A LEVEL MATHS 9709 · PAPER 5 PROBABILITY & STATISTICS 1 · 5.1 · ~5 MIN READ

Representation of data

Stem-and-leaf diagrams, box-and-whisker plots, histograms with unequal class widths and cumulative frequency graphs; median, quartiles and IQR; mean and standard deviation from raw, grouped and coded data, and choosing the right measures.

Stem-and-leaf | Box plots | Histograms & frequency density | Cumulative frequency | Median & quartiles | Mean & standard deviation | Coded data Σ(x − a) | Comparing distributions

RS by Riyaz Sir
mentorriyaz.in

Paper tag = **where it is examined** FORMULA LIST (MF19) given in the exam M1 A1 B1 how marks are awarded

highlight = comes up again and again Test yourself at the end

IN 30 SECONDS

- **Histogram: frequency density = frequency ÷ class width.** The area of each bar is the frequency, so bars of unequal width are fair.
- **Standard deviation** measures spread about the mean: $\sigma^2 = \Sigma x^2/n - \bar{x}^2$.
- **Coded data:** if you know $\Sigma(x - a)$ and $\Sigma(x - a)^2$, the mean is $a + \Sigma(x - a)/n$ and the standard deviation is unchanged by the shift.
- **Choose measures to suit the data:** median and IQR resist outliers and skew; mean and standard deviation use every value.

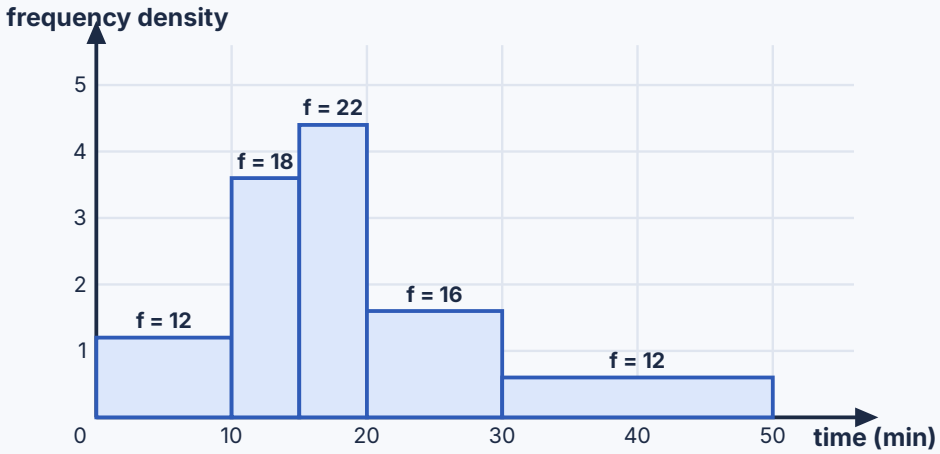
BY THE END YOU CAN

- Draw and interpret stem-and-leaf diagrams, box plots, histograms and cumulative frequency graphs.
- Find and interpret the median, quartiles and interquartile range.
- Calculate the mean and standard deviation from raw data, grouped data and $\Sigma(x - a)$ totals.
- Choose suitable measures and compare data sets in context.

Riyaz Sir

1 Diagrams

Histograms



Frequency density = frequency ÷ class width, so the AREA of each bar is the frequency.

Unequal class widths: height is frequency density, so each bar's area is its frequency.

- **Class boundaries matter.** For "time to the nearest minute, 10–14", the boundaries are 9.5 and 14.5, so the width is 5.
- **Reading a histogram:** frequency = frequency density × width. For part of a class, assume values are spread evenly.

WORKED EXAMPLE

A histogram has classes 0–10, 10–15, 15–20, 20–30 and 30–50 minutes with frequencies 12, 18, 22, 16, 12. Find the frequency densities, and estimate how many values exceed 25.

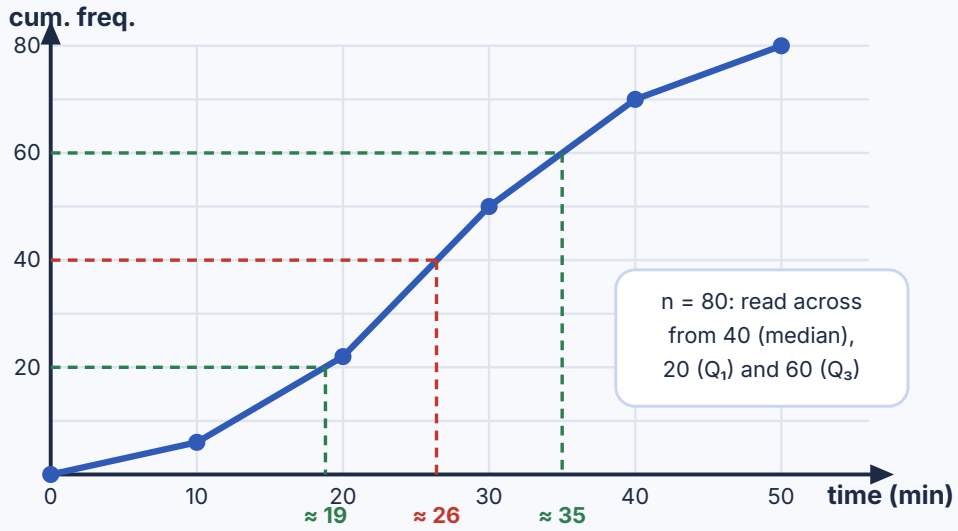
1. densities: 12/10 = 1.2, 18/5 = 3.6, 22/5 = 4.4, 16/10 = 1.6, 12/20 = 0.6 M1 A1
2. above 25: half of 20–30 gives 8, plus 12 M1
3. **20** A1

COMMON MISTAKE

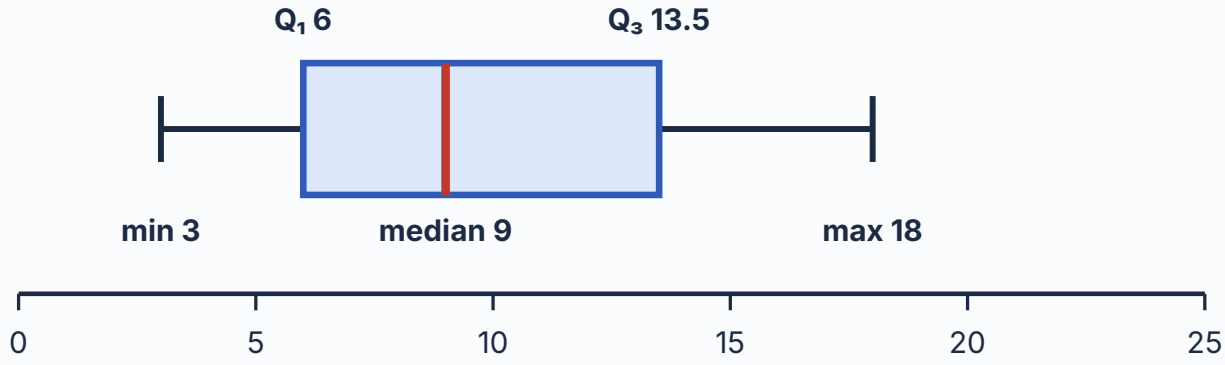
Plotting frequency as height for unequal widths. The 2024 Paper 5 report found this was still the most common histogram error.

Riyaz Sir

Box plots and cumulative frequency



Read across at n/2 for the median, n/4 and 3n/4 for the quartiles.



IQR = 13.5 − 6 = 7.5. Outliers lie beyond $Q_1 - 1.5 \times IQR$ or $Q_3 + 1.5 \times IQR$.

A box plot shows the median, quartiles and extremes at a glance.

- **Stem-and-leaf** keeps the raw data and needs a key (e.g. 3 | 7 means 37). A back-to-back diagram compares two sets.
- **Box plots** compare two sets quickly: compare medians (average) and IQRs (spread).
- **Skew:** if the median is nearer Q_1 than Q_3, the data are positively skewed.

2 Measures

Mean, median, standard deviation

$$\bar{x} = \Sigma x/n = \Sigma fx/\Sigma f \; ; \; \sigma^2 = \Sigma x^2/n - \bar{x}^2 = \Sigma fx^2/\Sigma f - \bar{x}^2$$ FORMULA LIST (MF19)

grouped data use mid-points: the result is an estimate

$$\bar{x} = a + \Sigma(x - a)/n \; ; \; \sigma^2 = \Sigma(x - a)^2/n - [\Sigma(x - a)/n]^2$$ LEARN IT

coding: subtracting a moves the mean but not the spread

WORKED EXAMPLE

For the histogram data above, estimate the mean and the median.

1. $\Sigma fx = 5 \times 12 + 12.5 \times 18 + 17.5 \times 22 + 25 \times 16 + 40 \times 12 = 1550$; $n = 80$ M1
2. mean ≈ **19.4 minutes** A1
3. median is the 40th value: 30 values below 15, 22 in 15–20 M1
4. **median ≈ 15 + (10/22) × 5 = 17.3 minutes** A1

WORKED EXAMPLE

For 20 values, $\Sigma(x - 50) = -36$ and $\Sigma(x - 50)^2 = 1462$. Find the mean and standard deviation.

1. mean = $50 + (-36/20)$ = **48.2** B1
2. $\sigma^2 = 1462/20 - (-1.8)^2 = 73.1 - 3.24 = 69.86$ M1
3. **$\sigma = 8.36$** A1

WORKED EXAMPLE

Two groups: 12 values with mean 15 and $\Sigma x^2 = 2900$; 8 values with mean 20 and $\Sigma x^2 = 3400$. Find the mean and standard deviation of all 20 values.

1. $\Sigma x = 180 + 160 = 340 \Rightarrow$ mean 17 M1 A1
2. $\Sigma x^2 = 6300 \Rightarrow \sigma^2 = 6300/20 - 17^2 = 315 - 289 = 26$ M1
3. **$\sigma = 5.10$** A1

COMMON MISTAKE

Forgetting to square the mean, or using $\Sigma(x - a)^2$ as if it were Σx^2. Also, comparing data sets with no context: "the median time is greater for group A, so they usually took longer".

LINKING QUESTION

Probability from data begins in Probability; spread and the normal model are in The normal distribution.

Test yourself — cover the answers first

QUICK CHECK

1. Find the frequency density for a class of width 4 containing 18 values.
2. What does the area of a histogram bar represent?
3. Which measures are best for skewed data?

4. $\Sigma x = 60$, $n = 12$, $\Sigma x^2 = 400$. Find the variance.

5. What does a key like 4 | 2 mean in a stem-and-leaf diagram?

EXAM-STYLE

6. 10 values have $\Sigma(x - 3) = 12$ and $\Sigma(x - 3)^2 = 50$. Find the mean and the standard deviation.

7. The heights in a class of width 20 cm have frequency density 1.5 per cm. How many values are in the class?

8. Data set A: median 24, IQR 6. Data set B: median 31, IQR 15. Compare.

Show answers

1. 4.5
2. frequency
3. median and IQR
4. 400/12 − 25 = 8.33
5. 42 (with the stated units)
6. mean 4.2; $\sigma^2 = 5 - 1.44 = 3.56 \Rightarrow \sigma = 1.89$
7. 30
8. B is on average larger and much more spread out than A

Stuck on a question?

Send it to me on WhatsApp. I read every message.

Want regular help? Send this to your parent →

Share with a friend

Ask Riyaz Sir