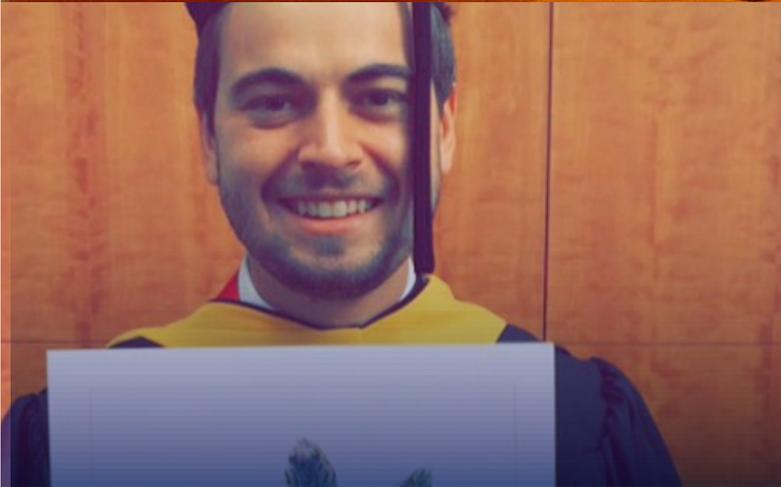


NISS

Parameters

WINTER | JANUARY 2018





NISS has taken up the challenge of living up to its name: the National Institute of Statistical Sciences. It aims to reach beyond the traditional statistical sciences to become a catalyst for maintaining statistical ideas and thinking as part of the big data and data science movement. With the ever-increasing size of data available to analyze and explore, many of the traditional tools of the statistics profession have been supplanted and replaced by new algorithms in machine learning and artificial intelligence. These new tools have brought great advances to the analysis of large data sets for prediction and decision-making. However, challenges remain for statisticians to engage as active participants in defining the foundation of data science.

Forums for industry sectors

Another new focus for NISS is to organize gatherings, both meet-ups and workshops, which bring together representatives from various sectors. For example, HR representatives from industry could meet to share common challenges and recommendations found to be effective for hiring data scientists. Another example would bring CIOs together, from a particular industry sector, such as banking or insurance, to discuss best practices for organizing teams in structuring the division.

Workshops and training short courses

NISS is expanding its offering of short courses for members of affiliated groups, individual industry sectors, government agencies, and/or national laboratories. This past year we have sponsored several R & Spark workshops, which provided training on tools for analyzing very large distributed data repositories, without the need to combine the data into one central computer center or location. This strategy for analyzing data provides considerable cost savings and privacy protection. Privacy and confidentiality protection result by keeping the source data in its original location protected by its owners' firewall. See the [R & Spark](#) article in this newsletter.

Conferences to address cutting-edge issues

NISS is planning for two conferences to bring together experts on data-enabled research and using evidence-based real-world experience to address public policy issues. The potential benefits from combining traditional survey techniques with the use of privacy protected administrative data sets is gaining support. Recent reports by the National Academy of Sciences, Engineering and Medicine (2017), Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy, Washington, DC 20001, The National Academies Press, and a report by the [bipartisan](#) Commission on Evidence-Based Policymaking, The Promise of Evidence-Based Policymaking (2017), describe in detail the potential benefits and also propose strict guidelines for implementing these ideas. As a result, a bill, [H.R. 4174](#), has already been proposed in Congress to create an agency to implement these ideas. See [Foundations for Evidence-Based Policymaking Act of 2017](#). This is truly a dramatic change from business as usual and will require strong advocacy to implement, and diligence to guarantee its implementation--while protecting privacy and confidentiality, and retaining the public's confidence that their privacy is being preserved. This [new proposal](#) about evidence-based policy will certainly generate strong debate in both the [legislative](#) and [executive](#) branches, and require neutral and independent voices to provide reconciliation for the underlying benefits and concerns. Hopefully, NISS can contribute to this debate.

Curriculum needed to create data science programs

Working closely with our academic affiliates, the ASA, and an international team, NISS plans to organize several regional affiliate meetings to address the challenges and the need for evolving the curriculum to meet the demands of Industry and Government for training data scientists. Several statistics departments have been renamed statistics and data science, e.g. [Carnegie Mellon](#) University, [University of Texas-Austin](#), and [Yale](#). Other departments of Statistics have simply created new programs in Data Science, e.g. M.S. in Data Science at [Harvard](#), a track within their MS Statistics at [Stanford](#), or an option in their MPS in Statistical Science at [Cornell](#). The Statistics profession needs to continue discussing these issues. At Penn State, an undergraduate major in data science was initiated as a joint program with three departments - Statistics, Computer Science and Information Science & Technology - in three different colleges.

ABOUT NISS

NISS is a national institute that delivers high-impact research in science and in public policy by leveraging the rich expertise of its staff with that of its base of affiliated organizations in academia, industry, and government. NISS works on issues where information and quantitative analysis are keys to solutions and decisions. NISS functions in three ways: as an expert advisor, as a basic researcher, and as a collaborator.

OUR MISSION

The National Institute of Statistical Sciences (NISS) is an independent research organization that serves as a neutral, objective expert in delivering research in science and public policy to its affiliates in academia, industry and government. NISS identifies, catalyzes and fosters high-impact cross-disciplinary and cross-sector research involving the statistical sciences.

There are many advocates for including data science and analytics in the statistical sciences driven primarily by the size and complexity of many available data sources. Knowledgeable persons have seen this evolving trend occurring long before the title data scientist became widespread. David Donoho in a presentation given at John Tukey's Memorial conference laid out the history and early origins of the importance of Data based on Tukey's work in the evolution of statistical research and practice. Leo Breiman, in an early paper nicely described the opposing perspectives of statisticians and computer scientists when analyzing data--the former fitting models to understand the process, the latter focusing on the success of prediction regardless of the underlying algorithm being used.

Post Docs working with NISS government affiliates

NISS continues its strong Post Doc program with the National Center for Education Statistics (NCES) and the National Agricultural Statistics Service (NASS). A session highlighting this ongoing research is scheduled for the [SDSS](#) meeting in Reston, VA, May 16-19, 2018. Presentations by Nell Sedransk joint with Andrew White, of NCES, will describe the program and two NISS postdocs will present their research working with NASS.

Sincerely,



James L. Rosenberger

What are Some of the Benefits of Becoming an Affiliate?

- Reserved spaces for Affiliates at each NISS conference and workshops
- Annual Affiliates Meeting held in conjunction with the National Statistical Meeting
- Affiliate Award Fund for travel expenses to NISS sponsored workshops
- Monthly Affiliate Highlights
- NISS Parameters Quarterly Newsletter highlight
- Jobs postings on the NISS website with unlimited word count

Cover Page Photos

- Top photo: Inside the R and Spark workshop room. Read the full story on [page 7](#).
- Below, left photo: Joseph Rodhouse, Research Associate, NISS, holding his diploma. Read the full story on [page 8](#).
- Below, right photo: NISS booth at the Joint Statistical Meetings (JSM) conference. Read the full story on [page 10](#).

The first virtual meet-up jointly hosted by NISS and the pharmaceutical giant Merck is an effort to introduce NISS to statisticians involved in the research and development of pharmaceutical products. This was an hour-long web meeting which took place on September 12, 2017. [The second NISS-Merck Virtual Meet-up was scheduled on Tuesday, January 23, 2018.](#)

The focus of the meet-up was to discuss the FDA draft guideline on [“Multiple Endpoints in Clinical Trials”](#) which was recently published in January 2017. The virtual meet-up was organized by distinguished researchers from the field of multiple comparisons. “This is a timely and important topic for late-stage clinical trials in support of product development,” says Dan Holder, Executive Director, Biometrics Research, Merck & Co., Inc., adding, “Multiplicity adjustments can make or break the analysis of a clinical trial, and therefore proper multiplicity adjustment is required for these trials if statistical inference is planned for multiple endpoints. Multiplicity adjustment is also necessary if the trial sponsor wishes to make a label claim of a positive drug effect on multiple endpoints.”

The meet-up was aimed to not only share ideas but also toward building connections. “We hope that these meetings foster a sense of community among statisticians who are engaged in the research and development of pharmaceutical products,” Holder adds.

Raymond Bain, Senior Vice President, Biostatistics, and Research Decision Sciences, Merck, and an elected member of the NISS Board introduced participants to NISS’s mission which is dedicated to serving the U.S. and international statistics community by engaging the NISS community in applied research driven by problems facing government and industry. “NISS is a communication enabler, co-sponsor, and coordinator of workshops and conferences,” said Bain in his discussions. While discussing the six characteristics of NISS, Bain emphasized how NISS’ role as an independent research organization could be valuable to pharmaceutical companies.

The meeting benefited from the participation of knowledgeable and influential thought leaders in the pharmaceutical industry and the Food and Drug Administration (FDA) in the United States (US). The meeting began with two presentations by notable experts on multiple comparison procedures. Alex Dmitrienko, Founder and President, Mediana Inc., discussed the Multiplicity Guidelines. Frank Bretz, Global Head, Statistical Methodology and Consulting, Novartis, discussed the implementation of the FDA Guidance. The presentations were followed by a panel discussion.

Walt Offen, Distinguished Research Fellow, Global Head of Statistical Sciences, AbbVie, moderated the discussions. Lisa LaVange, Director of the Office of Biostatistics in the Center for Drug Evaluation and Research (CDER), US FDA, joined Alex and Frank as a panelist. Lisa offered insightful comments from her role as the Biostatistics Head with CDER at the FDA.

All four participants, as well as Ray Bain, are Fellows of the American Statistical Association (ASA).

There are many technical challenges facing the pharmaceutical industry today. They include efficient trial designs for rare diseases, quantitative benefit-risk analysis, and the use of real-world evidence for comprehensive safety assessment and comparative effectiveness. Christy Chuang-Stein, Independent Statistical Consultant, who is also an elected member of the NISS Board, says, “There are other more strategic questions that a pharmaceutical sponsor should ask of themselves. These questions include increasing the success rate of product development and improving the efficiency in selecting product candidates for further development. NISS’s ability to act as a communication enabler, co-sponsor and coordinator of workshops and conferences, and as an independent research organization could be helpful to pharma companies in tackling these challenges.”

“NISS has access to skills and resources to catalyze cross-functional research that could be helpful to a variety of industries. Besides providing information and discussion on important topics, the meet-ups were conceived to increase NISS’s visibility among industries that might find their resources useful,” says, Holder.

In all, more than 190 people attended this virtual meet-up. NISS and Merck together plan to run similar virtual meet-ups once every quarter, covering varied topics such as machine learning and adaptive treatment strategies in clinical trials. The next meet-up is currently targeted for January 2018 with a theme on the recent Addendum (R1) to ICH E9 (Statistical Principles for Clinical Trials). This Addendum discusses how to handle the missing data problem in clinical trials. We plan to announce the details of this second meet-up in the near future.



“NISS’s ability to act as a communication enabler, co-sponsor and coordinator of workshops and conferences, and as an independent research organization could be helpful to pharma companies in tackling these challenges,” **CHRISTY CHUANG-STEIN**, Independent Statistical Consultant, who is also an elected member of the NISS Board



“NISS has access to skills and resources to catalyze cross-functional research that could be helpful to a variety of industries,” **DAN HOLDER**, Executive Director, Biometrics Research, Merck & Co.

WESTERN MICHIGAN UNIVERSITY'S DEPARTMENT OF STATISTICS BECOMES A NISS AFFILIATE

NISS



The Western Michigan University's Department of Statistics has become an Affiliate of NISS, effective November 2017.

"We are very happy to become a new academic affiliate of NISS. Inclusion in NISS' Affiliate program will open up new developmental and collaboration opportunities, especially for junior faculty. An individual [Affiliate Award Fund](#) is a great benefit. Being included in the NISS' Affiliate program should also increase the department's visibility," says, Kevin Lee, Assistant Professor, Department of Statistics, Western Michigan University. Lee is the Junior NISS Affiliate Liaison.

The Western Michigan University's Department of Statistics offers M.S. and Ph.D. programs in Statistics. Expanding its curriculum to meet industry demands, the department has recently started offering Bachelor's and Master's degree programs in Data Science. "Curricula of the programs offered by the department were modernized to strengthen students' computing skills. Our recent faculty hires have expertise in data mining, network analysis, and data visualization," says, Magdalena Niewiadomska-Bugaj, Department Chair, Department of Statistics, Western Michigan University. Niewiadomska-Bugaj is also a NISS Affiliate Primary Liaison.

The NISS Affiliate Program is a communication enabler for the statistical and data science community. "We are pleased Western Michigan University's Department of Statistics is the latest department to join our academic affiliate program" says, James Rosenberger, NISS Director. "The NISS Affiliate program brings together statistical and data science professionals from all sectors - academia, industry, government/national labs - to support research discussions, information dissemination, human resource development, and networking," Rosenberger adds.

Catering to the growing demands of Data Science

The world is becoming more and more quantitative and data driven. Organizations are making decisions based on analyzing large data sets. Hence there is an ever-increasing need to understand data and the ability to draw an insight from that data is even more important. "The department of statistics at Western Michigan University is fully aware of these changes and adapts the programs' curricula. We monitor what skills employers list in job announcements and where our graduates are hired. We prepare our students for careers in academia, industry and government by teaching the skills that are most useful and we also provide foundations which allow our

students to acquire new skills and learn new developing methodologies," says Niewiadomska-Bugaj.

Statistics is a rapidly developing discipline. The department of statistics at Western Michigan University is catering to these evolving needs. The research focus of the department is slightly shifting toward new problems from working with big data, such as data mining, data visualization and network analysis. "We want to be at its forefront," says, Niewiadomska-Bugaj. She adds, "We envision the department to continue preparing students for successful careers in academia, industry or government. While providing necessary mathematical rigor, our programs and research projects will respond and adapt to the specific needs of a quickly changing world."

With this NISS affiliation, the department of statistics at Western Michigan University looks forward to enhancing the chance of multi-disciplinary research collaboration, help with finding the complex yet interesting real-world problems, and promote developing new statistical methodologies to address these problems and challenges.

"As an affiliate of this national statistical organization, we expect to gain more interest in our programs. Also, as our department occasionally host conferences, for example, the [International Conference on Robust Rank-Based and Nonparametric Methods in 2015](#), we hope that in the near future NISS could co-sponsor a similar conference," says Niewiadomska-Bugaj.

Research efforts

The department of statistics at the Western Michigan University is strong in rank-based inference. Recently Professor [Joseph McKean](#) of the department coauthored a textbook titled "Nonparametric Statistical Methods Using R" where some of the robust methods are presented at an applied level.

Some of the research work that the university is working on include diagnostics for comparative efficiency of the log-rank test versus Wilcoxon; a regression approach to testing for bioequivalence of two measurements; statistical properties of the population stability index; extensions of the Bonferroni approach to multiple comparisons; algorithms for robust analyses of massive data sets; methodological research on zero-inflated data; and biological applications of spatial statistics.



"Curricula of the programs offered by the department were modernized to strengthen students' computing skills. Our recent faculty hires have expertise in data mining, network analysis, and data visualization." - [MAGDALENA NIEWIADOMSKA-BUGAJ](#), Department Chair, Department of Statistics, Western Michigan University



"We are pleased Western Michigan University's Department of Statistics is the latest department to join our academic affiliate program." - [JAMES ROSENBERGER](#), NISS Director.



"Inclusion in NISS' Affiliate program will open up new developmental and collaboration opportunities, especially for junior faculty." - [KEVIN LEE](#), Assistant Professor, Department of Statistics, Western Michigan University



"It wasn't until I took mathematical statistics in grad school and learned the underpinnings of all the methodology that I really started to enjoy statistics." - MICHAEL ROBBINS, Former NISS Postdoc

Born and raised in Greenville, South Carolina, Michael Robbins was always interested in quantitative research. He went to high school at JL Mann Academy in Greenville, South Carolina. He earned his bachelor's degree in Mathematics and Economics from Duke University in Durham, North Carolina, and then went on to earn his master's degree in Mathematical Sciences from Clemson University. Robbins also pursued a Ph.D. degree in Mathematical Sciences from Clemson University and received his degree in 2009. Immediately following his Ph.D., he joined NISS as a Postdoctoral fellow from 2009 to 2011.

Robbins always had a passion for statistics. He enjoyed probability theory and has always been quantitatively inclined. "Probability theory was something that clicked for me in high school and college. I took some statistics in high school and college, but found it rather enigmatic," says Robbins. He goes on to say, "You were taught how to perform an F-test, use a T-table, etc., with no idea of how or why any of it works. It wasn't until I took mathematical statistics in grad school and learned the underpinnings of all the methodology that I really started to enjoy statistics."

Robbins owes his successes to his mentors who he says played an important role guiding him all through his career. His Ph.D. advisor from Clemson University, Colin Gallagher, was proud to see Robbins transform from a student to an independent researcher while at Clemson. "Michael was great to work with," says Gallagher. Sharing his experiences working with Robbins, Gallagher says that Robbins showed up for research meetings with a candy bar and a soda, and he was always ready to dig in and talk mathematics. He enjoyed going to the whiteboard and deriving things on the spot. He liked to prove and explain the methodologies and the theoretical results, and most importantly, he constantly improved himself as a researcher. "If he didn't understand or know something he would go study it until he mastered it, then come back and explain it in a research meeting," says Gallagher.

Gallagher guided his mentees on how to formulate and attack a research problem, how to write a research paper, and how to give a research talk. As Robbins' mentor for four years, Gallagher remembers listening to Robbins practice

research talks and guiding him through the process.

"Initially we had a mentor-mentee relationship but by the end, I was able to treat Michael as a collaborator rather than a student researcher," says Gallagher.

As a Postdoc at NISS, Robbins developed an imputation algorithm for the USDA's Agricultural Resource Management Survey (ARMS). ARMS are a high-dimensional survey of more than 1000-2000 questions asked to 30,000-40,000 respondents who in this project were farmers from Iowa, Illinois, and other agricultural producing states. An imputation method is used to address the missing values because respondents don't usually answer all questions in the survey. This project titled, "Imputation in High-Dimensional Economic Data as Applied to the Agricultural Resource Management Survey" was administered by NISS and the National Agricultural Statistical Service (NASS). "It means that I have multiple numbers of items in the questionnaire that have multiple types of missing values and I want to impute those missing values using statistical methodology... that's exactly what we developed," says Dr. Sujit Ghosh, Professor at NC State University, Department of Statistics who mentored Robbins on this project.

Sharing his experience working with Robbins, Dr. Ghosh says, "Michael has the capability of taking on new projects. When he joined the program, he was not an expert on the methodology of imputation. He didn't have that background, but he worked hard, picked up the literature, read it thoroughly, and not only that, he came up and shared some creative research ideas too. He has the capability of picking up a topic and working on it, even if he has no background in the subject."

Speaking about the challenges of moving from school to a real-job at NISS, Robbins says, "You're accustomed to working under heavy guidance in school and now have to adapt to working largely on your own. The biggest adjustment for me from after my postdoc experience was learning to juggle many research projects simultaneously."

Prior to Robbins work at NISS, another Ph.D. advisor from Clemson University, Robert Lund, was happy to mentor Robbins for more than four years. All Lund had to do was give Robbins a technical problem to challenge him. Robbins would come back with the answer to the problem the

HIGHLIGHTS

EQUIPPING HIGH SCHOOL TEACHERS WITH CURRICULA REQUIRED TO TEACH STATISTICS AND DATA SCIENCE

NISS and the American Statistical Association (ASA) have joined hands to sponsor an international team of statisticians and computer scientists in 2018 to draft curricula for teaching statistics and data science to high school students. This team will have representatives from the USA, New Zealand, Australia, and Canada. They need \$10,000 to fund their first meeting, and NISS would like to provide assistance with your help.

Let's prepare our teachers! You can help make this happen.

MAKE A DONATION.

very next day. "He's a great abstract thinker and an astute student capable of so much," says Lund.

Robbins is now a statistician at RAND Corporation and has been extensively working on surveys. He was instrumental in building RAND's American Education Panels which are survey panels of over 30,000 educators. Robbins also uses his broad statistical toolkit and applies it to different areas such as health, education, criminology, military, defense, including others.



Robbins is working on an R package called "Microsynth" in collaboration with Steve Davenport, Assistant Policy Researcher at RAND Corporation. Davenport is also a doctoral candidate at the Pardee RAND Graduate School and Robbins is mentoring Davenport on this project. Microsynth is a product of a RAND Corporation project that aims to improve the matching of treatments to synthetic controls and to construct better estimators for the variance within synthetic control methods.

The synthetic control method is a statistical method used to evaluate the effect of an intervention in comparative case studies. There are many research studies that have applied the synthetic control methods, and there are packages already available in languages such as R that facilitate the use of synthetic controls. But the existing packages have built-in limitations that were appropriate in a relatively data-scarce environment, with a small number of unit-observations.

With the arrival of Big Data, "Microsynth is built to better take advantage of big data and counter these limitations," says Davenport, adding, "Microsynth exploits an abundance of data to improve the matching of different treatments and synthetic controls. Additionally, it also enhances the method by adapting the algorithms for use in data-rich environments, so that research can be conducted on many micro-level observational units."

Speaking about working with Robbins on the Microsynth project, Davenport says, "Michael has been invaluable as a mentor and partner: tremendously responsive and knowledgeable. That he has developed such a methodology is impressive and makes the rest of the project much easier."

UPCOMING EVENTS

NISS-MERCK VIRTUAL MEET-UP

WHEN: TUESDAY, JANUARY 23, 2018 - 11:00AM TO 12:00PM

WHERE: VIRTUAL MEET-UP

[MORE EVENT INFORMATION](#)

2018 CONFERENCE ON STATISTICAL PRACTICE (CSP)

WHEN: THURSDAY, FEBRUARY 15, 2018 - 8:00AM TO SATURDAY, FEBRUARY 17, 2018 - 5:30PM

WHERE: MARRIOTT PORTLAND DOWNTOWN WATERFRONT, 1401 SW NAITO PARKWAY, PORTLAND, OREGON, 97201, UNITED STATES

[MORE EVENT INFORMATION](#)

R & SPARK: TOOLS FOR DATA SCIENCE WORKFLOWS

WHEN: THURSDAY, FEBRUARY 22, 2018 - 9:00AM TO FRIDAY, FEBRUARY 23, 2018 - 5:00PM

WHERE: H2O HEADQUARTERS, 2301 LEGHORN ST., MOUNTAIN VIEW, CALIFORNIA, UNITED STATES

[MORE EVENT INFORMATION](#)

THE UNIVERSITY OF GEORGIA/ CLEMSON UNIVERSITY JOINT SEMINAR

WHEN: THURSDAY, MARCH 29, 2018 - 4:30PM TO 6:00PM

WHERE: BROOKS HALL ATHENS, GEORGIA, 30602 UNITED STATES

[MORE EVENT INFORMATION](#)

NISS/CANSSI WORKSHOP ON R & SPARK: TOOLS FOR DATA SCIENCE WORKFLOWS

WHEN: THURSDAY, APRIL 12, 2018 - 9:00AM TO FRIDAY, APRIL 13, 2018 - 5:00PM

WHERE: ROOM 212 AT VICTORIA COLLEGE, 73 QUEEN'S PARK CRESENT E, TORONTO, ON M5S 2C3, CANADA

[MORE EVENT INFORMATION](#)

SIXTH NOGGINS 2018 WORKSHOP

WHEN: FRIDAY, APRIL 13, 2018 - 9:00AM TO 5:00PM

WHERE: UNIVERSITY OF GEORGIA, RICHARD B. RUSSELL BUILDING SPECIAL COLLECTIONS, LIBRARIES, ROOM 277, ATHENS, GEORGIA, 30602, UNITED STATES

[MORE EVENT INFORMATION](#)

SYMPOSIUM ON DATA SCIENCE & STATISTICS (SDSS)

WHEN: WEDNESDAY, MAY 16, 2018 - 7:00AM TO SATURDAY, MAY 19, 2018 - 3:30PM

WHERE: HYATT REGENCY RESTON, 1800 PRESIDENTS STREET, RESTON, VIRGINIA, 20190, UNITED STATES

[MORE EVENT INFORMATION](#)

R & SPARK: TOOLS FOR DATA SCIENCE WORKFLOWS

WHEN: WEDNESDAY, MAY 30, 2018 - 9:00AM TO THURSDAY, MAY 31, 2018 - 5:00PM

WHERE: BUREAU OF LABOR STATISTICS CONFERENCE AND TRAINING CENTER. CONFERENCE ROOMS 1-2, POSTAL SQUARE BUILDING, 1ST STREET, NE, WASHINGTON, DISTRICT OF COLUMBIA, 20212-0001 UNITED STATES

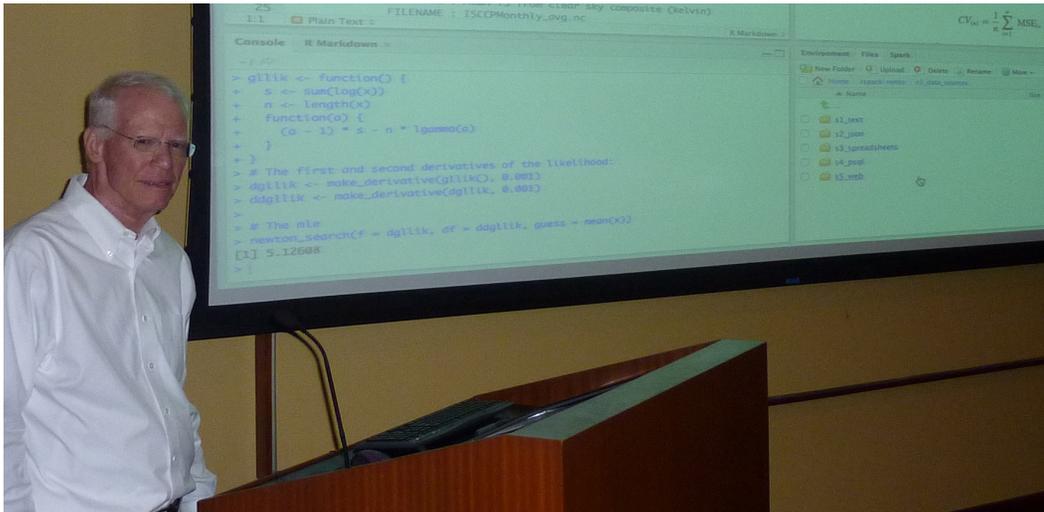
[MORE EVENT INFORMATION](#)

2018 JOINT STATISTICAL MEETING

WHEN: SATURDAY, JULY 28, 2018 - 7:30AM TO THURSDAY, AUGUST 2, 2018 - 12:20PM

WHERE: VANCOUVER CONVENTION CENTER, 999 CANADA PL, VANCOUVER, BC V6C 3C1, CANADA.

[MORE EVENT INFORMATION](#)



E. JAMES HARNER, Data Scientist and R and Spark Workshop Instructor

The two-day “R and Spark: Tools for Data Science Workflows” course was introduced by NISS in September 2017. The course was delivered by data science expert E. James Harner, Professor Emeritus of Statistics and Adjunct Professor of Management Information Systems at West Virginia University. The first course was held at the American Statistical Association in Alexandria, Virginia, and the second one at the University of California, Riverside campus. Upcoming R and Spark courses are scheduled at the H2O Headquarters in Mountain View, California, on Feb 22-23, 2018; in Toronto, Canada on April 12-13, 2018; and in Washington, DC on May 30-31, 2018.

Both R and Spark big data environments primarily work on rectangular data structures called data frames. R does analyses on a single computer core, whereas Spark distributes its computation over hundreds or thousands of nodes. “As a result, Spark’s computational load and storage capacity are both very large, but the applicable algorithms are limited. On the other hand, R has a huge number of useful algorithms, but its storage capacity and computing resources are limited,” says, Harner, as he speaks about how the R and Spark data frameworks complement each other.

The R and Spark course held at the University of California, Riverside campus, although open to the public, was attended by many students, since the course was priced at a discounted rate to make it affordable for students. In addition, UCR’s faculty partly supported the student registration fees for UCR students. “The course is beneficial to multiple audiences,” says Dan Jeske, Professor of Statistics at the University of California, Riverside. He adds, “The course is valuable to statisticians and data scientists trying to understand the distinction between statistics and data science; to students

about to enter the workforce; and to employers who want to encourage their in-house data analysts toward a valuable form of continuing education.” Additionally, this course is valuable to people from a non-statistics background, especially those in a substantive area that uses statistics. The only pre-requisite is that everyone attending the course should have some statistical training and basic knowledge of R.

Big Data and Data Science in statistics

There’s a lot of emphases and talk around Big Data and Data Science in the statistical world today. While Big Data is a catch phrase for any type of data - sensors, images, videos, natural language - Data Science is the study of Big Data workflows. Explaining the life cycle of extracted or captured data, Harner says, “First data is extracted from a source. It is then cleaned and transformed into a structured form (perhaps with preliminary analyses) before various machine learning algorithms are applied. The workflow proceeds to graphics and reproducible documents and perhaps to the creation of a data product before cycling back to the source for more data.”

Takeaway from the R and Spark course

The course presents a new paradigm of thinking about the analysis of data. The emphasis is on building workflows from data sources to data products with the use of prediction metrics rather than p-values (probability values) as the measure of success.

“The course helps participants, especially students, to make a shift in the way they think about data,” says Harner. Giving an example, Harner says, “Participants can start thinking from a procedural to a functional style of programming,

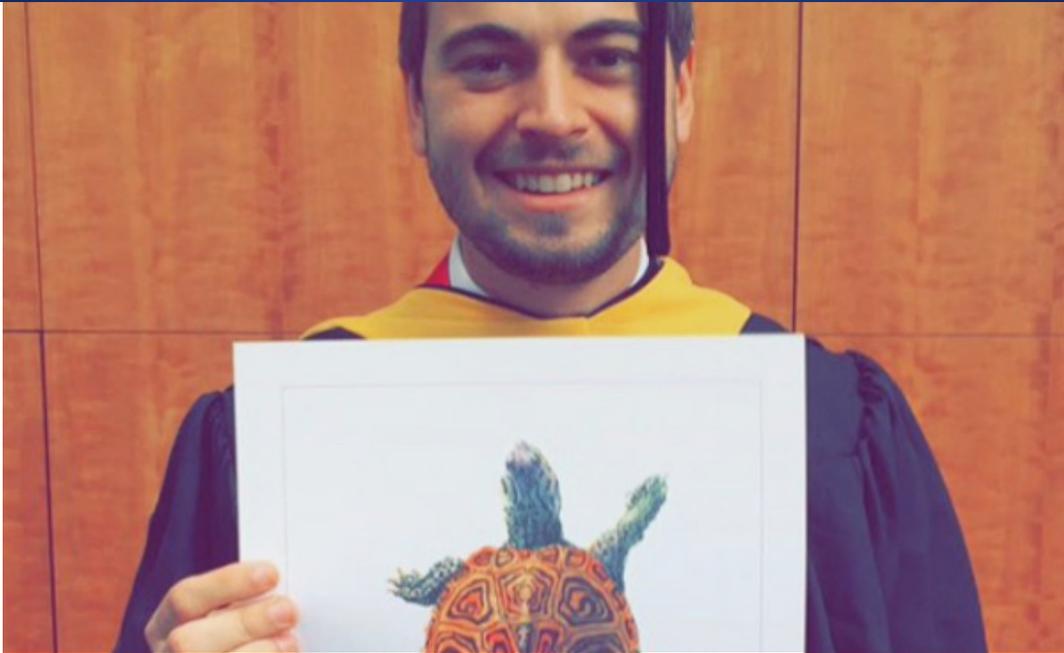
from a single file system to distributed storage models, and from a distributed computational model brought to the data rather than moving big data to the computation.”

The course is taught in a virtual environment called rspark. The rspark environment contains R, Hadoop, Spark, a relational database and many other data technologies, which can be deployed locally on a student’s computer or on Amazon’s Web Services cloud, so that participants get a hands-on training on it. “In fact, rspark was built to teach courses like this one,” says Harner, adding, “Thus enabling students to have a learning environment that operates identically to the actual large-scale computer clusters.”



“The course is valuable to statisticians and data scientists trying to understand the distinction between statistics and data science; to students about to enter the workforce; and to employers who want to encourage their in-house data analysts toward a valuable form of continuing education,” says **DAN JESKE** (right), Professor of Statistics at the University of California, Riverside, seen here with **MADELINE BAUER**, USC Statistician/Data Scientist (Retired).





“There is a wealth of research on the factors impacting survey response. However, much is still to be known and I hope to contribute scientifically to this body of knowledge.”- **JOSEPH RODHOUSE**, Research Associate, NISS, seen here holding his diploma.

[Joseph Rodhouse](#) joined NISS in October 2017 as a Research Associate and is working on the National Agricultural Statistics Service (NASS) project. The work involves investigating survey methodological phenomena, such as respondent burden and response likelihood to surveys. Declining response rates are a trend impacting the survey research field as a whole and so he is focusing on factors that could lead to increasing responses to surveys.

The goal of the project that Rodhouse is working on is to gain a better understanding of how a person’s likelihood to respond is influenced by characteristics of the survey. Most research has shown that among various other factors that contribute to respondent burden, the burden caused by the complexity of the survey and the number of surveys a respondent has received contribute the most. This study examines the effect of subsequent sampling on high and low burden surveys with respect to response outcomes. “The hope is that this information can be used to tailor future sampling strategies that could lead to improving response rates,” says Joseph Rodhouse. This study is scheduled to be presented at the Federal Committee on Statistical Methodology (FCSM) Conference in March 2018.

Additionally, Rodhouse is working on two other NISS-NASS projects:

1. In the first one, he is preparing an internal research report on the use of paradata in web surveys. Survey paradata usually consists of administrative data about the conduct of the survey, which includes characteristics such as the time of the interview, duration of the interview, number of contacts with the interviewee, the mode of communication, etc. This paradata report will help NASS statisticians investigate and improve data quality in their web surveys.
2. On the second NASS project, Rodhouse is working with a group of statisticians doing behavior coding and cognitive interviewing. The purpose of this study is to identify where in the survey interview process the errors are occurring.

Sources of errors in the process that are found to be nonrandom could then be prevented or corrected either in the survey design stage or in the data analysis stage.

Rodhouse enjoys his work on the NISS-NASS projects, especially the opportunities for meeting with other survey statisticians. “These projects have not only given me an opportunity to meet other survey statisticians and discuss with them about the future of survey methods but also given me an opportunity to meet and interview farmers face-to-face,” Rodhouse says.

Heather Ridolfo, Lead on the behavior coding and cognitive testing projects at NASS says, “I really like working with Joe. It’s fun to have someone new with a fresh perspective. Coming out of grad school he is up to speed with the latest advancements in the field, so he brings some great insight to our team.”

Ridolfo has assigned Rodhouse to projects where he works closely with other members of the staff and to projects where he works independently. She says, “Joe’s greatest strengths are his inquisitiveness and his thoughtfulness. I know if I give Joe a new project, he will spend the time researching the method, and reaching out to others when assistance is needed.”

Before joining NISS, Rodhouse was a JPSM Fellow. The JPSM ([Joint Program in Survey Methodology](#)) Fellow Program is a unique internship experience with a mission to educate the next generation of survey researchers, survey statisticians and survey methodologists. As a JSPM Fellow, Rodhouse worked on an experimental research project for the Census Bureau testing varying questions designed to capture the prevalence of “gig work” in the Current Population Survey. Gig workers who take on short-term jobs, some of them as a self-employment contractor, are usually not identified in surveys of employment. This research project was designed to capture data on workers in the gig economy. “The project was very interesting yielding some very interesting results,” says Rodhouse. Adding, “We found that the prevalence of gig work is perhaps not ignorable and that the specific

HIGHLIGHTS

NISS SLOGAN CONTEST: SUBMIT YOUR ENTRY FOR A CHANCE TO WIN

The National Institute of Statistical Sciences (NISS) is pleased to announce a competition to create a slogan/tagline that embodies the mission and purpose of NISS.

The competition starts today. Keeping the NISS Mission in mind, we invite you to send us your submissions by 5 p.m. on Wednesday, January 31, 2018. The winning slogan will be part of the NISS logo and all organization marketing and promotional material.

[CLICK HERE TO DOWNLOAD CONTEST INSTRUCTIONS](#)

[CLICK HERE TO SUBMIT YOUR ENTRY](#)



JOSEPH RODHOUSE graduates from JPSM (May 2017), seen here shaking hands with **FRAUKE KREUTE**, Director of JPSM.

question type yielded more reports of gig work.” He also worked on a project at JPSM analyzing whether questions designed to assess the cognitive decline in the Health and Retirement Study (HRS) were reliable measures for non-native English speakers.

Rodhouse earned his bachelor’s degree in Sociology from Whitman College in Walla Walla, Washington. As an undergrad, he developed an interest in quantitative social research which sparked his desire to pursue a master’s in Survey Methodology from the University of Maryland. While pursuing his undergrad degree, Rodhouse gained skills on topics such as web survey data collection methods, questionnaire design, and statistical analysis of survey data, by working on an independent study to conduct a survey for Whitman’s Alumni Office. He further strengthened those skills in the five years between undergrad and grad school by working as a survey research professional at a prominent market research firm. He says, “I was first galvanized to pursue statistics as an undergrad while learning about the importance of quantitative social research to our understanding of society. I found it very inspiring that statistics could be so illuminating.”

In order for societies to effectively and efficiently function, they need facts. Surveys and statistics are one important area that helps us understand what the facts are. Giving an example from the agricultural sector Rodhouse says: “We need to know the quantity of our agricultural production so we know how much food we have available in the country. This type of information impacts a lot of things, like the prices we pay for our food or how much we have available to send in aid to places in need worldwide. However, finding this information out is very difficult. It’s simply not possible to take a census of every agricultural producer whenever we need to know what the facts are. As a result, we rely on sample surveys and their statistical estimates to provide us with the best information possible. If it weren’t for statistics, determining agricultural production facts in the scope that we need them would be extraordinarily difficult.”

HIGHLIGHTS

The NISS Corporation Member Dr. Nicholas Horton received the 2017 Lagakos Distinguished Alumni Award. Dr. Horton graduated from the Department of Biostatistics in 1999 and is currently a Professor of Statistics at Amherst College. The award was presented to Dr. Horton on October 19, 2017. [READ THE FULL STORY HERE.](#)

NISS Corporation Member and Professor of Statistics at Amherst College, Dr. Nicholas Horton, has been named a fellow of the prestigious American Association for the Advancement of Science (AAAS). Horton was recognized for his “distinguished research contributions to statistical and data sciences, creativity in statistics education and professional service developing curriculum guidelines for statistical education and computing.” Horton is one among only 395 other scientists who have won the AAAS fellowship. The tradition of awarding AAAS Fellowships began in 1874. [THE FULL STORY HERE.](#)



NISS 2018 JEROME SACKS AWARD CALL FOR NOMINATIONS

SAVE THE DATE JOIN NISS AT JSM 2018 IN VANCOUVER, CANADA

Nominations are being sought for the 2018 National Institute of Statistical Sciences' (NISS) Jerome Sacks Award for Outstanding Cross-Disciplinary Research. The prize recognizes sustained, high-quality, cross-disciplinary research involving the statistical and data sciences.

An award of \$1,000 will be presented during the National Institute of Statistical Sciences reception at the Joint Statistical Meetings (JSM) in Vancouver, July 28 – August 2, 2018.

For more information, including a list of previous award winners, click [Jerome Sacks Award for Outstanding Cross-Disciplinary Research](#)

Procedure

To nominate an individual, submit as one PDF document the following information to officeadmin@niss.org by April 1, 2018:

- Nomination letter (maximum two pages)
- Supporting letters from two individuals (other than nominator)
- The Nominee's CV

Send questions or comments to officeadmin@niss.org

The theme of the 2018 Joint Statistical Meeting (JSM) hosted by the American Statistical Association is #LeadWithStatistics. The conference will cover topics ranging from statistical applications to methodology and theory to the expanding boundaries of statistics, such as analytics and data science. The 2018 JSM Conference will be held at the Vancouver Convention Centre from July 28 to August 2, 2018.

NISS has organized two invited sessions at the conference. The first session is the Jerry Sacks Award session focused on Genomics and the second session is on Big Data. Four distinguished speakers will be speaking at each of these two sessions. An additional topic-contributed session is also proposed in which NISS current and former postdocs will be speaking about their work. More information about these sessions will be available very soon.

"We will be showcasing our expansion plans at the 2018 JSM," says James Rosenberger, NISS Director. He adds, "NISS is aiming to reach beyond the traditional statistical sciences to become the catalyst for maintaining statistical ideas and thinking as part of the big data and data science movement."

NISS will also have a booth at the JSM 2018. The booth is an opportunity for JSM visitors to meet NISS Board members and senior Affiliate Liaisons and also chat with Junior Liaisons from NISS Affiliates to learn about NISS activities. Also, it is a great opportunity to learn how to become a NISS Affiliate and to learn about the NISS Postdoctoral Program. "Please stop by and visit us at our booth to express your preference for new opportunities and/or locations for NISS activities," Rosenberger adds.

The NISS Annual Reception is traditionally held at the JSM Conference. Each year, the Jerome Sacks Award for Cross-Disciplinary Research is presented at the JSM NISS Reception. This award honors Jerry Sacks, who was the first Director of NISS and a distinguished statistician who made significant contributions to the Statistics bridge theory, methodology and collaborative science and engineering. Sacks Award [Nominations](#) can still be made until April 1. More information about the event will be announced on our [JSM event page](#).

NISS

www.NISS.org

For more information about the National Institute of Statistical Sciences,

CONTACT:

1150 CONNECTICUT AVENUE NW, 9TH FLOOR, WASHINGTON, DC 20036

PHONE: (202) 862-4316 | EMAIL: COMMUNICATIONS@NISS.ORG