**Linear Regression Code**

/\* Missing value calculation \*/

**proc** **format**;

value $missfmt ' ' = 'Missing' other = 'Not missing';

value missfmt **.** = 'Missing' other = 'Not missing';

**quit**;

**proc** **freq** data = sashelp.cars;

format \_char\_ $missfmt.;

tables \_char\_/missing missprint nocum nopercent;

format \_numeric\_ missfmt.;

tables \_numeric\_/missing missprint nocum nopercent;

**quit**;

/\* Creating dummy variables & Dev/Val creation \*/

**data** develop validat;

set sashelp.cars;

if Type = 'Hybrid' then do;

T1 = **0**;T2=**0**;T3=**0**;T4=**0**;T5=**0**; end;

else if Type = 'SUV' then do;

T1 = **1**;T2=**0**;T3=**0**;T4=**0**;T5=**0**; end;

else if Type = 'Sedan' then do;

T1 = **0**;T2=**1**;T3=**0**;T4=**0**;T5=**0**; end;

else if Type = 'Sports' then do;

T1 = **0**;T2=**0**;T3=**1**;T4=**0**;T5=**0**; end;

else if Type = 'Truck' then do;

T1 = **0**;T2=**0**;T3=**0**;T4=**1**;T5=**0**; end;

else do;

T1 = **0**;T2=**0**;T3=**0**;T4=**0**;T5=**1**; end;

/\* Origin \*/

if Origin = 'Asia' then do;

O1=**0**;O2=**0**; end;

else if Origin = 'Europe' then do;

O1=**1**;O2=**0**; end;

else do;

O1=**0**;O2=**1**; end;

if DriveTrain = 'All' then do;

D1=**0**;D2=**0**; end;

else if DriveTrain = 'Front' then do;

D1=**1**;D2=**0**; end;

else do;

D1=**0**;D2=**1**; end;

randnum = round(rand ('UNIFORM'),**.1**);

if randnum in (**0**,**0.1**,**0.2**,**0.4**,**0.8**) then output validat;

else output develop;

**run**;

/\* Removing multi collinearity \*/

**proc** **reg** data = develop;

model MPG\_City =

MPG\_Highway

/\*MSRP\*/

/\*Invoice\*/

/\*EngineSize \*/

/\*Cylinders \*/

/\*Horsepower \*/

/\*Weight\*/

/\*Wheelbase\*/

Length

T1

/\*T2\*/

/\*T3\*/

T4

/\*T5\*/

O1

/\*O2\*/

/\*D1\*/

/\*D2\*/

/vif

;

/\* Checking IID assumption \*/

**proc** **reg** data = develop;

model MPG\_City = MPG\_Highway Length T1 T4 O1/dw spec;

output out = resids r = res;

**run**;

**proc** **univariate** data = resids

normal plot;

var res;

**run**;

**data** validat2;

set validat;

PredMPG\_City =

(**1.19550**) +

(MPG\_Highway\*(**0.93686**))+

(Length\*(-**0.03527**))+

(T1 \* (**2.29253**))+

(T4\*(**2.51997**))+

(O1\*(-**0.62102**))

;

**run**;

**proc** **corr** data = validat2;

var MPG\_City PredMPG\_City ;

**quit**;