Symbols and Meanings

X = dataset X
$x = x - values$ OR data values
x_{mid} = class midpoint of $x - values$ = class midpoint of the data values
Σ (pronounced as uppercase Sigma) = $summation$
Σx = summation of the $x - values$
$f = frequency$
$F = frequency$
Σf = summation of the frequencies
Σfx = summation of the product of the $x - values$ and their corresponding frequencies
$(\Sigma x)^2$ = square of the summation of the $x - values$

Σx^2 = summation of the squared of the $x - values$

$\bar{x}$ is sample mean of the $x - values$

μ = population mean

n = sample size

N = population size

$\tilde{x}$ = median

$\widehat{x}$ = mode

AM = assumed mean

D = deviation from the assumed mean

x_{MR} = midrange

LCL = lower class limit

UCL = upper class limit

min = minimum data value

max = maximum data value

LCB_{med} = lower class boundary of the median class

CW = class width

f_{med} = frequency of the median class

CF_{bmed} = cumulative frequency of the class before the median class

LCB_{mod} = lower class boundary of the modal class

f_{mod} = frequency of the modal class

f_{bmod} = frequency of the class before the modal class

f_{amod} = frequency of the class after the modal class

R = range	Q_3 = upper quartile or third quartile
s = sample standard deviation	P_{75} = 75th percentile or third quartile
s^2 = sample variance	IQR = interquartile range
σ = population standard deviation	$SIQR$ = semi-interquartile range
σ^2 = population variance	MQ = midquartile
CV = coefficient of variation	LF = lower fence
$z = z - score$	UF = upper fence
Q_1 = lower quartile or first quartile	TM = trimmed mean
P_{25} = 25th percentile or first quartile	Π (pronounced as uppercase Pi) = $product$
Q_2 = middle quartile or second quartile or median	Πx = product of the $x - values$
P_{50} = 50th percentile or median	GM = geometric mean

Formulas: Measures of Central Tendency

Raw Data and Ungrouped Data

$\underline{Sample\ \ Mean}$

(1.) $\bar{x} = \frac{\Sigma x}{n}$

(2.) $n = \Sigma f$

(3.) $\bar{x} = \frac{\Sigma f x}{\Sigma f}$

$\underline{Given\ \ an\ \ Assumed\ \ Mean}$

(4.) $D = x - AM$

(5.) $\bar{x} = AM + \frac{\Sigma D}{n}$

(6.) $\bar{x} = AM + \frac{\Sigma f D}{\Sigma f}$

$\underline{Population\ \ Mean}$

$$(7.)\ \ \mu = \frac{\Sigma x}{N}$$

$$(8.)\ \ N = \Sigma f$$

$\underline{Given\ \ an\ \ Assumed\ \ Mean}$

$$(9.)\ \ D = x - AM$$

$$(10.)\ \ \mu = AM + \frac{\Sigma D}{N}$$

$$(11.)\ \ \mu = AM + \frac{\Sigma fD}{\Sigma f}$$

$\underline{Median}$

$$(12.)\ \ \tilde{x} = \left(\frac{\Sigma f + 1}{2}\right) th\ \ for\ \ sorted\ \ odd\ \ sample\ \ size$$

$$(13.)\ \ \tilde{x} = \left(\frac{\Sigma f}{2}\right) th\ \ for\ \ sorted\ \ even\ \ sample\ \ size$$

$\underline{Mode}$

$$(14.)\ \ Mode = x - value(s)\ \ with\ \ highest\ \ frequency$$

$\underline{Midrange}$

$$(15.)\ \ x_{MR} = \frac{min + max}{2}$$

$\underline{Geometric\ \ Mean}$

$$(16.)\ \ GM = \sqrt[n]{\prod_{x=1}^{n} x}$$

Grouped Data

$\underline{Class\ \ Midpoint}$

$$(1.)\ \ x_{mid} = \frac{LCL + UCL}{2}$$

$Equal\ \ Class\ \ Intervals\ (Same\ \ Class\ \ Size)$

$\underline{Mean}$

$$(2.)\ \ \bar{x} = \frac{\Sigma f x_{mid}}{\Sigma f}$$

$Equal\ \ Class\ \ Intervals\ (Same\ \ Class\ \ Size)$

$\underline{Given\ \ an\ \ Assumed\ \ Mean}$

$$(3.)\ \ D = x_{mid} - AM$$

$$(4.)\ \ \bar{x} = AM + \frac{\Sigma f D}{\Sigma f}$$

$\underline{Median}$

$$(5.)\ \ \tilde{x} = LCB_{med} + \frac{CW}{f_{med}} * \left[\left(\frac{\Sigma f}{2}\right) - CF_{bmed}\right]$$

Mode

$$(6.)\ \widehat{x} = LCB_{mod} + CW * \left[\frac{f_{mod} - f_{bmod}}{(f_{mod} - f_{bmod}) + (f_{mod} - f_{amod})}\right]$$

Formulas: Measures of Dispersion

Raw Data and Ungrouped Data

$\underline{Range}$

(1.) $Range = max - min$

$\underline{Using\ \ Assumed\ \ Mean}$

(2.) $D = x - AM$

$\underline{Sample\ \ Variance}$

$First\ \ Formula$

(3.) $s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$

(4.) $s^2 = \frac{\Sigma f(x - \bar{x})^2}{\Sigma f - 1}$

Second Formula

$$(5.)\quad s^2 = \frac{n(\Sigma x^2) - (\Sigma x)^2}{n(n-1)}$$

$$(6.)\quad s^2 = \frac{\Sigma f(\Sigma f x^2) - (\Sigma f x)^2}{\Sigma f(\Sigma f - 1)}$$

Using Assumed Mean

$$(7.)\quad s^2 = \frac{\Sigma D^2}{n-1} - \left(\frac{\Sigma D}{n-1}\right)^2$$

$$(8.)\quad s^2 = \frac{\Sigma f D^2}{\Sigma f - 1} - \left(\frac{\Sigma f D}{\Sigma f - 1}\right)^2$$

Population Variance

First Formula

$$(9.)\ \sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$$

$$(10.)\ \sigma^2 = \frac{\Sigma f(x-\mu)^2}{\Sigma f}$$

Second Formula

$$(11.)\ \sigma^2 = \frac{N(\Sigma x^2) - (\Sigma x)^2}{N^2}$$

$$(12.)\ \sigma^2 = \frac{\Sigma f(\Sigma f x^2) - (\Sigma f x)^2}{(\Sigma f)^2}$$

Using Assumed Mean

$$(13.)\ \sigma^2 = \frac{\Sigma D^2}{N} - \left(\frac{\Sigma D}{N}\right)^2$$

$$(14.)\ \sigma^2 = \frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2$$

Sample Standard Deviation

First Formula

$$(15.)\ s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$(16.)\ s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f - 1}}$$

Second Formula

$$(17.)\quad s = \sqrt{\frac{n(\Sigma x^2) - (\Sigma x)^2}{n(n-1)}}$$

$$(18.)\quad s = \sqrt{\frac{\Sigma f(\Sigma f x^2) - (\Sigma f x)^2}{\Sigma f(\Sigma f - 1)}}$$

Using Assumed Mean

$$(19.)\quad s = \sqrt{\frac{\Sigma D^2}{n-1} - \left(\frac{\Sigma D}{n-1}\right)^2}$$

$$(20.)\quad s = \sqrt{\frac{\Sigma f D^2}{\Sigma f - 1} - \left(\frac{\Sigma f D}{\Sigma f - 1}\right)^2}$$

Population Standard Deviation

First Formula

$$(21.)\quad \sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

$$(22.)\quad \sigma = \sqrt{\frac{\Sigma f(x-\mu)^2}{\Sigma f}}$$

Second Formula

$$(23.)\quad \sigma = \frac{\sqrt{N(\Sigma x^2)-(\Sigma x)^2}}{N}$$

$$(24.)\quad \sigma = \frac{\sqrt{\Sigma f(\Sigma f x^2)-(\Sigma f x)^2}}{\Sigma f}$$

Using Assumed Mean

$$(25.)\ \ \sigma = \sqrt{\frac{\Sigma D^2}{N} - \left(\frac{\Sigma D}{N}\right)^2}$$

$$(26.)\ \ \sigma = \sqrt{\frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2}$$

Range Rule of Thumb

Approximate Value of Calculating Standard Deviation

$$(27.)\ \ s = \frac{Range}{4} = \frac{max - min}{4}$$

$\underline{Interquartile\ \ Range}$

(28.) $IQR = Q_3 - Q_1$

$\underline{Coefficient\ \ of\ \ Variation\ \ for\ \ Sample}$

(29.) $CV = \frac{s}{x} * 100 \ldots in\ \ \%$

$\underline{Coefficient\ \ of\ \ Variation\ \ for\ \ Population}$

(30.) $CV = \frac{\sigma}{x} * 100 \ldots in\ \ \%$

$\underline{Mean\ \ Absolute\ \ Deviation}$

(31.) $MAD = \frac{\Sigma|x - \bar{x}|}{n}$

$\underline{Mean\ \ Absolute\ \ Deviation}$

(32.) $MAD = \frac{\Sigma f|x - \bar{x}|}{\Sigma f}$

Grouped Data

$\underline{Class\ \ Midpoint}$

$$(1.)\ \ x_{mid} = \frac{LCL + UCL}{2}$$

$\underline{Using\ \ Assumed\ \ Mean}$

$$(2.)\ \ D = x_{mid} - AM$$

$\underline{Sample\ \ Variance}$

$First\ \ Formula$

$$(3.)\ \ s^2 = \frac{\Sigma f(x_{mid} - \bar{x})^2}{\Sigma f - 1}$$

$Second\ \ Formula$

$$(4.)\ \ s^2 = \frac{\Sigma f(\Sigma f x_{mid}^2) - (\Sigma f x_{mid})^2}{\Sigma f(\Sigma f - 1)}$$

Using Assumed Mean

$$(5.)\ \ s^2 = \frac{\Sigma D^2}{n-1} - \left(\frac{\Sigma D}{n-1}\right)^2$$

$$(6.)\ \ s^2 = \frac{\Sigma f D^2}{\Sigma f - 1} - \left(\frac{\Sigma f D}{\Sigma f - 1}\right)^2$$

Sample Standard Deviation

First Formula

$$(7.)\ \ s = \sqrt{\frac{\Sigma f(x_{mid} - \bar{x})^2}{\Sigma f - 1}}$$

Second Formula

$$(8.)\quad s = \sqrt{\frac{\Sigma f(\Sigma f x_{mid}^2) - (\Sigma f x_{mid})^2}{\Sigma f(\Sigma f - 1)}}$$

Using Assumed Mean

$$(9.)\quad s = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n-1}\right)^2}$$

$$(10.)\quad s = \sqrt{\frac{\Sigma f D^2}{\Sigma f - 1} - \left(\frac{\Sigma f D}{\Sigma f - 1}\right)^2}$$

Population Variance

First Formula

$$(11.)\ \sigma^2 = \frac{\Sigma f(x_{mid} - \bar{x})^2}{\Sigma f}$$

Second Formula

$$(12.)\ \backslash\text{sigm}^2 = \frac{\Sigma f(\Sigma f x_{mid}^2) - (\Sigma f x_{mid})^2}{\Sigma f(\Sigma f)}$$

Using Assumed Mean

$$(13.)\ \sigma^2 = \frac{\Sigma D^2}{N} - \left(\frac{\Sigma D}{N}\right)^2$$

$$(14.)\ \sigma^2 = \frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2$$

Population Standard Deviation

First Formula

$$(15.)\ \sigma = \sqrt{\frac{\Sigma f(x_{mid} - \bar{x})^2}{\Sigma f}}$$

Second Formula

$$(16.)\ \sigma = \sqrt{\frac{\Sigma f(\Sigma f x_{mid}^2) - (\Sigma f x_{mid})^2}{\Sigma f(\Sigma f)}}$$

Using Assumed Mean

$$(17.)\ \sigma = \sqrt{\frac{\Sigma D^2}{N} - \left(\frac{\Sigma D}{N}\right)^2}$$

$$(18.)\ \sigma = \sqrt{\frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2}$$

Relationship Between Quantiles

Fill in the blank

Percentile	**Decile**	**Quintile**	**Quartile**
10th	1st	📖	📖
20th	2nd	1st	📖
25th	2.5th	1.25th	1st
30th	3rd	📖	1.2nd
40th	📖	2nd	📖
50th	5th	2.5th	2nd
75th	📖	3.75th	3rd
100th	10th	5th	4th

Formulas: Measures of Location

A **data value is usual** if $-2.00 \leq z - score \leq 2.00$

A **data value is unusual** if $z - score < -2.00$ OR $z - score > 2.00$

$\underline{Sample}$

$$Minimum\ \ usual\ \ data\ \ value = \bar{x} - 2s$$

$$Maximum\ \ usual\ \ data\ \ value = \bar{x} + 2s$$

$\underline{Population}$

$$Minimum\ \ usual\ \ data\ \ value = \mu - 2\sigma$$

$$Maximum\ \ usual\ \ data\ \ value = \mu + 2\sigma$$

$\underline{z\ \ score\ \ for\ \ Sample}$

$$(1.)\ \ z = \frac{x - \bar{x}}{s}$$

$\underline{z\ \ score\ \ for\ \ Population}$

$$(2.)\ \ z = \frac{x - \mu}{\sigma}$$

$\underline{Quantiles(Percentiles,\ Deciles,\ Quintiles,\ and\ Quartiles)}$

$Convert\ \ a\ \ Data\ \ value\ \ to\ \ a\ \ Quantile$

$x\ \ and\ \ y\ \ are\ \ two\ \ different\ \ variables$

$$(3.)\ \ Percentile\ \ of\ \ x = \frac{number\ \ of\ \ values\ \ less\ \ than\ \ x}{total\ \ number\ \ of\ \ values} * 100 = yth\ \ Percentile$$

$$(4.)\ \ Decile\ \ of\ \ x = \frac{number\ \ of\ \ values\ \ less\ \ than\ \ x}{total\ \ number\ \ of\ \ values} * 10 = yth\ \ Decile$$

$$(5.)\ \ Quintile\ \ of\ \ x = \frac{number\ \ of\ \ values\ \ less\ \ than\ \ x}{total\ \ number\ \ of\ \ values} * 5 = yth\ \ Quintile$$

$$(6.)\ Quartile\ of\ x = \frac{number\ of\ values\ less\ than\ x}{total\ number\ of\ values} * 4 = yth\ Quartile$$

Convert a Quantile to a Data Value

Calculate the xth position of the yth Quantile

$$(7.)\ xth\ position = \frac{yth\ Percentile}{100} * total\ number\ of\ values$$

$$(8.)\ xth\ position = \frac{yth\ Decile}{10} * total\ number\ of\ values$$

$$(9.)\ xth\ position = \frac{yth\ Quintile}{5} * total\ number\ of\ values$$

$$(10.)\ xth\ position = \frac{yth\ Quartile}{4} * total\ number\ of\ values$$

If the xth position	then,
is an integer	$xth\ \ position = \frac{xth\ \ position + (x+1)th\ \ position}{2}$ In other words, find the value of the xth position; find the value of the next position; and determine the mean of the two values.
is not an integer	xth position is **rounded up**

$\underline{The\ \ Five - Number\ \ Summary\ \ of\ \ Data}$

(11.) $Minimum\,(min)$

(12.) $Lower\ \ Quartile\,(Q_1)$

(13.) $Median\ \ or\ \ Middle\ \ Quartile\,(Q_2)$

(14.) $Upper\ \ Quartile\,(Q_3)$

(15.) $Maximum\,(Max)$

Other Statistics from Quantiles

(16.) $IQR = Q_3 - Q_1$

(17.) $SIQR = \frac{IQR}{2} = \frac{Q_3 - Q_1}{2}$

(18.) $MQ = \frac{Q_3 + Q_1}{2}$

(19.) $Upper\ Quartile\ (Q_3)$

(20.) $LF = Q_1 - 1.5(IQR)$

(21.) $UF = Q_3 + 1.5(IQR)$

The **Empirical Rule** also known as the 68–95–99.7 percent Rule applies to distributions that are unimodal and symmetric.

It states that:

In a unimodal symmetric distribution (normal distribution) dataset, approximately:

(a.) 68% of the distribution lie within (below and above) one standard deviation from the mean

(b.) 95% of the distribution lie within (below and above) two standard deviations from the mean

(c.) 99.7% of the distribution lie within (below and above) three standard deviations from the mean

In other words, in a unimodal normal distribution (bell-shaped curve distribution):

(a.) 68% of the distribution lie from $\mu - \sigma$ to $\mu + \sigma$

(a.) 95% of the distribution lie from $\mu - 2\sigma$ to $\mu + 2\sigma$

(a.) 99.7% of the distribution lie from $\mu - 3\sigma$ to $\mu + 3\sigma$

Let us review an example:

Empirical Rule (68 — 95 — 99.7 percent Rule)