

Analyzing Categorical (Nominal) Data

Contingency Tables (Crosstabs)

- generated from 2 (sometimes 3) nominal level variables
- show the distribution of responses as counts in table "cells"
- observations must be independent
- found in SPSS under Analyze → Descriptives → Crosstabs
- Crosstab tables display observed counts, expected counts, and percentages

Sample SPSS Crosstab (2x4) Table

developed parkland * gender Crosstabulation

			gender		Total
			male	female	
developed parkland	decrease amt devel parkland/res	Count	7	6	13
		Expected Count	6.9	6.1	13.0
		% within gender	5.3%	5.2%	5.2%
maintain current amt dev pkind/res		Count	81	70	151
		Expected Count	80.7	70.3	151.0
		% within gender	60.9%	60.3%	60.6%
increase amt dev prkind/res		Count	40	34	74
		Expected Count	39.5	34.5	74.0
		% within gender	30.1%	29.3%	29.7%
no opinion		Count	5	6	11
		Expected Count	5.9	5.1	11.0
		% within gender	3.8%	5.2%	4.4%
Total		Count	133	116	249
		Expected Count	133.0	116.0	249.0
		% within gender	100.0%	100.0%	100.0%

Calculation of Expected Counts

Expected value for a cell = $\frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$

developed parkland * gender Crosstabulation

			gender		Total
			male	female	
developed parkland	decrease amt devel parkland/res	Count	7	6	13
		Expected Count	6.9	6.1	13.0
		% within gender	5.3%	5.2%	5.2%
maintain current amt dev pkind/res		Count	81	70	151
		Expected Count	80.7	70.3	151.0
		% within gender	60.9%	60.3%	60.6%
increase amt dev prkind/res		Count	40	34	74
		Expected Count	39.5	34.5	74.0
		% within gender	30.1%	29.3%	29.7%
no opinion		Count	5	6	11
		Expected Count	5.9	5.1	11.0
		% within gender	3.8%	5.2%	4.4%
Total		Count	133	116	249
		Expected Count	133.0	116.0	249.0
		% within gender	100.0%	100.0%	100.0%

Example
 $(13 \times 133) / 249 = 6.9 \text{ expected}$

Pearson Chi-square Statistic

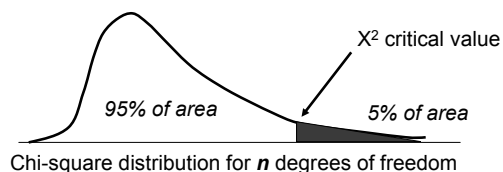
- tests whether there is an association between two categorical variables
- sums the differences between observed and expected cell counts:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

- found in SPSS under:
Analyze-->Descriptive Stats-->Crosstabs
[then click on Statistics box]

Statistical Significance with Chi-square

- Family of Chi-square distributions, each based on degrees of freedom
- Each distribution has a critical X^2 value at the .05 level of significance



Interpreting SPSS Output

developed parkland * gender Crosstabulation

		gender		Total
		male	female	
developed parkland	decrease amt devel parkland/res	Count 7	6	13
		Expected Count 6.9	6.1	13.0
		% within gender 5.3%	5.2%	5.2%
maintain current amt dev pkland/res	Count 81	70	151	
	Expected Count 80.7	70.3	151.0	
	% within gender 60.9%	60.3%	60.6%	
increase amt dev pkland/res	Count 40	34	74	
	Expected Count 39.5	34.5	74.0	
	% within gender 30.1%	29.3%	29.7%	
no opinion	Count 5	6	11	
	Expected Count 5.9	5.1	11.0	
	% within gender 3.8%	5.2%	4.4%	
Total	Count 133	116	249	
	Expected Count 133.0	116.0	249.0	
	% within gender 100.0%	100.0%	100.0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.296 ^a	3	.961
Likelihood Ratio	.296	3	.961
Linear-by-Linear Association	.069	1	.793
N of Valid Cases	249		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.12.

Chi-square statistic

Significance level (p-value)

Check on whether assumptions violated

Interpreting Chi-square Results

- p-values less than .05 indicate statistical significance
- sig. results indicate an “association” between the categorical variables (variables do not appear independent)
- check that Chi-square assumptions are not violated—all expected cell counts should be greater than 5 (2x2). In larger contingency tables, acceptable to have up to 20% of cells with expected frequencies below 5 but *in no case should expected frequencies be below one.*

Strength of Association Measures (Phi and Cramers V)

- Calculated statistics range from 0 to 1.
- Higher calculated values for statistics indicate stronger associations
- Zero indicates virtually no association. Values of 1 indicates near perfect association.
- *Phi statistic*. Use for 2 x 2 tables. Not accurate for larger tables.
- *Cramers V*. Use for greater than 2 x 2 tables.