



清华大学 统计学研究中心
Center for Statistical Science, Tsinghua University

年度报告

Annual Report

2020.07-2021.06

July 2020 to June 2021

北京·清华园



CONTENTS

目录

1

中心概况

2

组织架构

- 学术委员会
- 顾问委员会
- 行政团队

学科团队

- 中心教员
- 博士后
- 博士研究生

3

学科发展与人才引进

- 新晋升教员
- 新入职教员
- 访问学者
- 合作研究

4



6

学术活动

- 主办学术活动
- 统计学与数据科学论坛
- 参加学术活动

8

社会服务及影响

- 社会服务
- 学术杂志服务
- 统计咨询服务

5

学术成果

- 学术论文
- 科研项目
- 奖励荣誉

人才培养

- 本科生培养
- 研究生培养
- 优秀大学生夏令营

7

水木数据科学访谈

9



国际著名统计学家、清华校友、美国哈佛大学刘军教授和林希虹教授一直密切关注中心的发展动态，并从人才培养、团队建设、学科发展等多个方面给予支持和指导。

中心概况

清华大学统计学研究中心是学术独立的校级研究机构，统筹规划清华大学统计学科的发展与建设，行政事务挂靠工业工程系。中心自 2015 年成立以来，始终秉承并践行“开拓创新、争创一流”的发展理念，以“建立高水平师资队伍，开展高水平学术研究，推动跨学科交叉合作，建设国际一流学科”为发展目标，推动学科发展建设。

目前，中心已初步建成一支以优秀青年人才为主的朝气蓬勃的研究团队，并在清华园建立起完整的统计学人才培养体系，涵盖从本科到博士、博士后的各个层次。中心现有教授 1 人、杰出访问教授 1 人、客座教授 3 人、副教授 7 人、助理教授 4 人、讲师 3 人、博士研究生 38 人（截至 2021 年 6 月 30 日）。

在学术研究方面，中心以统计学理论和方法研究为基础，着重推动生物医学统计、工业统计、经济与金融统计、人工智能和大数据处理等交叉研究前沿方向，并取得了丰硕的学术研究成果。中心还组织专业力量创立统计咨询中心，为清华师生和社会各界提供高质量的统计咨询和数据分析服务。

上述多方面努力推动清华大学统计学科快速发展，学术声誉不断提升，国内外影响力持续提高；在 QS 国际学科排名中，清华大学“统计学与运筹学”由第 26 名（2016 年）提升至 16 名（2020 年），并保持稳定，达到国际一流学科水平；近两年泰晤士中国学科评级均为 A+。



- 6月，中心成立仪式
- 9月，首届博士研究生入学



- 3月，发起成立“中国现场统计研究会计算统计分会”
- 5月，建立“清华大学统计咨询中心”
- 9月，入选教育部“双一流”学科建设名单



- 7月，成立“中国统计咨询合作联盟”
- 10月，聘请汤家豪院士出任“杰出访问教授”



- 1月，清华大学“统计学”本科一学位通过教育部备案
- 3月，“统计学与运筹学”学科在QS世界学科排名保持第16名

2015

2016

2017

2018

2019

2020

2021



- 9月，开设本科生“统计学辅修”项目



- 11月，承办“国际计算统计协会亚洲分会25周年大会暨中国现场统计学会计算统计分会第二届年会”



- 7月，“泰晤士高等教育中国学科评级” A+
- 12月，中心师生在“清华大学抗击新冠肺炎疫情表彰大会”荣获表彰

组织架构





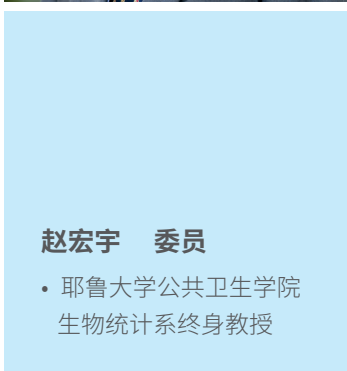
李志忠 委员

- 清华大学工业工程系
长聘教授



林希虹 委员

- 哈佛大学生物统计系终
身教授
- 美国国家医学院院士



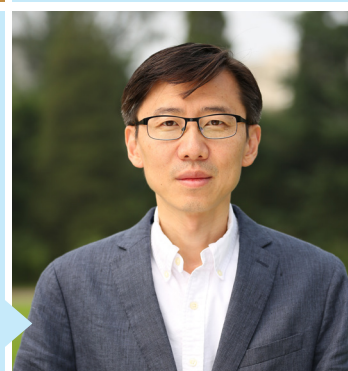
赵宏宇 委员

- 耶鲁大学公共卫生学院
生物统计系终身教授



朱 宇 委员

- 普渡大学统计系终身教授



Prof. Donald Rubin

- 清华大学特聘教授
- 美国科学院院士



王永雄 教授

- 斯坦福大学统计系终身
教授
- 美国科学院院士



汤家豪 院士

- 伦敦政治经济学院
荣休教授
- 挪威科学与文学院
外籍院士

杰出访问教授



马 腾

行政助理



田 芸

行政助理



宋希婷

科研助理



学科团队

- 中心教员 -



杨立坚 教授

- 清华大学统计学研究中心长聘教授
- 北卡罗来纳大学教堂山分校统计学博士
- 美国统计协会会员 (ASA Fellow)
- 国际统计学会当选会员 (ISI Elected Member)
- 国际数理统计学会会士 (IMS Fellow)
- 国际工程技术协会杰出会士 (IETI Distinguished Fellow)
- 国家级人才计划入选者

研究方向: 时间序列、函数型及高维数据的统计推断, 以及统计学在经济学、金融学、农学、食品科学、地理学、遗传学、神经科学和管理科学的应用

- 清华大学统计学研究中心长聘副教授
- 北京大学统计学博士
- 哈佛大学统计系博士后、副研究员
- 北京智源人工智能研究院“智源研究员”
- 国家级人才计划入选者

研究方向: 贝叶斯统计、统计计算、生物信息、文本分析、人工智能方法



邓柯 副教授



李东 副教授

- 清华大学统计学研究中心长聘副教授
- 香港科技大学统计学博士
- 香港科技大学数学系博士后
- 爱荷华大学统计与精算系博士后

研究方向: 金融计量经济学、非线性时间序列分析、网络与大数据

- 北京大学统计学博士
 - 耶鲁大学生物统计系博士后、副研究员
- 研究方向: 统计遗传学、生物信息学、应用统计

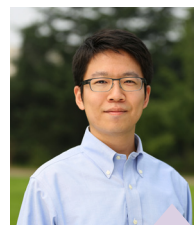


侯琳 副教授

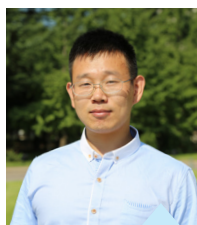
- 乔治华盛顿大学系统工程 (运筹学) 博士

- 哈佛大学医学院博士后

研究方向: 医学统计、自然语言处理、电子病历数据分析、医学知识提取、临床决策支持



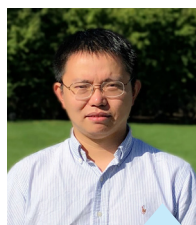
俞声 副教授



刘汉中 副教授

- 北京大学统计学博士
- 加州大学伯克利分校统计系联合培养博士
- 加州大学伯克利分校统计系博士后

研究方向: 高维数据统计推断、因果分析



林乾 副教授

- 麻省理工学院数学博士
 - 哈佛大学统计系博士后
 - 北京智源人工智能研究院“青年科学家”
- 研究方向: 降维方法、函数型 / 拓扑型数据分析、蒙特卡洛方法

- 多伦多大学统计学博士
 - 伦敦大学学院统计系博士后
 - 鲁尔波鸿大学数学系博士后
- 研究方向: 时间序列、变点推断、M 估计、网络数据



吴未迟 副教授

- 德克萨斯 A&M 大学统计学博士
 - 哥伦比亚大学生物统计系博士后
- 研究方向: 分位数回归、测量误差分析、高维数据统计分析、流行病学与生物遗传学的统计分析、电子医疗病历数据分析

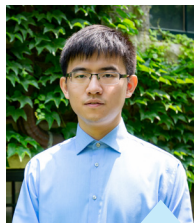


王天颖 助理教授



张静怡 助理教授

- 乔治亚大学统计学博士
- 研究方向: 数据融合、数据降维、去中心化网络、最优传输理论



杨朋昆 助理教授

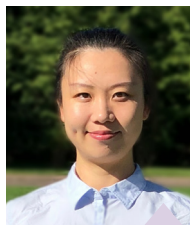
- 伊利诺伊大学香槟分校电子与计算机工程博士
 - 普林斯顿大学电子工程系博士后
- 研究方向: 高维统计理论、机器学习、算法及优化

- 哈佛大学统计学博士
- 研究方向: 贝叶斯统计、统计计算、系统发生学、生物信息学



胡志睿 助理教授

- 北京大学统计学博士
- 宾夕法尼亚大学联合培养博士
- 复旦大学上海数学中心博士后



邓婉璐 讲师



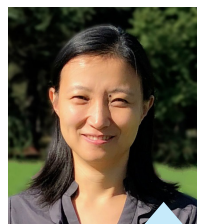
周在莹 讲师

- 清华大学统计学博士



王江典 讲师

- 北卡罗来纳州立大学统计学博士



高梦昭 高级咨询师

- 北京交通大学管理学博士
- 宾夕法尼亚州立大学应用统计硕士

- 博士后 -



蒋家坤

- 西南财经大学经济学博士
- 2018 年 7 月加入中心
- 2020 年 7 月出站
- 现就职于北京师范大学珠海校区

研究方向: 函数型数据分析、半参数模型、门槛模型

- 博士研究生 -

2016 级: 单娜阳、徐嘉泽

2017 级: 郭瀚民、李 杰、刘朝阳、袁 正、张心雨、钟 晨

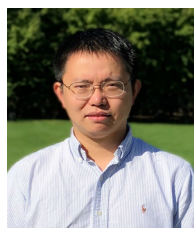
2018 级: 黄 昆、罗赛迪、罗声旋、潘长在、沈 翀、宋泽宁、王 掣、杨萱铃、余 博、朱 珂

2019 级: 任吉杨、宋 爽、孙 爽、陶宇心、王海洋、吴方维、郑思捷、周墨钦

2020 级: 白露佳、冯永真、胡祺睿、李冬煜、卢伟灏、卢 鑫、王 达、徐曼芸、余 成、苑洪意、张卓婧

学科发展与人才引进

- 新晋升教员 -



林 乾
晋升准聘副教授



吴未迟
晋升准聘副教授

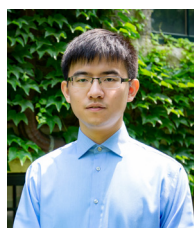
- 新入职教员 -



王天颖 助理教授



张静怡 助理教授



杨朋昆 助理教授



胡志睿 助理教授

- 访问学者 -



韩 杨 副教授
曼彻斯特大学数学学院
来访时间 :2021 年 3 月-9 月



- 合作研究 -

清华大学统计学研究中心高度重视产学研的有效结合。自 2016 年起，先后为原国家质量监督检验检疫总局（现国家市场监督管理总局）、国家食品安全风险评估中心、海关总署、国家市场监督管理总局等多个政府部门的政策决策和改革方案制定提供重要技术支持，为我国进出口食品安全监管改革、食品安全评估不确定性分析、日本输华食品核污染风险评估、新冠疫情防控、食品安全评价性抽检方案制定等工作做出重要贡献。

中心与北京协和医院、北京清华长庚医院、北京大学第一医院、北京大学第三医院等多个医院建立深入合作，运用统计学优势助力智能医学诊断系统、电子病历数据分析及挖掘等应用的开发。

此外，中心与国内外知名人文社科机构联合进行了大量交叉研究，运用数据科学手段协助人文历史学者整理海量中国经典历史文献和非物质文化遗产资料，提高研究效率，促进文化传承与创新。



学术成果

- 学术论文 -

- ◆ Shan Yu, Guannan Wang, Lily Wang and **Lijian Yang*** (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica* **31** (3): 1463-1487.
- ◆ **Chen Zhong** and **Lijian Yang*** (2021). Simultaneous confidence bands for comparing variance functions of two samples based on deterministic designs. *Computational Statistics* **36** (2): 1197-1218.
- ◆ Lijie Gu, Suojin Wang and **Lijian Yang** (2021). Smooth simultaneous confidence band for the error distribution function in nonparametric regression. *Computational Statistics and Data Analysis* **155**: 107106.
- ◆ Songtao Li, Ruoran Chen, **Lijian Yang**, Dinglong Huang and Simin Huang (2020). Predictive modeling of consumer color preference: using retail data and merchandise images. *Journal of Forecasting* **39** (8): 1305-1323.
- ◆ **Yuanyuan Zhang**, Rong Liu, Qin Shao and **Lijian Yang*** (2020). Two-step estimation for time varying ARCH models. *Journal of Time Series Analysis* **41** (4): 551-570.
- ◆ Junni Zhang, Feifei Wang, **Yichao Li**, **Ke Deng** and **Jun S. Liu** (2021). Bayesian text classification and summarization via a class-specified topic model. *Journal of Machine Learning Research* **22** (89): 1-48.
- ◆ **Haifeng Wang**, **Changzai Pan**, Xiao Guo, Chunlin Ji* and **Ke Deng*** (2021). From object detection to text detection and recognition: a brief evolution history of optical character recognition. *WIREs Computational Statistics* **13** (5): e1547. (Co-corresponding author)
- ◆ **Yang Yang**, Chunlin Ji* and **Ke Deng*** (2021). Rapid design of metamaterials via multi-target bayesian optimization. *The Annals of Applied Statistics* **15** (2): 768-796. (Co-corresponding author)
- ◆ Hao Xu, **Tingxuan Han**, **Haifeng Wang**, Shanggui Liu, Guanghao Hou, Guanchao Jiang, Jian Zhou* and **Ke Deng*** (2021). Detection of blood stains using computer vision based algorithm and it's association with postoperative outcomes in thoracoscopic lobectomy. *2021 AATS International Thoracic Surgical Oncology Summit*. (Co-corresponding author)
- ◆ Pan Wang, **Qi Li**, Nan Sun, Yibo Gao, **Jun S. Liu**, **Ke Deng*** and Jie He* (2021). MiRACLe: an individual-specific approach to improve microRNA-target prediction based on a random contact model. *Briefings in Bioinformatics* **22** (3) : bbaa117. (Co-corresponding author)
- ◆ Ruiyun Li, Bin Chen, Tao Zhang, Zhehao Ren, Yimeng Song, Yixiong Xiao, **Lin Hou**, Jun Cai, Bo Xu, Karen Kie Yan Chan, Ying Tu, Mu Yang, Jing Yang, **Zhaoyang Liu**, **Chong Shen**, **Che Wang**, Lei Xu, Qiyong Liu, Shuming Bao, Jianqin Zhang, Yuhai Bi, Yuqi Bai, Ke Deng, Wusheng Zhang, Wenyu Huang, Jason D Whittington, Nils Chr. Stenseth, Dabo Guan, Peng Gong and Bing Xu (2020). Suppression of future waves of COVID-19: global pandemic demands joint interventions. *Proceedings of the National Academy of Sciences of USA* **117** (42): 26151-26157.
- ◆ Rong Zhang, Ming Hu, Yu M. Zhu, Zhaohui S. Qin, **Ke Deng** and **Jun S. Liu** (2020). Inferring spatial organization of individual topologically associated domains via piecewise helical model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **17**(2): 647-656.
- ◆ Chunlin Ji, Bin Liu, **Yingkai Jiang** and **Ke Deng** (2020). Sequential learning for Dirichlet process mixtures.





Proceedings of the 2nd Symposium on Advances in Approximate Bayesian Inference.

- ◆ **Huan Gong** and **Dong Li** (2020). On the three-step non-Gaussian quasi-maximum likelihood estimation of heavy-tailed double AR models. *Journal of Time Series Analysis* **41** (6): 883-891.
- ◆ **Dong Li** and **Howell Tong** (2020). On an absolute autoregressive model and skew symmetric distributions. *Statistica* **80** (2): 177-198.
- ◆ Jing Zhou, **Dong Li**, Rui Pan and Hansheng Wang (2020). Network GARCH model. *Statistica Sinica* **30** (4): 1723-1740.
- ◆ Youhong Ben and **Feiyu Jiang** (2021). A note on Portmanteau tests for conditional heteroscedastic models. *Economics Letters* **192**: 109159.
- ◆ Wei Jiang, **Shuang Song**, **Lin Hou** and Hongyu Zhao (2021). A set of efficient methods to generate high-dimensional binary data with specified correlation structures. *The American Statistician* **75** (3): 310-322.
- ◆ Zehao Ren, Ruiyun Li, Tao Zhang, Bin Chen, **Che Wang**, Miao Li, **Shuang Song**, Yixiong Xiao, Bo Xu, **Zhaoyang Liu**, **Chong Shen**, Dabo Guan, **Lin Hou**, **Ke Deng**, Yuqi Bai, Peng Gong and Bing Xu (2021). Reduction of human mobility matters during early COVID-19 outbreak: Evidence from India, Japan and China. *International Journal of Environmental Research and Public Health* **18** (6): 2826.
- ◆ **Hanmin Guo**, James J. Li, Qiongshi Lu and **Lin Hou** (2021). Detecting local genetic correlations with scan statistics. *Nature Communications* **12** (1): 2033.
- ◆ Alexander Chi, Xia He, **Lin Hou**, Nam P. Nguyen, Guangying Zhu, Robert B. Cameron and Jay M. Lee (2021). Classification of non-small cell lung cancer's tumor immune micro-environment and strategies to augment its response to immune checkpoint blockade. *Cancers* **13** (12): 2924.
- ◆ Jiaqi Guan, Runzhe Li, **Sheng Yu** and Xuegong Zhang* (2021). A method for generating synthetic electronic medical record text. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **18** (1): 173-182.
- ◆ Yuanren Tong#, Keming Lu#, Yingyun Yang, Ji Li, **Yucong Lin**, Dong Wu, Aiming Yang, Yue Li*, **Sheng Yu*** and Jiaming Qian (2020). Can natural language processing help differentiate inflammatory intestinal diseases in China? Models applying random forest and convolutional neural network approaches. *BMC Medical Informatics and Decision Making* **20** (1): 248. #contributed equally, *contributed equally.
- ◆ **Zheng Yuan**, Yuanhao Liu, Qiuyang Yin, Boyao Li, Xiaobin Feng, Guoming Zhang, and **Sheng Yu*** (2020). Unsupervised multi-granular Chinese word segmentation and term discovery via graph partition. *Journal of Biomedical Informatics* **110**: 103542.
- ◆ **Yucong Lin**, Yang Li, Keming Lu, Cheng Ma, Peng Zhao, Daiqi Gao, Zihao Fan, Zijie Cheng, Zheyu Wang and **Sheng Yu*** (2020). Long-distance disorder-disorder relation extraction with bootstrapped noisy data. *Journal of Biomedical Informatics* **109**: 103529.
- ◆ **Lishan Yu** and **Sheng Yu*** (2020). Developing an automated mechanism to identify medical articles from Wikipedia for knowledge extraction. *International Journal of Medical Informatics* **141**: 104234.



- ◆ Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan Jr., Victor M. Castro, Dmitriy Dligach, Sean Finan, **Sheng Yu**, Vivian S. Gainer, Nancy A. Shadick, Guergana Savova, Shawn N. Murphy, Tianxi Cai, Scott T. Weiss and Rose Du* (2020). Age and morphology of posterior communicating artery aneurysms. *Scientific Reports* **10**: 11545.
- ◆ Aaron Sonabend#, Winston Cai#, Yuri Ahuja, Ashwin Ananthakrishnan, Zongqi Xia, **Sheng Yu*** and Chuan Hong* (2020). Automated ICD coding via unsupervised knowledge integration (UNITE). *International Journal of Medical Informatics* **139**: 104135. #contributed equally, *contributed equally.
- ◆ **Hanzhong Liu**, Xin Xu and Jingyi Jessica Li* (2020). A bootstrap lasso + partial ridge method to construct confidence intervals for parameters in high-dimensional sparse linear models. *Statistica Sinica* **30**: 1333-1355.
- ◆ **Hanzhong Liu** and Yuehan Yang* (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika* **107** (4): 935-948.
- ◆ **Qian Lin**, Xinran Li, Dongming Huang and **Jun S. Liu** (2021). On the optimality of sliced inverse regression in high dimensions. *Annals of Statistics* **49** (1): 1-20.
- ◆ **Qian Lin**, Zhigen Zhao and **Jun S. Liu** (2021). Testing model utility for single index models under high dimension. *Festschrift in Honor of R. Dennis Cook*, 65-86.
- ◆ Prahatha Venkatraman, Ishara Mills-Henry, Karthik Ramaswamy Padmanabhan, Pete Pascuzzi, Menna Hassan, **Jingyi Zhang**, Xinlian Zhang, Ping Ma, Chi Pui Pang, John E. Dowling, Mingzhi Zhang and Yuk Fai Leung (2020). Rods contribute to visual behavior in larval zebrafish. *Investigative Ophthalmology & Visual Science* **61** (12): 11-11.
- ◆ Yihong Wu and **Pengkun Yang** (2020). Polynomial methods in statistical inference: theory and practice. *Now Publisher*.
- ◆ Shaopeng Wang, Forest Isbell, **Wanlu Deng**, Pubin Hong, Laura E. Dee, Patrick Thompson and Michel Loreau (2021). How complementarity and selection affect the relationship between ecosystem functioning and stability. *Ecology* **102** (6): e03347.
- ◆ 崔鹏, **邓柯**, 王国豫, 王禹皓, 张江, 朱军, 朱占星 (2020)。安全可信智能的可能技术路径。《计算机学会通讯》**16** (11): 11-16。
- ◆ 徐嘉泽, 潘长在, 贺莉丽, 王宏甦, 张力伟, **邓柯*** (2020)。基于 TopWORDS方法的古文献专名识别:以《汉书》和《明史》为例。《数字人文》**2**: 90-105。
- ◆ 李建军, 高梦昭, **邓柯**, 李艺超, 郭瀚民, 方海萍, 赵明刚 (2021)。进口鸡肉传带禽伤寒沙门氏菌的定量风险评估。《中国动物检疫》**38** (1): 30-34。

Online Publications

- ◆ Jiakun Jiang, Li Cai and **Lijian Yang*** (2020). Simultaneous confidence band for the difference of regression functions of two samples. *Communications in Statistics-Theory and Methods*.
- ◆ Yan Fang, Lan Xue, Carlos Martins-Filho and **Lijian Yang** (2020). Robust estimation of additive boundaries with quantile regression and shape constraints. *Journal of Business and Economic Statistics*.
- ◆ **Wanchuang Zhu**, Yingkai Jiang, **Jun S. Liu** and **Ke Deng*** (2021). Partition-mallows model and its inference for rank aggregation. *Journal of the American Statistical Association*.
- ◆ Qingru Wu, Yi Tang, Shuxiao Wang, Liang Li, **Ke Deng**, Guigang Tang, Kaiyun Liu, Dian Ding and Hui Zhang (2020). Developing a statistical model to explain the observed decline of atmospheric mercury. *Atmospheric*



Environment.

- ◆ **Nayang Shan**, Yuhan Xie, **Shuang Song**, Wei Jiang, Zuoheng Wang and **Lin Hou** (2021). A novel transcriptional risk score for risk prediction of complex human diseases. *Genetic Epidemiology*.
- ◆ **Shuang Song**, **Nayang Shan**, Geng Wang, Xiting Yan, **Jun S. Liu** and **Lin Hou** (2021). Openness weighted association studies: leveraging personal genome information to prioritize noncoding variants. *Bioinformatics*.
- ◆ Holger Dette, Subhra Sankar Dhar and **Weichi Wu** (2021). Identifying shifts between two regression curves. *Annals of the Institute of Statistical Mathematics*.
- ◆ Holger Dette and **Weichi Wu** (2020). Prediction in locally stationary time series. *Journal of Business & Economic Statistics*.
- ◆ **Jingyi Zhang**, Huolan Zhu, Yongkai Chen, Chenguang Yang, Huimin Cheng, Yi Li, Wenxuan Zhong and Fang Wang (2021). Ensemble machine learning approach for screening of coronary heart disease based on echocardiography and risk factors. *BMC medical informatics and decision making*.
- ◆ Cheng Meng, Jun Yu, **Jingyi Zhang**, Ping Ma and Wenxuan Zhong (2020). Sufficient dimension reduction for classification using principal optimal transport direction. *Proceedings of NeurIPS*.

To Appear

- ◆ **Jie Li** and **Lijian Yang*** (2021). Statistical inference for functional time series. *Statistica Sinica*.
- ◆ **Yichao Li**, Wenshuo Wang, **Ke Deng*** and **Jun S. Liu*** (2021). Stratification and optimal resampling for sequential Monte Carlo. *Biometrika*. (Co-corresponding author)
- ◆ **Changzai Pan**, Hongsu Wang and **Ke Deng*** (2021). Exploration and statistical modeling on inheritance probability of the biographical addresses of historical figures from 7th to 19th centuries. *Digital Humanities and East Asian Studies*.
- ◆ **Yingkai Jiang**, Lixing Zhu, Xinshu Zhao, **Jun S. Liu** and **Ke Deng*** (2021). Total-effect test is superfluous for establishing complementary mediation. *Statistica Sinica*.
- ◆ **Tianying Wang*** and Alex Asher (2020). Improved semiparametric analysis of polygenic gene-environment interactions in case-control studies. *Statistics in Biosciences*.
- ◆ **Jingyi Zhang**, Wenxuan Zhong and Ping Ma (2021). A review on modern computational optimal transport methods with applications in biomedical research. *Emerging Topics in Statistics and Biostatistics*.
- ◆ Cong Fang, Jason Lee, **Pengkun Yang** and Tong Zhang (2021). Modeling from features: a mean-field framework for over-parameterized deep neural networks. *Conference on Learning Theory (COLT)*.
- ◆ **朱珂**, **姜瑛恺**, 王祥, 史志呈, 杨超 *, **刘汉中 ***, **邓柯 *** (2021)。领域知识与大数据推断相结合的定制化生产产品认证新模式研究。《应用概率统计》
- ◆ **王海洋**, 陈静 *, **邓柯 *** (2021)。基于卷积神经网络的民国广告图片聚类方法。《数字人文》
- ◆ **李艺超**, **徐崇元**, 方海萍, **高梦昭**, **牛晓月**, 宋志刚, 李建军, **邓柯 *** (2021)。进口动物产品潜在入侵疫病风险评估的统计学模型和方法。《数理统计与管理》
- ◆ **徐崇元**, 陈孟裕, 金锋, 丛林晔, 焦阳, **邓柯 *** (2021)。进口食品社会关注度专家评分一致性的统计学研究。《数理统计与管理》
- ◆ **朱珂**, **姜瑛恺**, 王祥, 史志呈, 杨超 *, **刘汉中 ***, **邓柯 *** (2021)。基于非负罚项逻辑回归的定制化产品风险评估研究。《数理统计与管理》

- 科研项目 -

项目来源	项目类型	项目期限	项目金额	负责人
国家自然科学基金	数学天元基金项目	2021 年 -2022 年	20 万（元）	杨立坚
索元生物医药公司	合作研究	2021 年 -2022 年	50 万（元）	杨立坚
国家自然科学基金	面上项目	2018 年 -2021 年	48 万（元）	杨立坚
清华大学国强人工智能研究院	校内委托项目	2021 年 -2023 年	100 万（元）	邓 柯
国家自然科学基金	重点项目课题	2020 年 -2024 年	50 万（元）	邓 柯
科技部	国家重点研发计划子课题	2020 年 -2023 年	54 万（元）	邓 柯
北京市自然科学基金	重点项目课题	2020 年 -2023 年	120 万（元）	邓 柯
上海起承文化发展有限公司	合作研究	2020 年 -2021 年	25 万（元）	邓 柯
广西产品质量检验研究院	国家机关委托项目	2020 年 -2020 年	10 万（元）	邓 柯
国家社会科学基金	重大项目课题	2019 年 -2022 年	10 万（元）	邓 柯
清华大学国强人工智能研究院	子课题	2019 年 -2021 年	30 万（元）	邓 柯
北京智源人工智能研究院	研究员专项	2019 年 -2021 年	140 万（元）	邓 柯
海关总署卫生检疫司	国家机关委托项目	2019 年 -2020 年	40 万（元）	邓 柯
国家自然科学基金	面上项目	2018 年 -2021 年	48 万（元）	邓 柯
深圳光启高等理工研究院	合作研究	2018 年 -2021 年	96 万（元）	邓 柯
科技部	国家重点研发计划子课题	2017 年 -2020 年	43 万（元）	邓 柯
国家自然科学基金	面上项目	2020 年 -2023 年	48 万（元）	李 东
科技部	国家重点研发计划子课题	2019 年 -2021 年	50 万（元）	李 东
国家自然科学基金	面上项目	2018 年 -2021 年	48 万（元）	李 东
国家自然科学基金	面上项目	2021 年 -2024 年	51 万（元）	侯 琳





项目来源	项目类型	项目期限	项目金额	负责人
科技部	国家重点研发计划子课题	2020 年 -2025 年	50 万（元）	侯 琳
江苏先声医学诊断有限公司	合作研究	2020 年 -2022 年	50 万（元）	侯 琳
中国科学院上海生命科学研究院	上海市科技重大专项	2019 年 -2020 年	27.5 万（元）	侯 琳
粤港澳大湾区数字经济研究院 (福田)	合作研究	2021 年 -2024 年	224 万（元）	俞 声
北京市自然科学基金	重点项目课题	2019 年 -2023 年	72 万（元）	俞 声
国家自然科学基金	青年项目	2019 年 -2021 年	24 万（元）	俞 声
清华大学	理科发展双 E 计划科研基金	2019 年 -2020 年	54 万（元）	俞 声
清华大学	医工结合创新支持计划	2019 年 -2020 年	40 万（元）	俞 声
北京逸风金科软件有限公司	合作研究	2018 年 -2020 年	100 万（元）	俞 声
科技部	国家重点研发计划子课题	2018 年 -2020 年	30 万（元）	俞 声
国家自然科学基金	面上项目	2021 年 -2024 年	52 万（元）	刘汉中
清华大学国强人工智能研究院	校内委托项目课题	2021 年 -2023 年	30 万（元）	刘汉中
国家自然科学基金	青年项目	2018 年 -2020 年	24 万（元）	刘汉中
科技部	国家重点研发计划子课题（参与）	2017 年 -2020 年	43 万（元）	刘汉中
国家自然科学基金	面上项目	2020 年 -2023 年	52 万（元）	林 乾
北京市自然科学基金	重点项目课题	2019 年 -2023 年	65 万（元）	林 乾
北京智源人工智能研究院	青年科学家专项	2019 年 -2021 年	100 万（元）	林 乾
北京字节跳动网络技术有限公司	合作研究	2019 年 -2020 年	25 万（元）	林 乾
国家自然科学基金	青年项目	2020 年 -2022 年	28.9 万（元）	吴未迟



- 奖励荣誉 -

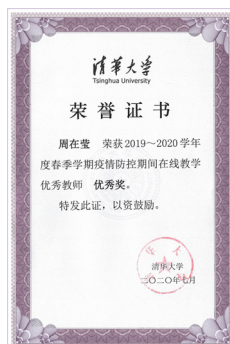
杨立坚教授荣获“2020年北京市普通高校优秀本科毕业设计(论文)优秀指导教师”

杨立坚教授指导的本科毕业设计(论文)《多维稠密函数型数据均值函数的同时置信区域构造》，获评“2020年北京市普通高校优秀本科毕业设计(论文)”，杨立坚教授获评优秀指导教师。



邓婉璐、周在莹两位老师荣获“2019-2020学年度春季学期疫情防控期间在线教学优秀教师”称号

疫情防控期间，清华大学统计学研究中心教学团队全体教师迅速反应，积极应对全新的教学模式和挑战，不断优化完善，打开线上线下融合式教学的新局面。其中，我中心邓婉璐、周在莹两位老师荣获“2019-2020 学年度春季学期疫情防控期间在线教学优秀教师”称号。



王天颖助理教授荣获“2020年国际数理统计协会(Institute of Mathematical Statistics, IMS)颁发的青年学者旅行奖(New Researcher Travel Awards)”



邓婉璐、周在莹两位老师在“清华大学第九届青年教师教学大赛”获奖

2020年12月，“清华大学第九届青年教师教学大赛”成功举办，我中心邓婉璐老师荣获“清华大学第九届青年教师教学大赛”(理科、医科组)一等奖；周在莹老师荣获“清华大学第九届青年教师教学大赛”(理科、医科组)二等奖。





中心师生在“清华大学抗击新冠肺炎疫情表彰大会荣获表彰”

2020年11月，清华大学举行抗击新冠肺炎疫情表彰大会。地球系统科学系宫鹏教授、徐冰教授领衔的“流行病学传播预测与对策”科技抗疫突击队荣获“清华大学抗击新冠肺炎疫情先进集体”荣誉称号。我中心邓柯副教授、侯琳副教授及博士生刘朝阳、沈翀、王掣、宋爽、余博作为突击队骨干成员，共同出席表彰大会接受表彰。



邓柯副教授荣获“世界华人数学家联盟最佳论文奖银奖”

2020年12月27-29日，“世界华人数学家联盟年会”在安徽合肥举行，清华大学统计学研究中心邓柯副教授作为第一作者的学术论文“On the unsupervised analysis of domain-specific Chinese texts”荣获“2020世界华人数学家联盟最佳论文奖-银奖”。该论文是邓柯副教授与美国哈佛大学 Peter Bol 教授、哈佛大学刘军教授和萨福克大学李佳漪副教授共同完成，论文发表于美国科学院院刊 PNAS 杂志。

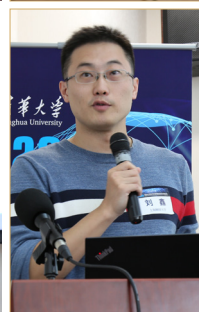
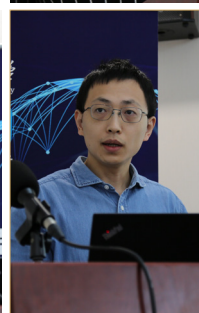
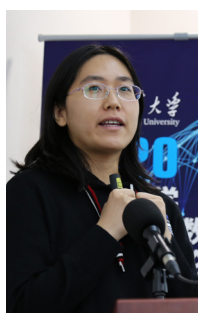


学术活动

- 主办学术活动 -

2020清华大学统计学与数据科学青年学者论坛

2020年11月14日，中心举办“2020 清华大学统计学与数据科学青年学者论坛”。论坛旨在促进国内青年统计和数据科学学者间的学术交流与合作，更好地推动统计学和数据科学的发展，同时加强与兄弟院校间的协同合作。会议由吴未迟、王天颖、张静怡三位助理教授发起并主持。来自清华大学，北京大学，中国人民大学，复旦大学，南开大学，浙江大学及上海财经大学等四十余名学者通过线上线下结合方式出席本次会议。





肿瘤多组学数据分析项目启动

2020年12月24日,清华大学统计学研究中心和先声诊断正式宣布达成合作,并举行“先声诊断-清华大学统计学研究中心合作项目启动会”。双方表示将致力于肿瘤前沿研究中的肿瘤免疫治疗标志物探索,就肿瘤异质性和分子分型以及肿瘤多组学数据整合分析等生信统计算法研究进行深入合作。清华大学统计学研究中心执行主任邓柯副教授、清华大学统计学研究中心侯琳副教授、清华大学统计咨询中心高级咨询师高梦昭博士、先声诊断常务副总裁崔欢喜先生、先声诊断生物信息高级总监郭昊博士、先声诊断市场部和品牌医学部高级总监宋超先生等嘉宾出席了本次活动。此次合作将共同推动肿瘤多组学生物信息统计分析前沿技术的探索和创新,助力精准医疗产业的发展,造福更多中国患者。



第五届北大-清华统计论坛

2021年6月10日,“第五届北大-清华统计论坛”在北京大学成功举办。北京大学、清华大学两校统计学科百余名师生参加了本次会议。北大-清华统计论坛是北大清华两校统计学科的传统学术活动,由北京大学统计科学中心和清华大学统计学研究中心联合发起,至今已成功举办五届。



第二届清华大学统计学教学改革研讨会

2021年6月25日,清华大学统计学研究中心发起并召开“2021第二届统计学教学改革研讨会”。会议旨在优化统计学相关课程教学体系,并为统计学科教学工作者搭建沟通与交流的平台。来自北京大学、中国人民大学、中央财经大学、北京师范大学等23家高校近五十名学者出席了本次研讨会。与会专家学者充分肯定了本次研讨会召开的必要性,并表示希望通过此次研讨会能切实地推动全国高校统计学的教学改革工作。



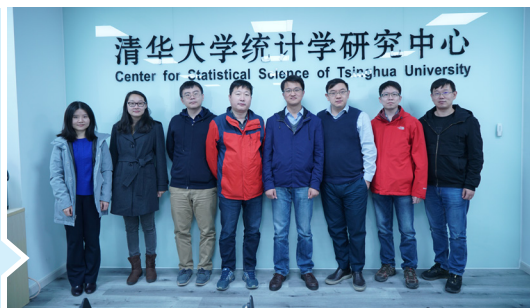


- 统计学与数据科学论坛 -

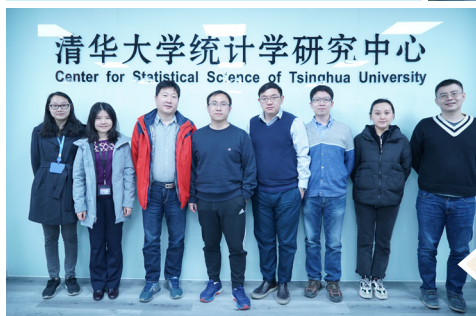
时间	主讲人	工作单位及职位	报告题目
2020/8/10	张日权	华东师范大学统计学院教授	函数型数据统计分析
2020/8/17	樊晓丹	香港中文大学统计系教授	Statistical Analysis for Tree-shaped Data
2020/9/3	孟 澄	中国人民大学统计与大数 据研究院助理教授	More Efficient Approximation of Smoothing Splines via Space-filling Basis Selection
2020/10/8	刘妍岩	武汉大学数学与统计学院教授	Histopathological Imaging-based Cancer Hetero- geneity Analysis via Penalized Fusion with Model Averaging
2020/10/19	秦国友	复旦大学公共卫生学院教授	Subgroup Analysis with Measurement Error
2020/10/26	Jun Yan	康涅狄格大学统计系教授	Moving-Resting Process with Measurement Error in Animal Movement Modeling
2020/11/6	刘玉坤	华东师范大学统计学院教授	估计总体大小的半参数全似然方法
2020/11/9	陈 钊	复旦大学大数据学院 青年研究员	Asset Selection Based on High Frequency Sharpe Ratio
2020/11/16	Yannis G. Yatracos	清华大学丘成桐数学 中心教授	Learning with Matching in Data-Generating Experiments
2020/11/27	韩 杨	曼彻斯特大学数学系 助理教授	Confident and Logical Selection of the Cut-point of a Biomarker for Patient Targeting
2020/11/30	朱复康	吉林大学数学学院教授	Softplus INGARCH Models
2020/12/7	Ying Wei	哥伦比亚大学生物统计系教授	Zero-inflated Quantile Rank-score Based Test with Application to scRNA-seq Differential Gene Expression Analysis
2020/12/14	俞章盛	上海交大学生物信息与 生物统计系教授	临床研究中基于影像的治愈率模型

时间	主讲人	工作单位及职位	报告题目
2020/12/21	李 扬	中国人民大学统计学院教授	Adaptive Randomization via Mahalanobis Distance
2021/3/5	王禹皓	清华大学交叉信息研究院助理教授	Debiased Inverse Propensity Score Weighting for Estimation of Average Treatment Effects with High-Dimensional Confounders
2021/3/26	Bing Li	宾州州立大学统计系教授	Sufficient Graphical Models
2021/4/12	苏良军	清华大学经济与管理学院教授	Uniform Inference in Linear Panel Data Models with Two-Dimensional Heterogeneity
2021/4/19	Peng Zeng	奥本大学数学与统计系副教授	Huber Regression and its Degrees of Freedom
2021/4/26	杨世昊	佐治亚理工学院工业系统学院助理教授	Inference of Dynamic Systems form Noisy and Sparse Data via Manifold-constrained Gaussian Process
2021/5/7	何子淮	斯坦福大学神经科学研究所助理教授	Identification of Putative Causal Loci in Whole-genome Sequencing Data via Knockoff Statistics
2021/5/14	李 根	密西根安娜堡大学生物统计系助理教授	New Amalgamation-Based Methods for Microbiome Compositional Data Analysis
2021/5/17	王 丹	上海纽约大学助理教授	Estimating and Forecasting Volatility using Leverage Effect
2021/5/24	余 睿	西南财经大学统计学院副教授	Location-adaptive Change-point Testing for Time Series
2021/5/31	Miaoyan Wang	威斯康辛麦迪逊大学统计系助理教授	Beyond the Signs: Nonparametric Tensor Completion via Sigh Series
2021/6/7	Wenyang Zhang	约克大学统计系教授	High Dimensional Dynamic Covariance Matrices with Homogeneous Structure
2021/6/11	柯 峥	哈佛大学统计系助理教授	Optimal Adaptivity of Signed-Polygon Statistics for Network Testing
2021/7/22	邬荣领	宾州州立大学统计遗传中心教授	Learning High-order Dynamical Interactome Networks from Big Static Data

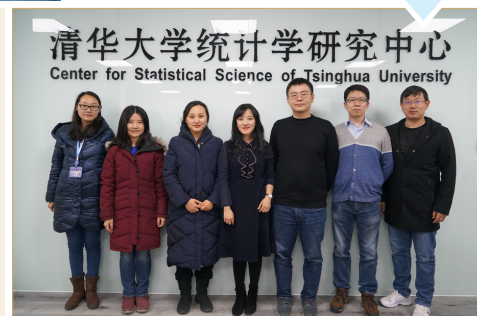
华东师范大学刘玉坤教授
访问我中心，并做学术报告



曼彻斯特大学韩杨助理教授
访问我中心，并做学术报告



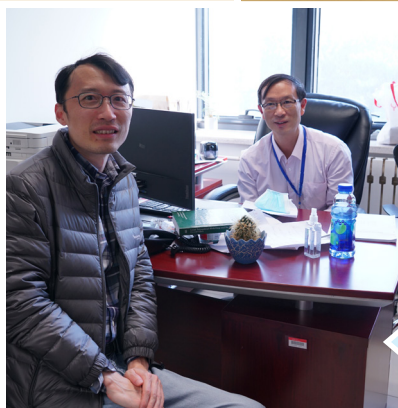
复旦大学陈钊研究员访
问我中心，并做学术报告



清华大学Yannis G. Yatracos
教授访问我中心，并做特邀
报告



复旦大学秦国友教授访
问我中心，并做学术报告



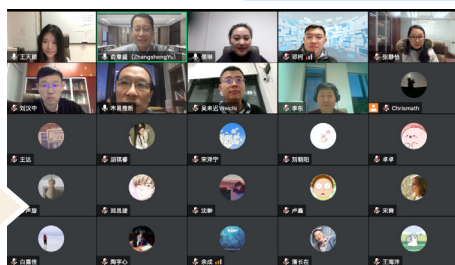
清华大学苏良军教授访问我
中心，并做特邀报告



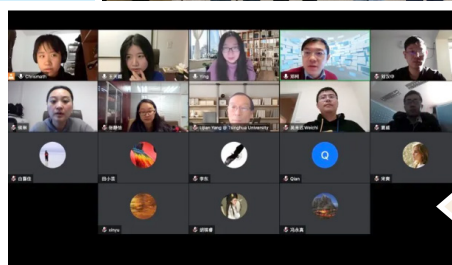
清华大学交叉信息研究院王禹
皓助理教授访问我中心，并做
学术报告



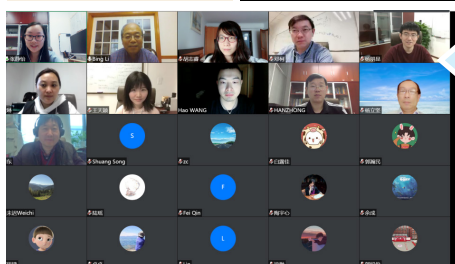
上海交通大学俞章盛
教授线上特邀报告



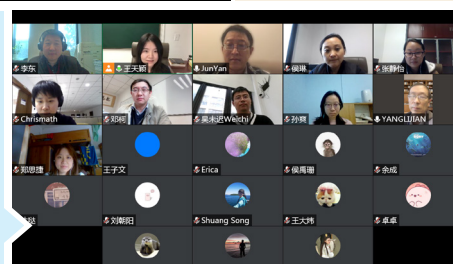
哥伦比亚大学Ying Wei
教授线上特邀报告



宾州州立大学Bing Li教授
线上特邀报告



康涅狄格大学Jun Yan教授
线上特邀报告



- 参加学术活动 -

杨立坚

- 2021 年 05 月 中国·西安 西安交通大学“西安统计与数据科学论坛”（邀请报告）
- 2021 年 04 月 中国·北京 北京大学公共卫生学院生物统计系（邀请报告）
- 2021 年 03 月 中国·上海 上海交通大学数学科学学院，统计论坛暨学科建设研讨会（发言）
- 2021 年 01 月 中国·上海 上海交通大学数学科学学院（邀请报告）
- 2021 年 01 月 中国·成都 西南财经大学“统计与计量经济学术报告”（邀请报告）
- 2020 年 12 月 中国·杭州 浙江工商大学统计与数学学院，统计学科建设研讨会（发言）
- 2020 年 11 月 中国·北京 中国人民大学统计与大数据研究院，统计学与数据科学学科建设咨询会（发言）
- 2020 年 11 月 中国·厦门 厦门大学邹志庄经济研究中心（邹志庄邀请报告）
- 2021 年 11 月 中国·上海 上海财经大学统计与管理学院（邀请报告）
- 2021 年 11 月 中国·北京 中国科学院数学院（邀请报告）
- 2020 年 10 月 中国·北京 中国人民大学统计学院，新中国统计教育 70 年：回顾与展望暨中国人民大学统计学科建设研讨会
- 2020 年 10 月 中国·北京 狗熊会（狗熊会学术报告）
- 2020 年 08 月 中国·北京 International Biometric Conference（特邀报告）
- 2020 年 08 月 中国·成都 西南财经大学统计学院，统计学科“十四五”规划专家咨询会（发言）
- 2020 年 07 月 中国·北京 清华大学长聘教授讲坛（邀请报告）

邓 柯

- 2021 年 06 月 中国·深圳 南方科技大学统计学系（邀请报告）
- 2020 年 12 月 中国·合肥 2020 世界华人数学家联盟年会（邀请报告）
- 2020 年 12 月 中国·上海 第二届中国数字人文年会（邀请报告）
- 2020 年 11 月 中国·长春 东北师范大学数学与统计学院（邀请报告）
- 2020 年 11 月 中国·深圳 南方科技大学统计学系（邀请报告）
- 2020 年 11 月 中国·广州 暨南大学经济学院（邀请报告）
- 2020 年 10 月 中国·北京 2020 中国计算机大会（大会论坛）
- 2020 年 10 月 中国·上海 上海交通大学数学科学学院（邀请报告）
- 2020 年 07 月 中国·开封 河南大学统计系（邀请报告）

李 东

- 2021 年 04 月 中国·北京 中科院数学与系统科学研究院（邀请报告）
- 2020 年 12 月 中国·贵阳 全国工业统计教学研究会数字经济与区块链技术协会成立大会
- 2020 年 12 月 中国·厦门 厦门大学现代统计学研讨会（邀请报告）
- 2020 年 06 月 中国·长春 吉林大学数学学院（邀请报告）

侯 琳

- 2021 年 06 月 中国·上海 中科院上海营养与健康研究所（邀请报告）
- 2021 年 05 月 中国·青岛 山东大学数学与交叉科学研究中心（邀请报告）
- 2021 年 04 月 中国·北京 北京大学统计科学中心（邀请报告）
- 2021 年 03 月 中国·线上 狗熊会（邀请报告）
- 2020 年 11 月 中国·北京 中国数学会医学数学专业委员会成立仪式暨第一届医学数学论坛（邀请报告）
- 2020 年 10 月 中国·线上 BioAsia2020（邀请报告）



俞 声

- 2021 年 06 月 中国·西安 Statistical Learning Methods in Modern AI (邀请报告)
- 2020 年 12 月 中国·线上 第 13 届中国 R 会议 (邀请报告)
- 2020 年 11 月 中国·北京 北京协和医院信息中心 (邀请报告)
- 2020 年 11 月 中国·北京 中国临床神经峰会 (邀请报告)
- 2020 年 10 月 中国·线上 中华医学会第 26 次全国医学信息学术年会 (邀请报告)
- 2020 年 10 月 中国·北京 2020 中国罕见病大会 (邀请报告)

刘汉中

- 2021 年 06 月 中国·北京 首都师范大学 (邀请报告)
- 2020 年 07 月 中国·线上 上海交通大学 (邀请报告)

林 乾

- 2020 年 12 月 中国·北京 北京交通大学 (邀请报告)
- 2020 年 12 月 中国·北京 清华大学 - 电子科技大学降维方法研讨会 (组织者)
- 2020 年 11 月 中国·北京 首都师范大学 (邀请报告)
- 2020 年 11 月 中国·珠海 北京师范大学珠海分校 第二届自然科学交叉前沿科学论坛 (邀请报告)

吴未迟

- 2020 年 12 月 英国·线上 CFE-CMStatistics (session chair)
- 2020 年 11 月 中国·北京 2020 清华大学统计学与数据科学青年学者论坛 (组织者)

王天颖

- 2021 年 06 月 巴西·线上 Department of Statistics, Universidade Federal de Pernambuco (邀请报告)
- 2021 年 06 月 中国·杭州 浙江大学 (邀请报告)
- 2021 年 03 月 中国·线上 Eastern North American(ENAR) Spring Meeting (邀请报告)
- 2020 年 12 月 中国·线上 ICSA 2020 Applied Statistics Symposium (邀请报告)
- 2020 年 11 月 中国·线上 全国青年统计学家论坛 (邀请报告)
- 2020 年 11 月 中国·北京 2020 清华大学统计学与数据科学青年学者论坛 (组织者)

张静怡

- 2020 年 11 月 中国·北京 2020 清华大学统计学与数据科学青年学者论坛 (组织者)
- 2020 年 10 月 中国·北京 中国人民大学青年统计学者论坛 (邀请报告)
- 2020 年 08 月 中国·线上 全国青年统计学家论坛 (邀请报告)

杨朋昆

- 2021 年 06 月 中国·北京 北京大学统计科学中心 (邀请报告)
- 2021 年 05 月 中国·北京 北京智源人工智能研究院 (邀请报告)
- 2020 年 08 月 中国·线上 全国青年统计学家协会 (邀请报告)

胡志睿

- 2021 年 05 月 中国·武汉 生物信息学与智能信息处理 2021 年学术会议 (BIIP2021)



人才培养

- 本科生培养 -

统计学专业辅修学位

2016 年开始，清华大学统计学研究中心开设统计学辅修项目，为学有余力的清华学子开设优质统计学课程。课程设计参考国际一流统计学科本科生培养方案，理论和应用并重，力争培养兼具统计学理论基础和应用能力的跨学科人才。近年来，辅修项目受到同学们的热烈欢迎，开设课程供不应求，多次响应同学们呼吁增加课程容量，年度新增课程修读申请超二百人次。2021 年初，清华大学工业工程系统统计学本科专业通过教育部审批备案，辅修统计学专业并满足培养方案培养要求的同学可授予统计学辅修专业学位。

课程编号	课程名称	学分	学期	先修要求
必修课程 16 学分				
40160713	初等概率论	3	秋	微积分、线性代数
30160263	统计推断	3	秋	微积分、线性代数
40160763	多元统计分析	3	春	统计推断、初等概率论
40160803	线性回归分析	3	春	统计推断、初等概率论
30160294	统计计算与软件	4	秋	统计推断、初等概率论





“数据思维与实践”课程证书项目

2020 年开始，清华大学统计学研究中心开设“数据思维与实践”课程证书项目。项目对修读学生的专业及年级要求放宽，以满足对统计学科有热情但无法达到辅修学位要求的同学们的修读需求。项目必修课两门，共 6 学分；选修课从课组中的 15 门课程自由选择修读，学分要求不少于 9 学分。完成项目要求并通过资格审查的同学可授予课程项目证书。

通识课建设：《统计学引论：数据分析的科学与艺术》

2018 年，清华大学统计学研究中心对标国际顶尖院校的统计学通识课程，筹备建设了清华大学首门统计学通识课，以应对大数据时代各专业领域对统计学的强烈需求。课程不但面向全校各专业本科学生开放，更加入清华 - 北大互选课名单，预留部分名额，向兄弟院校的学生开放。课程由中心长聘副教授邓柯、李东两位老师讲授。内容包括统计学基本思想、基础理论和基本方法的理论，横跨工程、生物医学、经济金融、人文社科等多个专题，融合数据分析实践和统计编程入门的实践。

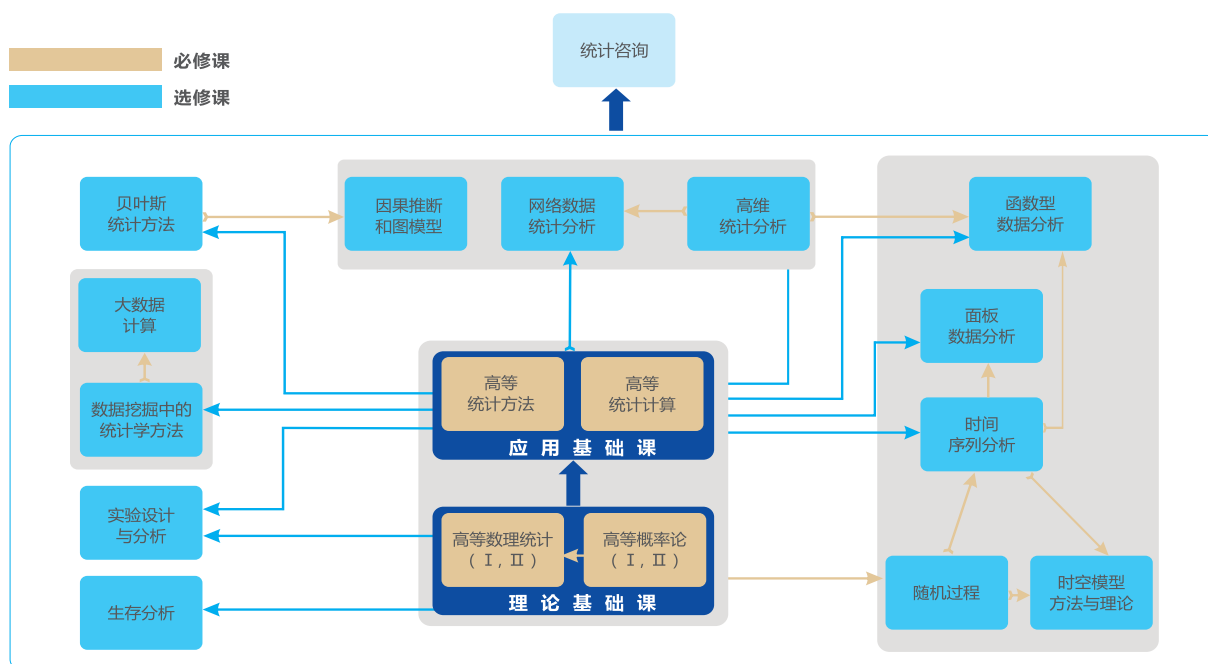


- 研究生培养 -

统计学博士

培养德智体全面发展,掌握扎实统计学基础理论和系统专业知识,具有独立从事统计学原创性研究和应用能力的统计学人才;使得学生具有统计学素养,掌握学术规范,独立开展学术研究和进行学术交流的能力;指导学生应用统计学知识解决实际问题,在有关的统计学研究方向上做出有重要理论或实际应用的创新性成果;毕业以后,适合于在高等学校、科研机构、政府部门、企事业单位中从事统计学及其相关领域的教学、科研、管理等方面的研究和工作。主要研究方向:数理统计、生物与医学统计、计量经济与金融统计、大数据统计、工业统计、统计计算等。

课程设置

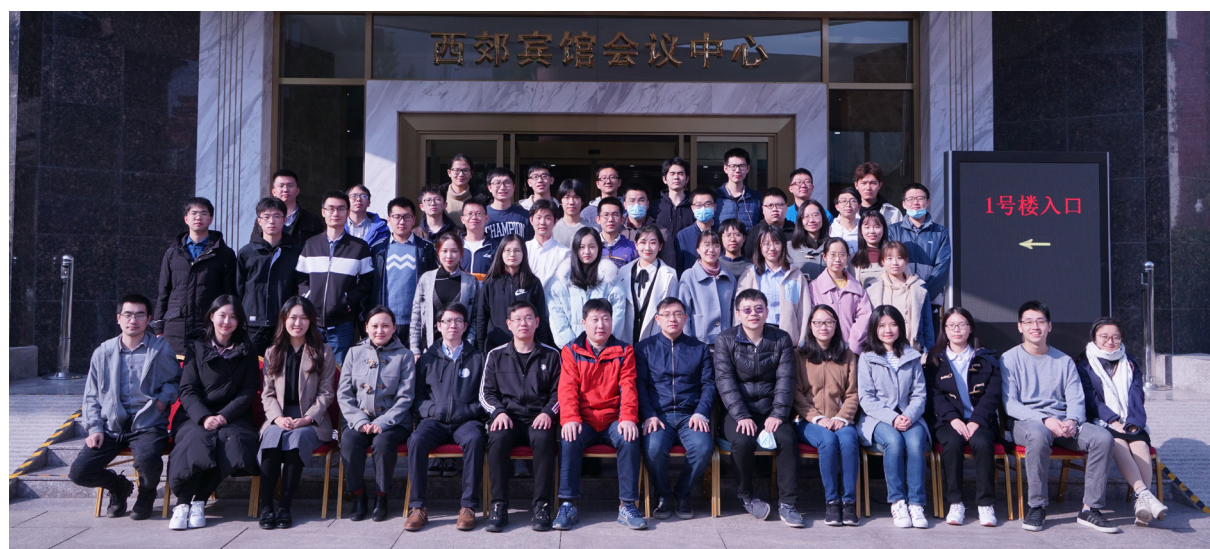




博士生论坛

2020年10月30日，由清华大学统计学研究中心发起并组织的“2020年清华大学统计学博士生论坛”在北京西郊宾馆成功举办。统计学博士生论坛是清华大学统计学研究中心的传统学术活动，旨在为青年统计学者提供学术交流的平台，以提高统计学者的专业知识及专业素养。按照清华大学统计学博士研究生培养方案，中心二年级及以上的博士生每年都要汇报自己的研究进展。

来自清华大学统计学研究中心的40余名在读博士生参与了此次论坛，青年学者们根据各自研究方向，分享最新研究成果以及在研究中遇到的问题，切磋技艺，相互交流，受益匪浅。



学生获奖情况

奖项名称	获奖者	具体信息	时间	导师
 “2020 数字人文年会” 优秀论文二等奖	王海洋	《基于卷积神经网络的近代广告图片聚类方法》	2020 年 10 月	邓 柯
 第六届全国高校研究生统计论坛十佳论文奖	李艺超	Stratification and Optimal Resampling in Sequential Monte Carlo	2020 年 11 月	邓 柯
 第六届全国高校研究生统计论坛十佳论文奖	蒋斐宇	Adaptive Inference for a Semiparametric Generalized Autoregressive Conditional Heteroskedasticity Model	2020 年 11 月	李 东
 中国现场统计研究会第四届全国统计学博士研究生学术论坛论文二等奖	李 杰	Statistical Inference for the Stationary Mean of Functional Time Series	2020 年 11 月	杨立坚
 中国现场统计研究会第四届全国统计学博士研究生学术论坛论文优秀奖	胡祺睿	Simultaneous Confidence Regions of the Mean Function based on Multivariable Functional Data	2020 年 11 月	杨立坚
 中国现场统计研究会第四届全国统计学博士研究生学术论坛论文优秀奖	张心雨	On the Least Squares Estimation of Multiple-Threshold-Variable Autoregressive Models	2020 年 11 月	李 东
 2020 年度百济神州青年论文二等奖	袁 正	CODER: A Pre-trained Cross-lingual Knowledge embedded Medical Term Representation for Term Normalization	2020 年 12 月	俞 声



奖项名称	获奖者	具体信息	时间	导师
 2020 年度百济神州青年论文优秀奖	宋 爽	Openness Weighted Association Studies: Leveraging Personal Genome Information to Prioritize Non-coding Variants	2020 年 12 月	侯 琳
 2021 年度汉南研究生差旅奖 (2021 IMS Hannan Graduate Student Travel Award)	钟 晨	Statistical Inference for Functional Time Series: Autocovariance Function	2021 年 4 月	杨立坚
  2021 年国际统计学协会简·丁伯根奖 (2021 ISI Jan Tinbergen Award)	李 杰 胡祺睿	Prediction Interval of Air Pollutants Concentration by Nonparametric Regression Analysis	2021 年 5 月	杨立坚
 第五届北大 - 清华统计论坛 “优秀毕业生”	蒋斐宇	北大 - 清华统计学科优秀毕业生	2021 年 6 月	李 东
 第五届北大 - 清华统计论坛 “优秀海报奖”	黄 昆	Inference for Dependent Error Functional Data with Application to Event Related Potentials	2021 年 6 月	杨立坚
 第五届北大 - 清华统计论坛 “优秀海报奖”	张心雨	On the Smooth Transition Moving Average Models	2021 年 6 月	李 东
 第二十五届清华大学研究生 “学术新秀”	蒋斐宇	清华大学研究生学术新秀	2021 年 6 月	李 东

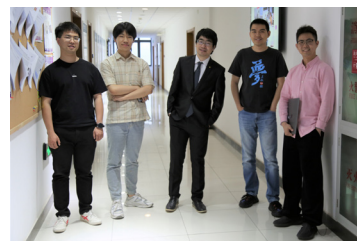


毕业生专栏

统计学研究中心 2020 年 7 月 -2021 年 6 月共七名同学顺利毕业：

2020 届：黄伟清、杨 洋

2021 届：余丽珊、张园园、蒋斐宇、李艺超、林毓聪



黄伟清 导师：侯琳 / 毕业去向：华为

在统计中心学习五年，我感到非常的幸运。统计中心是一个充满活力的大家庭，在这里，我们能与顶尖的学术大师进行交流，能感受到严谨的科研氛围。同时，我们也享受着自由的学术环境，可以尝试着去研究自己感兴趣的科研问题。除了科研活动，我们也有丰富的业余活动，让生活多姿多彩。读博不是一件很轻松的事，来到统计中心是一个无悔的选择，在这里，我们能克服种种困难，更加坚定自己的信念，不断追逐自己的梦想。愿统计中心的发展日新月异，培养出更高层次的人才！

杨 洋 导师：邓柯 / 毕业去向：腾讯公司微信事业部

在清华园度过了五年直博生活，如今我已毕业工作一年，蓦然回首，感慨万分。这五年间，中心给我们提供了丰富的学习资源和良好的科研环境。中心各位老师同学都非常优秀，在学术上造诣深厚，在生活中平易近人，他们都是我学习的榜样。这五年里我非常感谢我的导师邓柯老师，科研上大到研究方向的指引，小到论文文稿逐字逐句修改，邓老师都事无巨细给予我耐心指导；当我在生活中遇到困难时，邓老师也给予我莫大的帮助。希望我能一直秉持清华人“自强不息、厚德载物”的精神，为祖国健康工作五十年，最后衷心祝愿清华统计学研究中心越办越好。

余丽珊 导师：俞声 / 毕业去向：北京雁栖湖应用数学研究院

感谢统计中心对我的培养。中心频繁的学术活动给我们提供了一个非常优越的学术科研环境。每年开展的博士生论坛、每学期开展的清北统计论坛以及每周至少一次的学术报告给我们提供了一个很好的学术学习和交流的机会，让我们能够及时了解学术界各领域的前沿研究，不仅拓展了学术视野，也增强了学术能力。中心每学期的出游也极大地增加了同学们和师生间的感情。中心老师们不遗余力地打造了一个优良的学术软环境，使同学们安心学术、受益良多。

张园园 导师：杨立坚 / 毕业去向：苏州大学

统计学研究中心为我们营造了良好的学术氛围，在中心学习的过程中，我真切感受到了学术共同体对于学术研究的重要性，身处这样的氛围中，我对“科研工作者”这个标签的认知越来越全面、立体。我想，笃守研究兴趣，严谨从事研究，拥有一颗“平常心”是科研工作者的基本操守。不仅如此，我们也更应该厚植家国情怀，勇担社会责任，能够将统计学的研究成果应用于实际需求。衷心希望中心的发展能够蒸蒸日上，人才辈出。



蒋斐宇 导师：李东 / 毕业去向：复旦大学

最近也是顺利毕业了，非常感谢清华大学统计中心五年来的培养。作为中心正式招收的第一届博士生，我从课程设置、博士生培养、交流访问的机会等多个方面都感受到了中心的不断投入。我记得最开始中心只有几位老师在给我们开设博士生必修课，大部分时候只有我们 16 级入学的五个博士生在听，但老师们都一丝不苟，认真备课。现在随着师资队伍的不断壮大，中心也开了很多统计学前沿方向的选修课，大大拓宽了我们的学术视野，在学校里也不断收到其他同学的肯定，旁听的、选修的、辅修的人数一年比一年多，我也是由衷为中心高兴。统计中心在师资力量、学术资源和科研环境等多方面来说都在逐渐成为全国的领跑者，这里也希望统计中心越办越好，能够引领清华乃至中国统计学科的发展，为国家各行各业培养和输送统计学领域的顶尖人才。

李艺超 导师：邓柯 / 毕业去向：清华大学

在清华度过的九年时间注定是我生命中不可思议的一段经历。九年间，我逐渐摆脱了学术上的稚气走向成熟，特别是在统计中心的五年博士研究生涯，更加坚定了我从事科研工作的决心和信心。中心为我们提供的不仅仅是丰富的科研资源，更重要的是一个充满活力的科研环境和广阔的平台。中心的各位老师同学都非常优秀，他们都是我的榜样。其中，我的导师邓柯老师，不仅仅在科研学习上指导我前行，当我遇到困难时也为我提供了很多帮助与指导。接下来的日子里，我会继续和中心一起成长。希望自己成长为一个独当一面的科研学者，也祝愿清华统计能够稳健地发展成世界一流的统计学科。

林毓聪 导师：俞声 / 毕业去向：北京理工大学

很幸运人生能有机会在统计中心接受五年的学术熏陶！若要谈及给同学的建议，我建议大家多花些时间思考。博士有许多的时间可供支配，因此可以多去找你敬佩的学长、老师和朋友聊聊各个职业的利弊是什么；多去读读你钦佩的人的传记，看看这是不是你想要的人生（想成为高斯欧拉的话 emmm，想想就行了）；在做研究时，想想除了写论文让你毕业以外，这个研究更深层次的意义是什么。聪明能走得快，但多思考才能走得更远。



- 优秀大学生夏令营 -

2021年5月28-30日，“清华大学2021全国优秀大学生统计与数据科学交流会（暨优秀大学生夏令营）”活动成功举办。本次活动共收到来自全国99所高校的390份报名申请，经过层层筛选，来自清华大学、北京大学、北京师范大学、浙江大学、上海财经大学、西安交通大学等多所国内知名高校的共43名同学通过线上线下结合的方式参加了此次活动。





社会服务及影响

- 社会服务 -

杨立坚

- 2021 年 7 月 教育部“长江学者奖励计划”评审专家
- 2021 年 4 月 国家自然科学基金项目评审专家
- 2020 年 11 月 教育部“长江学者奖励计划”评审专家
- 2020 年 7 月 西安交通大学全球健康研究院兼职教授
- 2020 年 7 月 科学出版社统计与数据科学系列丛书编委
- 2020 年 7 月 苏州市现场统计研究会第六届理监事会顾问
- 2020 年 7 月 中国现场统计研究会生存分析分会第四届理事会常务理事
- 2020 年 7 月 中国现场统计研究会高维数据分会第一届理事会常务理事
- 2019 年 10 月 - 至今 清华大学第十二届学位评定数学分委员会委员
- 2019 年 9 月 - 至今 中国指挥与控制学会医工结合专业委员会委员
- 2020 年 7 月 清华大学理科学术委员会委员

邓 柯

- 2021 年 02 月 - 至今 External Reviewer of the Research Grants Council of Hong Kong
- 2021 年 01 月 - 2023 年 12 月 国家抗肿瘤药物临床应用专家委员会委员
- 2020 年 01 月 - 2022 年 12 月 清华大学教学顾问组成员
- 2020 年 01 月 - 2022 年 12 月 清华大学信息化用户代表委员会委员
- 2019 年 04 月 - 2021 年 03 月 北京智源人工智能研究院研究员
- 2019 年 04 月 - 至今 中国国家自然科学基金评审专家
- 2019 年 03 月 - 2023 年 03 月 中国现场统计研究会数据科学与人工智能分会副理事长
- 2018 年 12 月 - 2022 年 12 月 北京生物医学统计与数据管理研究会副会长
- 2018 年 12 月 - 2022 年 12 月 中国青年统计学家协会副会长
- 2018 年 12 月 - 2022 年 12 月 北京市统计学会理事
- 2018 年 10 月 - 2023 年 09 月 高等教育出版社第二届“现代统计学系列丛书”编委会成员
- 2018 年 01 月 - 2021 年 12 月 Board Member of IASC-ARS (国际计算统计学会亚太地区分会理事)
- 2017 年 11 月 - 2021 年 11 月 中国现场统计研究会理事
- 2017 年 03 月 - 2021 年 03 月 中国现场统计研究会计算统计分会理事长
- 2017 年 03 月 - 2022 年 02 月 中国研究型医院学会医疗信息分会医疗和临床科研大数据应用专委会委员



李 东

- 2021年09月-2026年09月 北京应用统计学会理事
- 2020年12月-2024年12月 全国工业统计学教学研究会数字经济与区块链技术协会常务理事
- 2019年10月-2023年10月 北京大数据协会常务理事
- 2019年04月-2023年04月 中国青年统计学家协会常务理事
- 2018年12月-2022年12月 全国工业统计学教学研究会第9届理事会常务理事
- 2018年10月-2022年10月 中国概率统计学会第11届理事会副秘书长
- 2017年03月-2021年03月 中国现场统计研究会计算统计分会理事

侯 琳

- 2017年03月-至今 中国现场统计研究会计算统计分会常务理事、秘书长

俞 声

- 2020年12月-2024年12月 中国现场统计研究会中国旅游大数据分会理事
- 2019年05月-至今 全国工业统计学教学研究会中国青年统计学家协会理事
- 2019年03月-至今 中国现场统计研究会数据科学与人工智能分会理事
北京生物医学统计与数据管理研究会副秘书长
- 2017年03月-至今 中国现场统计研究会计算统计分会理事

刘汉中

- 2019年-至今 全国工业统计学教学研究会青年统计学家协会第一届理事会理事
- 2019年-至今 北京应用统计学会理事
- 2016年-至今 中国现场统计研究会计算统计分会副秘书长

林 乾

- 2019年04月-至今 北京智源人工智能研究院“青年科学家”
- 2018年11月-至今 中国现场统计研究会计算统计分会副秘书长

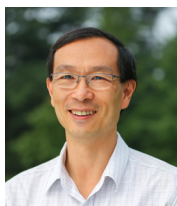
吴未迟

- 2020年12月-2024年12月 全国工业统计学教学研究会数字经济与区块链技术协会理事



- 学术杂志服务 -

统计学研究中心师资团队均在各自领域有很深厚的学术造诣和学术地位，长年为数理统计领域、计量经济学领域、医学信息学领域、生物医学领域、数据科学领域等国际知名学术期刊审稿。如 Annals of Statistics、Journal of the American Statistical Association、Journal of the Royal Statistical Society: Series B、Biometrika、Journal of Econometrics、Journal of Time Series Analysis、Statistica Sinica、PLOS、Genetics、Briefings in Bioinformatics、Journal of Machine Learning Research、Bernoulli、BMC Bioinformatics 等。



杨立坚 教授

- Stat
- Statistica Sinica

副主编

副主编



邓柯 副教授

- Statistica Sinica
- 《数字人文》
- 《应用概率统计》
- 《应用数学和力学》

副主编

编委

编委

编委



侯琳 副教授

- Statistics in Biosciences
- Quantitative Biology

副主编

编委

- 统计咨询服务 -

统计咨询中心概况

清华大学统计咨询中心由清华大学统计学研究中心发起并成立，于 2017 年 5 月开始运行，面向清华各院系及社会各界提供数据分析和统计咨询服务。

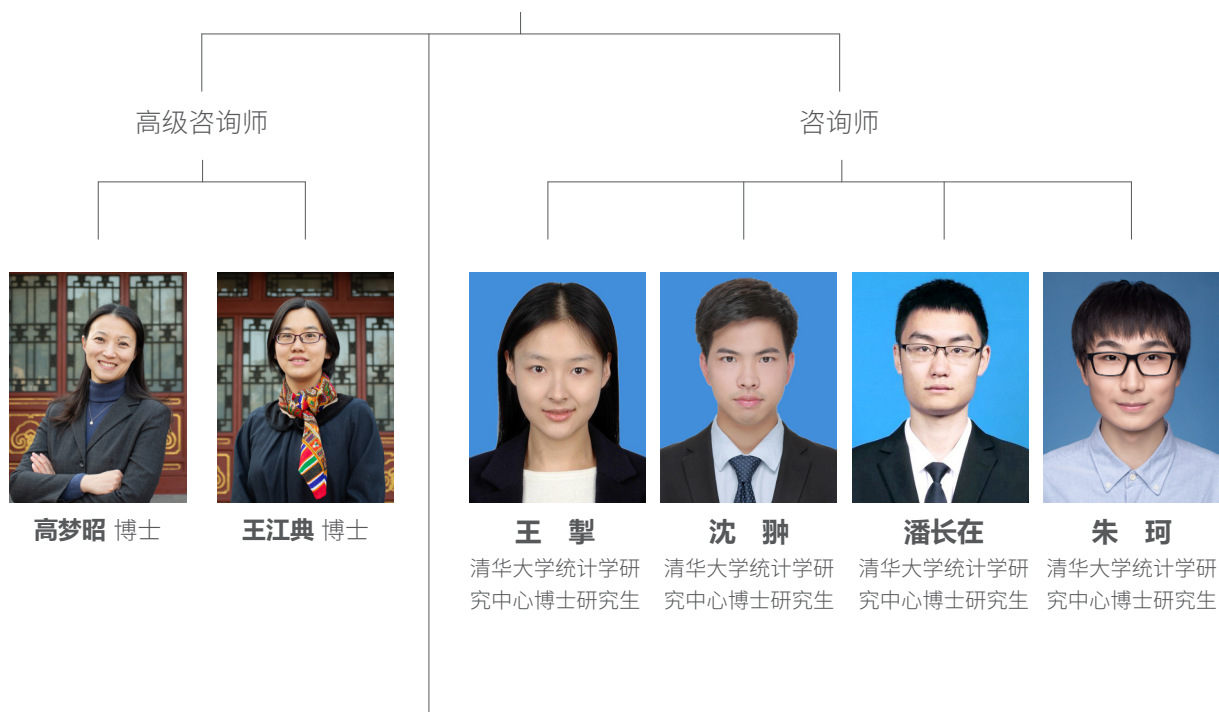
统计咨询中心一方面为清华大学各院系的广大师生、校外实体及企业提供高水平的统计建议与支持；另一方面，以跨专业为基础，通过合作鼓励统计学者和其他领域人员之间的交叉研究；同时，培养统计学研究生与其他领域研究者的交流互动能力，成为能够运用统计学技巧解决实际问题的合作者。

业务团队

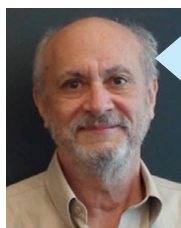


邓 柯 主任

清华大学统计学研究中心长聘副教授



学术顾问



Donald Rubin 教授

- 清华大学特聘教授
- 美国科学院院士



Wing Hung Wong 教授

- 斯坦福大学统计系终身教授
- 美国科学院院士



许宪春 教授

- 清华大学经济管理学院教授
- 国家统计局原副局长

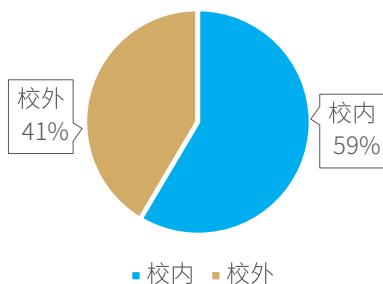


年度业务总览

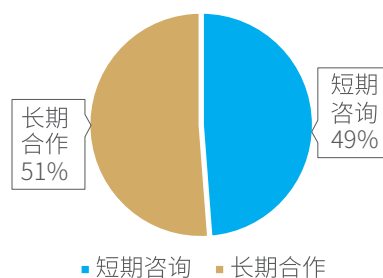
本年度统计咨询中心共承接咨询案例 41 例，包括秋季学期为 21 例，春季学期为 20 例。其中校内咨询 24 例，占比 56%；校外咨询 17 例，占比 44%。在校内咨询中，短期咨询占比 62.5%，长期合作占比 37.5%；校外咨询中短期咨询占比 29%，长期合作占比 71%。无论项目来源于校内还是校外，长期合作所占比重越来越大，说明咨询中心的业务逐渐转向更加紧密深入的合作方式。

咨询中心持续为校内外客户提供高质量的专业咨询服务。客户组成也更加多元化，特别是校外咨询，本年度客户来自政府、高校、医院、科研单位、公司等各行各业。

2020~2021 学年项目来源



2020~2021 学年项目类型



校内项目来源	项目数量
工程物理系	1
教评中心	2
人文学院	2
社会科学学院	5
土木工程系	1
写作中心	1
医学院	2
车辆与运载学院	1

校内项目来源	项目数量
工业工程系	1
环境学院	1
建筑学院	3
交叉信息研究院	1
深圳国际研究生院	1
组织部	1
长庚医院	1

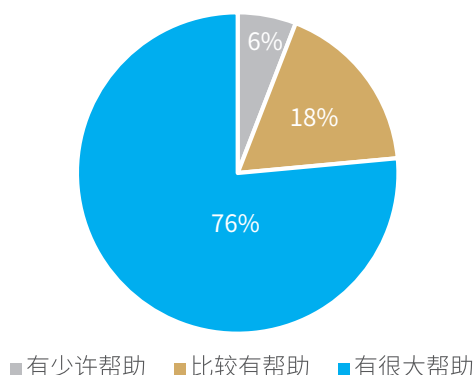
校外项目来源	项目数量
高校	2
公司	3
科研单位	1
医院	3
政府	8

满意度调查

本为了提高咨询服务质量，完善服务内容，在咨询结束后对客户进行服务满意度调查。根据本学年调查反馈，咨询客户对咨询中心服务满意度高达 100%，其中 83% 的客户认为咨询中心提供的咨询建议有很大帮助。

	咨询整体评价	咨询会议评价	会议纪要评价
很满意	76%	76%	76%
满意	18%	18%	24%

满意度调查-咨询建议评价



发表论文



共同发表 1:

Ruiyun Li, Bin Chen, Tao Zhang, Zhehao Ren, Yimeng Song, Yixiong Xiao, **Lin Hou**, Jun Cai, Bo Xu, Karen Kie Yan Chan, Ying Tu, Mu Yang, Jing Yang, **Zhaoyang Liu**, **Chong Shen**, **Che Wang**, Lei Xu, Qiyong Liu, Shuming Bao, Jianqin Zhang, Yuhai Bi, Yuqi Bai, **Ke Deng**, Wusheng Zhang, Wenyu Huang, Jason D Whittington, Nils Chr. Stenseth, Dabo Guan, Peng Gong and Bing Xu (2020). Suppression of future waves of COVID-19: global pandemic demands joint interventions. *Proceedings of the National Academy of Sciences of USA* **117**(42): 26151-26157.



共同发表 2:

Chen Chen, Mingyao Yao, Xu Luo, Yulin Zhu, **Zhaoyang Liu**, Hanchen Zhuo and Bin Zhao (2020). Outdoor-to -indoor Transport of Ultrafine Particles: Measurement and Model Development of Infiltration Factor. *Environmental Pollution* **267**,115402.



共同发表 3:

李建军, 高梦昭, 邓柯, 李艺超, 郭瀚民, 方海萍, 赵明刚 (2021)。进口鸡肉传带禽伤寒沙门氏菌的定量风险评估。《中国动物检疫》 **38**(1):30-34.



共同发表 4:

Feng Zhang, Changqing Jiang, **Yichao Li**, **Xiaoyue Niu**, Tiangang Long, Changgeng He, Jianqi Ding, Linze Li and Luming Li (2021). Investigation of Artifacts and Optimization in Proton Resonance Frequency Thermometry towards Heating Risk Monitoring of Implantable Medical Devices in Magnetic Resonance Imaging. *IEEE Transactions on Bio-medical Engineering*, DOI: 10.1109/tbme.2021.3081599 PMID: 34003743.



支持发表 1:

Chen Chen, Shan Liu, Wei Dong, Yi Song, Mengtian Chu, Junhui Xu, Xinbiao Guo, Bin Zhao and Furong Deng (2021). Increase Cardiopulmonary Effects of Ultrafine Particles at Relatively Low Fine Particle Concentrations. *Science of the Total Environment* **751**: 141726.



支持发表 2:

Xianli Lv and Shikai Liang (2020). Update Onyx Embolization for Plexiform Arteriovenous Malformation: Ante-grade Drifting Technique. *The Neuroradiology Journal* **33**(5):386-392.



年度优秀咨询师

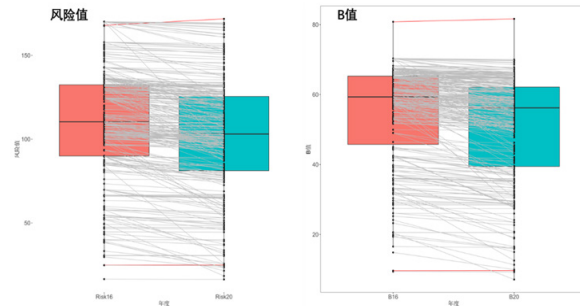


典型案例

案例一：海关进出口食品安全风险分析

在经济全球化、贸易全球化、分工精细化、产业链复杂化的今天，随着我国进口食品贸易的不断增长，如何监管进口食品安全风险是我国食品安全和国家安全面临的重大挑战。海关总署国际检验检疫标准与技术法规研究中心联合清华大学统计学研究中心邓柯副教授统计咨询中心团队，合作开展了“进出口食品安全风险评估模型研究”项目工作，深入分析了近十年来全国进口食品抽检检测结果数据，并结合进出口贸易数据、年度抽检计划等数据，构建了基于贝叶斯理论的进口食品安全风险模型，分别对产品、项目和国家（地区）这三个维度的风险效应进行稳健评估。研究成果帮助相关部门全面掌握进口食品安全情况，制订相应的风险防范措施；并且促进抽检资源的有效利用，实现资源的合理分配。

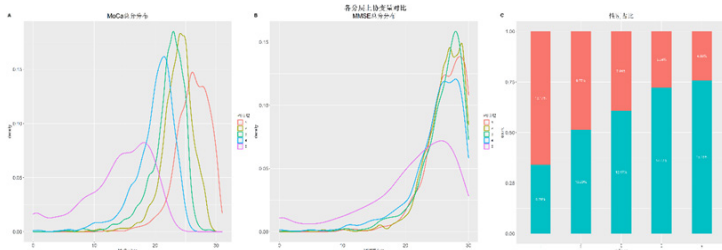
➤ 2016与2020年食品小类风险值与B值比较



案例二：广西省食品安全评价性抽检数据分析

为贯彻党的十九大精神，落实党中央、国务院以及自治区党委、政府关于食品安全监管的决策部署，广西壮族自治区从2018年起开展评价性抽检工作。为了更好地分析利用抽检数据，更加科学、客观反映广西食品安全状况，广西质检院委托清华大学统计学研究中心邓柯副教授统计咨询中心团队开展“广西食品安全评价性抽检数据分析研究”，构建多维度指标评价体系，创新性地采用了经验贝叶斯方法来计算食品安全评价指数，实现了包括地市、食品种类、检测项目，或地区与食品种类的交叉维度等多种指标的计算，从而为食品安全监管决策需要提供更多数据支持；并结合评估结果，对未来抽检计划的制定与部署产生指导性作用。

案例三：北医三院认知功能障碍缺失数据填补



阿尔兹海默症是一种常见于老年人的神经退行性疾病，我国目前是全球该病患者最多的国家。为了克服流行病学研究中无反应（non-response）现象对估计该病患率的负面影响，从2020年10月，统计咨询中心接受北医三院委托，对认知功能障碍数据中的缺失数据进行填补。在

邓柯副教授的带领下，咨询中心团队结合领域专家建议确定了数据的缺失机制，并基于统计学原理对缺失数据进行填补。经实践验证，基于倾向性得分分层的填补，各方法估计的患病率较为一致，弥补了该领域解决无反应现象的空白，保障了对我国阿尔兹海默症和轻度认知功能障碍群体人数的准确估计，为“健康中国2030”规划纲要提出的促进健康老龄化提供了有力的数据支持。

水木数据科学访谈 - 清华大学杰出访问教授 汤家豪院士

Talking with Great Minds-Howell Tong

采访 | 邓柯、张心雨 编辑 | 张心雨、陶宇心、侯禹珊



工业工程系党委书记李志忠教授(右)和中心杨立坚教授(左)为汤家豪教授颁发“杰出访问教授”聘书

1. Childhood and study experience oversea.

Q: Can you talk about your childhood?

A: In my childhood, my family and I were faced with general tough life conditions, but eventually we went through the hardship and became stronger. I am also very lucky that I always have good teachers. One of the teachers that I remember particularly once told us a story about Hua Luogeng when I was very young. That was the time Hua Luogeng returned to China. My teacher told me how he became famous after studying, and that actually made a quite impression on me. I think it's because of his story that I decided to study Mathematics. I moved to England in 1961 when my father worked there. The secondary school I attended was not a top school, but was very comprehensive. Under the support of the school and headmaster, I picked up English quickly. I was the only boy from that school who went to a university.

Q: In that period, did you encounter any challenge in your study or life?

A: Yes. First of all, I need to get used to the English way of schooling, for example moving across different classrooms to have lessons, and different dietary habit on campus. But luckily the students around were all very nice, and we became good friends. Because of this experience, I was able to know different culture and their way of speaking. It's quite a big challenge to adapt from the Hong Kong school system to a working class environment in London at that time.

Q: Is there any particular reason that you choose Statistics as your major?

A: Well as I said I decided to study Mathematics, and I graduated in Mathematics in Manchester. We had good statistics teachers and received many statistics courses, which was unusual at that time in England. I also got a chance to listen to a lecture about probability theory given by an eminent probability researcher. I was impressed by his lecture and became interested in probability theory. Because of my family, I decided to suspend the post-graduate study and took a job. During that time, I started to read some papers, and I came across one paper about time series by my former teacher in Manchester. That's how I became interested in time series. I wrote to him and went back to Manchester. Due to some reasons, I accidentally became a university teacher teaching statistics instead of a post-graduate student. I was very lucky.



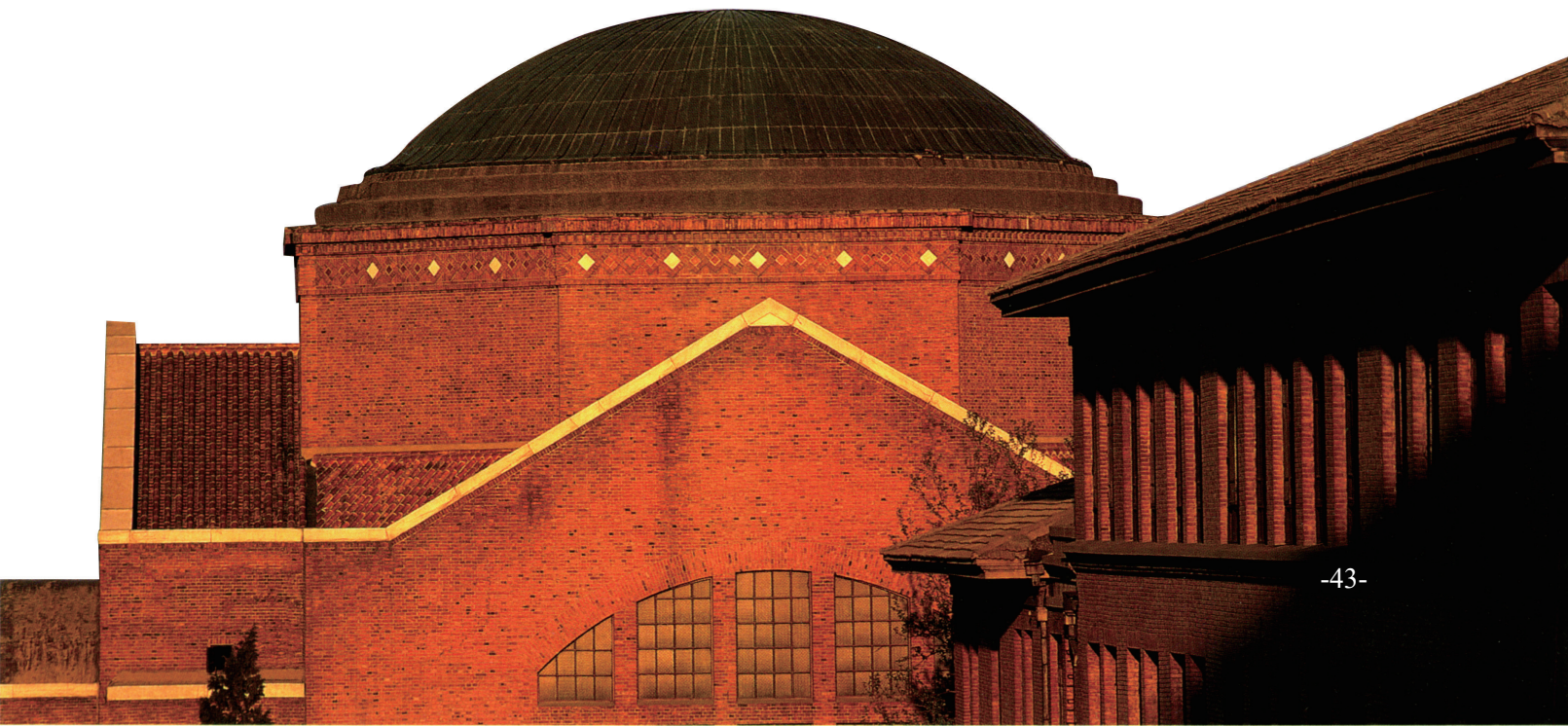
2. Early career as a statistician.

Q: How did you finally decide to become a statistics researcher?

A: Once I returned to Manchester, I became quite clear that statistics is the career I want to pursue. Thanks to my school, I had an opportunity to meet with other scientists and technologists, and became interested in control engineering and stochastic control. So, time series became quite a natural subject for me. My early career was mainly oriented in frequency domain, and I changed to time domain later on when I met Akaike. He visited us in Manchester for half a year, working on multivariate control system using multivariate linear AR model, as well as some aspect of AIC(Akaike Information Criterion). We became very good friends, and I wanted to learn more from him, so I applied a Royal Society Japan Fellowship, and went to Japan for 6 months. During that visit, I read a number of papers he collected, and learned a lot from not only the papers, but also the marks and personal notes he made. It was very valuable for me. By talking to him, I learned the background of why he did certain research. He did research not in front of the desk, but went out and met other scientists. He did not publish many papers in the first ten years of his career, but did a lot of great works later. He spent time cultivating friendship with engineers and other people. Because of this, he was asked to solve a problem of selecting a suitable model from a number of models in the field of predicting. That's the original problem behind AIC. So, I got a deep understanding of the whole idea of his research besides reading papers.

Q: We know that you published many great papers in your early career, so what's your secret for this fantastic achievements?

A: I remember the words of Mr. Yang Zhenning. He said do something that you are really passionate about. My father never interfered in my study, and I never interfered in my children's career either. Let the person choose what he or she is really interested in. My mentor is a time series analyst, but he never pushed me, so I had the chance to choose my own area. The reason why I choose statistics is because I want to produce something new, so I am lucky to be in the right environment where there is no pressure. I am also very lucky to have a good wife taking care of my family, and lucky to have the chance meeting with other scientists. I am a good learner, and I am able to pick up the things I want to learn. I think passion is very important rather than any secret. Remember to be observant and passionate.





3.About the threshold model

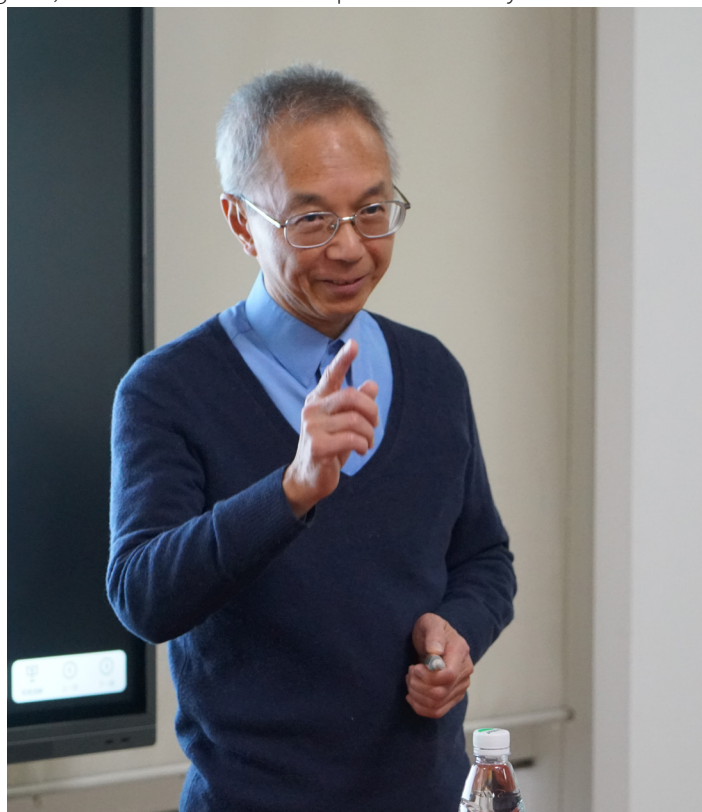
Q: Now let's talk about one of your most important work in non-linear time series, the threshold model. Where did the idea come from?

A: When I was visiting Akaike, I learned the way he produced the spectral density estimate. So, I used the approach on the lynx data, which I was very interested in. There was a session in the Royal Statistic Society and I presented this paper. During that discussion, there was one gentleman who made a very, very important comment. He said that the data is cyclical, but the cycle is not symmetric. The lynx population would rise slowly but fall rapidly. If you use a linear Gaussian model, you would never be able to capture it. Also, he said that from the point view of dynamical system, the cycle should be considered limit cycle. So, if you can produce a model that leads to limit cycle, it would be ideal. And David Cox and Akaike also made some similar comments. However, it is very difficult and is a big challenge. So, I decided to work on the problem. But my entire education up to that time was all in linear. So, I need to teach myself nonlinear dynamical systems.

Then, one day I was in my garden and mowing the lawn. When you mow the lawn, you go strip by strip. Suddenly, the idea of piecewise linearity came into my mind. This is because I was subconsciously thinking of the problem all the time.

Then I started working on the idea and a student did programs. One day she brought me some results which were too perfectly periodic to be possible. Then I found that she forgot the noise. This was the first time I saw limit cycle. Then, I said we could also see whether this model can produce other nonlinear phenomena, such as subharmonics, higher harmonics, amplitude-frequency dependency and so on. And it turned out that the model could do that.

Q: Did you encounter difficult times with the



汤家豪先生给清华学生授课



model?

A: Yes. A lot of people discussed the paper but I could not say everybody liked it, maybe because the idea was so new. I also got one or two people attacking. The model was invented in 1980s but has remained fairly quiet for 10 years. It was in about the 1990s that the model attracted a lot of attention. So, the beginning was not easy.

Q: From your experience, how to find a good research problem?

A: First of all, you have to be social. To me, statisticians are toolmakers. What tool you want to invent must be dictated by practical needs from people on the ground. So, we should go out, interact and collaborate with other scientists. We should be members of scientific teams. Don't follow fashion blindly. I never want to follow fashion. When I did nonlinear time series, almost none of the leaders in time series worked on that.

There are probably two types of research. One is the run-of-mill research, which means you have an incremental improvement. Those things do not take us long and you can publish these very quickly. The other one is the revolutionary research. Of course, in one's lifetime, one would probably not have more than a couple of such revolutions. But you must always keep them in mind, work on them in any spare time.

4.About the leadership

Q: You have been Chair of Statistics at several universities. How can you do good jobs in both academic and management? What's your secret?

A: I adopted the principle I learned from Lao Tzu (老子) and Sun Tzu's "Art of War" (孙子兵法). I cannot micromanage, so if there is any big job I will identify a suitable person. Then I will give the person my full support. So if you use one person you need to trust him (用人不疑, 疑人不用).

5.About statistics in the future

Q: Do you worry about the future of statistics given the competition from Machine Learning and AI?

A: As Lao Tzu has said, behind every good fortune there is a misfortune, and misfortune leads to good fortune (祸兮福之所倚, 福兮祸之所伏). I think the two aspects are certainly true for what challenge statistics is facing in the domain of data science. But if we sensibly steer our ship of statistics, we can benefit. Machine learning is certainly a powerful tool, but some of the ideas are not unknown or uncommon in statistics. Because in statistics, the basic training is how to handle randomness, and for anything that requires that, statistics has advantages. But on the other hand, we have to be fully prepared and liberate our minds. Some of the old ideas may be too restrictive. We used to deal with small data set in days of Fisher, but now we have to deal with large data sets. To defeat the new challenge, we have to adopt the attitude in Chinese culture: when foreigners come, we absorb them.

So, I don't worry. As long as we are broad-minded and ready to adapt, we can survive and grow.

编辑：邓 柯
执行编辑：侯禹珊

行勝
於言

清华大学统计学研究中心

地址：北京市海淀区清华大学伟清楼 212 室 (100084)

电话：010-62786091 传真：010-62783842

网址：www.stat.tsinghua.edu.cn

Center for Statistical Science of Tsinghua University

Address: Weiqing Building 212, Tsinghua University, Beijing 100084, China

Tel: +86 10-62786091 Fax: +86 10-62783842

[Http://www.stat.tsinghua.edu.cn](http://www.stat.tsinghua.edu.cn)



扫码关注微信公众号
水木数据派