AMOD 5240H – Assignment 4

Nov. 13, 2018

Due: Thursday, Nov. 29 at 10 pm

This assignment is intended to cover the material of Chapter 5 and the beginning of Chapter 6 of the text (*Introductory Statistics with Randomization and Simulation*). It is due on or before Nov. 29 at 10 pm via electronic submission as a **pdf file** on Blackboard.

From the textbook, complete the following questions (8 points each)

1. Exercise 5.4
2. Exercise 5.8. Explain
3. Exercise 5.10
4. Exercise 5.22
5. Exercise 5.30
6. Exercise 5.32

In addition, complete the following (22 points each)

7. Is there a relationship between the number of stories a building has and its height? Some statisticians compiled data on a set of $n = 60$ buildings reported in the 1994 World Almanac. You will use the data set to decide whether height can be predicted from the number of stories.

   (a) Load the data from buildings.txt
   (Note that this is a text file, so use the appropriate instruction. If you are having trouble uploading the data, open it to see its contents and type the data in: one vector for heights and one vector for stories. Ignore the year data.)
   (b) Draw a scatterplot with stories in the x-axis and height in the y-axis. Does there seem to be a linear relationship between the two variables?
   (c) Find the linear correlation coefficient between these variables. What does it tell you about the linear relationship?
   (d) Obtain the linear model and summary. Write down the regression equation that relates height with stories. Add the line to the scatterplot.
   (e) Test for significance of the regression at $\alpha = 0.05$. State the null and alternative hypotheses. Can the model be used for predictions? Justify your conclusion using the summary in (d).
   (f) State the coefficient of determination. What percentage of variation in height is explained by the number of stories?
   (g) Draw diagnostic plots (a plot of stories vs. residuals, and a normal probability plot for the residuals). Do assumptions appear to be satisfied?

8. The **openintro** package contains a data set called **bdims**, which consists of the body dimensions of 507 physically active individuals. Complete a full multivariate regression analysis, predicting the variable wgt (weight) using all significant elements. You should do a stepwise variable selection procedure, and explore the data.

Formatting of answers: 8 points