Title: The Accuracy of Body Fat Percentage Estimation With Body Composition Measurements

Research Question: Can different body measurements be used to accurately predict body fat percentage and the Adiposity index of a person?

Thesis: I will analyze the Fitting Body Fat % dataset to determine if body composition measurements can accurately predict body fat percentage, based on the Brozek formula, using Age, Weight, and Height along with Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Bicep, Forearm, and Wrist measurements. I will then determine the optimal combination of mining algorithm and variable selection to yield the most precise prediction.

Abstract: I would like to study the Fitting Body Fat % dataset for my research project. This dataset provides underwater weighing density, age, weight, height, and many different body measurements for over 252 men. Determining a person’s density with underwater weighing is an accurate method of determining body fat. For this project, I will examine the variations between different transactions to determine if there are any associations with measurements of the body and given body fat percentages and Adiposity index as an alternative method for accurately predicting these values. The measurement variables I will be examining are neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, and wrist circumference. I will start the project with the assumption that age is a completely independent variable and does not have any association with body fat percentage or the Adiposity index. Additionally, I will focus my initial analysis on the notion that the abdomen circumference will have the highest correlation with body fat percentage.

The Fitting Body Fat % data set is located at:

<http://people.stat.sfu.ca/~tim/assgts/bodyfat>

-Brandon

Outline:

1. Abstract
2. Introduction
3. Prior Research
   1. Overview of at least two prior studies with similar data
   2. Data
      1. Description
      2. Preparation
4. Methods
   1. Model Selection
      1. Explain why models were choose
   2. Multiple Linear Regression
      1. Description of method
   3. Recursive Partitioning Decision Tree (rpart)
      1. Description of method
   4. Naïve Bayes Classification
      1. Description of method
5. Experiments and Results Analysis
   1. Multiple Linear Regression
      1. Analyze different models
      2. Assess with R2, RMSE, F-statistic, etc…
   2. Recursive Partitioning Decision Tree (rpart)
      1. Analyze different models
      2. Construct confusion matrix and ROC
   3. Naïve Bayes Classification
      1. Analyze different models
      2. Construct confusion matrix and ROC
6. Conclusion
   1. Comparison of models
      1. Compare MLR, rpart, and Naïve Bayes
   2. Comparison of other prior research
      1. Compare MLR, rpart, and Naïve Bayes results to prior research section
   3. Final model selection
      1. Determine best model
   4. Lesson learned
      1. Describe personal lessons learned from project
   5. Final Thoughts
      1. Provide final thoughts, where to go from this project
7. References