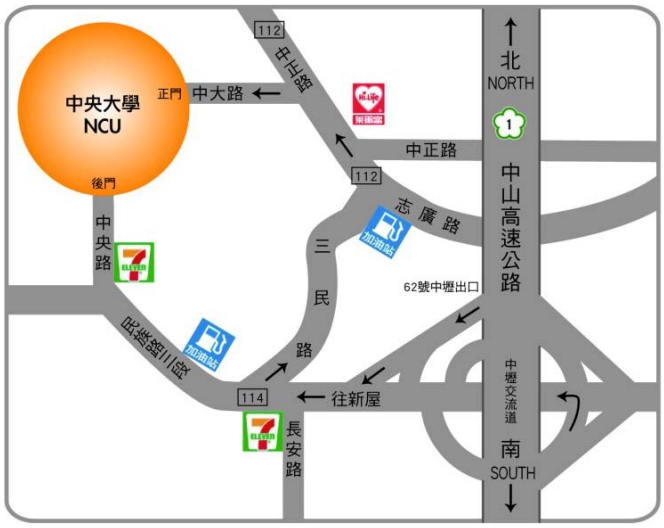
目　　錄

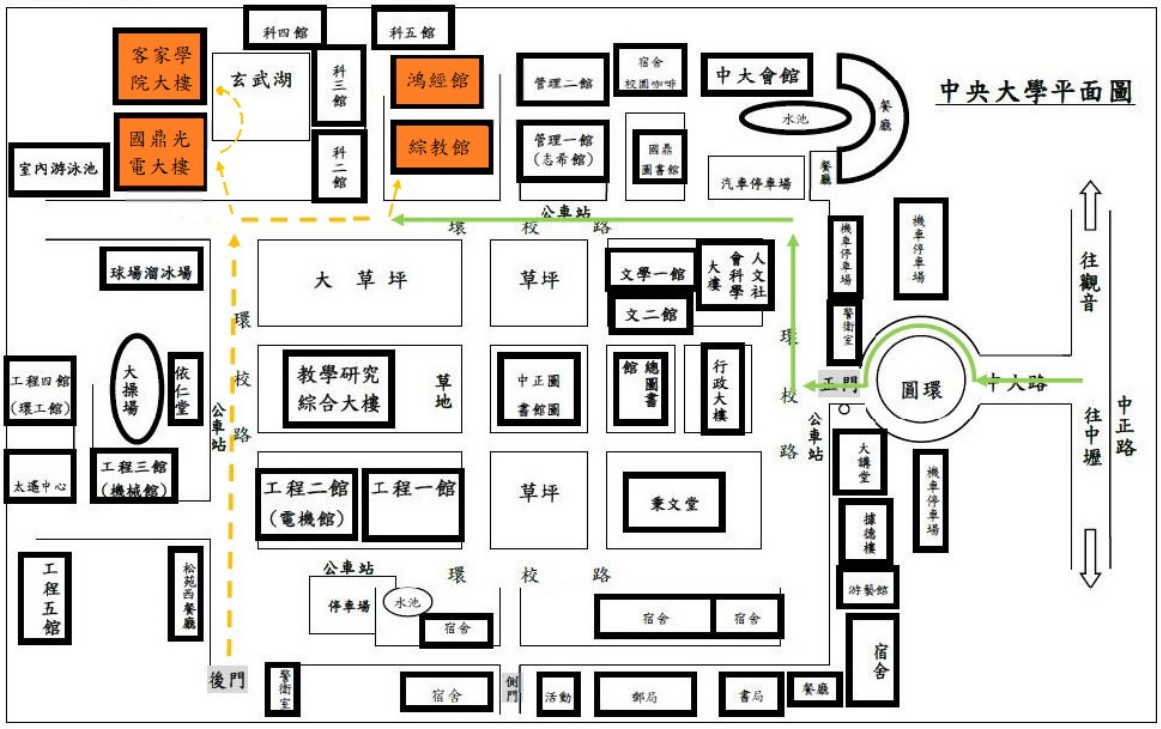
107 年統計學術研討會暨中央大學統計所 40 週年國際學術研討會

交通資訊

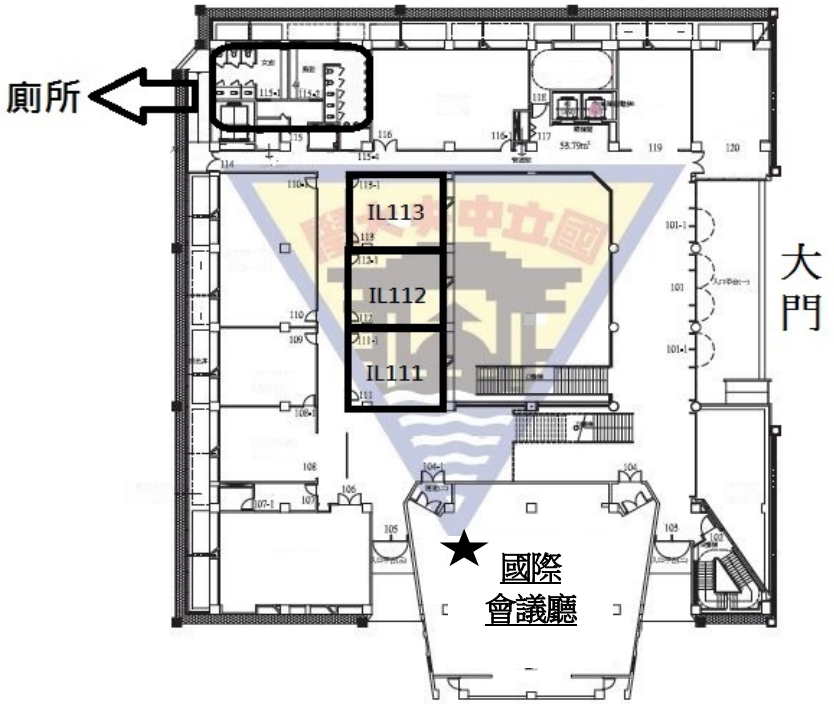
- 國道 1 號（中山高速公路）
  中壢交流道（62 公里）出口，往新屋方向行駛，沿民族路至三民路右轉，中正路左轉，中大路左轉即可抵達本校前門。車程約 5~10 分鐘。
- 國道 3 號(福爾摩沙高速公路）
  大溪交流道（62 公里）出口，往中壢方向行駛，轉台 66 線快速公路(往中壢、觀音方向），接國道 1 號（北上），於 62 公里中壢交流道出口，往新屋方向行駛，沿民族路至三民路右轉，中正路左轉，中大路左轉即可抵達本校前門。車程約 20 分鐘。
- GPS 衛星導航
  北緯 24.96828；東經 121.195474



校園平面圖

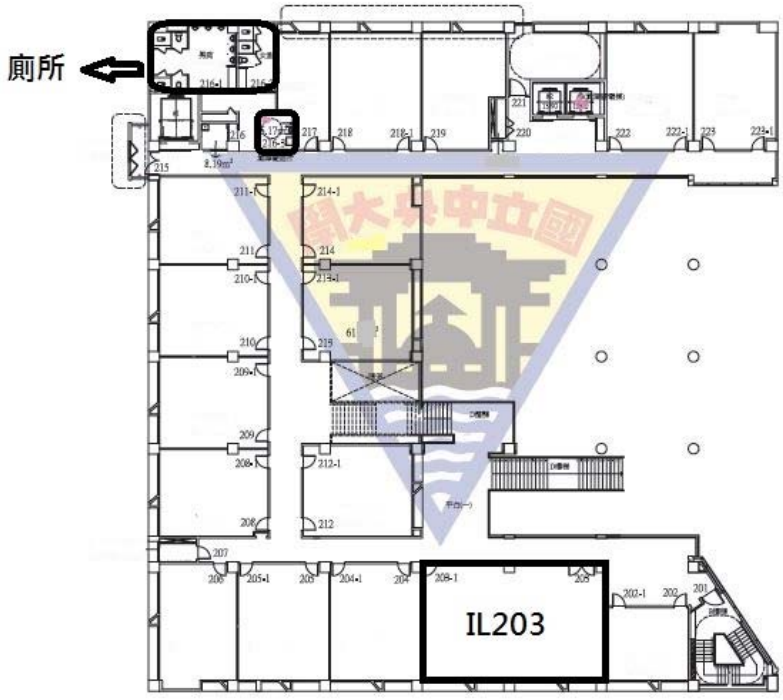


國立中央大學國鼎光電大樓一樓平面圖



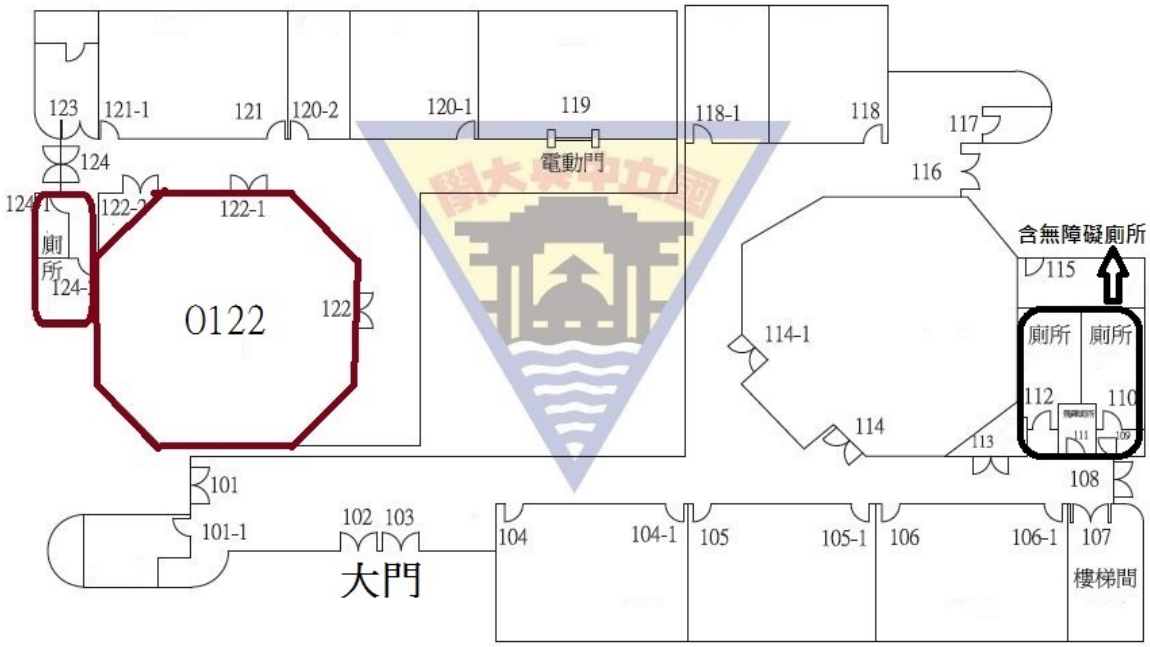
一樓

國立中央大學國鼎光電大樓二樓平面圖



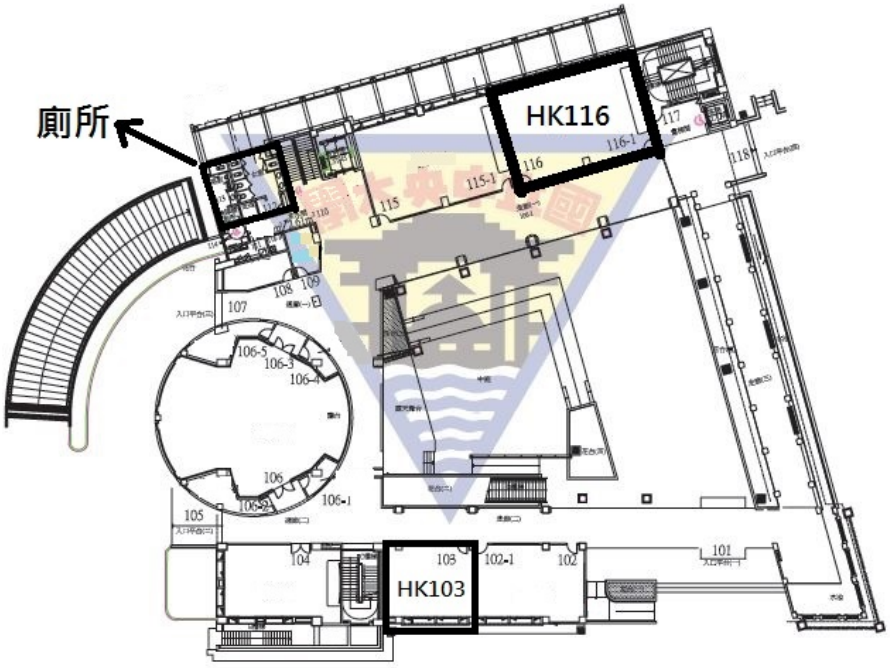
二樓

國立中央大學綜教館一樓平面圖



一 樓

國立中央大學客家學院平面圖



一樓

107 年統計學術研討會暨中央大學統計所 40 週年國際學術研討會

時間	議程表 I　2018 年 11 月 9 日（星期五）		
08：30－09：30	報到（地點：光電大樓國際會議廳）		
09：30－10：20	開幕典禮		
10：20－10：50	茶敘（光電大樓）		
10：50－12：00	Chair：Jiunn Tzon Hwang（黃俊宗） Keynote Speaker：Dr. James O. Berger（Conference Hall）		
12：00－13：00	午餐時間（IL111、IL112）		
13：00－14：00	Chair：Cheng-Der Fuh（傅承德） Keynote Speaker：Dr. Samuel Kou（Conference Hall）		
14：00－14：30	茶敘（光電大樓）		
14：30－15：50	學術研討會		
	A－1（IL111）	A－2（IL112）	A－3（IL113）
	Data Analytics in Statistical Genomics Chair：楊欣洲 1. 蕭朱杏 2. 謝文萍 3. 黃美菊	Spatial Statistics Chair：黃郁芬 1. 陳春樹 2. 林培生 3. 黃信誠	Statistical Inference Chair：許順吉 1. 程爾觀 2. 盧鴻興 3. 黃逸輝
	A－4（IL203）	A－5（HK103）	A－6（HK116）
	統計論文獎 Chair：陳君厚所長 1. 陳宜均 2. 楊郁成	Industry Statistics I Chair：吳碩傑 1. 李俊毅 2. 俞一唐 3. 蔡志群	Medical Statistics Chair：江村剛志 1. 程毅豪 2. 張　中 3. 江村剛志
15：50－16：10	中場休息		
16：10－17：30	學術研討會		
	B－1（IL111）	B－2（IL112）	B－3（IL113）
	International session I Chair：Henry Horng-Shing Lu 1. John Bailer 2. Peter Guttorp 3. Richard Barker	International session II Chair：Shih-Feng Huang 1. Mark Podolskij 2. Juergen Symanzik 3. Irena Križman	Experimental Design I Chair：蔡風順 1. 羅夢娜 2. 潘建興 3. 林長鋆
	B－4（IL203）	B－5（HK103）	B－6（HK116）
	Biostatistics Chair：黃怡婷 1. 黃怡婷 2. 黃佳慧 3. 張馨文	Financial Engineering I Chair：高竹嵐 1. 鄭宏文 2. 羅盛豐 3. 翁新傑	Survival Analysis Chair：溫啟仲 1. 黃彥棕 2. 黃名鉞 3. 蘇佩芳
18：00－	晚宴（新陶芳）		

時間	議程表Ⅱ　　2018 年 11 月 10 日（星期六）		
08：30－09：00	報到（光電大樓國際會議廳）		
09：00－10：20	學術研討會		
	C－1（IL111）	C－2（IL112）	C－3（IL113）
	Applied Probability I Chair：陳冠宇 1. 陳隆奇 2. 黃建豪 3. 劉聚仁	Financial Engineering II Chair：林良靖 1. 黃士峰 2. 蔡秉真 3. 繆維中	Industry Statistics II Chair：彭健育 1. 李宜真 2. 王義富 3. 彭健育
	C－4（IL203）	C－5（HK103）	C－6（HK116）
	High-dimensional Data Analysis Chair：呂恒輝 1. 顏佐榕 2. 江其衽 3. 洪　弘	Contributed Session Chair：黃世豪 1. 羅　琪 2. 章坤鎔 3. 趙維雄	Student Session I Chair：張明中 1. 林晏禎 2. 陳韋宏 3. 趙元唯 4. 藍丹璟 5. 許竣瑝 6. 施嘉翰
10：20－10：50	茶敘（光電大樓）		
10：50－12：10	學術研討會		
	D－1（IL111）	D－2（IL112）	D－3（IL113）
	中大統研所 40 週年紀念 Chair：陳春樹 1. 洪文良 2. 薛慧敏 3. 陳錦華 4. 張玉媚	中大統研所 40 週年紀念 Chair：王義富 1. 蔡政安 2. 李仁佑 3. 李昀寰	Experimental Design II Chair：張明中 1. 蔡欣甫 2. 許湘伶 3. 黃世豪
	D－4（IL203）	D－5（HK103）	D－6（HK116）
	Applied Probability II Chair：須上苑 1. 洪芷漪 2. 蕭守仁 3. 陳定立	Student Session II Chair：施嘉翰 1. 許雅娟 2. 翁琬婷 3. 林妤軒 4. 王尹辰 5. 黃昕蔚 6. 林庭羽	中大統研所 40 週年紀念 Chair：許根寧 1. 林文明 2. 沈仲維 3. 許根寧
12：10－	中央大學統計所成立 40 週年餐會鴻經館 107 室		

107 年政府統計學術研討會

時間	議程表 I　　2018 年 11 月 9 日（星期五）
09：00－09：30	報到（地點：光電大樓國際會議廳）
09：30－10：20	開幕典禮
10：20－10：50	茶敘（光電大樓）
11：00－12：00	政府統計 I（綜教館 122 教室） 主持人：楊貴顯 1. 許志銘 2. 李書惠 3. 林淑敏
12：00－13：00	午餐時間
13：00－14：00	政府統計 II（綜教館 122 教室） 主持人：林麗貞 1. 何宗欣 2. 陳韻潔 3. 黃敏慈
14：00－14：30	茶敘（綜教館）
14：30－15：50	政府統計 III（綜教館 122 教室） 主持人：饒志堅 1. 賴威宇 2. 黃騰皜 3. 謝之婕 4. 紀柏宇
15：50－16：10	中場休息
16：10－17：30	政府統計 IV（綜教館 122 教室） 主持人：劉瑞文 1. 郭昌儒 2. 洪裕鑫 3. 范宜鴻 4. 歐長潤
18：00－	晚宴（新陶芳）

學術研討會 I

107 年 11 月 09 日（星期五）

[A-1] Data Analytics in Statistical Genomics　教室：IL111

Chair：楊欣洲（中央研究院統計科學研究所）

14:30～15:50

- 蕭朱杏（國立台灣大學）
  Current Approaches and Challenges in Gene-set and Pathway Analysis
  當代基因集合與路徑分析之研究與挑戰
- 謝文萍（清華大學統計研究所）
  Ranking the Evolutionary Order of Recurrent Somatic Mutations
- 黃美菊（中央研究院統計科學研究所）
  An integrated analysis tool for big genomic data from a biobank

[A-2] Spatial Statistics　教室：IL112

Chair：黃郁芬（國立成功大學）

14:30～15:50

- 陳春樹（國立彰化師範大學統計資訊研究所）
  On fixed effects estimation for spatial regression with spatial confounding
- 林培生（國衛院群健所）
  Cluster-Temporal Models for Disease Surveillance with an Application to Dengue Fever Infection in Taiwan
- 黃信誠（中央研究院統計所）
  symptotic Properties of the Maximum Likelihood Estimators in Misspecified Linear Mixed-Effects Models with Fixed Numbers of Clusters

[A-3] Statistical Inference　教室：IL113

Chair：許順吉（中央大學統計所）

14:30～15:50

- 程爾觀（研究員）
  Constructing Generalized Linear Models by Data Information
- 盧鴻興（國立交通大學統計學研究所）
  Online Learning for Multiclass Classification with Applications
- 黃逸輝（淡江大學數學系）
  On the model checking for the single-index model

`

學術研討會 I

107 年 11 月 09 日（星期五）

[A-4] 統計論文獎　　教室：IL203

Chair：陳君厚（中央研究院統計科學研究所）

14:30～15:50

- 陳宜均（成功大學統計學系）
  An Embedding Learning-based Approach to Predict Influencers for Location Promotion in Social Networks（基於嵌入學習之社群網路地點行銷影響力預測）
- 楊郁成（逢甲大學統計學系）
  具羅吉斯函數鏈結共變量之混合式 t 線性混合效應模型

[A-5] Industry Statistics I　　教室：HK103

Chair：吳碩傑（淡江大學統計學系）

14:30～15:50

- 李俊毅（國立成功大學統計系）
  A New Demerit Control Chart for Monitoring the Quality of Multivariate Poisson Processes
- 俞一唐（東海大學統計系）
  Non-homogeneous cumulative link models for ordinal data
- 蔡志群（淡江大學數學學系）
  Solar panel lamination with extreme value regression model

[A-6] Medical Statistics　　教室：HK116

Chair：江村剛志（國立中央大學・統計研究所）

14:30～15:50

- 程毅豪（中央研究院統計科學研究所）
  Assessing wage dynamics and stagnation using quantile dynamics
- 張中（中山大學應數系）
  Estimating the Optimal Number and Location of Cut Points
- 江村剛志（國立中央大學・統計研究所）
  Univariate feature selection and compound covariate for predicting survival

學術研討會 I

107 年 11 月 09 日（星期五）

[B-1] International session Ⅰ　　教室：IL111

Chair：Henry Horng-Shing Lu 盧鴻興（國立交通大學統計學研究所）

16:10～17:30

- John Bailer (University of Miami)
  Going from the classroom to the studio: the birth and infancy of Stats + Stories
- Peter Guttorp (University of Washington)
  Decision making under uncertainty: the case of sea level rise adaptation
- Richard Barker (Department of Mathematics and Statistics,Pro-Vice-Chancellor, Sciences, University of Otago, New Zealand)
  Continuous Time Capture-Recapture

[B-2] International session Ⅱ　　教室：IL112

Chair：Shih-Feng Huang 黃士峰（國立高雄大學統計研究所）

16:10～17:30

- Mark Podolskij (Aarhus University)
  High frequency data in statistics
- Juergen Symanzik (Utah State University)
  Eye-tracking in practice: the EyeTrackR R package and its use in a study on human postures
- Irena Križman (International Association for O cial Statistics)
  Partnership between official statistics and academia

[B-3] Experimental Design Ⅰ　　教室：IL113

Chair：蔡風順（中央研究院統計科學研究所）

16:10～17:30

- 羅夢娜（國立中山大學應用數學系）
  Optimal designs for comparison of response surfaces
- 潘建興（中央研究院）
  A Construction of Cost-Efficient Designs with Guaranteed Repeated Measurements on Interaction Effects
- 林長鋆（中興大學）
  Stochastic search variable selection for definitive screening designs in split-plot and block structures

學術研討會 I

107 年 11 月 09 日（星期五）

[B-4] Biostatistics　教室：IL203

Chair：黃怡婷（國立臺北大學統計學系）

16:10～17:30

- 黃怡婷（國立臺北大學統計學系）
  Misspecification of a dependent variable in the joint model of the logistic model and random effect models
- 黃佳慧（臺北大學統計學系）
  Estimating treatment effects for semicompeting risks data with treatment switching
- 張馨文（中央研究院統計科學研究所）
  Nonparametric testing for multiple survival functions with non-inferiority margins

[B-5] Financial Engineering I　教室：HK103

Chair：高竹嵐（交通大學統計學研究所）

16:10～17:30

- 鄭宏文（東吳大學財務工程與精算數學系）
  Variance Components Risk Premium: Evidence from Option Valuation and Stock Returns
- 羅盛豐（中原財金系）
  On Corrected Diffusion Approximation with Two-flat Boundaries
- 翁新傑（中研院經濟所）
  Robust test of stock return predictability under heavy-tailed innovations

[B-6] Survival Analysis　教室：HK116

Chair：溫啟仲（淡江大學數學系）

16:10～17:30

- 黃彥棕（中央研究院統計科學研究所）
  A nonparametric approach to semi-competing risks via causal mediation modeling
- 黃名鈺（中央研究院統計科學研究所）
  Effective Dimension Reduction in Multivariate Baseline Proportional Hazards Model
- 蘇佩芳（成功大學統計學系）
  Response-adaptive treatment allocation for survival trials with clustered right-censored data

政府統計學術研討會

107 年 11 月 09 日（星期五）

政府統計Ⅰ	綜教館 122 教室
	Chair：楊貴顯（教育部統計處）
11:00～12:00	● 許志銘（教育部統計處） 解析式教育統計動態圖之思與用 ● 李書惠（教育部統計處） 運用 Google Analytics 解析教育統計網站之服務效能 ● 林淑敏（交通部統計處） 交通 e 級診療室

政府統計Ⅱ	綜教館 122 教室
	Chair：林麗貞（經濟部統計處）
13:00～14:00	● 何宗欣（經濟部統計處） 美中貿易摩擦對我國產業之影響 ● 陳韻潔（財政部關務署統計室） 從海關統計看國際經貿時事二三事 ● 黃敏慈（財政部關務署統計室） 運用大數據把關海關統計資料品質

政府統計Ⅲ	綜教館 122 教室
	Chair：饒志堅（內政部統計處）
14:30～15:50	● 賴威宇（內政部統計處） 以基地台信令觀察人口移動特性及常住行為 ● 黃騰皜（內政部統計處） 如何促進民眾使用統計資料及擴大應用範圍 ● 謝之婕（臺北市政府主計處） 淺談臺北市房價負擔能力指標與婚育之關聯 ● 紀柏宇（高雄市政府主計處） 高雄市人口遷徙及未來發展

政府統計學術研討會

107 年 11 月 09 日（星期五）

政府統計Ⅳ	綜教館 122 教室
	Chair：劉瑞文（交通部統計處）
16:10～17:30	● 郭昌儒（交通部統計處） 從電子票證數據觀察運具轉乘 ● 洪裕鑫（交通部公路總局） 證據基礎決策—以高齡駕駛人駕照管理制度為例 ● 范宜鴻（內政部警政署統計室） 以大數據探討道路交通事故肇事原因 ● 歐長潤（桃園市政府主計處） 運用巨量租賃資料—探討桃園 YouBike 站點及需求配置

學術研討會 Ⅱ

107 年 11 月 10 日（星期六）

[C-1] Applied Probability Ⅰ　教室：IL111

Chair：陳冠宇（國立交通大學應用數學系）

09:00～10:20

- 陳隆奇（國立政治大學應用數學系）
  Critical two-point function for long-range self-avoiding walks with power-law couplings
- 黃建豪（台灣大學數學系）
  The localized phase transition of a polymer
- 劉聚仁（國立成功大學 數學系）
  Waiting Time Analysis for a Restricted-Length M/G/1 Queue with Multiple Vacations

[C-2] Financial Engineering Ⅱ　教室：IL112

Chair：林良靖（國立成功大學）

09:00～10:20

- 黃士峰（國立高雄大學統計研究所）
  A Multivariate Compound Poisson Model with Copula and its Application on Limit Order Book
- 蔡秉真（南台科技大學財金系）
  How Stochastic are the Innovations to a Comprehensive Volatility Model? A Point Process Analysis on High-Frequency Data
- 繆維中（台灣科技大學財金所）
  Risk Structure in an Extended Version of Merton's Jump-Diffusion Model Where Jump Magnitudes Follow an Autoregressive Process of Order 2

[C-3] Industry Statistics Ⅱ　教室：IL113

Chair：彭健育（中央研究院 統計科學研究所）

09:00～10:20

- 李宜真（成大統計系）
  Planning of Accelerated Degradation Tests
- 王義富（中正大學數學系）
  Degradation Analysis on Trend Gamma Process
- 彭健育（中央研究院 統計科學研究所）
  Optimal Doubling Burn-in Policy Based on Tweedie Processes with Applications to Degradation Data

學術研討會 II

107 年 11 月 10 日（星期六）

[C-4] High-dimensional Data Analysis　教室：IL203

Chair：呂恒輝（東海大學 統計系）

09:00～10:20

- 顏佐榕（中央研究院統計科學研究所）
  An Attention Algorithm for Solving Large Scale Structured $\boldsymbol{l_0}$-norm Penalized Estimation Problems
- 江其衽（中央研究院）
  Inverse Regression for Multivariate Functional Data
- 洪弘（台灣大學流行病學與預防醫學研究所）
  Sufficient dimension reduction via random-partition for large-p-small-n problem

[C-5] Contributed Session　教室：HK103

Chair：黃世豪（國立中央大學數學系）

09:00～10:20

- 羅琪（中華大學餐旅管理學系）
  以生活型態區隔顧客前往酒吧的消費行為
- 章珅鎔（中央研究院）
  檢定羅倫茲曲線之優勢（A Likelihood Ratio Test for Lorenz Dominance）
- 趙維雄（東華大學應用數學系）
  Goodness-of-fit Statistics for Multinomial Logistic Regression Models

[C-6] Student Session I　教室：HK116

Chair：張明中（國立中央大學統計研究所）

09:00～10:20

- 林晏禎（銘傳大學-應用統計與資料科學學系）
  應用文字探勘技術探討財金新聞對股價影響之研究-以半導體產業為例
- 陳韋宏（高雄大學統計學研究所）
  指數隨機圖模型的吉氏取樣改良
- 趙元唯（高雄科技大學）
  Quantile autoregression model in real exchange rate reversion of Chinese Yuan
- 藍丹璟（國立高雄科技大學財務管理系碩士班）
  台灣上市公司外匯暴險受多因子影響之探討
- 許竣瑝（國立中央大學統計研究所）
  Performance of a two-sample test with Mann-Whitney statistics
- 施嘉翰（國立中央大學統計研究所）
  Pretest and shrinkage estimation of the mean under a univariate normal distribution

學術研討會 Ⅱ

107 年 11 月 10 日（星期六）

[D-1] 中大統研所 40 週年紀念　　教室：IL111

Chair：陳春樹（國立彰化師範大學統計資訊研究所）

10:50～12:10

- 洪文良（國立清華大學）
  Parametrizing the Kepler Exoplanet Period-Radius Distribution with the Bivariate Normal Inverse Gaussian Distribution
- 薛慧敏（政治大學統計系）
  Hi-C Data Normalization
- 陳錦華（臺北醫學大學）
  The Normality Test in Meta-Analysis of Binary Outcome
- 張玉媚（東海大學統計系）
  Bayesian Adaptive Randomization Based on Time-to-Event Outcomes of Efficacy and Toxicity

[D-2] 中大統研所 40 週年紀念　　教室：IL112

Chair：王義富（中正大學數學系）

10:50～12:10

- 蔡政安（國立台灣大學農藝學系）
  Integrative Gene Set Analysis and Visualization in Genome-wide Association Study
- 李仁佑（逢甲大學統計學系）
  Testing no Treatment Effect in Meta-Analysis
- 李昀寰（銘傳大學財務金融學系）
  風險偏好、VIX 期貨基差與 S&P 500 期貨報酬

[D-3] Experimental Design Ⅱ　　教室：IL113

Chair：張明中（國立中央大學統計研究所）

10:50～12:10

- 蔡欣甫（國立台灣大學農藝學系）
  Detection of Location and Dispersion Effects from Partially Replicated Two-Level Factorial Designs
- 許湘伶（高雄大學統計學研究所）
  Simultaneous selection of models and designs for optimal forecasting in possibly misspecified models
- 黃世豪（國立中央大學數學系）
  Optimal designs for binary response models with multiple nonnegative variables

學術研討會 Ⅱ

107 年 11 月 10 日（星期六）

[D-4] Applied Probability Ⅱ　教室：IL203

10:50～12:10

Chair：須上苑（國立中央大學數學系）

- 洪芷漪（國立政治大學應用數學系）
  Some Limit Distributions of Discounted Branching Random Walks
- 蕭守仁（彰化師大數學系）
  Gaussian moments conjecture and Jacobian conjecture
- 陳定立（中研院統計所）
  Evaluate the rate of convergence of Markov chain Monte Carlo by expected commute time

[D-5] Student Session Ⅱ　HK103

10:50～12:10

Chair：施嘉翰（國立中央大學統計研究所）

- 許雅娟（銘傳大學應用統計與資料科學學系）
  不動產在土壤液化區之研究— 以臺北市及高雄市為例
- 翁琬婷（銘傳大學應用統計與資料科學學系）
  2008 年全球金融危機對臺灣類股影響之研究 A Study about 2008 Global Financial Crisis Influence on Taiwan Stocks
- 林妤軒（國立高雄科技大學財務管理系碩士班）
  台灣產業價量關係隨因子變化之探討－分量迴歸分析
- 王尹辰（中央統計研究所）
  A new model for dependent competing risks data in reliability
- 黃昕蔚（國立中央大學統計研究所）
  Review on joint frailty copula models for recurrent event times
- 林庭羽（國立中央大學統計研究所）
  Robust ridge regression: applications to the NIKKEI NEEDS data

[D-6] 中大統研所 40 週年紀念　HK116

10:50～12:10

Chair：許根寧 PAREXEL International（百瑞精鼎）

- 林文明（昱冠資訊股份有限公司）
  深度學習/機器學習/統計建模在智慧工廠的應用
- 沈仲維（中正大學數學系）
  Model selection for semiparametric marginal mean regression accounting for within-cluster subsampling variability and informative cluster size.
- 許根寧 PAREXEL International（百瑞精鼎）
  Testing Two Primary Endpoints for a Confirmatory Clinical Study

James O. Berger

- Arts and Sciences Professor of Statistics, Duke University;
- 美國國家科學院院士
- 統計 COPSS (Committee of Presidents of Statistical Societies) Presidents' Award 得主 (1984)

Samuel Kou

- Professor ,Department of StatisticsHarvard University
- 統計 COPSS (Committee of Presidents of Statistical Societies) Presidents' Award 得主 (2012)

Gaussian Process Emulation of Computer Models with Massive Output

Prof. James O. Berger
(Arts and Sciences Professor of Statistics,Duke University; 美國國家科學院院士)

Abstract

Often computer models yield massive output; e.g., a weather model will yield the predicted temperature over a huge grid of points in space and time. Emulation of a computer model is the process of finding an approximation to the computer model that is much faster to run than the computer model itself (which can often take hours or days for a single run). Many successful emulation approaches are statistical in nature, but these have only rarely attempted to deal with massive computer model output; some approaches that have been tried include utilization of multivariate emulators, modeling of the output (e.g., through some basis representation, including PCA), and construction of parallel emulators at each grid point, with the methodology typically based on use of Gaussian processes to construct the approximations. These approaches will be reviewed, with the startling computational simplicity with which the last approach can be implemented being highlighted and its remarkable success being illustrated and explained; in particular the surprising fact that one can ignore spatial structure in the massive output is explained. All results will be illustrated with a computer model of volcanic pyroclastic flow, the goal being the prediction of hazard probabilities near active volcanoes.

Statistical Analysis of Single-Molecule Protein-Targeting Experiments via Hierarchical Hidden Markov Models

Samuel Kou
Department of Statistics
Harvard University

Abstract

Recent technological advances allow scientists to follow a biological process on a single-molecule basis. These advances also raise many interesting data-analysis problems. In this talk we will focus on recent single-molecule experiments on protein targeting. To maintain proper cellular function, proteins often need to be transported inside or out of a cell. The detailed molecular mechanism behind such a process (often referred to as protein targeting) remains unclear. Single-molecule experiments are designed to unveil the detailed mechanism and reveal the functions of different organelles involved in the process. The experimental data consist of hundreds of stochastic time traces (from the fluorescence recording of the experimental system). We introduce a Bayesian hierarchical model on top of a hidden Markov model (HMM) to analyze these data and use the statistical analysis results to answer the biological questions. We will discuss model selection, the construction of the hierarchical HMM, their biological meaning as well as our new understanding of the detailed molecular mechanism of protein transportation.

Current Approaches and Challenges in Gene-set and Pathway Analysis
當代基因集合與路徑分析之研究與挑戰

蕭朱杏、張宏卿
國立台灣大學公共衛生學院流行病學與預防醫學研究所

Abstract

The identification of a biological process associated with a complex disease is often carried out through the gene-set or pathway analysis. This analysis was originally developed to summarize the effect of a group of genes, and has had many applications in genomic studies. The first type of such analyses utilizes the marginal information from each marker via the corresponding test statistic, p-value, or effect size. Later, the focus shifts to the construction of relationship among genes within a pathway. This is because the genes in the set participate in the same biological function, and therefore some interplay and relationship among them are expected to exist. This leads to several development of topology-based pathway analysis; most of them characterize the topology with network properties. This type of analysis, however, does not receive wide recognition as the first type of approaches. Several issues, common to both types or specific to one of them, remain to be resolved. In this research, we discuss the advantages of these analyses, outline the challenges ahead of us, and suggest solutions which may be useful for future research.

Keywords: correlation、*gene-set analysis*、*network*、 *pathway analysis*

[A-1]

Ranking the Evolutionary Order of Recurrent Somatic Mutations

Wen-Ping Hsieh
Institute of Statistics, National Tsing Hua University

Abstract

Tumors are well-known for its massive accumulation of mutations along its development. Computational biologists are especially interested in reconstructing the evolutionary history of the somatic mutations in tumors. The issue differs from the phylogenetic/coalescence tree construction of human population because the data is a mixture of cells and there is no clear trace of the coexistence of variants in the same cells. We propose a procedure to decompose a tumor into well-separated subclones according to the allele frequencies of different sets of mutations. The mutations representing the emergence of the same subclone are arranged in an evolutionary tree with the neighbor-joining method according to the subclone proportions in each tumor. The root of the tree is selected by the orders derived from the RankNet across multiple samples. The strategy helps to identify the recurrent mutations across multiple tumors and to trace the driver mutations that potentially propagate the downstream variations.

[A-1]

An integrated analysis tool for big genomic data from a biobank

Mei-Chu Huang, Chia-Wei Chen, and Hsin-Chou Yang
Institute of Statistical Science, Academia Sinica

Abstract

An integrated analysis of big genomic data from a biobank provides an unmet opportunity to study health-related issues but must face a challenge in batch effect. There is an urgent demand for analysis tools that can simultaneously adjust for batch effect and provide an unbiased analysis of allele frequency (AF), allelic imbalance (AI), loss of heterozygosity (LOH), long contiguous stretch of homozygosity (LCSH), and copy number variation or alteration (CNV/CNA). In this study, we extended our software ALICE (AF/LOH/LCSH/AI/CNV/CNA Enterprise) by incorporating a new plate-based tiling array method to reduce batch effect. Local references for different types of chromosomal aberrations were generated based on local samples in the nearest neighbor plates to a sample to be examined, and then used to detect relative chromosomal aberrations. A parallel computing algorithm was implemented to increase computational efficiency. In addition, the results of our qPCR experiments for previous copy number validation were used to examine the performance of ALICE compared to PennCNV. We applied our batch effect reduction method to a large genomic dataset of 10,245 samples, genotyped with the Axiom TWB arrays, from the Taiwan Biobank Project. A dramatic reduction, 68.6% – 100.0% of the between-batch variation, was observed. The first draft of genomic maps of AF, AI, LOH/LCSH, and CNV/CNA in a Taiwan population and its three subpopulations (Minnan, Hakka, and Mainlanders) were established. The qPCR experiments showed ALICE outperformed PennCNV with regard to sensitivity in detecting copy number aberrations. In conclusion, ALICE with a user-friendly interface provides a powerful tool for an integrated analysis of big genomic data from biobanks.

[A-2]

On fixed effects estimation for spatial regression with spatial confounding

Chun-Shu Chen
Institute of Statistics and Information Science,
National Changhua University of Education

Abstract

Spatial regression models are often used to analyze the ecological and environmental data sets over a continuous spatial support. Issues of collinearity among covariates are often considered in modeling, but only rarely in discussing the relationship between covariates and unobserved spatial random processes. Past researches have shown that ignoring this relationship (or, spatial confounding) would have significant influences on the estimation of regression coefficients. To improve this problem, an idea of restricted spatial regression is used to ensure that the unobserved spatial random process is orthogonal to covariates, but the related inferences are mainly based on Bayesian frameworks. In this talk, an adjusted generalized least squares estimation method is proposed to estimate regression coefficients, resulting in estimators that perform better than conventional methods. Statistical inferences of the proposed methodology are justified both theoretically and numerically. (This is a joint work with Yong-Hui Qiu)

Cluster-Temporal Models for Disease Surveillance with an Application to Dengue Fever Infection in Taiwan

Pei-Sheng Lin
Institute of Population Health Sciences, National Health Research Institutes
Department of Mathematics, National Chung Cheng University

Abstract

Cluster analysis is a useful tool to explore underlying structures of a stochastic process and relations between observations by grouping items into same categories according to their similarity. Of particular interest is to cluster spatial-temporal units with elevated risks with accounting for potential risk factors and spatial-temporal correlation simultaneously. A motivation for this paper is from the study and surveillance of dengue fever (DF) infection in Taiwan. Since there is no effective vaccine or specific medicine to treat the dengue infection, a surveillance system, which can map clusters of cases, identify virus serotypes, and evaluate impact of environmental factors, is thus essential to prevent DF epidemic. We first develop an integrated cluster-temporal model and related parameter estimation. Then, an iterative procedure for identification of spatial-temporal clusters is devised. We adapt a deviance criterion for model selection such that the identified clusters can reflect different diffusion patterns. The proposed method is applied to the DF data of Taiwan for illustration.

Keywords: cluster analysis、disease surveillance、scan statistics、spatial-temporal data

[A-2]

Asymptotic Properties of the Maximum Likelihood Estimators in Misspecified Linear Mixed-Effects Models with Fixed Numbers of Clusters

Hsin-Cheng Huang
Institute of Statistical Science, Academia Sinica

Abstract

We consider a linear mixed-effects model with clustered structure, where the parameters are estimated by maximum likelihood (ML) based on possibly unbalanced data. Inference of this model is typically done based on the asymptotic theory assuming that the number of clusters tends to infinity with the sample size. However, when the number of clusters is fixed, the traditional asymptotic theory developed under a divergent number of clusters is no longer valid. In this research, we establish the asymptotic properties of the ML estimators of the random-effects parameters under a general setting, including models with fixed numbers of clusters. Our asymptotic theorems allow both the fixed effects and the random effects to be misspecified, and the dimensions of both effects to go to infinity with the sample size. This is joint work with Chih-Hao Chang and Ching-Kang Ing.

Constructing Generalized Linear Models by Data Information

Philip E. Cheng and Michelle Liou
Institute of Statistical Science
Academia Sinica

Abstract

Orthogonal decomposition of information identities of data likelihood is developed as a schematic procedure to identify the least indispensable predictors and interaction effects. This establishes a basic approach to constructing parsimonious generalized linear models with multivariate data of arbitrary sizes and variables before selecting a regression model. Empirical study of several data is used to illustrate the key analysis of data information, as exemplified in the construction of log-linear models and logistic models.

Keywords: Data information identity, logistic models, log-linear models, mutual information.

[A-3]

Online Learning for Multiclass Classification with Applications

Henry Horng-Shing Lu, National Chiao Tung University

Abstract

Supervised learning based on the methods of support vector machine (SVM) and the related techniques are very useful for the classification of complex data. However, the computation cost is very high when the training data set is massive. Online learning problems will need to handle the problems of memory limitation and computational complexity. In this study, the online learning methods by SVM and the related techniques for multi-class problems in massive data are developed. The empirical performance of these methods will be evaluated by real data applications, include communication network traffic management and others.

[A-3]

On the model checking for the single-index model

黃逸輝
淡江大學數學系

Abstract

The single-index model is a popular semiparametric model. It has the flexibility as nonparametric regression does and avoids the curse of dimensionality. It has been widely used and only a few diagnostic tools are available for model checkings. Most of the existent methods are based on deleted residuals, bootstrapping samples or martingal transforms to make the test being asymptotical distribution free (ADF).

In this talk, we propose another approach that computes the bias directly and then amends the test statistics to have ADF property. The construction of the test statistics is straightforward. It requires no bootstrapping or deleted residuals. Our test could be preferable regarding the performance of power through a simulation study.

[A-4]

An Embedding Learning-based Approach to Predict Influencers for Location Promotion in Social Networks
基於嵌入學習之社群網路地點行銷影響力預測

李政德、陳宜均
成大統計所

Abstract

Users in online social media tend to share their life via social networks and location checkin actions. Thus Location-based Social Network (LBSN) data can be formed accordingly. While LBSN has been exploited for applications such as Point-of-Interest (PoI) recommendation and social link prediction, an emerging task is Location Promotion, i.e., finding opinion leaders to promote a sepcific PoI. In this work, we propose and tackle two novel tasks, Targeted Influencer Prediction (TIP) and Targeted Visitor Prediction (TVP), in the context of Location Promotion. Given a target POI l to be promoted, TIP aims at predicting a set of influential users who can attract more users to visit l in the future, while TVP is to find a set of potential users who will visit l in the future. To deal with TIP and TVP, we propose a novel graph embedding method, LBSN2vec. The main idea of LBSN2vec is to learn a lowdimentional feature representation for each user and each location in an LBSN. In order to effectively find the reasonable context of each node for LBSN2vec, we devise a new weighted and penalized random walk mechanism. Equipped with the learned embedding vectors, we propose two similarity-based measures, Attractiveness and Visiting scores, to predict the influencers and potential visitors. Experiments conducted on a large-scale Instagram LBSN dataset exhibit that LBSN2vec and its variant can significantly outperform state-of-the-art graph embedding methods in both tasks of TIP and TVP.

關鍵字：Graph embedding, Feature learning, Information networks

具羅吉斯函數鏈結共變量之混合式 t 線性混合效應模型

王婉倫、楊郁成
逢甲大學統計學系統計與精算碩士班應用統計暨計量財務組

摘要

長期追蹤資料的基於模式化之分類與分群議題在近年越來越受關注，有限混合 t 線性混合模型（FM-tLMM）已經成為執行當資料包含離群值時之常用分析工具之一。在本文中，我們提出了一個延伸的有限混合 t 線性混合效應模型（Extended FM-tLMM），其中分群潛在變數（成份指標）被假設為受固定觀測到的共變量之影響。相較於傳統的有限混合 t 線性混合模型，其假設混合機率是固定但未知，我們所提出的延伸型有限混合 t 線性混合效應模型利用邏輯斯迴歸模型鏈結先驗的分群機率與感興趣的共變量之關聯性。因此，延伸型有限混合 t線性混合效應模型提供更準確的模型參數估計與更好的分群表現。為了獲得模型參數的最大概似估計，我們將模型重新建構為階層結構型態，並發展了交替的期望條件最大化（AECM）演算法。同時，我們利用一種基於信息之方法來計算參數估計量之標準誤。針對配適模型下個體間分群之問題也做了探討。我們利用一組有關愛滋病臨床試驗的實際資料與模擬研究以探究所提出模型之表現，並與現存模型在模型選擇、參數估計和分群方面作比較。

關鍵字：AECM 演算法; 厚尾現象； 訊息矩陣； 長期追蹤資料； 基於模式化之分群； 多變量 t 分佈

[A-5]

A New Demerit Control Chart for Monitoring the Quality of Multivariate Poisson Processes

Chung-I Li, Jen-Nan Pan and Min-Hung Huang
Department of Statistics, National Cheng Kung University

Abstract

This study aims to develop a new demerit control chart suitable for monitoring the quality of manufacturing processes with multiple characteristics subject to multivariate weighted Poisson distribution. Considering the correlation among different quality characteristics and their degrees of influence on the final product, we propose a new statistic for demerit scheme which gives different weights to different quality characteristics. Then, a new demerit control chart for multivariate weighted Poisson distribution (WMP chart) is developed accordingly. Moreover, a simulation study is conducted to evaluate the detecting performances of our proposed WMP chart and multivariate Poisson control chart (MP chart). Finally, a numerical example with a two dimensional telecommunication data set is given to demonstrate the usefulness of our proposed WMP chart. Both the simulation results and numerical example show that the detecting ability of our proposed WMP chart outperforms that of the MP chart when a process shift occurs. Hopefully, the results of this research can provide a better alternative for detecting the mean shifts occurred in a multivariate Poisson processes.

Keywords: Demerit control chart、Average run length、Multivariate Poisson control chart、Multiple quality characteristics、Multivariate weighted Poisson distribution

[A-5]

Non-homogeneous cumulative link models for ordinal data

俞一唐
東海大學統計學系

Abstract

Ordinal responses commonly arise in industrial experiments. To analyze data with ordinal responses, cumulative link modes are constantly used. These models assume the homogeneity among transferred cumulative probabilities, for example the proportional odds assumption. In some real examples, however, this homogeneity assumption does not hold. In this work, we propose using Gaussian processes to construct a non-homogeneous cumulative link model. In the proposed model, several Gaussian processes are considered simultaneously. Under Bayesian framework, a foam experiment is finally analyzed.

Keywords: Bayesian analysis、cumulative link model、Gaussian process、ordinal data

[A-5]

Solar panel lamination with extreme value regression model

Chih-Chun Tsai
Department of Mathematics, Tamkang University

Abstract

During the lamination process of solar module, the performance of the solar modules has been greatly relevant with the degree of cross-linking for EVA sheet. Traditionally, the degree of cross-linking for EVA sheet is obtained by using the extraction method to measure the gel content of EVA sheet. Motivated by lamination tests on solar modules, this study first constructed the statistical model with extreme value residuals to describe the relationship between the degree of cross-linking for EVA sheet and lamination time. Then, under the specification upper and lower limits of the degree of cross-linking for EVA sheet, the optimal lamination time of solar modules will be derived, and the optimal sample allocation for measuring EVA sheets destructively will be addressed. Currently, the new method is differential scanning calorimetric that measures the curing degree of EVA sheet as the degree of cross-linking for EVA sheet. This study also determined the specification upper and lower limits of the degree of cross-linking for EVA sheet under differential scanning calorimetric method.

Keywords: Solar module、cross-linking degree、extraction method

[A-6]

Assessing wage dynamics and stagnation using quantile dynamics

Yi-Hau Chen
Institute of Statistical Science, Academia Sinica

Abstract

Workers in Taiwan overall have been suffering from long-lasting wage stagnation since the mid-1990s. It is of interest to see if certain groups of workers, such as female, lower educated, and younger generation workers, suffer from the problem more seriously than the others. This work tries to apply a systematic statistical approach to study this issue, based on the longitudinal data from the Panel Study of Family Dynamics (PSFD) survey conducted in Taiwan since 1999. We propose the quantile dynamics regression model, generalizing recent methodology for quantile association, to assess the wage dynamics with respect to the marginal wage quantiles over time, as well as the effects of certain demographic and job factors on the wage dynamics. Estimation of the model can be based on the composite likelihoods utilizing the binary or ordinal information regarding the quantile dynamics, with the associated asymptotic theory established. A goodness-of-fit procedure for the proposed model is also developed. The performances of the estimation and the goodness-of-fit procedures for the quantile dynamics model are illustrated through simulation examples. The application of the proposed methodology to the PSFD survey data reveals that female, younger generation, private sector workers with education below postgraduate level experience more severe wage stagnation than the others. (Joint work with Chih-Yuan Hsu, Ruoh-Rong Yu, and Tsung-Wei Hung at Academia Sinica.)

Keywords: longitudinal data、panel study、quantile association、quantile regression、transition probability

[A-6]

Estimating the Optimal Number and Location of Cut Points

張 中
中山大學應數系

Abstract

In clinical practice, researchers usually categorize continuous variables for risk assessment. Many algorithms have been developed to find one optimal cut point to group variables into two halves; however, there is often need to determine the optimal number of cut points and their locations at the same time. In this proposal we propose a new AIC criterion, where the AIC values are corrected with crossvalidation and Monte Carlo method, to select the optimal number of cut points. In addition, the cross-validation and Monte Carlo methods will be used to correct the p-value and relative risk. To provide the biomedical researchers with an easy tool, we aim to develop an R function that utilized the genetic algorithm to find the location of the optimal cut points. In addition, we plan to conduct simulations to study the performance of our proposed method. In the end, we will apply our method to some real data examples.

Keywords: Monte Carlo method、*Cross-validation*、*cut points*、*AIC*

Univariate feature selection and compound covariate for predicting survival

Takeshi Emura,
Graduate Institute of Statistics, National Central University

Abstract

Univariate feature selection is one of the simplest and most commonly used techniques to develop a multigene predictor for survival. Presently, there is no software tailored to perform univariate feature selection and predictor construction. We develop the compound. Cox R package that implements univariate significance tests (via the Wald tests or score tests) for feature selection [1]. We provide a cross-validation algorithm to measure predictive capability of selected genes and a permutation algorithm to assess the false discovery rate. We also provide three algorithms for constructing a multigene predictor (compound covariate, compound shrinkage, and copula-based methods), which are tailored to the subset of genes obtained from univariate feature selection. We demonstrate our package using survival data on the lung cancer patients. We examine the predictive capability of the developed algorithms by the lung cancer data and simulated data. The developed R package, compound. Cox, is available on the CRAN repository. The statistical tools in compound. Cox allow researchers to determine an optimal significance level of the tests, thus providing researchers an optimal subset of genes for prediction. The package also allows researchers to compute the false discovery rate and various prediction algorithms.

Keywords: cancer prognosis、copula、Cox regression、cross-validation、dependent censoring、false discovery rate、gene expression、high-dimensional data、multiple testing

[B-1]

Going from the classroom to the studio: the birth and infancy of Stats + Stories

A. John Bailer
ISI President-Elect
Miami University, Oxford Ohio

Abstract

There are statistics behind the stories that are reported and there are stories behind these statistics. In this talk, I describe the collaboration between statisticians and journalists to develop a podcast to address this issue. The history and context for the 5 year old Stats + Stories podcast will be presented. The steps for implementing this podcast along with its evolution and future will be described. A consideration of what is needed to develop this in other languages will be discussed.

Decision making under uncertainty: the case of sea level rise adaptation

Peter Guttorp
ISI Vice-President
University of Washington

Abstract

There are several levels of uncertainty involved in making adaptation decisions to protect against sea level rise: the uncertainty in sea level rise projections, the uncertainty in flooding costs, the change in flooding costs as sea level rises etc. It is essential to take into account each of these uncertainties, as replacing uncertainties with, say, median values, easily can lead to substantial underestimation of the total costs with and without adaptation. We illustrate the issues with a case study from Norway.

[B-1]

Continuous Time Capture-Recapture

Richard Barker
Division of Sciences
University of Otago

Abstract

Motivated by field sampling of DNA fragments, we describe a general model for capture-recapture modeling of samples drawn one at a time in continuous-time. Our model is based on Poisson sampling where the sampling time may be unobserved. We show that previously described models correspond to partial likelihoods from our Poisson model and their use may be justified through arguments concerning S- and Bayes-ancillarity of discarded information. We demonstrate a further link to continuous-time capture-recapture models and explain observations that have been made about this class of models in terms of partial ancillarity. We illustrate application of our models using data from the European badger (Meles meles) in which genotyping of DNA fragments was subject to error.

High frequency data in statistics

Mark Podolskij
Council Member of Bernoulli Society (BS)
Aarhus University

Abstract

In this talk we present a short survey of recent developments in the statistical analysis of high frequency data. This type of observations appears in various applied fields such as economics, stereology, biology or physics. The notion “high frequency” refers to the fact that the time step between two adjacent observations converges to zero. We highlight some statistical problems that are particular related to high frequency data. We demonstrate various asymptotic results in the framework of diffusion and fractional type models.

[B-2]

Eye-tracking in practice: the EyeTrackR R package and its use in a study on human postures

Jürgen Symanzik
President-Elect of International Association for Statistical Computing (IASC)
Utah State University

Abstract

Eye tracking has been used in many scientific fields, such as behavioral sciences, education, marketing, and sports. Visualization usually plays an important role in the analysis of eye tracking data. In this presentation, I will first introduce the EyeTrackR R package that was developed for processing and visualizing eye tracking data from people looking at scientific posters. I will then present our first results of a study that tries to determine where people are looking when ranking the stability of a model holding certain postures. A portable eye-tracker is used to record the original video data of people looking at human postures. Image processing is used to extract statistical information from the video data.

This is joint work with Chunyang Li, Boyu Zhang, Eric McKinney, and Breanna Studenka.

Partnership between offcial statistics and academia

Irena Križman
Past President of International Association for Offcial Statistics (IAOS)

Abstract

In the Agenda 2030 statistical organisations and, especially those at the national level, were mandated with the ambitious task to monitor the implementation of the Sustainable Development Goals (SDGs). After intensive discussion of statisticians of different data communities, SDG Indicator framework has been adopted by the UN ECOSOC in 2017. To fulfil the requirements for data and statistics, among other activities different data communities have been encouraging to form new partnerships. Besides, the traditional partners are expected to strengthen their co-operation. Statistical societies at national, regional and global level play a very important role in this process. In my presentation I will focus on the partnership between academia and offcial statistics. Some good practices from Slovenia will be presented.

[B-2]

Optimal designs for comparison of response surfaces

Mong-Na Lo Huang[1], Tzu-Lun Yuan[1], Chi-Hsiang Chu[2]
Department of Applied Mathematics, National Sun Yat-sen University[1]
Clinical Trial Center, Kaohsiung Chang Gung Memorial Hospital

Abstract

In this work, we discuss the optimal design problems for comparison of response surfaces, which may be used for discriminating the expected responses between two or multiple experimental groups. In some cases in order to be able to discriminate the response surface models, estimation of nonlinear functions of the unknown parameters in the models may be needed. Then prior information about the unknown parameters in the response surface models would be needed for efficient design of experiments on group discrimination. In this talk, the locally-optimal and Bayesian-optimal designs for comparison of response surface models from two experimental groups on the two-dimensional space are presented, under certain prior information of the unknown parameters. Later cases with multiple experimental groups will also be discussed. The results are illustrated with some examples and the Bayesian optimal designs are compared with the locally optimal designs.

Keywords: Bayesian optimality、discrimination、equivalence theorem、locally optimality

A Construction of Cost-Effcient Designs with Guaranteed Repeated Measurements on Interaction Effects

Frederick Kin Hing Phoa
Institute of Statistical Science, Academia Sinica,

Abstract

This work introduced a useful class of cost-effcient designs for two-level multi-factor experiments. It provided guaranteed repeated measurements on all 2-tuples from any two factors and the number of repetitions was adjusted by the experimenters. Given the number of factors of interest, it utilized less resources than an orthogonal array while its repeated measurement provided a resistance towards outliers that a covering array failed to achieve. To bridge the wide spectrum between two extreme settings (orthogonal arrays and covering arrays) in terms of the number of repeated measures of tuples, we developed a systematic method to construct families of these designs, namely (supersaturated) repeated coverage design, with small run sizes under different number of factors and number of repetitions.

[B-2]

Stochastic search variable selection for definitive screening designs in split-plot and block structures

Chang-Yun Lin
Department of Applied Mathematics and Institute of Statistics,
National Chung Hsing University,

Abstract

Split-plot definitive screening (SPDS) and block definitive screening (BDS) designs have been developed for detecting active second-order effects in screening experiments when split-plot and block structures exist. In the literature, the multistage regression (MSR) and forward stepwise regression (FSR) methods were proposed for analyzing data for the two types of designs. However, there are some limitations and potential problems with the regression approaches. First, the degrees of freedom may not be large enough to estimate all active effects. Second, the restricted maximum likelihood (REML) estimate for the variances of whole-plot and block errors can be zero. To overcome these problems and to enhance the detection capability, we propose a stochastic search variable selection (SSVS) method based on the Bayesian theory. Different from the existing Bayesian approaches for split-plot and block designs, the proposed SSVS method can perform variable selections and choose more reasonable models which follow the effect heredity principle. The Markov chain Monte Carlo and Gibbs sampling are applied and a general WinBUGS code that can be used for any SPDS and BDS designs is provided. Simulation studies are conducted and results show that the proposed SSVS method well controls the false discovery rate and has higher detection capability than the regression methods.

Keywords: Bayesian、effect heredity、false discovery rate、generalized least squares、Gibbs sampling、Markov chain Monte Carlo、restricted maximum likelihood、WinBUGS

Misspecification of a dependent variable in the joint model of the logistic model and random effect models

Chun-Chao Wang, Associate Professor, Department of Statistics, National Taipei University
Yi-Ting Hwang, Professor, Department of Statistics, National Taipei University
Chung-Chuan Chou, Dr., Division of Cardiology, Department of Internal Medicine and College of Medicine, Chang Gung University
Hui-Ling Lee, Dr., Department of Anesthesia, Chang Gung Memorial Hospital

Abstract

Misspecification of an outcome can occur owing to many reasons in many applications. A misspecified dependent variable in the logistic or probit model might result in biased or inconsistent estimators (Hausman, Abrevaya and Scott-Morton, 1997). By incorporating the latent construction proposed in Hausman et al. (1997), the bias of the estimators can be reduced. This paper revises the joint model setting for simultaneously considering the multiple serial measurement and a binary outcome that proposed by Hwang et al. (2015) to take into account of the possible misspecification in the outcome. The joint likelihood approach along with the EM algorithm is used to find the estimates. Monte Carlo simulations are conducted to compare the impact of misspecification on the estimates. A retrospective data for the recurrence of AF is used to illustrate the usage of the proposed model.

Keywords: Atrial fibrillation、joint model、logistic regression、misspecification random effect model

[B-4]

Estimating treatment effects for semicompeting risks data with treatment switching

Chia-Hui Huang, Yi-Hau Chen, Jing-Li Wang and May Wang
National Taipei University, Academic Sinica, Taipei Medical University and
Taiwan FDA

Abstract

In clinical trials, it often arises that some patients randomized to control group switch to the alternative treatment after tumor progression. In such a study design, it is very likely that the patient's initiating the alternative treatment is related to his/her prognosis and may change the hazards of death. In this work, we propose a class of semicompeting risks models to estimating the effects of the combination treatments. The framework consists of a copula model for predicting the survival times of the two events at the first stage, and a conditional hazard model for predicting the death after progression with treatment switching. The maximum likelihood inference procedure is developed for the proposed models. Simulation studies and application to a clinical study are conducted to demonstrate the utility of the proposed methods.

Keywords: Maximum likelihood method、treatment switching、dependent censoring、semicompeting risks

Nonparametric testing for multiple survival functions with non-inferiority margins

張馨文
中央研究院統計科學研究所

Abstract

New nonparametric tests for the ordering of multiple survival functions are developed with the possibility of right censorship taken into account. The motivation comes from non-inferiority trials with multiple treatments. The proposed tests are based on nonparametric likelihood ratio statistics, which are known to provide more powerful tests than Wald-type procedures, but in this setting have only been studied for pairs of survival functions or in the absence of censoring. We introduce a novel type of pool adjacent violator algorithm that leads to a complete solution of the problem. The limit distributions can be expressed as weighted sums of squares involving projections of certain Gaussian processes onto the given ordered alternative. A simulation study shows that the new procedures have superior power to a competing combined-pairwise Cox model approach. We illustrate the proposed methods using data from a three-arm non-inferiority trial.

[B-5]

Variance Components Risk Premium: Evidence from Option Valuation and Stock Returns

[1]Hung-Wen Cheng, [2]Pei-Jie Xiao
[1]Soochow University
[2]National Taiwan University

Abstract

This study nests volatility components, jump intensity dynamics and non-monotonic pricing kernels in a single option model and contributes to describe index returns and option data. We construct a GARCH option pricing model with these three features yields a closed-form option valuation formula. We study the relation between variance risk premium and future market returns. In regressions, the model-implied long-run variance risk premium has predictive power for future market returns up to twelve months. The total variance risk premium is mainly attributed to long-run variance risk premium.

Keywords: Long-run、Short-run、Variance risk premium、Non-monotonic pricing kernel、Return predictability

On Corrected Diffusion Approximation with Two-flat Boundaries

Sheng-Feng Luo
Department of Finance, Chung Yuan Christian University

Abstract

We extend the method of Keener [Bernoulli 19, pp. 137--153 (2013)] to achieve a corrected diffusion approximation for boundary-crossing problems with two flat boundaries simultaneously, one is upper and the other is lower. The central idea is to appropriately apply the principle of smooth fit, originated from free-boundary problems, to both boundaries simultaneously. The approximation attained can be used to improve the rate of weak convergence between associated first-passage times monitored discretely and continuously. The resulting correction form still involves adjustments in the two levels of boundary, but the amounts adjusted can be different for the upper one and for the lower one. More interestingly, the way to shift the boundaries can also be either in the same direction or in the reverse direction. We further apply our correction method to the pricing of discrete double-barrier options to illustrate its application. Numerical examples show that our theoretical achievement performs quite well.

Keywords: first-passage time、corrected diffusion approximation、discrete option

[B-5]

Robust test of stock return predictability under heavy-tailed innovations

Cheng-Der Fuh and Meng-Hua Chung
Graduate Institute of Statistics, National Central University
Tianxiao Pang
School of Mathematical Sciences, Zhejiang University
Hsin-Chieh Wong
Institute of Economics, Academia Sinica

Abstract

Conventional tests of the predictability of stock returns are usually based on the normal innovation assumption, which does not fit the empirical data well. To remedy this ideal hypothesis, this paper considers a predictive regression model $y_t = \beta_0 + \beta_1 x_{t-1} + u_t,\ x_t = \rho x_{t-1} + e_t$ with possible heavy tails. To explore the predictability under heavy-tailed innovations, we construct a robust statistic based on this model. And then study asymptotic behaviour of the estimator of $(\beta_0,\ \beta_1)^T$ for stationary as well as local to unity cases. Our results show that when $|\rho| < 1$ or ρ tends to unity but slowly enough, the proposed robust statistic is indeed pivotal, and can be used directly to test the predictability of stock returns. Next, in the case of local to unity, we propose a modified test based on the celebrated Bonferroni Q-test in Campbell and Yogo (2006), and present the efficiency of this test. Finally, based on our proposed modified Bonferroni Q-test, we study the impact of heavy-tailed innovations to predictability. Numerical simulations and empirical studies are given for illustration.

Keywords: Domain of attraction of the normal law、Heavy-tailed、Least squares estimator、Predictive regression、Unit root

A nonparametric approach to semi-competing risks via causal mediation modeling

黃彥棕
中央研究院 統計科學研究所

Abstract

The semi-competing risk problem arises when one is interested in the effect of an exposure or treatment on both intermediate (e.g., having cancers) and primary events (e.g., death) where the intermediate event may be censored by the primary event but not the opposite. Here we propose a nonparametric approach by casting the semi-competing risk problem in the framework of causal mediation modeling. We set up a mediation model with the intermediate and primary events, respectively as the mediator and the outcome, and define indirect effect (IE) as the effect of the exposure on the primary event mediated by the intermediate event and direct effect (DE) as that not mediated by the intermediate event. A time-varying weighted Nelson-Aalen type of estimators are proposed for direct and indirect effects where the counting process at time tt of the primary event $N_{2n_1}(t)$ and its compensator $A_{n_1}(t)$ are both defined conditional on the status of the intermediated event right before t, $N_1(t^-) = n_1$. We show that $N_{2n_1}(t) - A_{n_1}(t)$ is a zero-mean martingale. Based on this, we further establish the asymptotic unbiasedness, consistency and asymptotic normality for the proposed estimators. Numerical studies including simulation and data application are presented to illustrate the finite sample performance and utility of the proposed method.

Keywords: Causal inference、Causal mediation model、Martingale、Nelson-Aalen estimator、Semi-competing risk

[B-6]

Effective Dimension Reduction in Multivariate Baseline Proportional Hazards Model

Ming-Yueh Huang
Institute of Statistical Science, Academia Sinica

Abstract

In this talk, we consider a nested family of multivariate baseline proportional hazards model for analyzing survival data. The family contains the Cox proportional hazards model and the continuously stratified proportional hazards model as special cases. It maintains the practically desirable hazard-ratio interpretation of target parameters, while allowing the control of multi-dimensional covariates in a nonparametric manner. The model also allows data-adaptive dimension reduction to reduce the effect of curse of dimensionality. Our goal is to strike a balance between flexibility and parsimony. Under the proposed model, we propose a novel criterion to implement the partial suffcient dimension reduction with right censored data. Further, we characterize the semiparametric effciency bound for estimating the target parameters and propose an effcient estimator. The effciency gain compared to the continuously stratified proportional hazards model is also proved.

Keywords: information bound、optimal bandwidth、partial suffcient dimension reduction、survival analysis

[B-6]

Response-adaptive treatment allocation for survival trials with clustered right-censored data

Pei-Fang Su
Department of Statistics, National Cheng Kung University

Abstract

In a clinical study, to compare two treatments with survival outcomes, treatment randomization may be required to be performed on clusters of multiple units with correlated responses. For example, for patients with both ears suffering from otitis media, a specific treatment is normally given to a single patient and hence the two ears will constitute a cluster. Statistical procedures are available for the comparison of treatment efficacies. Conventional approach for treatment allocation is the adoption of a balanced design with which half of the patients are being assigned to one of the treatment arms. However, with responsive-adaptive designs being more acceptable in recent years due to its desirable features, we derive a response-adaptive treatment allocation scheme for survival trials with clustered data. Our proposed treatment allocation scheme is superior to the balanced design in allowing more patients to be treated by the better treatment. The advantage of using our proposed randomization procedure is supported by a simulation study and the redesigning of a clinical study.

Keywords: Doubly adaptive biased coin design、test power、allocation function、copula model

解析式教育統計動態圖之思與用

許志銘
教育部統計處

摘要

簡潔視覺圖能瞬間打動人心突顯資訊亮點，惟目前廣泛所採行的主題式導向建構模式，其構思階段之簡約化美觀化處理程序，常使圖像所傳達觀點受限縮或隱蔽部分重要資訊，致未能普遍滿足各界需求。鑒此，在兼融主題式及資料導向的「解析式」思維下，參酌大數據分析之展示潮流，應用 D3（Data-Driven documents）為核心，瀏覽器為平台，搭配 Html 及 CSS 語法，針對學術上具深度解析典範圖樣（如 treemap、spiral、sankey 等等），撰寫具可攜性跨平臺跨載具之程式模組，將「視覺化線上分析」融入外界矚目的教育議題，具體視覺化資料查詢概念，並搭配複分類篩選功能提供階層探勘功能，引導使用者深入探索數據間隱藏的依存關聯，提高統計附加價值及決策品質。

[政府統計 I]

運用 Google Analytics 解析教育統計網站之服務效能

李書惠
教育部統計處

摘要

教育統計主要陳示我國教育發展動態及教育環境特徵變化，以供施政決策及學校輔導之參據，故本處自力建置各類資料查詢系統、電子書、動態視覺化圖表、教育統計簡訊等網路介面或單元，提供多元化的統計服務。為期提升服務效能，本文運用 Google Analytics 分析工具瞭解本處網站各類流量與使用情形，並定義三種關鍵指標用以評估教育統計查詢網之服務概況，進一步提出網站版面與功能優化建議，同時參考「政府網站版型與內容管理規範」逐項檢核網站功能與版面配置，提出未來整體網站改版之網頁配置設計示意，以貼近使用者需求之服務導向。

交通 e 級診療室

林淑敏
交通部統計處

摘要

過去交通政策多由專業人員擬訂與執行，而往往決策執行者認為有效並積極投入大量時間與成本後，才發現民眾感受不到，為創造交通運輸執行者與使用大眾雙贏局面，以「拉近兩者間感受差距」為設計核心，並聚焦在交通部運輸政策核心重點「公共運輸」，進一步打造透明化之「交通 e 級診療室」服務，讓實際公共運輸使用者參與績效診斷與政策擬訂角色，以提高外界對政策的好感度。

我們設計一系列「交通類有感統計指標」，採用醫療診斷流程做為推廣交通運輸設計概念，將交通各業統計資料以視覺化圖像呈現，讓實際公共運輸使用者提供精進良方，創建「交通 e 級診療室」作為交通運輸擬訂醫(e)療等級政策之服務平台。

美中貿易摩擦對我國產業之影響

何宗欣
經濟部統計處

摘要

今年以來美國對中國大陸價值 340 億美元及 160 億美元之商品，分別於 7 月 6 日及 8 月 23 日加徵 25%關稅，9 月 24 日再次針對 2,000 億美元加徵 10%之關稅，中國大陸亦以對等反制，兩大經濟體之貿易戰持續在進行中，對全球經濟恐造成不利影響，尤其我國以出口為導向，與中國大陸貿易往來密切，所受衝擊值得探討。

本文除就我國與中國大陸貿易依存關係、在陸台商對我國原材物料之仰賴程度等相關資料初步解析外，並就美國對中國大陸加徵關稅 6,842 項之美國海關產品碼與我國工業產品分類對照，歸納加稅產品所屬行業別，俾瞭解各行業受衝擊情形，最後利用 OECD 附加價值貿易 TIVA 資料庫，拆解美國各業別自中國大陸進口金額中，由我國貢獻之附加價值，俾評估對我經濟之直接影響，以充分發揮統計支援決策功能。

從海關統計看國際經貿時事二三事

陳韻潔、張莉玲
財政部關務署統計室

摘要

海關進出口貿易統計隱含國際貿易實務環境的諸多資訊，除總體經濟層面的應用外，亦是反映各國匯率政策、國際貿易政策（自由或保護）的重要資訊。本文由以下實例，概述應用情形：

1. 美國 2016 年 10 月「美國主要貿易夥伴外匯政策報告」(Foreign Exchange Policies of Major Trading Partners of the United States），將我國列入藉由匯率政策從事不公平貿易競爭觀察名單，本室為配合中央銀行與美國洽商需要，就臺美雙邊貿易統計差異進行分析，報部（財政部）後提供央行，並賡續提供相關資料。

2. 美國貿易擴張法 232 條款對全球鋼鐵、鋁產品加徵關稅，我國積極爭取關稅豁免，本室即分析我國相關產品出口情形，供本署長官參考。

運用大數據把關海關統計資料品質

黃敏慈
財政部關務署統計室

摘要

海關進出口貿易統計為我國重要經濟統計，由關務署統計室專責編製，以海關進出口報單為主要資料來源。近年海關順應國際趨勢，推動通關自動化與便捷化，「免審免驗」成為主要通關方式，加上「先放後核」等制度，使貿易統計完整性、正確性及即時性之確保益顯不易。

海關每月報單量約百萬份(不含快遞簡易報單)，貨品分類及貿易夥伴國組合達 280 萬種，此龐大數據量的量值檢核工作，實難以全面兼顧。本室運用大數據技術進行資料探勘，將紙本檢核作業全面電子化，融合統計學、歐盟技術文件與我國貿易統計資料特性，擘建自動化資料分析模型，偵測異常情形，配合視覺化結果呈現及自動派送作業，大幅提高檢核效率，確保資料品質。

以基地台信令觀察人口移動特性及常住行為

賴威宇
內政部統計處

摘要

戶籍登記為我國各項施政措施及資源分配之重要參據，舉凡交通、水利、用電等民生建設，治安、消防等公共安全，人文、教育及社會福利，皆以此作為規劃考量。但根據99年人口及住宅普查顯示，目前有超過2成國人呈現戶籍與居住地不同之情況，此差異恐影響資源分配之合適性。

本研究將以雙北市為例，蒐集電信業者基地台的信令資料，分析連續7日各時段之移動及使用行為，利用信令所在的剖面特徵，以集群分析方法將相似行為者歸類分群，初步推估各行政區之手機族常住人口數。並透過市話及手機調查，估計各行政區之手機族比率，最後再以比值估計法推估各行政區之常住人口，進而推估雙北的常住人口數。

如何促進民眾使用統計資料及擴大應用範圍

陳瑛璟、黃騰皜
內政部統計處

摘要

內政部今(107)年首次舉辦「2018 資料創新應用競賽－社會經濟空間資料暨戶政應用」活動，由統計處與戶政司共同規劃辦理，本次吸引學界及業界等不同領域團隊參賽，參賽作品題材多元、創意十足，內容均圍繞與民眾生活息息相關的議題，包含醫療照護、教育文化、居家環境、戶政服務、創業選址等，為社會經濟資料應用開啟新視角、注入新活力。

為展現內政資料之應用範圍，本文另特以人口、建物串接應用之案例，探討臺北市潛在風險住宅及其分布狀況，並加入實價登錄資料作為進一步找尋有利都市更新區域之所在。

淺談臺北市房價負擔能力指標與婚育之關聯

謝之婕
臺北市政府主計處

摘要

國人普遍將購屋視為婚姻與育兒之必須，高房價使得適婚年齡人口推遲婚育計畫，敲響「人口衰退」之警鐘。本文藉由內政部住宅及人口統計資料，並透過 Pearson 相關係數分析民國 91 年至 106 年臺北市房價與婚育之關聯性。

分析發現臺北市貸款負擔率與粗結婚率呈中度正相關，貸款負擔率雖高，但因結婚會增加購屋意願，並未對結婚率造成很大影響。又「助妳好孕」政策實施前，貸款負擔率與總生育率呈高度負相關；政策實施後，則呈中度正相關，係政府補助誘因，對市民生育意願的提升有所助益。

近三年臺北市生育率呈下滑趨勢，政府應持續觀察房價與婚育之關聯，將有助於相關單位作為研擬住宅政策之參據，進而提高市民之婚育率。

高雄市人口遷徙及未來發展

紀柏宇
高雄市政府主計處

摘要

報載本市 25 至 34 歲青年人口 6 年來大減 8 萬 2 千餘人，並以青年人口外流為題說明本市人口發展，實屬統計資料誤用之典型。簡言之，不同時間點之 25 至 34 歲人口非為同一族群，僅能約略比較期間數量差異，說明不同世代的人口結構組成，斷無法逕為推論人口外移等因素。

103 年 12 月桃園市升格成為第 6 個直轄市，因近雙北及機捷通車效應加持，在具地利之便的契機下，該市人口成長速度高居全國之冠；又 106 年 7 月臺中市人口總數超越本市，躍居全國第 2 大都市，區域的競爭使本市在城市發展面臨至為重要的關鍵時刻，尤應深入了解人口流動特性及趨勢。

回顧過去，看看現在，展望未來，本文將運用統計技術嘗試推估本市未來十年人口發展概況，俾供未來施政重要參考。

從電子票證數據觀察運具轉乘

郭昌儒
交通部統計處

摘要

完善之公共運輸系統為國家發展重要基礎，在交通部積極推動並鼓勵民眾使用電子票證搭乘公共運具政策下，電子票證使用比率有逐年成長之趨勢。然而，電子票證除了提供乘客便利之付費方式外，亦提供客運業者安全快速之票款計算收取功能，最為重要是電子票證記錄每位乘客搭乘各種交通運具之旅運行為資訊，因此電子票證數據資料對於公共運輸營運規劃管理有實質重要的助益。

因此，本篇研究透過分析大量電子票證交通交易數據，深入採掘近 3 年累積超過 60 億筆之電子票證交通大數據，瞭解民眾公運具搭乘旅運需求，萃取更具政策意涵之資訊，以增進電子票證資料之加值應用價值，為交通部下階段公運施政重點所需 KPI 構思規劃，藉由電子票證刷卡資料串聯，清楚呈現民眾搭乘不同運具之轉乘行為，創編運具轉乘率指標，對運具間無縫接軌推動目標，提出科技數據之新觀察。

證據基礎決策—以高齡駕駛人駕照管理制度為例

呂麗慧、周淑芬、洪裕鑫
交通部公路總局主計室

摘要

我國已於 107 年 3 月邁入高齡社會，基於關懷及維護年長者行車安全，公路總局自 102 年即研議規劃高齡駕駛人駕照管理制度，期間歷經 105 年之試辦及評估修正，至 106 年 7 月 1 日起正式實施。本案統計人員配合政策需求，適時產出正確、有效之統計資訊，可謂將證據基礎決策精神充分發揮之重要案例。所提供的證據資料包括：

1. 分析試辦資料，檢討修正認知功能測驗內容，降低政策偏差。
2. 透過民眾對高齡駕駛人管理意見調查，確立政策實施對象及條件。
3. 繪製認知功能測驗施測管道分布地圖，評估施測能量充足性。
4. 藉由制度實施半年之滿意度調查，探析民眾看法，作為政策檢視及調整之重要參據。

以大數據探討道路交通事故肇事原因

江欣容、黃逸勤、范宜鴻
內政部警政署統計室

摘要

警政署交通業務之兩大主軸為道路交通事故處理及執法。本研究利用近 6 年 A1、A2 類道路交通事故案件及當事人原始資料檔與臺北市違規資料檔，透過 Tableau 軟體串接資料進行大數據分析，內容分為「國道交通事故分析」及「交通違規與肇事之關連性分析」，前者就國道 1 號各路段肇事率、肇事原因、時段與車種別等交叉分析，研究發現肇事率與車流密度呈高度相關並提出降低事故之應用策略；後者以臺北市為例，探討交通事故與用路人違規之相關性，提供防制道路交通安全、警察加強執法及制定交通罰責之參考。研究結果並製作成 3 分鐘微電影短片，透過影片擴大宣導，提醒用路人小心駕駛，減少事故發生。

運用巨量租賃資料—探討桃園 YouBike 站點及需求配置

歐長潤、簡呈澔
桃園市政府主計處

摘要

受全球溫室效應的影響，世界各地氣候開始出現高溫炎熱或嚴寒，以及旱澇不均造成乾旱或水患等異常現象。再者，石油資源有限，能源耗竭的問題亦長期受到全球的關注。

桃園市政府為鼓勵市民使用低污染、低耗能的公共自行車作為短程接駁運具，以達降低環境污染及能源損耗，並響應全球節能減碳風潮，同時亦可改善都市道路交通擁擠，爰自 105 年 2 月 4 日啟用公共自行車租賃系統(以下簡稱 YouBike)，並持續建置服務站點，截至 106 年底已開通 180 個站點。

為了解桃園市民對 YouBike 之使用需求，本文運用 105 年 2 月至 106 年 12 月桃園市 YouBike 租賃資料，藉由列聯表及對應分析觀察平假日、租用時段及租用時間之分布及關聯性，再藉由集群分析探討租借站設置之供需情形，進一步分析高頻率騎乘之租借站平假日借還次數差異，了解現行停車位數及車輛運補機制下之須加強調度時段，最後透過平假日主要騎乘路線觀察市民騎乘範圍，供相關機關制定施政決策之參考。

[C-1]

Critical two-point function for long-range self-avoiding walks with power-law couplings

Lung-Chi Chen of Mathematical Sciences, National Chengchi University

Abstract

Consider the long-range self-avoiding walks on Z^d, whose one step distribution $D(x)$ decays as $|x|^{-d-\alpha}$ for some α > 0. In our previous work (2015), we have shown that, for $\alpha \neq 2$, the critical two-point function $G_{p_c}(x)$ decays as $|x|^{\alpha^2-d}$ above the upper-critical dimension $d_c := 2(\alpha^2)$. In this talk, we show that $G_{p_c}(x)$ for α = 2 decays as $|x|^{2-d}/log\,|x|$ whenever $d \geq d_c$ (including equality). This solves the conjecture in (2015), extended all the way down to $d = d_c$, and confirms a part of predictions in physics (2014). The proof relies on the lace expansion and new convolution bounds on power functions with log corrections.

[C-1]

The localized phase transition of a polymer

Chien-Hao Huang
Department of Mathematics
National Taiwan University

Abstract

A polymer is penalized by long excursions. We discuss the free energy and the order of phase transitions. The behavior of infinite-volume system is also discussed.

Keywords: Polymers、phase transitions、localization

Waiting Time Analysis for a Restricted-Length M/G/1 Queue with Multiple Vacations

Gi-Ren Liu
National Cheng Kung University

Abstract

The random Sleep-Awake schedule design has been used extensively in the design of modern communication systems for reducing the energy cost. In view of that there exists a trade-off between the energy saving ratio and the service\communication delay and the waiting time analysis can provide a guideline for the operators to set up the length of sleep time for the Sleep-Awake scheme to maintain reasonable service delay and reduce the impact of buffer overflow, a recent work on the waiting time analysis for a restricted-length M/G/1 queue with multiple vacations will be presented in this talk. The details of its application on the design of green routers can be referred to [1] and [2].

Keywords: Waiting Time Analysis、Sleep-Awake scheme、M/G/1 queue、multiple vacations

[C-2]

A Multivariate Compound Poisson Model with Copula and its Application on Limit Order Book

Shih-Feng Huang and Ling-Yu Kuo
National University of Kaohsiung

Abstract

A multivariate compound Poisson model with copula is proposed to depict the dynamics of Limit Order Book (LOB), where the intensity rates of order arrivals are assumed to have autocorrelations. The distribution of the first-passage time when the best ask (or bid) price moves is derived under the proposed model. A method of short-term stock price prediction is also developed. The LOB data of Intel, Microsoft, Johnson & Johnson, Yahoo and Taiwan Semiconductor Manufacturing stocks on five different days during 2008 and 2016 are employed for the empirical study. The numerical results indicate that the proposed model has satisfactory performances on modeling LOB and predicting short-term stock prices.

Keywords: Compound Poisson、Limit Order Book、short-term stock price prediction

How Stochastic are the Innovations to a Comprehensive Volatility Model? A Point Process Analysis on High-Frequency Data

Ping-Chen TSAI
Department of Finance
Southern Taiwan University of Science and Technology

Abstract

Stochastic Volatility (SV) models differ from ARCH-type models in treating volatility as a process with its own dynamics, rather than the conditional variance of returns. By estimating several versions of a comprehensive volatility model for realised volatility measures, we construct the innovations to such models as point processes and study their temporal dependence. The innovations to a baseline model show a higher degree of self-clustering, but less so after important stylized facts are added to the volatility model. We also investigate the properties of innovations which can be explained by leverage effect and long memory of volatility. The empirical results are made robust to intraday volatility pattern and price jumps.

Keywords: Heterogeneous Auto-Regressive Model、Realized Bi-Power Variation、Hawkes Process、Realized Range、Leverage Effect

[C-2]

Risk Structure in an Extended Version of Merton's Jump-Diffusion Model Where Jump Magnitudes Follow an Autoregressive Process of Order 2

繆維中
台灣科技大學財務金融研究所

Abstract

Merton's jump-diffusion model is a popular model for asset price dynamics as it incorporates jumps to capture price discontinuity. However, the independent increments in both the diffusion and jump parts make it unsuitable to describe a real market with momentum and reversal effects. To model these inherent market characteristics, we extend the jump structure and assume the sequential jump magnitudes follow an autoregressive process of order 2, or an AR (2) process. By introducing two additional parameters, the proposed model allows for a wide range of correlation structure among the successive jumps. A mathematical analysis is given for our model and some analytical results are derived for the asset return distributions and the European option prices. Using these results, we provide numerical examples to investigate how the two autoregressive coefficients influence the asset return distributions, the European option prices, and the risk measures such as value-at-risk (VaR) and expected shortfall (ES). We demonstrate that both the autoregressive coefficients have a significant influence on the risk structure but they affect the above risk measures in a different way.

Keywords: jump-diffusion models、auto-regressive processes of order 2、option pricing、value-at-risk、expected shortfall

[C-3]

Planning of Accelerated Degradation Tests

李宜真
國立成功大學統計學系

Abstract

The accelerated degradation test (ADT) is an efficient tool for assessing the lifetime information of highly reliable products. Because conducting an ADT is very expensive, how to plan an ADT is a challenging issue for reliability analysts. By taking the experimental cost into consideration, this study proposes an efficient algorithm to determine the total sample size, testing stress levels, the measurement frequencies, and the number of measurements based on a class of exponential dispersion (ED) degradation models. For an ADT plan, the proposed method provides some design insights, and we further discuss the properties from the viewpoints of design.

[C-3]

Degradation Analysis on Trend Gamma Process

Yi-Fu Wang and Bo-Wei Su
Department of Mathematics, National Chung Cheng University, Chiayi, Taiwan

Abstract

Degradation analysis is widely used to address the lifetime information of highly reliable products by fitting the quality characteristics which degrade over time. When the degradation path has monotone pattern, the most commonly used model is the gamma process model. For dealing with the ageing phenomenon in the degradation dataset, this article introduces a more general gamma process model, called Trend Gamma Process (TGP). Then the reliability inference of TGP model are provided. Finally, a case study is presented to illustrate the benefits of the proposed model.

Keywords: degradation analysis、gamma process、trend function

Optimal Doubling Burn-in Policy Based on Tweedie Processes with Applications to Degradation Data

Chien-Yu Peng and Kun-Hung Lin
Institute of Statistical Science, Academia Sinica

Abstract

In the current competitive marketplace, manufacturers need to screen weak products in a short period of time. It is a challenge for manufacturers to implement a burn-in test that can screen out the weak products quickly and efficiently. When selecting an approach to determine the duration of the burn-in, one could build a criterion aiming to minimize the burn-in cost. In practice, when the optimal burn-in time is unreasonable (e.g., time 0) due to minimizing the cost, this means that the burn-in procedure is unnecessary to perform for manufacturers. In this study, we propose an optimal doubling burn-in policy to improve the predicament without additional experiments. The advantage of the proposed policy is to simultaneously determine the optimal burn-in time and the optimal cutoff point for classifying weak and strong components from the production. In addition, a new degradation model based on a Tweedie mixture process is used for a burn-in test. The proposed burn-in procedure is applied to a real data.

[C-4]

An Attention Algorithm for Solving Large Scale Structured l_0-norm Penalized Estimation Problems

Tso-Jung Yen
Institute of Statistical Science
Academia Sinica

Yu-Min Yen
Department of International Business
National Chengchi University

Abstract

Technology advances have enabled researchers to collect large amounts of data with lots of covariates. Because of the high volume (large n) and high variety (large p) properties, model estimation with such kind of big data has posed great challenges for statisticians. In this paper we focus on the algorithmic aspect of these challenges. We propose a numerical procedure for solving large scale regression estimation problems involving a structured l0-norm penalty function. This numerical procedure blends the ideas of randomization, proximal operators and blockwise coordinate descent algorithms. In particular, it adopts an attention-based sampling distribution for picking up regression coe_cients for updates based on a closed form representation of the proximal operator of the structured l0-norm penalty function. Simulation study shows the proposed numerical procedure is competitive to the benchmark algorithm for sparse estimation in terms of runtime and statistical accuracy when both the sample size and the number of covariates become large.

Keywords: Attention mechanism、Blockwise coordinate descent algorithms、Nonconvex optimization、Proximal operators、Randomized algorithms

Inverse Regression for Multivariate Functional Data

Lu-Hung Chen
National Chung-Hsing University

Ci-Ren Jiang

Academia Sinica

Abstract

Inverse regression is an appealing dimension reduction method for regression models with multivariate covariates. Recently, it has been extended to the cases with functional or longitudinal covariates. However, the extensions simply focus on one single functional or longitudinal covariate. Motivated by a real application, we extend functional inverse regression to the cases with multiple functional covariates, whose domains could be different. The asymptotical properties of the proposed estimators are investigated. The computational issues are taken care with data binning, the fast Fourier transformation and random projections on a multi-core computation platform. In addition to simulation studies, the proposed approach is applied to predict the wind power capacity factor of the next day with the weather forecasts made today. Both demonstrate the good performance of our method.

Keywords: Big Data、Multidimensional/Multivariate Functional Data Analysis、Inverse Regression、Parallel Computing、Smoothing

[C-4]

Sufficient dimension reduction via random-partition for large-p-small-n problem

Hung Hung
Institute of Epidemiology and Preventive Medicine, National Taiwan University

Abstract

Sufficient dimension reduction (SDR) is continuing an active research area nowadays. However, conventional SDR methods can fail to apply when the number of covariates is larger than the available sample size. To overcome the problem of high dimensionality, some works are developed to project the covariates onto a lower dimensional envelope subspace, on which conventional SDR methods can be directly applied without losing information. On the other hand, random-partition of covariates has been shown to be effective in assisting the detection of influential variables. In this work, we propose a new SDR method to overcome the problem of high-dimensionality, which we call random-partition SDR (RP-SDR). The main idea of RP-SDR is to use random-partition to construct the envelope subspace. The procedure will then be repeated many times, each corresponds to a realization of random-partition. Finally, an integration method is applied to average out the effect of random-partition. Comparing with existing methods, RP-SDR is less affected by the selection of tuning parameters. Moreover, the estimation procedure of RP-SDR does not involve the determination of the structural dimension until the final stage, which makes it more robust in estimating the target of interest.

以生活型態區隔顧客前往酒吧的消費行為

羅琪、陳廷侑、姜智鈞、葉家菱、曾文怡、楊承宇、陳儷心、吳家綺

中華大學餐旅管理學系

摘 要

本研究以酒吧消費者為調查對象，以 EKB 模式作為本研究組織架構，採用 AIO 生活型態作為市場區隔之基礎，並以酒吧之屬性、人口統計變數作為投入變數，探討不同集群之酒吧消費者區隔之特徵。研究結果顯示，酒吧消費者可分為「獨立與經常運動」、「追求新知精打細算」、「社交門面」三個集群。各區隔市場消費者，在人口統計變數（性別、每月可支配之所得）、消費決策行為變數（消費目的）與消費實態變數（消費時段、消費金額），具有顯著差異。最後分別整理出其集群特徵並針對各集群特徵結合本研究之研究目的提出產品、價格與行銷策略之建議。

關鍵字:生活型態、消費行為、市場區隔

[C-5]

檢定羅倫茲曲線之優勢
(A Likelihood Ratio Test for Lorenz Dominance)

章珅鎔，劉長萱，程爾觀
中央研究院統計科學研究所

摘要

羅倫茲曲線（Lorenz Curve, 1905）常用來衡量個別經濟區域的人均所得分配狀況。若 A 區域的羅倫茲曲線座落在 B 區域的羅倫茲曲線之上方，則代表 A 區的分配比較 B 區的分配來得均衡，定義為"A 分配對 B 分配有羅倫茲優勢（Lorenz Dominance）。已知文獻中對於檢定兩條羅倫茲曲線是否有羅倫茲優勢的關係，或者有交點，通常是用比較兩條曲線的樣本隨機過程之理論為根據（Barrett et al., 2014）。本研究的目的是按照羅倫茲優勢（或有交點）的基本曲線性質，構造兩個概似比檢定；一者，檢定虛無假設為「兩條羅倫茲曲線有一個交點」；另者，檢定虛無假設為「兩條羅倫茲曲線無交點」。若前者顯著拒絕，而後者未拒絕，則表示「兩條曲線無交點，即有羅倫茲優勢」；反之，則表示「兩條曲線有交點」；若兩者檢定皆拒絕，而顯著拒絕的程度不一樣，則表示「兩條羅倫茲曲線之間超過一個交點」，或者是「兩條曲線很接近、而可能是無交點」。我們的模擬分析使用「對數常態分配、韋伯分配及帕芮托分配」；設定參數分配、代表不同的羅倫茲優勢組合。於每一對分配組合，計算兩個檢定的拒絕概率（按一般檢定水準 α=0.05）。譬如，兩個對數常態分配有相同位置參數 μ，接近的尺度參數 σ and σ'；理論上，兩條羅倫茲曲線無交點，但很接近； 於是兩條樣本曲線產生多數交點，兩個檢定的拒絕率都高。以真實資料驗證檢定時，使用了「The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988」資料中 133 個國家的 GDP 數據;「羅倫茲優勢」及「兩條曲線有一個交點」的兩種狀況都有檢定成立。

關鍵字：羅倫茲曲線、羅倫茲優勢、概似比檢定

Goodness-of-fit Statistics for Multinomial Logistic Regression Models

Wei-Hsiung Chao
Department of Applied Mathematics, National Dong Hwa University

Abstract

Polytomous regression models are often fitted through the use of maximum likelihood to study the relationship between a categorical response and some covariates. To assess the fit of these models with only categorical covariates, it is appropriate to use the Pearson-Fisher's test for product multinomials. In the presence of continuous covariates, there exist ad-hoc statistics of Pearson-Fisher's type based on grouping strategies for assessing the adequacy of the fitted models. These statistics are often formed as a sum of Pearson's statistics over all groups in which the within-group observations are in fact heterogeneous. No asymptotic result was available to show that these ad-hoc statistics are chi-squared distributed when the fitted model is correct. Under certain situations, these ad-hoc statistics were found to be chi-squared distributed through simulation studies. Without using any grouping strategies, we recently proposed a Pearson-like statistic W based on pooled observations that is useful in assessing the fit of a polytomous regression model with non-natural link. The W statistic is a quadratic form in the differences between the observed totals and fitted totals over response categories. Under certain rank condition, the asymptotic null distribution of the proposed statistic was shown to be chi-squared with appropriate degrees of freedom. Because the multinomial logit link is the natural link for nominal response, these differences are identical to zero when fitting a multinomial logistic regression model to the data. Thus, the W statistic has no value in assessing the goodness of fit of multinomial logistic regression model. To overcome this difficulty, we extend the W statistic to the WG statistic which involves partitioning the covariate space into groups, and show that it is asymptotically chi-squared distributed.

Keywords: Goodness of fit、Pearson's chi-square test、natural link、multinomial logit

[C-6]

應用文字探勘技術探討財金新聞對股價影響之研究
-以半導體產業為例

林晏禎、林秋華
銘傳大學應用統計與資料科學研究所

摘要

近年來，隨著經濟與資訊科技的發展，愈來愈多人開始使用新科技（如：理財機器人、Fintech）來獲取理財相關的資訊。在網路上也出現大量投資理財的方法，提供民眾做為參考依據。對於股市漲跌之研究方法繁多，且各具其特色，但多數只注重基本面與技術面分析，這兩種分析方式卻疏忽了外界資訊變化對股價造成的影響。本研究嘗試以文字探勘技術針對財金新聞進行股價預測，進行斷詞處理及詞頻統計，找出其中特徵及關聯性，並以過去學者常用之影響股價相關變數及文字探勘之結果作為影響因子，進一步探討臺灣半導體類股股價與上述變數間之關係。

關鍵字：文字探勘、股價預測、詞頻統計、臺灣半導體類股

指數隨機圖模型的吉氏取樣改良

陳韋宏、郭鋸霖
國立高雄大學統計學研究所

摘要

本研究提供新的演算法來產生指數隨機圖模型的網絡樣本，基於傳統的吉氏取樣法，我們給出合適的初始網絡圖形，來增加收斂速度，再利用 Chatterjee (2013)及 Bhamidi (2011)定義的指數隨機圖模型，簡化吉氏取樣的疊代過程，能大幅減少計算量，進而更有效的產生網絡樣本。

關鍵字：網絡、指數隨機圖模型、吉氏取樣

[C-6]

Quantile autoregression model in real exchange rate reversion of Chinese Yuan

Yuan-Wei Chao、Luke Lin
Department of Finance National Kaohsiung University of Science and Technology

Abstract

The research utilized the Quantile Autoregression (QAR) technique to examine the mean reversion properties of the Chinese Yuan (CNY) against the currency of China's three primary trade partners, United Stated, Europe and Japan from the period of 1998 to 2017. In this period, the CNY was both in a de facto peg to the US dollar (USD) and a Basket Band Crawling (BBC) regime period. Under the QAR empirical result, we can find the more clear behavior of RER: (i) the PPP is partially existed, (ii) the recovery times of recent peg to BBC regime are longer than peg to USD regime, (iii) the RER recovery patterns were different when confronted between large deviation and neutral deviation, but there were less significant difference in the recent currency regime period, (iv) the recovery pattern of CNY/USD is similar on large appreciation and depreciation in the recent currency regime period.

Keywords: Purchasing power parity、Quantile autoregression、Basket band crawling、Real Exchange rate

台灣上市公司外匯暴險受多因子影響之探討

藍丹璟、林育志
國立高雄科技大學財務管理系

摘要

過去與外匯暴險相關的研究中，學者們多半使用 Jorion (1990) 的外匯市場模型 近年來的模型考慮了更多層面的因素，有學者將匯率變動加入至 Fama and French (1992) 的三因子模型來檢測其結果的差異與顯著程度，本文將 Fama and French (2014) 提出的五個因子當作模型的控制變數，構建一個新的外匯暴險估計模型，並將其與外匯市場模型與採用三因子構建之模型進行比較。在決定因素的部分參考楊聲勇等人(2009) 加入衍生性金融商品的使用程度，檢驗其對外匯暴險之影響，此外也以虛擬變數的方式檢驗正負暴險與產業特性對外匯暴險的影響。

實證結果發現加入五因子的模型在估計暴險時顯著公司家數較少，有可能是因為考量了多個因子後暴險被稀釋掉，進而影響外匯暴險決定因素與衍生性商品交互作用效果的結果，只有出口比率呈現顯著，代表出口比率越高，公司接觸外幣交易的機會也越多，外匯暴險越大，其餘避險活動指標的因子則都不顯著。

關鍵字：五因子、外匯暴險、衍生性商品

[C-6]

Performance of a two-sample test with Mann-Whitney statistics

Chun Huang Hsu
Graduate Institute of Statistics, National Central University

Abstract

The Mann-Whitney statistics is an intuitive measure for testing the equality of two survival distributions in the two sample survival problem which is independent right censoring data. We introduce the asymptotically inference produces based on standard normal, bootstrap, and permutation methods and compare them by the simulated coverage probability. According to the result, the permutation is better method than others. We also consider the performance of the methods when censoring time is dependent on survival time. Finally, we analyze a real data from the package and illustrate all procedures.

Keywords: Mann-Whitney effect、Kaplan-Meier estimator、Asymptotic、Bootstrap、Permutation technique

Pretest and shrinkage estimation of the mean under a univariate normal distribution

Jia-Han Shih and Takeshi Emura
Graduate Institute of Statistics, National Central University

Abstract

This paper first reviews the pretest estimation of the mean under a univariate normal distribution with known variance when uncertain non-sample prior information is available. We define a fairly general pretest estimator which incorporates both sample and non-sample information and includes many existing pretest and shrinkage estimators as special cases. Important statistical properties of the general pretest estimator such as its exact distribution, bias, and mean square error (MSE) are derived. Our expressions for the bias and the MSE are simpler and more straightforward than the existing expressions. In addition, we propose a new shrinkage estimator based on the cumulative distribution function of the normal distribution. We demonstrate that the proposed shrinkage estimator is competitive with the positive-part Stein-rule estimator in terms of the MSE. Finally, estimation with unknown variance is also discussed.

Keywords: Biased estimation、Mean squared error、Mixture distribution、Randomized test、Uncertain non-sample prior information

[D-1]

Parametrizing the Kepler Exoplanet Period-Radius Distribution with the Bivariate Normal Inverse Gaussian Distribution

Wen-Liang Hung
Center for Teacher Education, National Tsing Hua University
Department of Computer Science, National Tsing Hua University

Abstract

This talk presents a simple and robust method for obtaining a comprehensive understanding of the joint period and radius distribution in Kepler exoplanets. The proposed method is based on particle swarm optimization and bivariate Normal Inverse Gaussian distribution. Furthermore, in the construction of the probability density function, this study selects planet-host stars with the GK-type. The injecting approach is also employed to solve the survey completeness of sample. The resulting occurrence rate of Earth analogs is 0.025 with a 95% bootstrap confidence interval between 0.023 and 0.032.

Hi-C Data Normalization

Siao-Cyuan Wei and Huey-Miin Hsueh
Department of Statistics National ChengChi University

Abstract

In recent year, it has been shown that gene activities are strongly correlated with chromatin topology. To understand the three-dimensional structures of chromosomes genome wide, the chromatin conformation capture with high-throughput sequencing (Hi-C) technique is developed. The procedure takes several steps, and systematic technical biases are induced during the experiment. To have an unbiased identification on chromatin interactions, data normalization to minimize these biases are essential and necessary. In this study, we aim to develop a normalization method via using a negative binomial regression model. The findings in Yaffe and Tanay (2011) are adapted in the regression model. Numerical studies are performed to justify the proposed method. For illustration, our method is applied on several real data sets.

Keywords: Chromatin topology、Hi-C、Negative binomial、Normalization

[D-1]

The Normality Test in Meta-Analysis of Binary Outcome

Jin-Hua Chen
Graduate Institute of Data Science / Research Center of Biostatistics,
College of Management,
Taipei Medical University

Abstract

Evidence-based medicine (EBM) is very crucial for medical development, especially in clinical trial and medical education. Meta-analysis method is the useful tool in EBM to integrate some statistics from several similarly studies. There are two steps in meta-analysis procedure. First, we need to examine the heterogeneity between studies which we collected from different sources, for example, PubMed, MEDLINE…etc. Second, according the results of heterogeneity test, we could fit the fixed effect model or random effect model to combine the information from different studies. Then we would estimate and test the overall effect in fixed or random effect model. These are the traditional and useful models in meta-analysis reports. This method is two stage tests. The overall significance level might be higher than specific significance level (0.05). The random or fixed effect model is based on the normal distribution assumption, but we discuss it rarely in practice. To obtain the accuracy overall estimator, we need to concern the normality property in meta-analysis. We proposed the mixed effect model to measure the overall estimator. Before estimating the effect, we need to examine the normality assumption. The useful methods of normality test include Anderson-Darling test, Cramér–von Mises test, Shapiro–Wilk test. In our study, we would discuss the type I error rate in normality assumption and the power performance in meta-analysis of binary outcome. We generated data from 2x2 tables in real world position and calculations were by the bootstrap procedure.

Bayesian Adaptive Randomization Based on Time-to-Event Outcomes of Efficacy and Toxicity

張玉媚
東海大學統計系

Abstract

There are two main objectives in treating patients in Phase II clinical trials. The primary objective is to test whether a new treatment benefits the patients and the second objective is to evaluate the toxicology of new treatments. Developing an adaptive randomization procedure that takes into account both efficacy and toxicity is essential for ethical and practical reasons. In this article, we propose a Bayesian adaptive randomization procedure based on both efficacy and toxicity response, which are considered as time-to-event outcomes and their association is modelled by a common random effect (shared frailty model). Under such model, the randomization probability is allowed to depend on patients' specific covariates. Moreover, after filtering out excessive toxicity or futility, we assign each new patient to the superior treatment arm. We conduct simulation studies under different scenarios to examine the finite sample performance of our proposed method and compare it with the Bayesian adaptive randomization procedures that consider efficacy only.

Keywords: Bayesian adaptive randomization、frailty、phase II trial、random effect、time-to-event outcome、toxicity

[D-2]

Integrative Gene Set Analysis and Visualization in Genome-wide Association Study

蔡政安
國立臺灣大學農藝學系

In DNA microarray studies, gene-set analysis (GSA) has become the focus of gene expression data analysis. GSA utilizes the gene expression profiles of functionally related gene sets in Gene Ontology (GO) categories or priori-defined biological classes to assess the significance of gene sets associated with clinical outcomes or phenotypes. Many statistical approaches have been proposed to determine whether such functionally related gene sets express differentially (enrichment and/or deletion) in variations of phenotypes. GSA provides a direct approach to the analysis of gene sets of interest and the results are relatively easy to interpret. Furthermore, microarray experiments inherit various sources of biological and technical variability, and analysis of a gene set is expected to be more reproducible than an individual gene analysis. In this talk, I will introduce a general strategy of gene set association analysis (GSAA) for integrating prior biological knowledge with gene expression data. An integrative analysis is proposed to discover the relationships between gene sets significantly associated with phenotypes. In addition, a graphical technique is used for visualizing and simultaneously exploring the associations of between and within gene sets and their interaction and network.

關鍵詞：microarray，gene-set analysis (GSA)，Gene Ontology (GO)，gene set association analysis (GSAA)

Testing no Treatment Effect in Meta-Analysis

JY Lee[1], JH Chen[2] and KF Cheng[2, 3]
[1]Department of Statistics, Feng Chia University, Taichung Taiwan.
[2]Biostatistics Center, Taipei Medical University, Taipei Taiwan.
[3]Department of Business Administration, Asia University, Taichung Taiwan.

Abstract

In a meta-analysis of multiple trials, the fundamental problem is to test whether a new treatment of interest is better than the placebo or an active treatment. Because the data of a meta-analysis may be heterogeneous, fixed-effects approach and random-effects approach are often use. The fixed-effects approach assumes that the all trial effects are the same, while the random-effects approach usually assumes the effects are random and follow a normal distribution. Under the normal random-effects model, Han-Eskin test was proposed for testing the averaged effect and variance of the effects are both zero, that is, all treatment effects (log odds ratios) are zeros. In this paper, we compare this test with T_{REM} test, which was originally designed for detecting disease association signals in sequencing studies. The T_{REM} test has greater advantage in that no distributional assumption of random effects is required. Furthermore, a simulation study indicated that it was more powerful than Han-Eskin test under wide range of simulation conditions. Two examples are given to illustrate applications of these two tests.

Keywords: Association test、bootstrap、meta-analysis、random-effects model、type I error

[D-2]

風險偏好、VIX 期貨基差與 S&P 500 期貨報酬

童寶瑩、李昀寰、李修全
銘傳大學財務金融學系

摘要

本研究以報酬分配、52 週高點比率與歷史高點比率捕捉投資人的風險偏好，透過線性迴歸為基礎的決策樹模型探討風險偏好程度是否會影響 VIX 期貨基差對 S&P 500 期貨報酬的影響，並進一步分析在考慮了投資人的風險偏好程度後，是否會能改善 VIX 期貨基差對 S&P 500 期貨報酬的預測績效。實證結果顯示，在相對低的報酬、遠離 52 週高點及歷史高點時，投資人的風險偏好程度較低，此時 VIX 期貨基差對 S&P 500 期貨報酬會產生正向影響；在相對高的報酬、接近 52 週高點及歷史高點時，投資人的風險偏好程度較高，此促使 VIX 期貨基差對 S&P 500 期貨報酬產生負向影響。在預測方面，以報酬分配或 52 週高點比率作為區分風險偏好程度的變數較能提高 VIX 期貨基差對 S&P500 期貨報酬的預測績效。

關鍵字：風險偏好、VIX 期貨基差、S&P 500 期貨報酬

Detection of Location and Dispersion Effects from Partially Replicated Two-Level Factorial Designs

Shin-Fu Tsai
Department of Agronomy, National Taiwan University

Abstract

During the preliminary stage of a quality improvement process, identification of active location and dispersion effects is an important issue. After understanding the impacts of different factorial effects on the system response, a quality engineer can improve the system performance by adjusting the levels of identified factors. In this talk, I will introduce a new testing procedure for identifying active location effects from partially replicated two-level factorial designs. In addition, a two-stage procedure will be introduced for integrating the analyses of location and dispersion effects. Some numerical results will be presented to demonstrate that the proposed method is a promising alternative for practical applications. (This is a joint work with Professor Chen-Tuo Liao.)

Keywords: Generalized inference、Factorial design、Fiducial generalized pivotal quantity、Quality improvement、Screening experiment

[D-3]

Simultaneous selection of models and designs for optimal forecasting in possibly misspecified models

Hsiang-Ling Hsu[1], Mong-Na Lo Huang[2], Ching-Kang Ing[3]
[1]National University of Kaohsiung, Taiwan
[2]National Sun Yat-sen University, Taiwan
[3]National Tsing Hua University, Taiwan

Abstract

The conventional optimal design methods have the capability of determining the data points for achieving estimation or prediction efficiency in situations where the working model is correctly specified. However, it is unlikely for these data to possess the desirable optimal properties when the working model is wrong. While this dilemma can be somewhat relieved by considering a set of candidate working models and applying a model selection approach to choose the most appropriate one, it is still difficult for most practitioners to claim the true model is included among the candidate models. Hence an optimal design method that takes model misspecification into account is called for. A three-stage procedure based on the model-design selection method constructs a workable principle to tackle the concerned issue for optimal prediction. Firstly among the candidate models, it is effective in choosing the true or most approximate model by a model selection method. The design relied on the selected model is picked from the competition designs for improving the ability of accurate prediction. The selected model and optimal design aggregated information completely at the final stage make the forecasts to attain the minimum predicted loss. The advantages of the simultaneous model-design selection methodology are also revealed obtaining the better prediction in a variety of scenarios of simulation studies.

Optimal designs for binary response models with multiple nonnegative variables

Shih-Hao Huang[1], Mong-Na Lo Huang[2], and Cheng-Wei Lin[2]
[1]Department of Mathematics, National Central University
[2]Department of Applied Mathematics, National Sun Yat-sen University

Abstract

In this work, we consider optimal approximate designs for binary response models with nonnegative explanatory variables. With respect to the Schur ordering, we construct an essentially complete class consisting of designs with a simple structure. In particular, we explicitly identify locally D-optimal designs within the class for logit and probit models. When the nonnegative explanatory variables have more restrictions, such as factorial and mixture experiments, we also provide an informative iteration algorithm to search an optimal design.

Keywords: D-optimality、ϕ_p*-optimality*、*Essentially complete class*、*Logit model*、 *Probit model*、*Schur ordering*

[D-4]

Some Limit Distributions of Discounted Branching Random Walks

Jyy-I Hong
National Chengchi University

Abstract

We consider a Galton-Watson discounted branching random walk $\{Z_n, \zeta_n\}_{n\geq 0}$, where Z_n is the population of then nth generation and ζ_n is a collection of the positions on $\mathbb{R}$ of the the Z_n individuals in the nth generation, and let Y_n be the position of a randomly chosen individual from the nth generation and $Z_n(x)$ be the number of points in ζ_n that are less than or equal to x, for $x \in \mathbb{R}$. In this talk, we present the limit theorems for the distributions of Y_n and $\frac{Z_n(x)}{Z_n}$ in both supercritical and explosive cases.

Keywords: branching random walks、branching processes、coalescence supercritical、explosive

Gaussian moments conjecture and Jacobian conjecture

Shoou-Ren Hsiau
Department of Mathematics, National Changhua University of Education

Abstract

In this talk, we introduce the connection between Gaussian moments conjecture and Jacobian conjecture. Here Gaussian moments conjecture asserts that

if $X = (X_1, X_2, \dots, X_n)$ with $X_1, X_2, \dots, X_n \overset{i.i.d}{\sim} N(0,1)$, and $P(x_1, x_2, \dots, x_n) \in C[x_1, x_2, \dots, x_n]$ is a complex-valued polynomial such that the moments $E(P(X)^m)$ are equal to 0 for all $m \geq 1$, then for every polynomial $Q(x_1, x_2, \dots, x_n) \in C[x_1, x_2, \dots, x_n]$, $E\big(P(X)^m Q(X)\big) = 0$ for $m \gg 0$. On the other hand, Jacobian conjecture asserts that if $f_1, f_2, \dots f_n \in C[x_1, x_2, \dots, x_n]$, and the Jacobian determinant of $f_1, f_2, \dots f_n$, $J(f_1, f_2, \dots f_n) = 1$, then $C[f_1, f_2, \dots f_n] = C[x_1, x_2, \dots, x_n]$. Recently, Derksen, van den Essen, and Zhao (2017) proved that if Gaussian moments conjecture holds, then Jacobian conjecture holds. It seems interesting to connect a conjecture in probability with a conjecture in algebra.

[D-4]

Evaluate the rate of convergence of Markov chain Monte Carlo by expected commute time

Ting-Li Chen
Institute of Statistical Sciences, Academia Sinica

Abstract

There are many criteria to evaluate the performance of Markov chain Monte Carlo, such as spectral gap, asymptotic variance, mixing time, large deviations, and geometric ergodicity. In this talk, I will introduce another criterion, expected commute time. First, I will discuss the relation between the asymptotic variance and the expected commute time. Based on this criterion, I will present the optimal transition matrix. I will also revisit the minimization problem of the optimal transition matrix with respect to the averaged asymptotic variance (Chen et al. 2012), and give a much shorter proof via the expected commute time.

Keywords: Markov chain Monte Carlo、commute time、asymptotical variance、optimal transition matrix

不動產在土壤液化區之研究— 以臺北市及高雄市為例

林真真、許雅娟
銘傳大學應用統計與資料科學學系

摘要

臺灣地區時常因地震影響，許多房屋倒塌，釀造不少災情，土壤液化議題頻頻受到大眾關注。政府為預防及減少災害所帶來的損失，設計了臺灣土壤液化潛勢圖，作為國土計畫、都市防災之用，並於 2016 年公開此項土壤液化資訊。本研究係探討土壤液化資訊揭露，對於不動產成交價之影響。資料來源為經濟部中央地質調查所公布之土壤液化經緯度資料，以及內政部不動產交易實價查詢服務網，數據期間為土壤液化資訊公開之前後一年。本研究，將採用 R 語言套件解讀 JSON 格式，並利用文字探勘技術，獲得土壤液化 潛勢區在各行政區之分布情形，並加入房屋的各種屬性（總面積、住宅類別、移轉層次、 房廳衛浴數、總樓層數等），建立特徵價格法之迴歸模型，並比較土壤液化資訊公開前後之差異。

關鍵字：土壤液化、房價、 JSON 、文字探勘、特徵價格法

[D-5]

2008 年全球金融危機對臺灣類股影響之研究
A Study about 2008 Global Financial Crisis Influence on Taiwan Stocks

翁琬婷[1] 、周子敬[1]
銘傳大學應用統計與資料科學學系

摘要

本研究旨在進行 2008 年全球金融危機對臺灣類股影響之研究。研究之具體內容包含了解本研究選擇之臺灣類股對臺灣加權股價指數（Taiwan Capitalization Weighted Stock Index, TAIEX）之現況、分析本研究選擇之臺灣類股對 TAIEX 於金融危機發生前後之差異、驗證 2008 年全球金融危機對臺灣類股影響之關係。為達上述研究目的，本研究收集臺灣證券交易所之公開資料進行分析，使用敘述統計、因素分析、信度分析、單因子多變量變異數分析及結構方程模式。本研究期待建構出模式，提出結論與具體建議，以供未來研究之參考。

關鍵字：2008 年全球金融危機、臺灣加權股價指數、結構方程模式、SmartPLS

台灣產業價量關係隨因子變化之探討－分量迴歸分析

林妤軒、林育志
國立高雄科技大學財務管理系

摘要

本文使用 Koenker and Bassett（1978）提出的分量迴歸（Quantile Regression, QR）方法，以及利用 Fama and French（2014）五因子模型之五個因子當作控制變數，來探討台灣八大產業的價量關係，以達到成交量與報酬率的係數更準確的邊際效應。首先，全文立基於莊家彰與管中閔（2005）之實證結果，此文獻得出美國股市的報酬率和成交股數有「價量齊揚」和「價量背離」現象，並在$\theta=0.5$ 分量反轉有「對稱Ｖ字型」關係，但台灣並無此對稱現象。我們經由加入單因子、三因子和五因子當控制變數之後，得出台灣產業的價量迴歸係數（β）反轉的分量（θ）會隨著加入控制因子的多寡來增加或減少。本文發現台灣股市比重最多的電子產業的β值在不加任何因子的價量關係模型下，在$\theta=0.86$ 由低分量的正相關反轉為高分量的負相關，接著加入控制變數的模型下，單因子在$\theta=0.47$ 反轉、三因子在$\theta=0.49$ 反轉，而加入五因子的結果，是在$\theta=0.5$ 反轉，有「倒Ｖ型反轉」現象。

關鍵字：八大產業、分量迴歸、價量關係、五因子、倒Ｖ型反轉

[D-5]

A new model for dependent competing risks data in reliability

Yin Chen, Wang
Graduate Institute of Statistics, National Central University

Abstract

We consider a new model based on Copula and frailty for dependent competing risks data. The purpose of our model is making the model more widely applicable. Models are no longer limited to independent situations. A commonly used copula is in Archimedean copulas, for example (Clayton copula, Gumbel copula, …). In this paper the frailty term follow the one parameter Gamma distribution. The Frailty model also has a good description of the dependent of observations. We derive data generation methods and Kendall’ s tau under the new model.

Keywords: Joint model、Copula、Frailty、Survival Analysis

[D-5]

Review on joint frailty copula models for recurrent event times

Xinwei Huang and Prof. Dr. Takeshi Emura
Graduate Institute of Statistics, National Central University

Abstract

Joint frailty-copula models have been proposed to deal with both the dependence across different studies and the association between time of event times (e.g., cancer relapse) and time of terminal event (e.g., death). We review two existing approaches that can be also extended to recurrent event data, the one parameter gamma distribution was used as the density of frailty term to describe the dependence on patient-level, and the Clayton copula was applied for describing the association between recurrent time and terminal time. The two methods differ in terms of how the gap time was applied to terminal event. An example of colorectal cancer data was analyzed is given for illustration.

Keywords: Recurrent data、Dependent censoring、Joint model、Copula、Frailty

[D-5]

Robust ridge regression: applications to the NIKKEI NEEDS data

Ting-yu Lin and Takeshi Emura
Graduate Institute of Statistics, National Central University

Abstract

In multiple linear regression models $\boldsymbol{y} = x\boldsymbol{\beta} + \varepsilon$, where $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$ and X is the n by p design matrix and $\boldsymbol{\beta}(\beta_1, \beta_2, , \dots, \beta_P)^T$ are unknown parameters and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{o}, \sigma^2\boldsymbol{I})$, the least squares estimator $\widehat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\boldsymbol{y}$ is usually used, but this method is not suitable for the models with collinearity. Therefore, the ridge regression estimator $\widehat{\boldsymbol{\beta}} = (X^TX + \lambda\boldsymbol{I})^{-1}X^T\boldsymbol{y}$ is used here to solve this problem. In addition, data may contain outliers. We review a robust ridge regression method and apply it to the Nikkei NEEDS data [1,2] to study the relationship between y(dividend) and x_1(capital)、x_2(income)、x_3(retain).

Keywords: Ridge regression estimator、collinearity、outlier

深度學習/機器學習/統計建模在智慧工廠的應用

林文明
昱冠資訊股份有限公司

人工智慧是一個系統，他不會因為演算法存在而有用。演算法必須與系統整合，才能發揮類似人的智慧。人工智慧的核心演算法就像是人的眼睛與大腦，而系統平台就是人的軀幹與肌肉，整合在一起才能發揮智慧。本演講重點放在眼睛與大腦，並且分享昱冠資訊在智慧工廠三大領域的應用：瑕疵檢測(DIS)/安全監控(ISS)/維修預測(PMQ)。這三大領域涉及的資料科學技術包含深度學習，機器學習及統計建模。本演講會簡單介紹何謂卷積類神經網路(Convolutional neural network, CNN)及遞迴神經網路(Recurrent neural network, RNN)。而後，我們會將重點介紹實現在智慧工廠的實例。從這些例子中，我會談到機器學習與統計建模如何與深度學習結合，成為真的可以應用在工廠的技術。

[D-6]

Model selection for semiparametric marginal mean regression accounting for within-cluster subsampling variability and informative cluster size.

沈仲維[1] 、程毅豪[2]

1. Department of Mathematics, National Chung Cheng University

2. Institute of Statistical Science, Academia Sinica

Abstract

We propose a model selection criterion for semiparametric marginal mean regression based on generalized estimating equations. The work is motivated by a longitudinal study on the physical frailty outcome in the elderly, where the cluster size, that is, the number of the observed outcomes in each subject, is "informative" in the sense that it is related to the frailty outcome itself. The new proposal, called Resampling Cluster Information Criterion (RCIC), is based on the resampling idea utilized in the within-cluster resampling method (Hoffman, Sen, and Weinberg, 2001, Biometrika 88, 1121-1134) and accommodates informative cluster size. The implementation of RCIC, however, is free of performing actual resampling of the data and hence is computationally convenient. Compared with the existing model selection methods for marginal mean regression, the RCIC method incorporates an additional component accounting for variability of the model over within-cluster subsampling, and leads to remarkable improvements in selecting the correct model, regardless of whether the cluster size is informative or not. Applying the RCIC method to the longitudinal frailty study, we identify being female, old age, low income and life satisfaction, and chronic health conditions as significant risk factors for physical frailty in the elderly.

Keywords: Clustered data、Longitudinal data、Resampling、Subsampling、Variable selection

Testing Two Primary Endpoints for a Confirmatory Clinical Study

Ken-Ning Hsu
PAREXEL International

Abstract

In a clinical trial, the hypothesis for the primary variable will be defined in the protocol, that is a postulation, assumption, or statement about the target patient population regarding the efficacy of experimental drug. The statement of the hypothesis is usually a scientific question that needs to be investigated. In some confirmatory clinical trials, more than one primary variable may be used, e.g., objective response rate and overall survival for clinical trial of anticancer drugs, which will lead to the multiplicity due to testing the two hypotheses. At the design stage of clinical trial, a biostatistician will propose several test strategies for controlling the overall type I error, and work with medical experts to select one test strategy. In this talk, I am going to introduce several test strategies for one real-life confirmatory clinical trial of anticancer drug with two primary endpoints.