**STAT 501** **Final Exam Solutions and Comments** **Spring 2001**

1. a. (3 points) (1.49/7) x 100% = 21.3%

b. (3 points) 0

c. (3 points) $(0.453)\sqrt{3.82} = 0.89$

d. (3 points) $\hat{\lambda}_1 = 3.82$

e. (5 points) The formula for the scores use the standardized values of the measured traits,

$$\hat{Y}_{1j} = 0.285\left(\frac{X_{1j}-\overline{X}_1}{\sqrt{s_{11}}}\right) + 0.211\left(\frac{X_{2j}-\overline{X}_2}{\sqrt{s_{22}}}\right) + 0.294\left(\frac{X_{3j}-\overline{X}_3}{\sqrt{s_{33}}}\right) + 0.435\left(\frac{X_{4j}-\overline{X}_4}{\sqrt{s_{44}}}\right)$$

$$+ 0.453\left(\frac{X_{5j}-\overline{X}_5}{\sqrt{s_{55}}}\right) + 0.453\left(\frac{X_{6j}-\overline{X}_6}{\sqrt{s_{66}}}\right) + 0.434\left(\frac{X_{7j}-\overline{X}_7}{\sqrt{s_{77}}}\right)$$

This is an overall size component with more emphasis on body length and height measurements than on head size measurements. This component accounts for about 54.6% of the total variance of the standardized measurements.

f. (5 points) The second principal component compares head size with length of body measurements. It assumes large positive values for tall criminals with relatively small heads, and it assumes extreme negative values for short criminals with relatively large heads. This component accounts for about 21.3% of the total variance of the standardized measurements.

The third principal component is a head shape component that compares head length with head width. It assumes large positive values for criminals with relatively long and narrow heads, and it assumes extreme negative values for criminals with relatively short and wide heads. This component accounts for about 9.3% of the total variance of the standardized measurements.

g. (3 points) The first principal component corresponds to the positive correlations between all of the measurements. The higher loadings of the body length measurements on the first component reflects that correlations among the body length measurements are stronger than correlations among the head size measurements. The second principal component corresponds to the tendency for correlations between head size and body length measurements to be weaker than either the correlations among the head size measurements or the correlations among the body length measurements. The third principal component reflects weaker correlations between head length and head width measurements than the correlation between the head width measurements.

2.a. (4 points) The communality is $h_7^2 = (0.799)^2 + (0.221)^2 = 0.687$ .

The specific variance is $\psi_7^2 = 1 - 0.687 = 0.313.$

b. (3 points) The main objective of a varimax rotation is to create a new set of orthogonal factors where each of the measured attributes has a high loading on only one factor.

c. (4 points) The first rotated factor is essentially a body length factor. It assumes large values for tall criminals with long left arm, foot and forefinger, and it assumes small values for short criminals with relatively short left arm, foot and forefinger.

The second rotated factor is essentially a head size factor. It assumes large values for criminals with relatively large heads, and it assumes small values for criminals with relatively small heads. This component accounts for about 9.3% of the total variance of the standardized measurements.

d. (3 points) $(0.975)(0.812) + (0.122)(0.256) = 0.808$

e. (3 points) The relatively large value of the Tucker and Lewis reliability coefficient indicates that after conditioning on the factor scores there are no large partial correlations between any pair of measurements. This suggests that two factors are adequate for describing the correlations among the measured traits.

f. (5 points) This method is based on the assumption that the observations on the measured traits and the corresponding values of the unobserved factors have a joint normal distribution, i.e.,

$$\begin{bmatrix} X_j \\ F_j \end{bmatrix} = N\left( \begin{bmatrix} \mu_j \\ 0 \end{bmatrix}, \begin{bmatrix} L\,L^T + \psi & L \\ L^T & I \end{bmatrix} \right).$$

Then, factor scores are obtained by computing the conditional mean of $F_j$ given $X_j$ , and replacing $\mu_j$ with the vector of sample means, i.e.,

$$\hat{F}_j = \hat{L}^T(\hat{L}\,\hat{L}^T + \hat{\psi})^{-1}(X_j - \overline{X})$$

3. a. (6 points, 1 for each item) A, B, C, D, F

b. (6 points, 1 for each item) C

b. (4 points, 1 for each item) A, B

4. a. (5 points) $T^2 = n(C\overline{X})'(CSC')^{-1}(C\overline{X})$ where $n = 25$ and

$$\overline{X}' = (\overline{X}_{10}, \overline{X}_{20}, \overline{X}_{30}, \overline{X}_{11}, \overline{X}_{21}, \overline{X}_{31}, \overline{X}_{12}, \overline{X}_{22}, \overline{X}_{32}, \overline{X}_{13}, \overline{X}_{23}, \overline{X}_{33})$$

is the transpose of the sample mean vector (note that 12 measurements are taken on each rabbit) and $S$ is the 12x12 sample covariance matrix, and

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

b. (3 points) (3,22) degrees of freedom

c. (5 points) The null hypothesis can be written as $H_o$: $C\beta M = 0$, where $C=1$, $\beta$ is a $1\times12$ vector of means arranged the same way as $\overline{X}'$ in part (a), and M has six columns corresponding to six linearly independent interaction contrasts. Then, k = 1, r = 1, p = 12, u=6, and n=25 and the degrees of freedom are ( 6, 19). This can also be done as a one sample $T^2$ test. One selection for M is

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

5. a. (3 points) Compute the values of the linear discriminants at Age=27, EF=40, HV=60, R1=R2=0, MI=0.

$$d_1^L(\underset{\sim}{x}) = -53.22 + (0.18)(27) + (10.42)(0) + (10.63)(0) + (0.71)(40) + (0.64)(60) + (14.05)(0) = 18.44$$

$$d_2^L(\underset{\sim}{x}) = -54.24 + (0.19)(27) + (10.75)(0) + (11.65)(0) + (0.65)(40) + (0.63)(60) + (16.34)(0) = 14.69$$

Since $d_1^L(\underset{\sim}{x}) > d_2^L(\underset{\sim}{x})$, classify as a VT case.

b. (3 points) Compute $\log\left(\frac{c(1\,|\,2)p_2}{c(2\,|\,1)p_1}\right) = \log\left(\frac{(1)(.9)}{(20)(.1)}\right) = \log(0.45) = -0.799$ and classify the patient as VT case because $d_1^L(\underset{\sim}{x}) - d_2^L(\underset{\sim}{x}) > -0.799$.

c. (3 points) When it is a result of a likelihood ratio criterion. That would be the case if (AGE, R1, R2, EF, HV, MI) has a multivariate normal distribution for both the VT and non-VT populations of patients and the covariance matrices are the same for those two populations. Since several of the variables are binary, these conditions would not be satisfied. Nevertheless, a linear discriminant rule could still be a very good classification rule, even if it is not optimal.

d. (3 points) One observation is set aside from the training samples and the linear discriminant rule is fit to the remaining observations in the training sample. The fitted rule is applied to the observation that was set aside and the classification result is recorded. The observation that was set aside is put back into the training sample and another observation is set aside and the process is repeated. This continues until each observation in the training samples has been set aside once and classified. The misclassification rates are estimated as the proportions of cases from the training samples that are misclassificd by this procedure.

e. (3 points) Describe the partial F-test

f. (3 points) The plots of the crossvalidation results indicate that a tree with three terminal nodes is a good choice for minimizing the probability of misclassifying a patient who is not a member of one of the training samples available for this study.

g. (3 points) Non-VT case.

h. (3 points) The crossvalidation results indicate that overall the probability of misclassification is about 42/186=0.226. The crossvalidation results do not provide any information about the probabilities of different types of misclassification. The resubstitution results indicate that you may be 3 times more likely to misclassify a VT case as a non-VT case than to misclassify a non-VT case as a VT case.

Final exam scores are given in the following stem-leaf display:

```
9 |
8 | 6 7
8 | 0 1 1 2 2 3 4
7 | 5 5 5 5 7 7 8
7 | 1 1 3 4
6 | 6 9
6 | 0 1 1
5 | 6 8
5 | 2
4 | 5 8
4 |
```