د. ناهد الجميلي

Descriptive Biostatistics

A descriptive measure is a single number that is used to describe a set of data. Descriptive measures include measures of central tendency and measures of dispersion. The best way to work with data is to summarize and organize them. Numbers that have not been summarized and organized are called raw data.

Central Tendency:

It is a property of the data that they tend to be clustered about a centers point and the various methods of determining the actual value at which the data tend to concentrate are called measures of central Tendency or averages. Hence, an average is a value which tends to sum up or describe the mass of the data. The measures of central tendency include:

– Mean (generally not part of the data set)
– Median (may be part of the data set)
– Mode (always part of the data set).

Measures of Dispersion

Dispersion is a property of the data that they tend to be spread out. In statistics, dispersion (also called variability, scatter, or spread) is the extent to which a distribution is stretched or squeezed.

it is included:

o Range
o Variance
o Standard deviation
o Coefficient of variation

Central Tendency

1. The mean or arithmetic mean: is the "average" which is obtained by adding all the values in a sample or population and dividing them by the number of values.

The *mean* is:

the arithmetic average of all the scores
$(\Sigma X)/N$

The mean of a population is represented by the Greek letter μ;
the mean of a sample is represented by $\bar{x}$

For example:

Calculate the mean of the following data:
1 5 4 3 2

Sum the scores (ΣX):
1 + 5 + 4 + 3 + 2 = 15

Divide the sum ($\Sigma X = 15$) by the number of scores ($N = 5$):
15 / 5 = 3

Mean = $\bar{X}$=3

2. Median: is the value that divides the set of data into two equal parts. It is the midpoint of the data set.

Finding the Median

Conceptually, it is easy to calculate the median:

-Sort the data from highest to lowest

-Find the score in the middle

-middle = (N + 1) / 2

What is the median of the following scores:
10 8 14 15 7 3 3 8 12 10 9

Sort the scores:
15 14 12 10 10 9 8 8 7 3 3

Determine the middle score:
middle = (N + 1) / 2 = (11 + 1) / 2 = 6

Middle score = median = 9

What is the median of the following scores:
24 18 19 42 16 12

Sort the scores:
42 24 19 18 16 12

Determine the middle score:
middle = (N + 1) / 2 = (6 + 1) / 2 = 3.5

Median = average of 3rd and 4th scores:
(19 + 18) / 2 = 18.5

3. Mode: The mode is primarily used with nominally scaled data. It is the only measure of central tendency that is appropriate for nominally scaled data. A particular disadvantage is that, with a small number of observations, there may be no mode. The mode is not often used in biological or medical data.

the mode is the most frequently occurring number in a distribution
if X = [1,2,4,7,7,7,8,10,12,14,17]
then 7 is the mode.

Measures of Dispersion

.
Range (R): In statistics, the range of a set of data is the difference between the largest and smallest values. The range of a set of data is the result of subtracting the smallest value from largest value. The range is used when you have ordinal data or you are presenting your results to people with little or no knowledge of statistics

The range is rarely used in scientific work as it is fairly insensitive

Range=(Highest value) – (Lowest value)
R = XL – XS
Example: Range of {1, 3,6,11, 14}
14-1=13
Note: The range is poor measure of dispersion? Because it only takes into account two of the values.

Variance:

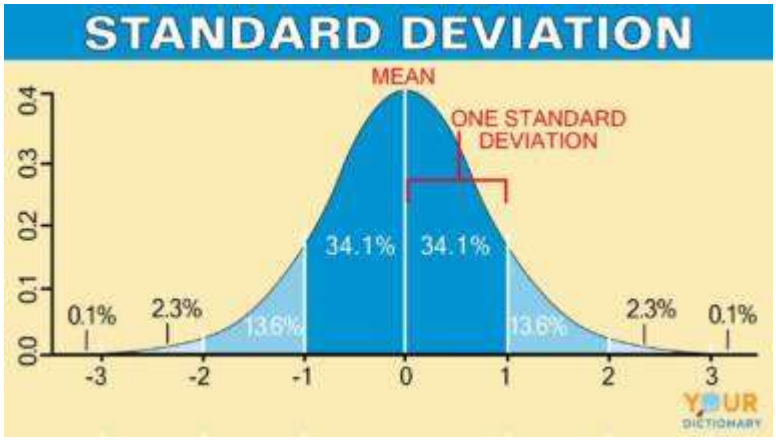
The variance is the most commonly used to measure of spread in biological statistics. The variance of a set of values is a measure of variation equal to the square of the standard deviation. The term variance refers to a statistical measurement of the spread between numbers in a data set. More specifically, variance measures how far each number in the set is from the mean and thus from every other number in the set.

Standard Deviation (s) or (sd): Is a number used to tell how measurements for a group are spread out from the average (mean or expected value). A low standard deviation means that most of the numbers are close to the average, while a high standard deviation means that the numbers are more spread out. Note that the sum of the deviations of the individual observations of a sample about the sample mean is always 0.

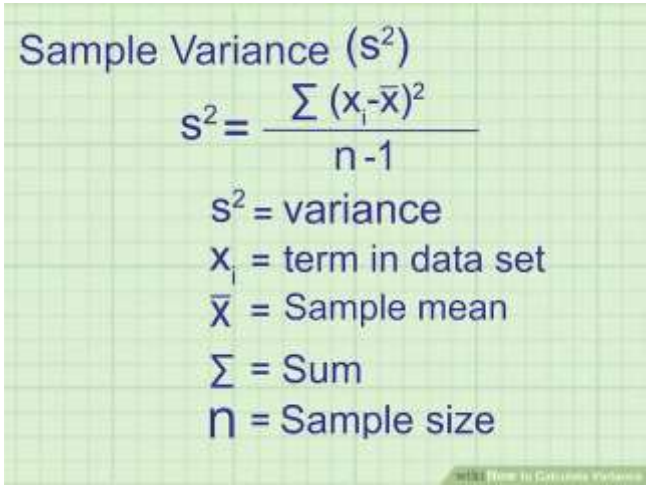
It is defined as a positive square root of variance
There are two standard deviation; Sample standard deviation and Population standard deviation.

Standard deviation = $\sqrt{}$variance

Variance = standard deviation^2

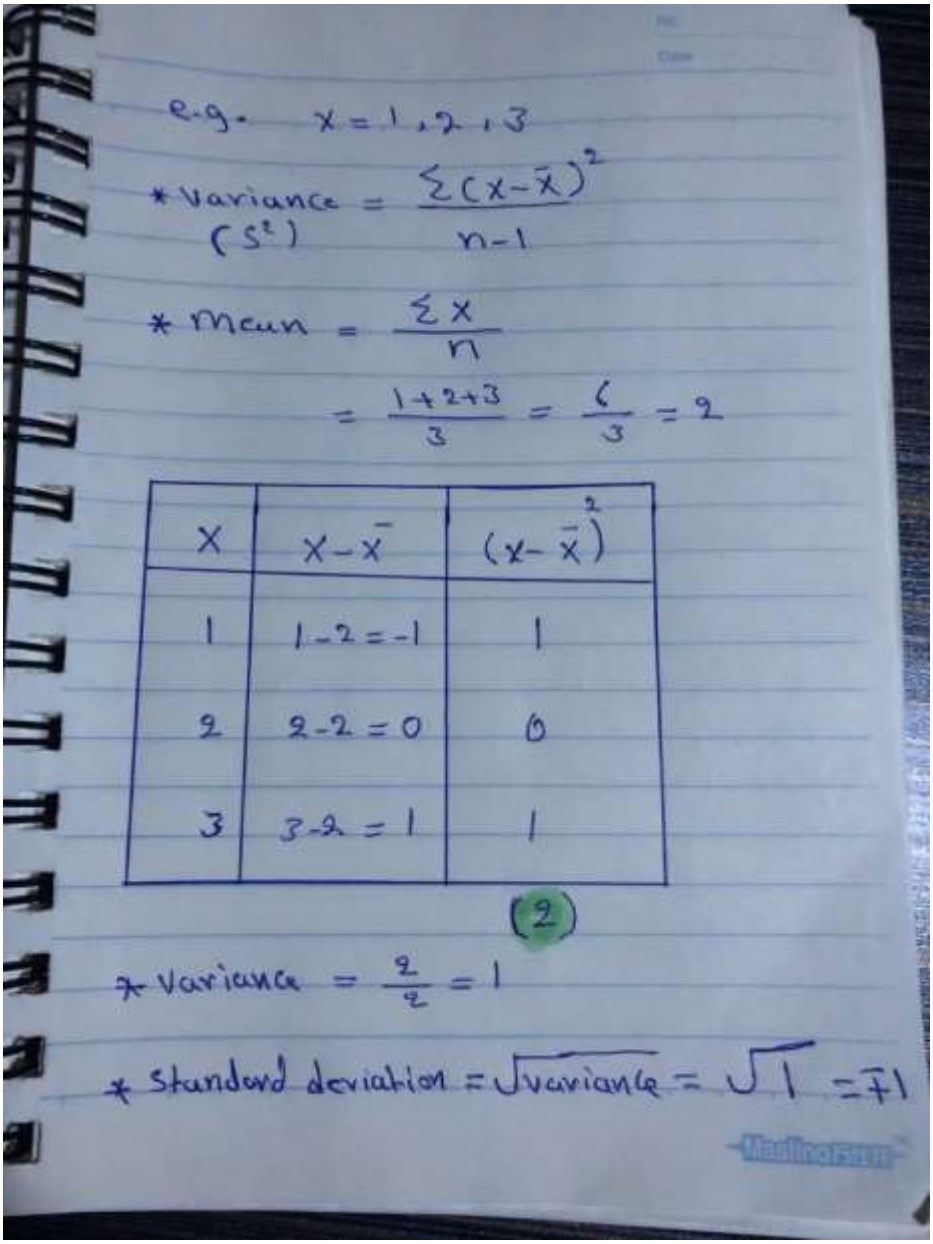


Standard deviation measures how far results spread from the average value. You can find the standard deviation by finding the square root of the variance.



In order to determine standard deviation:

1. Determine the mean (the average of all the numbers) by adding up all the data pieces (**xi**) and dividing by the number of pieces of data (**n**).
2. Subtract the mean (**$\bar{x}$**) from each value.
3. Square each of those differences.
4. Determine the average of the squared numbers calculated to find the variance. (In sample sizes, subtract 1 from the total number of values when finding the average.)
5. Find the square root of the variance. That's the standard deviation.

6. 
e.g. x = 1, 2, 3
* Variance (S²) = Σ(x − x̄)² / n − 1
* mean = Σx / n
= 1+2+3 / 3 = 6 / 3 = 2
x | x − x̄ | (x − x̄)²
1 | 1 − 2 = −1 | 1
2 | 2 − 2 = 0 | 0
3 | 3 − 2 = 1 | 1
(2)
* Variance = 2 / 2 = 1
* Standard deviation = √variance = √1 = ∓1