**Central tendency and** **dispersion**

**Central tendency (or measure of central tendency):**

is a central or typical value for a probability distribution, authors use central tendency to denote "the tendency of quantitative data to cluster around some central value." Quantitative data can be described by measures of central tendency, dispersion, and "shape".

 Central tendency is described by median, mode, and the means (there are different means- geometric and arithmetic).

the data can be subject to the operations of arithmetic. In particular, we can add or subtract observation values, we can sort them and rank them from lowest to highest.

The following may be applied to one-dimensional data

**Arithmetic mean or simply, mean:-**

the sum of all measurements divided by the number of observations in the data set.

**Median**

the middle value that separates the higher half from the lower half of the data set. The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.

**Mode**

the most frequent value in the data set. This is the only central tendency measure that can be used with nominal data, which have purely qualitative category assignments.

**Geometric mean**

the nth root of the product of the data values, where there are n of these. This measure is valid only for data that are measured absolutely on a strictly positive scale.

**Harmonic mean**

the reciprocal of the arithmetic mean of the reciprocals of the data values. This measure too is valid only for data that are measured absolutely on a strictly positive scale.

**Weighted arithmetic mean**

an arithmetic mean that incorporates weighting to certain data elements.

**Truncated mean or trimmed mean**

the arithmetic mean of data values after a certain number or proportion of the highest and lowest data values have been discarded.

**Interquartile mean**

a truncated mean based on data within the interquartile range.

**Midrange**

the arithmetic mean of the maximum and minimum values of a data set.

**Midhinge**

the arithmetic mean of the first and third quartiles.

**Trimean**

the weighted arithmetic mean of the median and two quartiles.

**Winsorized mean**

an arithmetic mean in which extreme values are replaced by values closer to the median.

Any of the above may be applied to each dimension of multi-dimensional data, but the results may not be invariant to rotations of the multi-dimensional space. In addition, there are the Geometric median, Quadratic mean (often known as the root mean square) ,Simplicial depth and Tukey median.

**Measure of central tendency:-**

This is the average of a population - allowing the population to be represented by a single value.

**Median:** is the middle number in a data set that is ordered from least to greatest.

-median : take the middle value for x1, x2, . . . , xn after the data has been sorted from smallest to largest,

 x(1), x(2), . . . , x(n)

**Example**

We asked 10 people to report the number of books they each read last month.

They reported the following:

5, 3, 4, 3, 0, 1, 4, 4, 8, 0.

**Step 1:** list all values in ascending order.

0, 0, 1, 3, 3, 4, 4, 4, 5, 8.

**Step 2:** Do I have an odd or even total number of reported values?

*Even total*: There are two middle values. The median is the average between those two values. (3,4)

In the number of books example, the median is: Median=3+4/2=3.5 Median=3+4/2=3.5

*Odd total:* There is only one middle value. The median is this middle value.

**Mean:-**

Arithmetic mean is the average of a set of numerical values,

- Means For a collection of numeric data, x1, x2, . . . , xn, the sample mean is the numerical average x¯ = 1 n (x1 + x2 + . . . + xn) = 1 n / ∑ X

Easy way to calculate the example above : write the following table:

Values in order: 0, 0, 1, 3, 3, 4, 4, 4, 5, 8.

| **X** | **f** | **fx** |
| --- | --- | --- |
| 0 | 2 | 0 |
| 1 | 1 | 1 |
| 3 | 2 | 6 |
| 4 | 3 | 12 |
| 5 | 1 | 5 |
| 8 | 1 | 8 |
| ∑x=21 | ∑f=10 | ∑fx=32 |

Easier way to calculate the mean

¯¯¯x= ∑xi x ∑fi =32/10=3.2

Arithmetic mean = sum of measures of values / total number of values.

**Mode** is the number that occurs most often in a data set.

Or is the value with highest frequency (highest f).

The reason for the difference in the number of repeated values, so the data is classified according to the type of mode into three types, which are:

 null data: meaning that there is no value that is repeated more than others. Uniform data: where there is only one value that is more frequent than another. Multi-mode data: where there are two or more values in the data that are repeated.

How about the following:

| **x** | **f** |
| --- | --- |
| 0 | 2 |
| 1 | 3 |
| 3 | 3 |
| 4 | 3 |
| 5 | 1 |
| 8 | 1 |

| **x** | **f** |
| --- | --- |
| 0 | 1 |
| 1 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 8 | 1 |