THE MASON STATISTICIAN

2026

A SUMMARY OF STUDENT CAPSTONES

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

Issue 3

TABLE OF CONTENTS

edited by Jonathan Auerbach
Department of Statistics
George Mason University
ISSN 3071-5210 (Online)
3071-5202 (Print)

MESSAGE FROM THE CHAIR

Dear Graduating Class,

Congratulations on completing your capstone projects in 2026! These projects are the seeds of transformation, reflecting your hard work, problem-solving skills, intelligence, and statistical knowledge. They demonstrate the power of statistical thinking and data science, showcasing your mentors' and professors' dedication to preparing students for the real world in this evolving AI landscape.

As you step beyond your undergraduate studies, 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 or a data scientist, but you can become an analyst, biostatistician, AI or machine learning specialist, consultant, researcher, professor, and more. You can grow and evolve no matter how the world changes as long as you remain curious and engaged with 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, clarity, 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

*guest contributor

Abdullah Alshubaily

Does Expected Goals xG Predict Actual Goals?

I examined whether the statistical metric ‘expected goals (xG)’ predicts the actual number of goals scored in a soccer match. The goal of this project was to understand how well xG explains actual goal scoring and to determine whether it is necessary to combine xG with additional information to obtain accurate predictions.

I used several statistical models to analyze a large dataset of 24,580 matches from several European leagues available from the Understat Football Database. The models include linear regression and nonlinear regression. I tested models that included additional variables such as defensive and pressing statistics. I compared models using measures such as R-squared and overall model fit to determine which model performed the best.

I found that expected goals is a strong predictor of goals scored. The best models explained a substantial portion of the variation in goals scored. Adding other variables produced only a modest improvement in prediction error. I conclude that xG is useful overall, but combining it with additional information can be necessary for more accurate predictions.

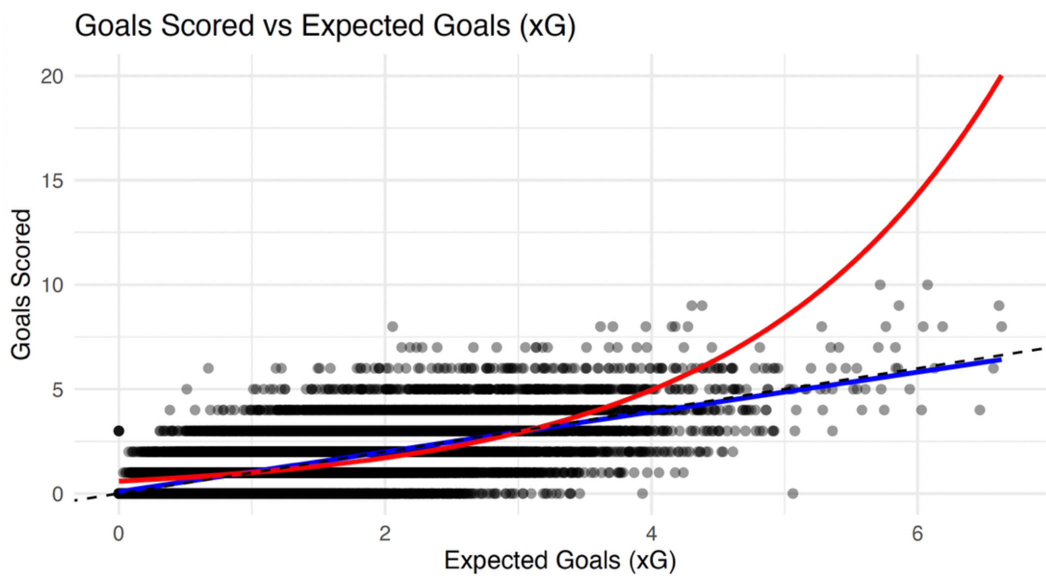

Figure: A scatter plot showing the number of goals scored in each match against the xG metric. Linear regression (blue line) agrees with the identity line (dashed line), suggesting xG is an unbiased measure of goals scored. In contrast, Poisson regression (red line) suggests high values of xG underpredict actual goals scored.

Why did you choose statistics?

I chose statistics because I enjoy collecting data and finding patterns. I am especially interested in sports, and statistics helps me analyze games and player performance. It allows me to understand how data can be used to make better decisions.

Acknowledgments

I would like to thank my professors in the statistics department for their guidance and support throughout my studies. I also want to thank my family for always supporting me and encouraging me to succeed.

Cole Boller-Pinkham

Analysis of Referee Bias in the English Premier League

I investigated whether referees in the English Premier League are biased, and how that bias might manifest in on-field decisions like fouls and cards. For example, I consider questions such as: Is there a home-team bias? Do certain referees show bias for or against specific teams? How large is that bias?

To answer these questions, I propose a metric called ‘adjusted bias’ that measures a referee’s bias toward each team. I then use statistical models such as logistic regression, random forests, and other machine learning techniques to evaluate how that bias influences the probability that a team will win a match.

The results of my analysis suggest that the adjusted bias difference—that is, home bias minus away bias—is an important and statistically significant predictor of the probability of a home win. In extreme cases, the predicted probability of a home win can shift by as much as 40 percentage points. However, these extreme cases are relatively rare. Most games show only a small change in home-win probability.

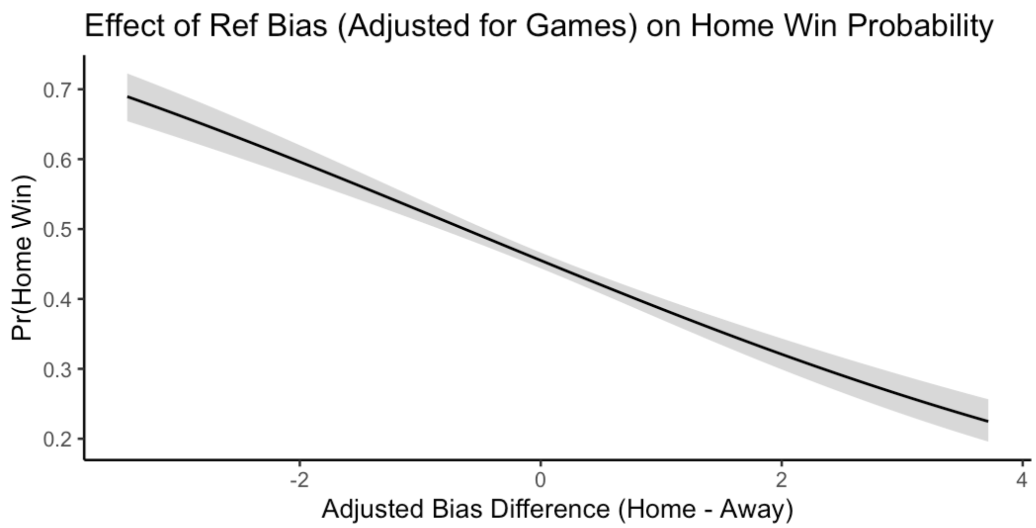

Figure: The probability that the home team wins (y-axis) as a function of the adjusted bias difference (x-axis), fit using logistic regression. The grey regions represent a 95% confidence region.

Why did you choose statistics?

I love numbers and understanding what they might mean. Statistics is a medium through which I can do so.

Statistics allows you to learn interesting things about the world around you. 'Hidden' patterns are everywhere, and often right in your face.

Acknowledgments

I would like to acknowledge my family and friends for the best support in the world. I would also like to thank my professors and advisors for offering their knowledge and wisdom to help me on my academic journey and beyond.

Lydia Gillespie

The Powerball is Random; People's Choices Are Not

I studied whether human number-selection bias influences Powerball outcomes in terms of how often drawings have winners. Specifically, I examined whether draws containing more 'low' numbers (1–31), commonly associated with birthdays, are more likely to have at least one winner. The analysis used Powerball drawing data from 2015 onward, including the five white ball numbers, the Powerball number, draw dates, and the number of winners for each drawing.

I first constructed features such as the count of low numbers, the presence of consecutive numbers, and other characteristics of each draw. Statistical methods included two-sample t-tests to compare draws with/without winners and chi-square tests to check whether drawn numbers are consistent with a uniform lottery process.

The results suggest that while the lottery draws appear uniformly random, draws with more low numbers are significantly more likely to have winners, consistent with players disproportionately choosing birthday-related numbers. Overall, the findings suggest that human biases do not affect which numbers are drawn, but they do influence how often a drawing produces winners.

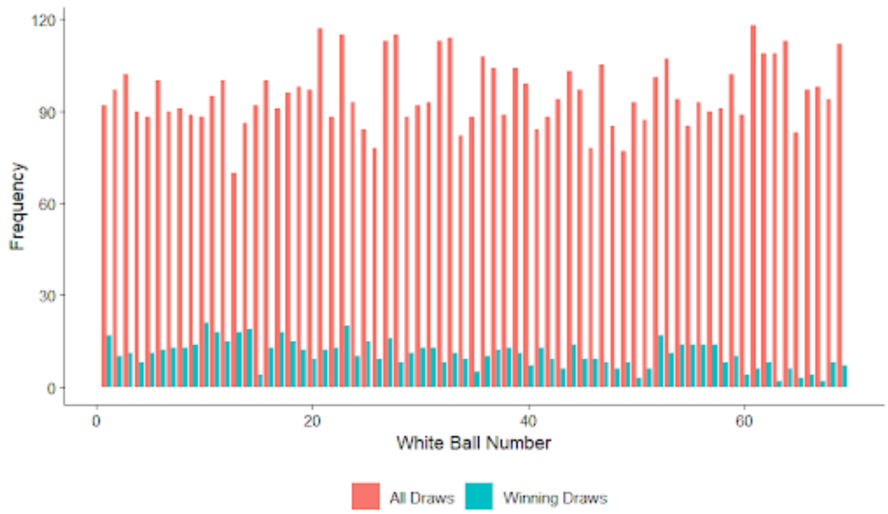

Figure: Histogram showing the distribution of white balls (numbered 1 to 69) is approximately uniform (red bars), while the distribution of white balls among the winning draws (blue bars) is skewed towards lower numbers.

Why did you choose statistics?

I chose to major in statistics because I enjoy working with data and searching for patterns.

The foundation of countless fields revolves around statistics; the versatility of the degree is an added benefit, opening up countless opportunities.

Acknowledgments

I would like to thank all of my professors for their invaluable wisdom.

Thank you to my friends and family for keeping me on track all these years, and thank you, Kyle.

Bethany Gopinath

Translating Multi-Omic Predictive Models into Clinical Practice

I worked with Cassandra Kujawa (page 16) to turn the multi-omic predictive framework we developed in prior work into a real-time, interactive tool for breast cancer prediction. The goal was to make a model that uses both gene activity and clinical information to predict whether a patient's cancer will respond to treatment—and to communicate the results in a simple web app.

We used data from 1,095 breast cancer patients in a public research database (TCGA) to build and test a prediction model. We then developed an R Shiny app where users can upload gene expression and clinical data. The app returns an estimated probability of treatment response and related information.

In testing, the combined gene-plus-clinical model was much better at detecting treatment success than a model using clinical information alone, suggesting gene expression contains important information that standard staging may miss. This tool is a research prototype and would need independent validation and clinical testing before any real-world use, including checks for performance across demographic groups.

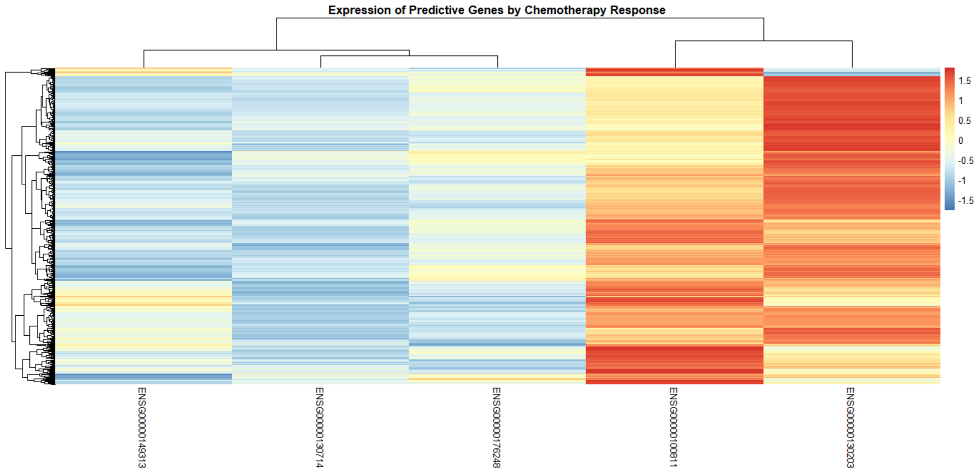

Figure: Hierarchical clustering of gene expression values for the top LASSO-selected predictive genes across patient samples. Red indicates high expression and blue indicates low expression, with distinct clustering patterns visible across patient subgroups.

Why did you choose statistics?

I've always been interested in medicine and healthcare, but I realized I was more drawn to the analytical side of it—understanding patterns in data rather than working directly in a clinical setting. Statistics felt like the perfect bridge between the two. It gave me a way to contribute to medical research and patient outcomes without losing the quantitative rigor I loved, and projects like this one—where the data literally comes from cancer patients—remind me why I chose it.

Acknowledgments

I would like to thank my parents, Gobi Gopinath and Margery Carazzone, for supporting me through my journey. A special thanks as well to Dr. Fadahunsi, Dr. Izmirli, Dr. Auerbach, Dr. Rios, and Dr. Kepplinger for their support, mentorship, and teaching.

Marie Han

The Effect of the Stock Market and Emotion on Political Ideology

I investigated whether emotional media sentiment and the Dow Jones Industrial Average (Dow) affect the political ideology of elected officials. I gathered headline sentiment and stock market data from 1990–2020, along with NOMINATE ideology scores for members of the U.S. House, Senate, and the president. I then fit a vector autoregression (VAR) model to examine how party-level ideology evolves over time in response to sentiment and changes in the Dow.

I find strong evidence that media sentiment Granger-causes stock market movements at medium-term horizons (approximately 1–9 months). Evidence that stock market changes predict shifts in political ideology is weaker and sensitive to model specification.

Overall, media sentiment appears to have predictive power for financial markets, while its relationship with political ideology is less clear and warrants further study. One limitation is that NOMINATE ideology scores update slowly and are fairly coarse, so even with lagging/aggregation the analysis may miss shorter-term changes that would be better captured by more granular measures like roll-call votes or legislative actions.

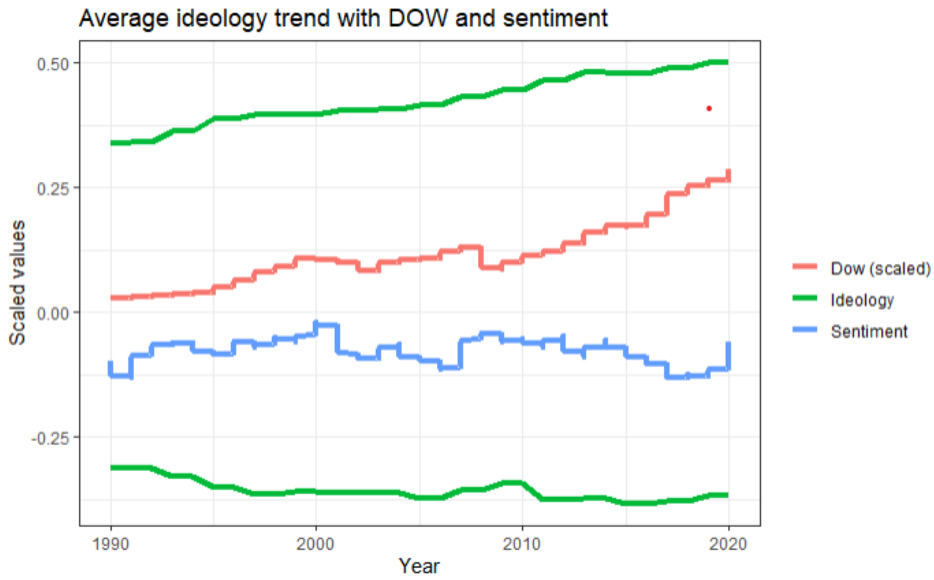

Table: A visualization of three trends between 1900 and 2020. The Dow Jones Industrial Average (red), NOMINATE ideology scores (top green Republican Party, bottom green Democratic Party), and sentiment of headlines in the news (blue).

Why did you choose statistics?

I originally had not intended to be a statistics major. I originally came into college as a sociology major, but a random statistics elective made me change my mind. I like the versatility of statistics. There is not a single field in research, industry, government, etc. where you aren't going to need to know statistics.

Acknowledgments

I'd like to thank my professors for being incredible sources of knowledge and patience. Also my dear friends for all of their support in late night study session and taco bell runs. And of course my family and partner for being supportive while I figure out how to be an adult.

Joshua Kornell

The House Always Wins: Gambler's Ruin and NFL Betting Markets

I revisited the classic Gambler's Ruin problem in the context of modern online sports betting. With sports gambling more accessible than ever, I asked whether the main ideas of Gambler's Ruin still apply and how much a small house edge matters over time.

To answer this question, I used the R package nflreadr to collect NFL game results and closing money-line odds, focusing on the 2023 regular season. I simulated a bettor who starts with a $100 bankroll and places $1 bets on every game, repeating until the bankroll reaches $0 (ruin) or $200 (doubling). I compared three strategies—random picks, always betting the favorite, and betting underdogs in close games—under four house-edge levels (0%, 2.5%, 5%, 10%).

Even with a 2.5% edge, over 85% of simulated bettors eventually went broke, showing that small edges can dominate long-run outcomes. Betting favorites often lost faster because favorites pay less when they win. I also found mild patterns in odds accuracy, but none produced a strategy that reliably overcame the house edge.

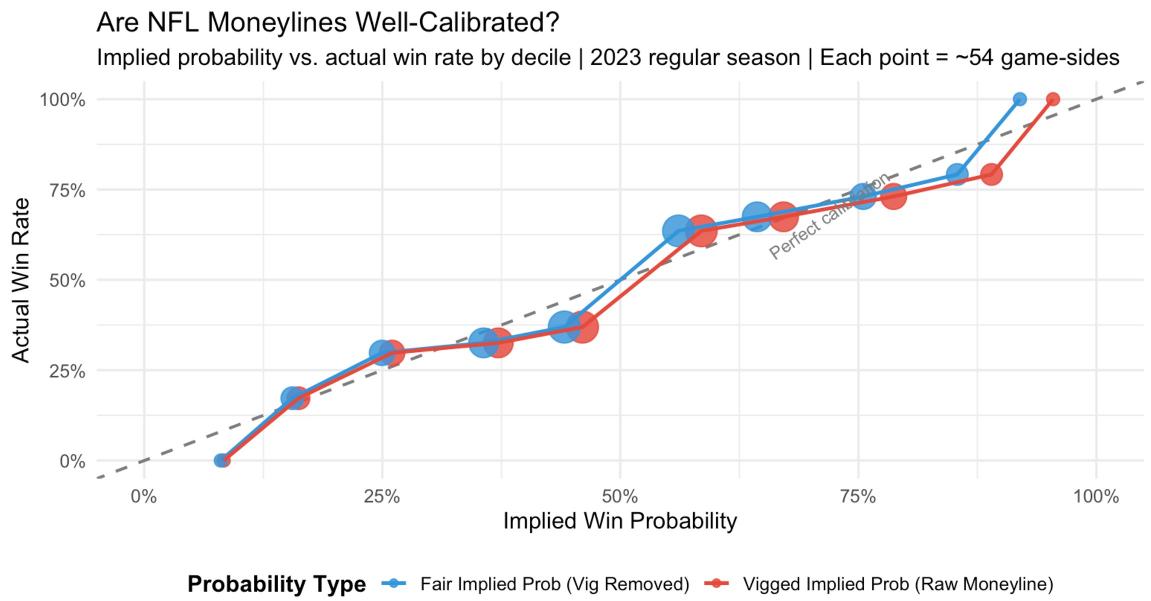

Figure: Bubble plot showing the win probability implied by the betting odds before a game (x-axis) and the actual win rate of those games (y-axis). Points on the identity line (dashed) indicate the advertised odds equal the actual odds. The points deviate from the line near 0%, 50%, and 100%.

Why did you choose statistics?

Probability and statistics were one of the few concepts I could really seep my teeth into when I was in high school. I would help out my friends with the key concepts before our exams, and explaining how it worked made me enjoy it even more. I decided to pursue it as an undergrad in hopes of exploring what I found so interesting about it.

Acknowledgments

All my love goes to my wonderful family and friends who've supported me throughout these four years whether they know it or not, as well as a big thank you to my professors and mentors at George Mason.

Cassandra Kujawa

Translating Multi-Omic Predictive Models into Clinical Practice

I worked with Bethany Gopinath (page 10) to investigate whether breast cancer treatment outcomes, specifically chemotherapy response and patient survival, could be predicted by combining clinical data with gene expression profiles. To answer this, we utilized a multi-omic approach that integrated clinical variables such as age, tumor stage, and treatment status with RNA sequencing data. We analyzed a dataset of approximately 1,100 patients from The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma cohort.

Due to genetic variation between patients and other complexities, predicting treatment success is often difficult. Traditional methods frequently lack the precision to make personalized clinical decisions. In our analysis, we applied several statistical machine learning methods, including a Cox proportional hazards model to identify mortality risk factors, Kaplan-Meier curves to visualize survival probabilities, LASSO logistic regression with ROC curves to classify chemotherapy responders, and KEGG pathway analysis to identify biological mechanisms driving breast cancer treatment outcomes. We found that combining gene expression data with clinical metadata significantly improved predictive power compared to using genomic data alone.

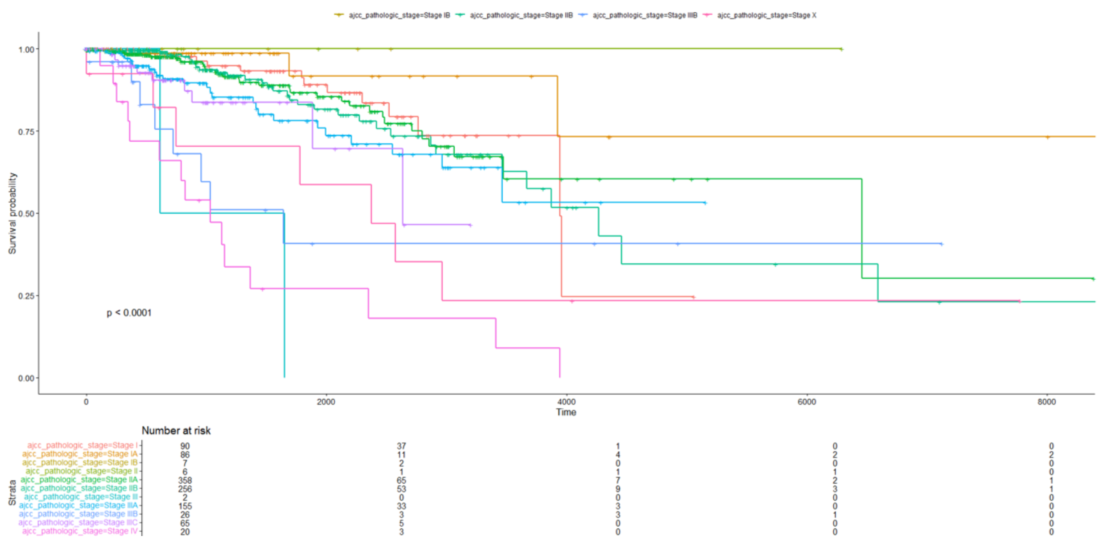

Figure: The Kaplan-Meier curves show how long patients survive after treatment, broken down by cancer stage. Each line represents a sample of individuals with different stages of breast cancer being treated with chemotherapy. The y-axis shows survival probability, and the x-axis tracks time since diagnosis in days. The table below the graph shows how many patients are still alive and are being tracked at each time point.

Why did you choose statistics?

I chose to major in statistics because of my passion for mathematical applications that provide solutions to complex, real-world problems. After taking an introductory statistics course during my first year, I realized that statistics offered the perfect balance between mathematics and programming. I saw it as a discipline that would provide me with the technical skills to build models and explain their results.

Acknowledgments

I would like to thank my family for their support and encouragement.

Isabela Leon

Modeling Time until Adoption of Shelter Dogs using Survival Analysis

I examined which factors influence how long dogs remain in animal shelters before being adopted. Using publicly available data from the Austin Animal Center, I restricted the dataset to dogs and measured each dog's length of stay until adoption. I focused on whether adoption timing differs by characteristics such as age, sex, and coat color.

I used survival analysis to model the time until adoption. I first constructed Kaplan–Meier curves to estimate and compare adoption patterns across different groups. I then fit Cox proportional hazards models to quantify how each characteristic is associated with the adoption rate, allowing direct comparisons of effect size while controlling for other variables.

Age was the strongest predictor: adoption rates decreased substantially as dogs got older. Senior dogs were adopted at roughly half the rate of puppies, indicating a strong preference for younger animals. Sex had a smaller but statistically significant association, with male dogs adopted slightly more slowly than females. Coat color showed little to no meaningful relationship with adoption timing.

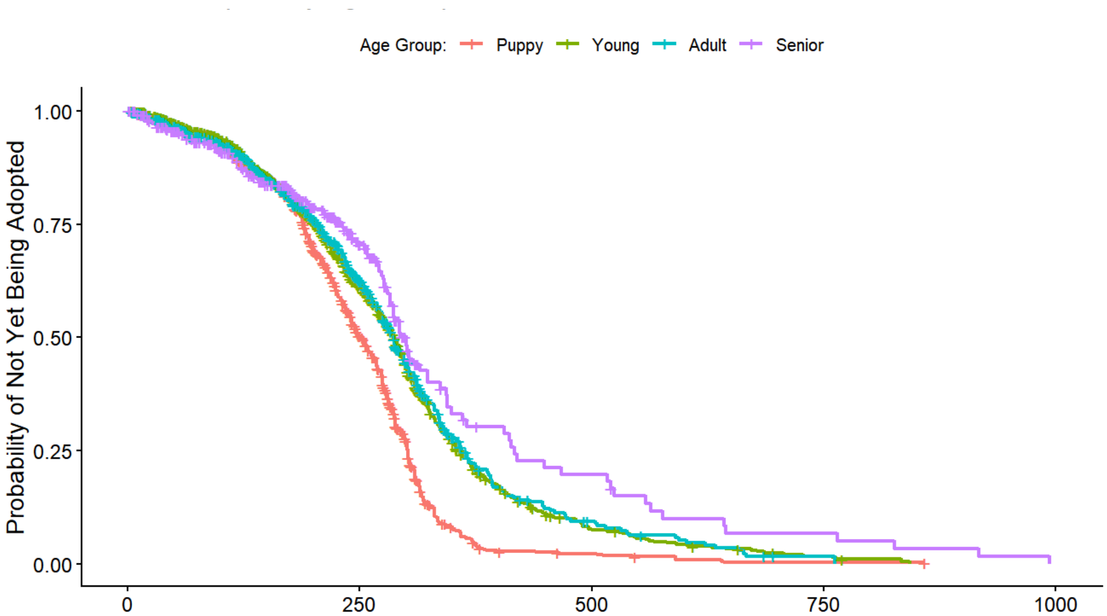

Figure: Kaplan-Meier curves showing the probability of a shelter dog not yet being adopted (y-axis) after a given number of days (x-axis) for four different age groups.

Why did you choose statistics?

I chose to major in statistics following a meaningful experience in my senior year of high school. After being involved in a serious car accident, I faced a difficult recovery, and my statistics teacher went out of his way to support me both academically and personally. His encouragement and kindness inspired me to engage more deeply with the class material, and I found that I genuinely enjoyed the subject.

Acknowledgments

I would like to thank my former statistics teacher, Alex Fogelson, for showing me kindness and inspiring me to explore the world of statistics. I would also like to thank my family and my boyfriend for being my biggest supporters throughout my time at George Mason.

Salma Makhlouf

An Analysis of Why Governments Request User Data from Meta

I investigated the relationship between government requests for user data from Meta between 2019 to 2023 and country-level economic and governance characteristics. Specifically, I examined whether variables such as GDP per capita, unemployment, inflation, and corruption levels (Corruption Perceptions Index) can help explain differences in request activity across countries.

I found that economic development and governance quality are significant predictors of government data requests. PCA revealed that these variables can be reduced to three components. The first component represents overall development and governance. The second component primarily captures inflation-related factors, and the third reflects labor market and economic growth dynamics.

Regression indicates that wealthier and less corrupt countries tend to have higher levels of request activity. While changes in governance over time have a statistically significant effect, their impact is relatively small compared to long-term structural differences across countries. Overall, the findings suggest that persistent economic and institutional characteristics more strongly influence government data requests than short-term changes.

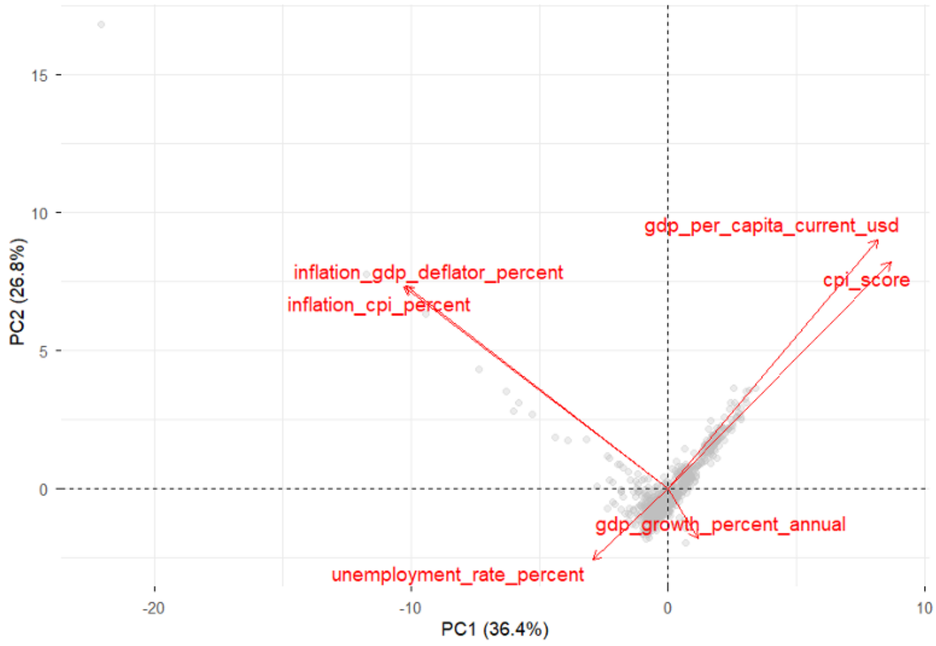

Figure: Biplot illustrating the relationships among economic and governance variables. GDP per capita and CPI score align strongly along PC1, while inflation measures are associated with PC2.

Why did you choose statistics?

I chose to major in statistics because I've always enjoyed math and have been interested in understanding patterns and making sense of data. I like that statistics is both analytical and practical, allowing me to solve real-world problems and draw meaningful conclusions from complex information.

Acknowledgments

I would like to thank my parents and my siblings, especially my sisters, for their constant support and encouragement throughout my academic journey. As a first-generation college graduate, their sacrifices and belief in me have meant everything. I'm especially grateful for the motivation and strength they have given me to reach this milestone.

Annie Nguyen

Measuring Narrative Bias in U.S. Historical Accounts

I developed a quantitative method to analyze perspective in U.S. history texts, using natural language processing (NLP) to measure how closely a narrative aligns with different reference viewpoints.

I analyzed U.S. history periods as recounted in Wikipedia articles alongside two reference texts—*A People's History* (Zinn) and *A Patriot's History* (Schweikart)—which represent contrasting historiographical angles. I embedded the texts using large language models to generate numerical vectors in a high-dimensional space. I then applied PCA to reduce dimensionality and used cosine similarity and regression-based comparisons to place each narrative relative to the two reference texts. This produces interpretable measures of similarity, dissimilarity, and a 'line of neutrality' defined between the references.

I found that defining a line of neutrality makes it possible to study how narrative perspective varies by time and source. A key assumption is that the reference texts capture distinct viewpoints, which may change with context. Future work will apply this framework to high school textbooks across U.S. regions.

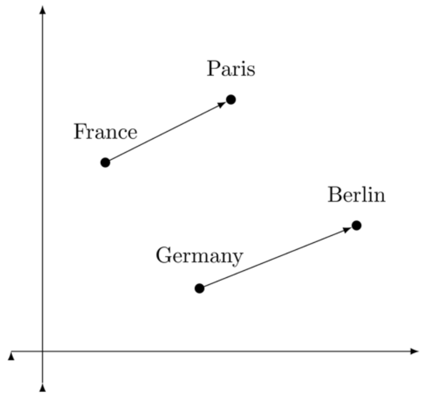

Figure: An illustration of word embeddings from Wikipedia (https://tinyurl.com/364xxn7x). Each word is a point in space, which can capture relationships like country and capital.

Why did you choose statistics?

I chose statistics because I enjoy both learning and applying statistical concepts and methods. After switching my major five times during my undergraduate career, I feel confident that I've finally found the right fit. I hope to continue to pursue life with empathy and passion, and this seems like the right step in that direction!

Acknowledgments

I'm deeply thankful to my parents, my partner Chris, and my friends for being unconditionally supportive of me. I also want to thank Dr. Auerbach, Dr. Fadahunsi, and all my professors and peers for their kindness, guidance, encouragement, and belief in me—especially as I found my place in statistics. And thank you Sue for helping Aleah and I run ASA student chapter events, even with my many last-minute emails. I am so grateful to have found such an amazing community!

Mitchell Nicolai

Isolating the Protective Effect of Physical Activity on Mental Health

I investigated whether physical activity is associated with lower mental distress among adults in Virginia. Using the 2024 Behavioral Risk Factor Surveillance System (BRFSS), I analyzed a representative sample of 6,593 residents. I measured mental distress using self-reported 'poor mental health days' in the past 30 days and focused on frequent mental distress (FMD), defined as 14 or more poor days per month.

I used survey-weighted regression models and, to reduce confounding from observable differences between adults who do and do not exercise, I applied propensity score weighting (PSW), balancing groups on income, education, and baseline health.

I found that Virginia adults who do not exercise have 60% higher odds of FMD (OR = 1.60). After PSW, the estimated gap increased to 3.29 poor mental health days per month, with a larger gap in rural Virginia (2.78 days) than in urban areas (1.34 days). Overall, these results are consistent with physical activity being protective—especially in rural settings—and suggest that improving access to physical activity in rural Virginia could help reduce geographic disparities in mental distress.

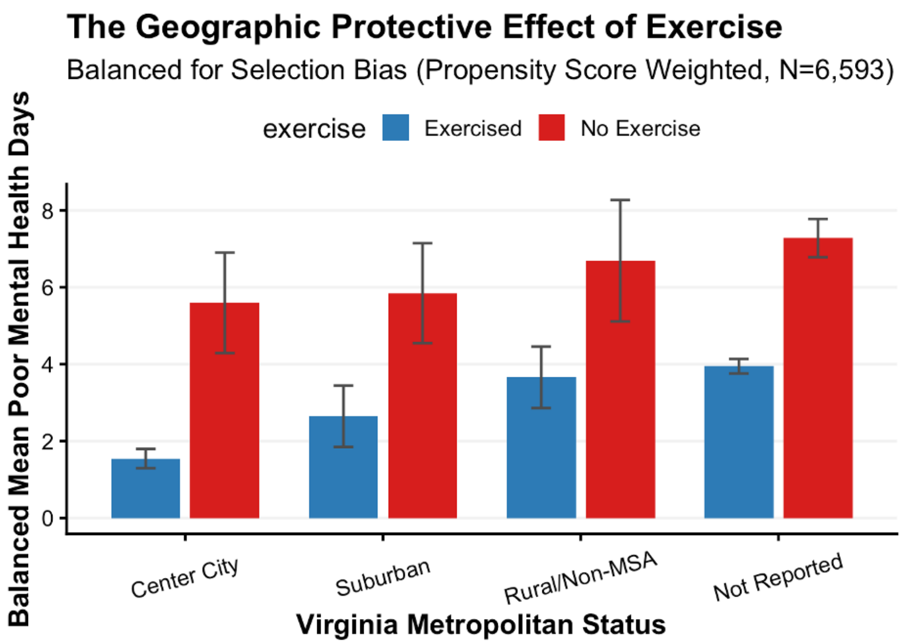

Figure: This chart compares the average number of poor mental health days for those that exercise (blue) and those that do not (red) by location after using propensity score weighting to adjust for baseline differences in income, education, and health.

Why did you choose statistics?

What I like most about statistics is the science of turning uncertainty into a clearer picture of how the world works. I like that it provides the power to quantify 'invisible' barriers, factors like geographic isolation or socioeconomic status that profoundly impact lives but often go unmeasured.

Acknowledgments

I am deeply grateful to my parents for their investment in my education and for giving me the opportunity to pursue my passion for statistics. Finally, to my wife, thank you for being my greatest motivator and for pushing me to cross the finish line. I am incredibly lucky to have you in my corner.

Aleah Talley

Modeling Homelessness as a Dynamic System

I investigated how homelessness is measured in the United States, focusing on the difference between prevalence and incidence.

Point-in-time (PIT) counts—the primary national measure—use a single-night snapshot to estimate prevalence: how many people are experiencing homelessness at one moment. While useful, PIT counts can miss incidence: the ongoing flow of people entering and exiting homelessness over time. As a result, short or repeated episodes may be underrepresented, making some experiences effectively ‘invisible’ in prevalence-only statistics.

To address this, I modeled homelessness data from the U.S. Department of Housing and Urban Development (HUD) using an Erlang A model (M/M/s+M queue). I incorporated observable measures such as shelter capacity, occupancy, and unsheltered counts to infer unobserved system features, including entry (arrival) rates and exit rates (including leaving before being served).
I found that prevalence-based measures can fail to reflect high turnover and instability. A queueing approach provides a complementary way to quantify flow, with implications for measuring need and allocating resources.

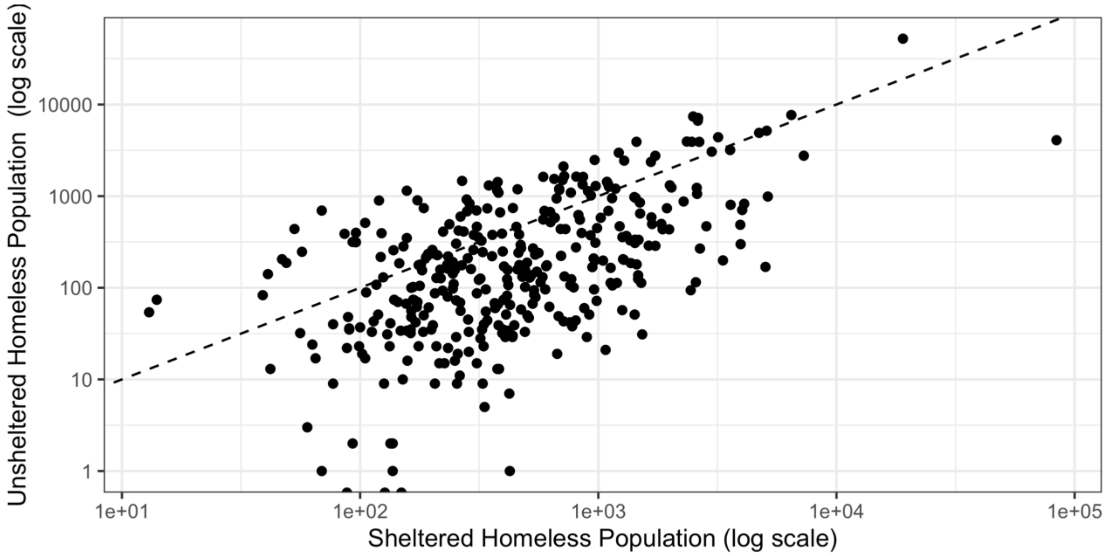

Figure: Scatter plot showing the size of the unsheltered (y-axis) and sheltered homeless populations (x-axis) in 2023 as reported in CoC data from the U.S. Department of Housing and Urban Development. The identity line (dashed line) is provided for comparison.

Why did you choose statistics?

I chose to major in statistics because I was drawn to its ability to make a science of uncertainty. It represents the closest thing the STEM world has to a crystal ball—one that relies not on certainty, but on patterns in historical data to inform the future.

Acknowledgments

I would like to thank Mary Grace Prinster and Dr. Jonathan Auerbach for their expertise and guidance throughout this project. I would also like to thank my dad, whose intuitively scientific curiosity about the world—and frequent reminder that "if you can't measure it, you can't control it"—shaped my understanding of education as a form of empowerment. I will always carry with me the time we spent together working through math problems.

Andrew Ward

Are NFL Referees Biased?

I investigated whether referees in the National Football League (NFL) are biased towards certain teams. For example, I examined whether teams get more favorable calls if they have a larger social media following (measure of popularity). I also test if being the home or the road team affects the number of penalty flags a team gets.

I used NFL referee data from the 2025 season available from nflpenalties.com. I also used data on team Instagram followers before the 2025 season found on List Wire. I applied two statistical tests to determine whether teams are favored based on social media following and whether the home team gets a different number of penalty flags compared to the road team. I used Kendall's tau to test if there is an association between social media following and net penalty yards. I used a two-sample t-test to test the difference of means between home penalty flags and road penalty flags.

Both tests indicated that social media following is not associated with net penalty yards and that there is not strong enough evidence to show that the mean of home penalty flags is different from the mean of road penalty flags.

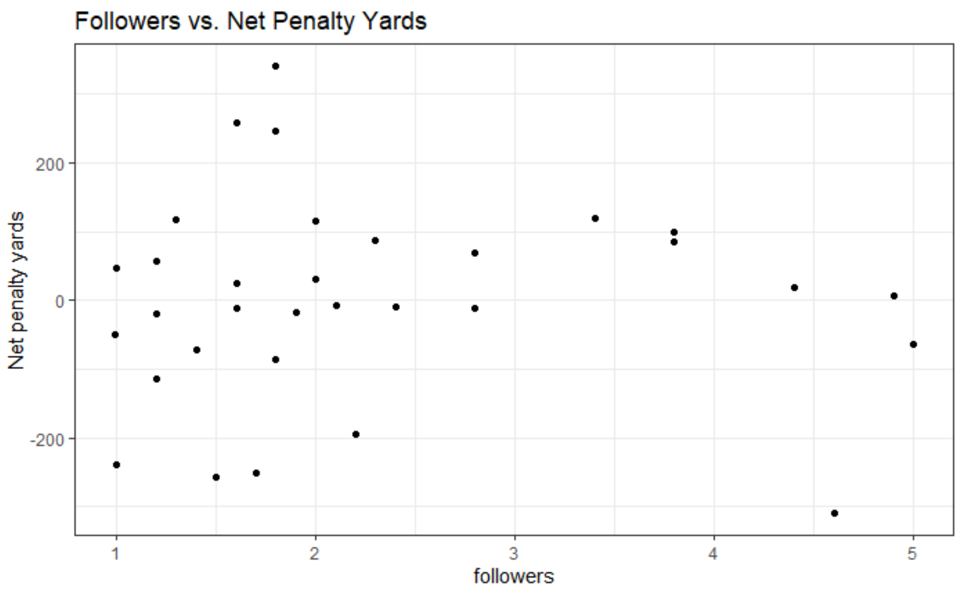

Table: Scatter plot showing the relationship between the net amount of penalties (y-axis measured in net yards) in favor of the home team given by a referee and the number following that team on Instagram (x-axis measured in millions).

Why did you choose statistics?

I choose statistics because I love looking at and analyzing data on different topics. I like how statistics can explain so much about the world around you in a simple way. It helps us answer important questions using data that we encounter in our everyday lives, and I think that's beautiful.

Acknowledgments

I would like to thank my parents for encouraging me to pursue my passions and supporting me throughout my life. I would also like to thank my professors and school teachers for helping me learn the material and opening me up to the relevance of statistics in society.

Cody Salenga

The Business Dilemma: Organic Growth or Merger?

I collaborated with a team of four students from diverse business concentrations under the group name 'Designer Consulting' to analyze a real-world strategic dilemma faced by RHI Magnesita: whether to pursue organic growth or engage in another major merger. Our analysis incorporated financial performance, industry competition, global market pressures, legal constraints, and vertical integration considerations. After evaluating the risks of increased debt, reduced profitability, and integration challenges, we determined that an additional merger would likely introduce more financial and operational strain than long-term value.

Instead, Designer Consulting recommended a focused organic growth strategy centered on strengthening core operations, expanding product offerings, and leveraging existing global assets. We identified opportunities in key industries such as steel, cement, and energy while emphasizing innovation, cost efficiency, and sustainability. Our findings highlighted that strategic internal growth would allow the company to maintain competitive advantage without the risks associated with market over consolidation, regulatory pressure, and cultural misalignment often seen in large-scale mergers.

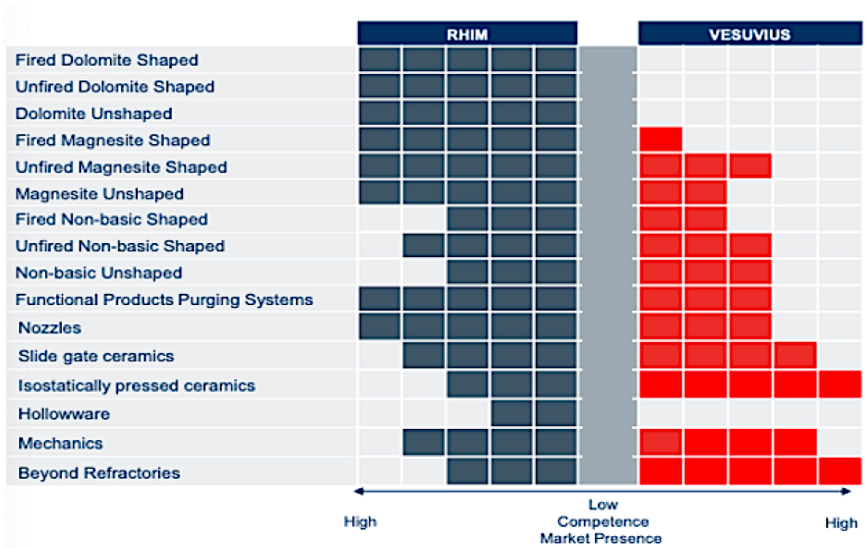

Figure: An assessment of business overlap between two firms, RHI Magnesita (RHIM) and Vesuvius, a potential merger.

Our team, Designer Consulting, advanced through multiple rounds of judging out of more than 75 competing teams, ultimately earning a cash prize of $500 for getting 2nd-place in a case competition hosted by the GMU TV Production Team. We delivered our final presentation live in front of an audience of over 300 attendees online, including students, faculty, and even the Dean of the Costello College of Business.

Acknowledgments

I'd like to thank the editor for giving me a spot in *The Mason Statistician 2026*, as an undergraduate student from the Costello College of Business, now a Full-Time Classified Staff member of the Department of Statistics.

Always build your network, especially with those outside of your discipline; you'll find amazing opportunities for doors to open in your own discipline and beyond!

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 2026 at five locations around the world: Washington D.C., USA; Kyoto, Japan; Vancouver, Canada; and Liestal-Weideli, Switzerland.

Students network in DC-area hackathon

On February 21, 2026, Mason students joined others from across the DC-area to share insights and pose for a photo (image below).

The event was held jointly with the Washington Statistical Society and featured speaker Scott Olesen, Lead Data Scientist at the Center for Forecasting and Outbreak Analytics in the Centers for Disease Control and Prevention (back left between TV and whiteboard in image).

See Fairfax County Times article: https://tinyurl.com/ywfr7n5d and competition website: https://competition.statistics.gmu.edu/

AWARDS & RECOGNITION

Chair's Award for Academic Excellence presented to Cole Boller-Pinkham

for overall excellence as determined by a nomination from the Department of Statistics Chair and vote of the faculty.

Senior Leadership and Service Award presented to Annie Nguyen and Aleah Talley

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

Statistics Faculty Award presented to Sarah Magdeburg

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

Best Prediction from Cassandra Kujawa

who, among Mason contestants in the Cherry Blossom Prediction Competition, most accurately predicted the bloom dates of cherry trees at five locations around the world.

ASA CHAPTER AT MASON

Images: Mason alums Abby Finch (left) and David Buxton (right) recognized with the All-Star award at the Sports Analytics Networking event on November 5, 2025 (image above). They joined Cole Boller-Pinkham on a panel moderated by Brett Hunter (top image on facing page). A second panel from Sports Management included moderator Craig Esherick, Derek Vigon, and Faith Brown (from left to right in bottom image on facing page). Source: https://tinyurl.com/2s45snan

Event organized by ASA Chapter leadership Aleah Talley (president) and Annie Nguyen (vice-president).

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