Name: Connor Raymond Stewart
Student Number: 101041125

STAT 2509 Assignment Three:

① Solve the following Questions Reguarding the relation between Profits & Capital investment/Ads:

a) State the MLR model & all assumptions:

the model has three variables: (one y-intercept w/ 2 Slopes)

$$y = B_0 + B_1X_1 + B_2X_2 + E, \text{ Where } E \text{ is the error}$$

The Assumptions are:

I) All x's are observed without error
II) y's are independently distributed with mean: $E(y) = B_0 + B_1X_1 + B_2X_2$
III) Varience of y's Constant
IV) y's or errors are $N \sim (E(y), \sigma^2)$

b) Find the estimates of the population parameters B_0, B_1, B_2 & get the least-Squares line:

$$(X^TX)^{-1}(X^Ty) = \begin{bmatrix} 1.62\ldots & -0.0445\ldots & -0.269\ldots \\ -0.0445\ldots & 0.00169\ldots & 0.00584\ldots \\ -0.269\ldots & 0.00584\ldots & 0.0587\ldots \end{bmatrix}\begin{bmatrix} 98 \\ 1433 \\ 383 \end{bmatrix}$$

[Note: $(X^TX)^{-1}$ is Shortened for brevity, Please See Q1 for $(X^TX)^{-1}$ in full]

$$[\hat{B}] = \begin{bmatrix} 1.62(98) - 0.0445(1433) - 0.269(383) \\ -0.0445(98) + 0.00169(1433) + 0.00584(383) \\ -0.269(98) + 0.00584(1433) + 0.0587(383) \end{bmatrix} = \begin{bmatrix} -8.1770184764 \\ 0.2921319818 \\ 4.4343028367 \end{bmatrix} = \begin{bmatrix} B_0 \\ B_1 \\ B_2 \end{bmatrix}$$

thus:

$$y = -8.1770184764 + 0.2921319818X_1 + 4.4343028367X_2$$

$$\doteq -8.1770 + 0.2921X_1 + 4.4343X_2$$

c) Find the predicted value of y for the new X vector, $X = \begin{vmatrix} 1 \\ 2 \\ 6 \end{vmatrix}$:

$$y \doteq -8.1770X_0 + 0.2921X_1 + 4.4343X_2$$
$$\doteq -8.1770(1) + 0.2921(2) + 4.4343(6)$$
$$\doteq -8.1770 + 0.5842 + 26.6058$$
$$\doteq 19.013$$

d) Setup the ANOVA table and test for the Significance of the model for $\alpha = 0.05$:

Profits $= B_0 + B_1(\text{investment}) + B_2(\text{adds})$

$$TSS = y^Ty - \frac{(\Sigma y_i)^2}{n} = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 1392 - \frac{(98)^2}{10} = \boxed{431.6}$$

$$SSR = \hat{B}^T(X^Ty) - \frac{(\Sigma y_i)^2}{n} \doteq [-8.1770 \quad 0.2921 \quad 4.4343]\begin{bmatrix} 98 \\ 1433 \\ 383 \end{bmatrix} - \frac{(98)^2}{10}$$

$SSR = -8.1770184764(78) + 0.2921319818(1433) + 4.4343028367(383) - 760.4$

$= 355.2153057$

$SSE = TSS - SSR$

$= 431.6 - 355.2153057$

$\doteq 76.3846943$

$MSE = SSE/(n-K-1)$

$= 76.3846943/(10-2-1)$

$\doteq 10.91209919$

$MSR = SSR/(n-K-1)$

$= 355.2153057/(10-7-1)$

$\doteq 177.6076529$

$K = 2$ as there's two conditions

$df_R = K = 2$

$df_E = n-K-1 = 10-1-2 = 7$

$df_T = n-1 = 9$

$r^2 = \frac{SSR}{TSS} = \frac{355.2153057}{431.6} \doteq 0.823019707$

$$F = \frac{r^2/K}{(1-r^2)/(n-K-1)} = \frac{0.823019707/2}{(1-0.823019707)/(10-2-1)}$$

$$= \frac{0.411509853}{(0.176980292)/(7)} = \frac{0.411509853}{0.025282898} = 16.27621318$$

$\doteq 16.28$, $F_{16.28;(2,7)}$ is less then 0.005, roughly at 0.002

ANOVA Table:

Source	df	Sum of Squares	Mean Square	F Value	Pr > F
Regression	2	355.2153057	177.6076529	16.28	0.002
Error	7	76.3846943	10.91209919		
Corrected Total	9	431.6			

Test Significance:

I) Claim: The model is Statistically Significant at predicting Profits based on Capital investment & advertising expendature, Given $\alpha = 0.05$.

$H_0: P(F) > 0.05$ → Null: the model is Not significant to 95% Confidence

$H_a: P(F) \leq 0.05$ → Alternative: the model's Significant to 95% Confidence

II) F-value:

from ANOVA-table, we see $F = 16.28$

III) Rejection Region:

$P(F) \leq 0.05$

$P(F = 16.28) \doteq 0.002$ (See ANOVA table)

thus: $0.002 \leq 0.05$ therefore we reject the null hypothesis

IV) Conclusion:

Since $P(F) \doteq 0.002 \leq 0.05 = \alpha$ we reject the null hypothesis we thus take the alternative hypothesis which states that the model is acurate within a 95% level of Confidence.

e) Find the Std. error for $\hat{B}_j$'s, where $j=0,1,2$:

B_j's Std. error $\rightarrow Var(\hat{B}_j)=\sqrt{V_{jj}MSE}$

$$S.E.\ B_0=\sqrt{V_{00}MSE}=\sqrt{1.6211(10.7121)}\doteq 4.2059$$

$$B_1=\sqrt{V_{11}MSE}=\sqrt{0.001688(10.7121)}\doteq 0.1357$$

$$B_2=\sqrt{V_{22}MSE}=\sqrt{0.05869(10.7121)}\doteq 0.8003$$

Where MSE is from SAS output on Sheet 1

f) Test wheather X_2 term Contributes to the model. Use t-test with $\alpha=0.05$:

$y=B_0+B_1X_1+B_2X_2$

I) Claim: The term X_2 Contributes to the model

$H_0: B_2=0 \rightarrow$ Null: X_2 doesn't Contribute to the model

$H_a: B_2\neq 0 \rightarrow$ Alternative: X_2 Contributes to the model

II) T-test:

$$t=\frac{\hat{B}_j}{\sqrt{V_{jj}MSE}}=\frac{\hat{B}_2}{\sqrt{(0.0586860556)(10.7121)}}=\frac{4.4343028367}{0.80024263}$$

$\doteq 5.54$

III) Rejection Region:

$$t>|t_{\alpha/2;n-k-1}|\rightarrow t>|t_{0.025;7}|\rightarrow t>|\pm 2.365|$$

IV) Conclusion:

Since $t\doteq 5.54>|\pm 2.365|$ we reject the null hypothesis. We thus take the alternative hypothesis which states that X_2 (advertising expenditure) Contributes to the model, with a 95% level of Confidence.

g) Find the values of the coefficient of determination, r^2, & adjust r^2 & interpret their meanings:

$$r^2 = \frac{SSR}{TSS} = \frac{355.2153057}{431.6} \doteq 0.82302 \doteq 82.30\%$$

$$MSE = SSE/(n-k-1)$$

$$r^2_{adj} = \frac{MSE}{TSS/n-1} = 1 - \frac{76.384694/(10-2-1)}{431.6/(10-1)} = 1 - \frac{10.91209857}{(2158/45)} \doteq 0.77245$$

$$\doteq 0.7725 \doteq 77.25\%$$

Since $r^2 \doteq 82.30\%$ & $r^2_{adj} = 77.25\%$, we see that r^2 & r^2_{adj} have a roughly 5.05% difference which means the model can use some improvment.

↳ Too improve the model we could add more variables

That said, the model — though very rough — is still good enough to pass for many experimental purposes. Calculations requiring high precision shouldn't use this model.

h) Use SAS to verify results, what's the conclusion about the goodness of the model:

We find & plot the residuals with the following model:

$$y_i - \hat{y}_i = y_i - (\hat{B}_0 + \hat{B}_1 x_1 + \hat{B}_2 x_2)$$

$$= y_i - (-8.1770184764 + 0.2321319818 x_1 + 4.4343028367)$$

See Graph 4 for more info.

the residual plot checks the assumption of independence (Graph 1)

We see in the residual plot all the values are independent since there's no clear trend.

this means the x-values are all independent from each other

Thus, the Assumption of independence (Assumption II) is valid.

Check graph 1 for the SAS output saying the same as above

Check anotations over graph 2.1/2.2 & three to see info on the Assumptions three & four.

↳ Assumption three NOT violated

↳ Assumption four violated BUT fixable

i) Use the SAS output to answer part δ using the Partial F-test with $\alpha = 0.05$:

I) State the hypothesis:

$H_0: B_2 = 0$

$H_a: B_2 \neq 0$

II) Find the F_{drop} using the full & reduced model:

$$F_{drop} = \frac{SSE_r - SSE_f / df_{SSE_r} - df_{SSE_f}}{MSE_F} \qquad f: y = B_0 + B_1X_1 + B_2X_2 + \mathcal{E}$$

$$r: y = B_0 + B_1X_1 + \mathcal{E}$$

$$= \frac{(411.4396 - 76.38486)/(8-7)}{10.91212} = \frac{335.05474}{10.91212}$$

$\doteq 30.70$

III) Rejection Region:

$$F_{drop} > F_{0.05(1,7)} = \boxed{5.59}$$

thus $F_{drop} > F_{0.05(1,7)}$ meaning we reject the null hypothesis

$\therefore$ We conclude that advertising expenditure (X_2) contributes to the given model, with a 95% level of confidence.

Results from b, d, g are verified in table one

See table one for proof

② Run SAS to test if the 3 lines are Parallel, using $\alpha = 0.05$.

$$y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \underbrace{B_4X_1X_2}_{B_4} + \underbrace{B_5X_1X_3}_{B_5} + \mathcal{E}$$

checks to see if we need these

We have 3 drugs:

$Drug A = B_0 + B_1X_1 + \mathcal{E} \longrightarrow$ No $X_2=0, X_3=0$

$Drug B = B_0 + B_1X_1 + B_2 + B_4X_1(1) \longrightarrow X_3=0, X_2=1$

$Drug C = B_0 + B_1X_1 + B_3 + B_5X_1 \longrightarrow X_2=0, X_3=1$

$\Downarrow$

$Drug B = (B_0 + B_2) + X_1(B_1 + B_4)$

$Drug C = B_0 + B_3 + X_1(B_1 + B_5)$

Claim: We Claim that the 3 lines are parallel

I) State the hypothesis:

$H_0: B_4 = B_5 = 0$

$H_a:$ at least one of B's $\neq 0$

Here, we are testing if the lines are Parallel, if they are then H_0 is true, otherwise H_a is true

II) Find $F_{drop} = [(SSE_r - SSE_f)/(df_{SSE_r} - df_{SSE_f})]/MSE_f$:

$$F_{drop} = \frac{(SSE_r - SSE_f)/(df_{SSE_r} - df_{SSE_f})}{MSE_f}$$

$$= \frac{(7.13833 - 0.689)/(8-6)}{0.11483} = \frac{(6.44933)/2}{0.11483}$$

$$= \frac{3.224665}{0.11483} \doteq 28.0821 \doteq 28$$

III) Find the Rejection Region:

$$F_{0.05(2,8)} = 4.46 < F_{drop} = 28$$

IV) Conclusion:

Since $f_{drop} > 5$ we reject H_0, thus we Conclude to 95% Confidence that the lines of the three drugs are not Parallel.

③ Given MSE is an unbiased estimator of σ^2, under what conditions will MSR be one: under what conditions can $E(MSR) = \sigma^2$:

By the given formula:

$$E(MSR) = \sigma^2 + \frac{1}{2}\left[B_1^2 \sum (X_{i1} - \bar{X})^2 + B_2^2 \sum (X_{i2} - \bar{X})^2 + 2B_1B_2 \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)\right]$$

We can find the solution if we let $B_1 = B_2 = 0$:

$$E(MSR) = \sigma^2 + \cancel{\frac{1}{2}\left[(0)^2 \sum (X_{i1} - \bar{X})^2 + (0)^2 \sum (X_{i2} - \bar{X})^2 + 2(0)(0) \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)\right]}$$

$$E(MSR) = \sigma^2 + 0 = \sigma^2$$

thus, $E(MSR)$ is an unbiased estimator of σ^2 when both B_1 & B_2 are zero.

$\boxed{\therefore B_1 = B_2 = 0}$

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: profits

Number of Observations Read	10
Number of Observations Used	10

indicates the null hypothesis is rejected for $\alpha=0.05$ with the ANOVA model.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	355.21514	177.60757	16.28	0.0023
Error	7	76.38486	10.91212		
Corrected Total	9	431.60000			

SSR MSR MSE SSE TSS

Root MSE	3.30335	R-Square	0.8230
Dependent Mean	9.80000	Adj R-Sq	0.7725
Coeff Var	33.70766		

r^2 r^2_{adj}

$\hat{B}_0$ $SE\,B_0$ $SE\,B_1$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-8.17702	4.20599	-1.94	0.0930
investment	1	0.29213	0.13571	2.15	0.0684
ads	1	4.43430	0.80024	5.54	0.0009

$\hat{B}_2$ $SE\,B_2$ $\hat{B}_1$

Connor, 101041125

i) verify b,d,g with SAS

verify b:
See $\hat{B}_0, \hat{B}_1, \hat{B}_2$

verify d:
See MSE, MSR, SSE, SSR

verify g:
See r^2 & r^2_{adj}

full model

The SAS System

The REG Procedure
Model: MODEL2
Dependent Variable: profits

Number of Observations Read	10
Number of Observations Used	10

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	20.16040	20.16040	0.39	0.5487
Error	8	411.43960	51.42995		
Corrected Total	9	431.60000			

SSR$_r$ MSR$_r$ MSE$_r$ SSE$_r$

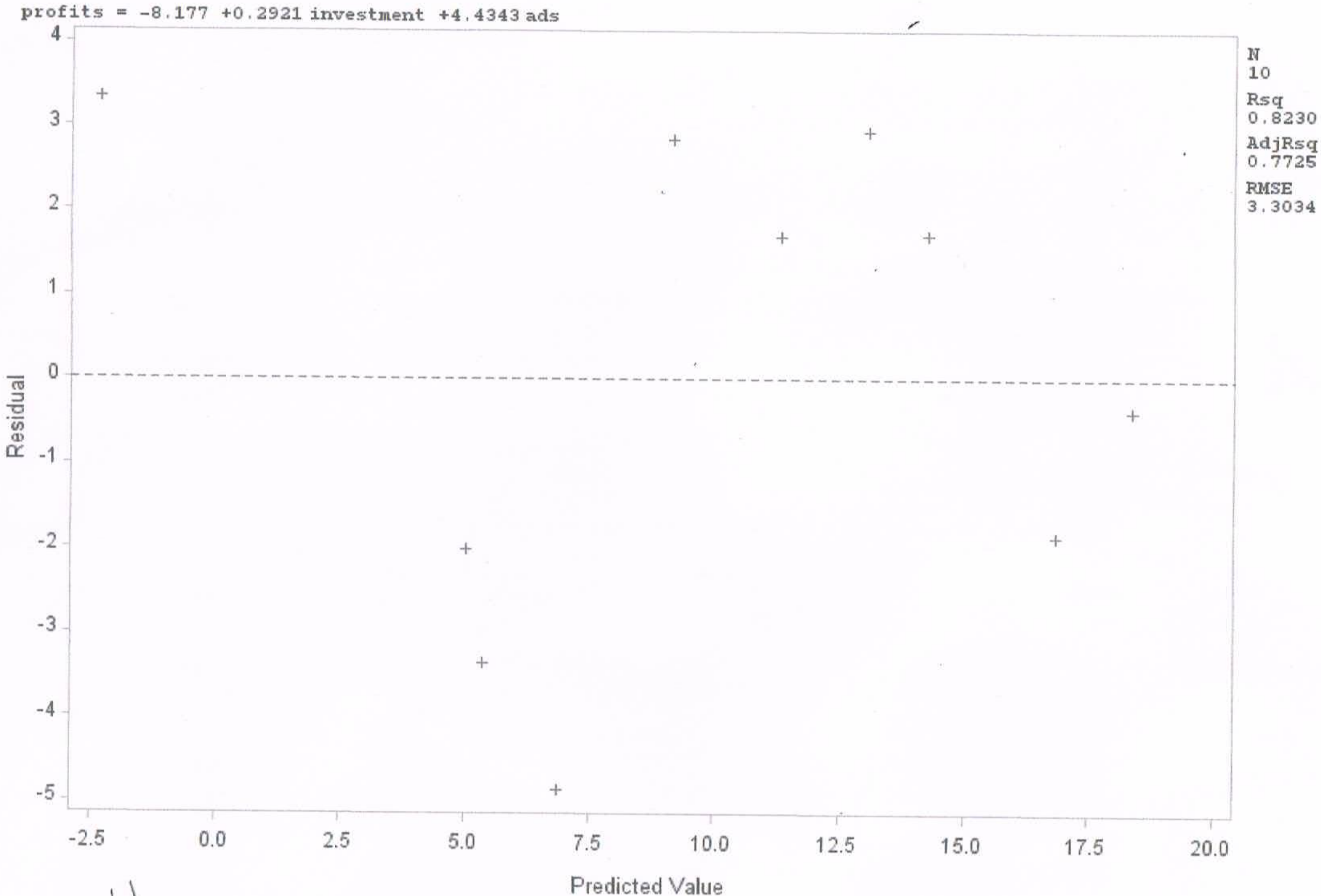
Root MSE	7.17147	R-Square	0.0467
Dependent Mean	9.80000	Adj R-Sq	-0.0725
Coeff Var	73.17824		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.18938	4.43928	2.75	0.0252
investment	1	-0.14934	0.23852	-0.63	0.5487

Connor, 101041125

Reduced model

Graph 1:



h)

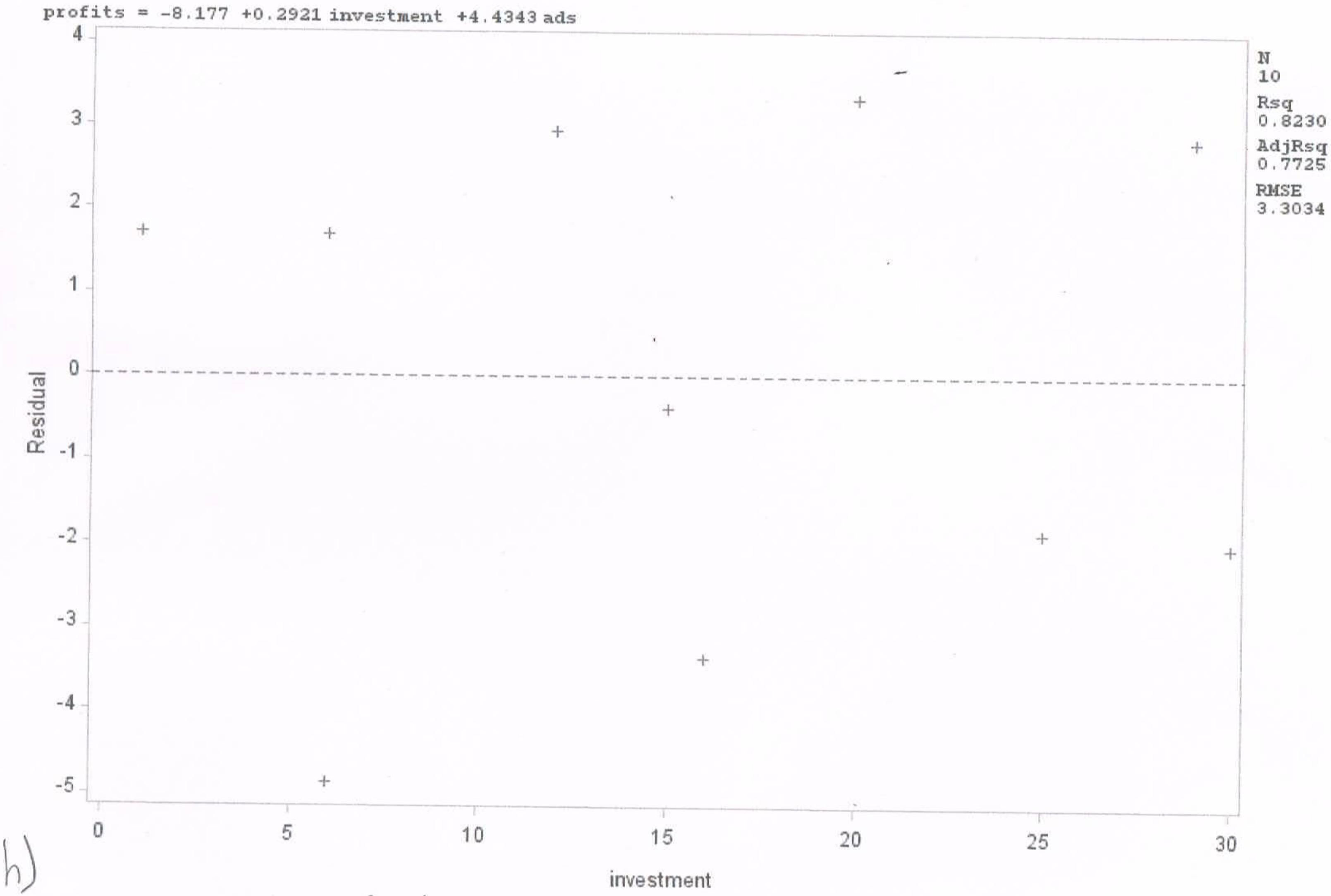
Connor, 101041125

This graph tests for the assumption of independence

No pattern among y → ~~MM~~ Assumption two NOT violated

Graph 2.i

The REG Procedure



Connor, 101041125

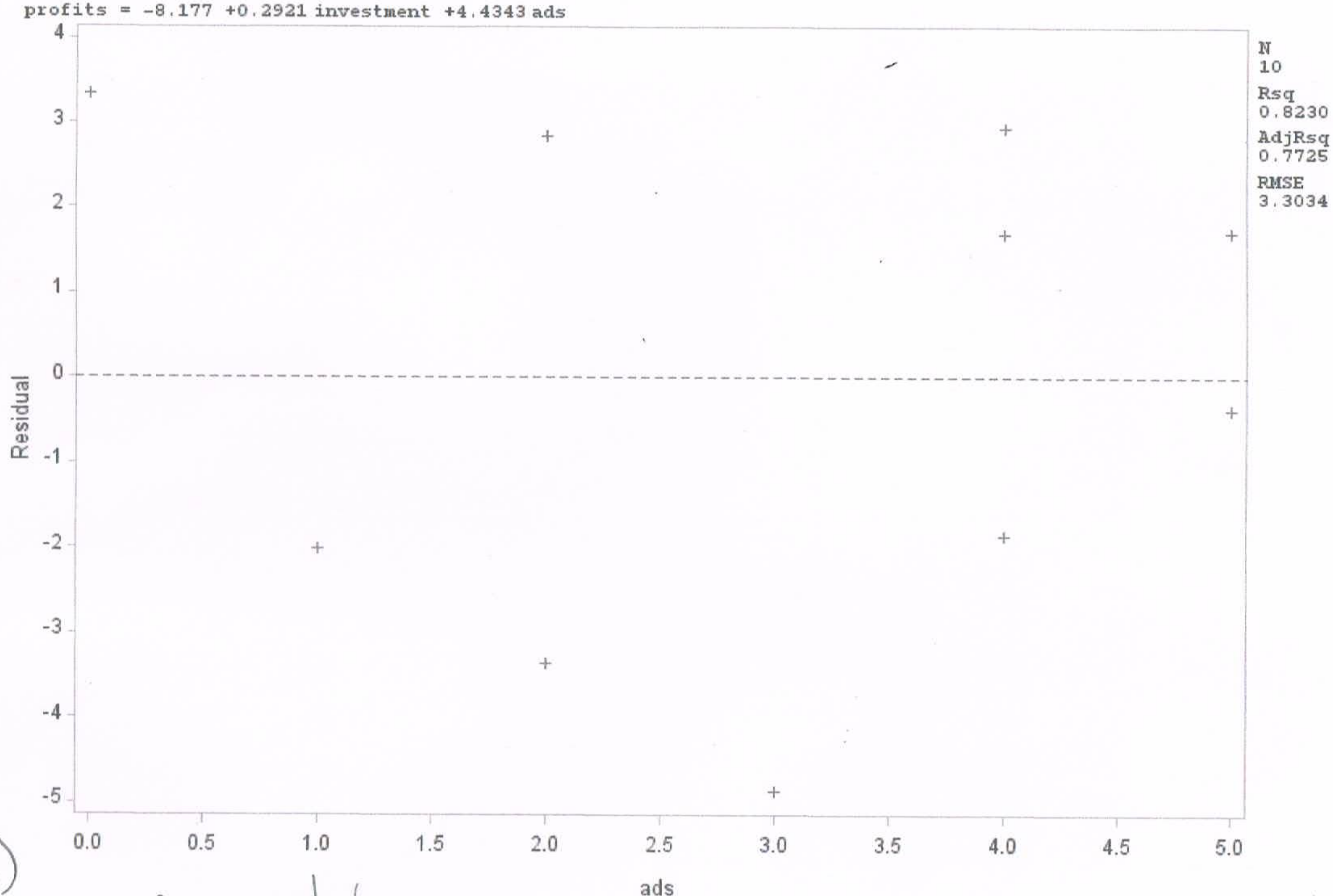
h)

Anylizes the varience of y's by Showing our residuals vs. the x-value for investment

nxt. Page

Graph 2.2:

The REG Procedure



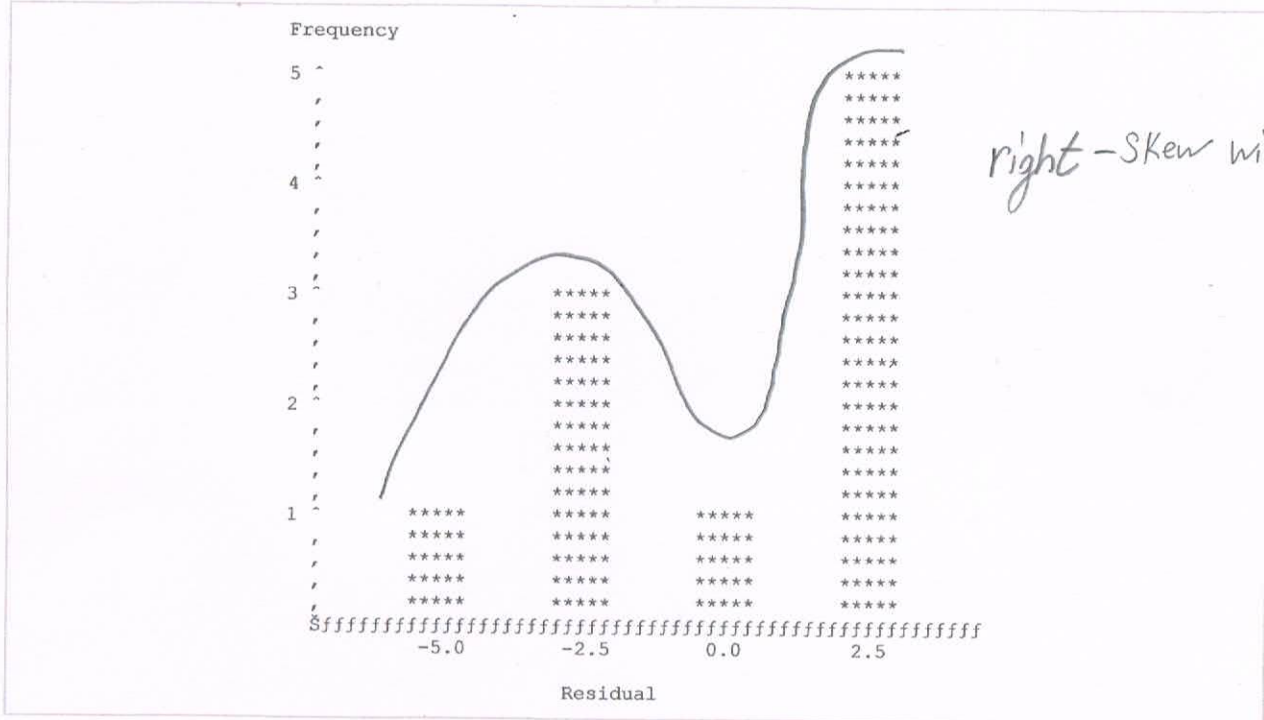
Connor, 101041125

h) Same as last page, but plots residuals vs. the x-value for ads

The y-values in Graphs 2.1 & 2.2 have no pattern → Assumption three's not violated

Graph 3:

The SAS System



right-Skew with depression

Connor, 101041125

h)

The residual Hist. tests for normality

↳ Not normal ∴ violates normality

↳ However, changing the data point scale (i.e. to logrithmic or exponential) can fix this

So, Although the histogram violates assumption four, the experiment can be easily altered so it does not.

Graph 4:

Profits(Y)	Capital Investment(X1)	Advertising Expenditure(X2)	Predicted Values	Residuals
15	25	4	16.86349242	-1.863492415
16	1	5	14.28662769	1.713372311
2	6	3	6.878681925	-4.878681925
3	30	1	5.021243814	-2.021243814
12	29	2	9.163414669	2.836585331
1	20	0	-2.33437884	3.33437884
16	12	4	13.06577665	2.934223348
18	15	5	18.37647543	-0.376475434
13	6	4	11.31298476	1.687015239
2	16	2	5.365698906	-3.365698906

Sum of Residuals

-1.7E-05

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: profits

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	15	16.8635	2.0907	11.9197	21.8072	7.6193	26.1077	-1.8635
2	16	14.2866	1.9929	9.5742	18.9990	5.1641	23.4092	1.7134
3	2	6.8787	1.7126	2.8290	10.9284	-1.9199	15.6772	-4.8787
4	3	5.0212	1.9216	0.4774	9.5650	-4.0154	14.0579	-2.0212
5	12	9.1634	1.7851	4.9423	13.3845	0.2846	18.0422	2.8366
6	1	-2.3344	2.3705	-7.9397	3.2710	-11.9487	7.2799	3.3344
7	16	13.0658	1.2314	10.1541	15.9775	4.7296	21.4020	2.9342
8	18	18.3765	1.8483	14.0060	22.7470	9.4257	27.3272	-0.3765
9	13	11.3130	1.5160	7.7282	14.8978	2.7185	19.9075	1.6870
10	2	5.3657	1.3159	2.2541	8.4773	-3.0424	13.7738	-3.3657

Sum of Residuals	0
Sum of Squared Residuals	76.38486
Predicted Residual SS (PRESS)	166.56530

Connor, 101041125

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: potency

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	55.29350 SSR	11.05870 MSR	96.30	<.0001
Error	6	0.68900 SSE	0.11483 MSE		
Corrected Total	11	55.98250 TSS			

indicates null hypothesis is rejected for $\alpha = 0.05$ with the ANOVA model.

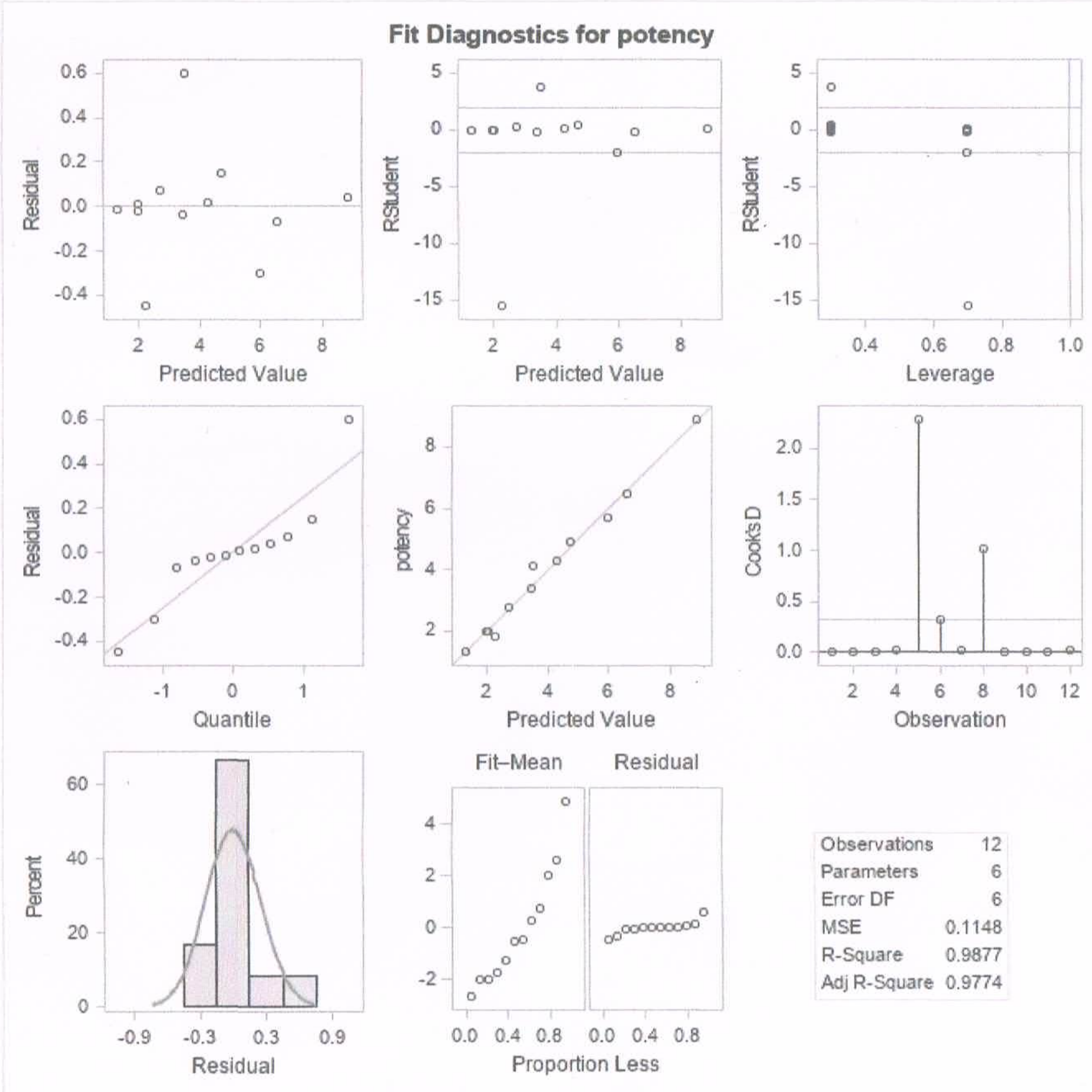
Root MSE	0.33887	R-Square	0.9877 r^2
Dependent Mean	3.97500	Adj R-Sq	0.9774 r^2_{adj}
Coeff Var	8.52505		

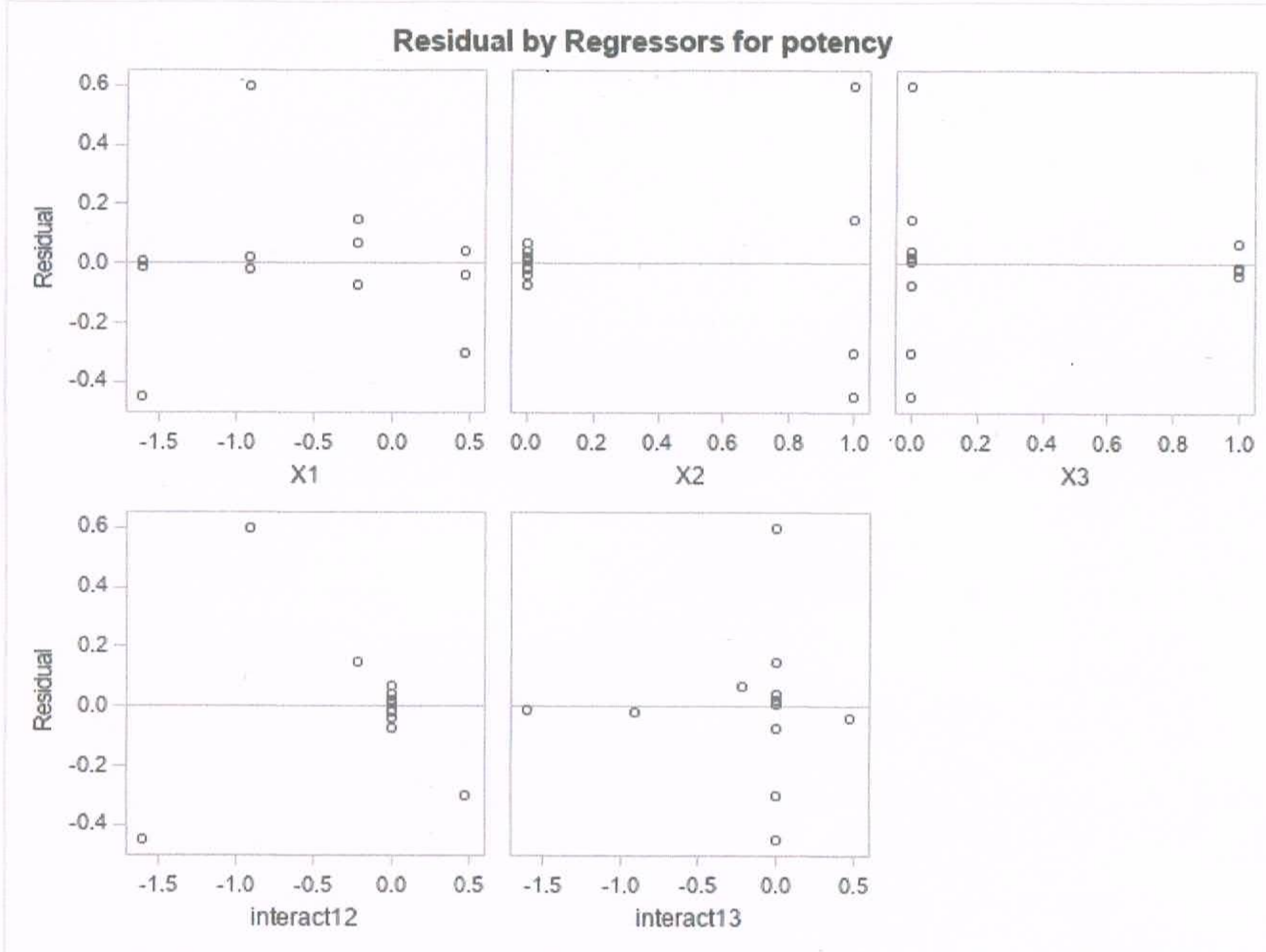
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.30722	0.21029	34.75	<.0001
X1	1	3.30377	0.21864	15.11	<.0001
X2	1	-2.15481	0.29740	-7.25	0.0004
X3	1	-4.34865	0.29740	-14.62	<.0001
interact12	1	-1.50040	0.30920	-4.85	0.0028
interact13	1	-2.27946	0.30920	-7.37	0.0003

Connor, 101041125 ods graphics off

The REG Procedure
Model: MODEL1
Dependent Variable: potency





Connor, 101041125 ods graphics off

The SAS System

The REG Procedure
Model: MODEL2
Dependent Variable: potency

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	48.84417 SSR_r	16.28139 MSR_r	18.25	0.0006
Error	8	7.13833 SSE_r	0.89229 MSE_r		
Corrected Total	11	55.98250 TSS_r			

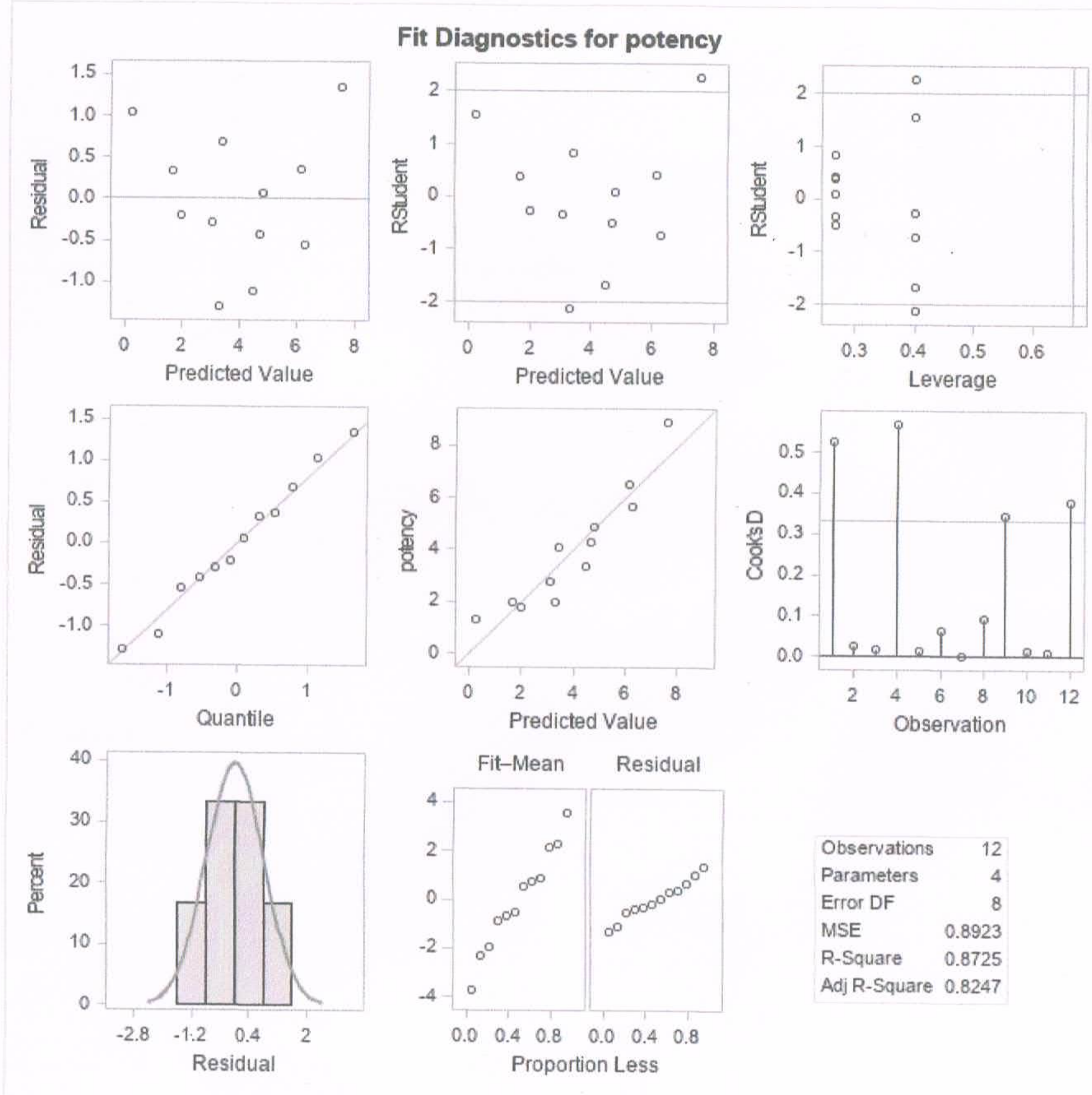
Root MSE	0.94461	**R-Square**	0.8725 r_r^2
Dependent Mean	3.97500	**Adj R-Sq**	0.8247 $r_{r\,adj}^2$
Coeff Var	23.76382		

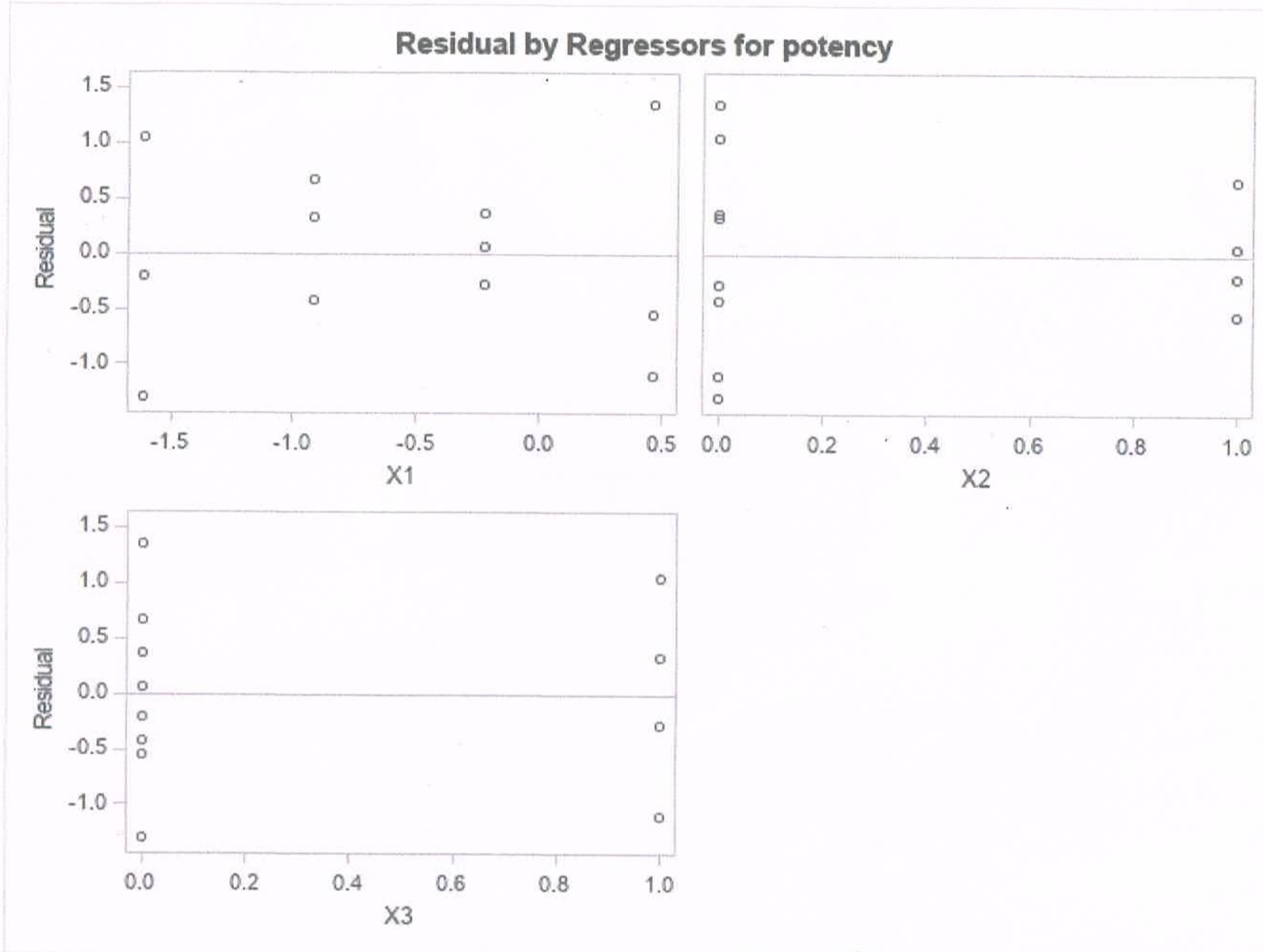
Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	6.58940	0.51309	12.84	<.0001
X1	1	2.04382	0.35187	5.81	0.0004
X2	1	-1.30000	0.66794	-1.95	0.0875
X3	1	-3.05000	0.66794	-4.57	0.0018

Connor, 101041125 ods graphics off

The SAS System

The REG Procedure
Model: MODEL2
Dependent Variable: potency





Connor, 101041125 ods graphics off

```
Footnote 'Connor, 101041125';
ods graphics off;
Data corporate;
input profits investment ads @@;
Cards;
15 25 4
16 1 5
2 6 3
3 30 1
12 29 2
1 20 0
16 12 4
18 15 5
13 6 4
2 16 2
run;
proc reg;
model profits=investment ads;
model profits=investment;
run;
run;
Proc Reg;
model profits=investment ads/CLM CLI;
run;
Proc Reg;
Model profits=investment ads;
Plot R.*P.;
Plot R.*ads;
Plot R.*investment;
Output out=res R=resids;
run;
Proc Chart data=res;
vbar resids;
run;
```

Footnote:
Name: Connor Raymond Stewart
Student Number: 101041125
File: SAS Code for Question 1

Footnote:
Name: Connor Raymond Stewart
Student Number: 101041125
File: SAS Code for Question 2

```
Footnote 'Connor, 101041125'
ods graphics off;
Data drug;
input dose X2 X3 potency;
X1=log(dose);
interact12=X1*X2;
interact13=X1*X3;
Cards;
0.2 0 0 2.0
0.4 0 0 4.3
0.8 0 0 6.5
1.6 0 0 8.9
0.2 1 0 1.8
0.4 1 0 4.1
0.8 1 0 4.9
1.6 1 0 5.7
0.2 0 1 1.3
0.4 0 1 2.0
0.8 0 1 2.8
1.6 0 1 3.4
run;
proc reg;
model potency=X1 X2 X3 interact12 interact13;
model potency=X1 X2 X3;
run;
```