ISSN 1391 4395 May -Aug 2023

IASSL NEWSLETTER

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.iassl.lk
http://www.facebook.com/iassl2020/
https://www.linkedin.com/company/iassl/

"Statistics is the grammar of science"

Karl Pearson

Featured Segments

From IASSL President's pen

Dr. Niroshan Withanage
President/IASSL

Welcome to the second issue of the IASSL Newsletter for the year 2023. It is my great pleasure to share some thoughts with you through the pages of the IASSL Newsletter. I would like to extend my sincerest appreciation to the dedicated Editorial Board and the distinguished authors who have consistently provided valuable articles, ensuring that the IASSL Newsletter is published on time.

In my capacity as President, I am delighted to report that IASSL is making remarkable progress across all its sub-committees. The Editorial Board has done a commendable job, and our Sri Lankan Journal of Applied Statistics (SLJAS) is now published seamlessly without any delays. I want to express my gratitude to all the local and international authors who have submitted high-quality papers.

Our Statistics Popularization Committee has been actively engaged in conducting numerous seminars for schoolchildren, aiming to enhance their knowledge in Statistics and equip them with the confidence to tackle Statistics-related questions in the Advanced Level Combined Mathematics paper. I extend my heartfelt thanks to all the dedicated resource persons who have volunteered their time for this crucial initiative. I also appreciate the tremendous support provided by school principals and teachers, without which these seminars would not have been a success.

Our Research and Development Committee has been collaborating with government and private organizations to promote and support the advancement of Statistics in their research and planning efforts. The Academic Training Committee has introduced new short courses in response to high demand, which has not only enriched our offerings but also contributed significantly to improve the standards of the stakeholder.

I would like to take this opportunity to express my gratitude to all the chairpersons of subcommittees and their members for the tremendous work they have done to enhance IASSL's standing.

The National Statistics Olympiad competition is just around the corner, and I invite all senior and junior students to participate and showcase their talents. Winners of this national competition may have the opportunity to represent Sri Lanka in the Statistics Olympiad competition conducted by the C.R. Rao Advanced Institute of Mathematics, Statistics, and Computer Science (AIMSC) in India.

I am excited to announce that the highly anticipated International Statistics Conference – 2024 is scheduled to take place on the 28th and 29th of December 2024 at the Mount Lavinia Hotel. The organizing committee is working diligently to ensure the success of this significant event.

As we move forward, I am confident that our collaborative efforts will continue to drive the advancement of statistics in research, planning, development, education, and policy-making in Sri Lanka.

Thank you for your continued support and I look forward to the exciting journey ahead.

Warm regards,
President, IASSL.

Editorial

Prof. 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 2023. In this issue, we will recount various events, projects and activities in which IASSL members were actively involved from the 1st of May 2023 until the 31st August 2023. Basically this issue contains articles from senior academics and industry professionals in the field of Statistics, One-Act-Play article from an emirates professor and Stat Undergrad Column with articles from undergraduates. Further this issue comprises News in brief which cover all events of IASSL during the considered period of this newsletter.'

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 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, secretory, all subcommittee chairpersons and executive council members of IASSL in providing information relating to the events conducted by them during the period May to August 2023. Last but not least, I would like to thank the editorial board members 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 consider in the Issue 3, 2023 of the IASSL

newsletter (editor@iassl.lk) and hope you all will enjoy reading this issue.

Editor/IASSL.

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.lk

One Act Play: Power Game

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

[The following discussion takes place between a Chemistry lecturer(C), a Biologist (B) and a Statistician (S), during an intermission at a research seminar].

B - I was astonished that most of the research papers presented up to now used parametric methods of data analysis even for discrete data.

C - What do you mean by discrete data?

B - I am not referring to count data but to data that is not continuous (ranked data).

C - What is the problem? Parametric methods have the highest power when testing Hypothesis!

S - That is true only if the data is continuous. However, if the sample size is large parametric tests can be applied even for discrete data. But, parametric tests were used by many researchers for discrete data even for small sample sizes.

C - What are the assumptions necessary to validate the use of parametric methods?

S- If the data can be assumed to be normally distributed, then all commonly used methods such as 't' tests, 'F' tests, analysis of variance, regression analysis can be used. Even if the normality assumption cannot be used, if the distribution is known one can construct the test statistics. In any case the data has to be continuously distributed.

B - How can we test the assumption of normality?

S - There are visual methods such as box plots and also more complicated analytical methods.

C - So, what methods can we use for discrete data?

S - The tests that have been used for discrete data analysis, do not assume any distribution. The methods used, are called distribution free statistics. The data are measured on a rank scale.

C - Can you explain what you mean by the rank scale?

S - All right. There are 4 scales of data measurements, namely; nominal, rank (or ordinal), interval and ratio scales. We will restrict ourselves to rank data, i.e. data that can be categorized with relationship between the categories.

C- Can you give us some examples of such data?

S - Letter grades given in evaluating the performance of students in a given subject, usually A, B, C, D and E, where the relationship is A>B>C>D>E. Another example is grading of food quality, where excellent, very good, good, poor, very poor are used in grading. Obviously the relationship is excellent>very good>good>poor>very poor.

C - O.K., I get the idea. But, what about the power of the test?

S - The tests used for rank scale data are called non-parametric tests. They have lesser power than parametric tests. But power is of no consideration, when we cannot satisfy the assumption made in parametric testing.

B - This is what puzzles me. Why do researchers use keep on using parametric tests for discrete data?

S - I believe that it is lack of knowledge about non-parametric methods.

B - Are there analogous methods in non-parametric statistics for every parametric method?

S - Of course. In fact there also a few non-parametric methods which are not found under parametric methods.

B - Are there regression and correlation procedures for non-parametric data?

S - Yes, of course.

B - Is there a difference in power for different non-parametric methods?

S- Let me make a general statement at this juncture. More information used in the measurement of data, more powerful tests can be found. For example the Wilcoxon signed rank test (analog for the paired ‘t’ test) looks at the sign as well as the magnitude of the difference and is therefore more powerful than the sign test, which only looks at the sign of the difference(+ for positive difference and – for negative difference) for comparing paired data. More information used in a test leads to more power in the test. Parametric procedures are more powerful since, they also assume a distribution (usually the normal distribution) additionally. Also, if continuous data are categorized we lose power in testing.

B - You mentioned that letter grades used in evaluating student performance are dealt using non-parametric techniques. How about the raw marks itself? Aren't they continuous? Can't we use parametric methods where raw marks are concerned?

S - Parametric techniques are used for raw marks when sample sizes are large. However, raw marks do not possess the attributes of interval or ratio data which are truly continuous data. The difference between 50 and 40 marks (i.e. 10 marks) is not the same as the difference between 95 and 85 marks although the difference is 10 in magnitude. In the interval scale, equal differences are comparable. Moreover 60 marks is not the double of 30 marks (a property of ratio scale data). Strictly speaking raw marks are non-parametric and we should use non-parametric methods only, unless the sample sizes are large. We can discuss other complications at a different time.

C - Thank you very much. I should learn more about non-parametric methods.

B - I also realize that my knowledge, especially of non-parametric methods is very poor. Thank you for the discussion.

MODELLING CHARACTERISTICS OF CHOICES – DISCRETE CHOICE MODELS

Senior Professor S. Samita
Department of Crop Science, Faculty of Agriculture,
University of Peradeniya,
Sri Lanka.

Binary responses are usually modelled using binomial generalized linear models, most commonly the linear logistic model. With more than two response categories, i.e., a discrete set of alternatives, multinomial generalized linear models, most commonly the multinomial logit model is fitted. Further, when the response categories are nominal–scale, the multinomial logit model fitted is the baseline category logit model, which is the generalization of the binomial logit model. The parameter estimation for all these models is made using maximum likelihood method. For a categorical response variable, Y, with a fixed set of J categories and $\mathbf{x}$ predictors, let $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$ for $j = 1, 2, \cdots, J$. Then the baseline category logit model can be defined by

$$ln\frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T\mathbf{x}, \quad j = 1, 2, \cdots, J-1 \tag{1}$$

where $\boldsymbol{\beta}$ is the parameter vector. As an example, one might fit such a model to explore whether preference to hobby is associated with, age, gender, profession etc.

Discrete Choice Model

Now, consider a situation where the interest is whether choice to brand of a product is associated with price of the brand. If we compare this situation with a situation where the interest is choice to the brand of product is associated with gender of the user, here the type of predictor price is different from, type of predictor gender. The difference is that, variable gender does not vary depending on the choice of brand, whereas price varies depending on the choice of brand. It can be seen that the predictor gender, is a characteristic of the chooser but price is a characteristic of the choice. In fact, this kind of situations are quite common. As another example, when modelling choice of mode of transport to go to work (train, bus, or car) on predictors, time taken for travel, and cost, time and cost vary depending on the choice. Thus, whenever it is a characteristic of the chooser, this issue does not arise, and whenever it is a characteristic of a choice, this issue arises.

Actually, one of the assumptions taken in fitting the multinomial logit models, such as baseline category logit model, is that the value of the predictor variable from an observation does not vary depending on the choice. However, when the predictors are characteristics of choices, this assumption cannot be fulfilled, and an alternative modelling approach has to be sought. The type of multinomial models in general, specifically the type of multinomial logit models, fitted for this type of situation is called Discrete Choice Model (DCM) (Johnson, and Hensher, 2019). The DCM not only provides a solution to above issue but also to the issue that for some individuals all choices are not available. As an example, to travel to work, for instance, it is possible the option train is not available for some, and option bus is not available for some others. DCMs themselves are of several types, and among them the most commonly used type is the conditional logit model.

Conditional logit model

The conditional logit model in most general form is defined of the form

$$ln\frac{\pi_a(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)} = \boldsymbol{\beta}^T(\mathbf{x}_{ia} - \mathbf{x}_{ib}) \tag{2}$$

where a and b are each pair of choices in the choice set C_i of the i^{th} subject, $\mathbf{x}_{ia}$ and $\mathbf{x}_{ib}$ are vectors of predictors for a^{th} and b^{th} choices respectively.

Modelling multinomial responses on characteristic of choices is quite common in Econometrics and Psychology. Especially in economics, when modelling utility of choices, conditional logit model is often used. In Statistics literature, McFadden (1974) described the model for the utility of j^{th} choice of i^{th} individual, U_{ij}, of the form

$$U_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \epsilon_{ij} \tag{3}$$

where ϵ_{ij} are identically independently standard extreme value distributed. Note that standard extreme value distribution is also known as the type I extreme value distribution or Gumbel distribution. In the utility theory, if the response is j for the i^{th} subject, then the i^{th} subject has maximum utility for the choice j. McFadden (1974) showed that the model of equation (3) with Gumbel distribution for $\boldsymbol{\epsilon}_{ij}$ is equivalent to the conditional logit model (equation 2), and it paved the way to study utility of different choices. Note that in (2) and (3), compared to (1), $\boldsymbol{\beta}$ is constant across choices. This is because there is a constant marginal utility for an additional unit of X variables. For instance, engine power varies depending on the car (choice dependent) but utility for an additional unit of engine power is constant across all cars.

Independence from irrelevant alternatives

From equation (2) it is clear that the ratio of probabilities for choices a and b are independent from other alternatives. In fact, from equation (1) also it can be seen that $ln[\pi_a(\mathbf{x})/\pi_b(\mathbf{x})]$ is independent from other alternatives. This property, which can be seen in conditional logit model and multinomial logit model, is called Independence from Irrelevant Alternatives (IIA) (Agresti, 2013). For instance, in an election with 3 candidates, if one withdraws, ratio of probabilities between other two does not change. It is expected that voters of withdrawn candidate will vote the other two in the same ratio as before the withdrawal. The IIA is an important assumption we take in fitting those models, and it is a follow–up from the assumption that $\boldsymbol{\epsilon}_{ij}$ are distributed independently and identically.

The IIA is unrealistic in some applications. For instance, to travel to work, suppose 60% choose train and the rest SLTB bus. Then the odds for train is $(0.6/0.4) = 1.5$. Now imagine private buses were also introduced to the area. Then intuitively we expect, some of those who travel by SLTB bus will shift to private buses and thereby odds of train instead of SLTB bus to be about

3.0, with about 20% using each bus option. For such situation conditional logit model may not be appropriate.

Nested Logit Model

Sometimes we find choices are naturally grouped. For instance, for mode of transport, with choices, train, bus and car, busses could be a group of two types, such as government and private. (Fig 1).

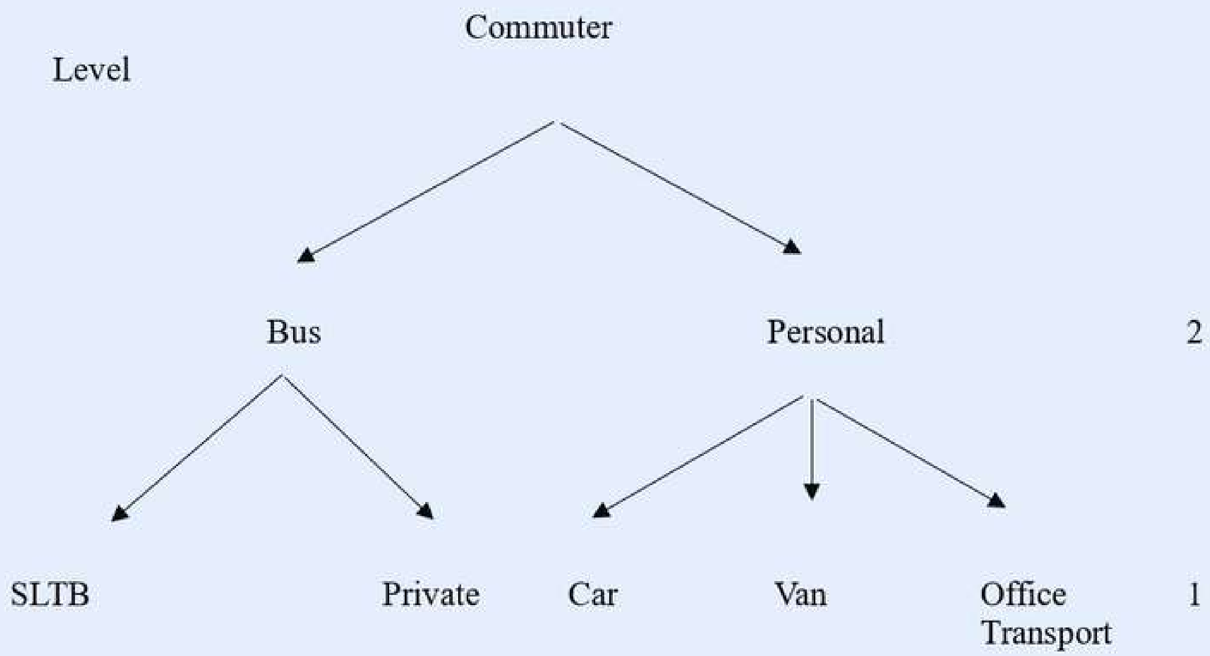


Fig 1. Two–level decision tree

In Fig. 1, private could also be a group with two companies providing the service. Then the arrangement becomes a three–level decision tree. Likewise, the decision tree could be with any number of levels. In Statistics, an arrangement such as in Fig 1 is called nested or multilevel. In this kind of a situation, the IIA cannot be assumed for all pairs in the DCM, and the IIA restriction may be relaxed to have it of the form, for some pairs it holds but not for all pairs. When choices can be partitioned into subsets, then it could be assumed that the IIA holds within subsets but not across subsets. The DCM that can accommodate this situation is nested logit model. The utility model for the nested logit model is similar that of conditional logit model (equation 3) but the joint distribution of errors, ϵ_{ij} are assumed to be generalized extreme value, which is a generalization of Gumbel distribution. Moreover, ϵ_{ij} are distributed identically, and within each subset they are correlated.

Let C_{kh} is the k^{th} choice set at h^{th} level with M_h number of choice sets at h^{th} level. The nested probability is decomposed into two logit models, probability of j^{th} choice within the k^{th} nest at h^{th} level and probability of choosing of k^{th} nest at h^{th} level. Here probability of choosing k^{th} nest at h^{th} level is associated with a quantity called inclusive value which is the average utility a subject can expect from the nest below. Since different nests can have different variances, factors are scaled, so that effects can be discussed in terms of standard deviation units. The ratio of scale factors between upper level and lower level provides the estimate of the correlation between choices within a nest. Thus, the impact of characteristic of choices can be studied under different levels of correlation within a nest using the nested logit model. If there is no correlation, then the nested logit model is equivalent to conditional logit model and IIA holds.

Mixed logit model

Another alternative to conditional logit model when errors are not independent, and not identical (heteroscedastic), is the mixed logit model. By means of utility, the model is of the form

$$U_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \xi_{ij} + \epsilon_{ij}$$

where ϵ_{ij} are identically independently standard extreme value distributed, and ξ_{ij} is also an error term, and a general distribution is assumed for these errors. These errors take account of correlation among choices and heteroscedasticity for each individual.

Mixed logit is derived by integrating logistic model function with respect to a distribution of model parameters (Train, 2009). Let the density function of the error component ξ_{ij} is denoted as $f(\xi_{ij}|\boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is a parameter vector of the distribution of ξ_{ij}. Then the probability of choice j for individual i is given as

$$\pi_i(j) = \int Q_i\left(j|\xi_{ij}\right) f\left(\xi_{ij}|\boldsymbol{\delta}\right) d\xi_{ij}$$

where $Q_i(j|\xi_{ij})$, the conditional probability of choice j for a given value of ξ_{ij}, is the logit

$$Q_i(j|\xi_{ij}) = \frac{\exp\left(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \xi_{ij}\right)}{\sum_{h \in C_i} \exp(\mathbf{x}_{ih}^T \boldsymbol{\beta} + \xi_{ih})}.$$

In fact, because of the choice probability is a mixture of logits with a function of ξ_{ij}, this model is called a mixed model. Due to the flexibility of the model, it is used under various circumstances. For instance, modelling food choices, hobbies etc. this model may be more appropriate as expectations from them vary widely between individuals. In general, maximum likelihood estimates are not available and estimates are obtained through simulation methods.

Other alternatives for conditional logit model

Some other possible alternatives when IIA is not applicable are heteroscedastic extreme-value model and multinomial probit model (Hensher, Rose and Greene, 2005). In terms of utility, both models are similar to conditional logit model. However, in the ϵ_{ij} of heteroscedastic extreme-value model are assumed to be independently nonidentically heteroscedastic extreme-value distributed. However, in the multinomial probit model we assume $\boldsymbol{\epsilon} \sim N_J(\mathbf{0}, \boldsymbol{\Sigma})$, i.e., multivariate normally distributed with parameters $\mathbf{0}$ and $\boldsymbol{\Sigma}$. Similarly, with multinomial probit model, ϵ_{ij} are assumed to be correlated and nonidentically multivariate normally distributed. With both methods parameter estimates are obtained using numerical methods or simulation.

Implementation

Although DCM became an established statistical technique quite some time ago, use of the model or the practice was rather limited, mainly because of difficulty in implementing it through the software packages. However, current standard software packages can easily be used to fit this model. In fact, free software packages such as SAS (SAS on Demand for Academics) and R can easily be used to fit this model. Moreover, three SAS Procedures, namely, PROC GENMOD, PROC PHREG and PROC MDC are available to fit this model. With R, packages DPLYR, AER and MLOGIT facilitate DCM fitting.

When modelling a response variable consisting a discrete set of options, not that all predictor variables are always characteristic of choices. It is possible some are characteristic of choosers. Since characteristic choosers are usually modelled by fitting multinomial logit models, in a situation with both type of variables, characteristics of choosers are modelled using the multinomial logit approach while characteristic of choices are modelled using conditional logit, or nested logit etc. depending on the situation, in the same model. Simply by organizing the data of characteristics of choosers in the appropriate way, they can be accommodated along with characteristics of choices in DCM fitting.

References

Agresti, A. (2013) Categorical Data Analysis, Third edition, Wiley, New York.

Hensher, D. A., Rose, J.M., & Greene, W.H. (2005). *Applied choice analysis: A primer*. Cambridge: Cambridge University Press.

Johnson, L. W. and Hensher, D. A. (2019). Applied Discrete-Choice Modelling, Routledge, UK

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. New York: Academic Press.

Train, K. E. (2009) Discrete Choice Methods with Simulation, Second edition, Cambridge University Press

Research study:

-A view on how to do it?

D. J. C. Suriyaarachchi
Former Senior Lecturer,
Department of Statistics & Computer Science,
University of Kelaniya,
Sri Lanka.

Introduction

This article provides a view of how a research study can be done. It starts with what is research and the purpose of research. Second section describes what is a research problem statement and how to formulate it. Third section describes how to setup test criteria. Fourth section indicates a procedure to select an experimental design for data collection. Final section explains how to perform inferential data analysis.

What is research and the purpose of research

What is research?

Research is considered as a careful consideration of study regarding a particular concern. It involves inductive and deductive methods. Inductive method associates with qualitative research while deductive method more commonly associates with quantitative research. In inductive method events are observed and then analyzed while in deductive method events are observed and then verified.

Purpose of research

There are three main purposes namely; Exploratory, Descriptive and Explanatory/Cause and effect.

Exploratory:

Exploratory to simply explore the research questions with its answers and analytics but not to provide final and conclusive solutions to existing problems. It uses to lay the foundation for more conclusive data collection and analysis. It also uses to determine the nature of the problem and helps better understanding of the problem.

Descriptive:

Descriptive to aim the accurate and systematic description of a population, or phenomenon. It uses to answer the questions of the forms “what”, “where”, “when” and “how”, but not the questions of the form “why”. Descriptive applies to a wide variety of research methods for investigating one or more variables and to focus on expanding knowledge on current issues through a process of data collection. Descriptive also uses to describe the behavior of a sample population and also to explain and validate the findings.

Explanatory/Cause an effect:

Explanatory uses to discover details about why something occurs and helps to understand the impact of specific changes in existing standard procedures.

Comparative analysis chart			
	Exploratory	**Descriptive**	**Explanatory/Cause an effect**
Method	Unstructured	Structured	Highly structured
How to conduct	Asking questions	Asking questions	By using hypothesis
Time	Early stages of decision making	Later stages of decision making	Later stages of decision making

In descriptive research, a research problem/question cannot be formulated because it answers the questions in the forms of "what", "where", "when" and "how" but not in the form "why". Explanatory/cause and effect is able to find details about why something occurs. Thus, a research problem/question can be formulated in this situation. By saying a research problem, it means a specific question, problem, or difficulty that needs to be investigated or analyzed during the research study. It helps to identify the key concepts and terms of research study. As of this reason a research problem should be clear, concise, and specific enough to guide the process and contribute to the definition of research project objectives, methods, and outcomes. This is the foundation of any research study.

Research problem statement and formulation of a research problem

Research problem statement

A research problem statement starts with a broad topic in the research area and then narrows down to a specific research question which ultimately leads to a framework for constructing a hypothesis.

Final statement of the research problem should be a brief and precise description of the problem that helps to construct appropriate hypothesis. This is usually written in interrogative tone.

Formulation of a research problem

Here are the five basic steps to formulate a research problem:

- Start with identifying a broad research area based on the research project.
- Break down the broad area into sub-areas.
- Choose a specific research topic from among sub-areas.
  Specific research topic selected should be manageable and feasible for the researcher and should have some research significance.
- Formulate a research question.
  Start with noting down important questions that considered important for the research study. If many questions found, narrow down and choose the one most important and impactful to the study and also fit to the expected length of the research study.
- Set research objectives.
  Draw a plan on the objectives of the research desire to explore.
  This should be aim oriented.
  It usually on the topics such as "examining", "investigating", "exploring" and "finding" etc.

Step by step process for setting up test criteria

Step 1: Research topic
Write down the topic of the study in black and white. By saying black and white it means that the topic should be very clear and precise to make easier to formulate a research question related to the study.

Step 2: Aims and objectives of the study
State the aims and objectives of the study and formulate main study hypothesis.

Step 3: Identification of the variables
Identify the independent and dependent variables through which the data have to be collected.

Step 4: Nature of the variables
Explore the nature of the variables.
For this:
Check whether qualitative or quantitative.
If qualitative check whether binary or categorical.
If categorical check further for nominal or ordinal.
If quantitative check whether discrete or continuous.

Step 5: Formulation of null hypothesis and identification of the test parameters.
In terms of selecting a statistical test, the most important question is "what is the main study hypothesis?"

In some cases, there is no hypothesis; the investigator just wants to "see what is there". For example, in an exploratory or descriptive research, a study hypothesis cannot be formulated and thus there is no null hypothesis to test. If there is no null hypothesis, then there is no statistical test.

On the other hand, in explanatory or cause and effect research a study hypothesis can be formulated.

Based on the main study hypothesis, formulate the null hypothesis which is going to be tested using an appropriate statistical test. Null hypothesis is usually tested against another hypothesis known as alternative hypothesis. Alternative hypothesis can be formulated after selecting a suitable statistical test.
Identify the parameters involve in the null hypothesis.

Step 6: Selection of a suitable statistical test.
Select a statistical test based on

- the number of variables,
- nature of the variables (continuous, binary, categorical) and
- the type of study design (paired or unpaired).

The statistical test should be selected before the study begins to ensure that the study results do not influence the test selection.

Step 7: Selection of a suitable test statistic.
Before selecting a statistical test, it is better to observe the followings:

- Number of independent variables that covary (vary in the same time period) with the dependent variable
- Level of measurement of the independent variables as well as dependent variables (for external validity).
- Whether the observations are independent or dependent and paired or unpaired.
- Whether the comparisons involve
    - populations to populations,
    - sample to population,
    - are two or more samples compared?
- Whether the hypothesis being tested comparative or relationship?

Choose a suitable test statistic for test parameter.
Identify the distribution of the test statistic and state the assumptions and conditions to be satisfied.

Step 8: Formulation of the alternative hypothesis
Assign a pre assumed fixed value for the test parameter under null hypothesis.

Make a decision whether to use two-tailed test or one-tailed test.

Use a two-tailed test when no particular direction of the estimate of the test statistic is assumed.

Use a one-tailed test only if there is clear evidence for the action of the intervention is only in one direction.

Formulate the alternative hypothesis as:

- H1: test parameter ≠ pre assigned fixed value if two-tailed is using,
- *H*1: test parameter > pre assigned fixed value
  or < pre assigned value if one-tailed is using,

Step 9: Setting the level of significance of the test
Level of significance is the value that represents likelihood of making Type I error, i.e., the criteria or threshold value based on which investigator can reject the null hypothesis. This is usually known as the p-value. The most common value set for p-value is 0,05.

In most sensitive cases like in medicine this is set to be 0.01.

Selection of an experimental design for data collection

Plan the experimental design based on the following key issues:

- Variables and how they are related,
- Hypotheses,
- Test statistics and, assumptions and conditions on it,
- How the design experimental treatments manipulate the independent variables,
- How the subjects are assigned to groups; within-subjects or between subjects,
- Plan for measuring the dependent variables.

First three key issues discussed earlier above.

How the design experimental treatments manipulate the independent variables

How the treatments manipulate the independent variables is very important for the external validity of the study.
Decide on how widely independent variables can vary and how finely it can be done.

Assigning subjects to treatment groups

Decide on study size, i.e., how many subject (individuals) should be involved in the experiment.

Distribute randomly these subjects to different treatment groups.
It can be done through one of the following experimental designs:

- Completely randomized design or randomized block design,
- Between-subjects design or within-subjects design.

Completely randomized design
In completely randomized design every subject is assigned to a treatment group at random.

Randomized block design

In a randomized block design (aka stratified random design) subjects are first grouped according to a characteristic they share and then randomly assigned to treatments within those groups.

Between-subjects design
This is also known as independent measures design or classic ANOVA design. In this type of design only one of the levels of a treatment is given to any one of the subjects.

Within-subjects design
This is also known as repeated measures design. In this type design every individual receives each of the experimental treatments consecutively, and their responses to each treatment are measured.

To avoid the order of treatment applications to influence the results of the experiment one can use the counterbalancing (randomizing or reversing the order of treatments among subjects).

Plan for measuring the dependent variables.
When planning, aim for valid and reliable measures of the dependent variable to minimize the research bias or error.
As experiments are always context-dependent, for a good experimental design, it is required that the data produced be accurate and relevant to the study.

Inferential data analysis
Data collection can be done using a suitable experimental design.
Collected data should be cleaned or scrubbed.
This can be done by

- removing major errors, duplicates, and outliers,
- removing unwanted data points,
- bringing structure to your data,
- filling in major gaps.

Then, carryout an exploratory analysis to identify initial trends and characteristics using descriptive data analysis techniques.

Compute the estimates of the relevant parameters.

Estimate the value of the test statistic.

Check the validity of the assumptions and the conditions of the proposed statistical test.

If these assumptions and conditions are satisfied perform the test and find the computed p-value and then compare this value with the level of significance to decide whether it is significant or non-significant.

If these assumptions and conditions are not satisfied then select an alternative test corresponding to the statistical test and then perform it to obtain the relevant results.
Make your inference based on the above results.

An Application of Statistical Techniques in Setting up Mortality Rates for the Valuation of Life Insurance Portfolios

H.A.M. Devindi Samaranayake, ASA,
Senior Actuarial Associate,
Milliman Limited,
Sri Lanka.

Background

Insurance companies in Sri Lanka should assess and report the solvency position quarterly as per the guidelines provided by Insurance Regulatory Commission of Sri Lanka (IRCSL). The current framework is Risk Based Capital (RBC) and accordingly, a company's liability for long-term business should be calculated using the Best Estimate Liability (BEL) which is based on best estimate assumptions. Paragraphs 28 and 29 of REGULATION OF INSURANCE INDUSTRY ACT, No. 43 OF 2000 is stated below.

28. *Every insurer shall use, in order to calculate the BE liabilities, appropriate actuarial and statistical techniques (such as analytical techniques, deterministic techniques, and simulation methods) to determine the mean of possible outcome, taking into account all relevant information about the insurer's business.*

29. *(1) Every insurer shall calculate the cash flows used to determine the BE liabilities on the basis of reasonable, supportable, and explicit BE assumptions to estimate the mean of possible outcomes.*
(2) BE assumptions shall:
   - *(a) be made using judgment and be based on experience;*
   - *(b) take into account relevant statistical and other information; and*
   - *(c) be neither overstated nor understated.*

(3) Every insurer shall use, subject to the provisions of paragraph (4), their own experience as the starting point in determining BE assumptions for future experience.
(4) If their experience is not sufficiently credible, such insurers shall use appropriate industry data, data from reinsurers, population statistics, or the assumptions used in a recent business planning exercise, to set suitable BE assumptions.

There are several key elements which have become quite significant under this process. Of all these assumptions, "Mortality Assumption" can be considered as one of the most important assumptions in life Insurance, if not the most important, that needs to be determined with extreme care. This is due to the nature of the life insurance business. The value of benefit payments and the timing of these payments are linked to a predefined risk event (typically death). Therefore, the value and the timing of the cashflows are uncertain and need to be estimated by using appropriate probabilities.

Developing a mortality rate table by using a company's own experience will be the best option that an Actuary would consider in setting up the best estimate mortality assumption. However, the experience of most of the insurance companies is insufficient to produce a company specific mortality rate table to adequate accuracy. They are often therefore compelled to use industry level data or standard mortality rate tables. Typically, industry mortality tables or standard mortality rate tables will be adjusted by a suitable multiplier to reflect recent experience of the company. The credibility of company experience should play a vital role in determining this multiplier.

The practical approach

When the Actuarial Department carries out a mortality experience investigation on a policy count basis, the output will be the crude death rate for age x which is determined by the ratio of "actual number of deaths at age x" to "count exposed to death risk". The calculation procedure of the "count exposed to death risk" is kept out of this article but can be thought of as number of lives which have been exposed to the risk of death of policyholders aged x. The crude death rates may have jumps and other anomalies which need to be smoothed in order to create mortality rates that are a smooth function of age.

The process of using statistical/actuarial techniques to construct a smooth set of mortality rates by refining the underlying estimated crude rates is called "Graduation". The aims of graduation are to produce a smooth set of rates, to remove random sampling errors and to use the information available from adjacent ages to improve the reliability of the estimates. There are various standard actuarial methods used for graduation such as graduation by parametric formula, graphical graduation, and graduation with reference to a standard table.

This article illustrates an instance of graduation with reference to a standard table.

Data Used

1. Crude death rates were calculated by using the number of deaths and exposed to risk counts based on Munich Re mortality study for Male lives.

   The mortality study conducted in 2013 by Munich Re, on Sri Lankan Assured Lives is the only extensive, industry-wide analysis that had been performed in the recent past. According to the "Sri Lanka Assured Lives Mortality Study Report - 2013" eight of the leading life insurance companies in Sri Lanka, covering over 95% of the life insurance market by written premium in year 2011, participated in this study.

2. UK Standard mortality rate table A67-70

Methodology

Step 1 – Select a standard table.

Ideally the class of lives involved in the graduation needs to be sufficiently similar to the class of lives whose experience forms the basis of the standard table. Even if overall levels of mortality differ between the two, it would still be desirable that the overall progression of rates from age to age would be similar.

Step 2 – Find a simple link to the standard table.

Let q_x^s be the rate of mortality of the selected standard table. Then we try to exploit the assumed similarity of the experiences by seeking a reasonably simple function f() such that q_x^0 = f(q_x^s). The search for suitable function f() can be aided by making some simple plots.

Step 3 – Determine parameter values.

Once a possible relationship has been identified, estimate model parameters by using suitable parameter estimation techniques such as maximum likelihood estimation, least squares method, etc.

Step 4 – Calculate graduated rates.

The assumed relationship with the standard table is then used to calculate the graduated rates.

Step 5 – Test

The resulting graduation would be subjected to tests of goodness-of-fit to the data.

Analysis

The UK Standard mortality rate table A67-70 has been selected for this illustration in the absence of any relevant standard mortality table for Sri Lanka insured lives. However, the A67-70 mortality rate table is one of the commonly used tables by Sri Lankan Insurance Industry for their actuarial calculations.

The scatter plot of crude death rates of male lives against standard mortality rate A67-70 indicated a positive linear relationship and the correlation coefficient was found as 0.974.

The simple link to the standard table used in this instance is:

Mortality (Graduation) = a + b* Standard mortality rate

Parameters were estimated by using least square method and the model is:

Mortality (Graduation) = 0.001 + 0.535* Standard mortality rate

The Chi-square test was performed to assess the overall goodness of fit of graduated rates and the test passed at 1% level of significance.

Further Explanation

The crude death rates, UK standard mortality rates (A67-70) and the resulting mortality rates after graduation is plotted for integer ages 18-60 in the above graph. The graph depicts the minor deviations of standard mortality rate A67-70 and crude death rates for the age band 18-42, but higher deviations for ages greater than 42. A sudden increase of crude death rate is visible at age 22, which is common characteristic in mortality tables called the "accident hump". This is due to the sudden increase in mortality rates for young individuals once they start to drive motorcars or ride motorbikes or start drinking alcohol.

The dotted line in the above plot is the crude death rates determined using industry experience, which is more relevant, but the irregularities are hard to justify in practice. The dashed line in the above plot is the UK standard mortality rates which consist of essential characteristics of mortality rate progression but do not carry any information regarding industry experience. Finally, by using a link between standard death rates and crude death rates, the graduated rates (the line in the above plot), is determined which has both characteristic smoothness and adherence to industry experience. These rates can be used as the best estimate mortality assumption in valuation of liabilities for Sri Lanka Life insurance company portfolios.

HOW TO BECOME A MEMBER OF IASSL?

Please visit

http://www.iassl.lk

Apply for the Membership

Application form

Unraveling the Enigmatic: Exploring the Will Rogers Paradox

Ms. Ishara Wijayaratne,
Fourth year student, BSc (honours in Statistics),
Department of Statistics, University of Sri Jayewardenepura,
Sri Lanka.

Abstract

This article provides a concise overview of the Will Rogers paradox, highlighting its real-world applications across different fields.

Keywords: Will Rogers paradox, applications, fields.

1. Introduction

In the realm of human understanding, the mysteries and contradictions of life often leave us astounded and intrigued. One such enigma that has captivated the minds of scholars and philosophers for centuries is a paradox—a concept that defies conventional wisdom and challenges our perception of reality.

What is a paradox, you might wonder? At its core, a paradox is a concept or situation that appears to defy logic or common sense, presenting us with contradictory ideas or outcomes. It challenges our fundamental beliefs and assumptions, coaxing us to re-evaluate our understanding of the world.

As we delve further into the paradoxical landscape, we encounter an intriguing paradox known as the Will Rogers Paradox. Named after the beloved American humorist, actor, and social commentator Will Rogers, this paradox weaves together seemingly incompatible principles, leaving us pondering its implications. With the stage now set, the audience primed, and the paradoxes unveiled, let us embark on a journey of discovery, where the unexpected becomes the norm and where contradictions hold the keys to profound wisdom. Join me as we unravel the enigmatic and delve into the realm of the Will Rogers Paradox, where truths collide, and mysteries arise.

2. The Discovery by Will Rogers

In the midst of the Great Depression in 1930, when the nation was grappling with economic turmoil, a famous comedian of the time, Will Rogers, uttered a seemingly harmless joke about migration. Little did he know that his playful remark would later become the centerpiece of a paradox that would captivate the minds of mathematicians and statisticians alike.

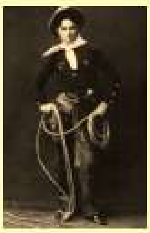


Figure 1: Will Rogers

Roger's witty remark went as follows: "When the Okies left Oklahoma and moved to California, they raised the average intelligence in both states". At first glance, it appears to be a humorous observation on the intelligence levels of the two groups. However, beneath the surface of this jest lies a profound mathematical insight that would later be known as the Will Rogers Paradox or the Okie Paradox.
What Rogers was implying, in his own unique comedic style, was that in his view, even the least intelligent Okie surpasses the intelligence of the most intelligent Californian.

Let's develop an intuitive understanding of this phenomenon by analyzing the representation in Figure 2. The pivotal factor in play here is the relative positioning of the data points. When a data point is re-classified from one group to another, a critical scenario emerges: if it positions itself below the departing group's average but above the receiving group's average, the outcome will be an elevation in the averages of both groups. This occurs because the reclassified data point effectively boosts the average of the group it is joining, while simultaneously lifting the average of the group it is departing from.

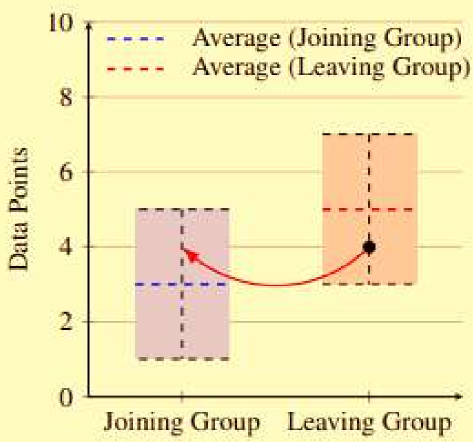


Figure 2: Re-classification of Data Points and Will Rogers Paradox

In the case of the Okies migrating from Oklahoma to California, some individuals from Oklahoma with intelligence levels (measured using some standardized method) below the state's average moved to California, a state with lower average intelligence. By doing so, the average intelligence of California increased, as the new arrivals raised the collective intelligence of the state. Simultaneously, as the less intelligent individuals left Oklahoma, the average intelligence of Oklahoma will also increase.

3. Relevance of Will Rogers Paradox in Different Fields

3.1. Medical Field

In the field of Oncology, the Will Rogers Phenomenon coined by Alvan Feinstein in 1985, reveals an intriguing observation about cancer patients. It highlights how changes in the criteria for classifying patients into different disease stages can result in misleading improvements in prognosis, even though individual outcomes remain unchanged.

Let's further explore this phenomenon with the assistance of Figure 3. Which considers a scenario involving two groups: individuals with cancer and those without. The left plot of Figure 3 portrays the initial mortality rates of these two groups, using blue and red points to represent the individuals in each group, respectively. It's important to take into account that cancer patients generally have a lower life expectancy compared to healthy individuals.

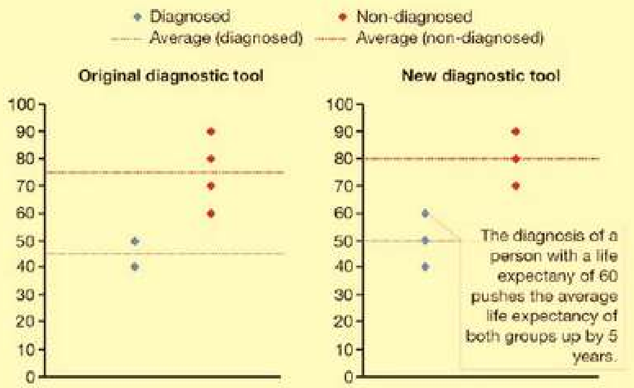


Figure 3: Cancer detection and Will Rogers Paradox

Now, imagine a newly developed diagnostic tool that has the capability to detect even the most minute traces of cancer, including less severe cases that were previously unidentifiable.

The introduction of this new diagnostic tool could lead to the re-classification of individuals initially considered cancer-free by the original tool. Now, they might fall into the group with cancer due to the detection of less severe cases that went unnoticed before.

The scenario, where the re-classified individuals have lower life expectancy than the average life expectancy of the group without cancer, yet a higher life expectancy than the average life expectancy of the group with cancer, leads to the emergence of the Will Rogers Paradox. Consequently, the reclassification of these individuals leads to an increase in the average life expectancy of the group diagnosed with cancer, simultaneously contributing to an increase in the average life expectancy of the group without cancer.

The manifestation of the Will Rogers Paradox in this scenario becomes apparent when comparing the left and right plots in Figure 3, which effectively explain the phenomenon through the use of numerical representations.

It's important to note that the underlying disease remains the same, and individual life expectancies are unaffected by the choice of diagnostic tool. The only change is our improved ability to accurately detect cancer, leading to the re-classification of previously undetected cases. Therefore, any apparent improvement in average life expectancy is solely due to the enhanced diagnostic tool and not due to the improvements in the health status of the individuals [1].

3.2. Sports Field

The Will Rogers paradox, can also be observed in professional sports such as basketball. It was detected when examining the case of college basketball player Alex Barcello in the United States.
Starting his collegiate career at the University of Arizona, Barcello experienced a mix of success during his freshman and sophomore seasons. However, in search of more playing time and improved team chemistry, he made the decision to transfer to Brigham Young University. Surprisingly, Barcello's move resulted in statistical improvements for both teams involved, which goes against the expectation that a player's transfer would benefit one team at the expense of the other. By transferring to Brigham Young University, Barcello transitioned from being an under-performer at Arizona to an over-performer at BYU. In his second season at the University of Arizona, he was ranked as the tenth-best player out of fourteen. However, at BYU, he became the fourth-best player, thus raising the average performance level for both teams, resulting in a win-win situation.

An analysis of Field Goal % reveals a noticeable improvement for both teams. Arizona's Field Goal % increased from 42.7% in Barcello's final season (2018-19) to 44.8% in the season following his departure (2019-20), as indicated in Table 1. Similarly, BYU's Field Goal % rose from 46.8% before Barcello joined the team (2018-19) to 50.4% after his arrival (2019-20), as shown in Table 2.

Table 1: Field Goal % Comparison at the University of Arizona

	UA 2018-19 season with Barcello	UA 2019-20 season without Barcello
Field Goal %	42.70%	44.80%

Table 2: Field Goal % Comparison at BYU

	BYU 2018-19 season without Barcello	BYU 2019-20 season with Barcello
Field Goal %	46.80%	50.40%

Table 3 further demonstrates Barcello's impact. During his sophomore season at Arizona, his Field Goal % was below the team average (Arizona 42.7% >Barcello 39.3%). However, in his junior season at BYU, his Field Goal % closely matched the team average (BYU 50.4% ≈ Barcello 49.3%).

	Barcello's sophomore season at Arizona (2018-19)	Barcello's junior season at BYU (2019-20)
Field Goal %	39.30%	49.30%

Table 3: Field Goal % Comparison for Barcello.

These findings exemplify the Will Rogers paradox and highlight the positive influence of Barcello's transfer on both teams. However, it is important to note that although the teams' performance improved after the transfer, it does not necessarily indicate a personal enhancement in Alex Barcello's basketball skills [2].

4. Summary and Conclusions

The Will Rogers paradox, originating from a humorous remark by the comedian Will Rogers, reveals the following fascinating phenomenon: when a data point is reclassified from one group to another and it falls below the leaving group's average but above the receiving group's average, with the average of the leaving group being generally greater than the receiving group, both groups ultimately experience an increase in averages.
This paradox has implications in various fields. In the medical field, it is demonstrated by a scenario, where changes in disease stage classification can lead to misleading improvements in prognosis without actual changes in individual outcomes. Similarly, in the sports field, the paradox is evident through the case of basketball player Alex Barcello, whose transfer improved the statistics of both the original and receiving teams. This challenges the expectation of a one-sided benefit. This paradox emphasizes the complexity of data analysis and highlights the importance of considering group dynamics. It warns us to be cautious when interpreting improvements resulting from changes in classifications or diagnostic statuses, as they may suffer from the Will Rogers paradox.

Acknowledgements

I would like to express my sincere gratitude to Dr. Rajitha M. Silva for granting me this invaluable opportunity to explore into the captivating world of paradoxes.

References

[1] Galak, J. (2021). Will Rogers Paradox Explained (YouTube Video). Retrieved from : Will Rogers Paradox Explained.

[2] Valerdi, R. (2021). Will Rogers Paradox in Basketball Analytics. Iris publishers. Retrieved from : Will Rogers Paradox in Basketball Analytics .

Summary of events and activities conducted by IASSL during May - August, 2023

Awareness sessions on Statistics and Probability for A/L Combined Mathematics school children

As the professional body of Statisticians in Sri Lanka, IASSL has well understood the importance of disseminating Statistics knowledge to A/L students who will sit for the exam in the near future as well as fostering statistical thinking among them.

During May - August, 2023 several seminars were conducted by the IASSL with University Senior Lecturers as resource persons with these aims.

The first seminar for A/L combined mathematics students was conducted at the Ramanathan Girls College on 22nd May from 8.30 am to 1.30 pm with the participation of over 100 students from Hindu College, Ramanathan College, Muslim Boys and Girls Colleges, Colombo. Mr. Dias and Prof. Sivananthawerl served as resource persons and the seminar was conducted in English and Tamil languages. Mr. K Manimarrphan, Principal of the Hindu College provided a great support to initiate and organize this first seminar of the year.

Another two school seminars were conducted on the 17th and 18th August, 2023 at St. Joseph's Girls' College, Gampola and Ginigathhena Central College in Ginigathhena respectively with the participation of over 70 students each day. Students in the Gampola and Ginigathhena education zones participated in the seminars. Mr. P. Dias and Dr. D.C. Wickramarachchi served as resource persons. Both the students and teachers appreciated the commitment, contribution and support extended by the IASSL to empower their knowledge on Statistics.

IASSL would like to thank the Chair and the sub-committee members of IASSL Statistics Popularization Committee for their immense contribution in planning and organizing these sessions.

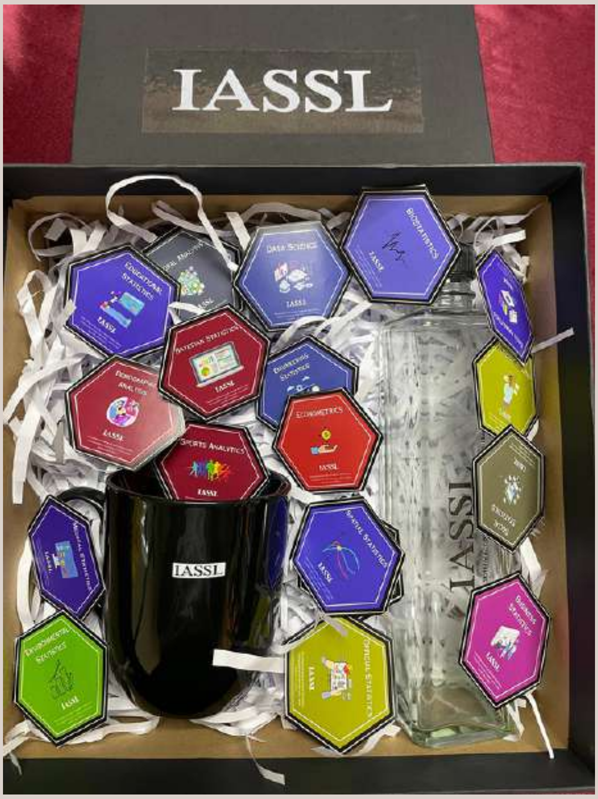
Commemorating the achievements of the Senior Category Winner of Statistics Olympiad Competition Organized by C.R. Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

The National Statistics Olympiad is a highly anticipated event that allows students to showcase their knowledge in Statistics and compete with talented minds from different parts of the country.

The 2022 national competition organized by the IASSL consisted of two categories of competitors: Junior and Senior. It was held on the 18th of December, 2022. The top three scorers in each category along with the merit pass holders from this national level competition then took part in the 13th Statistics Olympiad - 2023 organized by the C. R. Rao Advanced Institute of Mathematics, Statistics, and Computer Science (AIMSCS), India which was held on the 29th of January, 2023.

Ms. S. A. Sukhithi Chamali, a first year student from the University of Colombo, emerged as the First Place winner in the Senior category of the prestigious Statistics Olympiad 2023 competition organized by AIMSCS. Ms. Chamali's outstanding performance has made the entire nation proud and highlights the exceptional talent of Sri Lankan students.

To commemorate her achievements, IASSL organized an event on the 27th of July, 2023 at their premises. At the event IASSL gifted the Olympiad winner a plaque, a cash prize and some attractive IASSL souvenirs.



Fostering Statistical Thinking among School Children at SCIMATICS 23

Together with the young statisticians of the Statistics Society, University of Sri Jayewardenepura IASSL successfully conducted a session at the SCIMATICS 23 Science Day Exhibition hosted by Musaeus College, Colombo, with the intention of igniting statistical thinking in Young Minds.

Establishing collaborations with leading Sri Lankan Financial Institutions

Recently IASSL initiated a discussion with the National Development Bank (NDB), a leading financial institution in Sri Lanka with the aim of providing various tailormade training sessions in Statistics and Data Science to the employees of NDB facilitating NDB's long-term goal of building data literacy among their employees.
If your institution wish to collaborate with us to fulfill your statistical needs please get in touch with us.

Upcoming events and activities

Get ready for the International Statistics Conference 2024 (ISC 2024)!!

International Statistics Conference 2024 (ISC 2024), jointly organized by the IASSL and the Department of Statistics, University of Sri Jayewardenepura (USJ), Sri Lanka will be held on 28 - 29 December, 2024 at the Mt. Lavinia Hotel, Colombo, Sri Lanka. The Simon Fraser University (SFU), Canada will be the International collaborator of the Conference.
The theme of the conference is "Unleashing the power of data: Harnessing the Synergy between Statistics and Data Science". Keynote speech of the conference will be delivered by the renowned Statistician Professor Tim Swartz, SFU, Canada. The president of the American Statistical Association will be the chief guest of the inauguration.

A Pre-conference workshop will also be held on 27 December, 2024 at USJ premises. It will be conducted by Professor Tim Swartz. The theme will be related to sports analytics.

Important dates to keep in mind:

Calling for papers: 15th February, 2024

Paper submission deadline: 15th May, 2024

Notification sent to author: by 15th July, 2024

Submission of final camera-ready copy: by 15th of August, 2024

Early-bird registration closes: 31st August, 2024

Registration deadline: by 30th November, 2024

Tracks:
Probability and statistical inference, Statistical modelling, Time series and forecasting, Sampling and surveys, Experimental designs and analysis, Spatial and longitudinal data analysis, Actuarial statistics, Data mining and Big data analytics, Machine Learning and AI, Data visualization & predictive analytics, Econometric modelling, Sports analytics, Biostatistics, bioinformatics and healthcare analytics, Developments in statistical software, Ecological modelling, Financial and marketing data analytics, Manufacturing, telecommunication, transport and energy data analytics, Education and social science data analytics, Agriculture, food and environment data analytics, Operations research applications, and other relevant areas.

For more details stay tuned with www.isc24.iassl.lk

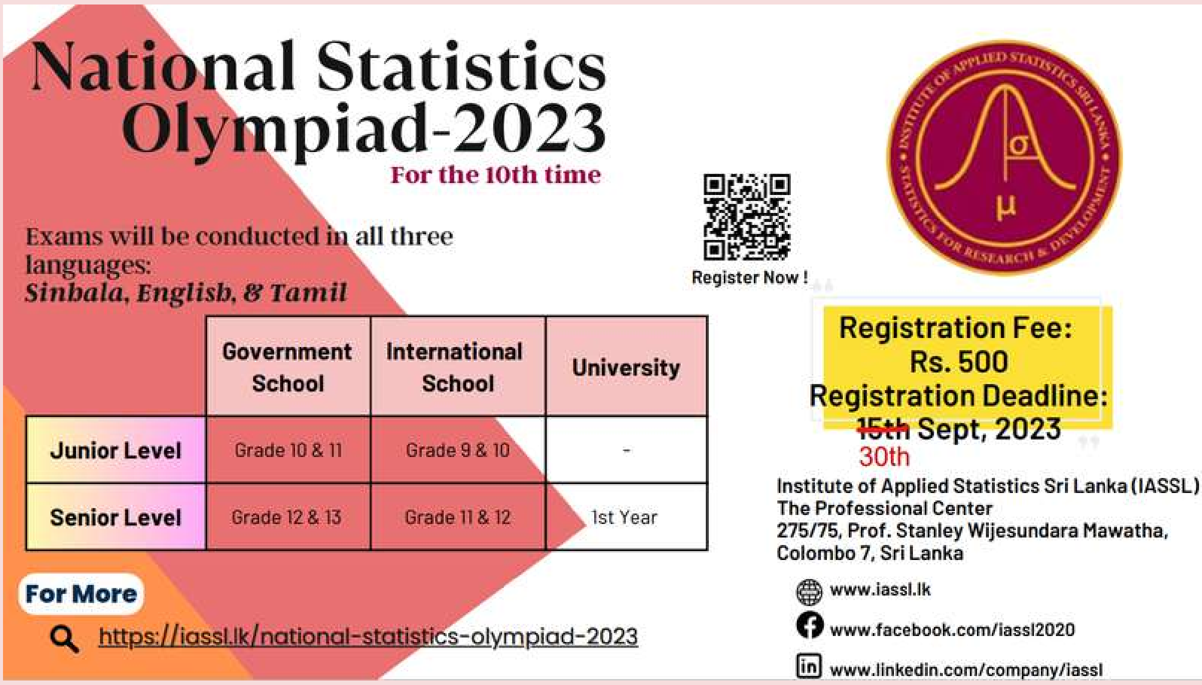
Co-chairs
Prof. Chandima Tilakaratne, University of Colombo.
Dr. Niroshan Withanage, University of Sri Jayewardenepura.

Co-secretaries
Dr. Chathuri Jayasinghe, University of Sri Jayewardenepura.
Dr. Neluka Devpura, University of Sri Jayewardenepura.

National Statistics Olympiad 2023

For the 10th time the National Statistics Olympiad competition will be organized by the IASSL later this year. We have already begun registering students. The closing date of applications will be the 30th of September, 2023. As usual it will be held in two categories: Junior and Senior. The exam will be conducted in all three mediums: Sinhala, Tamil and English.

For further information please visit the IASSL website iassl.lk.



Up coming Courses on offer by the IASSL

The following courses will be on offer by the IASSL during September - December, 2023.

- Systematic Literature Review (SLR) with Bibliometric Analysis: a way of manuscript writing with PRISMA-6th Batch (In October)
- Business Analytics with Power BI (In October)
- Data Analysis with Python (In November)
- Basic Statistics for Managers & Researchers (In November)
- Business Analysis with Excel
- Survey Design and Analysis
- Structural Equation Modelling (SEM) with AMOS/ R
- Multivariate Data Analysis with R
- Repeated Measure Data Analysis with R and SAS Studio (SAS OnDemand for Academics)
- Academic Writing with LaTeX
- Artificial Neural Networks with R

Welcome to the IASSL family!

During May - August 2023 the following statisticians became life members of IASSL:

- Dr. Neluka Devpura
- Dr. Dushan Kumarathunga
- Mrs. K.A. N. L. Kuruppuarachchi
- Mrs. A. Sanjaya G. Jayasinghe

Call for papers SLJAS Volume 25

SLJAS is a two star, open-access, international, double-blind peer-reviewed journal published by the Institute of Applied Statistics, Sri Lanka (IASSL). The main purpose of the journal is to publish 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 Sciences, Physical Sciences, Social Sciences, Economics and Actuarial Science fall within the scope of the journal. The SLJAS publishes three issues annually.
Authors are warmly invited to submit their Original Research Articles, Review Articles, Methodology Articles, Case Studies for publication in SLJAS Issues of volume 25.

We also consider publishing of full conference papers as a Special Issue. Please inquire by sending an email to editor@iassl.lk.

AUTHOR BENEFITS:

- No Submission, Processing and Publication Charges.
- Fast Track Publication.
- Provides Individual DOI (Digital Object Identifier) to each Published Article.

Published Papers are Indexed by: DOAJ, Google Scholar, Sri Lanka Journals Online.

Submit your manuscript via JMS (Journal Management System). We accept manuscripts submitted only via JMS and do not accept submissions via email.

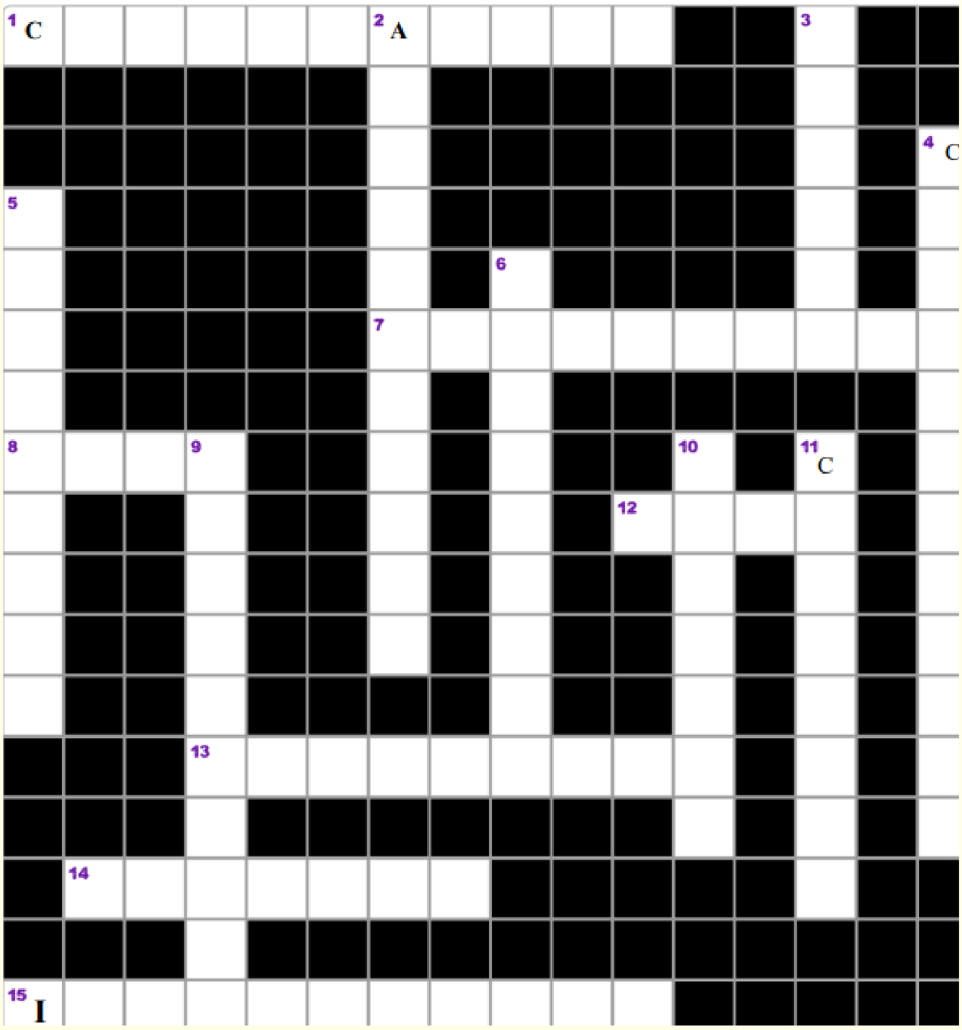
Author guidelines:

Authors are encouraged to use latex when preparing the manuscript for submission. Please visit https://sljastats.sljol.info/about/submissions for further details.

Please send your resume to become a Reviewer of SLJAS.

We look forward to receiving your valuable research article. We also request you to encourage your colleagues and students to contribute articles to SLJAS. Feel free to reach us for any query via editor@iassl.lk.

Sudoku Puzzle Competition



Please email your submission to appstatsl@gmail.com on or before 15th of December 2023. The draw will be held on the 28th of December, 2023. 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 2023 IASSL newsletter.

Thank you WIRETREE for sponsoring Issue 2 2023 puzzle competition!

Issue 1 2023 Puzzle Competition winners:
1st Place:
Dilki Kaushalya
2nd Place:
Tharuni Kavishka
3rd Place:
Taniya Fernando

DOWN

2 AU(BUC) = (AUB)UC. This law is called as ....

3 Any subgroup or subset of a population is called as:

4 This theorem says that the computed values of the average will be distributed according to a normal distribution

5 A numerical term that summarizes or describes a population.

6 Response variable is also called as .... variable

9 $\bigcup_{i=1}^{n} B_i = \Omega$; this is called as:

10 If Pearson correlation coefficient is equal to +1 or -1, then it is a.......... relationship.

11 An event is said to be a .... event, if it consists of two or more simple events.

ACROSS

1 The association exhibited between x and y coordinates in generally referred to as .... between the variables x and y.

7 For any event A, AUA = A and A∩A = A. The law is called as...................... law.

8 A measure of central tendency.

12 If C is a constant in a continuous probability distribution, then P(X=C) is always equal to:

13 If mean, median & mode of the distribution are the same then that distribution is ....

14 The distribution that mean and variance is always same

15 In Bernoulli model, the trials are ....

From the sponsor of the Issue 2 2023 Sudoku Puzzle Competition

CONTRIBUTIONS TO THE SEP-DEC (ISSUE 3) 2023 NEWSLETTER:

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

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

EDITORIAL BOARD/IASSL

IASSL NEWSLETTER

Official Newsletter of the Institute of Applied Statistics Sri Lanka

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

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