chapter 2

Statistics 1

Section 2.5 Measures of variability

PROJECT MATHS
Text & Tests 5
LEAVING CERTIFICATE
HIGHER LEVEL
STRAND 1
PROBABILITY & STATISTICS

70

When dealing with **averages** in the previous section, we were looking for a data value that was typical or representative of all the data values.

In this section, we will discuss the measure of the spread of the data about the mean to help us describe the data more fully.

The three most common ways of measuring the spread or **variability** of data are the **range**, the **interquartile range** and **standard deviation**.

1. The range

The **range** of a set of data is the highest value of the set minus the lowest value.

The range of a set of data is the largest value minus the smallest value.

It shows the **spread** of the data.

It is very useful when comparing two sets of data.

The range is a crude measure of spread because it uses only the largest and smallest value of the data.

The range of the numbers 14, 18, 11, 27, 21, 19, 33, 24 is

Range $= 33 - 11 = 22$

2. Quartiles and Interquartile range

When data is arranged in order of size, we have already learned that the median is the value halfway into the data. So we can say that the median divides the data into two halves. The data can also be divided into four quarters.

When the data is arranged in ascending order of size:

- the **lower quartile** is the value one quarter of the way into the data
- the **upper quartile** is the value three quarters of the way into the data
- the upper quartile minus the lower quartile is called the **interquartile range**.

The lower quartile is written Q_1; the median is Q_2; the upper quartile is Q_3.

Consider the following data which is arranged in order of size. It contains 15 numbers.

The lower quartile is the value one quarter of the way along. Written as Q_1

The median is the value halfway along. Written as Q_2

The upper quartile is the value three quarters of the way along. Written as Q_3

0 2 4 $\boxed{5}$ 7 8 10 $\boxed{12}$ 12 12 13 $\boxed{14}$ 14 15 16

Q_1 is the 4th value

Q_2 is the 8th value

Q_3 is the 12th value

The interquartile range is
upper quartile − lower quartile
$= Q_3 - Q_1$

The lower quartile $Q_1 = 5$.
The median $Q_2 = 12$.
The upper quartile $Q_3 = 14$.
The interquartile range $= Q_3 - Q_1 = 14 - 5 = 9$.

When n data values are written in order:

- the lower quartile, Q_1, is the $\frac{1}{4}(n + 1)$th value
- the median, Q_2, is the $\frac{1}{2}(n + 1)$th value
- the upper quartile, Q_3, is the $\frac{3}{4}(n + 1)$th value.

The interquartile range is
upper quartile − lower quartile
$= Q_3 - Q_1$

If you get a non-integer value for $\frac{1}{4}(n + 1)$ or $\frac{3}{4}(n + 1)$, e.g., $3\frac{3}{4}$, find the mean of the 3rd and 4th pieces of data.

If you get a non-integer for $\frac{1}{2}(n + 1)$, take the mean of the two data values on either side. For example if $\frac{1}{2}(n + 1) = 3\frac{1}{2}$, find the mean of the 3rd and 4th data values.

Example 1

These are the test marks of 11 students:

~~52~~, ~~78~~, ~~61~~, ~~49~~, ~~79~~, ~~47~~, ~~54~~, ~~58~~, ~~72~~, ~~62~~, ~~73~~

Find (i) the median (ii) the lower quartile (iii) the upper quartile (iv) the interquartile range. (v) Range

order

1 2 3 4 5 6 7 8 9 10 11

47, 49, 52, 54, 58, 61, 62, 72, 73, 78, 79

Q_1 lower quartile

Q_2 median

Q_3 upper quartile

middle of 11 is $\frac{11+1}{2} = 6$

$Q_1 \Rightarrow \frac{11+1}{4} = 3$

(i) $Q_2 = 61$

(ii) $Q_1 = 52$

(iii) $Q_3 = 73$

(iv) I.Q.R. $= 73 - 52 = 21$

(v) Range $= 79 - 47 = 32$

3. Standard deviation

One of the most important and frequently-used measures of spread is called **standard deviation**. It shows how much variation there is from the average (mean). It may be thought of as the average difference of the scores from the mean, that is, how far they are away from the mean. A low standard deviation indicates that the data points tend to be very close to the mean; a high standard deviation indicates that the data is spread out over a large range of values.

The Greek letter σ is used to denote standard deviation.

Take, for example, all adult men in Ireland. The average height is about 177 cm with a standard deviation of about 8 cm.

For this large population, about 68% of the men have a height within 8 cm of the mean.

> If the mean is $\bar{x}$ and σ is the standard deviation of a large sample, then 68% will lie between $\bar{x} + \sigma$ and $\bar{x} - \sigma$

Example 2

Find the standard deviation of the numbers 6, 9, 10, 12, 13.

use calculator

	X	FREQ.
1	6	1
2	9	1
3	10	1
4	12	1
5	13	1

$\sigma = 2.45$

Example 3

Find the standard deviation of the following frequency distribution:

Variable (x)	1	2	3	4	5	6
Frequency (f)	9	9	6	4	7	3

on calculator

	X	FREQ
1	1	9
2	2	9
3	3	6
4	4	4
5	5	7
6	6	3

$\sigma = 1.65$

Using a calculator to find the standard deviation of a frequency distribution

We will use a **Casio fx-83 ES** to find the standard deviation of a frequency distribution.

The following frequency distribution table shows the number of birdies scored per round of golf.

No. of birdies	0	1	2	3	4	5	6
Frequency	5	6	4	6	3	1	0

Find the mean and standard deviation, correct to one decimal place.

Key in MODE and select 2 for statistics mode.

Then select 1 for 1 − VAR and input variables.

0 = 1 = 2 = 3 = 4 = 5 = 6 = REPLAY 5 = 6 = 4 = 6 = 3 = 1 = 0 =

CASIO *fx-83ES*

	X	FREQ
1	0	5
2	1	6
3	2	4
4	3	6
5	4	3
6	5	1
7	6	0

For answers key in

AC SHIFT 1 5 2 = 1.96 = 2.0 = mean (birdies per round) $\bar{X}$

AC SHIFT 1 5 3 = 1.4554… = 1.5 = standard deviation σx

∴ Mean = 2.0 and standard deviation = 1.5

these nos are different in your calculator

Example 4

Find (a) the mean (b) the standard deviation of the following set of numbers:
(i) 5, 3, 1, 8, 2 (ii) 10, 6, 2, 16, 4

on calculator

	X	FREQ
1	5	1
2	3	1
3	1	1
4	8	1
5	2	1

$\bar{X} = 3 \cdot 8$

$\sigma = 2 \cdot 48$

	X	FREQ
1	10	1
2	6	1
3	2	1
4	16	1
5	4	1

$\bar{x} = 7 \cdot 6$

$\sigma = 4 \cdot 96$

Example 5

Here are the marks of 24 students in a science test:

48	54	76	34	82	67	76	92	54	72	86	47
80	73	64	57	68	36	82	74	71	62	46	52

(i) Find P_{60}
(ii) Find P_{75}
(iii) If Sinead scored 74 in the test, find on what percentile is her score.

order

14·4
14 | 15 17 18

34, 36, 46, 47, 48, 52, 54, 54, 57, 62, 64, 67, 68, 71, 72, 73, 74, 76, 76, 80, 82, 82, 86, 92

$P_{60} = 60^{th}$ percentile ?

$24(60\%) = 14 \cdot 4$

$24(75\%) = 18$

each person is
$\frac{100}{24} = 4 \cdot 16$

no. 17 $= 4 \cdot 16(17) = 70 \cdot 83$
≈ 71

(i) $P_{60} = 72$

(ii) $P_{75} = Q_3 = 76$

(iii) Sinead is P_{71}

9. Use your calculator, or otherwise, to show that the standard deviation of the numbers

3 4 6 2 8 8 5 is 2.17.

	X	FREQ
1	3	1
2	4	1
3	6	1
4	2	1
5	8	1
6	8	1
7	5	1

$\sigma = 2.166 \approx 2.17$ (2 dp)

11. Verify that 2 is the mean of this distribution.
Hence calculate the standard deviation, correct to 1 decimal place.

Variable	0	2	3	4
Frequency	4	3	2	3

x	0	2	3	4	Σ
f	4	3	2	3	12
xf	0	6	6	12	24

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{24}{12} = 2$$

using calculator

	x	freq
1	0	4
2	2	3
3	3	2
4	4	3

σx "standard deviation"

$\bar{x}$ "mean"

$\sigma = 1.581 \approx 1.6$ (1 dp)