Chapter 10
Association between Two Categorical Variables
Contingency Tables and c^2 (Chi-Square) Tests

What we have seen so far:

- In Chapters 3 and 11 we searched for association between two quantitative variables.

- In Chapter 12 we added one or more categorical variables to the quantitative predictors.

- In Chapter 13 we had a quantitative response and one or more categorical predictor variables each with 2 or more categories.

- This was an extension of what we saw in Chapter 9, where we had a quantitative response and a categorical predictor that had 2 categories (populations) [with dependent or independent samples].

- In Chapters 8 and 9 we looked at the difference between proportions (i.e., response is a categorical variable with 2 categories) of two populations (categorical predictor with 2 categories) with random samples (dependent or independent) from these populations.

We now extend this to the case of a **categorical predictor** with 2 or more categories and **a categorical response** with 2 or more categories, where data from a random sample are summarized in an **r □ c contingency table**.

Example: Last semester after the week-end when Gator Basketball team won the game that put them in the Final Four (which ended at 11:30 p.m.), 101 students in a Statistics class were asked to report their gender and whether or not have watched the whole game, part of it or not at all. The following table summarizes the responses:

Watched?	Gender		Total
	Male	Female	
Whole game	10	21	31
Part of Game	12	24	36
None	4	30	34
Total	26	75	101

To compare the differences in how much each gender watched the game, we need to find percentages in each category; but first we have to decide which variable is the response and which one is the predictor, so that we can decide what to put in the denominator of these proportions.

In this example,

- The **response** is how much each student **watched** the game and
- The **predictor** is gender.
- To compare the two genders we will divide the numbers in each "cell" of the above table by the total number of students of each gender, i.e., **divide the number of observations in each cell by the total in each predictor (gender) category**
- Such a division will give how much of the game watched by gender, i.e., the **conditional distribution of response:**

Conditional Distribution of Response

Watched?	Gender		Total
	Male	Female	
Whole game	38.5% (10/26)	28.0% (21/75)	30.7% (31/101)
Part of Game	46.2% (12/26)	32.0% (24/75)	35.6% (36/101)
None	15.4% (4/26)	40.0% (30/75)	33.7% (34/101)
Total	100.0% (26/26)	100.0% (75/75)	100.0% (101/101)

- In the above table, we see that male students watched more of the game than the females.