Year 11 Mathematics Advanced

Topic: Probability and Data (4 of 5)

Review Questions (Worked Solutions): Data

Term 3, 2026

Data

1. Frequency tables, mode and median

Q1 Killara 2022 Prelim Q2 · **1 mark**

A spinner is spun 50 times: Red 13, Blue 12, Yellow 10, Green 15. Find the relative frequency of spinning a red or green.

Criteria	Marks
Correct option (A)	1

Solution.

Red or green occurs 13 + 15 = 28 times out of 50 spins.

$$\text{Relative frequency} = \frac{28}{50} = \frac{14}{25} = 0.56.$$

The answer is **(A)**.

Q2 North Sydney Girls 2024 Trial Q6 · **1 mark**

From the dot plot, choose the row giving a correct statistical measure and the correct shape.

Criteria	Marks
Correct option (B)	1

Solution.

The data are clustered at the higher values with a tail running towards the lower values, so the distribution is **negatively skewed**. This immediately rules out (A) and (C), which both claim a positive skew. The range of the data is clearly greater than 4 (the values span from the smallest to the largest score), so (D) is incorrect. The remaining row, with the median at 7 and a negative skew, is consistent with the plot.

$$\text{Median} = 7, \qquad \text{Negatively skewed.}$$

The answer is **(B)**.

Q3 Caringbah 2024 Trial Q9 · **1 mark**

From the cumulative frequency table of student heights (total 800), find the height class of a student in the 7th decile.

Criteria	Marks
Correct option (C)	1

Solution.

The total number of students is 800. The 7th decile is the value below which $\frac{7}{10}$ of the data lie:

$$0.7 \times 800 = 560.$$

Reading the cumulative frequencies, the count first reaches or exceeds 560 in the class where the cumulative frequency jumps from 420 (at 151–160) to 680 (at 161–170). So the 560th student lies in the 161–170 cm class. The answer is **(C)**.

Q4 James Ruse 2022 Trial Q4 · **1 mark**

From the back-to-back stem and leaf plot of 24 test results, find the median and the mode of the whole class.

Criteria	Marks
Correct option (C)	1

Solution.

Combine the boys' and girls' scores into one ordered list of all 24 values. With 24 scores, the median is the average of the 12th and 13th values when the data are arranged in order:

$$\text{Median} = \frac{\text{12th value} + \text{13th value}}{2} = \frac{45 + 47}{2} = 46.$$

The most frequently occurring score across both sides of the plot is 49, so the mode is 49.

$$\text{Median} = 46, \qquad \text{Mode} = 49.$$

The answer is **(C)**.

Q5 Glenwood 2025 Trial Q1 · **1 mark**

Fourteen students' lolly counts are shown in a stem and leaf plot; two absent students are then added (total 16). The cumulative frequency of the 10–19 class is X. Find a possible value of X.

Criteria	Marks
Correct option (A)	1

Solution.

From the stem and leaf plot the 14 recorded students fall into the classes as follows:

$$0\text{–}9 : 4, \qquad 10\text{–}19 : 5, \qquad 20\text{–}29 : 4, \qquad 30\text{–}39 : 1.$$

The cumulative frequency at the end of the 10–19 class is currently $4 + 5 = 9$. Two more students are then added. Each added student in the 0–19 range increases this cumulative figure by 1; any added student of 20 or more leaves it unchanged. Hence

$$X \in \{9, 10, 11\},$$

depending on how many of the two extra students score under 20. The only listed value that lies in this set is 9 (both added students score 20 or more). The answer is **(A)**.

Q6 2022 HSC Q11 · **2 marks**

The complaint frequency table gives a cumulative frequency column (A for "Delivery fee") and a cumulative percentage column (B for "Damaged item"), out of a total of 200. Find A and B.

Criteria	Marks
Correct value $A = 160$	1
Correct value $B = 96$	1

Solution.

A is the running total of the frequencies up to and including "Delivery fee":

$$A = 98 + 62 = 160.$$

(Check: $\dfrac{160}{200} \times 100 = 80\%$, matching the given cumulative percentage.) B is the cumulative percentage at "Damaged item", whose cumulative frequency is 192:

$$B = \frac{192}{200} \times 100 = 96\%.$$

$$\boxed{A = 160, \qquad B = 96.}$$

Q7 Abbotsleigh 2023 Trial Q12 · **7 marks**

(a) Complete the frequency and cumulative frequency table of tree heights; (a)(ii) state the modal class; (a)(iii) find $P(0.45 \le h < 1.25)$. (b) For events with $P(A) = 0.4$, $P(B) = 0.55$, $P(B \mid A) = 0.6$: (i) test independence; (ii) find $P(A \cup B)$.

Criteria	Marks
(a)(i) completes the frequency and cumulative frequency columns correctly	2
(a)(ii) modal class 1.7–2.0 m	1
(a)(iii) $P = \frac{13}{43}$	1
(b)(i) shows $P(A \cap B) \neq P(A)P(B)$ (or $P(B \mid A) \neq P(B)$), so not independent	1
(b)(ii) computes $P(A \cap B) = 0.24$	1
(b)(ii) $P(A \cup B) = 0.71$	1

Solution.

(a)(i) The "Height less than" cumulative frequencies give a running total. Working through:

$$\text{cum}(< 0.45) = 2 \implies f(0.1\text{–}0.4) = 2.$$

$$\text{cum}(< 0.85) = 2 + 6 = 8.$$

$$\text{cum}(< 1.25) = 15 \implies f(0.9\text{–}1.2) = 15 - 8 = 7.$$

$$\text{cum}(< 1.65) = 15 + 11 = 26.$$

$$\text{cum}(< 2.05) = 40 \implies f(1.7\text{–}2.0) = 40 - 26 = 14.$$

$$\text{cum}(< 2.45) = 40 + 3 = 43 \quad \text{(total number of trees)}.$$

The completed tables are:

Height (m)	Frequency	Height less than	Cumulative Frequency
0.1–0.4	2	0.45	2
0.5–0.8	6	0.85	8
0.9–1.2	7	1.25	15
1.3–1.6	11	1.65	26
1.7–2.0	14	2.05	40
2.1–2.4	3	2.45	43

(a)(ii) The modal class is the class with the greatest frequency. The largest frequency is 14, so the modal class is $\boxed{\text{1.7–2.0 m}}$.

(a)(iii) "Less than 1.25 m but not less than 0.45 m" means $0.45 \le h < 1.25$, which covers the 0.5–0.8 and 0.9–1.2 classes:

$$\text{count} = \text{cum}(< 1.25) - \text{cum}(< 0.45) = 15 - 2 = 13.$$

$$P(0.45 \le h < 1.25) = \frac{13}{43} \approx 0.302.$$

(b)(i) First find $P(A \cap B) = P(B \mid A)\,P(A) = 0.6 \times 0.4 = 0.24$. For independence we require $P(A \cap B) = P(A)\,P(B)$:

$$P(A)\,P(B) = 0.4 \times 0.55 = 0.22 \neq 0.24 = P(A \cap B).$$

(Equivalently, $P(B \mid A) = 0.6 \neq 0.55 = P(B)$.) Therefore A and B are **not independent**.

(b)(ii) Using the addition rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.55 - 0.24 = 0.71.$$

2. Cumulative-frequency graphs, histograms and polygons

Q8 Penrith 2022 Trial Q6 · **1 mark**

A cumulative frequency polygon shows the heights of 400 students. Read off the median height.

Criteria	**Marks**
Correct option (B)	1

Solution.

For a cumulative frequency polygon, the median is the value corresponding to half of the total frequency:

$$\tfrac{1}{2} \times 400 = 200.$$

Drawing a horizontal line across from a cumulative frequency of 200 until it meets the polygon, then reading straight down to the height axis, gives a height of about 156 cm. (At 150 cm the cumulative frequency is below 200, and at 160 cm it is above 200, so the median lies just past 150 cm.)

$$\text{Median height} \approx 156 \text{ cm}.$$

The answer is **(B)**.

Q9 North Sydney Girls 2025 Trial Q1 · 1 mark

Four frequency graphs are shown. Identify the one representing a negatively skewed distribution.

Criteria	Marks
Correct option (D)	1

Solution.

A distribution is **negatively skewed** when it has a longer tail towards the lower (left-hand) scores, so the bulk of the data (and the peak) sits at the higher scores on the right.

- Graph (A) is symmetric (a bell shape), so no skew.
- Graph (B) is uniform (flat), so no skew.
- Graph (C) has its tail to the right, so it is positively skewed.
- Graph (D) has its peak on the right and its tail running to the left, so it is negatively skewed.

The answer is **(D)**.

Q10 Hurlstone 2021 Trial Q7 · 1 mark

Four graphs are shown. Identify which one cannot be a cumulative frequency polygon.

Criteria	Marks
Correct option (D)	1

Solution.

A cumulative frequency total can only stay the same or increase as the variable increases, since adding more scores can never reduce the running total. Hence a genuine cumulative frequency polygon must be **non-decreasing** (it never goes down as you move to the right).

- Graphs (A), (B) and (C) rise (or stay level) the whole way, so each is a possible cumulative frequency polygon.
- Graph (D) rises, then *falls* before rising again. A decreasing section is impossible for a cumulative frequency graph.

Therefore (D) CANNOT be a cumulative frequency polygon. The answer is **(D)**.

Q11 Hurlstone 2021 Trial Q15 · **1 mark**

A cumulative frequency histogram shows the number of hours worked per week by the respondents to Cole's survey. Find the range of responses giving the greatest 40% of hours worked.

Criteria	Marks
Correct range 36 to 44 hours per week (top 40% = 12 respondents)	1

Solution.

The total number of respondents is the height of the final bar, which is 30. The "greatest 40% of hours worked" are the respondents who work the most hours, i.e. the top 40%:

$$0.4 \times 30 = 12 \text{ respondents.}$$

Reading the cumulative frequencies from the histogram (bars of width 4), the running total reaches 18 at 36 hours. So $30 - 18 = 12$ respondents work more than 36 hours, which is exactly the top 40%. These respondents fall in the two highest classes, 36–40 and 40–44 hours.

$$\text{Range} = 36 \text{ to } 44 \text{ hours per week.}$$

Q12 Killara 2023 Trial Q4 · **1 mark**

A frequency histogram is given. Identify the corresponding cumulative frequency histogram.

Criteria	Marks
Correct option (B)	1

Solution.

Read the frequencies off the given histogram and form the running (cumulative) totals:

Class (centre)	Frequency	Cumulative frequency
5	2	2
10	4	6
15	8	14
20	3	17
25	3	20

The correct cumulative frequency histogram must have bars that never decrease, rising through the heights 2, 6, 14, 17, 20 and reaching the total of 20 at the last bar. Option (B) is the only histogram with these increasing bar heights ending at 20; (A) does not reach the total and is not increasing, and (C) is an ogive drawn against a percentage axis rather than a cumulative frequency histogram. The answer is **(B)**.

Q13 2021 HSC Q4 · **1 mark**

A bar graph shows the daily number of downloads of a song over twenty days. Identify the graph best showing the cumulative number of downloads.

Criteria	Marks
Correct option (C)	1

Solution.

The daily downloads start small, rise to a peak around day 10–11, then fall away. The cumulative total is found by adding each day's downloads to the running total, so it can only ever increase. Its shape reflects the daily rate:

- small daily amounts at the start $\Rightarrow$ the cumulative curve rises *slowly* at first;
- large daily amounts in the middle $\Rightarrow$ the cumulative curve rises *steeply* around days 9–12;
- small daily amounts at the end $\Rightarrow$ the cumulative curve *flattens* again.

This gives an S-shaped (steep-in-the-middle) increasing graph, which is option (C). The answer is **(C)**.

Q14 Trinity Grammar 2022 Trial Q29 · **2 marks**

A cumulative frequency histogram shows the number of alcoholic drinks per week consumed by 60 respondents. Find how many drinks per week the top decile of respondents consume.

Criteria	Marks
Identifies the top decile as the top 10% (6 respondents), i.e. the 90th percentile	1
Reads/interpolates the value: top decile consume more than about 7.6 drinks per week (up to 10)	1

Solution.

The total number of respondents is the height of the final bar, 60. The **top decile** is the top 10% of respondents (those consuming the most):

$$\tfrac{1}{10} \times 60 = 6 \text{ respondents.}$$

Equivalently, the top decile begins at the 90th percentile, i.e. the value below which 90% of the data lie:

$$0.9 \times 60 = 54.$$

From the histogram the cumulative frequency is 45 at 6 drinks and 56 at 8 drinks, so a cumulative frequency of 54 lies inside the 6–8 class. Interpolating across that bar,

$$x = 6 + \frac{54 - 45}{56 - 45} \times 2 = 6 + \frac{9}{11} \times 2 = 6 + \frac{18}{11} \approx 7.6.$$

So the top decile of respondents consume more than about 7.6 drinks per week (up to the maximum of 10).

3. Relative frequency and types of data

Q15 Hurlstone 2022 Prelim Q2 · **1 mark**

Choose which of the four quantities (kilometres per litre; number of motorists; number of passengers; number of refills per week) is *not* a discrete variable.

Criteria	Marks
Correct option (A)	1

Solution.

A discrete variable can only take separated (whole-number) values, whereas a continuous variable can take any value on a scale, measured to any level of accuracy. The number of motorists (B), the number of passengers (C) and the number of refills per week (D) are all counts, so they are discrete. The number of kilometres travelled per litre of petrol is a measurement that can take any value on a continuous scale, so it is **continuous, not discrete**.

(A) is a continuous variable.

The answer is **(A)**.

Q16 Sydney Girls 2023 Trial Q1 · **1 mark**

Choose which option is an example of discrete numerical data: the weight of baby goats; the month a birthday falls; the number of biscuits in a packet; a grade of A to D.

Criteria	Marks
Correct option (C)	1

Solution.

Numerical data is data that is a number; it is *discrete* when it can only take separated (counting) values. The weight of baby goats (A) is numerical but **continuous**. The month of a birthday (B) and a grade of A to D (D) are **categorical**, not numerical. Only the number of biscuits in a packet (C) is a count that takes whole-number values, so it is discrete numerical data.

Number of biscuits in a packet: discrete numerical.

The answer is **(C)**.

Q17 Hurlstone 2021 Prelim Q19 · **1 mark**

From the sibling frequency table (frequencies $2, 8, 5, 3, 1$ for 0 to 4 siblings), find the relative frequency of students who have 2 siblings.

Criteria	Marks
Correct option (B)	1

Solution.

The total number of students is the sum of all frequencies:

$$2 + 8 + 5 + 3 + 1 = 19.$$

The frequency of students with exactly 2 siblings is 5, so

$$\text{Relative frequency} = \frac{\text{frequency}}{\text{total}} = \frac{5}{19}.$$

The answer is **(B)**.

Q18 Barker 2022 Trial Q24 · **3 marks**

24 critics rate a film (0–5 stars) with frequencies $3, 4, 5, 7, 3, 2$. A partly filled two-way table records gender against "3 or more stars" / "less than 3 stars" (given: less-than-3 Female = 9, Male total = 11). (a) Complete the table. (b) Find $P(\text{male} \mid \text{3 or more stars})$.

Criteria	**Marks**
(a) correct row/column totals: "3 or more" = 12, "less than 3" = 12, Female total = 13	1
(a) correct interior cells: (3+ Male 8, Female 4); (less-than-3 Male 3, Female 9)	1
(b) $P(\text{male} \mid \text{3 or more}) = \frac{8}{12} = \frac{2}{3}$	1

Solution.

(a) From the frequency table, the number giving "3 or more stars" is

$$7 + 3 + 2 = 12,$$

and the number giving "less than 3 stars" is

$$3 + 4 + 5 = 12,$$

with an overall total of 24. The Male total is 11, so the Female total is $24 - 11 = 13$. In the "less than 3 stars" row, Female = 9, so Male = $12 - 9 = 3$. Down the Male column, "3 or more" Male = $11 - 3 = 8$. Down the Female column, "3 or more" Female = $13 - 9 = 4$. The completed table is:

	Male	Female	Total
3 or more stars	8	4	12
Less than 3 stars	3	9	12
Total	11	13	24

(b) Given the critic gave 3 or more stars, we restrict to the "3 or more stars" row (total 12), of whom 8 are male:

$$P(\text{male} \mid \text{3 or more stars}) = \frac{8}{12} = \frac{2}{3} \approx 0.667.$$

Q19 Hurlstone 2024 Prelim Q9 · **1 mark**

In a survey, 24 out of 30 students said they take a train to school. Based on this relative frequency, how many of the 960 students take a train?

Criteria	Marks
Correct option (B)	1

Solution.

The relative frequency of taking a train is $\frac{24}{30} = \frac{4}{5}$. Applying this proportion to the whole student body:

$$\frac{24}{30} \times 960 = \frac{4}{5} \times 960 = 768.$$

The estimate is 768, which is option **B**.

Q20 North Sydney Girls 2024 Trial Q13 · **3 marks**

From a Pareto chart of event costs (Venue \$8000, cumulative percentages rising 32%, 52%, 68%, ... across Venue, Food, Waitstaff): (a) estimate the percentage spent on Waitstaff; (b) find the total cost; (c) find the cost spent on Waitstaff.

Criteria	Marks
(a) Waitstaff percentage $\approx 68\% - 52\% = 16\%$ (accept $\pm 1\%$)	1
(b) total cost $= 8000 \div 0.32 = \$25\,000$	1
(c) Waitstaff cost $= 16\% \times \$25\,000 = \$4000$	1

Solution.

(a) On a Pareto chart the cumulative-percentage line gives the running total of the percentages. Reading the line, the cumulative percentage rises to 52% after Food and to 68% after Waitstaff, so the percentage for Waitstaff alone is the jump:

$$68\% - 52\% = 16\% \quad (\text{accept } \pm 1\% \text{ for reading the graph}).$$

(b) The Venue bar is \$8000 and its cumulative percentage is 32%, so \$8000 represents 32% of the total cost:

$$\text{Total} = \frac{8000}{0.32} = \$25\,000.$$

(c) Waitstaff is 16% of the total:

$$0.16 \times 25\,000 = \$4000.$$

Q21 Normanhurst Boys 2022 Trial Q20 · **3 marks**

30 Transit Officers issue infringements with frequencies $12, 5, 6, 3, 3, 1$ (for 0 to 5 infringements). (a) Find the median and describe the skewness. (b) For 120 officers, use the sample to estimate how many issue more than two infringements.

Criteria	**Marks**
(a) median = 1	1
(a) describes the shape as positively (right) skewed	1
(b) estimate = $\frac{7}{30} \times 120 = 28$ officers	1

Solution.

(a) There are 30 scores, so the median is the average of the 15th and 16th values in order. The cumulative frequency reaches 12 at "0 infringements" (positions 1–12) and 17 at "1 infringement" (positions 13–17), so both the 15th and 16th values are 1:

$$\text{Median} = \frac{1+1}{2} = 1.$$

Most of the data is bunched at the low values (0 and 1) with a tail stretching out towards the higher values $(3, 4, 5)$, so the distribution is **positively (right) skewed**.

(b) "More than two" infringements means 3, 4 or 5, with frequency $3 + 3 + 1 = 7$ out of 30. The relative frequency is $\frac{7}{30}$, so for 120 officers:

$$\frac{7}{30} \times 120 = 28 \text{ officers.}$$