Numerical measures

Numerical Measures

Standard deviation and variance calculated on ungrouped and grouped data.	Where raw data are given, candidates will be expected to be able to obtain standard deviation and mean values directly from calculators. Where summarised data are given, candidates may be required to use the formula from the booklet provided for the examination. It is advisable for candidates to know whether to divide by n or $(n-1)$ when calculating the variance; either divisor will be accepted unless a question specifically requests an unbiased estimate of a population variance.
Linear scaling.	Artificial questions requiring linear scaling will not be set, but candidates should be aware of the effect of linear scaling on numerical measures.
Choice of numerical measures.	Candidates will be expected to be able to choose numerical measures, including mean, median, mode, range and interquartile range, appropriate to given contexts. Linear interpolation will not be required.

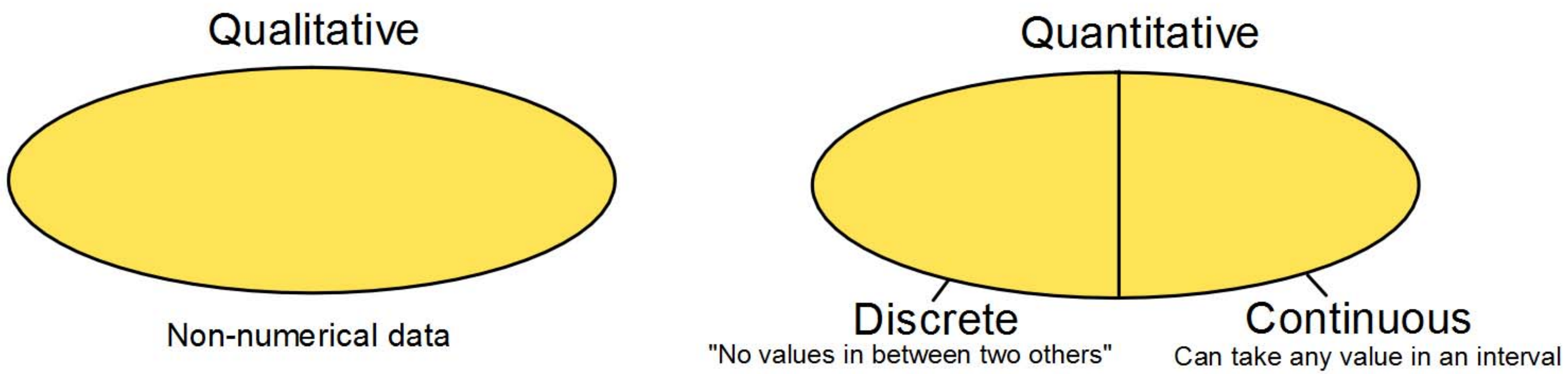
Introduction to statistics

Identify different types of variables and distinguish between primary ans secondary data.
Understand the terms population, sample, parameter, statistic.
Understand the concept of a simple random sample.

Statistics is about all aspect of dealing with data:
how to collect,
how to summarise it numerically,
how to present it pictorially
how to draw conclusions from it.

Statistics deals with events which have more than one possible outcome.
The quantity which varies is called a VARIABLE.

Types of variables



Primary and secondary data

Data collected for a specific investigation are known as PRIMARY Data.

Data read from other sources (publications, TV, ...) are called SECONDARY Data.

Population and Samples

A POPULATION is the entire collected data.
A SAMPLE is only part of these data.

Parameter and statistic

A numerical property (the mean, the median, etc..)
of a population is called a PARAMETER
of a sample is called a STATISTIC.

> A South American sports journalist intends to write a book about football in his home country. He will analyse all first division matches in the season. He records for each match whether it is a home win, an away win or a draw. He also records for each match the total number of goals scored and the amount of time played before a goal is scored. Reference books showed that in the previous season the mean number of goals per game was 2.317. On the first Saturday of the season he recorded the number of goals scored in each match and calculated the mean number of goals per match as 2.412. For the whole season the mean number of goals per match was 2.219.

Read the passage and identify an example of :

a) a population

b) a sample

c) a parameter

d) a statistic

e) a qualitative variable

f) a discrete variable

g) a continuous variable

h) a primary data

i) a secondary data

Random samples - Biased samples

A sample is random if every member of the population has an equal chance of being selected.
Moreover, all random samples (of the same size) should have an equal chance of being selected.

This second condition will depend on how you select the members of your sample.

A random sample chosen without replacement is called a SIMPLE RANDOM SAMPLE.

Numerical measures

Three measures of location can be used to describe the centre of a set of data – mode, median and mean.

The mode

- The mode is the value that occurs most often.

The median

- The median is the middle value when the data is put in order.
- In general if there are n observations first divide n by 2. When $\frac{n}{2}$ is a whole number find the mid-point of the corresponding term and the term above. When $\frac{n}{2}$ is not a whole number round the number up and pick the corresponding term.

The mean

- The mean is the sum of all the observations divided by the total number of observations.
- You can use the symbol Σ (the Greek letter s) instead of writing 'the sum of', and you can use $\bar{x}$ instead of writing 'the mean of the observations $x_1, x_2 \ldots x_n$'.
- The mean is given by

$$\text{Mean} = \frac{\text{Sum of observations}}{\text{Number of observations}} = \frac{\Sigma x}{n}$$

$\bar{x}$ is the symbol for the mean of a sample. μ is the symbol used for the mean of a population.

You should know which is the correct measure of location to use in each situation.

- **Mode** This is used when data is qualitative, or quantitative with either a single mode or bimodal. It is not very informative if each value occurs only once. The mode was not asked for in example 11 as every number occurs once only.
- **Median** This is used for quantitative data. It is usually used when there are extreme values, as in example 11, as they do not affect it.
- **Mean** This is used for quantitative data and uses all the pieces of data. It therefore gives a true measure of the data. However, it is affected by extreme values.

Exercises:

1 Meryl collected wild mushrooms every day for a week. When she got home each day she weighed them to the nearest 100 g. The weights are shown below.

500 700 400 300 900 700 700

a Write down the mode for these data.

b Calculate the mean for these data.

c Find the median for these data.

On the next day, Meryl collects 650 g of wild mushrooms.

d Write down the effect this will have on the mean, the mode and the median.

2 Joe collects six pieces of data x_1, x_2, x_3, x_4, x_5 and x_6. He works out that $\sum x$ is 256.2.

a Calculate the mean for these data.

He collects another piece of data. It is 52.

b Write down the effect this piece of data will have on the mean.

3 A small workshop records how long it takes, in minutes, for each of their workers to make a certain item. The times are shown in the table.

Worker	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Time in minutes	7	12	10	8	6	8	5	26	11	9

a Write down the mode for these data.

b Calculate the mean for these data.

c Find the median for these data.

d The manager wants to give the workers an idea of the average time they took. Write down, with a reason, which of the answers to **a**, **b** and **c** he should use.

1 **a** 700 g
b 600 g
c 700 g
d The mean will increase; the mode will remain unchanged; the median will decrease.

2 **a** 42.7
b The mean will increase.

3 **a** 8 minutes
b 10.2 minutes
c 8.5 minutes
d The median would be best. The mean is affected by the extreme value 26.

Frequency tables

Rebecca records the shirt collar size, x, of the male students in her year. The results are as follows.

x	Number of students, f
15	3
15.5	17
16	29
16.5	34
17	12

A frequency table is a quick way of writing a long list of numbers. It tells us that three students had a collar size of 15 and seventeen students had a collar size of 15.5 etc.

Find for these data

a the mode **b** the median **c** the mean.

Exercises:

1 The marks scored in a multiple choice statistics test by a class of students are:

5	9	6	9	10	6	8	5	5	7	9	7
8	6	10	10	7	9	6	9	7	7	7	8
6	9	7	8	6	7	8	7	9	8	5	7

a Draw a frequency distribution table for these data.

b Calculate the mean mark for these data.

c Write down the number of students who got a mark greater than the mean mark.

d Write down whether or not the mean mark is greater than the modal mark.

2 The table shows the number of eggs laid in 25 blackbirds' nests.

Number of eggs	0	1	2	3	4	5	6	7
Number of nests	0	0	0	1	3	9	8	4

Using your knowledge of measures of location decide what number of eggs you could expect a blackbird's nest to contain. Give reasons for your answer.

3 The table shows the frequency distribution for the number of petals in the flowers of a group of celandines.

Number of petals	Frequency (f)
5	8
6	57
7	29
8	3
9	1

a Work out how many celandines were in the group.

b Write down the modal number of petals.

c Calculate the mean number of petals.

d Calculate the median number of petals.

e If you saw a celandine, write down how many petals you would expect it to have.

4 The frequency table shows the number of breakdowns, b, per month recorded by a road haulage firm over a certain period of time.

Breakdowns b	Frequency f	Cumulative frequency
0	8	8
1	11	19
2	12	31
3	3	34
4	1	35
5	1	36

a Write down the number of months for which the firm recorded the breakdowns.

b Write down the number of months in which there were two or fewer breakdowns.

c Write down the modal number of breakdowns.

d Find the median number of breakdowns.

e Calculate the mean number of breakdowns.

f In a brochure about how many loads reach their destination on time, the firm quotes one of the answers to **c**, **d** or **e** as the number of breakdowns per month for its vehicles. Write down which of the three answers the firm should quote in the brochure.

5 A company makes school blazers in eight sizes. The first four sizes cost £48. The next three sizes cost £60 and the largest size costs £76.80. Write down, with a reason which of the mean, mode, or median cost the company is likely to use in advertising its average price.

1 a

Mark	5	6	7	8	9	10
Frequency	4	6	10	6	7	3

b 7.42
c 16
d The mean is greater.

2 5 eggs (mode 5, median 5, mean 5.44)

3 a 98
b 6
c 6.31
d 6
e 6

4 a 36
b 31
c 2
d 1
e 1.47
f the median

5 The company would use the mode (£48), since it is lower than the median (£54) and the mean (£56.10).

Grouped frequency tables

The mean, modal class and median can be calculated for grouped data. This section deals with grouped data. Original data may have been discrete or continuous, but all grouped data is treated as continuous data.

When data is presented as a grouped frequency table, specific data values are lost. This means that we work out **estimates** for the mean and median and the modal class rather than a specific value.

- **The mean of a sample of data that is summarised as a grouped frequency distribution is $\bar{x}$ where:**

  $\bar{x} = \frac{\sum fx}{\sum f}$ **and x is the mid-point of the group**

- **To find the median divide n by 2 and use interpolation to find the value of the corresponding term.**

- **The modal class is the class with the highest frequency.**

We do not need to do any rounding as we are dealing with continuous variables.

Since we do not have specific data values we call the class with the highest frequency the modal class.

Example:

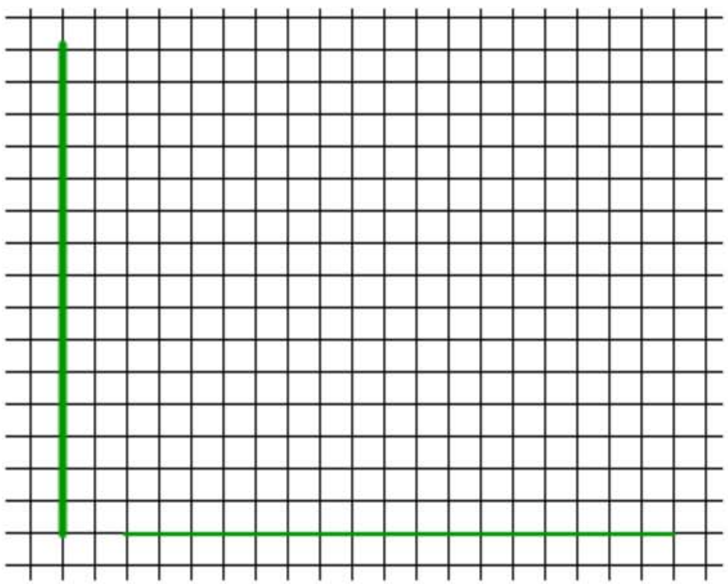
The length x mm, to the nearest mm, of a random sample of pine cones is measured. The data is shown below.

Length of pine cone (mm)	Number of pine cones, f	Cumulative frequency
30–31	2	2
32–33	25	27
34–36	30	57
37–39	13	70

a Write down the modal class.

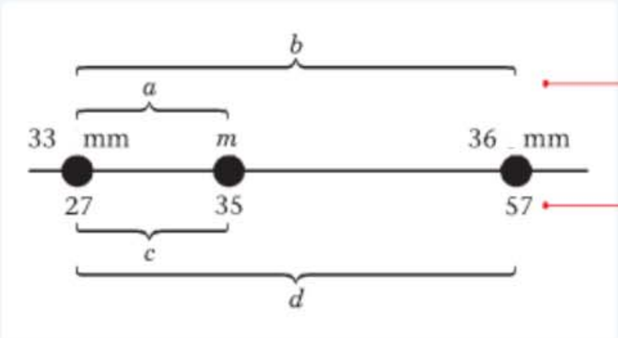
b Estimate the mean length of pine cone.

c Estimate the median length of pine cone.



Estimating the median - Linear interpolation

To find an estimate for the median we need to use **interpolation**. First set up the diagram below.



Practice: Using linear interpolation, work out the missing values

x	44	50	68
y	10		30

x	27	30	42
y	9.5		19.5

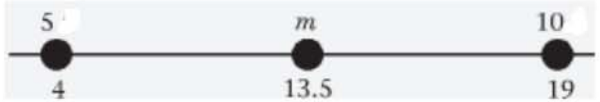
x	63	75	83
y	5		9

x	80	95	100
y	57		91

Example 2:

The numbers of questions answered correctly by children taking a general knowledge test are shown in the following frequency distribution.

Number of correct answers	Frequency
0–5	4
6–10	15
11–15	5
16–20	2
21–60	0
61–70	1



Estimate

a the mean number of correct answers,

b the median number of correct answers.

Mean = 10.98
Median = 8.17

Exercises:

1 A hotel is worried about the reliability of its lift. It keeps a weekly record of the number of times it breaks down over a period of 26 weeks. The data collected are summarised in the table opposite.

Number of breakdowns	Frequency of breakdowns (f)
0–1	18
2–3	7
4–5	1

a Estimate the mean number of breakdowns.

b Use interpolation to estimate the median number of breakdowns.

c The hotel considers that an average of more than one breakdown per week is not acceptable. Judging from your answers to **b** and **c** write down with a reason whether or not you think the hotel should consider getting a new lift.

2 The weekly wages (to the nearest £) of the production line workers in a small factory is shown in the table.

Weekly wage £	Number of workers, f,
175–225	4
226–300	8
301–350	18
351–400	28
401–500	7

a Write down the modal class.

b Calculate an estimate of the mean wage.

c Use interpolation to find an estimate for the median wage.

3 The noise levels at 30 locations near an outdoor concert venue were measured to the nearest decibel. The data collected is shown in the grouped frequency table.

Noise (decibels)	65–69	70–74	75–79	80–84	85–89	90–94	95–99
Frequency (f)	1	4	6	6	8	4	1

a Calculate an estimate of the mean noise level.

b A noise level above 82 decibels was considered unacceptable. Estimate the number of locations that had unacceptable noise levels.

4 DIY store **A** considered that it was good at employing older workers. A rival store **B** disagreed and said that it was better. The two stores produced a frequency table of their workers' ages. The table is shown below

Age of workers (to the nearest year)	Frequency store A	Frequency store B
16–25	5	4
26–35	16	12
36–45	14	10
46–55	22	28
56–65	26	25
66–75	14	13

By comparing estimated means for each store decide which store employs more older workers.

5 The speeds of vehicles passing a checkpoint were measured over a period of one hour, to the nearest mph. The data collected is shown in the grouped frequency table.

Speed (mph)	21-30	31-40	41-50	51-60	61-65	66-70	71-75
No. of vehicles (f)	4	7	38	42	5	3	1

a Write down the modal class.

b Calculate the difference, to two decimal places, between the median and the mean estimated speeds.

c The speed limit on the road is 60 mph. Work out an estimate for the percentage of cars that exceeded the speed limit.

1 **a** 1.19
b 0.722
c The median is less than 1 and the mean is only a little above 1 – it is 1 if rounded to the nearest whole number... The hotel need not consider getting a new lift yet but should keep an eye on the situation.

2 **a** £351 to £400
b £345
c £355

3 **a** 82.3 decibels
b 16

4 Store B (mean 51 years) employs older workers than store A (mean 50 years).

5 **a** 51 mph to 60 mph
b 0.71 mph (2 d.p.) (mean 50.03 mph, median 50.74 mph)
c 9%

Measures of dispersions

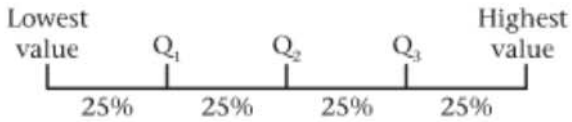
Range

The range of a set of data is very simple to find.

range = highest value − lowest value

Quartiles

Quartiles, Q_1, Q_2, Q_3, split the data into four parts.



Interquartile range

interquartile range = upper quartile − lower quartile

Method used for discrete data

This method is the same as we used for finding the median in Chapter 2.

- To calculate the **lower quartile,** Q_1, divide n by 4. When $\frac{n}{4}$ is a whole number find the mid-point of the corresponding term and the term above. When $\frac{n}{4}$ is not a whole number round the number up and pick the corresponding term.
- To calculate the **upper quartile**, Q_3, divide n by four and multiply by three. When $\frac{3n}{4}$ is a whole number find the mid-point of the corresponding term and the term above. When $\frac{3n}{4}$ is not a whole number round the number up and pick the corresponding term.

Method used for continuous data

- To calculate the **lower quartile**, Q_1, divide n by four and use interpolation to find the corresponding value.
- To calculate the **upper quartile**, Q_3, divide n by four and multiply by three and use interpolation to find the corresponding value.

Examples:

- Discrete data

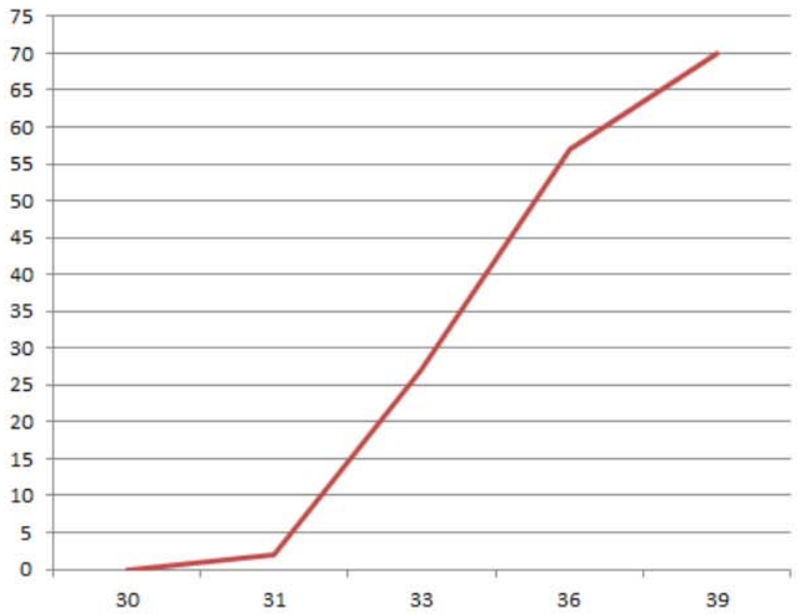
Find the range and interquartile range of the following data.

7 9 4 6 3 2 8 1 10 15 11

- Grouped frequency table

Length of time spent on internet (minutes)	Number of students	Cumulative frequency
30–31	2	2
32–33	25	27
34–36	30	57
37–39	13	70

The length of time (to the nearest minute), spent on the internet each evening by a group of students is shown in the table below. Calculate the interquartile range.



- Frequency table

x	Number of students, f
35	3
36	17
37	29
38	34
39	12

Rebecca records the number of CDs in the collections of students in her year.

The results are in the table opposite.

Find the interquartile range.

Exercises:

1 15 students do a mathematics test. Their marks are shown opposite.

7	4	9	7	6
10	12	11	3	8
5	9	8	7	3

a Find the value of the median.

b Find Q_1 and Q_3.

c Work out the interquartile range.

2 A group of workers were asked to write down their weekly wage. The wages were:

£550 £400 £260 £320 £500 £450 £460 £480 £510 £490 £505

a Work out the range for these wages.

b Find Q_1 and Q_3.

c Work out the interquartile range.

3 A superstore records the number of hours overtime worked by their employees in one particular week. The results are shown in the table.

a Fill in the cumulative frequency column and work out how many employees the superstore had in that week.

b Find Q_1 and Q_3.

c Work out the interquartile range.

Number of hours	Frequency	Cumulative frequency
0	25	
1	10	
2	20	
3	10	
4	25	
5	10	

4 A moth trap was set every night for five weeks. The number of moths caught in the trap was recorded. The results are shown in the table.

Number of moths	Frequency
7	2
8	5
9	9
10	14
11	5

Find the interquartile range.

5 The weights of 31 Jersey cows were recorded to the nearest kilogram. The weights are shown in the table.

Weight of cattle (kg)	Frequency	Cumulative frequency
300–349	3	
350–399	6	
400–449	10	
450–499	7	
500–549	5	

a Complete the cumulative frequency column in the table.

b Find the lower quartile, Q_1.

c Find the upper quartile, Q_3.

d Find the interquartile range.

6 The number of visitors to a hospital in a week was recorded. The results are shown in the table.

Number of visitors	Frequency
500–1000	10
1000–1500	25
1500–2000	15
2000–2500	5
2500–3000	5

Giving your answers to the nearest whole number find:

a the lower quartile Q_1, **b** the upper quartile Q_3, **c** the interquartile range.

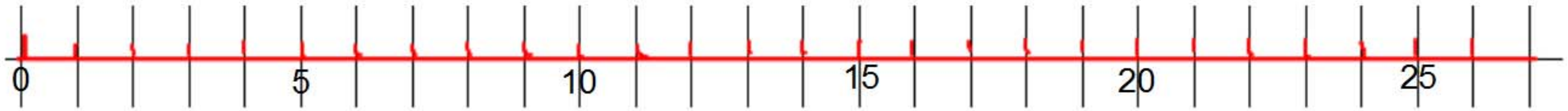
1 **a** 7
b 9
c 4
2 **a** £290
b $Q_1 = 400$, $Q_3 = 505$.
c £105
3 **a** 25, 35, 55, 65, 90, 100; total 100
b $Q_1 = 0.5$, $Q_3 = 4$.
c 3.5 hours
4 1 ($Q_1 = 9$, $Q_3 = 10$)
5 **a** 3, 9, 19, 26, 31
b 389 kg
c 480 kg
d 90.8 kg
6 **a** 1100
b 1833
c 733

Standard deviation and variance

1) Deviation from the mean

On a number line, plot the points

$x_1 = 15$, $x_2 = 9$, $x_3 = 23$, $x_4 = 12$, $x_5 = 17$



a) Work out μ. Plot it on the line.

b) Work out all the $x_i - \mu$ for i from 1 to 5

c) Work out the "average" deviation

What do you notice?

$x_i - \mu$ is called the deviation of x_i from the mean.

2) The variance

Deviations cancel each other out, because some are positive and some are negative.
To overcome this issue, we could study the SQUARE of the deviations.

$x_1 = 15$, $x_2 = 9$, $x_3 = 23$, $x_4 = 12$, $x_5 = 17$ $\mu=15.2$

a) Work out all the $(x_i-\mu)^2$.

b) Work out the "average" of these numbers.

The POPULATION VARIANCE is denoted σ^2. *It is given by:* $\sigma^2 = \frac{1}{n}\sum(x-\mu)^2$

The SAMPLE VARIANCE is denoted s^2. *It is given by:* $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$

This is sometimes referred as the UNBIASED estimate of the population variance

Note: The larger the variance is, the more variation there is in the x-values

Alternative formula:

A more friendly version of this formula is

- **variance** $= \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$

The following saying may help you remember this formula. Variance = 'mean of the squares minus the square of the mean'.

"The variance is the mean of the squares minus the square of the mean"

Example:

The marks gained in a test by seven randomly selected students are

3 4 6 2 8 8 5

Use the friendly formula to work out the variance and standard deviation of the marks of the seven students.

There is on issue with the variance: The unit !
if x is in cm, the variance is in cm^2,
if x is in kg, the variance is in kg^2! which does not make a lot of sense.

So we introduce the : ■ standard deviation $= \sqrt{\text{Variance}}$

Exercises:

1 Given that for a variable x:

$\sum x = 24 \quad \sum x^2 = 78 \quad n = 8$

Find:

a The mean. **b** The variance σ^2. **c** The standard deviation σ.

2 Ten collie dogs are weighed (w kg). The following summary data for the weights are shown below:

$\sum w = 241 \quad \sum w^2 = 5905$

Use this summary data to find the standard deviation of the collies' weights.

3 Eight students' heights (h cm) are measured. They are as follows:

165 170 190 180 175 185 176 184

a Work out the mean height of the students.

b Given $\sum h^2 = 254\,307$ work out the variance. Show all your working.

c Work out the standard deviation.

4 For a set of 10 numbers:

$\sum x = 50 \quad \sum x^2 = 310$

For a set of 15 numbers :

$\sum x = 86 \quad \sum x^2 = 568$

Find the mean and the standard deviation of the combined set of 25 numbers.

5 The number of members (m) in six scout groups was recorded. The summary statistics for these data are:

$\sum m = 150 \quad \sum m^2 = 3846$

a Work out the mean number of members in a scout group.

b Work out the standard deviation of the number of members in the scout groups.

6 There are two routes for a worker to get to his office. Both the routes involve hold ups due to traffic lights. He records the time it takes over a series of six journeys for each route. The results are shown in the table.

Route 1	15	15	11	17	14	12
Route 2	11	14	17	15	16	11

a Work out the mean time taken for each route.

b Calculate the variance and the standard deviation of each of the two routes.

c Using your answers to **a** and **b** suggest which route you would recommend. State your reason clearly.

1 **a** 3
b 0.75
c 0.866

2 3.11 kg

3 **a** 178 cm
b 59.9 cm^2
c 7.74 cm

4 mean 5.44, standard deviation 3.25

5 **a** 25
b 4

6 **a** The mean for both routes is 14.
b Route 1 has variance 4 and standard deviation 2. Route 2 has variance 5.33 and standard deviation 2.31.
c Route 1 would be best. Although the means are the same, the standard deviation for route 1 is lower, so this route is more reliable.

Variance and grouped frequency tables

Calculation of the variance and standard deviation for a frequency table and a grouped frequency distribution where x is the mid-point of the class.

If you let f stand for the frequency, then $n = \sum f$ and

$$\text{Variance} = \frac{\sum f(x - \bar{x})^2}{\sum f} \quad \text{or} \quad \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

Example: Calculate an estimate of the standard deviation

x	Number of students, f
35	3
36	17
37	29
38	34
39	26

Length of telephone call	Number of occasions
$0 < l \leqslant 5$	4
$5 < l \leqslant 10$	15
$10 < l \leqslant 15$	5
$15 < l \leqslant 20$	2
$20 < l \leqslant 60$	0
$60 < l \leqslant 70$	1

1.09
11.35

Exercises:

1 For a certain set of data:

$\sum fx = 1975 \qquad \sum fx^2 = 52\,325 \qquad n = 100$

Work out the variance for these data.

2 For a certain set of data

$\sum fx = 264 \qquad \sum fx^2 = 6456 \qquad n = 12$

Work out the standard deviation for these data.

3 Nahab asks the students in his year group how much pocket money they get per week. The results, rounded to the nearest pound, are shown in the table.

Number of £'s (x)	Number of students f	fx	fx^2
8	14		
9	8		
10	28		
11	15		
12	20		
Totals			

a Complete the table.

b Using the formula $\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$ work out the variance for these data.

c Work out the standard deviation for these data.

4 In a student group, a record was kept of the number of days absence each student had over one particular term. The results are shown in the table.

Number days absent (x)	Number of students f	fx	fx^2
0	12		
1	20		
2	10		
3	7		
4	5		

a Complete the table.

b Calculate the variance for these data.

c Work out the standard deviation for these data.

5 A certain type of machine contained a part that tended to wear out after different amounts of time. The time it took for 50 of the parts to wear out was recorded. The results are shown in the table.

Lifetime in hours	Number of parts	Mid-point x	fx	fx^2
$5 < h \leq 10$	5			
$10 < h \leq 15$	14			
$15 < h \leq 20$	23			
$20 < h \leq 25$	6			
$25 < h \leq 30$	2			

a Complete the table.

b Calculate an estimate for the variance and the standard deviation for these data.

1 133

2 7.35

3 **a**

Number of £'s (x)	Number of students (f)	$f\,x$	$f\,x^2$
8	14	112	896
9	8	72	648
10	28	280	2800
11	15	165	1815
12	20	240	2880
Totals	85	869	9039

b 1.82

c £1.35

4 **a**

Number of days absent (x)	Number of students (f)	$f\,x$	$f\,x^2$
0	12	0	0
1	20	20	20
2	10	20	40
3	7	21	63
4	5	20	80

b 1.51

c 1.23

5 **a**

Lifetime in hours	Number of parts	Mid-point (x)	f x	f x^2
5 < h = 10	5	7.5	37.5	281.25
10 < h = 15	14	12.5	175.0	2187.50
15 < h = 20	23	15.5	402.5	7043.75
20 < h = 25	6	22.5	135.0	3037.50
25 < h = 30	2	27.5	55.0	1512.50

b variance 22.0, standard deviation 4.69 hours

Summary of key points

1 The **range** of a set of data is the difference between the highest and lowest value in the set.

2 The **quartiles**, Q_1, Q_2, Q_3, Q_4, split the data into four parts. To calculate the **lower quartile, Q_1,** divide n by 4.

3 For discrete data for the lower quartile, Q_1, divide n by 4. To calculate the upper quartile, Q_3, divide n by 4 and multiply by 3. When the result is a whole number find the mid-point of the corresponding term and the term above. When the result is not a whole number round the number up and pick the corresponding term.

4 For continuous grouped data for Q_1, divide n by 4 and for Q_3 divide n by 4 and multiply by 3. Use interpolation to find the value of the corresponding term.

5 The **interquartile range** is $Q_3 - Q_1$.

7 The **deviation** of an observation x from the mean is given by $x - \bar{x}$.

- **Variance** $= \dfrac{\sum(x-\bar{x})^2}{n} = \dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2$
- **Standard deviation** $= \sqrt{\textbf{Variance}}$

8 To work out the variance and standard deviation for a frequency table and a grouped frequency distribution where x is the mid-point of the class, let f stand for the frequency, then $n = \sum f$

- Variance $= \dfrac{\sum f(x-\bar{x})^2}{\sum f}$ or $\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f}\right)^2$

9 When the data values are large you can use coding to make the numbers easier to work with. To find the standard deviation of the original data, find the standard deviation of the coded data and either multiply this by what you divided by, or divide this by what you multiplied by.

source:
Edexcel live textbook-Statistics 1