

Instructions: This is a closed book exam. No books or notes are allowed. We have not provided enough space for your answers on this exam. **Please write your answers on other sheets of paper.** Be sure to write your name on every sheet of paper that you submit.

If you do not remember a definition or result, or you do not recall exactly how to complete an argument or derivation, do not give up. Write down what you do know and describe how you would proceed if you could complete the missing step.

You may present formulas in answers using matrix notation. Be sure to define any notation you use that is not defined on the exam.

1. When gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. A company has developed a device that can be installed in the gas tank of a car to prevent vapors from escaping when the gas tank is filled. A small study was performed to examine the effectiveness of this device. Four cars were used in the study, and the device was installed in the gas tank of two of the cars. Gasoline was pumped into the tank of each car and the amount of gas vapor that escaped ( $Y$ ) was measured. Since the temperature of the gasoline ( $X_1$ ) can affect the outcome, two gasoline temperatures were used. The data are shown below. In this table  $X_2 = 1$  if the device was used and  $X_2 = -1$  if the device was not installed in the gas tank.

Amount of vapor that escapes $Y$	Gasoline temperature ( $^{\circ}\text{C}$ ) $X_1$	Use of device $X_2$
$Y_1$	0	-1
$Y_2$	30	-1
$Y_3$	0	1
$Y_4$	30	1

Consider the model

$$Y_i = \beta_1(X_{1i}) + \beta_2(X_{2i}) + \epsilon_i$$

where  $\epsilon_i$  is a random error with  $E(\epsilon_i) = 0$  for  $i = 1, 2, 3, 4$ .

Answer the following questions.

- (a) Under what conditions would this be a normal theory Gauss-Markov model?
- (b) Interpret  $\beta_2$  in this model? What does it represent in the context of the gasoline vapor study?
- (c) Let SSE denote the sum of squared residuals for this model. Assuming the conditions you listed in Part (a) are satisfied, what is the distribution of SSE? (No derivation is necessary, just state your answer.)
- (d) Let  $\tilde{b}$  be a solution to the normal equations. What are the properties of  $\tilde{b}$ ? Assume that the conditions you listed in Part (a) are satisfied.
- (e) To test for a device effect, the researchers propose the following test statistic

$$F = \frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{2SSE}$$

Assuming that the assumption you listed in Part (a) are satisfied, show that this statistic has an F-distribution. Report degrees of freedom.

- (f) With respect to  $\beta = (\beta_1 \beta_2)^T$ , describe the null hypothesis that can be tested with the F-test in Part (e). What is the alternative hypothesis?
- (g) The researchers want to use other test statistics of the form

$$F = \frac{2 \left( \sum_{i=1}^4 a_i Y_i \right)^2}{\left( \sum_{i=1}^4 a_i^2 \right) SSE} = \frac{2 \begin{pmatrix} \tilde{a}^T & \tilde{Y} \\ \tilde{a} & \tilde{a} \end{pmatrix}}{\begin{pmatrix} \tilde{a}^T & \tilde{a} \\ \tilde{a} & \tilde{a} \end{pmatrix} SSE} .$$

Note that the test statistic in Part (e) is a special case of this test statistic. Does this test statistic have an F-distribution for any vector of constants

$\tilde{a} = (a_1, a_2, a_3, a_4)^T$ ? Explain. If your answer is no, describe the set of  $\tilde{a}$  vectors for which this test statistic has an F-distribution.

2. Two varieties of corn (variety A and variety B) were compared in a field trail. In addition to the varieties, three levels of nitrogen were used (100, 150, and 200 pounds per acre (lb/a)). Six different fields were used, and the six combinations of varieties and nitrogen levels were randomly assigned to the fields. Let  $Y_{ij}$  denote the yield (in bushels per acre) of the  $i$ -th variety of corn when the  $j$ -th level of nitrogen is applied. Throughout this question,  $\varepsilon_{ij}$ ,  $i = 1, 2$ ,  $j = 1, 2, 3$ , denote independent  $N(0, \sigma^2)$  random variables where  $\sigma^2$  is an unknown variance.

The following two models were proposed:

$$\text{Model 1: } \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -50 & 2500 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 50 & 2500 \\ 0 & 1 & -50 & 2500 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 50 & 2500 \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{11} \\ \mathbf{e}_{12} \\ \mathbf{e}_{13} \\ \mathbf{e}_{21} \\ \mathbf{e}_{22} \\ \mathbf{e}_{23} \end{bmatrix}$$

$$\text{Model 2: } \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 0 & -2 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 0 & -2 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{m} \\ a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{11} \\ \mathbf{e}_{12} \\ \mathbf{e}_{13} \\ \mathbf{e}_{21} \\ \mathbf{e}_{22} \\ \mathbf{e}_{23} \end{bmatrix}$$

In model 1, note that the third column of the model matrix are the values of the nitrogen level minus 150 lb/a, and the fourth column contains the squares of the elements in the third column.

- (a) With respect to the effects of varieties and nitrogen levels on corn yields, interpret the parameters  $\gamma_1$  and  $\delta_1$  in Model 1.
- (b) For Model 1, indicate which of the following quantities are estimable
- $\gamma_1 - \gamma_2$
  - $\gamma_1 - 10\delta_1 + 100\delta_2$

Give a brief explanation, to support your conclusions.

- (c) For Model 2, determine if  $\mu + \alpha_1$  is estimable?  
Give a brief explanation to support your conclusion.
- (d) Expressing Model 2 as  $\tilde{Y} = X \tilde{\mathbf{b}} + \tilde{\mathbf{e}}$ , a solution to the normal equations is  $\tilde{\mathbf{b}} = (X^T X)^- X^T \tilde{Y}$ . Explain how a generalized inverse  $(X^T X)^-$  can be computed. (You are not expected to obtain a numerical value for  $(X^T X)^-$ , just briefly outline a procedure for how it can be computed.)
- (e) Using  $\tilde{\mathbf{b}} = (X^T X)^- X^T \tilde{Y}$  from Part (d), define the estimator

$$\hat{\mathbf{a}}_1 - \hat{\mathbf{a}}_2 = [0 \ 1 \ -1 \ 0 \ 0] \tilde{\mathbf{b}}.$$

What are the properties of this estimator?

- (f) Would the residual sum of squares from fitting models (1) and (2) be the same? Explain.
- (g) Using  $\tilde{Y} = W \tilde{\mathbf{g}} + \tilde{\mathbf{e}}$  to express Model 1 in matrix notation, show how to construct a 95% prediction interval for the corn yield that would be observed if variety 1 is grown with a nitrogen application of 120 lb/acre.
- (h) With respect to Model 1, show how to test the null hypothesis

$$H_0: \gamma_1 = \gamma_2 \text{ and } \delta_2 = 0.$$

against the alternative that  $H_0$  is not true. This null hypothesis is equivalent to the statement that the same straight line model for the effect of nitrogen on mean corn yield is appropriate for both varieties. Give a formula for your test statistic and explain how to use it. (Just state the formula, you do not have to prove anything or show any derivations.)