MA217 Fall 2009
Project Part #1: Review Chapters 2 and 3 – Collecting and describing data sets
Due: By noon in BNW 111 on Friday, October 23, 2009 (no late papers accepted).
Papers must be stapled! Total 50 points

Form a group of up to 4 students. Each group is responsible for handing in one report. Do not wait until the last minute to do this assignment as it will take you a lot of time to collect the data. Each person in the group should participate in all components of the project.

Important Dates:

- September 21: Get project and read it.
- September 24: Pick group members (up to 4 people per group) and submit draft of Project Proposal Form.
- October 1: Submit final draft of Project Proposal Form (typed) and begin collecting data.
- October 8: In-class Project Work Day. Complete data sets must be brought to class.
- October 23: Final draft of Project I due by noon in Bannow 111. Please be sure your project is stapled! I encourage you to print on both sides of the paper, if possible. The printers in the library have this capability. No late projects accepted.

All responses should be written using complete sentences, except when a graph or table 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.

Problem 1: In this question you will compare two groups using numerical data. First you need to determine two groups and decide what you'd like to compare. Then you need to collect data for a quantitative variable for each of the two different groups. You must collect data yourself from friends, classmates, coworkers or strangers or by making observations or taking measurements. You should have at least 40 observations for each group and it is OK if your two data sets have different numbers of observations. While you have wide latitude in choosing a topic, you may not choose a topic that involves alcohol, drugs, or any other topic which is illegal or violates the ethic codes of FFU.

Some examples of research questions to explore with this assignment:
1. Who studies more hours per week, business majors or non-business majors?
2. On which nights do waiters/waitresses earn bigger tips, Thursdays or Fridays?
3. Who has more money on their Stagcard, underclassmen or upperclassmen?

[1] 1.) Print out the raw data from Excel. You should have a total of 2 labeled columns (one for each group) with at least 40 observations for **each** column. This printout should not exceed one side of paper.

Type (or cut and paste) your answers for questions 2 through 4 in a Word document. Be sure the first page of your report lists all group members. You should resize your graphs so that you can fit both on one page. Once you finished answering questions 2 through 4 for Problem 1, print out your report and staple.

[2] 2.) Make a table that summarizes the mean, median, standard deviation, range, min, max and counts/number for each group. If you are copying the descriptive statistics output from Excel you should be sure to delete any information that is not asked for. Include appropriate units for each statistic.

[10] 3.) In Excel, make a clearly labeled and informative histogram (using histogram.xls) for each data set.

- Your graphs should have informative axis labels/titles (both on the x- and y-axes)
- Your histograms should have the same x-axis (ie: bin min and bin width).
- All data is represented on your graph.
- Choose a nice bin width so that you don't get too many empty bins or bins with only 1 observation. You also don't want too few bins.
- Copy your graphs into your word document and resize them. Both graphs should fit on one sheet of paper.

4.) You must submit a typed report (stapled, of course!) describing the differences in your data sets. You should include the following, in paragraph form (i.e. no bulleted lists) using complete sentences and correct spelling and grammar.

[5] a. Write a brief introduction that gives the who, what, where, why, when and how's.

What groups are you comparing and why?
Who did you collect data from?
When, where and how did you collect your data?

[6] b. Discuss the shape of the distribution and important features of each data set, referring to your histogram and numerical summaries to support your claims. Be sure to include if there are potential outliers and give numerical evidence to support why you believe an observation is a potential outlier.

[6] c. Make a comparison between the two groups. Which group has larger "typical" values? Which group's data is more spread out? Again, refer to your graphs and numerical summaries. Remember that there may not be a clear cut answer, but you should provide evidence to justify your claims.

5.) Save your data set and your report.

Remember to save your work often as you work on the project!!

Problem 2: In this question you will compare two groups using categorical data. First you need to determine two groups and decide what characteristic you'd like to compare. If you want, you can use the same two groups that you used in Problem 1. Then you need to collect data for a categorical variable for each of the two different groups. Create a survey question that involves a categorical variable with between 2 and 4 possible responses. Remember you need to make sure that every person surveyed can choose one of the 2 to 4 responses. You will then ask this survey question to two different groups of and compare the responses of these two groups to your survey question. You must collect data yourself from friends, classmates, coworkers or strangers. You should have at least 40 observations for each group and it is OK if your two data sets have different numbers of observations. While you have wide latitude in choosing topics, you may not choose a topic that involves alcohol, drugs, or any other topic which is illegal or violates the ethic codes of FFU**.**

Some examples of research questions to explore with this assignment:
1. Survey Question: Which baseball team do you prefer: the Yankees, Mets, Red Sox or Other? Asked to males and females.
2. Survey Question: Do you plan to vote in the presidential election?
Asked to underclassmen and upperclassmen.
3. Survey Question: Which American Idol judge do you think is best: Randy, Paula, Kara or Simon? Asked to American Idol viewers and non-viewers.

Type (or cut and paste) your answers for questions 1 through 4 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.

[2] 1.) Summarize your responses into a contingency table of observed counts for the 2 groups.

[2] 2.) Summarize your responses into another contingency table with conditioned proportions for each of your two groups.

[5] 3.) In Excel, make a clearly labeled and informative side-by-side bar graph using the conditioned proportions. Be sure to label your y-axis. Cut and paste the graph into your report. You should resize the graph so that it is no more than half a page.

4.) You must submit a typed report (stapled, of course!) describing the differences in the responses of your two groups. You should include the following, in paragraph form (i.e. no bulleted lists) using complete sentences and correct spelling and grammar.

[5] a. Give a brief introduction that gives the who, what, where, why, when and how's.

What groups are you comparing and why?
Who did you collect data from?
When, where and how did you collect your data?

[6] b. Write a paragraph (using complete sentences and correct grammar) in which you determine whether there appears to be a difference in the responses of your two groups. Be sure to refer to your graphs and contingency tables. Remember that there may not be a clear cut answer, but you should provide evidence to justify your claims.

5.) Each person should have a copy of your data set and a copy of the final project report. You will be using both them later in Part II of the Project!

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 descriptive analyses will appear on the exams and 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.

4.) Save your work often as you work on the project and as you enter your data.

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).

Automatic Deductions for Each Problem:

Wrong graph used	-10 points (each instance)
Answers not written in sentences	-2 points (each instance)
Too few samples collected ($n < 40$)	-5 points (for each instance)
Project not stapled	-2 points

PROJECT PROPOSAL FORM

Draft due 9/24 Final draft due 10/1 (typed)

Group Members:

Problem 1:

What is the numerical random variable you are considering?

What are the two groups that you will compare?

In a sentence or two, state your research question for Problem 1.

Problem 2:

What is the categorical random variable you are considering? List the 2 to 4 subcategories.

What are the two groups that you will compare?

In a sentence or two, state your research question for Problem 2.

Describe how your group will "randomly" collect your data?