DP Analysis & Approaches (SL/HL)

Topic 4.2 – Presentation of Grouped Data

Maths with Lemon

1. Types of data and frequency tables

Discrete data are countable values (e.g. grapes per bunch). **Continuous data** are measurements (e.g. height). A frequency table records the number of observations in each class. Continuous classes should have no gaps or overlaps, for example $149.5 \le h < 152.5$ and $152.5 \le h < 155.5$.

Grapes per bunch	Frequency	Height h (cm)	Frequency
1–3	2	$149.5 \le h < 152.5$	3
4–6	6	$152.5 \le h < 155.5$	8
7–9	7	$155.5 \le h < 158.5$	6
10–12	12	$158.5 \le h < 161.5$	12
13–15	10	$161.5 \le h < 164.5$	7
16–18	5	$164.5 \le h < 167.5$	4
> 19	1	$167.5 \le h < 170.5$	2
		$170.5 \le h < 173.5$	4
		$173.5 \le h < 176.5$	5

The original lesson uses equal-width intervals for its continuous-data histogram; frequency density histograms are not required here.

2. Histograms

For equal class widths, draw touching bars with heights equal to frequencies. Label both axes and show class boundaries accurately. A discrete frequency diagram may have separated bars. **Example:** the tallest continuous-data bar above is $158.5 \le h < 161.5$, with frequency 12.

3. Five-number summary, quartiles and box plots

The five-number summary is $\min, Q_1, Q_2, Q_3, \max$, where Q_2 is the median. The interquartile range is

$$\boxed{\text{IQR} = Q_3 - Q_1}, \qquad \text{range} = \max - \min .$$

For sorted ungrouped data, find the median, then the medians of the lower and upper halves (state the convention used). For a box plot, draw a box from Q_1 to Q_3, a median line at Q_2, and whiskers. If outliers are identified using the $1.5\,\text{IQR}$ rule, show them as crosses and end whiskers at the most extreme non-outliers. A symmetric box and whiskers can suggest approximate symmetry but does not prove normality.

4. Estimating the mean from grouped data

Let m_i be a class midpoint and f_i its frequency. Then

$$\boxed{\bar{x} \approx \frac{\sum f_i m_i}{\sum f_i}}.$$

Manila example: for temperature classes $[24,26), [26,28), [28,30), [30,32), [32,34), [34,36)$, use midpoints $25, 27, 29, 31, 33, 35$ and frequencies $1, 8, 32, 107, 174, 43$.

$$\bar{t} \approx \frac{25(1)+27(8)+29(32)+31(107)+33(174)+35(43)}{365} = \frac{11733}{365} \approx 32.1^{\circ}\text{C}.$$

5. Cumulative frequency (CF)

CF is a running total. Plot cumulative frequency against the *upper class boundary*, starting at the lower boundary with CF zero. Join the points with a smooth increasing curve (ogive).

Upper height boundary	152.5	155.5	158.5	161.5	164.5	167.5	170.5	173.5	176.5
CF	3	11	17	29	36	40	42	46	51

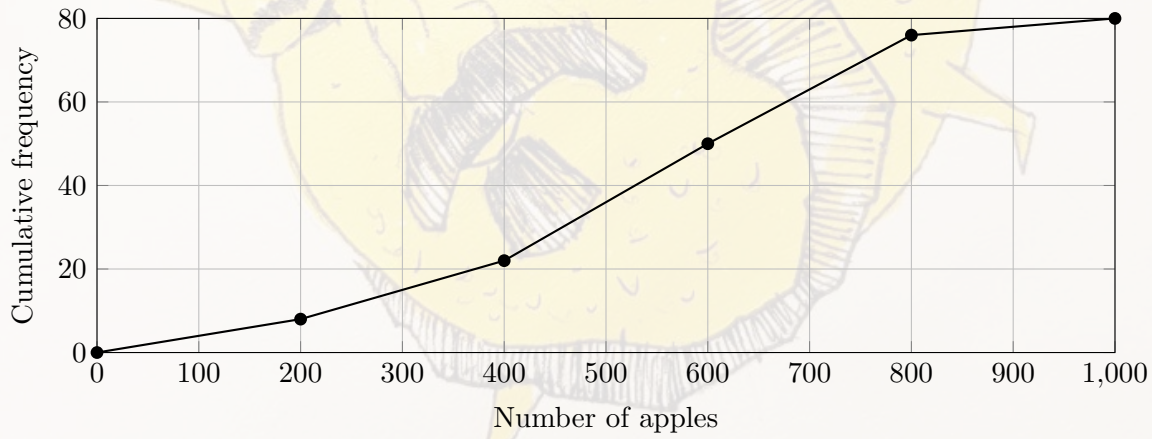
For N observations, read the median at $N/2$, Q_1 at $N/4$, and Q_3 at $3N/4$ on the vertical axis, then read the corresponding values horizontally. To estimate how many exceed x, calculate $N - \text{CF}(x)$.

6. Comparing two distributions

Compare **centre** (median), **spread** (IQR or range), and possibly **shape** (symmetry/skewness). For two hospitals, a larger median means a higher typical daily patient count; a larger IQR means greater variability in the middle half of days.

7. Worked cumulative-frequency example: apple trees

For frequencies $8, 14, 28, 26, 4$ in intervals $[0,200), [200,400), \ldots, [800,1000)$, the CF values are $8, 22, 50, 76, 80$. Plot $(0,0), (200,8), (400,22), (600,50), (800,76), (1000,80)$.



8. Exam tips

Show the class boundaries and cumulative totals. For grouped data, report estimates rather than exact values. For graph-reading questions, allow sensible rounding and state that the answer is approximate. For a box plot, make the scale uniform.



Maths with Lemon

Mathematics made clear.