# Data science Univariate

Skewness and kurtosis

Skewness

Skewness is a statistical measure that describes the degree of asymmetry of a distribution around its central tendency. It helps to identify whether the distribution leans towards the left (negative skew) or the right (positive skew).

The central tendency is represented by measures such as **mean**, **median**, and **mode**:

- A **positive skew** occurs when the tail of the distribution extends more towards the right, meaning the mean is greater than the median.

- A **negative skew** occurs when the tail extends more towards the left, meaning the mean is less than the median.

- A **symmetric distribution** has a skewness close to zero, where the mean, median, and mode are approximately equal.
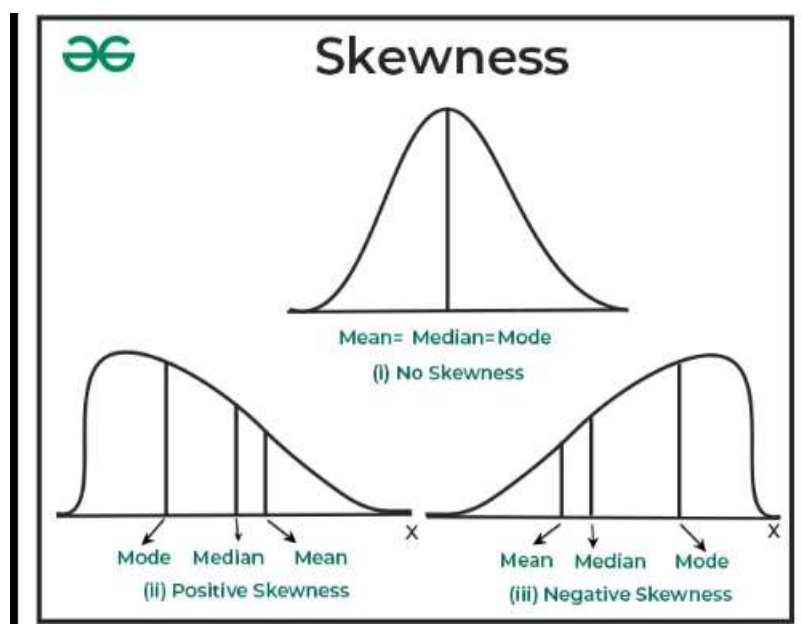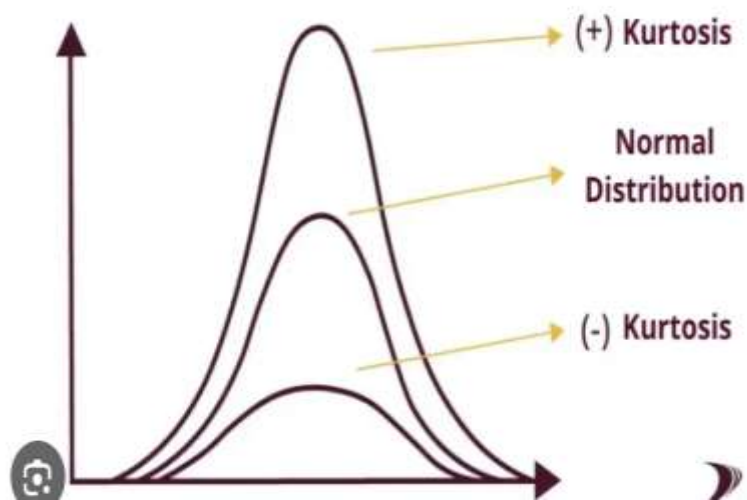
Kurtosis

Kurtosis measures the **tailedness** of the data distribution. It tells us how much of the data is in the tails and how much is concentrated around the mean, compared to a normal distribution.

Types of Kurtosis:

- **Leptokurtic (K > 3)**: Tails are heavier, meaning more extreme outliers.

- **Mesokurtic (K = 3)**: Tails are similar to a normal distribution (e.g., Gaussian distribution).

- **Platykurtic (K < 3)**: Tails are lighter, meaning fewer extreme values than a normal distribution.

Purpose: Kurtosis is used to understand the shape and distribution of data, particularly focusing on outliers.

Placement dataset skew and kurtosis

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
skew	0.0	-0.067359	0.116875	0.11676	0.282308	0.241681	0.528516
kurtosis	-1.2	-0.743673	-0.520352	-0.327652	-1.08858	-0.646232	-0.708724

"Both the skewness and kurtosis values for each individual column in the placement dataset."

```
#skew and kurtosis

dataset[columnName].kurtosis()

-0.708724111851927

dataset[columnName].skew()

0.5285158036669733
```

"Calculate both the skewness and kurtosis values for each individual column in the placement dataset."