

## PSTAT126 Homework 1: Due 1/26/2025

1. (3 pt) When asked to state the simple linear regression model, a student wrote it as follows:

$$E(Y_i) = \beta_0 + \beta_1 X_i + \epsilon_i.$$

Do you agree? Explain.

2. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- (a) (3 pt) Assume that  $X = 0$  is within the scope of the model. What is the implication for the regression function if  $\beta_0 = 0$ ? How would the regression function plot on a graph?
- (b) (4 pt) Under the assumption of  $\beta_0 = 0$ , derive the least squares estimate of  $\beta_1$ ?
- (c) (3 pt) How do you fit such a model using the *lm* function?

3. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- (a) (3 pt) What is the implication for the regression function if  $\beta_1 = 0$ ? How would the regression function plot on a graph?
  - (b) (4 pt) Under the assumption of  $\beta_1 = 0$ , derive the least squares estimate of  $\beta_0$ ?
  - (c) (3 pt) How do you fit such a model using the *lm* function?
4. (3 pt) The summation of residual equals zero for the simple linear model. Does that imply the summation of random errors in the model equals zero? Does the expectation of the summation of random errors equal zero? Comment.
5. (15 pt) Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Show that minimizing the sum of squared residuals lead to the following least squares coefficient estimates:

$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ b_0 &= \bar{Y} - b_1 \bar{X} \end{aligned}$$

where  $\bar{X} = \sum_{i=1}^n X_i/n$  and  $\bar{Y} = \sum_{i=1}^n Y_i/n$ . Show that  $b_0$  has minimum variance among all unbiased linear estimators.

6. **GPA.** The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year can be predicted from the ACT test score. All data are available in the "Homework Assignments and Dataests" section at Gauchospace.

- (a) (3 pt) In this problem, what is the predictor variable, and what is the response variable?
  - (b) (3 pt) Construct a scatter plot of the predictor and response variables. Comment on your impression of the relationship.
  - (c) (3 pt) Fit a simple linear regression model using *lm* function and report the estimated regression function.
  - (d) (3 pt) Add the fitted line to the scatter plot in (b). Does the estimated regression function appear to fit the data well?
  - (e) (3 pt) Obtain a point estimate of the mean freshman GPA for students with ACT test score equals 30.
  - (f) (3 pt) What is the point estimate of the change in the mean response when the entrance test score increases by one point?
  - (g) (3 pt) Obtain the residuals and check whether they sum to zero.
  - (h) (3 pt) Estimate  $\sigma^2$  and  $\sigma$ . What are the units of them?
7. **Airfreight breakage.** A substance used in biological and medical research is shipped by airfreight to users in cartons of 1,000 ampules. Data were collected on the number of times the carton was transferred from one aircraft to another over the shipment route and the number of ampules found to be broken upon arrival.
- (a) (6 pt) Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?
  - (b) (3 pt) Obtain a point estimate of the expected number of broken ampules when 1 transfer is made.
  - (c) (3 pt) Estimate the increase in the expected number of ampules broken when there are 2 transfers as compared to 1 transfer.
  - (d) (3 pt) Verify that your fitted regression line goes through the point  $(\bar{X}, \bar{Y})$ .