Descriptive Statistics

Descriptive Statistics

- These are used to describe the samples you are concerned with.
- They are used for – Getting the ‘feel’ of the data ; for use in the statistical tests themselves.
- Example – When you guide a friend coming for the first time to your home about the time it would take to reach from his/her home is a ‘mean’ value of travel time. The more often you have taken the journey earlier, the better would be the estimate.
- Say, you suggest that it would take 50 minutes to reach, give or take 10 minutes either side, traffic permitting. This is an estimate of ‘standard deviation’

Measures of central tendency

- Suppose the lengths of sample western painted turtles from a data set are – 28.5, 18.75, 22.9, 25.4, 25.4, 23.7, 23.9 cms.

- Mean: 24.078
- Median: 23.9
- Mode:25.4

Disadvantages of averages

- Averages do not tell the whole story
- Averages are susceptible to outliers. Outliers skew the averages and pull them in their direction
- Averages make us believe that data points are clustered around a point higher or lower than where they truly cluster
- Averages do not account for segments in data

Measures of dispersion

- Range is the simplest measure of variation. It is the difference between the highest value and the lowest value in the data set.
- Interquartile range describes the middle 50% of the values when ordered from lowest to highest.
- To calculate the IQR, we find the median of lower half and upper half of data. These are quartile 1 and quartile 3. The IQR is the difference between quartile 3 and quartile 1. IQR is resistant to outliers.
- Variance is the summed average squared difference of values from the mean.
- Standard deviation is calculated is the square root of variance.

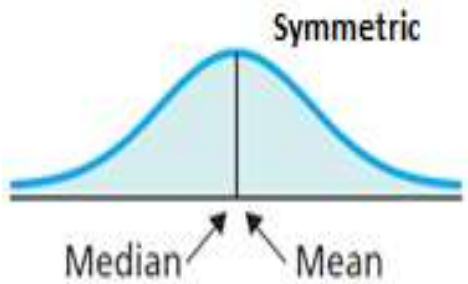
	Heights (cms)	Data value - Mean	(Data value - mean)^2	
	170.18	1.72	2.97	
	170.18	1.72	2.97	
	157.48	-10.98	120.47	
	157.48	-10.98	120.47	
	172.72	4.26	18.18	
	180.34	11.88	141.23	
	167.64	-0.82	0.67	
	165.64	-2.82	7.93	
	157.48	-10.98	120.47	
	185.42	16.96	287.78	
			823.15	**Sum**
Mean	**168.46**			
Range Max value - Min value 185.42 - 157.48	**27.94**			
Variance Sum/10	**82.31**			
Standard Deviation Sqrt (variance)	**9.07**			

	Heights (cms)		
	157.48		
	157.48		
	157.48	**157.48**	**Quartile 1**
	165.64		
	167.64		
	170.18	**168.91**	**Median (Quartile 2)**
	170.18		
	172.72	**172.72**	**Quartile 3**
	180.34		
	185.42		
IQR (Quartile 3 - Quartile 1)	**15.24**		

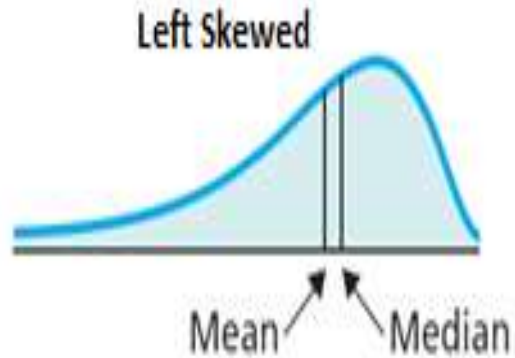
Skewness

- Skewness means lack of symmetry. In statistics, a distribution is called symmetric, if mean, mode and median coincide.
- If the right tail is longer, we get a positively skewed distribution.
- If the left tail is longer, we get a negatively skewed distribution

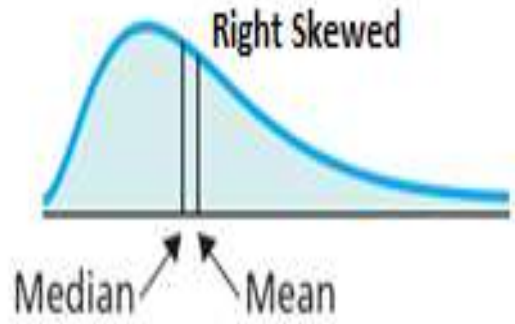
Symmetric and Skewed Distribution



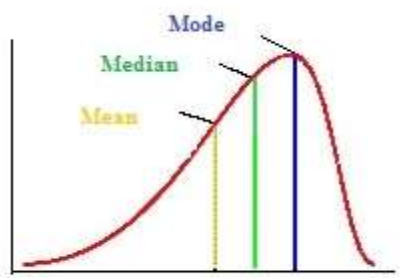
Normal Distribution



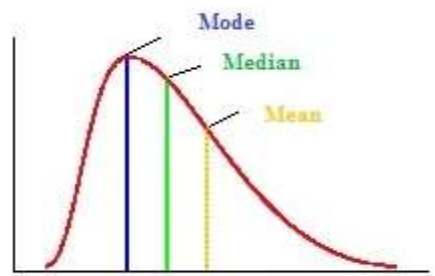
Negatively skewed



Positively skewed



Left-Skewed (Negative Skewness)

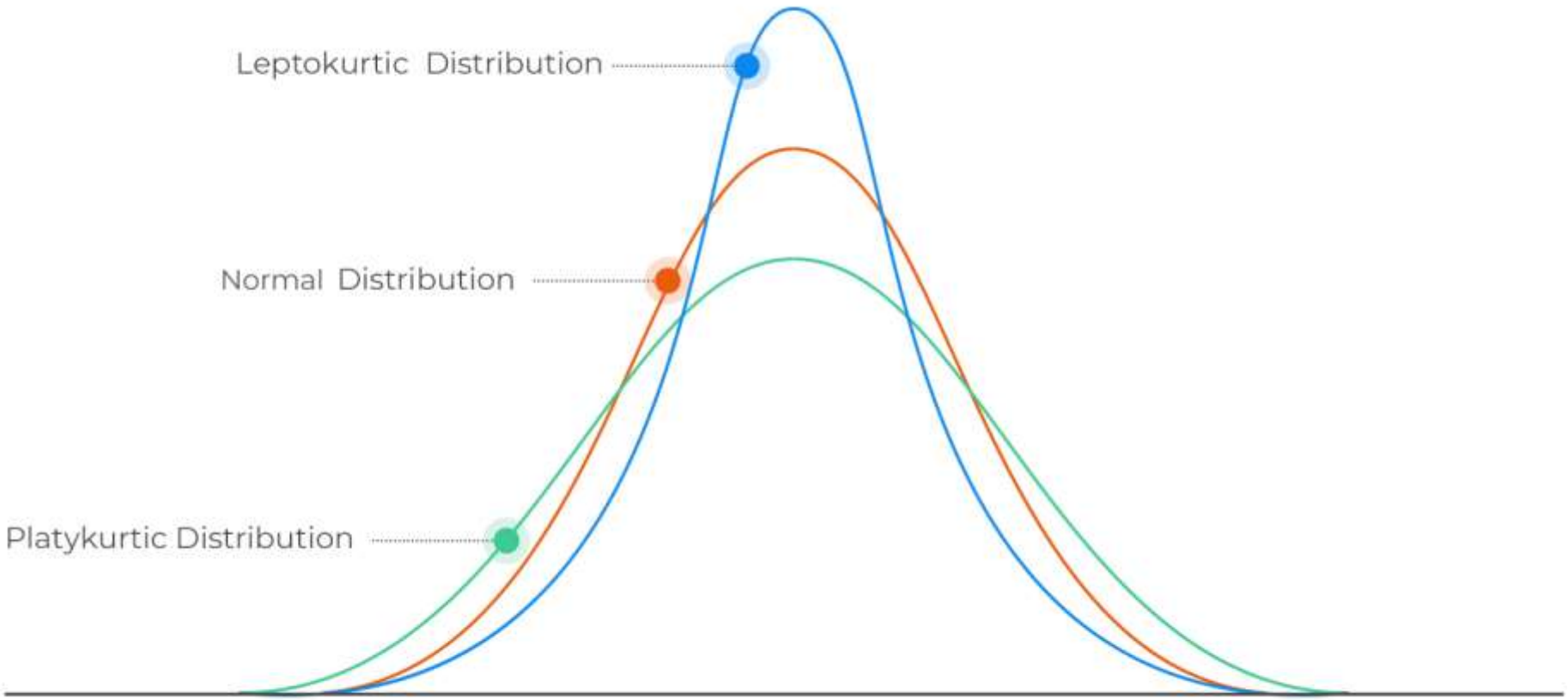


Right-Skewed (Positive Skewness)

Kurtosis

- Epistemology – Kurtos is a Greek word meaning curve or arch
- Kurtosis is a statistical measure that identifies whether the tails of a given distribution contain extreme values.
- Data sets with high kurtosis tends to have heavy tails or outliers
- Data sets with low Kurtosis tends to have light tails or outliers

Kurtosis Terminologies



MS- EXCEL descriptive Statistics toolbox

- https://www.socscistatistics.com/utilities/normaldistribution/default.aspx is an online calculator to produce some normally distributed data
- MS-EXCEL – Data – Data analysis
- Choose Descriptive statistics – Choose input Range, Output Range details
- Excel produces all descriptive statistics

Data	
Mean	-0.26
Standard Error	0.089
Median	-0.25
Mode	0.15
Standard Deviation	0.89
Sample Variance	0.792
Kurtosis	-0.14
Skewness	-0.24
Range	4.59
Minimum	-2.91
Maximum	1.68
Sum	-26
Count	100

Economic Applications

- Financial Markets
- Kurtosis isn't just a theory confined to mathematical textbooks; it has real life applications, especially in the world of economics. Fund managers usually focus on risks and returns, kurtosis (in particular if an investment is lepto- or platy-kurtic). According to stock trader and analyst Michael Harris, a leptokurtic return means that risks are coming from outlier events. This would be a stock for investors willing to take extreme risks. For example, real estate (with a kurt of 8.75) and High Yield US bonds (8.63) are high risk investments while Investment grade US bonds (1.06) and Small cap US stocks (1.08) would be considered safer investments.
- *Ref.: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/kurtosis-leptokurtic-platykurtic/*