



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY



统计与数据科学系
Department of Statistics and Data Science



理学院
COLLEGE OF SCIENCE



生物医学统计中心
Center for Biostatistics

2021 深圳 生物医学统计国际论坛

2021 SHENZHEN
INTERNATIONAL CONFERENCE ON BIostatISTICS



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY



统计与数据科学系
Department of Statistics and Data Science



理学院
COLLEGE OF SCIENCE



生物医学统计中心
Center for Biostatistics

会议手册 Conference Manual

南方科技大学统计与数据科学系、
南方科技大学理学院生物医学统计中心联合主办
2021年8月21日-8月22日

目录

Content

会议介绍	1
主要参会人员	2
组委会	3
论坛日程	4
会议摘要	6
论坛指南	22
主办方简介	24
会议通讯录	27
会议记录页	31



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY



统计与数据科学系
Department of Statistics and Data Science



理学院
COLLEGE OF SCIENCE



生物医学统计中心
Center for Biostatistics

附件一：《“2021深圳生物医学统计国际论坛”健康防疫安全参会通知》

会议介绍



南方科技大学（简称：南科大）是深圳在中国高等教育改革发展的宏观背景下创建的一所高起点、高定位的公办创新型大学，致力于服务创新型国家建设及深圳国际化现代化创新型城市建设。2012年教育部批准南科大建校，并赋予学校探索具有中国特色的现代化大学制度、探索创新人才培养模式的重大使命。

深圳作为我国改革开放政策和现代化建设先行先试的地区，位于粤港澳大湾区核心地带，是我国少数几个具有完善产业链集群的区域；是全球发展最快、中国经济最活跃的城市之一；是中国南方重要的高新技术发展重镇及创新基地；也是高素质人才和创新意识最为活跃的城市之一；是一座孕育梦想并成就梦想的先锋城市。

“2021年深圳生物医学统计国际论坛”于2021年8月21至22日在深圳南方科技大学举行，主办单位为南方科技大学统计与数据科学系及南方科技大学理学院生物医学统计中心。论坛将围绕生物医学统计相关主题设立主旨报告，特邀相关领域国内外著名专家学者进行学术报告。本次生物医学统计国际论坛在防控疫情常态化的条件下，将采取“线下主会场+线上分会场”的形式。

本次论坛旨在为来自国(境)内外的杰出专家与学者提供一个高层次、高水平的学术交流平台，深入探讨国际生物医学统计前沿领域的最新动态和发展趋势，进一步促进大数据背景下生物医学统计各领域的跨学科交流与合作。

主要参会人员

» 大会特邀报告嘉宾(Plenary talk)



Chris Holmes

Professor of Biostatistics, Department of Statistics, University of Oxford, Programme Director, Health and Medical Sciences Programme, The Alan Turing Institute, UK



Xihong Lin

Professor and former Chair of the Department of Biostatistics, Coordinating Director of the Program in Quantitative Genomics at the Harvard T. H. Chan School of Public Health, the 2006 Committee of Presidents of Statistical Societies (COPSS) Presidents' Award Laureate, USA

» 大会报告嘉宾 (Invited talk)

- | | |
|-----|------------------------------------|
| 方积乾 | 中山大学医学统计与流行病学系主任 |
| 赖颖蕾 | 乔治华盛顿大学教授，中国科学技术大学教授 |
| 吕 晖 | 上海交通大学生物信息学与生物统计学系主任，教授 |
| 欧春泉 | 南方医科大学公共卫生学院生物统计学系主任，教授 |
| 彭志行 | 南京医科大学生物统计学系教授 |
| 秦国友 | 复旦大学生物统计学教研室主任，教授 |
| 王学钦 | 中国科学技术大学教授 |
| 许王莉 | 中国人民大学应用统计科学研究中心生物统计与流行病学研究室主任，教授 |
| 薛付忠 | 山东大学公共卫生学院副院长，健康医疗大数据研究院院长 |
| 言方荣 | 中国药科大学生物统计教研室主任，生物统计与计算药学研究中心主任，教授 |
| 俞章盛 | 上海交通大学医学院临床研究中心主任，长聘教授 |
| 喻祖国 | 湘潭大学副校长，教授 |
| 赵 杨 | 南京医科大学生物统计学系主任，教授 |
| 周晓华 | 北京大学生物统计系主任，教授 |

» 南方科技大学特邀嘉宾

- | | |
|-----|--------------------------|
| 顾东风 | 南方科技大学代理副校长，讲席教授，中国科学院院士 |
| 杨学明 | 南方科技大学理学院院长，讲席教授，中国科学院院士 |

(按姓名首字母排序)

组委会

» 组委会主席

史建清 南科大统计与数据科学系教授、南方科技大学理学院生物医学统计中心主任
邵启满 南科大统计与数据科学系 系主任、讲席教授
田国梁 南科大统计与数据科学系 副系主任、教授

» 组委会成员

荆炳义 南科大统计与数据科学系 讲席教授（即将入职）
杨丽丽 南科大统计与数据科学系 教授
陈欣 南科大统计与数据科学系 副教授
胡延庆 南科大统计与数据科学系 副教授
蒋学军 南科大统计与数据科学系 副教授
李曾 南科大统计与数据科学系 副教授
焦熙云 南科大统计与数据科学系 助理教授
马一方 南科大统计与数据科学系 助理教授
徐匆 南科大统计与数据科学系 助理教授

» 组委会秘书

滕悦然 tengyr@mail.sustech.edu.cn
邱雅芳 qiuyf3@mail.sustech.edu.cn
宋安琪 songaq@mail.sustech.edu.cn
李兰 lil36@sustech.edu.cn
周琳 zhoul3@sustech.edu.cn

论坛日程

2021年8月21日（星期六）

线下会场：南科大会议中心-圆形会议室（4楼） 线上会场：腾讯会议ID 697 6434 2900（会议密码：0821）			
08:00-08:20	论坛报到		
08:20-09:00	开幕式-顾东风院士、杨学明院士、邵启满讲席教授致欢迎辞（主持人：史建清）		
时间	报告人	题目	主持人
09:00-09:50	Xihong Lin (Plenary talk)	Hypothesis Testing for a Large Number of Composite Nulls in Genome-wide Causal Mediation Analysis	荆炳义
09:50-10:10	茶歇		
10:10-10:45	周晓华	CSTE: An R Package for Estimating Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve	田国梁
10:45-11:20	方积乾	Estimating the Minimal Number of Repeated Examinations for Random Responsiveness -- a new topic of Biostatistics	
11:20-11:55	俞章盛	GapClust: a Light-Weighted Approach Distinguishing Rare Cells from Voluminous Single Cell Expression Profiles	
12:00-13:40	午餐		
13:45-14:20	赵杨	精准医疗与个体处理效应	李曾
14:20-14:55	许王莉	A Fast and Accurate Estimator for Large Scale Linear Model Via Data Averaging	
14:55-15:30	喻祖国	分形与统计方法在物种亲缘分析中的一些应用	
15:30-16:05	王学钦	Measuring and Predicting the Cognitive-impairment of Alzheimer's Disease Patients from Hippocampus Shape	
16:05-16:30	茶歇		

论坛日程

16:30-17:20	Chris Holmes (Plenary talk)	Machine Learning and Biostatistics: A Story of Robustness, Prediction and Inference	史建清
17:20-17:55	欧春泉	基于卫星监测数据估计我国近地面每日大气PM2.5浓度	
18:30-20:00	晚宴		

2021年8月22日 (星期日)

线下会场：南科大会议中心-圆形会议室（4楼）			
线上会场：腾讯会议ID 697 6434 2900（会议密码：0821）			
时间	报告人	题目	主持人
08:30-09:05	彭志行	不同阶段新冠疫情的模型研判分析与防控效果评价——从数据到决策	马一方
09:05-09:40	秦国友	Multiply Robust Estimation of Quantile Treatment Effects with Missing Responses with Application to CHARLS	
09:40-10:15	吕晖	生物医学数据全流程分析平台	
10:15-10:25	茶歇		
10:25-11:00	薛付忠	Causal Mediation Analysis using Summarized Genetic Data	焦熙云
11:00-11:35	言方荣	Shotgun: A Bayesian Seamless Phase I-II Design to Accelerate the Development of Targeted Therapies and Immunotherapy	
11:35-12:10	赖颖蕾	On the Efficient Genome-wide Concordance and Discordance Analysis of Multiple Large-scale Two-sample Expression Data Sets	
12:10-14:00	午餐		

会议摘要

Plenary talk

Machine Learning and Biostatistics: A Story of Robustness, Prediction and Inference



Chris Holmes

Department of Statistics and the Nuffield Department of Medicine Oxford University and Scientific Director for the Health Programme the Alan Turing Institute, UK

摘要:

We consider the potential for Machine Learning (ML) to assist in biostatistical studies. Traditional statistical methods tend to be explicit with an emphasis on uncertainty quantification and known robustness properties under model misspecification. On the other hand, ML methods tends to be more opaque with a focus on prediction performance. We consider some of the challenges and potential solutions in adapting ML for the biostatistician's toolkit.

Plenary talk

(按报告时间排序)

Hypothesis Testing for a Large Number of Composite Nulls in Genome-wide Causal Mediation Analysis



Xihong Lin (林希虹)

Department of Biostatistics and Department of Statistics
Harvard University

摘要:

In genome-wide epigenetic studies, it is often of scientific interest to assess whether the effect of an exposure on a clinical outcome is mediated through DNA methylation. Statistical inference for causal mediation effects is challenged by the fact that one needs to test a large number of composite null hypotheses across the genome. In this paper, we first study the theoretical properties of the commonly used methods for testing for causal mediation effects, Sobel's test and the joint significance test. We show the joint significance test is the likelihood ratio test for the composite null hypothesis of no mediation effect. Both Sobel's test and the joint significance test follow non-standard distributions, and they are overly conservative for testing mediation effects and yield invalid inference in genome-wide epigenetic studies. We propose a novel Divide-Aggregate Composite-null Test (DACT) for the composite null hypothesis of no mediation effect in genome-wide analysis. We show that the DACT method provides valid statistical inference and boosts power for testing mediation effects across the genome. We propose a correction procedure to improve the DACT method using Efron's empirical null method when the exposure-mediator or/and the mediator-outcome association signals are not sparse. Our extensive simulation studies show that the DACT method properly controls type I error rates and outperforms the Sobel's and the joint significance tests for genome-wide causal mediation analysis. We applied the DACT method to the Normative Aging Study to identify putative DNA methylation sites that mediate the effect of smoking on lung function. We also developed a computationally efficient R package DACT for public use.

CSTE: An R Package for Estimating Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve



周晓华 (ZHOU Xiaohua)

北京大学公共卫生学院生物统计系

摘要:

Personalized medicine has been attached much attention in modern biomedical research, which aims to tailor medical treatments based on individual's characteristics. A large number of statistical approaches have been developed to estimate an optimal individualized treatment rule (ITR) given the data collected from clinical trials or observational studies. These approaches can be classified into two general statistical methods and their hybrids. The first general approach intends to obtain the optimal ITR by a direct minimization of the population average outcome. The second general approach aims to model the heterogeneous treatment effects, using a covariate-specific treatment effect (CSTE) curve, and quantify variability associated with the estimated optimal individualized treatment rule. In this talk, I introduce the second approach and the CSTE R package for its implementation.

Estimating the Minimal Number of Repeated Examinations for Random Responsiveness -- a new topic of Biostatistics



方积乾 (FANG Jiqian)

中山大学医学统计与流行病学系

摘要:

Objective: The aim of this study was to develop a general method for estimation of the minimal number of repeated examinations needed to detect patients with random responsiveness, given a limited rate of missed diagnosis.

Methods: Basic statistical theory was applied to develop the method. As an application, 100 patients with disorders of consciousness (DOC) were assessed by the Coma Recovery Scale-Revised (CRS-R). DOC patients were supposed to be examined for 13 times over 20 days, while anyone who was diagnosed as a minimally conscious state (MCS) in a round would no longer be examined in the subsequent rounds. To test the validation of this method, a series of the stochastic simulation was completed by computer software under all the conditions of possible combinations of three kinds of distributions for p , five values of p , and four sizes of the sample and repeated for 100 times.

Results: A series of formula was developed to estimate the probability of a positive response to a single examination given by a patient and the minimal number of successive examinations needed based on the numbers of patients detected in the first i ($i = 1, 2, \dots$) rounds of repeated examinations. As applied to the DOC patients assessed by CRS-R, with a rate of missed diagnosis < 0.0001 , the estimate of the minimal number of examinations was six in traumatic brain injury patients while five in non-traumatic brain injury patients. The outcome of the simulation showed that this method performed well under various conditions possibly occurring in practice.

Interpretation: The method developed in this paper holds in theory and works well in application and stochastic simulation. It could be applied to any other kind of examinations for random responsiveness, not limited to CRS-R for detecting MCS; this should be validated by further researches though.

GapClust: a Light-Weighted Approach Distinguishing Rare Cells from Voluminous Single Cell Expression Profiles



俞章盛 (YU Zhangsheng)

上海交通大学医学院临床研究中心

摘要:

Single cell RNA sequencing (scRNA-seq) is a powerful tool in detailing the cellular landscape within complex tissues. Large-scale single cell transcriptomics provide both opportunities and challenges for identifying rare cells playing important roles in development and disease. We develop GapClust, a light-weighted algorithm to detect rare cell types from ultra-large scRNA-seq data set with unrivaled speed and memory efficiency. Benchmarking on diverse experiments demonstrates the superior performance of GapClust compared with existing approaches. When applied to intestine and 68 k PBMC data sets, GapClust recovered tuft cells and a novel subtype of monocyte, respectively.

精准医疗与个体处理效应



赵杨 (ZHAO Yang)

南京医科大学生物统计学系

摘要:

现代医学进入了个性化循证医学的时代，充分利用个体信息，预测个体处理效应，有助于医生针对病人特征采取精准治疗措施。本报告将首先介绍个体处理效应和随机森林的基本原理，讨论递归分割时混杂因素的控制和效应修饰因子的识别问题。

A Fast and Accurate Estimator for Large Scale Linear Model Via Data Averaging



许王莉 (XU Wangli)

中国人民大学应用统计科学研究中心
生物统计与流行病学研究室

摘要:

This work is concerned with the estimation problem of linear model when the sample size is extremely large and the data dimension can vary with the sample size. In this setting, the least square estimator based on full data is not feasible with limited computational resources. Many existing methods for this problem are based on sketching technique. We derive fine-grained lower bounds of the conditional mean squared error for sketching methods. For sampling methods, our lower bound provides an attainable optimal convergence rate. We propose a new sketching method based on data averaging. The proposed method reduces the original data to a few averaged observations. These averaged observations still satisfy the linear model and are used to estimate the regression coefficients. The asymptotic behavior of the proposed estimation procedure is studied. Our theoretical results show that the proposed method can achieve a faster convergence rate than the optimal convergence rate for sampling methods. Theoretical and numerical results show that the proposed estimator has good statistical performance as well as low computational cost.

分形与统计方法在物种亲缘分析中的一些应用



喻祖国 (YU Zuguo)

湘潭大学

摘要:

随着科学技术日新月异的发展和人类基因组计划的实施，人们越来越关心生命科学，越来越多的科学研究者进入生命科学或与生命科学相关的研究领域。处理海量生物数据的生物信息学是当前生命科学中非常重要而且活跃的分支之一，它也是生命科学、计算机科学、数学、统计学等多学科交叉的前沿研究领域。非线性科学是上世纪八十年代兴起的一门交叉学科，为刻画自然界复杂系统提供了许多有效的工具与方法。而分形又是非线性科学方法中最重要、应用最广泛的方法之一。分形和统计方法（如马尔科夫模型）已被广泛应用到DNA与蛋白质序列分析的研究中。在此报告中，将介绍分形及相关统计方法在生物信息学中物种亲缘分析问题中的一些应用。

Measuring and Predicting the Cognitive-impairment of Alzheimer's Disease Patients from Hippocampus Shape



王学钦 (WANG Xueqin)

中国科学技术大学管理学院统计与金融系

摘要:

We perform a metric kernel regression analysis on 373 Alzheimer's Disease (AD) patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. In trials, those patients suffer from cognitive impairment, measured by the Alzheimer's Disease Assessment Scale Cognitive Behavior Section (ADAS-Cog), which could have low sensitivity to measure AD progression in clinical trials. Our study goal is to scrutinize the subregions of the hippocampus and clinical factors that predict cognitive impairment levels. Nevertheless, it is rather challenging for modeling the brain hippocampus shape data, a typical non-Euclidean data as predictors. To overcome this critical issue, we develop a simple kernel regression estimation approach with the metric kernel, which can also grasp the features of the non-Euclidean structure. Based on our analysis, the left hippocampus's head and body regions with p values 0.0009 and < 0.0001 , and the body of the right hippocampus with p-value < 0.0001 are associated with the ADAS-Cog scale. The kernel regression method also shows that education length is statistically significantly associated with the ADAS-Cog scale at a significant level of 0.05. Therefore, with 3 subregions of the hippocampus and education length, we provide a reliable prediction model for cognitive impairment in AD patients. This model may facilitate personalized therapy-decision making for AD patients.

基于卫星监测数据估计我国近地面每日大气PM2.5浓度



欧春泉 (OU Chunquan)

南方医科大学公共卫生学院生物统计学系

摘要:

背景：详尽准确的大气污染数据是污染控制和健康效应评估的基础。然而，空气质量地面监测站的建设和维护相当昂贵，导致目前站点的分布非常稀疏且极为不均，很多小城市和农村甚至没有地面监测站，无法对污染状况进行全面准确地监测，更无法估计个体暴露情况。既往有研究采用卫星观测获取的气溶胶厚度（AOD）等数据估计地面颗粒物浓度，但模型预测效果不理想，且存在严重的缺失值问题。方法：采用地统计方法和随机森林等方法对AOD进行填补，基于填补后的AOD以及气象数据，采用三次样条基函数模拟它们和近地面PM2.5浓度的非线性暴露-滞后-反应关系，并结合XGBoost机器学习方法捕捉复杂的时空效应，以每1km*1km的分辨率对2014-2015年全国每日近地面PM2.5浓度进行估计，并用实际监测数据对模型进行交叉验证。结果：AOD观测值缺失率从84.54%下降到1.36%，无论从时间还是空间角度都取得对PM2.5较为理想的估计效果（交叉验证的R2达到0.88）。结论：该研究充分展示了卫星监测AOD数据的价值，为大气污染监测提供了新思路、新方法，也为后续大气颗粒物健康效应评估提供了基础数据支持。

不同阶段新冠疫情的模型研判分析与防控效果评价 ——从数据到决策



彭志行 (PENG Zhihang)

南京医科大学生物统计学系

摘要:

新冠肺炎的理论研究取得了很大进展，人们了解了病毒的流行潜力，包括平均潜伏期、重症患者占比等，并确定了新冠病毒传播的风险指标、疫情峰值、到达峰值的时间和最终感染规模等；评估了包括封城、密切跟踪隔离、检测和检出等重大防控策略的有效性和时效性，为制定科学防控措施奠定了坚实基础。疫情防控措施已经由单纯的非药物干预措施向“疫苗接种加非药物干预”的组合措施转变。然而，新冠疫苗接种与群体免疫有什么关系？新冠疫苗保护期与疫情流行规模有什么关系？新冠疫苗接种覆盖率与防控策略有什么关系？如何利用随机离散动力学模型刻画在不同国家、不同地区和人群中，不同强度的干预措施在疫情中的作用，并定量界定实施这些措施的目标人群和使用时机？如何研究在疫苗使用后，病毒变异、疫苗覆盖率变化，人们行为方式变化对疫苗作用影响的关键因素？如何探索疫情应对期间，各类干预措施对不同流行病学特征的呼吸道传染病传播是否具有短期和长期影响以及其差异性？如何评价在不同季节、不同传播能力、我国不同地区存在防控能力差异下，疫情预警及干预措施对疾病暴发的作用？这些问题的回答，都需要对全面的数据进行仔细挖掘与分析，结合疫情研判分析与防控效果评价，可为我国应对具有潜在大流行风险的呼吸道传染病防控策略和预警提供依据。

Multiply Robust Estimation of Quantile Treatment Effects with Missing Responses with Application to CHARLS



秦国友 (QIN Guoyou)

复旦大学公共卫生学院生物统计学教研室

摘要:

Causal inference and missing data have attracted significant research interests in recent years, while the current literature usually focus on only one of these two issues. Moreover, compared with the commonly used average treatment effect (ATE), the quantile treatment effect (QTE) is able to provide a complete picture of the difference between the treatment and control groups, as well as robustness to the outliers in the responses. Therefore, we develop a method to estimate the QTE in the context of missing data based on the idea of inverse probability weighting (IPW).

The proposed IPW estimator has the property of multiply robustness, that is, as long as the class of candidate models of propensity scores contains the correct model and so does the candidate models for the probability of being observed, the resulting QTE estimator is root- n consistent and asymptotic normal. Simulation studies are conducted to investigate the performance of the proposed method, and a real data from CHARLS is analyzed and different treatment effects are observed at various quantile levels of the response.

生物医学数据全流程分析平台



吕晖 (LV Hui)

上海交通大学生物信息学与生物统计学系

摘要:

生物医学数据分析是保障精准医学实施的重要支撑,但生物医学数据中组学数据来源多样,分析目的庞杂,数据格式繁多,分析软件达2.4万余个, workflow呈爆炸式组合,对研究结果的可靠性以及和重复性提出了重大挑战。为解决这一难题,保障临床和科研结果的客观性和科学性,构建了涵盖基因组、转录组、蛋白组、甲基化数据的生物医学大数据标准分析平台BMAP,以及数据分析评价体系,为不同研究中组学数据分析可靠性、一致性提供了重要保障。同时开发相关多组学算法,为相关领域的项目实施提供了重要支撑平台。

Causal Mediation Analysis using Summarized Genetic Data



薛付忠 (XUE Fuzhong)

山东大学公共卫生学院生物统计学系

摘要:

Background: Causal mediation analysis aims to investigate the mechanism linking an exposure and an outcome. Dealing with the impact of unobserved confounders among the exposure, mediator and outcome has always been an issue of great concern. Moreover, when multiple mediators exist, this causal pathway intertwines with other causal pathways, making it more difficult to estimate of path-specific effects (PSEs).

Methods: In this article, under the specific rank condition that the number of instrumental variables is larger than the number of mediators, we propose a method (PSE-MR) to identify and estimate PSEs of an exposure on an outcome through multiple causally ordered and non-ordered mediators using Mendelian randomization (MR), when there are unmeasured confounders among the exposure, mediators and outcome. Additionally, PSEs can also be estimated by PSEMR even when the assumptions of MR and traditional mediation analysis are violated.

Results: We conducted simulations to evaluate the finite sample performances of our proposed estimators in different scenarios. The results show that the causal estimates of PSEs are almost unbiased with good coverage and Type I error properties. We illustrate the utility of our method through a study of exploring the mediation effects of body mass index and pack years of smoking in the causal pathways from education to osteoarthritis risk.

Shotgun: A Bayesian Seamless Phase I-II Design to Accelerate the Development of Targeted Therapies and Immunotherapy



言方荣 (YAN Fangrong)

中国药科大学理学院生物统计教研室、
生物统计与计算药学研究中心

摘要:

Drug development of novel antitumor agents is conventionally divided by phase and cancer indication. With the advent of new molecularly targeted therapies and immunotherapies, this approach has become inefficient and dysfunctional. We propose a Bayesian seamless phase I-II “shotgun” design to evaluate the safety and antitumor efficacy of a new drug in multiple cancer indications simultaneously. “Shotgun” is used to describe the design feature that the trial begins with an all-comer dose finding phase to identify the maximum tolerated dose (MTD) or recommended phase II dose (RP2D), and then is seamlessly split to multiple indication-specific cohort expansions. Patients treated during dose finding are rolled over to the cohort expansion for more efficient evaluation of efficacy, while patients enrolled in cohort expansion contribute to the continuous learning of the safety and tolerability of the new drug. During cohort expansion, interim analyses are performed to discontinue ineffective or unsafe expansion cohorts early. To improve the efficiency of such interim analyses, we propose a clustered Bayesian hierarchical model (CBHM) to adaptively borrow information across indications. A simulation study shows that compared to conventional approaches and the standard Bayesian hierarchical model, the shotgun design has substantially higher probabilities to discover indications that are responsive to the treatment in question, and is associated with a reasonable false discovery rate. The shotgun provides a phase I-II trial design for accelerating drug development and to build a more robust foundation for subsequent phase III trials. The proposed CBHM methodology also provides an efficient design for basket trials.

On the Efficient Genome-wide Concordance and Discordance Analysis of Multiple Large-scale Two-sample Expression Data Sets



赖颖蕾 (LAI Yinglei)

乔治华盛顿大学、中国科学技术大学

摘要:

With the current microarray and RNA sequencing technologies, two-sample genome-wide expression data have been increasingly collected in biological and medical studies. Differential expression analysis and gene set enrichment analysis have been frequently conducted. Integrative analysis can be conducted when multiple data sets are available. In practice, concordant and discordant molecular behaviors among a series of data sets can be of biological and clinical interest. There is still a lack of statistical methods for these types of integrative analysis.

We have proposed a mixture model-based approach to the integrative analysis of multiple large-scale two-sample expression data sets. Since the mixture model is based on the transformed differential expression test P-values (z-scores), it is generally applicable to the expression data generated by either microarray or RNA sequencing platforms. The mixture model is simple with three normal distribution components for each data set to represent down-regulation, up-regulation and non-differential expression. However, when the number of data sets increases, the model parameter space increases exponentially due to the component combination from different data sets. To achieve a concordant and discordant integrative analysis for a series of data sets, we have introduced two model reduction strategies.

We demonstrate our methods on the recent TCGA RNA sequencing data. To illustrate a concordant integrative analysis, we apply our method to a series of data sets collected for studying two closely related types of cancer. To illustrate a discordant integrative analysis, we apply our method to a series of data sets collected for studying different types of cancer. Interesting disease-related pathways can be detected by our integrative analysis approach.

论坛指南

» 论坛安全防疫须知

健康、安全的举行大型集体活动，是当前新冠肺炎疫情防控常态下开展各类活动的前提和根本保障。鉴于近期国内新冠肺炎疫情发展形势变化，多地发现本土病例，升级为中高风险地区，为了保护您和他人健康安全，并推动此次会议安全顺利的开展，根据国家、广东省、深圳市、南科大疫情防控工作领导小组对疫情防控工作的最新要求，我们制定了《2021深圳生物医学统计国际论坛健康防疫安全参会通知》（详见附件），如转发国务院客户端公布的中高风险地区名单、现场参会的注意事项、入校申请流程等。希望您做好前期进入南科大校园现场参会的各项准备，以保障顺利进入校园。同时，在会议举办期间，希望您全程佩戴口罩，配合和支持会议疫情防控小组工作人员和志愿者的健康安全提示和指引。

“疫情防控，人人有责”。希望你我的守护，能推动2021深圳生物医学统计国际论坛健康、安全、圆满的举办！

会议疫情防控小组联系人：

宋老师：18620300662； 组委会邮箱：biostat@sustech.edu.cn

* 《2021深圳生物医学统计国际论坛健康防疫安全参会通知》详见附件一。

» 论坛报到

报到时间	报到地点
8月20日（星期五）14:00-17:30	专家公寓 注：仅适用于提前通过会务组登记住宿的嘉宾
8月21日（星期六）08:00-08:20	南科大会议中心 - 圆形会议室 (4楼)

注：校外线下参会者需提前联系会务组安全员（宋老师，18620300662）提交防疫信息申请入校。

» 会议安排

8月21日（星期六）：

- 开幕式及全体参会人员合影；
- 报告会场：南科大会议中心 - 圆形会议室 (4楼)；
- 欢迎晚宴：会务组统一安排。

8月22日（星期日）：

- 报告会场：南科大会议中心 - 圆形会议室 (4楼)。

论坛用餐安排

用餐由会务组统一安排，包括会议期间的午餐，晚餐和茶歇。详情请参阅以下安排：

日期	用餐	时间	用餐地点	用餐人员
8月21日	午餐	12:00-13:40	会务组统一安排	所有参会人员
8月22日	午餐	12:10-14:00		
8月21日	晚餐	18:30-20:00	会务组统一安排	

会场及交通

学校地址：广东省深圳市南山区学苑大道1088号

• **地铁站：**5号线（环中线）塘朗站

-塘朗地铁站C出口出站，步行约350米到南方科技大学正门（1号门）。

• **公交车站：**南方科技大学站（1号门，3号门）

-线路：43, 74, 81, M459。

会场地址：南科大会议中心-圆形会议厅



主办方简介

南方科技大学：

南方科技大学（简称：南科大）是深圳在中国高等教育改革发展的时代背景下创建的一所高起点、高定位的公办新型研究型大学。学校借鉴世界一流理工科大学的学科设置和办学模式，以理、工、医为主，兼具商科和特色人文社科的学科体系，在本科、硕士、博士层次办学，在一系列新的学科方向上开展研究，使学校成为引领社会发展的思想库和新知识、新技术的源泉。

南方科技大学将扎根中国大地，紧抓粤港澳大湾区、深圳先行示范区“双区”驱动，深圳经济特区、深圳先行示范区“双区”叠加的历史机遇，发扬“敢闯敢试、求真务实、改革创新、追求卓越”的创校精神，突出“创知、创新、创业”（Research, Innovation and Entrepreneurship）的办学特色，努力服务创新型国家建设及深圳国际化现代化创新型城市建设，快速建设成为聚集一流师资、培养拔尖创新人才、创造国际一流学术成果并推动科技应用的国际化高水平研究型大学，为尽早实现建成世界一流研究型大学的宏伟目标打下坚实基础。



主办方简介

南方科技大学统计与数据科学系：



南方科技大学统计与数据科学系成立于2019年4月，以建设国际一流的教育培养和研究基地为目的。本系志在为国家培养出具有扎实的科学基础，思想活跃，创新意识和能力强，有国际视野，脚踏实地，有朝气、有理想的拔尖人才。本系已经建立起本硕博人才培养体系，拥有统计学和数据科学与大数据技术两个本科专业，及数学学科下概率论和数理统计硕博学位授予权，主要研究领域包括数理统计、生物医学统计、金融统计和数据科学。

到目前为止，本系共有14位专任教师（其中3位即将入职）和4位双聘教师，其中有讲席教授4人，教授3人，副教授5人，助理教授6人。统计系拥有国际化、高水平的师资队伍，包括1名世界数学家大会邀请报告人，2名国家自然科学基金二等奖获得者，1名长江讲座教授，2名国际数理统计学会（IMS）会士、1名常务理事，1名英国皇家统计学会会士，1名IMS Medallion讲座演讲者，1名深圳市杰出人才培养对象，1名深圳市高层次国家级领军人才。

统计与数据科学系官网：<https://stat-ds.sustech.edu.cn/>



系官网二维码



系公众号二维码

南方科技大学理学院生物医学统计中心：

生物医学统计中心于2020年7月正式获批，设立于南方科技大学理学院，主办单位为南方科技大学理学院统计与数据科学系。目前中心共有21位骨干成员，其中有讲席教授3人，教授5人，副教授5人，副研究员5人，助理研究员2人，主治医师1人。现有成员单位包括南方科技大学统计与数据科学系、医学院、生物系、化学系，深圳儿童医院等。中心将统计理论和方法论的研究结合我校其他学科，以及医院、工业界和政府决策等，推进跨学科综合应用研究。

» 中心主要研究方向：

- * 复杂性数据分析，包括精准医疗和公众健康研究等
- * 适应性临床试验和数据分析
- * 统计组学数据分析，包括生物学、高效计算技术

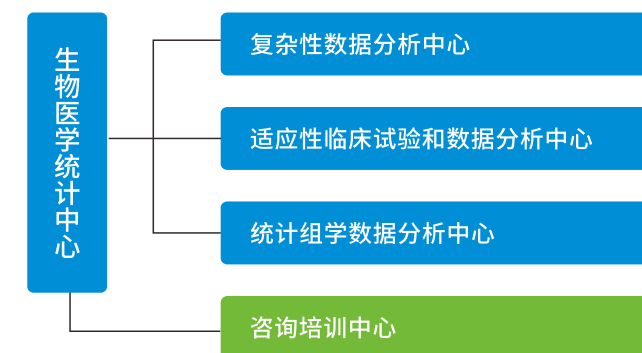
» 中心总体目标：

- * 发展创新理论和技术，探索对生命健康数据的分析和解释；
- * 交叉学科，攻克生命科学中的实际问题；
- * 复杂数据和大数据的创新研究，探索根本，创造价值；

» 中心定位：

用三至五年时间，初步建成国内领先并具有国际影响力，集基础研究，多学科交叉融合，咨询，专业人才培养以及产业转化的综合研究基地。

» 中心架构：



生物医学统计中心官网：<https://cbs.sustech.edu.cn/>

会议通讯录

(按姓名字母排序)

Chris Holmes	牛津大学	cholmes@stats.ox.ac.uk
Xihong Lin	哈佛大学	xlin@hsph.harvard.edu
Dandan Tang	Utrecht University Department of statistics and methodology	d.tang@uu.nl
毕祎琳	河南大学数学与统计学院	1326156644@qq.com
曾智	湖南省妇幼保健院	444239982@qq.com
查建栋	香港大学深圳医院肿瘤中心	827082263@qq.com
陈邦维	华南理工大学生物系	chenweibang@genomics.cn
陈晓平	福建师范大学数学与统计学院	xpchen@fjnu.edu.cn
陈欣	南方科技大学统计与数据科学系	chenx8@sustech.edu.cn
陈萱	首都医科大学	15901591528@163.com
陈亚菲	南方科技大学医学院	Caslinsky@163.com
陈玉媛	UEA business management	767660203@qq.com
翟一卜	南方科技大学	11812724@mail.sustech.edu.cn
董秋丽	上海师范大学数理学院	2804130949@qq.com
董晓歌	南方科技大学	11612133@mail.sustech.edu.cn
段明瑞	首都医科大学流行病学与卫生统计学系	dmr99056@163.com
范晓东	吉林化工学院数学系	fanxd14@mails.jlu.edu.cn
方积乾	中山大学医学统计与流行病学系	fangjq@mail.sysu.edu.cn
高洁	江南大学理学院	gaojie@jiangnan.edu.cn
高天栎	南方科技大学统计与数据科学系	11611529@mail.sustech.edu.cn
葛淑菲	上海科技大学数学科学研究所	geshf@shanghaitech.edu.cn
管诺	南方科技大学统计与数据科学系	nuoguan789@163.com
郭二林	徐州工程学院数学与统计学院	gel@xzit.edu.cn
郭小波	中山大学数学学院统计科学系	guoxb3@mail.sysu.edu.cn
郭阳	北京大学深圳医院	13120083056@163.com
过榴晓	江南大学理学院信计系	guoliuxiao@jiangnan.edu.cn
何娟	中国药科大学生物统计与计算药学研究中心	13505259611@163.com
侯锐	首都医科大学流行病学与卫生统计学	ruihou1@163.com
胡延庆	南方科技大学统计与数据科学系	huyq@sustech.edu.cn
黄嘉琳	中山大学管理学院	karron.w@qq.com
黄珊珊	煌途医药(无锡)有限公司	1511915065@qq.com
黄晓彦	北京大学深圳医院临床研究院	huangxiaoyan@pku.org.cn
纪宏超	南方科技大学化学系	ji.hongchao@foxmail.com
简学宁	首都医科大学流行病学与卫生统计学系	jian_0628@163.com
蒋如意	南方科技大学	11611407@mail.sustech.edu.cn
蒋学军	南方科技大学统计与数据科学系	jiangxj@sustech.edu.cn
焦熙云	南方科技大学统计与数据科学系	jiaoxy@sustech.edu.cn
金成	南方科技大学医学院	king-cheng@qq.com
荆炳义	南方科技大学统计与数据科学系	majing@ust.hk

赖颖蕾	乔治华盛顿大学、中国科学技术大学	ylai@gwu.edu
李曾	南方科技大学统计与数据科学系	liz9@sustech.edu.cn
李翠霞	徐州工程学院数学与统计学院	lcx@xzit.edu.cn
李海彬	首都医科大学附属北京朝阳医院心脏中心	15910681644@126.com
李嘉妍	首都医科大学流行病学与卫生统计学	lijiayan0507@163.com
李静	复旦大学公共卫生学院生物统计教研室	18211020063@fudan.edu.cn
李俊	香港大学深圳医院肝胆胰外科	lijun@hku-szh.org
李宁	首都医科大学 流行病学与卫生统计学	LN2939319100@163.com
李桃	深圳华大生命科学研究院精准健康研究所	litao2@genomics.cn
李小春	首都医科大学流行病学与卫生统计学系	xiaochunleo@163.com
李秀君	山东大学公共卫生学院生物统计学系	xjli1816@163.com
李亚鹏	复旦大学公共卫生学院	18211020006@fudan.edu.cn
李元	南方科技大学海洋系	liy2020@mail.sustech.edu.cn
李兆媛	香港中文大学(深圳)数据科学学院	lizhaoyuan@cuhk.edu.cn
梁宝生	北京大学公共卫生学院生物统计系	liangbs@hsc.pku.edu.cn
林李泽强	湖南师范大学统计系	linlizeqiang@qq.com
林子昭	南方科技大学生物系	11812534@mail.sustech.edu.cn
刘鹏渊	浙江大学转化医学研究院	pyliu@zju.edu.cn
刘薇	香港大学统计与精算系	liuweiv@hku.hk
刘显慧	江西财经大学统计学院	liuxh109@126.com
刘小源	北方工业大学理学院统计系	Liuxiaoyuan0611@163.com
刘晓玉	新加坡国立大学	xyliu0208@gmail.com
刘艺璇	南方科技大学统计与数据科学系	12032925@mail.sustech.edu.cn
卢悦祺	深圳华大生命科学研究院	luyueqi@genomics.cn
罗敬	中南民族大学数学与统计学学院	my007lj@163.com
吕晖	上海交通大学生物信息学与生物统计学系	huilu@sjtu.edu.cn
吕闻凭	香港中文大学深圳	lyuwade@gmail.com
马伶玉	香港大学深圳医院 临床肿瘤中心	838465865@qq.com
马一方	南方科技大学统计与数据科学系	mayf@sustech.edu.cn
孟晓妮	首都医科大学研究生	mengxiaoni385@163.com
孟钰	深圳大学总医院消化内科	297458077@qq.com
牟荣吉	上海交通大学医学院临床研究中心	rjmu@sjtu.edu.cn
欧春泉	南方医科大学公共卫生学院生物统计学系	ocq@fimmu.com
潘小青	上海师范大学数学系	xpan@shnu.edu.cn
彭春方	中山大学数学学院统计学系	pengchf@mail2.sysu.edu.cn
彭志行	南京医科大学生物统计学系	zhihangpeng@163.com
秦国友	复旦大学公共卫生学院	gyqin@fudan.edu.cn
邱建青	四川大学华西生物医学大数据中心	qiujianqing0219@163.com
邱世芳	重庆理工大学统计系	sfqiu@cqut.edu.cn

邱怡轩	上海财经大学统计与管理学院	qiuyixuan@sufe.edu.cn
闪莹	华大生命科学研究院	sylvia.shanboo@gmail.com
邵启满	南方科技大学统计与数据科学系	shaoqm@sustech.edu.cn
邵秀丽	四川大学华西医院	18980961263@163.com
施建华	闽南师范大学数学与统计学院	v0085@126.com
石建栋	香港浸会大学数学系	18481701@life.hkbu.edu.hk
史建清	南方科技大学统计与数据科学系	shijq@sustech.edu.cn
宋聪颖	北京大学深圳医院临床研究院	congying_song@126.com
宋爽	清华大学	songs15@tsinghua.org.cn
苏乐柔	中山大学数学学院	2902223448@qq.com
苏丽文	中国药科大学生物统计与计算药理学系	cpusullivan@163.com
孙健	复旦大学大数据学院	19110980003@fudan.edu.cn
孙思远	南方科技大学化学系	11930100@mail.sustech.edu.cn
孙韬	中国人民大学统计学院	sun.tao@ruc.edu.cn
孙文欣	首都医科大学公共卫生学院流行病学与卫生统计学系	1390559586@qq.com
孙杨	北京大学数学科学学院	sunyang@stu.pku.edu.cn
孙宇辉	深圳华大生命科学研究院智能硬件研究所	sunyuhui@genomics.cn
汤丹丹	荷兰Utrecht university methodology and statistics department	15626014993@163.com
汤敏丹	北京大学香港科技大学医学中心	2011210680@bjmu.edu.cn
唐鑫寅	厦门大学管理学院数据挖掘中心	tangxinyin1998@163.com
田国梁	南方科技大学统计与数据科学系	tiangl@sustech.edu.cn
田培馨	香港大学精算与统计学系	pxtian@connect.hku.hk
汪时嘉	南开大学统计与数据科学学院	shijia_wang@nankai.edu.cn
王纯杰	长春工业大学数学与统计学院	wangchunjie@ccut.edu.cn
王合玲	新疆财经大学统计与数据科学学院	601056505@qq.com
王玮玮	厦门大学健康医疗数据国家研究院	wangww0603@163.com
王学钦	中国科学技术大学管理学院统计与金融系	wangxq20@ustc.edu.cn
王梓	南方科技大学医学院	11930118@mail.sustech.edu.cn
翁福添	厦门大学医疗健康大数据研究院, 厦门大学医学院, 厦门大学数据挖掘研究中心	ftweng@163.com / xmftweng@163.com
吴密霞	北京工业大学统计与数据科学院	wumixia@bjit.edu.cn
吴淇	厦门大学数据挖掘研究中心	wuqi199509@163.com
吴小龙	厦门大学医学院、健康医疗大数据国家研究院	dragonwu@foxmail.com
向仕兵	南方科技大学统计与数据科学系	chibing.xiang@gmail.com
肖鸿民	西北师范大学数学与统计学院	xiaohm@nwnu.edu.cn
徐匆	南方科技大学统计与数据科学系	xuc6@sustech.edu.cn
徐思琦	香港大学	sqxu@hku.hk
许佩蓉	上海交通大学数学科学学院	prxu@shnu.edu.cn
许王莉	中国人民大学应用统计科学研究中心生物统计与流行病学研究室	xwlbnu@163.com
薛付忠	山东大学公共卫生学院生物统计学系	xuefzh@sdu.edu.cn

言方荣	中国药科大学理学院生物统计教研室, 生物统计与计算药理学研究中心	f.r.yan@163.com
杨丰凯	山东大学(威海)数学与统计学院	yfk@sdu.edu.cn
杨钧涵	首都医科大学	1718034@ccmu.edu.cn
杨丽丽	南方科技大学统计与数据科学系	yangll@sustech.edu.cn
杨晓洁	云南财经大学统计与数学学院	yangxiaojie@amss.ac.cn
尧旻昊	香港大学统计与精算系	yaominhao@pku.edu.cn
易晓巍	四川大学华西医院	zitong94yi@163.com
尹昌辉	南方科技大学医学院	12032579@mail.sustech.edu.cn
俞易安	南方科技大学统计与数据科学系	12031036@mail.sustech.edu.cn
俞章盛	上海交通大学医学院临床研究中心	yuzhangsheng@sjtu.edu.cn
喻祖国	湘潭大学	yuzg@xtu.edu.cn
袁平	北京大学深圳医院	yuanping_632@163.com
张弛	深圳大学经济学院统计系	chizhang@szu.edu.cn
张海丽	南方科技大学统计与数据科学系	zhanghl@sustech.edu.cn
张辉国	新疆大学数学与系统科学学院统计系	xjstat@qq.com
张润松	南方科技大学统计与数据科学系	11811223@mail.sustech.edu.cn
张小柳	深圳大学生物医学工程学院	xiaoliu_z@163.com
张晓宇	北京工业大学统计与数据科学系	zhangxiaoyu006@126.com
张溢潆	南方科技大学	12031339@mail.sustech.edu.cn
张英哲	哈佛公共卫生学院	yzhang1@hsph.harvard.edu
张园园	河南大学数学与统计学院	1348572962@qq.com
张子怡	南方科技大学统计与数据科学系	11712621@mail.sustech.edu.cn
章萌	首都医科大学公共卫生学院	sherryzhang1999@163.com
赵杨	南京医科大学生物统计学系	zhaoyang@njmu.edu.cn
赵振平	中国疾控中心慢病中心	zhaozhenping@ncncd.chinacdc.cn
郑德强	首都医科大学流行病学与卫生统计学系	dqzheng@ccmu.edu.cn
郑曼琪	首都医科大学流行病学与卫生统计学系	manqizh@foxmail.com
郑榕	中山大学数学学院	zhengr9@mail2.sysu.edu.cn
钟俊江	厦门理工学院应用数学学院	jjzhong@xmut.edu.cn
钟庭艳	上海交通大学	tyzhong@sjtu.edu.cn
周晓华	北京大学公共卫生学院生物统计系	azhou@bicmr.pku.edu.cn
周扬	北京大学统计科学中心	zhouyang@math.pku.edu.cn
朱建平	厦门大学管理学院, 健康医疗大数据国家研究院, 数据挖掘研究中心	xmjpzhu@xmu.edu.cn
朱映璇	首都医科大学公共卫生学院	zhu17801117166@163.com

会议记录页

附件一

“2021深圳生物医学统计国际论坛” 健康防疫安全参会通知

尊敬的参会者：

您好！

“2021深圳生物医学统计国际论坛”将于2021年8月21日—8月22日在广东省深圳市南方科技大学举办。论坛以“现场参会和线上参会”两种形式进行。我们诚挚的欢迎您参加此次论坛！

健康、安全的举行大型集体活动，是当前新冠肺炎疫情防控常态下开展各类活动的前提和根本保障。鉴于近期国内新冠肺炎疫情发展形势变化，多地发现本土病例，升级为中高风险地区。根据国家、广东省、深圳市对疫情防控工作的要求，南方科技大学（以下简称“南科大”）结合实际，发布了南科大疫情防控〔2021〕15号文，对安全进入南科大校园开展活动和工作做了明确规定。

为了保护您和他人的健康安全，并推动此次论坛安全顺利的开展，我们向您具体说明进入南科大校园、现场参会的注意事项和进校须知流程：

一、会前准备

（一）明确现场参会人员

现场参会人员必须全部来自深圳，报到前21天内没有中高风险地区、港台地区、国外旅居史，14天大数据行程中没有中高风险地区所在县市，持有绿色粤康码，已接种疫苗，持有入校前48小时内核酸证明（即8月19日参加核酸检测）。

（二）明确线上参会人员

满足现场参会人员条件外的其余人员均为线上参会人员。

（三）提前完成核酸检测

现场参会人员需要8月19日完成核酸检测。核酸检测阴性证明、健康绿码、大数据行程码，请您在8月20日10点前发送给宋老师（具体形式后续微信群通知）。

二、进校流程

具体形式后续微信群通知。

三、会议现场注意事项

现场参会人员必须全程佩戴医用外科口罩，隔座就坐。