THE MASON STATISTICIAN

2025

A SUMMARY OF STUDENT CAPSTONES

completed by graduates of the Department of Statistics at George Mason University

Issue 2

TABLE OF CONTENTS

edited by
Jonathan Auerbach
Department of Statistics
George Mason University

MESSAGE FROM THE CHAIR

Dear Graduating Class,

Congratulations on completing your capstone projects in 2025, during this historical moment! These projects are the seeds of transformation, reflecting your hard work, problem-solving skills, intelligence, and statistical knowledge. They demonstrate the power of data science and statistical thinking, showcasing your mentors' and professors' dedication to preparing students for the real world.

As you step into the world armed with knowledge, skills, and dreams, remember that you are in control of your career path. Mason Statistics will continue to support you as your academic family. Your statistical education is adaptable—you are not only a statistician but also a data scientist—and you can grow and evolve no matter how the world changes as long as you remain curious and inquisitive about the world around you. Remember: a statistician is defined not only by their training and statistical skills but also by a commitment to scientific inquiry coupled with presenting information with honesty and integrity.

Please take our survey to stay in touch!
https://tinyurl.com/GMU-STAT-Exit123

Jiayang Sun
Professor, Bernard J. Dunn Eminent Scholar, and Chair

CAPSTONE ABSTRACTS

Khalid Al-Falah

Maximizing GoFundMe Donations

I analyzed GoFundMe campaigns launched in 2022 to determine what makes a campaign successful. Specifically, I examined how the writing style of the description correlates with the amount of money the campaign raised. For example, does a more somber description result in more donations than a more hopeful one?

I analyzed data on approximately 13,700 different campaigns from the US that started in 2022. To capture the writing style of the description, I used sincerity scores, which measure the perceived truthfulness of the content, and included the scores in the data table alongside other factors like charity category, state, whether a cover photo was used, and number of photos in the main body. I then used a linear regression model with an elastic net penalty to model the relationship between these predictors and the amount raised in the campaign (on the log scale).

I found that the writing style of the description, as measured by the sincerity score, is an important factor in determining whether a campaign will be successful. GoFundMe users should be sure to carefully construct sincere narratives in order to maximize the benefit of the platform.

Figure: A word cloud of GoFundMe descriptions. Word size corresponds with the frequency a word is used, indicating descriptions often involve family and healthcare expenses.

Why did you choose statistics?

Statistics is about finding the truth in the world, and finding this truth is my goal in life. Numbers have deeper meanings than we give credit, representing the complex relationships all around us. The field of statistics has equipped me with the tools to discover these relationships. Sifting through the data and finding patterns that help explain a phenomenon is exhilarating.

Acknowledgments

I would like to thank the statistics department. I started my academic journey as a cybersecurity engineer, but I soon discovered that my passion was statistics. My meetings with Dr. Hunter and Dr. Izmirli helped me identify and achieve my goals, and classes with Dr. Auerbach and Dr. Rios, as well as work with Dr. Fadahunsi and Dr. Markaryan, helped me better understand the field. These experiences have inspired me to continue to pursue statistics after graduation.

Isaac Amouzou

How departure delays at airports affect one another

I investigated how one departure delay can spillover and cause delays in subsequent flights, using data from the Bureau of Transportation Statistics. I counted the number of departures delayed by15 minutes or more each hour between 2019 and 2024 at the top 50 US airports, which constitute about 85% of all air traffic in the US. I then modeled delays using vector autoregression with a 5-hour lag, i.e., a VAR(5). I then decomposed the forecast error variance into two components, corresponding with whether delays originated from within an airport or from other outside airports. I also conducted an impulse response analysis, and simulated airport delays to examine how those delays propagate throughout the system.

I found that approximately 50 percent of the variance in forecasted delays at an airport can be explained by past delays at that airport and the remaining 50 percent from delays originating from other airports. For example, BOS is a major contributor to delays at DCA, and a delay shock from BOS to DCA takes around10 hours to clear on average. Finally, I found that a shocking increase of 10 delays in an hour at DCA increased the expected number of delays by 6 over the next 15 hours.

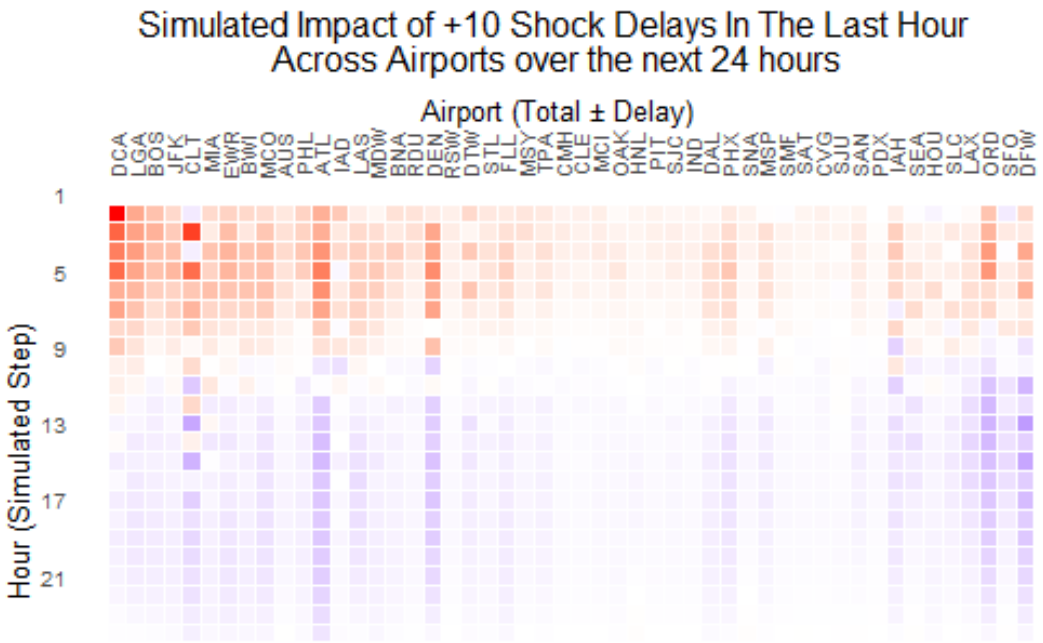

Figure: A heatmap of the simulated change in the number of delays at major airports in the United States in the hours following a delay at DCA. Darker colors represent more (red) or fewer (blue) delays.

Why did you choose statistics?

The field of statistics combines my love for numbers and programming with research. I enjoy learning how statistical models can be applied to evaluate claims, develop inferences, and how different methodologies can lead to varying interpretations of real-world phenomena.

Acknowledgments

I would like to thank Dr. Ben Seiyon Lee for his mentorship and guidance. I would also like to thank my Mom and Dad for inspiring me to love mathematics and research.

Sam Bahamonde

The Effect of Sports Gambling on Gambling Addiction

I investigated how the legalization of sports gambling in many states has affected gambling addiction in the United States. I used the number of calls to the National Problem Gambling Helpline as a measure of the prevalence of gambling addiction. The helpline reports data on the number of calls they have received each month over the last nine years.

I represented the data as a time series and used change point analysis to identify sudden changes in the number of calls. I then compared the timing of any significant jumps or dips to specific major sporting events or other factors that may explain the reason for the spikes. Changepoint analysis is useful because it identifies the most important changes, ignoring random fluctuations in the data that can be explained by chance.

I found that there are several significant jumps in calls to the center, almost all of which occur during the fall and winter, from October to February. This coincides with the professional football season in the United States, the sport with the greatest amount of betting in the country.

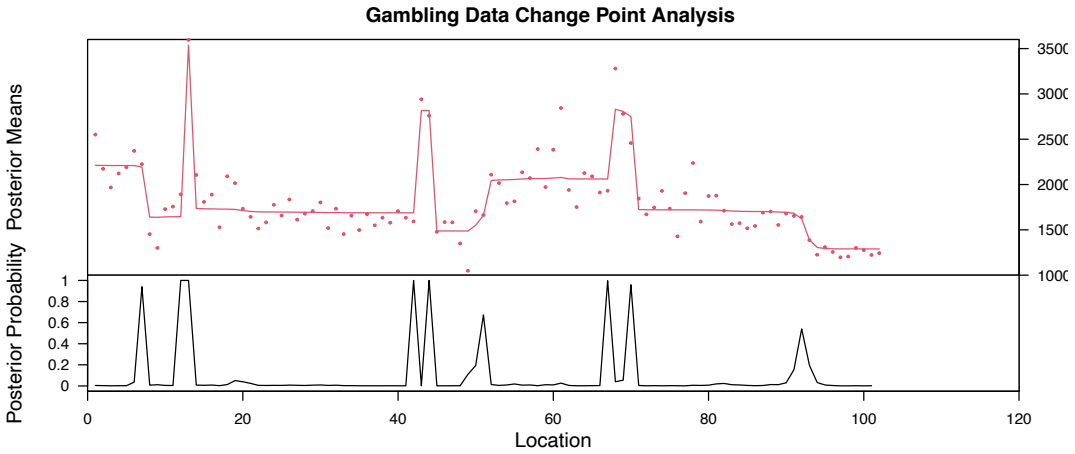

Figure: Calls (points) are decomposed into trend and changepoints (red line). The probability of a changepoint is included (black line).

Why did you choose statistics?

I have always enjoyed math, and I loved AP statistics in high school. I considered other majors, but I decided to apply as a statistics major at Mason. The rest is history.

My favorite part about statistics has always been how applicable it felt to the real world. Learning other math subjects is interesting and they obviously have their own important applications, but I think statistics classes in general do a good job of using real examples and datasets so you know what you are doing can help people solve real problems outside of a classroom.

Acknowledgments

I would like to thank my parents and stepparents, who support me in everything. I wouldn't be the person I am today without them. I'd also love to thank the Mason faculty and staff, especially in the statistics department. They have made college fun and interesting and have always been there if I've needed help along the way.

Jake Daunheimer

Predicting NFL Success for Defensive Ends: A College Performance Analysis

I investigated whether the current approach used to predict the success of football players in the National Football League can be improved. We see players drafted high who don't pan out, and others drafted lower but perform better than their peers. Is there a way to better predict their success using statistics?

I compiled a data set on one hundred defensive ends, sixty percent of whom were 5 or more years into their career. Variables included college stats and physical attributes such as height, weight, 40-time (time to complete the 40-yard dash), college sacks per game, college tackles for loss per game, NFL sacks per game, NFL tackles for loss per game, and NFL draft position.

I used multiple linear regression with stepwise selection to determine the most important predictors of success as measured by NFL sacks per game, the outcome variable. I found that the two most important predictors of NFL sacks per game were college sacks per game and college tackles for loss per game. I conclude that college performance matters more than physical attributes, 40-time, and draft position, although these measures are still valuable information for NFL Teams.

Image: John Ross running the 40-yard dash in 4.22 seconds in 2017. Source: nfl.com (https://tinyurl.com/yeym8v2w)

Why did you choose statistics?

My capstone builds on my research from high school, where I won an award for best presentation in statistics for showing that 40-time is not a good predictor of NFL success among quarterbacks. After that experience, I chose to major statistics with a concentration in Sports Analytics.

Acknowledgments

I will always first thank God for all he has done for me, especially in gifting me my wife, Alicia, and our first son, Matthew. Without Alicia's constant support in allowing me to fully focus on providing for us as I earned this degree, I would not be standing here today. I would also like to thank my parents and grandparents for their constant support in my dreams and always believing in me.

Abigail Finch

How Does Batting Stance Angle Influence Batted Ball Profile?

The aim of my project was to determine if different batting stances have a significant impact on a baseball player's batted ball profile. A hitter's swing is a chain reaction, and I wanted to know if there was an optimal set up in the box that produces the best results. To answer this question, I obtained data from baseballsavant.mlb.com, and combined data sets on batting stance, exit velocities, and batted ball types. I focused on batting stance angle, which measures how opened or closed a batter is in the box.

I used a general linear regression model to determine whether batting stance angle was a significant predictor of an individual's batted ball profile, defined by multiple outcomes: launch angle, exit velocity, hard hit percentage, and barrel. I also used mediation analysis to see if the relationship between batting stance angle and an outcome was mediated by another outcome.

I found that batting stance angle does not have a significant impact on batted ball profile at the major league level. I conclude that batting stance angle does not matter, and major league hitters should step to the plate in a way that makes them the most comfortable.

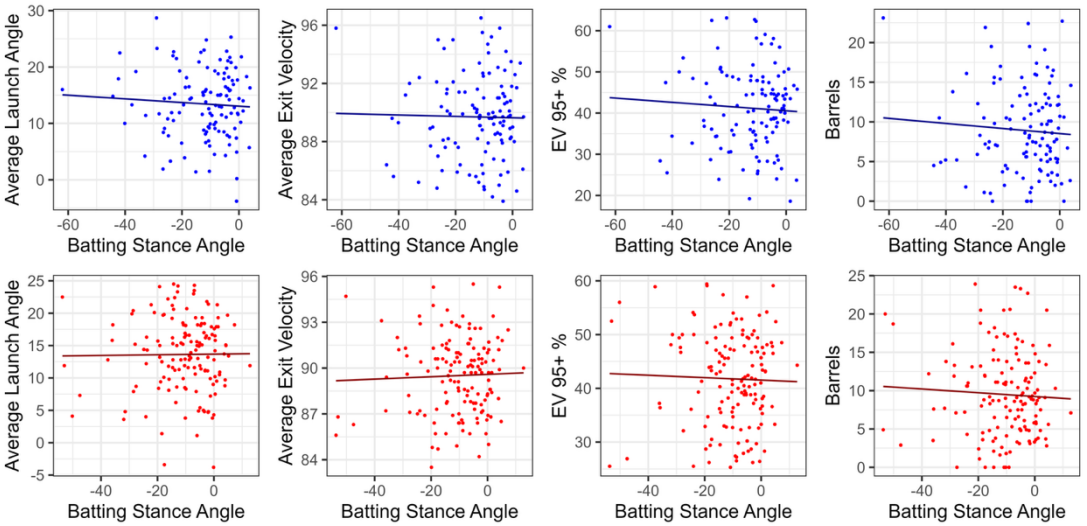

Figure: Batting angle stance and four different outcomes that make up batted ball profile: average launch angle, average exit velocity, hard hit percentage (EV 95+%), and percentage within the barrel zone (barrels). Left/Right-Handed indicates whether a baseball player is left or right-handed.

Why did you choose statistics?

I chose to major in statistics because I wanted to use data and analytics in baseball, which the sport analytics concentration allowed me to do. Learning how to code for statistics has set me up for success in any career path I take.

Acknowledgments

I would like to thank my parents, Jen and Darin, for their unwavering support and granting me the opportunity to go to college and pursue my goals. I would also like to thank the coaches and players of GMU and UCWV Baseball for taking care of me while I learn from them. Lastly, I would like to thank the New York Mets for taking a chance on me and allowing me to finish my degree before I start in their organization.

Emma Harris

Invisible in the Data: Ethical and Statistical Failures in Federal Disability Measurement

I investigated how survey instrument design and operational definitions contribute to statistical underrepresentation of disabled individuals in federal data systems. I conducted a secondary analysis of national disability prevalence estimates across sources such as the American Community Survey (ACS) and the Current Population Survey (CPS), highlighting how inconsistent question sets and measurement frameworks introduce coverage bias, measurement error, and compromised validity of population-level disability statistics.

I used a literature-based analytical approach, synthesizing findings from peer-reviewed statistical studies and federal datasets to compare disability prevalence estimates across national surveys. This included evaluating sources of measurement error, coverage bias, and instrument sensitivity using comparative prevalence data from ACS, CPS, and NHIS. This analysis emphasized how definitional inconsistencies and survey design choices impact statistical validity and data-driven policy decisions.

My findings indicate that federal disability data collection introduces significant measurement error, coverage bias, and misclassification, with prevalence estimates varying from 9.2 percent to 39.4 percent depending on the survey. Inconsistent operational definitions and limited instrument sensitivity compromise validity and underrepresent individuals with non-apparent, psychiatric, or episodic disabilities. I recommend adopting multi-item measurement frameworks, improving construct validity through collaboration with disability experts, and standardizing definitions to enhance data comparability and statistical accuracy.

Why did you choose statistics?

I've always enjoyed working with numbers, but statistics showed me how to turn that interest into something practical and impactful. I love how it helps uncover patterns, solve real-world problems, and make sense of complex systems. Statistics challenges me to think critically and communicate clearly, and I chose it because it gives me the tools to make informed decisions, support evidence-based work, and contribute to something bigger than myself.

Acknowledgments

I'd like to thank my parents for always believing in me and supporting me every step of the way. I also want to thank my best friend, Bea, for being there through the highs and lows and always reminding me what I'm capable of. And a huge thank you to my medical alert service dog, Kika, for being by my side every day and helping me stay safe, focused, and grounded through it all. I wouldn't have made it through this journey without them.

Michelle Harris

External Factors Affect Trends in Mental Health-Related Incidents on George Mason University Campus

I explored how different factors, such as holidays and academic breaks, affect the number of mental health-related incidents reported on campus. Specifically, I wanted to understand if events like exam periods and seasonal changes lead to more welfare checks, mental health calls, or self-harm calls. To study this, I analyzed five years of incident data, from 2019 to 2023.

I modeled the number of health-related incidents as a 3-state hidden Markov model, where the probability of transitioning between states changed according to observed covariates like exam periods, holidays, and seasonal breaks. I then used the fitted model to predict mental health incident for 2024.

I found that mental health-related incidents, especially welfare checks and self-harm calls, tend to increase during high-stress periods like exam seasons and decrease during breaks like summer. Based on these patterns, I recommend providing additional mental health resources during exams and late fall, while maintaining steady support year-round to address ongoing needs.

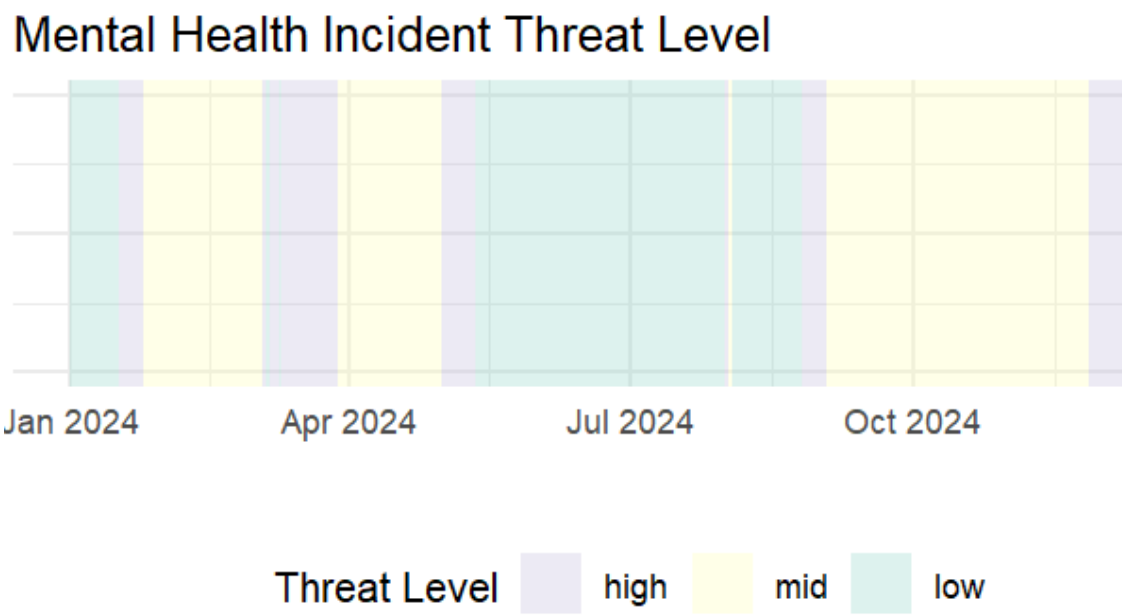

Figure: A 3-state hidden Markov model was fit to data on mental health incidents that occurred on George Mason University campus between 2019 and 2023. The model was then used to predict the state for each day (x-axis) in 2024. States are colored according to the expected number of mental incidents, from low (green) to high (purple), and are interpreted as threat levels. The predictions indicate mental health incidents align with the academic and holiday schedule.

Why did you choose statistics?

Statistics became the right path for me once I realized how it could support research in areas I care deeply about, like mental health. The skills I've gained—especially in data visualization and analysis—have helped me think more critically and communicate ideas clearly.

Acknowledgments

I would like to thank Drs. Auerbach, Izmirli, and Rios for their impactful role in my academic career. Their enthusiasm for teaching inspired a sense of curiosity and self-confidence, both academically and professionally.

Jake Lane

Assessing the Impact of Federal Funding on Degree Conferral at George Mason University

In light of recent political shifts and anticipated reductions in federal funding for universities, I investigated whether federally sourced funds had a measurable impact on the number of degrees conferred at George Mason University from 2015 to 2024. I analyzed institutional-level data from the National Science Foundation's Higher Education Research and Development Survey (2013–2023) alongside George Mason's Official Ten-Year Degrees Awarded Dashboard, which includes degree conferrals by college, major, concentration, student level, and demographics.

I used linear regression to model the number of each degree type (Bachelor's, Master's, and Doctorate) awarded each academic year as a function of the sum of the previous three years of federal funding. To assess predictive accuracy, I randomly split the data into 70/30 training and test sets and evaluated the models using root-mean squared error (RMSE). Finally, I simulated hypothetical funding reductions of 10, 25, and 50 percent to determine the impact on the number of degrees awarded by George Mason University.

I find that the effect of federal funding on the number of Bachelor's and Master's degrees awarded is statistically significant, although, surprisingly, the relationship between federal funding and the number of Doctorate degrees is not statistically significant. It follows that a simulated reduction in federal funding results in a corresponding decline in the number of Bachelor's and Master's degrees awarded by George Mason University.

Why did you choose statistics?

My interest in statistics began with a lifelong fascination for patterns and prediction, sparked early by fantasy football and deepened through advanced math studies. My military and defense contracting experience further developed my analytical skills and showed me the real-world impact of data-driven decision-making. Now, I'm eager to transition into data science to apply this passion and experience in solving complex, real-world problems.

Acknowledgments

I could not have achieved all that I have over the past three years at George Mason without the unwavering support of my wife, Rachel. On the long days when I was working full time and attending evening classes, she stepped up as a single parent in every way, managing our home, caring for our children, and making countless sacrifices so I could stay focused on my goals. Without her strength and dedication, none of this would have been possible.

Garrett Martin

Restaurant Tip Percentage and Its Factors

I investigated several factors that could explain why tip percentages vary among restaurant servers, such as sex, server mood, whether a server wears makeup, and a server's internal corporate rating, among others. I collected data from servers at a local restaurant from March 22nd through April 20th, and at the end of each shift, I logged each table's check total and tip total, as well as the previously mentioned variables. I modeled the log proportion of check totals and tip totals using a linear regression model with stepwise selection, and I checked the selected variables using linear regression with a lasso penalty. I then performed a simple linear regression with the most influential factors to obtain my final model.

I found that the factors I analyzed were statistically significant, but they did not explain much of the variation in tip percentages. The most influential factors were the server's mood and the check total, but these factors failed to produce an adjusted R-squared value greater than 0.014. Furthermore, there was no significant difference in tip percentages between male and female servers, or between makeup and non-makeup wearing female servers. I conclude that, of the variables I analyzed in my study, none can predict tip percentage with any meaningful accuracy.

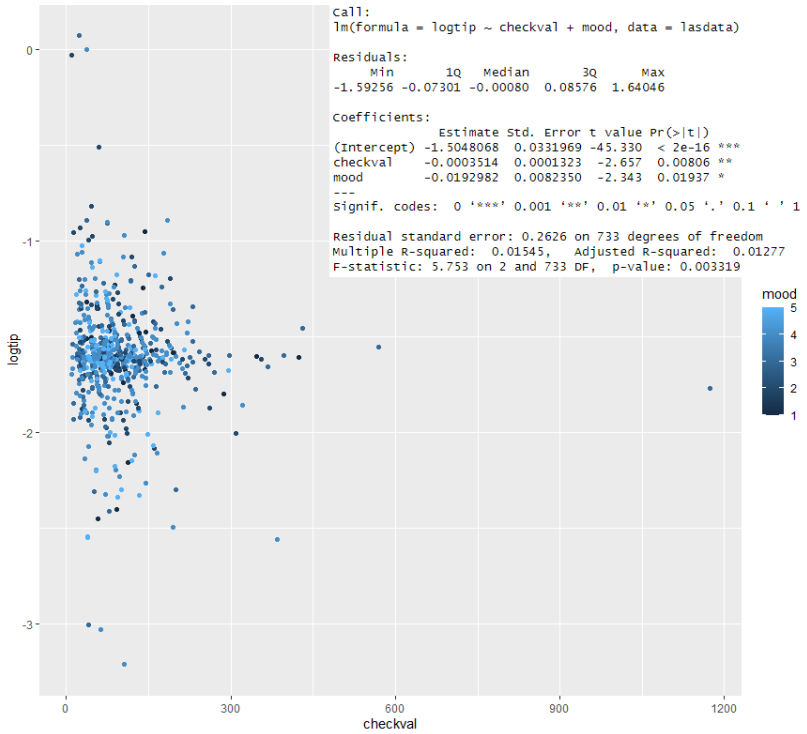

Figure: A scatter plot showing the relationship between the total amount of the check (x-axis) and tip percentage (on the log scale, y-axis). Colors correspond with the mood of the server. Both variables are significant, although much variation is unexplained.

Why did you choose statistics?

I've always enjoyed math and working with numbers, particularly in regard to probability, and I became interested in statistics after taking STAT 250 with Dr. Strazzeri. I found great enjoyment in it after taking further classes in the subject and decided to switch my major.

Acknowledgments

I'd like to thank Mrs. Judy Herbert for being a pleasure and a joy to work for these past 4 years, Ms. Shelley Santana, Ms. Gabby Parrillo, Mr. Nahom Daniels, and the rest of the servers at Ozzie's Good Eats for assisting me in my capstone, and Mr. Jonah Kossoy and Mr. Robbie Albritton for being a phenomenal pair of friends.

Logan Messer

Evaluating Football Prediction Models

I explored the use of predictive modeling in American football. Specifically, I examined the extent to which one can forecast whether a play will result in a touchdown using historical play-by-play data from the National Football League. The goal is to leverage patterns and insights from professional-level data to see if those same patterns can be effectively applied to the collegiate club football level, particularly using George Mason University's club team data. By training models using professional games and then applying those models to the collegiate level, the project examines both the robustness and adaptability of the models.

I used logistic regression and random forest to predict if a play resulted in a touchdown based on several predictors, including down, distance to go, field position, time remaining, and score differential. I found that field position, down, and yards to go were the strongest predictors of a touchdown. Plays occurring closer to the end zone, on earlier downs, and with fewer yards to go were significantly more likely to result in touchdowns. I conclude that these play characteristics can effectively be used to estimate touchdown probability, which could support decision-making for both coaching strategy and sports analytics tools at the collegiate level.

Method	Test Set	Accuracy (%)	Sensitivity (%)	Specificity (%)
Logistic Regression	NFL	79.08	79.18	77.01
Random Forest	NFL	78.71	79.71	57.85
Logistic Regression	College	78	79.56	61.54
Random Forest	College	76	80.29	30.77

Table: A summary of four different models for predicting whether a play will end in a touchdown. The models have similar accuracy and sensitivity but vary in specificity. Specificity is higher when using Random Forest or data from the National Football League (NFL) in comparison to logistic regression or data from the college level (George Mason University) .

Why did you choose statistics?

I have always found numbers and patterns interesting, and I thought a statistics major would be fun. I now enjoy statistics because it allows me to uncover hidden patterns in data and understand how different factors relate to each other.

There is something exciting about using logic and reasoning to make predictions and draw meaningful conclusions. I am drawn to the challenge of finding clarity in complexity and making sense of the bigger picture.

Acknowledgments

I would like to thank my parents and my fiancé Abby.

James Miller

Detecting Census Outliers at the Tract Level Using the Hypergeometric Distribution

My capstone project builds on my research at the U.S. Census Bureau, where I am developing a metric to assess whether a population estimate can be considered an outlier in that it exhibits large coverage error that cannot be explained by chance variation. My project focused on the decennial census where coverage error is determined from the post-enumeration survey. I sought a principled method for confirming the legitimacy of a potential outlier.

I used the hypergeometric distribution, which describes record matching between the decennial census and post-enumeration survey under ideal conditions, to develop a metric based on maximum likelihood theory. I then used this metric to calculate the consistency between the decennial census and post-enumeration survey in each census tract.

Preliminary findings among a small subset of tracts indicate that disagreement is a good predictor of whether the coverage error in a census tract is an outlier. Future work will consider all census tracts in the United States.

(Shaded in red: Matches, total in P Sample)		Enumerated in PES ? (P Sample)		
		YES	NO	Total
Enumerated in Census ? (E sample)	YES	N_{11}	N_{12}	N_{1+}
	NO	N_{21}	N_{22}	N_{2+}
	Total	N_{+1}	N_{+2}	N

Table: A demonstration of how the decennial census and post enumeration survey are combined to determine coverage. Under ideal conditions, the first cell follows a hypergeometric distribution when the margins are held fixed.
Source: census.gov (https://tinyurl.com/6dpcaw46)

Why did you choose statistics?

I chose to major in statistics for the love of painting stories with data and grounding those stories in reality. Statistics is a strong tool that is applicable among many fields and can be used to successfully propel me into a large variety of fields. I enjoy being able to work with data that has meaning, and piecing information together.

Acknowledgments

I will miss professors who are passionate about teaching statistics and truly care about their students.

Madeline Morman

Predicting Formula 1 Race Outcomes Using Ordinal Logistic Regression

I investigated how a Formula 1 driver's starting grid position influences their final race result. Specifically, I examined whether starting further forward significantly increases the likelihood of a better finishing position and how this relationship varies by driver and circuit. The analysis used race data from 2015 to 2024, including each driver's grid and finish positions, race location, year, and team.

I used ordinal logistic regression to model the relationship between grid position and finish position for each driver and each race. A B-spline transformation was applied to grid position to capture its non-linear relationship with finish position. Driver and race were included as categorical predictors (i.e., fixed effects) to control for individual and circuit-level variation.

This study provides clear evidence that starting grid position is a strong predictor of final race outcomes in Formula 1. Drivers who qualify near the front have a significantly higher probability of finishing in top positions, particularly at tracks with limited overtaking opportunities.

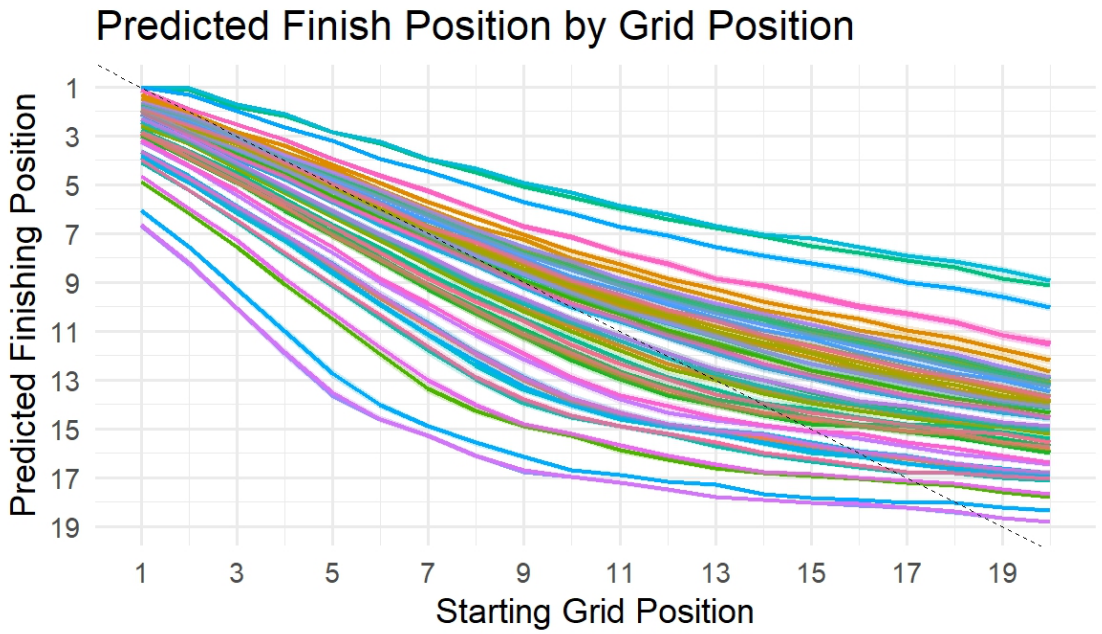

Figure: Each line depicts the non-linear relationship between grid position (x-axis) and finishing position (y-axis) for a Formula 1 driver. Except for the top two drivers, changing the grid position from 1 to 2 reduces the finishing position.

Why did you choose statistics?

I chose to major in statistics because I was inspired by my amazing AP Statistics teacher in junior year of high school. Ms. Horowitz made math and data analytics more exciting. It was during that class that I realized I could combine my lifelong love of sports and racing with the study of statistics and data analytics.

Acknowledgments

I would like to thank my parents for giving me amazing opportunities in life and who support me in everything I do. I would also like to thank all of the stats teachers and professors I have had who have allowed me to continue to learn and grow within this field.

Grayson Saccomando

Is There a Correlation between DUI Convictions and Alcohol-Related Car Crashes?

I examined whether the number of DUI convictions can be explained by the number of alcohol-related car crashes in Virginia each year. As the table (facing page) shows, the number of alcohol-related collisions has declined considerably in recent years. As a result, law enforcement may choose to prioritize other crimes. The question is whether the decline in alcohol-related collisions has been met with a corresponding decline in the number of DUI convictions.

To answer this question, I used linear regression to measure the relationship between DUI convictions and the number of alcohol-related car crashes. I also extrapolated this relationship to determine whether convictions are likely to continue to decline if alcohol-related crashes are further reduced.

I found a strong linear relationship between DUI convictions and the number of alcohol-related car crashes in recent years. This suggests that the decline in crashes is a plausible explanation for the decline in the number of convictions.

Virginia Alcohol-Related Motor Vehicle Statistics (Calendar Years 1999-2023)

Calendar Year	Alcohol-Related						DUI Convictions
	Total Crashes	% of Total	Fatalities	% of Total	Injuries	% of Total	
1999	10,942	7.8%	364	41.5%	8,359	10.3%	28,553
2000	11,085	7.8%	355	38.2%	8,251	10.3%	27,424
2001	11,265	7.8%	358	38.3%	8,211	10.2%	28,044
2002	11,788	8.0%	375	41.1%	8,465	10.7%	27,322
2003	11,388	7.4%	361	38.3%	7,819	9.9%	27,046
2004	11,504	7.5%	343	37.2%	7,911	10.1%	28,471
2005	11,495	7.5%	322	34.0%	7,512	9.9%	28,070
2006	11,736	7.7%	374	38.9%	7,543	10.3%	29,595
2007	11,215	7.7%	378	36.8%	7,130	10.4%	28,787
2008	10,294	7.6%	354	43.1%	7,000	10.1%	31,469
2009	9,366	8.0%	316	41.8%	6,256	9.9%	31,434
2010	8,221	7.1%	274	37.0%	5,578	9.1%	29,063
2011	8,411	7.0%	245	32.1%	5,465	8.6%	28,162
2012	8,777	7.1%	229	29.6%	5,861	8.8%	28,719
2013	8,047	6.6%	253	34.1%	5,288	8.1%	27,333
2014	7,667	6.4%	252	36.0%	5,003	7.9%	24,895
2015	7,592	6.0%	242	32.1%	4,917	7.6%	20,768
2016	7,482	5.8%	262	34.4%	4,855	7.2%	19,925
2017	7,285	5.7%	248	29.4%	4,430	6.8%	18,701
2018	7,181	5.4%	278	33.9%	4,475	6.7%	19,790
2019	7,048	5.5%	264	31.9%	4,402	6.7%	18,648
2020	6,624	6.3%	272	32.1%	3,986	7.6%	14,105
2021	6,749	5.7%	247	25.5%	4,224	7.2%	15,988
2022	6,910	5.6%	274	27.3%	4,174	7.0%	14,247
2023	6,979	5.5%	293	32.3%	4,400	6.9%	14,246

Source: Alcohol-Related Crashes - Virginia Traffic Crash Facts
DUI convictions - DMV Conviction File

Table: Alcohol-related collisions in Virginia between 1999 and 2023. Collisions have declined but the fatality rate has not.
Source: dmv.virginia.gov (https://tinyurl.com/y4tud287)

Why did you choose statistics?

I chose the statistics major because I enjoy analyzing data, and the major has given me the opportunity to analyze data and draw conclusions.

Acknowledgments

I would like to thank Dr. Brett Hunter, as well as my statistics professors and other professors throughout my time at Mason for helping me through this journey. I would also like to thank my family for supporting me throughout my college career.

Joshua Schnellenberger

Changes in 4th Down Decision Making In the National Football League

I examined how 4th down plays have evolved in the National Football League as analytics have become more integrated into the on-field decision-making process. I modeled the change in the proportion of 4th down plays in which teams “go for it,” and I estimated the field position where going for it becomes a 50/50 decision. I refer to this line as the “50/50 yard line.”

I fit a logistic regression model for each season, using yard line and yards to go as predictors. I used monotonic splines to capture the nonlinear relationship between both predictors and the outcome, ensuring that as distance from the opposing end zone or first down marker increased, the predicted probability of going for it decreased. I then used the models to determine the yard line at which teams were equally likely to go for it or not.

I find a clear, gradual increase in 4th down aggressiveness since 2000, with teams opting to go for it more frequently. Results also indicate that this shift towards an aggressive mindset has also correlated with an increase in success on 4th down attempts. Since 2018, a clear trend has emerged with the 50/50 yard line moving roughly 8 yards closer to a team’s own end zone each year.

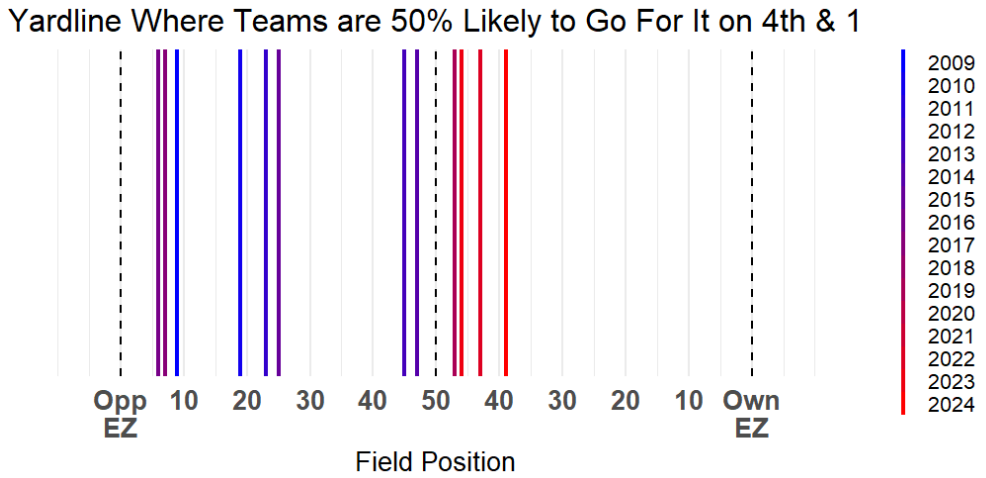

Figure: Each line represents the field position where the model predicts there is a 50/50 chance a team will attempt a 4th down conversion for a given season. This 50/50 threshold has gradually moved further away from the opposing endzone since 2009 and deeper into teams' own territory since 2018, demonstrating the increased aggressiveness on 4th down plays.

Why did you choose statistics?

I became interested in statistics due to my positive experience with AP Statistics in high school. I found that the real-world examples used in all of the problems I was solving made the subject far more interesting to me than any other math course I had taken previously.

Acknowledgments

I would like to thank my parents for giving me the opportunity to pursue my passion, as well as my sisters, Megan and Molly, for their love and support. I would also like to thank the statistics professors at Mason for helping me to get to where I am today, and my AP Statistics teacher Angelo Sciandra, for kickstarting my interest in the subject.

PREDICTION COMPETITION

When will the cherry trees bloom?

Mason students developed statistical models to predict the day that the cherry trees would first bloom in Spring 2025 at five locations around the world: Washington D.C., USA; Kyoto, Japan; Vancouver, Canada; and Liestal-Weideli, Switzerland.

Students share findings with the press

In a featured article published by the Fairfax County Times on March 7, statistics students discuss the advantages and limitations of artificial intelligence. They also share their thoughts on the “Beat the Bot” portion of the competition, in which students compete against AI generated models. Khalid Al-Falah and Madeline Morman are quoted.

AI joins George Mason University’s Cherry Blossom prediction competition

FAIRFAXTIMES.COM

Source: https://tinyurl.com/3p2vzdbw

Competition website: https://competition.statistics.gmu.edu/

AWARDS & RECOGNITION

Statistics Faculty Award presented to Joshua Schnellenberger

for outstanding achievement as an undergraduate student in statistics as determined by a vote of the Department of Statistics faculty.

Senior Leadership and Service Award presented to Isaac Amouzou

for leadership and service to the Department and wider community as determined by a vote of the Department of Statistics faculty.

Best Prediction in the Prediction Competition provided by Garrett Martin

who, among George Mason University contestants, most accurately predicted the bloom dates of cherry trees at five locations around the world: Washington D.C., USA; Kyoto, Japan; Vancouver, Canada; and Liestal-Weideli, Switzerland. The average prediction provided by Garrett was off by approximately one day on average, whereas expert predictions are commonly off by a week or more.

ASA CHAPTER AT MASON

Images: Director of the U.S. Census Bureau, Rob Santos (far right in image above) meets with statistics students at an event organized by the student chapter of the American Statistical Association at George Mason University. Chapter President, Isaac Amouzou, also shown (front in image above). Additional images of the event on facing page.


GEORGE
MASON
UNIVERSITY

STATISTICS AT MASON

UNDERGRADUATE PROGRAMS AT MASON

Undergraduate students majoring in statistics can pursue different concentrations as they work towards completing their degrees.

Statistical Analytics

This concentration blends computer science and statistics at the undergraduate level, two core disciplines of data science.

Applied Statistics

This concentration focuses on analytical methods applicable to a specific discipline of the student's choosing.

Mathematical Statistics

This concentration focuses on the theoretical underpinnings of statistics, preparing students for research and graduate study.

Sports Analytics

This concentration prepares students to work with sports teams and related industries, where data skills are increasingly in demand.

GRADUATE PROGRAMS AT MASON

There are several graduate degrees offered by the Department of Statistics. These programs can be supplemented by concentrations and certificates to create a unique graduate experience.

MS in Statistical Science

This degree program prepares students for statistics and data science occupations in industry and government. Two concentrations:

1. **Statistical Data Science**, which blends computer science and statistics, two core disciplines of data science, preparing students for the analysis of complex data sets.
2. **Modern Statistics**, which provides students with a rigorous curriculum encompassing theoretical underpinnings, sophisticated methodologies, and state-of-the-art techniques.

Qualified undergraduate students may obtain an Accelerated Statistical Science MS.

MS in Biostatistics

This degree program provides a similar training to the MS in Statistical Science, but it allows students to specialize in the design and analysis of health-related and biological studies.

Federal Statistics Graduate Certificate

This certificate is designed for current practitioners who wish to update their skills in statistics, survey methods, and data analysis, including graphics and data visualization.

PhD in Statistical Science

This degree program is a hybrid of probability, computation, and data analysis at the doctorate level.

Department of Statistics
College of Engineering and Computing
Nguyen Engineering Building, Suite 1705
Fairfax, Virginia 22030
703-993-3645