

RONG LI (李 嵘)

Date of Birth: August 1995

Email: rong.li.rl946@yale.edu

GitHub: <https://github.com/lirong95>

EDUCATION

- Yale University** New Haven, CT, United States
 - Postdoctoral Associate, Biostatistics, School of Public Health* 2022 – Present
 - 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, Statistics, 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, School of Science* 2013 – 2017

RESEARCH INTEREST

Statistical Method: Heterogeneity analysis, high dimensional analysis, model uncertainty, network analysis, variable selection

Statistical Application: Cancer heterogeneity, biomarker discovery, genetic epidemiology, multi-omics data integration, visualization

PAPERS

- Journal Articles:**
 - Li, R., Qin, Y.[†], Li, Y.* , Ma, S.* (2025) Uncertainty Assessment for Generalized Linear Models via Local Bootstrapping. *SCIENCE CHINA Mathematics* (in Chinese). Online.
 - Im, Y., Li, R., Ma, S.* (2025) Bayesian Modeling of Cancer Outcomes Using Genetic Variables Assisted by Pathological Imaging Data. *Statistics in Medicine*. 44(3-4):e10350.
 - Wang, J., Li, R., Chang, W., Hsiao, K., Shia, B., Ma, S.* (2025) Heterogeneous Network Analysis of Disease Clinical Treatment Measures via Mining Electronic Medical Record Data. *The Annals of Applied Statistics*. 19(1):637-654.
 - Li, R., Xu, S., Li, Y., Tang, Z., Feng, D., Cai, J., Ma, S.* (2025). Incorporating Prior Information in Gene Expression Network-based Cancer Heterogeneity Analysis. *Biostatistics*. 26(1):kxae028.
 - Li, R., Zhang, Q.* , Ma, S.* (2024). Regulation-incorporated Gene Expression Network-based Heterogeneity Analysis. *Statistica Sinica*. Online.

- Qin, Y., Wang, L., Li, Y., **Li, R.*** (2023). Visualization and Assessment of Model Selection Uncertainty. *Computational Statistics & Data Analysis*. 178:107598.
- **李嵘**, 张文丽, 李扬, 林存洁 * (2022). 基于深度学习的大规模肿瘤数据生存分析. *中国卫生统计*. 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.* Assessing Estimation Uncertainty under Model Misspecification. Major Revised, *Scandinavian Journal of Statistics*.
- Zheng, T., **Li, R.**, Wu, M., Ma, C.* Accommodating Spatial Heterogeneity in Geographically Weighted Regression with Group Penalty. Major Revised, *Statistics in Medicine*.
- **Li, R.**, Zhang, Q.*, Ma, S.* Conditional Graphical Models With A Hierarchical Sparse Estimation. Submitted, *Statistics and Computing*.

• Working Papers:

- Heterogeneity via Network-based Block-wise Regulation Analysis. *In progress. Joint work with Ma, S.*
We propose a heterogeneity analysis framework based on network and regulation accounting for block structures among predictors. This approach can simultaneously identify sample subgroups, block structures, as well as subgroup-specific networks and regulatory mechanisms.
- Unsupervised Clustering for Tumor Immune Archetypes in Pan-cancer Data. *In progress. Joint work with Tang, Z., John, B., Cai, J.*
This is a collaborative work with Boehringer Ingelheim, aiming at development of novel heterogeneity analysis method and its application on tumor immune archetypes for pan-cancer dataset.
- Bootstrap Hypothesis Testing for Model Selection. *In progress. Joint work with Qin, Y., Li, Y.*
In this work, we extend the hypothesis testing of parameters to models, using data-based bootstrapping algorithm to approximate 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. In: *Gene Expression Analysis: Method and Protocols (Second Edition)*. Springer Nature.
- 异质调查数据的特征选择. In: 复杂调查设计与建模. 科学出版社.
- 模型选择. In: 回归分析. 中国人民大学出版社.

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

Fellow

- *Yale University and Boehringer Ingelheim Biomedical Data Science Fellowship Program*

Multimodal network-based cancer heterogeneity analysis

2022 – Present

Develop more effective statistical learning methods for cancer heterogeneity analysis, effectively integrate multimodal data and incorporate gene regulatory network. Apply post-clustering techniques to identify representative genes as potential biomarkers for immunotherapy.

Research Associate

- *Renmin University of China*

Gas Station Evaluation System

2019

Collaborated with CPPEI on decision-making for gas stations. First, using factor analysis to build index. Second, developing K-means method with Gower distance for unsupervised clustering. Third, applying random forest and SVM for classification. Last, using Bootstrap method for interval prediction.

Influenza Early Warning System

2018

Collaborated with the Beijing Hospital of Traditional Chinese Medicine to monitor and predict influenza patterns. First, applying random forest to predict incidence cases. Second, developing time series prediction with rolling window technique.

- **Manager Assistant**

- *Statistical Consulting Center*

2019 – 2022

- Maintained for Linux Server, in charge of user management and routine maintenance
- Date crawling and extracting 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
- SCIENTIA SINICA Mathematics
- Briefings in Bioinformatics
- Statistics and Its Interface
- Biostatistics and Epidemiology
- Journal of Data Science