Linear Statistical Models — Quiz 04 — Full Name ______________________

1. Researchers conducted a blind taste test to compare coffee from different regions. Here is the dataset and the design matrix for the one factor model:

i	region	score	$j(i)$
1	Kenya	7	3
2	Kenya	9	3
3	Costa Rica	5	1
4	Costa Rica	7	1
5	Ethiopia	9	2
6	Ethiopia	9	2

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

(i) (3) Fill in the $j(i)$ column, using R's convention to order the levels of region aphabetically.

(ii) (3) Write down the one factor model for the scores, treating region as the factor. (define)

$$Y_i = b_0 + b_{j(i)} + \varepsilon_i$$

$\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$, b_0, b_1, b_2, b_3 unknown parameters

(iii) (3) In terms of the model parameters, write the expected scores for the three regions.

Kenya: ~~b_0+b_1~~ b_0+b_3

Costa Rica: ~~b_0~~ b_0+b_1

Ethiopia: b_0+b_2

(iv) (3) What is the rank of the design matrix and why?

(no derivation required, just give the number, and a brief explanation)

3. 4 columns. Last 3 add up to first

2. Consider $\widehat{\boldsymbol{b}} = \begin{bmatrix} 6 & 0 & 3 & 2 \end{bmatrix}^T$

(i) (3) Calculate $X^T X$ and $X^T \boldsymbol{y}$ (y = scores)

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 7 \\ 9 \\ 5 \\ 7 \\ 9 \\ 9 \end{bmatrix} = \begin{bmatrix} 16 + 12 + 18 \\ 12 \\ 18 \\ 16 \end{bmatrix} = \begin{bmatrix} 46 \\ 12 \\ 18 \\ 16 \end{bmatrix}$$

(ii) (3) Show that $\widehat{\boldsymbol{b}}$ is a solution to the normal equations $X^T X \widehat{\boldsymbol{b}} = X^T \boldsymbol{y}$

$$\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 36 + 6 + 4 \\ 12 \\ 12 + 6 \\ 12 + 4 \end{bmatrix} = \begin{bmatrix} 46 \\ 12 \\ 18 \\ 16 \end{bmatrix}$$

(iii) (3) Calculate the fitted values (Hint: you can check your work against the data!)

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 6 \\ 6 \\ 9 \\ 9 \end{bmatrix}$$

(iv) (3) Calculate the residuals and the residual sum of squares

$$\hat{\underline{e}} = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad rss = 4$$

(v) (3) Calculate $\hat{\sigma}^2$. You may leave your answer as a fraction.

$$\hat{\sigma}^2 = \frac{rss}{df} = \frac{4}{3}$$

(vi) (3) For this solution $\widehat{\boldsymbol{b}}$, how do you interpret the following quantities?

$\widehat{b}_0$: estimated EV for Costa Rica

$\widehat{b}_2$: estimated EV for Ethiopia - est. EV for Costa Rica

$\widehat{b}_3$: estimated EV for Kenya - estimated EV for Costa Rica.

3. (3 Bonus) What is the estimated standard error of $\widehat{B}_2$? (use the back)

$$X_0 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \qquad X_0^T X_0 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

$$\cancel{\left(2\begin{bmatrix} 6 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}\right)^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}}$$

$$\cancel{\hat{SE}(\hat{B}_2) = \sqrt{\frac{4}{3} \cdot 1} = \frac{2}{\sqrt{3}}}$$

$$(X_0^T X_0)^{-1} = \begin{bmatrix} 6 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}^{-1} = \left(2\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\right)^{-1}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

$$\hat{Var}(\hat{B}_2) = \frac{4}{3} \cdot \frac{1}{2} \cdot 2 = \frac{4}{3}$$

$$\hat{SE}(\hat{B}_2) = \frac{2}{\sqrt{3}}$$