Statistics and probability

OxfordAQA · International GCSE · Mathematics

OxfordAQA 9260 unit 3.4 (specification Version 5.1) – the smallest unit at 13% of the combined specimen marks. Core rows S1–S5 cover data types, reading tables, two-way tables, the full chart bank and the basic averages; the **Extension additions** add histograms with UNEQUAL class intervals, cumulative frequency diagrams, box plots, and quartiles with the inter-quartile range and percentiles. The 1E specimen carries a histogram and a cumulative-frequency correction to upper boundaries; the 2E specimen's statistics questions include the 40-men and 38% reads. Lines the specimens never sample are labelled specification originals.

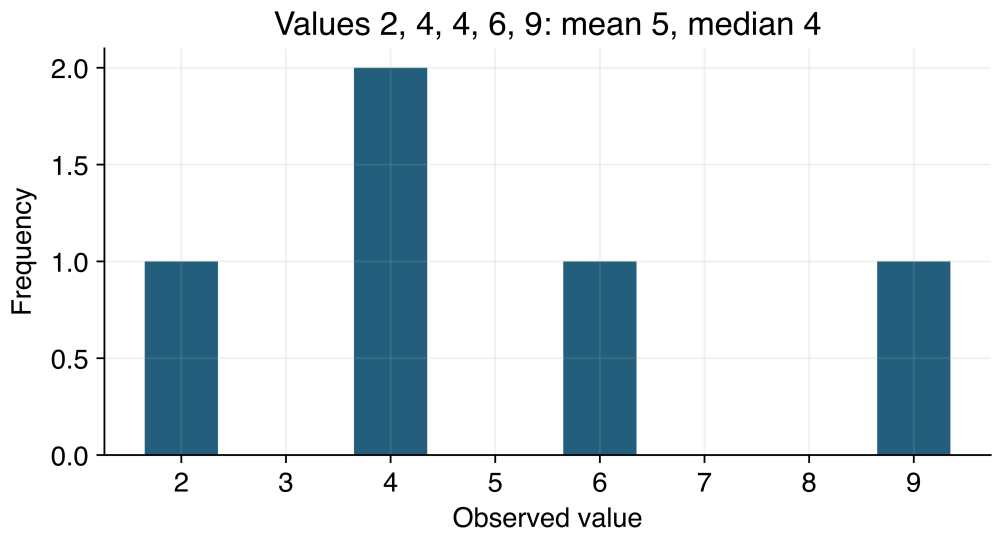
Must-know for unit 3.4:

Family	Formulae
histogram	frequency density $= \dfrac{\text{frequency}}{\text{class width}}$ – bar AREA is the frequency
cumulative frequency	plot at the UPPER class boundary; median at $\frac{n}{2}$, quartiles at $\frac{n}{4}$, $\frac{3n}{4}$
averages	mean $= \dfrac{\sum fx}{\sum f}$ (grouped: midpoints); IQR $= Q_3 - Q_1$

Data types, tables and charts

Qualitative data categorises (favourite fruit); **discrete** data counts (number of goals); **continuous** data measures (height), grouped or ungrouped. Extract data from printed tables and lists – read the row and column headers first, and quote units. **Two-way tables** cross-tabulate two variables; every row and column total checks the interior cells: 30 students, 18 cricket, 15 tennis and 7 both fills a 2×2 whose cells must sum to 30.

The Core chart bank: scatter graphs, stem-and-leaf, tally charts, pictograms, bar charts, dual and composite bar charts, pie charts, line graphs, frequency polygons and histograms with EQUAL class intervals. Choose by data type – categories take bars, comparisons take dual bars, parts-of-a-whole take pie charts, trends take line graphs, correlation takes scatter graphs.



Original native-lesson illustration; labels belong to its worked example.

Extension diagrams: unequal histograms, CF and box plots

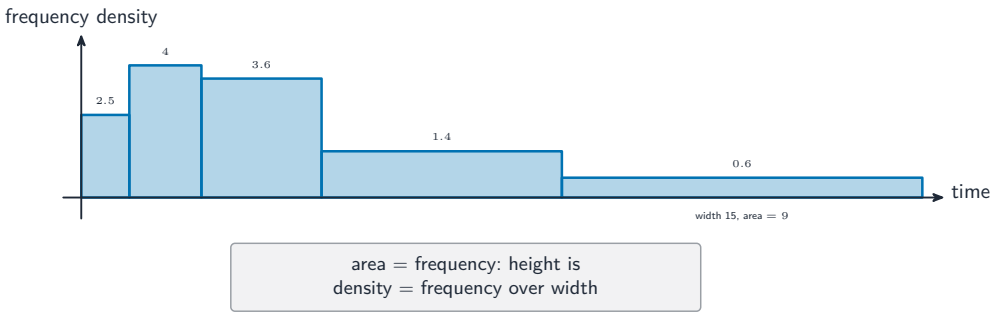
A **histogram** with **unequal class intervals** draws bar HEIGHT as the **frequency density** = frequency over class width – the AREA is the frequency (classes of widths 2, 3, 5, 10, 15 holding 5, 12, 18, 14, 9 give heights 2.5, 4, 3.6, 1.4, 0.6). A **cumulative frequency diagram** plots each running total at its UPPER class boundary as a smooth increasing curve (the 1E specimen's correction is exactly this upper-boundary point placement); the median reads at $\frac{n}{2}$, quartiles at $\frac{n}{4}$ and $\frac{3n}{4}$. A **box plot** (five-number summary: minimum, Q_1, median, Q_3, maximum) shows spread at a glance – the box spans the IQR, the whiskers the range, outliers by judgement.

Averages, quartiles and percentiles

The **mean** divides the total by the count (grouped data multiplies each MIDPOINT by its frequency first); the **median** orders and takes the middle; the **mode** is most frequent and the **modal class** is the class with the highest frequency DENSITY when widths differ. The Extension additions add **quartiles** and the **inter-quartile range** $= Q_3 - Q_1$ (the median of the lower and upper halves), and **percentiles** read like the median at any percentage – the pth percentile is the value $\frac{p}{100}n$ of the way up the ordered data or the cumulative-frequency curve.

Worked example (original). 60 workers' times have cumulative frequencies 6, 17, 27, 42, 60 at upper boundaries 10, 15, 20, 25, 30 (hours). The median reads at worker

30: between 20 (27) and 25 (42), so $20 + 5 \times \frac{30-27}{42-27} = 21$ hours. Q_1 at worker 15: $10 + 5 \times \frac{15-6}{17-6} = 14.1$; Q_3 at worker 45: $25 + 5 \times \frac{45-42}{60-42} = 25.8$; IQR ≈ 11.7 hours. Draw the reading lines – the marks are for the reads.



Checklist

- Data types qualitative/discrete/continuous; grouped and ungrouped (S1).
- Extract from printed tables and lists, quoting units (S2).
- Design and use two-way tables with row/column checks (S3).
- The Core chart bank, chosen by data type (S4).
- Unequal-interval histograms by frequency density (area is frequency); cumulative frequency diagrams at UPPER boundaries with median/quartile reads; box plots (S4 Extension; 1E histogram and CF-boundary anchors).
- Mean (grouped via midpoints), median, mode and modal class (S5).
- Quartiles, inter-quartile range and percentiles from data or the curve (S5 Extension).

Boundary notes: probability vocabulary sits in this specification's Number/Algebra work and earlier units' sets; the AO split is asserted by the specification. Lines not sampled by the specimens (stem-and-leaf, pie charts and several chart types on the Extension papers) are taught as specification originals; the specimens are Extension-tier only, so Core-tier demand has no local evidence. Worked examples are adapted from the 2016 specimen pair or original.