AQA

Please write clearly in block capitals.

Centre number | | | | | | Candidate number | | | | |

Surname _____

Forename(s) _____

Candidate signature _____

I declare this is my own work.

GCSE STATISTICS

H

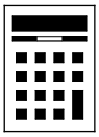
Higher tier Paper 2

Time allowed: 1 hour 45 minutes

Materials

For this paper you must have:

- a calculator
- mathematical instruments.



Instructions

- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Do not write outside the box around each page or on blank pages.
- If you need extra space for your answer(s), use the lined pages at the end of this book. Write the question number against your answer(s).
- Do all rough work in this book. Cross out any work you do not want to be marked.

Information

- The marks for questions are shown in brackets.
- The maximum mark for this paper is 80.
- You may ask for more answer paper and graph paper. These must be tagged securely to this answer booklet.

For Examiner's Use	
Question	Mark
1-4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
TOTAL	

Answer **all** questions in the spaces provided.

Do not write outside the box

1 A fair coin is tossed four times.

Circle the probability of getting 'tails' on all 4 tosses.

[1 mark]

$\frac{1}{2}$ $\frac{1}{4}$ $\frac{1}{8}$ $\frac{1}{16}$

2 Which one of these is **not** a measure of spread?

Circle your answer.

[1 mark]

interdecile range interpercentile range

standard deviation skewness

Do not write outside the box

3 Olivia wants to obtain a systematic sample of size 40 from a population of 2000 competitors in a race.

She numbers the competitors from 1 to 2000 and uses, as a starting point, competitor number 11

Circle the correct competitor number of the next person in her sample.

[1 mark]

12 21 51 61

4 Which statistical term means 'the extent to which something gives results that are consistent'?

Circle your answer.

[1 mark]

reliability validity

standard deviation unbiased

4

Turn over for the next question

Do not write outside the box

5 BMI stands for Body Mass Index and is based on your height and mass.

5 (a) Louise and William have this hypothesis,

"Our classmates have a lower BMI than other Year 11 students in the country."

They collect height and mass data from some of their classmates.

Gender	Height (m)	Mass (kg)
Female	1.38	51
Female	1.27	48
Female	1.31	39
Female	1.42	45
Female	1.52	55
Male	1.68	71
Male	1.60	65
Male	1.74	82
Female	1.38	51
Male	1.69	78

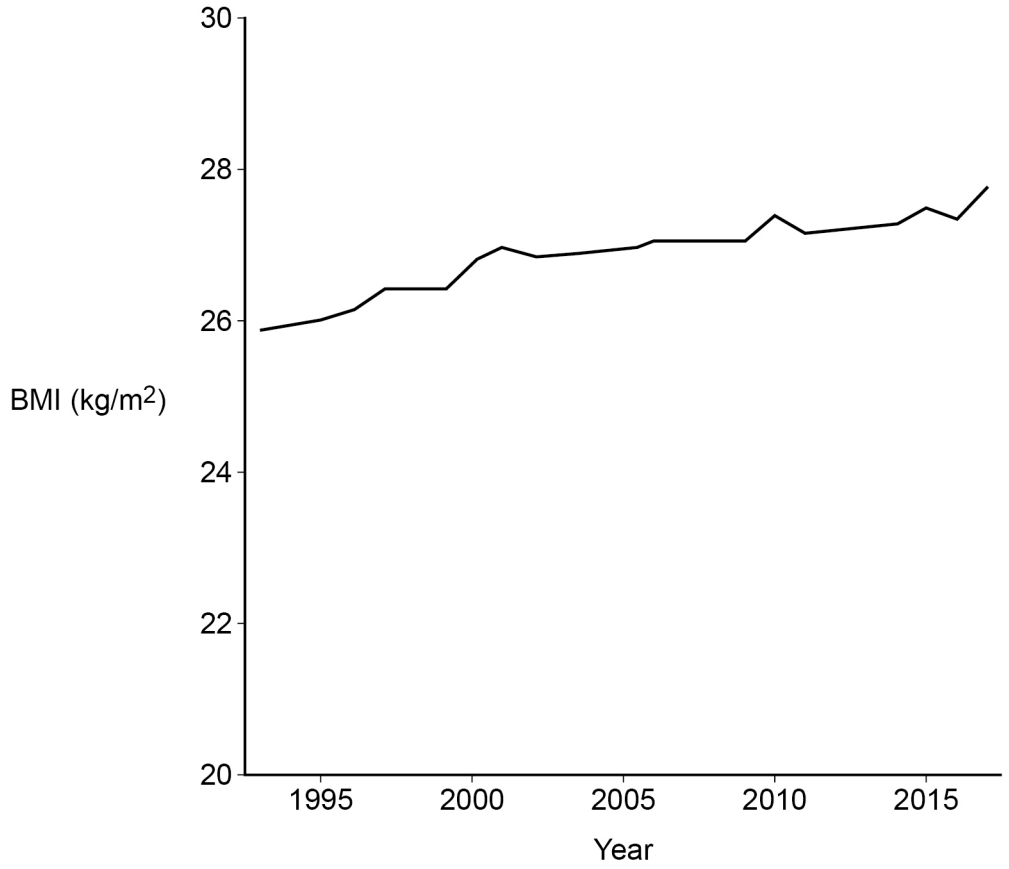
Give **one** criticism of the collected data.

[1 mark]

Do not write outside the box

5 (b) The graph shows the average adult BMI over recent years.

Adults' Average BMI (body mass index)



Source: NHS Digital

A healthy BMI is considered to be between 18.5 and 24.9

Make **two** distinct comments about the graph.

[2 marks]

1 _____

_____

_____

2 _____

_____

_____

3

Turn over ►

Do not write outside the box

6 HS2 (High Speed 2) is a faster train service that will link major cities in England.

Tom believes most people are against HS2 because it affects countryside and housing along its routes.

He decides to gather opinions about HS2.

6 (a) Write down a hypothesis Tom could use for his study.

[1 mark]

6 (b) Here is one of the questions from Tom's study.

How old are you?

Tick (✓) a box.

☐ under 21 ☐ 21 – 50 ☐ 51 – 60 ☐ 61 – 70

Write down **two** different problems with this question.

[2 marks]

Problem 1 _____

Problem 2 _____

6 (c) Here is an open question from Tom's study.

How much do you earn? £ _____

Write down a problem with this question.

[1 mark]

Do not write outside the box

6 (d) Tom reads that HS2 will link 29 stations.

He decides to take a random sample of 5 of the stations where he can ask people for their opinions.

Briefly describe a way Tom could achieve this.

[2 marks]

6 (e) One of the stations Tom gets in his random sample is Manchester Piccadilly.

To find opinions, he goes there one Saturday afternoon and asks his questions to the first 100 people who will answer.

6 (e) (i) Name this sampling method.

[1 mark]

Answer _____

6 (e) (ii) What is good about Tom finding opinions in this way?

[1 mark]

6 (e) (iii) What is not so good about Tom finding opinions in this way?

[1 mark]

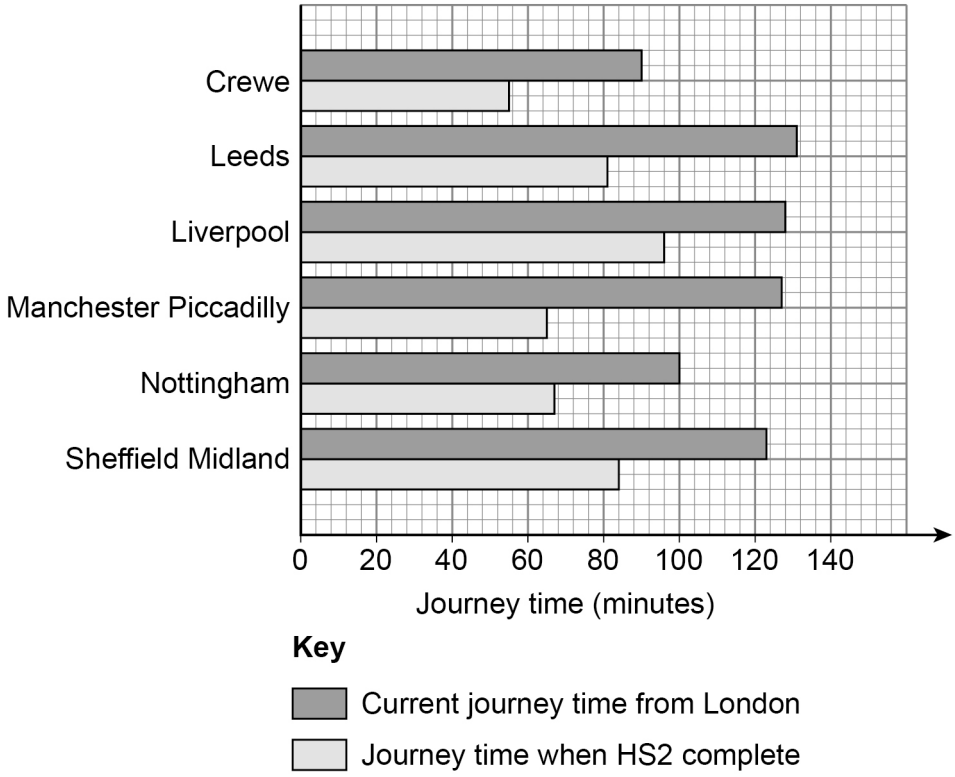
6 (e) (iv) Give a reason why Tom should also find opinions of people where HS2 will **not** have a station.

[1 mark]

Turn over ►

Do not write outside the box

6 (f) The Department of Transport produced this graph about HS2 in 2016 showing how journey times might change when HS2 is complete.



6 (f) (i) Write down the name of this type of diagram.

[1 mark]

Answer _____

Do not write outside the box

6 (f) (ii) Li Na says that the journey time between London and Manchester Piccadilly will be reduced by about an hour.

Is Li Na correct?

Tick (✓) a box.

Yes ☐ No ☐

Show working to justify your answer.

[2 marks]

6 (g) This table also shows information about reduced journey times from London.

London to:	Current journey time (mins)	Journey time after HS2 (mins)	Reduction time in minutes (% reduction)
Chesterfield	109	75	34 (31.2%)
Crewe	90	55	35 (38.8%)
Edinburgh	263	218	45 (17.1%)
Glasgow	272	218	54 (19.9%)
Liverpool	128	96	32 (25.0%)
Newcastle	172	139	33 (19.2%)
Preston	128	84	

Work out the missing time and percentage in the Preston row.

[3 marks]

Answer _____ mins _____ %

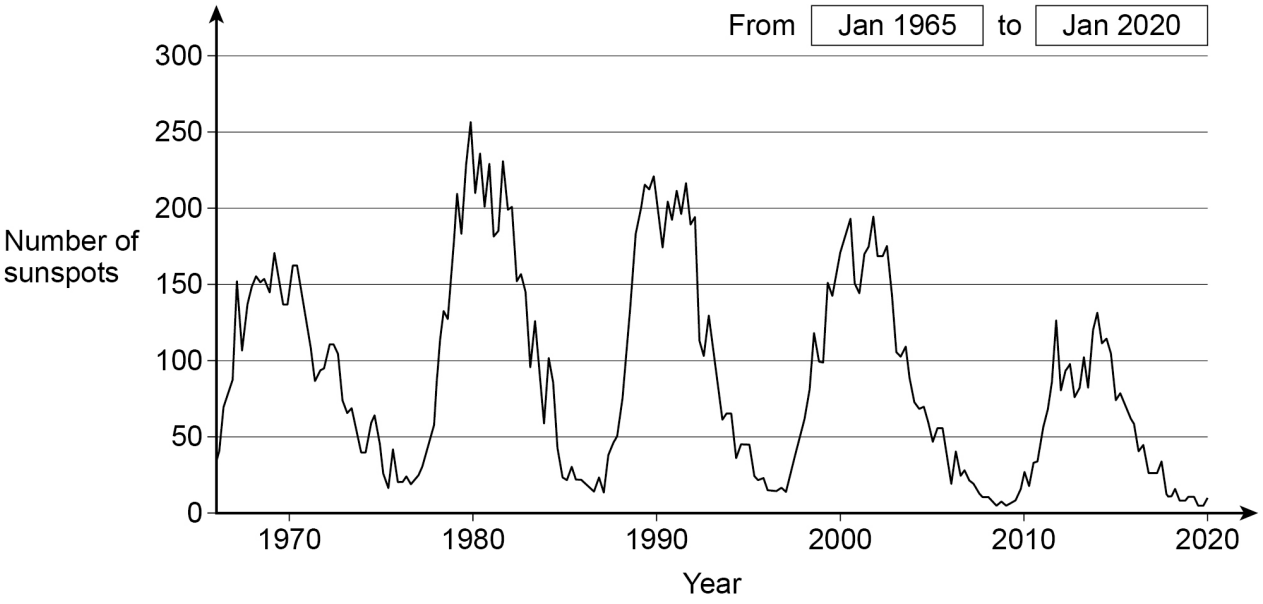
16

Turn over ►

Do not write outside the box

7 Sunspots are dark marks on the sun's surface which can affect things on Earth such as radio signals.

The number of sunspots recorded **monthly** from 1965 – 2020 is shown in the time series graph below.



Source: SpaceWeatherLive

7 (a) Estimate the year when the most sunspots were recorded for 1965 – 2020 and work out an estimate for the number of sunspots in that **year**.

[2 marks]

Year _____ **Number of sunspots** _____

7 (b) Describe **one** feature of the data.

[1 mark]

_____

_____

_____

7 (c) There are variations in the number of sunspots seen per month throughout the year.

How could you smooth out these variations?

[2 marks]

_____

_____

_____

_____

5

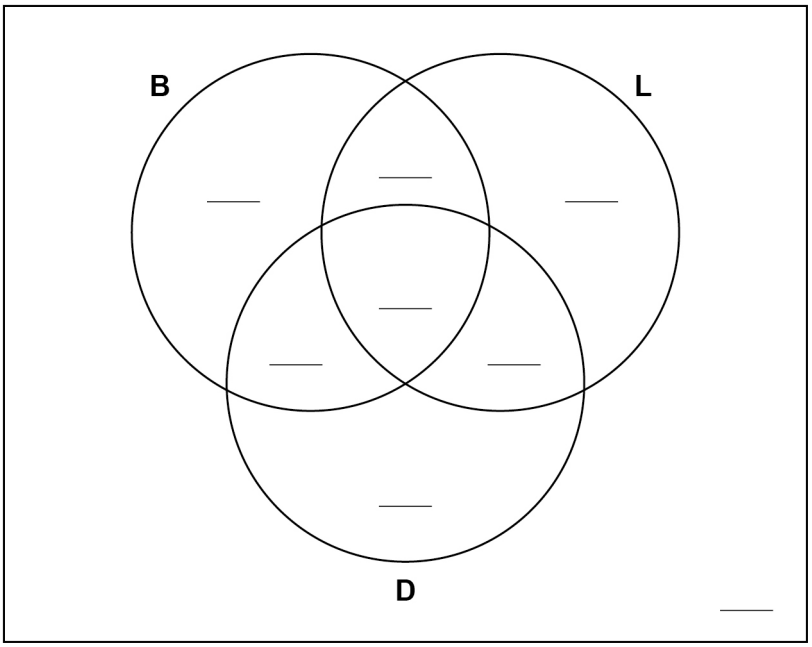
Do not write outside the box

8 100 people were asked whether they had hot food at breakfast (**B**), lunch (**L**) or dinner (**D**) yesterday,

- 54 **only** had hot food for dinner
- 1 person didn't have hot food for any meal
- no-one had hot food for all three meals
- a **total** of 4 people didn't have hot food for dinner
- the number of people who had hot breakfast **and** hot dinner is equal to the number of people who had hot lunch **and** hot dinner.

Complete the Venn diagram with a **possible** set of correct values.

[5 marks]



5

Turn over for the next question

Do not write outside the box

9 There are 7 players who can play for a snooker team,

Micky, Katie, Niles, Tommo, Paul, Jonno and Emma.

Each week **four** players are needed to make up the team.

9 (a) One week, Micky and Katie are chosen for the team and the other two players are chosen at random.

What is the probability that Niles is also in the team?

[3 marks]

Answer _____

9 (b) Paul is trying to work out the chances he will win a game.

He has the following sets of data available to him.

A How many of the last 5 games he won.

B How many of the last 20 games he won.

C How many of the last 100 games he won.

D How many, of all the games he's ever played, he won.

9 (b) (i) Give a statistical reason for using option **D**.

[1 mark]

Do not write outside the box

9 (b) (ii) Give a reason for choosing **one** of the other options.

State which option you choose.

[1 mark]

Option _____

Reason _____

_____

_____

5

Turn over for the next question

Do not write outside the box

10 A rare plant is found in only two forests in Scotland.

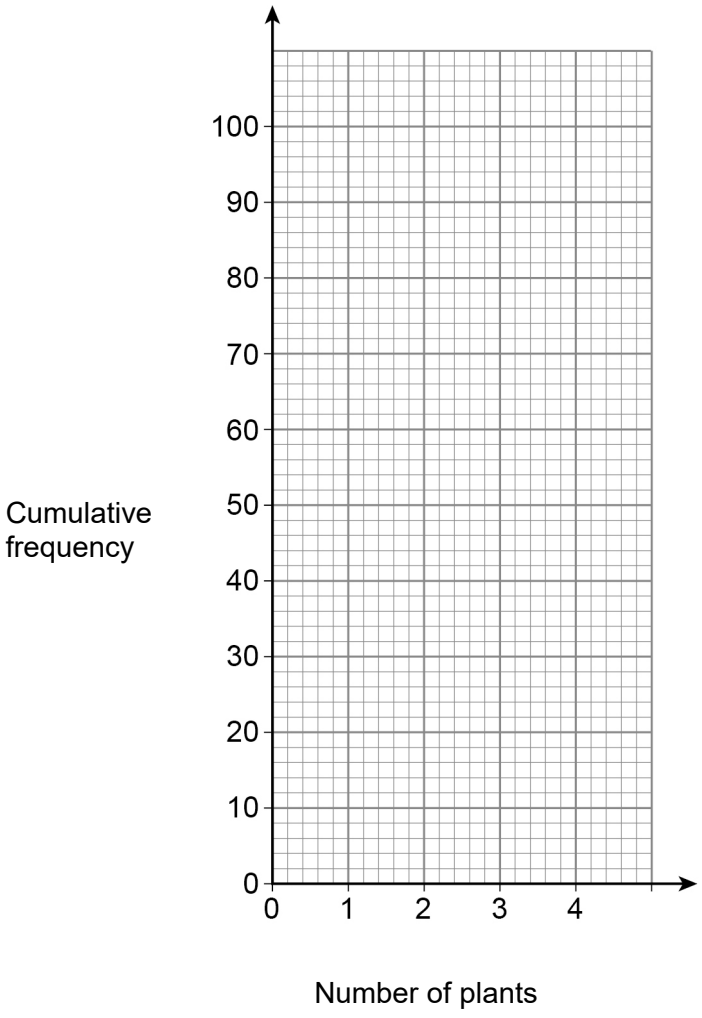
Loren is counting their numbers in same-sized areas in each of the forests.

10 (a) In forest **A**, the number of the plants in each of 100 areas is summarised below.

Number of plants	Frequency
0	34
1	39
2	12
3	10
4	5

10 (a) (i) Draw a cumulative frequency step polygon for forest **A** on the grid below.

[3 marks]



Do not write outside the box

10 (a) (ii) Give a reason why a cumulative frequency step polygon is appropriate for these data.

[1 mark]

10 (b) In forest **B**, the following information is obtained for 100 areas,

- the median number of plants = 2
- the interdecile range between the 1st and 9th deciles = 1

Use this information and the information in **part (a)** to compare, in context, the number of plants in forest A and forest B.

[4 marks]

Question 10 continues on the next page

Do not write outside the box

10 (c) A river in forest **B** is also home to a species of vole.

Loren wants to estimate the population of voles.

Here is her method,

1. She catches 30 voles and marks them with a dye.
2. She releases the voles on March 1st.
3. Loren returns to the forest on September 1st.
4. She catches 5 voles and sees how many are marked with the dye.

Criticise steps **3** and **4** in Loren's method.

[2 marks]

Step 3 criticism _____

_____

_____

Step 4 criticism _____

_____

_____

10

Do not write outside the box

Turn over for the next question

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

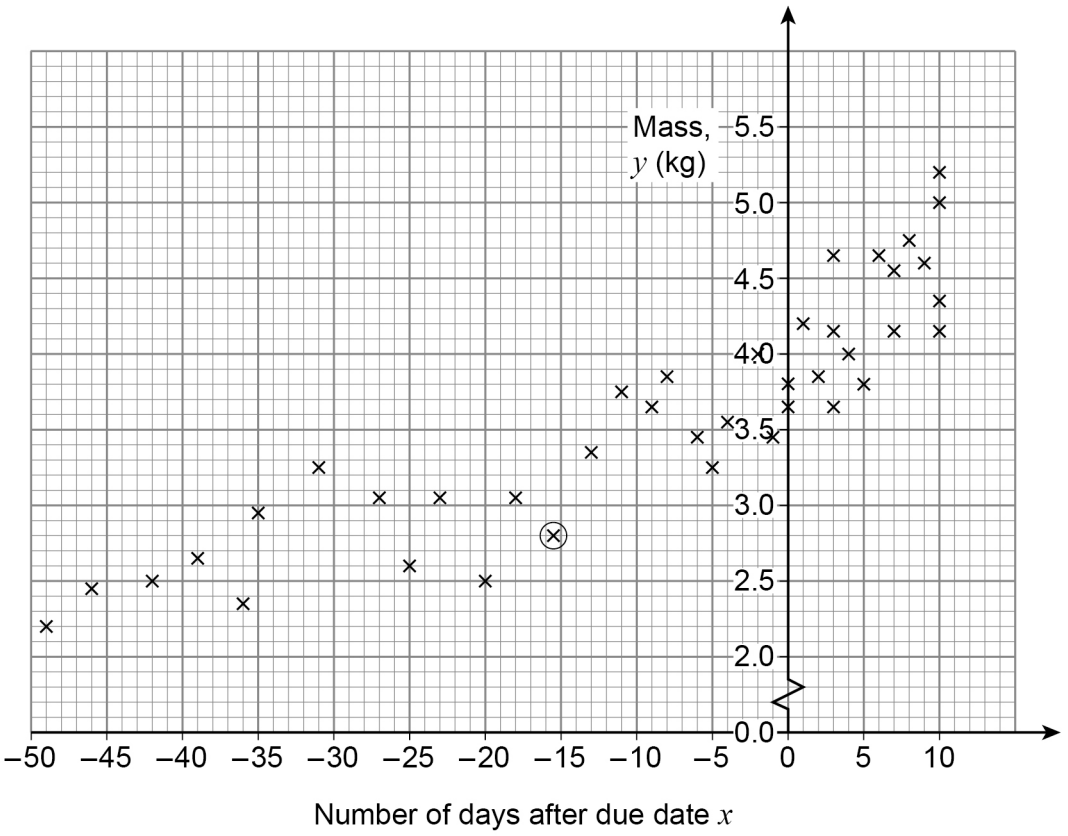
Do not write outside the box

11 The due date of a baby is the date on which it is expected to be born.

The scatter diagram shows the mass of 40 new-born babies **born** on March 20th 2021 plotted against the number of days the babies were born after their due date in whole days.

For example,

Value of number of days after due date	Interpretation
–10	Baby born 10 days **before** due date
5	Baby born 5 days **after** due date



11 (a) How many of **these** babies had a due date of March 10th 2021?

Circle your answer.

1 2 3 4

[1 mark]

Do not write outside the box

11 (b) The circled point is an incorrect plot.

Give a reason how you know this.

[1 mark]

11 (c) The error is corrected.

The equation of the line of best fit for the data is $y = 4.01 + 0.04x$

11 (c) (i) Interpret the value 4.01 in the context of this scatter graph.

[1 mark]

11 (c) (ii) Interpret the value 0.04 in the context of this scatter graph.

[1 mark]

11 (c) (iii) Draw the line of best fit on the graph.

[2 marks]

Question 11 continues on the next page

Do not write outside the box

11 (d) It is later found that there were two more babies born on March 20th.

- Sam was born 15 days before his due date.
- Nim was born 15 days after her due date.

Discuss the valid use of the scatter graph to estimate the birth mass of each of these babies.

Where an estimate is possible, explain your method and give the value.

[3 marks]

Sam

Nim

9

Do not write outside the box

Turn over for the next question

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

Do not write outside the box

12 A small factory produces windows.

- Each window has a 4% chance that it is damaged.
- Damaged windows cannot be sold.

On average, **one** window costs £50 to produce and is sold for £300.

12 (a) Each year, the factory produces 800 windows.

Work out the expected profit made from window sales.

[5 marks]

Answer £ _____

Do not write outside the box

12 (b) The quality control manager samples the next 5 windows produced to look for damage.

12 (b) (i) Comment on this data selection method.

[1 mark]

12 (b) (ii) Calculate the probability that **exactly one** of these 5 windows is damaged.

Assume that the number of damaged windows follows a Binomial distribution.

[3 marks]

Answer _____

9

Turn over for the next question

Do not write outside the box

13 A teacher marks some mock GCSE Statistics exam papers.

Each student sits two papers.

The teacher wants to see how the marks scored on the two papers are related.

She decides to calculate the value of Spearman's rank correlation coefficient (SRCC) using the formula

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

5 students sit the mock papers and the value of r_s = 0.8

13 (a) Interpret the value of SRCC in this context.

[1 mark]

13 (b) A sixth student now sits the papers.

Both papers are ranked 6th.

13 (b) (i) How will including the additional pair of papers change the value of SRCC?

[1 mark]

Do not write outside the box

13 (b) (ii) Calculate the new value of SRCC for all 6 students.

[4 marks]

Answer _____

6

Turn over for the next question

Do not write outside the box

14 The number of days between snow events (a day when it snows) in a Scottish town during winter 2020 – 2021 was recorded.

The data has been **ordered by size**.

0 0 0 1 1 3 4 4 10 28

For example, a value of 0 indicates it snowed on two consecutive days.

14 (a) Rhona says that it snowed on three consecutive days.

Is she correct?

Tick (✓) a box.

Yes ☐ No ☐ Cannot tell ☐

Give a reason for your answer.

[1 mark]

Do not write outside the box

14 (b) Use the formulae

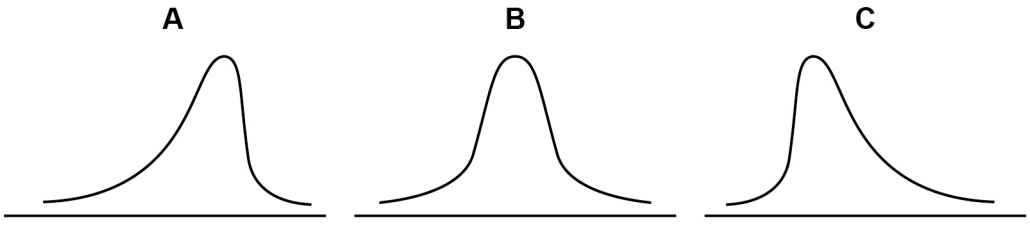
$$\text{skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}} \quad \text{and standard deviation} = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

to show that the skew of the data is +1.14 (to 2 decimal places).

[6 marks]

14 (c) Circle the letter of the diagram that shows data with a positive skew.

[1 mark]



8

END OF QUESTIONS

Do not write outside the box

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

Do not write outside the box

Question number	**Additional page, if required.** **Write the question numbers in the left-hand margin.**

Do not write outside the box

Question number	**Additional page, if required.** **Write the question numbers in the left-hand margin.**

Do not write outside the box

Question number	**Additional page, if required.** **Write the question numbers in the left-hand margin.**

Do not write outside the box

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED