Departamento de Matemática

Multivariate Analysis

Mater in Eng. and Data Science & Master in Mathematics and Applications

1^{st} Test — 1^{st} Semester – 2019/2020

Duration: 1.5 hours — 30/01/2020 – 11:30

Please justify conveniently your answers

Group I — 10.0 points

1. Assume that the random vector $(X_1, X_1 + X_2, X_3 - (X_1 + X_2))^t$ has multivariate normal distribution. Does $(X_1, X_2, X_3)^t$ has a multivariate normal distribution? (2.0)

2. Suppose $\boldsymbol{X} = (X_1, X_2, X_3, X_4)^t$ has a multivariate normal distribution with covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & & & \\ 1 & 2 & & \\ -1 & -1 & 3 & \\ 0 & -1 & 0 & 2 \end{pmatrix},$$

   (a) Find the variance of $X_1 + X_2 + X_3 + X_4$. (2.0)

   (b) Find all linear combinations of $aX_1 + bX_2 + cX_3 + dX_4$ which has zero correlation with $X_1 + X_2 + X_3 + X_4$. (3.0)

3. Verify that the Mahalanobis distance: (3.0)

$$D(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left((\boldsymbol{x}_i - \boldsymbol{x}_j)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)\right)^{1/2},$$

   is invariant to linear transformations of the form $\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$, where $\text{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$ is the population covariance matrix. Both $|\boldsymbol{\Sigma}$ and $\boldsymbol{A}$ matrices are nonsingular.

Group II — 10.0 points

Parkinson's Disease (PD) is a common neurological disorder, caused by the loss of brain cells that produce the chemical dopamine. A study was carried out to compare brain activity in PD sufferers and controls. A representative sample of 42 PD patients, at various stages of the disease, underwent brain imaging using Single Photon Emission Computerised Tomography. 14 controls of similar age to the patients, but with no known neurological disorder, were also recruited into the study and imaged in the same way.

A measure of brain activity was obtained for each of the Striatum (Sr), Caudate (Ca), and Putamen (Pu) regions on the left (L) side and right (R) side of the brain. These six regions are known to be important in dopamine production. This measure of activity was dimensionless but larger values indicated more brain activity. The Hoehn and Jahr (HY) disease stage was also recorded for each PD sufferer. This is a clinical indicator of the progress of the disease, which proceeds from Stage 1 to Stage 5 as the condition gets worse. In this study, Stage 0 was used to indicate normal controls.

The computer output displayed on the next page is from an analysis of the data for all 56 persons in the study.

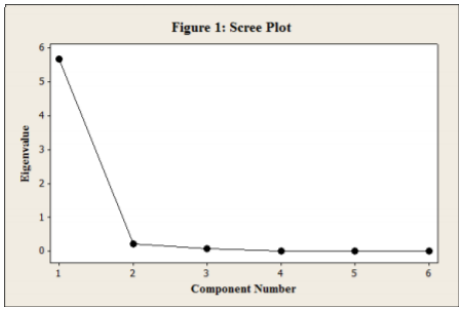
1. What does the sample correlation matrix (Table 1) reveal about the brain activity levels recorded in this study? (1.0)

Table 1: Sample Correlations.

	SrL	SrR	CaL	CaR	PuL	PuR
SrL	1.00	0.93	0.98	0.92	0.99	0.91
SrR	0.93	1.00	0.91	0.98	0.92	0.99
CaL	0.98	0.91	1.00	0.93	0.95	0.88
CaR	0.92	0.98	0.93	1.00	0.89	0.95
PuL	0.99	0.92	0.95	0.89	1.00	0.91
PuR	0.91	0.99	0.88	0.95	0.91	1.00

2. A principal component analysis of the data was carried out, based on the sample correlation matrix. Discuss some possible objectives of this analysis, and its limitations. (2.0)

3. A scree plot is given in Figure 1. What can be concluded from this plot? Having seen this plot, how would you proceed with the analysis? (2.5)



4. Table 2 lists the loadings of the first two principal components. Use these loadings to interpret each of the two components. (2.5)

Table 2: Loadings for the first two principal components.

	SrL	SrR	CaL	CaR	PuL	PuR
PC1	0.412	0.412	0.406	0.407	0.407	0.406
PC2	-0.406	0.413	-0.404	0.361	-0.411	0.449

5. Figures 2 and 3 show box and whisker plots (boxplots) of the first two principal component scores for the subjects in the study, according to their HY stage. Comment on these plots. (2.0)