R-by-*C* Tables

Background
Hypothesis Test

Background

This chapter considers an extension of the chi-square test introduced in the prior chapter. We now consider *R-by-C* cross-tabulations, where *R* represents the number of rows in the table and *C* represents the number of columns. For example, we may wish to consider a 4-by-2 table in which cases status and alcohol are cross-tabulated as follows:

Grams of Alcohol Consumed / Day	**Case Status** Esoph. Cancer	Control	Total
0 - 39	29	386	415
40 - 79	75	280	355
80 - 119	51	87	138
120+	45	22	67
Total	200	775	975

Notice that this 4-by-2 table could just as easily have been set up as a 2-by-4 table, having case status represented along rows and alcohol level represented along columns. However, this would *not* materially affect conclusions to follow.

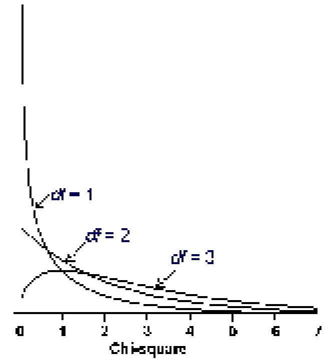
SPSS: To cross-tabulate data, click Analyze | Descriptive Statistics | Crosstabs.

A useful description comes from calculating proportions *within* groups. This proportions should be calculated in a way that makes sense. For example, the distribution of alcohol consumption in cases and controls for the illustrative data is:

Alcohol / day (gms)	**Esophageal Cancer** Yes	No	Total
0 - 39	14.5%	49.8%	43%
40 - 79	37.5%	36.1%	36%
80 - 119	25.5%	11.2%	14%
120+	22.5%	2.8%	7%
	100%	100%	100%

Notice the high percentage of cases that fall into the high alcohol consumption categories.

Hypothesis Test



We wish to test the hypotheses H_0: no association between the row and column variable vs. H_1: association between the row and column variables. A chi-square method, as discussed in the previous chapter, is used to perform the test. Chi-square distributions with 1, 2, and 3 degrees of freedom are illustrated in the figure to the right:

Expected frequencies under the null hypothesis (E_i) are calculated :

$$E_i = \frac{\text{row total} \times \text{column total}}{\text{total sample size}}$$

For the illustrative data example, expected frequencies are:

	CASE		
ALC	1	2	Total
0 - 39	(415*200) / 975 = 85.13	(415*775) / 975 = 329.87	415
40 - 79	(335*200) / 975 = 72.82	(355*775) / 975 = 282.18	355
80 - 119	(138*200) / 975 = 28.31	(138*775) / 975 = 109.69	138
120+	(67*200) / 975 = 13.74	(67*775) / 975 = 53.26	67
Total	200	775	975

The chi-square method should be used only when expected values exceed 5 in each cell.

To calculate the chi-square test statistic, we subtract expected counts (E_i) from the observed counts (O_i), square these difference, and divide by the expected counts in each table cell:

$$c^2_{\text{stat}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

For the illustrative example, χ^2_{stat} = [(29 - 85.1)^2 / 85.1 + (386 - 329.9)^2 / 329.9 + (75 - 72.82)^2 / 72.8 + (280 - 282.8)^2 / 282.8 + (51 - 28.3)^2 / 28.3 + (87-109.7)^2 / 109.7 + (45 - 13.7)^2/13.7 + (22 - 53.3)^2/ 53.3] = 36.98 + 9.54 + 0.07 + 0.02 + 18.21 + 4.70 + 71.51 + 18.38 = 159.41.

The test statistic has (R -1)(C -1) degrees of freedom, where R represents the number of rows in the table and C represents the number of columns. For the illustrative example, df = (4-1)(2-1) = 3. The p-value is determined as the area under the curve beyond the test statistic. In the case of the illustrative example, $p < .01$.