

Setting Up Your Computer and Folders to run SAS University Edition through Oracle VM Virtual Box

NOTE: Oracle VM Virtual Box and SAS University Edition can be downloaded for FREE at http://www.sas.com/en_us/software/university-edition/download-software.html

1. Create a directory on your flash or portable drive called "SASUniversityEdition"
2. Create a subdirectory under "SASUniversityEdition" called "myfolders"
3. You should now have something that looks like the following (let's assume your flash drive is "E" – be sure to substitute the correct drive letter as needed depending which computer you are currently using)

E:\SASUniversityEdition\myfolders

4. Run Oracle VM Virtual Box
 - a. *[If you are operating on your own computer, install Oracle VM Virtual Box first.]*
5. If you do not see the SAS University Edition Application, you will need to load it from the LRCLAB ("L") drive on the School of Nursing Network Drives.
 - a. Click on File, Import Appliance
 - b. Go to "L:\download_SASUniversityEd_07092014\" and choose the OVA file "unvbasicvapp_94110_ova_en_sp0_1.ova"
 - c. *OR LOAD THE OVA FILE FROM WHERE YOU HAVE IT ON YOUR OWN COMPUTER (if you are not logged into a School of Nursing computer, e.g. in 117 or the graduate student room).*
6. Once the SAS University Edition Application is loaded, you need to tell it where to find your "myfolders" directory
 - a. Click on Shared Folders and when the windows pops up, click on the little folder icon at the right with a plus + on it
 - b. Under folder path, click other and use the browser to locate the "myfolders" directory on your flash drive
 - c. The folder name will default to "myfolders"
 - d. Choose Auto-Mount to make sure this is turned on
 - e. It will default to Full (read/write) access
 - f. Click OK
7. Double click on the SAS University Edition to get the application running
8. Let it go through its set-up.
9. Once it says it is ready, do not close the window, just minimize it.
10. Open the browser (FIREFOX) and type in <http://localhost:10080>
 - a. *NOTE: I (Melinda Higgins, Ph.D.) have not gotten SAS University Edition to work under Internet Explorer – most likely due to security and plug-in issues that must be configured first. But it appears to just work under Mozilla Firefox. I have not yet investigated the OS X operating system and the associated browsers on a Mac.*
11. This will load the SAS University Application through the Browser window – click on Start SAS University
12. This will open another window/tab in the browser with SAS Studio loaded.
13. From here you will notice that under Folders/Myfolders you will be able to see what is on your flash drive, but you still have to create a SAS LIBRARY mapped to your drive to be able to read and write back to your flash drive.
14. SEE the example SAS program steps below.

A few examples to (A) see what you can do with SAS Studio running through SAS University Edition (via Oracle VM Virtual Box) and (B) so you can check to see if your set-up is working correctly.

1. Be sure you have a copy of the AgeOrionsData.sas7bdat SAS dataset
2. Open the mkh1.sas SAS program (shown below with each section described including the associated output files)

Setting Up Your LIBRARY

```
*****;  
* create a library to your "myfolders" directory;  
*****;  
  
libname mydata '/folders/myfolders';
```

Copy the dataset into the WORK library

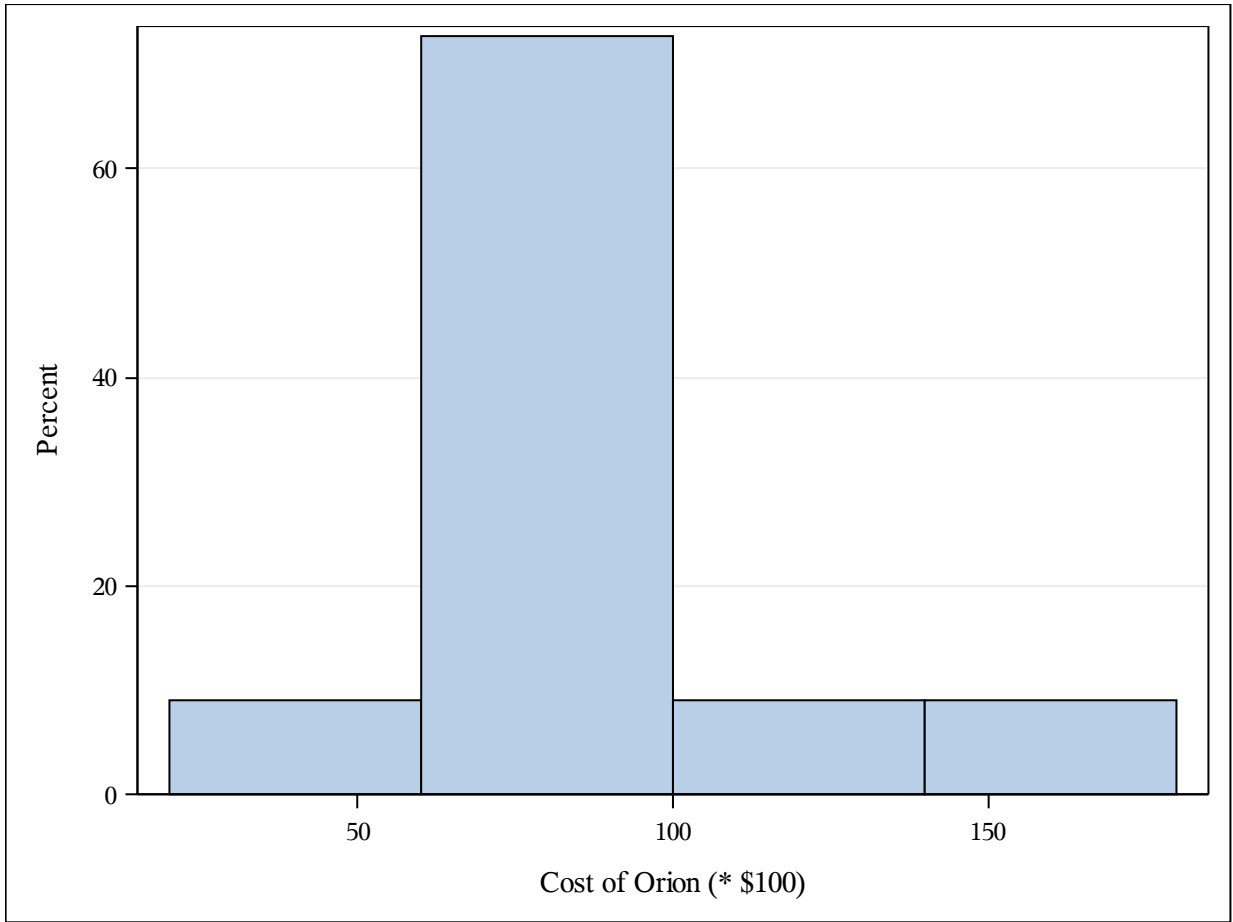
```
*****;  
* copy the ageorionsdata dataset into your work directory;  
* just remember if you change the dataset to copy the updated  
* dataset back to your MYDATA library so you have the  
* updated dataset since the WORK library id deleted  
* when you exit SAS;  
*****;  
  
data orions;  
  set mydata.ageorionsdata;  
run;
```

Create a Histogram of the Price of the cars

1. Click on Tasks, then Graph, then double click Histogram
2. Under the DATA tab
 - a. Select work.orions for data
 - b. Select Price for the Analysis variable under Roles
3. Under the OPTIONS tab
 - a. Do nothing – just keep the defaults – but you could add many options here
4. Then click RUN (the little running man icon)
5. Copy and Paste the code generated back into your SAS program (e.g. mkh1.sas)

```
*****;  
* example Tasks/Graph/Histogram *;  
*****;  
/*  
*  
* Task code generated by SAS Studio 3.1  
*  
* Generated on 'Sun Sep 14 2014 07:39:18 GMT-0400 (Eastern Standard Time)'  
* Generated by 'sasdemo'  
* Generated on server 'LOCALHOST'  
* Generated on SAS platform 'Linux LIN X64 2.6.32-431.11.2.el6.x86_64'  
* Generated on SAS version '9.04.01M1P12042013'  
* Generated on browser 'Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101  
Firefox/31.0'  
* Generated on web client 'http://localhost:10080/SASStudio/main?locale=en_US&zone=GMT-  
04%3A00'  
*  
*/  
  
/* Option group 5 (GRAPH SIZE) parameters. */  
/*--Set Graph Size (in inches)--*/  
ods graphics / reset width=6.4in height=4.8in imagemap;  
  
/*--SGPLOT proc statement--*/  
proc sgplot data=WORK.ORIONS;  
  /*--Histogram settings--*/  
  histogram Price /;  
  
  /*--Vertical or Response Axis--*/  
  yaxis grid;  
run;
```

RESULTS WINDOW (saved as a RTF (rick text format) and pasted here into MS WORD)

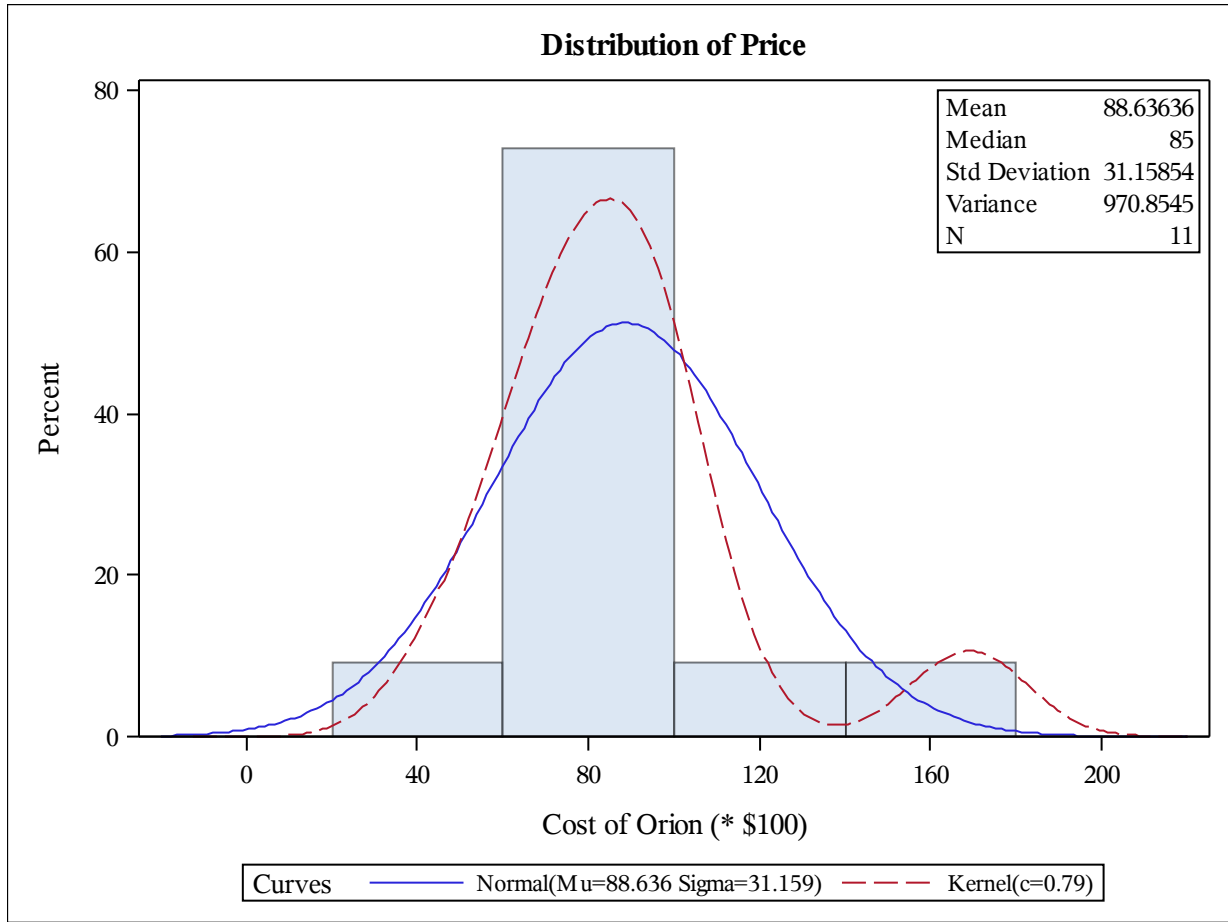


Run an analysis of the Distribution of the Price of the cars

1. Click on Tasks, then Statistics, then double click Distribution Analysis
2. Under the DATA tab
 - a. Select work.orions for data
 - b. Select Price for the Analysis variable under Roles
3. Under the OPTIONS tab
 - a. Under exploring data, click all 3: add normal curve; add kernel density estimate; add inset statistics
 - b. Under Inset Statistics, click: number of observations; mean; median; standard deviation; and variance
 - c. Under Checking for Normality, click: Normal probability plot and add inset statistics
 - d. Leave other options as is
4. Then click RUN (the little running man icon)
5. Copy and Paste the code generated back into your SAS program (e.g. mkh1.sas)

```
*****;  
* example Tasks/Statistics/Distribution Analysis *;  
*****;  
/*  
*  
* Task code generated by SAS Studio 3.1  
*  
* Generated on 'Sun Sep 14 2014 07:40:25 GMT-0400 (Eastern Standard Time)'  
* Generated by 'sasdemo'  
* Generated on server 'LOCALHOST'  
* Generated on SAS platform 'Linux LIN X64 2.6.32-431.11.2.el6.x86_64'  
* Generated on SAS version '9.04.01M1P12042013'  
* Generated on browser 'Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101  
Firefox/31.0'  
* Generated on web client 'http://localhost:10080/SASStudio/main?locale=en_US&zone=GMT-  
04%3A00'  
*  
*/  
  
ods noproctitle;  
ods select where=(lowcase(_path_) ? 'plot' or lowcase(_path_) ? 'gram');  
  
proc univariate data=WORK.ORIONS noprint;  
  histogram Price / normal kernel;  
  inset mean median std var n / position=ne;  
  probplot Price / normal(mu=est sigma=est);  
  inset n / position=nw;  
run;
```

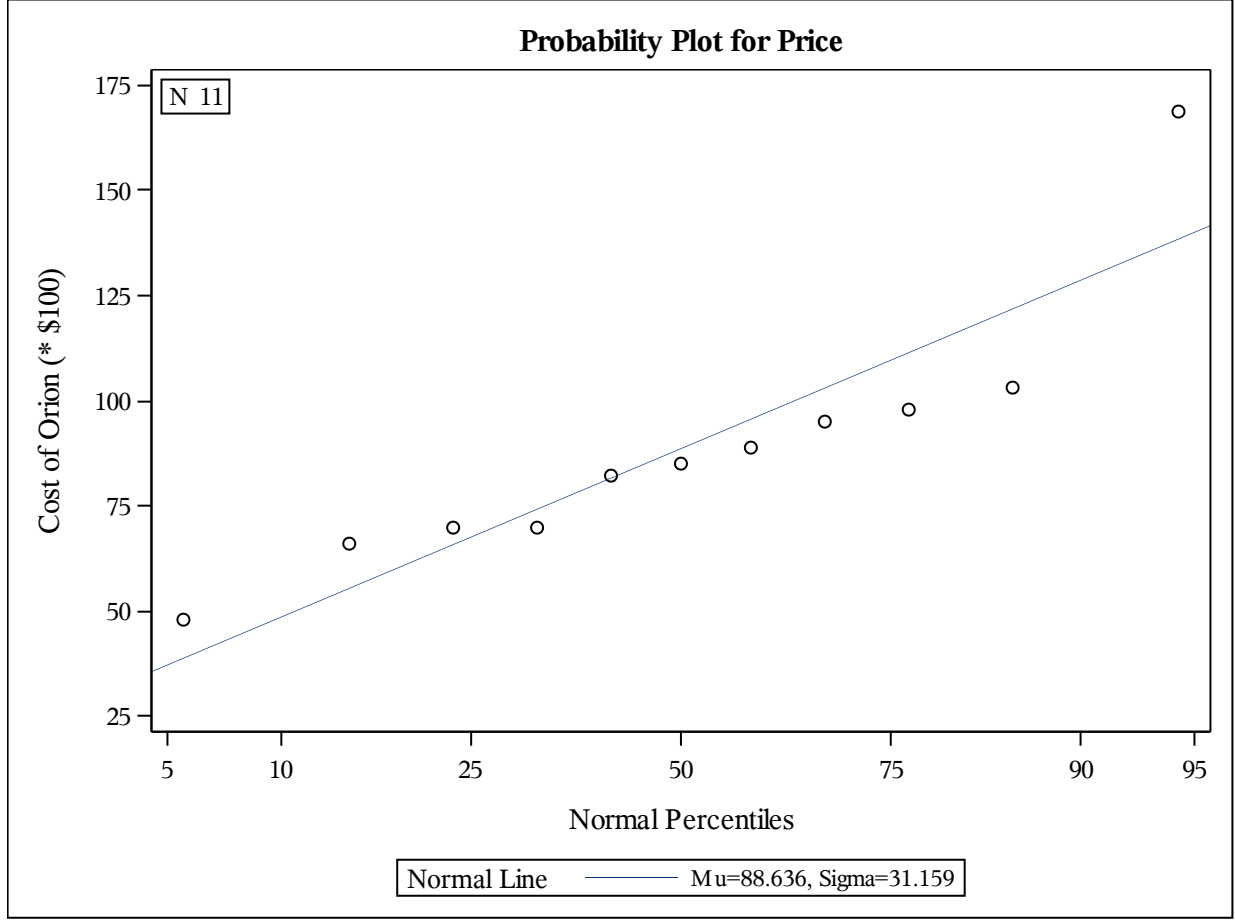
RESULTS WINDOW (saved as a RTF (rick text format) and pasted here into MS WORD)



Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	88.63636
Std Dev	Sigma	31.15854

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.23149539	Pr > D	0.096
Cramer-von Mises	W-Sq	0.10374973	Pr > W-Sq	0.090
Anderson-Darling	A-Sq	0.68532533	Pr > A-Sq	0.053

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	48.0000	16.1508
5.0	48.0000	37.3851
10.0	66.0000	48.7051
25.0	70.0000	67.6202
50.0	85.0000	88.6364
75.0	98.0000	109.6525
90.0	103.0000	128.5676
95.0	169.0000	139.8876
99.0	169.0000	161.1220

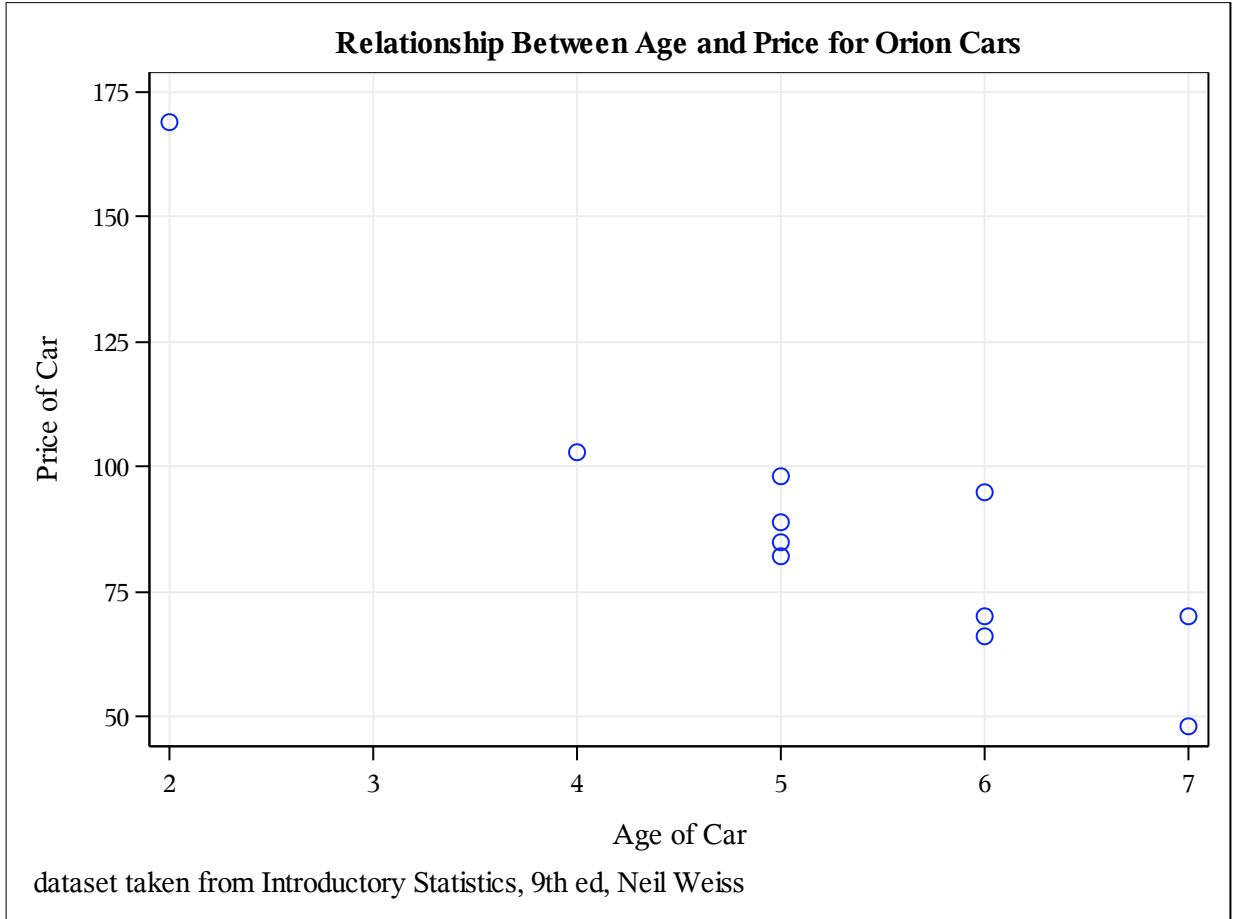


Create a Scatter Plot of the Price of the cars by the Age of the cars

1. Click on Tasks, then Graph, then double click Scatter Plot
2. Under the DATA tab
 - a. Select work.orions for data
 - b. Select Age as the X variable role
 - c. Select Price for the Y variable role
3. Under the OPTIONS tab
 - a. For Title type: Relationship Between Age and Price for Orion Cars
 - b. For Footnote type: dataset taken from Introductory Statistics, 9th ed, Neil Weiss
 - c. Under Marker Details, click apply marker color – choose one from the palette of colors, choose a symbol (I chose circle) and a marker size (here I choose 10).
 - d. Under X-axis add a Custom Label: Age of Car
 - e. Under Y-axis add a Custom Label: Price of Car
 - f. Leave other options as is
4. Then click RUN (the little running man icon)
5. Copy and Paste the code generated back into your SAS program (e.g. mkh1.sas)

```
*****;  
* example Tasks/Graph/Scatter Plot *;  
*****;  
/*  
*  
* Task code generated by SAS Studio 3.1  
*  
* Generated on 'Sun Sep 14 2014 07:57:35 GMT-0400 (Eastern Standard Time)'  
* Generated by 'sasdemo'  
* Generated on server 'LOCALHOST'  
* Generated on SAS platform 'Linux LIN X64 2.6.32-431.11.2.el6.x86_64'  
* Generated on SAS version '9.04.01M1P12042013'  
* Generated on browser 'Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101  
Firefox/31.0'  
* Generated on web client 'http://localhost:10080/SASStudio/main?locale=en_US&zone=GMT-  
04%3A00'  
*  
*/  
  
/*--Set output size--*/  
ods graphics / reset width=6.4in height=4.8in imagemap;  
  
/*--SGPLOT proc statement--*/  
proc sgplot data=WORK.ORIONS noautolegend;  
  /*--TITLE and FOOTNOTE--*/  
  title "Relationship Between Age and Price for Orion Cars";  
  footnote j=1  
    "dataset taken from Introductory Statistics, 9th ed, Neil Weiss";  
  
  /*--Scatter plot settings--*/  
  scatter x=Age y=Price / markerattrs=(symbol=Circle color=CX0824f7 size=10)  
    transparency=0.00 name='Scatter';  
;  
  
  /*--X Axis--*/  
  xaxis grid label="Age of Car";  
  
  /*--Y Axis--*/  
  yaxis grid label="Price of Car";  
run;
```


RESULTS WINDOW (saved as a RTF (rick text format) and pasted here into MS WORD)



Perform a Linear Regression for the Price of the Cars as “predicted by” the Age of the cars

1. Click on Tasks, then Statistics, then double click Linear Regression
2. Under the DATA tab
 - a. Select work.orions for data
 - b. Select Price for the Dependent variable role
 - c. Select Age as the Explanatory variable role
3. Under the METHODS tab:
 - a. Leave the defaults for 95% confidence level and Include Intercept
 - b. Under Model Selection Statistics, choose Adjusted R2
4. Under the OPTIONS tab
 - a. For Parameter Statistics, choose: standardized regression coefficients and confidence limits for estimates
 - b. Under plots, choose Diagnostic plots and Residuals for each explanatory variable
 - c. Also click Scatter plots: fit plot for a single explanatory variable, observed values by predicted values and (if you like) partial regression plots for each explanatory variable
 - d. Leave other options as is
5. Then click RUN (the little running man icon)
6. Copy and Paste the code generated back into your SAS program (e.g. mkh1.sas)

```
*****;
* example Tasks/Statistics/Linear Regression *;
*****;
/*
*
* Task code generated by SAS Studio 3.1
*
* Generated on 'Sun Sep 14 2014 08:01:35 GMT-0400 (Eastern Standard Time)'
* Generated by 'sasdemo'
* Generated on server 'LOCALHOST'
* Generated on SAS platform 'Linux LIN X64 2.6.32-431.11.2.el6.x86_64'
* Generated on SAS version '9.04.01M1P12042013'
* Generated on browser 'Mozilla/5.0 (Windows NT 6.1; WOW64; rv:31.0) Gecko/20100101
Firefox/31.0'
* Generated on web client 'http://localhost:10080/SASStudio/main?locale=en_US&zone=GMT-
04%3A00'
*
*/

ods noproctitle;

proc reg data=WORK.ORIONS alpha=0.05 plots(only)=(diagnostics residuals partial
    fitplot observedbypredicted);
    model Price=Age / stb clb partial adjrsq;
run;
quit;
```

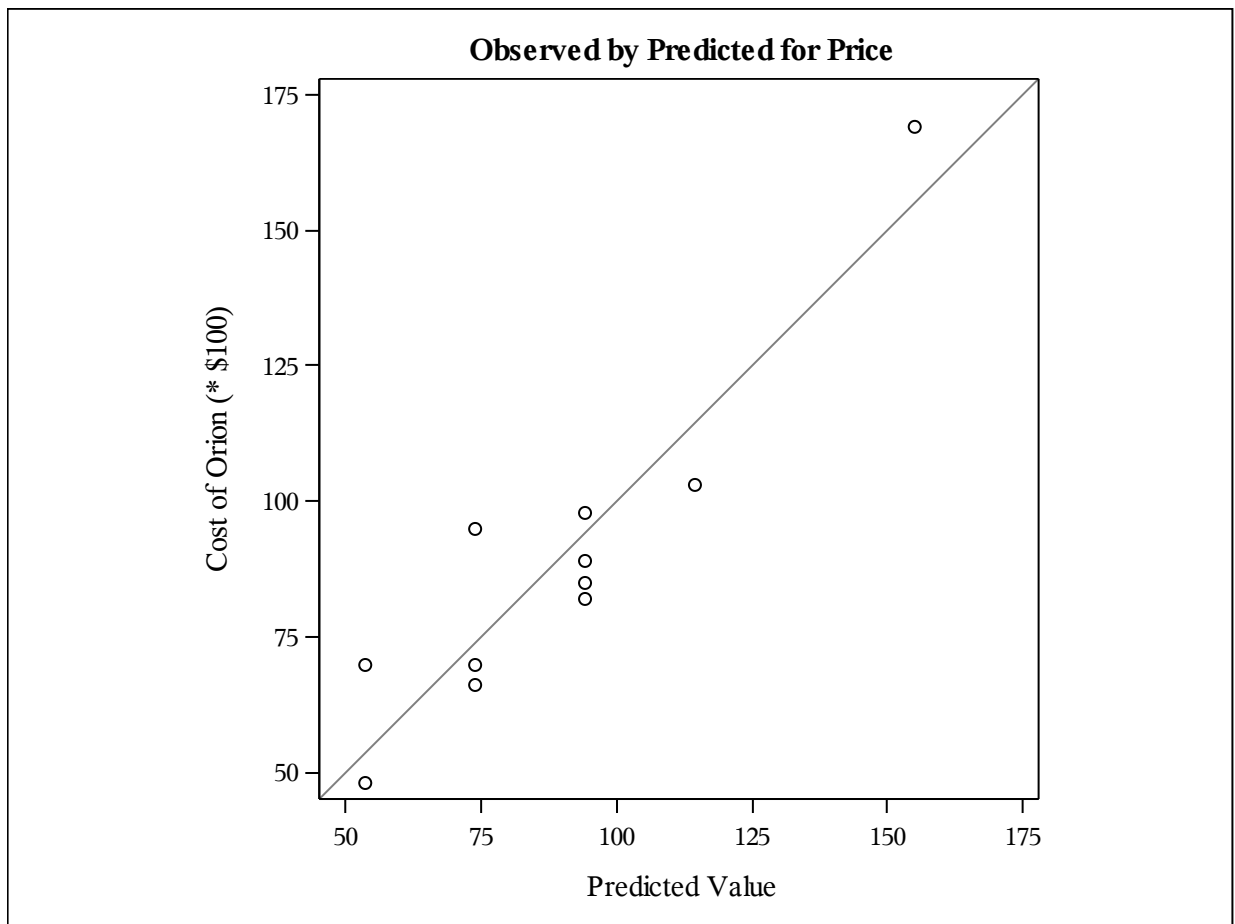
RESULTS WINDOW (saved as a RTF (rick text format) and pasted here into MS WORD)

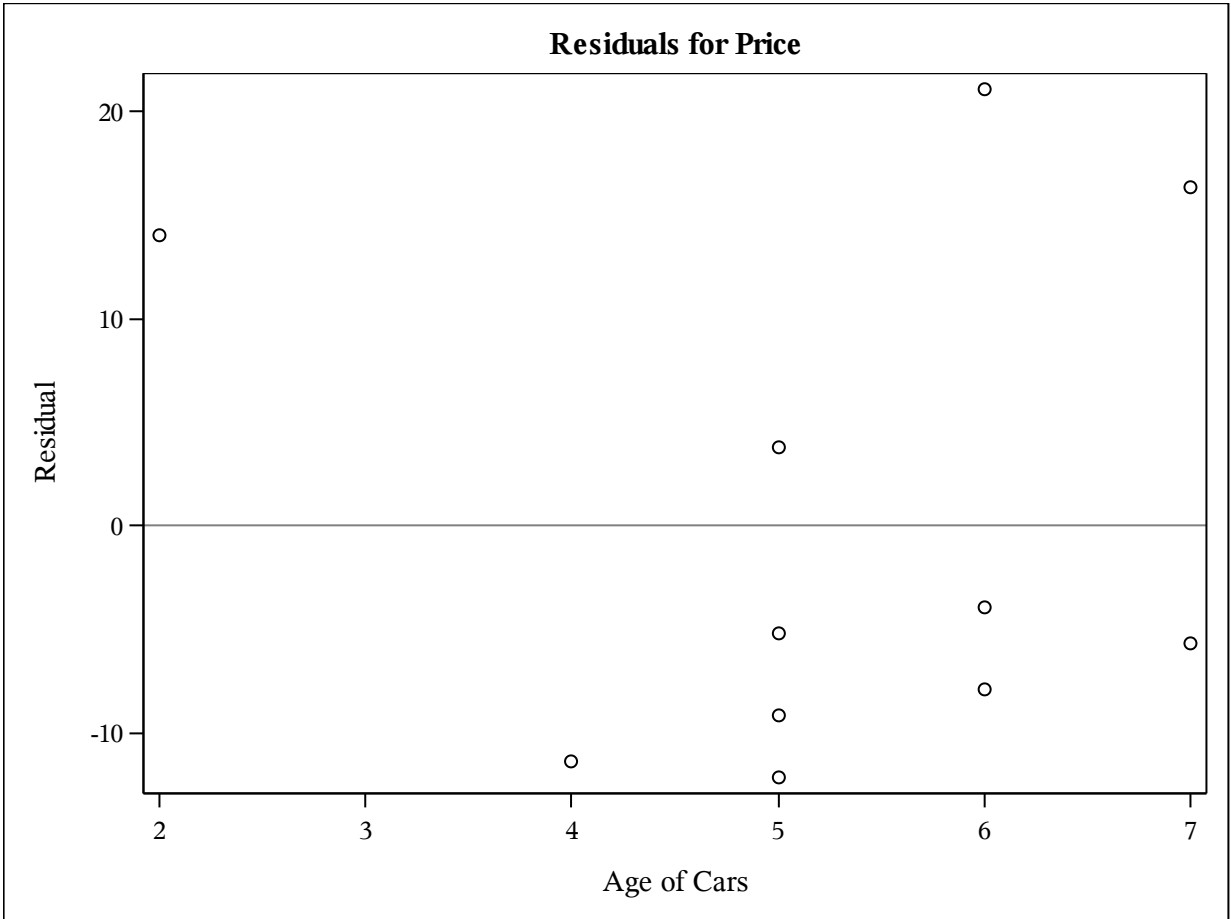
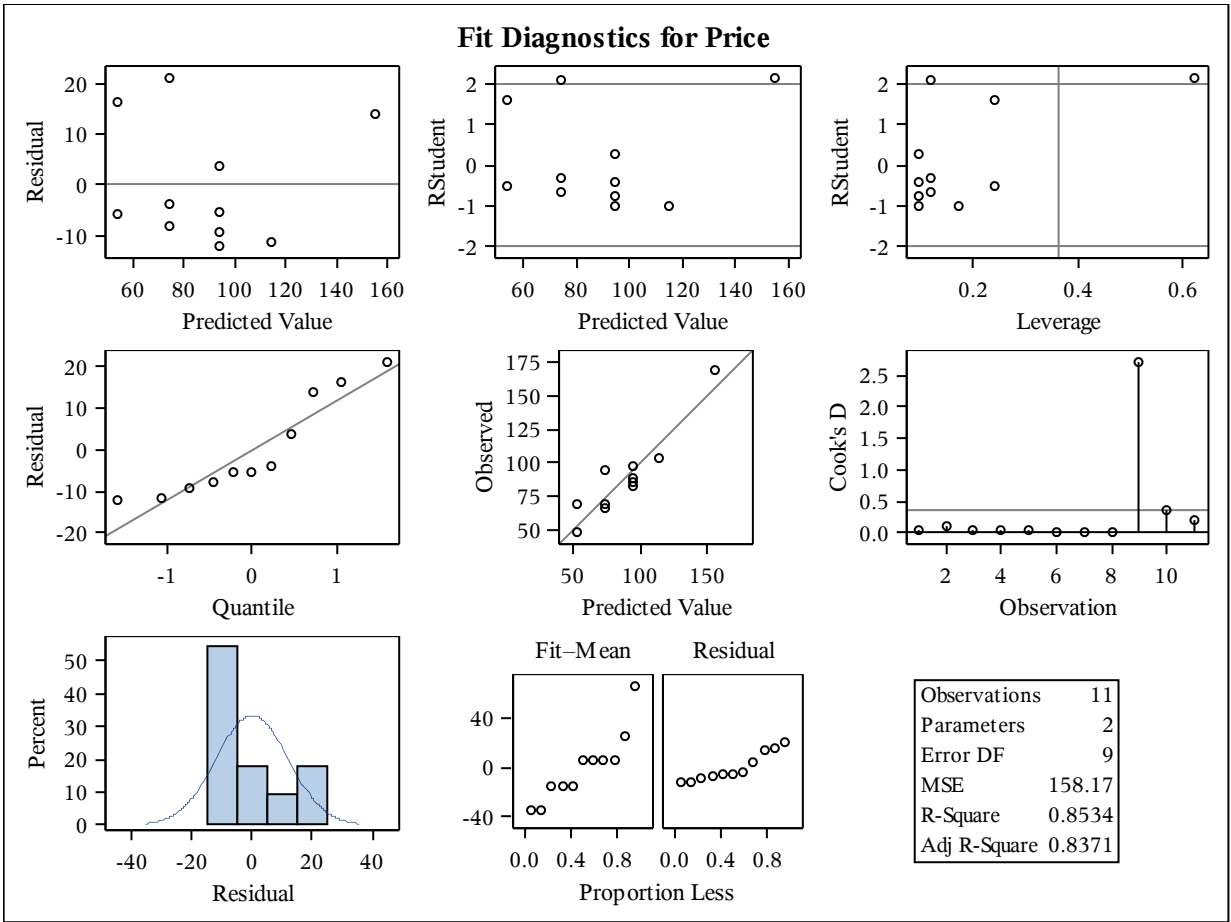
Number of Observations Read	11
Number of Observations Used	11

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8285.01392	8285.01392	52.38	<.0001
Error	9	1423.53153	158.17017		
Corrected Total	10	9708.54545			

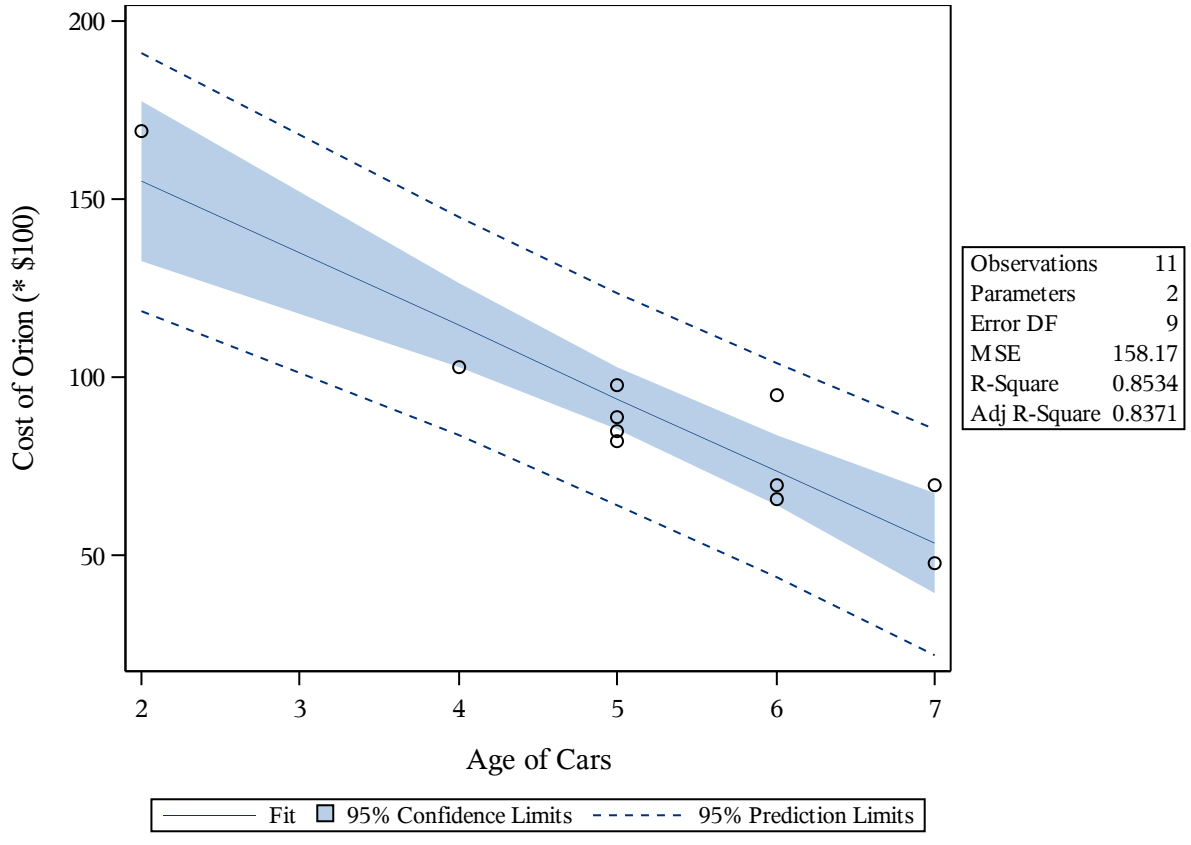
Root MSE	12.57657	R-Square	0.8534
Dependent Mean	88.63636	Adj R-Sq	0.8371
Coeff Var	14.18895		

Parameter Estimates									
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate	95% Confidence Limits	
Intercept	Intercept	1	195.46847	15.24034	12.83	<.0001	0	160.99243	229.94451
Age	Age of Cars	1	-20.26126	2.79951	-7.24	<.0001	-0.92378	-26.59419	-13.92833

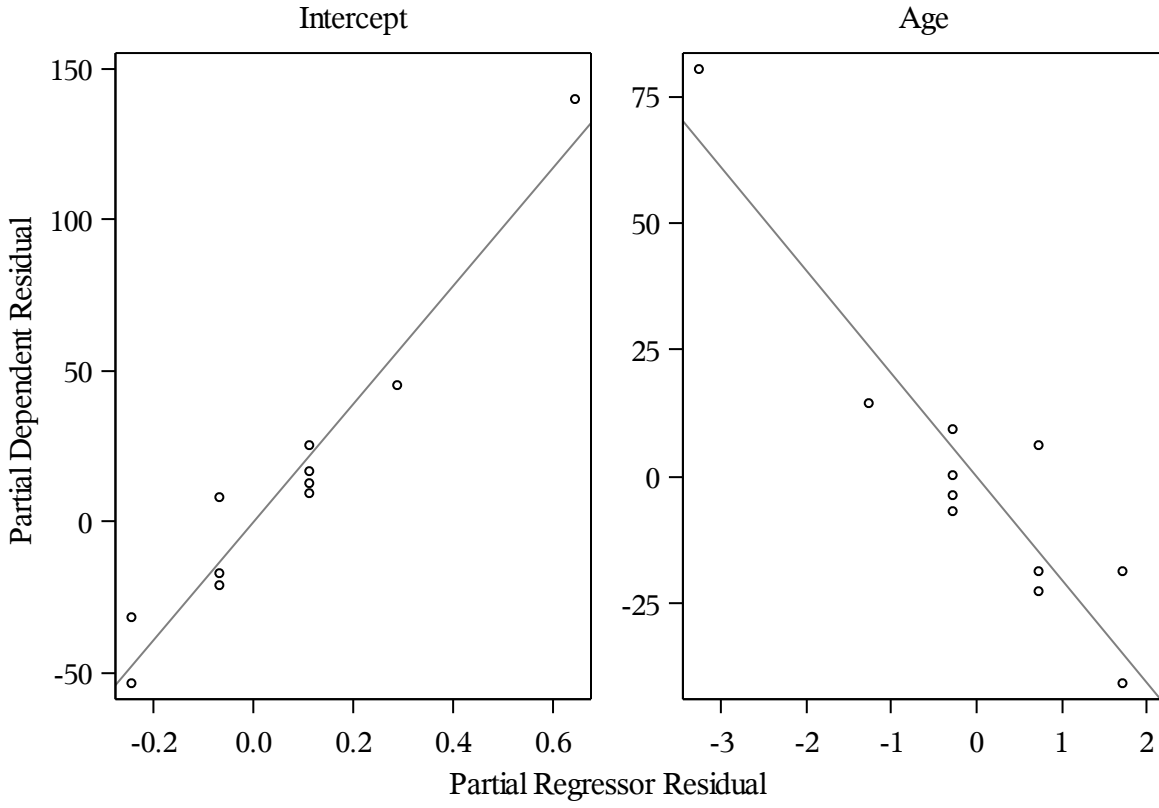




Fit Plot for Price



Partial Plots for Price



Let's modify the original dataset slightly by adding a new variable column of all 1s

```
*****;  
* If the dataset had changed make a copy of the updated  
* datafile *;  
*****;  
  
* example of adding a variable to the dataset;  
  
data orions2;  
  set orions;  
  newvar = 1; /* add a variable column of all 1s */  
run;
```

Save a copy of the modified dataset back to your original library (i.e. back to your flash drive)

```
* now save the updated dataset back into your library;  
  
data mydata.orions2;  
  set orions2;  
run;
```

NOTE: In SAS Studio, you can run other SAS procedures even though they are not available through the Tasks point-and-click interface. For example, suppose we code the Price of the Orion Cars into those less than or equal to \$85,000 and those greater than \$85,000. And then run a logistic regression for Price greater than \$85,000 using the Age of the car as the predictor.

```
data work.orions3;
  set orions;
  price_gt85 = price > 85;
run;

proc logistic data=orions3 plots=all;
  model price_gt85 = age;
run;
```

Here is the output that is generated to the results window

Model Information	
Data Set	WORK.ORIONS3
Response Variable	price_gt85
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	11
Number of Observations Used	11

Response Profile		
Ordered Value	price_gt85	Total Frequency
1	0	6
2	1	5

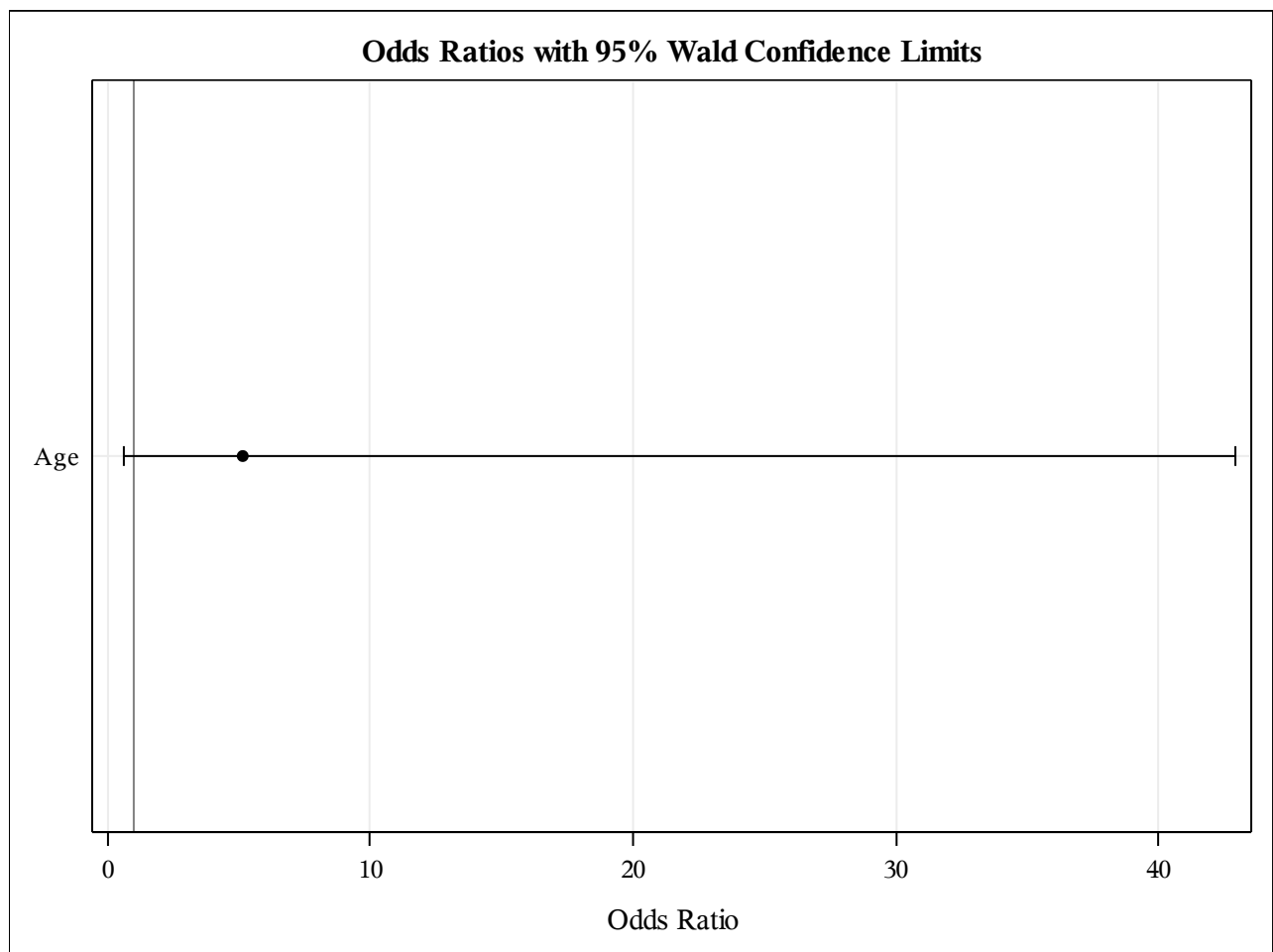
Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17.158	14.180
SC	17.556	14.976
-2 Log L	15.158	10.180

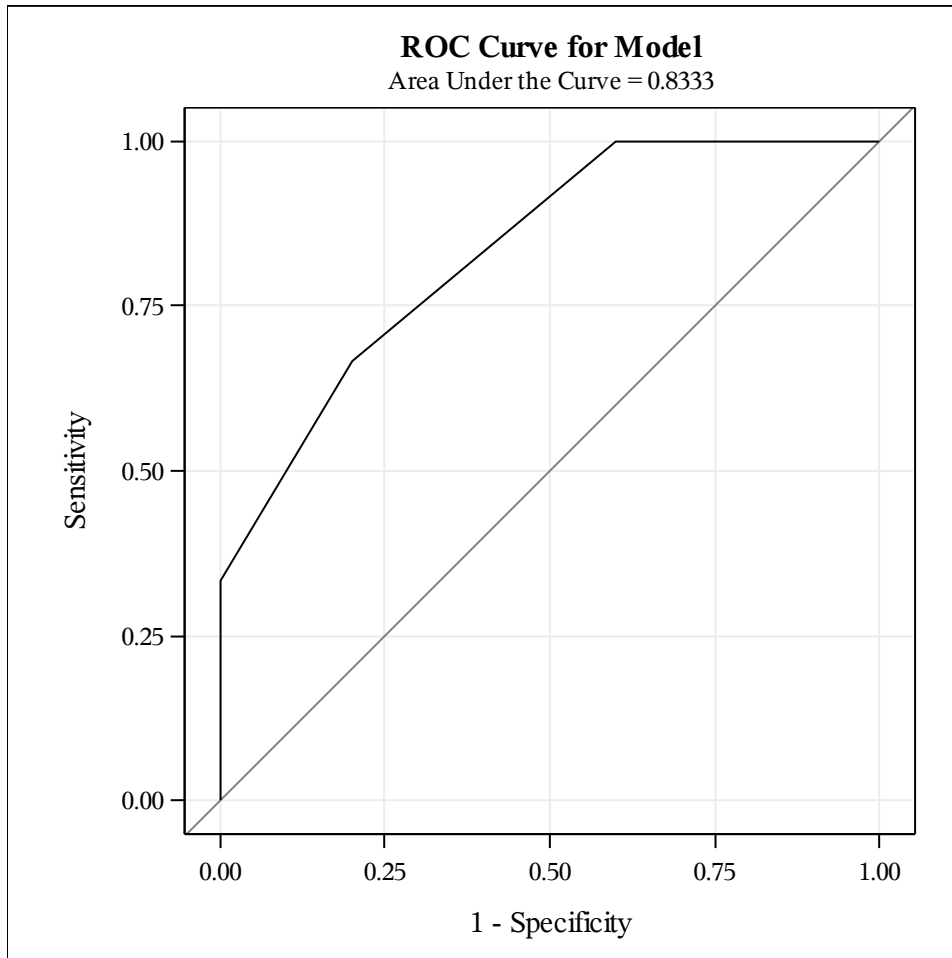
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.9782	1	0.0257
Score	3.8054	1	0.0511
Wald	2.2781	1	0.1312

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.5445	5.8481	2.1347	0.1440
Age	1	1.6357	1.0837	2.2781	0.1312

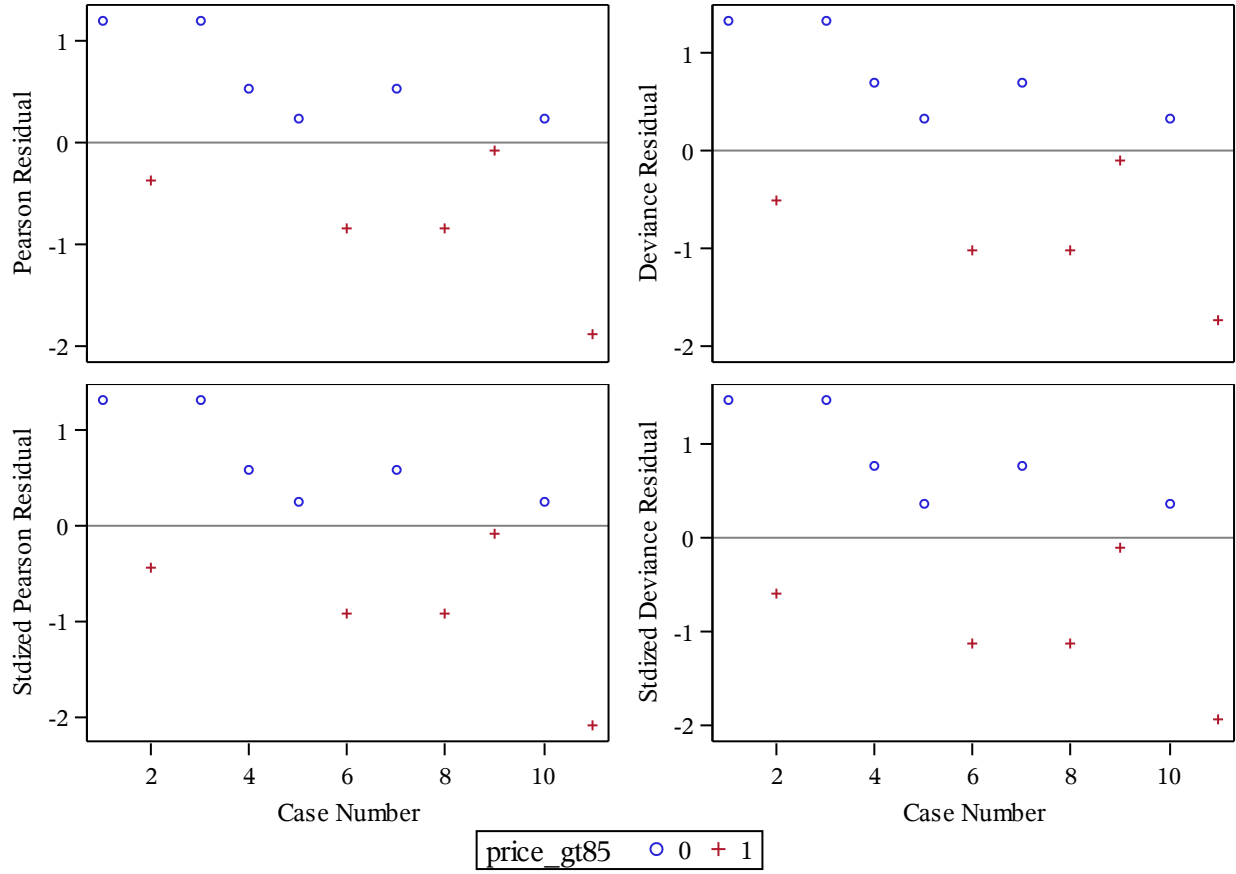
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	5.133	0.614	42.934



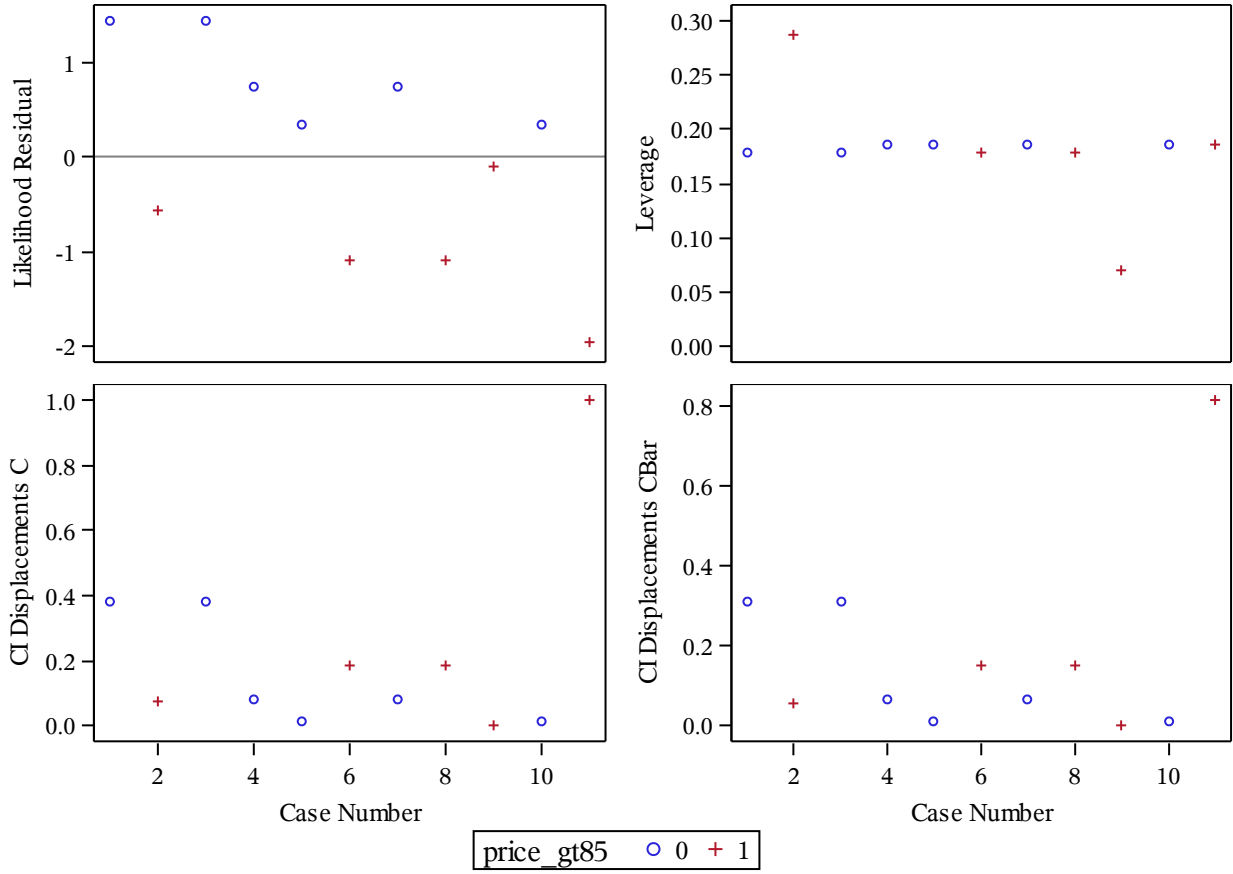
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	73.3	Somers' D	0.667
Percent Discordant	6.7	Gamma	0.833
Percent Tied	20.0	Tau-a	0.364
Pairs	30	c	0.833



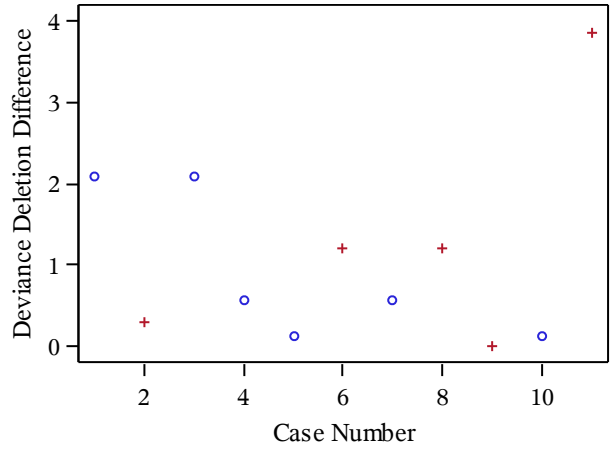
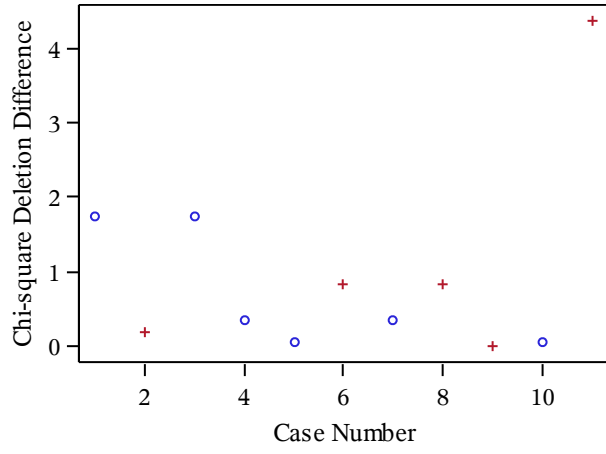
Influence Diagnostics



Influence Diagnostics

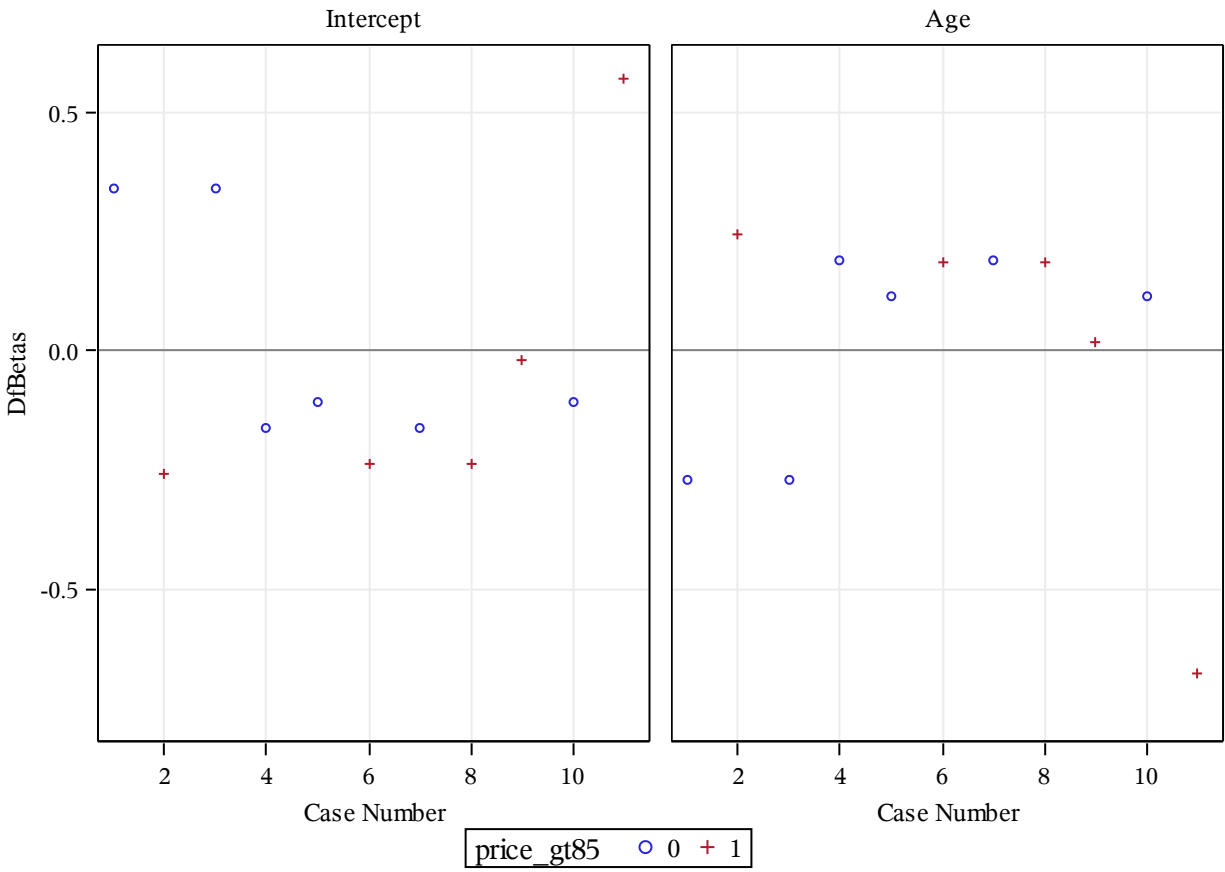


Influence Diagnostics

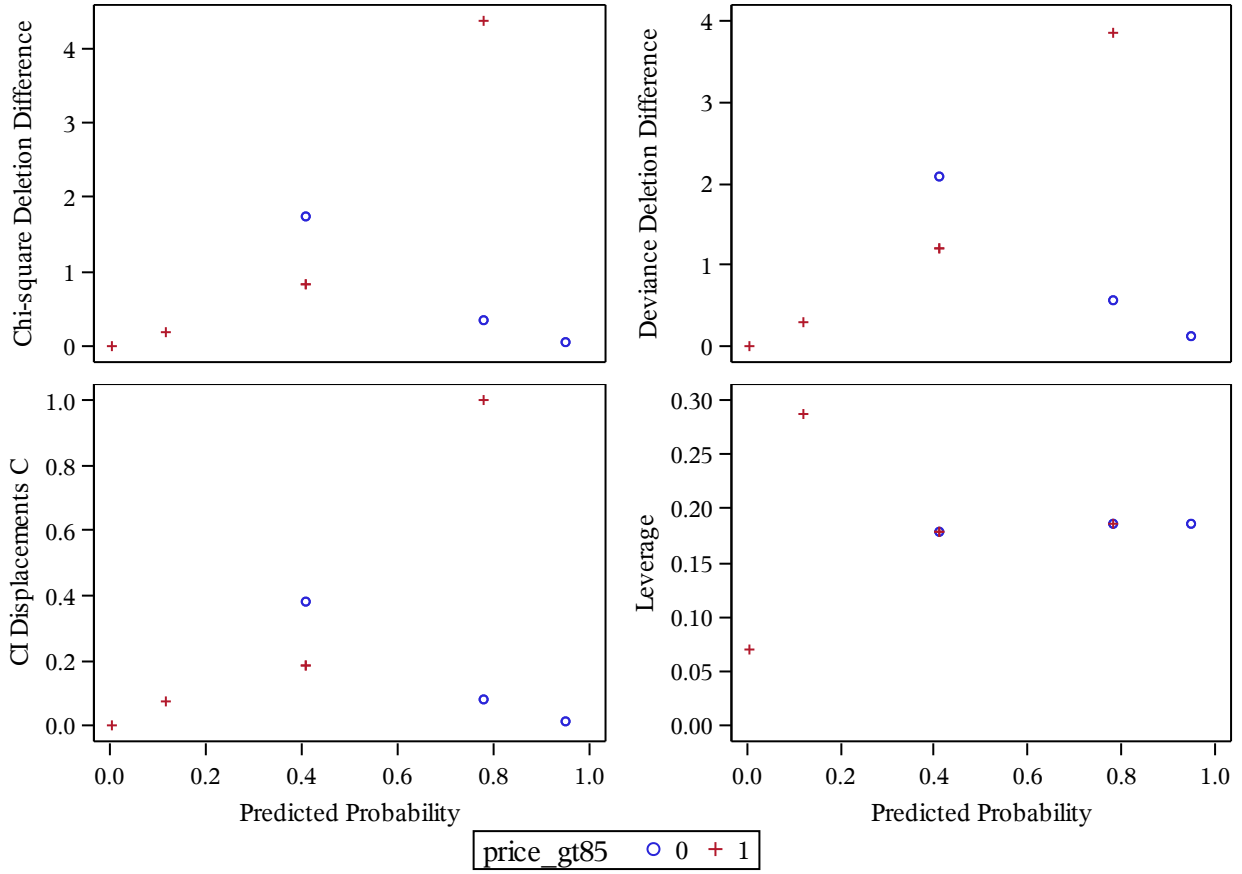


price_gt85 ○ 0 + 1

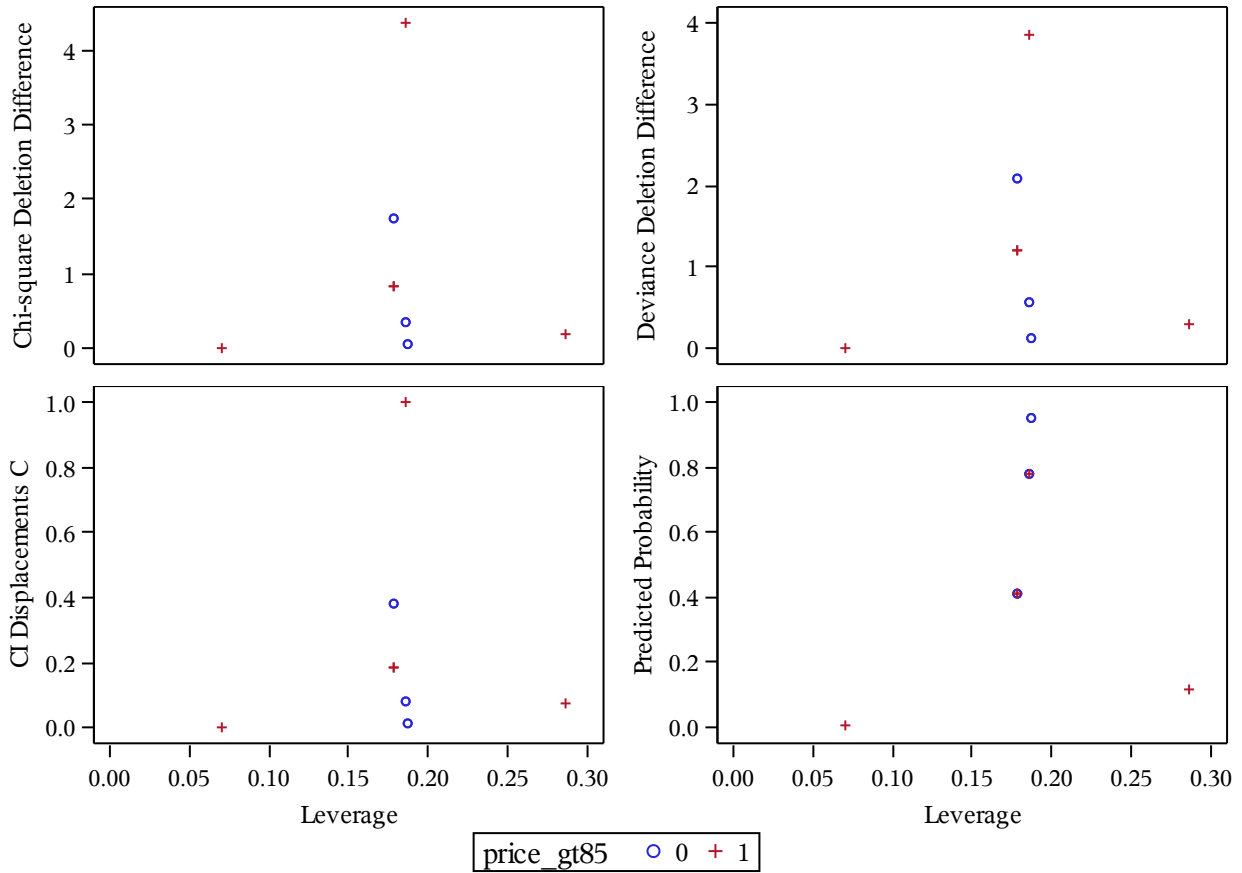
Influence Diagnostics



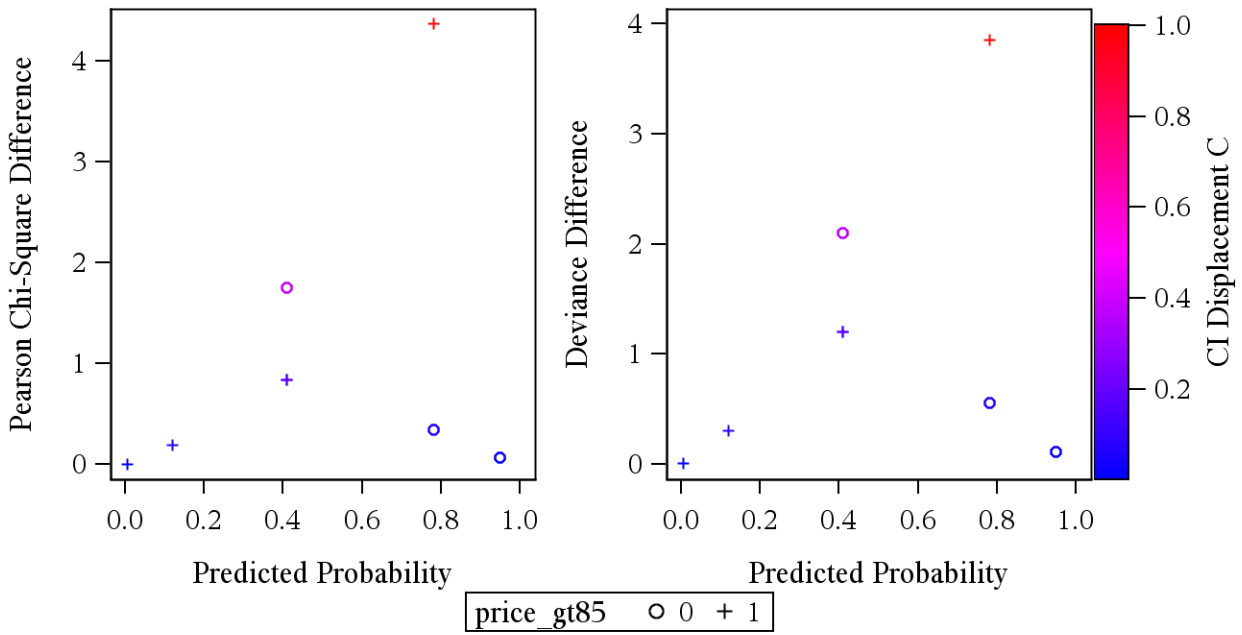
Predicted Probability Diagnostics



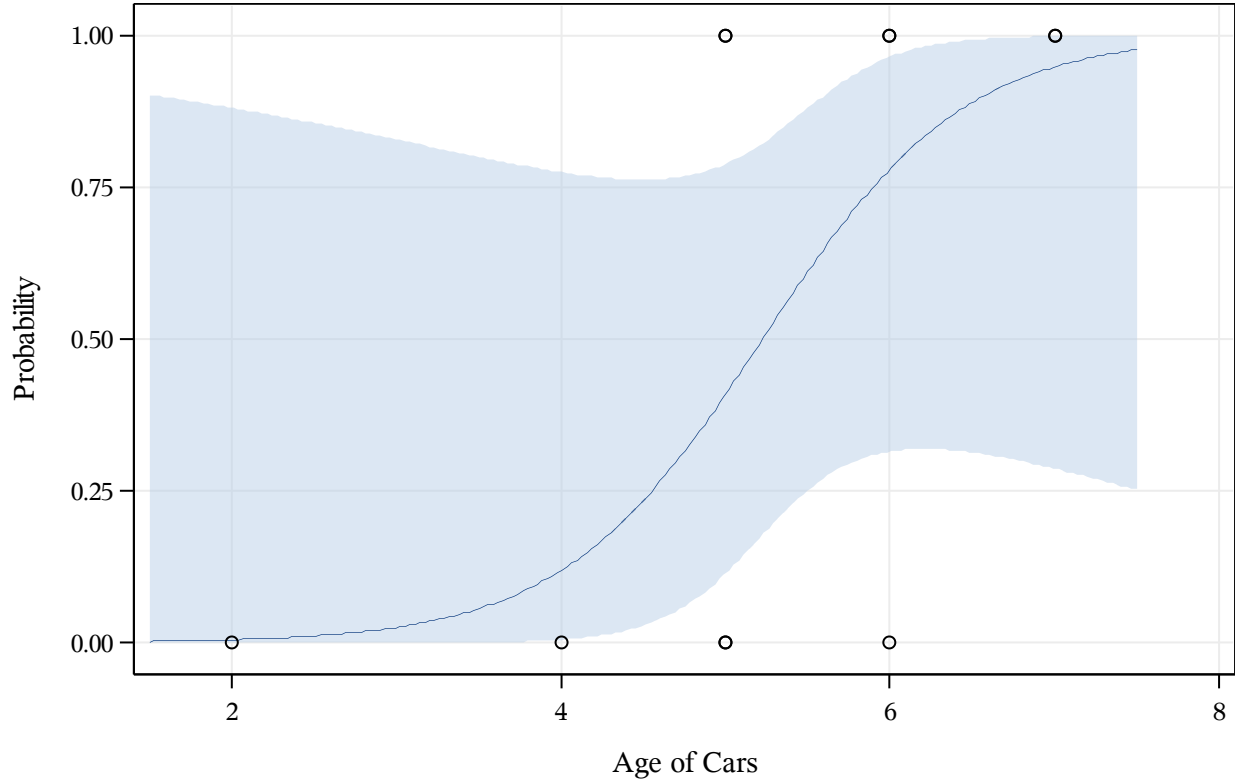
Leverage Diagnostics



Influence on the Model Fit and Parameter Estimates



Predicted Probabilities for price_gt85=0 With 95% Confidence Limits



○ Observed — Predicted