Lab 9: Linear Regression and Plots

The following worksheet is due by 8pm one day after this lab. You can find the submission dropbox in Brightspace by clicking on Content − > Lab Content.

0. Open a new R Markdown file.

Note: Your worksheet is to be submitted as the output of an R Markdown file (you can knit it to HTML and then convert it to PDF, or you can knit it to PDF if you have LaTeX on your computer, or you can knit it to Word and then convert that to a PDF).

1. Load the sales.csv dataset into R and save it to df.

(a) Define the relationship between month and sales in df. Please identify direction, form, and strength of the relationship.

(b) Perform a linear regression sales (y) and year (x1) and month(x2).

(c) Predict the sales in March 2010.

2. Use the built-in data set HairEyeColor to answer this question.

(a) Create a single table called hair_eye_totals which summarizes the total number of statistics students with each combination of hair and eye colour.

Note: The built-in data set consists of two tables with this information (one for women and one for men). The answer to part (a) is a single table combining the information from these two tables.

(b) Print out the hair_eye_totals table.

(c) Create a grouped bar plot which displays the information from the hair_eye_totals table. Your plot should include the following:

- a main title

- titles for the x-axis and y-axis

- colours to help differentiate the bars

- a legend to identify what each colour represents

(d) Create and print out a vector called percent_eye which contains the percent of statistics students with each eye colour (rounded to 2 decimal places). Show any additional code needed to create this vector.

(e) Create a pie chart displaying the information in the percent_eye vector. Your graph should include:

- a main title

- labels for each wedge displaying the eye colour

- a different colour for each eye colour

- the percentages displaying next to each wedge.

3. Use the mtcars data set in R to answer this question.

(a) Create a multi regression model for predicting miles per gallon (mpg) using weight (wt) and horsepower (hp).

(b) What is the interpretation of the coefficient estimates for the linear regression model part(a), which predicted miles per gallon (mpg) using weight (wt) and horsepower (hp) as predictor variables?