Week 7
Descriptive Statistics
Frequency

Generally a frequency distribution is used for looking at detailed information for nominal and ordinal (categorical) data that describes the results. Categorical data is for variables such as gender, i.e., males are coded as "1" and females are coded as "2." Frequencies options include a table showing counts and percentages, statistics including percentile values, central tendency, dispersion and distribution, and charts including bar charts and histograms. The steps for using the frequencies procedure is to click the Analyze menu, choose Descriptive Statistics then from the sub menu, choose Frequencies and select your variables for analysis. You can then choose statistics options, choose chart options, choose format options, and have IBM SPSS calculate your request.

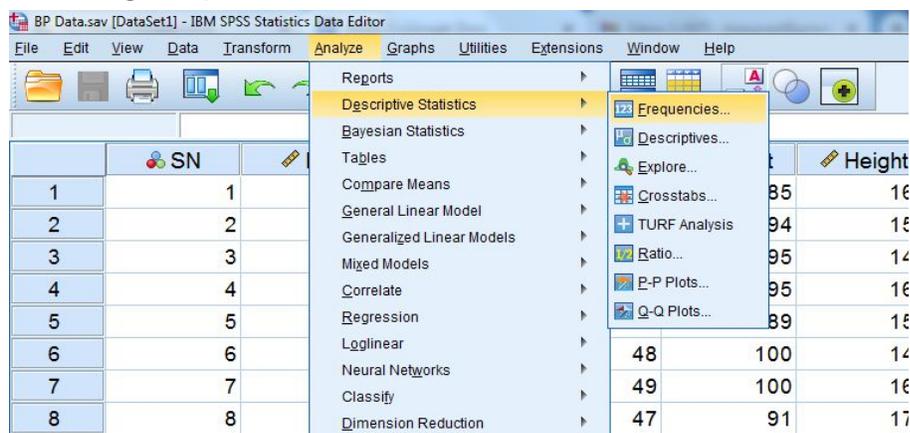**Choosing Frequencies Procedure:**

Fig: 7.1

From the Analyze menu, highlight Descriptive Statistics, as in above figure, then move your mouse across to the sub menu and click on Frequencies.

A Dialog box, as a screenshot below, will appear providing a scrollable list of the variables on the left, a Variable(s) choice box, and buttons for Statistics, Charts and Format options.
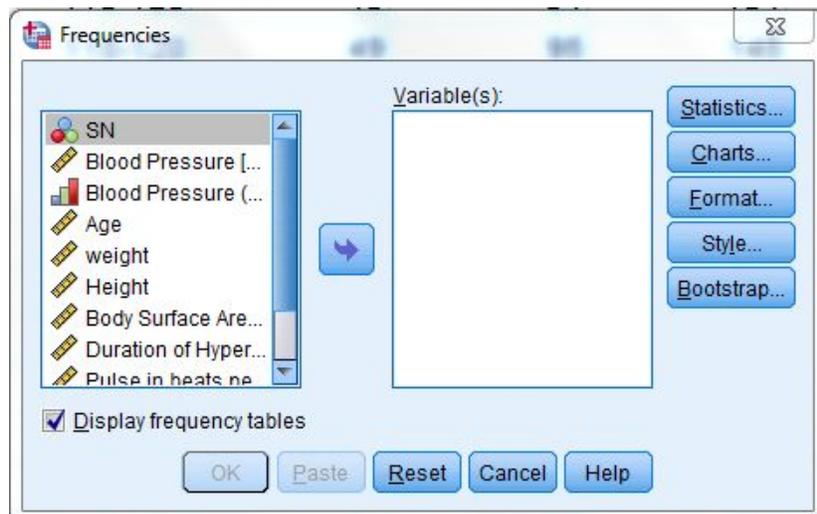

Fig 7.2

Selecting Variables for Analysis:

First select your variable from the main Frequencies Dialog box, Figure 7.2, by clicking the variable name on the left side. (Use the scroll bar if you do not see the variable you want.) Click the arrow on the right of the Variable List box, after selecting the desired variable to move the variable into the Variable(s) box. All variables selected for this box will be included in any procedures you decide to run. We could click OK to obtain a frequency and percentage distribution of the variables. In most cases we would continue and choose one or more statistics.

Choosing Statistics for Variables:

Click the Statistics button, right top of Figure 7.2, and a Dialog box of statistical choices will appear, as in Figure 7.3.
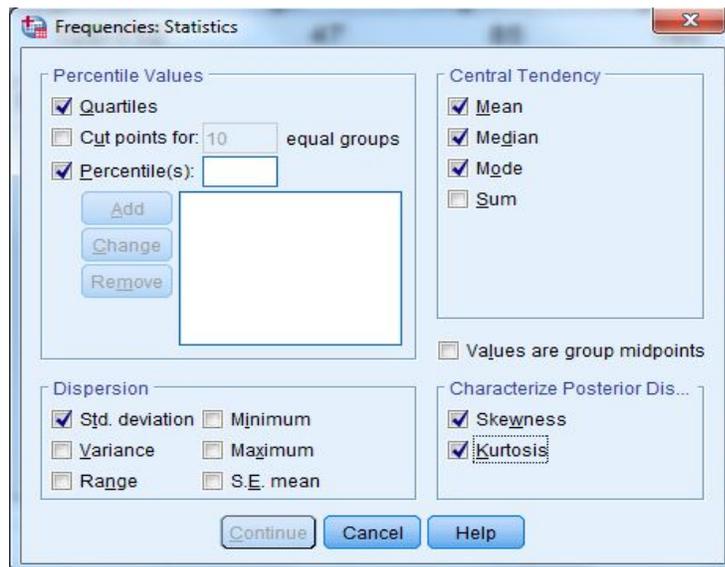

Fig: 7.3

For nominal (category) variables, we will select mode as a central tendency, for ordinal types of variable we will select median as a central tendency where as for remaining we use mean as a central tendency.

After clicking the appropriate central tendency, click the Continue button, bottom left, and we return to the main Frequencies dialog box, Figure 7.2. Similarly you can select the other statistical tools as your requirements.

We could now click OK and IBM SPSS would calculate and present the frequency and percent distribution (click OK if you want) but, in the more typical manner, we will continue and include choices for charts and check out the Options possibilities. If you clicked OK, just press the Analysis menu then choose Descriptive Statistics and then Frequencies from the sub menu and you will be back to this point with your variable and statistics chosen.

Choosing Charts for Variables:

On the main frequencies window, click the Charts button, Figure 7.2, and a Dialog box of chart choices, Figure 7.4, will appear.

Click Bar Chart, if the variable is categorical, then click Continue to return to the main Frequencies window box. If you have a continuous variable, choose Histograms and the With

Normal Curve option would be available. Choose the With Normal Curve option to have a normal curve drawn over the distribution so that you can visually see how close the distribution is to normal. Note: Frequencies is automatically chosen for chart values but if desired you could change that to Percentages, bottom Figure 7.4.
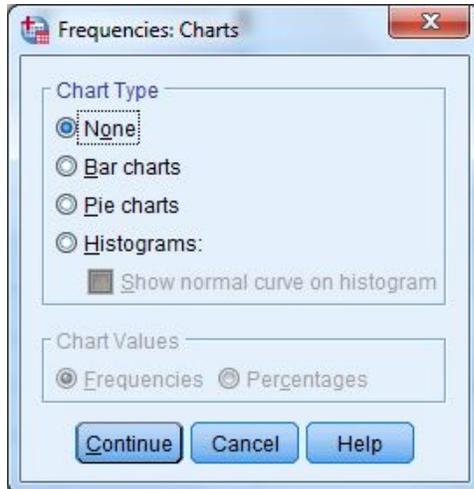

Fig 7.4

Now click OK on the main frequencies dialog box and IBM SPSS will calculate and present a frequency and percent distribution with our chosen format, statistics, and chart. (Note: We could look to see if additional choices should be made by clicking the Format button.)
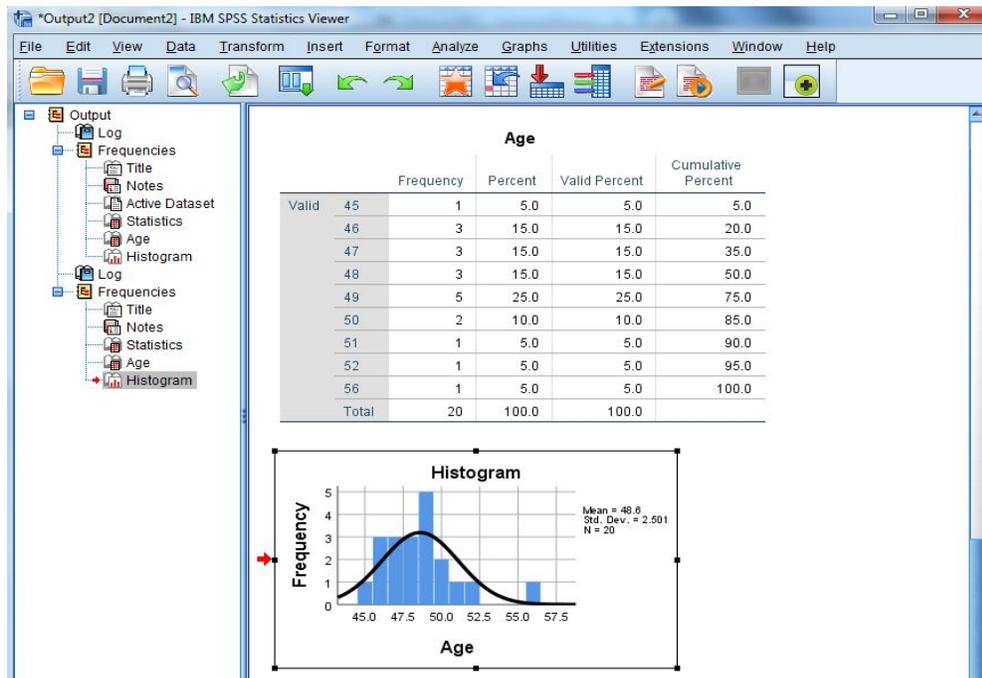

Fig 7.5

(Patience, processing time for IBM SPSS to perform the analysis in the steps above will depend on the size of the data set, the amount of work you are asking IBM SPSS to do and the CPU speed of your computer.) The output outline, left side, and the output, right side, will appear when IBM SPSS has completed its computations.

Descriptive Statistics
Descriptives
Descriptives (Analysis, Descriptive Statistics, Descriptives, Figure 7.6) is used to obtain summary information about the distribution, variability, and central tendency of continuous variables. Possibilities for Descriptives include mean, sum, standard deviation, variance, range, minimum, maximum, S.E. mean, kurtosis and skewness.

Choosing Descriptive Procedure:
First click the Analyze menu and select Descriptive Statistics, then move across to the sub menu and select Descriptives (see Figure 7.6). The Variable Choice dialog box will appear (see Figure 7.7).

Selecting Variables for Analysis:
First click on your variable of interest. Click the select arrow in the middle and IBM SPSS will place your variable in the Variable(s) box. Follow the similar steps to add different variables at the same time.
We could click OK and obtain maximum, minimum, mean and standard deviation (by default these statistical tools will be selected), but we will click the Options button and decide on statistics for our output.
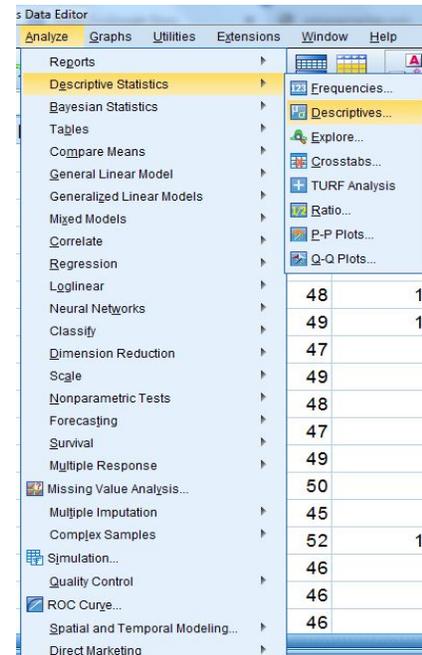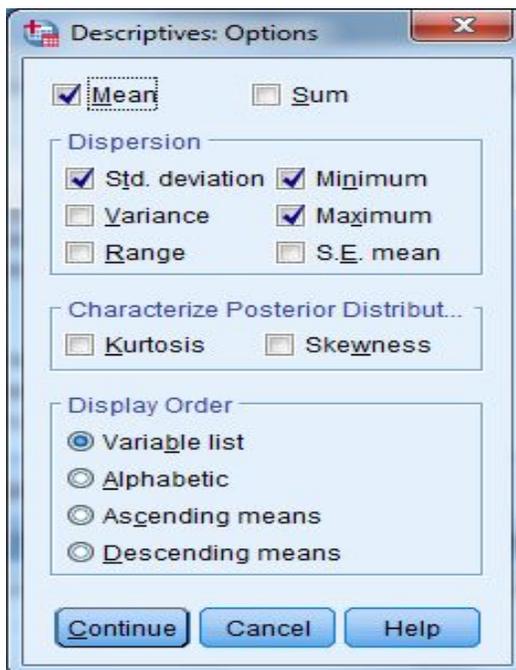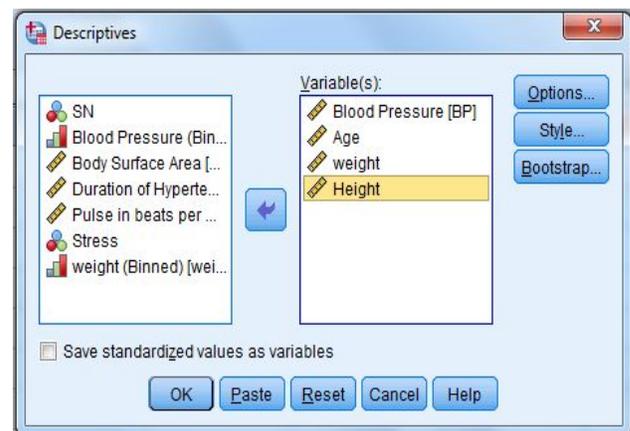

Fig: 7.6

Fig: 7.7

Fig: 7.8

The Options dialog box, Figure 7.8, will open. for variables with interval/ratio measures, choose: Mean, Std. deviation, Minimum and Maximum. We will leave the defaults for the Distribution and Display Order.
Next, click the Continue button to return to the main Descriptives dialog box, (Figure 7.7). Click OK in the Descriptives dialog box and IBM SPSS will calculate and display the output.

Explore

Explore is primarily used to visually examine the central tendency and distributional characteristics of continuous variables. Explore statistics include M-estimators, outliers, and percentiles. Grouped frequency tables and displays, as well as Stem-and-leaf and box-plots, are available. Explore will aid in checking assumptions with Normality plots and Spread vs. Level with the Levene test.

Choosing the Explore Procedure:

From the Analyze menu choose Descriptive Statistics, drag to the sub menu and select Explore.
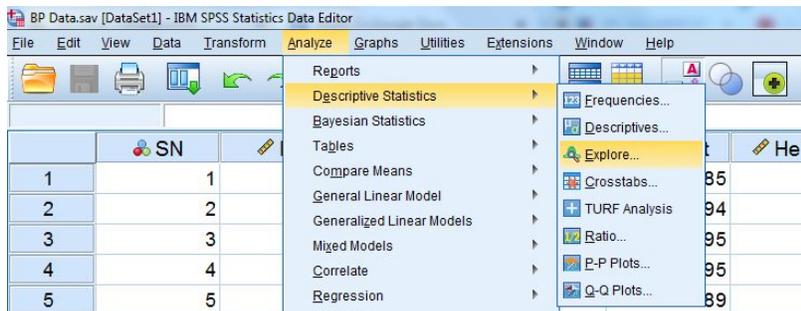

Fig: 7.9

Selecting Variables:

As in the other procedures, find and click the variable you want to explore, and then click the select arrow to include your variable in the Dependent List box. Choose the variable you want to explore and move into the Dependent List box. The dialog box should look like Figure 7.10.
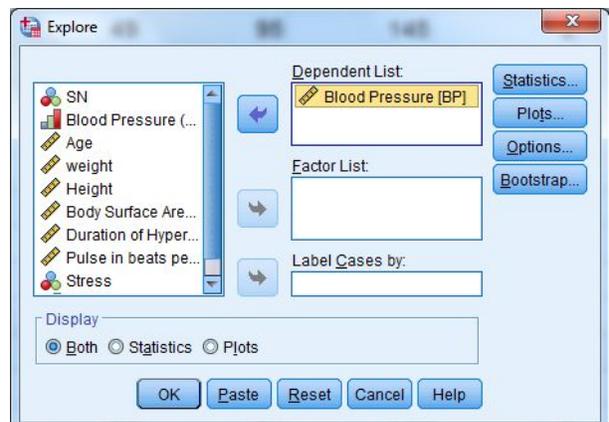
Selecting Displays:

In the Display box on the bottom left, you may choose either Both, Statistics, or Plots. We left the default selection, Both, to display statistics and plots.

Fig: 7.10

Selecting Statistics:

Click the Statistics button and the Explore: Statistics dialog box will open, Figure 7.11. Leave checked the Default box for Confidence Interval for the Mean 95%, and click the Outliers box so we can look at the extreme observations for our variable. Click Continue to return to the main explore dialog window.
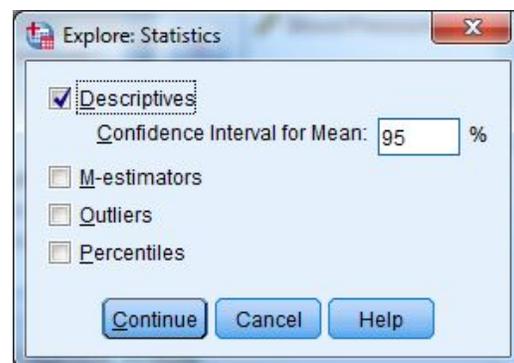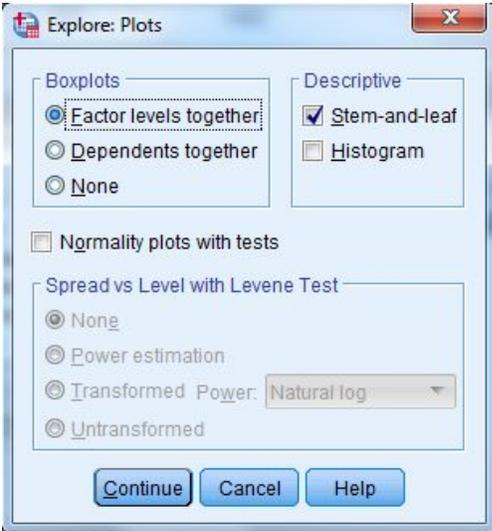

Fig: 7.11

Fig: 7.12

Selecting Plots:
Click the Plots button on the main Explore Dialog box, Figure 7.10, and the Explore: Plots dialog box, Figure 7.12, will open.
Click Stem-and-leaf and Histogram in the Descriptive box. Click on Normality Plots with Test so we can see how close the distribution of this variable is to normal. Click Continue to return to the main explore dialog box, Figure 7.10.

Fig: 7.13

Selecting Options:
Click the Options button in the main explore dialog box, Figure7.10, and the Explore: Options dialog box, Figure 7.13, will be displayed. No changes are needed here since the default of Exclude cases listwise is appropriate. Now click Continue to return to the main Explore dialog box, Figure 7.10. Click OK in the main Explore dialog box and IBM SPSS will perform the chosen tasks and display the data in the IBM SPSS Output.
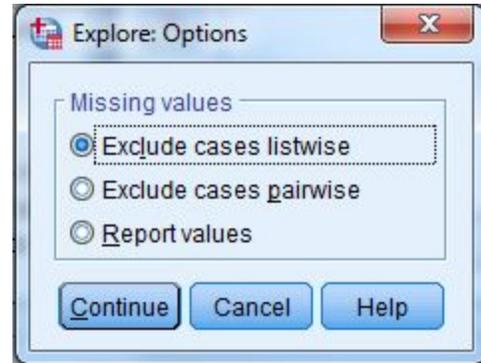
Interpretation of Explore Output:
Use the scroll bar to view any part of the output. The first part of the output is the Case Processing Summary, Figure 7.14. We can see that 20(100%) of our respondents answered this question with no missing.The total sample size was 20.

Explore

Case Processing Summary

| | Cases | | | | | |
| | Valid | | Missing | | Total | |
	N	Percent	N	Percent	N	Percent
weight	20	100.0%	0	0.0%	20	100.0%

Fig: 7.14

Descriptives

			Statistic	Std. Error
weight	Mean		93.09	.960
	95% Confidence Interval for Mean	Lower Bound	91.08	
		Upper Bound	95.10	
	5% Trimmed Mean		93.06	
	Median		94.15	
	Variance		18.446	
	Std. Deviation		4.295	
	Minimum		85	
	Maximum		101	
	Range		16	
	Interquartile Range		5	
	Skewness		.093	.512
	Kurtosis		-.355	.992

Fig: 7.15

The Descriptives statistics output should look like Figure 7.15. We can see all the typical descriptive statistics on this output: mean (93.09), lower bound (91.08) and upper bound (95.10) for a 95% confidence of the mean (in polling terminology this says that we are 95% confident that the mean for the population is between 91.08 and 95.10). Also shown, the median (94.15), variance (18.446),standard deviation (4.295), minimum (85), maximum (101), range (16), interquartile range (5), skewness (0.093), kurtosis (-0.355).

The Test of Normality is shown next (see Figure 7.16). This shows that this distribution follow the normal distribution. This is a pretty stringent test; most researchers would not require the distribution to be this close to

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
weight	.143	20	.200*	.963	20	.603

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

a. Lilliefors Significance Correction

Fig: 7.16

Fig: 7.17

```
weight Stem-and-Leaf Plot

    Frequency      Stem &   Leaf

        5.00        8 .   57799
       10.00        9 .   0012444444
        4.00        9 .   5599
        1.00       10 .   1

    Stem width:        10
    Each leaf:          1 case(s)
```
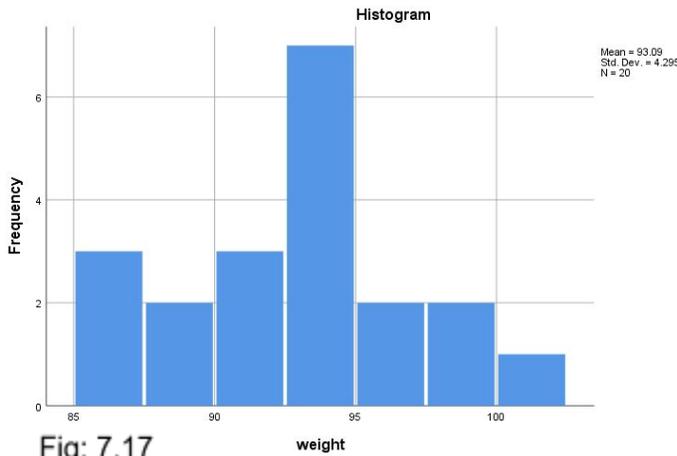
Fig: 7.18

normality.The histogram, Figure 7.17, shows a rough bell shaped distribution. IBM SPSS divided our distribution into seven groups.The statistics on the histogram tell us that the standard deviation is 4.295 with a mean of 93.09 for a total N of 20. The Stem-and-Leaf is next. Figure 7.18, again, shows a close but not quite normal distribution.

Interpretation of the Q-Q Plot :
Continue scrolling down the IBM SPSS Output Navigator to the Normal Q-Q Plot of weight (see Figure 7.19).
A Q-Q plot charts observed values against a known distribution, in this case a normal distribution. If our distribution is normal, the plot would have observations distributed closely around the straight line. In Figure 7.19, the expected normal distribution is the straight line and the line of little boxes is the observed values from our data. Our plot shows the distribution deviates somewhat from normality at the upper end.
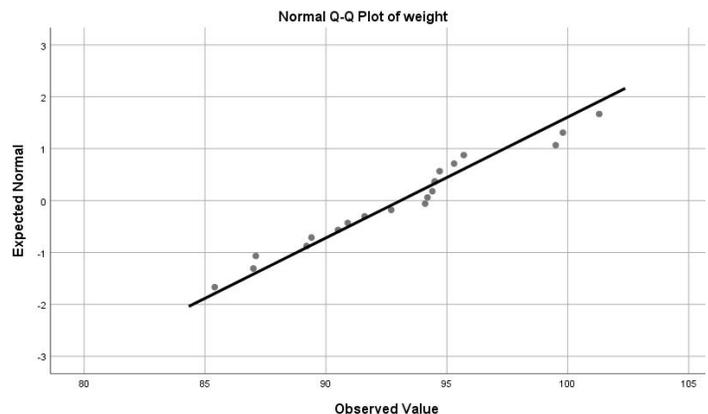

Fig: 7.19