S1·1

Representation of data

CHAPTER ASSIGNMENT · 33 MARKS

Name _____ Class _____ No. _____ Date _____

Answer every question in the space given, showing all your working. State appropriate units and degrees of accuracy. Give non-exact values correct to 3 significant figures unless stated otherwise.

Practice questions from *Cambridge International AS & A Level Mathematics: Probability & Statistics 1 Coursebook* (Cambridge University Press, Dean Chalmers), Chapter 1 (Representation of data), Exercises 1A to 1D and End-of-chapter review exercise 1, for classroom use. Where a question needs a diagram, data or descriptions are provided. The Check yourself and Chapter quiz questions are Guru Jazzy's own.

Exercise 1A · Stem-and-leaf diagrams

1A.1 Twenty people leaving a cinema are each asked, "How many times have you attended the cinema in the past year?" Their responses are: 6, 2, 13, 1, 4, 8, 11, 3, 4, 16, 7, 20, 13, 5, 15, 3, 12, 9, 26 and 10. Construct a stem-and-leaf diagram for these data and include a key.

1A.2 A shopkeeper takes 12 bags of coins to the bank. The bags contain the following numbers of coins: 150, 163, 158, 165, 172, 152, 160, 170, 156, 162, 159 and 175.

(a) Represent this information in a stem-and-leaf diagram.

(b) Each bag contains coins of the same value, and the shopkeeper has at least one bag containing coins with dollar values of 0.10, 0.25, 0.50 and 1.00 only. What is the greatest possible value of all the coins in the 12 bags?

1A.3 This stem-and-leaf diagram shows the number of employees at 20 companies:

Stem	Leaf
---	---
1	0 8 8 8 8 9 9
2	0 5 6 6 7 7 8 9
3	0 1 1 2 9

Key: 1 | 0 represents 10 employees.

(a) What is the most common number of employees?

(b) How many of the companies have fewer than 25 employees?

(c) What percentage of the companies have more than 30 employees?

(d) Determine which of the three rows in the stem-and-leaf diagram contains the smallest number of: (i) companies (ii) employees.

1A.4 Over a 14-day period, data were collected on the number of passengers travelling on two ferries, A and Z:

Ferry A (14)	Stem	Ferry Z (14)
---	---	---
8 7 6	2	
7 6 4 0	3	0 5 8
8 6 5 3	4	3 4 5 7 7 7
5 3 3	5	0 2 6 6 9

Key: 3 | 5 | 0 represents 53 passengers on A and 50 passengers on Z.

(a) How many more passengers travelled on ferry Z than on ferry A?

(b) The cost of a trip on ferry A is \$12.50 and the cost of a trip on ferry Z is \$x. The takings on ferry Z were \$3.30 less than the takings on ferry A over this period. Find the value of x.

(c) Find the least and greatest possible number of days on which the two ferries could have carried exactly the same number of passengers.

1A.5 The runs scored by two batsmen in 15 cricket matches last season were:

Batsman P: 53, 41, 57, 38, 41, 37, 59, 48, 52, 39, 47, 36, 37, 44, 59.

Batsman Q: 56, 48, 31, 64, 21, 52, 45, 36, 57, 68, 77, 20, 42, 51, 71.

(a) Show the data in a diagram that allows easy comparison of the two performances.

(b) Giving a reason for your answer, decide which of the batsmen performed: (i) better (ii) more consistently.

1A.6 The total numbers of eggs laid in the nests of two species of bird were recorded over several breeding seasons. The numbers of eggs laid in the nests of 10 wrens and 10 dunnocks are:

Wrens: 22, 18, 21, 23, 17, 23, 20, 19, 24, 13.

Dunnocks: 28, 24, 23, 19, 30, 27, 22, 25, 22, 17.

(a) Represent the data in a back-to-back stem-and-leaf diagram with rows of width 5.

(b) Given that all of these eggs hatched and that the survival rate for dunnock chicks is 92%, estimate the number of dunnock chicks that survived.

(c) Find the survival rate for the wren chicks, given that 14 did not survive.

1A.7 This back-to-back stem-and-leaf diagram shows the percentage scores of the 25 students who were the top performers in an examination:

Girls (12)	Stem	Boys (13)
---	---	---
4 1	8	2
8 6 6	8	5 9
3 2 1 0	9	0 1 3 3 4 4
8 7 7	9	5 6 6 9

Key: 1 | 8 | 2 represents 81% for a girl and 82% for a boy.

The 25 students are arranged in a line in the order of their scores. Describe the student in the middle of the line and find the greatest possible number of boys in the line who are not standing next to a girl.

Exercise 1B · Histograms and frequency density

1B.1 In a particular city there are 51 buildings of historical interest. The following table presents the ages of these buildings, given to the nearest 50 years:

Age (years)	50–150	200–300	350–450	500–600
No. buildings (f)	15	18	12	6

(a) Write down the lower and upper boundary values of the class containing the greatest number of buildings.

(b) State the widths of the four class intervals.

(c) Illustrate the data in a histogram.

(d) Estimate the number of buildings that are between 250 and 400 years old.

1B.2 The masses, m grams, of 690 medical samples are given in the following table:

Mass (m grams)	$4 \le m < 12$	$12 \le m < 24$	$24 \le m < 28$
No. medical samples (f)	224	396	p

(a) Find the value of p that appears in the table.

(b) On graph paper, draw a histogram to represent the data.

(c) Calculate an estimate of the number of samples with masses between 8 and 18 grams.

1B.3 The table below shows the heights, in metres, of 50 boys and of 50 girls:

Height (m)	1.2–	1.3–	1.6–	1.8–1.9
No. boys (f)	7	11	26	6
No. girls (f)	10	22	16	2

(a) How many children are between 1.3 and 1.6 metres tall?

(b) Draw a histogram to represent the heights of all the boys and girls together.

(c) Estimate the number of children whose heights are 1.7 metres or more.

1B.4 The heights of 600 saplings are shown in the following table:

Height (cm)	0–	5–	15–	30–u
No. saplings (f)	64	232	240	64

(a) Suggest a suitable value for u, the upper boundary of the data.

(b) Illustrate the data in a histogram.

(c) Calculate an estimate of the number of saplings with heights that are: (i) less than 25 cm (ii) between 7.5 and 19.5 cm.

1B.5 Each of the 70 trainees at a secretarial college was asked to type a copy of a particular document. The times taken are shown, correct to the nearest 0.1 minutes, in the following table:

Time taken (min)	2.6–2.8	2.9–3.0	3.1–3.2	3.3–3.7
No. trainees (f)	15	25	20	10

(a) Explain why the interval for the first class has a width of 0.3 minutes.

(b) Represent the times taken in a histogram.

(c) Estimate, to the nearest second, the upper boundary of the times taken by the fastest 10 typists.

(d) It is given that 15 trainees took between 3.15 and b minutes. Calculate an estimate for the value of b when: (i) $b > 3.15$ (ii) $b < 3.15$.

1B.6 A railway line monitored 15% of its August train journeys to find their departure delay times. The results are shown in a histogram with intervals 0–2 min (fd = 12), 2–4 min (fd = 18), 4–12 min (fd = 16), 12–18 min (fd = 20), and 18–20 min (fd = 8). It is given that 24 of these journeys were delayed by less than 2 minutes.

(a) How many journeys were monitored?

(b) Calculate an estimate of the number of these journeys that were delayed by: (i) 1 to 3 minutes (ii) 10 to 15 minutes.

(c) Show that a total of 2160 journeys were provided in August.

(d) Calculate an estimate of the number of August journeys that were delayed by 3 to 7 minutes. State any assumptions that you make in your calculations.

1B.7 A university investigated how much space on its computers' hard drives is used for data storage. The results are shown in a histogram with intervals 0–20 GB (fd = 2), 20–60 GB (fd = 3), 60–120 GB (fd = 4), and 120–200 GB (fd = 1). It is given that 40 hard drives use less than 20 GB for data storage.

(a) Find the total number of hard drives represented.

(b) Calculate an estimate of the number of hard drives that use less than 50 GB.

(c) Estimate the value of k, if 25% of the hard drives use k GB or more.

1B.8 The lengths of the 575 items in a candle maker's workshop are represented in a histogram with intervals 0–10 cm (fd = 5), 10–25 cm (fd = 25), 25–30 cm (fd = 10), and 30–50 cm (fd = 5).

(a) What proportion of the items are less than 25 cm long?

(b) Estimate the number of items that are between 12.4 and 36.8 cm long.

(c) The shortest 20% of the workshop's items are to be recycled. Calculate an estimate of the length of the shortest item that will not be recycled.

1B.9 The thicknesses, k mm, of some steel sheets are represented in a histogram with intervals 0.1–0.4 mm (fd = 4), 0.4–0.8 mm (fd = 2), and 0.8–1.1 mm (fd = 1). It is given that $k < 0.4$ for 180 sheets.

(a) Find the ratio between the frequencies of the three classes. Give your answer in simplified form.

(b) Find the value of n, given that frequency density measures sheets per n mm.

(c) Calculate an estimate of the number of sheets for which: (i) $k < 0.5$ (ii) $0.75 \leq k < 0.94$.

(d) The sheets are classified as thin, medium or thick in the ratio 1 : 3 : 1. Estimate the thickness of a medium sheet, giving your answer in the form $a \leq k < b$. How accurate are your values for a and b?

1B.10 The masses, in kilograms, of the animals treated at a veterinary clinic in the past year are illustrated in a histogram. The histogram has four columns of equal height. The following table shows the class intervals and the number of animals in two of the classes:

Mass (kg)	3–5	6–12	13–32	33–44
No. animals (f)	a	371	1060	b

(a) Find the value of a and of b, and show that a total of 2226 animals were treated at the clinic.

(b) Calculate an estimate of the lower boundary of the masses of the heaviest 50% of these animals.

1B.11 The minimum daily temperature at a mountain village was recorded to the nearest 0.5°C on 200 consecutive days. The results are grouped into a frequency table and a histogram is drawn. The temperatures ranged from 0.5°C to 2°C on n days, and this class is represented by a column of height h cm. The temperatures ranged from −2.5°C to −0.5°C on d days. Find, in terms of n, h and d, the height of the column that represents these temperatures.

1B.12 The frequency densities of the four classes in a histogram are in the ratio 4 : 3 : 2 : 1. The frequencies of these classes are in the ratio 10 : 15 : 24 : 8. Find the total width of the histogram, given that the narrowest class interval is represented by a column of width 3 cm.

1B.13 The percentage examination scores of 747 students are given in the following table:

Score (%)	p–50	51–70	71–80	81–q
No. students (f)	165	240	195	147

Given that the frequency densities of the four classes of percentage scores are in the ratio 5 : 8 : 13 : 7, find the value of p and of q.

Exercise 1C · Cumulative frequency graphs

1C.1 The reaction times, t seconds, of 66 participants were measured in an experiment and presented below:

Time (t seconds)	t < 1.5	t < 3.0	t < 4.5	t < 6.5	t < 8.5	t < 11.0	t < 13.0
No. participants (cf)	0	3	8	32	54	62	66

(a) Draw a cumulative frequency polygon to represent the data.

(b) Use your graph to estimate: (i) the number of participants with reaction times between 5.5 and 7.5 seconds (ii) the lower boundary of the slowest 20 reaction times.

1C.2 The following table shows the widths of the 70 books in one section of a library, given to the nearest centimetre:

Width (cm)	10–14	15–19	20–29	30–39	40–44
No. books (f)	3	13	25	24	5

(a) Given that the upper boundary of the first class is 14.5 cm, write down the upper boundary of the second class.

(b) Draw up a cumulative frequency table for the data and construct a cumulative frequency graph.

(c) Use your graph to estimate: (i) the number of books that have widths of less than 27 cm (ii) the widths of the widest 20 books.

1C.3 Measurements of the distances, x mm, between two moving parts inside car engines were recorded and are summarised in the following table. There were 156 engines of type A and 156 engines of type B.

Distance x (mm)	x < 0.10	x < 0.35	x < 0.60	x < 0.85	x < 1.20
Engine A (cf)	0	16	84	134	156
Engine B (cf)	0	8	52	120	156

(a) Draw and label two cumulative frequency curves on the same axes.

(b) Use your graphs to estimate: (i) the number of engines of each type with measurements between 0.30 and 0.70 mm (ii) the total number of engines with measurements that were less than 0.55 mm.

(c) Both types of engine must be repaired if the distance between these moving parts is more than a certain fixed amount. Given that 16 type A engines need repairing, estimate the number of type B engines that need repairing.

1C.4 The diameters, d cm, of 60 cylindrical electronic components are represented in a cumulative frequency graph passing through the points (0.05, 0), (0.1, 5), (0.2, 20), (0.3, 31), (0.4, 37), (0.5, 52), and (0.6, 60).

(a) Find the number of components such that $0.2 \leq d < 0.4$, and explain why your answer is not an estimate.

(b) Estimate the number of components that have: (i) a diameter of less than 0.15 cm (ii) a radius of 0.16 cm or more.

(c) Estimate the value of k, given that 20% of the components have diameters of k mm or more.

(d) Give the reason why 0.1–0.2 cm is the modal class.

1C.5 A cumulative frequency graph shows the masses, m grams, of 152 uncut diamonds, with curve starting at (2, 0) and ending at P(24, 152).

(a) Estimate the number of uncut diamonds with masses such that: (i) $9 \leq m < 17$ (ii) $7.2 \leq m < 15.6$.

(b) The lightest 40 diamonds are classified as small. The heaviest 40 diamonds are classified as large. Estimate the difference between the mass of the heaviest small diamond and the lightest large diamond.

(c) The point marked at P(24, 152) on the graph indicates that the 152 uncut diamonds all have masses of less than 24 grams. Each diamond is now cut into two parts of equal mass. Assuming that there is no wastage of material, write down the coordinates of the point corresponding to P on a cumulative frequency graph representing the masses of these cut diamonds.

1C.6 The densities, d g/cm³, of 125 chemical compounds are given in the following table:

Density d (g/cm³)	d < 1.30	d < 1.38	d < 1.42	d < 1.58	d < 1.70
No. compounds (cf)	32	77	92	107	125

Find the frequencies a, b, c, and d given in the table below:

Density d (g/cm³)	0–	1.30–	1.38–	1.42–1.70
No. compounds (f)	a	b	c	d

1C.7 The daily journey times for 80 bank staff to get to work are given in the following table:

Time (t min)	t < 10	t < 15	t < 20	t < 25	t < 30	t < 45	t < 60
No. staff (cf)	3	11	24	56	68	76	80

(a) How many staff take between 15 and 45 minutes to get to work?

(b) Find the exact number of staff who take $(x + y)/2$ minutes or more to get to work, given that 85% of the staff take less than x minutes and that 70% of the staff take y minutes or more.

1C.8 A fashion company selected 100 12-year-old boys and 100 12-year-old girls to audition as models. The heights, h cm, are represented in a cumulative frequency graph where the boys' polygon connects (145, 0), (155, 75), and (175, 100), and the girls' polygon connects (140, 0), (155, 25), and (180, 100).

(a) What features of the data suggest that the children were not selected at random?

(b) Estimate the number of girls who are taller than the shortest 50 boys.

(c) What is the significance of the value of h where the graphs intersect?

(d) The shortest 75 boys and tallest 75 girls were recalled for a second audition. On a cumulative frequency graph, show the heights of the children who were not recalled.

1C.9 The following table shows the ages of the students currently at a university, given by percentage. Ages are rounded down to the number of whole years:

Age (years)	< 18	18–19	20–21	22–24	25–28	29–35	36–44
Students (%)	0	27	51	11	5	4	2

(a) Represent the data in a percentage cumulative frequency polygon.

(b) The oldest 8% of these students qualify as ‘mature’. Use your polygon to estimate the minimum age requirement for a student to be considered mature. Give your answer to the nearest month.

(c) Of the 324 students who are 18–19 years old, 54 are not expected to find employment within 3 months of finishing their course. (i) Calculate an estimate of the number of current students who are expected to find employment within 3 months of finishing their course. (ii) What assumptions must be made to justify your calculations in part (c)(i)? Are these assumptions reasonable? Do you expect your estimate to be an overestimate or an underestimate?

1C.10 The distances, in km, that 80 new cars can travel on 1 litre of fuel are shown in the table:

Distance (km)	4.4–	6.6–	8.8–	12.1–	15.4–18.7
No. cars (f)	5	7	52	12	4

These distances are 10% greater than the distances the cars will be able to travel after they have covered more than 100000 km. Estimate how many of the cars can travel 10.5 km or more on 1 litre of fuel when new, but not after they have covered more than 100000 km.

1C.11 A small company produces cylindrical wooden pegs for making garden chairs. The lengths and diameters of the 242 pegs produced yesterday have been measured independently by two employees, and their results are given in the following table:

Length (l cm)	$l < 1.0$	$l < 2.0$	$l < 2.5$	$l < 3.0$	$l < 3.5$	$l < 4.0$	$l < 4.5$
No. pegs (cf)	0	0	8	40	110	216	242
Diameter (d cm)	$d < 1.0$	$d < 1.5$	$d < 2.0$	$d < 2.5$	$d < 3.0$		
No. pegs (cf)	0	60	182	222	242		

(a) On the same axes, draw two cumulative frequency graphs: one for lengths and one for diameters.

(b) Correct to the nearest millimetre, the lengths and diameters of n of these pegs are equal. Find the least and greatest possible value of n.

(c) A peg is acceptable for use when it satisfies both $l \geq 2.8$ and $d < 2.2$. Explain why you cannot obtain from your graphs an accurate estimate of the number of these 242 pegs that are acceptable. Suggest what the company could do differently so that an accurate estimate of the proportion of acceptable pegs could be obtained.

Exercise 1D · Comparing data representations

1D.1 Jamila noted each student's answer when her year group was asked to name their favourite colour.

(a) List the methods of representation that would be suitable for displaying Jamila's data.

(b) Jamila wishes to emphasise that the favourite colour of exactly three-quarters of the students is blue. Which type of representation from your list do you think would be the most effective for Jamila to use? Explain why you have chosen this particular type of representation.

1D.2 A large number of chickens' eggs are individually weighed. The masses are grouped into nine classes, each of width 2 grams, from 48 to 66 g. Name a type of representation in which the fact could be seen that the majority of the eggs have masses from 54 to 60 g. Explain how the representation would show this.

1D.3 Boxes of floor tiles are to be offered for sale at a special price of $75. The boxes claim to contain at least 100 tiles each.

(a) Why would it be preferable to use a stem-and-leaf diagram rather than a bar chart to represent the numbers of tiles, which are 112, 116, 107, 108, 121, 104, 111, 106, 105 and 110?

(b) How may the seller benefit if the numbers 12, 16, 7, 8, 21, 4, 11, 6, 5 and 10 are used to draw the stem-and-leaf diagram instead of the actual numbers of tiles?

1D.4 A charity group's target is to raise a certain amount of money in a year. At the end of the first month the group raised 36% of the target amount, and at the end of each subsequent month they manage to raise exactly half of the amount outstanding.

(a) How many months will it take the group to raise 99% of the money?

(b) Name a type of representation that will show that the group fails to reach its target by the end of the year. Explain how this fact would be shown in the representation.

1D.5 University students measured the heights of the 54 trees in the grounds of a primary school. The data are presented in a histogram with classes 2–3 m (fd = 12), 3–5 m (fd = 9), and 5–8 m (fd = 8), and also in a pie chart with sectors short (2–3 m), medium (3–5 m), and tall (5–8 m).

(a) Give one disadvantage of using each of the representations shown.

(b) Name and describe a different type of representation that would be appropriate for the audience, and that has none of the disadvantages given in part (a).

1D.6 The percentage scores of 40 candidates who took a Health and Safety test are given:

77 44 65 84 52 60 35 83 68 66 50 68 65 57 60 50 93 38 46 55

45 69 61 64 40 66 91 59 61 74 70 75 42 65 85 63 73 84 68 30

(a) Construct a frequency table by grouping the data into seven classes with equal-width intervals, where the first class is 30–39. Label this as Table 1.

(b) It is proposed that each of the 40 candidates is awarded one of three grades, A, B or C. Construct a new frequency table that matches with this proposal. Label this as Table 2.

(c) A student plans to display all three versions of the data (i.e. the raw data, the data in Table 1 and the data in Table 2) in separate stem-and-leaf diagrams. For which version(s) of the data would this not be appropriate? Suggest an alternative type of representation in each case.

1D.7 Last year Tom renovated an old building during which he worked for at least 9 hours each week. By plotting four points in a graph, (34, 49), (39, 50), (42, 51), (44, 52), where the vertical cf axis starts at 49 instead of 0, he has represented the time he spent working.

(a) What can you say about the time that Tom spent working on the basis of this graph?

(b) Explain why Tom's graph might be considered to be misleading.

(c) Name the different types of representation that are suitable for displaying the amount of time that Tom worked each week throughout the year. Consider the benefits of each type of representation and then fully describe (but do not draw) the one you believe to be the most suitable.

1D.8 The following table shows the focal lengths, l mm, of the 84 zoom lenses sold by a shop. For example, there are 18 zoom lenses that can be set to any focal length between 24 and 50 mm:

Focal length l (mm)	24–50	50–108	100–200	150–300	250–400
No. lenses (f)	18	30	18	12	6

(a) What feature of the data does not allow them to be displayed in a histogram?

(b) What type of diagram could you use to illustrate the data? Explain clearly how you would do this.

1D.9 The following table shows estimates, in hundred thousands, of the number of people living in poverty and the populations, in millions, of the countries where they live (Source: World Bank 2011/12):

Country	Chile	Sri Lanka	Malaysia	Georgia	Mongolia
No. living in poverty ($\times 10^5$)	24.8	18.4	10.8	7.91	7.50
Population ($\times 10^6$)	17.25	20.65	28.33	4.47	2.74

Represent, in a single diagram, the actual numbers and the relative poverty that exists in these countries. In what way do the two sets of data in your representation give very different pictures of the poverty levels that exist? Which is the better representation to use and why?

End-of-chapter review 1

R1.1 The weights of 220 sausages are summarised in the following table:

Weight (grams)	< 20	< 30	< 40	< 45	< 50	< 60	< 70
Cumulative frequency	0	20	50	100	160	210	220

(i) State how many sausages weighed between 50 g and 60 g. [1]

(ii) On graph paper, draw a histogram to represent the weights of the sausages. [4]



R1.2 The lengths of children's feet, in centimetres, are classified as 14–16, 17–19, 20–22 and so on. State the lower class boundary, the class width and the class mid-value for the lengths given as 17–19. [2]

R1.3 The capacities of ten engines, in litres, are given rounded to 2 decimal places, as follows: 1.86, 2.07, 1.74, 1.08, 1.99, 1.96, 1.83, 1.45, 1.28 and 2.19. These capacities are to be grouped in three classes as 1.0–1.4, 1.5–2.0 and 2.1–2.2.

(a) Find the frequency of the class 1.5–2.0 litres. [1]

(b) Write down two words that describe the type of data given about the engines. [2]

R1.4 Over a 10-day period, Alina recorded the number of text messages she received each day. The following stem-and-leaf diagram shows her results:

Stem	Leaf
---	---
0	9
1	0 3 4 4
1	5 5 5 6 8

Key: 1 | 5 represents 15 messages.

(a) On how many days did she receive more than 10 but not more than 15 messages? [1]

(b) How many more rows would need to be added to the stem-and-leaf diagram if Alina included data for two extra days on which she received 4 and 36 messages? Explain your answer. [2]

R1.5 The following stem-and-leaf diagram shows eight randomly selected numbers between 2 and 4:

Stem	Leaf
---	---
2	1 3 3 a
3	b 5 6 7

Key: 2 | 1 represents 2.1.

Given that a − b = 7 and that the sum of the eight numbers correct to the nearest integer is 24, find the value of a and of b. [3]

R1.6 Eighty people downloaded a particular application and recorded the time taken for the download to complete. The times are given in the following table:

Download time (min)	< 3	< 5	< 6	< 10
No. downloads (cf)	6	18	66	80

(a) Find the number of downloads that completed in 5 to 6 minutes. [1]

(b) On a histogram, the download times from 5 to 6 minutes are represented by a column of height 9.6 cm. Find the height of the column that represents the download times of 6 to 10 minutes. [2]

R1.7 A histogram is drawn with three columns whose widths are in the ratio 1 : 2 : 4. The frequency densities of these classes are in the ratio 16 : 12 : 3, respectively.

(a) Given that the total frequency of the data is 390, find the frequency of each class. [3]

(b) The classes with the two highest frequencies are to be merged and a new histogram drawn. Given that the height of the column representing the merged classes is to be 30 cm, find the correct height for the remaining column. [3]

(c) Explain what problems you would encounter if asked to construct a histogram in which the classes with the two lowest frequencies are to be merged. [1]

R1.8 The histograms below illustrate the number of hours of sunshine during August (31 days) in two regions, A and B. Neither region had more than 8 hours of sunshine per day:

Region A: 0–3 h (fd = 6), 3–4 h (fd = 5), 4–5 h (fd = 3), 5–6 h (fd = 2), 6–8 h (fd = 1).

Region B: 0–1 h (fd = 3), 1–2 h (fd = 0), 2–3 h (fd = 3), 3–4 h (fd = 4), 4–5 h (fd = 3), 5–8 h (fd = 6).

(a) Explain how you know that some information for one of the regions has been omitted. [2]

(b) After studying the histograms, two students make the following statements:

Bindu: There was more sunshine in region A than in region B during the first 2 weeks of August.

Janet: In August there was less sunshine in region A than in region B.

Discuss these statements and decide whether or not you agree with each of them. In each case, explain your reasoning. [3]

R1.9 A hotel has 90 rooms. The table summarises information about the number of rooms occupied each day for a period of 200 days:

Number of rooms occupied	1–20	21–40	41–50	51–60	61–70	71–90
Frequency	10	32	62	50	28	18

(i) Draw a cumulative frequency graph on graph paper to illustrate this information. [4]

(ii) Estimate the number of days when more than 30 rooms were occupied. [2]

(iii) On 75% of the days at most n rooms were occupied. Estimate the value of n. [2]

Check yourself

C1 Distinguish clearly between discrete quantitative data and continuous quantitative data, providing two everyday examples of each.

C2 The lengths of 120 metal rods are measured to the nearest millimetre and grouped into classes: 20–29 mm (frequency 24), 30–44 mm (frequency 45), and 45–69 mm (frequency 51). Find the true class boundaries, class widths, and frequency densities for each of the three classes.

C3 The heights, in centimetres, of 15 girls and 15 boys in a sports club are recorded. Construct a back-to-back stem-and-leaf diagram with rows of width 10, determine the median height for each group, and state which group has greater variation.

Girls: 152, 155, 158, 160, 162, 163, 165, 167, 168, 170, 171, 172, 174, 176, 180.

Boys: 156, 159, 161, 164, 166, 168, 170, 171, 173, 175, 177, 181, 184, 186, 192.

C4 A grouped frequency table records the waiting times of 80 patients at a clinic: 0–10 min (12 patients), 10–20 min (28 patients), 20–35 min (24 patients), and 35–50 min (16 patients). Construct the cumulative frequency table, state the coordinates of all plotted points on a cumulative frequency graph, and estimate the median waiting time.

C5 A dataset has lower quartile $Q_1 = 42$ and upper quartile $Q_3 = 66$.

(a) Calculate the interquartile range (IQR).

(b) Determine the lower and upper outlier boundaries using the 1.5 × IQR rule.

(c) State whether the values 3 and 105 are classified as outliers.

Chapter quiz

Q1 Which of the following is an example of continuous quantitative data?

A	B	C	D
Number of students in a class	Daily maximum temperature in degrees Celsius	Shoe sizes sold in a footwear shop	Number of goals scored in a football match

Q2 Exam marks are rounded to the nearest integer. What are the true boundaries for the class 60–69?

A	B	C	D
$59.5 \le \text{mark} < 69.5$	$60.0 \le \text{mark} < 69.0$	$59.5 \le \text{mark} \le 69.5$	$60.5 \le \text{mark} < 70.5$

Q3 What is the frequency density of a class with interval 20–50 and frequency 75?

A	B	C	D
2.5	2500	2250	0.4

Q4 In a histogram, what does the area of each column represent?

A	B	C	D
The class width	The cumulative frequency	A value proportional to the class frequency	The frequency density

---

Q5 A histogram bar has width 4 cm and height 6 cm. If 1 cm² represents 5 items, find the class frequency.

A	B	C	D
24	120	48	30

Q6 In a stem-and-leaf diagram with key 3 | 5 = 35, why must the leaves in each row be equally spaced?

A	B	C	D
To make row lengths act like a bar chart representing frequency	To ensure every leaf value is positive	To calculate the mean directly from leaf positions	Because leaves cannot exceed 9

Q7 When constructing a cumulative frequency graph from grouped data, at which points are frequencies plotted?

A	B	C	D
At the class mid-points	At the upper class boundaries	At the lower class boundaries	At the class widths

---

Q8 For 120 data values, at which cumulative frequency is the lower quartile Q_1 located on a curve?

A	B	C	D
30	40	60	90

Q9 A dataset has $Q_1 = 30$ and $Q_3 = 42$. Using the 1.5 × IQR rule, which of these values is an outlier?

A	B	C	D
8	15	36	55

Q10 A box plot has a much longer whisker on the right than on the left. What does this indicate about the data?

A	B	C	D
Positively skewed (skewed to the right)	Negatively skewed (skewed to the left)	Symmetrical distribution	The interquartile range is zero

---

Q11 Why cannot data with overlapping class intervals (such as 20–40 and 30–50) be displayed in a histogram?

A	B	C	D
Values in the overlap would be double-counted with ambiguous columns	The area of the columns would sum to zero	Frequency density cannot be calculated for intervals of equal width	Histograms only allow integer frequencies

---

Q12 What is a major advantage of a stem-and-leaf diagram over a histogram?

A	B	C	D
The original raw data values are retained and can be retrieved	It can represent thousands of data values without clutter	It automatically calculates the standard deviation	It does not require equal class widths

---

End of assignment · Check your answers against the chapter at gurujazzy.com/subjects/as-a-level-mathematics/s1-1-representation-of-data/