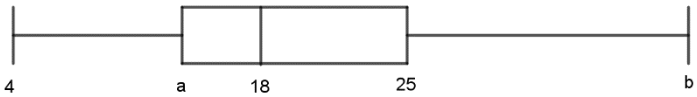
Statistics Test Week 34

Non-Calculator Questions:

1. A school administrator wants to research the views of the students on the school lunch service by asking a sample of 60 students to fill out a questionnaire. There are 400 students in the lower school and 200 students in the upper school.
   (a) Describe how she should select the students in the sample by: [3]
      (i) quota sampling;
      (ii) stratified random sampling.
   (b) Describe one advantage and one disadvantage of using quota sampling in this context. [2]

   [Total 5 marks]

2. The following diagram is a box and whisker diagram for a set of data.



   The interquartile range of the data is 10 and the range is 35.
   (a) Write down the median value. [1]
   (b) Find the values of (i) a
      (ii) b [2]
   (c) Show, with an appropriate calculation, whether the data contains an outlier. [2]

   [Total 5 marks]

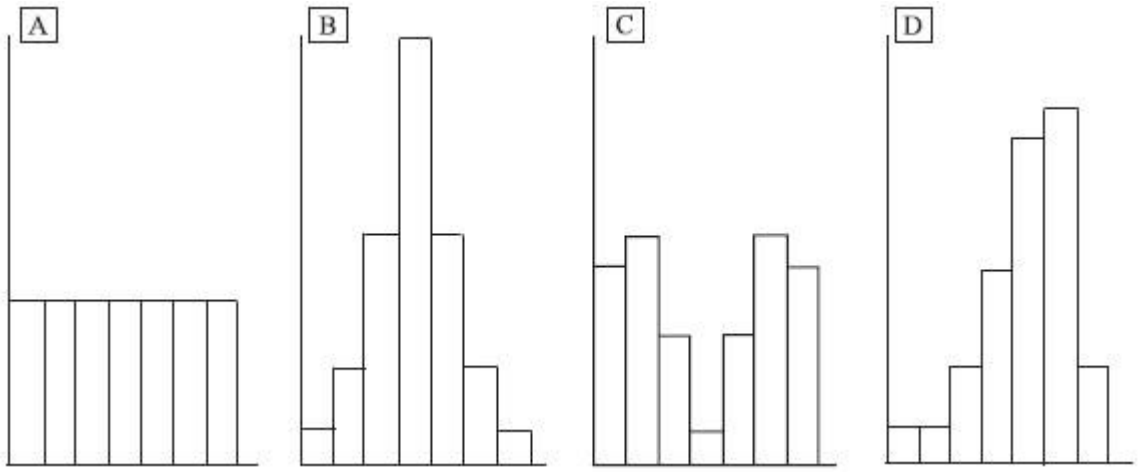
3. A class completes a test which is out of 60 marks. The teacher calculates the mean score as 42 and the variance as 25. She decides to convert the scores into percentages by subtracting 10 marks from all the scores, and then multiplying by 2.
   (a) Find the new values for (i) the mean [2]
      (ii) the variance [2]

   An extra student then completes the test late, and scores 30 out of 60.
   (b) Giving your reasons, state whether this will cause an increase or a decrease in
      (i) the mean [1]
      (ii) the variance [1]

   [Total 6 marks]

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

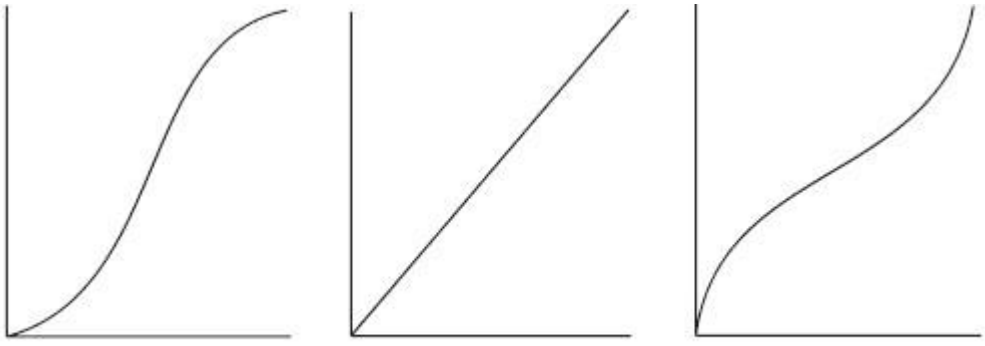
Frequency histograms for the four populations are given below.



(a) Each of the three box and whisker plots below corresponds to one of the four populations. Write the letter of the correct population under each plot.

....

(b) Each of the three cumulative frequency diagrams below corresponds to one of the four populations. Write the letter of the correct population under each diagram.



....

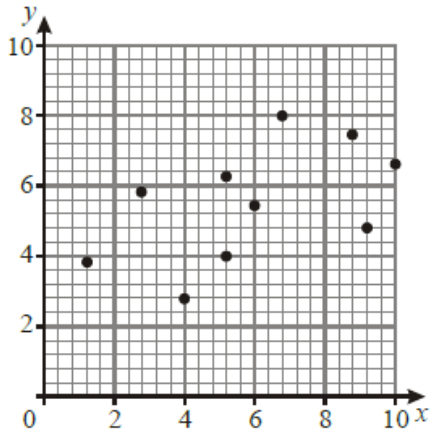
[Total 6 marks]

5. Statements i, ii, iii, iv and v represent descriptions of the correlation between two variables.

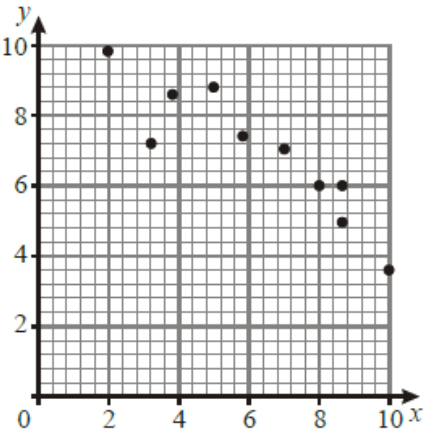
i High positive linear correlation
ii Low positive linear correlation
iii No correlation
iv Low negative linear correlation
v High negative linear correlation

Which statement best represents the relationship between the two variables shown in each of the scatter diagrams below?

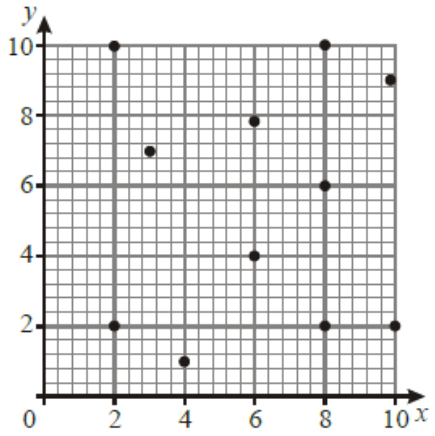
(a)



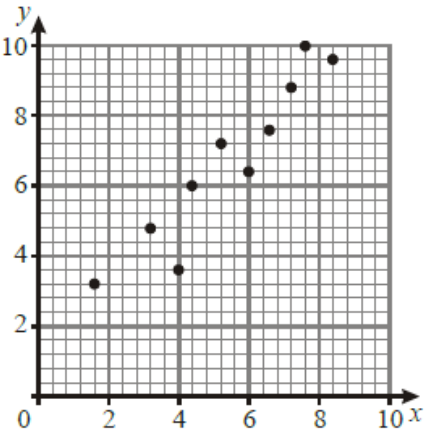
(b)



(c)



(d)



(Total 4 marks)

6. (a) The sum of five test scores, and the sum of their squares, are given by $\sum x = 103$ and $\sum x^2 = 2614$. Giving your answers to 3 significant figures, find:

(i) the mean score; [1]

(ii) the standard deviation. [1]

(b) A different set of five scores had a mean of 22.4 and a standard deviation of 5.22. Find the values of:

(i) $\sum x$ [1]

(ii) $\sum x^2$ [1]

[Total 4 marks]

7. The following table shows the times, to the nearest minute, taken by 100 students to complete a mathematics task.

Time (t) minutes	11–15	16–20	21–25	26–30	31–35	36–40
Number of students	7	13	25	28	20	7

(a) Construct a cumulative frequency table. (Use upper class boundaries 15.5, 20.5 and so on.)

[2]

(b) On graph paper, draw a cumulative frequency graph, using a scale of 2 cm to represent 5 minutes on the horizontal axis and 1 cm to represent 10 students on the vertical axis.

[3]

(c) Use your graph to estimate:

(i) the number of students who completed the task in less than 17.5 minutes;

(ii) the time it will take for $\frac{3}{4}$ of the students to complete the task.

[2]

[Total 7 marks]

Calculator Questions:

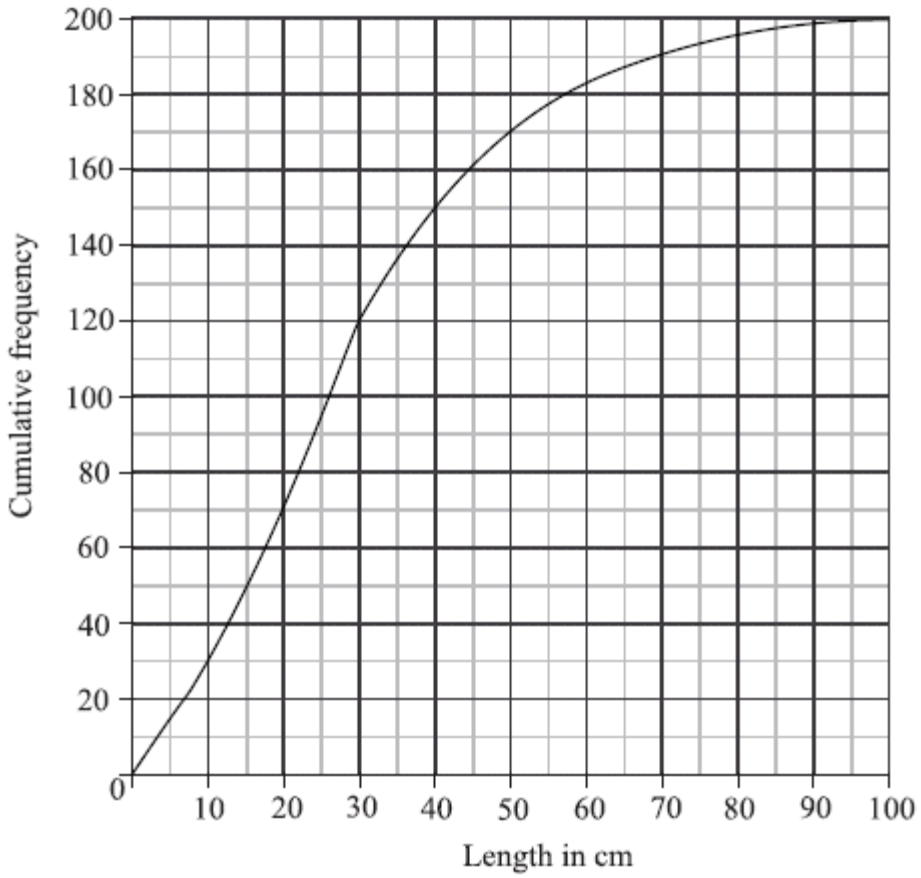
8. A fisherman catches 200 fish to sell. He measures the lengths, l cm of these fish, and the results are shown in the frequency table below.

Length l cm	$0 \le l < 10$	$10 \le l < 20$	$20 \le l < 30$	$30 \le l < 40$	$40 \le l < 60$	$60 \le l < 75$	$75 \le l < 100$
Frequency	30	40	50	30	33	11	6

(a) Calculate an estimate for the standard deviation of the lengths of the fish.

[3]

(b) A cumulative frequency diagram is given below for the lengths of the fish.



Use the graph to answer the following.

(i) Estimate the interquartile range.
(ii) Given that 40% of the fish have a length more than k cm, find the value of k.

[6]

In order to sell the fish, the fisherman classifies them as small, medium or large.

Small fish have a length less than 20 cm.
Medium fish have a length greater than or equal to 20 cm but less than 60 cm. Large fish have a length greater than or equal to 60 cm.

(c) Write down the probability that a fish is small. [2]

[11 marks]

9. The Type Fast secretarial training agency has a new computer software spreadsheet package. The agency investigates the number of hours it takes people of varying ages to reach a level of proficiency using this package. Fifteen individuals are tested and the results are summarised in the table below.

Age (x)	32	40	21	45	24	19	17	21	27	54	33	37	23	45	18
Time (in hours) (y)	10	12	8	15	7	8	6	9	11	16	13	13	9	17	5

(a) (i) Calculate the Pearson product-moment correlation coefficient r for this data.

[4]

(ii) What does the value of the correlation coefficient suggest about the relationship between the two variables?

[1]

(b) Find the equation of the regression line for y on x in the form $y = ax + b$.

[3]

(c) Use your equation for the regression line to predict:

(i) the time that it would take a 33-year-old person to reach proficiency, giving your answer correct to the nearest hour;

[2]

(ii) the age of a person who would take 8 hours to reach proficiency, giving your answer correct to the nearest year.

[2]

[Total 12 marks]