MATH 484/564 HOMEWORK #1

Due: September 12, Thursday

How to submit: submit in class

Problem 1 Derive intercept b_0 and slope b_1 for the simple linear regression model $\hat{y} = b_0 + b_1 x$ using the least square criterion.

Problem 2

1) Show that the regression line from the simple linear regression model passes through the point $(\bar{x}, \bar{y})$.

2) Show that $\sum_{i=1}^{n} (\hat{y_i} - \bar{y}) = 0$.

3) Show that $\sum_{i=1}^{n} (y_i - \hat{y_i}) = 0$.

Problem 3 Show that when the regression model becomes $y_i = \beta_1 x_i + \epsilon_i$, the estimate for β_1 becomes $b_{RTO} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$. (This is called regression through the origin (RTO).)

Problem 4 Use regression through the origin $\hat{y_i} = b_{RTO} x_i$ and ordinary linear regression $\hat{y_i} = b_0 + b_1 x_i$ to fit the same data, and report what is the value of $\sum_{i=1}^{n} e_i$ and r^2 for each type of regression. For this task, do not use the linear regression function $lm()$ from R. Use the formulas instead.

The dataset is generated from a model $y = 2x + \epsilon$, where ϵ follows normal distribution $\mathcal{N}(0, 0.1^2)$. x is distributed in the range of [-1, 1].

Problem 5 In the provided dataset **skincancer2.txt**, the response variable y is the mortality due to skin cancer (number of deaths per 10 million people) and the predictor variable x is the latitude. This is a modified dataset from the one used in class.

Inference on the slope of the data model $y = \beta_0 + \beta_1 x + \epsilon$ from the provided dataset to conclude whether there is linear relationship between y and x. The hypotheses are:

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

1) Report the point estimate b_1 for β_1, and the 95% confidence interval.

2) Report SST, SSR, SSE.

For these tasks, first use formulas you've learned to calculate the results, and then use R functions to verify them.

- use *lm()* to fit the data
- use *summary()* to get the regression results
- use *confint()* to provide the confidence interval
- use *plot()* to display a scatterplot of the data with the simple linear regression line