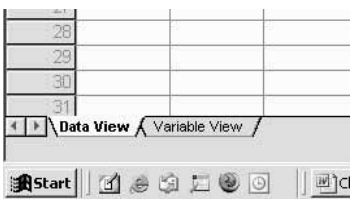
3.8.1. Are my data parametric?

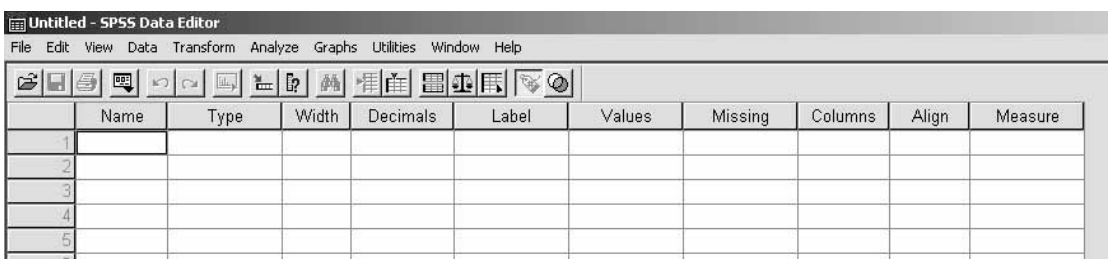
EXAMPLE 3.2. Length (mm) of two-spot ladybirds (*Adalia bipunctata*)

BOX 3.2. How to check if your data are normally distributed (parametric)

Step 1. Set up the variables.
When SPSS starts, select the 'Type in Data' option. Then choose 'Variable View' from the tabs at the bottom left of the screen.

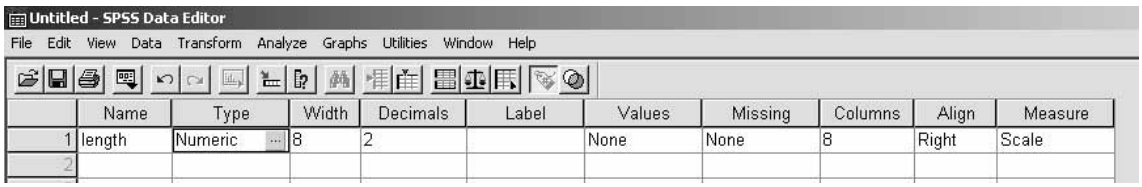


You will see a screen something like this:

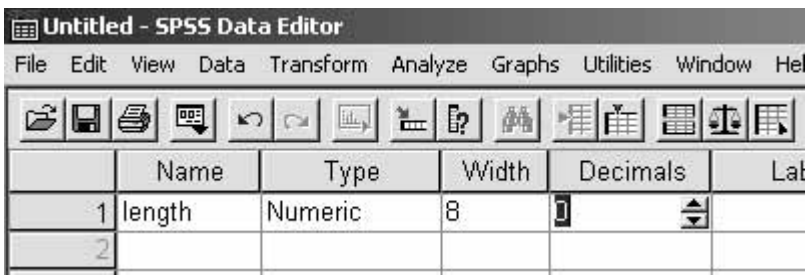


Each row represents a variable for the analysis.

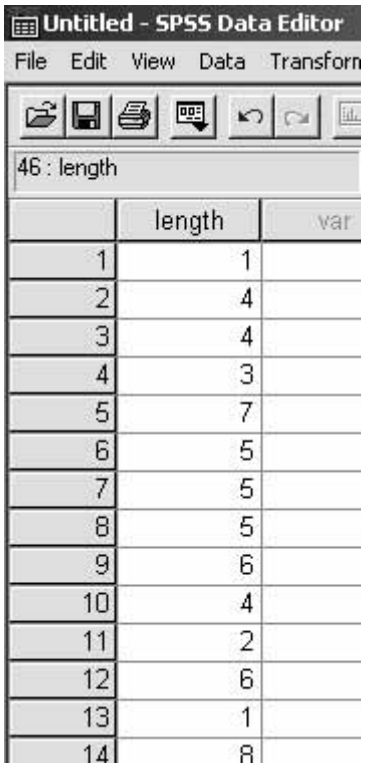
In the name for variable 1, type 'length' (SPSS won't accept capital letters as parts of a Name). Most of the other characteristics of the variable will be give default values as below:



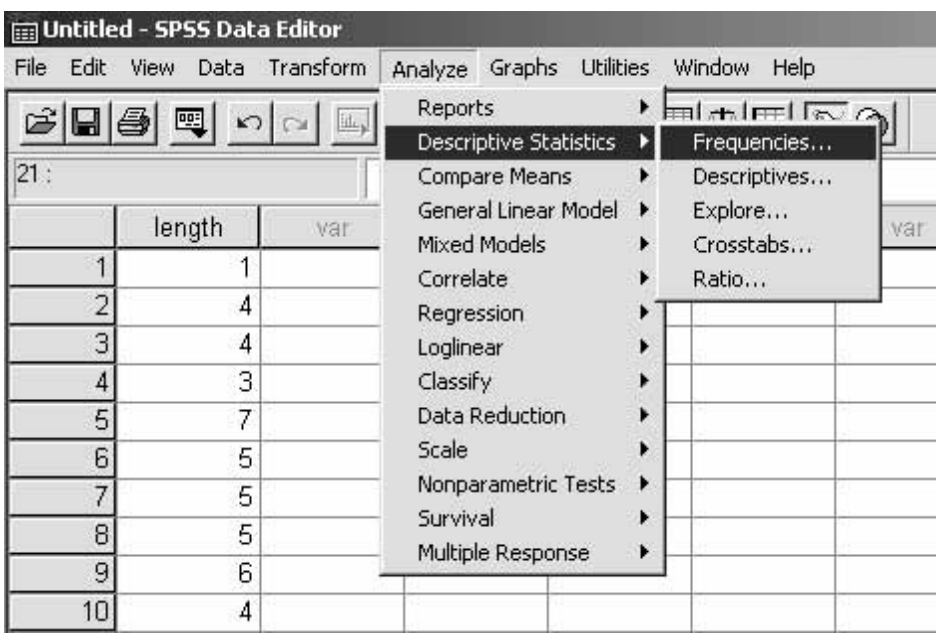
The lengths are to the nearest mm, so we change the 'Decimals' property to zero by clicking in the 'Decimals' window and using the 'up' and 'down' arrows that appear to make the adjustment.



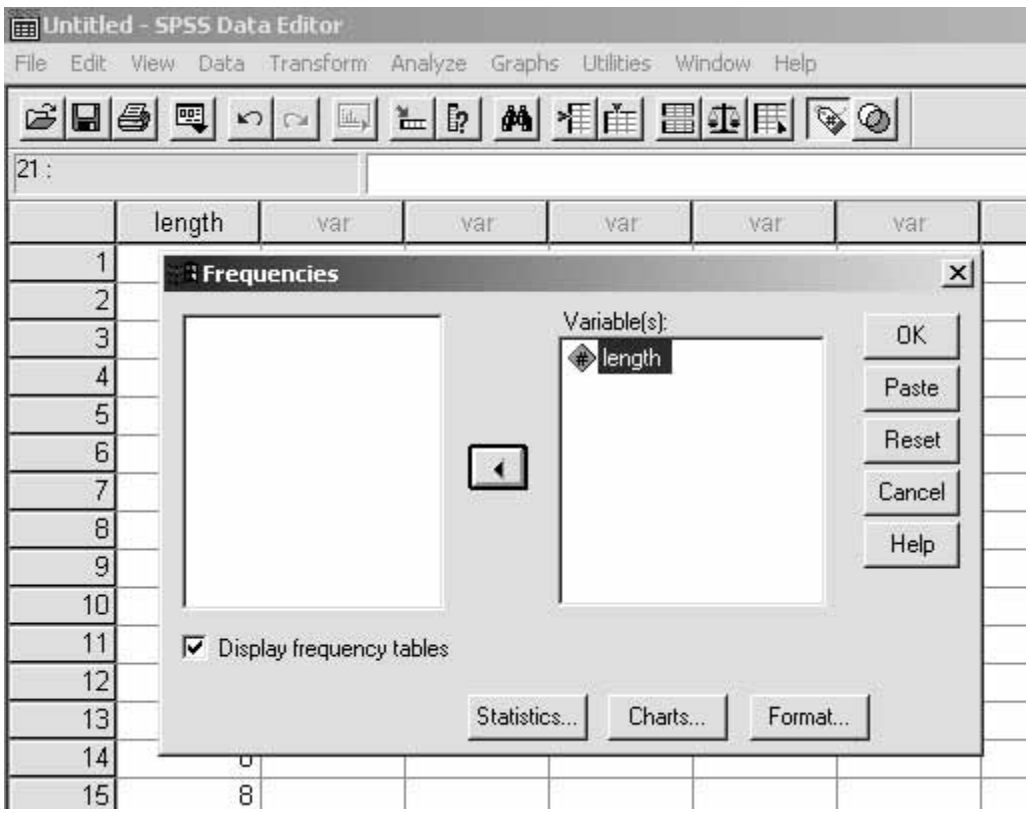
Step 2. Enter the data.
Transfer to data view using the tabs at the bottom left.
Enter the numerical data into column 1.



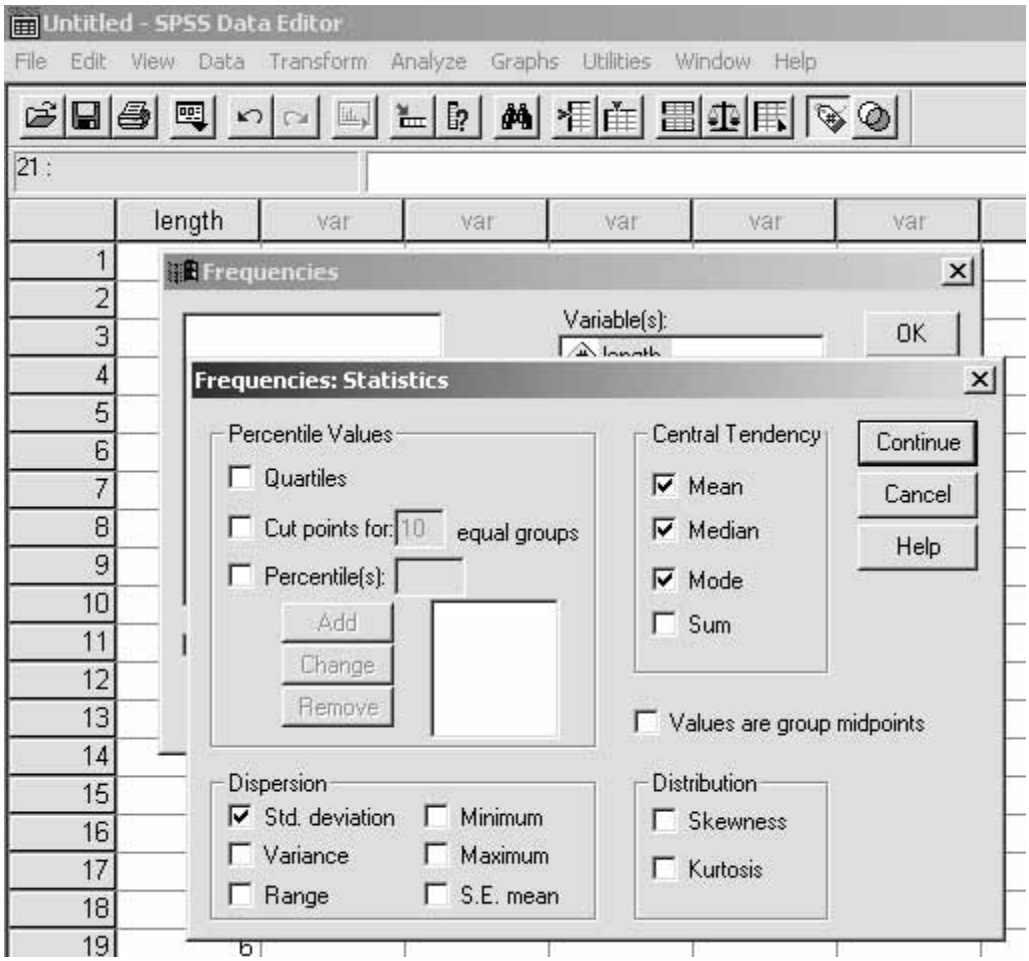
Step 3. Perform the necessary analysis.
Go to 'Analyze', 'Descriptive Statistics', 'Frequencies'.



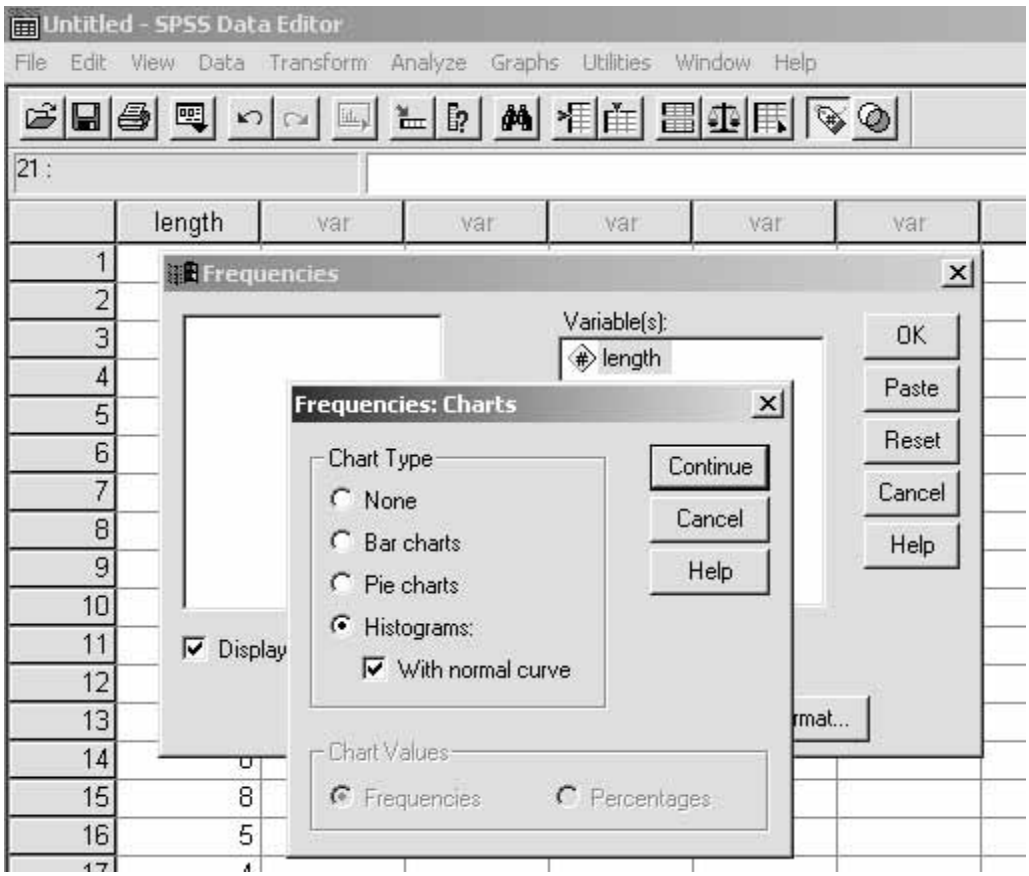
Highlight 'length' by clicking on it, then transfer it to the 'Variable' window by clicking on the arrow.



Click on 'Statistics', and select 'Std. deviation', 'Mean', 'Median' and 'Mode'.



Click on ‘Continue’.
Click on ‘Charts’, and select ‘Histograms’ and ‘With normal curve’.



Click on 'Continue', then click on 'OK'. The output will appear in a separate window.

Frequencies

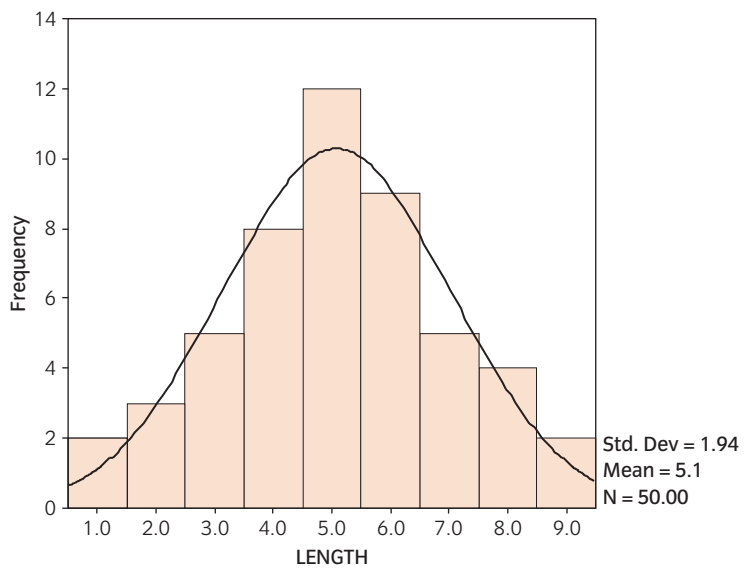
Statistics

LENGTH

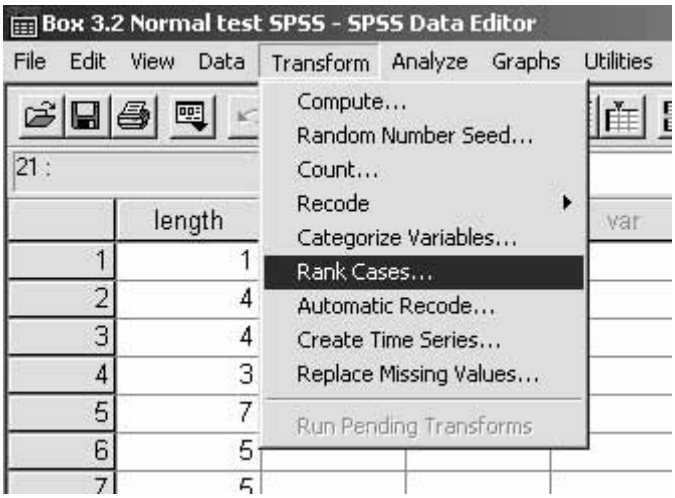
N	Valid	50
	Missing	0
Mean		5.08
Median		5.00
Mode		5
Std. Deviation		1.936

LENGTH

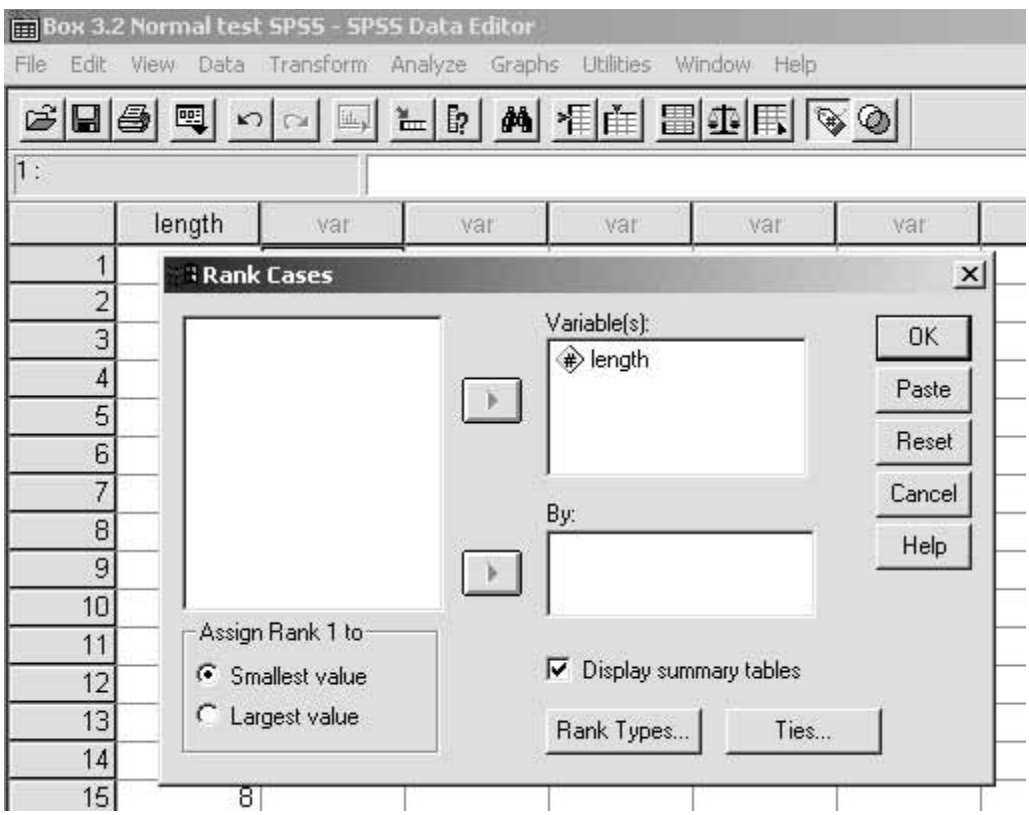
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	4.0	4.0	4.0
	2	3	6.0	6.0	10.0
	3	5	10.0	10.0	20.0
	4	8	16.0	16.0	36.0
	5	12	24.0	24.0	60.0
	6	9	18.0	18.0	78.0
	7	5	10.0	10.0	88.0
	8	4	8.0	8.0	96.0
	9	2	4.0	4.0	100.0
	Total	50	100.0	100.0	



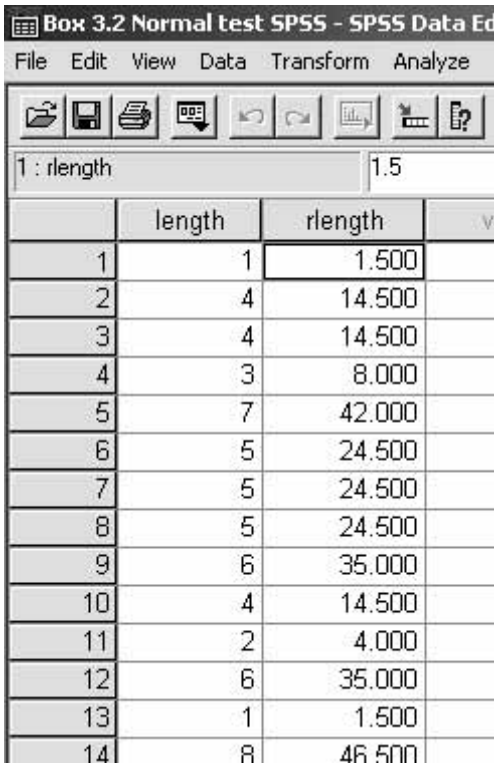
We also need the data in ranked order. Go to 'Transform', 'Rank Cases'.



Click on 'length' to highlight it, then click on the arrow to transfer it into the 'Variable(s)' window.



Click on 'OK'. A new variable will be generated with the ranks, and the summary table will appear in the 'Output' window.



	length	rlength
1	1	1.500
2	4	14.500
3	4	14.500
4	3	8.000
5	7	42.000
6	5	24.500
7	5	24.500
8	5	24.500
9	6	35.000
10	4	14.500
11	2	4.000
12	6	35.000
13	1	1.500
14	8	46.500

```
From          New
variable      variable      Label
---------     ---------     ------
LENGTH        RLENGTH       RANK of LENGTH
```

Step 4. Go through the individual tests.

(a) Are the data on an interval scale? Yes – they are measured in mm.
(b) Does the distribution appear to be a 'bell-shaped' curve? Looking at the graph, the answer is 'yes'.
(c) Do about 68% of your observations fall within the range $\overline{x} \pm 1\ s$?

The mean is 5.080 mm, and the standard deviation is 1.936 mm.

$\overline{x} \pm 1\ s = 5.080 + 1.936 = 7.016$

$\overline{x} - 1\ s = 5.080 - 1.936 = 3.144$

The measurements between these are 4, 5, 6, and 7 mm. From the ranking table in column 2, these correspond to ranks of 14.5, 24.5, 35.0, and 42.0. Looking at the table (or reading from the graph) there are 8, 12, 9, and 5 ladybirds in these categories, a total of 34 ladybirds. This is $(34/50) \times 100 = 68\%$ of the total. Thus about 68% of our observations lie in the range $\overline{x} \pm 1\ s$.

(d) Does median = mode = mean?

Median = 5.0 mm, mode = 5.0 mm, and mean = 5.080 mm. Thus they are all close together.

(e) For the goodness of fit chi-squared test, see Box 5.2.

From the information we can obtain at this stage, it looks as if the data are normally distributed.