256

THE ANALYSIS OF VARIANCE WITH VARIOUS BINOMIAL TRANSFORMATIONS

PROFESSOR SIR RONALD FISHER
Department of Genetics, Cambridge

1. *Introductory.*

Much experimental data are in the form, sometimes termed quantal, in which out of n independent trials a particular response, e.g. the death of an experimental animal, or the growth of a microbial culture, is observed a times, and in which such pairs of values n and a have been obtained under a variety of conditions, the variation being perhaps to some extent deliberately imposed, as by a variation of dosage, and to some extent out of experimental control, as is the variation in response of different batches of material.

In such cases it is usually desirable to interpret each pair of values as supplying information about a variate functionally connected with the probability of which a/n is an empirical estimate, and such that it is, so far as possible, additive in respect of the effects of varying conditions, and linear in deliberately imposed measures such as dosage, when these are given an appropriate metric.

Examples of such transformations of the probability, which have been widely used are

$$p = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du \tag{1}$$

where x, or, for computational convenience sometimes $x + 5$, is termed the Probit;

$$p = \sin^2 \varphi \qquad q = \cos^2 \varphi \tag{2}$$

where φ runs from 0 to 90°,
or, sometimes using the double angle,

$$2p = 1 - \cos\theta \qquad 2q = 1 + \cos\theta$$

where φ, or θ is spoken of as an Angular transformation; the logistic transformation

$$z = \tfrac{1}{2} \log (p/q) \tag{3}$$

where z, or sometimes $2z$, has been termed the Logit; the log log transformation,

$$x = \log \log (1/p), \tag{4}$$

and a variety of others appropriate to special situations, such as that devised for the interpretation of gene ratios in a situation involving diffusion and selection in equilibrium, defined by the differential equation

$$\frac{d^2p}{\partial x^2} = 4pqx \tag{5}$$

with the "boundary" conditions

$$p = \tfrac{1}{2}, \qquad x = 0$$

$$p \to 0 \quad \text{as} \quad x \to \infty,$$

defining a function x of p which I have termed a Legit (7).

No doubt many other such transformations will be developed for special purposes. The five mentioned have, however, all been the subject of mathematical investigation, and the necessary numerical tables have been supplied for their convenient use.

2. *The maximum likelihood procedure.*

The practical procedure of fitting by Maximum Likelihood as a means of obtaining a correct analysis of variance is an application of the general method of efficient scores. The probability of what has been observed in any set of n trials is

$$\frac{n!}{a!(n-a)!} p^a q^{n-a},$$

and the log likelihood, so far as concerns the unknown p, is

$$a \log p + (n - a) \log q;$$

then its rate of change for variation of the transformed variate, say x, is

$$\left(\frac{a}{p} - \frac{n-a}{q}\right) \frac{dp}{dx},$$

which is the efficient score with respect to x for this set of n observations. We notice that the n trials may be scored individually, with scoring coefficients

$$\frac{1}{p}\frac{dp}{dx}$$

for the events with expected frequency p, called "successes", and

$$-\frac{1}{q}\frac{dp}{dx}$$

for the alternative events, or "failures". The mean of such scores is zero, for

$$p\left(\frac{1}{p}\frac{dp}{dx}\right) - q\left(\frac{1}{q}\frac{dp}{dx}\right) = 0$$

and their mean square, the Amount of Information is

$$p\left(\frac{1}{p}\frac{dp}{dx}\right)^2 + q\left(\frac{1}{q}\frac{dp}{dx}\right)^2,$$

so

$$i = \frac{1}{pq}\left(\frac{dp}{dx}\right)^2$$

This is the amount of information about x for a single trial; for the set of n trials it comes naturally to a value n times as great.

Corresponding with any set of observations we may now construct a variate

$$y = \frac{1}{ni}\left(\frac{a}{p} - \frac{n-a}{q}\right)\frac{dp}{dx} = \frac{(a-np)}{n\dfrac{dp}{dx}}$$

by dividing the score by the amount of information on which it is based, which gives the linear adjustment required by the observations to any proposed value x. The variance of this variate will be exactly $1/ni$, when a takes the binomial distribution

$$(p+q)^n,$$

so in further analysis the variate will be given the weight ni. Parameters such as regression coefficients, class differences, etc., fitted by using such variates with their proper weights will necessarily satisfy the conditions of maximal likelihood. For example, if the transformed values, x, are believed to be linearly related to some observable, t, with an equation

$$x = A + Bt$$

in which A, B are to be adjusted to suit the data, we may note that

$$\frac{\delta x}{\delta A} = 1, \qquad \frac{\delta x}{\delta B} = t$$

so that A, B will take the pair of values of maximal likelihood, when the sum of the scores, and the sum of the products of each score by the corresponding t, are both zero. In other words, when

$$\sum wy = 0, \qquad \sum wyt = 0$$

where y is the working deviation from the expected value x.

Using as weights the reciprocals of the exact variances of the variate in each set, all residual sums of squares are χ^2 values of the appropriate degrees of freedom, and as such are available to test any questionable aspect of the hypothetical formulation on which the analysis is based.

3. *Practical apparatus.*

For each type of transformation used, we need only tabulate against x,

the maximum working value

$$x + \frac{1}{ip}\frac{dp}{dx} = x + q\bigg/\left(\frac{dp}{dx}\right),$$

the minimum working value

$$x - \frac{1}{iq}\frac{dp}{dx} = x - p\bigg/\left(\frac{dp}{dx}\right),$$

the weighting coefficient

$$i = \frac{1}{pq}\left(\frac{dp}{dx}\right)^2.$$

So for probits, if z stands for dp/dx,

$$z = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},$$

the maximum and minimum working probits are

$$x + \frac{q}{z}, \qquad x - \frac{p}{z},$$

the range is

$$\frac{1}{z}$$

and the amount of information is

$$\frac{z^2}{pq}.$$

Again, if instead of p being given explicitly in terms of x, x is given in terms of p, as for example,

$$-x = \log(-\log p),$$

then

$$\frac{dx}{dp} = \frac{-1}{p \log p},$$

and so,

$$\frac{dp}{dx} = -p \log p,$$

$$i = \frac{p}{q}(\log p)^2,$$

and the maximum and minimum working variates are

$$x - \frac{q}{p \log p}, \qquad x + \frac{1}{\log p},$$

remembering that $\log p$ is always negative.

Tabular apparatus for (1) Probits and (2) angular values have been published in *Statistical Tables*, (8) where are also given the formulae, which need no special tabulation, for (3) Logits. For (5) Legits, tables in similar form have been published more recently (7) in Biometrics. I do not know that the values for (4) given by the formulae above have been tabulated, but in any case they are not difficult to calculate using only a table of natural logarithms. If u is log dose taken to the base e, of a living infective agent, and if each living particle is supposed to have an equal and independent chance of establishing an infection, then the regression of x on u will be linear with regression coefficient equal to unity. The experimental verification of this equality is *pro tanto* confirmation of the theory that this probability is independent of the subject, as in another way is the χ^2 value of the sum of squares of deviations from the fitted line.

4. *"Corrections" to angular and square root transformations.*

Just as in the limit for large sample size and small probability the binomial distribution,

$$(q + p)^n,$$

tends to the Poisson limit with

$$m = pn,$$

so the angular transformation

$$p = \sin^2 \varphi$$

becomes

$$x \alpha \sqrt{m}$$

where x from different populations is proportional to the original angular measure φ. The working variate is then

$$\sqrt{m} + \frac{a - m}{2\sqrt{m}} = \frac{a + m}{2\sqrt{m}},$$

where a is the observed and m the expected frequency; the weighting coefficient is a constant, 4.

It should be emphasized that if an analysis of variance is in view the choice of what transformation to use should be governed by the prospective additiveness of the transformed variate when various controllable, or uncontrollable factors, the effects of which are to be analysed, are varied. In the author's view, conformity with this and other presuppositions of the method chosen, after accurate fitting, can be satisfactorily confirmed by a series of χ^2 tests. Significant heterogeneity appearing at this stage in spite of exact analysis gives serious grounds for doubting the appropriateness either of the transformation, or of the adequacy of control of the experimental material.

In the use of the angular and square root transformations the near constancy of the variances due to purely sampling error of the transformed variates has exercised a certain fascination, and has sometimes seemed to be the reason for choosing this type of transformation.

The fact is that the *amount of information* about φ,

$$i = \frac{1}{pq}\left(\frac{dp}{d\varphi}\right)^2$$

reduces, when p is equated to $\sin^2 \varphi$, to

$$\frac{1}{\sin^2 \varphi \cos^2 \varphi} (2 \sin \varphi \cos \varphi)^2 = 4,$$

and is exactly constant for all values of φ. The variance of the working angle is therefore absolutely constant.

Similarly, the amount of information about $\sqrt{m}$, where m is the parameter of the Poisson series is

$$\frac{1}{m} (2\sqrt{m})^2 = 4,$$

and the variance of the working value

$$\sqrt{m} + \frac{a - m}{2\sqrt{m}} = \frac{a + m}{2\sqrt{m}}$$

is exactly 1/4, for all values of m.

For a *normal* distribution the amount of information is the inverse of the variance; for other distributions this reciprocal equivalence does not hold, and the constancy of the amount of information supplied about $\sqrt{m}$ does not imply that the variance of the distribution of $\sqrt{a}$, where a is a Poisson variate will be constant. Such a connection is only to be looked for in large samples where the Poisson distribution approaches the normal.

In spite of the exact constancy of the amount of information which should, I think, have served as a warning, certain authors (a) thinking that the approximate constancy of the variance of $\sqrt{a}$ was the object of the transformation, and (b) observing that such constancy is imperfect, have suggested various troublesome modifications, which have now been available for some time. Thus in 1936 Professor M. S. Bartlett (2) in a paper entitled *The square root transformation and the analysis of variance,* proposed the use of $\sqrt{\alpha + 1/2}$, when α is an observed variate, and therefore a sufficient estimate of the parameter m of a Poisson Series, and pointed out that its sampling variance was somewhat more constant than that of $\sqrt{\alpha}$. Twelve years later, F. J. Anscombe in Biometrika (6) indicated that $\sqrt{\alpha + 3/8}$ was even better, at least when m is large, a result he ascribed to A. H. L. Johnson. Neither author seemed to realise that in the correct process of using these transformations, as set out above, the variance of each working value is exactly equal to $1/i$, and needs no adjustment.

It is difficult to judge just what influenced Bartlett in putting forward his proposal for adjusting the Poisson variate. In his 1936 paper he compares the addition of 1/2 to the variate before taking the square root, to Yates's correction for continuity, which Yates had introduced for tests of significance with χ^2 having one degree of freedom, but the comparison is very tenuous. Yates had the exact test of significance at the time, and could demonstrate empirically that his adjustment, (a) was easy to apply, and (b) did in fact improve the test of significance. Bartlett does not attempt to show that an improved analysis of variance results from his adjustment. In the more general case of the angular transformation he suggests

$$\sin^{-1}\sqrt{(a \pm \tfrac{1}{2})/n}$$

the signs being determined by whether a is less or greater than $n/2$. This awkward form was replaced by Anscombe (1948), by the more rational proposal to use

$$\sin^{-1}\sqrt{(a + \tfrac{3}{8})/(n + \tfrac{3}{4})},$$

which is at least consistent over the whole range of observations.

In seeking a transformation having constant variance, Bartlett may also have been influenced by the transformation of the correlation coefficient,

$$r = \tanh z,$$

which I had shown in 1921 (1) to give distributions for z sufficiently nearly normal for the use of the Gaussian distribution in tests of significance and sufficiently constant in variance for the (unknown) true correlation to be an unimportant factor in such tests. These advantages are such that for practical purposes, tabulation of the exact distribution has been entirely unnecessary, but to suppose that there are corresponding advantages in attempting to make the variance of some function of a Poisson variate as constant as possible, suggests that the non-normal character of this discontinuous distribution has been ignored, and even that it was proposed to use the empirical transforms as variates in the final analysis.

The unsuitability of using empirical transforms was early made clear in the case of Probits, where experiments in which all of the tests react alike would empirically be given infinite variates with zero weights. The exact treatment was given in 1935 in an Appendix to C. I. Bliss (9). The full table for obtaining the correct working values was given by Bliss (3) in 1938 *The determination of the dosage mortality curve from small numbers.* In the same year Fisher and Yates (8) *Tables for Statisticians,* gave corresponding tables both for the Probit, and for the angular transformation, in which the need for the use of a correct working variate had not been forced on the notice of statisticians by the appearance of infinite values.

In 1940 W. G. Cochran (4) considered Bartlett's adjustment in a paper in the Annals of Mathematical Statistics. He refers to the method exhibited in the preface of *Tables for Statisticians,* and for which provision had been made in tables XII and XIV, and evidently recognises this as more correct than the use of empirical angles. Yet he seems to assume that the latter may be used without inaccuracy save in special and particularly in terminal cases, for which however he mentions and does not totally disavow the proposal (11) to substitute 1/4 for 0 and $n - 1/4$ for n. It is a great pity that Cochran in this paper does not clearly point out that such adjustments have no useful function, at least finally, if it is intended to perform a correct analysis. The subsequent papers (5, 6) by Bartlett (1947) and Anscombe (1948), show no such consciousness of the situation as they would have obtained had Cochran expressed himself more definitely.

Arising from the idea that the empirical transforms can be used in the final analysis, instead of being always of a tentative and provisional character, is the emphatically advocated notion that great differences in computational effort are required in the use of different transformations. In particular the logistic transformation has been advocated by J. Berkson, as though this were a major consideration, and in a recent review of Finney's *Probit Analysis* (J.A.S.A. 47, 687), K. A. Brownlee (10) repeats Berkson's extraordinary claim that the logistic curve can be fitted thirty times as rapidly as the normal. I have fitted many cases of both over the last fifteen years, and there is little to choose between the two procedures, in cases that require careful fitting, i.e. when the different test batches are broken up in small groups, as must often happen when many factors are brought into the analysis. It is true that the working Logits, and their precision are given by simple formulae which need no special tabulation beyond the use of a readily available table of hyperbolic tangents, but the work of calculating each value is not less or appreciably different from that of looking up the values appropriate to other transformations, in the tables already available. In no case are the computations unduly onerous, and they are as likely to be lengthy with logits as with the other transforms, if the number of classes is large. To choose one transformation rather than another on the supposition that the labour will be less, without regard to its conformity with theoretical considerations, seems to be a very mistaken policy, seeing that the estimates, which are always an intrinsic part of the analysis of variance, are in such cases estimates only of mathematical artifacts. The appropriateness of our choice is, however, open to confirmation by the χ^2 test, and where this test shows the angular transformation to have been usually successful in like material, we may gain some real computational advantage by assigning in advance equal or proportional weights to the different entries in a two-way or three-way analysis.

BIBLIOGRAPHY

(1) R. A. Fisher (1921). On the "probable error" of a correlation coefficient deduced from a small sample. *Metron I, 4,* 1-32.
(2) M. S. Bartlett (1936). The square root transformation in the analysis of variance. *J. Roy. Stat. Soc., Suppl., 3.1,* 68-78.
(3) C. I. Bliss (1938). The determination of the dosage mortality curve from small numbers. *Q. J. of Pharm. and Pharmacol., 11,* 192-216.
(4) W. G. Cochran (1940). The analysis of variance when experimental errors follow the Poisson or Binomial laws. *Annals Math. Stat., 40,* 335-347.
(5) M. S. Bartlett (1947). The use of transformations. *Biometrics, 3,* 39-52.
(6) F. J. Anscombe (1948). The transformation of Poisson, Binomial and negative-binomial data. *Biometrika 35,* 246-254.
(7) R. A. Fisher (1950). Gene frequencies in a cline determined by selection and diffusion. *Biometrics, 6,* 353-361.
(8) R. A. Fisher and F. Yates (1938). *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd.
(9) C. I. Bliss (1935). The calculation of the dosage-mortality curve. Appendix by R. A. Fisher. *Ann. Applied Biol., 22,* 134-167.
(10) K. A. Brownlee (1952). Review of Finney's *Probit Analysis. J.A.S.A., 47,* 687.
(11) C. Eisenhart (1947). *Selected techniques of statistical analysis.* Columbia Univ., New York.