STAT 512

CARBON MONOXIDE FROM FREEWAY

Linear Regression Analysis

May 1, 2015

PURDUE UNIVERSITY

Overview

- Background
- Data Set
- Project Questions
- Best Model
- Questions

PURDUE UNIVERSITY

BACKGROUND

- Hourly carbon monoxide (CO) emissions were recorded on summer weekdays
- Measuring station by EPA @ 25 feet from the San Diego Freeway
- Winds from 145 to 325 degrees transport CO emissions from highway towards measuring station
- Measurements were recorded for each hour of the day 1 to 24

PURDUE UNIVERSITY

Data Set

Response Variable

- **Carbon Monoxide Emission, CO (ppm)**

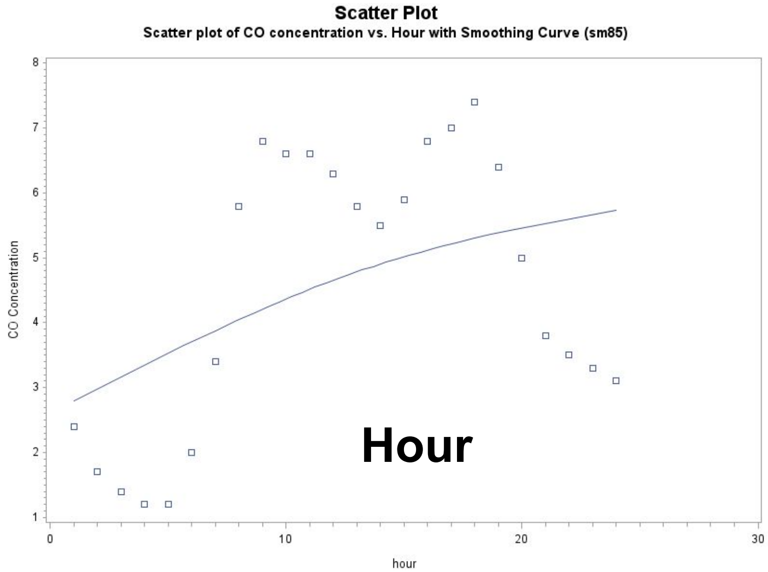
Predictor Variables

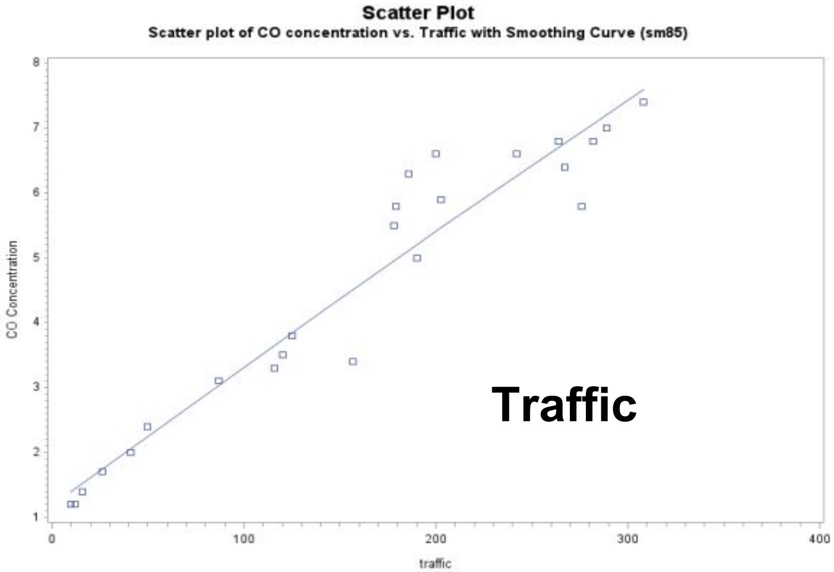
- **Hour**: hour of the day, from midnight to midnight
- **Traffic**: average weekday traffic density (traffic count/traffic speed)
- **Wind**: average perpendicular wind-speed component

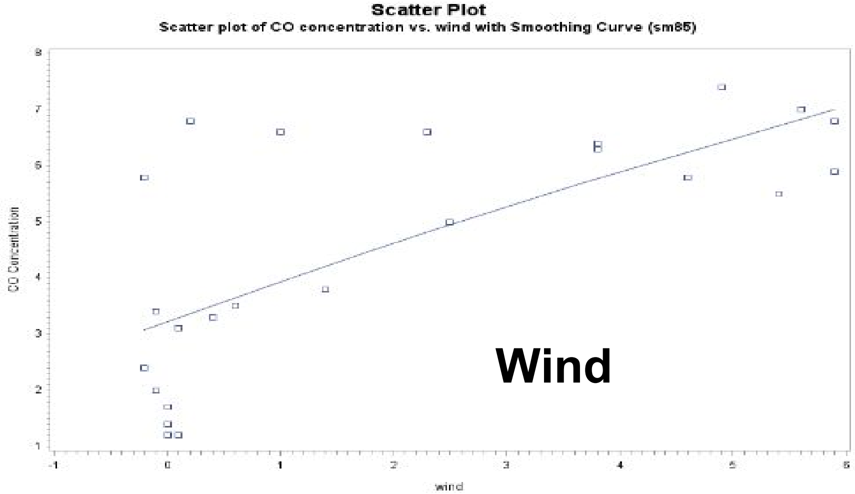
Number of Observations = 24

Question 1

Plot the data



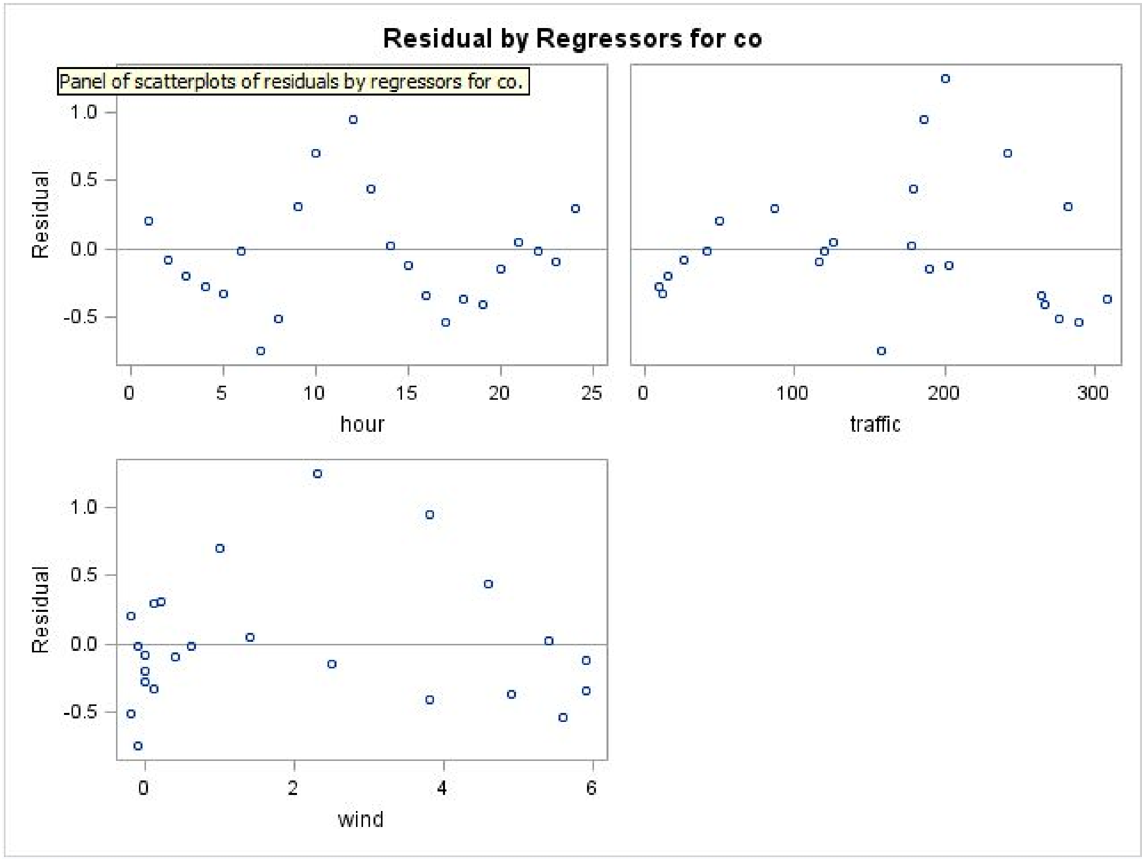




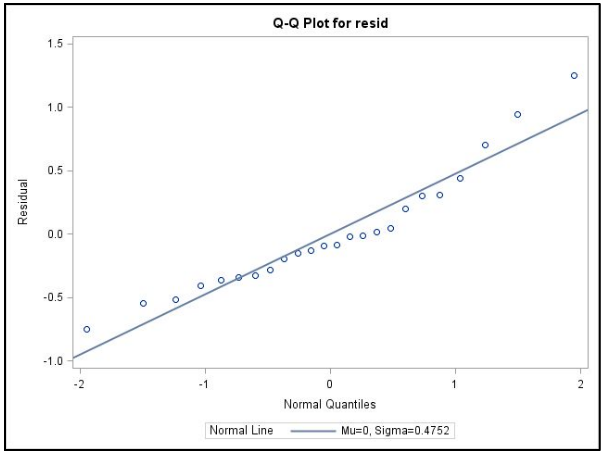
Question 1

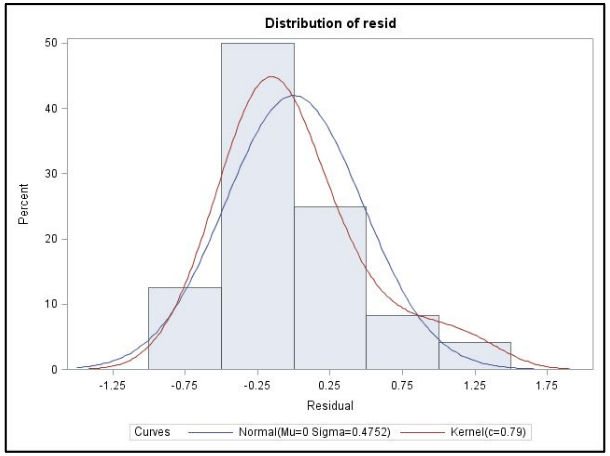
Assumptions: Residuals

Residual Plots:



QQ Plot and Histogram:





PURDUE UNIVERSITY

Data Set

Modification of Model

From the data file page online:

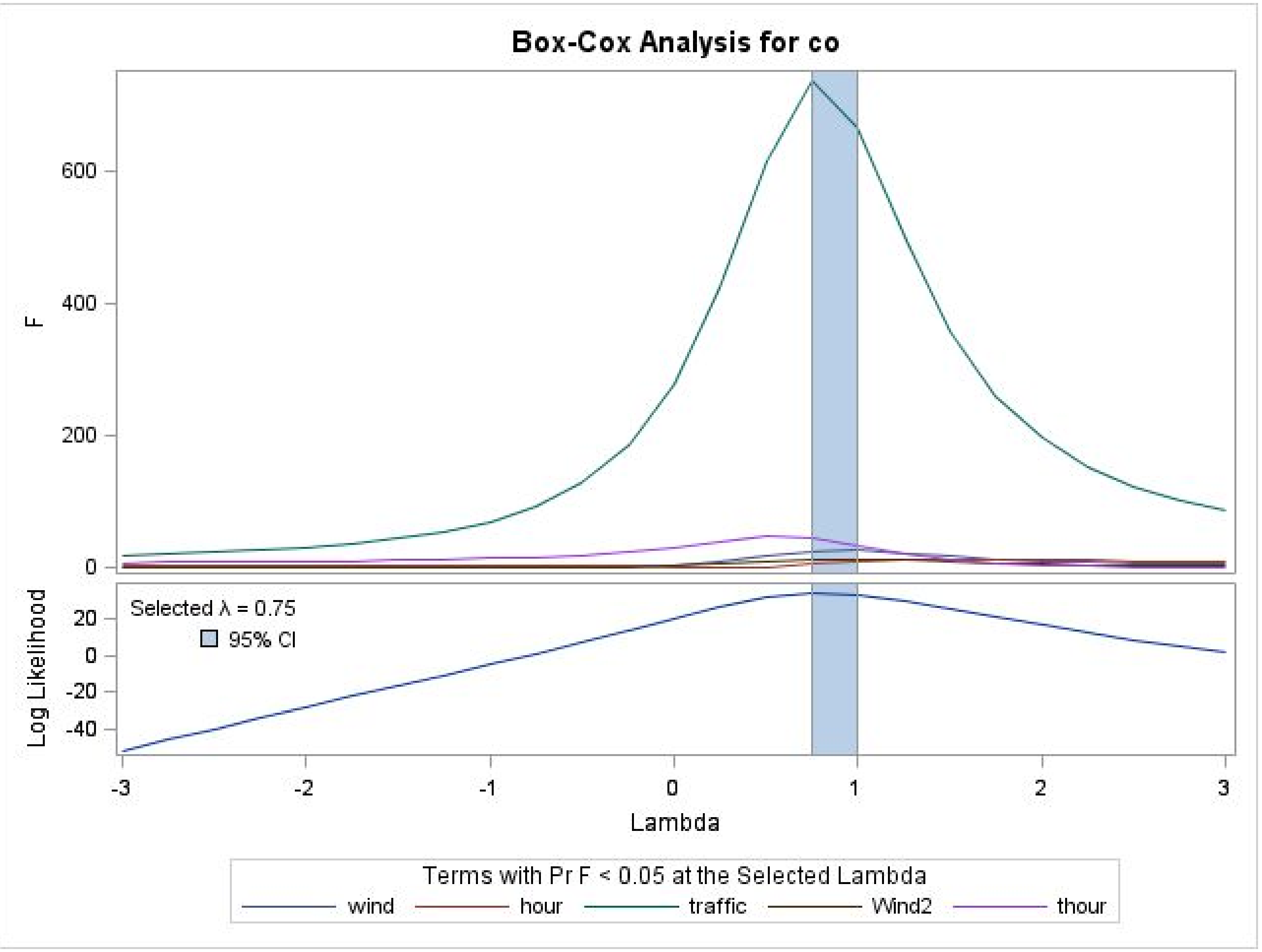
```
Coefficients:
                          Value Std. Error  t value Pr(>|t|)
            (Intercept)  1.7523    0.7793    2.2484   0.0390
                Traffic  0.0179    0.0013   13.7864   0.0000
                   Wind  0.3178    0.3447    0.9220   0.3702
             I(Wind^2) -0.0774    0.0268   -2.8851   0.0108
sin((2 * pi)/24 * Hour) -0.5437    0.9324   -0.5831   0.5680
cos((2 * pi)/24 * Hour) -0.7300    0.6574   -1.1106   0.2832
sin((4 * pi)/24 * Hour)  0.3664    0.4331    0.8460   0.4100
cos((4 * pi)/24 * Hour)  0.3110    0.1424    2.1838   0.0442
```

- **Wind^2: Quadratic term of wind**
- **Thour: cos ((4 * pi)/24 * Hour)**

PURDUE UNIVERSITY

Question 1

Box-cox to look for Transformation



Lamda is suggested to be 0.75 only for hour, so we will keep it to be 1.

Question 2

Piecewise Simple Linear Regression

From the previous problem, it was found that there exists that wind has a quadratic relationship with the response variable. So the explanatory variable "wind" was chosen to model the relationship with the response variable. The point chosen is wind=2.

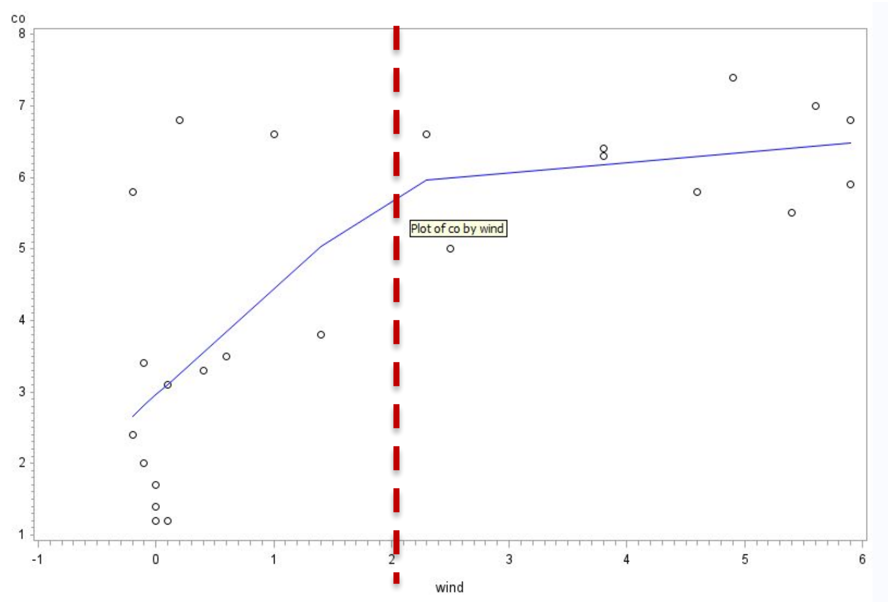
The piecewise regression line is

$$\text{co} = \beta_0 + \beta_1 wind + \beta_2 Xnew(wind - 2) + \epsilon$$

And Xnew=0 when wind<2 when Xnew=1

$$\text{co} = \beta_0 + \beta_1 wind \ \ (wind \leq 2, Xnew = 0)$$

$$\text{co} = (\beta_0 - 2\beta_2) + (\beta_1 + \beta_2) wind \ \ (wind > 2, Xnew = 1)$$



PURDUE UNIVERSITY

Question 2

Piecewise Simple Linear Regression

<u>Same line test</u>: $H_0 : \beta_2 = 0$ Ha: β_2 not equal to 0

Test sameline Results for Dependent Variable co				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	6.33594	2.96	0.1002
Denominator	21	2.14309		

p value= 0.1002 : fail to reject so $\beta_2 = 0$ This is same line. And the plot of wind vs co is attached.

PURDUE UNIVERSITY

Question 3

Extra Sums of Squares

SUM = Traffic + Wind

(a) The prediction of reduced model CO from thour, wind2 and hour is

$$CO=2.78542+0.09220wind2+0.11370thour+0.07275$$

The prediction of full model CO from sum, thour, wind2 and hour is

$$CO=1.24490+0.02151wind2+0.55578thour+0.02017sum-0.01269$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	46.91495	15.63832	5.53	0.0062
Error	20	$\boxed{56.54130}$	2.82706		
Corrected Total	23	103.45625			

Root MSE	1.68139	R-Square	0.4535
Dependent Mean	4.53750	Adj R-Sq	0.3715
Coeff Var	37.05538		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.78542	0.71238	3.91	0.0009
wind2	1	0.09220	0.02955	3.12	0.0054
thour	1	0.11370	0.49727	0.23	0.8215
hour	1	0.07275	0.05316	1.37	0.1863

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	100.70043	25.17511	173.57	<.0001
Error	19	$\boxed{2.75582}$	$\boxed{0.14504}$		
Corrected Total	23	103.45625			

Root MSE	0.38085	R-Square	0.9734
Dependent Mean	4.53750	Adj R-Sq	0.9678
Coeff Var	8.39329		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.24490	0.18010	6.91	<.0001
sum	1	0.02017	0.00105	19.26	<.0001
wind2	1	0.02151	0.00763	2.82	0.0110
thour	1	0.55578	0.11495	4.83	0.0001
hour	1	-0.01269	0.01283	-0.99	0.3350

Question 3

Extra Sums of Squares

$$\underline{SSM(sum|thour, wind2, hour)}$$
$$\underline{= SSE(thour, wind2, hour) - SSE(sum, thour, wind2, hour)}$$
$$\underline{= 56.5413 - 2.75582 = 53.785}$$

The degrees of freedoms for the F test is (1, 19), so the F statistic could be calculated

$$F = \frac{SSM(sum|thour, wind2, hour)/1}{MSE(sum, thour, wind2, hour)} = \frac{53.785}{0.14504} = 370.832$$

The p value for the test is much less than 0.0001 so the null hypothesis is accepted. So variable sum does contain additional useful information that is useful for predicting CO in a linear model which includes thour, hour and wind2. Therefore, the coefficient of the sum variable is not zero.

The test statement output from SAS is attached, which has the same result in part (a).

Question 3

Extra Sums of Squares

(b) The test statement output from SAS is attached, which has the same result in part (a).

Test 1 Results for Dependent Variable co				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	53.78548	370.82	<.0001
Denominator	19	0.14504		

(c)The test statistic is 19.26, which is very close to the root square of F statistic 370.82. The p-value from the two tests are both much smaller than 0.0001.

Question 4

Extra Sums of Squares

$$\sum(Type\ I\ SS) = \boxed{102.27}$$

$$= \quad SSM$$

$$\sum(Type\ II\ SS) = \boxed{49.057}$$

The last predictor thour has same type I SS and type II SS which is calculated by SS(traffic| wind2 thour wind traffic hour)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	102.27626	20.45525	312.03	<.0001
Error	18	1.17999	0.06556		
Corrected Total	23	103.45625			

Root MSE	0.25604	R-Square	0.9886
Dependent Mean	4.53750	Adj R-Sq	0.9854
Coeff Var	5.64270		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	1.33976	0.12262	10.93	<.0001	494.13375	7.82655
hour	1	-0.02864	0.00922	-3.11	0.0061	18.98266	0.63249
traffic	1	0.01907	0.00073929	25.80	<.0001	76.92455	43.63161
wind	1	0.58144	0.11448	5.08	<.0001	2.35433	1.69108
wind2	1	-0.07042	0.01944	-3.62	0.0019	1.77322	0.86028
thour	1	0.46475	0.07948	5.85	<.0001	2.24149	2.24149

PURDUE UNIVERSITY

Question 5

Predicting the Response Variable

No.	Variables	R^2	Adjusted-R^2
1	hour	0.1835	0.1464
2	sum	0.9318	0.9287
3	traffic	0.9267	0.9234
4	wind	0.5037	0.4812
5	wind2	0.3966	0.3692
6	hour sum wind	0.9498	0.9423
7	thour sum wind	0.9775	0.9742
8	thour traffic wind2	0.9710	0.9667
9	thour traffic wind	0.9775	0.9742
10	thour traffic hour wind wind2	0.9886	0.9854
11	sum thour traffic hour wind wind2	0.9886	0.9854

The model with hour, thour, traffic, wind, wind2 explains the largest percent of variation, for both R-square and adjusted R-square.

Question 6

Sameline, parallel, and same intercept test

(a) Give the equation of the fitted regression line using X, Xnew and their interaction.

The traffic is chosen as X. The center of traffic is 159.33.

$$co = \beta_0 + \beta_1 \times X_{new} + \beta_2 \times traffic + \beta_3 \times traffic \times X_{new} + \epsilon$$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.90592	0.60209	6.49	<.0001
traffic	1	0.01016	0.00251	4.05	0.0006
Xnew	1	-2.68776	0.63758	-4.22	0.0004
tXnew	1	0.00774	0.00350	2.21	0.0390

$$co = 3.9059 - 2.688 \times X_{new} + 0.01016 \times traffic + 0.00774 \times traffic \times X_{new}$$

PURDUE UNIVERSITY

Question 6

Sameline, parallel, and same intercept test

(b) Give the fitted regression line for X values< center (use part a).

When X < center, Xnew=1. So the fitted regression line is:

$$\text{co} = 3.9059 - 2.688 + 0.01016 \times traffic + 0.00774 \times traffic$$

$$\boxed{\text{co} = 1.2179 + 0.0179 \times traffic}$$

(c) Give the fitted regression line for X values $\geq$ center(use part a).

When X $\geq$ center, Xnew=0. So the fitted regression line is:

$$\boxed{\text{co} = 3.9059 + 0.01016 \times traffic}$$

Question 6

Sameline, parallel, and same intercept test

(d) Test whether the lines are same or parallel.

$$co = \beta_0 + \beta_1 \times X_{new} + \beta_2 \times traffic + \beta_3 \times traffic \times X_{new} + \epsilon$$

$$\begin{cases} X_{new} = 1, & co = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) traffic + \varepsilon \\ X_{new} = 0, & co = \beta_0 + \beta_2 traffic + \varepsilon \end{cases}$$

Sameline test: $H_0: \beta_1 = \beta_3 = 0$

$H_a: \beta_1 \text{ and } \beta_3 \text{ are not both } 0$

Null hypothesis rejected, so they are not same line.

Test sameline Results for Dependent Variable co				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	2.08573	12.24	$\boxed{0.0003}$
Denominator	20	0.17047		

Question 6

Sameline, parallel, and same intercept test

Parallel test: $H_0: \beta_3 = 0$
$H_a: \beta_3 \neq 0$

Null hypothesis rejected, so they are not parallel.

Test parallel Results for Dependent Variable co				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.83232	4.88	0.0390
Denominator	20	0.17047		

Same intercept test: $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

Null hypothesis rejected, so their intercept is different.

Test sameintercept Results for Dependent Variable co				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	3.02934	17.77	0.0004
Denominator	20	0.17047		

Question 7

Selection by C_p Criterion

Two Criteria:

1. $C_p \leq p$
2. Smallest C_p for a given p

p is the number of regression coefficients including the intercept.

Number in Model	C(p)	R-Square	Parameter Estimates					
			Intercept	Hour	TD	WS	Wind2	thour
5	6.0000	0.9886	1.33976	-0.02864	0.01907	0.58144	-0.07042	0.46475
4	13.6482	0.9825	1.13566	.	0.01861	0.45532	-0.05167	0.43631
4	17.1231	0.9803	1.29190	-0.01826	0.01953	0.18113	.	0.53477
3	19.4665	0.9775	1.15683	.	0.01912	0.16736	.	0.50154
3	29.7438	0.9710	1.15516	.	0.01985	.	0.02396	0.53512
4	29.7965	0.9722	1.24317	-0.01203	0.02019	.	0.02488	0.55856
4	38.1924	0.9669	1.38068	-0.02243	0.01797	0.73701	-0.09807	.
3	42.1898	0.9631	1.21669	.	0.01765	0.62935	-0.08184	.
2	50.3996	0.9567	1.09997	.	0.02157	.	.	0.51803
3	51.8153	0.9570	1.14661	-0.00653	0.02179	.	.	0.53041
2	61.6820	0.9495	1.27446	.	0.01829	0.17475	.	.
3	63.2417	0.9498	1.31897	-0.00569	0.01840	0.17919	.	.
2	78.0379	0.9391	1.27398	.	0.01921	.	0.02226	.
3	80.0059	0.9392	1.26232	0.00150	0.01917	.	0.02216	.
1	95.6397	0.9267	1.21905	.	0.02083	.	.	.
2	97.1555	0.9270	1.17500	0.00581	0.02065	.	.	.
3	668.1842	0.5665	2.81953	0.01871	.	1.44724	-0.15364	.

Best Model: **All Variables**

$\hat{y}$ = 1.33976-0.02864Hour+0.01907Traffic+0.58144Wind-0.07042Wind2+0.46475thour

Question 8

Selection by Stepwise Regression

Stepwise Selection: Step 1

Variable TD Entered: R-Square = 0.9267 and C(p) = 95.6397

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	95.87547	95.87547	278.24	<.0001
Error	22	7.58078	0.34458		
Corrected Total	23	103.45625			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.21905	0.23224	9.49419	27.55	<.0001
TD	0.02083	0.00125	95.87547	278.24	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable thour Entered: R-Square = 0.9567 and C(p) = 50.3996

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	98.97230	49.48615	231.76	<.0001
Error	21	4.48395	0.21352		
Corrected Total	23	103.45625			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.09997	0.18547	7.51021	35.17	<.0001
TD	0.02157	0.00100	98.93563	463.35	<.0001
thour	0.51803	0.13603	3.09683	14.50	0.0010

Bounds on condition number: 1.0399, 4.1595

Stepwise Selection: Step 3

Variable WS Entered: R-Square = 0.9775 and C(p) = 19.4665

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	101.13124	33.71041	289.98	<.0001
Error	20	2.32501	0.11625		
Corrected Total	23	103.45625			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.15683	0.13749	8.23023	70.80	<.0001
TD	0.01912	0.00093326	48.80012	419.78	<.0001
WS	0.16736	0.03883	2.15894	18.57	0.0003
thour	0.50154	0.10044	2.89856	24.93	<.0001

Bounds on condition number: 1.656, 12.909

Stepwise Selection: Step 4

Variable Wind2 Entered: R-Square = 0.9825 and C(p) = 13.6482

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	101.64377	25.41094	266.38	<.0001
Error	19	1.81248	0.09539		
Corrected Total	23	103.45625			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.13566	0.12488	7.88929	82.70	<.0001
TD	0.01861	0.00087360	43.29490	453.86	<.0001
WS	0.45532	0.12912	1.18626	12.44	0.0023
Wind2	-0.05167	0.02229	0.51253	5.37	0.0318
thour	0.43631	0.09524	2.00216	20.99	0.0002

Bounds on condition number: 21.631, 176.61

Stepwise Regression: starting with no variables in the model, testing the addition of each variable using a chosen model comparison criterion, adding the variable that improves the model the most, and repeating this process until none improves the model.

Question 8

Selection by Stepwise Regression

Stepwise Selection: Step 5

Variable Hour Entered: R-Square = 0.9886 and C(p) = 6.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	102.27626	20.45525	312.03	<.0001
Error	18	1.17999	0.06556		
Corrected Total	23	103.45625			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.33976	0.12262	7.82654	119.39	<.0001
Hour	-0.02864	0.00922	0.63249	9.65	0.0061
TD	0.01907	0.00073929	43.63161	665.57	<.0001
WS	0.58144	0.11448	1.69109	25.80	<.0001
Wind2	-0.07042	0.01944	0.86029	13.12	0.0019
thour	0.46475	0.07948	2.24149	34.19	<.0001

Bounds on condition number: 24.743, 254.7

All variables left in the model are significant at the 0.1500 level.

All variables have been entered into the model.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	TD		1	0.9267	0.9267	95.6397	278.24	<.0001
2	thour		2	0.0299	0.9567	50.3996	14.50	0.0010
3	WS		3	0.0209	0.9775	19.4665	18.57	0.0003
4	Wind2		4	0.0050	0.9825	13.6482	5.37	0.0318
5	Hour		5	0.0061	0.9886	6.0000	9.65	0.0061

Best Model: **All Variables**

$$\hat{y} = 1.33976-0.02864Hour+0.01907Traffic+0.58144Wind-0.07042Wind2+0.46475thour$$

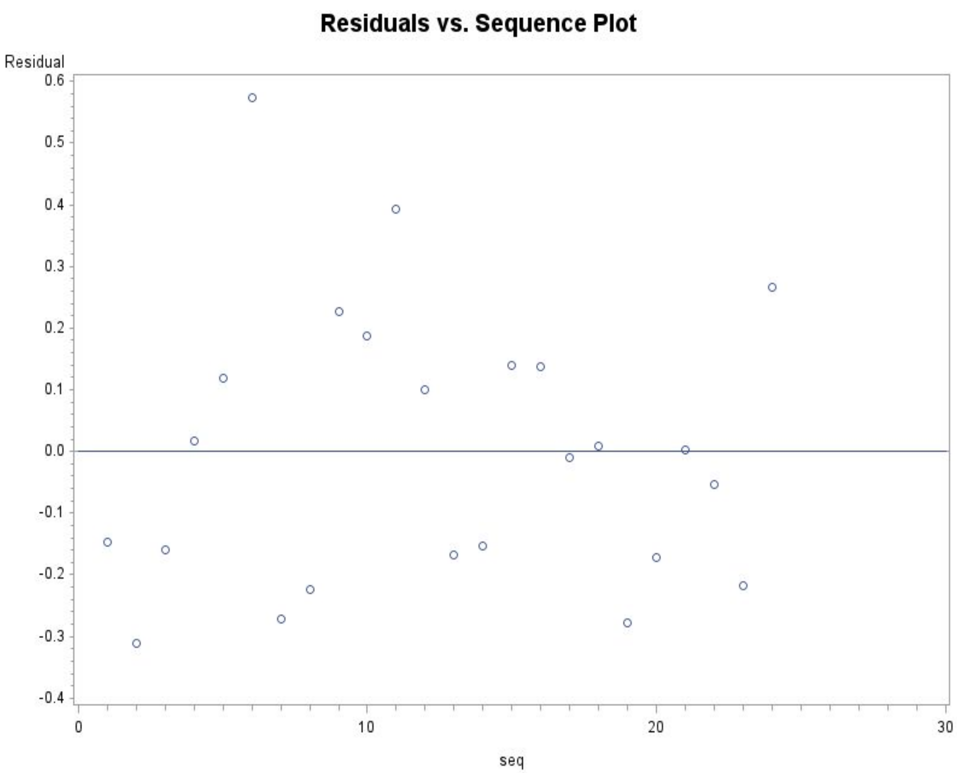
PURDUE UNIVERSITY

Question 9

Selection Check 4 Assumptions: Independence, constant variance, Linearity, Normality.

1. Independence

To check independence, plot the residuals against the order of observation (sequence). A pattern that is not random suggests lack of independence.

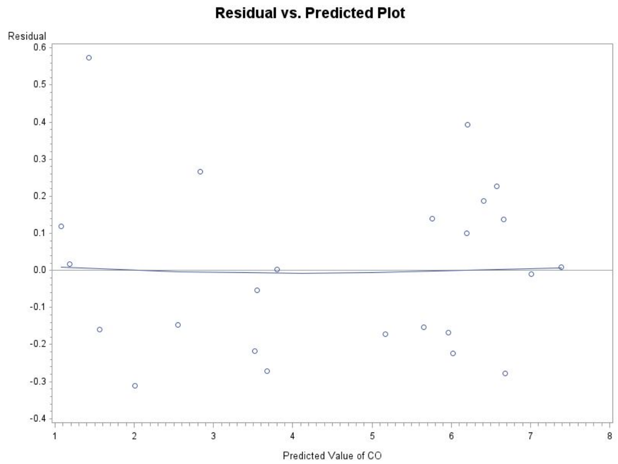


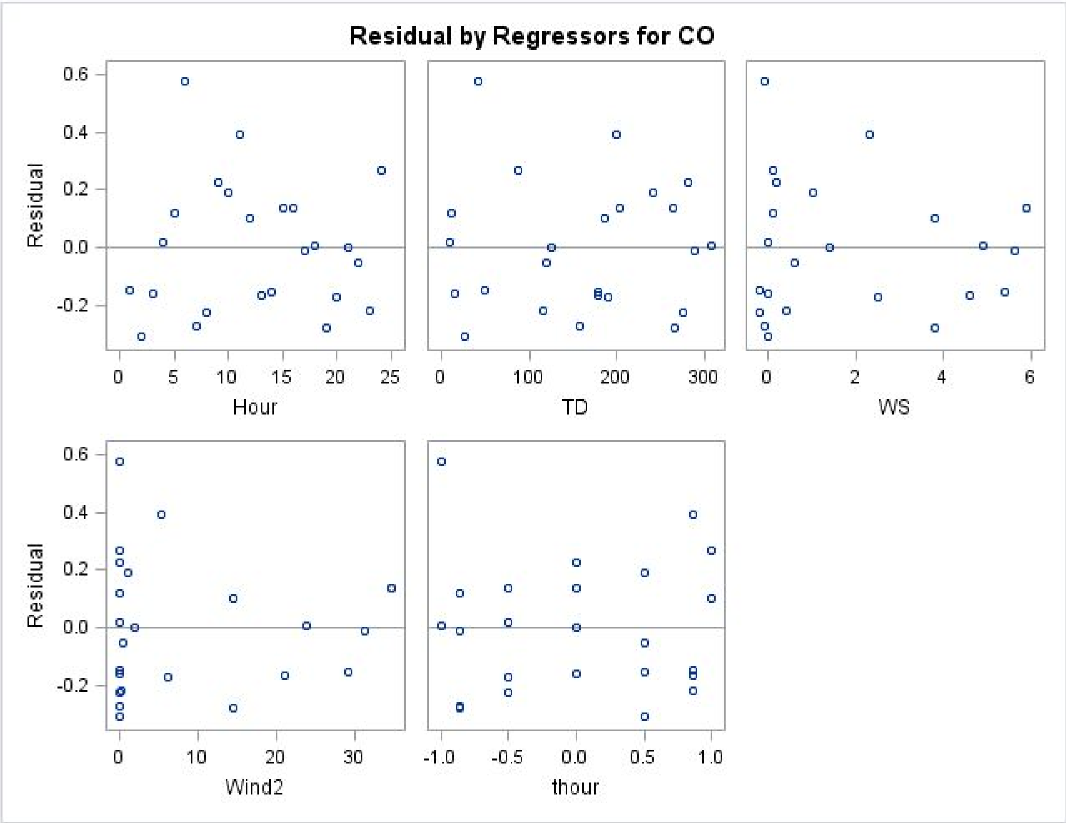
Question 9

Selection Check 4 Assumptions: Independence, constant variance, Linearity, Normality.

2. Constant Variance and Linearity

To check for the constant variance and the linearity, plot residuals against fitted values.





PURDUE UNIVERSITY

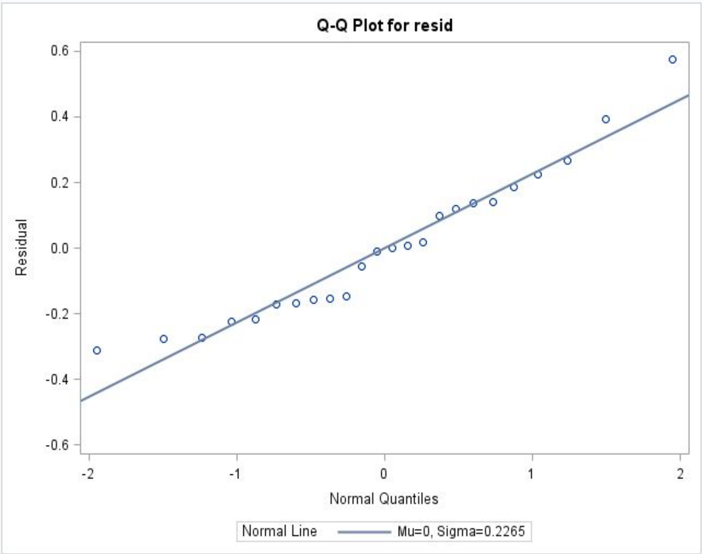
Question 9

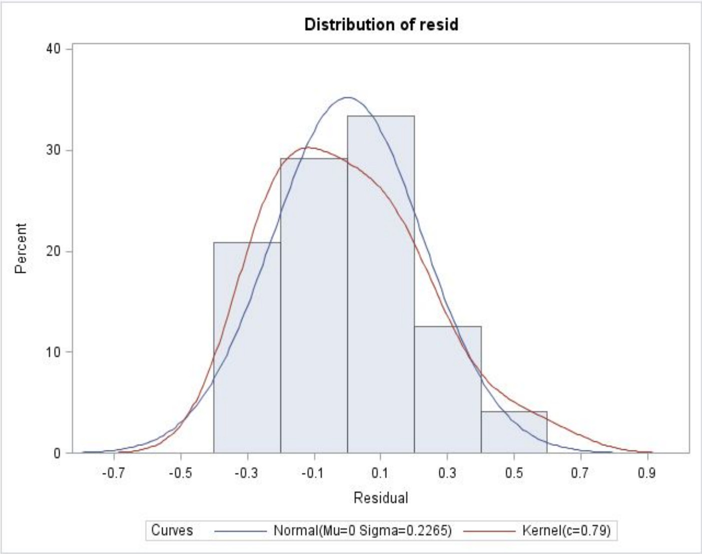
Selection Check 4 Assumptions: Independence, constant variance, Linearity, Normality.

3. Normality

The histogram shows that there is a negligibly small skewness on right. The QQ-plot shows that the residuals are approximately normal.

Moments			
N	24	Sum Weights	24
Mean	0	Sum Observations	0
Std Deviation	0.22650391	Variance	0.05130402
Skewness	0.73709818	Kurtosis	0.23388732
Uncorrected SS	1.17999253	Corrected SS	1.17999253
Coeff Variation	.	Std Error Mean	0.04623492





PURDUE UNIVERSITY

Question 10

Influential or Outlier test

- Studentized Residuals
- Studentized Deleted Residuals
- Cook’s D

Multicollinearity

- Tolerance and VIF

Usefulness

- Partial residual plots

The REG Procedure
Model: MODEL1
Dependent Variable: co

Output Statistics

Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS Intercept	hour	traffic	wind	wind2	thour
2.4000	2.5481	0.1302	-0.1481	0.220	-0.672	| *| |	0.026	-0.6613	0.2584	1.6315	-0.3904	-0.2786	0.1904	-0.0077	0.1047	-0.1069	-0.2272
1.7000	2.0108	0.1106	-0.3108	0.231	-1.346	| **| |	0.069	-1.3790	0.1865	0.9169	-0.6603	-0.6062	0.3327	0.1877	0.0148	-0.0328	-0.2122
1.4000	1.5590	0.1040	-0.1590	0.234	-0.680	| *| |	0.015	-0.6692	0.1651	1.4439	-0.2976	-0.2926	0.1156	0.1455	-0.0163	0.0074	0.0205
1.2000	1.1836	0.1139	0.0164	0.229	0.0717	| | |	0.000	0.0696	0.1979	1.7537	0.0346	0.0309	-0.0078	-0.0204	0.0039	-0.0031	-0.0149
1.2000	1.0804	0.1289	0.1196	0.221	0.541	| |* |	0.017	0.5297	0.2534	1.7107	0.3085	0.2381	-0.0418	-0.1880	0.0653	-0.0598	-0.1951
2.0000	1.4263	0.1214	0.5737	0.225	2.545	| |***** |	0.313	3.0912	0.2249	0.1251	1.6652	1.1189	-0.0441	-0.7519	-0.0384	0.0151	-1.1297
3.4000	3.6724	0.1059	-0.2724	0.233	-1.168	| **| |	0.047	-1.1811	0.1709	1.0589	-0.5363	-0.1829	0.0631	-0.1611	0.1728	-0.1080	0.2576
5.8000	6.0233	0.1582	-0.2233	0.201	-1.109	| **| |	0.126	-1.1164	0.3816	1.4905	-0.8770	0.0739	0.1111	-0.7146	0.4567	-0.3218	-0.0139
6.8000	6.5740	0.1447	0.2260	0.211	1.070	| |** |	0.090	1.0747	0.3196	1.3960	0.7365	-0.0789	-0.1655	0.6513	-0.2565	0.1381	0.1612

Studentized Residual and Studentized-deleted Residual

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D	RStudent
6	2.0000	1.4263	0.1214	0.5737	0.225	2.545	| |***** |	0.313	3.0912

Observation 6 has the largest studentized residual, Cook's D and studentized-deleted residual (Rstudent)

Using Bonferroni correction, $t\text{-}_{n-p-1}\ \text{-}(1\text{-}alpha/2n)=$
$t_{17}(.99896)= 3.627$

Thus, the largest studentized residual of 2.545 is < 3 and studentized-deleted residual of 3.0912<3.627
Thus, observation 6 not an influential point or outlier.

COOK'S D

Obs	Dependent Variable	Predicted Value	Error Mean Predict	Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D	RStudent
6	2.0000	1.4263	0.1214	0.5737	0.225	2.545	| |***** |	0.313	3.0912

$F_{p,n-p} = F_{6,18}(0.5) = 0.93$

Cook's D value of 0.313< 50th percentile value of F which is 0.93

In combination with the previous 2 tests, we can say that observation 6 is not an outlier or an influential point and the dataset doesn't have an outlier or an influential point.

Variance Inflation Factor (VIF) and Tolerance

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1.33976	0.12262	10.93	<.0001	.	0
Hour	1	-0.02864	0.00922	-3.11	0.0061	0.67066	1.49107
Traffic	1	0.01907	0.00073929	25.80	<.0001	0.54266	1.84278
Wind	1	0.58144	0.11448	5.08	<.0001	0.04041	24.74335
wind2	1	-0.07042	0.01944	-3.62	0.0019	0.04607	21.70568
transhour	1	0.46475	0.07948	5.85	<.0001	0.86481	1.15633

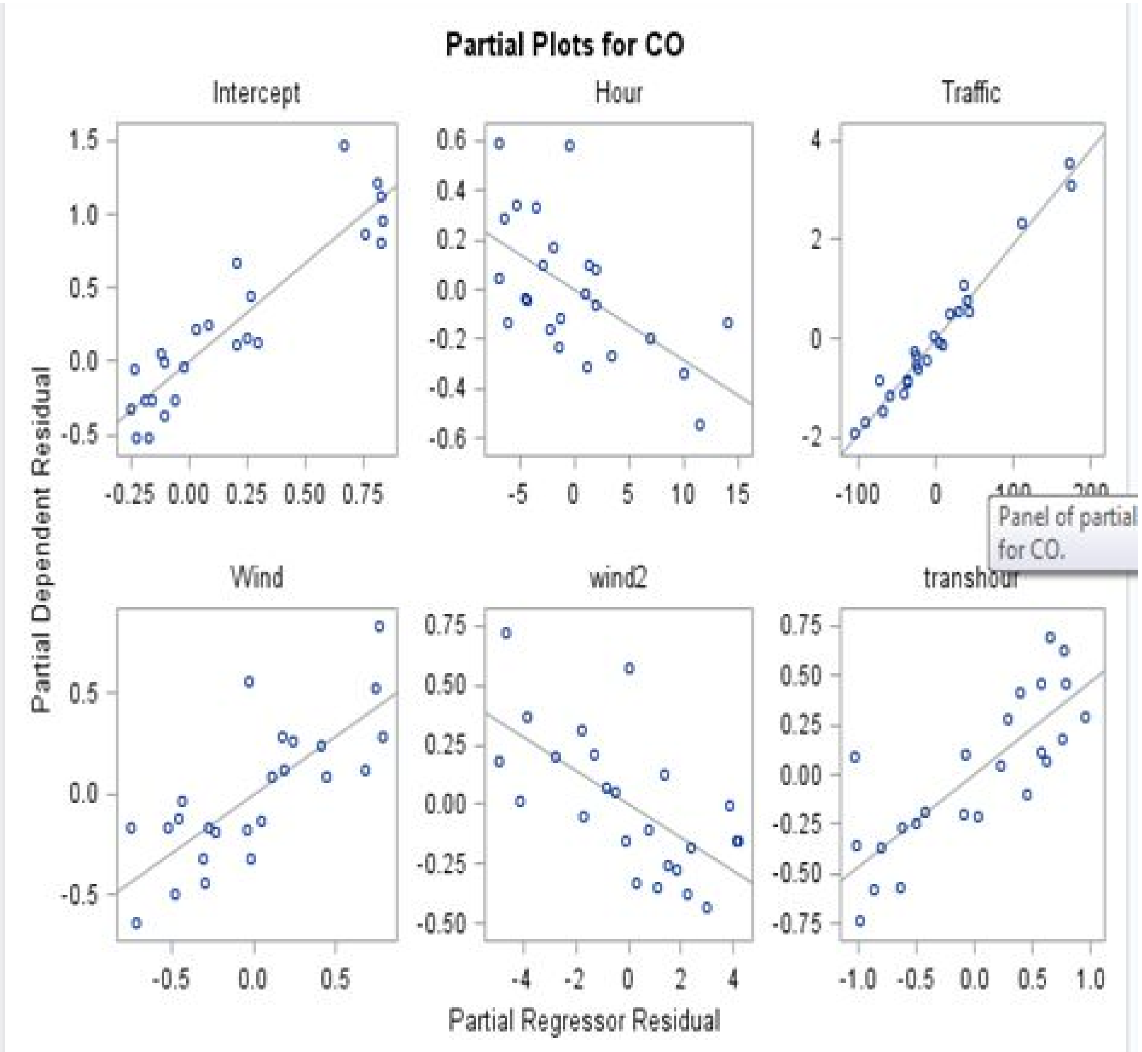
Variables Wind and Wind2 have tolerance < 0.1 and consequently VIF > 10 which puts them at a risk of multicollinearity.
But, all the variables have significant t values.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	102.27625	20.45525	312.03	<.0001
Error	18	1.18000	0.06556		
Corrected Total	23	103.45625			

While all the variables have significant t values, the whole model F value is significant as well.

Even though the VIF and Tolerance tests may show that variables Wind and Wind2 are at a risk of mutlicollinearity, we can conclude that multicollinearity isn't an issue in this model.

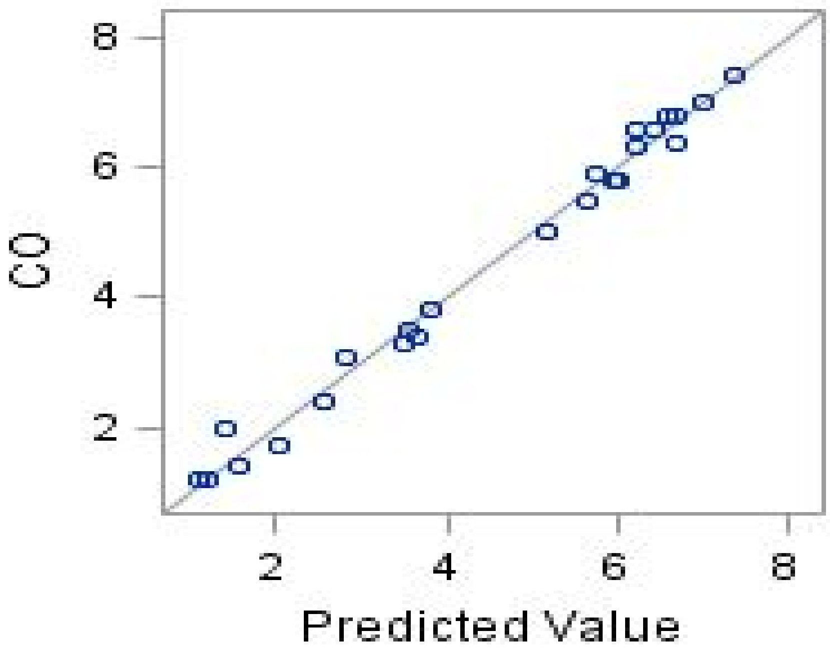
Partial Residual Plots



The partial plots shows that a linear pattern exists for all 5 predictors. And, thus all the variables are useful in the regression model.

"Best" Model Regression Equation

- CO=1.3398 – 0.0286*Hour + 0.0191*Traffic + 0.5814*Wind – 0.0704*Wind2 + 0.4647*transhour.



Question 11

Final Regression Model

Regression Equation:

CO=1.33976-0.0286 X1 + 0.46475 X2 + 0.01907 X3 + 0.5814 X4 - 0.07042 X5

X_1 = Hour, X_2 = Thour, X_3 = Traffic, X_4 = Wind, X_3 = Wind^2

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	90% Confidence Limits	
Intercept	1	1.33976	0.12262	10.93	<.0001	1.12714	1.55239
hour	1	-0.02864	0.00922	-3.11	0.0061	-0.04462	-0.01265
transhour	1	0.46475	0.07948	5.85	<.0001	0.32693	0.60257
traffic	1	0.01907	0.00073929	25.80	<.0001	0.01779	0.02035
wind	1	0.58144	0.11448	5.08	<.0001	0.38293	0.77995
wind2	1	-0.07042	0.01944	-3.62	0.0019	-0.10413	-0.03671

Question 11

90% CL for Mean of Response and individual observations

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	90% CL Mean		90% CL Predict	
1	2.4000	2.5481	0.1302	2.3224	2.7738	2.0501	3.0462
2	1.7000	2.0108	0.1106	1.8190	2.2025	1.5271	2.4944
3	1.4000	1.5590	0.1040	1.3786	1.7394	1.0798	2.0382
4	1.2000	1.1836	0.1139	0.9861	1.3811	0.6976	1.6695
5	1.2000	1.0804	0.1289	0.8569	1.3039	0.5833	1.5775
6	2.0000	1.4263	0.1214	1.2158	1.6369	0.9349	1.9177
7	3.4000	3.6724	0.1059	3.4888	3.8560	3.1919	4.1528
8	5.8000	6.0233	0.1582	5.7490	6.2975	5.5014	6.5451
9	6.8000	6.5740	0.1447	6.3230	6.8250	6.0640	7.0840
10	6.6000	6.4124	0.1187	6.2066	6.6181	5.9230	6.9017
11	6.6000	6.2066	0.1335	5.9751	6.4381	5.7059	6.7073
12	6.3000	6.2010	0.1326	5.9711	6.4309	5.7010	6.7010
13	5.8000	5.9685	0.1169	5.7657	6.1713	5.4804	6.4566
14	5.5000	5.6525	0.1174	5.4488	5.8561	5.1640	6.1409
15	5.9000	5.7611	0.1378	5.5222	6.0000	5.2569	6.2653
16	6.8000	6.6635	0.1353	6.4289	6.8982	6.1613	7.1657
17	7.0000	7.0101	0.1234	6.7962	7.2240	6.5173	7.5030
18	7.4000	7.3922	0.1178	7.1879	7.5965	6.9035	7.8809
19	6.4000	6.6782	0.1306	6.4517	6.9046	6.1798	7.1766
20	5.0000	5.1719	0.1326	4.9420	5.4019	4.6719	5.6719
21	3.8000	3.7985	0.1151	3.5988	3.9981	3.3117	4.2853
22	3.5000	3.5544	0.1171	3.3514	3.7574	3.0662	4.0425
23	3.3000	3.5173	0.1386	3.2770	3.7577	3.0125	4.0222
24	3.1000	2.8340	0.1664	2.5454	3.1225	2.3045	3.3635

This concludes our presentation.

Questions?