Mixed practice 16

1 A sample of discrete data is drawn from a population and given as 115, 108, 135, 122, 127, 140, 139, 111, 124.

Find:

a the interquartile range

b the mean

c the variance.

2 $\Sigma(x-\bar{x})^2 = 12$ and $n = 3$. Find the value of the standard deviation.

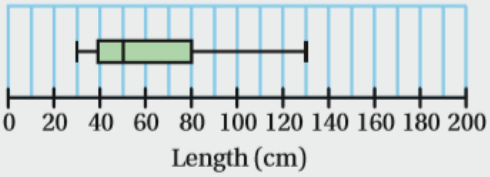
3 A student takes the bus to school every morning. She records the length of the time, in minutes, she waits for the bus on 12 randomly chosen days. The data set is summarised by $\Sigma x = 49$ and $\Sigma x^2 = 305.7$.

a Find the mean.

b Find the variance.

4 This box-and-whisker plot shows the lengths of corn snakes (x) in cm.

Use the definition that an outlier is more than 1.5 IQR from the closest quartile to find the range of values that would be outliers. Hence show that there are no outliers for this data.



5 **a** Use a counter example with two data items to show that $\overline{x^2} = \bar{x}^2$ is not always true.

b If $\overline{x^2} = \bar{x}^2$ find the standard deviation of the data. Hence provide an example of a set of data containing two items which has $\overline{x^2} = \bar{x}^2$.

6 40 people were asked to guess the length of a certain road. Each person gave their guess, l km, correct to the nearest kilometre. The results are summarised below.

l	10–12	13–15	16–20	21–30
Frequency	1	13	20	6

i **a** Use appropriate formulae to calculate estimates of the mean and standard deviation of l.

b Explain why your answers are only estimates.

ii A histogram is to be drawn to illustrate the data. Calculate the frequency density of the block for the 16–20 class.

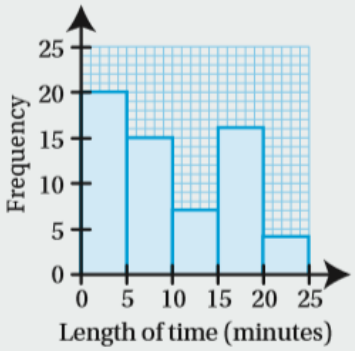
iii Explain which class contains the median value of l.

iv Later, the person whose guess was between 10 km and 12 km changed his guess to between 13 km and 15 km. Without calculation, state whether the following will increase, decrease or remain the same:

a the mean of l, **b** the standard deviation of l.

© OCR, AS GCE Mathematics, Paper 4732, January 2010

7 **a** From this chart calculate an estimate of the mean and the standard deviation of the data.



b The chart shows the distribution of the length of time, in minutes, Tom has to wait for a bus in the morning. Find the probability that, on a randomly selected day, he has to wait between 15 and 25 minutes.

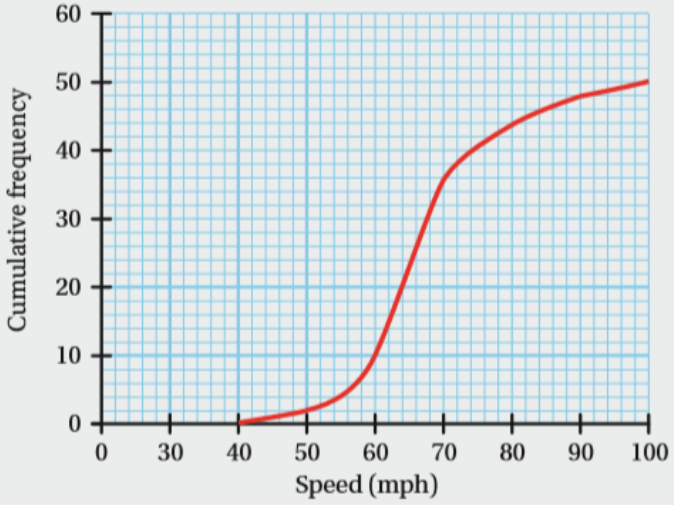
8 Jenny must sit 4 papers for an exam. All papers have an equal weight when their marks are combined. The mean of the first 3 papers Jenny has sat is 72% with a standard deviation of 8%.

a If she wants to get a mean of 75% overall what is the lowest percentage she can get in her fourth paper?

b What is the highest possible mean she can get?

c If she does get the highest possible mean, what is her new standard deviation?

9 This cumulative frequency diagram gives the speed, v, of 50 cars in mph as they travel past a motorway checkpoint.



a From the diagram find the median speed.

b Any car travelling above 75 mph will be stopped by the police. How many of these cars will be stopped?

c The middle 50% of speeds lie between a and b where $a < b$. Find the values of a and b.

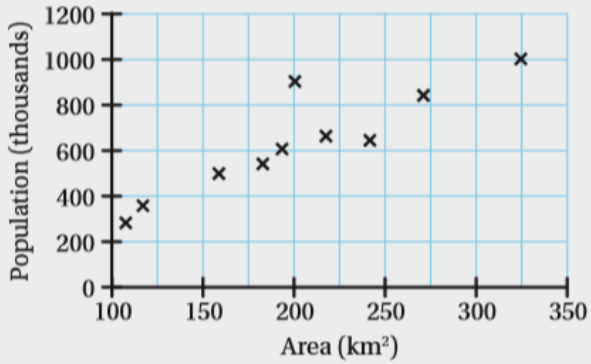
d Copy and complete the following frequency table.

v	Frequency
$40 < v \leqslant 50$	2
$50 < v \leqslant 60$	9
$60 < v \leqslant 70$	
$70 < v \leqslant 80$	
$80 < v \leqslant 90$	
$90 < v \leqslant 100$	

e Hence estimate the mean and the standard deviation in the speeds.

f Use the definition of an outlier as anything more than 2 standard deviations from the mean to show that some of the observed speeds are outliers. Decide, with justification, whether these speeds need to be removed for a valid analysis.

10 A geographer is studying data on the area (A) and population (P) of various cities in a country. He displays his data on a scatter diagram.



The mean population of the cities studied is 604 thousand.

a What is the advantage of displaying the data on a scatter diagram rather than two histograms?

b Describe the correlation of the data shown by the scatter graph.

c One of the cities completely fills an island so that as its population has grown it has not been able to expand. What is the area of the island?

d In the rest of the analysis, the city on the island is removed from the data. What effect does this have on the mean population of the cities studied? Explain your answer.

e The regression line of the remaining cities has equation $P = 3.05A - 3.26$.
Interpret the value 3.05 in the context of the cities.

f The capital of the country has an area of 1600 km^2. Why would it not be valid to use the regression line to predict the population of the capital?

11 The test marks of 14 students are displayed in a stem-and-leaf diagram, as shown below.

0	
1	2 6
2	1 3 5
3	w x 4 8 y z
4	6 7 7

Key: 1|6 means 16 marks

i Find the lower quartile.

ii Given that the median is 32, find the values of w and x.

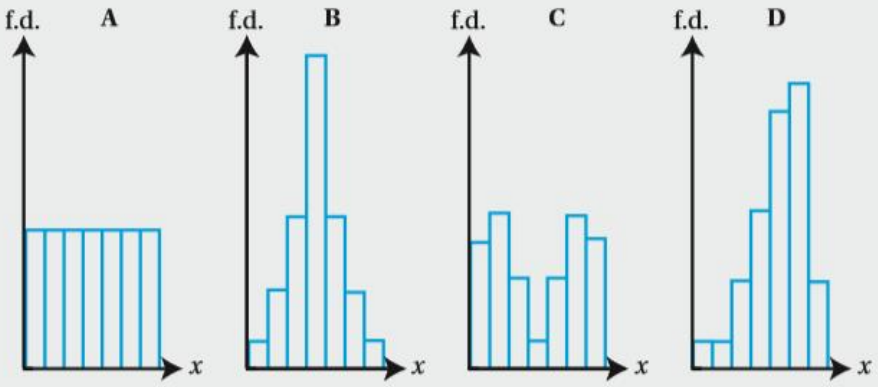
iii Find the possible values for the upper quartile.

iv State one advantage of a stem-and-leaf diagram over a box-and-whisker plot.

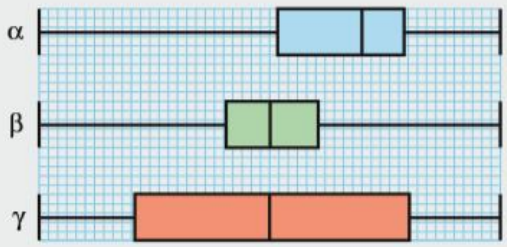
v State one advantage of a box-and-whisker plot over a stem-and-leaf diagram.

12 The four populations A, B, C and D are the same size and have the same range.

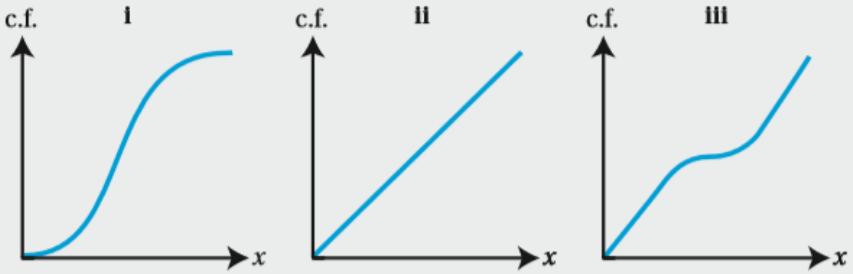
Histograms for the four populations are shown.



a Each of these three box-and-whisker plots corresponds to one of the four populations. Write the letter of the correct population for each of α, β and γ.



b Each of these three cumulative frequency diagrams corresponds to one of the four populations. Write the letter of the correct population for each of **i**, **ii** and **iii**.



13 The mean of a set of 10 data items is 115 and the variance is 154. Another piece of data is discovered and the new mean is 114. What is the new variance?

14 If the sum of 20 pieces of data is 1542, find the smallest possible value of Σx^2.

Mixed practice 16

1 **a** 24 **b** 125 **c** 124

2 2

3 **a** 4.08 minutes **b** 8.80 minutes^2

4 $x > 140$ or $x < -20$

5 **a** e.g. 1, 0 **b** 0, e.g. 1, 1

6 **a** **i** Mean = 17.7, SD = 3.84
ii Exact data values unknown
b 4 **c** 16–20
d **i** Increase **ii** Decrease

7 **a** $\bar{t} = 10.0, \sigma = 6.71$ **b** 0.323

8 **a** 84% **b** 79% **c** 14.0%

9 **a** 66 mph **b** 11
c $a = 61$ mph, $b = 73$ mph
d 224, 7, 3, 5 **e** $\bar{x} = 67.5, \sigma = 11.3$
f Do not remove – they are extreme but no reason to believe they are incorrect

10 **a** Shows relationships
b Strong positive
c 200 km^2
d It will go down
e the increase in population for each area increase of 1 km^2
f Extrapolating from the data

11 **a** 3 **b** $x = w = 30$
c 38, 39 **d** Shows all data
e Easier to find median and quartiles

12 **a** α:D, β:B, γ:C
b **i** B **ii** A **iii** C

13 150

14 118 888.2