University of Ferrara

Degree Course in "Economics, Markets and Management"

STATISTICAL METHODS for ECONOMICS and BUSINESS – 20 September 2016

Q01

Given the matrices $A = \begin{pmatrix} 1 & 2 \\ -3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1/7 & -2/7 \\ 3/7 & 1/7 \end{pmatrix}$, and the vector $c = \begin{pmatrix} -3 \\ 10 \end{pmatrix}$, what's the result of the product Ac?

$\begin{pmatrix} 17 \\ 19 \end{pmatrix}$ $\qquad\qquad\qquad$ $(19 \quad 17)$ $\qquad\qquad\qquad$ Impossible

$\qquad$ (a) $\qquad\qquad\qquad\qquad\qquad$ (b) $\qquad\qquad\qquad\qquad\qquad$ (c)

Q02
What is the determinant of A?
  a) 0.14.
  b) 7.
  c) 0.

Q03
Is c an eigenvector of B?
a) Yes.
b) No.
c) Yes, under some conditions.

Q04
What's the numerator of the coefficient of determination in the multiple linear regression analysis?
a) total sum of squares.
b) error sum of squares.
c) regression sum of squares.

Q05
Let us consider the following results of a multiple linear regression analysis, where the sale price of used cars (thousands of euros) is function of the production year (X_1), of the mileage (X_2: thousands of kms) and of the type of fuel (X_3: 0=gasoline; 1=diesel fuel), ($\alpha=0.05$):

	Coefficients	p-value
Intercept	12.350	0.000
X_1	2.35	0.024
X_2	-0.32	0.008
X_3	-0.02	0.078

Which of the following statements is false?
a) The type of fuel does not affect the price.
b) A 1000 kms increase in the mileage implies a predicted decrease of sale price equal to 320€.

c) The effect of the production year on the sale price is not significant.

Q06
Which of the following properties of the errors is not an assumption on the multiple linear regression model?
a) Autocorrelation.
b) Normality.
c) Homoscedasticity.

Q07
What is the aggregation rule in Principal Component Analysis?
a) Multiplicative aggregation.
b) Additive aggregation.
c) Fisher combination.

Q08
After the application of the rescaling method, what is the minimum and the maximum value for the transformed variable?
a) -1 and +1.
b) 1 and n.
c) 0 and 1.

Q09
The following table reports some descriptive statistics related to seven states of U.S.A. (2011, American Community Survey).

Informative variables	X_1	X_2	X_3	X_4
Description	Median age	Median household income	% above poverty	% with health insurance
Weights	0.35	0.25	0.30	0.10
California	35.4	57 287	83.4	81.9
Florida	41.1	44 244	83.0	79.1
Georgia	35.5	46 007	80.9	80.4
Illinois	36.8	53 234	85.0	86.9
Lousiana	35.9	41 734	79.6	82.5
Massachusets	39.3	62 859	88.4	95.7
Minnesota	37.6	56 954	88.1	91.2

After the application of the Nonparametric Combination methodology (rescaling type transformation with $c_1=0.01$, $c_2=0.02$, and Fisher combining function), what is the non-normalized value of the quality of life composite index for Minnesota?
a) 1.624.
b) 0.056.
c) 6.820.

Q10
In Factor Analysis, how many factors should be extracted from a dataset with *10* informative variables?
a) 10.
b) 2.

c) It depends on the problem and on the observed data.

Q11

Which of the following goals cannot be achieved with the Principal Components Analysis?
a) Dimensionality reduction of the dataset.
b) Detection heterogeneous groups of homogeneous statistical units.
c) Detection of new informative variables which can replace the observed original variables.

Q12

Which of the following sentences is not true?
a) The factor loadings are the coefficients (weights) of the linear combinations where the original observed variables are expressed as function of the unobserved factors.
b) The factor loadings are measures of the linear dependence between the original unobserved variables and the unobserved factors.
c) The factor loadings are the variances of the unique factors.

Q13

Let us consider the table of Q09. Means and standard deviations of the four variables are shown in the following table:

Informative variables	X_1	X_2	X_3	X_4
Description	Median age	Median household income	% above poverty	% with health insurance
Mean	37.37	51759.86	84.06	85.39
Standard deviation	2.14	7884.52	3.35	6.16

What is the Chebichev or Lagrange distance between California and Florida in terms of quality of life by considering the standardized variables?
a) 2.66.
b) 1.65.
c) 0.12.

Q14

Which of the following sentences about the quality of life distance between California and Florida is true?
a) The Manhattan distance is greater than or equal to the Chebichev distance.
b) The Manhattan distance is less than or equal to the Chebichev distance.
c) The Manhattan distance cannot be computed.

Q15

In the presence of n statistical units, which of the following cluster analysis methods is based on a algorithm with no more than n aggregations?
a) K-means method.
b) Non-hierarchical method.
c) Hierarchical method.