EXECUTIVE SUMMARY

Problem: Data for profit measures on 25 consecutive days has been collected. We are trying to predict the profit for a day, given the amount of material A (in gallons), material B (in lbs) and the day it was collected (1-25). In addition, we have been asked to give a predicted profit on day 26 when material A = 35 gallons and material B = 95lbs, and a 90% confidence interval for this prediction.

Variable	Mean	Std. Dev	Min	Max
Profit($100)	221.2	53.29	127.6	322.4
Material A (Gallons)	55.3	10.1	41.6	74.8
Material B (Pounds)	128.5	5.667	120	138.9
Day	13	7.35	1	25

Recommended Model: P=167.4+4.104*(D)
Profit P in units of 100s of dollars. Day (D) being the day that the datum was collected.
Sp.d=45.7, R^2=.322.

Using this reduced first order model, for day 26 where material A=35 gallons and material B = 95lbs we would predict a profit of 274 with a 90% confidence interval for this prediction given by the interval (172, 376).

ANALYSIS

Problem: Data for profit measures on 25 consecutive days has been collected. We are trying to predict the profit for a day, given the amount of material A (in gallons), material B (in lbs) and the day it was collected (1-25). In addition, we have been asked to give a predicted profit on day 26 when material A = 35 gallons and material B = 95lbs, and a 90% confidence interval for this prediction.

Data: The data has been entered into the computer and printed out (p.1a). The data has been checked for accuracy and has been verified to be the same as the data provided to us.

Profit: (in $), the dependent variable, has an average of $221.2, standard deviation of 53.29, ranges from a minimum value of $127.6 to a maximum of $322.4. The data appears to be randomly distributed, there appears to be no correlation between the distance from the mean and the frequency (p.1c).

Material A: an independent variable, has an average of 55.34, standard deviation of 10.15, ranges from a minimum value of 41.6 to a maximum of 74.8. The shape of the distribution appears unimodal, skewed to the right (p.1c).

Material B: an independent variable, has an average of 128.55, a standard deviation of 5.66, ranges from a minimum value of 120 to a maximum value of 138.9. The shape of the distribution appears unimodal (p.1c).

Order: an independent variable, has an average of 13, and a standard deviation of 7.35. Ranges from a minimum value of 1 to a maximum value of 25. The shape of the distribution is uniform (p.1c).

Profit (y) vs independent variables. Examining the correlation matrix (p.1d) we see the following significant results: the corr between:

Profit vs. Material B: R=.36, about 13% of the variability in the y scores about y-bar is explained by the simple regression between profit and material B. The std. dev. of the y scores about a simple linear regression using material B is approximately .93 times the std dev. of the y scores about y bar. As material B increases, y-hat increases in this fitted simple regression model.

Profit vs. Day: R=.57, about 32.5% of the variability in the y scores about y-bar is explained by the simple regression between Profit and Day. The std. dev. of the y scores about a simple linear regression using Day is approximately .82 times the std dev. of the y scores about y bar. As Day increases, y-hat increases in this fitted simple regression model.

Correlations between pairs of independent variables: examining the correlation matrix (p.1d), we see the following significant results:

Day vs. Material B: R=.71, about 50.4% of the variability in the Day scores about Day-bar is explained by the simple regression between Material B and Day. The std. dev. of the Day scores about a simple linear regression using Material B is approximately .70 times the std dev. of the Day scores about Day bar. As Material B increases, Day-hat increases in this fitted simple regression model.

Material A vs. Material B: R=.313, about 9.7% of the variability of the Material A scores about Material A- bar is explained by the simple regression between Material A and Material B. The std dev of the Material A scores about a simple linear regression using Material B is approximately .95 times the std dev of the Material A scores about material A-bar. As material B increases, material A-hat increases in this fitted simple regression model.

Scatterplots of Profit vs Independent variables: Describe the relationship:
Profit vs Material B: There appears to be a positive linear relationship between these two variables.
Profit vs Day: There appears to be a positive linear relationship between these two variables.

Scatterplots between pairs of independent variables:
Day vs Material B: There appears to be a strong positive linear relationship between these two variables.
Day vs. Material A: There doesn't appear to be a relationship between these two variables
Material vs. Material B: There appears to be a positive linear relationship between these two variables.

Fit of the first order model: E(P) = Bo + B1*D + B2*A +B3*B (p.2a) Examining the EXCEL results for this model, we find Sp.dab = 47.7 vs Sp = 53.29. R^2=.32. For testing Ho B1=B2=B3=0 since B holds, we find p for Global F= .042<.05=a. Thus the first order model is a significant improvement over E(P) = Bo.

Fit of the second order model:
E(P)=Bo + B1(D) +B2(A) +B3(B) +B4(A^2)+B5(B^2) +B6(D^2) +B7(DA) +B8(DB) +B9(AB)
We see that the standard deviation of the residuals about the second order model is 52.07, which we note is higher than the standard deviation of the y scores about the first order model (current best). For testing Ho: B4=B5=B6=B7=B8=B9=0, since C holds, we find the partial F is obtained by [(45524-37955)/(20-14)]/2711=.465. We reject this value if it is too large. Note since the cutoff point with a=.05 on an F(6;14) is approximately 2.85. (Reject Ho if Fobs >2.85, accept Ho otherwise) we accept Ho, and thus p>.05. We conclude the second order model is not significantly better than the first order model.

Residual plot for the highest order model fitted: Examining the residual plot for the second order model (p.2c) we note: the mean of the residuals seems to be 0 regardless of the value of the predicted. If assumption 1 was violated we would consider fitting a higher order model than has been fit so far and compare it to the last acceptable model.

Return to the first order model: We now examine the first order model to see if we can drop any variables. We only look at the higher order terms, and individually test; (p2b) (A holds since we are testing for single B=0)
Ho: B1=0 vs. H1: not Ho (A holds) [p=.02<.05, Reject Ho]
Ho: B2=0 vs. H1: not Ho (A holds) [p=.76>.05, Accept Ho]
Ho: B3=0 vs. H1: not Ho (A holds) [p=.79>.05, Accept Ho]
If we find one or more P values above a = .05, we drop the one with the highest P value and refit the model without this term. In this case, we drop the Material B and refit.

Fit of reduced first order of form: E(P) = Bo+B1(D) + B2(A)
The standard deviation of the y scores about the model is 46.6. R^2 = .327. We only look at the higher order terms and individually test:
Ho: B1=0 vs. H1: not Ho (A holds) [p=.004<.05; Reject Ho];
Ho: B2=0 vs. H1: not Ho (A holds) [p=.68>.05; Accept Ho]
Since we find one term with a p value>.05, we drop the term with the highest p value and refit the model without this term. In this case we drop Material A and refit.

Fit of further reduced first order: E(P) =B0 + B1(D)
The standard deviation of the y scores about the model is 45.7. R^2 = .322. We only look at the higher order terms and individually test
Ho: B1=0 vs. H1: not Ho (A holds) [p=.003<.05; reject Ho];
In this case we have no more terms to drop.

Residual plot for the FINAL (reduced second order) model: Examining the residual plot for this model (p.3b-2), we note:

1. The mean of the residuals, regardless of the value of the predicted, seems to be 0

2. The variance of the residuals, regardless of the value of the predicted, seems to be constant

Histogram of residuals for the final (reduced first order) model: Examining the histogram of the residuals (p.3b-3), we see that they appear to be more or less symmetric and unimodal, and we can't reject that they are normally distributed.

Independence of the residuals for the final (reduced first order) model: Given that the data was ordered by the order that it was collected, we examine the plot of the residual t vs residual t-1 (p.3b-4). It seems that there is some relationship between t vs. t-1. If we regress the residual t on residual t-1, we get a correlation of r=-.42 with a p of .044 and we conclude that the coefficient is significantly different from 0 (p.3b-5).

The prediction model is given by P=167.39 + 4.1038(D)
The predicted profit for day 26 when material A = 35 gallons and material B = 95 lbs is 274, obtained by substituting the day 26 for the D variable in the above equation. A 90% confidence interval is for this prediction is given by:
Prediction +/- tdferror,a/2*Sp.d * Sqrt(1+(1/n) +[(x*-xbar)^2/((n-1)s^2x)])
274+/- 2.074*45.8*1.08 =
274 +/- 102= (172, 376)

Conclusion: We recommend using the reduced first order model given above to describe the relationships.

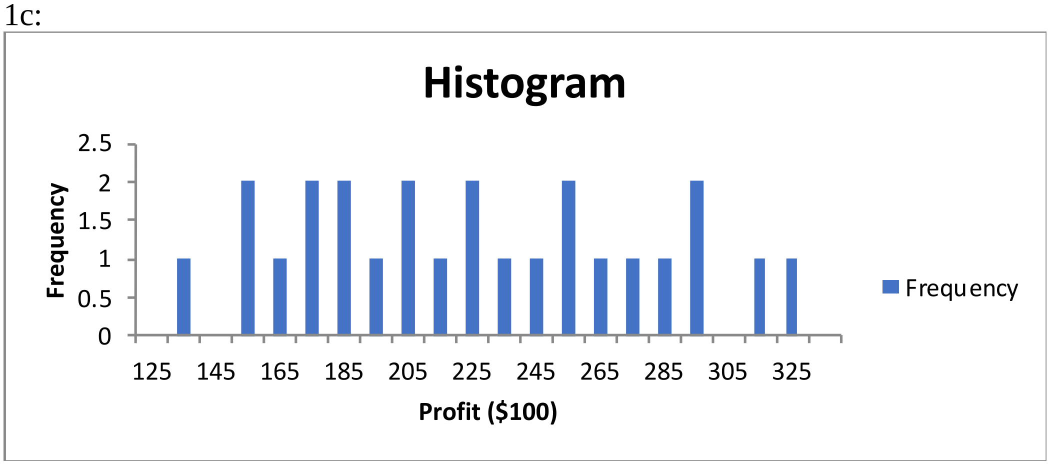
1a:

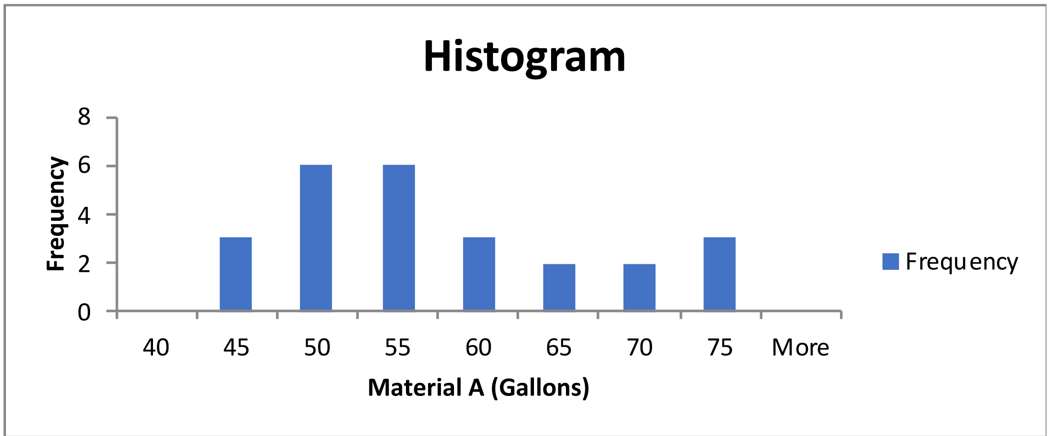
Profit	Day	Material A	Material B	A^2	B^2	D^2	DA	DB	AB
160	1	47.9	128.1	2294.41	16409.61	1	47.9	128.1	6135.99
154.1	2	47.9	125.4	2294.41	15725.16	4	95.8	250.8	6006.66
193.1	3	41.6	128.1	1730.56	16409.61	9	124.8	384.3	5328.96
202.8	4	69.8	125.4	4872.04	15725.16	16	279.2	501.6	8752.92
253.8	5	56	122.7	3136	15055.29	25	280	613.5	6871.2
182.4	6	74.1	122.7	5490.81	15055.29	36	444.6	736.2	9092.07
170.4	7	47.9	122.7	2294.41	15055.29	49	335.3	858.9	5877.33
221.2	8	47.9	122.7	2294.41	15055.29	64	383.2	981.6	5877.33
127.6	9	60.4	128.1	3648.16	16409.61	81	543.6	1152.9	7737.24
283.2	10	54.1	122.7	2926.81	15055.29	100	541	1227	6638.07
153.6	11	47.9	120	2294.41	14400	121	526.9	1320	5748
256.6	12	47.9	122.7	2294.41	15055.29	144	574.8	1472.4	5877.33
225.3	13	51	122.7	2601	15055.29	169	663	1595.1	6257.7
213.8	15	74.8	136.2	5595.04	18550.44	225	1122	2043	10187.76
171	16	51	133.5	2601	17822.25	256	816	2136	6808.5
285.7	17	63.5	128.1	4032.25	16409.61	289	1079.5	2177.7	8134.35
203.6	18	67.9	133.5	4610.41	17822.25	324	1222.2	2403	9064.65
177.3	19	51	128.1	2601	16409.61	361	969	2433.9	6533.1
310.4	20	41.6	133.5	1730.56	17822.25	400	832	2670	5553.6
289.3	21	57.3	136.2	3283.29	18550.44	441	1203.3	2860.2	7804.26
253.6	22	74.8	138.9	5595.04	19293.21	484	1645.6	3055.8	10389.72
215.8	23	57.3	133.5	3283.29	17822.25	529	1317.9	3070.5	7649.55
322.4	24	54.1	136.2	2926.81	18550.44	576	1298.4	3268.8	7368.42
266.8	25	51	133.5	2601	17822.25	625	1275	3337.5	6808.5

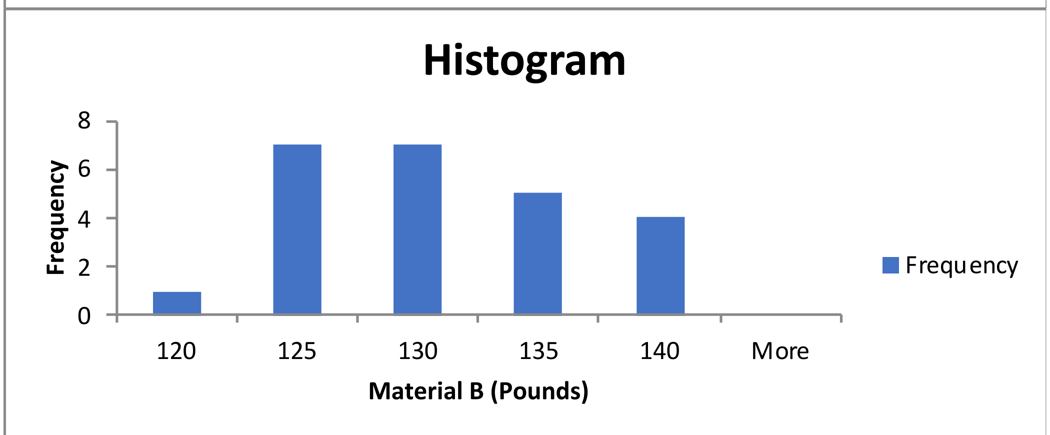
1b:

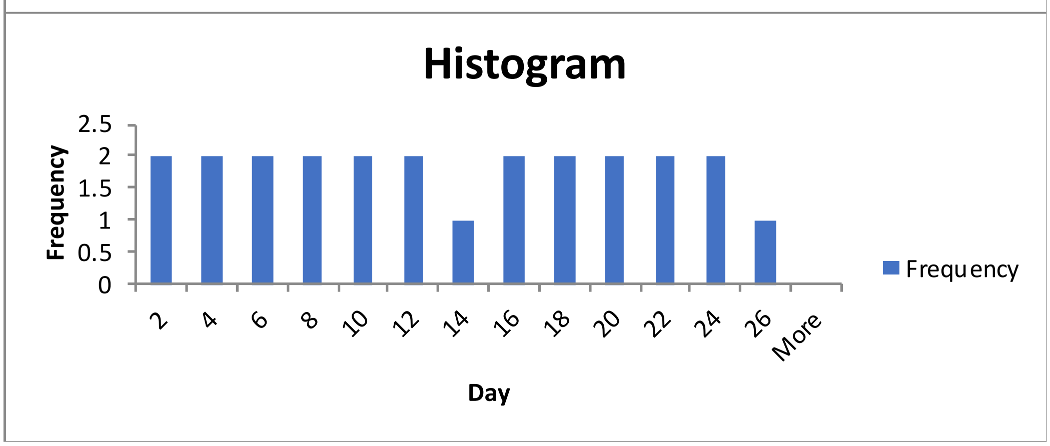
Profit		*Day*		*Material A*		*Material B*	
Mean	220.575	Mean	12.9583333	Mean	55.7791667	Mean	128.55
Standard Dev	54.3443789	Standard Dev	7.51508146	Standard Dev	10.1295083	Standard Dev	5.66729749
Minimum	127.6	Minimum	1	Minimum	41.6	Minimum	120
Maximum	322.4	Maximum	25	Maximum	74.8	Maximum	138.9
Count	24	Count	24	Count	24	Count	24

1c:

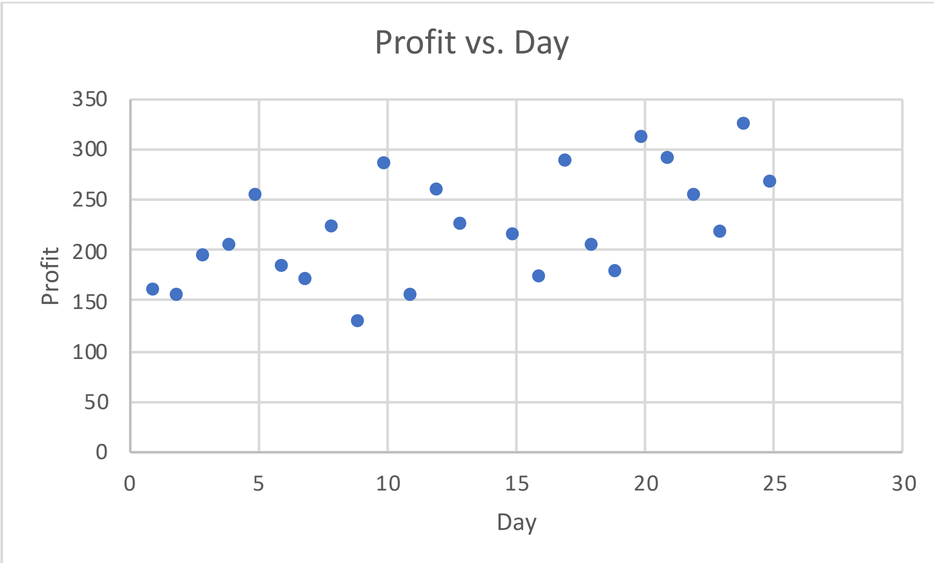


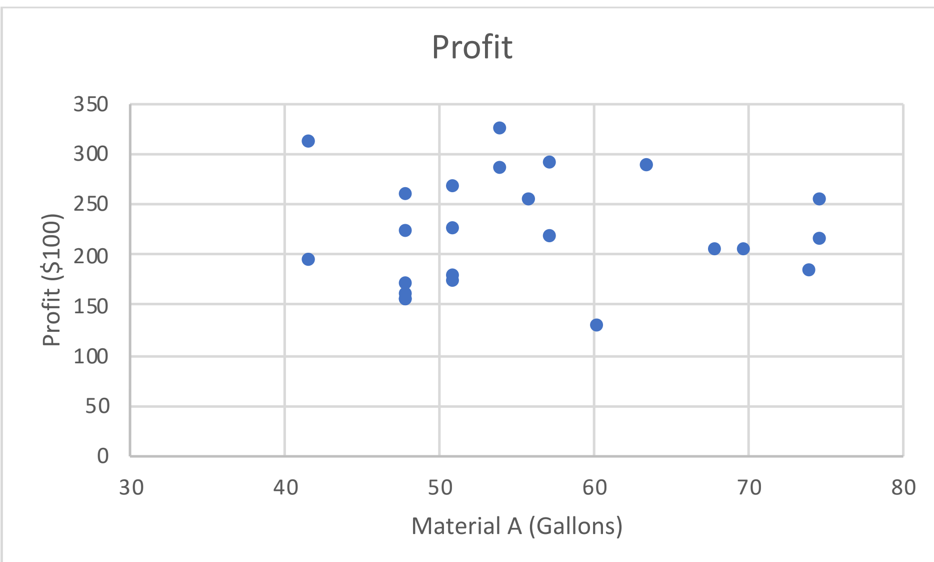

Histogram
Frequency
8
6
4
2
0
40
45
50
55
60
65
70
75
More
Material A (Gallons)
Frequency

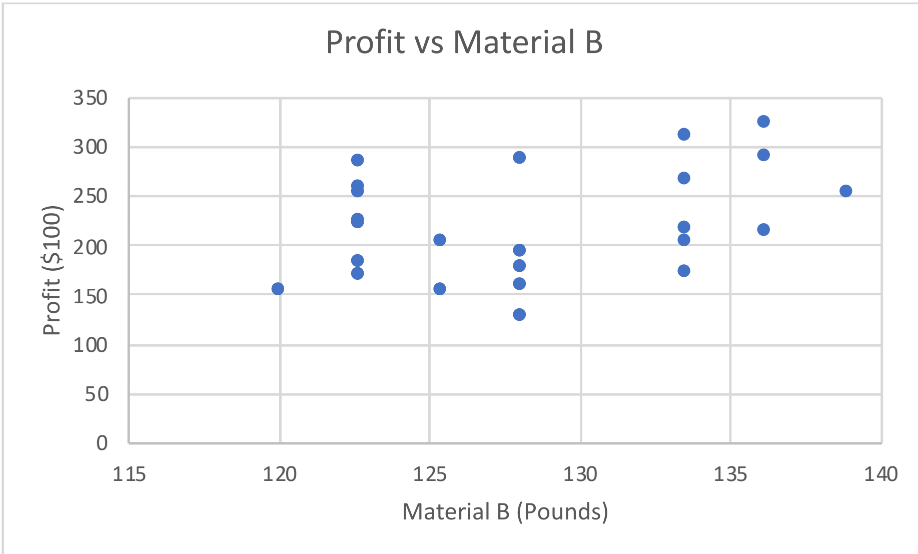

Histogram
Frequency
8
6
4
2
0
120
125
130
135
140
More
Material B (Pounds)
Frequency

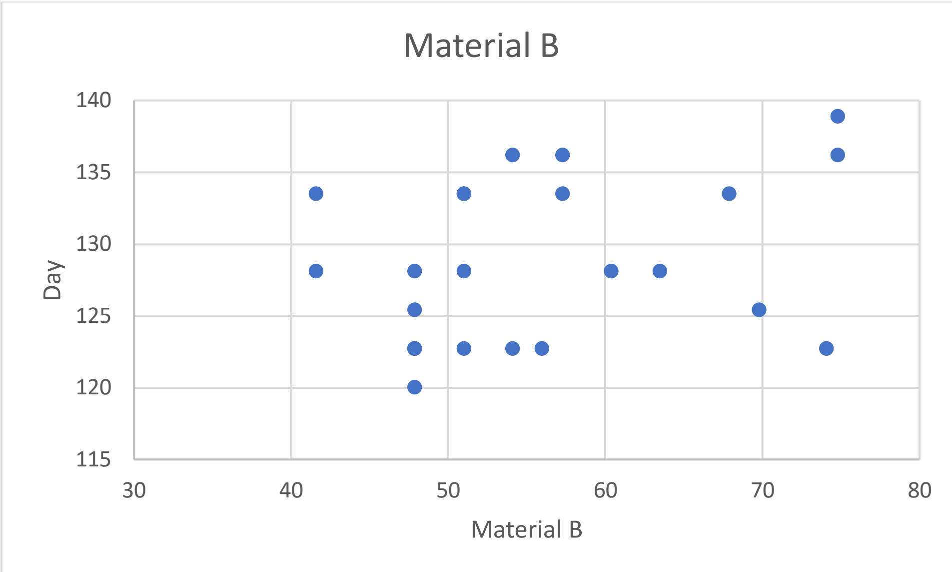

Histogram
Frequency
2.5
2
1.5
1
0.5
0
2
4
6
8
10
12
14
16
18
20
22
24
26
More
Day
Frequency

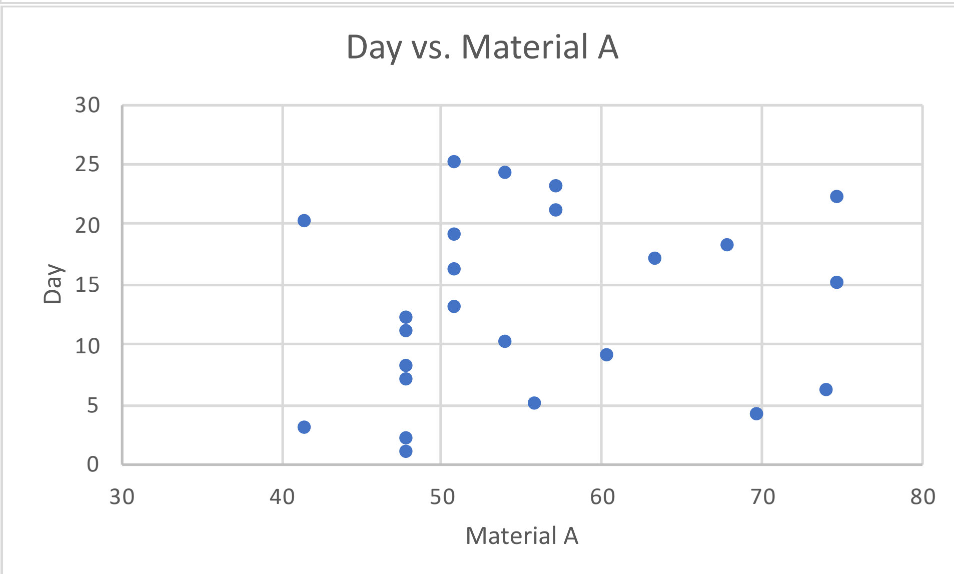
1e

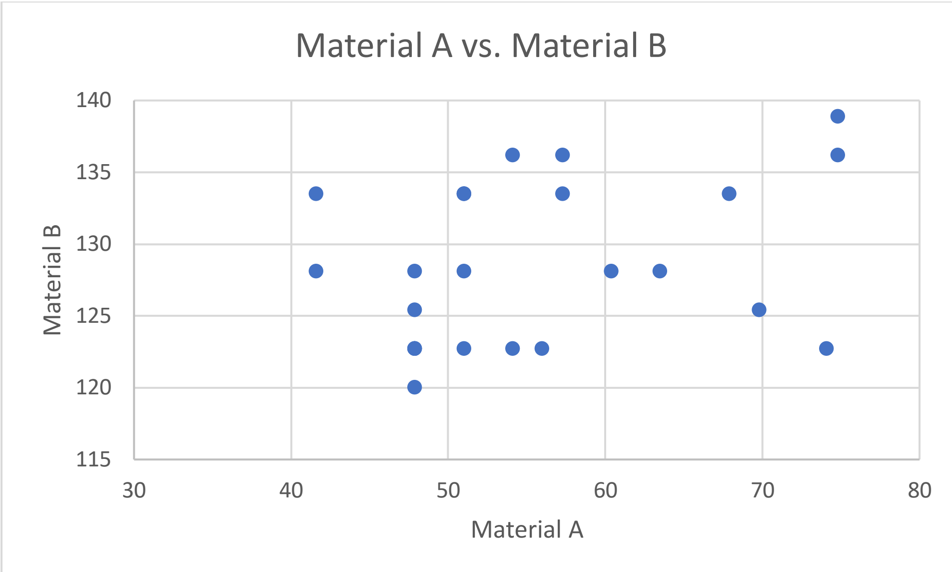

Profit vs. Day
350
300
250
200
150
100
50
0
Profit
0
5
10
15
20
25
30
Day


Profit
350
300
250
200
150
100
50
0
Profit ($100)
30
40
50
60
70
80
Material A (Gallons)


Profit vs Material B
Profit ($100)
350
300
250
200
150
100
50
0
115
120
125
130
135
140
Material B (Pounds)


Material B
140
135
130
125
120
115
Day
30
40
50
60
70
80
Material B


Day vs. Material A
30
25
20
15
10
5
0
Day
30
40
50
60
70
80
Material A



1d

	Profit	Day	Material A	Material B
Profit	1			
Day	0.56750024	1		
Material A	0.01564459	0.15631183	1	
Material B	0.35840307	0.71434066	0.31303973	1

2a-1

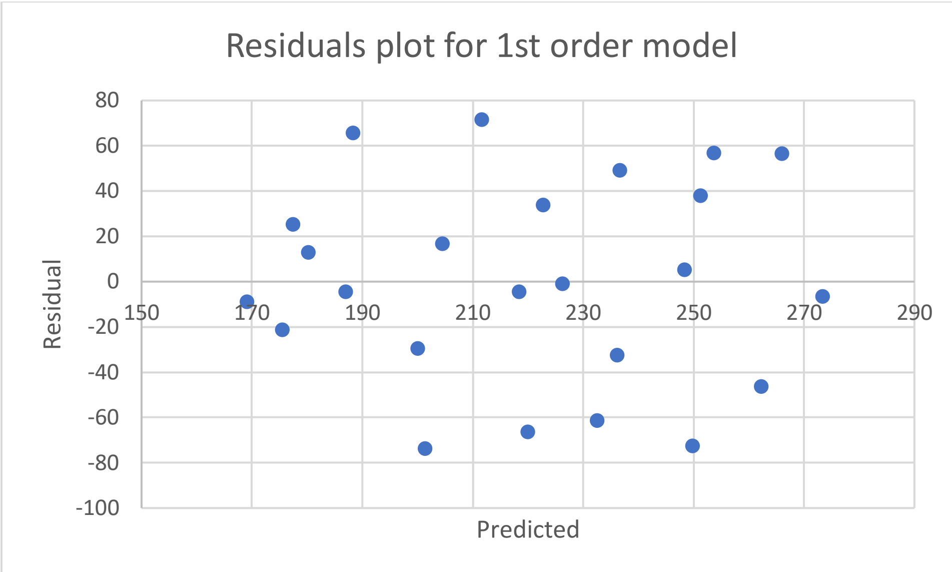
Regression Statistics	
Multiple R	0.57428177
R Square	0.32979955
Adjusted R S	0.22926948
Standard Err	47.7096147
Observations	24

ANOVA

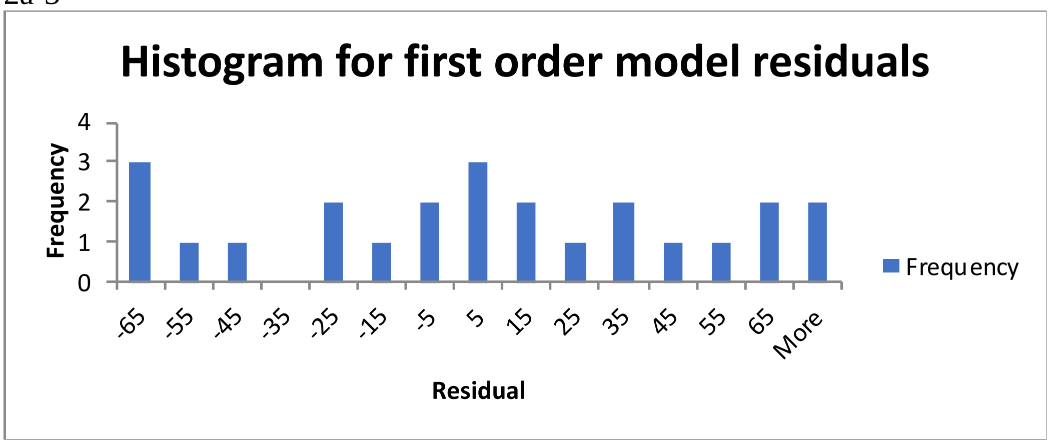
	df	SS	MS	F	Significance F
Regression	3	22402.0184	7467.33946	3.28060602	0.04218799
Residual	20	45524.1466	2276.20733		
Total	23	67926.165			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	267.518036	307.900294	0.86884631	0.39523833	-374.75072	909.786795	-374.75072	909.786795
Day	4.53959466	1.90140177	2.38749892	0.02694817	0.57334008	8.50584925	0.57334008	8.50584925
Material A	-0.3229543	1.03941088	-0.310709	0.75923678	-2.4911274	1.84521875	-2.4911274	1.84521875
Material B	-0.6826487	2.6221365	-0.2603406	0.79726255	-6.1523296	4.78703217	-6.1523296	4.78703217

2a-2



2a-3



2b-1:

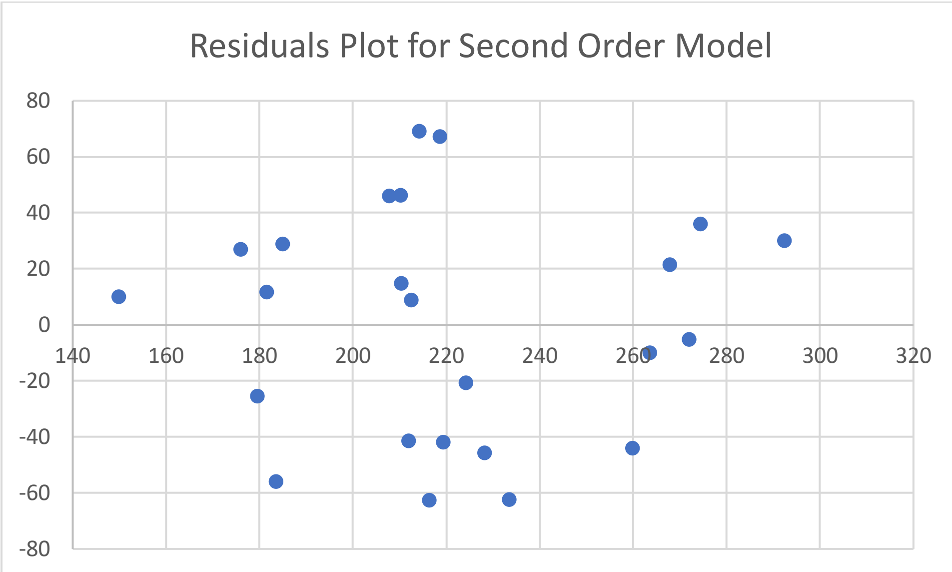
Regression Statistics	
Multiple R	0.66424303
R Square	0.4412188
Adjusted R S	0.08200232
Standard Err	52.0685425
Observations	24

ANOVA

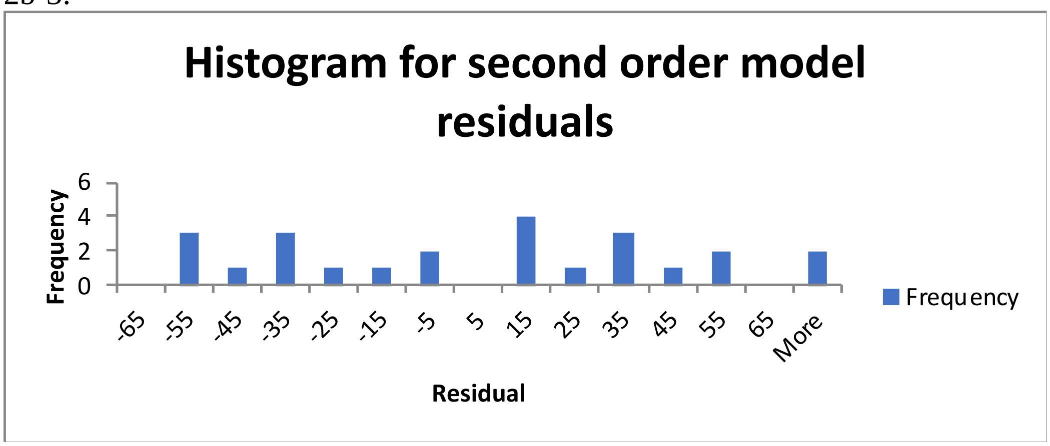
	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	29970.3013	3330.03347	1.22828106	0.35221856
Residual	14	37955.8637	2711.13312		
Total	23	67926.165			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1354.30527	15262.2652	0.08873553	0.93054896	-31379.998	34088.6086	-31379.998	34088.6086
Day	-103.47834	136.316479	-0.7591036	0.46038136	-395.84811	188.891434	-395.84811	188.891434
Material A	31.7720905	57.7373077	0.55028701	0.59079504	-92.062118	155.606299	-92.062118	155.606299
Material B	-20.195195	263.305313	-0.0766988	0.9399485	-584.92893	544.538534	-584.92893	544.538534
A^2	0.04175686	0.13907979	0.30023669	0.76840858	-0.2565396	0.34005335	-0.2565396	0.34005335
B^2	0.09781876	1.12409572	0.08701996	0.93188804	-2.3131268	2.50876431	-2.3131268	2.50876431
D^2	-0.2122283	0.45998734	-0.4613785	0.6516139	-1.198803	0.77434646	-1.198803	0.77434646
DA	0.12465339	0.32127308	0.38799822	0.70385299	-0.5644088	0.81371561	-0.5644088	0.81371561
DB	0.82484339	1.12596018	0.73256889	0.4759083	-1.590101	3.23978779	-1.590101	3.23978779
AB	-0.3012005	0.44857006	-0.6714682	0.51285023	-1.2632876	0.66088658	-1.2632876	0.66088658

2b-2:


Residuals Plot for Second Order Model
80
60
40
20
0
-20
-40
-60
-80
140
160
180
200
220
240
260
280
300
320

2b-3:


Histogram for second order model residuals
Frequency
6
4
2
0
-65
-55
-45
-35
-25
-15
-5
5
15
25
35
45
55
65
More
Residual
Frequency

3b-1

Final Model:

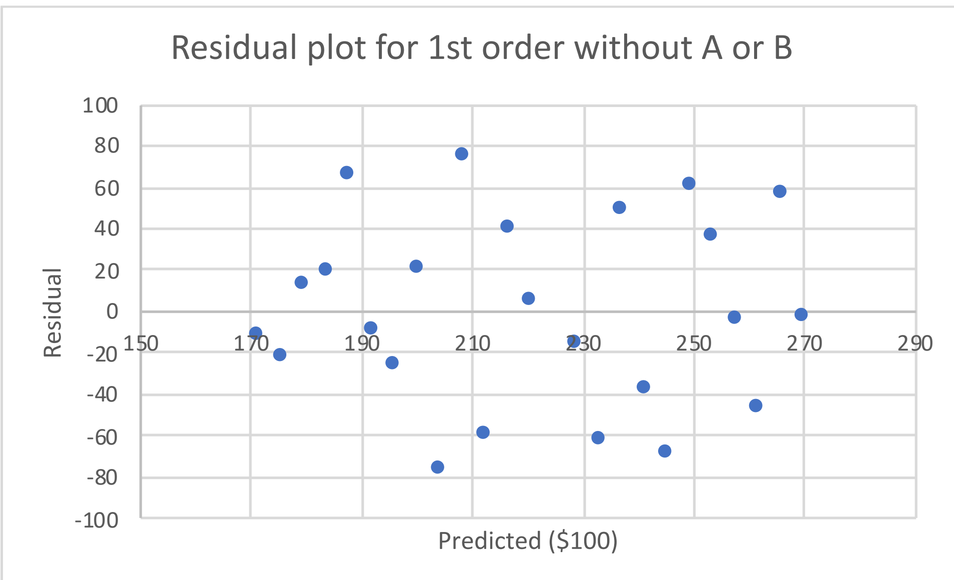
Regression Statistics	
Multiple R	0.56750024
R Square	0.32205652
Adjusted R S	0.29124091
Standard Err	45.7513541
Observations	24

ANOVA

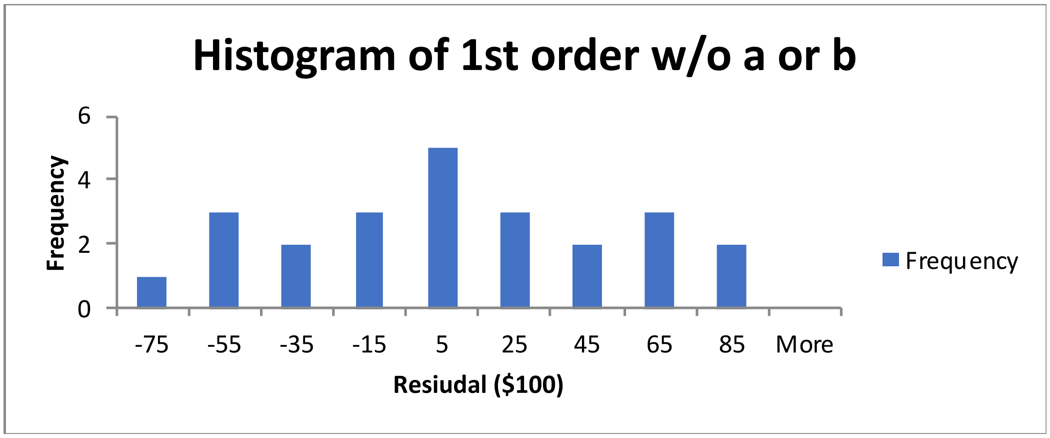
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	21876.0642	21876.0642	10.4510827	0.00382433
Residual	22	46050.1008	2093.1864		
Total	23	67926.165			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	167.396494	18.9157507	8.84958239	1.0622E-08	128.167628	206.62536	128.167628	206.62536
Day	4.10380754	1.26942294	3.23281344	0.00382433	1.47118549	6.73642958	1.47118549	6.73642958

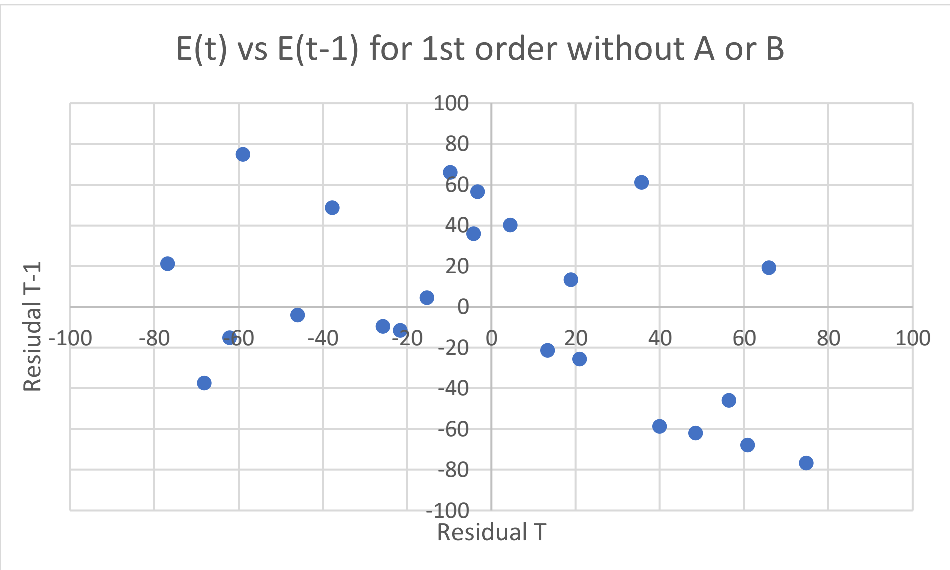
3b-2:



3b-3:



3b-4:



3b-5:

Use regression and fit E(t) = Bo + B1(E(t-1)) yields r = -.424 and p = .043

Regression Statistics	
Multiple R	0.4241709
R Square	0.17992095
Adjusted R S	0.14086957
Standard Err	42.3430131
Observations	23

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.55879331	8.82917103	0.06328944	0.95013458
res t-1	-0.4235837	0.19734067	-2.1464593	0.04367422