Project 1: Describing relationships

The purpose of this assignment is to tell the story of set of bivariate data using descriptive statistics and exploratory data analysis. You will first examine and describe each variable separately, then describe the relationship between the two variables. This will involve making different representations of each data set, calculating measures of center and variability, as well as looking at the data together using scatterplots and correlation. Your goal is to tell the story of what your data represent and how the two variables are related. You may use a calculator or computer for any part of this assignment.

Step 1: First, you need a problem that will lead to the collection and analysis of a set of bivariate data. Here are some sample questions:

1. Do you think that people from large families tend to marry people from large families, and people from small families tend to marry people from small families, or do you think that there is no relationship between the family sizes of married couples?

2. Do you think that there is a relationship between the price of a breakfast cereal and its sugar content (grams/oz?)

3. Do you think there is a relationship between the number of hours of sleep students get on average and the average number of caffeine drinks they consume each day?

4. Do you think that there is a relationship between the number of pages in a textbook and its price?

Email Progress Report 1: Posing a problem

<table>
<thead>
<tr>
<th>What problem will you investigate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where and how will you obtain your data?</td>
</tr>
<tr>
<td>How many pairs of data value will you collect?</td>
</tr>
</tbody>
</table>
Step 2: Collecting and analyzing your data
After your topic has been approved, go ahead and collect your data.

| Progress Report 2: Show me your data | Due date: |

Data Analysis

You will enter your data into Data Desk to generate graphs and statistics. You need to cut and paste or attach printouts or copies of all of the analyses as described below. You will need to:

1. Make three different representations of each variable (e.g., dot plot, histogram, boxplot). You may do some of these by hand or on the computer.
2. Calculate mean, median, mode, range, standard deviation, and interquartile range for each variable.
3. Make a scatterplot of the variables and calculate the correlation coefficient.

You may want to show me your data analysis at this point to check that it looks ok.

Step 3: Writing your report

Use the following format to interpret your data analyses and write the report. Use the four headings and attach copies of your analyses.

1. The Problem
   State the problem you set out to investigate.

2. Method
   Tell what your data represent, how many cases you collected, and where/how you obtained your sample of data. (Include a copy or list of your data at the end of your report.)

3. Analysis/Results
Describe your first variable. Talk about the shape of the data, the center, and the spread, and any interesting feature (e.g., outliers) in the context of the variable. Be sure to refer to the different graphs you generated.

Repeat for the second variable.

Describe the relationship between the variables as shown in your scatterplot and correlation analysis. Be sure to comment on any interesting features such as outlier or influential values.

4. Conclusion and Critique

What did you learn about the problem you investigated?

What might have affected your results? What would you do differently if you did this project again?
Attach this cover sheet to the front of your project with your name filled in and make sure the project is firmly stapled or clipped.

Name________________

Evaluation of Project 1

___ Clear statement of the problem investigated. (2 pt)

___ Clear description of the data: what they represent, number of cases, and how they were collected. (2 pt)

___ Analyses are complete and correctly done. (6 pt)

___ Graphs are completed and correctly done. (6 pt)

___ Each variable is described separately, and these descriptions are correct and complete. (6 pt)

___ The relationship between variables is correctly and completely described. (6 pt)

___ The conclusion and critique are complete and seem reasonable. (5 pt)

___ A list or copy of the data is included. (1 pt)

___ The paper is legible (either typed or printed) and easy to read. (1 pt)