Name: Mrinal Chaudhari **Homework-1** **STAT-641**

Part I

Q1: Ans: E: All of the statements are true
Q2: Ans: B: y decreases by 3 when x increases by 4
Q3: Ans: E: All of the above statements are true.
Q4: Ans: A: least-squares method
Q5: Ans: B: the observed values of the response variable y and the estimated values ˆy
Q6: Ans: C: 1.600
Q7: Ans: D: 9.76
Q8: Ans: E: All of the above statements are true
Q9: Ans: C: 0.667
Q10: Ans: C: 0.970
Q11: Ans: A: The coefficient of determination, denoted by R2 is interpreted as the proportion of observed y variation that cannot be explained by the simple linear regression model.
Q12: Ans: B: $\sqrt{SSE}\,/(n\ -2)$
Q13: Ans: C: $-2.878 \leq\ t\ \leq\ 2.878$
Q14: Ans: A: .01

Part II

Q.1) Table B.1 gives data concerning the performance of the 26 National Football League teams in 1976. It is suspected that the number of yards gained rushing by opponents (x_8) has an effect on the number of games won by a team (y).
a. Fit a simple linear regression model relating games won y to yards gained rushing by opponents x_8.
b. Construct the analysis-of-variance table and test for significance of a regression.
c. Find a 95% CI on the slope.
d. What percent of the total variability in y is explained by this model?
e. Find a 95% CI on the mean number of games won if opponents' yards rushing is limited to 2000 yards.

Solution:

A: Regression Equation

y=21.79 - 0.00703 x_8

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

TABLEB.1

Regression Analysis: y versus x8

Regression Equation

y = 21.79 - 0.00703 x8

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	21.79	2.70	8.08	0.000	
x8	-0.00703	0.00126	-5.58	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.39287	54.47%	52.72%	47.81%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	178.1	178.092	31.10	0.000
x8	1	178.1	178.092	31.10	0.000
Error	26	148.9	5.726		
Total	27	327.0			

B:

Since P-value <0.05 we reject the null hypothesis, H_0

Hence regression is significant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	178.1	178.092	31.10	0.000
x8	1	178.1	178.092	31.10	0.000
Error	26	148.9	5.726		
Total	27	327.0			

C: Find a 95% CI on the slope.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	21.79	2.70	8.08	0.000	
x8	-0.00703	0.00126	-5.58	0.000	1.00

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

α=1-95/100=0.05
p=1-α/2=0.975
df=n-2=28-2=26
tcrit=2.056; SE coef= 0.00126

-0.00703±2.056(0.00126)=(-0.0096,-0.0044)
95% slope ont the C.I (-0.0096,-0.0044)

D:

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.39287	54.47%	52.72%	47.81%

$$Total\ variability\ in\ y\ explained\ by\ the\ model\ is: R^2 = 54.47\%$$

TABLEB.1

Regression Analysis: y versus x8

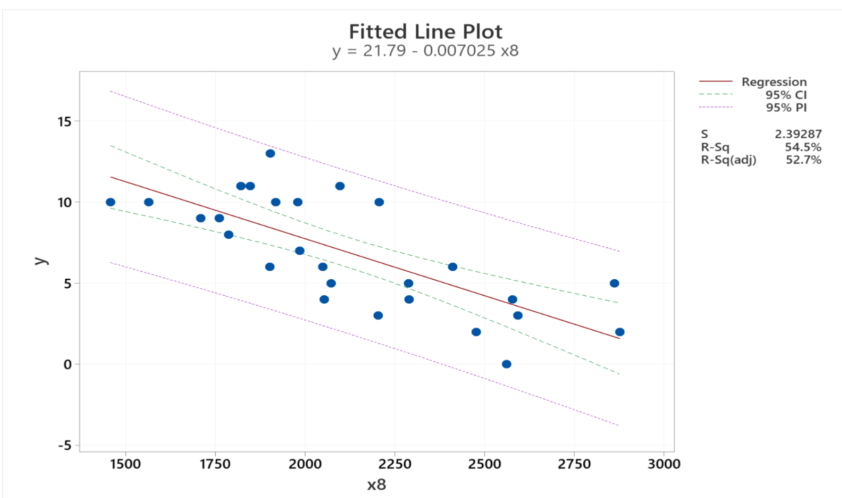
The regression equation is
y = 21.79 - 0.007025 x8

Model Summary

S	R-sq	R-sq(adj)
2.39287	54.47%	52.72%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	178.092	178.092	31.10	0.000
Error	26	148.872	5.726		
Total	27	326.964			



E:

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

TABLEB.1

Prediction for y

Regression Equation

y = 21.79 - 0.00703 x8

Settings

Variable	Setting
x8	2000

Prediction

Fit	SE Fit	95% CI	95% PI
7.73805	0.473015	(6.76575, 8.71035)	(2.72425, 12.7519)

The 95% CI on the mean number of games won by opponent yard rushing a limited to 2000 is (6.76575,8.71035)

Q.3) Table B.2 presents data collected during a solar energy project at Georgia Tech.
a. Fit a simple linear regression model relating total heat flux y (kilowatts) to the radial deflection of the deflected rays x_4 (milliradians).
b. Construct the analysis-of-variance table and test for significance of a regression.
c. Find a 99% CI on the slope.
d. Calculate R^2.
e. Find a 95% CI on the mean heat flux when the radial deflection is 16.5 milliradians.

Solution:
A:
Regression Equation

y=607.1 - 21.40 x4

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

TABLEB.2

Regression Analysis: y versus x4

Regression Equation

y = 607.1 - 21.40 x4

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	607.1	42.9	14.15	0.000	
x4	-21.40	2.57	-8.34	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.3277	72.05%	71.02%	66.93%

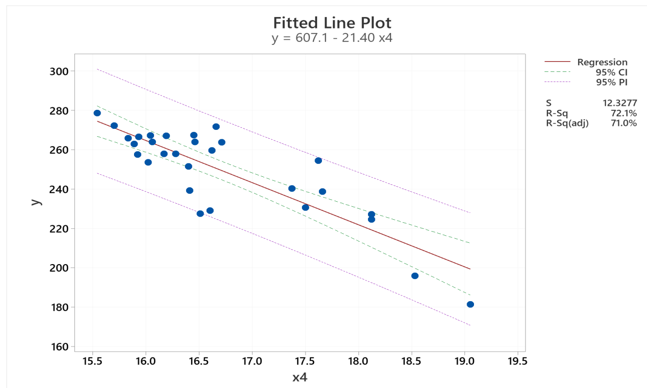
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	10578.7	10578.7	69.61	0.000
x4	1	10578.7	10578.7	69.61	0.000
Error	27	4103.2	152.0		
Lack-of-Fit	26	4100.1	157.7	50.46	0.111
Pure Error	1	3.1	3.1		
Total	28	14681.9			

Fits and Diagnostics for Unusual Observations

Obs	y	Fit	Resid	Std Resid	
22	254.50	229.99	24.51	2.06	R
24	181.50	199.39	-17.89	-1.70	X
25	227.50	253.75	-26.25	-2.17	R

R Large residual
X Unusual X



B:

Since p-value <0.05, we reject the null hypothesis, H_0
The regression is significant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	10578.7	10578.7	69.61	0.000
x4	1	10578.7	10578.7	69.61	0.000
Error	27	4103.2	152.0		
Lack-of-Fit	26	4100.1	157.7	50.46	0.111
Pure Error	1	3.1	3.1		
Total	28	14681.9			

C: Find a 99% CI on the slope

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	607.1	42.9	14.15	0.000	
x4	-21.40	2.57	-8.34	0.000	1.00

α=1-99/100=0.01
p=1-α/2=0.995
df=n-2=27
tcrit=2.771; SE coef= 2.57

-21.40±2.771(2.57)=(-28.58, -14.29)
99% slope ont the C.I (-28.58, -14.29)

D:Calculate R^2.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.3277	72.05%	71.02%	66.93%

R^2=72.05%

E:

TABLEB.2

Prediction for y

Regression Equation

y = 607.1 - 21.40 x4

Settings

Variable	Setting
x4	16.5

Prediction

Fit	SE Fit	95% CI	95% PI
253.963	2.34715	(249.147, 258.779)	(228.214, 279.711)

95% CI when radial deflection is: (249.147, 258.779)

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

Q.7) The purity of oxygen produced by a fractional distillation process is thought to be related to the percentage of hydrocarbons in the main condensor of the processing unit. Twenty samples are shown below.
a. Fit a simple linear regression model to the data.
b. Test the hypothesis H0: β1 = 0.
c. Calculate R^2.
d. Find a 95% CI on the slope.
e. Find a 95% CI on the mean purity when the hydrocarbon percentage is 1.00

Solution:
A:
Regression Equation

Purity%=77.85 + 11.81 Hydrocarbon %

QUESTION7

Regression Analysis: Purity% versus Hydrocarbon %

Regression Equation

Purity% = 77.85 + 11.81 Hydrocarbon %

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	77.85	4.20	18.52	0.000	
Hydrocarbon %	11.81	3.49	3.38	0.003	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.60164	38.86%	35.47%	26.02%

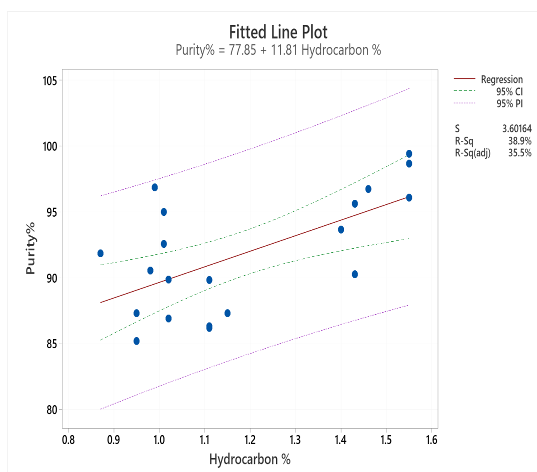
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	148.43	148.433	11.44	0.003
Hydrocarbon %	1	148.43	148.433	11.44	0.003
Error	18	233.49	12.972		
Lack-of-Fit	10	195.07	19.507	4.06	0.029
Pure Error	8	38.43	4.803		
Total	19	381.93			

Fits and Diagnostics for Unusual Observations

Obs	Purity%	Fit	Resid	Std Resid	
18	96.85	89.54	7.31	2.12	R

R Large residual



B: Test the hypothesis H0: β1 = 0

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	148.43	148.433	11.44	0.003
Hydrocarbon %	1	148.43	148.433	11.44	0.003
Error	18	233.49	12.972		
Lack-of-Fit	10	195.07	19.507	4.06	0.029
Pure Error	8	38.43	4.803		
Total	19	381.93			

Since the p-value <0.05, we reject the null hypothesis,
The regression is significant.

C : Calculate R^2

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.60164	38.86%	35.47%	26.02%

$R^2 = 38.86\%$

D: Find a 95% CI on the slope.

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	77.85	4.20	(69.02, 86.69)	18.52	0.000	
Hydrocarbon %	11.81	3.49	(4.47, 19.14)	3.38	0.003	1.00

95% CI on slope is : (4.47,19.14)

E:

QUESTION7

Prediction for Purity%

Regression Equation

Purity% = 77.85 + 11.81 Hydrocarbon %

Settings

Variable	Setting
Hydrocarbon %	1

Prediction

Fit	SE Fit	95% CI	95% PI
89.6599	1.02678	(87.5028, 91.8171)	(81.7917, 97.5282)

95% CI on the mean purity when hydrocarbon percentage is 1=(87.5028,91.8171)

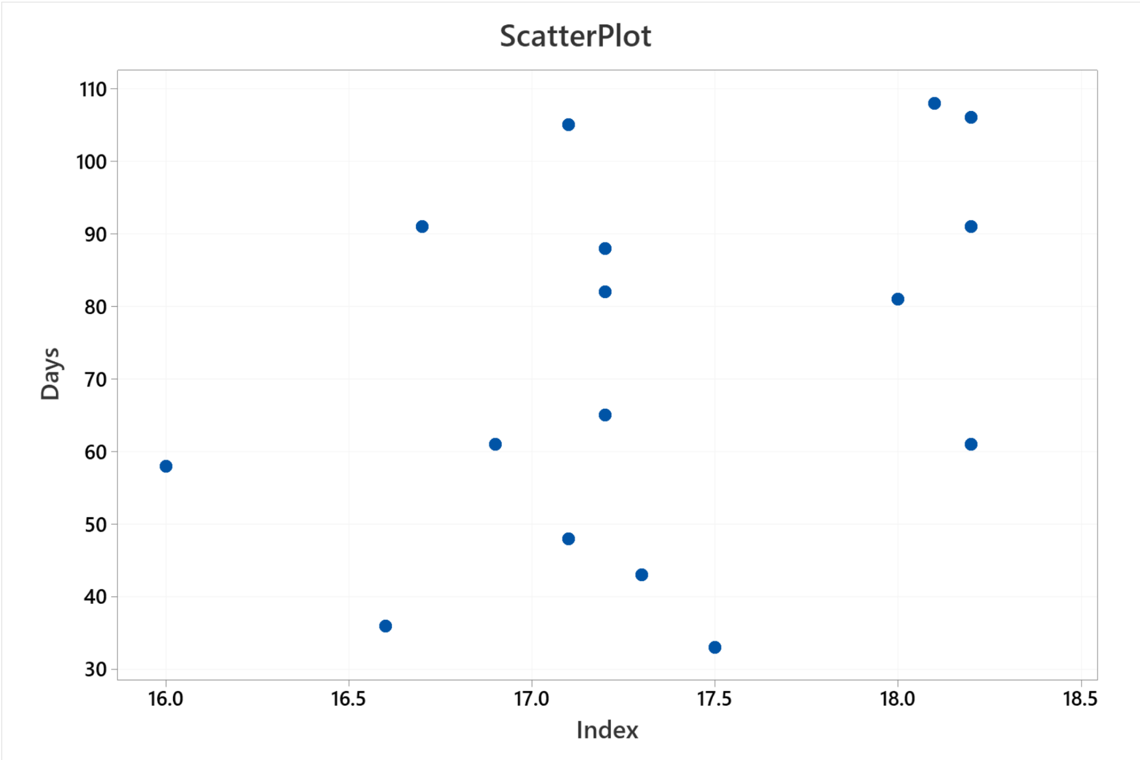
Q.13) Davidson ("Update on Ozone Trends in California's South Coast Air Basin," Air and Waste, 43, 226, 1993) studied the ozone levels in the South Coast Air Basin of California for the years 1976– 1991. He believes that the number of days the ozone levels exceeded 0.20 ppm (the response) depends on the seasonal meteorological index, which is the seasonal average 850-millibar temperature (the regressor). The following table gives the data.
a. Make a scatterplot of the data.
b. Estimate the prediction equation.
c. Test for significance of a regression.
d">† NELSON. Since the II. Calculate and plot the 95% confidence and prediction bands.
Solution:

A:

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

QUESTION13

Scatterplot of Days vs Index



B: Regression Equation

Y= -193 + 15.30x

Name: Mrinal Chaudhari **Homework-1** **STAT-641**

QUESTION13

Regression Analysis: Days versus Index

Regression Equation

Days = -193 + 15.30 Index

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-193	164	-1.18	0.258	
Index	15.30	9.42	1.62	0.127	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
23.7950	15.85%	9.84%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1493	1492.6	2.64	0.127
Index	1	1493	1492.6	2.64	0.127
Error	14	7927	566.2		
Lack-of-Fit	9	4968	552.0	0.93	0.564
Pure Error	5	2959	591.8		
Total	15	9419			

C: Test for significance of a regression

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1493	1492.6	2.64	0.127
Index	1	1493	1492.6	2.64	0.127
Error	14	7927	566.2		
Lack-of-Fit	9	4968	552.0	0.93	0.564
Pure Error	5	2959	591.8		
Total	15	9419			

Since the p-value is >0.05, we reject the null hypothesis H_0.
This indicates that the regression is not significant.

D:

The 95% confidence and prediction band plot is,

Name: Mrinal Chaudhari **Homework-1** **STAT-641**