Chi-square for within-subjects: McNemar's test

For a binary dependent variable, there is a form of the chi-square test for within-subjects designs called McNemar's chi-square. As with the paired t-test (or, later we'll discuss the within-subjects ANOVA), the McNemar test is used whenever the same individuals are measured (or surveyed) twice, matched on some variable (e.g., yoked by age), participants are paired in some way (e.g., twins or married couples), or responses on two measures are used (e.g., favorability toward gun control compared to favorability toward abolishing the second amendment).

Below is a 2 × 2 table of clinical depression classification of respondents to a national survey of older adults at two different time points (six months apart).[1]

		Time 2		
		Not Depressed	Depressed	
Time 1	Not Depressed	146	155	301
	Depressed	47	303	350
		193	458	651

It is important to note that this table differs from the non-matched 2 × 2 table from the poll data that we examined previously. That table used party identification by candidate choice for the purpose of asking whether independent voter identification meant that the voter was more (or less) likely to support a particular candidate—a comparison between two groups on a binary dependent variable. The above table involves two measurements of the same variable repeated at two different time points to ask whether there is an increase or decrease in the binary response over time.

To compute McNemar's, the classic formula uses the "discordant cells" in the design (i.e., not depressed then depressed and depressed then not depressed). This comparison is algebraically equivalent to comparing the marginal percentages (% depressed at Time 1 vs. % depressed at Time 2; see Agresti, 2013, p. 414). The marginal percentages make much more sense as a way of framing the research question, however.

$$McNemar's \; \chi^2 = \frac{(n_{21} - n_{12})^2}{n_{21} + n_{12}} = \frac{(c - b)^2}{c + b}$$

The c, b, and d come from a commonly used convenient way of labeling the cells, as in the table shown below.

		Time 2	
		Not	Depressed
Time 1	Not	a	b
	Depressed	c	d

Let's consider whether there was a significant change in the proportion of respondents who were depressed, 350/651 = .538 at Time 1 and 458/651 = .704 at Time 2. If there is a change overall, either decrease or increase, from one time point to another, relative to what is expected by chance, the result will be significant. Using the discordant counts from the table above,

---

[1] The data for the examples below come from the Late Life Study of Social Exchanges (LLSSE; Sorkin & Rook, 2004). The depression diagnosis is determined by the recommended cutoff scores for the brief 9-item version (Santor & Coyne, 1997) of the Center for Epidemiologic Studies-Depression scale (Radloff, 1977).

$$McNemar's\ \chi^2 = \frac{\left(n_{21} - n_{12}\right)^2}{n_{21} + n_{12}}$$

$$= \frac{\left(47 - 155\right)^2}{47 + 155}$$

$$= \frac{\left(108\right)^2}{202}$$

$$= 57.742$$

df in this test is 1, the critical value is 3.84 (from the standard chi-square table), and because the calculated value of 57.74 exceeds this value, there is a significant difference between Time 1 and Time 2 responses.

Extensions
The McNemar test was generalized by Stuart (1955) for square tables larger than the 2 × 2 case (e.g., yes, no, undecided), a test usually referred to as the Stuart-Maxwell statistic for testing marginal homogeneity. For repeated variables on 3 or more occasions, the Cochran's Q test (Cochran, 1950). I will not detail these here, but see Agresti (2013) for more detail on analyzing larger designs for matched pairs.

Software Examples
There are several ways to call a McNemar's test in SPSS and I'll illustrate just one.

SPSS Syntax
It is first good to get descriptives to see the frequencies and proportions. I request both row and column proportions, but we are really only concerned with the proportions in the margins.[2]

```
crosstabs /tables w1dep by w2dep
  /cells=count row column.
```

w1dep Time 1 depression * w2dep Time 2 depression Crosstabulation

			.00 not depressed	1.00 depressed	Total
			w2dep Time 2 depression		
w1dep Time 1 depression	.00 not depressed	Count	146	155	301
		% within w1dep Time 1 depression	48.5%	51.5%	100.0%
		% within w2dep Time 2 depression	75.6%	33.8%	46.2%
	1.00 depressed	Count	47	303	350
		% within w1dep Time 1 depression	13.4%	86.6%	100.0%
		% within w2dep Time 2 depression	24.4%	66.2%	53.8%
Total		Count	193	458	651
		% within w1dep Time 1 depression	29.6%	70.4%	100.0%
		% within w2dep Time 2 depression	100.0%	100.0%	100.0%

From the crosstabs, be sure to look at and report the marginal total percentages rather than the percentages within particular cells (i.e., not the conditional probabilities called "row" or "column percentage" in SPSS that we used for the more common contingency chi-square test).

---

[2] You can also request the McNemar test as a subcommand of CROSSTABS with /STATISTICS=MCNEMAR. The McNemar chi-square value is not printed, however, just the p-value. This approach because this command does not use a continuity correction. Two other methods are available in SPSS for the McNemar test but both use continuity correction.
`npar tests mcnemar=w1dep w2dep.`
`nptests /related test (w1dep w2dep) mcnemar.`
In this case and when examples have a substantial sample size, there will be no difference in conclusions.

To get the McNemar's significance test, use this PROPORTIONS command method which is a z test equivalent to mcnemar's test without continuity correction (square the z to obtain the chi-square values and use the "Two-Sided p" for significance). It provides the essential marginal frequencies and proportions (and you can ignore the "Asymptotic Standard Error"). Another table with confidence limits for the marginal proportions is also printed, but I have omitted it here.

```
proportions
 /paired samples w1dep w2dep
 testtypes=mcnemar.
```

Paired-Samples Proportions Statistics

		Successes	Trials	Proportion	Asymptotic Standard Error
Pair 1	w1dep Time 1 depression = 1.00 depressed	350	651	.538	.027
	w2dep Time 2 depression = 1.00 depressed	458	651	.704	.021

Paired-Samples Proportions Tests

	Test Type	Difference in Proportions	Asymptotic Standard Error	Z	Significance One-Sided p	Significance Two-Sided p
Pair 1: w1dep Time 1 depression - w2dep Time 2 depression	McNemar	-.166	.021	-7.599	<.001	<.001

If you want to report the McNemar's as a chi-square (which is most common), then computation is simple: $z^2 = \chi^2 = (-7.599)^2 = 57.74$.

R
I use the `mcnemar.test` base R function and turn off the continuity correction, with `correct = FALSE` (which is the same as the hand computation formula I used, and what I recommend). First, to obtain counts and proportions, you can use the lessR `Barchart` function (*ignore any statistical tests*)

```
> #get the cell and marginal frequencies (ignore significance tests)
> BarChart (w1dep, by=w2dep, do_plot=FALSE)

Joint and Marginal Frequencies
------------------------------
       w1dep
w2dep     0    1 Sum
  0     146   47 193
  1     155  303 458
  Sum   301  350 651


> #get marginal frequencies and proportions (ignore significance test)
> BarChart (w1dep, stat_x = "proportion", do_plot=FALSE)
--- w1dep ---
                 0      1     Total
Frequencies:    301    350     651
Proportions:  0.462  0.538    1.000


#get marginal frequencies and proportions (ignore significance test)
> BarChart (w2dep, stat_x = "proportion", do_plot=FALSE)
--- w2dep ---
                 0      1     Total
Frequencies:    193    458     651
Proportions:  0.296  0.704    1.000
```

```
> #make array of frequencies for the mcnemar.test function
> counts <-array(
>    c(146,155,47,303),
>    dim=c(2, 2),
>    dimnames=list(w1dep=c("not", "depressed"),
                   w2dep =c("not", "depressed"))
>  )
> counts   #checks that the matrix is right

> mcnemar.test(counts, y=NULL, correct = FALSE)
```

        McNemar's Chi-squared test
data: counts
McNemar's chi-squared = 57.743, df = 1, p-value = 0.00000000000002988

Example write-up. A McNemar's test was used to investigate whether the percentage of participants who were clinically depressed changed over the six-month interval. At baseline, 54% of participants met the clinical cutoff for depression, whereas, at follow-up, 70% of participants met the clinical cutoff for depression. The difference was significant, McNemar's $\chi^2(1)$ = 57.74, p < .001, and was a moderate-sized effect, w = .30.[3]

References

Agresti, A. (2013). *Categorical Data Analysis, Third Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey.
Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*, 256–266
Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, *42*, 412-416

---

[3] Though I have not seen authors report Cohen's w with McNemar's test, I see no reason we cannot compute it in this case,

$$w = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{57.74}{651}} = .30$$