

**Worth:** 30 points

**Project Due:** **Wednesday, November 15, 2006 at 5pm.** Turn it in during class, or in my mailbox in Griffin Floyd 103. The mailbox is the one UNDER the label Ripol. **Late Projects** will be accepted until Friday, November 17, in class, but will result in a 10-point penalty.

**Purpose:** To study the relationship between two categorical variables for UF students and learn how to analyze Contingency Tables using Minitab.

### 1. Choose your Topic.

Everyone should pick a topic that interests them – different topics for each student. You need to define your questions so that you have two categorical variables. One will be your explanatory variable that defines the groups you want to compare. The other one will be your response variable, which is the one you want to draw conclusions about. Here are some examples, where I list the explanatory variable first:

Study the association between:

- Gender and Drinking Habits
- Smoking and Exercise Habits
- “Greek” Affiliation and Party Habits
- Race and Level of Satisfaction with UF

**NOTE – You need to define your OWN questions.** These are just examples to get you started thinking of some ideas. In fact NO ONE should use any of those examples I listed above, or the same as their friends.

### 2. Collect the data

You should collect data from enough UF students so that all of the observed counts are 5 or more. Explain to the participants what the data will be used for. The data should be treated confidentially. If the subject is sensitive, you should prepare slips of paper with the two questions, ask the students to complete them without anyone watching, and place them in an envelope or box, where they will remain anonymous. You may collect data on yourself, your friends, classmates, or students that you don't know – this is not a random sample, but it would be very difficult to get one.

**3. Get ready to use Minitab** – you can buy it, rent it, download it for free (30 days) or use it in the CIRCA Labs. See course website for more details. I have requested a reservation for one of the CIRCA lab classrooms for **Tuesday, 11/7 from 10:40 to 1:40. Please check course website for confirmation.** Tezcan Ozrazgat, the TA for our class, will be there to help people who need it.

**4. Perform the statistical analysis –**

**Instructions for Minitab Version 14**

- a. Summarize the results of your data in a Two Way Table of counts.
- b. **Enter your counts on the worksheet**, using the necessary rows and columns. **Label the columns** (the row labels do not appear in the analysis, but the column labels do).
- c. On the top menu, go to **Stat – Tables –  $\chi^2$  Chi-Square Test (Table in Worksheet)**.
- d. Select **whichever columns contain your data**, then click OK.
- e. The analysis appears on the Session window.

**5. Prepare your report using a word processor. The report should be ONE page long, and include the following (see example on last page):**

- a. **Your NAME and UFID number, and the course information.**
- b. **The PURPOSE of your study – what are you trying to find out? (5 pts)**
- c. **A TABLE with the DATA collected. (5 pts)**
- d. **The statistical analysis performed by Minitab (4 pts)**– with the left mouse button, select the analysis from the Session window, then use the “copy” function in Minitab, and paste it into your document.
- e. **All the steps of the statistical analysis: (8 pts)**
  - null and alternative hypotheses
  - test statistic
  - the p-value associated with it.
- f. **Your conclusions**, in terms of the variables you are studying, including an explanation in plain English. **(4 pts)**
- g. **A statement about any problems with the assumptions – sample size and randomness of the data (4 pts)**

**Purpose:** To study the association between Gender of UF students, and the Presidential Candidate they Voted for in the 2004 Presidential Election.

**Data:**

	Bush	Kerry
Females	7	13
Males	9	7

**Chi-Square Test: Bush, Kerry**

Expected counts are printed below observed counts  
Chi-Square contributions are printed below expected counts

	Bush	Kerry	Total
1	7	13	20
	8.89	11.11	
	0.401	0.321	
2	9	7	16
	7.11	8.89	
	0.502	0.401	
Total	16	20	36

Chi-Sq = 1.626, DF = 1, P-Value = 0.202

**Ho:** There is NO association between Gender and Presidential Vote in the population of UF students

**Ha:** There IS an association between Gender and Presidential Vote in the population of UF students

**TS:**  $X^2 = 1.626$

**p-value** = 0.202

**Conclusion:** Since the p-value is large, we fail to reject the null hypothesis. Could not prove that the presidential vote was significantly different for the two genders.

**Problems** – There were no problems with the sample size. The expected counts for all cells were at least 5. However, the data was not quite randomly selected, since it consists mostly of my friends.