

Central Tendency & Dispersion

CHEAT SHEET (फररा)

Central Tendency

meaning of central tendency

- central value of all observation
- Representative of entire series

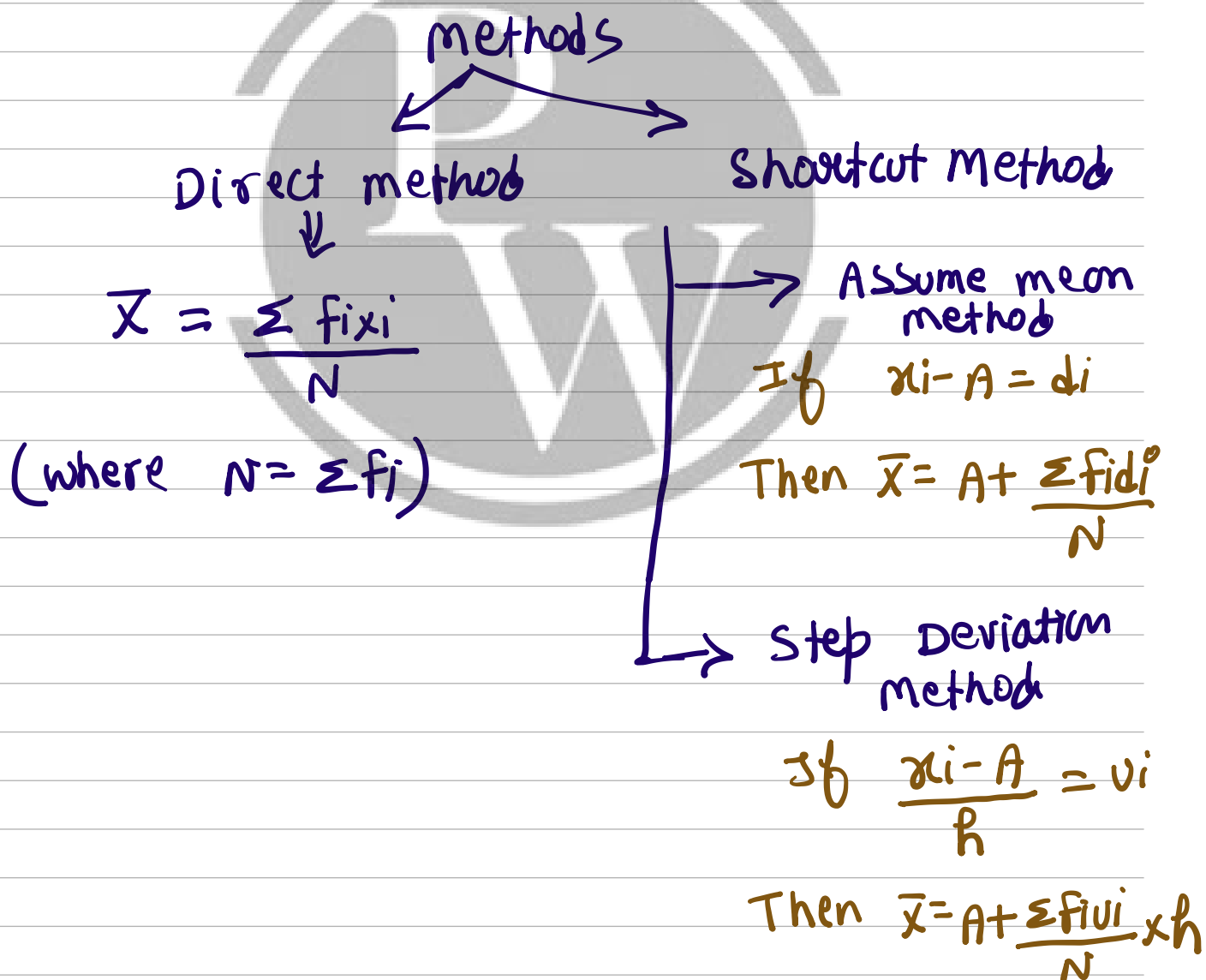
Characters of Good central tendency

- Easy To calculate → mean & mode
- Easy To understand → mean
- Based on all observation → $\bar{m}$, $\bar{u}$, $\bar{H}$
- Rigidly Defined → mean
- least affected by extreme values → median
- Have mathematical properties → mean, $\bar{u}$ & $\bar{H}$

Note : $\rightarrow$ AM is Best central Tendency
 $\rightarrow$ median is best for open ended series

Arithmetic mean ($\bar{x}$)

$$AM = \frac{\text{Sum of all observation}}{\text{Total no. of observations}}$$



Properties

1) If all observations are same (let k)
then mean also k

mean of 5, 5, 5, 5, 5 will be 5

2) Sum of Deviations from their arithmetic mean is always zero

i.e.
$$\sum (x_i - \bar{x}) = 0$$

3) Sum of the squares of Deviation is minimum when Deviations are taken from their arithmetic mean

i.e.
$$\sum (x_i - k)^2 \text{ is minimum}$$

when $k = \bar{x}$

4) Combined mean

$$\bar{x}_{12} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$$

5) Arithmetic mean changes with change of origin & change of scale

i.e. If $y_i = ax_i + b$

Then $\bar{y} = a\bar{x} + b$

$$1+2+3+\dots+n = \frac{n(n+1)}{2}$$

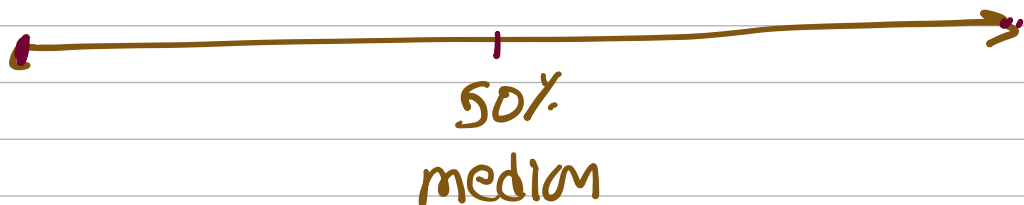
$$1^2+2^2+3^2+\dots+n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^3+2^3+3^3+\dots+n^3 = \left[\frac{n(n+1)}{2} \right]^2$$

A.M. of first 'n' natural numbers = $\frac{n+1}{2}$

Median

A number which divides entire Distribution in two parts is known as median. It represents half (50%) of total numbers.



→ It is not based on all observations so it can be used for open ended series.

methods

for individual &
Discrete series

for continuous
series



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

→ make cf column

→ locate $\frac{N}{2}$ in cf

→ select median class

→ use formula

$$\text{median} = l + \left[\frac{\frac{N}{2} - cf}{f} \right] \times h$$

→ median changes with change
of origin & change in scale
i.e. if $y = a + bx$

Then med. of $y = a + b(\text{med. of } x)$

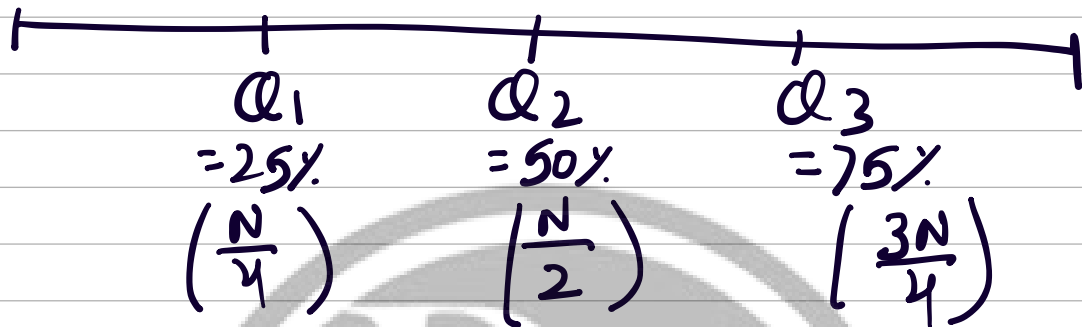
→ Sum of absolute deviation is
minimum when deviations are taken
from the median

i.e. $\sum |x_i - k|$ is minimum
when $k = \text{median}$

Partition values (Fractiles)

Quartiles ($Q_1, Q_2 \text{ \& } Q_3$)

Divides entire series in 4 parts)



For Individual & Discrete series

$$Q_1 = \left(\frac{n+1}{4}\right)^{\text{th}} \text{ term}$$

$$Q_2 = \text{median} = \left(\frac{n+1}{2}\right)^{\text{th}}$$

$$Q_3 = \left[3\left(\frac{n+1}{4}\right)\right]^{\text{th}} \text{ term}$$

In continuous series

for Q_1

→ Locate $\frac{N}{4}$ in CF

→ select ' Q_1 ' class

$$\rightarrow Q_1 = l + \left[\frac{\frac{N}{4} - CF}{f} \right] \times h$$

for Q_3

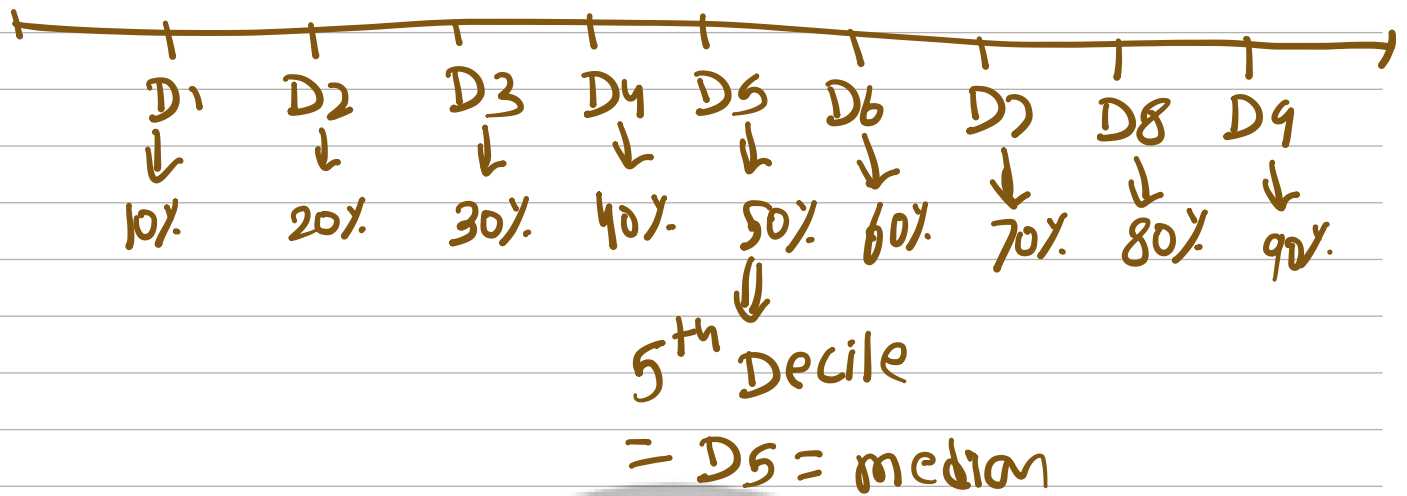
→ Locate $\frac{3N}{4}$ in CF

→ select Q_3 class

$$\rightarrow Q_3 = l + \left[\frac{\frac{3N}{4} - CF}{f} \right] \times h$$

Deciles ($D_1, D_2, \dots, D_9$)

Divides entire series in 10 parts



For individual & Discrete series

$$D_1 = \left(\frac{n+1}{10} \right)^{\text{th}} \text{ term}$$

$$D_2 = \left[2 \left(\frac{n+1}{10} \right) \right]^{\text{th}}$$

$$D_k = \left[k \left(\frac{n+1}{10} \right) \right]^{\text{th}} \text{ term}$$

For Continuous series

For D_1

→ locate $\frac{N}{10}$ in cf

→ select D_1 class

$$\rightarrow D_1 = l + \left[\frac{\frac{N}{10} - cf}{f} \right] \times h$$

for D_3

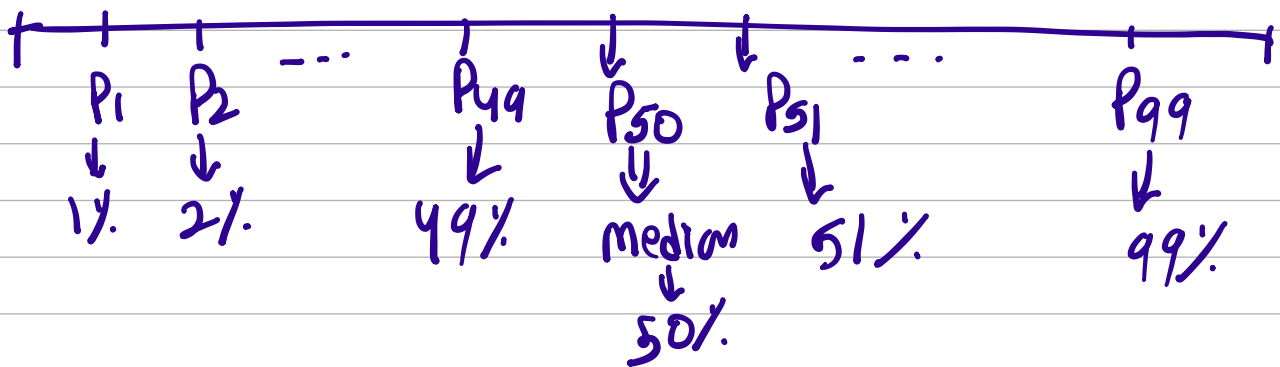
→ locate $\frac{3N}{10}$ in cf

→ select D_3 class

$$\rightarrow D_3 = l + \left[\frac{\frac{3N}{10} - cf}{f} \right] \times h$$

Percentiles ($P_1, P_2, \dots, P_{99}$)

Divides entire series in 100 parts



For Individual & Discrete series

$$P_1 = \left(\frac{n+1}{100} \right)^{\text{th}} \text{ term}$$

$$P_6 = \left[6 \left(\frac{n+1}{100} \right) \right]^{\text{th}} \text{ term}$$

$$P_k = \left[k \left(\frac{n+1}{100} \right) \right]^{\text{th}}$$

For Continuous series

For P_1

→ locate $\frac{N}{100}$ in CF

→ select P_1 class

$$\rightarrow P_1 = l + \left[\frac{\frac{N}{100} - cf}{f} \right] \times h$$

For P_7

→ locate $\frac{7N}{100}$ in CF

→ select P_7 class

$$\rightarrow P_7 = l + \left[\frac{\frac{7N}{100} - cf}{f} \right] \times h$$

mode

An observation with highest frequency

Individual series

g marks: 2, 1, 3, 1, 4, 3, 2, 3, 5, 2, 3, 1, 3, 1, 3, 5, 3, 4, 3

mode = 3

Discrete series

g

x_i	f_i
2	6
3	12
4	3
5	5

→ mode = 3

For continuous series

→ Check the class interval with highest frequency (It is modal class)

→ Use formula

$$\text{mode} = l + \left\{ \frac{f_1 - f_0}{2f_1 - f_2 - f_0} \right\} \times h$$

note: mode also changes with change of origin or change in scale
ie. $y = a + bx$

Then mode of $y = a + b(\text{mode of } x)$

Geometric mean

n^{th} Root of the Product of n observations

for individual series

$$GM = (x_1 \times x_2 \times \dots \times x_n)^{1/n}$$

used for

→ Average Dep

→ Average of %

for Discrete & Continuous series

$$GM = \left[(x_1)^{f_1} \times (x_2)^{f_2} \times \dots \times (x_n)^{f_n} \right]^{\frac{1}{N}}$$

Properties

1) If all items are same (let k)
then $GM = k$

$$2) \log(GM) = \frac{\sum \log x_i}{N}$$

$$GM = AL \left[\frac{\sum \log x_i}{N} \right]$$

$$3) \quad \text{GM}(xy) = \text{GM}(x) \times \text{GM}(y)$$

$$4) \quad \text{GM}\left(\frac{x}{y}\right) = \frac{\text{GM}(x)}{\text{GM}(y)}$$

5) combined Geometric mean

$$\text{GM} = \left[(\text{GM}_1)^{N_1} \times (\text{GM}_2)^{N_2} \right]^{\frac{1}{N_1 + N_2}}$$

#

Harmonic mean

use for
→ Avg Speed
→ Avg of Rates

Reciprocal of Average of
Reciprocal of all 'n' observations

→ find reciprocal of all items

→ find Average of these reciprocals

→ find Reciprocal of Average

For individual series

$$Hm = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{N}{\sum \left(\frac{1}{x_i} \right)}$$

for discrete & continuous series

$$Hm = \frac{N}{\sum \left(\frac{f_i}{x_i} \right)}$$

properties

1> If all items are same (let K)
then $Hm = K$

2> Combined $Hm = \frac{N_1 + N_2}{\frac{N_1}{H_1} + \frac{N_2}{H_2}}$

Relation b/w Am, Gm & Hm

$$Am \geq Gm \geq Hm$$

If all items are different

$$Am > Gm > Hm$$

For any two items a & b

$$Am \times Hm = Gm^2$$

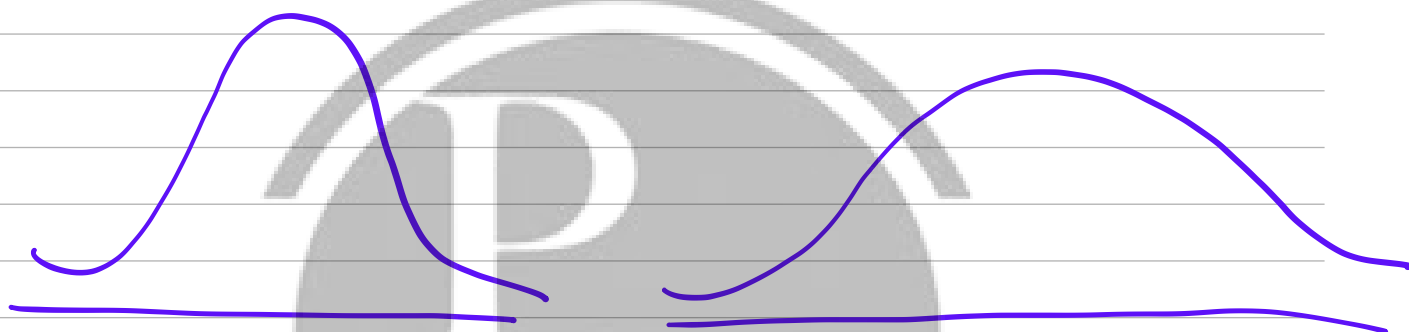
$$\text{weighted } Am = \frac{\sum w_i x_i^2}{\sum w_i}$$

$$\text{weighted } Hm = \frac{\sum w_i}{\sum \left(\frac{w_i}{x_i} \right)}$$

$$\text{weighted } Gm = AL \left[\frac{\sum w_i \log x_i}{\sum w_i} \right]$$

Dispersion

Statistical Technique to find degree consistency (or variability) in all the observation.



- low Dispersion
- concentrated data
- less variability
- more consistency

- High Dispersion
- scattered data
- more variability
- less consistency

measures

Absolute

- Range
- Q.D
- M.D
- S.D.

Relative

- coefficient of Range
- coefficient of Q.D.
- coefficient of M.D
- coefficient of variance

Range

Difference b/w largest & smallest observation

$$R_x = L - S$$

$$\text{Coefficient of Range} = \frac{L - S}{L + S} \times 100$$

→ Range does not change with origin

→ Range changes with scale

$$\begin{array}{l} \text{If } y_i = a + b x \\ \text{then } R_y = |b| \times R_x \end{array}$$

Quartile Deviations

$$\Rightarrow \text{interquartile Range} = Q_3 - Q_1$$

$$\Rightarrow \text{semi quartile Range} = \frac{Q_3 - Q_1}{2}$$

(Quartile Deviation)

$$\Rightarrow \text{coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

or

$$\frac{\text{Q.D.}}{\text{median}} \times 100$$

$\Rightarrow$ only for symmetrical Distribution

$\rightarrow$ QD does not change with origin

$\rightarrow$ QD changes with scale

$$y_i = a + bx$$

$$\text{QD of } y = |b| \times \text{QD of } x$$

mean Deviation (MD)

Average of Absolute Deviations
taken from mean, median or mode



$$MD_{\bar{x}} = \frac{\sum f_i |x_i - \bar{x}|}{N}$$

$$MD_m = \frac{\sum f_i |x_i - m|}{N}$$

Coefficient of M.D. = $\frac{MD}{\text{mean}} \times 100$
or
 $\frac{MD}{\text{median}} \times 100$

$\Rightarrow$ MD does not change with origin

$\Rightarrow$ MD changes with scale

If $y = a + bx$
then $MD_y = |b| \times MD_x$

Standard Deviation & variance

$$\text{variance} = \sigma^2$$

$$\text{S.D.} = \sigma$$

$$\text{S.D.} = \sqrt{\text{variance}}$$

$$\text{variance} = (\text{S.D.})^2$$

$$\# \text{ S.D. } (\sigma) = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}}$$

$$\# \text{ S.D.} = \sqrt{\frac{\sum f_i x_i^2}{N} - \left(\frac{\sum f_i x_i}{N}\right)^2}$$

$$\# \text{ S.D.} = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2}$$

$$\text{where } d_i = x_i - A$$

$$\# \text{ S.D.} = \sqrt{\frac{\sum f_i u_i^2}{N} - \left(\frac{\sum f_i u_i}{N}\right)^2} \times h$$

$$\text{where } \frac{x_i - A}{h} = u_i$$

$\Rightarrow$ S.D. Does not change with origin

$\Rightarrow$ S.D. Change with change of scale

$$\text{If } y_i = a + bx$$

$$\text{S.D of } y_i = |b| \times \text{S.D of } x$$

$$\text{variance of } y_i = b^2 \times \text{variance of } x_i$$

$$\# \left\{ \begin{array}{l} \text{Coefficient of variance} \\ \text{(CV)} \end{array} = \frac{\sigma}{\bar{x}} \times 100 \right\}$$



will be used for consistency & variability

Higher CV $\Rightarrow$ Higher variability

lesser CV $\Rightarrow$ Higher consistency

$$\# \text{ S.D. of first 'n' natural number} = \sqrt{\frac{n^2-1}{12}}$$

$$\# \text{ S.D. of first 'n' even natural no.} = \sqrt{\frac{n^2-1}{3}}$$

$$\# \text{ S.D. of first 'n' odd natural no.} = \sqrt{\frac{n^2-1}{3}}$$

$$\# \text{ S.D. of two numbers a \& b} = \frac{|a-b|}{2}$$

$$\# \text{ Q.D : M.D : S.D} = 10 : 12 : 15$$

$$\frac{\text{Q.D}}{\text{M.D}} = \frac{10}{12} \quad \bigg| \quad \frac{\text{M.D}}{\text{S.D}} = \frac{12}{15} \quad \bigg| \quad \frac{\text{Q.D}}{\text{S.D}} = \frac{10}{15}$$

Combined S.D.

$$= \sqrt{\frac{N_1 (\sigma_1^2 + d_1^2) + N_2 (\sigma_2^2 + d_2^2)}{N_1 + N_2}}$$

where $d_1 = \bar{x}_{12} - \bar{x}_1$

$$d_2 = \bar{x}_{12} - \bar{x}_2$$

$$\bar{x}_{12} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$$