People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research
Mohamed Khider University of Biskra

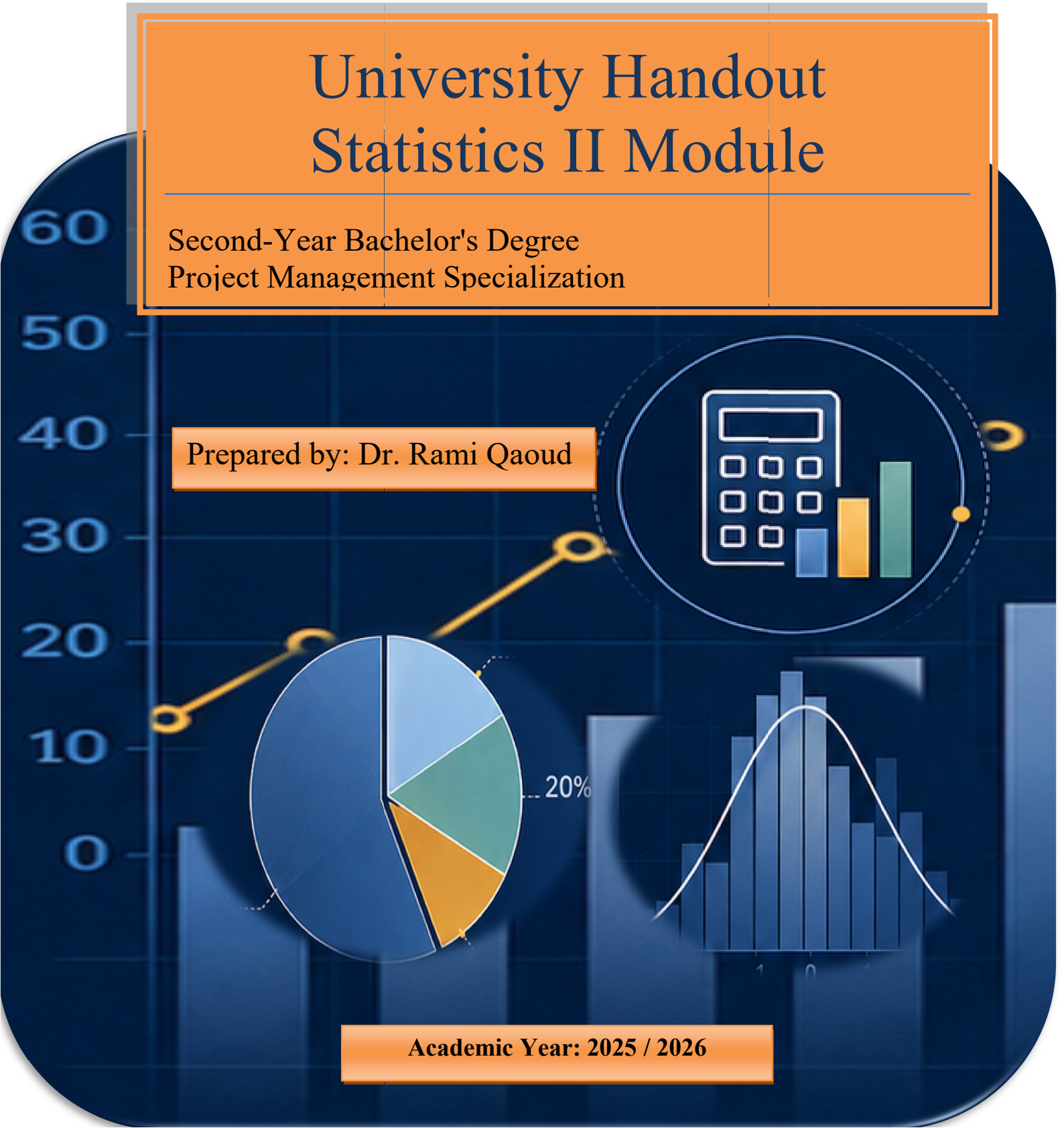
Faculty of Architecture, Urban Planning, Civil Engineering and Hydraulics.

Department of Architecture

University Handout
Statistics II Module

Second-Year Bachelor's Degree
Project Management Specialization

Prepared by: Dr. Rami Qaoud



Academic Year: 2025 / 2026

Table of Contents:

CHAPTER III : Measures of Central Tendency.

CHAPTER IV : Measures of Dispersion.

List of figures:

General Introduction

Descriptive statistics is a fundamental branch of statistics that focuses on the systematic organization, summarization, and presentation of data in a clear and meaningful way. Its main purpose is to transform raw and unstructured data into a simplified form that can be easily understood, interpreted, and communicated. Unlike inferential statistics, which aims to draw conclusions and make predictions about a population, descriptive statistics is strictly concerned with describing the characteristics of the observed dataset as it is, without extending interpretations beyond the available data.

This field relies on a combination of numerical, graphical, and tabular techniques to analyze and present information effectively. Numerical methods include measures of central tendency such as the mean, median, and mode, which describe the central position of the data. They also include measures of dispersion such as the range, variance, and standard deviation, which provide valuable insight into the spread, variability, and consistency of the dataset. In addition to numerical summaries, descriptive statistics uses graphical representations such as bar charts, pie charts, histograms, and line graphs, as well as tabular methods like frequency distribution tables, to visually and

systematically present data. Together, these tools help reveal patterns, trends, and structural characteristics within the data.

Descriptive statistics plays a crucial role in a wide range of academic and professional disciplines, including science, engineering, economics, social sciences, and environmental studies. It forms the initial step in any statistical analysis by providing a clear overview of the data and helping researchers identify important features before applying more advanced inferential methods. By offering a structured summary of information, it enhances understanding and supports more informed analysis and interpretation.

In academic and professional contexts, descriptive statistics is particularly valuable for presenting data in a clear, concise, and effective manner. It enables researchers, students, and professionals to communicate complex information in a simplified form, thereby improving the clarity of reports, studies, and decision-making processes.

In the field of architecture, descriptive statistics is especially useful for analyzing and interpreting various types of spatial and environmental data. For instance, architects may use statistical measures to examine the distribution of room sizes in residential or commercial buildings, helping to identify average dimensions and common design patterns. Data collected from user satisfaction

surveys can also be analyzed using descriptive statistics to assess the functionality, comfort, and efficiency of architectural spaces.

Furthermore, descriptive statistics is widely applied in environmental and urban analysis. Architects and urban planners can analyze data related to temperature variations, daylight availability, energy consumption, population density, land use distribution, and traffic flow. By calculating averages, identifying variability, and presenting results visually, they can make more informed and evidence-based design decisions. These insights contribute to improving energy efficiency, enhancing sustainability, and optimizing spatial organization in urban environments.

Overall, descriptive statistics provides a powerful and essential foundation for understanding and interpreting data. It supports effective decision-making, enhances the clarity of communication, and bridges the gap between raw data and meaningful insights. In both academic research and professional practice, it remains an indispensable tool for analyzing information and guiding well-informed conclusions and design strategies.

CHAPTER I:

GENERAL INTRODUCTION OF SUBJECT.

CHAPTER I Content

- **Introduction**

1. **Functions of Statistics**
2. **Branches of Statistics**
3. **Types of Data and Measurement Methods**
4. **Data Collection**
5. **Data Collection Methods**
6. **Types of Samples**
7. **Stages of Statistical Research.**

- **Conclusion**

- Introduction.

The field of statistics focuses on methods for collecting, quantifying, summarizing, and organizing data in a way that allows for its use in describing and analyzing information. This is essential for aiding decision-makers in making informed choices in their areas of work. Statistics is an independent discipline and serves as a fundamental pillar in most studies related to field research in the domains of science, engineering, economics, and the humanities. It is often mistakenly perceived solely as the counting and numerical expression of objects, which is a misconception about statistics. In this section of the lecture, we will explore the components of statistics that are relevant to researchers and university students in their research endeavors and academic studies, as well as in their university lives.

Descriptive statistics is a branch of statistics that deals with the collection, organization, analysis, and presentation of data in a meaningful way. Unlike inferential statistics, which aims to make predictions or generalizations about a population based on a sample, descriptive statistics focuses on summarizing and describing the main features of a dataset.

This field involves the use of numerical measures, such as mean, median, mode, and standard deviation, to provide insights into the central tendency, variability, and distribution of data. Graphical representations like histograms, bar charts, and pie charts are also commonly used in descriptive statistics to visualize data trends and patterns.

By simplifying complex data into understandable summaries, descriptive statistics plays a crucial role in fields like research, business, and social sciences, where making sense of large amounts of information is essential for informed decision-making.

1- Functions of Statistics

Statistics encompasses three main functions:

- Descriptive data.
- Statistical Inference
- Forecasting.

1.1. Descriptive data.

One of the most important functions of statistics is the description of data, which involves collecting, organizing, and summarizing the data. This is crucial because raw data cannot be effectively utilized or understood unless it is collected, tabulated, and presented in the form of graphs and statistical indicators that help in understanding the phenomenon.

Data description is a fundamental function of statistics that involves the systematic collection, organization, and summarization of data to provide a clear understanding of the information being analyzed. This process is essential for transforming raw data into a format that is interpretable and useful for decision-making. Here are the key components of data description:

1. **Collection of Data**:
    - This involves gathering data from various sources, which may include surveys, experiments, observational studies, and existing databases. The quality and reliability of the data collected are crucial for accurate analysis.
2. **Organization of Data**:
    - Once data is collected, it must be organized in a structured manner. This can include sorting data into categories, creating tables, or using software tools for data management. Proper organization helps in identifying patterns and trends within the dataset.

3. **Summarization of Data**:
    - Summarizing data involves using statistical measures to provide a concise overview of the main features of the dataset. Common statistical measures include:
        - **Measures of Central Tendency**: Such as the mean, median, and mode, which represent the average or typical values in the data.
        - **Measures of Dispersion**: Such as range, variance, and standard deviation, which describe the spread or variability of the data.
4. **Graphical Representation**:
    - Data can be visually represented using various types of charts and graphs, such as histograms, bar charts, pie charts, and line graphs. These visual aids help to convey complex information in a more accessible and comprehensible way.
5. **Statistical Indicators**:
    - Key indicators are derived from the data to highlight important aspects of the phenomenon under study. These may include proportions, percentages, or frequency distributions that assist in interpreting the data more effectively.

1.1.2. Importance of Descriptive data.

- **Understanding Phenomena**: By describing data, researchers and decision-makers can gain insights into the underlying patterns and trends, which is essential for understanding the phenomena they are studying.
- **Facilitating Informed Decisions**: Well-organized and summarized data provides a solid foundation for making informed decisions based on evidence rather than intuition.

- **Identifying Anomalies**: Data description helps in detecting outliers or anomalies within the data, which may require further investigation or correction.

In summary, data description is a crucial step in the statistical analysis process, serving as the gateway to more advanced analyses and interpretations. It lays the groundwork for drawing conclusions and making informed decisions based on the data at hand.

1.2. Statistical Inference

In scientific research, statistical inference is one of the most important functions of statistics. It aims to select a subset of the population, known as the sample, in a scientifically appropriate manner. The goal is to use and analyze the data from that sample to reach valid scientific conclusions that can be generalized to the rest of the population.

Statistical inference is based on two main principles:

- **Estimation**

  This involves estimating and calculating the indicators of the sample data, referred to as **Statistics**. The indicators of the population are called **Parameters**. The statistical measures derived and calculated from the sample are known as **Point Estimates**. These can also be used to estimate a range, which is referred to as **Interval Estimates**.
- **Tests of Hypotheses**

  In this process, sample data is used to make a scientific decision regarding hypotheses about the studied indicators.

Statistical Inference is a branch of statistics that focuses on drawing conclusions about a population based on data collected from a sample. It plays a crucial role in research and data analysis by allowing researchers to make informed decisions and predictions about larger groups without having to collect data from every member of that group.

1.2.1. Key Components of Statistical Inference.

1. **Estimation**:
    - **Point Estimation**: This involves calculating a single value (statistic) from the sample data to estimate a population parameter. For example, using the sample mean to estimate the population mean.
    - **Interval Estimation**: This provides a range of values (confidence interval) within which the population parameter is expected to lie. It gives more information than point estimates by indicating the uncertainty associated with the estimation.
2. **Hypothesis Testing**:
    - This involves making assumptions (hypotheses) about a population parameter and using sample data to test these assumptions.
    - **Null Hypothesis (H0)**: A statement that there is no effect or no difference, and it is what the test seeks to disprove.
    - **Alternative Hypothesis (H1 or Ha)**: This represents the statement that there is an effect or a difference.
    - Statistical tests (e.g., t-tests, chi-square tests) are used to determine whether to reject the null hypothesis based on the sample data.
3. **Sampling Distribution**:
    - The distribution of a statistic (like the sample mean) obtained from multiple samples of the same size from the same population.

Understanding the sampling distribution is essential for conducting hypothesis tests and creating confidence intervals.

4. **P-Value**:
    - A p-value helps determine the significance of the results from hypothesis testing. It represents the probability of observing the sample data, or something more extreme, assuming that the null hypothesis is true. A small p-value (typically less than 0.05) suggests strong evidence against the null hypothesis.

1.2.2. Importance of Statistical Inference

- **Generalization**: It allows researchers to generalize findings from a sample to the broader population, making it a powerful tool for scientific research.
- **Decision Making**: Statistical inference provides a framework for making decisions based on data, reducing reliance on intuition alone.
- **Understanding Relationships**: It helps in understanding relationships between variables and identifying significant patterns in data.

1.2.3. Limitations and Considerations

- **Sampling Bias**: If the sample is not representative of the population, the inferences drawn may be misleading.
- **Assumptions**: Many statistical methods rely on certain assumptions (e.g., normality, independence) that, if violated, can affect the validity of the results.
- **Type I and Type II Errors**: Type I error occurs when a true null hypothesis is incorrectly rejected, while Type II error occurs when a false null hypothesis is not rejected.

In conclusion, statistical inference is a fundamental aspect of statistical analysis that enables researchers to make meaningful conclusions about populations based on sample data. It is widely used across various fields, including social sciences, healthcare, and economics, to inform policy and practice.

1.3. Forecasting.

By using data from statistical inference that indicates the behavior of a phenomenon in the past, we can predict its behavior in the present and future, which is known as forecasting.

Forecasting is the process of using historical data and statistical methods to predict future events or trends. It plays a crucial role in various fields, including economics, finance, business planning, and social sciences. By analyzing past behaviors and trends, forecasting helps organizations and individuals make informed decisions and strategize for the future.

1.3.1. Key Components of Forecasting:

1. **Data Collection**:
    - Gathering historical data relevant to the phenomenon being predicted is essential. This data can come from various sources, including surveys, market reports, sales data, and economic indicators.
2. **Statistical Techniques**:
    - Various statistical methods are used for forecasting, including:
        - **Time Series Analysis**: Analyzing data points collected or recorded at specific time intervals to identify trends, seasonal patterns, and cyclic behavior.

- **Regression Analysis**: Assessing the relationships between variables to make predictions. This can involve simple linear regression or multiple regression models.
- **Exponential Smoothing**: A technique that applies decreasing weights to past observations, making it useful for forecasting time series data.

3. **Model Selection**:
   - Choosing the appropriate forecasting model is critical. Factors to consider include the nature of the data, the desired accuracy, and the specific context of the prediction.
4. **Validation and Evaluation**:
   - Once a forecasting model is developed, it is essential to validate its accuracy using techniques such as cross-validation or out-of-sample testing. Performance metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) help evaluate the model's predictive power.
5. **Continuous Improvement**:
   - Forecasting is not a one-time process. As new data becomes available, models can be updated and refined to improve their accuracy. Continuous monitoring and adjustment of forecasting methods are crucial for maintaining relevance.

1.3.2. Applications of Forecasting :

- **Business Planning**: Companies use forecasting to estimate future sales, inventory needs, and market demand, aiding in resource allocation and strategic planning.

- **Economics**: Economists forecast economic indicators such as GDP growth, unemployment rates, and inflation to guide policy decisions and investment strategies.
- **Weather Predictions**: Meteorologists rely on forecasting models to predict weather patterns and natural disasters, helping communities prepare and respond effectively.
- **Healthcare**: Forecasting is used in public health to predict disease outbreaks and allocate resources efficiently.

1.3.3. Limitations of Forecasting :

- **Uncertainty**: Forecasts are inherently uncertain, and unexpected events (e.g., natural disasters, market shifts) can significantly impact outcomes.
- **Dependence on Historical Data**: Forecasting relies on the assumption that past trends will continue into the future, which may not always hold true.
- **Model Limitations**: Each forecasting model has its strengths and weaknesses, and selecting the wrong model can lead to inaccurate predictions.

In summary, forecasting is a vital tool that leverages statistical methods and historical data to predict future outcomes. It is widely used across various fields and plays a crucial role in planning and decision-making processes.

2. Branche of Statistics

The fields of application for statistics are diverse and extensive across all areas of social, human, and natural sciences. In this context, researchers classify the branches of statistics into two main categories:

1. **Descriptive Statistics**
2. **Inferential Statistics**

As we can see in Figure following.

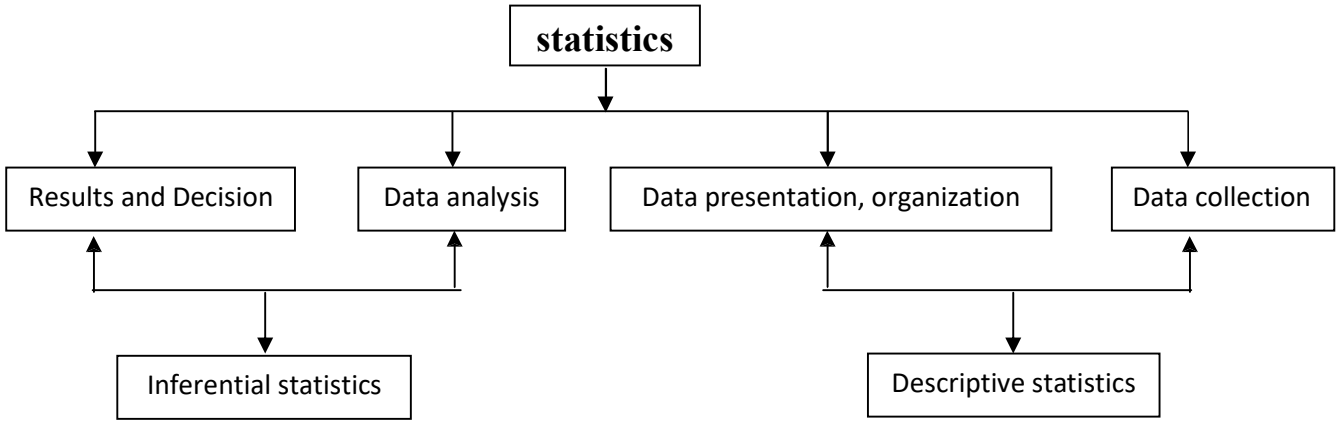


Figure 1-1 . Branche of Statistics

2.1. Descriptive Statistics.

Descriptive statistics is defined as a branch of statistics that focuses on collecting, presenting, and describing data without making conclusions or inferences. This branch of statistics is characterized by the abundance of statistical methods that can be used to process various data types in the field. Therefore, descriptive statistics often requires more time and effort compared to inferential statistics.

Among the most important measures in descriptive statistics are the arithmetic mean, standard deviations, and variance. Descriptive statistics includes the collection, presentation, and description of numerical data through the data gathered from samples. This description is done through several methods, including:

- **Statistical tables**, the most important of which are frequency tables.
- **Graphical representation**, including charts such as bar charts, pie charts, histograms, and cumulative frequency plots (both ascending and descending).
- **Measures of central tendency**, which include the arithmetic mean, geometric mean, median, mode, and quartiles.
- **Measures of dispersion**, which include range, variance, and standard deviation.

2.2. Inferential Statistics.

Inferential statistics is a branch of statistics that consists of a set of statistical methods used to draw conclusions and make decisions based on sample data extracted from a population under study.

3. Types of Data and Measurement Methods :

Data comes in various forms, including numerical and non-numerical. Non-numerical data refers to descriptions of levels or categories. For example, data can pertain to students' grades in a subject during a school term, which would be considered numerical data. Alternatively, data can involve temperature or humidity levels in a city during summer, which is also numerical. On the other hand, data can refer to characteristics such as gender at birth, family status, or attitudes toward comfort, which would be classified as non-numerical.

Therefore, scientists have structured data into two main categories based on these distinctions:

- **Qualitative Data**: This type of data describes characteristics or qualities and is often non-numerical.

- **Quantitative Data**: This type of data consists of numerical values that can be measured and quantified.

3.1. Qualitative Data.

Qualitative data refers to non-numerical data or numerical data presented in the form of levels or categories. In this context, qualitative data is measured according to two criteria:

3.1.1. Nominal Scale.

This consists of non-numerical data that is organized into mutually exclusive groups, where each group has characteristics that distinguish it from other groups. These groups cannot be compared in a meaningful way. For example, social status (single, married, widowed, divorced) is a descriptive variable measured using a nominal scale. Similarly, the variable of gender (male, female) and the variable of nationality (Algerian, non-Algerian) also fall into this category.

In the case of variables measured on a nominal scale, codes can be assigned. For instance, the number 01 could represent male, and 02 could represent female.

3.1.2. Ordinal Scales.

This type of data consists of levels or categories that can be arranged in either ascending or descending order. For example, the variable of educational level has indicators such as (illiterate, primary, secondary, university). This descriptive variable is measured using an ordinal scale. Similarly, the descriptive

variable for student grades (pass, good, very good, excellent) can also be measured according to the ordinal scale.

3.2. Quantitative Data.

Quantitative data refers to data expressed in numerical values that represent the actual value of a phenomenon. This type of data is divided into two main categories:

3.2.1.Interval Data.

This consists of numerical data measured on an interval or range from zero, where zero indicates the presence of the phenomenon. For example, student grades are a quantitative variable measured on an interval scale, where the student's level is assessed based on how far their grade is from zero. The further the grade is from zero, the better the level. Additionally, receiving a grade of zero does not mean that the student has no level at all. To clarify, consider the variable of air temperature, which is also a quantitative variable measured on an interval scale. A temperature of zero does not mean the absence of the phenomenon.

3.2.2. Ratio Data.

This type of data consists of quantitative variables where a value of zero indicates the absence of the phenomenon, which contrasts with interval data. For example, the production of milk from dairy farms means that a value of zero signifies no milk production. Similarly, for cement production, a value of zero indicates no cement production. In the case of broadcasting matches on a certain day, a value of zero means no matches were broadcast on that day.

Note: Interval data cannot undergo arithmetic operations like multiplication or division because it reflects a specific phenomenon. In contrast, ratio data can undergo such operations because it represents absolute numerical values.

4. Data Collection.

Data collection in statistics is of great importance; accurate collection leads to correct results, and the reverse is also true. In this context, the most important methods for collecting data include (1) data sources, (2) data collection methods, (3) types of samples, and (4) means of data collection.

4.1. Data Sources.

There are two primary sources of data collection: (1) primary sources and (2) secondary sources.

4.1.1.Primary Sources.

These are sources from which the researcher obtains data directly. For example, when conducting a study on families, data can be collected through direct interviews with the head of the household, gathering information about monthly income, family size, educational level, and the area in which they live, among other things.

- **Advantages and Disadvantages**:
  - **Advantages**: Primary data is considered reliable and accurate because it is collected directly by the researcher.
  - **Disadvantages**: It can be costly in terms of time, effort, and resources.

4.1.2.Secondary Sources.

Secondary sources are those from which the researcher obtains data indirectly. This can include publications from ministries or departments, weather reports, or data collected through specific devices or from other individuals, among other sources.

- **Advantages and Disadvantages**:
  - **Advantages**: These sources save the researcher money, time, and effort.
  - **Disadvantages**: They are generally considered to be less reliable than primary sources.

5. Data Collection Methods.

There are two methods for collecting data in statistics:

1. **Census Method**
2. **Sampling Method.**

5.1. Census Method.

This method is characterized by comprehensive data collection, impartiality, and accuracy of results. However, it is criticized for being more costly, requiring more effort, and taking more time. This method is used for a complete enumeration of the population. For example, a complete census of all dairy factories or all tiger farms in the country, among others.

5.2. Sampling Method.

This method relies on taking a sample from the population for study without the need to examine all members of the population. Therefore, the results obtained from the sample can be generalized to the population. The sample is selected using scientific and appropriate methods to ensure accurate results. This method is characterized by:

1. Reducing time and effort.
2. Reducing costs.
3. Obtaining more preferred data.

However, it is criticized for being less accurate than the results obtained through the census method.

6. Types of Samples.

There is a distinction between the study population and the sample drawn from this population. Thus, the sample is divided into two main types:

1. **Population**: The study population includes all members of the study. For example, when studying the population of potato farmers, sheep breeding farms, or architecture students, among others.
2. **Sample**: A sample is a part of the population selected using various methods for the purpose of studying this population. The following Figure illustrate the difference between the population and the sample.

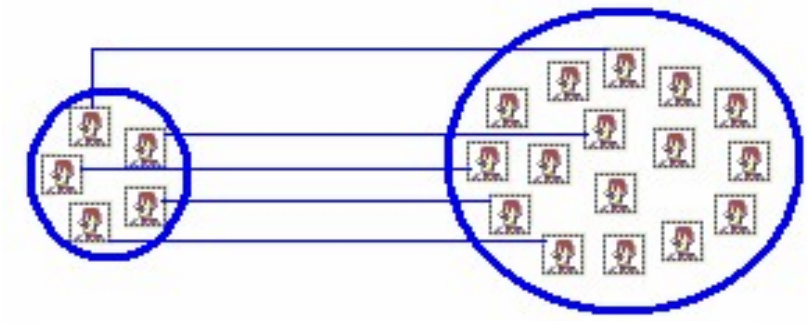


Figure 1-2 . Types of Samples

Samples can be divided based on the selection method into two types:

1. **Probability Samples**
2. **Non-Probability Samples**

6.1. Probability Samples.

These are samples selected according to probability rules, where the selection of its members is made randomly from the study population to avoid bias in the selection of items. The most important types of probability samples are:

- **Systematic Random Sample**: A sample where every nth individual is selected from a list.
- **Cluster Sample**: A sample where the population is divided into clusters (or groups), and entire clusters are randomly selected.
- **Simple Random Sample**: A sample where every member of the population has an equal chance of being selected.
- **Stratified Random Sample**: A sample that involves dividing the population into strata (groups) and randomly selecting samples from each stratum.

6.2. Non-Probability Samples.

These are samples where members are selected through non-random methods, allowing the researcher to choose items in a way that meets the sample's objectives. For example, a sample could be chosen from date palm farms that only produce a specific type of dates. Non-probability samples include types such as:

- **Purposive Sample**: A sample selected based on specific characteristics or criteria.
- **Quota Sample**: A sample where the researcher ensures equal representation of different subgroups.

The following Figure illustrates the types of samples in statistics5 (In Arabic).

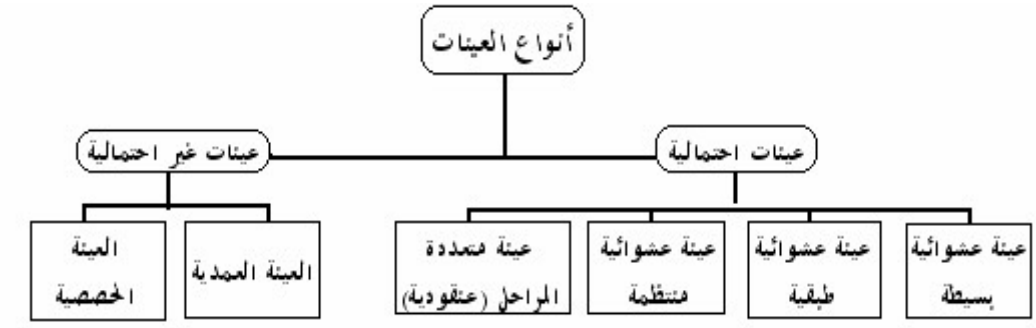


Figure 1-3 . Types of samples in statistics

7. Stages of Statistical Research.

Statistical research includes several stages, which can be summarized as follows:

1. Defining the Problem
2. Formulating Hypotheses
3. Data Collection
4. Tabulation
5. Data Classification
6. Statistical Description
7. Data Presentation

8. Statistical Analysis
9. Interpretation
10. Reporting.

- conclusion

The General Introduction of the Subject represents a fundamental and essential component of any academic course, lecture, or scientific document. It is not merely an introductory section, but rather a conceptual framework that establishes the foundation upon which all subsequent chapters or topics are built. By providing a clear overview of the subject, it enables learners to understand the general direction of the course before engaging with its detailed content.

This section plays a key role in defining the scope of the subject, identifying its main themes, and clarifying its objectives. It helps students and researchers recognize what the subject covers, why it is important, and how it contributes to a broader scientific or practical context. In doing so, it ensures that learners develop a coherent understanding of the subject's structure and purpose from the very beginning.

Moreover, the general introduction contributes to linking theoretical knowledge with real-world applications. By highlighting the relevance of the subject in various fields, it enhances the learner's motivation and interest. Whether in statistics, engineering, architecture, or social sciences, the introduction demonstrates how the subject can be applied to solve real problems, analyze phenomena, and support decision-making processes.

Another important function of the general introduction is to provide a roadmap of the course content. It outlines the sequence of topics and explains how each part contributes to the overall learning process. This logical organization helps learners build knowledge gradually and systematically, reducing confusion and improving comprehension.

In addition, the general introduction strengthens the continuity of learning by connecting previous knowledge with new concepts. It acts as a bridge between basic ideas and advanced topics, ensuring a smooth transition throughout the course. This structured approach is particularly important in scientific disciplines where concepts are often interrelated and cumulative.

Overall, the General Introduction of the Subject is a crucial element in academic writing and teaching. It enhances understanding, improves organization, and increases learner engagement. A well-written introduction not only prepares students for what they are about to study but also lays the groundwork for deeper analysis and critical thinking. For these reasons, it is considered an indispensable part of any well-structured educational or scientific work.

CHAPTER II:

METHODS OF PRESENTING STATISTICAL DATA

CHAPTER II Content:

• Introduction:

The stage of presenting data in the field of statistics is equally important as the stage of data collection. This is aimed at benefiting from the data in understanding the phenomenon under study, as well as assessing the central tendency and degree of homogeneity of the data. Statistical data are collected to provide meaningful information about a phenomenon, population, or process under study. However, raw data in their original form are often difficult to interpret, compare, or analyze. Therefore, statisticians organize and present data in a systematic manner that facilitates understanding, interpretation, and decision-making. The process of presenting statistical data is an essential step in statistical analysis because it transforms large volumes of numerical information into a clear and understandable format. The presentation of statistical data enables researchers, students, policymakers, and decision-makers to identify patterns, trends, relationships, and variations within a dataset. A well-designed presentation helps summarize complex information, highlights important findings, and improves communication between researchers and their audiences. Moreover, it allows readers to compare values, detect similarities and differences, and draw meaningful conclusions from the data. There are several methods for presenting statistical data, each serving a specific purpose depending on the nature of the information and the objectives of the analysis. Generally, statistical data can be presented through textual presentation, tabular presentation, and graphical presentation. Textual presentation describes data using written statements and is suitable for small datasets. Tabular presentation organizes data into rows and columns, providing a precise and systematic summary of numerical information. Graphical presentation uses visual tools such as bar charts, pie charts, histograms, frequency polygons, and line graphs to illustrate patterns, comparisons, and trends more effectively.

The choice of an appropriate presentation method depends on several factors, including the type of data, the size of the dataset, and the intended audience. In many cases, a combination of tables and graphs provides the most effective way of communicating statistical information. By presenting data clearly and accurately, researchers can enhance the quality of statistical analysis and facilitate a deeper understanding of the results. In summary, methods of presenting statistical data play a crucial role in transforming raw data into useful information. They improve data interpretation, support statistical analysis, and contribute to informed decision-making in various fields such as economics, business, engineering, social sciences, and scientific research.

- **Tabular Presentation of Data.**
- **Graphical Presentation of Data.**

Through the following illustration, we can demonstrate how to present statistical data using various statistical tools, including frequency tables and graphical displays.

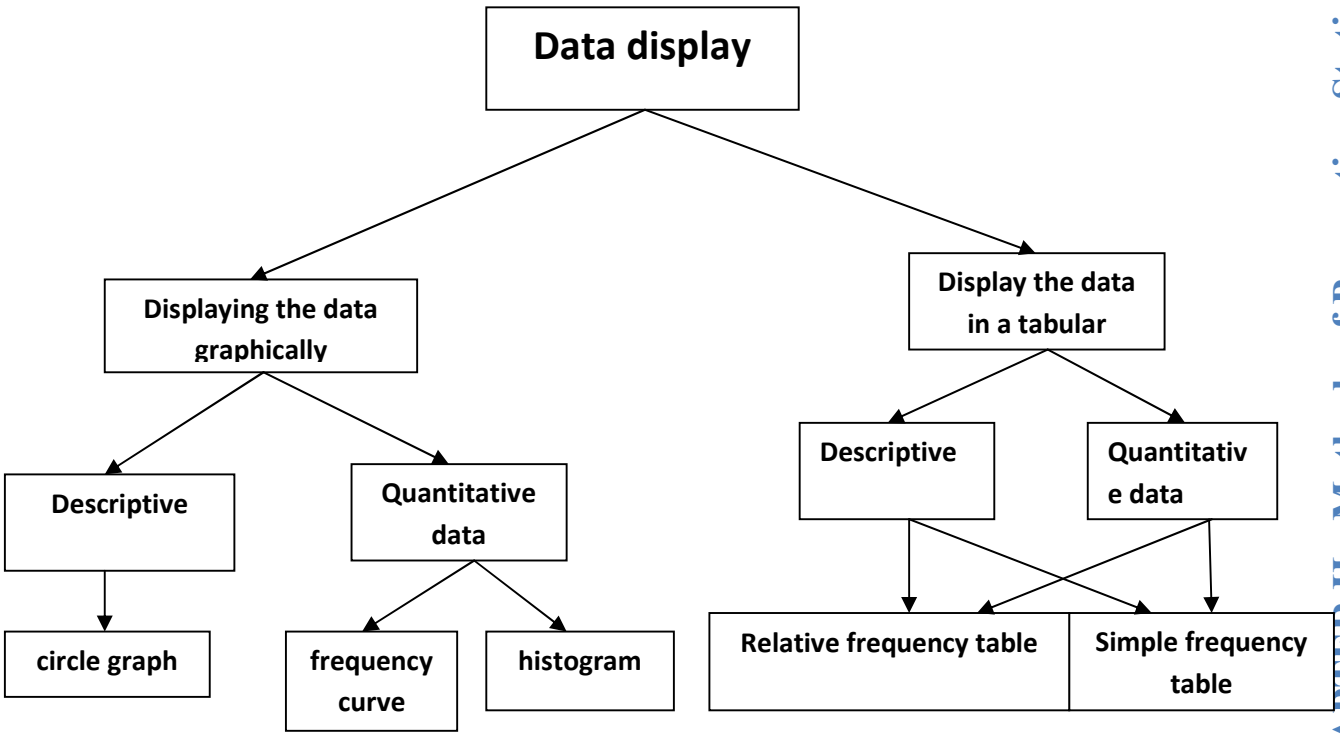


Figure 2-1 . Present Statistical Data Using Various Statistical tools

1. Tabular Presentation of Data.

Thus, data presentation can take the form of a frequency table, which can vary according to the type of data and the number of variables. In this regard, we can present below the data for a variable (categorical or quantitative) in the form of a simple frequency table.

1.1. Presenting of description Statistical Variables Datas in a Simple Frequency Table.

This table consists of two columns: one containing the levels (groups) of the variable and the other showing the frequency (counts) for each level (group). For example, if we have a study of a particular phenomenon that includes one categorical variable, we can present its data in the form of a simple frequency table. The relative frequency is calculated by dividing the frequency of each group by the total frequency, which means that:

$$\left(\frac{f}{\Sigma f}\right) = \frac{\textbf{Group frequency}}{\textbf{Total frequency}} = \textbf{relative frequency}$$

We can explain this with the following example:

Below are data on the educational levels of a sample of 50 individuals.

Elementary	Int-med	Ad-University	Secondary	Int-med	Secondary	Red-writes	Int-med
Int-med	Elementary	Secondary	Int-med	Secondary	Secondary	Int-med	Red-writes
Secondary	Red-writes	Elementary	Secondary	University	Red-writes	Secondary	Elementary
Int-med	University	Int-med	Elementary	Secondary	Int-med	Elementary	Int-med
Elementary	Secondary	Elementary	Red-writes	Secondary	Elementary	Int-med	Secondary
Secondary	Secondary	Ad-University	University	Elementary	University	Secondary	University
						Red-writes	Int-med

Required:

1. Present the data in the form of a frequency table.
2. Construct the relative frequency distribution, then comment on the results.

Solution.

- **First / Presenting the Data in the Form of a Frequency Table:**

The educational level (Reads and Writes - Primary - Intermediate - Secondary - University - Postgraduate) is an ordinal categorical variable. The above data can be presented in the form of a frequency table by following these steps:

- **Creating a Data Summary Table.**

Educational level	Statistical indicators	frequency
Reads and writes	/ 𝍸	6
Elementary	𝍸 𝍸	10
Intermediate	// 𝍸 𝍸	12
Secondary	/// 𝍸 𝍸	15
University	𝍸	5
Advanced to university	//	2
Sum		**50**

- **Constructing the Frequency Table.**

The frequency distribution for a sample of 50 individuals based on educational level.

Educational level	frequency	relative frequency
Reads and writes	6	0.12
Elementary	10	0.20
Intermediate	12	0.24
Secondary	15	0.30
University	5	0.10
Advanced to university	2	0.04
Sum	**50**	**1.00**

- **Constructing the Relative Frequency Distribution.**

By applying the previous equation, we can calculate the relative frequencies, and the third column in the previous table illustrates this distribution. From the relative distribution, it is observed that approximately 30% of the sample individuals have a secondary qualification, while the percentage of individuals with less than a secondary education (Intermediate, Primary, Reads and Writes) is more than 5%. The percentage of individuals holding a qualification higher than a university degree is about 4%, which is the lowest percentag.

- **Notes on the Table.**

When constructing a table to present data, the following should be considered:

1. Write a number for the table.
2. Provide a title for the table.
3. Each column in the table should have a title that indicates its content.
4. The source of the data should be included in the table.

1.2. Presenting Quantitative statistical Variable Data in a Simple Frequency Table.

This table consists of two columns: the first contains ascending categories for the readings taken by the variable, and the second includes the frequencies or the counts of observations that belong to the appropriate category. Thus, we can present quantitative variable data in the form of a simple frequency table. In the following example, we will explain how to present quantitative data graphically.

Example:

Below are the scores of 70 students in the final exam for the course in Applied Statistics.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

Required:

1. Construct the frequency distribution of the students' scores.
2. Construct the relative frequency distribution.
3. What is the percentage of students who scored between 70 and less than 80?
4. What is the percentage of students who scored less than 70?
5. What is the percentage of students who scored 80 or more?

Solution:

- **First / Constructing the Frequency Distribution:** The student's score in the exam is a continuous quantitative variable. To organize the data in the form of a frequency table, the following steps should be followed:

- **Range calculation** (**R**).

$$\textbf{Range = Maximum –Minimum}$$

$$\mathbf{R = 94 - 55 = 39}$$

- **Determining the number of categories** (**C**) :

Determining the number of categories is based on several considerations, including the researcher's opinion, the purpose of the study, and the size of the data. Many researchers believe that the optimal number of categories should range between 5 and 15. Assuming the number of categories is 8, we have (C = 8).

- **Calculating category length** (**L**) :

$$\mathbf{L} = \frac{Range}{Classes} = \frac{R}{C} = \frac{39}{8} = 4.875 \approx 5$$

- **Determining the Categories.**

Each category starts with a value called the lower limit and ends with a value called the upper limit. Thus, we find that:

- The lower limit of the first category is the lowest reading (score), which means the lower limit of the first category = 55.

The upper limit of the first category = lower limit + class width = L + 55 = 5 + 55 = 60.

Therefore, the first category is "55 to less than 60," read as "from 55 to less than 60."

- The lower limit of the second category = the upper limit of the first category = 60.

The upper limit of the second category = lower limit + class width = 5 + 60 = 65.

Therefore, the second category is "60 to less than 65," read as "from 60 to less than 65."

- In the same way, the boundaries of the other categories are determined as follows:

category Third :**65 to lessthan 70** **Category Four**:**70 to lessthan 75**

category Fifth :**75 to lessthan 80** **Category Six**:**80 to lessthan 85**

Category Seven :**85 to lessthan 90** **Category Eighth** :**90 to lessthan 95**

- Categories can be written in different forms, as shown in the data summary table:

Creating a Data Summary Table: Data Summary Table

Different Forms of category			Statistical indicators	frequency
Category	**Category**	**Category**		
55 to lessthan 60	55- 60	55-	𝍸 𝍸	10
60 to lessthan 65	60-65	60-	𝍸 𝍸 //	12
65 to lessthan 70	65-70	65-	𝍸 𝍸 ///	13
70 to lessthan 75	70-75	70-	𝍸 𝍸 𝍸 /	16
75 to lessthan 80	75-80	75-	𝍸 𝍸	10
80 to less than 85	80-85	80-	////	4
85 to lessthan 90	85-90	85-	///	3
90 to lessthan 95	90-95	90-	//	2
Sum				**70**

• Creating the frequency table.

Frequency distribution of 70 students according to their scores in the statistics exam.

Degree	Number of students frequency (f)	Relative Frequency
55 – 60	10	0.143
60 – 65	12	0.171
65 – 70	13	0.186
70 – 75	16	0.229
75 – 80	10	0.143
80 – 85	4	0.057
85 – 90	3	0.043
90 - 95	2	0.028
Sum	**70**	**1.00**

Relative frequency distribution

f/n = Relative frequency

Explanation.

* The percentage of students scoring between 70 and less than 80 is the sum of the relative frequencies for categories four and five:

0.372 = 0.143 + 0.229 = Percentage of students scoring between (70, 80)

That is, approximately 37.2% of students scored between (70, 80).

* The percentage of students who scored less than 70 is the sum of the relative frequencies of the first, second and third categories:

0.5 = 0.186 + 0.171 + 0.143 = the percentage of students who scored less than 70, meaning that approximately 50% of the students scored less than 70.

* The percentage of students who scored 80 or higher is the sum of the relative frequencies of the last three categories:

0.128 = 0.028 + 0.043 + 0.057 = the percentage of students who scored 80 or higher, meaning that approximately 12.8% of students scored 80 or higher.

2- Quantitative Statistical Datas Presentation.

Graphical presentation of data is one of the methods that can be used to describe data, in terms of the shape of the distribution and the duration of data concentration. In many practical aspects, graphical presentation is easier and faster in describing the phenomenon under study. The methods of graphically presenting data vary according to the type of data grouped in the form of a frequency table. Below is a presentation of the different graphical forms.

2-1 Histogram.

A histogram is a graphical representation of a simple frequency table for continuous quantitative data. It consists of adjacent bar graphs, where frequencies are represented on the vertical axis, while the values of the variable (category boundaries) are represented on the horizontal axis. Each category is represented by a bar, the height of which is the frequency of the category, and the length of its base is the length of the category.

For Example.

The following is the frequency distribution of weights in grams of a 100-gram sample of poultry selected from a farm after 45 days.

the weight	600–	620–	640–	660–	680–	700–720	Sum
Number of chickens	10	15	20	25	20	10	**100**

The Task Is:

1. What is the categroy size?

2. Draw the histogram.

3. Draw a histogram for the ratio and then comment on the graph.

Solution.

1. Length of category (L).

$$L = upper - Lower$$

$$L = 620 - 600 = 640 - 620 = \ldots = 720 - 700 = 20$$

$$20 =$$

2. Drawing the Histogram:

To draw a histogram, follow these steps:

Draw two perpendicular axes: the vertical axis represents the frequencies, and the horizontal axis represents the weights.

- Each class is represented by a bar whose height is the frequency of that class, and whose base is the class width.

• Each bar starts where the previous class bar ended.

The figure shows the histogram of chicken weights.

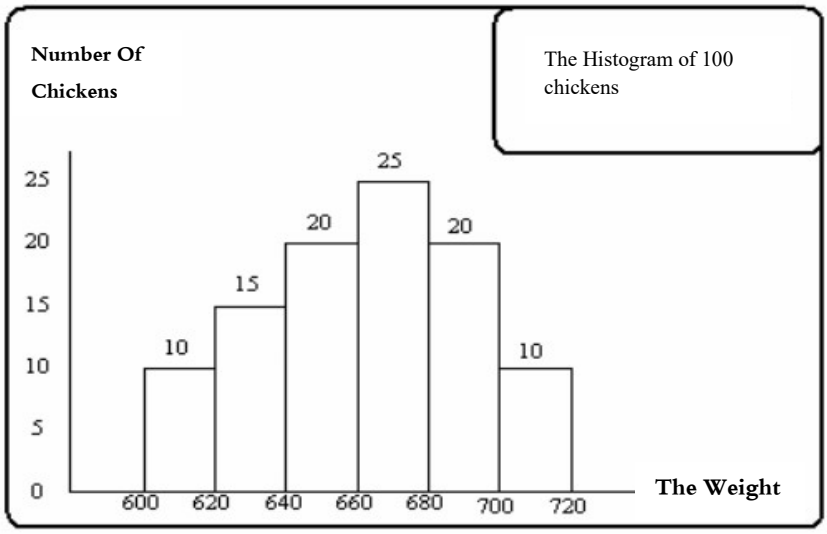


Figure 2-2 . The histogram shows the weights of a sample of 100 chickens

3. Drawing the relative frequency histogram:

Calculating the relative frequencies:

The Weight	600–	620–	640–	660–	680–	700–720	**Sum**
Number Of Chickens	10	15	20	25	20	10	**100**
Relative Frequency	0.10	0.15	0.20	0.25	0.20	0.10	**1.00**

Following the same steps as before when drawing the frequency graph, the relative frequency histogram is drawn, replacing the absolute frequencies with relative frequencies on the vertical axis, as shown in the following figure:

Figure 03 Shows The Relative Frequency Distribution Of Weights For A Sample Of 100 Chickens.

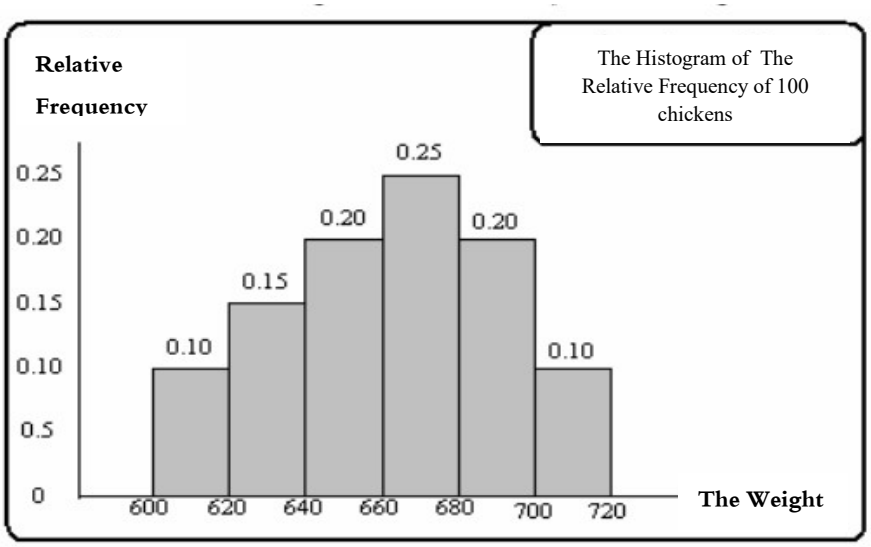


Figure 2-3 . The Relative Frequency Distribution of samples of 100 chickens

From the figure above, the following can be observed:

• 25% of the chickens weigh between 660 and 680 grams, which is the largest percentage.

* Notes on the Histogram Shape

a. The area under the histogram is equal to the sum of the frequencies (n).

b. The area under the relative histogram represents the sum of the relative frequencies and is equal to one.

c. Common values can be estimated; these are the values with the highest elevation. In the two previous figures, the common weight falls within the range (660-680) and is called the mode.

2-2- Frequency Curve.

A frequency curve is also a graphical representation of a simple frequency table, where frequencies are represented on the vertical axis, class midpoints on the horizontal axis, and then the coordinates are connected by curved lines. The ends of the curve are then connected to the horizontal axis.

The class midpoint is the value located in the middle of the class and is calculated using the following equation:

$$Midpoint = \frac{Lower + Upper}{2}$$

By following the previous steps, the frequency curve can be drawn. The lines can be smoothed into a curve so that it passes through the largest number of points. The frequency curve can be drawn using the data from the previous example.

Example:

Use the frequency table data in the example (poultry sample) to plot the frequency curve.

Solution:

To plot the frequency curve, follow these steps:

• Calculate the class midpoints by applying the previous equation.

The Weight	Number Of Chickens (Frequency).	Category Center (x)
600–	10	(600+620)/2= 610
620–	15	(620+640)/2= 630
640–	20	650
660–	25	670
680–	20	690
700– 720	10	(700+720)/2= 710
Total	**100**	

• The coordinate points are:

Category Center–(x)	590	610	630	650	670	690	710	730
Frequency (y)	0	10	15	20	25	20	10	0

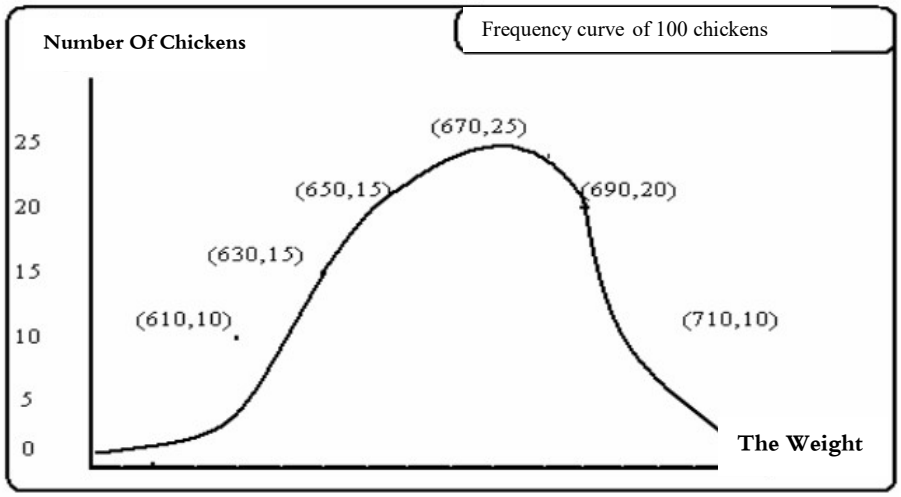


Figure 2-4 . Frequency curve for the weights of a sample of 100 chickens

The relative frequency curve can also be drawn by representing the relative frequencies on the vertical axis instead of the absolute frequencies, and then this curve takes the following form:

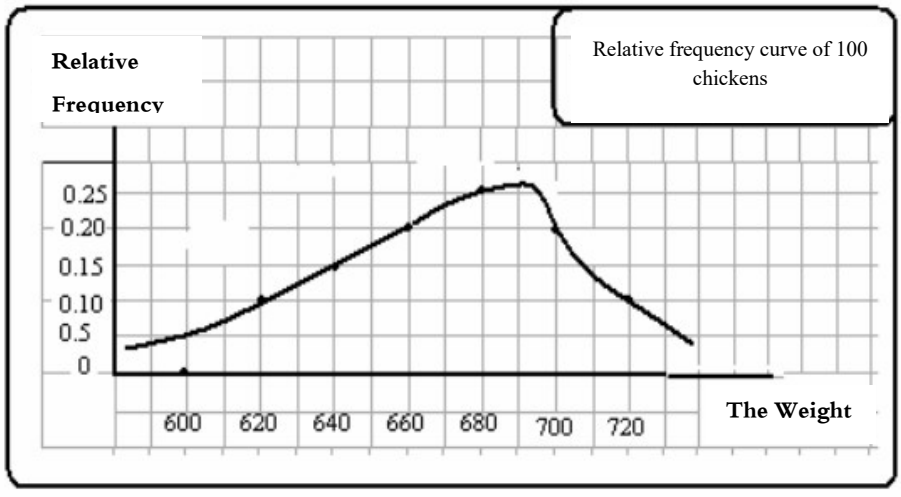


Figure 2-5 . Relative frequency curve for the weights of a sample of 100 chickens

2-3- Cumulative Frequency Distributions.

Cumulative frequency distributions are used to determine the number of observations that are less than or greater than a certain value. Researchers then construct ascending or descending cumulative tables. The following explains how to construct each of these types:

- Ascending Cumulative Frequency Distribution

To construct an ascending cumulative frequency table, the sum of frequencies (number of values) that are less than each class boundary is calculated.

Example

The following frequency table shows the distribution of 40 cows on a farm according to the amount of milk produced by the cow per day in liters.

Quantity Of Milk	18–	22–	26–	30–	34–38	**Total**
Number of cows	4	9	15	8	4	**40**

The task is to:

1. Construct the ascending cumulative frequency distribution.
2. Construct the relative ascending cumulative frequency distribution.
3. Construct the descending cumulative frequency distribution.

Solution.

1 And 2 -Constructing the ascending cumulative frequency distribution:

- Relatively ascending cumulative frequency and Ascending cumulative frequency distribution. As we see in the following tables.

Production quantity in liters	Number of cows
18-	4
22-	9
26-	15
30-	8
34-38	4
Total	**40**

Less than	Ascending cumulative frequency	Relatively ascending cumulative frequency
Less than18	0	0.00
Less than22	4	0.10
Less than26	13	0.325
Less than 30	28	0.70
Less than34	36	0.90
Less than38	40	1.00

3.. Construct the descending cumulative frequency distribution .

As we see in the following tables.

frequency distribution

Production quantity in liters	Number of cows
18-	4
22-	9
26-	15
30-	8
34-38	4
Sum	**40**

Cumulative frequency descending

More than or equal	Cumulative frequency descending	Relatively Cumulative frequency descending
More than or equal 18	40	1.00
More than or equal 22	36	0.90
More than or equal 26	27	0.675
More than or equal 30	12	0.30
More than or equal 34	4	0.10
More than or equal 38	0	0.00

3- The Graphical presentation of Descriptive datas.

The Variable of Descriptive data can be displayed in the form of a pie chart or bar chart, through which the groups or levels of this variable can be described and compared.

3-1- Pie Chart

To display the data of the descriptive variable in the form of a pie chart, a 360° pie chart is used, based on the relative frequency of the variable's groups. The angle corresponding to group r is determined by applying the following equation:

Example

The following frequency table shows the distribution of a sample size of 500 households according to the region they belong to.

Area	Riyadh	alsharqia	Al-Qassim	algharbia	**Total**
Number of families	150	130	50	170	**500**

Represent the data above in the form of a pie chart.

Solution

1. Determine the magnitude of the specialized angle for each region, applying the equation:

The relative frequency of a set × 360° = The Magnitude Of The Angle

Area	Number of families	Relative frequency	Angle
Riyadh	150	0.30	360 × 0.30 = 108°
alsharqia	130	0.26	360 × 0.26 = 93.6°
Al-Qassim	50	0.10	360 × 0.10 = 36°
algharbia	170	0.34	360 × 0.34 = 122.4°
Total	**500**	**1.00**	**360°**

2. Drawing the circle

From the previous table, we draw the circle and divide it into four parts, each area having a part proportional to the angle assigned to it, as shown in the following figure:

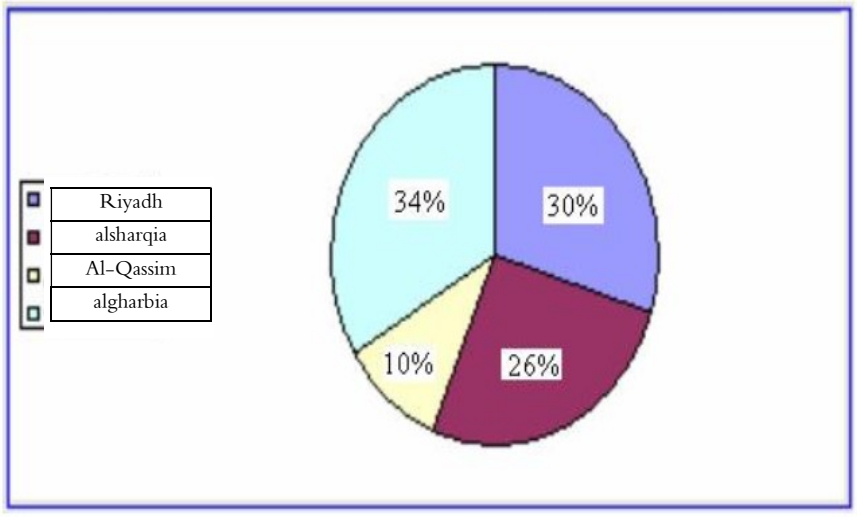


Figure 2-6 . Pie chart for a sample size of 500 households distributed by region

From the previous figure, we can observe that the percentage of families belonging to the western region is around 34%, which is the largest percentage in the sample, while the percentage of families in the sample is around 10%, which is the lowest percentage in the sample.

Finally, we can give other examples of methods used in statistical data.

1. Tabular Presentation (Statistical Table)

TEMPO	f_i
[48, 51[	5
[51, 54[	6
[54, 57[	4
[57, 60]	5
TOTAL	20

2. Bar Chart (Bar Diagram)

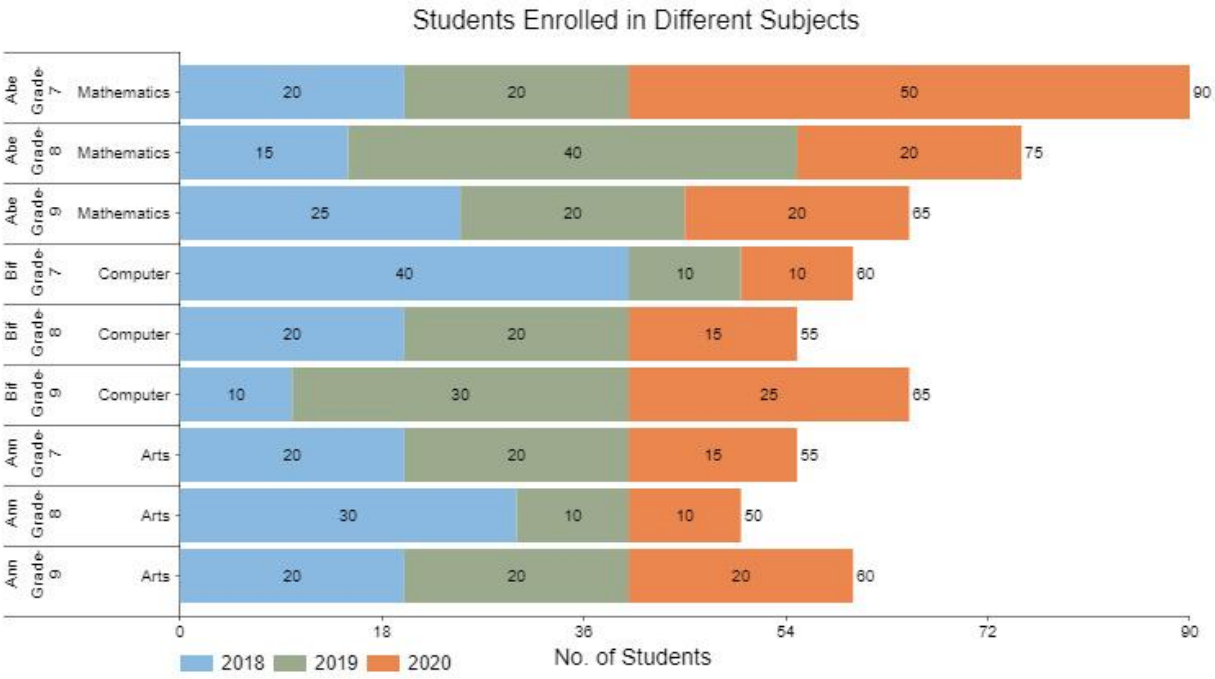


Figure 2-7 . Bar Chart

3. Pie Chart.

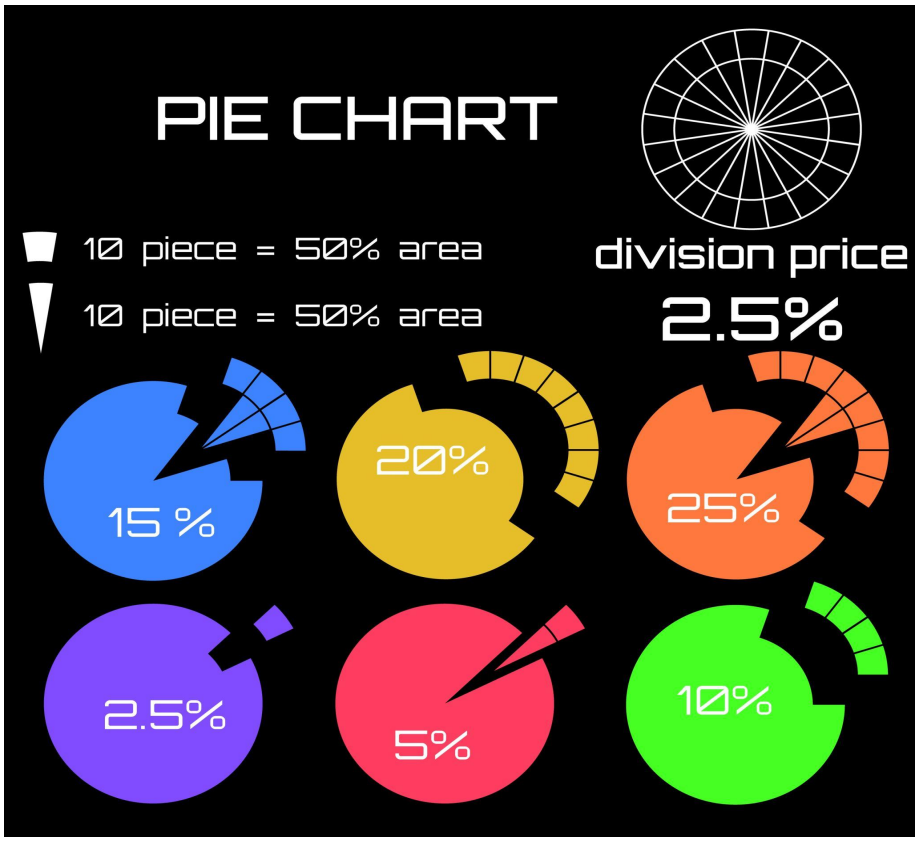


Figure 2-8 . Pie Chart

4. Histogram.

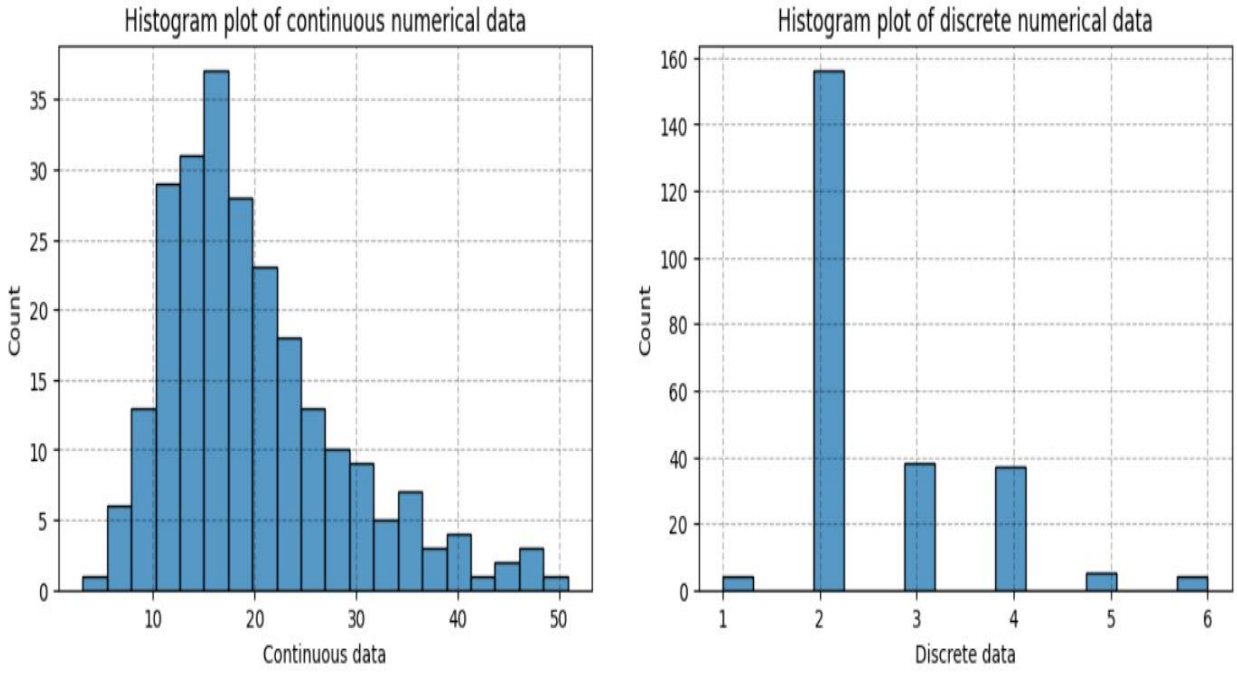


Figure 2-9 . Histogram.

4. Line Graph.

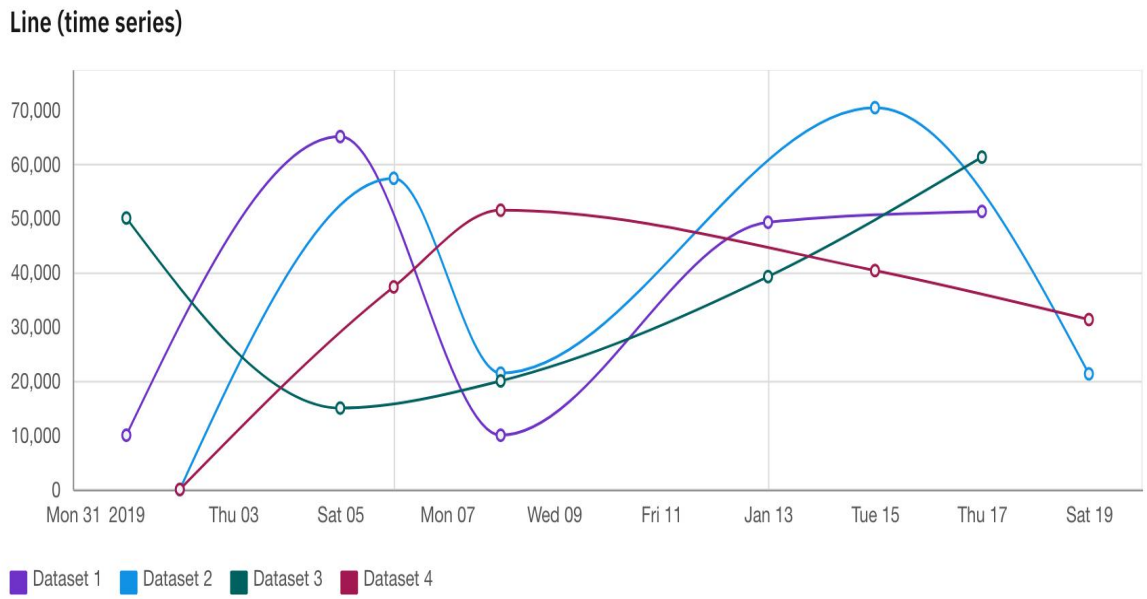


Figure 2-10 . Line Graph.

5. Frequency Polygon.

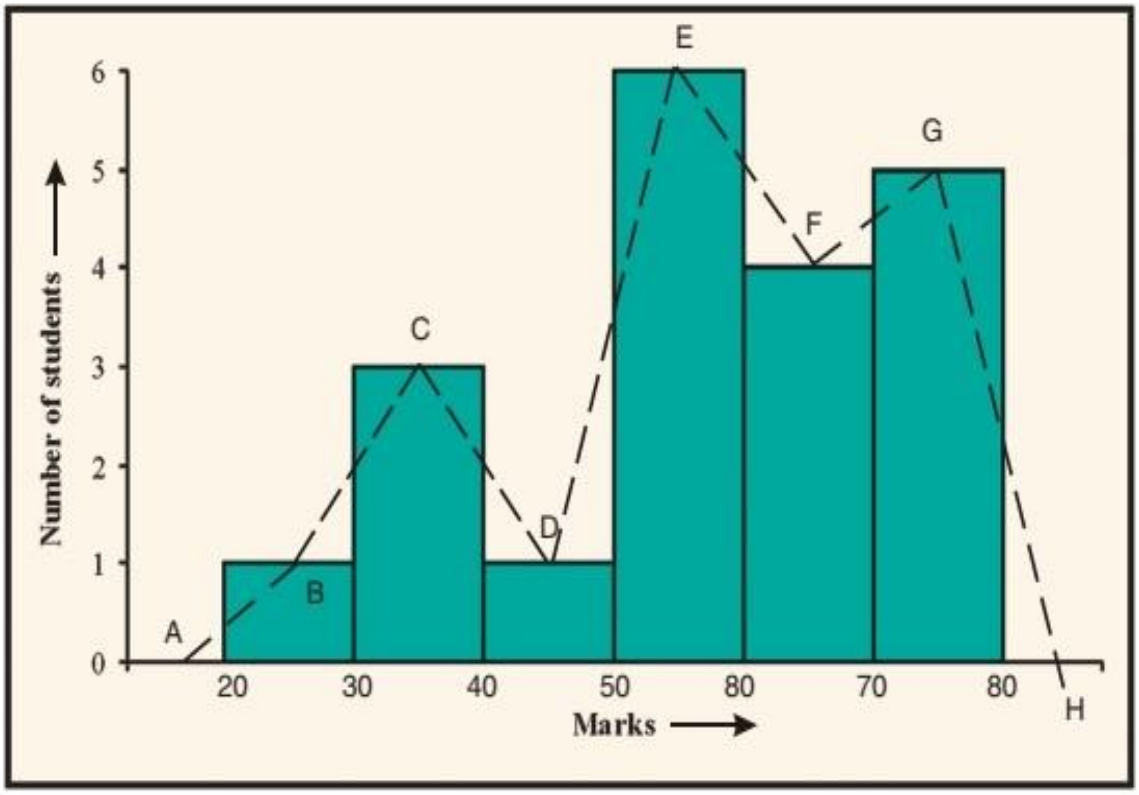


Figure 2-11 . Frequency Polygon.

- **conclusion**

Statistical data presentation is a fundamental stage in the process of data analysis, as it transforms raw numerical information into an organized and interpretable format. After data collection, researchers are often faced with large and complex datasets that are difficult to understand in their original form. Therefore, appropriate methods of presentation are used to simplify, summarize, and communicate the essential features of the data clearly and effectively.

In general, there are three main methods of presenting statistical data: textual presentation, tabular presentation, and graphical presentation**.** Each method serves a specific purpose depending on the nature of the data and the objective of the analysis.

Textual presentation involves describing data in written form. It is typically used for small datasets or when only key values and findings need to be highlighted. Although simple, it becomes less efficient when dealing with large or complex datasets.

Tabular presentation organizes data into rows and columns, allowing for systematic arrangement and easy comparison of values. Statistical tables, such as frequency distribution tables, are widely used because they provide a clear and precise summary of data while preserving numerical accuracy. Tables are especially useful for detailed analysis and for preparing data for further statistical calculations.

Graphical presentation uses visual tools to represent data, such as bar charts, pie charts, histograms, line graphs, and frequency polygons. This method is particularly effective for identifying patterns, trends, and relationships within the

data. Graphical methods make statistical information more accessible and easier to interpret, even for non-specialists. They are widely used in scientific research, reports, presentations, and decision-making processes.

Each method has its advantages and limitations. While textual presentation is simple and descriptive, it lacks efficiency for large datasets. Tabular presentation offers precision and structure but may require careful interpretation. Graphical presentation enhances clarity and visual understanding but may sometimes oversimplify detailed information.

In practice, these methods are often used together to provide a comprehensive view of the data. Tables may be used to present exact values, while graphs are used to illustrate trends and comparisons. The choice of presentation method depends on the type of data, the purpose of analysis, and the intended audience.

In conclusion, methods of presenting statistical data play a crucial role in statistical analysis by improving clarity, enhancing interpretation, and supporting effective communication of information. They are essential tools in fields such as economics, engineering, social sciences, and scientific research, where data-driven decision-making is required.

CHAPTER III

Measures of Central Tendency

CHAPTER III Content

- **Introduction.**

1- **Arithmetic Mean**

1-1- Calculating the Arithmetic Mean of Ungrouped Data.

1-2- Calculating the Arithmetic Mean of Grouped Data.

2- **Median**

2-1 Median of Ungrouped Data

2-2 Median of Grouped Data.

3- Mode

3.-1-Calculating the mode for ungrouped data

3-2- Calculating the mode for grouped data (Differences method)

- **Conclusion**

- **Introduction.**

This lesson presents some statistical measures that can be used to identify the characteristics of the phenomenon under study, as well as to compare two or more phenomena. These measures help determine the homogeneity of values within a variable and whether there are any outliers. In many practical applications, researchers need to calculate indicators that can be used to describe the phenomenon, specifically the value that is the midpoint or the value towards which the values tend. This highlights the importance of measures of central tendency, as relying solely on graphical representation is insufficient. Among the most important of these measures are measures of central tendency and dispersion. Measures of central tendency are also called measures of position or averages. They represent the values around which the values tend to cluster. These measures include the arithmetic mean, mode, median, geometric mean, harmonic mean, quartiles, and constants. The following is a presentation of the most important of these measures.

1. Arithmetic Mean

The arithmetic mean is one of the most important and widely used measures of central tendency in various applications. This measure can be calculated for both grouped and ungrouped data. The arithmetic mean has advantages and disadvantages, which are as follows:

- Advantages of the Arithmetic Mean:
- It is easy to calculate.
- It takes all values into account.
- It is the most widely used and understood measure.

- Disadvantages of the Arithmetic Mean:
- It is affected by outliers and extreme values.
- It is difficult to calculate in the case of descriptive data.
- It is difficult to calculate in the case of open frequency tables.

1.1. Calculating the Arithmetic Mean for Ungrouped Data.

The arithmetic mean can be generally defined as the sum of the values divided by the number of values. For example, if we have n values, denoted as x1, x2, ..., xn,

then the arithmetic mean of these values is denoted by $\bar{x}$ and is calculated using the following equation:

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Where the symbol $\sum$ represents the sum.

- **Example**

The following are the scores of 8 students in statistics:

34	32	42	37	35	40	36	40

The task is to find the arithmetic mean of the students' scores on the exam.

- **Solution**

To find the arithmetic mean of the scores, apply the above equation as follows:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

That is, the average score of the students in the statistics test is 37.

1.2. Calculating the Arithmetic Mean of Grouped Data.

The arithmetic mean is calculated using the following equation:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

If k is the number of groups and x1, x2, ……, xk are the midpoints of these groups, and f1, f2, ……, fk are the frequencies, and it is known that the original values cannot be determined from the frequency distribution table, then each value within a group is represented by the midpoint of that group. Therefore, the midpoint of the group is considered the estimated value of each item within that group.

-Example.

The following table shows the distribution of 40 students according to their weights.

Weight Categories	32-34	34-36	36-38	38-40	40-4	42-44
Number of Students	4	7	13	10	5	1

The task is to find the arithmetic mean.

- **Solution**

To calculate the arithmetic mean, we use the previous equation and follow these steps:

- Find the sum of the frequencies ∑/f.
- Calculate the categories indices x.
- Multiply the categories indices by the corresponding frequency (xf), and calculate the sum ∑.

Calculate the arithmetic mean by applying the previous equation.

Weight Categories (C)	Frequencies f	Category Centers x	xf
32-34	4	33=2 ÷ (34+32)	132 = 33 × 4
34-36	7	35	245 = 35 × 7
36-38	13	37	481= 37 × 13
38-40	10	39	390= 39 × 10
40-42	5	41	205= 41 × 5
42-44	1	43	43 = 43 × 1
Total	**40**		**1496**

SO the arithmetic mean of the students' weights is:

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{1496}{40} = 37.4 \ k.g$$

That is, the average weight of the students is 37.4 kg.

2. **Median**

The median is considered one of the most important measures of central tendency. This measure takes into account the order of values and defines the median as the value below which half the values (n/2) are less, and above which the other half (n/2) are greater; that is, 50% of the values are greater than it. The median measure has advantages and disadvantages, including:

- **Advantages of the median:**

1) It is not affected by outliers or extreme values.

2) It is easy to calculate.

3)The sum of the absolute deviations from the median is less than the sum of the absolute deviations from any other values, i.e., that:

$$\sum |x - Med| \leq \sum |x - a| \ , \ a \neq Med$$

- **Disadvantages of the median:**

1) It does not consider all values when calculating it; it relies on only one or two values.

2) It is difficult to calculate with metadata measured by a nominal criterion.

We can review below how to calculate the median for ungrouped and grouped data.

2.1. Median for Ungrouped Data.

In this context, to calculate the median for ungrouped data, we follow these steps:

- calculate the values in ascending order.
- Determine the median's rank, which is: Median rank = ((n+1)/2

If the number of values (n) is odd, then the median is:

$$\left(\frac{n+1}{2}\right) \text{ Value} = \text{Med}$$

The number of values (n) is even, then the median lies between the value (n/2) and the value ((n/2)+1), and then the median is calculated by applying the following equation.

$$\frac{\left(\frac{n}{2}+1\right) \text{ Value} + \left(\frac{n}{2}\right) \text{ Value}}{2} = \text{Med}$$

2.2. Median for Grouped Data

To calculate the median from grouped data in a frequency distribution table, the following steps are followed:

- Construct the ascending cumulative frequency distribution table.
- Determine the median rank: (n/2) = ((∑/f)/2)
- Determine the median class as shown in the following figure:

Previous ascending cumulative frequency f1: Lower boundary of median class (A)

Mediator rank (n/2): ⟶ Median Med

Subsequent ascending cumulative frequency f2: Upper boundary of median class

• The median is calculated by applying the equation:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L$$

Where:

L is the class length of the median, and is calculated using the following equation:

$$L = Upper - Lower$$

- **Example**

The following is the distribution of 50 medium-sized calves, according to their daily dry feed requirements in kilograms.

Daily Categories	1.5 -	4.5 -	7.5 -	10.5 -	13.5 – 16.5
Number of Calves f	4	12	19	10	5

The task is to calculate the median:

a. mathematically.

b. graphically.

- **Solution**

- Calculating the median mathematically

The Median rank: **n/(2) = (∑/f)/2 = 50/2 = 25**

Ascending cumulative frequency table:

Less than	Ascending cumulative repetition
1.5	0
4.5	4
A 7.5	f_1 16
Med (الوسيط)	25 — **The Median rank**
10.5	f_2 35
13.5	45
16.5	50

- Determining the median class:

This is the class that represents the median value, and it is a value less than (n/2) of the values. It can be determined by identifying the two ascending cumulative frequencies between which the median rank (n/2) lies. In the table above, we find that the median rank (25) lies between the two cumulative frequencies (35-16). The lower limit of the median class corresponds to the preceding ascending cumulative frequency of 7.5, and the upper limit of the median class corresponds to the following ascending cumulative frequency of 10.5. Therefore, the median class is (10.5-7.5).

• Applying the median formula to this example, we find that:

$$\mathbf{A= 7.5,\ f_1= 16,\ f_2= 35,\ L= 10.5 - 7.5 = 3}$$

The Median Value Is:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{25-16}{35-16} \times 3$$

$$= 7.5 + \frac{9}{19} \times 3 = 7.5 + \frac{27}{19} = 7.5 + 1.421 = 8.921 \ k.g$$

2.2.1. Calculating the median graphically.

• Graphically representing the ascending cumulative frequency distribution table

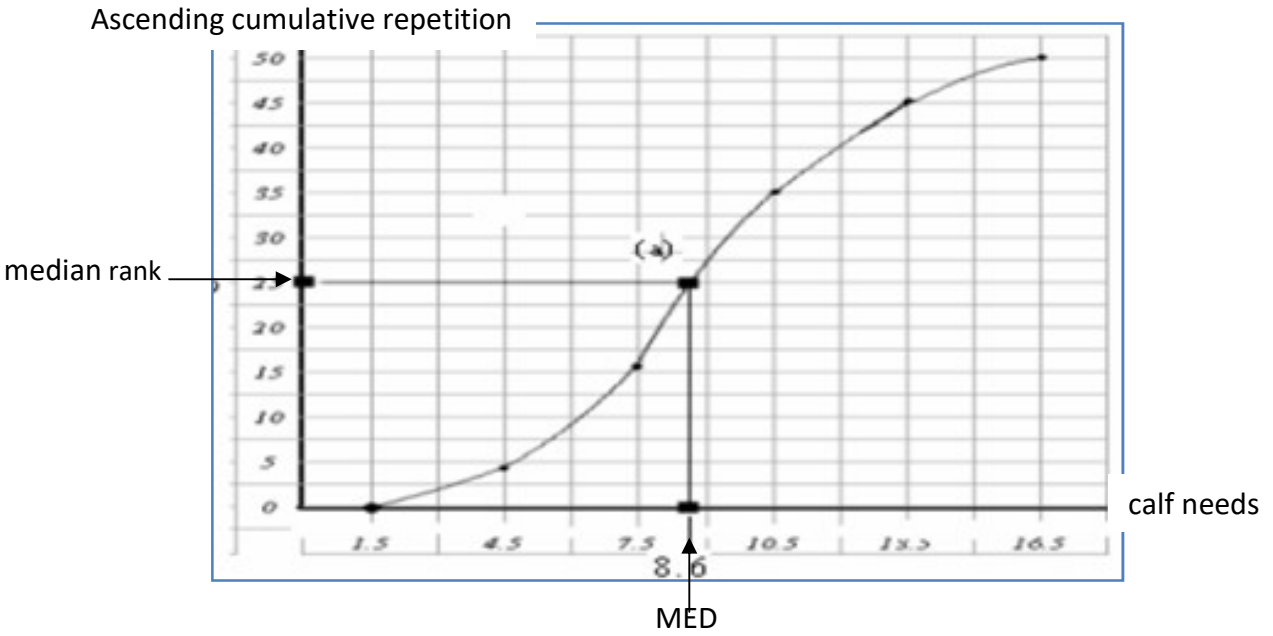


Figure 3-1 . Graphically representing the ascending cumulative frequency

• Determine the median (25) on the ascending cumulative frequency curve, then draw a horizontal straight line until the curve intersects at point (a).

• Project a vertical line from point (a) onto the horizontal axis.

• The point where the vertical line intersects the horizontal axis gives the median.

• The median, as shown in the figure, is Med = 8.6.

3. Mode

The mode is defined as the most common or frequently occurring value. It is commonly used in descriptive data to identify the prevalent pattern (level). It can be calculated for both grouped and ungrouped data as follows:

3.1. Calculating the mode for ungrouped data.

MOD = Most frequently occurring value

3.2. Calculating the mode in the case of grouped data (differences method).

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L$$

Where:

A: The minimum category of the mode (the category corresponding to the highest frequency)

d1: The first difference = (frequency of the category of mode – previous frequency)

d2: The second difference = (frequency of the category of mode – subsequent frequency)

L: The length of the category of mode.

Category Mod = category corresponding to the highest frequency

d1 = (Frequency of category mode – Previous frequency)

d2 = (Frequency of category mode – Next frequency)

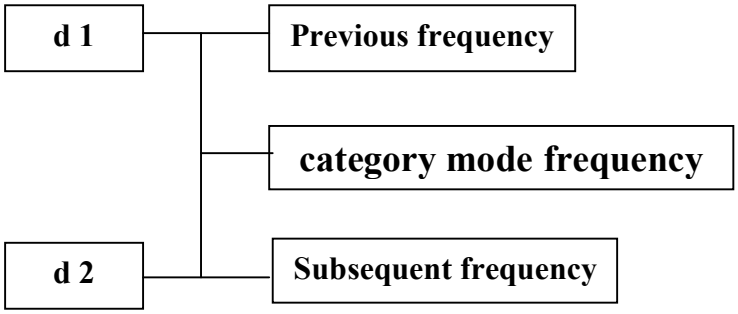


Figure 3-2 . Explanation of **d1** . **d2**

Example 01.

Random samples of students from some departments within the College of Food and Agricultural Sciences were tested, and their grades in Applied Statistics were recorded. The results were as follows:

Department of Plant Protection	80	77	75	77	77	77	65	70	58	67
Department of Food Science	88	68	60	75	93	65	77	85	95	90
Department of Economics	80	65	69	80	65	88	76	65	86	80
Department of Animal Production	85	73	69	85	73	69	69	73	72	85

The task is to calculate the mode of scores for each section.

- **Solution:**

- This data is ungrouped, therefore: Mode = Most frequently occurring value

The following table shows the grade pattern for each section:

Department	frequently value	mod value
Department of Plant Protection	77 was repeated 4	mod77 =
Department of Food Science	All values are unique	NO mod
Department of Economics	3 was repeated 65 3 was repeated 80	2 mod: 1 mod65 = 2 mod80 =
Department of Animal Production	3 was repeated 69 3 was repeated 73 3 was repeated 85	3 mod: 1 mod69 = 2 mod73 = 3 mod85 =

Example 02.

The following is a breakdown of 30 families according to their monthly consumption expenditure in thousands of dinars.

Spending categories	2 -	5 -	8-	11 -	14 - 17
Number of families f	4	7	10	5	4

The task is to calculate the monthly spending pattern of the family, using the differences method.

-Solution

To calculate the mode of this data, the previous equation is used, and the following steps are followed:

• Determine the category mod.

The Categories of mod is the class corresponding to the highest frequency.

Categories	frequency
2 -	4
5 -	7
8 -	10
11 -	5
14 - 17	4

category mode

A= 8

d_1=10−7= 3

d_2=10−5= 5

• Calculate the differences of **d**:

where

$$d_1 = (10-7) = 3 \qquad d_2 = (10-5) = 5$$

• Determine the lowest Categorie of the mod **(A=8)** and the Categorie width **(L=3)**.

• Applying the formula for calculating the mod to grouped data, we find that

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L$$

$$= 8 + \frac{3}{3+5} \times 3 = 8 + 1.125 = 9.125$$

- **Determining the data distribution pattern using measures of central tendency.**

The arithmetic mean, median, and mode can be used to describe the frequency curve, which expresses the shape of the data distribution as follows:

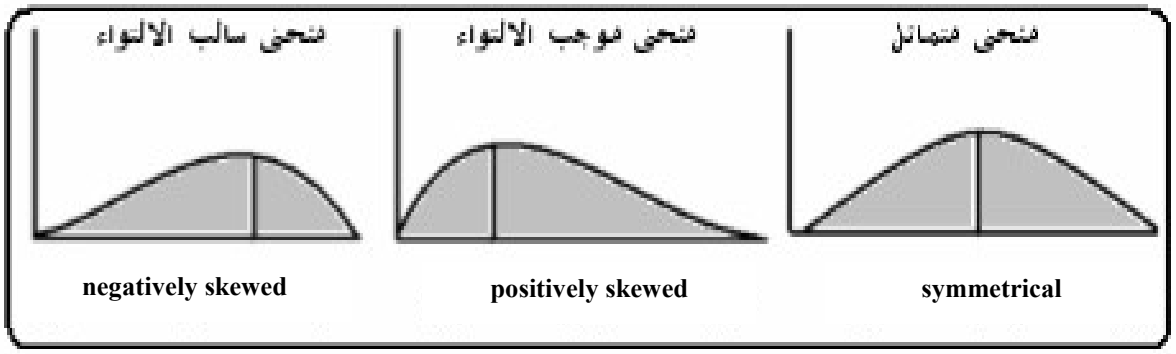


Figure 3-3 . The shape of the data distribution.

- A curve is symmetrical if: mean = median = mode.
- A curve is positively skewed (skewed to the right) if: mean < median < mode.
- A curve is negatively skewed (skewed to the left) if: mean > median > mode.

-Example with grouped data

The following frequency table shows the distribution of 100 farm workers by daily wage in dinars.

Salary	50 -	70 -	90 -	110 -	130 -	150 -	170 - 190
Number of Workers	8	15	28	20	15	8	6

The task is to: Calculate the mean, median, and mode.

Describe the wage distribution pattern on this farm.

-Solution.

-First: Calculate the arithmetic mean.

Salary Classes	Frequencies (f)	Class Midpoints (x)	F x
50 – 70	8	60	480
70 – 90	15	80	1200
90 – 110	28	100	2800
110 – 130	20	120	2400
130 – 150	15	140	2100
150 – 170	8	160	1280
170 - 190	6	180	1080
Total	100		11340

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{11340}{100} = 113.4 \ R.S$$

-Second: Calculation of the median (Med).

The rank of the median is (n/2 = 100/2 = 50).

Construction of the cumulative frequency distribution (ascending).

The rank of the median is (50).

Less Than	Ascending Cumulative Frequency
Less Than 50	0
Less Than 70	8
Less Than 90	Rang of med(50) 23 f_1
Less Than 110	51 f_2
Less Than 130	71
Less Than 150	86
Less Than 170	94
Less Than 190	100

$\frac{n}{2} = 50,$ $f_1 = 23, f_2 = 51,$ A= 90, L= 110–90= 20

So, the median value is:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 90 + \frac{50 - 23}{51 - 23} \times 20$$

$$= 90 + \frac{27}{28} \times 20 = 90 + \frac{540}{28} = 90 + 19.286 = 109.3 \ R.S$$

-Third: Mode (Mod) .

The modal class is the class corresponding to the highest frequency.

The highest frequency is 28, which corresponds to the class interval (90–110).

Calculation of the differences:

(d_1 = 28 - 20 = 8), and (d_2 = 28 - 15 = 13).

Lower limit of the class: (A = 90)

Class width: (L = 110 - 90 = 20)

Thus, the mode is calculated using the following formula:

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L = 90 + \frac{13}{13+8} \times 20 = 90 + \frac{260}{21} = 1024 \ R.S$$

- **Distribution shape description:**

From the previous results, we find that:

Mean: X' = 113.4

Medium: Med = 109.3

Mode: Mod = 102.4

That is: Mean < Median < Mode, therefore the wage data distribution is positively skewed, as shown in the following figure.

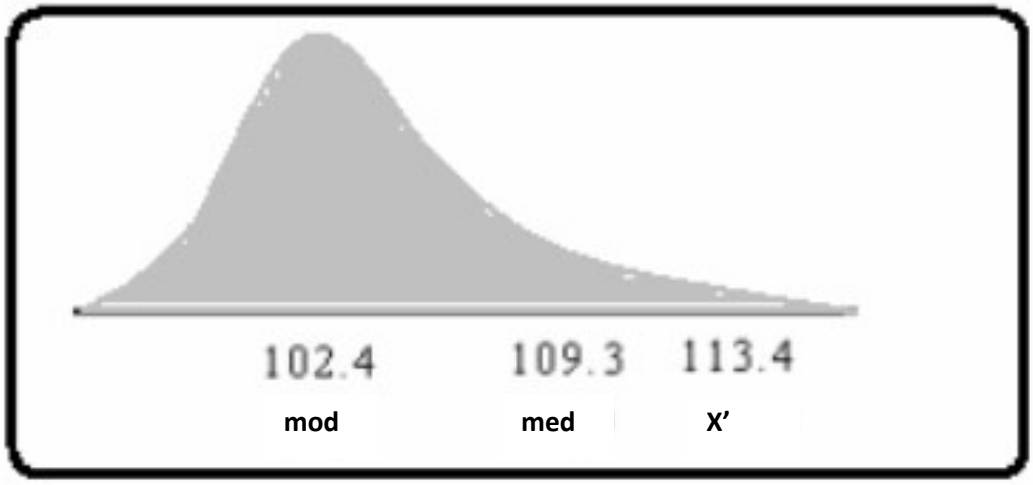


Figure 3-4 . The shape of the data distribution positively skewed

4. Quartiles

When a dataset is divided into four equal parts, three descriptive statistics known as **quartiles** are obtained:

- **First Quartile (Q1):**
  The value below which 25% of the observations fall.
- **Second Quartile (Q2):**
  The value below which 50% of the observations fall. This quartile is also referred to as the **median**.
- **Third Quartile (Q3):**
  The value below which 75% of the observations fall.

The following figure illustrates the positions of the three quartiles within a dataset.

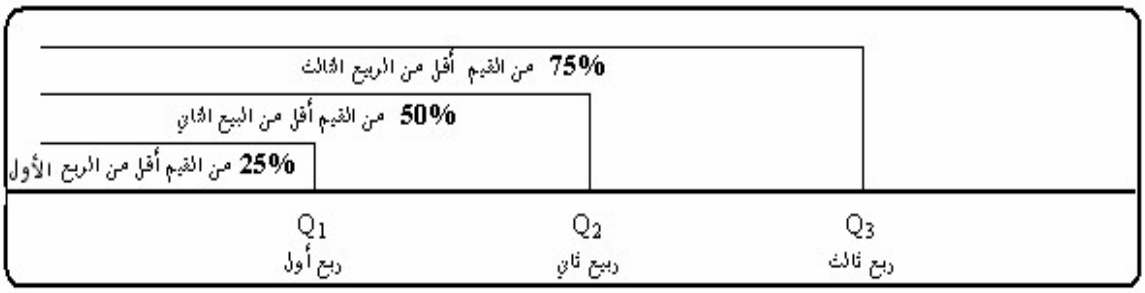


Figure 3-5 . The illustrates the positions of the three quartiles within a dataset.

Summary Table

Quartile	Meaning
Q1	25% of the data are below this value
Q2	Median
Q3	75% of the data are below this value
IQR	Difference between Q3 and Q1

-Interquartile Range (IQR)

The interquartile range measures the spread of the middle 50% of the data.

-Importance of Quartiles

- Understanding data distribution
- Measuring variability
- Detecting outliers
- Constructing box plots

-Quartile Position Formulas.

$$\text{Position of } Q1 = (n + 1) / 4$$

$$\text{Position of } Q2 = (n + 1) / 2$$

$$\text{Position of } Q3 = 3(n + 1) / 4$$

- Important Note

Before calculating quartiles, the dataset must always be arranged in ascending order.

Through the following figures, we can further interpret the quartiles.

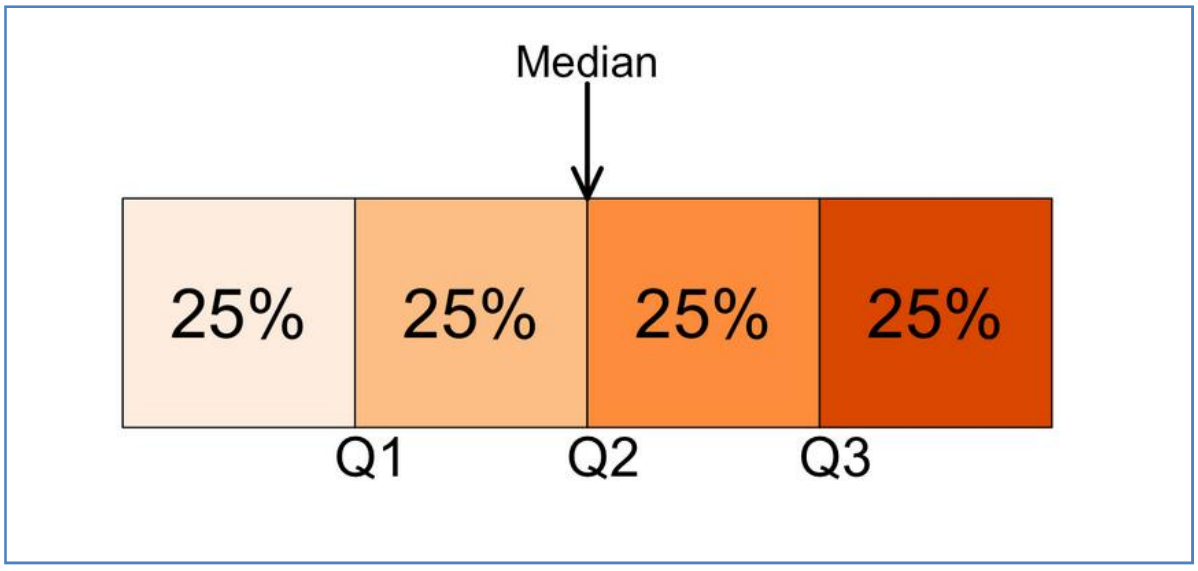


Figure 3-6 . The interpret the quartiles

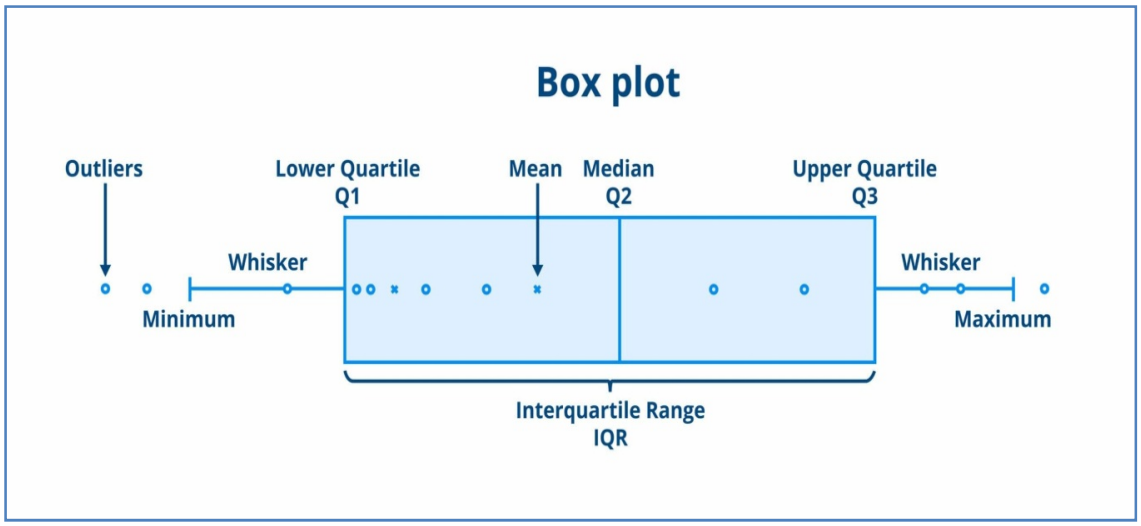


Figure 3-7 . The interpret the quartiles

To calculate any of the three quartiles, the following steps are followed:

- Assume that the number of observations is nnn, and that the data are arranged in ascending order as follows:

$$X(1)<X(2)<X(3)<\cdots<X(n)$$

where the ranks are:

$$1,\ 2,\ 3,\ \cdots,\ \cdots,n$$

- Determine the rank of the quartile Qi using:

$$R = (n+1)\times\left(\frac{i}{4}\right)$$

- If R is an integer, then the quartile value is:

$$Q_i = X_{(R)}$$

- If R is a fractional number, then the quartile Qi lies between:

$$X_{(l)} < Q_i < X_{(u)}$$

and Qi is calculated using the following formula:

$$Q_i = X_{(l)} + (R - l)(X_{(u)} - x_{(l)})$$

Example:

The following data represent the daily milk production quantity (in liters per cow) for a sample of 10 cows selected from a certain farm:

25, 23, 29, 32, 34, 29, 20, 18, 27, 30

Calculate the three quartiles of milk production quantity, and what is your interpretation?

Solution:

To calculate the three quartiles, the following steps are applied:

First - Arrange the values in ascending order and assign them ranks.

Data Table

Values	**18**	**20**	**23**	**25**	**27**	**29**	**29**	**30**	**32**	**34**
Rank	1	2	3	4	5	6	7	8	9	10
Rank of Quartile	2.75			5.5			8.25			

Quartile Ranks:

Q1=2.75Q2=5.5Q3=8.25

-Calculation of the First Quartile (Q1).

The position of the first quartile is:

$$R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{1}{4}\right) = 2.75$$

between the two values:

20<Q1<23

Applying Equation, we obtain:

$$l = 2,\ R = 2.75\ ,x_{(l)} = 20\ .x_{(u)} = 23$$

Therefore:

$$Q_1 = x_{(l)} + (R - l) \times (x_{(u)} - x_{(l)}) = 20 + 0.75(23 - 20) = 22.25$$

-Calculation of the second quartile (median) Q2.

So, the rank (position) of the second quartile is Q2.

$$R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{2}{4}\right) = 5.5$$

The second quartile (Q2) lies between the two values.

$$(27 < Q_2 < 29)$$

by applying the formula, we find that.

$$l = 5,\ R = 5.5\ ,x_{(l)} = 27\ .x_{(u)} = 29$$

So.

$$Q_2 = x_{(l)} + (R - l) \times (x_{(u)} - x_{(l)}) = 27 + 0.5(29 - 27) = 28$$

-Calculation of the third quartile (Q3).

The rank (position) of the third quartile is:

$$R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{3}{4}\right) = 8.25$$

The third quartile (Q3) lies between the two values.

$$(30 < Q_3 < 32)$$

By applying the formula, we get

$$l = 8, \quad R = 8.25 \ , x_{(l)} = 30 \ . x_{(u)} = 32$$

SO.

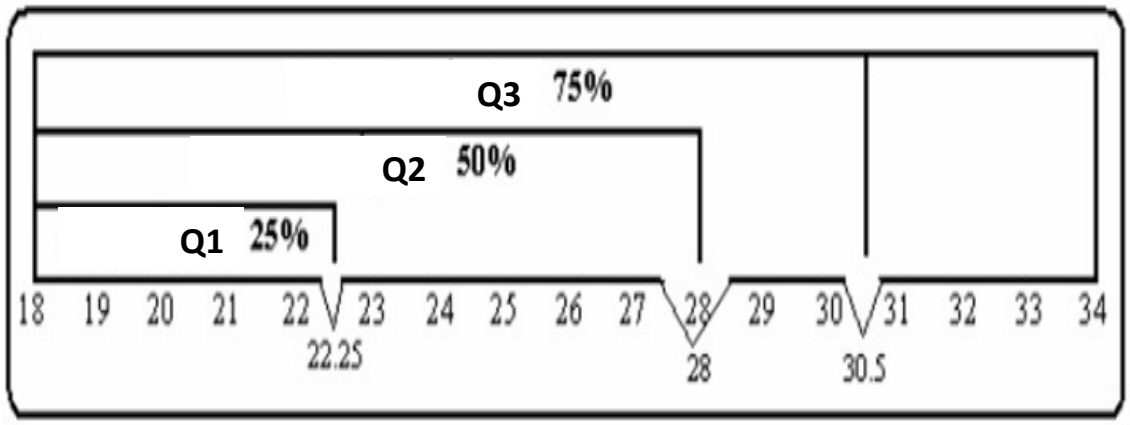


Figure 3-8 . The rank (position)

- 25 percent of the cows have a milk production of less than 22.25 liters per day.
- 50 percent of the cows have a milk production of less than 28 liters per day.
- 75 percent of the cows have a milk production of less than 30.5 liters per day.

-

- conclusion

Measures of central tendency are fundamental concepts in descriptive statistics used to identify the central or typical value within a dataset. They provide a single value that represents the overall distribution and help simplify large amounts of data for interpretation and comparison.

The three main measures of central tendency are the **mean**, **median**, and **mode**. The **mean** (arithmetic average) is calculated by summing all observed values and dividing by the total number of observations. It is widely used because it includes all data points; however, it is sensitive to extreme values (outliers), which can distort its representation of the data.

The **median** is the middle value of an ordered dataset. It divides the data into two equal parts, where 50% of the observations lie below it and 50% lie above it. The median is particularly useful in skewed distributions because it is not affected by extreme values.

The **mode** is the value or class that appears most frequently in the dataset. It is especially useful for categorical data or for identifying the most common value in a distribution. Together, these measures provide a comprehensive understanding of the data distribution. The relationship between mean, median, and mode can also indicate the shape of the distribution. For example, if the mean is greater than the median, and the median is greater than the mode, the distribution is positively skewed. Conversely, if the mean is less than the median and the median is less than the mode, the distribution is negatively skewed. The measures of central tendency are essential tools in statistical analysis, widely applied in economics, engineering, social sciences, and research to summarize and interpret data effectively.

CHAPTER IV :

Measures of Dispersion

Range, Mean Deviation, Variance, Standard Deviation.

CHAPTER IV Content

- **Introduction**

1. **Range**

- Range for Ungrouped Data
- Range for Grouped Data

2. Mean Deviation (MD)

- Mean Deviation for Ungrouped Data
- Mean Deviation for Grouped Data.

3. Variance

- Population Variance
- Sample Variance

4. Standard Deviation

- Standard Deviation of Ungrouped Data
- Standard Deviation of Grouped Data.

- **Conclusion**

- **Introduction.**

Measures of dispersion are used to measure the homogeneity of data. Data can be so closely related that measures of central tendency cannot be applied to compare data from two similar groups. Therefore, statisticians resort to using other measures of homogeneity, including measures of dispersion. These measures include the range, interquartile deviation, mean deviation, variance, and standard deviation.

1. Range

Range is the simplest measure of dispersion. It represents the difference between the largest and the smallest values in a dataset. The range provides a basic indication of the spread or variability of the data. A larger range indicates greater dispersion, while a smaller range suggests that the data values are more closely clustered together.

- Advantages of range measure:

It is preferred for announcing weather conditions, temperature, climate, pressure, etc. It is used in quality control.

- Disadvantages of range measure.

The range relies on only two values, disregarding other factors. It can be affected by outliers. The range is calculated for

1.1. Range for Ungrouped Data. using the following equation.

$$Rang = Max - Min$$

In the case of grouped data, the range scale is calculated using the following relationship.

1.2. The Range Of Grouped Data.

The Center Of The Last Group - The Center Of The First Group

Example,

The following frequency table shows the distribution of 60 farms according to the area planted with maize in hectares.

the category of Area	20-15	25-20	30-25	35-30	35-40	40-45
Number of farms	3	9	15	18	12	3

The required is calculation the Rang of area planted with maize.

The solution.

Rang = Last Category Center - First Category Center

Rang= (2/(20+15) – 2/(40+45))

= 17.5 – 42.5 =25

If the value of Rang is 25 hectares.

2. Mean Deviation (MD).

This is one of the measures of dispersion and is expressed as the average of the absolute deviations of the values from their arithmetic mean. The following equation is used to calculate the mean deviation.

This equation is used for ungrouped data.

$$MD = \frac{\sum |x - \bar{x}|}{n}$$

- To calculate the mean deviations, four basic steps must be taken.

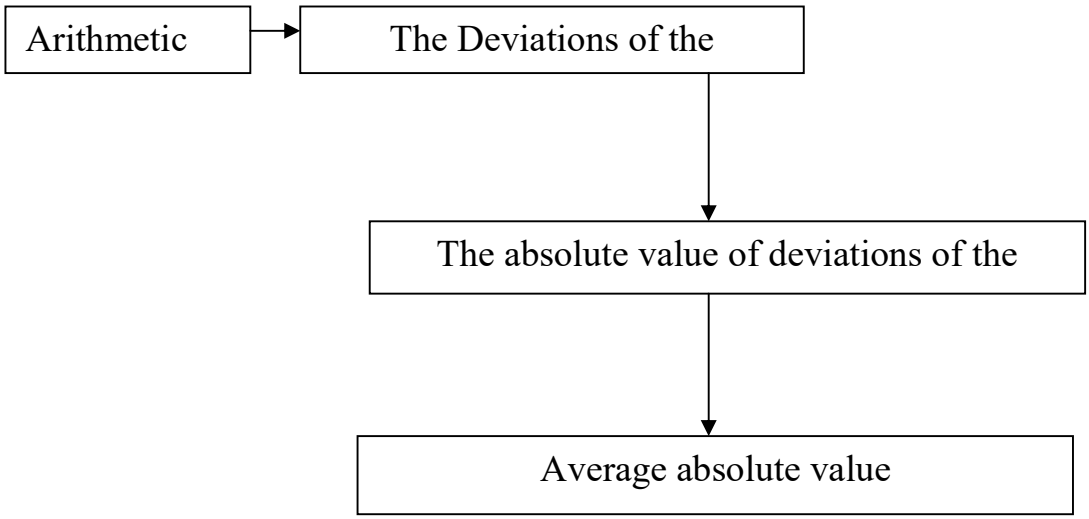


Figure 4-1 . four basic steps To calculate the mean deviations

- Advantages and disadvantages of the mean deviation.

One of its advantages is that it takes all values into account. However, its disadvantages include its susceptibility to outliers and the difficulty in mathematically manipulating it.

2.1. Mean Deviation (MD) of ungrouped data.

Example.

If the export capacity of five desalination plants is in million cubic meters as follows.

5 .2 . 4. 10 . 7.

Find the value of the mean deviation of the exported energy.

The solution.

To calculate the mean deviations, four basic steps must be taken: -

1-Arithmetic mean.

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5.6$$

x	$(x-\bar{x})=(x-5.6)$	$\lvert x-5.6\rvert$
4	4 - 5.6 = -1.6	1.6
5	5 - 5.6 = -0.6	0.6
2	2 - 5.6 = -3.6	3.6
10	10 - 5.6 = 4.4	4.4
7	7 - 5.6 = 1.4	1.4
Sum	0	11.6

So we are have M.D.

$$MD = \frac{\sum |x-\bar{x}|}{n} = \frac{11.6}{5} = 2.32$$

2.2. Mean Deviation (MD) of Grouped Data.

Example-

In the case of grouped data, the mean deviation is calculated using the following equation.

$$MD = \frac{\sum |x - \bar{x}| f}{n}$$

F = Category frequency Category center

X= Category frequency Category center

$\bar{X}$= Arithmetic mean.

Where.

The following frequency table shows the distribution of 40 Families according to their monthly expenditure in thousands of Algerian dinars.

Spending	2 - 5	5 - 8	8 - 11	11 – 14	14 – 17
N. Families	1	8	13	10	8

The required is the value of MD.

The solution.

category	f	x	$x\,f$	$\bar{x}$	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$
2-5	1	3.5	3.5	$\bar{x}=\frac{\sum x}{n}$	7.2	7.2
5-8	8	6.5	52		4.2	33.6
8-11	13	9.5	123.5		1.2	15.6
11-14	10	12.5	125	$=\frac{428}{40}=10.7$	1.8	18
14-17	8	15.5	124		4.8	38.4
sum	40		428			112.8

$$MD = \frac{\sum |x - \bar{x}| \, f}{n} = \frac{112.8}{40} = 2.82$$

3. Variance.

Variance is a fundamental measure of statistical dispersion that quantifies the extent to which data values deviate from their arithmetic mean. It is calculated as the average of the squared deviations from the mean and is widely applied in statistical analysis and scientific research. Two forms of variance are commonly distinguished: population variance and sample variance.

-Advantages of Variance

1. Uses All Observations

Variance is calculated using every value in the dataset, making it a comprehensive measure of dispersion.

2. Measures Data Variability Accurately

It provides a precise indication of how far observations are spread around the mean.

3. Fundamental to Statistical Analysis

Variance is the basis for many statistical techniques, including:

- Standard deviation
- Correlation analysis
- Regression analysis
- Analysis of Variance (ANOVA)
- Hypothesis testing

4. Suitable for Mathematical Operations

Because it is based on squared deviations, variance possesses useful algebraic properties that facilitate advanced statistical calculations.

5. Sensitive to Changes in Data

Any change in the dataset affects the variance, making it a useful indicator of variability.

-Disadvantages of Variance

1. Difficult Interpretation

Variance is expressed in squared units rather than the original units of measurement.

For example:

- If height is measured in meters (m),
- Variance is measured in square meters (m^2),

which can make interpretation less intuitive.

2. Sensitive to Extreme Values

Because deviations are squared, outliers have a large impact on the variance.

3. Not Easily Understood by Non-Specialists

Compared with the standard deviation, variance is less straightforward to explain and interpret.

4. Depends on the Mean

Since variance is calculated relative to the arithmetic mean, it may not be appropriate when the mean is not a representative measure of central tendency.

3.1. Variance in Population.

If the studied items are about a population, let's say x1, x2, x3..., to study the variance, we first denote it by the symbol s^2 sigma² and calculate it using the following relationship.

$$\sigma^2 = \frac{\sum (x-u)^2}{N}$$

Where

$$m = \sum x / N$$

Example.

We have a food packaging factory with 15 employees, where the number of years of experience of these employees is as follows:

5 . 13 . 7 . 14 . 12 . 9 . 6 . 8 . 10 . 13 . 14 . 6 . 11 . 12 . 10 .

Assuming this data was collected from some of members the community,

find the variance in the number of years of experience.

The Answer.

1- Calculating the arithmetic mean. $m = \frac{1}{N}\sum x$

$$= \frac{1}{15}(5+13++7+...+12+10) = \frac{1}{15}(150) = 10$$

2- Calculating deviations squared.

$$\sum (x-m)^2$$

$$\sum (x-m)^2 = 130$$

So the variance in the number of years of experience is.

x	$(x-m)$	$(x-m)^2$
5	5-10 = -5	25
13	3	9
7	-3	9
14	4	16
12	2	4
9	-1	1
6	-4	16
8	-2	4
10	0	0
13	3	9
14	4	16
6	-4	16
11	1	1
12	2	4
10	0	0
150	0	130

$$s^2 = \frac{\sum (x-u)^2}{N} = \frac{130}{15} = 8.67$$

3.2. The Variance of Sample.

The variance of a sample can be calculated as an estimate of the variance of the entire population. If the number of random samples is x1, x2, x3…xn and the sample size is n, then the sample variance is calculated using the following formula and is denoted by the symbol s^2.

Where :

$$\bar{x} = \sum x/n$$

$$S^2 = \sum (X - \bar{X})^2 / n$$

Variance is a powerful and essential measure of dispersion in statistics because it quantifies variability and serves as the foundation for many analytical methods. However, its interpretation can be challenging due to the use of squared units and its sensitivity to extreme values. For this reason, the standard deviation is often preferred for reporting and interpreting variability, while variance remains indispensable for statistical calculations and theoretical analysis.

4. Standard Deviation.

In short, the standard deviation can be defined as the positive square root of the variance, as shown in the following figure 2.

$$\text{Standard Deviation} = \sqrt{\ \text{Variance}}$$

Figure 4- 2. Standard Deviation

Therefore, the standard deviation is calculated according to the following relationship.

$$S = \sqrt{\sum (X - \bar{X})^2 / n}$$

-Interpretation of Standard Deviation

- Data are concentrated around the mean.
- Low variability.

-Advantages of Standard Deviation

- Uses all observations.

- Widely used in statistical analysis.

- Essential for regression and correlation analysis.

- Provides a precise measure of variability.

4.1. Standard Deviation of Ungrouped Data.

Example:

Find the variance and standard deviation of the following data.

22	15	13	12	8

The Answer.

X	$X - \bar{X}$	$(X - \bar{X})^2$
8	-6	36
12	-2	4
13	-1	1
15	1	1
22	8	64
$\sum X = 70$	$\sum (X - \bar{X}) = 0$	$\sum (X - \bar{X})^2 = 106$

$$\bar{X} = \sum X / n$$
$$= 70/5 = 14$$
$$S^2 = \sum (X - \bar{X})^2 / n$$
$$= 106/5 = 21.2 \text{ ----------> [Variance]}$$
$$S = \sqrt{\sum (X - \bar{X})^2 / n}$$
$$= \sqrt{106}/5 = \sqrt{21.2} = 4.6 \text{ ------> Standard deviation (S.D)}$$

4.2. Standard Deviation of Grouped Data.

In the case of of grouped data., the relationships become as follows.

1- Variance.

$$S^2 = \frac{\sum f(X - \bar{X})}{\sum f}$$

2-Standard Deviation.

$$S = \sqrt{\frac{\sum f(X - \bar{X})}{\sum f}}$$

Example.

We have the following data; we need to calculate the variance and standard deviation.

Grades	0-10	10-20	20-30	30-40	40-50	50-60
frequency	1	4	10	19	11	5

The Answer.

category	f	(X) C.categor	f * X	X - $\bar{X}$	(X- $\bar{X}$)^2	f (X- $\bar{X}$)^2
0 – 10	1	5	5	-30	900	900
10 – 20	4	15	60	-20	400	1600
20 – 30	10	25	250	-10	100	1000
30 – 40	19	35	665	0	0	0
40 – 50	11	45	495	10	100	1100
50 – 60	5	55	275	20	400	2000
$\sum$	50	-	1750	-	-	6600

$$\bar{X} = \frac{\sum fX}{\sum f}$$

$$\bar{X} = 1750 / 50 = 35$$

$$S^2 = \frac{\sum f(X-\bar{X})}{\sum f}$$

$S^2 = 6600/50 = 132$ ----------> [variance]

$S = \sqrt{6600/50} = \sqrt{132} = 11.48$ (S.D)

-Relationship Between Variance and Standard Deviation.

Variance-(Standard Deviation)^2

$$S^2 = \frac{\sum f(X-\bar{X})}{\sum f}$$

and

$$s\sqrt{s^2}$$

-Summary.

Measure	Symbol	Formula	Purpose
Variance	(s^2)	$S^2 = \frac{\sum f(X-\bar{X})}{\sum f}$	Measures average squared dispersion
Standard Deviation	(s)	√s2	Measures average dispersion in original units

- Variance measures the spread of data using squared deviations.
- Standard deviation is the square root of variance.
- A larger standard deviation indicates greater variability.
- A smaller standard deviation indicates more homogeneous data.
- Standard deviation is generally preferred because it is expressed in the same unit as the original data.

•Conclusion.

Variance and standard deviation are fundamental statistical measures used to assess the dispersion and variability of data around the arithmetic mean. While measures of central tendency such as the mean, median, and mode describe the central location of a dataset, they do not provide information about the extent to which observations differ from one another. Variance and standard deviation address this limitation by quantifying the spread of data values.

Variance measures the average of the squared deviations from the mean, providing a rigorous mathematical representation of variability. It serves as the foundation for numerous statistical methods and analyses, including regression, correlation, hypothesis testing, and analysis of variance (ANOVA). However, because variance is expressed in squared units, its direct interpretation may be less intuitive.

To overcome this limitation, the standard deviation is used. As the square root of the variance, it is expressed in the same units as the original data, making it easier to interpret and communicate. The standard deviation indicates the typical distance of observations from the mean and is widely employed in research, business, economics, engineering, and the social sciences to evaluate data consistency and reliability.

Together, variance and standard deviation provide valuable insights into the structure and distribution of data. Small values indicate that observations are closely clustered around the mean, reflecting greater homogeneity and stability, whereas large values suggest substantial variability and dispersion among observations. Consequently, these measures play a crucial role in descriptive and inferential statistics, helping researchers understand data behavior, compare

datasets, evaluate uncertainty, and make informed decisions based on statistical evidence. They remain indispensable tools for analyzing and interpreting quantitative information across a wide range of scientific and applied disciplines.

General Conclusion.

Descriptive statistics is a fundamental branch of statistics that deals with the systematic process of collecting, organizing, summarizing, and presenting data in a clear and meaningful way. Its primary objective is to transform large and complex datasets into simplified and interpretable information, allowing researchers and decision-makers to understand the essential characteristics of the data without losing its key meaning.

This field is based on three main components. The first is data organization, which involves arranging raw data into structured forms such as tables, frequency distributions, and classifications. This step is essential for making data manageable and preparing it for further analysis. The second component is numerical summarization, which includes measures of central tendency such as the mean, median, and mode, as well as measures of dispersion such as range, variance, and standard deviation. These statistical measures provide important insights into the behavior, distribution, and variability of the data. The third component is graphical representation, which involves the use of visual tools such as bar charts, pie charts, histograms, and line graphs to present data in a way that is easy to interpret and visually intuitive.

Unlike inferential statistics, descriptive statistics does not aim to make predictions or generalize results beyond the dataset being analyzed. Instead, it focuses strictly on describing and summarizing the available data in its current form. This makes it a crucial first step in any statistical analysis, as it provides a solid foundation for deeper statistical investigations.

Descriptive statistics is widely applied in many academic and professional fields. In architecture, for example, it is used to analyze spatial configurations, evaluate user behavior, assess environmental conditions, and measure design performance. It also plays an important role in engineering, economics, social sciences, and environmental studies, where understanding data patterns is essential for informed decision-making.

The, descriptive statistics is an indispensable tool in data analysis. By organizing, summarizing, and visually representing data, it enhances clarity, improves understanding, and facilitates effective communication of results. It ultimately supports better interpretation of information and contributes to more accurate and informed decision-making processes across various disciplines.

- Bibliography.

1- **statistics** – by David Freedman et al.

2- **Introductory Statistics** – by Barbara Illowsky & Susan Dean.

3- **Statistics for Business and Economics** – by Paul Newbold et al.

4- **Applied Statistics and Probability for Engineers** – by Douglas C. Montgomery.

- **مراجع باللغة العربية.**

1- د- شرف الدين خليل, كتاب الاحصاء الوصفي- مكتبة شبكة الابحاث والدرسات الاقتصادية

2- د- علي احمد السقاف, الاحصاء الوصفي والاستدلالي ، المركز الديمقراطي العربي-برلين- المانيا

3- د- احمد سعودي ,جامعة محمد بوضياف المسيلة-كلية العلوم الإنسانية والاجتماعية- محاضرات في تحليل الانحدار الخطي المتعدد.