Mean Median mode can be negative, Standard Deviation is always positive

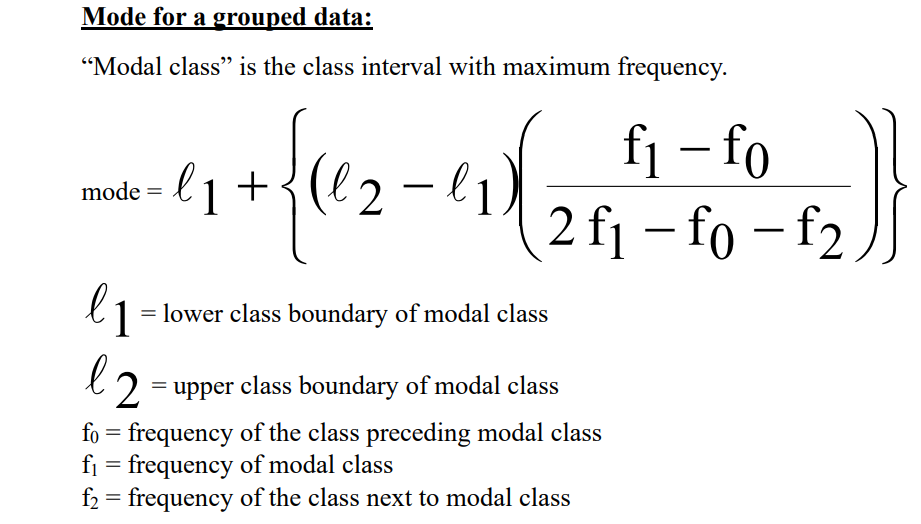
Page no. 60

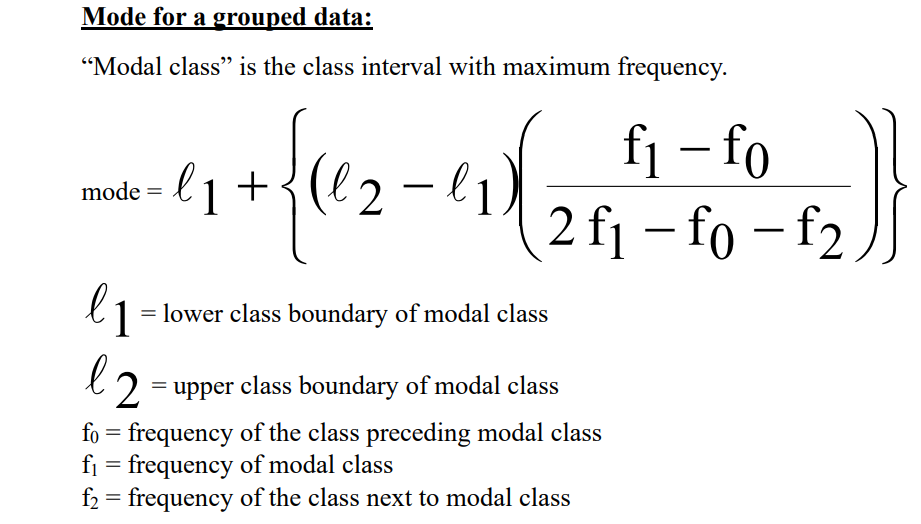
**Mode:** Greatest Frequency

The value which occurs most frequently in the dataset.

**Note**:

1. Sometimes Mode may not exist ( all values are unique )
2. If exists , it may not be single unique (single value)





Page no. 65

Mode = 3median – 2 mode

<https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/nominal-ordinal-interval-ratio/>

**Nominal Ordinal Interval Ratio:**

"Nominal," "ordinal," "interval," and "ratio" are the four levels of measurement in statistics.

Each level provides different information about the data and determines what kind of statistical analysis is appropriate to do on it.

### **1. Nominal** ( It has only categories ) ( No comparison between values )

* **Definition:** Data is categorized without any order or ranking.
* **Characteristics:** Labels or names only.
* **Examples:**
  + **Eye color**: blue, green, brown
  + **Gender**: male, female, non-binary
  + **Marital status**: single, married, divorced
* **Key point:** You **can't** say one category is "**greater**" or "**less than**" another.

**2. Ordinal** ( comparison between data )

* **Definition:** Data is categorized **with a meaningful order**, but the intervals between values are not equal or known.
* **Characteristics:** Order matters, but not the difference between values.
* **Examples:**
  + Education level: high school < college < graduate degree
  + Satisfaction rating: dissatisfied < neutral < satisfied
  + Military ranks
* **Key point:** You can rank the data, but **can't** measure exact differences between ranks.

**3. Interval** ( No absolute zero , Ratios are meaningless)

* **Definition:** Data is ordered, and the **difference between values is meaningful**, but there is **no true zero point**.
* **Characteristics:** Equal intervals; **zero is arbitrary**
* **Examples:**
  + Temperature in Celsius or Fahrenheit
  + IQ scores
  + Calendar years (e.g., 1990, 2000)
* **Key point:** You can add/subtract values, but **ratios don’t make sense** (e.g., 20°C is not “twice as hot” as 10°C).

**4. Ratio** (Absolute Zero present, Ratios have meaning)

* **Definition:** Data has all the properties of interval data, **plus a true zero point**, allowing for meaningful ratios.
* **Characteristics:** Ordered, equal intervals, and **absolute zero**
* **Examples:**
  + Height, weight, age
  + Distance
  + Income
  + Time duration
* **Key point:** You can compare using **multiplication/division** (e.g., 10 kg is twice as heavy as 5 kg).

Page no.70

**Measures of Dispersion : ( How spread data is )**

(1) Comparative study :

**Smaller** the **magnitude** (value) of dispersion, **higher** is the **consistency** or **uniformity** and vice-versa.

(2) Reliability of an average :

Small dispersion means less variation between values and averages. It means average is reliable and a good representation of the data.

(3) Control the variability :

Different measures of dispersion provide us data of variability from different angles,

and this knowledge can prove helpful in controlling the variation.

Especially in the financial analysis of business and medicine, these measures of

dispersion can prove very useful.

(4) Basis for further statistical analysis :  
 • Measures of dispersion provide the basis for further statistical analysis like computing

correlation, regression, test of hypothesis, etc.

**Range :** If L is the largest observation in the data and S is the smallest observation, then

range is the difference between L and S. Thus,

Range = L-S 🡨It is an absolute measure

L 🡨 Largest value of class

S 🡨 Smallest value of class

Ex1. 1,2,3,4,5,6 🡪 Range = 6-1 = 5

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | 20 | 30 | 40 |
| 2 | 3 | 4 | 5 |

Ex2.

Range = 40-10 = 30

Coefficient of range = 𝐿 − 𝑆 / L + S

**NOTE :**  greater Range = greater Variation = Less consistency = Less reliability

**Quartile Deviation**

We have seen earlier that range, as a measure of dispersion, is based only on two extreme

values and fails to take into account the scatter of remaining observations within the range.

To overcome this drawback to an extent, we use another measure of dispersion called InterQuartile Range. It represents the range which includes middle 50% of the distribution. Hence,

**Inter-Quartile Range = Q3 – Q1**

where, Q3 and Q1 represent upper and lower quartiles respectively.

**Half of Inter-Quartile-Range** i.e. **Semi-Inter-Quartile Range = (𝑄3 − 𝑄1) / 2**

is also used as absolute measure of dispersion. The semi-inter-quartile range is popularly known as

Quartile Deviation (Q.D.)

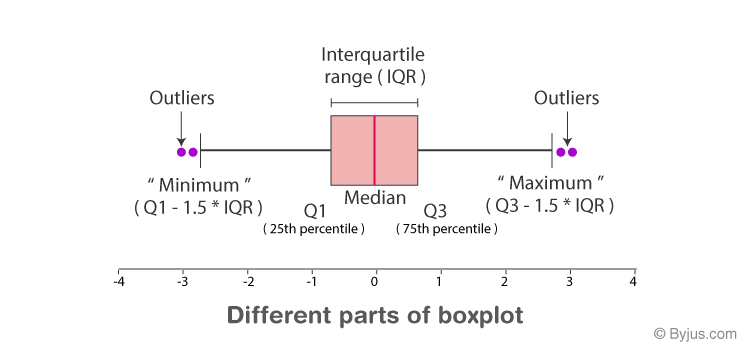
𝑸𝑫 = (𝑸𝟑 – 𝑸𝟏) / 𝟐

The corresponding relative measure of dispersion is called coefficient of quartile deviation

and is defined as **Coefficient of Q.D = 𝑄3 − 𝑄1 / 𝑄3 + 𝑄1**

**BoxPlot :** Box and whisker plot

To check symmetry of data and to check outliers presence



If red line is in center, then data is symmetric.

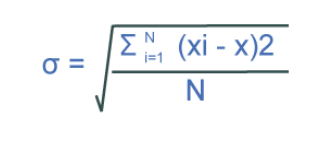
If it is not is center data is skewed.

Page no.73

**Standard Deviation :**

**SD =** sqrt(**Variance**)

Variance is measure of dispersion ( widely used )



SD is +ve square root of arithmetic mean of square of deviations of observations from their arithmetic mean

Observations 🡪 xi

1st arithmetic mean in sentence 🡪 (summation of something / N)

Deviation of observations = x – something

Their arithmetic mean means observations’ A.M 🡪 x bar

Deviation of observation from their A.M 🡪 xi – x bar

Square of deviations of observations from their arithmetic mean 🡪 (xi - x bar)2

Now consider 1st A.M. in sentence : sum((xi - x bar)2)/ N

Square root : sqrt ( sum (( xi – xbar)2 )/ N )

xbar = mean(x) # sum(x) / length(x)

deviations = sum((x - xbar)^2)

N = length(x)

sd = sqrt(deviations/N)

**Coefficient of Variance :**

Its relative measure corresponding to standard deviation

c.v. = (SD / xbar ) / 100