

جامعة العربي بن مهيدي - أم البواقي -

كلية العلوم الاجتماعية والإنسانية

اللجنة العلمية لقسم العلوم الاجتماعية

أم البواقي في: 2026/06/09

أم البواقي في: 2026/06/09

مستخرج من محضر اللجنة العلمية للقسم

المنعقد 08 جوان 2026

الموضوع: المصادقة على مطبوعة بيداغوجية

بناء على التقارير الإيجابية للخبراء الآتية أسماؤهم: د. قادري عادل (جامعة أم البواقي)،
د. سويبي دحمان (جامعة باتنة 1)، حول مطبوعة بيداغوجية بعنوان: " descriptive statistics " موجه
لطلبة السنة أولى جذع مشترك " للأستاذة حيواني كريمة، فقد تم اعتماد المطبوعة البيداغوجية،
كسند علمي.

رئيس اللجنة العلمية



أ.و. خالد عبد الوهاب
رئيس اللجنة العلمية

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Larbi Ben M'Hidi University

Faculty of Social and Human Sciences
Department of Social Sciences



A Pedagogical Handout in

Descriptive Statistics

A Pedagogical Handout Intended for First-Year Students
First Semester – Common Field in Social Sciences

Academic Year 2025–2026

Table of Contents

Table of Contents	
Introduction	6
First: General Introduction	
1. Introduction to Statistics	7
2- Concept of Statistics	7
2-1. Descriptive Statistics:	9
2-2. Inferential Statistics (Inductive Statistics)	10
3.The Development of Statistics	11
4.The Relationship Between Statistics and Other Sciences	14
5.Statistical Concepts and Terminology	18
6. The Importance of Statistics	23
7.Objectives of Statistics	23
8.Characteristics of Statistics	24
9. Sources of Statistical Data Collection	25
10. Methods of Collecting Statistical Data	26
11.Statistical Measurement Scales	27
Second: Organization and Presentation of Data	
1. Tabular Presentation of Statistical Data	37
2. Graphical Representation of Data	39
1.Presentation of Data in the Case of a Quantitative Variable	42
2.Graphical Representation of the Absolute and Relative Frequency Distribution of a Discrete Quantitative Variable	48
3.Graphical Representation of the Absolute and Relative Frequency Distributions of a Continuous Quantitative Variable	49
4.Presentation of Data in the Case of a Continuous Quantitative Variable	50
5.Graphical Representation of the Frequency Distribution for a	57

Continuous Quantitative Variable	
6.Graphical Representation of the Cumulative Frequency Distribution:	58
7.Presentation of Data in the Case of an Ordinal Qualitative Variable:	59
8.Graphical Representation of an Ordinal Qualitative Variable:	60
9.Presentation of Data for a Nominal (Non-Ordinal) Qualitative Statistical Variable:	61
10.Graphical Representation of a Nominal (Non-Ordinal) Qualitative Variable	62
Third: Measures of Central Tendency	
1. Concept of Central Tendency	67
2. Arithmetic Mean	
2.1.Uses of the Arithmetic Mean in Social Research	68
2.2.Characteristics and Limitations of the Arithmetic Mean	69
2.3.Methods of Calculating the Arithmetic Mean	70
1. Arithmetic Mean for Ungrouped Data	70
2. Arithmetic Mean for Grouped Data Using a Frequency Distribution	71
3.Determining the Arithmetic Mean Using Graphical Representations	74
3. Median	
3.1 Uses of the Mode in the Social Sciences	90
3.2.Characteristics of the Median	91
3.3.Computation of the Median	91
1.Median for Ungrouped Data	91
2.Median for Grouped Data	93
3.Graphical Method for Determining the Median	97
4. Mode	
4.1. Uses of the Median in the Social Sciences	99

4.2.Characteristics of the Mode	100
4.3.Calculation of the Mode	100
1.In the case of ungrouped data	100
2.In the Case of Grouped Data (Frequency Distribution)	102
3.Graphical Determination of the Mode	104
5.Relationship between Measures of Central Tendency (Arithmetic Mean, Median, and Mode)	105
Fourth: Measures of Dispersion	
1) Quartiles, Deciles, and Percentiles	
1.Quartiles	
1.2. Uses of Quartiles in the Social Sciences	114
1.3.Calculation of Quartiles	115
1. In the Case of Ungrouped Data	115
2. Quartiles in the Case of Grouped Data Using a Frequency Distribution	117
2. Deciles	
2.2.Uses of Deciles in the Social Sciences	123
2.3.Computing Deciles	124
1. In the Case of Ungrouped Data	124
2. Deciles in the Case of Grouped Data Using a Frequency Distribution	126
3.Percentiles	
3.1. Uses of Percentiles in the Social Sciences	132
3.2.Computing Percentiles	133
1. In the Case of Ungrouped Data	133
2.Percentiles in the Case of Grouped Data Using a Frequency Distribution	135
2) Standard Deviation and Variance	
1.1.Measures of dispersion	143
1.2.Statistical properties of variance	143

1.3.Statistical interpretation	144
2.Variance	
1.2.Sample variance (S^2)	144
2.2.Properties of variance	145
2.3.Uses of variance and standard deviation in the social sciences	145
2.4.Calculation of variance	146
1.Ungrouped data	146
2.Grouped data using a frequency distribution	147
3.Standard Deviation	
3.1.Definition of the standard deviation	150
3.2.Properties of the standard deviation	150
3.3.Standard Deviation Calculation	151
4. Standard Error	
1.The standard error	155
1.1.The statistical definition of the standard error	155
1.2.Calculation of the standard error	155
1.3.The difference between the standard error and the standard deviation	156
1.4.The relationship between the standard error and sample size	156
1.5.The importance of the standard error and its applications	156
Fifth: Measures Relating Central Tendency and Dispersion	
Coefficient of variation	

1.Absolute Measures of Dispersion	159
2.Relative Measures of Dispersion	159
3.Definition	159
4.Advantages of the Coefficient of Variation	160
5.Disadvantages:	160
6.Interpretation of the Coefficient of Variation Results	160
Sixth: Measures of Shape	
1.Definition of perfect symmetry	167
1.Skewness Coefficient	
2.Definition of skewness	167
2.1.Coefficient of skewness	168
2.2.Measurement of skewness	169
2. Kurtosis Coefficient	
3.Measures of kurtosis	169
Conclusion	176
Reference	177

Introduction

Statistics is regarded as one of the fundamental pillars of modern scientific knowledge, given its central role in the systematic and objective study and analysis of various phenomena. It is no longer viewed merely as a numerical tool for presenting data; rather, it has developed into an integrated scientific discipline that contributes to interpreting reality, understanding its dimensions, and supporting decision-making across different fields, particularly in the social sciences, where phenomena are characterized by complexity and interdependence.

This publication begins with a general overview of statistics, addressing its emergence and development, and highlighting its relationship with other scientific disciplines. Special emphasis is placed on the basic statistical concepts and terminology that form the foundation of any sound statistical study, such as statistical population, sample, variables, data types, and statistical measures. The book also discusses methods of organizing and presenting data, whether quantitative or qualitative, and gives particular attention to graphical representation as an effective means of simplifying and clarifying information.

The publication then examines measures of central tendency, namely the arithmetic mean, median, and mode, in terms of their concepts, significance, and applications in the social sciences. It explains methods for calculating these measures for both grouped and ungrouped data, in addition to their graphical determination. Separate sections are devoted to measures of dispersion, which complement measures of central tendency by assessing the degree of data spread and variability, including quartiles, deciles, percentiles, standard deviation, variance, and standard error.

This work is not limited to a descriptive presentation of statistical measures; it also explores the relationship between measures of central tendency and dispersion through the coefficient of variation. Furthermore, it addresses measures of distribution shape, such as skewness and kurtosis, which help in understanding the nature of statistical distributions and analyzing their characteristics. Overall, this publication aims to provide readers with a solid scientific foundation and clear practical skills, enabling them to apply statistical methods systematically in scientific research and to interpret and analyze results with a critical and informed perspective.

1. Introduction to Statistics

Statistics emerged during the Middle Ages as a result of governments' need to obtain accurate information about their populations. Population censuses were primarily conducted to determine the number of individuals who could be mobilized to form strong armies capable of defending the state against external threats arising from ambitions of territorial expansion and economic gain. In parallel, governments were also interested in estimating individuals' wealth in order to impose taxes and collect the financial resources necessary for funding the army and managing public affairs.

As data collection activities expanded, they gradually came to include information on births, deaths, production, and consumption. This expansion highlighted the necessity of organizing and summarizing the collected data in a systematic and accessible manner, often through the use of tables, to facilitate understanding, analysis, and practical use. Initially, these practices were referred to as "*the science of the state*" or "*the science of rulers*." Over time, they developed into an independent scientific discipline known today as Statistics.

The term Statistics is derived from the Latin word *State*, meaning "the state" (Al-Saifo, 2010, p. 23). Initially, statistics was understood as the discipline concerned with the scientific methods used for collecting, organizing, and presenting data, either in graphical or tabular form.

However, with the development of probability theory during the seventeenth and eighteenth centuries, the use of statistical analysis expanded significantly. Statistics began to be employed not only to describe data, but also to derive meaningful results that support decision-making, forecasting, estimation, and inference about a set of variables larger than those that had been directly observed.

2- Concept of Statistics

The term "ahsā" (to enumerate or count) is mentioned several times in the Holy Qur'an. Among these references are verse 34 of Surat Ibrahim, in which Allah says: "And if you were to count the favors of Allah, you could not enumerate them"; verse 6 of Surat Al-Mujādilah: "Allah has enumerated it, while they have forgotten it"; and verse 28 of

Surat Al-Jinn: “And He has encompassed whatever is with them and has enumerated all things in number.”

Among the commonly held perceptions of statistics, it is often viewed as a collection of numbers and data, such as population figures, numbers of births and deaths, and similar information. Consequently, statistics has traditionally been associated with counting, enumerating phenomena, and expressing them numerically. This represents a limited and narrow understanding of the concept of statistics.

Several definitions of statistics have been proposed. The following is a concise presentation of one such definition:

Statistics is a science concerned with the methods of collecting facts related to scientific and social phenomena that are represented through multiple cases or observations. It also deals with the recording of these facts in standardized numerical form and their summarization in a manner that facilitates the identification of trends and relationships among these phenomena. Furthermore, statistics involves the study of these relationships and trends and their use in understanding the true nature of phenomena and identifying the laws governing them (Mohamed Bouheza, 2011, p. 06).

Statistics is defined as “the science that studies the various methods and techniques for collecting quantitative data related to economic, social, and other phenomena, as well as arranging, classifying, analyzing, interpreting, and presenting these data in appropriate forms and formats, with the aim of facilitating sound decision-making” (Ratoul, 2006, p. 07).

Statistics is also defined as “the science that investigates appropriate scientific methods and techniques for collecting data, classifying, organizing, analyzing, and interpreting them in order to reach the results necessary to enhance knowledge or to make appropriate decisions, and to generalize, analyze, and interpret these results” (Al-Daima and Al-Basha, 2013, p. 16).

In addition, statistics may refer to the enumeration of objects. For example, the number of universities in Algeria in 2025 reached 66 universities; the total production of the Appliances for Monitoring and Measurement (AMC) factory between 2005 and 2025

amounted to 150,000 electricity meters; and the number of university hospitals in Algeria by 2025 reached 52 hospitals.

As for the technical definition of statistics, it is defined as “the science concerned with collecting, organizing, presenting, analyzing, and interpreting data in order to reach results and laws that govern them, as well as to make appropriate decisions.”

Accordingly, statistics consists of four main steps:

1. Data Collection: This step involves obtaining measurements or values derived from observations and experiments conducted by the researcher.

2. Organization and Presentation of Data: This refers to arranging the data obtained in the first step into specific tables designed for this purpose and presenting them using appropriate methods, such as graphical representations or frequency distributions.

3. Data Analysis: This involves applying various statistical methods to analyze the collected and presented data in order to provide a descriptive explanation of the phenomena under study.

4. Inference and Decision-Making: This step refers to the conclusions drawn in the form of estimates or predictions. Although this step is primarily associated with inferential statistics, it remains a fundamental outcome of the statistical process.

Based on the preceding definitions, two main types of statistics can be distinguished: Descriptive Statistics (Statistique Descriptive) and Inferential Statistics.

2-1. Descriptive Statistics:

It includes a set of statistical principles that help describe human and social phenomena through descriptive measures. This enables the researcher to present data in a form that is easy to understand and interpret, and to determine the degree to which these phenomena are present within the original population (Ahmed, 2008, p. 17).

It also focuses on organizing and presenting data in the form of tables, charts, and graphical representations, as well as calculating measures of central tendency (mean,

median, and mode) and measures of dispersion (range, standard deviation, variance, percentiles, quartiles, and deciles), in addition to other statistical measures.

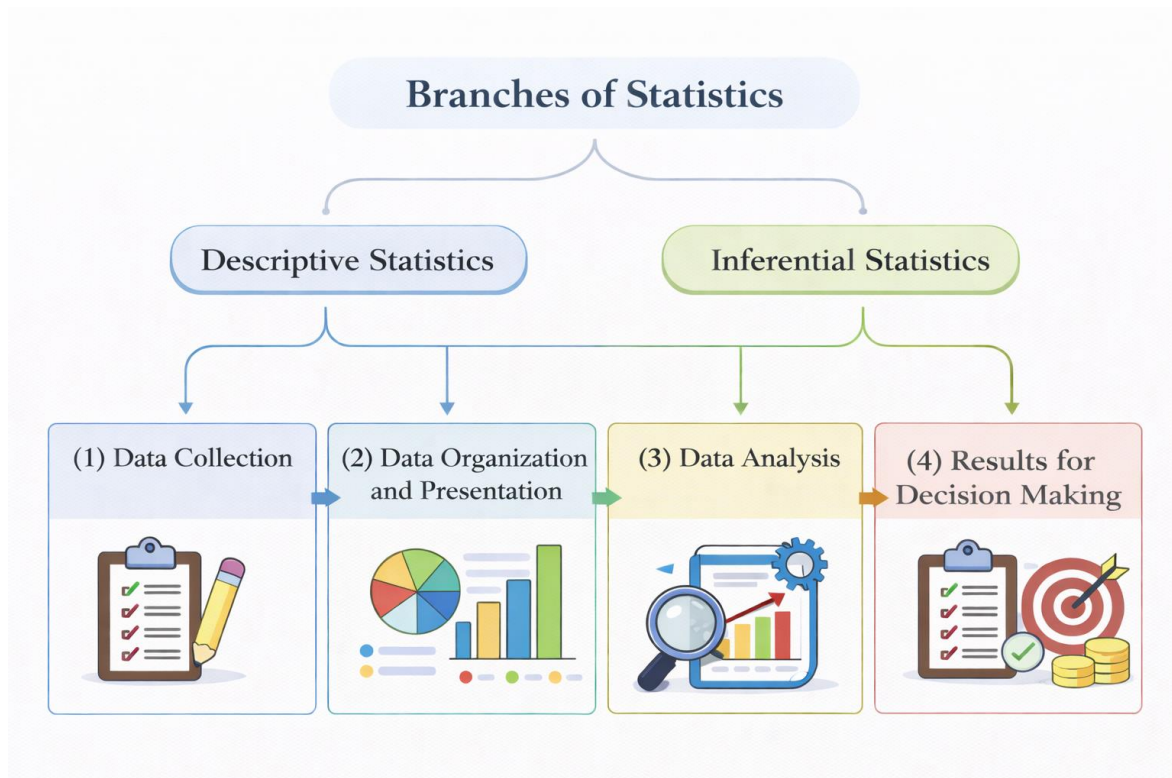
2-2. Inferential Statistics (Inductive Statistics)

It encompasses the statistical methods used to reach decisions, judgments, and conclusions about a population based on a sample drawn from that population (Boulquas, 2014, p. 8).

In other words, inferential statistics is concerned with generalizing results obtained from a sample to the entire population. That is, the researcher makes inferences about the population based on the data available from the selected sample.

The following figure summarizes the concept of statistics.

Figure (01) illustrates the concept of statistics



3. The Development of Statistics

The development of statistics represents a historical journey from a simple administrative tool used for collecting state data to an advanced mathematical science that supports decision-making across various scientific and social fields.

1. Early Beginnings: Statistics as an Administrative Tool

The emergence of statistics is closely linked to the fundamental functions of ancient societies, where the need to count populations, estimate resources, and collect taxes was central to state organization. In ancient Egypt, historical sources indicate that the pharaohs conducted an early population census around 3000 B.C. in order to estimate the labor force (Pearson & Kendall, 1970, p. 11).

Similarly, China developed a relatively advanced census system as early as the twentieth century B.C., aimed at maintaining accurate records of births, deaths, and cultivated land (Stigler, 1986, p. 4). Although these practices did not yet constitute statistics in its methodological and scientific sense, they laid the foundational basis for the idea of collecting quantitative data and using it in political and administrative decision-making.

2. The Middle Ages and the Renaissance: Record Keeping

During the Middle Ages, statistics maintained its practical role in administration. The Catholic Church played a prominent role in this field by keeping records of births, marriages, and deaths, which later became an important database for studying population development (Porter, 1986, p. 27).

Similarly, Islamic states relied on administrative bureaus (*Diwans*) to record taxes and land ownership. The Renaissance period witnessed the first early signs of statistical analysis, particularly in the work of Italian mathematicians related to insurance data and pension systems. This stage represented a transition of statistics from a purely authoritative administrative practice to a tool used for monitoring demographic changes.

3. The Seventeenth Century: The Transition to Scientific Analysis

The seventeenth century represents a pivotal turning point, during which quantitative analysis began to assume a structured scientific character. John Graunt is considered the founder of biostatistics. In his renowned study *Natural and Political Observations* (1662), he analyzed mortality data in London and identified regular patterns in birth and death rates. This work demonstrated that social phenomena are governed by measurable statistical laws (Graunt, 1662, p. 45).

During the same period, William Petty developed what he termed “*Political Arithmetic*,” emphasizing that numerical data constitute the foundation for understanding economic conditions and managing the state (Pearson & Kendall, 1970, p. 38).

4. The Eighteenth Century: Mathematical Foundations and Probability Theory

During the eighteenth century, the mathematical foundations of statistics became firmly established through the development of probability theory. Jacob Bernoulli published his seminal work *Ars Conjectandi* (1713), in which he laid the foundation for the concept of the *Law of Large Numbers*, demonstrating that the repetition of random experiments leads to the stabilization of relative frequencies (Stigler, 1986, p. 64).

Meanwhile, Pierre de Moivre introduced an approximation of the binomial distribution using the normal distribution, later known as the *bell-shaped curve*. Subsequently, Pierre-Simon Laplace further developed probability theory and expanded its applications across various scientific fields.

5. Statistics in the Social Sciences: The Emergence of Quetelet (Early 19th Century)

At the beginning of the nineteenth century, newly developed mathematical tools were applied to the study of human and social behavior. *Social physics* emerged through the work of Adolphe Quetelet, who employed statistical methods to analyze human phenomena and introduced the concept of “*the average man*” (*L’homme moyen*).

Quetelet applied the normal distribution to the measurement of individuals and behaviors, emphasizing that major social phenomena—such as crime and marriage rates—

are governed by statistical laws that allow for prediction. This approach laid the foundation for the use of statistics in the social sciences and public policy.

6. The Nineteenth Century: Modern Statistics and Its Academic Institutionalization

This century witnessed a major qualitative shift marked by the establishment of statistical societies, such as the Royal Statistical Society in 1834, which significantly strengthened the academic status of statistics. Among the most prominent contributors was Francis Galton, who introduced the concepts of regression and correlation toward the mean.

In addition, Karl Pearson made fundamental contributions by founding the journal *Biometrika* in 1901 and by developing the correlation coefficient and the chi-square test (χ^2). These advances played a crucial role in promoting the application of statistical methods in the biological and social sciences (Porter, 1986, p. 114).

7. The Twentieth Century: The Statistical Revolution and Intellectual Debate

Statistics reached a new level of maturity through the work of Ronald A. Fisher (R. A. Fisher), who established the theoretical foundations of experimental design, developed analysis of variance (ANOVA), and formulated the principles of hypothesis testing (Fisher, 1925, p. 12).

This period also witnessed the development of Bayesian statistics, along with the emergence of nonparametric methods, which enabled researchers in psychology and sociology to move beyond strict assumptions regarding data distributions (Bernardo & Smith, 2000, p. 5).

8. The Twenty-First Century: Big Data and Methodological and Ethical Challenges

In the twenty-first century, statistics has become closely integrated with computer science, emerging as a fundamental pillar of Data Science and Big Data analytics.

- **Ethical and social challenges:**

Issues related to data bias and its impact on predictive models—particularly in the context of social justice—have become increasingly prominent. Moreover, the protection of privacy has become a critical concern, leading to the emergence of the field of Differential Privacy.

- **Methodological integration:**

This century has also witnessed a growing use of mixed methods, as quantitative statistics alone is no longer sufficient. Instead, it is now effectively integrated with qualitative analysis to better understand the complex social contexts underlying numerical data, thereby enhancing the accuracy and depth of social research.

The historical development of statistics represents a transition from a simple tool for tax collection to a science of inference. It evolved from the description of phenomena (Graunt and Quetelet) to their prediction (Laplace and Fisher), and has now reached a stage where it is integrated with artificial intelligence and machine learning to process large-scale data.

The essence of statistics in the twenty-first century lies in its ability to transform uncertainty into actionable knowledge. This makes statistics not only a foundation of scientific research, but also a key compass for decision-making in governments, businesses, and the formulation of effective social policies.

4. The Relationship Between Statistics and Other Sciences

Statistics is considered one of the interdisciplinary sciences, as it intersects with a wide range of scientific and human disciplines. It provides quantitative tools that assist in interpreting phenomena and making evidence-based decisions. Statistics is not merely a set of numerical calculations; rather, it constitutes a framework that helps transform data into actionable knowledge across various fields (Gelman & Nolan, 2017, p. 12).

Moreover, statistics offers tools for measuring phenomena, interpreting relationships, and testing hypotheses, which makes it a fundamental pillar of scientific research in the social, natural, and applied sciences.

1. Its Relationship with the Social and Human Sciences

Statistics plays a central role in the study of human behavior and social phenomena. In

sociology, for example, researchers rely on surveys and sampling techniques to understand public opinion and social change, while carefully considering coverage and selection errors that may affect the credibility of research findings (Babbie, 2020, p. 195). Statistics is also an essential tool for analyzing societal changes such as poverty, unemployment, and migration, thereby enabling the formulation of more realistic and evidence-based social policies (Babbie, 2021, p. 50).

In psychology, statistics plays a fundamental role in analyzing the results of psychological and educational tests, evaluating the effectiveness of educational or therapeutic programs, and measuring psychological attributes such as intelligence, anxiety, and job satisfaction. This enables researchers to test hypotheses and understand human behavior objectively (Coolican, 2018, p. 45), thus enhancing the validity of results and their generalizability.

In the field of education, statistics is considered a key tool for measuring academic achievement, evaluating curricula, and analyzing attitudes, values, and differences between groups. It also contributes to the development of psychological and educational measurement instruments through the study of validity and reliability (O'Dwyer & Bernauer, 2014, p. 47). This supports informed decision-making in educational policy and enhances research objectivity while ensuring the generalizability of results (Creswell, 2012, p. 201).

2. The Relationship Between Statistics and the Economic and Administrative Sciences

In economics, statistics constitutes a primary tool for the study of economic indicators such as unemployment and inflation rates, as well as growth indicators. It is also used in the analysis of financial markets and the forecasting of economic crises. Moreover, statistics enables the formulation of economic forecasts based on accurate quantitative data and market predictions, as it is widely applied in financial data analysis, risk measurement, and the study of consumer behavior (Al-Hassadi, 2014, p. 45).

Accurate economic decision-making relies heavily on statistical models such as time series analysis and multiple regression. In modern management, decision-makers depend on statistics to support strategic decisions based on reliable data. Statistical methods are also applied in the fields of operations research, risk analysis, and administrative quality

control, which helps organizations improve efficiency and ensure the sustainability of their decisions (Research Academy, 2023).

3.The Relationship Between Statistics and the Health and Medical Sciences

Biostatistics constitutes a fundamental pillar in the design of medical studies and the analysis of their results. It is the primary tool used in designing clinical trials, comparing treatment methods, and evaluating the effectiveness of drugs and vaccines (Dawson & Trapp, 2009, p. 47).

Statistics also plays a vital role in epidemiology by estimating prevalence and mortality rates (Pagano & Gauvreau, 2018, p. 30), and in genetics through the analysis of the frequency of genetic mutations across different populations. In this way, statistics becomes an essential tool for making evidence-based therapeutic and preventive decisions. Without statistics, it would be difficult to assess the validity of medical findings or to generalize them to broader populations.

4. The Relationship Between Statistics and the Natural and Engineering Sciences

Statistics plays a pivotal role in physics, chemistry, and the biological sciences through the analysis of laboratory and experimental data. It is also widely used in engineering for quality control, process optimization, and the development of simulation models. The application of statistical methods in molecular biology and industrial engineering has enabled discoveries and applications that would not have been possible without this quantitative approach (Al-Fayez, 2021, p. 119).

Moreover, statistics contributes to industrial product quality control, the development of technological models, and the analysis of environmental and climatic phenomena through the prediction of long-term fluctuations (Montgomery & Runger, 2014, p. 8).

5.The Relationship Between Statistics, Computer Science, and Artificial Intelligence

The contemporary relationship between statistics and computer science is particularly evident in the fields of machine learning and artificial intelligence. Most algorithms used in big data processing, pattern recognition, and computer vision are based on probabilistic

models and statistical inference techniques. Dwivedi and colleagues emphasize that statistics represents “*the backbone that ensures the reliability and transparency of artificial intelligence outputs*” (Dwivedi et al., 2023, p. 22).

Similarly, Hastie, Tibshirani, and Friedman argue that many algorithms—such as logistic regression and neural networks—are essentially advanced statistical models developed for data classification and prediction (Hastie, Tibshirani et al., 2017, p. 119). Big data analytics relies heavily on statistical methods to extract patterns, support decision-making, and forecast future trends. This enables intelligent systems to detect regularities and predict outcomes, making statistics a core component of the digital revolution (James et al., 2021, p. 3).

The relationship between statistics and other sciences is therefore organic and integrative, positioning statistics as a common language of scientific research and a truly interdisciplinary methodology. Statistics is not limited to being a computational tool; rather, it constitutes a philosophical and methodological framework for understanding the world and constructing scientific knowledge across all fields. It provides the social sciences with the capacity for quantitative interpretation, equips economics with predictive tools, supports management in forecasting and decision-making, supplies medical sciences with rigorous experimental methods to ensure the reliability of results, and assists engineering and natural sciences in model development and simulation. Finally, statistics provides artificial intelligence with the foundations of credibility, objectivity, and the ability to handle highly complex data.

Thus, statistics emerges as a “servant science” (Servant Science) that contributes to the advancement of both human knowledge and practical applications. Accordingly, statistics constitutes a fundamental pillar in the construction of scientific knowledge across diverse disciplines

5. Statistical Concepts and Terminology

The importance of statistics did not emerge only in recent years; rather, it became evident toward the end of the eighteenth century, when Laplace (1749–1829) and Gauss (1777–1855) began developing statistical analysis and formulating the laws of probability. From that period onward, statistics started to expand into other disciplines such as

demography, biology, economics, sociology, and particularly psychology.

2. Statistical Terminology

2-1. Statistical Unit

A statistical unit refers to an element of the population under study that possesses one or more characteristics. It also represents the basic entity on which the statistical study or sampling process is conducted. A statistical unit must be defined in a precise and clear manner. It may represent a living entity such as a person, a student, or an employee; a material object such as an institution, a vehicle, or a container; or even an abstract entity such as an idea (Azzouz, 2010, p. 15).

In statistical terminology, a unit refers to any movable or immovable entity that can be measured or counted, and it is commonly denoted by the symbol U . For example, a worker in the population of workers, or a student in the population of students.

2-2. Sample

A sample is a group consisting of a number of units or individuals that represent the statistical population, meaning that it reflects the characteristics present in the population. The members of a sample must be selected randomly, such that every individual in the population has an equal chance of being included in the sample. The size of the sample varies depending on the importance of the study and the available material and human resources required to conduct it.

Sampling is one of the most commonly used approaches in field studies, due to the practical impossibility of collecting statistical information from all units that constitute the population under study, a process known as a complete census (Jallatou, 2010, p. 5). A sample is denoted by the symbol n .

It should be noted that the methods used to select a sample depend on the nature of the study.

2-3. Population

A population refers to the entire set of individuals or objects that share the characteristics of the study. It includes an unlimited number of possible samples, each of which consists of a group of individuals. The population is denoted by the symbol N .

The statistical population can be divided into two types:

1- Target Population

The target population refers to the entire group of individuals or elements that are intended to be studied and to whom the results of the study are meant to be generalized, such as university students, public sector employees in the health field within a given country, and similar groups.

2- Study Population

The study population consists of the group of individuals from whom it is actually possible to obtain data or information related to the phenomenon or problem under investigation.

In many cases, the statistical population is large, making it difficult or impractical to study all its elements. Therefore, researchers resort to studying a subset of the population, known as a sample.

2-4. Parameter

A parameter is a measure or constant that describes certain characteristics of a population. It is obtained through the analysis of population data and is usually derived from a complete census (Al-Daima & Al-Basha, 2013, pp. 20–21).

For example, the average individual income in a given country is considered a parameter, as it reflects the standard of living of the population. Parameters are typically denoted by Latin symbols, such as the population mean μ and the population variance σ^2 .

2-5. Statistical Phenomenon

A statistical phenomenon refers to a characteristic of elements that varies from one element to another in terms of form, type, or quantity. The characteristic under study is referred to as a variable (Ali Al-Omari & Ali Al-Ajili, 2013, p. 3).

Examples include height, age, weight, job satisfaction, performance, and leadership.

2-6. Variables

The term *variable* is frequently used in psychology, education, research studies, and textbooks on research methodology, statistics, and measurement. It is difficult to engage with research or communicate scientifically without using this concept, as variables constitute the building blocks of research questions and hypotheses.

- **Variable:**

A variable is a characteristic, behavior, trait, or measurable attribute that describes individuals within a sample. These characteristics vary from one individual or element to another (Al-Daima & Al-Basha, 2013, p. 21). Variables are measured using a specific scale and are observable by the researcher. They are commonly denoted by the symbol x_i .

For example, gender is a variable because it includes categories such as male and female; marital status is a variable because it includes categories such as single, married, divorced, widowed, and separated; and exam anxiety is a variable because it includes different levels (high anxiety, moderate anxiety, low anxiety).

- **Constant:**

A constant is a characteristic that does not vary and shows no differences among individuals.

Example:

There is a relationship between **job satisfaction** and **organizational commitment** among employees of industrial institutions.

Job satisfaction: Variable

Organizational commitment: Variable

Type of institution: Constant (*industrial institutions only*) **Employees:** Constant (*permanent employees only*)

- **Classification of Variables:**

Variables are divided into two main categories:

1- Qualitative (Categorical) Variables

Qualitative variables refer to variables or phenomena that cannot be measured numerically (quantitatively), but whose frequency of occurrence can be counted. These variables are descriptive in nature, and numerical values do not carry quantitative meaning in this context (Al-Saifo et al., 2010, p. 29). They represent attributes or categories rather than numerical quantities and are classified as follows:

- **Ordinal Qualitative Variables:**

These are descriptive data that do not represent true numerical values but can be

ordered according to degree, level, or rank, either in ascending or descending order. In this case, it is possible to state that one category is higher or lower than another, but the magnitude of the difference between them cannot be measured numerically. Examples include educational level (primary, middle, secondary, university) and levels of satisfaction (low, medium, high).

- **Nominal Qualitative Variables:**

These are descriptive data that represent categories or attributes that cannot be ordered. In other words, individuals can be classified into different categories, but there is no hierarchy or ranking among them.

Examples include gender (male – female), where neither category can be described as “higher” or “lower” than the other; blood type (A – B – AB – O); marital status (single – married – divorced – widowed); nationality; and academic specialization. These values cannot be ranked, as they serve solely for identification and classification rather than comparison.

In this type of variable, classification is based on whether or not an individual possesses a particular attribute or characteristic.

2- Quantitative (Numerical) Variables

Quantitative variables are those that can be expressed numerically using real numbers and measured objectively. They are the most commonly used and widespread type of variables in statistical studies, as the language of statistics is fundamentally numerical. Examples include weight, height, number of workers, number of working hours, and similar measures. Quantitative variables are further divided into two main types (Jallatou, 2002, p. 7):

2-1. Continuous Quantitative Variables

Continuous quantitative variables are variables that can take any real value within a given range, whether these values are whole numbers such as (50, 72, 100) or fractional/decimal values such as (72.5, 100.3, 42.75). Examples include temperature, weight, height, age, and wage.

These variables result from measurement rather than counting, and their units of measurement can be continuously divided into smaller parts. As a result, they can take an almost infinite number of possible values within a specified range.

2-2. Discrete Quantitative Variables

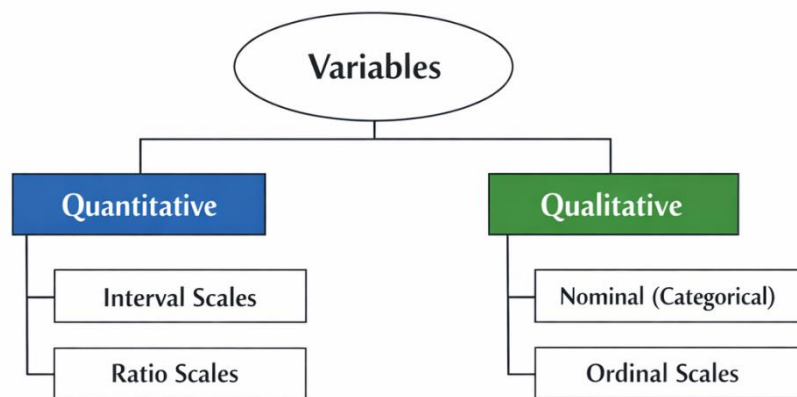
Discrete quantitative variables are variables that take only whole numerical values and cannot be divided into fractional parts within their unit of measurement. They usually result from counting rather than measurement, meaning they are expressed using integer values only.

Examples include the number of workers, the number of students at different levels of education, the number of rooms in a house, and the number of private economic institutions, among others.

With regard to statistical analysis, parametric statistical methods are commonly applied to this type of data.

The following figure illustrates the types of variables and measurement scales.

Figure (02): Classification of Variables and Measurement Scales



6. The Importance of Statistics

Statistics is considered one of the most important methods and tools in scientific research. Through the application of its rules, principles, and techniques, researchers are able to collect the necessary data and information required for a study and analyze them in order to reach the intended research results.

Statistics also plays a prominent role in developing future plans by forecasting outcomes across various sectors. It can be applied in numerous fields, including economic sciences, human and social sciences, such as education, industry, medicine, management, and even psychology.

In order for a researcher to achieve the objectives of a study, it is essential to master various scientific research methods and to be able to coordinate different variables, particularly during the stages of data collection and analysis. This necessity highlights the importance of relying on statistical methods in research.

7.Objectives of Statistics

In the modern era, statistics represents one of the fundamental pillars of human knowledge. It is no longer limited to computational techniques; rather, it has become an integrated intellectual framework for understanding and analyzing natural and social phenomena.

1. Description and Classification

Statistics aims to provide an accurate description of the phenomena under study through the collection, organization, and classification of data. This process enables the production of numerical summaries and graphical representations that help highlight the main characteristics of the data.

2. Inference and Prediction

Statistics is not limited to description; it extends to generalizing results from samples to the original populations and building predictive models. This field has developed significantly with the emergence of Bayesian inference, which offers greater flexibility in dealing with uncertainty and estimating the probabilities of future events (McElreath, 2020, p. 67).

3. Decision Making

Statistics serves as a practical tool for decision-makers across various fields, including economics, medicine, and education. It allows for the evaluation of alternatives based on objective quantitative criteria. Consequently, statistics provides decision-makers with

reliable tools to select the most appropriate options under conditions of uncertainty (Gregory, 2019, p. 88).

4. Discovery and Exploration

Statistics contributes to the discovery of hidden patterns within data and the identification of relationships between variables. This process facilitates the generation of new hypotheses and broadens the scope of scientific research.

8.Characteristics of Statistics

1. Quantitative Nature

Statistics is based on quantitative measurement and numerical analysis of phenomena, which distinguishes it from purely descriptive approaches and enables it to produce results that are testable and verifiable (Anderson et al., 2020, p. 52).

2. Flexibility and Adaptability

Statistics is characterized by its ability to adapt to various fields, ranging from the social sciences to medicine and engineering, and by developing methods tailored to specific research problems (Al-Fayez, 2021, p. 73).

3. Objectivity

Statistics establishes scientific objectivity by relying on systematic methods and verifiable results, thereby reducing the influence of subjective bias in scientific research (Gregory, 2019, p. 95).

4. Continuous Evolution

Statistics is marked by dynamism and continuous development, whether through the advancement of data analysis techniques or through the integration of technological innovations such as artificial intelligence and big data analytics (Al-Fayez, 2021, p. 102).

9. Sources of Statistical Data Collection

Statistical data sources refer to the origins from which statisticians obtain the data under study. Researchers generally rely on two main sources to acquire statistical information related to a particular phenomenon (Ratoul, 2006, p. 13).

9-1. Primary (Direct) Sources

Primary sources are those from which data are obtained directly. In this case, the researcher personally collects data from the sample under study. For example, when a researcher seeks to collect data on wage satisfaction, they may conduct an interview with a worker and obtain first-hand information regarding the worker's income, such as salary amount, allowances, needs, and expenditures.

Among the most important methods used to obtain information through this approach are direct contact and interaction with the units of the statistical population, or reliance on documents in which the information is still in its raw form. In this context, questionnaires and interviews represent two of the most important tools for collecting data from primary sources.

This type of data source is characterized by a high level of accuracy and reliability, since the researcher personally collects the data from the study sample. However, its main disadvantages are that it requires a significant amount of time and effort, and it is also financially costly.

9-2. Secondary (Indirect) Sources

Secondary sources are those from which data are obtained indirectly. In this case, the researcher acquires statistical information from previous studies and investigations, where the data have already been classified and organized by earlier researchers or by official institutions and agencies (such as the National Office of Statistics), or by non-official bodies. These data may be published in special bulletins or journals, or stored in traditional or electronic archives, such as security service reports, newspaper reports, and records of workplace accidents, among others.

One of the main advantages of this type of data source is the saving of time, effort, and financial resources. However, the degree of confidence in secondary sources is generally lower than that associated with primary sources.

10. Methods of Collecting Statistical Data

When conducting a statistical study of a particular phenomenon, it is necessary to collect data and information related to the units of the population under investigation. Statistical

data are collected from the field using one of the following two methods:

10-1. Complete Enumeration Method (Census)

This method involves collecting data from all units that constitute the statistical population. When using this method, it is assumed that the population is limited in size, such as students in a psychology department, workers in a GICA factory, or victims of medical errors. This method is considered one of the best approaches to data collection, as it provides comprehensive data about the research problem (Al-Mashhadani & Al-Abidi, 2013, p. 28).

This method provides highly accurate information when the conditions of scientific research are met. Despite its advantages—such as precision, lack of bias, and comprehensive coverage—it often remains impractical or even impossible to implement. This is due to the high costs involved, as complete enumeration requires substantial material and human resources and a longer period of time. Moreover, as the size of the population increases, the probability of error also increases.

This method is used when a very high level of accuracy is required, as in general censuses related to population, agriculture, or the economy. Therefore, it is most often the responsibility of governments.

10-2. Sampling Method

This method is used when it is difficult to conduct a study on all members of the population, or when it is sufficient to obtain information about a part of the population instead of the entire population (Al-Daima & Al-Basha, 2013, p. 22). In this case, a subset (sample) of the population elements under study is selected using a sound scientific approach. By statistically analyzing the sample data, the results can be generalized to the entire population.

It should be noted that the results obtained from a selected sample become closer to the true characteristics of the population as the sample size increases and as scientifically valid sampling procedures are followed.

Sampling is particularly useful when studying large or infinite populations, such as

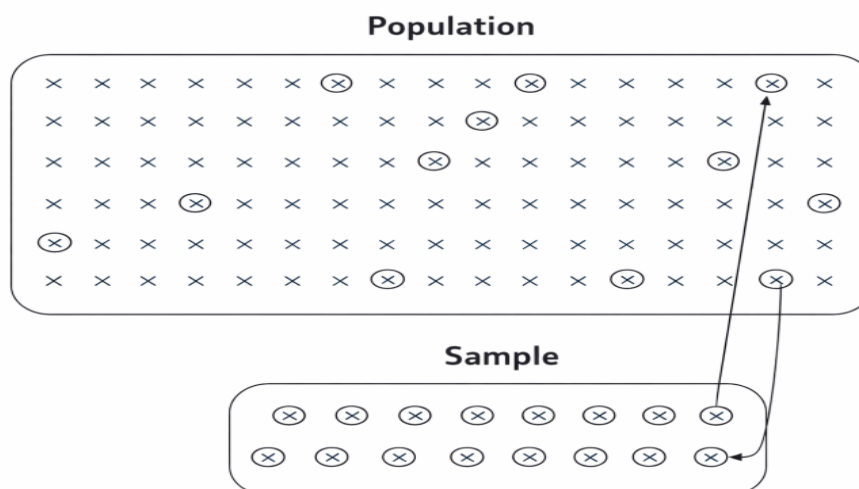
workers in the education sector, employees of Sonatrach, or women working in the health sector. However, the accuracy of results in such cases is relatively lower, since not all individuals are included in the study.

Among the main reasons for adopting the sampling method are the following (Al-Mashhadani & Al-Abidi, 2013, p. 28):

- Saving the money, effort, and time required to conduct research.
- The difficulty of conducting a complete census due to the nature of the population, which may be unlimited, very large, or composed of valuable or hazardous units, or similar constraints.

The objective of sampling is to draw conclusions about the population from which the sample is selected. The first step in sampling is to determine the sample size, followed by identifying the sampling frame from which the sample will be drawn. Subsequently, one of the appropriate sampling procedures or methods is applied.

Figure (03): Relationship Between the Population and the Sample



10.3.Types of Samples and Sampling Techniques

There are numerous types of samples, and through the procedures used to select them, the researcher can choose the type that best suits the conditions of the study being conducted. In general, samples can be classified into two main types:

1.Probability Samples

Probability samples are samples drawn from the statistical population in such a way

that each element of the population has an equal chance of being included in the sample. Consequently, these samples are representative of the population from which they are drawn, are suitable for statistical analysis, and allow the results to be confidently generalized to the population they represent (Al-Omari & Al-Ajili, 2000, p. 07).

More precisely, probability samples are those in which the sampling units are selected randomly from the study population in order to avoid bias resulting from the selection of units. Probability samples are classified as follows:

Simple Random Sample

A simple random sample is one in which each individual in the population has an equal probability of being included in the selected sample. This type of sampling is particularly useful and effective when the population is homogeneous and shares common characteristics (Kafi et al., 2012, p. 34).

For example, a population of executives may be considered homogeneous in terms of income levels. To obtain a simple random sample, researchers commonly use a random number table or the lottery method.

Stratified Random Sample

This type of sampling is used for heterogeneous populations, that is, populations composed of several social strata based on criteria such as income level, expenditure level, educational level, and other similar characteristics.

In this approach, the population is divided into homogeneous strata, and the proportion of each stratum relative to the total population is determined. The size of each stratum is denoted by N_1, N_2, \dots, N_i , where i represents the number of strata in the population.

To draw a stratified random sample, the following steps are followed (Jallatou, 2002, p. 10):

- Determine the proportion of each stratum relative to the total population $\frac{N_i}{N}$.
- Determine the desired sample size n .

- Determine the number of statistical units to be selected from each stratum n_{i_ini} , according to the proportions identified in Step (1), where:

$$n_i = n \times \frac{N_i}{N}$$

- Randomly select n_i units from each stratum N_i using a random number table. After completing this process, all selected units are combined to form a stratified random sample.

Systematic Random Sample

Systematic random sampling is used to select a sample from a population whose size is finite or known. In this method, a sampling interval (step) is determined, and a random starting number is chosen from the first interval. The sampling interval is then added successively to the initial number (Al-Daima & Al-Basha, 2013, p. 28).

For example, if we assume that we want to select a sample consisting of 100 individuals from a population of 2000 individuals, we begin by dividing the population size by the sample size:

$\frac{2000}{100}$ Thus, the sampling interval (step) equals 20 individuals. We then select the first number randomly (for example, number 9), and subsequently add the sampling interval (20) each time until one hundred numbers are obtained. Accordingly, the selected numbers would be: (9, 29, 49, 69, ...).

Cluster Random Sample

In some populations, units are organized into groups or clusters that are often similar with respect to the characteristic being studied, such as cities, colleges, or schools. These groups are referred to as clusters, each of which contains a number of elements from the original population that are often relatively homogeneous.

In such cases, cluster sampling is applied by selecting a simple random sample from among these clusters.

Table No. (01): Types of Cluster Sampling in Field Research

Type	Scientific Definition	Implementation Steps	Simplified Example	Fields of Use	Advantages	Disadvantages
Single-Stage Cluster Sampling	Clusters are selected randomly, and all elements within each selected cluster are studied without further internal selection.	<ul style="list-style-type: none"> - Dividing the population into clusters - Randomly selecting clusters - Studying all individuals within each cluster 	Selecting 6 schools randomly and studying all students within them	<ul style="list-style-type: none"> - Educational studies - Health studies - Local population studies 	<ul style="list-style-type: none"> - Simple - Less costly than a census - Fast 	Accuracy may decrease if clusters are highly heterogeneous
Two-Stage Cluster Sampling	Clusters are selected first, then a random sub-sample is selected from within each cluster.	<ul style="list-style-type: none"> - Selecting clusters - Selecting a sample within each cluster 	Selecting provinces → selecting institutions → selecting a subset of workers	<ul style="list-style-type: none"> - Social research - Organizational studies - Field psychological studies 	<ul style="list-style-type: none"> - Balanced between accuracy and cost - Flexible 	<ul style="list-style-type: none"> - Requires careful design - Higher sampling error
Multi-Stage Cluster Sampling	Sampling is conducted across multiple levels (three stages or more), moving progressively from larger to smaller units.	Country → provinces → municipalities → households → individuals	National household income and expenditure survey	<ul style="list-style-type: none"> - National statistics - Population studies - Public health - Government studies 	<ul style="list-style-type: none"> - Highly flexible - Very economical compared to a census - Suitable for large populations 	<ul style="list-style-type: none"> - Complex design - Requires statistical expertise - Risk of error if not applied accurately
Area Sampling	Relies on natural geographical divisions rather than administrative units.	Dividing an area geographically → selecting areas → studying households or individuals	Dividing a city into zones and selecting some for study	<ul style="list-style-type: none"> - Urban studies - Social studies - Urban planning - National surveys 	<ul style="list-style-type: none"> - Highly practical - Suitable for wide and scattered areas 	<ul style="list-style-type: none"> - Depends on map accuracy - Possible population density variation

Multi-Stage Random Sampling

This method is used when it is difficult to reach all elements of the target population directly, or when it is difficult to construct a complete sampling frame that includes all population elements. Accordingly, it is not necessary under this method to obtain a full

sampling frame for all population elements, especially in the final stage. This type of sampling is commonly used in agricultural statistics.

2. Non-Probability Samples

Non-probability samples are those in which sampling units are selected using non-random methods. In this case, the researcher selects sample elements in a way that serves the purpose of the study. This method is typically used when the members of the original population are unknown. Selection is therefore non-random and involves researcher intervention, whereby some individuals are chosen and others are excluded based on predefined criteria.

Examples include selecting samples of drug addicts, conducting psychological studies on intellectually disabled primary school students, or studying prison inmates. There are several types of non-probability sampling methods, which vary according to researchers' orientations. Here, only two types are presented (Al-Omari & Al-Ajili, 2000, p. 11):

Convenience (Accidental) Sample

This type of sample relies purely on chance. More precisely, it is selected without prior planning or systematic arrangement. One of its main advantages is that it saves time and reduces costs. It may provide reliable information if the target population is highly homogeneous. However, if the population elements are heterogeneous, this method may lead to selection bias.

A major limitation of this type of sampling is that its elements do not accurately represent the study population, making it difficult to generalize the study results to the entire population.

Quota Sample

In quota sampling, the researcher divides the study population into categories and then selects a number of individuals from each category proportional to its size within the population. In this respect, quota sampling resembles stratified sampling. However, it differs in two main aspects:

- It is used when the study population is unlimited.

- The researcher personally selects individuals by name without being bound by strict selection rules, meaning that the researcher intervenes in the selection process, whereas no such intervention occurs in stratified sampling.

This method reduces potential bias associated with non-probability sampling and is practically useful when sampling frames for population strata are unavailable. However, it still carries the risk of bias if proportional balance among strata is not accurately achieved, and it does not precisely represent the study population.

. Snowball Sample

In some studies, researchers may not clearly know which individuals should be included in the data collection process. In such cases, snowball sampling is used. This method begins by selecting one individual, and based on that person's response, the researcher identifies the next participant. The first individual thus serves as the starting point, and the process continues until the sample is completed.

11. Statistical Measurement Scales

1. Nominal Scale

The nominal scale is the simplest level of measurement. It is based on classifying data into categories or groups without any order or preference among them. At this level, it is only possible to calculate frequencies and percentages, and numerical comparisons between categories cannot be made (there is no concept of higher or lower).

Examples:

Type of institution (public / private), gender (male / female), type of contract (permanent / temporary), work department (human resources / production / services).

2. Ordinal Scale

The ordinal scale adds the property of ranking among categories, allowing the identification of higher and lower positions. However, the distances between ranks are not equal. In other words, it is possible to know which category is better or higher, but not the exact magnitude of the difference between them.

Examples:

Levels of job satisfaction (low – medium – high), performance evaluation (poor – good – excellent), level of organizational commitment (weak – moderate – strong).

At this level of measurement, the median and ranks can be calculated, and some nonparametric statistical tests can be applied.

3. Interval Scale

The interval scale represents a more advanced level of measurement. Data can be ordered, and equal intervals exist between values. However, the zero point is not a true zero but rather a symbolic one; therefore, it is not possible to state that one value is twice another. This scale allows operations such as subtraction, calculation of the mean, and standard deviation.

Examples:

Scales measuring professional attitudes toward technology, occupational stress scales ranging from 1 to 10, and some Likert-type scales that are often treated as interval scales in statistical analysis.

4. Ratio Scale

The ratio scale is the highest and most precise level of measurement. It possesses all the properties of the previous scales (ordering and equal intervals), in addition to having a true zero that indicates the complete absence of the characteristic being measured. This scale allows all mathematical operations, including ratio comparisons such as “twice” or “half.”

Examples:

Number of working hours, number of days of absence, number of occupational accidents, years of experience, number of hours of artificial intelligence usage at work per week.

Suggested exercises :**Exercise 1:**

Answer True or False and justify your answer:

1. Statistics is a science limited only to data collection without analysis.
2. Descriptive statistics aims to generalize results from a sample to the population.
3. Inferential statistics is based on probability theory.
4. A parameter describes the characteristics of a sample.
5. A variable is a fixed attribute that does not change among individuals.

Exercise 2:

Classify the following statements as Descriptive Statistics or Inferential Statistics, with justification:

1. Calculating the average age of students in the sociology department.
2. Predicting the unemployment rate in Algeria in 2030 based on previous data.
3. Presenting the distribution of students by gender in a statistical table.
4. Testing whether job satisfaction affects organizational commitment.

Exercise 3:

For a study on job satisfaction among healthcare workers in Algiers Province:

1. Identify:
 - The statistical population
 - The target population
 - The study population
2. Is it preferable to use:
 - A complete census
 - Orsampling?Justify your answer.

Exercise 4:

Consider the following study:

Studying the relationship between exam anxiety and academic achievement among first-year students.

1. Identify the variables in the study.
2. Classify each variable as:
 - Quantitative / Qualitative
 - Independent / Dependent
3. Give an example of a possible constant in this study.

Exercise 5:

Classify the following variables as Qualitative (Nominal / Ordinal) or Quantitative (Discrete / Continuous):

1. Gender
2. Number of daily study hours
3. Level of job satisfaction (low – medium – high)
4. Number of rooms in a house
5. Weight
6. University major

Exercise 6:

Identify the appropriate measurement scale for each of the following variables (Nominal – Ordinal – Interval – Ratio):

1. Type of contract (permanent / temporary)
2. Temperature measured in Celsius
3. Years of professional experience
4. Students' rank in class
5. Level of occupational stress from 1 to 10

Exercise 7:

Answer the following questions:

1. Mention two advantages and two disadvantages of primary data sources.
2. Give two examples of secondary data sources.

3. When is sampling preferred over a complete census?

Exercise 8:

For each of the following situations, identify the most appropriate sampling method and justify your choice:

1. Studying students' performance in a city divided into schools.
2. Studying workers' opinions in a small company with 40 employees.
3. Studying household income in a large country.
4. Selecting every tenth name from an employee list.

Exercise 9:

Answer the following questions:

1. Define a convenience (accidental) sample.
2. Why is it difficult to generalize results obtained from a convenience sample?
3. In which type of studies is snowball sampling used? Give an example.

Exercise 10:

For a study on the use of artificial intelligence at work among administrative employees:

1. Propose:
 - The type of population
 - The appropriate sampling method
2. Identify:
 - Two quantitative variables
 - One qualitative variable
3. Suggest the appropriate measurement scale for each variable.
4. Determine whether the study is closer to:
 - Descriptive statistics
 - Or inferential statistics
 - Justify your answer.

After collecting data related to a specific phenomenon, the researcher needs to organize and structure these data in a way that facilitates understanding their implications and clarifies their meanings through acceptable methods that allow for effective use. The importance of data becomes evident only when they are presented in an appropriate form that enables the researcher to benefit from them to the fullest extent. This is because the purpose of data presentation is to clearly and accurately highlight their main characteristics so that they can be used in making sound decisions. Achieving this requires the researcher to exert effort and to understand the nature of the field from which the data are collected, in order to help select appropriate methods for their presentation. Therefore, the researcher must organize the data using one of the following methods: statistical tables and graphical representations.

1. Tabular Presentation of Statistical Data:

The tabular presentation of data consists of arranging the primary data related to the phenomenon under study into statistical tables after they have been collected, organized, and classified, with the purpose of clarifying the main features of the phenomenon and facilitating its interpretation. This method is known for its accuracy and simplicity, as it allows the researcher or reader to grasp the general characteristics of the phenomenon through a rapid examination of the table.

The form and structure of a table vary according to the nature of the data being presented; however, tables commonly include two columns or two rows. The first column (or row) displays the values of the studied variable, which may take the form of attributes, discrete values, or class intervals, depending on the type of data. The second column (or row) contains the corresponding frequencies associated with these values or categories.

2. Rules for Constructing a Table:

For a statistical table to have validity and credibility, the following rules must be observed in its construction:

- Providing a clear title for the table.
- Indicating the source of the table's data.
- Stating the unit of measurement used, if applicable.
- Providing a title for each column and each row.
- Assigning a number to the table.

There are two types of tables: simple tables and composite tables.

1. Simple Frequency Tables

This type of table is used when analyzing a phenomenon that involves a single qualitative variable. In such cases, the data are usually organized in the form of a simple frequency table in order to summarize the distribution of the variable clearly. This table typically consists of two columns: the first identifies the categories or levels of the variable, while the second presents the number of observations corresponding to each category, expressed in the form of frequencies or percentages

2. Requirements to Be Considered When Designing a Data Table

When constructing a data table, several essential requirements must be taken into account to ensure clarity and accuracy.

- **Table Title:** It provides a concise indication of the data contained in the table.
Example: Number of workers in the electricity and gas distribution company in 2010.
- **Table Header:** It consists of vertically or horizontally arranged square or rectangular cells, each containing the headings corresponding to the data presented.
Example: Year, number of male workers, number of female workers, total number.
- **Table Body:** The table body represents the main component in which the data are recorded, and it is composed of cells that reflect the intersections between rows and columns, ensuring an organized and coherent display of information

Table Title: Table No. (2): Distribution of employees in the production company

Table 2.
Distribution of Employees by Department in a Production Company

Department	Number of Employees	Percentage (%)
Human Resources Department	7	10%
Finance Department	15	21%
Production Department	32	46%
Marketing Department	16	23%
Total	70	100%

Table Title: Clearly describes the content of the table in a concise manner.

Table Header: Identifies the variable categories and statistical measures using clear, bolded labels.

Table Body: Contains the statistical data obtained from the research, arranged under the appropriate header labels. **Table Body:**

2. Composite (Multiple) Frequency Tables

Composite frequency tables are tables that contain more than one variable, or multiple variables, or that examine two characteristics simultaneously within a given population.

2. Graphical Representation of Data

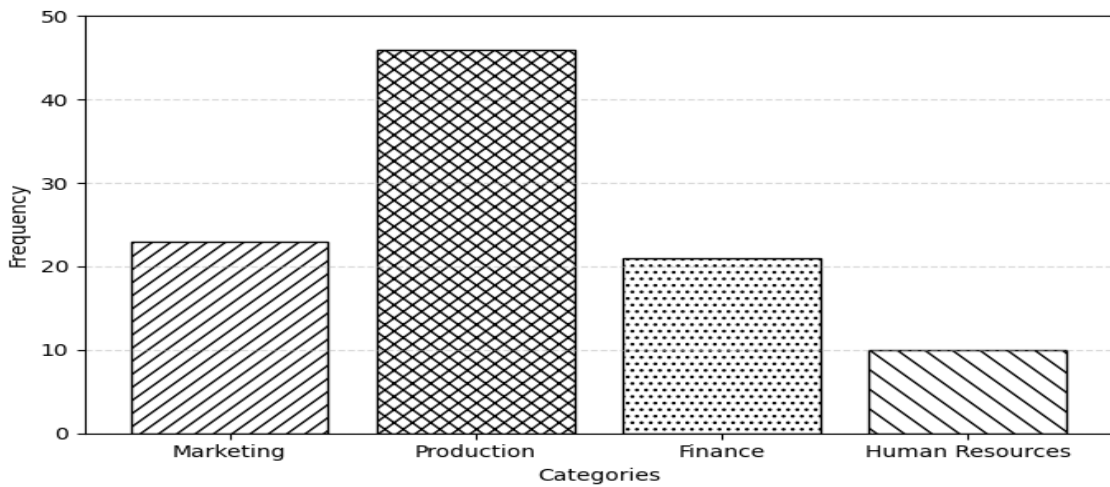
Graphical representation refers to the transformation of statistical data into graphical forms and charts that help highlight the characteristics of the studied phenomenon in a clear and rapid manner, and facilitate the easy understanding of its general trends. It also allows for comparisons between different variables or across multiple time periods, and assists in revealing patterns, variations, and essential differences that may not appear through numerical tables alone. Among the most widely used graphical tools in statistical analysis are histograms, frequency polygons, frequency curves, bar and rectangular charts, as well as pie charts.

The proper use of graphical representations contributes to improving the understanding of results, supporting the decision-making process, and simplifying the presentation of data, whether in academic or professional contexts.

Among the methods used to represent data from frequency tables graphically are the following:

1. Histogram (Frequency Histogram):

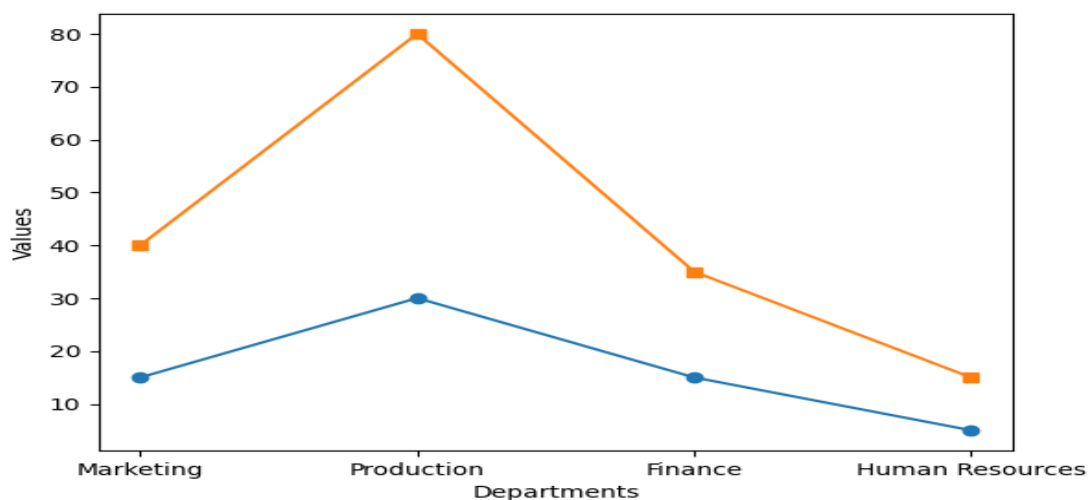
A histogram is the graphical representation of a simple frequency table. It is composed of adjacent bars, each of which corresponds to a specific frequency or class value. The height of each bar reflects the magnitude of the frequency, whereas the bar width corresponds to the size of the related class interval.

Figure (04) graphical representation based on a histogram**2- Frequency Polygon:**

It is also a graphical representation of the simple frequency table, where it consists of a set of connected straight line segments forming a broken line, defined by points whose coordinates are the class midpoints and their corresponding frequencies.

3- Frequency Curve:

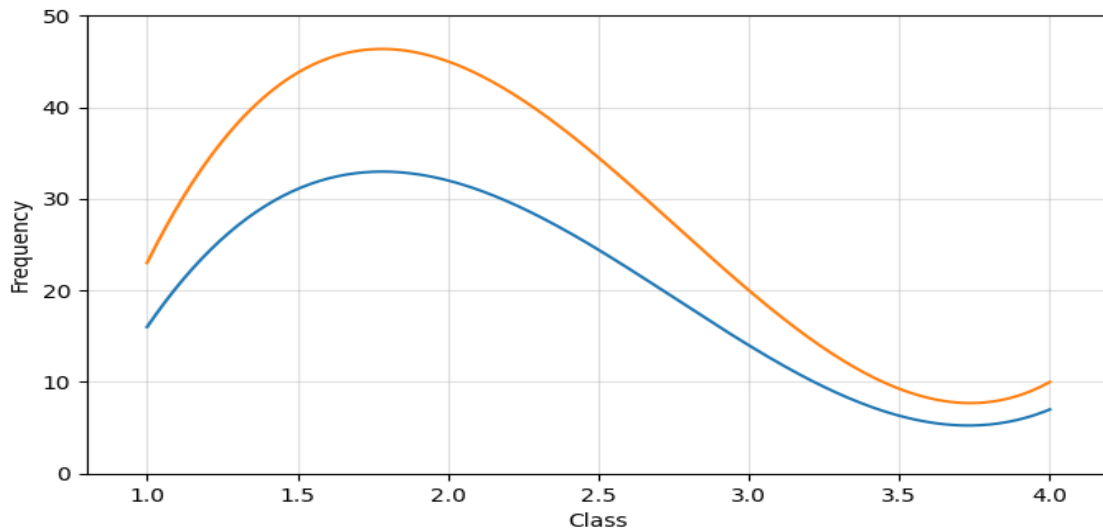
By following the same steps used to draw the frequency polygon, the frequency curve can be constructed. In this case, the broken line is smoothed into a continuous curve that passes through the maximum number of points.

Figure (05) Graphical representation using a frequency polygon.

3- Frequency Curve:

By following the same steps used to draw the frequency polygon, the frequency curve can be constructed. In this case, the broken line is smoothed into a continuous curve that passes through the maximum number of points.

Figure (06) Graphical representation using the frequency curve



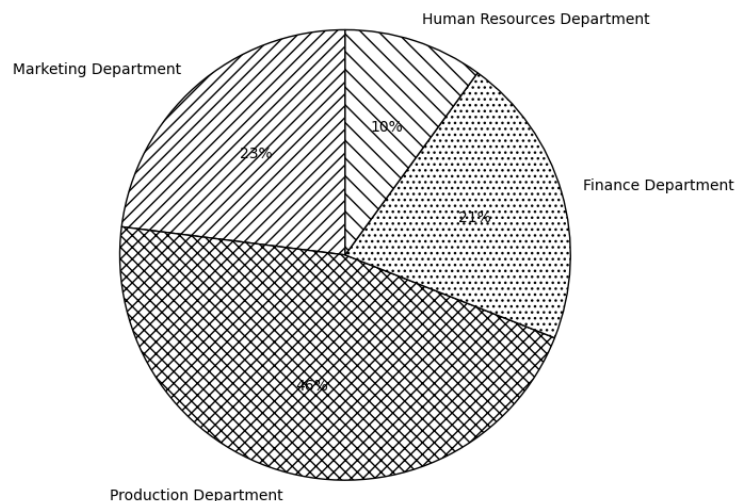
4- Pie Chart:

A pie chart is a circular graphical representation divided into several sectors, where each sector corresponds to a central angle proportional to the frequency of each studied characteristic. In order to construct the pie chart, an additional column is added to the data table containing the central angles associated with each frequency.

The central angles are calculated using the following formula:

$$\text{Central angle} = \frac{n_i}{n} \times 360$$

Figure (07): Graphical representation using a pie chart



5- Advantages and Disadvantages of Graphical Representations:

1- Advantages:

- **Attracting attention:** Graphs attract attention and are more closely associated with memory. Regardless of the interest given to presenting tables, many people tend not to pay much attention to them.
- **Simplicity in reading data:** Especially when the number of observations is large.
- **Ease of remembering results:** It is well known that graphical representations provide a more lasting impression than numbers and words.

1- Disadvantages:

- **Loss of data precision:** Graphs illustrate only general changes and do not show full detailed accuracy. Therefore, it is always preferable to accompany graphs with tables.
- **Possible complexity:** Graphs may sometimes be complex if they include multiple data sets, or costly if they require large-scale measurements for accurate drawing.

1. Presentation of Data in the Case of a Quantitative Variable

1.1. Presentation of Data in the Case of a Discrete Quantitative Variable:

This refers to the organization of the numerical values of a quantitative variable that takes discrete values in the form of an organized statistical table. Through this process, data are transformed from their initial raw and unordered state into a clearer and more readable format. This organization helps describe the statistical phenomenon under study without

losing any information and enables analysis and the extraction of its main characteristics.

This type of table is usually used to present and interpret data related to a single variable only. In its simplest form, it consists of two basic columns:

1. Values of the statistical variable:

These are displayed in the first column of the table and correspond to the different values assumed by the variable under study. The values are arranged in ascending order to facilitate reading and analysis. Each value of the variable is represented by the symbol x_i , where the index indicates the row number or the position of the value within the table.

($x_1, x_2, x_3 \dots \dots x_k$)

2. Absolute frequency:

This appears in the second column of the table and expresses the number of times each value of the statistical variable occurs within the observed data. It is represented by the symbol n_i , where the index refers to the value of the variable to which it corresponds ($n_1, n_2, n_3 \dots \dots n_k$)

Values of the statistical variable x_i	Absolute frequency n_i
x_1	n_1
x_2	n_2
x_3	n_3
.	.
.	.
.	.
x_k	n_k
Total	$n. = \sum n_i$

This table represents the most basic and simplest form of frequency distribution, as it includes only the values of the statistical variable and their corresponding absolute frequencies. However, it can be expanded to incorporate additional information of greater relevance to the study.

Example:

The following data represent the number of siblings for a sample of 20 students. The recorded values are as follows:

2, 3, 1, 4, 2, 0, 1, 3, 2, 4, 1, 2, 3, 0, 2, 1, 4, 3, 2, 1

Required:

- Identify the different values taken by the variable “number of siblings” and arrange them in ascending order.
- Calculate the absolute frequency of each value.
- Organize the results in a simple frequency distribution table including:
 - The values of the statistical variable x_i
 - The corresponding absolute frequency for each value n_i .
- Verify that the sum of the frequencies equals the total number of individuals in the sample.

Solution:

- **Identification and ordering of the different values:**

The distinct values of the variable are:

0 – 1 – 2 – 3 – 4

- **Calculation of frequencies (number of occurrences of each value):**
- The value 0 occurs 2 times.
- The value 1 occurs 5 times.
- The value 2 occurs 6 times.
- The value 3 occurs 4 times.
- The value 4 occurs 3 times.

Simple Frequency Distribution Table:

Number of siblings x_i	Number of students n_i
0	2
1	5
2	6
3	4
4	3
Total	20

Interpretation examples:

- n_3 : There are 6 students out of 20 whose number of siblings is equal to 2.
- n_4 : There are 4 students out of 20 whose number of siblings is equal to 3.

2. Relative and Percentage Frequency Distribution:

In many situations, it is preferable to express the frequency distribution in percentage form rather than relying solely on absolute frequencies, as this helps emphasize the relative weight of each variable value compared to the total number of observations. The relative frequency f_i is calculated by dividing the absolute frequency of each value by the total sum of all frequencies :

$$f_i = \frac{n_i}{\sum n_i} \quad \sum f_i = 01$$

The use of relative frequency enhances data clarity and allows for easier comparisons between values, particularly when absolute frequencies differ across variables or samples. It also facilitates the graphical presentation of results within a unified statistical report.

The relative frequency f_i can be converted into a percentage relative frequency in order to express the proportion of each value relative to the total frequencies in percentage form. This is done by multiplying the relative frequency by 100, and it is usually denoted by $f_i \%$:

$$f_i \% = f_i \times 100 = \frac{n_i}{\sum n_i} \times 100$$

Example:

Based on the previous example, calculate the relative frequency and the percentage relative frequency, and interpret f_2 and $f_4 \%$

Number of siblings x_i	Number of students n_i	f_i	$f_i \%$
0	2	0.1	10%
1	5	0.25	25%
2	6	0.30	30%
3	4	0.20	20%

4	3	0.15	15%
Total	20	1	100%

Interpretation:

- $f_2 = 25\%$ of the students have one sibling.
- f_4 %: 15% of the students have four siblings.

Cumulative Frequency Distribution:

In certain cases, it is necessary to determine the number of observations whose values are less than or equal to a specific limit, or greater than that limit. Such information plays an important role in statistical analysis, especially when examining distributions and making decisions based on threshold values. These data are obtained by calculating cumulative frequencies of two types:

Ascending Absolute Cumulative Frequency Distribution:

The ascending cumulative frequency is defined as the sum of the frequencies of statistical values that are less than or equal to a given value. It is denoted by $N_i\uparrow$ and is calculated progressively by adding the frequencies from the top of the table downward.

$$\begin{aligned}
 N_1\uparrow &= n_1, \\
 N_2\uparrow &= n_1 + n_2 = N_1\uparrow + n_2, \\
 N_3\uparrow &= n_1 + n_2 + n_3 = N_2\uparrow + n_3, \\
 &\vdots \\
 N_i\uparrow &= n_1 + n_2 + \dots + n_i = N_{i-1}\uparrow + n_i
 \end{aligned}$$

Descending Absolute Cumulative Frequency Distribution:

The descending cumulative frequency represents the total number of observations whose values exceed a specified threshold. It is symbolized by $N_i\downarrow$ and is determined sequentially, starting from the highest values and moving downward by subtracting frequencies step by step.

$$\begin{aligned}
 N_{1\downarrow} &= n, \\
 N_{2\downarrow} &= n - n_1 = N_{1\downarrow} - n_1, \\
 N_{3\downarrow} &= n - n_1 - n_2 = N_{2\downarrow} - n_2, \\
 &\vdots \\
 N_{i\downarrow} &= n - n_1 - n_2 - \cdots - n_{i-1} = N_{i-1\downarrow} - n_{i-1}
 \end{aligned}$$

Example:

Based on the data of the previous example, compute the ascending and descending absolute cumulative frequencies, and interpret $N_{4\downarrow}$ and $N_{5\uparrow}$

x_i	n_i	$N_i\uparrow$	$N_i\downarrow$
0	2	2	20
1	5	2+5=7	20-2=18
2	6	7+6=13	18-5=13
3	4	13+4=17	13-6=7
4	3	17+3=20	7-4=3
Total	20		

Interpretation:

- $N_{4\downarrow}$: The number of students whose number of siblings is greater than or equal to 3 is 7 students.
- $N_{3\uparrow}$: The number of students whose number of siblings is less than or equal to 4 is 20 students.

Relative ascending cumulative frequency:

It is calculated following the same procedure as the ascending absolute cumulative frequency, but it is based on relative frequencies f_i rather than absolute frequencies n_i

$$F_i\uparrow = \frac{N_i\uparrow}{\Sigma n_i}$$

$$F_i\uparrow = F_{i-1}\uparrow + f_i$$

Percentage relative ascending cumulative frequency:

It is obtained by multiplying the relative ascending cumulative frequency by 100:

$$F_i\uparrow\% = F_i\uparrow \times 100$$

Relative descending cumulative frequency:

It is computed using the same approach as the descending absolute cumulative frequency, but relies on relative frequencies f_i instead of absolute frequencies. n_i .

$$F_i\downarrow = \frac{N_i\downarrow}{\Sigma n_i}$$

Percentage relative descending cumulative frequency:

It is obtained by multiplying the relative descending cumulative frequency by 100:

$$F_i\downarrow\% = F_i\downarrow \times 100$$

Example:

Based on the data of the previous example, compute the relative cumulative frequencies and the percentage relative cumulative frequencies, and interpret $F_i\uparrow\%$ and $F_i\downarrow\%$

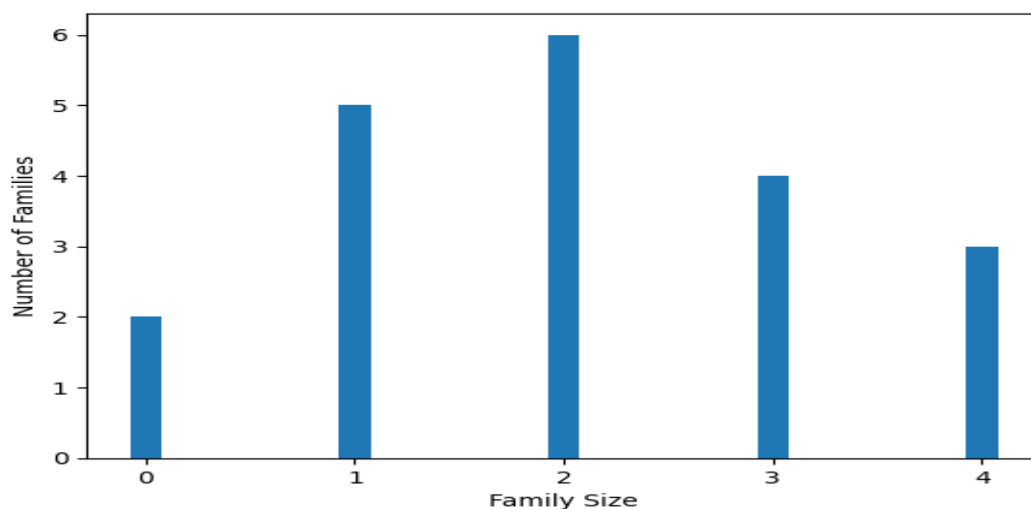
x_i	n_i	F_i	$F_i\%$	$F_i\uparrow$	$F_i\downarrow$	$F_i\uparrow\%$	$F_i\downarrow\%$
0	2	0.1	10%	0.1	1	10%	100%
1	5	0.25	25%	0.35	0.9	35%	90%
2	6	0.30	30%	0.65	0.65	65%	65%
3	4	0.20	20%	0.85	0.35	85%	35%
4	3	0.15	15%	1	0.15	100%	15%
Total	20	1	100%	/	/	/	/

- $F_3\uparrow\%$: There are 65% of the students whose number of siblings is less than or equal to 2.
- $F_4\downarrow\%$: There are 35% of the students whose number of siblings is greater than or equal to 4.

2. Graphical Representation of the Absolute and Relative Frequency Distribution of a Discrete Quantitative Variable:

The frequency distribution—whether absolute or relative—of a discrete quantitative variable can be illustrated using bar charts. This representation is achieved by placing the values of the variable along the horizontal axis and drawing vertical bars for each value, where the height of each bar corresponds proportionally to its frequency on the vertical axis. This form of graphical representation helps clarify the distribution of the data and visually emphasizes the differences between the variable values in a clear and accessible manner

Figure (08): Simple bar chart representing the number of siblings among 20 students.



3. Graphical Representation of the Absolute and Relative Frequency Distributions of a Continuous Quantitative Variable:

The ascending cumulative absolute or relative frequency is represented graphically by ascending straight line segments corresponding to the increasing cumulative frequencies associated with each value of the statistical variable under study.

As for the descending cumulative frequency, it is represented graphically by descending straight line segments that reflect the decreasing cumulative frequencies. The first line segment corresponds to the total frequency and the smallest value of the studied variable, while the second segment represents the total frequency minus the first simple frequency, combined with the second value of the statistical variable, and the process continues accordingly.

Example :

Based on the data provided in the previous example, graphically represent the ascending and descending cumulative frequency distributions.

.Solution :

Figure (09): Graphical representation of the ascending cumulative frequency of the distribution of the number of siblings among 20 students.

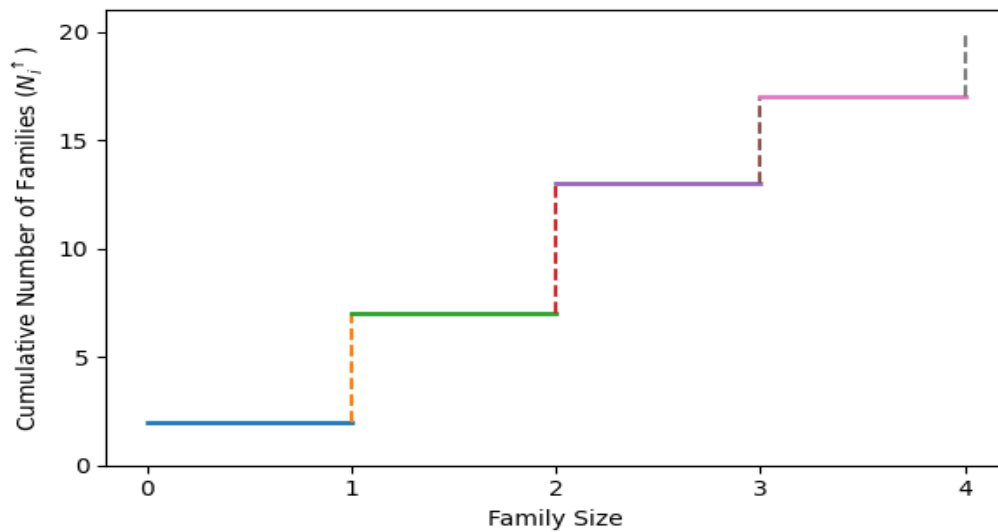
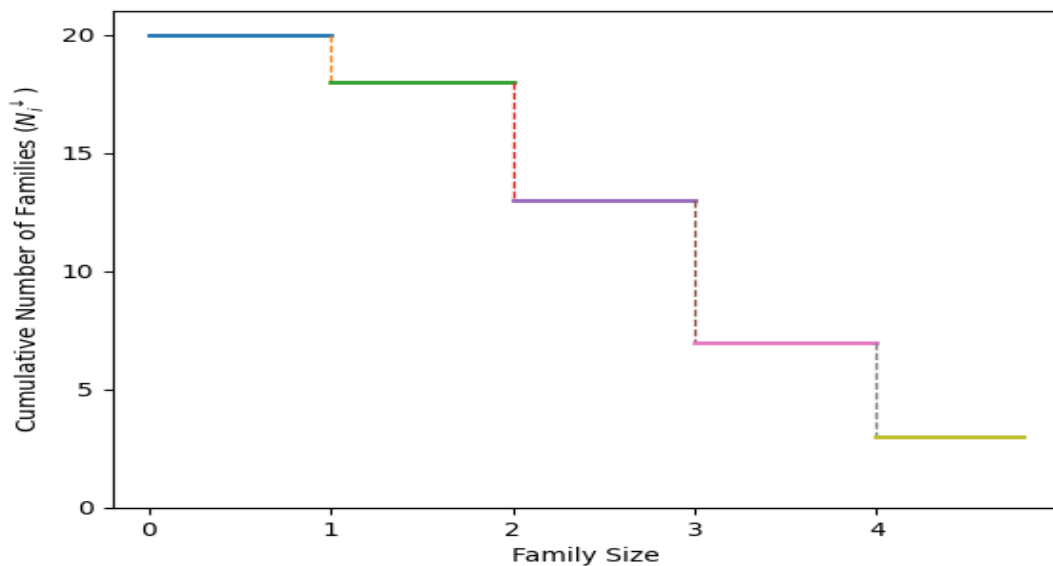


Figure (10) : Graphical representation of the descending cumulative frequency of the distribution of the number of siblings among 20 students.



4. Presentation of Data in the Case of a Continuous Quantitative Variable:

Continuous quantitative variables are among the most frequently encountered types of variables in statistical studies. They may assume integer or fractional numerical values and can take any value within a given interval, owing to their continuous nature without interruptions between values.

Absolute Frequency Distribution:

Since a continuous quantitative variable can take an unlimited number of values within a specific range, presenting these values individually becomes complex and difficult to read when compared to a discrete quantitative variable. Therefore, the data are grouped into classes representing successive intervals of the overall data range. Based on these classes, a frequency distribution table is constructed, which clearly and systematically displays the number of observations falling within each class.

Determining the Number of Classes:

Using a small number of classes in a frequency distribution table simplifies calculations; however, it simultaneously reduces the level of accuracy and may result in the loss of some information.. Conversely, increasing the number of classes enhances accuracy and provides a clearer representation of the data, but it requires greater computational effort. Consequently, there is no strict theoretical rule for determining the number of classes; rather, it depends on the nature of the phenomenon under study and the researcher's objective.

Nevertheless, it is generally recommended that the number of classes should not be fewer than five, so that the table does not lose essential details, and should not exceed fifteen, so as to avoid excessive complexity and difficulty of analysis (Ben Farhat & Qattoush, 2014, p. 30).

To construct frequency distribution tables for continuous quantitative variables, it is necessary to determine both the number of classes and the class width. In this regard, the following two formulas may be used as a guideline (Abdeljawad, 2013, p. 52):

Several researchers have proposed approximate formulas to help determine an appropriate number of classes. Among the most important are the following:

Yule's Formula:

This formula is applied when the number of observations is less than one thousand. The following steps are applied:

Determining the number of classes using the following formula:

$$K = 2.5 \sqrt[4]{n}$$

Where:

K: number of classes

n: number of observations

- Determining the class width by subtracting the minimum value from the maximum value to obtain the range, then dividing the result by the number of classes.

Sturges' Formula:

This formula is used when the number of observations exceeds one thousand:

$$K = 1 + 3.322 \log(n)$$

Where:

K: number of classes

n: number of observations

log: common (base-10) logarithm

These rules are considered approximate rather than mandatory; however, they provide a practical framework that assists the researcher in selecting a balanced number of classes that ensures both clarity and accuracy in data presentation..

Class Width:

To determine the class width in a frequency distribution table for a continuous quantitative variable, the total range of the data is divided by the chosen number of classes.

The total range is defined as the difference between the maximum and minimum values of the variable:

$$\text{Class width} = \frac{\text{Range}}{\text{Number of classes}}$$

$$L = \frac{R}{K}$$

Range = Maximum value – Minimum value

$$R = X_{\max} - X_{\min}$$

This calculation allows the data to be divided into classes of equal width, which facilitates organization and analysis and ensures clarity in tabular presentation as well as balance in graphical representation.

It is generally preferable to use equal-width classes when constructing a frequency distribution table for continuous quantitative variables, as this approach supports effective analysis and comparison between classes. However, in certain situations, unequal-width classes may be used.

The most common situations include:

- **Focusing on specific classes:** If the objective of the study is to give greater importance to certain classes while giving less attention to others, the less important classes may be combined into a single class in order to simplify the presentation.
- **Variation in frequencies:** If the frequencies of some classes are very small compared to others, these classes may be merged to avoid scattered classes with very low frequencies.

Accordingly:

- A frequency distribution composed of classes with equal widths is referred to as a regular frequency distribution.
- A frequency distribution composed of classes with unequal widths is referred to as an irregular frequency distribution

Determining Class Limits:

Each class in a frequency distribution is characterized by a lower limit and an upper limit. In most cases, the upper limit of a class is considered an apparent (or nominal) limit rather than an actual included value. For example, if the class is [a, b], then b is treated as the apparent upper limit of the class.

Classes are defined based on a predetermined width. The first class begins at the minimum value of the variable X_{\min} , while the upper limit of the final class may exceed the maximum

observed value X_{\max} , in order to ensure that all observations are included within the frequency distribution table.

Class Midpoint:

The class midpoint represents the central value of a class and is located exactly halfway between its lower and upper limits. It is calculated by summing these two limits and dividing the result by two, and it is denoted by the symbol C :

$$\text{Class midpoint} = \frac{\text{Upper class limit} + \text{Lower class limit}}{2} =$$

Determining the Number of Frequencies or the Number of Values in Each Class:

After defining the classes in a frequency distribution, the number of observations that fall within each class is calculated. This quantity is referred to as the absolute frequency and is denoted by n_i . When grouping data into classes, it is necessary to ensure that each observation belongs to one and only one class, so as to avoid overlap and maintain the accuracy of the frequency distribution.

Example:

The following data represent the statistics test scores of 40 students, as listed below:
45, 46, 47, 48, 48, 49, 50, 50, 51, 51, 52, 52, 53, 53, 54, 54, 55, 55, 56, 56, 57, 57, 58, 58, 59, 59, 60, 60, 61, 61, 62, 62, 45, 46, 47, 48, 49, 50, 51, 52

Required:

Construct the frequency distribution table.

Solution:

To construct the frequency distribution table, the following steps are followed:

Determining the number of classes:

To determine the number of classes, either **Yule's formula** or **Sturges' formula** can be applied.

Yule's Formula:

$$K = 2.5 \sqrt[4]{n}$$

$$K = 2.5 \sqrt[4]{40}$$

$$K = 2.5 \sqrt{6.32}$$

$$K = 2.5 \times 2.51$$

$$K = 6.29 \approx 6$$

Sturges' Formula:

$$K = 1 + 3.322 \log(n)$$

$$K = 1 + 3.322 \times 1.60$$

$$K = 6.32 \approx 6$$

Calculation of Class Width:

To determine the class width, the following formula is applied:

$$L = \frac{R}{K}$$

Before calculating the class width, the range must be computed:

$$R = X_{\max} - X_{\min}$$

$$R = 62 - 45 = 17$$

$$L = \frac{17}{6} = 2.83 \approx 3$$

Frequency Distribution Table:

Classes	Values belonging to the class	Class intervals X_i	Frequency n_i	Class midpoint C_i
F1	45, 46, 47]48 – 45]	6	$C1 = \frac{(45 + 48)}{2} = 46.5$
F2	48, 49, 50]51 – 48]	8	$C2 = \frac{(48 + 51)}{2} = 49.5$
F3	51, 52, 53]54 – 51]	8	$C3 = \frac{(51 + 54)}{2} = 52.5$
F4	54, 55, 56]57 – 54]	6	$C4 = \frac{(54 + 57)}{2} = 55.5$
F5	57, 58, 59]60 – 57]	6	$C5 = \frac{(57 + 60)}{2} = 58.5$
F6	60, 61, 62]63 – 60]	6	$C6 = \frac{(60 + 63)}{2} = 61.5$

Interpretation:

- n_5 : There are 6 students whose scores range between 57 and 59.

Cumulative and Relative Frequency Distribution:

The ascending and descending cumulative frequencies, relative frequencies, percentage relative frequencies, as well as cumulative relative frequencies for classes are determined using the same procedures applied to discrete quantitative variables, taking into consideration that each class represents a group of values rather than a single observation.

Example:

In a methodology test involving 50 students, the scores were distributed as follows:

Table (01): Distribution of the scores of 50 students in Methodology

Scores	n_i
]4 – 0]	5
]8 – 4]	8
]12 – 8]	14
]16–12]	15
]20–16]	8
Total	50

Required:

Calculate the absolute frequencies, relative frequencies, and percentage relative frequencies, and present them.

Solution:**Table (02): Distribution of absolute frequencies, relative frequencies, and percentage relative frequencies for the scores of 50 students in Methodology**

Classes	n_i	f_i	$f_i\%$	$N_i\uparrow$	$N_i\downarrow$	$F_i\uparrow$	$F_i\downarrow$	$F_i\uparrow\%$	$F_i\downarrow\%$
]4 – 0]	5	0.1	10%	5	50	0.1	1	10%	100%
]8 – 4]	8	0.16	16%	13	45	0.26	0.9	26%	90%
]12 – 8]	14	0.28	28%	27	37	0.54	0.74	54%	74%
]16–12]	15	0.3	30%	42	23	0.84	0.46	84%	46%
]20–16]	8	0.16	16%	50	8	1	0.16	100%	16%
Tota	50	1	100%	50	50	1	1	100%	100%

Interpretation:

- $N_4\uparrow$: The number of students whose scores are less than or equal to 16 is 42 students.
- $N_3\downarrow$: The number of students whose scores are greater than or equal to 8 is 37 students.
- $F_4\uparrow$: The percentage of students whose scores are less than or equal to 8 is 26%.
- $F_3\downarrow$: The percentage of students whose scores are greater than or equal to 16 is 16%.

5. Graphical Representation of the Frequency Distribution for a Continuous Quantitative Variable:

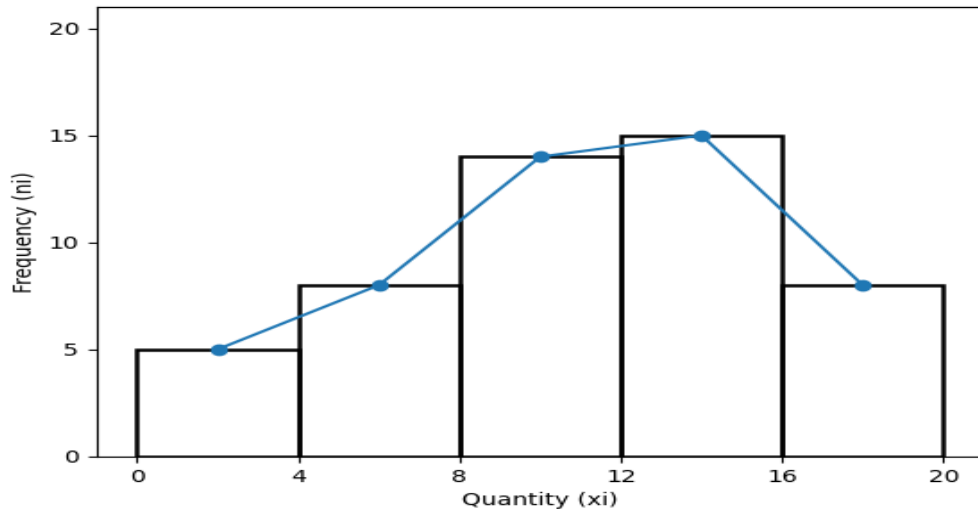
The absolute or relative frequency distribution of a continuous quantitative variable, when equal-width classes are applied, is commonly represented using a histogram.

- This type of chart consists of adjacent rectangles, where each rectangle corresponds to a single class, and the area of each rectangle is proportional to the frequency of that class.
- The horizontal axis represents the class intervals, while the vertical axis corresponds to the frequencies.

If the midpoints of the classes are connected by straight line segments, we obtain the Frequency Polygon:

- It is a closed polygon formed by plotting points whose coordinates are given by the class midpoint on the horizontal axis and the corresponding frequency on the vertical axis.
- These points represent the vertices of the polygon, and they are joined by straight lines to form the frequency polygon.

Figure (11) Graphical representation of the distribution of the scores of 50 students in the methodology test using a histogram and a frequency polygon.

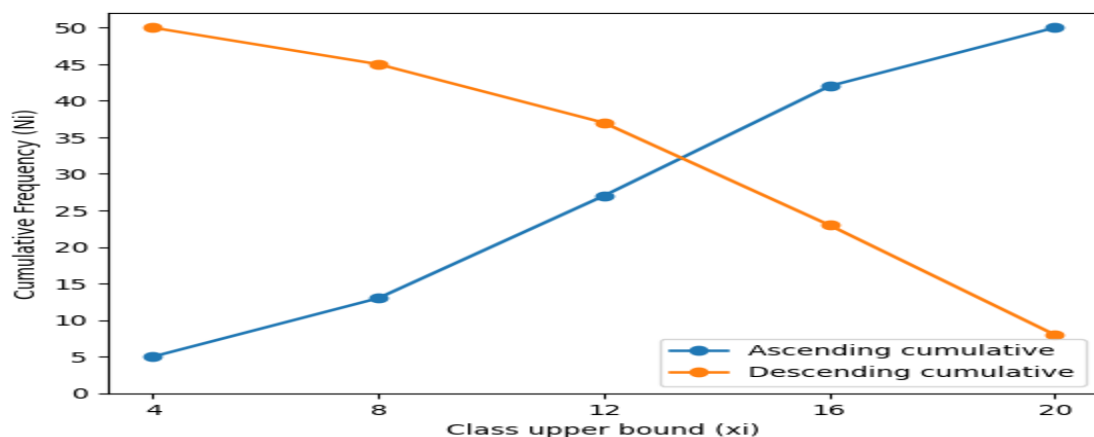


6. Graphical Representation of the Cumulative Frequency Distribution:

The ascending cumulative frequency is illustrated graphically by a curve obtained through plotting a series of points whose coordinates correspond to the upper class limits on the horizontal axis and the associated ascending cumulative frequencies on the vertical axis. These points are then connected using straight or smooth curves to form the cumulative frequency curve.

As for the descending cumulative frequency, it is also represented by a curve constructed by plotting points whose coordinates consist of the lower class limits on the horizontal axis and the corresponding descending cumulative frequencies on the vertical axis. These points are subsequently connected to obtain the descending cumulative frequency curve (Jelatto, 2010, p. 26)

Figure (12) Graphical representation of the cumulative frequency of the distribution of the scores of 50 students in the methodology test.



7. Presentation of Data in the Case of an Ordinal Qualitative Variable:

Ordinal qualitative variables are variables that are not expressed in numerical values, but rather take the form of descriptive categories or verbal attributes. However, they are characterized by the possibility of being arranged according to a specific logical or hierarchical order.

Frequency Distribution of an Ordinal Qualitative Variable:

To construct a frequency distribution table for ordinal qualitative data, a table is prepared that includes at least two basic columns: one dedicated to the categories of the variable arranged according to a logical order, and the other dedicated to the absolute frequency of each category. The table may be expanded by adding the relative frequency and the percentage relative frequency, as well as the ascending and descending cumulative frequencies in both absolute and relative forms. These indicators provide more detailed information about the distribution of the data.

Example:

The administration of the Faculty of Social and Human Sciences conducted a survey of the opinions of 60 students regarding their attitude toward the e-learning platform used by the faculty. The responses were as follows:

Strongly disagree, Neutral, Agree, Strongly agree, Neutral, Strongly disagree, Agree, Neutral, Strongly agree, Agree, Neutral, Disagree, Neutral, Agree, Strongly agree, Neutral, Disagree, Neutral, Agree, Strongly agree, Strongly disagree, Neutral, Agree, Agree, Neutral, Disagree, Strongly agree, Neutral, Agree, Strongly disagree, Neutral, Agree, Strongly agree, Neutral, Disagree, Neutral, Agree, Strongly agree, Neutral, Strongly disagree, Neutral, Agree, Agree, Neutral, Disagree, Strongly agree, Neutral, Agree, Strongly agree, Neutral, Disagree, Neutral, Agree, Strongly agree, Neutral, Strongly disagree, Neutral, Agree.

Required:

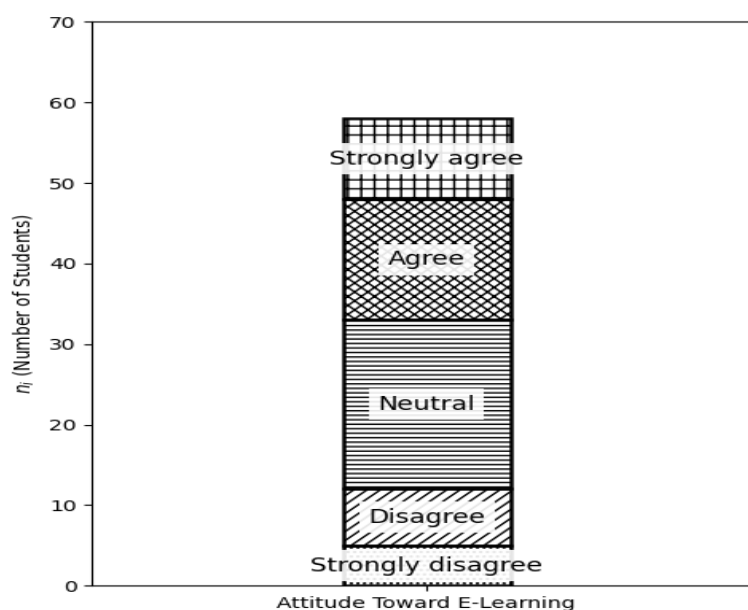
Prepare a frequency distribution table using absolute frequency, relative frequency, percentage relative frequency, and ascending and descending cumulative frequencies in both absolute and relative forms.

Solution:**Table (03): Distribution of the attitudes of 60 students toward the e-learning platform used by the faculty**

Categories	n_i	f_i	$f_i\%$	$N_{i\uparrow}$	$N_{i\downarrow}$	$F_{i\uparrow}$	$F_{i\downarrow}$	$F_{i\uparrow}\%$	$F_{i\downarrow}\%$
Strongly disagree	5	0.08	8.62%	5	58	0.08	1	8.62%	100%
Disagree	7	0.12	12.07%	12	53	0.20	0.914	20.69%	91.38%
Neutral	21	0.36	36.21%	33	46	0.56	0.793	56.90%	79.31%
Agree	15	0.25	25.86%	48	25	0.82	0.431	82.76%	43.10%
Strongly agree	10	0.17	17.24%	58	10	1	0.172	100%	17.24%
Total	58	1	100%	-	-	-	-	-	-

8. Graphical Representation of an Ordinal Qualitative Variable:

The graphical representation of an ordinal qualitative variable is known as the segmented bar chart. It consists of a rectangle divided into several segments, where each segment corresponds to a specific frequency of the studied attribute. When drawing the segmented bar, it is preferable to use the percentage relative frequencies corresponding to each category, since the total length of the bar represents 100% (Jelatto, 2010, p. 21).

Figure (13) Graphical representation of the distribution of the attitudes of 50 students toward e-learning using a stacked bar chart.

9. Presentation of Data for a Nominal (Non-Ordinal) Qualitative Statistical Variable:

Frequency Distribution:

When analyzing a nominal qualitative variable (that is, a variable that cannot be ordered), the frequency distribution table is restricted to listing the categories of the variable in the first column, along with the absolute frequency, relative frequency, and percentage relative frequency for each category. Ascending or descending cumulative frequencies are not applied in this case, as they do not have statistical meaning for non-ordered qualitative variables.

Example:

A researcher studied the preferred recreational activities of 50 university students and recorded the following responses:

Practicing sports, watching movies, reading, video games, watching movies, arts, watching movies, video games, reading, practicing sports, watching movies, practicing sports, video games, watching movies, arts, watching movies, practicing sports, video games, reading, watching movies, video games, practicing sports, arts, watching movies, reading, practicing sports, watching movies, video games, arts, watching movies, practicing sports, watching movies, reading, video games, watching movies, arts, practicing sports, video games, watching movies, reading, watching movies, practicing sports, video games, arts, watching movies, reading, watching movies, practicing sports, video games, watching movies.

Required:

Prepare a frequency distribution table using absolute frequency, relative frequency, and percentage relative frequency.

Solution:

Recreational activity	n_i	f_i	$f_i\%$
Watching movies	15	0.30	30%
Practicing sports	12	0.24	24%
Reading	8	0.16	16%
Video games	10	0.20	20%
Arts	5	0.10	10%
Total	50	1.00	100%

10. Graphical Representation of a Nominal (Non-Ordinal) Qualitative Variable:

A nominal qualitative variable can be represented graphically using either a pie chart or independent (separate) bar charts.

Pie Chart:

A pie chart consists of a circle divided into several sectors, where each sector corresponds to a specific category of the variable. The size of each sector is proportional to the frequency or percentage of that category in relation to the total dataset.

To determine the central angle associated with each category, the following formula is applied:

$$\text{Central angle} = \frac{n_i}{n} \times 360$$

Using this method allows the proportions of each category to be represented visually within the circle in a clear and accurate manner.

Independent Bar Charts:

Independent bar charts are composed of separate rectangles with equal base widths and uniform spacing between them. The height of each bar is proportional to the frequency corresponding to the category it represents.

Example:

Based on the data from the previous example, represent graphically the given frequency distribution.

Solution:

Since the statistical variable under study is a nominal qualitative variable, it may be represented either by a pie chart or by independent bar charts..

Using the pie chart:

First, the central angles are calculated as follows:

- Central angle for Watching movies

$$= \frac{15}{50} \times 360 = 108$$

- Central angle for Practicing sports

$$= \frac{12}{50} \times 360 = 86$$

- Central angle for Reading

$$= \frac{08}{50} \times 360 = 58$$

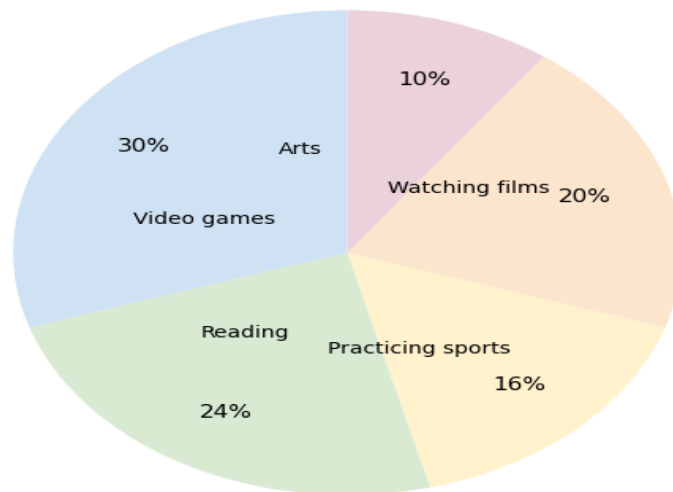
- Central angle for Video games

$$= \frac{10}{50} \times 360 = 72$$

- Central angle for Arts

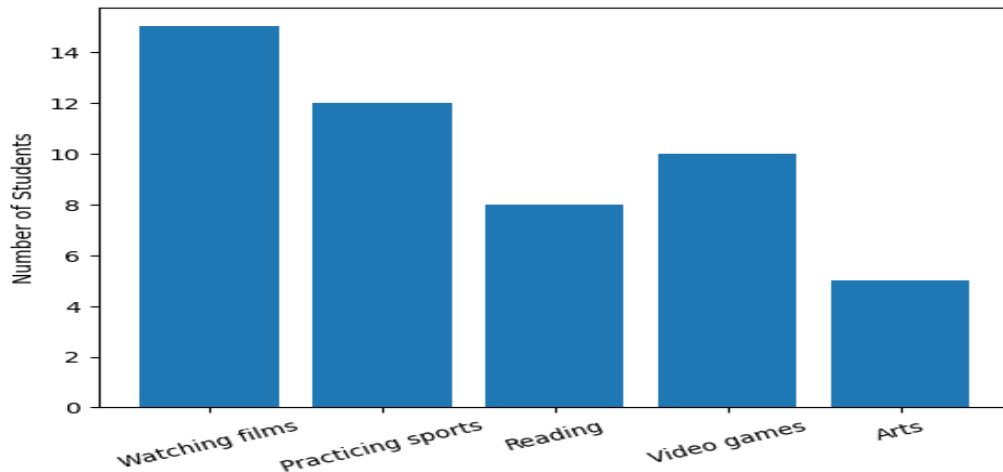
$$= \frac{5}{50} \times 360 = 86$$

Figure (14): Pie chart representing the relative distribution of preferred recreational activities among 50 university students.



Using Independent Bar Charts:

Figure (15): Independent bar chart representing the preferred recreational activities of 50 university students.



Suggested exercises :

Exercise 1:

The following data represent the number of absences recorded for a sample of 30 students during a semester:

2, 0, 1, 3, 2, 1, 4, 0, 2, 3, 1, 2, 0, 1, 3, 2, 4, 1, 2, 3, 0, 1, 2, 3, 4, 2, 1, 0, 2, 3

Required:

1. Identify the different values taken by the variable “number of absences” and arrange them in ascending order.
2. Calculate the absolute frequency of each value.
3. Construct a simple frequency distribution table.
4. Interpret n_2 .
5. Calculate the relative frequency f_i for each value.
6. Calculate the percentage relative frequency f_i
7. Interpret f_1 and f_3
8. Calculate the ascending absolute cumulative frequency $N_i\uparrow$.
9. Calculate the ascending absolute cumulative frequency $N_i\uparrow$.
10. Calculate the descending absolute cumulative frequency $N_i\downarrow$.
11. Interpret $N_3\uparrow$ and $N_2\downarrow$
12. Calculate the ascending and descending relative cumulative frequencies, followed by the percentage cumulative frequencies

Exercise 2:

Using the frequency distribution table:

1. Represent the absolute frequency distribution using bar charts.
2. Represent the relative frequency distribution using bar charts.
3. Draw the frequency polygon.
4. Comment on the overall shape of the distribution..

Exercise 3:

The following values represent the heights (in cm) of 40 students:

150, 152, 153, 155, 156, 158, 159, 160, 161, 162, 163, 165, 166, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194

Required:

1. Determine the number of classes using Sturges' formula.
2. Calculate the range and the class width.
3. Construct the frequency distribution table.
4. Calculate the class midpoints.

Exercise 4:

A survey was conducted on the level of satisfaction of 40 employees. The responses were as follows:

Very dissatisfied, dissatisfied, neutral, satisfied, very satisfied (with different frequencies).

Required:

1. Construct a frequency distribution table including n_i, f_i .
2. Calculate the ascending cumulative frequencies.
3. Represent the results using a segmented bar chart.

Exercise 5:

A researcher studied the preferred means of transportation of 60 students (bus, car, walking, bicycle).

Required:

1. Construct a frequency distribution table.
2. Calculate the central angles.
3. Represent the data using a pie chart and independent bar charts.

Exercise 6 :

Data were collected on the number of weekly study hours of a sample of 50 students.

Required:

1. Identify the type of the statistical variable.
2. Construct an appropriate frequency distribution table (absolute, relative, and percentage).
3. Calculate the ascending and descending cumulative frequencies.
4. Represent the data graphically using:
 - A histogram
 - A frequency polygon
 - A cumulative frequency curve
5. Interpret the results statistically and provide a general conclusion.

When a researcher completes the process of data collection, classification, tabulation, and presentation in statistical tables and charts, the next step involves examining the characteristics of the data and drawing conclusions. This is accomplished by the application of a set of statistical measurements known as measures of central tendency.

In this section, we will first address the measures of central tendency by focusing on their main characteristics, methods of calculation, and a comparison of their respective applications.

1. Concept of Central Tendency

Central tendency refers to the leaning of individual notes or statistics values to cluster or concentrate around a specific numerical value within a distribution. Accordingly, this central value around which the data are grouped is considered representative of the outstanding values and helps as a means of describing the data and emphasizing the most important characteristics of the phenomenon under study.

Therefore, measurements of central tendency are sometimes referred to as measures of location or medians, as they represent the values around which data tend to concentrate. These measures include the arithmetic mean, mode, median, geometric mean, harmonic mean, as well as quartiles and percentiles. In what follows, these measures will be presented, with a focus on their relevance and applications within our field of specialization.

In other words, measures of central tendency are among the fundamental statistical tools used to identify the value that represents the central point of a dataset. They provide an indication of the typical or most common value within a distribution, thereby facilitating the summarization and effective description of data. These three primary measurements of the central tendency are the arithmetic mean, median, and mode.

2. Arithmetic Mean

The arithmetic mean, also referred to as the mean, is one of the most well-known and widely used measures of central tendency in statistics. It represents the point of balance of a distribution for any given phenomenon. The arithmetic mean can be defined as:

“A statistical measure used to determine the central value of a set of statistical data. Simply put, it is found by dividing the sum of the data values by their number, indicated by n in the case of a sample and by N in the case of a population”(Al-Saifo et al., 2010, p. 102).

The arithmetic mean of a sample is represented by the symbol \bar{X} (read as: X-bar)

2.1. Uses of the Arithmetic Mean in Social Research

The arithmetic mean is widely used in social and human sciences research because it condenses a large amount of data into a single value that is easy to interpret and compare.

1. Data Summarization (Central Tendency)

The arithmetic mean assists as a measure of central tendency used to provide an overall picture of a dataset through a single value that summarizes the location of data points within a distribution. Examples include the average academic achievement, average income, average age, or average level of anxiety or satisfaction.

This facilitates understanding the general trend of a variable rather than dealing with a long list of numerical values, and it constitutes one of the fundamental tools for comprehending the statistical phenomenon itself (Gravetter & Wallnau, 2017, pp. 84–85).

2. Basis for Measures of Dispersion and Inferential Analysis

The arithmetic mean is not merely a descriptive measure; it forms the core of other statistical measures such as the standard deviation and variance, which assess the extent to which data values are dispersed around the central value (Field, 2018, pp. 100–102).

Moreover, the arithmetic mean serves as an initial element in many inferential statistical tests, such as the t-test and analysis of variance (ANOVA), which are used to compare means across multiple groups (Field, 2018, pp. 520–575).

3. Comparison Between Groups

In social and educational studies, the arithmetic mean is widely used to compare different groups in order to identify differences in variables. For example:

- Comparing the average academic achievement between male and female students.

- Comparing average anxiety levels across different age groups.
- Comparing average performance between experimental and control groups (Gravetter & Wallnau, 2017, pp. 84–85).

Such comparisons enable researchers to determine whether differences in means across groups are statistically significant when appropriate inferential analyses (such as the t-test or ANOVA) are applied (Field, 2018, pp. 520–575).

4. Analysis of General Trends

The arithmetic mean contributes to monitoring general trends of the phenomenon under study over time. For instance, it can be used to track changes in average job satisfaction across several academic terms or to compare average anxiety levels before and after an educational or psychological intervention.

This application is particularly important in analyzing trend changes in educational, psychological, and sociological research (Field, 2018, pp. 520–575).

2.2.Characteristics and Limitations of the Arithmetic Mean

Characteristics:

1. The arithmetic mean accounts for all data values in the distribution.
2. It is easy to calculate, and is characterized by conceptual simplicity and clearness.
3. It is the most usually used measure of central tendency.
4. It is affected by all values within the distribution; therefore, it incorporates all available information about the phenomenon, unlike the median and the mode, which rely on fewer values.
5. It is considered one of the most precise measures of central tendency and is widely used in various statistical computations, particularly in inferential statistics, except in the case of open-ended class distributions.
6. It is relatively stable, as it is only slightly affected by variations in the sample selected from the population.
7. It obeys algebraic and mathematical properties, which makes it a powerful tool for more complex statistical analyses.

Limitations

1. The arithmetic mean is highly sensitive to extreme or outlier values, especially in small datasets. The presence of very large or very small values may cause it to misrepresent the true central tendency of the data.
2. It cannot be calculated for qualitative (categorical) data.
3. It cannot be computed from frequency distributions with open-ended classes, due to the impossibility of determining the midpoints of open intervals.
4. Frequency tables that are open at the beginning or end cannot be used for this calculation, as it relies on class midpoints. It cannot be determined graphically (through visual representation).
5. It cannot be calculated for data measured on nominal or ordinal scales.

2.3.Methods of Calculating the Arithmetic Mean

The arithmetic mean can be calculated for both ungrouped (unclassified) and grouped (classified) data, as follows:

1. Arithmetic Mean for Ungrouped Data

Ungrouped data refer to data presented in the formula of a statistical sequence. If x_1, x_2, \dots, x_k represent the values of the phenomenon under study, then the arithmetic mean of this series is obtained by dividing the sum of these values by their number.

$$\text{Arithmetic Mean} = \frac{\text{Sum of values}}{\text{Number of values}}$$

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_k}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Where the symbol \sum denotes summation.

Example

The following are the exam scores of 8 students in the Applied Statistics course:

09 11 18 15 12 13 16 10

Required:

From the students' exam scores calculate the arithmetic mean.

Solution:

In order to calculate the arithmetic mean, we apply the formula as follows:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\bar{X} = \frac{09 + 11 + 18 + 12 + 15 + 13 + 16 + 10}{8} = \frac{104}{8} = 13$$

Thus, the “arithmetic mean” of the students' scores in the statistics exam is 13 points.

2. Arithmetic Mean for Grouped Data Using a Frequency Distribution**1- Case of a Discrete Quantitative Variable**

In some cases, each value of a variable may have a relative importance known as weights. Ignoring these weights when calculating the arithmetic mean may result in an incorrect measure of “central tendency”. The precision of the mean lean on the distribution of values and their corresponding weights. If we have ungrouped data organized in a frequency table (that is, each value has its corresponding frequency), or if we have weighted data (coefficients), or weighted data, and we wish to calculate the arithmetic mean, we first identify the values $x_1, x_2, x_3, \dots, x_k$ and their corresponding frequencies or weights $n_1, n_2, n_3, \dots, n_k$. The arithmetic mean is then calculated using this exact formula:

$$\bar{X} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^k n_i x_i}{\sum n_i}$$

The arithmetic mean can also be calculated using relative frequencies, according to the following formula:

$$\bar{X} = \sum f_i x_i$$

Example

The following frequency table indicates the number of study hours per day for 40 first-year students:

Number of Hours x_i	Number of Students n_i
0	3
1	6
2	10
3	11
4	7
5	3
Total	40

Required:

Calculate the arithmetic mean of daily study hours using both absolute frequency and relative frequency.

Solution

x_i	n_i	$n_i x_i$	f_i	$f_i x_i$
0	3	0	0.07	0
1	6	6	0.15	0.15
2	10	20	0.25	0.5
3	11	33	0.28	0.84
4	7	28	0.18	0.72
5	3	15	0.07	0.35
Σ .	40	102	1	2.55

Calculation Using Absolute Frequency

$$\bar{X} = \frac{\sum_{i=1}^k n_i x_i}{\sum n_i}$$

$$\bar{X} = \frac{0+6+20+33+28+15}{40} = \frac{102}{40} = 2.55$$

Calculation Using Relative Frequency

$$\bar{X} = \sum f_i x_i = 2.55$$

2. In the Case of a Discrete Quantitative Variable

It is well known that the original individual values cannot be directly identified from a frequency distribution table, since these values are grouped into classes. As a result, the class midpoint represents every value that falls inside its boundaries. As a result, the class midpoint is taken as the anticipated value for each observation belonging to that class.

If k represents the number of classes, and c_1, c_2, \dots, c_k represent the midpoints of these classes, while n_1, n_2, \dots, n_k represent their corresponding frequencies, then the arithmetic mean can be calculated using the following formula:

$$\bar{X} = \frac{n_1 c_1 + n_2 c_2 + \dots + n_k c_k}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^k n_i c_i}{\sum n_i}$$

The arithmetic mean can also be calculated using relative frequencies according to the following formula:

$$\bar{X} = \sum f_i c_i$$

Example

The following frequency table shows the number of daily hours of smartphone use for 60 students:

Daily Smartphone Use Classes	Number of Students
x_i	n_i
[0-2[5
[2-4[12
[4-6[20
[6-8[15
[8-10[8
Total	60

Required:

Calculate the arithmetic mean of daily smartphone use hours based on the above data, using both absolute frequency and relative frequency.

Solution

To calculate the arithmetic mean using the previous formula, the following steps are followed:

- Calculate the sum of frequencies $\sum n_i$
- Determine the class midpoints c_i
- Multiply each class midpoint by its corresponding frequency $n_i c_i$, then compute the total $\sum n_i c_i$
- Calculate the arithmetic mean by applying the formula.

x_i	n_i	c_i	f_i	$c_i n_i$	$f_i c_i$
[0-2[5	1	$\frac{5}{60} = 0.09$	$5 \times 1 = 5$	$0.09 \times 1 = 0.09$
[2-4[12	3	0.2	36	0.6
[4-6[20	5	0.33	100	1.65
[6-8[15	7	0.25	105	1.75
[8-10[8	9	0.13	72	1.17
\sum	60		1	318	5.3

Calculation of the Arithmetic Mean Using Absolute Frequency

$$\bar{X} = \frac{\sum_{i=1}^k n_i c_i}{\sum n_i}$$

$$\bar{X} = \frac{5+36+100+105+72}{60} = \frac{318}{60} = 5.3$$

Calculation of the Arithmetic Mean Using Relative Frequency

$$\bar{X} = \sum f_i c_i = 5.3$$

The average number of daily smartphone usage hours for first-year students is 5.30 hours.

3. Determining the Arithmetic Mean Using Graphical Representations

The position of the arithmetic mean can be approximately identified through graphical displays of quantitative data, particularly histograms. In such representations, the mean is interpreted as a balance point around which the data values are concentrated. This estimation is influenced by the shape of the distribution: when the distribution is symmetric, the mean lies near its center, whereas in a skewed distribution, it shifts in the direction of the

skewness. Graphical estimation is mainly intended to support visual interpretation of data and does not substitute for the exact numerical computation of the mean.

Steps for Graphical Estimation of the Mean

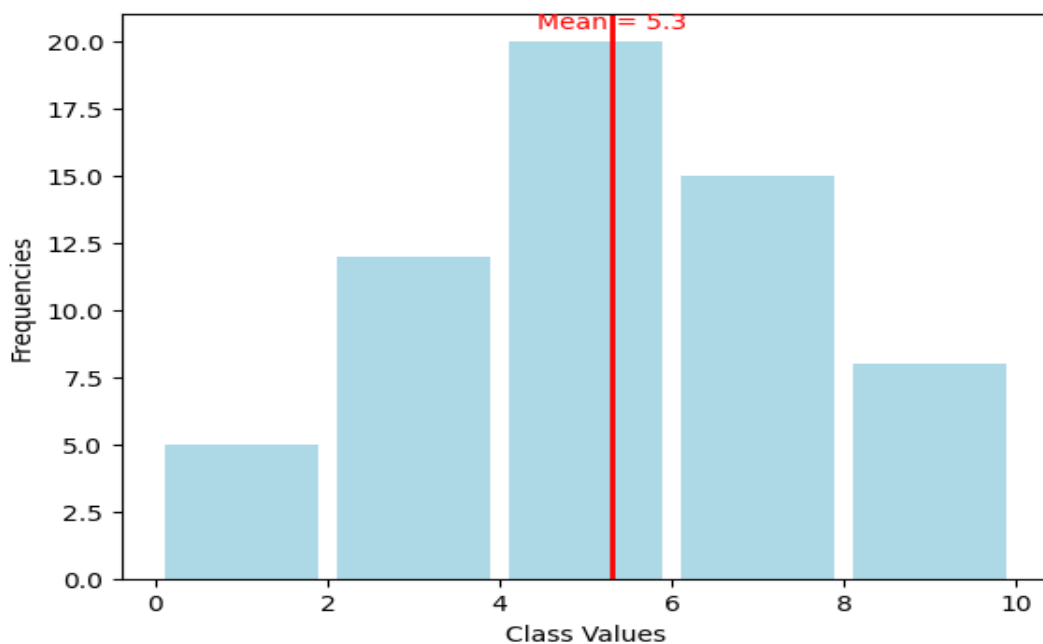
1. Display the data using a histogram.
2. Examine the overall form of the distribution (symmetric or skewed).
3. Locate the balance point where frequencies on both sides appear approximately equal.
4. Project this point onto the horizontal axis.
5. The resulting value represents an approximate location of the arithmetic mean.

It is important to note that graphical estimation of the arithmetic mean is only approximate and is used for explanatory and educational purposes, while precise determination of the mean requires mathematical calculation.

Example:

Based on the data from the previous example, determine the arithmetic mean graphically.

Figure (16): Illustrates the graphical representation of the arithmetic mean.



2.4. Weighted Arithmetic Mean of Means

When numerous groups are joined to form a single group, the weighted arithmetic mean is employed to calculate the overall mean. For example, if we have a first group consisting of n_1 values with an arithmetic mean \bar{X}_1 , and a second group consisting of n_2 values with an arithmetic mean \bar{X}_2 , then the arithmetic mean of the both groups is given by (Haidoushi, 2016, p. 55):

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\bar{X} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{\sum_{i=1}^k n_i}$$

Example

A company has four work teams in a factory producing electronic devices. Each team differs in the number of workers and the average weekly productivity per worker, as shown below:

Team	Number of Workers (n_i)	Average Weekly Productivity per Worker (units/week) \bar{X}_i
1	10	50
2	12	60
3	8	55
4	15	65

Required:

Calculate the overall average weekly productivity of all workers in the factory.

Solution

$$\bar{X} = \frac{(10 \times 50) + (12 \times 60) + (8 \times 55) + (15 \times 65)}{10 + 12 + 8 + 15}$$

$$\bar{X} = \frac{500 + 720 + 440 + 975}{45} = \frac{2635}{45} = 58.55$$

The average weekly productivity of all workers in the factory is 58.55 units.

2.5.Types of Means

Different types of means are used to summarize a set of numerical values, and each type is appropriate for a particular context of use. After the “arithmetic mean”, the “geometric mean” and the “harmonic mean” are among the most essential of these averages. The following section provides a detailed explanation of each.

1.Geometric Mean

The geometric mean of n values of a statistical variable is calculated as the n th root of the product of these values. If n_1, n_2, \dots, n_k represent possible values of the statistical variable, the geometric mean is commonly used when data are relative in nature or related to growth processes. It is represented by the symbol \bar{X}_G .

The fields of application of the geometric mean are relatively limited compared to the arithmetic mean. One of its most important applications is in the economic field, where it is used to calculate average rates of change for certain phenomena, such as population growth rates and interest rates in institutions (Jallatou, 2010, p. 37).

Characteristics of the Geometric Mean

- The geometric mean only applies to positive numbers and cannot be used for zero or negative values.
- It is less impacted by extreme (outlier) values than the arithmetic mean.
- Large numbers of values make computation difficult.

Calculation of the Geometric Mean

1. Unsorted Data

For unsorted data, the geometric mean is the n th root of the product of the values. If we have n values x_1, x_2, \dots, x_n , the geometric mean is computed using this formula (Binya, 2018, p. 70):

$$\bar{X}_G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

To simplify calculations, logarithms are frequently utilized, according to the following formula:

$$\bar{X}_G = 10^{\frac{1}{n} \sum \log x_i}$$

Example

The following are the scores of 8 students in the Applied Statistics course:

09 11 18 15 12 13 16 10

Required:

Calculate the geometric mean of the students' scores.

Solution:

$$\bar{X}_G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

$$\bar{X}_G = \sqrt[8]{10 \times 16 \times 13 \times 15 \times 12 \times 18 \times 11 \times 9}$$

$$\bar{X}_G = 12.67$$

$$\bar{X}_G = 10^{\frac{1}{n} \sum \log x_i}$$

$$\bar{X}_G = 10^{\frac{1}{8} (\log 10 + \log 16 + \log 13 + \log 15 + \log 12 + \log 18 + \log 11 + \log 9)}$$

Using logarithms:

$$\text{Log } G = \frac{1}{8} (\log 10 + \log 16 + \log 13 + \log 15 + \log 12 + \log 18 + \log 11 + \log 9)$$

$$\text{Log } G = \frac{1}{8} (1 + 1.20 + 1.11 + 1.18 + 1.08 + 1.26 + 1.04 + 0.95)$$

$$\text{Log } G = \frac{8.82}{8}$$

$$\text{Log } G = 1.103$$

$$\bar{X}_G = 10^{1.103}$$

$$\bar{X}_G = 12.67$$

Thus, the geometric mean of the students' scores in the statistics exam is 12.67.

2. Grouped Data Using a Frequency Table**1. Case of a Discrete Quantitative Variable**

The geometric mean of values grouped in a frequency table can be calculated using the following formula:

$$\bar{X}_G = \sqrt[n]{x_1^{n_1} \times x_2^{n_2} \dots \dots x_i^{n_i}}$$

To simplify calculations, logarithms are used as follows:

$$\bar{X}_G = 10^{\frac{1}{n} \sum n_i \log x_i}$$

Or equivalently:

$$\text{LogG} = \frac{\sum \log x_i n_i}{\sum n_i}$$

Where:

- x_1, x_2, \dots, x_i are the values of the statistical variable.
- n_1, n_2, \dots, n_i are the corresponding absolute frequencies.

Example

The following frequency table shows the number of study hours for 37 first-year students:

x_i	n_i
1	6
2	10
3	11
4	7
5	3
\sum	37

Required:

Calculate the geometric mean of the distribution.

Solution:

x_i	n_i	$x_i^{n_i}$	$\text{Log } x_i$	$n_i \log x_i$
1	6	1	0	0
2	10	1024	0.30	3
3	11	177147	0.48	5.24
4	7	16384	0.60	4.21
5	3	125	0.69	2.09
\sum	37			14.57

$$\bar{X}_G = \sqrt[n]{x_1^{n_1} \times x_2^{n_2} \dots \dots x_i^{n_i}}$$

$$\bar{X}_G = \sqrt[37]{1 \times 1024 \times 177147 \times 16384 \times 125}$$

$$\bar{X}_G = 2.5$$

or

$$\bar{X}_G = 10^{\frac{1}{n} \sum n_i \log x_i}$$

$$\bar{X}_G = 10^{\frac{1}{37} \sum 14.57}$$

$$\bar{X}_G = 2.5$$

or

$$\text{LogG} = \frac{\sum \log x_i n_i}{\sum n_i} = \frac{14.57}{37} = 0.4$$

$$\bar{X}_G = 10^{0.4} = 2.5$$

2. Case of a Continuous Quantitative Variable

To compute the geometric mean for grouped continuous data, the following steps are followed:

1. Determine the midpoint of the class.
2. Calculate the logarithm of the midpoint of each class.
3. Multiply the logarithm of each midpoint by the appropriate frequency.
4. Sum the products received.
5. To calculate the logarithm of the geometric mean, divide the total by the sum of frequencies.

$$\text{Log G} = \frac{1}{n} (\log c_1 n_1 + \log c_2 n_2 \dots + \log c_i n_i)$$

$$\text{Log G} = \frac{\sum \log c_i n_i}{\sum n_i}$$

$$\bar{X}_G = 10^x$$

Example

The next frequency table shows the daily number of smartphone usage hours for 60 students:

Daily Smartphone Use Classes x_i	Number of Students n_i
[0-2[5
[2-4[12
[4-6[20
[6-8[15
[8-10[8
Total	60

Solution:

Students Number of	n_i	c_i	Log c_i	Log $c_i \times n_i$
[0-2[5	1	0	0
[2-4[12	3	0.48	5.72
[4-6[20	5	0.70	13.99
[6-8[15	7	0.84	12.67
[8-10[8	9	0.95	7.63
Σ .	60			40.02

$$\text{Log } G = \frac{\Sigma \log c_i n_i}{\Sigma n_i} = \frac{40.02}{60} = 0.67$$

$$\bar{X}_G = 10^{0.67}$$

$$\bar{X}_G = 4.67$$

Thus, the geometric mean of daily smartphone usage hours is 4.67 hours.

2. Quadratic Mean ‘Root Mean Square’

The quadratic mean is determined as the square root of the average of the squared values. It is represented by the symbol \bar{X}_Q .

The arithmetic mean is the most prevalent metric in social research; though, the quadratic mean is only employed where there is a clear statistical justification, such as in the analysis of errors and deviations. It constitutes the basis for calculating the standard deviation, since the latter depends on squared differences (Montgomery & Runger, 2014, p. 41).

The quadratic mean is used in a limited scope, particularly in:

- The analysis of response intensity,
- Psychological measurements associated with fluctuation or variability.

Characteristics of the Quadratic Mean

1. It is always greater than or equal to the arithmetic mean, and it is equal to it only when all values are identical.
2. It is highly affected by large values due to the squaring process.
3. It can never be negative.
4. It is appropriate for continuous quantitative variables.
5. It is considered more sensitive to dispersion than the arithmetic mean (Gravetter & Wallnau, 2017, p. 87).

Limitations of the “Quadratic Mean”

The quadratic mean is not applicable in the following situations:

- when extreme values exist without a scientific justification,
- when analyzing qualitative or ordinal variables,
- In traditional educational and psychological studies without a methodological rationale.

Computation of the Quadratic Mean

1. Ungrouped (Raw) Data

If we have n values x_1, x_2, \dots, x_n , the quadratic mean is calculated according to the following formula:

$$\bar{X}_Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_i^2}{n}}$$

$$\bar{X}_Q = \sqrt{\frac{\sum x_i^2}{n}}$$

Example

The following are the scores of 8 students in the Applied Statistics course:

09 11 18 15 12 13 16 10

Required:

Calculate the quadratic mean of the students' exam scores.

Solution:

To compute the quadratic mean of the scores, we apply the formula as follows:

$$\bar{X}_Q = \sqrt{\frac{\sum x_i^2}{n}}$$

$$\bar{X}_Q = \sqrt{\frac{10^2+16^2+13^2+15^2+12^2+18^2+11^2+9^2}{8}}$$

$$\bar{X}_Q = \sqrt{\frac{100+256+169+225+144+324+121+81}{8}}$$

$$\bar{X}_Q = \sqrt{\frac{1420}{8}}$$

$$\bar{X}_Q = \sqrt{177.5}$$

$$\bar{X}_Q = 13.33$$

Thus, the quadratic mean of the students' exam scores is 13.33.

2. Grouped Data (Frequency Distribution)**Case of a Discrete Quantitative Variable**

When data are collected in a frequency table, the quadratic mean is calculated as the “square root” of the ratio of the summation of the products of the squared values x_i and their corresponding frequencies n_i to the total frequency, according to the following formula:

$$\bar{X}_Q = \sqrt{\frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_i^2 n_i}{n}}$$

$$\bar{X}_Q = \sqrt{\frac{\sum x_i^2 n_i}{\sum n_i}}$$

Example

The following frequency table shows the number of daily study hours for 40 first-year students:

Number of Hours x_i	Number of Students n_i
0	3
1	6
2	10
3	11
4	7
5	3
Total	40

Required:

Calculate the quadratic mean of the number of study hours.

Solution:

Compute the quadratic mean of the number of study hours.

x_i	n_i	x_i^2	$x_i^2 n_i$
0	3	0	0
1	6	1	6
2	10	4	40
3	11	9	99
4	7	16	112
5	3	25	75
Σ	40		332

$$\bar{X}_Q = \sqrt{\frac{\Sigma x_i^2 n_i}{\Sigma n_i}}$$

$$\bar{X}_Q = \sqrt{\frac{332}{40}}$$

$$\bar{X}_Q = \sqrt{8.3}$$

$$\bar{X}_Q = 2.89$$

Thus, the quadratic mean of daily study hours is 2.89.

Case of a Continuous Quantitative Variable

When the data are grouped, the quadratic mean is calculated as the square root of the ratio of the sum of the products of the squared class midpoints C_i and their corresponding frequencies n_i , according to the following formula:

$$\bar{X}_Q = \sqrt{\frac{\sum c_i^2 n_i}{\sum n_i}}$$

Example

The following table shows the daily smartphone usage hours for 60 students:

Daily Smartphone Use Classes x_i	Number of Students n_i
[0-2[5
[2-4[12
[4-6[20
[6-8[15
[8-10[8
Total	60

Required:

Calculate the quadratic mean of the daily smartphone usage hours of students.

Solution:

Compute the quadratic mean of the daily smartphone usage hours of students.

x_i	n_i	c_i	c_i^2	$n_i c_i^2$
[0-2[5	1	1	5
[2-4[12	3	9	108
[4-6[20	5	25	500
[6-8[15	7	49	735
[8-10[8	9	81	648
\sum	60			1996

$$\bar{X}_Q = \sqrt{\frac{\sum c_i^2 n_i}{\sum n_i}}$$

$$\bar{X}_Q = \sqrt{\frac{1996}{60}}$$

$$\bar{X}_Q = 5.77$$

Thus, the quadratic mean of daily smartphone utilization hours is: (5.77).

3. Harmonic Mean

The “harmonic mean” of a set of values is known as the reciprocal of the arithmetic mean of the reciprocals of the values (Salem & Ababneh, 2007, p. 98). It is signified by the symbol \bar{X}_H .

The use of the “harmonic mean” is limited and it requires the presence of a clear statistical justification related to the nature of the variable. The harmonic mean is used only when an inverse relationship exists between two phenomena. It is also applied in the analysis of:

- absence rates,
- frequency rates,
- performance rates.

Characteristics of the Harmonic Mean

- The harmonic mean can not be computed if a zero (0) value exists in the statistical series, since the reciprocal of zero (1/0) is indeterminate.
- It is always less significant than both the “arithmetic mean” and the “geometric mean”.
- It is highly sensitive to lesser values.
- It is not computable when a zero value is present.
- It is suitable for data that represent rates or ratios.
- It is not appropriate for qualitative or ordinal variables.
- It is used more frequently in applied analysis than in descriptive analysis (Gravetter & Wallnau, 2017, p. 89).

Limitations of the Harmonic Mean

The harmonic mean cannot be used in the following cases:

- When zero values are present.
- When analyzing achievement scores or direct psychological measures.
- In distributions that do not represent rates.
- When the values show very high variability without a clear justification.

Computation of the Harmonic Mean

1. Harmonic Mean for Raw (Ungrouped) Data

If we have values x_1, x_2, \dots, x_i , and their number is n , the harmonic mean is calculated according to the following formula:

$$\bar{X}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_i}}$$

$$\bar{X}_H = \frac{n}{\sum \frac{1}{x_i}}$$

Example

The following are the scores of 8 students in the Applied Statistics course:

09 11 18 15 12 13 16 10

Required:

Calculate the harmonic mean of the students' exam scores.

Solution

To compute the harmonic mean of the scores, we apply the formula as follows:

$$\bar{X}_H = \frac{n}{\sum \frac{1}{x_i}}$$

$$\bar{X}_H = \frac{8}{\frac{1}{10} + \frac{1}{16} + \frac{1}{13} + \frac{1}{12} + \frac{1}{18} + \frac{1}{11} + \frac{1}{9}}$$

The harmonic mean of the students' scores in the statistics exam is 13 points.

Case of Grouped Data Using a Frequency Distribution

Discrete Quantitative Variable

If the data are grouped in a frequency table, or if the values are associated with weights, and we have values x_1, x_2, \dots, x_i with a total number n , and corresponding frequencies or weights n_1, n_2, \dots, n_i , then the harmonic mean is calculated in this case according to the following formula:

$$\bar{X}_H = \frac{n_1 + n_2 + \dots + n_i}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_i}{x_i}}$$

$$\bar{X}_H = \frac{\sum n_i}{\sum \frac{n_i}{x_i}}$$

Example

The following frequency table shows the number of daily study hours for 37 first-year students:

Number of Hours x_i	Number of Students n_i
1	6
2	10
3	11
4	7
5	3
Total	37

Required:

Calculate the harmonic mean of the daily study hours.

Solution**Calculation of the Harmonic Mean**

x_i	n_i	n_i/x_i
1	6	6
2	10	5
3	11	3.66
4	7	1.75
5	3	0.6
\sum	37	17.01

$$\bar{X}_H = \frac{\sum n_i}{\sum \frac{n_i}{x_i}}$$

$$\bar{X}_H = \frac{37}{6+5+3.66+1.75+0.6}$$

$$\bar{X}_H = \frac{37}{17.10}$$

$$\bar{X}_H = 2.17$$

Thus, the harmonic mean of the daily study hours is 2.17.

Continuous Quantitative Variable

If the data related to a given phenomenon are grouped into classes within a frequency table, and we wish to calculate the harmonic mean, we may rely on the class midpoints c_1, c_2, \dots, c_n , along with their corresponding frequencies n_1, n_2, \dots, n_n . In this case, the harmonic mean is computed using the sum of the ratios of each frequency n_i to its corresponding class midpoint c_i , according to the following formula:

$$\bar{X}_H = \frac{n_1 + n_2 + \dots + n_i}{\frac{n_1}{c_1} + \frac{n_2}{c_2} + \dots + \frac{n_i}{c_i}}$$

$$\bar{X}_H = \frac{\sum n_i}{\sum \frac{n_i}{c_i}}$$

Example

The following frequency table shows the daily number of smartphone usage hours for 60 students:

Daily Smartphone Use Classes	Number of Students
x_i	n_i
[0-2[5
[2-4[12
[4-6[20
[6-8[15
[8-10[8
Total	60

Required:

Calculate the harmonic mean of the daily smartphone usage hours of students.

Solution

x_i	n_i	c_i	$\frac{n_i}{c_i}$
[0-2[5	1	5

[2-4[12	3	4
[4-6[20	5	4
[6-8[15	7	2.14
[8-10[8	9	0.88
Total	60		52.02

$$\bar{X}_H = \frac{5+12+20+15+8}{5+4+4+2.14+0.88}$$

$$\bar{X}_H = \frac{60}{52.02}$$

$$\bar{X}_H = 1.15$$

Thus, the harmonic mean of the daily smartphone usage hours of students is: 1.15.

3. Median

The median is known as one of the measures of central tendency which take into account the “ranks of values” and divides the phenomenon, or the set of values, into two equal parts. It is considered the second measure of central tendency and is also referred to as the 50th percentile (Percentile 50th), since it is the value below which half of the observations ($n/2$) fall and above which the other half ($n/2$) lie. In other words, percentage of 50% of the values is fewer than the median and 50% are superior to it. The median is denoted by the symbol Me.

3.1. Uses of the Mode in the Social Sciences

In the social sciences, the mode is mainly used to identify the most common social characteristics or patterns within a population or a sample, especially when dealing with qualitative or nominal variables.

Main uses include:

1. Identifying dominant social characteristics

The mode helps determine the most prevalent social attributes, such as:

- The most common level of education.
- The most frequent occupation in a given society.
- The dominant family structure.

2. Analyzing qualitative social variables

The mode is particularly suitable for variables such as:

- Gender.
- Marital status.
- Occupational or social affiliation.

3. Supporting social planning and policy decisions

By identifying the most widespread categories, the mode assists policymakers in:

- Recognizing the most representative social groups.
- Directing social programs toward the most common needs.

4. Interpreting grouped data in social surveys

In field studies, the modal class indicates where the social phenomenon under study is most concentrated.

3.2.Characteristics of the Median

- ✓ “The median” is not affected by extreme or outlier values.
- ✓ It does not allow for direct algebraic operations.
- ✓ It is easy and quick to calculate.
- ✓ It can be derived from open-ended frequency distribution tables.
- ✓ It can be determined graphically.
- ✓ It depends on the ordering of values rather than their numerical magnitude.
- ✓ It may be determined by one value or two values, depending on the number of observations.
- ✓ It is used with ordinal, interval, and ratio scales, but it is not appropriate for the nominal scale.
- ✓ It represents the average of ranks, not the arithmetic average of values.
- ✓ It lies between the arithmetic mean and the mode (Gravetter & Wallnau, 2017, pp. 83–88).

3.3.Computation of the Median

There are numerous methods for computing the median, including the following:

1.Median for Ungrouped Data

If the “data” are presented as an ungrouped statistical series, the first step in determining

the median is to arrange the values in ascending or descending order. Then, two cases are distinguished according to the number of observations.

Case 1: Odd Number of Observations (n is odd)

- The rank “position” of the median is computed using the following formula:

$$\frac{n + 1}{2}$$

- The median is determined by identify the value which rank corresponds to this position.
- The median is the value whose rank is $\frac{n+1}{2}$, and it is expressed mathematically by the formula:

$$M_e = x \left(\frac{n+1}{2} \right)$$

Example

Given the following values:

4, 7, 9, 6, 12, 10, 5, 8, 11

Solution:

Finding the median

1. Arrange the data in ascending order:

4 5 6 7 **8** 9 10 11 12

2. The number of values is n=9 (an odd number). Consequently, the rank of the median is:

$$\frac{9+1}{2} = 5 = \frac{n+1}{2}$$

3. The median is the value corresponding to the Fifth position:

$$.M_e = x (5)$$

$$M_e = 8$$

Case 2: Even Number of Observations (n is even)

If the number of values is even, the median is any value lying between the two central

values, which occupy positions $\frac{n}{2}$ and $\frac{n+1}{2}$. Therefore, the median is equal to the arithmetic mean of the two values corresponding to these positions, as follows:

$$M_e = \frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right)}{2}$$

Example

Given the following values:

Find the median of the following data:

6 8 10 3 1 13 7

Required:

Find the median of these values.

Solution

Arrange the data (values) in ascending order:

1 3 6 7 8 10 13 16

The number of values is: $n=8$ (an even number).

Accordingly, the rank of the median is $\frac{8}{2}$ and $\frac{8}{2} + 1$,

As a result, the median value is equal to the arithmetic mean of the two values that correspond to the median ranks, as follows:

$$M_e = \frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right)}{2} = \frac{x(04) + x(05)}{2} = \frac{7 + 8}{2} = 7.5$$

2. Median for Grouped Data

To calculate the median from grouped data displayed in a frequency distribution table, two cases are distinguished:

1. Case of a Discrete Quantitative Variable

To calculate the median for grouped data in a frequency distribution table (discrete

quantitative variable), the following steps are followed:

1. Compute the ascending cumulative frequency.
2. Determine the median rank $(\frac{N}{2})$
3. We look in the column (row) of the ascending cumulative frequency for the value whose cumulative frequency is equal to the median rank $(\frac{N}{2})$ or immediately greater than it

$$.N_i \uparrow \geq \frac{n}{2}$$

4. The value x_i corresponding to this cumulative frequency is the median value.

Example

The following table represent the distribution of the number of study hours spent by 37 students:

Number of Hours x_i	Number of Students n_i
1	6
2	10
3	11
4	7
5	3
Total	37

Required:

Find the median of these data.

Solution:

In order to calculate the median, we follow these steps:

1. Compute the ascending cumulative frequency:

x_i	n_i	$N_i \uparrow$
1	6	6
2	10	16
3	11	27
4	7	34

5	3	37
Σ .	37	/

2. Determine the median rank:

$$\frac{N}{2} = \frac{37}{2} = 18.5$$

3. Search for the cumulative frequency that is equivalent to or immediately superior to the median rank. Since 27 is the first cumulative frequency greater than 18.5

$$([27 > \frac{N}{2} = 18.5 > 16])$$

4. The value of the median is the value (x_i) corresponding to 27.

Thus, the median value is: **Me=3**

2. Case of a Continuous Quantitative Variable

To determine the median for grouped data presented in a frequency distribution table with class intervals (case of a continuous quantitative variable), the following steps are followed:

- Calculate the ascending cumulative frequency.
- Determine the median rank ($\frac{N}{2}$).

Determine the median class (the class to which the median belongs), which is the class that corresponds to the ascending cumulative frequency that is either immediately above or equal to the median rank;

$$N_i \uparrow \geq \frac{n}{2}$$

- Determine and compute the median by applying the statistical formula of the median (Azzouz, 2010, p. 119):

$$M_e = L_{Me} + \frac{\frac{N}{2} - N_{\uparrow Me-1}}{n_{Me}} \times K_{Me}$$

Where:

- L_{Me} : The lower boundary of the median class.
- N : The total frequency, where $N = \sum_{i=1}^k n_i$

- $N \uparrow_{Me-1}$: The ascending cumulative frequency of the class preceding the median class.
- n_{Me} : The absolute frequency of the median class.
- K_{Me} : The class width (length) of the median class.

Example

Exercise

The table below shows the distribution of a group of students according to their daily study hours:

Daily Study Hours x_i	Number of Students n_i
[0 -2[10
[2 -4[18
[4 -6[22
[6 -8[15
[8 -10[9
[10 -12[6
Total	80

Required:

Calculate the median of the distribution.

Solution

To determine the "median value", we follow these steps:

1. Compute the ascending cumulative frequency

x_i	n_i	$N_i \uparrow$
[0 -2[10	10
[2 -4[18	28
[4 -6[22	50
[6 -8[15	65
[8 -10[9	74
[10 -12[6	80
Σ	80	

2. Determine the median rank

$$\frac{N}{2} = \frac{80}{2} = 40$$

Where

$$N = \sum_{i=1}^k n_i = 80$$

3. Identify the median class

The class that corresponds to the ascending cumulative frequency that is either immediately higher than or equal to the median rank is known as the median class. Since 65 is the first cumulative frequency immediately greater than 40: ($[28 > 40 - \frac{N}{2} > 65]$)

The median class is [4–6[.

4. Calculate the median using the statistical formula

$$M_e = L_{Me} + \frac{\frac{N}{2} - N_{Me-1}}{n_{Me}} \times K_{Me}$$

$$M_e = 4 + \frac{\frac{80}{2} - 28}{22} \times 2 = 5.09$$

The median daily study time is approximately 5.09 hours.

3. Graphical Method for Determining the Median

The median can be easily determined graphically plotting the ascending and descending cumulative curves on the same graph. The vertical coordinate of their intersection point is half of the entire cumulative frequency, while the horizontal coordinate is the median value.

The median can also be determined graphically from either the ascending or descending cumulative curve by following these steps:

- Construct an ascending or descending cumulative frequency table.
- Draw the ascending or descending cumulative curve.
- Place the median position on the vertical axis.

To determine the value of the median, a horizontal line is drawn from the median rank point, parallel to the horizontal axis, intersecting the cumulative curve at point (A). A perpendicular is then dropped from this point onto the horizontal axis at point (B), which represents the median value.

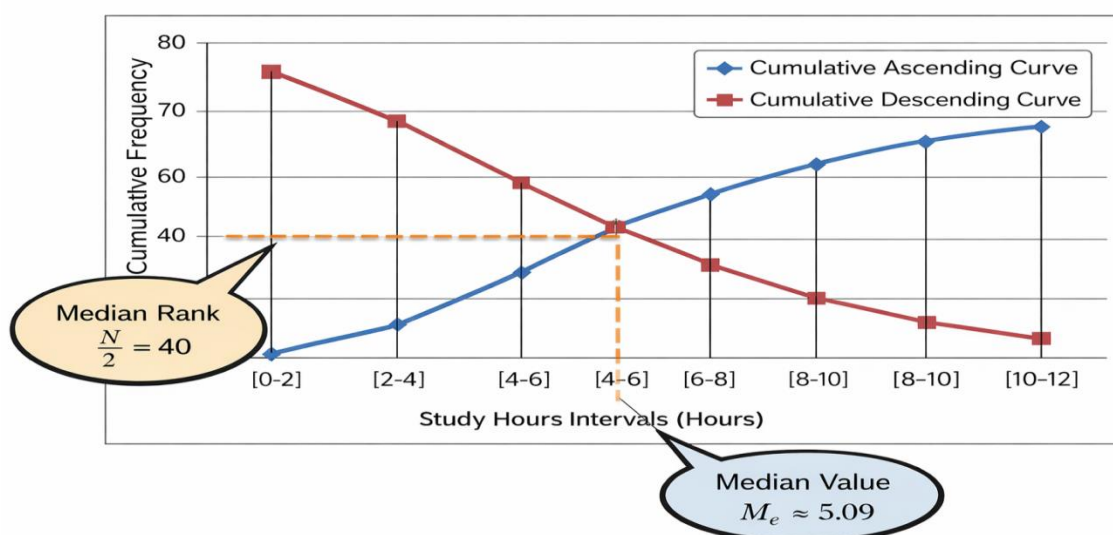
Exercise:

Use the graphical way to calculate the median based on the data in the preceding example.

Solution :

x_i	n_i	$N_i \uparrow$	$N_i \downarrow$
[0 -2[10	10	80
[2 -4[18	28	70
[4 -6[22	50	52
[6 -8[15	65	30
[8 -10[9	74	15
[10 -12[6	80	6
Σ	50		

Figure (17): Illustrates the graphical determination of the median using cumulative ascending and descending curves.

**4.Mode:**

The mode is the simplest measure of central tendency. It is determined as the value or attribute that occurs most frequently compared to the other values or attributes in a given distribution. Based on this definition, a dataset may have one mode, more than one mode, or no mode at all. The mode is considered the most appropriate measure for describing qualitative (categorical) data, and it is denoted by the symbol M_0 .

(Al-Saifo et al., 2010, p. 122)

4.1. Uses of the Median in the Social Sciences

The median is one of the most widely used measures in the social sciences because it provides a realistic representation of the center of social data, particularly when distributions are not symmetrical.

Main uses include:

- **Analyzing uneven economic and social variables**

The median is especially appropriate for variables such as:

- Income.
- Wages.
- Wealth distribution.

- **Reducing the influence of extreme values**

In social data sets that include very high or very low values, the median offers a more stable and representative measure than the mean.

- **Assessing living standards and social well-being**

The median income or median expenditure often reflects the actual conditions of the majority of the population more accurately.

- **Working with ordinal social variables**

The median is suitable for variables that can be ranked but do not have equal numerical intervals, such as:

- Levels of satisfaction.
- Social attitudes and opinions.

4.2.Characteristics of the Mode:

- The mode is more reliable and consistent than the arithmetic mean or median since it reflects the most commonly occurring value.
- It is not affected by extreme values or by intermediate values in the frequency distribution.
- Its value is influenced by the number of classes and their widths in grouped frequency tables.
- A dataset may contain more than one mode within the same frequency table.
- The mode is not suitable for algebraic operations such as addition or division.

- It can be determined graphically using frequency curves.
- It can also be identified from open-ended frequency distribution tables.
- Its value depends on the importance of the observation in terms of its frequency or magnitude.
- The mode is considered the most appropriate average for describing qualitative (categorical) phenomena, and it is regarded as a good measure of central tendency, especially when the distribution is symmetrical. (Al-Omari & Al-Ajeeli, 2000, p. 67)
- It is easy to determine, whether numerically or graphically.

Limitations of the Mode:

- The mode is often determined approximately, particularly in grouped frequency distributions. If the distribution is skewed, the mode may appear far from the center of the distribution, which reduces its effectiveness as a measure of central tendency.
- It is difficult to determine in frequency distributions with unimodal curves.
- It is based on a simple concept that does not necessarily represent all observations in the dataset.
- Its estimation becomes difficult when the number of observations increases significantly and when high frequencies is equal across consecutive classes.
- It is not appropriate for use with open-ended frequency distributions, as determining whether a frequency is large or small requires knowledge of the class width.
- Its value is affected by the method used to define the class intervals; therefore, it may vary from one researcher to another depending on the chosen class width.
- It does not provide a reliable representation of values in highly skewed distributions.
- It cannot be relied upon in problems that require algebraic manipulation.

4.3. Calculation of the Mode:

The mode can be calculated for both quantitative variables and qualitative variables.

1. In the case of ungrouped data:

The mode is defined as the value that occurs most frequently. It is widely used in descriptive data to identify the most common pattern or level.

To present this concept, several practical examples are provided to facilitate understanding

and clarify the idea.

Example 1:

The following data represent the grades achieved by 10 students:

Excellent- Good- Very Good- Good- Average- Above Average-Good-Weak- Very Good- and Good.

Required: Determine the mode of the data.

Solution:

We observe that the category *Good* occurs most frequently among the given categories. Therefore, the mode is “*Good*”.

Thus, the mode is: (**Good = Mo**).

Example 2:

Let the following values be given:

2, 4, 8, 6, 12, 10.

Required: Is there a mode for these values?

Solution:

There is no mode for these values because none of them is repeated.

Example 3:

Let the following values be given:

8 - 4 - 8 - 5 - 13 - 10 .

Required: Determine the mode of the data.

Solution:

The mode of this dataset is 8 ($Mo = 8$), as it is the most frequently occurring value. This dataset is unimodal.

Example 4:

Let the following values be given:

2, 4, 2, 10, 5, 13, 10.

Required: Determine the mode of the data.

Solution:

The modes of this dataset are 2 and 10 ($M_o = 2, 10$).

This distribution is referred to as a bimodal distribution.

Conclusion:

Statistical data may have no mode when all values occur with the same frequency. In other cases, distributions may be classified as follows:

- **Unimodal distribution:** when the dataset has one mode.
- **Bimodal distribution:** when the dataset has two modes.
- **Multimodal distribution:** when the dataset has more than two modes.

2. In the Case of Grouped Data (Frequency Distribution)

Discrete Quantitative Variable

When the statistical variable under study is a discrete quantitative variable, the mode is determined directly from the frequency distribution table. It is the value x_i that corresponds to the highest frequency. A dataset may have more than one mode. (Benya, 2018, p. 78)

Example:

The following frequency table shows the number of hours spent studying by 37 first-year students:

Number of hours x_i	Number of students n_i
1	6
2	10
3	11
4	7
5	3
Σ	37

Solution:

From the table above, we observe that “the highest frequency” is 11, which corresponds to the value 3. Therefore, the mode of the distribution is:

$$M_0 = 3$$

This means most of the students spend 3 hours studying, making it the most common number of study hours among the students.

Discrete Quantitative Variable (Grouped Data)

The mode for grouped data is derived after determining the modal class, which is the class with the highest frequency relative to the other classes. In this situation, the mode is computed with the following formula:

$$\text{Mod} = L_1 + \left(\frac{d_1}{d_1 + d_2} \right) \times k$$

Where:

- L_1 : lower limit of the modal class
- d_1 : difference between the frequency of the modal class and the frequency of the previous class
- d_2 : difference between the frequency of the modal class and the frequency of the following class
- K : class width

Example:

Based on the following data about a group of student 's daily study hours:

Daily Study Hours x_i	Number of Students n_i
[0 -2[10
[2 -4[18
[4 -6[22
[6 -8[15
[8 -10[9
[10 -12[6
Total	80

The required :

Calculate the mode of the previous table.

Solution:

From the table above, we observe that the class intervals are of equal width. Therefore,

the following steps are useful to calculate the mode:

1. Identification of the modal class:

The modal class is the one with the highest frequency.

Since the highest frequency is 22, the modal class is: [4–6[

2. Calculation of the mode:

The mode is calculated by using this formula:

$$M_o = L_1 + \left(\frac{d_1}{d_1 + d_2} \right) \times k$$

$$M_o = 4 + \frac{(22-18)}{(22-18)+(22-15)} \times 2$$

$$M_o = 4 + \frac{4}{11} \times 2$$

$$M_o = 4.73$$

Therefore, the mode value is:

$$M_o = 4.73$$

3. Graphical Determination of the Mode:

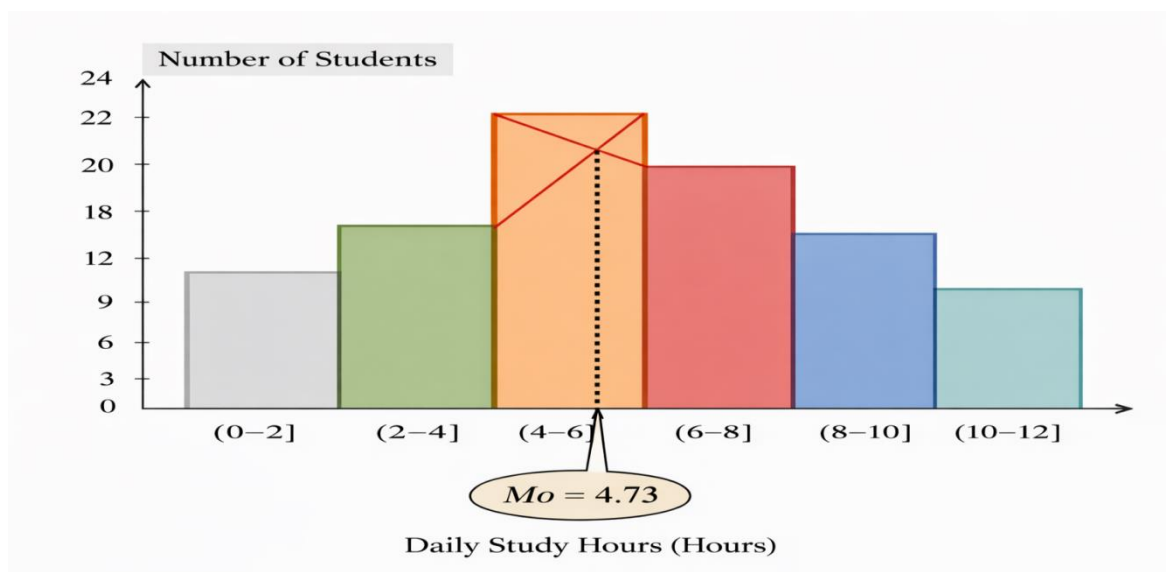
The mode can also be identified graphically by plotting the histogram of the modal class with the preceding and following classes. A line is drawn from the top left corner of the modal class rectangle to the top right corner of the preceding class rectangle, followed by another line from the top right corner of the modal class rectangle to the top left corner of the succeeding class rectangle. A vertical line is drawn from the intersection to the horizontal axis (x-axis). The point at which it intersects the axis represents the mode value.

Example:

Based on the previous table which shows the distribution of a group of students according to their daily study hours, determine the mode graphically.

Solution:

Figure (18): Illustrates the graphical representation of the mode in grouped data.



5. Relationship between Measures of Central Tendency (Arithmetic Mean, Median, and Mode)

If a dataset has a single mode, the arithmetic mean, median, and mode are related as follows:

- The curve is symmetric with a single peak and has a bell-shaped form if the frequency distribution is symmetrical. In this case: $\bar{X} = M_e = M_o$

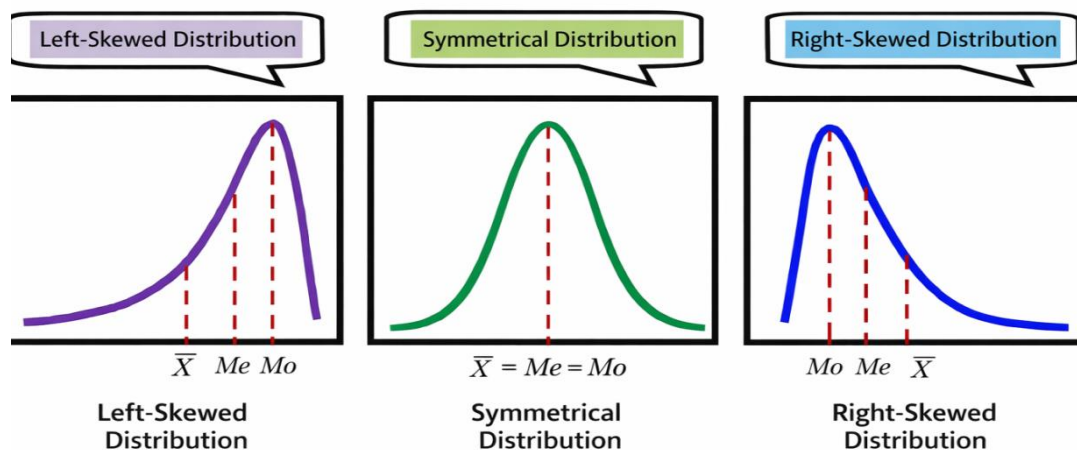
- If the frequency distribution is slightly skewed, the following relationship holds:

$(\bar{X} - M_e) 3 = M_o - \bar{X}$, This relationship is useful for estimating the arithmetic mean when the frequency distribution table is opened at one end.

- If the “arithmetic mean” is greater than the” median” and the” median” is greater than the “mode”, the distribution is positively skewed (right-skewed): $M_o < M_e < \bar{X}$.

- If the arithmetic mean is less than the median and the median is less than the mode, the distribution is negatively skewed (left-skewed): $M_o > M_e > \bar{X}$

Figure (19): Illustrates the relationship between the mean, median, and mode according to the type of distribution.



Suggested exercises :

Exercise 1

Answer True or False, and justify your answer:

1. The arithmetic mean is the most affected measure of central tendency by extreme values.
2. The median depends on the order of the values rather than their numerical magnitude.
3. The harmonic mean is used when the relationship between variables is direct.
4. The mode is the most appropriate measure for describing qualitative data.

Exercise 2

Indicate the most appropriate measure of central tendency in each of the following cases, with justification:

1. Studying the marital status of individuals.
2. Analyzing monthly salaries that include very high extreme values.
3. Calculating the average economic growth rate.
4. Analyzing the intensity of errors in physical measurements.

Exercise 3

The following values represent the daily study hours of 10 students:

4 – 6 – 5 – 8 – 7 – 6 – 5 – 9 – 4 – 6

Required:

1. Calculate the arithmetic mean.
2. Interpret the result statistically.

Exercise 4

Number of hours x_i	Number of students n_i
1	5
2	9
3	14
4	12
5	10
Total	50

Required:

1. Calculate the arithmetic mean using absolute frequencies.
2. Calculate it using relative frequencies.

Exercise 5

The following values represent the growth rates of an economic institution over four years:

1.05 – 1.10 – 1.08 – 1.12

Required:

1. Calculate the geometric mean.
2. Explain why the geometric mean is preferred in this case.

Exercise 6

Values x_i	Frequency n_i
2	6

4	10
6	14
8	10

Required:

1. Calculate the geometric mean using logarithms.
2. Interpret the result.

Exercise 7

The following values represent the intensity of psychological responses:

6 – 8 – 10 – 7 – 9

Required:

1. Calculate the quadratic mean.
2. Compare it with the arithmetic mean.

Exercise 8

Class interval	Number of students
[0–2[8
[2–4[12
[4–6[20
[6–8[10

Required:

1. Calculate the quadratic mean using class midpoints.
2. Explain its sensitivity to large values.

Exercise 9

The following values represent task completion rates (per hour):

2 – 4 – 8 – 16

Required:

1. Calculate the harmonic mean.
2. Explain why it is appropriate for this type of data.

Exercise 10

Speed	Number of cases
40	10
60	20
80	10

Required:

1. Calculate the harmonic mean.
2. Compare it with the arithmetic mean.

Exercise 11

The following values represent test scores:

12 – 15 – 9 – 18 – 14 – 11 – 16

Required:

1. Arrange the values in ascending order.
2. Calculate the median.

Exercise 12

Number of hours	Number of students
1	6
2	10
3	11
4	7
5	3
Total	37

Required:

1. Calculate the: ascending cumulative frequency.
2. Determine the « median rank ».
3. Extract the « median value ».

Exercise 13

The next values represent the number of absences:

2 – 3 – 1 – 2 – 4 – 2 – 3 – 5

Required:

1. Determine the mode.
2. Identify the type of distribution.

Exercise 14

The following values represent the number of books read by a group of students during one month:

3, 5, 2, 4, 5, 6, 5, 3, 4, 5

Required:

1. Identify the type of the statistical variable.
2. Calculate the mode.
3. Determine whether the distribution is unimodal, bimodal, or has no mode.

Exercise 15

The next data represent the grades obtained by a group of students in Statistics:

Good, Average, Very Good, Good, Pass, Good, Average, Very Good, Good

Required:

1. Determine the mode.
2. Interpret the result statistically.
3. Explain why the mode is the most appropriate measure in this case.

Exercise 16

Number of absences x_i	Number of students n_i
0	5
1	12
2	18
3	10
4	5

Required:

1. Identify the modal value.
2. Calculate the mode.
3. Interpret the result.

Exercise 17

The following table represents the monthly wages (in thousand dinars) of a group of workers:

Wage classes	Frequency
[20 – 30[14
[30 – 40[26
[40 – 50[32
[50 – 60[18
[60 – 70[10

Required:

1. Identify the modal class.
2. Calculate the
3. Interpret the result economically.

Exercise 18

The following are three statistical distributions with different relationships between the arithmetic mean (\bar{X}), the median (Me), and the mode (Mo):

a. $\bar{X} = Me = Mo$

b. $Mo < Me < \bar{X}$

c. $Mo > Me > \bar{X}$

Required:

1. Identify the type of each distribution (symmetrical, positively skewed, negatively skewed).
2. Draw an approximate curve for each distribution.
3. Explain the position of each measure of central tendency.

Exercise 19

The following data represent the daily study hours of a sample of students:

Hours	Number of students
[0–2[8
[2–4[15
[4–6[20
[6–8[12
[8–10[5

Required:

1. Calculate the mode numerically.
2. Identify the modal class.
3. Determine approximately the type of distribution.
4. Relate the result to the shape of the frequency curve.

Measures of positional tendency are considered fundamental concepts in descriptive statistics. Their purpose is to determine the relative position of values within a statistical distribution after the data have been arranged in ascending or descending order. Among the most important of these measures are quartiles, deciles, and percentiles, which are used to divide the dataset into equal parts. These measures enable the researcher to better understand the distribution and concentration of values, rather than depending mainly on measurements of primary pattern for instance the arithmetic mean or the median.

Such measurements are particularly relevant in the social, psychological, and educational sciences, where data are often non-symmetrical or influenced by extreme values, making reliance on the arithmetic mean alone insufficient for accurately describing the phenomenon under study. Quartiles, deciles, and percentiles contribute to a more in-depth description of the statistical distribution, whether in terms of analyzing dispersion, comparing individuals or groups, or determining the relative position of values within the statistical population. This makes them indispensable tools in statistical analysis.

1. Quartiles

1. Definition of Quartiles

Quartiles are descriptive statistical measures concerned with determining the position of values within ordered data. They aim to divide the dataset into four equal parts in order to identify the concentration of values in terms of the number of observations. These measures are represented by the symbols $Q1$ (first quartile), $Q2$ (second quartile), and $Q3$ (third quartile). The range of values between each quartile represents 25% (one quarter) of the total number of observations in the statistical series.

Quartiles are widely used in the social sciences to summarize and describe data distributions in psychological, social, and educational research.

According to the literature, quartiles can be defined as follows:

- **First Quartile ($Q1$):** Commonly referred to as the lowest quartile, it is the range of values beneath which 25% of the ordered data set resides, while 75% of the data exceeds it.

- **Second Quartile (Q_2):** This refers to the value at the middle of the statistical pattern and divides it into two equal parts; therefore, it is exactly equal to the median (Me).
- **Third Quartile (Q_3):** Also known as the higher quartile, it is the value follows which 75% of the outcomes correspond, while only 25% of the data lie above it.

1.2. Uses of Quartiles in the Social Sciences

Quartiles perform an essential part in statistical analysis since they are used to calculate measures of dispersion. The interquartile range ($Q_3 - Q_1$), which is particularly preferred when extreme values—either very small or very large—are present. This is due to its ability to reduce the influence of outliers on the description of data variability. In addition, the semi-interquartile range (quartile deviation) can be calculated; it corresponds to fifty percent of the gap between the first and third quartiles, as well as the dispersion of the middle 50% of the values in the statistical variation.

Quartiles are widely used in the social sciences (Gravetter & Wallnau, 2017, p. 76; Healey, 2015, p. 113), particularly in the following areas:

1. Analysis of the Distribution of Social and Psychological Variables

Quartiles help researchers understand how values are concentrated within a distribution, such as:

- distributions of academic achievement scores,
- levels of anxiety or job satisfaction,
- scores related to attitudes or social orientations.

They enable the identification of the proportion of individuals whose values fall within the lower, middle, or upper levels of the studied variable (Gravetter & Wallnau, 2017, p. 76).

2. Studying dispersion using the Interquartile Range

Quartiles tend to be the basis for calculating the interquartile range (IQR), an important measure of dispersion that is widely used in the social sciences because it:

- focuses on the central 50% of the data,
 - is less affected by extreme values compared to the range or the standard deviation.
- (Healey, 2015, p. 113)

3. Comparison Between Social Groups

Quartiles are used to compare distributions across different groups, such as:

- comparing anxiety levels between males and females,
- comparing job performance across occupational categories,
- comparing academic achievement between schools or departments.

These measures allow for an understanding of differences in distributional patterns, not merely differences in the mean (Moore et al., 2017, p. 99).

4. Detection of Outliers

Quartiles play a fundamental role in identifying outliers, especially when using the boxplot, which is particularly important in:

- sensitive psychological studies,
- field-based social research,
- large-scale survey analysis. (Gravetter & Wallnau, 2017, p. 78)

5. Graphical Representation of Social Data

Quartiles are used in constructing the boxplot, which is considered an effective visual tool for:

- summarizing the statistical distribution,
- displaying the median, dispersion, and outliers,
- facilitating the interpretation of results for researchers and decision-makers. (Moore et al., 2017, p. 101)

1.3. Calculation of Quartiles

1. In the Case of Ungrouped Data

For ungrouped data, quartiles are calculated by following specific steps that mainly depend on arranging the data and determining the number of observations (n), as outlined below:

- **Ordering the data:**

The first initial step is to arrange the data in the reverse order (from the smallest to the largest value).

- **Determining the position (rank) of the quartile:** The rank of the quartile (i) is calculated using the following formula:

$$\frac{i(n + 1)}{4}$$

- **Determining the quartile value:** The quartile value is the observation that occupies the remainder rank directly:

$$Q_k = x_{\frac{i(n+1)}{4}}$$

In the case where the quartile has two ranks, the arithmetic mean of their corresponding values is calculated in order to obtain the value of the i -th quartile.

Example

Suppose we have the following values representing the scores obtained by a group of students in a psychological test:

6 - 9 - 4 - 12 - 7 - 10 - 5

Required: Calculate Q_1 , Q_2 , and Q_3 . What can you conclude?

Solution

- Arrange the data in ascending order

4 - 5 - 6 - 7 - 9 - 10 - 12

$n=7$

Calculation of the First Quartile Q_1

$$Q_k = \frac{i(n+1)}{4}$$

$$Q_1 = \frac{1(n+1)}{4}$$

$$Q_1 = \frac{1(7+1)}{4}$$

$$Q_1 = x_{(2)}$$

The rank of the first quartile Q_1 is 2, and its value is the number situated at the second position in the systematic series, which is 5.

Calculation of the Second Quartile Q_2

$$Q_2 = \frac{2(n+1)}{4}$$

$$Q_2 = \frac{2(7+1)}{4} = \frac{16}{4} = 4$$

$$Q_2 = x_{(4)}$$

$$Q_2 = 7$$

The rank of the second quartile Q_2 is 4, and its value is the number sited at the fourth position in the ordered series, which is 7.

(Note that Q_2 is equal to the median.)

Calculation of the Third Quartile Q_3

$$Q_3 = \frac{3(n+1)}{4}$$

$$Q_3 = \frac{3(7+1)}{4} = \frac{24}{4} = 6$$

$$Q_3 = x_{(6)}$$

$$Q_3 = 10$$

The rank of the third quartile Q_3 is 6, and its value is the number situated at the sixth position in the ordered series, which is 10.

Interpretation of the Results

The results indicate that:

- 25% of the students gained a score less than or equal to 5,
- 50% of the students obtained a score less than or equal to 7,
- 75% of the students gained a score less than or equal to 10.

2. Quartiles in the Case of Grouped Data Using a Frequency Distribution

1. Discrete Quantitative Variable

In the case of a discrete quantitative variable, quartiles are calculated by following the standard steps and procedures described in statistical references, as follows:

- Compute the ascending cumulative frequency in order to determine the position of the rank among the observed values.
- Determine the quartile rank, which is calculated using the formula: $\frac{in}{4}$.
- Search in the column of the ascending cumulative frequency for the value equal to $\frac{in}{4}$ or the first value greater than it, such that:

$$N_{i\uparrow} \geq \frac{in}{4}$$

- The value n_i corresponding to the previously identified ascending cumulative frequency represents the quartile value.

Example

Suppose we have the following values, representing the number of books read in one year by a sample of students:

Number of books read x_i	Frequency n_i
1	3
2	7
4	12
6	15
8	9
10	4
12	2
Σ	52

Required: Calculate Q_3, Q_2, Q_1 . What can you conclude?

Solution

First, we compute the ascending cumulative frequencies:

Number of books read x_i	Frequency n_i	$N_1 \uparrow$
1	3	3
2	7	10
4	12	22
6	15	37
8	9	46
10	4	50
12	2	52
Σ	52	

Computing Q_1

- Determine the rank (position) of the first quartile:

$$\frac{in}{4} = \frac{52}{4} = 13$$

- In the upward average frequency column, find the value equal to 13 or the first value larger than 13.
- The value n_i corresponding to the ascending cumulative frequency 22 is 4.

$$Q_1 = 4$$

Computing Q_2 (the median)

- Determine the rank (position) of the second quartile:

$$\frac{in}{4} = \frac{2(52)}{4} = 26$$

- In upward average frequency column, look for the value that is equal to 26 or the first value bigger than 26.
- The value n_i corresponding to the ascending cumulative frequency 37 is 6.

$$Q_2 = 6$$

Computing Q_3

- Determine the rank (position) of the third quartile:

$$\frac{in}{4} = \frac{3(52)}{4} = 39$$

- In the ascending increased frequency column, look for the value that is equal to 39 or the first value greater than 39.
- The value n_i corresponding to the ascending cumulative frequency 46 is 8.

$$Q_3 = 8$$

The previous results indicate that:

- 25% of the students read four books or fewer during the year.
- 50% of the students read six books or less during the year.
- 75% of the students read eight books or fewer during the year.

Discrete Quantitative Variable

In the case of a discrete quantitative variable, quartiles are computed by following the standard steps and procedures presented in statistical references, as outlined below:

- Calculate the increasing average frequency in order to establish the position of the rank among the observed values.
- Determine the quartile rank, which is calculated using the formula: $\frac{in}{4}$
- In the column of the ascending cumulative frequency, search for the value equal to $\frac{in}{4}$ or the first value immediately greater than it, such that:

$$N_{\uparrow} \geq \frac{in}{4}$$

- Compute the quartile value using the following formula:

$$Q_I = L_{QI} + \frac{\frac{IN}{4} - N_{\uparrow QI-1}}{n_{QI}} \times K_{QI}$$

- L_{QI} : The inferior bound of the quartile class.
- N : The total frequency, $N = \sum_{i=1}^k n_i$

- $N \uparrow_{Q_{I-1}}$: The ascending cumulative frequency of the class preceding the quartile class.
- n_{Q_I} : The absolute frequency of the quartile class.
- K_{Q_I} : The width (length) of the quartile class.

Example

Suppose we have the following values, representing the monthly income level for a sample of workers:

Monthly income level x_i	Number of workers n_i
[20 - 30 [6
[30 - 40 [14
[40 - 50 [22
[50 - 60 [18
[60 - 70 [12
[70 - 80 [2
Total	80

Required: Compute Q_1 , Q_2 , and Q_3 . What can you conclude?

Solution

Monthly income level x_i	Number of workers n_i	$N_1 \uparrow$
[20 - 30 [6	6
[30 - 40 [14	20
[40 - 50 [22	42
[50 - 60 [18	60
[60 - 70 [12	72
[70 - 80 [2	80
Total	80	

Computing Q_1

- Determine the quartile position:

$$\frac{in}{4} = \frac{80}{4} = 20$$

- In the table, locate the first class whose ascending cumulative frequency is equal to or greater than 20.

- The quartile class is: [30 - 40 [
- Compute the first quartile using:

$$Q_I = L_{Q_I} + \frac{\frac{IN}{4} - N \uparrow_{Q_I - 1}}{n_{Q_I}} \times K_{Q_I}$$

$$Q_1 = 30 + \frac{20 - 6}{14} \times 10$$

$$Q_1 = 30 + 10$$

$$Q_1 = 40$$

Computing Q_2 (Median)

- Determine the quartile position:

$$\frac{in}{4} = \frac{2(80)}{4} = 40$$

- The quartile class is: [40 - 50 [
- Computation of the second quartile

$$Q_I = L_{Q_I} + \frac{\frac{IN}{4} - 20}{n_{Q_I}} \times K_{Q_I}$$

$$Q_2 = 40 + \frac{40 - 20}{22} \times 10$$

$$Q_2 = 49.1$$

Computing Q_3

- Determine the quartile position:

$$\frac{in}{4} = \frac{3(80)}{4} = 60$$

- The quartile class is: [50 - 60 [
- Computation of the third quartile

$$Q_I = L_{Q_I} + \frac{\frac{IN}{4} - 20}{n_{Q_I}} \times K_{Q_I}$$

$$Q_3 = 50 + \frac{60 - 42}{18} \times 10$$

$$Q_3 = 60$$

Interpretation

- 25% of the workers earn a monthly income less than or equal to 40 thousand.
- 50% of the workers earn a monthly income less than or equal to 7 thousand.
- 75% of the workers earn a monthly income less than or equal to 10 thousand.

2. Deciles

Deciles are among the descriptive statistical measures that are concerned with determining the position of values within ordered data. The objective is to divide the data into ten equally sized segments to establish the centering of values based on the number of observations. These values are denoted by the symbols D_1 (the first decile), D_5 (the fifth decile), up to D_9 (the ninth decile).

The distance between each decile and the next represents 10% (one tenth) of the total data in the statistical series. Deciles help determine the relative position of values within a statistical distribution with greater precision compared to quartiles. (Gravetter & Wallnau, 2017, p. 77)

Where:

- D_1 : The value below which 10% of the data fall.
- D_2 : The value under 20% of the data fall.
- D_3 : The value below which 30% of the data fall.
- D_4 : The value below which 40% of the data fall.
- D_5 : Equivalent to the median (50%).
- D_6 : The value under 60% of the data fall.
- D_7 : The value below which 70% of the data fall.
- D_8 : The value below which 80% of the data fall.
- D_9 : The value below which 90% of the data fall. (Moore et al., 2017, p. 100)

2.2.Uses of Deciles in the Social Sciences

Deciles are used in the social sciences because they provide a more detailed breakdown of the distribution of values, especially in studies that require the classification of individuals or phenomena into multiple levels. Among their most important uses are:

- **Detailed analysis of the distribution of psychological and social variables:** such as classifying levels of anxiety, intelligence, attitudes, or job satisfaction into ten categories.
- **Comparison between individuals or groups:** by determining the position of an individual (or a group) within the distribution; for example, stating that a given individual falls in the eighth decile in terms of achievement or performance.
- **Study of distributional asymmetry:** where deciles contribute to revealing the concentration of values in specific parts of the distribution.
- **Use in large-scale educational and social studies:** particularly in standardized tests and social surveys that require precise classification of results. (Healey, 2015, p. 114)

2.3.Computing Deciles

1. In the Case of Ungrouped Data

Deciles are computed in the case of ungrouped data by following specific steps that mainly depend on ordering the data and the number of observations(n), as outlined below:

1. **Ordering the data:** The process must begin by arranging the data in ascending order (from the smallest to the largest) as an essential preliminary step.
2. **Determining the position (rank) of the decile:** The following relation is used to determine rank of the decile (i) is calculated using:

$$D_k = \frac{i(n+1)}{10}$$

3. **Determining the value :**The decile value is the observation that occupies this rank directly:

$$D_k = x_{\frac{i(n+1)}{10}}$$

4. If $\frac{i(n+1)}{10}$ is an integer, the decile value is the observation whose rank corresponds exactly to that position.

5. If $\frac{i(n+1)}{10}$ is not an integer, the decile value is taken as the observation occupying the immediate rank.
6. In the case where two ranks correspond to the decile, the arithmetic mean of the two corresponding values is calculated to obtain the decile value i

Example:

Suppose we have the following values, which represent the scores of a group of students on a psychological test:

6, 9, 4, 12, 7, 10, 5

Required: Compute D_1 , D_5 , and D_8 . What can you conclude?

Solution

- **Ordering the values in ascending order:**

12, 10, 9, 7, 6, 5, 4

$n=7$

Computing D_1

$$D_k = \frac{i(n+1)}{10}$$

$$D_1 = \frac{1(n+1)}{10}$$

$$D_1 = \frac{1(7+1)}{10}$$

$$D_k = 0.8$$

$$D_1 = x_{(0.8)}$$

The rank of the first decile D_1 is 0.8, and its value is the number that occupies the first position in the ordered series, which is 1.

The results indicate that:

- 10% of the students obtained a score less than or equal to 1.
- 90% of the students obtained a score greater than 1.

Computing D_5

$$D_5 = \frac{2(n+1)}{10}$$

$$D_5 = \frac{5(7+1)}{10} = \frac{16}{10} = 4$$

$$D_5 = x_{(4)}$$

$$D_5 = 7$$

The rank of the fifth decile D_5 is 4, and its value is the number that occupies the fourth position in the ordered series, which is 7.

The previous results indicate that:

- 50% of the students obtained a score less than or equal to 7.
- 50% of the students obtained a score greater than 7.

2. Deciles in the Case of Grouped Data Using a Frequency Distribution

Discrete Quantitative Variable

In the case of a discrete quantitative variable, deciles are calculated by following the steps and procedures described in statistical references, as follows:

- Calculate the ascending cumulative frequency in order to determine the position of the rank among the values.
- Determine the rank, which is calculated using the same formula: $\frac{in}{10}$
- In the column of the ascending cumulative frequency, search for the value equal to $\frac{in}{10}$ or the first value immediately greater than it, such that:

$$N_i \uparrow \geq \frac{in}{10}$$

- The value x_i corresponding to the previously identified ascending cumulative frequency represents the decile value.

Example:

Suppose we have the following values, which represent the number of books read during one year by a sample of students:

Number of books read x_i	Frequencies n_i
1	3
2	7
4	12
6	15
8	9
10	4
12	2
Σ	52

Required: Compute D_1 , D_3 , and D_5 . What can you conclude?

Solution :

x_i	n_i	$N_i \uparrow$
1	3	3
2	7	10
4	12	22
6	15	37
8	9	46
10	4	50
12	2	52
Σ	52	

Computing D_1

- Determine the rank of the first decile:

$$\frac{in}{10} = \frac{52}{10} = 5.2$$

- In the ascending cumulative frequency column, search for the value equal to or immediately greater than 5.2.
- The value n_i corresponding to the ascending cumulative frequency 10 is 2.

$$D_1 = 2$$

- 10% of the students obtained a value less than or equal to 2.
- 90% of the students obtained a value greater than 2.

Computing D_5 (Median)

- Determine the fifth decile's rank:

$$\frac{in}{10} = \frac{2(52)}{10} = 26$$

- In the increasing frequency column, search for the value equal to or immediately greater than 26.
- The value n_i corresponding to the ascending cumulative frequency 37 is 6.

$$D_5 = 6$$

- 50% of the students obtained a value less than or equal to 6.
- 50% of the students obtained a value greater than 6.

Computing D_8

- Determine the rank of the eighth decile:

$$\frac{in}{4} = \frac{8(52)}{4} = 41.6$$

- In the ascending cumulative frequency column, search for the value equal to or immediately greater than 41.6.
- The value n_i equivalent to the ascending cumulative frequency 46 is 8.

$$D_8 = 8$$

- 80% of the students achieved a value less than or equal to 8.
- 20% of the students obtained a value superior than 8.

Discrete Quantitative Variable

To calculate deciles for a discrete quantitative variable, follow the processes and techniques given in statistical references, as outlined below:

- ✓ Compute the ascending cumulative frequency in order to determine the position of the rank among the values.
- ✓ Determine the rank, which is calculated using the same formula: $\frac{in}{10}$

- ✓ In the column of the ascending cumulative frequency, search for the value equal to $\frac{in}{10}$ or the first value immediately greater than it, such that:

$$N_{i\uparrow} \geq \frac{in}{10}$$

- ✓ The decile value is calculated using the following formula:

$$D_i = L_{DI} + \frac{\frac{IN}{10} - N_{DI-1\uparrow}}{n_{DI}} \times K_{DI}$$

Where:

- L_{DI} : the lower bound of the decile class.
- N : the total frequency,

$$N = \sum_{i=1}^k n_i$$

- $N_{DI-1\uparrow}$: The ascending cumulative frequency of the class preceding the decile class.
- n_{DI} : The absolute frequency of the decile class.
- K_{DI} : The width (length) of the decile class.

Example

Suppose we have the following values, which represent the monthly income level of a sample of workers:

Monthly income level x_i	Number of workers n_i
[20 - 30 [6
[30 - 40 [14
[40 - 50 [22
[50 - 60 [18
[60 - 70 [12
[70 - 80 [2
Total	80

Required: Calculate D_1 , D_5 , and D_8 . What can you conclude?

Solution:

Monthly income level x_i	Number of workers n_i	$N_1 \uparrow$
[20 - 30 [6	6
[30 - 40 [14	20
[40 - 50 [22	42
[50 - 60 [18	60
[60 - 70 [12	72
[70 - 80 [2	80
Total	80	

Computing D_1

- Determining the decile class:

$$\frac{in}{10} = \frac{80}{10} = 8$$

- Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 20.
- The decile class is: [30–40[
- Computing the first decile:

$$D_1 = L_{D_1} + \frac{\frac{IN}{10} - N_{\uparrow D_1 - 1}}{n_{D_1}} \times K_{D_1}$$

$$D_1 = 30 + \frac{8-6}{14} \times 10$$

$$D_1 = 30 + 1.43$$

$$D_1 = 31.43$$

Computing D_5 (Median)

- Determining the decile class:

$$\frac{in}{10} = \frac{2(80)}{10} = 40$$

- Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 40.
- The decile class is: [40–50[
- Computing the fifth decile:

$$D_5 = L_{D_5} + \frac{\frac{IN}{10} - N_{\uparrow D_5 - 1}}{n_{D_5}} \times K_{D_5}$$

$$D_5 = 40 + \frac{40-20}{22} \times 10$$

$$D_5 = 49.1$$

Computing D_8

- Determining the decile class:

$$\frac{in}{10} = \frac{8(80)}{10} = 64$$

- Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 64.
- The decile class is: [60–70[
- Computing the eighth decile:

$$D_I = L_{Di} + \frac{\frac{IN}{10} - N_{\uparrow DI-1}}{n_{DI}} \times K_{DI}$$

$$D_8 = 60 + \frac{64-60}{12} \times 10$$

$$D_8 = 63.33$$

3. Percentiles

Percentiles are considered measures of positional tendency in descriptive statistics. They are used to divide data arranged in ascending order into one hundred approximately equal parts, such that each part contains 1% of the total number of values. Percentiles aim to determine the relative position of values within a statistical distribution with very high precision, which makes them an effective tool for analyzing both individual and group data in the social sciences. (Gravetter & Wallnau, 2017, p. 78)

Percentiles are denoted by the symbols:

$${}^iP P_1, P_2, P_3, \dots, P_{99}$$

Where:

- P_1 : the value below which 1% of the data fall.
- P_{50} : equivalent to the median (50%).
- P_{90} : the value below which 90% of the data fall.

1.3. Uses of Percentiles in the Social Sciences

Percentiles are used in the social sciences in the following ways:

1. **Determining the relative position of individuals or phenomena:** Percentiles are used in psychological and educational tests to determine the position of a student or an individual within the distribution of results (for example, identifying a student's performance relative to their study group).
2. **Detailed data classification:** Percentiles enable researchers to divide data into very precise categories, which facilitates the analysis of subtle differences between groups.
3. **Performance and achievement evaluation:** Percentiles are used in educational reports and standardized tests to assess individuals' performance at different levels (such as identifying the highest or lowest twenty percent).
4. **Detection of outliers:** Percentiles help analyze the distribution of data and identify extreme or anomalous values within a precise range. (Healey, 2015, p. 115)

Percentiles are considered the most precise measures of positional tendency and are frequently used in major social and psychological research, where analyzing an individual's position within the distribution is of great importance, such as in educational studies or standardized psychological tests.

Percentiles are highly precise partitioning tools that provide researchers with microscopic-level details about data distribution. While the median divides the data into two halves, percentiles divide it into 100 parts, allowing for extremely accurate identification of the position of any observation relative to the rest of the population.

2. Relationship Between Percentiles and Other Measures

The literature indicates the existence of fixed relationships linking percentiles with the median, quartiles, and deciles, as follows:

- **The fiftieth percentile (C_{50}):** is equivalent to the median (Me), the second quartile (Q_2), and the fifth decile (D_5).
- **The twenty-fifth percentile (C_{25}):** is equivalent to the first quartile (Q_1).
- **The seventy-fifth percentile (C_{75}):** is equivalent to the third quartile (Q_3).

- **The tenth percentile (C_{10}):** corresponds to the first decile (D_1).

3.2. Computing Percentiles

1. In the Case of Ungrouped Data

Percentiles are computed in the case of ungrouped data by following specific steps that mainly depend on ordering the data and the number of observations (n), as follows:

1. **Ordering the data** The process must begin by arranging the data in ascending order (from the smallest to the largest) as an indispensable preliminary step.
2. **Determining the position (rank) of the percentile** The rank of the percentile (i) is calculated using the following relation:

$$P_k = \frac{i(n+1)}{100}$$

3. **Determining the value** The percentile value is the observation that occupies this rank directly:

$$P_k = x_{\frac{i(n+1)}{100}}$$

- If $\frac{i(n+1)}{100}$ is an integer, the value whose rank corresponds directly to that position is taken.
- If $\frac{i(n+1)}{100}$ is not an integer, the value whose rank corresponds directly to that position is taken.
- In the case where there are two ranks for the percentile, the arithmetic mean of the two corresponding values is calculated in order to obtain the percentile value i .

Example

Suppose we have the following values, which represent the scores of a group of students on a psychological test:

6, 9, 4, 12, 7, 10, 5

Required: Compute P_{10} , P_{50} , and P_{70} . What can you conclude?

Solution

- **Ordering the values in ascending order:**

12, 10, 9, 7, 6, 5, 4

$n=7$

Computing P_I

$$P_I = \frac{i(n+1)}{100}$$

$$P_{10} = \frac{10(n+1)}{100}$$

$$P_{10} = \frac{10(7+1)}{100}$$

$$P_k = 0.8$$

$$P_I = x_{(0.8)}$$

The rank of the tenth percentile P_{10} is 0.8, and its value is the number that occupies the first position in the ordered series, which is 1.

The results indicate that:

- 10% of the students obtained a score less than or equal to 1.
- 90% of the students obtained a score greater than 1.

Computing P_{50}

$$P_{50} = \frac{50(n+1)}{100}$$

$$P_{50} = \frac{50(7+1)}{100} = \frac{400}{100} = 4$$

$$P_{50} = x_{(4)}$$

$$P_{50} = 7$$

The rank of the fiftieth percentile P_{50} is 4, and its value is the number that occupies the fourth position in the ordered series, which is 7.

The previous results indicate that:

- 50% of the students obtained a score less than or equal to 7.
- 50% of the students obtained a score greater than 7.

2. Percentiles in the Case of Grouped Data Using a Frequency Distribution

1. Discrete Quantitative Variable

Percentiles are computed in the case of a discrete quantitative variable by following the steps and procedures described in statistical references, as follows:

- Compute the ascending cumulative frequency in order to determine the position of the rank among the values.
- Determine the rank, which is calculated using the same formula: $\frac{in}{100}$
- In the column of the ascending cumulative frequency, search for the value equal to $\frac{in}{100}$ or the first value immediately greater than it, such that:

$$N_1 \uparrow \geq \frac{in}{100}$$

- The value n_i corresponding to the previously identified ascending cumulative frequency represents the percentile value.

Example

Suppose we have the following values, which represent the number of books read during one year by a sample of students:

Number of books read x_i	Frequencies n_i
1	3
2	7
4	12
6	15
8	9
10	4
12	2
Σ	52

Required: Compute P_{12} , P_{50} , . What can you conclude?

Solution :

x_i	n_i	$N_i \uparrow$
1	3	3
2	7	10
4	12	22
6	15	37
8	9	46
10	4	50
12	2	52
Σ	52	

Computing P_i

- Determine the rank of the twelfth percentile:

$$\frac{in}{100} = \frac{12(52)}{100} = 6.24$$

- In the ascending cumulative frequency column, search for the values that are equal to or immediately greater than 6.42.
- The value n_i corresponding to the ascending cumulative frequency 10 is 2.

$$P_{12} = 2$$

- 12% of the students have a number of books read less than or equal to 2.
- 88% of the students have a number of books read greater than 2.

Computing P_{50} (Median)

- Determine the rank of the fiftieth percentile:

$$\frac{in}{100} = \frac{50(52)}{100} = 26$$

- In the ascending cumulative frequency column, search for the values that are equal to or immediately greater than 26.
- The value n_i corresponding to the ascending cumulative frequency 37 is 6.

$$P_{50} = 6$$

- 50% of the students have a number of books read less than or equal to 6.
- 50% of the students have a number of books read greater than 6.

2.3. Discrete Quantitative Variable

Percentiles are computed in the case of a discrete quantitative variable by following the

steps and procedures described in statistical references, as follows:

- Compute the ascending cumulative frequency in order to determine the position of the rank among the values.
- Determine the rank, which is calculated using the same formula:

$$\frac{in}{100}$$

- In the column of the ascending cumulative frequency, search for the value equal to $\frac{in}{100}$ or the first value immediately greater than it, such that:

$$N_{i\uparrow} \geq \frac{in}{100}$$

- The percentile value is computed using the following formula:

$$P_I = L_{PI} + \frac{\frac{IN}{100} - N_{i\uparrow PI-1}}{n_{PI}} \times K_{PI}$$

where:

- L_{PI} : The lower bound of the percentile class.
- N : The total frequency,

$$N = \sum_{i=1}^k n_i$$

- $N_{i\uparrow PI-1}$: the ascending cumulative frequency of the class preceding the percentile class.
- n_{PI} : the absolute frequency of the percentile class.
- K_{PI} the width (length) of the percentile class.

Example:

Suppose we have the following values, which represent the monthly income level of a sample of workers:

Monthly income level x_i	Number of workers n_i
[20 - 30 [6
[30 - 40 [14
[40 - 50 [22
[50 - 60 [18
[60 - 70 [12
[70 - 80 [2
Total	80

Required: Compute P_{78} , P_{50} , and P_{13} . What can you conclude?

Solution

Monthly income level x_i	Number of workers n_i	$N_1 \uparrow$
[20 - 30 [6	6
[30 - 40 [14	20
[40 - 50 [22	42
[50 - 60 [18	60
[60 - 70 [12	72
[70 - 80 [2	80
Total	80	

Computing P_{13}

- Determining the percentile class:

$$\frac{in}{100} = \frac{13(80)}{100} = 10.4$$

- Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 20.
- The percentile class is: [30–40[
- Computing the thirteenth percentile:

$$P_i = L_{Pi} + \frac{\frac{iN}{100} - N_{\uparrow DI-1}}{n_{DI}} \times K_{DI}$$

$$P_{12} = 30 + \frac{10.4 - 6}{14} \times 10$$

$$P_1 = 30 + 1.43$$

$$P_1 = 33.14$$

Computing P_{50} (Median)

- Determining the percentile class:

$$\frac{in}{100} = \frac{50(80)}{10} = 40$$

Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 40.

- The percentile class is: [40–50[
- Computing the fiftieth percentile:

$$P_I = L_{Pi} + \frac{\frac{IN}{100} - N_{DI-1}}{n_{DI}} \times K_{DI}$$

$$P_{50} = 40 + \frac{40 - 20}{22} \times 10$$

$$P_{50} = 49.1$$

Computing P_{78}

- Determining the percentile class:

$$\frac{in}{100} = \frac{78(80)}{100} = 62.4$$

- Search in the table for the class whose ascending cumulative frequency is equal to or immediately greater than 64.
- The percentile class is: [60–70[
- Computing the seventy-eighth percentile:

$$P_I = L_{Pi} + \frac{\frac{IN}{100} - N_{DI-1}}{n_{DI}} \times K_{DI}$$

$$P_{78} = 60 + \frac{62.4 - 60}{12} \times 10$$

$$P_{78} = 62$$

Suggested exercises :

Exercise 1:

The following values represent the test scores of a group of students:

5, 7, 9, 4, 6, 8, 10, 3

Required:

1. Compute:
 - P_{25}
 - P_{50}
 - P_{75}
2. What can you conclude from the results?

Exercise 2

The following values represent the number of weekly study hours for a sample of students:

2, 4, 6, 3, 5, 8, 7, 4, 6, 5

Required:

1. Compute:
 - D_1
 - D_5
 - D_9
2. Interpret the results statistically.

Exercise 3

The following table represents the number of annual absences for a sample of students:

Number of absences x_i	Frequencies n_i
0	4
1	6
2	10
3	12
4	8
5	5

Required:

1. Compute:
 - Q_1
 - Q_2
 - Q_3
2. Provide an interpretation for each quartile.
- 3.

Exercise 4

The following table represents the number of books read during one year by a sample of students:

Number of books x_i	Frequencies n_i
1	5
2	9
3	11
4	15
5	10

Required:

1. Compute:
 - D_2
 - D_5
 - D_8
2. Interpret the results in terms of the distribution of the number of books read.

Exercise 5

The following table represents the distribution of monthly wages (in thousands) for a sample of workers:

Monthly wage x_i	Number of workers n_i
[20–30[8
[30–40[14
[40–50[20
[50–60[18
[60–70[10

Required:

1. Identify the quartile classes.
2. Compute:
 - Q_1
 - Q_2
 - Q_3
3. What do the results indicate about wage levels?

Exercise 6

The following table shows the results of exam scores for a sample of students:

Scores x_i	Number of students n_i
[0–5[6
[5–10[12
[10–15[18
[15–20[14
[20–25[10

Required:

1. Compute:
 - P_{10}
 - P_{50}
 - P_{90}
2. Interpret the results from an educational perspective.

Exercise 7

The following table represents the monthly income level (in thousands) for a sample of employees:

Monthly income x_i	Number of employees n_i
[25–35[10
[35–45[18
[45–55[22
[55–65[20
[65–75[10

Required:

1. Calculate:
 - Q_1, Q_2, Q_3
 - D_4, D_7
 - P_{25}, P_{60}, P_{85}
2. Provide a comprehensive statistical analysis that highlights:
 - Income central tendency
 - Income dispersion
 - The relative position of individuals within the distribution

When comparing two data sets, the frequency distribution or the frequency curve can be used, as well as some measurement of central tendency, for example the “arithmetic mean, median, and mode”. Though, the use of these methods alone is not sufficient for comparison, especially when the measures of central tendency of the two data sets are equal. There may be a significant difference between both data groups in terms of the degree of closeness or dispersion of the data from one another, or in terms of how far the values deviate from the measure of central tendency.

If we calculate the arithmetic mean for two data sets, it is possible to find that the arithmetic mean of both data sets is equal, even though the values of the second data set are more homogeneous than those of the first data set. Therefore, statisticians have resorted to using other measures to evaluate the degree of data homogeneity or the extent of data dispersion around measures of central tendency. These measures can be used in comparing two or more data sets, including measures of dispersion and kurtosis, and this section will focus on these measures.

Measures of dispersion: Dispersion indicates the extent to which data values depart from their arithmetic mean. The higher the values of the dispersion measures, the greater the degree of separation and variation among the data values; conversely, smaller values indicate that the differences amongst the data values are limited. As a result, these measurements provide information about the degree of homogeneity or variability in the data, as well as its distribution.

There are several statistical methods for measuring dispersion, which differ in terms of accuracy and ease of use. Among these measures are the range, mean deviation, variance, and standard deviation.

Some measures evaluate the closeness or dispersion of values relative to one another, such as the range and the quartile deviation, other measurements, such as the arithmetic mean, mean deviation, and standard deviation, examine the proximity or distance of data from a certain value.

1.2. Statistical properties of variance:

- Variance is always non-negative.
- It equals zero only when all values are identical.
- It is highly affected by extreme values.

- Its unit is the square of the original unit of measurement, which makes it less straight forward to interpret (OpenIntro Statistics, 2019).

1.3. Statistical interpretation

The standard deviation is characterized by having the same unit as the original variable, which makes it one of the most:

- widely used
- interpretable
- accepted measures in applied research (Triola, 2019).

Applied significance

- A small standard deviation → high homogeneity.
- A large standard deviation → clear individual variability.

Therefore, the standard deviation is widely used in:

- interpreting the results of psychological measurements,
- comparing group performance,
- Evaluating the stability of behavioral phenomena (Field, 2018).

2. Variance:

Variance is a measure of dispersion that is commonly utilized in practical applications. It represents the average deviation of values from their arithmetic mean. In calculating variance, squared deviations are used in order to avoid the use of absolute values, as is the case with mean deviation. Variance is signified by σ^2 for population data and by S^2 for sample data. The superior the value of the variance, the more dispersed the distribution.

1.2. Sample variance (S^2):

In many cases, the population variance σ^2 is unknown. In such situations, a sample is drawn from the population, and the variance is calculated from the sample data as an estimate of the population variance. If a random sample of size n has the following observations: x_1, x_2, \dots, x_n , the sample variance (S^2) is given by:

2.2.Properties of variance:

- It has the same properties as the standard deviation.
- It is not affected by adding or subtracting a constant from the variable x .
- It is affected by multiplying or dividing the variable x by a constant.
- To determine variance, subtract the squares of the root and arithmetic mean.

2.3.Uses of variance and standard deviation in the social sciences:

The importance of variance and standard deviation in the social sciences is evident in several areas, most notably:

- Variance and standard deviation contribute to the analysis of individual differences among individuals or groups in various psychological and social variables, such as attitudes, values, motivation, and job satisfaction.
- They are used to assess the degree of homogeneity of samples in psychological, educational, and social studies, allowing judgments about the degree of similarity of data within a sample.
- They enable comparisons between statistical distributions of similar social phenomena, particularly when analyzing the results of different studies.
- Variance is considered a foundation for many inferential statistical methods, such as analysis of variance (ANOVA), making it a fundamental element in interpreting statistical differences.
- They assist researchers in interpreting statistical results and assessing the trustworthiness of the arithmetic mean as a descriptive indication of data.
- Standard deviation measures data distribution around the arithmetic mean. A smaller standard deviation indicates minimal dispersion.
- Conversely, when data are widely spread, with the possible presence of extreme values on the two sides of the mean, the standard deviation is large, reflecting a superior degree of dispersion.
- Thus, the standard deviation clearly expresses the extent to which the data of a given phenomenon are close to or distant from their arithmetic mean.

2.4. Calculation of variance:**1. Ungrouped data:**

The variance for n observations x_1, x_2, \dots, x_n is calculated using the following formula:

For a population:

$$\sigma^2 = \frac{\sum(x_i - X)^2}{n}$$

For a sample:

$$s^2 = \frac{\sum(x_i - X)^2}{n - 1}$$

Example:

The following values represent a sample drawn from a population:

45, 60, 55, 70, 65, 80, 75, 50, 90, 85

Required: Calculate the variance in this case.

Solution:

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{45+60+55+70+65+80+75+50+90+85}{10}$$

$$\bar{X} = \frac{675}{10} = 67.5$$

$$s^2 = \frac{\sum(x_i - X)^2}{n - 1}$$

$$s^2 = \frac{(45 - 67.5)^2 + (60 - 67.5)^2 + (55 - 67.5)^2 + (70 - 67.5)^2 + (65 - 67.5)^2 +$$

$$= \frac{(80 - 67.5)^2 + (75 - 67.5)^2 + (50 - 67.5)^2 + (90 - 67.5)^2 + (85 - 67.5)^2}{10 - 1}$$

$$s^2 = \frac{(-22.5)^2 + (-7.5)^2 + (-12.5)^2 + (2.5)^2 + (-2.5)^2 + (12.5)^2 + (7.5)^2 + (17.5)^2 + (22.5)^2 + (17.5)^2}{10 - 1}$$

$$+ \frac{506.25 + 56.25 + 156.25 + 6.25 + 6.25 + 156.25 + 56.25 + 306.25 + 506.25 + 306.25}{9}$$

$$s^2 = \frac{2062.5}{9} = 229.17$$

2. Grouped data using a frequency distribution

Discrete variable:

If X_1, X_2, \dots, X_k (are the values of the variable x), and $n_1, n_2, n_3, \dots, n_k$ (are their frequencies), the variance is defined by the following formula:

For a population:

$$\sigma^2 = \frac{\sum n_i(x_i - X)^2}{n}$$

For a sample:

$$s^2 = \frac{\sum n_i(x_i - X)^2}{n - 1}$$

Example:

The following table represents the data:

Score x_i	Frequency n_i
50	2
60	3
70	5
80	4
90	1
Total	15

Required: Calculate the variance.

Solution:

Score x_i	Frequency n_i	$n_i x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$n_i(x_i - \bar{X})^2$
50	2	100	50 - 69.33 = -19.33	373.54	747.08
60	3	180	9.33	87.05	261.15
70	5	350	0.67	0.45	2.25
80	4	320	10.67	113.78	455.12
90	1	90	20.67	427.07	427.07
Σ	15	1040			1892.67

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum n_i x_i}{\sum n_i}$$
$$\bar{X} = \frac{100 + 180 + 350 + 320 + 90}{15}$$

$$\bar{X} = \frac{1040}{15} = 69.33$$

Population variance:

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n}$$
$$\sigma^2 = \frac{1895.67}{15} = 126.18$$

Sample variance:

$$s^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n - 1}$$
$$s^2 = \frac{1892.67}{14} = 135.19$$

Continuous variable:

If X_1, X_2, \dots, X_k (are the values of the variable x) and $n_1, n_2, n_3, \dots, n_k$ (are their frequencies), the variance is defined by the following formula:

For a population:

$$\sigma^2 = \frac{\sum n_i (c_i - \bar{X})^2}{n}$$

For a sample:

$$s^2 = \frac{\sum n_i (c_i - \bar{X})^2}{n - 1}$$

Where c_i is the class midpoint of class i

Example: The table below shows the distribution of job satisfaction scores for a group of workers.

Class interval x_i	Frequency n_i
[49-40]	6
[59-50]	10
[69-60]	12
[79-70]	8
[89-80]	4
Total	40

Required: Calculate the variance.

Solution:

Class interval x_i	Frequency n_i	c_i	$c_i n_i$	$c_i - \bar{X}$	$(c_i - \bar{X})^2$	$n_i(c_i - \bar{X})^2$
[49-40]	6	44.5	267	44.5-63= -18.5	342.25	6x342.25 = 2053.5
[59-50]	10	54.5	545	-8.5	72.25	722.5
[69-60]	12	64.5	774	1.5	2.25	27
[79-70]	8	74.5	596	11.5	132.25	1058
[89-80]	4	84.5	338	21.5	462.25	1846
Σ	40		2520			5709

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum c_i x_i}{\sum n_i}$$

$$\bar{X} = \frac{2520}{40} = 63$$

Population variance:

$$\sigma^2 = \frac{\sum n_i (c_i - \bar{X})^2}{n}$$

$$\sigma^2 = \frac{5709}{40}$$

$$\sigma^2 = 142.73$$

Sample variance:

$$s^2 = \frac{\sum n_i(c_i - X)^2}{n - 1}$$

$$s^2 = \frac{5709}{39}$$

$$s^2 = 146.38$$

3. Standard Deviation:

When variance is used as a measure of dispersion, it is based on the sum of squared deviations; thus, this measure is incompatible with the units of measurement of the variable under study. For this reason, statisticians adopted a logical measure that takes into account the square root of the variance so that it becomes consistent with the units of measurement of the variable. This measure is the standard deviation.

3.1. Definition of the standard deviation:

The standard deviation is regarded as the most essential and commonly used measure of dispersion. It is defined as the square root of the mean of the squared deviations of values from the arithmetic mean, and it is calculated according to the nature of the data under consideration. The standard deviation represents the square root of the variance, and it allows one to determine the quantity to which data are spread about their arithmetic mean. The lesser the standard deviation, the inferior the dispersion of values in the distribution; the superior the standard deviation, the larger the dispersion. The standard deviation of a population is signified by the symbol σ , while the standard deviation of a sample is denoted by the symbol S .

Thus, the standard deviation is the positive square root of the variance, that is:

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

3.2. Properties of the standard deviation:

- It is one of the most widely used measures of dispersion in descriptive statistics.
- It measures the extent to which statistical values deviate from their arithmetic mean.

- Its calculation is based on all available observations without exception.
- It has the same unit of measurement as the original variable, which facilitates the interpretation of its results.
- It is amenable to algebraic operations; therefore, it is widely used in statistical formulas and models.
- It is an important indicator of the degree of data homogeneity; small values indicate closeness of observations and low dispersion, whereas large values indicate high dispersion.
- The model is heavily influenced by extreme values because to its reliance on squared deviations from the mean, while values near to the mean have minimal impact.
- It is influenced by the arithmetic mean, as it represents the reference point relative to which deviations are calculated.
- It does not change when a constant value is added to or subtracted from all observations, as the dispersion of data around the mean remains unchanged.
- Its value is always positive, or equal to zero only when all values are identical.
- It can be used to compare two statistical distributions of the same type that have the same arithmetic mean.
- It can not be computed from open-ended frequency distribution tables.
- Deviation considered easy to understand, calculate, and apply in descriptive studies.

3.3. Standard Deviation Calculation:

The standard deviation is calculated as follows:

Ungrouped Data:

The standard deviation of n observations X_1, X_2, \dots, X_n is calculated using the following formula:

For a population:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}}$$

For a sample:

$$s = \sqrt{\frac{\sum(x_i - X)^2}{n - 1}}$$

Example: Based on the results of the previous example:

Population variance:

$$\sigma^2 = \frac{1895.67}{15} = 126.18$$

$$\sigma = \sqrt{126.18}$$

$$\sigma = 11.23$$

Sample variance:

$$S^2 = \frac{1892.67}{14} = 135.19$$

$$s = 11.62$$

Continuous Variable:

If X_1, X_2, \dots, X_k are the values of the variable x and $n_1, n_2, n_3 \dots \dots, n_k$ are their frequencies, the standard deviation is defined by the following formula:

For a population:

$$\sigma = \sqrt{\frac{\sum n_i(c_i - X)^2}{n}}$$

For a sample:

$$s = \sqrt{\frac{\sum n_i(c_i - X)^2}{n - 1}}$$

where c_i is the midpoint of class i

Example: Based on the results of the previous example:

Population variance:

$$\sigma^2 = \frac{5709}{40}$$

$$\sigma^2 = 142.73$$

$$\sigma = \sqrt{142.73}$$

$$\sigma = 11.94$$

Sample variance:

$$s^2 = \frac{5709}{39}$$

$$s^2 = 146.38$$

$$s = \sqrt{146.38}$$

$$s = 12.09$$

Suggested exercises

Exercise 1

The following values represent the number of hours a group of students spend studying per week:

6, 8, 9, 10, 10, 11, 12, 14

Required:

1. Calculate the range.
2. Compute the arithmetic mean.
3. Calculate the variance.
4. Calculate the standard deviation.
5. Interpret the value of the standard deviation in the context of the data.

Exercise 2

The following table shows the frequency distribution of household income (in monetary units):

Income Intervals	Frequency
10,000 – 20,000	5
20,000 – 30,000	9
30,000 – 40,000	14
40,000 – 50,000	8

Required:

1. Calculate the arithmetic mean for the grouped data.
2. Compute the variance for the grouped data.
3. Calculate the standard deviation.
4. Comment on the degree of income dispersion.

Inferential statistics is considered one of the fundamental pillars of contemporary scientific research, as it aims to draw conclusions about the characteristics of a statistical population based on data collected from a limited sample. The scientific value of these conclusions stems from their degree of accuracy and their ability to represent the statistical reality of the population, which necessitates reliance on statistical measures that estimate the degree of uncertainty associated with the process of generalization. In this context, the standard error (SE) emerges as a pivotal statistical tool used to assess the precision of sample statistics—most notably the arithmetic mean—in comparison with the true but unknown population values (Gravetter & Wallnau, 2017, pp. 12–114).

1.The standard error

The standard error is a statistical measure that represents the level of precision in estimating a population parameter, such as the mean, using data from a sample. It represents the expected amount of variability in the value of the sample mean if a large number of equivalent samples were randomly picked from the same population. (Field, 2018, pp. 52–53).

1.1.The statistical definition of the standard error :

From a statistical perspective, the standard error is defined as the standard deviation of the sampling distribution of sample means from the true population mean. In other words, it assesses how closely the means of different samples obtained from the same population converge or diverge from the true value of the overall mean. (Howell, 2013, p. 68).

The lower the standard error, the more the sample mean is considered an accurate and credible representation of the population mean., which enhances the possibility of generalizing the results with a higher level of statistical confidence (Kirkwood & Sterne, 2003, pp. 46).

1.2.Calculation of the standard error :

The standard error of the mean is determined by the inverse correlation between the sample standard deviation and the sample size, whereby the standard error diminishes as the sample size enlarges. In practical applications, the population standard deviation is frequently unknown.

;therefore, it is replaced by the sample standard deviation according to the following formula:

$$SE = \frac{s}{\sqrt{n}}$$

where:

SE: the standard error of the mean.

s: the sample standard deviation (Standard Deviation).

n: the sample size (number of individuals).

The difference between the standard error and the standard deviation :

The distinction between the standard deviation (SD) and the standard error (SE) represents one of the conceptual issues that are frequently confused by novice researchers. Despite the mathematical relationship between the two measures, each serves a distinct and specific statistical function (Field, 2018, pp. 56).

Table (1): Illustrates the differences between the Standard Deviation (SD) and the Standard Error (SE).

Aspect	Standard Error (SE)	Standard Deviation (SD)
Measurement aim	Indicates the accuracy of the sample mean as an estimate of the population mean and reflects the dispersion of sample means	Indicates the dispersion of individual observations within the sample around their mean
Applied context	Inferential statistics (confidence intervals, hypothesis testing)	Descriptive statistics
Relationship to sample size	Decreases as sample size increases according to \sqrt{n}	Not directly affected by sample size
Statistical interpretation	Accuracy of external estimation	Internal variability of the data

1.4.The relationship between the standard error and sample size;

The presence of the square root of the sample size in the denominator of the mathematical expression causes the standard error to drop inversely as the sample size increases. This means that large samples provide more accurate and stable estimates of the population mean compared with small samples (Howell, 2013, p. 73).

It should be noted that this decrease is not linear but follows the square root law, which implies that achieving a doubling in precision requires a relatively significant increase in sample size.

1.5. The importance of the Standard Error and its applications

The importance of the standard error resides in its role as an essential tool in inferential statistical analysis, as it is utilized in various fundamental contexts, among them:

- **Measuring estimation accuracy:** It provides an indicator of how near the sample mean is to the actual population mean.
- **Constructing confidence intervals:** It is a key component in calculating confidence intervals, such as the 95% confidence interval, which is approximately estimated as:
Sample mean $\pm 1.96 \times SE$
- **Statistical hypothesis testing:** It is directly involved in the calculation of test statistics such as t and z tests, as well as analysis of variance (ANOVA).
- **Comparing means:** It helps determine whether differences between group means are statistically significant or merely due to chance.
- **Interpreting result stability:** It allows for the assessment of the stability of estimates when sampling is repeated or when multiple samples are compared (Gravetter & Wallnau, 2017, p. 118).

Example:

A researcher collected data on the number of weekly study hours for 25 students from a particular university. The results (in hours) were as follows: 12, 15, 14, 10, 13, 16, 12, 14, 11, 13, 15, 12, 14, 13, 16, 14, 12, 15, 13, 14, 12, 15, 13, 14, 12

Required:

Calculate the standard error of the mean.

Solution:

Calculation of the arithmetic mean:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{335}{25}$$

$$\bar{X} = 13.4$$

Calculation of the standard deviation:

$$s^2 = \frac{\sum(x_i - \bar{X})^2}{n - 1}$$
$$s^2 = \frac{54}{24} = 4.63$$
$$s = \sqrt{2.25} = 1.5$$

Calculation of the standard error:

$$SE = \frac{s}{\sqrt{n}}$$
$$SE = \frac{1.5}{\sqrt{25}}$$
$$SE = \frac{1.5}{5}$$
$$SE = 0.3$$

The standard error is an indispensable statistical concept in scientific research, as it represents the link between sample characteristics and the possibility of generalizing results to the statistical population. Understanding its nature as a measure of estimation accuracy, clearly distinguishing it from the standard deviation, and comprehending its inverse relationship with sample size are all elements that enable students and researchers to design more robust and reliable studies and to make inferential decisions based on sound statistical foundations.

Exercise

The following data represent the test scores of a group of students:

12, 14, 15, 16, 18, 20, 22, 23

Required:

1. Calculate the arithmetic mean.
2. Calculate the standard deviation.
3. Compute the coefficient of variation.
4. Interpret the value of the coefficient of variation in terms of data dispersion.

In descriptive statistics, measures of dispersion aim to show the extent to which values differ around a measure of central tendency—such as the arithmetic mean or the median. Methodologically, they are divided into absolute measures of dispersion and relative measures of dispersion. The following is a presentation of these measures.

1. Absolute Measures of Dispersion:

They express dispersion in the original units of measurement of the variable (cm, kg, points, etc.); therefore, they are affected by the scale of measurement.

2. Relative Measures of Dispersion:

They express dispersion as a percentage and are useful for comparing variables with different units or measurement scales. One of the most important of these measures is the coefficient of variation.

3. Definition:

Absolute measures of dispersion generally depend on the units of measurement associated with the variable under study, which makes their use inappropriate when there is a need to compare the degree of dispersion between statistical distributions that differ in units of measurement, such as comparing the dispersion of lengths with the dispersion of weights. This limitation is considered one of the most prominent methodological issues in descriptive statistical analysis, which necessitates resorting to alternative measures that overcome the effect of the unit of measurement.

The coefficient of variation is considered one of the most important relative measures of dispersion. It is defined as a measure that reflects the magnitude of the standard deviation relative to the arithmetic mean and is usually expressed as a percentage. The coefficient of variation is calculated according to the following mathematical relationship (NIST, 2023):

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100 \%$$

That is:

$$CV = \frac{S}{\bar{X}} \times 100 \%$$

This measure is known by several terms in the statistical literature, including the coefficient of variation, coefficient of change, and relative standard deviation. It is

commonly abbreviated as CV and is primarily used to compare the degree of variability in two or more datasets or samples, even when their units of measurement differ. It is also considered an effective tool for estimating the degree of homogeneity within data, as low values of the coefficient of variation indicate a high level of homogeneity, whereas high values reflect increased dispersion and weak homogeneity among values.

Note: The greater the dispersion, the larger the coefficient of variation.

4. Advantages of the Coefficient of Variation:

- It allows comparison of the degree of dispersion between variables that differ in units of measurement or in the nature of the measurement scales.
- It is a relative measure that is neutral with respect to the unit of measurement.
- It shows dispersion relative to the level of the mean.
- It is excellent for comparing series with different means and units.
- It is easy to calculate and easy to interpret since it is expressed as a percentage.
- It takes all values constituting the distribution into account because it is based on the standard deviation.
- It is suitable for comparing different tests and variables that are heterogeneous in terms of their means (OpenIntro Statistics, 2019).

5. Disadvantages:

- It is used only with quantitative data measured on a ratio scale.
- It is not valid when the arithmetic mean is close to zero.
- It is not recommended for use with variables that have an artificial zero (such as some Likert scales) (Field, 2018).
- Its interpretive accuracy depends on the accuracy of both the standard deviation and the arithmetic mean used in its calculation.
- Any error in estimating the standard deviation is directly reflected in the value of the coefficient of variation and its interpretation.

6. Interpretation of the Coefficient of Variation Results:

- $CV < 10\%$: Low dispersion
- $10\% \leq CV < 30\%$: Moderate dispersion

- $CV \geq 30\%$: High dispersion

Calculating the coefficient of variation: The coefficient of variation is calculated as follows:

Ungrouped data:

The coefficient of variation for ungrouped data is calculated according to the following steps:

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum X_i}{n}$$

Calculating the variance:

$$s^2 = \frac{\sum(x_i - \bar{X})^2}{n - 1}$$

Calculating the standard deviation:

$$CV = \frac{s}{\bar{X}} \times 100\%$$

Example:

The level of occupational stress was measured among 12 employees in the Human Resources Department at Sonelgaz, using a scale from 0 to 10, where a higher score indicates greater stress. The recorded data for each employee are as follows: 2, 5, 7, 3, 6, 4, 8, 5, 7, 3, 6, 9

Required: Calculate the coefficient of variation and interpret it.

Solution:

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{9+6+3+7+5+8+4+6+3+7+5+2}{12}$$

$$\bar{X} = \frac{65}{12} = 5.42$$

$$s^2 = \frac{\sum(x_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{(9 - 5.42)^2 + (6 - 5.42)^2 + (3 - 5.42)^2 + (7 - 5.42)^2 + (5 - 5.42)^2 + (8 - 5.42)^2 + (4 - 5.42)^2 + (6 - 5.42)^2 + (3 - 5.42)^2 + (7 - 5.42)^2 + (5 - 5.42)^2 + (2 - 5.42)^2}{12 - 1}$$

$$s^2 = \frac{11.69 + 0.17 + 2.50 + 5.86 + 0.33 + 2.01 + 6.65 + 0.17 + 2.50 + 5.86 + 0.33 + 12.81}{11}$$

$$s^2 = \frac{50.94}{11} = 4.63$$

$$s = \sqrt{4.63} = 2.15$$

Calculating the coefficient of variation:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

$$CV = \frac{2.15}{5.42} \times 100\%$$

$$CV = 39.66$$

Interpretation: CV = 39.66% indicates a moderate to high dispersion in the level of occupational stress among employees.

Grouped Data Using a Frequency Distribution

Discrete Variable:

If X_1, X_2, \dots, X_k are the values of the variable x , and $n_1, n_2, n_3, \dots, n_k$ are their frequencies, the coefficient of variation can be calculated according to the following steps:

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum n_i x_i}{\sum n_i}$$

Calculating the standard deviation:

$$s^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n - 1}$$

$$s = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{n - 1}}$$

Calculating the coefficient of variation:

$$CV = \frac{S}{\bar{X}} \times 100\%$$

Example:

The following table presents the frequency distribution of the variable “workers’ job satisfaction” among 40 workers at Sonelgaz (Production Unit), based on a five-point rating scale (from 1 to 5):

Score x_i	Frequency n_i
1	5
2	8
3	12
4	10
5	5
Total	40

Required: Calculate the coefficient of variation and interpret it.

Solution:

Score x_i	Frequency n_i	$n_i x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$n_i(x_i - \bar{X})^2$
1	5	5	1-3.05 = -2.05	4.20	21.01
2	8	16	1.05-	1.10	8.82
3	12	63	0.05-	0	0.03
4	10	40	0.95	0.9	9.02
5	5	25	1.95	3.80	19.01
Σ	40	122			57.9

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum n_i x_i}{\sum n_i}$$

$$\bar{X} = \frac{5 + 16 + 63 + 40 + 25}{40}$$

$$\bar{X} = \frac{122}{40} = 3.05$$

Calculating the standard deviation:

$$s^2 = \frac{\sum n_i(x_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{57.9}{40-1} = 1.48$$

$$s = \sqrt{1.48}$$

$$s = 1.21$$

Calculating the coefficient of variation:

$$CV = \frac{s}{\bar{X}} \times 100\%$$

$$CV = \frac{1.21}{3.05} \times 100\%$$

$$CV = 39.67$$

Interpretation: 39.67% indicates moderate to high dispersion in workers' job satisfaction around the mean. That is, job satisfaction among employees is highly heterogeneous, and there is a clear disparity between more satisfied and less satisfied workers.

Continuous Variable:

If X_1, X_2, \dots, X_k are the values of the variable x , and $n_1, n_2, n_3, \dots, n_k$ are their frequencies, the coefficient of variation can be calculated according to the following steps:

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum c_i X_i}{\sum n_i}$$

Calculating the standard deviation:

$$s^2 = \frac{\sum c_i (x_i - \bar{X})^2}{n - 1}$$

$$s = \sqrt{\frac{\sum n_i (c_i - \bar{X})^2}{n - 1}}$$

Calculating the coefficient of variation:

$$CV = \frac{s}{\bar{X}} \times 100\%$$

Example:

A social skills assessment was conducted among 30 employees in the Finance Department using a quantitative scale ranging from 0 to 9, and the results were distributed into the following classes:

Skill range x_i	Frequency n_i
[2 -0]	3
[4 -3]	5
[6 -5]	12
[8 -7]	8
[10 -9]	2
Total	

Required: Calculate the coefficient of variation and interpret it.

Solution:

Skill range x_i	Frequency n_i	c_i	$c_i n_i$	$c_i - \bar{X}$	$(c_i - \bar{X})^2$	$n_i(c_i - \bar{X})^2$
[2 -0]	3	1	3	1-5.51= -4.51	20.34	3x20.34= 61.02
[4 -3]	5	3.5	17.5	- 2.01	4.04	20.2
[6 -5]	12	5.5	66	0.01	0.0001	0.0012
[8 -7]	8	7.5	60	1.99	3.96	31.68
[10 -9]	2	9.5	19	3.99	15.92	31.84
Σ	30		165.5			144.74

Calculating the arithmetic mean:

$$\bar{X} = \frac{\sum c_i x_i}{\sum n_i}$$

$$\bar{X} = \frac{165.5}{30} = 5.51$$

Calculating the standard deviation:

$$s^2 = \frac{\sum n_i(c_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{144.74}{29}$$

$$s^2 = 4.99$$

$$s = \sqrt{4.99}$$

$$s = 2.23$$

Calculating the coefficient of variation:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

$$CV = \frac{2.23}{5.51} \times 100\%$$

$$CV = 40.47$$

Interpretation: CV = 40.47% indicates moderate to high dispersion in social skills among employees. The overall conclusion, based on the coefficient of variation (CV) as an indicator of relative dispersion, is that a precise comparison could be made between social skills, job satisfaction, and occupational stress despite differences in measurement scales, which supports the methodological soundness of the procedure.

- Social skills: (CV = 40.47%)
- Workers' job satisfaction: (CV = 39.67%)
- Occupational stress: (CV = 39.66%)

The comparison among the coefficients of variation for social skills, job satisfaction, and occupational stress shows a notable convergence in the level of relative dispersion, as all variables reflect moderate to high dispersion within the sample.

Exercise :

The following data represent the test scores of a group of students:

12, 14, 15, 16, 18, 20, 22, 23

Required:

1. Calculate the arithmetic mean.
2. Calculate the standard deviation.
3. Compute the coefficient of variation.
4. Interpret the value of the coefficient of variation in terms of data dispersion.

In addition to measures of central tendency and measures of dispersion, we will address two measures that illustrate the shape of the statistical distribution (skewness and kurtosis) in comparison with a reference distribution (the symmetric and normal distribution). These statistical tools are referred to as measures of shape.

Before addressing this issue, let us become acquainted with the most important forms that any frequency distribution may take. This is naturally determined by the nature of the distribution, or in other words, by the distribution of observations around a specific central value, which is often the arithmetic mean; thus, the frequency distribution may be.

Symmetric, if:

That is, 50% of the observations lie to the right and 50% to the left of these measures.

This form is considered a standard form, relative to which all remaining forms are measured.

Positively skewed: if it is more extended toward the right, and in this case it is

Negatively skewed: if it is more extended toward the left, and in this case it is

1. Definition of perfect symmetry:

It refers to a curve which, if divided into two halves, these two halves coincide perfectly with one another. That is, the frequency curve is symmetric, such that the left side with respect to the horizontal axis of the coordinate system represents the exact central point of the distribution, known as the point of symmetry. This point is equal to the arithmetic mean, the median, and the mode, and therefore provides a good representation of the distribution.

2. Definition of skewness:

Skewness represents the degree of symmetry or asymmetry of the data of any phenomenon under study. Through the frequency curve, the shape of the distribution becomes apparent. Skewness refers to the lack of symmetry of the phenomenon around its central value (the arithmetic mean), whereby the conditions of perfect symmetry are not satisfied.

If the frequencies are concentrated at the smaller values, the curve becomes positively skewed toward the right.

The distribution (curve) is positively skewed to the right (arithmetic mean $>$ median $>$ mode).

However, when the frequencies are concentrated at the larger values, the curve in this

case becomes negatively skewed toward the left.

The distribution is negatively skewed to the left (arithmetic mean < median < mode).

The graphical forms of frequency distributions vary: some are symmetric, where the measures of central tendency (arithmetic mean, median, and mode) are equal, while others are asymmetric, being skewed either to the right or to the left. This allows for the summarization and comparison of distributions in addition to measures of central tendency and measures of dispersion, which alone are insufficient to perform this task (that is, the process of comparison). Two distributions may be equal in terms of the mean and the standard deviation, yet differ in terms of skewness; they may have skewness in the same direction but with different magnitudes, or they may have equal degrees of skewness but differ in sign. Moreover, the type of skewness (positive or negative) and its degree (mild or pronounced) can be inferred from the shape of the curve itself; however, this does not provide a numerical measure of skewness. For this purpose, it has become important to identify a statistical tool used to measure skewness, known as the coefficient of skewness.

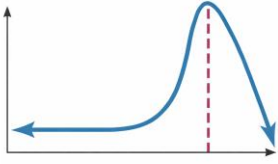
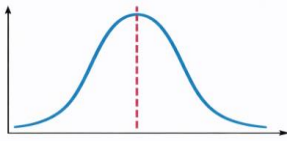
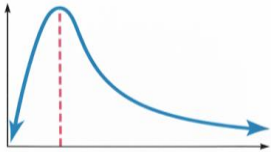
When representing the data of a phenomenon in the form of a frequency curve, the curve may take different shapes. It may be symmetric, meaning that it has a central value, and if a vertical line is dropped from its peak onto the horizontal axis, it divides the curve into two identical halves.

2.1.Coefficient of skewness:

Measures of skewness refer to the deviation of the frequency curve from symmetry. Symmetry means that dividing the frequency curve vertically at its midpoint yields two perfectly identical parts, whereas asymmetry implies that the two parts do not coincide, in which case the distribution is skewed either to the right or to the left. Graphically, a frequency distribution curve is skewed to the right when extreme values on the right influence the arithmetic mean and pull it toward the right; consequently, the mean is greater than the median. Conversely, if the distribution is skewed to the left, the arithmetic mean is smaller than the median. These cases can be illustrated as follows:

- Right-skewed distribution: $Mo < Me < \bar{X}$
- Symmetric distribution: $Mo = Me = \bar{X}$
- Left-skewed distribution: $Mo > Me > \bar{X}$

Figure (20): Illustrates the types of skewness in statistical distributions (positive skewness, symmetrical distribution, and negative skewness).

<p>Negative Skew (Left-Skewed)</p>  <p>Most of the data is concentrated on the right and a little on the left</p>	<p>Symmetrical Distribution</p>  <p>Data is equally distributed on both sides of the distribution.</p>	<p>Positive Skewness (Skewed to the Right)</p>  <p>Most of the data is focused on the left and little of it on the right.</p>
<p>The distribution is asymmetric and skewed to the left. $M_o > M_e > \bar{X}$</p>	<p>The distribution is symmetric. $M_o = M_e = \bar{X}$</p>	<p>The distribution is asymmetric and skewed to the right $M_o < M_e < \bar{X}$</p>

The shape of the bars or the histogram often makes it possible to perceive and clarify the nature of symmetry or asymmetry in a given distribution. Measures of skewness may also be used, as they assess the degree of skewness of the frequency curve and the imbalance in the distribution. The objective is to compare the shapes of distributions, and such comparisons are meaningful when they are carried out using the same parameters, even if they are applied to different distributions. That is, when comparing the skewness of different distributions, the same coefficient must be used. Measures of skewness take different forms, whether simple or composite, among the most important of which are:

2.2.Measurement of skewness:

Skewness can be measured by the coefficient of skewness, which helps determine the degree of symmetry or skewness of a distribution. Although there are several measures of skewness, among the most important are the following:

1.Yule–Kendall coefficient:

Also known as the quartile coefficient of skewness, Yule’s coefficient measures the skewness of a distribution by taking into account the relative positions of the quartiles with respect to the median. It is calculated using the following formula:

$$Y_k = (Q_1 + Q_3 - 2Me) / (Q_3 - Q_1)$$

For a symmetric distribution, the distances satisfy $(Me - Q_1) = (Q_3 - Me)$; therefore, $k_k = 0$

The distribution is skewed to the left when $(Me - Q_1) > (Q_3 - Me)$; therefore, $Y_k < 0$

The distribution is skewed to the right when $(Me - Q_1) < (Q_3 - Me)$; therefore, $Y_k > 0$

2. Coefficient of skewness:

To determine the value of skewness for a given phenomenon, the following formula is used, which is based on the arithmetic mean minus the mode:

$$WA = \bar{X} - Mo$$

3. Pearson's coefficient of skewness:

The formula of Pearson's coefficient of skewness is based on dividing the amount of skewness by the standard deviation. It is particularly useful in the case of grouped data and can be calculated using the following formula:

$$CA = \frac{\bar{X} - Mo}{S}$$

Pearson's coefficient of skewness can also be expressed using the median, as follows:

$$CA = \frac{3(\bar{X} - Me)}{S}$$

The coefficient of skewness is bounded between +1 and -1. If it is positive, this indicates that the distribution extends to the right; if it is negative, this indicates that the distribution extends to the left. When it equals zero, this means that the curve (distribution) is symmetric. The division of the skewness value by the standard deviation is intended to eliminate units of measurement, thereby enabling the comparison of skewness between two different phenomena.

Example:

The following values represent a sample drawn from a population:
45, 60, 55, 70, 65, 80, 75, 50, 90, 85

The required :task is to determine the value of skewness using Pearson's coefficient.

Solution:

Calculation of the arithmetic mean:

$$\bar{X} = (\sum x_i) / n = (45 + 60 + 55 + 70 + 65 + 80 + 75 + 50 + 90 + 85) / 10$$

$$\bar{X} = 675 / 10 = 67.5$$

Calculation of the standard deviation:

$$s^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{2062.5}{9} = 229.17$$

$$s = \sqrt{229.17}$$

$$s = 15.13$$

Calculation of the median:

Arrange the values in ascending order:

45, 50, 55, 60, 65, 70, 75, 80, 85, 90

Determination of the median position:

Since the data are ungrouped and the number of values is even, the median is the average of the two middle values:

$$Me = (x(n/2) + x(n/2 + 1)) / 2$$

$$Me = (70 + 65) / 2$$

$$Me = 67.5$$

Calculation of the skewness value:

$$WA = 3(\bar{X} - M_e)$$

$$WA = 3(67.5 - 67.5)$$

$$WA = 0$$

The distribution is symmetric and balanced.

Pearson's coefficient of skewness:

$$CA = \frac{3(\bar{X} - M_e)}{s}$$

$$CA = \frac{3(67.5 - 67.5)}{15.13} = 0$$

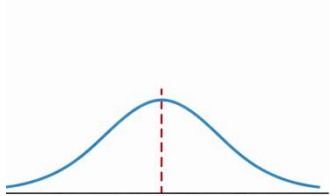
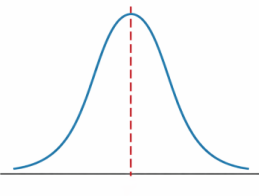
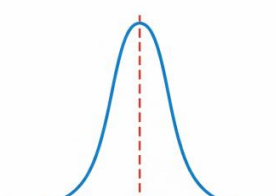
Accordingly, the distribution is symmetric.

3.Measures of kurtosis (flatness):

The coefficient of kurtosis is a statistical measure that expresses the degree of peakedness or flatness of a frequency distribution curve in comparison with the symmetric normal curve. It reflects the extent to which values are concentrated around the arithmetic mean and the manner in which the curve extends. The more dispersed the distribution, the more extended and less peaked the curve becomes, and vice versa. Although the coefficient of skewness is used to determine whether a frequency curve is symmetric or not, symmetry alone is insufficient to judge the normality of a distribution, as a curve may be symmetric yet non-normal, being either too peaked or too flat. Hence arises the need for the coefficient of kurtosis as a complementary measure that determines the nature of the curve's shape in terms of its degree of convexity or flatness.

The term *kurtosis* is used to describe the extent of elongation or lack thereof of the frequency curve or frequency polygon. If the curve is flatter than the normal curve, it is considered a platykurtic curve, characterized by a wider center and a lower peak compared to the peak of the normal curve. Conversely, if it is more peaked than the normal curve, it is referred to as a leptokurtic curve, distinguished by a narrower center and a higher peak. When the curve coincides with the normal (moderate) curve, it is described as a symmetric normal curve, and in this case the value of the coefficient of kurtosis equals (3), which is the characteristic value of the normal distribution.

Figure (21): Illustrates the types of kurtosis in statistical distributions (platykurtic, mesokurtic, and leptokurtic).

 <p style="text-align: center;">Platykurtic</p>	 <p style="text-align: center;">Normal (Gaussian)</p>	 <p style="text-align: center;">Leptokurtic</p>
<p>A platykurtic distribution means that: CF > 3</p>	<p>A mesokurtic (moderately peaked or symmetric) distribution means: CF = 3</p>	<p>A leptokurtic distribution means: CF < 3</p>

The kurtosis of frequency curves can be measured using several measures, among which

are:

1. Fisher's coefficient:

It is given by the following formula:
the fourth central moment divided by the standard deviation raised to the fourth power:

$$CF = U_4 / S^4$$

It should be noted in this regard that the mesokurtic curve (moderate kurtosis) is considered a reference for determining the nature of kurtosis. Practically, it has been found that the kurtosis coefficient of a symmetric (normal) distribution equals 3. Accordingly, the number 3 constitutes the basis for distinguishing among curves in terms of peakedness and flatness, as follows:

- A mesokurtic (moderately peaked) distribution means: $CF = 3$
- A leptokurtic (more peaked) distribution means: $CF < 3$
- A platykurtic (flatter) distribution means: $CF > 3$

If Fisher's coefficient is positive, this indicates that the distribution is less flat than the normal distribution; if it is negative, this indicates that the distribution is more flat, that is, more peaked than the normal distribution.

2. Kelly's coefficient:

Kelly's coefficient of kurtosis can be calculated using the following relationship:

$$CK = \frac{Q_3 - Q_1}{D_9 - D_1}$$

Example:

The following table shows the distribution of job satisfaction scores for a group of workers:

Class interval x_i	Frequency n_i
[49-50]	6
[50-55]	10
[55-60]	12
[60-65]	8
[65-70]	4

Total	40
-------	----

Required: Study the shape of the following frequency distribution curve using Fisher's coefficients.

Solution:

Class interval x_i	n_i	c_i	$c_i n_i$	$c_i - \bar{X}$	$n_i (c_i - \bar{X})$	$n_i (c_i - \bar{X})^2$	$n_i (c_i - \bar{X})^3$	$n_i (c_i - \bar{X})^4$
[49-40]	6	44.5	267	18.5-	111-	2053.50	-37 989.75	702 822.38
[59-50]	10	54.5	545	8.5-	85-	722.5	-6 141.25	52 200.63
[69-60]	12	64.5	774	1.5	18	27	40.50	60.75
[79-70]	8	74.5	596	11.5	92	1058	12 167	139 920.50
[89-80]	4	84.5	338	21.5	86	1849	39 793.50	855 600.38
Σ	40		2520	0		5709	7 870	1 750 604.63

Calculation of the arithmetic mean :

$$\bar{X} = \frac{\sum c_i x_i}{\sum n_i}$$

$$\bar{X} = \frac{2520}{40} = 63$$

Calculation of the standard deviation:

$$s^2 = \frac{\sum n_i (c_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{5709}{39}$$

$$s^2 = 146.38$$

$$s = \sqrt{146.38}$$

$$s = 12.09$$

$$U_4 = \frac{\sum n_i (c_i - \bar{X})^4}{\sum n_i}$$

$$U_4 = \frac{1750604.63}{40}$$

$$U_4 = 43765.11$$

Calculation of the coefficient of kurtosis using Fisher's coefficient:

$$CF = \frac{U_4}{S^4}$$
$$CF = \frac{43765.11}{21365.11}$$
$$CF = 2.05$$

Reference value for the normal distribution:

$$CF = 3$$

Since Fisher's coefficient of kurtosis for this distribution is 2.05, which is less than 3, the distribution is platykurtic, that is, it is less peaked than the normal distribution.

Exercise 1:

The following data represent the monthly income (in monetary units) of a small group of individuals:

10, 12, 14, 15, 16, 18, 25

Required:

1. Calculate the arithmetic mean.
2. Determine the median.
3. Identify the mode (if any).
4. Based on the relationship between the mean, median, and mode, determine the type of skewness (positive, negative, or symmetric).

Exercise 2:

Consider the following two data sets representing test scores of two groups of students:

- **Group A:** 10, 12, 14, 16, 18
- **Group B:** 14, 15, 16, 17, 18

Required:

1. Compare the concentration of values around the center for each group.
2. Identify which group shows a higher degree of peakedness.
3. State whether each distribution is likely to be:
 - Leptokurtic (peaked),
 - Mesokurtic (normal),
 - Platykurtic (flat).

Conclusion

Statistics constitutes an essential methodological tool that is indispensable for the scientific analysis of social phenomena based on accuracy and objectivity. This work has sought to present statistics in a comprehensive manner, beginning with its fundamental concepts, moving through methods of data organization and presentation, and culminating in the various statistical measures that enable researchers to describe data, analyze it, and understand the nature of its distribution.

The study has addressed measures of central tendency, dispersion, association, and distribution shape as interconnected components of statistical analysis, emphasizing that reliance on a single measure may lead to a partial or even misleading interpretation of the phenomenon under investigation. Particular attention has been given to the practical applications of these measures in the field of social sciences, thereby linking theoretical foundations with empirical practice and enhancing the reader's ability to apply statistical methods effectively in scientific research.

While this publication does not claim to cover all aspects of statistics exhaustively, it aims to establish a solid conceptual and methodological foundation that enables learners to progress confidently toward more advanced levels of study. It also seeks to develop the capacity for responsible and critical interpretation of statistical results. Ultimately, statistics remains a means for advancing human knowledge and understanding reality, rather than an end in itself—an objective this work aspires to achieve for both readers and researchers alike.

References:

Abdeljawad, M. (2013). *Social statistics*. Dar Al-Masira for Publishing, Distribution, and Printing.

Ahmed, S. (2008). *Principles of statistics: Applications and practical exercises using SPSS*. International House for Cultural Investments.

Al-Bahth Academy. (2023). *Statistics in modern management*.

Al-Basha, M. H., & Al-Da'mah, A. (2013). *Fundamentals of statistics with SPSS applications*. Dar Al-Manahij for Publishing and Distribution.

Al-Boulqawas, Z. (2014). *Lectures in descriptive statistics*. Faculty of Humanities and Social Sciences, Mohamed Khider University of Biskra.

Al-Fayez, A. (2021). *Statistics in the age of big data*. University Publishing House.

Al-Mashhadani, K., & Al-Obaidi, R. (2013). *Scientific research methodology*. Dar Al-Ayyam for Publishing and Distribution.

Al-Omari, A., & Al-Ajeeli, A. (2000). *Statistics and probability: Theory and application*. ELGA Publications.

Al-Saifo, W. (2010). *Fundamentals of statistical methods*. Zamzam Publishers and Distributors.

Al-Saifo, W., Bakr, A., & Al-Rifai, G. (2010). *Fundamentals of statistical methods for business and their applications in financial, administrative, and economic sciences*. Zamzam Publishers and Distributors.

Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2020). *Statistics for business & economics* (14th ed.). Cengage Learning.

Azzouz, A. (2010). *The complete guide to statistics: Detailed lessons, exercises, and solved problems* (Part 1). National Office of University Publications.

Babbie, E. (2020). *The practice of social research*. Cengage Learning.

Babbie, E. (2021). *The basics of social research*. Cengage Learning.

Ben Farhat, S., & Qattoush, A. (2013–2014). *Lectures in statistics I supported with exercises and solved examinations*. Setif, Algeria.

Benyah, S. (2017–2018). *Lectures in descriptive statistics (Statistics 01)*. Faculty of Economic Sciences, Commercial Sciences, and Management Sciences, Tiaret.

Bernardo, J., & Smith, A. (2000). *Bayesian theory*. Cambridge University Press.

- Coolican, H. (2018). *Research methods and statistics in psychology*. Psychology Press.
- Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Dawson, B., & Trapp, R. (2009). *Basic & clinical biostatistics*. McGraw-Hill Medical.
- Dwivedi, Y., Hughes, L., Baabdullah, A., & Koohang, A. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver & Boyd.
- Gelman, A., & Nolan, D. (2017). *Teaching statistics: A bag of tricks*. Oxford University Press.
- Graunt, J. (1662). *Natural and political observations made upon the bills of mortality*. John Martyn.
- Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the behavioral sciences* (10th ed.). Cengage Learning.
- Haidoushi, A. (2015–2016). *Lectures in descriptive statistics*. Akli Mohand Oulhadj University, Bouira.
- Hammoudi, S. (2009). *Principles of statistics and their applications in the educational and social fields*. Dar Al-Thaqafa for Publishing and Distribution.
- Hassadi, A. (2014). *Statistics in economics*. Arab Publishing House.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning*. Springer.
- Healey, J. (2015). *Statistics: A tool for social research* (9th ed.). Cengage Learning.
- Howell, D. (2013). *Statistical methods for psychology* (8th ed.). Cengage Learning.
- Jellato, J. (2002). *Statistics with exercises and solved problems*. National Office of University Publications.
- Jellato, J. (2010). *Statistics with exercises and solved problems* (8th ed.). National Office of University Publications.
- Kafi, M., et al. (2012). *Scientific research methods: Data collection and analysis techniques*. Dar Al-Thaqafa.

- Kirkwood, B. R., & Sterne, J. A. C. (2003). *Essential medical statistics* (2nd ed.). Blackwell Science.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.
- Montgomery, D., & Runger, G. (2014). *Applied statistics and probability for engineers*. John Wiley & Sons.
- Moore, D., McCabe, G., & Craig, B. (2017). *Introduction to the practice of statistics* (9th ed.). W. H. Freeman.
- O'Dwyer, L., & Bernauer, J. (2014). *Quantitative research for education: A guide to research design, data analysis, and presentation*. Pearson.
- OpenIntro. (2019). *Descriptive statistics*. <https://www.openintro.org/book/os>
- Pagano, M., & Gauvreau, K. (2018). *Principles of biostatistics*. Chapman and Hall/CRC.
- Pearson, E., & Kendall, M. (1970). *Studies in the history of statistics and probability: A series of papers* (Vol. 1). Charles Griffin & Co Ltd.
- Porter, T. (1986). *The rise of statistical thinking, 1820–1900*. Princeton University Press.
- Ratoul, M. (2006). *Descriptive statistics* (2nd ed.). National Office of University Publications.
- Salem, A. B., & Ababneh, A. (2007). *Principles of descriptive and inferential statistics* (1st ed.). Dar Al-Masira for Publishing, Distribution, and Printing.
- Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.
- Triola, M. (2019). *Elementary statistics* (13th ed.). Pearson.