MEAN :

RAW DATA : $\bar{x} = \frac{\sum x}{n}$

GROUPED DATA : $\bar{x} = \frac{\sum fx}{N} = \frac{\sum fx}{\sum f}$

Note : The sum of the deviation of the observation from the AM is always zero.

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

COMBINED Mean : $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

Weighted Arithmetic Mean : $\bar{x} = \frac{\sum wx}{w}$

where w is the weight.

Geometric Mean : $\sqrt[n]{x_1 \times x_2 \times \cdots\cdots x_n}$

RAW DATA : $\text{Antilog}\left[\frac{1}{n} \sum \log(x_i)\right]$

GROUPED DATA : $\text{Antilog}\left[\frac{1}{\sum f_i} \sum f_i \log(x_i)\right]$

Harmonic Mean : $\frac{n}{\sum 1/x}$ where $x \neq 0$.

Raw data : $\frac{n}{\sum 1/x}$ $\{x \neq 0\}$

Grouped data : $\frac{N}{\sum f/x}$ or $\frac{F}{\sum f/x}$ $\{x \neq 0\}$

MEDIAN :

median (m) = $\left(\frac{n+1}{2}\right)^{th}$

where n is number of observations.

Continuous frequency distribution :

$$L_1 + \frac{\left(N/2 - Pcf\right) * C}{F}$$

where L_1 = lower limit of median class

pcf = previous cumulative frequency.

C = class width.

f = class frequency.

N = total number of observation.

Mode :

(Z)= most frequent observation.

Continuos frequency distribution :

$$(Z) = L_1 + \frac{(f_1 - f_0) * c}{(2f_1 - f_0 - f_2)}$$

where, L_1 : lower limit of modal class
f_1 : freq. of modal class
f_0 : freq. of preceding class
f_2 : freq. of next class
c : Class width.

	Discrete	Continuos.
Quartile	$Q_i = \{i(n+1)/4\}^{th}$	$l + \frac{(iN/4 - Pcf)}{f} * c$
Decile	$D_i = \{i(N+1)/10\}^{th}$	$L + \frac{(iN/10 - Pcf)}{f} * c$
Percentile.	$P_i \{i(N+1)/100\}^{th}$.	$L + \frac{(iN/100 - Pcf)}{f} * c$

Scanned by CamScanner

Range :

Highest value - lowest value.

Co-efficient of Range : $\frac{H - L}{H + L}$

Quartile Deviation : $\frac{Q_3 - Q_1}{2}$

Co-efficient of Quartile Deviation : $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Interquartile Range : $Q_3 - Q_1$.

Mean Deviation about Mean.

RAW DATA : $\bar{x} = \sum \left| \frac{x - \bar{x}}{n} \right|$

tabulated data : $\bar{x} = \frac{\sum f \left| x - \bar{x} \right|}{N}$

Relative Measure : $\frac{M.D.(\bar{x})}{\bar{x}}$

co-efficient of $M.D(\bar{x})$

Mean deviation about Median

for Rawdata : $\frac{\sum |x - M|}{n}$

for tabulated data : $\frac{\sum f |x - M|}{N}$

Relative Measure : → $\frac{M.D.(M)}{M}$
coeff. of M.D (M)

Standard deviation.

	Rawdata	Tabulated data.
for population :	$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$	$\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{N}}$
for sample :	$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$	$\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{N-1}}$

Co-efficient of Variation (C.V)

$$C.V = \frac{\text{standard deviation}}{\text{Arithemetic Mean}} \times 100$$

$$= \frac{\sigma}{\bar{x}} \times 100.$$

Central Moments :

Raw data :- $\mu_r : \frac{\sum (x-\bar{x})^r}{n}$

Tabulated data - $\mu_r : \frac{\sum f (x-\bar{x})^r}{N}$

Raw Moments

Raw data : $\mu_r' = \frac{\sum (x-a)^r}{n}$

tabulated data : $\mu_r' = \frac{\sum f (x-a)^r}{N}$

where 'a' is any constant expect 'm'.

∴ mean = median = mode { symmetric }

mode < median < mean { +ve skewed }

mean < median < mode { -ve skewed }

Skewness Based on Moments

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \text{or} \quad \gamma_1 = +\sqrt{\beta_1}$$

Karl pearson's co-efficient of skewness

$$S = \frac{\text{mean} - \text{mode}}{S.D} = \frac{\bar{x} - Z}{\sigma}$$

If mode is ill defined,

$$S = \frac{3\,(\text{mean} - \text{median})}{S.D} = \frac{3\,(\bar{x} - M)}{\sigma}$$

since we know, $Z = 3M - 2\bar{x}$

Bowley's Co-efficient of sknewness (Galton's).

$$S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

Kurtosis (Based on Moments)

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad \text{OR} \quad \gamma_1 = \beta_2 - 3$$

β_2	
3	Mesokurtic (Normal)
< 3	Platykurtic
> 3	leptokurtic

Sampling Distribution :- Mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Where the population SD is known

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Where sample size is large, and population SD is not known

$$\bar{x} - Z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

- Where sample size is ≤ 30, student t-distribution is used (degrees of freedom $= n-1$)

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

In the Case of sampling distribution of the mean, the margin of error is:

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Therefore

$$n = \left(Z_{\frac{\alpha}{2}} \cdot \sigma / E\right)^2$$

- Variance

$$\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}$$

Confidence interval :- Variance and SD

Chi-squared CDF is given by:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

C.I.

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right)^{1/2} < \sigma \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)^{1/2}$$

Confidence Interval :- Proportions:-

Test statistics:-

$$Z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}}$$

C.I.

$$\hat{p} - Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

C.I. Difference between means

- population variances known

$$(\bar{X}-\bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$< \mu_1 - \mu_2$$

$$< (\bar{X}-\bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- population variances not known ; Sample > 30

$$(\bar{X}-\bar{Y}) - Z\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$< \mu_1 - \mu_2$$

$$< (\bar{X}-\bar{Y}) + Z\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- population variances not known; Sample < 30

$$(\bar{X}-\bar{Y}) - t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$< \mu_1 - \mu_2$$

$$< (\bar{X}-\bar{Y}) + t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

D.O.F.

$$\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2-1)}}$$

Confidence Interval

:- Difference between Means

$$S^2_{pooled} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$$

if $\frac{S_1}{S_2}$ = between 0.5 and 2 we are using above formula

$$T = \frac{(X-Y)-(\mu_1-\mu_2)}{S_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$D.O.F. = (n_1+n_2-2)$$

C.I. =

$$(\bar{X}-\bar{Y}) - t_{\alpha/2} \cdot S_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{X}-\bar{Y}) + t_{\alpha/2} \cdot S_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Date : Subject : Topic :

C.I. of Difference of Means of "paired samples":

$$d_i = X_i - Y_i, \quad i = 1, 2, \cdots n$$

$$\bar{d} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} d_i, \text{ and}$$

$$S_d^2 = \left\{\frac{1}{(n-1)}\right\} \sum_{i=1}^{n} (d_i - \bar{d})^2$$

$$T = \frac{\bar{d} - (\mu_1 - \mu_2)}{S_d/\sqrt{n}} \quad D.O.F. = n-1$$

$$\bar{d} - t_{\alpha/2} S_d/\sqrt{n} < \mu_d < \bar{d} + t_{\alpha/2} \frac{S_d}{\sqrt{n}}$$

Confidence Interval: Ratio of Variances

$$\text{Test statistic} = \left(\frac{S_1^2}{\sigma_1^2}\right) \Big/ \left(\frac{S_2^2}{\sigma_2^2}\right) = F = \frac{U/r_1}{V/r_2}$$

$$U = \frac{(n_1-1)S_1^2}{\sigma_1^2} \qquad V = \frac{(n_2-1)S_2^2}{\sigma_2^2}$$

$$r_1 = (n_1 - 1) \text{ \& } r_2 = (n_2 - 1)$$

C.I.

$$\frac{S_2^2}{S_1^2} \times F(\alpha/2, r_2, r_1) < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} \times F(1-\tfrac{\alpha}{2}, r_2, r_1)$$

Hypothesis Testing about one proportion

- Best Estimate $\hat{p} = x/n$
- standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

- Conditions
  - The sample is a simple random sample
  - Sample values are independent of each other.
  - $np(1-p) \geq 10$

- Test statistic for proportions

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Hypothesis Testing: single parameter

- The mean, when variance is known
  - Test statistic

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- The Mean, when variance is unknown

- Large Sample size $z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$ (normal dist)

- Small sample size

$$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$ (t-distribution)

The Variance

- Test statistic

$$\chi^2 = \frac{(n-1) s^2}{\sigma^2}$$ (Chi-Squared)

- The Proportion

Test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Between-Column variance

$$\hat{\sigma}^2 = \frac{\sum n_j (\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Variances within the Sample

$$\hat{\sigma}^2 = \sum \left(\frac{n_j - 1}{n_T - k}\right) s_j^2$$

$$F = \frac{MSB}{MSE}$$