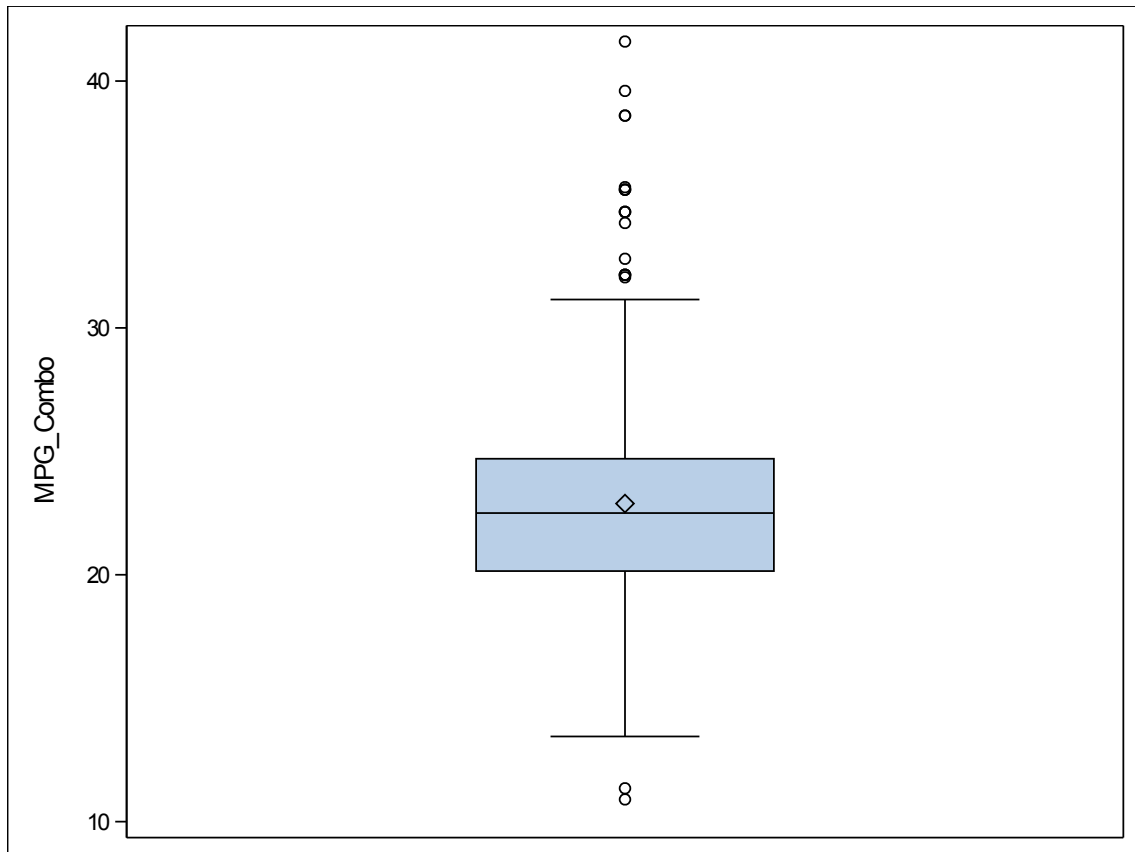


Stat 448, homework 1
Shuhui Guo

Exercise 1

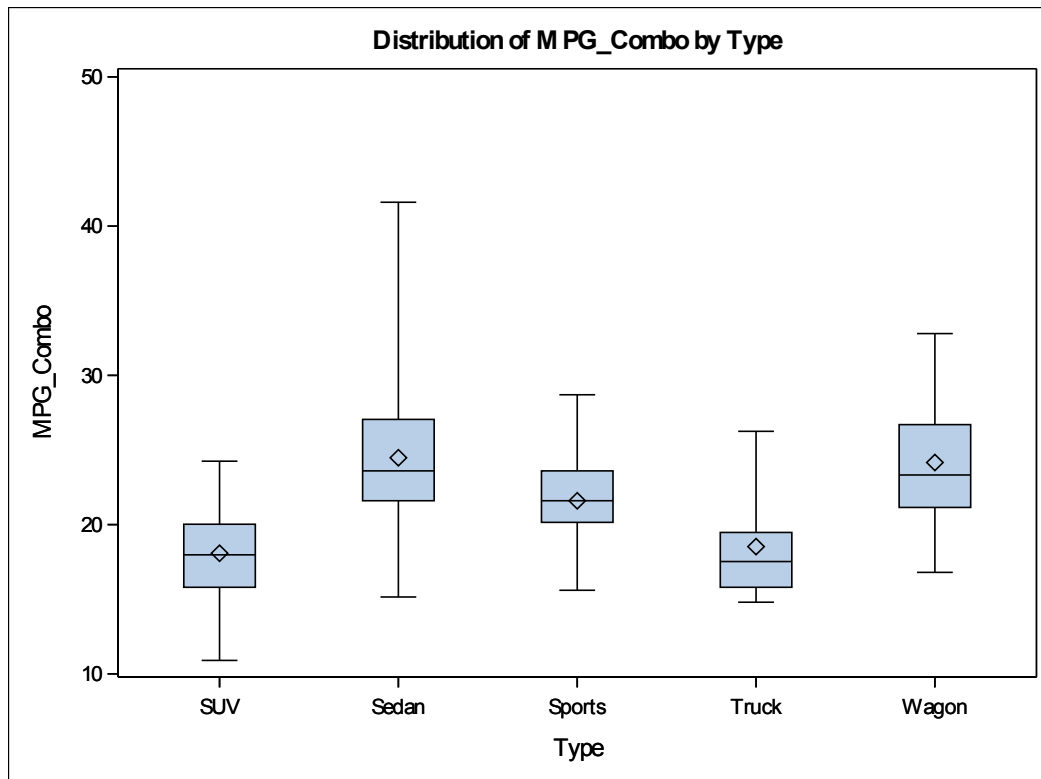
a) I created the variable **MPG_Combo** and the box plot is plotted as below:



There are 425 vehicles recorded in this dataset. **MPG_Combo** describes the combined fuel efficiency of each vehicle during its 55% of city driving and 45% of highway driving.

Based on the above plot, the median and the mean of combined fuel efficiency are close, which are around 23. The minimum is around 13 and the maximum is around 32 (except outliers). So we can say the normal level of combined fuel efficiency is from 13 to 32. The skew is not obvious, so the number of vehicles with higher fuel efficiency is similar to the number of vehicles with lower fuel efficiency. Nevertheless, there are outliers. The outliers higher than normal are much more than those lower than normal. The highest fuel efficiency could reach 40 and the lowest efficiency is about 10.

b) The box plots for **MPG_Combo** by **Type** are shown as below:



Based on this plot, we could conclude:

i) There are differences in the mean and median of combined fuel efficiency between the types. The mean of Sedan and Wagon are similar and they are comparatively higher. While the mean of SUV and Truck are similar and they are comparatively lower. The median of SUV is the lowest among the five types.

ii) There are differences in the variation of combined fuel efficiency between the types. The variation of Sedan is the biggest, which ranges approximately from 15 to 40. The variation of Truck is the smallest, which ranges approximately from 15 to 25. Also, for the types with similar mean, their variations are quite different.

iii) There are differences in the distribution of combined fuel efficiency in the types. For Truck, the long upper whisker shows the combined fuel efficiencies are varied amongst the highest quartile part whereas the short lower whisker shows they are similar amongst the lowest quartile part. For Sports, the length of upper whisker and lower whisker are similar, which indicates that combined fuel efficiencies distribute similarly amongst the two parts.

c) The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	425	Sum Weights	425
Mean	22.8847059	Sum Observations	9726
Std Deviation	4.6358734	Variance	21.4913221
Skewness	0.71081	Kurtosis	1.25934384
Uncorrected SS	231688.97	Corrected SS	9112.32059
Coeff Variation	20.2575179	Std Error Mean	0.22487289

Basic Statistical Measures			
Location		Variability	
Mean	22.88471	Std Deviation	4.63587
Median	22.50000	Variance	21.49132
Mode	21.15000	Range	30.70000
		Interquartile Range	4.55000

Tests for Location: $\mu_0=0$				
Test	Statistic		p Value	
Student's t	t	101.7673	Pr > t 	<.0001
Sign	M	212.5	Pr >= M 	<.0001
Signed Rank	S	45262.5	Pr >= S 	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	41.60
99%	35.70
95%	30.95
90%	29.15
75% Q3	24.70
50% Median	22.50
25% Q1	20.15
10%	17.25
5%	15.70

Quantiles (Definition 5)	
Level	Quantile
1%	13.80
0% Min	10.90

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
10.90	23	35.7	291
11.35	13	38.6	290
13.45	40	38.6	292
13.80	32	39.6	152
13.80	31	41.6	305

Based on the above tables, we could see:

i) In the table of tests for location, p-value for each test is less than 0.05, so we could conclude that it is unlikely that the mean or median of **MPG_Combo** is 0. According to the table of Basic Statistical Measures, the mean is 22.88 and median is 22.50, which validates the conclusion made before.

ii) The total number of observations is 425. The standard deviation is 4.64 and the range is 30.70, which is from 10.90 to 41.60 according to the tables of Quantiles and Extreme Observations.

iii) The skewness is 0.71, which is positive. So the distribution of **MPG_Combo** is right-skewed. Under such circumstance, the mean is greater than the median and the median is greater than the mode, which correspond to the results in the table of Basic Statistical Measures.

The basic descriptive statistics for **Invoice** are as below:

Moments			
N	425	Sum Weights	425
Mean	30096.48	Sum Observations	12791004
Std Deviation	17677.3562	Variance	312488924
Skewness	2.82591763	Kurtosis	13.8776543

Moments			
Uncorrected SS	5.17459E11	Corrected SS	1.32495E11
Coeff Variation	58.735627	Std Error Mean	857.477729

Basic Statistical Measures			
Location		Variability	
Mean	30096.48	Std Deviation	17677
Median	25672.00	Variance	312488924
Mode	14207.00	Range	163685
		Interquartile Range	16804

Tests for Location: $\mu_0=0$				
Test	Statistic		p Value	
Student's t	t	35.09885	Pr > t 	<.0001
Sign	M	212.5	Pr >= M 	<.0001
Signed Rank	S	45262.5	Pr >= S 	<.0001

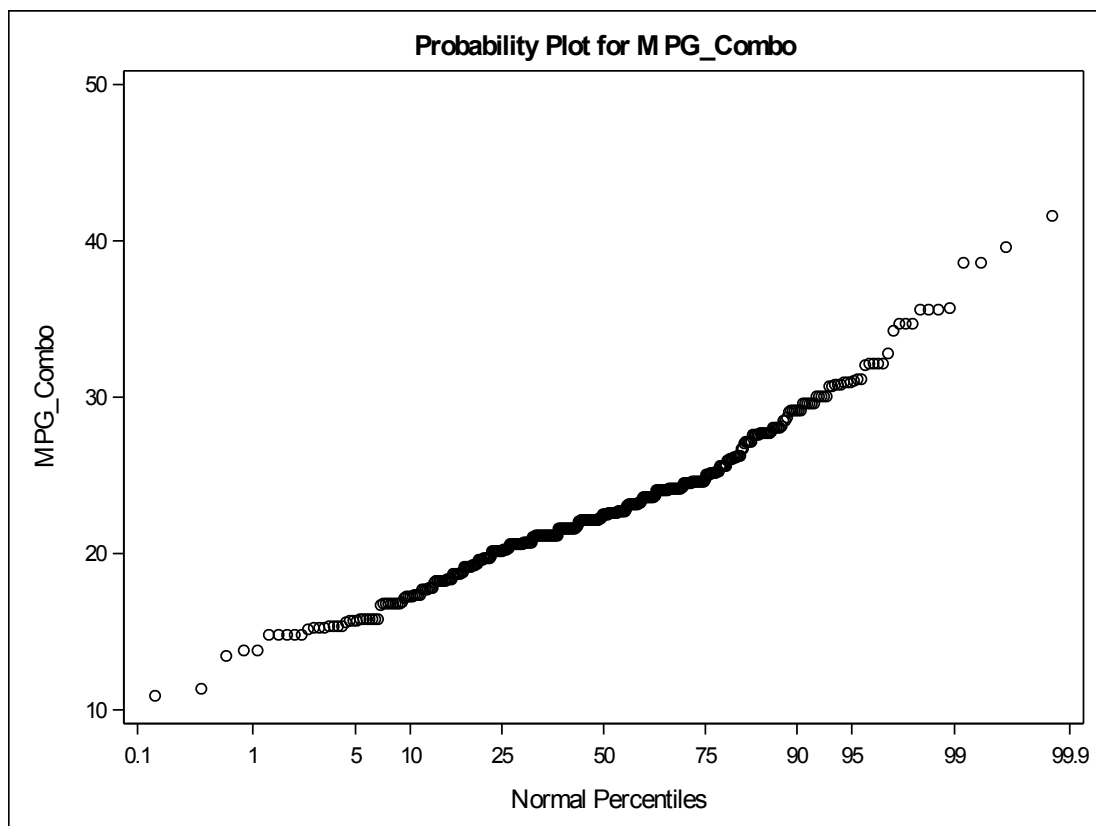
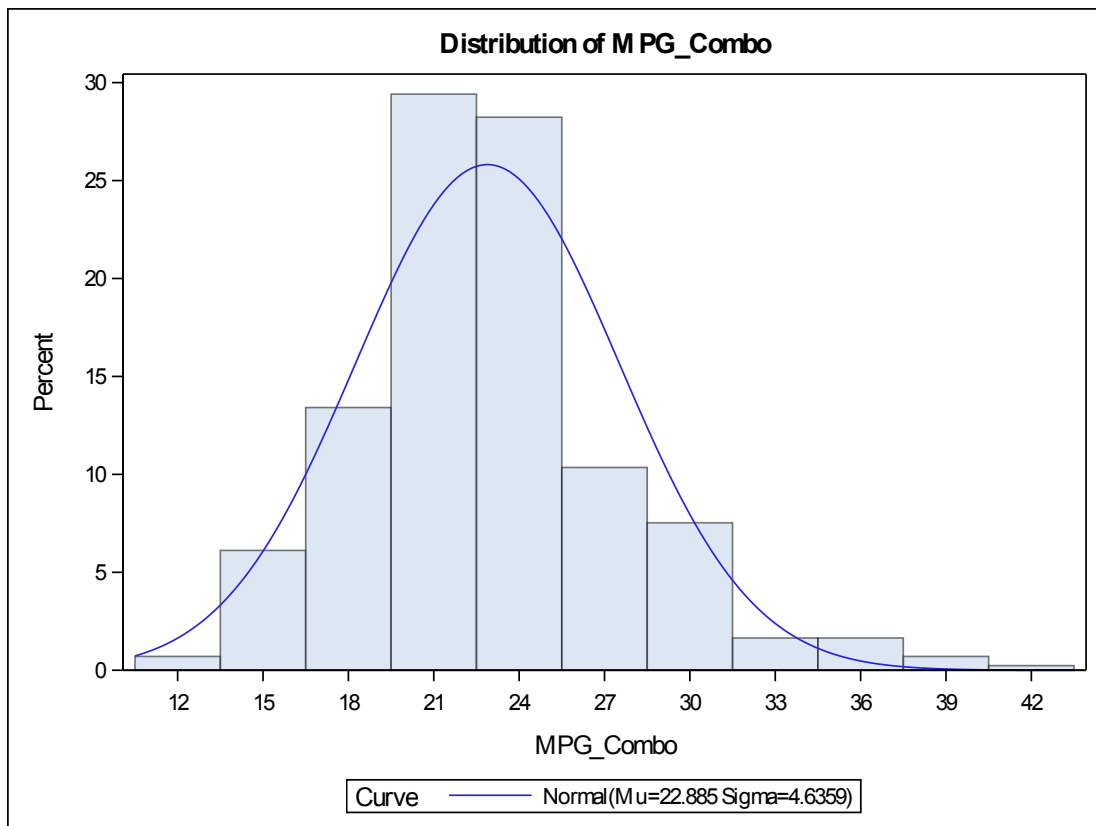
Quantiles (Definition 5)	
Level	Quantile
100% Max	173560
99%	88324
95%	66830
90%	48377
75% Q3	35777
50% Median	25672
25% Q1	18973
10%	14375
5%	12830
1%	10642
0% Min	9875

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
9875	187	88324	222
10107	162	113388	352
10144	290	117854	353
10319	269	119600	223
10642	292	173560	365

Based on the above tables, we could see:

- i) In the table of tests for location, p-value for each test is less than 0.05, so we could conclude that it is unlikely that the mean or median of **Invoice** is 0. According to the table of Basic Statistical Measures, the mean is 30096.48 and median is 25672.00, which validates the conclusion made before.
- ii) The total number of observations is 425. The standard deviation is 17677 and the range is 163685, which is from 9875 to 173560 according to the tables of Quantiles and Extreme Observations.
- iii) The skewness is 2.83, which is positive. So the distribution of **Invoice** is right-skewed. Under such circumstance, the mean is greater than the median and the median is greater than the mode, which correspond to the results in the table of Basic Statistical Measures.

The tests of normality for **MPG_Combo** are shown as below:



According to the histogram and probplot, the histogram of **MPG_Combo** seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.967372	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.103925	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.696027	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.776247	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **MPG_Combo**.

d) Repeat the analysis in part (c) by Type. The results are shown as below:

i) Type=SUV

The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	60	Sum Weights	60
Mean	18.08	Sum Observations	1084.8
Std Deviation	2.99846288	Variance	8.99077966
Skewness	0.03642022	Kurtosis	-0.2034269
Uncorrected SS	20143.64	Corrected SS	530.456
Coeff Variation	16.5844186	Std Error Mean	0.38709989

Basic Statistical Measures			
Location		Variability	
Mean	18.08000	Std Deviation	2.99846
Median	17.97500	Variance	8.99078
Mode	18.25000	Range	13.35000
		Interquartile Range	4.22500

For the variable **MPG_Combo** in the type of SUV, the total number of observations is 60. The mean, median and mode are 18.08, 17.98 and 18.25, respectively. The range is 13.35, which is comparatively small.

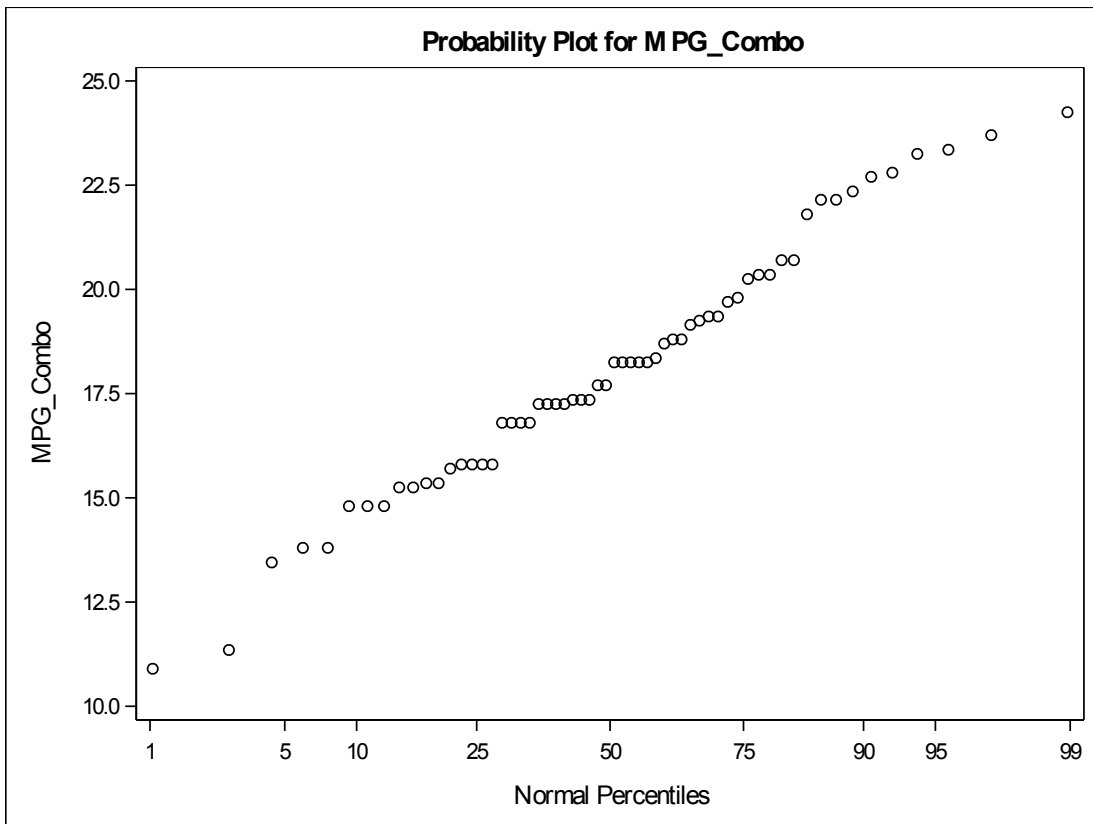
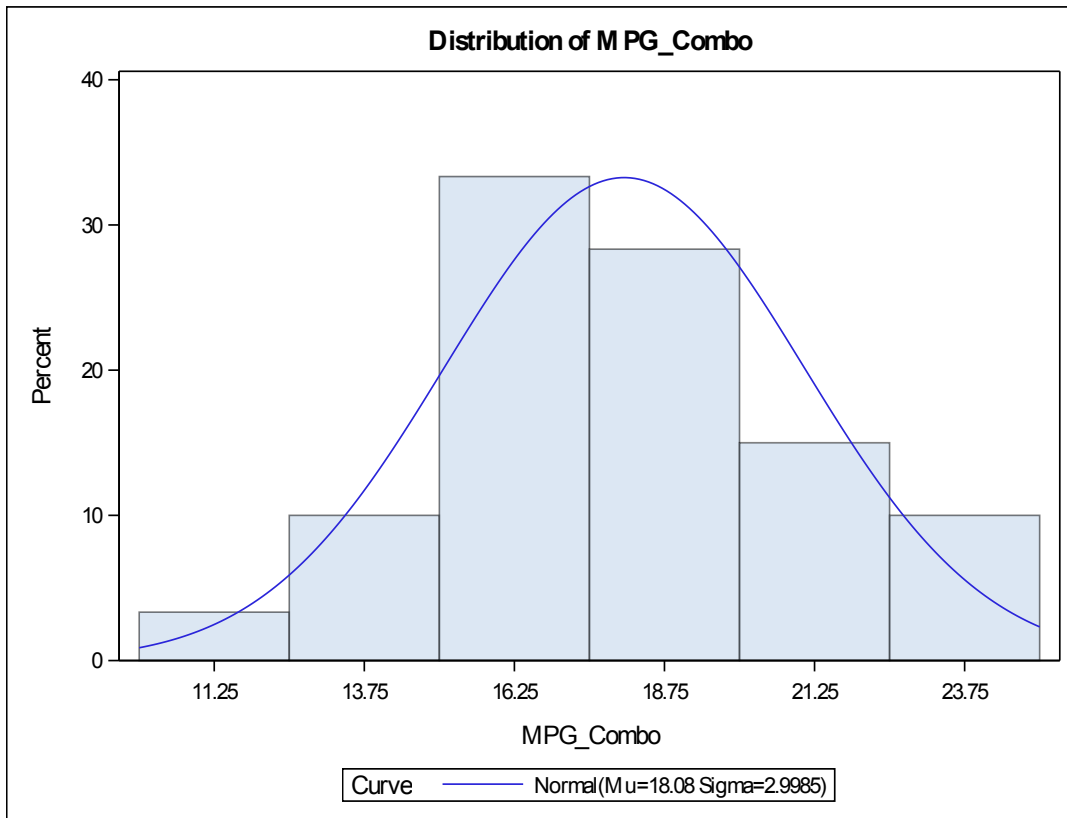
The basic descriptive statistics for **Invoice** are as below:

Moments			
N	60	Sum Weights	60
Mean	31625.35	Sum Observations	1897521
Std Deviation	12062.3877	Variance	145501196
Skewness	1.14910764	Kurtosis	1.44136756
Uncorrected SS	6.85943E10	Corrected SS	8584570588
Coeff Variation	38.1415152	Std Error Mean	1557.24755

Basic Statistical Measures			
Location		Variability	
Mean	31625.35	Std Deviation	12062
Median	29724.50	Variance	145501196
Mode	.	Range	54591
		Interquartile Range	16530

For the variable **Invoice** in the type of SUV, the total number of observations is 60. The mean and median are 31625.35 and 29724.50, respectively. The range is 54591, which is comparatively small.

The tests of normality for **MPG_Combo** are shown as below:



According to the histogram and probplot, the histogram of **MPG_Combo** seems to fit the curve of normal distribution. Also, the probability plot appears to be straight. Furthermore, the test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.982475	Pr < W	0.5417
Kolmogorov-Smirnov	D	0.064125	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.048658	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.340069	Pr > A-Sq	>0.2500

In the table of Test for Normality, the p-value of tests is more than 0.05. It does not reject the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would be reasonable for **MPG_Combo**.

ii) Type=Sedan

The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	262	Sum Weights	262
Mean	24.4795802	Sum Observations	6413.65
Std Deviation	4.28319726	Variance	18.3457788
Skewness	1.13146336	Kurtosis	1.76814625
Uncorrected SS	161791.707	Corrected SS	4788.24825
Coeff Variation	17.4970209	Std Error Mean	0.26461681

Basic Statistical Measures			
Location		Variability	
Mean	24.47958	Std Deviation	4.28320
Median	23.60000	Variance	18.34578
Mode	21.15000	Range	26.45000
		Interquartile Range	5.45000

For the variable **MPG_Combo** in the type of Sedan, the total number of observations is 262. The mean, median and mode are 24.48, 23.60 and 21.15, respectively. The range is 26.45, which is comparatively large.

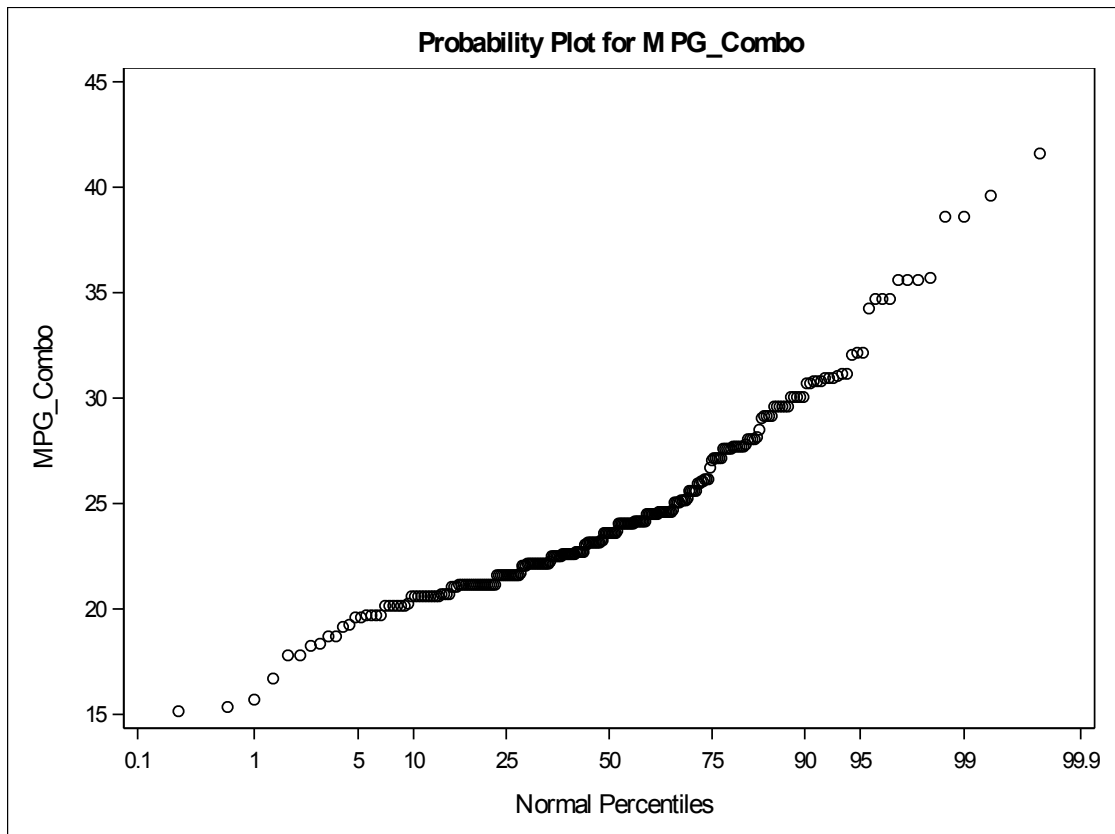
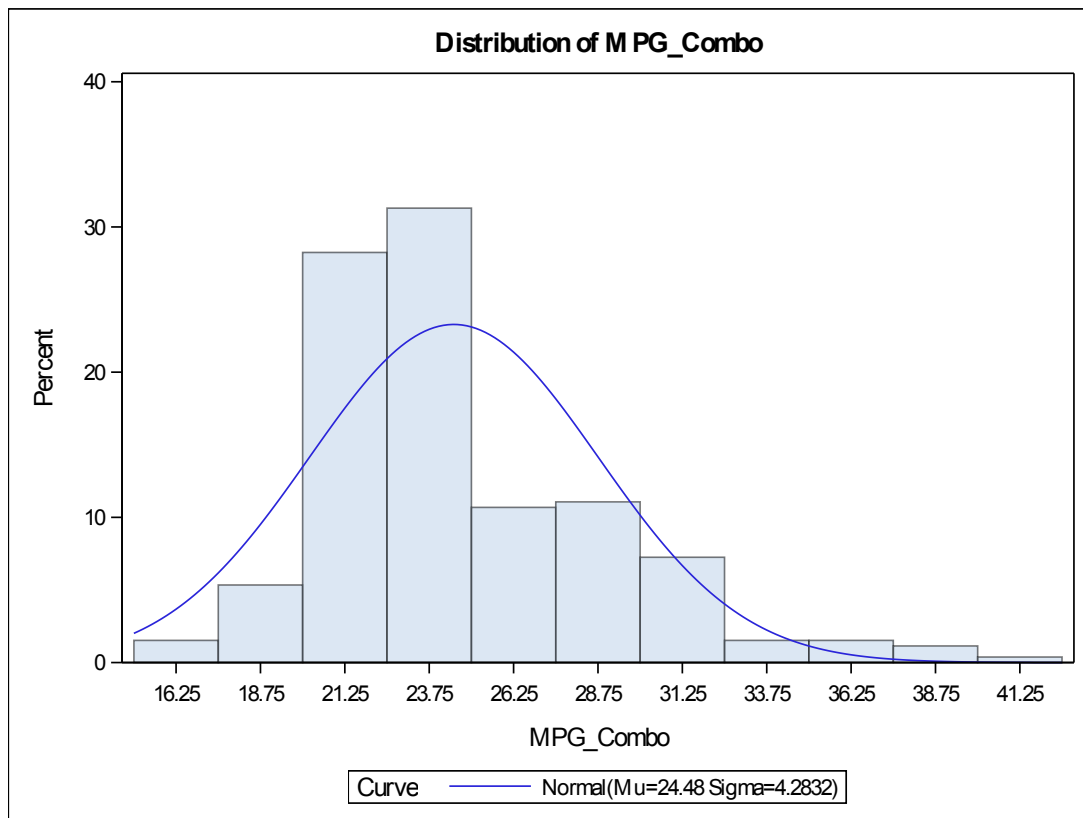
The basic descriptive statistics for **Invoice** are as below:

Moments			
N	262	Sum Weights	262
Mean	27389.7977	Sum Observations	7176127
Std Deviation	14305.1516	Variance	204637363
Skewness	2.15778851	Kurtosis	8.08015271
Uncorrected SS	2.49963E11	Corrected SS	5.34104E10
Coeff Variation	52.2280295	Std Error Mean	883.775218

Basic Statistical Measures			
Location		Variability	
Mean	27389.80	Std Deviation	14305
Median	24063.00	Variance	204637363
Mode	14207.00	Range	109725
		Interquartile Range	15194

For the variable **Invoice** in the type of Sedan, the total number of observations is 262. The mean, median and mode are 27389.80, 24063.00 and 14207.00, respectively. The range is 109725, which is comparatively large.

The tests of normality for **MPG_Combo** are shown as below:



According to the histogram and probplot, the histogram of **MPG_Combo** seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.922362	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.149091	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.175865	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.436773	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **MPG_Combo**.

iii) Type=Sports

The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	49	Sum Weights	49
Mean	21.594898	Sum Observations	1058.15
Std Deviation	2.71631347	Variance	7.37835884
Skewness	0.21533497	Kurtosis	0.51232696
Uncorrected SS	23204.8025	Corrected SS	354.161224
Coeff Variation	12.5784964	Std Error Mean	0.38804478

Basic Statistical Measures			
Location		Variability	
Mean	21.59490	Std Deviation	2.71631
Median	21.60000	Variance	7.37836
Mode	21.60000	Range	13.10000
		Interquartile Range	3.45000

For the variable **MPG_Combo** in the type of Sports, the total number of observations is 49. The mean, median and mode are 21.59, 21.60 and 21.60, respectively. The range is 13.10, which is comparatively small.

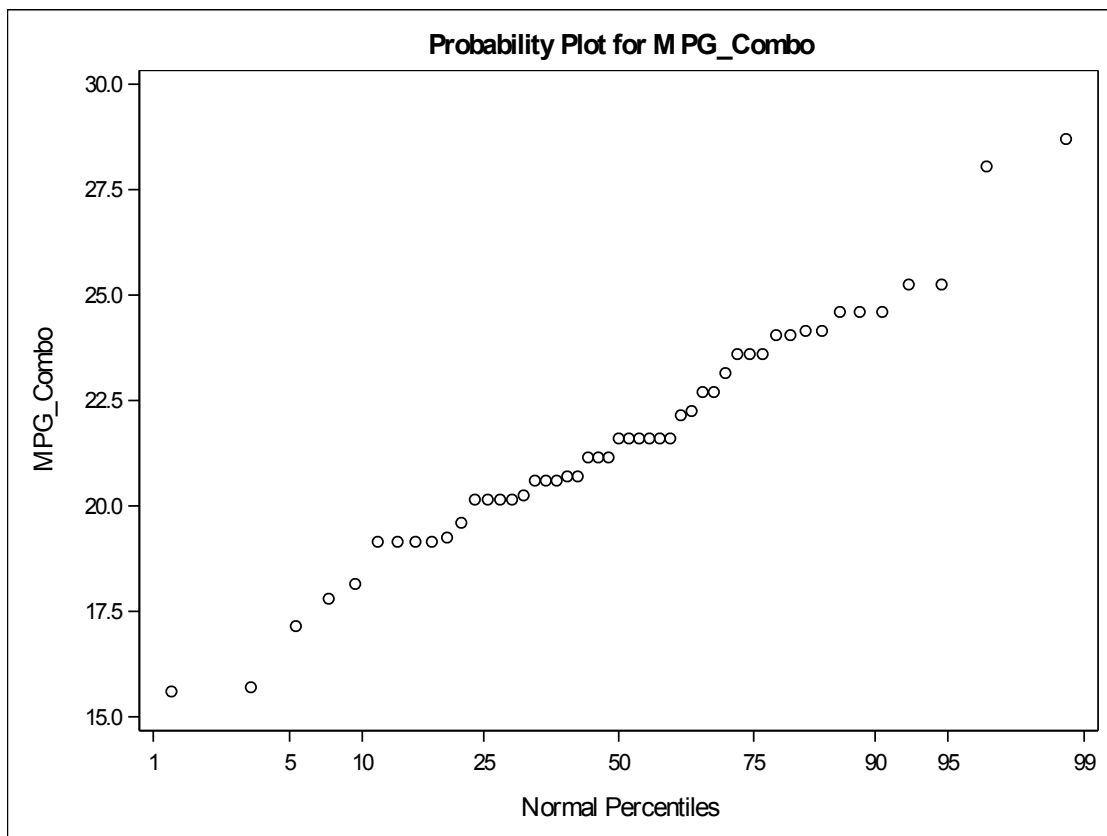
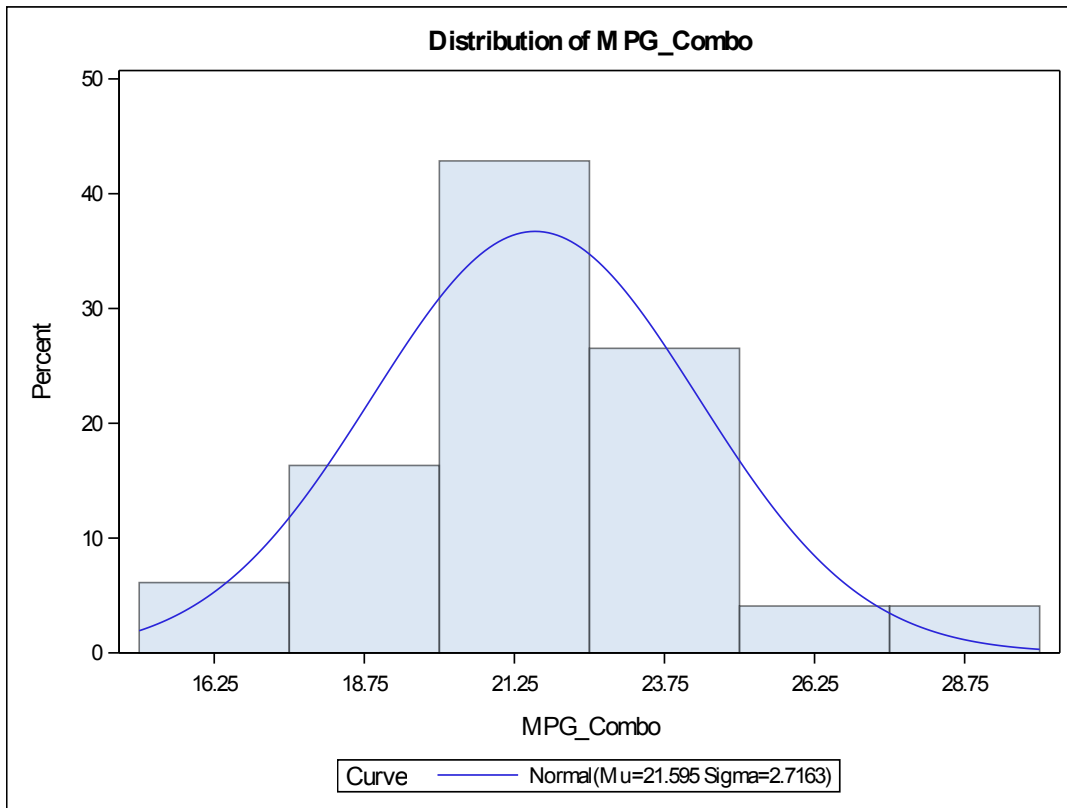
The basic descriptive statistics for **Invoice** are as below:

Moments			
N	49	Sum Weights	49
Mean	48473.1633	Sum Observations	2375185
Std Deviation	30609.5087	Variance	936942020
Skewness	1.8834277	Kurtosis	4.84925442
Uncorrected SS	1.60106E11	Corrected SS	4.49732E10
Coeff Variation	63.1473306	Std Error Mean	4372.78695

Basic Statistical Measures			
Location		Variability	
Mean	48473.16	Std Deviation	30610
Median	37548.00	Variance	936942020
Mode	.	Range	156617
		Interquartile Range	41431

For the variable **Invoice** in the type of Sports, the total number of observations is 49. The mean and median are 48473.16 and 37548.00, respectively. The range is 156617, which is comparatively large.

The tests of normality for **MPG_Combo** are shown as below:



According to the histogram and probplot, the histogram of **MPG_Combo** seems to fit the curve of normal distribution. Also, the probability plot appears to be straight. Furthermore, the test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.978152	Pr < W	0.4905
Kolmogorov-Smirnov	D	0.111496	Pr > D	0.1297
Cramer-von Mises	W-Sq	0.060028	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.38303	Pr > A-Sq	>0.2500

In the table of Test for Normality, the p-value of tests is more than 0.05. It does not reject the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would be reasonable for **MPG_Combo**.

iv) Type=Truck

The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	24	Sum Weights	24
Mean	18.525	Sum Observations	444.6
Std Deviation	3.47847418	Variance	12.0997826
Skewness	1.1919733	Kurtosis	0.44375964
Uncorrected SS	8514.51	Corrected SS	278.295
Coeff Variation	18.7771885	Std Error Mean	0.71004057

Basic Statistical Measures			
Location		Variability	
Mean	18.52500	Std Deviation	3.47847
Median	17.52500	Variance	12.09978
Mode	15.80000	Range	11.45000
		Interquartile Range	3.67500

For the variable **MPG_Combo** in the type of Truck, the total number of observations is 24. The mean, median and mode are 18.53, 17.53 and 15.80, respectively. The range is 11.45, which is comparatively small.

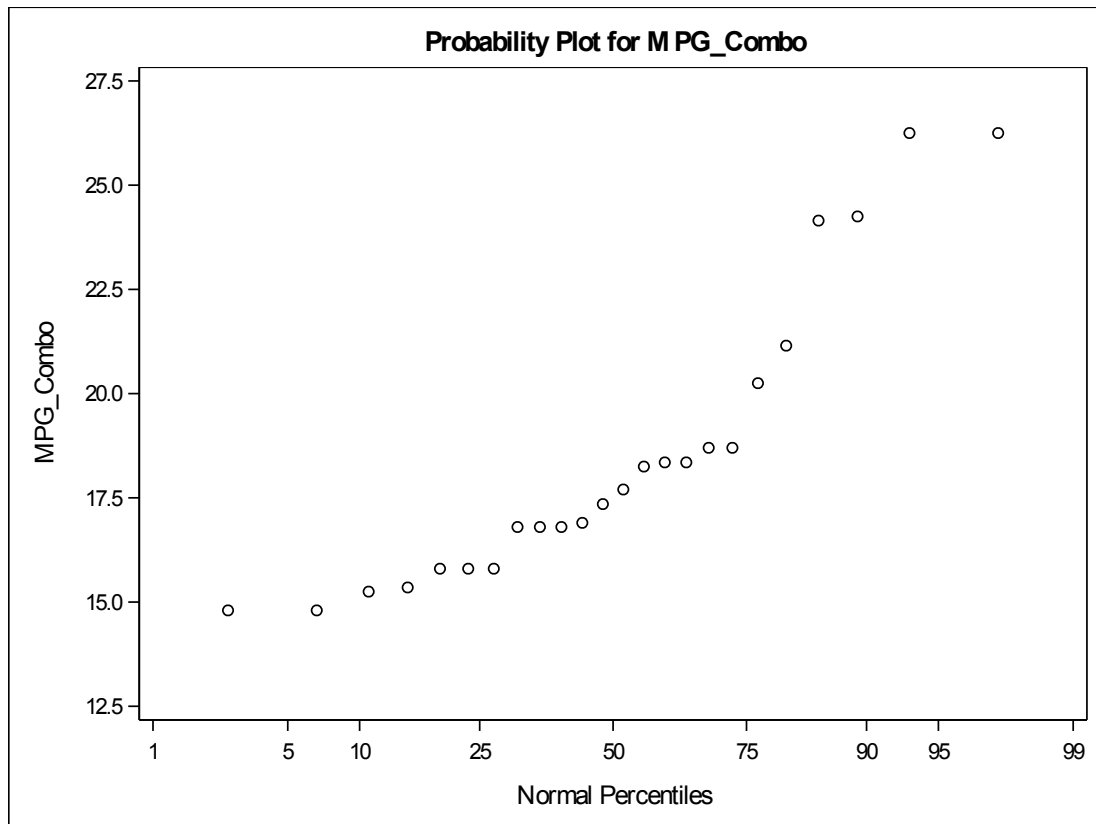
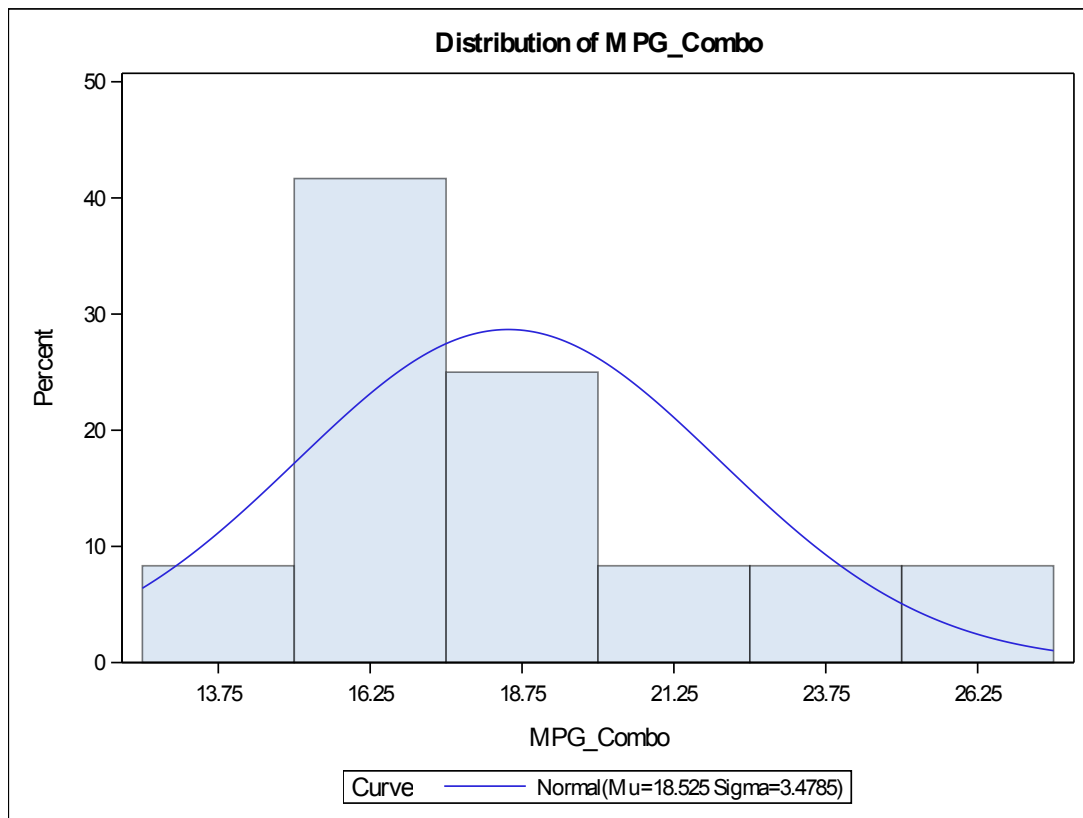
The basic descriptive statistics for **Invoice** are as below:

Moments			
N	24	Sum Weights	24
Mean	22616.75	Sum Observations	542802
Std Deviation	8852.13194	Variance	78360239.8
Skewness	1.43665175	Kurtosis	2.09054813
Uncorrected SS	1.40787E10	Corrected SS	1802285517
Coeff Variation	39.139717	Std Error Mean	1806.93387

Basic Statistical Measures			
Location		Variability	
Mean	22616.75	Std Deviation	8852
Median	19986.00	Variance	78360240
Mode	.	Range	36662
		Interquartile Range	8676

For the variable **Invoice** in the type of Truck, the total number of observations is 24. The mean and median are 22616.75 and 19986.00, respectively. The range is 36662, which is comparatively small.

The tests of normality for **MPG_Combo** are shown as below:



According to the histogram and probplot, the histogram of **MPG_Combo** seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.844193	Pr < W	0.0017
Kolmogorov-Smirnov	D	0.229938	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.228643	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.374134	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **MPG_Combo**.

v) Type=Wagon

The basic descriptive statistics for **MPG_Combo** are as below:

Moments			
N	30	Sum Weights	30
Mean	24.16	Sum Observations	724.8
Std Deviation	4.24727905	Variance	18.0393793
Skewness	0.44078339	Kurtosis	-0.4225572
Uncorrected SS	18034.31	Corrected SS	523.142
Coeff Variation	17.5797974	Std Error Mean	0.77544351

Basic Statistical Measures			
Location		Variability	
Mean	24.16000	Std Deviation	4.24728
Median	23.32500	Variance	18.03938
Mode	22.15000	Range	16.00000
		Interquartile Range	5.55000

For the variable **MPG_Combo** in the type of Wagon, the total number of observations is 30. The mean, median and mode are 24.16, 23.33 and 22.15, respectively. The range is 16.00, which is comparatively small.

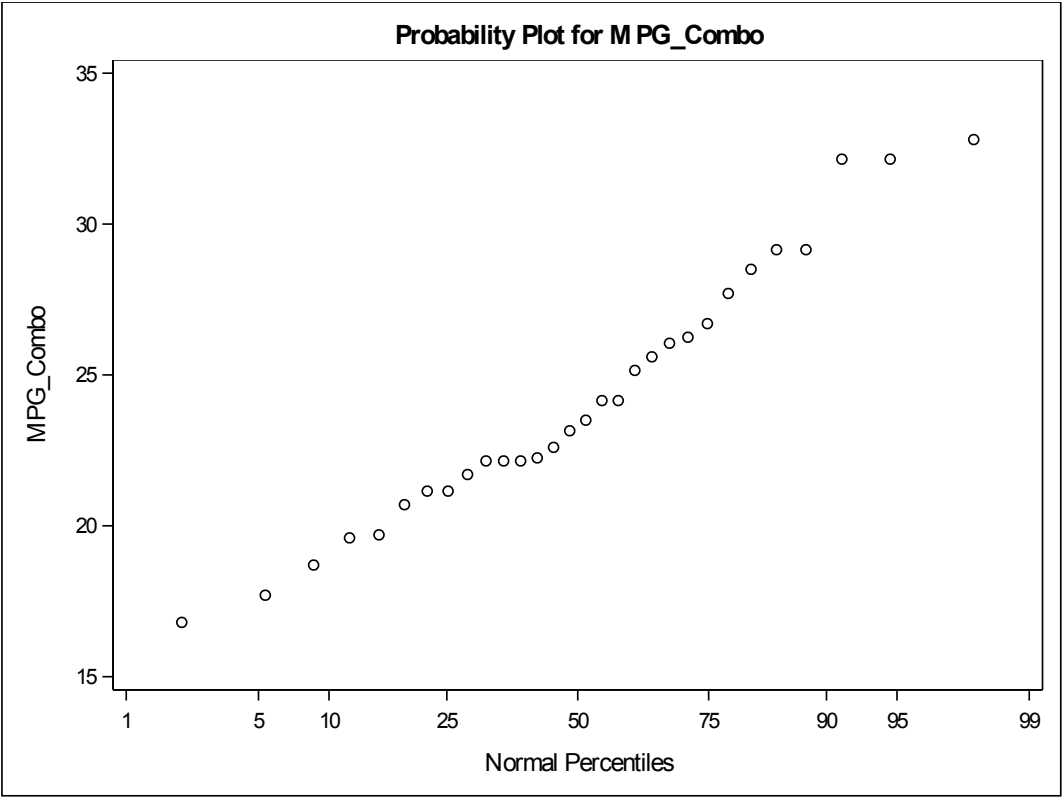
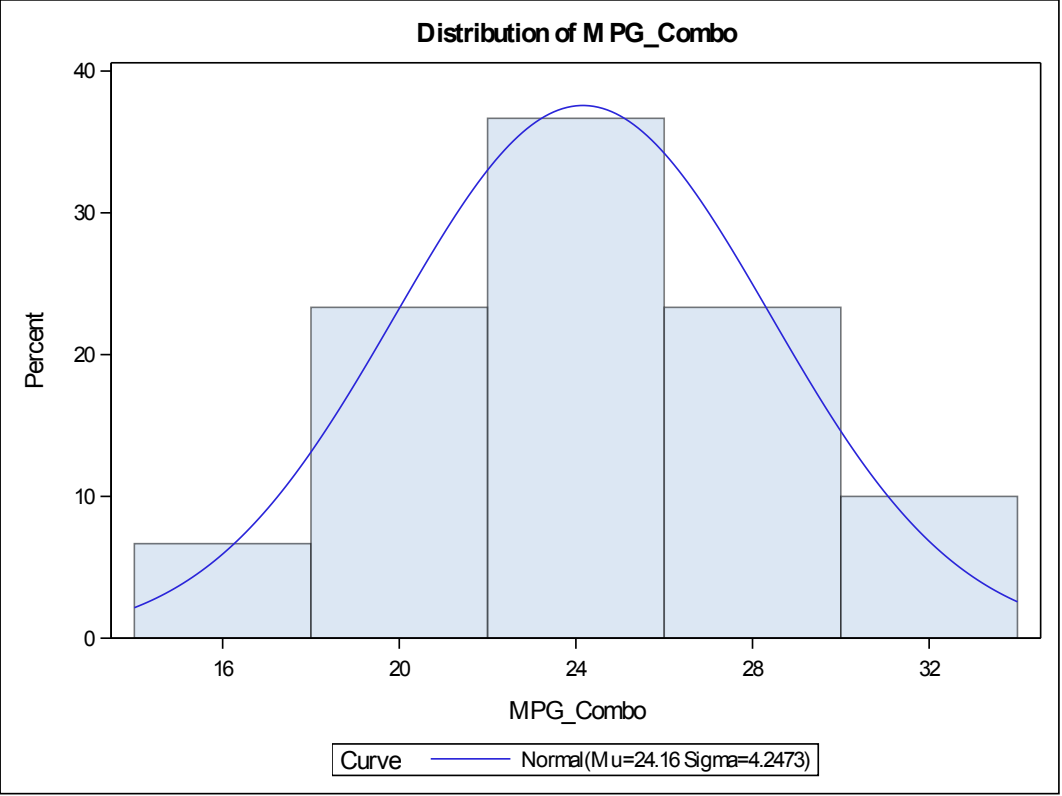
The basic descriptive statistics for **Invoice** are as below:

Moments			
N	30	Sum Weights	30
Mean	26645.6333	Sum Observations	799369
Std Deviation	10856.1149	Variance	117855231
Skewness	0.91116089	Kurtosis	0.57217698
Uncorrected SS	2.47175E10	Corrected SS	3417801701
Coeff Variation	40.7425666	Std Error Mean	1982.04634

Basic Statistical Measures			
Location		Variability	
Mean	26645.63	Std Deviation	10856
Median	23721.00	Variance	117855231
Mode	.	Range	45064
		Interquartile Range	15685

For the variable **Invoice** in the type of Wagon, the total number of observations is 30. The mean and median are 26645.63 and 23721.00, respectively. The range is 45064, which is comparatively small.

The tests of normality for **MPG_Combo** are shown as below:



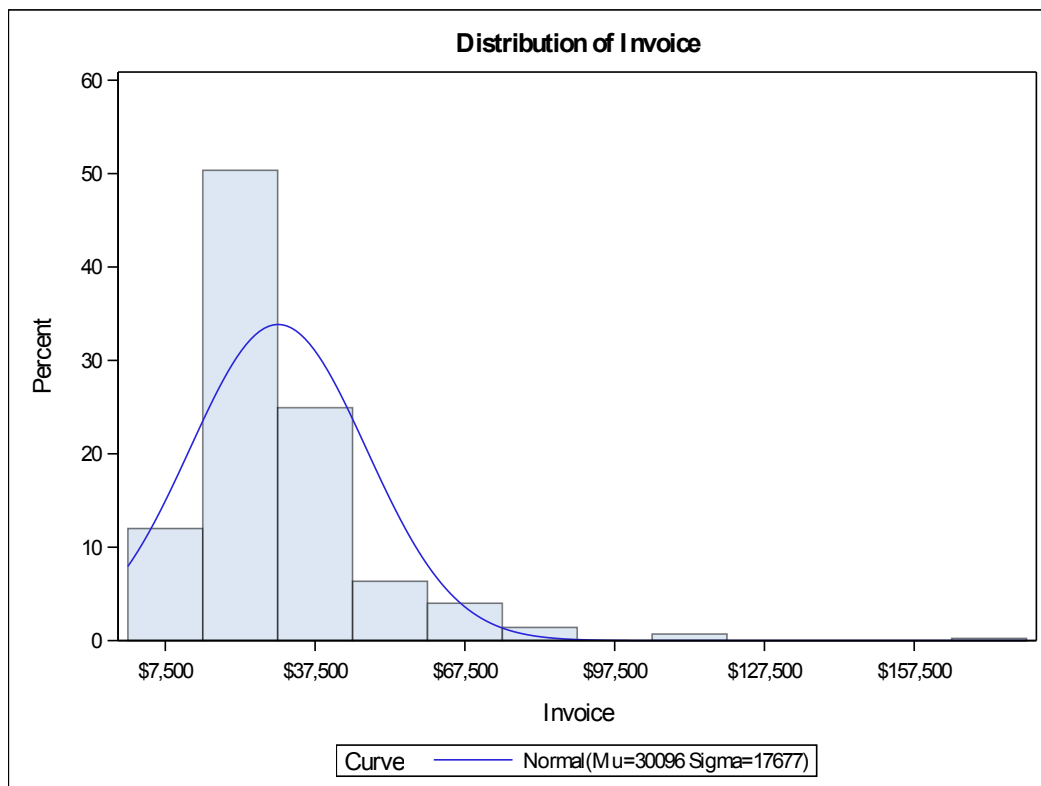
According to the histogram and probplot, the histogram of **MPG_Combo** seems to fit the curve of normal distribution. Also, the probability plot appears to be straight. Furthermore, the test statistics for normality should be checked.

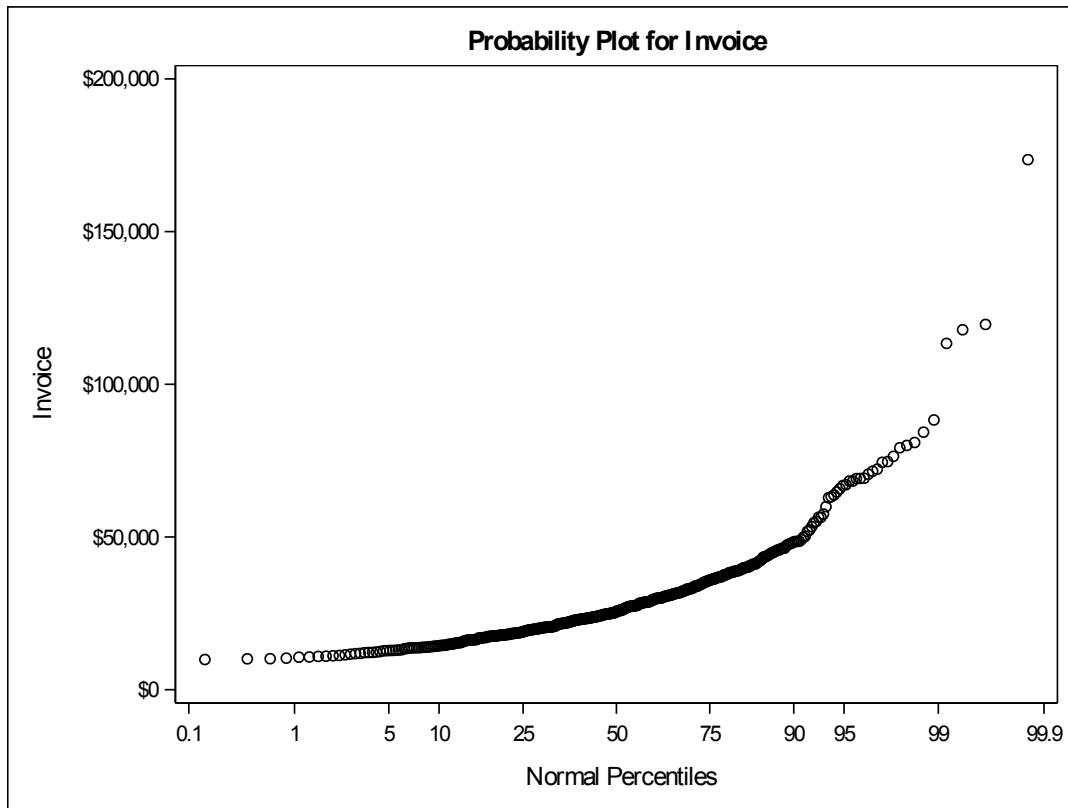
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.964051	Pr < W	0.3914
Kolmogorov-Smirnov	D	0.109967	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.057368	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.361301	Pr > A-Sq	>0.2500

In the table of Test for Normality, the p-value of tests is more than 0.05. It does not reject the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would be reasonable for **MPG_Combo**.

Exercise 2

a) First the normality of **Invoice** should be checked:





According to the histogram and probplot, the histogram of **Invoice** seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.77353	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.140604	Pr > D	<0.0100
Cramer-von Mises	W-Sq	3.393462	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	20.06351	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **Invoice**.

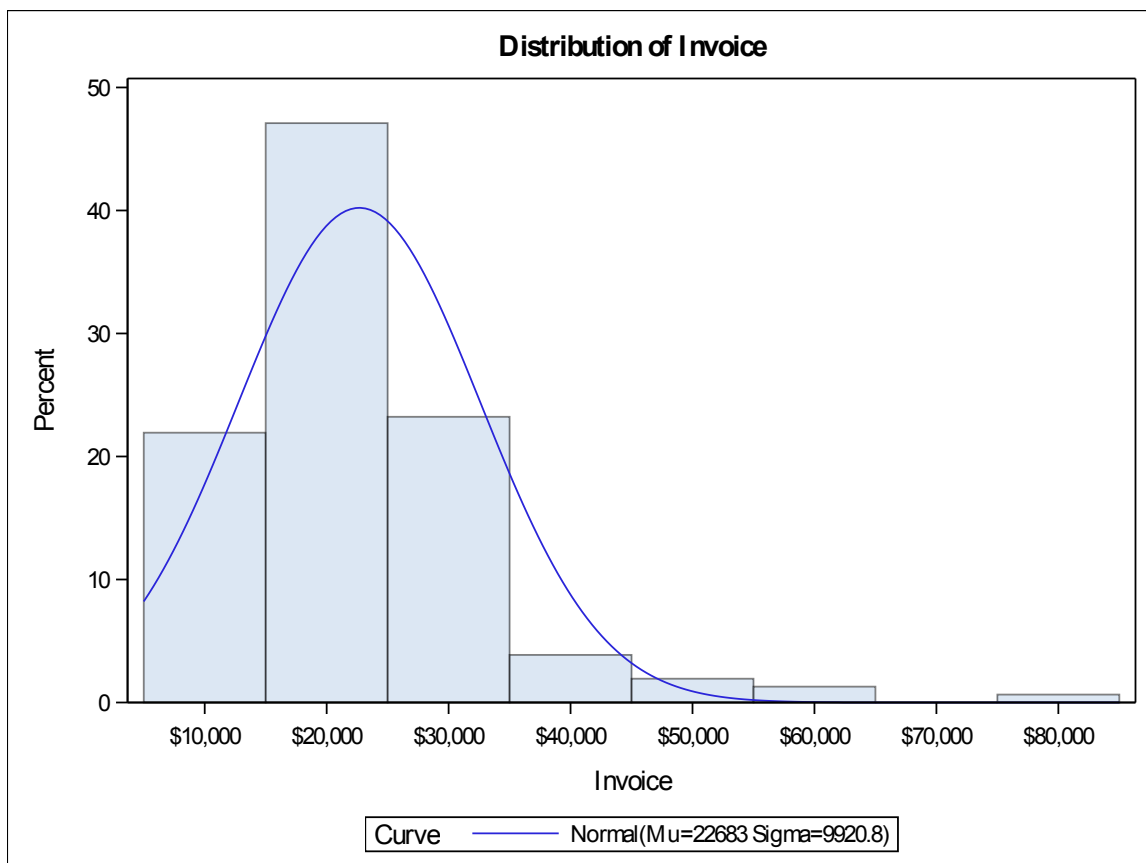
Then test whether the true mean or median Invoice price is \$22000. The test result is:

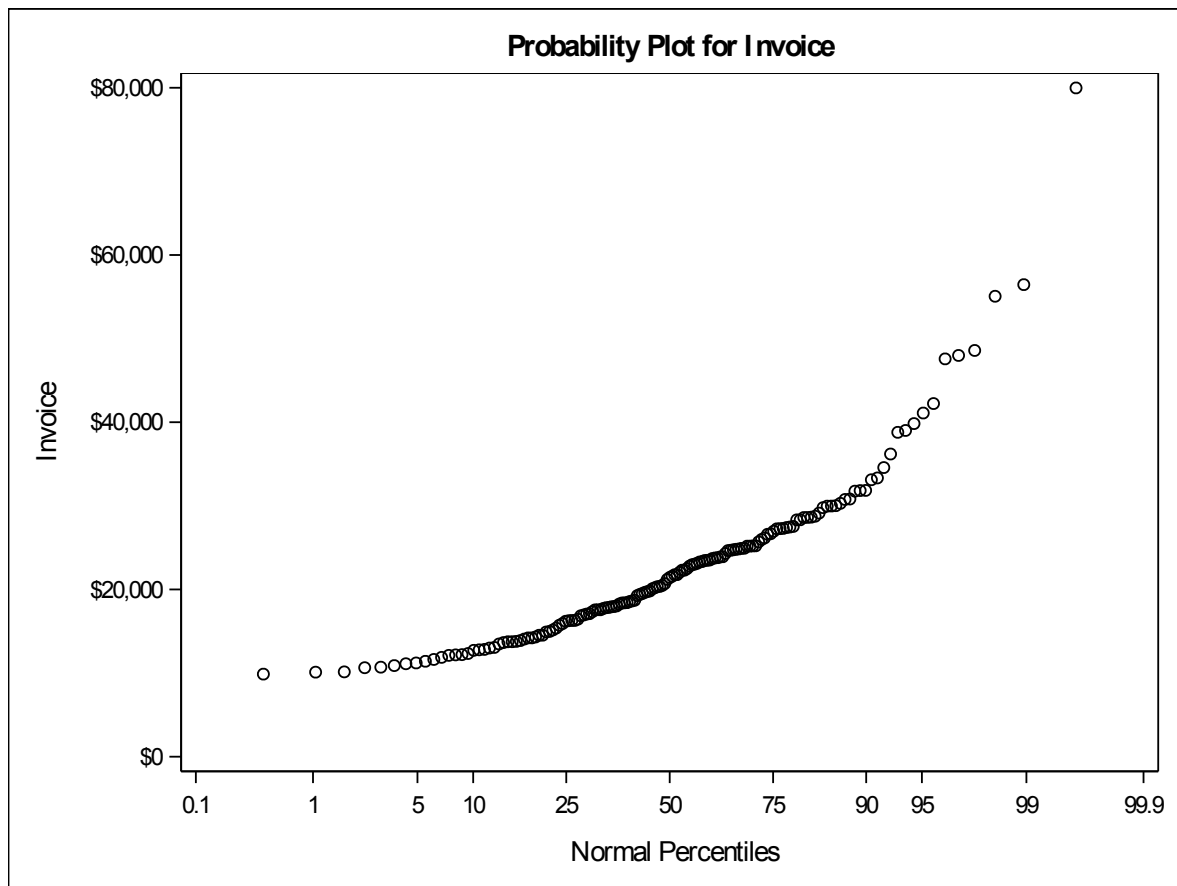
Tests for Location: $\mu_0=22000$				
Test	Statistic		p Value	
Student's t	t	9.442204	$\Pr > t $	<.0001
Sign	M	58.5	$\Pr \geq M $	<.0001
Signed Rank	S	21930	$\Pr \geq S $	<.0001

Based on the normality tests, **Invoice** does not follow a normal distribution. Therefore, the signed rank test should be chosen. The p-value is less than 0.05, so the null hypothesis is rejected. So it is reasonable to say the true mean or median Invoice price is not \$22000.

b) First the normality of **Invoice** in Asia and Europe should be checked.

The tests of normality for **Invoice** in Asia are shown as below:



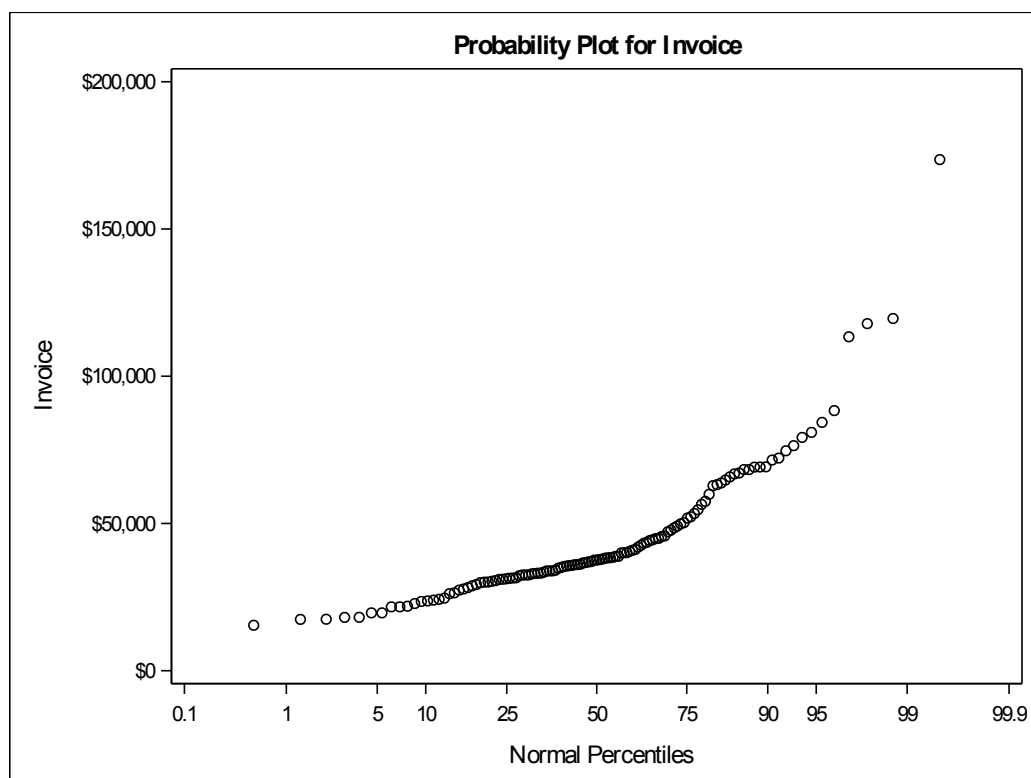
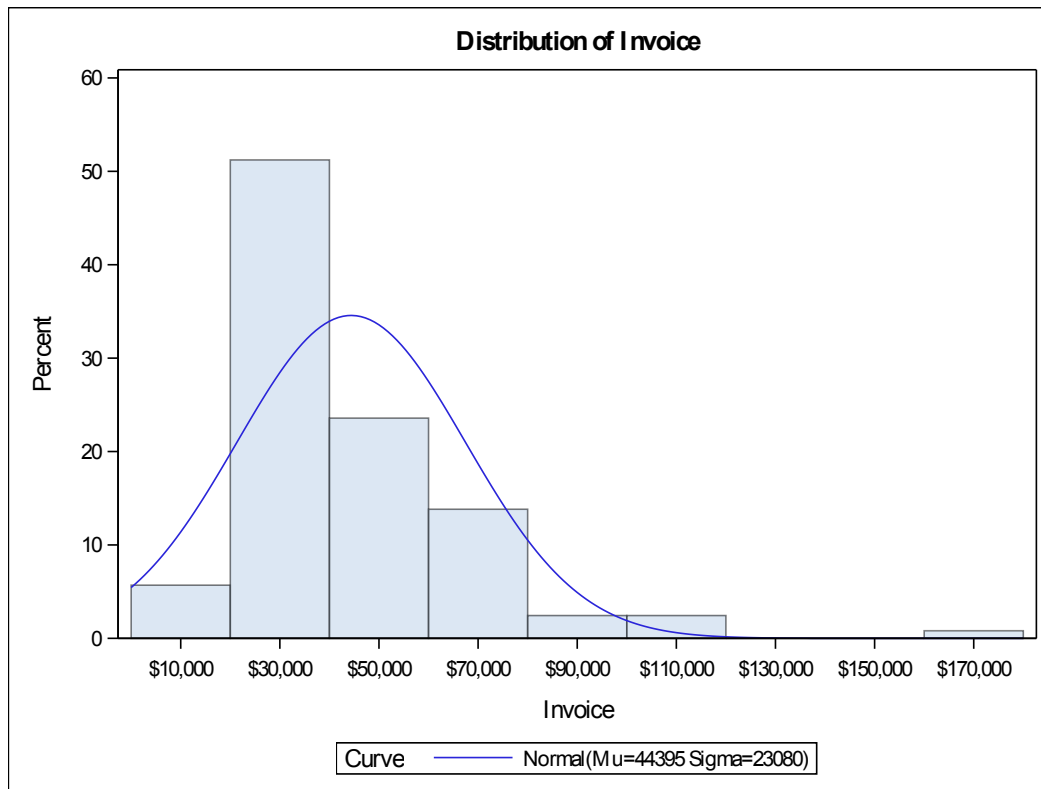


According to the histogram and probplot, the histogram of **Invoice** in Asia seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.846964	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.115286	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.601988	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.140188	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **Invoice** in Asia.

The tests of normality for **Invoice** in Europe are shown as below:



According to the histogram and probplot, the histogram of **Invoice** in Europe seems not fit the curve of normal distribution. Also, the point pattern in probability plot is curved with slope increasing from left to right. The test statistics for normality should be checked.

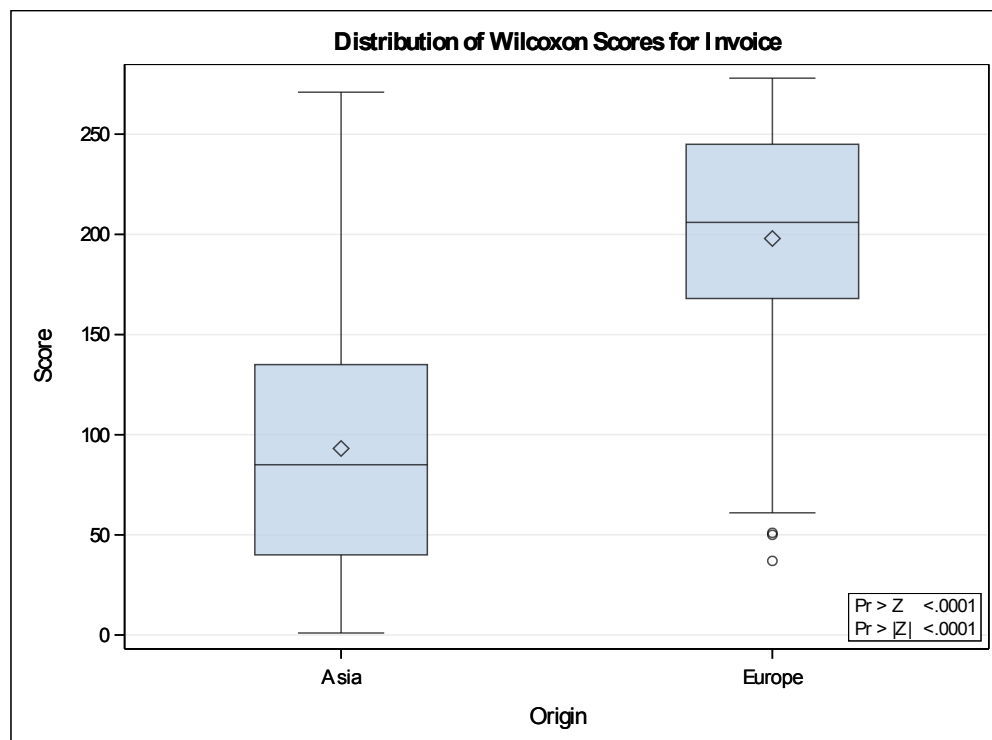
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.798086	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.175505	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.116219	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.068997	Pr > A-Sq	<0.0050

In the table of Test for Normality, the p-value of tests is less than 0.05. It rejects the null hypothesis, which is that there is no significant departure from normality. Therefore, the assumption of normality would not be reasonable for **Invoice** in Europe.

Based on the normality tests, **Invoice** does not follow a normal distribution in Asia and Europe. Therefore, the nonparametric test should be used. Results are shown as below:

Wilcoxon Scores (Rank Sums) for Variable Invoice Classified by Variable Origin					
Origin	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Asia	155	14434.0	21622.50	665.778401	93.122581
Europe	123	24347.0	17158.50	665.778401	197.943089
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	24347.0000
Normal Approximation	
Z	10.7964
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
Z includes a continuity correction of 0.5.	



The above tables and plot show the results of the Wilcoxon analysis. The Wilcoxon two-sample test statistic equals 24347.0, which is the sum of the Wilcoxon scores for Europe. This sum is greater than 17158.50, which is the expected value under the null hypothesis of no difference between the two groups. The one-sided p-value is less than 0.05, so the null hypothesis is rejected. Therefore, cars originated in Europe have significantly greater invoice price than Asian cars.

Exercise 3

a) The Pearson correlation matrix is

Pearson Correlation Coefficients, N = 425 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.82581 <.0001	0.14515 0.0027	0.16206 0.0008
Horsepower	0.82581 <.0001	1.00000	0.38199 <.0001	0.37198 <.0001
Wheelbase Wheelbase (IN)	0.14515 0.0027	0.38199 <.0001	1.00000	0.88882 <.0001
Length Length (IN)	0.16206 0.0008	0.37198 <.0001	0.88882 <.0001	1.00000

Based on the above table, we could know:

- i) The correlation between **Invoice** and **Horsepower** is 0.82581, which is significantly positive. So there is a strong tendency for **Horsepower** to increase with **Invoice**.
- ii) The correlation between **Invoice** and **Wheelbase** is 0.14515, which is significantly positive. So there is a small tendency for **Wheelbase** to increase with **Invoice**.
- iii) The correlation between **Invoice** and **Length** is 0.16206, which is significantly positive. So there is a small tendency for **Length** to increase with **Invoice**.
- iv) The correlation between **Horsepower** and **Wheelbase** is 0.38199, which is significantly positive. So there is a moderate tendency for **Wheelbase** to increase with **Horsepower**.
- v) The correlation between **Horsepower** and **Length** is 0.37198, which is significantly positive. So there is a moderate tendency for **Length** to increase with **Horsepower**.
- vi) The correlation between **Wheelbase** and **Length** is 0.88882, which is significantly positive. So there is a strong tendency for **Length** to increase with **Wheelbase**.

b) Perform the same correlation analysis by **Type**. The results are shown as below:

Type=SUV

Pearson Correlation Coefficients, N = 60 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.76589 <.0001	0.44030 0.0004	0.45230 0.0003
Horsepower	0.76589 <.0001	1.00000	0.70261 <.0001	0.69163 <.0001
Wheelbase Wheelbase (IN)	0.44030 0.0004	0.70261 <.0001	1.00000	0.93951 <.0001
Length Length (IN)	0.45230 0.0003	0.69163 <.0001	0.93951 <.0001	1.00000

Type=Sedan

Pearson Correlation Coefficients, N = 262 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.85442 <.0001	0.59361 <.0001	0.42909 <.0001
Horsepower	0.85442 <.0001	1.00000	0.66364 <.0001	0.55883 <.0001
Wheelbase Wheelbase (IN)	0.59361 <.0001	0.66364 <.0001	1.00000	0.85600 <.0001
Length Length (IN)	0.42909 <.0001	0.55883 <.0001	0.85600 <.0001	1.00000

Type=Sports

Pearson Correlation Coefficients, N = 49 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.80000 <.0001	-0.05278 0.7187	0.34370 0.0156
Horsepower	0.80000 <.0001	1.00000	0.31025 0.0300	0.56576 <.0001
Wheelbase Wheelbase (IN)	-0.05278 0.7187	0.31025 0.0300	1.00000	0.67539 <.0001
Length Length (IN)	0.34370 0.0156	0.56576 <.0001	0.67539 <.0001	1.00000

Type=Truck

Pearson Correlation Coefficients, N = 24 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.84423 <.0001	0.48689 0.0158	0.49544 0.0138
Horsepower	0.84423 <.0001	1.00000	0.74923 <.0001	0.71352 <.0001
Wheelbase Wheelbase (IN)	0.48689 0.0158	0.74923 <.0001	1.00000	0.94404 <.0001
Length Length (IN)	0.49544 0.0138	0.71352 <.0001	0.94404 <.0001	1.00000

Type=Wagon

Pearson Correlation Coefficients, N = 30 Prob > r under H0: Rho=0				
	Invoice	Horsepower	Wheelbase	Length
Invoice	1.00000	0.83032 <.0001	0.60500 0.0004	0.47532 0.0079
Horsepower	0.83032 <.0001	1.00000	0.64068 0.0001	0.49217 0.0057
Wheelbase Wheelbase (IN)	0.60500 0.0004	0.64068 0.0001	1.00000	0.83381 <.0001
Length Length (IN)	0.47532 0.0079	0.49217 0.0057	0.83381 <.0001	1.00000

The relationships between vehicle attributes for each type are:

- i) For SUV, all of the correlations are significant. The correlations between **Invoice** and **Horsepower**, **Horsepower** and **Wheelbase**, **Horsepower** and **Length**, **Wheelbase** and **Length** are strong. The correlations between **Invoice** and **Wheelbase**, **Invoice** and **Length** are moderate.
- ii) For Sedan, all of the correlations are significant. The correlations between **Invoice** and **Horsepower**, **Wheelbase** and **Length** are strong. The correlations between **Horsepower** and **Wheelbase**, **Horsepower** and **Length**, **Invoice** and **Wheelbase**, **Invoice** and **Length** are moderate.
- iii) For Sports, all other pairs are significant except the correlation between **Invoice** and **Wheelbase**. The correlation between **Invoice** and **Horsepower** is strong. The correlations

between **Invoice** and **Length**, **Horsepower** and **Wheelbase**, **Horsepower** and **Length**, **Wheelbase** and **Length** are moderate.

iv) For Truck, all of the correlations are significant. The correlations between **Invoice** and **Horsepower**, **Horsepower** and **Wheelbase**, **Horsepower** and **Length**, **Wheelbase** and **Length** are strong. The correlations between **Invoice** and **Wheelbase**, **Invoice** and **Length** are moderate.

v) For Wagon, all of the correlations are significant. The correlations between **Invoice** and **Horsepower**, **Wheelbase** and **Length** are strong. The correlations between **Horsepower** and **Wheelbase**, **Horsepower** and **Length**, **Invoice** and **Wheelbase**, **Invoice** and **Length** are moderate.

The differences with the results in part (a) are:

i) The weak correlation between **Invoice** and **Wheelbase**, **Invoice** and **Length** are not found in the results of each type.

ii) All of the six significant correlations in part (a) are also found in SUV, Sedan, Truck, Wagon but only five of them holds in Sports.