

Form a group of up to 3 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 and complete the project. Each person in the group should participate in all components of the project.

Important Dates:

- September 27: Get project and read it.
- September 30: In-class work day, pick group members (no more than 3 people per group) and begin to work on proposal
- October 4: Submit draft of Project Proposal Form (neatly handwritten or typed).
- October 6: Begin collecting data, Make any changes to Project Proposal (if needed)
- October 13: In-class project work day. Complete data sets must be brought to class.
- November 4: Final draft of Project 1 is due at the beginning of class. 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 emailed projects and no late projects will be accepted.
- You will be doing both Problems 1 and 2!!

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 | -5 points (for each problem) |
| • Project not stapled | -2 points |
| • Contingency table set up incorrectly for Problem 2 | -10 points |
| • Convenience or judgment sampling scheme is used | -10 points |
| • Typos/spelling errors/grammatical errors | -1 point (each instance) |

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. Everyone in your group should proofread the project before handing it in.

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. When you collect the data, you must use a systematic or simple random sampling scheme. You should have at least 40 observations for each group and it is OK if your two data sets have different numbers of observations. (**So, you should have at least 80 observations total for this part.**) 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. You may not use the same types of variables that were collected

in the class survey (with the exception of gender) or those used in class or homework examples.

Some examples of research questions to explore with this assignment:

1. Is there an association between major (business and non-business) and the amount of hours/week students study? Or who studies more hours per week, business majors or non-business majors?
2. Is there an association between day of the week (Thursday or Friday) and the amount of tips waiters/waitresses get? Or on which nights do waiters/waitresses earn bigger tips, Thursdays or Fridays?

Note: The research question is different from the survey question!

[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/group. [So you should have at least 80 observations total.] 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. Include appropriate units for each statistic. If you are copying the descriptive statistics output from Excel you should be sure to delete any information that is not asked for and be sure to add the units.

[6] 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 specific to your research problem (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 so that both graphs 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.

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

[2] What two groups are you comparing and why? What is your research question?

[2] Who did you collect data from? When and where did you collect your data?

[2] What type of sampling scheme (systematic or simple random sample) did your group use? Explain, in detail, how you chose subjects to participate and how you collected your data.

[3] What potential types of bias (sampling, response and/or nonresponse) might be present due to your sampling scheme? Discuss all 3 types. If you don't believe you had a particular type of bias, explain the steps your group took to minimize that type of bias.

[6] b. Discuss the shapes, centers and spread of each distribution and refer to your histogram and numerical summaries to support your claims. You should include at least 2 measures of center and spread for each distribution. Be sure to identify potential outliers and give numerical evidence to support why you believe an observation is a potential outlier.

[6] c. Use the results of part b to answer your research question. Does there appear to be an association between your 2 variables? Remember that there may not be a clear cut answer, but you should provide numerical 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. When you collect the data, you must use a systematic or simple random sampling scheme. You should have at least 40 observations for each group and it is OK if your two data sets have different numbers of observations. (**So, you should have at least 80 observations total for this part.**) 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. You may not use the same types of variables that were collected in the class survey (with the exception of gender) or those used in class or homework examples.

Some examples of research questions to explore with this assignment:

1. Research Question: Is there an association between gender and favorite baseball team?
Survey Question: Which baseball team do you prefer: the Yankees, Mets, Red Sox or Other? Asked to males and females.
2. Research Question: Is there an association between grade level and whether students own an iPhone?
Survey Question: Do you own an iPhone? Asked to underclassmen and upperclassmen.

Note: The research question is different from the survey question!

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.) Define your explanatory and response variable. Summarize your responses into a contingency table of observed counts for the 2 groups. **Please note that if the table is set up incorrectly (given your explanatory and response variable) then you will automatically lose 10 points since parts 1 – 4 and the subsequent analysis will be incorrect.**

[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?

What is your research question?

If different from the first part, please include when and where you collected your data.

If different from the first part, explain what type of sampling scheme (systematic or simple random sample) your group used. Explain, in detail, how you chose subjects to participate and how you collected your data.

[6] b. Write a paragraph (using complete sentences and correct grammar) in which you answer your research question. Does there appear to be a difference in the responses of your two groups or does there appear to be an association between your 2 variables? Remember that there may not be a clear cut answer, but you should provide numerical evidence to justify your claims. Be sure to refer to your graphs and contingency tables. Remember to provide numerical evidence to justify your claims.

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. Each person in the group should have a copy of the data and final project. You'll need it for Part 2 of the project.
- 5.) You need to hand in both Problem 1 and 2!!

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).

This is a copy of the rubric I will use when grading your group's project.

| Grading Rubric (Project Part 1) | | |
|---|-----------------|---------------|
| | Possible Points | Points Earned |
| Problem 1 | | |
| 1.) Raw Data | 1 | |
| 2.) Descriptive Stats (w/units) | 2 | |
| 3.) Graphs (2 histograms with labels on 1 page) | 6 | |
| 4.) Intro description (groups, research questions) | 2 | |
| Who, when, where | 2 | |
| Sampling scheme description | 2 | |
| Biases (sampling, nonresponse, response) | 3 | |
| 5.) Shapes of distributions (shape, center, spread, outliers) | 6 | |
| 6.) Answers research question, supports with numerical evidence | 6 | |
| Comments: | | |
| | | |
| | | |
| | | |
| Problem 2: | | |
| 1.) Contingency Table (raw) | 2 | |
| 2.) Contingency Table (Conditioned Proportions) | 2 | |
| 3.) Side-by-side bar charts (labeled) | 5 | |
| 4.) Intro description (groups, why, who, what, where) | 5 | |
| 5.) Answers research question, supports with numerical evidence | 6 | |
| Comments: | | |
| | | |
| | | |
| | | |
| Automatic Deductions: | | |
| Wrong graph used (each instance) | -10 | |
| Answers not written in sentences (each instance) | -2 | |
| Too few samples collected (each instance) | -5 | |
| Project not stapled | -2 | |
| Contingency table set up incorrectly for Problem 2 | -10 | |
| Convenience or judgment sampling scheme is used | -10 | |
| Typos/spelling errors/grammatical errors | -1 | |
| | | |
| Project Total (out of 50) | 50 | |

Additional Comments:

PROJECT PROPOSAL FORM

Due October 4, 2010 (typed or neatly written)

Group Members:

Section (Circle one): 8am or 9am

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. [Note: This is NOT the survey questions your group will use to collect the data. The research question is what you hope to answer and learn about your 2 groups from the data. Refer to the top of page 2 of the project.]

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. [Note: This is NOT the survey questions your group will use to collect the data. The research question is what you hope to answer and learn about your 2 groups from the data. Refer to the bottom of page 3 of the project.]

In this section, describe how your group will collect your data.

- When will you collect the data?
- Where will you collect the data?
- Who will collect the data?
- Submit a copy of your survey. [You should test your survey on a few people just to check that your survey questions/instructions are clear, all units are defined, etc.]
- Describe how you will select your sample. Remember that your sample must be either a systematic or simple random sample. **Be specific.**

An answer like: “We will just randomly ask 40 males and 40 females entering the library” is not acceptable since this is not a random sample, but a judgment sample. I want something like “We will ask every 10th person who enters the library until we have observations from 40 males and 40 females” or “We will use a random number generator with the numbers 1 to 78 and select 10 numbers. Then we will go to those dorm rooms in Jogues and give the survey to the person who answers the door”.