Optimal Data Analysis
Vol. 4 (April 22, 2015), 24-26

2155-0182/10/$3.00

UniODA *vs.* McNemar's Test for Correlated Proportions: Diagnosis of Disease Before *vs.* After Treatment

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

McNemar's test is used to assess the significance of the difference between two correlated proportions: for example when the two proportions are based on the same sample of subjects or on matched-pair samples. This methodology is compared with UniODA using two examples in disease diagnosis before *vs.* after treatment.

McNemar's test is applied to a 2x2 contingency table to assess the significance of the difference between two correlated proportions: for example when the two proportions are based on the same sample of subjects or on matched-pair samples.[1] The test determines whether row and column marginal frequencies are equal (referred to as "marginal homogeneity"). The null hypothesis of marginal homogeneity states that the two marginal probabilities for each outcome are identical. If 25 or more observations fall in the minor diagonal of the contingency table, the test has a chi-square distribution with one degree of freedom. If the result of the chi-square test is statistically significant, then one rejects the null hypothesis in favor of the alternative hypothesis that the marginal proportions are significantly different from each other. If either of the cell entries in the minor diagonal is "small" then an exact binomial test is used to evaluate imbalance in the minor diagonal cells.[2]

For expository purposes this report compares analysis conducted using McNemar's test versus using UniODA.[3] Data for this exposition are presented elsewhere.[4]

The first example assesses whether or not a drug has an effect upon a disease. Counts of individuals with the diagnosis of disease (either present or absent) before treatment are given in the rows, and counts of individuals with the diagnosis of disease after treatment (either present or absent) are given in the columns. This is known as a matched pairs design because the same subjects are included in before-and-after measurements.

Table 1: Data for First Example[4]

Disease Diagnosis		
	After Treatment	
Before Treatment	Present	Absent
Present	101	121
Absent	59	33

2155-0182/10/$3.00

Note that 59/92 (64.1%) of the observations without disease before treatment had the disease after treatment, while 121/222 (54.5%) of the observations with the disease before treatment were without disease after treatment. For these data the McNemar test statistic is chi-square (N=314, df=1) = 21.35, $p<0.00001$. Thus the null hypothesis is rejected, and it was concluded[4] that "...the test provides strong evidence to reject the null hypothesis of no treatment effect."

When UniODA was applied to these data[5] the model was: if disease is present before treatment then predict no disease after treatment (78.6% accurate classification, 54.5% predictive value), and if disease is absent before treatment then predict disease is present after treatment (36.9% accurate classification, 64.1% predictive value).[3] The UniODA finding was statistically significant (exact $p<0.003$), and the ESS statistic[3] of 15.4 reveals that the effect was relatively weak, yielding only 15.4% of the theoretical gain in classification accuracy that it is possible to achieve beyond chance.

The second example compares UniODA with the asymptotic McNemar test and several alternative variations.

Table 2: Data for Second Example[4]

Disease Diagnosis

	After Treatment	
Before Treatment	Present	Absent
Present	59	6
Absent	16	80

Note that 16/96 or 16.7% of the observations without disease before treatment had the disease after treatment, while 6/65 or 9.2% of the observations with the disease before treatment were without disease after treatment. Described previously[4] for these data the exact binomial test gives $p<0.053$; McNemar's test with continuity correction gives chi-square (N=161, df=1) = 3.68, $p<0.055$; asymptotic McNemar's test gives chi-square (N=161, df= 1) = 4.55, $p<0.033$; and the mid-P McNemar's test gives $p<0.035$. It was concluded[4] that "…the McNemar's test and mid-P version provide stronger evidence for a statistically significant treatment effect in this second example."

When UniODA was applied to these data the model was: if disease is present before treatment then predict disease is present after treatment (93.0% accurate classification, 83.3% predictive value), and if disease is absent before treatment then predict disease is absent after treatment (78.7% accurate classification, 90.8% predictive value). Note that a greater percentage of (and more) observations without disease acquired the disease after treatment, as compared with observations with disease who were without disease after treatment. The UniODA finding was statistically significant (exact $p<0.06 \times 10^{-20}$), and the ESS statistic[3] of 71.7 reveals that the effect was relatively strong, yielding 71.7% of the theoretical gain in classification accuracy that it is possible to achieve beyond chance.

A theoretical limitation of McNemar's methodology involves the use of inherently exploratory chi-square in testing confirmatory hypotheses. To accomplish this objective McNemar advocated the testing of directional hypotheses by dividing the chi-square Type I error rate in half.[6] In contrast, UniODA allows an exact test of both exploratory and confirmatory (weighted) hypotheses for every analysis and design.[3]

References

[1]McNemar Q (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153-157. DOI: 10.1007/BF02295996

2155-0182/10/$3.00

[2]Fagerland MW, Lydersen S, Laake P (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Medical Research Methodology, 13*: 91. DOI: 10.1186/1471-2288-13-91

[3]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*, Washington, DC, APA Books.

[4]http://en.wikipedia.org/wiki/McNemar%27s_test

[5]The UniODA syntax used to conduct this analysis is:

```
OPEN DATA;
OUTPUT EXAMPLE 1;
CATEGORICAL ON;
TABLE 2;
CLASS COL;
MCARLO ITER 25000 TARGET .001 STOP 99.999;
DATA;
101 121
59 33
END DATA;
GO;
```

[6]McNemar Q (1969). *Psychological statistics*. Hoboken, NJ, Wiley.

Author Notes

This study involved secondary data analysis of published de-identified data and was exempt from Institutional Review Board review.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646
USA