

BIOS 600: Principles of Statistical Inference

$r \times c$ Contingency Tables: Part I

Fall 2012

Reading

- ▶ Pagano and Gauvreau, Chapter 15

Motivating Example: *Streptococcus pneumoniae*

Infections due to *Streptococcus pneumoniae* remain a substantial source of morbidity and mortality in both developing and developed countries despite a century of research and the development of therapeutic interventions such as multiple classes of antibiotics and vaccination. The World Health Organization estimates that in developing countries 814,000 children under the age of five die annually from invasive pneumococcal disease (IPD), with an estimated 1.6 million deaths affecting all ages globally.

Several recent studies have identified associations between pneumococcal serotypes (species variations) and patient outcomes from IPD. We consider data from a Scottish study of pneumococcal serotypes and mortality.

Contingency Tables

A *contingency table* is a display format for showing the relationship between two categorical variables. Below is a contingency table for a subset of serotypes from the Scottish study.

Serogroup	Died	Survived	Total
31	10	24	34
10	7	37	44
15	12	60	72
20	9	97	106
Total	38	218	256

Contingency Tables

Typical questions of interest:

- ▶ Is there an association between the row variable (indexed by r) and the column variable (indexed by c)?
- ▶ How strong is any association?

Pneumococcal Serotypes and IPD Mortality

First, we can generate the proportion of infected individuals who died for each serotype.

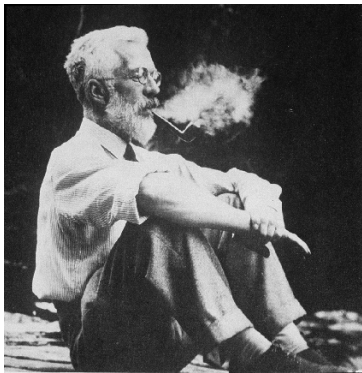
```
. tabi 37 7 \ 60 12 \ 97 9 \ 24 10 , row
```

Key			
frequency			
row percentage			

row	col		Total
	1	2	
1	37 84.09	7 15.91	44 100.00
2	60 83.33	12 16.67	72 100.00
3	97 91.51	9 8.49	106 100.00
4	24 70.59	10 29.41	34 100.00
Total	218 85.16	38 14.84	256 100.00

We would like to test H_0 : pneumococcal serotype is unrelated to mortality against the alternative H_A : pneumococcal serotype is related to mortality

Fisher's Exact Test



"Sir Ronald Fisher could be regarded as Darwin's
greatest twentieth-century successor." Richard Dawkins,

River out of Eden

Fisher's exact test is a great first choice for testing a relationship between two variables in a contingency table. While it has been around for almost 100 years, it was originally used only for very small samples due to the computational burden involved (this concern has been largely alleviated by modern computing).

Fisher's Exact Test

Fisher's exact test is fairly intuitive. The way it works is that we assume the column and row totals are fixed (so for our pneumococcus example, we assume we have 38 deaths and 218 survivors and that we have 34 in serogroup 31, 44 in serogroup 10, 72 in serogroup 15, and 106 in serogroup 20). Then, we construct all possible contingency tables with the same margins, and then sum up the probabilities of all tables as extreme or more extreme than our own table to get the p-value (recall the p-value is the probability of the observed data, or more extreme data, occurring under the null hypothesis).

Margins: row and column totals

Obviously, this was no fun before modern computing.

Fisher's Exact Test

Our Table

Serogroup	Died	Survived	Total
31	10	24	34
10	7	37	44
15	12	60	72
20	9	97	106
Total	38	218	256

Another Table With Same Margins

Serogroup	Died	Survived	Total
31	9	25	34
10	8	36	44
15	12	60	72
20	9	97	106
Total	38	218	256

Fisher's Exact Test for Pneumococcus Data

```
. tabi 37 7 \ 60 12 \ 97 9 \ 24 10 , exact
```

Enumerating sample-space combinations:

stage 4: enumerations = 1

stage 3: enumerations = 12

stage 2: enumerations = 114

stage 1: enumerations = 0

row	col		Total
	1	2	
1	37	7	44
2	60	12	72
3	97	9	106
4	24	10	34
Total	218	38	256

Fisher's exact = 0.027

What do we conclude?

Alcoholism Risk in 9/11 First Responders

Among firefighters and other “first responders” to the World Trade Center on September 11, 2001, there have been reports of increased alcohol-related difficulties (e.g., DUI). A survey of NYC firefighters (Bacharach, 2004) found the following.

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	309	793	1102
Did not participate	110	441	551
Total	419	1234	1653

Is alcoholism risk significantly related to 9/11 responder status?

Alcoholism Risk in 9/11 First Responders

Here the predictor (X) is 911 responder status, and the response (Y) is alcoholism risk level. We can look at the % high risk by 911 responder status.

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	309	793	1102
Did not participate	110	441	551
Total	419	1234	1653

9/11 responders: $\frac{309}{1102} = 28\%$ high risk

Other firefighters: $\frac{110}{551} = 20\%$ high risk

We can test H_0 : high risk status is unrelated to first-responder status against H_A : that the two are related using Fisher's exact test as follows.

Alcoholism Risk in 9/11 First Responders

```
. tabi 309 793 \ 110 441, exact
```

row	col		Total
	1	2	
1	309	793	1,102
2	110	441	551
Total	419	1,234	1,653

```
          Fisher's exact = 0.000  
1-sided Fisher's exact = 0.000
```

What do we conclude?

Chi-Square Test

Because we do have a fairly large sample, we can also test our hypothesis using a chi-square (χ^2) test. This test is valid in sufficiently large samples (cell sizes all > 10 for an 0.05-level test, with larger counts needed for smaller α levels), but Fisher's exact test is always valid. However, for very large samples, Fisher's exact test can still be too computationally expensive.

The chi-square test has a very nice motivation in terms of comparing observed proportions to proportions we would expect if H_0 were true.

Chi-Square Test

Suppose there is no association between 9/11 first responder status and alcoholism risk (H_0 true). From our observed table, we can see that $\frac{1102}{1653} = \frac{2}{3}$ or 67% of our study participants were 9/11 responders. Looking at alcoholism risk, we see that $\frac{419}{1653} = 0.2534$ or 25.34% of our subjects are at high risk of alcoholism.

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	309	793	1102
Did not participate	110	441	551
Total	419	1234	1653

Chi-Square Test

Overall probability a 9/11 responder: $\frac{1102}{1653}$

Overall probability high risk for alcoholism: $\frac{419}{1653}$

Using these two probabilities (and leaving the row and column totals fixed), we can calculate *expected* cell counts if H_0 is true. (When H_0 is true, the 9/11 responder status is independent of alcohol risk level.)

Chi-Square Test

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	?		1102 (67%)
Did not participate			551
Total	419 (25.34%)	1234	1653

? = Expected number of 9/11 responders at high risk if H_0 true

? = Probability of being both 9/11 responder and high risk \times number of subjects in study

$$? = \frac{1102}{1653} \left(\frac{419}{1653} \right) (1653) = 279.3$$

Chi-Square Test

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	279.3	?	1102
Did not participate			551
Total	419	1234	1653

? = Expected number of low risk 9/11 responders

? = Probability of being both 9/11 responder and low risk \times number of subjects in study

? = 1102 - expected number who are 9/11 responders at high risk

? = 1102 - 279.3 = 822.7

Chi-Square Test

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	279.3	822.7	1102
Did not participate		?	551
Total	419	1234	1653

? = Expected number of low-risk non-responders

? = Probability of being both 9/11 non-responder and low risk \times number of subjects in study

? = $1234 - \text{expected number who are 9/11 responders at low risk}$

? = $1234 - 822.7 = 411.3$

Chi-Square Test

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	279.3	822.7	1102
Did not participate	?	411.3	551
Total	419	1234	1653

? = Expected number of high risk non-responders

? = Probability of being both 9/11 non-responder and high risk \times number of subjects in study

? = $419 - \text{number of 9/11 responders at high risk} = 419 - 279.3 = 139.7$

? = $551 - \text{number of non-responders at low risk} = 551 - 411.3 = 139.7$

Chi-Square Test

	Alcoholism Risk		
	High	Low	
Participated in 9/11 response	279.3	822.7	1102
Did not participate	139.7	411.3	551
Total	419	1234	1653

So we can see that we had more 9/11 responders turn out to be at high risk for alcoholism than we expected. Is this difference statistically significant? (OK, we checked that already, but now let's try the chi-square test!)

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

- The chi-square test compares the observed frequencies (O) in each cell of the table to the expected frequencies (E) if H_0 is true.

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

- ▶ The chi-square test compares the observed frequencies (O) in each cell of the table to the expected frequencies (E) if H_0 is true.
- ▶ If differences between what we observe and expect ($O - E$) are large enough, we reject H_0 .

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

- ▶ The chi-square test compares the observed frequencies (O) in each cell of the table to the expected frequencies (E) if H_0 is true.
- ▶ If differences between what we observe and expect ($O - E$) are large enough, we reject H_0 .
- ▶ To combine differences across table cells, we need to square them (so that extra high-risk 9/11 responders are not cancelled out by fewer low-risk 9/11 responders) before adding them up.

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

- In addition, we need to *scale* the differences. That is, seeing 29 'extra' high-risk 9/11 responders is a big deal if our study only contains 100 subjects and is not a big deal if our study contains 100,000 subjects, so we divide by E to examine relative differences

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Low	Alcoholism Risk High	Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

- ▶ In addition, we need to *scale* the differences. That is, seeing 29 'extra' high-risk 9/11 responders is a big deal if our study only contains 100 subjects and is not a big deal if our study contains 100,000 subjects, so we divide by E to examine relative differences
- ▶ Our test statistic is $X^2 = \sum_{i=1}^{rc} \frac{(O_i - E_i)^2}{E_i}$, where rc is the number of cells in the table.

Chi-Square Test

- ▶ The distribution of this sum is approximated by a chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom, written $\chi^2_{(r-1)(c-1)}$ – see P&G Table A.8

Chi-Square Test

- ▶ The distribution of this sum is approximated by a chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom, written $\chi^2_{(r-1)(c-1)}$ – see P&G Table A.8
- ▶ Like the F distribution, there is a different χ^2 distribution for each degrees of freedom, and chi-square distribution is not symmetric

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

Let's carry out the test!

$$\blacktriangleright \chi^2 = \frac{(309-279.3)^2}{279.3} + \frac{(793-822.7)^2}{822.7} + \frac{(110-139.7)^2}{139.7} + \frac{(441-411.3)^2}{411.3} = 12.68$$

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

Let's carry out the test!

$$\blacktriangleright \chi^2 = \frac{(309-279.3)^2}{279.3} + \frac{(793-822.7)^2}{822.7} + \frac{(110-139.7)^2}{139.7} + \frac{(441-411.3)^2}{411.3} = 12.68$$

► OR: Stata

```
tabi 309 793 \ 110 441, all
```

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

Let's carry out the test!

►
$$\chi^2 = \frac{(309-279.3)^2}{279.3} + \frac{(793-822.7)^2}{822.7} + \frac{(110-139.7)^2}{139.7} + \frac{(441-411.3)^2}{411.3} = 12.68$$

► OR: Stata

```
tabi 309 793 \ 110 441, all
```

► We compare this to a $\chi^2_{(2-1)(2-1)} = \chi^2_1$ distribution which has critical value 3.84 at $\alpha = 0.05$ and reject H_0

Chi-Square Test

	Observed Data		Expected under H_0	
	Alcoholism Risk High	Alcoholism Risk Low	Alcoholism Risk High	Alcoholism Risk Low
Participated in 9/11 response	309	793	279.3	822.7
Did not participate	110	441	139.7	411.3

Let's carry out the test!

►
$$\chi^2 = \frac{(309-279.3)^2}{279.3} + \frac{(793-822.7)^2}{822.7} + \frac{(110-139.7)^2}{139.7} + \frac{(441-411.3)^2}{411.3} = 12.68$$

► OR: Stata

```
tabi 309 793 \ 110 441, all
```

► We compare this to a $\chi^2_{(2-1)(2-1)} = \chi^2_1$ distribution which has critical value 3.84 at $\alpha = 0.05$ and reject H_0

► What do we conclude?

Paired Data: McNemar's Test

- ▶ If data are paired and the outcome is a proportion, McNemar's test is often used

Paired Data: McNemar's Test

- ▶ If data are paired and the outcome is a proportion, McNemar's test is often used
- ▶ **Examples:**

Paired Data: McNemar's Test

- ▶ If data are paired and the outcome is a proportion, McNemar's test is often used
- ▶ Examples:
 - ▶ Case-control studies matched on a 1:1 basis (i.e., each case has a corresponding matched control)

Paired Data: McNemar's Test

- ▶ If data are paired and the outcome is a proportion, McNemar's test is often used
- ▶ Examples:
 - ▶ Case-control studies matched on a 1:1 basis (i.e., each case has a corresponding matched control)
 - ▶ **Twin studies**

Paired Data: McNemar's Test

- ▶ If data are paired and the outcome is a proportion, McNemar's test is often used
- ▶ Examples:
 - ▶ Case-control studies matched on a 1:1 basis (i.e., each case has a corresponding matched control)
 - ▶ Twin studies
 - ▶ Pre-/post data (baseline and follow-up on same individual)

Paired Data: McNemar's Test

- Contingency tables are restructured to accommodate matched pairs design

Paired Data: McNemar's Test

- ▶ Contingency tables are restructured to accommodate matched pairs design
- ▶ Cell frequencies are counts of *pairs*, not counts of individuals

Paired Data: McNemar's Test

- ▶ Contingency tables are restructured to accommodate matched pairs design
- ▶ Cell frequencies are counts of *pairs*, not counts of individuals
- ▶ Next we'll look at sample tables for a couple of settings

Matched Case-Control Study

Case	Control	
	Exposed	Unexposed
Exposed	a	b
Unexposed	c	d

The counts in this table for a case-control study are the number of *matched pairs* not the number of subjects, so there are $2(a + b + c + d)$ subjects.

H_0 : The proportion of subjects exposed to the risk factor is the same for cases and controls.

Twin Study

	Unexposed Twin	
	Diseased	Healthy
Exposed Twin Diseased	a	b
Healthy	c	d

The counts in this table are the number of sets of twins, not the number of individuals, so there are $2(a + b + c + d)$ individuals.

H_0 : The proportions diseased are equal for exposed and unexposed twins.



Baseline/Follow-up Data on Individuals

	Follow-Up	
Baseline	Diseased	Healthy
Diseased	a	b
Healthy	c	d

The counts in this table are the number of individuals, not the number of measurements (which is twice the number of individuals), so there are $2(a + b + c + d)$ measurements on $a + b + c + d$ subjects.

H_0 : The proportion of diseased subjects is the same at baseline and follow-up.

McNemar's Test

- ▶ Cells b and c are called *discordant* because they represent pairs with a difference

McNemar's Test

- ▶ Cells b and c are called *discordant* because they represent pairs with a difference
- ▶ Cells a and d are called *concordant* and do not contribute information about any difference

McNemar's Test

- ▶ Cells b and c are called *discordant* because they represent pairs with a difference
- ▶ Cells a and d are called *concordant* and do not contribute information about any difference
- ▶ If the null hypothesis is true, then we expect $b = c$

McNemar's Test

- ▶ Cells b and c are called *discordant* because they represent pairs with a difference
- ▶ Cells a and d are called *concordant* and do not contribute information about any difference
- ▶ If the null hypothesis is true, then we expect $b = c$
- ▶ McNemar's test statistic $X^2 = \frac{(b-c)^2}{b+c}$ measures how far b is from c . Under H_0 , $b - c = 0$.

McNemar's Test

- ▶ Cells b and c are called *discordant* because they represent pairs with a difference
- ▶ Cells a and d are called *concordant* and do not contribute information about any difference
- ▶ If the null hypothesis is true, then we expect $b = c$
- ▶ McNemar's test statistic $X^2 = \frac{(b-c)^2}{b+c}$ measures how far b is from c . Under H_0 , $b - c = 0$.
- ▶ We compare X^2 to the χ^2_1 distribution and reject H_0 for large values of X^2

Example: Health Insurance Mandate

A Gillings School of Global Public Health graduate student conducted a poll of 250 heterosexual married couples in the RTP area and asked whether each partner supported the health insurance mandate or not, obtaining the following data.

	Wife	
Husband	For	Against
For	75	45
Against	74	56

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.



Example: Health Insurance Mandate

	Wife	
Husband	For	Against
For	75	45
Against	74	56

► What is H_0 ?

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

Example: Health Insurance Mandate

	Wife	
	For	Against
Husband		
For	75	45
Against	74	56

- ▶ What is H_0 ?
- ▶ What H_0 is tested by McNemar's test?

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

Example: Health Insurance Mandate

	Wife	
	For	Against
Husband		
For	75	45
Against	74	56

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

- ▶ What is H_0 ?
- ▶ What H_0 is tested by McNemar's test?
- ▶ $\chi^2 = \frac{(b-c)^2}{b+c} = ?$

Example: Health Insurance Mandate

	Wife	
Husband	For	Against
For	75	45
Against	74	56

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

- ▶ What is H_0 ?
- ▶ What H_0 is tested by McNemar's test?
- ▶ $\chi^2 = \frac{(b-c)^2}{b+c} = ?$
- ▶ Conduct the test (hint: mcci 75 74 45 56)

Example: Health Insurance Mandate

	Wife	
	For	Against
Husband		
For	75	45
Against	74	56

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

- ▶ What is H_0 ?
- ▶ What H_0 is tested by McNemar's test?
- ▶ $\chi^2 = \frac{(b-c)^2}{b+c} = ?$
- ▶ Conduct the test (hint: **mcci**
75 74 45 56)
- ▶ What do you conclude?

Example: Health Insurance Mandate

	Wife	
Husband	For	Against
For	75	45
Against	74	56

The graduate student would like to know whether the opinions of the husbands and wives are independent, or whether they are related.

- ▶ What is H_0 ?
- ▶ What H_0 is tested by McNemar's test?
- ▶ $\chi^2 = \frac{(b-c)^2}{b+c} = ?$
- ▶ Conduct the test (hint: **mcci 75 74 45 56**)
- ▶ What do you conclude?
- ▶ What should the graduate student do?

Odds Ratio (OR)

Suppose we have a disease D (e.g., lung cancer) and two groups, E and E^c (e.g., E =smokers and E^c =nonsmokers).

$$OR = \frac{\left\{ \frac{Pr(D|E)}{1-Pr(D|E)} \right\}}{\left\{ \frac{Pr(D|E^c)}{1-Pr(D|E^c)} \right\}}$$

The OR ranges from 0 to ∞ . When $OR = 1$, there is no association between two variables.

Odds Ratio (OR)

Consider the following contingency table.

	Exposed	Unexposed	Total
Disease	a	b	$a + b$
No Disease	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$$\widehat{Pr}(D | E) = \frac{a}{a + c}$$

$$\widehat{Pr}(D | E^c) = \frac{b}{b + d}$$

Odds Ratio (OR)

$$\widehat{Pr}(D | E) = \frac{a}{a + c}$$

$$\widehat{Pr}(D | E^c) = \frac{b}{b + d}$$

$$OR = \frac{\left\{ \frac{Pr(D|E)}{1 - Pr(D|E)} \right\}}{\left\{ \frac{Pr(D|E^c)}{1 - Pr(D|E^c)} \right\}}$$

$$\widehat{OR} = \frac{\left\{ \frac{\frac{a}{a+c}}{\frac{c}{a+c}} \right\}}{\left\{ \frac{\frac{b}{b+d}}{\frac{d}{b+d}} \right\}} = \frac{ad}{bc}$$

For 9/11 responder data data, $\widehat{OR} = \frac{309(441)}{793(110)} = 1.56$. In Stata, use `cci 309 793 110 441` to get the estimated OR and 95% CI, which is 1.56 (1.21, 2.02). We interpret this by saying that 9/11 responders have 1.56 times the odds of being high risk for alcoholism as their firefighter colleagues who were not 9/11 responders.

Caution: Berkson's Fallacy

Due to ease of data collection, a restricted sample is often used to test a hypothesis of interest. We illustrate a common problem with this approach.

Researchers are interested in the special population of HIV+ women on antiretroviral therapy in sub-Saharan Africa. They would like to know whether in this population, a new pregnancy is related to the probability of having an AIDS-defining event.

In order to test their hypothesis, they recruit women from a large network of health care clinics and find the following.

Pregnant	AIDS		Total
	Yes	No	
Yes	31	44	75
No	124	99	223
Total	155	143	298

Their estimated OR and 95% CI are 0.56 (0.32, 0.99). Should HIV+ women on antiretroviral therapy try to get pregnant?

Caution: Berkson's Fallacy

Now we consider all HIV+ area women, not just those who visited a health care clinic during the study period.

Pregnant	AIDS		Total
	Yes	No	
Yes	44	175	219
No	248	990	1238
Total	292	1165	1457

Their estimated OR and 95% CI are 1.00 (0.68, 1.44). What happened?

Caution: Berkson's Fallacy

The bias in the clinic-based sample comes from the fact that not all HIV+ women are equally likely to visit a health clinic.

Diagnosis	# with Clinic Visit	Total Women	Pr(Visit)
Pregnant+AIDS	31	44	$\frac{31}{44} = 0.70$
Pregnant only	44	175	0.25
AIDS only	248	124	0.50
Neither	99	990	0.10

The observed spurious relationship, observed only because of how we chose the sample, is called *Berkson's fallacy* and is often a danger with clinic- or hospital-based samples.