Example 1: A data shows the scores of the exam which was taken by a group of students.

- **a)** Specify the type of data (scale)
- **b)** Build up frequency table for the data.
- **c)** Explain the highest frequency and the lowest percentile in the table.
- **d)** Draw a suitable plot of data.
- **e)** Find central measurements (mean, median, mode) using from raw data.
- **f)** Find sample mean using frequency table.
- **g)** Find and explain Q1, Q2, Q3 quartiles.
- **h)** Calculate dispersion measurements (variance, standard deviation)
- **i)** Calculate skewness and kurtosis measurements for the data

Table 1. Students' exam scores

14.5	46.6	59.5	70.5	75.5
18.5	48.4	62.4	70.5	77.5
20.6	50.5	63.4	71.0	83.5
25.3	51.5	65.4	71.5	84.0
28.8	54.8	65.5	71.6	84.4
40.6	54.8	66.5	71.8	87.4
42.5	55.0	69.0	72.0	88.5
43.0	56.8	69.9	75.0	92.0
43.5	57.8	70.0	75.3	98.4

$\sum_{i=1}^{n} x_i = 2765.6$ $\sum_{i=1}^{n} x_i^2 = 187277.6$ $\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^3 = -198789{,}9$ $\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^4 = 19294316{,}2$

ANSWERS OF EXAMPLE 1:

- **a)** Quantitative data-(Continuous data) the scale of it is "interval"

- **b)**

Table 2. Frequency table for the students' exam scores data **(answer of b)**

Group No /Class No	Lower Limit (LL))	Upper Limit (UL)	Class Value (s_i)	Frequency (f_i)	Relative Frequency ($p_i=f_i/n$)	$f_i \times s_i$	$f_i \times s_i^2$
1	14.5	28.4	(14.5+28.4)/2=21.45	4	4/45=0.09	85.80	1840.41
2	28.5	42.4	35.45	2	2/45=0.04	70.90	2513.405
3	42.5	56.4	49.45	10	0.22	494.50	24453.03
4	56.5	70.4	63.45	11	0.24	697.95	44284.93
5	70.5	84.4	77.45	14	0.31	1.084.30	83979.04
6	84.5	98.4	91.45	4	0.09	365.80	33452.41
			TOTAL	**n=45**	**1**	**2799.25**	**190523.2**

- **c)** The highest frequency is 14 that means in this exam the most of the students (or 14 of the students) got the scores between 70.5 and 84.4. The lowest percentile is 0.04 that means in this exam few of the students (or 4% of the students) got the scores between 28.5 and 42.4.

d)

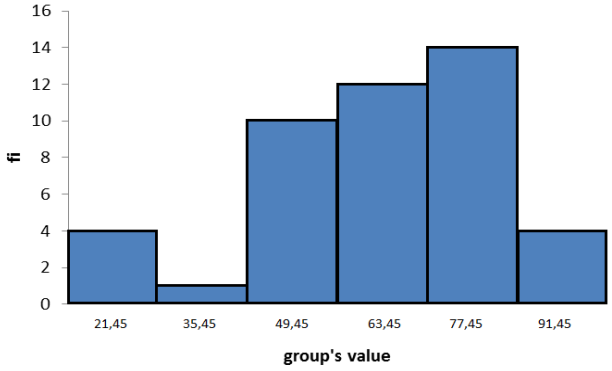


Figure 1. Histogram chart for the data given in Table 2. **(answer of d)**

e) Find central measurements (mean. median. mode) using from raw data.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2765.6}{45} = 61.46$$

$\bar{x}' = x_{23} = 65.5$ since for n=45. $i = \frac{n+1}{2} = \frac{45+1}{2} = 23$ if n is odd. That means the half of the students' exam scores are smaller/lower than 65.5 and the other half of the students' exam scores are larger/higher than 65.5. In other words. 50% of the students' exam scores are smaller/lower than 65.5 and 50% of the students' exam scores are larger/higher than 65.5. In other words 50%percent of students got score less than 65.5 and %50 percent of students got score greater than 65.5.

$\hat{x} =$ 54.8 and 70.5 (two times repeated), that is there are two modes.

f) Find sample mean using frequency table:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i s_i}{n} = \frac{2799.25}{45} = 62.2056$$

g) Find and explain Q1. Q2. Q3 quartiles.

$Q1 = \frac{x_{11} + x_{12}}{2} = \frac{48.4 + 50.5}{2} = 49.45$ for n=45. $i = \frac{n+1}{4} = \frac{45+1}{4} = 11.5$ if n is odd.

Q1 – First Quartile – 25% of the students' exam scores are smaller/lower than 49.45 and 75% of the students' exam scores are larger/higher than 49.45.

$Q2 = \bar{x}' = x_{23} = 65.5$ for n=45. $i = \frac{n+1}{2} = \frac{45+1}{2} = 23$ if n is odd.

Q2 – Second Quartile- 50% of the students' exam scores are smaller/lower than 65.5 and 50% of the students' exam scores are larger/higher than 65.5. Same as the median.

$Q3 = \frac{x_{34} + x_{35}}{2} = \frac{72 + 75}{2} = 73.5$ for n=45. $i = \frac{3 \times (n+1)}{4} = \frac{3 \times (45+1)}{4} = 34.5$ if n is odd.

Q3 – Third Quartile-75% of the students' exam scores are smaller/lower than 73.5 and 25% of the students' exam scores are larger/higher than 73.5.

h) Calculate dispersion measurements (variance. standard deviation).

For raw data:

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{187277.6 - 45\times(61.46)^2}{44} = 393.13$$

$$s = \sqrt{s^2} = \sqrt{393.13} = 19.82750$$

Another formula for s^2 is: $s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$, it could also be used and a close value will be obtained.

For grouped data (from frequency table):

$$s^2 = \frac{\sum_{i=1}^{k} f_i s_i^2 - \frac{\left(\sum_{i=1}^{k} f_i s_i\right)^2}{n}}{n-1} = \frac{190523.2 - \frac{(2799.25)^2}{45}}{44} = 372.598$$

$$s = \sqrt{s^2} = \sqrt{372.598} = 19.3028$$

i) Calculate skewness and kurtosis measurements for the data.

$$skewness = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{nS^3} = \frac{-198789.9}{45\times\left(19.82750\right)^3} \cong -0.57$$ since skewness<0 the distribution of the data is **skewed left.**

$$kurtosis = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4}{nS^4} - 3 = \frac{19294316.2}{45\times\left(19.82750\right)^4} \cong 2.8 - 3 = -0.2$$ since kurtosis<0 the distribution of the data is platykurtic (basık) according to a standard normal distribution.

Kurtosis Value Range

- Normal distribution kurtosis = 0
- A distribution that is more peaked and has fatter tails than normal distribution has kurtosis value greater than 0 (the higher kurtosis, the more peaked and fatter tails). Such distribution is called ***leptokurtic*** or ***leptokurtotic***.
- A distribution that is less peaked and has thinner tails than normal distribution has kurtosis value less than 0. Such distribution is called ***platykurtic*** or ***platykurtotic***.

Example 2: A data set shows the books sales (daily) of a publishing house during a year. The book sales are given for randomly selected 22 days in a year.

Table3. Books sales data.

19	39	58	75	135	195	196	200	235	254	255
286	312	314	356	370	371	373	373	430	433	490

$\sum_{i=1}^{n} x_i = 5769$ $\sum_{i=1}^{n} x_i^2 = 1900063$

a) Specify the type of data (scale)
b) Build up frequency table for the data.
c) Explain the highest frequency and the lowest percentile in the table.
d) Draw a suitable plot of the data.
e) Find central measurements (mean. median. mode) using from raw data.
f) Find sample mean using the frequency table.
g) Find and explain Q1. Q2. Q3 quartiles.
h) Calculate dispersion measurements (variance. standard deviation)

ANSWERS OF EXAMPLE 2:

a) Quantitative data-(Discrete data) the scale of it is "ratio".
b)

Table 4. Frequency table for the books sales data **(answer of b).**

Group No /Class No	Upper Limit (UL)	Lower Limit (LL)	Class Value (s_i)	Frequency (f_i)	Relative Frequency ($p_i=f_i/n$)	$f_i \times s_i$	$f_i \times s_i^2$
1	19	77	(19+77)/2=96/2=48	4	0.18	192	9216
2	78	136	(78+136)/2=214/2=107	1	0.05	107	11449
3	137	195	166	1	0.05	166	27556
4	196	254	225	4	0.18	900	202500
5	255	313	284	3	0.14	852	241968
6	314	372	343	4	0.18	1372	470596
7	373	431	402	3	0.14	1206	484812
8	432	490	461	2	0.09	922	425042
			Total	**n=22**	**1**	**5717**	**1873139**

c) The highest frequency is 4 but it has seen three times. That means 4 of the days randomly selected in a year number of books sales are between 19 and 77. Another 4 of the days randomly selected in a year the number of books sales are between 196 and 254. Another 4 of the days randomly selected in a year the number of books sales are between 314 and 372. The lowest percentile is 0.05 but it has seen two times. That means 5% of the days for randomly selected 22 days in a year the number of books sales are between 78 and 136. Another 5% of the days for randomly selected 22 days in a year the number of books sales are between 137 and 195.

d)

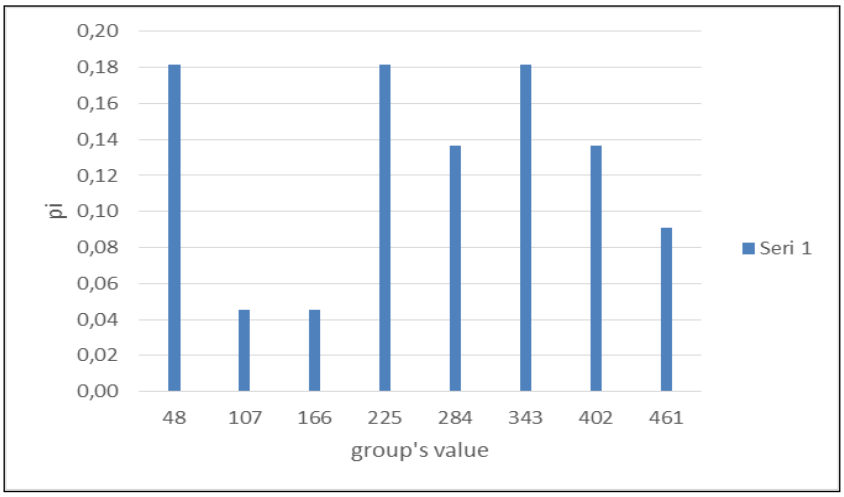


Figure 2. Line chart for the data given in Table 4 **(answer of d).**

e) Find central measurements (mean. median. mode) using from raw data.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5769}{22} = 262.2272$$

$\bar{x}' = \frac{x_{11} + x_{12}}{2} = \frac{255+286}{2} = 270.5 \cong 271$ for n=22. $i = \frac{n}{2} = \frac{22}{2} = 11$ if n is even. That means 11 of the days randomly selected in a year. the number of books sales are lower than 270.5 and the other 11 of the days randomly selected in a year. number of books sales are higher than 270.5. In other words. 50% of the days number of books sales are lower than 270.5 and 50% of the days number of books sales are higher than 270.5.

$\hat{x} =$ 373 (observed 2 times)

f) Find sample mean using frequency table.

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i s_i}{n} = \frac{5717}{22} = 259.8636$$

g) Find and explain Q1. Q2. Q3 quartiles.

$Q1 = \frac{x_5 + x_6}{2} = \frac{135+195}{2} = 165$ for n=22. $i = \frac{n}{4} = \frac{22}{4} = 5.5$ if n is even.

Q1 – First Quartile – 25% of the days number of books sales are lower than 165 and 75% of the days number of books sales are higher than 165.

$Q2 = \bar{x}' = \frac{x_{11} + x_{12}}{2} = \frac{255+286}{2} = 270.5$ for n=22. $i = \frac{n}{2} = \frac{22}{2} = 11$ if n is even

Q2 –Second Quartile – 50% of the days number of books sales are lower than 270.5 and 50% of the days number of books sales are higher than 270.5.

$Q3 = \frac{x_{16} + x_{17}}{2} = \frac{370+371}{2} = 370.5$ for n=22. $i = \frac{3 \times n}{4} = \frac{3 \times 22}{4} = 16.5$ if n is even.

Q3 – Third Quartile - 75% of the days number of books sales are lower than 370.5 and 25% of the days number of books sales are higher than 370.5.

h) Calculate dispersion measurements (variance. standard deviation).

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{1900063 - 22 \times (262.2273)^2}{21} = 18441.6155$$

$$s = \sqrt{s^2} = \sqrt{18411.6155} = 135.79989$$

$$s^2 = \frac{\sum_{i=1}^{k} f_i s_i^2 - \frac{\left(\sum_{i=1}^{k} f_i s_i\right)^2}{n}}{n-1} = \frac{1873139 - \frac{(5717)^2}{22}}{21} = 18452.31385 \qquad s = \sqrt{s^2} = \sqrt{18452.31385} = 135.8393$$

Example 3: The distribution of computers in lab 1 is given in below. Create cumulative frequency and cumulative relative frequency columns and draw pie chart for data.

Table 5. Frequency table for the computers in lab 1.

speed	f_i	p_i	Cumulative Frequency (F_i)	Cumulative Relative Frequency (P_i)	Angles
low	5	0.25	5	0.25	360×0.25=90º
medium	6	0.30	11	0.55	360×0.30=108 º
high	9	0.45	20	1.00	360×0.45=162 º
Total	**20**	**1.00**			**360**

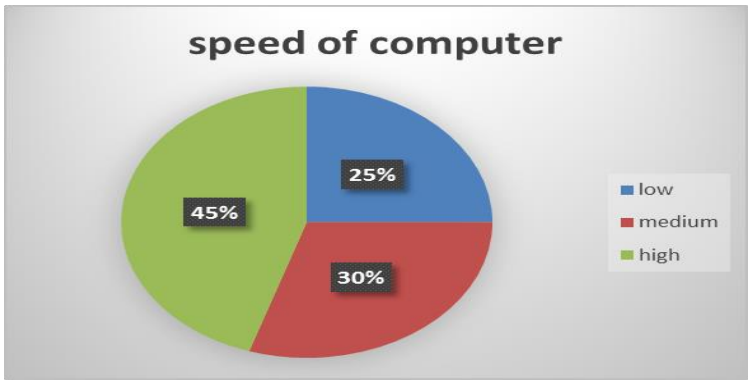


Figure 3. Pie graph for the data given in Table 5.

PERSONAL STUDY QUESTIONS

Example 4: A data set shows the amount of boron reserves (gr/1000) in computer processors (bilgisayar işlemcileri) produced by a producer.

a) Specify the type of data (scale)
b) Build up frequency table for the data.
c) Explain the highest frequency and the lowest percentile in the table.
d) Draw a suitable plot of data.
e) Find central measurements (mean. median. mode) using from raw data.
f) Find sample mean using frequency table.
g) Find and explain Q1. Q2. Q3 quartiles.
h) Calculate dispersion measurements (variance. standard deviation)
i) Calculate skewness and kurtosis measurements for the data

Table 6. Boron reserves data.

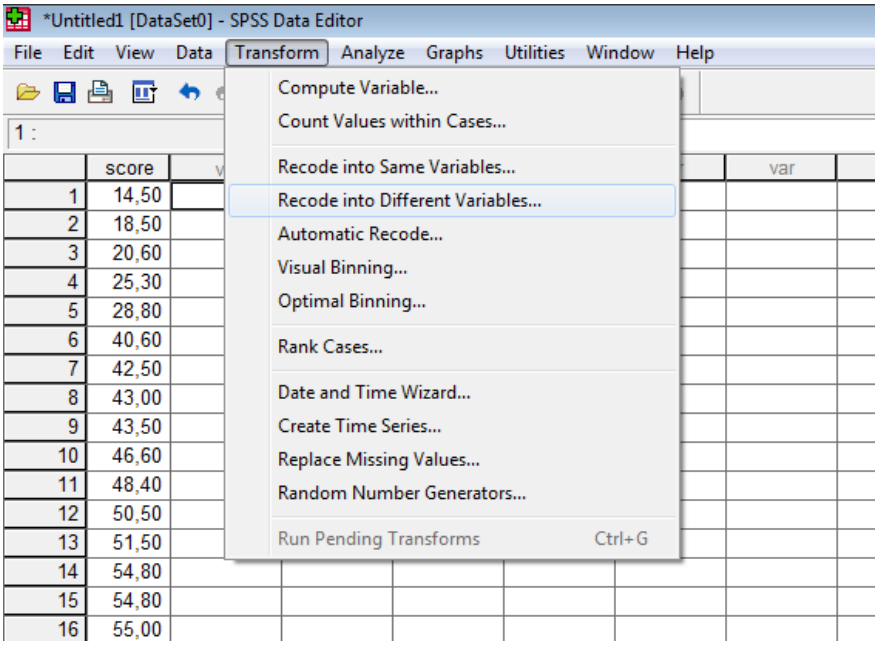
83.4	88.5	90.4	92.6
84.3	88.6	90.6	92.7
87.5	89.0	90.7	93.0
87.8	89.2	90.9	93.7
87.9	89.6	91.0	94.4
88.2	89.7	91.2	94.7
88.3	89.9	91.6	96.5
88.3	90.1	91.8	98.8
88.4	90.4	92.2	

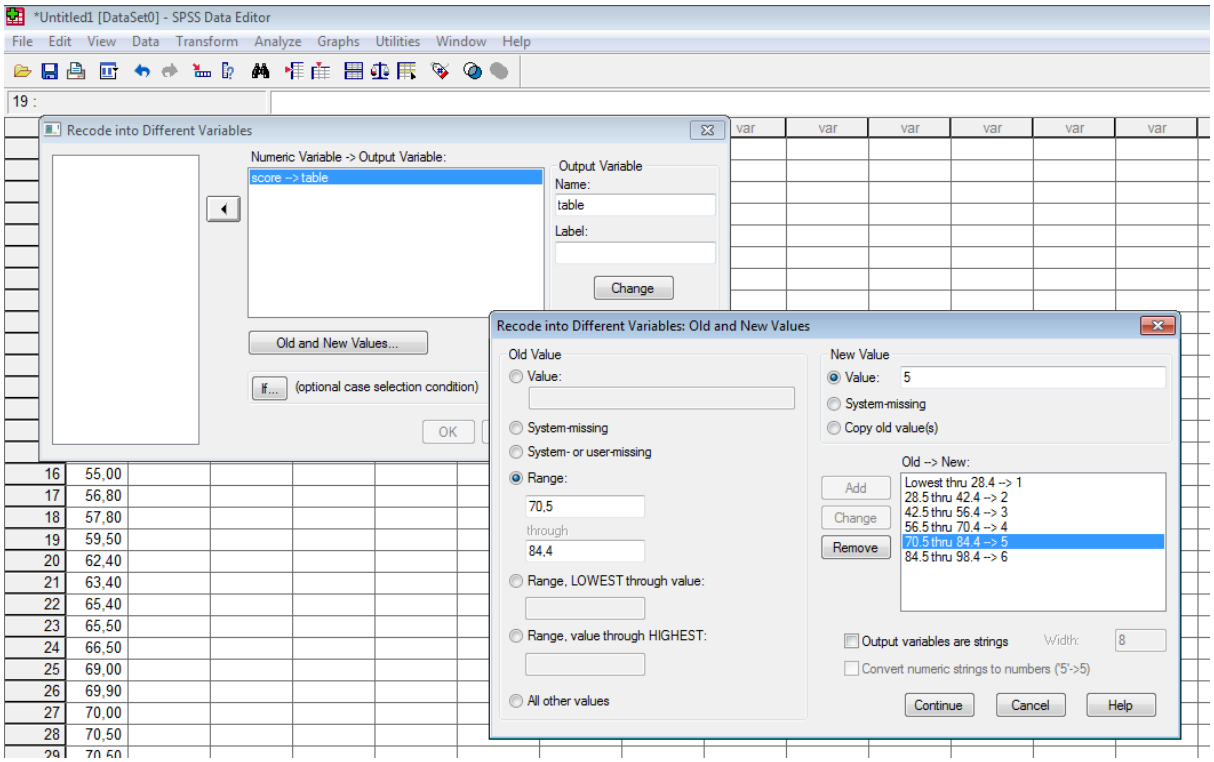
Example 5: A research is done for describing the computer users (for which purpose/purposes they use computers). 263 computer users attended to this research and 112 of them stated that they use computer for playing games. 57 of them using computer for utilizing internet. 82 of them using computer for their work. 12 of them using computer for mixed purpose (playing games. work. internet etc.). Build up frequency table for the data. Create cumulative frequency and cumulative relative frequency columns and draw pie chart for data.

Table 7. Frequency table for describing computer users.

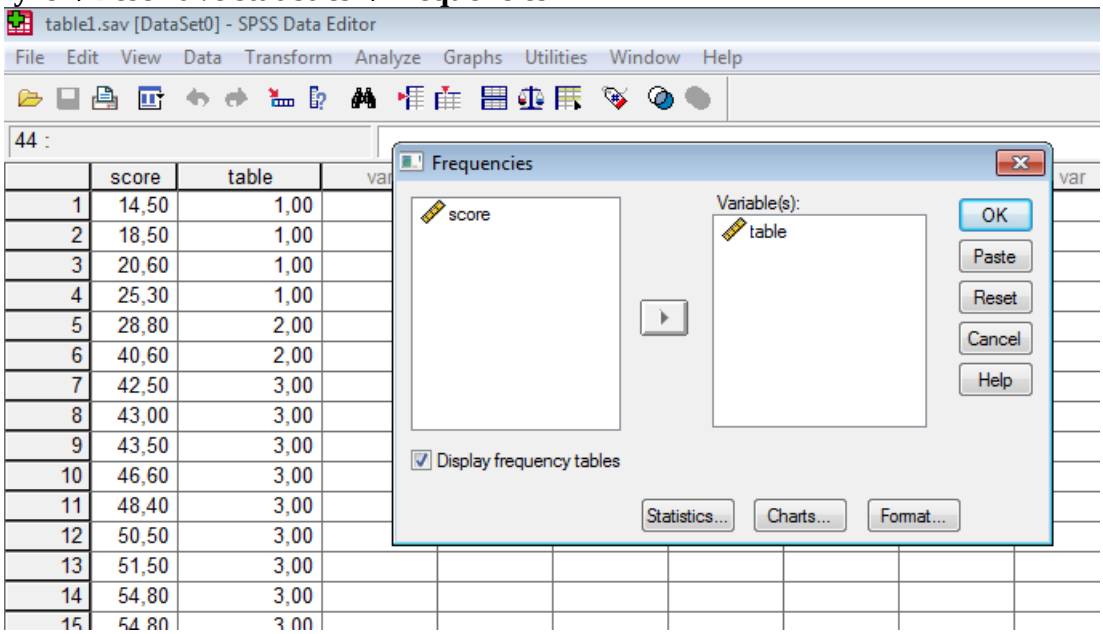
Purpose of Computer Usage	f_i	p_i	**Cumulative Frequency (F_i)**	**Cumulative Relative Frequency (P_i)**	**Angles**
Game					
Internet					
Work					
Mixed Purpose					
Total					

APPLICATION WITH SPSS FOR EXAMPLE 1 STUDENTS' EXAM SCORES DATA SET





Analyze→Descritive statistics→Frequencies



Statistics

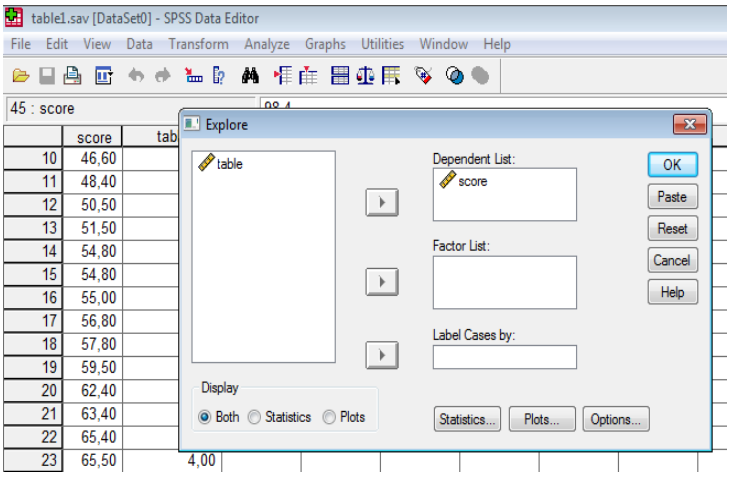
table

N	Valid	45
	Missing	0

table

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	4	8.9	8.9	8.9
	2.00	2	4.4	4.4	13.3
	3.00	10	22.2	22.2	35.6
	4.00	11	24.4	24.4	60.0
	5.00	14	31.1	31.1	91.1
	6.00	4	8.9	8.9	100.0
	Total	45	100.0	100.0	

Analyze→Descriptive statistics→Explore…



Descriptives

			Statistic	Std. Error
score	Mean		61,4556	2,95781
	95% Confidence Interval for Mean	Lower Bound	55,4945	
		Upper Bound	67,4166	
	5% Trimmed Mean		62,0944	
	Median		65,5000	
	Variance		393,688	
	Std. Deviation		19,84156	
	Minimum		14,50	
	Maximum		98,40	
	Range		83,90	
	Interquartile Range		24,05	
	Skewness		-,605	,354
	Kurtosis		,027	,695

Note: SPSS is using modified the kurtosis formula, so $kurtosis = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4}{nS^4} - 3$ **gives a different result from SPSS output.**

score Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1,00	Extremes	(=<15)
1,00	1 .	8
3,00	2 .	058
,00	3 .	
6,00	4 .	023368
8,00	5 .	01445679
7,00	6 .	2355699
12,00	7 .	000111125557
5,00	8 .	34478
2,00	9 .	28

Stem width: 10,00
Each leaf: 1 case(s)