EXECUTIVE SUMMARY

Problem: Data for a random sample of profit measurements on 25 individuals. However, observation 14 has a missing entry, so we are excluding the entire observation. Input the remaining 24 observations. Our object is to characterize the dependent variable PROFIT (P) (in $100). Along with the value of Profit for 24 individuals, we have been provided with the value of two independent variables: Material A (gallons), Material B (lbs), and the days that the data was collected over 25 consecutive days and is listed in DAY. Besides characterizing the relationship between these four variables, we wish to predict PROFIT for an individual with Material A=35 gallons and Material B=95 lbs on day 26, and give a 90% confidence interval for these predictions.

Variable	Mean	Std.Dev.	Minimun	Maximum	Shape of Distribution
Profit($100)	222.575	54.34437893	129.6	324.4	unimodal
DAY	12.95833333	7.515081455	1	25	uniform
A(gallons)	43.775	9.714769124	30.2	62	unimodal, skewed to the right
B(lbs)	71.13333333	2.9385987	66.7	76.5	unimodal, skewed to the right

Recommended Model w DAY: $\widehat{PROFIT}$= 169.39+ 4.1038* DAY
Profit (P) in units of $100; DAY (D). $S_{P,A}$=$45.75, R^2= 0.32.
Using this first order model, for an individual with Material A=35 gallons and Material B=95 lbs on day 26 is 275.99 ($100) with 90% confidence interval for this prediction given by the interval (190.93,361.05)(in $100).
Recommended Model w/o DAY: $\widehat{PROFIT}$= 222.575.
Profit (P) in units of $100; S_P=$ 54.26, R^2= 0.047.
Using this first order model, for an individual with Material A=35 gallons and Material B=95 lbs on day 26 is 275.99 ($100) with 90% confidence interval for this prediction given by the interval (203.56,241.59)(in $100).

ANALYSIS

Problem: Data for a random sample of profit measurements on 25 individuals. However, observation 14 has a missing entry, so we are excluding the entire observation. Input the remaining 24 observations. Our object is to characterize the dependent variable PROFIT (P) (in $100). Along with the value of Profit for 24 individuals, we have been provided with the value of two independent variables: Material A (gallons), Material B (lbs), and the days that the data was collected over 25 consecutive days and is listed in Day. Besides characterizing the relationship between these four variables, we wish to predict PROFIT for an individual with Material A=35 gallons and Material B=95 lbs on day 26, and give a 90% confidence interval for these predictions.

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 $100), the dependent variable, has an average of 222.58, standard deviation of 54.34, ranges from a minimum value of 129.60 to a maximum of 324.40. The shape of the distribution appears unimodal. (p.1c)

Day: an independent variable, has an average of 12.96, standard deviation of 7.52, ranges from a minimum value of 1 to a maximum of 25. The shape of the distribution appears uniform. (p.1c)

Material A: an independent variable, has an average of 43.78 gallons, standard deviation of 9.71 gallons, ranges from a minimum value of 30.20 gallons to a maximum of 62 gallons. The shape of the distribution appears unimodal, skewed to the right. (p.1c)

Material B: an independent variable, has an average of 71.13 lbs, standard deviation of 2.94 lbs, ranges from a minimum value of 66.70 lbs to a maximum of 76.50 lbs. The shape of the distribution appears unimodal, skewed to the right. (p.1c)

PROFIT (in $100) vs independent variables: examining the corr. matrix (p.1d), we see the following significant results: the corr. between:
<u>PROFIT vs DAY</u>: r= 0.57, about 32.49% of the variability in the y scores around y-bar is explained by simple regression between PROFIT and DAY. The std. dev. of the y scores about a simple linear regression using DAY is approximately 0.82 times the std. dev. of the y scores about y-bar. As DAY increases, y-hat increases in this fitted simple regression model.

<u>PROFIT vs Material B</u>: r= 0.36, about 12.96% of the variability in the y scores around y-bar is explained by simple regression between PROFIT and Material B. The std. dev. of the y scores about a simple linear regression using Material B is approximately 0.93 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 Material A: r= 0.016, about 0.0256% of the variability in the y scores around y-bar is explained by simple regression between PROFIT and Material A. The std. dev. of the y scores about a simple linear regression using Material A is approximately 0.99 times the std. dev. of the y scores about y-bar. As Material A increases, y-hat increases in this fitted simple regression model.

Correlations between pairs of independent variables: Examining the correlation matrix (p1d), we see the following significant results: for DAY vs Material B, r= 0.71, about 50.41% of the variability in the y scores around y-bar is explained by simple regression between DAY and Material B. The std. dev. of the y scores about a simple linear regression using Material B is approximately 0.70 times the std. dev. of the y scores about y-bar. As Material B increases, y-hat increases in this fitted simple regression model.
For Material B vs Material A, r= 0.31, about 9.61% of the variability in the y scores around y-bar is explained by simple regression between Material B and Material A. The std. dev. of the y scores about a simple linear regression using Material A is approximately 0.95 times the std. dev. of the y scores about y-bar. As Material A increases, y-hat increases in this fitted simple regression model.
For DAY vs Material A, r= 0.16, about 2.56% of the variability in the y scores around y-bar is explained by simple regression between DAY and Material A. The std. dev. of the y scores about a simple linear regression using Material A is approximately 0.99 times the std. dev. of the y scores about y-bar. As Material A increases, y-hat increases in this fitted simple regression model.

Scatterplots of PROFIT vs independent variables: Describe the relationship:
PROFIT vs Material A: There doesn't appears to be a linear relationship between these two variables.
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:
Material A vs DAY: Describe the relationship: There appears to be a slightly positive linear relationship between these two variables.
Material B vs DAY: Describe the relationship: There appears to be a positive linear relationship with a jump between these two variables.
Material A vs Material B: Describe the relationship: There appears to be a slightly positive linear relationship between these two variables.

There is significant relationship found between DAY and Material B and PROFIT after examining correlation matrix and scatterplots.

Fit of zero order model w/o DAY: $E(P)=\beta_0$ (p.4c). Examining the EXCEL results for this model, we find s_P= \$ 52.78 vs S_P=54.34, R^2= 0.047. For testing Ho: $\beta_0=\mu=0$, (since A holds), we find the p for the t_{obs}= 0<0.05=α. Thus, the zero order model is a significant improvement.

Fit of first order model w/o DAY: $E(P)=\beta_0+\beta_1*A+\beta_2*B$ (p.4a). Examining the EXCEL results for this model, we find $s_{P,A,B}$= $ 54.26 vs S_P=54.34, R^2= 0.14. For testing Ho: $\beta_1=\beta_2=0$, (since AB holds), we find the p for the Global F= 0.21> 0.05=α. Thus, the first order model is not a significant improvement over the model E(P)= β_0.

Fit of second order model w/o DAY: (p.4b)
$E(P)=\beta_0+\beta_1*A+\beta_2*B+\beta_3*A^2+\beta_4*B^2+\beta_5*A*B$.
We see the standard deviation of the residuals about the second order model is $ 50.03, which we note is not significantly lower than the standard deviation of the y scores about the first order model. R^2= 0.34. For testing H_0: $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$, (since ABC holds), we find the partial F is obtained by [(64767-45056)/(22-18)]/ 2503= 1.97. We reject this value if it is too large. Note since the cutoff point with α=0.05 on an $F_{5,18}$ is approximately 2.77. (Reject H_0 if F_{obs}>2.77, accept H_0 otherwise.) We accept H_0, and thus P>0.05. We conclude the second order model is not significantly better than the zero order model.

Residual plot for the Highest order model fitted: Examining the residual plot for the second model (p.4b), we note: the mean of the residuals seems to be zero regardless of the value if 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 zero order model (the current best model w/o DAY): We now examine the zero order model, to see if we can drop any of variables. $E(P)=\beta_0$, where β_0 is equal to the mean of PROFIT, $\beta_0=\mu=222.575$.

We only look at the zero order terms, i.e. terms not included in higher order terms in the model, and individually test; p.4c (A holds since we are testing a single β=0.)
H_0: $\beta_0=\mu=0$ vs H_0: not H_0; (A holds) [p= 0 <0.05, reject H_0]
If we find one or more p values above α=0.05, we drop the one with the highest p-value, and refit without this term. In this case, we drop do not drop any term.
Examining the residual plot for zeroth order model, the mean of the residuals is 0. However, we can't check the variability of the residuals for A2 since there is only one value.
Histogram of residuals for the FINAL model w/o DAY: Examining the histogram of the residuals (p.4c-3), we see that they appear to be more or less symmetric ad unimodal, and we can't reject that they are normally distributed.
Independence of the residuals for the FINAL model w/o DAY: Given that the data has been ordered by the order it was collected, we examine the plot of residual t vs residual t-1 (p.4c-4). There doesn't appear to be a line through this data which has a very positive or negative slope, so we conclude the residuals are independent. Indeed, if we regress residual t on residual t-1, we get a correlation of r=-0 with a p=0.73>0.05=α and conclude this is not significantly different from zero (p.4b-5). We could also calculate the Durbin-Watson statistics =133438.771/67926.165=1.96, which ranges from 0-4 with values near 2 indicating no correlation, 0 pos. corr., and 4 neg. corr. Therefore, there is

enough evidence to prove the independence of the residuals for the final model, at most a very slightly positive correlation.

The **prediction model w/o DAY** is given by $\widehat{PROFIT}$= 222.575.
The predicted profit for an individual with Material A=35 gallons and Material B=95 lbs on day 26 is 222.575 ($100), obtained by substituting this value in the above prediction equation. A 90% confidence interval for this prediction is given by the above prediction plus or minus $t_{dferror,\alpha/2}*S_{\beta o}$, 222.575+/-1.717*11.08=(203.56,241.59)(in $100).

Conclusion: We recommend using the zero order model given above to describe the relationships if do not include DAY variable.

However, since we found a significant relationship between DAY and Material B and PROFIT after examining correlation matrix and scatterplots. We should also consider to include DAY as an independent variable into our model.

Fit of first order model w DAY: $E(P)=\beta_o+\beta_1*A+\beta_2*B+\beta_3*D$ (p.2a). Examining the EXCEL results for this model, we find $s_{P,A,B,D}$= $47.71, R^2= 0.33. For testing Ho: $\beta_1=\beta_2=\beta_3=0$, (since AB holds), we find the p for the Global F= 0.042 < 0.05=α. Thus, the first order model is a significant improvement over the model $E(P)=\beta_o$.

Fit of second order model w DAY: (p.2b)
$E(P)=\beta_o+\beta_1*A+\beta_2*B+\beta_3*D+\beta_4*A^2+\beta_5*B^2+\beta_6*A*B+\beta_7*A*D+\beta_8*B*D$.
We see the standard deviation of the residuals about the second order model is $52.08, which we note is not significantly lower than the standard deviation of the y scores about the first order model. R^2= 0.66. For testing H_o: $\beta_4=\beta_5=\beta_6=\beta_7=\beta_8=0$, (since ABC holds), we find the partial F is obtained by [(45529-37966)/(20-14)]/ 2712= 0.4648 . We reject this value if it is too large. Note since the cutoff point with α=0.05 on an $F_{9,14}$ is approximately 2.65. (Reject H_o if F_{obs}>2.65, accept H_o otherwise.) We accept H_o, and thus P>0.05. We conclude the second order model is not significantly better than the first order model.

Residual plot for the Highest order model fitted w DAY: Examining the residual plot for the second model (p.2b), we note: the mean of the residuals seems to be zero regardless of the value if 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 w DAY: We now examine the first order model, to see if we can drop any of variables. We only look at the higher order terms, i.e. terms not included in higher order terms in the model, and individually test; p.2a (A holds since we are testing a single β=0.)
H_o: β_3=0 vs H_o: not H_o; (A holds) [p= 0.027<0.05, reject H_o]
H_o: β_2=0 vs H_o: not H_o; (A holds) [p= 0.79>0.05, accept H_o]
H_o: β_1=0 vs H_o: not H_o; (A holds) [p= 0.76>0.05, accept H_o]
If we find one or more p values above α=0.05, we drop the one with the highest p-value, and refit without this term. In this case, we drop A term and refit.

Fit of reduced first order of form w DAY: $E(P)=\beta_0+\beta_1*B+\beta_2*D$ (p.3a). Examining the EXCEL results for this model, we find $s_{P,B,D}$= \$ 46.67, R^2= 0.33.
We only look at the higher order terms and individually test
H_o: $\beta_2=0$ vs H_o: not H_o; (A holds) [p= 0.02<0.05, reject H_o]
H_o: $\beta_1=0$ vs H_o: not H_o; (A holds) [p= 0.71>0.05, accept H_o]
If we find one or more p values above $\alpha=0.05$, we drop the one with the highest p-value, and refit without this term. In this case, we drop B term and refit.
Fit of reduced first order of form w DAY: $E(P)=\beta_0+\beta_1*D$ (p.3b). Examining the EXCEL results for this model, we find $s_{P,D}$= \$ 45.75, R^2= 0.32. We only look at the higher order terms and individually test
H_o: $\beta_1=0$ vs H_o: not H_o; (A holds) [p= 0.004<0.05, reject H_o]
If we find one or more p values above $\alpha=0.05$, we drop the one with the highest p-value, and refit without this term. In this case, we have no more terms to drop.

Residual plot for the FINAL model w DAY: Examining the residual plot for this model (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 model w DAY: Examining the histogram of the residuals (p.3b-3), we see that they appear to be more or less symmetric ad unimodal, and we can't reject that they are normally distributed.
Independence of the residuals for the FINAL model w DAY: Given that the data has been ordered by the order it was collected, we examine the plot of residual t vs residual t-1 (p.3b-4). a) It only appears to be a line through this data with a slight negative slope. b) Indeed, if we regress residual t on residual t-1 to fit $E(e_t)=\beta_0+\beta_1 e_{t-1}$, we get a correlation of r=-0.43 with a p=0.044<0.05= α. Therefore, we reject H_o: $\beta_1=0$, and conclude this is significantly different from zero, and $\beta_1<0$ (p.3b-5). c) calculate the Durbin-Watson statistics =130957.706/46039.91=2.84, which ranges from 0-4 with values near 2 indicating no correlation, 0 pos. corr., and 4 neg. corr. Therefore, there is not enough evidence to prove the independence of the residuals for the final model. There appears a possible negative correlation of the residuals. Given a,b,c we conclude that it is very possible that the A4 violated and there exists negative corr. residuals.

The **prediction model w DAY** is given by $\widehat{PROFIT}$= 169.39+ 4.1038* DAY.
The predicted profit for an individual with Material A=35 gallons and Material B=95 lbs on day 26 is 275.99 (\$100), obtained by substituting this value in the above prediction equation. A 90% confidence interval for this prediction is given by the above prediction plus or minus $t_{dferror,\alpha/2}*S_{P,D}$*sqrt(1+(1/n)+0.1309)), 275.99 +/- 1.717*45.75*1.0829=(190.93,361.05)(in \$100).

Conclusion: We recommend using the first order model given above to describe the relationships if include DAY variable.

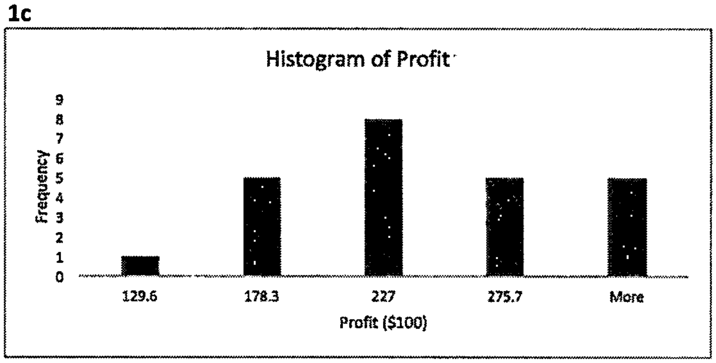
1

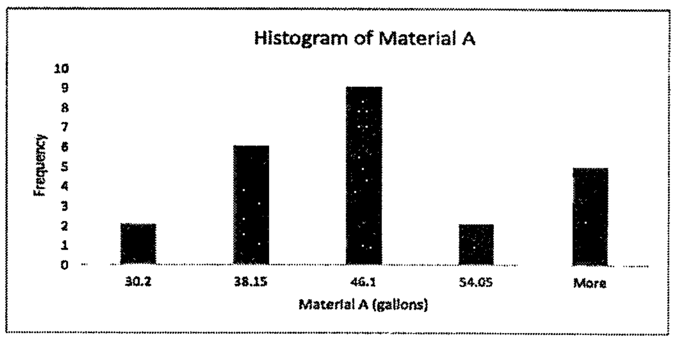
Profit($100)	A(gallons)	B(lbs)	Day	A^2(gallons^	B^2(lbs^2)	Day^2	A*B(gallons*	A*D(gallons)	B*D(lbs)
162	36.2	70.9	1	1310.44	5026.81	1	2566.58	36.2	70.9
156.1	36.2	69.5	2	1310.44	4830.25	4	2515.9	72.4	139
195.1	30.2	70.9	3	912.04	5026.81	9	2141.18	90.6	212.7
204.8	57.2	69.5	4	3271.84	4830.25	16	3975.4	228.8	278
255.8	44	68.1	5	1936	4637.61	25	2996.4	220	340.5
184.4	61.4	68.1	6	3769.96	4637.61	36	4181.34	368.4	408.6
172.4	36.2	68.1	7	1310.44	4637.61	49	2465.22	253.4	476.7
223.2	36.2	68.1	8	1310.44	4637.61	64	2465.22	289.6	544.8
129.6	48.2	70.9	9	2323.24	5026.81	81	3417.38	433.8	638.1
285.2	42.2	68.1	10	1780.84	4637.61	100	2873.82	422	681
155.6	36.2	66.7	11	1310.44	4448.89	121	2414.54	398.2	733.7
258.6	36.2	68.1	12	1310.44	4637.61	144	2465.22	434.4	817.2
227.3	39.2	68.1	13	1536.64	4637.61	169	2669.52	509.6	885.3
215.8	62	75.1	15	3844	5640.01	225	4656.2	930	1126.5
173	39.2	73.7	16	1536.64	5431.69	256	2889.04	627.2	1179.2
287.7	51.2	70.9	17	2621.44	5026.81	289	3630.08	870.4	1205.3
205.6	55.4	73.7	18	3069.16	5431.69	324	4082.98	997.2	1326.6
179.3	39.2	70.9	19	1536.64	5026.81	361	2779.28	744.8	1347.1
312.4	30.2	73.7	20	912.04	5431.69	400	2225.74	604	1474
291.3	45.2	75.1	21	2043.04	5640.01	441	3394.52	949.2	1577.1
255.6	62	76.5	22	3844	5852.25	484	4743	1364	1683
217.8	45.2	73.7	23	2043.04	5431.69	529	3331.24	1039.6	1695.1
324.4	42.2	75.1	24	1780.84	5640.01	576	3169.22	1012.8	1802.4
268.8	39.2	73.7	25	1536.64	5431.69	625	2889.04	980	1842.5

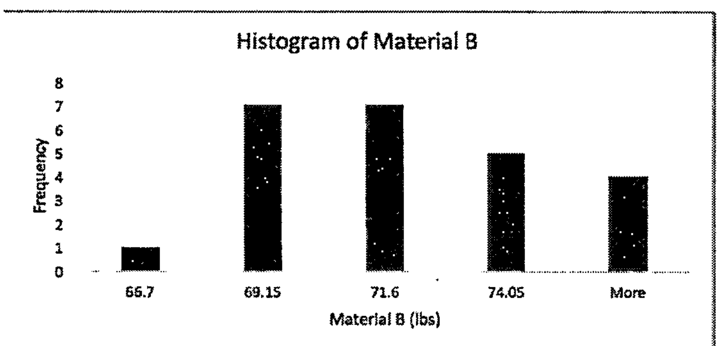
1b

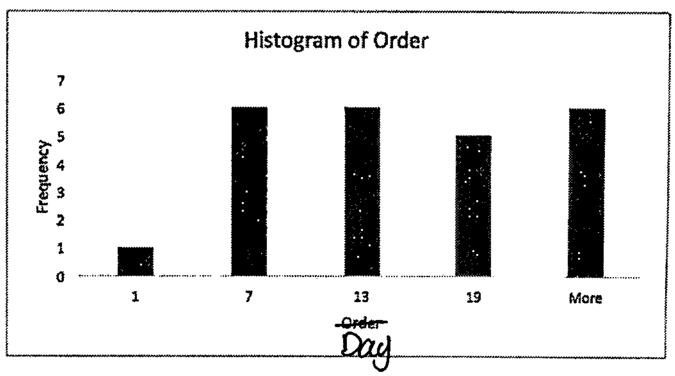
Day		*Profit($100)*		*A(gallons)*		*B(lbs)*	
Mean	12.9583333	Mean	222.575	Mean	43.775	Mean	71.1333333
Standard Dev	7.51508146	Standard Deviation	54.3443789	Standard Dev	9.714769124	Standard Dev	2.9385987
Minimum	1	Minimum	129.6	Minimum	30.2	Minimum	66.7
Maximum	25	Maximum	324.4	Maximum	62	Maximum	76.5
Count	24	Count	24	Count	24	Count	24

1c







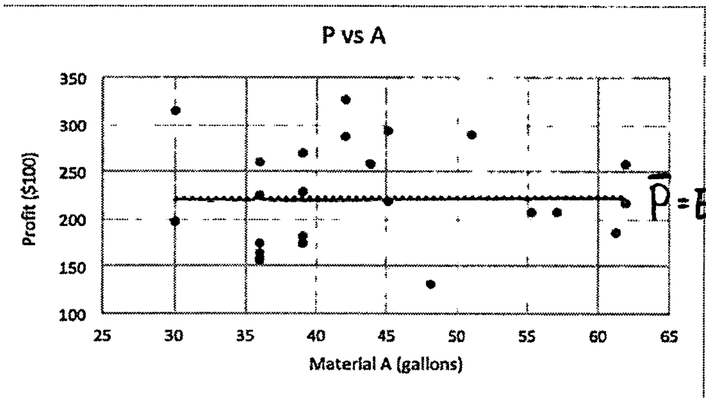


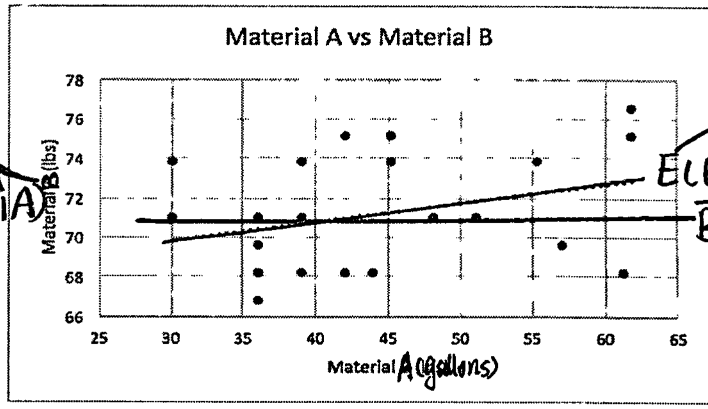
②

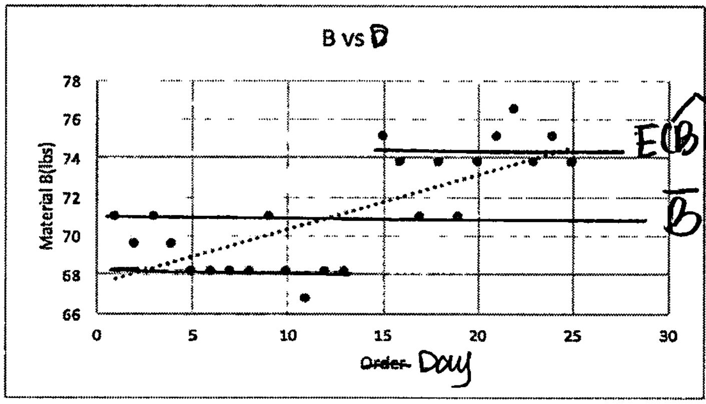
1d

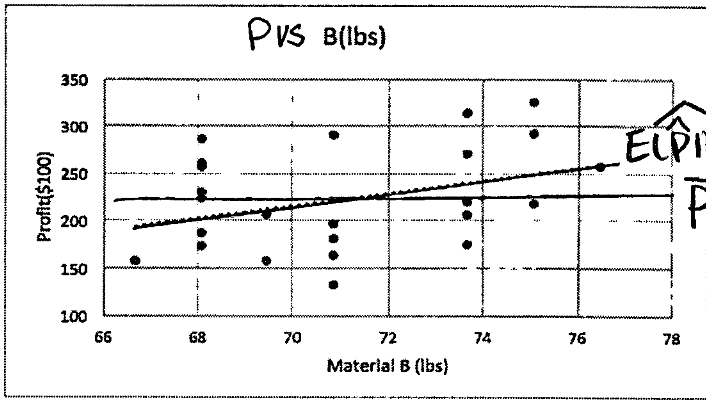
	Profit($100)	A(gallons)	B(lbs)	Day
Profit($100)	1			
A(gallons)	0.01637405	1		
B(lbs)	0.35840307	0.312793376	1	
Day	0.56750024	0.156372134	0.71434066	1

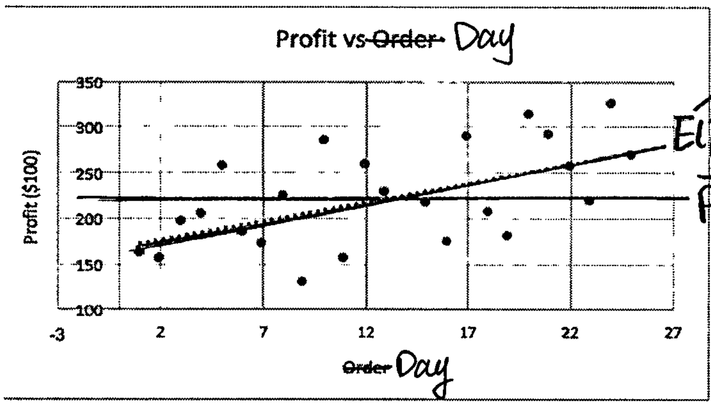
1e

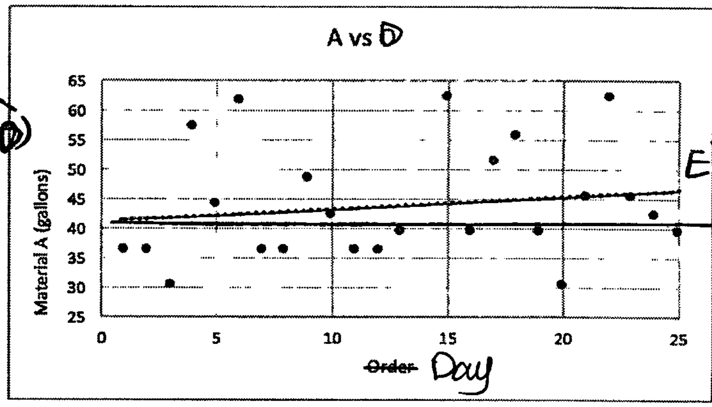












③

First Order Model w Day
SUMMARY OUTPUT
2a-1

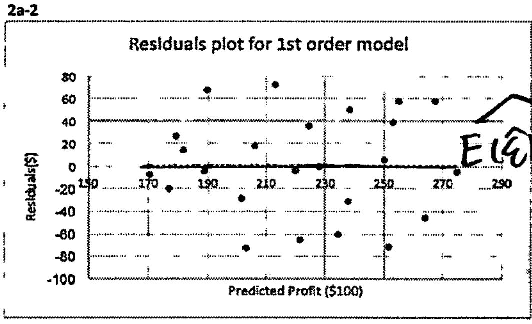
Regression Statistics	
Multiple R	0.57421156
R Square	0.32971891
Adjusted R Sq	0.22917675
Standard Erro	47.7124846
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	22396.5413	7465.51377	3.27940939	0.042234753
Residual	20	45529.6237	2276.48119		
Total	23	67926.165			

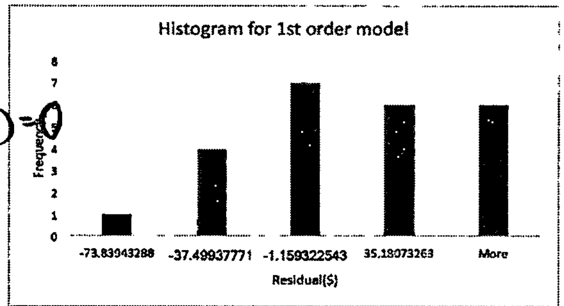
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	272.399322	331.7624372	0.8210674	0.42128757	-419.6449949	964.443639	-419.64499	964.443639
A(gallons)	-0.3324779	1.083717276	-0.306794	0.76217152	-2.593072552	1.9281167	-2.5930726	1.9281167
B(lbs)	-1.3229829	5.056612643	-0.2616342	0.79627899	-11.87089207	9.22492621	-11.870892	9.22492621
Day	4.54055957	1.901445429	2.38795156	0.02692255	0.574213912	8.50690524	0.57421391	8.50690524

2a-2



$E(\hat{\varepsilon}|\hat{P})=0$

2a-3



Second Order Model w Day
SUMMARY OUTPUT
2b-1

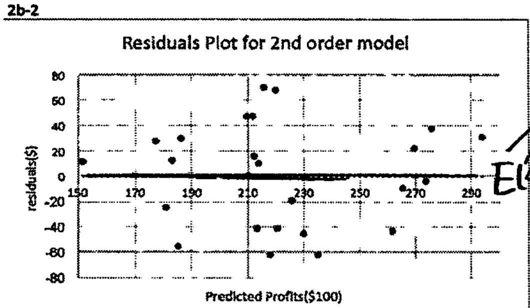
Regression Statistics	
Multiple R	0.6641303
R Square	0.44106906
Adjusted R Sq	0.08175631
Standard Erro	52.0755188
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	29960.12975	3328.90331	1.22753524	0.352588158
Residual	14	37966.03525	2711.85966		
Total	23	67926.165			

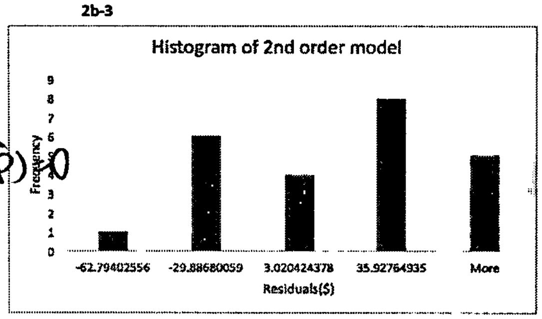
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1952.22284	17938.08646	0.10883116	0.91488089	-36521.14622	40425.5919	-36521.146	40425.5919
A(gallons)	37.0673626	63.26548711	0.5859018	0.56726767	-98.62361196	172.758337	-98.623612	172.758337
B(lbs)	-50.122969	549.4860487	-0.0912179	0.92861176	-1228.653331	1128.40739	-1228.6533	1128.40739
Day	-108.97089	145.8259405	-0.7472669	0.46726857	-421.7364307	203.794641	-421.73643	203.794641
A^2(gallons^	0.04418618	0.151292252	0.29205845	0.77452516	-0.280303427	0.36867579	-0.2803034	0.36867579
B^2(lbs^2)	0.37981221	4.180872988	0.09084519	0.92890259	-8.587268519	9.34689294	-8.5872685	9.34689294
Day^2	-0.2118594	0.460166426	-0.4603973	0.65230083	-1.198818184	0.77509947	-1.1988182	0.77509947
A*B(gallons*	-0.6091048	0.90130462	-0.6758035	0.51017528	-2.542210994	1.32400131	-2.542211	1.32400131
A*D(gallons)	0.13109845	0.335105708	0.39121521	0.7015275	-0.587631813	0.84982871	-0.5876318	0.84982871
B*D(lbs)	1.58471861	2.171457706	0.72979483	0.47754979	-3.072594967	6.24203219	-3.072595	6.24203219

2b-2



$E(\hat{\varepsilon}|\hat{P})=0$

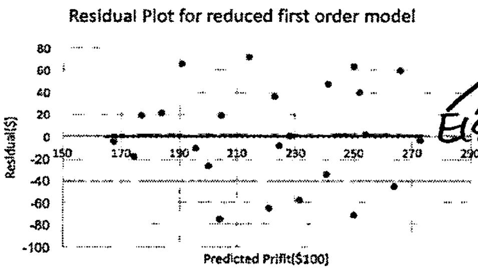
2b-3



④

Refit of First order Model w D
SUMMARY OUTPUT
3a-1

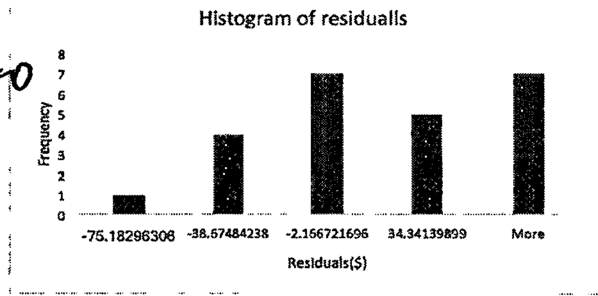
Regression Statistics	
Multiple R	0.57145821
R Square	0.32656449
Adjusted R Sc	0.26242777
Standard Errc	46.6720534
Observations	24



$\widehat{E(\hat{\varepsilon}|\hat{P})} = 0$

3a-2

3a-3



ANOVA

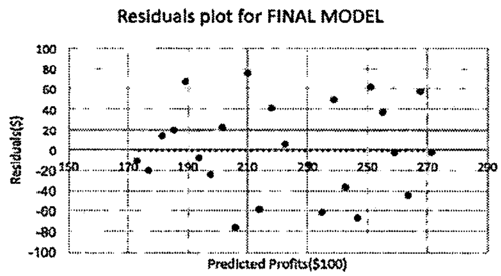
	df	SS	MS	F	Significance F
Regression	2	22182.27311	11091.1366	5.09169329	0.015743749
Residual	21	45743.89189	2178.28057		
Total	23	67926.165			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	289.187439	320.0830326	0.90347632	0.3765208	-376.4616668	954.836545	-376.46167	954.836545
B(lbs)	-1.7743198	4.732383471	-0.3749315	0.71147123	-11.61584997	8.0672104	-11.61585	8.0672104
Day	4.59942187	1.850489046	2.48551694	0.021438	0.751119227	8.44772451	0.75111923	8.44772451

Refit of First order Model w Day
SUMMARY OUTPUT
3b-1

Regression Statistics	
Multiple R	0.56750024
R Square	0.32205652
Adjusted R Sc	0.29124091
Standard Errc	45.7513541
Observations	24

3b-2

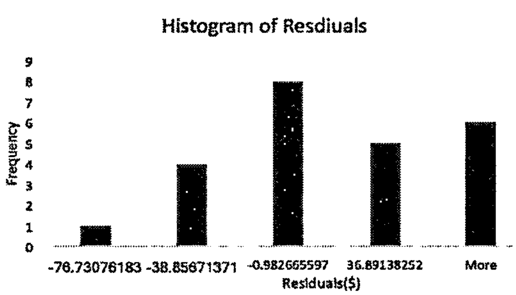


ANOVA

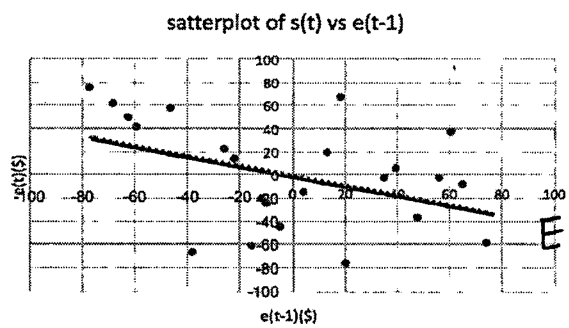
	df	SS	MS	F	Significance F
Regression	1	21876.06425	21876.0642	10.4510827	0.003824333
Residual	22	46050.10075	2093.1864		
Total	23	67926.165			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	169.396494	18.91575067	8.95531438	8.6376E-09	130.1676281	208.62536	130.167628	208.62536
Day	4.10380754	1.269422938	3.23281344	0.00382433	1.471185494	6.73642958	1.47118549	6.73642958

3b-3



3b-4



$\widehat{E(\hat{e}(t)|e(t-1))}$

5

RESIDUAL OUTPUT OF THIRD ORDER MODEL

Use Regression and fit E(e(t))=beta0+beta1*e(t-1) yields r=-0.43 with p=0.044

SUMMARY OUTPUT

3b-5

Regression Statistics	
Multiple R	0.43305135
R Square	0.18753347
Adjusted R Sc	0.14691015
Standard Errc	42.9480949
Observations	22

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8515.12681	8515.12681	4.61639872	0.044095699
Residual	20	36890.77707	1844.53885		
Total	21	45405.90388			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.78784571	9.157543239	0.19523202	0.84718105	-17.31445475	20.8901462	-17.314455	20.8901462
e(t-1)	-0.4307243	0.200469245	-2.1485806	0.0440957	-0.848895853	-0.0125528	-0.8488959	-0.0125528

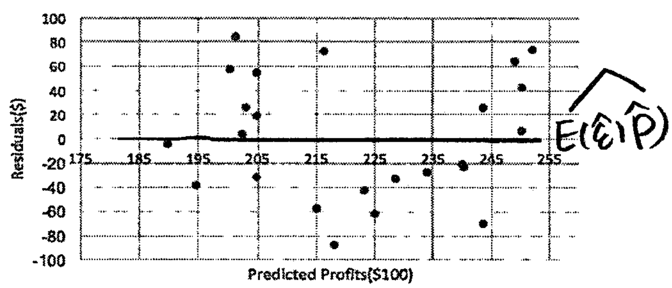
First Order Model w/o Day

SUMMARY OUTPUT

4a-1

4a-2

Residuals plot for the 1st order w/o Day



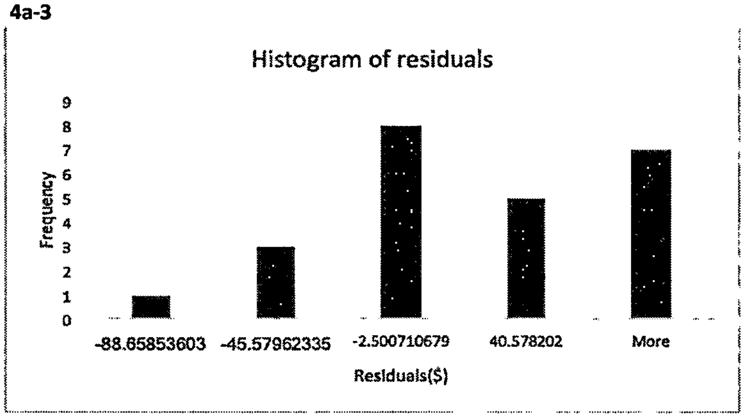
Regression Statistics	
Multiple R	0.37230538
R Square	0.1386113
Adjusted R Sc	0.05657428
Standard Errc	52.7847521
Observations	24

ANOVA

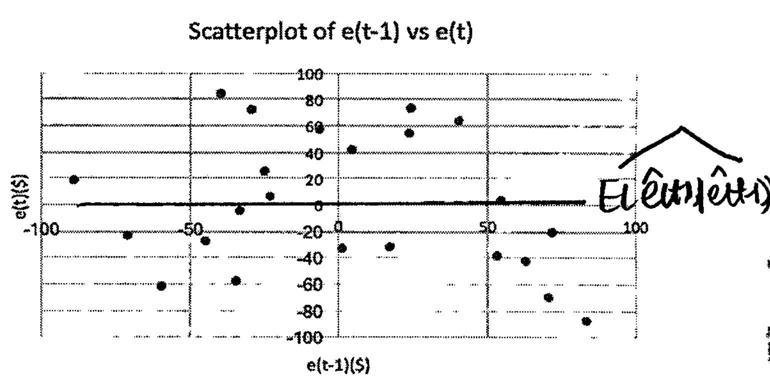
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9415.333861	4707.66693	1.68961889	0.208733089
Residual	21	58510.83114	2786.23005		
Total	23	67926.165			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-266.57917	269.0002479	-0.9909997	0.33296776	-825.9958145	292.837465	-825.99581	292.837465
A(gallons)	-0.5936027	1.192807133	-0.4976518	0.62390048	-3.074180883	1.88697557	-3.0741809	1.88697557
B(lbs)	7.24188094	3.943323703	1.83649162	0.08048276	-0.958709626	15.4424715	-0.9587096	15.4424715

4a-3



4a-4



6

Second Order Model w/o Day
SUMMARY OUTPUT
4b-1

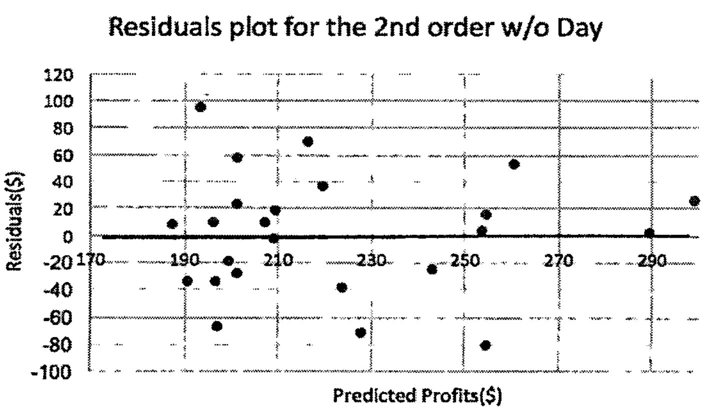
Regression Statistics	
Multiple R	0.58024517
R Square	0.33668446
Adjusted R Sc	0.15243014
Standard Errc	50.0313685
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	22869.684	4573.9368	1.82728124	0.15815313
Residual	18	45056.481	2503.13783		
Total	23	67926.165			

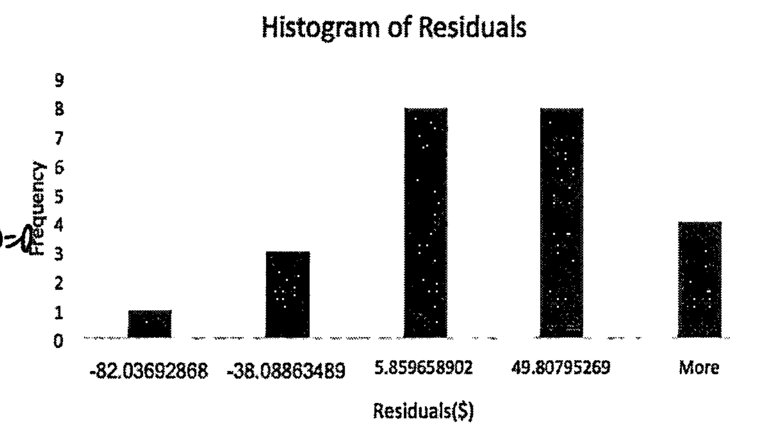
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	18573.6722	9136.284779	2.0329568	0.05707532	-620.9498276	37768.2943	-620.94983	37768.2943
A(gallons)	57.4815995	31.7575214	1.8100153	0.08702179	-9.238477101	124.201676	-9.2384771	124.201676
B(lbs)	-558.45985	265.7418135	-2.1015129	0.04994214	-1116.762687	-0.1570208	-1116.7627	-0.1570208
A^2(gallons^:	-0.0859377	0.122832156	-0.6996354	0.49309839	-0.343998506	0.17212306	-0.3439985	0.17212306
B^2(lbs^2)	4.19088345	1.932319281	2.16883591	0.04373553	0.131231284	8.25053562	0.13123128	8.25053562
A*B(gallons*l	-0.7078293	0.42854674	-1.6516969	0.11593386	-1.608172595	0.19251399	-1.6081726	0.19251399

4b-2



$\widehat{E(\hat{\varepsilon}|\hat{P})}=0$

4b-3



8

Zero Order Model w/o Day
SUMMARY OUTPUT
4c-1

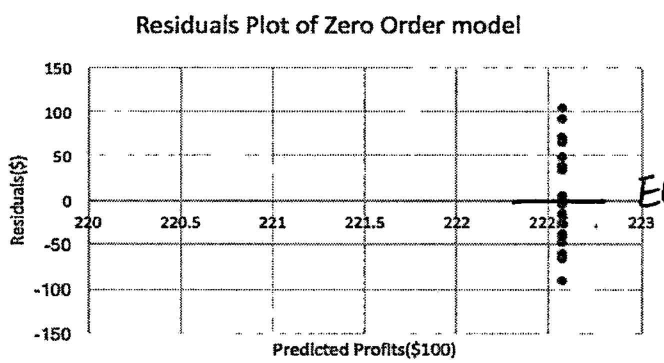
Regression Statistics	
Multiple R	0.21565834
R Square	0.04650852
Adjusted R Sc	0.003168
Standard Errc	54.2582292
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3159.145314	3159.14531	1.07309549	0.311500283
Residual	22	64767.01969	2943.95544		
Total	23	67926.165			

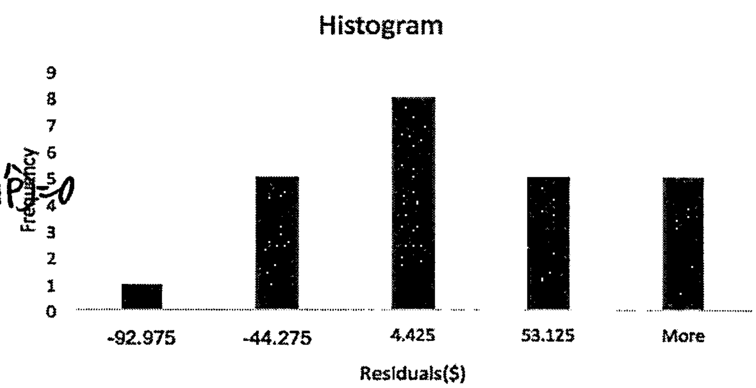
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	222.575	11.07541467	20.096313	1.2051E-15	199.6059958	245.544004	199.605996	245.544004
x0	0	0	65535	#NUM!	0	0	0	0

4c-2

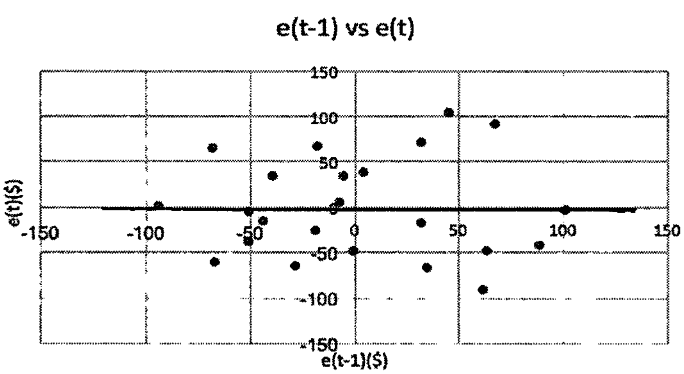


$E(\hat{\varepsilon}|\hat{P})=0$

4c-3



4c-4



$E(\hat{e}(t)|\hat{e}(t-1))$

9

RESIDUAL OUTPUT OF THIRD ORDER MODEL

Use Regression and fit E(e(t))=beta0+beta1*e(t-1) yields r=0 with p=0.73

4c-5

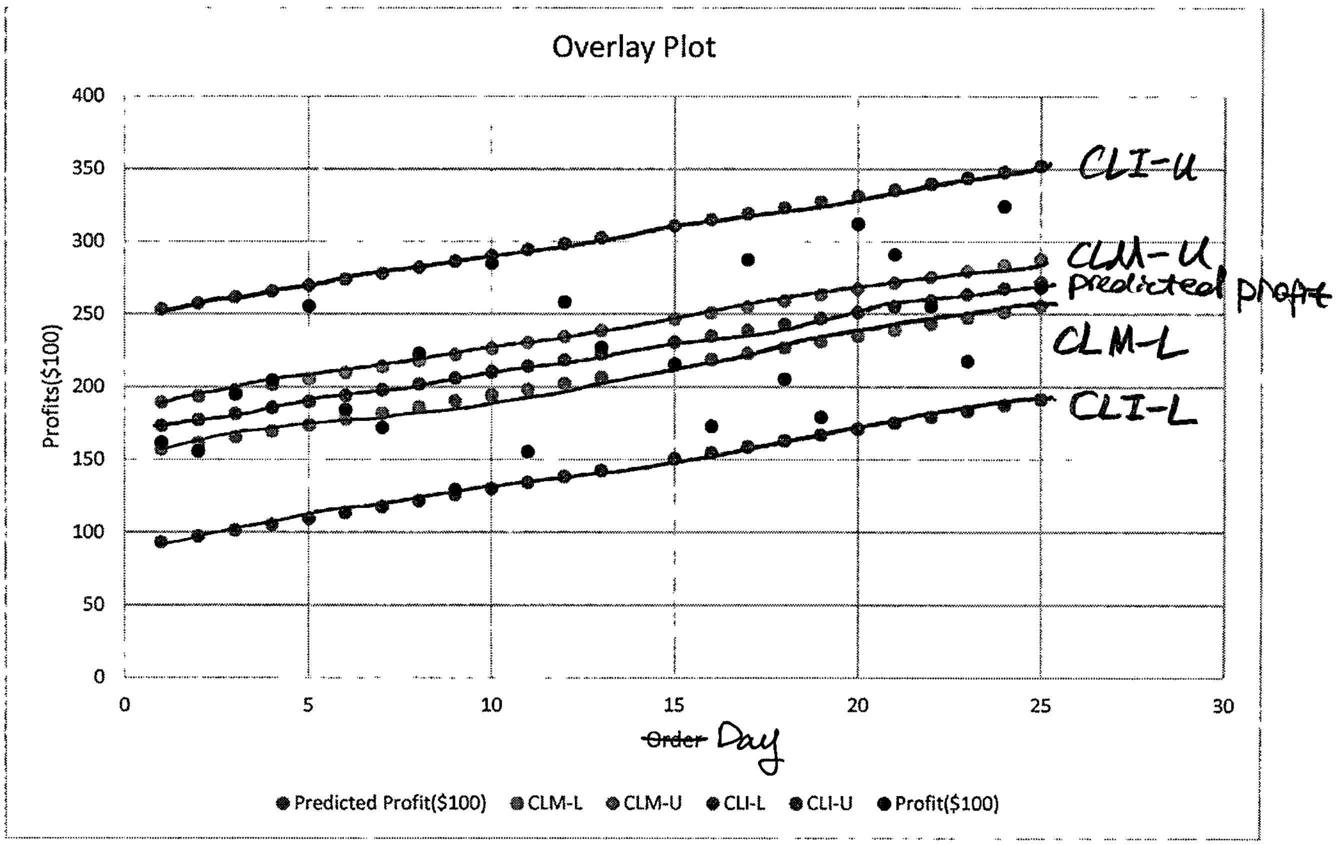
Regression Statistics	
Multiple R	0.07819067
R Square	0.00611378
Adjusted R Sc	-0.0435805
Standard Errc	54.195408
Observations	22

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	361.3501089	361.350109	0.12302779	0.729441042
Residual	20	58742.84489	2937.14224		
Total	21	59104.195			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	5.82475194	11.5553704	0.50407315	0.61971776	-18.27932834	29.9288322	-18.279328	29.9288322
e(t-1)	-0.0762748	0.217459906	-0.3507532	0.72944104	-0.529888166	0.37733866	-0.5298882	0.37733866

Day	Predicted Profit(	CLM-L	CLM-U	CLI-L	CLI-U	Profit($100)
1	173.5003015	157.465314	189.535289	93.3253639	253.675239	162
2	177.6041091	161.569122	193.639097	97.4291715	257.779047	156.1
3	181.7079166	165.672929	197.742904	101.532979	261.882854	195.1
4	185.8117241	169.776737	201.846712	105.636787	265.986662	204.8
5	189.9155317	173.880544	205.950519	109.740594	270.090469	255.8
6	194.0193392	177.984352	210.054327	113.844402	274.194277	184.4
7	198.1231468	182.088159	214.158134	117.948209	278.298084	172.4
8	202.2269543	186.191967	218.261942	122.052017	282.401892	223.2
9	206.3307618	190.295774	222.365749	126.155824	286.505699	129.6
10	210.4345694	194.399582	226.469557	130.259632	290.609507	285.2
11	214.5383769	198.503389	230.573364	134.363439	294.713315	155.6
12	218.6421844	202.607197	234.677172	138.467247	298.817122	258.6
13	222.745992	206.711004	238.78098	142.571054	302.92093	227.3
15	230.9536071	214.91862	246.988595	150.778669	311.128545	215.8
16	235.0574146	219.022427	251.092402	154.882477	315.232352	173
17	239.1612221	223.126235	255.19621	158.986285	319.33616	287.7
18	243.2650297	227.230042	259.300017	163.090092	323.439967	205.6
19	247.3688372	231.33385	263.403825	167.1939	327.543775	179.3
20	251.4726447	235.437657	267.507632	171.297707	331.647582	312.4
21	255.5764523	239.541465	271.61144	175.401515	335.75139	291.3
22	259.6802598	243.645272	275.715247	179.505322	339.855197	255.6
23	263.7840674	247.74908	279.819055	183.60913	343.959005	217.8
24	267.8878749	251.852887	283.922862	187.712937	348.062812	324.4
25	271.9916824	255.956695	288.02667	191.816745	352.16662	268.8



Yes, the data consists with calculated results since the predicted model is inside the 95% CI of both mean & individual.

Programmer Instruction for Computer Assignment 2

1. We were given 25 observations. However, observation 14 has a missing entry, so we are excluding the entire observation. Input the remaining 24 observations consisting of the two variables: one dependent variable, profit ($100), and two independent variables, material A(gallons), material B (lbs), and ~~order~~ Day (referred to below as P, A, B, and D respectively).
2. Get the univariate descriptive statistics including mean, std.dev., min, max, and a histogram for each of these four variables.
3. Get scatterplots of all pairs of these variables (6 plots in total). Display P on the y-axis when Pis included. If there is an D variable, display it on the x-axis when included. Specifically plot (using the notation "y" vs "x"): P vs A, P vs B, P vs D, A vs B, A vs D, and B vs D. Adjust the min and max on the x and y axes to remove blank areas from each of plots.
4. Get the correlation matrix for these four variables.
5. Fit the first order regression model $E(I)=b_0+b_1A+b_2B$ plus the residual plot and the histogram of the residuals.
6. Repeat step 5 for the 2nd order regression model: $E(I)=b_0+b_1A+b_2B+b_3A^2+b_4B^2$.
7. Save this material. Further output maybe requested based on these materials.

Comment :

Looks good except in steps 5 and 6 replace I with P in E()

Revised Programmer Instruction for Computer Assignment 2:

1. We were given 25 observations. However, observation 14 has a missing entry, so we are excluding the entire observation. Input the remaining 24 observations consisting of the two variables: one dependent variable, profit ($100), and two independent variables, material A(gallons), material B (lbs), and order(referred to below as P, A, B, and O respectively).
2. Get the univariate descriptive statistics including mean, std.dev., min, max, and a histogram for each of these four variables.
3. Get scatterplots of all pairs of these variables (6 plots in total). Display P on the y-axis when Pis included. If there is an O variable, display it on the x-axis when included. Specifically plot (using the notation "y" vs "x"): P vs A, P vs B, P vs O, A vs B, A vs O, and B vs O. Adjust the min and max on the x and y axes to remove blank areas from each of plots.
4. Get the correlation matrix for these four variables.
5. Fit the first order regression model $E(P)=b_0+b_1A+b_2B$ plus the residual plot and the histogram of the residuals.
6. Repeat step 5 for the 2nd order regression model: $E(P)=b_0+b_1A+b_2B+b_3A^2+b_4B^2$.
7. Save this material. Further output maybe requested based on these materials.

Listing timing:
Pre-writeup: 10min
Computer work: 100min
Write up: 100min

I have neither given nor received communication about his project with the following sole exceptions:
Professor Szatrowski helped with comments on programming instruction, and feedback for CP2 redo.

Sign: Yifeng Lan

Stat. 463/563 Computer 2 Assignment-Prof Szatrowski YIFENG LAN SEAT NO. 25

Include this page as the last page of your assignment-Do NOT discuss this assignment with anyone other than the instructor. Do - not show this assignment to anyone other than the instructor.

1

Prog. Inst. Due 6/13; Final report due at start of 6/25 class Suppose you had a programmer working for you who knew how to run SAS or Excel programs (but could not give you statistical advice). Write out what computer runs you would request this programmer do after you read this assignment but before you see any computer output. (Note that you anticipate asking for additional computer runs before completing this assignment). Include a copy with your final report.

You are to work on this assignment without the assistance of others without exception. You may use your class notes, class text, homework materials and videotapes without citation. You must cite specifically in detail, any other assistance given or received, (including assistance by me) You must write, sign and date: *I have neither given nor received communication about this project with the following sole exceptions:*

Time: You must keep track of total time spent on this assignment, and also list the three separate components: pre-writeup but not doing computer work, computer work, and writeup. I would hope that you would finish this assignment in about three hours. **You may not spend more than seven hours on this assignment.*** The time spent on each of these components, and total time must be clearly listed.

Scenario: I just received some data last night that my assistant collected on a random sample of profit measures on 25 consecutive days. The data consists of profit (in $100) and two covariates, (an independent variables), material A (in gallons) and material B (in lbs.). Assume I will be away for a week. When I come back, I want your report analyzing how to predict profit on my desk. I will then run to the Board meeting. I hope to have time to read your Executive summary and recommendation (if not, I may just read it as my presentation). If I have more time, I will review your enhanced explanation of your analysis and backup materials. Don't embarrass me!

I will be asked to give a 90% confidence interval for the predicted profit on day 26 when Material A = 35 gallons and Material B = 95 lbs. Include the formula and substitutions for calculating the CI's.

Report: The Executive Summary and Analysis should be at most three pages. It will be followed by a Table of Contents of the Appendix, the first page of which will include citation information and hours spent. It is strongly recommended that you use the format that I have given you in the several reports that we have covered in class. It will start with a brief executive summary which should include a summary of the problem, your recommended model, information about the requested predictions and confidence intervals and any limitations or concerns that you may have that I should use in a brief presentation so I don't overstate the accuracy of the results. Following the Executive summary (which should not be longer than 1/2 a page), you should mention to me what analyses you did and your conclusions with reference to the appendix backup materials. The appendix should have a table of contents and all exhibits should be neatly labeled. Don't show me all your fumblings and ramblings, just show me a straightforward path through what you did that was significant so that I have enough information to decide on whether or not I concur with your analysis. The last two pages will be the programmer's instructions followed by this page.

Use the format of the sample reports given out in class. Do not put a cover on this report. Your name and seat number should appear just above the executive summary on the first page.

On the first page of the appendix, give me the time spent on this project, both total and by the categories mentioned above. Also, give a clear statement of any materials and/or assistance that needs to be cited on this page. Include a table of contents for your appendix which details the computer output of the appendix. Make specific reference to page numbers of this output in your report. **Use α=0.05.** Check that your data has been entered accurately.

THE DATA:

Day	Profit	A	B
1	162	36.2	70.9
2	156.1	36.2	69.5
3	195.1	30.2	70.9
4	204.8	57.2	69.5
5	255.8	44	68.1
6	184.4	61.4	68.1
7	172.4	36.2	68.1
8	223.2	36.2	68.1
9	129.6	48.2	70.9
10	285.2	42.2	68.1
11	155.6	36.2	66.7
12	258.6	36.2	68.1
13	227.3	39.2	68.1
14	238.7	33.2	
15	215.8	62	75.1
16	173	39.2	73.7
17	287.7	51.2	70.9
18	205.6	55.4	73.7
19	179.3	39.2	70.9
20	312.4	30.2	73.7
21	291.3	45.2	75.1
22	255.6	62	76.5
23	217.8	45.2	73.7
24	324.4	42.2	75.1
25	268.8	39.2	73.7

Note that the observation on day 14 has a missing entry. You should exclude this observation from the analysis.

*You need not count time spent before 6/20