Program Summary - Final_Code.sas

Execution Environment

Author:	u61480438
File:	/home/u61480438/BAN110/PROJECT/Final_Code.sas
SAS Platform:	Linux LIN X64 3.10.0-1062.12.1.el7.x86_64
SAS Host:	ODAWS02-USW2-2.ODA.SAS.COM
SAS Version:	9.04.01M7P08062020
SAS Locale:	en_GB
Submission Time:	20/03/2023, 23:46:08
Browser Host:	BRAS-BASE-OSHWON9551W-GRC-33-174-94-60-131.DSL.BELL.CA
User Agent:	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/111.0.0.0 Safari/537.36
Application Server:	ODAMID00-USW2-2.ODA.SAS.COM

Code: Final_Code.sas

```
/* Importing Auto_Mpg.data file and analyzing the variables dataset*/
LIBNAME PROJECT "/home/u61480438/BAN110/PROJECT";

Data PROJECT.AUTO_MPG;
    INFILE '/home/u61480438/BAN110/PROJECT/auto-mpg.data';
    INFORMAT Car_Name $30.;
    INPUT @1 Mpg 4.
      @8 Cylinders 1.
      @12 Displacement 5.
      @23 Horsepower 5.
      @34 Weight 5.
      @45 Acceleration 4.
      @52 Model_Year 2.
      @56 Origin 1.
      @58 Car_Name & $30.;
    FORMAT Mpg 4.1
           Displacement 5.1
           Horsepower 5.1
           Weight 6.1
           Acceleration 4.1;
Run;

TITLE "Analyzing the Auto MPG data";
PROC CONTENTS DATA=PROJECT.AUTO_MPG VARNUM;
RUN;

/* Printing the first ten observations */
TITLE 'First ten observations of the raw Auto_Mpg dataset';

PROC PRINT DATA=PROJECT.AUTO_MPG (OBS=10);
RUN;

/* Descriptive statistics  of Dependent/Target variable MPG */
Title 'Descrpitive Statitics for Dependent variable MPG';

Proc Means Data=project.Auto_Mpg;
    Var mpg;
Run;

/* Histogram of Dependent/Target variable MPG */
Title 'Histogram of MPG';

Proc Sgplot Data=project.Auto_Mpg;
    Histogram Mpg;
    Density Mpg;
    Density Mpg / type=kernel;
Run;

/* Working with Categorical Values*/
options nolabel;
Title 'Listing Frequencies for Cylinders Model_year and Origin';

Proc Freq Data=project.Auto_Mpg;
    Tables Cylinders Model_Year Origin / nocum missing;
Run;

/* Checking for missing values in categorical variables by using informat method*/
Proc Format;
    Value Origin_Check 1, 2, 3='Valid' other='Invalid';
    Value Cyl_Check 3, 4, 5, 6, 8='Valid' other='Invalid';
    Value Year_Check 70-82='Valid' other='Invalid';
```

```
Run;

Data _null_;
    File Print;
    Set Project.Auto_Mpg (Keep=Car_Name Cylinders Model_Year Origin);

    If put(Cylinders, Cyl_Check.)='Invalid' then
        put 'Missing observation of Cylinders = ' _n_ Car_name Cylinders=;
    Else if Put(Model_Year, Year_Check.)='Invalid' then
        put 'Missing observation of Model_Year = ' _n_ Car_name Model_Year=;
    Else if put(Origin, Origin_Check.)='Invalid' then
        put 'Missing observation of Origin = ' _n_ Car_name Origin=;
Run;

Title 'Checking for Missing values in Categorical variables';

Proc Freq Data=project.Auto_Mpg;
    Tables Cylinders Model_Year Origin / nocum nopercent;
    Format Cylinders Cyl_Check. Model_Year Year_Check. Origin Origin_Check.;
Run;

/* Converting Date from 2 digit number to Date9. format for full year*/
Data project.Auto_Mpg;
    Set project.Auto_Mpg;
    Year_new=Cat('03/01/19', Model_Year);
    Model_Year=year(input(Year_new, mmddyy10.));
    Drop Year_new;
Run;

Title 'Printing first 5 observations after date conversion';

Proc Print Data=project.Auto_Mpg (obs=5);
    Var Car_Name Model_Year;
Run;

/*  Deriving Vehicle brand name and Model from car_Name variable and drop the car_name*/
Data project.Auto_Mpg;
    Set project.Auto_Mpg;
    Car_Name=Propcase(Compress(car_Name, '"'));
    Array model_n [6] $20. Model1-Model6;

    Do i=1 to 6;
        Model_n [i]=compress(Scan(Car_Name, i), "'");
    End;

    If _n_=293 then
        Model3='';
    Brand=Model1;
    Model=Catx('', Model2, model3, model4, model5, model6);
    Drop Model1-Model6 Car_Name i;
Run;
PROC PRINT DATA=project.auto_mpg (obs=5);
RUN;

/*Checking Errors in Brand Variable*/
Title 'Checking errors in Brand variable';

Proc Freq Data=project.Auto_Mpg;
    Tables Brand / nocum nopercent;
Run;

/*Checking and Correcting Spelling errors in Brand variable */
Data project.Auto_Mpg;
    Set project.Auto_Mpg;
    Brand=Tranwrd(Brand, 'Chevy', 'Chevrolet');
    Brand=Tranwrd(Brand, 'Chevroelt', 'Chevrolet');
    Brand=Tranwrd(Brand, 'Hi', 'Honda');
    Brand=Tranwrd(Brand, 'Maxda', 'Mazda');
    Brand=Tranwrd(Brand, 'Vw', 'Volkswagen');
    Brand=Tranwrd(Brand, 'Vokswagen', 'Volkswagen');
    Brand=Tranwrd(Brand, 'Toyouta', 'Toyota');
Run;

Title 'Corrected Brand variable';

Proc Freq Data=project.Auto_Mpg;
    Tables Brand / nocum nopercent;
Run;

/*Working with Numerical Variables*/
options nolabel;
```

```
Proc MEans Data=project.Auto_Mpg n nmiss min max mean median mode stddev var;
    Var mpg acceleration displacement weight horsepower;
Run;

/* Checking Missing Numeric Observations */
Title 'Identifying Missing numeric values';

Data _null_;
    File print;
    Set project.Auto_Mpg;
    Array Numeric [*] _NUMERIC_;

    Do i=1 to Dim(Numeric);

        If missing(numeric(i)) then
            put 'Missing Observation '
                Brand=Model=Mpg=Cylinders=Displacement=Horsepower=Weight=Acceleration=;
    End;
Run;


/* Treating the missing values of horsepower by impuation of mean value.
    Checking Mean Horsepower for various Cylinder categories */

Proc Means Data=project.Auto_Mpg;
    Class Cylinders;
    Var Horsepower;
Run;

/* Replacing missing horsepower with mean horespower grouped by Cylinders */
Proc Sort Data=project.Auto_Mpg;
    by Cylinders;
Run;


Proc Stdize data=project.Auto_Mpg out=project.Auto_Mpg reponly method=mean;
    by cylinders;
Run;

/* A new dervived variable Power-Weight Ratio */
Data project.Auto_Mpg;
    Set project.Auto_Mpg;
    PWR=horsepower/weight;
Run;
TITLE "Listing of Auto MPG(PWR-new derived variable)";
PROC PRINT DATA=project.auto_mpg(obs=5);
RUN;

/* Detecting outliers for numeric variables by using Standarad deviation method(Proc Univariate)
   and checking whether normally distributed*/
Proc Univariate Data=project.Auto_Mpg plots;
    Var mpg acceleration displacement weight horsepower;
Run;

/* After Checking we see variable Acceleration has normal distribution. Hence, we will use
Standard Deviation method to detect Outliers        */
Proc Means Data=project.Auto_Mpg noprint;
    Var Acceleration;
    Output out=Means (drop=_type_ _freq_) Mean=Std= / autoname;
Run;

Proc Means Data=project.Auto_Mpg noprint;
    Var pwr;
    Output out=IQR (drop=_type_ _freq_) Q1=Q3=Qrange= / autoname;
Run;

/* Detecting Outliers for Acceleration */
Title 'Listing Outliers for Acceleration';

Data _NULL_;
    Set project.Auto_Mpg (keep=Acceleration Brand Model);
    File Print;

    If _n_=1 then
        set Means;

    If Acceleration <=Acceleration_Mean - 2*Acceleration_StdDev or
        Acceleration > Acceleration_Mean + 2*Acceleration_StdDev then
            Put 'Outlier detected for ' Brand Model ' where Acceleration = '
            Acceleration;
Run;
```

```
Title 'Listing Outliers for Acceleration';

Data project.Auto_Mpg;
    Set project.Auto_Mpg;

    If _N_=1 then
        set means;

    If Acceleration < Acceleration_Mean - 2*Acceleration_StdDev or
        Acceleration > Acceleration_Mean + 2*Acceleration_StdDev then
            delete;
    Drop Acceleration_MEan Acceleration_StdDev;
Run;

Proc Univariate Data=project.Auto_Mpg plots;
    Var Acceleration;
Run;

/* Detecting Outliers for Power-Weight Ration using Inter Quartile Range */
Title 'Listing Outliers for Power-Weight Ratio';

Data _NULL_;
    Set project.Auto_Mpg (keep=pwr Brand Model);
    File Print;

    If _n_=1 then
        set IQR;

    If pwr < pwr_Q1 - 1.5*pwr_Qrange or pwr > pwr_Q3 + 1.5*pwr_Qrange then
        Put 'Outlier detected for ' Brand Model ' Power-Weight ratio = ' pwr;
Run;

Title;
Title 'Listing Outliers for Power-Weight Ratio';

Data project.Auto_Mpg;
    Set project.Auto_Mpg;

    If _n_=1 then
        set IQR;

    If pwr < pwr_Q1 - 1.5*pwr_Qrange or pwr > pwr_Q3 + 1.5*pwr then
        delete;
    Drop pwr_Q1 pwr_Q3 pwr_Qrange;
Run;

Title;

/* Checking Skewness of Variable Horsepower using QQplot and Histogram */
Title 'Histogram for Horsepower';

Proc sGplot Data=project.Auto_Mpg;
    Histogram horsepower;
    Density horsepower;
    Density horsepower / type=kernel;
Run;

Proc Gchart Data=project.Auto_Mpg;
    vbar horsepower;
    Run;
    Title 'QQ-Plot for Horsepower';

Proc Univariate Data=project.Auto_Mpg;
    Var horsepower;
    qqplot;
Run;

/* Applying Log10 transformation on Horsepower */
Data Log_test;
    Set project.Auto_Mpg;
    LogHP=Log(horsepower);
Run;

Title 'Histogram of Horsepower after Log Transformation';

Proc sGplot Data=log_test;
    Histogram loghp;
    Density loghp;
    Density loghp/ type=kernel;
Run;

Title 'QQ-Plot of Horsepower after Log Transformation';
```

```
Proc Univariate Data=log_test plots;
    Var Loghp;
Run;

Title 'Listing First 5 Observations from Final Dataset';

Proc Print Data=project.Auto_mpg (obs=5);
Run;

Data project.Auto_Mpg;
    Set project.Auto_Mpg;
    Label Brand='Brand of the Vehicle' Model='Model name of vehicle' Cylinders='Number of Cylinders. Categorical Variable
                  4, 6 or 8'
        Model_Year='The year in which the vehicle was manufactured' Origin='Country of Origin of the Vehicle Brand. Has t
                  Unites States  = 1
                  Germany =2
                  Japan = 3' MPG='City fuel cycle measured in miles/gallon'
        Displacement='Engine size of vehicle measured in cubic centimetres(CC)'
        Horsepower='Horsepower of the vehicle' Weight='Weight of vehicle in lbs'
        Acceleration='Time taken to reach from 0-60 mph'
        PWR='Power to weight ratio of vehicle measured as hp/lbs';
Run;
*   Test for normality using histogram and QQ plot for Target variable(MPG) Vs
    Independent variables(horsepower,weight,pwr,displacement) ;
options label;

Proc Contents Data=project.Auto_Mpg;
    ODS Select variables;
Run;

Proc sgplot data=project.Auto_mpg;
    histogram mpg;
    density mpg;
    density mpg / type=kernel;
Run;

Proc sgplot data=project.Auto_mpg;
    reg x=horsepower y=mpg / cli clm;
Run;

Proc sgplot data=project.Auto_mpg;
    reg x=weight y=mpg / cli clm;
Run;

Proc sgplot data=project.Auto_Mpg;
    reg x=pwr y=mpg / cli clm;
Run;

Proc sgplot data=project.Auto_Mpg;
```

Log: Final_Code.sas

Notes (118)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
71
72         /* Importing Auto_Mpg.data file and analyzing the variables dataset*/
73         LIBNAME PROJECT "/home/u61480438/BAN110/PROJECT";
NOTE: Libref PROJECT was successfully assigned as follows:
      Engine:        V9
      Physical Name: /home/u61480438/BAN110/PROJECT
74
75         Data PROJECT.AUTO_MPG;
76         INFILE '/home/u61480438/BAN110/PROJECT/auto-mpg.data';
77         INFORMAT Car_Name $30.;
78         INPUT @1 Mpg 4.
79           @8 Cylinders 1.
80           @12 Displacement 5.
81           @23 Horsepower 5.
82           @34 Weight 5.
83           @45 Acceleration 4.
84           @52 Model_Year 2.
85           @56 Origin 1.
86           @58 Car_Name & $30.;
87         FORMAT Mpg 4.1
88            Displacement 5.1
89            Horsepower 5.1
90            Weight 6.1
91            Acceleration 4.1;
92         Run;

NOTE: The infile '/home/u61480438/BAN110/PROJECT/auto-mpg.data' is:
      Filename=/home/u61480438/BAN110/PROJECT/auto-mpg.data,
```

```
      Owner Name=u61480438,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=11 August 2022 13:51:28,
      File Size (bytes)=30286

NOTE: Invalid data for Horsepower in line 33 23-27.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0

33  CHAR  25.0   4   98.00      ?          2046.      19.0   71  1."ford pinto" 69
    ZONE  33232223222333232222223222222222233332222223323222332230266762766762
    NUMR  25E0000400098E00000000F0000000000002046E00000019E000071001926F24009E4F2
Car_Name="ford pinto" Mpg=25.0 Cylinders=4 Displacement=98.0 Horsepower=. Weight=2046.0 Acceleration=19.0 Model_Year=71 Origin=1
_ERROR_=1 _N_=33
NOTE: Invalid data for Horsepower in line 127 23-27.

127 CHAR  21.0   6   200.0      ?          2875.      17.0   74  1."ford maverick" 72
    ZONE  33232223222333232222223222222222233332222223323222332230266762667676662
    NUMR  21E00006000200E0000000F0000000000002875E00000017E000074001926F240D165293B2
Car_Name="ford maverick" Mpg=21.0 Cylinders=6 Displacement=200.0 Horsepower=. Weight=2875.0 Acceleration=17.0 Model_Year=74 Origin=1
_ERROR_=1 _N_=127
NOTE: Invalid data for Horsepower in line 331 23-27.

331 CHAR  40.9   4   85.00      ?          1835.      17.3   80  2."renault lecar deluxe" 79
    ZONE  33232223222333232222223222222222233332222223323222332230276667672666672667762
    NUMR  40E9000400085E00000000F0000000000001835E00000017E3000800029225E15C40C5312045C5852
Car_Name="renault lecar deluxe" Mpg=40.9 Cylinders=4 Displacement=85.0 Horsepower=. Weight=1835.0 Acceleration=17.3 Model_Year=80
Origin=2 _ERROR_=1 _N_=331
NOTE: Invalid data for Horsepower in line 337 23-27.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0

337 CHAR  23.6   4   140.0      ?          2905.      14.3   80  1."ford mustang cobra" 77
    ZONE  33232223222333232222223222222222233332222223323222332230266762677662666762
    NUMR  23E60004000140E0000000F0000000000002905E00000014E300080001926F240D5341E703F2212
Car_Name="ford mustang cobra" Mpg=23.6 Cylinders=4 Displacement=140.0 Horsepower=. Weight=2905.0 Acceleration=14.3 Model_Year=80
Origin=1 _ERROR_=1 _N_=337
NOTE: Invalid data for Horsepower in line 355 23-27.

355 CHAR  34.5   4   100.0      ?          2320.      15.8   81  2."renault 18i" 70
    ZONE  33232223222333232222223222222222233332222223323222332230276667672 3362
    NUMR  34E50004000100E0000000F0000000000002320E00000015E8000810029225E15C401892
Car_Name="renault 18i" Mpg=34.5 Cylinders=4 Displacement=100.0 Horsepower=. Weight=2320.0 Acceleration=15.8 Model_Year=81 Origin=2
_ERROR_=1 _N_=355
NOTE: Invalid data for Horsepower in line 375 23-27.

375 CHAR  23.0   4   151.0      ?          3035.      20.5   82  1."amc concord dl" 73
    ZONE  33232223222333232222223222222222233332222223323222332230266266666762662
    NUMR  23E00004000151E0000000F0000000000003035E00000020E500082001921D303FE3F2404C2
Car_Name="amc concord dl" Mpg=23.0 Cylinders=4 Displacement=151.0 Horsepower=. Weight=3035.0 Acceleration=20.5 Model_Year=82
Origin=1 _ERROR_=1 _N_=375
NOTE: 398 records were read from the infile '/home/u61480438/BAN110/PROJECT/auto-mpg.data'.
      The minimum record length was 65.
      The maximum record length was 95.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.05 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              769.46k
      OS Memory           29096.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        54  Switch Count  1
      Page Faults                       0
      Page Reclaims                     184
      Page Swaps                        0
      Voluntary Context Switches        40
      Involuntary Context Switches      0
      Block Input Operations            96
      Block Output Operations           272


93
94         TITLE "Analyzing the Auto MPG data";
95         PROC CONTENTS DATA=PROJECT.AUTO_MPG VARNUM;
96         RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.06 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              3057.87k
      OS Memory           29868.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        55  Switch Count  0
      Page Faults                       0
      Page Reclaims                     232
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           16
```

```
97
98          /* Printing the first ten observations */
99          TITLE 'First ten observations of the raw Auto_Mpg dataset';
100
101         PROC PRINT DATA=PROJECT.AUTO_MPG (OBS=10);
102         RUN;

NOTE: There were 10 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              850.46k
      OS Memory           29864.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        56  Switch Count  0
      Page Faults                       0
      Page Reclaims                     87
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           24


103
104         /* Descriptive statistics  of Dependent/Target variable MPG*/
105         Title 'Descrpitive Statitics for Dependent variable MPG';
106
107         Proc Means Data=project.Auto_Mpg;
108         Var mpg;
109         Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              6867.12k
      OS Memory           36044.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        57  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1740
      Page Swaps                        0
      Voluntary Context Switches        26
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


110
111         /* Histogram of Dependent/Target variable MPG */
112         Title 'Histogram of MPG';
113
114         Proc Sgplot Data=project.Auto_Mpg;
115         Histogram Mpg;
116         Density Mpg;
117         Density Mpg / type=kernel;
118         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.29 seconds
      user cpu time       0.07 seconds
      system cpu time     0.02 seconds
      memory              14607.90k
      OS Memory           42412.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        58  Switch Count  1
      Page Faults                       0
      Page Reclaims                     3689
      Page Swaps                        0
      Voluntary Context Switches        269
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           960

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.

119
120         /* Working with Categorical Values*/
121         options nolabel;
122         Title 'Listing Frequencies for Cylinders Model_year and Origin';
123
124         Proc Freq Data=project.Auto_Mpg;
125         Tables Cylinders Model_Year Origin / nocum missing;
126         Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.03 seconds
      user cpu time       0.04 seconds
```

```
      system cpu time     0.00 seconds
      memory              909.84k
      OS Memory           42928.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        59  Switch Count  2
      Page Faults                       0
      Page Reclaims                     315
      Page Swaps                        0
      Voluntary Context Switches        16
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


127
128        /* Checking for missing values in categorical variables by using informat method*/
129        Proc Format;
130
130      !  Value Origin_Check 1, 2, 3='Valid' other='Invalid';
NOTE: Format ORIGIN_CHECK has been output.
131
131      !  Value Cyl_Check 3, 4, 5, 6, 8='Valid' other='Invalid';
NOTE: Format CYL_CHECK has been output.
132
132      !  Value Year_Check 70-82='Valid' other='Invalid';
NOTE: Format YEAR_CHECK has been output.
133        Run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              302.62k
      OS Memory           42664.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        60  Switch Count  2
      Page Faults                       0
      Page Reclaims                     63
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           64


134
135        Data _null_;
136        File Print;
137        Set Project.Auto_Mpg (Keep=Car_Name Cylinders Model_Year Origin);
138
139        If put(Cylinders, Cyl_Check.)='Invalid' then
140        put 'Missing observation of Cylinders = ' _n_ Car_name Cylinders=;
141        Else if Put(Model_Year, Year_Check.)='Invalid' then
142        put 'Missing observation of Model_Year = ' _n_ Car_name Model_Year=;
143        Else if put(Origin, Origin_Check.)='Invalid' then
144        put 'Missing observation of Origin = ' _n_ Car_name Origin=;
145        Run;

NOTE: 0 lines were written to file PRINT.
NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
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              799.09k
      OS Memory           43180.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        61  Switch Count  1
      Page Faults                       0
      Page Reclaims                     90
      Page Swaps                        0
      Voluntary Context Switches        8
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


146
147        Title 'Checking for Missing values in Categorical variables';
148
149        Proc Freq Data=project.Auto_Mpg;
150        Tables Cylinders Model_Year Origin / nocum nopercent;
151        Format Cylinders Cyl_Check. Model_Year Year_Check. Origin Origin_Check.;
152        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              913.96k
      OS Memory           43440.00k
```

```
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        62  Switch Count  2
      Page Faults                       0
      Page Reclaims                     159
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


153
154        /* Converting Date from 2 digit number to Date9. format for full year*/
155        Data project.Auto_Mpg;
156        Set project.Auto_Mpg;
157        Year_new=Cat('03/01/19', Model_Year);
158        Model_Year=year(input(Year_new, mmddyy10.));
159        Drop Year_new;
160        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              960.37k
      OS Memory           43440.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        63  Switch Count  1
      Page Faults                       0
      Page Reclaims                     158
      Page Swaps                        0
      Voluntary Context Switches        40
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


161
162        Title 'Printing first 5 observations after date conversion';
163
164        Proc Print Data=project.Auto_Mpg (obs=5);
165        Var Car_Name Model_Year;
166        Run;

NOTE: There were 5 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              636.78k
      OS Memory           43180.00k
      Timestamp           21/03/2023 03:46:01 AM
      Step Count                        64  Switch Count  0
      Page Faults                       0
      Page Reclaims                     66
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           0


167
168        /*Deriving Vehicle brand name and Model from car_Name variable and drop the car_name*/
169        Data project.Auto_Mpg;
170        Set project.Auto_Mpg;
171        Car_Name=Propcase(Compress(car_Name, '"'));
172        Array model_n [6] $20. Model1-Model6;
173
174        Do i=1 to 6;
175        Model_n [i]=compress(Scan(Car_Name, i), "'");
176        End;
177
178        If _n_=293 then
179        Model3='';
180        Brand=Model1;
181        Model=Catx('', Model2, model3, model4, model5, model6);
182        Drop Model1-Model6 Car_Name i;
183        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              988.65k
      OS Memory           43440.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        65  Switch Count  1
```

```
      Page Faults                       0
      Page Reclaims                     152
      Page Swaps                        0
      Voluntary Context Switches        43
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


184        PROC PRINT DATA=project.auto_mpg (obs=5);
185        RUN;

NOTE: There were 5 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              757.28k
      OS Memory           43180.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        66  Switch Count  0
      Page Faults                       0
      Page Reclaims                     69
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      1
      Block Input Operations            288
      Block Output Operations           24


186
187        /*Checking Errors in Brand Variable*/
188        Title 'Checking errors in Brand variable';
189
190        Proc Freq Data=project.Auto_Mpg;
191        Tables Brand / nocum nopercent;
192        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              823.00k
      OS Memory           43440.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        67  Switch Count  2
      Page Faults                       0
      Page Reclaims                     122
      Page Swaps                        0
      Voluntary Context Switches        19
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


193
194        /*Checking and Correcting Spelling errors in Brand variable */
195        Data project.Auto_Mpg;
196        Set project.Auto_Mpg;
197        Brand=Tranwrd(Brand, 'Chevy', 'Chevrolet');
198        Brand=Tranwrd(Brand, 'Chevroelt', 'Chevrolet');
199        Brand=Tranwrd(Brand, 'Hi', 'Honda');
200        Brand=Tranwrd(Brand, 'Maxda', 'Mazda');
201        Brand=Tranwrd(Brand, 'Vw', 'Volkswagen');
202        Brand=Tranwrd(Brand, 'Vokswagen', 'Volkswagen');
203        Brand=Tranwrd(Brand, 'Toyouta', 'Toyota');
204        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              1016.34k
      OS Memory           43440.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        68  Switch Count  1
      Page Faults                       0
      Page Reclaims                     119
      Page Swaps                        0
      Voluntary Context Switches        37
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


205
206        Title 'Corrected Brand variable';
207
208        Proc Freq Data=project.Auto_Mpg;
```

```
209         Tables Brand / nocum nopercent;
210         Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              822.68k
      OS Memory           43440.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        69  Switch Count  2
      Page Faults                       0
      Page Reclaims                     123
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           288


211
212         /*Working with Numerical Variables*/
213         options nolabel;
214
215         Proc MEans Data=project.Auto_Mpg n nmiss min max mean median mode stddev var;
216         Var mpg acceleration displacement weight horsepower;
217         Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.05 seconds
      user cpu time       0.05 seconds
      system cpu time     0.01 seconds
      memory              6521.65k
      OS Memory           48576.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        70  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1447
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


218
219         /* Checking Missing Numeric Observations */
220         Title 'Identifying Missing numeric values';
221
222         Data _null_;
223         File print;
224         Set project.Auto_Mpg;
225         Array Numeric [*] _NUMERIC_;
226
227         Do i=1 to Dim(Numeric);
228
229         If missing(numeric(i)) then
230         put 'Missing Observation '
231         Brand=Model=Mpg=Cylinders=Displacement=Horsepower=Weight=Acceleration=;
232         End;
233         Run;

NOTE: 8 lines were written to file PRINT.
NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              887.25k
      OS Memory           43436.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        71  Switch Count  0
      Page Faults                       0
      Page Reclaims                     94
      Page Swaps                        0
      Voluntary Context Switches        2
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


234
235
236         /* Treating the missing values of horsepower by impuation of mean value.
237         Checking Mean Horsepower for various Cylinder categories */
238
239         Proc Means Data=project.Auto_Mpg;
240         Class Cylinders;
241         Var Horsepower;
242         Run;
```

```
NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              8916.01k
      OS Memory           51904.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        72  Switch Count  1
      Page Faults                       0
      Page Reclaims                     2044
      Page Swaps                        0
      Voluntary Context Switches        17
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           16


243
244        /* Replacing missing horsepower with mean horespower grouped by Cylinders */
245        Proc Sort Data=project.Auto_Mpg;
246        by Cylinders;
247        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              914.28k
      OS Memory           43952.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        73  Switch Count  1
      Page Faults                       0
      Page Reclaims                     144
      Page Swaps                        0
      Voluntary Context Switches        39
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


248
249
250        Proc Stdize data=project.Auto_Mpg out=project.Auto_Mpg reponly method=mean;
251        by cylinders;
252        Run;

NOTE: No VAR statement is given. All numerical variables not named elsewhere make up the first set of variables.
NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 10 variables.
NOTE: PROCEDURE STDIZE used (Total process time):
      real time           0.02 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              930.34k
      OS Memory           44464.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        74  Switch Count  1
      Page Faults                       0
      Page Reclaims                     546
      Page Swaps                        0
      Voluntary Context Switches        47
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           264


253
254        /* A new dervived variable Power-Weight Ratio */
255        Data project.Auto_Mpg;
256        Set project.Auto_Mpg;
257        PWR=horsepower/weight;
258        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 398 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              943.15k
      OS Memory           44464.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        75  Switch Count  1
      Page Faults                       0
      Page Reclaims                     101
      Page Swaps                        0
      Voluntary Context Switches        48
      Involuntary Context Switches      0
```

```
      Block Input Operations            288
      Block Output Operations           264


259        TITLE "Listing of Auto MPG(PWR-new derived variable)";
260        PROC PRINT DATA=project.auto_mpg(obs=5);
261        RUN;

NOTE: There were 5 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              710.46k
      OS Memory           44204.00k
      Timestamp           21/03/2023 03:46:02 AM
      Step Count                        76  Switch Count  0
      Page Faults                       0
      Page Reclaims                     70
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           0


262
263        /* Detecting outliers for numeric variables by using Standarad deviation method(Proc Univariate)
264           and checking whether normally distributed*/
265        Proc Univariate Data=project.Auto_Mpg plots;
266        Var mpg acceleration displacement weight horsepower;
267        Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           1.61 seconds
      user cpu time       0.68 seconds
      system cpu time     0.03 seconds
      memory              4526.26k
      OS Memory           46432.00k
      Timestamp           21/03/2023 03:46:03 AM
      Step Count                        77  Switch Count  0
      Page Faults                       0
      Page Reclaims                     2052
      Page Swaps                        0
      Voluntary Context Switches        1262
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           2648


268
269        /* After Checking we see variable Acceleration has normal distribution. Hence, we will use
270        Standard Deviation method to detect Outliers*/
271        Proc Means Data=project.Auto_Mpg noprint;
272        Var Acceleration;
273        Output out=Means (drop=_type_ _freq_) Mean=Std= / autoname;
274        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set WORK.MEANS has 1 observations and 2 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              7103.03k
      OS Memory           51668.00k
      Timestamp           21/03/2023 03:46:03 AM
      Step Count                        78  Switch Count  3
      Page Faults                       0
      Page Reclaims                     1682
      Page Swaps                        0
      Voluntary Context Switches        33
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


275
276        Proc Means Data=project.Auto_Mpg noprint;
277        Var pwr;
278        Output out=IQR (drop=_type_ _freq_) Q1=Q3=Qrange= / autoname;
279        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set WORK.IQR has 1 observations and 3 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              7161.15k
      OS Memory           51668.00k
      Timestamp           21/03/2023 03:46:03 AM
```

```
      Step Count                        79  Switch Count  3
      Page Faults                       0
      Page Reclaims                     1679
      Page Swaps                        0
      Voluntary Context Switches        32
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


280
281        /* Detecting Outliers for Acceleration */
282        Title 'Listing Outliers for Acceleration';
283
284        Data _NULL_;
285        Set project.Auto_Mpg (keep=Acceleration Brand Model);
286        File Print;
287
288        If _n_=1 then
289        set Means;
290
291        If Acceleration <=Acceleration_Mean - 2*Acceleration_StdDev or
292        Acceleration > Acceleration_Mean + 2*Acceleration_StdDev then
293        Put 'Outlier detected for ' Brand Model ' where Acceleration = '
294        Acceleration;
295        Run;

NOTE: 21 lines were written to file PRINT.
NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: There were 1 observations read from the data set WORK.MEANS.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1205.56k
      OS Memory           45488.00k
      Timestamp           21/03/2023 03:46:03 AM
      Step Count                        80  Switch Count  0
      Page Faults                       0
      Page Reclaims                     105
      Page Swaps                        0
      Voluntary Context Switches        1
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           40


296
297        Title 'Listing Outliers for Acceleration';
298
299        Data project.Auto_Mpg;
300        Set project.Auto_Mpg;
301
302        If _N_=1 then
303        set means;
304
305        If Acceleration < Acceleration_Mean - 2*Acceleration_StdDev or
306        Acceleration > Acceleration_Mean + 2*Acceleration_StdDev then
307        delete;
308        Drop Acceleration_MEan Acceleration_StdDev;
309        Run;

NOTE: There were 398 observations read from the data set PROJECT.AUTO_MPG.
NOTE: There were 1 observations read from the data set WORK.MEANS.
NOTE: The data set PROJECT.AUTO_MPG has 377 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1293.15k
      OS Memory           45748.00k
      Timestamp           21/03/2023 03:46:03 AM
      Step Count                        81  Switch Count  1
      Page Faults                       0
      Page Reclaims                     139
      Page Swaps                        0
      Voluntary Context Switches        38
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


310
311        Proc Univariate Data=project.Auto_Mpg plots;
312        Var Acceleration;
313        Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.26 seconds
      user cpu time       0.11 seconds
      system cpu time     0.01 seconds
      memory              3356.40k
```

```
      OS Memory           46432.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        82  Switch Count  0
      Page Faults                       0
      Page Reclaims                     428
      Page Swaps                        0
      Voluntary Context Switches        282
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           488


314
315        /* Detecting Outliers for Power-Weight Ration using Inter Quartile Range */
316        Title 'Listing Outliers for Power-Weight Ratio';
317
318        Data _NULL_;
319        Set project.Auto_Mpg (keep=pwr Brand Model);
320        File Print;
321
322        If _n_=1 then
323        set IQR;
324
325        If pwr < pwr_Q1 - 1.5*pwr_Qrange or pwr > pwr_Q3 + 1.5*pwr_Qrange then
326        Put 'Outlier detected for ' Brand Model ' Power-Weight ratio = ' pwr;
327        Run;

NOTE: 2 lines were written to file PRINT.
NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.
NOTE: There were 1 observations read from the data set WORK.IQR.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              1023.00k
      OS Memory           45488.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        83  Switch Count  0
      Page Faults                       0
      Page Reclaims                     97
      Page Swaps                        0
      Voluntary Context Switches        5
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           48


328
329        Title;
330        Title 'Listing Outliers for Power-Weight Ratio';
331
332        Data project.Auto_Mpg;
333        Set project.Auto_Mpg;
334
335        If _n_=1 then
336        set IQR;
337
338        If pwr < pwr_Q1 - 1.5*pwr_Qrange or pwr > pwr_Q3 + 1.5*pwr then
339        delete;
340        Drop pwr_Q1 pwr_Q3 pwr_Qrange;
341        Run;

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.
NOTE: There were 1 observations read from the data set WORK.IQR.
NOTE: The data set PROJECT.AUTO_MPG has 377 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1299.12k
      OS Memory           45748.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        84  Switch Count  1
      Page Faults                       0
      Page Reclaims                     138
      Page Swaps                        0
      Voluntary Context Switches        38
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


342
343        Title;
344
345        /* Checking Skewness of Variable Horsepower using QQplot and Histogram */
346        Title 'Histogram for Horsepower';
347
348        Proc sGplot Data=project.Auto_Mpg;
349        Histogram horsepower;
350        Density horsepower;
351        Density horsepower / type=kernel;
```

```
352        Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.14 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              2162.06k
      OS Memory           46132.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        85  Switch Count  1
      Page Faults                       0
      Page Reclaims                     358
      Page Swaps                        0
      Voluntary Context Switches        263
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           552

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

353
354        Proc Gchart Data=project.Auto_Mpg;
355        vbar horsepower;
356        Run;

357        Title 'QQ-Plot for Horsepower';
358

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE GCHART used (Total process time):
      real time           0.17 seconds
      user cpu time       0.16 seconds
      system cpu time     0.02 seconds
      memory              7400.43k
      OS Memory           49948.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        86  Switch Count  5
      Page Faults                       0
      Page Reclaims                     1777
      Page Swaps                        0
      Voluntary Context Switches        53
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           288


359        Proc Univariate Data=project.Auto_Mpg;
360        Var horsepower;
361        qqplot;
362        Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.20 seconds
      user cpu time       0.12 seconds
      system cpu time     0.01 seconds
      memory              8114.87k
      OS Memory           53264.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        87  Switch Count  0
      Page Faults                       0
      Page Reclaims                     1822
      Page Swaps                        0
      Voluntary Context Switches        184
      Involuntary Context Switches      8
      Block Input Operations            0
      Block Output Operations           440


363
364        /* Applying Log10 transformation on Horsepower */
365        Data Log_test;
366        Set project.Auto_Mpg;
367        LogHP=Log(horsepower);
368        Run;

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set WORK.LOG_TEST has 377 observations and 12 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              978.31k
      OS Memory           47536.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        88  Switch Count  2
      Page Faults                       0
      Page Reclaims                     112
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264
```

```
369
370         Title 'Histogram of Horsepower after Log Transformation';
371
372         Proc sGplot Data=log_test;
373         Histogram loghp;
374         Density loghp;
375         Density loghp/ type=kernel;
376         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.15 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2047.03k
      OS Memory           48180.00k
      Timestamp           21/03/2023 03:46:04 AM
      Step Count                        89  Switch Count  1
      Page Faults                       0
      Page Reclaims                     368
      Page Swaps                        0
      Voluntary Context Switches        252
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           592

NOTE: There were 377 observations read from the data set WORK.LOG_TEST.

377
378         Title 'QQ-Plot of Horsepower after Log Transformation';
379
380         Proc Univariate Data=log_test plots;
381         Var Loghp;
382         Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.25 seconds
      user cpu time       0.13 seconds
      system cpu time     0.01 seconds
      memory              3550.03k
      OS Memory           48480.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        90  Switch Count  0
      Page Faults                       0
      Page Reclaims                     377
      Page Swaps                        0
      Voluntary Context Switches        268
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           496


383
384         Title 'Listing First 5 Observations from Final Dataset';
385
386         Proc Print Data=project.Auto_mpg (obs=5);
387         Run;

NOTE: There were 5 observations read from the data set PROJECT.AUTO_MPG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              898.43k
      OS Memory           47276.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        91  Switch Count  0
      Page Faults                       0
      Page Reclaims                     75
      Page Swaps                        0
      Voluntary Context Switches        2
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           48


388
389         Data project.Auto_Mpg;
390         Set project.Auto_Mpg;
391         Label Brand='Brand of the Vehicle' Model='Model name of vehicle' Cylinders='Number of Cylinders. Categorical Variable
391       ! which can take following values:
392         4, 6 or 8'
393         Model_Year='The year in which the vehicle was manufactured' Origin='Country of Origin of the Vehicle Brand. Has the
393       ! following categories:
394         Unites States  = 1
395         Germany =2
396         Japan = 3' MPG='City fuel cycle measured in miles/gallon'
397         Displacement='Engine size of vehicle measured in cubic centimetres(CC)'
398         Horsepower='Horsepower of the vehicle' Weight='Weight of vehicle in lbs'
399         Acceleration='Time taken to reach from 0-60 mph'
400         PWR='Power to weight ratio of vehicle measured as hp/lbs';
```

```
401        Run;

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.
NOTE: The data set PROJECT.AUTO_MPG has 377 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              948.50k
      OS Memory           47536.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        92  Switch Count  1
      Page Faults                       0
      Page Reclaims                     98
      Page Swaps                        0
      Voluntary Context Switches        37
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


402        *Test for normality using histogram and QQ plot for Target variable(MPG) Vs
403        Independent variables(horsepower,weight,pwr,displacement) ;
404        options label;
405
406        Proc Contents Data=project.Auto_Mpg;
407        ODS Select variables;
408        Run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              950.81k
      OS Memory           47536.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        93  Switch Count  1
      Page Faults                       0
      Page Reclaims                     96
      Page Swaps                        0
      Voluntary Context Switches        20
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           16


409
410        Proc sgplot data=project.Auto_mpg;
411        histogram mpg;
412        density mpg;
413        density mpg / type=kernel;
414        Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.14 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              2263.40k
      OS Memory           48180.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        94  Switch Count  1
      Page Faults                       0
      Page Reclaims                     401
      Page Swaps                        0
      Voluntary Context Switches        257
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           568

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

415
416        Proc sgplot data=project.Auto_mpg;
417        reg x=horsepower y=mpg / cli clm;
418        Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.25 seconds
      user cpu time       0.08 seconds
      system cpu time     0.02 seconds
      memory              4615.46k
      OS Memory           50604.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        95  Switch Count  1
      Page Faults                       0
      Page Reclaims                     910
      Page Swaps                        0
      Voluntary Context Switches        270
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           752
```

```
NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

419
420         Proc sgplot data=project.Auto_mpg;
421         reg x=weight y=mpg / cli clm;
422         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.19 seconds
      user cpu time       0.09 seconds
      system cpu time     0.00 seconds
      memory              5170.43k
      OS Memory           52652.00k
      Timestamp           21/03/2023 03:46:05 AM
      Step Count                        96  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1118
      Page Swaps                        0
      Voluntary Context Switches        269
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           824

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

423
424         Proc sgplot data=project.Auto_Mpg;
425         reg x=pwr y=mpg / cli clm;
426         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.19 seconds
      user cpu time       0.07 seconds
      system cpu time     0.01 seconds
      memory              4613.56k
      OS Memory           52396.00k
      Timestamp           21/03/2023 03:46:06 AM
      Step Count                        97  Switch Count  1
      Page Faults                       0
      Page Reclaims                     926
      Page Swaps                        0
      Voluntary Context Switches        270
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           832

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

427
428         Proc sgplot data=project.Auto_Mpg;
429         reg x=displacement y=mpg / cli clm;
430         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.21 seconds
      user cpu time       0.07 seconds
      system cpu time     0.01 seconds
      memory              3275.12k
      OS Memory           50348.00k
      Timestamp           21/03/2023 03:46:06 AM
      Step Count                        98  Switch Count  1
      Page Faults                       0
      Page Reclaims                     535
      Page Swaps                        0
      Voluntary Context Switches        272
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           776

NOTE: There were 377 observations read from the data set PROJECT.AUTO_MPG.

431
432         Proc Univariate Data=project.Auto_mpg plots;
433         Var mpg;
434         Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.24 seconds
      user cpu time       0.12 seconds
      system cpu time     0.01 seconds
      memory              3192.03k
      OS Memory           50272.00k
      Timestamp           21/03/2023 03:46:06 AM
      Step Count                        99  Switch Count  0
      Page Faults                       0
      Page Reclaims                     416
      Page Swaps                        0
      Voluntary Context Switches        276
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           552
```

```
435
436         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
447
```

Results: Final_Code.sas

Analyzing the Auto MPG data

The CONTENTS Procedure

Data Set Name	PROJECT.AUTO_MPG	**Observations**	398
Member Type	DATA	**Variables**	9
Engine	V9	**Indexes**	0
Created	20/03/2023 23:46:01	**Observation Length**	96
Last Modified	20/03/2023 23:46:01	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1363
Obs in First Data Page	398
Number of Data Set Repairs	0
Filename	/home/u61480438/BAN110/PROJECT/auto_mpg.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	4856950187
Access Permission	rw-r--r--
Owner Name	u61480438
File Size	256KB
File Size (bytes)	262144

Variables in Creation Order

#	Variable	Type	Len	Format	Informat
1	Car_Name	Char	30		$30.
2	Mpg	Num	8	4.1	
3	Cylinders	Num	8		
4	Displacement	Num	8	5.1	
5	Horsepower	Num	8	5.1	
6	Weight	Num	8	6.1	
7	Acceleration	Num	8	4.1	
8	Model_Year	Num	8		
9	Origin	Num	8		

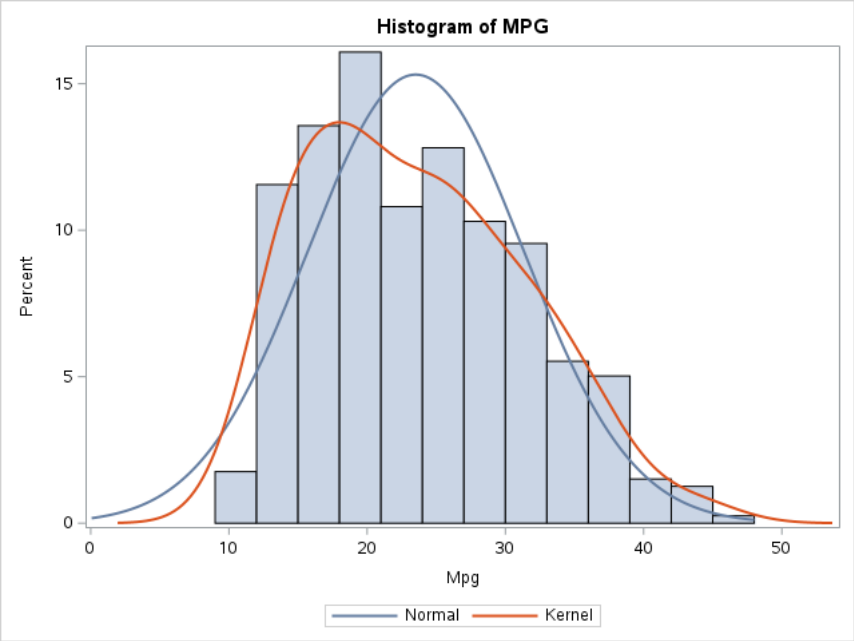
First ten observations of the raw Auto_Mpg dataset

Obs	Car_Name	Mpg	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model_Year	Origin
1	"chevrolet chevelle malibu"	18.0	8	307.0	130.0	3504.0	12.0	70	1
2	"buick skylark 320"	15.0	8	350.0	165.0	3693.0	11.5	70	1
3	"plymouth satellite"	18.0	8	318.0	150.0	3436.0	11.0	70	1
4	"amc rebel sst"	16.0	8	304.0	150.0	3433.0	12.0	70	1
5	"ford torino"	17.0	8	302.0	140.0	3449.0	10.5	70	1
6	"ford galaxie 500"	15.0	8	429.0	198.0	4341.0	10.0	70	1
7	"chevrolet impala"	14.0	8	454.0	220.0	4354.0	9.0	70	1
8	"plymouth fury iii"	14.0	8	440.0	215.0	4312.0	8.5	70	1
9	"pontiac catalina"	14.0	8	455.0	225.0	4425.0	10.0	70	1
10	"amc ambassador dpl"	15.0	8	390.0	190.0	3850.0	8.5	70	1

Descrpitive Statitics for Dependent variable MPG

The MEANS Procedure

Analysis Variable : Mpg				
N	Mean	Std Dev	Minimum	Maximum
398	23.5145729	7.8159843	9.0000000	46.6000000



Listing Frequencies for Cylinders Model_year and Origin

The FREQ Procedure

Cylinders	Frequency	Percent
3	4	1.01
4	204	51.26
5	3	0.75
6	84	21.11
8	103	25.88

Model_Year	Frequency	Percent
70	29	7.29
71	28	7.04
72	28	7.04
73	40	10.05
74	27	6.78
75	30	7.54
76	34	8.54
77	28	7.04
78	36	9.05
79	29	7.29
80	29	7.29
81	29	7.29
82	31	7.79

Origin	Frequency	Percent
1	249	62.56
2	70	17.59
3	79	19.85

Checking for Missing values in Categorical variables

The FREQ Procedure

Cylinders	Frequency
Valid	398

Model_Year	Frequency
Valid	398

Origin	Frequency
Valid	398

Printing first 5 observations after date conversion

Obs	Car_Name	Model_Year
1	"chevrolet chevelle malibu"	1970
2	"buick skylark 320"	1970
3	"plymouth satellite"	1970
4	"amc rebel sst"	1970

Obs	Car_Name	Model_Year
5	"ford torino"	1970

Printing first 5 observations after date conversion

Obs	Mpg	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model_Year	Origin	Brand	Model
1	18.0	8	307.0	130.0	3504.0	12.0	1970	1	Chevrolet	Chevelle Malibu
2	15.0	8	350.0	165.0	3693.0	11.5	1970	1	Buick	Skylark 320
3	18.0	8	318.0	150.0	3436.0	11.0	1970	1	Plymouth	Satellite
4	16.0	8	304.0	150.0	3433.0	12.0	1970	1	Amc	Rebel Sst
5	17.0	8	302.0	140.0	3449.0	10.5	1970	1	Ford	Torino

Checking errors in Brand variable

The FREQ Procedure

Brand	Frequency
Amc	28
Audi	7
Bmw	2
Buick	17
Cadillac	2
Capri	1
Chevroelt	1
Chevrolet	43
Chevy	3
Chrysler	6
Datsun	23
Dodge	28
Fiat	8
Ford	51
Hi	1
Honda	13
Maxda	2
Mazda	10
Mercedes	3
Mercury	11
Nissan	1
Oldsmobile	10
Opel	4
Peugeot	8
Plymouth	31
Pontiac	16
Renault	5
Saab	4
Subaru	4
Toyota	25
Toyouta	1
Triumph	1
Vokswagen	1
Volkswagen	15
Volvo	6
Vw	6

Corrected Brand variable

The FREQ Procedure

Brand	Frequency
Amc	28
Audi	7
Bmw	2
Buick	17
Cadillac	2
Capri	1
Chevrolet	47
Chrysler	6
Datsun	23
Dodge	28
Fiat	8
Ford	51
Honda	14
Mazda	12
Mercedes	3
Mercury	11
Nissan	1
Oldsmobile	10
Opel	4
Peugeot	8
Plymouth	31
Pontiac	16
Renault	5
Saab	4

Brand	Frequency
Subaru	4
Toyota	26
Triumph	1
Volkswagen	22
Volvo	6

Corrected Brand variable

The MEANS Procedure

Variable	N	N Miss	Minimum	Maximum	Mean	Median	Mode	Std Dev	Variance
Mpg	398	0	9.0000000	46.6000000	23.5145729	23.0000000	13.0000000	7.8159843	61.0896108
Acceleration	398	0	8.0000000	24.8000000	15.5680905	15.5000000	14.5000000	2.7576889	7.6048482
Displacement	398	0	68.0000000	455.0000000	193.4258794	148.5000000	97.0000000	104.2698382	10872.20
Weight	398	0	1613.00	5140.00	2970.42	2803.50	1985.00	846.8417742	717140.99
Horsepower	392	6	46.0000000	230.0000000	104.4693878	93.5000000	150.0000000	38.4911599	1481.57

Identifying Missing numeric values

```
Missing Observation Brand=Ford Model=Pinto Mpg=25.0 Cylinders=4 Displacement=98.0 Horsepower=. Weight=2046.0 Acceleration=19.0
Missing Observation Brand=Ford Model=Maverick Mpg=21.0 Cylinders=6 Displacement=200.0 Horsepower=. Weight=2875.0 Acceleration=17.0
Missing Observation Brand=Renault Model=Lecar Deluxe Mpg=40.9 Cylinders=4 Displacement=85.0 Horsepower=. Weight=1835.0
Acceleration=17.3
Missing Observation Brand=Ford Model=Mustang Cobra Mpg=23.6 Cylinders=4 Displacement=140.0 Horsepower=. Weight=2905.0
Acceleration=14.3
Missing Observation Brand=Renault Model=18i Mpg=34.5 Cylinders=4 Displacement=100.0 Horsepower=. Weight=2320.0 Acceleration=15.8
Missing Observation Brand=Amc Model=Concord Dl Mpg=23.0 Cylinders=4 Displacement=151.0 Horsepower=. Weight=3035.0 Acceleration=20.5
```

Identifying Missing numeric values

The MEANS Procedure

Analysis Variable : Horsepower						
Cylinders	N Obs	N	Mean	Std Dev	Minimum	Maximum
3	4	4	99.2500000	8.3016063	90.0000000	110.0000000
4	204	199	78.2814070	14.5230992	46.0000000	115.0000000
5	3	3	82.3333333	18.5831465	67.0000000	103.0000000
6	84	83	101.5060241	14.3104716	72.0000000	165.0000000
8	103	103	158.3009709	28.4535517	90.0000000	230.0000000

Listing of Auto MPG(PWR-new derived variable)

Obs	Mpg	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model_Year	Origin	Brand	Model	PWR
1	19.0	3	70.0	97.0	2330.0	13.5	1972	3	Mazda	Rx2 Coupe	0.041631
2	18.0	3	70.0	90.0	2124.0	13.5	1973	3	Mazda	Rx3	0.042373
3	21.5	3	80.0	110.0	2720.0	13.5	1977	3	Mazda	Rx 4	0.040441
4	23.7	3	70.0	100.0	2420.0	12.5	1980	3	Mazda	Rx 7 Gs	0.041322
5	24.0	4	113.0	95.0	2372.0	15.0	1970	3	Toyota	Corona Mark Ii	0.040051

Listing of Auto MPG(PWR-new derived variable)

The UNIVARIATE Procedure
Variable: Mpg

Moments			
N	398	Sum Weights	398
Mean	23.5145729	Sum Observations	9358.8
Std Deviation	7.81598431	Variance	61.0896108
Skewness	0.45706634	Kurtosis	-0.5107813
Uncorrected SS	244320.76	Corrected SS	24252.5755
Coeff Variation	33.2388955	Std Error Mean	0.39177989

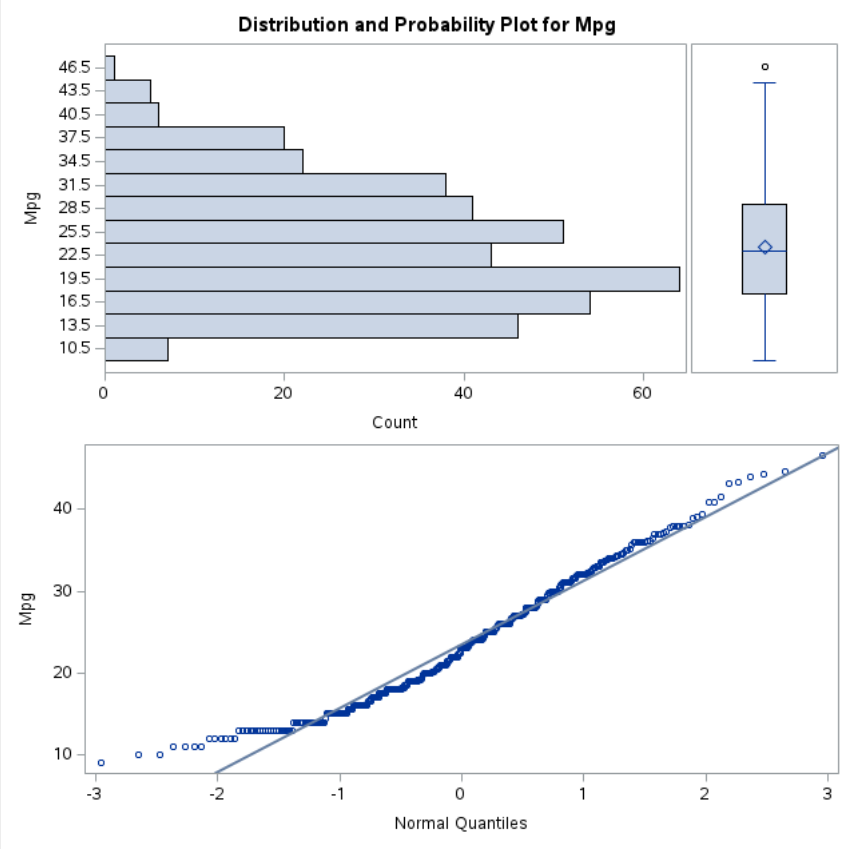
Basic Statistical Measures			
Location		Variability	
Mean	23.51457	Std Deviation	7.81598
Median	23.00000	Variance	61.08961
Mode	13.00000	Range	37.60000
		Interquartile Range	11.50000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	60.01986	Pr > |t|	<.0001
Sign	M	199	Pr >= |M|	<.0001
Signed Rank	S	39700.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	46.6
99%	44.0
95%	37.2
90%	34.4

Quantiles (Definition 5)	
Level	Quantile
75% Q3	29.0
50% Median	23.0
25% Q1	17.5
10%	14.0
5%	13.0
1%	11.0
0% Min	9.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
9	313	43.4	151
10	311	44.0	205
10	310	44.3	150
11	353	44.6	153
11	346	46.6	147



Listing of Auto MPG(PWR-new derived variable)

The UNIVARIATE Procedure
Variable: Acceleration

Moments			
N	398	Sum Weights	398
Mean	15.5680905	Sum Observations	6196.1
Std Deviation	2.75768893	Variance	7.60484823
Skewness	0.27877684	Kurtosis	0.41949688
Uncorrected SS	99480.57	Corrected SS	3019.12475
Coeff Variation	17.7137263	Std Error Mean	0.13823046

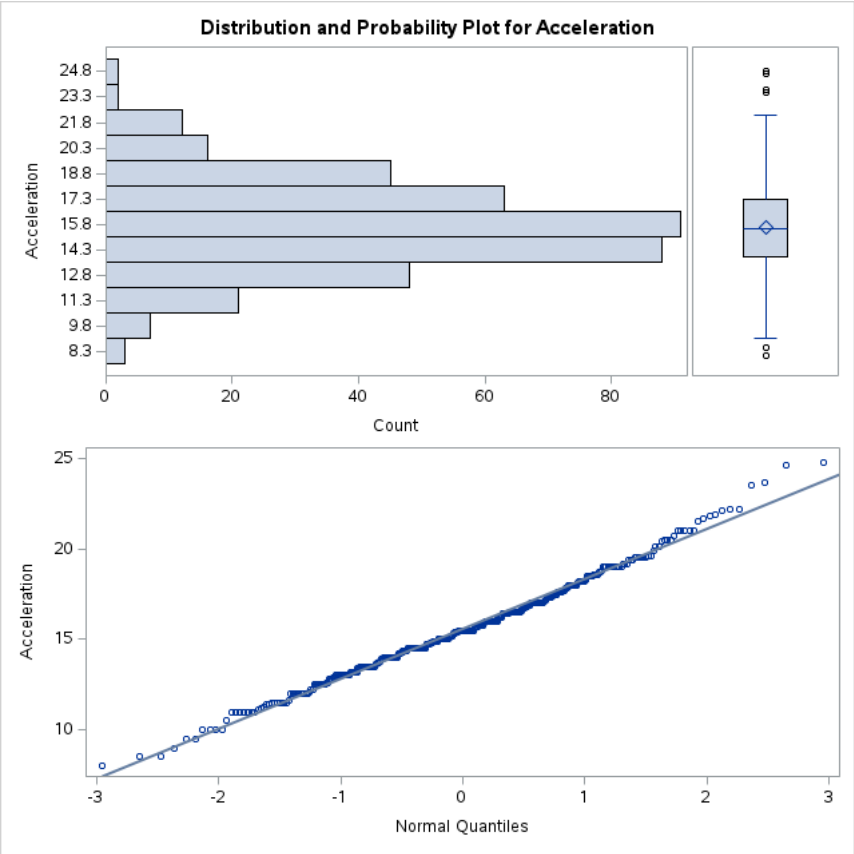
Basic Statistical Measures			
Location		Variability	
Mean	15.56809	Std Deviation	2.75769
Median	15.50000	Variance	7.60485
Mode	14.50000	Range	16.80000
		Interquartile Range	3.40000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	112.6242	Pr > |t|	<.0001
Sign	M	199	Pr >= |M|	<.0001
Signed Rank	S	39700.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	24.8

Quantiles (Definition 5)	
Level	Quantile
99%	23.5
95%	20.5
90%	19.0
75% Q3	17.2
50% Median	15.5
25% Q1	13.8
10%	12.0
5%	11.2
1%	9.0
0% Min	8.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
8.0	307	22.2	397
8.5	305	23.5	27
8.5	303	23.7	151
9.0	302	24.6	205
9.5	351	24.8	128



Listing of Auto MPG(PWR-new derived variable)

The UNIVARIATE Procedure
Variable: Displacement

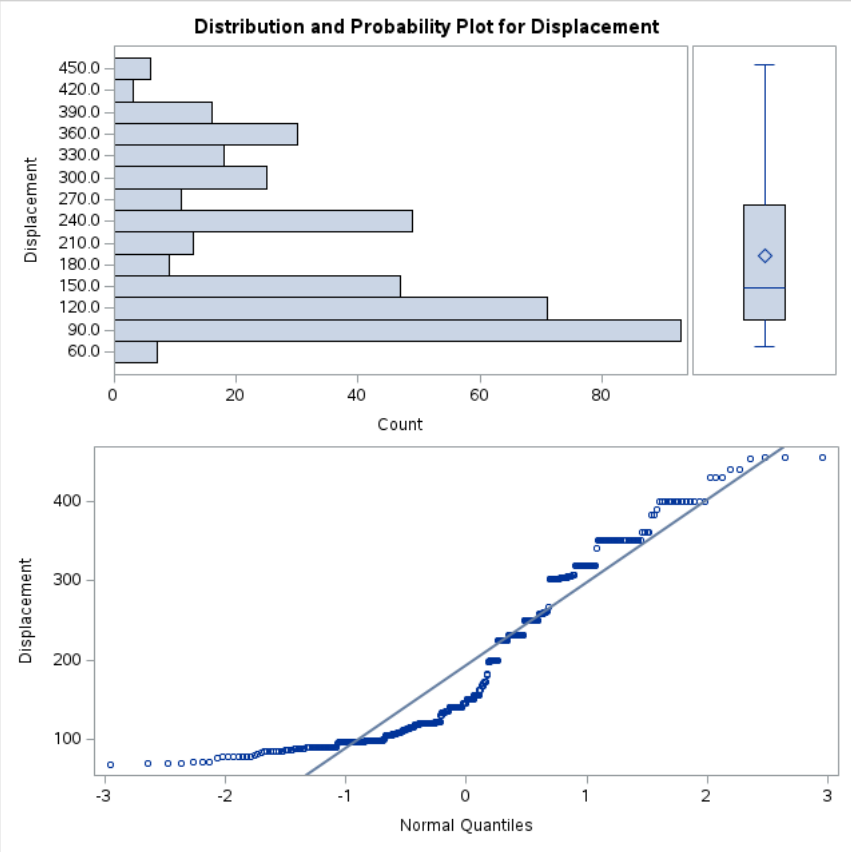
Moments			
N	398	Sum Weights	398
Mean	193.425879	Sum Observations	76983.5
Std Deviation	104.269838	Variance	10872.1992
Skewness	0.71964516	Kurtosis	-0.7465966
Uncorrected SS	19206864.3	Corrected SS	4316263.06
Coeff Variation	53.9068704	Std Error Mean	5.22657472

Basic Statistical Measures			
Location		Variability	
Mean	193.4259	Std Deviation	104.26984
Median	148.5000	Variance	10872
Mode	97.0000	Range	387.00000
		Interquartile Range	158.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	37.00815	Pr > |t|	<.0001
Sign	M	199	Pr >= |M|	<.0001
Signed Rank	S	39700.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	455.0
99%	454.0
95%	400.0
90%	350.0
75% Q3	262.0
50% Median	148.5
25% Q1	104.0
10%	90.0
5%	85.0
1%	70.0
0% Min	68.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	45	440	343
70	4	454	302
70	2	455	304
70	1	455	309
71	52	455	344



Listing of Auto MPG(PWR-new derived variable)

The UNIVARIATE Procedure
Variable: Weight

Moments			
N	398	Sum Weights	398
Mean	2970.42462	Sum Observations	1182229
Std Deviation	846.841774	Variance	717140.991
Skewness	0.53106251	Kurtosis	-0.7855289
Uncorrected SS	3796427105	Corrected SS	284704973
Coeff Variation	28.5091151	Std Error Mean	42.4483425

Basic Statistical Measures			
Location		Variability	
Mean	2970.425	Std Deviation	846.84177
Median	2803.500	Variance	717141
Mode	1985.000	Range	3527
		Interquartile Range	1386

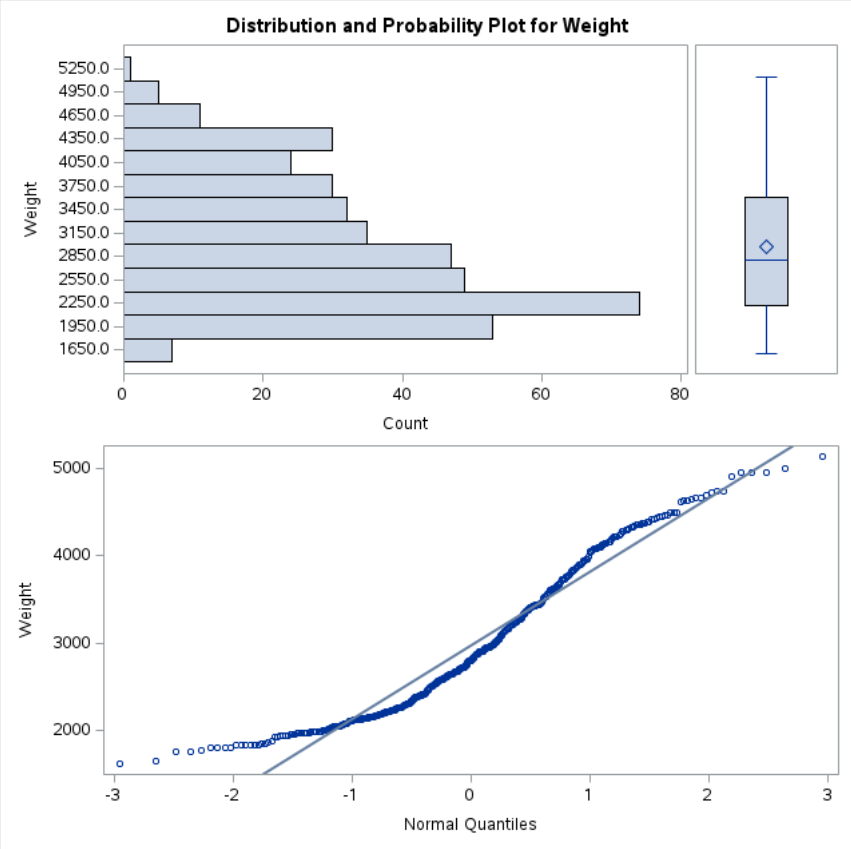
Note: The mode displayed is the smallest of 2 modes with a count of 4.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	69.9774	Pr > |t|	<.0001
Sign	M	199	Pr >= |M|	<.0001

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Signed Rank	S	39700.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5140.0
99%	4952.0
95%	4464.0
90%	4278.0
75% Q3	3609.0
50% Median	2803.5
25% Q1	2223.0
10%	1985.0
5%	1915.0
1%	1760.0
0% Min	1613.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1613	22	4951	344
1649	57	4952	339
1755	164	4955	318
1760	166	4997	346
1773	21	5140	320



Listing of Auto MPG(PWR-new derived variable)

The UNIVARIATE Procedure
Variable: Horsepower

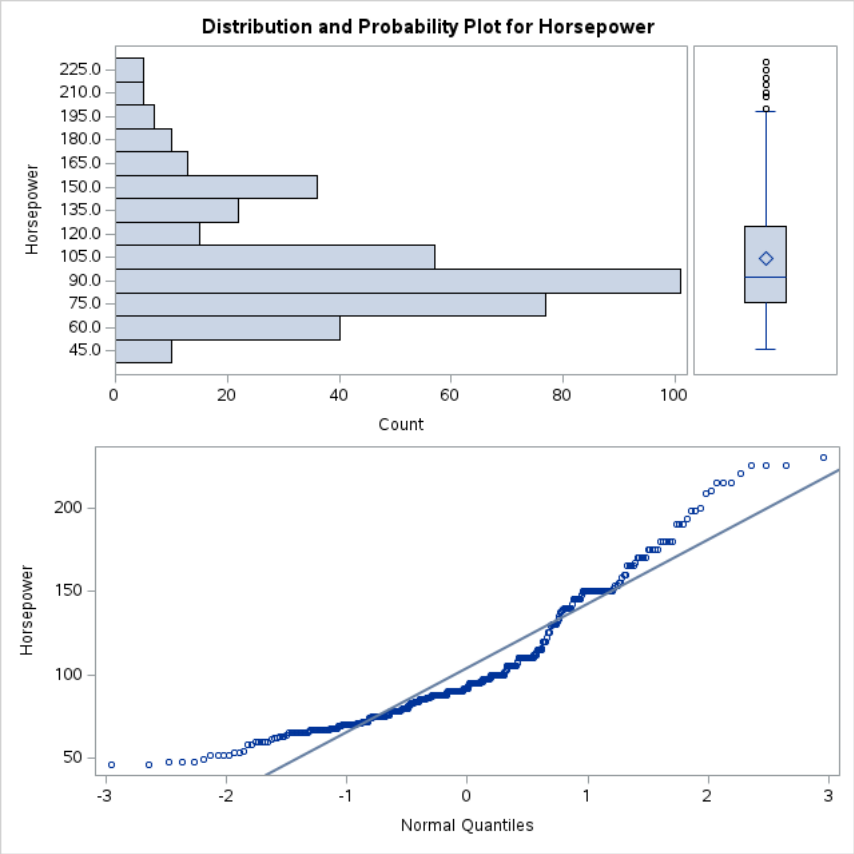
Moments			
N	398	Sum Weights	398
Mean	104.132947	Sum Observations	41444.9131
Std Deviation	38.3108858	Variance	1467.72397
Skewness	1.10843297	Kurtosis	0.7513134
Uncorrected SS	4898467.37	Corrected SS	582686.415
Coeff Variation	36.7903596	Std Error Mean	1.92035118

Basic Statistical Measures			
Location		Variability	
Mean	104.1329	Std Deviation	38.31089
Median	92.0000	Variance	1468
Mode	150.0000	Range	184.00000
		Interquartile Range	49.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	54.22599	Pr > |t|	<.0001
Sign	M	199	Pr >= |M|	<.0001
Signed Rank	S	39700.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	230
99%	225
95%	180
90%	158
75% Q3	125
50% Median	92
25% Q1	76
10%	67
5%	60
1%	48
0% Min	46

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
46	39	220	302
46	7	225	304
48	151	225	309
48	150	225	344
48	106	230	351



Listing Outliers for Acceleration

```
Outlier detected for Volkswagen Type 3  where Acceleration = 23.5
Outlier detected for Chevrolet Chevette  where Acceleration = 22.2
Outlier detected for Chevrolet Woody  where Acceleration = 22.1
Outlier detected for Peugeot 504  where Acceleration = 21.9
Outlier detected for Volkswagen Rabbit Custom Dies  where Acceleration = 21.5
Outlier detected for Peugeot 504  where Acceleration = 24.8
Outlier detected for Volkswagen Rabbit C Diesel  where Acceleration = 21.7
Outlier detected for Volkswagen Dasher Diesel  where Acceleration = 23.7
Outlier detected for Mercedes Benz 240d  where Acceleration = 21.8
Outlier detected for Volkswagen Pickup  where Acceleration = 24.6
Outlier detected for Ford Galaxie 500  where Acceleration = 10.0
Outlier detected for Chevrolet Impala  where Acceleration = 9.0
Outlier detected for Plymouth Fury Iii  where Acceleration = 8.5
Outlier detected for Pontiac Catalina  where Acceleration = 10.0
Outlier detected for Amc Ambassador Dpl  where Acceleration = 8.5
Outlier detected for Dodge Challenger Se  where Acceleration = 10.0
Outlier detected for Plymouth cuda 340  where Acceleration = 8.0
Outlier detected for Chevrolet Monte Carlo  where Acceleration = 9.5
Outlier detected for Buick Estate Wagon Sw  where Acceleration = 10.0
Outlier detected for Pontiac Grand Prix  where Acceleration = 9.5
Outlier detected for Oldsmobile Cutlass Salon Brou  where Acceleration = 22.2
```

Listing Outliers for Acceleration

The UNIVARIATE Procedure
Variable: Acceleration

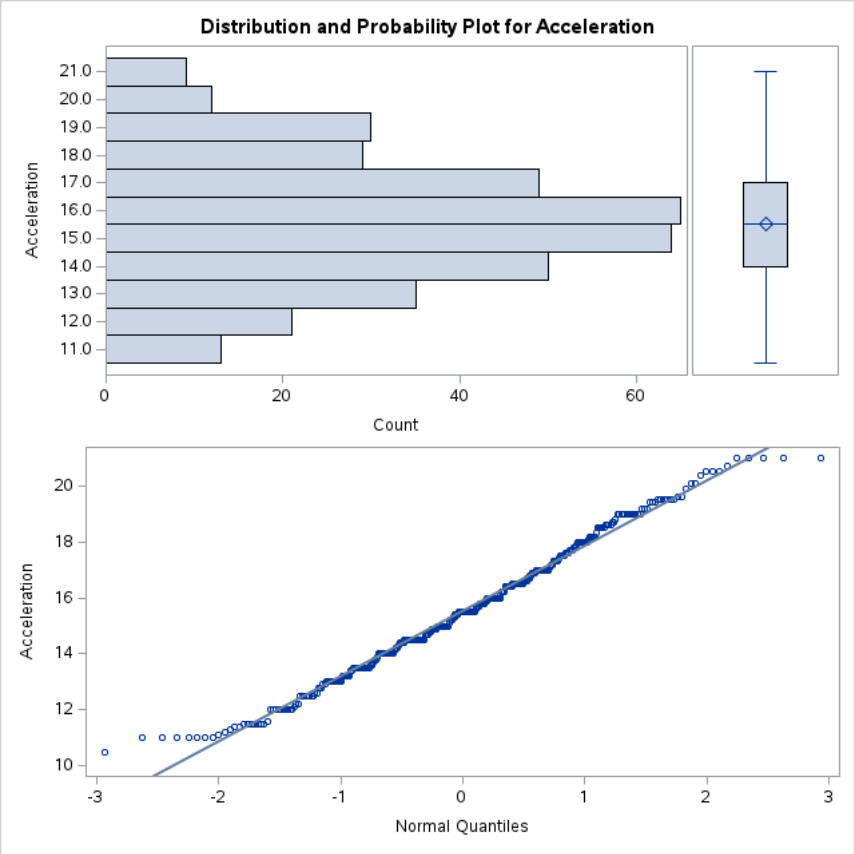
Moments			
N	377	Sum Weights	377
Mean	15.5254642	Sum Observations	5853.1
Std Deviation	2.33056446	Variance	5.4315307
Skewness	0.16335236	Kurtosis	-0.4993031
Uncorrected SS	92914.35	Corrected SS	2042.25554
Coeff Variation	15.0112385	Std Error Mean	0.12003018

Basic Statistical Measures			
Location		Variability	
Mean	15.52546	Std Deviation	2.33056
Median	15.50000	Variance	5.43153
Mode	14.50000	Range	10.50000
		Interquartile Range	3.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	129.3463	Pr > |t|	<.0001
Sign	M	188.5	Pr >= |M|	<.0001
Signed Rank	S	35626.5	Pr >= |S|	<.0001

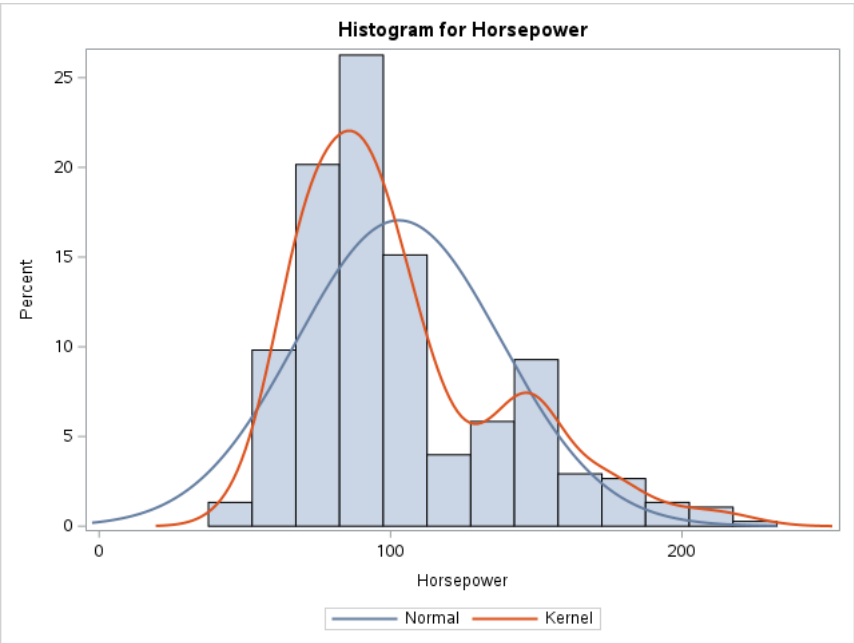
Quantiles (Definition 5)	
Level	Quantile
100% Max	21.0
99%	21.0
95%	19.5
90%	19.0
75% Q3	17.0
50% Median	15.5
25% Q1	14.0
10%	12.5
5%	11.5
1%	11.0
0% Min	10.5

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
10.5	290	21	38
11.0	333	21	51
11.0	332	21	231
11.0	326	21	233
11.0	325	21	246

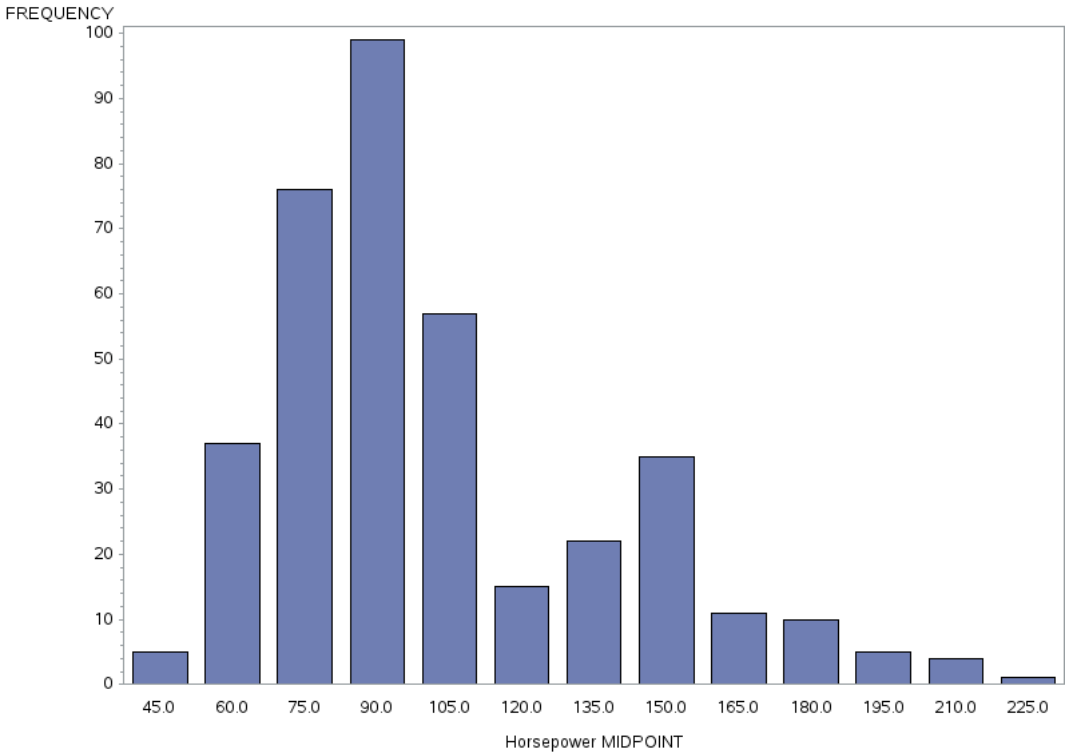


Listing Outliers for Power-Weight Ratio

```
Outlier detected for Bmw 2002  Power-Weight ratio = 0.0505819158
Outlier detected for Oldsmobile Omega  Power-Weight ratio = 0.0491266376
```



Histogram for Horsepower



QQ-Plot for Horsepower

The UNIVARIATE Procedure
Variable: Horsepower

Moments			
N	377	Sum Weights	377
Mean	102.875101	Sum Observations	38783.9131
Std Deviation	35.1016531	Variance	1232.12605
Skewness	1.05272099	Kurtosis	0.5868629
Uncorrected SS	4453178.37	Corrected SS	463279.395
Coeff Variation	34.12065	Std Error Mean	1.80782711

Basic Statistical Measures			
Location		Variability	
Mean	102.8751	Std Deviation	35.10165
Median	92.0000	Variance	1232
Mode	150.0000	Range	179.00000
		Interquartile Range	43.00000

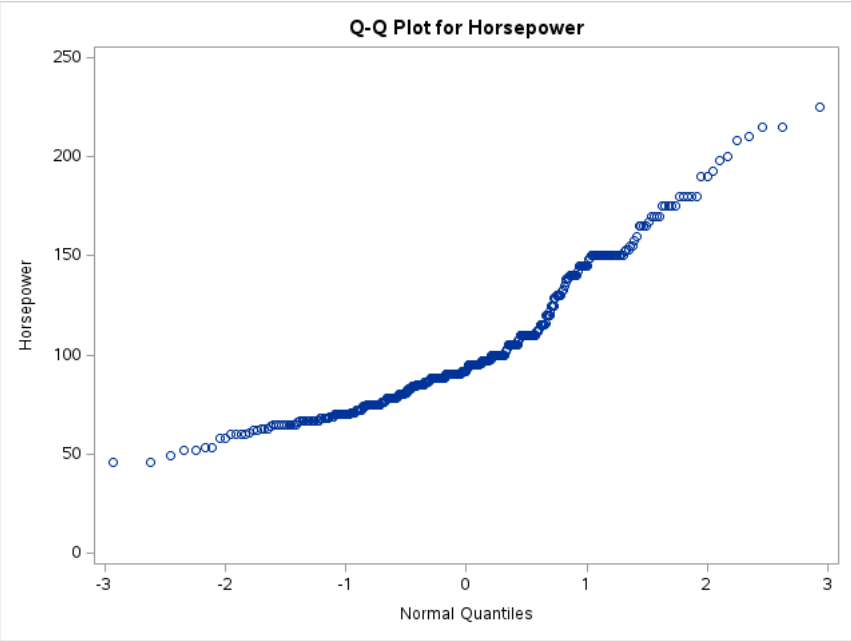
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	56.90539	Pr > |t|	<.0001
Sign	M	188.5	Pr >= |M|	<.0001
Signed Rank	S	35626.5	Pr >= |S|	<.0001

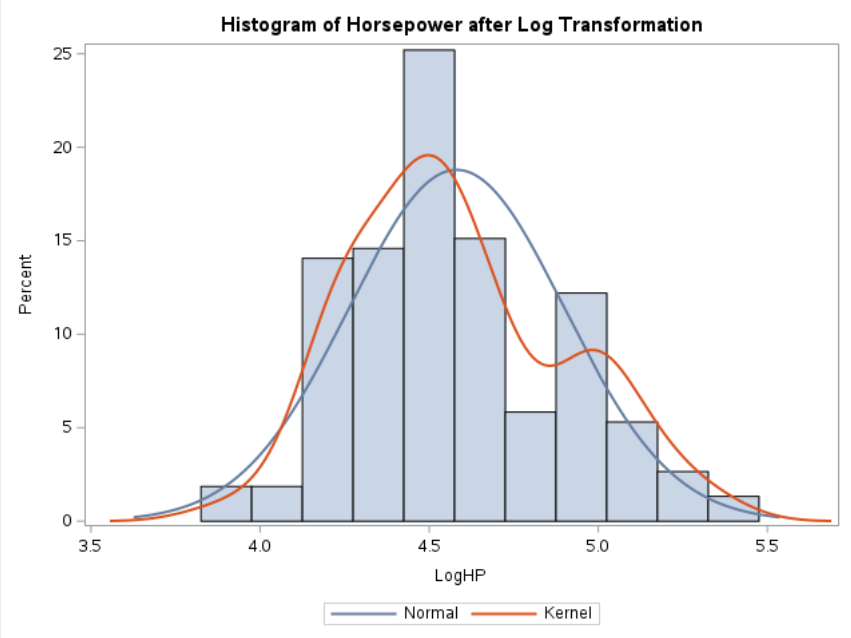
Quantiles (Definition 5)	
Level	Quantile
100% Max	225
99%	210
95%	175
90%	150
75% Q3	120
50% Median	92
25% Q1	77
10%	67
5%	63
1%	52
0% Min	46

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
46	38	208	307
46	7	210	293
49	44	215	291
52	103	215	324
52	56	225	325

QQ-Plot for Horsepower

The UNIVARIATE Procedure





QQ-Plot of Horsepower after Log Transformation

The UNIVARIATE Procedure
Variable: LogHP

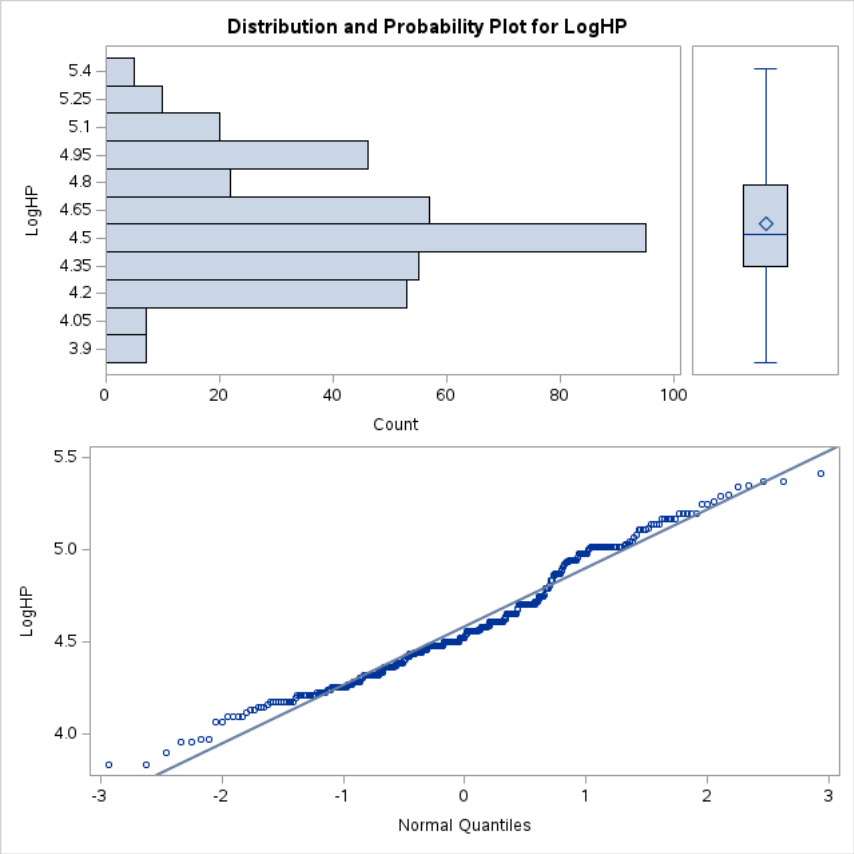
Moments			
N	377	Sum Weights	377
Mean	4.58104878	Sum Observations	1727.05539
Std Deviation	0.31840874	Variance	0.10138413
Skewness	0.4032149	Kurtosis	-0.4329167
Uncorrected SS	7949.8454	Corrected SS	38.1204312
Coeff Variation	6.95056428	Std Error Mean	0.01639888

Basic Statistical Measures			
Location		Variability	
Mean	4.581049	Std Deviation	0.31841
Median	4.521789	Variance	0.10138
Mode	5.010635	Range	1.58746
		Interquartile Range	0.44369

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	279.3512	Pr > |t|	<.0001
Sign	M	188.5	Pr >= |M|	<.0001
Signed Rank	S	35626.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5.41610
99%	5.34711
95%	5.16479
90%	5.01064
75% Q3	4.78749
50% Median	4.52179
25% Q1	4.34381
10%	4.20469
5%	4.14313
1%	3.95124
0% Min	3.82864

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
3.82864	38	5.33754	307
3.82864	7	5.34711	293
3.89182	44	5.37064	291
3.95124	103	5.37064	324
3.95124	56	5.41610	325



Listing First 5 Observations from Final Dataset

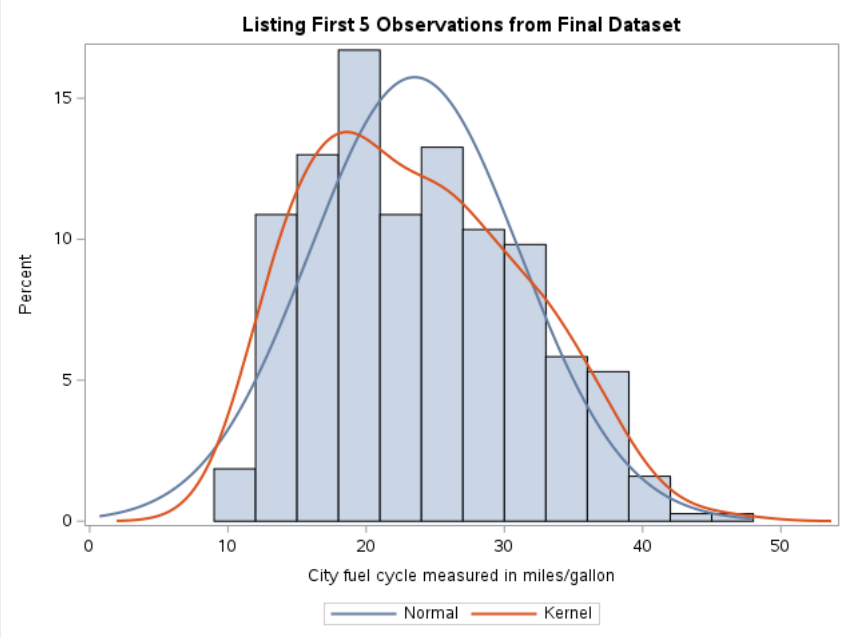
Obs	Mpg	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model_Year	Origin	Brand	Model	PWR
1	19.0	3	70.0	97.0	2330.0	13.5	1972	3	Mazda	Rx2 Coupe	0.041631
2	18.0	3	70.0	90.0	2124.0	13.5	1973	3	Mazda	Rx3	0.042373
3	21.5	3	80.0	110.0	2720.0	13.5	1977	3	Mazda	Rx 4	0.040441
4	23.7	3	70.0	100.0	2420.0	12.5	1980	3	Mazda	Rx 7 Gs	0.041322
5	24.0	4	113.0	95.0	2372.0	15.0	1970	3	Toyota	Corona Mark Ii	0.040051

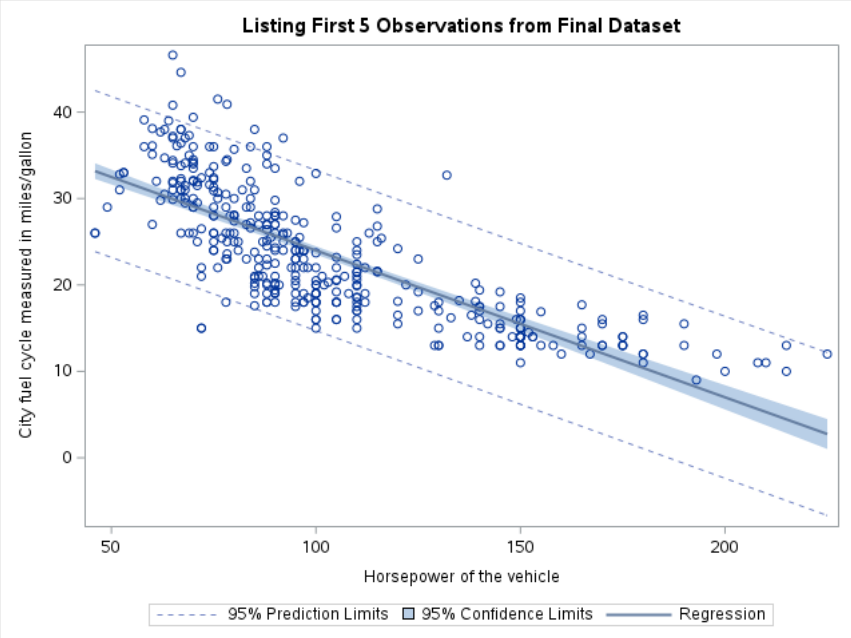
Listing First 5 Observations from Final Dataset

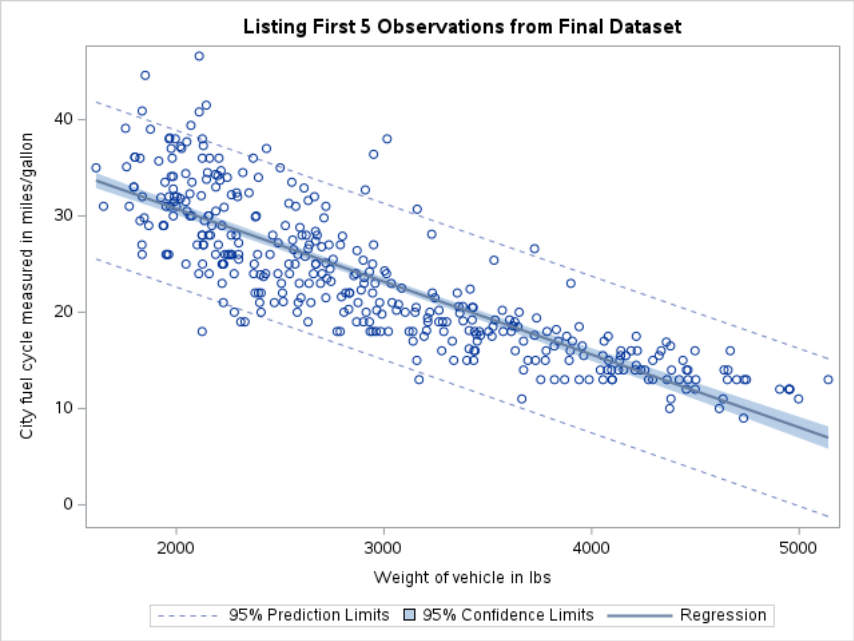
The CONTENTS Procedure

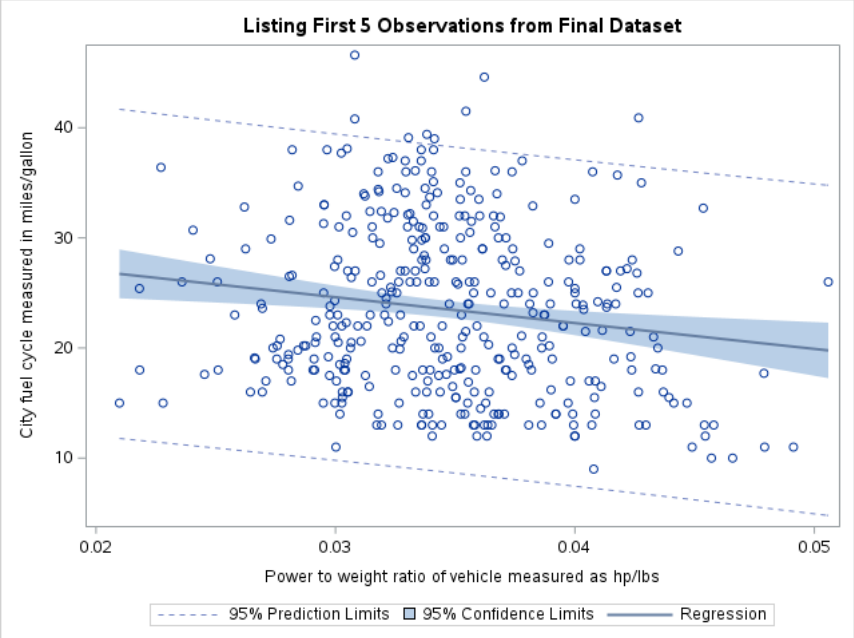
Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
6	Acceleration	Num	8	4.1	Time taken to reach from 0-60 mph
9	Brand	Char	20		Brand of the Vehicle
2	Cylinders	Num	8		Number of Cylinders. Categorical Variable which can take following values: 4, 6 or 8
3	Displacement	Num	8	5.1	Engine size of vehicle measured in cubic centimetres(CC)
4	Horsepower	Num	8	5.1	Horsepower of the vehicle
10	Model	Char	200		Model name of vehicle
7	Model_Year	Num	8		The year in which the vehicle was manufactured
1	Mpg	Num	8	4.1	City fuel cycle measured in miles/gallon
8	Origin	Num	8		Country of Origin of the Vehicle Brand. Has the following categories: Unites States = 1 Germany =2 Japan = 3

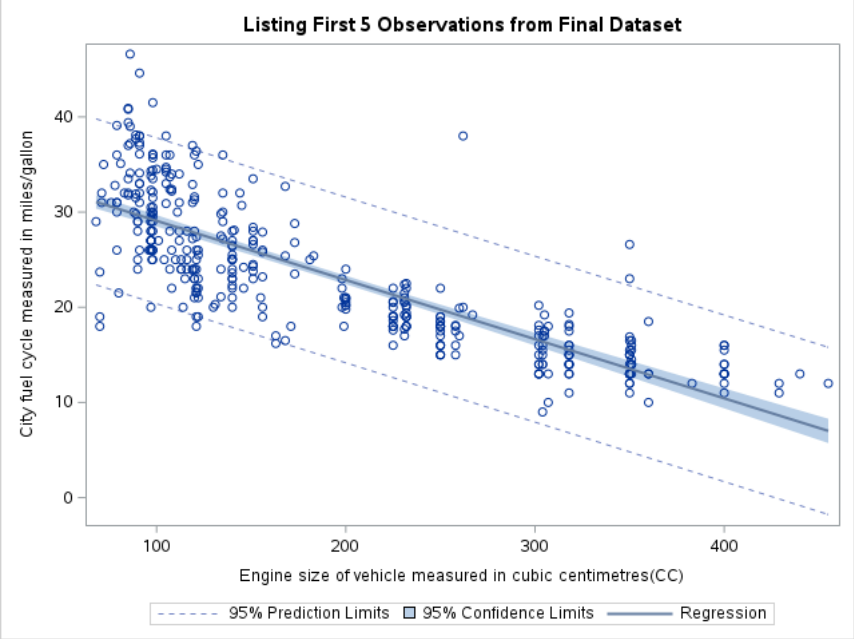
Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
11	PWR	Num	8		Power to weight ratio of vehicle measured as hp/lbs
5	Weight	Num	8	6.1	Weight of vehicle in lbs






Listing First 5 Observations from Final Dataset
City fuel cycle measured in miles/gallon
0
10
20
30
40
2000
3000
4000
5000
Weight of vehicle in lbs
95% Prediction Limits
95% Confidence Limits
Regression


Listing First 5 Observations from Final Dataset
City fuel cycle measured in miles/gallon
10
20
30
40
0.02
0.03
0.04
0.05
Power to weight ratio of vehicle measured as hp/lbs
95% Prediction Limits
95% Confidence Limits
Regression


Listing First 5 Observations from Final Dataset
City fuel cycle measured in miles/gallon
0
10
20
30
40
100
200
300
400
Engine size of vehicle measured in cubic centimetres(CC)
95% Prediction Limits
95% Confidence Limits
Regression

Listing First 5 Observations from Final Dataset

The UNIVARIATE Procedure
Variable: Mpg (City fuel cycle measured in miles/gallon)

Moments			
N	377	Sum Weights	377
Mean	23.5050398	Sum Observations	8861.4
Std Deviation	7.59955913	Variance	57.753299
Skewness	0.36963984	Kurtosis	-0.6473754
Uncorrected SS	230002.8	Corrected SS	21715.2404
Coeff Variation	32.3316157	Std Error Mean	0.39139721

Basic Statistical Measures			
Location		Variability	
Mean	23.50504	Std Deviation	7.59956
Median	23.00000	Variance	57.75330
Mode	13.00000	Range	37.60000
		Interquartile Range	11.40000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	60.05418	Pr > |t|	<.0001
Sign	M	188.5	Pr >= |M|	<.0001
Signed Rank	S	35626.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	46.6
99%	40.9
95%	37.0
90%	34.2
75% Q3	29.0
50% Median	23.0
25% Q1	17.6
10%	14.0
5%	13.0
1%	11.0
0% Min	9.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
9	294	40.8	143
10	292	40.9	145
10	291	41.5	129
11	333	44.6	144
11	327	46.6	141