STAT 509 Homework 7

Answer all questions in the order they are posed, and be sure to show your work to receive credit. Please **submit your work via one PDF file**. FYI, only **selected problems will be graded**.

1. Answer the following questions below:

   - a) What is the purpose of linear regression?
   - b) What assumptions are made in simple linear regression?
   - c) Write the equation of a simple linear regression model and explain each term.
   - d) Why is it important to examine the residuals of a regression model?

2. Answer the following questions below:

   - a) Suppose you have created a simple linear regression model predicting house prices based on square footage. How would you evaluate whether the model is a good fit for the data? What kind of diagnostic checks or plots would you use to assess the model's assumptions?
   - b) Consider a study that uses simple linear regression to predict the life expectancy of individuals based on their daily water intake. How would you approach collecting data for this study? What other variables might influence life expectancy that should be included in the analysis?
   - c) Simple linear regression makes several assumptions about the data. If a scatter plot of the data reveals a non-linear pattern, what does this suggest about the appropriateness of using simple linear regression? What alternative methods could you consider?

3. Explain the purpose of a confidence interval for the population mean in the context of simple linear regression. How does it differ from a prediction interval?

4. A researcher examined the controversial issue of the human vomeronasal organ, regarding its structure, function, and identity. The data in the table below shows the ***age of fetuses(x), in weeks, and length of crown-rum (y), in millimeters***. *Note: if x=10, then y=66 and so on.*

x	10, 10, 13, 13, 18, 19, 22
y	66, 66, 108, 106, 161, 166, 88

   Answer the following questions below:

- a) Find and interpret the linear correlation coefficient.
- b) Find and interpret the coefficient of determination.
- c) Find the least-squares estimates for the regression line. Also, interpret the value of the slope and intercept in the context of the problem.
- d) Predict y if x is 15. Is this value an overestimate or underestimate for y = 150?
- e) Construct a 90% confidence interval for the slope and interpret your interval. Compare and comment on the results obtained to that of hypothesis testing.
- f) Use R to compare your results from parts (c) and (e) above.

5. A recent graduate moving to a new job collected a sample of monthly rent (dollars) and size (square feet) of 2-bedroom apartments in one area of a Midwest city. Use the data below to answer the following questions below:

Size	Rent
1077	1075
1288	1270
842	760
1210	1175
820	845
945	885
1192	1035

- a) Is this simple or multiple linear regression? Explain.
- b) Find and interpret the intercept and slope in the regression model in context to this problem. In addition, state the units of the slope and intercept.
- c) What happens if the estimated line is used to predict rent when size is 2000? Why is this not a good idea?
- d) If the rent is actually 1000 (dollars), what is the value of the residual if size is 950? Is this regression model underestimating or overestimating?
- e) Construct a 95% confidence interval for the slope of the regression line, and interpret the resulting interval.
- f) Compute the 95% confidence interval for the population mean monthly rent when the size of the apartment is 950.
- g) State all the necessary assumptions needed for linear regression.