Statistics: Skewness, Coefficient of Variation and Kurtosis

Skewness:

- **Definition**:
  - Skewness is a statistical measure that describes the asymmetry of a distribution.
  - The presence of extreme lower or higher values (outliers) in datasets
  - Mainly lower extreme values pull the distributions towards the right-hand side (median > Mean), and the entire distribution looks skewed negatively
  - Higher extreme values pull the distribution towards the left side (mean>median), forming positive skewness.

- **Types of Skewness:**
  - **Positive Skewness (Right-Skewed):**
    - The right tail (larger values) is longer or fatter than the left tail (smaller values).
    - Most of the data points are concentrated on the left side of the distribution.
    - The mean is typically greater than the median.
    - **Example**: Income distribution, where a small number of people earn significantly more than the majority.
  - **Negative Skewness (Left-Skewed):**
    - The left tail (smaller values) is longer or fatter than the right tail (larger values).
    - Most of the data points are concentrated on the right side of the distribution.
    - The mean is typically less than the median.
    - **Example**: Age at retirement, where a few people retire very early, but most retire around a certain age.
  - **Zero Skewness (Symmetrical Distribution)**
    - The tails on both sides of the mean are balanced.
    - The mean and median are approximately equal.
    - Example: Normal distribution (bell curve).

- **Calculation of Skewness**
    - Skewness can be calculated using different formulas, but one common method is the Pearson's moment coefficient of skewness:
    - Skewness= $= \frac{1}{N} * \sum_{i=1}^{N} (\frac{x_i - \bar{x}}{\sigma})^3$

  Where,

    - N is the number of observations.
    - x_i is the i-th value in the data set.
    - $\bar{x}$ is the mean of the data set.
    - σ is the standard deviation of the data set.

- **Interpretation of Skewness Values**
  - **Skewness > 0**: The distribution is positively skewed (right-skewed).

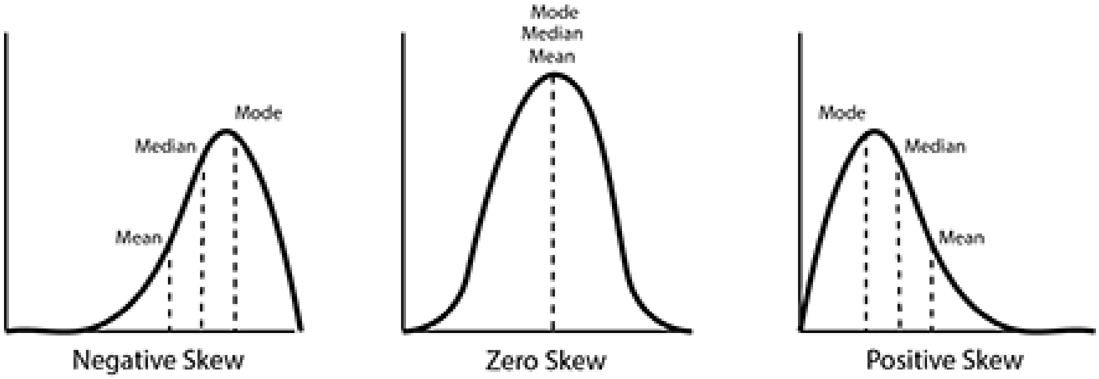
  - **Skewness < 0**: The distribution is negatively skewed (left-skewed).

  - **Skewness = 0**: The distribution is symmetrical.

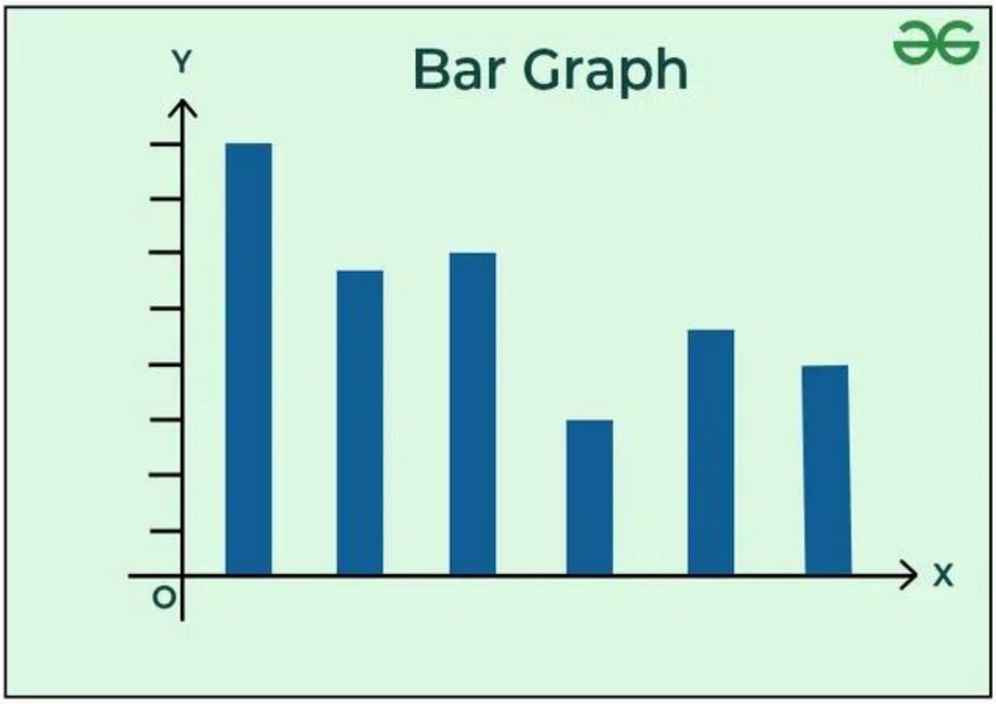
- **Practical Examples and Implications**
  - **Financial Data Analysis**
    - Positive skewness in asset returns indicates that there are frequent small losses and rare large gains.
    - Negative skewness suggests frequent small gains and rare large losses, which might be risky.
  - **Quality Control**
    - In manufacturing, skewness can indicate if the production process is consistently producing parts that are not within the desired specifications.
  - **Healthcare**
    - In epidemiology, positively skewed data may indicate the presence of a rare disease that affects a small portion of the population more severely.

➢ **<u>Visual Representation</u>**

- **Histograms**: By plotting a histogram, one can visually inspect the skewness. A right-skewed distribution will have a tail stretching to the right, while a left-skewed distribution will have a tail stretching to the left.



- **Box Plots**: Box plots can also show skewness. A right-skewed distribution will have a longer whisker on the right side.



- **Adjusting for Skewness**
  - **Transformation**: Data transformations like the logarithmic, square root, or cube root can reduce skewness and make the data more symmetrical.

  - **Robust Statistics**: Instead of relying on mean and standard deviation, using median and interquartile range (IQR) can provide better insights for skewed distributions.

  - Understanding skewness is crucial in statistical analysis as it affects the interpretation of data and the conclusions drawn from statistical tests. Recognizing and adjusting for skewness ensures more accurate and reliable analysis.

Coefficient of Variation (CV)

- **Definition:**
  - The Coefficient of Variation (CV) is a standardized measure of the dispersion of a probability distribution or frequency distribution.
  - It is often used to compare the degree of variation from one data series to another, even if the means are drastically different from each other.
- **Formula**
  - CV= $\frac{\sigma}{\mu}$
  - Where:
  - σ is the standard deviation of the data set.
  - μ is the mean of the data set.

- **Interpretation**
  - **High CV**: Indicates a high level of dispersion relative to the mean.
  - **Low CV**: Indicates a low level of dispersion relative to the mean.

- **Applications**
  - **Finance**: Comparing the risk (standard deviation) relative to the return (mean) of different investments.
  - **Quality Control**: Assessing the variability of production processes relative to the average production level.
  - **Economics**: Comparing the economic stability of different countries by analysing the variation in GDP growth rates.

- **Advantages**
  - Dimensionless: The CV is unitless, making it useful for comparing variability between datasets with different units or scales.
  - Relative Measure: Provides a relative measure of variability, which is more informative than the absolute measure (standard deviation) when comparing datasets with different means.

Kurtosis:

- **Definition**:
  - Kurtosis is a statistical measure that describes the shape of a distribution's tails in relation to its overall shape.
  - It indicates how outlier-prone a distribution is.

- **Types of Kurtosis:**
  - **Mesokurtic**:
    - Kurtosis is close to 0 (often considered to be 3 in some definitions).
    - The distribution has a similar tail shape to the normal distribution.
    - Example: Standard normal distribution.
  - **Leptokurtic**:
    - Positive kurtosis (>0, often >3 in some definitions).
    - The distribution has heavier tails and a sharper peak than the normal distribution.
    - Indicates a higher probability of extreme values.
    - Example: Data with frequent extreme values.
  - **Platykurtic**:
    - Negative kurtosis (<0, often <3 in some definitions).
    - The distribution has lighter tails and a flatter peak than the normal distribution.
    - Indicates a lower probability of extreme values.
    - Example: Data with few extreme values.

- **Formula**
    - Kurtosis $= \frac{1}{N} * \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$

  Where,

    - N is the number of observations.
    - x_i is the i-th value in the data set.
    - $\bar{x}$ is the mean of the data set.
    - σ is the standard deviation of the data set.

- **Interpretation**
  - **Kurtosis > 0** (often >3 in the standard form): Leptokurtic distribution with heavy tails.
  - **Kurtosis < 0** (often <3 in the standard form): Platykurtic distribution with light tails.
  - **Kurtosis = 0** (often =3 in the standard form): Mesokurtic distribution similar to the normal distribution.

- **Applications**
  - **Finance**: Assessing the risk of extreme market movements.
  - **Quality Control**: Identifying production processes with frequent extreme deviations.
  - **Environmental Science**: Analysing the occurrence of extreme weather events.