109年
統計學術研討會

http://www3.stat.sinica.edu.tw/csa2020

研討會主題

統計與全球健康福祉

2020 12/19 Sat.

活動地點 中央研究院 人文社會科學館
中央研究院 學術活動中心

109 年統計學術研討會會議手冊

會議日期： 2020 年 12 月 19 日 (星期六)

會議地點： 中央研究院

主辦單位： 中央研究院統計科學研究所
中國統計學社

合辦單位： 行政院主計總處
教育部
數學研究推動中心
主計協進社
中華機率統計學會

會議主席： 黃信誠教授、楊欣洲教授

籌備委員： 王婉倫教授、李百靈教授、李育杰教授、
郭美惠教授、陳春樹教授、陳瑞彬教授、
曾勝滄教授、程毅豪教授、黃文瀚教授、
黃怡婷教授、黃郁芬教授、銀慶剛教授、
盧鴻興教授、蕭守仁教授、蕭朱杏教授、
蕭金福教授 (依姓氏筆畫順序排列)

目錄

簡介

統計學術研討會為中國統計學社每年最大規模之統計學術活動，於每年 12 月由各統計學術機構輪流舉辦。2020 會議在中央研究院舉行，由中央研究院統計科學研究所與中國統計學社聯合主辦，並與行政院主計總處、教育部、科技部數學研究推動中心、主計協進社、及中華機率統計學會合辦。

會議主要包含社員大會及統計學術演講兩部分。社員大會中除將介紹中國統計學社過去一年以來業務推動情形，也將頒發學社獎項，包含終身成就獎、感謝獎、論文獎、及大學獎學金等。終身成就獎為表彰長期對我國統計事業有卓越貢獻者，論文獎頒給國內各大學統計及相關系所碩士班在學生論文優秀者，大學獎學金頒給各大學統計暨相關學系學生學業成績優秀者。

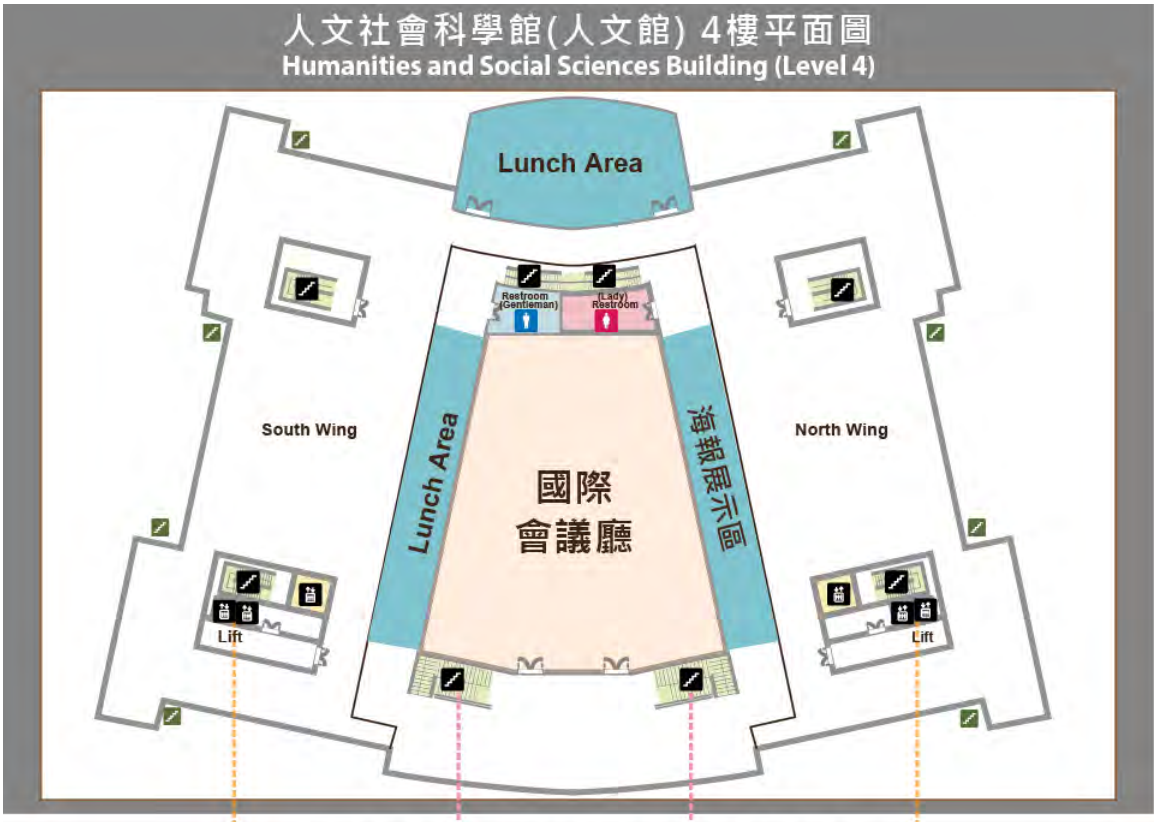
今年研討會的主題為「Advancing Statistics for Global Health and Well-being (統計與全球健康福祉)」，因此主題演講部分，將邀請陳建仁院士擔任此次大會主講人，藉此幫助統計學界了解在全球公衛議題上目前發展的趨勢與近況，也開啟雙方聯繫、合作的橋梁。除此之外，今年統計學術研討會將約有 20 場次邀請演講，邀請統計與相關領域的老師、研究人員、實務工作者，就政府統計、工業統計、生物統計、財務統計、及 COVID-19 等主題發表演講，並公開徵求海報論文發表。

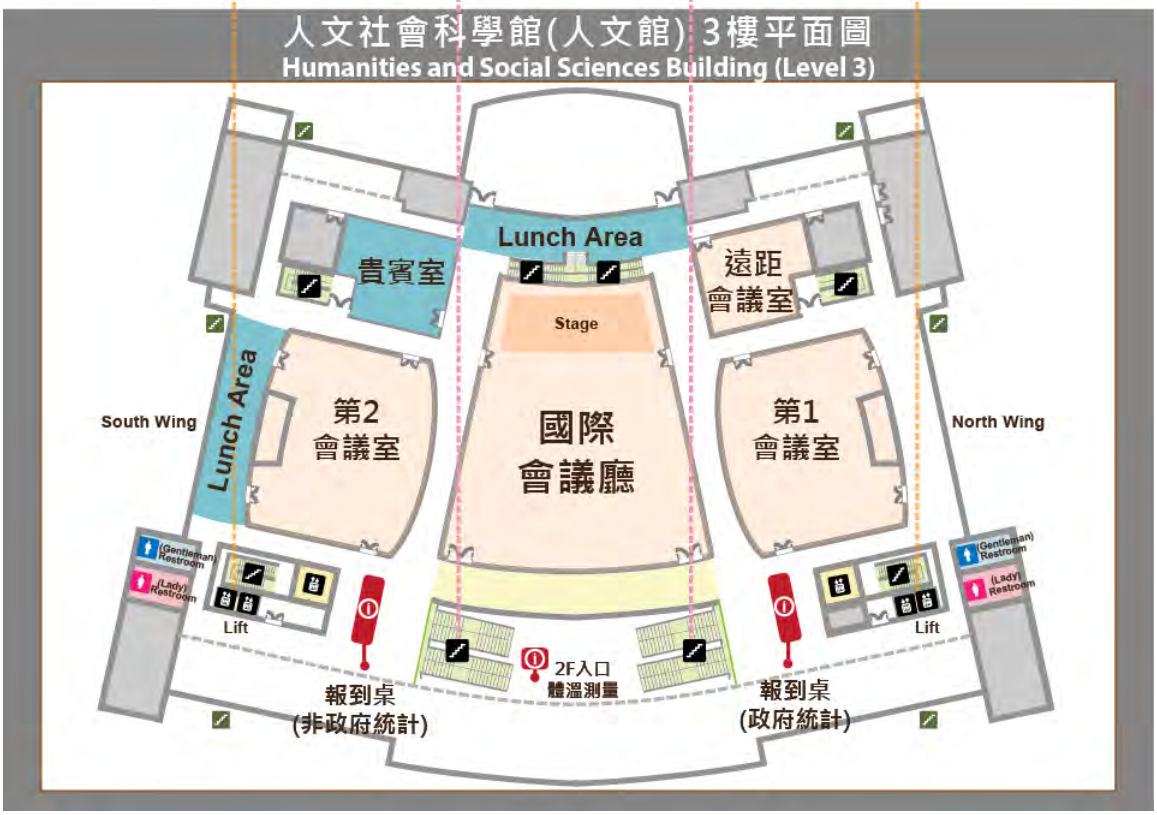
109 年統計學術研討會議程委員會

黃信誠　博士、楊欣洲　博士

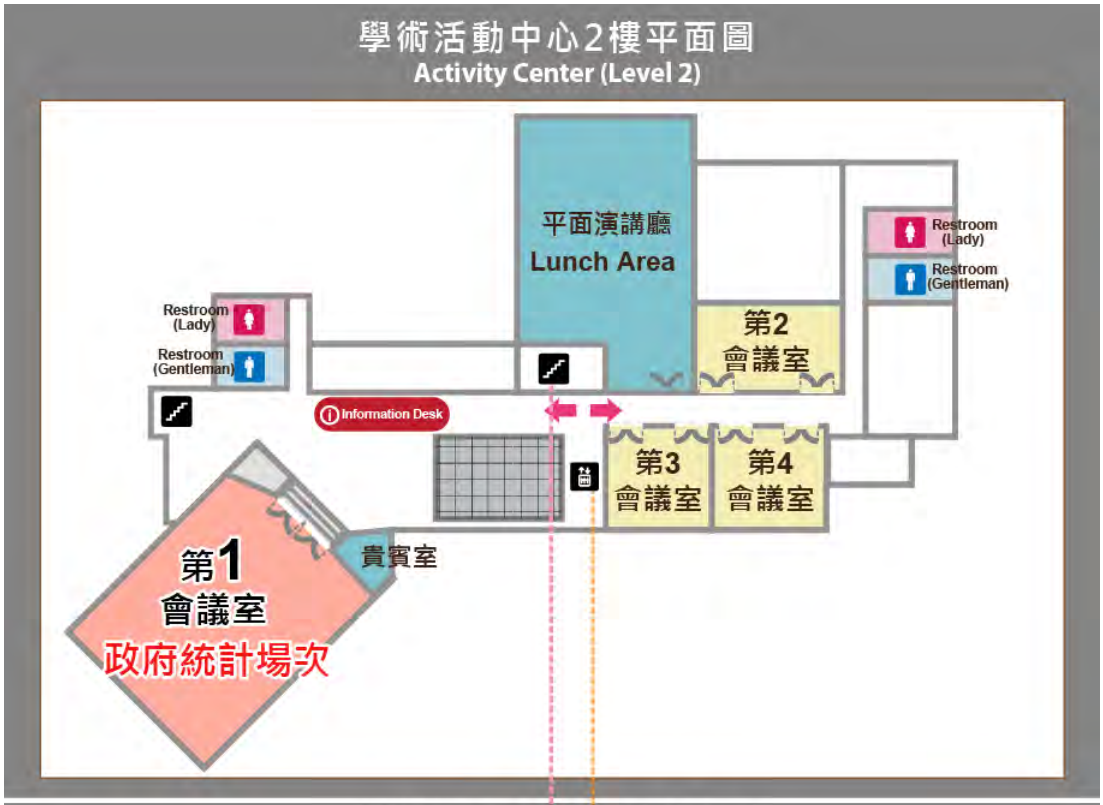
會場資訊

人文社會科學館會場圖





活動中心會場圖


學術活動中心2樓平面圖
Activity Center (Level 2)
平面演講廳
Lunch Area
第2
會議室
Restroom
(Lady)
Restroom
(Gentleman)
Information Desk
第3
會議室
第4
會議室
貴賓室
第1
會議室
政府統計場次


學術活動中心1樓平面圖
Activity Center (Level 1)
Auditorium
微風美饌
Restroom
(Lady)
Restroom
(Gentleman)
體溫測量
(政府統計)
Lobby
Activity Center
Information Desk
ATM
微風美饌
Entrance

中央研究院院區圖

01 院區大門
02 生物醫學科學研究所
03 總務處環安科
04 細胞與個體生物學研究所
05 生物多樣性研究博物館—動物標本館
06 分子生物研究所
07 生物化學研究所／生命科學圖書館
08 跨領域科技研究大樓
09 生態時代館、小森林復育區
10 院本部行政大樓（家驊館）
11 黃樓
12 植物暨微生物學研究所
13 資訊科技創新研究中心
14 永續科學中心
15 蔡元培紀念館、Trine & Zen Café
16 統計科學研究所
17 郵局、車廂、員工福利社及萊爾富便利商店
18 生態池
19 基因體研究中心
20 農業科技大樓

20 學術活動中心（四分溪書坊、大禮堂、會議室、佳福、餐廳-微風美饌）
21 中國文哲研究所
22 地球科學研究所
23 綜合體育館
24 人文社會科學館（人文館）
24 人文社會科學際合圖書館（人文館 1~2 樓）
26 植物分子育種溫室
27 溫室大樓
27 生物多樣性研究中心
28 環境變遷研究中心（環境變遷研究大樓C棟）
28 統計科學研究所（環境變遷研究大樓A棟2~6樓）
28 地球科學研究所（環境變遷研究大樓A棟7~9樓）

30 化學研究所
31 人文社會科學研究中心
31 圖書研究專題中心
32 資訊科學研究所
33 物理研究所
33 吳大猷紀念館
34 胡適紀念館
35 近代史研究所
36 歐美研究所
37 歷史語言研究所／歷史文物陳列館
38 傅斯年圖書館
39 經濟研究所
40 民族學研究所／民族學研究所博物館
41 嶺南美術館（近美大樓）
42 近史所檔案館
43 臺灣考古館

50 中央研究院宿舍群
60 中央研究院國際研究生教學行政大樓（教研大樓）
60 國際研究生學程辦公室及外籍人士服務之行政辦公室（教研大樓 2-3 樓）
60 中央研究院附設幼稚園（教研大樓 1 樓）
61 白樓
61 生物多樣性研究博物館一植物標本館（地下 1 樓）
62 國家生技研究園區
62 生醫轉譯研究中心（國家生技研究園區A棟）

會議事項

1. 注意事項：

   - 為防疫考量，敬請配合量測體溫，會場內也請配戴口罩。
   - 為尊重主講者與其他與會人員，請將手機關機或靜音。
   - 會場內禁止飲食。
   - 為避免影響會議進行，請準時入場。

2. 會場內提供無線網路：

   帳號：CSA
   密碼：20201219

3. 演講摘要請掃描 QR Code。

4. 本會議敬備午餐，請憑餐券至人文館四樓交誼廳或活動中心二樓平面演講廳領取餐點，並於會議規劃地點 (第 2、3 頁中藍色區塊 Lunch Area) 用餐：

   - 人文館：四樓交誼廳、四樓南側長廊或三樓第二會議室外用餐；
   - 活動中心：二樓平面演講廳。

5. 為響應環保，建議自行攜帶餐具與水杯。

109 年統計學術研討會議程表

<table>
<tr><td>時間</td><td colspan="5">12 月 19 日 (星期六)</td></tr>
<tr><td>08:30 ~ 09:00</td><td colspan="5">報到時間</td></tr>
<tr><td>09:00 ~ 09:10</td><td colspan="5">會議主席：黃信誠　研究員
開幕致詞 (中央研究院　廖俊智院長)
人文社會科學館　國際會議廳</td></tr>
<tr><td>09:10 ~ 09:50</td><td colspan="5">主持人：梁賡義　院士
大會主講：陳建仁　院士
風險估算公式在推動精準健康的重要性
Importance of Calculators in the Promotion of Precision Health</td></tr>
<tr><td>09:50 ~ 10:00</td><td colspan="5">合影</td></tr>
<tr><td>10:00 ~ 10:30</td><td colspan="5">社員大會</td></tr>
<tr><td>10:30 ~ 10:50</td><td colspan="5">茶點</td></tr>
<tr><td rowspan="5">10:50 ~ 12:05</td><td colspan="5">Session I</td></tr>
<tr><td colspan="4">人文社會科學館</td><td>學術活動中心</td></tr>
<tr><td>I-1 國際會議廳</td><td>I-2 第一會議室</td><td>I-3 第二會議室</td><td>I-4 遠距會議室</td><td>I-5 第一會議室</td></tr>
<tr><td>不同面向的統計分析與 COVID-19</td><td>碩士論文獎</td><td>長期資料分析</td><td>Machine Learning and Data Mining</td><td>政府統計 1</td></tr>
<tr><td>Chair：盧子彬
1. 方啓泰
2. 林煜軒
3. 張筱涵
Organizer：蕭朱杏</td><td>Chair：杜憶萍
1. 林亭妤
空間統計
Chair：陳春樹
1. 林培生
2. 張雅梅
3. 邱詠惠</td><td>Chair：王婉倫
1. 蔡秒玉
2. 沈仲維
3. 王婉倫</td><td>Chair：李百靈
1. 史玉山
2. 林松江
3. 高君豪</td><td>Chair：蔡美娜
1. 蔡宗顯
2. 劉佳佩
3. 徐　靖</td></tr>
<tr><td>12:05 ~ 13:20</td><td colspan="5">午餐及海報展示
地點：人文社會科學館 4 樓海報展示區</td></tr>
</table>

劃底線者為該場次之 Organizer。

下頁接續

接續上頁

13:20 ~ 14:35	Session II				
	人文社會科學館				學術活動中心
	II-1 國際會議廳	II-2 第一會議室	II-3 第二會議室	II-4 遠距會議室	II-5 第一會議室
	Advanced Machine Learning	資料探勘	生物統計 1	財務時間序列	政府統計 2
	Chair：李育杰 1. 李宏毅 2. 李靜沛 3. 邱維辰	Chair：黃怡婷 1. 林伯星 2. 須上英 3. 黃柏勳 4. 陳婉淑	Chair：黃郁芬 1. 吳宏達 2. 謝進見 3. 劉聚仁	Chair：郭美惠 1. 黃士峰 2. 孫立憲 3. 林良靖	李秋嬿 1. 陳雅俐 2. 呂東浩 3. 黃珮琪
14:35 ~ 14:45	休息				
14:45 ~ 16:00	Session III				
	人文社會科學館				學術活動中心
	III-1 國際會議廳	III-2 第一會議室	III-3 第二會議室	III-4 遠距會議室	III-5 第一會議室
	Recent Advances in Dimension Reduction	金融市場決策分析	生物統計 2	Industrial Statistics and Its Application	政府統計 3
	Chair：黃文璋 1. 陳定立 2. 陳素雲 3. 盧鴻興	Chair：吳牧恩 1. 顏汝芳 2. 許家豪 3. 林政憲 Organizer：黃文瀚	Chair：王秀瑛 1. 陳錦華 2. 林惠文 3. 溫啟仲 Organizer：程毅豪	Chair：陳瑞彬 1. 張明中 2. 黃世豪 3. 許湘伶	Chair：潘寧馨 1. 范汝欣 2. 陳垠伊 3. 曾仁人
16:00 ~ 16:20	茶點				
16:20 ~ 17:35	Session IV				
	人文社會科學館				學術活動中心
	IV-1 國際會議廳	IV-2 第一會議室	IV-3 第二會議室	IV-4 遠距會議室	IV-5 第一會議室
	模型選擇及高維資料分析的新發展	工業統計	Biostatistics for Public Health and Its Related Fields: In Honor of Professor Chao A. Hsiung's Retirement	應用機率	政府統計 4
	Chair：銀慶剛 1. 蔡恆修 2. 洪　弘 3. 黃學涵	Chair：羅夢娜 1. 鄭順林 2. 彭健育 3. 汪上曉 Organizer：曾勝滄	Chair：陳君厚 1. 杜憶萍 2. 蕭金福 3. 許志成 4. 熊　昭	Chair：蕭守仁 1. 符麥克 2. 黃建豪 3. 李重毅	Chair：饒志堅 1. 范宜鴻 2. 孫曉筠 3. 何宇卿
17:45 ~ 17:55	晚宴接駁				
18:10 ~ 20:30	晚宴 (By invitation only)				

劃底線者為該場次之 Organizer。

Session I

[I-1]：不同面向的統計分析與 COVID-19　　人文館-國際會議廳

Organizer：蕭朱杏 (國立臺灣大學流行病學與預防醫學研究所)

Chair：盧子彬 (國立臺灣大學流行病學與預防醫學研究所)

[I-2]：碩士論文獎 †、空間統計 ‡　　人文館-第一會議室

Organizer：陳春樹 ‡ (國立中央大學統計研究所)

Chair：杜憶萍 † (中央研究院統計科學研究所)

陳春樹 ‡ (國立中央大學統計研究所)

[I-3]：長期資料分析　　人文館-第二會議室

Organizer：王婉倫 (逢甲大學統計學系)

Chair：王婉倫 (逢甲大學統計學系)

[I-4]：Machine Learning and Data Mining　　人文館-遠距會議室

Organizer：李百靈 (淡江大學統計學系)

Chair：李百靈 (淡江大學統計學系)

[I-5]：政府統計 1　　活動中心-第一會議室

Chair：蔡美娜

10:50 ∼ 12:05

Abstract No:I-1-1

Eliminate COVID-19 without a lockdown: the Taiwan model explained

陳怡諠[1], 方啓泰[1]

[1]National Taiwan University

Abstract

Background

Never using a lockdown, Taiwan has no locally acquired cases of coronavirus disease 2019 (COVID-19) for more than 220 days. Here we report the theoretical basis behind the highly successful combination strategies used in the Taiwan model.

Methods

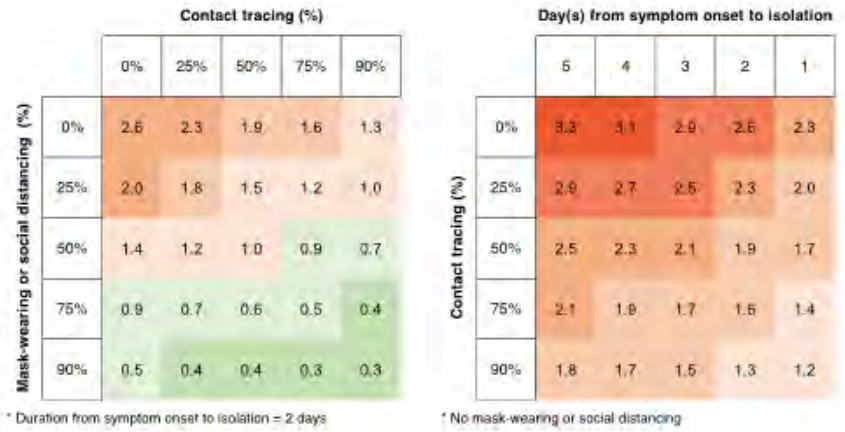
We constructed an SEIR model, with parameters based on large clinical and epidemiological studies, to assess effects of different intervention strategies, alone or in combination, to block the transmission of SARS-CoV-2 virus. Success in containment is defined as suppression of basic reproduction number (R0) to less than 1. We also modelled the medical and public health capacity required for practical implementation, under different influx rate of COVID-19 cases imported from other countries.

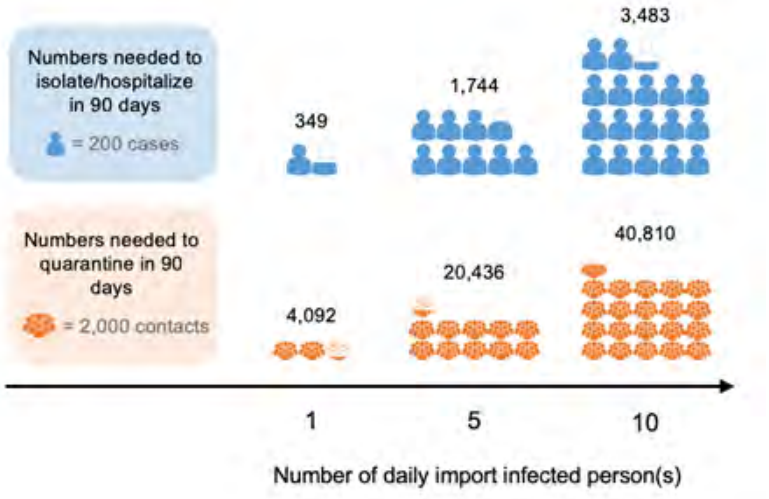
Results

The modelling results showed that a successful containment requires public surgical mask-wearing, in addition to rapid isolation and contact tracing (Figure 1). Furthermore, even with such combination interventions, a continuous influx of imported cases not detected at border will cause a proportional occurrence of secondarily transmitted local cases. Without a strict border control, when numbers of daily imported cases escaping quarantine increase from one to just ten, the pressure on medical and public health systems will rapidly escalate to a level beyond the upper limit of real-world capacity (Figure 2).

Conclusions

The two unique features of the Taiwan model, universal surgical mask-wearing and strict border quarantine, explain the successful elimination of COVID-19 in Taiwan without using lockdown. (Note: This work has been published on Journal of Formosan Medical Association)





Keyword: COVID-19, Modeling, Pandemic, Epidemic Control

Abstract No:I-1-2

快速即時調查疫情中各國群眾的心理與行為

林煜軒[1]

[1]財團法人國家衛生研究院

Abstract

我們的研究團隊在疫情爆發之初，率先採用 Google Trends 資料比較全世界先爆發疫情的 21 個國家在網路搜尋「洗手」及「口罩」的狀況，結果發現搜尋「洗手」愈多的國家，疫情爆發的速度較緩慢；而比較沒有搜尋「洗手」的國家，疫情爆發速度較快。

在三月中各國開始實施封城時，我們也發現網路搜尋「失眠」是受疫情影響心理健康最敏感的指標；而「自殺」的搜尋量通常在媒體報導後激增。。採用 Google Trends 資料，分析全世界各國疫情對心理健康的影響。利用網路搜尋來預測疫情，比傳統問卷調查更加有效率，可以搜集最即時、長期記錄的資料。

Keyword: Google Trends，洗手，口罩，失眠，自殺

Abstract No:I-1-3

Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan

張孟群[1], Rebecca Kahn[2], 李育安[1], 李政昇[1], Caroline Buckee[2], 張筱涵[1]

[1]國立清華大學, [2]Harvard T.H. Chan School of Public Health

Abstract

As COVID-19 continues to spread around the world, understanding how patterns of human mobility and connectivity affect outbreak dynamics, especially before outbreaks establish locally, is critical for informing response efforts. In Taiwan, most cases to date were imported or linked to imported cases. In collaboration with Facebook Data for Good, we characterized changes in movement patterns in Taiwan since February 2020, and built metapopulation models that incorporate human movement data to identify the high risk areas of disease spread and assess the potential effects of local travel restrictions in Taiwan. We found that mobility changed with the number of local cases in Taiwan in the past few months. For each city, we identified the most highly connected areas that may serve as sources of importation during an outbreak. We showed that the risk of an outbreak in Taiwan is enhanced if initial infections occur around holidays. Intracity travel reductions have a higher impact on the risk of an outbreak than intercity travel reductions, while intercity travel reductions can narrow the scope of the outbreak. The timing, duration, and level of travel reduction together determine the impact of travel reductions, and multiple combinations of these can result in similar impact. We developed an interactive application that allows users to vary inputs and assumptions and shows the spatial spread of the disease. Our results can be used readily if local transmission occurs in Taiwan after relaxation of border control, providing important insights into future disease surveillance and policies for travel restrictions.

Keyword: COVID-19, metapopulation model, mobility data, travel restrictions

Abstract No:I-2

Mixtures of t Factor Analyzers with Censored Data

林亭妤[1]

[1]國立中興大學

Abstract

混合 t 因子分析器已經被運用來處理具有複雜特徵的高維度資料並進行穩健性分群，此模型被認為是一種有效的建模工具。然而，在某些實際情況下，資料可能含有設限值，導致傳統方法在計算上變得複雜，甚至不可行。本文在 MFAC 架構下提出一個可以處理設限資料的擴展，簡稱 MtFAC。在參數估計方面，我們發展兩種交換性的期望條件最大化演算法(AECM)，其中，E 步驟須依賴截切多變量 t 分配的第一、二階動差。對於模型參數的混合比例與成份的平均向量的估計值之漸近標準差是由訊息遺失原理之方法或稱“路易斯”方法所得出。我們執行一些模擬實驗來檢驗最大概似估計量在有限樣本下的性質以及所提出模型在處理設限值和離群效應方面的性能。最後，我們應用一組具有實際設限值的真實資料來闡述所提出方法之有效性和實用性。

Keyword: AECM 演算法, 設限資料, 因子分析, 最大概似估計, 遺失訊息原則, 截切多變量 t 分佈

Abstract No:I-2-1

A Heterogeneity Measure for Cluster Identification with Application to Disease Mapping

林培生 [1]

[1] 國家衛生研究院

Abstract

Mapping of disease incidence has long been of importance to epidemiology and public health. In this paper, we consider identification of clusters of spatial units with elevated disease rates and develop a new approach that estimates the relative disease risk in association with potential risk factors and simultaneously identifies clusters corresponding to elevated risks. A heterogeneity measure is proposed to enable the comparison of a candidate cluster and its complement under a pair of complementary models. A quasi-likelihood procedure is developed for estimating the model parameters and identifying the clusters. An advantage of our approach over traditional spatial clustering methods is the identification of clusters that can have arbitrary shapes due to abrupt or non-contiguous changes while accounting for risk factors and spatial correlation. Asymptotic properties of the proposed methodology are established and a simulation study shows empirically sound finite-sample properties. The mapping and clustering of enterovirus 71 infection in Taiwan are carried out for illustration.

Keyword: clustering analysis, estimating equations, non-proximity cluster, spatial statistics

Abstract No:I-2-2

半參數空間模型於所得分配不均資料之應用

張雅梅[1]，黃佩新[1]

[1]淡江大學

Abstract

本研究提出一個非平穩半參數空間模型(non-stationary semi-parametric spatial model)來描述所得分配不均於空間上的相依性。該模型為數個基底函數(basis function)及平穩過程(stationary process)的線性組合，由於此模型有大量參數需要估計，我們使用 Tibshirani(1996)提出的最小絕對壓縮與篩選運算法(least absolute shrinkage and selection operator, lasso)進行參數估計，該方法可以同時估計參數及作變數選取。本研究將估計結果繪製成空間分佈圖，透過空間分佈圖來描述歐洲地區所得分配不均資料在空間上的分佈情形。根據研究結果顯示，波羅地海三小國:愛沙尼亞(Estonia)、拉脫維亞(Latvia)及立陶宛(Lithuania)的變異程度較大;所得分配不均於愛沙尼亞(Estonia)和瑞典(Sweden)附近有較高的相依性，在德國(Germany)、英國(UK)及西班牙(Spain)附近相依性較低。

Keyword: 所得分配不均，非平穩空間模型，最小絕對壓縮與篩選運算法，最小角迴歸法

Abstract No:I-2-3

Modeling Nonstationary Covariance Using a Linear Combination of Stationary Processes with Varying Coefficients

邱詠惠[1], 黃信誠[2], 陳春樹[1]

[1] 國立中央大學, [2] 中央研究院

Abstract

Spatial data over a large domain generally shows nonstationary features. However, how to appropriately specify a nonstationary covariance function is difficult and the computation of the corresponding inverse matrix in kriging is also thorny, especially when the data set is massive. In this paper, we develop a methodology for modeling nonstationary process based on a linear combination of stationary processes with spatially varying coefficients. A likelihood-based method, called iterative marginal optimization, is developed for efficiently estimating model parameters. Numerical results show that the proposed methodology provides a satisfactory approximation for the underlying covariance structure and the spatial prediction is also compatible.

Keyword: Cross-covariance, Nonstationarity, Parameter estimation, Positive semidefinite matrix, Spatial prediction

Abstract No:I-3-1

New model-averaged estimators of concordance correlation coefficients for longitudinal overdispersed Poisson data

蔡秒玉 [1]

[1]National Changhua University of Education

Abstract

The concordance correlation coefficient (CCC) can be used to assess agreement among multiple observers for continuous and discrete responses. We consider not only subject, observer and time effects, but also interaction effects in extended overdispersed three-way Poisson mixed-effects models for longitudinal overdispersed Poisson data. To avoid fitting data with a misspecified model, thus yielding biased CCC estimates, this research proposes new model-averaged estimators of CCC by combining the estimators of the variance components (VC) approach with model selection via corrected conditional Akaike information criterion (CAICC) and corrected conditional Bayesian information criterion (CBICC) measures under extended overdispersed three-way Poisson mixed-effects models. Simulation studies are conducted to compare the performance of VC with and without model selection via CAICC and CBICC and the new model-averaged approach for longitudinal Poisson and overdispersed Poisson data sets. An application of corticospinal diffusion tensor tractography study is presented for illustration. It can be concluded that the new model-averaged approach is a reliable procedure yields small mean square errors and nominal 95% coverage rates.

Keyword: CAICC, Concordance correlation coefficient, Model averaging, Poisson mixed-effects model, Variance components

Abstract No:I-3-2

Model selection based on resampling approaches for cluster longitudinal data with missingness in outcomes

沈仲維[1]

[1]國立中正大學

Abstract

In medical and health studies, longitudinal and cluster longitudinal data are often collected, where the response variable of interest is observed repeatedly over time and along with a set of covariates. Model selection becomes an active research topic but has not been explored largely due to the complex correlation structure of the data set. To address this important issue, in this paper, we concentrate on model selection of cluster longitudinal data especially when data are subject to missingness. Motivated from the expected weighted quadratic loss of a given model, data perturbation and bootstrapping methods are used to estimate the loss and then the model that has the smallest expected loss is selected as the best model. To justify the proposed model selection method, we provide various numerical assessments and a real application regarding the asthma data set is also analyzed for illustration.

Keyword: bootstrap, data perturbation, generalized estimating equations, variable selection

Abstract No:I-3-3

Clustering multiple longitudinal data via finite mixtures of multivariate t nonlinear mixed model

王婉倫[1]

[1]逢甲大學 (Feng Chia University)

Abstract

The multivariate *t* nonlinear mixed-effects model (MtNLMM) has been shown to be effective for analyzing multi-outcome longitudinal data following nonlinear growth patterns with fat-tailed noises or potential outliers. This paper considers the problem of clustering heterogeneous longitudinal profiles in a mixture framework of MtNLMM. A finite mixture of multivariate *t* nonlinear mixed model is proposed, and this new model allows accommodating complex features of longitudinal data. Intermittent missing values frequently occur in the data collection process of multiple repeated measures. Under a missing at random mechanism, a pseudo-data version of the alternating expectation conditional maximization (AECM) algorithm is developed to carry out maximum likelihood estimation and impute missing values simultaneously. The techniques for clustering of incomplete multiple trajectories, recovery of missing responses, and estimation of random effects are also provided. The utility of the proposed methods is illustrated through a simulation study and a real-data example coming from a pregnant women study.

Keyword: Heterogeneity, Missing data, Mixture models, Multiple nonlinear profiles, Outliers

Abstract No:I-4-1

Regression Trees for Multivariate Count Data

史玉山[1]

[1]中正大學

Abstract

We propose a regression tree method for multiresponse count data. At each node, a univariate count regression model is fitted and its associated partial score residuals are obtained. We then use the conditional independence tests of analysis of contingency tables of the residual patterns to select the split variable. A likelihood approach is applied to the selection of the split point at each node and to the tree pruning process. Our tree method is shown to be free from selection bias. Its advantage in prediction is demonstrated through some simulations and a real data analysis on the demand for medical care.

Keyword: Conditional independence test , Hurdle model, Partial Score residual, Zero-inflated model

Abstract No:I-4-2

Multi-label Classification based on GLMM

林松江 [1]

[1] 國立臺北教育大學

Abstract

Recently, more and more researchers focus on solving multi-label classification problems. Traditional approaches aim to transform the original multi-label classification problem into several single-label classification problems and then assemble information from single-label classifier to predict multi-label results. However, these strategies will the loss of information among labels resulting in damaging the classification performance. In other words, the performance of the algorithms may be improved through using the information from the association among labels, since labels may not be independent but may be correlated. In this study, we proposed a novel hybrid label-based meta-learning algorithm for multi-label classification based on an ensemble of a cluster algorithm and generalized linear mixed model (GLMM). The numerical results show that the proposed algorithm outperforms others, especially for cases with relatively large number of labels.

Keyword: Clustering, generalized linear mixed model (GLMM), meta-learning, multi-label classification

Abstract No:I-4-3

高維度象徵性資料之地圖視覺化

高君豪[1]

[1]淡江大學

Abstract

由於資料時代的來臨，資料產生與蒐集方式變得更加容易，使得資料規模急速成長。而巨量資料之資料分析與運算為目前重要之趨勢，其中象徵性資料分析(Symbolic Data Analysis)為巨量資料分析之一種重要之統計分析方法。另外地理資料之 EDA 圖形化工具，於現今資料科學中已為一經常使用之地理相關資料呈現方式，但主要還是呈現以單一變量之數值大小為主。透過色階對應其資料數值，將地理資訊以其對應色階中之顏色加以呈現。然而當象徵性資料中，若含有地理資訊時，其地圖視覺化便無法輕易實現。

本研究針對高維度象徵性資料之地圖視覺化提出一新的呈現方式。透過單一顏色呈現各個地理樣本於高維度變數中之相關性，樣本顏色的相似程度呈現高維度象徵性資料之綜觀表徵。並搭配象徵性矩陣資料視覺化方法及叢聚分析，探索樣本與變數及兩者之間的關係。最後以 Covid-19 於疫情期間各國每日染疫人數為實際範例，呈現高維度象徵性資料之地圖視覺化之分析結果。

Keyword: 象徵性資料分析，地圖視覺化，矩陣視覺化，巨量資料

Abstract No:I-5-1

我國出口領先廠商在總體經濟及稅收貢獻之探討

蔡宗顯[1]

[1]財政部統計處

Abstract

我國經濟成長動能高度仰賴出口，隨著近年美中貿易紛擾，催化全球供應鏈之重組，以及 COVID-19 疫情衝擊全球貿易，在在考驗我國出口廠商之應變能力。本文擇取 108 年排名前 20 大出口廠商，觀察其出、進口貨品與市場結構，並透過公開資訊財報資料計算營收與獲利情況，探討其在國內創造之附加價值、就業機會、對國內整體經濟之貢獻比重。另串接營利事業所得稅、綜合所得稅、營業稅等 11 項稅目資料，進而瞭解出口領先廠商對政府財源挹注之影響程度。

Keyword: 出口領先廠商, 總體經濟, 稅收貢獻

Abstract No:I-5-2

從外銷訂單看我國廠商海外生產布局變動

劉佳佩[1]

[1]經濟部統計處

Abstract

隨全球化的浪潮，各國透過國際分工獲取經濟利益，我國亦在此波浪潮下，透過三角貿易分工模式，逐步建立起全球性的分工體系。惟近年為因應美中貿易紛爭，企業開始移動生產基地，加上今年受 COVID-19 疫情衝擊，更加速全球供應鏈重組步伐，這波供應鏈縮短、破碎化的趨勢，勢將影響我國產業結構及相關產業鏈的全球布局。

本文利用外銷訂單海外生產實況調查，探究近年廠商海外生產變動情形，分別從廠商海外生產之模式、生產地之配置及生產考量之因素等面向剖析，再就各貨品、各規模別進行交叉分析，並針對我國在中國大陸及香港擁有生產線之廠商，觀察其生產線新增或移轉情形，藉以分析美中貿易戰對我國在中國大陸及香港之生產配置挪移走向，俾利掌握我國台商生產布局及經營策略之轉變，亦為本部擬訂或調整產業政策之重要參據，發揮統計支援決策功能。

Keyword: 海外生產, 外銷訂單

Abstract No:I-5-3

運用追蹤資料觀察轉職對低薪族群薪資之影響

徐靖[1]

[1]行政院主計總處國勢普查處

Abstract

薪資成長為我國各界所關注之重要議題，其中低薪現象及青年待遇問題尤甚，惟目前研究鮮少利用追蹤資料觀察低薪族群、青年族群之薪資現況及趨勢。本研究連結財稅資料建置 101 年至 107 年追蹤資料，觀察低薪族群及 23-29 歲青年低薪族群之薪資變化，進行同一族群跨時期研究。為降低外生衝擊之影響、抽樣誤差及調查偏誤，應用差異中之差異法（Difference in difference, DID）分析，發現轉職 1 至 2 次對於低薪族群及青年低薪族群皆有提升薪資之效果，並以轉職 1 次提升效果較為顯著；若進一步以固定效果模型（Fixed effect model）分析，則可發現在控制人口特徵、廠商規模、地區、行業等變數後，不論對於哪一類族群來說，轉職次數持續增加將使薪資負成長。

Keyword: 低薪, 青年族群, 追蹤資料, 差異中之差異法, 固定效果模型

Session II

[II-1]：Advanced Machine Learning　　人文館-國際會議廳

Organizer：李育杰 (國立交通大學應用數學系)

Chair：李育杰 (國立交通大學應用數學系)

[II-2]：資料探勘　　人文館-第一會議室

Organizer：黃怡婷 (國立臺北大學統計學系)

Chair：黃怡婷 (國立臺北大學統計學系)

[II-3]：生物統計 1　　人文館-第二會議室

Organizer：黃郁芬 (國立成功大學數學系)

Chair：黃郁芬 (國立成功大學數學系)

[II-4]：財務時間序列　　人文館-遠距會議室

Organizer：郭美惠 (國立中山大學應用數學系)

Chair：郭美惠 (國立中山大學應用數學系)

[II-5]：政府統計 2　　活動中心-第一會議室

Chair：李秋嬿

13:20 ~ 14:35

Abstract No:II-1-1

機器能否無師自通學習人類語言？

李宏毅[1]

[1] 台灣大學

Abstract

隨著深度學習技術突飛猛進，AI 的能力越來越強，但多數時候其學習仍需要人類對資料進行適當的標註才能學習，例如：今日要訓練語音辨識系統不只是需要蒐集大量的語音訊號，還要對這些語音訊號進行標註，告訴機器每一段語音訊號對應的文字，機器才能學習，因此僅有具備大量標註資料的語言，才能取得較高的辨識正確率。但在這個大數據的時代，要蒐集大量未標註資料是相對容易的，因此如何使用這些未標註資料，已經成為關鍵的研究議題。本演講將探討一系列減少標註資料量的語音和自然語言處理相關技術，包含以生成式對抗網路利用未成對資料、以自督導式學習利用未標註資料，以及元學習。

Keyword: Self-supervised learning, Meta learning, Generative Adversarial Network

Abstract No:II-1-2

Manifold identification for ultimately communication-efficient distributed optimization

LEE Ching-pei[1], LI Yu-Sheng[2], CHIANG Wei-Lin[3]

[1]Academia Sinica, [2]National Taiwan University, [3]University of California Berlekey

Abstract

This work proposes a progressive manifold identification approach for distributed optimization with sound theoretical justifications to greatly reduce both the rounds of communication and the bytes communicated per round for partly-smooth regularized problems such as the ℓ_1- and group-LASSO-regularized ones. Our two-stage method first uses an inexact proximal quasi-Newton method to iteratively identify a sequence of low-dimensional manifolds in which the final solution would lie, and restricts the model update within the current manifold to gradually lower the order of the per-round communication cost from the problem dimension to the dimension of the manifold that contains a solution and makes the problem within it smooth. After identifying this manifold, we take superlinear-convergent truncated semismooth Newton steps computed by preconditioned conjugate gradient to largely reduce the communication rounds by improving the convergence rate from the existing linear or sublinear ones to a superlinear rate. Experiments show that our method can be orders of magnitudes lower in the communication cost and an order of magnitude faster in the running time than the state of the art. This is a joint work with LI Yu-Sheng and CHIANG Wei-Lin, and the paper is published in the proceedings of the international conference of machine learning, 2020.

Keyword: Communication-efficient distributed optimization, Partly smooth functions, Manifold identification, Nonsmooth optimization

Abstract No:II-1-3

Demystifying the Black Box of Deep Neural Networks: Showcase on Medical Image Translation

邱維辰 [1]

[1]National Chiao Tung University

Abstract

Recent advances in deep learning have brought the magic leap to various research areas. Especially, with the growth of medical needs in the aging society, the integration of deep learning techniques into medical applications for assisting in the medical data analysis or imaging is one of the most important topics. Medical imaging, as the most prominent approach to the acquisition of spatially-resolved information of organs and tissues in vivo, has attracted great attempts to develop various deep-learning-based toolkits on processing the medical images. Particularly, the image-to-image translation across different medical image domains could provide additional diagnostic scans and benefit the clinical decision thus being an active research problem. However, the deep-learning-based translation models in themselves are inscrutable black boxes. We therefore propose to demystify the image translation process from a medical point of view, where we particularly focus on the translation from T1-MR images to PET ones, via adopting the representational similarity analysis. We discover that the process of T1-MR to PET image translation includes stages of recognition on various brain parts. Based on our findings, we build up an explainable model to demonstrate the capability of deep learning models for extracting medically meaningful information of brain imaging data, which untangles the biological plausibility hidden in deep learning models.

Keyword: Deep Learning, Image-to-Image Translation, Medical Images

Abstract No:II-2-1

Smart Medicine

林伯星 [1]

[1] 國立臺北大學

Abstract

WHO defines "eHealth is the use of information and communication technologies (ICT) for health. The eHealth unit works with partners at the global, regional and country level to promote and strengthen the use of ICT in health development, from applications in the field to global governance. The unit is based in the Department of Service Delivery and Safety in the Cluster of Health Systems and Innovation". In recent years, the rapid evolution of artificial intelligence (AI) has accelerated the development of smart medicine. The speaker, Prof. Lin Bor-Shing, will first briefly introduce the concepts of AI, including convolution neural networks (CNN), recurrent neural network (RNN), supervised learning, unsupervised learning, reinforcement learning, and generative adversarial network (GAN). Next, he will take the research results of his own intelligent assistive devices and AI diagnosis in rehabilitation projects as examples, to let the audience have a more practical understanding of intelligent medical care and future developments.

Keyword: eHealth, smart medicine, artificial intelligence, convolution neural network, recurrent neural network

Abstract No:II-2-2

Categorical Data Embedding

須上英[1], 陳定立[2]

[1] 台北大學, [2] 中央研究院

Abstract

Categorical data characterize samples by "categories", not by "quantities". Such a characterization makes it difficult for analysts to explore categorical data. In this talk, we will propose a method for embedding categorical data into the Euclidean space, so that the relationships between categories and between samples can be directly exhibited. The method is to find the subspace that can mostly discriminate between categories. The subspace can be considered as the underlying feature space behind the categorical observations. Then the samples are assigned according to the closest positions to their categories in the feature spaces. The effectiveness of this method will be evaluated by simulations and a clustering analysis.

Keyword: categorical data, embedding, discriminant analysis

Abstract No:II-2-3

機器學習在房顫術後復發預測的應用與比較

黃柏勳[1], 黃怡婷[1], 周宗川[2], 李慧玲[3]

[1]國立臺北大學, [2]林口長庚紀念醫院, [3]臺北長庚紀念醫院

Abstract

心房顫動(AF)可經由心導管射頻燒灼術(RFCA)來治療, 但仍有部分患者會復發, 患者需以斑點追蹤心臟超音波圖(STE)來追蹤, 並由心臟科醫生來判讀結果。STE 還會額外提供彎曲 M-mode 整體應變圖(GS)與對應的應變率圖(GSR), 但文獻尚未有使用該資料來協助臨床判讀術後狀態。

本研究會使用 GS 和 GSR 的彩色心臟超音波圖, 綜合 MPCA、SMOTE 演算法, 並以多變量統計方法、集成式學習模型與卷積神經網路來預測術後狀態, 最後以準確率、特異度等作為評估準則, 篩選最適方法, 期望提升臨床醫生對 RFCA 術後狀態的預測能力。

Keyword: 心導管射頻燒灼術, 斑點追蹤心臟超音波圖, 機器學習, 卷積神經網路

Abstract No:II-3-1

Exploring Spatial-Temporal Clustering, Aggregation, and Transmission of Avian Influenza Virus in Taiwan

吳宏達[1]

[1] 中興大學

Abstract

Background: Avian flu outbreak in poultry farms exhibited spatial clustering, which was conveniently explored by plotting the incident outbreaks and theoretically tested by *spatial scan statistics*.

Objectives: To explore the *second-order spatial clustering* according to various levels of resolutions, and estimate the transmission direction through a conventional *Standard deviational Elliptic* (SDE) method.

Methods: Likelihood-ratio statistics were constructed at different levels of spatial resolution with 'element areas' being a fine partition of townships decided by the smallest administrative units. Different 'null' spaces used as reference populations were compared to form an aggregation map (called *AGC map*) depicted to reveal the secondary aggregation of geo-clustering of outbreaks. A *Knox statistic* was used to decide, *a priori*, suitable partitions which formed independent clusters for subsequent explorations for spatial transmission.

Results: The AGC map offers second-order aggregation of disease spatial clustering, which also implies transmission directions. Knox statistic is useful for making partitions on time axis and on spatial expansion. Transmission direction can be explored by simply connecting the centers decided by the SDE method, contrasted with the *time-dependent* AGC map.

Keyword: Avian flu outbreak, spatial scan statistic, aggregation, transmission, Knox statistic

Abstract No:II-3-2

Dependence model selection for semi-competing risks data

謝進見 [1]

[1] 中正大學

Abstract

In this talk, we consider the model selection problem of the dependency between the terminal event and the non-terminal event under semi-competing risks data. When the relationship between the two events is unspecified, the inference on the non-terminal event is not identifiable. We can not make inference on the non-terminal event without extra assumptions. Thus, an association model for semi-competing risks data is necessary, and it is important to select an appropriate dependence model for a data set. We construct the likelihood function for semi-competing risks data to select an appropriate dependence model. From simulation studies, it shows the performance of the proposed approach is well. Finally, we apply our method to a bone marrow transplant data set.

Keyword: Copula model, Likelihood function, Model selection, Semi-competing risks data

Abstract No:II-3-3

Large Scale Assessment of Consistency in Sleep Stage Scoring Rules among Multiple Sleep Centers Using an Interpretable Machine Learning Algorithm

劉聚仁[1], 林定佑[2], 吳浩榳[3], 許元春[4], 劉景隆[5], 劉文德[6], 楊美貞[7], 倪永倫[8], 周昆達[9], 羅友倫[2]

[1]國立成功大學數學系, [2]林口長庚紀念醫院, [3]杜克大學, [4]國立交通大學, [5]台北馬偕紀念醫院, [6]雙和醫院睡眠中心, [7]台北慈濟醫院, [8]台中慈濟醫院, [9]臺北榮民總醫院

Abstract

STUDY OBJECTIVES: Polysomnography is the gold standard in identifying sleep stages; however, there are discrepancies in how technicians use the standards. Because organizing meetings to evaluate this discrepancy and/or reach a consensus among multiple sleep centers is time consuming, we developed an artificial intelligence (AI) system to efficiently evaluate the reliability and consistency of sleep scoring, and hence the sleep center quality.

METHODS: An interpretable machine learning algorithm was used to evaluate interrater reliability (IRR) of sleep stage annotation among sleep centers. The AI system was trained to learn raters from one hospital, and applied to subjects from the same or other hospitals. The results were compared with the experts' annotation to determine IRR. Intra-center and intercenter assessments were conducted on 679 subjects without sleep apnea from six sleep centers in Taiwan. Centers with potential quality issues were identified by the estimated IRR.

RESULTS: In the intra-center assessment, the median accuracy ranged from 80.3% to 83.3% with the exception of one hospital (designated E) with an accuracy of 72.3%. In the inter-center assessment, the median accuracy ranged from 75.7% to 83.3% when hospital E was excluded from testing and training. The performance of

the proposed method was higher for N2, awake, and REM, compared to N1 and N3. The significant IRR discrepancy of hospital E suggested a quality issue. This quality issue is confirmed by the physicians in charge of hospital E.

CONCLUSIONS: The proposed AI system proved effective in assessing IRR and hence the sleep center quality.

Keyword: interrater reliability, intra-center assessments, inter-center assessments, machine learning, sleep stage scoring

Abstract No:II-4-1

A Network Autoregressive Model with GARCH Effects and it's Applications

黃士峰[1], 江欣翰[1], 林昱君[1]

[1]National University of Kaohsiung

Abstract

This study proposes a network autoregressive model with GARCH effects, denoted by NAR-GARCH, to depict the return dynamics of stock market indices. The GARCH effects of each index are deleted marginally and the NAR model with the Granger causality test is used to capture the joint effects caused by other indices with the most updated market information. The proposed model has fewer parameters and is more flexible in changing model sizes than classical vector autoregressive models. The returns of 20 global stock indices from 2006 to 2020 are employed for our empirical investigation. The numerical results reveal that the NAR-GARCH model has satisfactory performances in both fitting and prediction for the 20 stock indices, especially when a market has strong upward or downward movements.

Keyword: GARCH, Granger causality test, network autoregressive model

Abstract No:II-4-2

Mean Field Games with Heterogeneous Groups: Application to Banking Systems

孫立憲 [1]

[1] 國立中央大學

Abstract

We study the system of heterogeneous lending and borrowing based on the relative average of log-capitalization given by the linear combination of the average within groups and the ensemble average and describe the evolution of log-capitalization by a system of coupled diffusions. The existence of closed- and open-loop Nash equilibria for the two-group case is guaranteed by the solvability for the coupled Riccati equations. Both equilibria consist of the mean-reverting term identical to the homogeneous game and all group averages owing to heterogeneity. The comparison of the obtained open-and closed loop Nash equilibria is also discussed. Finally, in financial implications, we study the influence of the incentive and relative parameters and also the number of banks on the corresponding liquidity rates through numerical analysis.

Keyword: Systemic risk, inter-bank borrowing and lending system, heterogeneous group , Nash equilibrium, Mean Field Game

Abstract No:II-4-3

Symbolic interval-valued data analysis for time series based on auto-interval-regressive models

林良靖[1], 簡湘霖[1], 宋豪[1], Sangyeol Lee[2]

[1] 國立成功大學, [2]Seoul National University

Abstract

This study considers interval-valued time series data. To characterize such data, we propose an auto-interval-regressive (AIR) model using the order statistics from normal distributions. Furthermore, to better capture the heteroscedasticity in volatility, we design a heteroscedastic volatility AIR (HVAIR) model. We derive the likelihood functions of the AIR and HVAIR models to obtain the maximum likelihood estimator. Monte Carlo simulations are then conducted to evaluate our methods of estimation and confirm their validity. Real data examples from the S&P 500 Index and PM2.5 are used to demonstrate our method.This study considers interval-valued time series data. To characterize such data, we propose an auto-interval-regressive (AIR) model using the order statistics from normal distributions. Furthermore, to better capture the heteroscedasticity in volatility, we design a heteroscedastic volatility AIR (HVAIR) model. We derive the likelihood functions of the AIR and HVAIR models to obtain the maximum likelihood estimator. Monte Carlo simulations are then conducted to evaluate our methods of estimation and confirm their validity. Real data examples from the S&P 500 Index and PM2.5 are used to demonstrate our method.

Keyword: AIR model, HVAIR model, Interval-valued time series, Order statistics, Symbolic data analysis

Abstract No:II-5-1

死因統計國際化新里程

陳雅俐 [1]

[1] 衛生福利部統計處

Abstract

死因統計係由死亡證明書上多項疾病因果關係所編製而成，為實證醫學最重要醫療成效評估來源，亦是衛福部網站關鍵字熱搜第一名。國際上利用疾病因果關係溯源至原死因主流系統有二，一為美國 ACME，一為歐盟 IRIS。近年美國因預算逐年緊縮，不再維護及更新 ACME，我國原採用版本為 2000 年版，版本過舊，故參考德、法等 20 餘國擬改採 IRIS。

為接軌國際潮流，提高國際資料可比較性，以利於融入 2022 年 WHO 即將發布之 ICD-11，爰進行 IRIS 試算及轉換比較，並邀集資料使用機關（單位）評估改版影響與因應措施。2020 年 6 月，完成死因判別系統由 ACME 改版 IRIS 作業，並協同機關及業務單位召開「2019 年國人死因統計結果」記者會，說明系統轉換重大訊息及發布變革結果，為亞洲第一個完成改版國家。

Keyword: 死因統計, 死亡證明書, 疾病因果關係, 原死因, ICD-11

Abstract No:II-5-2

桃園市一般生育率影響因素之探討

呂東浩[1]

[1]桃園市政府主計處

Abstract

臺灣生育率全球第 3 低、年輕人不婚不生是國安危機均為經常出現於報章媒體的標題，持續低迷的生育率亦為政府及民眾關心的議題，惟低生育率的成因眾說紛紜，低薪高房價導致適婚年齡男女不敢婚育，似乎為當前社會的主流意見，惟從經濟分析的角度出發，女性教育程度的提升，以及工商社會日益發達，女性工作機會的增加，這些現象均使生兒育女的機會成本上升，連帶降低育齡婦女生育的意願，2 種不同的說法似乎都言之成理。

爰此，為探求少子化現象的真正成因，本文運用 91 至 108 年桃園市的一般生育率資料，結合育齡婦女有偶占比、女性勞動力參與率及房地產可能成交價指數等多個總體變數，建立迴歸模型，量化上開變數對一般生育率之影響，藉以釐清低生育率現象背後的因果關係，期能作為政府部門施政決策之參考。

Keyword: 一般生育率, 迴歸

Abstract No:II-5-3

由健保就診統計談高雄市民健康維護情形

黃珮琪 [1]

[1] 高雄市政府主計處

Abstract

近年高雄市民平均壽命逐年增加，健保醫療費用核付金額亦逐年遞增，且 45 歲以上中高齡者健保醫療費用占全市比重超過七成二，隨著人口老化對醫療資源需求是市府社會福利及醫療衛生政策重要施政主軸，及公私部門醫療保健支出持續攀升，市府亟需剖析探討整體全面性疾病就醫資訊，爰以透過全民健保醫療統計資料觀察本市市民健保就診情形變動，探究市民就醫維護健康情形，並就本市市民就診率及醫療費用支出之變動趨勢、性別、年齡與疾病別差異分析，以提供做為市府規劃衛生醫療及預防保健等相關政策參據。

Keyword: 全民健保醫療統計, 健保就診率, 健保醫療費用

Session III

[III-1]：Recent Advances in Dimension Reduction　　人文館-國際會議廳

Organizer：盧鴻興 (國立交通大學統計學研究所)

Chair：黃文璋 (國立高雄大學統計學研究所)

[III-2]：金融市場決策分析　　人文館-第一會議室

Organizer：黃文瀚 (國立中興大學統計學研究所)

Chair：吳牧恩 (臺北科技大學資訊與財金管理系)

[III-3]：生物統計 2　　人文館-第二會議室

Organizer：程毅豪 (中央研究院統計科學研究所)

Chair：王秀瑛 (國立交通大學統計學研究所)

[III-4]：Industrial Statistics and Its Application　　人文館-遠距會議室

Organizer：陳瑞彬 (國立成功大學統計學系)

Chair：陳瑞彬 (國立成功大學統計學系)

[III-5]：政府統計 3　　活動中心-第一會議室

Chair：潘寧馨

14:45 ~ 16:00

Abstract No:III-1-1

Dimension reduction for mixed-type data

陳定立[1], 須上英[2]

[1] 中央研究院, [2] 台北大學

Abstract

Dimension reduction is often an important data pre-processing procedure before further analysis. In this talk, we will present a new approach for dimension reduction. We propose to assign the data in a lower dimensional space so that some pre-defined measurement error is minimized. It can be shown that this approach is equivalent to Principal Component Analysis (PCA), one of the most popular dimension reduction methods, for continuous type of data, while this approach can be directly generalized to dimension reduction for categorical data as well as mixed-type data. A simulation study will be given to demonstrate the result of this dimension reduction method.

Keyword: Dimension reduction, pricipal component analysis, caetegorical data

Abstract No:III-1-2

On asymptotic normality of cross data matrix-based PCA in high dimension low sample size

陳素雲[1], 王紹宣[2], 陳定立[3]

[1]Academia Sinica, [2]National Central University, [3] Academia Sinica

Abstract

Principal component analysis in high dimension low sample size setting has been an active research area in recent years. Yata and Aoshima (2010) proposed a cross data matrix-based method and showed the asymptotic normality for estimates of spiked eigenvalues and also consistency for corresponding estimates of PC directions. However, the asymptotic normality for estimates of PC directions is still lacking. In this article, we have extended Yata and Aoshima (2010)'s work to include the investigation of the asymptotic normality for the leading CDM-based PC directions and to compare it with the asymptotic normality for the classical PCA. Numerical examples are provided to illustrate the asymptotic normality.

Keyword: principal component analysis , high dimension, cross data matrix, asymptotic normality

Abstract No:III-1-3

Toward Image Understanding by Deep Learning and Influential Scores

Hung-Hsun Chen[1], Ssu-Ting Fang[1], Yi-Chu Li[1], Henry Horng-Shing Lu[1], Shaw-Hwa Lo[2]

[1]National Chiao Tung University, [2]Columbia University

Abstract

Explainable AI (XAI) is considered highly important due to its ability to explain the high predictive accuracy that is inherent in AI models. Images commonly have background characteristics and contain noise, which may be mistaken for prediction criteria. Consequently, explaining and improving the prediction ability of a deep learning model is a challenging task requiring an innovative approach. This study proposes the simple and feasible approach by influential score (I-Score) for deep learning to eliminate the effect of background information and noise on image classification. Hence, a reliable prediction model can be obtained from foreground features via deep learning. The performance is evaluated by simulation and empirical studies reported in this study.

Keyword: Deep Learning, Influential Score (I-Score) , Image Understanding, Explainable AI (XAI)

Abstract No:III-2-1

The Effect of Institutional Ownership on Firm Policies

洪茂蔚[1], 顏汝芳[2]

[1]國立臺灣大學, [2]國立臺北大學

Abstract

This paper provides evidence on the causal relationship of institutional ownership on corporate behavior and performance. Our analyses use the annual reconstitution of the Russell 1000 and 2000 indexes as a source of exogenous variation in institutional ownership to study the effect of institutional ownership on firm outcomes. The characteristics of firms near the Russell 1000/2000 threshold are similar, except that firms just include in the top of the Russell 2000 index have discontinuously higher institutional ownership, predominantly stemming from quasi-indexing institutions. Using this discontinuity, we examine whether additions to the Russell 2000 affect firm policies by implementing fuzzy regression discontinuity (RD) designs.

Keyword: Institutional ownership, Russell indexes, Regression discontinuity, Selection bias

Abstract No:III-2-2

An association rule-based trading framework with Kelly criterion

許家豪[1], 吳牧恩[2]

[1] 國立臺灣大學, [2] 國立臺北科技大學

Abstract

Technical analysis aims to utilize price and volume information to generate trading signals, which is suitable for association rule mining. However, the signals from association rules may contain uncertainty, and lack the portfolio management. We propose an association rule-based trading framework with Kelly criterion for portfolio management. Experimental results show that the framework provides valuable pattern with confidence of 63.5% and support of 15.9%. For trading performance, the framework obtain a shape ratio of 1.946, and the Kelly criterion increase the Sharpe ratio and win ratio by 0.3 and 3.5%, and reduces the drawdown risk by 80% during the training period. In the testing period, the framework obtain Sharpe ratio and win rate of 5.659 and 81.250%, and the Kelly criterion decrease the drawdown risk to 26%. In summary, the proposed framework extract valuable patterns through association rule mining, and the Kelly criterion for portfolio management can stabilize performance.

Keyword: association rule, trading framework, Kelly criterion, stop-loss, take-profit

Abstract No:III-2-3

交易機器人在期貨市場的應用

林政憲[1]

[1]拓思資訊股份有限公司

Abstract

因應程式交易風氣日益興盛，機器人操盤時代來臨，怎樣在期貨交易上導入自動化交易系統，透過定義交易的需求、收集與分析資料，應用適當的統計分析方法，尋找商品資訊的規律性與顯著性的條件，歸納後建構成一個系統模型，再進行回測與動態部位管理，逐步構建適合自身交易條件的策略，搭配資金管理演算法、實時反饋的智能化交易系統優化模組等技術，逐步完善期貨交易流程。

Keyword: 程式交易，期貨

Abstract No:III-3-1

Application of Deep Learning-based Methods in Tuberculosis Disease Classification

陳錦華[1], 陳芬芳[1], 李枝新[2]

[1]台北醫學大學, [2]市立萬芳醫院

Abstract

Tuberculosis (TB) is an infectious disease that causes the illness and death in millions of people for each year worldwide. Patients with cavitary TB are more contagious, and chest X-rays could help to diagnose cavities. This pilot study aims to conduct the performances of TB types classification on chest X-rays using deep learning methods. The dataset consisted of 270 chest X-rays totally, single image per patient, selected based on CT scan date. Three different seeds were setting to sample the chest X-rays and lung regions images into 50% for training, 25% for validation and 25% for testing. The chest X -rays and lung regions images extracted by an U-Net model separately were used to train the model in experiment 1 and 2. In experiment one, a CNN model was trained by chest X-rays. Also, in experiment two, an U-Net model was used to extract lung regions images to train a CNN model. The VGG-19 pre-trained models were used to classify chest X-rays and lung regions images to assist diagnosis of TB. The best area AUC values of the VGG-19 pre-trained model training by the first and the second group of lung regions were 0.7045 and 0.6719 on the test sets, respectively. The best AUC values of the fine-tuning VGG-19 pre-trained model trained by the third group and the first group were 0.6413 and 0.6621 on the test sets, respectively. In the future, the capabilities of the model can be further verified by increasing the number of chest X-rays.

Keyword: Tuberculosis, CNN, AUC, chest X-ray, cavity

Abstract No:III-3-2

Utilize external big data to improved two-stage design method on zero-inflated model

林惠文 [1]

[1] 東吳大學

Abstract

Analysis of big data using population-based medical and epidemiological studies is growing in popularity. However, health care databases or administrative data often lack personal information, and these missing variables are often important factors for studies. In addition, these databases are often involving random events with excessive zero count data. For these reasons, we propose a two-stage calibration zero-inflated method to handle excessive zero count data and missing confounding factors, and utilize external large data such as health care databases and administrative data to enhance testing power. Our method does not require model assumptions to still obtain robust estimates. In the simulation study, our method is unbiased and can effectively improve the test power by 30%. Finally, we illustrate an application of our method to explore the risk of Outpatient visit by using the National Health Interview Survey database.

Keyword: missing confounding, zero-inflated mode, two-stage model

Abstract No:III-3-3

Discrete time-to-event data with longitudinal covariates

温啟仲[1], 程毅豪[2]

[1]Tamkang University, [2]Academia Sinica

Abstract

Survival analysis has been conventionally performed on a continuous time scale. In practice, the survival time is often recorded or handled on a discrete scale; when this is the case, the discrete-time survival analysis would provide analysis results more relevant to the actual data scale. Besides, data on time-dependent covariates in the survival analysis are usually collected through intermittent follow-ups, resulting in the missing and mis-measured covariate data. In this work, we propose the sufficient discrete hazard (SDH) approach to discrete-time survival analysis with longitudinal covariates that are subject to missingness and mismeasurement. The SDH method employs the conditional score idea available for dealing with mis-measured covariates, and the penalized least squares for estimating the missing covariate value using the regression spline basis. The SDH method is developed for the single event analysis with the logistic discrete hazard model, and for the competing risks analysis with the multinomial logit model. Simulation results revel good finite-sample performances of the proposed estimator and the associated asymptotic theory. The proposed SDH method is applied to the scleroderma lung study data, where the time to medication withdrawal and time to death were recorded discretely in months, for illustration

Keyword: Competing risks, Measurement error, Right censored data, Survival analysis

Abstract No:III-4-1

M-aberration: a unified approach to select factorial designs

張明中 [1]

[1] 國立中央大學

Abstract

Minimum aberration has been ubiquitously adopted for selecting fractional factorial designs. Much work has been done on its various extensions, from which many fields of experimental design have benefited, including multi-stratum designs, multi-group designs, and multi-platform designs. However, most of these extensions are ad hoc and are developed on case-by-case bases without strong statistical justifications and a unified rationale. In this talk, we develop a new criterion referred to as M-aberration, derived from good statistical properties. Our theory not only features a unified framework for minimum aberration and is easily applied to many situations, but also enables experimenters to derive their own aberration criteria.

Keyword: Multi-stratum experiment, Nonregular design, Supersaturated design, Optimal design, Functional prior distribution

Abstract No:III-4-2

Testing Independence Between Two Spatial Random Fields

黃世豪[1]

[1]國立中央大學

Abstract

We consider testing independence between two spatial Gaussian random fields evaluated respectively at p and q locations with sample size n, where p and q are allowed to be larger than n. Our approach is based on canonical correlation analysis (CCA), without imposing any spatial stationarity and parametric structure for the two random fields. Instead of applying CCA directly to the two random fields, which is not feasible for high-dimensional testing considered, we adopt a dimension-reduction approach using a special class of multiresolution spline basis functions. These functions are ordered in terms of their degrees of smoothness. By projecting the data to the function space spanned by a few leading basis functions, the spatial variation of the data can be effectively preserved. The test statistic is constructed from the first sample canonical correlation coefficient in the projected space and is shown to have an asymptotic Tracy-Widom distribution under the null hypothesis. Our proposed method automatically detects the signal between the two random fields and is designed to handle irregularly spaced data directly. In addition, we show that our test is consistent under mild conditions and provide simulation experiments to demonstrate its powers. Moreover, we apply our method to investigate whether the precipitation in continental east Africa is related to the sea surface temperature (SST) in the Indian Ocean, and whether the precipitation in west Australia is related to the SST in the North Atlantic Ocean. (Work done jointly with H.-C. Huang, R. S. Tsay, and G. Pan.)

Keyword: canonical correlation analysis, high-dimensional test, irregularly spaced data, teleconnection, Tracy-Widom distribution

Abstract No:III-4-3

Optimum Designs for Parameter Estimation in a Multi-response Mixture Experiment

許湘伶 [1]

[1] 國立高雄大學

Abstract

A mixture experiment in the (q-1)-dimensional probability simplex is an experiment in which the q factors are non-negative and subject to the simplex restriction, which means the sum of all factors is equal to one. In this talk, we investigate the issue of the optimal designs for parameter estimation with the considered k responses models, consisted of the Scheffé's mixture polynomial models. Initially, we characterize the structure of candidate designs based on the complete classes of the weighted centroid designs for the considered multi-response mixture experimental models with the given covariance structure. According to the well-known equivalence theorem, we demonstrate that the obtained allocation measures at the support points are the D-optimal designs for particular multi-response mixture models. Specifically, some results of the D-optimal designs in multi-response scenarios are demonstrated to be independent of the covariance structure between the k responses, but depend on the allocation of the underlying polynomial models.

Keyword: Complete class, Design optimality, Kiefer ordering, Weighted centroid design

Abstract No:III-5-1

臺北市土壤液化潛勢區建築物、人口概況

土木建築科曾俊傑、統計室范汝欣[1]

[1]臺北市政府工務局

Abstract

臺北市政府工務局先後於民國 105 年、107 年公布 2 次土壤液化潛勢圖資，方便民眾查詢住家附近土壤液化資訊，並作為都市規劃和防災參考。本報告利用土壤液化潛勢圖資，結合戶籍人口數資料、建築物使用執照資料，計算各行政里高、中、低土壤液化潛勢範圍之建築物數量與屋齡，以及人口年齡結構；透過集群分析將各里分群，分析各群液化潛勢區建築物、人口概況，並進行土壤液化災害防救風險程度評估。

臺北市約有 21%面積屬於中、高度土壤液化潛勢區域，65%棟建築物座落於其中。經集群分析發現，各群大致呈層層向外擴散現象，其中評估土壤液化災害風險程度較高的前 20 個行政里，多位於民生東路和信義路區間，屬捷運線密集的高度發展區域。建議相關防災規劃併同集群分析結果，將人口及屋齡因素納入考量。

Keyword: 土壤液化

Abstract No:III-5-2

未來 20 年各類榮民人數與服務量影響推估

陳垠伊 [1]

[1] 國軍退除役官兵輔導委員會統計資訊處

Abstract

退除役官兵輔導工作是國家永續的責任，本會遵循憲法對軍人「崇功報勳」精神及政府照顧退除役官兵旨意，持續提供最妥適的安置與服務照顧。惟隨著資深榮民（民國 23 年以前出生者）逐漸凋零及遷臺後新增榮民年齡亦漸長，致近年來亡故人數遠超過新進人數，每年平均約減少 1.1 萬人，由於榮民總人數於 108 年底已低於 35 萬人，未來若榮民人數持續大量減少，勢必對本會各項施政措施規劃及資源有效分配造成極大影響。

為了解未來榮民人數可能演變情形，本研究採用與國發會人口預測相同之年輪法，參考應用近年榮民死亡率及新增人數變化數據、原因，除滾動式推估未來 20 年各年榮民總人數外，亦同時推估不同年齡別、地區別以及重要屬性（身心障礙、中低收入、特較需求及失智）之榮民人數；另外並利用現有員工、訪視人員人數，推算本會未來服務量變化情形，提供本會各業務主管單位及所屬單位妥善因應與規劃運用參考。

Keyword: 榮民人數推估, 年輪法

Abstract No:III-5-3

以分段迴歸及時間序列分析優化大專新生數預測模型

曾仁人[1]

[1]教育部統計處

Abstract

各教育階段學生人數預測攸關教育資源管控及招生名額調整等重要政策之擘劃與推展，其中大專校院一年級學生人數預測結果，因近年少子女化趨勢致生源減少，復以大專轉型退場議題，更為各界所關注。因現行推估方式之理論基礎改進空間仍存，爰導入分段式迴歸（segmented regression）及時間序列分析（ARIMA）統計方法，優化推估模型，提升預測確度，並進而建立逐年滾動修正之標準作業流程，俾充分支援決策需求。

Keyword: 大專新生數, 預測, 分段迴歸分析, 時間序列分析

Session IV

[IV-1]：模型選取及高維資料分析的新發展　　人文館-國際會議廳

Organizer：銀慶剛 (國立清華大學統計學研究所)

Chair：銀慶剛 (國立清華大學統計學研究所)

[IV-2]：工業統計　　人文館-第一會議室

Organizer：曾勝滄 (國立清華大學統計學研究所)

Chair：羅夢娜 (國立中山大學應用數學系)

[IV-3]：Biostatistics for Public Health and Its Related Fields: In Honor of Professor Chao A. Hsiung's Retirement　　人文館-第二會議室

Organizer：杜憶萍 (中央研究院統計科學研究所)

蕭金福 (國家衛生研究院群體健康科學研究所)

Chair：陳君厚 (中央研究院統計科學研究所)

[IV-4]：應用機率　　人文館-遠距會議室

Organizer：蕭守仁 (國立彰化師範大學數學系)

Chair：蕭守仁 (國立彰化師範大學數學系)

[IV-5]：政府統計 4　　活動中心-第一會議室

Chair：饒志堅

16:20 ~ 17:35

Abstract No:IV-1-1

Model Averaging for High-dimensional Linear Regression Models with Dependent Observations

蔡恆修[1], 銀慶剛[2], 余定宏[3]

[1]Academia Sinica, [2]National Tsing Hua University, [3] University of Iowa

Abstract

We introduce the orthogonal greedy algorithm (OGA) to screen out the nested set of signal variables under a high-dimensional linear regression framework with dependent observations. To gain the prediction performance, we propose the high-dimensional Mallow model averaging (HDMMA) criteria to determine the weight for averaging these nested high-dimensional linear regression models. We further analyze rates of convergence of prediction error for the averaging model under different sparsity conditions. Our contribution has three folds. First, we show that our procedure, named OGA+HDMMA, can achieve optimal convergence rates of prediction error discussed in Ing (2019). Second, we use simulation to show that the out-sample prediction of OGA+HDMMA can outperform the MCV method proposed in Ando and Li (2014) when the covariates are highly correlated or contain time-series effects. Third, the out-sample prediction of OGA+HDMMA performs comparably or even better than many well-known high-dimensional variables selection methods in some scenarios.

Keyword: High-dimensional Mallow model averaging, orthogonal greedy algorithm, sparsity conditions, high-dimensional linear regression models, optimal rate of convergence

Abstract No:IV-1-2

A generalized information criterion for high-dimensional PCA rank selection

洪弘[1], 陳素雲[2], 銀慶剛[3]

[1]台灣大學, [2]中央研究院, [3]清華大學

Abstract

Principal component analysis (PCA) is a commonly used statistical procedure for dimension reduction. An important issue for PCA is to determine the rank, which is the number of dominant eigenvalues of the covariance matrix. Among information-based criteria, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two most common ones. Both use the number of free parameters for assessing model complexity, which may suffer the problem of model misspecification. To alleviate this difficulty, we propose using the generalized information criterion (GIC) for PCA rank selection. The resulting GIC model complexity takes into account the sizes of eigenvalues and, hence, is more robust to model misspecification. The asymptotic properties and selection consistency of GIC are derived under the high-dimensional setting. Compared to AIC and BIC, the proposed GIC is better capable than AIC in excluding noise eigenvalues, and is more sensitive than BIC in detecting signal eigenvalues. Moreover, we discuss an application of GIC to selecting the number of factors for factor analysis. Our numerical study reveals that GIC compares favorably to the methods based on (deterministic) parallel analysis.

Keyword: GIC, high-dimensionality, model selection under misspecification, parallel analysis, PCA

Abstract No:IV-1-3

Sparse Matrix Estimation Based on Greedy Algorithms and Information Criteria

黃學涵[1], 銀慶剛[1], 蔡瑞胸[2]

[1]國立清華大學,[2] 芝加哥大學

Abstract

We consider the problem of estimating the covariance matrix of serially correlated vectors whose dimension is allowed to be much larger than the sample size. We propose using the orthogonal greedy algorithm (OGA) together with a high-dimensional Akaike's information criterion (HDAIC) to estimate the matrix, and show that the proposed estimate achieves the optimal rate under a sparsity condition more flexible than those in the existing literature. When the covariance matrix is bandable, we use the banding and tapering estimates instead and propose the first information criterion for choosing the banding and tapering parameters that can attain the optimal rate.

Keyword: Covariance estimation, High-dimension, Greedy algorithm, Information criterion, Time series

Abstract No:IV-2-1

Defect Pattern Clustering and New Pattern Classification of Wafer Bin Maps by Subset Learning with Convolutional Neural Networks

鄭順林[1]，張尹甄[1]

[1] 國立成功大學

Abstract

A wafer bin map (WBM) is the result of a circuit probe test (CP test) on a wafer after the completion of the manufacturing process in semiconductor industry. The specific defect patterns on WBMs provide crucial information for engineers to trace the failure causes in the complicated manufacturing process. Many research for WBM image recognition were done using statistical and deep learning methods. The statistical methods are often followed by additional image transformation processes. As for the deep learning method, they often focus on dealing the classification task with many labelled data. Their approach is difficult to apply for the real applications because usually only few labeled data are available.

In this study, a small subset of data with labels is used to train convolutional neural networks (CNN) models and extract features from the models without additional transformations on WBMs. We then take the features to cluster the WBMs. We also use the CNN models to separate the new pattern from the typical defect patterns by utilizing the help of untypical defect patterns from other WBM products. The evaluation of our procedure is conducted with simulation data.

Keyword: Wafer Bin Map, Clustering, Classification, Image Recognition, Convolution Neural Network

Abstract No:IV-2-2

Optimal Test Plan for Degradation Tests Based on Student-*t* Processes

彭健育[1]

[1] 中央研究院

Abstract

Stochastic processes are broadly used to assess reliability of high quality products, and the Wiener process is a particularly common one in degradation analysis. As an extension of the Wiener process, the Student-*t* process possesses the applicability and flexibility of degradation data with the heavy-tailed characteristic. This study discusses the D-optimal test plan (such as the inspection frequency, the number of measurements, and sample size) for a degradation test based on a Student-t process. Under a total experimental cost constraint, the D-optimal test plan not only utilizes to control experimental test time and cost efficiently but also increases the precision of the parameter estimators. The resistance data is presented as an illustrative example and the corresponding sensitivity analysis indicates robustness of the D-optimal test plan as the deviation of the estimated values is within 90% confidence interval of the model parameters.

Keyword: Conjugate distribution, Cauchy process, First passage time, Volterra integral equation

Abstract No:IV-2-3

Monitoring of a trajectory in industrial grade transition

汪上曉[1], 鄭英[2], Zhao Jing Wang[2]

[1] 國立清華大學, [2] 華中科技大學

Abstract

Many continuous industrial processes will operate in different steady states that produce products with different grades or event different products. The switching between two steady states is called transition. Transition consists of a series of operation changes that should be carry out in proper order, within certain magnitudes and time window. Faulty operation may lead to increase in inferior products, or more importantly, hazard events. Monitoring of the transition process is desired. In this work, a transition identification and monitoring scheme is proposed based on slow feature analysis (SFA). Two monitoring statistics which represent the location of the trajectory and the speed of transition are proposed. Using a numerical example, and the mode 4-to-2 transition of the Tennesse-Eastman process which exhibits catastrophic failure, operating faults were generated based on the guidewords of HAZOP. In addition to missed detection rate and false alarm rate, an early detection performance index was introduced. The advantages of proposed method were benchmarked against a stage-based principle component analysis approach using these indices. transition monitoring, operating faults, transition identication, trajectory-based method

Keyword: transition monitoring, operating faults

Abstract No:IV-3-1

Statistical Genetics and 3D Protein Structures

杜憶萍 [1]

[1] 中研院

Abstract

In this talk, I will share the story of studying statistical genetics to pursue my Ph.D. degree and my journey to study 3D protein structures. Some review of genetic linkage analysis, its mathematical modeling and sequential analysis will be provided. For the past decade, my research has focused on the image analysis of cryo-electronic microscopy (cryo-EM). Cryo-EM has become a powerful technique to solve the 3D structures of macro protein particles with high efficiency to provide crucial medical insight for the pharmaceutical industry. However, the data characteristics include strong noise, huge dimension, large sample size and high heterogeneity with unknown orientations have made analysis very challenging. A couple of self-developed algorithms employing statistical insights to face these challenges will be presented.

Keyword: statistical genetics, cryo-electronic microscopy (cryo-EM), Protein Structures

Abstract No:IV-3-2

Statistical Issues on the Conduct of Clinical Trials During the COVID-19 Pandemic

蕭金福[1], 陳啟天[2], 姜杰[1], 陳祈安[1]

[1]National Health Research Institutes, [2]StatPlus Inc

Abstract

The COVID-19 outbreak is impacting clinical trials in many ways, such as patient recruitment, data collection and data analysis. To proceed in this difficult time, the adoption of new technologies and new approaches for conducting clinical trials needs to be accelerated. Simultaneously, regulatory agencies such as the US FDA and EMA have issued guidance to help the pharmaceutical industry conduct clinical trials of medical products during the COVID-19 pandemic. In this talk, we will address some statistical issues in the conduction of clinical trials during the COVID-19 pandemic. Specifically, statistical issues related to protocol modifications caused by COVID-19 will be raised.

Keyword: COVID-19, protocol modifications, heterogeneity

Abstract No:IV-3-3

Big data research and policy translation

許志成[1]

[1]NHRI

Abstract

台灣的糖尿病發生率過去十年來增加 11%，盛行率增加了近五成，歷年來台灣透析的發生率與盛行率也都名列世界前茅。近十年糖尿病與慢性腎臟病防治工作最重要的項目，首推整合照護方案，包含:糖尿病共同照護網，糖尿病醫療給付改善方案，初期慢性腎臟病醫療給付改善方案(early-CKD 計畫)，末期腎臟病前期之病人照護與衛教計畫(pre-ESRD 計畫)等。對糖尿病及慢性腎臟病患病人而言，加入整合照護方案可以獲得減少醫療花費、降低死亡率，以及降低或延緩透析的發生率等的正向效益，但從文獻上也發現，病情較為嚴重、年齡較大及經濟較為弱勢者，常常會被排除在整合照護方案之外，且目前這些整合照護方案的收案人數也有待加強。這一類全國性之衛生政策實務與其他臨床監測與應用必須依賴大數據分析，才能確實掌握全貌與趨勢變化。本次演講，我將說明如何以健保資料庫來做臨床暨醫療政策之轉譯研究並藉此促進跨領域的合作。

Keyword: chronic kidney disease, pre-ESRD program, big data research

Abstract No:IV-3-4

Risk prediction models in cancer—statistical learning viewpoints

熊昭 [1]

[1]National Health Research Institutes

Abstract

In this talk, I will compare three statistical methods to build absolute risk models for predicting cancer occurrence using three examples. BCRAT, the well-known Gail model predicting breast cancer occurrence, is constructed using nested case-control dataset embedded in and composite hazard ratio from a large cohort and forms the framework for the development of competing risk models for predicting cancer occurrence. PLCO m2012, an absolute risk model predicting lung cancer in the next 6 years for an ever-smoker, is a logistic regression model constructed using all the ever-smokers in the control arm of the PLCO trial. TNSF-SQ, an absolute risk model constructed by our group predicting lung cancer in the next 6 years for a never-smoking female in Taiwan, is a logistic regression model constructed using the case-control components of GELAC , Taiwan biobank served as part of the controls and age-specific population incidence rates derived from *Cancer Registry* . Bias-variance tradeoff will also be discussed. Other risk models will be mentioned.

Keyword: absolute risk, bias-variance tradeoff, cancer occurrence, risk prediction, statistical learning

Abstract No:IV-4-1

Asymptotic normality for the size of graph tries built from M-regular tree labelings

符麥克 [1]

[1] 國立政治大學

Abstract

Graph tries (G-tries), a generalization of classical tries, have been proposed in 2014 by P. Jacquet who considered mean and variance of the size of G-tries and conjectured a normal limiting distribution, as the number of independent label functions approaches infinity. This is one of the rare examples of a shape parameter of a random structure with variance considerably larger than the mean, but one nevertheless expects that a central limit theorem holds. In this talk, we will explain how to use the method of moments to prove this conjecture. The talk is based on joint work with my postdoc Tsan-Cheng Yu (National Chengchi University).

Keyword: probabilistic analysis of algorithms, tries, shape parameters, limit laws, method of moments

Abstract No:IV-4-2

One-dimensional polymers in random environments

黃建豪[1]

[1]國立政治大學

Abstract

We put the range of one-dimensional simple random walk in the disordered environments in the Gibbs' setting. The walk affects by forces such as random potentials and external fields. Those effects may push the walk away (super-diffusive) or pull the walk around the origin (sub-diffusive). Under a suitable scaling of those forces with heavy-tailed distributions of random potentials, one can see a rich phase diagram. Especially, for the typical case which researchers are interested in, we get the end-to-end fluctuation exponent is 2/3. This is a joint work with Q. Berger, N. Torri and R. Wei. For details, please see arXiv: 2002.06899.

Keyword: Random polymer, Range, Heavy-tailed distributions

Abstract No:IV-4-3

An elementary approach to the asymptotic normality of the Stirling numbers

李重毅[1]

[1]Academia Sinica

Abstract

In this study, we mention an elementary approach to display the asymptotic normality of the form of the inclusion-exclusion. One of the most representative cases is the Stirling numbers of the second kind, which is used in numerous fields such as the combination, probability, and even statistics connecting with the left-truncated generalized Poisson distribution. It has many great properties so that it was carefully studied with various technical approaches to the estimates of the value, the estimate of the summation, i.e. Bell number, and the asymptotic behavior of the local limit theorem, etc. The interesting thing is that the elementary approach we mention is enough to cover all of the above results. We thus apply this approach to other general forms of the inclusion-exclusion.

Keyword: Stirling numbers of the second kind, asymptotic estimate, Bell number, local limit theorem, inclusion-exclusion

Abstract No:IV-5-1

防制酒駕大數據分析

范宜鴻[1]

[1] 內政部警政署統計室

Abstract

酒駕零容忍為政府至高原則。為遏止酒駕發生，除透過多次修法提高罰則，最有效方式即增加酒駕勤務（提高見警率）。分析發現102 年修法後，酒駕初犯已逐年減少，而酒駕累犯並沒有明顯減少，顯示酒駕者仍存僥倖心態，惟警力有限，若大規模投入酒駕勤務，將對其他勤務產生排擠效應。因此，在警力不變的情況下，如何選擇適合路檢點提高勤務效率為當前重要的議題。

警察在執行酒駕勤務（設置路檢點或機動巡邏），主要依過去曾發生酒駕事故地點、時間，及考量當地的交通道路特性。為提升酒駕攔檢效率，創建「NPA酒駕支援勤務平台」，除整合酒駕事故及酒駕違規資料外，平台加入了新的元素-「酒駕違規者發生事故位址」，擴大可能發生酒駕預警熱點，再配合員警酒駕執法經驗，聚焦勤務熱區或查察地點，提升執法效率。開發多元互動式酒駕事故及違規儀表板，視覺化呈現酒駕事故斑點圖、熱力圖及酒駕熱時，並以最小統計區為統計單元導入酒駕熱區斑點圖，作為酒駕勤務規劃之參考。

Keyword: 酒駕, 路檢點, 斑點圖, 勤務平台

Abstract No:IV-5-2

由臺北捷運客運量解析商圈人流變化

孫曉筠 [1]

[1] 臺北市政府捷運工程局會計室

Abstract

捷運營運與民眾的日常生活緊密結合，改變了臺北都會區民眾的生活型態和活動人口分布，民眾逐捷運而居，捷運路線周邊的經濟特性也隨之轉變，有新商圈興起，也有舊商圈没落，臺北市整體的商業環境隨著消費人潮的流動而轉變。商圈存續的首要因素就是消費人數的多寡，若只能滿足周邊消費者的需求，生意穩定但發展有侷限性，必須要能吸引外地消費者，才能促進產業發展。能影響外地消費者前來的意願，除了商圈本身的特色吸引力外，交通的可及性也是因素之一。

本文運用民國 108 年各捷運站分日分時進出站人次資料，交叉分析平日和假日各捷運站乘客流量的多寡，探討捷運站周邊商圈人流變化。研究發現各捷運站假日平均進出站人次普遍較平日為少，但仍有一些周邊生活機能具特色的捷運站例外。對於不同特色型態的商圈，建議採用差異化的輔導措施，以使政府資源發揮最大效能。

Keyword: 捷運客運量, 分日分時進出站人次, 商圈

Abstract No:IV-5-3

揭開臺鐵隱性旅客需求之面紗

何宇卿 [1]

[1] 交通部臺灣鐵路管理局主計室

Abstract

旅客預訂票紀錄為建構旅運需求分析之重要資料，完整訂購票作業資料包括:查詢、訂票、付款、取票、退/換票、驗/補票等，過往票務系統僅記錄成功完成訂票作業的票務資料流，未囊括不成功之紀錄，即旅客無法成功訂票之訂票未果資料。有鑑於此，臺鐵新（第四代）票務系統首度納入訂票未果資料，以完備旅客預訂票資訊。

每逢連續假期，臺鐵熱門區段或車次總是上演搶票攻防戰的戲碼，然而在有限的運量供給下，確實無法滿足所有旅客最佳的旅運需求，故臺鐵局亟需瞭解以往票務資訊系統所無法獲得旅客第一時間訂不到火車票之訂票未果資料。本研究藉由蒐集預訂票各構面資料（如:提前預訂天數、預約訂票時段別、起訖站熱區及熱門車次等），並串聯臺鐵「列車時刻表」等檔案，以深入探勘隱性旅客需求情形，刻畫隱性旅客的圖像特質，發掘出需求缺口，以及第一時間所流失之隱性旅客客收，俾利業務單位研擬票務及列車排點等相關措施參酌，進而提供旅客更臻完善的旅運服務，創造旅客與臺鐵雙贏局面。另運用統計檢定方法，探討臺鐵 e 訂通 APP 提供線上退票功能前後，訂票未果客群使用 APP 訂票的占比情形，結果其占比顯著提升，顯示 APP 退票優化功能是顯著便民措施。

Keyword: 訂票未果, 隱性旅客

海報論文

No.	作者	單位	講題
1.	戴志達	永析統計諮詢顧問	探討統計諮詢重要因素 - 以永析諮詢顧問為例
2-1.	江淑娥	以斯帖統計顧問公司	以斯帖統計 - 以專業貼近你心
2-2.			Cancer Risk on Residents of ^{60}Co-Contaminated Buildings in Taiwan - Analyzed Using the Two-stage Sampling Approach
3.	曹宇翔	高雄醫學大學	高醫大附設醫院醫學統計分析與諮詢服務介紹
4.	熊嘉妮	資料科學統計合作社	Introduction to Academia Sinica Data Science Statistical Cooperation Center - Case Studies
5.	蔡嘉軒	中央研究院	The Decomposite T^2-test when the Dimension is Large
6.	陳祈安	國家衛生研究院	Balancing Benefits and Risks in Dynamic Treatment Regimes
7.	羅琪	中華大學	Relationships among Brand Awareness, Brand Image, Perceived Value, Service Quality and Brand Loyalty - A Study of MOS Berger
8.	江村剛志	長庚大學	A Copula-based Markov Chain Model for Serially Dependent Event Times with a Dependent Terminal Event
9.	李易昀	國立中央大學	Optimal Strategies for Index Tracking with Risky Constrains

下頁續接

接續上頁

No.	作者	單位	講題
10.	雷亭儀	輔仁大學	Linear Heteroscedastic Measurement Error Models with Partial Replication
11.	黃清駿	華岦工程有限公司	Economic Design of Max Control Chart by Considering Asymmetric Loss Function
12.	蔡承翰	國立中山大學	A Kriging-Correlation Based Approach to Unsupervised Feature Selection Problems
13.	鍾冬川	逢甲大學	OUxy: an R Package for Inferring Statistical Phylogenetic Adaptive Trait Evolution
14.	林威均	國立政治大學	以基於 Spline 模型建構辨別呼吸中止症發作的特徵
15.	黃岱瑋	國立清華大學	Collateral Effect of Non-pharmaceutical Interventions Against COVID-19 on Other Infectious Diseases
16.	陳小英	國立臺北科技大學	婆媳和諧關係之探討
17.	林素秋	國立臺灣師範大學	如何避免成為下流老人 - 人力相關調查分析

12:05 ~ 13:20

Abstract No: 1

探討統計諮詢重要因素-以永析諮詢顧問為例

戴志達[1]

[1] 永析諮詢顧問有限公司

Abstract

永析諮詢顧問從 2011 年草創，正式立案於 2017 年，從事統計相關工作已達 10 年的經歷，服務過數千個專案。

統計諮詢除了對統計方法需要有足夠的理解，對於研究方法、文獻蒐集、實驗設計、資料處理、程式設計上也需要有一定的理解，
實務上面對諮詢的客戶時，除了了解客戶的需求，我們也會針對客戶遇到的困難給予相關議題的解答，根據客戶領域給予最適當的解法。

我們會建議需要統計諮詢的夥伴們從研究目的及研究方法上著手， 之後才是針對需求進行統計方法的建置，方才能適當的解決問題。在分析過程中如果有遇到特殊情況，可以從專業的統計文獻中找到解決方式，也可以透過不同的統計程式語言進行處理。

我們希望能透過這次海報讓大家對永析統計有更多的了解，以及對統計諮詢領域有更多的認識。

Keyword: 永析統計，統計分析，統計諮詢，研究方法，實驗設計

Abstract No: 2-1

以斯帖統計－以專業貼近你心

江淑娥[1]

[1] 以斯帖統計顧問股份有限公司

Abstract

以斯帖統計顧問公司登記於 2005 年 8 月, 2006 年 4 月 1 日開始正式對外營運, 主要的服務項目為統計教學、研究設計與統計方法的諮詢, 並承接公家機關、國內外藥廠、醫院, 或臨床醫療人員等合作案, 多年來已累積大量的服務客群, 以協助各單位學術研究的進行。

「以斯帖」設立的主要宗旨, 除了以專業的心成為各研究單位的貼心好伙伴, 更希望藉由清晰明確的教學方式, 讓學習統計不再是研究者的苦惱, 而是成就與快樂, 因此展開一系列的教學課程。統計教學的對象包括醫院、學校、公家機關等教學研究部委託之在職進修課程, 或與外商藥廠合作之教學方案, 以及公開授課多年的「應用生物醫學統計研習營」、「健保資料分析研習營」, 以及近期開設的「AI 人工智慧研習營」, 至今學員已散佈在各個醫學與學術領域中。

其中「應用生物醫學統計研習營」的課程包括:基礎醫學統計、流病資料分析方法、存活分析、統合分析、重覆測量方法、傾向分數加權及配對、階層線性模式、結構方程模式及樣本數估算等統計方法, 讓學習者收穫滿滿, 進而應用於學術研究中。

Keyword: 統計教學, 統計諮詢, 學術研究委託

Abstract No: 2-2

Cancer Risk on Residents of ^{60}Co-Contaminated Buildings in Taiwan - analyzed using the two-stage sampling approach

Shu-O Chiang[1], Chien-Hao Chen[2], Chuan-Jong Tung[3], Pesus Chou[4], Jacqueline Whang-Peng[5], Hung-Yi Chiou[6], Yi-Hau Chen[5], Li-Hui Yu[7], Lukas Jyuhn-Hsiarn Lee[8], Saou-Hsing Liou[1, 8]

[1]Graduate Institute of Life Sciences, Academia Sinica, National Health Research Institutes and National Defense Medical Center, Taipei, Taiwan., [2]ESTAT Statistical Consulting Co., Ltd., Taipei, Taiwan., [3]Department of Medical Imaging and Radiological Sciences, School of Medicine, Chang-Gang University., [4]Institute of Public Health, Yang-Ming University., [5]Academia Sinica, Taipei, Taiwan., [6]School of Public Health, Taipei Medical University., [7]Department of Health, Taipei City Government., [8]National Health Research Institutes, Miaoli, Taiwan.

Abstract

Purpose: This study attempted to investigate the effects of long-term, low-dozing ionizing radiation on the incidence of cancer in people living in radio-contaminated buildings using the two-stage sampling approach after correcting confounding factors.

Methods: This is a retrospective cohort study from 1982 to 2010. Demographic variables and migration data were collected from people who lived, worked or studied in 1,660 radio-contaminated buildings in northern Taiwan since their construction from 1982 onwards; the cumulative dose of radiation in the environment was estimated using the exposure dose data collected by the Atomic Energy Council. Stage-one cancer risk information was established by combining the basic information of the residents with the nationwide cancer registration data, national health insurance data, and national mortality data. The control cohort was established using a 1:10 matching ratio. Stage-two cancer risk information was established using epidemiological data on some of the RCB residents and the control cohort. The two-stage sampling approach was applied to combine data from both stages, also to control confounding factors, so as to further explore the correlation between radiation

and cancer.

Results: The incidence of cancers before correction, leukemia was the only cancer type that presented statistical significance (excluding chronic lymphocytic leukemia) (IDR = 1.66; 95% CI = 1.14-2.41). After correcting confounding factors with the two-stage sampling approach, risk of all cancers in the exposed cohort was found to be 0.89 times (ORa=0.89; 95%CIa=0.67-1.20) that of the control cohort. Although there was no increase in the overall risk of cancer incidence, there was a statistically significant increase in the risk of leukemia among RCB residents, which was 2.02 times (ORa=2.02; 95% CIa=1.25-3.27) that of the control cohort and showed a dose-response relationship, but there was no significant increase in risk of other types of cancer.

Conclusion: After correcting confounding factors with the two-stage sampling approach, when radio-contaminated buildings residents were exposed long-term to low-dose of Cobalt-60, there was no increase risk of overall cancer; however the risk of leukemia did have a statistically significant increase, indicating that long-term, low-dose radiation does have risks for developing a particular type of cancer.

Keyword: radiation, cancer, two-stage sampling approach

Abstract No: 3

高醫大附設醫院醫學統計分析與諮詢服務介紹

曹宇翔[1], 謝慧敏[1], 黃毓婷[1], 王昱蒼[1], 黃天祥[2], 黃聖棋[2]

[1] 高雄醫學大學附設醫院(Kaohsiung Medical University Hospital), [2] 高雄醫學大學

Abstract

高醫大附設醫院醫學統計分析與生物資訊研究室 及 高醫大大數據研究中心提供校院體系同仁統計相關問題之諮詢及分析服務, 提昇研究論文發表能量。

業務概況:
開放固定時間預約討論研究方向與分析方法, 由申請者、統計諮詢老師與統計分析師三方組成專案小組進行研究論文之剖析與撰寫。另定期舉辦統計協作分析與資料庫申請之說明會, 藉由服務提供來提升同仁論文發表質與量。

服務項目包括:
1.高醫體系醫院研究資料庫諮詢與臨床資料擷取
2.研究設計、計畫書撰寫諮詢
3.統計分析協作
4.統計軟體操作與分析結果判讀
5.舉辦統計分析及統計軟體之訓練課程。

Keyword: 高醫體系醫院研究資料庫, 臨床資料擷取, 統計分析協作, 統計諮詢

Abstract No: 4

Introduction to Academia Sinica Data Science Statistical Cooperation Center – Case Studies

熊嘉妮[1], 陳淑君[1], 蔡宜庭[1]

[1]Institute of Statistical Science, Academia Sinica

Abstract

Academia Sinica Data Science (or Data Intelligence) Statistical Cooperation Center (AS DISC) was established in January 2019 as the unique data-science-centric core facility at Academia Sinica. DISC provides statistical consultation services for the clients and real-world questions from various fields. In addition to the close cooperation with domain experts, DISC also provides the "statistical clinics" services aiming to nurture the students who want to learn more statistical core thinkings and data science perspectives through real problems. Our clients come from not only academic institutes and governmental agencies but also medical centers and industries. In this poster, we share our consulting experience in several case studies from Academia Sinica and hospitals.

Keyword: Data Science, Consulting Case, Core Facility

Abstract No: 5

The decomposite T^2-test when the dimension is large

蔡嘉軒 [1]

[1]Academia Sinica

Abstract

In this paper, we propose a test for the mean vector of high-dimensional data when data dimension p is a function of sample size n. The proposed test, called the decomposite T^2-test, in the high dimensional testing problem is constructed based on the estimation work of Ledoit and Wolf (2018), which is an optimal orthogonally equivariant estimator of the inverse of population covariance matrix under Stein loss function. The asymptotic distribution function of the proposed test statistic and its modified version are investigated. Under a sequence of local alternatives the asymptotic distribution and its asymptotic power function can then be obtained. An application of the proposed decomposite T^2-test statistic is in testing significance for the effect of COVID-19 to public transportation, in which the data are from Taipei Metro System.

Keyword: Decomposite T^2-test, High-dimensional covariance matrix, Orthogonally equivariant estimator

Abstract No: 6

Balancing Benefits and Risks in Dynamic Treatment Regimes

陳祈安[1]，蕭金福[1]

[1] 國家衛生研究院

Abstract

Dynamic treatment regime (DTR) is an area which use the concept of data driven, statistical methods, and machine learning techniques in precision medicine. A DTR is a set of sequential decision rules, each corresponding to a key decision point in a disease or disorder process. Each of rules takes as input patient information and returns the treatment option he or she should receive. In the last decade, DTR is said to be optimal if it maximizes the mean of a single outcome. However, this framework oversimplifies the goal of clinical decision making which include balancing several potentially competing outcomes and accommodating heterogeneous patient preferences across these outcomes. In this study, the proposed framework allows for much general definitions of optimality in which the optimal regime maximizes a single scalar feature subject to constraints on other features in the multivariate normal outcome distribution. By varying the optimization functions and constraints, patients and clinicians can interactively explore tradeoffs across treatment strategies and select a treatment plan that suits the individual need of each patient.

Keyword: Precision medicine, Constrained Q-learning, Dynamic treatment regimes

Abstract No: 7

Relationships among Brand Awareness, Brand Image, Perceived Value, Service Quality and Brand Loyalty-A Study of MOS Berger

孫子鈞[1], 羅琪[1]

[1]Chung Hua University

Abstract

In 1990, MOS Burger entered into Taiwan's market. It created another new brand image emphasizing the principles of “material selection strictly” and “meal production after ordering”. Its better product quality performance and fresh taste are favored by many consumers. Nowadays, there are many studies related to brand awareness, brand image, perceived value, service quality and brand loyalty, but few scholars have done a complete study on the relationship between these five latent variables. This research takes the customers of MOS Burger as the survey object, uses the structural equation model to explore and establish the relationship between brand awareness, brand image, perceived value, service quality and brand loyalty. Convenience sampling was used for data collection, and a total of 450 usable questionnaires were obtained. The major findings and suggestions of this study are as follows. Brand awareness has positive direct effect on service quality and brand image, Service quality and brand image have positive direct effect on perceived value. Perceived value and brand image have positive direct effect on brand loyalty. However, brand image is not significantly related to perceived value and service quality is not positively related to brand loyalty.

Keyword: brand awareness, brand image, perceived value, service quality, brand loyalty

Abstract No: 8

A copula-based Markov chain model for serially dependent event times with a dependent terminal event

黃昕蔚[1], 王維菁[1], 江村剛志[2]

[1] 國立交通大學, [2] 長庚大学

Abstract

Copula modeling for serial dependence has been extensively discussed in a time series context. However, fitting copula-based Markov models for serially dependent survival data is challenging due to the complex censoring mechanisms. The purpose of this paper is to develop likelihood-based methods for fitting a copula-based Markov chain model to serially dependent event times that are dependently censored by a terminal event, such as death. We propose a novel copula-based Markov chain model for describing serial dependence in recurrent event times. We also apply another copula model for handling dependent censoring. Due to the complex likelihood function with the two copulas, we propose a two-stage estimation method under Weibull distributions for fitting the survival data. The asymptotic normality of the proposed estimator is established through the theory of estimating functions. We propose a jackknife method for interval estimates, which is shown to be asymptotically consistent. To select suitable copulas for a given dataset, we propose a model selection method according to the 2nd stage likelihood. We conduct simulation studies to assess the performance of the proposed methods. For illustration, we analyze survival data from colorectal cancer patients. We implement the proposed methods in our original R package "*Copula.Markov.survival*" that is made available in CRAN (https://cran.r-project.org/). The full paper is available in [1].

[1] Huang, X., Wang, W. & Emura, T. (2000). A copula-based Markov chain model for serially dependent event times with a dependent terminal event. Japanese Journal of Statistics and Data Science. https://doi.org/10.1007/s42081-020-00087-8

Keyword: Survival analysis, Copula, Recurrent event, Censoring, Weibull distribution

Abstract No: 9

Optimal Strategies for Index Tracking with Risky Constrains

李易昀[1], 孫立憲[1]

[1] 國立中央大學

Abstract

Index tracking is a popular passive investment strategy in finance. It refers to the problem of reproducing the performance of a stock market index by considering a portfolio of assets comprised on the index. This paper mainly attempts to construct a model based on the technique of the portfolio optimization problem through the linear quadratic regulator to trace closely an index. We obtain the optimal strategy using the dynamic programming and the corresponding HJB equation. However, we consider the problem of tracking instability when tracking the index through portfolio optimization. In this case would cause the excessive tracking error. Therefore, this research specifically joins the penalty quadratic term in risky assets and attempts to capture the tracking of unstable situations to weaken the tracking error. We show that the proposed model controls the tracking instability and compare the performance with the model that without joining the penalty quadratic term in risky assets using an empirical study of the S&P 500 and several individual stocks in the U.S.

Keyword: Market tracking, portfolio optimization, dynamic programming principle, Hamilton-Jacobi-Bellman equation, exact penalty function

Abstract No: 10

Linear Heteroscedastic Measurement Error Models with Partial Replication

雷亭儀[1]，蔡嘉仁[1]

[1] 輔仁大學

Abstract

When an adjusted or new method (test method) is developed, the linear measurement error model is commonly used to compare it with another reference method. For this routine practice, the measurements on the reference method can be placed on the y-axis, whereas those of the test method on the x-axis, then the slope of this linear relationship indicates the agreement between them and also the performance of the test method. In estimates of plant disease severity, there have the replicated observations in human raters (test method) but images of the leaves measured (reference method) do not replicate. How to evaluate the agreement for visual estimates by human raters and the accuracy of estimates of mean disease severity? This study focuses on investigating partially replicated observations in linear models with heteroscedastic measurement errors. We develop the modified least squares approach and apply it in plant disease severity data. The covariance matrix of parameters will be constructed using the sandwich method and the model-based expectation method, based on unbiased estimating equations. This approach can be applied to different fields based on both functional model and structural model.

Keyword: heteroscedastic errors, measurement error model, modified adjusted squares, partial replication

Abstract No: 11

Economic design of Max control chart by considering asymmetric loss function

黃清駿[1]

[1] 華岦工程有限公司

Abstract

A single control chart, Max chart, has been shown to simplify and perform as effectively as joint and *S* charts for monitoring both the process mean and standard deviation (Chen and Cheng (1998); Chen and Huang (2006); Lee and Khoo (2017, 2018)). This paper develops an economic model for the design of Max control charts by considering asymmetric quality loss functions. Feng and Kapur (2006) developed economic models of specifications for 100% inspection based on asymmetric quality loss functions. An asymmetric quality loss function allows the performance of a product to deteriorate in different ways as the quality characteristic deviates to one or other side of the target value. The objective of economic design is to find the optimal settings of chart parameters that minimize the associated cost of operating control charts in process environment by considering asymmetric quadratic loss function. A numerical example is presented to illustrate the application of the chart.

Keyword: control chart, economic design, quality loss function

Abstract No: 12

A Kriging-Correlation Based Approach to Unsupervised Feature Selection Problems

蔡承翰[1], 郭美惠[1], 黃士峰[2]

[1]國立中山大學, [2]國立高雄大學

Abstract

Feature selection for unsupervised learning is a challenging problem. Laplacian Score is one of the well-known locality preserving filtering methods for unsupervised feature selection. In this study, we propose a Kriging-Correlation (KC) Score, which integrates the Automatic Fixed Rank Kriging method with the correlation analysis, to extract important features for unsupervised high dimensional clustering problems. We design a data generating process which generates high dimensional ($p > n$) data with different dropout rates mimicking high throughput gene expression measurement of individual cells. The single-cell interpretation via multi-kernel learning (SIMLR) proposed by Wang *et al.* (2017) is used to obtain similarity matrix and the t-distributed stochastic neighbor embedding (t-SNE) is applied for dimension reduction and clustering visualization. The results of our simulation study show that the variables selected by the KC Score achieve higher accuracy and Normalized Mutual Information than the ones selected by the Laplacian Score. We further apply the proposed method to four single-cell data sets, the results again support the superiority of the KC Score over the Laplacian Score.

[1] Wang, B., Zhu, J., Pierson, E., Ramazzotti, D. and Batzoglou, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning, *Nature Methods,* **14,** 414–416

Keyword: Feature selection, Unsupervised learning, SIMLR, Laplacian Score

Abstract No: 13

OUxy: an R Package for Inferring Statistical Phylogenetic Adaptive Trait Evolution

鍾冬川 [1]

[1] 逢甲大學

Abstract

The development of an adaptive trait simulator package for inferring trait evolution along a phylogenetic tree is shown. Stochastic processes of the continuous type are broadly applied to modeling trait evolution when the evolutionary relationship among species and traits of study interest are present. By including several popular stochastic processes, evolutionary information embedded in a dataset can be revealed. The highlights of the method include: 1. The implementation of the popular Cox-Ingersol-Ross process for modeling rate evolution within the package to prevent rates from becoming negative and thus is potentially a useful extension to study adaptive trait evolution in randomly evolved environment. 2. The established trait simulator approach along with approximate Bayesian computation procedure provides a feasible statistical inference without model likelihood. 3. The procedure proposed for trait simulator along phylogenetic tree can be applied to all established models of trait evolution in literature, thus providing users an alternative option to analyze their data.

Keyword: Statistical modeling, Brownian motion, Cox-ingersoll-ross process, Ornstein-uhlenbeck process, Phylogenetic comparative method

Abstract No: 14

以基於 spline 模型建構辨別呼吸中止症發作的特徵

林威均[1], 黃子銘[1], 黃佳慧[1], 楊素芬[1], 蕭又新[1]

[1]政治大學

Abstract

睡眠呼吸中止症(sleep apnea)是近年來健康議題的重點討論對象，是一種年長者及男性民眾容易罹患的睡眠障礙疾病，患者會在睡眠時反覆出現暫時性的呼吸停止現象，嚴重影響自身的睡眠品質。

本次研究對呼吸氣流訊號進行週期劃分，並以基於 spline 的方法來建立呼吸週期的模型，觀察各週期的配適情況及計算相關參數，發現在呼吸停止現象發生時，確實會有一些在正常呼吸狀況下較少見的模式，此外，我們將這些參數去建構辨別呼吸中止的羅吉斯迴歸模型，得到還不錯的判斷結果，相信這些參數能夠作為辨別呼吸中止症的有效特徵。

Keyword: spline, 睡眠呼吸中止症, 異常偵測, sleep apnea

Abstract No: 15

Collateral effect of non-pharmaceutical interventions against COVID-19 on other infectious diseases

黃岱瑋[1]

[1]國立清華大學

Abstract

COVID-19 has quickly spread throughout the world and changed the way we live. Unprecedented measures have been taken in response to this pandemic. Several studies suggested that control measures taken against COVID-19 have been substantially effective on reducing its transmission. However, the relationship between population behavior changes against COVID-19 and transmission of other infectious diseases is rarely discussed and has not been examined on a large scale.

Our goal is to study the impact of non-pharmaceutical interventions against COVID-19 on transmission of other infectious diseases. We first quantified the correlation between temporal changes of COVID-19 and acute respiratory diseases including influenza, influenza-like illness (ILI), and acute respiratory infections (ARI) by using data from multiple countries. Our preliminary results show a positive correlation between the transmission of COVID-19 and acute respiratory diseases. Areas where there are a decrease in COVID-19 transmission also show a declining trend of acute respiratory disease confirmed cases. These results can indirectly support our hypothesis: control measures taken against COVID-19 have collateral impact on other infectious diseases.

Our future works are to estimate the reduction in transmission rates of other infectious diseases, compare the level of reduction between countries, and test whether transmission dynamics of a specific infectious disease can serve as an indicator to estimate spread risk of COVID-19 by using epidemiological assessment tools. Our analytical results will not only provide insights into the realistic impact of non-pharmaceutical interventions against COVID-19 on our lives, but also establish

a new possible epidemiological model for COVID-19 risk assessment.

Keyword: COVID-19, non-pharmaceutical intervention, collateral effect, infectious disease

Abstract No: 16

婆媳和諧關係之探討

陳小英[1]

[1] 國立臺北科技大學

Abstract

婆媳互動仍是現代女性的困擾嗎？隨著時代的變遷，家庭結構的轉變，華人宗法規範長久以來的限制是否受到時空的牽引，對現今社會的婆媳關係產生變化呢？過往的婆媳研究多著重於衝突面，然而婆媳互動真的僅有衝突而無和諧嗎？在兩兩存續的關係中，原生家庭與丈夫的支持為已婚女性與婆婆產生和諧關係的關鍵呢？文化規範在現今婆媳互動中的影響力呢？

本研究的研究目的，探究促使婆媳和諧關係的關鍵因素。故採分層叢集非隨機取樣之方式，共計抽取 42 所學校，2339 名學生家長填答問卷，回收問卷 1752 份，剔除無效問卷 495 份，餘得有效問卷 1257 份，有效率 71.75%。

本研究採結構方程模式來驗證婆媳和諧關係之概念模式，結果為透過夫妻互動，獲得丈夫的支持越多則已婚女性的婆媳關係越好。其次，原生家庭支持可促進婆媳和諧關係，原生家庭協助教養子女，可減低女性因教養子女的態度與方式形成的摩擦，此實質的協助解決可降低婆媳衝突的機會。另外，當已婚女性越認同角色規範的女性其婆媳關係越和諧的，除婆婆的觀念是傳統或是習俗典範的傳遞者外，現代女性本身仍扮演傳承的角色。因此現代社會中某些傳統的特質仍會繼續保存而未明顯的消退，亦非所有現代文化特徵或價值元素能取代傳統文化特徵或價值元素。

Keyword: 角色規範，原生家庭支持，婆媳和諧關係，丈夫支持

Abstract No: 17

如何避免成為下流老人—人力相關調查分析

林素秋[1]

[1] 國立臺灣師範大學

Abstract

高齡者難以維持一般生活水準的 2 項重要因素為:經濟支持、人際網絡支持。其中，老人經濟資源來源分 3 大類:個人投入勞動市場獲得的生命週期內移轉、家庭內獲得的代間移轉、社會或政府資源重分配獲得的社會移轉(Ryder, 1988)，3 類中，無子女者缺乏後代奉養的代間資源移轉;未婚單身者較缺乏社會移轉能獲得的經濟資源。故未婚單身無子女者，老年經濟來源較少，成為「下流老人」風險較高。

我國非婚生子女比例極低，故未婚單身人口接近未婚單身無子女人口。當前我國進入高齡社會，關注老年貧窮問題，未婚單身族群不可忽視。故本文運用人力資源調查、97-106 年人力運用調查等報告，分析中高齡未婚單身族群的勞動力參與率。

結果發現，未婚單身人口因家庭經濟負擔較低，比其他婚姻狀態，退出勞動力市場的比例較高、年齡較早。而未婚單身人口中，55 歲以上教育程度較低者的勞動力參與率較高，而隨著年齡增長的生理退化，體能、認知能力與學習成長都相對不利，又受限於學歷，面對知識經濟與科技快速發展，年齡越高越易面臨新型態從業能力不足(黃春長、王維旎, 2016)，因此與其探討如何提高中高齡未婚單身族群勞動力參與率，思考如何幫助目前的勞動市場參與者，更能幫助避免其未來落入老後貧窮。

Keyword: 老年貧窮，中高齡未婚單身，勞動力參與率

主辦單位

中央研究院
統計科學研究所

中國統計學社

合辦單位

行政院
主計總處

教育部

數學研究
推動中心

主計
協進社

中華機率
統計學會