Elizabeth Binkina

Dr. Kristi McDuffie

RHET 233

23 July 2023

Narrative

I decided to write a newsletter for undergraduates currently majoring in statistics, as well as students thinking about majoring in statistics at the University of Illinois. I do know that students, especially when they are searching for the right program, often read these newsletters to better understand what this degree entails and also what opportunities and career options are presented within this college. This newsletter is a great way to speak to the student audience about why they should specifically study statistics, what real-world problems they will get the opportunity to solve, and what kind of careers they can possibly have in the future. Students, including myself, often face tough decisions when it comes to making the best choice with their education.

When students read this newsletter, I want students to get the big picture about what statistics is and how it relates to the real world, as well as become encouraged and excited to study in this field. Picking a school and major can be one of the most overwhelming experiences in a student's career. Therefore, I want this newsletter to be one of their resources. Another purpose of this newsletter is to ameliorate students from feeling intimidated when pursuing this major. The word "statistics" is often associated with words and phrases like: hard, contains heavy math and calculus, rigorous, coding, etc. While those words are an accurate description of this

major, there is a lot more good to it that students do not really have the knowledge about. I want to emphasize the good and rewarding parts of statistics to these students as someone who has experience pursuing this degree. I want students to understand that they will learn to solve big real-world problems using certain statistics techniques.

To optimize the amount of students reading this newsletter, I would publish it in the Statistics@ Illinois Newsletter as well as the Department of Statistics Newspaper. This is a platform that many students, including myself, check fairly often. They also send out email notifications so no statistics undergraduate will have the chance to miss it. It will also be sent out to students who are not currently enrolled at UIUC when they ask for more information during the UIUC application process.

I made the rhetorical choice to include specific examples of how statistical techniques and tests are used to solve very complex real-world problems, such as predicting Alzheimer's and heart diseases. Using these examples, students should feel inspired and motivated to see that this could be in their future if they decide to pursue this field of study at UIUC. With a newsletter that includes research on statistical techniques that can very well solve real-life complicated problems that affect millions of people, students are likely to stay engaged and read until the end. I find that students are likely to become more engaged with reading this specific genre because it is written by a student that is passionate about this topic. They will find it as an extremely helpful and resourceful tool. Students often seek advice from mentors, older students, and generally people with experience and knowledge in the area as it provides them a sense of trust.

In my genre analysis, I discussed how authors of scholarly journals and articles write their reports. I talked about how authors present their research with evidence from their statistical

research techniques and methodologies. In a way, I also present the research I found with evidence from statistical tests. But, the biggest difference is that I go into significantly more detail about why these statistical tests provide exciting and riveting results when it comes to answering a research question. I talk a little about the research process, but spend a great amount explaining why solving these problems are incredibly rewarding.I also talk about how students have the opportunity to improve and save the lives of millions of people to encourage them to pursue this field of study.

I conducted a great amount of research for this project. I already had a bit of knowledge of all of the statistical research methods I used in the project, as I have experience implementing and using these algorithms and conducting research studies of my own. Therefore, I conducted research on how these algorithms apply to bigger real-life examples such as life-threatening diseases. I wanted to incorporate these high profile examples in my genre production because it would give readers a bigger picture of how these algorithms and tests work on a much larger scale.

I originally wanted to write my own scholarly article but I figured that my intended audience would find it more interesting if I wrote to them in a more informal tone, as well as published in a more accessible location. I wanted this genre production to encourage and inspire students while also being relatable, which would be incredibly hard to accomplish with a formal scholarly paper. Therefore, I diverted to writing a newsletter. I wanted to publish my findings in a network that would reach the most statistics undergraduate students, and I thought that the UIUC Statistics Newsletter would be the perfect fit.

The most difficult part of this project was explaining the research in my own words, while also trying to keep the audience understand and engage in the material. I also had to watch how I present my information because part of my intended audience is students that may not have the knowledge of these specific statistical terms and tests yet. Therefore, I explained the specific statistical tests in terms inexperienced students could still understand. I am proud of how I present my enthusiasm with this field as well as how I encourage others to become inspired with the research I provided, as well as the language I used.

# STATISTICS @ ILLINOIS NEWSLETTER

## Why Study Statistics? Rewarding Real World Problem Solving using Statistical Research Techniques and Methodology that Save Lives.

You might wonder what is so special about studying statistics. But did you know that it is a powerful tool used to predict anything from the number of people predicted to show up for a flight all the way to predicting an incurable disease? If you've ever been bumped from a flight, it's because the airline underestimated the number of people that would show up. Airlines do this to not lose potential profit. Statistics is not just a branch of mathematics; it is a booming area of study in which real-world problems are being solved every day in a variety of different fields. This entails using a vast amount of data along with previous trends to carefully and quickly predict future patterns and events. Statistics is all about collecting specific data that will help answer your research question, performing diagnostics to best fit the data to an appropriate model, and conducting several statistical tests and regressions to find trends and patterns in the data set.

Studying statistics is incredibly rewarding, as you are given the opportunity to create many kinds of models and graphs that explain various complex situations. You get to use software to analyze large amounts of data very quickly, which would be nearly impossible to do by hand, as it can help you solve very complicated real-world problems. Another exciting part of studying statistics is that you are never going to be presented with the same real-world problem. New data sets

are continuously being created, techniques and interpretation methods are constantly being advanced, and new statistical models and model selection processes are always being developed. Being a statistics undergraduate will help you develop critical problem-solving and thinking skills, which are of high value in the workforce. In this newsletter, I will specifically talk about the riveting real-world applications of statistical research techniques that help solve complex problems in today's world. It will hopefully get you more excited and motivated about the extraordinary possibilities that you will entail with this undergraduate degree.

# Example 1: Logistic Regression in Alzheimer's Research

One real-world example is in Alzheimer's research. Alzheimer's disease is an incurable and detrimental neurological disorder that affects more than fifty-five

million people, causing hundreds of thousands of deaths every year. You may even know someone, maybe a family member, with Alzheimer's. It is extremely difficult to watch someone you love still be physically alive but with loss of all of their identity. However, there are statistical techniques continuously evolving that can better predict this disease earlier, helping healthcare professionals take action sooner.

Machine learning and linear regression are used to predict Alzheimer's disease based on certain variables such as age, sex, fitness level, etc. These are both statistical techniques that you will learn in your undergraduate statistics classes. A team of doctors, researchers, and specialists in data analytics led by Park et al. collected a vast amount of data from health records and utilized machine learning algorithms and logistic regression to identify the variables

that contribute the most to the development of Alzheimer's. This is an incredibly powerful tool as it uses data from a vast number of patients to identify the most significant causes of an incurable disease. Their findings guide healthcare professionals in creating "better therapeutic strategies for delaying the onset of AD" (Park et al. 2) that could improve the lives of patients with Alzheimer's. In their research study, the logistic regression led to the conclusion that "hemoglobin level is the most significant predictor" (Park et al. 3). This information is crucial and incredibly valuable because now healthcare professionals can understand that a patient's hemoglobin level is a significant predictor when assessing their likelihood of developing Alzheimer's, enabling them to take appropriate measures to help increase their quality of life much sooner. This is the kind of research you would be conducting in the work force. Imagine being able to give your loved one who has been diagnosed with this disease a better life. Is that not extraordinary?

The power of statistics lies in being able to use a massive amount of data, and these statistical techniques are specifically designed for that purpose. For example, "Machine learning is an optimal choice of analytics for analyzing large-scale administrative health data containing thousands of descriptors from hundreds of thousands of individuals...Given the recent rapid growth of machine learning technology, the application of AI technology to clinical predictive modeling is likely to have a deep impact on medicine" (Park et al. 1). With these efficient and accurate algorithms and statistical methodologies constantly being improved, which is something that you as a statistician will always be a part of, you will help improve and save many many lives. As a statistician, you contribute a great deal to

improving patient care, as well as making significant progress in finding cures for other detrimental and incurable neurological diseases like Parkinson's and Huntington's disease. This is incredibly rewarding and powerful because doctors would not be able to better treat their patients without this statistical research.

# Example 2: Machine Learning Algorithms in Heart Disease Research

Another way you can use statistical research methods is to predict heart diseases. Heart disease is one of the leading causes of death in America, yet it is vastly more predictable and avoidable than many other conditions. You may know someone, maybe a friend or family member, with some sort of heart condition. It is not that uncommon. There are many conditions that can be congenital, including ones that are genetic, but many are developed later in life. These are the most preventable. Many acquired heart conditions can be prevented if victims knew the early symptoms and sought treatment sooner. All patient data is kept in medical records, which could be used by statisticians to create models to help identify the main contributors and symptoms. This could allow healthcare professionals to catch the disease sooner and come up with better options for prevention and treatment for the condition. Since the volume of medical data is rather large, statistical algorithms like machine learning are a perfect tool to efficiently analyze this vast amount of data. "The healthcare field is an application area of data mining since it has vast data resources that are difficult to be handled manually...One of the reasons for fatality due to heart disease is that the risks are either not identified, or they are identified only at a later stage" (Latha et al. 1). Using this already

available data could give thousands of future patients an opportunity to prevent heart disease and help them live longer and healthier lives.

For example, if a research study concluded that an increase in exercise and better nutrition significantly reduced the chances of developing heart disease, many individuals would have numeric and scientifically backed data to motivate them to start practicing these healthier habits. Another example would be if the research study concluded that shortness of breath is the most statistically significant early symptom of heart disease. This would prompt individuals who have experienced shortness of breath to get checked out by a doctor sooner, giving them a significantly better chance of receiving treatment before the disease progresses to an untreatable stage. Yes, data analytics is not such an easy job as it requires countless hours of coding, extraordinary willpower and effort, debugging, fixing errors, figuring out correct models, testing, testing again, etc. But, Imagine the research you put your blood, sweat, and tears into saves millions of patients from cardiac problems. It would feel good, would it?

This research process would be nearly impossible to perform any other way. For instance, "The K-means clustering algorithm was utilized to extract data from the dataset, and the frequent patterns were mined using the Maximal Frequent Itemset Algorithm (MAFIA) for predicting heart disease based on different weightage assigned to different factors. The frequent patterns having a value greater than a specific threshold were found to be precise in detecting the occurrence of a myocardial infarction" (Latha et al.2). Myocardial infarction (MI), otherwise known as a heart attack, can go completely

undetected and lead to catastrophic events, including death. Because the MAFIA algorithm is so precise in using patients' data to detect frequent patterns, it can potentially prevent the occurrence of a heart attack and save a life. This is a very impressive technique because even the greatest and most experienced doctors cannot predict such cardiac events with the symptoms the patients provide, as heart attacks are often "silent and go undetected" (Ojha 1). Doctors, especially those that work in emergency medicine, where many heart attacks occur, can utilize the MAFIA algorithm to quickly come up with a course of action to help save a trauma patient's life before it is too late. This includes providing them with certain medications, oxygen therapy, early testing and labs, etc. By creating algorithms like MAFIA, you can indirectly save the lives of countless individuals. Doctors are not robots, they very rarely, but still do, make mistakes. When you

create a successful algorithm like MAFIA, you have the potential to not only save lives, but also significantly reduce inevitable human error. How impressive!

## Example 3: Data Mining Used in Kidney Dialysis Research

Another statistical research method, data mining, is an exciting statistical process that sorts through extremely large data sets and identifies different patterns and relationships between variables to predict certain trends. You will take a class solely dedicated to learning this process, as it is a very commonly used practice. Data mining is defined as a "non-trivial extraction of implicit, previously useful, and unknown information about data" (Bala 960). In the medical field, this process is used to obtain information from various research reports in order to carefully extract important and useful information that helps healthcare

professionals make informed decisions when treating their patients. "Due to its ability to discover the relationship and pattern of the medical database, early detection or prediction of pathological conditions through mining has become feasible" (Sriram et al. 1). This field deals with a high volume of both numerical and categorical data, which makes statistical software and methodology the best way to properly analyze it and draw the most accurate conclusions. These relationships between variables are incredibly important as they lead to the discovery of which symptoms, medications, lab results, etc., have the most significant impact on the development of certain diseases.

Doctors heavily rely on data from lab results and scans to diagnose their patients, as well as taking medical history into consideration. Taking a massive data set that contains millions of other patients with similar symptoms and lab results into consideration would significantly improve the accuracy of diagnosing and treating patients. For example, Sriraam et al. presented a "data mining approach for parametric evaluation to improve the treatment of kidney dialysis patients. The experimental result shows that classification accuracy using Association mining between the ranges of 50–97.7% is obtained based on the dialysis parameter combination. Such a decision-based approach helps the clinician decide the level of dialysis required for individual patients" (Bala 963). Kidney dialysis is used when either the patient is quite low on a transplant list for a kidney or it too weak to undergo surgery. It is incredibly comforting for the patient to know they are in the best possible hands even when they are going through one of the hardest times in their lives, all because of your analysis. Data mining is an incredible life-saving technique that carefully predicts how

to treat every person on a case-by-case basis. The model is incredibly valuable as it can help many healthcare workers develop an extremely accurate treatment plan, giving patients the best chance of recovery. Thanks to the statistical models you will work with in your career, you have the capability to improve the quality of life for many people. This accurate science consistently reduces the chance of human error, avoiding unnecessary mistakes that can even lead to death.

## Takeaways:

Overall, as you can see from the examples of using statistical research methods to solve real-world problems, they are quite powerful. Being able to solve problems as complex as finding the main cause of complicated diseases and contributing those skills to save lives is quite a rewarding process. This is why statistics is such an amazing area of study, as it improves the quality of life for many individuals, helps people look at trends and make better decisions for the future, and makes analyzing large quantities of data very efficient and accurate. As a statistics undergraduate, you have all of these rewarding moments to look forward to. Being able to say "wow, I did that" at the end of a project that solved a major issue is an extremely exciting moment.

Work Cited

Park JH;Cho HE;Kim JH;Wall MM;Stern Y;Lim H;Yoo S;Kim HS;Cha J; "Machine Learning

Prediction of Incidence of Alzheimer's Disease Using Large-Scale Administrative Health Data."

*NPJ Digital Medicine*, 26 Mar. 2020, https://doi.org/10.1038/s41746-020-0256-0

Latha, C. Beulah, and S. Carolin Jeeva. "Improving the Accuracy of Prediction of Heart Disease

Risk Based on Ensemble Classification Techniques." *Informatics in Medicine Unlocked*, vol. 16,

2 July 2019, p. 100203, https://doi.org/10.1016/j.imu.2019.100203 .

Ojha, Niranjan, and Amit S Dharmoon. "Myocardial Infarction - Statpearls - NCBI Bookshelf."

*National Library of Medicine*, 8 Aug. 2022, www.ncbi.nlm.nih.gov/books/NBK537076/.

Sriraam, N., et al. "Data Mining Approaches for Kidney Dialysis Treatment." *Journal of*

*Mechanics in Medicine and Biology*, 6 July 2005, www.worldscientific.com/doi/abs/10.1142/

S0219519406001893?journalCode=jmmb .