A review of descriptive statistics

Prepared by: Prof. GUECHARI Yasmina

II. Types of 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 flatness of the data they are analyzing in econometric studies.

III. Preliminary analysis of the Data

Examples of central tendancy measures

- **Mean**: The average income of a group of people.
- **Median**: The middle value of house prices in a city.
- **Mode**: The most common shoe size sold in a store.

Examples of spreadness' measures

- Variance
- Standard deviation

Examples of asymmetry and flatness measures of the data

- Skewness
- Kurtosis

III. Preliminary analysis of the Data

i. **Mean:** the mean or average. 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 B2 to B11,

1. in a cell where you want the result to appear (let's say B13), type: = the word average and the average function will automatically appear choose it,

2.Select the dada set B2:B11 you get the formula **= average(B2:B11)**

then press Enter

3.The cell B13 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 B2 to B11,

1. in a cell where you want the result to appear (let's say B13), type: = and the word Median and the Median function will automatically appear choose it,

2.Select the dada set B2:B11 you get the formula **= Median(B2:B11)**

then press Enter

3.The cell B13 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, or the most common values in a distribution.

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 B from B2 to B11.

1.Use the MODE.SNGL function:

1. In an empty cell, type the formula =MODE.SNGL(B2:B11) where B2:B11 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

- $Variance(X) = \sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$
- **To calculate variance using excel we follow the following step**
- For example, if your data is in cells B2 to B11,

1. In a cell where you want the result to appear (let's say B13), type = and the word var and the var function will automatically appear choose it,
2. Select the dada set B2:B11 you get the formula **= VAR(B2:B11)** then press Enter
3. The cell B13 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.
- $std.Dev = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}}$
- 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 B2 to B12,

1. in a cell where you want the result to appear (let's say B13), type: = and the word STDEV and the STDEV function will automatically appear choose it,

2.Select the dada set B2:B11 you get the formula **= STDEV(B2:B11)**

then press Enter

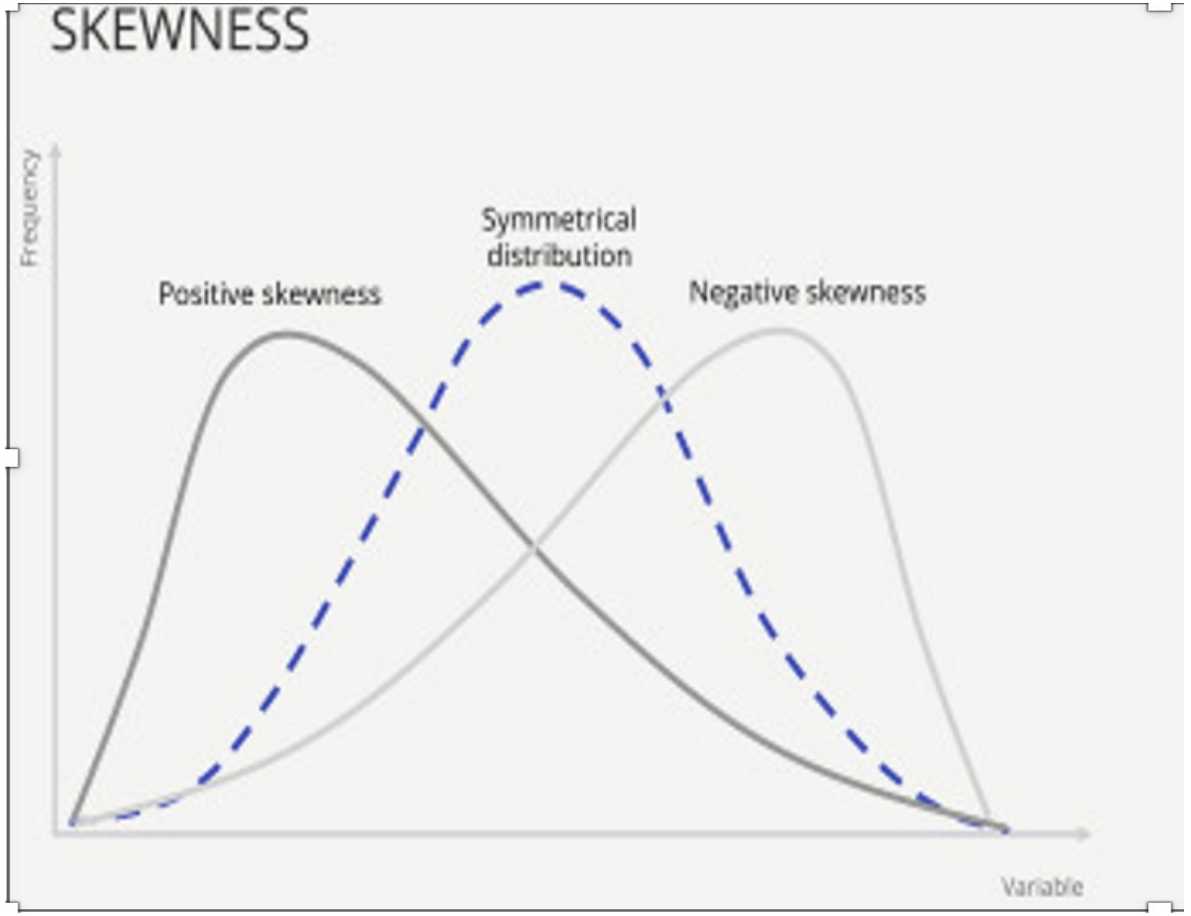
3.The cell B13 will now display the standard deviation of your dataset

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)
- Negative skewness indicates a left-skewed distribution.
- Skew=0 means the data is asymmetric.

Preliminary analysis of the Data



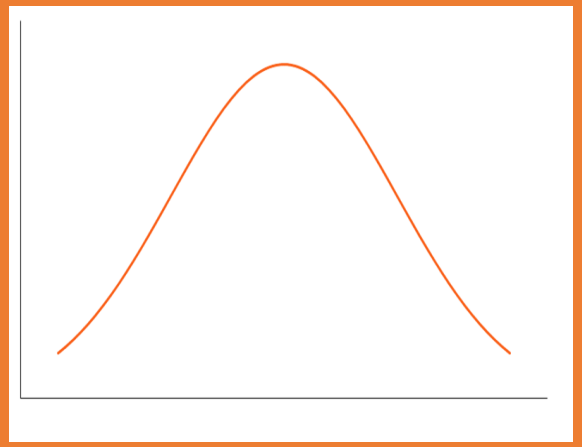
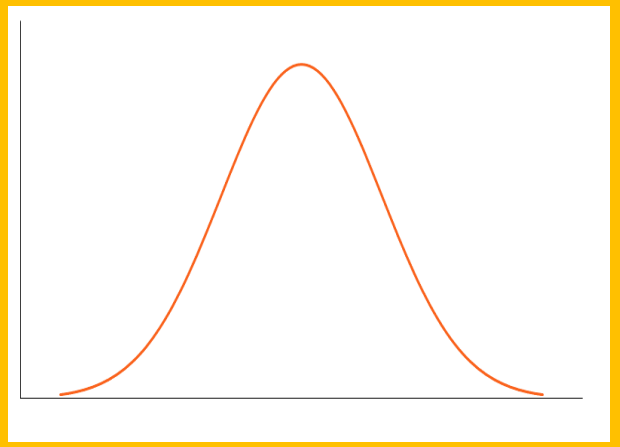
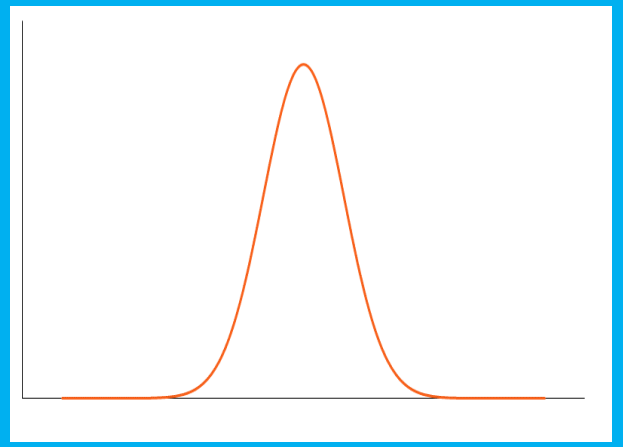
III. Preliminary analysis of the Data

- 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

vii. **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. If the distribution has a flat peak, it is said to be platykurtic, and if the top of the distribution is average (not pointed and not flat), it is called mesokurtic.

Kurtosis

	Normal	
< 3	3	> 3
Platykurtic	Mesokurtic	Leptokurtic. High peak
		

III. Preliminary analysis of the Data

How to calculate kurtosis in excel

- For example, if your data is in cells B2to B11, you would enter =KURT(B2:B11) into another cell (let's say B13) and press Enter.
- The result in cell B13 will be the kurtosis of the data in cells B2 to B11.

III. Preliminary analysis of the Data

- The kurtosis of a normal distribution equals 3.
- The distribution is paltykurtic, when the kurtosis coefficient is less than 3.
- The distribution is leptokurtic, when the flatness coefficient is greater than 3.

Excel Application

- **Exercise: Sector-wise Returns:** Consider the annual returns of three sectors (Technology, Finance, and Healthcare) over the past ten years (in percentage): as in the following table

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.

Excel Application

obsno	Tech	Fin-	Health	
1	15,2	8,3	5,6	
2	10,5	6,5	4,1	
3	5,7	2,1	2,9	
4	12,3	8,5	7,8	
5	8,6	12,1	3,9	
6	19,8	4,6	5,3	
7	7,1	7,2	9,5	
8	11,4	10,3	6,3	
9	14,7	8,3	5,6	
10	15,2	6,5	6,7	
11	11,6	12,1	8,1	