Analysis of Paired Dichotomous Data: A Gentle Introduction to the McNemar Test in SPSS

Omolola A. Adedokun
Purdue University

Wilella D. Burgess
Purdue University

Background: Although McNemar Test is the most appropriate tool for analyzing pre-post differences in dichotomous items (e.g., “yes” or “no”, “correct” or “incorrect”, etc.), many scholars have noted the inappropriate use of Pearson’s Chi-square Test by researchers, including social scientists and evaluators, for the analysis of related or dependent dichotomous variables.

Purpose: The goal of this paper is to promote the use of McNemar Test among evaluators by providing a gentle introduction to the method.
Setting: Not applicable.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Using data from 506 6th grade students’ responses to a pre-post science test; this contribution illustrates how to conduct McNemar Test in SPSS..

Findings: This contribution provides a non-technical introduction to McNemar test and illustrates its use in an applied research/evaluation context.

Keywords: *McNemar Test; paired dichotomous data; Chi-square test; equality of proportions.*

____

Program evaluation often involves the examination of pretest-posttest differences in dichotomous items (e.g., yes or no, correct or incorrect, etc.). Although the McNemar test is one of the very few statistical procedures available for pretest-posttest analysis of related dichotomous variables, it remains largely unpopular among evaluators. We suspect that this unpopularity is due to a lack of knowledge of this statistical method among evaluators—with the exception of practitioners in the subfield of health evaluation among whom the method is relatively popular (Berenson & Koppel, 2005).

The unpopularity of the McNemar test has been noted by other scholars. For example, Levin and Serlin (2000) as well as Berenson and Koppel (2005) noted that hypotheses testing in most statistics courses for social scientists—including evaluators—often center on mainstream tests of mean differences (e.g., dependent and independent *t*-tests, ANOVA, etc.) and independent sample proportions (e.g., chi-square analysis) with little or no attention paid to the McNemar test of correlated proportions. In the same vein, extant and recent studies (e.g., Hoffman, 1976; Berenson & Koppel, 2005) have commented on the improper use of Pearson's chi-square test, instead of the McNemar test, for the analysis of correlated or dependent dichotomous variables. Indeed, the inappropriate use of Chi-square analysis for dependent dichotomous variables may lead evaluators to make misleading conclusions and recommendations to program stakeholders.

The purpose of this contribution is to promote the use of the McNemar test for the analysis of correlated proportions among evaluators. We do not intend to provide technical details of the McNemar test. Rather, our goal is to provide a gentle introduction to its application in a relatively non-technical manner. Using data from a science test, we illustrate the application of the McNemar test, and how to interpret and present the findings. We chose SPSS because most social scientists, including evaluators, are familiar with its user friendly interface (Heck, Thomas & Tabata, 2010). We begin with a general description of the McNemar procedure followed by an illustration of how the analysis is conducted in SPSS.

Description of the McNemar Test

The McNemar test (1947) is best described as a 2×2 cross classification of paired (or matched) responses to a dichotomous item. In simple terms, the McNemar test can be viewed as a type of chi-square test that uses dependent (i.e., correlated or paired) data rather than independent (unrelated) samples. The McNemar test is a non-parametric statistical test; i.e., it is distribution free and can be used with data sets and samples that are not normally distributed (Ciechalski, et al., 2002).

For illustration, we use data from a pretest-posttest science test that was administered to 506 6th grade students to evaluate the impact of a curricular unit/ module on their understanding of Mixtures and Solutions. Students' responses to each of 15 questions on the test were scored as correct (1) or incorrect (0). Table 1 describes response patterns to one of the questions in a typical 2×2 format.

Table 1
Example of 2×2 Classification Table for McNemar Analysis

		Posttest		
		Correct (1)	Incorrect (0)	
Pretest	Correct (1)	a = 55	b = 35	$n_1 = 90$
	Incorrect (0)	c = 199	d = 217	
		$n_2 = 254$		$n = 506$

Where:

- a = number of students who gave correct responses in the pretests and posttests
- b = number of students who gave correct response in the pretest but an incorrect response in the posttest
- c = number of students who gave incorrect responses in the pretest but correct in the posttest
- d = number students who gave incorrect responses in both the pretest and posttest
- n = total number of matched pairs (i.e., a + b + c + d)
- n_1 = total number of students who provided correct responses in the pretest (i.e., a + b)
- n_2 = total number of students who provided correct responses in the posttest (i.e., a + c)
- p_1 = proportion of correct responses in the pretest, i.e., $n1/n$ or $(a + b)/n$
- p_2 = proportion of correct responses in the posttest, i.e., $n2/n$ or $(a + c)/n$

Hypothesis Testing

Suppose we wish to examine pretest-posttest changes in the proportion of students that reported correct responses before and after the intervention, then, we will test the null hypothesis that $p_1 = p_2$. In a sense, hypothesis testing for the McNemar Test uses data from the two discordant cells b & c (see Table 1) where change has occurred to test the equivalence of the two proportions (i.e., marginal homogeneity). The uncorrected test statistic[1] for the McNemar procedure is a chi-square test (with 1 degree of freedom) denoted as $(b - c)^2/(b + c)$ and the corrected test statistic is $(|b - c|-1)^2/(b + c)$ (Interested readers should see Berenson & Koppel, 2005; McNemar 1947; Lehr, 2010; Feuer & Kessler, 1989; & Marascuilo et al., 1979 for detailed explanations of the derivation of the test statistic).

Conducting McNemar Test in SPSS

1. Inspect data set to ascertain that responses are appropriately coded as 1 or 0. For example, run frequency counts of pretest and posttest responses to the item(s) (Analyze→Descriptive Statistics→ Frequencies and then select variables of interest and click "OK").
2. Conduct the analysis. There are two ways to run the McNemar analysis using the user friendly drop down menu in SPSS: (1) the binomial approach (described in Figure 1) that is better suited to small samples and/or when b + c is less than 10, and (2) the original McNemar approach (described in Figure 2) that is appropriate for large samples. While the SPSS output for the binomial approach includes an exact binomial p-value and does not provide a test statistic value, the output for the original method includes the value of the

[1] A limitation of the McNemar test is that it was designed for use with large samples. For small sample sizes, a correction formula like the Yates correction formula should be used instead of the McNemar test (Ciechalski, Pinkney and Weaver, 2002). Also, the asymptotic 2×2 McNemar test assumes that the number of discordant pairs (i.e., b+c) is equal to or larger than 10. Hence, use of an exact binomial test is recommended if discordant pairs are less than 10 (Rufibach, 2011).

test statistic and incorporates a correction for continuity in the analysis.

3. Interpret results. The results/outputs for both methods include the *p*-value for the test (in this case, $p < 0.05$ indicating statistically significant differences between pretest and posttest responses). Thus, we can reject the null hypothesis that $p_1 = p_2$.

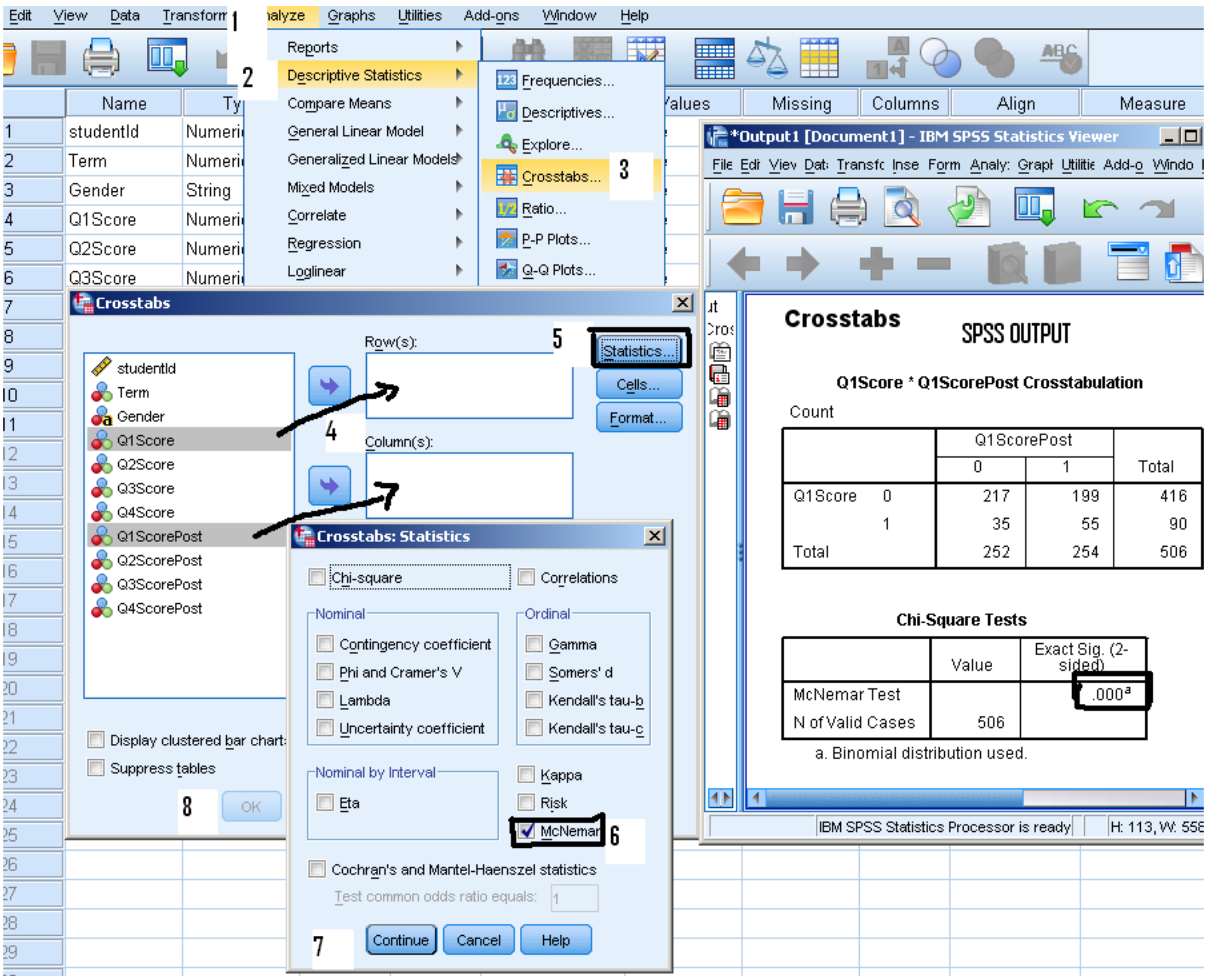

Figure 1. Conducting McNemar Test in SPSS (Binomial Method)

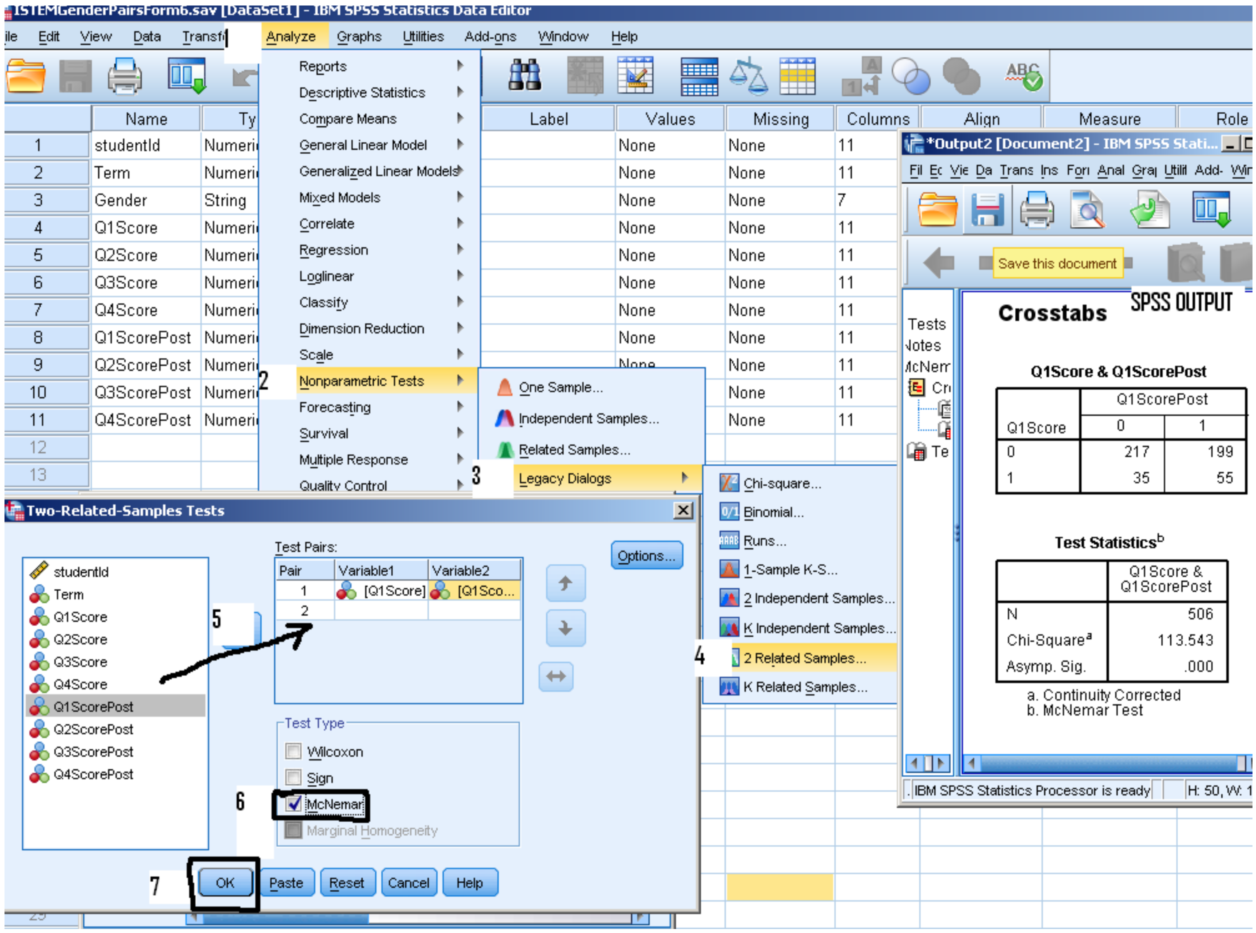

Figure 2. Conducting McNemar Test in SPSS (Traditional Method with Yates Continuity Correction)

Two Sample McNemar Test

McNemar analysis has been extended to multiple group designs (Feuer & Kessler, 1989; Marascuillo & Serlin, 1979) to compare changes in pretest and posttest responses between experimental, gender, racial or age groups, when the researcher wishes to test the hypothesis that the marginal change is the same across groups. For example, in the case of the science test illustrated above, we may wish to examine whether the proportion of pretest-posttest change is the same for boys and girls.

The first step in testing the two-sample hypothesis is to conduct separate McNemar analyses for each group (Marascuilo & Serlin, 1979) to identify the discordant or “change” cells as shown in Table 3. To perform the separate McNemar analysis in SPSS, split the data by group (e.g., “boys” and “girls” data sets) and repeat the steps described above to generate Table 2.

Table 2
Two Sample McNemar Test

		Posttest			
		Boys (284)		Girls (222)	
		Correct (1)	Incorrect (0)	Correct (1)	Incorrect (0)
Pretest	Correct (1)	29	20	26	15
	Incorrect (0)	106	129	93	88

The values of the discordant pairs for each sample are then compiled into a separate 2×2 table (see Table 3) and a chi-square test of independence or *z*-test of the equality of two proportions is performed to examine if the probability of change (or marginal probabilities) is the same across both groups (Marascuilo & Serlin, 1979; Levin & Serlin, 2000; Howell, 2008). Fortunately, online calculators (e.g., http://www.graphpad.com/quickcalcs/contingency1.cfm) are available for such analysis. For the example considered here, the chi-square analysis revealed no statistically significant gender difference in students' pretest-posttest responses to the question.

Table 3
2×2 Classification Table for Two Sample McNemar Test

Discordant Pairs	Groups	
	Boys	Girls
Correct Pretest & Incorrect Posttest	20	15
Incorrect Pretest & Correct Posttest	106	93

Conclusion

This contribution provides a non-technical introduction to the McNemar test with the overarching goal of promoting its use and application among evaluators. The article illustrates the use of the McNemar test in an applied research context and seeks to discourage the use of chi-square analysis for examining paired or dependent variables—for which McNemar is the most appropriate. As illustrated above, the method is simple, quick and easy to perform, has high practical power and "it enables an appropriate confirmatory data analysis for situations dealing with paired dichotomous responses to surveys or experiments" and, provides a useful addition to evaluators' analyses (Berenson & Koppel, 2005, p. 134). In addition to its use for testing significance of changes in related proportions, McNemar analysis also has potential as an exploratory tool in dichotomous item analysis. Indeed, the method provides a useful addition to evaluators' analysis toolkit.

This article is by no means an exhaustive introduction to McNemar analysis. There are other advanced topics related to the method that may be of interest to evaluators. For example, although the McNemar test works best when data is complete, (i.e., with no missing data), systems or methods (including maximum likelihood strategies) for handling the analysis with missing data have also been developed (Marascuillo, Omelich, & Gokhale, 1988). Hopefully, this non-technical illustration will serve as a guide or resource for evaluators who are often more interested in the appropriate application of statistical methods and not necessarily their mathematical derivations and technical details. We also hope that this illustration will spur more evaluators to gain deeper knowledge of the method and its extensions.

Acknowledgements

Data for this study come from the Evaluation of the Indiana Science Initiative Project implemented by the Indiana Science Technology Engineering and Mathematics (I-STEM) Network and funded by the Lilly Foundation.

The authors wish to thank Ann Bessenbacher for her role in data preparation.

References

Berenson, M. L., & Koppel, N. B. (2005). Why McNemar's procedure needs to be included in the Business Statistics Curriculum. *Decision Sciences Journal of Innovative Education*, *3*(1), 125-136.

Ciechalski, J. C., Pinkney, J. W., & Weaver, F. S. (2002). *A method for assessing change in attitude: The McNemar Test*. Poster presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge, NY.

Hoffman, J. I. E. (1976). The Incorrect use of Chi-square analysis for paired data. *Clinical Experiment Immunology*, *24*, 227-229.

Lehr, R. G. (2010). McNemar Test. In S. Chow (Ed.), *Encyclopedia of biopharmaceutical statistics* (pp. 740-744). Informa Healthcare.

Levin, J. R., & Serlin, R. C. (2000). Changing students' perspectives of McNemar's test of change. *Journal of Statistics Education*, *8*(2). Retrieved March 15, 2011 from www.amstat.org/publications/jse/secure/v8n2/levin.cfm

Lachenbruch, P. A., (2005) McNemar test. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics* (pp. 3062-3063). New York, NY: John Wiley and Sons.

Marascuilo, L. A., Omelich, C. L., & Gokhale, D. V. (1988). Planned and post hoc methods for multiple-sample McNemar (1947) tests with missing data. *Psychological Bulletin*, *103*, 238-245.

Marascuilo, L. A. & Serlin, R. C. (1979). Tests and contrasts for comparing change parameters for a multiple sample McNemar data model. *British Journal of Mathematical and Statistical Psychology*, *32*, 105-112.

Howell, D. C. (2008). *Testing change over two measurements in two independent groups*. Retrieved June 21, 2011 from http://www.uvm.edu/~dhowell/methods7/Supplements/Testing%20Dependent%20Proportions.pdf