Skewness and Kurtosis

Histogram

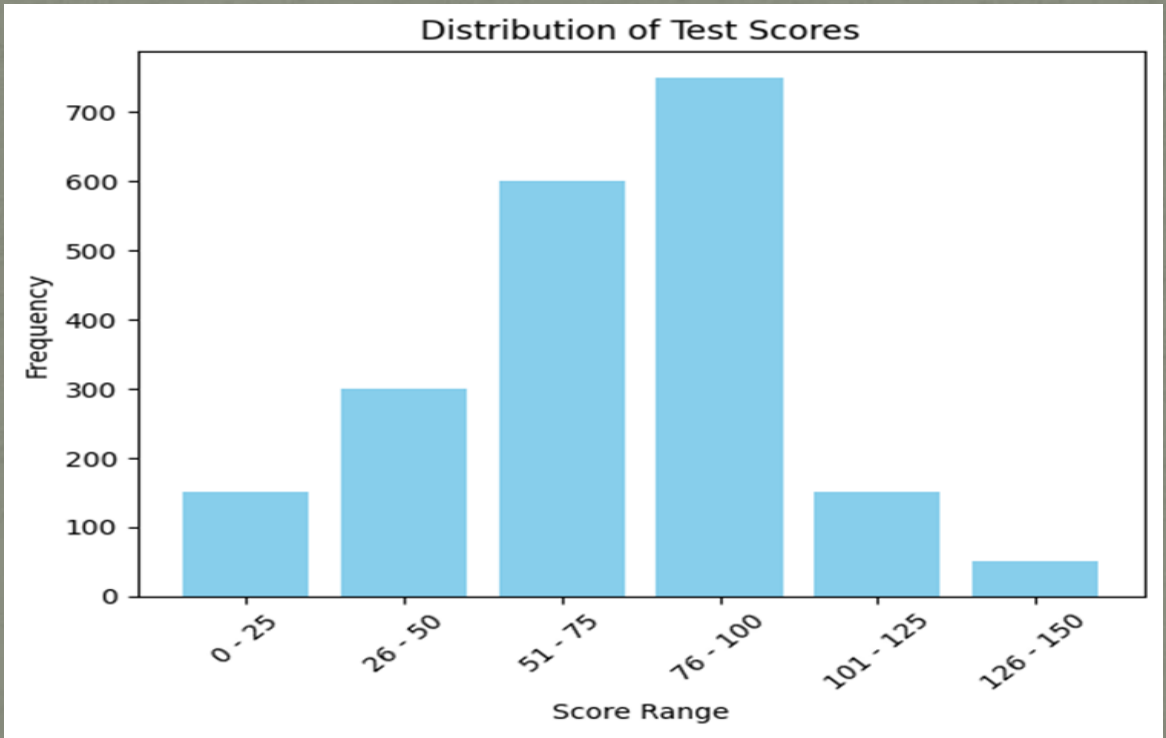
What is Histogram?

- Histograms are graphical representations of data distributions. They consist of bars, each representing the frequency or count of observations falling within specific intervals, known as bins.

- The histogram graphically shows the following:
  - ✓ Frequency of different data points in the dataset.
  - ✓ Location of the centre of data.
  - ✓ The spread of dataset.
  - ✓ Skewness/variance of dataset.
  - ✓ Presence of outliers in the dataset

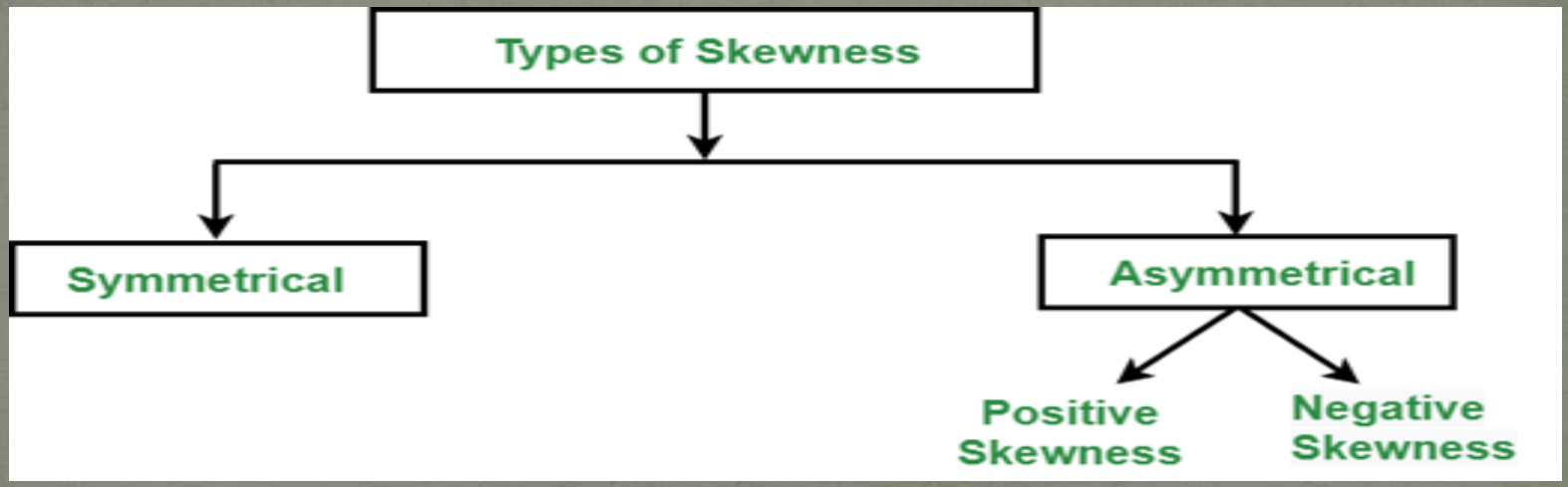
Histogram contains the following axes:

- Vertical Axis: Frequency/count of each bin.
- Horizontal Axis: List of bins/categories.

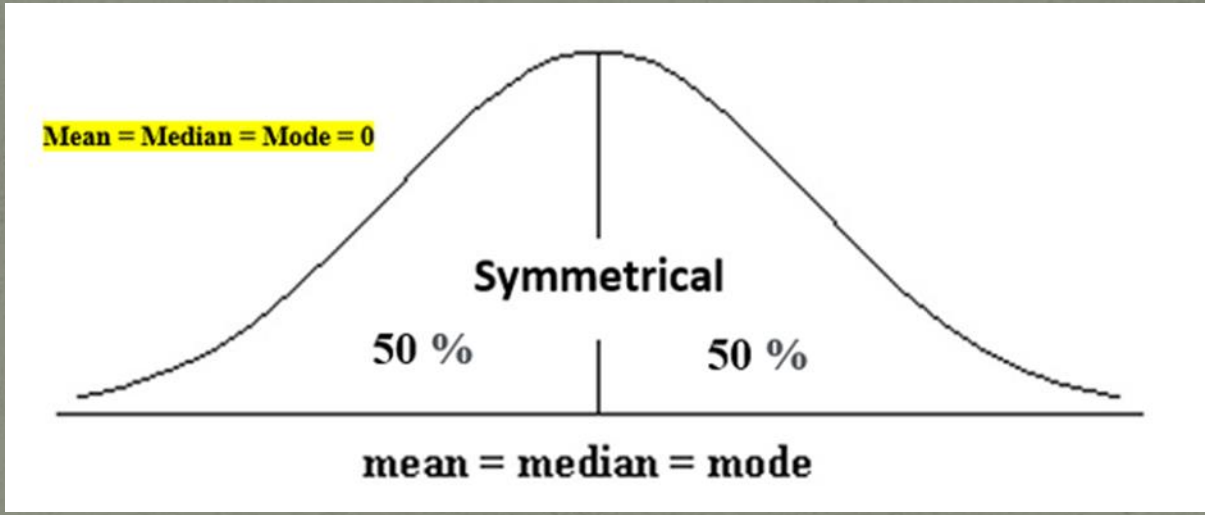


What Is Skewness?

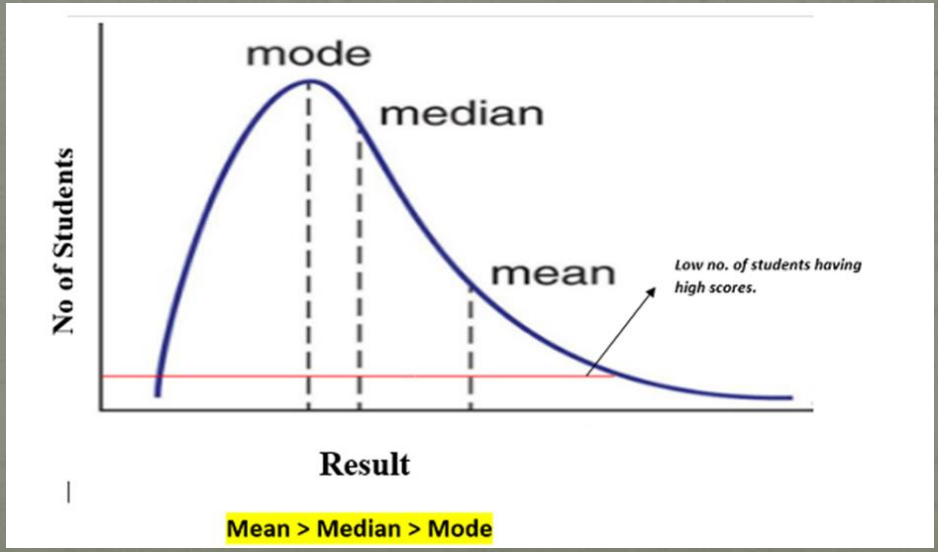
- Skewness is an important statistical technique that helps to determine asymmetrical behaviour than of the frequency distribution, or more precisely, the lack of symmetry of tails both left and right of the frequency curve.
- A distribution or dataset is symmetric if it looks the same to the left and right of the centre point.



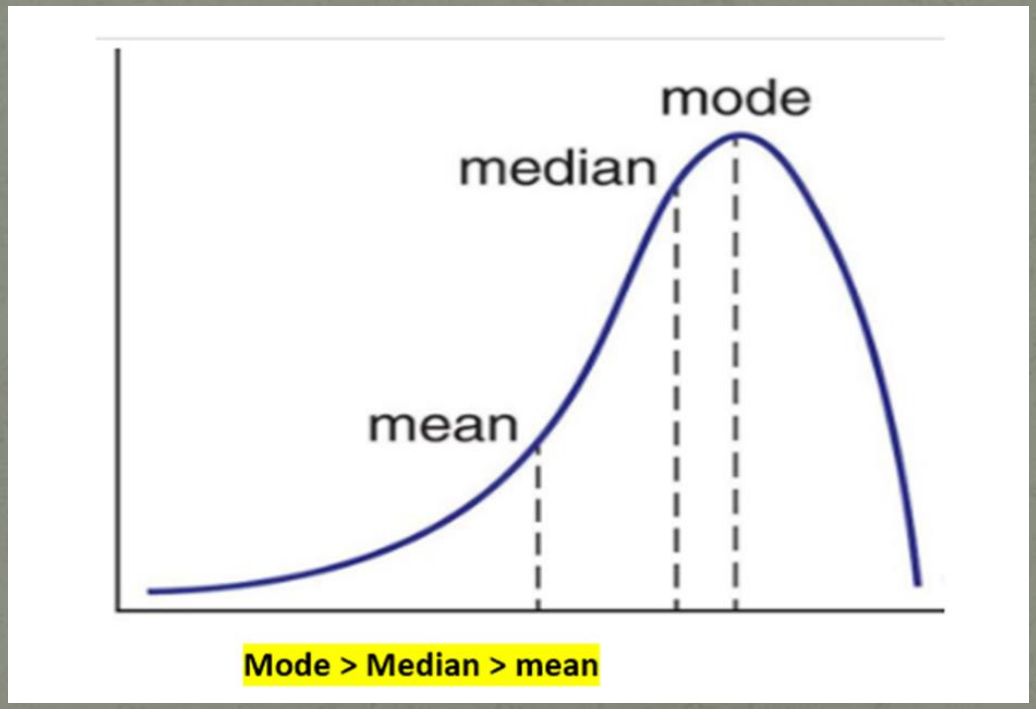
- Symmetric Skewness:
  - ✓ A perfect symmetric distribution is one in which frequency distribution is the same on the sides of the centre point of the frequency curve.
  - ✓ In this, Mean = Median = Mode. There is no skewness in a perfectly symmetrical distribution



- Asymmetric Skewness
  - ✓ A asymmetrical or skewed distribution is one in which the spread of the frequencies is different on both the sides of the centre point or the frequency curve is more stretched towards one side or value of Mean.
  - ✓ Median and Mode falls at different points.
- Positive Skewness: In this, the concentration of frequencies is more towards higher values of the variable i.e. the right tail is longer than the left tail.

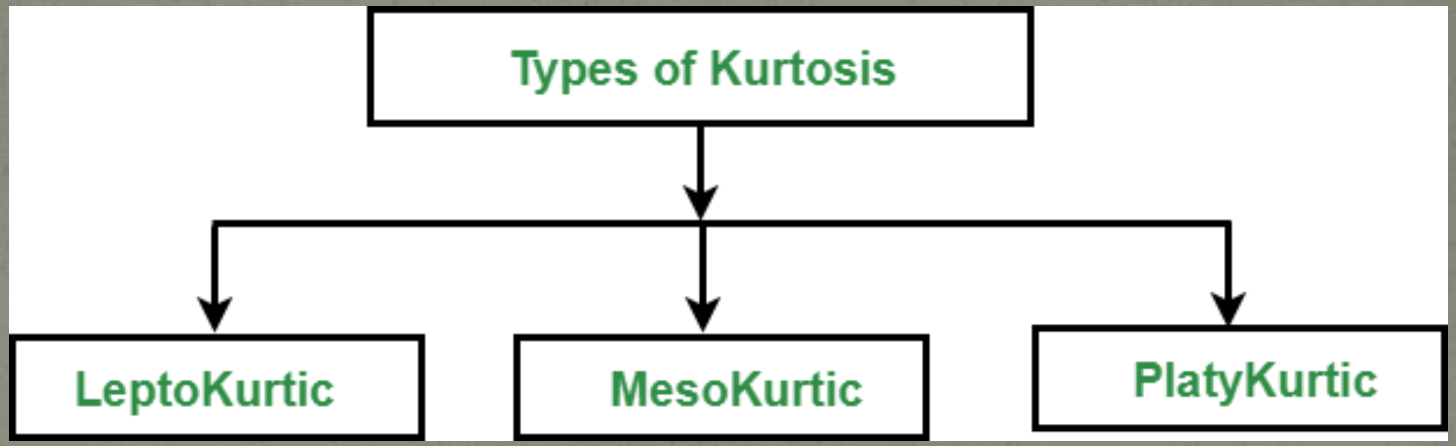


- Negative Skewness: In this, the concentration of frequencies is more towards the lower values of the variable i.e. the left tail is longer than the right tail.

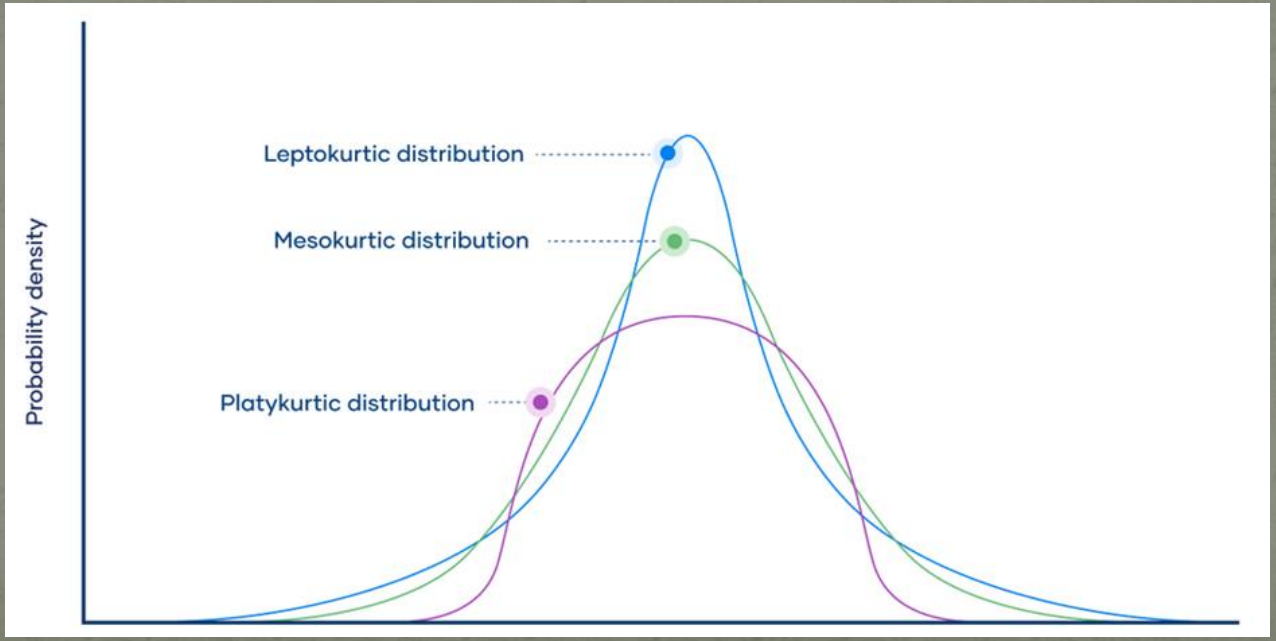


Kurtosis

- It is also a characteristic of the frequency distribution. It gives an idea about the shape of a frequency distribution.
- Basically, the measure of kurtosis is the extent to which a frequency distribution is peaked in comparison with a normal curve.
- It is the degree of peakedness of a distribution.



- Leptokurtic(>3): Leptokurtic is a curve having a high peak than the normal distribution. In this curve, there is too much concentration of items near the central value.
- Mesokurtic(=3): Mesokurtic is a curve having a normal peak than the normal curve. In this curve, there is equal distribution of items around the central value.
- Platykurtic(<3): Platykurtic is a curve having a low peak than the normal curve is called platykurtic. In this curve, there is less concentration of items around the central value.



Report for Skew and Kurtosis

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	108	67.3034	66.3347	66.3586	72.1006	62.2782	277649
Median	108	67	65	66	71	62	265000
Mode	1	62	63	65	60	56.7	300000
Q1:25%	54.5	60.6	60.9	61	60	57.945	240000
Q2:50%	108	67	65	66	71	62	265000
Q3:75%	161.5	75.7	73	72	83.5	66.255	300000
99%	212.86	87	91.129	83.86	97	76.1142	NaN
Q4:100%	215	89.4	91.15	88.5	98	77.89	390000
IQR	107	15.1	12.1	11	23.5	8.31	60000
1.5rule	160.5	22.65	18.15	16.5	35.25	12.465	90000
Lesser	-106	37.95	42.75	44.5	24.75	45.48	150000
Greater	322	98.35	91.15	88.5	118.75	78.72	390000
Min	1	40.89	42.75	50	50	51.21	200000
Max	215	89.4	91.15	88.5	98	77.89	390000
Kurtosis	-1.2	-0.60751	0.0869008	-0.0974897	-1.08858	-0.470723	-0.239837
Skew	0	-0.132649	0.162611	0.204164	0.282308	0.313576	0.8067

Result Set of Placement.csv

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Skew	Normal or Symmetric Skewness	Negative Skewness	Normal or Symmetric Skewness	Normal or Symmetric Skewness	Normal or Symmetric Skewness	Normal or Symmetric Skewness	Normal or Symmetric Skewness
Kurtosis	Platykurtic	Platykurtic	Platykurtic	Platykurtic	Platykurtic	Platykurtic	Platykurtic

Skew -> Except ssc_p rest values are equal to 0 so it is normal or symmetric skewness and ssc_p has -0.13 so it is negative skewness

Kurtosis -> Result of all column values are <3 so it is platykurtic.