Review Article

Open Access

An alternative approach to cochran Q test for dichotomous data

Abstract

This paper proposes, develops and presents a statistical method for the analysis of sample data where these data are dichotomous responses assuming only two possible mutually exclusive values such as 1, representing positive response say, and 0 representing negative response say. The Chi-square test statistic based on the Chi-square test for independence is developed as an alternative to the usual Cochran Q test for dichotomous data to test the null hypothesis of equal positive response rate by subjects to a set of test or treatment. The proposed method is illustrated with some sample data and shown to be at least as powerful as the usual Cochran Q test when applied to the same sample observations.

Keywords: chi-square, dichotomous data, cochran q test, subjects, responses, success

Volume 4 Issue 4 - 2016

Okeh UM,[1] Oyeka ICA,[2] Igwenagu CM[3]

[1]Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Nigeria
[2]Department of Statistics, Nnamdi Azikiwe University, Nigeria
[3]Department of Industrial Mathematics and Applied Statistics and Demography, Enugu State University of Science and Technology, Nigeria

Correspondence: Okeh UM, Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Nigeria, Email uzomaokey@ymail.com

Received: April 03, 2016 | **Published:** May 07, 2016

Introduction

Sometimes a researcher may perform an experiment involving repeated observations or blocks, in which the variable of interest is dichotomous, meaning that it can assume only one of two possible mutually exclusive values. One of these two possible values is considered a ‘success’, positive response, present, well, etc. This is often coded as ‘1’,while the other value may be considered a failure, negative response, absent, bad, etc often coded a ‘0’.[1]

Research interest would then be to determine the proportions of subjects responding positive if the sampled blocks of subjects are the same across all treatments or tests. In this situation Cochran Q test[2–4] for the dichotomous data may be applied.

To adjust, and make allowance for some situations in which test or trial outcomes are not just dichotomous assuming only two mutually exclusive options such as 1 or 0,but when there may be some intermediate and third outcome such as unknown, indeterminate, non-definitive, etc, that may be code with say a minus sign(-). Oyeka, et al.,[5] introduced a third category of response and developed a Chi-square (χ^2) test statistic for independence to test the null hypothesis of equal positive response rates under various treatments or test.

We will here however construct and alternative test statistic similar to Cochran Q test assuming that there are only two possible response outcomes or options that may be coded as either 1 or 0.

Methods

The proposed method would then be compared with the usual Cochran Q test as shown below.

The proposed method

Sometimes a researcher may be interested in comparing responses of blocks of subjects to a set of treatments in a diagnostic screened test or clinical trials. Specifically, suppose a researcher has collected a random sample of n block of subjects matched on some demographic characteristics such as age, sex, body weight, etc, where each block of subjects contain some c matched subjects and interest of the researcher is to administer each of these c subjects randomly one of c treatments. The subject responses to each of the treatments are all dichotomous assuming only one of two possible and mutually exclusive response options such as positive, or negative, present or absent, good or bad; success or failure, dead or alive, etc.

Let x_{ij} be the response by a randomly selected subject from the i[th] block of subjects administered treatment T_j for i=1,2,…,n;j=1,2,…,c where each x_{ij} is dichotomous, either positive or negative, present or absent etc. To develop a test statistic to help determine whether on the average subject responses are the same for all treatments or conditions we may

Let

$$u_{ij}=\begin{cases}1, \text{if a randomly selected subject from the ith block of subjects respond} \\ \quad \text{positive,indicating condition present,good when administered treatment } T_j \\ 0, \text{if these same subject responds negative,indicating condition absent,bad when} \\ \quad \text{adminiterd treatment } T_j\end{cases} \quad (1)$$

For i=1,2,..,n; j=1,2,..c.

Let

$$W_j = f_{.j}^{+} = \sum_{i=1}^{n} u_{ij} \quad (2)$$

and

$$W_j = f_{.j}^{+} = \sum_{i=1}^{n} u_{ij} \quad (3)$$

Be the total number of positive responses, that is total number of 1's by subjects administered treatment T_j and

$$f_{.j}^{0} = n - f_{.j}^{+} \quad (4)$$

Be the total number of negative response that is number of 0's obtained when subjects are administered treatment T_j.

Now the total number of positive responses, that is the total number of 1's and the total number of negative responses, that is the total number of 0's for all the c treatments are respectively

$$f_{..}^{+} = \sum_{j=1}^{c} f_{.j}^{+}; f_{..}^{0} = \sum_{j=1}^{c}\left(n - f_{.j}^{+}\right) = nc - f_{..}^{+} \quad (5)$$

Note that the expected value and variance of u_{ij} are respectively

$$E\left(u_{ij}\right) = \pi_{.j}^{+}, Var\left(u_{ij}\right) = \pi_{.j}^{+}\left(1 - \pi_{.j}^{+}\right) \quad (6)$$

The sample estimate of π_j^{+} is

$$\hat{\pi}_{.j}^{+} = p_{.j} = \frac{f_{.j}^{+}}{n} \quad (7)$$

Similarly, the overall sample proportion of positive responses for all the c treatments is

$$p_{..} = \frac{f_{..}^{+}}{nc} \quad (8)$$

To obtain a test statistic for the null hypothesis that subjects on the average do not differ in their positive responses to all the c treatments, it would be easier to use the Chi-square test for independence.

Under this approach, we note that the observed frequencies of positive and negative responses by subjects at treatment T_j that is the total number of 1's and 0's at treatment T_j are respectively

$$o_{1j} = f_{.j}^{+}; o_{2j} = f_{.j}^{0} = n - f_{.j}^{+} \quad (9)$$

Now, under the null hypothesis of equal positive responses for all the c treatments, the expected frequencies of positive and negative responses for T_j are respectively

$$E_{1j} = \frac{n.f_{..}^{+}}{nc} = \frac{f_{..}^{+}}{c}; O_{1j} = \frac{n.f_{..}^{0}}{nc} = \frac{f_{..}^{0}}{c}; O_{2j} = \frac{nc - f_{..}^{0}}{c}; E_{2j} = \frac{nc - f_{..}^{+}}{c} \quad (10)$$

Hence, the required χ^2 test statistic is

$$\chi^2 = \sum_{j=1}^{c}\sum_{i=1}^{2}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad (11)$$

With c-1 degrees of freedom under the null hypothesis H_0

Substituting equations (9) and (10) in equation (11), we obtain

$$\chi^2 = \sum_{j=1}^{c}\left(f_{.j}^{+} - \frac{f_{..}^{+}}{c}\right) \Big/ \frac{f_{..}^{+}}{c} + \frac{\sum_{j=1}^{c}\left((n - f_{.j}^{+}) - \frac{(nc - f_{..}^{+})}{c}\right)^2}{\frac{nc - f_{..}^{+}}{c}}$$

Which when further simplified becomes

$$\chi^2 = \frac{n.c^2}{f_{..}^{+}\left(nc - f_{..}^{+}\right)}\sum_{j=1}^{c}\left(f_{.j}^{+} - \frac{f_{..}^{+}}{c}\right) \quad (12)$$

Which under the null hypothesis H_0 of equal positive response rate has approximately the χ^2 distribution with c-1 degrees of freedom for sufficiently large n and c. Equation (12) when expressed in terms of sampled proportion becomes

$$\chi^2 = \frac{n}{p_{..}(1 - p_{..})}\sum_{j=1}^{c}\left(p_{.j}^{+} - p_{..}\right)^2 = \frac{n\left(\sum_{j=1}^{c}\left(p_{.j}^{+2} - p_{..}^{2}\right)\right)}{p_{..}(1 - p_{..})} \quad (13)$$

Illustrative example: The effects of four drug presentations on patients are to be studied. Interest is to determine whether or not the four drugs equally improved patients' condition. Sixty patients are selected and grouped into 15 blocks so that the four patients in each block are approximately identical in age, initial condition, sex, etc. Patients in each block are randomly selected for treatment with only one of the four experimental drugs. After the specified medication period, the patients are classified as either improved (success) or not improved (failure) under a given drug and coded with a 1 or 0 respectively. The results are shown in Table 1. Can it be concluded on the basis of these data that patients improve equally on all the four drugs?

An illustration of proposed alternative to Cochran Q test: To illustrate the proposed method we apply equation 1 to obtain values of u_{ij} for the data of Table 1,for i=1,2,…,15;j=1,2,…,4.The summary values of u_{ij} that is of $f_{.j}^{+}$ *and* $f_{.j}^{0}$ with the corresponding sample proportions $p_{.j}$ of 1's are shown at the bottom of Table 1, for j=1,2,3,4. Using the values of $p_{.j}$ in equation 13 we obtain the Chi-square test statistic for the null hypothesis of no difference in possible response rates, that is the proportion of subjects, or patients improving under the four drugs as

$$\chi^2 = \frac{(15)\left((0.60)^2+(0.67)^2+(0.33)^2+(0.33)^2-40(0.48)^2\right)}{(0.48)(0.52)}$$

$$\chi^2 = \frac{(15)(1.027-0.922)}{0.250} = \frac{(15)(0.105)}{0.25} = \frac{1.575}{0.25} = 6.300(p-value = 0.0984),$$

Which with 4-1=3 degrees of freedom is not statistically significant $\left(\chi^2_{0.99;3}=11.35\right)$, leading to the conclusion that patients may not have deferred in their improvement rates over the four drugs.

Table 1 Patients' response to four drug preparations (1=improved, 0=Not improved)

	Drugs(treatments)				
Patients(Blocks)	**Drug 1**	**Drug 2**	**Drug 3**	**Drug 4**	**Total**
1	1	1	0	0	2
2	1	1	0	1	2
3	1	0	0	0	1
4	1	1	1	1	4
5	1	1	0	1	3
6	0	1	0	0	1
7	0	1	1	0	2
8	1	1	1	0	3
9	0	0	1	0	1
10	0	0	0	0	0
11	0	0	0	0	0
12	1	1	0	1	3
13	1	0	0	1	2
14	0	1	1	0	2
15	1	1	0	0	2
Total(n)	15	15	15	15	$60(n.c)$ $29(=f^{+})$
$f^{+}_{.j}$	9	10	5	5	
$f^{0}_{.j}=n-f^{+}_{.j}$	6	5	10	10	$31=(f^{0}_{..})$
$p_{.j}$	0.6	0.67	0.33	0.33	$0.48=(p_{..})$

Now that if one had instead applied the usual Cochran Q test to analyze the same data, we would obtain the required test statistic as

$$Q = \frac{(c-1)\left[\sum_{j=1}^{4} T_{.j}^2-(\sum_{j=1}^{4} T_{.j})^2 \Big/ c\right]}{\sum_{i=1}^{15} B_i-\left(\sum_{i=1}^{15} B_i^2 \Big/ c\right)}$$

$$=\frac{(4-1)\left((9)^2+(10)^2+(5)^2+(5)^2\right)\Big/4}{(2+3...+2)-\left((2)^2+(3)^2+...+(2)\right)^2\Big/4}$$

$$=\frac{(3)(231-210.25)}{29-18.75} = \frac{62.25}{10.25} = 6.073(p-value = 0.1057)$$

Which with 4-1=3 degrees of freedom is also not statistically significant $\left(\chi^2_{0.99;3}=11.35\right)$.However the Chi-square value of 6.073 obtained using the usual Cochran Q test is slightly less than the corresponding Chi-square value of 6.300 obtained using the proposed method. Hence the usual Cochran Q test statistic is likely to lead to an acceptance of false null hypothesis (Type II error) more frequently and is therefore likely to be less powerful than the present method at least for the present data.

Summary and conclusion

We have discussed and presented above an alternative modified and probably easier method for the analysis of sample data that may be appropriate for use with the usual Cochran Q test.

A test statistic based on the Chi-square test for independence is developed for testing the null hypothesis that subjects or blocks of matched subjects on the average do not differ in proportions responding positive when administered a number of tests or treatments in a diagnostic screening test or clinical trial.

Citation: Okeh UM, Oyeka ICA, Igwenagu C. An alternative approach to cochran Q test for dichotomous data. *MOJ Public Health.* 2016;4(4):114–117. DOI: 10.15406/mojph.2016.04.00086

The proposed test method is illustrated with some sample data and shown to be at least as powerful as the usual Cochran Q test when applied to data of equal sizes. This is because the Chi-square value of 6.073 obtained using the usual Cochran Q test is slightly less than the corresponding Chi-square value of 6.300 obtained using the proposed method. Hence the usual Cochran Q test statistic is likely to lead to an acceptance of false null hypothesis (Type II error) more frequently and is therefore likely to be less powerful than the present method at least for the present data.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Oyeka ICA, Nnanatu CC. Pairwise Comparison in Repeated Measures. *Journal of Modern Applied Statistical Methods*. 2014;13(2):8.
2. Oyeka CA. *An Introduction to Applied Statistical Methods*. 8th ed. Nigeria: Nobern Avocation Publishing Company; 2010.
3. Cochran WG. The comparison of percentages in matched samples. *Biometrika*. 1950;37(3–4):256–266.
4. Freund JE. *Mathematical Statistics*. 5th ed. USA: Prentice–Hall Internal Editions; 1992.
5. Oyeka CA, Utazi CE, Nwosu CR, et al. A Statistical Comparison of Test Scores: A Non–Parametric Approach. *Journal of Mathematical Sciences*. 2010;21(1):77–87.

Citation: Okeh UM, Oyeka ICA, Igwenagu C. An alternative approach to cochran Q test for dichotomous data. *MOJ Public Health.* 2016;4(4):114–117. DOI: 10.15406/mojph.2016.04.00086