Contingency-Table Method

Same Example as last class: Suppose we sample 1000 students in the SPHTM and 400 are male. Suppose we sample 600 students in the SOM and 300 are male. Use a two-tailed test to test if the proportions of students who are male are different in the two schools. Conduct the test at the 5% level of significance.

	Male	Female	Total
SPHTM	40	60	100
SOM	30	30	60
Total	70	90	160

Each population is a row in the table. Each gender is a column. Each person falls into one of the cells.

Note: Number of students at SPHTM is a row total called a row marginal.

Note: There are a total of 70 men. This is called a column marginal.

Note: There are a total of 160 people in the study. This is called the grand total.

The above table is a 2 x 2 contingency table. There are two categories each at two levels. It is arbitrary which variable is the row variable and which is the column variable. Each cell is the number of subjects with that particular value for each variable.

Note: These are the Observed values. These are sometimes labeled O_{11}, O_{12}, O_{21}, O_{22} where O_{ij} stands for the i^{th} row and the j^{th} column.

For this example, we will be conducting a test for homogeneity of two binomial proportions. Is the proportion of Males the same for each population?

Expected Table—We need to calculate the number of subjects in each cell which we would expect if there was no relationship between gender and school; i.e., $p_1 = p_2 = p$.

Where $p_1 = \Pr(M \mid SPHTM)$
and $p_2 = \Pr(M \mid SOM)$

To generalize:

Let n_1 = sample size of 1st sample
n_2 = sample size of 2nd sample
X_1 = number exposed (with characteristic) in the 1st sample
X_2 = number exposed (with characteristic) in the 2nd sample

General Contingency Table

	Variable 1		
Population	**1**	**2**	**Total**
1	X_1	$n_1 - X_1$	n_1
2	X_2	$n_2 - X_2$	n_2
Total	$X_1 + X_2$	$n_1 + n_2 - (X_1 + X_2)$	$n_1 + n_2$

If H_0 is true, best estimate of p is

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} \text{ or } \frac{X_1 + X_2}{n_1 + n_2}$$

Under H_0, for (1, 1) cell we expect

$n_1\hat{p} = \frac{n_1(X_1 + X_2)}{n_1 + n_2}$ **in that cell**

This is the row total * column total / grand total
Similarly, in (2, 1) cell we expect