

IASSL

NEWSLETTER

“Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write.”

H. G. Wells

CONTACT INFORMATION

Institute of Applied Statistics
Sri Lanka

🏠 The Professional Center,
275/75, Prof. Stanley
Wijesundara Mawatha,
Colombo 07,
Sri Lanka.

☎ +94 11 2588291

✉ appstatsl@gmail.com

🌐 [http:// www.iappstat.lk](http://www.iappstat.lk)

📘 [https://www.facebook.institute.of.applied.statistics Sri Lanka](https://www.facebook.institute.of.applied.statistics.Sri.Lanka)

🌐 <https://www.linkedin.com/company/iassl/>

IASSL EDITOR:

Dr. Vasana Chandrasekara

✉ editor.iassl@gmail.com

Stories inside...

- ▶ Editorial
- ▶ One Act Play
- ▶ Interactive teaching and learning tools...
- ▶ Modelling: prediction is the name of the game
- ▶ Replacing data scientists...
- ▶ Data Visualization
- ▶ News in Brief
- ▶ Photo Gallery
- ▶ New Council – 2022
- ▶ From Stat Undergrads **NEW!**
- ▶ Upcoming courses/Events
- ▶ Announcements

IASSL President's Message



Dr. D.C. Wickramarachchi

President/IASSL

It is with immense pleasure that I bring this message for the January-April 2022 IASSL Newsletter. IASSL newsletter always be a live wire of the institute which brings various news on events carried out by the institute, informative articles presented in an attractive way, puzzles, messages etc. The editorial board should be highly appreciated for release of yet another newsletter. I would like to extend my gratitude for all the resource persons for providing valuable articles for the newsletter.

In March, 2022, IASSL held its 10th Annual General meeting. I was appointed as the president for the second consecutive year and I take this opportunity to thank all for your confidence in me to hold the Presidency of IASSL. The members should be appreciated for appointing a strong executive council for another year. To celebrate the 10th Anniversary of IASSL, the members and the staff of IASSL who have rendered the yeoman services to the Institute were felicitated at the AGM and which became the highlight of the AGM 2022. On behalf of IASSL, I would like to offer my heartiest congratulations for the members who have tirelessly worked during the last decade for the betterment of the Institute.

IASSL has been moving fast to achieve the set targets despite the prevailing condition in the country. The council has executed many events which is a remarkable achievement under the circumstances. The short courses are continuously running via online mode. Several webinars on vital topics have been conducted. The Sri Lanka Journal of Applied Statistics is now almost up to date. Based on the request from OPA, IASSL submitted its proposal on overcoming the prevailing situation in the country. Many more events have been planned to conduct throughout the year.

Please stay in touch with us and do not hesitate to contact us with any questions, concerns. You are welcome to the IASSL Website 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 first 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 January 2022 until the 30th of April of 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 cover all events of IASSL during the considered period of this newsletter followed by a photo gallery and statistics undergraduate column. A puzzle completion is included for all readers to relish and win prizes. Finally, the upcoming events of IASSL are listed for your information.

A huge thank you 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, secretary, all subcommittee chairpersons and executive council members of IASSL for providing information relating to the events conducted by them during the period January to April 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 creation of this issue of the IASSL newsletter.

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

ONE ACT PLAY

The Correlation Problem



Emeritus Prof. R.O. Thattil
Founder - Association of the Applied
Statistics Sri Lanka

The following discussion about the correlation coefficient is between 3 graduate students Anton (A), Bento (B) and Charles (C) and a Statistician (S).

A: I cannot understand why a correlation coefficient (r) as small as 0.2 in simple linear regression is statistically significant as reported in some research articles.

B: Rubbish! For a correlation coefficient to be significant it should be larger than $\frac{1}{2}$.

S: Both of you are right or wrong given the situation.

C: I don't understand your statement, Sir.

S: Using the magnitude of the correlation coefficient alone is not enough to talk about the significance of the correlation coefficient. Suppose you use only 2 observations in your study what will be the correlation coefficient?

A: Obviously a perfect correlation with $r = 1$, since you can draw a straight line to pass through both observations.

S: Right. Suppose you had 3 observations, what will be the value of r ?

A: Greater than 0.9, I guess.

S: Right again. Even if the variables are not correlated the correlation coefficient will be very high. You can now see that with a few observations the value of r will be close to 1. Now imagine you have 1000 observations. What will be the value of a significant correlation coefficient?

B: Definitely greater than $\frac{1}{2}$!

S: You are definitely wrong! It is great error to talk about a magnitude for a significant r value. It all depends on the number of observations. When r is large the value of r for significance will drop. Even $r = 0.2$ can be significant when the sample size n is large. Some people cheat by displaying a large value for r , while not mentioning the number of observations in the study.

C: So how would I know if a given r value is significant or not?

S: There are statistical tables giving significant values of r for different sample sizes.

C: I did not know this. I always thought that a significant value for r should be at least $\frac{1}{2}$.

S: You can now see that this is not true.

A, B and C: (All together). Yes. We understand now.

C: Wait. This is not the end of the story. We have been talking about linear regression. i.e. Straight line models. If the nature of regression is not a straight line you can have a linear correlation coefficient, which is not significant even when there exists a strong relationship between the variables.

A: Can you give us an example, Sir.

S: Suppose the relationship between the independent and dependent variable is a quadratic function, you may end up with a r value close to zero, if you fit a straight line model.

A: Yes, Sir. I can see that. (B and C also agree with A)

C: So how can we determine if the straight line is the best model?

S: Do not forget the simple scatter diagram or scatter plot which many people don't look at but, directly go and calculate the correlation coefficient. The mistake made by many persons is to input the data using a statistical software package and believe in whatever output is displayed by the computer. The scatter plot will give you a visual picture of whether the relationship is straight line or not. There are more complicated methods of finding the best fitting model but, do not forget the scatter plot. When we go to multiple regression, matters get more complicated with the number of observations as well as the number of independent variables used, affecting the multiple correlation coefficient. But, that is for another day!

A, B and C: Thank you very much, Sir.
(Curtain comes down)

"Statistics' real contribution to society is primarily moral, not technical"

– Steve Vardeman and Max Morris

Interactive Teaching and Learning Tools for Applications in Statistics



Prof. Pushpa Wijekoon

Department of Statistics and Computer Science,
Faculty of Science, University of Peradeniya,
Peradeniya, Sri Lanka

In earlier days, teachers had to rely on limited tools for teaching, and most of these methods are teacher centered. One of the weaknesses of these old methods is the difficulty of identifying the strengths and weaknesses of the students. Further, when a question is asked in class, it is answered by the same group of students, and it is difficult to know whether the silent students have understood what is being taught. To overcome these barriers, the student-centered approach for teaching is very popular in recent years in which students are engaged and motivated in the class.

The modern technology can easily be used to change the teacher centered methods to student centered approach. Since we live in a digital era, many digital tools are available for teachers to communicate with students. Some of these tools can be easily used for class assignments, and some can be used to enhance their knowledge. In this article we discuss some of the digital tools that can be used as effective teaching and learning methods for understanding statistical concepts.

Some of these tools are developed by the reputed associations related to the field of statistics. One such association is “American

Statistical Association”. In their website (<https://www.amstat.org/education/k-12-educators/>), they present classroom resources, publications in statistics education, guidelines and reports, workshops and webinars for teachers, and student competitions. The following are some selected interesting teaching tools shown in this website for which the java applets are used:

(i) Rossman/Chance Applet Collection (<http://www.rossmanchance.com/applets/>)

gives different types of applets for demonstrating introductory statistics concepts such as Data Analysis, Sampling Distribution Simulations, Probability, etc. Students can easily run these applets and can learn the relevant concepts independently.

(ii) Statkey (<http://lock5stat.com/statkey/>) website provides applets for descriptive statistics and graphs, bootstrap confidence intervals, and randomization of hypothesis tests.

(iii) Statlets

(<http://www.math.usu.edu/~schneit/CTIS/>) is also a very informative web page to learn statistical concepts interactively using the built in applets.

(iv) Rice Virtual Lab in Statistics (<http://onlinestatbook.com/rvls.html>)

also contains applets for simulations/demonstrations, case studies with example data, etc. Students can learn statistical concepts easily by using this site.

In addition to the above applets, many YouTube videos are available as independent learning tools. For example, the YouTube Video series introduced by Duke University professors and students (<https://www.youtube.com/user/profdstangl/playlists>) is an informative resource to teach basic descriptive statistical concepts. The YouTube videos included in the website, “Data to Insight”

(<https://www.stat.auckland.ac.nz/~wild/d2i/4StatEducators/>) also gives a hands-on introduction to statistical data analysis. Many such YouTube videos can now be found just by browsing the internet.

As an alternative to the java applets introduced earlier by the researchers, currently, the Shiny Apps are used as interactive teaching and learning tools for data analysis. Shiny (<https://shiny.rstudio.com/>) is an R package from which we can build interactive web apps. R Programming is one of the most widely used programming languages in the world today to analyze data. R was developed in 1995 by Ross Ihaka and Robert Gentleman when they worked together at the Department of Statistics, University of Auckland, New Zealand. The name 'R' was derived from the first letters of their names. One of the major reasons that R is immensely popular is that it is open-source and is continuously expanding since the developers and programmers from the whole world are contributing to its development. The Shiny Apps developed by many researchers can be found in the shiny Gallery (<https://shiny.rstudio.com/gallery/>). Some other shiny apps developed for data analysis are available on the Rsquared Academy website (<https://apps.rsquaredacademy.com/>). These apps can locally launch from the R packages mentioned in this website: <https://blog.rsquaredacademy.com/shiny-apps/>.

Another interesting R package that can be used for the teaching learning process is the "swirl" R package. Swirl (<https://swirlstats.com/>) turns the R console into an interactive learning environment. Students are guided through R programming exercises where they can answer questions in the R console. Currently, six interactive courses; R Programming E, Advanced R Programming, Exploratory Data Analysis, Getting and Cleaning Data, Regression Models, Statistical Inference; are available in "swirl". To use these courses, first install the "swirl" package in the R environment. Then, load it by using the code

library("swirl"). For example, to install Exploratory Data Analysis course, run the code `swirl::install_course ("Exploratory Data Analysis")` in the R console and continue to work in that course. Similarly, the other courses can be installed to the computer, and use them to learn the respective course. To begin a lesson, just type `swirl()`, and give a username, when prompted. Then, the learner can select a course. Following the instructions given step by step, the learner can independently complete the course easily. The teachers can also create their own interactive courses and share them freely with students in their classroom or around the world. The swirlify package can be used to write and edit swirl courses. To start writing a swirl course, it is recommended to read the Quick Start guide (<http://swirlstats.com/swirlify/quickstart.html>). Refer <http://swirlstats.com/swirlify/> for a more comprehensive documentation about writing swirl courses. Some other courses designed by the teachers can be found here: https://github.com/swirldev/swirl_courses.

If a teacher needs to prepare an interactive tutorial, the recently introduced learnr package (<https://rstudio.github.io/learnr/>) is the best option. The interactive tutorials will help teachers to examine the understanding and skills of students after completing a lesson. To prepare such an interactive tutorial, this link provides more information: <https://rstudio.github.io/learnr/articles/learnr.html>.

When performing data analysis, the usual practice is to run all necessary codes to get the results. However, it is very helpful in the teaching and learning process if the students can see the incremental code-output evolution to learn how the output changes step-by-step when adding R codes one by one. To fulfill this gap in the data analysis practical classes, now

the flipbookr R package can be used from which the user can present the code step-by-step and side-by-side with its output. Some details regarding how this package can be used are shown here: <https://www.rstudio.com/blog/building-code-movies-with-flipbookr/>.

Writing a professional report is also a challenging task in the applications in Statistics. Over the past years, R markdown documents have been used to prepare reports based on data analysis. Plenty of online materials are available to get an insight of R markdown. One such online book is R Markdown: The Definitive Guide (<https://bookdown.org/yihui/rmarkdown/>). Recently, the R Studio community introduced quarto (<https://quarto.org/>), which is an open source scientific and technical publishing system built on Pandoc. Quarto can be used to create dynamic content with R including the other programming languages; Python, Julia and Observable. Quarto is the best option to publish high-quality articles, reports, presentations etc based on applications in Statistics. An introduction to quarto is briefly explained in this video: <https://www.youtube.com/watch?v=6p4v0KS6Xls>.

Modelling: prediction is the name of the game



Prof. B.M.L.D.B. Suriyagoda
Department of Crop Science, Faculty of
Agriculture, University of Peradeniya,
Peradeniya, Sri Lanka

Prediction of the behaviour of biological systems, economy and markets, climate patterns, and many

other processes is done through modelling. Models need to be derived to make predictions. Models used for prediction can broadly be categorized either as statistical models or simulation models. The basic difference between these two models is that, statistical models make definite predictions with an associated probability while the simulation models make predictions without an associated probability. When making predictions of a particular system, for example under the Sri Lankan context, one can use an already existing model elsewhere in the world or can develop a model from the basic principles. Irrespective of these methods, before use in making predictions, three steps need to be followed. Those are; (1) model parameterization, (2) calibration and (3) validation. Most models fail in making accurate predictions due to the deficiencies in either one or more steps in above approaches.

Parameterization can be done either by changing/modifying the parameters of an already existing model for local conditions. In this process, model parameters need to be collected from local literature or need to be derived by the researcher, if they do not exist. If an already used model is not available elsewhere for the same purpose, one has to develop a model from basic principles. Once parameterisation is complete (i.e. model is developed) this model can be further fine-tuned by altering those parameter estimates slightly, considering their confidence intervals, to make more accurate predictions. This process is called as model calibration. In some instances, model calibration may take place once the parameterisation is over or otherwise model parameterisation and calibration processes may occur simultaneously. Once these two steps are complete, model is now ready for internal evaluation and this step is called as the model validation. In this process, model predictions

are tested against an independent data set of diverse nature. If there is a close agreement between the model predicted values and the actual observations such a model can be called a good model. Only at this stage can this model be used to make predictions. These predictions may be useful in making various decisions such as research knowledge synthesis, policy analysis or scenario comparison.

Once a model is developed for a particular system, it is expected to predict the performance of that system with high accuracy while that model may fail in making accurate predictions outside the range used to develop the model. Moreover, if the parameterisation and validation were not performed using a representative and adequate sample then the accuracy of model predictions can be questioned and may not be guaranteed. As a result, model predictions may not represent what can actually happen/observed in the field or nature. Then the trust on these predictions become weak and models fail. Therefore, modelling has to be done with much care and effort.

Replacing Data Scientists with Algorithms for Application of Scale



Dr. Sankha Muthu Poruthotage

VP Data Science and Product Management
Algonomy Software Pvt Ltd,
Colombo, Sri Lanka

Time Series Forecasting (TSF) is probably the most utilized data science use case in the world. For centuries, enterprises in both public and private sectors have applied TSF for various planning

related activities. Due to its wide popularity all stake holders generally grasp the concept of TSF and Data Scientists are regularly called upon to develop TSF models.

When faced with developing a TSF model, Data Scientists typically work on three aspects of the problem as below:

1. Data Preparation

This activity generally involves converting data to required formats and structures, detection and treatment of sparse-noisy-abnormal data and iterative feature engineering.

2. Model Fitting

This is where the data scientists will look at various modeling options. The options can vary from parametric time series models, ML algorithms, DL algorithms, smoothing techniques to simple rule-based methods.

3. Model Evaluation and Selection

There are several methods to select the best candidate model out of several competing models. While some of the methods are based on statistical inferences, methods that are based on training-testing paradigm are more popular due to their wide applicability.

While the above approach is somewhat generic to any predictive modeling problem (Classification and Regression) TSF tends to have some unique challenges due to the nature of the problems it is trying to solve.

One such challenge is the large scale at which TSF needs to be done. Following are a few examples of large scale;

- A supermarket chain which intends to forecast the demand for each of the products at each of the location:

A supermarket chain can typically have 100s of locations and 1000s of distinct products at each

location. Hence the scale of this problem can easily run in to millions of TSF!

- A telecommunication provider which intends to forecast the data consumption at each of its transmitting locations:

A telecommunication provider can have thousands of locations scattered across the large geography. The scale of this problem too can run in to tens of thousands of TSF!

- A FMCG distribution company which wants to forecast sales for all products at most granular sales region:

A distribution company can have 100s of sales regions and 100s of products that they distribute. Therefore, the scale of this problem too can be in thousands.

If a data scientist needs to provide a solution to a large scale TSF problem as the ones mentioned above, he/she will have to abandon the customized approach of Data Preparation, Model Fitting, Model Evaluation that was mentioned at the beginning of this article. That is simply because, such customized approaches cannot be replicated thousands of times. In such scenarios, data scientists typically resort to more simplistic approaches such as smoothing techniques. Alternatively, they may use a single parametric model with the same parameter structure. They may also use a single machine learning or a deep learning algorithm with the same hyper-parameter configuration.

However, such “One Size Fits All” approaches compromise on model fit and accuracy of results. In other words, accuracy is compromised for scale. This is what is called as the Accuracy Vs Scale dilemma in TSF!

A simple minded yet a novel approach to scaled TSF problem is algorithmizing the three steps mentioned above. In a way, this can be thought of as data scientists trying to replicate themselves through an algorithm!

While the concept of algorithmizing the steps of developing a TSF model may look simple, it can involve some tedious and cumbersome problem solving. Following is a high-level depiction of such algorithm.

Pre-Modelling Steps

- Automated Anomaly handling considering time effect and variable effect.
- Considering historical patterns to handle Anomaly points and adjust them.
- Automated seasonality detection.
- There are seasonality patterns in data which are not detected from the default seasonality settings, we need to capture them automatically and incorporate them for modelling.
- Fourier terms to capture additional seasonality effects.
 - Incorporating Fourier terms to capture additional seasonality effects.
- Automated feature selection and feature engineering.
 - Correlation related feature selection.
 - Feature importance-based feature selection.
 - Creating time related variables if needed.
- Smoothing time series methods to handle exemption periods.
 - Exemption periods are smoothed according to the historical patterns.

Modelling Steps

Following are the steps required in modelling:

- Univariate Smoothing methods
- Moving Average
- Simple Average
- Holt winters
- Decomposition method
- Optimized classical time series model building

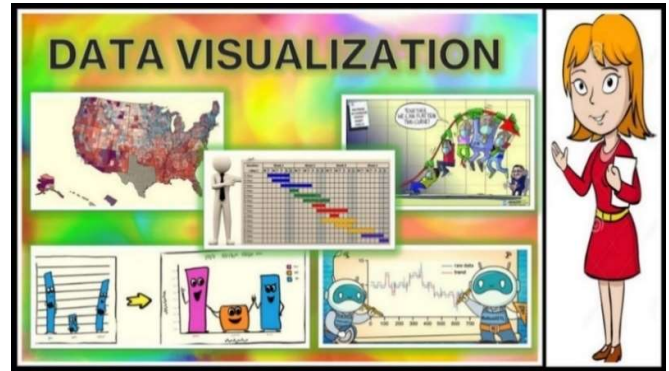
- Building large no of Arimax, Sarimax models with different parameters to select best fitted model.
- Ensemble Modelling
 - Classical time series (Arima/Sarima) + error (machine learning algorithm)
- Improved Machine learning and deep learning algorithms.
 - Random Forest, xgboost, Adaboost, KNNregression, NN
 - LGBM Regression
 - Improved prophet
- Capturing time related effects using
 - Time related variables,
 - Lag variables, and
 - Fourier terms.
- Automated Parameter tuning.

Model Selection Steps are as follows:

- Time related cross validation methods for model selection, and
- Considering different error measures for model selection.

An algorithmized approach to scaled TSF problem can resolve the Accuracy-Scale dilemma to a great extent. However, it should be noted that a good data scientist is still likely to beat the algorithm for a given forecasting problem or for a small set of forecasting problems. This situation however is not comparable since the algorithm is intended to solve the scaled TSF problem.

Finally, it should be noted that, a small number of forecasts out of a large forecasting exercise can turn out to be erroneous in an obvious way. This can generally happen in small datasets in which the testing period is limited to few data points. In order to minimize such forecasts, it is recommended to implement some guard rails around the outputs. Setting up guard rails is a fascinating algorithmic problem on its own, that can add immense value to a scaled TSF solution.



B. R. P. M. Basnayake

Lecturer

Department of Statistics and
Computer Science, Faculty of
Science, University of Peradeniya
Peradeniya, Sri Lanka

Data is beyond numbers, and it illustrates real life. In order to visualize data, one should understand what they represent. This is the fundamental path to data analysis. If you do not care what your data exactly represents, it will lead to misinterpretation. Here, the crucial point is that we should look beyond the mean, total, or median of data, as this kind of measurement implies only a slight part about the behavior of data. Hence, for storytelling and decision-making, we need to consider the aggregated data as they are showing interesting characteristics.

Data visualization is:

- **Representation** – Effective and suitable charts for visual perception of data.
- **Presentation** – Effective and suitable design principles (typeface, coordinate systems, scale, position, length, direction, angle, shape, color, scale, etc.) for charts and graphs.
- **Facilitate understanding** – Facilitate the viewer by removing obstacles related to design as you are unable to force the observer into the understanding procedure.

Design principles for charts and graphs

A figure may consist of one or several charts. A chart comprises a data area, axis names, axis labels, legends and headings. The foremost important feature of the design principle is the correct perception or the presentation form of the data. This is depending on the information one wants to deliver to the viewers. Unfortunately, typefaces or the design of letters based on size, weight, slope, etc. are neglected, but they have a significant contribution to the visualization of the figures. Hence, the application of an accurate typeface leads to much more clarity. The coordinate systems identify the dimensions of visualization, and the scale directs how data maps in the dimension. Generally, a rule of thumb in scaling is such that the direction fluctuates mainly around 45 degrees. However, this rule is not easy to maintain for real-world data. Hence, the appropriate task is to begin with this suggestion and adjust according to the situation. For instance, if there is a significant slight change, then it may be suitable to extend the scale to see the real shift. In contrast, if there is no significant slight change, it is better not to stretch the scale to make the change dramatic. Further, visual cues such as position, length, direction, angle, shape, area, volume and color are important aspects of data visualization. Color selection should be done consciously as the color blindness of some viewers may lead to misinterpretations.

There are many tools for data visualization such as Microsoft Excel, Tableau Software, Gephi, R software, etc. To conclude, data visualization helps viewers to obtain insights and patterns in data such that they can pay attention to required observations.

References

- Yau, N. (2013). *Data Points: Visualization that means something*. Wiley, ISBN: 978-1-118-46219-5
 Kirk A. (2013). *Data Visualisation: A Handbook for Data Driven Design 1st Edition*, ISBN-13: 978-1473912144
 Rahlf T. (2017). *Data Visualization with R*, Springer, ISBN: 978-3-319-49750-1

News in Brief

► 10th AGM of IASSL

The 10th AGM of IASSL was held on the 27th of March 2022 at the auditorium of the Organization of Professional Associations of Sri Lanka. Dr. D.M.J.B. Senanayake, Director of Rice Research and Development Institute was invited as the chief guest of the ceremony.

► Felicitating Yeoman Service to IASSL

The following IASSL members were felicitated at the 10th AGM of IASSL to felicitate yeoman service to IASSL for more than a decade:

- Prof. R.O. Thattil
 Mr. Jayantha De Silva
 Dr. (Mrs.) D.D.M. Jayasundara
 Mr. P. Dias
 Prof. K.A.P. Siddhisena
 Mrs. D.A.B.N. Amarasekara
 Prof. (Mrs.) N.R. Abeynayaka
 Prof. T.S.G. Peiris
 Dr. R.A.B. Abeygunawardana
 Mrs. Padma Yatapana
 Dr. D.C. Wickramarachchi
 Prof. C.D. Tilakaratne
 Dr. A.P.G.S. De Silva
 Mr. P. Sarukkali
 Mr. Priyath De Silva
 Prof. L.D.B. Suriyagoda
 Prof. S. Samitha
 Prof. M.R. Sooriyarachchi
 Dr. M.A.C.S.S. Fernando
 Mr. M.D.N. Gunarathne
 Prof. L.H.P. Gunarathne
 Mr. R.B.M. Korale

Prof. W.N. Wickramasinghe
 Prof. (Mrs.) Pushpa Wijekoon
 Dr. (Ms.) Chandanie Navaratna
 Prof. Sarath Bannaheka
 Ms. Chandrika Fernando

► From IASSL subcommittees

Certificate short courses conducted in Jan - April 2022

► Academic and Training Committee has conducted the following short courses in the months of Jan - April:

- **Basic Statistics for managers and researchers:**
Resource person: Prof. N. Rupika Abeynayake,
Department of Agribusiness management,
Wayamba University of Sri Lanka.
- **Business analysis with power BI:**
Resource person: Ms. Samudra
Bandaranayake, Senior Data Analyst, Wiley in
Sri Lanka.
- **Modeling binary, ordinal and nominal**
outcomes:
Resource person: Dr. Niroshan Withanage,
Senior Lecturer, Department of Statistics,
University of Sri Jayewardenepura.

► Winners of the Best Research Project Awards 2021

► Undergraduate Category

Winner – Gold Medal
 Shared Frailty Model for Joint Survival Data-A
 Simulation Study
 J.C. Liyanage, Department of Statistics, University
 of Colombo

1st Runner-up – Silver Medal
 Classification of Chronic Kidney Disease of
 Unknown Etiology Using Data Science
 Techniques
 L.T.H. Liyanage, Department of Statistics &
 Computer Science, University of Kelaniya

2nd Runner-up – Bronze Medal
 Bayesian Synthetic np Control Chart
 S.D. Silva, Department of Statistics, University of
 Sri Jayewardenepura

Merit Awards: D. H. N. Perera, G.C.J. Piyatilake,
 K. M. U. Gunarathne, S. R. N. Gunaratne, H. H. M.
 Hettiarachchi

► Postgraduate Category

Winner - Gold Medal
 A study of employment among Arts graduates
 in Sri Lanka with model-assisted approaches
 using complex survey data
 I.T. Jayamanne, Department of Statistics,
 University of Colombo

1st Runner-up – Silver Medal
 Three-way analysis methods to detect panel
 disconsensus in tea sensory evaluation
 D.R. Fernando, Postgraduate Institute of
 Agriculture, University of Peradeniya, Sri Lanka

2nd Runner-up – Bronze Medal
 Long Memory in Stock Market Returns:
 Evidence from Sri Lanka
 P.T.M. Gunathilake, Department of Business
 Finance, Faculty of Management, University of
 Peradeniya

► Open Category

Winner – Gold Medal
 Joint Modelling of two count variables using a
 shared random effect model in the presence
 of clusters for complex data
 Prof. M. R. Sooriyarachchi, Department of
 Statistics, Faculty of Science, University of
 Colombo, Sri Lanka

1st Runner-up – Silver Medal

Statistical tests for feature selection in plant pre-microRNA prediction

S. Lokuge, Department of Computer Engineering,
Faculty of Engineering, University of Peradeniya

2nd Runner-up – Bronze Medal

A Novel Hybrid Grey Exponential Smoothing Approach to Predicate Transmission Dynamics of the COVID-19 outbreak in Sri Lanka

D.M.K.N. Seneviratna, Department of Interdisciplinary Studies, Faculty of Engineering, University of Ruhuna

All awardees were awarded their prizes and/or certificates at the 2022 AGM.

► Felicitation Book

Launching Ceremony of Emeritus Prof. R.O. Thattil

The Launching of the Felicitation Book in Honor of Emeritus Professor R.O Thattil was held on 29th April 2022 at Prof RR Appadurai Auditorium of the Faculty of Agriculture, University of Peradeniya. He is the founder of the Applied Statistics Association Sri Lanka (ASASL) established in 1999. Institute of Applied Statistics Sri Lanka (IASSL) is the successor of the former ASASL.



► Diploma Awardees 2020/2021

IASSL Awarded Diploma and Applied Statistics to H.A.L.V. Hettiarachchi, P. Lakshika and L.D.A. Gimhana at the 2022 AGM.

► New Members of the IASSL: Jan – April 2022

1. Miss. U.M. Weerasena
2. Mr. S. Vaikunthan
3. Mr. K. Manimarrphan
4. Dr. C.L. Jayasinghe
5. Mr. R.M.P.S. Rathnayaka
6. Mr. K.A.P.M. Weerathunga
7. Mr. P.L.J.U. Sampath
8. Mrs. D.W.L.U. de Silva

► New members appointed for the OPA Young Professionals Forum

At the 122th Executive Council meeting held on the 27th of April, 2022, the following IASSL members were appointed as IASSL representatives in the OPA Young Professionals Forum:

- Dr. Rajitha M. Silva
- Dr. Vasana Chandrasekara
- Dr. Chathuri L. Jayasinghe

Photo Gallery

10th AGM of the IASSL – March 2022



Welcoming the Hon. Chief Guest Dr. D.M.J.B. Senanayake, Director, Rice Research and Development Institute



Lightning oil lamp (Left to right): Hon. Chief Guest, & Dr. D.C. Wickramarachchi (president of IASSL) & Prof. L. Suriyagoda (immediate past president of IASSL)



Address by the Chief Guest



Address by the President of IASSL



Awarding IASSL fellowships (clockwise from top-left): Dr. A.P.G.S. De Silva, Dr. C.D. Tilakaratne & Mrs. D.A.B.N. Amarasekara



Participants of the AGM

Felicitation Ceremony: Felicitating Yeoman Service to IASSL



From left to right row-wise;
 First row: Prof. R.O. Thattil, Prof. (Mrs.)
 N.R. Abeynayaka;
 Second Row: Dr. A.P.G.S. De Silva, Prof.
 L.D.B. Suriyagoda;
 Third Row: Prof. S. Samitha, Prof. C.D.
 Tilakaratne,
 Fourth Row: Prof. T.S.G. Peiris

Felicitation ceremony (continued)



From left to right row-wise;
 First row: Prof. M.R. Sooriyarachchi,
 Mrs. P. Sarukkali (on behalf of Mr. P.
 Sarukkali)
 Second Row: Mrs. D.A.B.N. Amarasekara,
 Mrs. Padma Yatapana,
 Third Row: Dr. R.A.B. Abeygunawardana,
 Mr. Priyath De Silva
 Fourth Row: Ms. R.P.K.Sandunika Nishani

Winners of the best research project 2021: Awards ceremony

Undergraduate Category



Winners of the undergraduate category (clockwise from top left): J.C. Liyanage (Gold Medal), L.T.H. Liyanage (1st Runner-up), S.D. Silva (2nd Runner-up)

Merit Awards





Merit Awardees of the undergraduate category (row-wise from left to right): first row: H. H. M. Hettiarachchi, S. R. N. Gunaratne; Second row: K. M. U. Gunarathne D. H. N. Perera,; Third row: G.C.J. Piyatilake

Postgraduate Category



Winners of the postgraduate category (clockwise from top left): I.T. Jayamanne (Gold Medal), D.R. Fernando (1st Runner-up), P.T.M. Gunathilake (2nd Runner-up)

Open Category



Winners of the open category (clockwise from top left): Prof. M. R. Sooriyarachchi (Gold Medal), S. Lokuge (1st Runner-up), D.M.K.N. Seneviratna (2nd Runner-up)

Diploma in Applied Statistics – 2020/2021: Awarding Ceremony



Diploma in Applied Statistics Awardees (clockwise from top left): H.A.L.V. Hettiarachchi, P. Lakshika, L.D.A. Gimhana

New Executive Council – 2022

President -	Dr. Chitraka Wickramarachchi
Immediate Past president-	Prof. L.D.B. Suriyagoda
Vice President, Chairperson/ House & Finance Committee	Dr. M.A.C.S.S. Fernando
Secretary -	Dr. R.A.B. Abeygunawardana
Asst. Secretary -	Dr. Isuru Hewapathirana
Treasurer -	Mr. Priyath De Silva
Asst. Treasurer -	Dr. C.W.C. Silva
Editor	Dr. N.V. Chandrasekera
Associate Editor	Dr. C.L. Jayasinghe
Chairperson/ Academic Training Committee -	Dr. Nirosan N. K. Perera
Chairperson/ Research & Development Committee -	Prof. Kapila T. Rathnayaka
Chairperson/ Statistics Popularization Committee-	Dr. Rajitha M. Silva
Executive Council Members-	Prof. C.D. Tilakaratne
	Prof. S. Samita
	Dr. A.P.G.S. De Silva
	Prof. (Mrs.) N.R. Abeynayake
	Mrs. D.A.B.N. Amarasekara
	Prof. T. Sivananthawerl
	Ms. Padma Yatapana

Upcoming Events by IASSL

All Island Competition in Statistics

IASSL is happy to inform that as an initiative of the Statistics and Popularization Committee of IASSL, an *all island competition in Statistics for School students* will be organized this year (November/December). 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.

Stat Undergrads Column

During the most recent T20 match between Sri Lanka and Australia which was held on 11th 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). 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. The accuracies vary from 53.08% (first ball of the match) to 97.65% (last ball of the match). This research study has

REAL - TIME PREDICTION AND STRATEGY FORMULATION IN T20 CRICKET

PREDICTION WITH REAL - TIME WIN PROBABILITY



- Predicts who wins a T20 match while the match is in-progress, from the first ball itself (on a ball-by-ball basis)
- Uses three techniques: Naive Bayes, Logistic Regression and SVM
- Gradually increasing prediction accuracy as the match progresses

STRATEGY FORMULATION

- Formulates strategies with the respective winning probabilities of the strategies for the two teams in the second innings
- Provides quantitative answers to the following types of what-if questions
 - What if we bat slowly for the next 5 five overs without losing wickets, and scoring only two boundaries? If we do so, what will be our chances of winning?
 - Can we hit 3 more boundaries in the next three overs while losing a wicket? If we do so, how will our winning chances look like?
- Imagine the following situation:
 - You are the coach of the chasing team
 - Your team is chasing a score of 145 in 20 overs
 - Your current score is 38 for the loss of two wickets after 6 overs.
 - You want to know what your chances of winning will be by the end of 10th over:
 - If you score 6 boundaries and no sixes inside the next 4 overs while eventually achieving a run rate of 7.2 for the loss of 1 more wicket.
 - Interestingly, the strategy formulator will be able to let you know your probability of winning if you follow the above strategy!
 - More importantly, the strategy formulator will give you insights into more and more such strategies with their respective winning probabilities.

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.

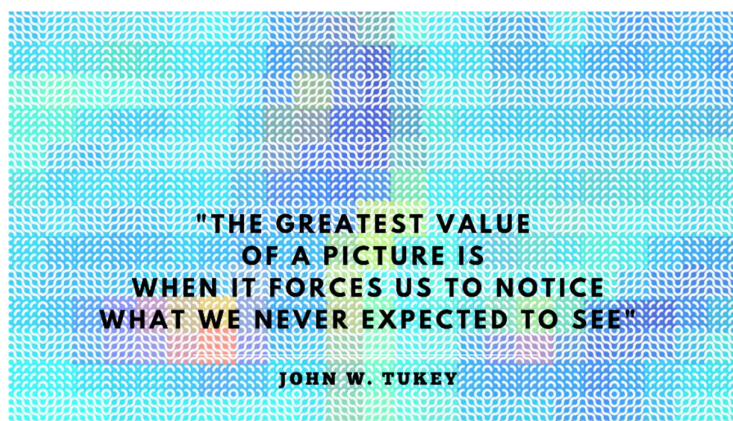
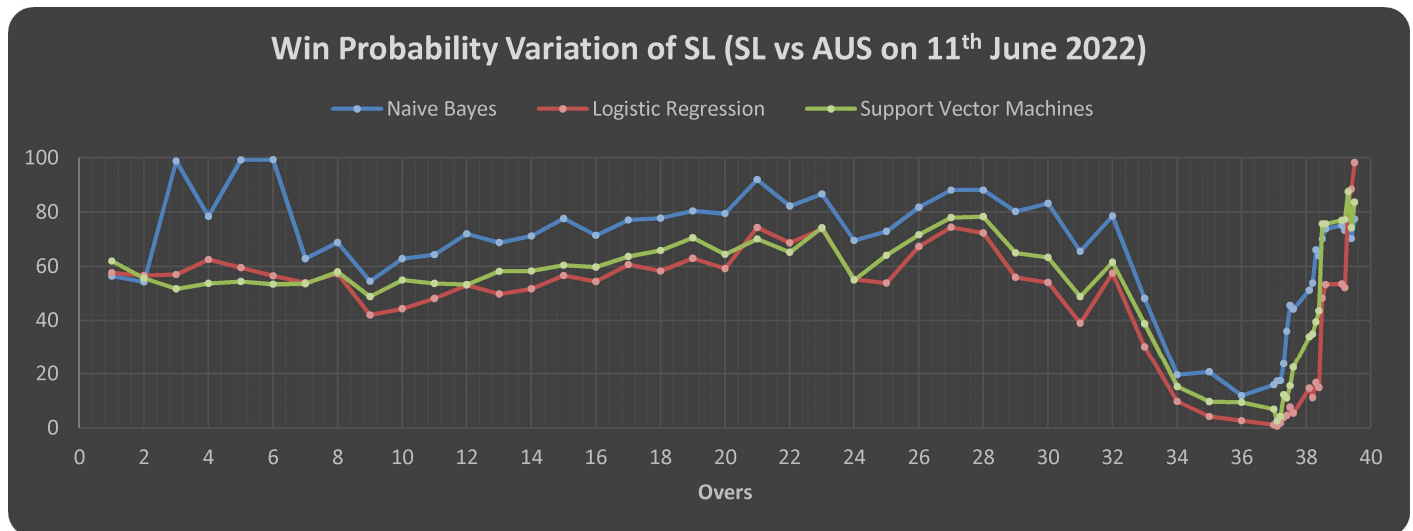
Authors: LCP Pussella¹, RM Silva², WCP Egodawatta³

¹Department of Physical Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka

²Department of Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka

³Department of Plant Sciences, Rajarata University of Sri Lanka, Anuradhapura, Sri Lanka

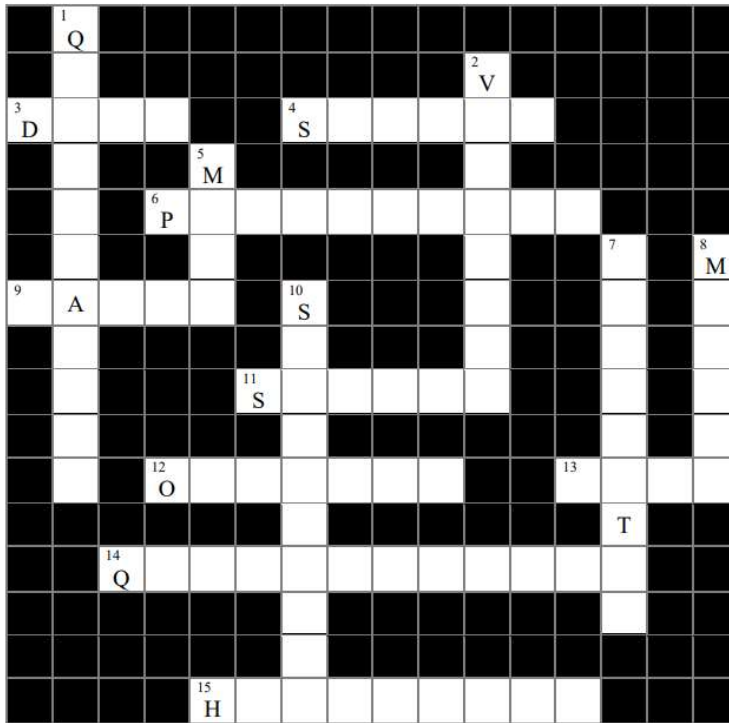
This article was written based on the first author's final year research project.



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 (by Janith Wanniarachchi, Fourth year student, BSc (honours in statistics), USJ).

Sudoku Puzzle Competition

STATISTICS CROSSWORD PUZZLE



INSTRUCTIONS

Across

3. The raw information from which statistics are created
4. A descriptive feature in which describes how similar or varied the set of observed values are for a particular variable
6. The entire pool from which a statistical sample is drawn
9. The difference between the lowest and highest values
11. A small part or quantity intended to show what the whole is like
12. The type of data in which the data values follow a natural order
13. The arithmetic average of a distribution, obtained by adding the

scores and then dividing by the number of scores

14. Numerical, you can perform meaningful math
15. A graph that displays the distribution of continuous data

Down

1. Data identified by something other than numbers
2. An attribute that describes a person, place, thing, or idea
5. The datum which occurs the most in a set of data
7. Any numerical quantity that characterizes a given population or some aspect of it
8. The middle score in a distribution; half the scores are above it and half are below it
10. Practice or science of collecting and analyzing numerical data in large quantities

First prize: Rs. 5000

Second prize: Rs. 3000

Third prize: Rs. 2000

Please email your submission to appstatsl@gmail.com on or before 30th of July, 2022. The draw will be held on the 15th of August, 2022 and the winners will be announced in the May-August 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.

Upcoming Certificate Short Courses by IASSL

The following short courses will be conducted in the coming months by IASSL:
July, 2022

- ▶ **Structural equation modeling (SEM) with AMOS and SmartPLS:**
Resource person: Dr. Chathurani Silva, Senior Lecturer, Department of Decision Sciences, University of Sri Jayewardenepura.

- ▶ **Analyzing Multivariate time series and modeling volatility with EViews/R:**
Resource person: Dr. Hasanthi Pathebariya, Senior Lecturer, Department of Statistics, University of Sri Jayewardenepura.

- ▶ **Data analysis for social sciences:**
Resource person: Mr. N.A.N.J Maduwansha, Senior Lecturer, Department of Social Statistics, University of Sri Jayewardenepura.

August, 2022

- ▶ **Research methodology for business and managerial researchers:**
Resource person: Dr. Chathurani Silva, Senior Lecturer, Department of Decision Sciences, University of Sri Jayewardenepura.

- ▶ **Basic Statistics for managers and researchers:**
Resource person: Prof N. Rupika Abeynayake, Department of Agribusiness management, Wayamba University of Sri Lanka.

September, 2022

- ▶ **Business analysis with power BI:**
Resource person: Ms. Samudra Bandaranayake, Senior Data Analyst, Wiley in Sri Lanka.

- ▶ **Modeling binary, ordinal and nominal outcomes:**
Resource person: Dr. Nirosan Withanage, Senior Lecturer, Department of Statistics, University of Sri Jayewardenepura.

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

Announcements

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.

Contributions to the May-August 2022 Newsletter:
 If you have any submissions, comments, suggestions & feedback, Please send them to editor.iassl@gmail.com.



Be a proud member of IASSL. The application can be downloaded from www.iappstat.lk

	<p>Institute of Applied Statistics, Sri Lanka has been incorporated by the Parliament of Sri Lanka on 20th September 2011 by the Institute of Applied Statistics, Sri Lanka (Incorporation) Act No. 38 of 2011</p>
<p style="text-align: center; color: #1F77B4;">BENEFITS</p> <ul style="list-style-type: none"> Recognition of Professional Competence Networking Opportunities Continuing Professional Development <p style="text-align: center; color: #1F77B4;">Journal</p> <p>The Sri Lankan Journal of Applied Statistics(SLJAS) is an open-access, international, double-blind peer-reviewed journal published by IASSL</p>	<p style="text-align: center; color: #8B0000; font-weight: bold;">INSTITUTE OF APPLIED STATISTICS, SRI LANKA</p> 
<p style="color: #1F77B4;">Facebook</p> <p>www.facebook.com/iassl2020</p>	<h2>Apply for Membership</h2>
<p style="text-align: center; color: #1F77B4;">Contact Us</p> <p style="text-align: center;">1The Professional Center, 275/75, Prof. Stanley Wijesundera Mawatha, Colombo 7, Sri Lanka</p>	<p>"The only professional organization committed to promote and assist the advancement of Statistics for the furtherance of research, planning, development, education, and training in Sri Lanka."</p>



We sincerely appreciate all who contributed to this issue and participated in its' production.

- Editorial Board/IASSL

IASSL Newsletter

Official Newsletter of the Institute of Applied Statistics Sri Lanka



**Institute of Applied Statistics, Sri Lanka
Level 02,
Professional Center
275/75 Prof. Stanley Wijesundera Mawatha,
Colombo 07,
Sri Lanka.**