Y12 Applied Chapter 3—Representations of Data

REPRESENTATIONS OF DATA

KEY WORDS & DEFINITIONS

1. Outlier

A data value that lies beyond expected extremities. These are usually calculated as a multiple of the interquartile range above the upper quartile or below the lower quartile.
ie. either greater than $Q_3 + k(Q_3 - Q_1)$
or less than $Q_1 - k(Q_3 - Q_1)$

2. Cleaning

The process of removing anomalies from the data set.

WHAT DO I NEED TO KNOW

Comparing 2 sets of data:

Calculate & compare the measures of location
Calculate & compare the measures of spread
Compare outliers if applicable
Mean & s.d go together
Median & IQR go together.
Ensure all comparisons are done IN CONTEXT

Histograms

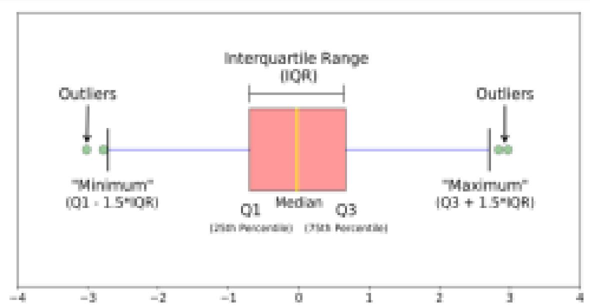
Area of bar ∝ Frequency so
Area of bar = k x Frequency
Area does NOT always = Frequency

BOX PLOTS

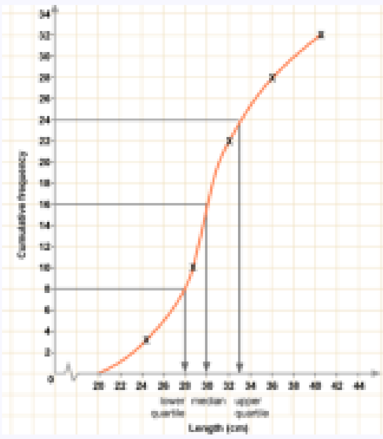
Box plots are rarely symmetrical

25% of the data lies within each section

Always use the same scale when comparing box plots



CUMULATIVE FREQUENCY



Plot points at the upper limits of group boundaries
Ensure it makes sense to extrapolate the curve at the beginning
Be careful of questions that ask "How many are **more** than…"

HISTOGRAMS

Time	Frequency	Frequency Density	
$0 < t \leq 10$	20	2	$20 \div 10$
$10 < t \leq 15$	15	3	$15 \div 5$
$15 < t \leq 20$	10	2	$10 \div 5$
$20 < t \leq 25$	9	1.8	$9 \div 5$
$25 < t \leq 40$	6	0.4	$6 \div 15$

Frequency Density = $\frac{\text{Frequency}}{\text{Class width}}$

Histograms are used to represent grouped continuous data
Area of bar = k x frequency
If k = 1, then frequency density = $\frac{\text{frequency}}{\text{class width}}$

You may need to find the areas of parts of bars if questions don't use the class boundaries.
Joining the middle of the tops of each bar in a histogram forms a frequency polygon