University of California, Davis
Department of Statistics

Name ....Solution....
PRINT PLEASE

Sample
STA 135 **Final**
(The final exam is on **Wednesday, March 16, 3:30-5:30pm in Socsci 1100**)

Instructions: 1. **WORK ALL PROBLEMS**. Please, give details and explanations and **SHOW ALL YOUR WORK** so that partial credits can be given.
2. You may use **Four** sheets of **notes** and a **calculator** but **no** other reference materials.

The final exam is going to have five questions. Two questions from the topics of midterm II, similar to the questions of midterm II. There will be three questions from the topics that is covered after midterm II. The followings provide examples of the three questions on the subjects covered after the second midterm.

Points

1. In a multivariate analysis of variance, we have interest in comparing the mean vectors for three treatment groups. For a sample of size 50 from each treatment, the sample mean vectors are:

$\bar{\underline{x}}_1 = [3.428 \quad .306]'$, $\bar{\underline{x}}_2 = [2.770 \quad 1.326]'$, $\bar{\underline{x}}_3 = [2.974 \quad 2.026]'$.

The between and within matrices of sum of squares are:

$$B = \begin{bmatrix} 11.344 & -21.820 \\ -21.820 & 75.352 \end{bmatrix}, \qquad W = \begin{bmatrix} 16.950 & 4.125 \\ 4.125 & 14.729 \end{bmatrix}$$

(a) Construct the multivariate analysis of variance table.
(b) Test the hypothesis of no treatment effect at 0.05 level. ($F_{4,292}(0.05) = 2.37$)
(c) Construct 95% Bonferroni confidence intervals for the treatment effects. ($t_{147}(.00042) = 2.675$)

(a) The MANOVA Table is

Source	Matrix SS	df
Treatments	B	$g-1$
Residual	W	$n-g$
Total	$B+W$	$n-1$

Using the given information, we have:

MANOVA Table

Source	Matrix SS	df
Treatment	$B = \begin{bmatrix} 11.344 & -21.820 \\ -21.820 & 75.352 \end{bmatrix}$	2
Residual	$W = \begin{bmatrix} 16.950 & 4.125 \\ 4.125 & 14.729 \end{bmatrix}$	147
Total	$B+W = \begin{bmatrix} 28.294 & -17.695 \\ -17.695 & 90.081 \end{bmatrix}$	149

(b)

$$\Lambda^* = \frac{|W|}{|B+W|} = \frac{232.64}{2235.64} = .104 \qquad n = 150$$

For $p \geq 1$ and $g = 3$, $\left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p,2(n-p-2)}$

$$\left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = \left(\frac{150-2-2}{2}\right)\left(\frac{1-\sqrt{.104}}{\sqrt{.104}}\right) = 153.3$$

$$F_{2p,2(n-p-2)}^{(.\alpha)} = F_{4,292}^{(.05)} = 2.37$$

Since $153.3 > 2.37$, we can reject $H_0: \tau_1 = \tau_2 = 0$ at .05 level.

(c) The $100(1-\alpha)\%$ Bonferroni intervals are

$$\bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}\left(\alpha/pg(g-1)\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)}$$

w_{ii} = i-th diag. elem. of W

for $i=1$: $t_{n-g}\left(\alpha/pg(g-1)\right)\sqrt{\frac{w_{11}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)} = 2.675\sqrt{\frac{16.950}{147}\left(\frac{2}{50}\right)} = .182$

for $i=2$: $= 2.675\sqrt{\frac{14.729}{147}\left(\frac{2}{50}\right)} = .169$

	Comparisons $i=1$	Comparisons $i=2$
1 vs 2	$(3.428-2.770) \pm .182 \Rightarrow (.476, .840)$	$(.306-1.326) \pm .169 \Rightarrow (-1.189, -.851)$
1 vs 3	$(3.428-2.974) \pm .182 \Rightarrow (.272, .636)$	$(.306-2.206) \pm .169 \Rightarrow (-2.069, -1.731)$
2 vs 3	$(2.770-2.974) \pm .182 \Rightarrow (-.386, -.022)$	$(1.326-2.206) \pm .169 \Rightarrow (-1.049, -.711)$

2. The covariance matrix for a two-dimensional random vector $\underline{X}$ is:

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

(a) Determine the population principal components Y_1 and Y_2 .
(b) What portion of the total population variance is explained by Y_1.
(c) Covert the covariance matrix to a correlation matrix and compute the principal components based on the correlation matrix.
(d) Compare the principal components obtained in parts (a) and (c) and discuss.
(e) Compute the correlations between Y_1 , Y_2 and Z_1 and Z_2 , where Z_1 and Z_2 are the standardized values of X_1 and X_2 .

(a) $|\Sigma - \lambda I| = 0 \Rightarrow \begin{vmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 - 7\lambda + 6 = 0$

The eigenvalues are $\lambda_1 = 6$ and $\lambda_2 = 1$

$(\Sigma - \lambda_i I)\underline{e}_i = 0 \Rightarrow$ eigenvectors $\underline{e}_1 = \begin{bmatrix} .894 \\ .447 \end{bmatrix}$ and $\underline{e}_2 = \begin{bmatrix} .447 \\ -.894 \end{bmatrix}$

Then, The Principal components are

$Y_1 = .894\, X_1 + .447\, X_2$

$Y_2 = .447\, X_1 - .894\, X_2$

(b) $Var(Y_1) = \lambda_1 = 6$. The portion of total variance explained by Y_1 is

$6/(6+1) = .86$ 86 %.

(c) $\rho = \begin{bmatrix} 1 & .6325 \\ .6325 & 1 \end{bmatrix}$

The eigen values are $\lambda_1 = 1.6325$, $\lambda_2 = .4595$ and the eienvectors are $\underline{e}_1 = \begin{bmatrix} .707 \\ .707 \end{bmatrix}$ and $\underline{e}_2 = \begin{bmatrix} .707 \\ -.707 \end{bmatrix}$. Then the Principal Components are

$Y_1 = .707\, Z_1 + .707 Z_2$

$Y_2 = .707 Z_1 - .707\, Z_2$

The proportion of total variance explained by $Y_1 = \frac{1.6325}{2} = .816$

(d) The results in parts (a) and (c) are different, since variances of X_1 and X_2 are different.

(e) The Correlation Coefficient between Y_i and Z_k is $\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i}$.

$\rho_{Y_1, Z_1} = e_{11}\sqrt{\lambda_1} = .707\sqrt{1.6325} = .9033$ | $\rho_{Y_2, Z_1} = e_{21}\sqrt{\lambda_2} = .707\sqrt{.3675} = .4286$

$\rho_{Y_1, Z_2} = e_{12}\sqrt{\lambda_1} = .707\sqrt{1.6325} = .9033$ | $\rho_{Y_2, Z_2} = e_{22}\sqrt{\lambda_2} = -.707\sqrt{.3675} = .4286$

3. The mean vector of a p-dimensional random vector $\underline{X}$ is $\underline{0}$ and its covariance matrix is Σ. Let the pairs $(\lambda_1,\underline{e}_1), \ldots, (\lambda_p,\underline{e}_p)$ represent the eigenvalues and eigenvectors of Σ. Define the orthogonal matrix Q as $Q = (\underline{e}_1, \ldots, \underline{e}_p)$, and let $\underline{Y} = Q'\underline{X}$.

(a) Find the mean vector and covariance matrix of $\underline{Y}$.
(b) Partition Q as $Q = (Q_1, Q_2)$ where Q_1 is a pxq matrix ($q < p$). let $\underline{Y}^{(1)}$ be a q-dimensional random vector constructed from the first q elements of $\underline{Y}$. Find a random vector $\underline{\varepsilon}$ such that $\underline{X} = Q_1 \underline{Y}^{(1)} + \underline{\varepsilon}$.
(c) Find $E(\underline{\varepsilon}' \underline{\varepsilon})$.
(d) Let $\underline{X}_1, \ldots, \underline{X}_n$ be a random sample of size n from the above population (a population with mean vector $\underline{0}$ and covariance matrix Σ). Find an estimate for $E(\underline{\varepsilon}' \underline{\varepsilon})$.

(a) $E(\underline{Y}) = E(Q'\underline{X}) = Q'E(\underline{X}) = Q'\underline{0} = \underline{0}$

$Cov(\underline{Y}) = E(\underline{Y}\,\underline{Y}') = E[(Q'\underline{x})(\underline{x}'Q)] = Q'E(\underline{x}\,\underline{x}')Q = Q'\Sigma Q = \Lambda$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$.

(b) $\underline{Y} = Q'\underline{X} \Rightarrow Q\underline{Y} = QQ'\underline{X} \Rightarrow \underline{X} = Q\underline{Y}$ since $Q'Q = I$

$\underline{X} = [Q_1, Q_2]\begin{bmatrix} \underline{Y}^{(1)} \\ \underline{Y}^{(2)} \end{bmatrix} = Q_1 \underline{Y}^{(1)} + Q_2 \underline{Y}^{(2)}$.

This leads to $\underline{\epsilon} = Q_2 \underline{Y}^{(2)}$.

(c) $E(\underline{\epsilon}\,\underline{\epsilon}') = E[(Q_2 \underline{Y}^{(2)})'(Q_2 \underline{Y}^{(2)})]$

$= E(\underline{Y}^{(2)\prime} Q_2' Q_2 \underline{Y}^{(2)})$

$= E(\underline{Y}^{(2)\prime}\underline{Y}^{(2)})$ since $Q_2'Q_2 = I$

$= E(\sum_{i=q+1}^{p} Y_i^2)$

$= \sum_{i=q+1}^{p} E(Y_i^2) = \sum_{i=q+1}^{p} Var(Y_i)$

$= \sum_{i=q+1}^{p} \lambda_i$

(d) Let S be the sample covariance matrix of $\underline{X}$,

$$S = \frac{1}{n-1}\sum_{j=1}^{n}(\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})'.$$

Then S is an unbiased estimator of Σ.

Let $\hat{\lambda}_i$ be the i-th eigenvalue of S. $\hat{\lambda}_i$ is an estimator for λ. Then an estimator for $E(\underline{\epsilon}'\underline{\epsilon})$ is $\sum_{i=q+1}^{p}\hat{\lambda}_i$.