

BIG DATA AND ITS PROCESSING USING STATISTICAL METHODS

Babanazarova S.A

Samarkand Institute of Economics and Service
Associate Professor of the Department of
“Economic Analysis and Statistics”

Abdullayeva E.J

Samarkand Institute of Economics and Service
Student of group 424, Banking specialization

Annotation: This article discusses the concept of Big Data, which is highly relevant in the current era, its main characteristics, and the processes of processing it using statistical methods. Due to the volume, velocity, and variety of Big Data, traditional analysis methods are insufficient, making it necessary to use specialized technologies and statistical methods for processing. The article analyzes the possibilities of deriving practical conclusions from Big Data using methods such as descriptive statistics, regression analysis, hypothesis testing, and clustering. Additionally, examples are provided of the application of Big Data in economics, medicine, marketing, banking, and public administration. The results show that effective use of Big Data is a key factor in optimizing management decisions and increasing competitiveness.

Keywords: Big Data, statistical analysis, data processing, descriptive statistics, inferential statistics, regression analysis, clustering, data mining, analytical methods, data volume, digital economy, forecasting, , data quality.

Annotatsiya: Ushbu maqolada hozirgi davrda katta ahamiyatga ega bo'lgan katta ma'lumotlar (Big Data) tushunchasi, ularning asosiy xususiyatlari hamda statistik usullar asosida qayta ishlash jarayonlari yoritilgan. Katta ma'lumotlar hajmi, tezligi va xilma-xilligi sababli an'anaviy tahlil usullari yetarli bo'lmagani bois, ularni qayta ishlashda maxsus texnologiyalar va statistik metodlardan foydalanish zarurligi ko'rsatilgan. Maqolada deskriptiv statistika, regressiya tahlili, gipotezalarni tekshirish, klassterlash kabi usullar orqali katta ma'lumotlardan amaliy xulosalar olish imkoniyatlari tahlil qilingan. Shuningdek, katta ma'lumotlarning iqtisodiyot , marketing, bank ishi va davlat boshqaruvidagi qo'llanilishiga doir misollar keltirilgan. Olingan natijalar katta ma'lumotlardan samarali foydalanish boshqaruv qarorlarini optimallashtirish va raqobatbardoshlikni oshirishda muhim omil ekanini ko'rsatdi.

Kalit so'zlar: Katta ma'lumotlar, Big Data, statistik tahlil, ma'lumotlarni qayta ishlash, deskriptiv statistika, inferensial statistika, regressiya tahlili, klassterlash, ma'lumotlar qazib olish (Data Mining), analitik usullar, ma'lumotlar hajmi, raqamli iqtisodiyot, bashorat qilish, ma'lumotlar sifati.

Introduction.

Big Data is one of the important concepts associated with the development of information technology and science in the modern world, which is characterized by an increase in its volume, type and speed. The term big data first appeared in the 1990s and refers to large volumes of data, many types and changing at high speed. This data usually represents large digital resources that are collected and analyzed by companies, governments and scientific organizations.

The concept of big data is largely associated with the development of modern information technology, and its main features and the process of processing using statistical methods constitute an important era in this field. Big data is usually characterized by its volume, variety and speed. Associated with these features, they show the limitations of traditional methods for analysis. The volume of big data is often measured in terabytes or even petabytes, which requires new approaches and technologies for statistical analysis.

They include many types of data: texts, image and video files, sensor data, etc. This data is collected at high speed, in real time or near real time, so its processing and analysis requires the use of advanced technologies, in particular statistics and computer science methods.

Research methodology.

Since traditional statistical methods are not sufficient for effective analysis of big data, new technologies and methodologies are being developed. Statistical methods, including descriptive statistics, regression analysis, hypothesis testing, and clustering, are important in working with big data, which allow us to draw practical conclusions from the data. These methods allow us to predict the internal structures, important properties, and future changes of data, which is of great importance in optimizing management decisions and increasing competitiveness. First of all, the volume of data is of great importance when working with big data. Traditional statistical methods, such as statistical probabilities or classification models, are usually designed for small and medium-sized data. For large data, computer resources and computing power are of great importance in their processing. To solve this problem, parallel processing technologies are widely used. These technologies allow you to divide data into pieces and process them simultaneously, using multiple resources. At the same time, special algorithms and models are used to integrate statistical methods with large volumes of data, to optimize them and adapt them to large volumes of data.

The issue of speed is also a big problem, especially when it is necessary to analyze data in real time or in very short periods of time. In this case, statistical methods and technologies, such as stream technologies, that is, systems that allow you to analyze data flows in real time, play an important role. With the help of these technologies, it is possible to identify important patterns and trends from large volumes of data flows. Such approaches are widely used, especially in the fields of financial markets, security, healthcare and communications.

As for the issue of diversity, big data comes from different formats and sources. This creates significant difficulties in applying statistical methods. Data can be unstructured, fragmented, or fabricated, making it difficult to work with. Statistical and database technologies are widely used to address this problem.

Analysis and results.

The analysis of big data using statistical methods has become an integral part of modern analytical processes, given its volume, diversity and speed. These methods allow you to draw practical and accurate conclusions from large data sets, as well as help make management decisions more effective and justified.

Descriptive statistics play an important role in expressing the general characteristics of big data. This method determines the distribution of data, central tendencies (mean, median, mode) and dispersion indicators (variance, standard deviation). It helps to understand the general state and structure of the data, and also uses graphs and charts to visually represent the main characteristics of a particular data set. For example, descriptive statistics are important for companies when performing customer segmentation, showing the distribution of geographical data, or performing statistical analysis on products.

Regression analysis is widely used to identify relationships between data and predict future changes. This method allows you to study the effect of one variable (independent variable) on another variable (dependent or dependent variable). For example, in the banking and finance sector, it is used to predict the ability to repay debts, the amount of deposits, and credit risk. This method also allows you to forecast future trends and changes, taking into account certain factors, which plays an important role in management and strategic planning.

The method of testing hypotheses is an important part of statistical theory. With this method, you can determine whether the assumptions or hypotheses made on the basis of data are true or false. This process is usually carried out using statistical probability theory and various testing methods. The clustering method allows you to group large data sets according to their similarities. With this method, objects in a data set are divided into segments based on their similarities and differences. For example, in banking, customer segmentation, marketing optimization, and credit risk management are essential for tailoring marketing strategies to individual customer groups. Clustering also provides opportunities to identify problems, understand the needs of specific groups, and optimize them.

The integration of big data and statistical methods continues today and in the future. With the development of artificial intelligence and machine learning technologies, statistical analysis processes are being automated and deepened. It allows you to draw practical and accurate conclusions from large data sets, and also helps to make management decisions more effectively and reasonably.

Conclusions and suggestions

This study comprehensively studied the process of processing big data (Big Data) and its statistical methods. Traditional analysis methods are becoming less effective in solving problems associated with the increase in the volume, type and speed of big data, and in particular, the possibilities of comprehensive analysis are expanding with methods such as descriptive statistics, regression analysis, hypothesis testing and clustering. With the help of these methods, the internal patterns, relationships and trends of big data are identified, which create the opportunity to make practical and strategic decisions. The integration of big data and statistical methods remains a key factor in the development of the field. This integration not only allows for a deeper analysis of data and optimizes decision-making processes, but also expands technological and scientific achievements. Future research in this area, especially in conjunction with artificial intelligence and machine learning, is expected to further develop statistical analysis. This, in turn, has a significant impact on social, economic and technological development and creates new opportunities for increasing efficiency in various fields. At the same time, technical and ethical issues arising in the analysis of big data, including data quality, privacy protection and limited computing resources, remain important problems. To solve these problems, there is a need to develop new methods and technologies, as well as to develop statistical methods in conjunction with machine learning and artificial intelligence. These areas determine the prospects for the development of the field and allow increasing efficiency in working with big data.

References:

1. Khan, S., & Ahmad, M. (2020). Katta ma'lumotlar tahlilida statistik usullar. O'zbekiston Davlat Universiteti nashriyoti.
2. Ismoilov, S. (2022). Katta ma'lumotlar: nazariya va amaliyot. Toshkent: Axborot texnologiyalari markazi.
3. O'zbekiston Respublikasi Statistikasi. (2021). Katta ma'lumotlar va analitik usullar. Toshkent: O'zbekiston Respublikasi Vazirlar Mahkamasi.



<https://ijmri.de/index.php/ijpse> , German international journals company

4. Statistika va ma'lumotlar tahlili – O'zbekiston Respublikasi Statistika Qo'mitasi:
<https://stat.uz/uz>
5. Tursunov, N. (2020). Katta ma'lumotlar tahlili va biznes qarorlar qabul qilish. Toshkent: Iqtisodiyot va boshqaruv nashriyoti.
6. Big Data va raqamli iqtisodiyot – Axborot texnologiyalari va kommunikatsiyalarini rivojlantirish vazirligi: <https://www.mf.uz/uz>