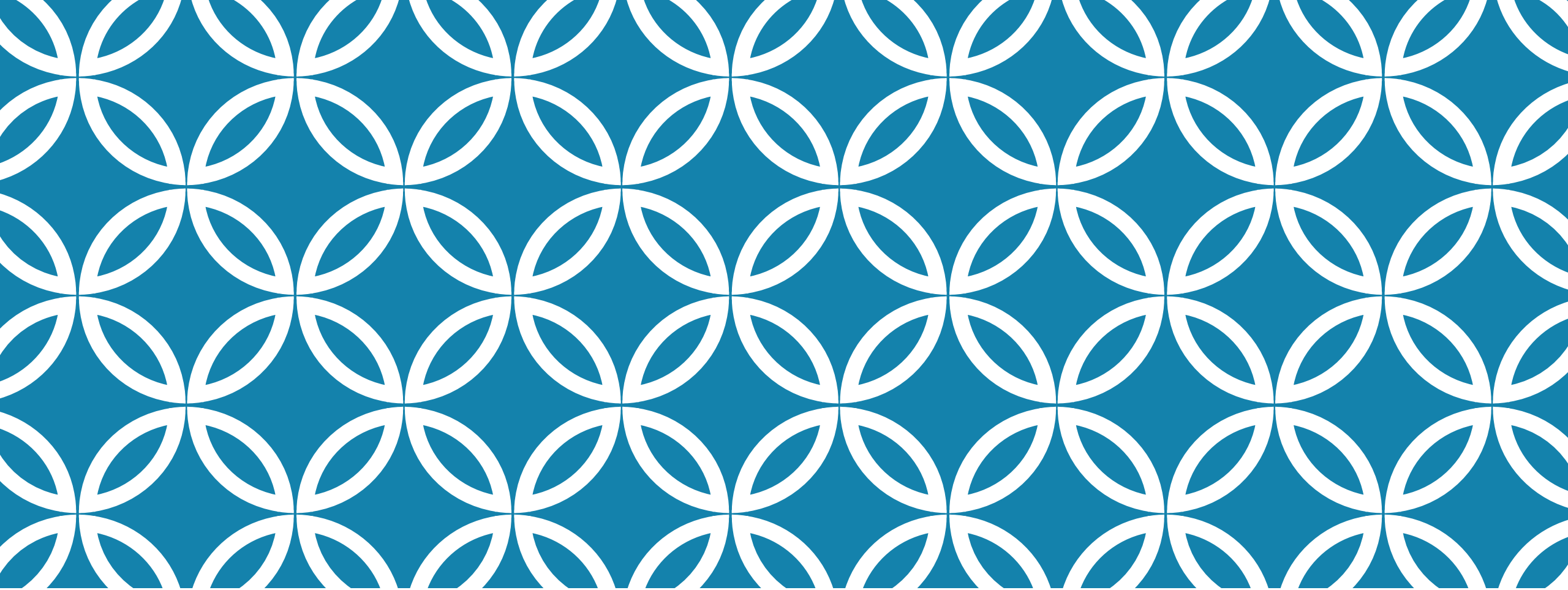


UNIVARIATE ANALYSIS

(DESCRIPTIVE STATISTIC ANALYSIS)

Maryam Najimigoshtasb

Meta Data

These dataset needs to be completed in order to be more clear. There some pints as following which needs to be considered.

1)- The unit of measurements of some variables are missing which are highlighted in the table.

2) The variable number 14 and 15 is not explanatory. Please indicate what they stand for?

3)- There some missing value for some attributes.

4) As our data set is small having this number of missing value is not recommended , I was wondering that if it I possible replace the missing value with mean or other sufficient values.

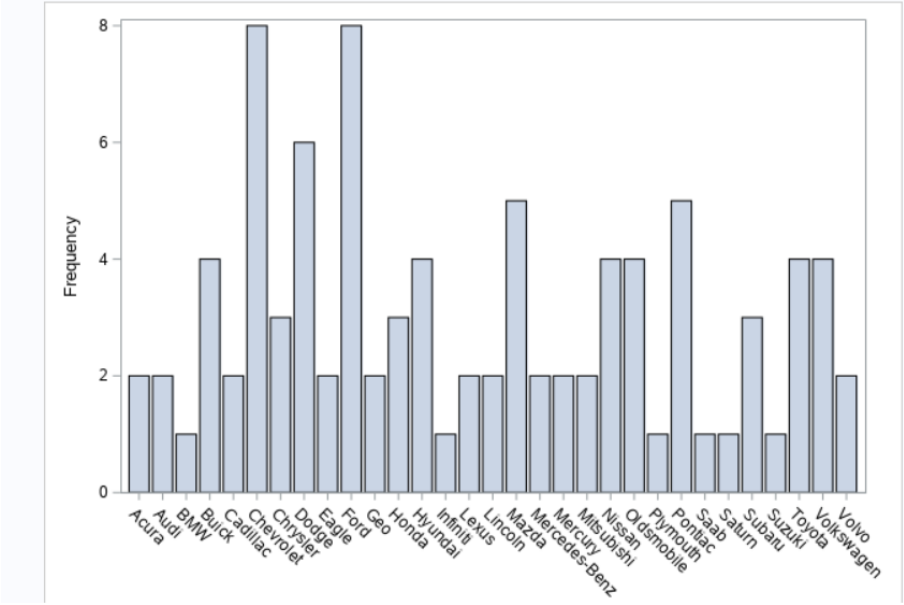
	Variable Name	Description	Type of variable	uniqe	primary key	missing value	Number of missing Value	informat
1	Manufacturer	32 distinct Automobile Manufactures in mentioned in this column	Nominal(Character)	No	No	No	0	$13.
2	Model	93 distinct Models of the Vehicles	Nominal(Character)	Yes	No	No	0	$14.
3	Category	6 distict category of cars interms of their size	Nominal(Character)	No	No	No	0	$7.
4	Min_Price	Minume Price of each car Is menioned here, **the price currency is not mentioned**	Number(continues)	No	No	Yes	14	
5	Mid_Price	Adverage Price of each car Is menioned here , mean of Min and Max price considering the one decimal point, (min_price+max_price)/2, 12 cars have the same avrage price, **the price currency is not mentioned**	Number(continues)	No	No	Yes	12	
6	Max_Price	Maximun Price of each car Is menioned here , **the price currency is not mentioned**	Number(continues)	No	No	No	0	
7	City_Fuel	The amount feul it burns in the city **(Missing Unit of Measurment)**	Number(Disterict)	NO	No	No	0	
8	Hwy_Fuel	The amount feul it burns in the Highway **(Missing Unit of Measurment)**	Number(Disterict)	NO	No	No	0	
9	Air_Bags	The number of Air Bag which is exist in the Car, 0 = none, 1 = driver only, 2 = driver & passenger	Ordinal(categorical)	NO	No	No	0	
10	Drive_Train	It Is Drive train type 0 = rear wheel drive, 1 = front wheel drive, 2 = all wheel drive	Ordinal(categorical)	NO	No	No	0	
11	Cylinders	The number of Cylinders	Number(Disterict)	NO	No	yes	1	
12	Engine_Size	The size of engin **(Missing Unit of Measurment)**	Number(continious)	No	No	No	0	
13	Max_HP	maximum horsepower	Number(Disterict)	No	No	No	0	
14	Max_HP_RPM	**The name is not standing out the concept or the meaning of variable**	Number(Disterict)	No	No	No	0	
15	RPM_high	**The name is not standing out the concept or the meaning of variable**	Number(Disterict)	No	No	No	0	
16	Manual	Manual Transmition is available yes=1, no=0	Nominal(Character)	No	No	No	0	
17	Fuel_Tank	Fuel Tank Capacity **(Missing Unit of Measurment)**	Number(continues)	No	No	No	0	
18	Passenger	Number of Passenger	Number(Disterict)	No	No	No	0	
19	Length	length of car**(Missing Unit of Measurment)**	Number(continious)	No	No	No	0	
20	Wheel_Base	wheel Base	Number(Disterict)	No	No	No	0	
21	Width	width of car**(Missing Unit of Measurment)**	Number(continious)	No	No	No	0	
22	U_Turn_Diam	u turn space **(Missing Unit of Measurment)**	Number(continious)	No	No	No	0	
23	Rear_Room	rear seat room **(Missing Unit of Measurment)**	Number(continious)	No	No	yes	2	
24	Luggage	**it is not clear it is number of luggage (not logical) or what?(Missing Unit of Measurment)**	Number(Disterict)	No	No	Yes	11	
25	Weight	weight of car **(Missing Unit of Measurment)**	Number(continious)	No	No	No	0	
26	Domestic	the car is manufactured in us yes =1 no =0	Ordinal(categorical)	No	No	No	0	

Univariate Analysis

a) Nominal variables

for these variable we can not do much about Univariate Analysis in Categorical variable unless checking the variable in terms of cleaning or finding missing variable.

Manufacturer: There is no any missing value

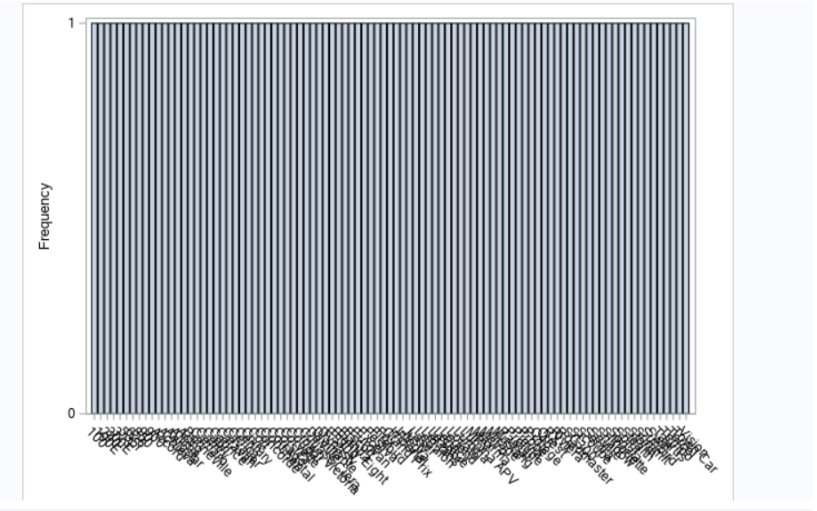


Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

Manufacturer				
Manufacturer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Acura	2	2.15	2	2.15
Audi	2	2.15	4	4.30
BMW	1	1.08	5	5.38
Buick	4	4.30	9	9.68
Cadillac	2	2.15	11	11.83
Chevrolet	8	8.60	19	20.43
Chrysler	3	3.23	22	23.66
Dodge	6	6.45	28	30.11
Eagle	2	2.15	30	32.26
Ford	8	8.60	38	40.86
Geo	2	2.15	40	43.01
Honda	3	3.23	43	46.24
Hyundai	4	4.30	47	50.54
Infiniti	1	1.08	48	51.61
Lexus	2	2.15	50	53.76
Lincoln	2	2.15	52	55.91
Mazda	5	5.38	57	61.29
Mercedes-Benz	2	2.15	59	63.44
Mercury	2	2.15	61	65.59
Mitsubishi	2	2.15	63	67.74
Nissan	4	4.30	67	72.04
Oldsmobile	4	4.30	71	76.34
Plymouth	1	1.08	72	77.42
Pontiac	5	5.38	77	82.80
Saab	1	1.08	78	83.87
Saturn	1	1.08	79	84.95
Subaru	3	3.23	82	88.17
Suzuki	1	1.08	83	89.25
Toyota	4	4.30	87	93.55
Volkswagen	4	4.30	91	97.85
Volvo	2	2.15	93	100.00

Model:
There is no any missing value.
All variable are unique .



Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

Model				
Model	Frequency	Percent	Cumulative Frequency	Cumulative Percent
100	1	1.08	1	1.08
190E	1	1.08	2	2.15
240	1	1.08	3	3.23
300E	1	1.08	4	4.30
323	1	1.08	5	5.38
535i	1	1.08	6	6.45
626	1	1.08	7	7.53
850	1	1.08	8	8.60
90	1	1.08	9	9.68
900	1	1.08	10	10.75
Accord	1	1.08	11	11.83
Achieva	1	1.08	12	12.90
Aerostar	1	1.08	13	13.98
Altima	1	1.08	14	15.05
Astro	1	1.08	15	16.13
Bonneville	1	1.08	16	17.20
Camaro	1	1.08	17	18.28
Camry	1	1.08	18	19.35
Capri	1	1.08	19	20.43
Caprice	1	1.08	20	21.51
Caravan	1	1.08	21	22.58
Cavalier	1	1.08	22	23.66
Celica	1	1.08	23	24.73
Century	1	1.08	24	25.81
Civic	1	1.08	25	26.88
Colt	1	1.08	26	27.96
Concorde	1	1.08	27	29.03
Continental	1	1.08	28	30.11
Corrado	1	1.08	29	31.18
Corsica	1	1.08	30	32.26
Corvette	1	1.08	31	33.33
Cougar	1	1.08	32	34.41
Crown Victoria	1	1.08	33	35.48
Cutlass Ciera	1	1.08	34	36.56

Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

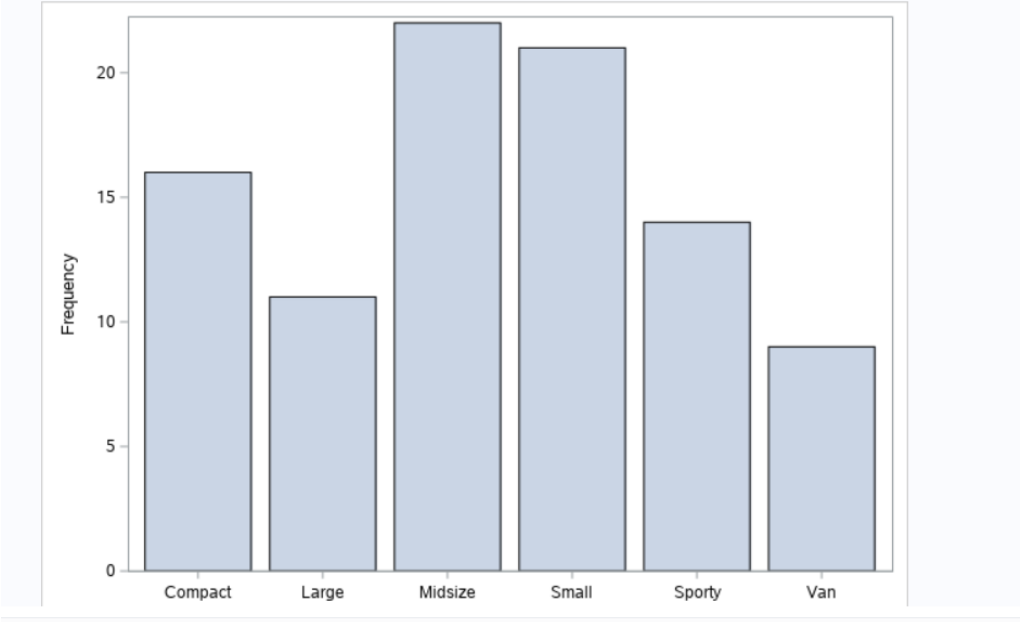
Model				
Model	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DeVille	1	1.08	35	37.63
Diamante	1	1.08	36	38.71
Dynasty	1	1.08	37	39.78
ES300	1	1.08	38	40.86
Eighty-Eight	1	1.08	39	41.94
Elantra	1	1.08	40	43.01
Escort	1	1.08	41	44.09
Eurovan	1	1.08	42	45.16
Excel	1	1.08	43	46.24
Festiva	1	1.08	44	47.31
Firebird	1	1.08	45	48.39
Fox	1	1.08	46	49.46
Grand Prix	1	1.08	47	50.54
Imperial	1	1.08	48	51.61
Integra	1	1.08	49	52.69
Justy	1	1.08	50	53.76
Laser	1	1.08	51	54.84
LeBaron	1	1.08	52	55.91
LeMans	1	1.08	53	56.99
LeSabre	1	1.08	54	58.06
Legacy	1	1.08	55	59.14
Legend	1	1.08	56	60.22
Loyale	1	1.08	57	61.29
Lumina	1	1.08	58	62.37
Lumina APV	1	1.08	59	63.44
MPV	1	1.08	60	64.52
Maxima	1	1.08	61	65.59
Metro	1	1.08	62	66.67
Mirage	1	1.08	63	67.74
Mustang	1	1.08	64	68.82
Passat	1	1.08	65	69.89
Prelude	1	1.08	66	70.97
Previa	1	1.08	67	72.04
Probe	1	1.08	68	73.12

Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

Model				
Model	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Protege	1	1.08	69	74.19
Q45	1	1.08	70	75.27
Quest	1	1.08	71	76.34
RX-7	1	1.08	72	77.42
Riviera	1	1.08	73	78.49
Roadmaster	1	1.08	74	79.57
SC300	1	1.08	75	80.65
SL	1	1.08	76	81.72
Scoupe	1	1.08	77	82.80
Sentra	1	1.08	78	83.87
Seville	1	1.08	79	84.95
Shadow	1	1.08	80	86.02
Silhouette	1	1.08	81	87.10
Sonata	1	1.08	82	88.17
Spirit	1	1.08	83	89.25
Stealth	1	1.08	84	90.32
Storm	1	1.08	85	91.40
Summit	1	1.08	86	92.47
Sunbird	1	1.08	87	93.55
Swift	1	1.08	88	94.62
Taurus	1	1.08	89	95.70
Tempo	1	1.08	90	96.77
Tercel	1	1.08	91	97.85
Town Car	1	1.08	92	98.92
Vision	1	1.08	93	100.00

Category:
There is no any missing value.



Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

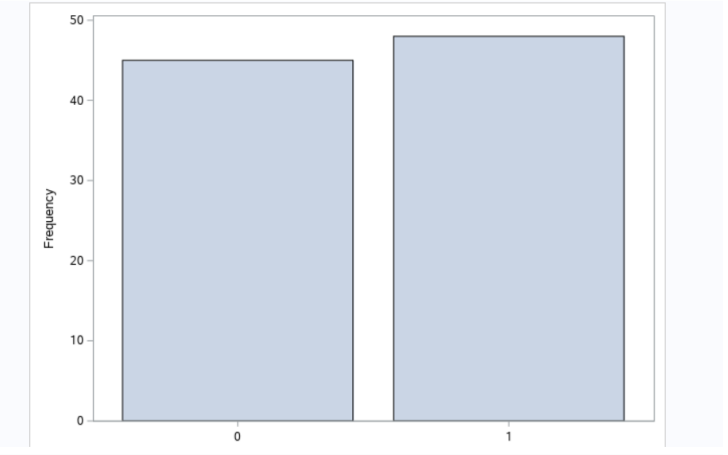
Category				
Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Compact	16	17.20	16	17.20
Large	11	11.83	27	29.03
Midsize	22	23.66	49	52.69
Small	21	22.58	70	75.27
Sporty	14	15.05	84	90.32
Van	9	9.68	93	100.00

There is no any missing values.

Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

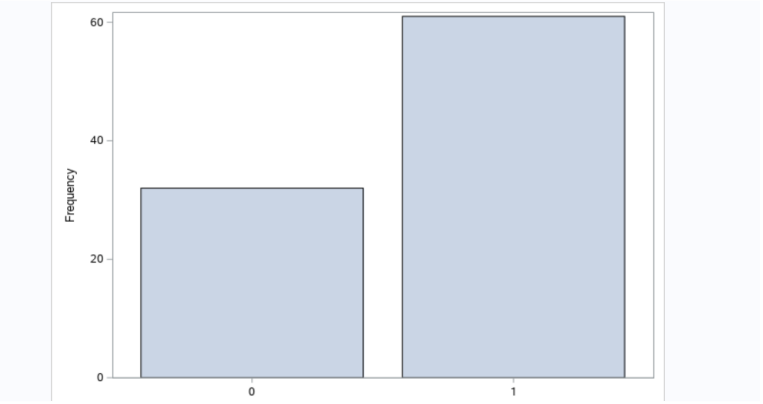
Domestic				
Domestic	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	45	48.39	45	48.39
1	48	51.61	93	100.00



Frequency of Qualitative Variables(Nominal) for Univariate Analysis

The FREQ Procedure

Manual				
Manual	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	32	34.41	32	34.41
1	61	65.59	93	100.00

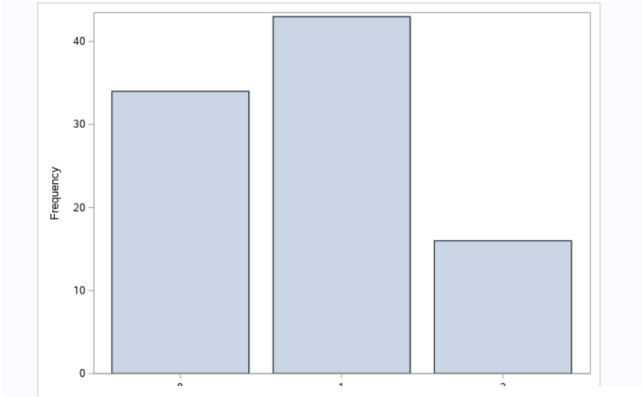


There is no any missing values.

Frequency of Qualitative Variables(Ordinal) for Univariate Analysis

The FREQ Procedure

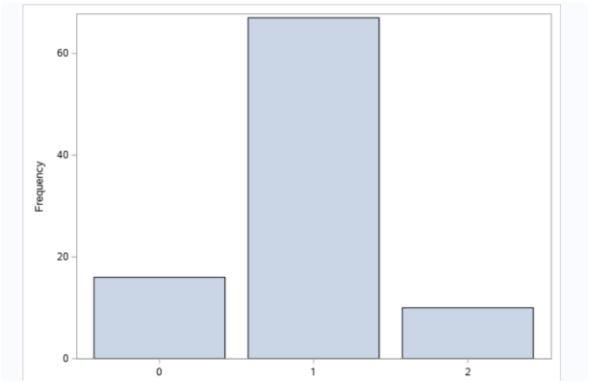
Air_Bags				
Air_Bags	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	34	36.56	34	36.56
1	43	46.24	77	82.80
2	16	17.20	93	100.00



Frequency of Qualitative Variables(Ordinal) for Univariate Analysis

The FREQ Procedure

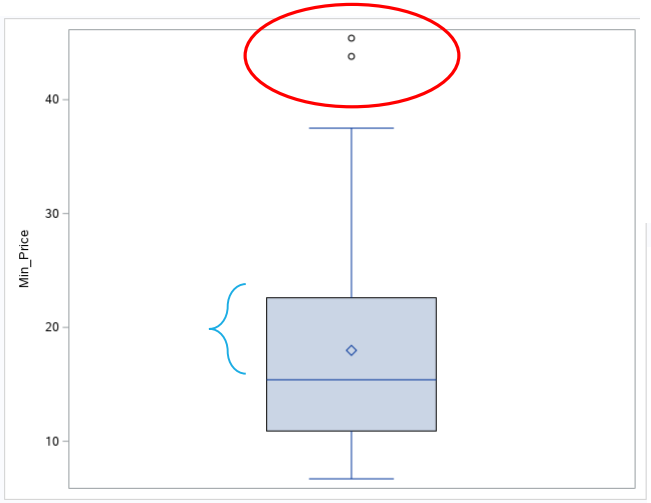
Drive_Train				
Drive_Train	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	16	17.20	16	17.20
1	67	72.04	83	89.25
2	10	10.75	93	100.00



MIN-PRICE

Looking at the figures we realize that mean Is greater that median. this means the data is spread out around the mean and highly likely there are outliers, but where is the outliers? Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is a little above one so we realized that data is skewed to the right moderately, Kurtosis is greater than zero which indicate that the distribution of data has a heavier tail. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there are a sudden jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Comparing N and number of observation in the outlier part we need either consider median or treat the outlier not by removing the data.

All this interpretation is completely clear in the boxplot and histogram of the data



Moments			
N	79	Sum Weights	79
Mean	17.9810127	Sum Observations	1420.5
Std Deviation	9.11123674	Variance	83.0146349
Skewness	1.01948628	Kurtosis	0.52496422
Uncorrected SS	32017.17	Corrected SS	6475.14152
Coeff Variation	50.6714327	Std Error Mean	1.02509422

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	17.54084	Pr > |t|	<.0001
Sign	M	39.5	Pr >= |M|	<.0001
Signed Rank	S	1580	Pr >= |S|	<.0001

Basic Statistical Measures			
Location		Variability	
Mean	17.98101	Std Deviation	9.11124
Median	15.40000	Variance	83.01463
Mode	.	Range	38.70000
		Interquartile Range	11.70000

Quantiles (Definition 5)	
Level	Quantile
100% Max	45.4
99%	45.4
95%	34.7
90%	33.0
75% Q3	22.6
50% Median	15.4
25% Q1	10.9
10%	7.9
5%	7.3
1%	6.7
0% Min	6.7

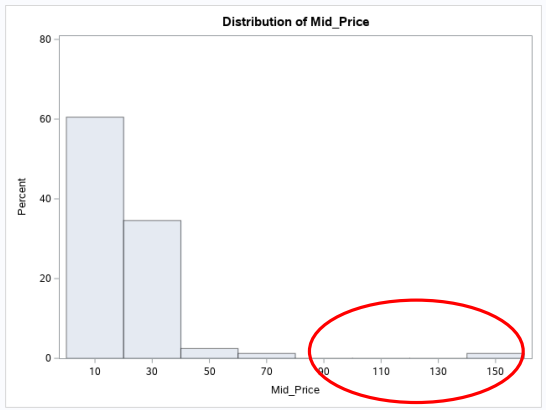
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
6.7	34	34.6	17
6.8	38	34.7	44
6.9	28	37.5	11
7.3	70	43.8	53
7.4	47	45.4	42

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	14	15.05	100.00

MID-PRICE

Looking at the figures we realize that mean Is greater that median. this means the data is spread out around the mean and highly likely there are outliers, but where is the outliers? Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is 5 so we realized that data is skewed to the right severely , Kurtosis is greater than zero which indicate that the distribution of data has a heavier tail. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there are a sudden jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Too treat the outlier we will remove the data here. Because it is small percentage of our data.

All this interpretation is completely clear in the boxplot and histogram of the data



The UNIVARIATE Procedure
Variable: Mid_Price (Mid_Price)

Moments			
N	81	Sum Weights	81
Mean	21.8148148	Sum Observations	1767
Std Deviation	17.6873621	Variance	312.842778
Skewness	5.15961146	Kurtosis	35.770306
Uncorrected SS	63574.2	Corrected SS	25027.4222
Coeff Variation	81.0795885	Std Error Mean	1.96526245

Basic Statistical Measures			
Location		Variability	
Mean	21.81481	Std Deviation	17.68736
Median	18.50000	Variance	312.84278
Mode	.	Range	143.60000
		Interquartile Range	13.60000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	11.1002	Pr > |t|	<.0001
Sign	M	40.5	Pr >= |M|	<.0001
Signed Rank	S	1660.5	Pr >= |S|	<.0001

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	12	12.90	100.00

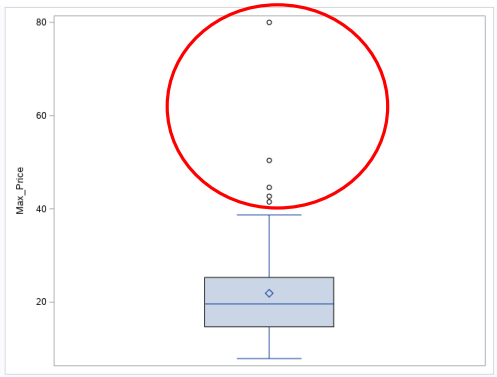
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
7.4	30	38.0	18
8.0	41	40.1	11
8.3	49	47.9	44
8.4	36	61.9	55
8.6	73	151.0	14

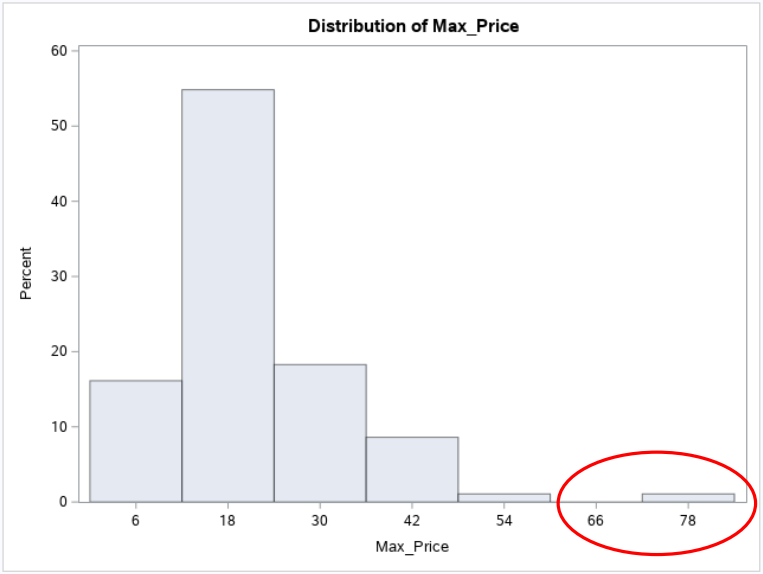
Quantiles (Definition 5)	
Level	Quantile
100% Max	151.0
99%	151.0
95%	38.0
90%	34.7
75% Q3	26.1
50% Median	18.5
25% Q1	12.5
10%	9.8
5%	8.6
1%	7.4
0% Min	7.4

MAX-PRICE

Looking at the figures we realize that mean Is greater that median. this means the data is spread out around the mean and highly likely there are outliers, but where is the outliers? Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is 2 so we realized that data is skewed to the right severely , Kurtosis is greater than zero which indicate that the distribution of data has a heavier tail. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there are a sudden jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Too treat the outlier we will remove the data here. Because it is small percentage of our data.

All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	21.8989247	Sum Observations	2036.6
Std Deviation	11.0304568	Variance	121.670977
Skewness	2.03385866	Kurtosis	7.4394262
Uncorrected SS	55793.08	Corrected SS	11193.7299
Coeff Variation	50.3698557	Std Error Mean	1.14380509

Basic Statistical Measures			
Location		Variability	
Mean	21.89892	Std Deviation	11.03046
Median	19.60000	Variance	121.67098
Mode	18.40000	Range	72.10000
		Interquartile Range	10.60000

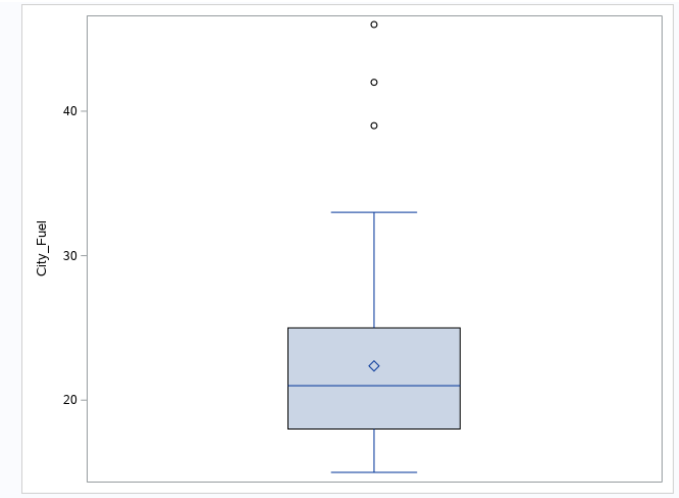
Quantiles (Definition 5)	
Level	Quantile
100% Max	80.0
99%	80.0
95%	41.5
90%	36.2
75% Q3	25.3
50% Median	19.6
25% Q1	14.7
10%	11.0
5%	9.5
1%	7.9
0% Min	7.9

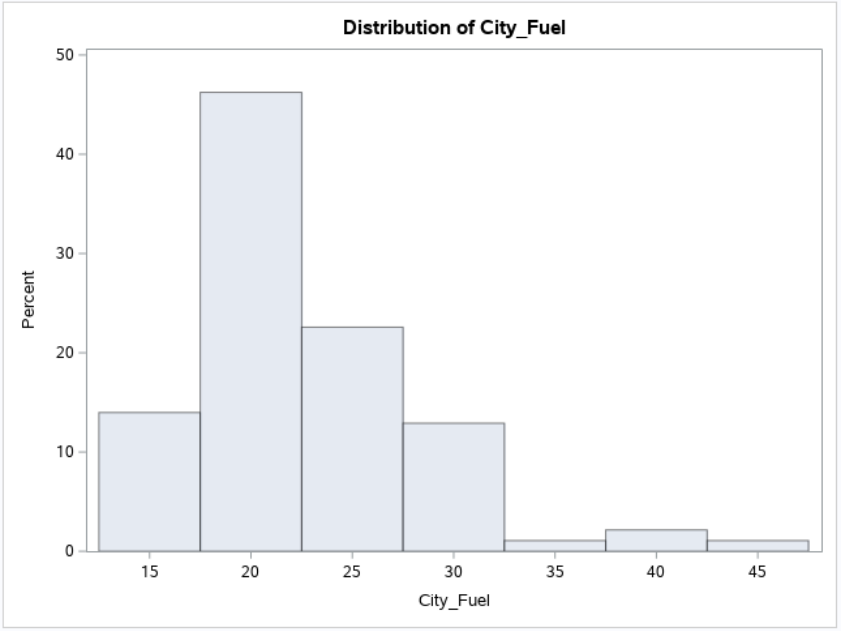
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	19.14568	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
7.9	31	41.5	19
9.1	53	42.7	11
9.2	44	44.6	4
9.5	88	50.4	48
9.5	80	80.0	59

City_Fuel

Looking at the figures we realize that mean Is a little bit greater that median. this means the data is spread out around the mean and highly likely there are outliers, but where is the outliers? Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is almost close to 2 so we realized that data is skewed to the right severely , Kurtosis is greater than zero which indicate that the distribution of data has a heavier tail. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there are a sudden jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Removing the data might not be a good idea here how ever there few here. We can treat them by taking log and compare then decide what to do.
All this interpretation is completely clear in the boxplot and histogram of the data





The UNIVARIATE Procedure
Variable: City_Fuel (City_Fuel)

Moments			
N	93	Sum Weights	93
Mean	22.3655914	Sum Observations	2080
Std Deviation	5.61981151	Variance	31.5822814
Skewness	1.70443015	Kurtosis	4.00430589
Uncorrected SS	49426	Corrected SS	2905.56989
Coeff Variation	25.1270419	Std Error Mean	0.58274731

Basic Statistical Measures			
Location		Variability	
Mean	22.36559	Std Deviation	5.61981
Median	21.00000	Variance	31.58228
Mode	18.00000	Range	31.00000
		Interquartile Range	7.00000

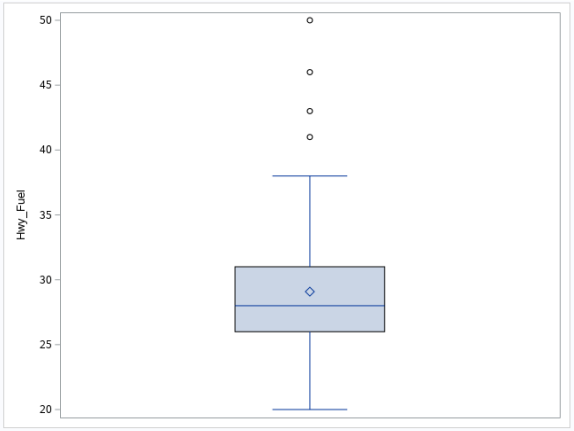
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	38.37957	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

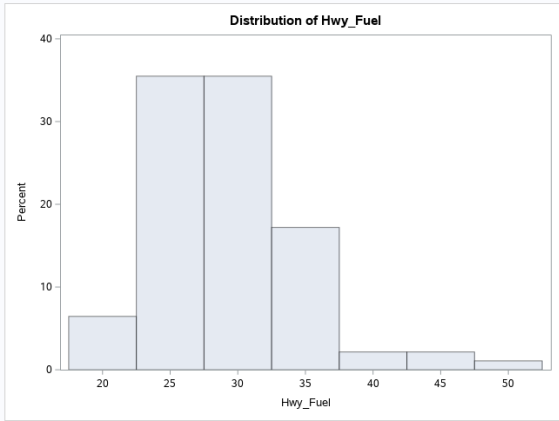
Quantiles (Definition 5)	
Level	Quantile
100% Max	46
99%	46
95%	32
90%	29
75% Q3	25
50% Median	21
25% Q1	18
10%	17
5%	16
1%	15
0% Min	15

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
15	2	32	89
15	1	33	90
16	5	39	91
16	4	42	92
16	3	46	93

Hwy_Fuel

Looking at the figures we realize that mean Is a little bit greater that median. this means the data is spread out around the mean and highly likely there are outliers, but where is the outliers? Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is 1.2 so we realized that data is skewed to the right severely , Kurtosis is greater than zero which indicate that the distribution of data has a heavier tail. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there are a sudden jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Removing the data might not be a good idea here how ever there few here. We can treat them by taking log and compare then decide what to do.
All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	29.0860215	Sum Observations	2705
Std Deviation	5.33172603	Variance	28.4273025
Skewness	1.22989674	Kurtosis	2.61420061
Uncorrected SS	81293	Corrected SS	2615.31183
Coeff Variation	18.330888	Std Error Mean	0.55287423

Basic Statistical Measures			
Location		Variability	
Mean	29.08602	Std Deviation	5.33173
Median	28.00000	Variance	28.42730
Mode	26.00000	Range	30.00000
		Interquartile Range	5.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	52.60875	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	50
99%	50
95%	38
90%	36
75% Q3	31
50% Median	28
25% Q1	26
10%	23
5%	22
1%	20
0% Min	20

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
20	36	38	79
20	17	41	73
21	89	43	83
21	26	46	42
22	87	50	39

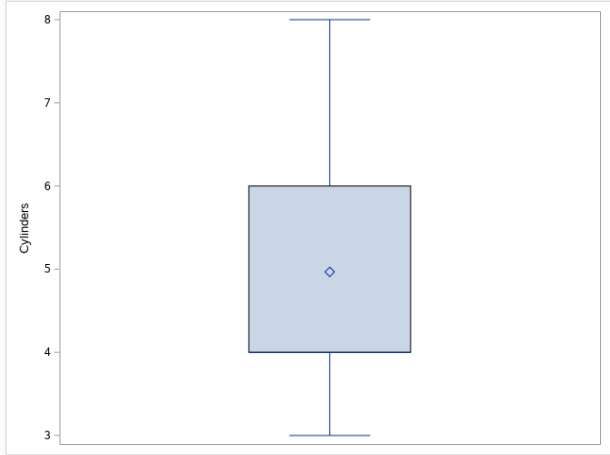
Cylinders

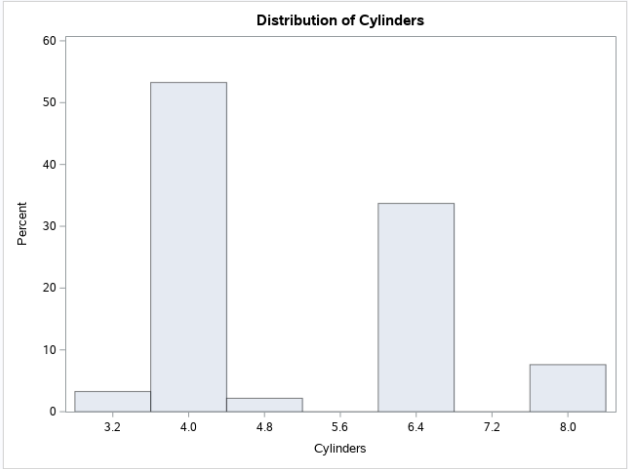
Respect to District type of variable

Looking at the figures we realize that mean, median and mode are very close to each other. this means the data is not spread out around the mean and highly likely there are no outliers.

Looking at skewness, data is skewed to the right lightly, Kurtosis shows light tail.
There is no sudden jump in the data by looking at the quartile and percentile .

All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	92	Sum Weights	92
Mean	4.9673913	Sum Observations	457
Std Deviation	1.30469218	Variance	1.70222169
Skewness	0.82019014	Kurtosis	-0.204359
Uncorrected SS	2425	Corrected SS	154.902174
Coeff Variation	26.265138	Std Error Mean	0.13602356

Basic Statistical Measures			
Location		Variability	
Mean	4.967391	Std Deviation	1.30469
Median	4.000000	Variance	1.70222
Mode	4.000000	Range	5.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	36.51861	Pr > |t|	<.0001
Sign	M	46	Pr >= |M|	<.0001
Signed Rank	S	2139	Pr >= |S|	<.0001

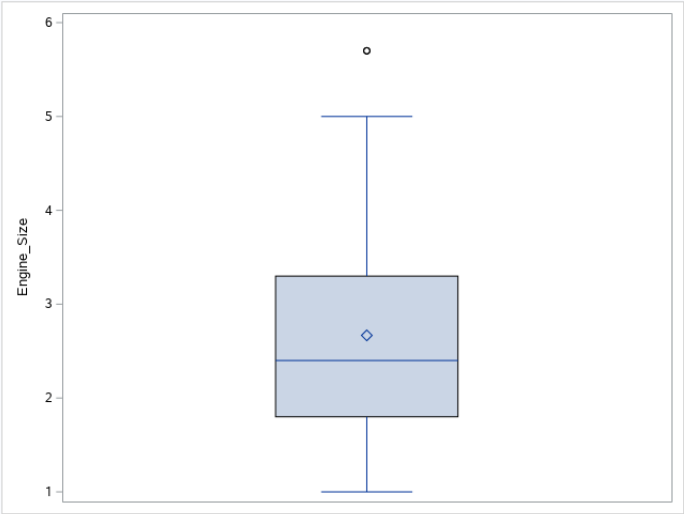
Quantiles (Definition 5)	
Level	Quantile
100% Max	8
99%	8
95%	8
90%	6
75% Q3	6
50% Median	4
25% Q1	4
10%	4
5%	4
1%	3
0% Min	3

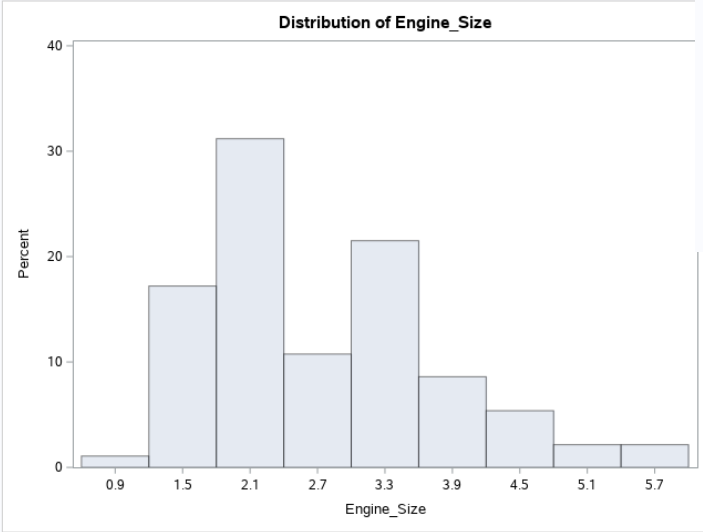
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
3	83	8	18
3	80	8	19
3	39	8	38
4	92	8	48
4	90	8	52

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	1	1.08	100.00

Engine_Size

Looking at the figures we realize that mean almost close to the median. It is a little bit higher that median. So highly likely there are outliers, outliers should be Out of mean+/- 3STD, also we are going to see the skewness, in the Normal distributions the Skewness should be 0 but here is .85 so we realized that data is skewed to the right lightly, Kurtosis is greater than zero which indicate that the distribution of data has a light heavy tail compare to normal distribution. Therefore, there is an outlier 100 percent, by looking at the quartile we will see there ia a small jump from 95 percent to 99 percentile of the data , so the outliers lie in this area of data , we need to go deeper to precisely see the outliers. Removing the data might not be a good idea here how ever there few here. We can treat them by taking log and compare then decide what to do.
All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	2.66774194	Sum Observations	248.1
Std Deviation	1.03736301	Variance	1.07612202
Skewness	0.85941842	Kurtosis	0.38102491
Uncorrected SS	760.87	Corrected SS	99.0032258
Coeff Variation	38.8854334	Std Error Mean	0.10756953

Basic Statistical Measures			
Location		Variability	
Mean	2.667742	Std Deviation	1.03736
Median	2.400000	Variance	1.07612
Mode	3.000000	Range	4.70000
		Interquartile Range	1.50000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	24.80016	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5.7
99%	5.7
95%	4.6
90%	3.8
75% Q3	3.3
50% Median	2.4
25% Q1	1.8
10%	1.5
5%	1.3
1%	1.0
0% Min	1.0

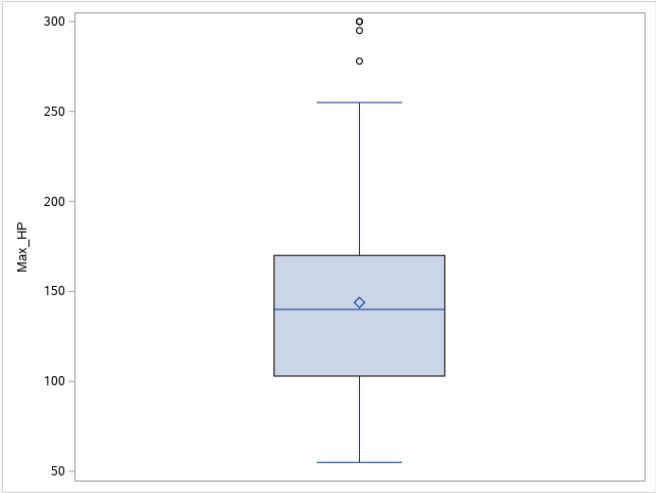
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1.0	39	4.6	52
1.2	80	4.9	10
1.3	83	5.0	18
1.3	57	5.7	8
1.3	31	5.7	19

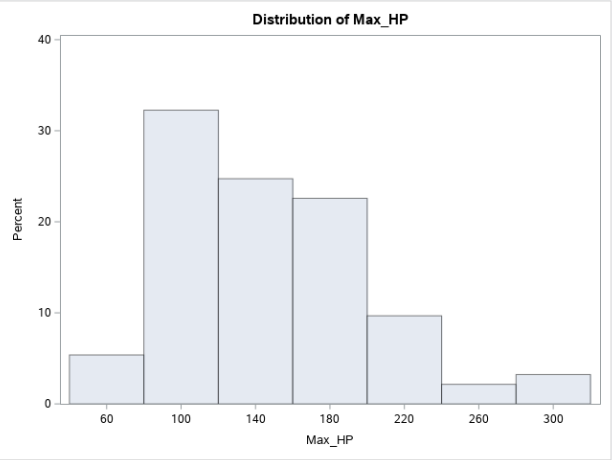
Max_HP

Respect to District type of variable

Looking at the figures we realize that mean and median are very close to each other. However, mode is far from them. So this is a sign the there might be an outlier. We can see from the skewness and kurtosis that data is skewed to the right and have heavy tail means that the number of data in the outlier might ne more. As we can see here compare o the number of observation 3 outlier is not small.

All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	143.827957	Sum Observations	13376
Std Deviation	52.3744095	Variance	2743.07878
Skewness	0.95172825	Kurtosis	1.11088259
Uncorrected SS	2176206	Corrected SS	252363.247
Coeff Variation	36.4146239	Std Error Mean	5.43097327

Basic Statistical Measures			
Location		Variability	
Mean	143.8280	Std Deviation	52.37441
Median	140.0000	Variance	2743
Mode	110.0000	Range	245.00000
		Interquartile Range	67.00000

Note: The mode displayed is the smallest of 2 modes with a count of 7.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	26.48291	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	300
99%	300
95%	255
90%	208
75% Q3	170
50% Median	140
25% Q1	103
10%	85
5%	74
1%	55
0% Min	55

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
55	39	255	57
63	31	278	48
70	83	295	11
73	80	300	19
74	73	300	28

Max_HP_RPM

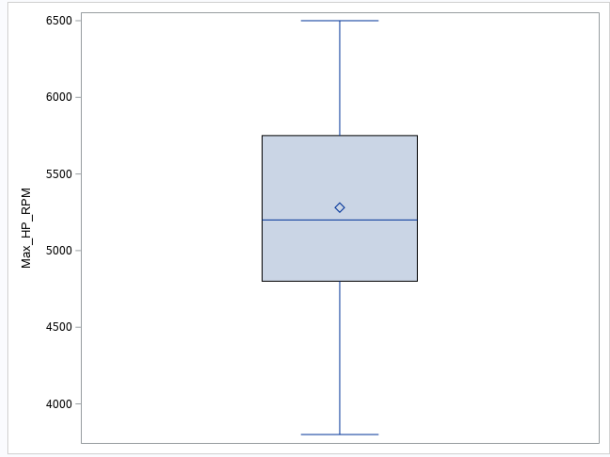
Respect to District type of variable

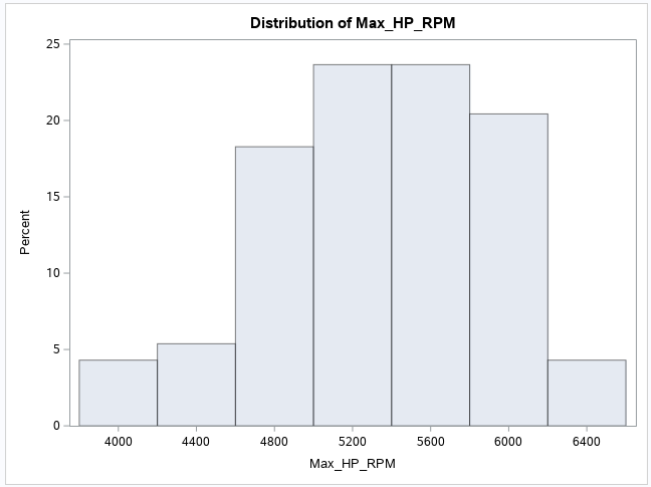
Looking at the figures we realize that mean, median and mode are very close to each other. this means the data is not spread out around the mean and highly likely there are no outliers.

Looking at skewness, data is skewed to the left very very light, Kurtosis shows small light tail.
There is no sudden jump in the data by looking at the quartile and percentile .

So there is no any outlier

All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	5280.64516	Sum Observations	491100
Std Deviation	596.73169	Variance	356088.71
Skewness	-0.2585327	Kurtosis	-0.409479
Uncorrected SS	2626085000	Corrected SS	32760161.3
Coeff Variation	11.3003558	Std Error Mean	61.8781937

Basic Statistical Measures			
Location		Variability	
Mean	5280.645	Std Deviation	596.73169
Median	5200.000	Variance	356089
Mode	6000.000	Range	2700
		Interquartile Range	950.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	85.33936	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	6500
99%	6500
95%	6000
90%	6000
75% Q3	5750
50% Median	5200
25% Q1	4800
10%	4600
5%	4200
1%	3800
0% Min	3800

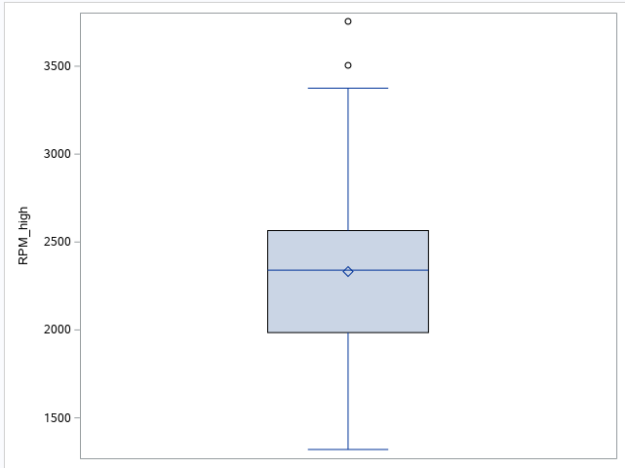
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
3800	61	6000	83
4000	17	6200	93
4000	8	6300	1
4100	10	6500	32
4200	38	6500	57

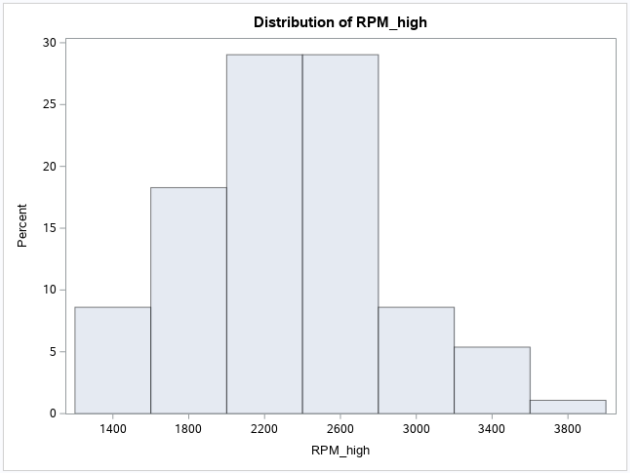
RPM_high

Respect to District type of variable

Looking at the figures we realize that mean and median are very close to each other. However, mode is far from them. So this is a sign the there might be an outlier. We can see from the skewness and kurtosis that data is skewed to the right lightly and have ticker tail than normal means that the number of data in the outlier might be in comparison to the N. As we can see here compare o the number of observation 2 outlier is not small. However we can ignore it.

All this interpretation is completely clear in the boxplot and histogram of the data





Moments			
N	93	Sum Weights	93
Mean	2332.2043	Sum Observations	216895
Std Deviation	496.506525	Variance	246518.73
Skewness	0.28154602	Kurtosis	0.22054386
Uncorrected SS	528523175	Corrected SS	22679723.1
Coeff Variation	21.2891523	Std Error Mean	51.4853283

Basic Statistical Measures			
Location		Variability	
Mean	2332.204	Std Deviation	496.50653
Median	2340.000	Variance	246519
Mode	1690.000	Range	2435
		Interquartile Range	580.00000

Note: The mode displayed is the smallest of 2 modes with a count of 3.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	45.29843	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

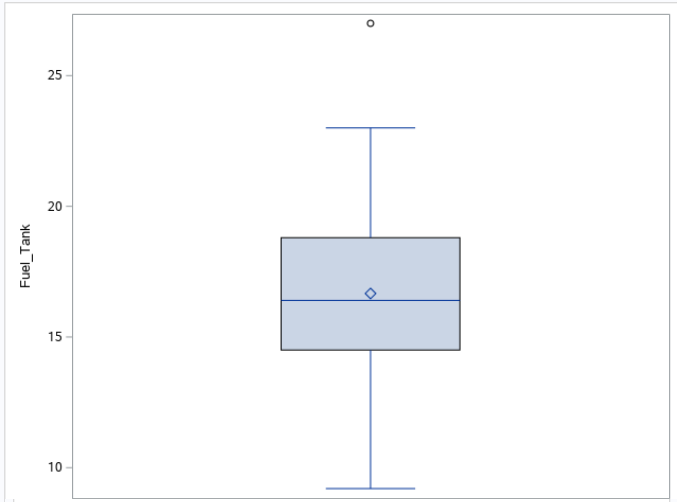
Quantiles (Definition 5)	
Level	Quantile
100% Max	3755
99%	3755
95%	3285
90%	2910
75% Q3	2565
50% Median	2340
25% Q1	1985
10%	1690
5%	1510
1%	1320
0% Min	1320

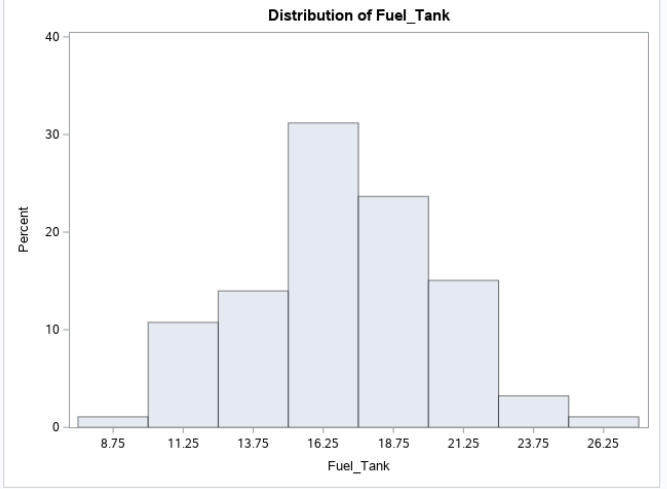
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1320	8	3285	23
1350	18	3360	83
1415	38	3375	81
1450	19	3505	84
1510	10	3755	39

Fuel_Tank

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	16.6645161	Sum Observations	1549.8
Std Deviation	3.27937047	Variance	10.7542707
Skewness	0.1081462	Kurtosis	0.1272065
Uncorrected SS	26816.06	Corrected SS	989.392903
Coeff Variation	19.678762	Std Error Mean	0.34005488

Basic Statistical Measures			
Location		Variability	
Mean	16.66452	Std Deviation	3.27937
Median	16.40000	Variance	10.75427
Mode	18.00000	Range	17.80000
		Interquartile Range	4.30000

Note: The mode displayed is the smallest of 2 modes with a count of 9.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	49.00537	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	27.0
99%	27.0
95%	21.1
90%	20.0
75% Q3	18.8
50% Median	16.4
25% Q1	14.5
10%	12.4
5%	11.1
1%	9.2
0% Min	9.2

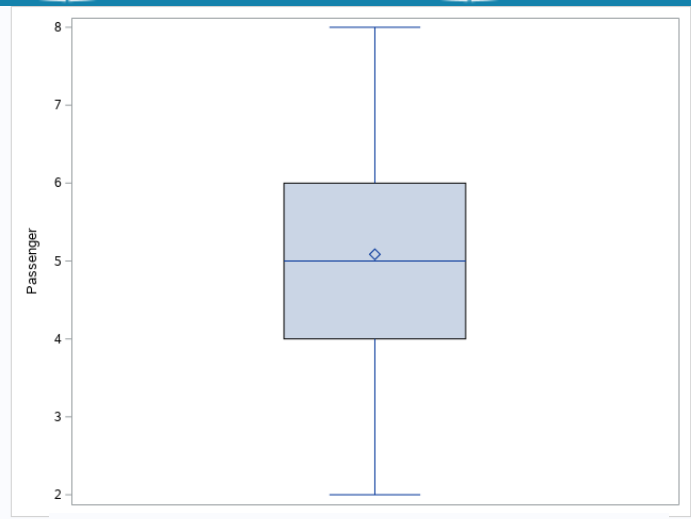
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
9.2	80	21.1	89
10.0	31	22.5	48
10.6	83	23.0	8
10.6	39	23.0	18
11.1	60	27.0	17

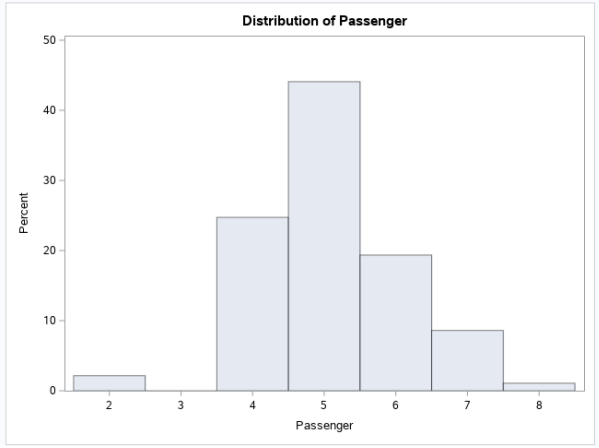
Passenger

Respect to District type of variable

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	5.08602151	Sum Observations	473
Std Deviation	1.03897853	Variance	1.07947639
Skewness	0.06251685	Kurtosis	0.9361926
Uncorrected SS	2505	Corrected SS	99.311828
Coeff Variation	20.4281191	Std Error Mean	0.10773705

Basic Statistical Measures			
Location		Variability	
Mean	5.086022	Std Deviation	1.03898
Median	5.000000	Variance	1.07948
Mode	5.000000	Range	6.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	47.20773	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

The FREQ Procedure

Passenger				
Passenger	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	2	2.15	2	2.15
4	23	24.73	25	26.88
5	41	44.09	66	70.97
6	18	19.35	84	90.32
7	8	8.60	92	98.92
8	1	1.08	93	100.00

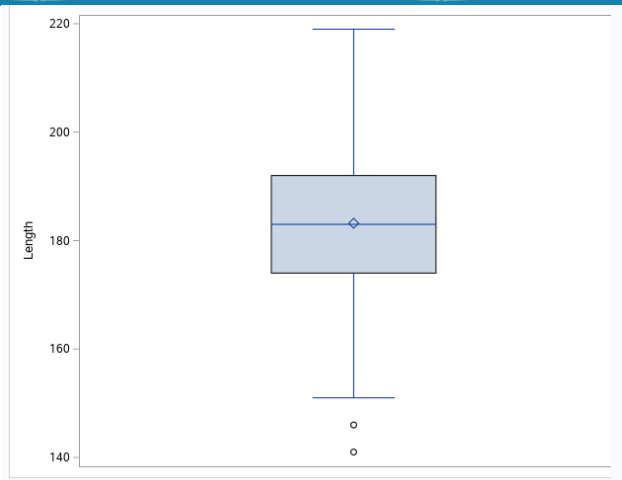
Quantiles (Definition 5)	
Level	Quantile
100% Max	8
99%	8
95%	7
90%	6
75% Q3	6
50% Median	5
25% Q1	4
10%	4
5%	4
1%	2
0% Min	2

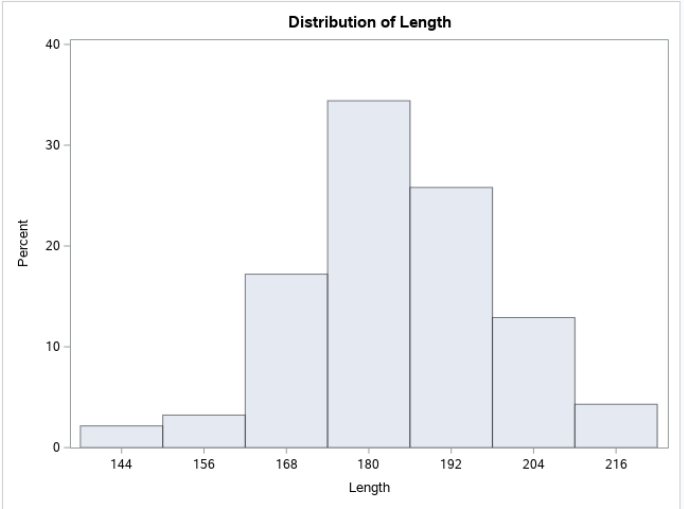
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
2	57	7	66
2	19	7	70
4	91	7	87
4	88	7	89
4	85	8	17

Length

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	183.204301	Sum Observations	17038
Std Deviation	14.6023815	Variance	213.229547
Skewness	-0.0900946	Kurtosis	0.44925041
Uncorrected SS	3141052	Corrected SS	19617.1183
Coeff Variation	7.97054515	Std Error Mean	1.51419643

Basic Statistical Measures			
Location		Variability	
Mean	183.2043	Std Deviation	14.60238
Median	183.0000	Variance	213.22955
Mode	184.0000	Range	78.00000
		Interquartile Range	18.00000

The mode displayed is the smallest of 2 modes with a count of 5.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	120.9911	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

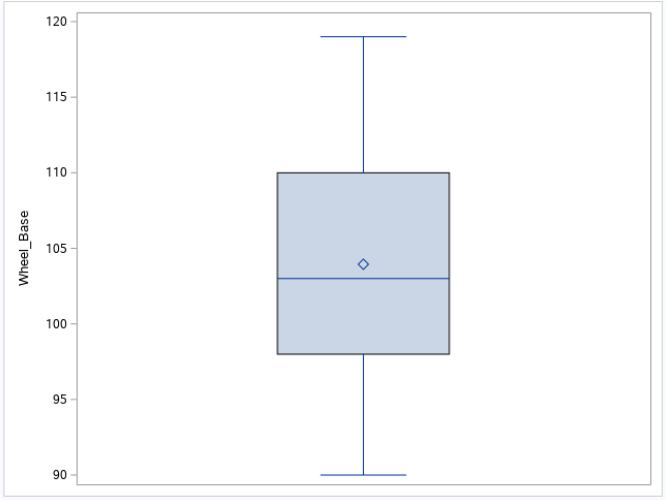
Quantiles (Definition 5)	
Level	Quantile
100% Max	219
99%	219
95%	206
90%	202
75% Q3	192
50% Median	183
25% Q1	174
10%	166
5%	161
1%	141
0% Min	141

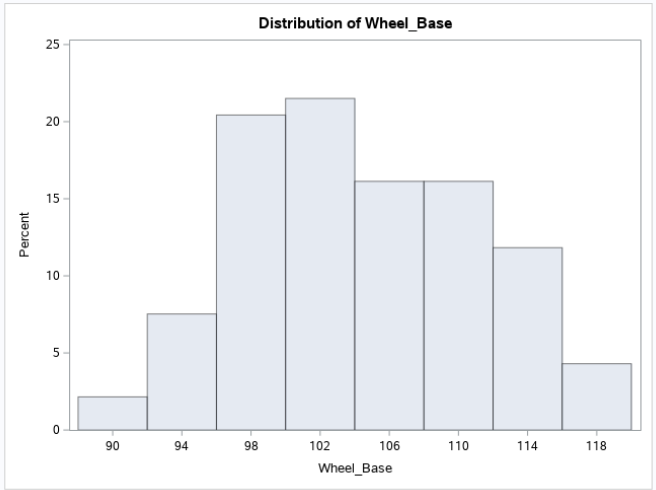
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
141	31	206	10
146	80	212	38
151	39	214	18
159	91	216	8
161	83	219	52

wheel_base

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	103.946237	Sum Observations	9667
Std Deviation	6.81967357	Variance	46.5079476
Skewness	0.11372684	Kurtosis	-0.7974557
Uncorrected SS	1009127	Corrected SS	4278.73118
Coeff Variation	6.56077007	Std Error Mean	0.70716721

Basic Statistical Measures			
Location		Variability	
Mean	103.9462	Std Deviation	6.81967
Median	103.0000	Variance	46.50795
Mode	103.0000	Range	29.00000
		Interquartile Range	12.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	146.9896	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

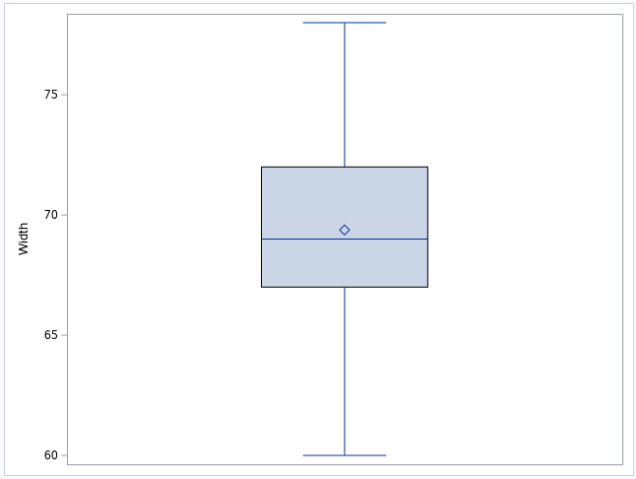
Quantiles (Definition 5)	
Level	Quantile
100% Max	119
99%	119
95%	115
90%	113
75% Q3	110
50% Median	103
25% Q1	98
10%	96
5%	93
1%	90
0% Min	90

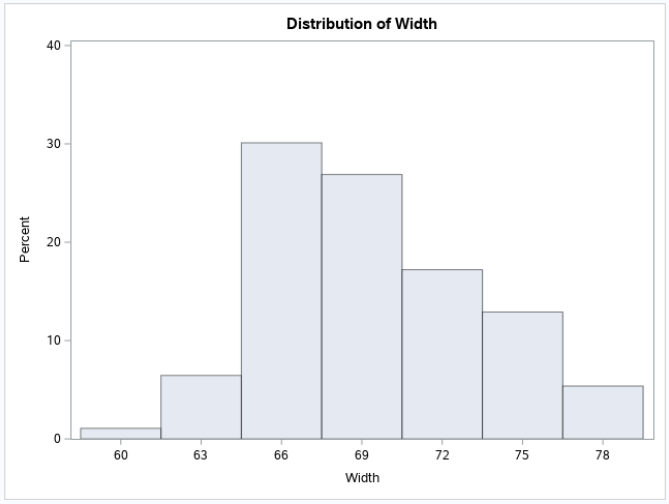
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
90	80	115	89
90	31	116	8
93	88	116	18
93	83	117	52
93	39	119	36

Width

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	69.3763441	Sum Observations	6452
Std Deviation	3.77898646	Variance	14.2807387
Skewness	0.26402738	Kurtosis	-0.2464277
Uncorrected SS	448930	Corrected SS	1313.82796
Coeff Variation	5.44708216	Std Error Mean	0.39186264

Basic Statistical Measures			
Location		Variability	
Mean	69.37634	Std Deviation	3.77899
Median	69.00000	Variance	14.28074
Mode	67.00000	Range	18.00000
		Interquartile Range	5.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	177.0425	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

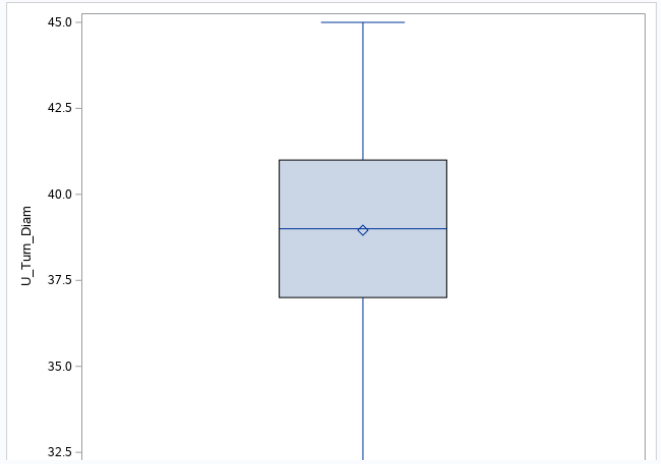
Quantiles (Definition 5)	
Level	Quantile
100% Max	78
99%	78
95%	77
90%	74
75% Q3	72
50% Median	69
25% Q1	67
10%	65
5%	63
1%	60
0% Min	60

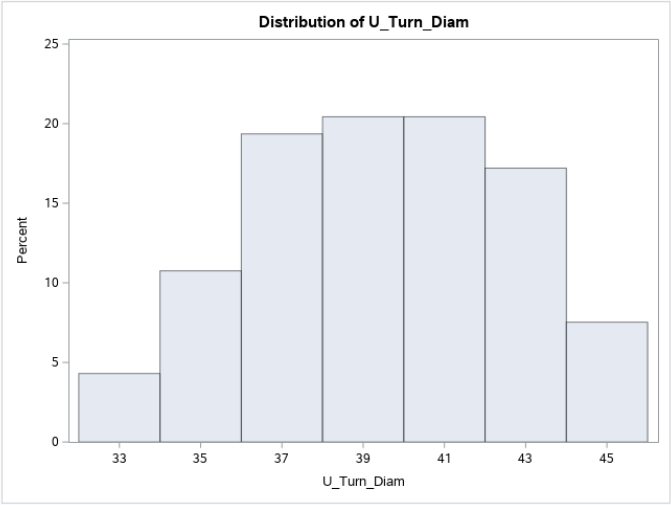
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
60	80	77	18
63	88	77	52
63	83	78	8
63	44	78	17
63	39	78	38

U_Turn_Diam

Same interpretation for this variable applies.

No outlier





Moments			
N	93	Sum Weights	93
Mean	38.9569892	Sum Observations	3623
Std Deviation	3.22326454	Variance	10.3894343
Skewness	-0.1335686	Kurtosis	-0.732204
Uncorrected SS	142097	Corrected SS	955.827957
Coeff Variation	8.27390567	Std Error Mean	0.33423696

Basic Statistical Measures			
Location		Variability	
Mean	38.95699	Std Deviation	3.22326
Median	39.00000	Variance	10.38943
Mode	39.00000	Range	13.00000
		Interquartile Range	4.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	116.555	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

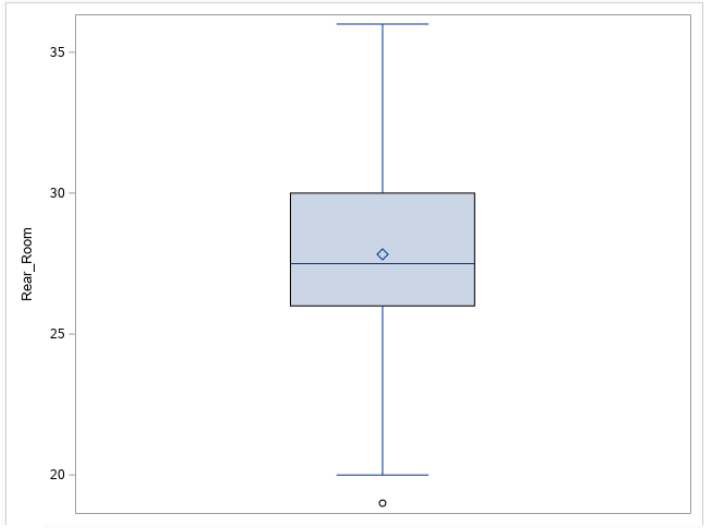
Quantiles (Definition 5)	
Level	Quantile
100% Max	45
99%	45
95%	44
90%	43
75% Q3	41
50% Median	39
25% Q1	37
10%	34
5%	34
1%	32
0% Min	32

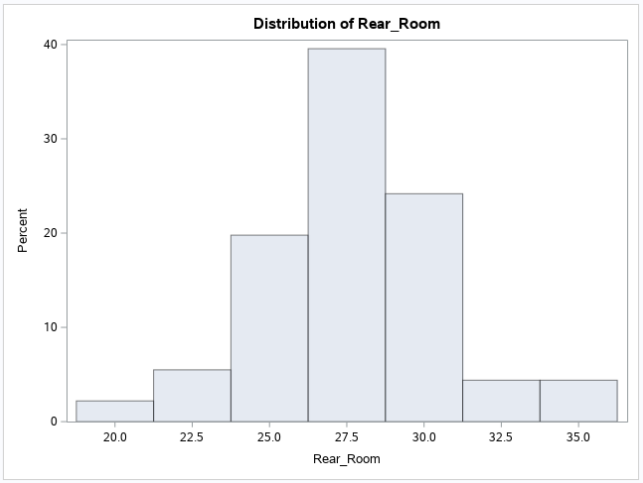
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
32	80	44	22
32	23	44	70
33	64	45	8
33	31	45	36
34	88	45	52

rear_room

Same interpretation for this variable applies.

No outlier





Moments			
N	91	Sum Weights	91
Mean	27.8296703	Sum Observations	2532.5
Std Deviation	2.98907249	Variance	8.93455433
Skewness	0.07825951	Kurtosis	0.89468223
Uncorrected SS	71282.75	Corrected SS	804.10989
Coeff Variation	10.7405961	Std Error Mean	0.31333994

Basic Statistical Measures			
Location		Variability	
Mean	27.82967	Std Deviation	2.98907
Median	27.50000	Variance	8.93455
Mode	26.50000	Range	17.00000
		Interquartile Range	4.00000

lote: The mode displayed is the smallest of 2 modes with a count of 9.

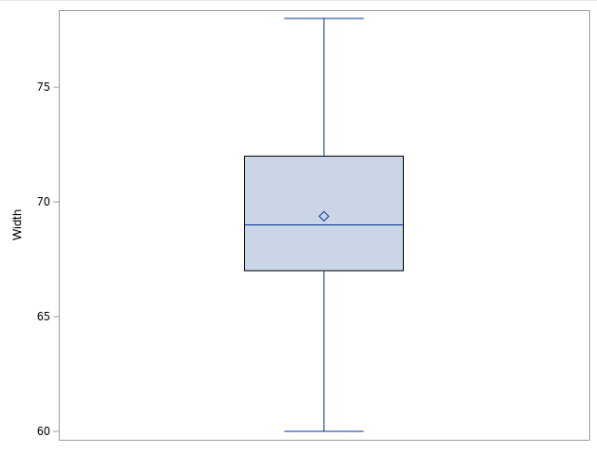
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	88.81623	Pr > |t|	<.0001
Sign	M	45.5	Pr >= |M|	<.0001
Signed Rank	S	2093	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	36.0
99%	36.0
95%	33.5
90%	31.0
75% Q3	30.0
50% Median	27.5
25% Q1	26.0
10%	24.5
5%	23.5
1%	19.0
0% Min	19.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
19.0	60	33.5	17
20.0	28	34.0	89
23.0	85	35.0	10
23.0	35	35.0	87
23.5	80	36.0	22

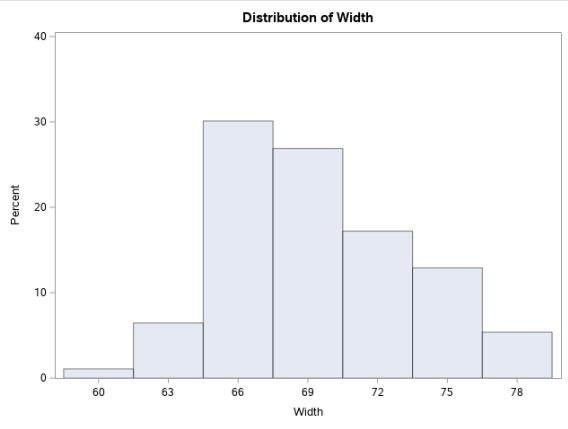
Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	2	2.15	100.00

Width



Same interpretation for this variable applies.

No outlier



Moments			
N	93	Sum Weights	93
Mean	69.3763441	Sum Observations	6452
Std Deviation	3.77898646	Variance	14.2807387
Skewness	0.26402738	Kurtosis	-0.2464277
Uncorrected SS	448930	Corrected SS	1313.82796
Coeff Variation	5.44708216	Std Error Mean	0.39186264

Basic Statistical Measures			
Location		Variability	
Mean	69.37634	Std Deviation	3.77899
Median	69.00000	Variance	14.28074
Mode	67.00000	Range	18.00000
		Interquartile Range	5.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	177.0425	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

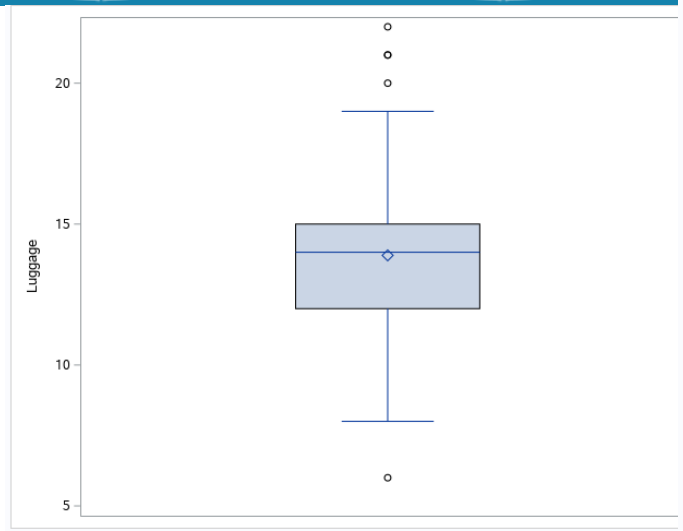
Quantiles (Definition 5)	
Level	Quantile
100% Max	78
99%	78
95%	77
90%	74
75% Q3	72
50% Median	69
25% Q1	67
10%	65
5%	63
1%	60
0% Min	60

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
60	80	77	18
63	88	77	52
63	83	78	8
63	44	78	17
63	39	78	38

Luggage

Same interpretation for this variable applies.

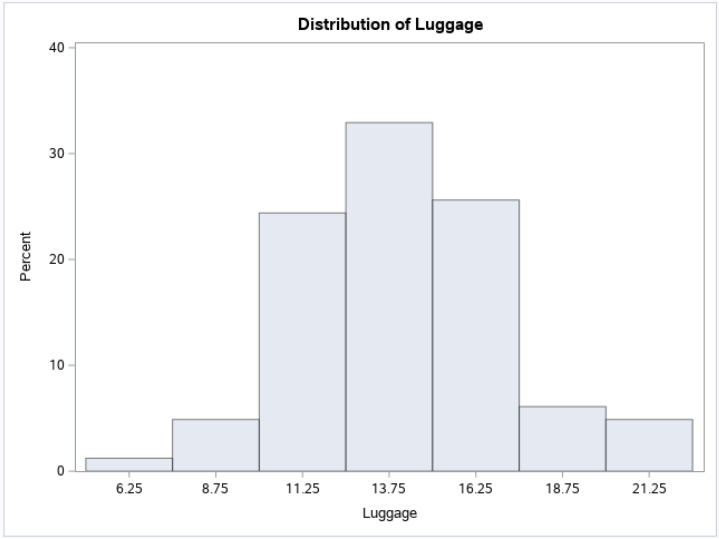
No outlier



Moments			
N	82	Sum Weights	82
Mean	13.8902439	Sum Observations	1139
Std Deviation	2.99796679	Variance	8.98780488
Skewness	0.22956577	Kurtosis	0.54968781
Uncorrected SS	16549	Corrected SS	728.012195
Coeff Variation	21.5832552	Std Error Mean	0.33107005

Quantiles (Definition 5)	
Level	Quantile
100% Max	22
99%	22
95%	19
90%	18
75% Q3	15
50% Median	14
25% Q1	12
10%	10
5%	9
1%	6
0% Min	6

Basic Statistical Measures			
Location		Variability	
Mean	13.89024	Std Deviation	2.99797
Median	14.00000	Variance	8.98780
Mode	14.00000	Range	16.00000
		Interquartile Range	3.00000

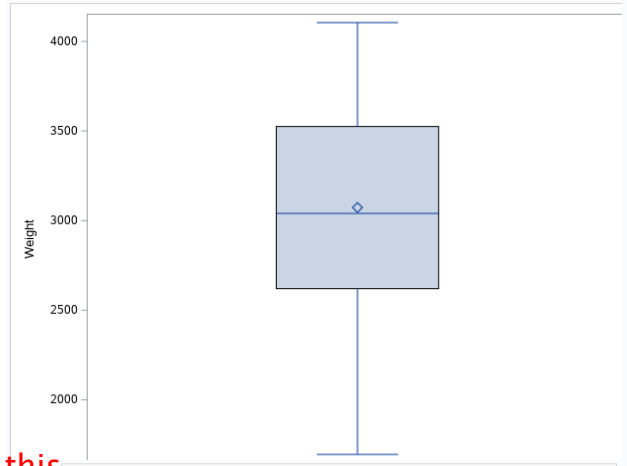


Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	41.9556	Pr > |t|	<.0001
Sign	M	41	Pr >= |M|	<.0001
Signed Rank	S	1701.5	Pr >= |S|	<.0001

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
6	60	19	51
8	72	20	18
8	41	21	8
9	50	21	38
9	46	22	52

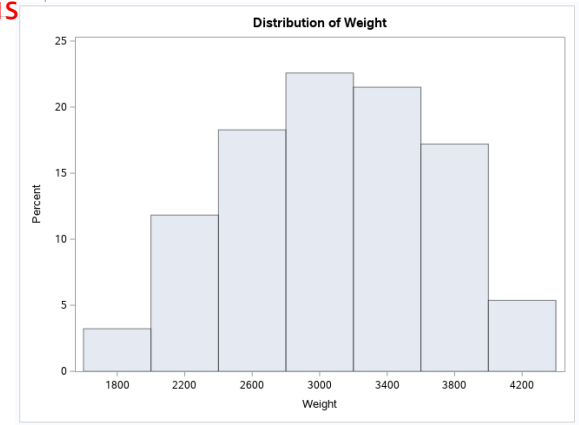
Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	11	11.83	100.00

Weight



Same interpretation for this variable applies.

No outlier



Moments			
N	93	Sum Weights	93
Mean	3072.90323	Sum Observations	285780
Std Deviation	589.89651	Variance	347977.893
Skewness	-0.143669	Kurtosis	-0.8551157
Uncorrected SS	910188250	Corrected SS	32013966.1
Coeff Variation	19.1967162	Std Error Mean	61.1694186

Basic Statistical Measures			
Location		Variability	
Mean	3072.903	Std Deviation	589.89651
Median	3040.000	Variance	347978
Mode	2295.000	Range	2410
		Interquartile Range	905.00000

Note: The mode displayed is the smallest of 12 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	50.23594	Pr > |t|	<.0001
Sign	M	46.5	Pr >= |M|	<.0001
Signed Rank	S	2185.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	4105
99%	4105
95%	4000
90%	3805
75% Q3	3525
50% Median	3040
25% Q1	2620
10%	2295
5%	2055
1%	1695
0% Min	1695

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1695	39	4000	48
1845	31	4025	17
1965	83	4055	52
2045	80	4100	66
2055	84	4105	8

Best Regards

Maryam Najimigoshtasb.