Case Study – Exploratory Data Analysis

Submitted by: Vishal Kumar

Due Date: 24th August, 2015

Date of Submission: 23rd August, 2015

Supervisor's Remarks

Late Submission:

Plagiarism:

Completeness:

Quality of Content:

Results and Interpretations:

Additional Remarks:

Exploratory Data Analysis

Exploratory data analysis covers the essential exploratory technique for summarizing data. It is an important part of Data Science. This technique is applied before formal modelling starts on the data and helps in the development of more complex statistical models. Exploratory techniques are also important for eliminating or establishing hypotheses about the world that can be addressed by the data. In this case study we will cover some plotting techniques in SPSS. We will also cover some of the common multivariate statistical techniques used to visualize high-dimensional data.

Data Source and Description

The dataset for our case study is "Motor Trend Car Road Tests". This data set is available in software R and has been exported in CSV format from there.

The data was taken from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Format of our data: A data frame with 32 observations on 11 variables.

1.	mpg	Miles/(US) gallon	(Continuous Data)
2.	cyl	Number of cylinders	(Discrete Data)
3.	disp	Displacement (cu.in.)	(Continuous Data)
4.	hp	Gross horsepower	(Continuous Data)
5.	drat	Rear axle ratio	(Continuous Data)
6.	wt	Weight (lb/1000)	(Continuous Data)
7.	qsec	1/4 mile time	(Continuous Data)
8.	vs	V/S	(Ordinal Data)
9.	am	Transmission (0 = automatic, 1 = manual)	(Ordinal Data)
10.	gear	Number of forward gears	(Ordinal Data)
11.	carb	Number of carburettors	(Ordinal Data)

This dataset however is available Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

EDA for Individual Variables

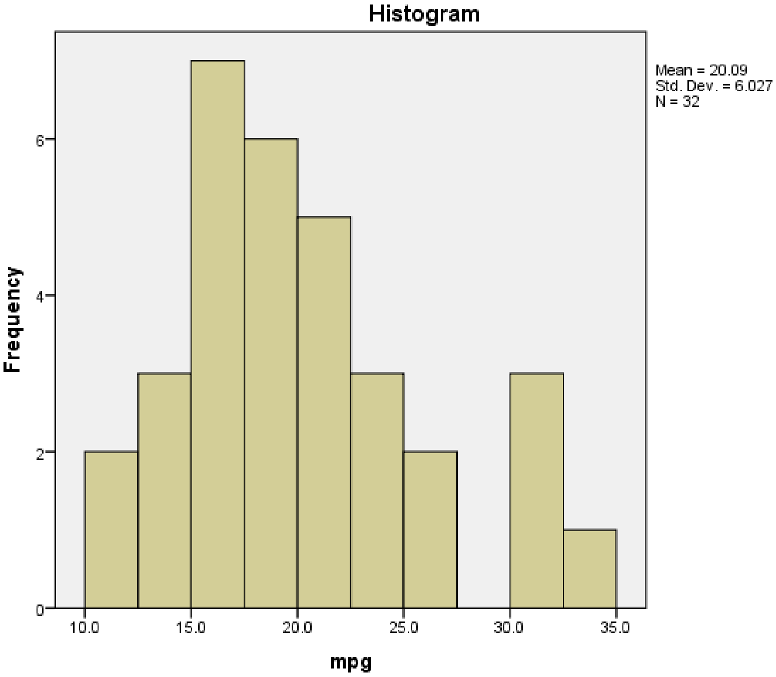
1. Miles per gallon

This variable is represented by **mpg**. It is a continuous variable.

Descriptive Statistics

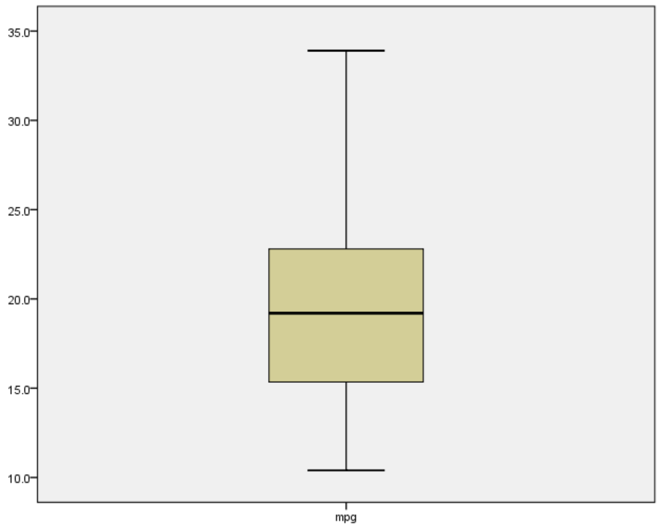
			Statistic	Std. Error
mpg	Mean		20.091	1.0654
	95% Confidence Interval for Mean	Lower Bound	17.918	
		Upper Bound	22.264	
	5% Trimmed Mean		19.893	
	Median		19.200	
	Variance		36.324	
	Std. Deviation		6.0269	
	Minimum		10.4	
	Maximum		33.9	
	Range		23.5	
	Interquartile Range		7.5	
	Skewness		.672	.414
	Kurtosis		-.022	.809

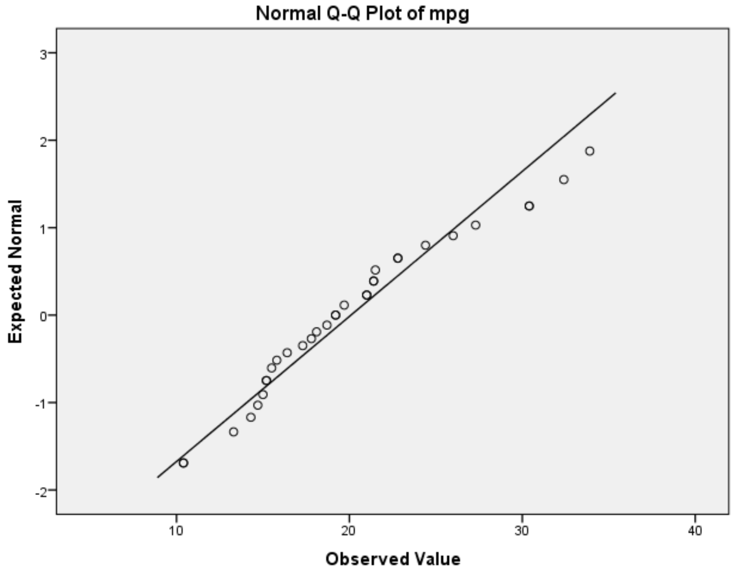
We set our hypotheses as:
H_0: The sample is from a normal population.
H_1: The sample is **NOT** from a normal population.




```
mpg Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     5.00        1 .  00344
    13.00        1 .  5555567788999
     8.00        2 .  11111224
     2.00        2 .  67
     4.00        3 .  0023

 Stem width:    10.0
 Each leaf:        1 case(s)
```

Box Plot





<u>Inference</u>

From histogram, stem and leaf plot and box plot we see that our data is slightly skewed to the right i.e. it is positively skewed. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data. No outliers were detected in the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
mpg	.126	32	.200*	.948	32	.123

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

<u>Inference</u>

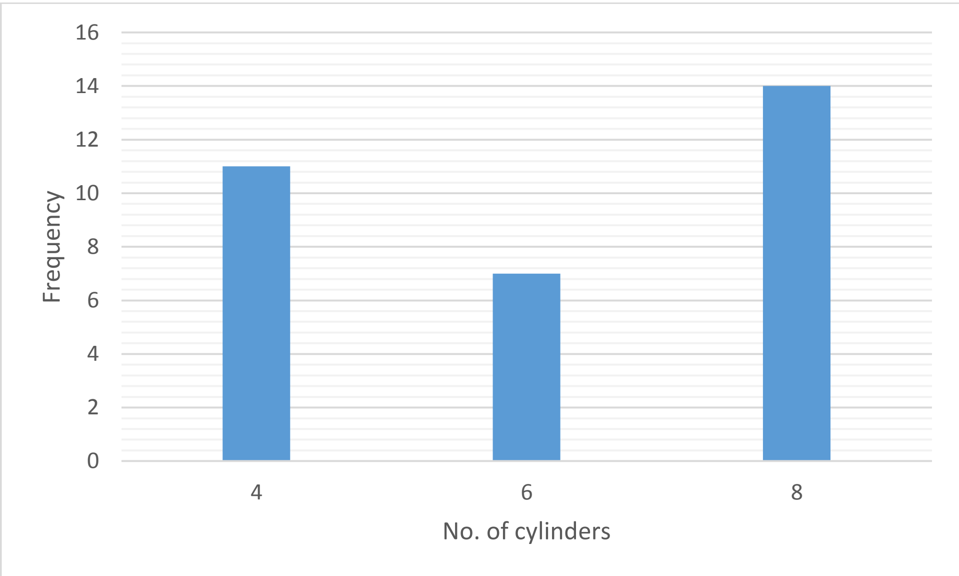
From Kolmogorov –Smirnov's test we see p-value is 0.200 > 0.05 and from Shapiro-Wilk's test the p-value is 0.123 > 0.05, hence we fail to reject our null hypothesis at 5% level of significance and conclude that the sample is from normal population.

2. **No. of cylinder**

This variable is represented by **cyl**. It is a discrete variable.

Descriptive Statistics

			Statistic	Std. Error
cyl	Mean		6.19	.316
	95% Confidence Interval for Mean	Lower Bound	5.54	
		Upper Bound	6.83	
	5% Trimmed Mean		6.21	
	Median		6.00	
	Variance		3.190	
	Std. Deviation		1.786	
	Minimum		4	
	Maximum		8	
	Range		4	
	Interquartile Range		4	
	Skewness		-.192	.414
	Kurtosis		-1.763	.809



<u>**Inference**</u>

No missing value has been observed. Data is discrete hence we do not plot Stem & leaf plot and Box plot. Also we do not go for test for normality or Q-Q plot.

We observe that most of the cars have 8 cylinder configuration.

3. (A) **Displacement**

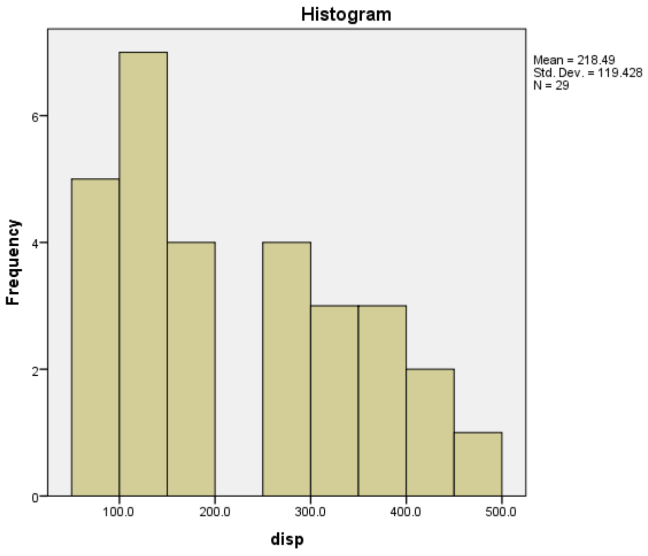
This variable is represented by **disp** (cu.in.). It is a continuous variable. We have missing values in our data. So first we will perform our analysis on missing value, then we will estimate missing value by the series mean and finally we will compare the two analysis.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
disp	29	90.6%	3	9.4%	32	100.0%

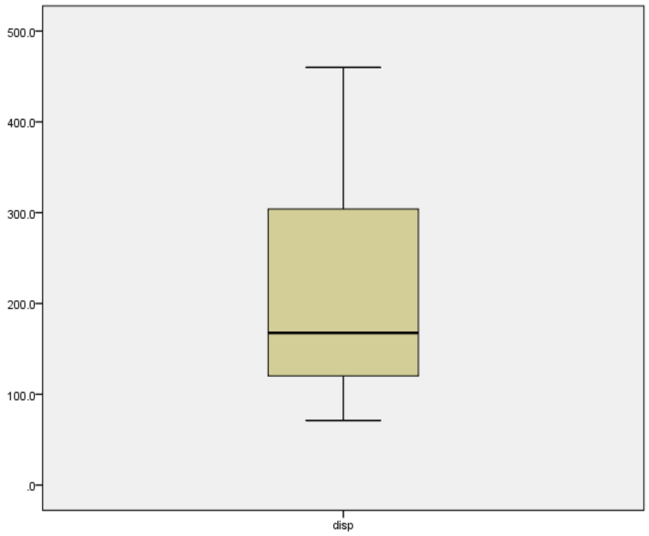
Descriptive Statistics

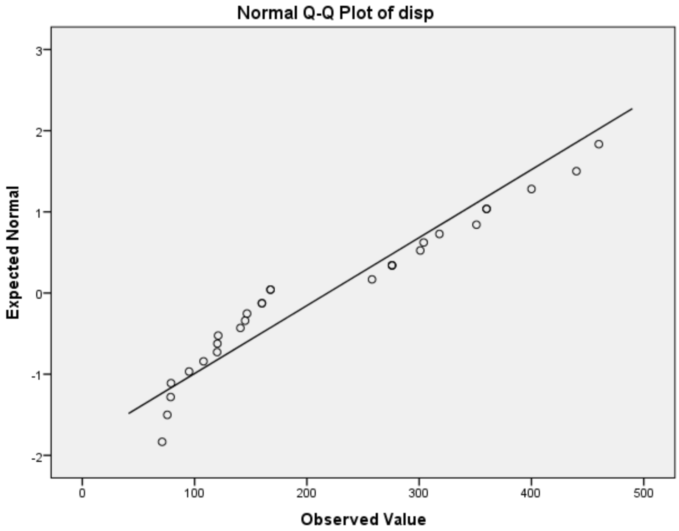
			Statistic	Std. Error
disp	Mean		218.486	22.1773
	95% Confidence Interval for Mean	Lower Bound	173.058	
		Upper Bound	263.914	
	5% Trimmed Mean		213.522	
	Median		167.600	
	Variance		14263.143	
	Std. Deviation		119.4284	
	Minimum		71.1	
	Maximum		460.0	
	Range		388.9	
	Interquartile Range		190.8	
	Skewness		.500	.434
	Kurtosis		-1.028	.845




```
disp Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     5.00        0 .  77779
     7.00        1 .  0222444
     4.00        1 .  6666
      .00        2 .
     4.00        2 .  5777
     3.00        3 .  001
     3.00        3 .  566
     2.00        4 .  04
     1.00        4 .  6

 Stem width:   100.0
 Each leaf:        1 case(s)
```

Box Plot





Inference

From histogram, stem and leaf plot and box plot we see that our data is slightly skewed to the right i.e. it is positively skewed. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data. No outliers were detected in the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
disp	.217	29	.001	.908	29	.016

a. Lilliefors Significance Correction

Inference

From Kolmogorov –Smirnov's test we see p-value is 0.001 < 0.05 and from Shapiro-Wilk's test the p-value is 0.016 < 0.05, hence we reject our null hypothesis at 5% level of significance and conclude that the sample is not from normal population.

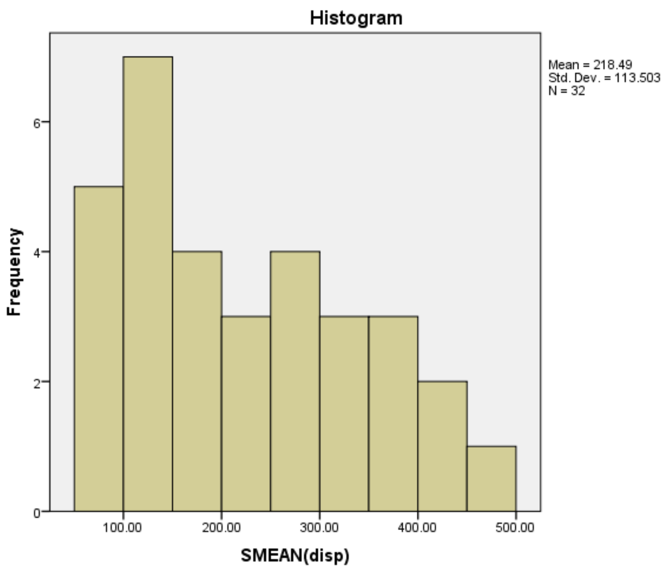
In the next section we will estimate the missing values by the series mean and perform similar analysis. After that we will compare the two analysis.

3. (B) **Displacement** (Missing value analysis)

Descriptive Statistics

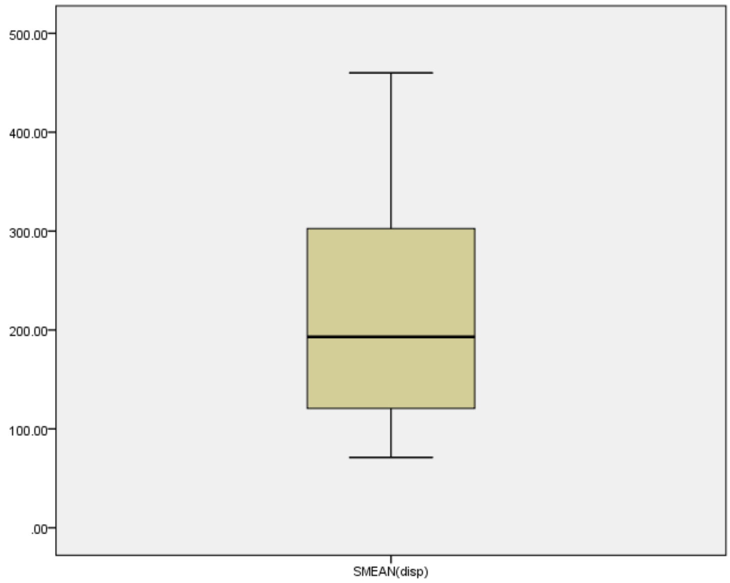
			Statistic	Std. Error
SMEAN(disp)	Mean		218.4862	20.06461
	95% Confidence Interval for Mean	Lower Bound	177.5642	
		Upper Bound	259.4083	
	5% Trimmed Mean		213.5777	
	Median		193.0431	
	Variance		12882.839	
	Std. Deviation		113.50260	
	Minimum		71.10	
	Maximum		460.00	
	Range		388.90	
	Interquartile Range		182.78	
	Skewness		.523	.414
	Kurtosis		-.796	.809

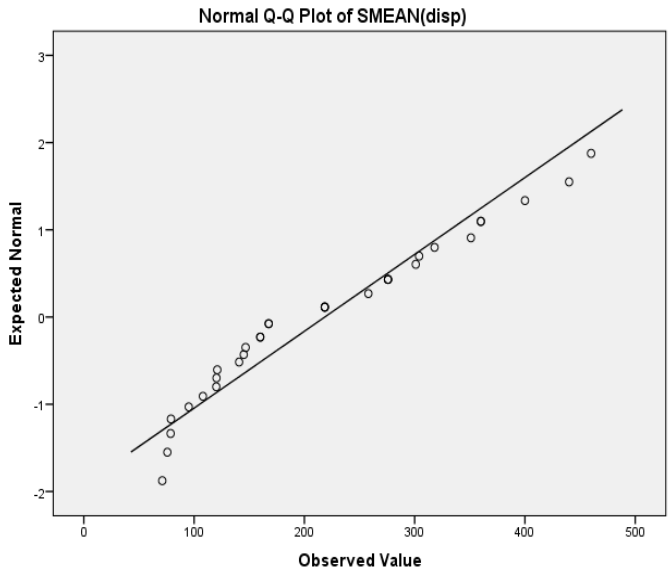
We set our hypotheses as:
H_0: The sample is from a normal population.
H_1: The sample is **NOT** from a normal population.




```
SMEAN(disp) Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     5.00        0 .  77779
     7.00        1 .  0222444
     4.00        1 .  6666
     3.00        2 .  111
     4.00        2 .  5777
     3.00        3 .  001
     3.00        3 .  566
     2.00        4 .  04
     1.00        4 .  6

 Stem width:   100.00
 Each leaf:        1 case(s)
```

Box Plot





Inference

From histogram, stem and leaf plot and box plot we see that our data is slightly skewed to the right i.e. it is positively skewed. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data. No outliers were detected in the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SMEAN(disp)	.173	32	.016	.932	32	.044

a. Lilliefors Significance Correction

Inference

From Kolmogorov –Smirnov's test we see p-value is 0.016 < 0.05 and from Shapiro-Wilk's test the p-value is 0.044 < 0.05, hence we reject our null hypothesis at 5% level of significance and conclude that the sample is not from normal population.

Comparison: Missing value analysis has reduced the skewness by some amount however both samples are still away from being normal.

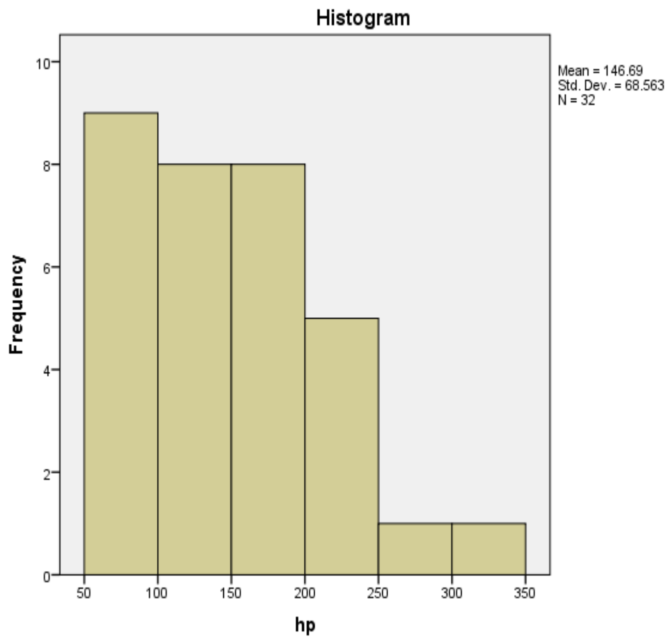
4. Gross Horsepower

This variable is represented **hp**. It is a continuous variable.

Descriptive Statistics

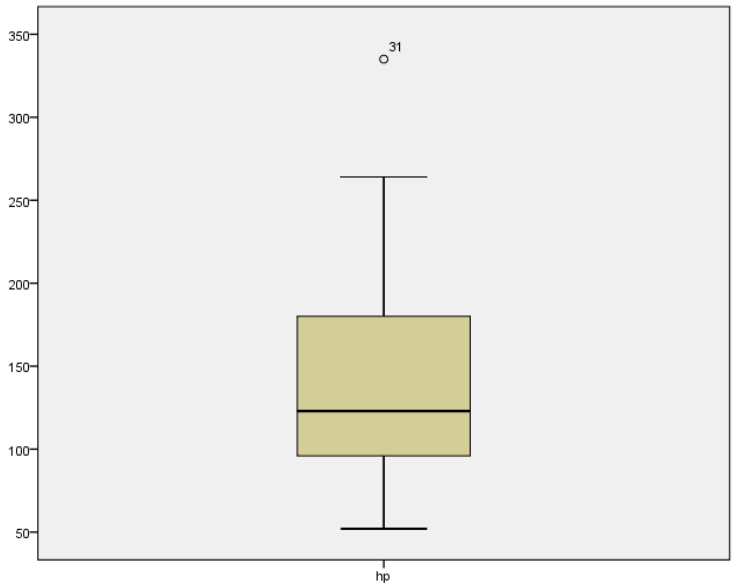
			Statistic	Std. Error
hp	Mean		146.69	12.120
	95% Confidence Interval for Mean	Lower Bound	121.97	
		Upper Bound	171.41	
	5% Trimmed Mean		142.76	
	Median		123.00	
	Variance		4700.867	
	Std. Deviation		68.563	
	Minimum		52	
	Maximum		335	
	Range		283	
	Interquartile Range		85	
	Skewness		.799	.414
	Kurtosis		.275	.809

We set our hypotheses as:
H_0: The sample is from a normal population.
H_1: The sample is **NOT** from a normal population.




```
hp Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     9.00        0 .  566669999
     8.00        1 .  00111122
     8.00        1 .  55777888
     5.00        2 .  01344
     1.00        2 .  6
     1.00 Extremes    (>=335)

 Stem width:   100
 Each leaf:        1 case(s)
```

Box Plot



Inference

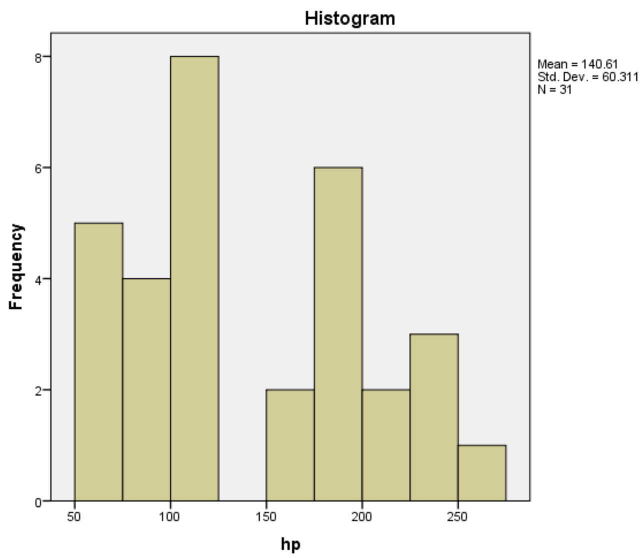
From histogram and stem and leaf plot we see that our data is skewed to the right i.e. it is positively skewed. However from Box-Plot we observe that we have an **outlier** in our data.

We will **remove** the outlier and perform the analysis again.

The 31st observation i.e. hp = 335 was removed and analysis was performed again as follows:

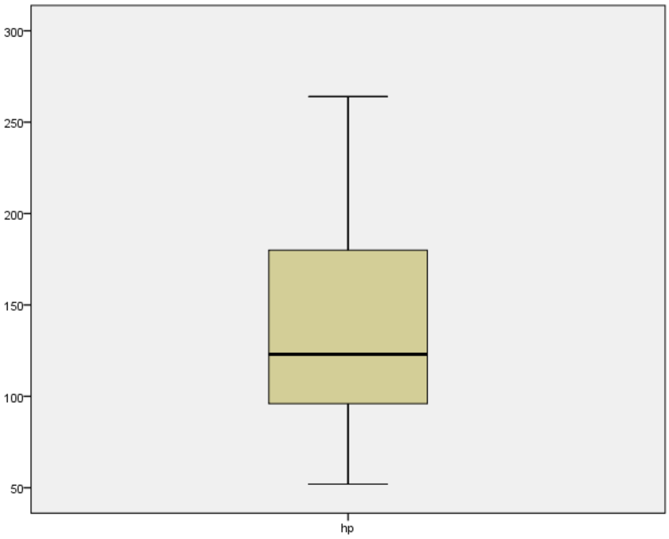
Descriptive Statistics

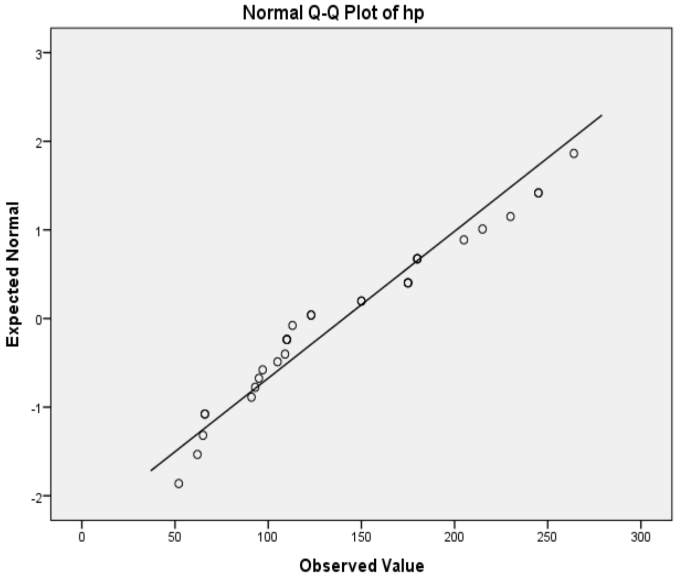
			Statistic	Std. Error
hp	Mean		140.61	10.832
	95% Confidence Interval for Mean	Lower Bound	118.49	
		Upper Bound	162.74	
	5% Trimmed Mean		138.86	
	Median		123.00	
	Variance		3637.378	
	Std. Deviation		60.311	
	Minimum		52	
	Maximum		264	
	Range		212	
	Interquartile Range		85	
	Skewness		.456	.421
	Kurtosis		-.826	.821




```
hp Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     9.00        0 .  566669999
     8.00        1 .  00111122
     8.00        1 .  55777888
     5.00        2 .  01344
     1.00        2 .  6

 Stem width:   100
 Each leaf:        1 case(s)
```

Box Plot





Inference

From histogram, stem and leaf plot and box plot we see that our data is slightly skewed to the right i.e. it is positively skewed. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data. No outliers were detected in the data now.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
hp	.163	31	.035	.938	31	.072

a. Lilliefors Significance Correction

Inference

From Kolmogorov –Smirnov's test we see p-value is 0.035 > 0.01 and from Shapiro-Wilk's test the p-value is 0.072 > 0.01, hence we fail to reject our null hypothesis at 1% level of significance and conclude that the sample is from normal population.

However for small samples we tend to prefer Shapiro-Wilk's test, so even at 5% level of significance we fail to reject the null hypothesis and conclude that sample is from normal population.

5. Rear Axle Ratio

This variable is represented by **Drat**. It is a continuous variable.

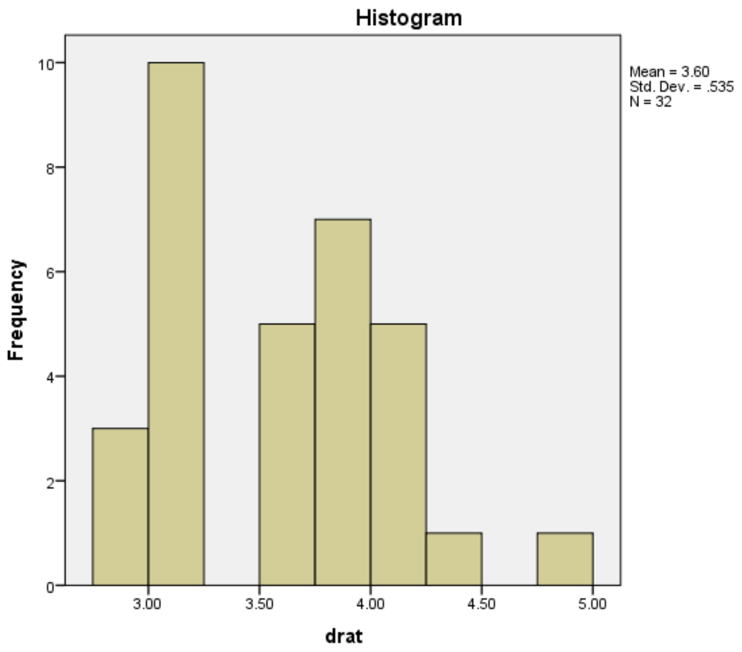
Descriptive Statistics

			Statistic	Std. Error
drat	Mean		3.5966	.09452
	95% Confidence Interval for Mean	Lower Bound	3.4038	
		Upper Bound	3.7893	
	5% Trimmed Mean		3.5794	
	Median		3.6950	
	Variance		.286	
	Std. Deviation		.53468	
	Minimum		2.76	
	Maximum		4.93	
	Range		2.17	
	Interquartile Range		.84	
	Skewness		.293	.414
	Kurtosis		-.450	.809

We set our hypotheses as:

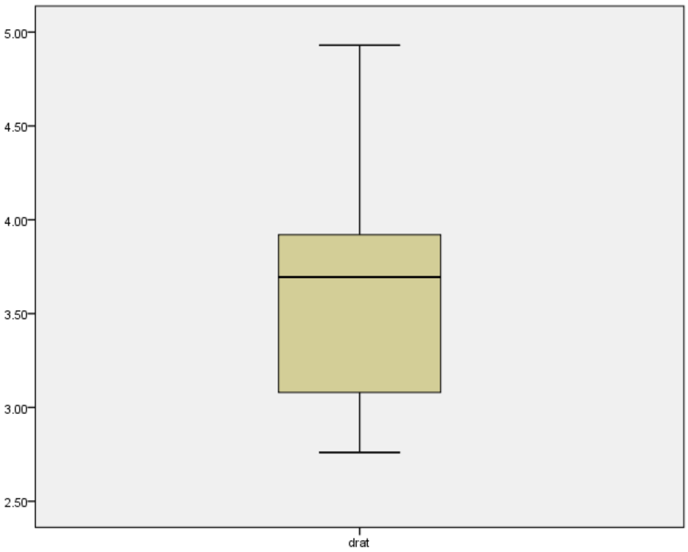
H_0: The sample is from a normal population.

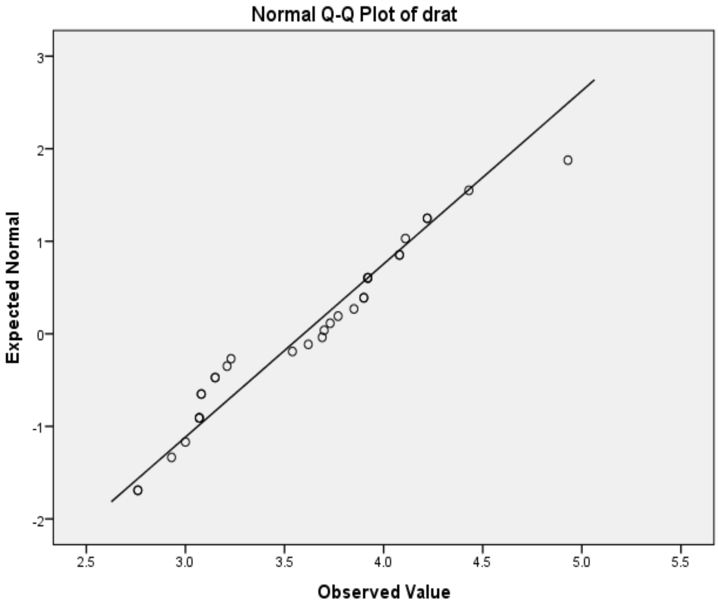
H_1: The sample is **NOT** from a normal population.




```
drat Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     3.00        2 .  779
    10.00        3 .  0000001122
    12.00        3 .  566777899999
     6.00        4 .  001224
     1.00        4 .  9

 Stem width:   1.00
 Each leaf:        1 case(s)
```

Box Plot





Inference

From histogram, stem and leaf plot and box plot we see that our data is skewed to the right i.e. it is positively skewed. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data. No outliers were detected in the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
drat	.160	32	.037	.946	32	.110

a. Lilliefors Significance Correction

Inference

From Kolmogorov –Smirnov's test we see p-value is 0.037 > 0.01 and from Shapiro-Wilk's test the p-value is 0.110 > 0.01, hence we fail to reject our null hypothesis at 1% level of significance and conclude that the sample is from normal population.

However for small samples we tend to prefer Shapiro-Wilk's test, so even at 5% level of significance we fail to reject the null hypothesis and conclude that sample is from normal population.

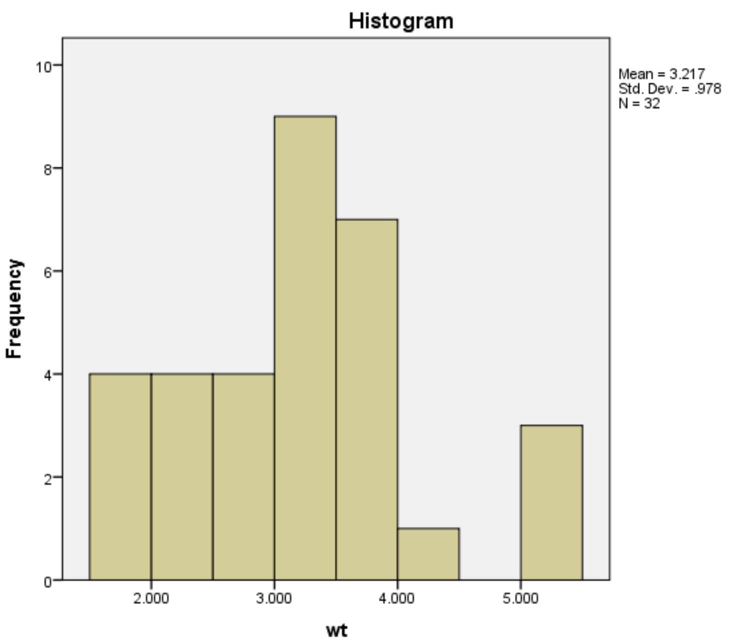
6. **Weight**

This variable is represented by **Wt**. It is a continuous variable.

Descriptive Statistics

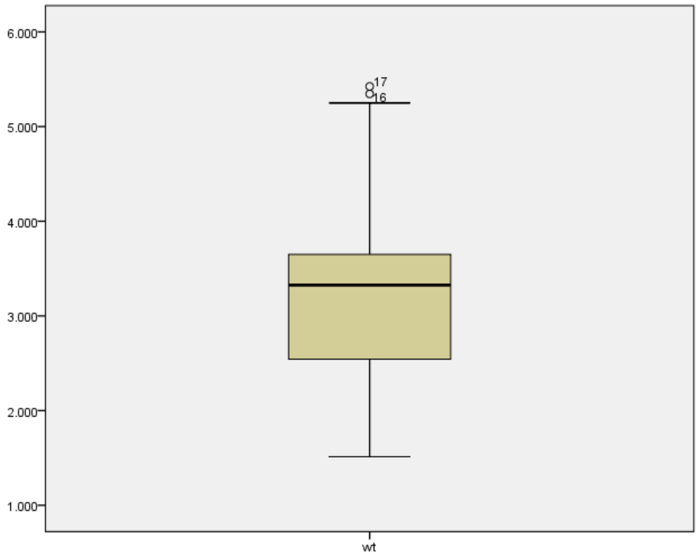
			Statistic	Std. Error
wt	Mean		3.21725	.172968
	95% Confidence Interval for Mean	Lower Bound	2.86448	
		Upper Bound	3.57002	
	5% Trimmed Mean		3.18885	
	Median		3.32500	
	Variance		.957	
	Std. Deviation		.978457	
	Minimum		1.513	
	Maximum		5.424	
	Range		3.911	
	Interquartile Range		1.186	
	Skewness		.466	.414
	Kurtosis		.417	.809

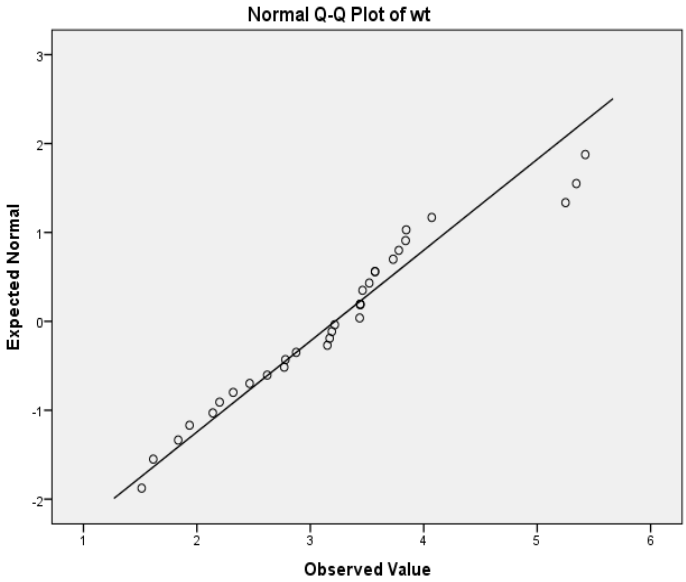
We set our hypotheses as:
H_0: The sample is from a normal population.
H_1: The sample is **NOT** from a normal population.




```
wt Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     4.00        1 .  5689
     4.00        2 .  1234
     4.00        2 .  6778
     9.00        3 .  111244444
     7.00        3 .  5557788
     1.00        4 .  0
      .00        4 .
     1.00        5 .  2
     2.00 Extremes    (>=5.3)

 Stem width:   1.000
 Each leaf:        1 case(s)
```

Box Plot





<u>Inference</u>

From histogram and stem and leaf plot we see that our data is pretty much normal. However from Box-Plot we observe that we have **outliers** in our data. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
wt	.136	32	.142	.943	32	.093

a. Lilliefors Significance Correction

<u>Inference</u>

From Kolmogorov –Smirnov's test we see p-value is 0.142 > 0.05 and from Shapiro-Wilk's test the p-value is 0.093 > 0.05, hence we fail to reject our null hypothesis at 5% level of significance and conclude that the sample is from normal population.

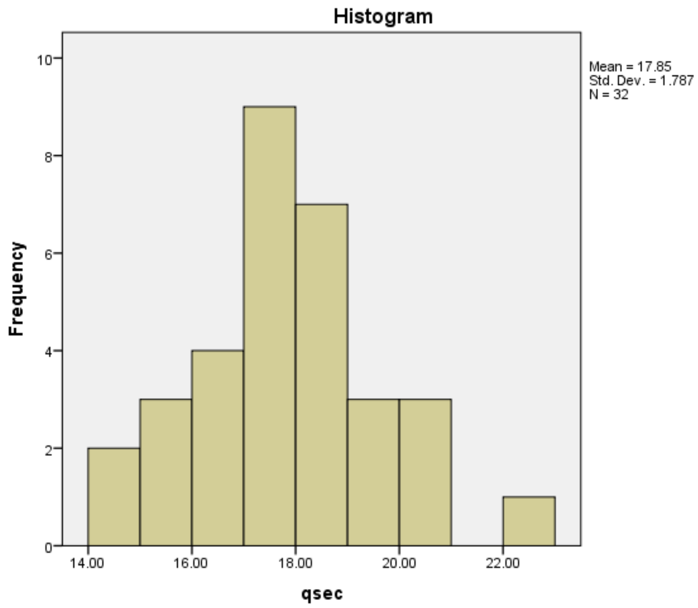
7. 1/4 Mile Time

This variable represented by **Qsec**. It is a continuous variable.

Descriptive Statistics

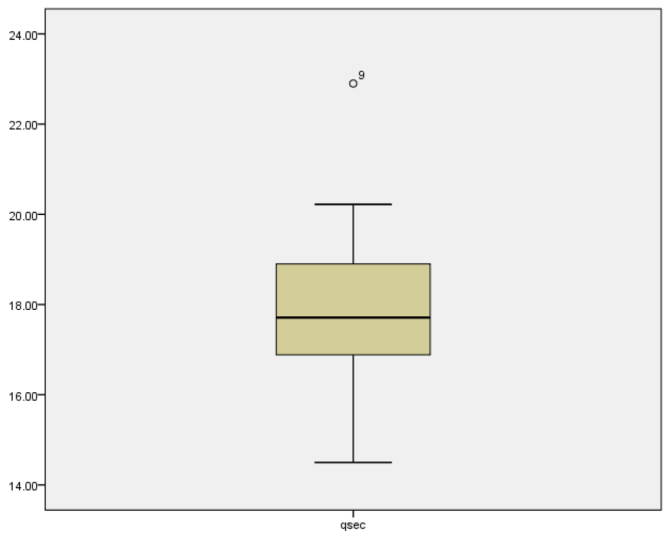
			Statistic	Std. Error
qsec	Mean		17.8488	.31589
	95% Confidence Interval for Mean	Lower Bound	17.2045	
		Upper Bound	18.4930	
	5% Trimmed Mean		17.8079	
	Median		17.7100	
	Variance		3.193	
	Std. Deviation		1.78694	
	Minimum		14.50	
	Maximum		22.90	
	Range		8.40	
	Interquartile Range		2.02	
	Skewness		.406	.414
	Kurtosis		.865	.809

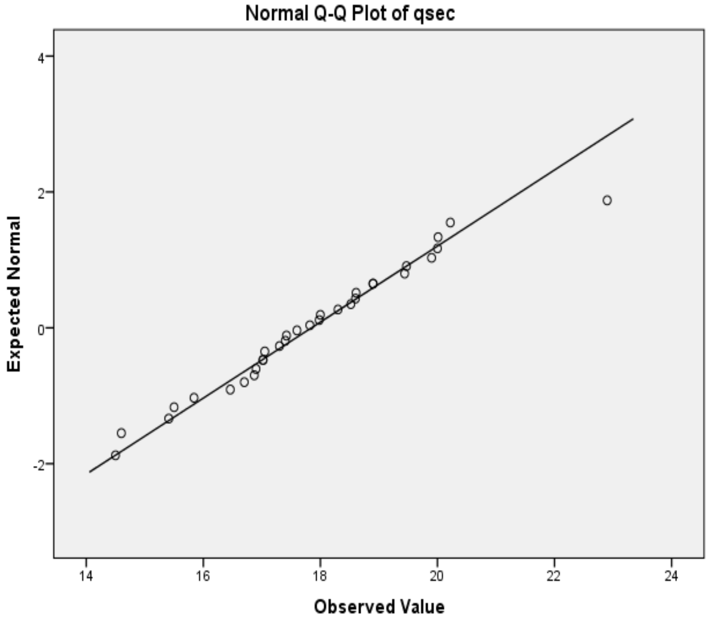
We set our hypotheses as:
H_0: The sample is from a normal population.
H_1: The sample is **NOT** from a normal population.




```
qsec Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2.00       14 .  56
     3.00       15 .  458
     4.00       16 .  4789
     9.00       17 .  000344689
     7.00       18 .  0356699
     3.00       19 .  449
     3.00       20 .  002
     1.00 Extremes    (>=22.9)

 Stem width:    1.00
 Each leaf:        1 case(s)
```

Box Plot





Inference

From histogram and stem and leaf plot we see that our data is pretty much normal. However from Box-Plot we observe that we have an **outlier** in our data. However these are only crude measures we will perform more sophisticated techniques available to us like **Kolmogorov-Smirnov Test** and **Shapiro-Wilk Test** in the next section to test the normality of the data.

Testing for Normality:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
qsec	.073	32	.200*	.973	32	.594

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Inference

From Kolmogorov –Smirnov's test we see p-value is 0.200 > 0.05 and from Shapiro-Wilk's test the p-value is 0.594 > 0.05, hence we fail to reject our null hypothesis at 5% level of significance and conclude that the sample is from normal population.

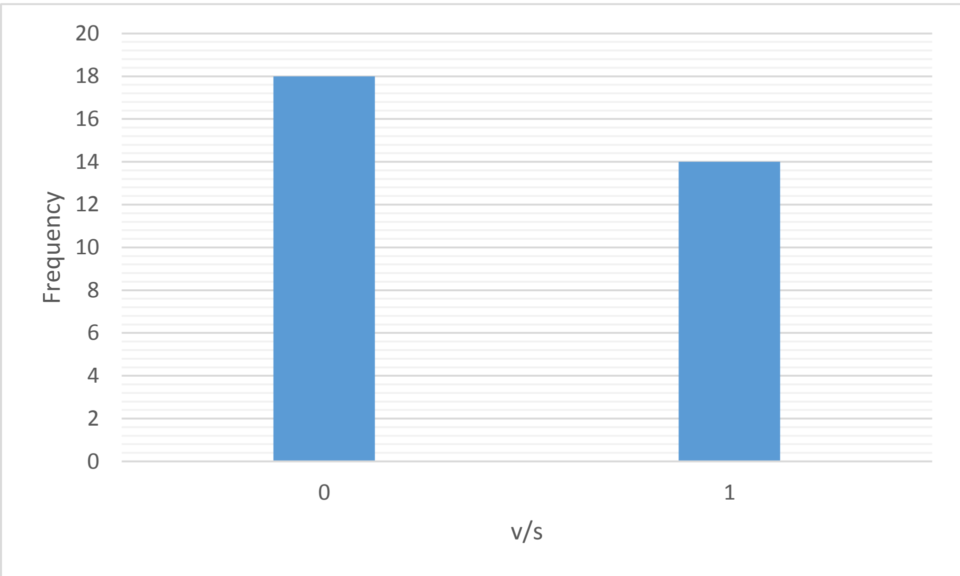
8. **V/S**

This variable is represented by 1/4 mile time **v/s**. It an ordinal variable.

Descriptive Statistics

			Statistic	Std. Error
vs	Mean		.44	.089
	95% Confidence Interval for Mean	Lower Bound	.26	
		Upper Bound	.62	
	5% Trimmed Mean		.43	
	Median		.00	
	Variance		.254	
	Std. Deviation		.504	
	Minimum		0	
	Maximum		1	
	Range		1	
	Interquartile Range		1	
	Skewness		.265	.414
	Kurtosis		-2.063	.809

Histogram



<u>Inference</u>

Histogram shows that ¼ mile time was not available for most of the cars.

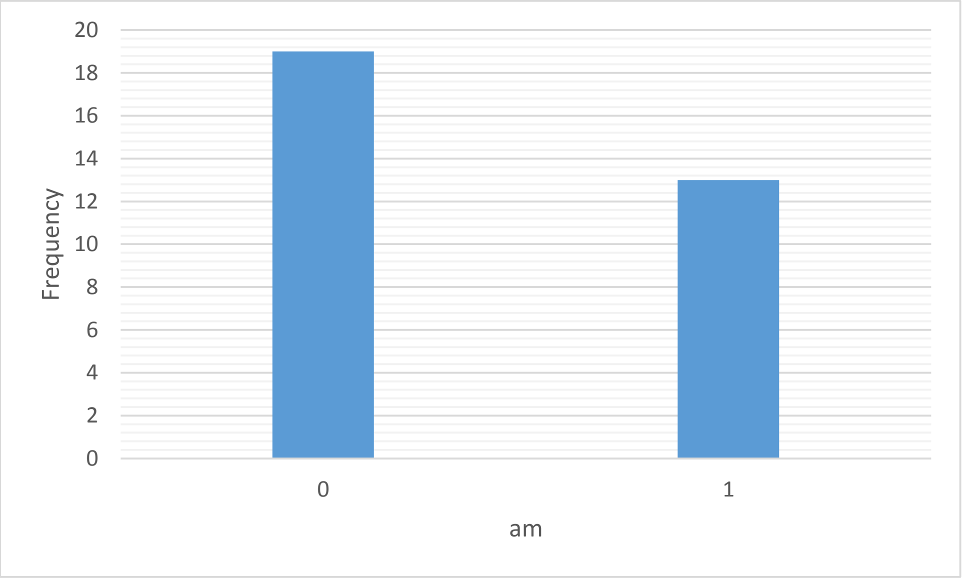
9. **Transmission**

This variable is represented by **am** (0 = automatic, 1 = manual). It is an ordinal variable.

Descriptive Statistics

			Statistic	Std. Error
am	Mean		.41	.088
	95% Confidence Interval for Mean	Lower Bound	.23	
		Upper Bound	.59	
	5% Trimmed Mean		.40	
	Median		.00	
	Variance		.249	
	Std. Deviation		.499	
	Minimum		0	
	Maximum		1	
	Range		1	
	Interquartile Range		1	
	Skewness		.401	.414
	Kurtosis		-1.967	.809

Histogram



<u>Inference</u>

We see that most of the cars have Automatic transmission configuration.

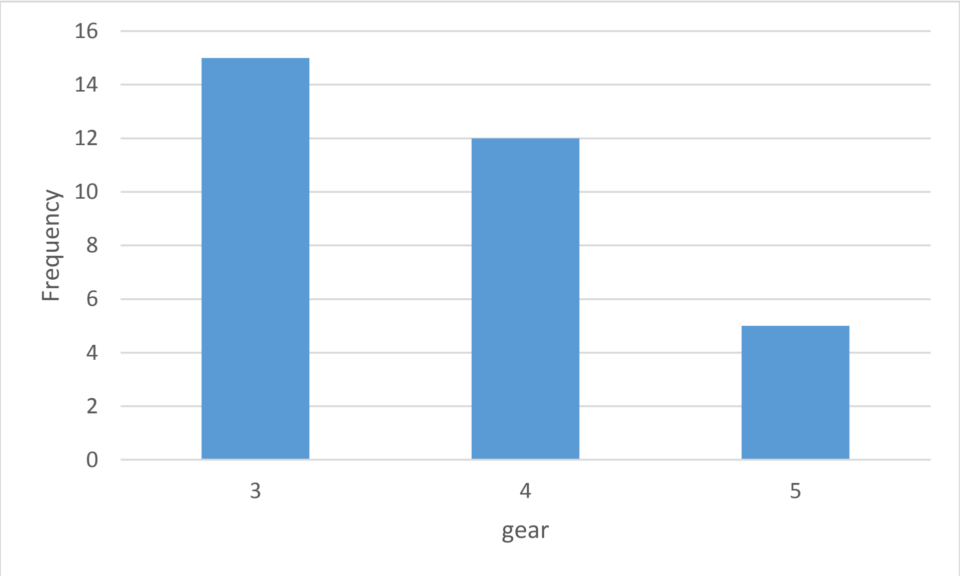
10. Number of Forward Gears

This variable is represented by **gear**. It is an ordinal variable.

Descriptive Statistics

			Statistic	Std. Error
gear	Mean		3.69	.130
	95% Confidence Interval for Mean	Lower Bound	3.42	
		Upper Bound	3.95	
	5% Trimmed Mean		3.65	
	Median		4.00	
	Variance		.544	
	Std. Deviation		.738	
	Minimum		3	
	Maximum		5	
	Range		2	
	Interquartile Range		1	
	Skewness		.582	.414
	Kurtosis		-.895	.809

Histogram



<u>Inference</u>

We see that most of the cars have 3 forward gear configuration.

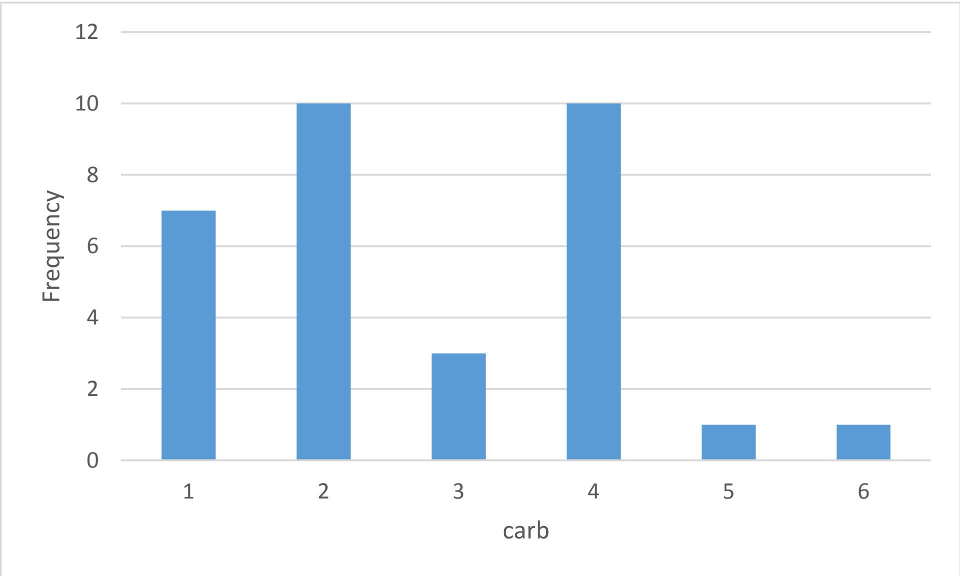
11. Number of carburettor

This variable is represented by **carb**. It is an ordinal variable.

Descriptive Statistics

			Statistic	Std. Error
carb	Mean		2.81	.286
	95% Confidence Interval for Mean	Lower Bound	2.23	
		Upper Bound	3.39	
	5% Trimmed Mean		2.67	
	Median		2.00	
	Variance		2.609	
	Std. Deviation		1.615	
	Minimum		1	
	Maximum		8	
	Range		7	
	Interquartile Range		2	
	Skewness		1.157	.414
	Kurtosis		2.020	.809

Histogram



<u>Inference</u>

We see that most of the cars have either 2 or 4 carburettors configuration.

EDA for Discrete Variables (Multivariate)

Correlations

			cyl	vs	am	Gear	carb
Spearman's rho	cyl	Correlation Coefficient	1.000	-.814**	-.522**	-.564**	.580**
		Sig. (2-tailed)	.	.000	.002	.001	.001
		N	32	32	32	32	32
	vs	Correlation Coefficient	-.814**	1.000	.168	.283	-.634**
		Sig. (2-tailed)	.000	.	.357	.117	.000
		N	32	32	32	32	32
	am	Correlation Coefficient	-.522**	.168	1.000	.808**	-.064
		Sig. (2-tailed)	.002	.357	.	.000	.726
		N	32	32	32	32	32
	gear	Correlation Coefficient	-.564**	.283	.808**	1.000	.115
		Sig. (2-tailed)	.001	.117	.000	.	.531
		N	32	32	32	32	32
	carb	Correlation Coefficient	.580**	-.634**	-.064	.115	1.000
		Sig. (2-tailed)	.001	.000	.726	.531	.
		N	32	32	32	32	32

**. Correlation is significant at the 0.01 level (2-tailed).

<u>Inference:</u>

The table below shows which pair of variables are highly correlated in either direction.

Variable	**Cyl**	**Vs**	**Am**	**Gear**	**carb**
Max +ve correlation with	Carb	Gear	Gear	Am	Cyl
Max –ve correlation with	vs	Cyl	Cyl	Cyl	vs

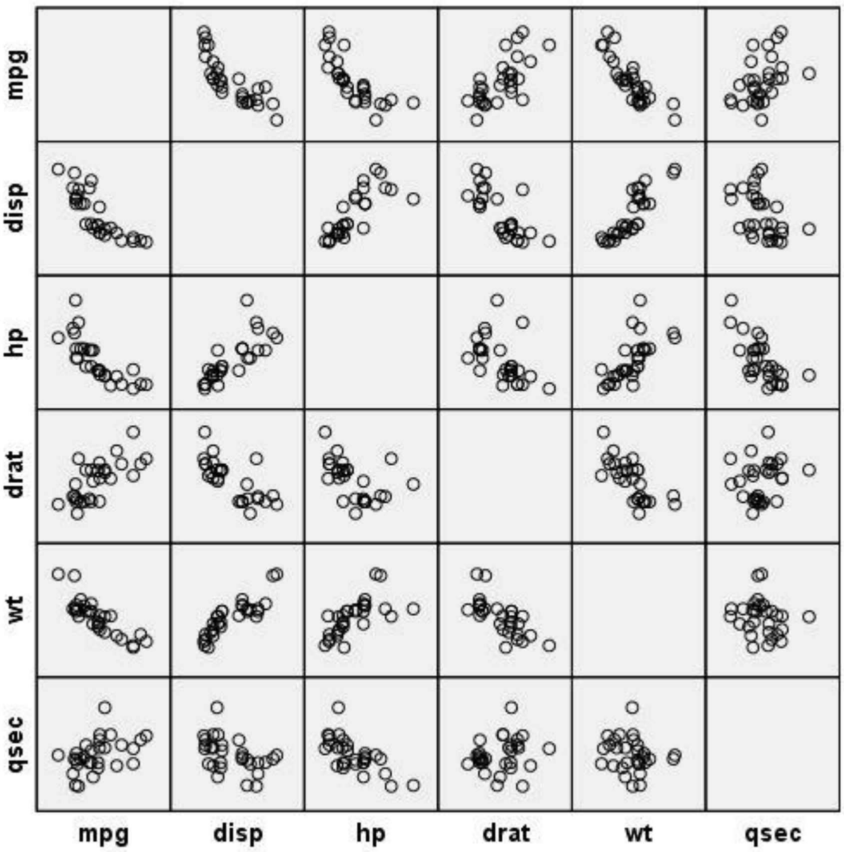
EDA for Continuous Variables (Multivariate)

Correlations

		mpg	disp	hp	drat	wt	Qsec
mpg	Pearson Correlation	1	-.826**	-.776**	.681**	-.868**	.419*
	Sig. (2-tailed)		.000	.000	.000	.000	.017
	N	32	29	32	32	32	32
disp	Pearson Correlation	-.826**	1	.793**	-.746**	.870**	-.458*
	Sig. (2-tailed)	.000		.000	.000	.000	.012
	N	29	29	29	29	29	29
hp	Pearson Correlation	-.776**	.793**	1	-.449**	.659**	-.708**
	Sig. (2-tailed)	.000	.000		.010	.000	.000
	N	32	29	32	32	32	32
drat	Pearson Correlation	.681**	-.746**	-.449**	1	-.712**	.091
	Sig. (2-tailed)	.000	.000	.010		.000	.620
	N	32	29	32	32	32	32
wt	Pearson Correlation	-.868**	.870**	.659**	-.712**	1	-.175
	Sig. (2-tailed)	.000	.000	.000	.000		.339
	N	32	29	32	32	32	32
qsec	Pearson Correlation	.419*	-.458*	-.708**	.091	-.175	1
	Sig. (2-tailed)	.017	.012	.000	.620	.339	
	N	32	29	32	32	32	32

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).



Inference:

The table below shows which pair of variables are highly correlated in either direction.

Variable	Maximum +ve correlation	Minimum –ve correlation
Mpg	Drat	Wt
Disp	Wt	Mpg
Hp	Disp	Mpg
Drat	Mpg	Disp
Wt	Disp	Mpg
qsec	Mpg	hp

Discrete vs Continuous Variable

In this section we try to understand relationship between discrete and continuous variable. For this we consider 2 discrete variables "Number of cylinders" and "No of forward gears" and 2 continuous variables "Miles per gallon" and "Displacement". We will compute Spearman's Rank correlation and analyse the results.

Correlation Table

			cyl	gear	mpg	disp
Spearman's rho	cyl	Correlation Coefficient	1.000	-.564**	-.911**	.927**
		Sig. (2-tailed)	.	.001	.000	.000
		N	32	32	32	29
	gear	Correlation Coefficient	-.564**	1.000	.543**	-.546**
		Sig. (2-tailed)	.001	.	.001	.002
		N	32	32	32	29
	mpg	Correlation Coefficient	-.911**	.543**	1.000	-.910**
		Sig. (2-tailed)	.000	.001	.	.000
		N	32	32	32	29
	disp	Correlation Coefficient	.927**	-.546**	-.910**	1.000
		Sig. (2-tailed)	.000	.002	.000	.
		N	29	29	29	29

**. Correlation is significant at the 0.01 level (2-tailed).

Inference

From the table we see there is significant correlation between some variables. No. of cylinders in a car is highly correlated with the displacement. However No. of cylinders in a car is oppositely correlated with the miles that a car covers with a gallon of fuel i.e. with more cylinders in car, miles that a car covers with a gallon of fuel decreases.