Hsuan Lee

Personal Website | LinkedIn | @ hsuan.lee1@gmail.com | GitHub | Taoyuan, Taiwan

Education

Utrecht University — Utrecht, the Netherlands
M.Sc. in Methodology and Statistics (Research Master)
Minor: Data Science — *Sep. 2021 – Jun. 2023*

- **Network Science Research Group**
- **Data Science Research Group**

Soochow University — Taipei, Taiwan
B.A. in Sociology — *Sep. 2016 – Jun. 2020*

- **Recipient of Departmental Award for the Best B.A. Thesis**
- **Speaker for the Office of International Academic Exchange**

Southeast University — Nanjing, China
University-Level Student Exchange Program in Social Science — *Feb. 2019 – Jun. 2019*

Charles University — Prague, Czech Republic
Faculty-Level Student Exchange Program in Social Science — *Sep. 2018 – Feb. 2019*

Technical Skills

Analytical Skills: Data Science & Machine Learning, Mathematical, Bayesian, Multivariate, & Biostatistics, Causal Inference, Structural Equation & Multilevel Modeling, Psychometrics, Survey Data Analysis

Languages: R (Advanced), Python, SQL

Statistical Software: Mplus, SPSS, JAGS, JASP

Research-facilitating Software: Power BI, Github, LaTeX, Zotero, Rmarkdown, Visual Studio, MS Office programs

Experience

Statistical Consultant — Utrecht, the Netherlands
Methodology and Statistics, Utrecht University — *Apr. 2023 – Jun. 2023*

- Provided statistical consulting for undergraduate and master students at Utrecht University.

Network Science Student Assistant — Utrecht, the Netherlands
Methodology and Statistics, Utrecht University — *Jan. 2023 – Mar. 2023*

- Assisted Dr. Mahdi Shafiee Kamalabad in preparing, updating, and inspecting materials for the social network section of the Network Science course.
- Assisted a PhD student in simulating social network data, i.e. Relational Event History Data.

Multivariate Statistics Teaching Assistant — Utrecht, the Netherlands
Methodology and Statistics, Utrecht University — *Aug. 2022*

- Assisted Dr. Dave Hessen in teaching multivariate statistics covering: t-test, one/two way ANOVA, Linear Regression, ANCOVA, Advanced Linear Regression (one-hot encoding), one/two way MANOVA, Repeated Measures Analysis, Logistic Regression, and Principle Component Analysis.

Projects

Changepoint Detection in Social Networks: an Extension of the Relational Event Model

Sep. 2022 – Jun. 2023 | *GitHub*

- Conducted independently. Addressed the limitations of the Relational Event Model by integrating various changepoint detection methods into the Moving Window Approach. This extension enhanced the model's capability to identify temporal changes within social networks, improving the detection of significant shifts in network dynamics over time.

Machine Learning Forecasting of IMDb Top 1000 Hits: 2022 Films

Oct. 2022 – Nov. 2022

- Collaborated with Ana Martins and Timo van Veghel. Analyzed IMDb data for 2022 films, examining 14 features including director, actors, genre, and film length. Applied machine learning techniques (logistic regression, bagging, random forest, boosting) to predict potential Top 1000 hits. Identified the most accurate prediction methods and assessed the potential of 2022 releases for Top 1000 ranking.

New York City Rat Sightings: Data Visualization Analysis

Sep. 2022

- Worked with Ana Martins and Timo van Veghel. Analyzed NYC rat sightings using data visualization techniques. Investigated correlations between sightings and months, and assessed the relationship between borough population density and rat sightings. Visualized results with interactive scatter plots, line and bar charts, and identified boroughs with the fastest rat sighting resolutions.

Bayesian Modeling of Body Fat Prediction: Age and Height Factors

May. 2022 – Jun. 2022

- Conducted independently. Analyzed body fat data from 252 men using Bayesian linear regression to assess the influence of age and height. Employed the Metropolis-Hastings sampler with non-standardized t-distributions as priors. Conducted convergence diagnostics and compared model effectiveness using DIC and Bayes factors.

Choletrapib Effect on HDL Cholesterol: Results from a RCT in Dyslipidemia Patients

May. 2022 – Jun. 2022

- Collaborated with Daniel Anadria and Lela Roos on a double-blinded RCT involving 485 dyslipidemia patients to assess Choletrapib's impact on HDL cholesterol. Analyzed HDL levels at baseline, 1 year, and 2 years using ANCOVA. Observed a significant increase in HDL cholesterol in the treatment group compared to the control.

Imputation Strategies for Household Income Data in the European Social Survey

Dec. 2021 – Jan. 2022

- Collaborated with Daniel Anadria, Florian van Leeuwen, and Katja Sonntag to explore imputation strategies for household income data ('hinctnta') in ESS Round 9 for Italy. Compared categorical and continuous imputation methods to address high item non-response rates. Found categorical imputation better preserved the decile structure and accuracy of income data.

Languages

- **Chinese** (Native)
- **English** (Proficient)
- **Taiwanese** (Elementary)