# Beyond Mean & Variance

# Contents

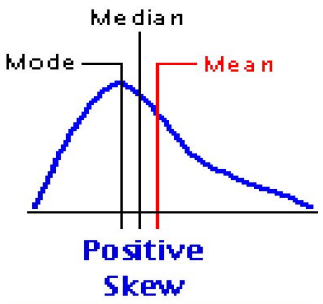
# Skewness

Skewness gives us the Shape of the data. It is the 'Lack of Symmetry'

## Positively Skewed

- Right Tail is longer
- Mass of the distribution is concentrated on the left

Mode<Median<Mean

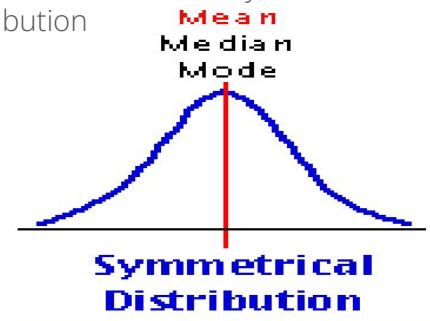



## Symmetric

- Both tails are equal
- Mass of the distribution is equally distributed

Mean=Median=Mode

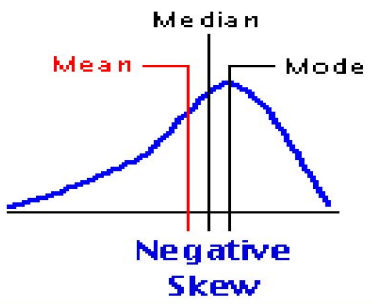
Normal Distribution is symmetric distribution




## Negatively Skewed

- Left Tail is longer
- Mass of the distribution is concentrated on the right

Mean<Median<Mode



# How to Calculate Skewness

$$\frac{(Mean - Mode)}{Standard\ Deviation}$$

Pearson Measure of Skewness

$$\frac{3(Mean - Median)}{Standard\ Deviation}$$

Pearson Measure of Skewness (Alternative Form)

$$\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1}$$

* where $Q_1$=First Quartile, $Q_2$=Second Quartile, $Q_3$= Third Quartile

Bowley's Coefficient of Skewness

# Skewness Based on Third Moment

- The most widely used measure of skewness is based on the third moment.

$$\frac{n}{(n-1)(n-2)}\sum\left(\frac{x_j-\bar{x}}{s}\right)^3$$

- Any threshold or rule of thumb is arbitrary, but here is one: If the skewness is greater than 1.0 (or less than -1.0), the skewness is substantial and the distribution is far from symmetrical. Value ‘zero’ indicates symmetric distribution.

# Kurtosis

Kurtosis is defined as a measure of 'peakedness'. It is generally measured relative to Normal distribution. (Which means 'excess of kurtosis' is measured)

## Mesokurtic

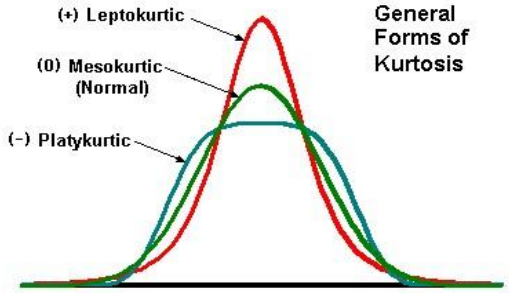
Normal distribution is termed as mesokurtic distribution

## Leptokurtic

A leptokurtic distribution has a more acute peak. (positive kurtosis)

## Platykurtic

A platykurtic distribution has a flatter peak. (negative kurtosis)



# How to Calculate Kurtosis

$$\frac{n(n+1)\sum_{i=1}^{n}(x_i-\bar{x})^4}{(n-1)(n-2)(n-3)s^4}-\frac{3(n-1)^2}{(n-2)(n-3)}$$

Measure of Kurtosis

Value zero indicates peakedness is same as that of  normal distribution.

# Moments

Moments are the constants which help us in knowing the characteristics of population and the graphic shape of a data.

$k^{th}$ raw moment $(\mu_k')$=

$$\sum_{i=1}^{n} \frac{x_i{}^k}{n}$$

The moments about zero are called RAW MOMENTS

$k^{th}$central moment $(\mu_k)$=

$$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^k}{n}$$

The moments about mean are called CENTRAL MOMENTS

# Get an Edge!

- Skewness :
  - If bulk of the data is at the left and the right tail is longer, distribution is skewed right or positively skewed.
  - If bulk of the data is at the right and the left tail is longer, distribution is skewed left or negatively skewed.
  - If skewness is less than −1 or greater than +1, the distribution is highly skewed.
  - If skewness is between −1 and −½ or between +½ and +1, the distribution is moderately skewed.
  - If skewness is between −½ and +½, the distribution is approximately symmetric.
- Kurtosis :
  - A distribution with kurtosis ≈3 (excess ≈0) is called mesokurtic.
  - A distribution with kurtosis <3 (excess kurtosis <0) is called platykurtic.
  - A distribution with kurtosis >3 (excess kurtosis >0) is called leptokurtic.

# Case Study

To learn Descriptive Statistics in R, we shall consider the below case as an example.

**Background**

Data of 100 retailers in platinum segment of the FMCG company.

**Objective**

To describe the variables present in the data

**Sample Size**

Sample size: 100
Variables: Retailer, Zone, Retailer_Age, Perindex, Growth, NPS_Category

# Data Snapshot

Retail Data

**Variables**

| Retailer | Zone | Retailer_Age | Perindex | Growth | NPS_Category |
|---|---|---|---|---|---|
| 1 | North | <=2 | 81.84 | 3.04 | Promoter |

**Observations**

| Columns | Description | Type | Measurement | Possible values |
|---|---|---|---|---|
| Retailer | Retailer ID | numeric | - | - |
| Zone | Location of the retailer | character | East, West, North, South | 4 |
| Retailer_Age | Number of years doing business with the company | character | <=2, 2 to 5, >5 | 3 |
| Perindex | Index of performance based on sales, buying frequency and buying recency | numeric | - | positive values |
| Growth | Annual sales growth | numeric | - | positive values |
| NPS_Category | Category indicating loyalty with the company | character | Detractor, Passive, Promoter | 3 |

# Skewness and Kurtosis in R

```
#Importing Data
retail_data <-read.csv("Retail_Data.csv", header=TRUE)
```

We have already seen that Growth variable is Positively Skewed, so we'll find out skewness & kurtosis value for the same

```
library(e1071)
```

Using package "e1071" in R is the easiest way to find skewness and kurtosis

```
# Skewness
skewness(retail_data$Growth,type = 2)
```

```
[1] 1.591236
```

- **skewness()** gives skewness of the variable.
- **type=**2 uses moment based formula.

```
# Kurtosis
kurtosis(retail_data$Growth,type = 2)
```

```
[1] 4.283886
```

- **kurtosis()** gives kurtosis of the variable.
- **type=**2 uses moment based formula.

```
# Skewness and Kurtosis by Zone
f <- function(x)c(skew = skewness(x,type = 2),kurt = kurtosis(x,type = 2))
aggregate(Growth~Zone,data = retail_data,FUN = f)
```

| | Zone | Growth.skew | Growth.kurt |
|---|---|---|---|
| 1 | East | 1.15293909 | 0.75091158 |
| 2 | North | -0.04046698 | -1.06571086 |
| 3 | South | 2.36833028 | 6.88519638 |
| 4 | West | 0.64121875 | -0.58961827 |

# Application Areas

Questions like :

- What is the shape of my data? Where are my data values concentrated and what is it's spread?

| | |
|---|---|
| Skewness | It is generally used to check how close the data is to normal distribution.<br>- It is used to decide appropriate statistical measure |
| Kurtosis | It is used to see the extent to which the data is concentrated about its mean.<br>- Not commonly reported but very helpful to assess the distribution of variable under study. |

# Quick Recap

In this session, we learnt the basics of knowing the Shape of the Data.

| | |
|---|---|
| Skewness | • Measures of Skewness<br>• Visualizing and Interpreting Skewness |
| Kurtosis | • Measures of Kurtosis<br>• Visualizing and Interpreting Kurtosis |