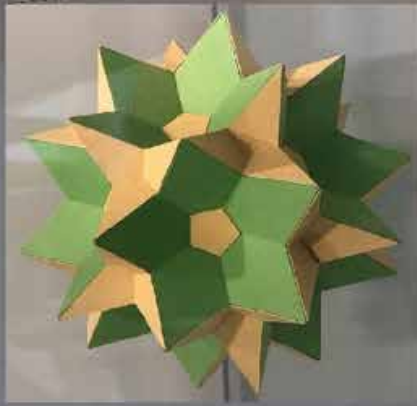




Charles Sturt
University

School of Computing
and Mathematics



Book of Abstracts

Applied Statistics and Policy Analysis Conference 2019

CSU Convention Centre, Wagga Wagga
September 5 and 6, 2019

ASPAC2019 Administrator
Data Science Research Unit (DSRU)
School of Computing and Mathematics

Charles Sturt University
Wagga Wagga, NSW 2678
AUSTRALIA



ASPAC2019 Organiser:

Dr Azizur Rahman
Leader of Statistics and Data Mining Research Group
School of Computing and Mathematics

Conference Administrative Assistant:

Emma Humphries
School of Computing and Mathematics

Edited by:

Dr Azizur Rahman
School of Computing and Mathematics

Published by:

Statistics and Data Mining Research Group
School of Computing and Mathematics, Faculty of
Business Justice and Behavioural Science
Charles Sturt University, Australia

Printed by:

CSU Print 2019

Welcome



It's a great pleasure for me to welcome you to the Applied Statistics and Policy Analysis Conference 2019 (ASPAC2019) which is being organised by the Statistics and Data Mining Research Group from the Faculty of Business, Justice and Behavioural Sciences (FBJBS) on September 05-06, 2019.

The Faculty of BJBS has an outstanding record of high achievement in research. With over 240 academic researchers and higher degree research students, the Faculty is thriving. The Faculty has a number of well-established areas of research excellence across its schools and centres. Our high-quality research is designed to have application and impact in our regions as well as the national and international level. The Applied Statistics and Data Mining Research area is a recognized Australian research priority and the Faculty incorporates the training of future research leaders.

I am very proud of the Faculty's involvement in the organization of ASPAC2019 as it will provide a cross-disciplinary venue for researchers and industry practitioners at local, national and international level to address the rich space of shaping effective policy through the use of big data, accurate estimates and modern computing tools and statistical modelling. The CSU is the only regional university to set up a Data Science Research hub, which aims to grow Australia's national capability in data driven research for helping evidenced based policy, development and commercialisation.

I strongly believe that this conference of about 60 delegates from Australia and overseas is an opportunity to collaborate through debate, insightful presentations, discussions and the sharing of technical ideas with colleagues and industry experts.

I hope you enjoy the ASPAC2019 and it provides you with the opportunity to engage and discover more about the exciting research being conducted by your peers, industry and international researchers.

A handwritten signature in white ink, appearing to read 'Tracey Green', written in a cursive style.

Professor Tracey Green
Executive Dean,
Faculty of Business, Justice and Behavioural Sciences

Welcome



It's my great pleasure to welcome you to the Applied Statistics and Policy Analysis Conference 2019 (ASPAC2019), organised by the Statistics and Data Mining Research Group from the Faculty of Business, Justice and Behavioural Sciences (FBJBS), at the CSU Convention Centre Wagga, September 05-06, 2019.

The Faculty of BJBS has a significant research record as a result of obtaining competitive grants, industry and government funding, and producing high quality research publications. The Faculty has about 80 academic researchers and over 100 research higher degree students concentrated in various research areas to benefit the communities in our regions as well as the communities we serve nationally and internationally. Applied Statistics and Data Mining research focuses on innovative, relevant and ground-breaking research across the wide spectrum of modelling knowledge discovery areas and is a recognized Australian research priority area.

The ASPAC2019 aims to promote research collaborations and exchange ideas between data scientists, applied statisticians and data modellers who will detail the latest innovations in research to gather and disseminate information from small to big data settings, and from policy analysts who will describe how they use existing information from increasing big data environments and indicate areas in which there need to be methodological and technological developments. Another aim is to establish connections between researchers at tertiary institutions and working in industry in Australia and overseas.

I strongly believe that this forum of 60 participants from Australia and overseas will provide an opportunity to collaborate through presentations, discussions and the sharing of technical ideas with colleagues and experts.

I hope you enjoy the ASPAC2019 and gain knowledge and insight from the presentations and exchange of innovative research ideas by your peers and expert senior researchers.

Professor Mark Morrison
Associate-Dean (Research), FBJBS

Welcome



On behalf of the School of Computing and Mathematics (SCM), it is my great pleasure to welcome you to the 2nd Applied Statistics and Policy Analysis Conference 2019 (ASPAC2019) which will take place at the Charles Sturt University Convention Centre, Wagga Wagga Campus, on 5 and 6 September 2019.

SCM, through its strong focus on and commitment to research, provides a high level of learning, mentoring and guidance in various Faculty Research Areas of Charles Sturt. Our academic staff are very successful researchers in terms of winning competitive research grants and research excellence awards (including prestigious Vice-Chancellor awards). Our School publication profile has also significantly improved due to a substantial increase in high ranking journal publications.

ASPAC2019 will provide a cross-disciplinary venue for researchers from government, industry, and academia at local national and international level together to create a forum for exchanging innovative research ideas, state-of-the-art, and highlight activities that deal with modern data modelling challenges and assisting various policy decisions. The ASPAC2019 keynote sessions involve increasing access to government microdata for research purposes to micro-geospatial targeting for precision public policy. ASPAC2019 also have a range of invited sessions which are specially focusing on industry's research issues and modelling opportunities.

I am confident that ASPAC2019 will successfully serve the purpose of creating an environment of cooperation and collaboration through insightful presentations, discussions and sharing the technical ideas with colleagues and experts at national and global level.

Once again, I welcome you all and hope you enjoy the ASPAC2019

A handwritten signature in cursive script that reads "Irfan Altas". The ink is dark and the signature is written in a fluid, connected style.

Associate Professor Irfan Altas
Head of School, School of Computing and Mathematics

Table of Contents

Preface		10
Acknowledgements		12
SIDO1:	Modeling for prospect of aman rice production in Dhaka division, Bangladesh Sayed Mohibul Hossen, Md. Takrib Hossain, Aditi Chakraborty and DR. Mohd Tahir Ismail	15
SIDO2:	The role of forecasting the size of school populations in planning for regional education in Australia Terence Mills, Leonardo Veliz and Julianne Lynch	16
SIDO3:	Basic income and optimal taxes. Efficiency, equity, losers and winners. Ugo Colombino and Nizamul Islam	17
SIDO4:	A tax benefit model for policy evaluation in Luxembourg: LuxTaxBen Nizamul Islam and Lennart Flood	18
SIDO5:	Delay effect and burden of weather-related tuberculosis cases in Rajshahi province, Bangladesh, 2007-2012 Md Abdul Kuddus, Emma McBryde and Oyelola Adegboye	19
SIDO6:	Increasing access to government microdata for research purposes* D. Trewin	20
SIDO7:	Statistical analysis of big data: The good, the bad and the ugly* Susan Wilson	20
SIDO8:	Technical efficiency and value chain analysis of potato: An empirical study in Bangladesh Mahfuza Afroj, Mohammad Mizanul Haque Kazal, Imtiaz Faruk Chowdhury and Md. Mahfuzar Rahman	21
SIDO9:	Statistical Data Analysis without Statistical Significance Gang Xie (John)	22
SIDO10:	Water Supply and Sanitation in Nepal: an Overview Jitendra Upadhyaya	23
SIDO11:	Missing continuous outcomes in stepped-wedge trials Sorif Hossain and Anower Hossain	24
SIDO12:	Assessing competition in the Indian banking sector using panel data Zhiheng Li, Shuangzhe Liu, Fanda Meng and Milind Sathye	25
SIDO14:	River-based Tilapia Cage culture in Bangladesh: Impacts and Determinants of Adoption Zhiheng Li, Shuangzhe Liu, Fanda Meng and Milind Sathye	

SID15:	Geospatial and temporal variation of prostate cancer incidence Win Wah, Arul Earnest, Susannah Ahern, Sue Evans Melanie Evans and Jeremy Millar	26
SID16:	Factors associated with coronary heart disease among elderly in Sylhet region in Bangladesh: Astudy based on different communities Kanis Fatama Ferdushi, Anton Abdulbasah kamil, Mohammad Nayeem Hasan and Tanjila Islam	27
SID17:	Finite mixture modelling approach to identify factors affecting children ever born for 15–49 year old women in Asian country M. Karimuzzaman, M. M. Hossain and A. Rahman	28
SID18:	An assessment of influencing factors for motherhood during childhood using factor analysis and logistic regression methods Mohammad Salim Zahangir and Mosammat Zamilun Nahar	29
SID19:	Applied Bayesian modeling for assessment of interpretation uncertainty in spatial domains Scott McManus, Azizur Rahman, Ana Horta and Jacqueline Coombes	30
SID20:	A spatially-explicit microsimulation analysis of the proposed sugar tax on population health in the ACT A. Richardson, A. Lal, NCEPH, C. Joseph and B. Phillips	31
SID22:	Predict neonatal intensive care unit (NICU) discharge date using machine learning techniques Arunkumar Namachivayam and Laxman Bidari	32
SID23:	A study on the impact of Alcoholism on EEG-based Cryptographic Key Generation Systems Dang Nguyen, Dat Tran, Wanli Ma	33
SID24:	Performance Analysis of Clustering algorithms for mineral exploration from Hyperspectral data Deepa C, Amba Shetty and A.V. Narasimhadhan	34
SID25:	Reviewing the past to shape the future of complex and chaotic IT project management in retail J. Hayes, M. R. Islam and A. Rahman	35
SID26:	A model based approach of measuring methane emission from different rice cultivation systems in a South-East Asian country Shamim Mia, Milton Kumar Saha, AKM Abdul Ahad Biswas, Md. Abdus Sattar and Feike A. Dijkstra	36
SID27:	Nutrient loading in the river systems around major cities in Bangladesh: A quantitative estimate with consequences and potential recycling options Shamim Mia, Md. Rushna Alam, Md. Abdus Sattar, and Feike Dijkstra	37
SID28:	Determining Risk Factors of Antenatal Care Attendance and its Frequency in Bangladesh: An Application of Count Regression Analysis Kakoli Rani Bhowmik, Sumonkanti Das, and Md. Atiqul Islam	38

SID31:	Estimation of Child Undernutrition at Disaggregated Administrative Tiers of a North-Eastern District of Bangladesh: An Application of Small Area Estimation Method Sumonkanti Das, Bappi Kumar, Md. Zakir Hossain, Sabbir Tahmidur Rahman and Azizur Rahman	39
SID32:	On Propensity Score Methodology Paul Dewick and Shuangzhe Liu	40
SID33:	Consumers adoption behavior prediction through technology acceptance model and machine learning models Xinying Li and Lihong Zheng	41
SID34:	Effect of women's education on Skilled Birth Attendants in South and South East Asia: A Cross-country assessment on Sustainable Development Goal 3.1 Raaj Kishore Biswas, Nurjahan Ananna and Jahar Bhowmik	42
SID35:	Asking Good Questions to Understand Voluntary Enrolments in Mathematics Ning Li	43
SID36:	Modelling and Analysis of Computer Experiments using a Simple Pendulum Model Kazeem Adewale Osuolale	44
SID37:	Modelling the arrival of plant pathogens: the role of import volumes and biosecurity interventions R. Duncan, B. Jennifer, P. Hulme, J. Cooper, P. Johnston and B. Sikes	45
SID38:	Can Data Fusion increase the Performance of Action Detection and Recognition in the Dark? Anwaar Ulhaq	46
SID39:	Seasonal occurrence of amoxicillin, ampicillin and penicillin residues in informal milk supply chains Naveed Aslam, Sosheel S Godfrey, Mateen Abbas, Muhammad Y Tipu, Muhammad Ishaq, David M McGill, Hassan M Warriach, Muhammad Husnain, Peter C Wynn	47
SID40:	Evaluating Faster-RCNN and YOLOv3 for Target Detection in Multi-sensor Data Anwaar Ulhaq, Asim Khan and Randall Robinson	48
SID41:	Semantic vegetation detection in repeat photography for environmental data analysis Asim Khan, Anwaar Ulhaq, Mobeen ur Rehman and Randall W Robinson	49
SID42:	Policy into practice; statistics the forgotten gatekeeper I. Hume	50
SID43:	Progress and controversy in statistical modelling of the genomic architecture of complex traits* David Balding	51
SID44:	Optimal policy modelling and welfare policy settings in Australia B. Phillips	52

SID45:	Using individual based modelling to inform policy: Where we have come from, and where big data will lead us* R. Tanton	52
SID46:	Sensor-centric source inference and image clustering for supporting various decision making process* C-T. Li	53
SID47:	Finding significant determinants and impacts of farm-level integrated pest management practices using statistical tools Md Sadique Rahman	54
SID48:	Parental perception on children online safety; a study on parents of different ethnic communities in Australia A. Imran	55
SID49:	Using a spatial farm microsimulation model for Victoria to estimate the impact of an external shock on farmer incomes Y. Vidyattama and R. Tanton	56
SID50:	Wavelet-Based Quantile density function estimation under random censorship Esmaeil Shirazi and Hassan Doosti	56
SID51:	Computing robust statistics via an EM Algorithm Maheswaran Rohan	57
SID52:	Performance Analysis of Clustering algorithms for mineral exploration from Hyperspectral data Deepa C, Amba Shetty and A.V. Narasimhadhan	58
SID53:	Data privacy and security in the cloud Peter Padiet, Rafiqul Islam and Azizur Rahman	59
SID54:	Working Zones 2016 Karen Malam	59
SID55:	Adapting truck GPS data for freight metrics David Mitchell	60
SID56:	The ABC of Big Data Siu-Ming Tam	60
SID57:	Herbicides and plant hormesis: a case study in Conyza sumatrensis Md Asaduzzaman, Eric Koetz and Hanwen Wu	61
SID57:	Current research and statistical demands within DPI Deb Slinger	61
SID59:	Multilevel modelling for precision public policy S.V. Subramanian	62
SID60:	Spatial microsimulation and multilevel methods for precision public policy A. Rahman	62

Preface

The overwhelming growth of data and its users is a reality, which has put new thoughts among the research community to devise new ideas for giving data driven evidenced-based policy decisions at local, state, national and international level. In recent years, applied statistics and data science have received renewed interest from a broad range of stakeholders ranging from Governments, to corporations and end users of data and its analysis or modelling tools. As a result, applied statistics and data science research such as data mining and policy analysis have been placed high as a national priority in many countries including Australia. In this data centric world with vastly growing demand situation, there is a need to ensure that reliable statistical and modelling solutions that address important and emerging policy issues at both public and private institutions are disseminated timely and widely amongst the research and industry community.

The 2nd Applied Statistics and Policy Analysis Conference 2019 (ASPAC2019) aims to promote research collaborations and exchange ideas between data scientists, applied statisticians and data modellers who will detail the latest innovations in research

to gather and disseminate information from small to big data settings, and from policy analysts who will describe how they use existing information from increasing big data environments and indicate areas in which there need to be methodological and technological developments. Another aim is to establish connections between researchers at tertiary institutions and working in industry in Australia and overseas. The theme of ASPAC2019 is “Effective policy through the use of big data, accurate estimates and modern computing tools and statistical modelling”.

This year, the main feature of the ASPAC2019 is that it has received a more than 60 good quality submissions (covering a wide range of topics including health statistics, small area estimation, microsimulation modelling, data science, public policy and agricultural modelling and policy) from 16 different countries across the globe. A truly international mix of HDR students to expert researchers in academia and industry are attending the 2019 conference. The keynote sessions involve increasing access to government microdata for research purposes to micro-geospatial targeting for precision public policy. ASPAC2019 also have a range of

invited sessions and panel discussion which are specially focusing on industry's research issues and modelling opportunities. It will illustrate at "how we can use the different modelling tools and statistics techniques to inform policy decision making processes in a better way."

In addition to this refereed abstracts book, ASPAC2019 will publish selected full papers as a volume with the Springer. A comprehensive extended version of very high quality papers would be considered for submission to special issue of journal as well. Authors are encouraged to write an expression of interest by email to the Conference Organising Committee Chair on aspac@csu.edu.au for contributing to journal publications.

Finally, thanks to all authors and participants who have contributed to the ASPAC2019 initiative. We are looking forward to welcome you in the next conference as well.

Dr Azizur Rahman

Leader, Statistics and Data Mining Research group
ASPAC2019 Organising Committee Chair



Acknowledgements

We would like to thank all the members of the Organising Committee, Program Committee, Publicity Team, Awards Committee and Referees for the peer reviews who have provided excellent supports to the Applied Statistics and Policy Analysis Conference (ASPAC2019). Thanks are also due to all of our valuable sponsors and partners for their significant contributions. Special thanks to Professor John Germov, the Charles Sturt Provost and Deputy Vice-Chancellor (Academic) for his valuable time and speech to open ASPAC2019.

Organising Committee:

Program Chairs:

Dr Azizur Rahman, *Charles Sturt University*
Professor Syed Ejaz Ahmed, *Brock University*

Organising Committee:

Dr Azizur Rahman (Chair), *Charles Sturt University*
Professor Raymond Chambers, *University of Wollongong*
Professor Syed Ejaz Ahmed, *Brock University*
Professor Terence Mills, *La Trobe University*
Professor Denny Meyer, *Swinburne University of Technology*
Associate Professor Md Rafiqul Islam, *Charles Sturt University*
Associate Professor Ben Phillips, *Australian National University*
Dr Nizamul Islam, *The Luxembourg Institute of Socio-Economic Research (LISER)*
Dr Ryan Ip, *Charles Sturt University*
Dr Siu-Ming Tam, *Australian Bureau of Statistics and University of Wollongong*
Associate Professor Shuangzhe Liu, *University of Canberra*
Dr Jahar Bhowmik, *Swinburne University of Technology*
Dr Md Musfequs Salehin, *Charles Sturt University*

Dr Cathy Honge Gong, *Australian National University*
Dr Dmitry Demskoy, *Charles Sturt University*
Dr Sumonkanti Das, *Maastricht University*
Associate Professor Md Zahidul Islam, *Charles Sturt University*
Dr Anwaar Ulhaq, *Charles Sturt University*
Prof. Marcin Szymkowiak, *Pozna University*
Dr Md Asaduzzaman, *NSW-DPI and CSU*
Associate Professor Andrew Robinson, *University of Melbourne*
Dr Maheswaran Rohan, *NSW Department of Primary Industries (DPI)*
Dr Mizanur Khondoker, *University of East Anglia, UK*

Program Committee:

A. Rahman, Prof. J. Hicks (*CSU*), Prof. S.E. Ahmed (*Brock*), M. R. Islam, Prof. T. Mills (*La Trobe Uni*), Dr S. Das (*Maastricht Uni*), R. Ip, Dr C. H. Gong (*ANU*), M. M. Salehin, Dr R. Yunus (*Uni. Malaya*), Dr Z. Li (*CSU*), Prof. D. Meyer (*Swinburne*), Dr A. Kabir (*CSU*), A/Prof. X. Huang (*CSU*), Dr L. Zheng (*CSU*), Dr S. Rehman (*CSU*), Dr J. Bhowmik (*Swinburne*), Dr N. Islam (*LISER, Luxembourg*), Prof. M. Szymkowiak (*Pozna Uni. of Economics and Business*), Dr S-M. Tam (*ABS & UoW*), A/Prof. B. Phillips (*ANU*), A/Prof. S. Liu (*UC*), Dr M. Asaduzzaman (*NSW-DPI & CSU*), Dr M. Rohan (*NSW-DPI*), A/Prof. A. Robinson (*Melbourne*), Prof. R. Chambers (*NIASRA, Wollongong*), Dr M. Khondoker (*UEA, UK*).

Awards Committee:

Dr Azizur Rahman (Chair), Associate Professor Md Rafiqul Islam, Dr Lihong Zheng and Professor Dennis Trewin.

Publicity Team:

Dr Anwaar Ulhaq (Co-chair), Dr Ashad Kabir (Co-chair), Dr Sabih Rehman, Dr Dmitry Demskoy, Dr Ryan Ip, Azizur Rahman

Sponsors



Australian Government

Department of Infrastructure, Regional Development and Cities

Bureau of Infrastructure, Transport and Regional Economics



**Charles Sturt
University**



Publication and Information Partners:



A special thank goes to Drs Dmitry Demskoy, Lihong Zheng, Ryan Ip and A/Prof. Md Rafiqul Islam for their outstanding local support, Amna Sabih and Emma Humphries for their utmost professionalism and enthusiasm in providing administrative supports, A/Profs Md Zahidul Islam, Md Rafiqul Islam and Tanveer Zia for chairing sessions, Prof. Mark Morison (Associate-Dean Research in BJBS) for his constant support, advice and encouragements and the Head of School of Computing and Mathematics, Prof. Irfan Altas, for his endless support, advice, leadership and stimulus in this conference.

Finally, we would like to thank everyone who has contributed to the ASPAC2019.

Book of Abstracts



Submission ID 1 (SIDO1):**Modeling for prospect of aman rice production in Dhaka division, Bangladesh**

Sayed Mohibul Hossen¹, ²Md. Takrib Hossain, ³Aditi Chakraborty and ⁴DR. Mohd Tahir Ismail

¹PhD student, School of Mathematical Science, Universiti Sains Malaysia, 11800 USM Penang, Malaysia. Email: m300cht@gmail.com

^{2,3}Department of Statistics, Faculty of Science, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh.

⁴Associate Professor, School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia.

Abstract

Rice is the dominant food crop of Bangladesh, about 75 percent of agricultural land is use for production and it contributes 28 percent of GDP. Aman is one of the main harvest crops and second largest rice crop in the country in respect to the volume of production. The main purpose of this study is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model by Box-Jenkin's approach that could be used to forecast the production of Aman Rice in Dhaka Division Bangladesh. The Secondary data were collected for the year 1972-73 to 2014-15 from the Bangladesh Agricultural Research Council (BARC) for the purpose of model identification and forecast up-to the year 2035 of the identified model. Data sets are checked for whether it is stationary or not through graphical method, correlogram and unit root test. Thus Box-Jenkins approach is applied for determination of ARIMA model. The best selected Box-Jenkin's ARIMA model for forecasting the production of Aman Rice is ARIMA (1,1,1). For residual diagnostics correlogram Q-statistic and histogram and normality test were used. The comparison between the original series and forecasted series shows the same manner which indicates the fitted model behaved statistically well and suitable to forecast the production of Aman Rice in Dhaka Division Bangladesh. We have found that the annual production of Aman Rice in Dhaka Division Bangladesh is slightly increasing.

Keywords: Aman, Correlogram, Unit Root test, ARIMA, Forecasting.

SIDO2:**The role of forecasting the size of school populations in planning for regional education in Australia**

Terence Mills¹, Leonardo Veliz² and Julianne Lynch²

¹Mathematics and Statistics, La Trobe University, Australia

²School of Education, Deakin University, Australia

Abstract

School education policy in Australia is shaped by the priorities of state and federal governments. However, there are signs of a growing interest by governments in Australia in considering regional perspectives in education. This case study is set in City of Greater Bendigo which is a major regional city in Victoria. A key factor in any educational planning exercise is the number of students involved. This is especially important when planning for resources required to meet local demand. The aim of the project is to investigate how the number of school- aged children in Bendigo has been changing over time, and is likely to change in the future. The data used were the estimated resident population of City of Greater Bendigo by 1-year age groups, over the years 2001–2017. These data are available from Australian Bureau of Statistics. Two forecasting methods were used to predict the numbers of children, aged 0–5, 6–12, and 13–18, in Bendigo up to 2028. The first method was Holt's time series method. This method readily leads to prediction intervals that enable us to assess the uncertainty associated with the forecasts. The second forecasting method was based on a simplified cohort component method; this model employs only high- school algebra. Comparisons were also made with official state- level forecasts. The two methods led to comparable forecasts which were consistent with less detailed official forecasts.

Results indicate strong population growth in the pre-school years 0–5, modest growth in the primary school years 6–12, and weak growth in the secondary school years 13–18 during the next decade in City of Greater Bendigo.

This project offers an approach to forecasting that could be readily adapted by other communities, and by the federal government in its National Regional, Rural and Remote Education Strategy.

Keywords: Educational planning, schools, regional policies, forecasting, time series, demography.

SIDO3:**Basic income and optimal taxes. Efficiency, equity, losers and winners.**

Ugo Colombino¹ and Nizamul Islam²

¹Professor Emeritus, Dipartimento di Economia e Statistica, University of Torino, Italy, ugo.colombino@unito.it.

²Research Fellow, Luxembourg Institute of Socio-Economic Research, Department of Living Conditions, Belval, Luxembourg, nizamul.islam@liser.lu (Corresponding author).

Abstract

We use a behavioural microsimulation model embedded in a numerical optimization procedure in order to identify optimal (social welfare maximizing) tax-transfer rules. The rules consist of a universal basic income and a flexible tax approximated by a 4th degree polynomial. We illustrate the emerging equity-efficiency trade-off, the solutions implied by alternative degrees of inequality aversion and the distribution of winners and losers with respect to the current rules. We also explore the design of optimal policies where the number of losers is constrained to be “sufficiently small” and/or “even” across deciles. We compare the results in three countries, Italy, Luxembourg and Ireland.

Keywords: Micro-simulation, Optimal tax, Polynomial approximation

SIDO4:**A tax benefit model for policy evaluation in Luxembourg: LuxTaxBen**

Nizamul Islam¹ and Lennart Flood²

¹Luxembourg Institute of Socio-Economic Research: LISER, Luxembourg; E-mail: nizamul.islam@liser.lu

²University of Gothenburg, Sweden; E-mail: Lennart.Flood@handels.gu.se

Abstract

This paper presents a new tax benefit microsimulation model for Luxembourg. The main distinguishing feature of that model, as compared to existing ones (among which EUROMOD) is that it includes a behavioural module for labour supply. We validate our model by comparing incomes, taxes, and transfers produced by the model with those of EU-SILC and EUROMOD. The labour supply model is validated by studying the fit (comparing observed hours of labour with those predicted by the model). Wage elasticities are reported and compared with similar exercises in the literature. Finally, the model is used for assessing a recent switch of the Luxembourg tax system from joint to individual taxation. By comparing simulated non-behavioural output it is shown that, as a whole, Lux-TaxBen produces an output very close to that produced by EUROMOD and EU-SILC. Further, the behavioural simulation suggests that the reform has no impact on the labour supply for males in couple while a significant number of inactive females change their inactivity status and start working due to the reform. It is also shown that the change in the labour income increases by 0.8% while the welfare dependence decreases by -0.6 due to move from joint to individual tax rules.

Keywords: Micro-simulation, Distributional and behavioural effect, Welfare.

SIDO5:**Delay effect and burden of weather-related tuberculosis cases in Rajshahi province, Bangladesh, 2007–2012**

Md Abdul Kuddus, AITHM, James Cook University, Australia
Emma McBryde, AITHM, James Cook University, Australia
Oyelola Adegboye, AITHM, James Cook University, Australia

Abstract

Tuberculosis (TB) is a potentially fatal infectious disease that continues to be a public health problem in Bangladesh. Each year in Bangladesh an estimated 70,000 people die of TB and 300,000 new cases are projected. It is important to understand the association between TB incidence and weather factors in Bangladesh in order to develop proper intervention programs. In this study, we examine the delayed effect of weather variables on TB occurrence and estimate the burden of the disease that can be attributed to weather factors. We used generalized linear Poisson regression models to investigate the association between weather factors and TB cases reported to the Bangladesh National TB control program between 2007 and 2012 in three known endemic districts of North–East Bangladesh. The associated risk of TB in the three districts increases with prolonged exposure to temperature and rainfall, and persisted at lag periods beyond 6 quarters. The association between humidity and TB is strong and immediate at low humidity, but the risk decreases with increasing lag. Using the optimum weather values corresponding to the lowest risk of infection, the risk of TB is highest at low temperature, low humidity and low rainfall. Measures of the risk attributable to weather variables revealed that weather–TB cases attributed to humidity is higher than that of temperature and rainfall in each of the three districts. Our results highlight the high linearity of temporal lagged effects and magnitudes of the burden attributable to temperature, humidity, and rainfall on TB endemics. The results can hopefully advise the Bangladesh National TB control program and act as a practical reference for the early warning of TB cases.

Keywords: Tuberculosis, Bangladesh, Distributed lag models, Weather

SIDO6:**Increasing access to government microdata for research purposes***

D. Trewin, Australian National University, Australia

Abstract

It is generally accepted that improved access to government held microdata for research and evaluation purposes can lead to better policy analysis and better policy outcomes. There have been very important steps to improving access. Institutes such as the Australian Bureau of Statistics and the Australian Institute of Health and Welfare have legislation that enables research access under prescribed conditions. This creates some limitations which can frustrate researchers. However, it is important for the sustainability of research access that there is an appropriate balance with privacy/confidentiality constraints. Nevertheless, there is scope to change the existing balance. In response to the Productivity Commission's inquiry into data access and data use, there have been other important initiatives to assist researcher access, led by the Department of Prime Minister and Cabinet. These are still largely in the exploratory stage. The main focus of the paper will be to summarise the current initiatives; discuss the legal, technical and cultural barriers; and make suggestions on how to improve researcher access that takes these barriers into consideration.

Keywords: Data access, research and evaluation purposes, microdata, legal, technical and cultural barriers.

SIDO7:**Statistical analysis of big data: The good, the bad and the ugly***

Susan Wilson, School of Mathematics and Statistics, University of New South Wales, Australia

Abstract

Technological advances are producing a tsunami of "big data" of very widely varying types and quality. Challenges abound on how to deal with such data, including wide-ranging deliberations regarding statistical modelling. A major, general, concern is the lack of reproducibility. This problem is ubiquitous, including where results are being used as evidence to drive health care and social policy. In the presentation, an overview will be given of various aspects of (big) data and statistical approaches to its analysis in the modern, interconnected, world.

Keywords: Technology, big data, statistical modelling, health care analysis, social policy evaluation.

SIDO8:**Technical efficiency and value chain analysis of potato: An empirical study in Bangladesh**

Mahfuza Afroj¹, Mohammad Mizanul Haque Kazal², Imtiaz Faruk Chowdhury³ and Md. Mahfuzar Rahman³

¹Department of Agribusiness and Marketing, Faculty of Agribusiness management, Sher-e-Bangla Agricultural University, Sher-e-Bangla Nagar, Dhaka-1207, Bangladesh.

²Department of Development and Poverty Studies, Faculty of Agribusiness management, Sher-e-Bangla Agricultural University, Sher-e-Bangla Nagar, Dhaka-1207, Bangladesh.

³Department of Agronomy, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Sher-e-Bangla Nagar, Dhaka-1207, Bangladesh.

Abstract

This study was conducted to assess the technical efficiency, capital financing and value chain analysis of potato in Bangladesh from November 2015 to April 2016. In total 252 farm households were selected as sample from top four potato growing districts Rangpur, Munshigonj, Bogra and Rajshahi. Additionally, 12-13 small and medium rural traders were selected from each district. In the case of retailers, 61 were chosen, 10 from each district while 21 were selected from the capital, Dhaka. Similarly for urban wholesalers 8 were selected from each district and 18 from Dhaka. Stochastic frontier model was used to analyse data and found that among the different variables K fertilizer, irrigation and seed cost is statistically significant at different levels. Additionally, technical inefficiency model find education, farm size and credit is statistically significant at 5% level of significant. These variables have great impact on efficiency of potato cultivation. The govt. should support on N fertilizer, pesticide and other statistically insignificant variables to increase the efficiency of the farmers at higher level. In potato value chain, farmers, small traders and large traders (*Bapari*) share was 28.57%, 13.07% and 11.58% respectively. The share of *Aratdar* was 15.87% and 8.44% for commission agent. The rural and urban retailers share is 9.23% and 11.48%. The share of the processing company and cold storage was 42.31% and 33.33% respectively. The information obtained in this study will help the policy maker to take such policy that helps to increase the efficiency at farmer's level and develop efficient potato value chain in Bangladesh.

Key words: Technical Efficiency, Marketing Channel, Value Chain.

SIDO9: Statistical Data Analysis without Statistical Significance

Gang Xie (John), Research Office, Charles Sturt University, Wagga Wagga, NSW 2678, Australia

Abstract

In March 2019, American Statistical Association (ASA) proposed a call for “stop using the term ‘statistically significant’ entirely and moving to a world beyond ‘ $p < 0.05$ ’” because many statistically significant results aren’t replicating. This presentation provided our understanding and suggestions on performing statistical data analysis without referring to statistical significance. First, the concerns related to p -value and statistical significance were briefly reviewed through a point-by-point explanation of the six principles of the 2016 ASA’s p -value statement. Simulation evidence and empirical examples were presented to illustrate the reasons why “By itself, a p -value does not provide a good measure of evidence regarding a model or hypothesis”, hence, dichotomization of the data analysis outcomes according to whether or not they possess statistical significance, typically determined by $p < 0.05$, was theoretically flawed; more importantly, the development of a dichotomization mind set was the fundamental reason for (the continuing) misinterpretation and misuse of p -value in specific and of statistical analysis outcomes in general. Finally, a couple of simple examples were demonstrated to show, without statistical significance, researchers could perform statistical data analysis correctly and better by focusing on the scientific interpretation of their results by referring to subject matter theory or the underlying mechanism or professional judgment rather than being blindly hinged on the p -value (which itself was subject to many assumptions that most likely not met in real life situation). A valid statistical data analysis should: do your study and perform your data analysis as planned; report the point estimate/effect size and the associated interval estimate (e.g., a 95% confidence interval) or the associated p -value and interpret them from your subject matter’s perspective.

Unless you could justify your data analysis results were based on a representative sample from a defined target population, any inferential statements would not be statistically grounded.

Keywords: data analysis, p -value, statistical significance, reproducibility.

SID10: Water Supply and Sanitation in Nepal: an Overview

Jitendra Upadhyaya¹

¹Research For Next Generation, Kathmandu University, Dhulikhel, Nepal

Abstract

Water is basic necessity of life. The earth has immense quantity of water, but usable water is very limited. The demand of water is continuously increasing due to increase in population, industrialization, urbanization and living standard of the people. According to a report of the World Bank, by 2025, fifty two countries having two-thirds of the total world population would face shortage of water. Of the total quantity of water, 97.5% is salt water in the sea. Only 2.5% is fresh water which is available in the form of surface and groundwater. The surface water is always the first choice as it is easily and readily available. Ground water is rechargeable to an extent and still it should be conserved as much as possible for the benefits of future generation. The existing coverage of water and sanitation services in Nepal is approximately 80% and 43% respectively. Currently, about 5.5 million people have not accessed safe water supply services and 16 million lack adequate sanitation facilities. Similarly, treated water is available to only 15% people. Remaining 85% population either depend on groundwater and/or surface water sources for their domestic requirements. On the other hand, the available water is getting polluted day by day mostly due to anthropogenic activities; also due to natural processes. Water born diseases have remained one of the major public health concerns in Nepal, which are primarily derived from poor water sanitation and lack of proper hygienic practices. As a consequence, thousands of people annually suffer from water related health problems and many of them loss their life. Poor sanitation and unsafe drinking water has been one of the leading causes of various infectious diseases in Nepal. Water and sanitation related health problems constitute 60% to 80% of total illnesses, respectively. Many of them such as diarrhea, dysentery, intestinal helminthes, jaundice, cholera are endemic in Nepal, particularly in the rural areas. Recently, the increased trend of urbanization has also created pressure on urban sanitation and environmental management. At present, about 60% of urban population has not connected to sewage system and 71% of rural population has not accessed improved sanitation system. To solve the existing situation of water and sanitation, the government has formulated long-term and mid-term plan with appropriate strategy for 2007 to 2020. The target noted in the strategy is to provide 100% coverage of both water supply and sanitation.

Keywords: health problems, sanitation, water pollution, water supply

SID11: Missing continuous outcomes in stepped-wedge trials

Sorif Hossain¹ and Anower Hossain²

¹Institute of Statistical Research and Training (ISRT)

²University of Dhaka, Bangladesh

Abstract

In stepped-wedge cluster randomised trials (SW-CRTs), clusters are sequentially randomised, so that each cluster randomised to the control condition for the unique length of time before switching to the intervention conditions for the remainder of the study. Two commonly used analysis approaches for SW-CRTs are cluster-level analysis and individual-level analysis using linear mixed models (LMM). Missing outcomes are a commonly occurring problem in SW-CRTs which can lead to invalid estimates and misleading inferences if ignored or handled inappropriately. It is plausible to have missing outcomes in SW-CRTs depending on baseline covariates. In this paper, we considered only continuous outcomes and missingness only in outcomes depending on baseline covariates. We investigate analytically and by simulations the validity of cluster level analysis and LMM using complete record analysis (CRA) and multiple imputed data sets. Cluster level analysis using CRA gives a biased estimate unless the missingness mechanism is the same between the two intervention groups. LMM using CRA gives valid estimate regardless of the missingness mechanisms. On the basis of the simulation study and analytical results, we give guidance on the conditions under which each approach is valid.

Keywords: stepped-wedge cluster randomised trials, missing outcome, complete case analysis, multilevel multiple imputation.

SID12:**Assessing competition in the Indian banking sector using panel data**

Zhiheng Li¹, Shuangzhe Liu¹, Fanda Meng¹ and Milind Sathye¹

¹Faculty of Science and Technology University of Canberra, Australia

Abstract

The paper aims to assess the level of competition in the Indian banking sector as well as the impact of the global financial crisis on the competition pattern. We use panel data for the period from 2005–2018. We found that the overall competition in the Indian banking sector is strong, though there are differences by type of bank ownership (that is, private sector, foreign banks and state-owned banks). The various policy measures taken by the Indian government in recent years have helped boost competition.

Keywords: Competition, Indian banking sector, global financial crisis, panel data

SID14: River-based Tilapia Cage culture in Bangladesh: Impacts and Determinants of Adoption

Mohammad Mizanul Haque Kazal¹, and Md. Sadique Rahman²

¹Department of Development & Poverty Studies, Sher-e-Bangla Agricultural University, Sher-e-Bangla Nagar, Dhaka, Bangladesh.

² Department of Management and Finance, Sher-e-Bangla Agricultural University, Sher-e-Bangla Nagar, Dhaka, Bangladesh.

Abstract

Cage culture, an aquaculture technique for tilapia production, expanding quickly in flowing water of rivers and canals in Bangladesh. The present study identifies the determinants of improve cage culture practices adoption and its impact on productivity and profitability. The study employed poisson regression, and inverse probability weighted regression adjustment techniques to achieve the objectives. The findings indicate that the decision to adopt was positively influenced by education (p -value <0.10), extension contact (p -value <0.05), societal membership (p -value <0.05), and number of working person (p -value <0.10) while inversely influenced by farm size (p -value <0.10). Adopters of improve practices received significantly higher productivity and profit compared to non-adopters. More research and investment, as well as modification in extension services and approaches are needed to improve the adoption level and sustain the production. Higher productivity implies higher income, thus may reduce poverty.

Keywords: Adoption; improve practices; impact evaluation; tilapia farming

SID15:**Geospatial and temporal variation of prostate cancer incidence**

Win Wah¹ Arul Earnest¹, Susannah Ahern¹, Sue Evans¹ Melanie Evans¹ and Jeremy Millar²

¹School of Public Health and Preventive Medicine, Monash University, Australia.

²Radiation Oncology unit, Alfred Health, Melbourne, Victoria, Australia.

Abstract

There has been insufficient and mixed findings on the association between geographic and temporal patterns of prostate cancer incidence and possible risk factors. To evaluate geographical and temporal variations in prostate cancer incidence in Victoria, Australia. This study analysed 105349 cases of incident prostate cancer diagnosed among Victorian residents between 1982 and 2016 from the population-based Victorian Cancer Registry. It performed Poisson regression analyses to identify an association between an annual number of prostate cancer counts, Prostate-Specific Antigen (PSA) tests and the elderly male population (≥ 65) after adjusting for population-at-risk and years. It also applied Bayesian spatial-temporal models to determine any association with prostate cancer incidence and area-level factors. The overall trend of the age-standardized prostate cancer incidence was increasing. The highest age-specific incidence was observed among 65–74 over 35 years. Every increase in 1000 PSA tests per 100,000 population, prostate cancer cases increased by 17% (RR=1.17, 95% CI=1.13–1.22). A 1% increase in the proportion of the male population (≥ 65) correlated with a 7% increase in prostate cancer cases (RR=1.07, 95% CI=1.06–1.10). This study showed substantial geographic variation in the age-standardized RR of prostate cancer across postal areas (POA) throughout 2010–2016. Men living in the most socio-economically advantaged POA had a decreased risk of prostate cancer (RR=0.92, 95%CI=0.86–0.98) independent of other areal factors, remoteness, population density, pollution, smoking, indigenous status, and origin of birth. The major limitation is an ecological bias that findings at the population-level cannot be generalizable to individuals. Men living in the most socio-economically advantaged POA were associated with decreased prostate cancer risk, independent of other areal factors. Age-specific risk of developing biological prostate cancer, temporal changes in PSA testing and an increasingly elderly population contributed to an increasing trend of prostate cancer incidence in Victoria, Australia.

Keywords: Cancer incidence; Prostate Cancer; Socio-economic; Spatial-temporal analysis

SID16:

Factors associated with coronary heart disease among elderly in Sylhet region in Bangladesh: A study based on different communities

Kanis Fatama Ferdushi¹, Anton Abdulbasah kamil², Mohammad Nayeem Hasan¹ and Tanjila Islam¹

¹Department of Statistics, Shahjalal University of Science and Technology, Bangladesh.

²Faculty of economics, administrative and social sciences, Istanbul Gelisim University, Turkey.

Abstract

Coronary Heart Disease (CHD) is one of the major causes of morbidity and mortality in many developing countries, including Bangladesh. A stratified random sampling with proportional allocation technique was used to collect data from elderly people from urban, rural, and ethnic areas of the Sylhet region. A total of 230 (110 women and 120 men) people aged 60 years or above from the above mentioned areas were included in this study. A multiple logistic regression model was used to evaluate major risk factors associated with CHD for this sample group. The prevalence of CHD was higher for males than females, at 47.80% and 52.20 % for females and males, respectively. Elderly people in urban areas were significantly (AOR = 4.03; 95% CI: 1.22–3.29) more likely to have CHD as compared to elderly persons living in rural areas. Elderly persons of ethnic origin were found to be less likely to suffer from CHD (AOR = 0.04, 95% CI: 0.01–0.17514.69) in comparison to urban elderly. The risk factors smoking (44.8%) and hypertension (51.3%) were also positively associated with CHD. Elderly persons who reported to exercise regularly were found to have 89% (AOR = 0.11, 95% CI: 0.03–0.50) less risk of suffering from CHD as compared to those who did not exercise regularly. The findings of this study further indicated that factors such as a high BMI (overweight/obese), high sugar intake, high soft drink consumption, diabetes, and mental stress have a significant influence on CHD.

Keywords: Coronary Heart Disease, Elderly, Risk Factor, Adjusted Odds Ratio.

SID17:

Finite mixture modelling approach to identify factors affecting children ever born for 15-49 year old women in Asian country

M. Karimuzzaman¹, M. M. Hossain¹ and A. Rahman²

¹Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

²School of Computing and Mathematics, Charles Sturt University, Australia

Abstract

The number of ever born children is one of the main components of population dynamics that determine the size, structure, as well as the composition of a countries' population. Children ever born refer to the number of children born alive to the person up to a specified reference date and served as a response variable here. A secondary dataset is used in this paper that is obtained from a countrywide representative survey entitled Bangladesh Demographic and Health Survey (BDHS) 2014. This study aims to identify the socioeconomic and demographic factors influencing children ever born to the women of 15-49 years old in Bangladesh. The first attempt of this paper is to identify the best-fitted model among generalized Poisson, Negative Binomial, truncated, COM and finite mixture regression model form. The results suggest that among the model considered in this study Finite Mixture Negative Binomial Regression with three components gives the best-fitted model to estimate the number of ever born children in Bangladesh. It reveals that respondents age, residential status, family size and intention of using contraception have shown positive impact and respondents education, drinking water, toilet facility, religious status, household head age, wealth index, age at first birth, and husband education shows a negative impact on ever born children.

Keywords: Children Ever Born; Poisson, Negative Binomial, Finite Mixture Regression; Bangladesh.

SID18:

An assessment of influencing factors for motherhood during childhood using factor analysis and logistic regression methods

Mohammad Salim Zahangir¹ and Mosammat Zamilun Nahar¹

¹Department of Statistics, University of Chittagong, Bangladesh.

Abstract

Though Bangladesh has achieved great success in family planning as well as maternal and child sector nowadays, it still needs further improvement. This study deals with popular phenomenon motherhood in childhood and its influential factors in Bangladesh. Data are obtained from the 2014 Bangladesh Demographic and Health Survey (BDHS). It can be seen that 62.1% of women age 18 or below become a mother or pregnant in childhood. The relationship between factors obtained by factor analysis and motherhood in childhood is assessed by both linear discriminant and logistic regression analyses. The covariates that are found to be significant by the χ^2 -test are also analysed by binary logistic regression technique for examining their effects on child-bearing in childhood. The analysis reveals that respondent's education, husband's education and age at first marriage are significantly negatively associated and respondent's current age is significantly positively associated with the chance of becoming a mother in childhood. Region, wealth index, husband's occupation and husband's age are also significant to some extent. In conclusion, the prevalence of motherhood in childhood can be reduced by educating women up to secondary or higher levels, alleviating poverty and limiting the provisions of early marriage.

Keywords: Motherhood in Childhood · Factor Analysis · Binary Logistic Regression.

SID19:**Applied Bayesian modeling for assessment of interpretation uncertainty in spatial domains**

Scott McManus¹, Azizur Rahman¹, Ana Horta¹ and Jacqueline Coombes²

¹Charles Sturt University, Australia

²Snowden Technologies, Perth, Australia

Abstract

In the mining industry, code compliant reporting standards for public announcements have been developed setting minimum standards for public reporting of exploration results and mineral resources. These include an assessment of the quality and confidence in the data and work carried out since public reporting aims to provide information that is *material, transparent and competent* to investors. There are four phases required to estimate an mineral resource (preparation, investigation, model creation and validation), and estimation is highly dependent on the accuracy of the preparation stage which is a result of the quality of the geological interpretation given for the mineralization process and current spatial location. Performance of feasibility studies in mining projects has been poor, with a 50% failure rate, 17% of failures are attributable to issues in geological interpretation. This interpretation seeks to spatially define geologically homogenous areas in the resource (spatial domains), corresponding to a single statistical population with a single orientation, where possible. In the estimation workflow, the creation of the spatial domain presents a challenge in terms of assessing the uncertainty in the geological interpretation often due to the manual and subjective interpretation used to guide its creation as well as in spatial domains with several mineralization overprint events. The proposed work investigates a Bayesian method using multivariate quantitative data combined with qualitative data to assess the interpretation uncertainty of classification of borehole intervals to a spatial domain defined by a 3D 'wireframe' or 'rock type' model interpretation using either implicit or explicit modeling techniques.

Keywords: Bayesian, Uncertainty, Spatial domain.

SID20:**A spatially-explicit microsimulation analysis of the proposed sugar tax on population health in the ACT**

A. Richardson, NCEPH, Australian National University (ANU), Australia

A. Lal, NCEPH, ANU, Australia

C. Joseph, Centre for Social Research and Methods, ANU, Australia

B. Phillips, Centre for Social Research and Methods, ANU, Australia

Abstract

The average Australian consumes nearly 10 kilograms of free sugar in beverages every year, yet the consumption of sugar-sweetened beverages (SSB) is known to be a risk factor in chronic health conditions including obesity, type 2 diabetes, and cardiovascular disease. These consumption rates are health impacts are known to vary geographically, and so therefore will the effect of any proposed sugar tax. Baseline estimates of the impacts of this tax need to account for this spatial patterning. This project aimed to construct those baseline estimates of the impact of the sugar tax proposed by the Australian Medical Association. Modelling was conducted for the whole of Australia at the SA2 level. Demographic data was sourced from the Australian Census 2016, and consumption data came from the most recent ABS Household Expenditure Survey. A spatially-explicit, microsimulation model was developed to assess the impact of the proposed sugar tax on tax paid, and SSB consumption. This modelling is feasible from the sugar tax point of view as it has been conducted for US sugar tax proposals. It is also feasible in the Australian context due to the availability of microsimulation models for the Australian tax-paying population. Results of this project shown in choropleth map form will indicate the extent of variability across SA3's in relation to important demographic variables such as socio-economic status. Corresponding maps of health outcomes will also be presented, and options for exploring associations will be described.

Keywords: Microsimulation, public health, spatial variation.

SID22:**Predict neonatal intensive care unit (NICU) discharge date using machine learning techniques**

Arunkumar Namachivayam¹ and Laxman Bidari¹

¹Ashwini Institute of Child Health & Research Centre, India

Abstract

Discharging babies from NICU depends on many factors including non-medical reasons like the need for supportive equipment, parental education, preparedness etc. Attempt is made to describe a model to predict the date in which babies will be ready for discharge in the next few days there by providing sufficient time to foresee the nonmedical reasons for delay in discharge. Data will be collected retrospectively from hospital records including daily clinical notes from NICU. These data will be used to build a supervised machine learning algorithm to predict days to discharge. Random forest will be trained by using observed features and International Classification of Diseases (ICD – 10). Models were trained with different combinations of diagnosis like premature babies, Cardiac problem, Neonatal sepsis, Gastrointestinal disorders and number of observed features. Algorithm will be built by randomly selecting half of the babies data to a training set, and the other half were used as the testing set. Data from daily clinical notes to predict the discharge date from NICU. This model can be improved as more clinical information is included there by precision of prediction improves. Algorithm built for various diagnosis very similarly. Supervised machine learning algorithm using random forest can predicts the discharge date from the NICU. This algorithm will help neonatologist to plan the timely discharge and has the possibility to prevent discharge delays for non- medical reasons.

Keywords: NICU, Machine Learning, Random Forest

SID23:

A study on the impact of Alcoholism on EEG-based Cryptographic Key Generation Systems

Dang Nguyen¹, Dat Tran¹, Wanli Ma¹

¹Faculty of Science and Technology, University of Canberra, Australia

Abstract

Alcoholism is one of the brain disorders that involves in electroencephalogram (EEG) signals and have impact on EEG-based systems. However, this issue has not been investigated. In this paper, we propose an EEG-based cryptographic key generation system using EEG signals and present experiments performed on two groups of subjects, alcoholic and non-alcoholic groups in the Alcoholism database. This EEG-based cryptographic key generation system is based on an assumption that EEG signal is quasi-stationary if the time window is sufficiently short. With this assumption, stable EEG features are extracted to generate cryptographic keys. The impact of alcoholism on performance of the system is analysed based on our experimental results.

Keywords: Data mining, EEG, Authentication, Cryptographic key generation, Alcoholism.

SID24:**Performance Analysis of Clustering algorithms for mineral exploration from Hyperspectral data**

Deepa C¹, Amba Shetty¹ and A.V. Narasimhadhan¹

¹National Institute of Technology Karnataka, India.

Abstract

Recent advances in sensor technology coupled with high computing power has fostered the need for development of efficient algorithms to process hyperspectral data. Hyperspectral remote sensing, with its wide range of application areas is proved to be a promising technology for analysing earth surface materials. The hyperspectral remote sensors simultaneously record data as contiguous narrow bands of information over the electromagnetic spectrum ranging from visible to short wave infrared region. The analysis and interpretation of hyperspectral data is a challenging task because of the need of calibration, information redundancy and dimensionality problem. Hence a systematic and generic approach has to be followed to extract and analyse such data. Clustering plays a vital role in exploring data, creating predictions and avoid anomalies in the data. Clusters with similar characteristics are grouped together using reiterative techniques. As the real world data is growing exponentially, large datasets with little or no prior information can be identified into interesting patterns with clustering. So in this paper, popular clustering algorithms are evaluated on the AVIRIS (Air-born Visible Infrared Imaging Spectrometer) sensor dataset for mineral identification. Principal Component Analysis (PCA) has been used as a dimensionality reduction technique to address the problem of data redundancy. These techniques plays a vital role in mineral exploration since in field observation is more expensive, time consuming and requires more man power. This is extremely important since mineral deposits reflects the economic status of a country.

Keywords: Hyperspectral data, clustering, PCA.

SID25:**Reviewing the past to shape the future of complex and chaotic IT project management in retail**

J. Hayes¹, M. R. Islam¹ and A. Rahman¹

¹Charles Sturt University, Australia

Abstract

IT projects are by nature, complex and chaotic with a significant proportion failing, assessed as 'not meeting requirements', experiencing overruns in time, budget or scope or not determined acceptable by sponsors and stakeholders. This paper presents a literature review, focused on defining a project initiation and governance framework rooted in complexity theory and bound to the Liminal Cynefin framework with the potential to transform IT project management by understanding projects from the intersection of chaos, complexity and constraints theories. The findings would assist decision makers in the project management industry to assess the potential complexity of a project in the Concept, Validate and Plan stages, matching these results with adaptive governance models and adaptive project management leadership in order to improve project outcomes. With the traditional hard paradigm of the project management industry advocating quantitative measures of project success criteria and projects still failing against these measures, an analysis of historical projects against an amalgamation of current developments in chaos, complexity and constraints theories, combined with alignment to the Cynefin framework is proposed.

Keywords: Retail, Complexity Theory, Theory of Constraints, Chaos Theory, Liminal Cynefin

SID26:**A model based approach of measuring methane emission from different rice cultivation systems in a South-East Asian country**

Shamim Mia^{1,2}, Milton Kumar Saha¹, AKM Abdul Ahad Biswas¹, Md. Abdus Sattar¹ and Feike A. Dijkstra²

¹Patuakhali Science and Technology University, Bangladesh.

²School of Life and Environmental Sciences, University of Sydney, Australia.

Abstract

Global warming due to anthropogenic emissions of greenhouse gases (GHGs) is and will be affecting the livelihood of millions of people. Rice cultivation is one of the dominant contributors to methane (CH₄) emission, a GHG that is 28 times more powerful than CO₂. Rice is one of the dominant crops in many Asian countries including Bangladesh. However, a reliable estimate of CH₄ emission from rice cultivation in Bangladesh is not available, particularly considering the different rice types (i.e., local land races, high yielding varieties (HYV), and hybrid varieties) grown under diverse conditions. In the present study, we estimated current CH₄ emission (both rate and amount) from different rice farming systems using the IPCC Tier 1 method. Additionally, we predicted future emission potentials using changing trends in land areas and seasons of rice cultivation. Across all rice types grown and seasons, (i.e., Aus (from March to August), Aman (July to December) and Boro (December to June), the estimated CH₄ emission in 2017 was at 1912 Gg CH₄ yr⁻¹ (with 95% Cis of 921–3353 Gg CH₄ yr⁻¹) while it was estimated at 1120 Gg CH₄ yr⁻¹ in the year 2060, a 41% decreased from 2017 with 1% land migration for non-agricultural activities. We also found statistically significant differences in CH₄ emission rates among types of rice cultivation ($p < 0.01$) and growing season ($p < 0.01$). Averaged across all seasons, the highest CH₄ emission was from hybrid varieties (203 Kg CH₄ ha⁻¹ yr⁻¹), while the lowest (134 Kg CH₄ ha⁻¹ yr⁻¹) from local land races. In contrast, the same local land races showed the highest emission rates when normalized against yield. Across all rice types, the CH₄ emission was estimated to be largest in Boro season accounting ~49% of the total annual emission. Our findings, therefore, provides a deeper insight into CH₄ emissions from rice cultivation systems in Bangladesh.

Keywords: Global warming; CH₄ estimation; Rice types; Rice growing season; Bangladesh

SID27:

Nutrient loading in the river systems around major cities in Bangladesh: A quantitative estimate with consequences and potential recycling options

Shamim Mia¹, Md. Rushna Alam¹, Md. Abdus Sattar¹, and Feike Dijkstra²

¹Patuakhali Science and Technology University, Bangladesh.

²School of Life and Environmental Sciences, University of Sydney, Australia.

Abstract

Biological organisms including human being acquire mineral nutrients for their growth and development. A significant amount of these nutrients remain unused in the left over materials, known as waste and in the metabolic residues i.e., urine and faeces. These nutrients could potentially be a resources for agriculture if recycled and reused. Therefore, it requires a critical examination of nutrient loading and find out possible recycling strategies. In the current study, we estimated the nutrient loading from municipal organic waste and human excreta using linear modelling, explored the potential consequences to ecosystem services and finally proposed several management strategies. Waste and human excreta generation were calculated from the per capita generation rate while nutrient concentrations were considered as the average of literature values. The daily carbon (C), nitrogen (N), phosphorus (P) and potassium (K) loading from municipal organic waste, urine and faeces to the water bodies around the major cities in Bangladesh, respectively, were 3290 t (~5351 t organic matter), 307 t (~591 t urea), 54 (~ 203 t TSP) and 94 (~ 117 MOP) t. The large amount of nutrients dumping into water often reduces dissolved oxygen content resulting loss of biodiversity. Combined pyrolysis and composting of municipal organic waste can reduce the volume of organic waste while nutrient harvesting through adsorption with activated carbon, biochar and zeolite could be potentially used for recycling of nutrients human faces and urine. Moreover, precipitation and recycling of N as isobutyraldehyde-diurea and P as struvite may also be practiced. Biological trapping with algae or other quick growing plants may contribute significantly in recycling of nutrients. Altogether, our study provides a comprehensive understanding of nutrient loading and its potentials recycling options that may help to attain environmental sustainability.

Keywords: mineral nutrients; organic waste; biodiversity; recycling of nutrients; environmental sustainability.

SID28:

Determining Risk Factors of Antenatal Care Attendance and its Frequency in Bangladesh: An Application of Count Regression Analysis

Kakoli Rani Bhowmik¹, Sumonkanti Das^{2,3}, and Md. Atiqul Islam³

¹Leiden University Medical Centre, Leiden, The Netherlands

²Quantitative Economics, Maastricht University, Maastricht, The Netherlands

³Department of Statistics, Shahjalal University of Science & Technology, Sylhet, Bangladesh.

Abstract

Standard Poisson and negative binomial regression models are the common count regression analysis tools for modelling the number of antenatal care (ANC) visits. Two-part (zero and count) models like zero-inflated and Hurdle regression models are recommended for modelling ANC visits with excess zeros. The intra-cluster correlation (ICC) can be accounted by incorporating cluster-specific random intercepts in the corresponding standard and two-part models. The existence of excess zeros in the distribution of ANC visits in Bangladesh raises the issue of identifying a proper count regression model for the number of ANC visits covering the issues of over dispersion, zero-inflation, and ICC in determining the risk factors of ANC use and its frequency. The data have been extracted from the 2014 Bangladesh Demographic and Health Survey. The hurdle negative binomial regression model with cluster-specific random effects at both zero- and count- parts is found as the best fitted model. Women who have poor education status, live in poor households, have less access to mass media, and belong to Sylhet and Chittagong divisions are less likely to use prenatal care and to have more ANC visits. In addition, women who live in rural areas, depend on other family members' decision for taking health care, and have unintended pregnancies had lower tendency to more ANC visits. The findings recommend incorporation of random community effects along with over dispersion and zero-inflation in modelling the ANC data of Bangladesh, and model selection should be model-driven rather than data-driven since practically assumption of structural zeros is tough to meet.

Keywords: Hurdle model, negative binomial model, prenatal care, random effects, uniformity test, zero-inflation.

SID31:**Estimation of Child Undernutrition at Disaggregated Administrative Tiers of a North-Eastern District of Bangladesh: An Application of Small Area Estimation Method**

Sumonkanti Das^{1,2}, Bappi Kumar¹, Md. Zakir Hossain¹, Sabbir Tahmidur Rahman¹ and Azizur Rahman³

¹Department of Statistics, Shahjalal University of Science & Technology, Bangladesh.

²Quantitative Economics, Maastricht University, The Netherlands.

³Charles Stuart University, Australia

Abstract

Children of Sunamganj district located in the north-eastern part of Bangladesh are highly vulnerable to undernutrition and chronic food insecurity due to its geographic location, long-time waterlog, frequent flash floods, and underdeveloped infrastructure. In this study, child undernutrition indicators stunting and underweight are estimated at district, sub-district (Upzila) and union level administrative tiers of Sunamganj district employing the World Bank small area estimation (SAE) method to a Sunamganj household level survey data collected in 2018 and the census 2011 data of Sunamganj. District level prevalence of stunting and underweight are estimated as about 48.5% (95% CI: 45.3–51.7%) and 37.0% (95% CI: 34.6–39.8%) based on the SAE method. At upzila level, stunting varied from 41.0% to 54.9% and underweight varied from 24.0% to 53.4%; while the indicators varied over 19.5–59.7% and 20.2–56.8% respectively at union level. A significant number of unions are found as hotspots of higher underweight and stunting over the north, north-eastern and north-western parts of Sunamganj. Though the southern part of Sunamganj was homogeneous in the upzila level maps of stunting and underweight; significant number of heterogeneous unions are found in the union-level maps. The upzilas belong to the north-ern part particularly closer to the Indian border and haor areas are mostly vulnerable to stunting and underweight. The study findings on disaggregate level prevalence of stunting and underweight might help the concerned government and non-government organizations to prepare and implement aid-related programs on public health and nutrition.

Keywords: Stunting, Sunamganj District, Underweight, Upzila, Union, World Bank Method.

SID32:
On Propensity Score Methodology

Paul Dewick¹ and Shuangzhe Liu¹

¹University of Canberra Canberra ACT 2601, Australia

Abstract

In an observational study, researchers are constantly required to distinguish the effects caused by the assignment of treatment. Propensity score methodology is one way to determine the effects of, and their probabilities, given a vector of observed covariates, which is particularly popular in the fields of medical, pharmaceutical and social sciences. However, there are mixed views for the best methodologies to use and an overall understanding of the propensity score methodology. Also, there is minimal literature for propensity score methods being used within the broader scientific community. Propensity score methodology can be suited to determine effects caused by, not only treatment of pharmaceutical medication, but for “treatment” of some external event, proposed event or interaction within the wider community. For example, the effect on a regional community due to business closure, or a road by-pass would be a reasonable case of how propensity score methods can be further used within the wider scientific community. The main objective of this paper is to demonstrate how propensity score methodology can be used to answer questions on effects caused by an external event or interaction on a community. The propensity score methodology will be given a framework that can be followed, explained and reported that will help allow for robust decision making, planning and policy decisions to be undertaken.

Keywords: Propensity scores, Methodology, Framework, Policy.

SID33:

Consumers adoption behavior prediction through technology acceptance model and machine learning models

Xinying Li¹ and Lihong Zheng²

¹Changchun University of Technology, Changchun, Jilin, China.

²School of Computing and Maths, Charles Sturt University, Australia.

Abstract

This paper is to uncover the key factors that influence purchase intention of customers through analysing technology acceptance theories/models, in the current online-to-offline (abbreviated as O2O) mobile commerce, and to improve the prediction accuracy of consumers adoption behaviour by utilizing machine learning based methods. With a huge amount of smart phone users, O2O mobile commerce derived from electronic commerce (abbreviated as e-commerce) has been growing vastly. There are many research interests has been attracted on online banking, digital wallet, E-tickets, order tracking, supply chain and so on. However, there is little specific study about O2O mobile APP consumers adoption behaviour. Motivated from the commonly used technology acceptance theories/models, especially, the Unified Theory of Acceptance and Use of Technology (UTAUT) model, this paper is to identify key influencing factors of O2O mobile APP consumers adoption behaviour. Then, a new model is proposed as an extended version of UTAUT. The new model has been validated through a survey questionnaire conducted in target groups. More significantly, treating consumers adoption behaviour as a binary classification problem, we apply two different types of machine learning based approaches (Linear Discriminant Analysis (LDA) and Logistic Regression (LR)) to predicate the possible action result by taking into consideration of all influencing factors from the collected survey data. Comparing against several other conventional approaches, Logistic regression shows the better predication accuracy. Hence, it will provide better guidance for promotion strategies in a more productive way.

Keywords: Logistic Regression, O2O, APP, Adoption Behavior.

SID34:**Effect of women's education on Skilled Birth Attendants in South and South East Asia: A Cross-country assessment on Sustainable Development Goal 3.1**

Raaj Kishore Biswas¹, Nurjahan Ananna² and Jahar Bhowmik³

¹School of Aviation, University of New South Wales, Australia

²Ibrahim Medical College, Bangladesh

³Department of Statistics Data Science and Epidemiology, Swinburne University of Technology, Australia.

Abstract

The Sustainable Development Goal (SDG) 3.1 is to “reduce the global maternal mortality ratio (MMR) to less than 70 per 100,000 live births” by 2030. One of the indicators of MMR is the proportion of births attended by a skilled health personnel. To achieve this goal low- and middle-income countries are required to increase the coverage of skilled birth attendants (SBA) for safe delivery during childbirth. This study used the Demographic and Health Surveys (DHS) data and assessed 1,171,731 women aged 15–49 years from 10 countries selected from South and Southeast Asian (SSEA) region to evaluate the status of SDG 3.1 in this region. This paper also evaluated the contribution of women's education on SBA coverage using surveys conducted during the period 1992 to 2017. Logistic regression models were fitted adjusting the survey clusters, strata and sampling weights. Meta-analyses were conducted collapsing the effect sizes and confidence intervals of education on SBA coverage. Cambodia, Indonesia and Philippines had over 80% SBA coverage after 2010, whereas Bangladesh had only 44.7% coverage among the selected countries in SEA. Education of women at all levels (primary, secondary and higher) were significantly associated with SBA coverage, suggesting that education is a key to skilled delivery cares in SSEA region.

Keywords: Sustainable Development Goal, Maternal mortality ratio, Skilled birth attendants, Education, Meta-analysis, Demographic and Health Surveys.

SID35:

Asking Good Questions to Understand Voluntary Enrolments in Mathematics

Ning Li¹

¹Australian Mathematical Sciences Institute, Melbourne Australia.

Abstract

It is of national concern that participation in higher-level mathematics subjects in senior secondary schools has been declining over the last few decades in Australia. As a gateway subject for tertiary studies in Science, Technology, Engineering, and Mathematics (STEM), the persistent decline can impact the country's economy in the long term. Understanding the causes of the decline can inform practice and help shed light on possible solutions. This paper describes the design and validation of a survey instrument for measuring factors that influence students' decisions to continue or discontinue studying mathematics beyond Year 10. Taking a social cognitive perspective, the instrument investigates motivation in subject selection under the assumption that what people think, believe and feel affects how they behave. An initial form of the instrument was developed and piloted to 564 Years 10 & 11 students. The responses were then used to analyze the reliability, factorial structure, and discrimination of the form. Psychometric evidences support the formation of a reduced form on scales of self-concept, self-efficacy, subjective value, anxiety and learning experience in mathematics. The refined form has reliable internal consistency and a clear structure.

Keywords: Instrument, Survey item analysis, Self-efficacy in mathematics

SID36:

Modelling and Analysis of Computer Experiments using a Simple Pendulum Model

Kazeem Adewale Osuolale¹

¹Nigerian Institute of Medical Research, Lagos, Nigeria

Abstract

A computer experiment is an experiment conducted using data obtained from a computer model or simulator in lieu of the physical process. A physics-based experiment known as a simple pendulum experiment was performed to demonstrate a computer experiment. A true model called a computer model of a simple pendulum was used to simulate a real life pendulum experiment. The inputs to the computer code were varied in order to determine the effect of different inputs on the output(s) of a pendulum experiment. The outputs of such computer model is used as a proxy for the real life observations of the study. The focus of this study is to determine the output which is the time it takes the pendulum bob to return to rest. This time is also called the stoppage time in this paper. MATLAB 2016a computer package (www.mathworks.com/) was used for the development of the program that generates the time it takes the pendulum to return to rest. .

Keywords: Computer experiment, Simulator, Simple Pendulum,.

SID37:

Modelling the arrival of plant pathogens: the role of import volumes and biosecurity interventions

R. Duncan, Institute for Applied Ecology, University of Canberra, Australia

B. Jennifer, P. Hulme, J. Cooper, & P. Johnston, Lincoln University, New Zealand

B. Sikes, University of Kansas, United States

Abstract

The movement of people and goods around the world can lead to the spread of unwanted pests and pathogens, which can have large economic and ecological impacts. Many countries try to limit pest and pathogen arrivals by screening goods and people before they enter. But how effective are these biosecurity measures? This can be difficult to assess because we rarely have data on key metrics such as the number of new pathogens arriving and establishing over time. Here, we use a database of all known New Zealand plant pathogen records to estimate the rate at which new fungal pathogens arrived and established on economically important plant species over the last 133 years. The observed rate at which new pathogens are recorded results from both an arrival and a discovery process. We modeled the discovery of new pathogens as a function of sampling effort, which varied over time, conditional on the underlying but unobserved rate of new pathogen arrival in a Bayesian framework. We then examined the underlying rate of pathogen arrival over time as a function of trade volume and passenger arrival rates, and in relation to the implementation of biosecurity measures in different economic sectors (crop, pasture, horticulture and forestry). Our modelling revealed that pathogen arrivals in New Zealand increased exponentially over the first 100 years, paralleling an increase in the volume of goods imported, but the arrival rate has slowed since 1980 despite increasing trade and passenger arrivals. However, these recent trends differed among economic sectors. Pathogen arrivals of crop and pasture plants have declined but continued to increase on forestry and fruit trees. This trend reflects difference in the biosecurity measures imposed, suggesting that targeted biosecurity can reduce the establishment of nonnative pathogens even while global trade and travel continue to increase.

Keywords: biosecurity, Bayesian modelling, invasion ecology, observation model

SID38:

Can Data Fusion increase the Performance of Action Detection and Recognition in the Dark?

Anwaar Ulhaq^{1,2}

¹School of Computing and Mathematics, Charles Sturt University, Australia.

²Centre of Applied Informatics, Victoria University, Australia.

Abstract

Automated human action detection and recognition is a challenging research problem due to the complexity of its data. Contextual data provides additional cues about actions like if we know car and man, we can short-list actions involving car and man i.e driving, opening the car door etc. Therefore, such data can play a pivotal role in modelling and recognizing human actions. However, the visual context during night is often badly disrupted due to clutter and adverse lighting conditions especially in outdoor environments. This situation requires the visual contextual data fusion of captured video sequences especially if such videos are used for automated recognition of any human activities. In this paper, we have explored the significance of contextual data fusion for automated human action recognition in night-time video sequences. For this purpose, we have proposed an action recognition framework based on contextual data fusion, spatiotemporal feature fusion and correlation filtering. We have performed experimentation on multi-sensor night vision video data from infra-red (IR) and visible (VIS) sensors. Experimental results show that contextual data fusion based on the fused contextual information and its colourization significantly enhances the performance of automated action recognition.

Keywords: Contextual data, outdoor environments, automated recognition, spatiotemporal feature fusion, correlation filtering

SID39:**Seasonal occurrence of amoxicillin, ampicillin and penicillin residues in informal milk supply chains**

Naveed Aslam^{1,5}, Sosheel S Godfrey¹, Mateen Abbas², Muhammad Y Tipu², Muhammad Ishaq³, David M McGill^{3,5}, Hassan M Warriach³, Muhammad Husnain⁴, Peter C Wynn¹

¹Graham Centre for Agricultural Innovation Charles Sturt University, Wagga Wagga 2650 NSW, Australia.

²Quality Operation Laboratory, University of Veterinary and Animal Sciences, Lahore 54000 Pakistan.

³ASLP Dairy Project, University of Veterinary and Animal Sciences, Lahore 54000 Pakistan.

⁴Department of Clinical Medicine and Surgery, University of Veterinary and Animal Sciences, Lahore 54000 Pakistan

⁵Faculty of Veterinary and Agricultural, University of Melbourne, Werribee 3030, VIC, Australia.

Abstract

Informal milk marketing chains provide the major milk supply link from smallholder dairy production systems to consumers in developing countries. Prevalence risk of antibiotic residues was investigated in milk samples (n=528) collected from different levels of informal supply chains in Pakistan from 2012–2013. After screening, all positive samples were further analyzed by High Performance Liquid Chromatography to quantify individual β -lactam residues. Fifteen percent of the total samples were found to be positive for at least one of the antibiotics. All positive samples (81/528) were positive for amoxicillin. Percentage of positive samples for ampicillin and penicillin was 12.1 and 6.4% respectively. Percentages of positive samples collected from farmers, small collectors, large collectors and retailers were 17.5, 15.1, 8.3 and 13.5 respectively. When relating to season of collection 11.3, 10.2, 19.1, 17.9 and 16.3% of samples in autumn, monsoon, spring, summer and winter were found positive. Concentrations of amoxicillin, ampicillin and penicillin in positive samples were 79.5 ± 12.15 , 106.6 ± 12.49 and $13.7 \pm 4.2 \mu\text{gkg}^{-1}$ respectively. Occurrence of these high levels of drug residues demand mass awareness programs associated with the implementation of legislation and its enforcement designed to protect the consumer.

Keywords: Beta lactams; Traditional milk marketing chains; smallholder dairy production; antimicrobial residues

SID40:

Evaluating Faster-RCNN and YOLOv3 for Target Detection in Multi-sensor Data

Anwaar Ulhaq^{1,2}, Asim Khan² and Randall Robinson²

¹School of Computing and Mathematics, Charles Sturt University, Australia.

²College of Engineering and Science, Victoria University, Australia.

Abstract

Intelligent and autonomous systems like driverless cars are seeking the capability to navigate around at any time of the day and night. Therefore, it is vital to have the capability to reliably detect objects to predict any situation. One way to capture such imagery is through multisensor data like FLIR (Forward Looking Infrared) and visible cameras. Contemporary deep object detectors like YOLOv3 (You Look Once Only) [1] and Faster-RCNN (Faster Region based Convolutional Neural Networks) [2] are well-trained for daytime images. However, no performance evaluation is available against multi-sensor data. In this paper, we argue that diverse contextual multi-sensor data and transform learning can optimise the performance of deep object detectors to detect objects around the clock. We explore how contextual multi-sensor data can play a pivotal role in modelling and recognizing objects especially at night. For this purpose, we have proposed the use of contextual data fusion on available training data before training these deep detectors. We show that such enhancement significantly increases the performance of deep learning based object detectors.

Keywords: Object Detection, YOLOv3, Faster-RCNN, Deep Learning

SID41:**Semantic vegetation detection in repeat photography for environmental data analysis**

Asim Khan¹, Anwaar Ulhaq^{1,2}, Mobeen ur Rehman³ and Randall W Robinson¹

¹Institute for Sustainable Industries and Liveable Cities, Victoria University, Australia

²Charles Sturt University, NSW, Australia

³Avionics Department, Air University, Islamabad, Pakistan.

Abstract

Environment change being one of the contemporary global issues needs special attention of the researchers. With the advancement in computer vision researchers are equipped enough to come up with algorithms accomplishing automated system for environment monitoring. This paper proposes an algorithm which can be used to observe the change in vegetation utilizing the images of a particular site over a period of time. This would help the environment experts to put on their efforts in a right direction and right place to improve the environmental protection. The proposed algorithm registers the images so that the comparison can be carried out in an accurate manner using unified frame- work for all the images. Registration algorithm aligns the new images with the existing images available in the record of the same particular environmental site by performing transformation. Registration process is followed by semantic segmentation process which segments out the vegetation region from the image. A novel approach towards segmentation is proposed that performs the classification between vegetation patches and non-vegetation patches. The proposed algorithm showed promising results with F-measure of 85.36%. The segmentation result leads us to robust calculation of vegetation index that can be used to make a vegetation growth record regarding a particular site.

Keywords: Vegetation Index, Image Registration, Image Segmentation, SVM, Flucker Post Dataset, F-Measure

SID42:**Policy into practice; statistics the forgotten gatekeeper**

I. Hume¹

¹NSW Department of Primary Industries, Australia

Abstract

The Australian Federal Government Carbon Farming Initiative (CFI) is a voluntary carbon offsets scheme. It allows land managers to earn carbon credits by changing land use or management practices to store carbon. The Carbon Farming Futures (CCF) program ran from 2012 to 2017 to identify where farmers can boost productivity and profitability; improve soil and reduce greenhouse gas emissions. The Action on the ground program was one component of the CCF which aimed to assist farmers to try new technologies and practices on their farms. Laboratory and field plot research indicated that ploughing nutrients into the soil with crop residues/stubble increased the amount of soil carbon stored. Our project was to see if this was possible using farm equipment. The challenge was to design experiments with plots large enough to be managed with commercial farm machinery that deliver enough precision to test different treatments. We tested carbon sequestration methods in 14 fields in different bioregions. Soil variability was estimated using two electromagnetic surveying instruments. Geostatistics and finite mixture models identified areas of similar soil in fields where randomized block experiments were established. The field experiments failed to reproduce the high sequestration rates of the earlier more controlled research. Reasons for this are proposed. However, the experimental design methods are applicable to this hybrid demonstration/research approach.

Keywords: Soil carbon, experimental design, large plots, demonstration, policy.

SID43:**Progress and controversy in statistical modelling of the genomic architecture of complex traits***David Balding¹¹Schools of BioSciences and of Mathematics and Statistics, University of Melbourne, Australia**Abstract**

One of the great challenges in science today is to model the relationship between the DNA sequence of an individual and outcomes of interest, such as yield in crops or health status in humans. The goals of this modelling include both prediction and understanding of mechanisms. The genome-wide association study (GWAS) has for over a decade been providing relevant data, but the underlying processes are complex, with many genetic variants each of weak effect. Initially attention was limited to one-variant-at-a-time analyses but the resulting models had limited explanatory power. Simultaneous analyses of GWAS variants provided a big step forward, but introduced new difficulties due to the problem of overfitting due to too many predictors. More recently, models that allow joint analyses of single-variant association statistics have provided a further advance, due to their effectively unlimited sample size and avoidance of confidentiality issues associated with individual-level data. However, these models introduce further difficulties that have not yet been fully resolved. I will review recent progress by our group and others in using genome-wide SNPs to assess the heritability of complex human traits, the distribution of that heritability across the genome, the genetic correlation of pairs of traits and the effect of confounding on GWAS test statistics.

Keywords: complex traits, statistical genetics, computational statistics, genomic prediction, heritability.

SID44:**Optimal policy modelling and welfare policy settings in Australia**

B. Phillips¹

¹Centre for Social Research and Methods, Australian National University, Australia

Abstract

The complexity of the social security system makes it challenging for policy makers to assess what changes should be made to the system to achieve policy objectives, and the implications of changes to the system. This paper describes the results of an initial attempt to develop a new methodology and modelling tool for optimising the social security system to achieve a particular outcome. The illustrative case used is minimising relative income poverty. We do this by using a microsimulation approach in which we alter welfare payments (or other parameters) to minimise household poverty, subject to a range of constraints, such as the overall social security budget or relationships between payment rates. The relationship between payment rate and poverty gap is then estimated using a linear regression model that provides parameter values for an equation that describes how changes in payment rates affect the poverty gap. This equation can be used to determine 'optimal' payment rates, subject to constraints such as a budget constraint or changes from current payment levels.

Keywords: Social security, household poverty, microsimulation approach, poverty gap, budget constraint.

SID45:**Using individual based modelling to inform policy: Where we have come from, and where big data will lead us***

R. Tanton¹

¹NATSEM, University of Canberra, Australia

Abstract

Modelling that uses individual records (eg, microsimulation, agent based modelling) is a powerful tool to inform policy, as it allows a policy maker to look at the distributional impacts of a change in policy, for example, the impact of a change in taxes on low income families, or couple families with children. This type of modelling also allows the introduction of behavioural change, estimated at the individual level. This address will look at how these types of models have been used to inform policy in the past, and how the greater availability of data (for example, linked datasets and big data) might be used in these models to inform policy in the future.

Keywords: microsimulation modelling, distributional impacts, agent based methods, behavioural change, big data.

SID46:**Sensor-centric source inference and image clustering for supporting various decision making process***C-T. Li¹¹School of Information Technology, Deakin University, Australia**Abstract**

Similar to people identification through human fingerprint analysis, digital forensics through device fingerprint analysis has attracted much attention amongst scientists, practitioners and law enforcement agencies around the world in the past decade. Device information, such as model and serial number, stored in the EXIF are useful for identifying the devices responsible for the creation of the images in question. However, stored separately from the content, the metadata in the EXIF can be removed and manipulated at ease. Device fingerprints deposited in the content by the device provide a more reliable alternative to aid forensic investigations. The hardware or software of each stage in the digital image acquisition process leaves artifacts in the content that can be used as device fingerprints to identify the source devices. This talk will start with an introduction to various types of device fingerprints, their applications and limitations. A more focused presentation on the application of sensor pattern noise, as a form of device fingerprint, to source device verification, common source inference, source device identification and source-oriented image clustering will then be delivered. An interesting and challenging clustering problem with the number of classes far greater than the size of classes will also be discussed. Finally, how all of these data science techniques and results can assist to policy decisions would be illustrated.

Keywords: fingerprint analysis, digital forensics, source devices, clustering methods, policy decisions.

SID47:**Finding significant determinants and impacts of farm-level integrated pest management practices using statistical tools**

Md Sadique Rahman¹

¹Department of Management and Finance, Sher-e-Bangla Agricultural University, Bangladesh.

Abstract

The study aims to identify the determinants of adoption of IPM and its impacts on productivity and pesticide applications. The study employed two-part hurdle poisson regression, propensity score matching (PSM), and inverse probability weighted regression adjustment (IPWRA) techniques to achieve the objectives. The findings indicated that level of integrated pest management (IPM) adoption was low. Around 15% and 10% of the respondents adopted poultry refuse and sex pheromone trap, respectively while about 11% adopted both the practices. Decision to adopt significantly influenced by training ($p < 0.01$), neighbour farmers adoption ($p < 0.01$) and distance from highway ($p < 0.05$) while extent of adoption depend on extension contact ($p < 0.05$) and neighbour farmers adoption ($p < 0.05$). Adoption of IPM significantly reduced pesticide applications and increased productivity compared to non-adopters based on kernel and radius matching. Adoption analysis suggested that more research and field demonstrations are required to improve the adoption level. Reduced pesticide applications may have some environmental benefits. Due to higher productivity, there is scope to boost the role of IPM in anti-poverty policies in Bangladesh.

Keywords: Impact evaluation, pesticide application, poisson hurdle model, propensity score matching, vegetable farming.

SID48:

Parental perception on children online safety; a study on parents of different ethnic communities in Australia

A. Imran¹

¹School of Information, Technology (IT) & Systems, University of Canberra, Australia

Abstract

The exponential growth and the use of internet in all spheres of human life poses new challenges and risks for the young children growing up in the digital age, with potential short and long-term ramifications to our society and the future. Parents play an integral role in the development of their children's attitudes and behaviours. Yet studies indicate that adults are not adequately mitigating the range of cyber risks posed to children where parent-oriented solutions to cyber risks are extremely inadequate. This study attempts to fill previous research voids by understanding the status and nature of parental perception of their children's online use and the risks. This study, as part of the Children Online Protection project, surveyed 205 parents of different ethnic communities in Australia. The results indicate that parents do not understand the scope of cyber threats in relation to their children and are ill equipped to monitor and mitigate the risks posed. The study also identifies the gaps in strategy and initiatives by government and schools about the cyber risks and discuss possible way outs and calls for parent inclusive solutions for different ethnic communities to better protect their children.

Keywords: Parental perception, children online safety, cyber risk.

SID49:**Using a spatial farm microsimulation model for Victoria to estimate the impact of an external shock on farmer incomes**

Y. Vidyattama¹ and R. Tanton¹

¹NATSEM, University of Canberra, Australia

Abstract

Greater uncertainty in climate and market conditions have raised real questions for Australian Government's at different levels on how to allocate and distribute support to reduce the levels of financial stress for farmers. Modelling of on and off farm incomes can provide information to help understand the spatial distribution of changes in income as a result of some policy change. A spatial model of farmer incomes operating at the unit record level allows us to identify policies to help support these farmers. This study discusses the development of a spatial microsimulation model to identify concentrations of financial stress in the Australian State of Victoria. It further shows how an external shock such as drop in the price of milk can be incorporated into the model to help understand its impact on the spatial distribution of farmer incomes and poverty rates across Victoria.

Keywords: microsimulation, farming, incomes.

SID50:**Wavelet-Based Quantile density function estimation under random censorship**

Esmail Shirazi¹ and Hassan Doosti²

¹Gonbad Kavous University, Iran

²Macquarie University, Australia

Abstract

In this paper, the estimation of a quantile density function in the presence of right censored data is investigated. We propose and develop a new wavelet-based methodology for this problem. In particular, an adaptive hard thresholding wavelet estimator is constructed. Under mild assumptions on the model, we prove that it enjoys powerful mean integrated squared error properties over Besov balls. While existing estimators of the quantile density function are not good at the tails, our proposed estimators perform well at the tails. The comparison of the proposed estimator has been made with estimators given by Jones (1992) and Soni (2012) graphically and in terms of the mean integrated square error (MISE) for the uncensored case.

Keywords: Random censorship, Wavelets, Quantile function, Rate of convergence

SID51:

Computing robust statistics via an EM Algorithm

Maheswaran Rohan¹

¹NSW Department of Primary Industries, Wagga Wagga, Australia.

Abstract

Maximum likelihood is perhaps the most common method to estimate model parameters in applied statistics. However, it is well known that maximum likelihood estimators often have poor properties when outliers are present. Robust estimation methods are often used for estimating the model parameters in the presence of outliers, but these methods lack a unified approach. We propose a unified method using EM algorithm to make statistical modelling more robust. In this paper, we describe the proposed method of robust estimation and demonstrate it using the example of estimating the location parameter. Well known real data sets with outliers were used to demonstrate the application of proposed estimator. Finally, the proposed estimator is compared with standard M-estimator. In this talk, the location case was considered for simplicity, but directly extends to the robust estimation of parameters in a broad range of statistical models. Hence this proposed method aligns with the classical statistical modelling, in terms of a unified approach.

Keywords: M-estimator, Mixture distributions, Improper distribution.

SID52:

Performance Analysis of Clustering algorithms for mineral exploration from Hyperspectral data

Deepa C¹, Amba Shetty¹ and A.V. Narasimhadhan¹

¹National Institute of Technology Karnataka, India.

Abstract

Recent advances in sensor technology coupled with high computing power has fostered the need for development of efficient algorithms to process hyperspectral data. Hyperspectral remote sensing, with its wide range of application areas is proved to be a promising technology for analysing earth surface materials. The hyperspectral remote sensors simultaneously record data as contiguous narrow bands of information over the electromagnetic spectrum ranging from visible to short wave infrared region. The analysis and interpretation of hyperspectral data is a challenging task because of the need of calibration, information redundancy and dimensionality problem. Hence a systematic and generic approach has to be followed to extract and analyse such data. Clustering plays a vital role in exploring data, creating predictions and avoid anomalies in the data. Clusters with similar characteristics are grouped together using reiterative techniques. As the real world data is growing exponentially, large datasets with little or no prior information can be identified into interesting patterns with clustering. So in this paper, popular clustering algorithms such as k-means, fuzzy c-means and self organising maps are evaluated on the AVIRIS (Air-born Visible Infrared Imaging Spectrometer) sensor dataset for mineral identification. Principal Component Analysis (PCA) has been used as a dimensionality reduction technique to address the problem of data redundancy. These techniques plays a vital role in mineral exploration since in field observation is more expensive, time consuming and requires more man power. This is extremely important since mineral deposits reflects the economic status of a country.

Keywords: Hyperspectral data, clustering, PCA, k-means, fuzzy c-means, self organising maps.

SID53:**Data privacy and security in the cloud**

Peter Padiet¹, Rafiqul Islam¹, Azizur Rahman¹,

¹Charles Sturt University, Australia.

Abstract

Cloud computing have slowly made it way and has gained popularity with its services and products and they have repeatedly increased, and many users are attracted to migrate their data into the cloud. But there are still issues which concern users when outsourcing their data and the business application into the cloud, security and privacy are very critical and created trust concern. There are 5 attributes related to security and privacy as follow; privacy-preservability, availability, integrity, confidentiality and accountability. The aim of this paper is to improve security in the cloud because the uses of cloud services has been more increasing and in the next few years would be likely the cloud services will be a big driving force, better to be prepare and develop majors to encounter and security loophole that may arise in the near future.

Keywords: Cloud, privacy, security, SaaS;

SID54:**Working Zones 2016**

Karen Malam¹

¹Bureau of Infrastructure, Transport and Regional Economics

Abstract

The Bureau of Infrastructure, Transport and Regional Economics (BITRE) have developed Australian 'Working Zone' (WZ) regions based on the 2016 Census Journey-to-Work data. WZs are mutually exclusive regions delineated to reflect the commuting patterns of Australian workers. WZs are useful for spatial analysis of economic, social and policy issues at a regional level because they reflect the actual geographic behaviour of individuals, as opposed to other administrative and political boundaries. They are particularly useful for analysing labour markets, because individual WZs have minimal work-based commuting flows either into or out of adjoining WZs.

Keywords: regional, economic, functional zones

SID55:**Adapting truck GPS data for freight metrics**David Mitchell¹¹Bureau of Infrastructure, Transport and Regional Economics**Abstract**

Road freight data to inform infrastructure planning and transport policy has historically been collected using traditional sample survey based methods, owing to the large size and varying scope of the industry. However, such surveys are financially costly, impose a significant burden on respondents, and the availability of results can significantly lag collection. Moreover, while such surveys can provide reasonable detail at large area level and major corridors, they do not provide sufficiently accurate information to address all policy and planning issues facing governments and industry. The increasing uptake of digital technologies in the freight transport sector offers an alternative data sources with which to inform policy. The Bureau of Infrastructure, Transport and Regional Economics (BITRE) has developed processes and systems to use telematics data from private sector firms to provide information about the Australian road freight network and industry to assist decision making by firms, government and the public. This paper provides an overview of the project, outlines how the BITRE collects, processes and stores telematics data, and presents examples of some of the measures and outputs that can be produced using such data.

Keywords: road freight, GPS data, telematics,**SID56:****The ABC of Big Data**Siu-Ming Tam¹¹Methodology Division, Australian Bureau of Statistics**Abstract**

The statistical challenges in using big data for making valid statistical inference in finite population have been well documented in literature. These challenges are due primarily to statistical bias arising from under-coverage in the big data source to represent the population of interest and measurement errors in the variables available in the data set. By stratifying the population into a big data stratum and a missing data stratum, we can estimate the missing data stratum by using a fully responding probability sample, and hence the population as a whole by using a data integration estimator. By expressing the data integration estimator as a regression estimator, we can handle measurement errors in the variables in big data and also in the probability sample. Finally, we develop a two-step regression data integration estimator to deal with non-response in the probability sample. An advantage of the approach advocated in this talk is that we do not have to make unrealistic missing-at-random assumptions for the methods to work. The proposed method is applied to an example using 2015-16 Australian Agricultural Census data.

Keywords: Measurement error; Non-response; Regression estimation; Selection bias.

SID57:**Herbicides and plant hormesis: a case study in *Conyza sumatrensis***

Md Asaduzzaman¹, Eric Koetz¹ and Hanwen Wu¹

¹NSW Department of Primary Industries, Wagga Wagga, Australia

Abstract

Herbicides are main tool for weed control in modern crop production system. However, weed resistance to herbicides is a topic of growing concern among farmers and weed scientists. Determination of herbicide resistance is the first step in management of resistant weeds, where levels of resistance needs to be assessed through herbicide dose response assay and other tactics. In herbicide dose response bioassay, some substances, although toxic at higher doses, can be stimulatory at low doses. This biphasic dose-response phenomenon is commonly termed hormesis. This study investigated the hormesis responses of tall fleabane (*Conyza sumatrensis* L) to herbicide paraquat. The results disclose that paraquat resistant *Conyza sumatrensis* shows hormetic growth increase at two lower doses of paraquat, leading to fitness enhancements of plant height, above-ground biomass and buds plant⁻¹ over untreated plants. The hormetic effects resulted in higher fitness at vegetative with an increase of 52%; 23% and 80%; 65% above-ground biomass over control for resistant biotypes Bio-1 and Bio-2 respectively. Consequently both populations generated higher numbers of buds (33 to 66% in Bio-1 and 46% to 78% in Bio-2) plant⁻¹ under two lower doses of paraquat. This enhanced relative fitness during the vegetative stage would increase the competitiveness in resistant weeds against neighbouring crops.

Keywords: Weeds, dose, stimulation, herbicides

SID57:**Current research and statistical demands within DPI**

Deb Slinger¹

¹NSW Department of Primary Industries, Wagga Wagga, Australia

The presentation will give examples of current research activities under her portfolios, promoting statistical methods and valuing statistical needs within DPI and examples of decision making, made in agricultural science based on statistical outputs. On the other hand, current development of centralising research data using database will be briefly discussed. Finally, the presentation of the challenges / importance of data from collection through to big data outputs using the research programs within this portfolio that influence adoption of agricultural practices for agricultural consultants and growers.

Keywords: statistical methods, big data outputs, agricultural practices, decision making.

SID59:

Multilevel modeling for precision public policy

S.V. Subramanian¹

¹Harvard University, USA

Abstract

Child sex ratio is a marker of disproportionate sex ratio at birth and discriminatory practices that lead to differential survival in early childhood by sex. This talk will demonstrate how we used the 2011 Census on rural India to present the first local analysis of CSR across 587,043 villages using our multilevel modelling methods. The magnitude of local variation in CSR was heterogeneous across states/union territories and districts. Our findings provide timely evidence to inform localized programmes to be implemented with greater precision.

Keywords: multilevel analysis, CSR, public policy, India.

SID60:

Spatial microsimulation and multilevel methods for precision public policy

A. Rahman¹

¹Charles Sturt University, Australia

Abstract

This presentation will address two major health related issues in Australia and provides an account of the methodologies that have been used by researchers for estimating small area health related characteristics across the globe. Findings reveal that both spatial microsimulation and multilevel modelling methods can be used for micro-geospatial targeting for precision public policy.

Keywords: spatial microsimulation modelling, multilevel methods, reweighting, health and obesity, Australia.

