

STAT 8260 Exam 2 – Tuesday, April 10
SHOW ALL WORK

Name: _____

1. Suppose we take a random sample of $n = 6$ adult father-son pairs and we measure the height of each man. Suppose that 3 of the sons in this sample were first-born children in their family and the other 3 happened to be second-born. Let (y_{ij}, x_{ij}) be the (son's height, father height) pair of measurements corresponding to the j th son of the i th birth order, $i = 1, 2$, $j = 1, 2, 3$. Consider the following five models for such data.

$$(i) \quad y_{ij} = \alpha_i + e_{ij}$$

$$(ii) \quad y_{ij} = \alpha + \beta x_{ij} + e_{ij}$$

$$(iii) \quad y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}$$

$$(iv) \quad y_{ij} = \mu + \alpha I(i = 2) + \beta x_{ij} + e_{ij}$$

$$(v) \quad y_{ij} = \mu + \alpha I(i = 2) + \beta x_{ij} I(i = 1) + \gamma x_{ij} I(i = 2) + e_{ij}$$

where $I(\cdot)$ denotes an indicator variable taking the value 1 when the condition inside the parentheses is true and the value 0 otherwise. All models have the same assumptions on the e_{ij} s: $e_{11}, \dots, e_{23} \stackrel{iid}{\sim} N(0, \sigma^2)$.

- a. **(8 pts.)** Which, if any, of these models are equivalent.

b. **(9 pts.)** Write down model (iii) in vector/matrix notation.

c. **(9 pts.)** Under the maintained hypothesis that model (v) holds, express the hypothesis that $H_0 : \{\text{model (ii) holds}\}$, in the form of the general linear hypothesis on the parameters of model (v).

- d. **(9 pts.)** Based upon the results in the table below, compute an appropriate test statistic (give me its numeric value for these data) and give its reference distribution for testing the hypothesis in part (c).

Model	MSE	R^2
(ii)	9.9268	.0432
(v)	3.2262	.8445

2. Consider the regression model

$$\begin{aligned} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} &= \beta \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \end{pmatrix} \\ &= \beta \mathbf{j}_4 + \mathbf{e} \end{aligned}$$

where $\begin{pmatrix} e_{11} \\ e_{12} \end{pmatrix} \sim N\left(\mathbf{0}, \sigma_1^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ independent of $\begin{pmatrix} e_{21} \\ e_{22} \end{pmatrix} \sim N(\mathbf{0}, \sigma_2^2 \mathbf{I})$ where $\sigma_2^2 = 2\sigma_1^2$ and $\rho = .5$.

- a. **(13 pts.)** Derive a simple, non-matrix formula for $\hat{\beta}_{\text{BLUE}}$, the best linear unbiased estimator of β in this model.

b. (13 pts.) Assuming the model is correct, compute the ratio

$$\frac{\text{var}(\bar{y}_{..})}{\text{var}(\hat{\beta}_{\text{BLUE}})},$$

where $\bar{y}_{..} = \frac{1}{4} \sum_{i=1}^2 \sum_{j=1}^2 y_{ij}$ is the sample mean.

3. Consider the classical linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

where \mathbf{X} is $n \times (k + 1)$ with $\text{rank}(\mathbf{X}) = k + 1 < n$. Suppose this model is fit to data \mathbf{y} yielding the ordinary least squares regression parameter estimate $\hat{\boldsymbol{\beta}}$ and MSE, s^2 . We wish to form a prediction interval for an additional observation y_0 conforming to this model. That is, $y_0 = \mathbf{x}_0^T \boldsymbol{\beta} + e_0$ where \mathbf{x}_0 is the a vector of explanatory variables corresponding to y_0 (\mathbf{x}_0^T can be considered as an additional row of \mathbf{X}) and $e_0 \sim N(0, \sigma^2)$ independent of \mathbf{e} .

(13 pts.) For $\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}}$, show that

$$\frac{y_0 - \hat{y}_0}{s \sqrt{1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0}} \sim t(n - k - 1).$$

4. Consider the classical linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

where \mathbf{X} is $n \times p$ with $\text{rank}(\mathbf{X}) = p < n$. Define $\tilde{\sigma}_\ell^2 = SSE/(n - \ell)$, an estimator of σ^2 . Recall the mean square error of an estimator T of a parameter θ is $E\{(T - \theta)^2\}$ but can also be represented as variance + bias².

- a. **(13 pts.)** Find the mean square error of $\tilde{\sigma}_\ell^2$, for an arbitrary value of the constant ℓ .

- b. **(13 pts.)** Find the value of ℓ that yields the estimator $\tilde{\sigma}_\ell^2$ with the smallest mean square error.