Deepthi.M
20BPS1133

CSE 3505-FDA DA2

Suppose you have the following dataset

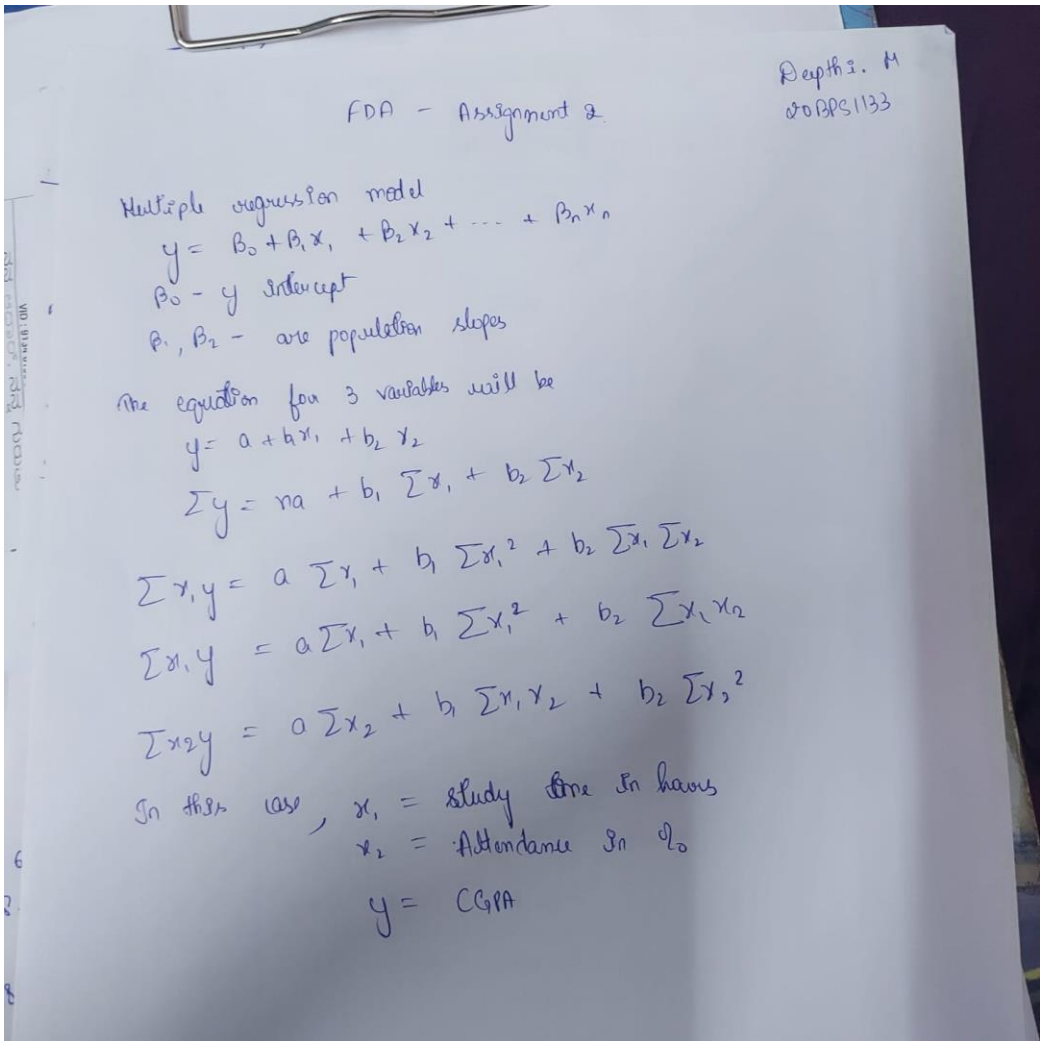
Registration Number	Study Time in Hrs	Attendance in %	CGPA
1	4	100	6.2
2	6	50	5.3
3	16	95	9.9
4	12	85	9.0
5	18	100	10.0
6	2	50	4.0
7	5	70	6.9
8	9	80	7.7
9	15	80	8.9
10	3	75	6.5
11	7	7.5	8.0

The above table shows the marks obtained by students based on their study hours and attendance in the class.

---

a) Derived the multiple regression equation to predict CGPA based on study Time and Attendance.

Deepthi.M
20BPS1133



FDA - Assignment 2

Deepthi. M
20BPS1133

Multiple regression model

$y = B_0 + B_1 x_1 + B_2 x_2 + \cdots + B_n x_n$

B_0 - y intercept

B_1, B_2 - are population slopes

The equation for 3 variables will be

$y = a + b_1 x_1 + b_2 x_2$

$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2$

$\sum x_1 y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 \sum x_2$

$\sum x_1 y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2$

$\sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2$

In this case, x_1 = Study time in hours

x_2 = Attendance in %

y = CGPA

Deepthi.M
20BPS1133

y	x_1	x_2	y^2	x_1^2	x_2^2	yx_1	x_1x_2	yx_2
6.2	4	100	38.44	16	10000	24.8	400	620
5.3	6	50	28.09	36	2500	31.8	300	265
9.9	16	95	98.01	256	9025	158.4	1520	940.5
9	12	85	81	144	7225	108	1020	765
10	18	100	100	324	10000	180	1800	1000
4	2	50	16	4	2500	8	100	200
6.9	5	70	47.61	25	4900	34.5	350	483
7.7	9	80	59.29	81	6400	69.3	720	616
8.9	15	80	79.21	225	6400	133.5	1200	712
6.5	3	75	42.25	9	5625	19.5	225	487.5
8.0	7	7.5	64	49	56.25	56	52.5	60
80.4	97	792.5	6539	1169	64631.25	823.8	7187.5	6149

$$80.4 = 11 \times a + b_1 \times 97 + b_2 \times 792.5$$

$$823.8 = 97a + 1169 b_1 + 76875 b_2$$

$$6149 = 792.5a + 64631.25 b_2 + 7687.5 b_1$$

from 3 equations we get

$$a = 4.78077$$

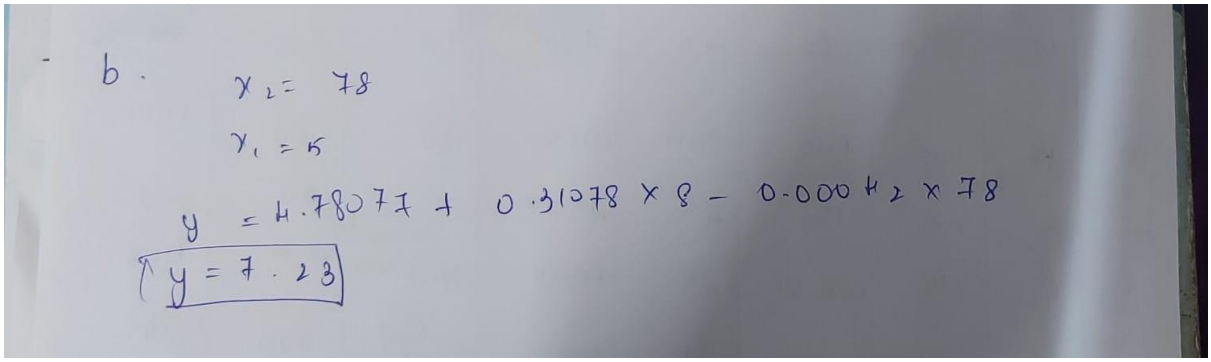
$$b_1 = 0.31078$$

$$b_2 = -0.00042$$

$$y = 4.78077 + 0.31078 x_1 - 0.00042 x_2$$

Deepthi.M
20BPS1133

b) Apply multiple regression to predict the CGPA of a student if he has 78% attendance and 8hr Study time.



c) Finally write an R script to perform the multiple regression and predict the CGPA of the student as per the condition given in bit (b).

Code:

```
x1<-c(4,6,16,12,18,2,5,9,15,3,7)

x2<-c(100,50,95,85,100,50,70,80,80,75,7.5)

y<-c(6.2,5.3,9.9,9,10,4,6.9,7.7,8.9,6.5,8)

input<-data.frame(y,x1,x2)

input

regM<-lm(y~x1+x2,data=input)

regM

summary(regM)
```

```
> regM<-lm(y~x1+x2,data=input)
> regM

Call:
lm(formula = y ~ x1 + x2, data = input)

Coefficients:
(Intercept)           x1           x2
  4.7950347    0.3114073   -0.0006964

> summary(regM)

Call:
lm(formula = y ~ x1 + x2, data = input)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3830 -0.4206  0.1886  0.5620  1.0303

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.7950347  0.8073421   5.939 0.000346 ***
x1           0.3114073  0.0572976   5.435 0.000620 ***
x2          -0.0006964  0.0116896  -0.060 0.953957
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9037 on 8 degrees of freedom
Multiple R-squared:  0.8217,    Adjusted R-squared:  0.7771
F-statistic: 18.44 on 2 and 8 DF,  p-value: 0.00101
```

Deepthi.M
20BPS1133

```
> cgpa=4.78077+0.31078*8-0.00042*78
> cgpa
[1] 7.23425
> 
```

d) Interpret the results and various statistics measures obtained after executing the script and attach the outputs.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.7950347  0.8073421   5.939 0.000346 ***
x1           0.3114073  0.0572976   5.435 0.000620 ***
x2          -0.0006964  0.0116896  -0.060 0.953957
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9037 on 8 degrees of freedom
Multiple R-squared:  0.8217,    Adjusted R-squared:  0.7771
F-statistic: 18.44 on 2 and 8 DF,  p-value: 0.00101
```