Departamento de Matemática

Multivariate Analysis

Master in Mathematics and Applications

1^{st} Exam | 1^{st} Semester – 2012/2013
Duration: 3 hours | 07/01/2013 – 3 pm

Please justify conveniently your answers

Group I — 4.5 points

1. The random vector $(X, Y)^t$ has bivariate normal distribution with $Var(X) = Var(Y)$. Show that $X + Y$ and $X - Y$ are independent random variables. (1.5)

2. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be two independent random vectors, where $\boldsymbol{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2$. Consider two independent random samples, with sizes n_1 e n_2 from each population.

   (a) Prove that (1.5)

$$\frac{n_1 n_2}{n_1 + n_2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^t \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \sim \chi^2_{(p)}.$$

   (b) A phycologist wants to compare two groups of people, characterized by three random variables $\boldsymbol{X} = (X_1, X_2, X_3)^t$, representing intelligence tests. Group 1 is formed by people ($n_1 = 20$) who do not present a senile factor, and group 2 by those ($n_2 = 30$) presenting a senile factor, having obtained (1.5)

$$\bar{\boldsymbol{x}}_1 = (23.5, 30.3, 40.4)^t, \quad \bar{\boldsymbol{x}}_2 = (25.5, 31.7, 42.3)^t.$$

   If we admit that the random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in each group have multivariate normal distributions with the same covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 10 & 6 & 0 \\ & 12 & 6 \\ & & 10 \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 0.1750 & -0.1250 & 0.0750 \\ & 0.2083 & -0.1250 \\ & & 0.1750 \end{pmatrix},$$

   test if the on average the two group have equal results on the 3 performed intelligence tests. State the hypotheses, test statistic, decision rule, and conclusion. Decide based on the p-value.

Group II — 5.5 points

Let U_1 and U_2 be two independent random variables with standard normal distribution. Suppose that $\boldsymbol{X} = (U_1, U_2, U_1 + U_2, U_1 - U_2)^t$.

1. Compute the correlation matrix of $\boldsymbol{X}$, $\boldsymbol{R}$. (1.5)

2. How many standardized principal components are of interest? Note that the non trivial eigenvalues of $\boldsymbol{R}$ are $\lambda_1 = \lambda_2 = 2$. (1.0)

3. Show that the first two eigenvectors of $\boldsymbol{R}$ are $\boldsymbol{\gamma}_1 = (\frac{1}{2}, \frac{1}{2}, \frac{\sqrt{2}}{2}, 0)^t$ and $\boldsymbol{\gamma}_2 = (\frac{1}{2}, -\frac{1}{2}, 0, \frac{\sqrt{2}}{2})^t$, respectively. (1.5)

4. Write and interpret the first two standardized principal components. (1.5)

Group III **6.0 points**

In a consumer-preference study, a random sample of customers were asked to rate several attributes of a new product. The responses, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The five variables under study are: X_1 - Taste, X_2 - Good buy for money, X_3 - Flavor, X_4 - Suitable for snack, and X_5 - Provides lots of energy. The original factor loadings (obtained by the principal components method and standardized variables) and the Varimax rotated factor loadings are shown in the following table:

	Estimated factor loadings		Rotated estimated factor loadings	
	f_1	f_2	f_1^*	f_2^*
Taste	0.56	0.82	0.02	0.99
Good buy for money	0.78	-0.52	0.94	-0.01
Flavor	0.65	0.75	0.13	0.98
Suitable for snack	0.94	-0.10	0.84	0.43
Provides lots of energy	0.80	-0.54	0.97	-0.02
Cumulative proportion of total (standardize) sample variance explained	0.574	0.935	0.507	0.935

1. Compute the communalities and the specific variances for the original solution. Comment the results. (1.0)
2. Estimate the sample correlation matrix? (1.0)
3. What is the solution leading to the easiest interpretation? Based on that choice, give an interpretation of each factor. (1.0)
4. Obtain the two first sample principal components and interpret them. (2.0)
5. Compare the results obtained by factor analysis and principal component analysis. (1.0)

Group IV **4.0 points**

An observation x comes from one of the two populations with prior probabilities $P(Y = 0) = 0.6$, $P(Y = 1) = 0.4$, and $X|Y = j \sim Binomial(10, q_j)$, where $j = 0, 1$, $q_0 = 0.3$, and $q_1 = 0.5$.

1. Obtain the classification rule that minimizes the total probability of misclassification. (2.0)
2. Calculate the total probability of misclassification, associated with the previous classification rule. (1.5)
3. Classify an item characterized by $x_0 = 5$. (0.5)