RONG LI

Email: rong.li.rl946@yale.edu
Github: https://github.com/lirong95

Education

- **Yale University** — New Haven, CT, United States
  Postdoctoral Associate, Department of Biostatistics — *2022 – Now*
  ***Supervisor**: Dr. Shuangge(Steven) Ma.*
  Fellow, Yale-Boehringer Ingelheim Biomedical Data Science Fellowship
  ***Mentors**: Dr. Shuangge(Steven) Ma, Dr. Zuojian Tang*
- **Renmin University of China** — Beijing, China
  PhD, School of Statistics — *2017 – 2022*
  ***Thesis**: Evaluation on Variable Selection Uncertainty under Model Misspecification.*
  ***Supervisor**: Dr. Yang Li.*
- **Minzu University of China** — Beijing, China
  B.S., Applied Statistics — *2013 – 2017*

Research interest

Statistical Method: Heterogeneity analysis, high dimensional analysis, model uncertainty, network analysis, variable selection
Statistical Application: Cancer heterogeneity, drug discovery, genetic epidemiology, multi-omics data integration, precision medicine, visualization

Papers

- **Journal Articles**:
  - **Li, R.**, Zhang, Q.*, Ma, S.* (2024+). Regulation-incorporated Gene Expression Network-based Heterogeneity Analysis. *Statistica Sinica*. Accepted.
  - **Li, R.**, Xu, S., Li, Y., Tang, Z., Feng, D., Cai, J., Ma, S.* (2024+). Incorporating Prior Information in Gene Expression Network-based Cancer Heterogeneity Analysis. *Biostatistics*. Accepted.
  - Qin, Y., Wang, L., Li, Y., **Li, R.*** (2023). Visualization and Assessment of Model Selection Uncertainty. *Computational Statistics & Data Analysis*. 178:107598.
  - **Li, R.**, Zhang, W., Li, Y., Lin, C.* (2022). Survival Analysis of Large-scale Tumor Data based on Deep Learning (in Chinese). *Chinese Journal of Health Statistics*. 39(1):84-86.
  - Li, Y., **Li, R.**†, Lin, C.*, Qin, Y., Yang, Y. (2021). Robust Group Variable Screening based on Maximum Lq-likelihood Estimation. *Statistics in Medicine*. 40(30):6818-6834.
  - Li, Y., Wang, F., **Li, R.**, Sun, Y.* (2020). Semiparametric Integrative Interaction Analysis for Non-small-cell Lung Cancer. *Statistical Methods in Medical Research*. 29(10):2865-2880.
  - Li, Y., **Li, R.**, Lin, C.*, Qin, Y., Ma, S. (2019). Penalized Integrative Semiparametric Interaction Analysis for Multiple Genetic Datasets. *Statistics in Medicine*. 38(17):3221-3242.
  - Li, Y., **Li, R.**, Wu, M.*, Qin, Y., Ma, S. (2019). Integrative Interaction Analysis using Threshold Gradient Directed Regularization. *Applied Stochastic Models in Business and Industry*. 35(2):354-375.

  *Remark: * corresponding author; † contributed equally to the first author*

- **Submitted & Revised Articles**:
  - **Li, R.**, Qin, Y.†, Li, Y.*, Ma, S.* Uncertainty Assessment for Generalized Linear Models via Local Bootstrapping. Revised, *SCIENCE CHINA Mathematics*.
  - Wang, J., **Li, R.**, Chang, W., Hsiao, K., Shia, B., Ma, S.* Heterogeneous Network Analysis of Disease Clinical Treatment Measures via Mining Electronic Medical Record Data. Revised, *The Annals of Applied Statistics*.
  - **Li, R.**, Qin, Y., Li, Y.* Assessing Estimation Uncertainty under Model Misspecification. Submitted, *Scandinavian Journal of Statistics*.

- Im, Y., **Li, R.**, Ma, S.* Bayesian Modelling of Cancer Outcomes Using Genetic Variables Assisted by Pathological Imaging Data. Submitted, *Statistics in Medicine*.

- **Working Papers**:
  - Hierarchical Sparse Graphical Model with Latent Variables. *In progress*
    In this work, we propose regularization-based method to identify multiple levels of gene expression networks (i.e., direct, after accounting for co-regulation, and after accounting for latent factors) in a single step.
  - Unsupervised Clustering for Tumor Immune Archetypes in Pan-cancer Data. *In progress*
    This is a collaborative work with Boehringer Ingelheim, develop novel heterogeneity analysis method and apply it to pan-cancer dataset. The goal is to discover the tumor immune archetypes for cancer patient, identify type-specific gene signatures and therapeutic targets.
  - Bootstrap Hypothesis Testing for Model Selection. *In progress*.
    In this work, we extend the hypothesis testing of parameters to models. We conceptualize the null hypothesis as $H_0 : m = m_0$, which contains only true actively predictors, and estimate the null distribution by a data-based bootstrapping algorithm. Depending on the survival models, we have an idea of the discrete distribution of selected model.

- **Chapters in Books**:
  - A Selective Review of Network Analysis Methods for Gene Expression Data. **Li, R.**, Yi, H., Ma, S.
  - Model Evaluation Visualization. In: *Visualization in Data Science* (in Chinese).
  - Feature Selection of Heterogeneous Survey Data. In: *Modern Survey Analysis* (in Chinese).
  - Model Selection. In: *Regression Analysis* (in Chinese).

Presentations

- **JSM 2024 (Joint Statistical Meetings)**: Poster Session, August 2024. Portland, OR, United States.
- **AIME 2024 (The 22nd International Conference Artificial Intelligence in Medicine)**: Drug Discovery Workshop, July 2024. Salt Lake City, UT, United States.
- **NESS 2024 (The 37th New England Statistics Symposium)**: Invited Session, May 2024. University of Connecticut, CT, United States.
- **ICSA 2023 China Conference**: Poster Session, June 2023. Chengdu, China.
- **ENAR 2023 Spring Meeting**: Invited Session, March 2023. Nashville, TN, United States.
- **The 7-th Academic Seminar of Beijing Biomedical Statistics and Data Management Research Association**: December 2021. Beijing, China.
- **The 7-th IMS-China International Conference on Statistics and Probability**: July 2019. Dalian, China
- **Big Data and Business Intelligence Seminar and Cross Strait Statistical Analysis Seminar**: May 2018. Taipei

Experience

- **Research Associate**
  Renmin University of China
  Cooperated with enterprises to solve concrete problems with data and statistical methods
  - Decision and Forecast
    Geographic Information Analysis and Sales Forecast for Gas Station *2019*
    Collaboration with CPPEI to develop an analytical model of a rating system for gas station using geographic POI information. Based on the operating data and the spatial information from gas stations to predict oil and non-oil sales using factor analysis, clustering and classification methods.
  - Biomedical Statistics
    Influenza and TCM Early Warning System Construction *2018*
    Collaboration with the Beijing Hospital of Traditional Chinese Medicine to monitor influenza patterns in the Beijing region and build a TCM early warning system. Forecast the future incidence of the four influenza viruses during the flu season based on the weather conditions using the random forest with rolling window.

- **Manager Assistant**
  Statistical Consulting Center *2019 - 2022*
  - Maintained for Linux Server, in charge of user management and routine maintenance

- Date crawling and downloading from publicly available databases, including The Cancer Genome Atlas Program (TCGA) and Surveillance, Epidemiology, and End Results (SEER)
- Assistance in the preparation of manuscripts and presentations for scientific meetings

- **Teaching Assistant**
  Renmin University of China *2018 - 2019*
  - Multivariate Statistical Analysis
  - Applied Regression Analysis

Honors and Awards

- Outstanding Graduates Award *2022*
- First Price of "BeiGene" Excellent Paper for Youth *2021*
- National Scholarship for Graduate Students of China (Master) *2018*
- Top Ten Papers of the 3-rd National Postgraduate Statistics Forum *2017*

Journal Review

- Statistics in Medicine
- Annals of the Institute of Statistical Mathematics
- Science China
- Statistics and Its Interface
- Biostatistics and Epidemiology
- Journal of Data Science