IOWA | Department of Biostatistics

Iowa Summer Institute in Biostatistics

2023

14th Annual
Iowa Summer Research Symposium

University of Iowa
Iowa City, Iowa

July 20, 2023

C217-CPHB
College of Public Health

Table of Contents

2023 Iowa Summer Institute in Biostatistics Presentations

Predicting Outcomes in a Population in Early Recovery from Alcohol Use Disorder …….. 1

Predictive Modeling for Body Fat Percentage Based on Anthropometric Measures 2

Predicting Time Until Failure Renal Failure for Newly Diagnosed C3G Patients .. 3

Why is learning so often difficult to achieve? ……. 4

Comparing Exposure Pathways of Enteric Pathogens to Infants Living in Low – to- Middle – Income Countries ……..……. 5

Statistical Models and Disparity Measures of Annual Incomes in the US …… 6

Piper Dean
Carleton College
Kelsey Mulligan
Wheaton College
Julian Soto
University of Puerto Rico, Mayaguez

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Predicting Outcomes in a Population in Early Recovery from Alcohol Use Disorder

In this project, we will focus on a cohort of survey respondents who identify as being in recovery from Alcohol Use Disorder (AUD). The survey collected data on a variety of variables including demographics, mental health, and use history. We will attempt to build Machine Learning models to evaluate recovery status from AUD. Specifically, the models will focus on predicting changes in recovery goals. The Machine Learning models used will include Single Decision Trees, Random Forests and Neural Networks. All the models will be evaluated using cross validation and then assessed for predictive power. We will then compare these models to Bayesian multinomial regression to quantify evidence for the association between the primary drivers of drinking outcomes and relapse.

Mentors of Research Group

Dr. Grant Brown, Associate Professor, Dept. of Biostatistics, University of Iowa

Dr. Paul Gilbert, Associate Professor, Project PI, Dept. of Community and Behavioral Health, University of Iowa

Mitira Schwab
College of William and Mary

Nadine Zayyad
California State University Channel Islands

Sam Ingram
Arizona State University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Predictive Modeling for Body Fat Percentage Based on Anthropometric Measures

The percentage of an individual's body fat is an important physiologic measure used to characterize fitness and health. However, accurate assessments of body fat percentages can be difficult and expensive to obtain. The dual-energy x-ray absorptiometry (DEXA) scan is considered the “gold standard,” and yet this procedure must be done in a laboratory setting. The purpose of this project is to develop models for predicting the percentage of body fat in males based on easily obtained anthropometric measures, such as weight and height, and circumferences for neck, chest, waist, hip, thigh, knee, ankle, bicep, forearm, and wrist. One of the challenges in developing such models is that anthropometric measures are inherently collinear, which can result in highly inaccurate model parameter estimates. To formulate our models, we consider approaches based on best-subsets variable selection guided by information criteria. Predictive efficacy will be investigated using split-sample validation. The models will be based on a dataset comprised of 250 records on male subjects who have had their percentage of body fat accurately obtained. The data was collected at the BYU Human Performance Research Center.

Mentor of Research Group
Dr. Joseph Cavanaugh, Professor, Department of Biostatistics, University of Iowa

Ashley Birnesser
Lake Forest College

Zoe Chafouleas
Dartmouth College

Bryan Salcedo
Kean University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Predicting Time Until Failure Renal Failure for Newly Diagnosed C3G Patients

Complement-c3 glomerulopathy (C3G) is a group of rare conditions that cause kidney disease through abnormal activation of the autoimmune system. The immune system of C3G patients lacks control of the protein C3 - a protein on the surface of red blood cells - resulting in a constitutive inflammatory response that causes damage to patient kidneys over time. The focus of this research is to develop a model to predict time until renal failure in newly diagnosed C3G patients using various measures of kidney function at the time of diagnosis. Our methods of modeling include Cox proportional hazards, Kaplan-Meier curves, and decision trees. These models identify predictor variables for the risk of reaching ESRD using data collected from a cohort of C3G patients by the University of Iowa. Methodology for predicting time until renal failure will be accomplished using a statistical survivorship function and decision tree based on the initial measurements of these variables in C3G patients.

Mentors of Research Group
Dr. Patrick Breheny, Associate Professor, Department of Biostatistics, University of Iowa
Logan Harris, Graduate Student, Department of Biostatistics, University of Iowa

Hannah Mullins
Wheaton College

Iramir Neto
Kean University

Andrés Vazquez
University of Puerto Rico-Mayagüez

Rachel Watson
Bradley University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Why is learning so often difficult to achieve?

A typical college student today has access to learning resources, but they may also face adverse learning experiences that lead to discomfort during learning, through challenged beliefs or failures. This discomfort causes some students to ruminate on negative thoughts and emotions which may lead to distal behaviors, such as dropping classes or accepting defeat. Researchers from the Department of Neuropsychology at the University of Iowa conducted a pilot study to determine the role that biofeedback plays in sustaining students' efforts to engage with ideas that induce cognitive and emotional discomfort. Our methods for identifying significant electrodermal activity (EDA) were based on change-point models. The change-point model was used to identify structural changes in the EDA data, while video evidence was deployed to validate the changes and ensure the activities engaged in have appropriately led to a surge, a decrease, or stabilization of EDA profile. The findings from the pilot study suggest that EDA may be used as a means for detecting learners' discomfort, particularly while facing failure. A correlation such as this will address a critical need neglected by prior research that uses self-report data and learners' overt talk as the sole metrics to construct claims regarding learners' cognitive and emotional discomfort. EDA technologies can augment people's capacity to monitor and reflect upon the emergence of their discomfort and thus deliver to learners a psychological buffer that supports their continued learning—transforming a liability into an asset.

Mentor of Research Group
Gideon K. D. Zamba, Professor, Department. of Biostatistics, University of Iowa

Other Contributor
Dosten Kpozehouen, Graduate Student, Department of Biostatistics, University of Iowa

Leah Valentiner
Wellesley College

Gretchen Guetzlaff
Wartburg College

Shalimar Vargas Grippo
University of Puerto Rico – Mayagüez

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Comparing exposure pathways of enteric pathogens to infants living in low – to-middle – income countries

Over half a million children die each year from diarrheal diseases. This burden disproportionately affects those living in low-to-middle-income countries. Children in these areas experience a wide range of exposures to pathogens, and it is unclear which exposure pathways are among the most important contributors. Observational data were collected on children's behavior in a study led by Drs. Kelly Baker and Daniel Sewell in two neighborhoods in Nairobi, Kenya. This research project aims to use Bayesian methodology to analyze these data. We generated four models for each behavior of interest using Bayesian methodology: Poisson models, negative binomial models, zero inflated Poisson models, and zero inflated negative binomial models. We fitted models without categorizing the individuals, as well as ones that grouped individuals based on age, neighborhood, and both age and neighborhood. The best model for each behavior was selected using the lowest DIC. With these results, we were able to estimate the rate of childrens' behaviors that could expose them to the different pathogens in the area. The long-term goal of this research project is to combine these data with microbiological data to estimate the probability of children becoming infected through various pathways. This ISIB project assists this goal by providing insight into the rates at which certain behaviors occur among infants, and whether these rates differ by age and by neighborhood. The results of the larger project will provide key information for developing effective intervention strategies for these diseases.

Mentors of Research Group
Dr. Daniel K. Sewell, Associate Professor, Department of Biostatistics, University of Iowa
Daniel I. Kakou, Graduate Student, Department of Biostatistics, University of Iowa

Jessica Lee
(New York University)
Elaina McGovern
(Boise State University)
Juleika Torres
(University of Puerto Rico- Mayagüez)

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Statistical Models and Disparity Measures of Annual Incomes in the US

Income disparity is an important social and economic issue in our current world. To better understand the extent of disparity in the US, we used information from the Internal Revenue Service Income Statistics and the United States Census Bureau's Annual ASEC survey to estimate quantiles (0.00001 to 0.99999) of individual 2022 income. We calculated disparity metrics geared for incomes (the Gini Index, the Decile Ratio, and the Palma Ratio), as well as some elementary measures of statistical variability (standard deviation and coefficient of variation). We examined how these measures are affected by various potential changes in the income distribution, such as additive and multiplicative transformations, as well as bringing the lowest-value incomes to a standard minimum. We found that raising incomes across the distribution decreased disparity even better than just adding income to the lowest tail, although at a higher cost. We also compared our estimated distribution to standard statistical models, including the normal, uniform, exponential, gamma, and the log-normal distributions. We found that a log-normal distribution was the best fit to our data, so we examined how changes to the parameters of this distribution can affect disparity metrics. We found that changes in the sigma parameter of the log-normalized data set were more impactful in reducing income disparity than changes in the mu parameter. Although we did not find a model that captured all aspects of the distribution, future work could be aimed at exploring additional parametric models, such as a zero-inflated distributions and other mixture models. It is our hope that our investigations will help provide quantitative frameworks for assessing programs aimed at narrowing the income gap and promoting equal opportunities, which requires concerted action on the part of governments, institutions, and society at large.

Mentor of Research Group
Dr. Jeffrey Dawson, Professor, Department of Biostatistics, University of Iowa