Practical Examinations-Sem-II (2022-23)

STAT-567-Regression Analysis-Cr (1+1)

Time: 2-Hours	Max. Marks: 30 [15 + 15](Practical Record: 10, Viva-voce: 10)

Note: *Please solve ANY TWO of the following questions. Practical Record and Viva-voce parts of examinations would be examined subsequently.* ***Please do not use lm() model unless allowed for***.

Problem # 1.

The following data in Table 1, "Cyliner's breaking strengh" were taken to test the effect of adding a small percentage of coal dust to the breaking strength of sand used for making concrete. Several batches were mixed under practically identical conditions except for the variation in the percentage of coal. From each batch, four cylinders were made and tested for breaking strength in pounds per square inch (lb/in^2). One cylinder in the third sample was defective, so there were only three items in this sample:

Table 1. Cyliner's breaking strengh

Sample	1	2	3	4	5
Perc.Coal	0	0.05	0.1	0.5	1
Breaking Strength	1690	1550	1625	1725	1530
	1580	1445	1450	1550	1545
	1745	1645	1510	1430	1565
	1685	1545		1445	1520

For the data of Table 1, "Cyliner's breaking strengh" carry out the following analyses: (15)

1. Enter the data of in a suitable format. (2)
2. Carry out the regression analysis along with requisite ***F***-tests of hypothesis of breaking strength of cylinder on the percent coal mixed into it, using formulae approach. (5)
3. Carry out the Barlett's test of the residuals and draw conclusion about the homogeneity of 5-classes of samples due to different groupgs of percentage coal mixed into the cylinders, using formulae approach. (4)
4. Carry out the Durbin-Watson test on the residuals of the above mentioned regression analysis, and draw conclusions of the test, using formulae approach. (4)

Problem # 2.

From a study on World Review of Nutrition and Dietetics in 1972 in USA, 72 observations in Table 2, "Boy's Weight Height Ratio" are given, in which the response variable is ***Y** = boy's weight:height ratio (W:H)*, for equally spaced values of the corresponding predictor variable ***X** = age in months*. Assume that the observations fall into three groups each of 24 observations in order. Assuming further, that the three groups of data can be explained by three straight lines of time trends.

Table 2. Boy's Weight Height Ratio

Group.1			**Group.2**			**Group.3**		
S.No.	W.H	Age	S.No.	W.H	Age	S.No.	W.H	Age
1	0.46	0.5	25	0.88	24.5	49	0.92	48.5
2	0.47	1.5	26	0.81	25.5	50	0.96	49.5
3	0.56	2.5	27	0.83	26.5	51	0.92	50.5
4	0.61	3.5	28	0.82	27.5	52	0.91	51.5
5	0.61	4.5	29	0.82	28.5	53	0.95	52.5
6	0.67	5.5	30	0.86	29.5	54	0.93	53.5
7	0.68	6.5	31	0.82	30.5	55	0.93	54.5
8	0.78	7.5	32	0.85	31.5	56	0.98	55.5
9	0.69	8.5	33	0.88	32.5	57	0.95	56.5
10	0.74	9.5	34	0.86	33.5	58	0.97	57.5
11	0.77	10.5	35	0.91	34.5	59	0.97	58.5
12	0.78	11.5	36	0.87	35.5	60	0.96	59.5
13	0.75	12.5	37	0.87	36.5	61	0.97	60.5
14	0.8	13.5	38	0.87	37.5	62	0.94	61.5
15	0.78	14.5	39	0.85	38.5	63	0.96	62.5
16	0.82	15.5	40	0.9	39.5	64	1.03	63.5
17	0.77	16.5	41	0.87	40.5	65	0.99	64.5

18	0.8	17.5	42	0.91	41.5	66	1.01	65.5
19	0.81	18.5	43	0.9	42.5	67	0.99	66.5
20	0.78	19.5	44	0.93	43.5	68	0.99	67.5
21	0.87	20.5	45	0.89	44.5	69	0.97	68.5
22	0.8	21.5	46	0.89	45.5	70	1.01	69.5
23	0.83	22.5	47	0.92	46.5	71	0.99	70.5
24	0.81	23.5	48	0.89	47.5	72	1.04	71.5

Using the methods of Dummy variables, and lm() model, carry out the following analyses: (15)

1. Enter the data of Table 2, "Boy's Weight Height Ratio" in a suitable format as required by analyses. (2)
2. Estimate a common regression line for the three groups combined. Give your interpretations of the estimated coefficients. (2)
3. Find the three regression lines from the above-estimated common regression line, separately for each of the groups. Give your interpretations of the estimated coefficients. (2)
4. Plot the data points as well as the fitted lines of the three groups along with their points of intersections ***on the same graph***. (3)
5. Also, provide the analysis of variance table. (2)
6. Find and analyze the residuals through diagnostic plots. (2)
7. Give your final interpretations (2)

Problem # 3.

Draper and Smith provides Hald data in Appendix 15A, in his celebrated book on Applied Regres- sion Analysis on page 348, which is reproduced below in Table 3, "Hald Data". The variables shown in the table are:

X_1 = amount of tricalcium aluminate, $3CaO.Al_2O_3$
X_2 = amount of tricalcium silicate, $3CaOSiO_2$.
X_3 = amount of tetracalcium alumino ferrite, $3CaO.Al_2O_3.Fe_2O_3$
X_4 = amount of dicalcium silicate, $2CaOSiO_2$,
(Response)Y = heat evolved in calories per gram of cement.

X_1, X_2, X_3 and X_4 are measured as percent of the weight of the clinkers from which the cement was made.

Table 3. Hald Data

x1	x2	x3	x4	y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

Using matrix approach of analysis, look at the above mentioned Table 3, "Hald Data" data,

1. Fit the eqution $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$ (03)
2. Compute the standard errors of estimators of b_1, b_2, b_3, b_4. (03)
3. Test the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus the alternative not so. (4.5)
4. Also, test the linear hypothesis $H_0 : \beta_1 = \beta_3; \beta_2 = \beta_4$ versus the alternative not so. (4.5)