DMDW

Assignment

BY

Abrar Ali

BBA(IT)III

121418408020

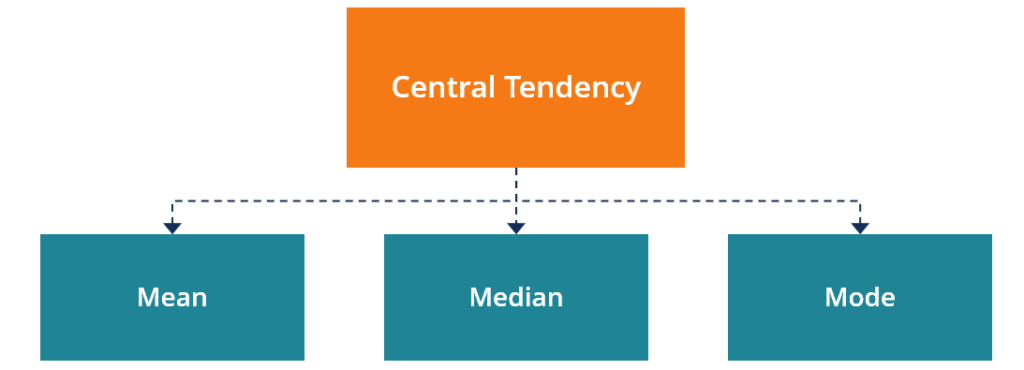
**STATISTICAL DESCRIPTION OF DATA:**

**MEASURING OF CENTRAL TENDENCY**:

CENTRAL TENDENCY:

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency but there are others, such as the median and the mode.

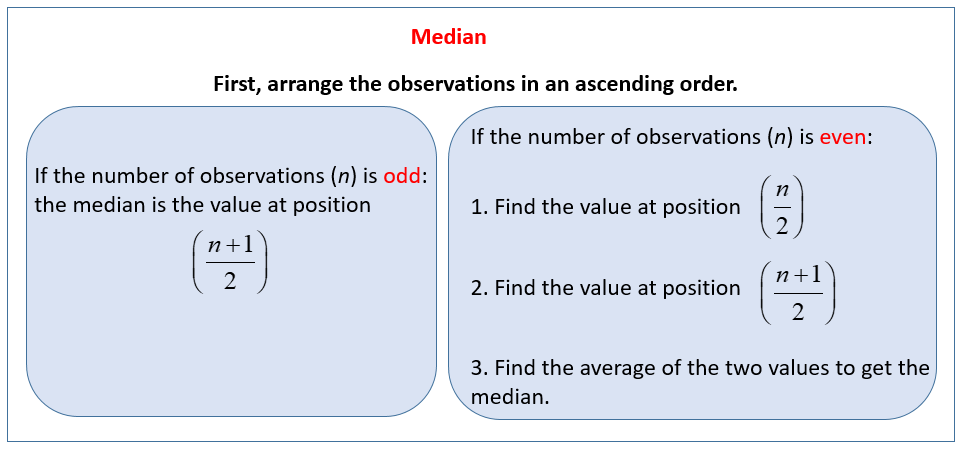
The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others.



The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have n values in a data set and they have values x1, x2, ..., xn, the sample mean, usually denoted by IMG_256 (pronounced x bar), is:

IMG_257

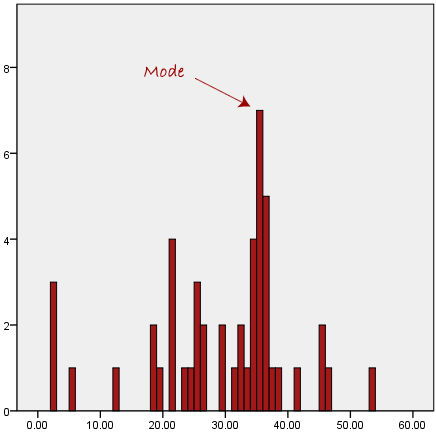
Median: The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data



Mode: Defines the most frequently occurring value in a dataset. In some cases, a dataset may contain multiple modes while some datasets may not have any mode at all.

On a histogram it represents the highest bar in a bar chart or histogram

Example:



Measures of Dispersion

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable is.

Types of Measures of Dispersion

There are two main types of dispersion methods in statistics which are:

* Absolute Measure of Dispersion
* Relative Measure of Dispersion

Absolute Measure of Dispersion

An absolute measure of dispersion contains the same unit as the original data set. Absolute dispersion method expresses the variations in terms of the average of deviations of observations like standard or means deviations. It includes range, [standard deviation](https://byjus.com/maths/standard-deviation/), quartile deviation, etc.

The types of absolute measures of dispersion are:

1. **Range:** It is simply the difference between the maximum value and the minimum value given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6
2. **Variance:** Deduct the mean from each data in the set then squaring each of them and adding each square and finally dividing them by the total no of values in the data set is the variance. Variance (σ2)=∑(X−μ)2/N
3. **Standard Deviation:** The square root of the variance is known as the standard deviation i.e. S.D. = √σ.
4. **Quartiles and Quartile Deviation:**The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.
5. **Mean and Mean Deviation:** The average of numbers is known as the mean and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is known as the mean deviation (also called mean absolute deviation).

Relative Measure of Dispersion

The relative measures of dispersion are used to compare the distribution of two or more data sets. This measure compares values without units. Common relative dispersion methods include:

1. Co-efficient of Range
2. Co-efficient of Variation
3. Co-efficient of Standard Deviation
4. Co-efficient of Quartile Deviation
5. Co-efficient of Mean Deviation

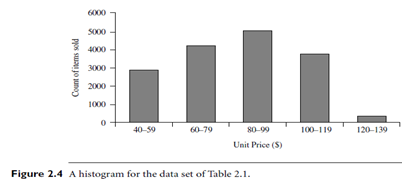
|  |  |
| --- | --- |
| Range | C.D. = (Xmax – Xmin) ⁄ (Xmax + Xmin) |
| Quartile Deviation | C.D. = (Q3 – Q1) ⁄ (Q3 + Q1) |
| Standard Deviation (S.D.) | C.D. = S.D. ⁄ Mean |
| Mean Deviation | C.D. = Mean deviation/Average |

**Graphic Displays of Basic Descriptive Data Summaries**

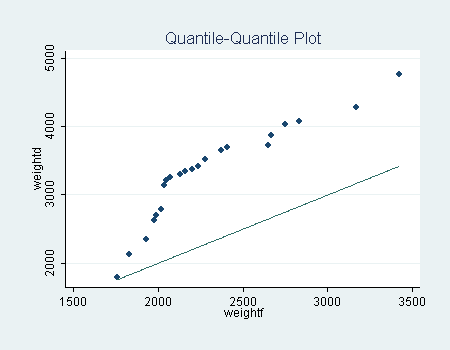
Aside from the bar charts, pie charts, and line graphs used in most statistical or graphical data presentation software packages, there are other popular types of graphs for the display of data summaries and distributions. These include histograms, quarantile plots, q-q plots, scatter plots, and loess curves. Such graphs are very helpful for the visual inspection of your data.

Plotting histograms, or frequency histograms, is a graphical method for summarizing the distribution of a given attribute. A histogram for an attribute A partitions the data distribution of A into disjoint subsets, or buckets. Typically, the width of each bucket is uniform. Each bucket is represented by a rectangle whose height is equal to the count or relative frequency of the values at the bucket.

Histogram plot:



A **quantile** plot is a simple and effective way to have a first look at a univariate data distribution. First, it displays all of the data for the given attribute allowing the user to assess both the overall behavior and unusual occurrences. Second, it plots quantile information.



A **scatter plot** is one of the most effective graphical methods for determining if there appears to be a relationship, pattern, or trend between two numerical attributes. To construct a scatter plot, each pair of values is treated as a pair of coordinates in an algebraic sense and plotted as points in the plane.

IMG_256