MA155 Statistical Reasoning
Chapter 4 Project and Scoring Guide
Points: 50
Be sure to turn in this sheet with your project. I will use this to score your project.

Through research, you will find two sets of data that you will analyze and compare as follows. It’s best to think of an idea first and then look for the data. The two data sets should have a common characteristic to compare. The data needs to be quantitative that makes sense to compute the mean, median, 5 number summary, etc. Data could involve sports (Does the American League score more runs due to a designated hitter than the National League? Home runs is the common characteristic.), cars (Do foreign cars get better gas mileage than American cars? Gas mileage is the common characteristic.), people (Are salaries lower in the Midwest than on the east coast? Salary is the common characteristic.) Some websites with data are: http://lib.stat.cmu.edu/DASL, http://www.usa.gov/Topics/Reference_Shelf/Data.shtml, http://www.fedstats.gov/. The text gives other sites for data along with sport sites, car sites, etc. It should be real data, not something made up for a problem. The paper that you turn in should be typed. The paper should be set up in the same order as the description of the project, but should be written as a paper and not numbered 1, 2, 3... If you need to type a mathematical expression or equation and do not have Equation Editor on Microsoft Word, you may download a 30 day free trial of MathType by going to www.dessci.com.

1. Write one or two good paragraphs explaining why you chose this topic and the corresponding data sets and what you think the graphs and statistics may show. It may be helpful to do a little bit of background research on the topic. Remember sources should always be stated.
   - Points: 5 or 3 or 0
2. Your data sets should be included in an organized form. Data sets should have at least 20 values per set. Use all the values in the data sets. Do not arbitrarily cut off the data set at 20 values. Source should be stated. Data sets must be quantitative and appropriate to analyze using the 5-number summary, mean, standard deviation, etc.
   - Points: 7 or 4 or 0
3. Use SPSS or Excel to calculate the 5-number summary, range, mean, and standard deviation of each data set. Copy the statistics provided into your paper.
   - Points: 9 or 6 or 3 or 0
4. Use SPSS to draw parallel boxplots. (Students in the ITV class may draw the boxplots by hand, but should be very neat. Use a ruler, set the scale and black ink. It should look like part of your paper.)
   - Points: 5 or 3 or 0
5. Show the calculations to determine outliers for each data set. Remember mild outliers are values less than $Q_1 - 1.5(IQR)$ or values greater than $Q_3 + 1.5(IQR)$. State any outliers.
   - Points: 5 or 3 or 0
6. Using the statistics and graphs from #3-5, compare the two distributions. Be sure to discuss shape, outliers, center and spread (SOCs). Be very specific. (For example: Data set 1 appears to have a bell-shaped distribution while the shape of data set 2 is skewed to the right. A measure of center, the median (5.6) of data set 1 is less than the median (7.8) of data set 2. In fact, the median of data set 1 is very close to the first quartile of data set 2 which indicates that 75% of the scores in data set 2 were better than 50% of the scores in data set 1, etc.)
   - Points: 9 or 6 or 3 or 0
7. Write a good concluding paragraph or two. Did the statistics and graphs support what you thought it might? Can you reach any conclusions, etc.? Any ideas for further study?
   - Points: 5 or 3 or 0
8. Paper has an organized, typed, easy to follow presentation.
   - Points: 5 or 3 or 0
MA155 Statistical Reasoning
Chapter 7 Project and Scoring Guide
Points: 50
Be sure to turn in this sheet with your project. I will use this to score your project.

Through research, you will find a set of paired data that you will analyze and compare. It’s best to think of an idea first and then look for the data. Think of a question that you would like to attempt to answer. “Will a woman every run a 4 minute mile?” You could look at the women’s winning time for the 1600 m race over the last 20 Olympics or World championships. “Is there a correlation between the number of homicides and the average income in a city?” You could find the homicide rate and the average income for 20 cities in a specified year. Set up your paper as follows. It should be typed.

1. Find two sets of data that have a correlation, r, greater than 0.5 or less than -0.5. Remember you are looking for pairs of values. You should have at least 20 pairs. Use all of the necessary data. Don’t arbitrarily cut it off at 20 pairs. The only place 100 pairs will take more time is typing it into Excel or SPSS and many times you can copy and paste the data. Be sure to check the value of r before proceeding with the project. If you have significantly more than 20 pairs of data, you can get by with a smaller value of r. If this is the case, discuss it with me before proceeding. It is best to think of an idea first and then look for the data. For ideas, think of things we discussed in class, look over examples and exercises in the text, sports sites, data sites, etc. Write a paragraph as to why you chose the data and what you think it might show.
   Points: 6 or 3 or 0

2. List your data in an organized form. State the source. State what the explanatory and response variables will represent and why. Example: Poverty level will be the explanatory variable because people cannot always choose their level of income. It will be the independent variable to see if I can predict the homicide rate.
   Points: 8 or 4 or 0

3. Use SPSS or Excel to calculate and state the value of r. Discuss the direction and strength of the linear relationship.
   Points: 6 or 3 or 0

4. Use SPSS or Excel to calculate and state the equation of the least-squares regression line.
   Points: 6 or 3 or 0

5. Use SPSS or Excel to draw a scatterplot, including the line of best fit. Paste this into the project.
   Points: 6 or 3 or 0

6. Use the least squares regression line to predict the response variable for a selected explanatory variable. Be sure to use a value that it is reasonable to use to make a prediction. Example: What will be the average winning time for the 2012 women’s 1600 meter run?
   Points: 6 or 3 or 0

7. Write a concluding paragraph. Was the correlation between the data in the direction and strength you expected? What did you learn about your topic? Is there causality, an underlying cause or is this relationship coincidental? Any ideas for further study?
   Points: 6 or 3 or 0

8. Project is typed, organized in the above order and easy to follow.
   Points: 6 or 3 or 0