

# Edexcel Statistics 1

## Working with data

### Section 2: Representation and summary of data - dispersion

#### Multiple choice test

For questions 1 – 4 use the following data set:

64 99 48 68 78 72 66 76 93 55 68 87 77 75 69

1. What is the range?

- (a) 51
- (b) 5
- (c) 73.5
- (d) 99
- (e) I don't know

2. What is the upper quartile?

- (a) 12
- (b) 78
- (c) 66
- (d) 87
- (e) I don't know

3. What is the interquartile range?

- (a) 6
- (b) 51
- (c) 8
- (d) 12
- (e) I don't know

4. The following ten random numbers were generated by my calculator:

7, 4, 5, 2, 3, 5, 4, 4, 9, 7

What is the variance of this set of data?

- (a) 2.11 (to 3 s.f.)
- (b) 4.44 (to 3 s.f.)
- (c) 4
- (d) 2
- (e) I don't know

5. The following are the marks of Alice, Brian, Charles and Dora in five tests.

Alice	12	8	17	18	15
Brian	15	14	5	20	16
Charles	14	14	20	8	14
Dora	12	19	9	14	16

Who has the most consistent set of marks?

## Edexcel S1 Working with data Section 3 MC test solutions

- (a) Alice  
(c) Charles  
(e) I don't know
- (b) Brian  
(d) Dora

6. A set of 15 numbers has mean 7 and standard deviation 4. A second set of 25 numbers has mean 9 and standard deviation 5. What is the standard deviation of the combined set of 40 numbers?

- (a) 4.75  
(c) 4.5  
(e) I don't know
- (b) 4.63 (to 3 s.f.)  
(d) 4.81 (to 3 s.f.)

7. A random sample of 400 people was selected from a telephone directory. The length of each surname was recorded with the following results:

Length of surname	2	3	4	5	6	7	8	9	10
Frequency	2	8	142	102	63	38	24	12	9

What is the standard deviation of the length of the surnames?

- (a) 1.57 (to 3 s.f.)  
(c) 2.48 (to 3 s.f.)  
(e) I don't know
- (b) 2.73 (to 3 s.f.)  
(d) 1.58 (to 3 s.f.)

8. A manufacturer made rods which had a nominal length of 45 cm. A sample of 60 rods was measured, each to the nearest millimetre. Their deviations ( $x$ ) from 45 cm were recorded in millimetres and summarised as  $\sum x = 90$ ,  $\sum x^2 = 370$ . (For example, a rod of length 44.3 cm would be recorded as  $-7$ ).

What is the standard deviation of the lengths of the rods?

- (a) 0.200 cm (to 3 s.f.)  
(c) 0.198 cm (to 3 s.f.)  
(e) I don't know
- (b) 2.00 cm (to 3 s.f.)  
(d) 1.98 cm (to 3 s.f.)

Questions 9 and 10 are about the following information:

The masses, measured to the nearest kilogram, of 100 boys were recorded.

<b>Mass (kg)</b>	51 - 55	56 - 60	61 - 65	66 - 70	71 - 75
<b>Frequency</b>	7	14	45	20	14

## Edexcel S1 Working with data Section 3 MC test solutions

9. What is an estimate for the standard deviation of the masses?

- (a) 5.37 kg (to 3 s.f.)                      (b) 7.91 kg (to 3 s.f.)  
(c) 28.8 kg (to 3 s.f.)                      (d) 5.34 kg (to 3 s.f.)  
(e) I don't know

10. What is an estimate of the 20<sup>th</sup> percentile?

- (a) 58 kg    (b) 60.6 kg  
(c) 60.1 kg                                      (d) 55 – 60 kg  
(e) I don't know

## Section 2: Representation and summary of data - dispersion

### Solutions to multiple choice test

1. The correct answer is (a)

Largest number is 99, smallest number is 48.

$$\text{Range} = 99 - 48 = 51.$$

2. The correct answer is (b)

Discarding the median value (72), the upper half of the data is

75 76 77 (78) 87 93 99

The upper quartile is the median of this set of data, which is 78.

3. The correct answer is (d)

Discarding the median value, the lower half of the data is

48 55 64 (66) 68 68 69

The lower quartile is the median of this set of data, which is 66.

$$\text{Interquartile range} = 78 - 66 = 12.$$

## Edexcel S1 Working with data Section 3 MC test solutions

4. The correct answer is (c)

$$\text{Sample mean } \bar{x} = \frac{7+4+5+2+3+5+4+4+9+7}{10} = 5$$

$$S_{xx} = 2^2 + 1^2 + 0^2 + 3^2 + 2^2 + 0^2 + 1^2 + 1^2 + 4^2 + 2^2 = 40$$

$$\text{variance} = \frac{S_{xx}}{n} = \frac{40}{10} = 4$$

5. The correct answer is (d)

$$\text{Alice: } \sum a = 70 \Rightarrow \bar{a} = 14, \quad \sum a^2 = 1046 \Rightarrow S_{aa} = 1046 - 5 \times 14^2 = 66$$

$$\text{Brian: } \sum b = 70 \Rightarrow \bar{b} = 14, \quad \sum b^2 = 1102 \Rightarrow S_{bb} = 1102 - 5 \times 14^2 = 122$$

$$\text{Charles: } \sum c = 70 \Rightarrow \bar{c} = 14, \quad \sum c^2 = 1052 \Rightarrow S_{cc} = 1052 - 5 \times 14^2 = 72$$

$$\text{Dora: } \sum d = 70 \Rightarrow \bar{d} = 14, \quad \sum d^2 = 1038 \Rightarrow S_{dd} = 1038 - 5 \times 14^2 = 58$$

Dora is the most consistent as she has the lowest value of  $S_{xx}$ .

6. The correct answer is (a)

For the set of 15 numbers:

$$\bar{x} = 7 \Rightarrow \sum x = 7 \times 15 = 105$$

$$\sqrt{\frac{S_{xx}}{15}} = 4 \Rightarrow S_{xx} = 240$$

$$\Rightarrow \sum x^2 - n\bar{x}^2 = 240$$

$$\Rightarrow \sum x^2 - 15 \times 7^2 = 240$$

$$\Rightarrow \sum x^2 = 975$$

For the set of 25 numbers:

$$\bar{y} = 9 \Rightarrow \sum y = 9 \times 25 = 225$$

$$\sqrt{\frac{S_{yy}}{25}} = 5 \Rightarrow S_{yy} = 625$$

$$\Rightarrow \sum y^2 - n\bar{y}^2 = 625$$

$$\Rightarrow \sum y^2 - 25 \times 9^2 = 625$$

$$\Rightarrow \sum y^2 = 2650$$

For the combined set of 40 numbers:

$$\sum z = 105 + 225 = 330 \text{ and } \sum z^2 = 975 + 2650 = 3625$$

$$\bar{z} = \frac{330}{40} = 8.25$$

$$S_{zz} = \sum z^2 - n\bar{z}^2 = 3625 - 40 \times 8.25^2 = 902.5$$

## Edexcel S1 Working with data Section 3 MC test solutions

$$\text{Standard deviation} = \sqrt{\frac{S_{zz}}{40}} = 4.75$$

7. The correct answer is (a)

Length, $x$	2	3	4	5	6	7	8	9	10	Total
$x^2$	4	9	16	25	36	49	64	81	100	
Frequency, $f$	2	8	142	102	63	38	24	12	9	400
$fx$	4	24	568	510	378	266	192	108	90	2140
$fx^2$	8	72	2272	2550	2268	1862	1536	972	900	12440

$$\begin{aligned} S_{xx} &= \sum fx^2 - n\bar{x}^2 \\ &= 12443 - 400 \times 5.35^2 \\ &= 991 \end{aligned}$$

$$\text{Standard deviation} = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{991}{400}} = 1.57 \text{ (3 s.f.)}$$

8. The correct answer is (c)

$$S_{xx} = \sum x^2 - n\bar{x}^2 = 370 - 60 \times 1.5^2 = 235$$

$$\text{Standard deviation of } x = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{235}{60}}$$

$$\text{Standard deviation of } l = \frac{1}{10} \sqrt{\frac{235}{60}} = 0.198 \text{ cm (3 s.f.)}$$

9. The correct answer is (d)

$$\text{using the coding } x = \frac{m - 63}{5}$$

Mass (kg)	51 - 55	56 - 60	61 - 65	66 - 70	71 - 75	Total
Mid-interval value, $m$	53	58	63	68	73	
Coded value, $x$	-2	-1	0	1	2	
$x^2$	4	1	0	1	4	
Frequency, $f$	7	14	45	20	14	100
$fx$	-14	-14	0	20	28	20
$fx^2$	28	14	0	20	56	118

$$\bar{x} = \frac{20}{100} = 0.2$$

## Edexcel S1 Working with data Section 3 MC test solutions

$$S_{xx} = \sum fx^2 - n\bar{x}^2 = 118 - 100 \times 0.2^2 = 114$$

$$\text{Standard deviation for } x = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{114}{100}}$$

$$\text{Standard deviation for the masses} = 5 \sqrt{\frac{114}{100}} = 5.34 \text{ kg (3 s.f.)}$$

10. The correct answer is (c)

The 20<sup>th</sup> percentile is the 20<sup>th</sup> data item. This lies in the 55.5 – 60.5 class interval, and is the 13<sup>th</sup> item of the 14 items in that class.

$$\text{Median} = 55.5 + \frac{13}{14} \times 5 = 60.1 \text{ kg.}$$