



## JMP Workshop on Data Handling, Analysis, and Visualization

11th-12th July 2024

Dr. Qurat-Ul-Ain Khaliq

Country Coordinator, Pakistan

### 1. INTRODUCTION

**Dr. Quratulain ,Khaliq** has expertly organized and designed the content for the **JMP Workshop on Data Handling, Analysis, and Visualization on July 11–12, 2024**. This workshop was online, free registration and certification. This initiative not only prepares students for upcoming poster competitions but also imparts essential knowledge and skills. I found JMP software is highly appealing and believe it's crucial to teach it to others. Our efforts have successfully educated participants not only in Pakistan but also in India, showcasing the workshop's extensive reach and impact.

It successfully gathered professionals, PhD's, university and college students, and enthusiasts from various fields to enhance their skills in data science and statistical analysis. We organized this workshop because JMP is still relatively new in the statistician community, and many individuals struggle with creating technical posters. A poster is an excellent tool for sharing research, allowing one to convey maximum information in a short time. To address this, we concluded the workshop with a poster competition. We received numerous submissions, not only from Pakistan but also from India. Many students used JMP to design their posters and present their ideas in an innovative way. The workshop also covered data handling, cleaning, and engineering, comparing methods in Python and JMP. Several professors joined us to share their research. The high-class posters submitted by your department were truly impressive, showcasing both rigorous academic work and a deep understanding of real-life social issues.

This workshop's success is largely due to ISLP's tremendous support. The honorable **Reija Helenius, ISLP Director**, provided us with immense support. This workshop is not only a success for us, but also for ISLP. We not only reached our target audience but also conducted a pilot competition. Most of our participants were unfamiliar with the JMP and poster design. During this workshop, they learned how to use JMP and design their own posters. Many participants have requested additional workshops to further equip statisticians with statistical tools and promote statistical literacy.

**Dr. Volker Kraft** conducted a two-hour session where he not only explained JMP features but also covered advanced topics like handling real-life data, data analysis, and graphical displays. This session was incredibly beneficial, providing us with numerous posters on JMP. Additionally, **Dr. Quratulain Khaliq** led six sessions, guiding through data engineering, cleaning, analysis, and graphical display. She also conducted a special session on poster design. Participants really enjoyed these sessions.

**Professor Dr. S. Gandhiya Vendhan**, Country Coordinator of India from Bharathiar University, Coimbatore, showed active participation and support in our recent JMP Workshop. His efforts were instrumental to the event's success. We received the registration form the day before the workshop. They prepared their students, who actively participated despite short notice.

## **2. PARTICIPANTS**

Our audience consisted of many undergraduate students, with over 90% of participants coming from the statistics community. We compared posters from Pakistan and India at the end and selected the top three winners. The long duration of the workshop did not deter attendees, and the feedback form reflected overwhelmingly positive responses.

We are thrilled to have participants from a diverse and prestigious array of institutions around the globe. We are also honored to include: Afghanistan National Public Health Institute, Kabul, Afghanistan; Dominion University College, Ghana; Shanghai Jiao Tong University, China; Sapienza Università di Roma, Italy; Tecnica da Comunicacao do Instituto Nacional de Estatística, Portugal; Universida de Aberta (UAb), Portugal; Universidad de Panamá; University of Bologna, Italy; Masaryk University, Brno, Czech Republic; Bharathiar University, Coimbatore, India; Masaryk National University of Lesotho, add to the global representation.

Further, we have participants from Bahauddin Zakariya University, Multan; CMH Institute of Medical Sciences, Multan; COMSATS University; Federal Urdu University of Arts, Science & Technology; Government Post Graduate College Baghbanpura, Lahore; Govt. Graduate College GT Road, Jhelum, Gomal University, Dera Ismail Khan; GSCWUH; Haqqania School, Higher Education Department, Government of the Punjab; Karachi University, Sindh, Lahore College for women University; NCBA&E, Lahore; Nishtar Hospital; NUST, Islamabad, Pakistan; Presidential Initiative for Artificial Intelligence & Computing; PMAS Arid Agriculture University; Quaid-I-Azam University, Islamabad; Rawalpindi Women University; Riphah International University, Lahore; University of Sargodha; Aga Khan University; The Women University Multan; University Management Technology; University of Baluchistan; University of Gujrat; University of Karachi; University of Punjab; University of South Asia; University of Science & Technology Bannu KPK, Pakistan; Viqar un Nisa Women's University are part of our diverse group. This wide range of participants reflects the global interest and commitment to advancing knowledge and collaboration in our field.

## **3. WORKSHOP OBJECTIVES**

The purpose of this workshop was to highlight the importance of creating posters and prepare the audience for an upcoming poster competition. Additionally, it aimed to teach the audience how to use the JMP software, which is completely new to our community. We wanted to familiarize everyone with the features of the JMP software. None of our audience members knew how to professionally create posters or had any idea about the JMP software.

We also organized a poster competition within the workshop to assess how much benefit the participants gained from the session and to gauge the success of the workshop. Initially, we created three categories to give every educational level a chance to showcase their outcomes. We received numerous submissions from India and Pakistan, with many participants using JMP software to design their posters and share innovative ideas.

To motivate our participants, we selected the top posters as winners, including national and international winners. This approach was intended to promote online learning, ensuring that attendees not only listened to the workshop but also contributed by applying what they learned. The workshop's primary objectives were as follows:

1. The aim is to acquaint the participants with the capabilities of JMP software.
2. The aim is to demonstrate practical techniques for data handling and preparation.
3. We aim to offer practical experience in data analysis using JMP.
4. To teach participants how to create meaningful visualizations to communicate findings effectively.
5. Prepare the audience for the upcoming poster competition of the International Statistical Literacy Project (ISLP).

## **4. IMPORTANCE OF RESEARCH POSTER**

Research posters play a crucial role in the academic and scientific community. Here are some key points highlighting their importance:

- ❑ **Visual Communication:** Research posters enable the visual presentation of complex ideas and data, facilitating a brief understanding of the research by the audience. This is particularly useful in conferences where attendees have limited time.
- ❑ **Engagement:** Posters facilitate direct interaction between researchers and their audience. This engagement can lead to valuable feedback, new ideas, and potential collaborations.
- ❑ **Accessibility:** Posters can make research findings accessible to a wider audience, including those who might not have the time or expertise to read full research papers. This can help disseminate knowledge more broadly.
- ❑ **Conciseness:** Creating a poster requires researchers to distil their work into its most essential components, which can help clarify their own understanding and improve their ability to communicate their findings succinctly.
- ❑ **Networking:** Presenting a poster at a conference or workshop provides an opportunity for researchers to network with peers, experts, and potential collaborators, enhancing their professional development.
- ❑ **Recognition:** A well-designed and impactful poster can increase the visibility of a researcher's work, leading to recognition within their field. This can be particularly beneficial for early-career researchers seeking to establish themselves.
- ❑ **Learning and Teaching Tools:** Posters can serve as educational tools, helping students and early-career researchers learn how to effectively present their work. They also promote skills in design, summarization, and public speaking.
- ❑ **Showcasing Innovation:** Posters serve as a platform to showcase innovative methodologies, tools, and findings, encouraging the exchange of new ideas and advancements in the field.

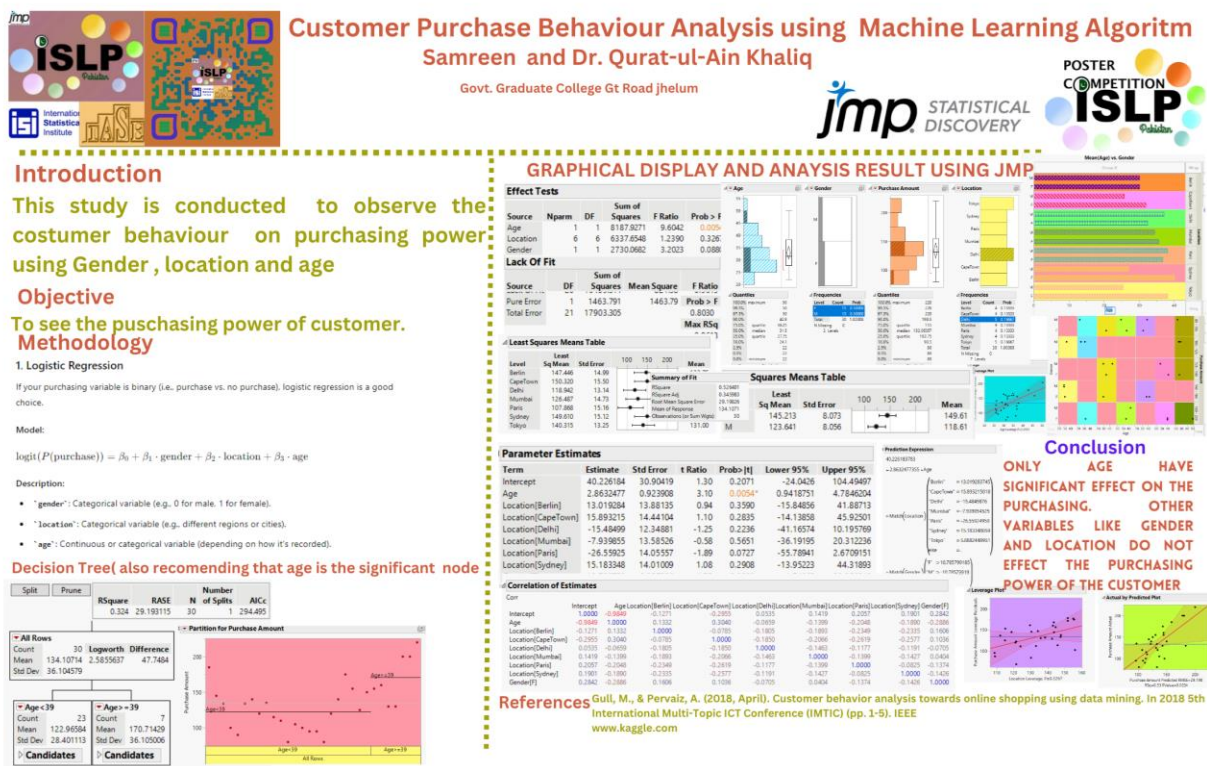


Figure 1: Poster Designed by Samreen and Dr. Quratulain Khaliq Showing the Customer Purchasing Behavior Analysis applying Machine Learning Algorithm using JMP Software.



Position	Pakistan			
	Category Wise Poster Winner			National Winners
	I	II	III	
1st	Samreen	Roman Zainab	Shahla Gul	Shahla GUI
2nd	Maryam Tahir	Maria Malik	Ismat Parveen	Samreen
3rd	Nimra Ali and Ammara	Haseeba and Syed Shafqat Hussain Gardezi	Sidra Gul	Roman Zainab
Position	India			
	Category Wise Poster Winners			National Winners
	I	II	III	
1st	Suryaprahasan. R	Swetha. M	Gayathri R	Gayathri R
2nd	C.Aravinth	Suryaprakash G	Sarathkumar S	Suryaprahasan .R
3rd	Sakthiswar. K	Anu K	Nandhini devi S	Swetha. M & C.Aravinth
International Winners				
1st	Shahla Gul			
2nd	Gayathri R			
3rd	Suryaprahasan .R and Samreen			

**Table-1: JMP Workshop Poster Winner's Country Wise Detail (Pakistani Team was Supervised by Dr. Quratulain Khaliq while Indian Team was Supervised by Dr. S. Gandhiya Vendhan).**

## 5.SPECIAL GUEST

Prof. Dr. Muhammad Asadi's presence as our special guest was an honor. Prof. Dr. Asadi has been a significant source of support, inspiration, and motivation for the organizing team and participants alike. His expertise and insights added immense value to the workshop, inspiring participants to leverage data science for innovative solutions.



**Figure 2: Details of All Sessions Organized in the JMP Workshop**



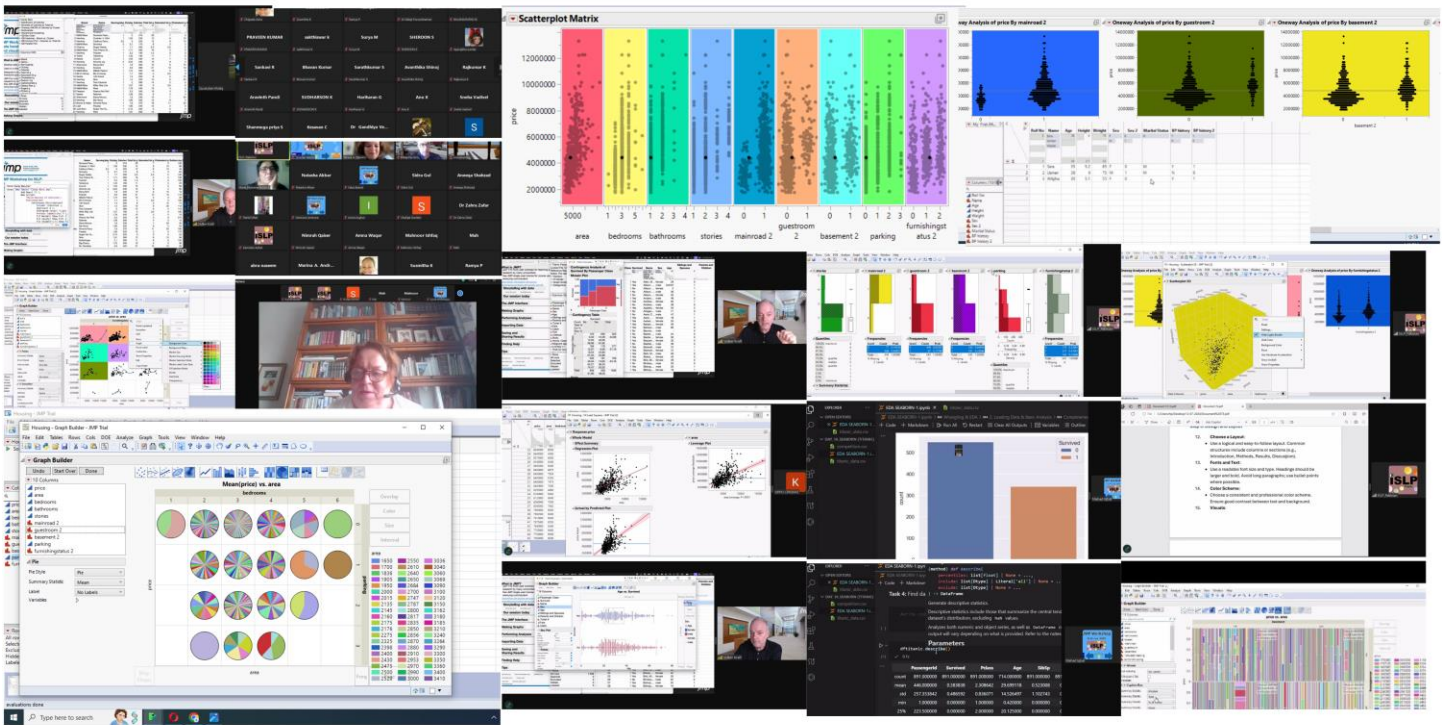


Figure 3: Honorable Reija Helenius, ISLP Director's Speech, JMP Training by Volker Kraft, Dr. Qurat-ul-Ain Khaliq, and Mahad Iqbal Online Sessions Overview.

### “Analysis of Protein Consumption Across Various Food Categories”

**1. Introduction**  
This dataset contains the consumption levels of various food categories (Red Meat, White Meat, Eggs, Milk, Fish, Cereals, Starch, Nuts, Fruits & Vegetables) for 25 different countries. The values represent the average daily intake in grams per person. This data can be used to analyze dietary patterns and nutritional habits.

**2. Objectives**  
To determine the correlation between different protein sources.  
To identify key patterns and relationships in protein consumption using Principal Component Analysis (PCA).

**3. Methodology**  
Correlation Analysis: Pearson correlation coefficients between all pairs of protein variables (Red Meat, White Meat, Eggs, Milk, Fish, Cereals, Starch, Nuts, Fruits & Vegetables).  
Principal Component Analysis (PCA): Performed PCA to reduce dimensionality and identify patterns in the dataset. PCA transforms the original variables into a set of linearly uncorrelated components,  $PC_1, PC_2, \dots, PC_n$ , that capture the variance in the data.  
1. Covariance matrix:  $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_j - \bar{X})$   
2. Eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  and corresponding eigenvectors  $V_1, V_2, \dots, V_n$   
3. The  $PC$ 's are given by  $PC_j = X \cdot V_j$ ,  $X$  is the centered data matrix.  
4. The proportion of variance explained by each principal component  $PC_j$  is  $\lambda_j / \sum \lambda_i$ .

**4. Results and Discussion**  
**5.1 Dietary Correlation Patterns**  
Positive Correlations: Red-meat correlates positively with Eggs, Milk, and to a lesser extent with Starch. Negative Correlations: Red-meat shows negative correlations with Cereals and Nuts. Weak Correlations: There are several weak correlations observed, such as between Red-meat and Fish, and Red-meat and Starch.  
**5.2 Major Findings from PCA**  
PC1 explains 44.5% of the variance, and PC2 explains 18.2%, totaling 62.7% of the variance. Red Meat (0.30281) and White Meat (0.31056) have relatively high loadings on PC1 (0.644852) and Fruits & Vegetables (0.31619) have the highest loadings on PC2.

**6. Reference**  
Data set from this website: <https://www.kaggle.com/>

Roman Zainabi Submitted to Dr. Quratulain Khaliq ISLP

### PRICE EVALUATION OF CERTAIN HOUSING SCHEME

SHAHALA GUL AND QURATULAIN KHALIQ  
SHAHALAGUL@GMAIL.COM, DR. MEHR ALI SHAH AHSAN AGRICULTURE UNIVERSITY, COUNTRY: GUJARAT, INDIA

**1. Introduction**  
A SIMPLE YET CHALLENGING PROJECT, TO PREDICT THE HOUSING PRICE BASED ON CERTAIN FACTORS LIKE HOUSE AREA, BEDROOMS, FURNISHED, NEARNESS TO MAINROAD, ETC. THE DATASET IS SMALL YET, IT'S COMPLEXITY ARISES DUE TO THE FACT THAT IT HAS STRONG MULTICOLLINEARITY. CAN YOU OVERCOME THESE OBSTACLES & BUILD A DECENT PREDICTIVE MODEL?

**2. Objective**  
1. BUILD REGRESSION MODELS TO PREDICT THE SALES W.R.T A SINGLE & MULTIPLE FEATURE.  
2. ALSO EVALUATE THE MODELS & COMPARE THEIR RESPECTIVE SCORES LIKE R2, RMSE, ETC

**3. Methodology**  
WE WANT TO SEE THE RELATIONSHIP PRICE AND BATHROOMS, STORIES, MAIN-ROAD, GUESTROOM, BASEMENT, HOT-WATER-HEATING, AIR-CONDITIONING, PARKING, PREFER-AREA, FURNISHING-STATUS

**4. Result and Discussion**  
We have excluded hot-water, air-conditioning, prefer-area  
Multiple Regression  
$$Y_i = f(X_1, \beta) + \epsilon_i$$
  
Estimates  
Prediction Expression  
$$Y = 290.18517365 \cdot \text{area} + 121024.41192 \cdot \text{bedrooms} + 102573.74542 \cdot \text{bathrooms} + 587678.56038 \cdot \text{stories}$$

**5. Reference**  
Harrison, D. and Rubinfeld, D.L. (1977) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 41-59.  
Berkley, D.A., Kohn, E. and Welch, J.E. (1988) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

### Countries Data Analysis

Sidra Gul (Sidragul567@yahoo.com)  
Bahauddin Zakariya University, Multan, Pakistan

**Introduction**  
This analysis is based on the prediction of the income level with different factors like import, export, child mortality, health, gross domestic production (GDP), total fertility and inflation of the different countries.

**Objective**  
Fit the regression model to predict the income effect on the single and multiple factors  
Formulate the model on the base of R2 and RMSE etc.

**Methodology**  
We find the relationship between income, import, export, mortality, fertility rate, health, GDP and inflation.  
$$y_i = f(X_i, \beta) + \epsilon_i$$
  
$$y_i = \text{income}$$
  
$$X_i = \text{Explanatory variables}$$
  
$$x_1 = \text{import}$$
  
$$x_2 = \text{Export}$$
  
$$x_3 = \text{Child mortality}$$
  
$$x_4 = \text{Health}$$
  
$$x_5 = \text{G.D.P}$$
  
$$x_6 = \text{Total fertility}$$
  
$$x_7 = \text{inflation}$$
  
$$\beta = \text{parameter}$$
  
$$\epsilon = \text{error term}$$

**Result and Discussion**  
According to the criteria of R2 square and mean squared error metrics export, total fertility and GDP has significant result while health, import, inflation, life expectation and total fertility have non significant effect.

**Reference**  
Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics. McGraw-Hill.

Submitted to Dr. Quratulain Khaliq ISLP

### Evaluating the Impact of Remote Work on Flexibility, Productivity, and Health: A Statistical Analysis

SYED SHAFQAT HUSSAIN GARDEZI  
BAHAUDDIN ZAKARIYA UNIVERSITY, \*bilsicareem@gmail.com

**ABSTRACT**  
The study explores the impact of working from home on flexibility, productivity, and health. It compares remote work to traditional office environments. The study uses a quantitative approach to analyze the relationship between remote work and various factors like productivity, health, and work-life balance. The study also explores the impact of remote work on flexibility, productivity, and health. The study uses a quantitative approach to analyze the relationship between remote work and various factors like productivity, health, and work-life balance.

**INTRODUCTION**  
The shift to remote work has become increasingly prevalent, especially in the wake of the COVID-19 pandemic. Understanding the impacts of remote work versus office work is crucial for both employers and employees.  
Key areas of interest include productivity, flexibility, time management, physical and mental health, social interactions, and overall work performance.  
This analysis aims to explore these variables using survey data and statistical tests to draw meaningful conclusions.

**RESULTS**  
The analysis revealed that the majority of respondents have experienced working from home. There was no significant association between working from home and productivity changes (Pearson Chi-Square = 0.794, p > 0.376). In terms of flexibility, there was a significant association between working from home and flexibility (Pearson Chi-Square = 9.188, p < 0.002). These findings suggest that remote work offers significant flexibility benefits, particularly in terms of work-life balance and autonomy. However, there was no significant association between working from home and health outcomes (Pearson Chi-Square = 0.794, p > 0.376), or mental health issues (Pearson Chi-Square = 0.188, p > 0.662). These results indicate that while remote work offers flexibility benefits, it does not appear to have a significant impact on overall health or mental well-being. Further research is needed to explore the long-term effects of remote work on these aspects.

**CONCLUSIONS**  
The study found that remote work offers significant flexibility benefits, particularly in terms of work-life balance and autonomy. However, there was no significant association between working from home and health outcomes or mental health issues. These findings suggest that while remote work offers flexibility benefits, it does not appear to have a significant impact on overall health or mental well-being. Further research is needed to explore the long-term effects of remote work on these aspects.

Figure 4: Some Appealing Posters Designed by Pakistani Participants Using JMP Software, Under the Supervision of Dr. Quratulain Khaliq.





Figure 5: Some Beautiful Posters Designed by Indian Participants Supervised by Dr. S. Gandhiya Vendhan.



Figure 6: Some Beautiful Posters Designed by Pakistani Participants Under the Supervision of Dr. Quratulain Khaliq.

## 6. AUDIENCE FEEDBACK

The workshop has active participation from a diverse group of attendees, including data analysts, researchers, students, and professionals from various industries. The interactive sessions and hands-on exercises ensured that participants could apply the concepts they learned during the workshop. We gradually collected feedback forms

from our participants and received very positive responses. Our participants thoroughly enjoyed our workshop and requested that ISLP conduct more such workshops to promote statistical literacy in our community. They also emphasized the need for this kind of healthy competition, which allows people from different regions to learn from each other's work. The feedback from participants was overwhelmingly positive. Attendees appreciated the practical approach of the workshop, the clarity of the instructions, and the opportunity to work on real-world data problems. Many participants expressed interest in further training and advanced workshops on specific topics within data science and JMP software. We are very grateful to ISLP for their support. Without them, we could not have achieved success in this volunteer work. Their support has been instrumental in our accomplishments.

## **7.CONCLUDING REMARKS**

Today, we are moving towards an era dominated by artificial intelligence and machine learning. All the models used in artificial intelligence are based on statistics. As a result, the importance of statistics has increased even more, and it is critical for our new generation to learn them. We can promote statistical literacy by bringing people from different disciplines to our platform and creating a learning environment. We should engage in such healthy activities and courses to promote statistics both locally and globally. Statistics are the pillars of any subject; without them, research is impossible. Statistics are important for all disciplines, not just statisticians. In this era, only the statistics community can effectively spread knowledge of statistics. Even after using software, understanding the results requires knowledge of statistics.