

Modified McNemar Test

I.C.A. Oyeka*

Department of Applied Statistics, Nnamdi Azikiwe University, Awka Nigeria

Abstract: This paper proposes and presents a test statistic that intrinsically and structurally adjusts the usual McNemar test statistic for the possible presence of tied responses between the paired populations of cases and control subjects that may be measurements on any scale. The method also enables the researcher readily estimates not only the chances that among a random selected pair of case and control subjects the case responds positive and the control responds negative, or the case responds negative and the control responds positive, but also even when both case and control subjects have similar responses, it enables one easily estimate the probability that both respond positive or both respond negative. The proposed method, which is shown to be relatively more efficient and hence likely to be more powerful than the usual McNemar test statistic is illustrated with some data.

Keywords: Case-control, treatment, subjects, population, procedure.

INTRODUCTION

The two sample statistical tests are used when the research interest is in determining whether there is statistical difference between two populations from which the samples are drawn. These two populations may be similar or independent, but our interest here is in when the two populations are related and not easily amenable to analysis using parametric methods because the data may not satisfy the necessary assumptions. Example of similar or related populations include those situations in which repeated measurements or observations are made on the same individuals or subjects at two different points in time or when individuals of similar characteristics are being compared [1, 2]. For example, the subjects in the two populations may be the same individuals, except that they may have received one treatment, drug, procedure, or some exposure at one occasion and then received a different treatment, drug, procedure or exposure at a second occasion. The two populations may also be such that subjects are first paired or matched on some characteristics before the administration of a procedure of interest. In each case the population which has undergone treatment or procedure is compared with the population which has undergone the other treatment or procedure to ascertain whether the two populations are statistically different. McNemar has developed a non-parametric statistical method for this purpose [3]. In this paper, we briefly present McNemar's procedure and then propose its modification.

THE McNEMAR TEST

The McNemar test for differences in responses in related or matched pairs is particularly used in study designs or experiments in which each subject serves as his own control or his own matched pair and in which measurements are in either nominal or ordinal scale. Thus it may be used to test the effectiveness of a particular treatment, drug, procedure, exposure or experience of a population of subjects. In each of these situations, subjects may serve as their own control because the experience or exposure by the subjects may be at two points in time or space. The observations being analyzed may be measurements on any scale.

To test the significance of any observed change using the McNemar test, one sets up a fourfold table of frequencies to represent the first and the second sets of responses from the same individuals. Thus suppose in a controlled clinical trial involving case and control subjects, each case is matched with a single control and the relative frequency of the outcome of interest among the cases is to be compared with that among the controls. Because of the matching of cases with controls, the proper unit of analysis is the matched pair rather than the individual subjects [4, 5]. Results from such case-control studies are usually presented in a format as in Table 1 for use with the McNemar test.

In Table 1 there are altogether $n = n_{11} + n_{12} + n_{21} + n_{22}$ pairs of case-control subjects studied. Of these n_{11} are such that case and control subjects both respond positive; n_{12} are such that the case responds positive and the control responds negative; n_{21} are such that the case responds negative and the control responds positive; and n_{22} are such that the case and control subjects both respond negative. The proportion of control subjects studied who respond positive is

*Address corresponding to this author at the Department of Applied Statistics, Nnamdi Azikiwe University, Awka Nigeria; Tel: +238052563956; E-mail: cyprianoyeka@yahoo.com

Table 1: Fourfold Table for the Presentation of Data on Matched Samples

| Case (Treatment T ₂) | Control(Treatment T ₁) | | |
|----------------------------------|--|--|---|
| | Response Positive (+) | Response Negative (-) | Total (n _i) |
| Response Positive (+) | n ₁₁ | n ₁₂ | n ₁ (=n ₁₁ +n ₁₂) |
| Response Negative (-) | n ₂₁ | n ₂₂ | n ₂ (=n ₂₁ +n ₂₂) |
| Total (n _i) | n _{.1} (=n ₁₁ +n ₂₁) | n _{.2} (=n ₁₂ +n ₂₂) | n _{..} (=n) |

$$p_1 = \frac{n_{11} + n_{21}}{n} \tag{1}$$

while the proportion of cases studied who respond positive is

$$p_2 = \frac{n_{11} + n_{12}}{n} \tag{2}$$

The difference between the proportions of case and control subjects who respond positive is

$$p_2 - p_1 = \frac{n_{12} - n_{21}}{n} \tag{3}$$

which is independent of n₁₁ and n₂₂ the number of outcomes in which the case and control subjects both respond positive or both respond negative respectively. The standard error of the difference between the two proportion is

$$Se(p_2 - p_1) = \frac{\sqrt{n_{12} + n_{21}}}{n} \tag{4}$$

which is also unaffected by n₁₁ and n₂₂. If π₁ and π₂ are respectively the proportions of control and cases in the sampled populations who respond positive then a hypothesis that may be of interest is

$$H_0 : \pi_2 - \pi_1 = 0 \tag{5}$$

versus either a one-sided or two sided alternative hypothesis. The McNemar test statistic [3, 6] for testing the null hypothesis of Eqn. 5 is

$$\chi^2 = \left(\frac{p_2 - p_1}{Se(p_2 - p_1)} \right)^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \tag{6}$$

which has a chi-square distribution with 1 degree of freedom. The null hypothesis of equal population proportions is rejected at the α level of significance in favour of the alternative hypothesis if

$$\chi^2 \geq \chi^2_{1-\alpha;1} \tag{7}$$

MODIFIED McNEMAR TEST

The McNemar test assumes that the data being analysed are frequency or count data measured on the nominal or ordinal scale of measurement already presented in summary form. However in controlled experiments like in many other studies involving data collection, the required data are not usually available in summary form, but mostly collected as raw data from the field which must be processed for subsequent summarization if necessary by the researcher. Also these data may be measurements on at least the interval scale that is they are quantitative rather than qualitative information and may furthermore be continuous data. This means that the probability of obtaining any tied observations is at least theoretically zero. However in practice ties do not occur. Just as the use of summary data leads to loss of information and hence loss of reliability in conclusions reached with the data, too many ties in the data if not adjusted for, also leads to reduction in the power of any test statistic used in data analysis.

Hence the following method may be used to modify the McNemar test statistic to allow for the possibility of ties in the data. Now let (y_{i2}, y_{i1}) be the responses or scores by the case and control subjects respectively for the ith pair of case and control subjects assumed to have been exposed to two possible experimental conditions or treatments T₂ and T₁ say respectively for i=1,2,...,n pairs of case and control subjects. We further assume that these measurements are at least on the interval scale.

$$Let u_i = \begin{cases} 1, & \text{if } y_{i2} \text{ and } y_{i1} \text{ are such scores that } y_{i2} \text{ indicates that case responds} \\ & \text{positive and } y_{i1} \text{ indicates that control responds negative} \\ & \text{to the condition of interest} \\ 0, & \text{if } y_{i2} \text{ and } y_{i1} \text{ are such scores that indicate that case and} \\ & \text{control subjects both respond positive} \\ & \text{or both respond negative.} \\ -1 & \text{if } y_{i2} \text{ and } y_{i1} \text{ are such scores that } y_{i2} \text{ indicates that case responds} \\ & \text{negative and } y_{i1} \text{ indicates that control responds positive} \end{cases} \tag{8}$$

For the ith pair of case and control subjects, for i=1,2,...,n.

Let

$$\pi^+ = P(u_i = 1) ; \pi^0 = P(u_i = 0); \text{ and } \pi^- = P(u_i = -1) \quad (9)$$

where

$$\pi^+ + \pi^0 + \pi^- = 1 \quad (10)$$

finally let

$$W = \sum_{i=1}^n u_i \quad (11)$$

Now under these specifications the expected value of u_i is

$$E(u_i) = \pi^+ - \pi^- \quad (12)$$

and the variance of u_i

$$Var(u_i) = \pi^+ + \pi^- - (\pi^+ - \pi^-)^2 \quad (13)$$

Hence from Eqn 11 and 12, we have that the expected value of W is

$$E(W) = n(\pi^+ - \pi^-) \quad (14)$$

and from Eqn 13, we have that

$$Var(W) = n(\pi^+ + \pi^- - (\pi^+ - \pi^-)^2) \quad (15)$$

Note that π^+, π^0 and π^- are respectively the probabilities that for a randomly selected pair of case and control subjects, on the average the case responds positive and the control responds negative or the case and control both respond positive or both respond negative, or the case responds negative and the control responds positive. Their sample estimates are respectively

$$\hat{\pi}^+ = \frac{f^+}{n}; \hat{\pi}^0 = \frac{f^0}{n} \text{ and } \hat{\pi}^- = \frac{f^-}{n} \quad (16)$$

where f^+, f^0 and f^- are respectively the number of 1's, 0's and -1's in the frequency distribution of these numbers in $u_i, i=1, 2, \dots, n$. In other words f^+, f^0 and f^- are respectively the number of case and control subject pairs in which the case respond positive and the control respond negative, or the case and control subjects both respond positive or both respond negative or the case responds negative and the control subject responds positive. These frequencies can therefore be expressed in terms of the case and control frequency pairs of Table 1 as respectively

$$f^+ = n_{12}; f^0 = n_{11} + n_{22} = f^{0+} + f^{0-}; f^- = n_{21} \quad (17)$$

where

$$f^{0+} = n_{11}; f^{0-} = n_{22} \quad (18)$$

are respectively the number of case and control subject pairs in which the case and control subjects both respond positive or both respond negative; with corresponding relative frequencies $\hat{\pi}^{0+}$ and $\hat{\pi}^{0-}$ respectively.

Now $\pi^+ - \pi^-$ is a measure of the differential rate of positive responses by subjects in the experimental or treatment condition T_2 namely case, and standard condition T_1 namely control and its sample estimate is

$$\pi^+ - \pi^- = \frac{W}{n} = \frac{f^+ - f^-}{n} \quad (19)$$

with estimated variance obtained from Eqn 15 as

$$Var(\hat{\pi}^+ - \hat{\pi}^-) = \frac{Var(W)}{n^2} = \frac{\hat{\pi}^+ + \hat{\pi}^- - (\hat{\pi}^+ - \hat{\pi}^-)^2}{n} \quad (20)$$

Note that the square of W namely

$$W^2 = (n(\hat{\pi}^+ - \hat{\pi}^-))^2 = (f^+ - f^-)^2 \quad (21)$$

is the numerator of the McNemar test statistic of Eqn 6. Now to develop a test statistic for the difference between positive response rates for case and control subjects we note that π^+ is the proportion of pairs of subjects out of a total of n pairs in which the subject (case) administered treatment T_2 say in a given pair responds positive and the subject (control) in the pair administered treatment T_1 say, responds negative; π^0 is the proportion of the total number of n pairs of subjects in which members of the pair both respond positive or both respond negative and π^- is the proportion of pairs out of a total of 'n' pairs in which the subject (case) administered treatment T_2 in a given pair responds negative and the subject (control) in the pair administered treatment T_1 responds positive. The case and control differential positive response rate is $\pi^+ - \pi^-$, whose sample estimate and variance are given respectively by Eqns 19 and 20. Note that in terms of the sampled proportion in Eqns 1 and 2

$$P_1 = \frac{n_{11} + n_{21}}{n} = \frac{f^{0+} + f^-}{n} = \hat{\pi}^{0+} + \hat{\pi}^- \quad (22)$$

and

$$P_2 = \frac{n_{11} + n_{12}}{n} = \frac{f^{0+} + f^+}{n} = \hat{\pi}^{0+} + \hat{\pi}^+ \quad (23)$$

where

$$\hat{\pi}^{0+} = \frac{f^{0+}}{n} \text{ and } \hat{\pi}^{0-} = \frac{f^{0-}}{n} \quad (24)$$

such that

$$\hat{\pi}^0 = \hat{\pi}^{0+} + \hat{\pi}^{0-} \quad (25)$$

which provides more detailed and additional useful information. Now testing the null hypothesis H_0 that the proportions of subjects responding positive under the two experimental or treatment conditions T_2 and T_1 differ by some value θ_0 is equivalent to testing

$$H_0 : \pi^+ - \pi^- = \theta_0 \text{ versus } H_1 : \pi^+ - \pi^- \neq \theta_0 \quad (-1 \leq \theta_0 \leq 1) \quad (26)$$

The null hypothesis of Eqn 26 may be tested using the test statistic given by

$$\chi^2 = \frac{(W - n\theta_0)^2}{n(\hat{\pi}^+ + \hat{\pi}^- - (\hat{\pi}^+ - \hat{\pi}^-)^2)} \quad (27)$$

or equivalently

$$\chi^2 = \frac{n((\hat{\pi}^+ - \hat{\pi}^-) - \theta_0)^2}{\hat{\pi}^+ + \hat{\pi}^- - (\hat{\pi}^+ - \hat{\pi}^-)^2} \quad (28)$$

which has approximately the chi-square distribution with 1 degree of freedom for sufficiently large 'n'. H_0 is rejected at the α level of significance if Eqn 7 is satisfied, otherwise H_0 is accepted.

Note that under H_0 the numerators of the test statistic of Eqns 27 and 28 are as in the McNemar test statistic independent of $n_{11} = f^{0+}$ and $n_{22} = f^{0-}$ the number of pairs in which case and control subjects in each pair both respond positive or both respond negative to the treatments of interest. The denominator of these Equations is also seen to be independent of n_{11} and n_{22} . Hence the proposed modified test statistic, like the McNemar test statistic is not affected by those pairs in which the subjects in each pair both respond positive or both respond negative to the treatments administered. However unlike the McNemar test statistic the present test statistic has nothing less by specifications been adjusted and corrected for the possible presence of ties in the data. Furthermore, the variance of the modified test statistic in Eqn.28 is smaller than the variance of the usual McNemar test statistic in Eqn 6. This is because from Eqns 20 and 4, we have that

$$\begin{aligned} \text{Var}(\hat{\pi}^+ - \hat{\pi}^-) &= \frac{\hat{\pi}^+ + \hat{\pi}^- - (\hat{\pi}^+ - \hat{\pi}^-)^2}{n} = \frac{n_{12} + n_{21}}{n^2} - \frac{(n_{12} + n_{21})^2}{n^3} \\ &\leq \frac{n_{12} + n_{21}}{n^2} = \text{Var}(P_2 - P_1), \end{aligned}$$

since

$$\frac{(\hat{\pi}^+ - \hat{\pi}^-)^2}{n} = \frac{(n_{12} + n_{21})^2}{n^3} \geq 0, \text{ for all } \pi^+ \neq \pi^-, \text{ or } n_{12} \neq n_{21}$$

Therefore, the modified test statistic is relatively more efficient and hence likely to be more powerful than the usual McNemar test statistic whenever case and control subjects have differential response rates ($\hat{\pi}^+ \neq \hat{\pi}^-$; or $P_1 \neq P_2$) to the treatments of interest. Note that $n(\hat{\pi}^+ - \hat{\pi}^-)^2$ is the reduction in the variance of the test statistic W due to the fact that this test statistic has by specifications been adjusted for the possible presence of ties in the data, that is for ties between the responses of case and control subjects. Thus in the presence of ties in the data, the McNemar test statistic is likely to have a larger estimated variance than the modified test statistic thereby yielding smaller calculated chi-square values and increasing the chances of accepting a false null hypothesis (Type II error) more frequently than the modified test statistic.

ILLUSTRATIVE EXAMPLE

A dietician is interested in comparing the effectiveness of a certain food ingredient T_2 (Treatment) with a standard food ingredient T_1 (control) in modulating body weight among subjects. She collected a random sample of 60 volunteer subjects. She matched these volunteers each with a single control of the same gender and comparable age. After measuring their initial body weights she then randomly assigned one member of each of the resulting 30 paired samples of subjects to the experimental treatment T_2 and the remaining member of the pair to the standard treatment or control T_1 . After the experimental period, she again measured the body weight of each member of the paired samples and thereafter determined the gain(+) or loss(-) in body weight of each member obtaining the following results (Table 2)

To apply the modified McNemar test, to the data of Table 2 we evaluate the values of u_i of Eqn 8 where y_{i1} and y_{i2} are respectively the weight gains by the subjects in the i th pair of case and control subjects given the new (T_2) and standard (T_1) diet preparations. For $i=1,2,\dots,30$ the results are shown in Table 2. From the values of u_i we have that $f^+ = n_{12} = 7$, $f^0 = n_{11} + n_{22} = f^{0+} + f^{0-} = 4 + 4 = 8$; $f^- = n_{21} = 15$

Hence from Eqn 16 we have that

$$\hat{\pi}^+ = \frac{7}{30} = 0.233; \hat{\pi}^0 = \frac{8}{30} = 0.267; \hat{\pi}^- = \frac{15}{30} = 0.500;$$

Note that

Table 2: Gain(+) or Loss(-) in Body weight of 30 matched Pairs of Subjects administered some diets

| | Treatment | Control | | | Treatment | Control | | | Treatment | Control | |
|------|----------------|----------------|----------------|------|----------------|----------------|----------------|------|----------------|----------------|----------------|
| Pair | T ₂ | T ₁ | u _i | Pair | T ₂ | T ₁ | u _i | Pair | T ₂ | T ₁ | u _i |
| 1 | - | - | 0 ⁻ | 11 | + | - | +1 | 21 | - | + | -1 |
| 2 | - | + | -1 | 12 | + | - | +1 | 22 | + | - | +1 |
| 3 | - | + | -1 | 13 | - | + | -1 | 23 | - | + | -1 |
| 4 | - | - | 0 ⁻ | 14 | - | - | 0 ⁻ | 24 | + | - | +1 |
| 5 | - | + | -1 | 15 | + | - | +1 | 25 | - | + | -1 |
| 6 | + | + | 0 ⁺ | 16 | + | - | +1 | 26 | - | + | -1 |
| 7 | + | + | 0 ⁺ | 17 | - | + | -1 | 27 | - | + | -1 |
| 8 | - | + | -1 | 18 | + | + | 0 ⁺ | 28 | - | + | -1 |
| 9 | - | + | -1 | 19 | + | - | +1 | 29 | - | + | -1 |
| 10 | + | + | 0 ⁺ | 20 | - | + | -1 | 30 | - | - | 0 ⁻ |

$$\hat{\pi}^0 = \frac{8}{30} = \frac{4}{30} + \frac{4}{30} = 0.133 + 0.133 = \hat{\pi}^{0+} + \hat{\pi}^{0-}$$

Also $W = f^+ - f^- = n_{12} - n_{21} = 7 - 15 = -8$.

Also from Eqn 15 we have that the estimated variance of W is

$$\begin{aligned} Var(W) &= (30)(0.233 + 0.500 - (0.233 - 0.500)^2) \\ &= (30)(0.733 - 0.071) = (30)(0.662) = 19.86 \end{aligned}$$

Hence to test the null hypothesis of Eqn 26 using the modified test statistic we have from Eqn 27 with $\theta_0 = 0$ that

$$\chi^2 = \frac{(7 - 15)^2}{19.860} = \frac{64}{19.860} = 3.223 (P - value = 0.0772)$$

which with 1 degree of freedom is not statistically significant showing that the new diet T₂ and the standard diet T₁ do not have differential effect on body weight of subjects. If we had used the usual or unmodified McNemar test statistic to analyse the data we would have from Eqn 4 that the estimated variance

of $P_2 - P_1$ is $Var(P_2 - P_1) = \frac{f^+ + f^-}{n^2} = \frac{n_{12} + n_{21}}{n^2}$
 $= \frac{7 + 15}{(30)^2} = \frac{22}{900} = 0.024$. The corresponding test statistic

for the null hypothesis of Eqn 26 with $\theta_0 = 0$ from Eqn 6

is $\chi^2 = \frac{(7 - 15)^2}{7 + 15} = \frac{64}{22} = 2.909 (P - value = 0.0910)$ which

with 1 degree of freedom is also not statistically significant. Although the proposed modified statistic and the usual McNemar test statistic have both lead to the acceptance of the null hypothesis, the relative sizes of the calculated chi-square values and the attained P-values show that the unmodified McNemar test statistic is likely to lead to an acceptance of a false null

hypothesis (Type II error) more frequently than the modified test statistic. Finally note that the estimated variance of $\hat{\pi}^+ - \hat{\pi}^-$ is $var(\hat{\pi}^+ - \hat{\pi}^-) = \frac{0.233 + 0.500 - (0.233 - 0.500)^2}{30} = \frac{0.733 - 0.071}{30} = \frac{0.662}{30} = 0.022$ which is

$$0.024 - 0.022 = 0.002 = \frac{0.071}{30} = \frac{(\hat{\pi}^+ - \hat{\pi}^-)^2}{n}$$

less as expected than the variance of $P_2 - P_1$ obtained when the unmodified McNemar test is used.

SUMMARY AND CONCLUSION

We have discussed briefly the McNemar test and proposed its modification which intrinsically and structurally adjusts the test statistic for the possible presence of tied responses between the sampled populations of case and control subjects which may be data on any scale of measurement. The proposed test statistic is shown to be more efficient and hence likely to be more powerful than the regular McNemar test statistic. The method also enables the researcher readily estimate not only the chances that among a randomly selected pair of case and control subjects, the case responds positive and the control responds negative; or the case responds negative and the control responds positive, but also even when both case and control subjects have similar responses, it enables one easily estimate the probability that both respond positive or both respond negative. The proposed method is illustrated with some data.

ACKNOWLEDGEMENT

I wish to appreciate Okeh Uchechukwu for typesetting and proofreading this work before it was ready for submission to the journal.

REFERENCE

- [1] Oyeka CA. An Introduction to Applied Statistical Methods. Nobern Avocation Publishing Company Enugu. Nigeria 2010.
- [2] Oyeka CA. An Introduction to Applied Statistical Methods (Third edition). Nobern Avocation Publishing Company Enugu. Nigeria 1996.
- [3] McNemar Q. Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika* 1947; 12: 153-57.
<http://dx.doi.org/10.1007/BF02295996>
- [4] Oyeka CA, Utazi CE, Nwosu CR, *et al.* A method of analyzing paired data intrinsically adjusted for ties. *Global J Math* 2009; 1(1): 1-6.
- [5] Freund JE. *Mathematical Statistics* (5th ed.). Prentice Hall Inc, New York 1992.
- [6] Gibbons JD. *Nonparametric Statistical Inferences*. Mc Graw Hill New York 1971.

Received on 24-08-2012

Accepted on 17-09-2012

Published on 24-09-2012

<http://dx.doi.org/10.6000/1929-6029.2012.01.01.07>