

# Categorical Data: Tests of Association

BIOS 6611

CU Anschutz

Week 5

- 1 Chi-Squared Test
- 2 Tests for Small Sample Sizes
- 3 Test for Paired Samples

# Chi-Squared Test

# Tests of Association

Each of the effect measures (RD, RR, OR) will have a respective associated confidence interval (and methods for obtaining it).

It should be noted that the CI for an effect measure will sometimes be at odds with the p-value for the test statistic, particularly when the results are marginally significant. This is a result of a mismatch between the ways in which the CI and the test statistic are formulated.

Regardless of the study design we can test the *hypothesis of no association between exposure and disease* using one (and the same) test statistic.

# The 2x2 Table for Testing Association

The test statistic used to test association is based on the assumption of independence between the row and column variables for *exposure* and *disease*.

Our  $H_0$  is that there is no association between exposure ( $E$ ) and disease ( $D$ ). This can be directly related to our definition for two independent events in terms of probability:  $P(E \cap D) = P(E) \times P(D)$ .

Based on our study, we can set up a 2x2 table to summarize the disease outcome based on a given exposure:

<i>Exposure</i>	<i>Disease</i>		
	Yes	No	
Yes	$O_{11}$	$O_{12}$	$n_1$
No	$O_{21}$	$O_{22}$	$n_2$
	$m_1$	$m_2$	$N$

# The Expected 2x2 Table

The expected number of events ( $E$  and  $D$ ) can be calculated as  $N \times P(E_i) \times P(D_j)$  for  $i = 1, 2; j = 1, 2$ . In our context the expected number of events in each 2x2 table cell can be calculated as

$$E_{ij} = N \times \frac{n_i}{N} \times \frac{m_j}{N} = \frac{n_i \times m_j}{N},$$

which results in the 2x2 table for what we would *expect* to observe:

<i>Exposure</i>	<i>Disease</i>		
	Yes	No	
Yes	$E_{11}$	$E_{12}$	$n_1$
No	$E_{21}$	$E_{22}$	$n_2$
	$m_1$	$m_2$	$N$

# The $\chi^2$ Test Statistic

To summarize how much the entries in the observed table deviate from the assumption of independence, we combine the information in the two tables in this way (from Karl Pearson):

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_1^2,$$

where the  $\chi_1^2$  is the square of a standard normal distribution, i.e.  $\chi_1^2 = Z^2$ .

Because we are applying a continuous distribution to discrete data, we often use the Yates-corrected version that corrects for continuity:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}} \sim \chi_1^2$$

# The $\chi^2$ Assumption and Generalizations

One assumption we must make for the chi-squared test to be valid is that *all of the expected values must be greater than or equal to 5*, otherwise we must consider other tests.

While we considered a  $2 \times 2$  table, the  $\chi^2$  test of independence can be generalized to an  $r \times c$  table, where the subscripts on  $O$  and  $E$  range from 1 to  $r$  and 1 to  $c$ , respectively.

Next we'll look at how we can use R to obtain the effect measures (RD, RR, OR) and evaluate the test of independence.



# Our Lung Cancer and Drinking Example

Recall our example examining the effect of heavy drinking (exposure) and lung cancer (disease):

<i>Drinking Status</i>	<i>Lung Cancer</i>		
	Yes	No	
Heavy	33	1667	1700
Not Heavy	27	2273	2300
	60	3940	4000

Let's create this in R and then leverage some functions for our calculations:

```
lc <- as.table(matrix(c(33, 1667, 27, 2273), ncol=2, byrow=T))
dimnames(lc) <- list(drinking.status=c("heavy", "non"),
                    lung.cancer=c("yes", "no"))
```

# Measures of Effect with epi.2by2

```
library(epiR)
epi.2by2(1c)
```

```
##           Outcome +      Outcome -      Total      Inc risk *      Odds
## Exposed +           33          1667          1700           1.94      0.0198
## Exposed -           27          2273          2300           1.17      0.0119
## Total                60          3940          4000           1.50      0.0152
##
## Point estimates and 95% CIs:
## -----
## Inc risk ratio                1.65 (1.00, 2.74)
## Odds ratio                    1.67 (1.00, 2.78)
## Attrib risk *                 0.77 (-0.02, 1.56)
## Attrib risk in population *   0.33 (-0.25, 0.91)
## Attrib fraction in exposed (%) 39.53 (-0.18, 63.49)
## Attrib fraction in population (%) 21.74 (-3.31, 40.72)
## -----
## Test that OR = 1: chi2(1) = 3.895 Pr>chi2 = 0.05
## Wald confidence limits
## CI: confidence interval
## * Outcomes per 100 population units
```

# $\chi^2$ Test of Independence

```
chisq.test(lc, correct=F)
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  lc  
## X-squared = 3.8947, df = 1, p-value = 0.04844
```

```
chisq.test(lc, correct=T)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  lc  
## X-squared = 3.3927, df = 1, p-value = 0.06548
```

**Overall Conclusion:**

# Tests for Small Sample Sizes

# Categorical Data with Small Expected Counts

If the assumption of each expected cell ( $E_{ij}$ ) having at least 5 is violated, the normal asymptotics (i.e.,  $\chi^2_1 = Z^2$ ) may not be close to the underlying true probability of independent association.

In these cases, **Fisher's exact test** or **Barnard's exact test** can be used.

# Fisher's Exact Test

Fisher's test is the two-sample (conditional) analog to the exact one-sample binomial test and gives exact p-value results for any 2x2 table based upon the *hypergeometric distribution*. Fisher's exact test assumes the margins of the 2x2 table are fixed.

As the sample size increases, Fisher's exact p-value  $\rightarrow \chi^2$  test p-value (especially for Yates-corrected  $\chi^2$ ). This is a nice property, because for large samples Fisher's exact test can be computationally intensive.

*Fun fact:* Fisher supposedly derived this test for the “lady tasting tea” experiment.

# Barnard's Exact Test

Barnard's test considers three types of designs with different distributions for each:

- Cross-sectional studies are based on a *multinomial* distribution.
- Case-control studies use the product of two independent *binomial* distributions.
- Designs that stop once a set number of events has been observed would use a *hypergeometric* distribution.

The p-values are calculated by allowing different parts of the margin to vary (depending on the design), which results in more combinations to explore than Fisher's test.

It was originally less popular than Fisher's exact test due to being more computationally intensive. However it has been shown to generally be more powerful than Fisher's exact test, and computational concerns for small samples are no longer a concern with modern computing.

## Small Expected Cell Count Example

Suppose a retrospective study is done on the deaths of all men aged 50-54 in a specific county over a 1-month period. Of the 35 men who died of cardiovascular disease, 5 were on a high salt diet before they died. Of the 25 men who died of other causes, 2 were on a high salt diet. Is there an association between a high salt diet and CVD?

<i>Salt Levels</i>	<i>Death</i>		
	CVD	Other	
High	5	2	7
Low	30	23	53
	35	25	60

Are the expected values too small for the  $\chi^2$  test?



# Fisher's Exact Test Example

```
# Create matrix for data results
cvd <- matrix(c(5 ,2 ,30 ,23),ncol=2,byrow=T,
  dimnames = list(salt=c("high","low"),death=c("cvd","other")) )
# fit Fisher's exact test
fisher.test(cvd)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  cvd
## p-value = 0.6882
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.278957 21.620483
## sample estimates:
## odds ratio
##  1.897126
```

**Conclusion:**

# Barnard's Exact Test Example

```
library(DescTools) # load package to run test
BarnardTest(cvd)
```

```
##
## Barnards Unconditional 2x2-test
##
## data: cvd
## Score statistic = -0.74773, p-value = 0.596
## alternative hypothesis: two.sided
## sample estimates:
## Nuisance parameter
##           0.901
```

**Conclusion:**

# Test for Paired Samples

## Motivating Example

A study was done to compare two chemotherapy regimens. Subjects were matched by age and stage of disease. A random member of each pair received treatment A and the other was assigned to treatment B. The patients were followed for 5 years with survival as the outcome variable. There were 1242 patients total, and 621 pairs.

- Treatment A: Survival rate =  $526/621 = 0.847$  (84.7%)
- Treatment B: Survival rate =  $515/621 = 0.829$  (82.9%)

Is this (small) difference in survival significant? We might first set up the usual 2x2 contingency table appropriate for a  $\chi^2$  analysis:

<i>Treatment</i>	<i>Survive</i>		
	Yes	No	
A	526	95	621
B	515	106	621
	1041	201	1242

# Accounting for the Pairs

The  $\chi^2$  test is valid only if the two samples are *independent*. Since the patients in our example have been matched on age and stage, the groups are **not** independent.

We can set up a different type of 2x2 contingency table with the matched pair as the unit of observation. Here are the results for the matched pairs data:

<i>Outcome Trt A</i>	<i>Outcome Trt B</i>		
	Survive	Die	
Survive >5 Yrs	510	<b>16</b>	526
Die in 5 Yrs	<b>5</b>	90	95
	515	106	621

Pairs where the outcomes differed are known as *discordant pairs* (bold numbers). In these discordant pairs we see that a larger number showed the patient on treatment A surviving and the matched patient on treatment B dying than the reverse.

# McNemar's Test

**McNemar's test** is a test for paired data in  $2 \times 2$  tables. Its general procedure is to

- Ignore concordant pairs (pairs where the patients had the same outcome), focus on if the discordant pairs occur in equal frequencies
- Let  $p = P(\text{patient on Trt A lived given that the paired patients had different outcomes})$ , i.e., the probability of a discordant pair
- Test  $H_0 : p = \frac{1}{2}$

As is common with proportions, we can test the hypothesis using either exact or normal approximation methods:

- Large sample: for  $n_D = \text{number of discordant pairs} \geq 20$  we can use a normal theory test
- Small sample: for  $n_D < 20$  use exact binomial test

# McNemar's Test Underlying Calculations

Let  $n_D$  be the number of discordant pairs and  $n_A$  be the number of discordant pairs where Trt A patient lived (or Trt B patient lived, we just need to choose one for the calculations).

For the **large sample** with a continuity correction:

$$X^2 = \frac{(|n_A - \frac{n_D}{2}| - \frac{1}{2})^2}{\frac{n_D}{4}} \sim \chi_1^2, \text{ under } H_0 : p = 0.5$$

where our p-value =  $P(\chi_1^2 > X^2)$ .

For the **small sample** find the exact binomial probabilities:

$$\text{If } n_A < \frac{n_D}{2} : p = 2 \times \sum_{k=0}^{n_A} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D}$$

$$\text{If } n_A > \frac{n_D}{2} : p = 2 \times \sum_{k=n_A}^{n_D} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D}$$

$$\text{If } n_A = \frac{n_D}{2} : p = 1.0$$

# McNemar's Test Example

```
chemopairs <- matrix(c(510 ,16 ,5 ,90),ncol=2,byrow=T,  
  dimnames=list(trtA=c("survive", "5yr"),trtB=c("survive", "5yr")))  
mcnemar.test(chemopairs, correct=F) # no correction
```

```
##  
## McNemar's Chi-squared test  
##  
## data:  chemopairs  
## McNemar's chi-squared = 5.7619, df = 1, p-value = 0.01638
```

```
mcnemar.test(chemopairs, correct=T) # w/continuity correction
```

```
##  
## McNemar's Chi-squared test with continuity correction  
##  
## data:  chemopairs  
## McNemar's chi-squared = 4.7619, df = 1, p-value = 0.0291
```

## Conclusion:



# Connections to Other Tests

McNemar's Test is related to other statistical tests:

- It is equivalent to a one-sample method for paired qualitative data, *the Sign test*.
- It is a special case of the Cochran–Mantel–Haenszel test for stratified or matched data, which can accommodate more than just paired data.