Introduction to Social Statistics

Day 21

Instructor Rob Kemp

MondayApril 1, 2013

Announcements

- Next Test: April 5th
  - No More lectures
  - Only one new statistical test!!
- Homework 3 Posted
  - Only 2 Chapters: 10 & 11

Agenda: Chi-Square

- Review Chi Square as a Statistical Test
- Limitations of the Chi-Square test

Chi-square test

Step 1: Research and Null Hypotheses SES and smoking

- H_0: There is no association between the two variables. That is, health and education are *statistically independent*.
- H_1: There is a relationship between the two variables. That is, health and education are *statistically dependent*.
- alpha = 0.05

Step 2: Chi-Square (Obtained)

- The test statistic that summarizes the differences between the observed (f_o) and the expected (f_e) frequencies in a bivariate table.

Calculating the Obtained- Chi-Square

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

f_e = expected frequencies
f_o = observed frequencies

Data

Table 11.7 Health by Educational Level: GSS 2006

	Educational Level			
Health	*Less Than High School*	*High School Degree*	*Some College or More*	*Total*
Poor	16 (12.5%)	26 (6.0%)	6 (2.1%)	48 (5.7%)
Moderate	44 (34.4%)	79 (18.1%)	39 (13.9%)	162 (19.2%)
Good	52 (40.6%)	213 (48.9%)	137 (48.9%)	402 (47.6%)
Excellent	16 (12.5%)	118 (27.1%)	98 (35.0%)	232 (27.5%)
Total	128 (100%)	436 (100.1%)	280 (99.9%)	844 (100%)

Chi-Square Calculations

Table 11.8 Calculating Chi-Square for Education and Health

Education and Health	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Less than high school/poor	16	7.3	8.7	75.69	10.37
Less than high school/moderate	44	24.6	19.4	376.36	15.30
Less than high school/good	52	61.0	−9.0	81.0	1.33
Less than high school/excellent	16	35.2	−19.2	368.64	10.47
High school/poor	26	24.8	1.2	1.44	0.06
High school/moderate	79	83.7	−4.7	22.09	0.26
High school/good	213	207.7	5.3	28.09	0.13
High school/excellent	118	119.8	−1.8	3.24	0.03
Some college or more/poor	6	15.9	−9.9	98.01	6.16
Some college or more/moderate	39	53.7	−14.7	216.09	4.02
Some college or more/good	137	133.4	3.6	12.96	0.10
Some college or more/excellent	98	77.0	21	441	5.73

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 53.96$$

F_e= (Row Marginal*Column Marginal)/N
Marginal=sub-total or the total for the row/column.

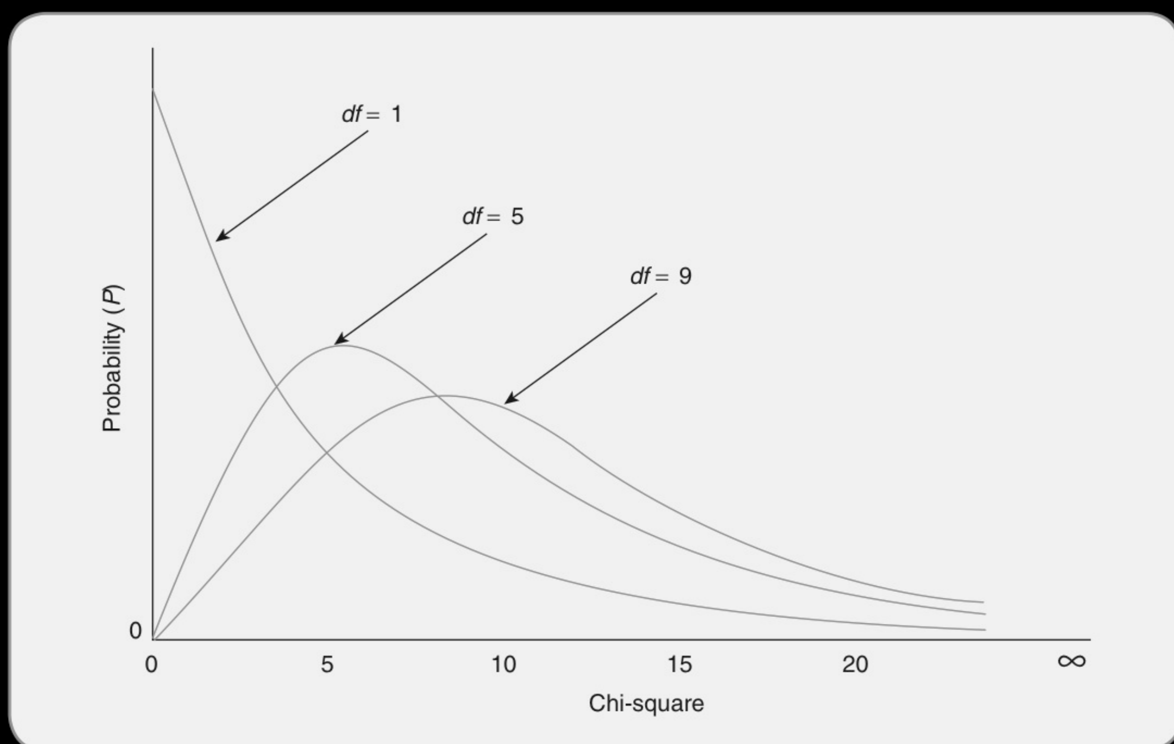
The Sampling Distribution of Chi-Square

- The sampling distribution of chi-square tells the probability of getting values of chi-square, assuming no relationship exists in the population.
- Chi-square values are always positive. The minimum possible value is zero, with no upper limit to its maximum value.

The Sampling Distribution of Chi-Square

- Similar to T-Statistic as it is determined by the DF of the table being used.

Figure 11.1 Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom



Step 3:Determining Degrees of Freedom

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

where

r = the number of rows

c = the number of columns

Calculating Degrees of Freedom

How many degrees of freedom would a table with 4 rows and 3 columns have (our example table)?

(4-1)*(3-1)=3*2=6

Therefore, this 4x3 table has 6 degrees of freedom.

Step 4: Comparing test statistic

Look up Critical Value in Appendix D: Chi-Square distribution

Strategy 1:

- Find the chi-square closest to the obtained chi-square you calculated, given your degrees of freedom.
- Find the p-value (across the top) associated with your obtained chi-square. This tells you the degree of certainty with which you can you reject the null hypothesis.

Strategy 2:

- Set alpha (e.g., at 0.05), then find the chi-square statistic associated with your degrees of freedom
- Compare your obtained chi-square to the associated chi-square; if obtained chi-square > chi-square statistic, then can reject the null hypothesis at alpha level of probability (e.g., 0.05)

Comparing test statistic

- Appendix D: Chi-Square distribution
  - X2=53.96
  - df=4
  - alpha-level=0.05

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Step 5: Determining results

- H_0: There is no association between the two variables. That is, health and education are *statistically independent*.
- H_1: There is a relationship between the two variables. That is, health and education are *statistically dependent*.

- X^2: $53.96 > 12.592$
  - *Therefore, we can reject the null hypothesis that there is no association between*

Limitations of the Chi-Square Test

- The Chi-Square test does <u>not</u> give us much information about the strength of the relationship or its substantive significance in the population.
- The chi-square test is sensitive to sample size, in fact the chi-square statistic is directly proportional to sample size.
- The chi-square test is also sensitive to small expected frequencies in one or more cell of the table.

Example 1

- 1. Pre-school Attendance and Pre-algebra Achievement (these are contrived data, based on a real study)
- In these times of educational reform, attention has been focused on pre-school for all children. Since many districts are facing budget cuts, funding pre-school programs may impact other offerings. Before making their recommendations, administrators in a large urban district take a random sample of 50 seventh graders and compare the pre-algebra achievement levels of those who attended pre-school and those who did not. If achievement is independent of attending pre-school then the proportions at each level should be equal.

Example 1

1. Null and Research Hypothesis set alpha at 0.05
2. Calculate Chi-Square
3. Calculate degrees of freedom and find Critical Value
4. Compare test statistic
5. Conclusions

	Below Grade Level	At Grade Level	Advanced	Total
Pre-School	8	6	6	20
No Pre-School	6	15	9	30
Total	14	21	15	50

Example 1

- Step 1: Null and Research Hypotheses
  - H_0: Pre-school attendance and Pre-algebra achievement at statistically independent.
  - H_1: Pre-school attendance and Pre-algebra achievement at statistically dependent.
  - Alpha= .05

Example 1

- Step 2: Calculate Chi-square

Observed	**Below Grade Level**	**At Grade Level**	**Advanced**	**Total**
Pre-School	8	6	6	20
No Pre-School	6	15	9	30
Total	14	21	15	50

	Fo	**Fe**	**Fo-Fe**	**(Fo-Fe)2**	**(Fo-Fe)2/Fe**
Pre-School/ Below	8				
Pre-School/ At	6				
Pre-School/ Advanced	6				
No Pre-School/ Below	6				
No Pre-School/ At	15				
No Pre-School/ Advanced	9				
				Chi-Square	

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

f_e = expected frequencies
f_o = observed frequencies

Example 1

- Step 2: Calculate Chi-square

Expected	Below Grade Level	At Grade Level	Advanced	Total
Pre-School	5.6	8.4	6	20
No Pre-School	8.4	12.6	9	30
Total	14	21	15	50

	Fo	Fe	Fo-Fe	(Fo-Fe)2	(Fo-Fe)2/Fe
Pre-School/ Below	8	5.6			
Pre-School/ At	6	8.4			
Pre-School/ Advanced	6	6			
No Pre-School/ Below	6	8.4			
No Pre-School/ At	15	12.6			
No Pre-School/ Advanced	9	9			
				Chi-Square	

$$F_e = \frac{\text{(Column Marginal)(Row Marginal)}}{N}$$

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

f_e = expected frequencies
f_o = observed frequencies

Example 1

- Step 2: Calculate Chi-square

	Fo	Fe	Fo-Fe	(Fo-Fe)2	(Fo-Fe)2/Fe
Pre-School/ Below	8	5.6	2.4	5.76	1.03
Pre-School/ At	6	8.4	-2.4	5.76	0.69
Pre-School/ Advanced	6	6	0	0	0.00
No Pre-School/ Below	6	8.4	-2.4	5.76	0.69
No Pre-School/ At	15	12.6	2.4	5.76	0.46
No Pre-School/ Advanced	9	9	0	0	0.00
				Chi-Square	**2.86**

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

f_e = expected frequencies
f_o = observed frequencies

Example 1

- Step 3: Calculate DF and Find Critical Value
  - DF=(r-1)*(c-1)=(2-1)*(3-1)=1*2=2
  - Critical Value=5.991
- Step 4: Compare Test Statistic
  - $X^2=2.86<5.991$, Insignificant
- Step 5: Conclusions
  - Therefore, pre-school attendance is *independent* of achievement in pre-algebra

Example 2

- 2. Gender and Suicide
(these are contrived data, based on a real study)
- Below is a 2x2 cross-tabulation that shows the relationship between gender and whether a suicide attempt was successfully completed. Test the null hypotheses that the relative frequency of women who successfully commit suicide is the same as the relative frequency of men who successfully commit suicide. What do your results indicate?

	Complete	Failed	Total
Male	22	7	29
Female	6	19	25
Total	28	26	54

Example 2 Answers

- Null and Research Hypotheses (Alpha: .05)
  - H_0: Suicide completion and gender are statistically independent.
  - H_1: Suicide completion and gender are statistically dependent.

	Fo	Fe	Fo-Fe	(Fo-Fe)2	(Fo-Fe)2/Fe
Male/Complete	22	15.04	6.96	48.48	3.22
Male/Failed	7	13.96	-6.96	48.48	3.47
Female/Complete	6	12.96	-6.96	48.48	3.74
Female/Failed	19	12.04	6.96	48.48	4.03
				Chi-Square	**14.46**

- DF=1, CV=3.841<14.46
- Therefore, suicide completion and gender are statistically dependent.