

Statistical Methods - Quiz #1(45 minutes)

September 15, 2025 (Monday)

Section(교반): __A1__ Cadet Number(교번): _____ Name(성명): _____ Score: _____

- All solutions must include a detailed step-by-step explanation.

1. Consider the following linear regression model.[50 points]

$$y_i = 1 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \text{ independently for } i = 1, \dots, n.$$

(1) Find the least squares estimate of β_1 , denoted $\hat{\beta}_1$, that minimizes

$$S(\beta_1) = \sum_{i=1}^n (y_i - 1 - \beta_1 x_i)^2.$$

(2) Let the fitted values be $\hat{y}_i = 1 + \hat{\beta}_1 x_i$. Using the result in (A), show that

$$\sum_{i=1}^n (y_i - 1)^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - 1)^2.$$

(3) Find $E(\hat{\beta}_1)$.

2. Suppose that we have the two multiple linear regression models.

$$\text{Model A: } y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

$$\text{Model B: } y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

for $i = 1, \dots, n$, where $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. Let $\hat{y}_i^{(A)}$ and $\hat{y}_i^{(B)}$ be the fitted values for Model A and Model B, respectively. Compare each of the following quantities. (Hint: use projection.) [30 points]

(1) $\sum_{i=1}^n (y_i - \hat{y}_i^{(A)})^2$ and $\sum_{i=1}^n (y_i - \hat{y}_i^{(B)})^2$

(2) $\sum_{i=1}^n (\hat{y}_i^{(A)} - \bar{y})^2$ and $\sum_{i=1}^n (\hat{y}_i^{(B)} - \bar{y})^2$

(3) R^2 for Model A (= R_A^2) and R^2 for Model B (= R_B^2)

3. Consider the regression of *heart rate after exercise* (y_i) on categorical *exercise type* (with three levels: Running, Cycling, Swimming) using two indicator variables x_{i1} and x_{i2} . [20 points]

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ independently. x_{i1} and x_{i2} are defined as:

$$x_{i1} = \begin{cases} 1 & \text{if participant } i \text{ did Running,} \\ 0 & \text{otherwise,} \end{cases} \quad x_{i2} = \begin{cases} 1 & \text{if participant } i \text{ did Cycling,} \\ 0 & \text{otherwise.} \end{cases}$$

Subject No. i	1	2	3	4	5	...	n
Exercise type	Swimming	Swimming	Cycling	Cycling	Running	...	Running
Heart rate y_i (bpm)	118	125	131	136	145	...	138

(1) Express the mean of *heart rate* after cycling in terms of $\beta_0, \beta_1, \beta_2$.

(2) Express the mean difference in *heart rate* between Running and Swimming (*heart rate after Running* - *heart rate after Swimming*) in terms of $\beta_0, \beta_1, \beta_2$.