University of Ferrara

Degree Course in "Economics, Markets and Management"

STATISTICAL METHODS for ECONOMICS and BUSINESS – 11 January 2016

Q01
Usually, if a $n \times k$ matrix represents the dataset of a statistical problem, then…
a) there are *n* statistical units and *k* variables
b) there are *n* variables and *k* statistical units
c) there are $n \times k$ statistical units

Q02
If $A = \begin{pmatrix} 2 & 1 & 0 \\ -3 & 1 & 3 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, then the result of the product *Ab* is…
a) $\begin{pmatrix} 2 & 2 & 0 \\ -3 & 2 & 9 \end{pmatrix}$
b) $\begin{pmatrix} 4 \\ 8 \end{pmatrix}$
c) impossible

Q03
If A is a $p \times p$ non-singular matrix and the $p \times p$ matrix B has two equal rows, then…
a) the determinant of the product *AB* is equal to zero
b) the determinant of the product *BA* is not equal to the determinant of the product *AB*
c) the determinant of the product *AB* is equal to one

Q04
Let x_{ij} be the value of the j-th explanatory variable, in a multiple linear regression model, observed on the i-th unit. The least squares estimates of the coefficients of the regression model consist of the values $b_0, b_1, \dots, b_k$ that minimize the function…
a) $\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - \cdots - b_k x_{ik})$
b) $\sum_{i=1}^{n}(y_i + b_0 + b_1 x_{i1} + \cdots + b_k x_{ik})^2$
c) $\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - \cdots - b_k x_{ik})^2$

Q05
Let us consider the following results of a multiple linear regression analysis and let α=0.05 be the significance level:

	Coefficients	p-value
Intercept	245.55	0.017
X1	58.45	0.280
X2	0.27	0.004
X3	28.75	0.042

Which of the following statements is true?
a) variables X1 and X3 significantly affect the dependent variable of the model
b) only variable X2 significantly affects the dependent variable of the model
c) variables X2 and X3 significantly affect the dependent variable of the model

Q06
The R^2 index of a regression model takes value zero when…
a) all the predicted values of the dependent variable are equal
b) all the residuals are equal to zero
c) the least square estimate of the intercept is equal to zero

Q07
The rescaling method is a transformation of a given variable X into a new variable Z …
a) with mean equal to zero and variance equal to one
b) that takes values in the interval *[0,1]*
c) that takes values in the interval *[1,n]* where *n* represents the number of units

Q08
Let us consider a composite indicator for performance evaluation, where the performance is measured according to k>1 aspects (informative variables): the greater the value the better the performance. Which of the following combining functions takes into account only the best of the *k* performances, according to the NonParametric Combinatio methodology?
a) Liptak
b) Logistic
c) Tippett

Q09
Let us consider the following data related to the marks of a student in the exams “Mathematics” (weight=0.4), “Statistics” (weith=0.4) and “Economics” (weight=0.2). The best and the worst performance for each exam correspond to the marks 30 and 18 respectively.

	Weight	Mark	Min	Max
Mathematics	0.4	27	18	30
Statistics	0.4	26	18	30
Eonomics	0.2	21	18	30

By applying the NonParametric Combination methodology, with c_1=0.01 (numerator) and c_2=0.02 (denominator) normalization constants and with the additive combining function, the final score of the composite performance indicator for the considered students is…
a) 0.38
b) 0.62
c) 0.85

Q10
In Factor Analysis, the factor loadings measure…
a) the linear dependence between the original informative variables and the factors
b) the linear dependence between the original informative variables
c) the correlations between the common and the unique factors

Q11
When we wish to reduce the dataset size from the point of view of the variables and all the original informative variables are uncorrelated between each other, then …
a) the Principal Component Analysis provides just one component that “explain” all the data variability
b) the Principal Component Analysis provides linearly dependent components that “explain” all the data variability
c) the Principal Component Analysis is useless

Q12
Let us consider the following output of a Principal Component Analysis related to the first three components:

Component	Eigenvalue	% Variance	Cumulative %
1	6.249	52	52
2	1.229	10	62
3	0.719	6	68

According to these results, which of the following statements is true?
a) the first two components should be excluded from the analysis because the eigenvalues are greater than one
b) the first component “explains” more than half of the global data variability
c) the remaining components “explain” more than half of the global data variability

Q13
In Cluster Analysis, according to the “Single Linkage” or “Nearest Neighbour” method, the distance between two clusters C_1 and C_2 correspond to…
a) the maximum distance between one unit of C_1 and one unit of C_2
b) the mean distance between one unit of C_1 and one unit of C_2
c) the minimum distance between one unit of C_1 and one unit of C_2

Q14
In a Cluster Analysis, given the partition of *100* units in *5* clusters, the proportion of variability explained by this partition is equal to 0.6 (60%). Which of the following statements is consistent with these data?
a) the deviance within the clusters (WD) corresponds to 60% of the deviance between the clusters (BD)
b) the deviance between the clusters (BD) corresponds to 60% of the total deviance (TD)
c) the deviance within the clusters (WD) corresponds to 60% of the total deviance (TD)

Q15
Which of the following graphs is commonly used for representing a hierarchical cluster analysis?
a) scatter diagram
b) scree plot
c) dendogram