

# Multiple Sample Chi Square

# This is useful for what situation?

- Basically, non-parametric. (No assumption about normal distribution)
- No need for central tendency measures like mean.
- Unlike t-test used for continuous data, this chi square test could be used for categorical data.
- If you are interested in the comparison of mid-term scores between female and men, then...
- If you are interested in the comparison of political attitude to Iraq war between female and men, then...

# Statistic Chi-square

- Here is the equation.

$$\chi^2 = \sum \frac{(f_{observed} - f_{expected})^2}{f_{expected}}$$

- Null hypothesis: There is no relationship between two variables. (In other words, independent in statistics) Even if there is, this is just because of sampling errors.
- Research hypothesis: not H0.

# How to get the statistic chi?

- $F_o$  is the frequency observed in each category of the dependent variable (what you are trying to explain)
- $F_e$  is the frequency expected in each category for the dependent if null hypothesis is true (that is, the sample and population is distributed equally)
- Sigma: just sum all of elements!
- Degree of freedom for chi distribution:  $(k-1)(n-1)$ , where  $k$  and  $n$  is the number of categories for each variable.
- Go to the table for chi square distribution

# Simple Example

- This is what you observed.
- This is called perfect association. (maximum splitting!) Note that the frequency for two off-diagonal cells is 0.

	Republican	Democrat
Agree on Iraq	10	0
Disagree on Iraq	0	10

# Simple Example

- This is what is expected when each frequency is randomly distributed. In other words, if your null hypothesis is true, there should be no association between party and attitude to Iraq attack.

	Republican	Democrat
Agree on Iraq	5	5
Disagree on Iraq	5	5

# Simple Example

- Let's calculate the statistic chi-square!

$$\chi^2 = \sum \frac{(f_{observed} - f_{expected})^2}{f_{expected}}$$
$$= \frac{(10 - 5)^2}{5} + \frac{(0 - 5)^2}{5} + \frac{(0 - 5)^2}{5} + \frac{(10 - 5)^2}{5}$$

- Statistic Chi-square is 20. What is DF?  $(2-1)(2-1)=1$ .
- Go to the table for this distribution in your textbook!

- Let's suppose that your confidence level is 95%. (Alpha is 0.05)
- Critical value is 3.841 at this level.
- So, what's your conclusion? You have to reject  $H_0$ . Two variables are not independent, which means that if you know one variable, say "party", you can guess another variable, say "attitude..." with some errors.

d.f.	Alpha level	
	0.05	0.01
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475
8	15.507	20.090
9	16.919	21.666
10	18.307	23.209

# What happened to Titanic?

- It has been said about the alleged discrimination against women and the third class passengers in the process of rescue.
- Your research hypothesis is "Men, first."
- Your null hypothesis is there is no difference 1) between men and women in terms of their survival (or death), which means that if you know one variable, "sex", you cannot estimate another variable, "survival" because there is no relationship between two variables!

# Men, first? Really? (N=1316)

Observed	Female	Male	Total
Death	123	694	817
Survival	324	175	499
Total	447	869	1316

Expected	Female	Male	Total
Death	277.5	539.5	817
Survival	169.5	329.5	499
Total	447	869	1316

- How to calculate the statistic chi-square? (343.53)

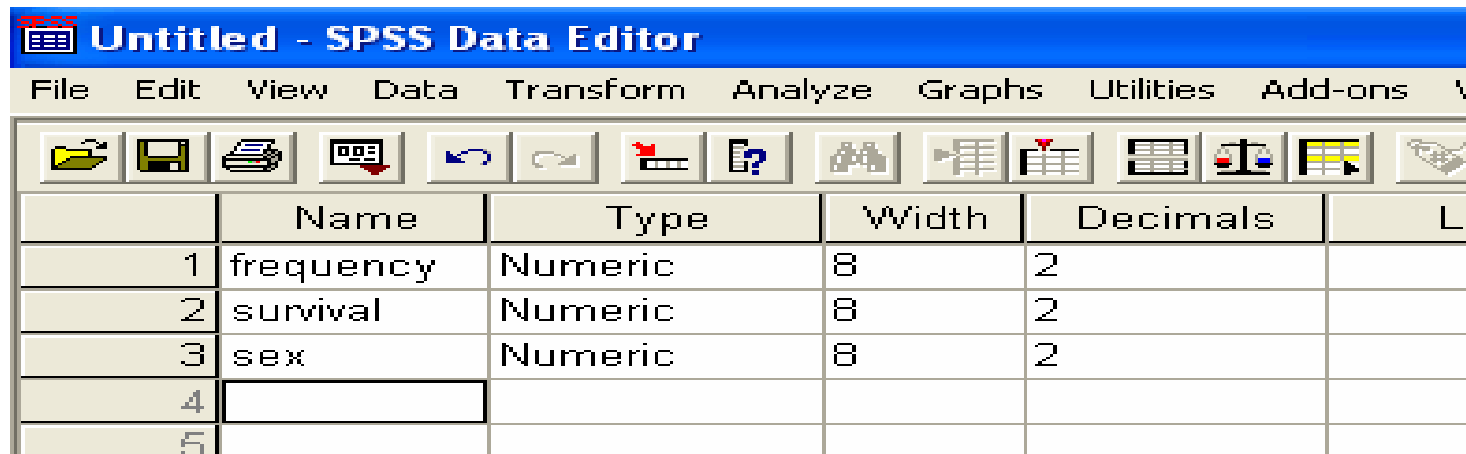
$$\chi^2 = \sum \frac{(f_{observed} - f_{expected})^2}{f_{expected}}$$

$$= \frac{(123 - 277.5)^2}{277.5} + \frac{(694 - 539.5)^2}{539.5} + \frac{(324 - 169.5)^2}{169.5} + \frac{(175 - 329.5)^2}{329.5}$$

- ☺ You might be curious about how I got each expected score. I considered probability based on random distribution, but you might as well use SPSS as follows. (Recommended)
- Critical value ( $\alpha=0.05$ ) is 3.841 again. The statistic is far away from 3.841. What's your conclusion?

# How to use SPSS? (Given frequency!)

- First, go to the “variable view” in SPSS data editor.
- There need to be trick because your data is about frequency for two variables. (Recall that column is about variables, while row is about cases usually)
- Just put “frequency”, “survival”, “sex” in the first column.

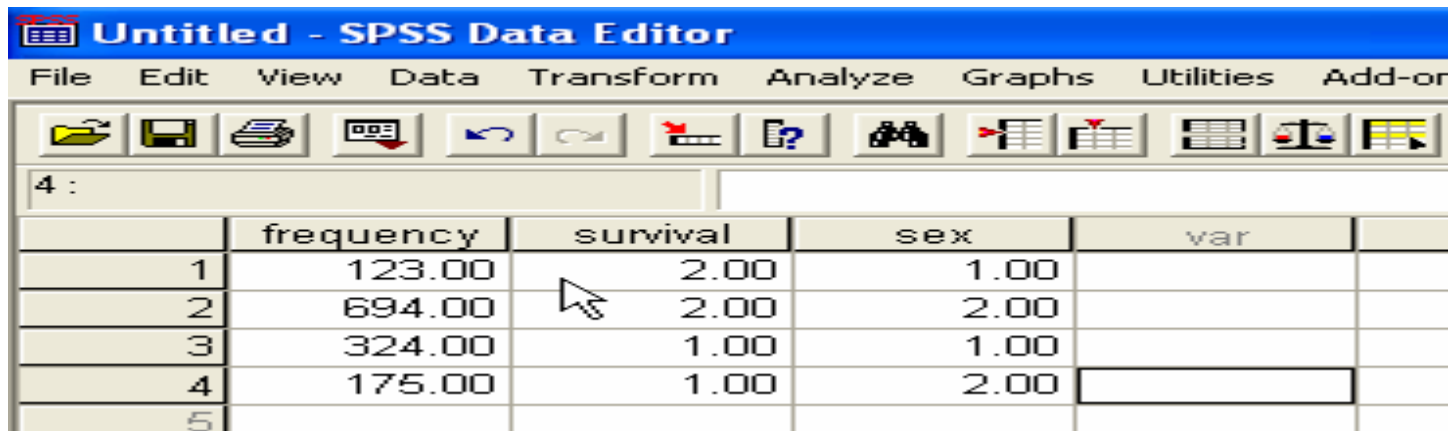


The screenshot shows the SPSS Data Editor interface in Variable View. The title bar reads "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, and Windows. The toolbar contains various icons for file operations and data manipulation. The main area is a table with columns for Name, Type, Width, Decimals, and L. The first three rows are filled with "frequency", "survival", and "sex", all set to Numeric type with a width of 8 and 2 decimal places. The fourth and fifth rows are empty.

	Name	Type	Width	Decimals	L
1	frequency	Numeric	8	2	
2	survival	Numeric	8	2	
3	sex	Numeric	8	2	
4					
5					

# Go to “data view” and input data

- Input your data in the following way.
- If survival, give 1; otherwise, 2.
- If female, give 1; otherwise, 2.
- E.g. the first row means that the frequency is 123 for female(1) and no survival(2)



The screenshot shows the SPSS Data Editor interface. The title bar reads "Untitled - SPSS Data Editor". The menu bar includes "File", "Edit", "View", "Data", "Transform", "Analyze", "Graphs", "Utilities", and "Add-on". The toolbar contains various icons for file operations and data manipulation. The data grid shows the following data:

	frequency	survival	sex	var
1	123.00	2.00	1.00	
2	694.00	2.00	2.00	
3	324.00	1.00	1.00	
4	175.00	1.00	2.00	
5				

# Go to Data > Weight Cases

- Because your data is about frequency for each variable, you have to require SPSS to regard your data as frequency.
- Move "frequency" to another small window. Click Ok.

The screenshot shows the SPSS Data Editor window with a data table and the 'Weight Cases' dialog box open. The data table has the following content:

	frequency	survival	sex	var	var	var	var
1	123.00	2.00	1.00				
2	694.00	2.00	2.00				
3	324.00	1.00	1.00				
4	175.00	1.00	2.00				

The 'Weight Cases' dialog box is open, showing the following options:

- Do not weight cases
- Weight cases by

Frequency Variable:

Current Status: Do not weight cases

Buttons: OK, Paste, Reset, Cancel, Help

# Go to Analyze > Descriptive > Crosstabs...

- Move two variables, survival and sex, to each small window.

The screenshot shows the SPSS Data Editor interface. The main window displays a data table with the following data:

	frequency	survival	sex	var	var	var	var
1	123.00	2.00	1.00				
2	694.00	2.00	2.00				
3	324.00	1.00	1.00				
4	175.00	1.00	2.00				
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

The Crosstabs dialog box is open, showing the following configuration:

- Row(s): survival
- Column(s): sex
- Layer 1 of 1: Previous, Next
- Display clustered bar charts:
- Suppress tables:
- Buttons: Exact..., Statistics..., Cells..., Format..., OK, Paste, Reset, Cancel, Help

# Click statistics tab and check chi-square

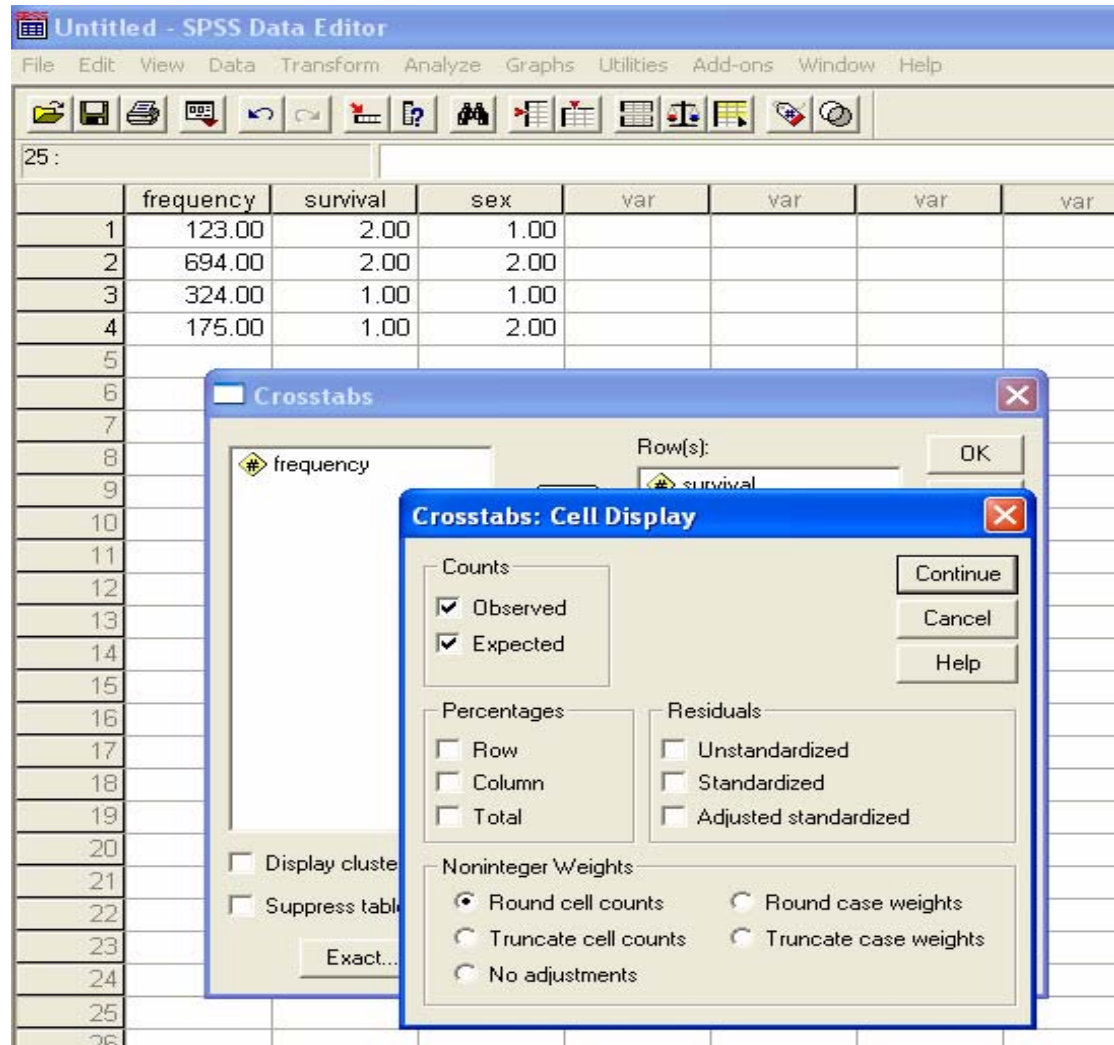
- On the bottom, the second tab is about statistics.
- Check chi-square, click continue, and get out of there.

The screenshot shows the SPSS Data Editor interface. The main window displays a data grid with columns labeled 'frequency', 'survival', 'sex', and 'var'. The 'frequency' column contains values 123.00, 694.00, 324.00, and 175.00 for rows 1 through 4. The 'survival' column contains values 2.00, 2.00, 1.00, and 1.00. The 'sex' column contains values 1.00, 2.00, 1.00, and 2.00. A 'Crosstabs' dialog box is open in the foreground, with the 'Statistics' sub-dialog box also open. In the 'Crosstabs: Statistics' dialog, the 'Chi-square' checkbox is checked. Other options like 'Nominal', 'Ordinal', 'Nominal by Interval', and 'Cochran's and Mantel-Haenszel statistics' are visible but not checked. The 'Continue' button is highlighted.

	frequency	survival	sex	var	var	var	var
1	123.00	2.00	1.00				
2	694.00	2.00	2.00				
3	324.00	1.00	1.00				
4	175.00	1.00	2.00				

# Click statistics tab and check chi-square

- On the bottom, the third tab is about cells.
- Check “expected”, click continue, and get out of there.
- Finally, click OK on the right.



Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help



Output

- Crosstabs
  - Title
  - Notes
  - Case Proc
  - survival \*
  - Chi-Squar

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
survival * sex	1316	100.0%	0	.0%	1316	100.0%

**survival \* sex Crosstabulation**

			sex		Total
			1.00	2.00	
survival	1.00	Count	324	175	499
		Expected Count	169.5	329.5	499.0
	2.00	Count	123	694	817
		Expected Count	277.5	539.5	817.0
Total		Count	447	869	1316
		Expected Count	447.0	869.0	1316.0

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	343.568 <sup>b</sup>	1	.000		
Continuity Correction <sup>a</sup>	341.348	1	.000		
Likelihood Ratio	347.771	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	343.307	1	.000		
N of Valid Cases	1316				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 169.

# Interpretation

- Almost the same as the calculation by hand.
- According to SPSS output, chi-square 343.568 (p=0.000) Df=1=(2-1)(2-1).
- What's your conclusion?

# One-way and Two-way

- $X \rightarrow Y$  or  $Y \rightarrow X$ : this is one-way relationship. In other words, causal relationship where one is independent variable and the other is dependent variable.
- $X \leftrightarrow Y$ : this is two-way relationship (double arrow) Both variables are called covariates.
- These things have nothing to do with symmetric or asymmetric measure!

# Asymmetric Measure (e.g. Lambda)

	X		total
Y	7	4	11
	3	8	11
total	10	12	22

- When x is located in independent variable. (column)
- E1 is 11.
- E2 is  $3+4=7$ .
- PRE is  $(11-7)/11=0.364$ .

# Continued

	Y		total
X	7	3	10
	4	8	12
total	11	11	22

- When Y is located in independent variable. (column)
- E1 is 10.
- E2 is  $3+4=7$ .
- $PRE=(10-7)/10=0.3$
- In this way, the two PRE are not the same, which means Lambda is asymmetric measure!

# Symmetric Measure (e.g. Chi-square)

Obse -rved	X		total
	Y	7	
	3	8	11
total	10	12	22

Expec- ted	X		total
	Y	5= (10*11) /22	
	5= (10*11) /22	6= (12*11) /22	11
total	10	12	22

# Continued

- Chi-square= $(7-5)^2/5+(4-6)^2/6+(3-5)^2/5+(8-6)^2/6=2.93$ .
- This result comes when X is located in independent variable (column.)

# Continued...

Observed	Y		total
X	7	3	10
	4	8	12
total	11	11	22

Expected	Y		total
X	5= $(11 \cdot 10)$ /22	5= $(11 \cdot 10)$ /22	10
	6= $(11 \cdot 12)$ /22	6= $(11 \cdot 12)$ /22	12
total	11	11	22

# Continued

- Chi-square= $(7-5)^2/5+(3-5)^2/5+(4-6)^2/6+(8-6)^2/6=2.93$ .
- This result comes when Y is located in independent variable (column.)
- The two results are the same, which means chi-square is symmetric.