ISSN 1391 4395 JAN-APR 2025

IASSL NEWSLETTER

Institute of Applied Statistics Sri Lanka
The Professional Center
275/75
Prof. Stanley Wijesundara Mawatha
Colombo 07
Sri Lanka

+94 11 2588291
admin@iassl.lk
http://www.iassl.lk
http://www.facebook.com/iassl2020/
https://www.linkedin.com/company/iassl/

"Statistical thinking is an essential component of informed citizenship."
- David Spiegelhalter

Featured Segments

INSIGHTS AND IMPACT: STATISTICAL RESEARCH TODAY

PUZZLE COMPETITION

NEWS IN BRIEF

ANNOUNCEMENTS

From IASSL President's pen

Dr. Rajitha M. Silva
President/IASSL

With great excitement and a deep sense of responsibility, I step into the role of President of the Institute of Applied Statistics, Sri Lanka (IASSL). Standing at this new threshold, I am reminded of a metaphor I shared at our recent Annual General Meeting—seeing IASSL as a living, breathing body. Each committee, each member, and each initiative functions like a vital organ, working together in dynamic harmony to keep our institute vibrant and strong.

.

The Research and Development Committee forms the backbone of our body, giving IASSL its posture and strength in the professional world. Their tireless work in mentoring, consulting, and recognizing emerging talent through awards and initiatives builds the very future of applied statistics in Sri Lanka.

Our Statistics Popularization Committee serves as the voice and face of IASSL, ensuring that statistics are not just a subject but a movement—alive, visible, and accessible. From the Statistics Olympiad to teacher training and outreach activities, they amplify our presence and proudly place IASSL on the national and global map.

The Academic and Training Committee is our lifeblood. Through their Diploma, Higher Diploma, and Certificate programs, they circulate knowledge, invigorate our intellectual vitality, and sustain our operations. Their work ensures that education and professional development remain central to our mission.

The House and Finance Committee stands as the muscle and metabolism of IASSL. They balance our resources with care and discipline, maintaining the strength and functionality of our institution through prudent management and constant upkeep.

The Editorial Board, meanwhile, acts as the eyes and brain of IASSL. Through the Sri Lankan Journal of Applied Statistics (SLJAS) and our newsletters, they nurture critical thought, promote scholarly dialogue, and project the intellectual brilliance of our members to the world.

And above all, the true heart of IASSL is you—our members, volunteers, and dedicated statisticians and data scientists. Your passion, commitment, and spirit keep this heart beating strong and steady. Together, united in purpose, we propel our institute forward.

I must also take a moment to recognize the extraordinary legacy of our founding members, past presidents, and longstanding council members. You are the DNA of IASSL, the architects of its growth, and the custodians of its values. It is your vision and dedication that allow us to dream bigger today.

Looking ahead, I invite every member—no matter which "body part" you represent—to move forward together. Let us collaborate, innovate, and inspire. Let us nurture new talent, strengthen our community, and spread the transformative power of statistics far and wide.

Because IASSL is not just surviving—it is thriving. And together, with unity, passion, and purpose, I believe our finest chapters are still to be written.

President,
Institute of Applied Statistics, Sri Lanka (IASSL)

Editorial

Dr. Chathuri L. Jayasinghe
Editor/IASSL

Dear Readers,

It is with great pleasure that we present the latest issue of the Institute of Applied Statistics, Sri Lanka (IASSL) Newsletter. This issue reflects the vibrant and diverse community that forms the heart of our Institute — bringing together contributions from distinguished academics, accomplished industry professionals, and emerging talents from among our undergraduate community.

Within these pages, you will find a rich selection of articles covering contemporary developments in statistical theory, applications in industry, and innovative research initiatives. We are proud to showcase the wide spectrum of expertise that continues to advance the role of statistics in solving real-world challenges.

In addition to scholarly articles, this issue features important news updates from IASSL during January to April, 2025, announcements of upcoming events and initiatives, and highlights of recent activities that further our mission to promote excellence in applied statistics.

Continuing one of our most popular traditions, we are pleased to present a new Sudoku puzzle for our readers to enjoy, along with the announcement of the winners from our previous edition. We encourage all members to participate and engage with us as we blend professional rigor with a spirit of community and collaboration.

We extend our sincere appreciation to all contributors and thank our readers for your continued support and engagement. Your participation strengthens the IASSL community and furthers our shared vision of advancing the practice and impact of statistics in Sri Lanka and beyond.

We look forward to your feedback and contributions to future editions.

Warm regards,
Editor,
Institute of Applied Statistics, Sri Lanka (IASSL)

CONTACT INFORMATION

Institute of Applied Statistics
Sri Lanka
The Professional Center
275/75, Prof. Stanley Wijesundara Mawatha
Colombo 07
Sri Lanka

+94 112588291

facebook.com/iassl2020

linkedin.com/company/iassl/

editor@iassl.lk

Why are Panel Data Models Important for Econometricians?

Dr. Neluka Devpura
B.Sc (Hons, UOC), M.Sc (NUS, Singapore), PhD (Deakin, Australia)
Senior Lecturer,
Department of Statistics, Faculty of Applied Sciences,
University of Sri Jayewardenepura

Econometrics is a branch of economics that uses mathematical and statistical theory to analyse economic data. Panel data models have a long history in econometrics over decades and they are getting more popular with empirical data analysis. Panel data consists of two dimensions: cross-sectional dimension and the time dimension. It is common to analyse a group of countries over a period of time to explore the relationships, similarities and differences. Further, another example would be a group of companies or firms during any specific period. Thus, this is where the panel data and panel regression techniques come into play widely. Hence, panel data regression models are highly useful in both economics and finance.

Panel data refers to pooling of observations accross a cross-section of countries, firms, households, or individuals over multiple time periods. over several time periods. In addition to these, surveying a number of households or individuals and following them over a period of time also can be considered as panel study. Two well-known examples of such data sets in Sri Lanka are the Household Income and Expenditure Survey (HIES) conducted periodically and Sri Lanka Labour Force Survey which are conducted quarterly by the Department of Census and Statistics (DCS). The latter survey follows the same individuals over time and is more aligned with the typical panel data structure used in econometric analysis.

Panel data models are important for econometricians because they provide rich information, more variability, compared to cross-sectional or time series data alone. The econometricians mainly deal with economic theory and data to come up with econometric models and make informed decisions. In this news article, we look fundamental concepts in econometrics and also, we review the literature and discuss the basic principles, theories, methods, models, and applications of panel data regression analysis. Panel regressions allow us to control individual and time-specific effects, thus providing more accurate and reliable results. One key application of panel regression is particularly useful in analyzing how economic policies impact to countries over time.

A panel data set is rich in data as multiple units are observed over a period of time and thus, it improves the efficiency and reliability of econometric estimates. Since two dimensions are considered that are cross-sectional and time series, the data consists of more observations and this leads to high degrees of freedom. When solving some econometric problems, we have to deal with endogenous explanatory variables as well as unobserved heterogeneity. However, panel regression can be used to overcome these issues or account for these when model building. Also, when there are multiple explanatory variables, the collinearity among these explanatory variables are reduced. Another important feature of the panel data set is the possibility of controlling unobserved heterogeneity. Omitting important variables makes parameter estimates to be biased but panel data allows researchers to control for individual-specific characteristics (like firm size, household features, or country policies) that are unobserved.

Types of panels are a balanced panel when the time period is the same for all the cross-sections and when different time periods are used then this becomes an unbalanced panel. The two types of panel regression models are linear and nonlinear panel data models, and for linear models, the common method of estimation is ordinary least squares whereas for nonlinear panel data, it is Maximum likelihood estimation.

There are various types of panel data regression models such as fixed effects, random effects, pooled, and first-differenced models etc.

The panel data studies received more popular due to recent developments and the availability of computer power and software to carry out sophisticated and rigorous computations. Econometricians commonly use software such as STATA, R, and EViews.

Econometric and Other Issues

The time dimension is a key feature of panel data sets and hence econometric issues (See, Reed and Ye, 2009) of serial correlation, non-stationarity, persistence due to unobserved heterogeneity, endogeneity and heteroskedasticity are major concerns. Moreover, parameter instability due to structural breaks is also another factor that affects the regression model. Structural instabilities may be present in the data due to disruptive events such as financial crises, COVID-19 etc. A very recent research work by Ditzen et al. (2024), develops the methods and asymptotic theory for the analysis of panel data with multiple structural breaks. Moreover, Karavias et al. (2023) explore structural breaks due to COVID-19 pandemic of stock returns of 61 countries with four explanatory variables over 38 weeks in 2020.

Economic panel data often exhibits the issue of cross-sectional dependence or interdependence. One cause of this issue is due to spillover effects among countries or the same sector firms etc.Chudik and Pesaran (2013) provide a survey of the literature on estimation and inference in large panel data models with cross-sectional dependence. Karabiyik et al. (2019) review the theory on estimation methods and statistical inference for both stationary and nonstationary panel data with cross-sectional dependence. Pesaran (2006) proposed the Common Correlated Effects (CCE) approach to the estimation of panel data models which is robust to different types of error cross section dependence.

Dynamic Analysis of Panel Models

Panel data can be used to study dynamics such as lagged effects which pure cross-sectional or time series models are not able to capture. Popular Autoregressive Distributed Lag (ARDL) models can be employed to build models in which the lag dependent variable is one of the explanatory variables. The dynamic model could be used to analyse the long and short terms effects of economic policy.

Some Applications of Panel regressions

Policy Evaluation:
Econometricians can evaluate policy impacts before and after interventions. In line with their research topic, Hsaio et al. (2012) proposed a panel data approach to assess the impact of a policy intervention of Hong Kong with China. In order to see the impact of political and economic integration, they used a panel of 24 countries and found very weak evidence on the growth of the Hong Kong economy.

Bond predictability:
Devpura et al (2021) employed a panel predictive regression model to study bond excess returns of 25 Organization for Economic Cooperation and Development (OECD) countries using a set of 12 predictor variables. These variables included macroeconomic, financial, and commodity-based indicators. The models used were,

$$r_{i,t} = \alpha_i + \beta_i x_{i,t-1} + u_{i,t}$$

$$u_{i,t} = \lambda_i f_t + \varepsilon_{i,t}$$

Where i is number of countries (i=1, 2, ..., 25), t is the time (from January 2000 to December 2016), $r_{i,t}$ is bond excess returns, $x_{i,t}$ is a mx1 vector of predictors, f_t is a common factor estimated as demeaned bond excess returns. λ_t is factor loading, and $\varepsilon_{i,t}$ is idiosyncratic error term.

In this paper, they found that commodity variables, both oil spot and futures prices, and the world commodity price index predict bond excess returns. Also, they confirmed that macroeconomic variables (term spread and T-bill yield) are successful predictors of bond excess returns.

Energy Economics:
Inglesi-Lotz (2016) studies the impact of renewable energy consumption on economic growth using panel data set of 34 OECD countries over the period of annual data from 1990 to 2010. The panel regression model suggests the impact is statistically significant.

Environmental Economics:
The impact of corruption on environmental performance measures was investigated by Lisciandra and Migliardo (2017) using 153 countries during the time span of 2002 to 2012. They employed panel VAR analysis and found significant evidence of corruption as an important determinant of environmental quality.

References

Chudik, A., & Pesaran, M. Hashem (2013), Large panel data models with cross-sectional dependence: A survey. SSRN Electronic Journal. 10.2139/ssrn.2316333.
Devpura, N. Narayan, P.K., & Sharma, S.S. (2021). Bond return predictability: Evidence from 25 OECD countries, Journal of International Financial Markets, Institutions and Money, 75, https://doi.org/10.1016/j.intfin.2021.101301.
Ditzen, J., Y. Karavias, and J. Westerlund (2024) Multiple Structural Breaks in Interactive Effects Panel Data Models. Journal of Applied Econometrics. https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3097.
Hsiao C, Steve Ching H, Wan S K. (2012) A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with the mainland China. Journal of Applied Econometrics, 27: 705 -740
Inglesi-Lotz, R. (2016) The impact of renewable energy consumption to economic growth: a panel data application, Energy Economics, 53,58-63, 10.1016/j.eneco.2015.01.003
Karabiyik, H., Palm, F. C., & Urbain, J. P. (2019). Econometric Analysis of Panel Data Models with Multifactor Error Structures. Annual Review of Economics, 11, 495-522. https://doi.org/10.1146/annurev-economics063016-104338
Karavias, Y., P. K. Narayan, and J. Westerlund. (2023). Structural Breaks in Interactive Effects Panels and the Stock Market Reaction to COVID-19. Journal of Business & Economic Statistics 41, 653–666.
Lisciandra, M. & Migliardo, C. (2017). An Empirical Study of the Impact of Corruption on Environmental Performance: Evidence from Panel Data, Environmental & Resource Economics, 68, 297-318.
Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. Econometrica 74, 967-1012.
Reed, W. R., & Ye, H. (2009). Which panel data estimator should I use? Applied Economics, 43(8), 985–1000. https://doi.org/10.1080/00036840802600087

Creating an AI Assisted Data Analysis Pipeline – without Coding

Dr. Sameera Viswakula
B.Sc (Hons, UOC), M.Sc (USA), PhD (USA)
Senior Lecturer,
Department of Statistics, Faculty of Science,
University of Colombo

Statisticians, Data Analysts, Data Scientists and research students analyze datasets at different stages as part of their work/research at various levels of complexities. However, the initial steps are the same in most cases. A Statistician may be fitting a model to predict the prevalence of a disease, an analyst may be finding some patterns of a couple of economic variables, a Data Scientist may be building a fraud detection algorithm, and a research student may be fitting a basic regression model to test the significance of the independent variable.

Nowadays, **AI Agents** are being used everywhere to make our lives easier. We have seen 'agents' in real life and most of us have used their services at least once. It can be an insurance agent, or ticketing agent etc., they act on the instructions or requests they receive and perform a task. An AI agent is a smart program that understands what you want, figures out how to help, and acts—like writing code or analyzing data. Here, the word 'smart' means it does the work without instructions.

How about automating the frequently used data pre-processing and initial analysis tasks using an AI agent? Once the agent finds a dataset, it will find the number of records, variables, missing values, outliers, it will do the missing imputation accordingly, remove outliers, do the encoding and scaling, suggest possible transformation etc. Sounds fantastic, right? But how do we build this AI agent? Cos' we are not professional programmers, right?

Now we do not have to worry about installing fancy software and a bunch of libraries anymore. We do not need to code extensively now thanks to **vibe coding**. Vibe coding means we just prompt our vibe to a LLM based tool and generate codes and develop apps with zero coding involved.

The following is a part of the prompt I submitted to Claude AI's free version to build an app to do the above discussed pre-processing steps. Claude AI is a generative artificial intelligence (AI) chatbot and a family of large language models (LLMs) developed by the Anthropic research firm. This is the initial prompt I submitted to Claude.

I want to create an AI agent to analyze data. It should do the following.

1. *Browse a dataset or should be able to specify any dataset.*

2. *Identify categorical and quantitative variables in the dataset.*

Is it possible for you to do this using RStudio and Shiny server?

Claude AI gave me a bunch of R codes and I just pasted them into RStudio, saved the file and clicked on the "Run App" button. Here is the app generated by the code written by Claude.ai for the above prompt.

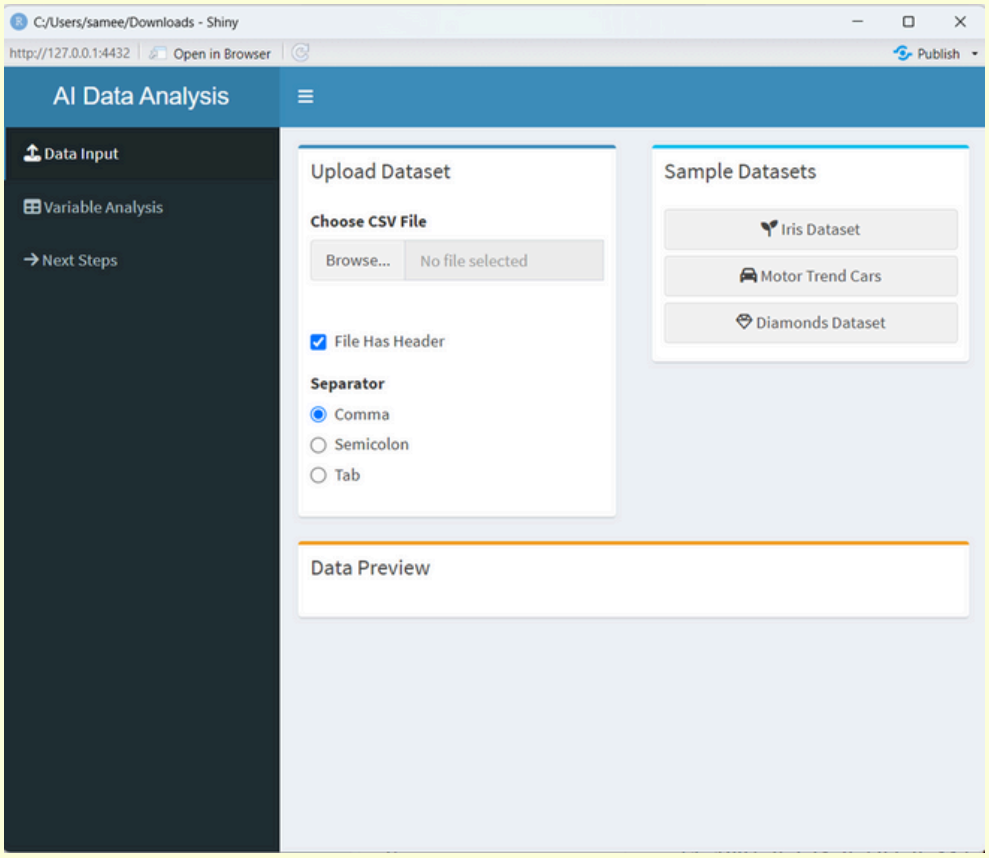


Figure 1: Landing page of the App

The "Advertising.csv" file contains a dataset with some advertisements related data. A snapshot of the dataset is as follows. Let's browse this dataset and feed into the app.

	A	B	C	D	E
1		TV	radio	newspape	sales
2	1	230.1	37.8	69.2	22.1
3	2	44.5	39.3	45.1	10.4
4	3	17.2	45.9	69.3	9.3
5	4	151.5	41.3	58.5	18.5
6	5	180.8	10.8	58.4	12.9
7	6	8.7	48.9	75	7.2
8	7	57.5	32.8	23.5	11.8
9	8	120.2	19.6	11.6	13.2
10	9	8.6	2.1	1	4.8
11	10	199.8	2.6	21.2	10.6
12	11	66.1	5.8	24.2	8.6
13	12	214.7	24	4	17.4
14	13	23.8	35.1	65.9	9.2

Figure 2: Snapshot of the dataset

The developed app identified the variables correctly and listed as follows.

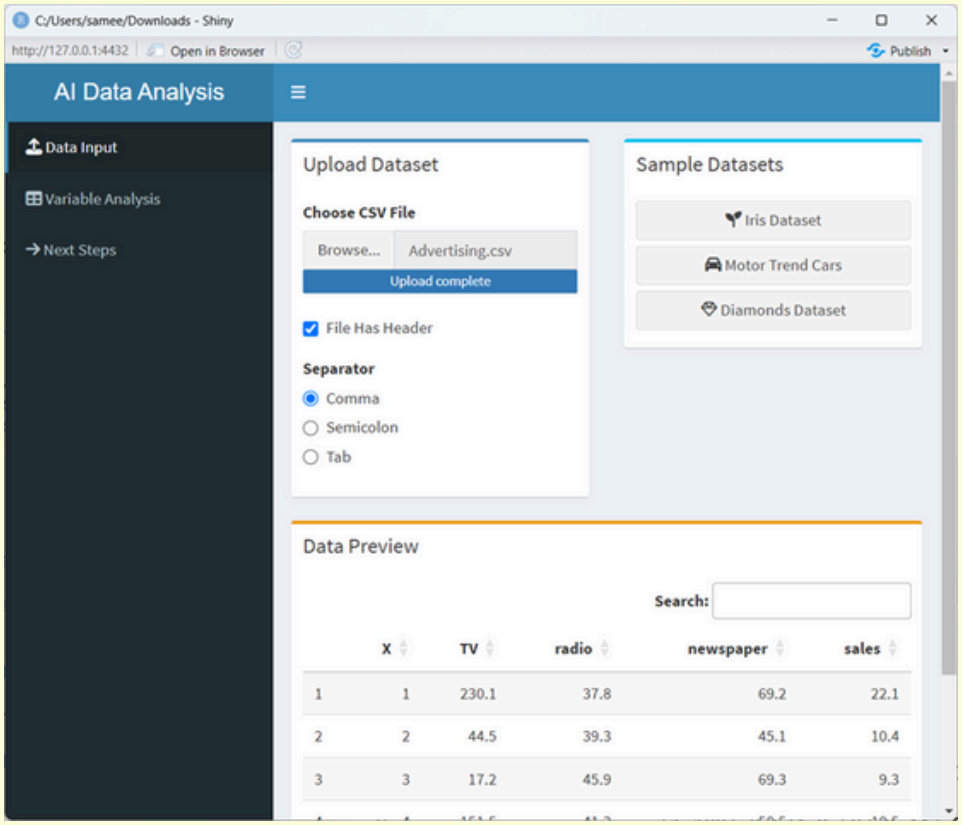


Figure 3: View of the Data Input tab after uploading the dataset

The variable view is as follows. It summarizes the categorical and quantitative variables.

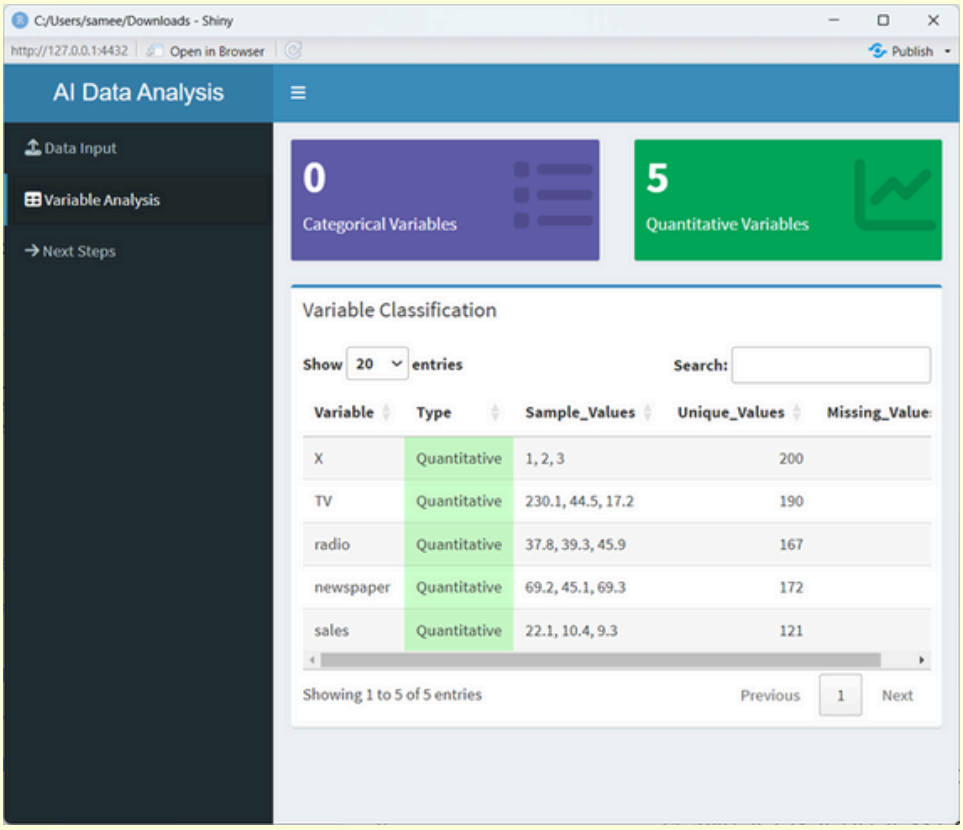


Figure 4: View of the Variable Analysis tab with Advertising dataset

The app generated from a couple of sentences works perfectly fine and it even suggested some next steps.

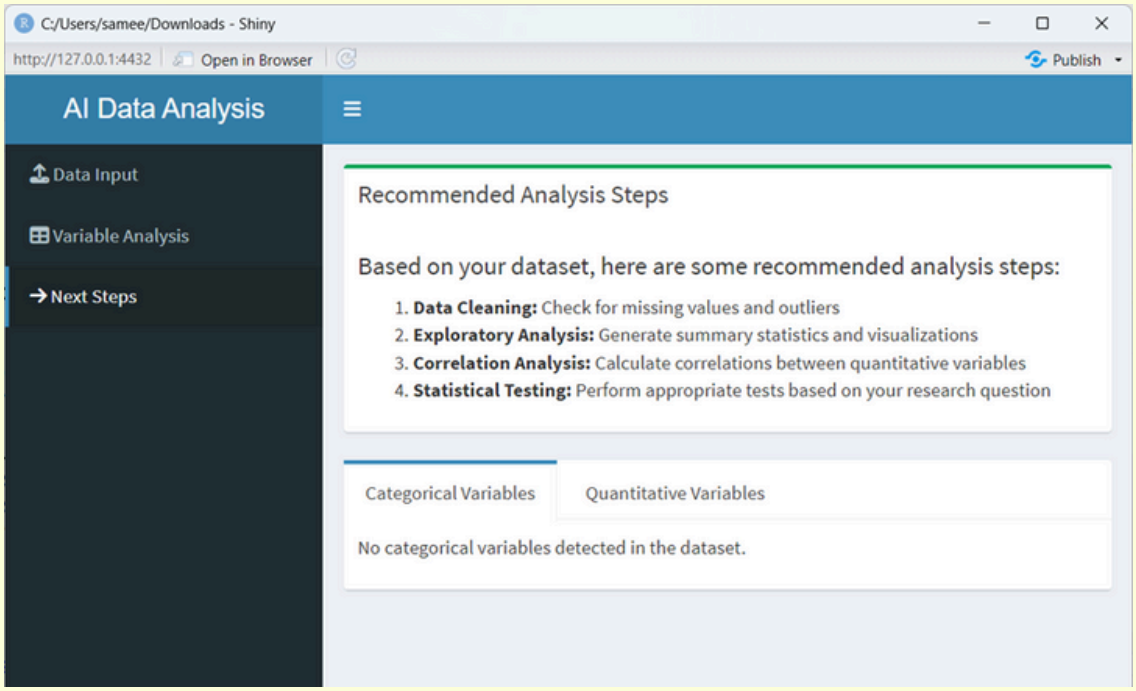


Figure 5: View of the Next Steps tab

As you have seen, we can very easily create an app without writing a single code. Next step is to write a prompt to incorporate some pre-processing steps into the app. One can use a GitHub repository and keep the versions controlled. This will help us to deal with the bugs better. If your app come out good, you can even create a login page, host on a Shiny server, incorporate a reliable payment gateway like Stripe and make some money. Happy vibe coding.

References

Anthropic. (2025). Claude 3.7 Sonnet [Large language model]. Retrieved May 7, 2025, from https://www.anthropic.com

Karpathy, A. (2025). There's a new kind of coding I call "vibe coding", where you fully give in to the vibes, embrace exponentials, and forget that the code even exists... [Tweet]. X. https://x.com/karpathy/]

Revolutionizing Data Analysis with AI: Google's Gemini Agent in Colab

Dr. M.M Mohamed Mufassirin
B.Sc (Hons, SEUSL), M.Sc (Peradeniya), PhD (Griffith, Australia)
Lecturer in Computer Science,
Department of Computer Science, Faculty of Applied Sciences,
South Eastern University of Sri Lanka

1. Introduction

Getting started with data science and artificial intelligence (AI) often involves a time-consuming setup processes. This includes installing and loading necessary libraries, organizing datasets, configuring computing environments, and ensuring compatibility between different tools and frameworks. These steps are important for a smooth workflow, but they can be time-consuming and distracting from the key tasks in data science, such as exploring datasets and finding insights. By reducing the burden of these initial setup tasks, data scientists and AI practitioners can focus more on problem-solving, model development, and innovation, which can ultimately enhance their productivity and the impact of their work.

Moreover, as the demand for data-driven decision-making continues to grow across industries, there is a parallel need for tools that make data science more accessible, efficient, and user-friendly. This has paved the way for the adoption of AI-assisted technologies that help automate repetitive tasks, guide users through complex processes, and translate natural language instructions into actionable codes The ultimate goal is to lower the entry barriers to data science and empower a wider range of professionals, from business analysts to domain experts, to engage in advanced data analysis, even with limited programming knowledge.

In this context, Google's recent introduction of the Gemini-powered Data Science Agent within its Colab Python platform represents a significant milestone. By embedding a conversational AI assistant directly into the collaborative and cloud-based environment of Colab, Google is transforming how users interact with data and build solutions. The agent utilizes the capabilities of Gemini 2.0, Google's most advanced large language model, to understand natural language prompts, generate Python code, create complete Jupyter Notebooks, and offer intelligent assistance throughout the analytical process. This article explores the features, benefits, and potential implications of the innovative tool Data Science Agent, considering how it may play a significant role in shaping the future of data analysis.

2. Google Colab: A Brief Overview

Google Colaboratory, commonly known as Google Colab, is a cloud-based platform that allows users to write and execute Python code directly in their browsers (Bisong, 2019). Launched in 2017, Colab has become a staple for data scientists and researchers, offering free access to computing resources, including GPUs and TPUs. Its integration with Jupyter Notebooks facilitates collaborative coding, data analysis, and machine learning model development without the need for complex local setups (Sherrer & Kara, 2022).

3. Introduction to the Gemini-Powered Data Science Agent

In March 2025, Google unveiled the Data Science Agent, an AI assistant powered by its advanced Gemini 2.0 model, within the Colab platform (Google Developers, 2025). This agent is designed to automate and enhance various aspects of data analysis by interpreting natural language instructions and generating fully functional Jupyter Notebooks. Initially launched for trusted testers in December 2024, the Data Science Agent became publicly available to users aged 18 and above in selected countries and languages in early 2025 (Carl F., 2025). Figure 1 shows a conceptual illustration of Google's Gemini-powered Data Science Agent in Colab, automating data analysis through natural language.

Figure 1: A conceptual illustration of Data Science Agent in Colab.

4. Key Features of the Data Science Agent

a.Natural Language Processing Capabilities

The Data Science Agent powers the multimodal capabilities of the Gemini 2.0 model to understand and process natural language inputs. Users can describe their data analysis objectives in plain English, and the agent translates these descriptions into executable Python code within a Jupyter Notebook. This feature significantly lowers the barrier to entry for individuals with limited programming experience, enabling a broader range of users to perform complex data analyses.

b. Automated Notebook Generation

Upon receiving a natural language description of a data analysis task, the agent automatically generates a comprehensive Jupyter Notebook. This notebook includes data import statements, preprocessing steps, analysis procedures, visualizations, and interpretations, providing a holistic approach to data exploration and insight generation.

c. Integration with Google's AI Ecosystem

The Data Science Agent is seamlessly integrated with Google's broader AI ecosystem. It utilizes the Gemini 2.0 model's enhanced reasoning and coding capabilities to assist users in planning and executing tasks on their computers and the web. This integration ensures that users have access to the latest advancements in AI, facilitating more efficient and effective data analysis workflows.

5. Benefits for Data Scientists and Researchers

a. Enhanced Productivity

By automating routine and complex coding tasks, the Data Science Agent allows data scientists and researchers to focus more on interpreting results and making data-driven decisions. This shift reduces the time and effort required to develop analysis scripts from scratch, thereby accelerating the research process.

b .Accessibility for Non-Programmers

The natural language interface democratizes data analysis by enabling individuals without extensive programming backgrounds to perform sophisticated analyses. Educators, business analysts, and other professionals can leverage the agent to gain insights from data without the steep learning curve traditionally associated with coding.

c. Consistency and Reproducibility

Automated notebook generation ensures that analyses are conducted in a consistent and reproducible manner. This standardization is particularly beneficial in collaborative environments, where multiple stakeholders need to understand and verify the analysis process.

6. Comparison with Existing Tools

Before the introduction of Google's Data Science Agent, tools like OpenAI's ChatGPT Advanced Data Analysis (formerly Code Interpreter) offered similar functionalities. However, Google's agent distinguishes itself through its deep integration with the Colab platform and the broader Google AI ecosystem. This integration provides users with a more cohesive and streamlined experience by utilizing Google's robust infrastructure and AI advancements (Carl F., 2025).

7. Implications for the Future of Data Analysis

The launch of the Gemini-powered Data Science Agent represents a major step forward in the development of data analysis tools. Traditionally, data analysis has required strong programming skills, familiarity with complex software, and a solid understanding of statistical techniques. These requirements have limited access to data science, especially for individuals and organizations without technical expertise.

However, with the introduction of AI-powered tools like the Data Science Agent, many of these challenges can now be overcome. The agent helps users perform data analysis by understanding natural language instructions and automatically generating the necessary code and visualizations. As AI technology continues to improve, it is expected that such tools will become even more powerful, accurate, and easier to use. This advancement will likely encourage a wider range of people to engage in data analysis, including professionals from non-technical fields such as education, healthcare, social sciences, and business.

In the long run, tools like the Gemini-powered agent may help build a more data-literate society, where individuals are better equipped to understand and use data in their daily lives and work. This could lead to more efficient problem-solving, increased innovation, and a stronger connection between data insights and real-world outcomes.

8. Conclusion

Google's Gemini-powered Data Science Agent represents a significant advancement in the democratization and enhancement of data analysis. By combining natural language processing with automated notebook generation, it empowers a diverse range of users to extract meaningful insights from data efficiently. As this technology continues to develop, it holds the promise of reshaping how we approach and utilize data in research, business, and beyond.

References

Bisong, E. (2019). Google Colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_7

Carl, F. (2025, March 3). Google launches free Gemini-powered data science agent on its Colab Python platform. VentureBeat. Retrieved from https://venturebeat.com/ai/google-launches-free-gemini-powered-data-science-agent-on-its-colab-python-platform

Google Developers. (2025, March 6). Data science agent in Colab with Gemini. Google Developers Blog. https://developers.googleblog.com/en/data-science-agent-in-colab-with-gemini/

Sherrer, Kara (May 25, 2022). "Google Colab vs Jupyter Notebook: Compare data science software". TechRepublic. Retrieved from https://qz.com/1113999/nerds-rejoice-google-just-released-its-internal-tool-to-collaborate-on-ai

How do statistics shape decisions in finance, business, and policy?

Ms. Jithmi Tharushika
Transfer Pricing Analyst
Ernst & Young Sri Lanka

Statistics has evolved to be a critical yet invisible component driving business strategies, financial analysis, economic forecasting, accounting, auditing, fraud detection and policy making globally.

This article explores how statistical methods guide crucial decisions across finance, business operations, and government policy, with real-world examples that demonstrate their profound and growing influence.

1. Financial decision-making: managing risk and optimizing investments

In the financial world, uncertainty is the norm. Predicting market movements, assessing investment risks, and ensuring financial stability all depend heavily on statistical models.

Statistical techniques such as regression analysis and time series modeling are primarily used in making a wide variety of investment decisions. For instance, portfolio managers utilize statistical methods like Capital Asset Pricing Model ("CAPM") to evaluate expected returns and manage risks effectively. Similarly, volatility modeling techniques such as ARCH and GARCH models are integral in financial econometrics, aiding traders and analysts in understanding and forecasting market movements.

The banking industry relies heavily on statistical models for the risk management practices. Credit scoring models like logistic regression or machine learning algorithms help banks and lending institutions assess the creditworthiness of borrowers, thereby minimizing default risks. Accounting professionals leverage statistical tools in auditing and fraud detection. Methods such as anomaly detection, Benford's Law, and various sampling techniques are routinely employed to identify irregularities and ensure compliance with financial reporting standards.

Risk management and VAR models

Banks and investment firms use Value-at-Risk ("VAR") models to estimate the maximum potential loss on an investment portfolio over a given time frame at a certain confidence level. These models rely on historical market data and statistical simulations such as Variance-Covariance Method and Monte Carlo simulations.

Example: JPMorgan Chase

JPMorgan Chase pioneered a firm-wide value-at-risk system during the late 1980s. This modeled several hundred key factors and a covariance matrix was updated quarterly from historical data. From this, the standard deviation of portfolio value was calculated subsequently employing various value-at-risk metrics such as 95% USD value-at-risk, which was calculated using an assumption that the portfolio's value was normally distributed.

With this value-at-risk measure, J.P. Morgan replaced a cumbersome system of notional market risk limits with a simple system of value-at-risk limits. [1]

Today, it models thousands of hypothetical scenarios daily, allowing the firm to set capital reserves and make real-time trading decisions while complying with regulatory stress tests.

Portfolio optimization

Statistical methods like Mean-Variance Optimization help investors build portfolios that maximize returns for a given level of risk. Mean variance analysis was introduced by the economist Harry Markowitz and is used to combine a portfolio of assets in order to maximize the expected return under a given level of risk. Based on the outputs from the model, investors are able to choose the best combination of asset risk and return based on the assessment of individual risk tolerance, to obtain the most ideal results.

Example: Blackrock's Aladdin platform

Aladdin platform is a software system built by Blackrock Solutions. In 2013, it handled about USD 11 trillion in assets, which was about 7% of the world's financial assets, and kept track of about 30,000 investment portfolios. As of 2020, Aladdin managed USD 21.6 trillion in assets[2]. The system processes millions of data points daily, using statistical forecasting and optimization techniques to help manage the asset portfolios, allowing clients to know what they own across their portfolio, identify opportunities, and make more informed decisions.

2. Business strategy: predicting markets and customer behavior

Beyond finance, statistics shapes corporate strategies by turning raw data into actionable insights. Systematic collection, analysis, interpretation and presentation of data relevant to business operations serves as a critical tool for organizations to gain insights into their performance, market dynamics, and customer behavior. By applying various statistical methods and techniques, businesses uncover patterns, trends, and relationships within their data, enabling them to make informed decisions, set goals, and optimize processes.

Predictive analytics for customer behavior

Businesses use statistical models to forecast customer needs, purchasing patterns, and churn risk. Techniques like logistic regression, time-series forecasting, and survival analysis enable companies to anticipate behavior before it happens.

Example: Amazon's dynamic pricing

Amazon uses statistical models to adjust product prices every 10 minutes. By analyzing historical purchase data, competitor prices, inventory levels, customer behavior and demand elasticity, Amazon optimizes prices in real-time to maximize revenue while maintaining customer loyalty.[3]

Operational optimization

Statistical forecasting models, like Auto-Regressive Integrated Moving Average ("ARIMA"), help businesses manage inventory, production schedules, and supply chain logistics.

Example: Walmart's supply chain analytics

Walmart applies demand forecasting models across thousands of stores, predicting which products will sell fastest and ensuring shelves are stocked efficiently. This statistical foresight reduces wastage and maximizes profitability in a highly competitive retail market.[4]

3. Accounting, auditing, tax and fraud detection: ensuring integrity

Financial audits are no longer about manual checking alone; modern auditing has become increasingly statistical. Accounting professionals leverage statistical tools in auditing and fraud detection. Methods such as anomaly detection, Benford's Law, and various sampling techniques are routinely employed to identify irregularities and ensure compliance with financial reporting standards. This improves transparency and reliability in financial disclosures.

Statistics also play a crucial role in taxation ensuring compliance with local and international tax regulations. Statistical analyses, including benchmarking studies, comparative analysis and descriptive statistics, are employed to justify pricing structures between related entities and defend positions during tax audits.

Audit sampling and statistical testing

Auditors use sampling techniques and hypothesis testing to verify large transaction sets without reviewing each item individually.

Example: KPMG's data-driven audits

Global audit firms like KPMG employ data analytics platforms that use statistical models to flag outlier transactions, deviations from historical trends, or patterns inconsistent with expected business behavior, enabling smarter audit insights and streamlined audit procedures[5].

Fraud detection using Benford's Law

Benford's Law which is the principle that in naturally occurring datasets, smaller digits appear more frequently is a technique used to detect financial fraud.

Example: Forensic accounting in corporate investigations

In corporate investigations, anomalies in financial records, such as an unusually high number of transactions starting with the digit 9often trigger deeper statistical testing and legal scrutiny, leading to the discovery of embezzlement or accounting fraud[6].

Benchmarking and arm's-length analysis

For multinational corporations, setting fair prices for intercompany transactions known as transfer pricing is a major compliance challenge. Statistics plays a pivotal role in justifying these prices to tax authorities worldwide. Corporations conduct benchmarking studies comparing the financial results of their transactions to those of independent, comparable companies. Statistical measures like interquartile ranges ("IQR") are used to identify acceptable price ranges. In cases where pure comparable data aren't available, companies apply multiple regression analysis to adjust for differences in markets, functions, or risk profiles, ensuring defensible pricing structures.

Example: OECD BEPS compliance

Under the OECD's Base Erosion and Profit Shifting ("BEPS") Action Plan 13, companies must defend their transfer pricing using statistically sound comparability analyses. Firms like Amazon, Apple, and Starbucks have invested heavily in global tax compliance teams that prepare detailed statistical documentation to minimize disputes and penalties[7].

4. Policymaking and public finance: evidence-based governance

At the macroeconomic level, statistics are fundamental for policy making. Governments and central banks utilize economic indicators derived from statistical analyses, such as GDP growth, unemployment rates, inflation rates, and consumer confidence indices, to craft monetary and fiscal policies. Predictive analytics, particularly time series forecasting, enables policymakers to predict economic downturns and make informed decisions to ensure economic stability.

Economic forecasting and fiscal policy

Macroeconomic models based on statistical methods forecast future economic conditions, helping set interest rates, tax policies, and government spending.

Example: The U.S. Congressional Budget Office ("CBO")

The CBO uses statistical models like Bayesian vector autoregression ("BVAR") structural vector autoregressions ("SVAR") and Monte Carlo simulations to generate economic predictions, to predict the impact of proposed tax reforms or healthcare changes, influencing the legislative decisions[8],[9].

5. Emerging trends: The Future of statistical decision-making

Globally, the use of statistics in finance and business is evolving rapidly, driven by technological advancements and data availability. The integration of statistical methods with big data analytics has enabled financial institutions to handle massive data sets, improving decision-making accuracy and efficiency. Platforms using machine learning algorithms are increasingly adopted for high-frequency trading, real-time risk assessment, and dynamic pricing strategies. Investors and regulators are increasingly using statistical Environmental, Social, Governance ("ESG") scores to drive sustainable finance and corporate governance reforms.

Additionally, the adoption of artificial intelligence ("AI") and machine learning ("ML") has significantly augmented the role of statistics. Predictive modeling and data-driven decision-making processes are now fundamental in financial forecasting, fraud detection, customer relationship management, and strategic planning. Statistical methods combined with AI are reshaping how businesses operate and compete globally.

6. Conclusion: Statistics as a driver of intelligent decisions

Across finance, business, and policy, statistics have become a frontline strategic asset which enables organizations to manage uncertainty, identify opportunities, mitigate risks, and deliver better outcomes.

As the financial world becomes increasingly data-driven, mastery of statistical applications will continue to offer competitive advantages, guiding effective decision-making and contributing to robust economic growth globally.

References

[1] Value-at-Risk - Theory and Practice -Second Edition Glyn. A. Holton
[2] Aladdin® by BlackRock - software for portfolio management
[3] How Amazon Uses Real-Time Data and Dynamic Pricing to Maximize Profits - Pricefy Blog
[4] Retail Supply Chain Systems Analysis: A Case of Walmar
[5] What is a Data-driven Audit?
[6] Benford's Law As a Useful Tool to Determine Fraud in Financial Statements
[7] Transfer Pricing Documentation and Country-by-Country Reporting, Action 13 - 2015 Final Report | OECD
[8] Conditional Forecasting With a Bayesian Vector Autoregression
[9] Structural Vector Autoregressive Models | Oxford Research Encyclopedia of Economics and Finance

The Hidden Algorithms of Human Fitness

Ms. Naethree Premnath
Fourth Year Undergraduate, BSc (Hons) in Data Science,
Department of Statistics,
University of Colombo

Introduction

In a world where 1.8 billion adults do not move enough to stay healthy, as highlighted by a 2024 Lancet Global Health study (Flaxman et al., 2024), the need to bridge fitness and technology has become more urgent than ever. In response to this challenge, a project was undertaken to explore how physical performance could be decoded using machine learning methods. The goal was set to transform physical metrics into meaningful insights that could redefine how fitness is understood and improved by individuals.

Dataset Description

The analysis was based on the 'Body performance Data' dataset, obtained from Kaggle (Kukuroo, 2023), which contained records from 13,393 individuals aged between 20 and 64 years. The dataset appears to originate from South Korea, potentially linked to national fitness assessments supported by the Korea Sports Promotion Foundation (KSPO), a government-funded organization established to promote public health and sports participation across the country . It includes a variety of attributes such as height, weight, gender, systolic and diastolic blood pressure, as well as performance metrics like sit-up counts, broad jump distances, grip strength, and sit-and-reach flexibility scores. Each individual is categorized into one of four performance levels, A, B, C, or D, where A represents the highest level of physical performance and D the lowest. These measurements and classifications provide a rich and diverse basis for assessing physical fitness across a broad adult population, encompassing both medical indicators and physical performance outcomes.

Data Preprocessing

1. Handling Outliers

As is often the case with real-world datasets, the raw data contained several anomalies. During the initial stages of exploration, some unrealistic entries were identified. For instance, there was a record of a 21-year-old male with a body fat percentage of 78.4%, which is physiologically implausible given his height and weight. Additionally, there were multiple instances of blood pressure readings recorded as 0/0, which are not biologically possible and likely resulted from faulty measurement equipment. To preserve the reliability and credibility of the analysis, such outliers were carefully cross-referenced with medical literature and subsequently removed from the dataset.

2. Feature Engineering

To enhance the dataset and extract deeper insights, new variables were engineered. Mean Arterial Pressure (MAP) was calculated using a standard formula that considers both systolic and diastolic blood pressure. This provided a more holistic view of cardiovascular health than either measure alone. Furthermore, Body Mass Index (BMI) was included as a variable despite its well-known limitations, such as its inability to distinguish between muscle and fat mass. Nevertheless, it remained a broadly accepted and useful measure for evaluating general body composition across large populations.

Exploratory Data Analysis

1. Key Patterns and Insights

Descriptive analysis revealed several notable patterns within the data. Men generally exhibited superior performance in grip strength and broad jumps by Figure 1 and Figure 2, while women tended to perform better in flexibility tests such as the sit-and-bend-forward as shown in Figure 3.

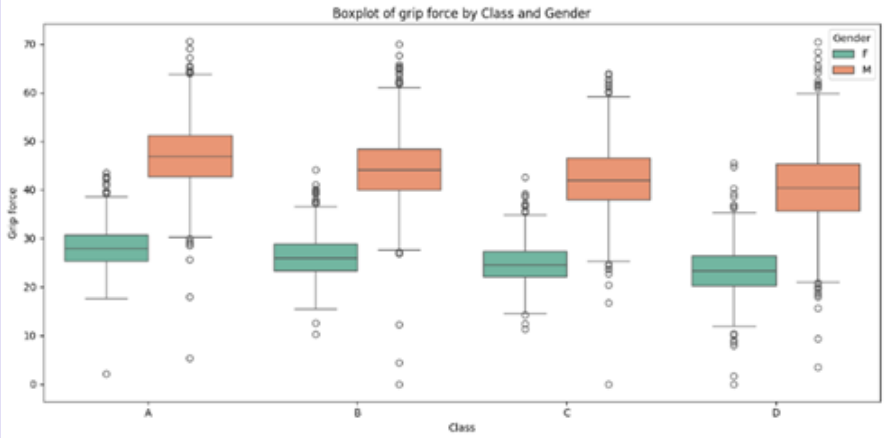


Figure 1: Box plot of Grip strength by Gender

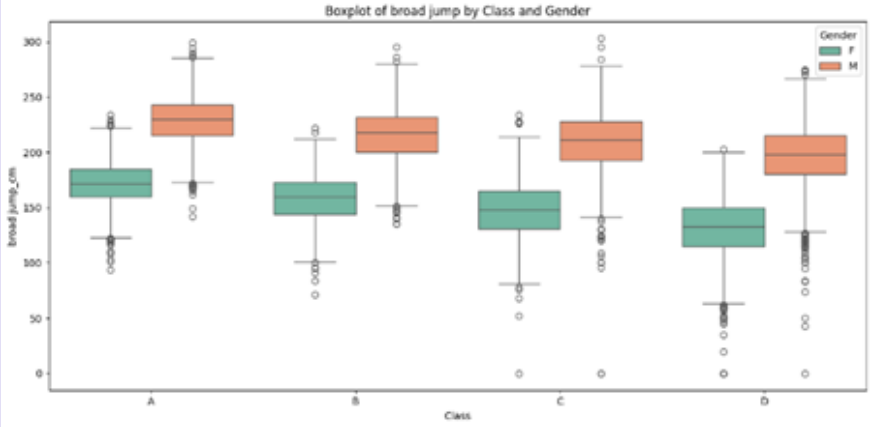


Figure 2: Box plot of Broad jump by Gender

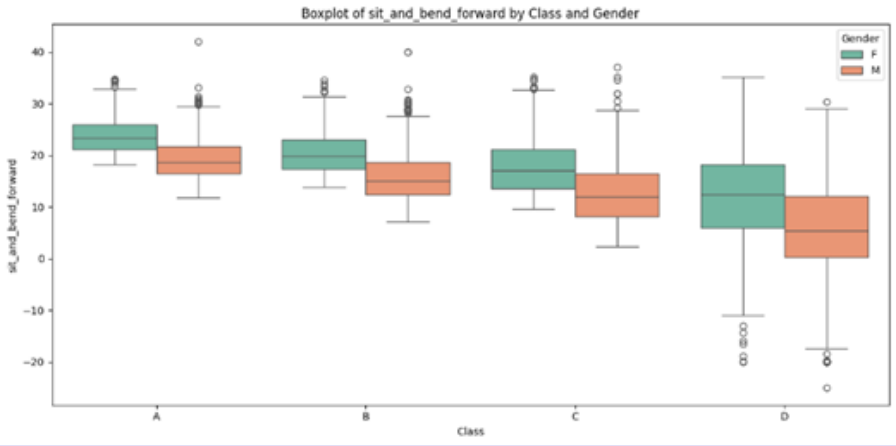


Figure 3: Box plot of Sit and bend forward by Gender

When analyzing the distribution across performance classes, individuals in "Class A," representing the highest level of physical fitness, demonstrated exceptional performance in sit-and-bend-forward, sit-ups and broad jumps by Figure 4.

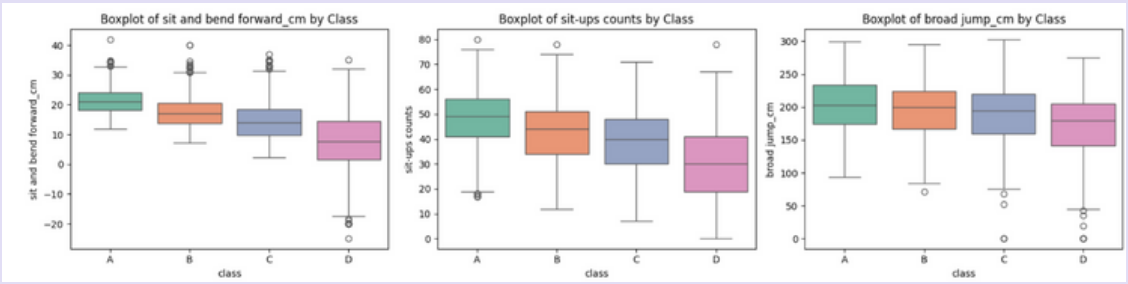


Figure 4: Box plots of physical metrics by Class

Further analysis showed that higher BMI and body fat percentages were generally associated with poorer performance outcomes in physical tests. However, there were exceptions, particularly among muscular individuals whose high BMI was due to increased muscle mass rather than excess fat. This complexity highlighted the nuanced relationship between body composition and physical performance.

2. Challenges Identified

Despite the richness of the dataset, several challenges emerged during the analysis. One significant issue was multicollinearity, where certain variables, such as BMI and body fat percentage, were highly correlated. This posed a risk of skewing model interpretations and needed to be carefully addressed. Additionally, Partial Least Squares Discriminant Analysis (PLS-DA) indicated considerable overlap between the different performance classes, making it difficult to draw sharp distinctions between them. This overlapping nature of the data introduced an added layer of complexity to the classification tasks.

Model Development

1. Selection of Algorithms

To tackle the classification problem, a variety of machine learning models were tested. Regularized models such as Logistic Ridge Regression and Lasso Regression were employed initially to manage multicollinearity by applying penalty terms that shrink the coefficients of correlated features. These methods provided more stable and interpretable models. However, non-linear ensemble methods such as Gradient Boosting, XGBoost, and Random Forest were also explored, given their ability to model complex, non-linear relationships within the data. While these ensemble models showed promising initial results, they also tended to overfit the training data, performing significantly worse on unseen test data.

2. Model Tuning and Feature Selection

To address issues of overfitting and improve the generalizability of the models, hyperparameter tuning was conducted using a grid search approach. In addition, feature importance analysis revealed that Mean Arterial Pressure (MAP) had minimal influence on model performance as shown in Figure 5. Consequently, it was removed from the final feature set to simplify the model and reduce noise. These adjustments helped refine the models and improve their real-world performance.

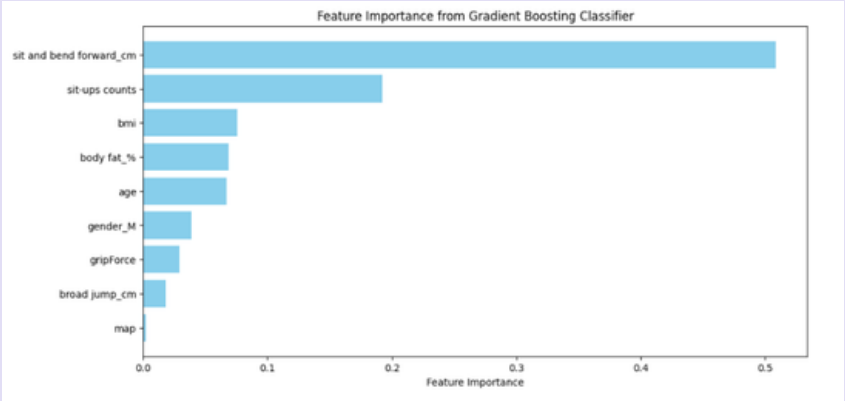


Figure 5: Feature Importance Graph

3. Final Model Performance

After extensive experimentation and tuning, the Gradient Boosting Classifier emerged as the best-performing model. The optimal parameters included 'learning_rate': 0.1, 'max_depth': 5, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 200. The model achieved an accuracy of approximately 75% on the unseen test data. Precision, recall, and F1 scores were consistent across both training and test datasets, indicating good generalization ability. The confusion matrix confirmed reliable classification performance across all four body performance categories as shown in Figure 6.

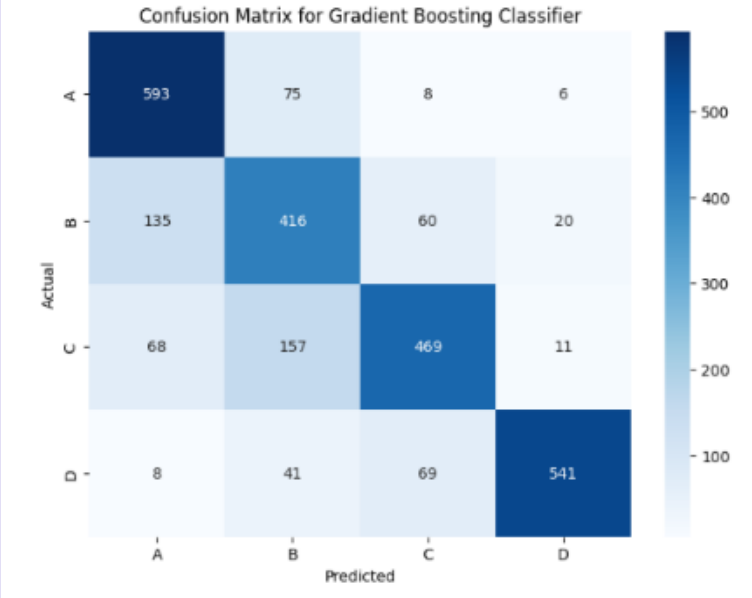


Figure 6: Confusion Matrix for the Best Model

The final model performance is summarized in the Table 1.

Table 1: Performance Metrics of the Gradient Boosting Model on Training and Test Sets

	Training Set				Test Set			
	Precision	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**
Gradient Boost	78%	77%	77%	77%	76%	75%	75%	76%

Practical Application: FitIQ

Building upon the findings from the model development phase, a user-oriented application called FitIQ was developed. FitIQ allows users to input their basic physical measurements and obtain an estimated performance class based on the trained Gradient Boosting model. Beyond classification, the platform also offers personalized fitness recommendations tailored to the user's specific needs, providing actionable insights for improving their fitness levels.

The demonstration of FitIQ can be accessed via the QR code found in Figure 7.

Figure 7: QR Code of Demo of FitIQ

Conclusion

In a time when health and fitness are more critical than ever, tools like FitIQ represent a meaningful step forward in personalizing fitness guidance through data-driven methodologies. By combining the predictive power of machine learning with real-world applications, this project highlights the immense potential for technology to enhance personal health and well-being. As machine learning models continue to evolve, and as more comprehensive data becomes available, the integration of data science into everyday health practices promises to open new avenues for individuals seeking to lead healthier, more informed lives.

Acknowledgements

This project would not have been possible without the contributions and dedication of the team members. Sincere appreciation is extended to:

Aslam Brantha
Fourth Year Undergraduate
BSc (Hons) in Statistics
Department of Statistics
University of Colombo

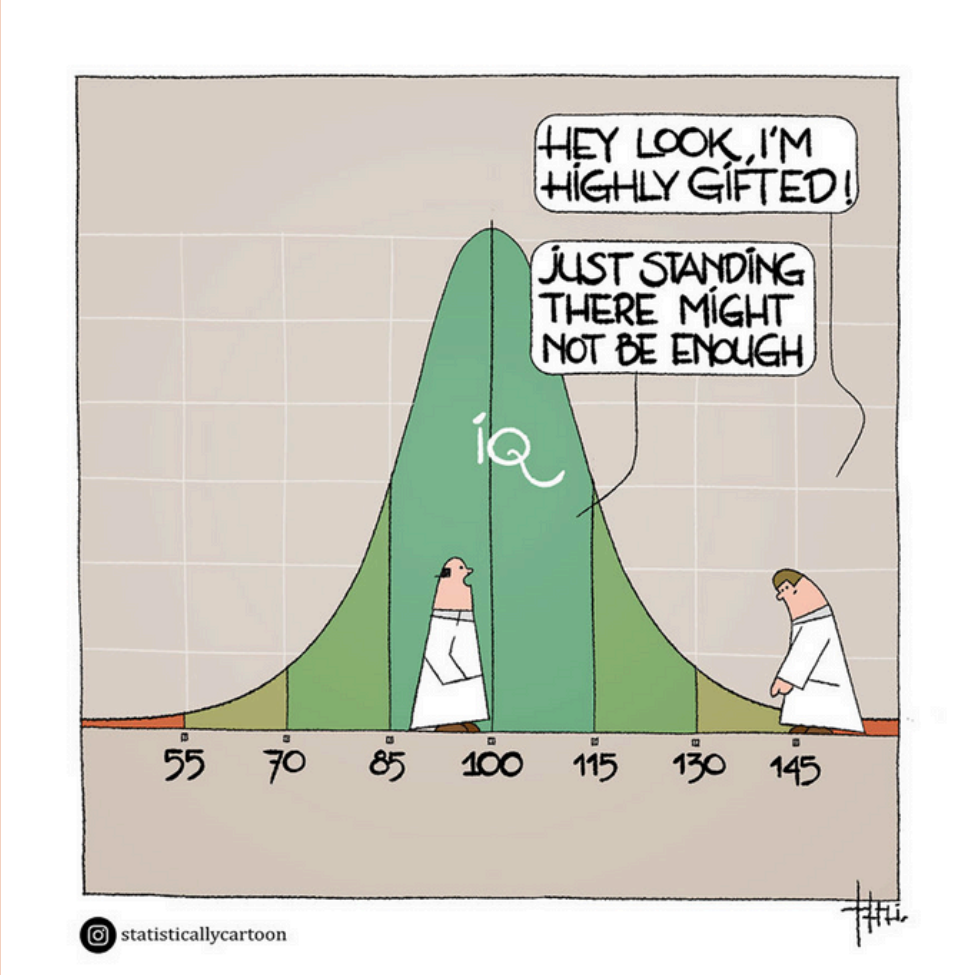
Dinula Ramuditha
Fourth Year Undergraduate
BSc (Hons) in Statistics
Department of Statistics
University of Colombo

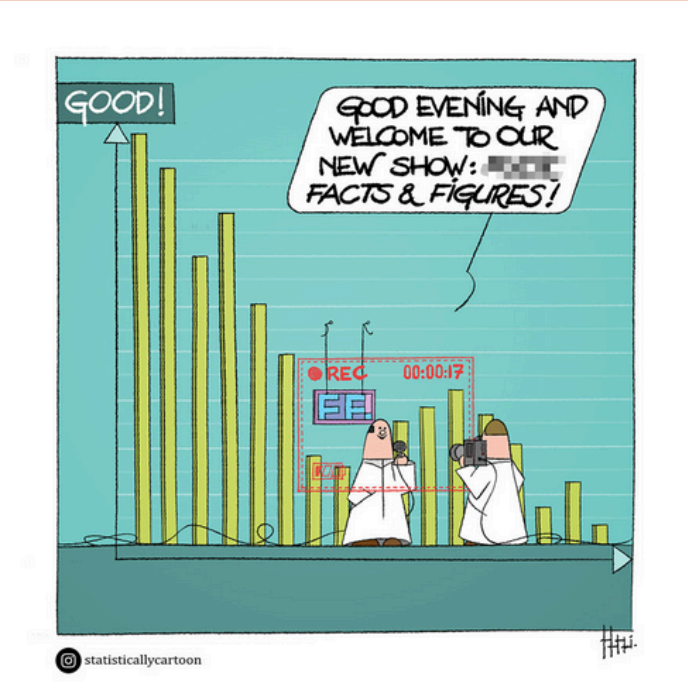
References

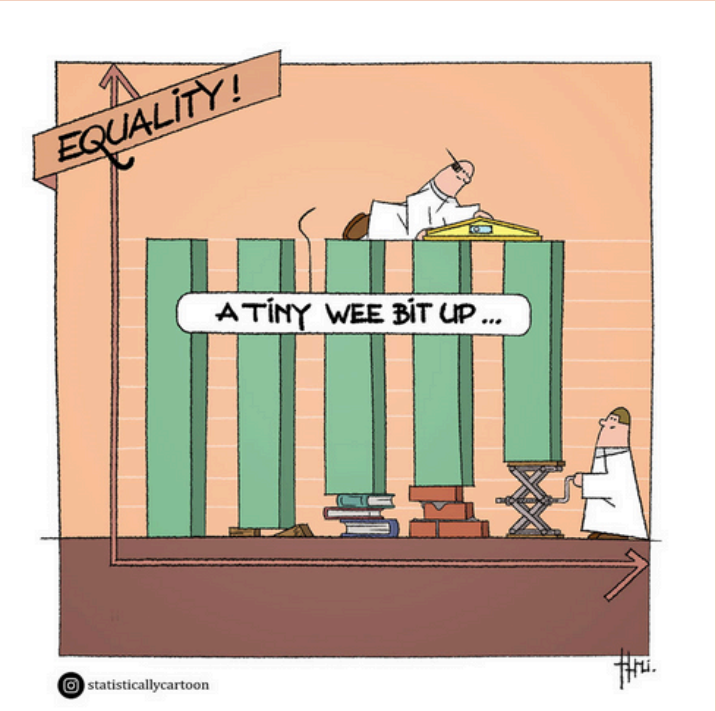
Flaxman, S., Semenova, E., Health, A., Strain, T., Flaxman, S., Guthold, R., Semenova, E., Cowan, M., & Riley, L. M. (2024). National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5Ã‚Â·7 million participants. The Lancet Global Health, 12, e1232–e1243. https://doi.org/10.1016/S2214-109X(24)00150-5

Kukuroo. (2023). Body performance data [Data set]. Kaggle. https://www.kaggle.com/datasets/kukuroo3/body-performance-data

STAT COMICS







Source: https://www.boredpanda.com/statistically-insignificant-comics-raf-schoenmaekers/
Accessed date: 29th April 2025

13th Annual General Meeting 2025

The 13th Annual General Meeting of the Institute of Applied Statistics Sri Lanka was held on 6th April 2025 at the Auditorium of the Professional Centre (OPA). All the life Members were Invited for the occasion. The New Executive council was appointed at the AGM for the year 2025.

President/ IASSL

Dr. Rajitha M. Silva
Department of Statistics,
Faculty of Applied Sciences,
USJ

Immediate Past President, Chairperson- Academic Training Committee/ IASSL

Dr. Niroshan Withanage
Department of Statistics,
Faculty of Applied Sciences,
USJ

Vice President, Chairperson- House & Finance MC/ IASSL

Dr. A.P.G.S. De. Silva
Former Director, Department of Census and Statistics, Sri Lanka

Secretary/ IASSL

Dr. Wasana Wijesuriya
Freelance Statistical Consultant

Treasurer/ IASSL

Dr. Priyath De Silva
Former Vice President / OPA
Former Director Investments,
Ministry of Nation Bulting

Editor/ IASSL

Dr. Chathuri Jayasinghe
Department of Statistics,
Faculty of Applied Sciences,
USJ

Assistant Secretary

Prof. Kapila T. Rathnayake

Department of Physical Sciences and Technology ,Faculty of Applied Sciences, SUSL

Assistant Treasurer

Mrs. D.A.B.N. Amarasekara

Department of Crop Science, Faculty of Agriculture, UOR

Associate Editor

Dr. K.A.D. Deshani

Department of Statistics, Faculty of Sciences, UOC

Chairperson/ Statistics Popularization Committee

Mr. H.A.C.S. Hapuarachchi

Department of Sports Sciences & Physical Education,Faculty of Applied Sciences SUSL

Chairperson/ Research & Development Committee

Dr. Chitraka Wickramarachchi

Department of Statistics, Faculty of Applied Sciences, USJ

Executive Council Member

Prof. S. Samita

Department of Crop Science, Faculty of Agriculture, UOP

Executive Council Member

Prof. T.S.G. Peiris

Department of Mathematics & Statistics, Faculty of Humanities and Sciences, SLIIT

Executive Council Member

Prof. T. Sivananthawerl

Department of Crop Science, Faculty of Agriculture, UOP

Executive Council Member

Prof. C. D. Tilakaratne

Department of Statistics, Faculty of Science, UOC

Executive Council Member

Mr. E.R. Anjana D. Bandara

Department of Statistics, Faculty of Science, UOC

Executive Council Member

Dr. Mahasen Dehideniya

Department of Statistics & Computer Science,Faculty of Science,UOP

Executive Council Member

Dr. Neluka Devpura

Department of Statistics, Faculty of Applied Sciences, USJ

Executive Council Member

Dr. G. A. C. N. Priyadarshani

Department of Statistics, Faculty of Science, UOC

Ex Officio

Department of Census and Statistics, Sri Lanka.

Subcommittee Members

Academic Training Committee

1. *Dr. Niroshan Withanage (Chairperson)* , Department of Statistics, Faculty of Applied Sciences, USJ.
2. *Mrs. D.A.B.N. Amarasekara, (Assistant Treasurer) (Ex Officio)* , Department of Crop Science, Faculty of Agriculture, UOR
3. *Dr. Neluka Devpura* , Department of Statistics, Faculty of Applied Sciences, USJ
4. *Dr. Dushan Kumarathunga,* Department of Agricultural Biology, Faculty of Agriculture, UOR
5. *Dr. Shanika Thathsarani,* Department of Economics and Statistics ,Sabaragamuwa University of Sri Lanka
6. *Dr. Mahasen Dehideniya* ,Department of Statistics & Computer Science,Faculty of Science,UOP
7. *Dr. Malima Attapattu* , Department of Statistics & Computer Science,Faculty of Science,UOP
8. *Ms. Menasha Senanayaka* , Department of Mathematics, Faculty of Applied Sciences, USJ

Editorial Board

1. *Dr. Chathuri Jayasinghe (Editor)* , Department of Statistics, Faculty of Applied Sciences, USJ
2. *Dr. K.A.D. Deshani (Associate Editor)* , Department of Statistics, Faculty of Science, UOC
3. *Prof. C. D. Tilakaratne* , Department of Statistics, Faculty of Science, UOC
4. *Dr. Jagath Senarathne,* Department of Statistics & Computer Science, Faculty of Science, UOP
5. *Dr. G. A. C. N. Priyadarshani* , Department of Statistics, Faculty of Science, UOC
6. *Dr. Lakshika S. Nawarathna* , Department of Statistics & Computer Science, Faculty of Science, UOP
7. *Ms. Menasha Senanayaka* , Department of Mathematics, Faculty of Applied Sciences, USJ

House & Finance Management Committee

1. *Dr. A. P. G. S. De Silva (Vice President/Chairperson)*, (Retired) Department of Census and Statistics, Sri Lanka
2. *Dr. Wasana Wijesuriya (Secretary) (Ex Officio)* Freelance Statistical Consultant
3. *Dr. Priyath De Silva, (Treasurer) (Ex Officio)* , Former Vice President OPA
4. *Mrs. D.A.B.N. Amarasekara (Assistant Treasurer) (Ex Officio),* Department of Crop Science, Faculty of Agriculture, UOR
5. *Prof. Kapila T. Rathnayake* , Department of Physical Sciences and Technology ,Faculty of Applied Sciences, SUSL
6. *Eng. S. Vaikunthan,* AMIE (SL)
7. *Dr. T.U.S. Peiris* , Department of Nutritio, Faculty of Livestock Fisheries and Nutrition,Wayamba University of Sri Lanka
8. *Dr. Sachith P. Abeysundara* , Department of Statistics & Computer Science, Faculty of Science, UOP
9. *Dr. N.A.D. N. Napagoda* ,Department of Mathematical Sciences, Faculty of Applied Sciences, WUSL
10. *Prof. S. Samitha* , Department of Crop Science, Faculty of Agriculture, UOP

Statistics Popularization Committee

1. *Mr. H.A.C.S. Hapuarachchi (Chairperson),* Department of Sports Sciences & Physical Education,Faculty of Applied Sciences SUSL
2. *Mrs. D.A.B.N. Amarasekara (Assistant Treasurer) (Ex Officio),* Department of Crop Science, Faculty of Agriculture, UOR
3. *Prof.LD.B. Suriyagoda,* Department of Crop Science, Faculty of Agriculture, UOP
4. *Dr. Chathuri Jayasinghe* ,Department of Statistics, Faculty of Applied Sciences, USJ
5. *Prof. T. Sivananthawer,* Department of Crop Science, Faculty of Agriculture, UOP
6. *Dr. Hemalika T.K. Abeysundara* , Department of Statistics & Computer Science, Faculty of Science, UOP
7. *Mr. Dilan Rodrigo* , Visiting Academics at The Open University of Sri Lanka
8. *Mr. Thilina Withanage* , Department of Science for Technology, Faculty of Technology,USJ

Research & Delevelopment Committee

1. *Dr. Chitraka Wickramarachchi (Chairperson)*, Department of Statistics, Faculty of Applied Sciences, USJ
2. *Mrs. D.A.B.N. Amarasekara (Assistant Treasurer) (Ex Officio),* Department of Crop Science, Faculty of Agriculture, UOR
3. *Mr. D. C. A. Gunawardene,*(Retired) Department of Census and Statistics, Sri Lanka
4. *Mr. Nandana Gunarathne,* ASMP (The World Bank Project)
5. *Prof . Siddhisena,* Professor Emeritus, Department of Demography, UOC
6. *Dr. A. P. G. S. De Silva* , (Retired) Department of Census and Statistics, Sri Lanka.
7. *Mr. Priyadarshana Dharmawardena,*Senior Statistician Department of Census and Statistics, Sri Lanka
8. *Mr. Thilina Withanage* , Department of Science for Technology, Faculty of Technology,USJ

Best Research Awards – IASSL 2024: Celebrating Excellence in Statistical Research

The Institute of Applied Statistics Sri Lanka (IASSL) proudly concluded its prestigious Best Research Awards – 2024 competition, recognizing outstanding contributions to the field of Statistics and Applied Statistics. The awarding ceremony was held on 6th April 2025 at the Organization of Professional Associations (OPA) premises in Colombo, drawing together a distinguished audience of academics, researchers, and students.

This annual competition, organized by IASSL, aims to encourage excellence in statistical research among undergraduates, postgraduates, and independent researchers. This year's awards were based on research completed during 2024, and submissions were accepted under three categories: Undergraduate, Postgraduate, and Open. The response to the call for applications was highly encouraging with 30 applications for the undergraduate category, 12 applications for the postgraduate category and 12 research articles for the open category.

Research submitted to the Undergraduate and Postgraduate categories was limited to those who had completed their degree requirements at Sri Lankan universities in 2024. Meanwhile, the Open category considered published research papers that appeared in reputed journals or were presented at recognized conferences between January 1 and December 31, 2024.

The awards were presented in a formal ceremony that highlighted the relevance and growing importance of Statistics in modern research, policy-making, and industry applications. The IASSL Best Research Awards competition continues to serve as a platform for promoting quality research and fostering innovation in the statistical sciences in Sri Lanka.

Undergraduate Category

Winner
Mr. R.A.S. Ouchithya

Winner
Ms. D.D.H. Tharushika

1st Runner up
Ms. G.W.S. Kavindika

2nd Runner up
Ms. N.H. Wanigasingha

Postgraduate Category

Winner
Ms. S. N. Kalutharage

1st Runner up
Ms. A.F. Thahara

2nd Runner up
Ms. Minasha Abeykoon

Open Category

Winner
Ms. Raveesha S. Illesinghe

1st Runner up
Mr. J.K.S. Sankalpa

2nd Runner up
Mr. Hasanka Wijesundara

Undergraduate Merit Awards

Ms. O.V. Kulasekara

Ms. A.G. S.N. Karunarathne

Ms. J.A.Y.R. Jayasekara

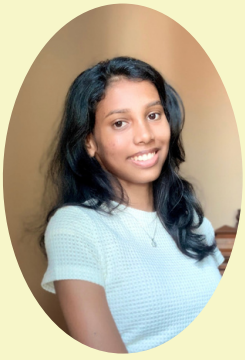
Ms. R. Irshani

Ms. S.A.D.N. Prabodini

Ms. O.M.D.N.W. Opatha

Mr. T.D.S. Ariyarathna

Mr. R. Hariskar



Ms. U.S. Wickramasinghe

Ms. A.M.P. Rajapakse

Members' Achievements

Dr. Thiyanga Talagala, Senior Lecturer in the Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura (USJ), has been ranked first in the prestigious Short Talks, Big Impacts Women in Science Spotlight Competition. This competition, organized by UNESCO – Organization for Women in Science for the Developing World (OWSD), included participants from the Asia-Pacific and Arab region, Latin America and the Caribbean region, and Africa region.

Her talk, Exploring Artistic Legacy: Analyzing Artifact Differences Across Kingdoms Using Machine Learning, showcased how machine learning can uncover hidden mathematical structures in historical paintings and artefacts. Drawing inspiration from Sri Lanka's rich artistic heritage—such as the moonstone, the intricate wooden carvings at Ambekke, and the vibrant wall paintings of ancient kingdoms—her research explores the deep connections between geometry, symmetry, and artistic expression. By analyzing visual elements like colour, pattern, and composition, she aims to identify unique artistic signatures across different civilizations, offering a new perspective on Sri Lanka's cultural history through data-driven insights.

Dr. Talagala emphasized the intrinsic link between art and mathematics, stating that "behind every artist, there is a mathematician, and behind every mathematician, there is an artist," with nature serving as the ultimate example of this harmony. This perspective drives her research at the intersection of data science and cultural heritage.

Congratulations

IASSL Expands Reach of Teacher Training Workshop Series Across Sri Lanka

The Teacher Training Workshop (TTW) series on Probability and Statistics, spearheaded by the Statistics Popularization Committee (SPC) of the Institute of Applied Statistics Sri Lanka (IASSL), has made commendable progress in enhancing mathematics and statistics education nationwide. The preparation of the teacher training manual in all three mediums is progressing concurrently with the ongoing workshop series. These initiatives are part of the efforts to achieve the goals and objectives outlined in the MOU signed last year between IASSL and the Ministry of Education. Dr. Chathuri Jayasinghe served as the key coordinator for this initiative on behalf of IASSL.

The workshop series on the "Probability" component has been successfully conducted across all provinces, with the exception of the Northern and Eastern Provinces during the past two years. Our qualified resource panel included, Prof. S. Banneheka, Dr. Niroshan Withanage, Dr. Rajitha M. Silva, Mr. P. Dias, Dr. D.C. Wickramarachchi, and Dr. Chathuri Jayasinghe.

Meanwhile, the workshop series on "Statistics" component has also begun, and the first workshop was conducted for the Southern Province on the February 21, 2025 in which Dr. Niroshan Withanage and Mr. P. Dias served as resource persons. The IASSL is actively working in coordination with the Ministry of Education (MOE) to roll out these sessions in the coming months for the remaining provinces.

This national initiative reflects IASSL's commitment to capacity building among school teachers, with the ultimate goal of improving statistical education and nurturing analytical thinking among future generations.

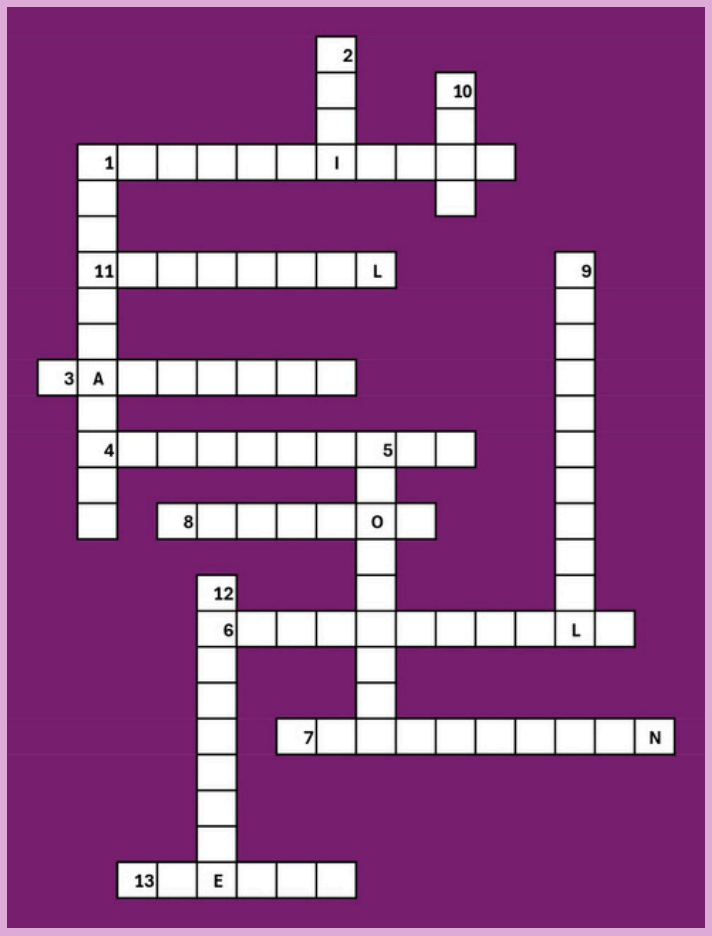
All-Island School Statistics Competition 2025: Inspiring the Next Generation of Statisticians

The Statistics Popularization Committee (SPC) of the Institute of Applied Statistics Sri Lanka (IASSL) has successfully launched the 2025 edition of the All-Island School Statistics Competition, a flagship initiative aimed at nurturing statistical literacy and interest in Statistics among school students nationwide.

The Senior Category examination of the competition was held online on March 16, 2025, and the Junior Category examination was held on March 30, 2025, also via the online platform. The digital format of the event enabled wide participation, with over 500 students from both Junior and Senior levels registering for the competition.

To ensure a smooth and secure examination process, the SPC partnered with an external service provider, eLearning.lk Pvt Ltd., under the coordination of Dr. Navodi Mekhala Hakmanage and Dr. Rajitha M. Silva. Their oversight ensured the technical integrity and operational success of the event. Dr. Chathuri Jayasinghe and Mr. Jasotharan, also contributed greatly to the success of the competition.

Puzzle Competition



Across

1. The probability of an event given that another event has occurred
3. What is a visual display of data with rectangles of varying heights called?
4. In probability, what is the name of the event that cannot occur?
6. This female statistician is famed for her work in nursing, healthcare, and the polar area diagram
7. The entire pool from which a statistical sample is drawn
8. A graphical method for comparing data distributions using quartiles
11. Error in prediction
13. A distribution that is not symmetrical

Down

1. What is the term for the measure of the relationship between two variables?
2. The name of the index used to measure the inequality of income distribution within a population
5. This resampling technique generates multiple samples from a single dataset to estimate the sampling distribution
9. This statistical distribution is often used to model waiting times between events in a Poisson process
10. The difference between the estimator's expected value and the true value of the parameter being estimated
12. Drawing conclusions from data

Please email your submission to appstatsl@gmail.com on or before 20th August 2025.
The draw will be held on the 28th August, 2025.

Correct submissions will be shortlisted and the winners will be selected randomly and will be announced in the Issue 2 of 2025 IASSL newsletter.

WINNERS FROM ISSUE 3, 2024

- **Binali Senaya Aluthge**
- **Gayathri Lakshani Jayawardena**
- **Prabodini Hansika**

SPONSER FOR THE PUZZLE COMPETITION
SG ACTURIAL

ACTUARIAL

"We are proud to announce that we have worked with 16 Organizations from 5 Countries for Consultancy Services and Participants from 105 Organizations in 22 Countries have registered for our Training Programmes"

Saroja Gunatilleke
Managing Director / Actuary

CORPORATE SOLUTIONS

INSURANCE COMPANIES

- Actuarial Consultancy
- Strategic Support
- Internal Audit
- IFRS 17

ALL COMPANIES

- Risk Management Consultancy
- Process Improvements
- Business Plans
- Analysis

ACTUARIAL & RISK MANAGEMENT TRAINING PROGRAMS

Training suitable for Insurance Companies, Banks, Finance Companies, Stock Brokers, Wealth Management Companies, Software Developers, Regulators, Insurance Brokers and Students.

- Actuarial Training for Non-Actuaries
- Risk Management
- Insurance Products from a Risk Angle
- Life Insurance Financial Aspects
- Introduction to Actuarial Concepts
- Practical Applications of Actuarial Science for Students

For more details please scan the QR code or visit us on **www.sgactuarial.com** or call **(+94) 713 146 349**

Courses conducted by IASSL during JAN - APR 2025

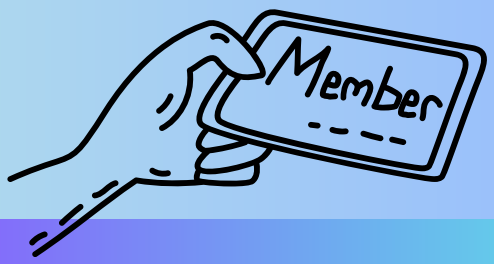
- Certificate course on Data Analysis using Power BI
- Certificate Course on Structural Equation Modeling (SEM) with AMOS & SmartPLS

Upcoming Courses (MAY-AUG 2025)

- Certificate course on Basic Statistics for Managers and Researchers
- Certificate Course on Systematic Literature Review (SLR) with Bibliometric Analysis: a way of manuscript writing with PRISMA
- Certificate course on Essential Training in Learning R
- Certificate course on Essential Training in Data Analysis with IBM SPSS
- Certificate course on Discover the Power of R in Data Analysis
- Certificate course on Qualitative Data Analysis
- Certificate course on Data Analytics with Pythons

New Life Members of IASSL

- **MS. K.S. SAUBHAGYA**
- **MR. C. VIDANAGE**
- **DR. E.D.N. MIHIRANI**
- **MR. V. KUNASEGARAM**
- **MS. N.I.M.B. SENANAYAKA**
- **MR. P.H.R. PRABUDDHA**
- **MR. T.R. WITHANAGE**
- **MR. L.G.D.S. LOKUGAMAGE**



HOW TO BECOME A MEMBER OF IASSL?
Please Visit: www.iassl.lk

Apply for the Membership
Membership Form

Sri Lankan Journal of Applied Statistics (SLJAS)

CALL FOR PAPERS

VOLUME 26

ABOUT SLJAS

SLJAS is an open-access, international, double-blind peer-reviewed journal published by the Institute of Applied Statistics, Sri Lanka (IASSL).

SLJAS ACCEPTS

- Original Research Articles
- Critical reviews including conceptual discussions, Opinions and book reviews.

Authors are warmly invited to submit their papers for publication in SLJAS

THE SCOPE OF THE JOURNAL INCLUDES THE FOLLOWING DOMAINS:

APPLICATIONS OF STATISTICS	THEORETICAL AND METHODICAL ASPECTS OF STATISTICS	DATA SCIENCE AND ANALYTICS

AUTHOR BENEFITS:

- No Article Processing Charges (No APC)
- Fast Track Publication
- Provides Individual DOI (Digital Object Identifier) to each Published Article
- Indexed in Google Scholar

Submit your manuscript via JMS (Journal Management System).

WEB ADDRESS www.iassl.lk
EMAIL editor@iassl.lk
PHONE 011 2 588 291

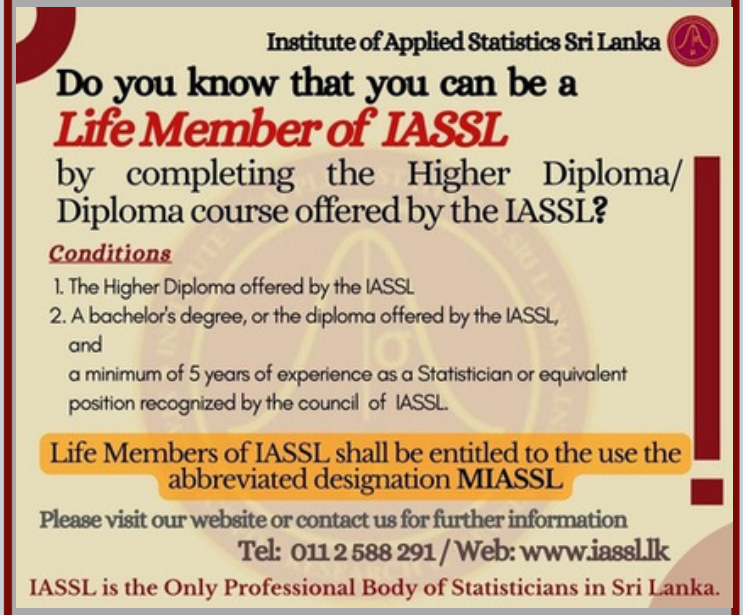
Click on the post to Go ahead....

HOW TO BECOME A MEMBER OF IASSL?

Please visit

http://www.iassl.lk

Apply for the Membership Application form



We are updating Contact Information of IASSL Members :

Please refer below link to update your details:
https://forms.gle/itfDcB5Buedbwp9F7

CONTRIBUTIONS TO THE MAY-AUG (ISSUE 2) 2025 NEWSLETTER:

If you have any submissions, comments, suggestions & or feedback, please send them to editor@iassl.lk.

WE SINCERELY APPRECIATE ALL WHO CONTRIBUTED TO THIS ISSUE, AND THOSE WHO PARTICIPATED IN THE PREPARATION OF IT.

EDITORIAL BOARD/IASSL

IASSL NEWSLETTER

Official Newsletter of the Institute of Applied Statistics Sri Lanka

Institute of Applied Statistics Sri Lanka
The Professional Center 275/75
Prof. Stanley Wijesundara Mawatha
Colombo 07 Sri Lanka

+94 11 2588291
admin@iassl.lk
http://www.iassl.lk
http://www.facebook.com/iassl2020/
https://www.linkedin.com/company/iassl/