

# Quiz 1: Linear Regression

Name: \_\_\_\_\_

**Instructions:** Answer all questions. For multiple choice questions, select the single best answer. For short answer questions, provide concise responses (1-3 sentences).

**Total Points:** 16 (Part I: 8 points, Part II: 8 points)

---

## Part I: Multiple Choice (1 point each)

**Question 1:** In the simple linear regression model  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ , which component represents the random error term?

- a.  $Y_i$
- b.  $\beta_0$
- c.  $\beta_1$
- d.  $\varepsilon_i$

**Question 2:** What is the possible range of Pearson's correlation coefficient  $r$ ?

- a.  $0 \leq r \leq 1$
- b.  $-1 \leq r \leq 1$
- c.  $-\infty < r < \infty$
- d.  $0 < r < \infty$

**Question 3:** A high correlation between variables  $X$  and  $Y$  indicates that:

- a.  $X$  causes  $Y$
- b.  $Y$  causes  $X$
- c. There is a strong linear association between  $X$  and  $Y$
- d. There is no relationship between  $X$  and  $Y$

**Question 4:** The residual for observation  $i$  in a regression model is defined as:

- a.  $\hat{Y}_i - Y_i$
- b.  $Y_i - \hat{Y}_i$
- c.  $Y_i - \bar{Y}$

d.  $\hat{Y}_i - \bar{Y}$

**Question 5:** In the simple linear regression model, which of the following is assumed to be normally distributed?

- a. The predictor variable  $X$
- b. The response variable  $Y$
- c. The error terms  $\varepsilon_i$
- d. The regression coefficients  $\beta_0$  and  $\beta_1$

**Question 6:** Log transformations are commonly used in regression to address:

- a. Missing data
- b. Multicollinearity between predictors
- c. Non-constant variance (heteroscedasticity) and/or nonlinearity
- d. Small sample sizes

**Question 7:** Homoscedasticity in regression means:

- a. The errors have different variances across all levels of  $X$
- b. The errors have constant variance across all levels of  $X$
- c. The predictors are uncorrelated
- d. The response variable is normally distributed

**Question 8:** The primary purpose of  $k$ -fold cross-validation is to:

- a. Estimate the regression coefficients
- b. Test for multicollinearity
- c. Assess the model's predictive performance on unseen data
- d. Transform non-normal variables

## Part II: Short Answer (2 points each)

**Question 9:** In one sentence, explain what it means if a regression model has  $R^2 = 0.75$ .

**Question 10:** What does a funnel-shaped pattern in a residual vs. fitted values plot suggest about the regression model?

**Question 11:** Give one example of a situation when linear regression would NOT be appropriate for modeling a relationship.

**Question 12:** Why should we generally avoid fitting very high-order polynomial regression models (e.g., degree 10 or higher)?