Question 1. [11 marks] In 1965, Statistics Canada reported that 15% of all graduating students from high school in Canada went on to post-secondary education (college or university). In 1999, they reported that 51% were continuing. A 2007 study examined a sample of 824 graduating students in Ottawa and found 465 of them were planning to continue on to college or university. In Hearst, however, only 221 of 460 were planning to continue.

Use the .01 level of significance for the following tests of hypotheses.

a) (3) Conduct a hypothesis test that allows you to decide if the 2007 proportion of students continuing on to postsecondary education in the Ottawa population is consistent with or different from the 1999 report on Canadian students.

b) (4) The head of the Hearst School Board maintains that the proportion of their students continuing on is less than in Ottawa by no more than 2%. You think this is ridiculous. Can you show that the rate in Ottawa is more than 2% above Hearst?

c) (4) In 1965, a random sample of 473 students entering university had an average of 73.4 and a standard deviation of 8.1. In 1999, a similar study based on a random sample of 383 found an average of 81.2 with a standard deviation of 6.6. Test a researcher's claim that the average entering grade has increased by more than 6 during the 1965-1999 period.

Question 2. [10 marks]

a) (5) A retiring professor notes that grade inflation in a course taken over by a junior colleague is so egregious that there are only half as many Bs as As and only half as many Cs as Bs, with no other marks. Test at the .05 level of significance whether the grade distribution of the following data fits the professor's description.

A	B	C
33	15	10

b) (5) While not accepting the retiring professor's assertion, the Assistant Dean looks at another course taught by a novice professor, one in mid-career, and one about to retire, and is concerned about differences between sections. Do these tabulated data below show real differences between the mark distributions awarded by the three professors? Test at the .05 level.

Marks	novice	Mid-career	old-fogey
F	11	15	19
B, C, D	30	30	30
A	19	15	11

Question 3. [12 marks] Megalaw, a New York corporate litigation firm, has 325 partners. You are asked to examine their billings for the last quarter of 2006. Ms. Battleaxe, the Managing Partner, thinks that the "average" billings are $200,000 per quarter, in line with the average at other major firms.

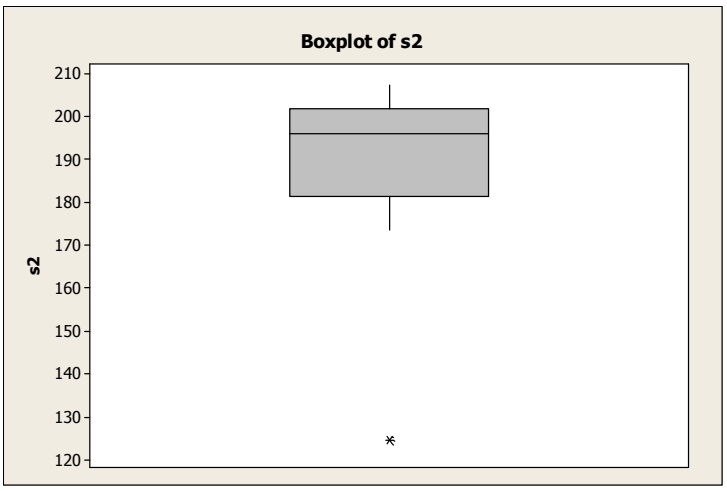
a) (3) To save work, you sample only 14 lawyer's billings at random and find

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
s2	14	0	188.96	5.67	21.20	124.72	181.55	195.93	201.88	207.26

Compute a 95% upper bound (i.e., one-sided "interval") for the true mean of the lawyer's billings from this data.

b) (2) Having taken ADM 2304, you are cautious enough to use Minitab to check your data and also prepare the following printout:

```
Stem-and-Leaf Display: s2

Stem-and-leaf of s2  N  = 14
Leaf Unit = 1.0

 1   12  4
 1   13
 1   14
 1   15
 1   16
 3   17  35
 5   18  36
(5)  19  44799
 4   20  1337
```



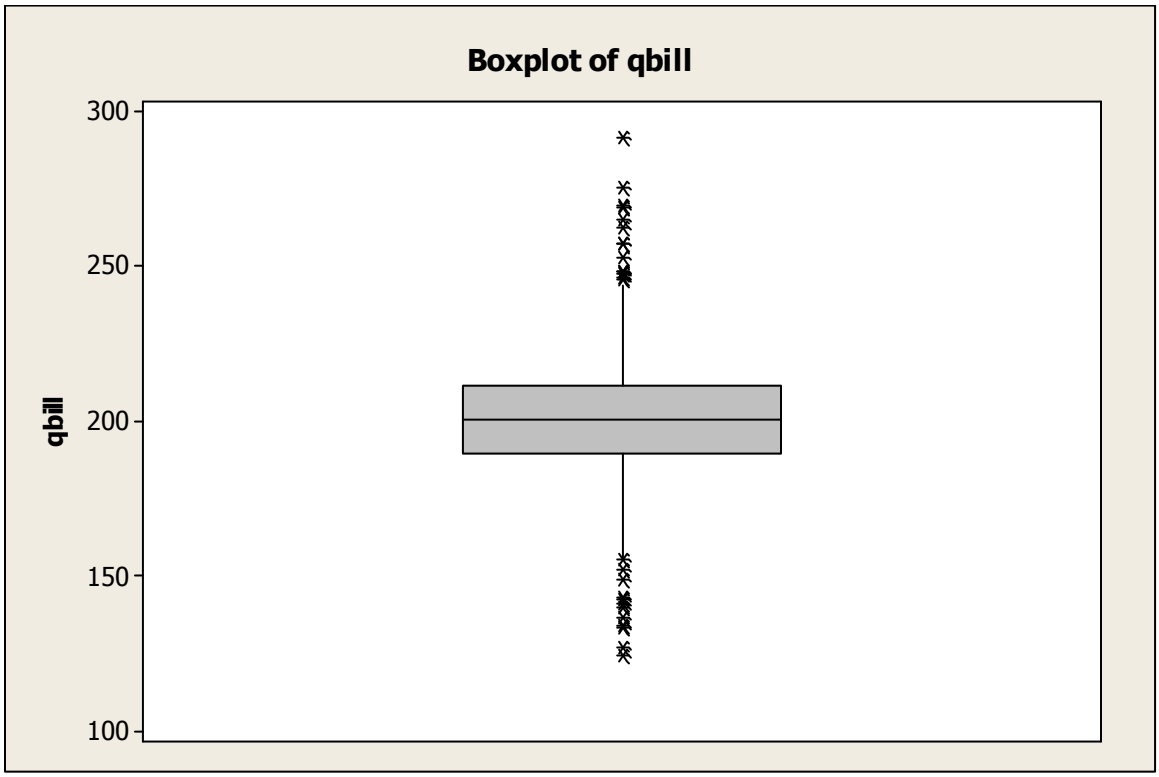
Question 3b (cont'd).

```
Wilcoxon Signed Rank Test: s2

Test of median = 200.0 versus median < 200.0

           N
         for   Wilcoxon         Estimated
     N  Test  Statistic      P     Median
s2  14    14       23.0  0.034      193.1
```

Interpret this output and explain whether this or the results from (a) are more appropriate to examining Ms. Battleaxe's position (no formal test is required).

Because Ms. Battleaxe is the managing partner and you are a junior lawyer, you are told to use the data from all the lawyers. A boxplot of their billings during the last quarter of 2006 in $1000s is given below:

```
Variable    N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3  Maximum
qbill     325   0  200.94     1.34  24.24   124.72  189.41  200.33  211.51   291.49
```



c) (3) Construct a 95% confidence interval for the mean quarterly billings, treating the data as a sample from the presumed large "population" of corporate lawyers.

d) (1) What assumptions, if any, apart from the use of the data as a sample, do you need to make to justify the interval in (c)?

e) (1) What might you conclude from the interval in (c)?

f) (2) Explain why there may be a difference between the results of parts (a) and (c) or between the data in parts (a) and (c) (Note: you do NOT need all the space provided.)

Question 4. [17 marks] You would like to model the weekday soup sales at the local Subway across from the new Demarais Building construction site according to the outdoor temperature. You have collected the daily maximum temperature in degrees Celsius (^{o}C) and daily sales data (in $ spent on soup) for the last 60 weekdays of the winter. The corresponding Minitab output is presented in **Appendix A, Model 1**.

a) (2) Referring only to the corresponding graphs, do you think the underlying assumptions of the model are satisfied? Explain briefly.

b) (1) Calculate the value of R-square for Model 1.

c) (3) Test whether the daily temperature variable is useful for predicting the daily sales? Use the .05 level of significance.

You realize that the sales of soup at this Subway may also be influenced by other factors. To better predict the sales of soup, you decide to include more variables in the model: 1) an indicator or dummy variable for when there is precipitation on the days in question (precip = 1 for yes, = 0 for no); and 2) whether Tim Horton's "Roll Up the Rim" promotion is going on (rollrim = 1 for yes, = 0 for no). The results of the corresponding Minitab analyses are given as Model 2 and Model 3 in Appendix A.

d) (3) Test for the usefulness or significance of **Model 3**. Use the .05 level of significance.

e) (3) Which of the three models should you use and why? Comment briefly on each model.

f) (1) Regardless of your answers above, use model 2 to provide a point prediction for the daily soup sales at the Subway later on a weekday when the temperature reaches 27 ^{o} C and Tim's is not “Rolling up the Rim”. Show all intermediate calculations.

g) (1) Comment on the validity of the prediction of Subway's soup sales when the temperature is 27 ^{o} C and Tim's is not "Rolling up the Rim".

h) (3) Compute a 95% interval for Subway’s soup sales on a weekday when it is 7^{o} C and Tim's is not “Rolling up the Rim”. Note that you are given the “SE Fit” at this point in the output.

Question 5. [11 marks] Westjet would like to conserve fuel on its flights. To study the impact of factors such as time and wind on the consumption of fuel (in litres), it has collected data on fuel consumption for 36 flights from Toronto to Calgary, on three different arrival times (≥ 10 minutes earlier than scheduled, less than 10 minutes from schedule or "on time", and ≥ 10 minutes later than scheduled, coded as -10 for early, 0 for "on time", and +10 for late, respectively), and three different wind conditions (1 for headwind, 0 for no wind, and -1 for tailwind). Each treatment combination has 4 replications or observations. Refer to Appendix B for the full output and, if needed, the studentized range table.

Parts (a), (b), (c), and (d) are based on Model B-1.

a) (1) Based on the graphical output, do the arrival time and wind factors interact with respect to their effect on fuel consumption? Explain briefly.

b) (3) Test for the presence of interaction between the two factors in this model, using a formal hypothesis test.

c) (2) Estimate the difference between the mean fuel consumption for flights which are 10 minutes late with a tailwind and for flights which are 10 minutes early with a headwind. You must show your calculations. (Be careful to identify the appropriate codings for the values of each factor.)

d) (2) Now calculate the appropriate standard error for the difference in (c) and use it to compute a 95% confidence interval estimate to see if the difference is statistically significant. [You looked at the graph of all the fuel consumption means before deciding to test the difference between the two means in (c).]

e) (3) Examine the three ANOVA models, B-1, B-2 or B-3. Which one would you use and why? Comment on each model.

Appendix A

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
temp	60	3.05	1.33	10.27	-10.15	-7.56	2.11	13.99	16.54
rollrim	60	0.50	0.065	0.5042	0.00	0.00	0.50	1.00	1.00
sales	60	173.95	7.87	60.97	79.00	123.25	165.00	240.25	269.00
precip	60	0.45	0.0648	0.5017	0.00	0.00	0.00	1.00	1.00

MODEL 1

Regression Analysis: sales versus temp

The regression equation is
sales = 191 - 5.66 temp

Predictor	Coef	SE Coef	T	P
Constant	191.199	2.503	76.37	0.000
temp	-5.6580	0.2355	-24.03	0.000

S = 18.5778 R-Sq = ???? R-Sq(adj) = 90.7%

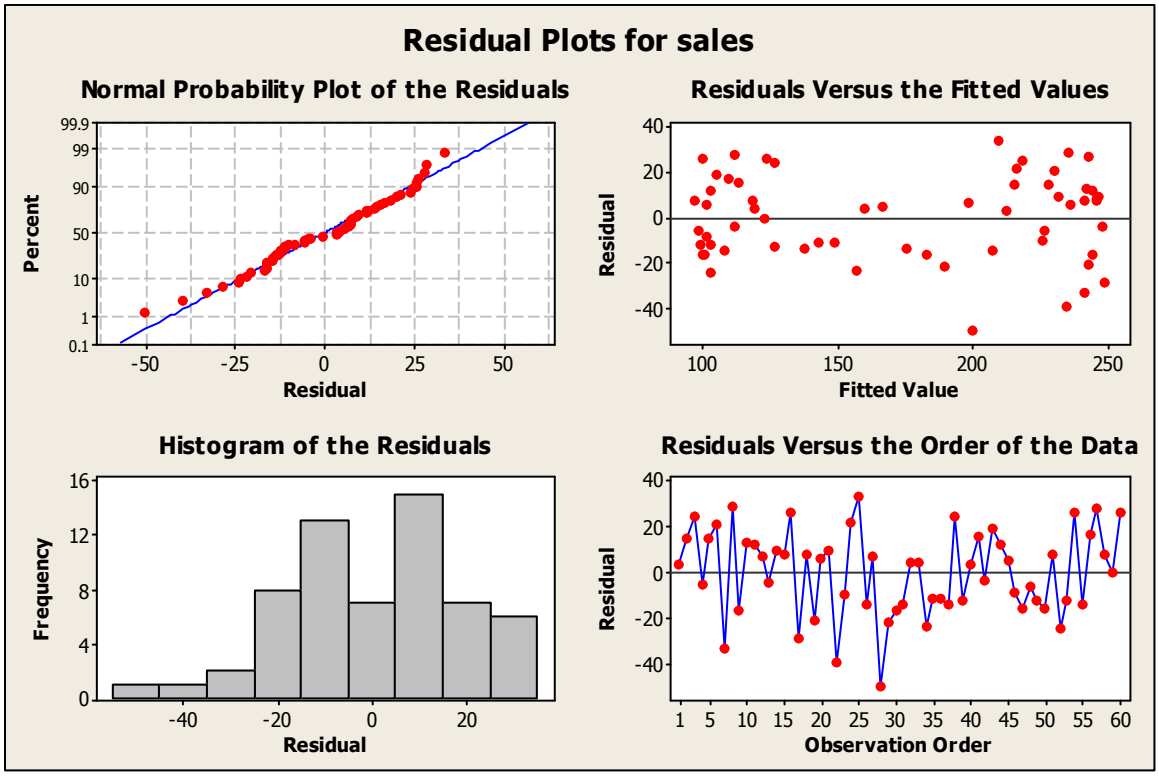
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	------	199301	577.46	0.000
Residual Error	58	20018	345		
Total	59	219319			

Unusual Observations

Obs	temp	sales	Fit	SE Fit	Residual	St Resid
22	-7.7	195.00	234.66	3.48	-39.66	-2.17R
28	-1.6	150.00	200.09	2.63	-50.09	-2.72R

R denotes an observation with a large standardized residual.



MODEL 2

Regression Analysis: sales versus temp, rollrim

The regression equation is
sales = 201 - 4.83 temp - 23.6 rollrim

Predictor	Coef	SE Coef	T	P	VIF
Constant	200.510	3.294	60.87	0.000	
temp	-4.8339	0.3002	-16.10	0.000	2.0
rollrim	-23.647	6.116	-3.87	0.000	2.0

S = 16.6798 R-Sq = 92.8% R-Sq(adj) = 92.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	203460	101730	365.65	0.000
Residual Error	57	15858	278		
Total	59	219319			

Source	DF	Seq SS
temp	1	199301
rollrim	1	4159

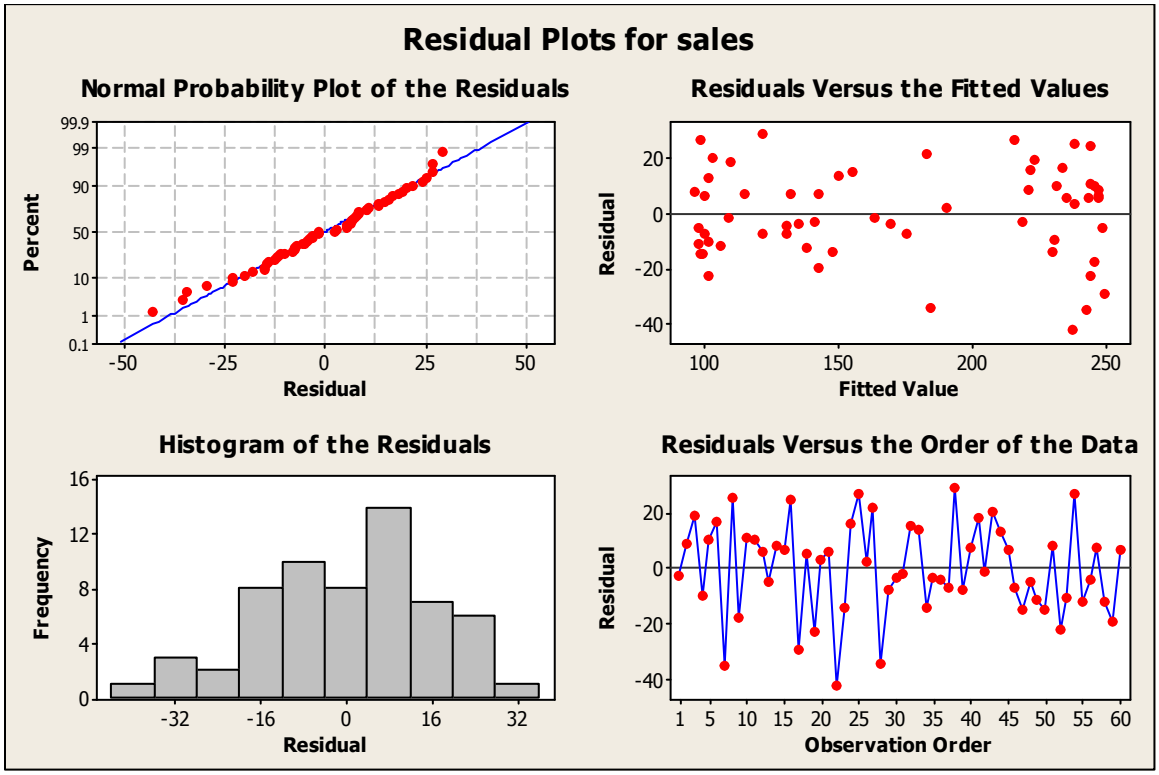
Unusual Observations

Obs	temp	sales	Fit	SE Fit	Residual	St Resid
7	-8.8	208.00	243.27	3.35	-35.27	-2.16R
22	-7.7	195.00	237.64	3.22	-42.64	-2.61R
28	-1.6	150.00	184.46	4.68	-34.46	-2.15R

R denotes an observation with a large standardized residual.

Predicted Values for New Observations using Model 2

New Obs	temp	rollrim	Fit	SE Fit	95% CI	95% PI
1	27.0	0.000000	yyyyy	9.84	(xxxxxx, xxxxxx)	(xxxxxxx, xxxxxx)
2	7.0	0.000000	166.67	4.53	(xxxxxx, xxxxxx)	(xxxxxxx, xxxxxx)



MODEL 3

Regression Analysis: sales versus temp, rollrim, precip

The regression equation is
sales = 199 - 4.87 temp - 23.5 rollrim + 3.81 precip

Predictor	Coef	SE Coef	T	P	VIF
Constant	198.828	3.833	51.88	0.000	
temp	-4.8719	0.3041	-16.02	0.000	2.1
rollrim	-23.477	6.133	-3.83	0.000	2.0
precip	3.807	4.407	0.86	0.391	1.0

S = 16.7171 R-Sq = 92.9% R-Sq(adj) = 92.5%

Analysis of Variance

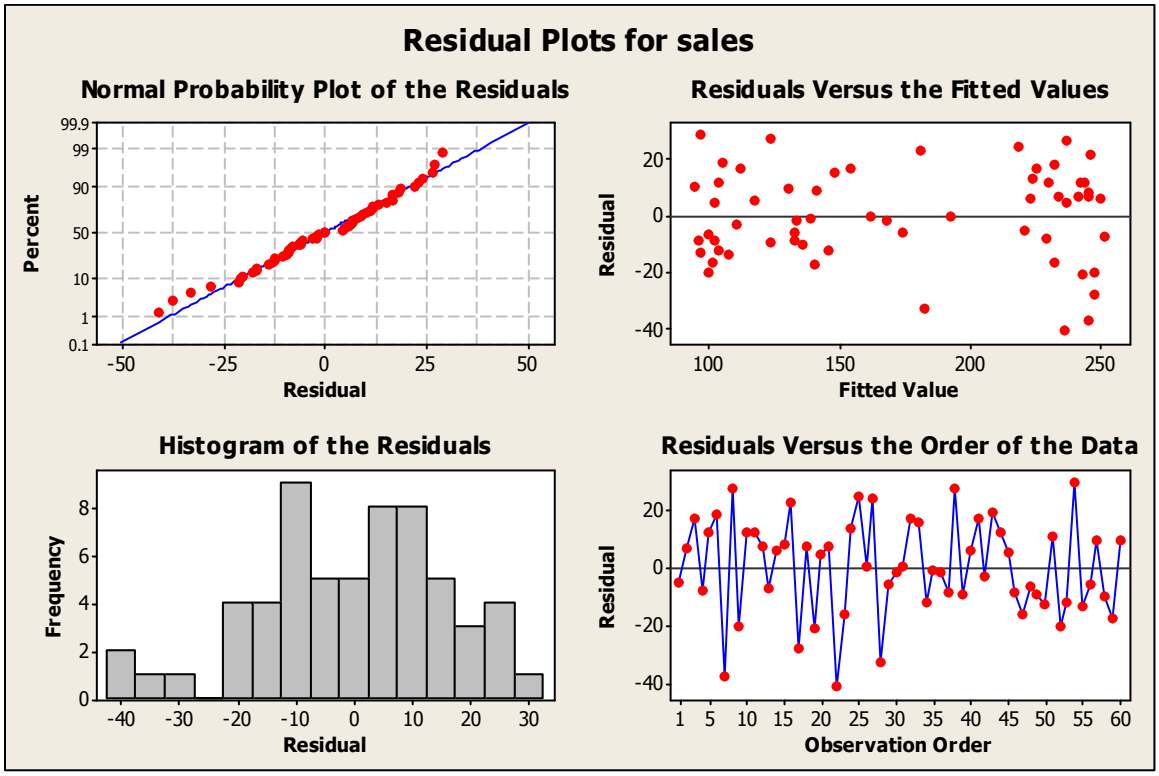
Source	DF	SS	MS	F	P
Regression	3	203669	67890	242.93	0.000
Residual Error	56	15650	279		
Total	59	219319			

Source	DF	Seq SS
temp	1	199301
rollrim	1	4159
precip	1	209

Unusual Observations

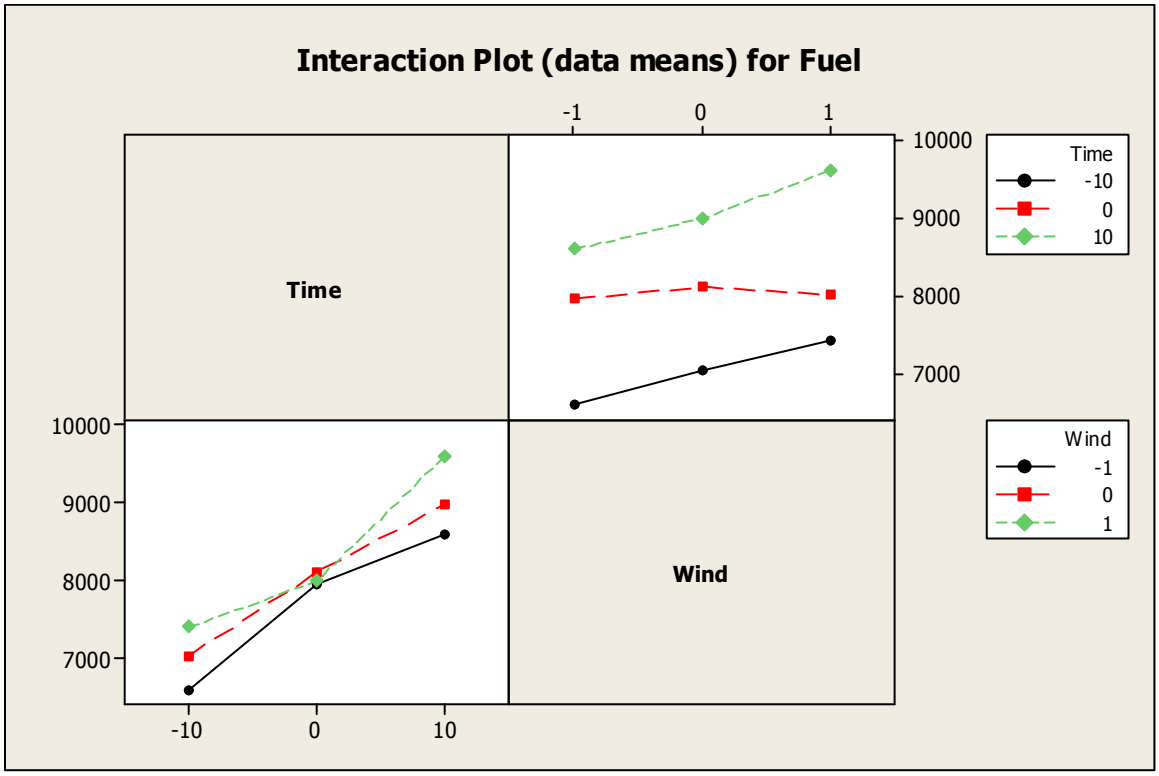
Obs	temp	sales	Fit	SE Fit	Residual	St Resid
7	-8.8	208.00	245.73	4.40	-37.73	-2.34R
22	-7.7	195.00	236.25	3.61	-41.25	-2.53R
28	-1.6	150.00	183.00	4.99	-33.00	-2.07R

R denotes an observation with a large standardized residual.



Appendix B

Row	Fuel	Time	Wind								
1	6336	-10	-1	13	8106	0	-1	25	8590	10	-1
2	6703	-10	-1	14	7883	0	-1	26	8598	10	-1
3	6564	-10	-1	15	7752	0	-1	27	8387	10	-1
4	6743	-10	-1	16	8053	0	-1	28	8748	10	-1
5	6859	-10	0	17	8139	0	0	29	9146	10	0
6	7062	-10	0	18	8254	0	0	30	8892	10	0
7	7173	-10	0	19	8228	0	0	31	8916	10	0
8	7028	-10	0	20	7801	0	0	32	8963	10	0
9	7283	-10	1	21	8055	0	1	33	9262	10	1
10	7590	-10	1	22	8063	0	1	34	9897	10	1
11	7348	-10	1	23	7620	0	1	35	9475	10	1
12	7386	-10	1	24	8227	0	1	36	9755	10	1



Model B-1.
Two-way ANOVA: Fuel versus Time, Wind

Source	DF	SS	MS	F	P
Time	2	25122413	12561206	351.63	0.000
Wind	2	2345945	1172972	32.84	0.000
Interaction	-	1138220	-----	-----	-----
Error	27	964512	35723		
Total	35	29571089			

S = 189.0 R-Sq = 96.74% R-Sq(adj) = 95.77%

Individual 95% CIs For Mean Based on Pooled StDev

Time	Mean	----+---------+---------+---------+----
-10	7006.25	(-*-)
0	8015.08	(-*)
10	9052.42	(-*-)
		----+---------+---------+---------+----
		7200	7800	8400	9000

```
                Individual 95% CIs For Mean Based on
                Pooled StDev
Wind     Mean   ------+---------+---------+---------+---
-1     7705.25  (---*----)
 0     8038.42              (----*---)
 1     8330.08                        (---*----)
               ------+---------+---------+---------+---
                  7750      8000      8250      8500
```

Model B-2.
Two-way ANOVA: Fuel versus Time, Wind

```
Source  DF        SS        MS       F      P
Time     2  25122413  12561206  185.19  0.000
Wind     2   2345945   1172972   17.29  0.000
Error   31   2102731     67830
Total   35  29571089

S = 260.4   R-Sq = 92.89%   R-Sq(adj) = 91.97%


                Individual 95% CIs For Mean Based on
                Pooled StDev
Time     Mean   ------+---------+---------+---------+---
-10    7006.25  (--*-)
  0    8015.08                (--*-)
 10    9052.42                                  (--*-)
               ------+---------+---------+---------+---
                  7200      7800      8400      9000


                Individual 95% CIs For Mean Based on
                Pooled StDev
Wind     Mean   --------+---------+---------+---------+-
-1     7705.25  (-----*-----)
 0     8038.42              (------*-----)
 1     8330.08                         (-----*-----)
               --------+---------+---------+---------+-
                    7750      8000      8250      8500
```

Model B-3.
One-way ANOVA: Fuel versus Time

```
Source  DF        SS        MS      F      P
Time     2  25122413  12561206  93.18  0.000
Error   33   4448676    134808
Total   35  29571089

S = 367.2   R-Sq = 84.96%   R-Sq(adj) = 84.04%


                        Individual 95% CIs For Mean Based on
                        Pooled StDev
Level   N    Mean  StDev  ---+---------+---------+---------+------
-10    12  7006.3  373.9  (--*--)
  0    12  8015.1  205.6            (---*--)
 10    12  9052.4  471.6                          (--*--)
                        ---+---------+---------+---------+------
                        7000      7700      8400      9100

Pooled StDev = 367.2
```