

Big Data and Statistics

Suwetha P^{1*}, Surulivel L² and Aruna S³

¹Ph.D. scholar, Department of Aquatic Environment Management,

² Assistant Professor, Department of Fisheries Extension, Economics and Statistics,

³ Assistant Professor, Department of Aquatic Environment Management,

Dr. MGR. Fisheries College and Research Institute, Ponneri, Tamil Nadu -601 204, India

Corresponding Author

Suwetha P

Email: suwethap0410@gmail.com



OPEN ACCESS

Keywords

Big data, Statistics, Application

How to cite this article:

Suwetha, P., Surulivel, L. and Aruna, S. 2026. Big Data and Statistics. *Vigyan Varta* 7 (03): 37-39.

ABSTRACT

The rapid growth of the digital era has resulted in the generation of huge and complex datasets from sectors such as healthcare, agriculture, environmental monitoring, finance, and social media. This has led to the emergence of Big Data, commonly characterized by the five V's: volume, velocity, variety, veracity, and value. However, data alone cannot provide meaningful knowledge without proper analysis. Statistics plays a crucial role in Big Data by offering systematic tools for organizing, modeling, and interpreting large-scale datasets. Statistical techniques such as regression analysis, hypothesis testing, clustering, factor analysis, and predictive modeling help in identifying patterns, relationships, and trends, thereby supporting evidence-based decision-making. Modern tools including R, Python, SPSS, SAS, Apache Spark, Tableau, and Power BI enhance the processing and visualization of Big Data. Despite its benefits, challenges such as computational complexity, storage issues, privacy risks, and ethical concerns remain significant. Future developments must focus on scalable and transparent analytical methods for sustainable Big Data applications.

INTRODUCTION

The growth of digital era has led to the generation of complex and massive datasets across various sectors such as healthcare, environmental monitoring,

agriculture, finance, and social media, giving rise to the concept of Big Data. Big Data is characterized by 5 V's which exceeds the capacity of traditional data processing systems

and demands advanced analytical approaches. Yet without systematic analysis, raw data by itself doesn't provide meaningful knowledge. Statistics offers a theoretical and methodological foundation for large-scale dataset organization, modeling, and interpretation. Big Data and statistical science are necessary to promote evidence-based decision-making in both research and applied areas, and to convert complex data into actionable insights.

What is Big Data

Big data refers to data collections that are too large for the majority of software tools to handle, such as exabytes or zettabytes (Matthew *et al.*, 2025). Very huge data sets with high volumes, velocity, veracity, and variability are involved. The data does not match the standard database design, is too large, or is too quick. It might need various approaches and instruments for processing, measurement, evaluation, and profiling. It is commonly characterized by the 5 V's (Volume, Velocity, variety, Veracity and Value).

Role of Statistics in Big Data

Statistics plays an important role in revealing hidden patterns, relationships and structures within large-scale and high-dimensional datasets in this big data era. Big Data needs advanced statistical analytics to support fast or real-time decision-making and effective data visualization. Statistical methodologies are employed in various fields, including genomics, online health systems, behavioral research, and administrative data analysis, to handle complex, varied, and extensive datasets. Advanced statistical approaches, including both Frequentist and Bayesian procedures, as well as novel pattern recognition algorithms, are being employed to examine these data sets (Fang *et al.*, 2015). The complexity of Big Data requires

computational statisticians who collaborate with computer scientists and engineers to create scalable models and analytical instruments. Ultimately, the value of Big Data depends not on its size, but on rigorous statistical processing and interpretation that ensure accuracy, validity, and meaningful inference.

Statistical Tools and techniques used

- Python & R
- SAS & SPSS
- Apache Spark
- Tableau & Power BI
- KNIME
- MATLAB
- Excel
- Regression Analysis
- Hypothesis Testing
- Cluster Analysis
- Data Mining/Predictive Modeling
- Factor Analysis

Applications of big data

Evidence-based decision-making in stock market prediction; Report writing for scientific and news articles, advertisements, emails, and blogs; Reduction in response time, such as healthcare, finance, marketing, and education; Finding positive or negative emotions in marketing, finance, and the political and social sciences; Feedback from customers or agents in customer call centers and healthcare; Placement of items in automated security and surveillance systems in the retail industry (Puri and Haritha, 2016).

Challenges and Future Directions

Big Data analytics has potential, but it also has drawbacks, such as limited data storage, computational complexity, privacy issues, and ethical dilemmas. Sensitive information

protection and data security are still major issues, especially when data is exchanged or processed across several platforms (Fan *et al.*, 2014). A major area of focus for statistical research is creating scalable algorithms that preserve inferential correctness. Future developments are anticipated to concentrate on combining statistical theory with high-performance computing and machine learning in order to handle the increasing complexity of Big Data. To ensure the long-term viability of Big Data analytics, it will be crucial to prioritize interpretability, transparency, and ethical responsibility.

CONCLUSION

The advantages of big data include significant improvements in computation time, significant cost reductions, and the ability to provide new services. Big data advancements allow us to answer questions that were previously unanswerable, draw knowledge and insights from data, increase company productivity, and generate significant value for the global economy. It should be highlighted, therefore, that big data's main benefits stem from its

processing and analysis rather than from its unprocessed state. Multidisciplinary collaborations will be necessary to assist decision-making and service innovation as a result of the profound developments in big data technology and management.

REFERENCES

- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, 1(2), 293-314.
- Fang, H., Zhang, Z., Wang, C. J., Daneshmand, M., Wang, C., & Wang, H. (2015). A survey of big data research. *IEEE network*, 29(5), 6-9.
- Matthew N. O. Sadiku, Uwakwe C. Chukwu and Janet O. Sadiku, Big Data: An Overview. *International Journal of Trend in Research and Development*, 12(6).
- Puri, G. D., & Haritha, D. (2016). Survey big data analytics, applications and privacy concerns. *Indian Journal of Science and Technology*, 9(17), 1-8.