A Simple Example of a Comparison Involving Quantal Data

D. R. Cox

Biometrika, Vol. 53, No. 1/2. (Jun., 1966), pp. 215-220.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28196606%2953%3A1%2F2%3C215%3AASEOAC%3E2.0.CO%3B2-S

Biometrika is currently published by Biometrika Trust.

Biometrika (1966), **53**, 1 *and* 2, *p.* 215
Printed in Great Britain

A simple example of a comparison involving quantal data

BY D. R. COX

Bell Telephone Laboratories Incorporated, Murray Hill, New Jersey
and Birkbeck College, London

SUMMARY

An example is given of the comparison of two matched groups of observations in which the response is quantal. Alternative methods of analysis are sketched. Under a logistic model, the theoretically appropriate procedure is a generalization of Fisher's 'exact' method for the 2×2 contingency table; Cochran (1954) gave the corresponding large-sample significance test.

1. INTRODUCTION

Data in which the response variable has only two possible forms, success or failure, defective or non-defective, etc., arise in many fields. The present note is concerned with a particular comparison problem occurring with this sort of data. The problem could be formulated fairly generally, but the discussion here will be very largely in terms of the following specific example.

Table 1 gives data of Gordon & Foss (1966) on the effect of rocking on the crying of very young babies. On each of 18 days, the babies not crying at a certain instant in a hospital nursery served as subjects. One baby selected at random was rocked for a set period, the remainder serving as controls. The numbers not crying at the end of a specified period are

Table 1. *Gordon & Foss's data on the crying of babies*

Day	No. of control babies	No. not crying	No. of experimental babies	No. not crying	Score, y
1	8	3	1	1	0·62
2	6	2	1	1	·67
3	5	1	1	1	·80
4	6	1	1	0	·17
5	5	4	1	1	·20
6	9	4	1	1	0·56
7	8	5	1	1	·38
8	8	4	1	1	·50
9	5	3	1	1	·40
10	9	8	1	0	−·89
11	6	5	1	1	0·17
12	9	8	1	1	·11
13	8	5	1	1	·38
14	5	4	1	1	·20
15	6	4	1	1	·33
16	8	7	1	1	0·12
17	6	4	1	0	−·67
18	8	5	1	1	·38

given in Table 1. Here, the data will be used solely to illustrate procedures of analysis; for the interpretation of the results, see Gordon & Foss's paper.

The data have the form of a paired comparison experiment with quantal response. There is no information as to the extent to which the same infant enters the experiment on a number of days; we shall treat responses on different days as independent.

2. Simple analysis

A possible analysis is to pool the data from the different days into the single 2×2 contingency table of Table 2. Standard methods for estimation and testing can then be applied. In particular, chi-squared with one degree of freedom, corrected for continuity, is 2·36, corresponding to a one-sided significance level of 0·063. It is convenient for comparison to give one-sided levels.

A criticism of this approach is that it ignores the grouping into days; this can lead both to systematic error and to underestimation of precision.

Table 2. *Contingency table formed from data of Table* 1

	Not crying	Crying	Total
Control group	77	48	125
Experimental group	15	3	18

One preferable simple procedure is to assign a 'score' for each day having an expectation measuring the difference between treatments. A simple way to do this is to define y_i for the ith day to be the difference between the proportions of successes (i.e. non-criers) in the experimental and control groups. We then consider an idealized model in which, however conditions may vary from day to day, the difference between groups in the probability of success remains constant, with value Δ, say. Then $E(y_i) = \Delta$, it is reasonable to treat the y_i's on different days as independent and, as an approximation, we shall regard the y_i's as having the same, but unknown, variance.

Standard methods for the analysis of a simple random sample can now be used. The estimated mean and variance are 0·23 and 0·189, so that Δ is estimated to be 0·23 with an estimated standard error of 0·11. If we use the standard t tables, the departure from the null hypothesis $\Delta = 0$ is significant at about the one-sided 2 % level. The analysis could, if appropriate, be extended to include the regression of the y_i's on suitable independent variables.

One advantage of this analysis is that it avoids assumptions, largely unverifiable, about the independence of different outcomes within one day. If, however, we postulate on the ith day probabilities θ_{1i}, θ_{2i} of success in the two groups, with the conditions of binomial sampling holding, the variance of y_i is

$$\frac{\theta_{1i}(1-\theta_{1i})}{n_{1i}} + \frac{\theta_{2i}(1-\theta_{2i})}{n_{2i}}. \tag{1}$$

Here n_{1i}, n_{2i} are the number of observations in experimental and control groups on the ith day; in the present application, $n_{1i} = 1$. The average of (1) over i is estimated in general by

$$\frac{1}{k}\sum_{i=1}^{k} \frac{r_{1i}(n_{1i}-r_{1i})}{n_{1i}^2(n_{1i}-1)} + \frac{1}{k}\sum_{i=1}^{k} \frac{r_{2i}(n_{2i}-r_{2i})}{n_{2i}^2(n_{2i}-1)}, \tag{2}$$

where k is the number of days and r_{1i}, r_{2i} are the numbers of successes on day i.

Formula (2) requires that all the sample sizes exceed unity and so cannot be applied to the example. The second term of (2) has the value 0·039 and the first and predominant term can be estimated approximately by

$$\frac{\Sigma r_{1i}}{k}\left(1-\frac{\Sigma r_{1i}}{k}\right), \tag{3}$$

which is 0·139. The theoretical variance is thus approximately 0·178 in good agreement with the observed variance of 0·189.

It may be objected that the use of the t distribution is invalid because the y_i's are not normally distributed. It is known, however, that the normality assumption is not critical. In the present case, where the n_{1i} and n_{2i} are very unequal and small, and the significance borderline, it is worth making an approximate assessment of the effect of non-normality. By the same sort of argument as used for (1) and (2) the third cumulant ratio γ_1 is estimated to be $-0{\cdot}7$; and the kurtosis γ_2 is negative. The tables of Gayen (1949) now show that the approximate effect of non-normality is to at most double the one-sided significance level, i.e. to convert the 2 % level into the 4 % level. The skewness arises because of the very unequal sample sizes in the two groups combined with the occurrence of probabilities appreciably different from one-half. Even so, the effect on the significance level is rather minor.

Note that the application of a non-parametric procedure such as the sign or Wilcoxon test would not be appropriate because the population mean score corresponds to the parameter Δ and is zero on the null hypothesis. The population median, on the other hand, has no such interpretation and is not in general zero when $\Delta = 0$.

The method of analysis based on the scores y_i is simple, direct and capable of generalization. There is, however, some loss of information, especially from treating all the y_i with equal weight. This would be serious if the different days contained very different numbers of observations. Also, the model in which the difference of probabilities is constant can at best hold over a limited range.

In the next section, we give an extension to this problem of Fisher's 'exact' treatment of the 2×2 contingency table. The analysis will be based on a model different from the above, in which the treatment effect is constant on the logistic scale rather than on a direct probability scale.

3. Logistic analysis

A central role in the analysis of quantitative data is played by linear models. If we had quantitative responses from the design of Table 1, we might interpret the data in the light of a model in which, possibly after transformation, an observation on the ith day has expectation

$$\beta_i+\lambda \quad \text{or} \quad \beta_i-\lambda, \tag{4}$$

depending on which treatment has been used.

In many ways, the simplest analogous model for quantal data is additive on a logistic scale. That is

$$\log\left(\frac{\theta_{1i}}{1-\theta_{1i}}\right)=\beta_i+\lambda, \quad \log\left(\frac{\theta_{2i}}{1-\theta_{2i}}\right)=\beta_i-\lambda. \tag{5}$$

If we assume that all trials are independent with probability of success given by (5), the

likelihood of the observations can be obtained. Yates (1955) gave the maximum-likelihood analysis of a number of problems connected with this sort of data, under the logistic and other models. In the following, rather simple methods are given for inference about λ.

The likelihood of the observations is, in the notation of § 2,

$$\prod_{i=1}^{k} \frac{\exp\{(\beta_i+\lambda)r_{1i}+(\beta_i-\lambda)r_{2i}\}}{(1+e^{\beta_i+\lambda})^{n_{1i}}(1+e^{\beta_i-\lambda})^{n_{2i}}}. \tag{6}$$

It follows from (6) that the jointly sufficient statistics for $(\beta_i, \ldots, \beta_k, \lambda)$ are the total numbers of successes per day, i.e. $r_{1i}+r_{2i}$ $(i = 1, \ldots, k)$ and $\Sigma(r_{1i}-r_{2i})$. The last quantity can equivalently be replaced by $r_{1.} = \Sigma r_{1i}$, the total number of successes in the experimental group.

Further, if we are concerned with inference about λ, regarding $\beta_1, \ldots, \beta_k$ as nuisance parameters, the usual conditionality arguments lead to the consideration of the observed value of $r_{1.}$ relative to the conditional distribution of the corresponding random variable $R_{1.}$, given $r_{.i} = r_{1i}+r_{2i}$ $(i = 1, \ldots, k)$.

We shall for simplicity consider mainly the null case, $\lambda = 0$. Given $r_{.i}$, the random variable R_{1i} giving the number of successes on the ith day has a hypergeometric distribution; it is the test statistic of Fisher's 'exact' test applied to the observations on the ith day. Thus the generating function is

$$G_i(z) = \Sigma_x \frac{\binom{r_{.i}}{x}\binom{n_{.i}-r_{.i}}{n_{1i}-x}}{\binom{n_{.i}}{r_{.i}}} z^x \tag{7}$$

and

$$E(R_{1i}) = n_{1i}r_{.i}/n_{.i}, \tag{8}$$

$$\text{var}\,(R_{1i}) = \frac{n_{1i}r_{.i}(n_{.i}-r_{.i})(n_{.i}-n_{1i})}{n_i^2(n_{.i}-1)}. \tag{9}$$

Since the data on different days are assumed independent, the required distribution of $R_{1.} = \Sigma R_{1i}$ has generating function

$$G(z) = \Pi G_i(z), \tag{10}$$

and

$$E(R_{1.}) = \Sigma E(R_{1i}), \quad \text{var}\,(R_{1.}) = \Sigma\,\text{var}\,(R_{1i}). \tag{11}$$

Because of the central limit theorem, the distribution of $R_{1.}$ will often be adequately approximated by a normal distribution with a continuity correction, and in fact a large-sample test based on $R_{1.}$ was obtained by Cochran (1954); see Birch (1964) for a general discussion of the detection of partial association in contingency tables and for its connexion with logistic models.

The distribution for $\lambda \neq 0$ is similarly the convolution of a number of generalized hypergeometric distributions (Fisher, 1935); see Gart (1962) for an account of methods for approximating to the distribution, several of which could be adapted to the present problem. For the remainder of the present paper, we concentrate on the null hypothesis distribution.

There are several special cases of (9)–(11). If $k = 1$, we have the results for a single 2×2 contingency table. If $n_{1i} = n_{2i} = 1$, we have a set of matched pairs. The test based on $R_{1.}$ is equivalent (Cox, 1958) to McNemar's (1947) test in which pairs in which both individuals give the same response are rejected. Both null and non-null distributions of $R_{1.}$ are binomial.

As a final special case, return to the example of § 1. Here $n_{1i} = 1$, $n_{2i} > 1$. The special feature now is that the individual R_{1i} are (0, 1) random variables with

$$G_i(z) = (q_i + p_i z), \tag{12}$$

where

$$p_i = \frac{r_{.i}}{n_{2i}+1} \quad (q_i = 1 - p_i), \tag{13}$$

$$E(R_{1i}) = p_i, \quad \text{var}\,(R_{1i}) = p_i q_i. \tag{14}$$

Thus, for $R_{1.}$

$$G(z) = \Pi(q_i + p_i z), \tag{15}$$

$$E(R_{1.}) = \Sigma p_i, \quad \text{var}\,(R_{1.}) = \Sigma p_i q_i. \tag{16}$$

In the particular example,

$$E(R_{1.}) = 6{\cdot}53, \quad \text{var}\,(R_{1.}) = 3{\cdot}420.$$

The observed value $r_{1.} = 3$, so that the standardized value corrected for continuity is $3{\cdot}03/\sqrt{3{\cdot}42} = 1{\cdot}64$, corresponding in the normal approximation to a one-sided test area of 0·050.

Since the distribution of $R_{1.}$ is in this case positively skew, the normal approximation overestimates the true probability. It is possible to develop more refined approximations based on (15). Alternatively, if the adequacy of the normal approximation is in genuine doubt, it is possible to obtain by computer the coefficients of the relevant powers of z in (15); a recursion formula for the coefficients in a product of k terms in terms of the corresponding coefficients in a product of $k-1$ terms can be used. In the present example, the coefficients of z^0, z^1, z^2, z^3 were obtained in this way; this sum, the one-sided significance level, is 0·045.

4. Discussion

The significance levels corresponding to a number of methods of analysis have been calculated fairly precisely for the example. This has been done to illustrate technique and it is not suggested that in practice the very precise evaluation of significance levels is called for. The different methods agree quite closely in the present case.

The analysis of § 3 is based on a rather special model for the data. However, the resulting significance test for the null hypothesis $\lambda = 0$ can be regarded as a pure randomization test, and as justified under the null hypothesis that the response of any individual does not depend on the treatment used. This is an advantage since in the present example assumptions about the probability of success on a particular day and about the independence of different trials are largely unverifiable.

With more extensive data, further problems would arise, such as examining the constancy of λ in the model (5); see Yates (1955) for a thorough analysis and discussion of such matters, based on maximum-likelihood arguments. Radhakrishna (1965) has made an interesting comparison of the asymptotic relative efficiencies of tests based on the assumptions of constant differences on the logistic, normal and other scales.

Essentially the same methods apply to more complex comparisons involving quantal responses.

I am grateful to Professor Foss and Miss Gordon for asking the question that led to this paper, to Mr R. L. Harris for some of the calculations, and to Dr J. J. Gart for some very helpful comments on a preliminary version of the paper.

References

Birch, M. W. (1964). The detection of partial association, I. The 2×2 case. *J. R. Statist. Soc.* B, **26**, 313–24.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**, 417–51.

Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika*, **45**, 562–5.

Fisher, R. A. (1935). The logic of inductive inference. *J. R. Statist. Soc.* **98**, 39–54.

Gart, J. J. (1962). Approximate confidence levels for the relative risk. *J. R. Statist. Soc.* B, **24**, 454–63.

Gayen, A. K. (1949). The distribution of Student's t in random samples of any size drawn from non-normal. *Biometrika*, **36**, 353–69.

Gordon, T. & Foss, B. M. (1966). The role of stimulation in the delay of onset of crying in the newborn infant. *J. Exp. Psychol.* (to appear).

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–7.

Radhakrishna, S. (1965). Combination of results from several 2×2 contingency tables. *Biometrics*, **21**, 86–98.

Yates, F. (1955). The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. *Biometrika*, **42**, 382–403.