Edexcel Collecting and interpreting data

Topic assessment

1. George records the time he spends per day surfing the internet for the first three weeks of May. The times, given to the nearest minute, are as follows:

0	26	13	5	18	12	35
24	61	16	10	26	15	0
0	73	21	17	16	42	32

(i) Illustrate the data using a sorted stem and leaf diagram with eight stems. Comment briefly on the shape of the distribution. [3]

(ii) Find the mode, median and mean, commenting on their relative usefulness as measures of central tendency for this data set. [5]

(iii) Calculate the standard deviation and hence find any outliers that are more than 2 standard deviations from the mean. [4]

(iv) George's Dad claims that he is spending too much time on the internet. He tells George to reduce his usage so that the mean daily time for May is 20% less than the current mean.

What is the maximum total time George can spend surfing the internet for the remaining 10 days of May? [3]

2. Over a period of time, a teacher recorded the number of time, x, each of the 20 students in the mathematics class was absent. The distribution was as follows.

Number of times absent, x	0	1	2	3	4	5	6	7	8	9	10	11 or more
Number of students, f	4	6	3	2	0	2	0	1	1	0	1	0

$$\sum f = 20, \quad \sum fx = 53, \quad \sum fx^2 = 299$$

(i) Illustrate the data using a suitable diagram. [2]

(ii) State the mode and find the median for the data set. [2]

(iii) Calculate the mean and the standard deviation of the data set. [3]

During this period of time, there were 30 mathematics lessons. The teacher needs to analyse the distribution of the number of times each student was *present* during the 30-lesson session.

(iv) Without creating a new frequency distribution, deduce values for the mean and standard deviation of the numbers of times students were present. Describe the shape of the new distribution. [3]

There are 12 boys and 8 girls in the class. The mean of the numbers of times boys were *absent* was 3, and the standard deviation was also 3.

(v) Show that the mean of the numbers of times girls were absent is 2.125. [2]

(vi) Find the standard deviation of the numbers of times girls were absent. [3]

3. Asha keeps hens so that she has a regular supply of fresh eggs. During March, the numbers of eggs she collected on each day were as follows.

4	5	8	4	6	5	6	7	7	10	11	18	12	9	5	6
5	6	4	5	5	6	7	8	8	13	10	11	14	9	10	

(i) Find the median number of eggs collected. [1]

(ii) Find the upper quartile, lower quartile and interquartile range. [3]

(iii) Draw a box and whisker plot for the data. [3]

(iv) Using your answers to part (ii), identify any outliers that are more than 1.5 times the interquartile range below the lower quartile or above the upper quartile. [3]

(v) Calculate the mean number of eggs laid per day. [2]

4. A motoring magazine carried out a survey of the value of petrol-driven cars that were five years old. In the survey, the value of each car was expressed as a percentage of its value when new. The results of the survey are summarised in the following table.

Percentage of original value (x)	Number of cars
$15 \le x < 20$	4
$20 \le x < 25$	12
$25 \le x < 30$	18
$30 \le x < 35$	13
$35 \le x < 40$	6
$40 \le x < 45$	5
$45 \le x < 55$	2

(i) Draw a histogram on graph paper to illustrate the data. [4]

(ii) Calculate an estimate of the median of the data. [2]

(iii) Calculate estimates of the mean and standard deviation of the data, giving your answers correct to 2 decimal places. Hence identify any outliers that are more than two standard deviations from the mean. [7]
[7]

A similar survey of 60 diesel driven cars produced a mean of 34.3% and a standard deviation of 11.7%.

(iv) Use these statistics to compare the values of petrol and diesel cars five years after they were purchased as new. [2]

Total 57 marks

Edexcel Data Assessment solns

Solutions to topic assessment

1. (i)

0	0 0 0 5
10	0 2 3 5 6 6 7 8
20	1 4 6 6
30	2 5
40	2
50	
60	1
70	3

Key: 20 | 6 means 26 minutes

The distribution has a positive skew.

[3]

(ii) Mode = 0

Median is 11th item = 17

$$\text{Mean} = \frac{462}{21} = 22$$

The mode is not very useful as it is not representative of the data. Most of the values appear only once.
The mean is skewed by two unusually large values.
The median is the most representative of the data.

[5]

(iii) $S_{xx} = \sum x^2 - n\bar{x}^2 = 17220 - 21 \times 22^2 = 7056$

$$\text{Standard deviation} = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{7056}{21}} = 18.33 \text{ minutes (2 d.p.)}$$

Outliers are more than 2 standard deviations from the mean
i.e. greater than $22 + 2 \times 18.33 = 59.66$
or less than $22 - 2 \times 18.33 = -14.66$
The outliers are therefore 61 and 73.

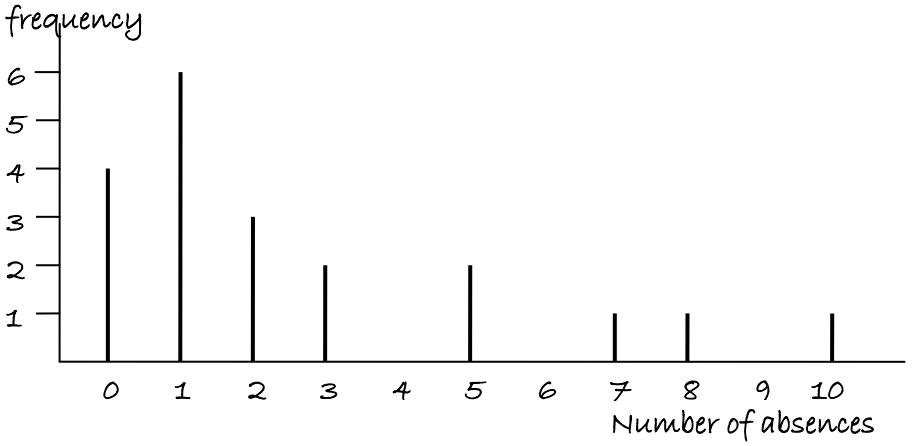
[4]

(iv) Target mean $= 0.8 \times 22 = 17.6$
Maximum total for the whole of May $= 17.6 \times 31 = 545.6$
Maximum total time available for the last 10 days of May
$= 545.6 - 462 = 83.6$ minutes

[3]

Edexcel Data Assessment solns

2. (i)



[2]

(ii) Mode = 1

Median is halfway between 10th and 11th values which are 1 and 2
Median = 1.5

[2]

(iii) $\text{Mean} = \dfrac{\sum fx}{n} = \dfrac{53}{20} = 2.65$ days

$S_{xx} = \sum fx^2 - n\bar{x}^2 = 299 - 20 \times 2.65^2 = 158.55$

$\text{Standard deviation} = \sqrt{\dfrac{S_{xx}}{n}} = \sqrt{\dfrac{158.55}{20}} = 2.82$ days (3 s.f.)

[3]

(iv) Number of times present = p, so $p = 30 - x$

$\bar{p} = 30 - \bar{x} = 30 - 2.65 = 27.35$

Standard deviation = 2.82 (3 s.f.) as before.

The new distribution is negatively skewed.

[3]

(v) Total number of absences in class = 53

Total number of times boys were absent $= 3 \times 12 = 36$

Total number of times girls were absent $= 53 - 36 = 17$

Mean number of times girls were absent $= \dfrac{17}{8} = 2.125$.

[2]

(vi) For boys: $3 = \sqrt{\dfrac{S_{xx}}{12}} \Rightarrow S_{xx} = 108$

$108 = \sum fx^2 - 12 \times 3^2 \Rightarrow \sum fx^2 = 216$

Total value of $\sum fx^2 = 299$

For girls: $\sum fx^2 = 299 - 207 = 83$

$S_{xx} = \sum fx^2 - n\bar{x}^2 = 83 - 8 \times 2.125^2 = 46.875$

$\text{Standard deviation} = \sqrt{\dfrac{S_{xx}}{n}} = \sqrt{\dfrac{46.875}{8}} = 2.42$ days (3 s.f.)

[3]

Edexcel Data Assessment solns

3. (i) In order:

4	4	4	5	5	5	5	5	5	6	6
6	6	6	7	7	7	8	8	8	9	9
10	10	10	11	11	12	13	14	18		

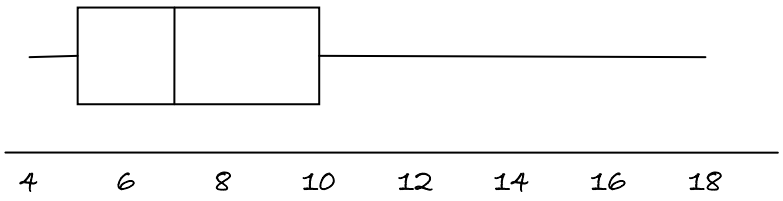
The median is the 16th item, which is 7.

[1]

(ii) There are 31 data items.
$\frac{1}{4} \times 31 = 7.75$, so the lower quartile is the 8th item. This is 5.
$\frac{3}{4} \times 31 = 23.25$, so the lower quartile is the 24th item. This is 10.
Interquartile range $= 10 - 5 = 5$.

[3]

(iii)



[3]

(iv) Outliers are more than 1.5 × the interquartile range above the upper quartile or below the lower quartile.
Outliers are therefore above $10 + 1.5 \times 5 = 17.5$ or below $5 - 1.5 \times 5 = -2.5$.
The only outlier is 18.

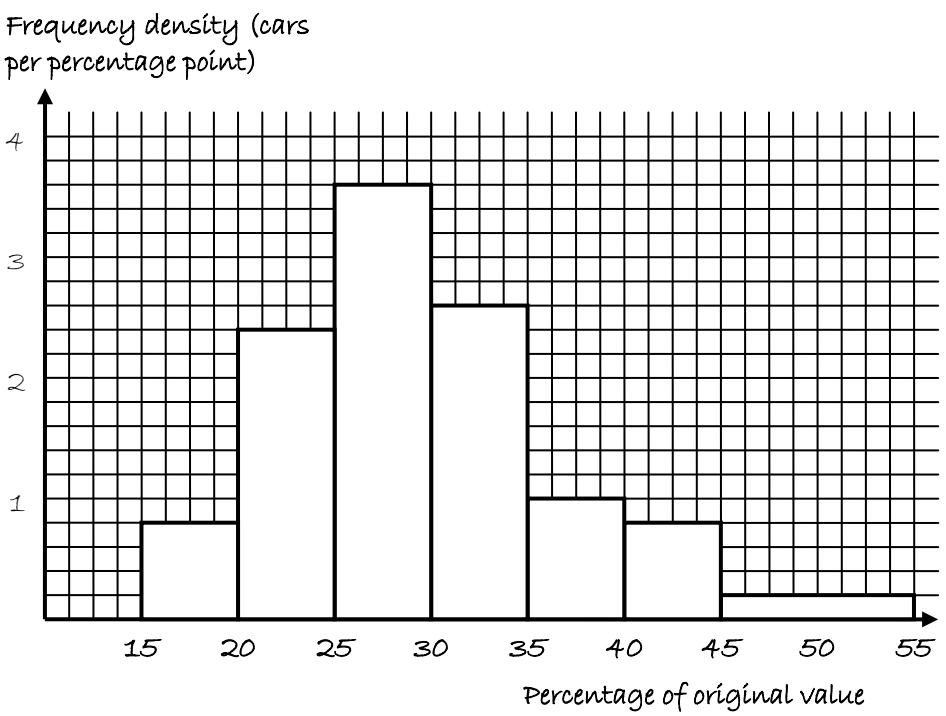
[3]

(v) Total number of eggs laid $= 244$.
Mean number of eggs laid $= \frac{244}{31} = 7.87$ (3 s.f.)

[2]

Edexcel Data Assessment solns

4. (i)



[4]

(ii) There are 60 cars, so the median is the 30th car. This is the 14th car out of the 18 cars in the 25 – 30 class interval.

Estimated median $= 25 + \frac{14}{18} \times 5 = 28.9$ (3 s.f.).

[2]

(iii)

Class interval	Mid-interval value, x	x^2	Frequency, f	fx	fx^2
15 – 20	17.5	306.25	4	70	1225
20 – 25	22.5	506.25	12	270	6075
25 – 30	27.5	756.25	18	495	13612.5
30 – 35	32.5	1056.25	13	422.5	13731.25
35 – 40	37.5	1406.25	6	225	8437.5
40 – 45	42.5	1806.25	5	212.5	9031.25
45 – 55	50	2500	2	100	5000
Totals			60	1795	57112.5

Estimated mean $= \frac{1795}{60} = 29.9\%$ (3 s.f.)

$S_{xx} = 57112.5 - 60 \times 29.9^2 = 3412.1$

Estimate of standard deviation $= \sqrt{\frac{3412.1}{60}} = 7.54$ (3 s.f.)

Outliers are more than 2 standard deviations from the mean.
Outliers are therefore above $29.9 + 2 \times 7.54 = 44.98$ or below

$29.9 - 2 \times 7.54 = 14.82$.
The two cars in the $45 - 55$ class interval are outliers.

[7]

(iv) The mean percentage value of the diesel-driven cars is higher than that for the petrol- driven cars, so on average cars with diesel engines hold their value better. The standard deviation for diesel-driven cars is higher, which means that there is greater variation in the values for diesel-driven cars.

[2]