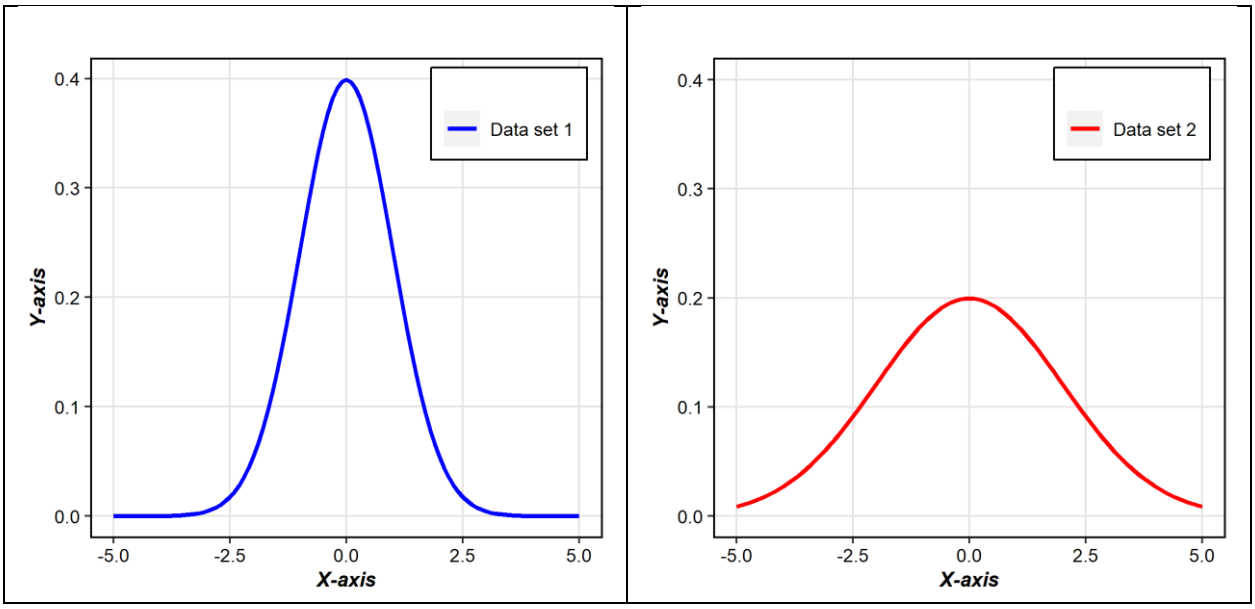
Shape of the distribution

In previous chapter, we learned about "Location" and "Dispersion" which are two important quantitative concepts.



Two data sets may have identical means and identical variance, but their graphical shapes may be different.

And, Mean & Variance fail to describe the shape of the data distribution or graphical shape of a frequency distribution.

Then sir, how do we describe our graphical shape of the distribution?

In this case, we need to learn a term "Moments"/ "Statistical Moments".

Sir, what is moments?

Moments are constant which used to determine some characteristics/properties of frequency distribution.

How many characteristics?

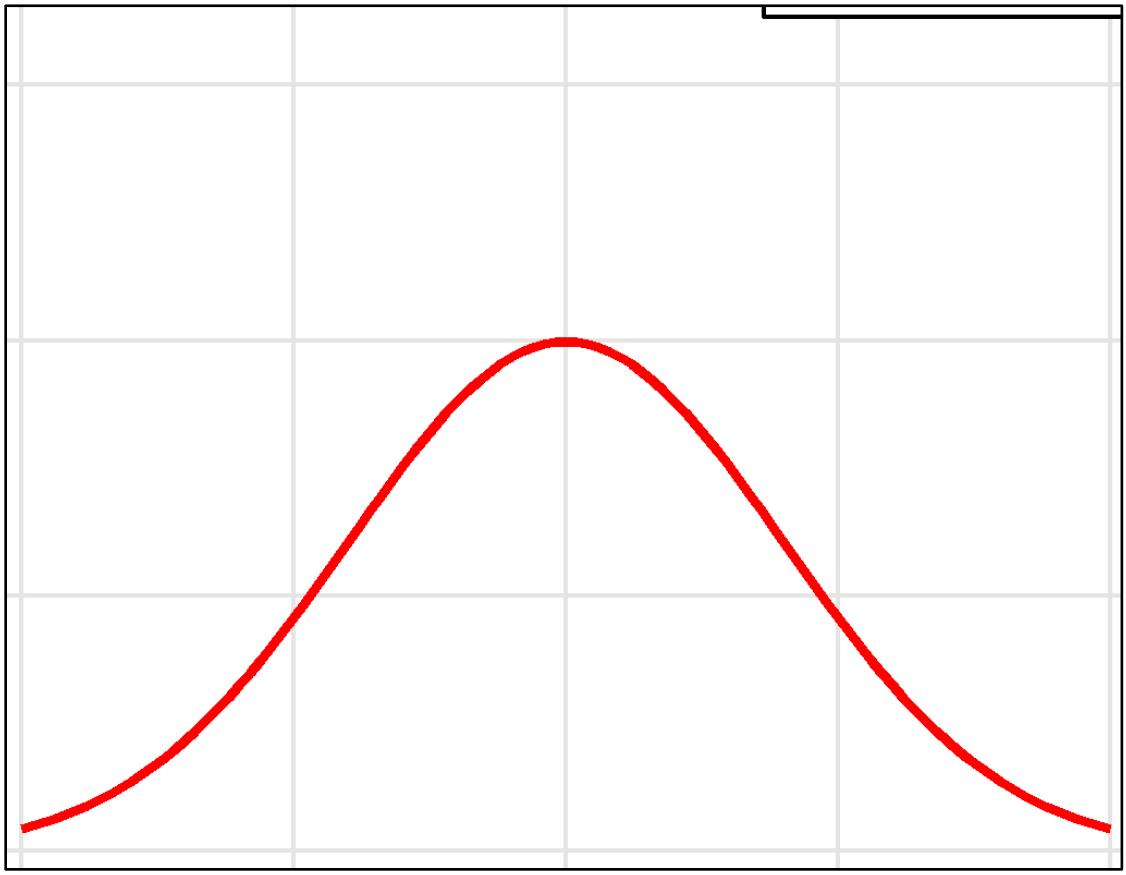
Two characteristics: 1) Skewness and 2) Kurtosis.

And, these characteristics actually called shape of the distribution, our new chapter.

Skewness: Lack of symmetry of a distribution

First questions, “What is symmetry of a distribution”.

Look at this picture.



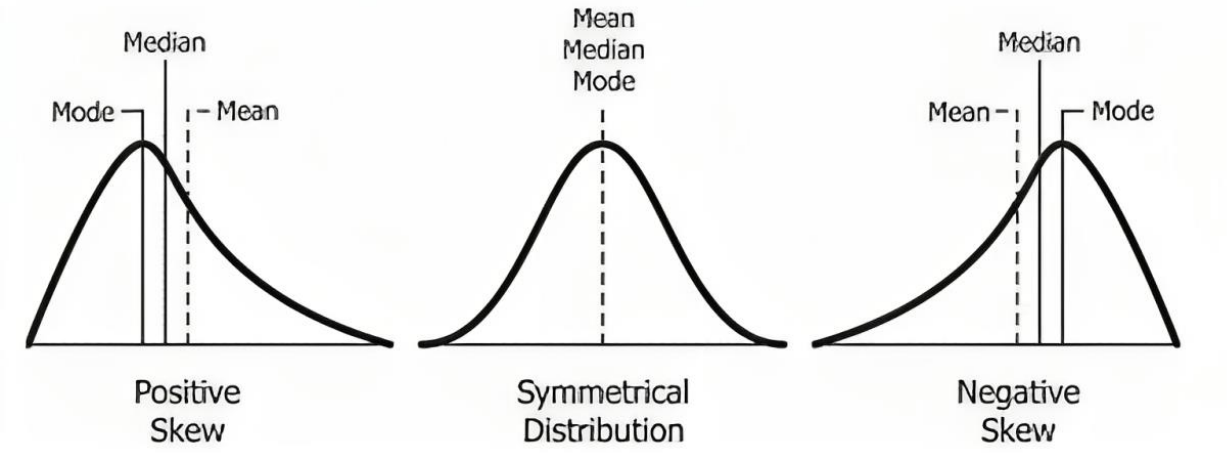
If you can divide this graph like this that both sides have equal percent of data. Suppose, you draw a perpendicular line exactly in middle part of the curve. See, there are two parts and both are equal. If you divide this curve in two equal parts, then it is called “Symmetrical Distribution”.

Then “What is skewness?” Lack of symmetry, right???

That means, if your curve deviate from this symmetrical position, either left side of right side, then it is called lack of symmetry, and this is called skewness.

There are two types of skewness:

1) Positive skewness: If your curve deviate towards Y axis
2) Negative skewness:



Mean>Median>Mode = Positive

Mean<Median<Mode = Negative

Coefficient of Skewness:

$S_k = \frac{3(Mean-Median)}{SD}$

- $S_k > 0$: ***Positively skewed***
- $S_k < 0$: ***Negatively skewed***
- $S_k = 0$: ***Symmetric***

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9

Median = 6

SD = 7.07

$$S_k = \frac{3(Mean - Median)}{SD} = \frac{3(9-6)}{7.07} = 1.27$$

Since $S_k > 0$, thus the distribution is positively skewed distribution.

Kurtosis: Degree of peaked or flatness of a distribution.

Three types:

1) Leptokurtic
2) Mesokurtic/Normal
3) Platykurtic

We can calculate Kurtosis by using $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Mathematical explanation is not so important for mid or quiz, but interpretation is important.

Shape of the distribution

Box & Whisker Plot:

A box-whisker plot, also known as a box plot, is a graphical representation of the distribution of a dataset.

It displays the five-number summary of the data:

Minimum value

First Quartile (Q1)

Second Quartile/Median (Q2)

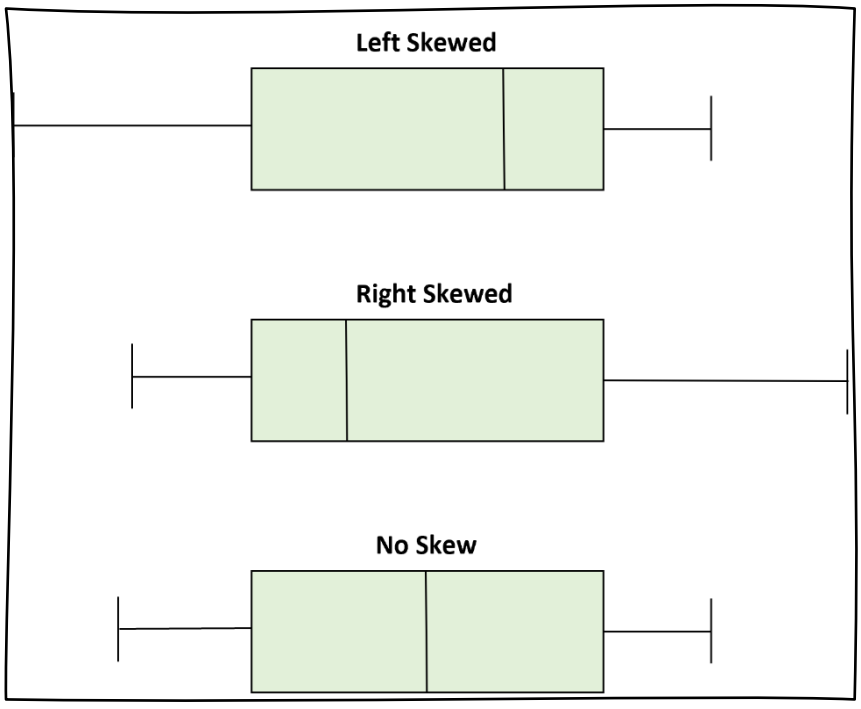
Third Quartile (Q3)

Maximum value

Importance of Box plot:

To get idea of the shape of the distribution

To detect outliers from the data.



IQR
$= Inter\ Quartile\ Range$

$$IQR = Q_3 - Q_1$$

Lower Fences:

$$\{Q_1 - 1.5 \times IQR\}$$

Upper Fences:

$$\{Q_3 + 1.5 \times IQR\}$$

Example:

2900, 2765, 2960, 2890, 2880, 2720, 2930, 2950, 2860, 3060, 3260, 3525

Organize the data into ascending order,

2720, 2765, 2860, 2880, 2890, 2900, 2930, 2950, 2960, 3060, 3260, 3525

Now,

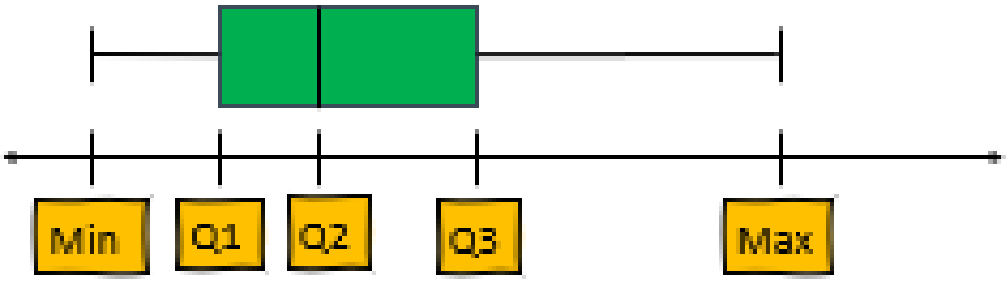
Min=2720

Q1=2870

Q2=2915

Q3=3010

Max=3525



$$\therefore IQR = Q_3 - Q_1 = 14$$

$$Lower = \{Q_1 - 1.5 \times IQR\} = 2660$$

$$Upper = \{Q_3 + 1.5 \times IQR\} = 3220$$