Classification Results[a,c]

		User_type	Predicted Group Membership		Total
			Heavy	Light	
Original	Count	Heavy	19	6	25
		Light	7	18	25
	%	Heavy	76.0	24.0	100.0
		Light	28.0	72.0	100.0
Cross-validated[b]	Count	Heavy	16	9	25
		Light	8	17	25
	%	Heavy	64.0	36.0	100.0
		Light	32.0	68.0	100.0

(a) 74.0% of original grouped cases correctly classified.

(b) Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

(c) 66.0% of cross-validated grouped cases correctly classified. (10)

6. (a) 'A factor analysis attempts to identify underlying variables or factors and justify dropping variables to shorten questionnaires' Explain.

(b) Describe Cluster Analysis. Explain how an insurer provider would use it to detect fraudulent claims and a bank will use it for credit scoring. (10)

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 2856

Unique Paper Code : 61013924

Name of the Paper : (SEC–4) Statistical Software Package

Name of the Course : **Bachelor of Management Studies (BMS), 2019 (CBCS)**

Semester : IV

Duration : 3 Hours Wednesday

Maximum Marks : 50

8/5/19

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt any **Five** questions.
3. **All** questions carry equal marks.
4. Parts of a question must be answered together.

1. State with reasons whether the following statements are True or False :

(a) For testing hypothesis for equality of the two means using t statistics, the p value as obtained in the output window is for a one tail test.

(b) An alternative hypothesis while testing the equality of two populations could be written as H_1: $\mu_1=\mu_2$

(c) ANOVA is an extension of the t test for comparing the means of two independent populations.

(d) For the application of a chi square test, the expected frequency in each cell should be at least five.

(e) Chi-Square test involves the population distribution to be normal.

(f) The significance of the individual regression coefficients is tested by a t statistic.

(g) There is no distinction between a dependent and independent variable while conducting factor analysis.

(h) Factor loading gives the correlation coefficient between a factor score and components extracted.

(i) In discriminant analysis, the dependent variable is measured as scale in SPSS.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Instagram	.991	.444	1	48	.509
Twitter	.815	10.898	1	48	.002
LinkedIn	.822	10.377	1	48	.002
Facebook	.986	.689	1	48	.411

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.366[a]	100.0	100.0	.517

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.732	14.331	4	.006

Canonical Discriminant Function Coefficients

	Function
	1
Instagram	.097
Twitter	-.088
LinkedIn	.539
Facebook	.093
(Constant)	-3.520

Unstandardized coefficients

(ii) Construct the equation of multiple regression

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	10.197	6	1.700	25.898	.011[a]
	Residual	.197	3	.066		
	Total	10.394	9			

a. Predictors: (Constant), CR, FATA, DDTD, NB, LOGtd, NBB

b. Dependent Variable: ROA.

Source: SPSS output result.

(b) Using the above table, state the hypothesis and comment on the rejection of same at 5 % level of significance. (10)

5. State the hypothesis and interpret the results for the following problem :

A study was conducted to identify the variables which distinguish between heavy/light users of social networking sites among students. A questionnaire was designed for the same. The social networking sites considered for the study were facebook, LinkedIn,Twitter and Instagram.The online survey was conducted on 50 students in the age group 20-30. They were asked to mention the number of hours they spent daily on these sites.

(j) A dendogram is a tree like diagram that graphically presents the cluster results. (10)

2. (a) What is a chi square test?

(b) What are the different types of chi square tests? Explain the steps for conducting each of them.

(c) State the hypothesis and interpret the results for the following problem :

Genetic theory states that children having one parent of blood type A and the other of blood type B will always be of one of three types, A, AB, B and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one parent with A blood group and other with B blood group , 30 per cent were found to be types A, 45 per cent per cent type AB and remainder type B. Test the claim that the ratio given by the genetic theory is correct

Test Statistics

	Blood_Type
Chi-Square	4.500[a]
Df	2
Asymp. Sig.	.105

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 75.0. (10)

3. (a) What are the different types of T tests that can be conducted on a sample data explain giving suitable examples?

(b) State the hypothesis and interpret the results for the following problem :

A study of the effect of caffeine on muscle metabolism used eighteen male volunteers who each underwent arm exercise tests. Nine of the men were randomly selected to take a capsule containing pure caffeine one hour before the test. The other men received a placebo capsule. During each exercise the subject's respiratory exchange ratio (RER) was measured. (RER is the ratio of CO_2 produced to O_2 consumed and is an indicator of whether energy is being obtained from carbohydrates or fats).

The question of interest to the experimenter was whether, on average, caffeine changes RER.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RER	Equal variances assumed	.322	.576	2.481	25	.020	6.333	2.553	1.076	11.590
	Equal variances not assumed			2.208	12.137	.047	6.333	2.869	.090	12.576

(10)

4. Using the following SPSS output tables answer the respective questions :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.139	2.808		2.186	.030
	FEMALE	5.493	.875	.289	6.274	.000
	reading score	.125	.065	.136	1.931	.055
	math score	.238	.067	.235	3.547	.000
	science score	.242	.061	.253	3.986	.000
	social studies score	.229	.053	.260	4.339	.000

a. Dependent Variable: writing score

(a) Using the above table answer the following:

(i) State the hypotheses and comment on rejection at 5% level of significance.