STAT 511 **EXAM I** **NAME ________________**
Spring 2001

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

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

   A. What is the minimal set of conditions for a model to be a Gauss-Markov model?

   B. 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).

   (i) $\alpha_2 + \alpha_3$

   (ii) $2\mu + 4\alpha_1 - \alpha_2 - \alpha_3$

C. Model B can be written as $\underset{\sim}{Y} = W\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}$, where

$$W = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\alpha} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

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 $\overline{Y}_{1.}$, $\overline{Y}_{2.}$, $\overline{Y}_{3.}$, $\overline{Y}_{4.}$, the sample means for milk production at the four levels of amino acid.

D. 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 $\overline{Y}_{1.}$, $\overline{Y}_{2.}$, $\overline{Y}_{3.}$, $\overline{Y}_{4.}$.

E. 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.

F. Write model A as $\underset{\sim}{Y} = X\underset{\sim}{b} + \underset{\sim}{e}$, where

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -2 \\ 1 & -2 \\ 1 & -2 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \\ 1 & 4 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

and define $P_X = X\left(X^T X\right)^{-1} X^T$. Also define $P_W = W\left(W^T W\right)^{-} W^T$ for Model B. Furthermore, recall the following results about the normal distribution and distributions of quadratic forms.

Result 1. Let A be a symmetric matrix with rank (A) = k, and let $\underset{\sim}{Y} \sim N\left(\underset{\sim}{\mu}\, ,\, \mathring{\mathbf{a}}\right)$.

If $A\Sigma$ is idempotent, then $\underset{\sim}{Y}^T A \underset{\sim}{Y} \sim \chi^2_k\left(\underset{\sim}{\mu}^T A \underset{\sim}{\mu}\right)$.

Result 2. Let $\underset{\sim}{Y} \sim N\left(\underset{\sim}{\mu}\, ,\, \mathring{\mathbf{a}}\right)$ and let A_1 and A_2 be symmetric matrices. If

$A_1 \Sigma A_2 = 0$, then $\underset{\sim}{Y}^T A_1 \underset{\sim}{Y}$ and $\underset{\sim}{Y}^T A_2 \underset{\sim}{Y}$ are independent random variables.

Result 3. If $\underset{\sim}{Y} \sim N\left(\underset{\sim}{\mu}\, ,\, \mathring{\mathbf{a}}\right)$ and A is a symmetric matrix of non-random constants, then

$$E\left(\underset{\sim}{Y}^T A \underset{\sim}{Y}\right) = \underset{\sim}{m}^T A \underset{\sim}{m} + \operatorname{tr}(A\Sigma)$$

$$\operatorname{Var}\left(Y^T A Y\right) = 4\underset{\sim}{m}^T A \Sigma A \underset{\sim}{m} + 2\operatorname{tr}(A\Sigma A\Sigma)$$

Result 4. If $\underset{\sim}{Y} \sim N\left(\underset{\sim}{\mu}\, ,\, \mathring{\mathbf{a}}\right)$ and A is a matrix of non-random constants, then

$A\underset{\sim}{Y} \sim N\left(A\underset{\sim}{\mu}\, , A\, \mathring{\mathbf{a}}\, A^T\right)$.

Result 5. If $\begin{bmatrix} \underset{\sim}{Y_1} \\ \underset{\sim}{Y_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \underset{\sim}{m_1} \\ \underset{\sim}{m_2} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$ then $\underset{\sim}{Y_1}$ and $\underset{\sim}{Y_2}$ are

independent random vectors if and only if $\operatorname{Cov}\left(\underset{\sim}{Y_1}\, ,\, \underset{\sim}{Y_2}\right) = \Sigma_{12} = 0$.

(i) Show that $\frac{1}{\sigma^2}\underset{\sim}{Y}^T(I - P_X)\underset{\sim}{Y}$ has a chi-square distribution under model B, i.e., when $\underset{\sim}{Y} \sim N\left(W\underset{\sim}{\gamma}, \sigma^2 I\right)$. Report degrees of freedom. Does it have a central chi-square distribution?

(ii) Show that $\frac{1}{\sigma^2}\underset{\sim}{Y}^T(I - P_X)P_W(I - P_X)\underset{\sim}{Y}$ has a chi-square distribution under model B. Report degrees of freedom. Does it have a central chi-square distribution?

G. Using the notation and results presented in Part (F), show that

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

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.

H. 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.

Model C: $\underset{\sim}{Y} = M\underset{\sim}{\gamma} + \underset{\sim}{\varepsilon}$ where $\underset{\sim}{\varepsilon} \sim N\left(\underset{\sim}{0}, \sigma^2 I\right)$ and

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

(A) Under what conditions, if any, would g_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 $H_0 : \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$ against the alternative that H_0 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 <u>not</u> 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 σ^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 $\underset{\sim}{Y} = X\underset{\sim}{\boldsymbol{b}} + \underset{\sim}{\boldsymbol{e}}$, where X is an $n \times k$ matrix with $\text{rank}(X)=k$ and $\text{Var}(\underset{\sim}{\varepsilon}) = \sigma^2 I + \delta^2 \underset{\sim}{1}\underset{\sim}{1}^T$, where I is the $n \times n$ identity matrix and $\underset{\sim}{1}$ 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 $\underset{\sim}{\beta}$ is estimable and $\underset{\sim}{c}^T\underset{\sim}{\beta}$ is estimable for any $\underset{\sim}{c} \neq \underset{\sim}{0}$. The ordinary least squares estimator for $\underset{\sim}{c}^T\underset{\sim}{\beta}$ is

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

(A) Show that $\underset{\sim}{c}^T\underset{\sim}{b} = \underset{\sim}{c}^T (X^TX)^{-1}X^T\underset{\sim}{Y}$ is an unbiased estimator for $\underset{\sim}{c}^T\underset{\sim}{\beta}$.

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

Exam Score ____________