Statistics

|Kurtosis (التفلطح)

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Agenda

- Skewness (الأنحراف - الالتواء)
- What is Kurtosis (التفلطح)
  - High kurtosis (leptokurtic)
  - Low kurtosis (platykurtic)
  - Normal kurtosis (mesokurtic)
- How to Calculate Kurtosis?
- Why it so important to identify kurtosis?
- Example

Introduction

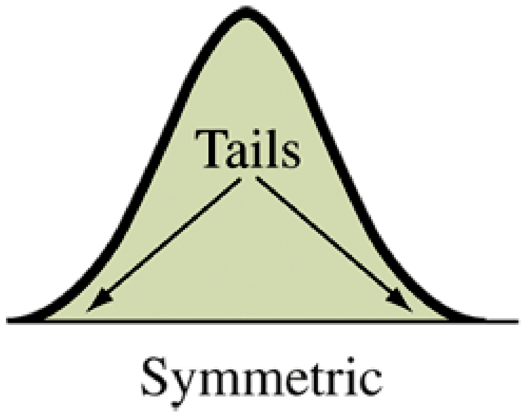
- In statistics, **kurtosis** and **skewness** are both measures of a distribution's shape, but they focus on different aspects.

- They're related in the sense that they help you understand how a dataset **deviates from the normal distribution**, but they describe distinct features.

→ Let's break it down and explore their relationship.

Symmetric Distribution

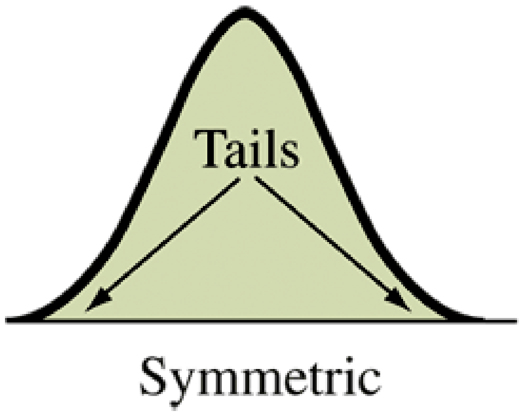
- The left and right sides **mirror** each other.
- Typically, the **mean = median = mode**.



- Example: A perfectly normal (bell-shaped) distribution of standardized **test scores** (under an **ideal scenario**) or measurement errors in a well-controlled experiment.

Symmetric Distribution

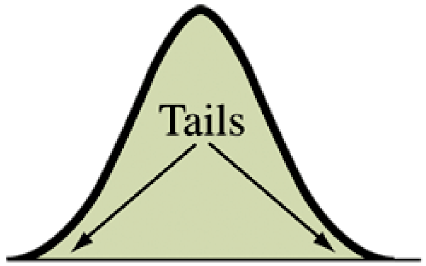
- The left and right sides mirror each other.
- Typically, the mean = median = mode.



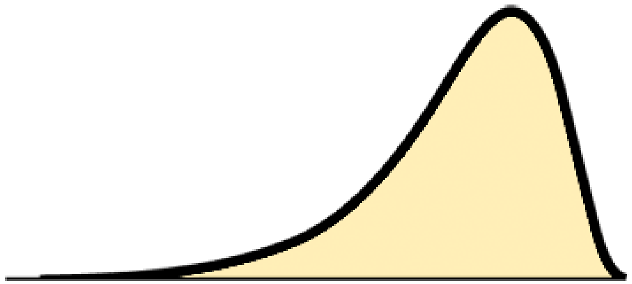
- If you measure the **heights of adult males** in a homogeneous population, you might get a fairly symmetric (normal-like) distribution centered around a mean (e.g., 175 cm) and the **distribution tails decreasing at about the same rate in both directions**.

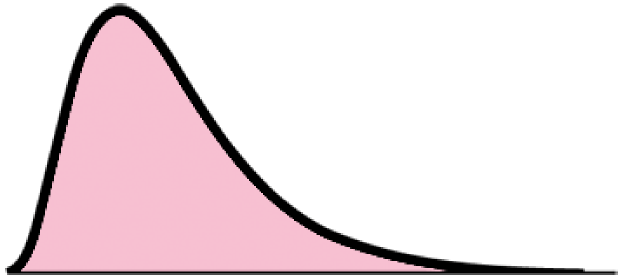
To conclude - Skewness

Symmetric Distributions: if both left and right sides of the histogram are mirror images of each other.



Symmetric


Skewed to the left


Skewed to the right

A distribution is **skewed to the left** if the left tail is longer than the right tail.

A distribution is **skewed to the right** if the right tail is longer than the left tail.

Heavier tail

- Heavier tail refers to the part of a probability distribution that tapers off more slowly than that of a normal distribution.

- This means that in a distribution with heavy tails, extreme values (far from the mean) occur more frequently than they would in a normal distribution.

→ Leading to a greater likelihood of extreme events

Examples of Heavy-Tailed Distributions:

- **Pareto Distribution**: Often used to model wealth distribution, where a small percentage of the population holds a large portion of the wealth.
- **Cauchy Distribution**: Known for its heavy tails and undefined mean and variance.
- **Log-Normal Distribution**: A distribution of a variable whose logarithm is normally distributed, leading to a heavy-tailed nature.

What is Kurtosis (التفلطح)?

- Kurtosis is a statistical measure that describes the shape of a distribution's tails and its peakedness relative to a normal distribution.

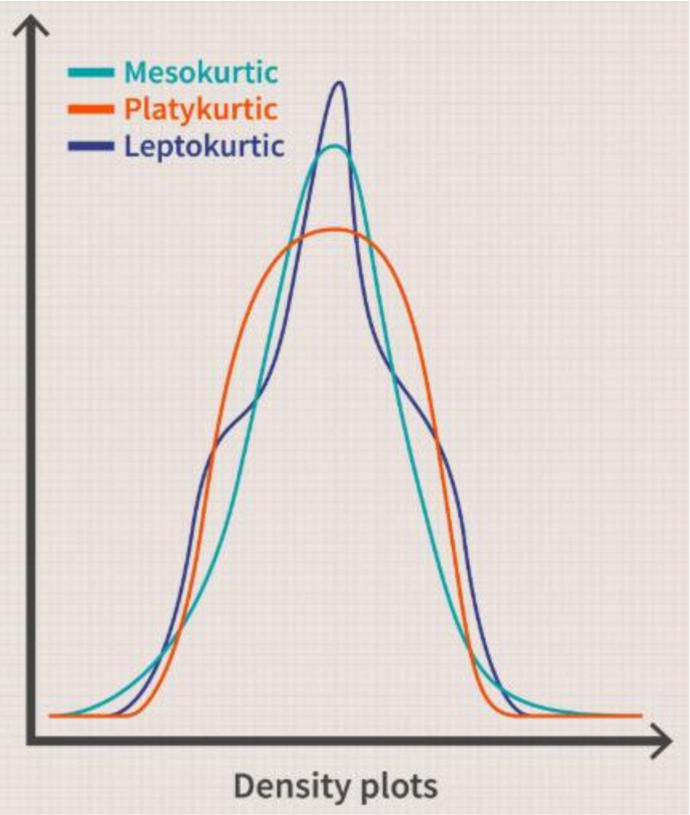
- هو مقياس إحصائي يعبّر عن حدة القمة وسمك الأطراف

- It tells you how much of the data is concentrated in the tails (extreme values) versus the center of the distribution.

What is Kurtosis (التفلطح)?

- **High kurtosis التوزيع المفرطح** (leptokurtic): Indicates a distribution with heavy tails and a sharp peak. This means more extreme outliers than a normal distribution

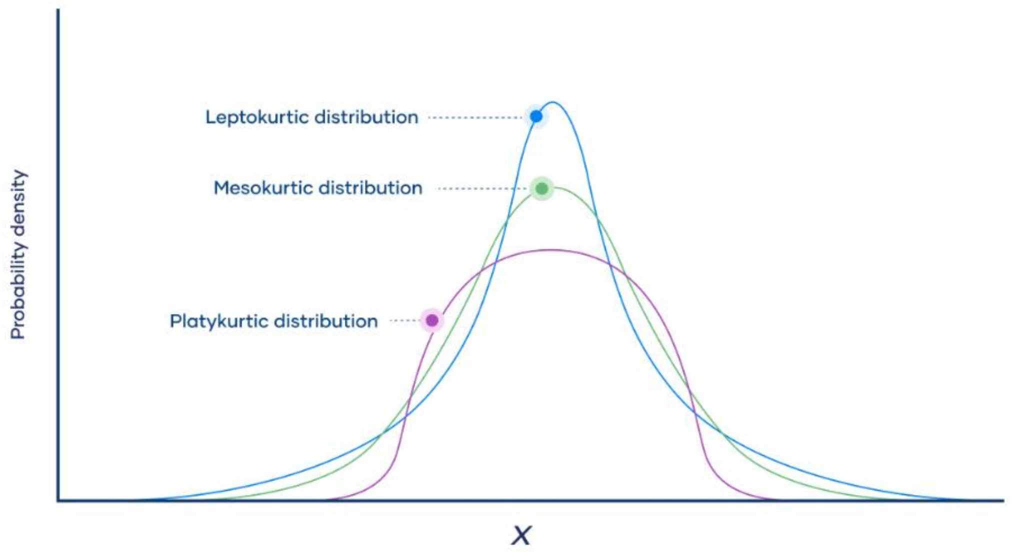
- التوزيع يحتوي على ذيول ثقيلة وقيم متطرفة أكثر من التوزيع الطبيعي



What is Kurtosis (التفلطح)?

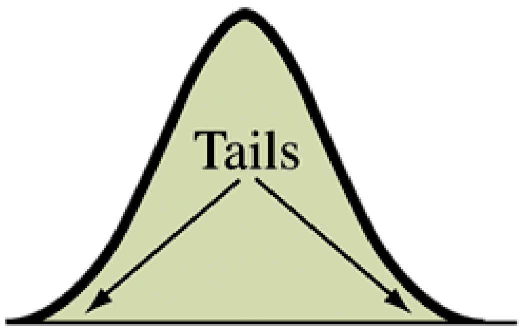
- **Low kurtosis التوزيع المنبسط** (platykurtic): Indicates a distribution with light tails and a flatter peak. Fewer extreme outliers.

- توزيع أكثر تسطحًا مع ذيول خفيفة مقارنة بالتوزيع الطبيعي



What is Kurtosis (التفلطح)?

- **Normal kurtosis** (mesokurtic): Matches the normal distribution, with moderate tails and peakedness.
- The standard benchmark for a normal distribution’s kurtosis is 3 (though some software subtracts 3 to report "excess kurtosis," where 0 represents a normal distribution).



Symmetric

Equation for Kurtosis Calculation:

- The sample kurtosis formula :

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})^4}{s^4} \right) - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Where:

- x_i= each data point
- $\bar{x}$: mean of the data
- In practice, you don't need to compute this by hand—most statistical software (like Python, R, or Excel) can calculate it for you.

How to Interpret Kurtosis?

- Kurtosis > 3 (Excess > 0): Heavy tails, more outliers (e.g., stock market returns often show this).
- Kurtosis < 3 (Excess < 0): Light tails, fewer outliers (e.g., a uniform distribution).
- Kurtosis = 3 (Excess = 0): Normal distribution-like behavior.

Practical Steps to "Know" Kurtosis

- Collect your data: Ensure you have a dataset to analyze.
- Use a tool:
  - Python: Use scipy.stats.kurtosis() or pandas.DataFrame.kurt().
  - R: Use kurtosis() from the moments package.
  - Excel: Use KURT() function.
- Interpret the result: Compare it to 3 (or 0 for excess kurtosis) to understand the distribution’s shape.

Kurtosis & Skewness

Key Differences

- **Aspect of Shape**:
  - Skewness = asymmetry (left vs. right).
  - Kurtosis = tail weight and central peakedness.
- **Independence**:
  - A distribution can have high kurtosis but zero skewness (e.g., a symmetric leptokurtic distribution like a t-distribution).
  - A distribution can be skewed but have normal kurtosis (e.g., a log-normal distribution).

Kurtosis & Skewness

- Mathematical Connection
- There's a theoretical bound linking skewness and kurtosis. For any distribution:

$$Kurtosis \geq Skewness^2 + 1$$

- This is known as the Pearson inequality for excess kurtosis.

- It shows that large skewness tends to imply at least some minimum level of kurtosis, because asymmetry often pushes data into the tails.

Kurtosis & Skewness

- Mathematical Connection
- There's a theoretical bound linking skewness and kurtosis. For any distribution:

$$Kurtosis \geq Skewness^2 + 1$$

- This is known as the Pearson inequality for excess kurtosis.

- Example: If skewness = 2, then kurtosis ≥ 5 (excess kurtosis ≥ 2). This reflects how extreme values driving skewness also boost tail weight.

Why It Matters

- Together, skewness and kurtosis give a fuller picture of a distribution:
  - Skewness tells you "which way" the data leans.
  - الالتواء يُخبرك عن "اتجاه ميل" البيانات
  - Kurtosis tells you "how extreme" the tails are and how peaked the center is.
  - In fields like finance, high kurtosis (risk of extreme events) with skewness (direction of those events) can signal potential volatility.

Why it so important to identify kurtosis?

- Identifying kurtosis in a dataset is important because it provides critical insights into the behavior of the data, particularly the presence and impact of extreme values (outliers) and the overall shape of the distribution.
- This has practical implications across various fields like finance, quality control, risk management, and scientific research.

→ Here’s why kurtosis matters and when it’s essential to pay attention to it:

Why it so important to identify kurtosis?

- **Understanding Extreme Events (Tail Behavior)**
- Kurtosis measures **how heavy or light** the tails of a distribution are compared to a normal distribution. This is crucial because:

- **High kurtosis (leptokurtic)**: Indicates a greater likelihood of extreme outliers. For example:
  - In finance, stock returns (عوائد الأسهم)with high kurtosis suggest a higher risk of crashes or booms (انهيارات أو طفرات مفاجئة), beyond what a normal distribution predicts.

Why it so important to identify kurtosis?

- **Understanding Extreme Events (Tail Behavior)**
- Kurtosis measures **how heavy or light** the tails of a distribution are compared to a normal distribution. This is crucial because:

- **Low kurtosis (platykurtic)**: Suggests fewer extreme values, implying stability or uniformity.
  - Why it matters: Many statistical models (e.g., regression, hypothesis testing) assume normality (kurtosis = 3). If kurtosis deviates significantly, those models may underestimate or overestimate risks or fail to fit the data properly.

Why it so important to identify kurtosis?

Assessing Risk and Uncertainty

- Kurtosis directly ties to the likelihood of rare, high-impact events:
- In risk management (إدارة المخاطر) , high kurtosis signals that "black swan" events (unexpected extremes) are more probable. Ignoring it could lead to under preparedness (مخاطر جسيمة).

- Example: During the 2008 financial crisis, many models assumed normal distributions and missed the high-kurtosis reality of market returns, leading to catastrophic miscalculations.

Why it so important to identify kurtosis?

Evaluating Data Fit for Statistical Methods

- Many statistical tools assume a normal distribution (or at least moderate kurtosis):
  - Why it matters: Identifying kurtosis helps you decide whether to transform the data (e.g., log transformation) or use robust, non-parametric methods instead of standard ones.

Key Takeaways

- High kurtosis = More extreme outliers = More variability & risk
- Low kurtosis = Fewer outliers = More stable & predictable behavior
- In applications, kurtosis helps in risk assessment, anomaly detection, and decision-making by revealing how much extreme data points influence the distribution.

Example

- Given Data

#	Data X_i
1	4
2	8
3	9
4	6
5	4
6	5

- Sample size: n=6
- Sample mean: $\bar{X} = 6$
- Sample standard deviation: $s = 2.097618$

Equation for Kurtosis Calculation:

- The sample kurtosis formula :

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})^4}{s^4} \right) - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Where:

- x_i= each data point
- $\bar{x}$: mean of the data
- In practice, you don’t need to compute this by hand—most statistical software (like Python, R, or Excel) can calculate it for you.

Example

- Compute $\sum_{i=1}^{n}\left(\frac{(x_i-\bar{x})^4}{s^4}\right)$

#	Data X_i	$\left(\frac{(X_i-\bar{X})^4}{s^4}\right)$
1	4	0.826446
2	8	0.826446
3	9	4.183884
4	6	0
5	4	0.826446
6	5	0.051653

- $\sum_{i=1}^{n}\left(\frac{(x_i-\bar{x})^4}{s^4}\right) = 6.714876$

Example

- Compute Each Term in the Formula
- $\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{(x_i-\bar{x})^4}{s^4}\right) = \frac{6(7)}{(5)(4)(3)} \times 6.714876$
- $\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{(x_i-\bar{x})^4}{s^4}\right) = 4.700413$
- Second Term:
- $\frac{3(n-1)^2}{(n-2)(n-3)} = \frac{3(5)^2}{(4)(3)} = 6.26$

Example

- Compute Kurtosis

- $K = 4.700413 - 6.25$
- $K = -1.54959$

- Since $K < 0$, the distribution is Platykurtic, meaning it is too flat with lighter tails than a normal distribution.

References

- Montgomery, D.C. and Runger, G.C., 2010. Applied statistics and probability for engineers. John wiley & sons.
- Devore, J.L., 2011. Probability and Statistics for Engineering and the Sciences. Cengage learning.
- Gubner, J.A., 2006. Probability and random processes for electrical and computer engineers. Cambridge University Press.
- Agresti, A. and Franklin, C., 2007. The art and science of learning from data. Upper Saddle River, New Jersey, 88.
- Navidi, W.C., 2008. Statistics for engineers and scientists (Vol. 2, No. 4). New York, NY, USA: McGraw-Hill Higher Education.