PRACTICAL TECHNIQUES IN DATA MANAGEMENT

November 11-12, 2019

COURSE DESCRIPTION

This course serves as an introduction to practical techniques in data management using RStudio. Specifically, it covers the following:

- importing and exporting data
- transforming variables
- subsetting and summarizing data
- cleaning and reshaping data

Prior experience in R programming is not needed in order to take this course as it will also cover the basics of such language. The course serves as a launch pad in preparation for more advanced statistical modeling.

TARGET AUDIENCE

The training will equip practitioners with skills in order to start using R for data management. This will be particularly useful those who are involved in (but not limited to) the analysis of the following data structures:

- ✓ cross-section data (e.g. opinion polls, satisfaction surveys, needs assessment)
- ✓ time-series data (e.g. sales forecasting, economic modeling, overlay analysis)
- ✓ longitudinal data (e.g. client monitoring, market research, demand forecasting)

PREREQUISITE KNOWLEDGE

Knowledge of basic data management using other software (e.g. MS Excel) or programming languages will be helpful but not necessary.

PRACTICAL TECHNIQUES IN DATA MANAGEMENT

November 11-12, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 1

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Introduction to R
	▪ R
	▪ Rstudio and its interface
	▪ Basic Operations in R
	R Data Frames
	▪ Datasets in R
10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Continuation of Datasets in R
	Workshop: Exploring a Dataset's Attributes (Economic Data)
12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Operations on Datasets
03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Workshop: Working on Variables (Socio-Demographic Data)

PRACTICAL TECHNIQUES IN DATA MANAGEMENT

November 11-12, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 2

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	R Data Frames
	Operations on Datasets (cont)
10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Workshop: Subsetting, Merging, and Reshaping Datasets
12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Using "Tidyverse" Packages for Data Management

- Tidyr for Reshaping
- Dplyr for Subsetting and Summarizing

03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Workshop: Using Tidyr and Dplyr (Sales Data)
	Exporting Datasets

PRACTICAL TECHNIQUES IN DATA MANAGEMENT

November 11-12, 2019

ABOUT THE RESOURCE PERSON

CHARLENE MAE CELOSO is an Assistant Professor from the UP School of Statistics. She obtained her BS Statistics degree from UP Diliman in 2015, graduating Magna Cum Laude. She has obtained her MS Statistics degree last year from the same institution. As an instructor, she has handled courses in elementary statistics, introduction to programming, introduction to exploratory data analysis, and introduction to regression analysis. She has also served as a resource person in workshops on introduction to R programming and time series analysis.

LOCATION ANALYTICS USING SPATIAL STATISTICS

November 14-15, 2019

COURSE DESCRIPTION

Many statistical techniques focus on the analysis of data based on its numeric or textual form. However, many of us overlook on one aspect of data: its location. Geography, which represents statistical data's spatial property, offers a lot of potential for gaining deeper insights on phenomena, such as population and demographics, among many others.

This session will focus on how to incorporate and leverage data's spatial property in doing statistical inquiry using a technology called Geographic Information Systems (GIS). The ArcGIS platform, a commercial industry-standard GIS software solution, and R, an open-source statistical package, shall be used separately and eventually integrated in this session. Specific techniques to be learned in this session include:

- Tabular data conversion and processing to spatial data
- Exploratory spatial data analysis
- Spatial and attribute-based querying
- Basic cartography
- Spatio-statistical analysis

LOCATION ANALYTICS USING SPATIAL STATISTICS

November 14-15, 2019

TARGET AUDIENCE

Statistical practitioners and students who do not yet have any prior knowledge or experience in using GIS, or current GIS users who want to learn the GIS capabilities of the ArcGIS platform in various statistical applications. Examples of scenarios to be presented include the following industries and domains to demonstrate the abovementioned techniques:

- Community-based socioeconomic profiling for local government offices
- Real property inventory and analysis for city and municipal assessors
- Field data collection for market research
- Spatial analytics for demographic studies

PREREQUISITE KNOWLEDGE

Participants must have basic usage experience in Microsoft Excel or any other office spreadsheet applications. Some understanding of tabular data organization, mapping, and statistics is favorable for quick understanding of the topics to be presented, but not required.

LOCATION ANALYTICS USING SPATIAL STATISTICS

November 14-15, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 1

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Introduction to the Course Topics and Expectation
10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	What is GIS?
	What is ArcGIS?
	"Bridging" ArcGIS and R
	Spatial data models
	Anatomy of a spatial dataset
	Overview of GIS applications in statistical practice
12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Feature extraction in ArcGIS
	Example: Community-based socioeconomic profiling for local government offices
	Processing tables into spatial datasets
	Example: Real property inventory for city and municipal assessors
	Field data collection in ArcGIS
	Example: Field data collection for market research
03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Map data visualization in ArcGIS
	Basic cartography in ArcGIS
	Example: Spatial analytics for demographic studies

LOCATION ANALYTICS USING SPATIAL STATISTICS

November 14-15, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 2

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Overview of spatial pattern analysis Descriptive vs. inferential spatial statistics Measuring spatial central tendency and spread
10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Measuring spatial clusters (global scale) in R Example: Real property analysis for city and municipal assessors
12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Measuring spatial clusters (local scale) in R Example: Customer behavior analysis for market research
03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Spatial modeling in R Example: Modeling the behavior of populations in demographic research Recap and conclusion

LOCATION ANALYTICS USING SPATIAL STATISTICS

November 14-15, 2019

ABOUT THE RESOURCE PERSON

LIZ LINDA ALINDOGAN is a GIS specialist in the Training Division of Geodata Systems Technologies, Inc. and is one of the senior instructors who conducts basic to advanced GIS and GPS courses that covers desktop, geodatabase, mobile, and enterprise GIS. She has already earned multiple technical certifications from Esri for her experience in geodatabase management and usage of ArcGIS Desktop. She already conducted classes for numerous local and national government offices, such as the National Mapping and Resource Information Authority (NAMRIA), Mines and Geosciences Bureau (MGB), and the Metropolitan Waterworks and Sewerages System (MWSS); and for private/non-government organizations such as Maynilad, the Philippine Disaster Resilience Foundation (PDRF), and Prime Water. Aside from teaching, she is also active is GIS project implementations, some projects with which she was engaged are the e-Tax Map Project of the Quezon City Assessor's Office, the GIS-Based Decision Support System of the Forest Management Bureau (FMB), and the Location Referencing System Project of the Road Board of the Philippines. She earned her degree on Bachelor of Science in Geodetic Engineering from the University of the Philippines, and is currently completing her master's degree on Geomatics Engineering in the same university. She also previously worked as a research assistant at the UP-Training Center for Applied Geodesy and Photogrammetry, and at the Philippine Institute of Volcanology and Seismology (DOST-PHIVOLCS).

TIME SERIES ANALYSIS AND FORECASTING
November 18-19, 2019

COURSE DESCRIPTION

The Time Series Analysis and Forecasting using R Studio training course provides participants with a working knowledge about the basic principles and statistical techniques in time series analysis and forecasting, specifically:

- Preliminary concepts in time series
- Descriptive tools and the basic components of a time series
- R Studio syntax for the different techniques in time series analysis and forecasting
- Exponential smoothing
- ARIMA modeling

TARGET AUDIENCE

The knowledge and skills learned in this training will enable a participant to make sound statistical forecasts based on time-series data. The training is specifically tailored-fit to those involved with decision-making processes, planning personnel, and monitoring and evaluation officers.

Applications of this training on time series analysis and forecasting include (but not limited to) the following:

1. Economic planning
2. Forecasting of sales
3. Supply chain planning
4. Production and capacity planning
5. Project monitoring
6. Financial risk management

PREREQUISITE KNOWLEDGE

Participants of this training must have at least a basic understanding of several statistical methods such as summary measures (e.g., mean, standard deviation, correlation). They must have a working background in using MS Excel. Knowledge in the R language is not required.

TIME SERIES ANALYSIS AND FORECASTING
November 18-19, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 1

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Some Preliminaries

- Basic Concepts in Forecasting
- The Time Series Data
- Forecast Accuracy Measures
- Time Series Decomposition

10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Introduction to R Studio

- The R Studio Interface
- Basic Data Management and Functions in R Studio
- Creating a line graph
- Workshop: Time Series Decomposition (using GDP Data)

12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Forecasting Using Exponential Smoothing

- What is Exponential Smoothing
- Types of Exponential Smoothing
- Performing Exponential Smoothing in R Studio

03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Forecasting Using Exponential Smoothing (cont.)

- Workshop I on Exponential Smoothing (using GDP data)
- Workshop II on Exponential Smoothing (using sales data)

TIME SERIES ANALYSIS AND FORECASTING

November 18-19, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 2

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	ARIMA Modeling

- Stationarity and Unit Root Tests
- Autocorrelation and Correlograms

10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	ARIMA Modeling (cont.)

- The Autoregressive Process
- The Moving Average Process

12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	ARIMA Modeling (cont.)

- Non-stationary and Integrated Processes
- Estimating the ARIMA Model in R Studio
- Workshop I on ARIMA modeling (using PSEi data)

03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	ARIMA Modeling (cont.)

- Workshop II on ARIMA modeling (using Dubai gas prices)

Recap and Conclusion

TIME SERIES ANALYSIS AND FORECASTING
November 18-19, 2019

ABOUT THE RESOURCE PERSON

JOHN CARLO P. DAQUIS is an Assistant Professor IV of the UP School of Statistics. He has been teaching undergraduate time series analysis and forecasting. He is also a resource person on several occasions on topics that include: regression analysis, time series analysis, forecasting, basic statistical methods using various statistical software. Prof. Daquis also worked with both public and private institutions like the Asian Development Bank and the Philippine Statistical Research and Training Institute as a consultant.

PREDICTIVE ANALYTICS USING SUPERVISED STATISTICAL LEARNING TECHNIQUES
November 21-22, 2019

COURSE DESCRIPTION

This training program is a survey of most commonly used supervised machine learning methods. Statistical machine learning refers to a wide range of algorithms and techniques of learning from observed data to construct models in order to make inferences and predictions and to make decisions. However, this course only focuses on the main methodologies in supervised machine learning which includes linear regression, classification problems, tree-based methods, neural networks, and support vector machines. With a little preview of the theoretical concepts of these tools, the course offers the relevant applications in different fields and areas.

TARGET AUDIENCE

Attendees who want to know the methods in the rising field of statistical machine learning are highly encouraged to participate. Also, those who want to expand their statistical programming skills and implementation in R are invited.

Market researchers, business managers, practitioners in the social sciences are to take pride in the illustrations which range in their respective fields.

PREREQUISITE KNOWLEDGE

Participants must have taken introductory courses on basic statistics covering both descriptive and inferential analysis.

PREDICTIVE ANALYTICS USING SUPERVISED STATISTICAL LEARNING TECHNIQUES

November 21-22, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 1

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Statistical Learning

- Supervised versus Unsupervised Learning
- Regression versus Classification Problems
- Introduction to R

Regression Methods
- Simple Linear Regression

10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Regression Methods (cont)

- Multiple Linear Regression
- Variable Selection
- Diagnostics Checking

12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Regression Methods (cont)

- Ridge Regression
- PC Regression
- Workshop: Regression Methods

03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Regression Methods (cont)
	Logistics Regression Workshop

PREDICTIVE ANALYTICS USING SUPERVISED STATISTICAL LEARNING TECHNIQUES

November 21-22, 2019

TOPICS TO BE COVERED AND SCHEDULE

DAY 2

08:00 am – 08:30 am	Registration
08:30 am – 10:00 am	Tree-based Methods

- Regression Trees
- Workshop: Regression Trees

10:00 am – 10:30 am	Break
10:30 am – 12:00 pm	Tree-based Methods (cont)

- Classification Trees
- Workshop: Classification Trees

12:00 pm – 01:00 pm	Lunch
01:00 pm – 03:00 pm	Tree-based Methods (cont)

- Random Forests
- Workshop: Random Forests

03:00 pm – 03:30 pm	Break
03:30 pm – 05:30 pm	Neural Networks

- Overview of Neural Network
- Case Study Review

Support Vector Machine
- Overview of SVM
- Case Study of Review

PREDICTIVE ANALYTICS USING SUPERVISED STATISTICAL LEARNING TECHNIQUES
November 21-22, 2019

ABOUT THE RESOURCE PERSON

STEPHEN JUN VILLEJO is currently an Assistant Professor and College Secretary of the School of Statistics, University of the Philippines Diliman. He graduated magna cum laude with a bachelor's degree in Statistics last 2013 and finished his Master of Science in Statistics last 2015 at the University of the Philippines Diliman. He handled courses on probability theory, statistical inference, regression analysis, multivariate analysis, and stochastic processes. His research interests are Spatial and Spatio-Temporal Analysis, Epidemiology, Bayesian Computation, and multivariate analysis. He has published papers in peer-reviewed journals and has presented papers in local and international conferences. He has handled several trainings on basic statistics, time series analysis, predictive analytics, technical report writing, and data management. In addition, he was involved in doing several research projects and in doing government and private consultancy work.