MA217 Fall 2009
Project Part #2: Reviewing Chapters 4, 9 to 11: Sampling and Hypothesis Testing
Due: By noon on Friday, December 11, 2009 in BNW 111(no late papers accepted).
Papers must be stapled! Total 50 points

The goal of this project is for your group to design and run two hypothesis tests using the data you collected for Part 1 of the project. Each group will hand in one project write-up. All responses should be written using complete sentences, except when a chart is asked for. Your sentences should be written so that they read nicely, are grammatically correct, and can be understood by a person not taking statistics. You may use information and output from Project I.

Automatic Deductions for Each Problem:

Wrong statistical test used	-10 points (each instance)
Answers not written in sentences	-2 points (each instance)
Project not stapled	-2 points
Missing evaluation(s)	-1 point for each one missing

Problem 1: In this part you will design and run a hypothesis test comparing the numerical data you collected from your two groups from the first part of Project I. You must have at least 40 (numerical) observations for each group. If not, your group will need to collect more samples. It is OK if your two data sets have different numbers of observations.

Type (or cut and paste from Excel) your answers for questions 1 to 5 below in a Word document. Be sure the first page of your report lists all group members. Once you finished answering the questions, print out your report for Problem 1 and staple.

[2] 1.) Identify the two populations (groups) that will be compared. Be precise. For example, if you only collected data from first-year students then your population would be all first-year students at Fairfield University, not all FFU students.

[2] 2.) Define, in words, the population parameters (ie: mean or proportion) for each group that you will compare. Remember that if you are studying differences between two groups, we can't compare the numbers or counts observed. You will either need to compare the **proportions** in each group or the **averages** of each group. Which you use will usually depend on whether the response variable is categorical (proportion) or numerical (average).

[5] 3.) State which significance level your group will use as well as the appropriate null and alternative hypotheses (in words) that your group would like to investigate. Explain why your group chose your alternative hypothesis.

Do not use symbols. Make sure your sentences are grammatically correct, read nicely and can be understood by someone not taking stats.

4.) Conduct the appropriate hypothesis test. Be sure to include the following:

- State which specific test you used. [Hint: Use the name on the tab you use on the Hypothesis Test Worksheet.]
- [4] Create a chart that lists your sample information from each population that are needed to run the test. Be sure to include units for all statistics in your chart.
- [3] Give the test statistic (with appropriate symbol) and p-value.
- [7] State whether you have enough information to support your alternative hypothesis or not and interpret what that means for your specific test. Again, make sure your conclusion reads nicely and can be understood by someone not taking statistics. You should be sure that your conclusion references the correct parameter(s).
- [6] Check if the assumptions of the test are met and include any computer output and/or labeled graphs needed to check the assumptions. Be sure to explain WHY the assumptions are met. Simply stating "The assumptions are met" is not sufficient. You should state "The assumptions are (not) met because……" and provide evidence to support your claim.

[3] 5.) Briefly describe how your group collected your data. Identify the type of sampling scheme you used and identify any potential sources of bias (sampling, response and/or nonresponse).

Example for Problem 1 (Partial Solution):
Original Research Question from Project I:
On which nights do the waitstaff earn bigger tips, Thursdays or Fridays?

1.) The two populations being considered are all the waitstaff that work on Thursday nights and all the waitstaff that work on Friday nights.

2.) The parameters of interest are the average of all tips received on Thursday nights and the average of all tips received on Friday nights.

3.) I will use a significance level of 5%. The null hypothesis is that the average of all the Thursday night tips is the same as the average of all Friday night tips. The alternative hypothesis is that the average of all Thursday night tips is less than the average of all Friday night tips.

Problem 2: In Problem 2 of Project I you collected responses of categorical data for two groups. Recall that the categorical variable had between 2 to 4 possible responses. You must have at least 40 observations for each group and it is OK if your two data sets have different numbers of observations. If you don't have at least 40 observations for each group you will need to collect more data.

For this question you will design and run an appropriate hypothesis test to determine if there is evidence of a relationship between your groups and the categorical variable. [Or equivalently, if there is evidence that the categorical variable depends on your groups.]

Type (or cut and paste from Excel) your answers for questions 1 and 2 below in a Word document. Be sure the first page of your report lists all group members. Once you finished answering the questions, print out the report for Problem 2 and staple.

[5] 1.) State which significance level your group will use as well as the appropriate null and alternative hypotheses (in word, no symbols) that your group would like to investigate. Make sure your sentences are grammatically correct, read nicely and can be understood by someone not taking stats. Explain which hypothesis your group believes is true (before running the test) and why.

2.) Conduct an appropriate hypothesis test. Be sure to include the following:

- State which specific test you used. [Hint: Use the name on the tab you use on the Hypothesis Test Worksheet.]
- [1] Create a contingency table which summarizes the responses to the categorical RV that you observed for each group.
- [4] Give the test statistic (with appropriate symbol), degrees of freedom and p-value.
- [6] State whether you have enough information to support your alternative hypothesis or not and interpret what that means for your specific test. Again, make sure your conclusion reads nicely and can be understood by someone not taking statistics.
- [2] Check if the assumptions of the test are met and include any computer output and/or labeled graphs needed to check the assumptions. Be sure to explain WHY the assumptions are met. Simply stating "The assumptions are met" is not sufficient. You should state "The assumptions are (not) met because……" and provide evidence to support your claim.

Example for Problem 2 (Partial Solution):
Problem I Survey Question: Which American Idol judge is your favorite: Randy, Paula, Kara or Simon? Asked to American Idol viewers and non-viewers.

1.) I will use a significance level of 5%. The null hypothesis for the test is that the ratings for favorite American Idol (AI) judge do not depend on whether the respondent watches American Idol or not. The alternative hypothesis is that the ratings for favorite AI judge depend on whether the respondent watches AI or not.

General Comments

1.) Remember that you are being graded on the creativity, quality and accurateness of your work, not the amount of time and effort spent.

2.) Each person should be involved in every part of this project. Since these types of test will appear on the final, you need to know how to do them.

3.) Clearly, each person in the group should proofread the entire final paper and check for grammatical mistakes and mathematical inaccuracies! I will deduct heavily for these types of errors.

Firing of group members: Just like in the business world, people who are not contributing to the group may be "fired" from the group. If your group decides on this issue, you must inform the "firee" and give them a chance to improve within the group. If this does not happen, your group should give the person the complete data set and questions to be answered so that he/she may complete the project on his own. You must also inform me that the person is no longer in the group. I hope this option is not needed.

Typed Report (Guidelines below):
Your double-spaced typed report should be written using complete sentences, correct grammar and spelling. You should assume your audience has a limited statistical background and knows nothing about your data set or your analysis. You should **not** use symbols (except where specifically mentioned). Instead of using symbols, write out what the symbols represent in terms of your problem. For example, instead of saying that your null hypothesis is $\mu = 25$, write out "The null hypothesis for this test is that Ford vehicles average 25 miles per gallon under city driving conditions. The alternative hypothesis is that the true average is less than 25 mpg."

Individual Evaluations:
These evaluations should be done individually and must be **handed to me separately** from the written report. It gives you a chance to **honestly** reflect on your contributions and how your group functioned. The evaluation form is on the last page. It will not be used in determining your grade, but rather it gives me feedback about the project.

Individual Evaluation of Group Project (Parts I and II)

To be done individually and handed by separately from the group project by each group member by noon on Friday, December, 11th in BNW 111

Your honest feedback will help me evaluate this project as well as how your group functioned. **Your responses will only be read by me and will NOT be used for grading purposes**. Thank you for your feedback! ☺

1.) What is your name?

2.) Who were the other members in your group?

3.) Was it helpful to work in a group on this assignment? Why or why not?

4.) How did your group function?

5.) How was the work in your group divided? Who did what? Please be specific.

6.) Estimate the total number of hours you worked on Part 1 and Part 2 of the project alone.

7.) Estimate the total number of hours your group worked together on Part 1 and Part 2 of the project.

8.) What numerical grade would you give your group and each group member? Why? [Remember that in college we grade the quality of work, not the quantity/amount of work.]

9.) What numerical grade would your give yourself and why?

10.) Did you like this project? Why or why not? Which parts did you like/dislike?

11.) What improvements can/should be made to this project? (Not doing a final project is not an option!)