**Statistics 301 A Project #1 Fall 2016**

The projects allow you to apply your data analysis skills to a practical problem. The project includes description of data, inference about answers to questions, and evaluation of assumptions. You will work in a team of 2 to 4 people, although please talk with me if you very much prefer to work as an individual. Below, I provide three topics, some with a choice of variables. Your team is to choose a topic and variable. I will provide you additional background information and the appropriate data set. Your team will write a technical report of not more than 8 typed (11 point font or larger) double-spaced pages. Reports will not be returned, so make a copy if you want a copy for yourself.

Your analysis should consider:

* characteristics of the variables you are working with
* the relationship between the explanatory and response variables
* inferences about parameters
* evaluation of model assumptions

Your report should include:

* An executive summary (one or two paragraphs) that briefly but completely summarizes the results of your analysis. Although this will appear at the beginning of your report, it is likely to be the last thing you write.
* A summary of relevant aspects of the study design.
* A description of the characteristics of the variables you are working with. Include any graphs, summary statistics, and analysis that support your description.
* A description of the relationship between your explanatory and response variables. Again, include any graphs, summary statistics, and analysis that support your description. You may need to transform one or more variable if the initial analysis has problems with assumptions.
* An evaluation of assumptions for the ``final’’ analysis.
* Answer(s) to topic-specific question(s), including inferences to the relevant population. These should contain specifics and be phrased in subject-matter terms.

Write your report with a minimum of statistical jargon. Jargon is especially inappropriate in the executive summary and your answers to topic-specific questions. For example: “X is statistically significant” is jargon and neither specific nor phrased in subject-matter terms.

Turning in a wad of JMP output is **not** an appropriate statistical analysis. The body of your report should include only those results that support your description or inference. You should cut/paste important figures into your report. Most numeric results will be better presented in a form other than the JMP output. Remember that JMP usually gives you more than you need so do not include graphs or output that is not needed. Including unnecessary JMP output will result in deductions from the final project score.

The project grading rubric will be distributed within a week. If everyone on the team contributes approximately equally, each team member will receive the same score. When the report is submitted, any team member can include a separate sealed envelope giving her/his assessment of the relative effort contributed by each team member. This should be expressed as a percentage of the total team effort for each team member. If effort was wildly unequal, I will discuss the issue with team members and if appropriate, adjust individual scores. If no envelopes are turned in, I will assume equal contributions.

**Due dates**:

Monday, 3 Oct, end of lecture: Each team submits a list of team members and their choice of topic and variables. I will provide each team its data set before the end of lab on Tuesday, 4 Oct.

Friday, 21 Oct, end of lecture: Turn in your report. No HW due that week.

The information below is provided to help you choose a project. Additional background information will be provided after you choose your topic.

**Topic 1: Air Pollution in US Cities**. These data are from 75 Standard Metropolitan Statistical Areas (the US Census Bureau designation for a city and its surrounding region) in the US. These data and other, similar data sets have been used to investigate many different relationships. For project 1, you will look at the relationship between the climate in a city and the average amount of an air pollutant.

Variables: Choose one climate variable and one air pollutant

 Climate variables: average July temperature **or** average annual rainfall

 Air pollutant: HydroCarbons (HC) **or** Sulphur Dioxide (SO2)

**Topic 2: Mercury in Bass.** Bass are a common freshwater fish. Because they are a top predator, they can have high concentrations of mercury. Thesedata are from 60 lakes in Florida. For each lake, the data set contains the average mercury concentration in three-year-old fish and four water characteristics. You will look at the relationship between fish mercury and one water variable.

Variables:

 Alkalinity, pH, Calcium, **or** Chlorophyll

**Topic 3: Crawling in babies.** Babies learn to crawl at about 6 months. In winter, babies are often bundled in clothes that restrict their movement. These data were collected to see whether this may delay learning to crawl. The data are the age (in weeks) when a baby learned to crawl and the average temperature six months after birth. The data are all from one US city, so the major variation in average temperature is that from winter to summer. There is no choice of X variable for this topic.