
Descriptive Biostatistics

Measures of Dispersion

Measures of dispersion including:

1. Range.
2. Variance.
3. Standard deviation.

Variability of data

Dispersion refers to the variability exhibited by the values of the data. The amount may be small when the values are close together.

Range:

The range is the difference between the largest and smallest values in the set of observations. These values are often called the maximum and the minimum.

Finding Range for Ungrouped Data

Range = Largest value – Smallest Value

Disadvantages:

The range, like the mean has the disadvantage of being influenced by outliers.

Its calculation is based on two values only: the largest and the smallest.

Variance and Standard Deviation

The standard deviation is the most used measure of dispersion. The value of the standard deviation tells how closely the values of a data set are clustered around the mean.

In general, a lower value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively smaller range around the mean.

In contrast, a large value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively large range around the mean. The Variance calculated for population

data is denoted by σ^2 (read as sigma squared), and the variance calculated for sample data is denoted by s^2.

The standard deviation calculated for population data is denoted by σ, and the standard deviation calculated for sample data is denoted by s.

Calculation of the sample variance

s^2: Sample variance.

x_i: Individual value.

$\bar{x}$ = Sample mean.

n = number of values.

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Variance of a population

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

σ^2 = population variance.

N = population size.

μ = population mean.

Sample standard deviation

The standard deviation is obtained by taking the positive square root of the variance.

Sample standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Population Standard Deviation

For a population, the standard deviation which is the square root of the population variance.

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

$$\sigma^2 = \sum_{i=1}^{N} (X_i - \mu)^2 / N \quad \text{or} \quad \sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right], \text{ (Population)}$$

$$s^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / n-1 \quad \text{or} \quad s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right], \text{ (Sample)}$$

Manual Calculation of a Standard Deviation

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16
		40

$\bar{x} = 6$

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

$$= \sqrt{\frac{40}{4}}$$

$$= \sqrt{10}$$

Coefficient of variation

Coefficient of variation is a measure of the relative amount of variation.

$$C.V. = \frac{s}{\bar{x}} (100)$$

C.V. is independent of the units of measure.

Standard Error of the mean (SE)

It measures the variability or dispersion of the sample mean from population mean.

It is used to estimate the population mean, and to estimate differences between populations means.

$SE = SD/\sqrt{n}$

Example:

A set of data (4, 6, 3, 4, 5, 2).

Compute: (Range,Variance,Standard deviation and Coefficient of variation).

R = 6 – 2 = 4

$$s^2 = \frac{1}{6-1}\left[(4^2 + 6^2 + \ldots + 2^2) - \frac{(4+6+\ldots+2)^2}{6}\right] = 2$$

$$s = \sqrt{2} = 1.414$$

$$C.V = \frac{1.414}{4}100\% = 35.35\%$$

Quintiles: It is a value below which certain proportion of observations occurred in the ordered set of data values.

Centiles:

Those values, in a series of observations arranged in ascending order of magnitude, which divide the distribution into 100 equal parts.

10th Percentile: it is the value below which 10% of the observations lie.

We also frequently used 3rd, 97th, and the 50th (median) percentile.

Quartiles:

These are the observations in an array that divide the distribution into four equal parts.

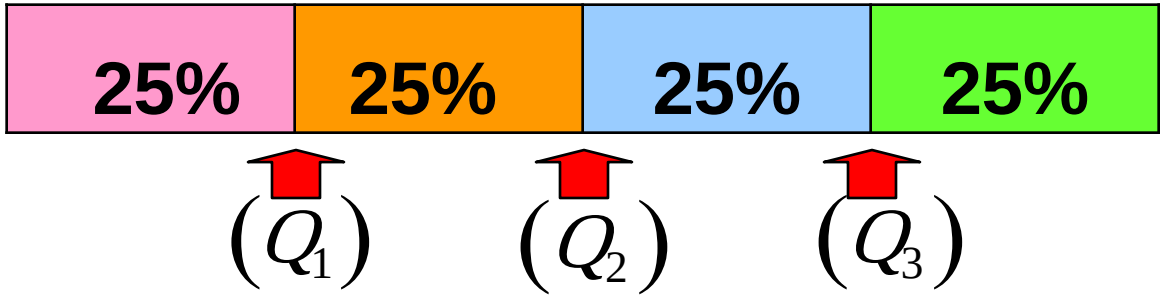
1st (lower Quartile): the value below which 25% of observations lie in an ordered array.

2nd quartile = Median = 50th percentile.

Upper Quartile = 75th percentile.

Quartiles

- Split Ordered Data into 4 Quarters



Q_1 = first quartile

Q_2 = second quartile = Median

Q_3 = third quartile

References

١. WAYNE W. DANIEL,BIOSTATISTICS A Foundation for Analysis in the Health Sciences.10th edition .2013. chapter II, page 38.

٢. Review of Preventive and Social Medicine (Including Biostatistics), 7 th edition, 2015. chapter 16, p: 860.

Thank you

Mean...

Median...

Mode...

Range...

What do they mean??