Yang Cai

ladyissy.github.io
yc3404@columbia.edu | 530-750-9867 | 31 River Ct, APT 2308, Jersey City, NJ, 07310

EDUCATION

COLUMBIA UNIVERSITY
Master of Arts in Statistics (MA)
Expected Dec 2018 | New York, NY
Cum. GPA: 3.51/4.3

NEW YORK UNIVERSITY
Bachelor of Arts in Economics (BA)
May 2017 | New York, NY
College of Arts and Science
Dean's List (Academic Year 2015)
Major GPA: 3.46 / 4.0

LINKS

LinkedIn:// www.linkedin.com/in/yang-cai-80165b66/

COURSEWORK

GRADUATE
Probability
Statistical Inference
Linear Regression Models
Statistical Computing
Statistical Machine Learning
Bayesian Statistics
Advanced Data Analysis
Elementary Stochastic Processe

UNDERGRADUATE
Statistics
Econometrics
Mathematics for Economics
Linear Algebra
Money and Banking
Intermediate Macroeconomics
Intermediate Microeconomics

SKILLS

COMPUTING
R• SAS• Python

OFFICE
Word• Excel• PowerPoint

LANGUAGE
English• Mandarin

ACADEMIC ACTIVITIES

A STUDY IN BIRTHWEIGHT This project analyzes 189 observations that collected at Baystate Medical Center during 1986 to study a causal relationship between the birth weights and the risk factors.

- Used linear regression technique to study the causal relationship between the birth weights and the risk factors.
- Divided birth weight into two categories (low and not low) and did exploratory data analysis by comparing the ratio between the two categories among the risk factors.

HIERARCHICAL LINEAR REGRESSION MODEL USING RSTAN This project used hierarchical model and linear regression to study how gross horse power and rear axle ratio affect miles per gallon for 10 types of cars.

- Employed Rstan to do the simulation and computation.
- Integrated Bayesian statistics by specifying likelihood and prior distributions.
- Used RShiny to check the result of simulation.

WAGE DIFFERENCES BETWEEN THREE RACES This project explored the data set of roughly 25,000 records for males between the age of 18 and 70 who are full time workers to study if African males had wage differences than other two races.

- Used linear regression model to analyze the data.
- Diagnostics and model validation was carried out by plotting with ggplot2.

GETTING AND CLEANING DATA PROJECT This project runs a piece of R code that cleans a human activity recognition dataset (with 10299 instances) by doing the following steps:

- Merge the training and test data set to create one data set.
- Extract only the measurements on the mean and standard deviation for each measurement.
- Use descriptive activity names to name the activities in the data set.
- Appropriately label the dataset with descriptive variable names.
- Create a second, independent tidy dataset with the average of each variable for each activity and each subject.

LINEAR REGRESSION MODEL PROJECT This project builds a model that explores the statistical differences of average male wages for the three race classes by doing the following steps:

- Fit a rough model, check interaction terms and make adjustments
- Model Selection, Diagnostics and Model Validation
- Answer the research question based on the model built

CERTIFICATES

2016 R Programming by Johns Hopkins University on Coursera
2016 Getting and Cleaning Data by Johns Hopkins University on Coursera

SOCIETIES

2013 Lambda Omicron Xi Community Service Sorority