Samin Naji – 9812399057

In both Dataset1 & Dataset2:

1. Determine the number of variables and observations in each dataset.
3. Calculate the appropriate statistics for each variable.

Dataset1:

File Information

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format
V1	1	<none>	Scale	Input	8	Right	F4	F4
age	2	<none>	Scale	Input	8	Right	F2	F2
marital	3	<none>	Nominal	Input	9	Left	A9	A9
income	4	<none>	Scale	Input	8	Right	F3	F3
gender	5	<none>	Nominal	Input	6	Left	A6	A6
jobsat	6	<none>	Nominal	Input	21	Left	A21	A21

Variables in the working file

Frequencies

Statistics

		V1	age	income
N	Valid	107	107	107
	Missing	0	0	0
Std. Error of Mean		178.656	1.204	9.947
Std. Deviation		1848.029	12.457	102.895
Variance		3415212.770	155.189	10587.279
Skewness		.031	.558	4.421
Std. Error of Skewness		.234	.234	.234
Kurtosis		-1.218	-.340	26.506
Std. Error of Kurtosis		.463	.463	.463
Range		6258	54	810
Minimum		127	19	9
Maximum		6385	73	819

Frequency Table

marital

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	51	47.7	47.7	47.7
	Unmarried	56	52.3	52.3	100.0
	Total	107	100.0	100.0	

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	52	48.6	48.6	48.6
	Male	55	51.4	51.4	100.0
	Total	107	100.0	100.0	

jobsat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Highly dissatisfied	17	15.9	15.9	15.9
	Highly satisfied	27	25.2	25.2	41.1
	Neutral	28	26.2	26.2	67.3
	Somewhat dissatisfied	15	14.0	14.0	81.3
	Somewhat satisfied	20	18.7	18.7	100.0
	Total	107	100.0	100.0	

Dataset2:

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
V1	34	56	4	60	33.65	2.984	17.399	302.720	-.106	.403	-1.229	.788
Before	34	27.82019103	26.94125562	54.76144665	44.58756469	1.168312451	6.812373702	46.408	-.728	.403	.606	.788
After	34	11.67845463	42.45801695	54.13647158	48.78590495	.4870769240	2.840122113	8.066	-.323	.403	-.046	.788
Valid N (listwise)	34											

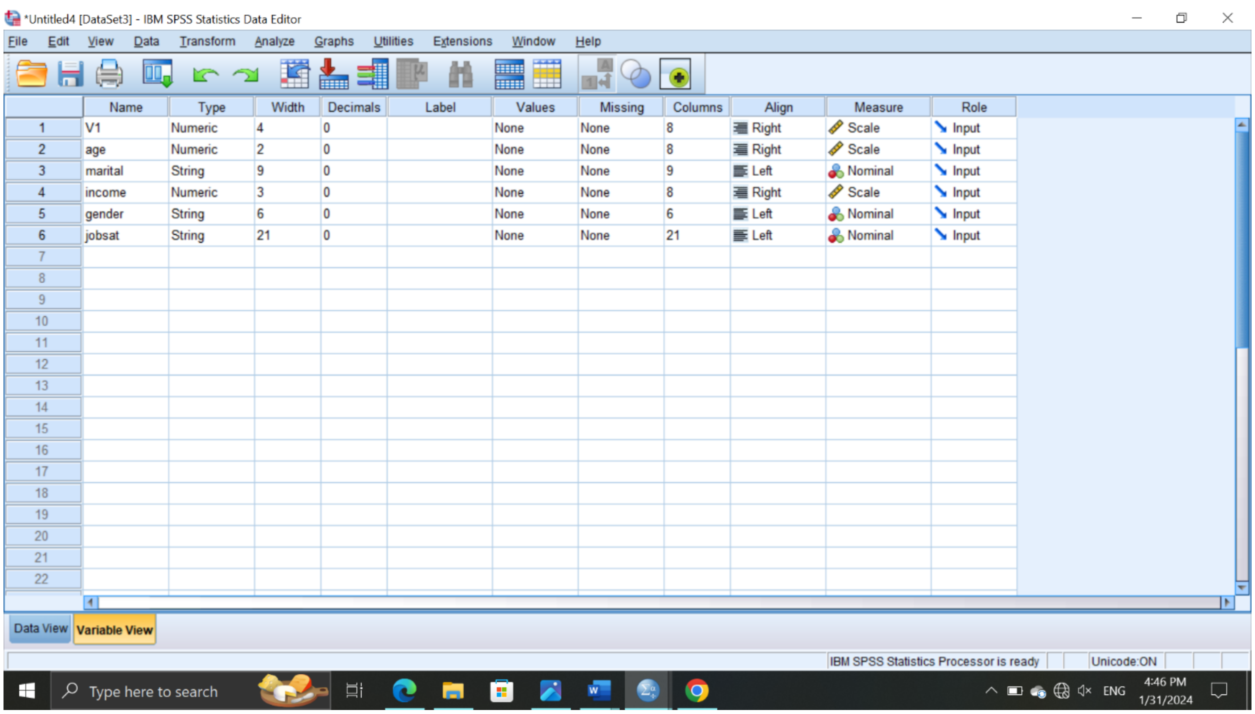
2. Identify which variables are quantitative and which are qualitative.

You can use the **Variable View** tab in SPSS to see the names, types, and labels of the variables in your dataset. The type of variable will determine what kind of statistical analysis you can perform on it.

some general rules to help classify our variables:

- If the variable has a decimal point, it is likely a quantitative variable.
- If the variable has only two possible values, it is likely a binary variable, which is a type of qualitative variable.
- If the variable has more than two possible values, but they are not ordered or ranked, it is likely a nominal variable, which is another type of qualitative variable.
- If the variable has more than two possible values, and they are ordered or ranked, it is likely an ordinal variable, which is the third type of qualitative variable.

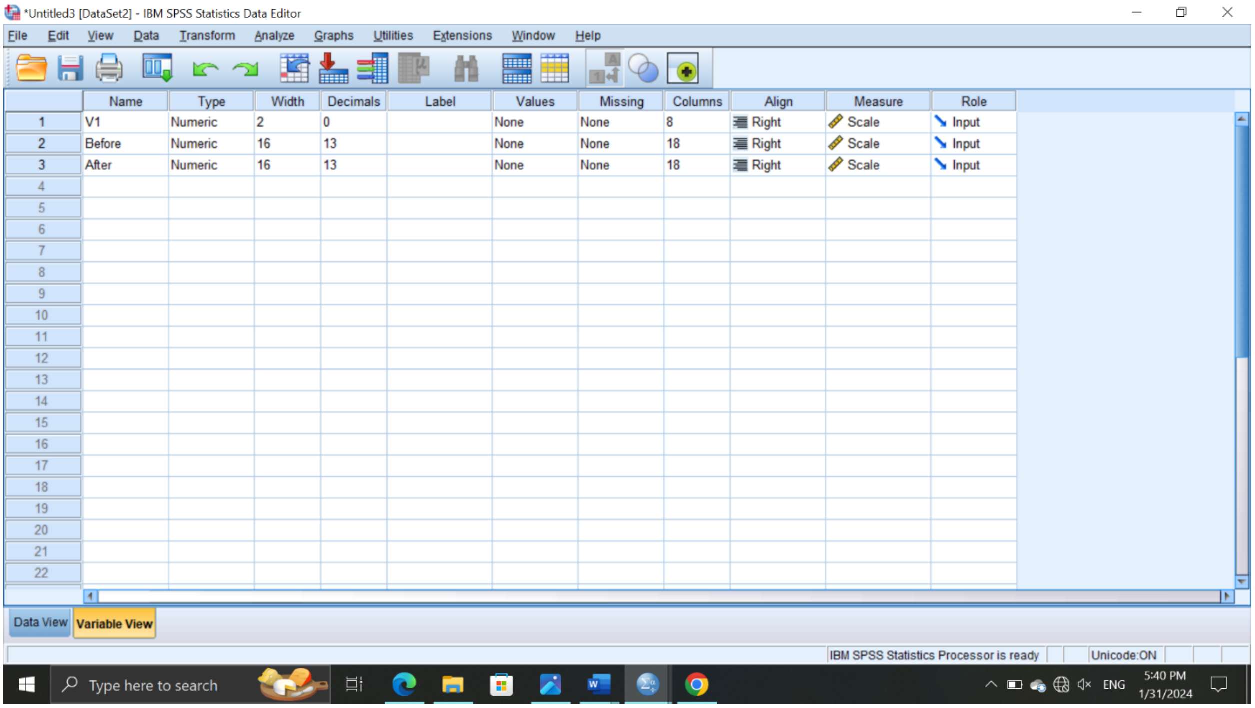
Dataset1:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	V1	Numeric	4	0		None	None	8	Right	Scale	Input
2	age	Numeric	2	0		None	None	8	Right	Scale	Input
3	marital	String	9	0		None	None	9	Left	Nominal	Input
4	income	Numeric	3	0		None	None	8	Right	Scale	Input
5	gender	String	6	0		None	None	6	Left	Nominal	Input
6	jobsat	String	21	0		None	None	21	Left	Nominal	Input

- **V1**: This variable is **quantitative**, because it is numeric and scale.
- **age**: This variable is **quantitative**, because it is numeric and scale.
- **marital**: This variable is **qualitative**, because it is string and nominal.
- **income**: This variable is **quantitative**, because it is numeric and scale.
- **gender**: This variable is **qualitative**, because it is string and nominal.
- **jobsat**: This variable is **qualitative**, because it is string and nominal.

Dataset2:



- **V1**: This variable is **quantitative**, because it is numeric and scale.
- **Before**: This variable is **quantitative**, because it is numeric and scale.
- **After**: This variable is **quantitative**, because it is numeric and scale.

4. Create suitable charts for each variable.

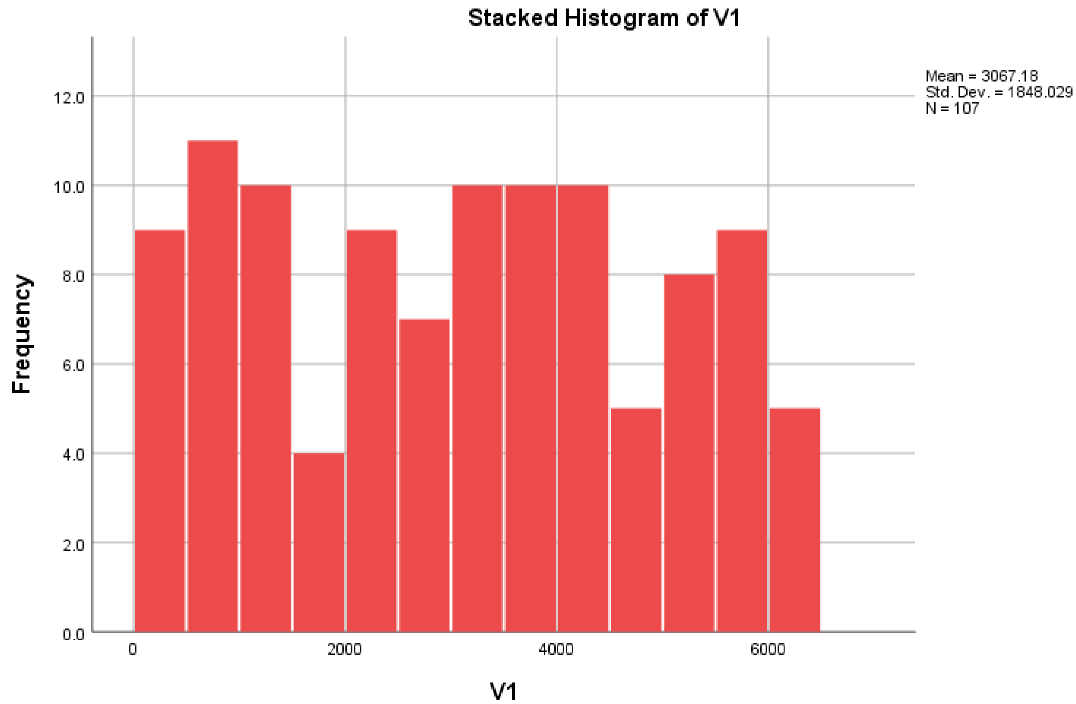
Scale variables are usually measured at the interval or ratio level of measurement. Therefore, the suitable graphs for scale variables in SPSS are the ones that can show the distribution, outliers, or relationship of your variables, such as histograms, boxplots, or scatterplots.

- Histograms show the frequency of values in intervals or bins, and you can also overlay a normal curve to check the normality of your data.
- Boxplots show the median, quartiles, and extreme values of your variables, and they can help you identify outliers or skewness.
- Scatterplots show the relationship between two variables by plotting them as dots on a coordinate plane.

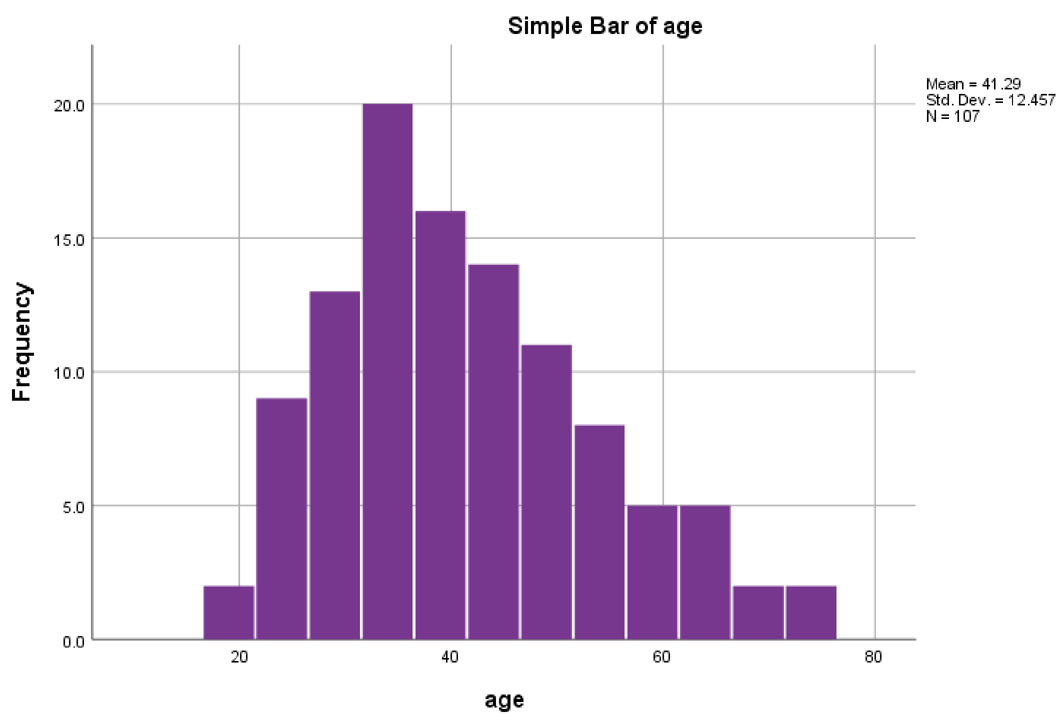
For nominal or ordinal variables, we can use bar graphs, pie charts, or histograms to show the frequency or percentage of each category.

- Bar graphs are useful for comparing multiple categories or groups, while
- pie charts are good for showing the proportion of each category in the whole.
- Histograms are similar to bar graphs, but they show the distribution of a single variable in intervals or bins.

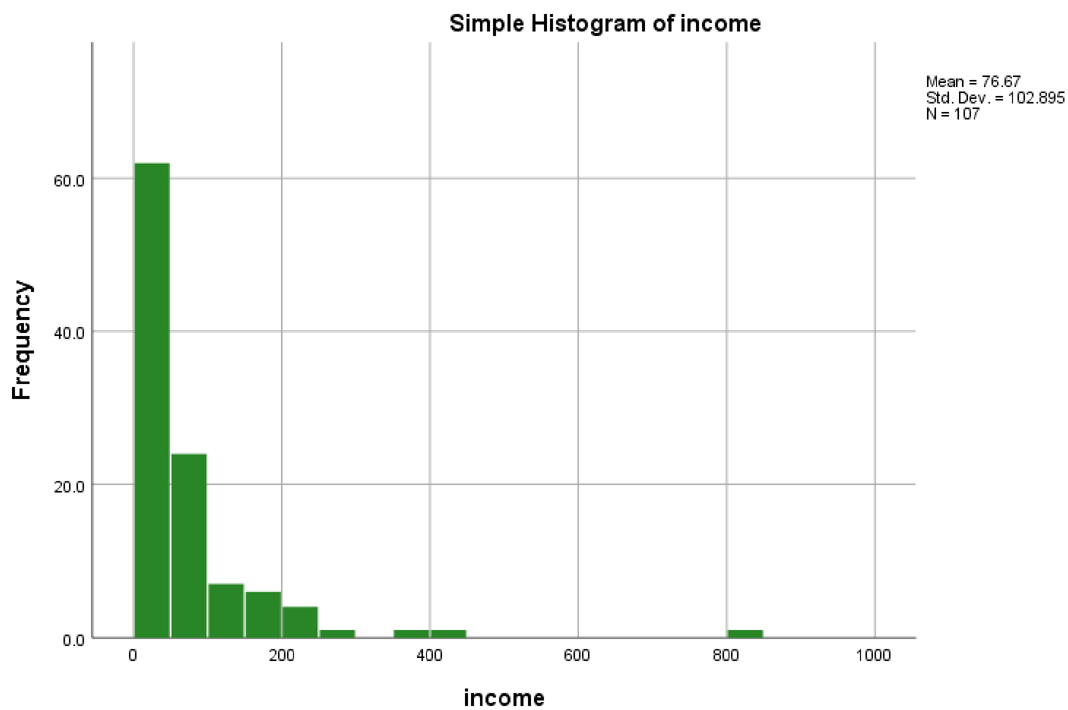
Dataset1:

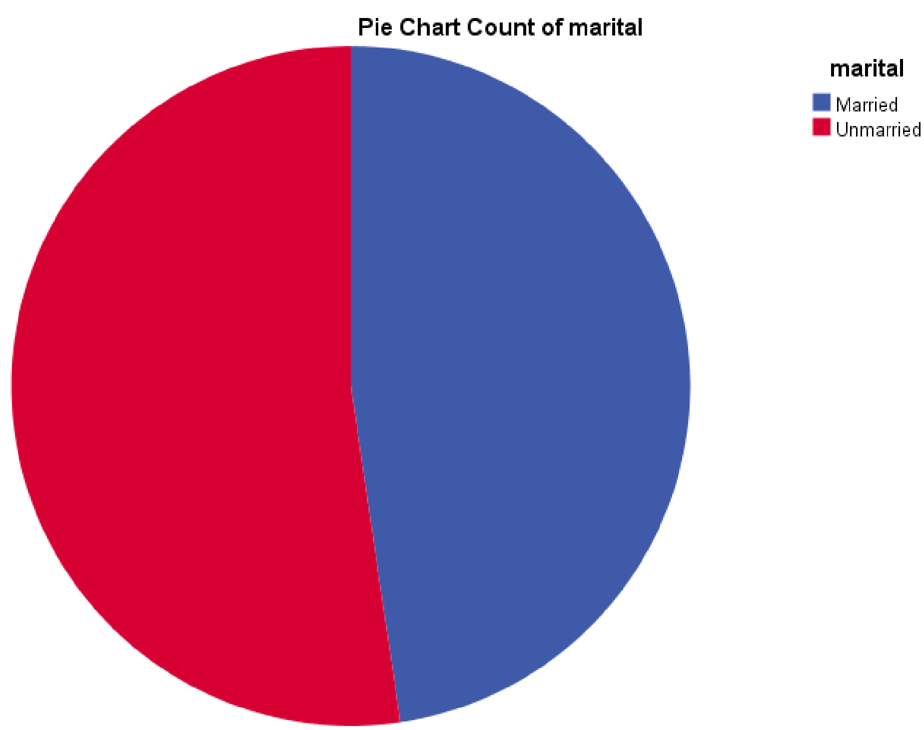


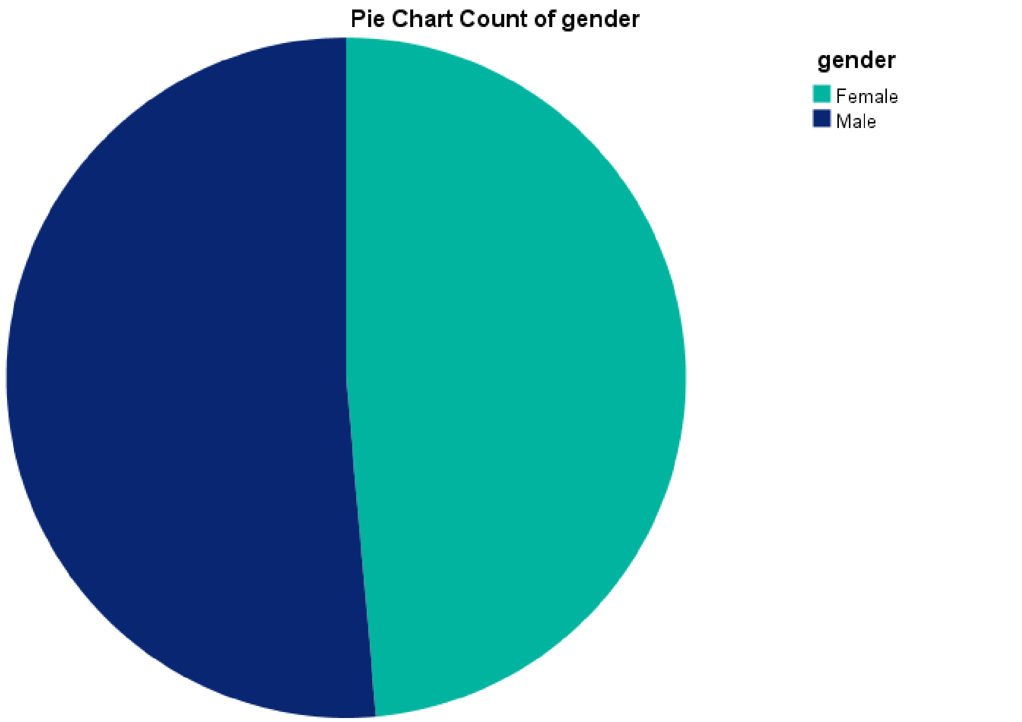
Stacked Histogram of V1
Mean = 3067.18
Std. Dev. = 1848.029
N = 107
Frequency
0.0
2.0
4.0
6.0
8.0
10.0
12.0
0
2000
4000
6000
V1

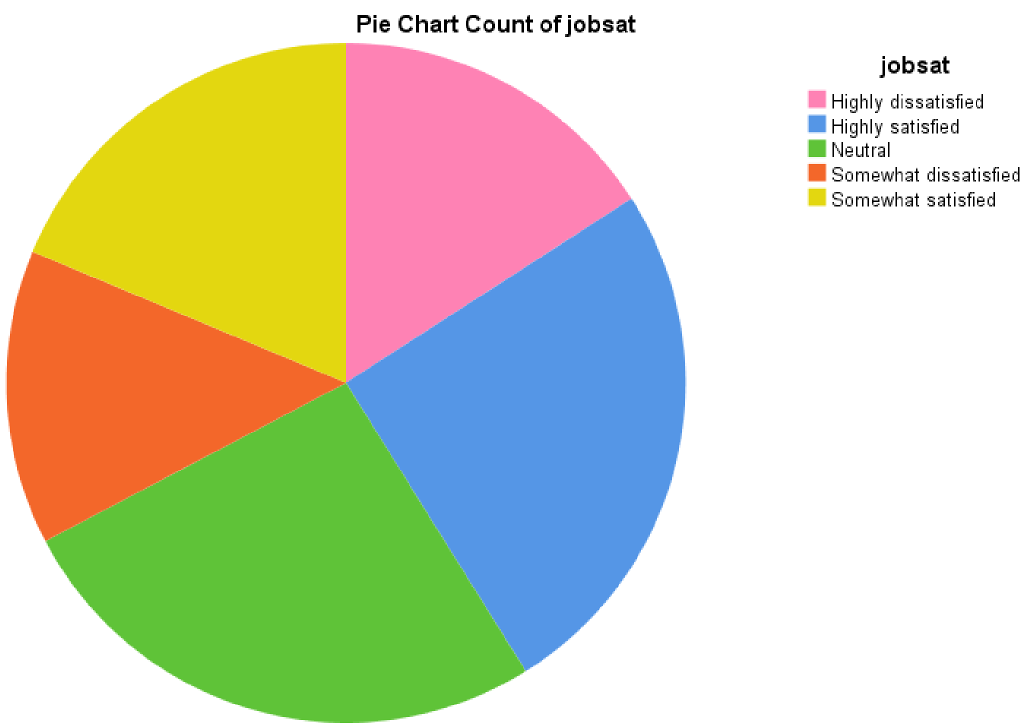


Simple Bar of age
Mean = 41.29
Std. Dev. = 12.457
N = 107
Frequency
0.0
5.0
10.0
15.0
20.0
20
40
60
80
age


Simple Histogram of income
Mean = 76.67
Std. Dev. = 102.895
N = 107
Frequency
60.0
40.0
20.0
0.0
0
200
400
600
800
1000
income


Pie Chart Count of marital
marital
Married
Unmarried

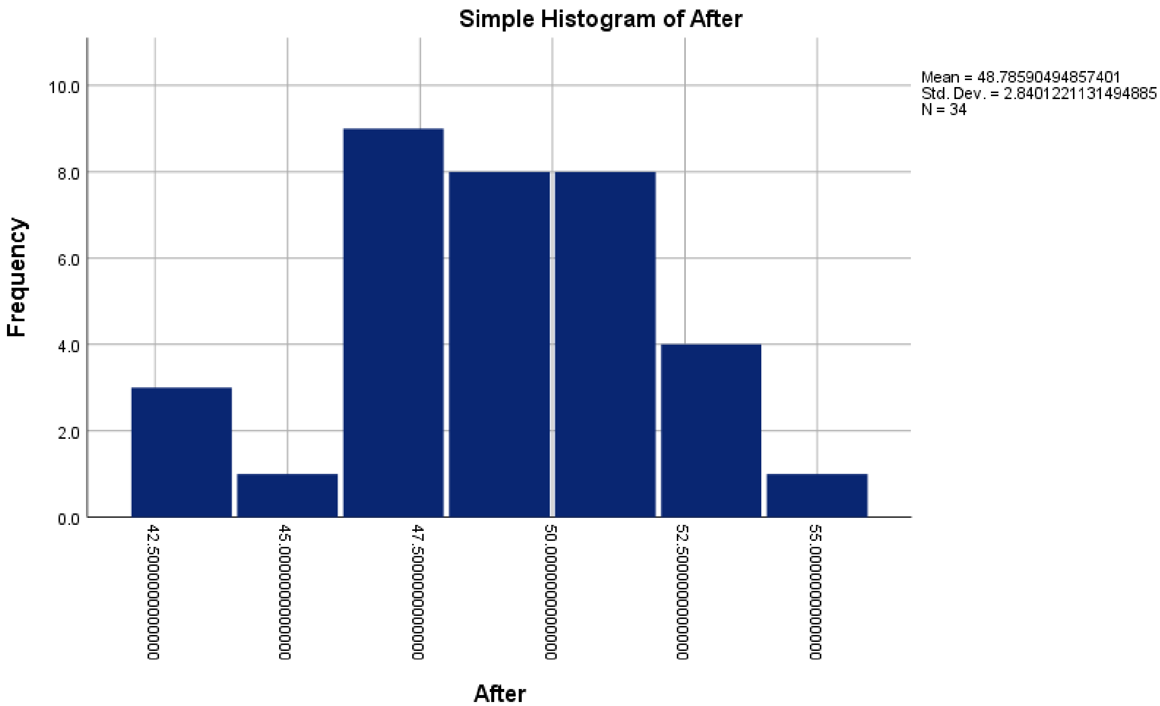

Pie Chart Count of gender
gender
Female
Male


Pie Chart Count of jobsat
jobsat
Highly dissatisfied
Highly satisfied
Neutral
Somewhat dissatisfied
Somewhat satisfied

Datasat2:


Simple Histogram of V1
Mean = 33.65
Std. Dev. = 17.399
N = 34
Frequency
5.0
4.0
3.0
2.0
1.0
0.0
0
20
40
60
V1


Simple Histogram of Before
Mean = 44.58756469197625
Std. Dev. = 6.812373701867224
N = 34
Frequency
12.0
10.0
8.0
6.0
4.0
2.0
0.0
25.000000000000
30.000000000000
35.000000000000
40.000000000000
45.000000000000
50.000000000000
55.000000000000
Before



5. In Dataset1, compare the income distribution between male and female categories. Check for normality if necessary.

T-Test

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
income	Male	55	82.11	87.944	11.858
	Female	52	70.92	117.262	16.261

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
income	Equal variances assumed	.232	.631	.560	105	.577	11.186	19.967	-28.405	50.777
	Equal variances not assumed			.556	94.441	.580	11.186	20.126	-28.772	51.144

To check for normality of income for each gender group in SPSS we will get an output table that shows the descriptive statistics and the results of two tests for normality: the Kolmogorov-Smirnov test and the Shapiro-Wilk test. These tests compare the observed distribution of income to a theoretical normal distribution and calculate the probability of observing a deviation as large or larger than the one you observed, if the data was normally distributed. Again, a common rule

of thumb is to reject the null hypothesis of normality if the p-value is less than 0.05. This means that the data is unlikely to be normally distributed.
we will also get a normality plot for each group, which is a graphical way of checking for normality. The normality plot shows the observed values of income on the x-axis and the expected values of a normal distribution on the y-axis. If the data is normally distributed, the points should fall along a straight diagonal line. If the data is skewed or has outliers, the points will deviate from the line.

Explore

Gender

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	gender	N	Percent	N	Percent	N	Percent
income	Female	52	100.0%	0	0.0%	52	100.0%
	Male	55	100.0%	0	0.0%	55	100.0%

Descriptives

	gender			Statistic	Std. Error
income	Female	Mean		70.92	16.261
		95% Confidence Interval for Mean	Lower Bound	38.28	
			Upper Bound	103.57	
		5% Trimmed Mean		52.62	
		Median		38.00	
		Variance		13750.308	
		Std. Deviation		117.262	
		Minimum		9	
		Maximum		819	
		Range		810	
		Interquartile Range		42	
		Skewness		5.411	.330
		Kurtosis		33.589	.650
	Male	Mean		82.11	11.858
		95% Confidence Interval for Mean	Lower Bound	58.33	
			Upper Bound	105.88	
		5% Trimmed Mean		70.52	
		Median		43.00	
		Variance		7734.099	
		Std. Deviation		87.944	
		Minimum		12	
		Maximum		440	
		Range		428	
		Interquartile Range		80	
		Skewness		2.214	.322
		Kurtosis		5.460	.634

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	gender	Statistic	df	Sig.	Statistic	df	Sig.
income	Female	.320	52	.000	.414	52	.000
	Male	.213	55	.000	.724	55	.000

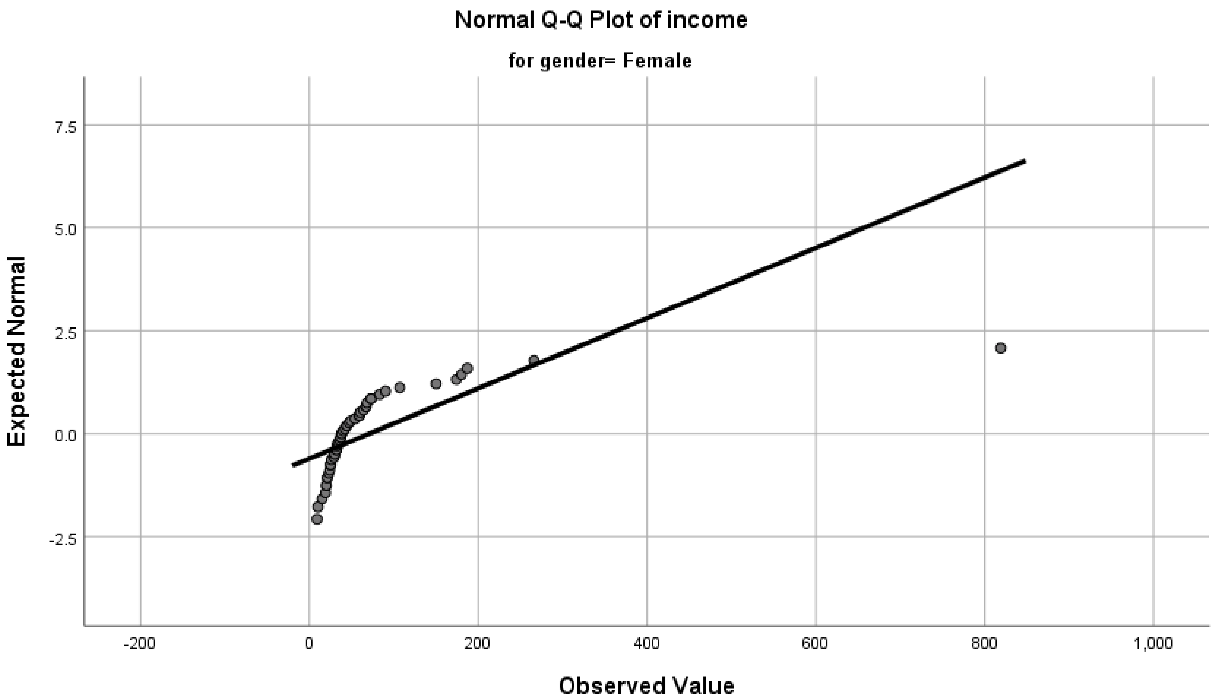
a. Lilliefors Significance Correction

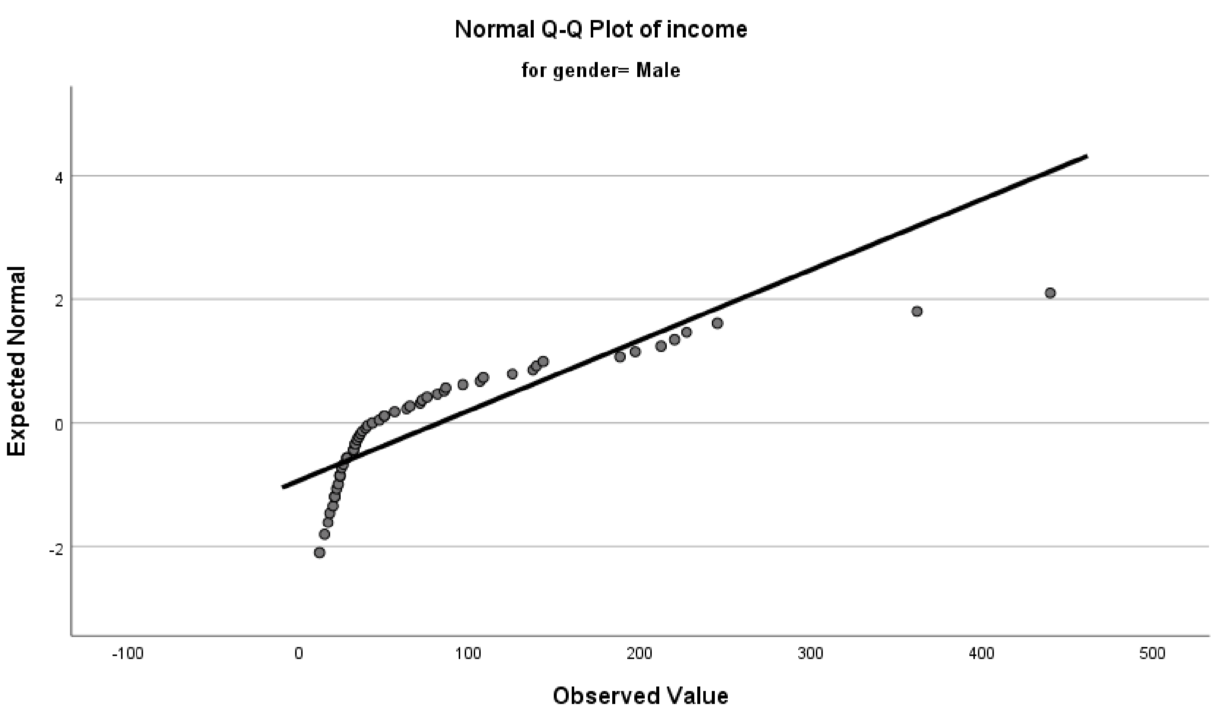
Income

Stem-and-Leaf Plots

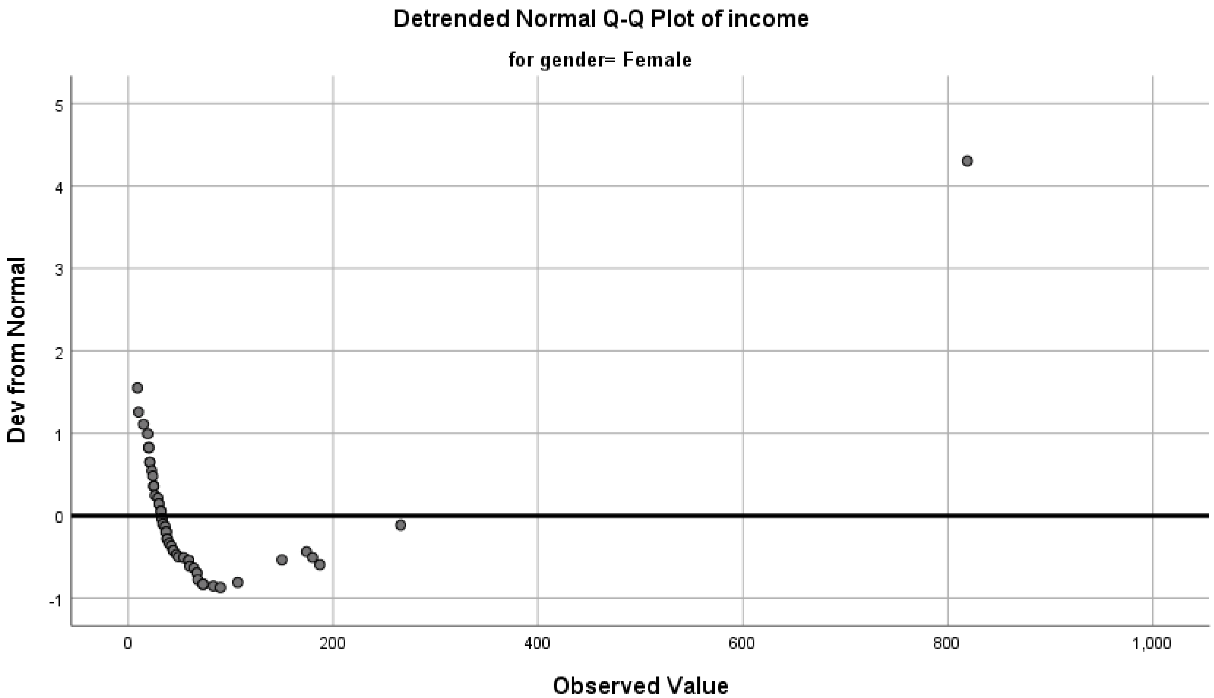
```
income Stem-and-Leaf Plot for
gender= Female

 Frequency    Stem &  Leaf

     1.00        0 .  9
     3.00        1 .  059
    11.00        2 .  00113455569
    12.00        3 .  002233467788
     6.00        4 .  024479
     3.00        5 .  499
     5.00        6 .  04778
     2.00        7 .  33
     1.00        8 .  3
     1.00        9 .  0
     1.00       10 .  7
     6.00 Extremes    (>=150)

 Stem width:   10
 Each leaf:        1 case(s)


income Stem-and-Leaf Plot for
gender= Male

 Frequency    Stem &  Leaf

    29.00        0 .  11112222222222222233333333444
    12.00        0 .  555667778889
     6.00        1 .  002334
     2.00        1 .  89
     1.00        2 .  1
     5.00 Extremes    (>=220)

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

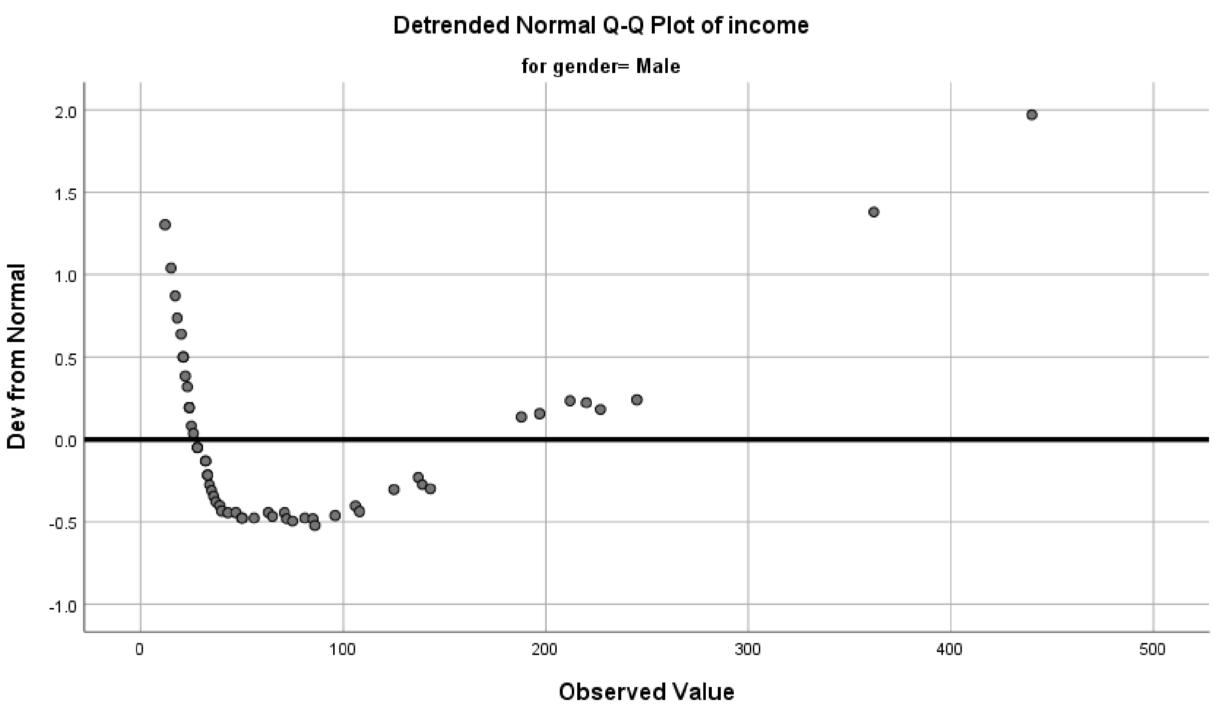
Normal Q-Q Plots

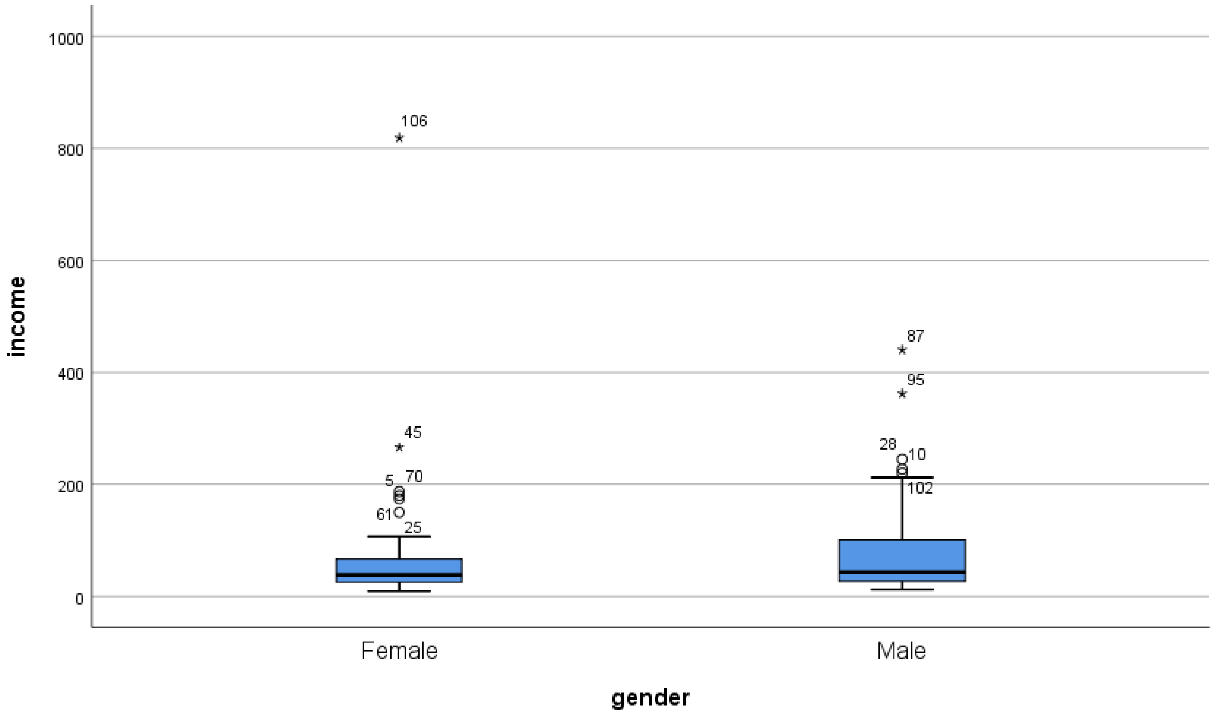




Detrended Normal Q-Q Plots







6. In Dataset1, compare the income mean to 65.

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
income	107	76.67	102.895	9.947

One-Sample Test

	Test Value = 65					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
income	1.173	106	.243	11.673	-8.05	31.39

7. In Dataset2, compare the 'Before' and 'After' columns using the appropriate test. Check for normality if necessary.

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	After	48.78590495	34	2.840122113	.4870769240
	Before	44.58756469	34	6.812373702	1.168312451

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	After & Before	34	-.154	.385

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	After - Before	4.198340257	7.773880257	1.333209465	1.485905205	6.910775308	3.149	33	.003

To check for normality of income for each gender group in SPSS, we will get an output table that shows the descriptive statistics and the results of two tests for normality: the Kolmogorov-Smirnov test and the Shapiro-Wilk test. These tests compare the observed distribution of income to a theoretical normal distribution and calculate the probability of observing a deviation as large or larger than the one you observed, if the data was normally distributed. Again, a common rule of thumb is to reject the null hypothesis of normality if the p-value is less than 0.05. This means that the data is unlikely to be normally distributed.

we will also get a normality plot for each group, which is a graphical way of checking for normality. The normality plot shows the observed values of income on the x-axis and the expected values of a normal distribution on the y-axis. If the data is normally distributed, the points should fall along a straight diagonal line. If the data is skewed or has outliers, the points will deviate from the line.

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Before	34	100.0%	0	0.0%	34	100.0%
After	34	100.0%	0	0.0%	34	100.0%

Descriptives

			Statistic	Std. Error
Before	Mean		44.58756469	1.168312451
	95% Confidence Interval for Mean	Lower Bound	42.21061514	
		Upper Bound	46.96451425	
	5% Trimmed Mean		44.96544033	
	Median		45.14309516	
	Variance		46.408	
	Std. Deviation		6.812373702	
	Minimum		26.94125562	
	Maximum		54.76144665	
	Range		27.82019103	
	Interquartile Range		7.322285228	
	Skewness		-.728	.403
	Kurtosis		.606	.788
After	Mean		48.78590495	.4870769240
	95% Confidence Interval for Mean	Lower Bound	47.79493950	
		Upper Bound	49.77687040	
	5% Trimmed Mean		48.85235719	
	Median		49.05939276	
	Variance		8.066	
	Std. Deviation		2.840122113	
	Minimum		42.45801695	
	Maximum		54.13647158	
	Range		11.67845463	
	Interquartile Range		3.581470420	
	Skewness		-.323	.403
	Kurtosis		-.046	.788

Tests of Normality

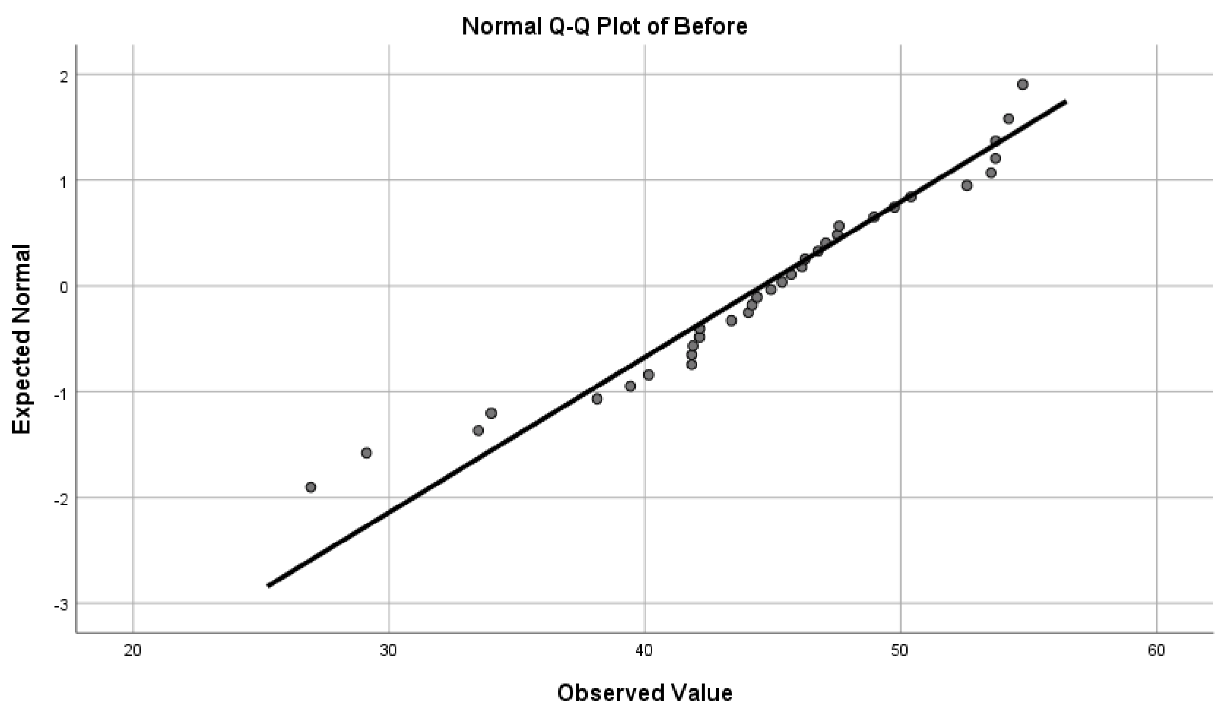
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Before	.137	34	.107	.944	34	.081
After	.077	34	.200*	.978	34	.703

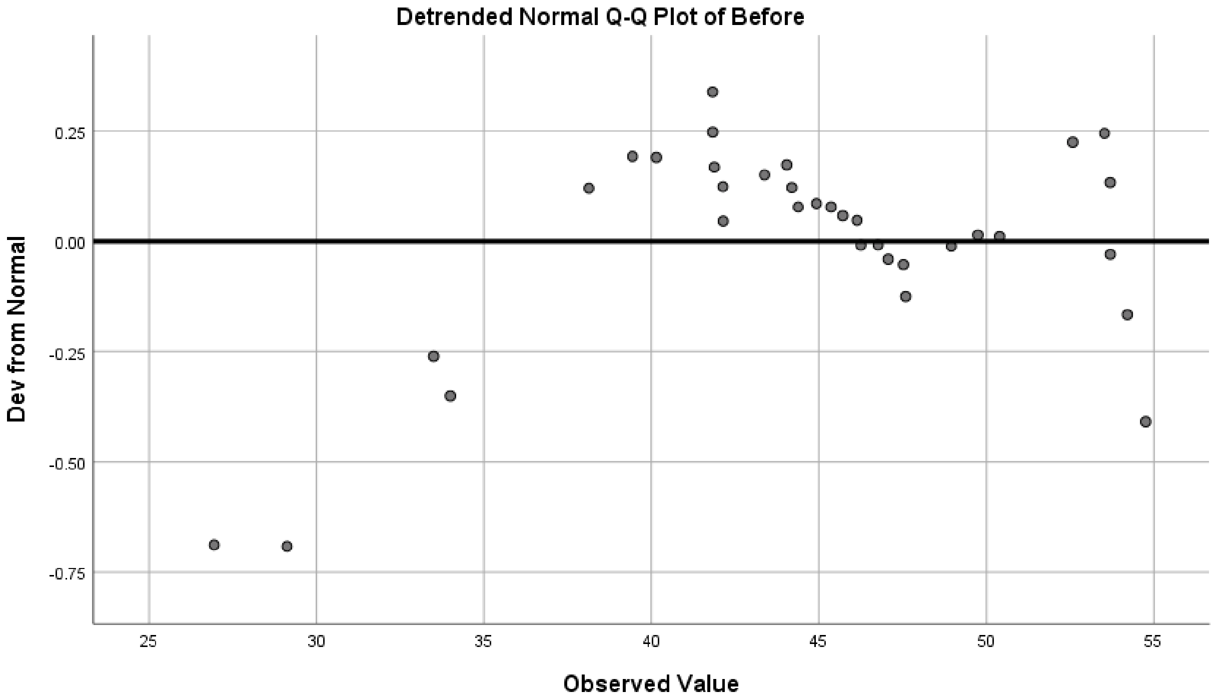
*. This is a lower bound of the true significance.

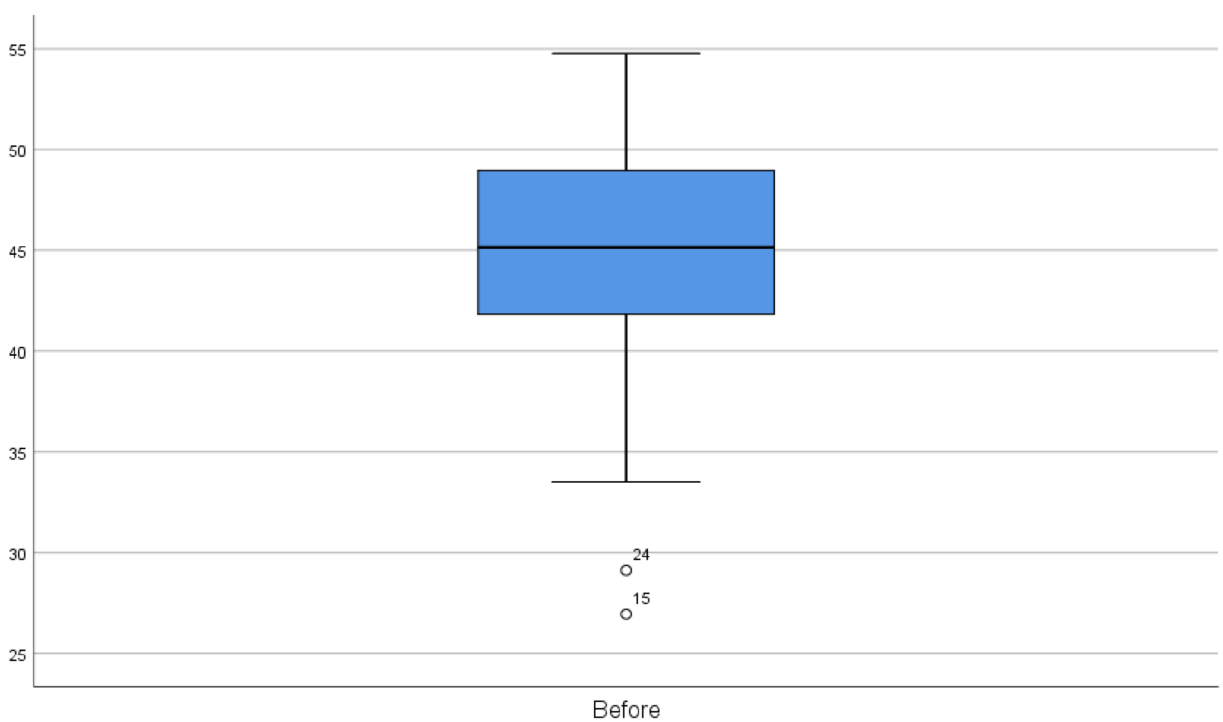
a. Lilliefors Significance Correction

Before

```
Before Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2.00 Extremes    (=<29)
     2.00        3 .  33
     2.00        3 .  89
    11.00        4 .  01112234444
    10.00        4 .  5566677789
     7.00        5 .  0233344

 Stem width:  10.00000
 Each leaf:        1 case(s)
```

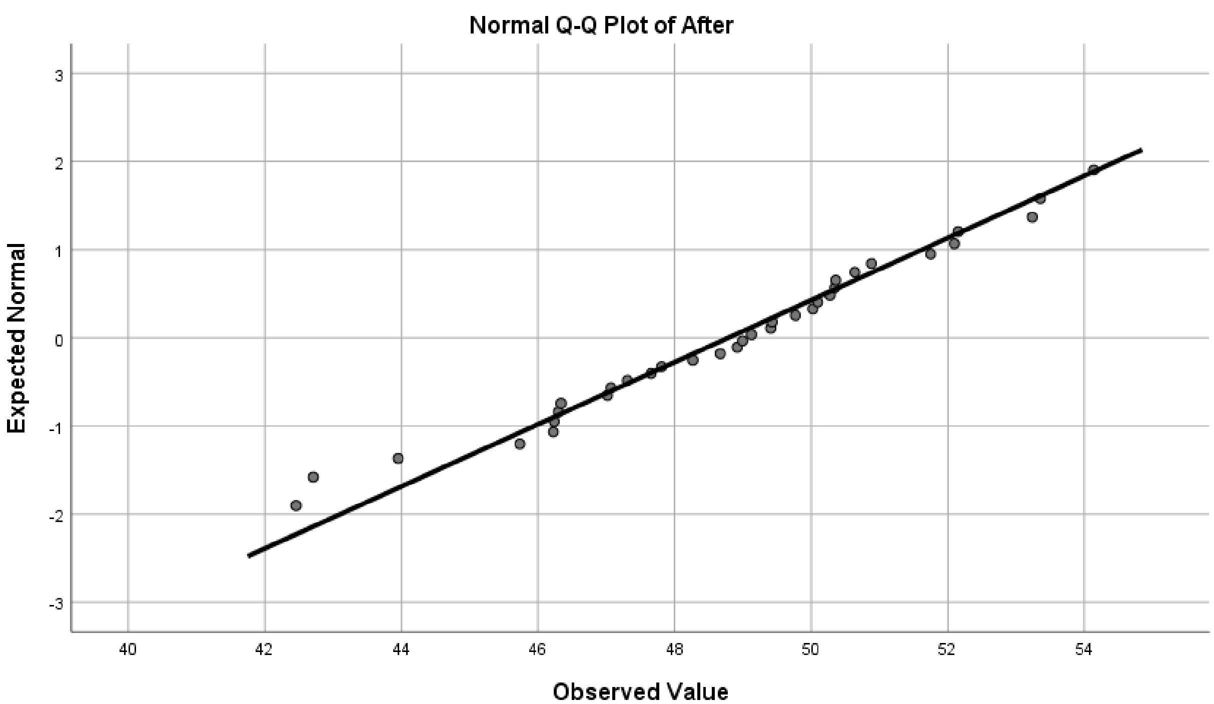


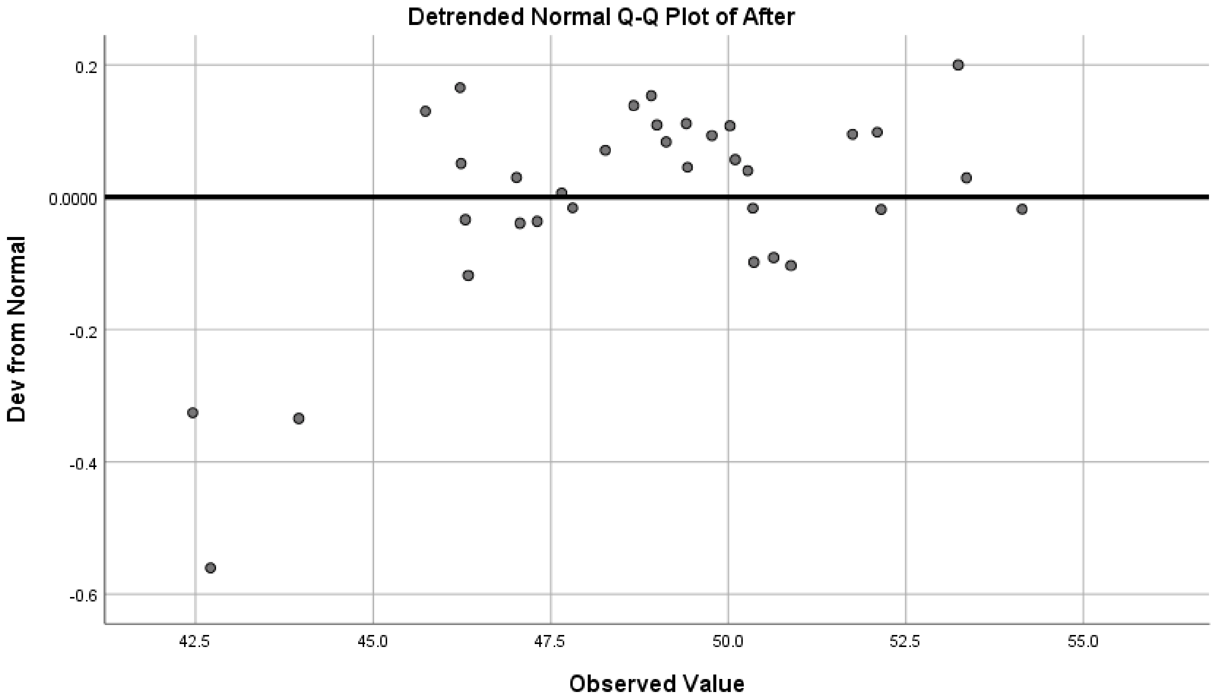

Detrended Normal Q-Q Plot of Before
Dev from Normal
0.25
0.00
-0.25
-0.50
-0.75
25
30
35
40
45
50
55
Observed Value

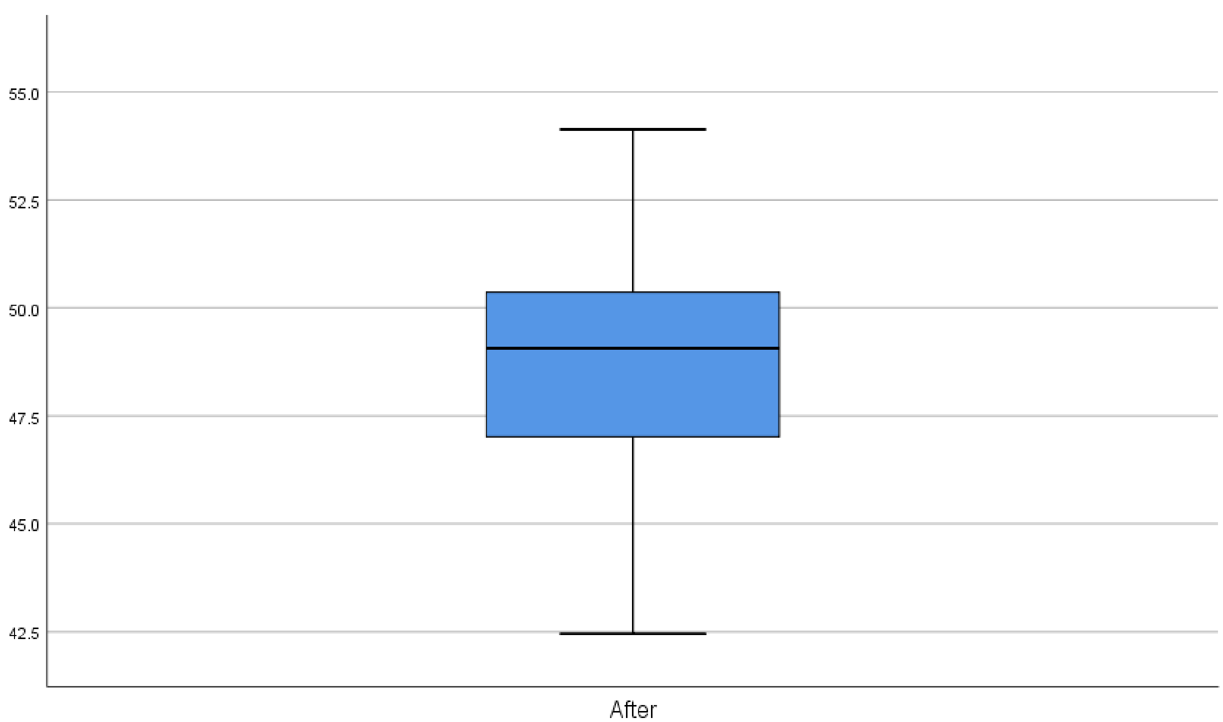

55
50
45
40
35
30
25
24
15
Before

After

```
After Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     3.00        4 .  223
     1.00        4 .  5
     9.00        4 .  666677777
     8.00        4 .  88889999
     8.00        5 .  00000001
     4.00        5 .  2233
     1.00        5 .  4

 Stem width:  10.00000
 Each leaf:        1 case(s)
```




Detrended Normal Q-Q Plot of After
Dev from Normal
0.2
0.0000
-0.2
-0.4
-0.6
42.5
45.0
47.5
50.0
52.5
55.0
Observed Value


55.0
52.5
50.0
47.5
45.0
42.5
After