9. Why Factor Analysis is required ? Describe the Factor Analysis in brief with stating suitable assumptions. How it is useful in the field of Biostatistics ? [10]

----- x -----

Question Paper Code : 6404

M.A./M.Sc. (Sem.-IV) Examination, 2018

BIOSTATISTICS

[Second Paper]

(Applied Multivariate Analysis)

Time : Three Hours] **[Maximum Marks : 70**

Note : Attempt total **five** questions by selecting **one** question from each of the unit. Question **No.1** is **compulsory**. Use of non-scientific calculator is allowed.

1. Define the following in short: [3x10=30]

   (a) Canonical Correlation

   (b) Principal Component Analysis

   (c) Factor Analysis

   (d) Cluster Analysis

   (e) Multiple correlation coefficient

   (f) Partial correlation coefficient

   (g) Scatter matrix plot

(h) Problem of classification

(i) Problem of Multicollinearity

(j) Dummy variable

UNIT-I

2. (a) What do you understand by multiple regression? Give the estimate of parameters for multiple regression model. Also state their properties.[5]

(b) For three variables X_1, X_2 andX_3 the simple correlation coefficients are r_{12}=0.59, r_{13}=0.67 and r_{23}=0.90. Compute all possible partial correlation coefficients. [5]

3. Compute all possible multiple correlation coefficient for the data given in question 2(b). Also test their significance assuming that the sample size is 15. Given $t_{13,0.05}$=2.13. [10]

UNIT-II

4. (a) What do you understand by Hotteling's T2 statistics ? Define the tests for mean vector under single sample problem. [5]

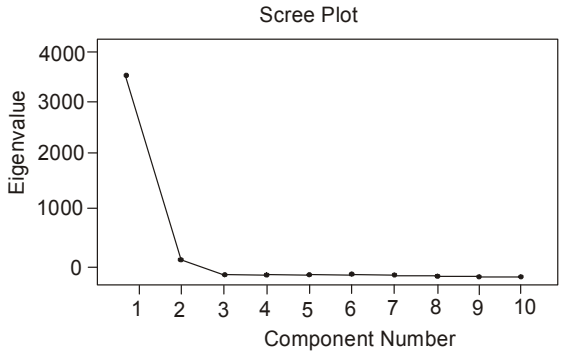
Component Loading						
Variable	Component					
	1	2	3	4	5	6
Vehicle type	0.02	0.26	0.08	0.74	0.17	-0.13
Price in thousands	0.85	-0.25	0.46	-0.05	0.00	0.00
Engine size	0.84	0.25	-0.04	0.19	0.13	0.07
Horsepower	1.00	-0.01	-0.03	0.00	0.00	0.00
Wheelbase	0.30	0.85	0.17	0.27	-0.28	-0.02
Width	0.54	0.58	0.03	0.21	0.09	0.56
Length	0.40	0.90	0.07	-0.15	0.06	-0.01
Curb weight	0.63	0.45	0.27	0.37	0.22	0.04
Fuel capacity	0.52	0.46	0.26	0.53	0.25	0.00
Fuel effeciency	-0.62	-0.28	-0.12	-0.56	-0.37	0.13

Component Score Coefficient Matrix [a]						
	Component					
Variable	1	2	3	4	5	6
Vehicle type	0.00	0.00	0.00	0.01	0.00	-0.01
Price in thousands	0.05	-0.24	1.89	-0.48	0.02	0.20
Engine size	0.00	0.00	0.00	0.01	0.02	0.02
Hosepower	0.94	-0.22	-1.84	0.09	-0.72	-0.42
Wheelbase	0.01	0.24	0.20	0.81	-1.84	-0.35
Width	0.00	0.03	0.01	0.13	0.11	1.63
Length	0.02	0.77	0.24	-1.37	1.21	-0.50
Curb weight	0.00	0.00	0.00	0.01	0.01	0.00
Fuel capacity	0.00	0.03	0.08	0.41	0.42	0.01
Fuel efficiency	0.00	-0.02	-0.04	-0.52	-0.76	0.57

UNIT-IV

8. Write the steps involved in performing the PCA using SPSS. By applying the PCA on a dataset using SPSS following results are obtained. Interpret these without rewriting the tables. [10]

Total Variance Explained			
Components	Initial Eigenvalues [a]		
	Total	% of Variance	Cumulative %
1	3426.42	91.97	91.97
2	212.42	5.7	97.67
3	50.59	1.36	99.03
4	19.89	0.53	99.56
5	9.14	0.25	99.81
6	4.10	0.11	99.92
7	2.67	0.07	99.99
8	0.20	0.01	100.00
9	0.08	0.00	100.00
10	0.04	0.00	100.00



(b) Given that using a sample of size 25, from a multivariate normal population with mean vector μ and variance-covariance matrix $\sum$, the mean vector and variance-covariance matrix are obtained as :

$$\bar{X} = \begin{bmatrix} 25 \\ 20 \\ 30 \end{bmatrix}, \sum = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 40 \end{bmatrix}$$

Test the null hypothesis that $H_0 : \mu = \begin{bmatrix} 28 \\ 25 \\ 32 \end{bmatrix}$.

Given that $F_{3.22,0.05}$ =3.05. [5]

5. (a) For the data given in question 4 (b) test the symmetry of the mean vector. $F_{3.25,0.05}$ =2.99.[5]

(b) For the data given in question 4 (b) test the null hypothesis $H_0 : \mu_1 + 2\mu_2 + 3\mu_3 = 75$ and $\mu_1 + \mu_2 = 45$, where μ_1, μ_2 and μ_3 are elements of mean vector μ.

Given that $F_{2,25,0.05}$ =3.39. [5]

UNIT-III

6. What do you understand by Discriminant analysis ? Also define its use in the field of Biostatistics. Obtain the classification rule for classifying an individual into one of two multivariate normally distributed populations. [10]

7. Write down the steps involve in performing the discriminant analysis using SPSS. On applying discriminant analysis on a dataset using SPSS following results are obtained. Interpret these results without rewriting the tables : [10]

Tests of Equality of Group Means				
	F	df1	df2	Sig.
Years with current employer	60.76	1.00	698	<0.01
Years at current address	19.40	1.00	698	<0.01
Household income in thousands	3.53	1.00	698	0.06
Debt to income ratio (x100)	124.9	1.00	698	<0.01
Credit card debt in thousands	44.47	1.00	698	<0.01

	Box's M	400.64
F	Approx.	26.42
	df1	15.00
	df2	484853.56
	Sig.	0.00

Classification Results					
			Predicted Group Membership		
Previously defaulted			No	Yes	Total
Original	Count	No	393	124	517
		Yes	43	140	183
	%	No	76.0	24	100.0
		Yes	23.5	76.5	100.0
Cross validated	Count	No	392	125	517
		Yes	46	137	183
	%	No	75.8	24.2	100.0
		Yes	25.1	74.9	100.0

Classification Function Coefficients		
	Previously Defaulted	
	No	Yes
Years with current employer	0.18	-0.01
Years at current address	0.13	0.08
Household income in thousands	0.04	0.05
Debt to income ratio (x100)	0.35	0.46
Credit card debt in thousands	-1.06	-0.66
(Constant)	-4.02	-4.55