STATISTICS

• Data is *discreet* or *continuous*: *discrete* data – info collected by counting *continuous* data – info collected by measurement (e.g. lengths, heights, time etc.)

UNGROUPED DATA (arrange in order of size)

Measures of central tendency	Measures of dispersion (or spread)
1. Mean (Average)	1. Range
$\checkmark \bar{x} = \frac{\text{sum of all data items}}{\text{number of items}}$	✓ Largest value – smallest value
	2. Interquartile range (spread around median)
2. Median	$\checkmark IQR = Q_3 - Q_1$
	QUARTILES: divide data spread into quarters
✓ middle value in a set of ordered	✓ Q_1 – Lower quartile
data	position $=\frac{1}{4}(n+1)$
	$\checkmark Q_2$ – Median
	position $=\frac{1}{2}(n+1)$
	$\checkmark Q_3$ – Upper quartile
	position $=\frac{3}{4}(n+1)$
	✓ Semi-interquartile range = $\frac{Q_3 - Q_1}{2}$
3. Mode	3. Standard deviation (σ) (spread around mean)
 ✓ the number/value that appears most frequently in the data set 	 ✓ Determine with a scientific calculator ✓ Small σ – data closely bunched around mean Larger σ – data spread away from mean

• *Percentiles*: Divide data spread into **one hundredths** E.g. 11th percentile of 70 values $=\frac{11}{100}(70 + 1) = 7,81 \approx \text{the 8}^{\text{th}}$ value

• Five number summary: (1) Min value; (2) Q_1 ; (3) Median; (4) Q_3 & (5) Max value

• Box and whisker diagram/plot:

*Represents the five number summary *Draw diagram to scale on a number line to observe skewedness of data



• Skewed data

SYMMETRICAL distribution (*mean = median*)



SKEWED TO THE LEFT / NEGATIVELY SKEWED (mean < median)



SKEWED TO THE RIGHT / POSITIVELY SKEWED (mean > median)



• *Outliers*: data values that are less than $(Q_1 - 1, 5 \times IQR)$, or.... data values that are greater than $(Q_1 + 1, 5 \times IQR)$

GROUPED DATA

• *Measures of central tendency:*

*Estimated mean *Modal class *Median class

- Cumulative frequency graph / OGIVE
 - ✓ Represent cumulative results from a cumulative frequency table.
 - ✓ Usually forms a smooth S-shaped curve.
 - \checkmark To plot the ogive:
 - *Anchor the graph on the $x axis \rightarrow$ use the lower limit of 1st interval
 - **x* coordinates of points \rightarrow use upper limit of each interval
 - **y* coordinates of points \rightarrow use cumulative frequency values
 - \checkmark From an ogive, the following can be asked:

*to complete the frequency table

*the median, lower and upper quartiles (work from the vertical/y – axis)

BIVARIATE DATA & SCATTER PLOTS/GRAPHS/DIAGRAMS

- Bivariate data means that the data is recorded as ordered number pairs and investigates the nature of a possible relationship between two variables.
- The ordered number pairs are plotted on a Cartesian plane (called a scatter plot)
- The line of best fit/regression line is the trend line drawn on a scatter plot.
- Mathematical way of finding the equation of the regression line is known as the *least squares regression line*.
 Equation of the least squares regression line is ŷ = a + bx

(Determine *a* and *b* with the aid of a scientific calculator)

- *INTERPOLATE* estimate by 'observing' the known data around a point (read off info based on the line of best fit of the known data).
- *EXTRAPOLATE* read off info from extended parts of the line of best fit.
- The <u>correlation</u> refers to how strong the relationship between the two variables being investigated, is. It can be weak, strong, etc.

- The <u>correlation coefficient</u> (*r*) indicates the strength of the relationship between the variables.
 - ✓ Determine the value of *r* with a scientific calculator (-1 < r < 1)
 - ✓ The value of r:



- > The closer r is to either -1 or 1, the stronger the relationship between the two variables.
- → r positive → the gradient of the line is positive
- > r negative → the gradient of the line is negative