

Cleveland State University

Department of Mathematics and Statistics

Independent Study: STA 597- Second-generation p-values and their Use in Statistical Practice

To be filled out by instructor and student. Submit electronic copy to mathematics@csuohio.edu. For projects that will take place in a semester for which enrollment is not yet open, please wait to submit form until enrollment has been opened for the academic year containing that semester.

Student Name _____
Student Email _____
Faculty Advisor **Dr. Linda Quinn**

CSU ID _____
Term/Year **FALL 2022**
Credit Hours # **1**

In each category below, be as precise as possible while recognizing the tentative nature of some decisions.

Pre-requisites: A course in inferential statistics or enrolled in tApplied Statistics Specialization

Topic of Study: Second-generation p-values and their Use in Statistical Practice

Jeffrey D. Blume and Megan H. Murray will be presenting a full day workshop on this topic on **Friday, October 14th from approximately 8:30 – 4:00 at the Holiday Inn in Independence** as part of the annual Fall Workshop of the Cleveland Chapter of the American Statistical Association. **Attendance is limited at the workshop. If you do not register in time, you will need to switch to another one credit course.**

This is an excellent opportunity for graduate students to potentially meet students from other colleges, socialize with local professional statisticians (typically from Lubrizol, Cleveland Clinic, Progressive, Goodyear, American Greetings, Key Bank, etc.), and hear distinguished speakers on a very relevant topic. The conference is aimed at the Master's level statistician. **It will involve a small separate registration fee of \$50 that includes breakfast, lunch, and breaks. Payment is required to complete your registration for this course.** Payment can be made to Dr. Quinn (checks should be made out to the Cleveland ASA). A PayPal link will be available as well.

Additionally, this class will have discussion and presentation forum on **two Thursdays (from 2:00 to 3:00pm)**. The first **Thursday, October 20th** will be to review the material from the conference and to discuss a short project to be completed for the last meeting. The last meeting, **Thursday, December 1st**, will have students presenting a brief report of their project. Projects are individual but may be coordinated. More details on October 20th.

Readings to be covered: Full attendance at the conference and readings in preparation for the conference

Papers to be written: One project presentation using skills from course.

Will there be an examination? No

Approximate number of meetings planned: 3 (2 Thursdays and 1 all day Workshop)

Basis for assignment of final grade: Entire day conference attendance is mandatory (70%). Attendance and active participation at Thursday sessions (15%). Presentation (15%).

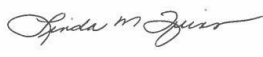
Grade: A= [96-100%] A- =[92-96%] B+ =[88-92%] B =[84–88%] B- =[80-84%] C = [70–80%] F= [0-70%]

I have discussed my proposal with the faculty member named above. I agree to all requirements and deadlines for this Independent Study.

Student signature _____

Date _____

I agree to supervise the student listed above. We have discussed the proposal, requirements, and deadlines for this Independent Study.

Faculty signature 

Date: **8/3/2022**

Department Use Only:

Course Number: **STA597**

Section: **1**

Class Number: **6290**

Description:

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.

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 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, observations 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, PhD 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 in the analysis of a wide array of biomedical research. He has won numerous awards for his teaching and mentoring, and for 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 PhD candidate in Biostatistics at Vanderbilt University working with Professor Blume. Her methods research focuses 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 a R package ("FDRestimation") for computing, estimating, and visualizing false discovery rates (currently on CRAN). Her biomedical research has focused in 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 shorter time to graduation. Megan has won several teaching awards in both Biostatistics and Data Science at Vanderbilt.