STATISTICAL MEASURES

Measures of Central Tendency

- The mode of a data set is the value that occurs most frequently.There can be no mode, one mode, or several modes.
- The median of a data set is the value that lies in the middle when the data are arranged in size. When there are two middle values then the median is the midpoint between the two values
- The mean of a data set is the sum of all the values divided by the number

SL 4.2	Interquartile range	$\text{IQR} = Q_3 - Q_1$
SL 4.3	Mean, $\overline{x}$, of a set of data	$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$, where $n = \sum_{i=1}^{k} f_i$
	Sample statistics Unbiased estimate of population variance s_{n-1}^2	$s_{n-1}^2 = \frac{n}{n-1} s_n^2$

Example 1:
The number of days of sunshine in Helsinki in January was recorded for a period of 35 years and the data is given in the frequency table:

Number of days of sunshine x	Number of years f	fx	cf
3	1	3	1
4	2	8	3
5	1	5	4
6	2	12	6
7	7	49	13
8	5	40	18
9	9	81	27
10	8	80	35

	n = 35	278	

a. State the modal number of days of sunshine in January in Helsinki for the period.
9

b. Calculate the mean number of days of sunshine in January in Helsinki for the period.
278/35 =7.94

c. Determine the median number of days of sunshine in January in Helsinki for the period.
8

d. Comment on how a "day of sunshine" might be defined.
-

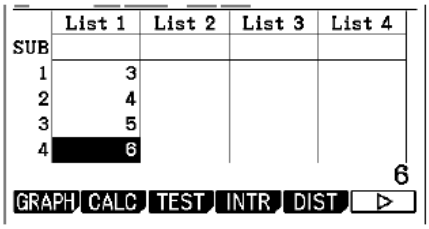
Calculator Support (Casio)

Press MENU 2 STAT to display the List Editor screen.

Type 3, 4, 5, … 10 in the first column.

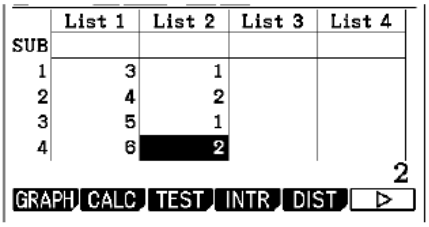
Press EXE after each number to move to the next cell.

Note: *If the list contains other numbers, you can clear it by pressing* F4 DEL-ALL.



Press ▶ to move to the next column.

Enter the frequencies of each of the ages in the second column.



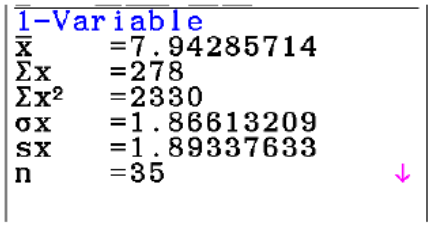
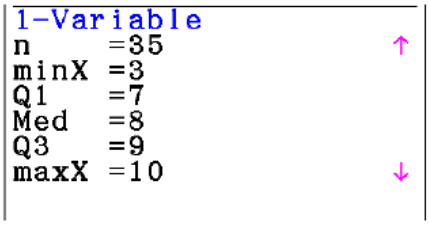
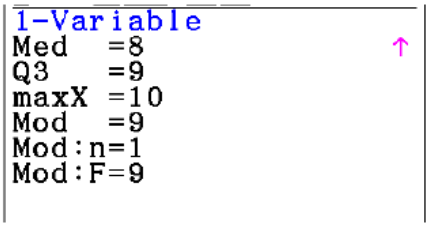
To calculate an estimate of the mean of the ages represented in the table.

Press F2 CALC.

Press F6 SET.

Change 1Var Freq to List2 by pressing F2 LIST and typing 2.

Press EXIT.

Press F1 1-VAR. The results show that the estimate of the mean ($\bar{x}$) is 7.94. So, the mean number of days of sunshine is 7.94 days.	
Scroll down using ▼. The table of statistics shows that the median is 8. So, the median number of days of sunshine is 8 days.	
Further scrolling reveals yet more statistics. The table of statistics shows that there is one mode which is 9 and the frequency is 9 as well.	

Measures of Dispersion

- Measures of dispersion measure how spread out a data set is.
- The most common measure of dispersion is the range, which is found by subtracting the smallest number from the largest number
- The standard deviation, σ_n , gives an idea of how the data values are related to the mean.The standard deviation is also known as the root mean-squared deviation.
- The variance is the standard deviation squared: σ_n^2
- An outlier is defined as a data item that is more than 1.5 × IQR below Q_1 or above Q_3
- Outliers are extreme data values, or the result of errors in reading data, that can distort the results of statistical processes

Example 2:
The number of days of precipitation in January in London for 2008–2017 is given in the table:

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Days of Precipitation	19	16	21	21	13	21	30	26	21	15

a. Write down the range of the number of days of precipitation in January in London for these years.
17

b. Calculate the interquartile range of the number of days of precipitation in January in London for these years.
5

c. Find the standard deviation of the number of days of precipitation in January in London for these years.
4.80

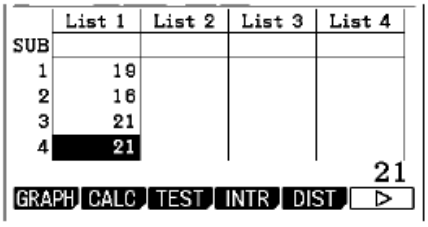
d. Find whether the 30 cm precipitation in January 2014 is an outlier.
Yes 30>28.5

Calculator Support (Casio)

Press MENU 2 to display the List Editor screen.

Type the numbers 19, 16, 21, ... 15 in the first column. Press EXE after each number to move to the next cell.

Note: *If the list contains other numbers, you can clear it by pressing* F4 DEL-ALL.



To calculate statistics for this dataset

Press F2 CALC.

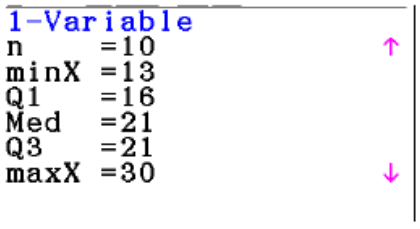
Press F6 SET.

1Var XList should be List1 and 1Var Freq should be 1.

Press EXIT •

Press F1 1-VAR.

The table of statistics shows that the minimum is 13 and the maximum is 30.



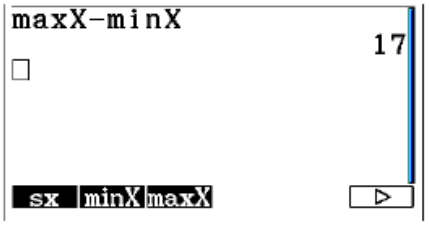
Press MENU 1 RUN·MAT to display the Run-Matrix screen for arithmetical calculations.

The statistics that you calculated earlier are all stored as variables.

To calculate the range Use Range = maxX – minX.

Press VARS F3 STAT F1 X F6 ▷ F3 maxX.

− F2 minx and press EXE.

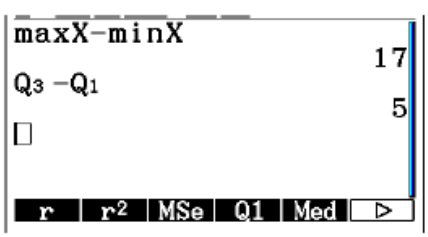


To calculate the interquartile range Use IQR = Q3 – Q1.

Press EXIT F3 GRAPH F6 ▷ F6 ▷ F1 Q3.

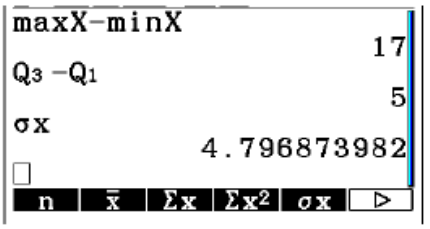
− F6 ▷ F6 ▷ F4 Q1 and press EXE.

The inter quartile range is 5.



To retrieve the standard deviation press EXIT F1 X F5 σx and press EXE.

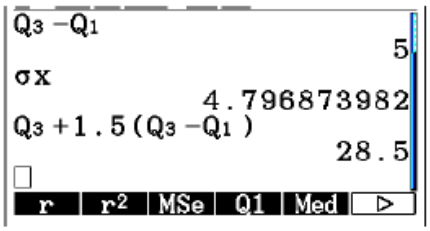
The standard deviation is 4.80 days.



To determine whether 30 is an outlier use Q3 + 1.5(IQR)

Select Q_3 and Q_1 from the list to enter the calculation Q3 + 1.5 (Q3 − Q1).

30 < 28.5, so 30 is an outlier.



Statistical measures of continuous data

Example 3:
Heights of 200 r trees are measured and the results recorded:

Height,h, (m)	$0 < h \leq 1$	$1 < h \leq 2$	$2 < h \leq 3$	$3 < h \leq 4$	$4 < h \leq 5$	$5 < h \leq 6$
x	0.5	1.5	2.5	3.5	4.5	5.5
Frequency	17	35	69	51	22	6

fx						

a. Find estimates for the mean and standard deviation of the height of these r trees.
   Mean = 2.72
   Standard deviation = 1.20
b. State why your calculations are estimates.

Unbiased Estimators

Example 4.

The number of breakdowns on a road in a city is studied. A sample of 20 observations is recorded on 20 different days. Let X be the number of breakdowns on a particular day. It is given that $s_n^2 = 3.8$.

(a) Find s_{n-1}^2.

[2]

Let $\bar{X}$ be the mean of the sample. It is also given that there are 60 breakdowns in these 20 days.

(b) Find an unbiased estimate of the population mean.

[2]