# Probability and Statistics

Slide set 5

# Skewness

Skewness is a measure of the asymmetry of a probability distribution. It helps to understand the shape of the distribution. There are two main types of skewness: population skewness and sample skewness.

1. **Population Skewness:**
   - The population skewness is a measure of asymmetry for an entire population.
   - The formula for population skewness ($\gamma_1$) is given by:
     $\gamma_1 = \frac{\mu_3}{\sigma^3}$
     where $\mu_3$ is the third central moment and $\sigma$ is the standard deviation.

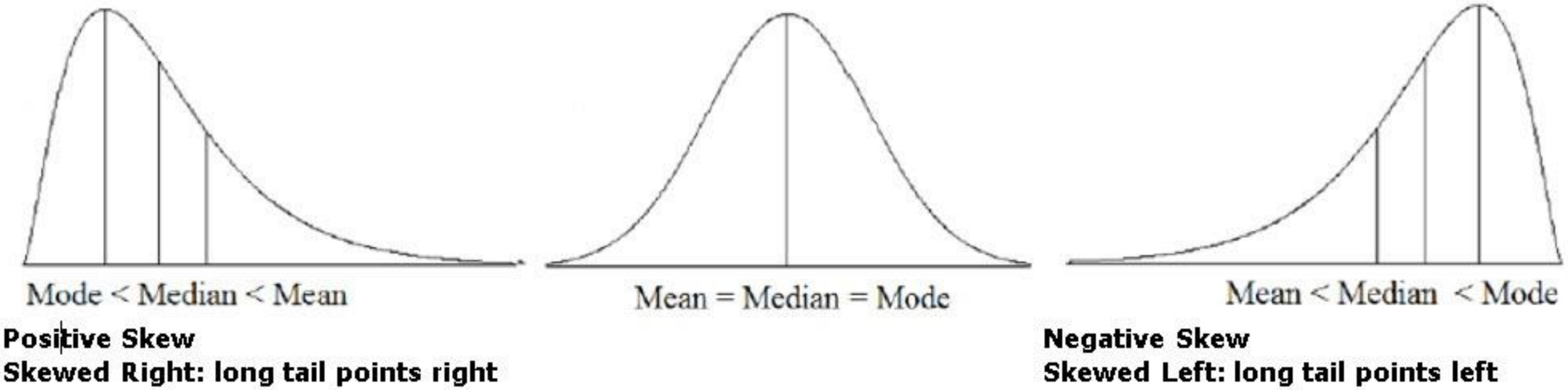
   **Example:**

   Consider a population of exam scores with the following values: 70, 75, 80, 85, 90.
   - Mean ($\mu$) = $\frac{70+75+80+85+90}{5} = 80$
   - Standard Deviation ($\sigma$) ≈ 6.71 (calculated using the population standard deviation formula).
   - Calculate $\mu_3$, the third central moment: $\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^3$
   - If $\mu_3$ is positive, it indicates positive skewness; if negative, it indicates negative skewness.
2. **Sample Skewness:**
   - The sample skewness is an estimate of the population skewness based on a sample from that population.
   - The formula for sample skewness ($G_1$) is given by:
     $G_1 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{s}\right)^3$
     where $n$ is the sample size, $\bar{X}$ is the sample mean, and $s$ is the sample standard deviation.

Interpretation:

- Skewness close to 0 indicates a roughly symmetric distribution.
- Positive skewness suggests a tail on the right side, while negative skewness suggests a tail on the left side.

# Skewness


Mode < Median < Mean
Mean = Median = Mode
Mean < Median < Mode
Positive Skew
Skewed Right: long tail points right
Negative Skew
Skewed Left: long tail points left

# Kurtosis

Kurtosis is a measure of the "tailedness" or the sharpness of the peak of a probability distribution. It describes the shape of the distribution's tails in relation to the tails of a normal distribution.

1. **Population Kurtosis:**
   - The population kurtosis measures the tailedness of the entire population. The formula for population kurtosis ($\beta_2$) is given by:
     $\beta_2 = \frac{\mu_4}{\sigma^4}$
     where $\mu_4$ is the fourth central moment and $\sigma$ is the standard deviation.
   - If $\beta_2 > 3$, it indicates leptokurtic (heavy-tailed) distribution.
   - If $\beta_2 < 3$, it indicates platykurtic (light-tailed) distribution.
   - If $\beta_2 = 3$, it indicates mesokurtic (normal) distribution.
2. **Sample Kurtosis:**
   - The sample kurtosis is an estimate of the population kurtosis based on a sample from that population. The formula for sample kurtosis ($G_2$) is given by:

     $$G_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$
   - Similar to population kurtosis, $G_2 > 3$ indicates leptokurtic, $G_2 < 3$ indicates platykurtic, and $G_2 = 3$ indicates mesokurtic.

Interpretation:

- A distribution with positive kurtosis (leptokurtic) has heavier tails and a sharper peak than a normal distribution.
- A distribution with negative kurtosis (platykurtic) has lighter tails and a flatter peak than a normal distribution.
- Mesokurtic distributions have kurtosis equal to 3, similar to a normal distribution.

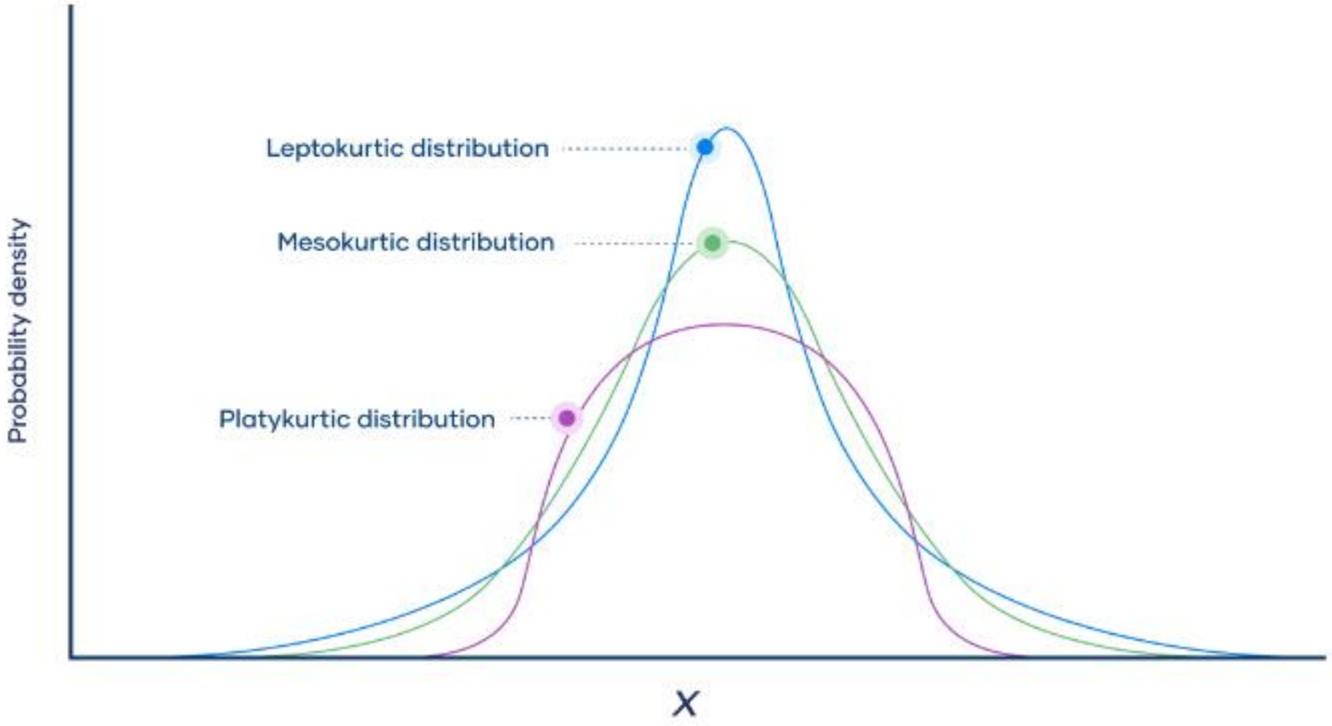
When calculating kurtosis, it's important to note that the formulas involve the fourth central moment, which is the average of the fourth power of the deviations from the mean. Positive values indicate more extreme tails, while negative values indicate less extreme tails compared to a normal distribution.

**Example: Exam Scores**

Consider a set of exam scores for a class:

$\text{Scores: } 70, 75, 80, 85, 90$

We'll calculate both sample and population kurtosis for this dataset.


Leptokurtic distribution
Mesokurtic distribution
Platykurtic distribution
Probability density
X