A NOTE ON AN UNCONDITIONAL ALTERNATIVE TO COCHRAN'S *Q*

D.J. BEST[1] AND J.C.W. RAYNER[1,2]

[1] School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia

[2] Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia

Abstract

We give an example to demonstrate that a conditional statistical test can give a possibly different p-value to an unconditional test, thereby altering the statistical conclusion. This same example also gives different p-values for the associated permutation test and parametric bootstrap test. When there is a choice, whether a test is conditional or unconditional may be important.

Keywords: binary data; parametric bootstrap; permutation test; randomised block design

2010 Mathematics Subject Classification: Primary 62G10

Secondary 62-07

1. Introduction

An unconditional alternative test to that based on Cochran's Q is based on the statistic S = $\{t/(t-1)\}Q$ in which t is the number of treatments or products being compared. The statistic Q is derived using the hypergeometric distribution and the statistic S is derived using a product Bernoulli model. Using the former implies marginal totals are fixed so that if an experiment or study was repeated the marginal totals for a 'success' would be the same. While inference based on this condition is valid it is thought by many statisticians that it can be improved upon by using an unconditional statistic. See [1].

We observe that if Monte Carlo or exact methods are used to obtain an accurate unconditional test based on S then the product Bernoulli method should be used to generate random data values or exact probabilities. A permutation test or hypergeometric probabilities should *not* be used. In the following we use the χ^2_{t-1} distribution and a parametric bootstrap to get p-values for S.

We now give a general definition of Q and specific definitions for t = 2 and 3. Suppose we have a randomised block design in which t treatments are applied to r blocks. Let X_{ij} be the outcome for treatment i on block j, and suppose that X_{ij} = 1 if the outcome is a 'success', and X_{ij} = 0 otherwise. Using standard 'dot' notation, $\sum_{i=1}^{t} X_{ij} = X_{\bullet j}$ and $\sum_{i=1}^{t}\sum_{j=1}^{r} X_{ij} = X_{\bullet\bullet}$. Cochran's Q is given by

$$Q = t(t-1)\frac{\sum_{i=1}^{t}(X_{i\bullet} - X_{\bullet\bullet}/t)^2}{\sum_{j=1}^{r} X_{\bullet j}(t - X_{\bullet j})^2}.$$

For $t = 2$ we can also define Q in terms of the frequencies n_1, n_2, n_3 and n_4 for the responses (1, 1), (1, 0), (0, 1) and (0, 0). This gives

$$Q = (n_2 - n_3)^2/(n_2 + n_3).$$

Similarly for $t = 3$ we can also define Q in terms of the frequencies $n_1, \ldots, n_8$ for the responses (1, 1, 1), (1, 1, 0), (1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1) and (0, 0, 0). This gives

$$3(n_2 + n_3 + n_4 + n_5 + n_6 + n_7)Q = (n_2 + n_3 + 2n_4 - 2n_5 - n_6 - n_7)^2$$
$$+ (n_2 - 2n_3 - n_4 + n_5 + 2n_6 - n_7)^2 + (-2n_2 + n_3 - n_4 + n_5 - n_6 + 2n_7)^2.$$

2. Example

The StatXact 6 User Manual [2] considers three treatment, three period cross-over clinical data for 11 patients. The three treatments were placebo, aspirin and a new drug. The data were binary with success and failure being the categories and the results for $(n_1, \ldots, n_8)$ were (0, 0, 2, 0, 3, 1, 3, 2).

To reinforce the meaning of a conditional test here, the outcome for the first patient is as in Table 1. The other patients have similar tables. Before applying the treatments we know that each patient will receive each of the three treatments. The unconditional test assumes no knowledge of the outcomes. The conditional test assumes that if the experiment were repeated (as, for example, in a permutation test) the first patient would have two successes and one failure. Similarly the other patients would have the same number of successes in every repetition. To many this is an unnatural restriction.

Routine calculation yields $Q = 6.222$ and $S = tQ/(t - 1) = 9.333$. P-values are as shown in Table 2. Two of the p-values are significant at the 0.05 level, one is significant at the 0.01 level, and one is not significant at the 0.05 level. Which p-value should be used?

TABLE 1. Data for the first patient

Patient 1	Success	Failure	Total
Placebo	0	1	1
Aspirin	1	0	1
New drug	1	0	1
Total	2	1	3

TABLE 2. P-values for the three treatments

P-value type	Statistic	P-value
Asymptotic χ^2_2	Q	0.045
Permutation test	Q	0.059
Asymptotic χ^2_2	S	0.009
Parametric bootstrap	Q or S	0.039

Before giving our suggested answer to this question, we briefly describe the parametric bootstrap procedure used. For each patient a probability of success was calculated. There are four possibilities: $p = 0$, 1/3, 2/3 and 1. Ignoring the $p = 0.0$ and $p = 1.0$ cases as these do not affect the value of Q, we calculate 10,000 data sets with nine values where five of these were sets of three random values from a Bernoulli distribution with probability of success 2/3 and four were three random values from a Bernoulli distribution with probability of success 1/3. For each of the 10,000 data sets a value of Q was calculated and the proportion greater than or equal to 6.222 was calculated. This proportion was 0.039 as in Table 1 above. The proportion of S values greater than or equal to 9.333 was, of course, also 0.039.

The reader should be able to generalise the procedure given to other, similar, data sets.

On p.364 of [2] it is stated that the asymptotic result of 0.045 for Q is erroneous in that 0.05 significance cannot be claimed because the permutation test result is 0.059. However this result is only true for a conditional test. If an unconditional parametric bootstrap test is used the p-value is 0.039 which correctly gives 0.05 significance. We suggest many statisticians would be happy to accept this unconditional p-value which indicates there is a treatment difference at the 0.05 level.

We also see that the asymptotic χ_2^2 p-value for S is 0.009. This is somewhat more liberal than the parametric bootstrap value. We suggest that with modern computing facilities the parametric bootstrap p-value should be calculated for all but quite large data sets.

References

[1] RAYNER, J.C.W. and BEST, D.J. (2017). Unconditional analogues of Cochran–Mantel–Haenszel tests. To appear in the *Australian & New Zealand Journal of Statistics*.

[2] STATXACT version 6 (2003). User Manual Volume 2. CYTEL Software, Cambridge, MA.