E-Tidings Newsletter

SCASA Events and News

Volume 11, Issue 4

April, 2023

IN THIS ISSUE

42ND APPLIED STATISTICS WORKSHOP REPORT

By Olga Korosteleva, CSULB, SCASA Treasurer and Editor of eNewlsletter

On Friday, April 21, between 9 AM and 4:15 PM, SCASA hosted the 42nd Annual Applied Statistics Workshop, in the form of a webinar via Zoom. A total of 37 people attended this workshop.

Our instructors, Josh Dunlap, Emmanuel Hernandez, Gauthier Kelly, and David Laredo work as data scientists and data engineers at Amazon Web Services. The topic of the workshop was *Cloud Computing and Sagemaker*. The morning session included general overview and no coding, whereas the afternoon session consisted primarily of hands-on coding. The workshop was saturated with useful information and received high praises and "five-star" reviews in the chat.

In addition, as it is our tradition by now, the winner of the first place in the SCASA Regional Data Visualization Poster Competition presented his research. Raymond Levine is a junior at Westview High School in San Diego and the title of his research is "Covid on the Border". Congratulations to Raymond!

And, as always, the event culminated in a book raffle, where 12 books were raffled off. Only people who were present were eligible for a prize.

We thank all the attendees for their time and support of the event and we would like to extend our gratitude to Amgen Inc., our faithful sponsor for many years now. Thank you, Amgen, for your continuous support!
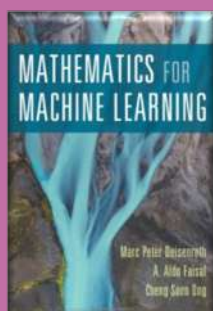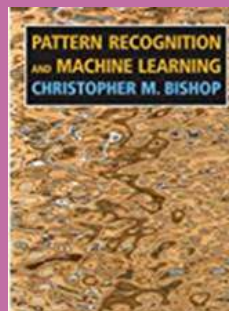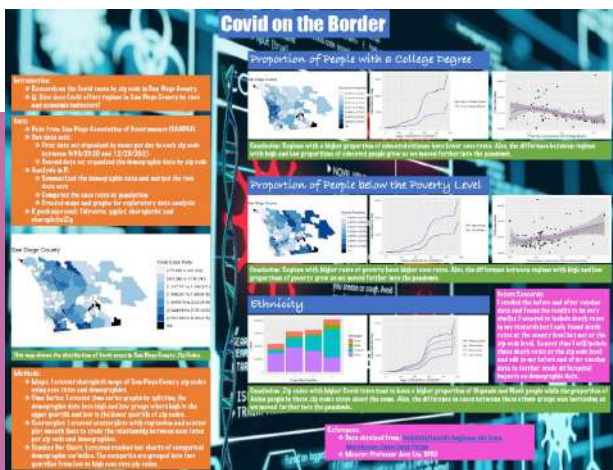
WOMEN IN DATA SCIENCE
RIVERSIDE

REPORT

By Varsha Kuppa, UCR

The 2023 WiDS Riverside Conference "Moving Beyond Data"

The Women in Data Science (WiDS) conference elevates women in the field by providing inspiration, education, community, and support. As part of the annual WiDS led by Stanford University at about 200+ locations worldwide, the WiDS 2023 Riverside virtual conference took place on April 15th, 2023. It started with a research talk by Dr. Vivian Li, whose area of expertise includes genomic data science and analysis. Dr. Li used her own and other scientists' research to demonstrate the power of data science methods in the biological sciences. Genomic data science, through the use of powerful computational and statistical methods, enables researchers to decode the information hidden in DNA and RNA sequences.

For the next portion of the conference, female data science professionals in industry and academia—were invited to participate in informational panels. The industry panel included accomplished individuals working at Spotify, Meta, Bank of America, Atlas AI, and Amazon. They shared valuable insights into how they found success in the private data science sector. For example, apart from the obvious necessary skills to be a successful data scientist, the industry panelists revealed a less-known but crucial skill that is vital in a data scientist's career path: strong technical writing competency. They also compared different working experiences in both big companies and small start-up companies.

Following the industry panel, a graduate student panel consisting of Ph.D. students from Statistics and Data Science majors was scheduled. The graduate student panelists revealed what a Ph.D. program looks like, what they did to get there, and why they chose to do a Ph.D. Some of the panelists were not data science or statistics majors when they started college but fell in love with the field after brief exposure.

The day ended with an intro workshop followed by an advanced workshop in both Python and R. With workshop leaders who are experts at using their respective technologies, they tailored their workshops to both beginner and advanced students, facilitating learning for all levels of learners.

Through a series of panels, workshops, and information sessions, WiDS 2023 Riverside successfully uplifted accomplished women in data science while teaching the next generation of data scientists what to expect. Those that attended the panels, the workshops, or both, gained not only information but also a humanized and less abstract perspective of the data science field.

UPCOMING STATISTICAL EVENTS IN OUR AREA AND BEYOND

SAS Innovate

May 8-10, 2023
Hyatt Regency Orlando

https://innovate.sas.com/event/8f50bb6d-ae11-4687-a7a9-f6c8204eba24/summary?RefId=PLN2755_1766711525

2023 Data Con LA

August 12, 2023
University of Southern California
https://www.dataconla.com/

2023 IDEAS Global AI Conference

October 14, 2023
Los Angeles Convention Center, 1201 S Figueroa St
https://www.joinideas.org/ai2023/

Western Users of SAS Software (WUSS) Conference

Oct 31-November 1, 2023
San Diego
https://www.wuss.org/cfp/

Regression Modeling Strategies Short Course 2023

The Regression Modeling Strategies (RMS) 4-day short course will be held on May 16-19, 2023 as a very interactive live web course using Zoom.

FRANK E. HARRELL, JR., PH.D., Professor, Department of Biostatistics, Vanderbilt University School of Medicine, Instructor

DREW G. LEVY, PH.D., GoodScience Inc., Guest Instructor and Moderator

TARGET AUDIENCE

Statisticians and other quantitative researchers who want to learn some modern flexible methods for developing and validating predictive models, including approaches to missing data imputation, data reduction, model specification and variable selection, importance of causal inference in formulating models, model validation, relaxing linearity assumptions, and how to choose between machine learning and statistical models.

PREREQUISITES

Good working knowledge of ordinary multiple regression models. Some individuals will want to take the free **Biostatistics for Biomedical Research** course in preparation (especially sessions on regression). Or take the 1-day workshop on May 12 to enhance **R** and **RStudio** skills, learn about multiple linear regression (a prerequisite for the 4-day course), and to get an introduction to the R rms package.

Registration for RMS 4-day short course (closes May 12)
https://hbiostat.org/doc/rms/4day.html

Registration for R Pre-RMS Workshop Friday May 12, 2023
https://hbiostat.regfox.com/pre-rms2023
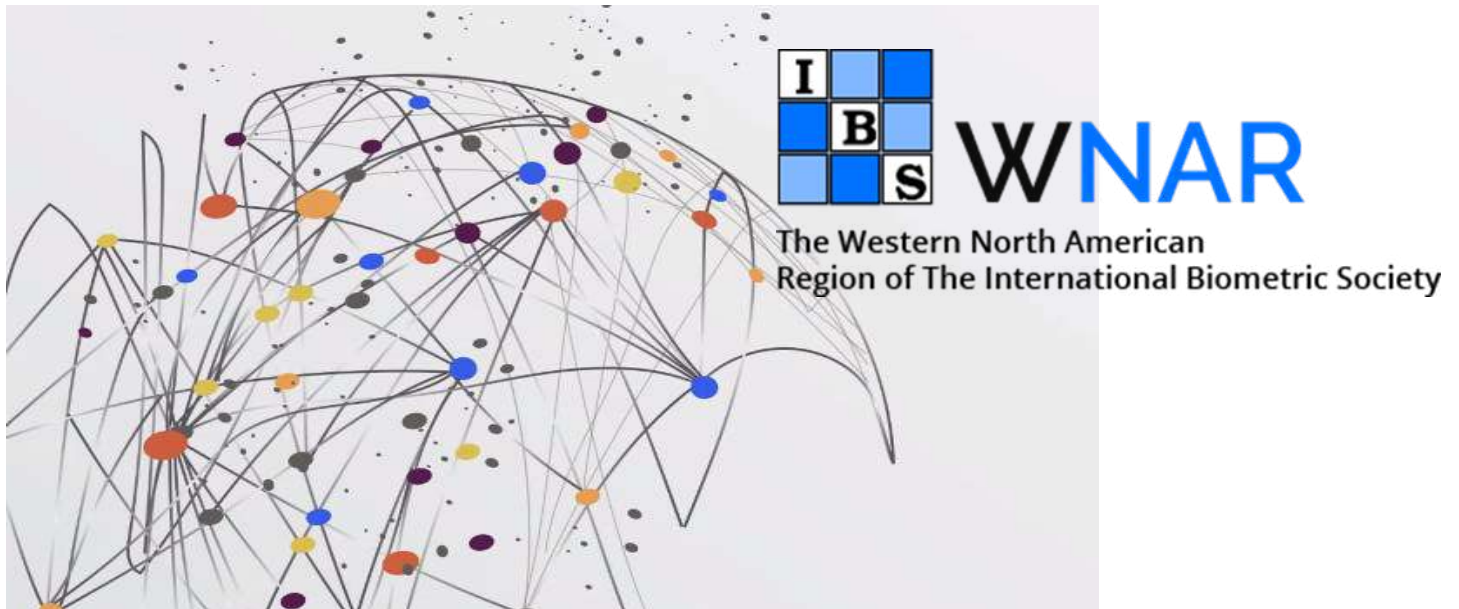
The University of Washington Department of Biostatistics invites you to explore the 50+ online and in-person short courses offered through the 2023 Summer Institutes, July 10-August 4.

Taught by leading educators and researchers from UW and around the world, these graduate-level courses in statistical methods, techniques and analyses that will be taught across four institutes:

Statistical Genetics (SISG) , July 10-28, In-person at University of Washington, Seattle

Statistics and Modeling in Infectious Diseases (SISMID) , July 10-26, In-person at University of Washington, Seattle

Statistics for Clinical & Epidemiological Research (SISCER) , July 10 – August 1, Online

Statistics for Big Data (SISBID) , July 24 – August 5, Online

Early rates available through June 5. Learn more at https://si.biostat.washington.edu/ .

This year, we are excited to also host the 2023 Symposium in Statistical and Quantitative Genetics , July 14-16. Leading researchers will present their current work in statistical genetics, genome sciences, and public health genetics. Applications are invited for contributed oral or poster presentations.

```
Editor - Untitled1 *

    /*Anatomy of Proc Report Statement and Options*/
    /* by */
    /*Sunil Gupta, SASSavvy.com*/
```

Key Design Factors **Page Style / Layout**	Listing/Table, Formatted Dataset/Summarize, Layout (Concatenate, Nest/Across, Group), Alias/Derive Variable, Subtotals/Totals, Style ODS Style, Orientation (Portrait/Landscape)
Formatted Dataset **Proc Report Options** **By Variables / Page** **Subtotal / Totals Lines**	Extract Datasets from Proc Means/Proc Freq, Build Row/Section by Row/Section, Display as Proc Report Listing nowindows nocenter missing headline headskip nofs list split='|' spacing break after sex / < options > ; break after page / page; rbreak after/ ol summarize ;
Model Listing **Model Summary Table**	**column** usubjid age; **define** usubjid / order; **define** age / display 'Age'; **column** sex age; **define** sex / group; **define** age / mean 'Mean Age';
Column Options	Row Variables Nest Columns , (Group Columns) Optional Optional , Required
Concatenate Columns **Derive Columns** **Alias Columns** **Nest / Across Columns** **Column By Groups** **Row By Groups** **Column Combinations**	**column** gender age; **define** gender / display; **define** age / display; **column** (weight height bmi); **define** bmi / computed format=4.1 'BMI'; * bmi is computed from weight and height; **column** sex, (age = agen age = agemean); **define** sex / across center; **define** agen / analysis n; **define** agemean / analysis mean ; **column** gender, age; **define** gender / across; **define** age / mean; **column** gender, (age weight); **define** gender / group; **define** age / mean; **define** weight / mean; **column** gender (age weight); **define** gender / group; **define** age / mean; **define** weight / mean; Concatenate, Derive, Alias, Nest / Across, Row / Column By Groups
Define Options **Define Summary Keywords**	format, 'label', style()={}, group, order, center, across (columns), analysis (num vars), noprint, width, right/left, spacing, 'label', flow categorical (pctn, pctsum), continuous (n, mean, min, max, std, sum, computed, nmiss, range)
Compute Block	by group table formatting processing, derive variable, before/after <variable>, line

```
compute after region;  * subtotals after by variable region ;
 region = 'Totals:';
endcomp;

compute SEX / CHAR;
 if SEX > ' ' then SEXHLD = SEX; * conditional statements are valid;
 if SEX = ' ' then SEX = SEXHLD;
endcomp;

define bmi / computed format=4.1 'BMI';
compute bmi;
 bmi = weight / (height*height) * 703;  * weight and height variables exist in input dataset;
endcomp;
```

SAS® Savvy
The One Stop SAS® & CDISC Solutions

Dr. Normalcurvesaurus, Ph.D. presents