Soci708 – Statistics for Sociologists

Module 9 – Analysis of Two-Way Tables[1]

François Nielsen

University of North Carolina
Chapel Hill

Fall 2009

[1] Adapted from slides for the course Quantitative Methods in Sociology (Sociology 6Z3) taught at McMaster University by Robert Andersen (now at University of Toronto)

Goals of This Module

- We look at two categories of applications of the χ^2 ("chi-squared") test:
  1. The χ^2 test for independence of categorical variables
     - Inference for several proportions simultaneously in a contingency table
  2. The χ^2 test for goodness of fit
     - Test of the fit of a theoretical distribution (of any kind) to categorical data

Introducing Inference for Contingency Tables

- Recall that last we week we discussed the z-statistic to test for difference between two proportions (a *two-sample z-test*)
- We can see this as a test for the difference between two categorical variables, each with only two categories
- In other words, we essentially had a contingency table with two columns and two rows, e.g.:

	Men	Women
Employed	298 (93%)	320 (80%)
Unemployed	22 (7%)	80 (20%)
Total (n)	320 (100%)	400 (100%)

- In this example we would test the difference in unemployment level for men and women

Two-sample Test for Proportions Revisited

- Assume we want to test the alternative hypothesis that men have lower unemployment:

$$H_a : p_1 < p_2$$

- The null hypothesis is that there is no difference in unemployment or level of men is greater:

$$H_0 : p_1 \geq p_2$$

- The pooled sample proportion is then:

$$\hat{p} = \frac{22 + 80}{320 + 400} = \frac{102}{720} = .1416667$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{.06875 - .20}{\sqrt{.1417(1 - .1417)\left(\frac{1}{320} + \frac{1}{400}\right)}} = -5.018526$$

Comparing Several Proportions: The χ^2 Test (Chi-squared Test)

- Since the one-sided P-value for $z = -5.018526$ is much less than .001 we have strong evidence against the null hypothesis and thus reject it
- The difference-of-proportions test works well if we have only two proportions to compare, but it will not suffice if we have more than two proportions
  - In other words, if we have variables with more than two categories or proportions from more than two samples, this test is not applicable
- This brings us to a new test called the χ^2 test (or chi-squared, pronounced "kai" squared)
  - The χ^2 test is *a difference of proportions test that can be generalized to any number of categories of the explanatory and response variables*

Chi-squared Test for Two Proportions

1. Calculate the expected count in each of the cells of the table
    - An expected count is is the number of cases that would fall into a category under the assumption that the null hypothesis is true and the row and column variables are independent of each other (i.e., the variables are unrelated)
    - So, following the previous example we would expect that *equal proportions* of men and women are unemployed. This is simply the pooled proportion we used earlier:

$$\hat{p} = \frac{22 + 80}{320 + 400} = \frac{102}{720} = .142$$

Chi-squared Test for Two Proportions (2)

- From this pooled proportion we can now easily determine the expected counts in each cell of the table
  - For *unemployed men the expected count* is simply the total number of men in the sample multiplied against the pooled proportion: $320 \times .142 = 45.44$
  - Similarly, for *unemployed women the expected count* is $400 \times .142 = 56.8$
  - We can also determine the expected counts for the other cells. So, for *employed men the expected count* is: $320 \times (1 - .142) = 274.56$
  - The *expected count for employed women* is: $400 \times (1 - .142) = 343.2$
- These expected counts are what *we would observe on average over the long run if there were no difference between men and women*

Chi-squared Test for Two Proportions (3)

- Another way to calculate expected counts is to start with joint probabilities expected under the assumption that gender and employment status are *independent*
- Expected joint probabilities are then the products of the marginal probabilities:

	Men	Women	Total
Employed	$.858 \times .444 = .3814$	$.858 \times .556 = .4769$	.858
Unemployed	$.142 \times .444 = .0630$	$.142 \times .556 = .0787$	.142
Total	.444	.556	1.0

- Expected counts are obtained by multiplying the expected joint probabilities by $N = 720$
  - E.g., $.3814 \times 720 = 274.63$ for employeed men, etc.
- This is also called *fitting the marginals* to the table

Chi-squared Test for Two Proportions (4)

2. A shortcut formula for calculating expected counts is as follows:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{\text{table total } (n)}$$

- Since expected counts are based on sample data, they should be thought of more accurately as *estimated* expected counts
  - That is, just as with other statistical inferences based on random sampling, we don't know what the population distributions for the variables are

Chi-square Test for Two Proportions (5)

3. We now proceed to see how different the observed counts are away from the expected counts
   - The chi-squared test allows us to compare these two values:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

   - The chi-squared test is simply a more complicated way of calculating the difference-of-proportions z-test
   - There are two differences, however
     3.1 The difference-in-proportions z-test can be used for a directional alternative hypothesis but the chi-squared is nondirectional
     3.2 The chi-squared test can be used for more than two proportions

Chi-squared Test for Two Proportions (6)

- As is often the case, it can be useful to calculate the chi-squared test using a table to keep things straight
- The chi-squared statistic for the example data is then:

Obs	Exp	Obs−Exp	$\frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}}$
298	274.6	23.4	2.001
22	45.4	−23.4	12.061
320	343.2	−23.2	1.568
80	56.8	23.2	9.476
720	720	0	$\chi^2 = 25.11$

- As usual, we now must proceed to determine whether this chi-square value is statistically significant
- We do so, by examining its place in the chi-square distribution

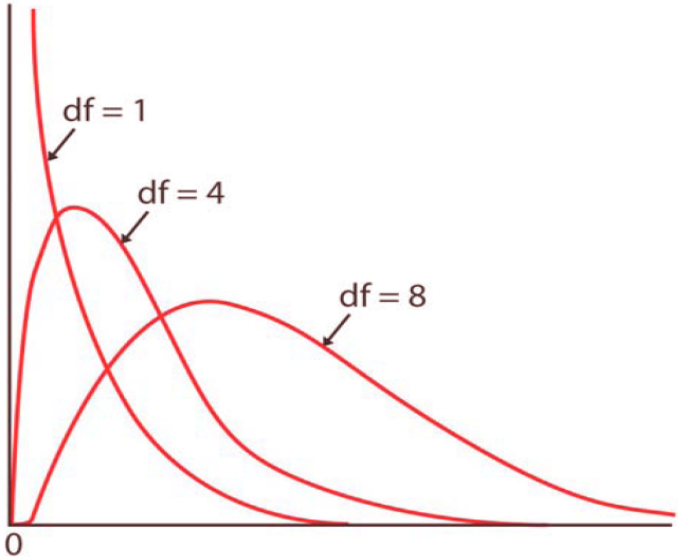
The Chi-square Distributions

- Like the t-distributions, the χ^2 distributions are a family of density curves indexed by degrees of freedom (df)
  - The χ^2 *distribution is not symmetric, however – it is positively skewed*
- For each df there is a different χ^2 distribution
  - As the df increase, the distribution becomes less skewed
  - Recall that the df is the number of values that are allowed to vary when computing the statistic. If we fill in all but the last row and column of the expected counts, we can calculate the remaining values by subtraction
  - Therefore df are calculated from the number of rows and columns in the table:

$$df = (r-1)(c-1)$$

    where r is the number of rows and c is the number of columns in the table

The Chi-square Distributions (2)



The Chi-square Distributions (3)

- Recall that our example gave a $\chi^2 = 25.11$
- Since we have two rows and two columns, the degrees of freedom are:

$$df = (r-1)(c-1) = (2-1)(2-1) = 1$$

- If we look at Table E in Moore et al., we see that this value corresponds to an area to the right of less than 0.0005. (The P-value is the area to the right of χ^2 under the density curve)
  - In other words, *the chi-square value is statistically significant, suggesting that the difference in proportions is unlikely due to chance alone*
- We can conclude, then, that there is a statistically significant relationship between gender and unemployment

χ^2 Test for 2 Proportions in Stata

Employment and Gender Example

```
. * in Stata
. tabi 298 320\ 22 80, chi2 col nofreq V

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |     93.13      80.00 |     85.83
         2 |      6.88      20.00 |     14.17
-----------+----------------------+----------
     Total |    100.00     100.00 |    100.00

          Pearson chi2(1) =  25.1856   Pr = 0.000
               Cramér's V =   0.1870

. * Note Cramér's V option spelled capital "V"
. * Cramér's V in 2 x 2 table is called ``phi''
. * V = .187 means gender and employment are
. * correlated .187
```

The Chi-square Distributions (4)

Relationship Between z Test and χ^2 Test

- Our example data (employment status by gender) gives a $\chi^2 = 25.11$
- Recall that for the same data the z test for the difference in unemployment (men minus women) was -5.018526 Is this a coincidence?
- It is not! In fact, a χ^2 variate with 1 df is equal to the the square of a z variate, so that the square root of the χ^2 is equal to z (adjusting the sign to that of the original difference in the proportions):

```
  . * in Stata
  . display sqrt(25.1856)
5.0185257
```

Chi-square Test – Tables Larger Than 2×2

An example

- The χ^2 test works exactly the same way for tables larger than 2×2
- Assume the following 3×3 table:

	Country			
Education	England	Scotland	Wales	Total
Degree	776	401	307	1484
Secondary	3306	1456	1053	5815
None	1667	927	789	3383
N	5749	2784	2149	10682

- Our goal is to test whether education level differs according to country
- We should always begin by calculating the percentages so that we can better see – and discuss – the pattern in the data

Chi-square Test – Tables Larger Than 2×2

An example (2)

- The table below contains the same information as the previous table, except now the counts show percentages for each education category by country

	Country			
Education	England	Scotland	Wales	Total
Degree	13.5	14.4	14.3	1484
Secondary	57.5	52.3	49.0	5815
None	29.0	33.3	36.7	3383
N	5749	2784	2149	10682

- The null hypothesis is that all of the proportions are equal:

$$H_0 : p_{\text{England}} = p_{\text{Scotland}} = p_{\text{Wales}}$$

- The alternative hypothesis is:

H_a : At least one of the proportions differs from the others

Chi-square Test – T Larger Than 2×2

An example (3)

- Expected counts are found in the usual way.
- Here is the expected count for the those with a *degree in England:*

$$\begin{aligned}\text{expected count} &= \frac{\text{row total} \times \text{column total}}{\text{table total } (n)} \\ &= \frac{1,484 \times 5,749}{10,682} \\ &= 798.7\end{aligned}$$

- As with the 2×2 table, we do this for every cell in the table
- We then compare the observed counts with the expected counts using the χ^2 test

Chi-square Test – Tables Larger Than 2×2

An example (4)

- The complete table of expected counts for these data is:

	Country			
Education	England	Scotland	Wales	Total
Degree	798.7	386.8	298.6	1484
Secondary	3129.6	1515.5	1169.9	5815
None	1820.7	881.7	680.6	3383
N	5749	2784	2149	10682

- It is important to note here that the expected counts sum (within rounding error) to the *observed row and column marginals*
  - In other words, if your calculations don't square with the observed total row and total column, you have done something wrong

Chi-square Test – Tables Larger Than 2×2

An example (5)

- We then calculate the χ^2 test and its degrees of freedom in the usual way:

$$
\begin{aligned}
\chi^2 &= \sum_{\text{all cells}} \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \\
&= \frac{(776 - 798.7)^2}{798.7} + \frac{(3306 - 3129.6)^2}{3129.6} \\
&+ \cdots + \frac{(798 - 680.6)^2}{680.6} \\
&= 57.93
\end{aligned}
$$

- The *df* are: $(r-1)(c-1) = (3-1)(3-1) = 2 \times 2 = 4$
- The P-value for a $\chi^2 = 57.93$ with 4 *df* is much less than 0.0005 so we can reject the null hypothesis that there is no relationship between the two variables

Breaking down the components of the Chi-square test:

Cell	Obs	Exp	Obs−Exp	$(\text{Obs}-\text{Exp})^2$	$\frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}}$
England					
Degree	776	798.7	−22.7	515.29	.645
Secondary	3306	3129.6	176.4	31116.96	9.94
None	1667	1820.7	−153.7	23623.69	12.97
Scotland					
Degree	401	386.8	14.2	201.64	0.52
Secondary	1456	1515.5	−59.5	3540.25	2.34
None	927	881.7	45.3	2052.09	2.33
Wales					
Degree	307	298.6	8.4	70.56	.24
Secondary	1053	1169.9	−116.9	13665.61	11.68
None	789	680.6	108.4	11750.56	17.27
$\sum$			0		57.93

χ^2 in R

Distribution of Education by Region of the UK

```
> # in R
> observed <- matrix(nrow=3, ncol=3,
           c(776,3306,1667,401,1456,927,307,1053,789))
> # a matrix specifying the observed counts
                        in each cell of the table
> chisq.test(observed)

        Pearson's Chi-squared test

data:  observed
X-squared = 57.9348, df = 4, p-value = 7.875e-12
```

χ^2 in Stata

Distribution of Education by Region of the UK

```
. * in Stata
. tabi 776 401 307\ 3306 1456 1053\ 1667 927 789, chi2 V

           |                 col
       row |         1          2          3 |     Total
-----------+---------------------------------+----------
         1 |       776        401        307 |     1,484
         2 |     3,306      1,456      1,053 |     5,815
         3 |     1,667        927        789 |     3,383
-----------+---------------------------------+----------
     Total |     5,749      2,784      2,149 |    10,682

          Pearson chi2(4) =  57.9348   Pr = 0.000
               Cramér's V =   0.0521

. * Note Cramér's V option is capital "V"
```

Measuring Strength of Association (Effect Size) (1)

Using Cramér's V or ϕ to Measure Effect Size

- The value of the χ^2 statistic depends on N and is thus not easily comparable across studies
- A more comparable measure of the strength of association between two categorical variables is provided by Cramér's V:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

  where k is the number of rows or the number of columns, whichever is less
- When $k = 2$ Cramér's V reduces to ϕ ("phi") given by

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Measuring Strength of Association (Effect Size) (2)

Using Cramér's V or ϕ to Measure Effect Size

- Cramér's V or ϕ ranges from 0 to 1 and can be interpreted as a correlation coefficient:
- In the gender and employment example

$$V = \phi = \sqrt{25.1856/720} = 0.1870$$

means that gender and employment status are correlated .187
- In the distribution of education by region in the UK example

$$V = \sqrt{57.9348/(10,682 \times 2)} = 0.0521$$

means that educational distribution and region are correlated only .052 (not very big!)
- For birth defect by mother diabetic status in Pima (next slide)

$$V = \phi = \sqrt{25.5111/1,207} = 0.1454$$

means that diabetic status of mother and one or more birth defect in the newborn are correlated .145

Another Example of Association Test

```
. tabi 31 754\ 13 362\ 9 38, chi2

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |        31        754 |       785 
         2 |        13        362 |       375 
         3 |         9         38 |        47 
-----------+----------------------+----------
     Total |        53      1,154 |     1,207 

          Pearson chi2(2) =  25.5111   Pr = 0.000
                 Cramér's V =   0.1454
```

Child Birth Defects by Mother's Diabetic Status Among Pima Indian Mothers

Data	1+ Defect	No Defect	Total
Nondiabetic	31	754	785
Prediabetic	13	362	375
Diabetic	9	38	47
Total	53	1154	1207

- Is there a statistically significant association between birth defect and diabetic status of mother?

Antireligionist Example (1)

Allow Antireligionist to Teach in College or University? Frequencies[2]

```
. tab degree colath, chi2

                |         ALLOW
                |  ANTI-RELIGIONIST TO
     RS HIGHEST |         TEACH
         DEGREE |   ALLOWED  NOT ALLOW |     Total
----------------+----------------------+----------
LT HIGH SCHOOL |       105        158 |       263
   HIGH SCHOOL |       524        397 |       921
JUNIOR COLLEGE |        91         53 |       144
      BACHELOR |       239         74 |       313
      GRADUATE |       108         22 |       130
----------------+----------------------+----------
         Total |     1,067        704 |     1,771

          Pearson chi2(4) = 112.4130   Pr = 0.000
```

[2]This and next 2 slides repeated from Module 4

Antireligionist Example (2)

Allow Antireligionist to Teach in College or University? Row Percentages

```
. tab degree colath, chi2 row nofreq

                |          ALLOW
                | ANTI-RELIGIONIST TO
     RS HIGHEST |         TEACH
         DEGREE |   ALLOWED  NOT ALLOW |     Total
----------------+----------------------+----------
LT HIGH SCHOOL |     39.92       60.08 |    100.00
   HIGH SCHOOL |     56.89       43.11 |    100.00
JUNIOR COLLEGE |     63.19       36.81 |    100.00
      BACHELOR |     76.36       23.64 |    100.00
      GRADUATE |     83.08       16.92 |    100.00
----------------+----------------------+----------
          Total |     60.25       39.75 |    100.00

          Pearson chi2(4) = 112.4130   Pr = 0.000
```

Antireligionist Example (3)

Allow Antireligionist to Teach in College or University? Table in publication form combining output from two Stata runs (frequencies only and row percentages)

Should [somebody who is against all churches and religions] be allowed to teach in a college or university, or not?

Highest degree	Yes	No	Total	N
Less than High School	39.9	60.1	100.00	263
High School	56.9	43.1	100.00	921
Junior College	63.2	36.8	100.00	144
Bachelor	76.4	23.6	100.00	313
Graduate	83.1	16.9	100.00	130
Total	60.3	39.8	100.00	
N	1067	704		1771

SOURCE: General Social Survey 1998

Things to Know about the χ^2 Test

1. The χ^2 test is used to test the null hypothesis that there is *no relationship between two categorical variables*
2. We assume that we have *simple random samples*
3. The χ^2 test should be used only when no more than 20% of the expected counts are smaller than 5 and all of the expected counts are 1 or greater
   - For a 2×2 (row $\times$ column) table all four cells should have expected counts of 5 or more
4. For a 2×2 table the χ^2 test is simply the square of the z-statistic for a test for proportions, and the P-value for the χ^2 is the same as the P-value for the two-sided z test
   - A z-test is preferred for 2×2 tables because it gives the choice of one or two sided tests
5. If observed and expected frequency are identical – i.e., no relationship – the $\chi^2 = 0$
6. The larger the difference between expected and observed counts, the larger the χ^2

χ^2 Test of Goodness of Fit

Jury Pool Example (Koopmans 1987, p.413ff)

Jury Pool Age Distribution

Age Group	Observed	Census
18-19	23	0.061
20-24	96	0.150
25-29	134	0.135
30-39	293	0.217
40-49	297	0.153
50-64	380	0.182
65+	113	0.102
Total	1336	1.000

- The fairness of representation of the jury pool in a large municipal court district is contested in a court case
- A SRS with $n = 1336$ is drawn from the jury pool of the district
- The age distribution of the sample is compared to the age distribution of the population of the whole district given by the census
  - Is the age distribution of the jury pool the same as the age distribution of the district?
  - For legal case, *any* variable will do to show non-representation

χ^2 Test of Goodness of Fit (2)

Jury Pool Example (Koopmans 1987, p.413ff)

- Hypotheses:

$$H_0 : p_1 = .061, p_2 = .150, \ldots, p_7 = .102$$
$$H_a : p_i \neq p_{i0} \text{ for some } i$$

- Calculations for χ^2 test:

Fit of Jury Pool Age Distribution With Census Distribution

Age Group	Observed	Census	Expected	$\frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}}$	z	Symbol
18-19	23	0.061	81.5	41.99	−6.480	@
20-24	96	0.150	200.4	54.39	−7.375	@
25-29	134	0.135	180.4	11.92	−3.452	@
30-39	293	0.217	289.9	0.03	0.181	.
40-49	297	0.153	204.4	41.94	6.476	@
50-64	380	0.182	243.2	77.02	8.776	@
65+	113	0.102	136.3	3.97	−1.994	O
Total	1336	1.000	1336.0	231.26		

χ^2 with $7 - 1 = 6$ df: 231.26. $P(\chi^2_6 > 231.26) = 0.000000$

  - We conclude that the jury pool is *not* representative of the district population

χ^2 Test of Goodness of Fit (3)

Jury Pool Example (Koopmans 1987, p.413ff)

- I have not yet figured out how to do this in Stata (if you find out, please let me know), but here is how it can be done in R:

```
> # in R
> freqs<-c(23, 96, 134, 293, 297, 380, 113)
> ps<-c(0.061, 0.150, 0.135, 0.217, 0.153, 0.182, 0.102)
> chisq.test(freqs, correct=FALSE, p=ps)

        Chi-squared test for given probabilities

data:  freqs
X-squared = 231.26, df = 6, p-value < 2.2e-16
```

Entering a Published Table in Stata

Replicating the Drinks by Religious Denomination Example

- In a spreadsheet enter the following file, then select and copy to the clipboard

relig	drinks	count
1	0	39
2	0	6
3	0	2
4	0	5
1	1	94
2	1	40
3	1	21
4	1	47

```
. * in Stata paste into Data Editor
. * then CLOSE Data Editor
(3 vars, 8 obs pasted into editor)

. tab relig drinks [fweight=count], all

           |        drinks
     relig |         0          1 |     Total
-----------+----------------------+----------
         1 |        39         94 |       133
         2 |         6         40 |        46
         3 |         2         21 |        23
         4 |         5         47 |        52
-----------+----------------------+----------
     Total |        52        202 |       254

          Pearson chi2(3) =  13.6827   Pr = 0.003
 likelihood-ratio chi2(3) =  14.4258   Pr = 0.002
               Cramér's V =   0.2321
                    gamma =   0.4915  ASE = 0.121
          Kendall's tau-b =   0.2105  ASE = 0.051
```

- Note use of the `[fweight=count]` clause