

- 1) Obtain the height-weight dataset from <http://www2.stetson.edu/~jrasp/data.htm> (go to that webpage, then look for bodyfat.xls).
- 2) Use Matlab's boxplot command to produce a boxplot of heights. You will see one outlier.
- 3) Compute the mean, the median, the standard deviation, and the interquartile range of the height data. Now remove the outlier, and compute the mean, the median, the standard deviation, and the interquartile range of the resulting data. What has happened? Why?
- 4) Use Matlab's boxplot command to produce a boxplot of weights. You will see two outliers.
- 5) Compute the mean, the median, the standard deviation, and the interquartile range of the weight data. Now remove the outlier, and compute the mean, the median, the standard deviation, and the interquartile range of the resulting data. What has happened? Why?
- 6) Are the weight outliers the same as the height outliers? What is going on here?
- 7) Use correlation to predict the age, weight, adiposity, neck, chest and ankle for a person 70.3cms high. How good is each prediction? Which one would you trust the most? Which one would you trust the least? Why?

This exercise should be done individually, and is due 10 Sep. Prepare a PDF document with the answers (including figures) and email it to me at daf@uiuc.edu Your email should have CS498 Homework in the subject, otherwise I might miss it.