STA 4163 FINAL PROJECT REPORT

Part I

1) Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
y	100	-359.0466	293.4573	7.991020	128.5962957
x1	100	-1.7919	2.8690	.278830	.9670970
x2	100	-2.1965	2.3702	.155535	1.0068672
x3	100	-2.2986	2.9802	.060189	1.0540885
x4	100	-2.8520	2.5768	.016009	1.1836155
x5	100	-2.6261	2.7164	.038161	1.0625001
Valid N (listwise)	100				

2) i)

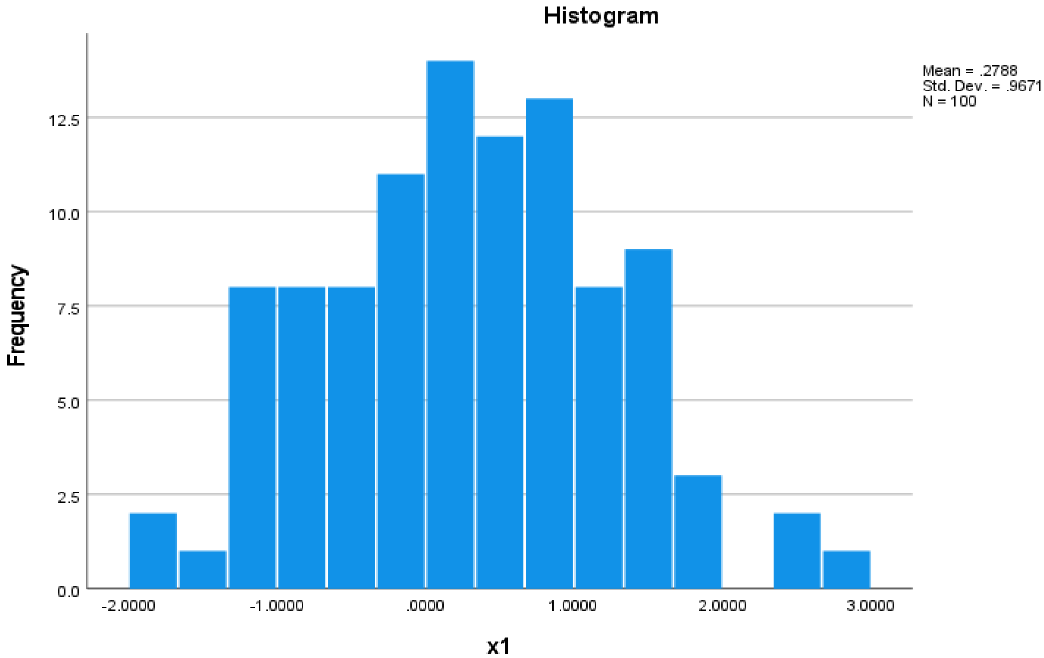
CI values Lower	.086937
Upper	.470723
IQR	1.3571
Skewness	.110

ii) Tests of Normality

Kolmogorov-Smirnova				Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
x1	.038	100	.200*	.993	100	.879

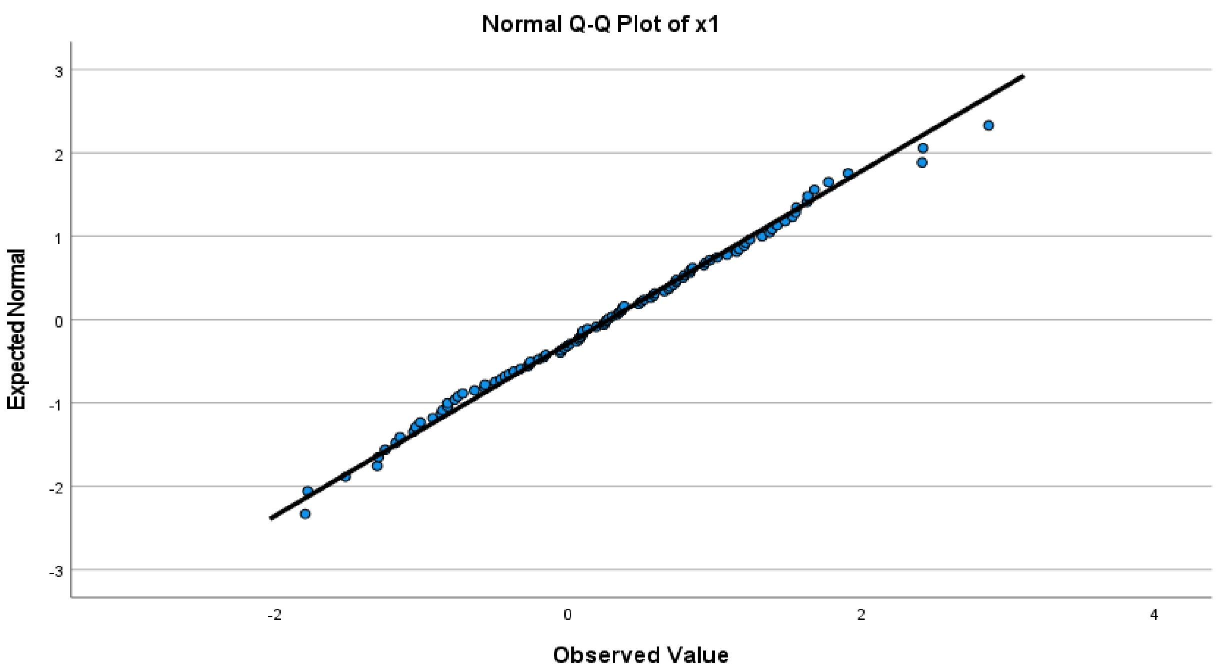
Since the p value in the Kolmogorov- Smirnova test is less than 0.05 and greater than 0.05 in the Shapiro-Wilk test, we can conclude that the data is normal.

iii)



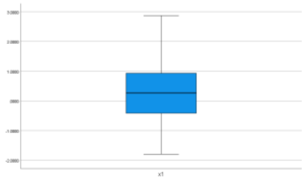
From the histogram of variable x1, we can also infer that the data is normal as the frequency values are distributed within the standard deviation of the data.

iv)



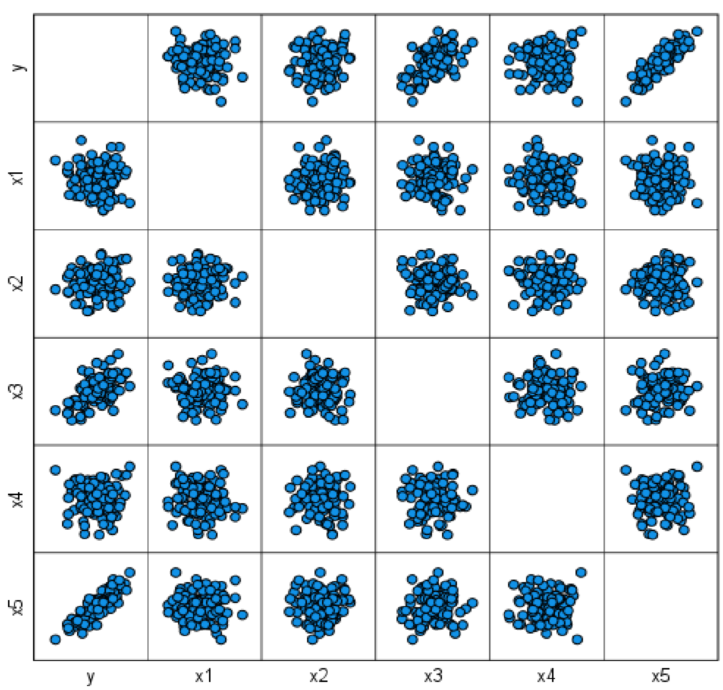
From the QQ-Plot of x1, we can also conclude that the data is normal, as most of the observed values follow the expected normal line.

v) <u>Box-Plot</u>



After analyzing the Box-plot and each quartile, we can conclude that the data is normal.

3) i) <u>Matrix Scatterplot</u>



From the scatterplot matrix above, we can see that there is a positive correlation between y and x5. Meanwhile, there is a low correlation between y and the rest of the quantitative variables.

ii)

Correlations

		y	x1	x2	x3	x4	x5
y	Pearson Correlation	1	-.045	.098	.557**	.002	.877**
	Sig. (2-tailed)		.657	.330	.000	.981	.000
	N	100	100	100	100	100	100
x1	Pearson Correlation	-.045	1	.043	.002	-.106	-.046
	Sig. (2-tailed)	.657		.674	.984	.294	.651
	N	100	100	100	100	100	100
x2	Pearson Correlation	.098	.043	1	-.054	.001	.032
	Sig. (2-tailed)	.330	.674		.596	.995	.753
	N	100	100	100	100	100	100
x3	Pearson Correlation	.557**	.002	-.054	1	-.068	.108
	Sig. (2-tailed)	.000	.984	.596		.504	.286
	N	100	100	100	100	100	100
x4	Pearson Correlation	.002	-.106	.001	-.068	1	.036
	Sig. (2-tailed)	.981	.294	.995	.504		.721
	N	100	100	100	100	100	100
x5	Pearson Correlation	.877**	-.046	.032	.108	.036	1
	Sig. (2-tailed)	.000	.651	.753	.286	.721	
	N	100	100	100	100	100	100

From the Pearson's correlation coefficients, we can see that x3 and x5 have a strong positive correlation coefficient with respect to y. While x2 and x4 have a low positive correlation and x1 has a negative low correlation coefficient.

<u>Part 2</u>

1) ➔<u>Linear regression model between y and x1</u>

➔Linear regression model between y and x1

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	x1[b]	.	Enter

a. Dependent Variable: y

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.045[a]	.002	-.008	129.1197848

a. Predictors: (Constant), x1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3315.675	1	3315.675	.199	.657[b]
	Residual	1633848.045	98	16671.919		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.660	13.443		.719	.474
	x1	-5.984	13.419	-.045	-.446	.657

a. Dependent Variable: y

➔Linear regression model between y and x2

|

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	x2[b]	.	Enter

a. Dependent Variable: y

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.098[a]	.010	.000	128.6227814

a. Predictors: (Constant), x2

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15869.370	1	15869.370	.959	.330[b]
	Residual	1621294.350	98	16543.820		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.035	13.016		.464	.644
	x2	12.574	12.839	.098	.979	.330

a. Dependent Variable: y

➔<u>Linear regression model between y and x3</u>

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	x3[b]	.	Enter

a. Dependent Variable: y

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.557[a]	.310	.303	107.3593749

a. Predictors: (Constant), x3

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	507612.252	1	507612.252	44.040	.000[b]
	Residual	1129551.468	98	11526.035		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x3

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.902	10.754		.363	.717
	x3	67.932	10.236	.557	6.636	.000

➔<u>Linear regression model between y and x4</u>

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	x4[b]	.	Enter

a. Dependent Variable: y

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.002[a]	.000	-.010	129.2503724

a. Predictors: (Constant), x4

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.161	1	9.161	.001	.981[b]
	Residual	1637154.559	98	16705.659		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x4

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.987	12.926		.618	.538
	x4	.257	10.975	.002	.023	.981

➔<u>Linear regression model between y and x5</u>

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	x5[b]	.	Enter

a. Dependent Variable: y

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.877[a]	.769	.767	62.0710390

a. Predictors: (Constant), x5

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1259587.959	1	1259587.959	326.927	.000[b]
	Residual	377575.761	98	3852.814		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x5

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.940	6.211		.634	.527
	x5	106.162	5.871	.877	18.081	.000

2) Only x3 and x5 are significant to predict y, given that x1, x2 and x4 have a very low correlation coefficient with respect to y.
3) X5 is the more important variable to predict y, as it has the higher correlation coefficient of .877 meaning there is a strong positive correlation between x5 and y.

<u>Part 3</u>

1)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.877[a]	.769	.767	62.0710390	.769	326.927	1	98	.000
2	.993[b]	.986	.985	15.5965027	.216	1455.207	1	97	.000
3	.998[c]	.995	.995	9.1629102	.009	185.034	1	96	.000
4	.998[d]	.995	.995	9.0635384	.000	3.117	1	95	.081
5	.998[e]	.995	.995	9.1015268	.000	.209	1	94	.649

a. Predictors: (Constant), x5

b. Predictors: (Constant), x5, x3

c. Predictors: (Constant), x5, x3, x2

d. Predictors: (Constant), x5, x3, x2, x1

e. Predictors: (Constant), x5, x3, x2, x1, x4

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.940	6.211		.634	.527
	x5	106.162	5.871	.877	18.081	.000
2	(Constant)	.738	1.563		.472	.638
	x5	100.060	1.484	.827	67.428	.000
	x3	57.060	1.496	.468	38.147	.000
3	(Constant)	-1.225	.929		-1.318	.191
	x5	99.610	.872	.823	114.173	.000
	x3	57.748	.880	.473	65.606	.000
	x2	12.468	.917	.098	13.603	.000
4	(Constant)	-.770	.955		-.806	.422
	x5	99.537	.864	.822	115.208	.000
	x3	57.763	.871	.473	66.339	.000
	x2	12.540	.908	.098	13.817	.000
	x1	-1.666	.944	-.013	-1.765	.081
5	(Constant)	-.789	.960		-.822	.413
	x5	99.521	.868	.822	114.621	.000
	x3	57.791	.877	.474	65.926	.000
	x2	12.540	.911	.098	13.759	.000
	x1	-1.621	.953	-.012	-1.701	.092
	x4	.356	.780	.003	.457	.649

a. Dependent Variable: y

ANOVA[a]

Double-click to activate

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1259587.959	[illegible]	[illegible]959	326.927	.000[b]
	Residual	377575.761	98	3852.814		
	Total	1637163.720	99			
2	Regression	1613568.383	2	806784.192	3316.675	.000[c]
	Residual	23595.337	97	243.251		
	Total	1637163.720	99			
3	Regression	1629103.664	3	543034.555	6467.860	.000[d]
	Residual	8060.057	96	83.959		
	Total	1637163.720	99			
4	Regression	1629359.686	4	407339.922	4958.627	.000[e]
	Residual	7804.034	95	82.148		
	Total	1637163.720	99			
5	Regression	1629376.968	5	325875.394	3933.898	.000[f]
	Residual	7786.752	94	82.838		
	Total	1637163.720	99			

a. Dependent Variable: y

b. Predictors: (Constant), x5

c. Predictors: (Constant), x5, x3

d. Predictors: (Constant), x5, x3, x2

e. Predictors: (Constant), x5, x3, x2, x1

f. Predictors: (Constant), x5, x3, x2, x1, x4

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	x3	.468[b]	38.147	.000	.968	.988
	x2	.071[b]	1.462	.147	.147	.999
	x1	-.005[b]	-.100	.921	-.010	.998
	x4	-.029[b]	-.604	.547	-.061	.999
2	x2	.098[c]	13.603	.000	.811	.996
	x1	-.008[c]	-.666	.507	-.068	.998
	x4	.004[c]	.333	.740	.034	.994
3	x1	-.013[d]	-1.765	.081	-.178	.996
	x4	.005[d]	.631	.529	.065	.993
4	x4	.003[e]	.457	.649	.047	.983

a. Dependent Variable: y

b. Predictors in the Model: (Constant), x5

c. Predictors in the Model: (Constant), x5, x3

d. Predictors in the Model: (Constant), x5, x3, x2

e. Predictors in the Model: (Constant), x5, x3, x2, x1

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-357.240570	297.011597	7.991020	128.2901135	100
Std. Predicted Value	-2.847	2.253	.000	1.000	100
Standard Error of Predicted Value	1.030	3.536	2.164	.538	100
Adjusted Predicted Value	-356.919525	297.568115	7.932183	128.2155439	100
Residual	-26.7521057	25.3348370	.0000000	8.8687126	100
Std. Residual	-2.939	2.784	.000	.974	100
Stud. Residual	-2.997	2.946	.003	1.005	100
Deleted Residual	-27.8074818	28.3713169	.0588366	9.4467555	100
Stud. Deleted Residual	-3.134	3.075	.003	1.019	100
Mahal. Distance	.278	13.953	4.950	2.927	100
Cook's Distance	.000	.173	.011	.022	100
Centered Leverage Value	.003	.141	.050	.030	100

a. Dependent Variable: y

Regression Equations:

$y = 106.16x_5 + 3.94$

$y = 100.06x_5 + 57.06x_3 + 0.74$

$y = 99.61x_5 + 57.75x_3 + 12.47x_2 - 1.225$

$y = 99.54x_5 + 57.76x_3 + 12.54x_2 - 1.67x_1 - 0.77$

$y = 99.52x_5 + 57.79x_3 + 12.54x_2 - 1.6x_1 + 0.36x_4 - 0.79$

2) The best model is the last model, where x5, x4, x3, x2 and x1 are used to predict y with the equation $y = 99.52x_5 + 57.79x_3 + 12.54x_2 - 1.6x_1 + 0.36x_4 - 0.79$, because this model has the highest statistical significance and the highest Adjusted R square.
3) **Prediction**:
   Let x1 = .0142
   x2 = 2.2825
   x3 = -.8427
   x4 = -.4225
   x5 = -.5634
   y = 40.03

   Using our model,

   y = $99.52(-.5634) + 57.79(-.8427) + 12.54(2.2825) - 1.6(.0142) + 0.36(-.4225) - 0.79$

   y = 35.025

   CI = (30.35 : 39.7)
   PI = (16.36 : 40.00)