

# Hui Lin

---

✉ [longqiman@gmail.com](mailto:longqiman@gmail.com) 🐇 [happyrabbit](https://github.com/happyrabbit) 🌐 [scientistcafe.com](https://scientistcafe.com)

---

## Education

**2011 - 2013**, PhD in Statistics, Department of Statistics, Iowa State University, USA

**2009 - 2011**, MS in Statistics, Department of Statistics, Iowa State University, USA

**2005 - 2009**, BS in Statistics, School of Mathematical Science, Beijing Normal University, CHINA

## Computer Skill:

Programming: expertise in R (8 yrs), Python (1+ yr), SQL (4+ yr), CSS, JavaScript, HTML5  
Software and system: SAS, Linux/Unix, Windows, AWS, limited experience with spark

## Working and Research Experience

**Sep.2017 - Present**, Data Scientist, DowDuPont, Johnston, IA, USA

Provide statistical leadership for a broad range of predictive analytics in Marketing

Apply models to drive decisions for billion-dollar marketing budget

*Bayesian Regression Model, Super Learner ensembles: Elastic Network Regression, AIC Stepwise Regression, Random Forest, Neural Network, Support Vector Machine, Gradient Boosting Tree, Partial Least Square Regression*

Lead project on social media analytics

*Network Analysis, Natural Language Processing (POS tagging, entity recognition, sentimental analysis, topic modeling)*

*Construct automatic social media data pipeline with data engineer: scrape data from Ag Forum or get data through API, build database in the cloud, develop dashboard for marketing users*

Provides support to evaluate marketing programs under the observational scenario

*Provide statistical training to other analysts on experimental design and propensity score analysis*

Automating models and dashboard that track business performance

*Identify key marketing metrics with the marketing analytics team*

*Develop and implement Shiny dashboard including those metrics*

**May.2013 - Aug.2017**, Data Scientist, DuPont Pioneer, Johnston, IA, USA

Provide statistical leadership for a broad range of predictive analytics and customer segmentation analyses focused on improving Sales and Marketing productivity

*Group Lasso Logistic Regression, Random Forest, Hierarchical Clustering*

Analyze market research survey data using psychometric models and natural language processing; quantitatively study customer perception and identify variables that are predictive of patterns and/or are leading indicators

*Factor analysis, Topic Model, Longitudinal Analysis*

Work with data engineers to design and build marketing database in the cloud to support current and future marketing analytics

*It evolves to the current project on social media analytics*

**Jun.2015 - Present**, Co-founder, [Central Iowa R User Group](#)

**Jun.2012 - Apr.2013**, Consultant, Statistical Consulting for Business College, Iowa State University

**Apr.2012 - Apr.2013**, Research Assistant, [Exact Test for Difference in Proportions Based on One-to-two Matched Binary Data](#)

**Nov.2011 - Apr.2013**, Research Assistant, [Distribution of Non-normal, Dependent Bivariates with Additive Measurement Error](#)

**Sep.2009 - Apr.2013**, Research Assistant, [Production Animal Disease Risk Assessment Program](#)

**Aug.2011 - May.2012**, Consultant, General University Statistical Consulting and Statistics Consulting for College of Veterinary Medicine, Iowa State University

**Spring.2010 Semester**, Teaching Assistant, TA for STAT 341, 342, Department of Statistics, Iowa State University

## Certifications

### **Structuring Machine Learning Projects**

Coursera, License QNMAPFJUBKAL

October 2017 – October 2017

### **Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization**

Coursera, License X85WN3PHKLS7

October 2017 – October 2017

### **Neural Networks and Deep Learning**

Coursera, License XAS5ZLMD234X

September 2017 – September 2017

### **Pioneer Young Professionals Leadership Exploration and Development Program**

DuPont Pioneer

January 2016 – December 2016

## Book

### [Machine Learning using R](#): By Hui Lin

Publisher: Publishing House of Electronics industry

Status: Published in Oct 2017

ISBN : 9787121326585

### [Introduction to Data Science](#): By Hui Lin and Ming Li

Publisher: CRC Press

Status: 2019 Dec 31 (Expected)

Github: <https://github.com/happyrabbit/IntroDataScience>

## Book Translation

[Applied Predictive Modeling](#): By Max Kuhn and Kjell Johnson

Chinese translator for chapters 1-4, 12-14, 16, 18-20

Publisher in China: Huazhang Publishing Inc.

Status: Published in May 2016

[R for Marketing Research and Analytics](#): By Chris Chapman and Elea McDonnell Feit

Chinese translator for the book

Publisher in China: Huazhang Publishing Inc.

Status: Published in Oct 2016

[Statistical Rethinking - A Bayesian Course with Examples in R and Stan](#): By Richard McElreath

Chinese translator for the book

Publisher in China: Huazhang Publishing Inc.

Status: Expect to finish by Nov 2017

## Journal Publications

Krysta Deitz, Kelly Makielski, Jackie M. Williams, **Hui Lin**, Jo Ann Morrison (2015). [Effect of 6-8 weeks of oral ursodeoxycholic acid administration on serum concentrations of fasting and postprandial bile acids and biochemical analytes in healthy dogs](#), Veterinary Clinical Pathology, DOI: 10.1111/vcp.12275.

**Hui Lin** (2013), [Statistical methods in disease risk analysis, disease testing and nutrition epidemiology](#), Graduate Theses and Dissertations. Paper 13380.

J.M. Williams, I.A. Krebs, E.A. Riedesel, C.K. Kelly, **H. Lin** (2013), [Comparison of fluoroscopy and computed tomography for tracheal lumen diameter measurement and determination of intraluminal stent size in healthy dogs](#), Veterinary Radiology & Ultrasound. DOI: 10.1111/vru.12111

**Hui Lin**, Chong Wang, Peng Liu, Derald J. Holtkamp (2012), [Construction of Disease Risk Scoring Systems using Logistic Group Lasso: Application to Porcine Reproductive and Respiratory Syndrome Survey Data](#), Journal of Applied Statistics. DOI:10.1080/02664763.2012.752449

Derald J. Holtkamp, **Hui Lin**, Chong Wang (2012), [Identifying questions in the American Association of Swine Veterinarian's PRRS risk assessment survey that are important for retrospectively classifying swine herds according to whether they reported clinical PRRS outbreaks in the previous 3 years.](#), Preventive Veterinary Medicine 106(2012) 42-52

Allen, Kelli; Coetzee, Johann; Edwards-Callaway, Lily; Glynn, Hayley; Dockweiler, Jenna; KuKanich, Butch; **Lin, Hui**; Wang, Chong; Fraccaro, Elena; Jones, Meredyth; Bergamasco, Luciana (2012), [The impact and effect of timing of oral meloxicam administration on physiological responses in calves after dehorning with local anesthesia](#), Journal of Dairy Science, DOI: <http://dx.doi.org/10.3168/jds.2012-6251>

Derald J. Holtkamp, **Hui Lin**, Chong Wang (2012), Evaluation of an objective risk scoring system for assessing the likelihood of virus introduction in porcine reproductive and respiratory syndrome virus-free breed-to-wean sow herds in the U.S., Open Journal of Veterinary Medicine. 3 168-175

Hui Lin, Cheng Hao, Xiao Wang, Qiang Yuan, Tong-Jie Zhang, and Bao-Quan Wang (2008), [Observational H\(z\) Data as a complementary to other cosmological probes](#), Modern Physics Letter A (No. MPLA-D-08-00168R1)

## Tutorials and Invited Talks

Preparing Statistician/Statistics Graduates to Be Data Scientist, 2017-07-30, [Joint Statistical Meeting 2017](#), [Course Website](#)

Interactive document and dashboard using R and shiny, 2017-07-12, [Central Iowa R User Group](#), [Blog](#)

Data Scientist Using R - Data Wrangling, 2017-05-02, [Central Iowa R User Group](#), [Blog](#)

Data Science Introduction, 2017-03-29, [Central Iowa R User Group](#), [Slides](#)

Unlock Unstructured Data, 2016-12-29, [Beijing Institute of Technology](#), [Slides](#), [Rcode](#)

Market Research Customer Perception Analytics, 2016-7-21, [Video](#), [Slides](#), [Rcode](#)

Data Wrangling in R and Model Technique, 2016-6-8, Capital of Statistics(COS)-North America, [Video](#), [Slides](#), [Rcode1](#), [Rcode2](#)

General Process of Data Analysis and Data Preprocessing, 2016-6-2, Capital of Statistics(COS) - North America, [Video](#), [Slides](#), [Rcode](#)

[Jewels Academy Math Camp](#), 2016.2.20, [Jewels Academy](#), Des Moines, Iowa

[What is data scientist](#), 2015.12.12, [Capital of Statistics\(COS\)](#), Beijing, CHINA, [Video](#), [Slides](#)

[Data Pre-processing Using R](#), 2015.11, [Central Iowa R User Group](#), Johnston, Iowa, USA

Analytics to Support Business Decision in Traditional Industry, 2014, 7th R conference, Beijing, CHINA