STAT 308 – Homework 2

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you will not be given full credit. Please upload your assignment by Thursday, September 15, 11:59 pm in a pdf file to Sakai.

- 1. In a simple linear regression problem where $n = 30$, we obtain

$$\sum_{i=1}^{n} x_i = 75, \sum_{i=1}^{n} y_i = 660, \sum_{i=1}^{n} x_i^2 = 240, \sum_{i=1}^{n} y_i^2 = 18000, \sum_{i=1}^{n} x_i y_i = -1200.$$

– a. Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$.

– b. Calculate SSE and $s^2_{Y|X}$.

- 2. Consider the dataset `AdRevenue.csv` on the course webpage. Suppose we are interested in modelling the ad revenue (in millions of dollars) of magazines based on the number of magazines in circulation (in millions).

– a. Draw a scatterplot of AdRevenue vs. Circulation. Comment on the four aspects of a scatterplot.

– b. Do you think a linear relationship between Circulation and AdRevenue is appropriate? Justify your response.

– c. Using `R`, find the equation of the least squares regression line.

– d. Add the least squares regression line to the scatterplot in (a).

– e. Interpret the slope of the regression line in the context of the given problem.

– f. Interpret the intercept of the regression line in the context of the given problem. Does this interpretation make sense? Why or why not?

– g. What do we expect the amount of ad revenue to be when there are 4 million magazines in circulation?

– h. Find the value of SSE and $s^2_{Y|X}$ for the least squares regression line.

– i. Determine if the assumption of homoscedasticity is violated.

– j. Determine if the assumption of normally distributed residuals is violated.