

Theory of Linear Models

Exam Question

April 2026

The following data on a response variable Y were collected from a small observational study with explanatory variables X_1 , X_2 , X_3 and X_4 .

X_1	X_2	X_3	X_4	Y
5.3	0	65	11.8	22.5
6.2	0	46	10.8	19.4
5.8	1	23	8.1	24.6
12.2	1	77	19.9	24.1
10.0	0	9	10.9	34.3
11.1	0	101	21.2	36.0
8.3	1	45	12.8	31.9
5.4	1	88	14.2	38.4

The proposed model is

$$E(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i},$$

with Y_i s uncorrelated and $V(Y_i) = \sigma^2$.

1. Express the model in matrix form, i.e. write down the design matrix $\mathbf{X}$ and the vector of parameters β .
2. Find the rank of the matrix $\mathbf{X}$. You can do this by hand or by computer.
3. Explain how to obtain a generalized inverse of $\mathbf{X}'\mathbf{X}$. You may use a computer for numerical calculations but must explain your method..
4. Using your generalized inverse, compute a least squares estimator. Clearly describe the steps taken.

5. Show that β_1 is not estimable. [Note that if you use a computer, zero vectors might not be identically zero.]
6. Show that $\theta = \beta_1 + 2\beta_2 + \beta_4$ is estimable and find its least squares estimate $\hat{\theta}$. [Note that if you use a computer, zero vectors might not be identically zero.]
7. (Slightly harder, but still by hand.) The columns of $\mathbf{X}$ are linearly dependent. Find a non-zero vector $a \in \mathbb{R}^5$ such that $\mathbf{X}a = \mathbf{0}$ and use it to reparameterize the model in terms of four identifiable parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ (give an explicit model for $E(Y_i)$ in terms of the γ 's). Express the corresponding least squares estimates in terms of $\hat{\beta}$ and evaluate them numerically.