# LINQUAN MA

+1 608 736 0567 ◇ mlqmlq1997@gmail.com

1300 University Avenue, Madison, WI, 53706, U.S.

Personal website: `https://mlqmlq.github.io`

## EDUCATION

**Department of Statistics, University of Wisconsin-Madison**  Aug. 2017 - May 2019

M.S. in Data Science

GPA: **4.0**/4.0; Rank: **1**/42 (out of all students in my program)

Core Courses: Statistical Inference I, II, Linear Models, Design and Analysis of Experiments, Time Series Analysis, Non-parametric Statistics, Classification and Regression Trees.

Ongoing Courses: Mathematical Statistics I (Ph.D. level), Data Science Computing Project.

**School of Gifted Young, University of Science and Technology of China** Aug. 2014 - May 2018

B.S. in Statistics

Major GPA: **3.9**/4.3; Rank: Top **5**/51 (out of all students studying statistics at School of Gifted Young)

Core Courses: Regression Analysis (A), Applied Statistical Software (A), Probability Theory (A), Mathematical Statistics (A), Stochastic Processes (A-), Real Analysis (A-), Complex Analysis (A), Functional Analysis (A-), Mathematical Analysis I, II (A, A), Linear Algebra I, II (A-, A-).

## RESEARCH INTERESTS

Sufficient dimension reduction, penalized regression, causal inference and missing data analysis.

## PUBLICATIONS

- **Ma, L.**, Yin, Y., Liu, L., Geng, Z. (2017). On the individual surrogate paradox. *Submitted to Biostatistics*. Preprint version: `http://arxiv.org/abs/1712.08732`

- **Ma, L.**, Liu, L., Yang, W. (2018). Envelope method with ignorable missing data. *To be submitted*. Slides: `https://mlqmlq.github.io/envelope.pdf`

- Shi, Y., **Ma, L.**, Liu, L.. Efficient multivariate mixed regression in task fMRI study. *In manuscript*.

## RESEARCH EXPERIENCE

**On the Individual Surrogate Paradox**  Apr. 2017 - Dec. 2017

*Adviser: Lan Liu, School of Statistics, University of Minnesota*

- Proposed a new methodology to evaluate the surrogate paradox based on an individual perspective.
- Proved the individual surrogate paradox can be excluded under certain criteria.
- Derived the sharp bounds of heterogeneity effect by applying simplex algorithm on the dual problem.
- Performed real data analysis on a clinical trial data.
- Built an R package for practitioners interested in our method.

**Envelope Method with Ignorable Missing Data**  Oct. 2017 - Nov. 2018

*Adviser: Lan Liu, School of Statistics, University of Minnesota*

- Developed a method for envelope estimation with both the predictors and the responses missing at random.
- Incorporated the envelope structure in the expectation-maximization (EM) algorithm.
- Proved the superiority of the EM envelope over the standard EM in the sense of asymptotic variance, and confirmed that in the simulations.

- Applied our method to the Chronic Renal Insufficiency Cohort (CRIC) study using bootstrap; Discovered two new significant biomarkers and got shorter confidence intervals.

### Efficient Multivariate Mixed Regression in Task fMRI Study                 Aug. 2018
*Adviser: Lan Liu, School of Statistics, University of Minnesota*

- Invited to realize the proposed method in R codes.

### Max Independent Component Analysis (Ongoing)                 May 2018 - Present
*Adviser: Zhengjun Zhang, Department of Statistics, University of Wisconsin-Madison*

- Recovering independent components in the existence of max operator in the linear system.
- Applied fast Fourier transformation to analyze signals in the frequency domain.
- Proposed a method of recovering the mixing matrix $\mathbf{A}_{p \times p}$ by the joint distribution when $p = 2$.
- Working on generalizing the method when $p \geq 3$.

## COURSE PROJECTS

### Machine Learning with Large Scale Data Set Containing Missing Value  Oct. 2017 - Dec. 2017
*Adviser: Wei-Yin Loh, Department of Statistics, University of Wisconsin-Madison*

- Handled high-dimensional National Birth Rate data set (several GB) through the server to predict whether a new born infant is underweight.
- Imputed missing values using MICE and GUIDE.
- Assessed various machine learning algorithms: logistic regression, regression trees, random forest, SVM, LDA, GUIDE, XgBoost and ensemble learning.
- Selected the best model by cross validation, and showed that ensemble learning outperforms other methods.

### Statistical Inference on the Gene Expression                 Dec. 2017
*Adviser: Zhengjun Zhang, Department of Statistics, University of Wisconsin-Madison*

- Applied cross validation for tuning the lasso penalty parameter.
- Utilized Generalized Measures of Correlation (GMC) method for model selection.
- Detected the global maxima of the non-convex objective function by grid search of the starting points.

## AWARDS

| | |
|---|---:|
| Student Academic Excellence Award, UW-Madison | 2017, 2018 |
| Outstanding Student Scholarship, USTC | 2014, 2015, 2016, 2017 |

## SKILLS

**Programming:** R, LaTeX, Matlab, Bash, Python, C++.

**Other:** Linux (Ubuntu), Mac OS, Windows OS.

## MISCELLANEOUS

**Talk:** Presented the envelope method in Prof. Zhengjun Zhang's group meeting.

**TA Experience:** Hold office hours for the graduate level course *Statistical Inference.*                 Fall 2018

**Seminars:** Got familiar with high-dimensional statistics, random linear algebra, non-linear optimization, graphical models, network analysis and deep learning.