

9.6: McNemar's test

Introduction

There are a number of scenarios in which subjects are **paired** or matched as part of the experimental design in order to control for confounding variables — a **matched pair case-control**. Subjects may be matched by age, or other criteria, or the observations are repeat measures of the same subjects (e.g., left hand vs. right hand). One member of each pair is then randomly assigned to a treatment, the remaining pair member then assigned to the other treatment group. This scenario should remind you of our standard contingency table problem, but instead of a random collection of subjects assigned to treatments, the data are paired nominal. Thus, paired means that experimental (sampling) units are not independent, which if ignored violates an assumption required to employ the χ^2 test. We use **McNemar's test** instead.

The possible results of such a design include just two outcomes: the pairs have the same outcome (agree, **concordant**) or the pairs have different outcomes (disagree, **discordant**).

McNemar's solution was to consider only the discordant pairs. Consider two kinds of tests or assays for a condition, and the doctor receives the results of both tests.

Table 9.6.1. Format of data where McNemar's test can be applied.

		Test 2		
		Positive	Negative	Row total
Test 1	Positive	a	b	$a + b$
	Negative	c	d	$c + d$
Column total		$a + c$	$b + d$	n

Null hypothesis is that marginal proportions are equal:

$$H_O = p_b = p_c$$

$$H_A = p_b \neq p_c$$

Then McNemar's test is given by

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

and the test has one degree of freedom.

If one of the cells is low, then a continuity correction would be applied (Edwards 1948, cited in Fagerland et al 2013). With this correction the equation becomes

$$\chi_c^2 = \frac{(|b - c| - 1)^2}{b + c}$$

If either b or c is small, then the McNemar's test statistic does not approximate a χ^2 distribution very well, so there is a binomial version that you would use (Cochran's Q test) in cases where there are three or more matched sets and is common in meta-analysis (Kulinskaya and Dollinger 2015).

R code

Example data: Approval ratings for President Trump at two important markers during the Covid-19 pandemic: in April 2020, deaths passed 10,000 persons in the U.S.; in October 2020, it was reported that President Trump tested positive for SAR-COV2 and was admitted to Walter Reed National Military Medical Center (admitted 3 Oct., released 5 Oct.). Surveys were conducted by [YouGov](#) (April, sponsored by The Economist; October, sponsored by Yahoo News; data extracted from [How Americans View Biden's Response To The Coronavirus Crisis](#))

Table 9.6.2. U.S. approval ratings for President Trump in 2020.

	Approve	Disapprove

	Approve	Disapprove
April survey	720	705
October survey	645	812

Enter the data as a matrix (note this would be a general approach for the contingency table problems, too, instead of entering via Rcmdr menu). The discordant pairs are $b = 645$ and $c = 705$.

```
covid19 <- matrix(c(720, 645, 705, 812), nrow = 2, dimnames = list("April survey" = c(
  covid19
                                October survey
April survey      Approve  Disapprove
      Approve      720      705
      Disapprove    645      812
```

Uncorrected:

```
mcnemar.test(covid19, correct=FALSE)

McNemar's Chi-squared test

data: covid19
McNemar's chi-squared = 2.6667, df = 1, p-value = 0.1025
```

Correction applied:

```
mcnemar.test(covid19, correct=TRUE)

McNemar's Chi-squared test with continuity correction

data: covid19
McNemar's chi-squared = 2.5785, df = 1, p-value = 0.1083
```

Conclusions?

No change in approval ratings. The correction for small sample size had little effect on p-value, unsurprisingly, given that the surveys included 1500 (April) and 1504 (October) persons.

Unconditional paired tests

McNemar's solution considers only the discordant pairs; it's a conditional test. The downside of these tests is that the concordant pairs are not considered. Thus, by in effect tossing out some portion of the experimental results, it shouldn't surprise you that the statistical power of the test is reduced (see [Chapter 11](#)). Thus, McNemar's test may no longer be the best choice. Alternative unconditional tests have been proposed, and the **mid-P** alternative shows promise (Routledge 1994; Fagerland et al 2013). The mid-P value is calculated as the standard p-value for a test statistic minus one half the difference between the standard p-value and the next lowest possible p-value. McNemar's mid-p test is available in package `contingencytables`. Try with the example data set in Fagerland et al 2013 (Table 1).

```
#create a 2x2 matrix
bentur <- rbind(c(1, 1), c(7, 12))
```

First run McNemar's test without correction for small sample size.

```
mcnemar.test(bentur, correct=FALSE)
```

R output follows:

```
McNemar's Chi-squared test

data: bentur
McNemar's chi-squared = 4.5, df = 1, p-value = 0.03389
```

Next, run McNemar's test with correction for small sample size.

```
mcnemar.test(bentur, correct=TRUE)
```

R output follows:

```
McNemar's Chi-squared test with continuity correction

data: bentur
McNemar's chi-squared = 3.125, df = 1, p-value = 0.0771
```

Last, run mid-p version of McNemar's test.

```
McNemar_midP_test_paired_2x2(bentur)
```

R output

```
[1] The McNemar mid-P test: P = 0.039063
```

See also `mcnemarExactDP` function in `exact2x2` package. Without explanation, here's the R code and results.

```
mcnemarExactDP(n = sum(bentur), m= bentur[1,2] + bentur[2,1], x = bentur[1,2])

      Exact McNemar Test (with central confidence intervals)

data: n=sum(bentur) m=bentur[1, 2] + bentur[2, 1] x=bentur[1, 2]
n = 21, m = 8, x = 1, p-value = 0.07031
alternative hypothesis: true difference in proportions is not equal to 0
95 percent confidence interval:
 -0.54549962 0.02044939
sample estimates:
      x/n      (m-x)/n  difference 
0.04761905 0.33333333 -0.28571429
```

Alternatively, use wrapper function `mnemar.exact()`.

```
mcnemar.exact(bentur)
```

R output:

```
Exact McNemar test (with central confidence intervals)

data: bentur
b = 1, c = 7, p-value = 0.07031
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.003169739 1.111975554
sample estimates:
odds ratio
 0.1428571
```

Note the alternative hypothesis: p-value is two-tailed.

Questions

1. Apply McNemar's test and mid-P exact test to [CDC example](#)

		Controls	
		Exposed	Not exposed
Cases	Exposed	58	89
	Not exposed	32	95

This page titled [9.6: McNemar's test](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.