STAT 501 EXAM 1 Name _____
SPRING 2001

Instructions: You may use only your calculator and the attached tables and formula sheet. You can separate the tables and formula sheet from the rest of the exam. Show your solutions and explanations in the space provided on this exam. You may not have time to complete all of the calculations. Credit will be given for describing an appropriate method or showing how to appropriately apply a formula, even when the final answer is missing or incorrect.

1. Researchers at Iowa State University have collected information from a sample of n=55 graduate students from a certain country where English is not the native language. The j-th student in the sample provided information on

X_{1j} = Score on a test of English proficiency (call it Test 1) taken in the student's native country.

X_{2j} = Score on a second test of English proficiency (call it Test 2) taken in the student's native country.

X_{3j} = Score on a third test of English proficiency (call it Test 3) given after the student arrives at Iowa State University, but before the student takes English 100D.

X_{4j} = Score on a fourth test of English proficiency (call this Test 4) taken after completing the English 101D course at ISU.

X_{5j} = Number of years the student studied English in their native country.

Denote the set of responses for the j-th student as $\underset{\sim}{X}_j = \left(X_{1j}\ X_{2j}\ X_{3j}\ X_{4j}\ X_{5j}\right)'$ and assume $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n \sim \text{NID}(\underset{\sim}{\mu}, \Sigma)$ for some unknown mean vector and covariance matrix

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

respectively.

Let

$$\underset{\sim}{\bar{X}} = \frac{1}{n}\sum_{j=1}^{n} \underset{\sim}{X}_j \qquad \text{and} \qquad S_x = \frac{1}{n-1}\sum_{j=1}^{n} (\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})'$$

denote the sample mean vector and sample covariance matrix for this sample of n=55 students.

(a) What is the distribution of the sample mean vector $\underset{\sim}{\bar{X}}$?

(b) What is the distribution of $Z_j = X_{3j} - (X_{2j} + X_{1j})/2$?

(c) What is the distribution of $W = (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})' S_x^{-1} (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})$?

(d) The maximum likelihood estimate of $\rho_{14\cdot 35}$, the partial correlation between the score on Test 1 and the score on Test 4 controlling for years of study (X_5) and the score on Test 3 (X_3), is 0.21. Can you reject the null hypothesis $H_0 : \rho_{14\cdot 35} = 0$ at the $\alpha = .05$ level of significance? Give a value of a test statistic and show how you reached your conclusion.

(e) Show how to test the null hypothesis

$$H_0 : (\mu_4 - \mu_1) = (\mu_4 - \mu_2) = (\mu_4 - \mu_3)$$

against the alternative that H_0 is false. Give a formula for your test statistic, degrees of freedom, and show how to obtain a p-value.

(f) The researchers constructed a likelihood ratio test of the null hypothesis

$$H_0 : \rho_{14} = \rho_{24} = \rho_{34} \text{ and } \rho_{12} = \rho_{23} = \rho_{13}$$

against the alternative that H_0 is not true. Here ρ_{ij} denotes the correlation between the scores for the i-th and j-th tests. The observed value for the ratio of optimized likelihoods is $\hat{\Lambda} = 0.18$. Is there sufficient evidence to reject the null hypothesis at the $\alpha = .05$ level of significance? Show how you reached your conclusion.

2. Another way to obtain an F-test of the null hypothesis in part(e) of problem 1 is to first compute differences in test scores, $D_{1j}=X_{4j}-X_{1j}, D_{2j}=X_{4j}-X_{2j}, D_{3j}=X_{4j}-X_{3j},$ and then compute a F-test from the ANOVA table for the model

$$D_{ij} = \mu + \alpha_i + S_j + \varepsilon_{ij} \qquad \text{for } j = 1, 2, \ldots, 55 \text{ subjects and } i = 1, 2, 3 \text{ differences.}$$

In this model, $S_j \sim \text{NID}(0, \sigma_S^2)$, j = 1, 2, …, 55, are random student effects, and $\varepsilon_{ij} \sim \text{NID}(0, \sigma_\varepsilon^2)$ are independent random errors. Any random error is independent of any of any random subject effect.

A. What are the degrees of freedom for this F-test?

B. Explain how the assumptions about the distribution of (D_{1j}, D_{2j}, D_{3j}) for this F-test differ from the assumptions about the distribution of (D_{1j}, D_{2j}, D_{3j}) used in the test in part (e) of problem 1. Describe the distributional assumptions that are used for both tests.

EXAM SCORE _____

3. The information in Problem 1 was also obtained from an independent sample of m = 22 students from another country. The sample mean vector and sample covariance matrix for the results from this sample are denoted by

$$\underset{\sim}{\bar{Y}} = \frac{1}{m}\sum_{j=1}^{m}\underset{\sim}{Y}_j \qquad \text{and} \qquad S_y = \frac{1}{m-1}\sum_{j=1}^{m}(\underset{\sim}{Y}_j - \underset{\sim}{\bar{Y}})(\underset{\sim}{Y}_j - \underset{\sim}{\bar{Y}})',$$

respectively.

(a) Explain how you would test the null hypothesis that the population covariance matrices are equal for students from these two countries. Give a formula for your test statistic and its degrees of freedom.

(b) Suppose the null hypothesis was rejected in Part (a). How would you test the null hypothesis that the mean vectors are the same for the populations of students from these two countries?

4. An animal ecologist wanted to know if a certain species of small mammal uses different types of habitat in proportion to availability in the animal's home range. The home range is the area in which the animal lives. Habitat was classified into one of the following four categories in this study:

    (1) Cultivated land (cornfields, wheat fields, etc.)
    (2) Grassland (not cultivated and few trees)
    (3) Woodland (not cultivated and many trees)
    (4) Wetlands (streams, lakes, ponds, marshes, swamps, etc.)

    Suppose, for example, that the home range for a particular animal consists of 70% cultivated land, 10% grassland, 15% woodland, and 5% wetlands. The animal would use the four types of habitat in proportion to availability if it spent 70% of its time in cultivated fields, 10% of its time in grassland, 15% of its time in woodland, and 5% of its time in wetlands. If this were not true, we would call the animal selective in its habitat use.

    Data were collected by capturing a sample of n=30 of these animals. A collar with a radio transmitter was placed on each animal before it was set free. A global positioning satellite system was used to locate the position of each animal at hourly time intervals during a two month study period. For each animal, these locations were plotted on a map and its home range was determined as the interior of the smallest circle that included all the points on the map. Using detailed maps of the habitat in each area, the researchers were able to determine $\underset{\sim}{X}_j = \left(X_{1j},\ X_{2j},\ X_{3j},\ X_{4j}\right)'$, the proportions of time that the i-th animal spent in each type of habitat, and $\underset{\sim}{Y}_j = \left(Y_{1j},\ Y_{2j},\ Y_{3j},\ Y_{4j}\right)'$ the proportion of the j-th animal's home range that is covered by each type of habitat. The animals were far enough apart to prevent overlap of the home ranges. Consequently, the researchers were confident that habitat use by one animal had no influence on habitat use by any other animal in the study.

    Show how to perform a T^2 test of the null hypothesis that on average these animals use habitat types in proportion to availability in their home ranges. Give a formula for your test statistic, degrees of freedom, and show how to obtain a p-value.