Edexcel Statistics 1

Working with data

Section 3: Representing data

Multiple choice test

For questions 1 and 2 use the information given below:

The following numbers are represented by a sorted stem-and-leaf diagram with seven branches.

23	49	38	31	45	17	56	22	33	31	48	49	54
64	28	44	67	34	53	26	49	38	23	34	35	75

1. What is the "stem" for the number 48?

(a) 8 (b) 4
(c) 7 (d) 0
(e) I don't know

2. How many "leaves" are there on the third "stem"?

(a) 6 (b) 7
(c) 8 (d) 9
(e) I don't know

Questions 3 – 5 are about the following data.

The weights of 50 newborn babies (in kg) in a particular hospital are as follows.

3.24	2.97	3.48	3.75	3.13	3.42	3.57	4.14	3.88	3.65
2.94	3.19	3.79	3.71	3.45	3.52	3.68	3.31	3.48	3.06
3.40	3.57	3.22	4.03	3.91	3.67	3.31	3.58	3.42	3.37
2.85	3.59	3.40	3.38	3.76	3.82	3.47	3.61	3.05	3.74
3.29	3.56	3.41	2.89	3.78	3.43	3.57	3.38	4.17	3.44

The data is to be grouped into the following class intervals:

2.80 – 2.99	3.00 – 3.19	3.20 – 3.29
3.30 – 3.39	3.40 – 3.49	3.50 – 3.59
3.60 – 3.69	3.70 – 3.79	3.80 – 3.99
4.00 – 4.19		

Edexcel S1 Data Section 3 MC test solutions

3. What is the frequency for the class interval 3.40 – 3.49?

(a) 11 (b) 12
(c) 10 (d) 9
(e) I don't know

4. What is the frequency density for the class interval 3.80 – 3.99?

(a) 30 (b) 3
(c) 0.6 (d) 15
(e) I don't know

5. Six students each made a statement about histograms.

Anne said "there are gaps between the columns".
Brian said "the columns are always of equal width".
Catherine said "the vertical axis is labelled 'frequency density'".
Diana said "the horizontal axis shows the variable being measured".
Edward said "the frequency is proportional to the height of each column".
Frank said "the frequency is proportional to the area of each column".

Who was correct?

(a) Brian, Diana and Edward only (b) Catherine, Diana and Frank only
(c) Catherine and Frank only (d) Anne, Catherine and Frank only
(e) I don't know

6. The times taken by a number of students to answer a mental arithmetic problem correctly were recorded and the following table produced:

Time (sec)	$0 \le t < 2$	$2 \le t < 6$	$6 \le t < 10$	$10 \le t \le 20$
Frequency density (number/sec)	7	5	8	3

How many students were there?

(a) 96 (b) 23
(c) 136 (d) 88
(e) I don't know

Edexcel S1 Data Section 3 MC test solutions

7. Using the definition of outliers as values more than 2 standard deviations from the mean, how many outliers are there in the following data?

47, 19, 25, 29, 4, 52, 22, 36, 21, 25

(a) 3
(b) 1
(c) 0
(d) 2
(e) I don't know

Questions 8 and 9 are about the data set

47, 70, 38, 53, 51, 27, 55, 18, 55, 42, 47, 45.

8. What is the interquartile range of the data set?

(a) 17
(b) 36
(c) 14
(d) 7
(e) I don't know

9. Using the definition of outliers relating to the interquartile range, which data items are outliers?

(a) no outliers
(b) 18
(c) 70
(d) 18 and 70
(e) I don't know

10. A box-and-whisker plot shows the following information: smallest value = 146 cm,
Q_1 = 217 cm, Q_2 = 253 cm, Q_3 = 325 cm and the largest value = 413 cm. What is the shape of the distribution?

(a) symmetrical
(b) negative skew
(c) uniform
(d) positive skew
(e) I don't know

Edexcel S1 Data Section 3 MC test solutions

Working with data

Section 3: Representing data

Solutions to multiple choice test

1. The correct answer is (b)

1	7
2	2 3 3 6 8
3	1 1 3 4 4 5 8 8
4	4 5 8 9 9 9
5	3 4 6
6	4 7
7	5

The "stem" is the figure on the left, and in this case this corresponds to the tens digit.
For the number 48, the "stem" is therefore 4.

2. The correct answer is (c)

The third "stem" is the "3". There are 8 "leaves" on this "stem" (i.e. 8 numbers in the 30's).

3. The correct answer is (a)

Class interval	Class width	Frequency	Frequency density
2.80 – 2.99	0.2	4	20
3.00 – 3.19	0.2	4	20
3.20 – 3.29	0.1	3	30
3.30 – 3.39	0.1	5	50
3.40 – 3.49	0.1	11	110
3.50 – 3.59	0.1	7	70
3.60 – 3.69	0.1	4	40
3.70 – 3.79	0.1	6	60
3.80 – 3.99	0.2	3	15
4.01 – 4.19	0.2	3	15

The frequency for the class interval 3.40 – 3.49 is 11.

Edexcel S1 Data Section 3 MC test solutions

4. The correct answer is (d)

$$\text{Frequency density} = \frac{\text{frequency}}{\text{class width}}$$

For class interval 3.80 – 3.99, frequency density $= \frac{3}{0.2} = 15$

5. The correct answer is (b)

Anne is wrong. A bar chart has gaps between the columns; a histogram does not.
Brian is wrong. Histograms can have columns of unequal widths.
Catherine is correct.
Diana is correct.
Edward is wrong, and Frank is correct. The frequency is proportional to the area of each column.
Catherine, Diana and Frank are correct.

6. The correct answer is (a)

Frequency = frequency density × class width

Time (sec)	$0 \le t < 2$	$2 \le t < 6$	$6 \le t < 10$	$10 \le t \le 20$
Class width	2	4	4	10
Frequency density (number/sec)	7	5	8	3
Frequency	14	20	32	30

The total number of students = 14 + 20 + 32 + 30 = 96.

7. The correct answer is (c)

$$\sum x = 280 \quad \Rightarrow \bar{x} = \frac{280}{10} = 28$$

$$S_{xx} = 19^2 + 9^2 + 3^2 + 1^2 + 24^2 + 24^2 + 6^2 + 8^2 + 7^2 + 3^2$$
$$= 1762$$

$$\text{Standard deviation} = \sqrt{\frac{S_{xx}}{n-1}} = \sqrt{\frac{1762}{9}} = 13.99$$

Outliers are more than 2 standard deviations from the mean.
In this case outliers are numbers which are zero or less, or 56 or more.
There are no outliers.

Edexcel S1 Data Section 3 MC test solutions

8. The correct answer is (c)

In order: 18 27 38 42 45 47 47 51 53 55 55 70

There are 12 data items.
$\frac{1}{4} \times 12 = 3$, so lower quartile is the mean of the 3rd and 4th data items.
Lower quartile = 40
$\frac{3}{4} \times 12 = 9$, so lower quartile is the mean of the 9th and 10th data items.
Upper quartile = 54
Interquartile range = 54 – 40 = 14

9. The correct answer is (b)

Outliers are 1.5 × IQR above the upper quartile or below the lower quartile.
1.5 × IQR = 1.5 × 14 = 21
Outliers are above 54 + 21 = 75, or below 40 – 21 = 19.

10. The correct answer is (d)

$Q_2 - Q_1 = 253 - 217 = 35$
$Q_3 - Q_2 = 325 - 253 = 72$
The median is nearer to the lower quartile than to the upper quartile, so the distribution is positively skewed.