Lab 10 - Regression and Correlation Methods - STA 570

# Project Goals

In this lab assignment, you will be asked to:

* Interpret a scatterplot
* Determine whether a linear correlation is present between two variables
* Determine a regression model and interpret it.

# Sample Data Set

Data Set 1 “Body Data” from Appendix B includes body mass index (BMI) and high-density lipoprotein (HDL) cholesterol measurements from 300 subjects.

As always, open a new RMarkdown file. Make sure that your lab includes the code, output, and graphs, along with your write up to each question below.

# Analysis of Data

1. Generate a scatterplot between the HDL and LDL variables.

# Import your data set   
body <- read.table(file = "01 - Body Data.txt", sep = "\t", header = TRUE)

*#You may need this later. This code converts numeric variables to categorial*

body$GENDER..1.M. = as.factor(body$GENDER..1.M.)  
  
# Scatterplot of BMI vs. HDL  
plot(x = body$BMI, y = body$HDL)

1. Does there appear to be a linear association between BMI and HDL?
2. Does it appear that the relationship between BMI and HDL is linear? Weak or strong?
3. What is the equation of the regression line?
4. Interpret the slope and intercept in the context of HDL and BMI.

# Information from correlation and regression  
# Use Multiple R-squared for correlation  
summary(lm(HDL~BMI, data = body))

1. If the measurements are separated according to gender, do the scatterplots, correlation, or regression model change?

# Plot separating male and female with corresponding regression lines. 0 = female, 1 = male:  
library(ggplot2)

ggplot(body, aes(x=BMI, y=HDL, color = GENDER..1.M., lty=GENDER..1.M.)) + geom\_point() +

geom\_smooth(method="lm", se=FALSE)

# Regression model using gender as a covariate  
# 0 = female, 1 = male  
# Using HDL as the response variable  
summary(lm(HDL~BMI + body$GENDER..1.M., data = body))

1. Write a brief conclusion on how gender affects or does not affect HDL .