DATA ORGANISATION & CENTRAL TENDENCY

SECTION 1: DATA ORGANISATION & PRESENTATION

CONCEPT	FORMULA / CALCULATION	KEY NOTES
1 STURGES' RULE	$K = 1 + 3.322 \log_{10} N$	Determines optimal number of classes (K) for sample size (N)
2 CLASS MARK (MID-POINT)	$m = \frac{L_1 + L_2}{2}$	L_1 = Lower Limit L_2 = Upper Limit

SECTION 2: MEASURES OF CENTRAL TENDENCY (AVERAGES)

1 ARITHMETIC MEAN ($\bar{X}$)

- Individual Series: $\bar{X} = \frac{\sum X}{n}$
- Discrete / Continuous Series: $\bar{X} = \frac{\sum fX}{\sum f}$
- Assumed Mean Method: $\bar{X} = A + \frac{\sum fd}{N}$ (d = X - A)
- Step-Deviation Method: $\bar{X} = A + \left(\frac{\sum fd'}{N}\right) \times h$ $\left(d' = \frac{X - A}{h}\right)$
- Weighted Mean ($\bar{X}_w$): $\bar{X}_w = \frac{\sum wX}{\sum w}$
- Combined Mean ($\bar{X}_{12}$): $\bar{X}_{12} = \frac{n_1 X_1 + n_2 X_2}{n_1 + n_2}$

2 POSITIONAL AVERAGES (MEDIAN & MODE)

MEDIAN (M) — CONTINUOUS SERIES

$$M = L + \left(\frac{\frac{N}{2} - c.f.}{f}\right) \times h$$

Where L = lower limit of median class, $c.f.$ = cumulative frequency of preceding class, f = frequency of median class, h = class length.

MODE (Z) — CONTINUOUS SERIES

$$Z = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

Where f_1 = modal class frequency, f_0 = preceding frequency, f_2 = succeeding frequency.

3 GEOMETRIC MEAN (GM) & HARMONIC MEAN (HM)

- Geometric Mean (GM):

$$GM = \left(\prod X\right)^{1/n} = \text{Antilog}\left(\frac{\sum \log X}{n}\right)$$

- Harmonic Mean (HM):

$$HM = \frac{n}{\sum\left(\frac{1}{X}\right)}$$

4 PARTITION VALUES (QUARTILES)

- First Quartile (Q_1):
  - Size of $\left(\frac{N+1}{4}\right)^{th}$ term (Individual)
  - $Q_1 = L + \left(\frac{\frac{N}{4} - c.f.}{f}\right) \times h$ (Continuous)
- Third Quartile (Q_3):
  - Size of $3\left(\frac{N+1}{4}\right)^{th}$ term (Individual)
  - $Q_3 = L + \left(\frac{\frac{3N}{4} - c.f.}{f}\right) \times h$ (Continuous)

CORE MATHEMATICAL RELATIONSHIPS

$AM \geq GM \geq HM$

$\text{Mode} = 3(\text{Median}) - 2(\text{Mean})$

MEASURES OF DISPERSION & CORRELATION

SECTION 1: MEASURES OF DISPERSION

MEASURE	ABSOLUTE FORMULA	RELATIVE FORMULA (COEFFICIENT)
1 Range (R)	$R = L - S$	Coeff of Range $= \frac{L-S}{L+S}$
2 Quartile Deviation (QD)	$QD = \frac{Q_3 - Q_1}{2}$	Coeff of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$
3 Mean Absolute Deviation (MD)	$MD = \frac{\sum \lvert X - \bar{X} \rvert}{n}$	Coeff of MD $= \frac{MD}{\bar{X}}$ or $\frac{MD}{Median}$
4 Variance (σ^2)	$\sigma^2 = \frac{\sum (X - \bar{X})^2}{n}$	—
5 Standard Deviation (σ)	$\sigma = \sqrt{\frac{\sum f d^2}{N} - \left(\frac{\sum f d}{N}\right)^2} \times h$	Coefficient of Variation (CV) $= \frac{\sigma}{\bar{X}} \times 100$

COMBINED STANDARD DEVIATION (σ_{12})

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

where $d_1 = \bar{X}_1 - \bar{X}_{12}$, $d_2 = \bar{X}_2 - \bar{X}_{12}$

SECTION 2: CORRELATION ANALYSIS (r)

1 COVARIANCE

$$Cov(X,Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n} = \frac{\sum XY}{n} - \bar{X}\bar{Y}$$

2 KARL PEARSON'S COEFFICIENT OF CORRELATION (r)

- Covariance Formula: $r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$
- Direct Computational Formula: $r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$
- Properties: $-1 \le r \le +1$

3 SPEARMAN'S RANK CORRELATION (ρ)

- No Tied Ranks: $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$ $(d = R_1 - R_2)$
- With Tied (Repeated) Ranks: $\rho = 1 - \frac{6\left[\sum d^2 + \frac{\sum (m^3 - m)}{12}\right]}{n(n^2-1)}$ (m = number of times a rank repeats)

4 ERROR & LIMITS OF CORRELATION

- Standard Error (SE_r): $SE_r = \frac{1 - r^2}{\sqrt{n}}$
- Probable Error (PE_r): $PE_r = 0.6745 \times SE_r = 0.6745 \times \left(\frac{1 - r^2}{\sqrt{n}}\right)$

INDEX NUMBERS & STATISTICAL TOOLS MATRIX

SECTION 1: PRICE & QUANTITY INDEX NUMBERS

1 UNWEIGHTED INDEX NUMBERS

• Simple Aggregative Method:	$P_{01} = \frac{\sum p_1}{\sum p_0} \times 100$
• Simple Average of Price Relatives:	$P_{01} = \frac{\sum R}{n}$ where $R = \left(\frac{p_1}{p_0}\right) \times 100$

2 WEIGHTED INDEX NUMBERS

METHOD	FORMULA	WEIGHT USED (W)
Laspeyres Index (P_L)	$P_L = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$	Base Year Quantity (q_0)
Paasche Index (P_P)	$P_P = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$	Current Year Quantity (q_1)
Fisher's Ideal Index (P_F)	$P_F = \sqrt{P_L \times P_P} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$	Geometric Mean of P_L & P_P

SECTION 2: COST OF LIVING & DEFLATING

1 CONSUMER PRICE INDEX (FAMILY BUDGET METHOD):

$$CPI = \frac{\sum WR}{\sum W}$$

where $R = \frac{p_1}{p_0} \times 100$, $W = p_0 q_0$

2 REAL WAGE (DEFLATED INCOME):

$$\text{Real Wage} = \left(\frac{\text{Money Wage}}{\text{CPI}}\right) \times 100$$

3 PURCHASING POWER OF MONEY:

$$\text{Purchasing Power} = \frac{1}{\text{CPI}}$$

SECTION 3: ESSENTIAL TESTS OF CONSISTENCY FOR INDEX NUMBERS

1 TIME REVERSAL TEST: $P_{01} \times P_{10} = 1$

Satisfied by: Fisher's Ideal Index, Marshall-Edgeworth, Kelly's Index.

2 FACTOR REVERSAL TEST: $P_{01} \times Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = V_{01}$ (Value Index)

Satisfied ONLY by: Fisher's Ideal Index.

3 CIRCULAR TEST: $P_{01} \times P_{12} \times P_{20} = 1$

Satisfied by: Simple Geometric Mean of Price Relatives, Kelly's Index.