AQA (GCSE Notes)

Chapter 6: Statistics

Q1. A random sample of 40 students was taken from a school of 500 students. Why might the sample not fully represent the whole school?

Answer: The sample might not fully represent the whole school because it may not include all types of students.

Solution:

Total students = 500

Sample size = 40

Percentage sampled = (40 ÷ 500) × 100 = 8%

Since only 8% of students are selected, some year groups, genders, or abilities might be underrepresented.

Therefore, the sample may not reflect the entire school accurately.

Q2. A survey was done using a sample of 100 people from a town. Explain one limitation of using this sample to represent the views of the entire town.

Answer: The sample may not include all groups in the town, leading to biased results.

Solution:

The town likely has a diverse population.

If the 100 people are from similar age, job, or area, their views might not match the entire town.

So, the results may not be accurate for everyone.

Q3. A student wants to investigate how much time people spend on homework. She chooses a sample of 10 friends. Explain why this sample might not be suitable.

Answer: The sample is biased because her friends may have similar homework habits.

Solution:

The student and her friends likely attend the same class or school.

They might have similar subjects and workloads.

This makes the sample unrepresentative of all students.

Q4. A company takes a sample of 50 products from a batch of 2000. Explain one reason why they take a sample instead of checking every product.

Answer: It saves time and resources.

Solution:

Checking all 2000 products would take a lot of time and cost more.

A random sample of 50 can still give an idea of the product quality.

This is more efficient for large batches.

Q5. A bar chart shows the number of pets owned by children in different classes. Describe one thing you can learn from the chart.

 Answer: You can learn which class has the most or fewest pets.

 Solution:

 Look at the height of each bar.

 The taller the bar, the more pets that class owns.

 This helps compare between classes.

Q6. A pie chart shows how a family spends its income. Describe how to work out the amount spent on rent if the angle is 90°.

 Answer: Divide 90° by 360° and multiply by total income.

 Solution:

 Fraction of income spent on rent = 90 ÷ 360 = 1⁄4

 Amount spent = Total income × 1⁄4

Q7. A frequency table shows how many hours students revise for a test. Explain how to find the modal class.

 Answer: The modal class is the class with the highest frequency.

 Solution:

 Look at the frequency column.

 Find the class with the largest number.

 That class is the modal class.

Q8. A vertical line chart shows the number of goals scored in 10 football matches. How would you describe the spread of the data?

 Answer: Look at the range and how values vary.

 Solution:

 Find the lowest and highest number of goals.

 Range = Highest − Lowest

 Check if the values are close or spread out.

 This tells you how consistent the scores are.

Q9. A pictogram shows how many books were read by students in each month. Describe how to interpret a symbol that represents 5 books.

 Answer: Count the symbols and multiply by 5.

 Solution:

 If 3 symbols = 3 × 5 = 15 books

 Use the key to find the number of books per symbol

 Then total = number of symbols × value of one symbol

Q10. A time series graph shows daily temperatures over a month. Describe one trend you can see from the graph.

 Answer: You can see if temperatures are increasing, decreasing, or staying the same.

Solution:

Look at the general direction of the line.

If it rises, temperature is increasing.

If it falls, temperature is decreasing.

Flat line means stable temperatures.

Q11. A line graph shows a company's profit over a year. What does a downward slope on the graph tell you?

Answer: It shows that the company's profit is decreasing.

Solution:

Downward slope = value is going down over time

So, profit is reducing month by month.

Q12. Draw a frequency table for the following data: 5, 7, 8, 7, 6, 5, 5, 8, 7, 6.

Answer:

Value	Frequency
5	3
6	2
7	3
8	2

Solution:

Count each number

5 appears 3 times
6 appears 2 times
7 appears 3 times
8 appears 2 times

Q13. Construct a bar chart for the number of rainy days in each week of a month: 3, 2, 5, 4.

Answer: Draw bars for each week with heights matching the number of rainy days.

Solution:

Label x-axis: Week 1, Week 2, Week 3, Week 4

Label y-axis: Rainy Days

Plot bars:

Week 1 = 3

Week 2 = 2

Week 3 = 5

Week 4 = 4

Q14. A pie chart has a 72° angle for one category. What fraction of the whole does this represent?

 Answer: 72 ÷ 360 = 1⁄5

 Solution:

 Fraction = Angle ÷ 360

 Fraction = 72 ÷ 360 = 0.2 = 1⁄5

Q15. A student records the number of push-ups done daily for a month. Suggest a suitable chart or graph to represent the data.

 Answer: A time series graph

 Solution:

 Push-ups are measured over time (daily).

 Time series graph shows trends and changes clearly.

Q16. A histogram shows the weight of people with class intervals of 10kg. Explain how the height of each bar is decided.

 Answer: Height = Frequency ÷ Class width

 Solution:

 Class width = 10 kg

 Height of bar = Frequency ÷ 10

 This gives frequency density

Q17. A histogram has unequal class intervals. Explain how the frequency density is calculated.

 Answer: Frequency density = Frequency ÷ Class width

 Solution:

 For each class, subtract lower limit from upper limit to get class width

 Then divide frequency by class width to get height of bar

Q18. Draw a histogram for this data: 0–10 (4), 10–20 (8), 20–30 (12), 30–50 (16).

 Answer:

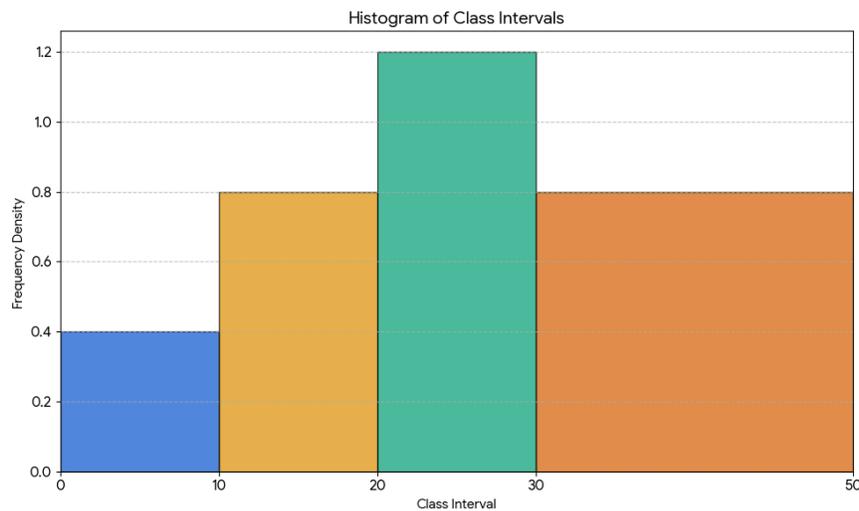| Class	Width	Frequency	Frequency Density
0 - 10	10	4	0.4
10 - 20	10	8	0.8
20 - 30	10	12	1.2
30 - 50	20	16	0.8

Solution:

Calculate the class width for each interval

0–10 = 10
10–20 = 10
20–30 = 10
30–50 = 20
Calculate frequency density = Frequency ÷ Width
Plot histogram with class intervals on x-axis and frequency density on y-axis

Q19. A grouped frequency table shows the time taken to run a race. Explain how to estimate the mean time.

 Answer: Use the midpoints and total frequency.
 Solution:

1. Find the midpoint of each class
2. Multiply midpoint × frequency for each class
3. Add all (midpoint × frequency) values
4. Add all frequencies
5. Mean = Total (midpoint × frequency) ÷ Total frequency

Q20. A cumulative frequency graph is plotted for the height of students. Describe how to find the median from the graph.

 Answer: Find the height at half of the total frequency.
 Solution:

1. Total frequency = last value on cumulative frequency axis
2. Divide by 2 to get median position
3. Locate this value on y-axis
4. Draw a horizontal line to meet the curve
5. From that point, draw a vertical line down to x-axis
6. The value on x-axis is the median height

Q21. A cumulative frequency graph is used to estimate the interquartile range. Describe how to do this.

Answer: Subtract the lower quartile (Q1) from the upper quartile (Q3) using the cumulative frequency graph.

Solution:

1. Find the total frequency from the highest cumulative value on the y-axis.
2. Q1 (Lower Quartile) is the value below which 25% of the data falls.
   Q1 position = 1⁄4 × total frequency.
3. Q3 (Upper Quartile) is the value below which 75% of the data falls.
   Q3 position = 3⁄4 × total frequency.
4. Use the graph to find the data values that correspond to Q1 and Q3.
   – For Q1: go to its y-position, draw across to the curve, then down to the x-axis.
   – For Q3: repeat the same process.
5. Interquartile Range (IQR) = Q3 − Q1.

Q22. A box plot shows the test scores of two classes. How can you compare the spread of scores between the classes?

Answer: Compare the range and interquartile range (IQR) of both box plots.

Solution:

1. Range = Maximum − Minimum.
2. Q1 (Lower Quartile) = value below which 25% of scores fall.
3. Q3 (Upper Quartile) = value below which 75% of scores fall.
4. IQR = Q3 − Q1.
5. Compare both classes using range and IQR.
6. A larger IQR or range shows greater spread.

Q23. A box plot shows an outlier. Explain what an outlier is and how it can affect the data.

Answer: An outlier is a value much higher or lower than the rest of the data. It can affect the mean and range.

Solution:

1. An outlier is a value far from the other data points.
2. It increases the range.
3. It can pull the mean up or down.
4. It usually has less effect on the median and IQR.

Q24. A dataset has values: 3, 4, 6, 8, 9. Find the range and explain what it tells you.

Answer: Range = 6. It tells how spread out the data is.

Solution:

1. Maximum = 9.
2. Minimum = 3.
3. Range = 9 − 3 = 6.
4. This shows the difference between the highest and lowest values.

Q25. Explain how to find the median from a set of 11 ordered numbers.
 Answer: Select the 6th number.
 Solution:

1. Arrange numbers in ascending order.
2. Count total values = 11.
3. Median position = (11 + 1) ÷ 2 = 6.
4. Median = 6th value in the list.

Q26. A frequency table shows the ages of people in a club. Explain how to find the mean age.
 Answer: Multiply each age by its frequency, add the results, then divide by the total frequency.
 Solution:

1. Multiply Age × Frequency for each row.
2. Add all the results = Total Age Sum.
3. Add all the frequencies = Total People.
4. Mean = Total Age Sum ÷ Total People.

Q27. A group of students record their walking times. Explain how to find the modal class.
 Answer: The modal class is the class with the highest frequency.
 Solution:

1. Look at the grouped frequency table.
2. Identify the class with the largest frequency.
3. That class interval is the modal class.

Q28. A bar chart shows the favourite subjects of students. Explain how you can tell which subject is most popular.
 Answer: The subject with the tallest bar is the most popular.
 Solution:

1. Each bar shows the number of students.
2. The tallest bar means the highest number.
3. That subject is chosen by the most students.

Q29. A student makes a pie chart from data collected. Explain why all the angles should add up to 360°.

 Answer: A full circle has 360°, so all category angles must total 360°.

 Solution:

1. A pie chart is a circle.
2. A circle = 360°.
3. Each category is a slice with an angle.
4. All angles must add to 360° to show full data.

Q30. Explain why box plots are useful for comparing two data sets.

 Answer: They show the median, range, and interquartile range, making comparisons easier.

 Solution:

1. Box plots display 5 values: minimum, Q1, median, Q3, maximum.
2. You can compare medians to see which set has higher values.
3. Compare ranges and IQRs to see spread.
4. Outliers are also visible.

Q31. A grouped frequency table has class intervals 0–10, 10–20, and so on. Explain why you can't find the exact mean.

 Answer: The exact data values are not known within each class.

 Solution:

1. Only class intervals and frequencies are given.
2. Individual data values are missing.
3. We estimate the mean using midpoints.
4. Therefore, the mean is approximate.

Q32. Explain how to use midpoints when estimating the mean from grouped data.

 Answer: Multiply each midpoint by frequency, add results, then divide by total frequency.

 Solution:

1. Midpoint = (Lower limit + Upper limit) ÷ 2.
2. Multiply Midpoint × Frequency for each class.
3. Add all Midpoint × Frequency values.
4. Add all frequencies.
5. Mean ≈ Total (Midpoint × Frequency) ÷ Total Frequency.

Q33. A histogram shows the number of cars passing a junction in different time intervals. Explain why histogram bars are drawn without gaps.

 Answer: Because the data is continuous with no breaks between intervals.

 Solution:

1. Time intervals are continuous.
2. There are no gaps between intervals.
3. Gaps would wrongly suggest categories.
4. Histogram bars touch to show continuous data.

Q34. A student wants to make a cumulative frequency graph. What values should they plot on the y-axis?
 Answer: The cumulative frequencies.
 Solution:

1. Add frequencies one by one down the table.
2. These running totals are cumulative frequencies.
3. Plot cumulative frequencies on the y-axis.
4. Plot them against the upper class boundaries on the x-axis.

Q35. Explain how to calculate the upper and lower quartiles from a list of data.
 Answer: Divide the ordered data into four equal parts to find Q1 and Q3.
 Solution:

1. Arrange data in ascending order.
2. Total number of values = n.
3. Q1 (Lower Quartile) = 1⁄4 × (n + 1).
4. Q3 (Upper Quartile) = 3⁄4 × (n + 1).
5. Find values at those positions.

Q36. A box plot has a long whisker on the right. What does this suggest about the distribution of the data?
 Answer: The data is positively skewed (skewed to the right).
 Solution:

1. Long right whisker = values are spread out at the high end.
2. Most data is at the lower end.
3. Few high values extend the upper tail.
4. This is called positive skew.

Q37. A set of data has a mean of 20 and a range of 50. What does this tell you about the spread of the data?
 Answer: The data is widely spread around the mean.
 Solution:

1. Mean = 20.
2. Range = 50 = Maximum − Minimum.

3. A large range shows high variation in values.
4. The data is not close together.

Q38. Explain why the median is sometimes a better measure than the mean.
Answer: The median is not affected by extreme values.
Solution:

1. Mean uses all data points.
2. One extreme value can change the mean a lot.
3. Median is the middle value.
4. It stays stable even with outliers or skewed data.

Q39. A set of data has a mode of 15. What does this tell you?
Answer: 15 is the most frequently occurring value.
Solution:

1. Mode = value that appears most often.
2. 15 occurs more than any other number in the set.
3. This is the most common value.

Q40. A bar chart shows the number of people choosing different drinks. Two bars are the same height. What does this mean?
Answer: The same number of people chose those two drinks.
Solution:

1. Height of bar = number of people.
2. Same height = equal number.
3. So, both drinks were chosen equally.

Q41. A frequency table shows 100 students and how many use a certain app. Explain how to calculate the percentage who use it.
Answer: Divide the number of users by 100 and multiply by 100 to get the percentage.
Solution:
Step 1: Use the formula: Percentage = (Number of users ÷ Total students) × 100
Step 2: Suppose 35 students use the app: Percentage = (35 ÷ 100) × 100
Step 3: Percentage = 35%

Q42. A pie chart shows transport methods. The 'bus' section is 120°. What does this angle represent in terms of the whole group?
Answer: It represents one-third or 33.3% of the group.
Solution:
Step 1: Use the formula: Percentage = (Angle ÷ 360°) × 100
Step 2: Percentage = (120 ÷ 360) × 100
Step 3: Percentage = 33.3%

Q43. A student calculates the interquartile range as 12. What does this number tell you about the data?

 Answer: The middle 50% of the data is spread over 12 units.

 Solution:

 Step 1: Interquartile range (IQR) = Q3 – Q1

 Step 2: If IQR = 12, it means Q3 is 12 units greater than Q1

 Step 3: This shows how spread out the central values are

Q44. A data set contains an outlier that is much higher than the rest. What effect does this have on the mean and median?

 Answer: It increases the mean but has little or no effect on the median.

 Solution:

 Step 1: The mean is sensitive to extreme values

 Step 2: An outlier pulls the mean upwards

 Step 3: The median is the middle value, so it remains almost unchanged

Q45. A time series graph shows sales increasing steadily. What can this trend suggest for the future?

 Answer: Sales may continue to increase in the future.

 Solution:

 Step 1: A steady upward trend indicates consistent growth

 Step 2: If current conditions stay the same, the trend is likely to continue

Q46. A line graph shows a sudden drop in values. Give two possible reasons for this change.

 Answer: A seasonal effect or a sudden event like a strike or shortage.

 Solution:

 Step 1: Sudden changes can be due to external events

 Step 2: Check for known disruptions or unusual time periods

Q47. A grouped frequency table is missing the class 20–30. Explain how this affects the results.

 Answer: The data will be incomplete and calculations like mean and median will be inaccurate.

 Solution:

 Step 1: Missing a class means some values are not represented

 Step 2: This affects total frequency and central measures

 Step 3: The final interpretation may be misleading

Q48. A student draws a histogram with incorrect bar widths. Explain how this affects interpretation.

 Answer: It gives a false impression of the data distribution.

 Solution:

 Step 1: In a histogram, area represents frequency

Step 2: Incorrect widths distort the area
Step 3: This misleads the viewer about the actual frequencies

Q49. A set of test scores has a mean of 65 and a range of 5. What does this tell you about the scores?
 Answer: The scores are close together around 65.
 Solution:
 Step 1: Range = Highest – Lowest = 5
 Step 2: A small range means low variation
 Step 3: Most scores are near the mean of 65

Q50. A box plot shows the test results of two groups. One has a larger interquartile range. What does this tell you about that group?
 Answer: That group has more variation in the middle 50% of scores.
 Solution:
 Step 1: A larger IQR means a wider spread of middle values
 Step 2: The scores are less consistent in that group

Q51. Draw a scatter graph for the following pairs of values: (2, 5), (3, 7), (4, 8), (5, 11), (6, 13), (7, 14). Describe the type of correlation.
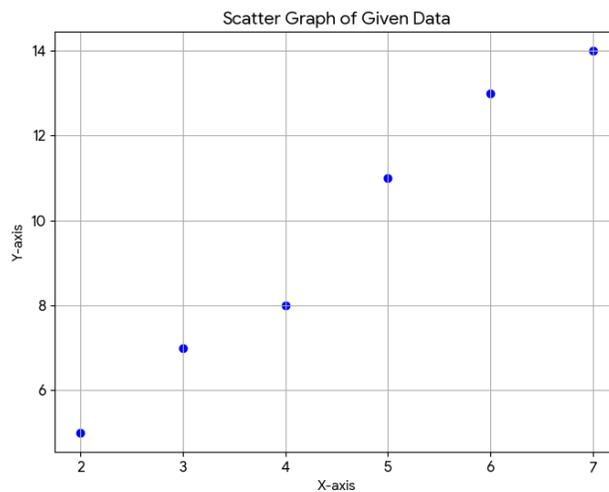 **Answer:** There is a positive correlation.
 Solution:
 Step 1: Plot the points on a graph
 Step 2: The points rise from left to right
 Step 3: This shows as one variable increases, so does the other

Scatter Graph of Given Data

Q52. The test scores of 15 students are: 34, 45, 50, 52, 56, 58, 60, 61, 65, 67, 70, 73, 75, 78, 80. Draw a box plot to represent this data.
 Answer: The box plot has minimum 34, Q1 = 52, median = 61, Q3 = 70, maximum = 80
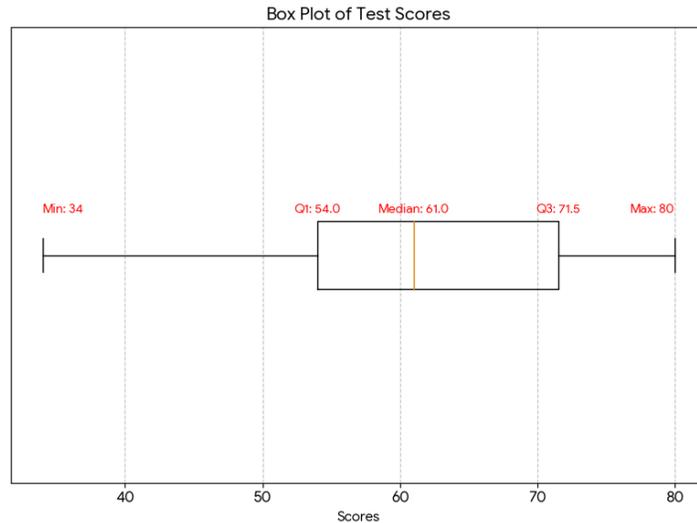 **Solution:**

Step 1: Order data (already in order)

Step 2: Median = 8th value = 61

Step 3: Lower half = 34 to 60 → Q1 = 4th value = 52

Step 4: Upper half = 65 to 80 → Q3 = 12th value = 70

Step 5: Minimum = 34, Maximum = 80

Box Plot of Test Scores

Min: 34   Q1: 54.0   Median: 61.0   Q3: 71.5   Max: 80

Scores

Q53. The number of hours studied and the corresponding test scores for six students are: (1, 35), (2, 40), (3, 50), (4, 55), (5, 65), (6, 70). Plot these values on a scatter graph. What type of correlation is shown?
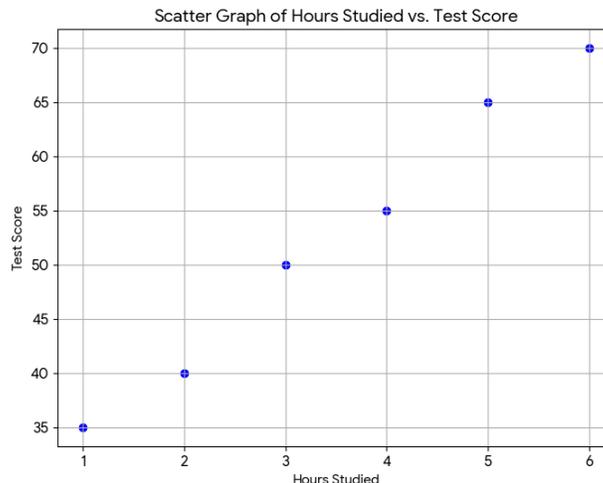
Answer: Positive correlation

Solution:

Step 1: Plot each point

Step 2: The values rise as hours increase

Step 3: Shows that more study hours lead to better scores

Scatter Graph of Hours Studied vs. Test Score

Q54. The ages of a group of people are: 23, 25, 26, 27, 28, 29, 30, 90. Explain how the outlier affects the mean and the median.

 Answer: The outlier increases the mean but not the median much.

 Solution:

 Step 1: Mean = (23+25+26+27+28+29+30+90) ÷ 8 = 278 ÷ 8 = 34.75

 Step 2: Median = average of 4th and 5th values = (27+28)/2 = 27.5

 Step 3: The mean is pulled higher by 90, median is not affected much

Q55. A grouped frequency table shows the marks of 30 students: 0–10: 3, 10–20: 7, 20–30: 10, 30–40: 6, 40–50: 4. Draw a cumulative frequency graph and estimate the median.

 Answer: Median is approximately 22

 Solution:

 Step 1: Cumulative frequency:
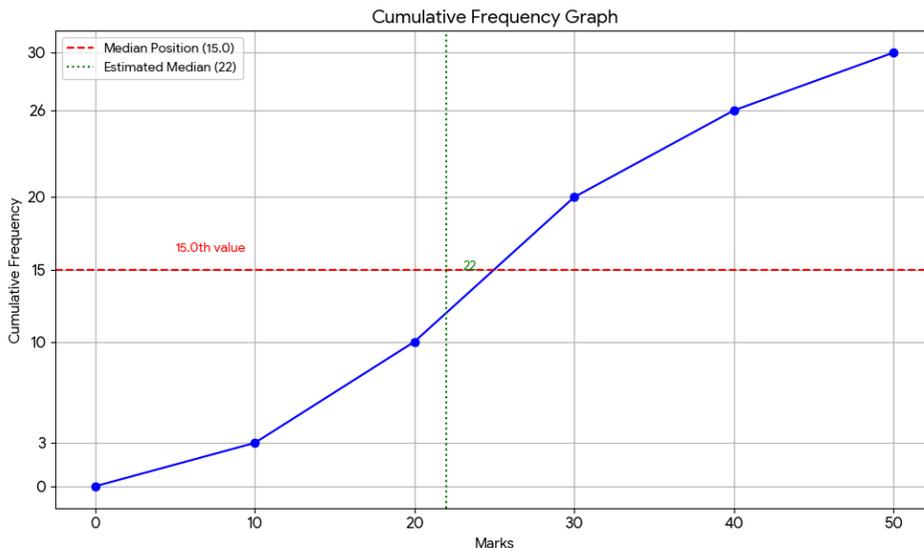
 0–10: 3

 10–20: 3+7=10

 20–30: 10+10=20

 30–40: 20+6=26

 40–50: 26+4=30

 Step 2: Median = 30 ÷ 2 = 15th value

 Step 3: Find class with cumulative frequency ≥ 15 → in 20–30 group

 Step 4: Estimate median using graph between 10–20 and 20–30

 Step 5: Approximate value = 22

Cumulative Frequency Graph

Q56. The following values represent test scores: 44, 49, 53, 56, 59, 60, 61, 62, 64, 65, 66. Draw a box plot and calculate the interquartile range.

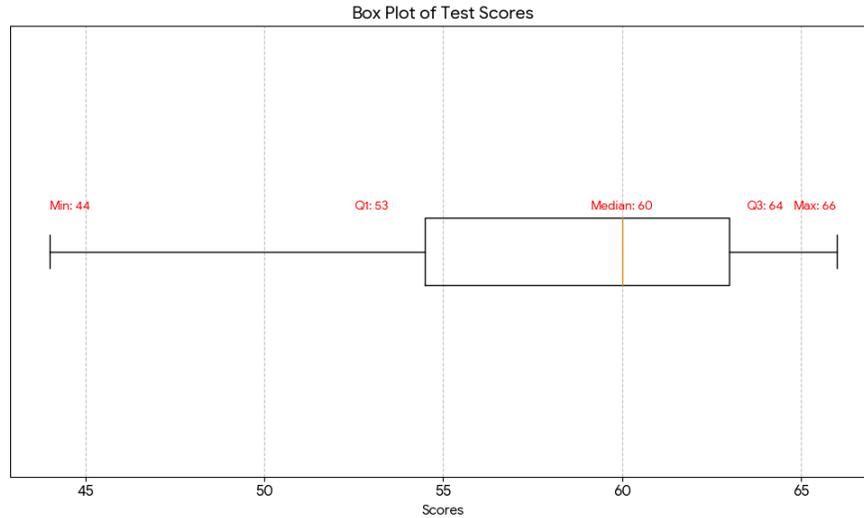 **Answer:** IQR = 64 – 53 = 11

 Solution:

 Step 1: Order data (already ordered)

Step 2: Median = 6th value = 60
Step 3: Lower half = 44 to 59 → Q1 = median of 1st 5 = 53
Step 4: Upper half = 61 to 66 → Q3 = median of last 5 = 64
Step 5: IQR = Q3 – Q1 = 64 – 53 = 11

Box Plot of Test Scores

Q57. Given this grouped frequency table, estimate the mean:

Class Interval	Frequency
0 - 10	5
10 - 20	8
20 - 30	12
30 - 40	5

Answer: Estimated mean = 18.5
 Solution:
Step 1: Find midpoints:
0–10 = 5
10–20 = 15
20–30 = 25
30–40 = 35
Step 2: Multiply midpoints by frequencies:
5×5 = 25
15×8 = 120
25×12 = 300
35×5 = 175
Step 3: Total = 25+120+300+175 = 620

Step 4: Total frequency = 5+8+12+5 = 30
Step 5: Mean = 620 ÷ 30 = 20.67

Q58. Write whether each of the following is discrete or continuous data and explain: Number of books read, temperature, height, number of pets.

Answer:

Number of books read – Discrete

Temperature – Continuous

Height – Continuous

Number of pets – Discrete

Solution:

Step 1: Discrete = counted values (whole numbers)

Step 2: Continuous = measured values (can include decimals)

Q59. Compare a bar chart and a histogram using this data:

Shoe Sizes: Size 5: 2, Size 6: 4, Size 7: 6, Size 8: 5

Answer: Bar chart is appropriate because shoe sizes are discrete data.

Solution:

Step 1: Bar charts are for discrete data

Step 2: Histogram is for continuous grouped data

Step 3: Shoe sizes are exact values, not intervals

Q60. The ages of 12 students are: 13, 14, 14, 14, 15, 15, 16, 16, 17, 17, 18, 18. Represent this data using a stem-and-leaf diagram.

Answer:

Solution:

Stem	Leaf
1	3 4 4 4 5 5 6 6 7 7 8 8

Q61. The following data shows a relationship between hours revised and exam marks: (1, 35), (2, 42), (3, 47), (4, 56), (5, 64), (6, 70). Draw a scatter graph and add a line of best fit. Predict the score for 7 hours of revision.

Answer: Positive correlation; predicted score is about 76

Solution:

Step 1: Plot the points on a graph

Step 2: Draw a line of best fit through the pattern

Step 3: Estimate where the line passes for x = 7

Step 4: From the trend, each extra hour increases marks by about 6

Step 5: Predicted score = 70 + 6 = 76

Q62. A school has 1,200 students. A sample of 120 is taken to investigate lunch preferences. Explain the difference between population and sample in this case.

Answer: Population = 1,200 students, Sample = 120 students taken to represent the group

Solution:

Step 1: Population means the entire group being studied

Step 2: Sample is a smaller part selected from the population

Step 3: In this case, 1,200 is the population, 120 is the sample

Q63. Given the frequency table:

Score	Frequency
1	2
2	5
3	8
4	3
5	2

Find the median score.

Answer: Median = 3

Solution:

Step 1: Total frequency = 2+5+8+3+2 = 20

Step 2: Median position = (20 + 1) ÷ 2 = 10.5th value

Step 3: Cumulative frequency:

Score 1: 2

Score 2: 2+5=7

Score 3: 7+8=15

Step 4: 10.5th value lies in score 3

Step 5: Median = 3

Q64. A set of data has the following values: 2, 2, 2, 5, 7, 9, 10

Discuss whether mode or mean is more appropriate as a measure of average.

Answer: Mode is more appropriate because most values are 2

Solution:

Step 1: Mean = (2+2+2+5+7+9+10) ÷ 7 = 37 ÷ 7 ≈ 5.29

Step 2: Mode = 2

Step 3: Data is skewed with repeated low values

Step 4: Mode better reflects the most common value

Q65. Given the data: (2, 12), (4, 20), (6, 25), (8, 32), (10, 39)
Plot a scatter graph and state if there is correlation. If yes, what kind?
Answer: Positive correlation
Solution:
Step 1: Plot the points
Step 2: Points rise from left to right
Step 3: This indicates as x increases, y increases
Step 4: Correlation is positive

Q66. Use this grouped data to draw a cumulative frequency curve:
0–20: 4
20–40: 10
40–60: 8
60–80: 6
80–100: 2
Estimate the median and upper quartile.
Answer: Median ≈ 45, Upper quartile ≈ 70
Solution:
Step 1: Cumulative frequency:
0–20: 4
0–40: 4+10=14
0–60: 14+8=22
0–80: 22+6=28
0–100: 28+2=30
Step 2: Median position = 30 ÷ 2 = 15th
Step 3: Upper quartile = ¾ × 30 = 22.5th
Step 4: 15th value lies in 40–60
Step 5: 22.5th value lies in 60–80
Step 6: Estimate from curve: Median ≈ 45, Upper quartile ≈ 70

Q67. The ages of ten people are: 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Find the interquartile range and explain what it tells you about the spread.
Answer: IQR = 28 – 24 = 4
Solution:
Step 1: Ordered data is already given
Step 2: Q1 = median of first 5 = 24
Step 3: Q3 = median of last 5 = 28
Step 4: IQR = Q3 – Q1 = 28 – 24 = 4
Step 5: This shows the middle 50% of values spread over 4 years

Q68. A population is very large. Explain why a random sample of 100 is more practical than using the whole population.
Answer: It saves time, cost, and effort while still giving reliable results

Solution:
Step 1: Collecting data from the whole population is difficult
Step 2: A sample is quicker and cheaper to analyse
Step 3: A random sample avoids bias and represents the population

Q69. The frequency table is:

Interval	Frequency
0 - 10	4
10 - 20	6
20 - 30	10
30 - 40	5

Explain how to draw a histogram with unequal class widths.
Answer: Use frequency density = frequency ÷ class width
Solution:
Step 1: Find class widths (all are equal here = 10)
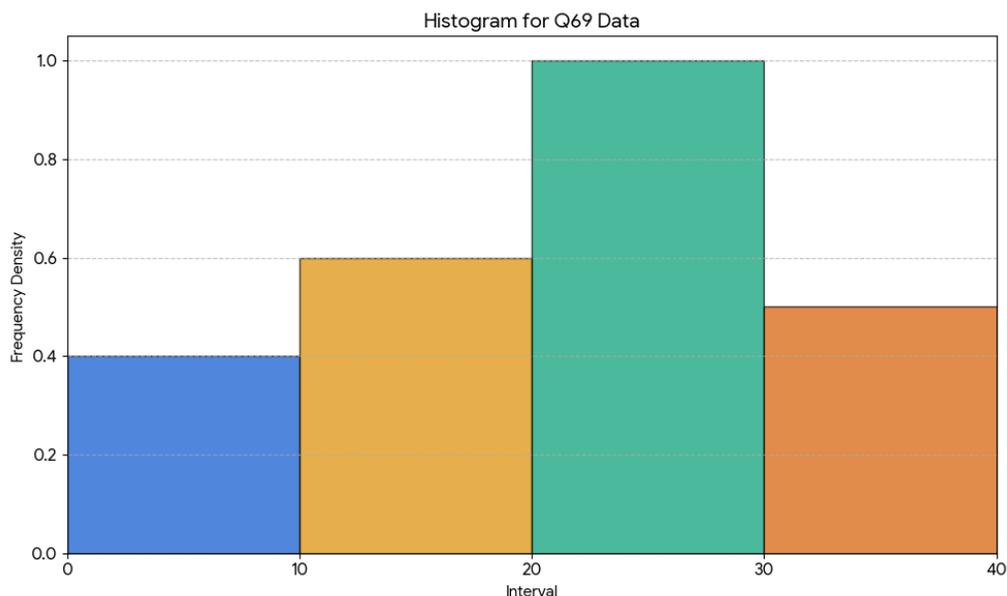Step 2: Frequency density = frequency ÷ width
0–10: 4 ÷ 10 = 0.4
10–20: 6 ÷ 10 = 0.6
20–30: 10 ÷ 10 = 1.0
30–40: 5 ÷ 10 = 0.5
Step 3: Plot bars using class width on x-axis and frequency density on y-axis

Histogram for Q69 Data

Q70. A scatter graph shows the values: (1, 10), (2, 9), (3, 8), (4, 7), (5, 6)
Draw the graph and describe the type of correlation.
 Answer: Negative correlation
 Solution:
 Step 1: Plot the points
 Step 2: As x increases, y decreases
 Step 3: The points fall from left to right
 Step 4: Correlation is negative

Q71. Calculate the range, mean, median, and mode for the following numbers: 6, 8, 8, 10, 12, 14, 16
 Answer: Range = 10, Mean = 10.57, Median = 10, Mode = 8
 Solution:
 Step 1: Range = 16 – 6 = 10
 Step 2: Mean = (6+8+8+10+12+14+16) ÷ 7 = 74 ÷ 7 ≈ 10.57
 Step 3: Median = 4th value = 10
 Step 4: Mode = value that appears most = 8

Q72. Box plots for Class A and Class B show the same median but different interquartile ranges. Explain what this tells you about the spread of scores.
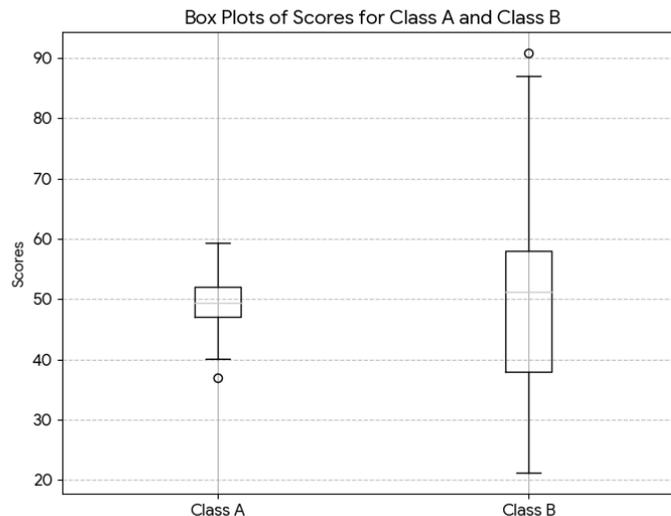 **Answer:** The class with the larger IQR has more varied scores
 Solution:
 Step 1: Same median = same middle value
 Step 2: Different IQR = different spreads
 Step 3: Larger IQR means less consistency

Box Plots of Scores for Class A and Class B

Q73. Scatter graphs for two different data sets show different patterns. One is tightly clustered along a line, the other is more spread out. Describe the strength of correlation in both graphs.
 Answer: Tightly clustered = strong correlation; spread out = weak correlation

Solution:

Step 1: Close to line = strong link between variables

Step 2: Spread out = weaker link

Step 3: Strength depends on how close the points follow a trend

Q74. A set of data includes: 1, 1, 1, 2, 2, 3, 3, 25

Explain whether the mean or median gives a better representation of the data.

Answer: Median is better because the outlier affects the mean

Solution:

Step 1: Mean = (1+1+1+2+2+3+3+25) ÷ 8 = 38 ÷ 8 = 4.75

Step 2: Median = (4th + 5th)/2 = (2+2)/2 = 2

Step 3: 25 is an outlier that increases the mean

Step 4: Median better represents typical value

Q75. The following scatter graph shows height (cm) against shoe size. Explain how the line of best fit can be used to estimate the shoe size of someone who is 175 cm tall.

Answer: Use the line to find the shoe size corresponding to 175 cm

Solution:

Step 1: Find 175 cm on the x-axis

Step 2: Move up to the line of best fit

Step 3: Move across to read the corresponding y-axis value

Step 4: That value is the estimated shoe size

Q76. The number of books read by 20 students last month is recorded.

Number of books: 0, 1, 2, 3, 4, 5

Frequencies: 2, 3, 7, 4, 2, 2

Present the data using a bar chart.

Answer: A bar chart with book numbers on x-axis and frequencies on y-axis

Solution:

Step 1: Plot bars for each number of books

0 → 2

1 → 3

2 → 7

3 → 4

4 → 2

5 → 2

Step 2: Label axes and title

Step 3: Bars should be equal width and spaced

Q77. Use the data below to draw a cumulative frequency graph and estimate the lower quartile.

Interval	Frequency
0 - 10	3
10 - 20	5
20 - 30	9
30 - 40	8
40 - 50	8

Answer: Lower quartile ≈ 17
Solution:
Step 1: Cumulative frequency:
0–10: 3
0–20: 3+5=8
0–30: 8+9=17
0–40: 17+8=25
0–50: 25+8=33
Step 2: Lower quartile = ¼ × 33 = 8.25th value
Step 3: Lies in 20–30 group
Step 4: Estimate value ≈ 17 from the graph

Q78. Why are the mean, median, and mode useful when comparing two different data sets?
Answer: They give different types of average and help compare central values and spread
Solution:
Step 1: Mean shows overall average
Step 2: Median shows middle value
Step 3: Mode shows most common value
Step 4: Together they give a full picture

Q79. Two data sets have the same mode but very different ranges. Explain what this tells you about the distributions.
Answer: The most common value is the same, but one set is more spread out
Solution:
Step 1: Same mode = same peak
Step 2: Different ranges = different variability
Step 3: One set may have outliers or more extreme values

**Q80. The heights (cm) of students are recorded as: 150, 153, 154, 155, 160, 162, 163, 170, 174
Is this data discrete or continuous? Explain.**
Answer: Continuous data
Solution:

Step 1: Heights are measured, not counted
Step 2: Can include decimals and vary smoothly
Step 3: Therefore, it is continuous data

Q81. The points plotted on a scatter graph are: (1, 5), (2, 6), (3, 5), (4, 6), (5, 5)
Describe the type of correlation and what it suggests.
 Answer: No correlation
 Solution:
 Step 1: Plot the points on a graph
 Step 2: Observe the pattern
 Step 3: The points do not show an increasing or decreasing trend
 Step 4: Therefore, there is no clear correlation between the variables

Q82. Two box plots show final exam scores for Group A and Group B. Group A has a wider range and lower median. What can you conclude about the performance of both groups?
 Answer: Group B performed better and more consistently
 Solution:
 Step 1: Wider range in Group A shows more variation in scores
 Step 2: Lower median in Group A means overall performance was lower
 Step 3: Group B had higher and more consistent results

Q83. The grouped data below shows the time taken (in minutes) to complete a task:

Time	Frequency
0 - 10	2
10 - 20	5
20 - 30	9
30 - 40	3

Draw a histogram to represent this data.
 Answer: Histogram with equal-width bars for each interval, height = frequency
 Solution:
 Step 1: Class widths are all 10
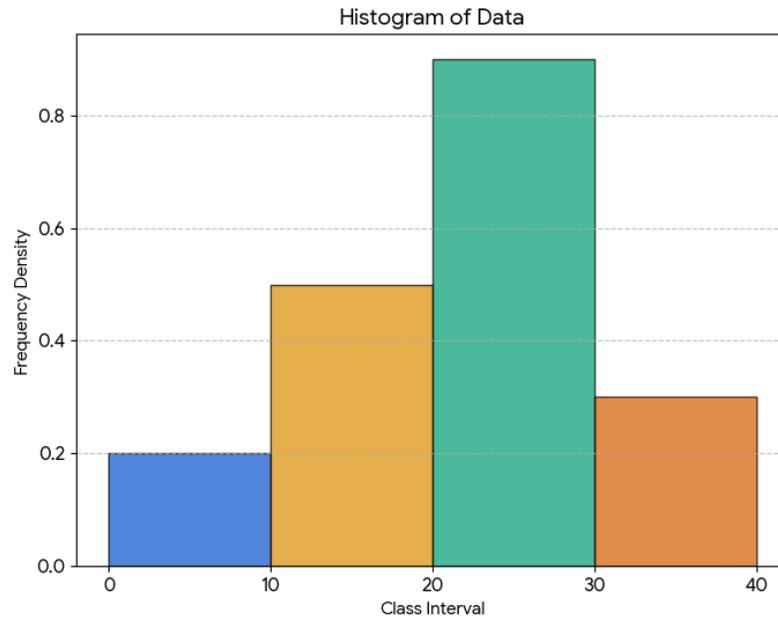 Step 2: Frequency density = frequency ÷ class width
 0–10: 2 ÷ 10 = 0.2
 10–20: 5 ÷ 10 = 0.5
 20–30: 9 ÷ 10 = 0.9
 30–40: 3 ÷ 10 = 0.3
 Step 3: Plot bars using class intervals on x-axis and frequency density on y-axis

Histogram of Data

Q84. A survey only includes people from one school. Explain why the sample may be biased when making conclusions about all students in the city.

 Answer: It does not represent the entire student population

 Solution:

 Step 1: One school may not reflect all types of students

 Step 2: Different schools may have different environments

 Step 3: A more diverse sample is needed for accurate results

Q85. The data values are: 12, 15, 17, 20, 22, 25, 30, 32

 Find the lower and upper quartiles and describe why quartiles are useful.

 Answer: Q1 = 15, Q3 = 30

 Solution:

 Step 1: Ordered data: already in order

 Step 2: Lower half: 12, 15, 17, 20 → Q1 = median of lower half = (15+17)/2 = 16

 Step 3: Upper half: 22, 25, 30, 32 → Q3 = median of upper half = (25+30)/2 = 27.5

 Step 4: Quartiles help describe the spread and detect outliers

Q86. The ages of 11 people are: 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31

 Calculate the interquartile range.

 Answer: IQR = 28 – 23 = 5

 Solution:

 Step 1: Median = 6th value = 26

 Step 2: Lower half = 21 to 25 → Q1 = median = 23

 Step 3: Upper half = 27 to 31 → Q3 = median = 28

 Step 4: IQR = Q3 – Q1 = 28 – 23 = 5

Q87. The points plotted are: (1, 2), (2, 4), (3, 6), (4, 8), (5, ?)
Use a line of best fit to predict the missing value.
Answer: 10
Solution:
Step 1: The pattern shows y = 2x
Step 2: For x = 5 → y = 2 × 5 = 10
Step 3: Missing value = 10

Q88. On a scatter graph, the points lie close to a straight line. Explain what this tells you about the relationship between the variables.
Answer: There is a strong correlation
Solution:
Step 1: Points close to a straight line = consistent trend
Step 2: A strong linear relationship exists
Step 3: This means as one variable changes, the other does too

Q89. The data below is about students' favourite colours:
Red: 5, Blue: 10, Green: 6, Yellow: 3
Explain why a bar chart is better than a histogram in this case.
Answer: Bar chart is for categories; histogram is for continuous data
Solution:
Step 1: Favourite colour is categorical data
Step 2: Bar charts are suitable for separate categories
Step 3: Histogram is only for grouped numerical data

Q90. A box plot shows a very high maximum value compared to the rest of the data. What might this indicate?
Answer: There may be an outlier
Solution:
Step 1: A large gap between Q3 and maximum suggests a high value
Step 2: This could be an extreme value
Step 3: May indicate an outlier or unusual result

Q91. A population of 1,000 workers is surveyed. Describe how to randomly select a fair sample of 100.
Answer: Use random number generator or lottery method
Solution:
Step 1: Assign a number to each of the 1,000 workers
Step 2: Use a random number generator to pick 100 unique numbers
Step 3: Select the corresponding people for the sample

Q92. Given the cumulative frequency table:

Marks	Cumulative Frequency
10	2
20	6
30	13
40	20
50	25

Estimate the median mark.
 Answer: Median ≈ 32
 Solution:
 Step 1: Total frequency = 25
 Step 2: Median position = 25 ÷ 2 = 12.5th value
 Step 3: Find interval containing 12.5th value → 30 to 40
 Step 4: Estimate value = about 32 from graph or interpolation

Q93. The scores are: 11, 14, 18, 19, 25, 27, 29
 Calculate the range and explain what it shows.
 Answer: Range = 29 – 11 = 18
 Solution:
 Step 1: Highest value = 29
 Step 2: Lowest value = 11
 Step 3: Range = 29 – 11 = 18
 Step 4: This shows the total spread of the data

Q94. The data is: 2, 2, 2, 3, 4, 100
 Would you use the mean or the median as the average? Explain why.
 Answer: Use the median
 Solution:
 Step 1: Mean = (2+2+2+3+4+100) ÷ 6 = 113 ÷ 6 ≈ 18.83
 Step 2: Median = (3+4)/2 = 2.5
 Step 3: Mean is affected by outlier 100
 Step 4: Median better reflects central value

Q95. A box plot shows the lower quartile close to the median but a large upper quartile. What does this suggest about the data?
 Answer: The upper half of the data is more spread out
 Solution:
 Step 1: Small gap between Q1 and median = values are close

Step 2: Large gap between median and Q3 = wider variation
Step 3: Suggests data is skewed to the right

Q96. From this frequency table, find the mean:

Marks	Cumulative Frequency
1	2
2	3
3	5
4	4

Answer: Mean = 2.85
 Solution:
 Step 1: Multiply marks by frequency
 1×2 = 2
 2×3 = 6
 3×5 = 15
 4×4 = 16
 Step 2: Total = 2+6+15+16 = 39
 Step 3: Total frequency = 2+3+5+4 = 14
 Step 4: Mean = 39 ÷ 14 ≈ 2.79

Q97. Two data sets have the same mean but different interquartile ranges. What does this tell you about the consistency of the data?
 Answer: The one with the smaller IQR is more consistent
 Solution:
 Step 1: Same mean = same average
 Step 2: Larger IQR = more spread out data
 Step 3: Smaller IQR = more consistent values

Q98. Define a representative sample and explain why it's important in data collection.
 Answer: A representative sample reflects the population and avoids bias
 Solution:
 Step 1: Representative sample includes all key groups in the population
 Step 2: Ensures accurate and fair results
 Step 3: Helps generalise findings to the full population

Q99. A scatter graph shows the values: (2, 4), (3, 6), (4, 8), (5, 10), (6, 12)
 Describe the type of correlation and what it tells you.
 Answer: Strong positive correlation

Solution:

Step 1: As x increases, y increases

Step 2: Points lie exactly on a straight line

Step 3: Tells us variables increase together in a fixed ratio

Q100. A national survey is too expensive to conduct for the entire population. Explain why a sample is used and how to ensure it's unbiased.

Answer: A sample is cheaper and quicker; it must be randomly selected

Solution:

Step 1: Whole population takes time and money

Step 2: A sample gives similar results if selected well

Step 3: Random selection avoids bias and covers all groups fairly