**Stat 301 -- Mini-Project 2**

**Designing and Analyzing a Randomized Experiment**

**Goal:**  To design an experiment to collect data which you will then describe and analyze using the methods of Chapters 3. (This limits you to a categorical response. If you really want to analyze a quantitative response variable, I can discuss this individually with your group.)

**Teams:** You are to work in teams of 3 people. It is up to the members of the group to make sure everyone contributes equally.  Plan your schedules so that you will have time to work together on the project outside of class. Teams should be formed and preliminary project topics selected by **Feb. 17**. You may be asked to share your proposals with the rest of the class. Please start early so you have time to ask questions. You should have your data collected by **Feb. 20** (guideline).

**The Study:**You are free to choose your own topic. You should think of two groups that you can compare with one binary response variable **through a randomized, comparative experiment**.  Make sure you choose a topic for which it is feasible to gather the data in a relatively short period of time. The question may be related to your major or some other topic of interest.  For example, you could have a group member ask someone for change, randomly determining how they are dressed at the time, or you could randomly assign people to take a survey with two different wordings and see whether they respond differently depending on how the question is asked.  Your study must obtain **at least 10 experimental units in each explanatory variable group**.  (Of course, more is better, because it gives your test more power = a better chance of detecting a difference between the treatments if there really is one.) Your experimental units do not have to be people! There will be credit for creativity. **You do not need to worry about selecting a representative sample**, you will focus more on comparing the groups to each other.

**Project Proposal (5 pts):**  Due **Feb. 17**. Identify the team members, your research question, and your data collection plan (experimental units, explanatory variable, response variable, how you will carry out random assignment). This proposal should be uploaded by one team member into Canvas.  Please include the names of the individuals in your group!

**Final Report (45 pts):**  Due **Feb. 27**. This should be a typed report, written collaboratively by all team members.  Your report should be written as to other student researchers.  Make sure it includes at least the following (and please use these section headings in your report):

*I. Introduction* (5 pts) – Why did you choose this topic?  What did you expect to find?  Have similar studies been done elsewhere?  Why should the reader be interested in your results and continue reading? Make sure you include a clear statement of your driving research conjecture. (What “direction”?) Identify your response variable and your explanatory variable.

*II. Summary of Data Collection Methods* (10 pts) – How (e.g., when, where) did you collect the data?  What were the experimental units? How many? How did you find them?  What groups did you compare, how did you form the groups (explanatory variable)? In particular, how did you carry out the random assignment process? What was your response variable?  How was this variable measured? (Include copies of any surveys or other documentation.) Any “operational definitions” (e.g., how defining handedness)? What additional “controls” did you exert in the study (e.g., where subjects blind to your treatment conditions)?  Did anything go wrong or unexpectedly during the course of the study? **Note:** You can never give me too much detail in this section!  In particular, there should be enough information that someone else could replicate your study on their own based only on your description (and hopefully improve upon it based on your suggestions in section 4).

*III. Analysis of Results* (20 pts)

*Descriptive statistics:* Include appropriate numerical and graphical summaries of your data.  **Write several sentences explaining what you found in these sample data**. Make sure you integrate the output into the body of the report and include discussions of how you are interpreting the message in these summaries.

* Categorical data
  + A two-way table of your results
  + Two different statistics to summarize your data
  + A well-labeled segmented bar graph or mosaic plot

*Discussion should include:*

In your discussion you should fully describe your sample, sample size, and *interpret* the sample statistics in context. (Remember to be careful in use of “of”, “and”, and “that” in interpreting proportions.) Do the sample data provide preliminary support for your research conjecture? You should also comment on any unexpected results in your data.

*Inferential:*

* Define the population(s) and parameter(s) in words.
* State the null and alternative hypotheses in symbols and in words.
* State what a type I and a type II error would represent in this context and their potential consequences. Which do you think is the more serious error?
* Use simulation, exact, and large-sample methods to analyze your data.

1. Simulation: Use a simulation to obtain an empirical p-value. Your simulation methods must be clear. Include copies of your applet input/output.
2. Exact method: Calculate an exact p-value using Fisher’s Exact Test
3. Large-sample method: Carry out a two-sample *z* test (test statistic and p-value) and obtain a confidence interval for a parameter of interest. Comment on the validity of this approach.

Include details of your technology use and include a copy of the computer output verbatim (e.g., screen capture or copy and paste- if the latter, watch the font and font size to make sure the layout is clean). Compare your p-value results across the two/three methods – Are they close? Do you expect them to be (i.e., is the large-sample method appropriate for your data)?

* Include a careful *interpretation* of what the p-value tells you in this analysis.
* Discuss the *decision*you will make from this p-value: Is the difference between the groups statistically significant?
* Include a careful interpretation of your confidence interval, in context, assuming it is a valid interval. (What does it tell you about your research study?)

(Note: All computer output should be included in the body of the report.  Make sure all figures and graphs are clearly labeled.)

*IV. Conclusion* (5 pts)

* Summarize the conclusions you would draw from these results (significance, confidence, causation, generalizability).  Be sure to reference the relevant results to support your conclusions and to refer back to the type of study conducted in explaining the scope of your conclusions.
* Reflect on your study: What did you learn?  Did the data behave as you expected?  What questions remain?
* Critique the methods used to collect the data.  Is there anything you would do differently next time?  How might this affect the conclusions of the study?  What similar questions might someone chose to investigate in the future to build on your results?

*Presentation*(5 pts) - style, organization, layout, grammar (e.g., affect vs. effect), presentation of report, creativity

**Final Project Submission:**You are encouraged to complete your project before exam 2 or at least make sure you could.

**Example Previous Project Topics:**

* Are people in an elevator more likely to initiate a conversation with a stranger if leaning against wall or standing straight up?
* Are males more likely to open a door for a female than for a male?
* Are students less likely to vote for a Panda Express on campus if it increases meal card cost?
* Does the quality of the picture influence someone’s food choice?
* Does slow or fast classical music affect color choice?
* Can students tell the difference between cookies made with sugar and Splenda?
* Does soaking popcorn kernels in water first reduce number of duds?
* Does tapping on the top of a soda can reduce the amount of fizz?