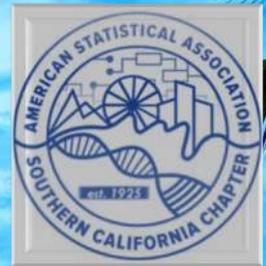


E-Tidings Newsletter

SCASA Events and News



Volume 10, Issue 7 –8

July-August, 2022

IN THIS ISSUE



- PAGES 2-3** **Virtual ASA Traveling Course, 10/22/2022**
- PAGE 4** **Post-Event Report: The R User Conference**
- PAGE 5** **Post Event Report: DS-PATH Summer Fellowship Program**
- PAGE 6** **JMP Discovery Summit 9/12/-9/14 in person, 9/19-9/22 online**
- PAGE 7** **SAS Explore – Virtual Event, 9/27-9/29**
- PAGES 8-11** **Interview with Jeffrey Gornbein**
- PAGE 12** **Dr. Normalcurvesaurus, Ph.D. Presents**



Virtual ASA Traveling Course

October 22, 2022, 10AM-1PM

An Introduction to Second-generation P-values and Their Use in Statistical Practice



Instructors: *Jeffrey D. Blume, School of Data Science, University of Virginia, and Megan H. Murray, Biostatistics, Vanderbilt University*

When: *Saturday, October 22, 2022, 10 AM – 1 PM*

Location: *via Zoom (link is posted on Eventbrite.com)*

Cost: *\$25 general/\$0 students and retirees*

Registration: *<https://www.eventbrite.com/e/404430220677>*



Abstract: Second-generation p-values were recently proposed to address the well-known imperfections of classical p-values. Their implementation can largely be thought of as codifying 'good standard practice' for interpreting and reporting classical p-values. Second-generation p-values maintain the favorable properties of classical p-values while emphasizing scientific relevance to expand their utility, functionality, and applicability. In particular, they can report evidence in favor of the alternative, in favor of the null hypothesis, or neither (inconclusive); they automatically incorporate an adjustment for multiple comparisons and multiple looks; they have lower false discovery rates than classical p-values; and they are easier to interpret. Second-generation p-values have been shown to work well in regularized models. They also lead to significantly improved model selection procedures in linear and generalized linear models. Also, second-generation p-values are non-denominational in the sense that they are readily applied in frequentist, likelihood and Bayesian settings.

Continued on the next page ...

... Continued from the previous page.

Outline and Objectives: This course will briefly revisit the history of p-values as originally envisioned in significance and hypothesis testing, and the resulting controversy over their scientific interpretation. The importance of distinguishing between three key inferential quantities (the measure of the strength of evidence, design error rates, and false discovery rates for observed data) will be illustrated. The second-generation p-value will be introduced and contrasted with standard methods. The workshop will explain how to design studies in which the second-generation p-value is used as the primary mode of inference. We will cover the computation of second-generation p-values (in R), guidelines for presenting results, and when appropriate, how to present accompanying false discovery rates. Multiple examples will be presented using data from clinical trials, observational studies, and high-dimensional analysis of large-scale data. Advanced applications in model selection, adaptive monitoring of clinical trials, and regularized models will be shown if time allows. Mathematical details will be kept to a minimum, e.g., statistical properties will be presented but without mathematical proof.

About the instructors:

Jeffrey D. Blume, Ph.D. is the Quantitative Foundation Associate Dean for Academic and Faculty Affairs and Professor of Data Science at the School of Data Science at the University of Virginia. He recently moved from Vanderbilt University where he served as Vice-Chair for Education in the Department of Biostatistics and Director of Graduate Education in the Data Science Institute. Professor Blume is a fellow of the American Statistical Association and has extensive experience in the operation, analysis, and methodological aspects of Clinical Trials and the analysis of a wide array of biomedical research. He has won numerous awards for his teaching and mentoring, and his work on racial disparities in lung cancer screening. His statistical research focuses on the foundations of statistical inference, mediation modeling, diagnostic clinical trials, and prediction modeling. Professor Blume has published extensively on the foundations of statistical inference, particularly on likelihood methods for measuring statistical evidence, and he authored the origin publications on second-generation p-values.

Megan H. Murray is a Ph.D. candidate in Biostatistics at Vanderbilt University working with Professor Blume. Her methods of research focus on evaluating and extending the statistical properties of second-generation p-values and on characterizing false discovery rate methodology. She recently co-authored a paper describing her development of an R package ("FDRestimation") for computing, estimating, and visualizing false discovery rates (currently on CRAN). Her biomedical research has focused on the areas of lung cancer, particularly in assessing racial disparities in lung cancer diagnosis, and of graduate student survey data, specifically evaluating if a positive mentor experience leads to more published papers or a shorter time to graduation. Megan has won several teaching awards in both Biostatistics and Data Science at Vanderbilt.



useR! 2022

POST-EVENT REPORT

The R User Conference

By Madeline Bauer, USC (retired)

What is UseR!2022? UseR! 2022 was an all-virtual conference, with 1227 registered participants from 96 countries. UseR!2022 is organized by R community volunteers and supported by the R Foundation. Attendees include R developers and users who are data scientists, business intelligence specialists, analysts, statisticians from academia and industry, and students. The conferences showcase applications of R software and developments in the software itself, as well as new and updated R packages that provide boundless additional functionality to R. Community contributions form the foundation of useR! Conferences.

Recordings of most of the UseR! 2022 presentations are available on YouTube, at https://www.youtube.com/c/useRConference_global

UseR!2022 Tutorial: One of the most useful talks was a tutorial on Regression Modeling Strategies by Frank Harrell. The video is available at <https://youtu.be/UcpkIJXlvY>.

SCASA members will remember Frank's workshops on Regression Modeling Strategies at the Applied Statistics Workshop in 2005 (based on the 1st edition of the book) and in 2019, based on the 2nd edition. Both editions of that book included chapters with comprehensive guidelines for preparing for regression analysis: regression assumptions, modeling interactions, dealing with missing data, model development, model complexity, dealing with large numbers of predictors. That is followed by chapters with worked examples on specific regression techniques, including binary and ordinal logistic regression, and non-parametric and parametric survival analysis. The S-Plus and R packages Hmisc and rms have been developed by the author reflecting years of experience using S-Plus and R extensively in biomedical research and clinical trials.

So What's New? The e-book R Workflow for Reproducible Data Analysis and Reporting <https://hbiostat.org/rflow/> is the essential book! It encourages and supports reproducible research reports using R with Rmarkdown and Quarto and facilitates understanding the data and their quality and completeness.

R Workflow incorporates the guidelines from those beginning chapters in the book, and, thanks to the novel online format using Rmarkdown and now Quarto it also provides the analysis project workflow for modern, high-quality, reproducible analysis. The book uses examples to illustrate the many functions in the analysis toolkit developed by the author for data import, processing, analysis, and tables. along with the actual code, output, and graphics (both static and interactive).

As we all know so well, ~80% of data science is "cleaning" the data, so you won't be surprised that "analysis" is in chapter 15! Thanks to Frank Harrell for sharing his "best statistical practices" and his well-developed "toolkit" – as Bill Cleveland says, "Tools Matter."

SOME RESOURCES

The book: Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer (2015). <https://link.springer.com/book/10.1007/978-3-319-19425-7>.

The E-book R Workflow:

<http://hbiostat.org/rflow/>

Tutorial Video:

<https://youtu.be/UcpkIJXlvY>.

Tutorial "handouts":

Main course notes:

<https://hbiostat.org/doc/rms.pdf>

Software: <http://hbiostat.org/R/rms/>

Other useful links: <https://hbiostat.org/>

Blog - Statistical Thinking: <https://www.fharrell.com/>

Blog - Statistical principles: <https://www.fharrell.com/post/principles/>

Blog - R Workflow Overview: <https://www.fharrell.com/post/rflow>

Quarto templates: <https://github.com/mcanouil/awesome-quarto>

Reproducible Research: <https://hbiostat.org/rr/index.htmltemplate>

Rviews review of R Workflow: <https://rviews.rstudio.com/2022/06/17/frank-s-workflow>

POST-EVENT SUMMARY

DS-PATH Summer Fellowship Program

The DS-PATH Summer Fellowship Program is a 10-week program that provides data science training and education for students from partnering institutions in the Inland Empire of Southern California. Community college, undergraduate, and M.S. graduate students work on team-based real-world projects in collaboration with faculty and industry partners while also attending talks and participating in technical workshops. Professional development workshops that focus on resume building, ethics, and written and oral communication are also part of the fellowship program. The first cohort of 35 summer fellows completed the DS-PATH program in Summer 2022. Applications for next summer's program will open in early 2023.



Teams presented project results at a Summer Symposium on Thursday, August 18, 2022 at UC Riverside's campus.

Projects Included:

- Anomaly Detection
- ESRI - GIS Case Study
- FAA Event Report Text Mining
- GIS Mapping - Client Map and Resource Gaps in Riverside County
- JPL CA Central Valley Land Subsidence Dashboard
- JPL CA Central Valley Land Subsidence Modeling
- Lexi, the Medical AI
- Monet Short Video Performance Score Prediction Algorithm
- San Bernardino County Traffic Congestion Analysis
- Social Distancing during COVID-19 Data Analysis

To learn more about DS-PATH initiatives, please visit dspathways.com



The DS-PATH Summer Fellowship Program is supported by the NSF under the Harnessing Data Revolution Data Science Corps grant. HDR DSC Award #2123444, #2123271, #2123313

AMERICAS



DISCOVERY SUMMIT



EXPLORING DATA
INSPIRING INNOVATION

<https://discoverysummit.jmp/en/2022/usa/home.html>

In-Person

Yes, it's really happening.
Discovery Summit is back in-person!

**September
12-14, 2022**

[Learn More](#)

Online

Can't make it in-person?
Join us online. *Can* make it in-person? Join us online too!

**September
19-22, 2022**

[Learn More](#)

SAS EXPLORE

A Virtual Event for Technologists
Sept. 27-29

<https://explore.sas.com/event/8c314b18-f97b-4156-ae95-2358df0ba862/?websitePage:670a9359-fd8c-4de9-ae07-090b4d8153d8>

This virtual event is complimentary.

[Register now!](#)

AGENDA AT A GLANCE

Start [previewing sessions](#) and check back often as we'll be adding more sessions soon!

We've planned global showtimes for a global audience. When you register, select your region to see sessions in your regional time zone. Sessions will start at 10 a.m. for all regions.

DAY 1	DAY 2	DAY 3
<p>9:45 - 10 A.M. Kickoff</p>		
<p>10 - 11:30 A.M. General Session and Keynote SAS CEO Jim Goodnight and CTO Bryan Harris will share the future of SAS technology. Technical experts will lead a demo of SAS® Viya® throughout the analytics life cycle. The session concludes with SAS Hackathon winners sharing their business cases.</p>	<p>10 - 11:30 A.M. General Session Learn about SAS technology advancements and how customers are using SAS to solve their analytics challenges. We'll recognize the SAS User Feedback Award winner and discuss how users are driving innovation.</p>	<p>10 A.M. - 4:30 P.M. Training Sessions by SAS Education Take advantage of free training courses:</p> <ul style="list-style-type: none"> • Building Advanced Dashboards Using SAS Visual Analytics for SAS® Viya® • Creating Reports Again and Again ... And Again With MacroLearning in Model Studio • More Analysis, Less Coding with SAS Viya for Learners • SAS Visual Text Analytics: Exploring Literature for Drug Safety Signals • Software Workshop: Advanced Tips and Tricks for SAS Model Studio • Software Workshop: Automated Large-Scale Forecasting • Software Workshop: End-to-End Machine Learning in Model Studio • Speaking to Microsoft Excel with PROC SQL in SAS Viya • Struggling With Syntax? SAS Studio to the Rescue! • Using Python for Data Analytics in SAS Viya
<p>11:30 A.M. - 12:15 P.M. Super Demos Choose from demos that provide a quick look at SAS software.</p>		
<p>12:15 - 12:45 P.M. Lunch break and time to engage!</p>		
<p>1 - 4 P.M. Breakout Sessions Expand your SAS knowledge and technical skills with how-to guides, road maps and demos on key topics, including advanced analytics, AI and machine learning, cloud and data management, and more.</p>		
<p>4 - 4:30 P.M. Keynote</p>		

The Floor is Yours!

Interviewed by Jennifer Lee



Interview with Jeffrey Gornbein

This month, I had the honor of interviewing SCASA member Jeffrey Gornbein. Jeffrey is a Professor in the Department of Computational Medicine (formerly Biomathematics) and is a Principal Statistician in the Statistics Core of the Department of Medicine at UCLA. His research in the field of biostatistics is not only interesting but also highlights a different area of statistics that is in demand these days. Although Jeffrey may not be as active in SCASA

as some of the previous interviewees, he has been a member since before 2000 and that dedication to SCASA shows in his enthusiasm for the services SCASA provides its members. *[Ed Note: Some questions were edited for clarity]*

What was your initial reason for joining SCASA?

[A] member since before 2000, [I] liked the workshops, wanted to network with, and learn from, colleagues.

What has been your general experience with SCASA?

They are very friendly and supportive, [but] I have only had time to attend the workshops.

We appreciate you coming to the workshops even with your busy schedule! What do you wish to see SCASA do more in the future?

Continue the workshops.

Those workshops are so very educational. I'm looking forward to the next one. How has SCASA enriched your career in statistics in any way?

With the workshops and, some time ago, some good mentoring/advice.

I would love to know how you think SCASA provided some good mentoring advice.

One of the "traditional" weaknesses of some academic statistics programs is that they did not provide enough one-on-one mentoring of students or exposure them sufficiently to the non-mathematical/non-technical aspects of statistical collaboration. The SCASA workshops and other activities help students and new graduates meet and interact with other statisticians, including those who work in industry. The academic programs are getting better at this, but SCASA can still make a contribution.

Continued on the next page.

The Floor is Yours!

Interview with Jeffrey Gornbein

...Continued from the previous page.

I absolutely agree with your perspective. I hope SCASA continues to help students and new graduates as they begin their careers.

Now, I would like to get to more personal things. I see you are a Professor at UCLA in the Department of Computational Medicine as well as the Principal Statistician in the Statistics Core of the Department of Medicine at UCLA. What made you go into the field of biostatistics?

I like math and applying this to medicine is both interesting, useful, and does not hurt anyone. [I] did not have the temperament or dexterity to be an MD.

Applying math to medicine is, to me, very interesting, since it is not an area I am very familiar. What do you love about your current position?

[I] get to help do medical research, use cool software, get to teach, [and] hang out with very intelligent folks.

What kind of software are you talking about?

There is a lot of conventional statistical software including SAS, R, Stata, Python, SPSS and others.

There is also a growing body of software for general advanced machine learning such as TensorFlow for neural nets. The toolbox just gets bigger and bigger.

Yes, the toolbox is absolutely getting bigger and I, for one, am excited about that.

You have worked in both academia and industry, in places like Seimans and Glaxo. What would you say is the biggest difference between these two fields? Any similarities?

Both obviously use the same stat methods and software but naturally, industry is oriented toward promoting their products. In industry, there is a greater chance of a conflict between seeking the truth and meeting industry goals, such as getting your product approved by the FDA. In academic research, it is much easier to speak the truth to power. [I] could argue that we should not allow for-profit groups to even do medical research.

Could you elaborate further on not allowing for-profit groups to even do medical research?

This is no secret and is not just my argument. For-profit groups have an obvious stake in the outcomes and have built-in conflicts of interest. The emphasis on profit instead of improving human health is one of the ongoing scandals in our society. Millions of dollars can be at stake when, for example, a drug for an important disease such as Alzheimer's disease is tested. If the same group

Continued on the next page.

The Floor is Yours!

Interviewed by Jennifer Lee

...Continued from the previous page.

that stands to profit also runs the clinical trial or runs it with little or no oversight, we get treatments that are not effective or are not better than cheaper alternatives that work just as well. Lobbying in Congress and weakening of the FDA are part of this. This is a tragedy when we spend what is often vast amounts of taxpayer money on things that do not work well or can even be harmful. We would save money if we strengthened the oversight and then spent less on poor treatments.

What was your first job out of school?

The now-defunct BMDP statistical software company. It was purchased by SPSS which in turn was purchased by IBM.

Can you tell us more about what you did at that job?

Some of the first relatively easy-to-use statistical software (that did not require the user to write a FORTRAN or COBOL program) was developed in the 1960s in what was then the UCLA Biomathematics Dept headed by Dr. Will Dixon, using an NIH grant. Dr. Dixon's goal was to give scientists the tools they needed so they could actually get analyses done without having to be statistical or programming experts. The group that started this eventually spun this off as a private company, BMDP Inc, in the early 1980s. The BMDP company was a model from which the SAS and SPSS corporations borrowed heavily.

I haven't heard FORTRAN and COBOL in such a long time! It's bringing back memories. SPSS is still going strong in some federal government agencies, interestingly enough.

You are a Bruin through and through, from undergraduate, graduate school and now as a professor. For those who are non-Bruins, any fond memories you would like to share?

As an undergrad, I lived in the Co-op, an off-campus housing cooperative. It was a wonderful place and helped me grow as an adult. Also got to see the UCLA basketball team when John Wooden was still the coach.

John Wooden is a legend. That must have been a wonderful time. Is there any advice you would like to give to students and/or people just beginning their career in biostatistics, or any statisticians at the start of their career for that matter?

Quantitative folks such as biostatisticians and data scientists will be in demand. Use your training and talent for good things that benefit humanity, even though it may not pay as well.

Continued on the next page.

The Floor is Yours!

Interview with Jeffrey Gornbein

...Continued from the previous page.

Helping better mankind is an honorable goal and something we all should take to heart. As a biostatistician, what has been your most memorable research discovery throughout your career?

Hard to pick just one. Also, biostatisticians often do not make the discoveries, they help medical investigators make the discoveries. The models I helped make for the risk of recurrent breast cancer and premature birth (the "labor curve") are actually being used, so this is exciting (and a bit scary).

Is there any exciting research you can share with us on which you are currently working?

[I am] currently working with Dr. Jeffrey Saver on models for stroke recovery and Dr. Holly Middlekauff on the negative effects of vaping.

Last, but definitely not least, what is your favorite restaurant in the LA area and why?

Very old school, love Canter's deli in the Fairfax district. But only go occasionally since this is not very healthy.

I've not been there before and now it is on my "to try" list. Thank you, Jeffrey, for this wonderful and informative interview.

A huge thank you to Jeffrey for taking the time out of your busy schedule to talk about an important area in the field of statistics. Canter's Deli has been in business since 1931 and its menu looks absolutely wonderful. Their website is <https://www.cantersdeli.com/> and I hope everyone gets a chance to try out this establishment.

Now that summer is at its tail end, I look forward to seeing everyone at a subsequent SCASA event. Even if life gets busy, as it often does, I hope this interview inspires you to come to an event or two in the future. SCASA has lots of wonderful tools to enrich your professional life.

Do you want to be featured in a future article? Or, do you know someone in SCASA you think should be featured? Email me at SCASA.Interviews@gmail.com.

Continued on the next page.



Dr. Normalcurvesarus, Ph.D. Presents

