St. Xavier's College (Autonomous), Ahmedabad - 380009

(Affiliated to Gujarat University)

BDA Semester – II ESE Practical June 2021

Subject Name**:** Advanced Statistical Methods   Sub. Code: PBD 2802L

Date & Time: 21/06/21 & 09:00 AM to 11:00 AM   Total Marks: 14+10(viva)

N.B: (1) Attempt any TWO full questions (including sub-options).

(2) Save r script file with name ESE 2802L Roll number (e.g. ESE 2802L PBD-01)

(3) Write appropriate hypothesis and conclusions whenever required.

Q.1   **Answer the following questions.**

(A) The following are the values of the cephalic index found in two samples of skulls, one consisting of 14 and the other of 12 individuals.

Sample – I									
74.1	77.7	74.0	74.4	73.8	79.3	75.8	82.8	72.2	75.2
74.2	77.1	78.4	76.3						

Sample – II									
70.8	73.9	74.2	70.4	69.2	71.2	75.8	72.4	76.4	78.1
72.8	73.3								

Assuming that the distribution in the two samples are normal, answer the following:

(a) Is it possible that average cephalic index in sample II is 75?

(b) Is it reasonable to say that the average of index in the two samples is the same?

(B) The manager of a car plant wishes to investigate how the plant's electricity usage depends upon the plant's production.

Month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Production ($ million)	4.5	3.5	4.3	5.1	5.6	4.9	5.3	5.8	4.7	5.6	4.9	4.2
Electricity usage (million kWh)	2.5	2.3	2.5	2.7	2.9	3.1	3.2	3.5	3.1	3.3	2.6	2.5

(i) Find regression line of electricity usage depends upon the plant's production.

(ii) Predict production ($ million) when electricity usage is 2.8 and 3.4 (million kWh).

(C) Find the optimum solution of the following LPP by simplex method:

Minimise $Z = 5x_1 - 6x_2 + 4x_3$

Subject to constraints: $3x_1 + 4x_2 + 6x_3 \geq 9$,

$\qquad\qquad\qquad\quad x_1 + 3x_2 + 2x_3 \geq 5$,

$\qquad\qquad\qquad\quad 7x_1 - 2x_2 - x_3 \leq 10$,

$\qquad\qquad\qquad\quad x_1 - 2x_2 + 4x_3 \geq 4$,

$\qquad\qquad\qquad\quad 2x_1 + 5x_2 - 3x_3 = 3$,

$\qquad\qquad\qquad\quad x_1, x_2, x_3 \geq 0$.

Q.2   **Answer the following questions.**

(A)   We are interested in how the heat evolved in the curing of cement is affected by the amounts of various chemical included in the cement mixture. The independent and dependent variables are listed below:

X_1 = amount of tricalcium aluminate,

X_2 = amount of tricalcium silicate,

X_3 = amount of tetracalcium alumino ferrite,

X_4 = amount of dicalcium silicate,

Y = heat evolved in calories per gram of cement.

X_1	7	1	11	11	7	11	3	1	2	21	1	11	10
X_2	25	28	55	30	51	54	70	30	53	46	39	65	67
X_3	6	15	8	8	6	9	17	22	18	4	23	9	8
X_4	60	52	20	47	33	22	6	44	22	26	34	12	12
Y	79	74	104	88	96	109	103	73	93	116	84	113	109

Fit multiple regression model. Apply forward selection method and check forward selection method model and original model are same or not?

(B)   The adjoining table gives the results of an experiment on the effects of five manual treatments I, II, III, IV, V on the yield of sugarcane.

I (552)	III (431)	IV (425)	V (572)	II (451)
II (405)	I (525)	V (463)	IV (441)	III (481)
V (471)	II (492)	I (472)	III (381)	IV (410)
IV (430)	V (469)	III (432)	II (467)	I (460)
III (325)	IV (445)	II (429)	I (413)	V (493)

Write down the ANOVA table. Test whether the treatments are equally effective and test for the differences between variety effects.

Q.3   **Answer the following questions.**

(A)   The given table show information about bike purchase (1 = purchase, 2 = would not purchase), score out of 100 for durability of the bike, performance of the bike and looks of the bike. Predict the model using binary logistic regression.

Buyer	Durability	Performance	Looks	Buyer	Durability	Performance	Looks
1	75	84	58	2	49	39	69
1	82	67	54	2	48	43	68
1	80	68	48	2	43	38	21
1	83	57	33	2	45	78	24
1	66	56	37	2	33	73	49
1	59	43	41	2	34	54	46
1	58	38	39	2	31	48	52
1	57	39	19	2	29	37	49
1	92	78	20	2	27	39	32
1	91	76	41	2	41	19	34

(B) Suppose there are two coaching classes that prepare students for entrance test for admission into professional courses. Let's assume that these students belong to arts, commerce and science faculty only. Marks Scored in Entrance Test will be the response variable. Classes and faculty will be two independent factors. Classes will have 2 levels and faculty will have 3 levels. 20 students are randomly selected taking coaching from different classes and have different faculties. The dataset is given below:

Marks	classes	Faculty	Marks	cclasses	Faculty
75	Class1	Arts	66	Class2	Arts
76	Class1	Commerce	65	Class2	Commerce
91	Class1	Science	77	Class2	Science
78	Class1	Arts	64	Class2	Arts
72	Class1	Commerce	62	Class2	Commerce
94	Class1	Science	76	Class2	Science
76	Class1	Arts	57	Class2	Arts
73	Class1	Commerce	60	Class2	Commerce
92	Class1	Science	72	Class2	Science
79	Class1	Arts	63	Class2	Arts

Analyse the above data using appropriate ANOVA method.