Inference for Two-Way Tables

- We can also use the Chi-Square method to make inferences for data in two-way tables.
- The formula to find the expected count in a two-way table is:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

- Where n is the grand total of all values.

- When conducting a Chi-square test of independence in a two-way table, the null and alternate hypothesis will be:
  - H_0: There is no association between the row and column variables.
  - H_a: There is an association between the two variables.
- The test statistic is: $\chi^2 = \sum_{allcells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
- The p-value: $P(\chi^2_k > \chi^2)$, where χ^2_k represents a Chi-square distribution with $df = (r - 1)(c - 1)$ degrees of freedom.

- The assumptions necessary for the test to be valid are:
  1. The observations constitutes a simple random sample from the population of interest, and
  2. The expected counts are at least 5 for each cell of the table.
- By itself, the chi-square test determines only whether the data provide evidence of a relationship between the two variables.
- If the result is significant, one can go on to identify the source of that relationship by finding the cells of the table that contribute most to the value (i.e. those cells with the biggest discrepancy between the observed and expected counts) and by noting whether the observed count falls above or below the observed count in those cells.

Example:

Num of rows: $r = 2$
num of cols: $c = 3$
$df = (1)(2) = 2$

Use the data below to determine if there is sufficient evidence to conclude that an association exists between car color and the likelihood of being in an accident.

	Red	Blue	White	Row tot
Obs: Car has been in accident	28	33	36	97
Obs: Car has not been in accident	23	22	30	75
	51	55	66	172 (n)

Exp: Car has been in acc. $97 \times 51/172$ $97 \times 55/172$ $97 \times 66/172$

Exp: Car has not been in acc $75 \times 51/172$ $75 \times 55/172$ $75 \times 66/172$

$$\chi^2 = \frac{(28-28.76)^2}{28.76} + \frac{(33-31.02)^2}{31.02} + \frac{(36-37.22)^2}{37.22} + \frac{(23-22.2)^2}{22.2} + \frac{(22-23.98)^2}{23.98} + \frac{(30-28.78)^2}{28.78}$$

$\chi^2 = .43$

pvalue: $p(\chi^2 > .43) = \chi^2cdf(.43, 10\text{^}6, 2) = .8065 > .05$

Fail to reject H_0 no association