Homework 2

MATH 4387/5387 Fall 2020

Subrata Paul

8/24/2020

Problem 1

The linear regression model can be written in matrix notation as $y = X\beta + \epsilon$. Create the table shown below and describe what each term (y, X, β, ϵ) represents (interpretation), specify the dimension of each term (size), indicate whether we model the term as random or non-random, and whether the term is observed or unobserved.

Term	Size	Interpretation	Random?	Observable?
y				
β				
X				
ϵ				

Problem 2

Assuming a simple linear regression model, derive the ordinary least squares estimators of β_0 and β_1. Do not use matrix notation in deriving your solution.

Problem 3

Let $H = X(X^TX)^{-1}X^T$ is the hat matrix. Prove that $I - H$ is a projection matrix (Symmetric + Idempotent).

Problem 3

While proving that $\hat{\beta}_1$ is an unbiased estimator of β_1, we represented the OLS estimate as $\hat{\beta}_1 = \sum k_i Y_i$, where $k_i = \frac{X_i - \overline{X}}{\sum (X_i - \overline{X})^2}$. Use it with the properties of k_i, that we already have proved, to derive the variance of $\hat{\beta}_1$. (Hint: in linear regression framework we assume $Var(\epsilon_i) = \sigma^2$ and $Cov(\epsilon_i, \epsilon_j) = 0$ when $i \neq j$)

Problem 4

Under simple linear regression model, the Mean Squared Error (MSE) is defined as $\frac{\sum (y_i - \hat{y}_i)^2}{n-2}$ where n is the number of observations. MSE is an unbiased estimator of σ^2, where σ^2 is the variance of ϵ_i. What is an unbiased estimator of the variance of $\hat{\beta}_1$?

Problem 5

The square root of the variance of an estimator is the standard error (SE). You can derive the SE $(\hat{\beta})$ from problem 4. According to theory,

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2}$$

where t_{n-2} represents a Student's t distribution with $n-2$ degrees of freedom. Find an expression for 95% confidence interval of β_1.

Problem 6

If $\hat{\beta}_1 = 2$, $SE(\hat{\beta}_1) = 0.02$, and $n = 50$ calculate 95% confidence interval for β_1.

Problem 7

Based on the confidence interval on problem 6, perform the hypothesis test,

$$H_0 : \ \beta_1 = 0 \quad \text{Vs.} \ H_1 : \ \beta_1 \neq 0$$

Problem 8

Use the `simu_hw1.txt` data and fit a multiple linear regression model with `response` as the response variable and `pred1`, `pred2`, and `pred3` as predictors. Write down the equation of the fitted line (fitted model).