K.A.MUKHITDINOVA

STATISTICS: A STUDY GUIDE

ISLOM KARIMOV NOMIDAGI TOSHKENT DAVLAT TEXNIKA UNIVERSITETI

GUVOHNOMA

O'QUV ADABIYOTINING NASHR RUXSATNOMASI

Toshkent davlat texnika universiteti Kengashining 4-son majlis qaroridan 10.02.2025 yildagi 01/9-05-97-sonli buyrug'iga asosan

Mukhitdinova K.A.

(muallif(lar)ning familiyasi, ismi-sharifi)

60310100-Economics (by industries and sectors), 60411200-Management

(ta'lim yo'nalishi (mutaxassisligi) shifri va nomi)

(management of innovative projects in industrial enterprises)

talabalari uchun tavsiya etilgan

Statistics

The manual (O'quv qo'llanma) ga

(o'quv adabiyotning nomi va turi: darslik,o'quv qo'llanma)

O'zbekiston Respublikasi Vazirlar Mahkamasi tomonidan litsenziya berilgan nashriyotlarda nashr etishga ruxsat berildi.

Rektor ______ **S.M.Turabdjanov**

(imzo)

Royxatga olingan raqami

02-2025-0002

MINISTRY OF HIGHER EDUCATION, SCIENCE AND INNOVATIONS OF THE REPUBLIC OF UZBEKISTAN

K.A.MUKHITDINOVA

STATISTICS:
A STUDY GUIDE

TASHKENT 2025

UO‘K: 38.45:330.320
KBK: 005.7(575.2)

M 43

Mukhitdinova K.A. Statistics. The manual. – T.: 2025.- 216 p.

“Statistika” fani bo’yicha tayyorlangan mazkur o’quv qo’llanmada iqtisodiyotda yuz berayotgan jarayonlarni miqdoriy baholash uchun dastlabki ma’lumotlarni to’plash, statistik ko’rsatkichlarni hisoblash, tahlil va talqin qilish, iqtisodiy diagnostika bilan shug’ullanish va iqtisodiy hodisa va jarayonlarni miqdoran va sifat jihatidan baholash bo’yicha chuqur nazariy va amaliy bilimlarni shakllantirish ko'rib chiqilgan. Bu o’quv qo’llanma “Korporativ boshqaruv” va “Iqtisodiyot (tarmoqlar va sohalar)” yo’nalishlari bakalavr talabalariga soha iqtisodiyoti bo’yicha muayyan bilimlarni egallashga yordamlashadi.

This manual, prepared on the subject "Statistics," considers the collection of preliminary data on the quantitative assessment of processes, which are taking place in economy, the calculation, analysis and interpretation of statistical indicators, the implementation of economic diagnostics and the formation of intimate theoretical and practical knowledge on the quantitative and qualitative assessment of economic processes and phenomena. This manual will help students of the directions of "Corporate Governance" and "Economics (branches and spheres)" master certain knowledge on economics of the sphere.

ISBN 978-9910-9339-9-8

Reviewers: Khalikulova G.T.- Professor of the Department of Economic Theory of TSEU, DSc;
Allaeva G.J.- Professor of the Department of Industrial Economics and Management of TSTU, DSc

Preface

In today's data-driven world, students gain a significant academic and professional advantage when they understand how to apply statistical thinking and analysis to real-world problems. Whether in economics, engineering, education, or public policy, statistical literacy has become essential for making informed decisions based on evidence.

This English-language edition was prepared with the aim of serving students enrolled in English-taught programs at universities across Uzbekistan, particularly those studying statistics and related fields. The book was originally written in Uzbek by Dr. Kamola Muhitdinova, Doctor of Science in Economics at Tashkent State Technical University. As an experienced academic and researcher, Dr. Muhitdinova identified a shortage of statistical textbooks in English that reflect the economic context and educational needs of Uzbekistan. Motivated by this gap, she undertook the translation of her own work to ensure that students could benefit from a textbook tailored to the national environment while also meeting the standards of international academic literature.

Statistics is a foundational discipline in both the natural and social sciences. It equips learners with tools for data collection, analysis, interpretation, and informed decision-making. This translated edition provides a precise and methodical presentation of key statistical concepts, from descriptive statistics to probability theory, hypothesis testing, and regression analysis. Each chapter is complemented by examples and exercises that foster a deeper understanding and encourage the practical application of the material.

A distinctive feature of this textbook is its integration of the statistical system of Uzbekistan as a case study for institutional development. By highlighting national practices and reforms—such as the State Statistics Agency under the President of the Republic of Uzbekistan—students are encouraged to explore how statistical methods operate not only in theory but in real policy-making environments.

This textbook introduces students to the fundamental principles and practical methods of statistics, with a focus on economic and social applications. It begins with the historical development of statistics and proceeds to cover key topics such as data collection, classification, tabulation, graphical representation, and the use of statistical indicators. The book emphasizes both descriptive and inferential techniques, guiding readers through real-world examples and exercises. Special attention is also given to the organization of statistical activities in Uzbekistan, providing valuable local context alongside globally accepted methodologies.

Behind the preparation of this book lies a personal and academic journey. Writing and translating this work was not only an academic endeavor, but also a response to a broader national need: to support Uzbekistan's educational modernization and the global orientation of its universities. The author's passion for the discipline, shaped by years of teaching and research, is reflected in the clarity, accessibility, and purpose of the material. By sharing this passion, the book aims to inspire students not just to study statistics, but to see it as a vital instrument in solving real-world challenges.

This translated edition is intended to support university-level instruction in statistics, enhance independent learning, and provide a resource for applied research. It can be effectively used in academic programs related to economics, public policy, management, and engineering, where statistical analysis plays a central role.

CONTENTS:

Introduction

Statistics is a science that deals with the collection, analysis, interpretation, and presentation of data. The main task of the science of statistics is to analyze the phenomena occurring in social, economic, and other fields and draw scientifically based conclusions.

With the help of statistics, large amounts of data can be easily understood and analyzed. This science is important in understanding, forecasting, and decision-making of socio-economic processes.

The development of the science of statistics is closely related to modern technologies and computer applications. Today, many programs and tools for collecting and analyzing statistical data can be used to analyze data quickly and accurately.

In this textbook, you will learn the basic concepts, methods, and practices of statistics. This tutorial will help you develop statistical analysis skills and solve real-life problems.

In recent years, a series of measures have been implemented in Uzbekistan to improve the statistical system. Notably, the Presidential Resolution of the Republic of Uzbekistan “On Additional Measures to Ensure Openness and Transparency in Public Administration and Enhance the Country's Statistical Potential” (PQ-4273[1]), issued on April 9, 2019, was adopted. As a result, several positive changes have been achieved in the statistical system. However, challenges remain in areas such as ensuring the openness of information for public use, improving the transparency and quality of open data, and refining the mechanisms for delivering this information to citizens and the global community. These shortcomings, which fail to fully meet international standards, have contributed to Uzbekistan's low position in the Statistical Capacity Index.

As President Sh. Mirziyoyev noted[2], "The more truthful and open the statistical system is, the more accurate and effective the reforms will be". The subject of "Statistics" plays a crucial role in training engineers

---

[1] https://lex.uz/docs/-4277342

[2] https://review.uz/uz/post/milliy-statistika-tizimini-rivojlantirish-islohotlar-samaradorligi-garovi

who understand both the theoretical and practical significance of this vital field and grasp all innovations and organizational changes in this area.

Acquiring knowledge in the science and practice of statistics requires developing skills in collecting, processing, and analyzing real statistical data. The science of statistics examines the general principles and methods for collecting, processing, analyzing, and interpreting data related to mass social phenomena and processes. Moreover, it studies, analyzes, and statistically forecasts events and processes occurring in the economy at both macro and micro levels. Additionally, it reflects the conditions, processes, and outcomes of the market economy, as well as the laws and trends of societal development, through statistical indicators.

CHAPTER 1. SUBJECT AND METHODS OF STATISTICS

1.1. The Emergence and Development of Statistics

The term "statistics" originates from the German word "statskunde" (state science), the Greek word "status" (stable, fixed), and the Latin word "status" (referring to the condition, state, or status of certain events or phenomena). Statistics is understood in two ways: as a science that studies quantitative aspects of developments that do not change in nature, and as an activity that records and expresses the quantitative aspects and numbers of mass phenomena.

In contemporary usage, the term statistics encompasses the following:

- A social science concerned with the methods of collecting, organizing, analyzing, and comparing mass phenomena and events.
- A practical activity focused on gathering, processing, and analyzing information about various events.
- Numerical data or information presented in a structured format.

Since ancient times, practical needs have led to the emergence of statistics. In order to determine the number of people qualified for the armed forces, to know the number of the population that can be taxed, there is a need for information on the age and gender structure of the population. This, in turn, led states to monitor changes in population size and composition. The progressive development of civil societal relations necessitated the systematic documentation of events directly connected to these relationships. As a result, the procedure for recording events such as births, deaths, marriages, and divorces were established. Subsequently, regulations were introduced to record the permanent migration of individuals from one place to another. In addition, the development of industry, agriculture, trade, craftsmanship and other areas, as well as economic relations, required regular monitoring of economic events and actions.

The history of the science of statistics is very ancient; its beginning goes back to the early stages of human civilization. Initially, statistics was used to collect and organize data. According to some sources, a population census was conducted in Egypt 3500 years before our era. In China, 2,000 years before our era, data were collected on the population size, age, and gender structure, as well as the state of industry and agriculture.

In ancient Egypt, Mesopotamia, and China, various data were collected for public administration purposes, such as population census, tax collection, and resource management. During this period, statistics were mainly used for state needs.

Statistics as a science originated in England in the 17th century. The founders of the school of "Political Arithmetic" U. Petty (1623-1687) and J. Graunt (1620-1674) are the founders of the science of statistics. The demographic direction dominated Graunt's scientific work, while U. Petty's work was in the economic-statistical direction.

Political arithmeticians sought to determine the state and development of events occurring in society based on quantitative calculations.

Almost simultaneously with the "Political Arithmetic" school, a descriptive school emerged in Germany. Representatives of this school (G. Achenwall, G. Conring, and A. Schloetzer) were engaged in describing, recording, and analyzing the state of the government.

German scholar and professor of philosophy and law, G. Achenwall (1719–1772), referred to his lectures on state studies at the University of Göttingen as “Statistics.” Thus, the term “statistics” began to be used in 1749.

In the 19th century, the science of statistics developed further and began to be used in various fields. The normal distribution and regression analysis introduced by Carl Gauss and Pierre-Simon Laplace made statistics widely used in scientific and applied research. During this period, statistics were integrated with social, economic, and natural sciences.

By the 40s of the 19th century, the use of descriptive approaches in statistics had nearly ceased and later ended.

In the first half of the 19th century, a mathematical approach to statistics began to emerge. The Belgian scientist A. Quetelet (1796–1874) greatly contributed to this direction. He called statistics "social physics", a science that studies social events using quantitative methods.

English scientists F. Talman, K. Pearson, and G. Fisher further developed the mathematical trend in statistics in the second half of the 19th century and the beginning of the 20th century.

In the 20th century, the development of the science of statistics accelerated with the introduction of new technologies and calculation methods. During this period, R.A. Scientists such as Fisher, Jerzy Neumann, and Egon Pearson developed statistical methodologies. According to Fisher's work, experimental design and special analysis methods were developed, which increased the accuracy and reliability of statistical research.

In modern times, statistical analysis is carried out with the help of computer technologies. Artificial intelligence and machine learning techniques have taken statistics to a new level as the possibilities of analyzing large volumes of data have expanded. Today, statistics are important in social, economic, natural, and technical sciences, as well as in business and industry.

1.2. Subject and Methods of Statistics

The science of statistics studies the quantitative aspect of mass-social events (economic, political, cultural, educational, educational, natural) within a specific time and region in an integral connection with the qualitative aspects.

The main features of the subject matter of statistics are as follows:

1) Statistics, as a science, studies not individual facts but mass social phenomena;

2) Statistics examines the quantitative aspect of mass-social phenomena in close relation to their qualitative aspects. The quantitative side of these social events means their number and scale, while the qualitative aspect denotes their internal character and fundamental essence;

3) Statistics as science identifies and elucidates the structural composition of social phenomena—their internal organizational patterns—and represents these through systematic statistical indicators and metrics.

The methodology of the statistics discipline refers to the set of methods, techniques, approaches, and tools used in the scientific study and management of mass social phenomena, as well as in the process of identifying the patterns manifested in these phenomena and applying them in practice. Statistical research aimed at calculating the quantitative relationships of mass social events and identifying the patterns they exhibit is carried out in several stages, which are further divided into phases. These stages and phases differ from one another in terms of their objectives, tasks, and characteristics. Therefore, each phase and stage employs specific investigation methods, approaches, and tools. At the same time, depending on the nature of the field under study and the specific problem at hand, a particular method (or a combination of methods) is used as the primary and leading tool in a specific phase and stage of the research.

In a broad sense, statistical research consists of the following two stages:

1. Descriptive statistics stage
2. Analytical statistics stage

In the descriptive statistics stage, the primary objectives and tasks include specifying the studied objects, collecting and processing data about them, calculating indicators that represent the quantitative norms of mass social phenomena, and presenting them in a clear and concise manner—sometimes accompanied by verbal descriptions when necessary.

In the analytical statistics stage, the main objectives and tasks involve conducting a statistical analysis of the indicators, identifying and evaluating cause-and-effect relationships between them, uncovering patterns in the distributions of the studied objects, assessing the reliability of scientific hypotheses, and drawing statistical conclusions.

Each stage, in turn, is divided into phases. For example, the descriptive statistics stage consists of several phases. First, it involves specifying the population of studied objects and mass social phenomena. Then, statistical observation is conducted on these objects. The next phase includes organizing and structuring the collected raw data, followed by the calculation of summary indicators. Finally, the results are presented clearly and concisely. The analytical statistics stage comprises multiple phases as well. It begins with constructing distributions of the studied objects based on various characteristics and identifying patterns within them. This is followed by quantifying relationships between different phenomena and analyzing trends in their development over time. Another essential phase involves evaluating scientific hypotheses and drawing statistical conclusions. The final phase consists of analyzing the interconnections between different aspects of complex processes as an integrated system.

In each phase of a given stage, various methods, tools, and approaches are used to study mass social phenomena. For example, in the phase of specifying objects, methods such as simple or complex classification, grouping based on elementary or hierarchical associations, various approaches to secondary grouping, and cluster analysis techniques are applied. During statistical monitoring, researchers use several methods to gather data. These include running experiments in production settings or laboratories, conducting special investigations and surveys, collecting data through questionnaires, and using sampling techniques for observations. In studying the interrelationships between phenomena, methods such as analytical grouping, constructing parallel series, visualizing data through diagrams, the balance method, correlation and regression analysis, analysis of variance (ANOVA), multivariate analysis methods, and other techniques are utilized.

The theory of statistics is a broad framework that applies to both stages of statistical research and all their phases. It explains the general principles and methodologies of statistical investigation, integrating the methods used with the nature of the mass social phenomena being studied.

When studying mass social phenomena quantitatively, statistics also utilizes mathematical methods. In particular, several statistical methods are based on algebra and number theory, including set theory, algebraic systems, systems of linear equations and inequalities, matrices, vectors and Euclidean spaces, determinants, and others. Statistical theory and mathematical statistics are usually considered as two independent disciplines. This distinction is based on differences in their mathematical apparatus. The theory of statistics employs elementary mathematical methods, while mathematical statistics uses more complex mathematical approaches.

Statistics emerges as an independent discipline and is applied in various fields because its subject—mass social phenomena—appears in all fields and possesses common characteristics. At the same time, in certain fields, these phenomena manifest under specific conditions and acquire particular attributes influenced by those conditions. Therefore, the way statistical laws emerge and their forms of development take place uniquely in each field. This necessitates adapting statistical methodologies to specific field conditions, applying appropriate methods, refining them as circumstances change, and replacing outdated techniques with new ones. As a result, the statistics of each field develop as a relatively independent discipline within the broader field of statistics while also serving as one of its essential branches and directions.

Economic statistics belongs to this category of sciences and constitutes a very important branch of the overall field of statistics. Similarly, within the framework of economic statistics, branch statistics are emerging and developing as independent directions - macroeconomic statistics, microeconomic statistics, and statistics of industry, agriculture, and other sectors and fields.

The subject of economic statistics is the study of the quantitative manifestation of economic laws in a specific time and area, examining the quantitative aspect of economic phenomena and events in close connection with their qualitative aspects.

Economic phenomena and processes consist of events at both macroeconomic and microeconomic levels in terms of their hierarchical structure.

Phenomena and processes that arise on the scale of the global, national, or large regional markets are referred to as macroeconomic processes. For example: the formation of supply and demand in global or national markets, the formation of national wealth, employment and unemployment rates, changes in price levels, depreciation of national currencies, inflation rates, investment processes, and their utilization.

Microeconomic processes refer to the flow of events that occur at the scale of an enterprise, firm, or within specific goods and services, labor, and financial markets. For example: the formation of demand or supply for a particular product in a market, the results of an enterprise's production and financial activities, the formation of a firm's economic assets and liabilities, and the utilization of labor, material, and financial resources.

Economic statistics is based on general statistical methods but has a methodology adapted to economic conditions.

In the process of studying macro and microeconomic phenomena and events, economic statistics makes extensive use of general statistical methods.

Economic statistics is an important tool of social cognition. It creates the opportunity to comprehend life events, understand the essence of economic theories, and examine their rules and concepts.

1.3 . Theoretical Foundations and Main Categories of Statistics

In a certain historical context of social development, these actions recur in a specific manner, meaning they are expressed through distinct patterns or laws. These laws characterize the recurring connections (or relationships) between certain actions or their functions. Such fundamental connections are called economic laws. Like any laws of nature or society, they can be classified into three categories (types): causal laws, structural laws, and functional laws.

Causal Laws – These laws express relationships in which one event is always followed by another specific event. This sequence is typically observed over time. The first event is called the cause, while the second event is the effect.

*Structural Law*s – These laws describe connections where two or more events always occur together. The consistent, simultaneous occurrence of such events forms a certain lawful structure, which is the basis for the name "structural laws."

Functional Laws – These laws represent relationships between quantifiable events, which can be expressed in the form of mathematical functions.

In economics, not every connection or relationship always appears perfectly. Sometimes, certain relationships may not be visible in every situation. For example, a law might generally hold true, but in some cases, outside factors interfere, making it seem like the law isn't working as expected. These interfering factors are called auxiliary or random connections, and they can create obstacles for economic laws to function smoothly. Mathematicians describe this interference as external noise, meaning outside influences that slightly distort or change the actual relationships in economic processes.

Stochastic laws are laws that manifest in a mass repetition of similar phenomena.

Thus, the categories of statistical science are as follows:

1) a statistical population refers to a collection of units or phenomena that share a common characteristic while differing in quantity. This variation, known as variation, is a key feature of a statistical population. For example, the variation in the population size of cities within a country arises due to historical, geographical, economic, social, and other factors;
2) a statistical law refers to a necessary and intrinsic relationship between two phenomena. A statistical law expresses the general, essential, and significant connections present in phenomena;
3) statistical regularity refers to the sequence, repetition, and order observed in events and phenomena. It represents the objective quantitative patterns of mass events and phenomena. The emergence of statistical regularity is linked to the application of

the law of large numbers. The essence of this law is that the larger the set of events or phenomena, the more the variations caused by random factors cancel each other out, making the general regularities of the essential relationships between events and phenomena more clearly evident.

4) Statistical indicator expresses the quantity, value, or volume of events and phenomena occurring under specific conditions. An integrated system that describes the unity of events and is interconnected with each other is called a system of indicators;

The science of statistics systematically explains the processes of data collection, analysis, interpretation, and presentation based on its theoretical foundations. These foundations include the following key components:

The core theoretical aspect of statistics is working with data. Data are typically collected through research and experiments and must be accurately recorded to ensure reliability and validity. Statistics examines variables, which represent different characteristics of the research object. Variables are categorized into qualitative variables (nominal and ordinal) and quantitative variables (interval and ratio). Statistical theory develops methods for comparing data and determining relationships between variables, exploring both differences and connections within datasets.

The normal distribution, binomial distribution, and other mathematical probability distributions play a fundamental role in statistical theory. These distributions are used to predict and interpret the results of the study.

Probability theory is a crucial component of statistical analysis, used to assess the randomness in data and estimate statistical errors. Statistical inference involves concluding data and includes methods such as hypothesis testing, confidence intervals, and regression analysis.

The *main methods* of statistics include:

— *Descriptive statistics* are used to summarize and describe the general characteristics of data. This involves using statistical measures such as mean, median, mode, spread, and variance to present data.

— *Inferential statistics* allows us to draw conclusions about a population based on sample data. This category includes methods like hypothesis testing, confidence intervals, and regression analysis.

— *Parametric statistic*s are employed when data follow a specific, fixed distribution. This method involves evaluating the fit of the data to a normal distribution and its parameters.

— *Non-parametric statistic*s are used when there are no assumptions about the distribution of the data. This method is often applied to data based on rankings, medians, or other ordered measures.

— *Hypothesis testing* involves testing the validity of hypotheses using statistical methods. This process helps in making decisions between a null hypothesis and an alternative hypothesis.

— *Regression analysis* is used to identify and predict the relationships between variables. This process involves determining whether the relationship is linear or non-linear.

The above theoretical foundations and basic methods provide the necessary tools for the application of statistics in research and practical activities. Accurate and reliable conclusions can be drawn with the help of statistics, and these processes play an important role in various areas of society and the economy.

1.4. Organization and Functions of Statistical Activities in the Republic of Uzbekistan

After the Republic of Uzbekistan gained independence, special attention was given to organizing statistical activities. In order to adapt general issues of statistics and forecasting to the market economy, on August 6, 1992, by the Decree of the President of the Republic of Uzbekistan and the Resolution of the Cabinet of Ministers, the State Committee on Statistics was merged with the Committee on Development Forecasting. As a result, the State Committee on

Forecasting and Statistics under the Cabinet of Ministers of the Republic of Uzbekistan (Uzstatforecast) was established.

To ensure that statistical activities are conducted following legal and international standards, Oliy Majlis adopted the Law "On State Statistics" on September 2, 1993. Additionally, in line with Resolution No. 433 of the Cabinet of Ministers, the "State Program" for transitioning to the accounting and statistical system recognized in international practice was approved on August 24, 1994.

On November 17, 1994, the Cabinet of Ministers adopted Resolution No. 555, "On the Unified State Register of Enterprises and Organizations (USRE&O) for All Forms of Ownership and Economic Management." This resolution was introduced to develop and maintain the USRE&O in line with the principles of international registers while considering the needs of the republic.

In 1995, the State Committee for Forecasting and Statistics of Uzbekistan was reorganized into the Committee for Macroeconomics and Statistics.

On May 15, 1997, the Ministry of Macroeconomics and Statistics was established based on the Committee for Macroeconomics and Statistics.

On October 11, 1997, under Resolution No. 474 of the Cabinet of Ministers, the Ministry of Macroeconomics and Statistics was reorganized into the Ministry of State Forecasting and Statistics.

On December 12, 2002, Oliy Majlis adopted a new edition of the Law "On State Statistics" of the Republic of Uzbekistan.

On December 24, 2002, according to the Presidential Decree "On the Reorganization of the Ministry of Macroeconomics and Statistics of the Republic of Uzbekistan," the Ministry of Economy of Uzbekistan and the State Committee of Statistics of Uzbekistan were established. According to this decree, the following main tasks of statistical agencies were defined:

- Developing and implementing of a unified policy in the field of statistics;

- Establishing a scientifically grounded, nationally integrated statistical methodology and system of indicators that comply with international statistical standards and ensure the compilation of national accounts.
- Ensuring the timeliness, reliability, and objectivity of statistical and reporting indicators;
- Equipping statistical agencies with modern computers and technologies;
- Maintaining the Unified State Register of Enterprises and Organizations (USREO) and the economic-statistical classification system.

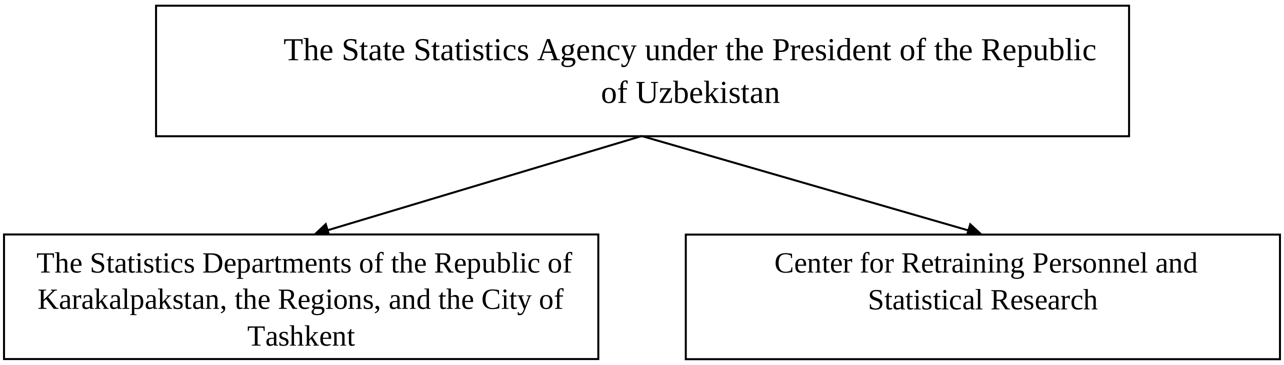


Figure 1.1. Structure of the State Statistics Agency under the President of the Republic of Uzbekistan.

In New Uzbekistan, statistical science and practice continue to receive significant attention. At the national level, decisions related to this area are being made and put into practice. Notably, the Presidential Decree PQ-3165, issued on July 31, 2017, outlined measures to improve the activities of the State Statistics Committee of the Republic of Uzbekistan. Additionally, the Presidential Decree PQ-4273, dated April 9, 2019, introduced further measures to ensure transparency and openness in public administration, as well as to strengthen the country’s statistical capacity.

Review questions

1. What is the origin of the term "statistics" and what did it originally mean?
2. What is understood by "statistics" in today's context?

3. Define the subject matter of the science of statistics.
4. Who are the founders of the science of statistics?
5. What are mass phenomena and processes, and what characteristics do they have?
6. What are the stages of statistical research?
7. What phases are involved in the analytical stage of statistics?
8. Define the subject matter of economic statistics.
9. What are the methods of statistics?
10. What are the functions of statistical bodies?

CHAPTER 2. METHODS OF COLLECTING AND PROCESSING STATISTICAL DATA

2.1. Concept and Significance of Statistical Observation, the Initial Stage of Statistical Research

The statistical analysis of social phenomena and processes is a complex and multifaceted task. Statistical observation represents the most critical stage of this research process. For the investigation to be effective, the data collected through observation must be unbiased and complete. Statistical observation must be conducted based on thorough scientific reasoning and well-prepared plans. This process involves directly gathering and verifying primary data.

Statistical observation refers to the systematic, planned, and scientifically organized collection of data on social phenomena and processes.

Statistical monitoring is the process of collecting, recording, and analyzing data. This process is the main part of the research, the success of which directly depends on the accuracy and precision of the observation. Observations are carried out using a variety of methods, including questionnaires, interviews, experiments, and the collection of official reports.

Statistical monitoring is important in the economy, society, health, education, and many other fields. Its importance is seen in the following aspects:

The data collected through statistical monitoring serves as the main source for research and analysis. This database plays an important role in further analysis and decision-making processes. In scientific research, the data collected through statistical observations increases the accuracy and reliability of the research. The results of observations are used to confirm or reject scientific hypotheses.

State bodies and business enterprises make strategic and tactical decisions based on the results of statistical observations. Statistics play a key role in these processes. In economic and social planning, future situations are predicted based on the results of statistical observations. This process helps ensure economic stability and efficiency.

The first stage of statistical research is the observation process. This stage includes several important steps:

The main purpose and tasks of the research are clearly defined. This process helps to choose the right research topic and define its scope.

Sources of data required for monitoring are identified. These sources can be official statistical reports, surveys, interviews or experiments.

In the process of creating a monitoring plan, the methods, time and place of data collection are determined. Also, the respondents or objects involved in the observation are selected. Data will be collected according to the monitoring plan. It is important to ensure the accuracy and reliability of the data throughout this process. Collected information is properly recorded and stored.

The collected data will be analyzed. During this process, statistical methods are used to identify relationships and trends between data.

Statistical observation and its initial stage play a decisive role in the success of the research. Scientific and practical conclusions can be drawn through proper data collection and analysis. This helps to optimize the decision-making process and increase the efficiency of planning.

Statistical observation must meet the following requirements:

1. Statistical observation must record statistical data in interconnection and as a whole;
2. All units within the observed population must be fully accounted for within the territorial boundary.
3. Data must be comprehensively collected over time.
4. The data obtained through statistical observation must reflect objective reality.
5. Statistical observation must be conducted based on a uniform program and methodology across all regions, institutions, organizations, and time periods.

6. The results of the observation should be used for operational management promptly.

2.2 Forms, Types, and Methods of Statistical Observation

Forms of statistical observation are divided into reporting and specially organized forms (Figure 2.1.). Reporting is a key method of continuously collecting information about the activities of enterprises and organizations in a prescribed manner and within a specified timeframe. According to the Law of the Republic of Uzbekistan "On State Statistics," all enterprises and organizations located within the country that have legal entity status—regardless of their ownership and organizational-legal form—are required to submit accurate and complete information about their activities to the relevant state statistical bodies following the established program, procedure, and deadlines. Reports consist of accounting reports and statistical reports.

Accounting reports reflect a company's production activities based on "debit" and "credit" calculations recorded in the accounts. In statistical reports, business activities are expressed in terms of the product name, unit of measurement, actual volume, planned volume, and the performance indicator of the plan (in percentages).

Specially organized statistical surveys refer to the process in which statistical organizations hire specific respondents to systematically or at predetermined times collect and record primary data about a studied phenomenon or process. Examples of such surveys include household budget statistics, consumer price statistics, informal trade statistics, and agricultural statistics.

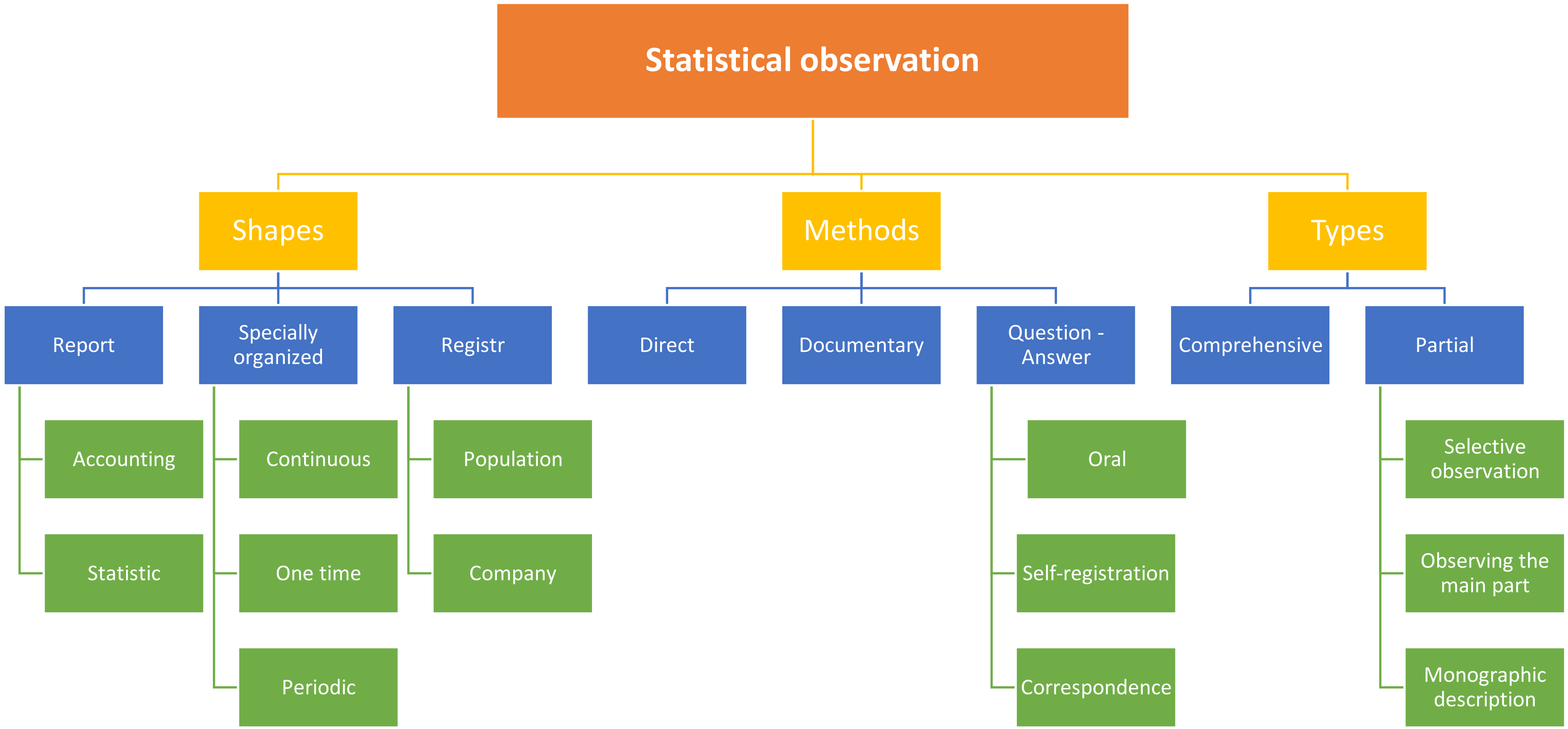


Figure 2.1. Statistical Observation

In continuous observation, an event is recorded immediately when it happens. For example, births, deaths, marriages, and divorces are registered right away in official civil registry offices or, in rural areas, by local administrative bodies such as village councils. Similarly, employees' attendance at work is recorded as soon as they check in, typically in payroll or attendance logs.

Periodic observation refers to the process of recording an event after a certain period has passed and conducting observations at regular intervals. Examples include conducting a population census every ten years and assessing students' academic performance every semester based on their rating scores.

A one-time observation is conducted when necessary to resolve a specific issue and can be repeated at different time intervals. Examples of such observations include the registration of housing stock, the inventory of perennial fruit trees, and the assessment of damages caused by natural disasters.

Statistical observations are classified based on the method of data collection into direct, documentary, and question-and-answer methods. In the direct method, statistical agency employees collect data by directly measuring and counting the results of a company's production activities. In the documentary method, statistical employees obtain data from relevant documents, such as compiling quarterly or annual reports from a company's monthly reports. The survey method is primarily used to gather public opinion on specific events or phenomena and can take the form of oral interviews, self-reporting, or correspondence surveys.

In the oral interview method, researchers verbally ask respondents survey questions and record their answers. In the self-reporting (self – self-registration) method, survey forms are distributed to respondents, who fill them out themselves. In the correspondence method, necessary information is collected in a structured manner by designated correspondents.

Statistical observations are also classified based on the scope of coverage into *comprehensive and partial observations*. In comprehensive observation, data is collected on every single unit of the studied population. In partial observation, data is collected only on a specific part of the population. Within partial observation, there are several approaches. Sampling observation gathers data from a representative subset that accurately reflects the characteristics of the whole population. The main part observation focuses on the largest and most significant portion of the population. In monographic description, a specific part of the population is studied in depth from multiple perspectives.

Statistical observation and its initial stage play a decisive role in the success of research. Proper data collection and analysis can lead to scientific and practical conclusions. This helps to optimize the decision-making process and increase the efficiency of planning.

Statistical monitoring is carried out in various forms. The main forms are:

Continuous Observation: In this form, observations are made continuously. As an example, we can cite economic statistics that are collected continuously.

Periodic Observation: In this form, observations are made between certain periods. For example, the population census is conducted every 10 years.

One-time Observation: In this form, observations are conducted only once. For example, surveys conducted on a specific event or event.

There are different types of statistical monitoring, each of which is carried out according to certain goals and tasks:

General Survey: This type of survey covers all respondents or objects. For example, the national census.

Sample Survey: This type of survey is conducted only through selected respondents or objects. For example, sample questionnaires on household income.

Monographic Survey: This type of survey is carried out for a detailed study of an event or object. An example is the analysis of the economic situation of an enterprise.

Indicator Survey (Indicator Survey): This type of survey is carried out on certain indicators. For example, tracking the unemployment rate or inflation indicators.

Statistical monitoring is carried out using various methods. The main methods are:

Questionnaire (Survey): The survey method is used to collect information from the respondents. In this method, a set of questions is prepared and presented to the respondents. The questionnaire can be written or oral.

Interview: the interview method is a method of collecting information through a direct conversation between the observer and the respondent. In this method, the observer asks questions verbally, and the respondent answers.

Experiment: the method of experiment is a method of gathering information by creating and changing certain conditions. This method is widely used in scientific research.

Official Documents: The method of collecting information through official documents is based on information provided by government bodies and organizations. For example, official statistical reports and annual reports.

Observation: the observation method is a method of gathering information through direct observation of events. In this method, the observer directly observes and records the event or process.

Forms, types, and methods of statistical observation are selected depending on the purpose, size, and other factors of the research. Properly selected methods and forms play an important role in ensuring the accuracy and reliability of research results.

2.3. Program-methodological and Organizational Aspects of Statistical Observation

A statistical observation plan is developed based on the nature of the research being conducted. This plan consists of program-methodological and organizational aspects. The purpose and objectives of the observation define the reasons for conducting the statistical observation and consider the quantitative representation of the observed characteristics.

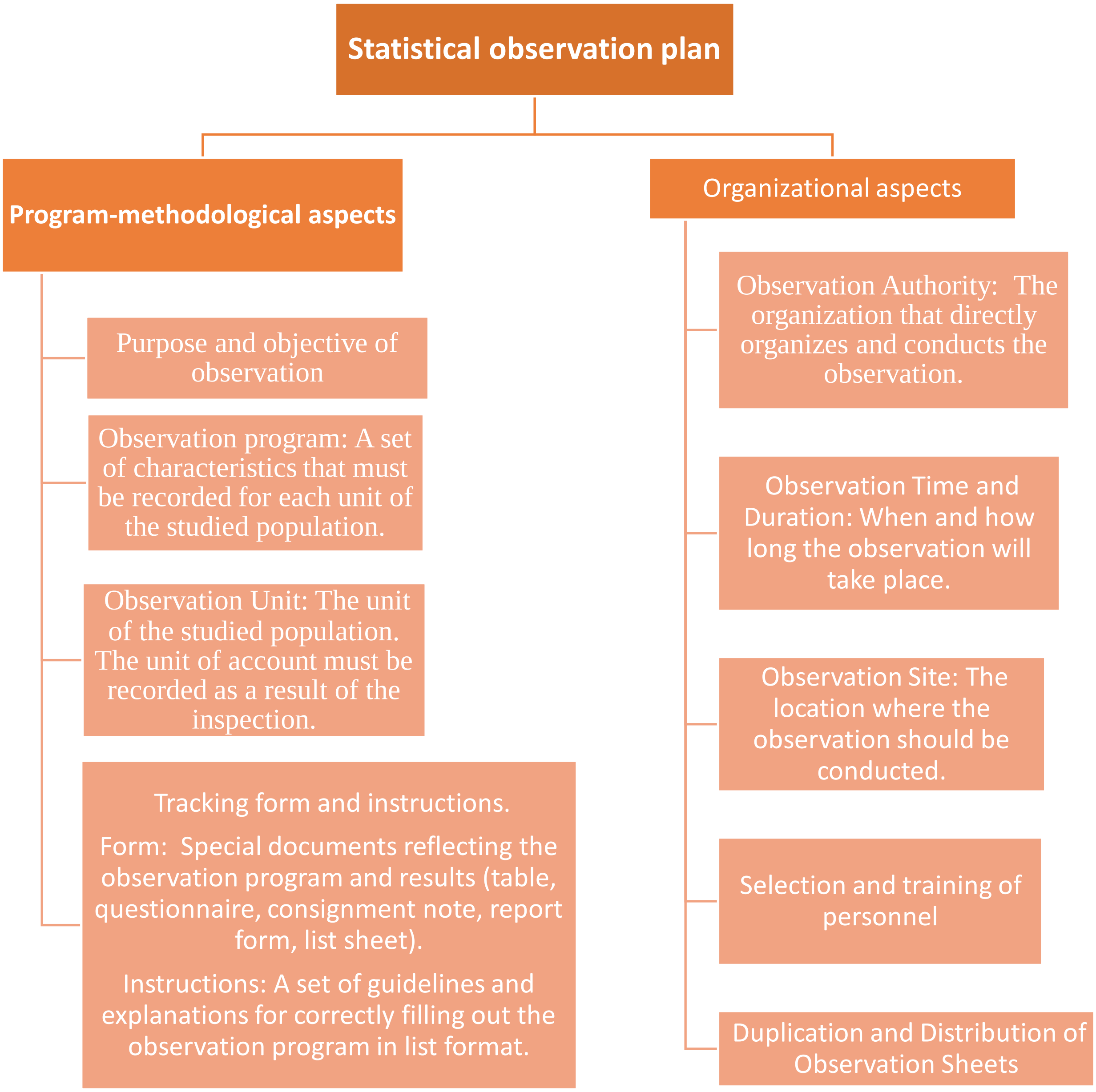


Figure 2.3. Statistical observation plan

The object of observation is a studied collection of events and phenomena, such as enterprises and organizations, farms, population, families, households, and others, in terms of their activities. Data on their key characteristics are collected, and generalized indicators that describe the dataset are calculated based on this information.

When determining the exact scale and scope of the observation object, it is necessary to specify its dimensions in terms of time, material structure, and measurement units.

Along with the object of observation, there is also its subject. The observation subject refers to legal or physical entities that record and collect data on the studied phenomena. In microeconomic statistics, for example, the observation subject includes individual enterprises, organizations, and institutions. As economic entities, they record events related to their activities and collect data. In macroeconomic statistics, statistical agencies serve as observation subjects. In special statistical surveys and censuses, these agencies record and collect primary data. In reporting, while recording primary data and presenting generalized information at the enterprise level is the task of individual economic entities, statistical organizations are responsible for receiving and generalizing this data within their territorial boundaries. In this case, economic entities, i.e., enterprises, firms, organizations, and institutions are called reporting units.

The observation unit refers to the smallest independent and indivisible structural element of the observed object. During the observation process, information about its key characteristics is collected. For example, when examining industrial activity, each enterprise and firm becomes an observation unit, and when studying agriculture, each cooperative farm, peasant farm, and farmer's household serves as an observation unit.

In observing certain issues, along with the observation unit, a unit of account is also defined. For example, in livestock census, detailed information is collected about each type of animal; in equipment inventory, about each machine, machine tool, and other equipment; and in perennial tree census, about each type of tree. In this case, while the farm or enterprise is the observation unit, each item being studied is the unit of account.

The guideline is developed to ensure uniform organization, which refers to creating a standardized, consistent approach to collecting, recording, and understanding data, and to facilitate the understanding of the indicators specified in the observation program. It outlines the purpose and objectives of the observation, its object and unit, the source and subject from which the data will be obtained, the duration of the observation, the procedures for documenting records, the reporting period to which the data pertains, the calculation methods for indicators, and other issues directly related to the observed object. Guidelines are typically developed with the participation of multiple experts.

Guidelines are prepared separately for each statistical form. The document in which observation data is recorded is called a statistical form. It is known by different names, such as a reporting form, registration sheet, questionnaire (survey), timesheet, invoice, waybill, and so on.

A statistical form is of two types:

1) Separate form;
2) List-type form.

A separate form is filled out separately for each respondent or object of observation. This form method is used in the following cases:

Used to get detailed information about each person or object. For example, in the population census, the personal information of each person is recorded separately.

When conducting special questionnaires or interviews, a separate form is filled out for each respondent.

A separate form is used in cases where the data must be detailed and accurate for analysis. The importance of a separate form can be explained as follows:

- Provides detailed information about each respondent.
- Helps to record data completely and accurately.
- Provides a high degree of reliability in the collection of personal or special data.

In the form of a list, the form is filled out in common for several respondents or objects of observation. This form method is used in the following cases:

Used to get general information about more than one person or object. For example, recording the names and grades of students in a class in the form of a list.

The form of the list is convenient for conducting short-term and quick observations, collecting and recording data quickly.

A list form can be effective when collecting and recording large amounts of data.

The importance of the list form can be explained as follows:

- Speeds up the data collection process.
- Allows you to store information on large groups or many objects in one place.
- Helps to record data simply and systematically.

Separate forms and list-type forms are important tools used in collecting the data required for statistical observations. Both types have their own advantages and fields of application, which are chosen depending on the purpose and content of the research. While a separate form is convenient for collecting precise and detailed information, a list form is useful for recording large amounts of information quickly and efficiently.

The place of observation refers to the location or, in other words, the place where the studied event occurs. In reports, equipment inventories, and other observations of economic resources in various forms and methods, the location where an enterprise, organization, or institution is situated is called the place of observation.

The time of observation is the period during which the event is studied, meaning the time interval from the start of the examination until the data collection is completed. In this case, it includes both the time to which the data pertains and the time when the data is collected. In reporting, this time interval is called the reporting period. The reporting period is measured from 00:00 on the first day to 24:00 on the last day. It is not possible to define such a period when determining the state or quantity of events because significant changes may occur in the state of the event, and thus in the number of occurrences, during this period.

Therefore, in such cases, it is necessary to use the method applied in photographic capture. As is known, a photograph captures the appearance of the subject (for example, a person) at the moment the camera shutter opens. Similarly, when organizing statistical observations on the state of events, it is essential to determine the exact moment of recording. This allows for identifying the number and structure of events present at a specific moment.

There are three methods for verifying statistical observation data:

1) External control – ensures the proper documentation of records, including compliance with guideline requirements and completeness of data;
2) Logical verification, where statistical data is checked for content (logically);
3) Arithmetic verification, where interrelated data is checked numerically.

Errors made during the data collection process are called observation errors. Even in single-subject observations conducted using highly accurate measuring instruments, errors can still occur.

Observation errors are divided into two types:

1) Recording errors;
2) Representative (authoritative) errors.

Recording errors occur as a result of misunderstanding and incorrectly recording the content of data. Recording errors are divided into random errors and systematic errors. Random errors include counting errors and omission errors in writing. Systematic errors always occur in a specific direction. Therefore, they strongly affect general indicators. Such errors are dangerous. Systematic errors can be made knowingly or unknowingly. Systematic errors, in turn, are of two types: measurement errors and personal errors, i.e., observer errors.

Observer errors can be intentional (biased) or unintentional (unbiased). Intentional errors have a specific purpose. Unintentional errors arise from the vague presentation of the observation method, especially the program and guidelines, and observers' partial deviation from established procedures to simplify work. At the same time, partial observations may contain specific errors, known as representativeness errors. A representativeness error refers to the difference between the summary indicators derived from sample observation data and the actual corresponding indicators of the full population being studied.

2.4. Understanding Data Summarization in Statistics

In the complex world of statistical research, transforming raw data into meaningful insights requires a sophisticated process known as summarization. Beyond simple data collection, summarization is a critical methodology that helps researchers understand patterns, trends, and underlying social phenomena. In statistics, summarization refers to the process of organizing, systematizing, and condensing collected data.

Statistical summarization is the process of combining data collected from separate units (observations) to obtain a general view of the dataset. It is an essential part of statistical analysis that allows one to identify common characteristics in the data, describe them, and analyze them. This process is particularly valuable when working with quantitative data or large datasets that need to be efficiently managed and interpreted.

There are several types of summarization in statistics, each used depending on the type of data and the purpose of the study. The main types of summarization include:

Simple Summarization: In simple summarization, the values of all units are combined to determine a grand total or another general measure (such as an average value). This approach is used to determine total volume, sums, or average values. For example, calculating the total grades or the average grade of students in a class.

Grouped Summarization: In grouped summarization, units are grouped by a specific attribute or category and summarized within each group. This method is used for analyzing and comparing data across different groups. For example, summarizing population or income by region.

Ranked Summarization: In ranked summarization, units are ranked according to a specific criterion and then summarized. It is used to analyze units based on their level of importance or other evaluative factors. For instance, ranking companies by revenue and then calculating the total revenue.

Cumulative Summarization: In cumulative summarization, the values of the units are added in a progressive manner, meaning the value of each subsequent unit is added to the sum of the previous ones. It is often used to analyze trends over time. For example, determining the cumulative change in annual earnings.

Time-Based Summarization: In time-based summarization, units are grouped by time period and summarized accordingly. This method is used to analyze time-dependent changes, such as monthly sales volumes or daily temperature variations.

Summarization performs several important functions in statistical analysis:

It converts raw data into a manageable and analyzable format. Through summarization, the general characteristics of data can be revealed and interpreted. This enables informed scientific, economic, and managerial decision-making based on the analytical results.

Summarization is a core aspect of statistical analysis. Through its various methods, it becomes possible to draw broader conclusions from the data and identify meaningful patterns. Based on computational methodology, summarization is classified into simple and complex summarization

- Simple summarization involves deriving overall summaries for the entire dataset.
- Complex summarization includes grouping observation units, deriving totals for each group and for the entire dataset, and displaying the results in structured tables.

From an organizational standpoint, summarization is further divided into centralized and decentralized summarization. Centralized summarization occurs when data is collected and processed at a single location. Decentralized summarization is conducted step by step across different levels or administrative units.

Stages of Summarization in Statistics:

1) A program and plan for data processing are developed. The program specifies the groups to be distinguished based on certain characteristics, the system of indicators to be calculated for each group, and their territorial boundaries. The plan defines who will conduct the summarization, the procedure to be followed, and the format of the tables that will present the results.
2) Data units are categorized into groups based on the relevant characteristics under study.
3) Summaries are calculated for each group and for the dataset as a whole.
4) The results are presented in tables and graphs for visualization and interpretation.

2.5. Grouping Method in Statistics: Concept, Significance, and Types

In statistics, grouping refers to the classification of mass social phenomena into homogeneous groups based on significant characteristics. For example, industrial enterprises can be grouped according to their dividend payment levels, or employees can be categorized based on work experience (e.g., those with less than three years of experience as inexperienced and those with more than three years as experienced). The grouping method facilitates the summarization and visualization of data concisely and explicitly. Additionally, it provides a foundation for data processing and analysis. To ensure methodological accuracy, selecting grouping criteria, determining the number of groups, and establishing group boundaries must be conducted based on scientific principles.

The grouping method is a method of dividing data into groups according to certain characteristics or characteristics in the process of statistical analysis. This method provides an opportunity to organize data, identify common features in them, and analyze them. Grouping is effectively used when the volume of data is large or when their analysis is complex.

The importance of the grouping method is manifested in several aspects:

- ✓ By grouping data, it becomes easier to organize and manage it. This simplifies and speeds up the analysis process.
- ✓ By grouping data, similarities and differences can be identified. This allows you to see and evaluate the features that are important in the analysis process.
- ✓ Grouping makes it easier to identify trends and relationships between data. This helps to better understand the research results.
- ✓ The ability to make statistical conclusions and make scientific, economic, or management decisions on the basis of grouped data increases.

The grouping method has several types, each of which has its characteristics and fields of application:

Simple Grouping: In this type, data is grouped according to simple criteria. For example, grouping by age: 0-18 years, 19-35 years, 36-60 years and over 60 years.

Interval Grouping: In this type, data is grouped by certain intervals. For example, grouping income by 1000 dollar interval: 0-1000, 1000-2000, 2000-3000, etc.

Nominal Grouping: In this type, data are grouped by specific names or categories. For example, grouping by gender: men and women.

Ordinal Grouping: In this type, data is grouped according to a specific order or level. For example, grouping by level of education: primary, secondary, higher.

Geographic Grouping: In this type, data is grouped by geographic regions. For example, grouping by provinces or regions.

Time-based Grouping: In this type, data is grouped by time interval. For example, grouping by year, month, day, or hour.

The grouping method is implemented through the following steps:

- ✓ Determining the criteria by which data should be grouped.
- ✓ Define ranges or categories for each group based on defined criteria.
- ✓ Sorting and sorting data into defined groups.
- ✓ Analyze grouped data and identify trends, relationships, and conclusions.

The grouping method is an important part of statistical analysis, it allows to organize data, identify common features, and analyze it. This method is used in various fields and helps to increase the accuracy and reliability of research results.

Statistics follows established scientific principles to ensure that data grouping is systematic and meaningful. The key principles include:

1. Grouping criterion;
2. Number of groups;
3. Interval length.

These principles help statisticians organize data effectively, making it easier to identify patterns, compare results, and draw accurate conclusions.

1) Types of Grouping Criteria

Grouping criteria are classified into the following categories based on their *nature*:

— Attributive criteria: These are qualitative characteristics that distinguish one group from another. For example, when grouping a population by nationality or university specializations, the number of groups will correspond to the number of specializations.
— Alternative (binary) criteria: These characteristics divide a dataset into two mutually exclusive groups which are usually opposite one to another. Examples include educated vs. uneducated individuals, students who attended classes vs. those who did not, and individuals who passed the final assessment vs. those who failed.
— Quantitative criteria: These characteristics are expressed numerically. For instance, students who score below 55 points receive an unsatisfactory grade, while those scoring between 55 and 71 receive a satisfactory grade, and so on.

Grouping criteria can also be classified based on their *purpose and function*:

— Non-essential criteria: These are external attributes that describe events or phenomena externally but are not used for grouping. Examples include a company's name, address, phone and fax numbers, and bank account details. While these attributes are important, they do not serve as a basis for grouping.

— Essential criteria: These represent the internal properties of events and phenomena. Examples include the number of employees in an industrial enterprise, their composition, and the value of fixed assets.

From the perspective of *significance*, grouping criteria are further divided into two types:

— Factor (causal) criteria: These are characteristics that influence the outcome variable.

— Outcome (resultant) criteria: These are characteristics that change due to the influence of factor criteria.

2) Determining the Number of Groups

Selecting an appropriate number of groups is crucial for effective grouping. This is best achieved by applying the multi-variant rule.

Some textbooks recommend determining the number of groups using G.A. Sturges' formula. However, this formula is specifically designed for constructing frequency distributions and should only be used when the dataset contains more than 100 observations ($N > 100$) and when grouping is based on a single criterion. Additionally, the values of the grouping criterion must exhibit uniform variability across individual units in the dataset.

3) Determining Group Intervals

The boundaries of the group intervals can be defined in two ways: first, by assigning an equal interval size to all groups. Example: Age groups (0-10, 11-20, 21-30, etc.). The second is by expressing the interval in an unequal form, where the sizes either increase or decrease. Example: Income brackets where lower-income groups have smaller intervals, and higher-income groups have larger intervals. In the first case, the group interval is determined using the following formula:

$$i = \frac{X_{\max} - X_{\min}}{K}$$

where: X_{max} - s the maximum value in the dataset;
X_{min} - is the minimum value,
K - is the number of groups to be formed.

Grouping is used to identify statistical relationships and patterns, analyze the structure of the population, and describe socio-economic categories. There are three main types of grouping based on purpose and objectives:

1) typological;
2) analytical;
3) structural grouping.

Typological grouping refers to the classification of the studied phenomena into socio-economic types. A socio-economic type is understood as a set of economic entities or individuals that share common functions and roles in the society and economy.

Analytical grouping is a method used to describe the relationships between two or more variables. This approach classifies variables into two categories: factor variables and resulting variables. A resulting variable is one that changes under the influence of other variables, while a factor variable is an independent variable that affects and determines changes in the resulting variable. Analytical grouping is typically structured based on factor variables, with each group characterized by its corresponding resulting variables.

Grouping can be based on one or multiple variables. A classification based on a single variable is referred to as simple or unidimensional grouping, while a classification involving two or more variables is termed complex or multidimensional grouping. Although analytical grouping aims to reflect the relationship between a factor and a resulting variable, this relationship does not occur in isolation. The factor variable itself is influenced by other external variables, leading to additional variability in the resulting variable due to external "noise" factors.

Structural grouping typically defines the composition of a set based on a specific characteristic. The process begins by constructing a distribution series based on a chosen attribute. Then, collective indicators that describe the structure of each group—such as the total size of the grouping attribute within the groups and the quantities of related characteristics—are calculated. These values are then used to determine the contribution of individual groups to the overall set.

Structural groupings and distribution series share similarities, but they differ in function and structure. The primary goal of structural grouping is to study the composition of a set. For this purpose, each structure must be characterized using multiple indicators, revealing its various aspects. In contrast, distribution series are used to identify the structural characteristics and patterns within a statistical set.

Structural groupings facilitate the statistical analysis and quantitative assessment of structural changes over time and across different regions. Two methods are used for this purpose: the first evaluates internal differences within a set, while the second assesses differences in structure between sets.

In the first case, the *internal structural differences* of each set are evaluated using their average absolute value, average quadratic value, and their coefficients. If the indicator being studied for certain groups is m_i, then their weight (share, part) $d_i = \frac{m_i}{\sum m_i}$. In that case: the average absolute weight difference $\Delta d = \frac{\sum |d_i - \bar{d}|}{K}$ (the sign is not considered). Its coefficient (K) is calculated $K\Delta = \frac{\Delta d}{\bar{d}}$; The average quadratic difference of weights $\sigma_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{K}}$, its coefficient $K_\sigma = \frac{\sigma_d}{\bar{d}}$, where $\bar{d} = \frac{\sum d_i}{K}$. These indicators are calculated for comparative analysis of the structure of the dataset in terms of dynamics or spatial aspects.

In the second case, the arithmetic and quadratic mean differences, as well as the coefficients, are calculated based on the differences between the structure weights (d_i) that express the composition of two comparable sets.

$$\theta = \frac{\sum |d_{i1} - d_{i0}|}{K}$$

$$S = \sqrt{\frac{\sum (d_{i1} - d_{i0})^2}{n}}$$

$$K_\theta = \frac{\theta}{\overline{d_1}}$$

$$K_S = \frac{S}{\overline{d_{i1}}}$$

where: d_{i1} - the proportion (share) of the collection groups (structure) in the current (or comparable) period;

d_{i0} – the proportion (share) of the collection groups in the past (or comparative) period;

θ- the average absolute value of the differences in the structure of the two sets (the sign is not considered);

S - mean square value of the differences in the structure of two sets;

K_θ - the level of average absolute differences;

K_S - level of average squared differences

Statistical grouping methods described above are implemented based on the raw data attributes of individual observational units. This initial classification process is termed primary grouping. However, alongside primary grouping, there exists a secondary grouping methodology that involves processing and reorganizing the results of primary groupings. This method is typically used when the results of the primary grouping are not suitable for the intended purpose or when it is necessary to adapt them for comparison. Secondary grouping can be executed through two distinct approaches:

1. Consolidation and optimization of primary group intervals to better align with research objectives;
2. Adjustment of relative proportions of specific groups relative to the overall total.

Review questions:

1. What are the programmatic and methodological considerations in the development of a statistical observation plan?

2. What are the organizational aspects involved in the preparation of a statistical observation plan?
3. Based on the timing of event recording, into what categories can statistical observations be classified, and what are the key components of each category?
4. What is meant by partial observation? How does it relate to monographic descriptions, basic mass observations, questionnaire-based observations, and sampling observations?
5. What is partial observation? What about monographic description, basic mass observation, observation via questionnaire, and sampling observation?
6. What kinds of errors can occur in statistical observation?
7. What is statistical grouping, and why is it significant in statistical analysis?
8. What is typological grouping, and how does it differ from analytical and structural grouping?
9. What is a grouping criterion, and what factors should be considered when selecting it?
10. How is the number of groups determined in statistical grouping, and what considerations are involved in defining group intervals? What methods of secondary grouping do you know?

CHAPTER 3. REPRESENTATION OF STATISTICAL DATA IN TABLES AND GRAPHS

3.1. Statistical Tables: Concept and Types

The results of summarization and grouping are presented in the form of tables. In statistics, tables consist of the intersection of horizontal and vertical lines and serve as an organized and visual representation of data.

The Russian scholar N.K. Kirilov was the first to use statistical tables in 1727. Every table has a header and a body. The header is placed in the horizontal rows, while the body is placed in the vertical columns. The part of the table that describes the subject of discussion is called the header. The indicators that describe the header are referred to as the body of the table.

According to the characteristics of the header, statistical tables can be classified into the following three types:

- Simple tables, where the header only lists years, months, events, or objects.
- Grouping tables, where the header is divided into groups based on a particular characteristic.
- If the groups in a grouped table are further subdivided, such tables are referred to as combinational tables.

When constructing a statistical table, the following principles must be adhered to:

1. The table should be of an appropriate size, avoiding excessive complexity.
2. The table's title, stub (row headings), and column headings must be clearly and concisely formulated in a comprehensible manner.
3. The units of measurement for all indicators, as well as the geographical and temporal scope of the data, should be explicitly stated.

4. The indicators within the table should be mutually verifiable to ensure coherence and reliability.
5. All numerical values must maintain a consistent level of precision throughout the table.
6. The table should include all relevant groups, subgroups, and overall totals where applicable.
7. If the header and body need to be divided into groups, the groups in the header should be labeled with alphabetic characters, and those in the body should be marked with numbers.

Statistical tables are one of the important tools used to organize, present, and analyze data. They simplify the process of analysis and drawing conclusions by presenting data in a visual form. Statistical tables perform the following main functions:

- Allows you to store and present data in an orderly and systematic manner.
- It makes it possible to compare different indicators with each other and determine their interrelationship.
- It helps to analyze the data in depth and draw informed conclusions.

Statistical tables are divided into several types, each of which is used depending on the type of data and the purpose of analysis. The main types of statistical tables are:

Simple Table: A simple table is one-dimensional and shows data on only one indicator or parameter. Example: A table showing the population by region.

Complex Table: A complex table shows information on several indicators or parameters. Example: A table showing population, gross domestic product (GDP), and average income by region.

Two-dimensional table: In a two-dimensional table, data is displayed on two dimensions (indicators). Example: A table showing the population by sex and age.

Multi-dimensional table: A multi-dimensional table displays data on several dimensions. Example: A table showing the population by sex, age, and level of education.

Dynamic Table: In a dynamic table, data shows changes over time. Example: Population change over the years.

Matrix Table: In a matrix table, data is displayed using cross-tabulation method. Example: Information on educational attainment and employment status.

Pivot table: In a pivot table, data can be rotated and displayed in different dimensions. Example: It is possible to quickly and easily analyze various indicators through pivot tables created in Excel.

The following rules should be observed when compiling statistical tables:

- The information must be clear and correct. It is necessary to check the data several times to avoid errors.
- The title of each table should be clear and fully express the information. Indicators must be clearly defined and units of measurement indicated.
- Information should be arranged in an orderly and systematic manner. This process ensures that the data is understandable and easily analyzed.

The design of the table should be aesthetically pleasing and easily comprehensible. The use of colors, lines, and other visual elements serves to enhance the clarity and facilitate the analytical process.

3.2. An Overview of Graphs and Their Types

Effectively presenting statistical data in a clear, engaging, and concise manner is essential for ensuring broad public understanding. The graphical method serves this purpose and is widely applied in practice.

In statistics, graphs refer to the representation of information about social life through various geometric shapes, lines, pictorial symbols, and conditional markings on geographic maps. These visual tools not only capture attention quickly but also enhance memory retention, providing a more comprehensive and in-depth understanding of the data.

Therefore, graphs serve as a vital tool for effectively communicating and disseminating information about processes across all areas of economic, cultural, and social development.

The use of graphical representation in statistics facilitates the simplification, clarification, summarization, and consolidation of data.

This method is particularly useful for comparing different phenomena, analyzing their dynamics and interrelationships, monitoring the implementation of plans and government orders, examining the structure of complex datasets, and visualizing the distribution density of specific events.

Graphs can be classified into various types based on their structure and the processes they represent.

A *diagram* refers to the depiction of statistical data using geometric shapes. When data are represented through conditional symbols on geographic maps, such graphs are called *cartograms*. *Cartodiagrams*, in turn, are a combination of diagrams and cartograms. In this case, geographic maps are divided into specific regions based on the spatial distribution of phenomena, and corresponding diagrams are placed within these regions to illustrate the data. Cartograms and cartodiagrams are particularly useful for visualizing the spatial distribution of various events or phenomena.

The main type of graphs is charts. Charts are primarily constructed using flat geometric shapes and lines.

To represent statistical data in charts, a scale is first determined, based on which a graduated axis is created.

A scale is a conditional standard that converts numerical indicators into proportional visual representations on a graph. It defines how a specific quantity of the studied phenomenon corresponds to one unit of length on the graph. For example, when creating a chart depicting cotton production in Uzbekistan, if 1 million tons of cotton is represented as 1 centimeter, this measurement defines the scale of the graph.

A graduated axis is a line where specific points correspond to certain quantities of the represented phenomenon, meaning each point can be read as a numerical value.

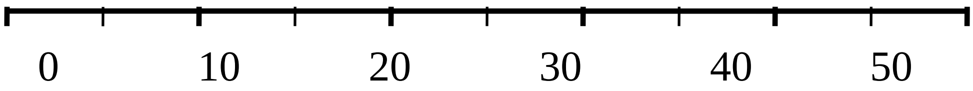


Figure 3.1. A graduated axis (uniform scale).

The graduated axis consists of the following elements:

1) A baseline, which serves as the foundation of the scale.
2) Marking points, positioned along the baseline in a specific order.
3) Numerical values, assigned to each marking point to indicate corresponding quantities.

Each point on the baseline of the scale corresponds to a specific quantity of the studied phenomenon, meaning every quantity is represented by a distinct point on the line. The baseline can be either a straight line or a curved line, classifying scales into linear and curved types. A standard millimeter ruler serves as an example of a linear scale, while a clock face (which displays time through the use of a flat dial with reference marks) represents a curved scale. Many measuring instruments feature arc-shaped curved scales.

When applying graphs in statistics, evenly graduated scales (uniform scales) are typically used. When the quantities being depicted differ widely in range, variable graduated scales (non–uniform scale) are used in constructing graphs. These include logarithmic and semi-logarithmic scales, where segment lengths along the baseline are proportional to the logarithms of the represented values. In simpler words, in a semi-logarithmic graph, population growth over decades may be plotted using a non-uniform scale, where the distance between 1 million, 10 million, and 100 million is the same, reflecting their logarithmic progression rather than absolute differences.

Types of Graphs Based on Visualization and Construction Methods

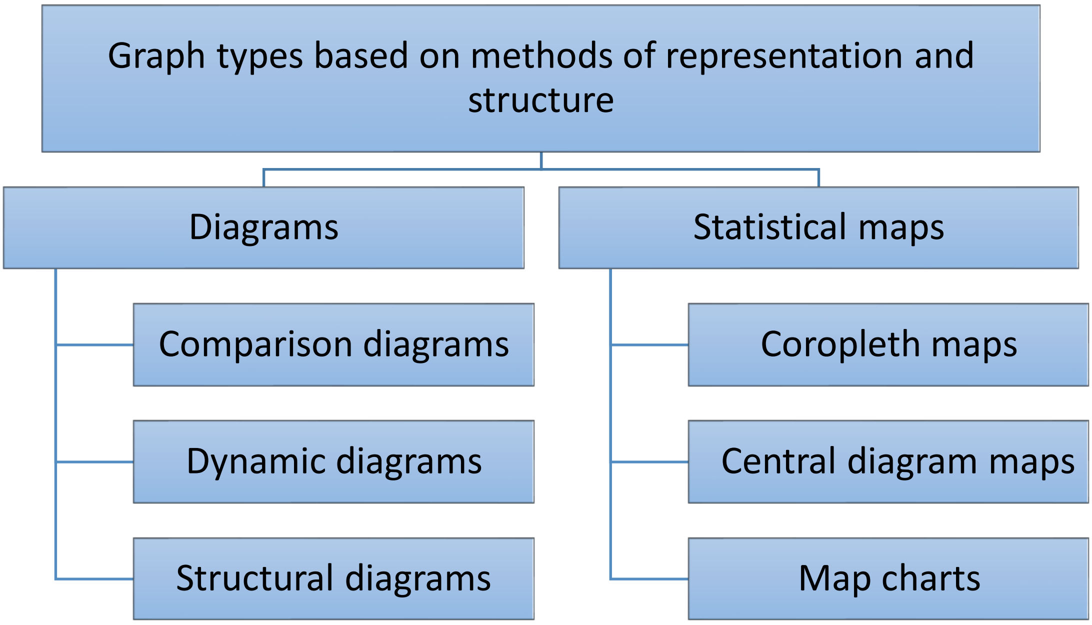


Figure 3.2. Types of Graphs

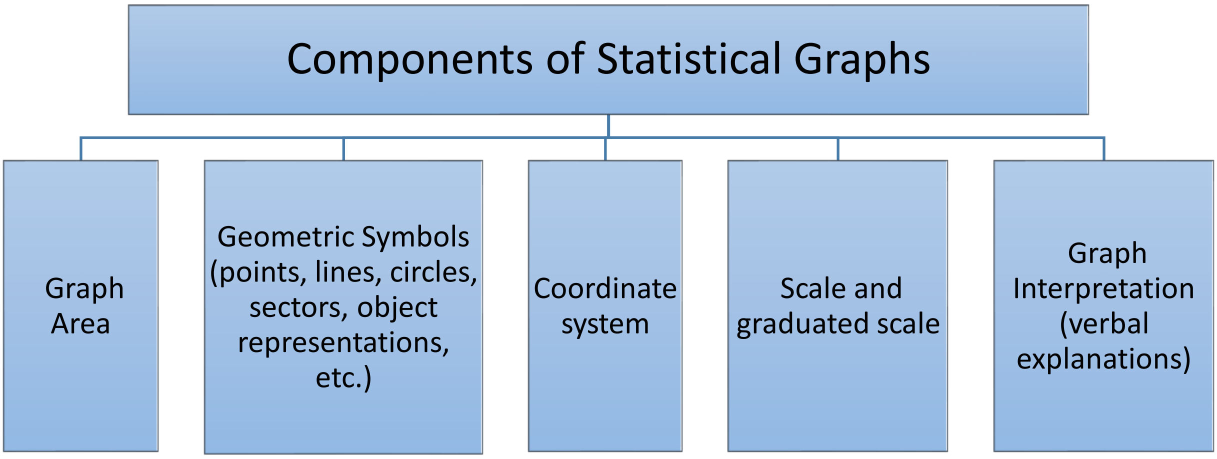


Figure 3.3. Components of Statistical Graphs

There are numerous types of line and 2- dimensional diagrams, with the most significant ones being the following:

Line graphs (aka line charts) are the most common type and are used to illustrate dynamic series, relationships between events, distribution series, and performance indicators of contracts (plans). These graphs are constructed based on a coordinate plane or numerical grid. On the vertical axis (ordinate axis), the indicators of a dynamic series or the values of a resultant feature are marked with dots according to a chosen scale. On the horizontal axis (abscissa axis), the periods (times) of the series or the values of the factor feature are plotted using dots. Then, perpendicular lines are drawn from each point on the axes toward the opposite axis. The intersection points (ordinate peaks) are connected, forming a broken line. This broken line represents the dynamic series or the interrelation of studied events or phenomena.

The graph created for an interval (range-based) variation series is called a distribution *histogram*. Line graphs are often constructed in the form of broken lines. However, this form is not always suitable for studying complex phenomena and processes. In some cases, a spiral-shaped line diagram is more appropriate.

Spiral-shaped diagrams are used when it is necessary to represent both the overall change of the studied set and the changes within its components or when showing periodic variations (e.g., year by year) along with intra-period changes (e.g., month by month).

Bar charts. Bar charts are the simplest way to visually represent statistical data. They are used to compare various indicators, illustrate the structure of complex phenomena, and track their changes over time.

Pie charts. A pie chart consists of a circle divided into segments, with the segment sizes corresponding to the represented numerical values. These charts help visualize the structure of complex phenomena by breaking them down into their components. They are constructed in the following order. A circle is drawn and considered as the total sum of the studied set (100%). Then, the circle is divided into sectors proportional to the number of components. To do this, the central angles of the sectors are determined and marked on the circle using a protractor. If whole and fractional numbers are expressed in absolute terms, the scale is determined by dividing 360° by the total whole number, and then it is successively multiplied by the fractional numbers. If the structure of the phenomenon is expressed in relative proportions (where the total set is taken as 100% and its individual parts are calculated as percentages of the total), the scale is equal to 3.60 (360° ÷ 100%), and each part's share (calculated in percentages) is multiplied by this value.

Square and Circular Diagrams. These diagrams belong to the category of 2 - dimensional diagrams and consist of a series of squares or circles, each representing the quantity of the depicted phenomenon by its area. Such diagrams are created to illustrate dynamic series and compare quantities related to a specific time (period). To construct square-shaped diagrams, the square roots of the represented quantities are taken, and squares are drawn with sides proportional to the obtained values. Circular diagrams are constructed similarly; however, in this case, circles are drawn with radii proportional to the square roots of the represented quantities.

Pictorial (Illustrative) charts. To enhance the impact of graphics, graphs are created by drawing illustrations (pictures) of the studied phenomena. A conditional representation of the phenomenon's image (figure) is taken as the unit of measurement. For example, to depict the dynamics of freight turnover on a railway, images of freight wagons are drawn.

Statistical Maps (A thematic map). Statistical decisions involving geographical structures are represented as statistical maps. These maps help to clarify the characteristics and patterns present in the territorial distribution of the studied phenomena. They are used to study the placement and development of productive forces by economic regions, as well as to analyze the use of economic resources in a regional context, determining quantitative proportions and patterns in this regard.

Statistical maps differ from economic-geographical maps. Economic-geographical maps qualitatively describe and estimate the distribution of productive forces, providing a highly detailed and precise depiction of their geographical location. In contrast, statistical maps offer a clearer spatial quantitative representation of statistical data related to these issues in a regional context. So, economic geographic maps describe the location of economic resources qualitatively, while statistical maps represent it quantitatively. In this sense, statistical maps can be defined as a visual representation of statistical tables structured in a geographical context. However, statistical maps are more informative than such tables because the visual language is better suited for marking quantities and proportions in space than tabular data.

Therefore, statistical maps enhance geographically structured tables, making the data more comprehensible and visually interpretable.

A statistical map is a contour-based geographical map that contains only the outlines of administrative divisions (in some cases, also water arteries), within which statistical data is represented using symbolic indicators.

Statistical maps are classified into three types based on their intended purpose and objectives: Choropleth maps, Proportional symbol maps, and Combination maps. If it is necessary to depict the distribution of a certain phenomenon by regions, then a Choropleth map (a cartogram) is used. However, if the goal is to provide a more detailed description of specific regions based on given indicators or to illustrate interregional connections, then a proportional symbol map is created.

A choropleth map is a type of statistical map in which the territorial distribution of the studied phenomenon is represented using symbolic indicators (such as hatching, color, dots, etc.) corresponding to specific value ranges. These symbols are placed within each administrative division according to the indicator's value. Choropleth maps are divided into shaded (color) and dot maps.

A shaded choropleth map represents a distribution series that describes the intensity of a phenomenon across regions (e.g., population density). In this case, the total sum of the represented quantities does not have economic significance.

A dotted choropleth map is used to show the spread of a phenomenon in specific areas, meaning its distribution by districts. Here, the total sum of the represented values (numbers) carries economic significance.

Combined maps refer to statistical maps with complete tables and statistical data. For example, statistical-geographical data, such as population dynamics by region. In this case, the data is visualized on contour maps, showing information for specific territorial divisions (such as provinces or regions).

Review questions:

1. What do you understand by statistical tables? Do they have advantages over verbal descriptions?
2. What elements make up a statistical table? What is meant by its "header" and "body"? What is meant by the "layout" of the table?
3. What types of statistical tables do you know?
4. What are simple tables? And group tables? What about combination tables?
5. What is meant by statistical graphs?
6. List the components of a graph and explain their functions.
7. What types of graphs exist?
8. How are changes in data over time represented in diagrams? What is their structure, and how do these diagrams relate to each other?
9. Explain the process of creating pic charts with an example. What processes do they represent?
10. When are circle and square-shaped diagrams used? Explain the process of creating them with an example.

CHAPTER 4. STATISTICAL INDICATORS

4.1. The Concept and Types of Statistical Indicators

Statistical indicators quantitatively express the qualitative aspects of mass social phenomena. These indicators characterize the number, size, ratio, and density of the object under study. In the context of the modern market economy, indicators that assess the implementation of business plans and the quality of services provided to both legal and individuals have become of significant importance.

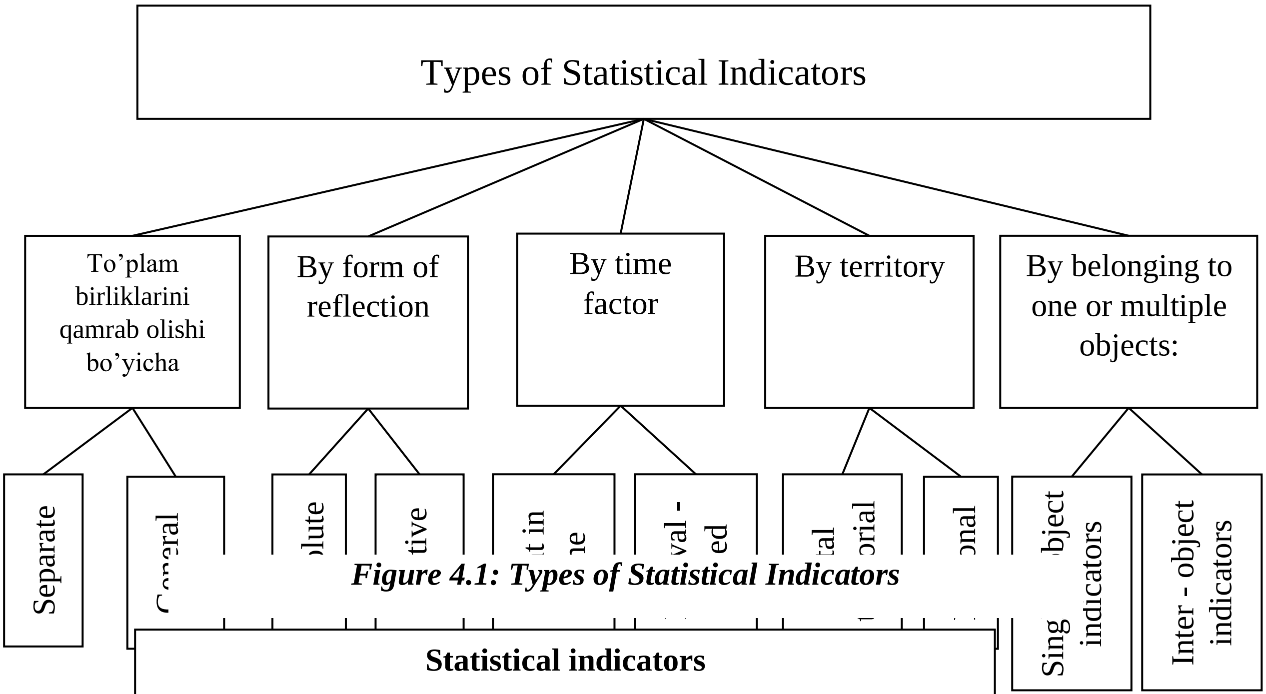


Figure 4.1: Types of Statistical Indicators

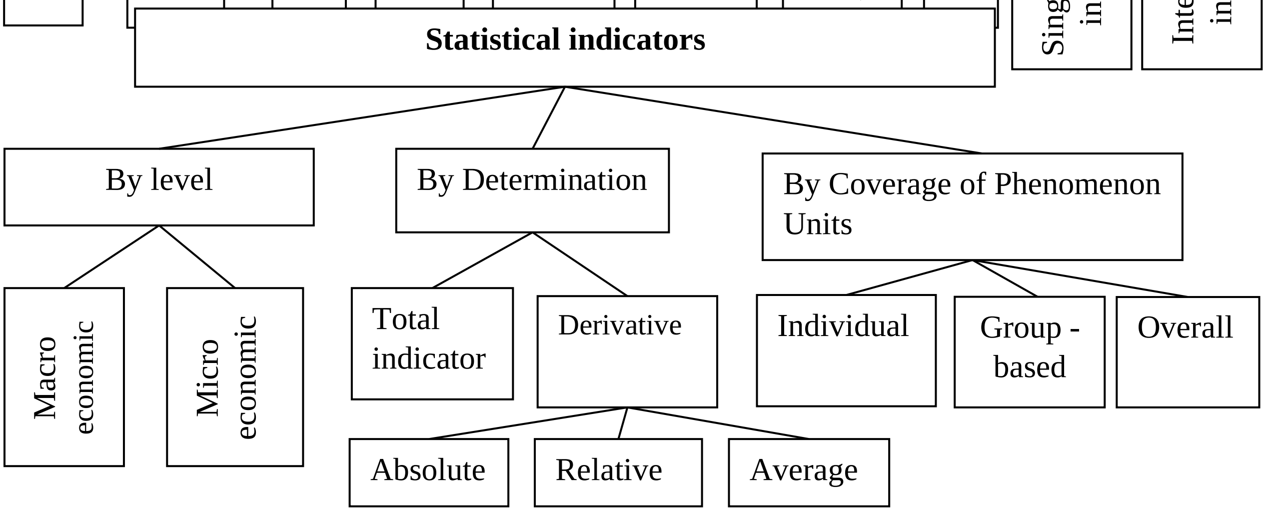


Figure 4.2: Classification of statistical indicators

Statistics are numerical values or measurements used to represent, describe, and analyze data. They play an important role in research, analysis, and decision-making. Using statistical indicators, it is possible to obtain accurate and reliable information about various processes and events, identify trends in them, and draw conclusions.

Statistical indicators are divided into several types according to different aspects and purposes. The main types are:

1. *Descriptive statistics*: It is calculated by dividing the sum of all values in the data set by their number.
2. *Inferential Statistics:* Hypothesis Testing is used to confirm or reject research hypotheses. Confidence Interval (Confidence Interval) shows the possible range for population parameters. P-value is a statistical measure used to verify a hypothesis different from the null in a hypothesis test.
3. *Variability Measures*: Dispersion (Variance) shows how far the values in the data set are from the average value. Standard Deviation is the square root of the variance and shows the average distance of the data from the mean. Mean Absolute Deviation shows the average deviation of the values in the data set from the mean value.
4. *Shape Measures:* Flatness (Kurtosis) shows the degree of skewness of the data distribution. Skewness shows the asymmetry of the data distribution.
5. *Relative Measures* refer to data values expressed as a percentage of the total volume, providing a proportional representation of the dataset. Relative Change quantifies the proportional variation in values within a dataset over time or between different conditions.

6. *Comparison Measures:* Indexes are used to compare data for different periods or regions. For example, the price index, and production index. The growth rate shows the rate of change of data over time.

4.2. Absolute Statistical (Indicator) Values

Absolute values refer to quantities that express the number, volume, or degree of mass social phenomena. These values are determined through two methods: 1) Aggregation; 2) Special calculations (For example: to calculate the population at year-end, births during the year are added to the beginning-year population, and deaths during the year are subtracted).

Absolute values are categorized into two types based on their representation:

1) Individual values (for instance: indicators of each student)
2) Aggregate values (for instance: indicators of the total students);

The units of measurement for absolute values comprise:

1. *Natural measurement units*: These include units of count, time, weight, and length.
2. *Conditional natural measurement units* – These are used to determine the total amount of qualitatively similar but structurally different phenomena. For instance, fuels such as wood, coal, and gas are converted into conditional fuel units based on their heat emission when burned. Similarly, dairy products are measured based on fat content, and animal feed (e.g., alfalfa, beet, rice, and fresh grass) is assessed based on nutritional caloric value.
3. *Value measurement units* – Used to calculate the volume of gross products and capital investments.
4. *Complex measurement units* – These consist of two or more measurement units. For example, the work performed by freight transport is measured in ton-kilometers, electricity consumption in kilowatt-hours, and labor time in man-hours.

4.3. Relative Statistical Indicators: Concept and Types

Relative indicators are numerical values derived from comparing two comparable measurements. The units of measurement for relative values are:

1) Coefficients: where the base of the ratio $\frac{a}{b}$ is normalized to one whole unit;

2) Percentages: where the base of the ratio $\frac{a}{b}$ is equated to 100 (%);
3) Per mille a/b ratio base is equal to 1000 (%$_0$);
4) Per ten thousand (Basis point): where the base of the ratio is equated to 10,000 (%$_{00}$).

The types of relative indicators are:

1. Relative quantity of contractual obligations (RQC) – This expresses the ratio of the planned level for the current year to the actual level of the previous year:

$$\textbf{RQC} = \frac{\textit{Planned level for the current year}}{\textbf{\textit{Actual level of the previous year}}}\ \mathbf{100\%;}$$

2. The following formula is used to find the relative amount of contract, plan execution (RQC $_{\text{per execution}}$):

$$\textbf{RQC}_{\textbf{per execution}} = \frac{\textit{Planned level for the current year}}{\textbf{\textit{Planned level for the current year}}}\ \mathbf{100\%;}$$

3. Relative quantities of dynamics express the change (increase or decrease) of events over time and are calculated using two methods:

Base method: each period's level is compared with the level of the first period;

Chain method: each period's level is compared with the level of the previous period;

4. Structural (composition) relative quantities (SRQ) determine the structure of a set and are calculated using the following formula:

$$\textbf{SRQ} = \frac{\textbf{\textit{Indicator of a specific group within the set}}}{\textbf{\textit{Total indicator of the set}}}\ \mathbf{100\%;}$$

Relative quantities of intensity determine the distribution density of events. For example: population density, productivity, labor efficiency.

5. Relative quantities of comparison (RCQ) characterize a particular event in different objects or regions during a specific period and are calculated as follows:

$$\textbf{RCQ} = \frac{\textit{Indicator describing object A}}{\textit{Indicator describing object b}}\ \textbf{100\%;}$$

Cr/q = (A pointer to the object A)/(A pointer to the object B)

Before comparing absolute and relative quantities, they must be converted into a comparable form. In the current market economy, solving this issue is highly relevant. The conditions for converting absolute and relative quantities into comparable form include:

1) Indicators must be expressed in the same unit of measurement.
2) Indicators must be calculated using the same methodology.
3) Attention must be given to ensuring that indicators are fully collected for a specific time and region. When using absolute and relative quantities, a differential approach is necessary, meaning that in addition to analyzing overall results, specific group indicators should also be examined.
4) It is essential to use absolute and relative quantities together. A single absolute or relative quantity alone does not provide a complete understanding of a phenomenon.

4.4. The Average Values: Concept and Types

Averages are an important part of statistical analysis and represent the central tendency of a set of data. They help to get an overview of the data, to identify similarities and differences, and to evaluate the results of the study. Means are used to determine the central tendency of data by showing the average value of the data.

Throughout our lives, we frequently encounter the concept of "average." For example, average gross product per capita or average salary per employee in an enterprise. Absolute and relative values do not provide a generalized characterization of a phenomenon; instead, average values summarize and describe the variability of a homogeneous set of events or indicators. Average values serve as a fundamental tool in socio-economic analysis. The rules for calculating averages are as follows:

1) The characteristics used for calculating the average should be relevant and meaningful to the analysis.
2) The average should be calculated for a group of similar and comparable events or data points.
3) The average should be calculated not only for the entire dataset but also for specific groups within it.
4) The accuracy of an average improves as the number of data points increases, following the law of large numbers.

Types of Average Values

A. Analytical averages: (Arithmetic mean, Geometric mean, Chronological mean);
B. Structural averages: (Mode, Median, Quantiles).

Arithmetic Mean: Arithmetic mean is calculated by dividing the sum of all data values by their number. The most commonly used quantity is the average, which is often used to determine the overall performance of a set of data.

Median: The median is the value in the middle when arranging a set of data. If the number of values is even, the arithmetic mean of the middle two values is taken. It is used for ordered data and is used when there is inequality between variables.

Mode: The mode is the most repeated value in the data set. Used to identify multiple repeated values and display the most common value in a data set.

Geometric Mean: The geometric mean is calculated by taking the root of the product of all values. Used for percentage changes, relative growth rates, and other multiplicative data.

Harmonic Mean: The harmonic mean is calculated by dividing the sum of the inversions of the values by the number of values. Used for ratios and correlated values, such as speed and efficiency measurements. Averages play an important role in expressing the central tendency of data and in their analysis. Their different types are selected depending on the type of data and the purpose of the analysis. While the arithmetic mean is the most common, the median and mode are useful for detecting skewness in the distribution of data. Geometric and harmonic means are used for relative and multiplicative data. A properly chosen average will help increase the accuracy of the data and the reliability of the analysis results.

4.5. Types of Analytical Averages

There are various methods for calculating the average value, the ***arithmetic mean*** is the most commonly used. The arithmetic mean has two forms: simple and weighted.

In an unweighted variation series, the simple arithmetic mean is calculated by summing all observed values of a variable and dividing the total by the number of observations. However, in a weighted series, where different values have varying levels of importance or frequency, the weighted arithmetic mean is used to ensure a more representative average. The simple arithmetic mean provides a general measure of central tendency, but it may not always reflect the true distribution of values when weights are significant.

$$\bar{X}_{simple\ arithmetic\ mean} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where: $\sum$ - denotes sum;

X - represents the individual values of the studied variable

n - numbers of values (number of row options)

If the variable X has n values $x_1, x_2, \ldots, x_n$ or x_i $(i = \overline{1, n})$ occurring $f_1, f_2, \ldots, f_n$ or f_i $(i = \overline{1, n})$ times respectively, then the weighted arithmetic mean is calculated as:

$$\bar{X}_{\text{weightedarith}} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Where f_i represents the weight (or frequency) of each value.

Properties of the Arithmetic Mean:

1) The sum of the differences between individual values and their arithmetic mean is always zero:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \quad \text{or} \quad \sum_{i=1}^{n}(x_i - \bar{x}) f_i = 0$$

2) The sum of the squared differences between individual values and their arithmetic mean is minimized:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \min \quad \text{or} \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 f_i = \min$$

3) If each value of the variable is multiplied or divided by a constant (arbitrary number), the mean is also multiplied or divided by that constant.

$$\frac{\sum_{i=1}^{n}\frac{x_i}{c} f_i}{\sum_{i=1}^{n} f_i} = \bar{x} * c$$

4) If a constant (arbitrary number) is added to or subtracted from each value of the variable, the mean also changes by that constant (arbitrary number).

$$\frac{\sum_{i=1}^{n}(x_i - c) f_i}{\sum_{i=1}^{n} f_i} = \bar{x} + c$$

5) The mean remains unchanged if all weights are multiplied or divided by the same constant.

$$\frac{\sum_{i=1}^{n} x_i \frac{f_i}{c}}{\sum_{i=1}^{n}\frac{f_i}{c}} = \bar{x}$$

The purpose of studying the properties of the arithmetic mean is to simplify, speed up, and smooth calculations. Using the properties of the arithmetic mean, the average value in the variation series is calculated by the 'moment' method.

In this method, a constant arbitrary number 'A' is subtracted from the series of variants, and the obtained result is divided by another arbitrary number 'B'. As a result, a new series $y_i = \frac{x_i - A}{B}$ is formed from the given x_i. The arithmetic mean is then calculated for this new series.

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i}$$

Then, it is multiplied by the number 'B', and the number 'A' is added to the obtained result. As a result, the true arithmetic mean of the initial series is obtained.

$$\bar{x} = \bar{y}B + A.$$

In series with equal-width intervals, it is recommended to take 'A' as the value of the variant with the highest frequency and 'B' as the interval width

Geometric Mean

In asymmetric distributions, especially those that are highly skewed (or peaked, elongated), the use of the geometric mean is more justified. In socio-economic life, many phenomena follow such distribution patterns.

The geometric mean X_{geom} is obtained by taking the nth root of the product of n terms in the series:

$$\bar{x}_{geom} = \sqrt[n]{x_1 * x_2 * x_3 ... x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

there: $\prod_{i=1}^{n}$ means the product of terms. For example, if the width of a house is 5 m, the length is 11.4 m, and the height is 4 m, what is the average side length of the house volume?

$\bar{x}_{geom} = \sqrt[3]{5*11,4*4} = \sqrt[3]{228} = 6,11m$.

In a distinctly asymmetric distribution (if it arises from the nature of the phenomenon rather than randomness), the arithmetic mean is always somewhat of a 'false' average.

In such cases, the geometric mean more accurately represents the central tendency of the distribution. Asymmetric distribution forms when the random variability of a trait combines with systematic, stable differences (for example, differences in wages among equally skilled workers). When converted to a logarithmic scale, such a distribution takes on a 'normal' shape, meaning that the logarithms of the values follow a normal distribution.

The nature and characteristics of such distribution series are accurately reflected in the geometric mean because it is based on the logarithms of the series terms.

$$\log \bar{x}_{od.geom} = \frac{\log x_1 + \log x_2 + ... + \log x_n}{n}.$$

In our example above $\log \bar{x}_{geom} = \frac{\log 5 + \log 11{,}4 + \log 4}{3} = 0{,}699 + 1{,}057 + 0{,}602$,

$\bar{x}_{geom} = 6{,}11$

Harmonic Mean and Its Applications

The harmonic mean is a type of average used in specific situations where rates, ratios, or proportional relationships are involved. It is particularly useful when averaging values expressed in terms of per-unit measurements, such as speed, efficiency, or financial ratios.

In the harmonic mean, the sum of the reciprocals of the variable values (1/x) $\sum_{i=1}^{n} \frac{1}{x_i} = const$, is considered a constant quantity. It is self-evident that when determining the average for economic phenomena, this rule must be justified from the perspective of the economic essence of the phenomenon. Otherwise, the obtained average value and its qualitative basis will not correspond to each other.

Simple harmonic mean: $\bar{X}_{simpleharm.} = \frac{1_1 + 1_2 + ... + 1_n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}$

or shortly, $\bar{X}_{simple.harm.} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

The weighted harmonic mean is used when the averaged values have different weights (Wi) and is calculated as follows:

$$\bar{X}_{weighted.harm.} = \frac{w_1 + w_2 + w_3 + ... + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \frac{w_3}{x_3} + ... + \frac{w_n}{x_n}} = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}};$$

It is known that any average value arises from the ratio of two indicators. The first indicator represents the total volume of the studied characteristic, while the second indicator determines the number of entities possessing this characteristic (its weight or frequency). If the data representing the total volume of the characteristic (i.e., the numerator of the ratio) and the individual levels of the characteristic are known, then the average value is calculated using the harmonic mean formula. However, if the total volume of the characteristic and the total number of entities are known, but the individual levels are unknown, then the aggregate average

4.6. Structural Average Values

The average value is the mean of varying quantities. It represents the general trend or pattern characteristic of the dataset while also masking individual values of the variable. A market economy requires reliance on precise values of a variable when solving real-life issues. For example, demand for clothing and footwear is determined not by their average size but by the exact quantity of each specific size. Therefore, forecasting supply prospects is also based on such specific data. The demand for gasoline, spare parts, and tires for automobiles is determined not by their average values but by their specific types. Similarly, supply is also based on these precise indicators.

The depreciation of the national currency and the emergence of inflationary processes are inherent features of a market economy. Studying these processes requires continuous monitoring of market prices. However, prices are subject to significant cyclical fluctuations and vary widely across retail outlets, individual sellers, buyers, and over time. As a result, for the same product in the market, different price levels can be observed. Since it is impractical to record all price variations, in practice, the most frequently occurring price level for each product in the market is recorded.

Clarifying the characteristics and patterns in the structure of datasets and analyzing the concentration of their units within specific intervals requires not only the use of average values but also the identification of descriptive parameters known as central tendency measures in distribution series. These measures include the ***mode, median, and quantiles.***

The *mode* refers to the most frequently occurring value in a dataset. In discrete series, it is identified as the value with the highest frequency.

In grouped (interval) series, the mode is determined using the following formula:

$$\mu_0 = x_0 + \frac{f_{\mu_0} - f_{\mu_{0-1}}}{(f_{\mu_0} - f_{\mu_{0-1}}) + (f_{\mu_0} - f_{\mu_{0+1}})} i_{\mu_0}$$

Here, μ_0 – Mode (the most frequent value);

x_0 – The lower boundary of the modal class (interval);

f_{μ_0} – Frequency of the modal class;

$f_{\mu_{0-1}}$ – Frequency of the class before the modal class;

$f_{\mu_{0+1}}$ – Frequency of the class after the modal class;

i_{μ_0} – The width of the modal class interval.

The *median* is the value that divides a dataset into two halves. In an ordered or ranked series, the median corresponds to the middle value. If the series has an odd number of values (e.g., 9 or 15 elements), the median is the value at the central position, such as the 5th or 8th element, respectively.

In grouped data with an odd number of intervals, the median is calculated using the following formula:

$\mu_e = x_0 + \dfrac{\dfrac{\sum_{j=1}^{K} f_j}{2} - f^{/}_{\mu_{e-1}}}{f_{\mu_e}} i_{\mu_e}$ and in grouped data with an even number of intervals: $\mu_e = x_0 + \dfrac{\dfrac{\sum_{j=1}^{k} f_j + 1}{2} - f^{/}_{\mu_{e-1}}}{f_{\mu e}} * i_{\mu_e}$

where: : μ_e -median;

x_0 - Lower boundary of the interval (group) containing the median;

$f'_{\mu_{e-1}}$ - the total number of units for the interval before the median;

$f_{\mu e}$ - the number of units in the median interval;

i_{μ_e} - Class width of the median interval;

k - number of intervals (groups);

$\sum f_j$ - sum of all frequencies.

A quantile is a statistical value that divides a dataset into equal-sized parts (4, 5, 10, and 100 parts). It helps to understand the distribution of data by splitting it into fractions or percentages.

A quantity (variant value) that divides the series into four equal parts is called a quartile, one that divides it into five parts is called a quintile, one that divides it into ten parts is called a decile, and one that divides it into one hundred parts is called a percentile. Each set of series has 3 quartiles, 4 quintiles, 9 deciles, and 99 percentiles. For example, if a dataset is divided into quintiles, it is separated into five equal parts based on its distribution: the lowest 20% of values fall below the first quintile (Q_1), the next 20% (20%–40%) lie between Q_1 and Q_2, the middle 20% (40%–60%) are between Q_2 and Q_3, the next 20% (60%–80%) fall between Q_3 and Q_4, and the top 20% of values lie above Q_4. Another example, the lower quartile (Q_1) is the value in an ordered series where one-quarter (1/4) of the total data points have values smaller than it, while three-quarters (3/4) have values greater than it. The upper quartile (Q_3) follows the opposite pattern, meaning that three-quarters (3/4) of the total data points have values smaller than it, while one-quarter (1/4) have values greater than it. The lower quartile is denoted as Q_1, and the upper quartile as Q_3.

$$Q_1 = X_{0(Q_1)} + \frac{\sum_{j=1}^{k} \frac{f_j}{4} - f'_{Q_1-1}}{f_{Q1}} * i$$

$$Q_3 = X_{0(Q_3)} + \frac{\sum_{j=1}^{k} \frac{f_j}{4} - f'_{Q_3+1}}{f_{Q_3}} * i$$

$$Q_2 = \mu_e.$$

Table below presents the formulas for calculating the first and last quintiles, deciles, and percentiles in grouped data.

Table 4.1.

Formulas for Determining the First and Last Quintiles, Deciles, and Percentiles in Grouped Data

Indicators	Initial indicator	Last indicator
Quintile (W)	$W_1 = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{5} - f'_{W_1-1}}{f_{W_1}} * i$	$W_4 = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{5} - f'_{W_4+1}}{f_{w_4}} * i$
Deciles (D)	$D_1 = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{10} - f'_{\text{Б}_1-1}}{f_{\text{Б}_1}} * i$	$D_9 = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{10} - f'_{\text{Б}_9+1}}{f_{\text{Б}_9}} * i$
Percentile(F)	$F_1 = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{100} - f'_{F_1-1}}{f_{F_1}} * i$	$F_{99} = X_0 + \dfrac{\sum_{j=1}^{k} \frac{f_j}{100} - f'_{F_{99}+1}}{f_{F_{99}}} * i$

In a symmetric distribution, the arithmetic mean, mode, and median are equal. However, in asymmetric distributions, they differ. In a right-skewed (positively skewed) distribution, they are arranged in the following order:

$\mu_o, \mu_e, \bar{x}_{arif}$ Mode < Median < Mean. In a left-skewed (negatively skewed) distribution, they follow the opposite order: $\bar{x}_{arif}, \mu_e, \mu_o$ Mean < Median < Mode.

Review Questions:

1. How are statistical indicators described?
2. Explain the concept of absolute values. What measurement units of absolute values do you know?
3. Define relative values.
4. What measurement units of relative values do you know?
5. What types of relative values exist?
6. List the types of average values.
7. By which formula is the arithmetic mean calculated?
8. When is the harmonic mean formula used?
9. What is meant by structural average values?
10. What is the essence of mode and median, and how are they calculated?

CHAPTER 5. VARIATION ASSESSMENT METHODS AND DISPERSION ANALYSIS PRINCIPLES IN STATISTICS

5.1. The Concept of Variation and Its Importance in Statistical Analysis

In a dataset, variation in a particular characteristic refers to the fluctuation or divergence of its values across individual units within a specific time and region. The elements of a dataset operate in different environments, leading to the emergence of variation. The primary cause of variation is the diversity of conditions, as numerous factors and forces exist, acting in different ways and influencing the outcome to varying degrees.

Variance is a statistical measure of the differences between values in a dataset. Shows how much the values in the data set differ from each other. Variance can be used to determine the spread of a set of data and how much it deviates from the mean.

Dispersion is one of the main metrics used to measure variation. It calculates how far each value is from the mean with a square mean. And the square root of the variance gives the standard deviation, which represents the deviation from the mean.

The importance of conducting a statistical analysis of variation can be outlined as follows:

- Variation plays an important role in determining the spread of values in a data set. This makes it possible to identify differences between values and evaluate them.
- By obtaining variation information, researchers and decision-makers can make inferences about the spread of data and the extent of variability. This helps in making scientific, economic, and managerial decisions.
- Variation helps in identifying and minimizing statistical errors. By knowing the distribution and variability of the data, the reliability of the research results can be increased.

- By studying variation, one can identify the causes of variation in a data set. This process allows studying the influence of internal and external factors.

Variation is used to assess the accuracy of statistical models and predict future events from them. A small variation indicates the stability and accuracy of the models. Variation is an important indicator in the analysis of the results of experiments and scientific studies. Through it, the level of variability in the data set is determined and the results are evaluated.

5.2. Measures of Variation

Variation, defined as the dispersion of characteristic values around central measures (levels of the characteristic), is most simply quantified by the range. The *range (*denoted as R) represents the difference between the maximum and minimum observed values of the characteristic: $R=X_{max} - X_{min}$.

X_{max} is the maximum value of the characteristic (upper bound) and X_{min} is its minimum value (lower bound). The range does not reflect the internal structure of the distribution, meaning it does not account for differences between individual values. The range may have the same value in both symmetric and asymmetric distributions (e.g., J shaped distributions) even though these distributions often differ significantly in terms of dispersion.

The mean squared deviation, or variance, is another measure of variation. It is defined as the arithmetic mean of the squared differences between individual values and their arithmetic mean.

This indicator is expressed using the following formulas:

For unweighted (simple) series: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$

For weighted (grouped) series: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{\sum f_i}$

Where, σ^2 represents the variance;

x_i denotes the values of the series variants;

$\bar{x}$ is the arithmetic mean of the variant, calculated using the following formula: $\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f_i}$, here, f_i refers to the frequency (number of units) of the variants. Thus, variance can be calculated using the following formula.

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2$$

Measures of variation are measurements that represent the spread of values (or, the degree of variability) in a set of statistical data. They are used to determine how far a set of data deviates from the central tendency, whether the data are similar or different. Indicators of variation play an important role in assessing the reliability and randomness of data during the analysis process.

Variance: Variance represents how far the values in a set of data are from the mean, using the mean squared. This metric measures the overall variability in the data set

Standard Deviation is the square root of the variance and indicates the average distance of the data set from the mean.

Mean Absolute Deviation represents the average deviation of the values in the data set from the average value. Unlike the variance, this metric is calculated without squaring.

The coefficient of variation expresses the variability of the data as a percentage compared to the average value. This indicator is used to estimate relative volatility.

Interquartile range (IQR) represents the spread of data by the difference between the first and third quartile values. This indicator shows the distribution of the main part of the data set.

Indicators of variation play an important role in determining the spread of data and the degree of variability. They are used in the analysis process to assess the reliability, randomness, and general distribution of the data. Correctly selected variation indicators ensure accurate and reliable analysis of the data set.

Table 5.1.

Statistical Variation Indicators

Indicators	**Meaning**	**Formula**
Dispersion (Variance)	The variance represents how far the values in a set of data are from the mean, using the root mean square.	$\sigma^{2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$
Standard Deviation	The standard deviation is the square root of the variance and represents the deviation from the mean.	$\sigma = \sqrt{\sigma^2}$
Mean Absolute Deviation	The standard deviation represents the average deviation of the values in the data set from the mean.	$MAD=\frac{\sum_{i=1}^{n}[X_i-\bar{X}]}{n}$
Interquartile Range	The interquartile range is the difference between the first and third quartile values and represents the spread between the data sets.	**$IQR=Q_3-Q_1$**

5.3. Properties of Variance and Quadratic Mean Deviation

Variance and quadratic mean deviation are the most convenient measures of variability for performing algebraic operations. The key properties of variance and quadratic mean deviation are as follows:

When σ^2_x and σ_x are calculated relative to the arithmetic mean, these indicators represent the smallest possible measure of variability, meaning that in this case $\sigma_X^2 \langle S_A^2$, $A \neq \bar{x}$. As proven above:

$$S_A^2=\frac{\sum (X-A)^2}{N}=\sigma^2_{\bar{x}}+d^2$$

Here, $d^2q(X\text{-}A)^{2,}$ thus $S_A^2 \rangle \sigma^2_{\bar{x}}$ from $\sigma^2_{\bar{x}}=S_A^2-d^2$

If all elements of the series are increased or decreased by a constant A i.e. X-A, this case does not affect the variance and the quadratic mean deviation. That is, for the new series Y = X - A, these indicators remain equal to those of the original series. $\sigma_y^2 = \sigma_x^2$

If all elements of a series are scaled by a constant factor k (either multiplied or divided), the variance decreases or increases by a factor of k^2, while the quadratic mean deviation changes by a factor of k.

$$\sigma^2_{\bar{x}} = \frac{N_1(\sigma_1^2 + d_1^2) + N_1(\sigma_2^2 + d_2^2)}{N_1 + N_2}$$

Determining the quadratic mean deviation for the first N natural numbers is also of practical importance. From algebra, it is known that the sum of the first N natural numbers is given by: N (N + 1)/2, and the sum of their squares is expressed N(N+1)(2N+1)/6. Thus, the arithmetic mean of the first N natural numbers is:

$$\frac{N(N+1)}{2} \div N = \frac{N+1}{2}$$

According to the formula, their variance is given by:

$$\sigma^2 = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4}$$

which simplifies to

$$\sigma^2 = \frac{N^2-1}{12}$$

This formula can be applied by organizing the dataset based on a common characteristic without directly measuring the levels of the attribute. By ranking the units accordingly and assigning ordinal numbers, an N-ranked series can be constructed and analyzed.

5.4. Simplified Methods for Calculating Variance and Quadratic Mean Deviation

Based on the properties of variance discussed earlier, the calculation of both variance and quadratic mean deviation can be simplified to some extent. One such method is known as the method of conditional moments.

In this approach, a constant value A is subtracted from each term x_i in the given series, and the resulting values are then divided by another constant K. This transformation results in a new series u_i replacing the original x_i series. If the series consists of evenly spaced variants, it is recommended to choose A as the variant with the highest frequency and K as the class interval width. This selection significantly simplifies calculations. After this transformation, the arithmetic mean of the new y_i series, along with the mean of its squared values, is computed.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{N} \quad or \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i f_i}{\sum f_i} \; and \quad \overline{y^2} = \frac{\sum_{i=1}^{n} {y_i}^2}{N} \quad or \quad \overline{y^2} = \frac{\sum_{i=1}^{n} {y_i}^2 f_i}{\sum f_i}$$

As a result, $\sigma_{\bar{y}}^2 = к^2(\overline{y^2} - \bar{y}^2)$

This indicator also determines the variance of the original x_i series, because

$$\sigma_{\bar{y}}^2 = \sigma_{\bar{x}}^2 \text{ or } \overline{y^2} - \bar{y}^2 = \overline{x^2} - \bar{x}^2$$

The Rule of Adding Variances and Its Application in Market Analysis

The total variance ($\sigma_{x_i}^2$) is the result of adding the average partial variance ($\overline{\delta_i^2}$) and the variance of partial means ($\sigma_{\bar{x}_i}^2$). This is called the rule of adding variances. According to this law, the total variance consists of two component variances: one measures the variability within the subsets of the dataset, while the other describes the differences between these subsets expressed through their partial means (variation). The essence of each variance will be clarified with the following example.

If the dataset units are grouped based on a significant characteristic, then the distribution series is defined by three types of variances: total variance, intergroup variance, and intragroup variance. Total variance represents how the studied characteristic varies under the influence of all factors, while intergroup variance measures which part of this variation is formed as a result of the grouping characteristic's influence. The remaining part of the total variation accounts for the contribution of all other factors and is determined by intragroup variances. As a result, total variance consists of intergroup variance and the average intragroup variance, i.e. $\sigma_{\bar{x}}^2 = \overline{\delta_{x_i}^2} + \sigma_{\bar{x}_i}^{\ 2}$, where $\sigma_{\bar{x}}^2$ represents the total variance

- Total variance: $\sigma_{\bar{x}}^2 = \frac{\sum(x-\bar{x})^2}{n}$ there, $\bar{x} = \frac{\sum x}{n}$
- Intergroup variance: $\sigma_{\bar{x}_i}^{\ 2} = \frac{\sum(\bar{x}_i - \bar{x})^2}{n_i}$ where i - is the number of groups and the average value of the symbol for each group;
- Average intragroup variance: $\overline{\delta^2}_i = \frac{\sum \delta_i^2 n_i}{\sum n_i}$ there $\delta_i^2 = \frac{\sum(x_i - \bar{x}_i)^2}{n_i}$;

  x - individual values of the characteristic in the dataset;
  x_i - individual values of the characteristic within each group;
  n_i - the number of units in each group;
  n - the total number of units in the dataset.

Alternative – derived from the Latin root “alter”, meaning "one of the two" – refers to each of the mutually exclusive possibilities or choices. An alternative characteristic is a property that is present in some units of the studied population while absent in others. For example, a portion of consumers may be inclined to consume a particular product, whereas another portion may not.

The values of an alternative characteristic are expressed as "1" for units possessing the characteristic and "0" for those that do not. In the total population, the proportion of units exhibiting the alternative characteristic is denoted by "p", while the proportion of units not exhibiting it is denoted by "q", with their sum equal to one:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1p + 0q}{p+q} = 1p + 0q = p$$

Thus, the mean value of an alternative characteristic is equal to the proportion of units possessing that characteristic in the total population. The variance for this characteristic is given by the formula:

$$\sigma_p^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = pq \text{ thus, } \sigma_p^2 = pq$$

The maximum value of the variance for an alternative characteristic is: pq=0,5*0,5=0,25. The three types of variances mentioned above are interrelated as follows: $\sigma_p^2 = \overline{\delta_{p_i}^2} + \sigma_{\bar{p}_i}^{\;2}$

In this case, the individual deviations are summed without considering their signs. The arithmetic mean of these absolute deviations is called the mean absolute deviation (MAD). This indicator is expressed in the following forms:

For ungrouped series: $\bar{d} = \frac{\sum |x - \bar{x}|}{n}$;

For weighted series: $\bar{d} = \frac{\sum_{i=1}^{n} |\; x_i - \bar{x} \;| f_i}{\sum f_i}$,

here, d (d – modulus) represents individual deviations of sequence elements. Just as the quadratic mean deviation reaches its minimum value when deviations are measured relative to the arithmetic mean, the absolute mean deviation also attains its minimum value when deviations are measured relative to the median.

In a symmetric distribution, the median is the point located in the middle of the distance between the first and third quartiles, dividing this distance into two equal parts, $\mu_e - Q_1 = Q_3 - \mu_e$

This difference can be interpreted as a measure of variation. However, since a perfectly symmetric distribution never exists, the measure of variation is usually taken as half of the difference between the third quartile and the median, as well as between the median and the first quartile. It is calculated using the following formula:

$$Q = \frac{(Q_3 - \mu_e) + (\mu_e - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

The semi-interquartile range describes the variability specific to only the central part of the dataset, ignoring variations in other sections. For this reason, in our example, it has a smaller value compared to the mean absolute deviation.

All the variation measures discussed above are expressed in the same units as the measured characteristic. However, it is not possible to compare the variation of datasets with different measurement units using these indicators. Even when the measurement units are the same, comparing the variation of datasets with different natures is not feasible. Therefore, statistics recommend using relative measures of variation. Since the standard deviation and mean absolute deviation are expressed in the same units as the measured characteristic, they should be divided by a certain measure of the characteristic level, such as: $R/\bar{x}$; $\bar{d}/\bar{x}$; $\sigma/\bar{x}$. As a result, the obtained indicators are called variation measures. The last expression above is usually calculated as a percentage and is referred to as the coefficient of variation.

$V = \frac{\sigma}{\bar{x}} * 100\%;$ here $\bar{x}$ - the arithmetic mean of the characteristic and σ - represents the standard deviation. When the average value is close to zero, this coefficient is considered somewhat unreliable.

Review Questions:

1. Define the concept of variation in statistical analysis. Why is it essential to measure variation in a dataset?
2. By which formula are the main variation indicators calculated in ranked (unweighted) series?
3. By which formula are the main variation indicators calculated in weighted series?
4. Explain the computation of total variance. What are its key advantages and limitations in statistical analysis?
5. Discuss the fundamental properties of variance and their implications in data analysis.
6. How is intergroup variance calculated?
7. How is the average internal variance calculated?
8. How is variance calculated using a simplified method?
9. List the types of variance and their calculation formulas.

10. How is the variance of an alternative (binary) variable determined? What is its significance in statistical inference?

CHAPTER 6. SAMPLE OBSERVATION

6.1. The Concept and Significance of Sample Observation

In real life, there are large-scale social phenomena where the total volume of data collected during a complete census may change over time (e.g., studying household incomes across a republic). In such cases, sample observation is applied. In statistics, sample observation refers to selecting a necessary number of units from a population using specific methods, examining them, and generalizing the findings to the entire population. The population under study is called the *parent population*, while the subset selected for examination is known as the sample population. The sample population must be *representative*, meaning it should accurately reflect all the characteristics of the parent population. Sample observation helps save time and resources, reduces the number of products that may degrade in quality or lose their value due to complete enumeration, and allows for a more in-depth study of the subject while verifying the results of a complete census.

Table 6.1.

Summary Indicators in Sampling

Indicators	Population	Sample
1. Number of Units	N	n
Units with Specific Characteristics: 1) count	M	m
b) proportion	$p = \frac{M}{N}$	$w = \frac{m}{n}$
2. Mean value	$\bar{x} = \frac{\Sigma x}{N}$	$\tilde{x} = \frac{\Sigma x}{n}$
3. Variance	$\overline{\sigma^2} = \frac{\Sigma(x-\bar{x})^2}{N}$	$\widetilde{\sigma^2} = \frac{\Sigma(x-\tilde{x})^2}{n}$
4. Proportion Varience	$\sigma^2_p = p*q$	$\sigma^2_w = w*(w-1)$

q - the proportion of units that do not possess the specified characteristic

A sample survey is a process of collecting information from a small part of the population (ie, the entire data set). This method is used when it is not possible to monitor the entire population or it is too expensive and complicated. Based on the results obtained through selective observation, general conclusions can be drawn about the population.

The main components of selective monitoring are:

- Population: A collection of all possible subjects or respondents of a study.
- Sampling: A small group selected from a population.
- Sampling Method: Methods used to form the sample (random sampling, stratified sampling, etc.).
- Sample Size: The number of units included in the sample.

The following methods of selective (sample) observation can be mentioned:

- *Random Sampling* is a method in which each population unit has an equal chance. This method is statistically the most reliable.
- *Stratified Sampling*: The population is divided into strata, and a random sample is taken from each stratum. This method increases the level of representation of different groups in the population.
- *Cluster Sampling* divides the population into clusters (groups) and collects data from randomly selected clusters. This method is convenient for large populations.
- *Systematic Sampling*: each part of the population is selected in a row (for example, every 10th unit). This method is simple and convenient, but can be used when the population is orderly.

Selective observation is widely used in economics, social science, medicine, marketing and other fields. The following aspects show its importance:

Less time, money, and human resources are spent, but tangible results are obtained.

Based on the results of selective observation, scientific, economic and management decisions are made.

When complete observation of large populations is not possible, general conclusions are drawn by selective observation.

General trends, problems, and opportunities in the population are identified and measures are developed accordingly.

Sample observation is an efficient and cost-effective method of statistical analysis, which allows obtaining accurate and reliable data in various fields. Using this method, it is possible to draw general conclusions about large populations and make strategic decisions.

6.2. Types of Sample Observation

In statistical sampling, the following methods are used to select units from the population and can be classified into random with replacement and random without replacement methods.

1) **In random sampling**, each unit of the population is assigned a number. These numbers are written on pieces of paper, placed in identical-looking containers, and thoroughly mixed. The required number of units is then randomly selected. If the selected unit's number is recorded and then returned to the population for possible reselection, the method is called random with replacement. If the selected unit is not returned to the population after being recorded, the method is termed random without replacement.
2) **Systematic sampling** is carried out in the following manner:
   a) The units of the population are arranged and numbered according to the research objective. For example, workers can be ordered based on the volume of products they have produced, students can be ranked according to their entrance exam scores, or individuals can be listed alphabetically by surname.
   b) Interval length is determined: $i = \frac{N}{n}$

Here: N - the number of units in the population.

n - the number of units in the sample.

1 - an arbitrary k unit is selected from the first interval;

2 - k + i unit from the second interval;

3 - k + 2*i unit from the third interval;
n - k + (n-1)*i unit from the n^{th} interval.

For example, selecting 200 workers from 10,000 using the mechanical method is carried out as follows:

a) Workers are arranged and numbered based on the studied characteristics.
b) interval length i=N/n=10000/200=50;

From the 1st interval, a random unit is selected. For example, if k = 25;

From the 2nd interval: $k+i$=25 +50=75

From the 3rd interval: k+2i=25+2x50=125, and so on.

3) **Stratified sampling** is conducted as follows: a) The population units to be studied are divided into homogeneous groups based on important typical characteristics and b) Units from each group are selected randomly in proportion to their weight in the total population. For example: To select 1,000 students from 15,000 students across 6 faculties of an institute using the stratified method, students in the institute are grouped by faculties. The proportion of each faculty's students relative to the total student population is determined. Then 1,000 students are selected according to these proportions, i.e. from 15,000 students:
XIF-30%, MIF-25%, KIF-20% and etc.

Thus, 30% of the 1,000 students will be selected from XIF, 25% from MIF, and so on.

4) **In cluster (group) sampling**, instead of selecting individual units from the population, entire groups (clusters) are selected using either a random or systematic method. For example, if 15% of 15,000 students in an institute need to be selected using cluster sampling:

15,000 ÷ 25 = 600 groups (assuming each group consists of 25 students).

15,000 represents 100%, and 15% of this needs to be selected:

(15,000 × 15) ÷ 100 = 2,250 students.

2,250 ÷ 25 = 90 groups.
Thus, 90 out of the 600 groups will be randomly or systematically selected.

In combinational sampling, multiple sampling methods are applied together. The main population is divided into groups based on an important characteristic, and necessary units are selected using different methods—some groups are sampled randomly, while others are selected systematically.

Sample survey data is used to characterize the population through summary indicators. For this purpose, the sample must accurately represent all significant characteristics of the population. If the sample reflects the key attributes of the population, it is considered representative (valid).

Regardless of how representative the sample is, there will always be differences between population and sample indicators. This is because the population includes units not selected in the sample. These differences are known as sampling errors. Sampling errors are classified into two types:

- Random errors
- Systematic (regular) errors.

Errors that occur unintentionally during the observation process, such as those resulting from the observer's fatigue, are considered random errors. According to the law of large numbers, as the sample size increases, random errors decrease.

Systematic errors, on the other hand, can occur both intentionally and unintentionally. Unintentional systematic errors arise due to inaccuracies in measuring instruments or deficiencies in sampling and observation methods. Intentional systematic errors occur when observation results are deliberately altered. For example, if the proportion of high-quality products in the sample is artificially increased to exaggerate the overall quality of production, a systematic error is introduced. Individuals who deliberately manipulate observation results may be held criminally liable.

Table 6.2.

Commonly Used Types of Sample Observation in Practice

Selective observation types	Meaning	Advantages	Drawbacks	Example
Random Sampling	A method in which each unit of the population has an equal chance	The level of representativeness of the data set is high, errors are minimal	It can be expensive and complicated for large populations	Randomly select 100 people from a population of 1000 people
Stratified Sampling	The population is divided into strata and a random sample is taken from each stratum	The more representation is provided from each layer, the more accurate the data set becomes	The layers can be difficult to identify and separate	Stratify a country's population by age group and take a random sample from each group
Cluster Sampling	The population is divided into clusters and data are collected from randomly selected clusters	Convenient and economical for large populations	Variability between clusters can be high	Creating clusters by cities and conducting population surveys within randomly selected cities
Systematic Sampling	Each part of the population is selected in a row	Simple and convenient, easy to use	Errors may occur in cases where the population is orderly	Choosing every 10th house on the streets of a city
Multistage Sampling	It is a multi-stage selection process, with different methods used at each stage	Suitable for large and complex populations	It is complicated and time consuming	First select regions, then select districts within the region, then select households within each district
Quota Sampling	The researcher selects respondents	Convenient and economical, it allows to take into	The selection process can be subjective,	Select a certain number of respondents for each gender and

	based on certain quotas	account many factors	which can affect the results	age group in the study
Adaptive Sampling	The selection process may change during the course of the study	Flexible and results can be more accurate	The process can be complicated and expensive	Sampling more in areas of high probability

Errors intentionally made to manipulate the results of an observation are classified as systematic errors committed knowingly. For example, a systematic error may occur if the sample contains a disproportionately high number of high-quality items compared to the overall population to artificially enhance the perceived quality of manufactured products. Individuals who deliberately falsify observational data may be subject to legal consequences, including criminal liability.

6.3. Sampling Errors

Any sampling estimate (a) has an upper error bound (Δ_a), which is equal to the product of its standard error (μa) and the confidence coefficient (t):

$$\Delta_a = t * \mu_a$$

The methods for determining the confidence coefficient and standard errors are as follows:
The confidence coefficient *t* and its relationship with probability P(t) is expressed using the following integral:

$$P(t) = \frac{1}{2\pi} \int_{-t}^{t} e^{-\frac{z^2}{2}} dz$$

To simplify the process of calculating probabilities for given confidence coefficient values, a table has been created that characterizes the relationship between them. This table allows for determining the probability corresponding to a given confidence coefficient and vice versa, identifying the confidence coefficient for a desired probability. In practical or academic problem-solving, the following confidence coefficient values are widely used:

t	1.00	1.96	2.00	2.58	3.00
P(t)	0.683	0.950	0.954	0.990	0.997

The standard errors of sample indicators (μ) vary depending on the sampling methods and structure.
The standard error (μx) of the sample mean ($\tilde{X}$) is calculated as follows, based on the sampling methods and structures:

Table 6.3.

Formulas for Determining the Standard Error of the Sample Proportion

T/R	**Sampling Methods and Their Types**	Sampling scheme	
		Repetitive	**Non - Repetitive**
1	Simple random sampling	$\mu_P = \sqrt{\frac{\omega(1-\omega)}{n}}$	$\mu_P = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$
2	Simple mechanical sampling	Not Applicable	$\mu_P = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$
3	Stratified (grouped) simple random sampling	$\mu_P = \sqrt{\frac{\overline{\omega}(1-\overline{\omega})}{n}}$	$\mu_P = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$
4	Stratified (grouped) mechanical sampling	Not Applicable	$\mu_P = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$
5	Systematic random sampling	$\mu_P = \sqrt{\frac{\delta_\omega^{\ 2}}{s}}$	$\mu_P = \sqrt{\frac{\delta_\omega^{\ 2}}{s}\left(1-\frac{r}{R}\right)}$
6	Systematic mechanical sampling	Not Applicable	$\mu_P = \sqrt{\frac{\sigma_\omega^{\ 2}}{s}\left(1-\frac{n}{N}\right)}$

In the given formulas, the average of the proportions (ω_j) within groups ($\bar{\omega}$) and the between-group variance ($\sigma_\omega{}^2$) are used:

$$\overline{\omega}_\mathrm{j} = \frac{\sum \omega_\mathrm{j} \mathrm{n}_\mathrm{j}}{\sum \mathrm{n}_\mathrm{j}},$$

$$\overline{\sigma_\omega{}^2} = \mathrm{w}(1 - \mathrm{w}) = \frac{\sum (\overline{\omega_\mathrm{j}} - \overline{\omega})^2 \mathrm{n}_\mathrm{j}}{\sum \mathrm{n}_\mathrm{j}}.$$

The confidence interval for the population proportion (P) is determined based on the sample proportion (ω) and its margin of standard error ($\Delta_\omega = t * \mu_\omega$). According to P.L. Chebyshev's theorem, the following inequality holds with probability P(t):

$$|m - \omega| \le \Delta_\omega.$$

From this, we obtain:

$$\omega - \Delta_\omega \le m \le \omega + \Delta_\omega$$

Or,

$$(\omega - \Delta_\omega) * 100\% \le m(\%) \le (\omega + \Delta_\omega) * 100\%$$

as the resulting inequalities.

Determining the Required Sample Size

Based on the margin of error (Δ_x) of the sample mean, the required sample size for random sampling is determined as follows:

If sampling is conducted with replacement, then:

$$t\sqrt{\frac{\sigma^2}{n}} \le \Delta_{\overline{x}},$$

From this, we obtain:

$$n \ge \frac{t^2 \sigma^2}{\Delta^2 \overline{x}}$$

This inequality indicates that the minimum required sample size must be:

$$n = \frac{t^2 \sigma^2}{\Delta^2 \overline{x}}$$

If sampling is conducted without replacement, then:

$$n = \frac{t^2\sigma^2 N}{\Delta^2_{\overline{X}} N + t^2\sigma^2}$$

Methods of Generalizing Sample Observations to the Population

The data from sample observations can be generalized to the population using two methods:

1) Recalculation Method

Suppose that in a sample observation, the confidence intervals for the mean ($\bar{x}$) and proportion (p) of a certain characteristic are determined as follows:

$\tilde{x} - \Delta_X \le \bar{x} \le \tilde{x} + \Delta_X$ and $\omega - \Delta_\omega \le p \le \omega + \Delta_\omega$.

Using these inequalities, the confidence intervals for the total population size (N), the sum of observed characteristic values (x) and the proportion of units (xN) in the population possessing the observed characteristic (pN) can be determined as:

$$\tilde{x}N - \Delta_X N \le \bar{x}N \le \tilde{x}N + \Delta_X N$$

and $$\omega N - \Delta_\omega N \le pN \le \omega N + \Delta_\omega N$$

These quantities' errors do not exceed Δ_xN and Δ_rN with probability P(t).

2) Coefficient Method

In some cases, complete enumeration data is verified using the sample survey method, and necessary adjustments are made accordingly.

Review Questions:

1. What are the different types of partial observation methods?
2. What is the sample survey methodology? How does it differ from other types of partial observation?
3. What are the main objectives and functions of sample observations?
4. In a market economy, which phenomena and processes are typically studied using sample surveys? Provide examples.
5. What do the terms population and sample refer to in statistics? How are they described using summary measures?
6. What is sampling error? How does it differ from recording error?
7. What sampling methods ensure the representativeness of a sample?

8. What is simple random sampling? What methods are used to implement it?
9. What are the stages of systematic sampling?
10. What is cluster sampling, and in what situations is it applied?

CHAPTER 7. STATISTICAL ANALYSIS OF RELATIONSHIPS

7.1. Relationships Between Phenomena and Processes, and Methods for Their Statistical Analysis

Various statistical methods have been developed to analyze the relationships between large-scale social phenomena and processes. The choice of a particular method depends on the research objectives and analytical requirements. When examining such relationships, variables are generally categorized into two main groups based on their role in the analysis:

When studying relationships, variables are generally categorized into two main groups based on their significance: 1) Factor (independent) variables, 2) Dependent (resulting) variables.

Factor variables influence changes in dependent variables, which vary in response to changes in factor variables.

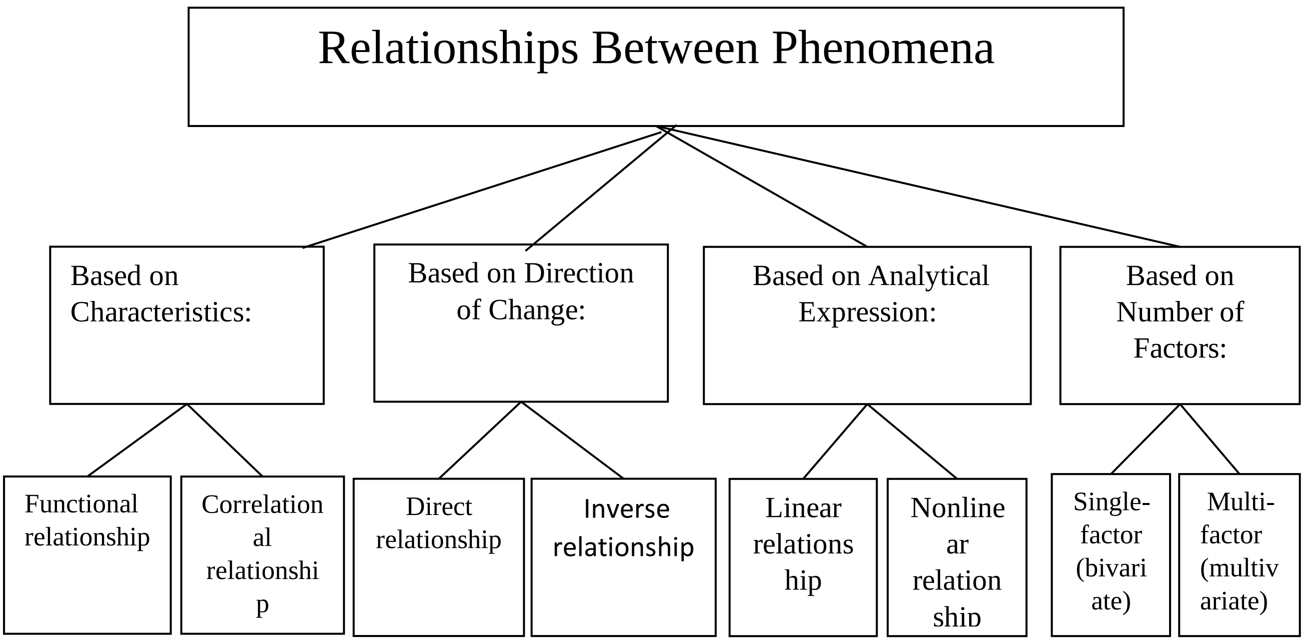


Figure 7.1. Types of Relationships Between Phenomena

Functional and Correlational Relationships

In a functional relationship, each value of the factor variable corresponds to exactly one value of the dependent variable. For example, the area of a rectangle is equal to the product of its base and height:

$S_{\square}$=a*h. In a *correlational relationship*, it is not possible to determine a complete list of all factor variables influencing the dependent variable. For instance, it is impossible to establish an exact list of all factors affecting crop yield.

If an increase (or decrease) in the factor variable leads to a corresponding increase (or decrease) in the dependent variable, the relationship is called a *direct relationship*. If an increase in the factor variable results in a decrease in the dependent variable, such a relationship is called an *inverse relationship*.

When represented on a coordinate plane, if the intersection points of the values of the factor and dependent variables form a pattern similar to a straight line, the relationship is classified as a *linear relationship*. If the points resemble the shape of a parabola or hyperbola, the relationship is referred to as a *nonlinear relationship*.

If the relationship between a single factor variable and a single dependent variable is studied, it is called a *bivariate relationship*. If the relationship between multiple factor variables and a single dependent variable is analyzed, it is referred to as a *multivariate relationship*.

To study functional relationships, methods such as the balance method, parallel series analysis, analysis of variance (ANOVA), regression-correlation analysis, and economic index methods are used. The balance method examines proportionality in relationships. In the analytical grouping method, classification is carried out based on a determining factor. Then, for each group, average and relative values describing the resulting indicators are calculated, and the obtained data are compared with changes in the dependent variable. In the parallel series method, the theoretical existence of a relationship between the compared variables is first determined. Then, both series are arranged in order (either increasing or decreasing) and placed side by side for comparison. For example, the relationship between labor productivity and wages.

7.2. Regression-Correlation Analysis

Regression and correlation are closely related: regression determines the form of a statistical relationship, while correlation measures the strength (tightness) of that relationship. Regression-correlation analysis involves determining the direction of relationships, expressing their analytical forms (formulas), and assessing their strength (intensity).

Regression correlational analysis is an important part of statistical analysis, which allows to determine connections between events and processes and draw conclusions based on them. In this method, the relationship between two or more variables is analyzed and it is possible to predict one variable using other variables.

Correlation analysis is used to determine the degree of relationship between two variables. The correlation coefficient (rrr) indicates the direction and strength of this relationship.

Correlation coefficient:

- r = 1: Positive complete correlation (when one variable increases, the other also increases).
- r =−1: Negative complete correlation (when one variable increases, the other decreases).
- r = 0: No correlation (no relationship between variables).

To apply regression-correlation analysis, the dataset must meet the following requirements:

- The average value of the dependent variable must be reliable and representative, accurately reflecting the overall trend without being distorted by errors, outliers, or biases.
- The dataset must be sufficiently large.
- The units in the dataset must be interrelated.
- The dependent variable must follow a normal distribution law for all factor values.

Regression equations are determined based on the type of relationship between events and phenomena:

1. The *linear regression equation* is calculated as follows: $\bar{Y}_x = a_0 + a_1 x$,
   here: a_0 – free term (intercept); a_1 – regression coefficient; X – independent variable (factor); Y – dependent variable (resulting factor); $\bar{Y}_x$- the average value of the dependent variable calculated

based on the independent variable. In turn, parameters a_0 and a_1 of the linear regression equation are calculated as follows:

$$\boldsymbol{a_0} = \frac{\sum \boldsymbol{Y} \cdot \sum \boldsymbol{X^2} - \sum \boldsymbol{XY} \cdot \sum \boldsymbol{X}}{\boldsymbol{n} \sum \boldsymbol{X^2} - (\sum \boldsymbol{X})^2}$$

$$\boldsymbol{a_1} = \frac{\boldsymbol{n} \sum \boldsymbol{XY} - \sum \boldsymbol{Y} \cdot \sum \boldsymbol{X}}{\boldsymbol{n} \sum \boldsymbol{X^2} - (\sum \boldsymbol{X})^2}$$

2. Curvilinear Regression Equations:
   a. Parabolic relationship: $\boldsymbol{Y_x}$=$\boldsymbol{a_0}$+$\boldsymbol{a_1 x}$+$\boldsymbol{a_2 x^2}$
   b. Hyperbolic relationship: $\boldsymbol{Y_x}$=$\boldsymbol{a_0}$+$\boldsymbol{\frac{a_1}{x}}$
   c. Exponential relationship: $\boldsymbol{Y_x}$=$\boldsymbol{a_0}$ $\boldsymbol{X^{a_1}}$

7.3. Determining the Strength of Relationships

To measure the strength (tightness) of relationships between phenomena, the following formulas are used:

Correlation Coefficient

$\boldsymbol{\tau} = \frac{\overline{\boldsymbol{XY}} - \overline{\boldsymbol{X}}\overline{\boldsymbol{Y}}}{\boldsymbol{\sigma_x} \cdot \boldsymbol{\sigma_y}}$ or $\boldsymbol{\tau} = \frac{\overline{\boldsymbol{XY}} - \overline{\boldsymbol{X}}\overline{\boldsymbol{Y}}}{\sqrt{(\overline{\boldsymbol{X}}^2 - (\overline{X})^2) - (\overline{\boldsymbol{Y}}^2 - (\overline{Y})^2)}}$ here, $-1 < \tau < 1$

The *correlation ratio* is calculated using the following formula: $\eta = \sqrt{\frac{\sigma^2 gr}{\sigma_{um}{}^2}}$

When a curvilinear relationship exists between variables, the strength of the relationship or the *correlation index* is calculated as follows:

$$R = \sqrt{1 - \frac{\boldsymbol{\sigma^2_{Y-Y_x}}}{\boldsymbol{\sigma^2_Y}}}$$

The *coefficient of determination* is calculated using the following formula:

$$\mathbf{D}=\frac{\boldsymbol{\sigma^2}_{\boldsymbol{Y_x}}}{\boldsymbol{\sigma^2_Y}}$$

In this formula: $\boldsymbol{\sigma^2}_{\boldsymbol{Y_x}}$ **-** represents the variation in the dependent variable caused by the studied independent factor 0<D<1,

$\boldsymbol{\sigma^2_Y}$ **-** represents the total variation in the dependent variable caused by all influencing factors.

The impact of unaccounted factors is calculated using the following formula:

$$R = \sqrt{\mathbf{1} - \frac{\boldsymbol{\sigma^2}_{Y-\bar{Y}_x}}{\boldsymbol{\sigma^2_Y}}}$$

Where $\sigma^2{}_{Y_x} = \frac{\sum(\bar{Y}_x - \bar{Y})^2}{n}$; $\sigma_Y^2 = \frac{\sum(Y-\bar{Y})^2}{n}$; $\sigma^2_{Y-\overline{Y_X}} = \frac{\sum(Y-\bar{Y}_x)^2}{n}$

The *Fechner coefficient* is calculated as follows: $F_x = \frac{M-H}{M+H}$

where - 1 $< F_x <$ 1 ;

M – The number of same-sign paired deviations, where both ($x - \bar{x}$) and (y- $\bar{y}$) have the same sign (either both positive or both negative);

N – The number of opposite-sign paired deviations, where both (x- x̅) and (y- y̅) have different signs (one positive and one negative).

Spearman's Rank Correlation Coefficient

$$\mathbf{P=1 - \frac{6\sum d^2}{n(n^2-1)}}$$

In this formula: *d* – The deviation between the ranked factor variable and the ranked dependent variable (results). d=$X_{tar} - Y_{tar}$; *n* – The number of ranked pairs.

Note: The term "ranked pair" (derived from the English word "range") refers to a pair consisting of the factor variable (X) and the dependent variable (Y).

Kendall's Coefficient

$$\boldsymbol{\tau = \frac{2\sum S}{n(n-1)}}\ \mathbf{; S=Q-P;}$$

where: Q– The number of positive ranked pairs, meaning values greater than Y_{tar};

P – The number of negative ranked pairs, meaning values smaller than Y_{tar}.

Concordance Coefficient

Concordance measures the strength of association among multiple variables and is calculated as: $\boldsymbol{\omega} = \frac{\mathbf{12S}}{\mathbf{m^2(n^3-n)}}$

where:

m – The number of factors (variables); n – The number of observations; S – The difference between the sum of squared rank totals and the squared mean of rank totals.

Table 7.1.

Calculation of the Concordance Coefficient

Number of companies	Value mln. s (X)	Number of issued shares (Y)	Busy departments in the enterprise (Z)	R_x X_{tr}	R_y Y_{tr}	R_z Z_{tr}	The sum of colors	Squares of sum of colors
1	2954	856	119	9	7	1	17	289
2	1605	930	125	1	9	2	12	144
3	4102	1563	132	10	10	3	23	529
4	2350	682	141	6	5	4	15	225
5	2625	616	150	7	3	5	15	225
6	1795	495	165	4	2	6	12	144
7	2813	815	178	8	6	7	21	441
8	1751	858	181	3	8	8	19	361
9	1700	467	201	2	1	9	12	144
10	2264	661	204	5	4	10	19	361
Total							165	2863

$S=2863-\frac{165^2}{10}=2863-2722{,}5=140{,}5$

$$\omega = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 140{,}5}{3^2(10^3 - 10)} = 0{,}19$$

Table 7.2.

Example of Weak Association Between Studied Variables

t/r	X	Y	Z	R_x	R_y	R_z	Sum up $\sum R_x+R_y+R_z$	The square of the sum
1	2954	856	119	4	3	1	8	64

2	1605	930	125	1	4	2	1	49
3	4102	1563	132	5	5	3	13	169
4	2350	682	141	2	2	4	8	64
5	2625	616	150	3	1	5	9	81
							45	427

Review Questions:

1. How are relationships between events classified?
2. What is a functional relationship? Provide an example.
3. What is a correlational relationship? Provide an example.
4. Based on their analytical representation, what types of relationships between events are categorized?
5. What is understood by correlation analysis?
6. What requirements must a dataset meet for regression-correlation analysis to be applied?
7. By which formula are linear and nonlinear regression equations calculated?
8. What is the essence of the correlation coefficient, and how is it calculated?
9. What is the significance of the determination coefficient?
10. How is the strength of relationships between events determined?

CHAPTER 8. METHODS OF STATISTICAL ANALYSIS OF TIME SERIES

8.1. Time Series: Essence, Classification, and Construction Rules

Social phenomena and processes are constantly in motion, evolving, and developing, with progress moving from simple to complex and from lower to higher levels. Statistics does not only examine social phenomena in their current state but also studies their changes over time.

In statistics, the change of social phenomena over time is referred to as dynamics, and the set of indicators describing this process is known as a *time series*. Any time series consists of the following two elements:

- A list of chronological moments (dates) or periods (years, months).
- Levels that describe the quantity, volume, or composition of the studied phenomenon.

The rules for constructing time series include:

1. Ensuring the comparability of indicators, meaning they must be expressed in the same unit of measurement and calculated using the same methodology across periods.
2. Maintaining territorial consistency, meaning that the observation object must remain the same in terms of geographical scope, and observation units must be uniformly defined.
3. Ensuring uniformity in terms of the length of the time periods considered and the level of precision in calculations (e.g., 0.1, 0.01, or 0.001 accuracy).

The results of statistical observation are expressed in two types of absolute values:

First type: Values representing the state of phenomena at a specific point in time. Examples include population size or the number of livestock at the beginning or end of a year. Time series based on such absolute values are called momentary time series.

Second type: Values representing the state of phenomena over a certain period. Examples include the number of births or the amount of harvested cotton within a given time frame. Time series based on such absolute values are called interval time series.

It is also possible to construct a *derivative time series* based on the original absolute series. Derived time series include relative and average values calculated from absolute data.

Key Differences Between Moment-Based and Period-Based Time Series

1. In moment-based time series, each level represents the state of a phenomenon at a specific time. In period-based time series, each level represents the cumulative amount of a phenomenon over a time interval. Hence, the level of a period-based time series depends on the length of the time interval.
2. The sum of levels in a moment-based time series does not carry economic meaning, whereas the sum of levels in a period-based time series has real significance and can be used to construct a new series.
3. The calculation of averages differs between moment-based and period-based time series.

8.2. Indicators for Analyzing Time Series

Time series are used to analyze changes in economic indicators over time. These lines show changes over a certain period of time and help to identify trends in economic events and processes. Time series are an important tool in economic analysis, forecasting and management decision-making.

There are several ways to classify dynamic lines:

1) By time intervals:
   - By calendar periods (annual, quarterly, monthly, weekly and daily)
   - At different time intervals (before and after a specific event);
2) According to the nature of the indicators:
   - Absolute indicators (quantity, volume);
   - Relative indicators (growth rate, share);
3) According to the structure:

- Simple dynamic series (one type of indicators)
- Complex dynamics series (several types of indicators)

The following rules must be followed to correctly compile the dynamic lines:

1. Information must be obtained from reliable sources and verified accurately.
2. The same time interval should be used for each period.
3. Indicators must be expressed in the same measurement units.
4. The data must be in a ready state for analysis, that is, invalid data must be removed and the necessary planes must be entered.
5. Data must be sequential and ordered.

The following steps are taken to create dynamic lines:

1. Collect the necessary data to create a dynamics series.
2. Data is sorted and arranged by time period.
3. Dynamics series are represented in the form of graphs and tables.
4. Data is analyzed and results are issued.
5. Based on the time series, future indicators are forecasted.

With the help of these rules and methods, the process of creating time series and analyzing them is carried out effectively.

Table 8.1.

The main indicators used in the analysis of time series

The name of the index	Comment
Average indicator	The average value of the indicators during the time series.
Average growth rate (%)	Growth rate of indicators in the average percentage form
Average annual growth rate (%)	Average annual growth rate
Average quantitative growth (Δ)	Average quantitative growth of indicators during the time series
Coefficient of variation	The ratio of the average square deviation of the indicators to the average value.
Trend	The general upward or downward trend of indicators over time.
Cyclicity	Regular or recurring changes over time

Seasonality	Regular change of indicators in different periods of the year
Absolute growth (Δ)	The absolute number of indicators increase during the time series
A chain pointer	The ratio of the indicator of each subsequent period to the previous period
Base indicator	The ratio of the indicator of each period to the indicator of the first period
Mean squared deviation	Deviation of indicators from the average value during the time series.
Dispersion	The mean of the squares of the deviations from the mean

Calculation of the average value of the indicators during the time series. This indicator is important for determining averages and understanding the general trend over time. Calculation of the growth rate in the average percentage form of the indicators during the dynamics series. This is used to understand how quickly metrics are increasing or decreasing. Calculate the average annual growth rate. This indicator is useful in determining the annual growth rate and forecasting in the future. Calculation of the average quantitative growth of indicators during the time series. This indicator shows the average increase or decrease for each year or period. The ratio of the standard deviation of the indicators to the average value. This indicator is used to evaluate the variability of the data.

Determining the general upward or downward trend of indicators over time. Trend analysis helps to understand long-term changes in indicators. Identify regular or recurring changes over time. Cyclical analysis helps to understand seasonal changes or economic cycles. Determining regular changes of indicators in different periods of the year. This indicator is used to determine seasonal effects. Calculation of the absolute number of indicators growth during the time series. This indicator shows the amount of increase or decrease in each period. The ratio of the indicator of each subsequent period to the previous period. Chain indicators are used for step-by-step evaluation of changes in indicators over time. The ratio of the indicator of each period to the indicator of the first period. This indicator is used to evaluate the change of data from the initial state. Calculation of the deviation of indicators from the average value during the time series. This indicator is important for determining the dispersion of data. The mean of the squares of the deviations from the mean. It is used to estimate the spread of dispersion indicators and the level of variability.

The analysis of time series is conducted to study the laws governing economic processes. For this purpose, the following indicators are calculated:

1) Absolute growth (or decline)
2) Growth (or decline) rate
3) Incremental growth (or decline)
4) Absolute value of 1% incremental growth (or decline)

This indicator is calculated by comparing the values of two different periods. Typically, the reference value for comparison is either the first observation in the series or the value from the previous year.

If each value (y_i) is compared with its preceding value (y_{i-1}) (i.e., year-to-year comparison), the resulting indicator is called a *chain-based index*. Conversely, if each value is compared only with the initial period value (y_0), the resulting indicator is referred to as a base-period index.

I. *Absolute Increment (Increase or Decrease)*

Absolute increment is determined by subtracting either the initial value or the preceding period's value from the current period's value:

$M_{change} = (Y_i - Y_0)$ for Base method;

$M_{change} = (Y_i - Y_{i-1})$ for Chain method.

Here, M_{change} represents the absolute increase or decrease.

II. *Growth (or Decline) Rate*

The growth or decline rate ($K_{g/d}$) indicates how many times a given period's value is greater or smaller compared to the initial or previous period's value. It is calculated as follows:

$K_{g/d} = (Y_i / Y_0)*100$ for Base method;

$K_{g/d} = (Y_i / Y_{i-1})*100$ for Chain method.

III. Relative Growth (or Decline) Rate

The relative growth (or decline) rate can be determined using two methods. It is calculated by subtracting the initial or previous period value from the current period value, multiplying the result by 100, and then dividing by the initial period value.

K_{rg} =(Y_i - Y_0)*100 /Y_0 for Base method;

$K_{rg'}$= (Y_i -Y_{i-1})*100/ Y_{i-1} for Chain method.

If growth or decline rates are already calculated, the relative growth (or decline) rate can be obtained using: $K_{rg} = K_{g/d} - 100$

To calculate the absolute value of a 1% incremental increase (or decrease), the chain absolute growth (or decline) value should be divided by the chain relative growth (or decline) rate.

$$\mathbf{A} = \frac{\text{Absolute Increment (or Decrease)}}{\text{Relative Growth (or Decline) Rate}} \quad \mathbf{or} \quad |1\%| = \frac{Y_{i-1}}{100}$$

Averages in Time Series Analysis

The calculation of the average value in time series primarily depends on the nature of the series. If the values in the time series are given at equal time intervals, the average level can be determined using the *simple arithmetic mean* formula as follows:

- Simple Arithmetic Mean (if observations are recorded at equal time intervals):

$$\overline{\boldsymbol{Y}} = \frac{(\mathrm{Y1+Y2+Y3+\cdots+Yn})}{n} = \frac{\sum \mathrm{Y}}{\boldsymbol{n}}$$

- Chronological Mean (for equidistant momentary time series):

$$\overline{\boldsymbol{Y}} = \frac{(\mathrm{Y1/2+Y2/2+Y3/2+\cdots+Yn/2})}{n-1}$$

- Average Absolute Increment:

$$\overline{\boldsymbol{M}} = \frac{\sum \mathrm{M}}{n}$$

- Average Growth Rate (calculated using the geometric mean formula):

$$\overline{\boldsymbol{K}}_{g/d} = \sqrt{K_1 * K_2 * K_3 * \ldots * K_n}$$

8.3. Methods for Processing Time Series

It is not always possible to directly determine the general patterns of a phenomenon based on empirical time series data. Such tasks are accomplished through methods of processing empirical time series data. These methods are categorized as follows:

1) Standardizing and linking time series;
2) Expanding time intervals;
3) Calculating moving averages;
4) Analytical smoothing of time series;
5) Applying interpolation;
6) Applying extrapolation.

1. Standardizing and Linking Time Series

To enable the comparison of different time series expressed in various measurement units, they are converted into relative series based on a common baseline. This is achieved by comparing each subsequent observation with the base period and expressing it as a percentage. Thus, different series are brought to the same reference point. A similar standardization effect can be achieved through the linking method.

2. Expanding Time Intervals

The expansion of time intervals refers to converting daily data into ten-day averages or monthly figures, monthly data into quarterly or annual values, and annual data into multi-year averages.

3. Calculating Moving Averages

The moving average method involves replacing actual time series values with their moving average equivalents. These averages can be computed over different periods to smooth out fluctuations.

4. Analytical Smoothing of Time Series

Analytical smoothing is conducted in the following steps:

a) Identifying the general trend and characteristics of the time series through economic analysis;

b) Selecting an appropriate mathematical equation based on the nature of the time series;

c) Estimating the parameters of the chosen equation using the least squares method;

d) Computing the smoothed series based on the selected equation.

5. Interpolation

Interpolation is the process of estimating unknown values within the range of a time series. This can be done using:

a) Averaging the two adjacent values;

b) Utilizing absolute or relative growth rates;

c) Applying additional or average incremental growth indicators.

6. Extrapolation

Extrapolation is used to estimate time series values for future (or past) periods. It is classified into two types:

a) Forward extrapolation (forecasting) – estimating unknown future values of the time series;

b) Backward extrapolation (retrospective analysis) – estimating missing historical values of the time series.

Additional Methods for Time Series Analysis

1. *Moving Average*

The moving average method smooths fluctuations in a data series by calculating the average value over a specified number of points. For example, a 3-day moving average is calculated as:

$$\text{Moving Average}_{t} = \frac{Y(t-1)+Y(t)+Y(t+1)}{3}$$

This technique helps to reduce short-term variations and highlight longer-term trends.

2. *Exponential Smoothing*

Exponential smoothing assigns greater weight to more recent observations, making it more responsive to recent changes. The formula is:

$$S_t = \alpha Y_t + (1 - \alpha)\, S_{t-1}$$

where α is the smoothing coefficient ($0 < \alpha < 1$). A higher α gives more importance to recent data, making the model more responsive to changes.

3. *Linear Trend Analysis*

Linear trend analysis examines whether a variable follows a consistent upward or downward pattern over time. The trend is represented by the equation:

$$Y_t = a + b_t$$

Where a is the intercept (starting value), b is the trend's growth rate, and t is the time period.

4. Non-Linear (Complex) Trends

More complex trends, such as quadratic or exponential trends, are used to model non-linear patterns in the data. For example, an exponential trend follows the formula:

$$Y_t = a \cdot e^{bt}$$

Where a is the initial value, and b represents the growth rate. These models capture accelerating or decelerating trends in time series data.

5. Seasonality Indicators

Seasonal decomposition methods divide a time series into repeating seasonal periods to identify seasonal patterns. Seasonality indices are calculated for each period to quantify recurring fluctuations in the data.

6. Regression Analysis

Regression analysis examines the relationship between time series variables, allowing for the prediction of future values based on explanatory factors. This method helps identify dependencies and trends in the data.

7. ARIMA Model (Autoregressive Integrated Moving Average)

The ARIMA model is a widely used forecasting technique that accounts for trends, autocorrelation, and moving average components in time series data. It is particularly effective in modeling time-dependent patterns and making short-term predictions.

8. First Differencing

The first differencing method transforms a time series into a stationary series by calculating the difference between consecutive observations:

$$Y_t^1 = Y_t - Y_{t-1}$$

This approach is useful for stabilizing data and removing trends.

9. Autocorrelation Coefficient

The autocorrelation coefficient measures the correlation between a time series and its past values at different time lags. It helps identify repeating patterns, dependencies, and the presence of cyclical behavior in the data.

8.4. Forecasting in Business

It is well known that economic conditions change over time. Therefore, managers must be able to forecast (predict) these conditions that affect their companies. One of the methods that ensures precise planning is forecasting. Despite the variety of methods available, all forecasting techniques are aimed at predicting future events. Since a company's development plans and strategies must account for these changes, forecasting plays a crucial role in decision-making.[3].

In modern society, the need for forecasting remains constant. For example, government officials must forecast unemployment rates, inflation, industrial output, and tax policies to implement appropriate economic strategies. An airline executive must predict passenger demand to plan workforce and technological investments. University or college administrations need to estimate the number of students enrolling in the coming year to ensure adequate housing and facilities. Forecasting methods allow analysts to predict future conditions based on past trends and historical data.

Categories of Quantitative Forecasting

Quantitative forecasting methods can be categorized into two main approaches:

- Time Series Analysis – Examining past data to identify patterns and trends over time.
- Causal forecasting analysis – Investigating relationships between different factors and their impact on forecasting accuracy.

Approaches to Forecasting

There are two main approaches to forecasting:

[3] David M.Levine, David F.Stephan. Statistics for managers. 2014, PHI Learning Private Limited, Delhi-110092. 610-p.

- Qualitative Forecasting
- Quantitative Forecasting

If the researcher lacks numerical data, *qualitative forecasting* methods (e.g., expert judgment, market research) can be used. These methods are subjective in nature. On the other hand, if historical data about the research subject is available, *quantitative forecasting methods* should be applied.

Time Series Analysis

A time series is a sequence of numerical data collected at successive time intervals. Time series analysis enables forecasting based on past and present numerical variables. For example, daily stock prices on the New York Stock Exchange form a time series. Other examples include the monthly Consumer Price Index (CPI) and quarterly Gross Domestic Product (GDP) measurements.

Causal Forecasting Analysis

Causal forecasting methods analyze cause-and-effect relationships influencing the variable being predicted. These methods help identify factors that impact the forecasted values. Examples of causal forecasting methods include: Multiple regression analysis, Econometric modeling, and Leading indicator analysis.

Trend Identification and Smoothing Techniques

If the long-term trend of income growth exhibits significant fluctuations, it may be difficult to visually identify a trend on a graph. In such cases, moving averages or exponential smoothing techniques are used to eliminate short-term variations.

The *moving average method* is influenced by the choice of period length (L), which determines the number of data points used to compute each average. This method helps reduce cyclical fluctuations, and to ensure accurate trend estimation, the period length should be an integer. A moving average for a given period is denoted as MA(L).

Calculating a Moving Average

Suppose we have data covering 11 years and need to compute a 5-year moving average (L = 5). The moving averages are calculated as follows:

The first 5-year moving average is obtained by summing the values of the first five observations and dividing by 5:

$$\text{MA (5)} = = \frac{\text{Y1+Y2+ Y3+Y4+Y5}}{5}$$

This process continues until the moving average is determined for the last 5 years. The computed moving average for a given year should be positioned at the center of the selected time interval. If the total number of years is 11, then L = 5.

In addition to moving averages, the exponential smoothing method is also widely used to identify long-term trends characterizing data variations. This method allows for making short-term forecasts, which gives it an advantage over moving averages.

The exponential smoothing method derives its name from the process of exponentially weighting moving averages. Each observation in the time series is influenced by previous observations, with more recent data points receiving greater emphasis.

One key advantage of exponential smoothing is that it inherently accounts for recent trends while gradually diminishing the influence of older observations. Unlike moving averages, this method does not require storing a fixed number of past values but instead assigns exponentially decreasing weights to historical data points. In exponential smoothing, the assigned weight (denoted as W) determines the importance of recent observations. As time progresses, older data points lose significance, meaning frequently occurring values receive higher weights, whereas less frequent values are assigned lower weights. For computational purposes, Microsoft Excel can be used to perform exponential smoothing efficiently.

The equation enables the smoothing of time series data over a given time interval i. It consists of three key components: Current observation (Yi): This represents the observed value in the time series at time i. Previous exponential smoothing value (Ei-1): This is the smoothed value from the preceding time period. Weight factor (W): This is the smoothing coefficient that determines the influence of past observations on the current smoothed value. The smoothing equation is expressed as:

$$E_i = WY_i + (1-W)E_{i-1}, \qquad i=2,3,4$$

Where E_i is the exponentially smoothed value at time i, E_{i-1} is the smoothed value from the previous period, Y_i is the observed value at time i and W is the smoothing coefficient, constrained within $0 < W < 1$. This formula ensures that more recent observations have a greater impact on the smoothed value, while older values gradually lose influence over time.

The selection of the smoothing coefficient or weight is crucial, as it directly affects the outcome. Unfortunately, this selection is somewhat subjective in nature. If the researcher wants to eliminate unwanted cyclical or random fluctuations from the time series, they should choose a small weight (close to zero). On the other hand, if the time series is being used for forecasting, a larger weight (W)—one that is closer to one—should be selected. Forecasting is of great importance for enterprises in making strategic decisions, in efficient allocation of resources and in being competitive in the market.

1. Qualitative methods

Expert opinions: in this method, forecasts are made based on the opinions and experiences of qualified experts. Experts assess future trends based on their knowledge and experience.

Delphi method: a group of experts exchange opinions anonymously and come to a common forecast through repeated evaluations several times. This method is used to include alternative opinions and reach consensus.

2. Quantitative methods

Time Series Analysis: future indicators are predicted based on changes in time series. In this method, smoothing, trend and non-seasonal components are identified.

Regression Analysis: in this method, relationships between variables are determined and future values are determined. Forecasts are made based on the regression equation.

Extrapolation: future values are projected based on the trends of indicators in the past periods. Extrapolation is used to continue trends in time series. Forecasting in business is an important tool for predicting future economic and financial performance. Forecasts are made using qualitative and quantitative methods, and these results are of great importance in making strategic decisions. The forecasting process includes the stages of data collection, method selection, model building, verification, forecasting and monitoring. This process plays a major role in the effective operation of the enterprise and success in market competition.

Review questions

1. What do you understand by time series?
2. What elements constitute a time series?
3. How are time series classified based on their essence and representation?
4. Provide the formulas for calculating the average level of a time series.
5. How are the indicators for analyzing time series calculated?
6. How are the average annual growth rate and incremental growth rate determined?
7. In what cases are time series brought to a common base?
8. How is the calculation of growth trends in time series performed?
9. What do you understand by smoothing in time series analysis?
10. Explain the concepts of interpolation and extrapolation.

CHAPTER 9. STATISTICAL INDEXES

9.1. The Essence, Importance, and Classification of Indexes

The term "index" is commonly encountered in everyday life, such as postal codes or the index of newspapers and journals. The word "index" originates from Latin and means "indicator" or "sign." Indexes are used in mathematics, economics, and many other disciplines. In statistics, an index is understood as a relative indicator that reflects the relationship of a particular phenomenon across different regions or periods or compares current data with planned or projected values. In international practice, specific indices are denoted by the letter "i," while general indices are denoted by the letter "I." When calculating indices in statistics, the current (reporting) period is designated by the number 1, while the base (previous) period is denoted by the number 0. Indexes serve the following functions:

- Identifying the dynamics of mass social phenomena over two or more periods;
- Calculating the dynamics of average economic indicators;
- Determining the ratio of indicators across different regions;
- Converting macroeconomic indicators into comparable values.

Indices are relative indicators used to monitor, compare, and analyze changes in economic and financial indicators. Indices allow the comparison of indicators in different periods or different places. They are used to determine trends in economic events and processes, as well as to assess general changes.

The importance of indices is reflected in the following factors:

- Indices simplify complex economic phenomena and provide easy understanding.
- They help identify changes and trends over time.
- Comparability - makes it possible to compare performance between different periods or locations.

Indices are used in the formation of the state's economic policy and evaluation of its results. Financial indices play an important role in investment portfolio management and financial decision-making.

Indexes are classified into several types based on their various characteristics:

1. *By type of indicators:*

— Price indices: These indices track price changes of goods and services. For example, the consumer price index (CPI) and, wholesale price index (WPI).

— Volume indices: These indices track changes in production or trade volumes.

— Wage indices: These indices track changes in wage levels.

— Financial indices: These indices track changes in the stock market. For example, the Dow Jones Industrial Average (DJIA), S&P 500, and FTSE 100.

2. *By time:*

— Average indices represent the average value for a certain period.

— Periodic indices represent a separate value for each period.

— According to the method of calculation: simple indices represent the average change of indicators in a certain period.

— Compositional indices are calculated according to the shares of its components.

Several methods are used to calculate indices:

Laspeyres index:

In this method, an index is calculated based on the amounts in the base period.

$$I_L = \frac{(\sum (P1 \times Q0))}{(\sum (P0 \times Q0))} \times 100$$

where P1 – prices in period t, P0 – prices in the base period, Q0 – quantities in the base period.

Paasche Index:

In this method, an index is calculated based on the amounts of the current period.

$$I_P = \frac{(\sum (P1 \times Q1))}{(\sum (P0 \times Q1))} \times 100$$

where P1 – prices in period t, P0 – prices in base period, Q1 – quantities in period 1.

Fisher index:

It is calculated as the geometric mean of Laspeyres and Paasche indices.

$$I_F = \sqrt{I_L * I_P}$$

Indices are an important tool for monitoring, analyzing and comparing changes in economic and financial indicators. They simplify complex economic phenomena, help identify trends, and enable comparisons between different periods or locations. Indices are of great importance in formulating public policy and making investment decisions.

The following conventional symbols are used in constructing statistical indexes:

Table 9.1.

Conventional Symbols Used in Constructing Statistical Indices

Indicator	Sign	Periods	
		Base	Current
1. Product volume	q	q_o	q_1
2. Product price	p	p_o	p_1
3. Production cost price	c	c_o	c_1
4. Total labor expenditure for production, number of registered workers	T	T_o	T_1
5. Time spent per unit of production (person-day, person-hour, etc.)	t	t_o	t_1

Statistical Indices Are Classified as Follows:

1) By the Scope of Coverage: a) Individual indices, b) General indices
2) By the Basis of Comparison: a) Dynamic indices, b) Regional indices,
3) By Composition: a) Fixed-composition index. b) Variable-composition index, c) Structural shift index
4) By Calculation Period: a) Annual indexes, b) Quarterly indexes, c) Monthly indexes, d) Weekly indexes.

9.2. Individual and General Indexes

Individual indexes are derived by comparing events or phenomena of the same type. For example, the individual index of a product's price is determined by dividing the current period price of a specific product by its base period price.

Individual indices are used to measure changes in a specific economic indicator or event. These indices allow for comparison and analysis of a single indicator.

Consumer Price Index (CPI) tracks changes in the prices of consumer goods and services.

The Wholesale Price Index (WPI) tracks changes in wholesale market prices.

The production index tracks changes in the volume of industrial production.

The average wage index tracks changes in wage levels.

Stock indices track changes in share prices on the stock market. For example, Dow Jones Industrial Average (DJIA), S&P 500, FTSE 100.

General indices are used to evaluate the general economic situation by combining the changes of several indicators. These indices cover various sectors of the economy and allow for the analysis of broader indicators. Including:

— It combines the prices of various goods and services to track overall changes in price levels.

— It observes the general changes in production volumes in various sectors, such as industry, agriculture, and the service sector. Monitors the increase or decrease of national income.

— Monitors general changes in economic activity in various sectors of the economy.

The main types of individual indices include:

$i_q = \frac{q_1}{q_0}$ - Index of the physical volume (quantity) of produced or sold goods;

$i_p = \frac{p_1}{p_0}$ - Price index;

$i_c = \frac{c_1}{c_0}$ - Cost price index;

$i_{pq} = \frac{p_1 q_1}{p_0 q_0}$ - Value (trade turnover) index;

$i_t = \frac{t_0}{t_1}$ - Labor productivity index.

General indices include:

$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}$ - Physical volume index;

$I_P = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ - Price index;

$I_c = \frac{\sum c_1 q_1}{\sum c_0 q_1}$ - Cost price index;

$I_{Pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$ - Value index;

$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1}$ - Labor productivity index.

The numerator and denominator of general indexes consist of the sum of the product of two quantities, where typically one of the quantities changes (the indexed quantity), while the other remains constant in both the numerator and denominator (the index weight). The indexed quantity refers to the characteristic being studied for its variation. The index weight is a quantity used to bring the indexed quantity into a common measure for comparison. When selecting the index weight, the following rule must be followed:

— If an index is being constructed for a quantitative indicator, the base period is used as the weight.
— If an index is being constructed for a qualitative indicator, the current period is used as the weight.

Qualitative indicator indexes include: Product price index, Product cost price index, Time expenditure index.

Let us discuss the index of product value (trade turnover), the index of physical volume (quantity) of products, and the index of product price. The trade turnover index indicates how many times the value of products in the current period has increased or decreased compared to the base period. If we subtract the denominator from the numerator of the fraction, we obtain the absolute change in product value in the current period compared to the base period (the increase or decrease in monetary terms):

$$\Delta_{qp}= \Sigma q_1 p_1 - \Sigma q_0 p_0$$

In the physical volume index of a product, the indexed quantity is the product expressed in natural measurement units, while the weight is the price. Different types of products become comparable only when multiplied by their respective prices. The price from the base period is used as the weight. In the numerator of the fraction, the conditional value of the current period's product volume is expressed at the base period's price. In the denominator, the actual value of the base period's product volume at the base period's price is represented. The quantity index shows how many times (or by what percentage) the product value has increased or decreased due to changes in the product quantity. If we subtract the denominator from the numerator of the fraction, we obtain the absolute change in product value due to the change in product quantity (in monetary terms).

$$\Delta q= \Sigma\, q_1 \cdot p_o - \Sigma\, q_o \cdot p_o$$

$$I_p = \frac{\Sigma\, p_1 \cdot q_1}{\Sigma p_0 \cdot q_1} \quad \Delta p=\Sigma\, p_1 \cdot q_1 - \Sigma p_0 \cdot q_1,$$

The following relationship exists between these indices:

$$I_{qp}=I_q \cdot I_p \;\; \text{or} \;\; \frac{\Sigma\, q_1 \cdot p_0}{\Sigma q_0 \cdot p_0} \cdot \frac{\Sigma\, p_1 \cdot q_1}{\Sigma p_0 \cdot q_1} = \frac{\Sigma\, q_1 \cdot p_1}{\Sigma q_0 \cdot p_0}.$$

In statistics, along with general indexes, average indexes are also widely used. An average index can be constructed in the form of either an arithmetic mean or a harmonic mean. Regardless of the calculation method, the average index yields the same result as the aggregate index. An average index must be derived from a general index. To achieve this, the indexed indicator in the numerator or denominator of the general index must be replaced with its identity value derived from the corresponding individual index.

$$I_q = \frac{\sum P_0 q_1}{\sum P_0 q_0} = \frac{\sum i_q P_0 q_0}{\sum P_0 q_0}$$

Thus, the average arithmetic index of the physical volume of a product is determined using the formula for the arithmetic mean, where individual indexes are weighted by the product values of the base period.

The dynamics of average indicators should not only reflect changes in the characteristic being averaged but also account for structural changes within the dataset. To study this, we use variable composition, fixed composition, and structural shift indexes.

a) Fixed composition index of product price:

$$I_p = \frac{\sum P_1 q_1}{\sum q_1} : \frac{\sum P_0 q_1}{\sum q_1}$$

b) Variable composition index of product price:

$$I_{\bar{p}} = \frac{\sum P_1 q_1}{\sum q_1} : \frac{\sum P_0 q_0}{\sum q_0} = \bar{p}_1 : \bar{p}_0$$

c) Structural shift index of product price:

$$I_{s/s} = \frac{\sum P_0 q_1}{\sum q_1} : \frac{\sum P_0 q_0}{\sum q_0}$$

The following relationship exists between a) fixed composition index, b) variable composition index, and c) structural shift index:

$$I_{\bar{p}} : I_p = I_{s/s}$$

An index is a multifaceted concept that is applied in various fields, serving specific purposes. In statistics, this term is used in the sense of a complex comparative economic indicator. In general terms, an index is derived by examining an economic phenomenon in two different states and measuring them using a special method, resulting in a comparative indicator. This result can be expressed in different forms: relative, average, or absolute values, as well as their unified representation.

An economic process occurs between two states of a phenomenon, leading to development. Indices serve as a measure of this development, expressing the relative, average, and absolute changes as a whole. For comparison, the states of a phenomenon can be examined from different perspectives, revealing various aspects of the development process. This includes changes over time, inter-object and interregional or international comparisons, the degree of fulfillment of plans, contracts, or economic norms, as well as internal shifts within economic structures. This provides the theoretical and methodological foundation for classifying indices into types such as dynamic indices, territorial indices, international indices, plan or contract performance indices, and structural change indices. Additionally, indices can also be classified based on other characteristics, such as: Coverage of statistical units, Structural form, Calculation method, etc. As a result, a complex, multi-tiered classification system of indices is formed.

For indices to have economic significance and interpretation, the underlying indicators must possess objectivity and materiality. Otherwise, they remain abstract, absolute numbers with no meaningful context. However, interpreting this fundamental requirement narrowly, by superficially ensuring that indicators have the same unit of measurement, is incorrect. Indices gain economic meaning only when they reflect real changes in phenomena under specific conditions and with the necessary degree of accuracy. Only in this case do they meet the primary requirement of economic analysis.

Although individual, unweighted, and fixed-weight general indices are structurally closer to relative measures, they differ in essence. This is because, in addition to capturing relative changes, they also allow for the determination of average and absolute changes, adhering to the requirements of objectivity and materiality.

These properties are not characteristic of variable-basis or currently weighted general indices, such as those constructed using the Laspeyres and Paasche methods. As a result, the theory of index tests emerged, which establishes that the above-mentioned properties should serve as key criteria in constructing ideal indices. Indices that meet these criteria are known as Fisher indices, named after the founder of this theory, Irving Fisher. Fisher indices are calculated as the geometric mean of the

Laspeyres and Paasche aggregate indices, ensuring compliance with the fundamental properties required for an ideal index.

The primary form of general indices involves the calculation of aggregate indices. A commonly promoted methodological approach in statistical textbooks and manuals suggests using the Paasche method (current weights) for qualitative indicators and the Laspeyres method (base-period weights) for quantitative indicators. However, this recommendation lacks both theoretical and practical justification. In any aggregate index, one of the variables in the numerator or denominator does not represent an actual, observable economic measurement but rather a hypothetical, computed value. This means that its substantiveness, materiality, and economic reality are conditional. The economic meaning of an index is not determined by which period's values are used as weights but rather by the context, relationships, and aspects of economic development it aims to measure. From this perspective, each index serves as a fundamental measure of change, allowing the determination of relative, average, and absolute variations. At the same time, every index has both theoretical and practical strengths and limitations, making the choice of index dependent on the specific economic analysis being conducted.

Review questions

1. What do you understand by the term "index"?
2. What are the functions of the index method?
3. What types of indices do you know?
4. What problems can be solved using indices?
5. What processes can be studied using the index method?
6. What is meant by an individual index?
7. What types of individual indices do you know?
8. How are the general indices of product value, product price, and product quantity (physical volume) calculated?
9. Provide information about average indices.
10. How are the indices of variable composition, fixed composition, and structural shifts in product prices calculated, and what is the relationship between them?

CHAPTER 10. POPULATION STATISTICS

10.1 . Population and Its Statistical Study

Demographic statistics is one of the oldest branches of statistics. Population counts have been conducted for military purposes and tax assessment since ancient times. The earliest statistical regularities identified in demographic studies were related to birth and mortality rates. Today, the population is a key subject of comprehensive research, as it (particularly the economically active population) directly participates in production processes and serves as a consumer of goods and services. Interest in demographic statistics continues to grow. Demographic statistics include the study of population size and composition within a defined geographic area, which is continuously affected by births, deaths, immigration, and emigration. The demographic composition of any country is heterogeneous and evolves over time. Therefore, population dynamics, structural changes, and other characteristics must be analyzed within their specific historical and socio-economic context. In population statistics, the unit of observation is typically an individual, though in certain cases, it may be a family. In international statistical practice, households are also considered observational units. A household is defined as a group of individuals who reside together and share economic resources (not necessarily related by kinship). Unlike a family, a household unit can also be a single person who is financially independent.

In demographic statistics, the statistical unit of observation can belong to various population groups, including:

1. The total population (either the resident population or the de facto population);
2. Specific population subgroups, such as the economically active population and the unemployed, urban and rural populations, or male and female populations, among others. Additionally, young families and elderly families, as well as individuals born or

deceased within a given year, can also be considered statistical units of observation.

During a population census, the total number of individuals in a given geographic area is determined at a specific reference point (critical moment). The population is categorized into the following four groups:

1. De facto population (DFP): This refers to all individuals who are physically present in the given area at the time of the census, regardless of whether they are permanent or temporary residents.
2. De jure population (DJP): This includes individuals who are officially registered as permanent residents of the given area, regardless of whether they are physically present at the time of the census.
3. Temporary residents (TR): These are individuals residing in the area for a period of up to six months due to specific reasons such as vacation, medical treatment, professional training, or education.
4. Temporary absentees (TA): These are individuals who have left the area for a period of up to six months for similar reasons.

$$DJP = DFP + TA - TR$$

$$DFP = DJP + TR - TA$$

Population statistics is a branch of science dealing with the collection, processing, analysis, and presentation of data on population size, structure, dynamics, and territorial location. Demographic phenomena and processes are studied with the help of population statistics.

The tasks of population statistics include:

- Collection and analysis of data on the population and its changes.
- Studying demographic indicators such as age, gender, nationality, marital status.
- Studying the location of the population in different regions and analyzing migration processes.
- Studying the socio-economic characteristics of the population, such as employment, income, and education level.

Methods of population statistics include methods of collecting and analyzing demographic data. The following main indicators are used in the analysis of population statistics:

The number of people living in a certain area.

The number of children born in a certain period.

The number of people who died in a certain period.

The difference in birth and death rates.

The difference between the number of people entering and leaving the area.

Ratio of working age population to young children and elderly.

Population statistics are important from the following aspects:

Determining the needs of the population and planning social services.

Distribution of economic resources according to population size and structure.

Improving population health and planning the health care system.

Planning educational institutions according to population and age structure.

Forming state policy and making decisions based on information about the population.

Population statistics is an important field of science for studying the demographic, economic and social characteristics of society. This area provides complete information about the number, structure, dynamics and territorial location of the population and is of great importance in socio-economic planning. With the help of population statistics, demographic processes and events are studied, and information necessary for the development of various areas of society is provided.

10.2. Statistical Analysis of Population Size and Composition

In macroeconomic statistics, the study of social capacity (resources) begins with an analysis of population size. The main sources for determining population size include: Periodic population censuses; data on natural population change, including births and deaths; data on migration movements, including immigration and emigration.

Natural population changes—such as births, deaths, marriages, and divorces—are recorded by civil registration offices. Migration movements, which include individuals relocating permanently to or from a specific country, region, or district, are recorded in urban household registration systems (for entries and exits) and in rural administrative records, such as household registers maintained by local councils.

Based on the processing of these records and previously conducted population censuses, annual estimates of the total population at the end of the year (A_1) for the entire country, as well as for individual regions, districts, cities, and rural areas, are calculated using the following formula:

$$A_1 = A_0 + (A_{born} - A_{deaths}) + (A_{arrivals} - A_{left}),$$

Where A_0 population at the beginning of the year; A_{born} = number of births during the year; $A_{dealths}$ = number of deaths during the year; $A_{arrivals}$ = number of immigrants (people who moved into the area for permanent residence during the year); A_{left} = number of emigrants (people who moved out of the area for permanent residence during the year).

Such calculations can also be performed based on place of residence, gender, and age groups. Through censuses and current demographic accounting, population size is typically determined as of the beginning of the year. However, population figures recorded on a specific date may not always be suitable for macroeconomic indicators. For example, indicators such as gross domestic product (GDP) per capita, national income, and gross income are not calculated based on the population at the beginning or end of the year but rather on the average annual population.

The average population size is calculated using different formulas depending on the reference periods or dates for which population data is provided:

1. If population figures are given for the beginning and end of the year, the average population is calculated using the simple arithmetic mean formula:

$$\overline{A} = \frac{A_0 + A_1}{2}$$

Where: $\overline{A}$ - average population, A_0 - population at the beginning of the year, A_1 - population at the end of the year.

2. If population figures are provided as a time series at equal intervals, the average population is determined using the chronological mean formula:

$$\overline{A} = \frac{\frac{1}{2}A_1 + A_2 + A_3 + ... + \frac{1}{2}A_n}{n-1}$$

3. If population figures are provided as a time series with unequal intervals, the average population is calculated using the weighted arithmetic mean formula:

$$\overline{A} = \frac{\Sigma \overline{A} \cdot t_i}{\Sigma t_i}$$

where $\overline{A}$ - the average of two consecutive population counts, t_i - the time interval (in months, days, etc.).

Statistics examines the composition of the population based on various characteristics, including general demographic indicators (sex, age, marital status), ethnic composition (nationality, native language, citizenship), labor force characteristics (place of employment, occupation, specialization), social structure (social group, living conditions), educational and cultural level (education level, participation in education, academic degree), and geographical distribution (urban or rural residence, region, district, etc.).

The ratio of men to women in the population and their age distribution significantly influence marriage rates and serve as a fundamental demographic factor in population growth. According to statistical data, on average, there are 105–106 male births per 100 female births, a biological pattern observed across all countries worldwide.

Population refers to the number of people living in a certain period or area. This indicator is the main parameter for population statistics. The following methods are used to determine the population. It will be necessary to enumerate the entire population and collect information on their demographic, economic and social characteristics.

Conducting observations on a certain part of the population (sample) and summarizing the results.

Continuous monitoring and recording of population demographic changes.

The composition of the population is the distribution of the population according to demographic, social, and economic indicators. Key demographic indicators include:

- Distribution of the population by age groups.
- Distribution of the population by the number of men and women.
- Distribution of the population by married, single, divorced and widowed status.
- Distribution of the population by nationality.
- Distribution of the population by level of education.
- Distribution of the population according to whether they are employed or not.

The following methods are used for statistical analysis of population size and structure:

1. *Descriptive Statistics*: Describing and summarizing information about the population.
2. *Analytical Statistics*: Determining and analyzing the relationships between demographic processes and events.
3. *Analysis of birth and death rates*: Analysis of natural population growth or decline.
4. *Analysis of migration:* study of territorial movement of the population and its causes.
5. *Analysis of age structure*: study of population distribution by age.

The following main indicators are used in the analysis of population statistics:

The number of people living in a certain area.

The number of children born in a certain period.

The number of people who died in a certain period.

The difference in birth and death rates.
The difference between the number of people entering and leaving the area.
Ratio of working-age population to young children and the elderly.

Demographic, economic and social characteristics of the society can be studied by statistical analysis of the number and composition of the population. With the help of these analyses, it is possible to obtain complete information about the number, structure, dynamics and territorial location of the population and to make decisions necessary for socio-economic planning.

The main source of population statistics is the current account (deed records and arrival and departure certificates) and sample observations. The current account of demographic processes is the current account of births, deaths, arrivals, departures, recorded marriages and divorces. Statistical data on the natural movement and migration of the population are collected on a monthly basis. Initial data is systematically calculated on a quarterly basis. The calculated results are observed and taken into account in statistical processes. Accuracy and reliability parameters are used in calculating the population on annual and quarterly basis.

The initial data are sufficiently close to the required parameters regarding statistical coverage, classifier, estimation and specific period of calculation. Births, deaths, marriages and divorces must be recorded in the Ministry of Justice's Register of Civil Status Records. The absolute number of processes is obtained on the basis of the processing of data from the second copies of records of birth, death, marriage and divorce records. The current registration of migrants is assigned to the Migration and Citizenship Departments of the regional Internal Affairs Departments (departments) of the Ministry of Internal Affairs, which carry out registration of residents by address of residence and address of migration. Statistical observations are taken into account in case of a change of permanent residence. The International System of Diseases and Statistical Classification of Health Problems is used to keep records of the dead by causes of death.

Data on the population covers the permanent residents of the Republic of Uzbekistan (without exceptions). The current calculation of the population is based on the results of the last population census, by adding to it the number of people born and immigrated to the area each year, and subtracting the number of deaths and people who moved from the area. The statistical methods used correspond to the basis of statistical processes and allow to reduce errors in data processing to a minimum. The process of data preparation brings errors to a minimum in the processing process, that is, in the processes of reducing errors in data coding, editing and preparation of tables.

10.3. Indicators of Population Movement

Population movement indicators encompass two distinct groups of measures. The first group includes natural population movement, which consists of birth rates, mortality rates, marriage, and divorce statistics. The second group involves migration indicators, which track population inflows and outflows due to migration.

Population size constantly changes due to births, deaths, and migration. The natural movement of the population refers to changes in population size resulting from births and deaths, independent of migration. Key statistical indicators describing natural movement include birth rates, which measure the number of births within a given period, and mortality rates, which track the number of deaths. Additionally, natural population growth is determined by the difference between the number of births and deaths, while marriage and divorce rates serve as important indicators of social and demographic trends. Absolute figures for births, deaths, and natural population growth are typically reported for a specific time period, providing a fundamental basis for demographic analysis and forecasting. However, absolute figures alone do not reflect birth, death, and natural change rates relative to the total population. To accurately assess the level of natural movement, relative indicators expressed per 1,000 people (‰ or promille) are used. One such measure is the birth rate coefficient (Kt), which is determined by dividing the number of births (B) by the average annual population (P). This allows for a standardized comparison of birth rates across different population sizes and regions, providing a clearer understanding of demographic trends.

$$K_t = \frac{A_{tug'}}{\overline{A}} * 1000\%_0$$

Table 10.1.

Indicators of population movement

№	Indicators	Calculation methods	Explanation
I. Natural Movement Indicators			
1.	Birth Rate Coefficient (K_t)	$K_b = \frac{A_{born} \cdot 1000}{\overline{A}}$	A_{born}-number of births $\overline{A}$-average number of populations
2.	Mortality Rate Coefficient ($K_{o'}$)	$K_d = \frac{A_{deaths} \cdot 1000}{\overline{A}}$	A_{deaths}-the number of dead people
3.	Absolute Natural Change (ΔNCh)	$\Delta NCh = A_{born} - A_{deaths}$	
4.	Natural Change Coefficient (K_{nv})	$K_{nv} = \frac{(A_{born} - A_{deaths}) \cdot 1000}{\overline{A}}$ or $K_{nv} = K_b - K_d$	
5.	Specific Birth Rate Coefficient (K_{br})	$K_{br} = \frac{A_{born} \cdot 1000}{\overline{W'}}$	W'-Average number of women aged 15-49
6.	Mortality rate of children under 1 year of age (K_{dch})	$K_{dch} = \frac{m \cdot 1000}{\frac{2}{3}N_1 + \frac{1}{3}N_0}$	m-Number of infants who died before 1 year in the current year; N_0 and N_1-the number of births previous and current year
II. Mechanical Movement Indicators			

7.	In-Migration Coefficient (K_a) (arrival rate)	$K_a = \frac{A_{\text{arrivals}} \cdot 1000}{\overline{A}}$	A_{arrivals}-the number of people who came to this settlement for permanent residence
8.	Out-Migration Coefficient (K_l) (leave rate)	$K_l = \frac{A_{\text{left}} \cdot 1000}{\overline{A}}$	A_{left} - the number of people who left from this settlement
9.	Net Migration Coefficient (K_{mp})	$K_{mp} = \frac{(A_{\text{arrivals}} - A_{\text{left}}) \cdot 1000}{\overline{A}}$ or $K_{mp} = K_a - K_l$	
10	Absolute Net Migration (ΔMCh)	$\Delta MCh = A_{\text{arrivals}} - A_{\text{left}}$	

The Vitality Coefficient is calculated by dividing the number of births by the number of deaths:

$$K_h = \frac{A_{\text{born}}}{A_{\text{deaths}}}$$

This indicator can also be determined by dividing the birth rate by the death rate:

$$\frac{K_b}{K_d} > 1$$

If the resulting ratio is greater than 1, it indicates natural population growth in the given area. Conversely, if the ratio is less than 1, it signifies population decline in that area.

Thus, an increase in the vitality coefficient leads to population growth, while a decrease results in population decline.

10.4. Estimating the Future Population Size

Planning the long-term macroeconomic outlook begins with forecasting population changes. This is because key planning indicators must be aligned with societal needs, particularly those of the population. Estimating future population growth involves analyzing demographic structural changes, predicting natural population growth and migration trends at the regional level, determining urban-to-rural population ratios, assessing future labor force composition, and forecasting demographic shifts over the next 15–20 years.

A market economy does not only function based on current or near-term interests but also considers the long-term impact of population and labor force growth based on the current productive capacity of economic resources. This enables policymakers to anticipate the socioeconomic consequences of demographic changes in advance.

The future population size is estimated using the global method and the cohort component (age-shifting) method.

In the global method, calculations are based on the initial population size (A0) at the beginning of the year and either the natural growth rate (K_{NV}) or the overall population change coefficient as key input parameters. The estimation follows the formula:

$$A_1 = A_0 \cdot \left(1 + \frac{K_{nv}}{1 - \frac{1}{2}K_{nv}}\right) \quad A_2 = A_1 \cdot \left(1 + \frac{K_{nv}}{1 - \frac{1}{2}K_{nv}}\right)$$

Here, A_1 and A_0 represent the population at the end of the current year and the beginning of the next year, respectively. K_{vn} - coefficients of natural change (or general change).

Let's examine the application of this method with the following hypothetical example: Initial population at the beginning of the year = 118,000 people, Natural change coefficient = 17‰ (per thousand). The population of the city after two years is calculated as follows:

1. $A_1 = A_0 \cdot \left(1 + \frac{K_{vn}}{1 - \frac{1}{2}K_{vn}}\right)$

$$A_1 = 118 \cdot \left(1 + \frac{0{,}0170}{1 - \frac{1}{2}0{,}0170}\right)$$

A_1=118(1+0,01701)

A_1=118x1,0171

A_1=120 thousand people

Δ1=120-118=2 thousand people

2. $A_2 = A_1 \cdot \left(1 + \frac{K_{vn}}{1 - \frac{1}{2}K_{vn}}\right)$

$$A_2 = 120 \cdot \left(1 + \frac{0{,}0170}{1 - \frac{1}{2}0{,}0170}\right)$$

A_2=120•(1+0,0172)

A_2=120·1,0171

A_2=122 thousand people

Δ_2=122-120=2 thousand people

In this calculation, the mechanical change coefficient has been disregarded. While this method is useful for estimating the future population, it does not allow for separate calculations based on age groups. Additionally, the rate of natural or overall growth, which is taken as the base value, may vary for each subsequent year, leading to differences in projected population changes.

In the age-shifting method, the future population is estimated using the following formula:

$$i_x \cdot P_x = i_{x+1}$$

i_x = Population at age x

P_x = Probability of surviving at age x

i_{x+1} = Population that is likely to reach age x+1

Thus, the population at each age (i_x) is "shifted" to the next age (i_{x+1}) using the survival probability coefficient (Px).

Let's examine this method using sample numerical data. Based on this information, we will estimate the future population for three consecutive generations over the next 1, 2, and 3 years (i.e., N, N+1, and N+2 years). To solve this, we will complete the last columns of the table.

Table 10.2.

Age-Shifting Method for estimating the future population.

Initial Population at the Start of the Year	Projected Population at the Beginning of the Next Year	Projected Population at the Beginning of Year 2	Projected Population at the Beginning of Year 3
i_{17} i_{18}	$i_{18} = i_{17} \cdot P_{17}$ $i_{19} = i_{18} \cdot P_{18}$	$i_{19} = i_{18} \cdot P_{18}$ $i_{20} = i_{19} \cdot P_{19}$	$i_{20} = i_{19} \cdot P_{19}$ $i_{21} = i_{20} \cdot P_{20}$

i_{19}	$i_{20}= i_{19} \cdot P_{19}$	$i_{21}= i_{20} \cdot P_{20}$	$i_{22}= i_{21} \cdot P_{21}$
$i_{17}+ i_{18}+ i_{19}$	$i_{18}+ i_{19}+ i_{20}$	$i_{19}+ i_{20}+ i_{21}$	$i_{20}+ i_{21}+ i_{22}$

This table demonstrates how each age group's population is progressively shifted to the next age group using age-specific survival probabilities.

Table 10.3.

The procedure for calculating the prospective number of the population using the age shift method

Age	Age Group Population at the Beginning of the Year	Probability Coefficient of Survival for Each Age	Estimated Population at the Beginning of the Year for the Next Age Group		
			N	N+1	N+2
			$i_{x+1}=i_x \cdot P_x$		
17	3786	0,99641	-	-	-
18	3764	0,99600	3786·0,99641 = 3772	-	-
19	3742	0,99562	3764·0,99600 = 3749	3772·0,99600=3757	-
20	-	0,99526	3742·0,99562 = 3726	3749·0,99562=3732	3757·0,99526=3740
21	-	0,99503	-	3726·0,99526=3708	3732·0,99516=3714
22	-	0,99498	-	-	3708·0,99503=3689

So, out of the 3,786 individuals aged 17 in the current year:

By year N: $i_{18}= i_{17} \cdot P_{17}$

By year N+1: $i_{19}= i_{18} \cdot P_{18}$

By year N+2: $i_{20}= i_{19} \cdot P_{19}$

i_{18}=	i_{19}=	i_{20}=
3786·99641=3772	3772·99600=3757	3757·99562=3740

Similarly, the population for other age groups has been calculated for the corresponding years.

Review Questions:

1. What is the object of study in population statistics, and what are the sources for calculating the population size?
2. Into which categories is the population divided during a census?
3. How is the average annual population calculated?
4. How are the indicators of natural population movement calculated?
5. How are the indicators of mechanical population movement calculated?
6. How are the special coefficients of population movement calculated?
7. What process does the population viability coefficient represent?
8. What is the significance of expressing population movement in absolute and relative indicators?
9. Provide information on calculating the future population size.
10. How is population forecasting carried out using the cohort (age-shifting) method?

CHAPTER 11. LABOR MARKET STATISTICS

10.1. Economically Active and Inactive Population

The permanent population of a country is divided into two groups: economically active and economically inactive populations. The economically active population offers its labor force to the labor market, where employers purchase it. In contrast, the economically inactive population cannot participate in the labor market due to objective reasons.

Population surveys (censuses) play a crucial role in collecting data on employment and unemployment. These surveys cover all population groups, providing comprehensive information on employment structure and its changes, the number of unemployed individuals, and various forms of labor activity that are not reflected in traditional statistical accounting methods.

Labor market statistics encompass a standard set of indicators, including the total labor force, distribution of employed individuals across economic sectors, workforce composition and size in enterprises, organizations, and institutions, utilization of working hours, labor productivity, wage levels and dynamics, and working conditions across different industries.

During the planned economy period, data on labor resources and employment were derived from population censuses conducted at intervals, as well as from mass statistical reports of enterprises, organizations, and institutions. Additionally, selective labor force surveys conducted for the study of wages and employment provided supplementary insights.

In the transition to a market economy, profound socio-economic transformations have imposed new demands on statistical analysis. Labor market research has become a key focus of statistics, including the measurement of employment and unemployment rates, vocational training for displaced workers and associated costs, and the study of forced migration trends. These challenges necessitate the alignment of Uzbekistan’s statistical methodologies with international standards.

The following new data sources are utilized to study this process:

— Statistical reports from employment service agencies, which reflect job search applications submitted by citizens, temporary job vacancies, and registered unemployment figures.
— Quarterly labor force surveys conducted in the last week of each quarter to assess employment levels among the population.
— Data from migration service agencies regarding foreign nationals employed in Uzbekistan’s national economy.

Comprehensive data on the number, composition, and distribution of labor resources can be obtained from population census records.

The following relative indicators are calculated to assess the state of the labor market:

- Working-age population coefficient – represents the share of the working-age population within the total population.
- Labor-force participation coefficient (Kt) – represents the proportion of the labor force within the working-age population.
- Employment coefficient – represents the proportion of employed individuals within the total population.
- Working-age employment coefficient – represents the proportion of employed individuals within the total working-age population.
- Labor resource employment coefficient – represents the proportion of employed individuals within the total labor resources.
- Total dependency ratio – represents the number of dependents (non-working-age individuals) per 1,000 working-age individuals.
- Labor resource replacement ratio – represents the number of children and adolescents (under 16 years old) per 1,000 working-age individuals.
- Pension burden ratio – represents the number of pensioners per 1,000 working-age individuals.
- Unemployment rate – the ratio of unemployed individuals to the economically active population, expressed as a percentage.

- Economic activity rate – represents the share of the economically active population within the total population.

The economically active population consists of individuals who are capable of working based on their age and health condition. According to the existing law in our country, the working-age population includes men aged 16 to 59 and women aged 16 to 54. Within this age group, individuals are classified based on their health status into two categories: employable and non-employable. The non-employable group includes unemployed individuals with disabilities classified as Category I and II, as well as retirees of working age who do not work and receive pensions under preferential conditions.

Labor resources refer to individuals who are both of working age and health condition and are actively engaged in the economy. They constitute a portion of the total permanent resident population. They include the following:

- Individuals of working age who are capable of working;
- Adolescents under the age of 16 who are actively employed;
- Individuals who are older than the official working age but continue to work;
- Foreign nationals who are actively employed in the country.

The number of labor resources (LR) is determined as follows:

$$LR = PW - (D + UP) + PT;$$

PW - labor resources, people of working age;

D - group I and II disabled people;

UP - unemployed population;

PT - employed pensioners and working teenagers under the age of 16.

When calculating labor resources for specific regions, commuter migration is also considered.

The economically active population is classified based on employment status. The economically active population (A_{eap}) consists of the sum of the employed population (A_{ep}) and the unemployed population (A_{up}):

$$A_{eap} = A_{ep} + A_{up}$$

The coefficient for this indicator is calculated using the following formulas:

1. Economically active population coefficient (K_{eap})

$$K_{eap} = \left(A_{eap} : \overline{\text{A}}\right) \cdot 100$$

2. Economically inactive population coefficient (K_{eip})

$$K_{\text{eip}} = \left(A_{\text{eip}} : \overline{\text{A}}\right) \cdot 100$$

The economically inactive population includes the following categories:

a) Students and cadets enrolled in full-time educational institutions;

b) Non-working individuals classified as Group III disabled;

c) Individuals engaged in household activities or caring for sick relatives;

d) Unemployed individuals earning income from personal property or real estate;

e) Individuals voluntarily choosing not to work, regardless of their economic status.

3. Employment coefficient (K_{er}),

$$K_{er} = \left(A_{er} : \text{A}_{\text{eap}}\right) \cdot 100 \text{ or } K_{er} = 100\% - \text{K}_{\text{up}}$$

4. Unemployment rate (K_{up}),

$$K_{up} = \left(A_{up} : \text{A}_{eap}\right) \cdot 100 \text{ or } K_{up} = 100\% - K_{er}$$

5. Coefficients characterizing the demographic burden of the working-age population:

   a) Potential replacement coefficient:

   $$K_{pch} = \left(A_{0-15} : \text{A}_{wap}\right) \cdot 1000\%_0,$$

   A_{0-15} - population under 0-15 years of age;

   A_{wap} –the number of working-age population.

   b) Pensioner burden coefficient:

   $$K_{p} = \left(A_{p} : \text{A}_{\text{wap}}\right) \cdot 1000\%_0,$$

   c) Overall dependency ratio

   $$K_{glf} = \left(\left(A_{0-15} + A_{p}\right) : \text{A}_{\text{wap}}\right) \cdot 1000 \text{ or } K_{glf} = K_{pch} + K_{p}$$

11.2. Composition, Workforce Size, and Mobility Indicators of Enterprise Employees

Economic activity refers to human efforts directed at producing material and intangible goods. The composition of the workforce is analyzed based on several categories:

- industry classification (primary and auxiliary activities),
- role in the work process (managers, specialists, employees, and workers), demographic indicators (such as gender and age),
- demographic indicators (gender, age);
- specialization,
- skill level;

Workforce statistics can be used to determine the average number of registered employees, new hires, and the actual workforce over a specific period. The average monthly headcount of registered employees ($\overline{H}$) is calculated by summing the number of registered employees (H) for each day of the month and dividing it by the number of calendar days in that period (K_k).

$$\overline{H} = \sum H : K_k \text{ or } \overline{H} = \sum H_{kk} : K_k$$

ΣH_{kk} - Total sum of employee-days worked and non-worked during the month, ΣH - is the total number of registered employees during the period, K_k – Number of calendar days in the period.

To determine the average number of employees, present at work, the total number of employees who reported to work (or the total employee-days worked) during the month (or period) is divided by the total number of working days (period). The formula is as follows:

$$\overline{E}_{came} = \Sigma E_{came} : P_{working.days}$$

Here, $P_{\text{working days}}$ = Number of working days in the period.

To determine the average number of employees who actually worked ($\bar{E}_{act.worked}$), the total number of employees ($\sum E_{act.worked}$) who effectively worked (excluding those who reported to work but remained idle) is divided by the total number of working days:

$$\overline{E}_{act.worked} = \sum E_{act.worked} : P_{working.days}$$

$E_{act.worked}$ - Number of employees who reported to work but remained idle the entire day.

The total number of enterprise employees fluctuates due to employee turnover, which includes new hires and terminations. Statistics examine workforce changes, i.e., workforce dynamics, using the following indicators:

1. Employee Dismissal Rate (K_{edr}) – This coefficient is calculated by dividing the number of employees dismissed during the reporting period ( ) by the average number of employees on the payroll during the same period ($\overline{H}$)

   $K_{edr} = H_{dismiss} : \overline{H}$
2. Employee Hiring Rate (K_{hiring}) – This coefficient is determined by dividing the number of employees hired during the reporting period (E_{hiring}) by the average number of employees on the payroll during that period:

   $$K_{hiring} = E_{hiring} : \overline{E}$$
3. Turnover Rate ($K_{turnover}$) – This coefficient is calculated by dividing the number of employees dismissed either voluntarily or due to disciplinary violations ($E_{dismiss}$) by the average number of employees on the payroll during the reporting period:

   $$K_{turnover} = E_{dismiss} : \overline{E}$$
4. Employee Retention Rate (K_{ret}) – This coefficient is obtained by dividing the number of employees who worked throughout the entire period (year) (Err) by the total number of employees at the end of the period (or the average number of employees on the payroll):

   $$K_{reten} = \frac{E_{reten}}{\overline{E}}$$
5. Overall Workforce Turnover Rate ( ):

   $$K_{\Sigma wtr} = (E_{hiring} – E_{dismiss}) : \overline{H}$$

The working time of enterprise employees is typically measured in man-hours and man-days. A man-day is calculated by multiplying the number of employees by the number of days worked. Similarly, a man-hour is determined by multiplying the number of employees by the total hours worked.

To assess the efficiency of labor utilization, the following working time fund indicators are determined:

1. Calendar Working Time Fund – This is calculated by multiplying the total number of employees on the payroll (or the average number of employees on the payroll during the period) by the total number of calendar days in the period.
2. Recorded Working Time Fund – This fund is derived by subtracting the number of public holidays and rest days from the calendar working time fund.
3. Maximum Possible Working Time Fund – This is obtained by subtracting the number of days allocated for scheduled annual leave from the recorded working time fund.

To determine working time utilization coefficients, the actual man-days worked are divided by the corresponding working time fund values.

Indicators of Working Time Utilization

1. Average Workday Duration – Determined by dividing the total man-hours worked by the total man-days worked.
2. Average Number of Workdays per Employee – Calculated by dividing the total man-days worked by the average number of employees on the payroll.
3. Average Number of Work Hours per Employee – Obtained by dividing the total man-hours worked by the average number of employees on the payroll.

One of the key indicators of production organization efficiency is the shift coefficient of employees, which is calculated by dividing the total man-days worked during the reporting period by the number of man-days worked at full shift capacity.

11.3 Labor Productivity Statistics

Labor productivity refers to the volume of output produced per unit of time or the amount of time required to produce a single unit of output.

The level of labor productivity is determined by the ratio of the total volume of produced goods to the time spent on production, expressed as $\omega = q/T$, or as the time spent per unit of output, expressed as $t = T/q$. These two indicators have an inverse proportional relationship: $t = 1/\omega$ and $\omega = 1/t$.

The dynamics of labor productivity are analyzed using indices, which are divided into individual and aggregate indices. To calculate labor productivity using indices, the following notations are used:

ω - labor productivity; q - volume of production; T - total time spent (number of workers); t - time spent per unit of production; P - price per unit of production; d - share of each indicator relative to the total.

Labor productivity in the production of a single type of product is determined using individual indices:

$$i_\omega = \omega_1 : \omega_0 = (q_1/T_1):(q_0/T_0)$$

This index can also be constructed using the inverse indicator. However, since labor demand (t) is inversely related to the amount of product produced per unit of time (ω), the index is formulated as follows:

$$i_t = t_0 : t_1 = (T_0/q_0):(T_1/q_1)$$

Labor productivity for enterprises (associations) producing the same type of product is calculated using the natural index as follows:

$$I_\omega = \omega_1 : \omega_0 = \left(\sum q_1 / \sum T_1\right):\left(\sum q_0 / \sum T_0\right)$$

If the product is measured in a conditional natural unit, then it is analyzed using the conditional natural index as follows:

$$I_{\omega k} = \omega_{Ik} : \omega_{Ok} = \left(\sum {}_{q1k} / \sum T_1\right):\left(\sum q_{ok} / \sum T_0\right)$$

Here, k is the coefficient for converting to a conditional natural unit.

When different types of products are produced, the level of labor productivity is calculated using the value method as follows:

$$\omega = \frac{\sum_{i=1}^{n} q_i p_i}{\sum_{i=1}^{n} T_i},$$

Here:

q - the quantity of product types;

P - the price per unit of product.

The index of labor productivity in the production of various types of products is analyzed using value indicators.

$$I_{\omega} = \frac{\sum q_{i1} p_{i0}}{\sum T_{i1}} : \frac{\sum q_{i0} p_{i0}}{\sum T_{i0}}$$

Here:

$\Sigma q_{i1} p_{io}$ and $\Sigma q_{i0} p_{io}$ – the volume of products at constant prices in the current and base periods; T_{i1} and T_{i0} – the time spent (average number of employees) in production during the current and base periods.

In enterprises producing a single type of product, the following aggregate indices of labor productivity are determined:

The variable composition index of labor productivity:

$$I_{\mathrm{var}(\overline{w})} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum W_{i1} T_{i1}}{\sum W_{i0} T_{i0}} = \frac{\sum W_{i1} d_{i1}}{\sum W_{i0} d_{i0}}$$

Here,

d_{i0}, d_{i1}– the shares of labor time expenditure in the base and current periods $(T_i : \sum_{i=1}^{n} T_i)$;

$\overline{w}_0, \overline{w}_1$– the average levels of labor productivity in the base and current periods;

W_{i0}, W_{i1}– the levels of labor productivity of the enterprise in the base and current periods.

This index is influenced by two factors:

a) Changes in the level of labor productivity in individual enterprises;

b) Changes in the share of labor time expenditure.

To separately determine the impact of each of these factors, fixed (constant) composition and structural shift indices are calculated.

The change in average labor productivity due to the first factor is characterized by the fixed composition labor productivity index.

$$I_{const(w)} = \frac{\sum W_{i1} T_{i1}}{\sum W_{i0} T_{i1}} = \frac{\sum W_{i1} d_{i1}}{\sum W_{i0} d_{i1}}$$

The change in average labor productivity due to the second factor is determined using the structural shift index.

$$I_{struct} = \frac{\sum W_{i0} T_{i1}}{\sum W_{i0} T_{i0}} = \frac{\sum W_{i0} d_{i1}}{\sum W_{i0} d_{i0}}$$

The following relationships exist between these indices:

$$I_{\mathrm{var}(\overline{w})} = I_{const(w)} \times I_{struct}$$

Review Questions:

1. What is the object of study in labor statistics?
2. How are the economically active and inactive populations determined?
3. Explain the concept of labor resources.
4. How is the demographic dependency ratio of the working-age population calculated?
5. What are the main indicators of unemployment statistics?
6. How is the number of employees in an enterprise (registered, recorded, etc.) calculated?
7. What is the significance of employee turnover coefficients?
8. How is the employee instability coefficient of an enterprise calculated?
9. What is labor productivity, and how is its individual index determined?How is the index analysis of labor productivity conducted?

CHAPTER 12. NATIONAL ACCOUNTING AND THE SYSTEM OF MACROECONOMIC INDICATORS

12.1 . Differences and Similarities Between the National Economy Balance and the System of National Accounts

The core essence of national accounting is determined by the system of macroeconomic balances. The System of National Accounts (SNA) is a system of interrelated indicators, classifications, and accounts used for recording and analyzing the results of economic activities (production, consumption, savings) at the macro level.

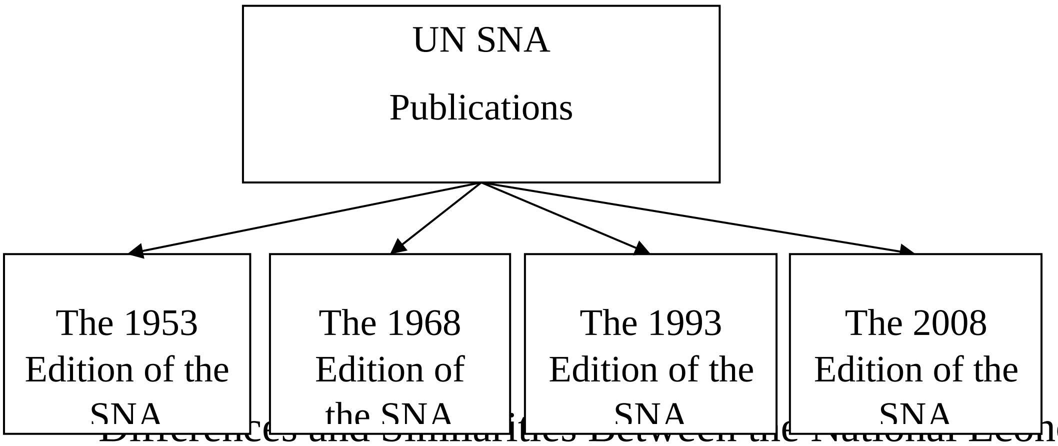
The SNA is a balance of the economy that conforms to international standards adopted in developed countries. The calculations in the SNA are made based on the principle of double-entry bookkeeping. The creation of the SNA was influenced by the global economic crisis of 1929-1933. It was proven in practice that under the conditions of a market economy, the market mechanism, based solely on the freedom of commodity producers and the competition among them, was not capable of leading a country out of crisis. During this period, the economic theory of the great British economist John Maynard Keynes emerged: To stabilize the economy, external forces controlling the market mechanism must influence supply and demand. This force is the state. For the state's macroeconomic policy to correspond to supply and demand, there needed to be a clear and comprehensive source of information about the state, development, and interrelations of the economy. This source took the form of the SNA.

In the 1930s, the SNA, initially created in the Netherlands, then in Germany and France, spread rapidly across Europe, Asia, and the Americas. However, differences in the methodology for calculating national account indicators and discrepancies in the number of calculations made international comparisons difficult. Therefore, creating international SNA standards became an urgent task. The first international standard for the System of National Accounts by the United

Nations (UN) was created in 1953. In 1968, the second standard was established by the UN. In 1969, based on this SNA, the first system of integrated European economic indicators was approved for the member states of the European Union. In 1993, the scientific work in this field was consolidated, and the third international standard for the System of National Accounts was created by the UN. In 1995, based on this SNA, the second system of integrated European economic indicators was established.

The 1993 United Nations System of National Accounts consists of 700 pages, 21 chapters, and nearly 600 calculations. These calculations reflect the following:

1) Standard calculations for domestic economic sectors;
2) Calculations for economic sectors;
3) Calculations for types of economic operations: (commercial, finance, credit, insurance, and other economic activities are divided into two types: a) market operations: based on monetary transactions; b) non-market operations: based on barter without the use of money);
4) Calculations for foreign economic relations;
5) Calculations that comprehensively describe the entire economy.



Differences and Similarities Between the National Economic Balance (NEB) and the System of National Accounts (SNA)

During the Soviet era, the National Economic Balance (NEB) was used to describe and analyze the economy.

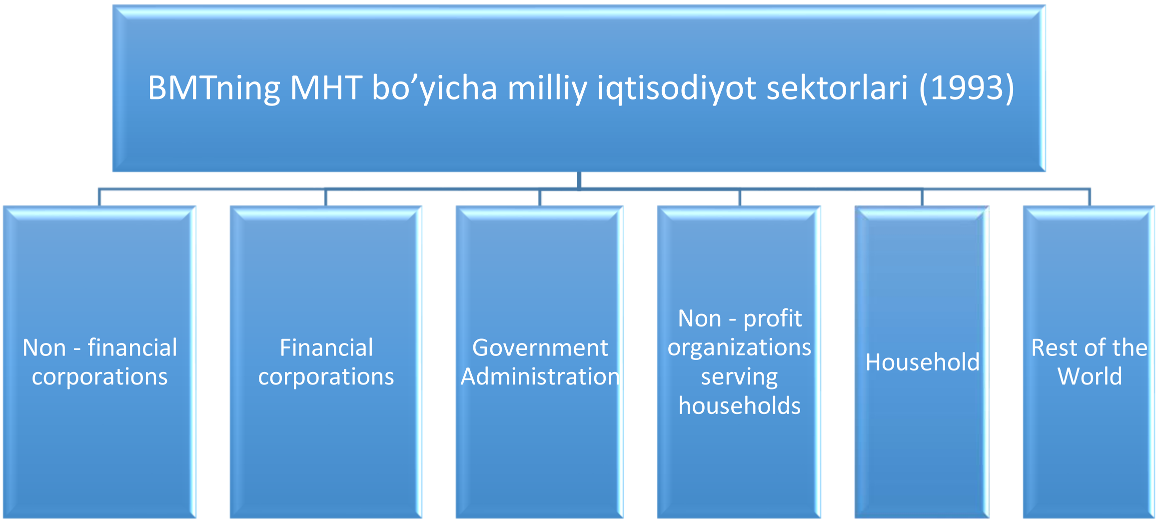
Table 12.1.

Differences Between the System of National Accounts (SNA) and the National Economic Balance (NEB)

National Economic Balance (NEB)	System of National Accounts (SNA)
1. Social property.	1. Various forms of property
2. Centralized management.	2. Self-management
3. Gross Domestic Product (GDP) is only created in material production sectors.	3. GDP is the sum of the value of goods produced and services provided
4. The economy is classified into two sectors: a) Material goods-producing sector b) Non-material goods-producing sector	4. The economy is classified into two sectors: a) Goods-producing sector b) Services sector

Figure 12.2

Sectors of the National Economy According to the UN SNA



Similarities Between the System of National Accounts (SNA) and the National Economic Balance (NEB):

1) Both systems recognize the processes of material production and the provision of services.
2) Both systems operate based on the principle of double-entry bookkeeping.

The SNA quantifies the indicators related to the production of material goods and services, the formation, distribution, and redistribution of income, and the use of income. It also characterizes macroeconomic processes related to foreign economic activity and their outcomes.

Characteristics of the System of National Accounts of the Republic of Uzbekistan

Currently, the Republic of Uzbekistan uses the 3rd version of the System of National Accounts (SNA) published by the United Nations in 1993. This system, while taking into account the specific features of Uzbekistan's economy, expresses the following types of economic operations:

1) Transactions involving goods and services;
2) Distribution operations;
3) Financial operations.

At the macro level, the calculations, schemes, and sectoral accounts of the national economy are consistent and interrelated from start to finish. In each calculation, the sums of economic operations are balanced using balancing items. These balancing items serve as the transition mechanism, enabling the shift from one calculation to another through responsible economic analysis.

12.2. Key Concepts and Categories of the SNA

Key Concepts in SNA:

1) *Marshall's "Economic Production" Concept*: The creation of income is not limited to material production sectors but also includes the services sector.
2) *J. Sey and Marshall's "Factors of Production" Concept*: The income generated through production belongs to all factors of production: land, capital, labor, and entrepreneurship.
3) *J. Hicks' "Income" Concept:* Income refers to the ability of a production unit to use its capital during a certain period and retain its maximum value at the end of that period.
4) *V. Leontief's "Economic Circulation" Concept*: Economic circulation comprises all production processes (production,

distribution, redistribution, final consumption, and savings processes).

Categories of SNA:

1) *Function (Objective):* The production of goods and services, the distribution of income, and their use for final consumption and saving.
2) *Economic Operations*: Cooperation between economic units in producing goods and services, distributing income, redistributing it, acquiring financial assets, and incurring debt obligations.
3) *Objects of Economic Operations*: Goods, services, money, financial documents, etc.
4) *Economic Units*: Economic agents involved in the cooperation of producing goods and engaging in trade relations.
5) *Economic Agents:* Independent economic decision-makers who make decisions regarding the economic activities they undertake.

Key Concepts in SNA:

- *Economic Area*: A national geographic area, including all territories where individuals, goods, and financial units move freely. This includes national space, territorial waters, mineral resources, and the country's enclaves abroad (embassies, military bases, scientific bases), excluding the enclaves of other states within the country.
- *Domestic Economy*: Refers to the economic activities and results of both residents and non-residents within the country's economic territory.
- *National Economy*: Refers to the economic activities of residents, regardless of their place of residence.
- *Residents*: Institutional units that are considered part of the economic territory of the country, engaged in economic activities for at least a year, regardless of their nationality or citizenship.
- *Economic Units* (Economic Agents): Legal and physical persons who manage their assets and conduct economic activities independently and are responsible for their accounts.

- *Non-residents*: Units that are considered non-residents of a given country, including foreign embassies, enclaves, tourists, seasonal workers, artists, and athletes who are in the country for less than a year.

12.3. J. Hicks’ Theory of Income

Different definitions of income can be found in economic literature. For example, in 1930, I. Fisher defined income as the final use. In 1898, the Dutch statistician F. Boss defined income as the sum of final use and the net increase of capital asset reserves, emphasizing that capital is the collection of resources with market value. American economist Samuelson suggested that land, natural resources, and other non-produced assets, when directed towards economic activity by labor, transform into capital. According to Lippe, income is the sum of monetary earnings that regularly come to economic units and influence the demand for money.

According to this theory, the income is the largest amount that can be used for consumption during a certain period, and at the end of the period, the value of the capital at the beginning of the period is maintained.

In the 1993 System of National Accounts (SNA), based on Hicks' theory of income, income is defined as the largest sum, which, when used for consumption during a certain period, allows the capital at the end of the period to maintain the value it had at the beginning.

According to Hicks' theory, the system of income indicators includes the following:

1) Primary income;
2) Value-form current transfers;
3) Discretionary income (1q + 2q);
4) Social transfers in kind;
5) Adjusted discretionary income (3q + 4q);
6) Gross National Income (GNI) = Gross Domestic Product (GDP) + ΔD;

7) Gross National Disposable Income (GNDI) = GNI + Δj.t. (Δj.t. – represents change in a specific economic indicator over time)

12.4. Relations Between Macroeconomic Indicators

The following relations exist between macroeconomic indicators:

1) Gross Domestic Product (GDP);
2) Gross Fixed Capital Formation (GFCF);
3) Net Domestic Product (NDP) = GDP - GFCF;
4) Balance of income ΔD = A - B;
5) Gross National Income (GNI) = GDP + ΔD;
6) Net National Income (NNI) = GNI - GFCF;
7) Balance of current transfers received and paid from foreign countries Δc.t;
8) Gross Disposable National Income (GDNI) = GNI + Δc.t;
9) Private Consumption (PC);
10) National Savings (NS) = GDNI - PC;
11) Balance of capital transfers Δcap;
12) Financing sources for investment (ISF) = NS + Δcap;
13) Gross Savings (GS);
14) Purchases of non-financial intangible assets;
15) Net Credit (or Net Debt) 15q = 12q - 13q - 14q.

Here, A represents the income earned by foreign residents in the given country; B represents the income earned by residents of the given country from foreign countries.

12.5. Other Income Indicators in SNA

Other income indicators in the SNA include the following:

1) Primary income;
2) Disposable income;
3) Adjusted disposable income.

Adjusted disposable income is divided into real final consumption and savings.

Income is categorized as follows:

- Nominal income (gross) NI = MBOD + MKD, MBOD = income from labor,

MKD = income from property;

- Net income (net) NI = NI + T_t - S,

  Tt = income from transfer payments,

  S = taxes and social insurance contributions;

- Real income (RI) represents the purchasing power of money income and serves as an aggregate indicator of the population's standard of living. Real income is calculated as:

  $\mathbf{RD=\frac{SD}{Ip}}$; (I_p is the general price index)

Current transfers in cash consist of:

1) Income tax;
2) Property tax;
3) Land tax on individuals;
4) Social insurance contributions and social benefits.

The generated national income is distributed among the members of society, enterprises, and the state. This process refers to the primary distribution of national income. The primary distribution of national income consists of:

a) Primary incomes of labor;

b) Primary incomes of production and service-providing enterprises.

The redistribution of national income includes:

- Government budget revenues;
- Revenues of institutions servicing households.

Statistical agencies analyze both absolute and relative indicators of national income through their reports. In this analysis, the population's income is grouped, and conclusions about the standard of living are drawn.

12.6. Concept of Gross Domestic Product (GDP)

Gross Domestic Product (GDP) is a key macroeconomic indicator that reflects the economic potential of a country and is used in international comparisons.

GDP represents the market value of all final goods and services produced within a country's territory during a specific period (quarter, year). It includes the income earned by all residents and non-residents in the country. GDP is composed of the results of economic activities in material production industries, the service sector, private household activities, and individual labor.

Gross Output (GO) is the initial indicator used in the calculation of GDP. In the System of National Accounts (SNA), Gross Output is analyzed in two groups:

1) Market-based production;
2) Non-market production.

Production consists of:

1) Goods and services produced by individuals and communities;
2) Goods that producers have retained for their own use or for gross savings;
3) Services provided by paid domestic workers.

Economic activity encompasses both production activities and non-market services performed by households for their own consumption.

In the system of national accounts, the value of goods and services consumed in the process of gross output is referred to as intermediate consumption (IC).

The composition of intermediate consumption includes:

1) Costs of material raw materials;
2) Fees for non-material services;
3) Travel expenses.

12.7. Methods of Calculating GDP

GDP is calculated using three methods:

1) Production Method (also called the output method);
2) Income Method (also called the distribution method);
3) Expenditure Method (also called the final use method).

Production Method:

In the production method, GDP is calculated at the sector level and for the entire economy.

In this method, GDP is the sum of the Gross Value Added (GVA) at the sector level, which can be expressed as:

$GDP = \sum \text{GVA}$,

Where GVA=Gross Output (GO) − Intermediate Consumption (IC).

For the economy as a whole, GDP is calculated as:

GDP = GO - IC + Taxes on Products (TP) + Subsidies on Products (SP).

Where:

GO = Gross Output;

IC = Intermediate Consumption;

TP = Taxes on Products;

SP = Subsidies on Products.

The value of taxes on products is calculated by subtracting subsidies on products from taxes on products.

The value of taxes on imports is calculated by subtracting subsidies on imports from taxes on imports.

Income or Distribution Method:

In the distribution method, GDP is calculated using the following formula:

GDP = Compensation of Employees (CE) + Gross Profit (GP) + Mixed Income (MI) + Taxes on Production (TP).

Where:

Compensation of Employees (CE) includes wages, salaries, and other benefits;

Gross Profit (GP) refers to the profits earned by businesses;

Mixed Income (MI) represents income from self-employed individuals who do not separate labor and capital income;

Taxes on Production and Imports (TP) are taxes on production and imports;

Expenditure Method (Final Use Method)

In the final use method, GDP is calculated using the following formula:

GDP = Final Consumption Expenditure (FCE) + Gross Capital Formation (GCF) + Export-Import Balance (EIQ) + Statistical Discrepancy (Δ_stat).

Where:

Final Consumption Expenditure (**FCE)** consists of household consumption, government consumption, and consumption by non-profit institutions serving households;

Gross Capital Formation **(GCF**) includes investments in fixed capital and changes in inventories;

EIQ is the difference between exports and imports (exports minus imports);

Statistical Discrepancy (Δstat) accounts for any differences in data sources.

Final consumption expenditure consists of three parts:

1) Government expenditure;
2) Household consumption expenditure;
3) Expenditure by non-profit organizations serving households.

Gross capital formation consists of:

1) Allocations for the development of fixed capital;
2) Changes in inventories of circulating capital;
3) Net purchases of valuables.

To calculate the export-import balance, the value of imports is subtracted from exports.

GDP Physical Index Volume:

The physical volume index of GDP is calculated as: $\mathbf{Iq}=\frac{\sum \mathbf{q_1 \cdot p_0}}{\sum \mathbf{q_0 \cdot p_0}}$

Where q_1 and q_1 represent the quantities of goods or services produced in the current and base periods; p_0 is the price of goods or services in the base period.

Methods of Expressing GDP in Constant Prices

GDP can be expressed in constant prices using the following methods:

1) Deflation Using the Price Index:

$$\sum \mathbf{q_1 p_1} : \mathbf{I_p} = \sum \mathbf{q_1 p_1} : \frac{\sum \mathbf{q_1 p_1}}{\sum \mathbf{q_1 p_0}} = \sum \mathbf{q_1 p_1} \cdot \frac{\sum \mathbf{q_1 p_0}}{\sum \mathbf{q_1 p_1}} = \sum \mathbf{q_1 p_0}$$

2) In this method, the gross output (GO) is calculated in constant prices, then intermediate consumption (IC) is also calculated in constant prices. The difference between constant-price GO and

constant-price IC gives the gross value added (GVA) at constant prices:

GO - IC = GVA $\sum$GVA = GDP.

3) Exploitation of Indicators at Base Period Prices: This method is used for expressing GDP using base period prices.

$$\sum \mathbf{q_0 p_0} \cdot \mathbf{I_q} = \sum \mathbf{q_0 p_0} \cdot \frac{\sum \mathbf{q_1 p_0}}{\sum \mathbf{q_0 \cdot p_0}} = \sum \mathbf{q_1 p_0}$$

4) Cost Element Deflation Method: In this method, indicators are deflated by cost elements to obtain constant prices.

Example: Calculation of GDP at Market Prices for a Country (Current Prices, Billion Soms):

1) Gross Output (GO) at basic prices = 28,054
2) Taxes on products and imports (TP) = 1,964
3) Subsidies on products and imports (SP) = 594
4) Intermediate Consumption (IC) = 13,124
5) Gross Profit and Mixed Income (GP + MI) = 7,364
6) Compensation of Employees (CE) = 7,078
7) Taxes on production and imports (TP_i.ch) = 2,454
8) Subsidies on production and imports (SP_i.ch) = 596
9) Final Consumption Expenditure (FCE) = 11,021
   a. Household Consumption = 7,627
   b. Government Expenditure = 3,056
   c. Non-profit Organizations Serving Households = 338
10. Gross Capital Formation (GCF) = 3,828
   a. Fixed Capital Formation = 3,294
   b. Change in Circulating Capital = 534
11. Net Export (EIQ) = 655
12. Statistical Discrepancy (Δ_stat) = ?

SOLUTION:

1. Using the Production Method:

$GDP = GO + TP - SP - IC$=28,054+1,964−594−13,124=16,300 billion soms.

2. Using the Income Method:

$GDP = GP + CE + TP - SP = 7{,}364 + 7{,}078 + 2{,}454 - 596 = 16{,}300$ billion soms.

3. Using the Expenditure Method:

$GDP = FCE + GCF + NX + \Delta stat = 11{,}021 + 3{,}828 + 655 + 796 = 16{,}300$ billion soms.

Statistical Discrepancy (Δ_stat) = 16,300 - 15,504 = 796 billion soms.

Review Questions:

1. What is the historical background of the System of National Accounts (SNA)?
2. Define the System of National Accounts (SNA) and its purpose in macroeconomic analysis.
3. How does the United Nations' SNA classify the national economy into institutional sectors?
4. What are the main categories of the System of National Accounts?
5. Identify and explain the key macroeconomic indicators used in the SNA framework.
6. Describe the interrelationships among macroeconomic aggregates.
7. What is Gross Domestic Product (GDP), and why is it a key economic indicator?
8. What are the three approaches to measuring GDP?
9. How is GDP measured at constant prices, and what are the methods used for deflation?
10. How does Keynesian economic theory relate to macroeconomic aggregates and national income accounting?

CHAPTER 13. NATIONAL WEALTH STATISTICS

13.1. Concept and Classification of National Wealth

National wealth is one of the fundamental indicators in economic statistics. Its total value, measured in constant prices, reflects a country's economic and social potential, while the per capita amount serves as a measure of the country's level of economic development.

Economists in various countries have engaged in measuring national wealth and conducting cross-country comparisons. Historical records of economics and statistics indicate that the first calculation of national wealth was carried out in England by W. Petty in 1664. This indicator was subsequently computed in France in 1789, in the United States in 1805, and in Russia in 1864.

National wealth is defined as the accumulation of tangible and intangible assets created through ancestral labor over centuries and aggregated within society, including natural resource endowments.

In terms of its formation, national wealth consists of two components:

1. Wealth generated by human labor;
2. Wealth derived from natural endowments.

By the late 20th century, defining the components of national wealth and developing a unified conceptual framework for assessing its value, composition, dynamics, and utilization became a pressing issue. To address this, the United Nations Statistical Commission and the World Bank formed a working group led by D. Dixon (USA) and K. Hamilton (Canada). This group conducted studies in 1997, estimating the value, structure, and per capita levels of national wealth for 92 countries using 1994 data. In 2000, a similar assessment was carried out for 10 countries based on 1997 data. The findings were presented at the 26th General Conference of the International Association for Research in Income and Wealth in August 2000, where aggregated national wealth calculations for different countries were discussed. At this conference,

K. Hamilton presented his analysis of national wealth per capita in 110 countries for the year 1997, along with its growth trends. His calculations indicated that in many countries—where per capita Gross National Product (GNP) was below the global average—the growth rate of total capital was lower than the population growth rate. This suggested a decline in overall well-being in these nations.

National wealth is typically expressed as the sum of the following components: Fixed Capital (FC), Circulating Capital (CC), and Personal Property of the Population (PPP). This relationship can be represented by the formula:

$$NW=FC+CC+PPP$$

In the System of National Accounts (SNA), national wealth is defined as the accumulation of non-financial and financial assets and is classified as follows:

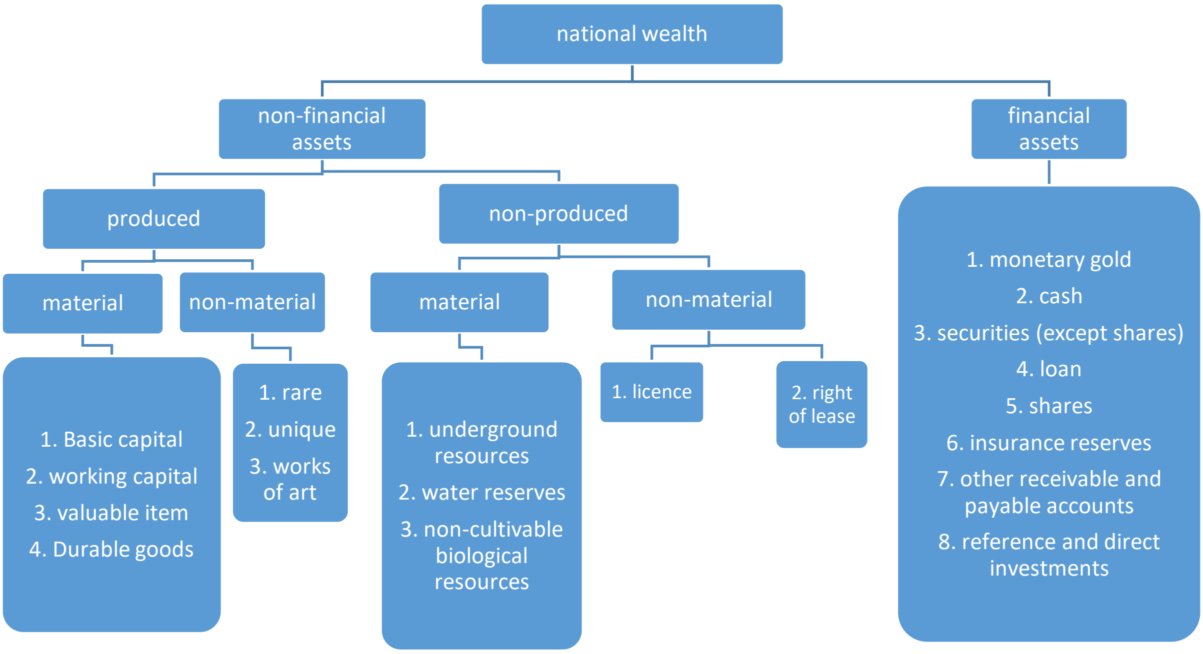


Figure 13.1. Classification of National Wealth

When calculating the net value of national wealth and its associated capital, a special table is prepared at the beginning of the year for each economic entity, as well as for the entire economy, to present the balance of assets and liabilities as follows:

Assets	Liabilities
1. Non-Financial Assets	3. Financial Liabilities
2. Financial Assets	4. Net worth of equity capital

The change in the value of assets is calculated using the following formula:

$$A_1 = A_0 + \Delta_{econ} + \Delta_{other+}\Delta_{inf/def}$$

where:

- A_1 and A_0 represent the value of assets at the beginning and end of the year, respectively.
- Δ_{econ} – Change in asset value due to economic transactions (e.g., production, buying, and selling).
- Δ_{other} – Other changes, including variations in natural resources, mineral deposits, and natural disasters.
- $\Delta_{inf/def}$ – Change in asset value due to inflation or deflation.

13.2. Fixed Capital Statistics

Fixed capital is a crucial component of a nation's wealth. Fixed capital refers to production assets created during the production process that serve for a long time and gradually transfer their value to the product.

Certain categories of durable goods are excluded from the composition of fixed capital, including such as small tools and equipment, some military items (destructive devices and their delivery mechanisms), and machinery and equipment purchased by households for final consumption (cars, motorcycles, bicycles, furniture, and household appliances). However, if these assets are utilized for production purposes, they are classified as part of fixed capital.

Expenditures on land improvement (its condition) and the enhancement of other non-produced tangible assets are not included in the composition of fixed capital. However, in addition to fixed capital, the following items are recorded on the balance sheet:

a) Historical monuments are included in the “Other Structures” subcategory, provided they have an acquisition cost;
b) "Machines and equipment" group, including military equipment used for civilian purposes (cars, computers, etc.);
c) "Transport vehicles" group, including space-faring apparatus;
d) "Animals" group, including rabbits, fish, poultry, and honeybees;
e) The value of unfinished assets (such as construction projects, production facilities, young animals, non-bearing perennial plantations, etc.) is recorded if ownership is transferred.

Fixed Capital Valuation:

It is known that the normative service life of fixed capital in international calculations, including by the World Bank, is set at 25 years. During this period, the valuation of initially accounted fixed capital changes. Fixed capital can be valued in historical (initial) value, current (revalued) value, residual value, depreciable value, scrap value, balance (mixed) value, or other methods. Each valuation method has its function in accounting, economic statistical analysis, financing (reinvestment), taxation, and calculating national wealth.

Full Initial Cost: This is the cost of acquiring fixed capital assets, including the purchase price, transportation, and installation expenses incurred by the enterprise.

Depreciated Initial Cost (Initial Residual Value) – The value obtained by deducting accumulated depreciation from the full initial cost of fixed capital.

Full Replacement Cost (Current Cost) – The expenditure required for an enterprise to acquire an identical fixed capital asset under current market conditions.

Depreciated Replacement Cost (Current Residual Value) – The value derived by subtracting accumulated depreciation from the full replacement cost.

Full Balance Sheet Value – The value of fixed capital as recorded in the enterprise’s balance sheet. Before the most recent revaluation, existing fixed capital is accounted for at full replacement cost, while newly acquired fixed capital is recorded at full initial cost.

In national wealth calculations, statistics and international practices recommend using the residual value indicator in the current value (replacement cost) for fixed capital. However, due to the varying speed of valuation changes and depreciation policies in different countries, it is difficult to conduct scientifically grounded international comparisons of fixed capital levels and per capita levels.

As is well known, the current (replacement) cost of fixed capital represents its value at prevailing market prices on a specific date. The revaluation of fixed capital is carried out using two main methods:

1. <u>Index-Based Method</u> – Applying indices such as the GDP deflator, Consumer Price Index (CPI), investment deflator, average index, specific indices for different fixed capital groups, or other deflators.
2. <u>Direct Valuation Method</u> – Determining market value directly, which may involve the use of real estate appraisal services.

In Uzbekistan, the revaluation of fixed capital (by groups) has historically been conducted using coefficient-based indexation, accounting for the degree of depreciation.

As a result of the revaluation of fixed capital, changes occur not only in its technological and sectoral structure but also in its degree of depreciation and usability (measured as the share of residual value in the total value). Thus, revaluation affects both the total value of fixed capital and its residual value in different ways. Consequently, this process has a direct impact on the total volume of national wealth.

The depreciation policy adopted in a country significantly influences the volume of fixed capital (its residual value) and, consequently, the size of physical capital. The method used to calculate depreciation—more broadly referred to as fixed capital consumption—has a substantial impact on GDP and reinvestment volumes. Therefore, it is essential to examine the methodologies applied in assessing physical capital, both globally and in Uzbekistan.

In the Republic of Uzbekistan, the levels of fixed capital depreciation, residual value, and other related indicators depend on the depreciation calculation method used in accounting, taxation, and statistics. Depreciation represents the normative value of fixed capital consumption rather than its actual value. International statistics recognize 11 different methods for calculating depreciation:

1) Age-Life Method – Based on the remaining useful life of an asset.
2) Annuity Method – Uses the annuity principle for calculating depreciation, applying compound interest to allocate costs evenly over time.
3) Combination of Production and Straight-Line Method – Depreciation is calculated using both the straight-line method and asset utilization rates (N = n - k).
4) Constant Percentage of Decreasing Balance Method – Depreciation is applied at a declining rate over time.
5) Double Declining Balance Method – A form of accelerated depreciation where the residual value is depreciated at twice the standard rate.
6) Labour-Hour Method – Depreciation is calculated based on the number of operating hours or total labor hours worked.
7) Multiple Straight-Line Method – A variation of the straight-line method where the rate is periodically adjusted.
8) 150% Declining Balance Method – Depreciation is calculated at 1.5 times the straight-line rate based on residual value.
9) Production Unit Basis Method – Depreciation is tied to the output produced rather than time. This method was widely used in Uzbekistan's extractive industries during the Soviet era.
10) Straight-Line Method – A widely adopted method where depreciation is allocated equally over the asset's useful life.
11) Sum-of-the-Digits (Sum-of-Years) Method – Accelerated depreciation where costs are allocated based on the sum of asset's lifespan years.

In developed countries such as the United States, Germany, Canada, the United Kingdom, and France, a widely used approach for calculating the residual value of fixed capital (Net Fixed Capital - NFC) is the Perpetual Inventory Method (PIM). This methodology, introduced by American economist R. Goldsmith, was first published in 1951 in his book "The Perpetual Inventory of National Wealth".

To apply this method, the following data is used:

1) Long-term dynamic series on gross capital investment;
2) Information on the normative service life of separate groups of fixed capital;
3) The current value of gross capital investments. This is done using the capital investment index, which aggregates indices for construction works, equipment, transportation vehicles, and construction materials. The overall index is calculated based on these specific indices.

The total depreciation amount is calculated using the following formula:

$$D_a = \frac{FC_{fiv} - FC_{dv}}{UL}$$

Where,

FC_{fiv}- full initial value of the fixed capital;

FC_{dis}- the disposal (or termination) value of fixed capital, can also be called as salvage value;

UL-useful life

The annual depreciation rate is calculated using the following formula:

$$D_{annual\,rate} = \frac{D_a}{FC_{fiv}} * 100$$

Indicators of Fixed Capital Movement:

- Fixed Capital Inflow Coefficient:

$$FC_{inflow\,coefficient} = \frac{FC\ inflow}{FC_{\ total\,value}} * 100\%$$

- Fixed Capital Renewal Coefficient:

$$FC_{renewal\ \ coefficient} = \frac{FC_{\ renewal}}{FC_{\ total\,value}} * 100\%$$

- Fixed Capital Outflow Coefficient:

$$FC_{outflow\ \ coefficient} = \frac{FC_{\ removed}}{FC_{\ total\,initial\,value}} * 100\%$$

Fixed Capital Condition Indicators

- Fixed Capital Usability Coefficient:

$$FC_{usability\ coefficient} = \frac{FC_{\ usable\ value}}{FC_{\ residual\ value}} * 100\%$$

- Fixed Capital Depreciation Coefficient:

$$FC_{depreciation\ coefficient} = \frac{FC_{\ tvbd} - FC_{\ rrv}}{FC_{\ tvbd}} * 100\%$$

  FC $_{tvbd}$ - Total Fixed Capital Value (Before Depreciation);
  FC $_{rrv}$ - Residual (Remaining) Value of Fixed Capital

- Indicates the extent of wear and tear or depreciation in fixed capital. These two coefficients sum to:

$$FC_{usability\ coefficient} + FC_{depreciation\ coefficient} = 100\%$$

Fixed Capital Utilization Indicators

- Fixed Capital Return Rate: Measures the output produced per unit of fixed capital.

$$FC_{return\ rate} = \frac{Q_{\ total\ production}}{FC_{\ total\ value}}$$

- Fixed Capital Intensity: Represents the amount of fixed capital required to produce one unit of output.

$$FC_{intensity} = \frac{FC_{\ total\ value}}{Q_{\ total\ production}}$$

- Fixed Capital Construction Coefficient: Evaluates the fixed capital stock relative to the duration of its utilization.

$$FC_{return\ rate} = \frac{FC_{\ total\ value}}{T_{\ total\ time\ period}}$$

13.3. Circulating Capital Statistics

Circulating capital refers to the material and financial resources that are fully transferred into the value of the product within a single production cycle.

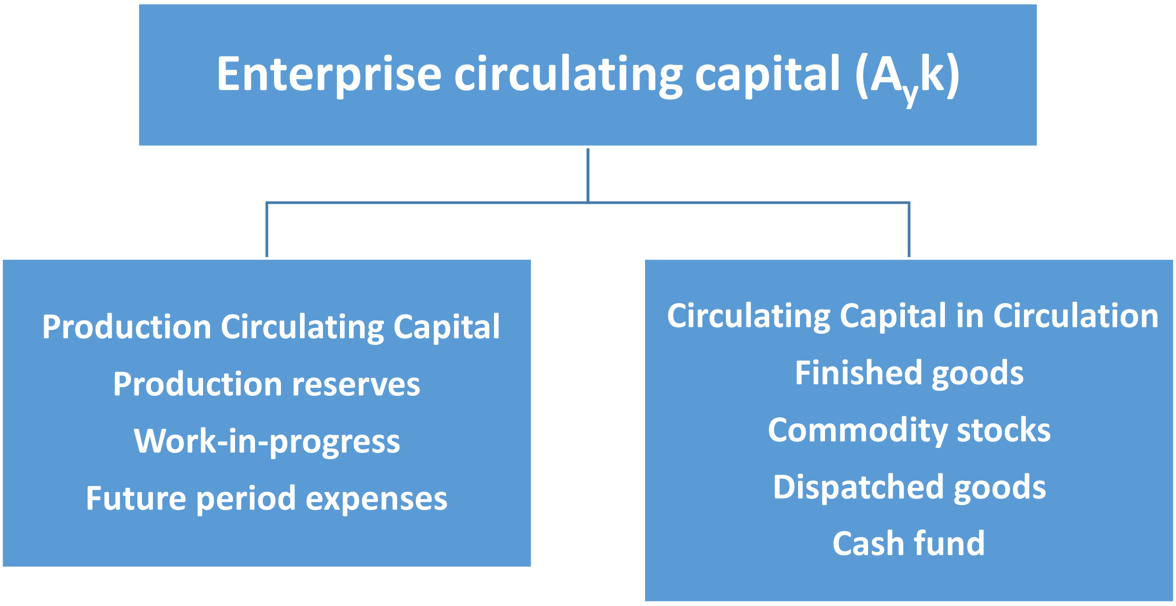


Figure 13.2. Enterprise circulating capital

Circulating Capital Turnover Ratio

The circulating capital turnover ratio (CC_{tr}) represents the volume of sold goods corresponding to one unit of circulating capital:

$$CC_{tr} = \frac{Q_r}{A\overline{yk}}$$

Circulating Capital Binding Ratio

The circulating capital binding ratio (CC_{br}) is determined as follows:

$$CC_{br} = \frac{A\overline{yk}}{Q_r}$$

Material Consumption and Indices:

$$ms = \frac{MS}{q} \qquad\qquad i_{ms} = \frac{ms_1}{ms_0}$$

$$\Delta_{ms} = ms_1 - ms_0$$

$$I_{ms} = \frac{\sum ms_1 q_1}{\sum ms_0 q_1}$$

$$\Delta_{ms} = \sum ms_1 q_1 - \sum ms_0 q_1$$

Raw Material Consumption Indices for Different Product Types

$$I_{ms} = \frac{\sum ms_1 p_1 q_1}{\sum ms_0 p_0 q_0}$$

Due to Changes in Product Quantity:

$$I_{ms} = \frac{\sum ms_0 p_0 q_1}{\sum ms_0 p_0 q_0}$$

Due to Changes in Unit Price of Products:

$$I_{ms} = \frac{\sum ms_0 p_1 q_1}{\sum ms_0 p_0 q_1}$$

Review Questions:

1. What is meant by national wealth in the System of National Accounts?
2. How is national wealth classified?
3. Explain the balance of assets and liabilities.
4. Define fixed capital and describe its classification.
5. What are the methods for calculating fixed capital depreciation allowances?
6. What do fixed capital movement indicators represent?
7. How are fixed capital condition indicators calculated?
8. Which formulas are used to determine fixed capital utilization indicators?
9. Define circulating capital and describe its classification.
10. How are the turnover coefficient, retention coefficient, and turnover period of circulating capital calculated, and what do they represent?

CHAPTER 14. FINANCIAL MARKET STATISTICS

14.1. Financial Statistics and Its Functions

The object of study in financial statistics comprises transactions and operations conducted for profit-making purposes. Its subject matter is the set of methods used to calculate the efficiency of these transactions and operations. The tasks of financial statistics are as follows:

1. To carry out monthly financial calculations;
2. To identify changes and trends in the financial market;
3. To determine indicators related to securities trading, insurance, taxation, banking, credit, and corporate stability.

Functions of Statistics

- To gather reliable information about the transactions and operations conducted in the financial market;
- To aggregate, classify, and represent the collected data in tables and graphs;
- To analyze all transactions and operations conducted in the financial market using all statistical methods;
- To identify trends;
- To determine the profitability of operations for participants, identify and introduce resources used in the market;
- To provide methodological assistance for the integration of the national financial market into the global financial market.

State Budget Statistics Functions

The state budget is the centralized monetary fund of the state used to meet national consumption needs. The main function of state budget statistics is to assess its main parameters:

— The total volume of state budget revenues and expenditures, the situation where expenditures exceed revenues (deficit) or where revenues exceed expenditures (surplus), official transfers, etc.;
— The composition of state budget revenues and expenditures;
— Sources of financing the budget deficit;

- The size of the state's internal debt;
- The size, composition, and other related aspects of taxation.
- The basis for studying the state budget is the economic classification of its revenues and expenditures.

The composition of budget revenues includes the following:

Tax receipts – mandatory, non-compensatory, non-refundable payments (e.g., value-added tax, excise duties, profit tax, income tax, etc.).

Non-tax revenue (revenue from the sale of state property, profits from managed enterprises, fines, etc.):

Non-compensatory revenue (transfers)

Budget expenditures are classified into the following categories depending on the functions being carried out:

- General government services;
- Public and social services;
- Government services related to economic activities;
- Other functions.

Economic classification is used to determine the activities of government bodies and their impact on the country's economy. The economic classification includes the following categories:

A. Current expenditures (expenditures on goods and services, interest payments, subsidies, and current transfer payments).

B. Capital expenditures (purchases of goods to create fixed capital and reserves, purchasing land and intangible assets, capital transfers, etc.).

The volume of redistribution processes resulting from the state's intervention is expressed as the ratio of the budget's revenue section to the Gross Domestic Product (GDP). The statistical study of budget revenues and expenditures is based on the dynamics of decision-making. Factors causing deviations from the plan are statistically evaluated. The main factors affecting tax revenue include the following:

- Changes in the tax base;
- Changes in the tax rate.

The effect of changes in the tax base on the absolute growth of tax revenue can be calculated using the following formula:

$$U_s = \Sigma\, S_{1i} \times K_{0i} - \Sigma\, S_{0i} \times K_{0i}$$

Where,

S_{1i} and S_{0i} represent the tax base volume for the current and base periods for the i-th indicator,

K_{1i} and K_{0i} represent the tax rate for the current and base periods for the i-th indicator.

The absolute growth of tax revenue is calculated as follows:

$$U_k = \Sigma\, S_{1i} \times K_{1i} - \Sigma\, S_{1i} \times K_{0i};$$

Where,

S_{1i} and S_{0i} are the tax base volumes for the current and base periods,

K_{1i} and K_{0i} are the tax rates for the current and base periods.

The absolute sum of taxes increased due to these two factors equals the total growth of tax revenue. For analytical purposes, alongside the main indicators, other indicators are also used to assess the impact of seasonal and other factors on the formation and expenditure of the budget.

Key Indicators of Bank Statistics

The banking system (central, commercial, and savings banks) plays a leading role in the financial and credit activities of a country. The system of indicators reflecting the activities of banks should assess their reliability and efficiency. The first group of indicators consists of absolute statistical indicators, which include the following:

a) Bank assets and resources;

b) Bank deposits;

c) Bank loans;

d) Capital or charter fund;

e) Profit.

The source of information about these indicators is the bank balance sheet published in the press. However, for in-depth analysis, these indicators are insufficient because such balance sheets do not specify the credit resources and their usage terms. Bank assets represent the utilized credit resources (deposits), which are equal to the total balance (currency) of the bank. Bank resources are the amount of funds available to the bank, which are used for lending and other operational activities and are also equal to the total balance (currency) of the bank. Bank resources can be divided into two groups:

1) Own Funds: This includes the bank's equity capital, which consists of shareholder and reserve capital. It is generated by the placement of bank shares in the securities market, as well as social funds, allocations from profits, and retained earnings.
2) Attracted and Borrowed Funds: The bank's attracted funds include resources held in correspondent and interbank deposit accounts; funds of enterprises and organizations transferred to the bank accounts (deposits); public savings funds; budget funds, and others. Borrowed (non-deposit) funds include loans from the central bank, interbank loans, temporary interbank financial assistance, and others.

To determine the adequacy of bank capital, the ratio of the bank's capital to the amount that could be lost is calculated using the following formula:

$$N_1 = \frac{C}{\mathrm{Ar} - \mathrm{Ru} - \mathrm{Rk} - \mathrm{Rd}}$$

Where:

C is the bank's capital; $A_r = \sum A_{0i} \times K_{pi}$ is the sum of risk-weighted assets, accounting for the potential loss risk; A_{0i} represents assets related to individual operations; K_{pi} is the risk coefficient for the i-th operation; R_u is the total amount of reserves created due to the depreciation of securities; R_k is the total amount of reserves that could be lost on loans; R_d is the total amount of reserves that could be lost from operations with other assets and receivables.

14.2. Advanced Financial Calculations

The goal of advanced financial calculations is to conduct a quantitative study based on a qualitative analysis of the content of financial processes.

The key indicators of advanced financial calculations are as follows:

1. Growth Value

Growth value refers to the increase in the initial sum of money due to added interest. The growth amount is calculated using simple (ordinary) interest and compound interest.

a) **Simple Interest Growth**

The growth amount with simple interest is calculated using the following formula:

$$S=P(1+in)$$

Where:

S is the increased value of the money, P is the initial sum, i is the interest rate, n is the number of interest periods, $1 + in$ is the multiplier for the initial sum.

b) **Compound Interest Growth**

For compound interest, the growth value is calculated using the following formula:

$S=P\ (1+i)^n$

If the loan term is less than a year, the growth value (S) is calculated as:

$$\mathbf{S=P(1+\frac{d}{k}i)}$$

Where: d is the number of days of the loan, k is the number of calendar days in the year.

2. Determining the Initial Sum

The initial sum can be determined using the following formulas:

For simple interest:	For compound interest:
$P=\frac{S}{1+in}$	$P=\frac{S}{(1+i)^n}$

3. Discount Amount

The discount amount (D), which is the benefit for the lender, can be calculated using the following formula:

$$D = S - P$$

Example Problem:

A loan of 10 million is provided for 3 years with an annual interest rate of 20%. The growth amount after 3 years is calculated as follows:

1. Simple Interest Growth: S=P(1+in)=10+10×0.2×3=16 mln
2. Compound Interest Growth:

$S=P(1+i)^n=10(1+0,2)^3=10*1.728=17.28$ mln

Finding Unknowns:

$$S=P(1+in) \qquad S=P\left(1+\frac{d}{K}i\right)$$

$$S=P+Pin \qquad S=P+P\frac{d}{K}i$$

$$Pin = S-P \qquad i=\frac{S-P}{pd}k$$

$$i=\frac{S-P}{Pn} \qquad d=\frac{S-P}{Pi}k$$

$$n=\frac{S-P}{Pi}$$

14.3. Monetary Circulation Statistics

Money circulation statistics is a system of statistical methods and indicators used to analyze economic processes and the circulation of money in the economy. Money circulation is important in assessing economic stability, inflation rate, interest rates and other macroeconomic indicators. This chapter provides detailed understanding of money circulation statistics and its main indicators.

According to general definitions, money is a special type of commodity that serves as a universal equivalent for all other goods. Money primarily fulfills the following five functions:

1) Medium of exchange;
2) Store of value;
3) Unit of account;
4) Means of payment;
5) Standard of deferred payment.

The first paper money appeared in China in the 9th century, in the United States in 1690, and in France in 1776. The subject of monetary circulation statistics is the system of indicators that characterize the level, dynamics, and composition of the money supply, as well as monetary relations carried out between economic entities in cash and non-cash forms. The money supply refers to the total amount of cash and non-cash funds available for transactions within a country, held by individuals and legal entities. The money supply (M) is composed of various monetary aggregates. The number of these aggregates varies depending on a country's economic structure and monetary management system. For example, in the United States, Russia, and Uzbekistan, the money supply is divided into four aggregates. In Japan and Germany, it consists of three aggregates. In the United Kingdom and France, there are two monetary aggregates.

The structure of monetary aggregates is as follows:

M_0 – Cash in circulation

$M_1 = M_0$ + demand deposits (non-interest-bearing deposits)

$M_2 = M_1$ + time deposits (up to 3 years)

$M_3 = M_2$ + targeted loans, insurance policies

$M_4 = M_3$ + marketable securities (stocks, bonds, promissory notes)

$M_5 = M_4$ + foreign currency held by households, businesses, and organizations.

The **money multiplier** describes the increase in the money supply as a result of an expansion in bank reserves and is calculated as: $M_m=\frac{C+D}{C+R}$

Where C – Cash in circulation, D – Deposits, and R – Required reserves of commercial banks.

The amount of money required for circulation (m) is determined using the following formula:

$$m = \frac{q\bar{p}}{v}$$

Where q – Quantity of goods sold, $\bar{p}$ – Average price of goods, and v – Velocity of money circulation.

Monetary equilibrium is determined by the following conditions: If

$mv = q\,\bar{p}$, the value of money remains stable.

$mv > q\,\bar{p}$, the value of money increases (inflationary pressure).

$mv < q\,\bar{p}$, the value of money declines (deflationary trend).

The turnover period (t) is calculated as:

$$t=\frac{D}{V}; \text{ or } \left(V=\frac{D}{t}\right)$$

Where D represents the number of calendar days in the period. The overall velocity of money depends on changes in the velocity of each monetary aggregate and the share of each aggregate in the total money supply.

The importance of money circulation statistics can be explained as follows:

Money circulation statistics play an important role in the formation and implementation of state economic policy.

The central bank's monetary policy is implemented by controlling the money supply and inflation.

Money circulation statistics are important in ensuring the stability of the financial system.

Financial crises can be prevented by monitoring and analyzing the liquidity level of banks and financial institutions.

Money circulation statistics play an important role in making investment decisions.

In the financing of investment projects, the money supply, interest rates and the inflation rate are analyzed.

Money circulation statistics is an important tool for analyzing economic processes and formulating economic policy. These statistical methods and indicators allow analysis of money circulation, liquidity, inflation and interest rates in the economy. It is possible to ensure economic stability and prevent financial crises through money circulation statistics. This will help ensure the efficient operation of the economy and the well-being of the people.

Review Questions:

1) What are the objects of study and tasks of financial statistics?
2) How are advanced financial calculations performed?
3) How is the initial principal calculated under simple and compound interest methods?
4) What are the different monetary aggregates?
5) Explain the concept of the money multiplier.

6) How is the required money supply for circulation determined, and how does it affect the value of money?
7) How is the index analysis of money circulation velocity conducted?
8) What does banking credit activity statistics study?
9) How are credit operation indicators calculated?
10) How is the statistical analysis of credit indicators conducted?

CHAPTER 15. FOREIGN ECONOMIC ACTIVITY STATISTICS

15.1. Understanding Foreign Economic Activity

Foreign economic activity is a system of economic relations that represents the processes of exchange of products, services, capital and technologies between countries and organizations. This activity plays an important role in ensuring economic development, competitiveness and integration of countries in the international arena.

Export: the sale of domestically produced goods and services abroad. Exports provide a flow of foreign currency to the national economy and open new markets for local producers.

Import: the introduction of goods and services produced abroad into domestic markets. Imports provide greater choice for consumers and encourage domestic producers to become more competitive.

Problems and limitations of foreign economic activity

Tariff and notary barriers can hinder international trade. These barriers take the form of customs duties, quotas, licenses and standards.

Changes in exchange rates affect export and import operations. A sharp change in the exchange rate can disrupt the trade balance.

Political instability and security issues can have a negative impact on international trade and investment.

Changes in international financial markets and financial crises can harm the national economy.

Foreign economic activity refers to economic relations conducted between different countries with the aim of obtaining economic benefits**.** These economic relations take the following forms:

1) **International trade**, which consists of stable commercial transactions between different countries. Currently, structural changes in global trade are becoming evident; the share of fuel, raw materials, and food products is relatively decreasing, while the share of products from scientific and technological sectors is increasing.

2) **International production cooperation**, which involves stable economic ties between enterprises of different countries aimed at developing industrial collaboration.

3) **Interstate capital migration**, referring to the transformation of financial reserves into investments across national borders.

4) **International labor migration**, where labor flows from countries with a surplus workforce to those with a high demand for labor.

5) **Mutual settlements and currency transactions**, encompassing financial operations related to international payments and foreign exchange settlements.

Foreign economic activity is a complex and multifaceted process that greatly contributes to the economic development of countries. Through this activity, countries participate in international trade, attract investments, and adopt new technologies. At the same time, there are various problems and restrictions in foreign economic activity, which require effective management. By analyzing and managing these processes, countries will have the opportunity to ensure economic stability and become active participants in the global market. Foreign economic activity is economic relations conducted between different countries for the purpose of economic benefit.

15.2. Key Indicators of Foreign Economic Activity

<u>Indicators describing foreign trade:</u>

1) Sales of export-local products and services to other countries.

The export volume shows the competitiveness of the national economy.

2) Import - products and services purchased from other countries.

The volume of imports is related to the demand and resources of the national economy.

3) Trade balance represents the difference between exports and imports.

A positive trade balance (trade surplus) indicates that exports exceed imports. A negative trade balance (trade deficit) indicates that imports exceed exports.

Trade balance = Export - Import

4) The balance of payments represents the general balance of all economic relations of the state. The calculation of the balance of payments is divided into two main sections:
-Current account: Balance of trade, balance of services, balance of income and balance of transfers.
-Capital and financial account: Balance of investments, loans and other financial transactions.

5) Export quota of the region (K_e). The export quota refers to the specific volume of production and delivery of certain goods and services for export. (Quota refers to the share of each participant in the social activity sectors based on an agreement). In Uzbekistan, quotas are applied to limit the export of certain essential consumer goods and strategic materials.

$$\boldsymbol{Q_e = \frac{E}{GDP} \cdot 100\%} \qquad \boldsymbol{I_e = Q_{e1} : Q_{e0}}$$

In this formula: E is the export volume, I_e is the export index, Q_{e1} and Q_{e0} are the export quota levels for the current and base periods.
This indicator does not reflect the quality of international economic relations. Therefore, the real export quota of the region (Q_{re}) is calculated as:

$$QK_{re} = \frac{E}{NP} \cdot 100\%$$

Here, NP is the net product of the region.
The region's net export quota index is calculated as:

$$I_{re} = Q_{re1} : Q_{re0}$$

Here, Q_{re1} and Q_{re0} are the real export quota levels for the base and current periods.

Import quota refers to the quantity of foreign products that are permitted to be brought into a country each year, with the amount being limited. Licenses are granted for a limited quantity, and import without a license is prohibited. The region's import quota (Qi) is calculated as:

$$\boldsymbol{Q_i = \frac{I}{GDP} \cdot 100\%}$$

$$\boldsymbol{I_i = Q_{i1} : Q_{i0}}$$

The share of imports in the national consumption is calculated as:

$$Q_{di} = \frac{I_i}{\sum IS_i} \cdot 100\%$$

Where: I_i is the import of product i, and $\sum IS_i$ is the total consumption of product i
in the country.

Determining the country's foreign trade quota (Q_{ft}) is carried out as follows.

$$Q_{ft} = \frac{0.5(E + I)}{GDP} \cdot 100\%$$

The international competitiveness coefficient is calculated as:

$$Q_{ic} = \frac{E - I}{FTT}$$

Where FTT – Foreign Trade Turnover (FTT = I +E) and E−I=Δ_{FTT} is the difference in foreign trade turnover.

The foreign trade turnover per capita coefficient is:

$$Q_P = \frac{FTT}{\overline{P}}$$

The region's foreign trade turnover share index is:

$$Q_{FTT} = \frac{Region_{FTT}}{Global_{FTT}}\ x\ 100\ \%$$

The foreign trade turnover index is calculated as:

$$I_{FTT} = K_{FTT_1} : K_{FTT_0}$$

Export elasticity:

$$\varepsilon_E = \frac{\Delta E}{E_0} : \frac{\Delta GDP}{GDP_0}$$

Where:

Δ GDP is the change in GDP

ΔE is the change in export volume $\Delta E = E_1 - E_0$

E_0 and E_1 are the export volumes in the base and current periods

GDP_0 and GDP_1 are the GDP values in the base and current periods.

Export advancement coefficient:

$$Q_E = I_E : I_{GDP}$$

Import elasticity:

$$\mathbf{E_I = \frac{\Delta I}{I_0} : \frac{\Delta GDP}{GDP_0}}$$

Import advancement:

$$\mathbf{Q_i = I_I : I_{GDP}}$$

Cover coefficient:

$$\mathbf{C_{cover} = E : I}$$

Coordination index:

$$\mathbf{I_{coordination} = I_E : I_I}$$

In the study of a country's foreign economic relations, foreign trade plays an important role, and statistics thoroughly analyzes foreign trade in all aspects.

The tasks of foreign trade statistics include:

—Studying the volume, composition, and dynamics of foreign trade.

—Analyzing factors affecting the development of foreign trade.

—Comparing and studying the foreign trade of various countries.

In studying foreign trade, customs statistics data are used as the basis.

Customs Statistics is a component of foreign economic relations statistics. It accounts not only for goods crossing the country's borders but also for certain specified services and completed works. The object of study in customs statistics is the country's foreign trade turnover. Its scope of observation includes recording the export and import of goods in both physical (natural) and monetary terms, as well as determining their geographical direction. The basis of customs statistics is the Customs Code of the country. The primary data source is the Customs Declaration Form (CDF). This document records information on goods exiting and entering the country's borders. The declaration includes details such as the direction of the goods (entry or exit), value, product code and name, net weight, customs regime, and other relevant data. Based on these details, statistical information on the country's foreign trade is compiled. Therefore, standardized units of measurement, as well as international classifiers and nomenclatures, are used in the CDF.

These provide the basis for determining the geography of foreign trade and serve as a key data source for customs statistics.

In the customs statistics of the Republic of Uzbekistan, the following categories of goods are taken into account:

Import:

- —Goods imported for free circulation;
- —Re-imported goods;
- —Goods imported for inward processing within the customs territory;
- —Goods imported for processing under customs control;
- —Re-exported goods;
- —Goods brought in by individuals and transferred to state ownership;
- —Foreign goods imported into duty-free shops;
- —Goods imported under lease agreements for a period of one year or more.

Export:

- —Goods exported under the customs export regime;
- —Goods exported after processing within the customs territory;
- —Goods exported under the re-export regime;
- —Goods exported from duty-free shops, including those of Uzbekistan and foreign origin;
- —Goods exported under lease agreements for a period of one year or more.

In customs statistics, all goods that increase or decrease a country's material wealth are included (excluding currencies in circulation).

Customs statistics do not account for the following categories of goods:

a) Goods whose value and weight do not meet the statistical observation requirements;

b) Currencies in circulation, including monetary gold. These are used by banks to carry out interbank settlements;

c) Goods used by foreign diplomatic and consular missions, as well as gifts presented to state institutions, international organizations, and heads of state;
d) Goods not subject to commercial transactions, such as items imported for personal use, postage stamps, wedding or ceremonial items, etc.

Foreign diplomatic missions located within the country are considered as part of their respective home countries. Therefore, the movement of goods between countries and their diplomatic missions is regarded as domestic movement.

Goods imported from abroad for commercial purposes by individuals are accounted for separately in statistical reporting. These goods make up approximately 20–25% of the country's total imports.

Customs statistics record goods both in physical units and in monetary terms. This, in turn, enables the determination of the total volume and structure of exports and imports.

The expression of foreign trade turnover in monetary terms is a key indicator of a country's economic status and its international relations. The ratio of a country’s exports and imports over a given period is used to determine its trade balance. If exports exceed imports, the trade balance is considered active (surplus); if imports exceed exports, the balance is passive (deficit). In Uzbekistan's statistics, export-import operations are recorded in both U.S. dollars and the national currency. Conversion to U.S. dollars is done based on the official exchange rate set by the Central Bank on the date the Customs Cargo Declaration (CCD) is submitted. An import quota is defined as a quantitative restriction on the volume of foreign goods allowed into the country annually. The state issues a limited number of licenses permitting imports, and importation without a license is prohibited.

The following services are also included in the foreign trade turnover: construction and installation works, repair of ships, airplanes, trains, automobiles, and equipment. These services are all performed on a commercial basis.

The following items are <u>not</u> included in the foreign trade turnover:

- humanitarian aid provided free of charge;
- products brought for exhibition purposes;
- personal goods sent via postal services.

Protectionism is a policy aimed at protecting the national economy from foreign competition. This includes measures such as:

- the state prohibiting the import of certain goods under its authority;
- increasing tariffs on imported goods;
- raising environmental safety standards for imported goods;
- limiting the volume of imports through the establishment of quotas;
- restricting the amount of goods allowed to be imported under these quotas;
- providing subsidies for exported goods to reduce costs and increase competitiveness abroad;
- establishing export quotas to promote the sale of domestic goods.

Protectionism is often accompanied by policies that stimulate production for export and partially substitute imported goods with domestic alternatives.

Dumping refers to the practice of exporting goods at prices below their production cost in order to capture foreign markets, or selling goods in foreign markets at lower prices than in the domestic market. Dumping is considered a competitive strategy aimed at weakening or eliminating rivals in foreign markets. The resulting losses are often covered by the state budget. The profits generated from higher domestic prices are redirected to the state budget.

15.3. Balance of Payments

The Balance of Payments (BoP) is one of the key tools for macroeconomic analysis and forecasting. BoP data reflects the current state of foreign trade development over a given period. Based on the BoP, it is possible to determine the forms in which foreign investments have been attracted, the investments directed abroad, and whether the country’s external debt obligations have been fulfilled on time or not. It also shows how the Central Bank has changed its international reserves. The balance of payments is used to shape the country's monetary policy and to regulate the exchange rate and domestic foreign exchange market. The BoP also serves as a source of information for national accounts (System of National Accounts).

Table 15.1.

Parts of the Balance of Payments	Description
1. Current account	
1.1. Trade balance	Difference between export and import goods
1.2. Balance of services	Difference between export and import services
1.3. Balance of income	Investment Income and Payments
1.4. Balance of transfers	One-way transfers (eg, remittances)
2. Capital and financial accounting	
2.1. Capital accounting	Capital transfers and operations related to non-produced, non-financial assets
2.2. Financial accounting	Foreign investments, portfolio investments, and other financial flows
2.2.1. Direct investments	Direct investments in enterprises
2.2.2. Portfolio investments	Investments in stocks and other financial assets
2.2.3. Reserve assets	Foreign currency reserves of the Central Bank
3. Errors and omissions	Undetected errors and omissions in the balance of payments
4. Remaining balance	Current account and total balance of capital and financial accounts

The procedure for compiling BoP was first published as an international standard by the United Nations in 1947. Since then, the task of improving the rules for compiling the BoP has been carried out by the International Monetary Fund (IMF).

The BoP is a statistical report that systematically records a specific country's economic transactions with the rest of the world over a defined period of time (month, quarter, year).

The structure of the Balance of Payments consists of the following two accounts:

1. Current account – This account records all economic transactions between residents and non-residents related to goods, services, income, and current transfers.
2. Capital account – This account includes transactions involving foreign financial assets and liabilities.

In BoP, economic transactions are recorded using the double-entry principle: entries are recorded as credits (+) and debits (–). The Balance of Payments is compiled by the External Economic Activity Department of the Ministry of Economy.

Review Questions:

1. What is foreign economic activity, and in what forms does it manifest?
2. What are the main indicators used to describe foreign trade?
3. How is export elasticity calculated?
4. Provide a definition of the Balance of Payments.
5. What accounts are included in the Balance of Payments?
6. Which international organization is responsible for developing the methodology for compiling the Balance of Payments?
7. In the System of National Accounts, in which accounts is foreign economic activity reflected?
8. Provide information about the “Rest of the World” sector’s account of incomes and current transfers.
9. Explain the concept of duties (taxes) collected by customs authorities.
10. What functions do customs tariffs perform?

CHAPTER 16. STATISTICS OF LIVING STANDARDS

16.1. Concept of Living Standards

Living standards refer to the scope of fulfillment of physical, moral, and social needs, as well as the opportunities created to meet these needs. It is a complex socio-economic category that reflects these aspects.

To obtain a more comprehensive understanding of standard of living, it is necessary to conduct a deeper study of the population's income and needs, consumption levels and structure, housing, property, cultural and household goods, and other factors.

In order to get more complete information about the standard of living, it is necessary to study in depth the level of income and needs of the population, the level and structure of consumption, housing, property, cultural and household goods, etc. Indicators representing living standards are diverse and can be categorized into the following groups:

- General and specific;
- Economic and socio-demographic;
- Objective and subjective;
- Monetary and physical (natural);
- Quantitative and qualitative;
- Indicators of consumption proportions and composition;
- Statistical and other indicators.

General indicators include:

1. National income;
2. Consumption fund;
3. Per capita share of the consumption fund of national wealth, etc.

These indicators reflect the overall level of the socio-economic development of society.

Specific indicators may include:

1. Level and patterns of consumption;
2. Working conditions;
3. Housing provision and household amenities;
4. Level of socio-cultural services;
5. Conditions for child upbringing;
6. Social security, and others.

Economic indicators reflect the economic capabilities of individuals and society. Indicators that represent the level of economic development of society and the well-being of each person (such as real income, employment, etc.) are classified as economic indicators.

Socio-demographic indicators reflect characteristics such as the age, gender, and professional-qualification structure of the population, as well as the physical reproduction of the labor force.

Objective indicators are based on objective (economic, technical, etc.) data, whereas subjective indicators take into account the level of satisfaction among population groups or individuals regarding their work, family relationships, income, employment, lifestyle, and similar factors.

Value-based (monetary) indicators include all income-related indicators and other monetary-based measures (such as the volume of trade turnover, services, transportation, monetary savings and investments, etc.).

Physical (natural) indicators represent specific material goods and services measured in physical units (e.g., pieces, square meters, cubic meters, etc.), such as the volume of food consumption, energy use, property ownership, housing, and availability of cultural and household goods.

Quantitative indicators determine the volume of consumption of specific material goods and services, while qualitative indicators allow for the assessment of population well-being in terms of quality.

In economic literature, there is no single universally accepted indicator that precisely defines the category of "standard of living of the population." Although systems of indicators have been proposed to describe it, considerable debate remains.

One widely used method of assessing the standard of living is to express it in terms of the aggregate goods and services available to an individual, household, or social group within the population. In this approach, the primary indicator is the household income and its ability to access goods and services.

To analyze the standard of living of the population, the following indicators are typically used:

- Indicators of population income;
- Indicators of consumption and expenditure on goods and services by the population;
- Savings;
- Indicators of housing availability;
- Indicators of income inequality (differentiation), poverty level, and poverty threshold;
- Socio-demographic characteristics of the population;
- Composite assessments of the population's standard of living.

The standard of living is a socio-economic concept that reflects the level to which individuals' material and spiritual needs are satisfied and the degree of improvement in living conditions.

In the System of National Accounts, the standard of living of the population is expressed through the following indicators:

- Primary income (PI): wages and salaries, mixed income, income from property;
- Disposable income (DI): calculated as $DI = PI + \Delta CT$,

Where ΔCT is the net current transfers received from other sectors of the economy (i.e., the balance of current transfers received and paid);

- Adjusted disposable income (ADI): calculated as $ADI = DI + ST$,

Where ST represents social transfers in kind—non-cash social benefits provided to households by non-profit institutions serving households and government bodies (e.g., free services in education, culture, healthcare, etc.).

The Human Development Index (HDI), used by the UN, allows to evaluate of the standard of living of the population, the level of education, and health indicators.

$$\mathbf{HDI} = \frac{\mathbf{1}}{\mathbf{3}}\ (\mathbf{Life\ expectancy\ index} + \mathbf{Education\ index} + \mathbf{Income\ index})$$

The consumer price index (CPI) is used to evaluate price changes by monitoring changes in the consumer basket.

$$\mathbf{CPI} = \frac{\mathbf{Current\ year\ consumption\ basket\ value}}{\mathbf{Base\ year\ consumption\ basket\ value}}\ \mathbf{x\ 100})$$

The Gini coefficient is used to measure income inequality. A value of 0 represents complete equality, and a value of 1 represents complete inequality.

$$\mathbf{Gini = 1 - \sum (L_i - L_{i-1})\ (P_i + P_{i-1})}$$

Where: L_i = cumulative income share up to group i,

Pi = cumulative population share up to group i

This formula is based on the Lorenz curve, which visually represents income distribution.

The standard of living of the population reflects the economic and social status of the society. With the help of these indicators and indices, it is possible to create a complete picture of the well-being, quality of life and social opportunities of the population. Various measures are taken by the state and social institutions to improve the standard of living. These measures are aimed at increasing the well-being of the population, ensuring social justice and strengthening economic stability.

16.2. Indicators Reflecting the Standard of Living in the System of National Accounts

The main system of indicators used to reflect the standard of living of the population includes the following:

1. Standard of Living Coefficient

$$\mathbf{C_{living\ standard.}} = \frac{\boldsymbol{NNI}}{\boldsymbol{\bar{A}}};$$

Where: NNI – Net National Income (NNI),
$\bar{A}$ – Average annual population.

2. Standard of Living Index

$$\mathbf{I_{living\ standard.}=K_{t.d1.}:K_{t.d0.}=\frac{NNI_1}{\bar{A}_1}:\frac{NNI_0}{\bar{A}_0}}$$

3. Income Indicators

—Nominal Income (NI) $\boldsymbol{NI = LWI + PI}$

Where: LWI – Labor-based income (wages, entrepreneurial income),
PI – Property income (profits, rent, interest, dividends).

—Real Income (RI) $\mathbf{RI} = \boldsymbol{\frac{NI+TP-T}{I_p}}$

Where: TP[4] – Transfer payments, T – Taxes, Ip – General price index.

—Disposable Income (DI) **DI=RI−(CM+C)**

Where: CM – Compulsory payments, C – Contributions (such as pension fund contributions, union dues, etc.).

4. Human Development Index (HDI)

$$HDI = I_{GDP\ pc} \cdot I_{edu} \cdot I_{life}$$

Where: I_{GDPpc} – Ratio of the country's GDP per capita to the global standard, I_{edu} – Ratio of the country's average years of schooling to the global standard, I_{life} – Ratio of life expectancy to the global standard.

Interpretation of HDI values:

HDI < 0.5 → Low development
0.5 ≤ HDI < 0.8 → Medium development
0.8 ≤ HDI ≤ 1.0 → High development

5. Poverty Coefficient

Poverty is defined as the state in which a person is unable to meet their basic needs. Methods for Measuring Poverty:

[4] Transfer income refers to non-earned income received by individuals without providing goods or services in return, such as pensions, unemployment benefits, remittances, and social assistance.

- Statistical method – distribution of the population by income level
- Normative method – calculation based on the cost of a minimum consumption basket.

Poverty Coefficient Formula: $C_{poverty}=\frac{P_{below}}{\bar{A}}$

Where: P_{below} – Number of people with income below the poverty line, $\bar{A}$ – Average total population.

Poverty Index (Change Over Time):

$$I_p = C_{p_1} : C_{p_0} = \frac{P_{below\ 1}}{\bar{A}_1} : \frac{P_{below_0}}{\bar{A}_0}$$

Where: Subscript 0 refers to the base period, and subscript 1 refers to the current period.

16.3. Indicators Describing the Material, Spiritual, and Social Conditions of the Population

To express the standard of living, both absolute and relative poverty indicators are used. *Absolute poverty* refers to the complete absence of income or the lack of sufficient income for an individual (or household) to meet their minimum living needs. It is most commonly observed among the homeless and similar vulnerable groups.

In Uzbekistan, *relative poverty* is defined by an income level below the subsistence minimum. In international practice, this indicator typically applies to individuals whose income is less than 40–60% of the national average income.

Household income refers to the total amount of monetary and in-kind resources received or produced by individuals or their family members over a certain period.

Income may be in **monetary or in-kind form**. **Monetary income** includes all funds received as wages, pensions, scholarships, various social benefits, and income from property (such as interest, dividends, profits, rent); as well as proceeds from the sale of shares, securities, real estate, livestock, household produce, personal belongings, and various goods; and payments for services provided, among others.

Wages are the main source of income for the majority of the population. In developed countries, wages make up 70–80% of total household income. Wages represent the earnings of hired workers and are used to purchase basic necessities for workers and their families, as well as to form personal savings. Wages consist of three components:

1. Base wage (main salary);
2. Bonuses or incentive payments;
3. Overtime pay (compensation for work beyond regular working hours).

Depending on how much of the wage reaches the worker, wages are divided into *gross (brutto)* and *net (netto)* wages. Gross wage refers to the total amount recorded before deductions — the full wage assigned to the employee. Net wage, on the other hand, is the amount that remains after deductions (such as taxes and contributions), which is the actual cash received by the employee.

Wages are also categorized based on how the work is measured into time-based and piece-rate (task-based) wages. Time-based wages are paid according to the time worked (days or hours) by an employee with a certain level of skill and experience. Piece-rate wages are paid based on the amount or quality of goods produced or the volume of work completed. The amount earned through piece-rate wages varies directly with the worker's productivity.

Wage levels are variable and change due to three main factors:

1. *Demand and supply of labor* – Since labor is a commodity, wages (its price) depend on the market. When labor demand increases, wages rise; when supply increases, wages fall. Equilibrium wages are established when demand and supply balance.

2. *Labor productivity* – Wages also change regardless of supply and demand. Under conditions of employment, higher productivity leads to increased wages and vice versa.

3. *Market relevance of work* – Wages are paid not merely for working, but for labor that meets market demand. If goods sell well, wages increase; if they do not, wages may decrease. The inability to sell goods may lead firms to delay or reduce wage payments.

Pensions are a special form of income paid by the government or firms to individuals who can no longer work due to age or disability.

Old-age pensions are considered deferred income earned during employment. Individuals contribute a portion of their income to pension funds, which is returned as monthly payments upon retirement. Disability pensions are not based on earned income but are provided by the state to support those who are unable to work. The amount of pension varies depending on previous income and the degree of disability.

Allowances are another form of special income, usually provided as social assistance to the poor in lump sums or on a regular basis by the state, firms, or charities. Their amount depends on available financial resources.

Scholarships are a form of income provided to students as financial support, based on the expectation that they will contribute to society in the future. They are funded by the state or private organizations.

Profit is the income entrepreneurs receive for taking financial risks in business. It is not guaranteed and varies depending on the success of the venture. If the business succeeds, the entrepreneur earns profit; if not, losses occur. Since few people are entrepreneurs, profit accounts for a relatively small share of total household income. For example, in the U.S., profit constitutes about 15–18% of household income.

Dividends are the share of profits distributed to shareholders. They depend on the profitability of the issuing company, so different companies pay different dividend amounts. As the market economy develops, dividends play a growing role in household income as more people become shareholders.

Interest is income earned by lending money to others. Interest rates depend on the demand and supply for credit.

Rent is the income received by property owners for leasing their assets, such as land, buildings, apartments, or houses.

In-kind income includes all agricultural outputs such as crops, livestock products, poultry, produce from personal garden plots, and products gathered from nature for personal or household use.

Table 16.1.

Balance of Population's Cash Income and Expenditures

Income	**Expenditures and Savings**
1. Wages 2. Income received by employees from organizations, excluding wages 3. Dividends (stocks) 4. Proceeds from product sales 5. Pension payments 6. Scholarships 7. Income from the financial system 8. Income from remittances 9. Total income from foreign currency sales 10. Other income Total Income Expenditures exceeding income	1. Purchase of goods and payment for services 2. Mandatory payments 3. Savings and securities 4. Purchase of a house 5. Expenditures for purchasing foreign currency 6. Money sent through remittances Total Expenditures Expenditures exceeding income
Balance	Balance

The consumer basket (basket of goods) is the minimum set of food products, non-food goods, and services necessary to maintain a person's health and ensure their daily activities. The basket of goods consists of three categories: food, non-food, and service baskets. The food basket holds a crucial place within the consumer basket. The food basket is a collection of products required for one person to eat for a month (calculated based on the minimum consumption standards of these products). This set corresponds to physical needs, provides the necessary calories, contains essential nutrients, and aligns with traditional skills for organizing meals.

Minimum living standard: This indicator is calculated based on the "basket of goods". The "basket of goods" includes the following products and services:

1. Food products, beverages, tobacco
2. Clothing, footwear
3. Rent, water, electricity payments
4. Household items
5. Medical services
6. Transportation and communication services
7. Recreation, education, culture
8. Personal hygiene

The *elasticity coefficient* is calculated to quantitatively reflect the relationship between income or price dynamics and the consumption level of individual goods. This coefficient indicates how the consumption level changes when income (or price) per capita changes by 1%.

$$C_{\text{elasticity}} = \frac{\Delta \bar{y}}{y_0} / \frac{\Delta x}{x_0}; \quad \Delta y = y_1 - y_0 \qquad \Delta x = x_1 - x_0$$

In this formula:

y_0 and y_1 represent the consumption levels in the base and current periods;

x_0 and x_1 represent the income (or price) per capita in the base and current periods.

The *Consumer Price Index* (CPI) is calculated using the following formula, which takes into account goods and services:

$$Ip = \frac{\sum i_p d_0}{\sum d_0}$$

Where: i_p is the individual price index of each good or service, and d_0 is the share (weight) of goods and services in the market basket.

Statistical agencies also calculate the *average per capita consumption* of certain food products, as well as indicators of the population's provision with non-food goods.

The volume and structure of consumption are not determined solely by the total or per capita average income but also by the state of the consumer market. For example, by the level of supply of specific goods in the market and the price variation of those goods. To quantitatively reflect such relationships, the indicator of the purchasing power (PP) of the population's monetary income is calculated. This indicator shows the volume of goods that can be purchased and the services that can be paid for with the average per capita income:

PP = $PD_{\bar{A}}/\bar{P}_i$

$PD_{\bar{A}}$ - is the average per capita monetary income and $\bar{P}_i$ is the average price of good i.

A qualitative indicator of the population's standard of living is represented by indicators such as housing provision and living conditions.

The stratification of the population by income, which has become especially evident in the life of our Republic recently, has highlighted the need to adopt a system of indicators used in international statistical practice for analyzing the socio-economic differentiation of the population.

This analysis primarily involves distributing the population into groups based on average per capita income to compare the well-being of different population segments. Special attention is given to low-income social groups, as studying these segments is essential for developing effective state social policy.

Several indicators are calculated to characterize the income distribution of the population, including:

— <u>Mode of income</u>: the income level most frequently observed among the population;

— <u>Median income</u>: the income level that lies in the middle of a ranked distribution series. Half the population has income above this level, and half has income below it;

— Income differentiation decile coefficient (C_{idd}): This shows how many times greater the minimum income of the wealthiest 10% of the population is compared to the maximum income of the least affluent

10%. The formula is $K_d=d_{10}/d_2$, where d_{10} and d_2 are the 10th and 2nd deciles, respectively.

—Funds coefficient (CF): This indicator represents the ratio of average monthly per capita incomes between the population groups in the 1st and 10th deciles and is used to assess income inequality.

$$K_D = \frac{\bar{d}_{10}}{\bar{d}_1}$$

Where: $\bar{d}_1$ is the average monthly per capita income of the lowest 10% income group (1st decile); $\bar{d}_{10}$ is the average monthly per capita income of the highest 10% income group (10th decile).

—Gini coefficient of income concentration (C_{Gini}): This coefficient reflects the distributional characteristics of the total volume of income among various population groups.

$$C_{GIni} = \sum p_i q_{i+1} - \sum p_{i+1}\, q_i$$

Where: p_i is the cumulative share of total income up to group i, q_i is the cumulative share of the population up to group i. The Gini coefficient ranges from 0 (perfect equality) to 1 (maximum inequality), and is one of the most widely used indicators to assess income distribution fairness.

The consumer basket is calculated with respect to the following periods:

1. Compared to the previous month;
2. Compared to the previous quarter;
3. Compared to December of the current year;
4. Compared to the same month of the previous year.

The calculation of the basket of goods (consumer basket) involves several steps:

1. Individual price indices for each good and service are calculated across cities:

$$i_{pj} = \frac{p_{1j}}{p_{o\,j}}$$

Where j refers to the types of goods and services, p_{1j} is the current price, p_{0j} is the base period price.

2. The arithmetic mean of the recorded prices is calculated for each item.

$$\overline{p}_j = \frac{\sum_{j=1}^{n} p_j}{n}$$

3. The aggregate indices are calculated for each region by combining the individual goods and services included in the consumer basket. For this, the Laspeyres Price Index is used:

$$I_p = \frac{\sum_{j=1}^{n} \frac{p_{1j}}{p_{0j}} p_{0j} q_{0j}}{\sum_{j=1}^{n} p_{0j} q_{0j}}$$

(This formula aggregates individual indices into a general index.)

The prices of goods included in the Consumer Basket (CB) are continuously monitored by the National Statistics Committee of the Republic Of Uzbekistan, and a deflator is calculated based on these prices:

$$D=\frac{CB_1P_1}{CB_1P_0} \cdot 100\%$$

Where: CB_1P_1 is the consumer basket at current prices, CB_1P_0 is the consumer basket at constant prices.

If the deflator is greater than 1, it indicates an inflationary process—money depreciates, and its purchasing power decreases. If the price index is less than 1, it reflects a decline in prices (deflation).

The deflator can be used to determine the standard of living of an individual, a household, and the population as a whole.

The value of money (also referred to as real value) refers to the purchasing power of a unit of currency, i.e., how many goods and services can be purchased with it.

The real value of money depends on price levels, which in turn are influenced by production costs and the money supply circulating in the market.

There is an inverse relationship between the value of money and the price level: When prices decrease, the value of money increases. When prices rise, money depreciates.

$$D_{\text{eflator GDP}} = \frac{D_{GDP_1}}{D_{GDP_0}} \cdot 100\%$$

Consumer Price Index (CPI)

In international statistical practice, Laspeyres and Paasche indices of goods and services produced are referred to as types of Consumer Price Indices (CPI). The Laspeyres Index shows how consumer expenditures would change if only prices changed in the current period relative to the base period, while keeping the consumption level and structure constant:

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

However, since the structure of consumer expenditures constantly changes, the Laspeyres Index tends to overstate inflation, while the Paasche Index tends to understate it.

According to R. Torvey, a well-known expert in the field of Labour Statistics from the UK, there is no “ideal” formula for calculating CPI. There is always a discrepancy between what the index measures and what it is intended to reflect. Additionally, the issue of changes in the quality of goods remains largely unresolved in index calculations. In international practice today, CPI estimates are increasingly based on 200 to 300 group indices, which provide a more accurate benchmark.

Review Questions

1. What is the system of key indicators used to express the standard of living of the population?
2. How is the standard of living measured in the System of National Accounts (SNA), and what does it signify?
3. Which formulas are used to calculate the living standard coefficient and index, and what do these indicators reflect?
4. Explain the indicators for nominal, real, and disposable income.
5. What does the Human Development Index (HDI) represent?
6. What is meant by the balance of household monetary income and expenditures?
7. What components make up the "basket of goods" used in the calculation of living costs?
8. How is the income elasticity coefficient of per capita income calculated, and what does it indicate?
9. How is the Gini coefficient for income concentration calculated?
10. What are your views on the debates related to calculating the Consumer Price Index (CPI)?

Glossary:

Population is the sum of all the units studied for research.

A sample is a small group selected from the population.

Arithmetic mean is the average value obtained by dividing the sum of data values by the number of values.

Median is the middle value in an ordered data set.

Fashion is the most repeated value.

Dispersion is the squared average of the deviation of the values from the mean value.

The standard deviation is the square root of the variance.

Coefficient of variation is the ratio of the standard deviation to the mean value.

A histogram is a rectangular chart that depicts the distribution of data.

Regression is a method of determining the relationship between one or more variables.

Correlation is an indicator that shows the degree of relationship between two or more variables.

Hypothesis Testing is a method of testing a research hypothesis.

The p-value is a metric used to evaluate the outcome of a hypothesis test.

A Confidence Interval is a range of possible values of a population parameter.

Quartiles are values that divide a data set into fourths.

Variance is the spread or variability of data values.

Anova (Analysis of Variance) is a method of comparing the mean values of several groups.

Proportion - the ratio of a certain value to the total sum.

Percentages are the percentage of a value multiplied by 100.

A nominal variable is a variable classified by qualitative characteristics.

An ordinal variable is a variable in which the data are ordered, but the difference between the measurements is not determined.

An interval variable is a variable that has clear differences between measurements but does not have an absolute zero point.

A ratio variable is a variable that is the difference between measurements and has an absolute zero point.

The chi-square test is a method of determining the relationship between nominal variables.
The z-value is a value that indicates the position of the data in the standard normal distribution.
The t-distribution is a distribution that represents the distribution of mean values for small samples.
Cumulative distribution is an aggregate representation of a set of values.
A histogram is a chart that shows the frequency distribution of data.
An experimental design is a method of conducting an experiment and collecting data.
A contingency table is a table that shows the relationship between two or more categories.
A Liker scale is a multi-point measurement scale used in surveys.
Skewness (Skewness) is an indicator that shows the asymmetry of the data distribution.
Flatness (Kurtosis) - an indicator showing the degree of linearity of the data distribution.
Cluster analysis is a method of classifying data into groups.
Factor analysis is a method of reducing many variables to a smaller set of factors.
Logistic regression is a regression analysis involving binary variables.
Bayesian theory is a method of updating probabilities.
Monte Carlo simulation is a method of analysis using random numbers.
Error (Error) - the difference between the estimated value and the actual value.
The bootstrap method is a method of estimating statistics by resampling the sample.
Resampling is a method of resampling a sample.
Probability is the probability of an event occurring.
Odds are the ratio of the probability of an event to occur.
Naive Bayes is a classification method based on a simplified probability model.
Principal component analysis (PCA) is a data reduction method.

$R2$ (Coefficient of Determination) is an indicator showing the level of accuracy of the regression model.

Multicollinearity is the degree of correlation between several independent variables.

Heteroskedasticity is the lack of constant variation in a set of data.

Autocorrelation is the relationship between observed values in a time series.

Comparative analysis is a method of comparing different data sets.

Tests

1. What is meant by population?
   - A) Only people in big cities
   - B) The number of people living in a certain area
   - C) Only people living in rural areas
   - D) All people
2. What does statistics study?
   - A) Eating habits of people
   - B) Social, economic and demographic phenomena
   - C) Only economic indicators
   - D) Demographics only
3. What does descriptive statistics do?
   - A) Data collection and analysis
   - B) Depicting and summarizing data
   - C) Just analyzing
   - D) Collection only
4. What is the mean value?
   - A) The greatest value
   - B) The smallest value
   - C) The average amount of values
   - D) Only the first value
5. What is the median?
   - A) The largest number of values
   - B) The average value of the values
   - C) The middle value between the ordered values
   - D) The smallest number of values
6. What is fashion?
   - A) The greatest value
   - B) The smallest value
   - C) The most repeated value
   - D) Only the first value
7. What is the function of dispersion?
   - A) Calculate the average value of the values
   - B) Determining the difference between values
   - C) Only determine the largest value

D) Only determine the smallest value

8. What is standard deviation?

A) The average value of the values

B) The square root of the dispersion

C) The greatest value

D) The smallest value

9. What are the methods of statistical observation?

A) Registration, selective monitoring, current monitoring

B) Registration only

C) Selective observation only

D) Current observation only

10. What is registration?

A) Determining the population

B) Observation only in large cities

C) Census of all population

D) Monitoring only in rural areas

11. What is selective tracking?

A) Observation only in large cities

B) Monitoring a certain part of the population

C) Census of all population

D) Monitoring only in rural areas

12. What is Current Observation?

A) Continuous data monitoring

B) Observation only in large cities

C) Monitoring only in rural areas

D) Only one observation

13. What does the system of national accounts include?

A) Only the state budget

B) Analysis of the results of economic activity

C) Only export and import data

D) Only the population

14. What does Gross Domestic Product (GDP) represent?

A) The total value of all final products and services produced in the national economy during a certain period

B) Value of exported products only

C) Value of imported products only

D) Government expenditure only

15. What is added value?

A) New value created in the production process

B) Costs only

C) Incomes only

D) Export only

16. What does national income represent?

A) Only the state budget

B) Net worth created in the national economy

C) Only export and import data

D) Only the population

17. What is natural population growth?

A) The difference in birth and death rates

B) Only birth rates

C) Mortality rates only

D) Only migration indicators

18. What does the migration indicator represent?

A) Only birth rates

B) The difference between the number of people who moved to the area and the number of people who left

C) Mortality rates only

D) Only natural growth

19.What does the overall growth rate represent?

A) Only natural growth

B) Growth as a result of natural growth and migration

C) Migration indicators only

D) Only birth rates

20. What does age-specific birth rate represent?

A) The number of births per 1000 people among women of a certain age group

B) Only birth rates

C) Mortality rates only

D) Only migration indicators

21. What is the age-related mortality rate?

A) Only birth rates
B) The number of deaths per population of a certain age group
C) Migration indicators only
D) Only natural growth

22. What are statistical analysis methods?
A) Descriptive statistics, analytical statistics
B) Descriptive statistics only
C) Analytical statistics only
D) Registration only

23. What is the median value?
A) The greatest value
B) The smallest value
C) Average value between ordered values
D) First value only

24. What is the value of fashion?
A) The greatest value
B) The smallest value
C) The most repeated value
D) First value only

25. What is the function of dispersion?
A) Calculation of the average value of the values
B) Determining the difference between the values
C) Determine only the largest value
D) Determine only the smallest value

26. What is standard deviation?
A) The average value of the values
B) Square root of dispersion
C) The greatest value
D) The smallest value

27.What do population statistics study?
A) Only economic indicators
B) Demographic indicators only
C) Social, economic and demographic events
D) Only social indicators

28. What does national income represent?

A) Only the state budget
B) Net worth created in the national economy
C) Only export and import data
D) Only the population

29. What does the system of national accounts include?
A) Only the state budget
B) Analysis of the results of economic activity
C) Only export and import data
D) Only the population

30. What does Gross Domestic Product (GDP) represent?
A) The total value of all final products and services produced in the national economy during a certain period
B) Value of exported products only
C) Value of imported products only
D) Government expenditure only

31. What is added value?
A) New value created in the production process
B) Costs only
C) Incomes only
D) Export only

32. What is natural population growth?
A) The difference in birth and death rates
B) Only birth rates
C) Mortality rates only
D) Only migration indicators

33. What does the migration indicator represent?
A) Only birth rates
B) The difference between the number of people who moved to the area and the number of people who left
C) Mortality rates only
D) Only natural growth

34. What does the overall growth rate represent?
A) Only natural growth
B) Growth as a result of natural growth and migration
C) Migration indicators only

D) Only birth rates

35. What does age-specific birth rate represent?

A) The number of births per 1000 people among women of a certain age group

B) Only birth rates

C) Mortality rates only

D) Only migration indicators

36. What is the age-related mortality rate?

A) Only birth rates

B) The number of deaths per population of a certain age group

C) Migration indicators only

D) Only natural growth

37. What does descriptive statistics do?

A) Data collection and analysis

B) Describing and summarizing data

C) Analysis only

D) Collection only

38. What does analytical statistics do?

A) Data collection and analysis

B) Describing and summarizing data

C) Identification and analysis of relationships between data

D) Collection only

39. What are the methods of statistical observation?

A) Registration, selective monitoring, current monitoring

B) Registration only

C) Selective observation only

D) Current observation only

40. What economic entities does the national accounts system cover?

A) Public sector only

B) Households, enterprises, public sector and foreign sector

C) Only households

D) Only businesses

Cases in statistics

Case 1: Population Dynamics Analysis

The birth, death and migration rates of the population in one region are presented in the table below. Based on these indicators, determine the natural and general growth indicators of the population of the region.

Year	Birth rate	Deaths	Migration (arrival)	Migration (left)	Population (initial)
2020	1500	800	300	500	100,000
2021	1600	900	350	450	101,000
2022	1550	850	400	400	101,600

Assignments

Calculate the natural growth rate for each year.

Calculate the migration rate for each year.

Calculate the total growth rate for each year.

Determine the population at the end of the year.

The solution

1. Natural growth rate: Natural growth = Births - Deaths
2. Migration indicator: Migration balance = Arrivals - Departures
3. Total growth indicator: Total growth = Natural growth + Migration balance
4. Population at the end of the year: Population (end) = Population (beginning) + Total growth

Case 2: Business Development Analysis

The following information is provided about the production volume and costs in one enterprise. Based on this information, calculate the added value of the enterprise.

Year	Production volume (mln. sum)	Value of intermediate products (mln. sum)
2020	500	200
2021	550	220
2022	600	250

Assignments

Calculate the value added for each year.

Determine the growth rate of value added.

The solution

1. Added value: Added value = Production volume - Value of intermediate products

2. Growth rate: Growth rate = (Current year value added - Last year value added) / Last year value added × 100

Case 3: Consumer Price Index Analysis

Below are the prices of the products in the consumer basket. Based on this information, calculate the consumer price index.

The name of the product	Quantity(kg)	Cost (sum) 2020	Cost (sum) 2021
Bread	10	3000	3500
Milk	5	4000	4500
Oil	2	8000	9000

Assignments

Calculate the total value of the consumer basket for each year.

Determine the consumer price index for 2021.

The solution

1. Total Value: Total Value = S (Quantity × Price)

2. Consumer price index: Index = (total value of 2021 / total value of 2020) × 100

Case 4: Salary Analysis

The following information is provided about the average salary and number of employees in one enterprise. Based on this information, calculate and compare the total salary of the enterprise.

Year	The number of workers	Average salary (sum)
2020	100	1,500,000
2021	120	1,600,000
2022	130	1,700,000

Assignments

Calculate the total salary for each year.

Determine the overall wage growth rate.

The solution

1. Total wages: Total wages = Number of workers × Average wages

2. Growth rate: Growth rate = (Current year's total salary - Last year's total salary) / Last year's total salary × 100

Case 5: Local Market Analysis

The following information is provided about the sales volume and price of two different products in the local market. Based on this information, calculate the total sales volume and growth rate for each product.

The name of the product	Trade volume (tonn) 2020	Cost (sum) 2020	Trade volume (tonn) 2021	Cost (sum) 2021
Product A	200	10,000	250	12,000
Product B	150	20,000	180	22,000

Assignments

Calculate the total sales volume of each product for each year.
Determine the growth rate for each product.

The solution

1. Total Sales Volume: Total Sales Volume = Sales Volume × Price
2. Growth rate: Growth rate = (Current year's sales volume - Last year's sales volume) / Last year's sales volume × 100

REFERENCES

I. Decrees and Resolutions of the Republic of Uzbekistan

1. O‘zbekiston Respublikasi, (2017). O‘zbekiston Respublikasi Konstitutsiyasi [Constitution of the Republic of Uzbekistan]. Tashkent: O‘zbekiston.
2. O‘zbekiston Respublikasi, (2002). Qonun “Davlat statistikasi to‘g‘risida” [Law "On State Statistics"], No. 441-II, 12 December. Tashkent.
3. O‘zbekiston Respublikasi Prezidenti, (2009). “Barkamol avlod yili” davlat dasturini ishlab chiqish va amalga oshirish bo‘yicha tashkiliy chora-tadbirlar to‘g‘risida Farmoyish [Order on the Development and Implementation of the “Year of Harmoniously Developed Generation” State Program], 9 December. Xalq So‘zi, No. 238.
4. O‘zbekiston Respublikasi Prezidenti, (2005). “Tadbirkorlik subyektlari tomonidan taqdim etiladigan hisobot tizimini takomillashtirish va uni noqonuniy talab etganlik uchun javobgarlikni kuchaytirish to‘g‘risida” Farmon [Decree on Improving the Reporting System Submitted by Business Entities and Increasing Accountability for Its Unlawful Demands], 15 June. Xalq So‘zi.

II. Resolutions of the Cabinet of Ministers of the Republic of Uzbekistan and Regulatory Legal Acts of Ministries

5. O‘zbekiston Respublikasi Vazirlar Mahkamasi (2008). *Eksport qiluvchi korxonalarni moliyaviy va kredit yo‘li bilan qo‘llab-quvvatlashga doir qo‘shimcha chora-tadbirlar to‘g‘risida* [On Additional Measures to Support Exporting Enterprises Financially and Through Credit], Resolution No. 290, 31 December. Tashkent.

K.A.MUKHITDINOVA

STATISTICS:
A STUDY GUIDE

Qog`oz bichimi 60x84 1/16. Times New Roman
garniturasida terildi.
Ofset uslubida oq qog`ozda chop etildi.
Nashriyot hisob tabog`i 13.5, Adadi 100. Buyurtma № 27-04
Bahosi kelishuv asosida

«ZUXRO BARAKA BIZNES» MChJ
bosmaxonasida chop etildi.
Manzil: Toshkent shahar Chilonzor tumani,
Bunyodkor shoh ko‘chasi 27 A–uy.