

# Introduction to Econometrics

Prepared by: Prof. GUECHARI Yasmina

# I. What is Econometrics?

- 1. Definition:** Econometrics can be defined as the application of statistical techniques to economic data in order to analyze and quantify the relationships between different economic variables.

# I. What is Econometrics?

**2. The purpose of econometrics** is to provide empirical evidence and quantitative analysis to better understand economic phenomena and make informed decisions.

By using econometric techniques, economists can assess the impact of various economic factors on outcomes, conduct policy evaluations, predict future trends, and enhance our overall understanding of economic behavior.

# I. What is Econometrics?

## **3. The scope of econometrics** is:

- Forecasting,
- Policy evaluation,
- Understanding economic behavior.

## II. Data collection

### 1. **Sources of Economic Data:** this could include:

- government agencies, which often collect and publish economic statistics.
- International organizations like the World Bank, World Trade Organisation (WTO) databases provide aggregate data on trade in goods and services, or CFI or private financial data providers.
- Surveys conducted by various organizations, both public and private, to gather specific economic information. These surveys may cover topics such as employment, consumer spending, business activities, etc.

## II. Data collection

- The data can be primary data or secondary data.
- Reliable data sources: use reputable and trustworthy sources to ensure the accuracy and reliability of economic data; unreliable data can lead to inaccurate conclusions, affecting policy-making, business strategies, and academic research.
- Data quality: assessing data quality, such as accuracy, completeness, consistency, and timeliness; the poor data quality can introduce biases and errors into analyses, impacting the reliability of economic assessments.

## II. Data collection

### 2. **Types of Data: There is three types of Economic data**

- i. **Cross-sectional data:** Data collected from different units at a specific point in time.
- ii. **Time series data:** Data collected from a single unit over a time period.
- iii. **Panel data:** A combination of cross-sectional and time series data, collected from multiple units over time.

# III. Preliminary analysis of the Data

**1. Descriptive Statistics:** descriptive statistics provide a comprehensive summary of key characteristics of a dataset, helping economists and researchers to better understand the central tendency\*, spread, asymmetry, and tail behavior of the data they are analyzing in econometric studies.

\* The central tendency is the statistical measure that identifies a single value as representative of an entire distribution. It aims to provide an accurate description of the entire data.



# III. Preliminary analysis of the Data

**i. Mean:** the mean or average, is a measure of the central tendency of a dataset. It is calculated by adding up all the values in a dataset and dividing by the number of observations.

- The mean provides a single numerical summary that represents the "typical" value in the dataset.

- The mean is:  $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$

- **Example:**

Time period	Jan-1	Jan-2	Jan-3	Jan-4	Jan-5	Jan-6	Jan-7	Jan-8	Jan-9	Jan-10
Money spent (X)	60	52	42	45	40	35	38	30	25	20

Mean:  $\bar{X} = \frac{387}{10} = 38.7$

# III. Preliminary analysis of the Data

- **How to calculate the mean using excel:**

- For example, if your data is in cells A1 to A10,

1. in a cell where you want the result to appear (let's say B1), type: = and the word average and the average function will automatically appear choose it,

2. Select the data set A1:A10 you get the formula **= average(A1:A10)**

then press Enter

3. The cell B1 will now display the Mean of your dataset

# III. Preliminary analysis of the Data

- ii. Median:** the median is the middle value of a dataset when it is ordered from lowest to highest.
  - Unlike the mean, the median is not affected by extreme values or outliers, making it a robust measure of central tendency.
  - It is especially useful when the dataset is not symmetrically distributed.

year	sales	
2010	141	
2011	157	
2012	177	
2013	191	203
2014	201	
2019	205	
2015	206	
2017	211	
2018	212	
2016	214	

### III. Preliminary analysis of the Data

- **How to calculate the median using excel:**
- For example, if your data is in cells A1 to A10,
  1. in a cell where you want the result to appear (let's say B1), type: = and the word Median and the Median function will automatically appear choose it,
  2. Select the data set A1:A10 you get the formula **= Median(A1:A10)** then press Enter
  3. The cell B1 will now display the Median of your dataset

# III. Preliminary analysis of the Data

**iii. Mode:** mode refers to the value or values that appear most frequently in a dataset. In other words, it is the data point or points that occur with the highest frequency. Unlike the mean (average) and median (middle value), which focus on central tendencies, the mode highlights the most common values in a distribution.

- A dataset may have:

- 1. Unimodal:** If it has one mode (one most frequent value).

- 2. Bimodal:** If it has two modes (two distinct most frequent values).

- 3. Multimodal:** If it has more than two modes.

- It's important to note that a dataset can also be "uniform" or "uniformly distributed" if no value stands out as being more frequent than others, resulting in no mode.

# III. Preliminary analysis of the Data

## How to calculate the Mode using Excel

**Enter your data into a column in Excel.**

1. Let's say your data is in column A from A1 to A10.

### **1. Use the MODE.SNGL function:**

1. In an empty cell, type the formula =MODE.SNGL(A1:A10) where A1:A10 is the range of your data.
2. Press Enter.

2. The result will be the mode of your dataset.

# III. Preliminary analysis of the Data

**iv. Variance:** the variance measures the degree of variation or fluctuation of data

- $Variance(X) = \sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$

- **To calculate variance using excel we follow the following step**

- For example, if your data is in cells A1 to A10,

1. In a cell where you want the result to appear (let's say B1), type = and the word var and the var function will automatically appear choose it,

2. Select the data set A1:A10 you get the formula **= VAR(A1:A10)** then press Enter

3. The cell B1 will now display the variance of your dataset.

# III. Preliminary analysis of the Data

**v. Standard deviation:** standard deviation measures the dispersion in a set of values. The **standard deviation** tells us, how far each value lies from the mean.

- A higher standard deviation indicates greater variability in the time series data, suggesting more volatility.

- $std. Dev = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}}$

- A high standard deviation means that values are generally far from the mean, while a low standard deviation indicates that values are clustered close to the mean.



### III. Preliminary analysis of the Data

- **How to calculate standard deviation using excel:**

- For example, if your data is in cells A1 to A10,

1. in a cell where you want the result to appear (let's say B1), type: = and the word STDEV and the STDEV function will automatically appear choose it,

2. Select the data set A1:A10 you get the formula **= STDEV(A1:A10)**

then press Enter

3. The cell B1 will now display the standard deviation of your dataset

# III. Preliminary analysis of the Data

**vi. Skewness:** skewness measures the asymmetry of the distribution of values in a dataset.

- Positive skewness indicates a right-skewed distribution (tail on the right), while negative skewness indicates a left-skewed distribution. Skew=0 means the data is asymmetric (تماثل).
- Understanding skewness helps to assess the shape of the distribution and the location of extreme values.
- **How to calculates the skewness of the specified data range in excel.**
- **For example**, if your data is in cells A1 to A10, you would enter =SKEW(A1:A10) into another cell (let's say B1) and press Enter. The result in cell B1 will be the skewness of the data in cells A1 to A10.

### III. Preliminary analysis of the Data

- If the measures of central tendency, the mode, the median, and the Mean do not match, the distribution is considered skewed.
- In the case of symmetrical distributions, the mean, median, and mode are equal, and the farther the curve is far from symmetry, the farther these values are far from each other.

### III. Preliminary analysis of the Data

**vii. Kurtosis: Kurtosis:** Flatness is a measure of the degree to which the peak of a distribution is high relative to a normal distribution . If the distribution has a high peak (larger than the normal distribution), it is said to be leptokurtic (مدبب). If the distribution has a flat peak, it is said to be platykurtic (مفلطح), and if the top of the distribution is average (not pointed and not flat), it is called mesokurtic.

# III. Preliminary analysis of the Data

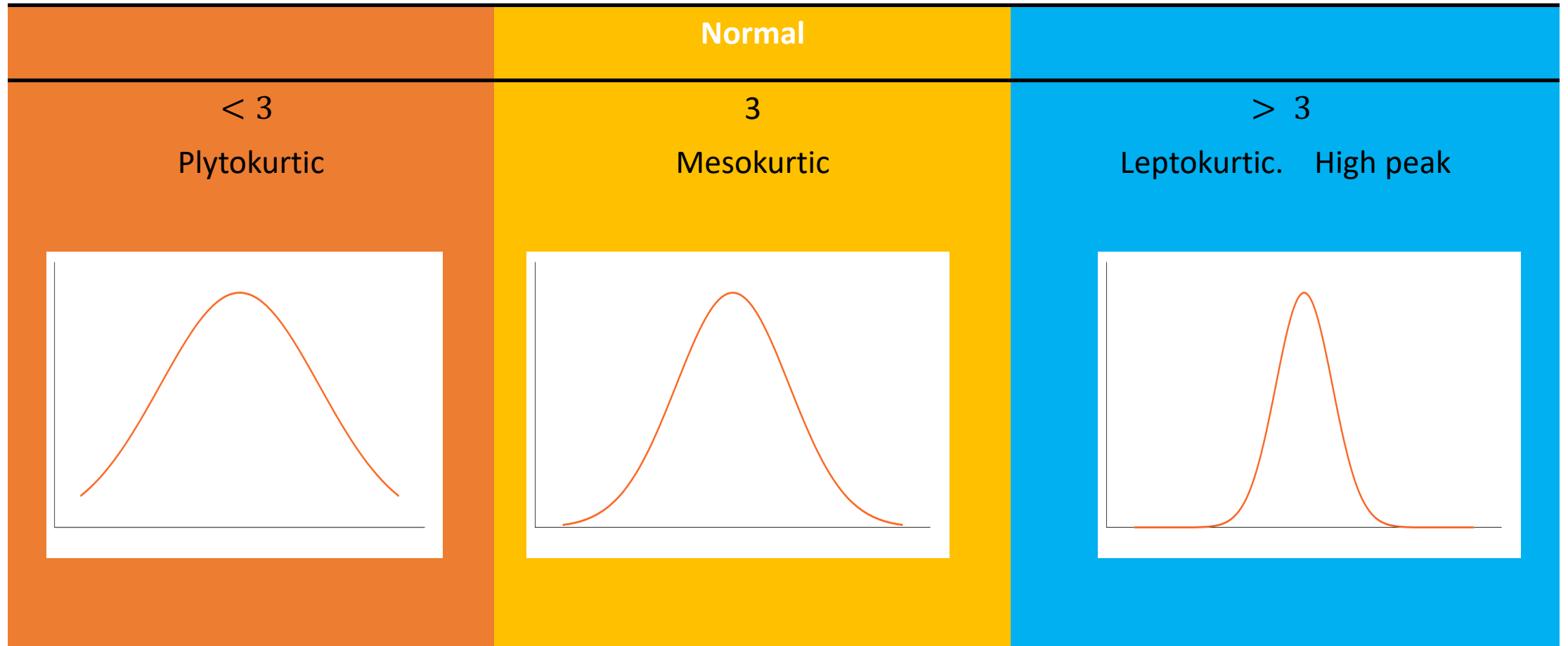
## How to calculate kurtosis in excel

- For example, if your data is in cells A1 to A10, you would enter `=KURT(A1:A10)` into another cell (let's say B1) and press Enter. The result in cell B1 will be the kurtosis of the data in cells A1 to A10.
- Keep in mind that interpreting kurtosis can be context-dependent. A positive kurtosis indicates heavier tails than a normal distribution, while negative kurtosis indicates lighter tails. A kurtosis value close to zero suggests a distribution with tails similar to a normal distribution.

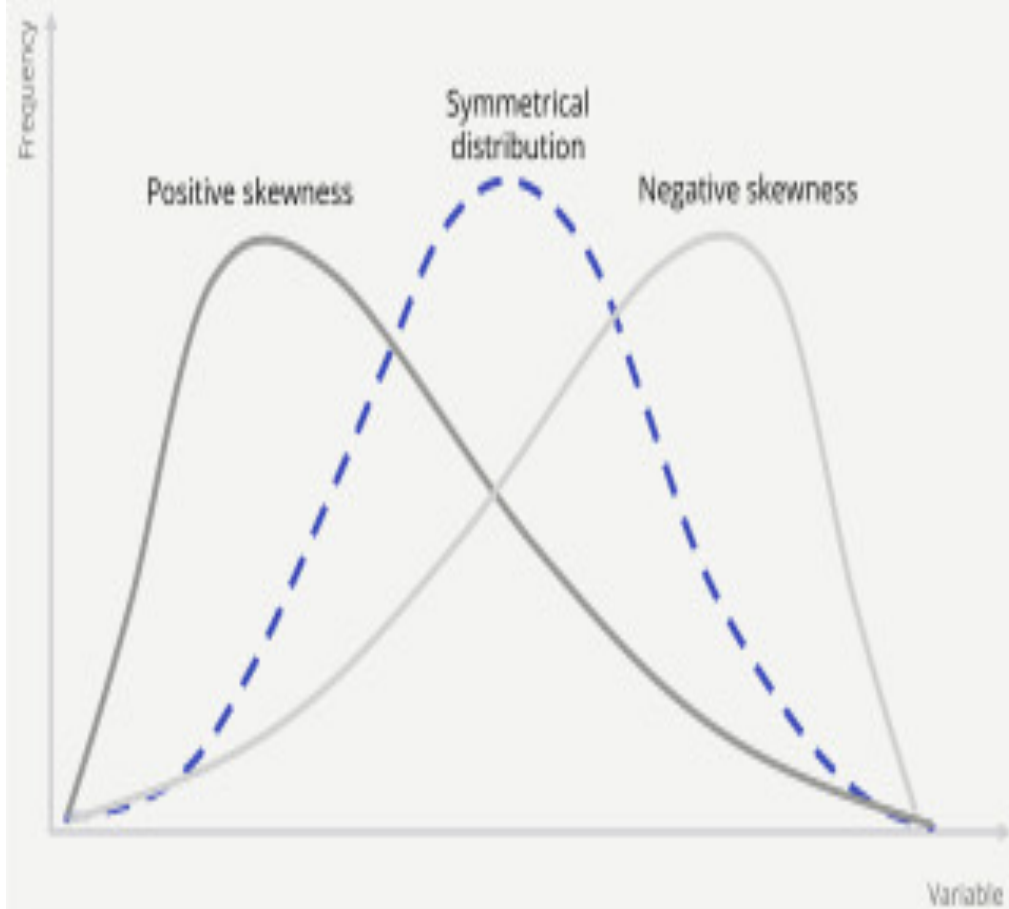
### III. Preliminary analysis of the Data

- The kurtosis of a normal distribution equals 3.
- The distribution is flattened (مفلطح), when the kurtosis coefficient is less than 3.
- The distribution is tapered (مدبب), when the flatness coefficient is greater than 3.

# Kurtosis



## SKEWNESS



## KURTOSIS





# Excel Application

- **Exercise: Sector-wise Returns:** Consider the annual returns of three sectors (Technology, Finance, and Healthcare) over the past ten years (in percentage):
    - **Technology**  
[15.2,10.5,-5.7,12.3,8.6,19.8,-3.2,7.1,11.4,14.7,15.2,10.5,-5.7,12.3,8.6,19.8,-3.2,7.1,11.4,14.7]
    - **Finance:** [8.3,6.5,-2.1,9.8,5.4,12.1,-1.5,4.6,7.2,10.3, 8.3,6.5,-2.1,9.8,5.4,12.1,-1.5,4.6,7.2,10.3]
    - **Healthcare:** [5.6,4.1,2.9,6.7,3.2,7.8,1.5,3.9,5.1,6.3, 5.6,4.1,2.9,6.7,3.2,7.8,1.5,3.9,5.1,6.3]
1. Calculate the mean return for each sector.
  2. Calculate the standard deviation of returns for each sector.
  3. Determine the sector with the highest kurtosis.