## AP Statistics Project
## Serving the Community through Statistics

**Project Overview:**

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. 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.

The research question(s) for this project will stem from the service agency that we partner with. For the 2015-2016 school year this will be Mobile Loaves and Fishes. Students will provide the following four services: meeting with agency and developing a survey instrument, conducting survey, compiling data and performing statistical inference procedures, and presenting results.

All students will compose a written report on the findings of the project to submit to the instructor. The instructor will select one report to be used as the basis of the presentation that will be given to the partnering agency.

The exact substance of the project will vary from year to year depending on the needs of the service agency. There are a few components that will be included in every project and a grading rubric based on these components is provided and explained below.

**Component 1: Selecting a Question and Designing a Study**

The process of developing a statistical project should demonstrate the scientific method and pose a focused question or questions, collect appropriate data, analyze the data thoughtfully, and draw correct conclusions.

Once a question is proposed, students should examine it. First, is it a question that can be answered? (The question "Is there intelligent life in the universe that does not come from Earth?" is an extremely interesting question, but not one that is likely to be answered in a short-term project.) Second, can students collect data to answer the question or has someone else already collected data that could be used to find the answer?

Once the question is chosen, data must be collected. If published data are used, students should understand how the data were obtained and record their source. Usually, students will need to collect their own data. Time should be spent deciding how to collect this data. If a survey is used, how are the people chosen to answer the questionnaire? How will bias be avoided? How will the data be recorded?

This is the proposal phase of the project. The first step in this project will be to meet with the instructor to review drafts of the proposal, receive feedback, and plan to move forward with data collection.

**Practical Advice from Real Statisticians (PARS): Selecting a Question**

Selecting a good question for a statistics project is important. Not only should the question be interesting, it should give rise to data that lend themselves to statistical treatment. For example, if the question leads to a categorical response (i.e., What is your favorite color?), one may be left with nothing more than a few counts (one for each category). This limits both the graphical and statistical analyses that can be used. Be sure the question can be answered with the data collected. Questions need to be stated clearly. If more than one question is posed, each should be answered. Finally, upon completion of the project, it should be reviewed to be certain the question being posed was actually answered.

## Component 2: Data Collection and Compilation

After the details have been worked out, students are ready to obtain the data. Great care should be exercised at every stage of data collection. Careless measurement or recording of data cannot be remedied in the analysis phase of a project.

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. These displays should effectively summarize what was learned from the survey that will be analyzed further to answer the question of interest.

PARS: **Collecting Data**

Collecting data properly is challenging. Students who find data that have already been compiled often do not realize the pitfalls and potential errors of data collection. As a consequence, they miss an opportunity to understand this vital phase of any project.

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. Haphazard or other unplanned sampling is not random sampling and can lead to biased results.

Replication is important in any study. For example, the purpose of a study may be to compare the growth of a corn plant with and without fertilizer. Suppose two pots are used and two corn seeds are planted in each pot. Then it is randomly determined which pot gets which treatment (fertilizer or no fertilizer). Even though there are two plants under each treatment, there is no replication. The reason for this is that treatments were assigned randomly to pots (not plants). More than one pot would have to be used for each treatment for there to be true replication.

If a survey is conducted, a copy of the survey should be included in an appendix. For all projects, raw data should be included as an appendix.

**Component 3: Statistical Inference and General Conclusions**

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.

Finally, consider the strengths and weaknesses of the project. What would be changed if the project was done again?

## Putting it all Together: The Written Report

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:

- Demonstrate how and why the particular topic was chosen
  - This introductory section of the report will draw heavily on the approved proposal and answer similar questions (Who, What, When, Where, Why, How).
- Show how the research was conducted
  - Again draw from the proposal and explain how the study was designed. Then comment on how it 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.
- Include the collected data and its analysis
  - Include a copy of the survey questions and tables summarizing the numeric results. Include a minimum of three different graphical displays (bar chart, comparative bar chart, histogram, box plot, etc.) that each display a unique aspect of the data.
- Delineate what conclusions were obtained
  - Conduct a minimum of one hypothesis test and one confidence interval using the obtained data and interpret the results.
- Discuss the strengths and weaknesses of the selected statistical methods
  - Reflect. No statistical study is perfect, even at a professional level. Be honest about the obtained results and the confidence with which they should be viewed. Always conclude by suggesting how another researcher could take this report and expand/improve upon it in the future.