

## Homework No. 4 – Multiple Linear Regression

### Introduction

Regression is a tool that allows us to find the best *conditional relationship* of a predictor variable given one or more response variables. We can use regression to explore data, to test hypotheses about relationships between variables, and to develop prediction equations. In this assignment you will practice fitting a regression model to tree height and age data, and evaluating the results.

### Instructions

On the class website there is a file called “heightage.csv”, which contains height and dbh measurements for a sample of red and jack pine trees.

Part A. Once, for red pine, and then again, for jack pine, complete the following:

1. Find the linear regression coefficients for the model where height is the response and dbh is the predictor.
2. Is the regression a good one? Explain.
3. Generate a scatter plot of the data, and superimpose the regression line.
4. Find the predicted height for a 15-inch diameter tree.

Part B. Create a new column in your spreadsheet that will represent species and set the value to 1 for all jack pine trees, and 0 for all red pine trees. Then, complete the following:

1. Find the linear regression coefficients for the model where height is the response and there are two predictors, dbh and species code (0 or 1).
2. Is the regression a good one? Explain.
3. Generate a scatter plot of the data, and superimpose the regression line.
4. Find the predicted height for a 15-inch diameter tree of each species.
5. How is the regression model in Part B different from the two regression models in Part A? Explain.

### Product

Submit your answers in professional memo format. You must put your Lab Section on your memo and address the memo to your TA!

Please, if you're having trouble, ask for help! It's our job to provide help if you need it.

### Due Date

This assignment is due at the beginning of class on Monday February 23, 2009.