quad I = 20{\cdot}23 \quad \Delta = 0{\cdot}02755$$

Hence a closer approximation to $\hat{h}$ is

$$h_2 = h_1 + \Delta = 2{\cdot}02755.$$

The calculations are now repeated using $h = 2{\cdot}02755$ to give

$$\theta = 0{\cdot}096878, \quad B = 2{\cdot}21253, \quad A = 0{\cdot}231297 \quad D = 0{\cdot}268839$$

$$S = 0{\cdot}009563, \quad I = 19{\cdot}64, \qquad \Delta = 0{\cdot}00049$$

An even better approximation to $\hat{h}$ is then given by

$$h_3 = h_2 + \Delta = 2{\cdot}02804,$$

which is correct to at least three decimal places.

https://doi.org/10.1017/S0022172400023792 Published online by Cambridge University Press

REFERENCES

ALLING, D. W. (1971). Estimation of hit number. *Biometrics* **27**, 605–13.

BOEYÉ, A., MELNICK, J. L. & RAPP, F. (1966). SV 40-adenovirus hybrids; presence of two genotypes and the requirement of their complementation for viral replication. *Virology* **28**, 56–70.

COOPER, P. D. (1961). The plaque assay of animal viruses. *Advances in Virus Research* **8**, 319–78.

DE MAEYER, E. (1960). Plaque formation by Measles Virus. *Virology* **11**, 634–8.

DULBECCO, R. (1952). Production of plaques in monolayer tissue cultures by single particles of an animal virus. *Proceedings of the National Academy of Sciences, U.S.A.* **38**, 747–52.

DULBECCO, R. & VOGT, M. (1954). Plaque formation and isolation of pure lines with Poliomyelitis viruses. *Journal of Experimental Medicine* **99**, 167–82.

FINNEY, D. J. (1964). *Statistical Method in Biological Assay*, 2nd edition. London: Griffin.

FISHER, R. A. (1970). *Statistical Methods for Research Workers*, 14th edition. Edinburgh: Oliver and Boyd.

KALBFLEISCH, J. G. & SPROTT, D. A. (1974). Inferences about hit number in a virological model. *Biometrics* **30**, 199–208.

KEMPTHORNE, O. (1969). *An Introduction to Genetic Statistics*. Ames, Iowa: State University Press.

KHERA, K. S. & MAURIN, J. (1958). L'etude par la methode des plaques du virus aphteux (type C) en couche monocellulaire de rein de porcelet. *Annales de l'Institut Pasteur* **95**, 557–67.

KJELLÉN, L. (1961). A study of adenovirus in host cell systems by the plaque technique. *Virology* **14**, 234–9.

REID, D. B. W., CRAWLEY, J. F. & RHODES, A. (1949). A study of fowl pox virus detection on the chorioallantois by the poch counting technique. *Journal of Immunology* **63**, 165–71.