

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69      /* importing data from Excel; I converted the .xlsx file to a .csv file */
70      /* setting column names to match Excel file */
71      data vehicles;
72      infile '/home/u63996162/Personal Research Projects/Vehicle Data.csv'
73      dlm=',' dsd firstobs=2 missover;
74      length Vehicle $50 driveWheels $20 vehicleType $20;
75      input Vehicle :$50. MSRP dealerCost engineSize numCylinders Horsepower
76      cityMPG hwyMPG Weight Wheelbase Length Width
77      driveWheels :$20. vehicleType :$20.;
78      run;

```

NOTE: The infile '/home/u63996162/Personal Research Projects/Vehicle Data.csv' is:
 Filename=/home/u63996162/Personal Research Projects/Vehicle Data.csv,
 Owner Name=u63996162,Group Name=oda,
 Access Permission=-rw-r--r--,
 Last Modified=28Apr2025:13:43:25,
 File Size (bytes)=32929

NOTE: 428 records were read from the infile '/home/u63996162/Personal Research Projects/Vehicle Data.csv'.
 The minimum record length was 52.
 The maximum record length was 104.

NOTE: The data set WORK.VEHICLES has 428 observations and 14 variables.

NOTE: DATA statement used (Total process time):

real time	0.02 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	764.87k
OS Memory	19620.00k
Timestamp	09/02/2025 09:12:45 PM
Step Count	24 Switch Count 2
Page Faults	0
Page Reclaims	138
Page Swaps	0
Voluntary Context Switches	23
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	264

```

79
80      /* printing the data set for a visual check */
81      /*
82      proc print data=vehicles;
83      run;
84
85
86      /* transform the driveWheels variable to be a numeric, indicator variable */
87      /* transform the vehicleType variable to be numeric indicator variables */
88      data codedVehicles;
89      set vehicles;
90      if upcase(driveWheels) = 'AWD' then updatedDriveWheels = 1;
91      else updatedDriveWheels = 0;
92      if upcase(vehicleType) = "SEDAN" then isSedan = 1;
93      else isSedan = 0;
94      if upcase(vehicleType) = "SPORTS CAR" then isSportsCar = 1;
95      else isSportsCar = 0;
96      if upcase(vehicleType) = "SUV" then isSUV = 1;
97      else isSUV = 0;
98      if upcase(vehicleType) = "WAGON" then isWagon = 1;
99      else isWagon = 0;
100     if upcase(vehicleType) = "MINIVAN" then isMinivan = 1;
101     else isMinivan = 0;
102     run;

```

NOTE: There were 428 observations read from the data set WORK.VEHICLES.

NOTE: The data set WORK.CODEDVEHICLES has 428 observations and 20 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	997.28k
OS Memory	19880.00k
Timestamp	09/02/2025 09:12:45 PM
Step Count	25 Switch Count 2
Page Faults	0
Page Reclaims	128
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

```

103
104      /* Base specification model */
105      proc reg data=codedVehicles plots=none;
106      model MSRP = engineSize numCylinders Horsepower cityMPG
107      hwyMPG Weight Wheelbase Length
108      Width updatedDriveWheels isSedan isSportsCar
109      isSUV isWagon isMinivan;
110      run;

111
112      /* Stepwise selection step */

```

NOTE: PROCEDURE REG used (Total process time):

real time	0.03 seconds
user cpu time	0.04 seconds
system cpu time	0.00 seconds
memory	3550.53k
OS Memory	22204.00k
Timestamp	09/02/2025 09:12:45 PM
Step Count	26 Switch Count 1
Page Faults	0
Page Reclaims	785
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	3
Block Input Operations	0
Block Output Operations	56

```

113      proc reg data=codedVehicles;
114      model MSRP = engineSize numCylinders Horsepower cityMPG
115      hwyMPG Weight Wheelbase Length
116      Width updatedDriveWheels isSedan isSportsCar
117      isSUV isWagon isMinivan / selection=stepwise slentry=0.1 slstay=0.1;
118      run;

119
120      /* Multicollinearity check */

```

NOTE: PROCEDURE REG used (Total process time):

real time	2.86 seconds
user cpu time	0.33 seconds
system cpu time	0.04 seconds
memory	14160.37k
OS Memory	35140.00k
Timestamp	09/02/2025 09:12:48 PM
Step Count	27 Switch Count 22
Page Faults	0
Page Reclaims	15518

```

Page Swaps                                0
Voluntary Context Switches               1084
Involuntary Context Switches             11
Block Input Operations                    0
Block Output Operations                   2240

```

```

121      proc reg data=codedVehicles;
122      model MSRP = Horsepower Wheelbase hwyMPG Weight
123      Width isSUV numCylinders engineSize isSportsCar
124      / vif collinoint;
125      run;

126
127      /* Suspected interaction between predictors step */

```

NOTE: PROCEDURE REG used (Total process time):

```

real time          0.54 seconds
user cpu time      0.20 seconds
system cpu time    0.03 seconds
memory            11730.68k
OS Memory          40928.00k
Timestamp          09/02/2025 09:12:48 PM
Step Count         28  Switch Count  22
Page Faults        0
Page Reclaims      12537
Page Swaps         0
Voluntary Context Switches  918
Involuntary Context Switches  7
Block Input Operations  0
Block Output Operations 1880

```

```

128      proc glm data=codedVehicles;
129      model MSRP = Horsepower Wheelbase hwyMPG Weight
130      Width isSUV numCylinders engineSize isSportsCar
131      hwyMPG*Weight isSportsCar*Horsepower;
132      run;

133
134      /* Final model analysis */

```

NOTE: PROCEDURE GLM used (Total process time):

```

real time          0.05 seconds
user cpu time      0.05 seconds
system cpu time    0.00 seconds
memory            1929.18k
OS Memory          34484.00k
Timestamp          09/02/2025 09:12:49 PM
Step Count         29  Switch Count  2
Page Faults        0
Page Reclaims      241
Page Swaps         0
Voluntary Context Switches  24
Involuntary Context Switches  2
Block Input Operations  0
Block Output Operations 312

```

```

135      proc glm data=codedVehicles;
136      model MSRP = Horsepower Wheelbase hwyMPG Weight
137      Width isSUV numCylinders engineSize isSportsCar isSportsCar*Horsepower;
138      output out=residuals predicted=predict residual=res;
139      run;

140
141      /* normality assumption check; normal probability plot of residuals */

```

NOTE: The data set WORK.RESIDUALS has 428 observations and 22 variables.

NOTE: PROCEDURE GLM used (Total process time):

real time	0.04 seconds
user cpu time	0.05 seconds
system cpu time	0.00 seconds
memory	2329.87k
OS Memory	34744.00k
Timestamp	09/02/2025 09:12:49 PM
Step Count	30 Switch Count 4
Page Faults	0
Page Reclaims	262
Page Swaps	0
Voluntary Context Switches	27
Involuntary Context Switches	3
Block Input Operations	0
Block Output Operations	568

```

142     proc univariate data=residuals;
143         var res;
144         probplot res / normal (mu=est sigma=est);
145         title "Normal Probability Plot of Residuals";
146     run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time	0.13 seconds
user cpu time	0.08 seconds
system cpu time	0.01 seconds
memory	8258.37k
OS Memory	37888.00k
Timestamp	09/02/2025 09:12:49 PM
Step Count	31 Switch Count 1
Page Faults	0
Page Reclaims	1430
Page Swaps	0
Voluntary Context Switches	176
Involuntary Context Switches	4
Block Input Operations	0
Block Output Operations	400

```

147
148     /* scatter plot of predicted versus residuals for constant variance test */
149     /* also the model specification test */
150     proc sgplot data=residuals;
151         scatter x=predict y=res;
152         refline 0 / axis=y;
153         title "Plot of Predicted Values Versus Residuals";
154     run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.06 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	1303.71k
OS Memory	34348.00k
Timestamp	09/02/2025 09:12:49 PM
Step Count	32 Switch Count 2
Page Faults	0
Page Reclaims	316
Page Swaps	0
Voluntary Context Switches	143
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	432

NOTE: There were 428 observations read from the data set WORK.RESIDUALS.

```

155
156     data residuals_time;
157         set residuals;
158         obs_number + 1; /* creates a time-like variable starting at 1 */
159     run;

```

NOTE: There were 428 observations read from the data set WORK.RESIDUALS.

NOTE: The data set WORK.RESIDUALS_TIME has 428 observations and 23 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	984.65k
OS Memory	33704.00k
Timestamp	09/02/2025 09:12:49 PM
Step Count	33
Page Faults	0
Page Reclaims	97
Page Swaps	0
Voluntary Context Switches	17
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

160
161     /* scatter plot of residuals versus the time that they come in */
162     /* independence of errors test */
163     proc sgplot data=residuals_time;
164         scatter x=obs_number y=res;
165         refline 0 / axis=y;
166         title "Independence of Errors Test";
167     run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.06 seconds
user cpu time	0.03 seconds
system cpu time	0.01 seconds
memory	2184.12k
OS Memory	34348.00k
Timestamp	09/02/2025 09:12:49 PM
Step Count	34
Page Faults	0
Page Reclaims	291
Page Swaps	0
Voluntary Context Switches	140
Involuntary Context Switches	2
Block Input Operations	0
Block Output Operations	416

NOTE: There were 428 observations read from the data set WORK.RESIDUALS_TIME.

```

168
169
170
171     OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
181

```