ISSN 1391 4395

May – August 2022

IASSL NEWSLETTER

Institute of Applied Statistics Sri Lanka
The Professional Center
275/75
Prof. Stanley Wijesundara Mawatha
Colombo 07
Sri Lanka
+94 11 2588291
appstatssl@gmail.com
http://www.iappstat.lk
http://www.facebook.com/iassl2020/
https://www.linkedin.com/company/iassl/

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

H. G. Wells

Featured Segments

A GUIDE TO RECOMMENDATION

"This article is primarily directed towards students seeking graduate admissions and young recommenders who are new to the task. "

DATA MINING FOR BUSINESS ANALYTICS

"The ability to analyze complex datasets in order to predict future trends opens greater opportunities. "

USAGE OF STATISTICS IN MACHINE LEARNING MODEL MONITORING

"Statistical tests are the key concept behind the ML model monitoring techniques. "

NEWS IN BRIEF

ADVANCING NETBALL THROUGH DATA SCIENCE

"While subjective opinions and improved thinking can drive a team towards a win, assistance from data-driven insights could be helpful to unravel the hidden techniques inside data . "

WHEN INCOMPLETE IS GOOD (ONE-ACT-PLAY)

"Won't there be a problem with pairwise comparisons of treatments? "

REAL-TIME PREDICTION AND STRATEGY FORMULATION IN T20 CRICKET

"What if we bat slowly for the next five overs without losing wickets, and scoring only two boundaries?"

ANNOUNCEMENTS

SUDOKU PUZZLE COMPETITION
UPCOMING EVENTS
UPCOMMING COURSES

IASSL President's Message

Dr. D.C. Wickramarachchi
President/IASSL

Welcome to this new issue of IASSL newsletter and it is my honor to write this issue's opening letter. It is pleasing to mention that IASSL's progress during May-August was highly satisfactory as all the ambitious goals that were intended to achieve were successfully met.

Possessing a sound knowledge in Statistics helps us to use proper methods when collecting data and then choose correct methods in the data analyzing stage and finally to make correct interpretations of the results. The importance of Statistics has now been realized by the society more than ever before. IASSL continuously provides its assistance to educate and popularize Statistics among the public and it is evident that the effort is well recognized and appreciated by the public. There is a high demand for short courses conducted by the IASSL and during the last four months IASSL was able to conduct ten, very engaging short courses in diverse areas of Statistics.

On behalf of IASSL, I would like to cordially welcome the new memberships. The continuous growth of the membership highlights that fresh graduates and industry experts alike have identified the importance of having the membership of this prestigious organization. If you have not joined with us so far please do apply for the membership and contact the office for further information.

Sri Lanka Journal of Applied Statistics is steadily progressing now with significantly reduced processing time. SLJAS reports significant and novel contributions in statistics with the particular focus on Applied Statistics. The journal has a distinguished Editorial Board with sound academic qualifications, to ensure the high scientific standards in publishing. You can access the journal through IASSL website free of charge and further I herewith kindly invite you to submit an article to SLJAS which is a peer-reviewed, open access journal that publishes original research articles.

I take this opportunity to thank all the resources persons, members of the Executive Council, Editorial Board, members of the subcommittees, staff of the office and all others who contributed to IASSL for their tireless work.

This newsletter brings you interesting and informative articles authored by academics and industry experts. Please do enjoy the IASSL Newsletter May-August 2022.

Please stay in touch with us and do not hesitate to contact us with any questions, concerns. You are welcome to the IASSL Web site (https://iassl.lk) and the Facebook page.

Editorial

Dr. Vasana Chandrasekara
Editor/IASSL

It is indeed a great honour to be the Editor of IASSL and it is an immense pleasure to launch this second issue of the newsletter for the year 2022. In this issue, we will recount various events, projects and activities in which IASSL members were actively involved from the 1st of May 2022 until the 31st of August 2022. Basically this issue contains articles from senior academics and industry professionals in the field of Statistics, One-Act-Play article from an emeritus professor, articles from IASSL members, news in brief which covers all events of IASSL during the considered period of this newsletter. This issue introduces the 'Stat Undergrad Column' which enables Statistics undergraduates of Sri Lanka to share their research findings among the community. As usual, the puzzle completion is included for all readers to relish and win prizes, and the winners of the puzzle competition of the last issue are announced in this issue. Finally, the upcoming events of IASSL are listed for your information.

A huge thank to all the professors, industry professionals, IASSL members and undergraduates who contributed to writing the valuable articles for this issue. Moreover, I appreciate the support extended by the President, Secretory, all subcommittee Chairpersons and the Executive Council members of IASSL in providing information relating to the events conducted by them during the period May to August 2022. Last but not least, I would like to thank the Editorial Board members, and especially the Associate Editor of IASSL, for their immense support throughout the preparation of this issue of the IASSL newsletter.

I invite all readers to submit articles and news to be considered in the next issue of the IASSL newsletter (editor.iassl@gmail.com) and hope you all will enjoy reading this issue.

CONTACT INFORMATION

Institute of Applied Statistics Sri Lanka
The Professional Center
275/75
Prof. Stanley Wijesundara Mawatha
Colombo07
Sri Lanka

+94 11 2588291
facebook.com/iassl2020
linkedin.com/company/iassl/
editor.iassl@gmail.com

ONE-ACT-PLAY:

When Incomplete is Good

Professor Emeritus R.O. Thattil
Founder President
Applied Statistics Association of Sri Lanka (ASASL)

The following conversation takes place at the postgraduate student common room. The participants are graduate students; one a Statistics student (Nimal), another a Biology student (Silva) and the third an Agriculture Engineering student (Perera).

Perera: I cannot understand why researchers insist on Incomplete Block Designs (IBD) sometimes, when Complete Block Designs are available for experiments!

Silva: In one of our courses the lecturer stated that IBD's are alternative designs, to complete block designs such as Randomized Complete Block Designs (RCBD).The word incomplete in IBD put me off.

Nimal: But my lecturer said that IBD's are essential when the number of treatments are large.

Perera: Why can't we use a RCBD even then?

Silva: What is your opinion Nimal?

Nimal: When the number of treatments are large, you cannot incorporate all the treatments into a complete block. By definition a complete block is where all the treatments are found inside a block.

Silva: So what? We can still find a block large enough to accommodate all the treatments.

Nimal: The problem is that large blocks are not homogeneous. The idea in blocking is to have less variability within the block. When the treatment number is large we cannot find homogeneous blocks to accommodate all the treatments within each block.

Perera: Can you give me an example?

Nimal: In Animal Science experiments, the natural block is the litter. All animals within the litter are more or less homogenous. However, the litter size is usually small except in animals like rats and pigs. If you have 6 treatments and an average litter size of 4 animals, obviously block size is smaller than the treatment number and therefore all the treatments cannot be accommodated within the block. If each animal is a plot or experimental unit, only 4 treatments can be included within a block, leading to IBD.

Perera: But litter sizes can be different and therefore block sizes are different. How do we use a design with different block sizes?

Nimal: It is good to have equal block sizes, but in a general IBD, having different block sizes is not a problem. Techniques are available to analyze such designs. Hand calculations are difficult but, we can use a statistical software program such as SAS in the general linear models (GLM) procedure to analyze such designs.

Silva: Won't there be a problem with pairwise comparisons of treatments?

Nimal: Yes, some pairs will occur more frequently than other pairs. In terms of precision of comparisons, the comparison of pairs that occur together more frequently will be more precise than pairs that occur together less frequently.

Perera: Suppose I want equal precision for all pairs of treatments?

Nimal: That is not a problem. There is a class of IBD's where all pairs occur together within the blocks, an equal number of times. Such a design is called the Balanced Incomplete Block Design (BIBD),

Silva: Can we construct a BIBD in all situations?

Nimal: If the number of blocks are adequate we can. But, when we don't have enough blocks we can obtain partial balance with let us say 2 levels of precision, leading to the partially balanced incomplete block design (PBIBD).

Perera: I think you have to give us an example.

Nimal: All right. Take an experiment with twins. Say we have 4 treatments (T1 to T4). We can accommodate only 2 treatments in a block (i.e. a pair of twins). We will need 6 blocks to obtain a BIBD with the following configuration;
Block 1– (T1 and T2)
Block 2 –(T1 and T3)
Block 3—(T1 and T4)
Block 4—(T2and T3)
Block 5—(T2 and T4) and
Block 6 –(T3 and T4).
All pairs now occur together in the blocks exactly once.
If we have 12 blocks (or twin pairs), we can have all the pairs occurring together exactly twice within the blocks. Of course you have to randomize the treatments within each block.

Perera: If we don't have even 6 blocks I can't see how we can have a BIBD.

Nimal: Exactly. Then, we have to use a PBIBD. You have to decide for which pairs you need higher precision.

Perera: I can see the problem clearly now. There is no alternative when block size is smaller than the number of treatments, other than going for an IBD.

Nimal: You got it right! You have now understood the need for an IBD. Furthermore, smaller blocks have less variability than larger blocks. Therefore, IBDs are better designs than RCBDs when block size is small.

--------Curtain Comes Down------

A Guide to Recommendation

Nimal Wickremasinghe, Ph.D
Professor and Consultant
Informatics Institute of Technology,
Former Head of Department of Statistics,
University of Colombo.

Suddenly there has been an emergence of requests from students and colleagues for recommendations. It may be correlated with some current issue prevailing in the country. Thus, it is thought that writing something on this topic is timely since it might help students and academics from a wide spectrum of disciplines including Statistics.

Usually there are three types of recommendations: Academic, Employment, and Character. These are also called letters of reference. An individual needs to know what exactly is needed depending on the case.

Character Recommendations:

These are usually needed by landlords when accepting tenants, in legal issues, in some specific types of government and industry jobs, and so on. One must realize that the word 'character' has a broader meaning than what most people in our society think. The 'character' (or the moral character) of an individual expresses a variety of attributes including the presence or absence of virtues (i.e. good moral qualities) and good habits. Character recommendations are usually given by well-known and recognized persons in society, like school principals, religious leaders, Deans of Faculties, and so on, who know the individual well enough. In my opinion, a university academic must think twice before giving such a recommendation to a student on request unless he/she personally knows the student well for a longer period. In most cases, the lecturer has contact with a student only in the class and knows only about student's academic performance (and may be attendance too). The lecturer may be clueless about what the student does outside class.

For the student:

Typically, you may request these recommendations from former teachers, program coordinators, heads of departments, research supervisors and coaches, who know the student and his/her performance well. Some hints for students are:

- Choose someone who likes you and knows you well; he/she would most probably give a strong recommendation.
- Don't send the request by e-mail; give a call and make an appointment with the potential recommender, and make the request in person.
- When meeting in-person is not possible, send a nice e-mail explaining the need and furnish all relevant information.
- Don't hesitate to mention any specific attributes that you would like to see included in the letter.
- Remember to send an e-mail thanking him/her, or call in-person and thank, once you receive the letter.
- Don't assume that your professor /lecturer/advisor remembers everything about you. In most cases they remember only your name, your face, and the class obtained; not even the date graduated or any research publications.
- You should provide a file (physical or electronic) containing:
  - Copy of transcript,
  - Copy of CV,
  - Courses you have taken with him/her and the grades obtained,
  - Research done with him/her and publications (if any),
  - Brief account of experience in the subject area,
  - Your professional goals,
  - Details of all degree programs applied for and the person to be contacted in each program along with addresses of departments and e-mails, and
  - Deadline to submit recommendations
- Avoid requesting recommendations from poor writers even if they like you and know you well enough.
- Avoid people who would say 'why don't you write what you want. I will sign'. Those are not genuine recommenders; they don't want to spend even 1 minute for you.
- Obtaining a recommendation is not a right or entitlement; you are lucky if you get one on time.

Employment Recommendations:

These are needed by those trying to get a new job. Usually they are written by current or former employers or immediate supervisors. Co-workers are not expected to write such recommendations. Generally, they include employment history, performance of the individual, work ethics, and personal achievements during the period in question. As far as universities are concerned, fresh graduates on temporary appointment usually need these to apply for permanent jobs or other temporary positions. A lecturer who has supervised such a graduate or the Head of Department, may give such a recommendation.

Academic Recommendations:

This is actually the focus of this article. Academic recommendations are an essential part of any application process for college admission or graduate admission by a student. In the case of college admission, it is the high school graduate (in the local context, 'Advanced Level passed') who seek academic recommendations to enter a college. Since this case is not very applicable in the local context, emphasis will be given in this article to academic recommendations for graduate admissions (i.e. for 'postgraduate' degrees) and is primarily directed towards giving some helpful advices and suggestions for students seeking graduate admissions and young recommenders who are new to the task.

Sources:

· Personal experience

· 'thoughtco.com/recommendation–letters..'

(Downloaded on July 7, 2022)

For the Young Recommender (who may be new to recommending):

• Think twice before giving a recommendation to a student who is a friend of your student but has not taken any courses with you. You have the right to decline in such cases.

• A strong recommendation does not necessarily mean number of pages; one page may be enough in most cases [note that universities rate the recommender too].

• Always edit the letter for grammar and spelling, before sending a pdf version.

• Keep away from 'traditional writing' with fancy words; only highlight the student's ability and potential for successfully completing the degree in question. Don't hesitate to use the word 'excellent' when it should be used.

• If you are not the research supervisor, keep away from highlighting his/her research area; (but you may mention that he/she has done publishable research with one of your colleagues); remember that there will be other recommenders to write about it.

• If the highest grade is B+ and the student is one of the three (say) who got B+ for your course, mention that it's relatively a very good grade.

• If you happen to recommend two students for the same program in same university department, you should be very careful when choosing the correct ratings; make sure your comparisons don't contradict students' actual standings.

Data Mining for Business Analytics:

Concepts, Techniques, and Applications

Professor R. M. Kapila Tharanga Rathnayaka
Department of Physical Sciences & Technology,
Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka.

Data mining is a process of discovering and excavating useful patterns from considerably large datasets or databases involving methods at the intersection of Machine Learning, Statistics, and Database Systems. Furthermore, it provides the tools and techniques to deal with large amounts of data in the business domain to increase sales, and market strategies effectively to make better decisions. At present, the biggest industries in the fields of banking, insurance, retail, and social media around the world have been gaining miscellaneous benefits from their data analytics. For example, some industries have been using different types of Data mining techniques to develop more effective marketing strategies to increase their sales and decrease costs.

As a practice, the predictive capacity of data mining has been positively involved in changing the design of business strategies as well as understanding the present to anticipate the future. So, it's critical to find the best data preparation tools for data mining analytics like Oracle Data Mining, Rapid Miner, Weka, MonkeyLearn, and etc.

The ability to analyze complex data sets in order to predict future trends opens greater opportunities. Generally, the data mining process is carried out under 5 major steps (Figure 1). They are:

- Data Understanding: Finding and identifying the source of valuable information and lord them into correct data warehouses; Determining the data that will be needed to solve the problem and gathering it from all available sources;
- Data Preparation: Extracting the information that is surely or possibly useful for business;
- Modelling: Using algorithms to identify patterns within the data.
- Evaluation: Determining whether and how well the results delivered by a given model will help achieve the goal.
- Model Development: Making the results of the project available to decision makers.

"The ability to analyze complex data sets in order to predict future trends opens greater opportunities."

Banking & finance is one of the most enticing and essential sectors that can be widely applied data mining techniques for identifying the meaningful characteristics for making future ad-judgments under modern technologies. Furthermore, making decisions under numerous types of economic policies and reforms that have been regarded as one of the biggest challenges in the modern economy is significant; especially, data mining and machine learning techniques have been widely using to retain credit card customers, customer retention in the competitive business environment as well as prediction of future financial events such as credit ratings, investment analysis, trading futures, and understanding and managing financial risk.

This article discusses the data mining techniques and their applications in customer relationship management (CRM), and fraud prevention and detection.

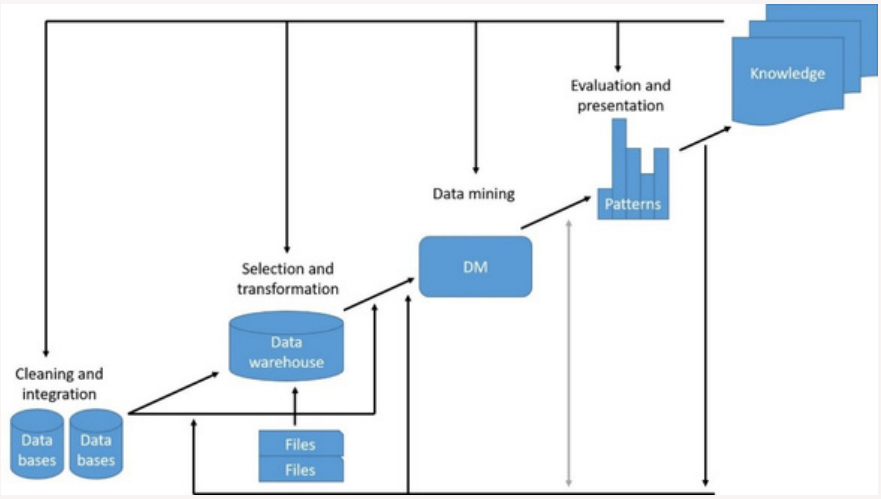


Figure 1: Data Mining Process

A. Customer Relationship Management (CRM)

Customer relationship management (CRM) is a combination of government practices, sound business strategies and modern technologies that can be used to provide the service, at the right time to manage customer interactions throughout their lifecycle. Furthermore, CRM help banks to understand their customers' needs providing targeted, timely and relevant information that can add value to their customers. For an example, k-Means clustering algorithm can apply to partition the customer profession data into separate clusters and rank the probability. In addition to that, Apriori algorithm can be used to perform market basket analysis as well as for discovering market associations.

B. Fraud Detection

Fraud Detection using data mining is another popular method which is used in the banking sector for the credit card transactions to identify the potentially fraudulent activity. According to the literature, two different approaches have been developed to detect fraud patterns. In the first approach, fraud patterns in their data warehouse are being identified based on the data mining methodologies. In the second phase, fraud pattern identification is mainly carried out strictly based on the bank's own internal information.

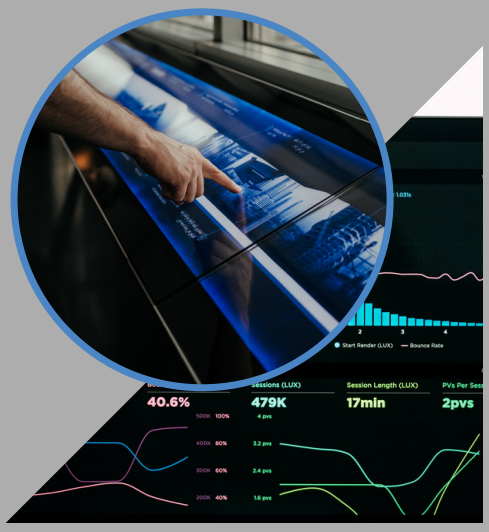
Keywords: Machine learning approaches, Business Analytics and Data mining techniques.

Usage of Statistics in Machine Learning Model Monitoring

Mr. Shanaka Chathuranga
Consultant – Product Engineering,
Emirates Group Technology Center,
Dubai, UAE.

Have you ever thought about the next step after deploying a model into the production environment? Is it okay to leave the model unmonitored after a production deployment? This is one of the main faults done in many machine-learning projects. If we forget to monitor the model after deploying it to the production, there is a high chance that our model performance will be degraded, and we will end up with a low performing model and might even have to re-invest in rebuilding the model. This article will be discussing about model monitoring which is an important aspect in machine learning projects after its deployment. Statistical hypothesis tests are the underlying key concept behind the ML model monitoring techniques.

Data used to build a certain model, change overtime due to the changes in macro and micro environments which means that distribution of the data may change from time to time. According to the fundamental theory in machine learning, data that we have used to make predictions should be coming from the same distribution as training data. Therefore, changes to these distributions might lead to an inaccurate prediction. That is why we should closely monitor or supervise our ML models. There are many model monitoring techniques used in ML projects. Data drift and concept drift are widely used in the industry. Statistical tests that are being used to find Data drift are discussed below.



- Monitoring a data drift

Data drift is one of the main reasons why model accuracies degrade over time. For machine learning models, data drift is the change in the model input data that leads to model performance degradation. Monitoring data drift helps to detect these model performance issues. Different techniques can be used to detect a data drift. Utilizing statistical tests to compare the distribution of training data with real data is the simplest method. If there is a considerable discrepancy between these two distributions, a drift has taken place.

- Monitoring the numerical data drift

A statistical test can be used to detect a numerical data drift. This can be used to compare numerical features, numerical target, and numerical prediction of the model. Two sample Kolmogorov–Smirnov test and Wasserstein metric are widely used to detect such a drift.

- Monitoring a categorical data drift

A Categorical drift is a drift in the categorical variables and categorical target in the model. Chi-squared test and Jensen–Shannon divergence can be used to measure the categorical data drift. Evidently, AI is one of the best open-source tools which can be used to do the model monitoring. Here model monitoring for simple churn model is implemented using Evidently AI. The full project link: https://github.com/shanakaChathu/churn_model

Feature	Type	Reference Distribution	Current Distribution	Data drift	P-Value for Similarity Test
number_vmail_messages	num			Not Detected	0.381711
number_customer_service_calls	num			Not Detected	0.4535
total_intl_minutes	num			Not Detected	0.757163
total_eve_charge	num			Not Detected	0.791611
total_eve_minutes	num			Not Detected	0.791611
total_day_calls	num			Not Detected	0.993117

6 rows 1-6 of 6

Figure 1: Detecting Drift in features (independent variables) of the model

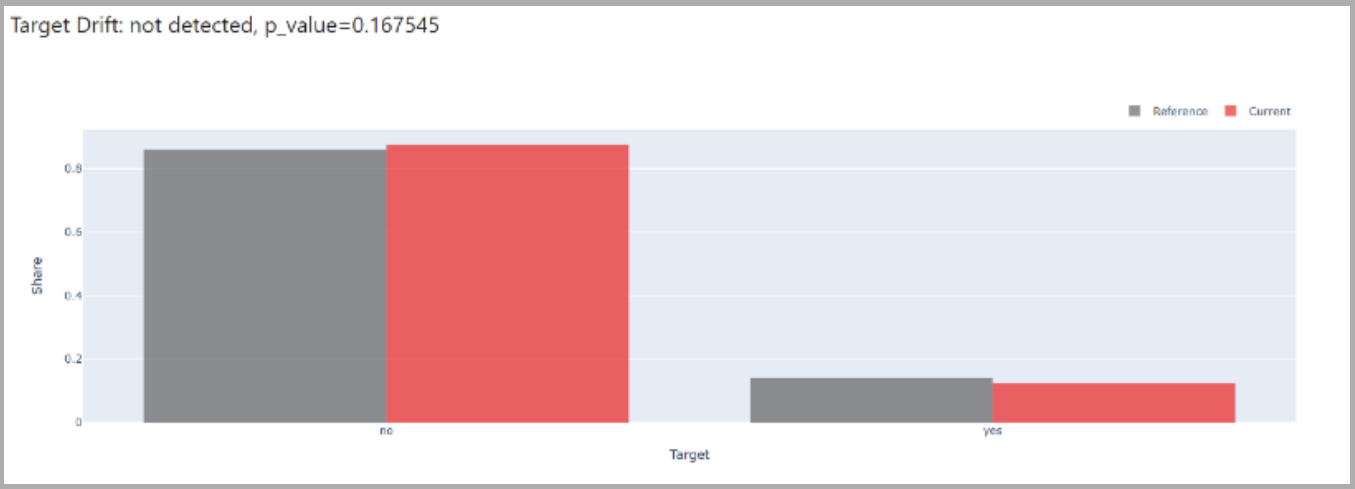


Figure 2: Detecting drift in Target variable of model

Figure 1 shows the independent features and compare the current distribution with reference distribution to detect any data drift. It shows the P-value for the similarity test so that data drift can be detected based on the P-value. In Figure 2, the target is a categorical variable (churn or not churn). Therefore chi-squared test is used to measure the data drift. The decision can be made based on the P-value of this test.

Above two reports are simple examples which can be used to monitor a ML model. There are many other kinds of reports which can be utilized for model monitoring. It is always a good practice to implement the model monitoring techniques in production environment if we want models to produce accurate results in the long run.

References:

1.https://evidentlyai.com/

Advancing Netball through Data Science

Mr. Janith Wanniarachchi, Ms. Udeshi Salgado,
Dr. Rajitha M. Silva, Dr. Chathuri L. Jayasinghe

USJ Sports AnalyzR

Statistics in Sports Research Group,
Department of Statistics,
University of Sri Jayewardenepura,
Sri Lanka.

Data science being a multidisciplinary area combining Mathematics, Statistics and Computer Science, has revolutionized the decision-making process of many industries in the past decade, ranging from art to medicine. Focusing on an area with a decision-making process that has long been running on gut instinct, the field of sports was a new addition to industries integrating data science into practice. This was after the realization that, while subjective opinions and improved thinking can drive a team towards a win, assistance from data-driven insights could be helpful to unravel the hidden techniques inside data that would help a team drive a strategic home run. From insights on own team performance, the strategies used by the other teams to the most optimal time to make substitutions, etc. data science has kept contributing to the advancement of the sports industry like no other. In addition to aiding game strategizing, data science has helped athletes all over the world discover their strengths and points of improvement, helping them exponentially improve their performance.

> Being a country with an abundance of talented and fierce athletes, Sri Lanka has hopped into the bandwagon of using data science to improve its sports industry as well.

Among the plethora of sports played in Sri Lanka and all around the world, Netball holds a prestigious place predominantly among women due to its non-contact nature and team coordination aspects of the game. Played between two teams of seven players, the objective of each team is to score goals by passing a ball down the court through players with assigned goals and areas and shooting the ball through a goal ring. In Sri Lanka, Netball has been popular since its inception in 1921. Despite the lack of support given by the general public towards Netball, the Sri Lankan Netball team has worked in bringing fame and glory to their home country. Recently, in 2018, the Sri Lankan Netball team reached great heights by achieving the championship in the 11th Asian Netball Championship against Singapore.

However, the strategies and approaches taken in coordinating the team players are still based on subjective decision making and an objective data-driven approach is yet to be taken.

The question would then become as to why Netball requires analytics. In other terms, we would need to understand the feasibility of applying analytics for Netball. Australia and New Zealand are two teams known for their rivalries in many sports and in 2014, the Australian team was able to win the gold medal in the 2014 Commonwealth Games in Glasgow, after a long period of drought in gold medals for Australia. Their secret relied on using analytics to analyze and understand the movement patterns among the different players. Their inspiration for using analytics came from looking at the way a shoal of fish move around in a small tank. They looked for cues in the behaviour of individual fish which might predict a pattern at the shoal level, while also looking for things individual players do that might predict a pattern of play at the team level. In addition to player positioning, there have been huge advancements made in automatically deriving data from available video footage.

> Statistics in Sports is a growing field in Statistics that provides specialized methodology for collecting and analyzing sports data in order to make decisions for successful planning and implementation of new strategies.

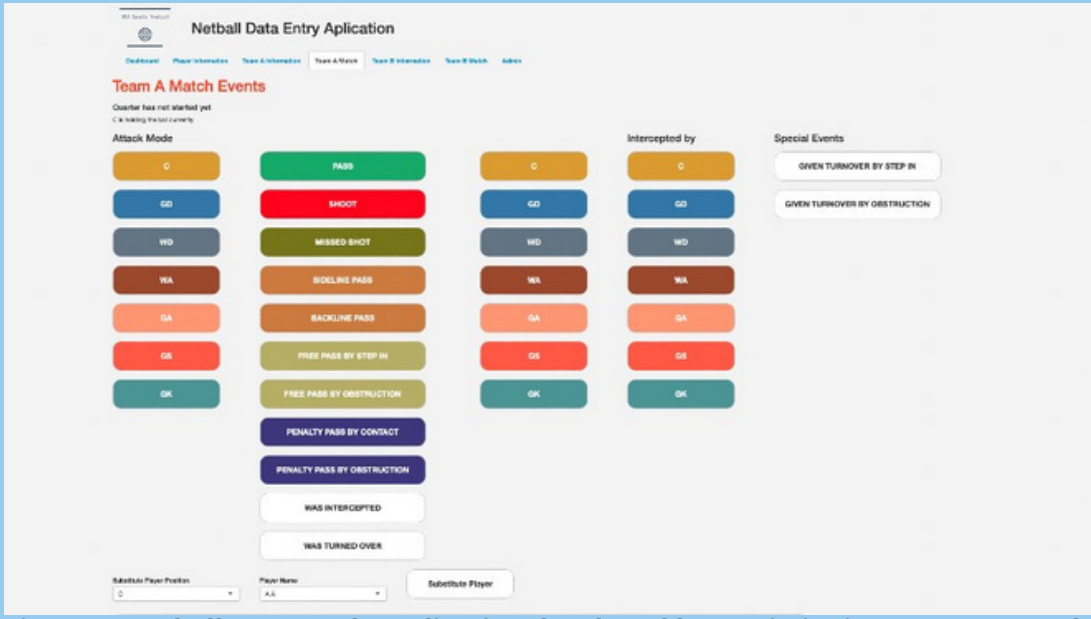


Figure 1: Netball event web application developed by Statistics in Sports Research Group, USJ.

Advancements in deep learning techniques have enabled analysts to analyze the posture and stance of each player as they are moving through the court to understand their strengths and weaknesses. With such a vast sea of opportunities available, data science would be a valuable addition to the arsenal of tools that Netball coaches and players can utilize.

Applying data science techniques for the decision-making processes in Netball would first require data to exist. It would be quite difficult to perform data science or any statistical techniques without real-world data at hand. In terms of data collection, there are several approaches available: both manual and automated.

As an automated procedure, it is possible to take video recordings of the matches played by players. Proper video recording equipment with adequate lighting and minimal disturbances would be required to ensure that high-quality data can be collected, thereby expediting the data processing stage. Another method is manually collecting summary performance data of each player (i.e. the number of shots, bad passes and penalties that were made) for each match. A fixed format for the data entry would be required which can be replicated for all matches. The granularity of the data collected would be quite small, as with manual labour only summary measures can be collected. But what if instead of obtaining only a high-level summary view of the data we could instead record each action that happens within a match manually with the input from domain experts?

The Statistics in Sports Research Group of the University of Sri Jayewardenepura has taken the initiative to develop a web application (Figure 1) to assist in collecting events occurring in a Netball match. Through an intuitive user interface, the users can record the type of event and which players contributed to the event from the perspective of both teams. For example, if a pass was made from the Center to the Wing attacker, then this would get recorded with three clicks. In addition to simple passes, a wide variety of actions such as penalty passes, free passes, throw-ins, and missed shots will be recorded, while also recording the scores of both the teams and most importantly the team that holds the Center pass. Built using R Shiny for faster prototyping and ease of use for statistical analysis, the beta version of the web application communicates with an Azure SQL cloud database to record and store the data safely and securely. While the match is proceeding, the web application is capable of showcasing in real-time whether the current Goal Shooter has been faltering in her shoots and therefore would need to be substituted. The collected data would then be used to identify the performance of players and replay the entire sequence of events that occur within a match to identify potential mistakes and associations between players within a match. It can indicate whether a certain player is always passing to a specific player resulting in a goal which can be investigated to identify team coordination tactics and player combinations. In addition to shooting and scoring, the web application gives the option to record the events that occur while the ball is in the hands of the opposing team. Through this data, insights on player combinations that work well for intercepting the ball from the other team can be identified. Currently, this web application is undergoing rigorous testing by recording test matches to ensure that real-time performance can be achieved amidst the fast-paced action of a Netball match.

Data science has surpassed the expectations of many as the field grows to assist in making the right decisions based on past data. However exciting the prospect may be, applying data science to Netball requires investments in time, money and effort to collect the necessary data and provide analytics throughout. The Sports Statistics Research Group of the University of Sri Jayewardenepura is currently researching innovative pathways for bringing data science research into the world of sports through modern technologies and methodologies.

Connect with us: http://science.sjp.ac.lk/sta/statistics-in-sports-research-group/

News in Brief

Certificate short courses conducted in May - August 2022 by Academic and Training Subcommittee:

1. Modelling Binary & Nominal Outcome Variables:
Resource person: Dr. Niroshan Perera, Department of Statistics, University of Sri Jayewardenepura.

2. Basic Statistics for Managers and Researchers:
Resource person: Prof. N. Rupika Abeynayake, Department of Agribusiness management, Wayamba University of Sri Lanka.

3. Business Analytics using Power BI:
Resource person: Ms. Samudra Bandaranayke, Senior Data Analyst, Wiley in Sri Lanka.

4. Structural Equation Modelling:
Resource persons: Dr. Sampath Fernando, Department of Economic & Statistics, University of Sabaragamuwa and Dr. Chathurani Silva, Department of Decision Sciences, University of Sri Jayewardenepura.

5. Multivariate Time Series and Modeling Volatility:
Resource person: Dr. Hasanthi Pathberiya, Department of Statistics, University of Sri Jayewardenepura.

6. Data Analysis for Social Science:
Resource person: Mr. N.A.N.J. Maduwansha, Department of Social Statistics, University of Sri Jayewardenepura.

7. Basic Statistics for Undergraduates:
Resource person: Prof. N. Rupika Abeynayake, Department of Agribusiness management, Wayamba University of Sri Lanka.

8. Research Methods for Management & Social Science Researchers:
Resource persons: Dr. Sampath Fernando, Department of Economic & Statistics, University of Sabaragamuwa and Dr. Chathurani Silva, Department of Decision Sciences, University of Sri Jayewardenepura.

9. Basic Statistical Analysis using STATA:
Resource person: Dr. M.G Nuwan Indika, Department of Business Economics, University of Colombo.

10. Basic Statistical Analysis using R:
Resource person: Dr. Chitraka Wickramarachchi, Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura.

A few testimonials from participants of recent workshops...

"Indeed I'm thankful to IASSL for organizing a workshop for Doctors in the field of Public Health + research and tailor-made to suit our requirements. It did exactly what it said it would, was interesting and interactive. The hands-on nature of the course, applied analysis with real data, and the clear explanation of concepts which have previously seemed impenetrable. Very good levels of participation"

Dr. A.F.F. Fazla
MBBS, MSc, MD (Community Medicine)
Senior Registrar
(Policy Analysis and Development)
Management Development and Planning Unit,
Ministry of Health.

"THANK YOU VERY MUCH IASSL FOR ORGANIZING A WORKSHOP FOR POSTGRADUATE DOCTORS IN COMMUNITY MEDICINE WHO ARE ENGAGED IN RESEARCH AROUND THE WORLD WITH DIFFERENT TIME ZONES. IT WAS VERY INTERESTING AND INTERACTIVE. EVERYONE ACTIVELY PARTICIPATED IN WORKSHOP AND GOT HANDS-ON EXPERIENCE IN DATA ANALYSIS."

DR. BUDDHINI HERATH DENUWARA,
HONORARY RESEARCH FELLOW,
UNIVERSITY OF MELBOURNE,
AUSTRALIA

Be a proud member of IASSL!

IASSL offers a variety of membership categories and below we mention eligibility criteria required for each membership category. For details of the procedure of obtaining a membership and the benefits of becoming a member please visit http://iassl.lk/

- Fellow members should have following qualifications:
  - A general (life) member of the Institute (or Applied Statistics Association of Sri Lanka) for a continuous period of ten years and
  - A distinguished person who has done immense service to Statistics discipline in general or to the Institute in special.

- General members should have at least one of the following qualifications:
  a. A Postgraduate degree (MSc, MPhil, PhD) in Statistics/ Applied Statistics
  b. A degree with a minimum of 30 credits in Statistics
  c. A degree with a minimum of 15 credits in Statistics and a project in Statistics/Applied Statistics carried out for a period of not less than one semester and
     i. A Postgraduate diploma in Statistics/Applied Statistics with 3 years of post experience as a statistician or equivalent position acceptable to the council or
     ii. The diploma offered by IASSL with 5 years of post experience as a statistician or equivalent position acceptable to the council.

- Honorary Fellow
  - Any distinguished person who has done an immense service to the Statistics discipline in general or to the IASSL in special are eligible to be an Honorary Fellow.

- Special Membership
  - Offered to someone who is recognized by the Executive Council of the IASSL and the general membership considering his/her service rendered to IASSL.

- Corporate Member
  - Any public sector or private sector organization that is engaged in statistical work shall be eligible to be enrolled as a corporate member. Corporate members can send one authorized representative to the Annual General Meeting.

- Student Membership
  - Any undergraduate of a degree program where courses related to Statistics are offered as a subject of the degree shall be eligible to be a student member.

NOTE:

- **Honorary Fellowships and Fellowships will be awarded at the Annual General Meeting or at an International Statistics Conference conducted by the Institute on the recommendation of the Executive Council of IASSL.**
- **Application for membership should be made in a prescribed form approved by the Executive Council of IASSL. It shall be proposed and seconded by general life members.**
- **The Executive Council of IASSL shall have the power to refuse any application for membership, without giving reasons for such action.**

New Members (May-Aug 2022)

Miss. K.T.N. De Silva
Miss. H.D.M.T. Karunaratne

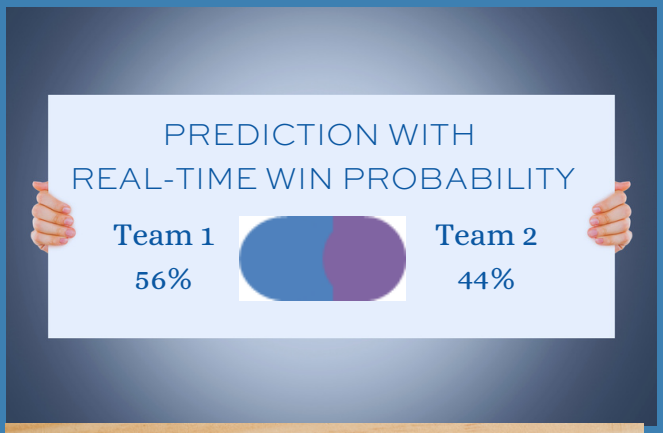
Stat Undergrad Column

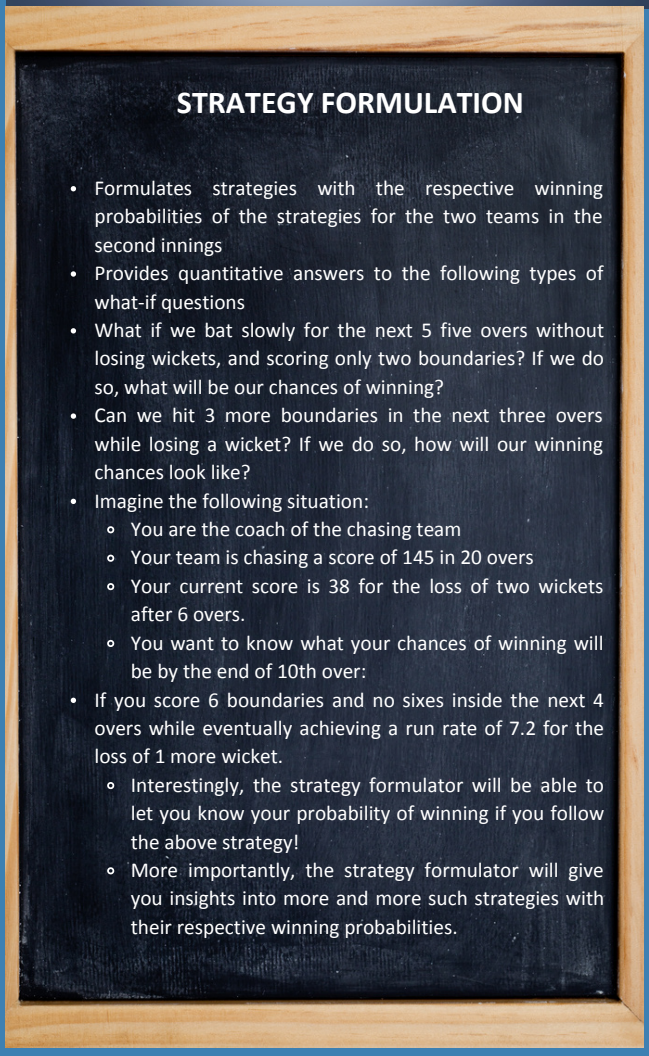
Real -Time Prediction and Strategy Formulation in T20 Cricket

Mr. L.C.P Pussella
BSc (honours in Industrial Mathematics)
Rajarata University of Sri Lanka

During the most recent T20 match between Sri Lanka and Australia which was held on 11th of June 2022, we were fortunate enough to witness one of the finest ever onslaughts in T20 cricket by Dasun Shanaka. One of the key aspects that we observed after the win was how Shanaka was able to defy the "odds" by carrying the team home with a win.

As many people were interested in learning how the live win probability applications played their part after the win, we decided to explore how our own real-time winner predictor fared during the match. This application provides the capability of choosing a preferred classifier out of Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM), and the plot given below illustrates how the application was able to "sense" the real-time match situation to provide acceptable probabilities. As shown in the plot, after the end of the 17th over, Sri Lanka's win probability rises up gradually with the beginning of Shanaka's heroics.





Apart from the real-time prediction capability, the application is also capable of formulating strategies during the second innings of the match. The strategies can be highly beneficial for making plans during the match to assist the captains, coaches, and the team management.

The models of the application were trained and tested by considering matches played in Indian Premier League from 2008 to 2020 based on 11 highly significant features which we statistically identified using the Least Absolute Shrinkage and Selection Operator (LASSO).

you may use this QR code to experience the interactive application yourself.

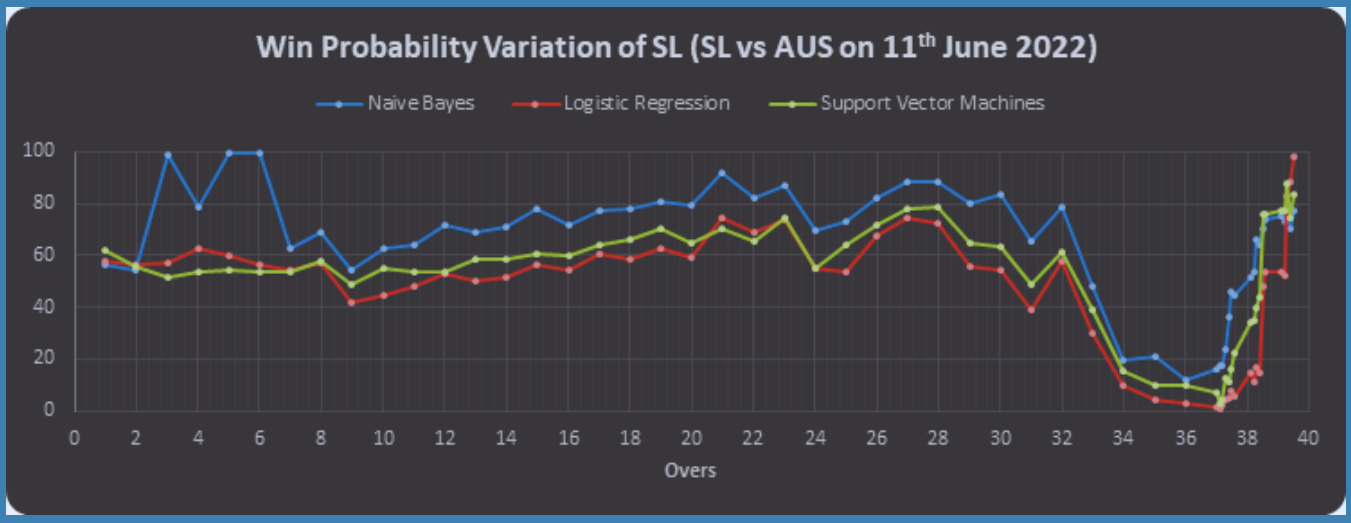


Figure 1: Win probability variation of SL

Since we offer the predictions and strategy formulation on a ball-by-ball basis, the model accuracies are also available on a ball-by-ball basis for each classifier (see Figure 1). The accuracies vary from 53.08% (first ball of the match) to 97.65% (last ball of the match). This research study has already been submitted for publication, and if you are interested in learning more about the application, you may use the given QR code to experience the interactive application yourself.

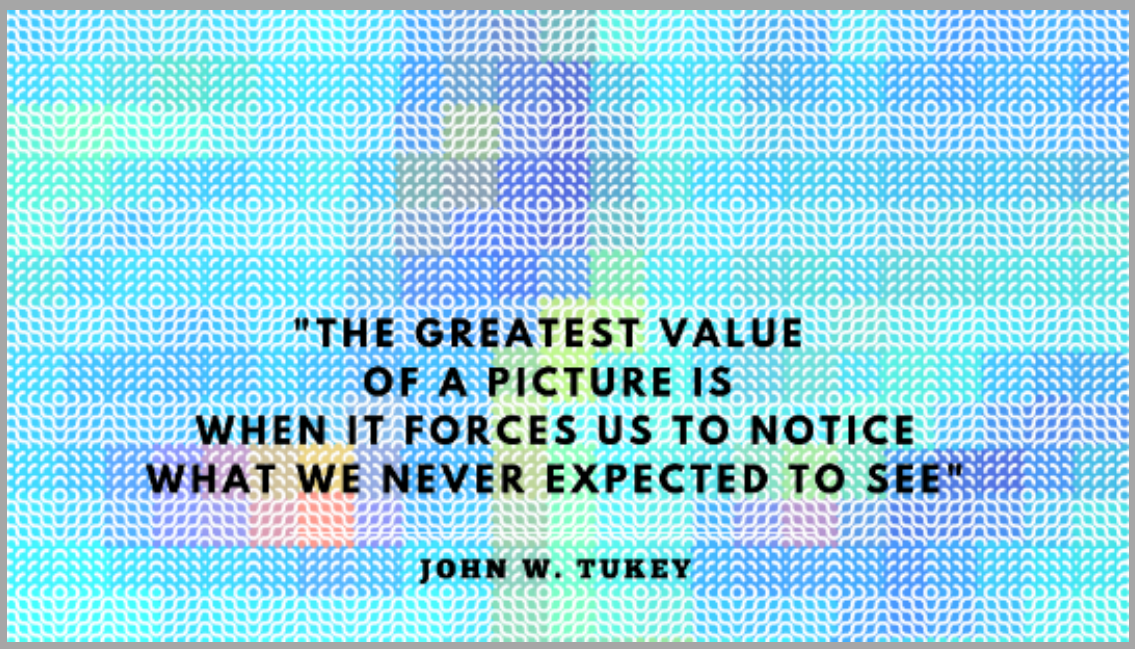
Authors: LCP Pussella[1] , RM Silva [2], WCP Egodawatta[3]
1 Department of Physical Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka
2 Department of Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
3 Department of Plant Sciences, Rajarata University of Sri Lanka, Anuradhapura, Sri Lanka

(This article was written based on the first author's final year research project.)

Stat Undergrad Column

Image Generation in R

Mr. Janith Wanniarachchi
BSc (honours in statistics)
University of Sri Jayewardenepura



The background of the above image was generated using R and the truchet package, by generating Truchet multi-scale mosaics based on a design matrix which is repeated over and over.

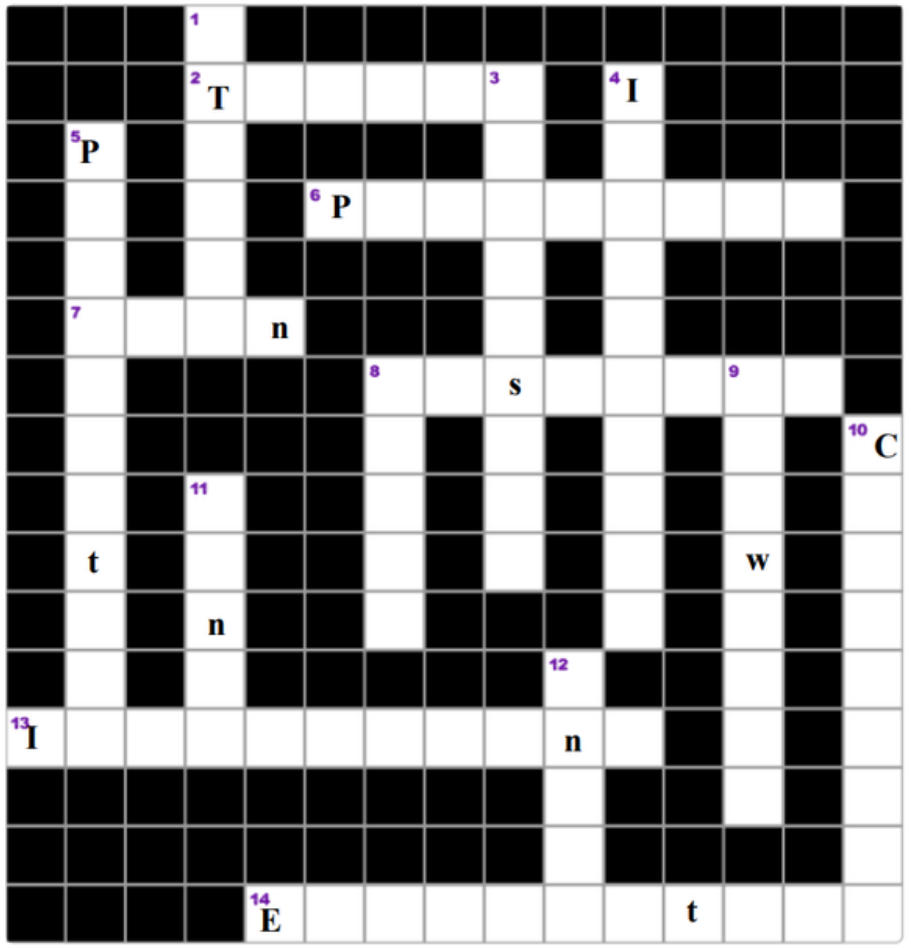
GitHub: https://paezha.github.io/truchet/index.html#assembling-a-mosaic-with-multi-scale-tiles

Sudoku Puzzle Competition

First prize: Rs. 5000
Second prize: Rs. 3000
Third prize: Rs. 2000

Please email your submission to appstatsl@gmail.com on or before 15th of December 2022. The draw will be held on the 30 th of December, 2022. Correct submissions will be short listed and the winners will be selected considering the order of submission and will be announced in the Issue 3 2022 IASSL newsletter.

We would like to sincerely thank a well-wisher for sponsoring this competition in memory of late Mr. Palitha Sarukkali the first President of IASSL.



INSTRUCTIONS

Across

2 In a random experiment, performances/executions are known as: .....

6 A numerical term that summarizes or describes a population.

7 An expected value of a random variable is also called as:.....

8 The variable of interest to be measured in the experiment outcome.

13 Predictor variable is also called as.....variable.

14 A variable is used to predict or explain differences in the response variable.

Down

1 In stratified sampling, the subpopulations are called as: .....

3 A numerical term that summarizes or describe a sample.

4 For any event A, AUA = A and A⋂A = A. This law is called as ..... law.

5 In mathematics, the notation of ..... relates to the number of ways of arranging all the members of a set into some sequence or order.

8 For a ..... skewed distribution, the mean is typically greater than the median.

9 Provide an idea about the shape of the distribution.

10 P(Ω) = 1. This is known as the as the axiom of .....

11 A measure of dispersion.

12 Set of outcomes that belong to either to A, to B, or both.

Announcements

- Upcoming Certificate Short Courses by IASSL
- Upcoming Events at IASSL

Discounted course fee for members:

10% reduction in the course fee for IASSL life members!

The following short courses will be conducted in the coming months by IASSL:

- Statistical Modelling with R:

Resource person: Dr. Vasana Chandrasekara, Department of Statistics & Computer Science, University of Kelaniya.
Date: 13th - 16th September, 2022

- Basic Statistics for Managers and Researchers:

Resource person: Prof. N. Rupika Abeynayake, Department of Agribusiness management, Wayamba University of Sri Lanka.
Date: 20th - 24th September, 2022

- Artificial Neural Networks with R:

Resource person: Dr. Vasana Chandrasekara, Department of Statistics & Computer Science, University of Kelaniya.
Date: 27th - 30thSeptember, 2022

- Panel Data Analysis:

Resource person: Dr. M.G Nuwan Indika, Department of Business Economics, University of Colombo.
Date: To be Announced

- Multivariate Time Series and Modeling Volatility:

Resource person: Dr. Hasanthi Pathberiya, Department of Statistics, University of Sri Jayewardenepura.
Date: To be Announced

- Modelling Binary & Nominal Outcome Variables:

Resource person: Dr. Niroshan Perera, Department of Statistics, University of Sri Jayewardenepura.
Date: To be Announced

For more information regarding the upcoming short courses please contact IASSL using the contact information on the last page of the Newsletter.

All Island Competition in Statistics

IASSL is happy to inform that as an initiative of the Statistics and Popularization Subcommittee of IASSL, an all island competition in Statistics for School/University students will be organized. The competition is scheduled to be held online with the intention of preparing students for upcoming Statistical Olympiad competition. The competition will be held under four categories: Grades 6-8, Grades 9-11, Grades 12-13, and Undergraduate. Announcements and updates related to the competition will be shared through the IASSL website in the near future.

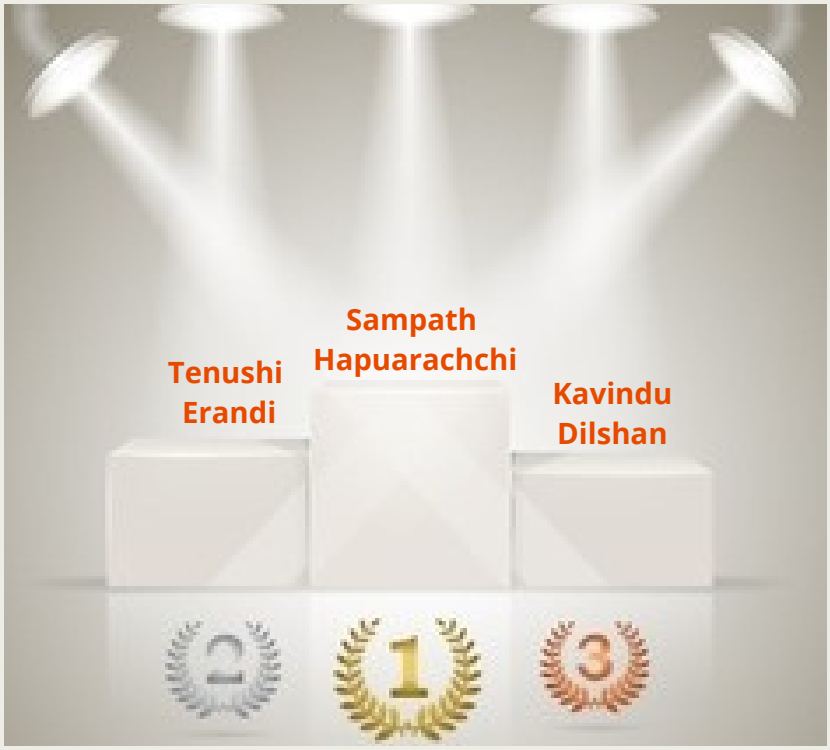
Announcements

- SLJAS: call for papers for volume 23
- Issue 1 2022 Sudoku Puzzle Competition Winners

Call for papers for Volume 23 of Sri Lanka Journal of Applied Statistics (SLJAS)

The Sri Lankan Journal of Applied Statistics (SLJAS) publishes the results of original work on applications of Statistics, and on theoretical and methodical aspects of Statistics. The journal also welcomes critical reviews including conceptual discussions, opinions and book reviews. Applications of Statistics in the area of Agriculture & Forestry, Medical, Dental and Veterinary Sciences, Natural, Physical Sciences, Social Sciences, Economics and Actuarial Science fall within the scope of the journal. Please visit https://sljastats.sljol.info/ for more information.

Winners of the Issue 1 2022 Newsletter Sudoku Competition:



Congratulations to all the winners and thank you to all who participated!

CONTRIBUTIONS TO THE SEPTEMBER-DECEMBER (ISSUE 3) 2022 NEWSLETTER:

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

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
appstatssl@gmail.com
http://www.iappstat.lk
http://www.facebook.com/iassl2020/
https://www.linkedin.com/company/iassl/