

1. Different levels of an amino acid (0, 1, 3, or 6 mg/kg) were added to the rations fed to dairy cows. Ten different cows were used in the study. The objective was to examine the possible effect of the amino acid on milk production. Four cows were randomly assigned to the control group (0 mg/kg of the amino acid) and two cows were randomly assigned to each of the other three levels of the amino acid. Aside from the level of this amino acid, the rations fed to the ten cows were identical. The cows were kept in separate pens, so you can assume that one cow responded independently of any other cow. Each cow was put on the assigned diet immediately after giving birth to a calf. The response was the amount of milk produced by the cow during the next 200 days. The data are displayed in the following table.

Observed Milk Production	Level of Amino Acid (mg/kg)
$Y_{11}$	$X_1 = 0$
$Y_{12}$	$X_1 = 0$
$Y_{13}$	$X_1 = 0$
$Y_{14}$	$X_1 = 0$
$Y_{21}$	$X_2 = 1$
$Y_{22}$	$X_2 = 1$
$Y_{31}$	$X_3 = 3$
$Y_{32}$	$X_3 = 3$
$Y_{41}$	$X_4 = 6$
$Y_{42}$	$X_4 = 6$

Consider the following linear models:

**Model A:**  $Y_{ij} = \beta_0 + \beta_1(X_i - 2) + \epsilon_{ij}$ , where  $\epsilon_{ij} \sim NID(0, \sigma^2)$

**Model B:**  $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ , where  $\epsilon_{ij} \sim NID(0, \sigma^2)$

- (a) State the definition of an estimable function of parameters for a linear model. Use the definition to determine if either of the following is estimable for **model B**. (In each case give an explanation. A simple yes or no answer is not sufficient)

**Definition:**

i.  $\alpha_2 + \alpha_3$

ii.  $2\mu + 4\alpha_1 - \alpha_2 - \alpha_3$

(b) **Model B** can be written as  $\mathbf{Y} = W\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ , where

$$W = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix}$$

Compute a generalized inverse for  $W^T W$  and use it to obtain a solution to the normal equations. Express your solution as a function of  $\bar{Y}_1$ ,  $\bar{Y}_2$ ,  $\bar{Y}_3$ ,  $\bar{Y}_4$ , the sample means for milk production at the four levels of amino acid.

(c) Verify that  $3\alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$  is an estimable function of the parameters in **model B**. Present a formula for the ordinary least squares estimator for  $3\alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$  as a function of  $\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \bar{Y}_{3\cdot}, \bar{Y}_{4\cdot}$ .

(d) State the Gauss-Markov theorem and describe what it implies about the properties of the ordinary least squares estimator for  $3\alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$  that you presented in part (d).

(e) Write **model A** as  $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -2 \\ 1 & -2 \\ 1 & -2 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \\ 1 & 4 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

and define  $P_X = X (X^T X)^{-1} X^T$ . Also define  $P_W = W (W^T W)^{-1} W^T$  for **Model B**.

i. Show that

$$\frac{1}{\sigma^2} \mathbf{Y}^T (I - P_X) \mathbf{Y}$$

has a chi-square distribution under **model B**, i.e.,  $\mathbf{Y} \sim \mathbf{N}(\mathbf{W}\boldsymbol{\alpha}, \sigma^2 \mathbf{I})$ . Report degrees of freedom. Does it have a central chi-square distribution?

ii. Show that

$$\frac{1}{\sigma^2} \mathbf{Y}^T (I - P_X) P_W (I - P_X) \mathbf{Y}$$

has a chi-square distribution under **model B**. Report degrees of freedom. Does it have a central chi-square distribution?

(f) Using the notation and results presented in Part (f), show that

$$F = \frac{c\mathbf{Y}^T(I - P_X)P_W(I - P_X)\mathbf{Y}}{\mathbf{Y}^T(I - P_W)\mathbf{Y}}$$

has an  $F$ -distribution for some constant  $c$  when **model B** is the correct model. Report a numerical value for  $c$  and degrees of freedom, and give a formula for the non-centrality parameter of the  $F$ -distribution.

(g) Can the  $F$ -statistic defined in Part (g) be used to test a null hypothesis of practical importance? Explain.

2. Consider the study in Problem 1 and suppose the researchers also had information on the age of each cow. Let  $Z_{ij}$  denote the age of the  $j$ -th cow fed the  $i$ -th level of amino acid. The following model was proposed.

$$\mathbf{Model\ C: } \mathbf{Y} = M\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

and

$$M\boldsymbol{\gamma} = \begin{bmatrix} 1 & -2 & Z_{11} & 0 & 0 & 0 \\ 1 & -2 & Z_{12} & 0 & 0 & 0 \\ 1 & -2 & Z_{13} & 0 & 0 & 0 \\ 1 & -2 & Z_{14} & 0 & 0 & 0 \\ 1 & -1 & 0 & Z_{21} & 0 & 0 \\ 1 & -1 & 0 & Z_{22} & 0 & 0 \\ 1 & 1 & 0 & 0 & Z_{31} & 0 \\ 1 & 1 & 0 & 0 & Z_{32} & 0 \\ 1 & 4 & 0 & 0 & 0 & Z_{41} \\ 1 & 4 & 0 & 0 & 0 & Z_{42} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{bmatrix}$$

(a) Under what conditions, if any, would  $\gamma_2$  be estimable in **model C**?

(b) Show how you would test the null hypothesis that the linear effect of age on average milk production is the same at every level of amino acid in the diet, i.e., show how to test the null hypothesis

$$H_0 : \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 \quad \text{against} \quad H_A : H_0 \text{ is false.}$$

Report a formula for your test statistic, degrees of freedom, and indicate how you would reach a conclusion. (Just state the formula for your test statistic, you do not have to prove anything about the distribution of your test statistic or show any derivation.)

(c) Suppose **model C** is correct. Show how you would construct a 95% confidence interval for  $\sigma^2$ .

- (d) Is **model C** a reparameterization of **model B** in Problem 1? Give an explanation. A simple "yes" or "no" response is not sufficient.

3. Consider the linear model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where  $X$  is an  $n \times k$  matrix with  $\text{rank}(X) = k$  and  $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 I + \delta^2 \mathbf{1}\mathbf{1}^T$ , where  $I$  is the  $n \times n$  identity matrix and  $\mathbf{1}$  is an  $n \times 1$  vector of ones. For this model, each observation has the same variance  $\sigma^2 + \delta^2$ , but the observations are correlated. Any pair of observations has correlation  $\delta^2/(\sigma^2 + \delta^2)$ . Furthermore, since  $X$  has full column rank, every element of  $\boldsymbol{\beta}$  is estimable and  $\mathbf{c}^T \boldsymbol{\beta}$  is estimable for any  $\mathbf{c} \neq \mathbf{0}$ . The ordinary least squares estimator for  $\mathbf{c}^T \boldsymbol{\beta}$  is

$$\mathbf{c}^T \mathbf{b} = \mathbf{c}^T (X^T X)^{-1} X^T \mathbf{Y}.$$

- (a) Show that  $\mathbf{c}^T \mathbf{b} = \mathbf{c}^T (X^T X)^{-1} X^T \mathbf{Y}$  is an unbiased estimator for  $\mathbf{c}^T \boldsymbol{\beta}$ .

- (b) Is  $\mathbf{c}^T \mathbf{b} = \mathbf{c}^T (X^T X)^{-1} X^T \mathbf{Y}$  a best linear unbiased estimator for  $\mathbf{c}^T \boldsymbol{\beta}$ ? Justify your answer.

Your Score: \_\_\_\_/100