Stat 501 **EXAM 1** Name ______________
Spring 2005

Instructions: Write your answers in the spaces provided on this exam. Use the back of the page or another sheet of paper if you need more space, but clearly indicate where this is done. Display formulas used in calculations and display quantities such as degrees of freedom, mean vectors, parameter matrices, formulas for test statistics, etc. This is more important than the final numerical answer. Multiple choice questions may have more than one correct answer or no correct answers. Circle each correct answer. You may use a calculator and the formula sheet and tables attached to this exam. No other materials are allowed.

1. Suppose $\mathbf{X} = (X_1 + X_2 + X_3 + X_4)' \sim N(\boldsymbol{\mu}, \Sigma)$ where

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 7 \\ 11 \\ 12 \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} 25 & 15 & 5 & 2 \\ 15 & 25 & 10 & 3 \\ 5 & 10 & 25 & 8 \\ 2 & 3 & 8 & 25 \end{bmatrix}$$

(a) What is the joint distribution of $\mathbf{Y} = \begin{bmatrix} X_1 - X_2 \\ X_3 - X_4 \end{bmatrix}$?

(b) What is the conditional distribution of X_3 given $(X_1, X_2) = (5, 8)$?

(c) Consider the region with boundary defined by

$$(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \chi^2_{(4).05} = 9.49.$$

Which of the following statements are correct? (Circle each correct statement.)

i. The region is centered at $\boldsymbol{\mu} = [3\ \ 7\ \ 11\ \ 12]'$.

ii. The probability that a randomly selected observation is in this region is 0.95.

iii. Since $\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44} = 25$, the region is a 4-dimensional sphere.

iv. The volume of this region is proportional to the determinant of Σ.

v. The value of the density function is larger for any point inside the boundary than for any point outside the boundary.

2. What is the Anderson-Darling test? Describe the null hypothesis and the basic motivation behind the construction and application of the test. (It is not enough to simply report a formula for the test statistic, in fact, you can answer this question without reporting a formula.)

3. Soil samples were taken at n randomly selected locations in corn fields in central Iowa. Each soil sample was obtained by drilling a hollow tube into the first 10 feet of soil. This allows measurements of the nitrogen concentration in the soil to be made at depths of 1, 3, 5, 7, and 9 feet from the surface. The five measurements from the j-th location can be arranged in a vector as

$$\mathbf{X}_j = (X_{1j}, X_{2j}, X_{3j}, X_{4j}, X_{5j})'$$

Let $\bar{\mathbf{X}}$ denote the vector of sample means and let $S = \frac{1}{n-1}\Sigma_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'$. Suppose $\mathbf{X}_j, j = 1, 2, \ldots, n$ are independent vectors of observations sampled from a multivariate normal population of potential vectors of measurements with unknown mean vector $\boldsymbol{\mu}$ and unknown covariance matrix Σ.

(a) Show how you would use T^2 to test the hypothesis that the mean nitrogen concentration is the same at all 5 depths.

(b) How would you test the null hypothesis that the mean nitrogen concentration decreases in such a way that the mean at one depth is half of the mean at the previous depth, i.e., $H_0 : (\mu_i/\mu_{i-1}) = .5, i = 2, 3, 4, 5$, where μ_i is the mean nitrogen concentration at $2i - 1$ feet below the surface?

(c) Suppose you computed Λ, the ratio of likelihoods, for the test of the null hypothesis that the correlation matrix for nitrogen concentrations at the 5 depths has the form

$$P = \begin{bmatrix} 1 & \alpha & \beta & \gamma & \delta \\ \alpha & 1 & \alpha & \beta & \gamma \\ \beta & \alpha & 1 & \alpha & \beta \\ \gamma & \beta & \alpha & 1 & \alpha \\ \delta & \gamma & \beta & \alpha & 1 \end{bmatrix}$$

against the general alternative of an arbitrary correlation matrix. Given a numerical value for Λ, show how you would test this null hypothesis.

4. Sixty young male pigs were used in an experiment to determine the effects of riboflavin on growth. All of the pigs were 19 days old when they entered the study. They were randomly divided into three groups with 20 pigs in each group. The three groups of pigs were randomly assigned to three diets. The diets were the same except for the level of riboflavin: 0 mg, 5 mg, or 10 mg per kg of feed. Each pig was kept in a separate pen. Increases in muscle tissue were recorded every two weeks after the beginning of the study. Sample means for the biweekly increases are shown below:

Average Biweekly Increases in Muscle Tissue

	Inspection Times			
Added Riboflavin	2 weeks	4 weeks	6 weeks	8 weeks
0 mg/kg	4.2	4.4	5.1	5.3
5 mg/kg	5.3	5.8	6.4	6.7
10 mg/kg	5.4	5.5	6.2	6.9

(a) What are the experimental units in this study?

(b) Which factor, if any, corresponds to repeated measurements on the the experimental units?

(c) i. Consider the null hypothesis of no interaction between level of riboflavin in the diet and time. Write this null hypothesis in the form $H_0 : C\boldsymbol{\beta}M = \mathbf{0}$. To answer this question you must specify a model that defines a matrix of parameters $\boldsymbol{\beta}$. Carefully define any notation you use to define this model.

ii. Suppose the value of Wilks criterion for testing this null hypothesis is 0.835. Show how to convert this value into an F-test. Report the degrees of freedom.

(d) Specify conditions on the distribution of biweekly weight gains under which the use of Wilks criterion in part (c) provides an F-test with reliable p-values. (Assume these conditions are true for the rest of the problem.)

(e) Assuming the conditions you outlined in part (d) are correct, is the F distribution for the test in part (c) exact or only an approximation when the null hypothesis is true?

(f) i. Consider the null hypothesis that for each two-week period the mean increases in muscle tissue are the same for all three levels of riboflavin. Mean increases in muscle tissue could differ across time periods, however. Using your model from part (c), write this null hypothesis in the form $H_0 : C\boldsymbol{\beta}M = \mathbf{0}$.

ii. Suppose the value of Wilks criterion for testing this null hypothesis is 0.632. Show how to convert this value into an F-test. Report the degrees of freedom.

(g) i. Consider the null hypothesis that the biweekly mean increases in muscle tissue follow a straight line trend across time when no riboflavin is added to the diet. Using your model from part (c), write this null hypothesis in the form $H_0 : C\boldsymbol{\beta}M = \mathbf{0}$.

ii. Suppose the value of Wilks criterion for testing this null hypothesis is 0.891. Show how to convert this value into an F-test. Report the degrees of freedom.

(h) An analysis of variance table with the following sources of variation was suggested by one researcher:

Source of Variation	d.f.
Time	3
Riboflavin Levels	2
Time $\times$ Riboflavin level interaction	6
Pigs within Riboflavin Levels	57
Residuals	171
Corrected total	239

i. Write out a formula for the mixed effects model that corresponds to this ANOVA table. Carefully define any parameters used in the model and describe the distributional assumptions for the random effects.

ii. To form an F-test, the mean square for riboflavin levels would be divided by which other mean square? What are the degrees of freedom for this test? How does the null hypothesis tested by this F-test differ from the null hypothesis tested in part (f)?

iii. Identify conditions under which the F-test for no interaction between time and riboflavin levels computed from this ANOVA table would be better than the test in part (c)? In what way would it be better?

5. A random sample of 800 fifth grade students were each given three exams. The first exam provided an IQ (intelligence quotient) score, the second exam measured mathematical ability, and the third exam measured verbal ability. The researchers want to test the hypothesis that the correlation between mathematical ability and IQ is the same as the correlation bewteen verbal ability and IQ. They ask you to help them do this test. Describe what you would do.

EXAM SCORE __________