

12.8: The McNemar Test

Suppose you've been hired to work for the *Australian Generic Political Party* (AGPP), and part of your job is to find out how effective the AGPP political advertisements are. So, what you do, is you put together a sample of $N=100$ people, and ask them to watch the AGPP ads. Before they see anything, you ask them if they intend to vote for the AGPP; and then after showing the ads, you ask them again, to see if anyone has changed their minds. Obviously, if you're any good at your job, you'd also do a whole lot of other things too, but let's consider just this one simple experiment. One way to describe your data is via the following contingency table:

	Before	After	Total
Yes	30	10	40
No	70	90	160
Total	100	100	200

At first pass, you might think that this situation lends itself to the Pearson χ^2 test of independence (as per Section 12.2). However, a little bit of thought reveals that we've got a problem: we have 100 participants, but 200 observations. This is because each person has provided us with an answer in *both* the before column and the after column. What this means is that the 200 observations aren't independent of each other: if voter A says "yes" the first time and voter B says "no", then you'd expect that voter A is more likely to say "yes" the second time than voter B! The consequence of this is that the usual χ^2 test won't give trustworthy answers due to the violation of the independence assumption. Now, if this were a really uncommon situation, I wouldn't be bothering to waste your time talking about it. But it's not uncommon at all: this is a *standard* repeated measures design, and none of the tests we've considered so far can handle it. Eek.

The solution to the problem was published by McNemar (1947). The trick is to start by tabulating your data in a slightly different way:

	Before: Yes	Before: No	Total
After: Yes	5	5	10
After: No	25	65	90
Total	30	70	100

This is exactly the same data, but it's been rewritten so that each of our 100 participants appears in only one cell. Because we've written our data this way, the independence assumption is now satisfied, and this is a contingency table that we *can* use to construct an X^2 goodness of fit statistic. However, as we'll see, we need to do it in a slightly nonstandard way. To see what's going on, it helps to label the entries in our table a little differently:

	Before: Yes	Before: No	Total
After: Yes	a	b	a+b
After: No	c	d	c+d
Total	a+c	b+d	n

Next, let's think about what our null hypothesis is: it's that the "before" test and the "after" test have the same proportion of people saying "Yes, I will vote for AGPP". Because of the way that we have rewritten the data, it means that we're now testing the hypothesis that the *row totals* and *column totals* come from the same distribution. Thus, the null hypothesis in McNemar's test is that we have "marginal homogeneity". That is, the row totals and column totals have the same distribution: $P_a + P_b = P_a + P_c$, and similarly that $P_c + P_d = P_b + P_d$. Notice that this means that the null hypothesis actually simplifies to $P_b = P_c$. In other words, as far as the McNemar test is concerned, it's only the off-diagonal entries in this table (i.e., b and c) that matter! After noticing this, the **McNemar test of marginal homogeneity** is no different to a usual χ^2 test. After applying the Yates correction, our test statistic becomes:

$$X^2 = \frac{(|b - c| - 0.5)^2}{b + c}$$

or, to revert to the notation that we used earlier in this chapter:

$$X^2 = \frac{(|O_{12} - O_{21}| - 0.5)^2}{O_{12} + O_{21}}$$

and this statistic has an (approximately) χ^2 distribution with $df=1$. However, remember that – just like the other χ^2 tests – it's only an approximation, so you need to have reasonably large expected cell counts for it to work.

12.8.1 Doing the McNemar test in R

Now that you know what the McNemar test is all about, let's actually run one. The `agpp.Rdata` file contains the raw data that I discussed previously, so let's have a look at it:

```
load("./rbook-master/data/agpp.Rdata")
str(agpp)
```

```
## 'data.frame':    100 obs. of  3 variables:
## $ id           : Factor w/ 100 levels "subj.1","subj.10",...: 1 13 24 35 46 57 68 79 90 100
## $ response_before: Factor w/ 2 levels "no","yes": 1 2 2 2 1 1 1 1 1 1 ...
## $ response_after : Factor w/ 2 levels "no","yes": 2 1 1 1 1 1 1 2 1 1 ...
```

The `agpp` data frame contains three variables, an `id` variable that labels each participant in the data set (we'll see why that's useful in a moment), a `response_before` variable that records the person's answer when they were asked the question the first time, and a `response_after` variable that shows the answer that they gave when asked the same question a second time. As usual, here's the first 6 entries:

```
head(agpp)
```

```
##      id response_before response_after
## 1 subj.1             no             yes
## 2 subj.2             yes             no
## 3 subj.3             yes             no
## 4 subj.4             yes             no
## 5 subj.5             no             no
## 6 subj.6             no             no
```

and here's a summary:

```
summary(agpp)
```

```
##      id      response_before response_after
## subj.1 : 1      no :70             no :90
## subj.10 : 1     yes:30             yes:10
## subj.100: 1
## subj.11 : 1
## subj.12 : 1
## subj.13 : 1
## (Other) :94
```

Notice that each participant appears only once in this data frame. When we tabulate this data frame using `xtabs()`, we get the appropriate table:

```
right.table <- xtabs( ~ response_before + response_after, data = agpp)
print( right.table )
```

```
##               response_after
## response_before no yes
##               no  65  5
##               yes 25  5
```

and from there, we can run the McNemar test by using the `mcnemar.test()` function:

```
mcnemar.test( right.table )
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data:  right.table
## McNemar's chi-squared = 12.033, df = 1, p-value = 0.0005226
```

And we're done. We've just run a McNemar's test to determine if people were just as likely to vote AGPP after the ads as they were before hand. The test was significant ($\chi^2(1)=12.04, p<.001$), suggesting that they were not. And in fact, it looks like the ads had a negative effect: people were less likely to vote AGPP after seeing the ads. Which makes a lot of sense when you consider the quality of a typical political advertisement.

This page titled [12.8: The McNemar Test](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.