

MEASURES OF LOCATION: DEFINITIONS

Location = Central Tendency

- ▶ **Parameter:** A key feature of a population (eg. population mean) *PP: Population Parameter*
- ▶ **Statistic:** A key feature of a sample (eg. sample mean) *SS: Sample Statistic*
- ▶ **Mean:** Average; Used to measure *central tendency*/ location.
Can be affected by extreme outliers. In such cases, median can be a better of location/ central tendency (Eg. 1, 2, 3, 99)
- ▶ **Median:** Middle value of scores when ordered from highest to lowest.
- ▶ **Mode:** Value that has the highest frequency & is most common.

MEASURES OF VARIABILITY: DEFINITIONS

Variation: Used to calculate covariance

- ▶ **Variation:** Measure of dispersion amongst data / Spread of data.
Squared deviation from the mean
- ▶ **Standard Deviation:** Distance of value / observation from the mean.
(eg. In an exam, the given mean is 65 & the SD. is 15, so the data can vary from 50-80)
Low SD = Low variability & High SD = High variability
- ▶ **Range:** A simple measure of variability; Can be misleading (Maximum - Minimum)
- ▶ **Percentiles:** Ranking of values to measure location.
- ▶ **Inter Quartile Range:** Difference between upper and lower quartiles. (Another measure of spread)

STANDARDISING DATA

Creating a **transformed** variable with zero mean & unit variance from any original quantitative variable. (eg. 1)

► **Z-scores:** Transformed variable that is free of units of measurement.

- Calculation: (Observation - Mean) / SD
- Allows direct comparisons across original variables measured in *different units*.

► **Coefficient of Variation:** Standardised Variation

- When you wish to measure variation *relative to sample size*.
- Calculation: SD / Mean

MEASURES OF ASSOCIATION: DEFINITIONS

Describes the relationship between **2 variables** (Usually scatter-plots)

► **Covariance:** Measure of linear relationship between 2 variables (How they affect each other)

Can be $>$ or $<$ 1 (eg. 50 = positive linear association, **but** 0 = no linear association)

► **Correlation Coefficient:** Standardised Covariance. Unit-free measure of Association

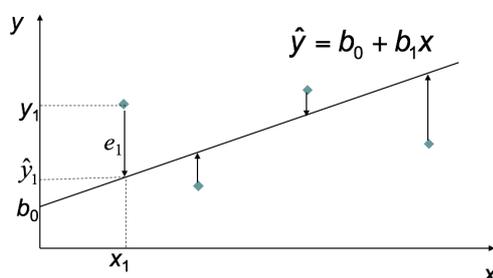
Values lie *between* -1 & 1. (1 being perfect positive correlation, -1 being perfect negative correlation and 0 meaning no correlation)

(eg. Time spent on studies & Score. i.e. Positive correlation: \wedge Time \wedge Score)

LEAST SQUARES

► To minimise the sum of squared residuals. (Based on regression analysis)

Determining the line of best fit whilst decreasing the distance between the actual value & the mean.



FORMULA:

- "b₀" represents the intercept & "b₁" represents the slope (b₁ has the **same sign** as covariance)

Both estimates are minimised as much as possible, so that the residual sum of squares are minimised.

- "e" represents the error (squares)
- Arrows represent the distance between mean & actual value (∴ error)

* In the case of zero covariance (correlation), $b_1 = 0$. There is no slope.

$$b_1 = \frac{s_{xy}}{s_x^2} \quad b_0 = \bar{y} - b_1\bar{x}$$

R - SQUARED STATISTIC

► Describes the fit of the 'model' to the actual data. (Measuring the 'best fit')

- Considers how much variation is in the residuals.
- Maximum value = 1 (perfect fit)
- Minimum value = 0 (no fit: cannot explain any relationship)
- A.K.A. **coefficient of determination**.