DP Analysis & Approaches (SL/HL)

Topic 4.2 – Presentation of Grouped Data

Maths with Lemon

Mathematician: _____

Student worksheet

1. Warm-up: electricity consumption. Monthly consumption in kWh:

245, 268, 298, 321, 341, 376, 409, 431, 459, 486, 502, 529, 546, 573, 597, 622, 645, 671

(a) Use your GDC to calculate one-variable statistics and explain their meanings. (b) State whether the data are discrete or continuous. (c) Draw a histogram or bar graph. [8]

2. Displaying grouped data. Create appropriate frequency diagrams from the tables below. Explain the difference between discrete and continuous data. [6]

Grapes per bunch	Frequency
1–3	2
4–6	6
7–9	7
10–12	12
13–15	10
16–18	5
> 19	1

Height h (cm)	Frequency
$149.5 \leq h < 152.5$	3
$152.5 \leq h < 155.5$	8
$155.5 \leq h < 158.5$	6
$158.5 \leq h < 161.5$	12
$161.5 \leq h < 164.5$	7
$164.5 \leq h < 167.5$	4
$167.5 \leq h < 170.5$	2
$170.5 \leq h < 173.5$	4
$173.5 \leq h < 176.5$	5

3. Five-number summary. For the electricity data in Q1, calculate the minimum, Q_1, median, Q_3, maximum and IQR. Draw a box-and-whisker diagram. [7]

4. Frequency and quartiles. Complete the cumulative-frequency column and determine the five-number summary and IQR for the following *discrete* values. [6]

Value	Frequency	CF
2	4	
4	7	
6	8	
7	5	
10	2	

5. Estimated mean. For Manila, estimate the mean daily maximum temperature. [5]

Temperature (°C)	$[24, 26)$	$[26, 28)$	$[28, 30)$	$[30, 32)$	$[32, 34)$	$[34, 36)$
Frequency	1	8	32	107	174	43

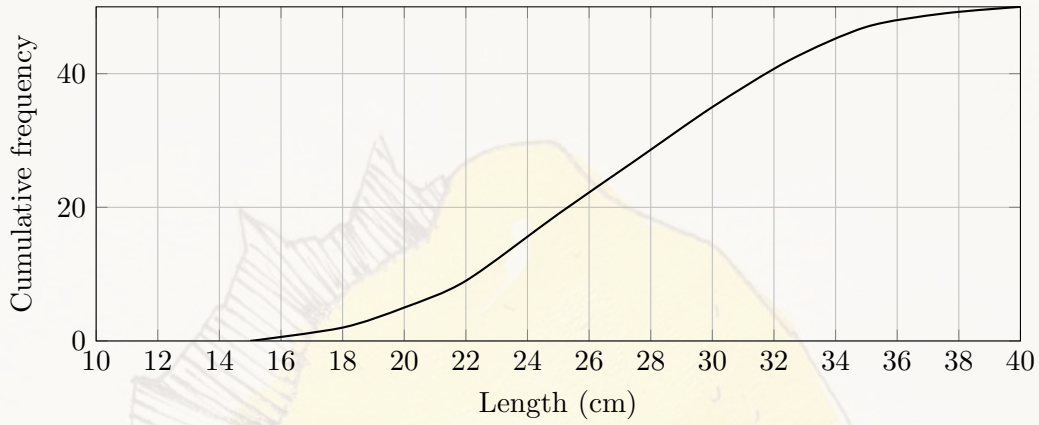
6. Height cumulative frequencies. Complete the CF column and plot an ogive. Estimate the median and IQR. [8]

Upper boundary	152.5	155.5	158.5	161.5	164.5	167.5	170.5	173.5	176.5
Frequency	3	8	6	12	7	4	2	4	5
CF									

. . . . . .

. . . . . .

7. Trout lengths (graph reading). The original lesson's cumulative-frequency curve reaches 50 trout at 40 cm. Read the curve to estimate: (a) total caught, (b) number shorter than 20 cm, (c) number longer than 30 cm, (d) median, (e) IQR. [6]



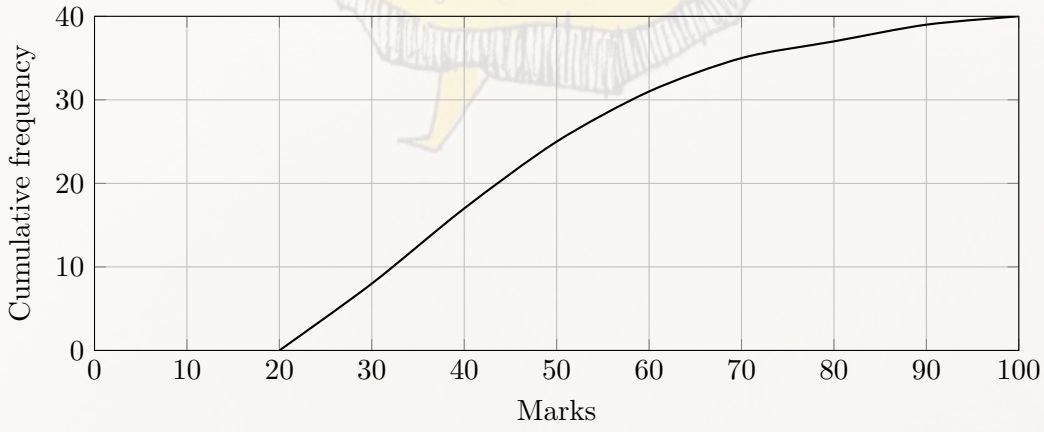
8. Apple harvest. Complete the missing frequencies and cumulative frequencies, then draw the cumulative-frequency graph. [7]

Number of apples	Frequency	CF
$0 \le x < 200$	8	
$200 \le x < 400$		22
$400 \le x < 600$	28	
$600 \le x < 800$		76
$800 \le x < 1000$		80

. . . . . .

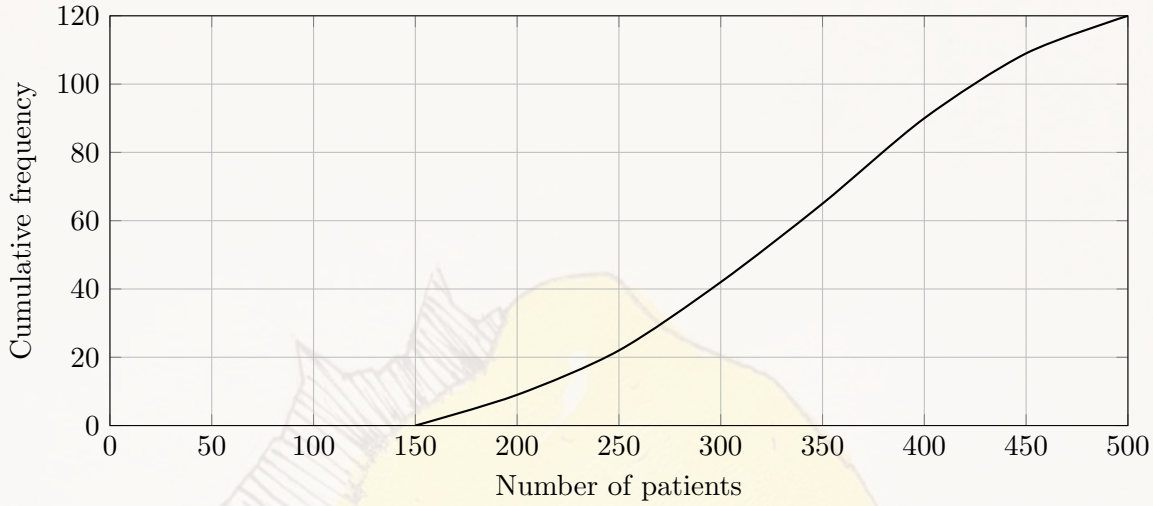
. . . . . .

9. Maths test marks. From the original lesson's ogive for 40 students, estimate (a) median, (b) IQR, and (c) number failing if the pass mark is 40. [6]



10. Hospital comparison. The original Sercombe Hospital ogive is represented approximately by the curve below. Leckwith Hospital frequencies for $[100, 200)$, $[200, 300)$, $[300, 400)$, $[400, 500)$ are $24, 32, 44, 20$.

(a) Plot Leckwith's cumulative-frequency points on the same axes. (b) Make two comparisons. (c) Estimate the percentage of Leckwith days with more than 420 patients. (d) Estimate x such that Sercombe had more than x patients on 20% of days. [10]



Detailed Markscheme

Topic 4.2 – Presentation of Grouped Data

Suggested allocations; numerical answers read from recreated curves are approximate.

1. Warm-up [8]

$n = 18$, $\sum x = 8319$, mean $\bar{x} = 8319/18 \approx 462.17$ kWh, median 472.5 kWh, sample standard deviation $s_x \approx 133.80$ kWh, and range $671 - 245 = 426$ kWh. The data are continuous measurements recorded to the nearest kWh. The mean and median describe centre; standard deviation and range describe spread. Accept a properly labelled graph using sensible equal-width classes. [Statistics 4; interpretation 2; graph 2]

2. Frequency diagrams [6]

The grape counts are discrete; heights are continuous. Plot frequencies $2, 6, 7, 12, 10, 5, 1$ for the grape categories, and $3, 8, 6, 12, 7, 4, 2, 4, 5$ for the contiguous height classes. Both have a maximum frequency of 12. [Classification 2; each diagram 2]

3. Five-number summary [7]

With $n = 18$, the median is the mean of observations 9 and 10:

$$Q_2 = (459 + 486)/2 = 472.5.$$

The lower nine values have median $Q_1 = 341$; the upper nine have median $Q_3 = 573$. Therefore

$$\boxed{(245, 341, 472.5, 573, 671)}, \qquad \boxed{\text{IQR} = 232}.$$

Draw the box from 341 to 573, median line at 472.5 and whiskers at 245 and 671. [Summary 5; IQR 1; plot 1]

4. Discrete frequencies [6]

Cumulative frequencies: $4, 11, 19, 24, 26$; total $n = 26$. The 13th and 14th values are 6, so $Q_2 = 6$. The lower 13 have median 7th value, $Q_1 = 4$; upper 13 have median 20th value, $Q_3 = 7$. Thus

$$\boxed{(2, 4, 6, 7, 10)}, \qquad \boxed{\text{IQR} = 3}.$$

[CF 2; quartiles 3; IQR 1]

5. Manila estimated mean [5]

Use midpoints $25, 27, 29, 31, 33, 35$:

$$\sum f = 365, \qquad \sum fm = 25 + 216 + 928 + 3317 + 5742 + 1505 = 11733.$$

$$\boxed{\bar{t} \approx 11733/365 \approx 32.1^\circ\text{C}}.$$

[Midpoints 2; products and totals 2; answer 1]

6. Height CF [8]

CF: $3, 11, 17, 29, 36, 40, 42, 46, 51$. Start ogive at $(149.5, 0)$ and plot upper boundaries against CF. With $N = 51$, interpolate at 12.75, 25.5, 38.25:

$$Q_1 \approx 155.5 + \frac{12.75 - 11}{6}(3) = 156.4,$$

$$Q_2 \approx 158.5 + \frac{25.5 - 17}{12}(3) = 160.6,$$

$$Q_3 \approx 164.5 + \frac{38.25 - 36}{4}(3) = 166.2.$$

Hence $\boxed{\text{IQR} \approx 9.8 \text{ cm}}$. [CF 2; graph 3; median 1; quartiles/IQR 2]

7. Trout curve [6]

The curve reaches 50, so 50 trout. At 20 cm, CF ≈ 5, so about 5 shorter. At 30 cm, CF ≈ 35, so about $50 - 35 = 15$ longer. Median is read at CF 25, about 27 cm. Quartiles are read at CF 12.5 and 37.5, approximately 23 cm and 31 cm, so IQR ≈ 8 cm. [1 each for total, each count, median, quartiles and IQR]

8. Apple harvest [7]

Complete rows as $(f, \text{CF}) = (8, 8), (14, 22), (28, 50), (26, 76), (4, 80)$. Plot $(0, 0), (200, 8), (400, 22), (600, 50), (800, 76), ($
[Table 4; graph 3]

9. Maths test [6]

For $N = 40$, median is at CF 20, approximately 44 marks. Q_1 at CF 10 is about 32 marks; Q_3 at CF 30 is about 58 marks; IQR ≈ 26 marks. CF at 40 is about 17, so approximately 17 students scored below 40. [Median 2; IQR 2; failures 2]

10. Hospitals [10]

Leckwith CF points: $(100, 0), (200, 24), (300, 56), (400, 100), (500, 120)$. For Leckwith, median CF 60 corresponds to about 309 patients; for Sercombe, median CF 60 is about 339 patients. Hence Sercombe has the higher median. Leckwith's middle 50% is approximately 219 to 355 patients, IQR ≈ 136; Sercombe's is approximately 270 to 370, IQR ≈ 100, so Leckwith is more variable in its middle half. [Plot 3; comparisons 3] At 420, Leckwith CF $\approx 100 + 20(20/100) = 104$; days above 420 $\approx 120 - 104 = 16$, giving $\boxed{13.3\%}$. [2] 20% above x means 80% at or below x: CF $= 0.8(120) = 96$. Read Sercombe curve: $\boxed{x \approx 415 \text{ patients}}$. [2]
Graph-based estimates can vary. The curves reproduced here are approximations of those in the uploaded lesson, not the original vector graphics.



Maths with Lemon

Mathematics made clear.