

**Statistics GCSE****Paper 2**

2025

Edexcel Foundation

Variant 4

1ST0/2F

**Instructions**

- Write all answers in the spaces provided.
- Answer all questions.
- You must show all your working.
- There may not be enough space to show all your working out.

**Information**

- This is a practise paper to aid your revision for your exams.
- This site, and all that work on it, have no affiliation or relationship with any exam board.
- This site is not endorsed by any company or charity, unless we state otherwise.

**Copyright**

- Sharing this PDF is strictly forbidden unless you have the author's permission.
- This paper is only authorised to be used by the person who has bought it.
- You may not store this online or any shared area such as an intranet.
- Please contact us if you have any queries.

**Advice**

- You can get support for all these questions at our website: [www.statsgcse.com](http://www.statsgcse.com)
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

1 Ms. Rivera asked the students in her art class which colour is their favourite from the choices Blue, Red, Green and Yellow.

Their answers are shown below:

Blue Green Blue Red Yellow Blue  
Blue Red Green Yellow Blue Red Blue  
Blue Blue Green Blue Yellow Red Red

(a) Complete the tally and frequency columns on the frequency table for the data.

(2 marks)

Colour	Tally	Frequency
Blue		
Red		
Green		
Yellow		

(b) Find the number of students in Ms. Rivera's art class.

(1 mark)

\_\_\_\_\_

(c) Compare the amount of students who chose Blue with the amount who chose Yellow.

(1 mark)

\_\_\_\_\_  
\_\_\_\_\_

(d) Ms. Rivera finds the mode from the data.

Give a reason why the mode is an appropriate average to use in this situation.

(1 mark)

---

---

---

---

(e) Give **one** advantage of displaying this data in a tally chart rather than leaving it as raw data.

(1 mark)

---

---

---

---

2 The table shows the results of a survey into Reading Habits.

For example, 95% of all respondents said "Read books" but only 92% of 16-24 year olds said this.

Some respondents participated in more than one option.

Reading Habit	All People %	Gender		Age			
		Male	Female	18- 26	27- 36	37- 56	57+
Read books	95%	94%	96%	92%	94%	96%	98%
Read newspapers	21%	20%	22%	4%	8%	28%	44%
Read online articles	54%	56%	52%	62%	58%	52%	44%
Read magazines	14%	11%	17%	13%	16%	17%	10%
Listened to audiobooks	12%	11%	13%	14%	13%	11%	10%
Visited the library	91%	90%	92%	84%	89%	96%	95%
Read e-books	37%	33%	41%	51%	45%	29%	23%
Subscribed to book services	7%	6%	8%	10%	8%	6%	4%
None of the above	6%	7%	5%	9%	7%	5%	3%

(a) A person who selected **Visited the library** is chosen at random.

Describe the most likely gender and age of the person.

(2 marks)

---

---

---

---

---

---

---

---

---

---

**(b)** Describe the relationship in the data between reading e-books and age.

(1 mark)

---

---

---

---

**(c)** Emily is starting a new library.

She suspects that over half of her customers will be female.

Explain how she used the information from the table to come to this conclusion.

(1 mark)

---

---

---

---

---

---

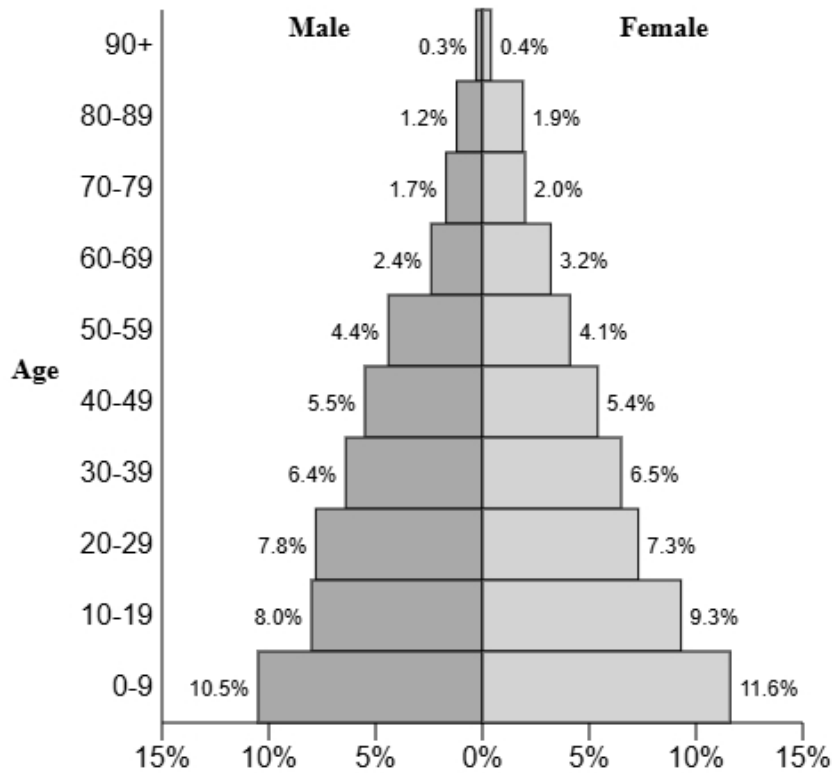
---

---

---

---

3 The population pyramid below shows the percentage of males and females in each age group for the town Marshcombe.



(a) Write down the percentage of females in the age group 20-29.

(1 mark)

\_\_\_\_\_ %

(b) Find the age group for males that has 6.4% of the population.

(1 mark)

---



---



---



---

(c) Find the age group that has 17.3% of the population.

(1 mark)

---

---

---

---

(d) Compare the percentage of the population aged 50-79 between males and females.

(1 mark)

---

---

---

---

(e) Give a reason why the sum of all the percentages is 99.9% and not 100%.

(1 mark)

---

---

---

---

- 4 A comedy club sells four different ticket types at different prices.  
The tickets for a weekend have been recorded in this table.

Ticket price	£5	£10	£20	£35	Total
Amount sold	290	120	75	15	500

The owner of the comedy club wants to open up a new comedy club in a different city.

She uses the data to predict how the tickets will sell in the new location.

Ticket price	£5	£10	£20	£35
Predicted proportion	58%	24%	15%	3%

- (a) The owner want to improve her predictions.  
Explain how she can do this.

(1 mark)

---

---

---

---

(b) The owner opens up the new comedy club.

The table below shows the tickets sales over the opening weekend at the new comedy club.

Ticket price	£5	£10	£20	£35
True proportion	51%	21%	19%	9%

Compare the sales of £35 tickets at the new location compared to the original location and suggest a reason for any difference.

(2 marks)

---

---

---

---

---

---

---

---

---

---

5 Emily owns a food truck that sells bagels.

Customers can choose either ham or cheese bagel **and** either a coffee or a hot chocolate drink.

She records her sales in a two-way table, but the table is incomplete.

	coffee	hot chocolate	Total
ham bagel		45	87
cheese bagel	33	38	
Total	75		158

(a) Complete the two way table.

(2 marks)

	coffee	hot chocolate	Total
ham		45	87
cheese	33	38	
Total	75		158



**6** A hospital is planning to introduce a new appointment booking system.  
Sophia wants to carry out a survey to find out what all patients think about the proposed change.

Sophia thinks that she should take a sample rather than a census.

**(a)** Give two reasons why Sophia might think this.

(2 marks)

---

---

---

---

---

---

---

---

**(b)** Sophia has decided to use the electoral register as a sampling frame.  
Explain what a sampling frame is.

(1 mark)

---

---

---

---

**(c)** Sophia has decided to use the electoral register as a sampling frame.  
State one problem Sophia may have using the electoral register as a sampling frame.

(1 mark)

---

---

---

---

(d) Give **two** reasons why Sophia should conduct a pilot survey.

(2 marks)

---

---

---

---

---

---

---

---

**7** Liam is a civil servant working for the Department for Work and Pensions (DWP). He is researching the average weekly working hours of UK employees and takes a simple random sample of 10 workers from various industries and asks them how many hours they worked last week.

The hours worked of the 10 people are listed:

59	54	56	55	90
56	55	59	58	58

Liam believes that one of the values is an outlier.

**(a)** Describe the meaning of the term ‘simple random sample’.

(1 mark)

---

---

---

---

**(b)** Write down the value that is most likely to be an outlier and explain why you think this value is an outlier.

(2 marks)

---

---

---

---

---

---

---

---

(c) Liam removes the outlier.

State whether the mean of the remaining nine values is greater than, is equal to or is less than the mean of all ten salaries.

Give a reason for your answer.

(2 marks)

---

---

---

---

---

---

---

---

---

---

(d) After calculating the mean of the nine values without the outlier, Liam uses this mean in a report to describe all the employees in the UK.

Describe two things that could affect the reliability of her conclusions.

(2 marks)

---

---

---

---

---

---

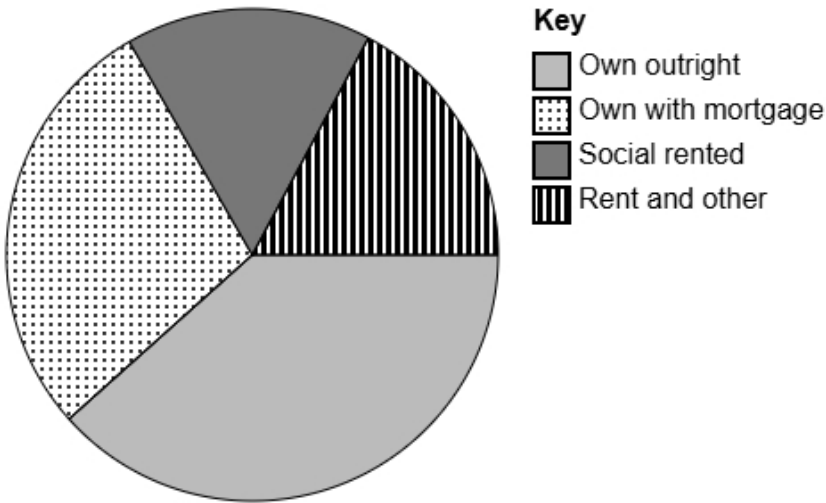
---

---

---

---

8 The accurately drawn pie chart shows information about the tenure types for people in Wales in 2021.



(a) Explain how you can tell that most households own their house outright in Wales in 2021 using the pie chart.

(1 mark)

---

---

---

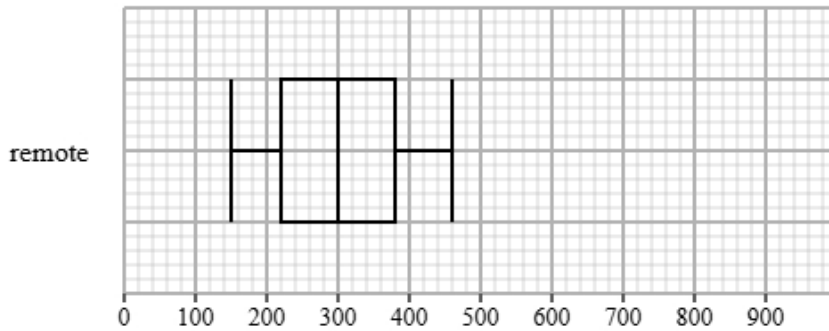
---

(b) In 2021 there was an estimated 130 thousand households.  
Calculate an estimate for the number of people in the UK in 2021 who's highest level of qualification was 'Own outright'.  
Round your answer to the nearest thousand.

(2 marks)

\_\_\_\_\_ thousand

- 9 Ethan collected the steps for remote and on-site workers in an hour within their day. Both groups recorded their steps over the same period. The box plot presents data on the steps for the remote workers.

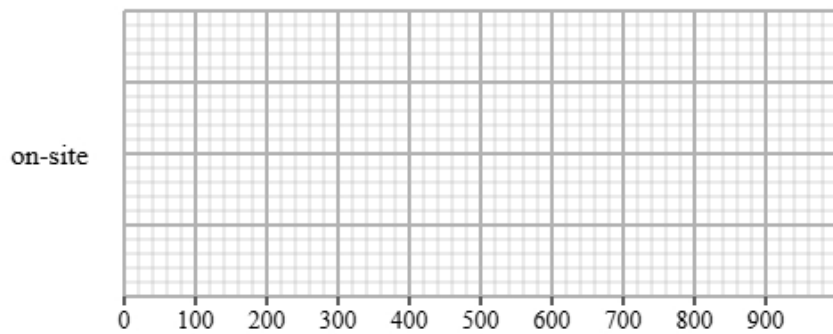


The table gives information about the steps for the on-site workers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
300	550	800	850	900

- (a) Draw a box plot for the steps for the on-site workers.

(2 marks)







(c) An airline has announced a new route at a price of £2000.  
Jamie is planning on using the line of best fit on the scatter diagram to predict the distance of the flight.

Explain whether or not it is appropriate to use the line of best fit for this prediction.

(2 marks)

---

---

---

---

**11** Noah investigates the reaction times (in milliseconds) of 120 people taking a driving test.

The times range from 203 ms to 281 ms.

Noah considers using one of the two possible grouped frequency tables for the results, Table A or Table B, shown below.

**Table A**

Time ( $t$ ms)	Frequency
$170 < t \leq 200$	0
$200 < t \leq 230$	31
$230 < t \leq 260$	67
$260 < t \leq 290$	22
$290 < t \leq 310$	0

**Table B**

Time ( $t$ ms)	Frequency
$200 < t \leq 220$	14
$220 < t \leq 240$	40
$240 < t \leq 260$	44
$260 < t \leq 280$	20
$280 < t \leq 300$	2

**(a)** Give **two** advantages of using grouped data rather than raw data.

(2 marks)

---

---

---

---

---

---

---

---

**(b)** Noah feels that Table B gives more detail than Table A about the results.  
Assess the appropriateness of Noah's claim.

(2 marks)

---

---

---

---

---

---

---

---

---

---

**(c)** Noah wants to work out the average reaction times of the 120 people taking a driving test.

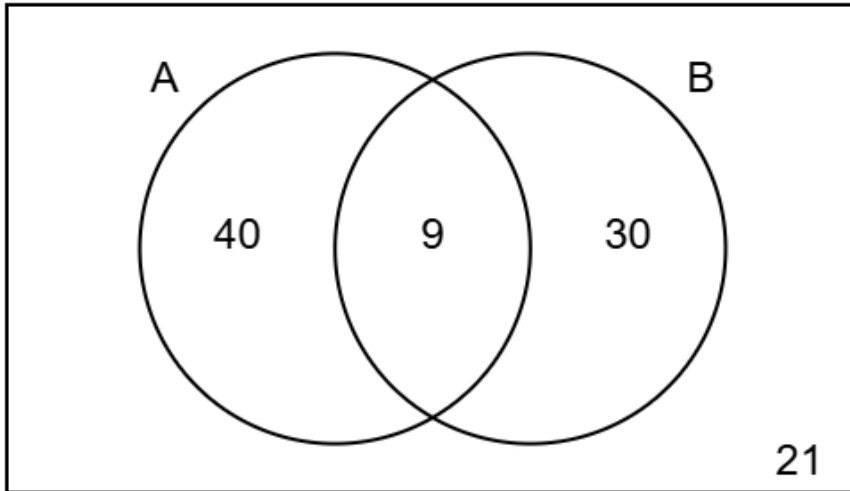
He decides to use Table B.

Calculate the average reaction times of the 120 people taking a driving test, giving your answer to 1 decimal place.

(3 marks)

\_\_\_\_\_ ms

- 12** The Venn diagram shows information about 100 customers in shopping centre.  
A is the event that the customer bought a product in the electronics shop.  
B is the event that the customer bought a product in the clothing shop.  
The numbers in the Venn diagram indicate the number of customers.



- (a)** In the Venn diagram, explain what the number 9 means.

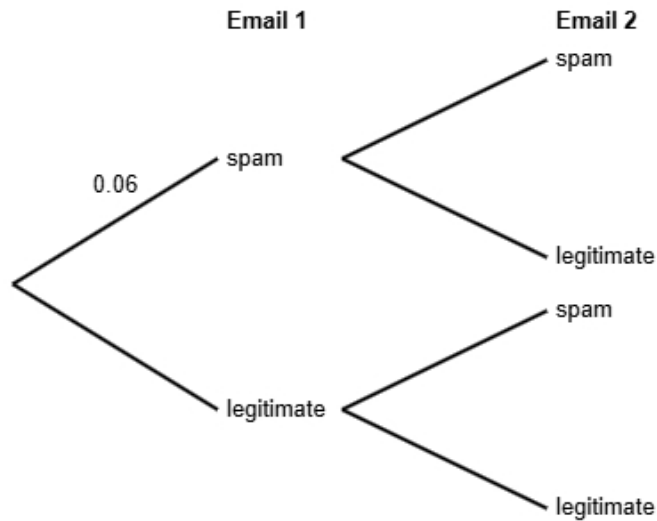
(1 mark)

---

---



- 13** A study shows that 6% of emails received by a certain email provider are spam. All other emails are legitimate. Maria receives two emails in her inbox. She does not know if each email is spam or legitimate.



- (a)** Complete the probability tree diagram.

(2 marks)

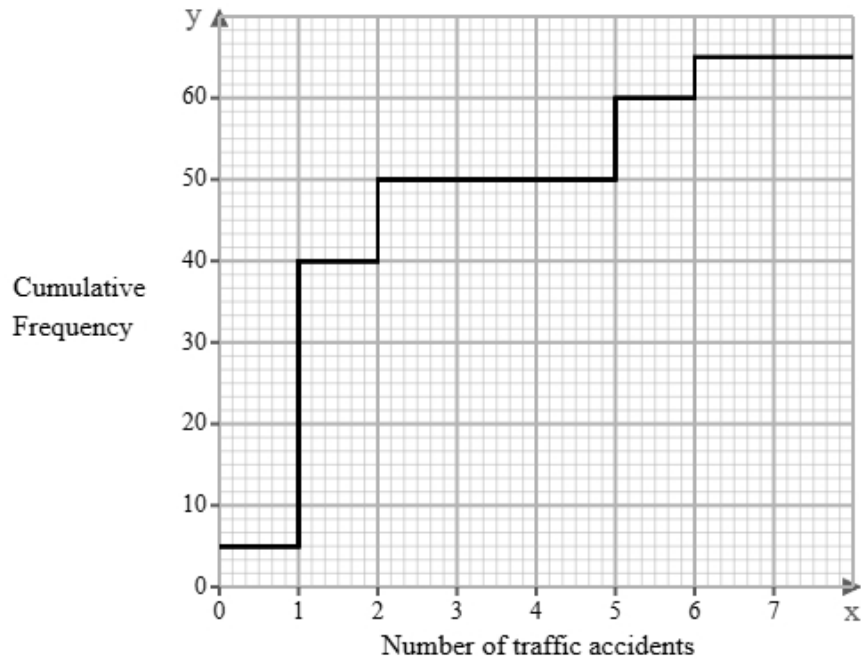
- (b)** Find the probability that both of Maria's emails are legitimate.

(2 marks)

\_\_\_\_\_



14 The cumulative frequency step polygon shows information about number of traffic accidents reported in a neighbourhood over 65 days.



(a) Give a reason why a cumulative frequency step polygon has been used to display this data.

(1 mark)

---

---

---

---

**(b)** Find the number of days where there were:

- i) exactly 4 traffic accidents.
- ii) more than 4 traffic accidents.

(3 marks)

i) Exactly 4 traffic accidents: \_\_\_\_\_

ii) More than 4 traffic accidents: \_\_\_\_\_

**(c)** In 60 days fewer than  $x$  traffic accidents were reported.

Find the value of  $x$

(1 mark)

\_\_\_\_\_

**(d)** Rosemary believes the interquartile range of number of traffic accidents reported is 8.

Explain why the interquartile range for this data cannot be 8.

(1 mark)

---

---

---

---