JIYE CHOI

Washington, D.C., USA • jiyechoi.data@gmail.com • +1 (703) 859-5867 • jiyechoi-data • jiyechoi-data.github.io

EDUCATION

Georgetown University **Washington, D.C., USA**
Master of Science, Biostatistics Aug. 2025 - Dec. 2026 (Expected)
Gyeongsang National University **Jinju, Korea**
Bachelor of Science, Information and Statistics Feb. 2019

- **Scholarship:** President's Recommendation Scholarship (Fall 2018); Globalization Scholarship (Fall 2018); Academic Excellence Scholarship (Fall 2016, Spring & Fall 2015, Fall 2014); Study-Abroad Scholarship (Spring 2015)
- **Relevant Coursework:** Survival Analysis; Statistical Inference; Linear Models & Multivariate Analysis; Categorical Analysis; Probability & Sampling; Principles of Epidemiology; Statistical Software(R, SAS)
- **Research Practicum** (in progress), Department of Biostatistics

TECHNICAL SKILLS

- **Programming**: R, SAS, Python, SPSS, AMOS
- **Database Management**: SQL, MySQL, PostgreSQL
- **Version Control**: Github
- **Data Visualization**: ggplot2, matplotlib, seaborn

PROJECT EXPERIENCES

Bank Customer Behavior Analysis Using SAS Enterprise Miner and Association Rule Discovery Nov. 2018

- Performed association rule mining in SAS Enterprise Miner, analyzing 7,991 customer transactions to uncover patterns between banking services and web navigation behaviors.
- Applied metrics such as support, confidence, and lift to identify key customer segments, revealing that customers using saving accounts were 1.02 times more likely to also hold checking accounts..
- Proposed data-driven marketing strategies based on association rules, suggesting targeted promotions for saving account users to increase engagement with checking account services by up to 1-2%.

Regional Birth Rate Disparities Analysis in South Korea Using R and Multivariate Statistical Methods Jun. 2018

- Executed Principal Component Analysis (PCA) and Factor Analysis in R, as part of a team of four students, developing an "Urban Development Index" and "Demographic Factors," explaining 80% of variance in birth rate disparities.
- Identified distinct clusters of districts using Cluster Analysis and Multidimensional Scaling, uncovering 2-4 socioeconomic groupings across 69 districts.
- Pinpointed key drivers such as urban development, financial autonomy, access to childcare facilities, and education quality, showing that districts with higher urban development indices and greater access to childcare and educational infrastructure had birth rates up to 15-20% higher than districts with lower development levels.
- Explained approximately 36% and 17% of the variance in birth rate disparities through higher financial autonomy and cultural infrastructure, respectively.

Forecasting Jeju Island Tourism Revenue Using SAS, ARIMA, and Transfer Function Models Jun. 2018

- Conducted time series analysis in SAS, utilizing ARIMA and transfer function models on 36 months of foreign tourist arrival data to forecast Jeju tourism revenue, incorporating variables such as tourist arrivals, exchange rates, and promotions..
- Developed a predictive model with an accuracy rate of over 85%, identifying key factors such as a 10% fluctuation in foreign tourist arrivals correlating with a 3.96% change in tourism revenue. Additionally, significant impacts were observed during national holidays and international promotional campaigns.
- Developed a predictive model with an accuracy rate of over 85%, identifying key factors such as foreign tourist arrivals, which showed that a 10% fluctuation correlated with a 3.96% change in tourism revenue.

Facebook Page Analysis of "National Geographic" Using R, Text Mining, and Topic Modeling on Linux Jun. 2016

- Extracted and processed real-time data from Facebook's API using R on a Linux platform, analyzing over 2,000 posts from the "National Geographic" Facebook page to identify key audience engagement themes.
- Applied word frequency analysis, word cloud generation, and Latent Dirichlet Allocation (LDA) topic modeling, uncovering 4 prominent topics such as wildlife, photography, environmental themes, and global exploration, with wildlife-related content comprising a significant portion of total engagement.
- Discovered that posts containing keywords related to wildlife and photography, such as 'photo' and 'animal', were among the top drivers of engagement, providing actionable insights for optimizing content strategy and audience targeting.

Analyzing Monthly Living Expenses Over Time Using AMOS and Latent Growth Curve Modeling Jun. 2016

- Applied Latent Growth Curve Modeling (LGCM) and Structural Equation Modeling (SEM) in AMOS to analyze 3 years of panel data, focusing on the effects of job status and savings on monthly living expenses.
- Modeled the trajectory of living expenses, revealing that individuals with higher initial savings had 0.399 higher baseline expenses, though savings did not significantly impact the rate of expense increase over time.
- Identified a consistent increase in living expenses over the 3-year period, with an average slope of 27.6, indicating a significant annual increase in living expenses regardless of savings levels.