Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080409
ISSN 2371-8838 Vol. 8 Num. 4

Construction of a Mathematical Statistics Experimental Platform Based on Mobile Platform and Embedded System

Xinyuan Sun[1], Wei Wang[2,*]

[1]*School of Mathematics & Statistics, Anhui Normal University, Wuhu, Anhui, 241000, China*
[2]*Basic Teaching College, Nantong Institute of Technology, Nantong, Jiangsu, 226000, China*
[*]*Corresponding author*

Keywords: Mathematical Statistics, Experimental Platform, Big Data, Hadoop, Data Analysis

Abstract: Mathematical statistics is a scientific method that studies how to collect, organize and analyze data to reveal the inherent quantitative regularity of things. It has a wide range of applications in agriculture, industrial production, and socio-economic fields, such as seed selection, process improvement, socio-demographic surveys and psychological analysis. This paper proposes a mathematical statistics experiment platform based on mobile platform and embedded system, which realizes data transmission and real-time monitoring through wireless communication technology and user interface, and introduces the key technology of Hadoop-based big data platform. The experimental results demonstrate the effectiveness of the platform in random number generation, sample data analysis, hypothesis testing and regression analysis. In the future, with the development of cloud computing and big data technologies, data analytics tools will be combined with artificial intelligence and machine learning to provide more advanced intelligent analytics and predictive capabilities, enabling automatic identification, classification and pattern discovery of data, improving analytical efficiency and accuracy, and supporting real-time decision-making and response.

1. Introduction

With the rise and development of modern management science, computer science, behavioral science and other disciplines, computer storage and processing of massive information has increasingly become a new understanding. Large-scale complex decision-making systems often contain a large amount of uncertainty, that is, a large number of semi-structured or unstructured problems, at this time, it is difficult to rely on individual experience and wisdom to make accurate decisions, and also carries a certain degree of risk. Therefore, decision makers need to use computers to assist in collecting and analyzing all kinds of historical and contemporary, others' and their own, successful and unsuccessful data, experiences, techniques, methods and so on. Using appropriate mathematical and statistical models, it is possible to realize solutions that may not be the best, but are optimal for the moment, which brings great convenience and great economic benefits to every area of our production and life.

This article first introduces the basic concepts of mathematical statistics and its importance in real life, including its wide application in agriculture, industrial production and socio-economic fields. Then, the article proposes an experimental platform for mathematical statistics based on mobile platform and embedded system as well as a big data platform based on Hadoop, and describes in detail the construction process and key technologies of these platforms. Then, the article shows the experimental results, including the results and discussions on random number generation, sample data analysis, hypothesis testing and regression analysis. Finally, the article provides an outlook on future trends, suggesting the possibility of combining with technologies such as cloud computing, artificial intelligence, and machine learning, as well as the potential of these technologies to improve the efficiency and accuracy of data analysis. The entire article is clear and well-structured, introducing both theoretical foundations and practical applications, providing readers with a comprehensive view of the application of mathematical statistics in modern science and technology.

2. Related Works

Experts have long conducted specialized research on laboratory platform analysis and data analysis platform technology. Kapici H O investigated the impact of designing inquiry-based technology-enhanced lesson plans on pre-service teachers' TPACK (Technological Pedagogical Content Knowledge) self-efficacy on a virtual platform. The analysis of lesson plans showed that they commonly used inquiry-based learning with moderate integration of online scaffolding tools and virtual labs. This suggests that effective use of technology in designing lesson plans enhances teachers' self-confidence and facilitates the integration of educational technology [1]. Nungu L investigated the role of online collaboration in supporting STEM (Science Technology Engineering Mathematics) learning in higher education, collecting data from 88 graduate students through oral interviews and questionnaires. The results showed that despite the challenges of internet connectivity, lab resources and technology skills, students successfully completed learning activities and learned to navigate technology from their peers through tools such as virtual labs and smartphones [2]. Vallejo W provided guidelines on how to run the virtual lab simulation and it can be modified and downloaded from the Github repository. He chose to use the Python programming language and Colab because they were free and widely recognized in academia [3]. Wieland H introduced PIOLab as a virtual laboratory to address the challenges of building gPIOT (General-Purpose Input /Output tables), utilizing cloud computing and collaborative research environments to build individual input-output tables by pooling resources and secondary data. He demonstrated the architecture and analytical capabilities of PIOLab with an example of a global steel supply chain [4].

Boss J M designed ICU (Intensive Care Unit) Cockpit. It is a secure, fast and scalable platform for collecting and visualizing multimodal waveform data in intensive care units and validating algorithms online [5]. Liu J considered the energy platform as a solution based on the history of the telecommunication industry that was committed to addressing the renewable energy challenges. The platform needs to involve large-scale energy storage and breakthroughs in new technology areas, such as efficient power electronics and deeply integrated energy technology and information science [6]. Jahn B outlined the modeling process in pandemic decision-making, from data collection to publication of results, emphasizing the use of mathematical, statistical, and decision analytic models, with particular attention to their application to COVID-19 [7]. Torres J A G presented an open source experimental design of a Bayesian optimization platform for multi-objective reaction optimization. He optimized the performance by fine-tuning the algorithmic components and applied it to chemical reaction optimization. He also developed a graphical user interface for non-specialists, enabling chemists to easily apply the platform to laboratory practice [8]. Cui M summarized existing and emerging high-throughput proteomics methods for biomedical research and clinical practice, covering methods such as mass spectrometry, protein pathway arrays, tissue microarrays, and single-cell proteomics, and discussed related computational and statistical algorithms [9]. Loh T P discussed the current status, challenges and needs of method evaluation from a clinical laboratory medicine perspective. He focused on key aspects such as standardized terminology, selection of performance indicators, experimental design, sample requirements, statistical evaluation and data reporting [10]. Verawati N improved the reasoning skills of STEM students in a modern physics course using virtual simulation integrated with LMS (Learning Management System) platform. The results showed that virtual simulation learning significantly improved reasoning skills and provided a significant advantage over traditional face-to-face learning [11]. Raman R conducted a bibliometric study of 9523 publications on virtual laboratories in higher education between 1991 and 2021. He also studied the Altmetrics of virtual labs for the first time and found that the correlation between citations and Altmetrics was low, but the percentage of publications of interest increased linearly [12].

Sabir Z utilized Artificial Neural Networks (ANN) and Levenberg-Marquardt backpropagation to solve the SIQ (susceptible-infective-quarantined) based COVID-19 mathematical model with a focus on blocking effects. By numerically analyzing three different types of authentication, test and training sample data, he showed that the mean square error of the nonlinear SIQ kinetic model was significantly reduced, which provided a useful reference to cope with COVID-19 [13]. Liu M presented the overall architecture of an online learning system based on big data to meet the growing demand [14]. Yano J suggested to emphasize the importance of data science in solving problems in chemistry experiments. Suggestions include strengthening ties between chemists and data scientists, developing chemical data science methods, co-designing chemical experiments, and integrating various data sources into a data network for chemical research [15]. Existing bottlenecks in research on statistical techniques for data mainly include difficulties in obtaining high-quality data, challenges in data integration and application of diversified tools, lack of interdisciplinary collaboration, applicability of statistical models in complex systems, and limitations in effectively communicating the research results and disseminating their applications. These issues limit the wide application and further development of statistical techniques in different fields.

3. Methods

3.1 Platform Construction Process

The process of building a mathematical and statistical experiment platform based on mobile platforms and embedded systems includes: selecting appropriate hardware platforms (e.g., Raspberry Pi, Arduino, etc.) and mobile devices (e.g., smartphones or tablets); developing or selecting statistical analysis software and applications to ensure compatibility and optimize performance; using wireless communication technologies (e.g., Bluetooth or Wi-Fi) to achieve data transmission and real-time monitoring; designing and implementing a user interface so that users can easily perform statistical experiments; and finally, conducting system testing and optimization to ensure the stability and reliability of the platform [16-17].

3.2 Key Technologies of Hadoop-Based Experimental Big Data Platform

When building a Hadoop-based intelligent experimental analysis big data platform, Python is mainly used for Web page data crawling, Flume is used for data aggregation and log collection, Kafka (Apache Kafka) is used to buffer the data, and Spark is used to calculate the real-time data and then store it in Hbase. For data that needs to be calculated and cleaned offline, it is stored in MySQL after MapReduce operation. For log information, it is analyzed and calculated using Hive and stored in MySQL. Finally, the data is visualized on a large data screen [18].

3.3 System Function Design

The main business process analyzed by the mathematical statistics platform is shown in Figure 1.

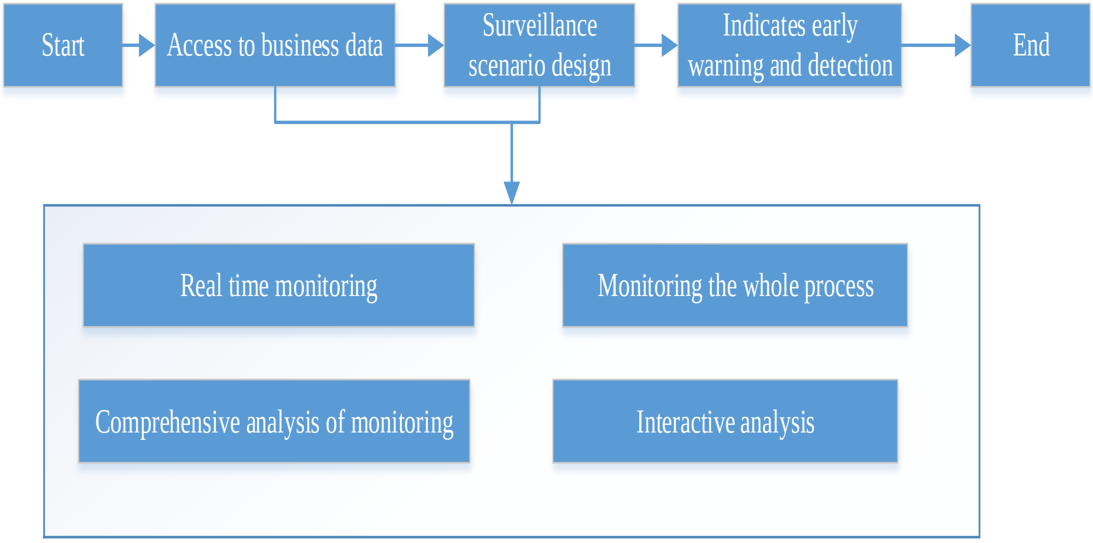

Figure 1: Laboratory Mathematical Statistical Analysis Business Process Diagram

1) In this study, real-time monitoring of the laboratory mathematical statistics analysis platform panorama. This study meets the application requirements by collecting the data of the laboratory mathematical statistics analysis platform. With the help of big data technology, this study can view the comprehensive information of the laboratory mathematical statistics analysis platform and monitor the running status of each experimental module in real time. This study analyzes the business processing of the mathematical statistics analysis platform in the laboratory in real time and monitors the whole process of the experiment [19].

2) This research comprehensive analysis laboratory mathematical statistics analysis platform. This study monitored the running status of the mathematical statistics analysis platform of the laboratory, the experimental processing of each department and the flow of the experimental process, and carried out real-time online monitoring of the mathematical statistics analysis of each module or the overall business indicators.

3) The whole process monitoring of laboratory mathematical statistics analysis platform. This study monitors modules, business types, regional experimental status, time distribution, laboratory mathematical statistics analysis platform and experimental status of each department. At the same time, on-site monitoring was carried out to make statistics on the details of the screening experiments, and supervision notifications were sent to the selected abnormal experiments.

4) Laboratory Mathematical and Statistical Analysis Platform Interaction Analysis: comprehensive analyzing user interaction data in the Laboratory Mathematical and Statistical Analytics platform, this includes the assessment of user behavior, quality of the experimental process, satisfaction, efficiency of problem solving, topic trends, and anomalous behavior. Through this analysis, the laboratory is able to gain a deeper understanding of user needs, improve service quality and adjust the system in a timely manner to meet changing needs.

4. Results and Discussion

4.1 Environment

The experiments were conducted in a laboratory with a wireless network connection to ensure smooth data transfer and communication between the devices. The laboratory was equipped with a variety of computing devices including desktop computers, laptops, tablets and smartphones to provide multi-platform support.

4.2 Devices

This system includes embedded systems using Raspberry Pi and Arduino, mobile devices such as smartphones and tablets for monitoring and displaying data, various sensor modules for generating random data including hardware and software random number generators for different types of random number generation, high performance desktop computers for data analysis and storage along with networking devices to ensure stable connection and data transmission.

4.3 Experimental Results

Table 1: Random Number Generator Testing

Random number type	Generator Type	Number generated	Average value	Standard deviation
Uniform distribution	Hardware RNG	1000	0.501	0.289
Normal distribution	Software RNG	1000	2.01	0.98
Exponential distribution	Hardware RNG	1000	0.027	0.015
Poisson distribution	Software RNG	1000	2.98	1.52

In the random number generator testing experiment in Table 1, we used hardware and software random number generators to generate different types of random number samples. For each sample, we calculated the mean and standard deviation as statistical indicators. The results show that the hardware random number generator produces uniformly distributed random numbers with a mean of about 0.501 and a standard deviation of about 0.289; the software random number generator produces normally distributed random numbers with a mean of about 2.01 and a standard deviation of about 0.98; the hardware random number generator produces exponentially distributed random numbers with a mean of about 0.027 and a standard deviation of about 0.015; the software random number generator produces Poisson-distributed random numbers with a mean of about 2.98 and a standard deviation of about 1.52.

In the sample data analysis experiment in Figure 2, we analyzed the samples of different sizes and calculated the mean and standard deviation of the samples, as well as the t-value and p-value as the test indicators. The results show that we obtained different t-values and p-values for samples of different sizes, where the t-value is used to measure the degree of difference between the sample mean and the overall mean, and the p-value is used to determine whether the difference is significant or not.

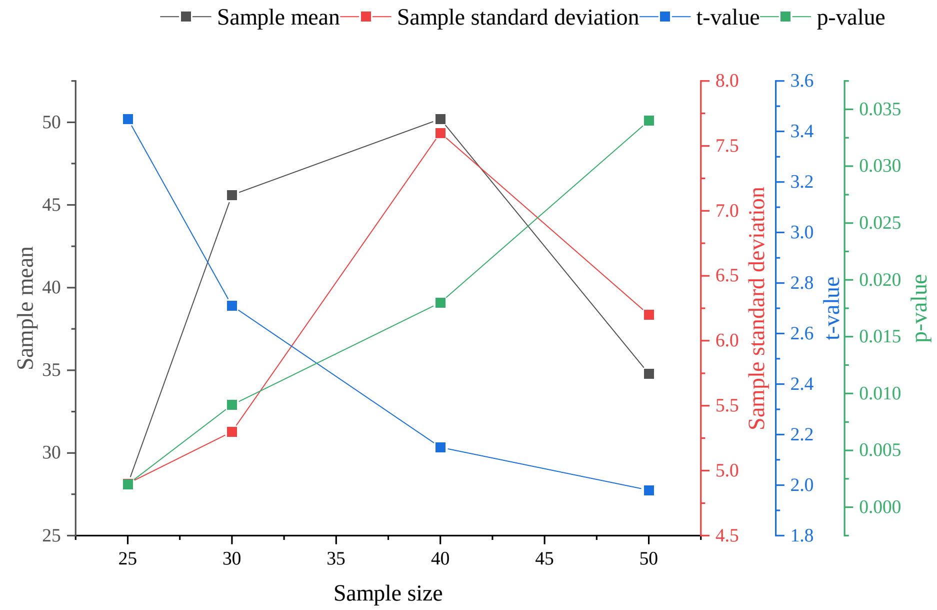

Figure 2: Sample Data Analysis

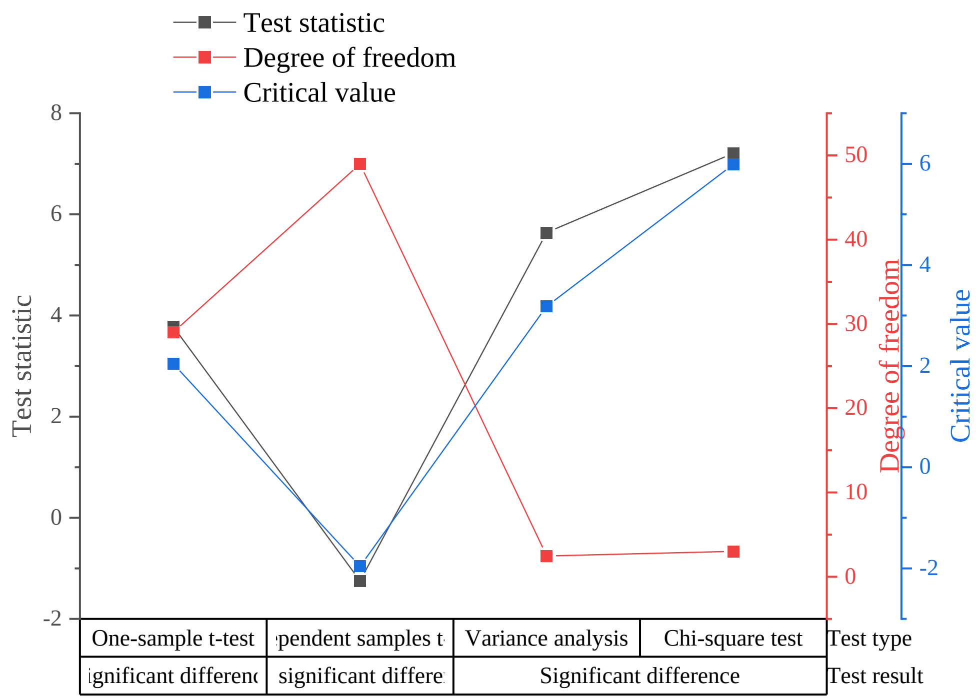

Figure 3: Hypothesis testing

In the hypothesis testing experiment in Figure 3, we used the one-sample t-test, independent samples t-test, ANOVA, and chi-square test to test for variability across samples. For each test, we calculated the test statistics, degrees of freedom, critical values, and test results. The results show that in different tests we obtained different test results, some differences are significant while others are not.

Table 2: Regression analysis

Regression equation	Correlation coefficient	Goodness of fit	Regression coefficient	Test of significance
$y = 2.5 + 0.8x$	0.76	0.58	Constant term $p < 0.05$	Significant Difference
$y = 1.2 + 1.4x$	0.89	0.79	Slope $p < 0.01$	Significant difference
$y = 3.6 + 0.6x$	0.67	0.45	Intercept $p < 0.1$	No significant difference
$y = 4.8 + 0.3x$	0.91	0.83	Slope $p < 0.001$	Significant difference

In the regression analysis experiment in Table 2, we used regression equations, correlation coefficients, goodness of fit, regression coefficients and significance tests to analyze the relationship between the variables. The results show that we obtained different regression equations and correlation coefficients for describing the degree of linear relationship between the variables. Meanwhile, we conducted significance tests on the regression coefficients to determine the degree of fit and predictive ability of the regression model. The results of these data analyses provide an important reference for the performance and validity of the experimental platform and help to assess the accuracy and reliability of the experiments.

5. Conclusion

With the continuous advancement of cloud computing and big data technologies, data analytics tools based on cloud platforms will have a broader application prospect in the future. With the continuous development of artificial intelligence and machine learning technologies, these technologies will be combined with data analytics tools to achieve higher levels of intelligent analysis and predictive capabilities. Cloud-based data analytics tools can integrate advanced machine learning algorithms to automate data identification, classification, and pattern discovery to more quickly analyze massive amounts of data and provide deeper insights. This will greatly improve the efficiency and accuracy of data analysis and provide users with more reliable decision support. In addition, with advances in mobile platforms and embedded systems technology, mathematical statistics lab platforms will become more flexible and accessible. Such technological advances will enable data analysis tools to run across a wide range of devices and environments, thus better supporting a wide range of application needs. Through wireless communication technology and user interface optimization, data transmission and real-time monitoring will become more efficient, further enhancing the practicality and user experience of the experiment platform. In conclusion, future data analysis will not only be limited to data processing and statistical analysis, but will also integrate various cutting-edge technologies to realize intelligent and automated data processing capabilities. These advances will promote the wide application of mathematical statistics experimental platforms in various fields, providing stronger support and services for scientific research, industrial production and social management.

Acknowledgement

This work was supported by The Ministry of Education's Industry University Collaborative Education Project, Project Number (230829282307299), Project Name: Construction of Higher Mathematics Teaching Reform Practice Base in Universities under the Background of Big Data.

References

[1] Kapici H O, Akcay H. Improving student teachers' TPACK self-efficacy through lesson planning practice in the virtual platform [J]. Educational Studies, 2023, 49(1): 76-98.

[2] Nungu L, Mukama E, Nsabayezu E. Online collaborative learning and cognitive presence in mathematics and science education. Case study of university of Rwanda, college of education[J]. Education and Information Technologies, 2023, 28(9): 10865-10884.

[3] Vallejo W, Díaz-Uribe C, Fajardo C. Google colab and virtual simulations: practical e-learning tools to support the teaching of thermodynamics and to introduce coding to students[J]. ACS omega, 2022, 7(8): 7421-7429.

[4] Wieland H, Lenzen M, Geschke A, et al. The PIOLab: Building global physical input–output tables in a virtual laboratory [J]. Journal of industrial ecology, 2022, 26(3): 683-703.

[5] Boss J M, Narula G, Straessle C, et al. ICU Cockpit: a platform for collecting multimodal waveform data, AI-based computational disease modeling and real-time decision support in the intensive care unit[J]. Journal of the American Medical Informatics Association, 2022, 29(7): 1286-1291.

[6] Liu J, Yang X, Wang X, et al. A laboratory prototype of automatic pavement crack sealing based on a modified 3D printer [J]. International Journal of Pavement Engineering, 2022, 23(9): 2969-2980.

[7] Jahn B, Friedrich S, Behnke J, et al. On the role of data, statistics and decisions in a pandemic[J]. AStA Advances in Statistical Analysis, 2022, 106(3): 349-382.

[8] Torres J A G, Lau S H, Anchuri P, et al. A multi-objective active learning platform and web app for reaction optimization [J]. Journal of the American Chemical Society, 2022, 144(43): 19999-20007.

[9] Cui M, Cheng C, Zhang L. High-throughput proteomics: a methodological mini-review[J]. Laboratory investigation, 2022, 102(11): 1170-1181.

[10] Loh T P, Cooke B R, Markus C, et al. Method evaluation in the clinical laboratory[J]. Clinical Chemistry and Laboratory Medicine (CCLM), 2023, 61(5): 751-758.

[11] Verawati N, Ernita N, Prayogi S. Enhancing the reasoning performance of STEM students in modern physics courses using virtual simulation in the LMS platform[J]. International Journal of Emerging Technologies in Learning (iJET), 2022, 17(13): 267-277.

[12] Raman R, Achuthan K, Nair V K, et al. Virtual Laboratories-A historical review and bibliometric analysis of the past three decades[J]. Education and Information Technologies, 2022, 27(8): 11055-11087.

[13] Sabir Z, Raja M A Z, Alhazmi S E, et al. Applications of artificial neural network to solve the nonlinear COVID-19 mathematical model based on the dynamics of SIQ[J]. Journal of Taibah University for Science, 2022, 16(1): 874-884.

[14] Liu M, Yu D. Towards intelligent E-learning systems[J]. Education and Information Technologies, 2023, 28(7): 7845-7876.

[15] Yano J, Gaffney K J, Gregoire J, et al. The case for data science in experimental chemistry: examples and recommendations [J]. Nature Reviews Chemistry, 2022, 6(5): 357-370.

[16] Aggarwal R, Verma J, Siwach M. Small files' problem in Hadoop: A systematic literature review[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(10): 8658-8674.

[17] Zhang J, Lin M. A comprehensive bibliometric analysis of Apache Hadoop from 2008 to 2020[J]. International Journal of Intelligent Computing and Cybernetics, 2023, 16(1): 99-120.

[18] Guan S, Zhang C, Wang Y, et al. Hadoop-based secure storage solution for big data in cloud computing environment[J]. Digital Communications and Networks, 2024, 10(1): 227-236.

[19] Zhang C, Liu X. An Analysis on the Learning Rules of the Skip-Gram Model. 2019 International Joint Conference on Neural Networks, 2019, pp. 1-8