UNIVERSITY OF FORT HARE

STS 504

HONOURS EXAMINATIONS

NOVEMBER 2018

... ...

Time: 3Hours

Subject: Applied Multivariate Statistics

Mark: 100

This paper consists of 5 pages including the cover page

Internal Examiners

Mr. A. Odeyemi

Prof.Y Qin

External Examiner

Prof. DJ deWaal

INSTRUCTIONS

- Answer ALL questions
- Statistical Tables will be provided.

Question 1

a) (i) Define principal component analysis and what is its purpose. [3]

(ii) Determine the population principal components Y_1 and Y_2 for the covariance matrix

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix} \qquad [6]$$

(iii) Also, calculate the proportion of the total population variance explained by the first principal component. [4]

b) Convert the covariance matrix in (a) above to a correlation matrix ρ. [5]

(i) Determine the principal components Y_1 and Y_2 from ρ and compute the proportion of total population variance explained by Y_1. [4]

(ii) Compare the components calculated in b(i) with those obtained in a(i) . Are they the same? Should they be? [2]

(iii) Compute the correlations ρ_{Y_1,Z_1}, ρ_{Y_1,Z_2}, and ρ_{Y_2,Z_1}. [6]

....[30mks]

Question 2

a) Show that the covariance matrix

$$\rho = \begin{bmatrix} 1 & .63 & .45 \\ .63 & 1 & .35 \\ .45 & .35 & 1 \end{bmatrix}$$

for the $p = 3$ standardized random variables Z_1, Z_2, and Z_3 can be generated by the $m = 1$ factor model

$$Z_1 = .9F_1 + \varepsilon_1$$

$$Z_2 = .7F_1 + \varepsilon_2$$

$$Z_3 = .5F_1 + \varepsilon_3$$

where $Var(F_1) = 1$, $Cov(\varepsilon, F_1) = 0$, and

$$\Psi = Cov(\varepsilon) = \begin{bmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{bmatrix}$$

That is, write ρ in the form $\rho = LL' + \Psi$. [6]

b) Use the information in a above

i) Calculate communalities $h_i^2, i = 1,2,3$, and interpret these quantities.

ii) Calculate $Corr(Z_i, F_1)$ for $i = 1,2,3$. Which variable might carry the greatest weight in "naming" the common factor? Why? [6]

c) The eigenvalues and eigenvectors of the correlation matrix ρ in a above are

$\hat{\lambda}_1 = 1.96$ $e_1' = [.625, \quad .593, \quad .507]$

$\hat{\lambda}_2 = .68$ $e_2' = [-.219, \quad -.491, \quad .843]$

$\hat{\lambda}_3 = .36$ $e_3' = [.749, \quad -.638, \quad -.177]$

i) Assuming an $m = 1$ factor model, calculate the loading matrix L and matrix of specific variances Ψ using the principal component solution method. Compare your results with those in 2a above. [5]

ii) What proportion of the total population variance is explained by the first common factor? [4]

....[21mks]

Question 3

a) Certain characteristics associated with a few United States presidents are listed in the table below.

President	Birthplace (region of U.S.)	Elected first term?	Party	Prior U.S. congregational experience?	Served as vice-president?
1. R.Reagan	Midwest	Yes	Republican	No	No
2. J.Carter	South	Yes	Democrat	No	No
3. G.Ford	Midwest	No	Republican	Yes	Yes
4. R.Nixon	West	Yes	Republican	Yes	Yes
5. L.Johnson	South	No	Democrat	Yes	Yes
6. J.Kennedy	East	Yes	Democrat	Yes	No

Introducing appropriate binary variables, calculate similarity coefficient,

$\frac{a+b}{p}$, where a represents the frequency of 1-1 matches, b is the frequency of 1-0 matches,

c is the frequency of 0-1 matches, d is the frequency of 0-0 matches, and $p = a + b + c + d$.[Hint: Equal weights for 1-1 matches and 0-0 matches]. **[12]**

b) Consider the matrix of distances

$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ 4 & 0 & & & \\ 6 & 9 & 0 & & \\ 1 & 7 & 10 & 0 & \\ 6 & 3 & 5 & 8 & 0 \end{bmatrix} \end{array}$$

Cluster the five items using Single-linkage hierarchical procedure. Draw the dendogram for each of the procedures. **[12]**

..........**[24mks]**

Question 4

a) Define Discriminant analysis. [3]

Consider the two data sets

$$X_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

For which

$$\overline{X_1} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \overline{X_2} = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad \text{and} \quad S_{pooled} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(i) Calculate the linear discriminant function. [5]

(ii) Classify the observation $X_0' = \begin{bmatrix} 2 & 7 \end{bmatrix}$ as population π_1 or population π_2 Using the estimated minimum ECM Rule for two normal populations. [5]

b) Define canonical correlation. Consider the covariance matrix

$$Cov\left(\begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \hline X_1^{(2)} \\ X_2^{(2)} \end{bmatrix}\right) = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

$$= \left[\begin{array}{cc|cc} 100 & 0 & 0 & 0 \\ 0 & 1 & 0.95 & 0 \\ \hline 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{array}\right]$$

Verify that the pair of canonical variate are

$$U_1 = X_2^{(1)}, V_2 = X_1^{(2)} \text{ with Canonical correlation } \rho^* = 0.95. \quad [12]$$

....[25mks]