Roadmap to Data Analysis

V. Comparing Counts – Chi-Square Test of Association

Learning Objectives

- Understand the type of data that are relevant for a chi-square analysis
- Understand the basics of how chi-square works
- Understand how to use a chi-square calculator with example data
- Understand how to interpret results

Type of Data for Chi-Square Analysis

- Categorical variables
- Counts of people, services, things
  - "Frequency" – count of the number of cases in a particular *category* of a variable
  - "Frequency distribution" – counts or percentages of the number of cases in each category of the variable
    - The number of women and men in a treatment group
    - The number of services used by clients in each ethnicity group
    - The percent of an agency's clients who are employed full time, employed part time, or not employed
    - The number of survey respondents who are "Dissatisfied", "Satisfied," or who have no opinion.

How does Chi-Square work?

Based on a "cross-tabulation table" – showing the frequency distributions of two categorical variables. Here's an example: To monitor outreach activities, you want to know if there is any relationship between gender and engagement in treatment.

Gender	In treatment		Not in treatment		Total
	N	%	N	%	N (%)
Male	20	29%	50	71%	70 (100%)
Female	40	36%	70	64%	110 (100%)
Total	60	33%	120	67%	180 (100%)

“Observed” vs. “Expected”

The chi-square statistic quantifies the relationship between the *observed* values (shown in the previous slide) compared to the *expected* values – what you would expect to see if there was no relationship between gender and engagement treatment.

Gender	In treatment (expected values)		Not in treatment (expected values)		Total
	N	%	N	%	N (%)
Male	35	50%	35	50%	70 (100%)
Female	55	50%	55	50%	110 (100%)
Total	90	50%	90	50%	180 (100%)

Using a Chi-Square Calculator

- Go to http://graphpad.com/quickcalcs/contingency1/
- Or, via the HealTorture.org website:

http://www.healtorture.org/content/basic-statistical-methods

  - Use the contingency table calculator
- Enter the observed counts or observed percentages in the empty cells. Also, enter your own column and row headings (see example, next slide)
- Choose "Chi-square with Yates correction"*
- Click "Calculate"

*For smaller samples, i.e. when at least one cell has fewer than 5 observations, choose "Fisher's exact test" for a more accurate analysis

Enter data (observed counts)

	In treatment	Not in treatment
Male	20	50
Female	40	70

Results—as shown in Calculator

	In treatment	Not in treatment	Total
Male	20	50	**70**
Female	40	70	**110**
Total	**60**	**120**	**180**

Chi-square with Yates correction

Chi squared equals 0.844 with 1 degrees of freedom.

The two-tailed P value equals 0.3581

The association between rows (groups) and columns (outcomes) is considered to be not statistically significant.

Interpret results

- Since the *p* value is greater than .05 (*p* = 0.36), you can conclude that there is no evidence of a relationship between gender and engagement in treatment (in other words, no statistically significant difference between the observed and expected values)
- To report chi-square results formally:

"*In our agency sample, there was no relationship between gender and engagement in treatment, $\chi^2 = 0.84$, df = 1, p = 0.36."*

- The symbol for chi-square is χ^2 (as chosen by the statistician Karl Pearson – hence the "Pearson chi-square" statistic)
- The "degrees of freedom" (df) has to do with how the chi-square statistic is distributed theoretically, based on the number of cells in the cross-tabulation table. It is required to calculate the *p* value and typically reported.