**AP Statistics Project**
**Serving the Community through Statistics**
**The Final Report**

**DUE: Friday, May 4**

In general, a statistical project is the process of answering a research question using statistical techniques and presenting the work in a written report. The research question may arise from any field of scientific endeavor, such as athletics, advertising, aerodynamics, or nutrition. The research question(s) for this project will stem from our class partnership with Mobile Loaves and Fishes. In a statistical project, a written report is used to present the findings. What follows in this document are some basic guidelines for developing a statistical project in general and some specific details about how to write the final paper for this particular project.

Now that we have administered the survey, our last step in the project is to analyze the data, write a report of our findings, and present the results. We will analyze the data together in class. The final report will be written **individually**. Think of this situation as analogous to a lab report – we ran the "lab" together as a class and we all have the same results, but you need to communicate those results and their implications in your own words. Great latitude may be taken in developing the written report. Students should plan how to communicate their work effectively. The longest report does not necessarily represent the best project. However, the report **must** accomplish the following:

1.  **Revise your proposal.**

The final report will build off of the proposal. In other words, the proposal will serve as the introduction to the final report answering the *who?*, *what?*, *when?*, *where?*, *why?*, and *how?* questions. **The proposal addresses steps 1-2 of the investigative process and the final paper will cover steps 3-6.**

The first step of writing the final paper will be to edit the proposal based on the instructor's feedback. The *how?* section of the proposal needs to be rewritten from a future tense "we will…" to a past tense "we did…" when discussing how data was collected. Comment on how it the proposal was actually implemented. If problems arose, how were they addressed? Reflect on how bias was handled effectively or could have been handled better. Make your case for why the obtained data is trustworthy.

Here is some practical advice from a real statistician (PARS), taken from American Statistical Association's guidelines for student projects:

> Research questions need to be stated clearly. Upon completion of the project, it should be reviewed to be certain the question being posed was actually answered.
>
> The data collection process should be described clearly, and the student's role in the data collection should be clear. The variables in the study should be defined clearly in terms of what is to be measured and how. If a random sample is taken, the randomization process should be given.

2. **To your revised proposal, add the following sections:**
   a. Data Summary (Step 3)
   b. Data Analysis (Step 4)
   c. Conclusions (Step 5)
   d. Reflections (Step 6)

3. **Data Summary: a summary of important data in the form of graphical displays.**

Thoughtful analysis of the data may take many forms and should be guided by the question and how the data were collected. Usually, it is best to begin by graphing the data.

Students will be required to create a minimum of **three** different (yet appropriate) graphical displays that demonstrate different aspects of the obtained data (bar chart, comparative bar chart, histogram, box plot, etc.). These displays should effectively summarize what was learned from the survey that will be analyzed further to answer the question of interest. The rules for creating graphical displays from previous units should be followed, making sure the graphs are well labeled, easy to compare, and help answer the question of interest. Tables summarizing the numeric results should be included with the graphical displays.

All graphs and tables must be computer generated. In the course folder, under "general course documents," there is a file: "How to do stuff in Excel." We will cover how to make some graphical displays in class, but this handout will be a useful reference.

PARS:

> Graphical displays provide insights into data. Many projects fail to take advantage of this important statistical tool. In projects using at least one graphical display, the graphs often are only the most rudimentary pie and bar charts. Stem-and-leaf, dot plots, box plots, and scatter plots are some of the methods that might provide more insight into the data. Displaying sample means with error bars also may be helpful. Care should be taken to use appropriate graphs. For example, line plots and scatter plots are used sometimes when bar charts would be better. Replication permits variability to be captured by the data; appropriate graphs make it visible.

4. **Data Analysis: discuss the inferences that can be drawn from the data.**

Delineate what conclusions were obtained. Students will be required to conduct at minimum the following methods of statistical inference:

- At least one inference must be drawn from proportions (z) and one from means (t)
- At least one inference method should follow a one-sample procedure and at least one method should follow a two-sample procedure
- At least one confidence interval should be constructed and at least one hypothesis test should be run
- At least one Chi-square test must be performed on variables across multiple survey questions

5. **Conclusions and discussion.**

<u>All conclusions should be interpreted in non-statistical language.</u> Remember the audience for your report is Mobile Loaves and Fishes. Don't use terms like "p-value," "alpha," or "null hypothesis" – you did all of that in the data analysis section. Here you need to explain what those results mean practically for MLF.

---

PARS:

Once analysis is complete, the question should be answered. The data may not be able to provide a conclusive answer. For example, one treatment may appear to be better than another, but the difference was not statistically significant. If the question has a definitive answer, that should be presented. A check should be made at this point to make certain the answer matches the question. It is easy to get caught up in the analysis phase and obtain many answers, none of which addresses the research question.

When a sample is drawn, inferential statistics usually are needed to answer a question. While useful, graphs and descriptive statistics alone are not sufficient in this instance. When using formal inferential statistical tests, the assumptions for any method should be checked. For example, variances should not be pooled if they are substantially different (which can be tested) and the sample sizes are reasonably large. Students should fully understand the methods they use, otherwise inappropriate statistical terminology may be used. It is better to use simpler (but appropriate) methods correctly than to apply more sophisticated procedures improperly.

For hypothesis tests, care should be taken to state the null and alternative hypotheses appropriately. Remember that in a subject-matter area, the hypothesis is what the researcher wants to prove. In statistics, this usually becomes the alternative hypothesis, as the strongest conclusions can be drawn from rejecting the null in favor of the alternative. Note the null hypothesis is never 'accepted.' Instead, it is traditional to say "we failed to reject the null hypothesis," which gives the proper impression that it is not known with

certainty that the null is true but that the data do not refute it. The reason for this is the probability of a type II error is not known.

Confidence intervals can be misinterpreted. For example, a confidence interval cannot confirm a test statistic because the test statistic is, by construction, the center of any confidence interval.

6. **Reflections.**

Discuss the strengths and weaknesses of the selected statistical methods. Reflect on your results and discuss why they are trustworthy or be honest about why they are not. Discuss what went well and what didn't. Look ahead to next year's stats classes doing a follow-up study – discuss how the study could be been improved if done again.

7. **Putting it all together and formatting.**

The final report should be no less than four and no more than eight typed pages with 1-inch margins, 12 pt. times new roman font, and 1.5 line spacing. Keep in mind that you are adding to the proposal, so you will already be starting on page 3 or 4. Also, this length takes into account that you will be including tables and graphs.

Please reference the rubric to be sure you have met all the requirements before submitting your paper.

PARS:

Font size should be at least 12 pt., and complete sentences and standard grammar should be used. The writing emphasis should be on the statistical aspects of the study. Background information should lead to a precise statement of the question to be considered. Some projects benefit from a more detailed description of the data collection phase. Details of the statistical analysis should be presented. The statistical methods should be outlined and discussed clearly. The analysis should serve as the foundation for any conclusions drawn. A "reflection on the process" should be a realistic self-evaluation of the work. Simply stating that all went well raises concerns, as few studies ever have *everything go right*.