R x C Contingency Tables

Tables with R rows and C columns that display the relationship between two variables. Row variable has R categories. Column variable has C categories.

Tables can be used if we want to conduct a test of homogeneity; i.e., we want to compare the proportion with some characteristic in 3 or more populations. An example is the proportion of obese children in Louisiana vs. Mississippi vs. Alabama vs. Arkansas.

They can also be used if we want to conduct a test of independence; i.e., we want to see if the distribution of one variable is the same no matter what the distribution of the other variable in a single population. An example of the latter is hair color and gender.

	Variable 1					
Variable 2	**1**	**2**	**3**	**…**	***c***	**Total**
1	n_{11}	n_{12}			n_{1c}	$n_{1.}$
2	n_{21}	n_{22}			n_{2c}	$n_{2.}$
3						
…						
r	n_{r1}	n_{r2}			n_{rc}	
Total	$n_{.1}$	$n_{.2}$				$n_{..}$

The above are the observed frequencies and can be labeled O_{11}, O_{12}, …, O_{1c}, O_{21}, O_{22}, …, O_{2c}, …, O_{rc}.

Homogeneity:

H_0: The proportion with some characteristic is the same in each population
H_1: The proportion with some characteristic is not the same in each population

Independence:

H_0: In a population, the 2 variables are independent
H_1: In a population the 2 variables are not independent

Generalizing from the 2 X 2 situation, the expected table can be found in a similar way. The expected number of units in the (i, j) cell = E_{ij} = (the total in the i^{th} row) * (the total in the j^{th} column) / the grand total.

The sum of the expected values for any row or column must equal the sum for the corresponding row or column in the observed table.

Test Statistic to Compare Observed with Expected

$$X^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)*(c-1)}$$

H_0 will be rejected for large values of X^2; that is, if the observed and expected counts are very dissimilar, the more likely we are to reject H_0.