South African Statistical Association Conference 2015

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

57th Annual Conference
29 November – 2 December 2015

57th South African Statistical Association Conference 2015

Programme and Abstracts

Statistics South Africa

The South Africa I know, the home I understand

sas THE POWER TO KNOW

COMMUNICATING THROUGH DATA VISUALISATION

National statistical offices are an important source of information for evidence-based decision-making. However, the standard methodology of releasing statistics makes it difficult for the average citizen to comprehend the importance and value of official statistics to their lives. The use of data visualisation techniques is a growing international trend that has made statistics more accessible to the person on the street.

Innovating dissemination: Census 2011 results

One of Statistics South Africa's (Stats SA) key strategic objectives is to develop new and innovative statistical products and services to respond to increased user demand. This innovation began with the release of the Census 2011 results. Instead of the usual text-heavy, static presentation, the Census 2011 presentation made use of animated graphics to bring the message across.

On 30 October 2012, Stats SA became the first NSO to release census data using an iPad application. The Stats SA application, which is available on the iTunes store, was first used to disseminate Census 2011 information. Its library has since expanded to include releases on both economic and social statistics. Stats SA plans to launch an Android app before the end of 2015.

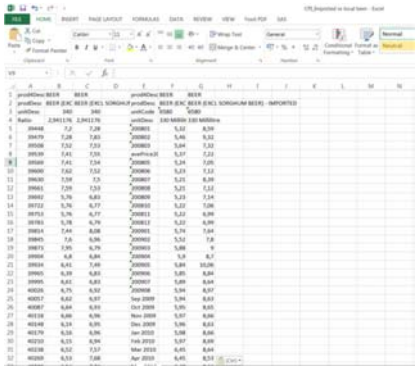
A light from afar: revamping the Stats SA website

Following the positive response to the way in which Census 2011 results were disseminated in terms of interactive presentation and the new products that were developed, the Statistician-General launched a project in February 2013 to radically redesign Stats SA's website in a way that would make it easier for data users to find the data that they were interested in, and that would address the concerns that they have.

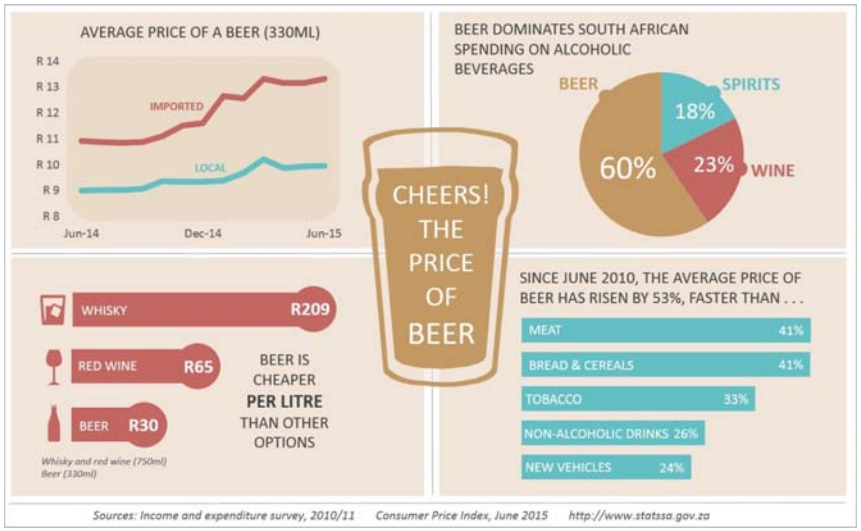
The new-look Stats SA website was launched on 01 August 2013. The innovative website was well received and garnered a number of mentions in online articles.

Taking statistics to the people

Data visualisation aims to take statistical publications that are text and graphic-heavy



and, using infographics, turn it into visualisations that are easily understood by all, especially those who do not have a background in statistics.



Telling the stories behind the data

There is often more to statistics than meets the eye. Behind the text and tables lurk interesting stories that are just waiting to be told. The Communications team works closely with subject matter specialists to identify and tell these stories, and sometimes find out interesting facts.

Reaching more platforms

The simple language and data visualisation lends itself to being used on various media platforms. Stats SA has active Facebook and Twitter accounts, which are well utilised to communicate with various market segments. The mainstream media also utilises the data stories and data visualisations to tell the story of statistics in a simple yet powerful way.

Creating a statistical literate society

In his speech at the opening of the 50th Anniversary Conference of the South African Statistical Association, former Finance Minister, Trevor Manuel, had the following to say:

Our best endeavours are not about numbers, they are about people and the quality of the lives of even the poorest. This is the measure of civilisation. Statistics can make an enormous difference to the quality of democracy. Innumeracy is the enemy of democracy. People familiar with numbers and facts can measure progress in their own lives and are empowered to speak about what remains to be done.

Data visualisation makes a contribution towards building a statistical literate society by translating statistics into an understandable format.

Statistics South Africa

The South Africa I know, the home I understand

This conference would not have been possible without the generous support of the following sponsors:

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The South Africa I know, the home I understand

CASIO

1. Introduction

Globally, 2015 is being recognised as a year of monumental movements, happenings, and remembrances. The United Nations declared 2015 the International Year of Light, it has now been 70 years since the end of WWII, and on the South African front, the first Statistics lecture at the University of Pretoria was given 90 years ago, Department of Statistics at the University of Pretoria celebrates its 76th birthday and hosts the 57th Annual Conference of the South African Statistical Association in collaboration with StatsSA.

The SASA Conference of 2015 promises to be a stimulating and invigorating experience for scholars, students, and industry experts alike, with a multitude of high-profile statisticians and academic leaders attending this conference from all over the world.

The LOC has embarked on revamping the previous format of the conference for 2015 in order to encourage high quality contributions by the 300-400 delegates who are expected to attend. Approximately 30 international experts are attending that will form an integral part of the plenary and parallel special sessions – thereby setting an unprecedented level of excellence that few academic conferences across South Africa could compare with favourably. This conference aims to promote the broad variety of statistical areas that can (and should) be studied in South Africa, and assist with the crucial development of analytical skills needed for the country. By having the international speakers present it can greatly assist to make sure that South Africa stay on top of the game of international academic trends in the field of statistics, and ensure a competitive and stimulating research- and industry-based environment for South Africa.

2. General Information

Registration
Registration for the conference is on Sunday 29 November 2015 3pm – 5pm before the official opening at 5pm in the Aula (building 17 on the map), Hatfield Campus, University of Pretoria.

The information desk will be open the rest of the conference in the Eng III foyer.

Parking
Please enter the University on University Road and park in the parkade on level 3 (building 82 on the map) - keep left at the university gate entrance on University Road. Note that you cannot enter this parkade from other entrances of the university. You will be given a SASA conference parking ticket for each day at registration. On your first day, take a parking ticket from the machine and replace it with the ticket received at registration. For the rest of the days use the SASA2015 parking ticket to enter and exit. Only return the ticket to security at the end of each day. There are only 300 spaces available. Additional cars can park in open parking on campus (note that all undercover parking is reserved for University of Pretoria staff).

WIFI Access
Delegates can connect to TuksGuest on the WIFI network. You will need to register each day to access this network.

Delegates can also connect to the Eduroam wireless network. If you are visiting UP and your home institution participates in Eduroam, you should be able to get free Internet access at UP by simply connecting to the Eduroam wireless network at any of our on-campus hotspots. You will need to authenticate with your home institution's credentials.

Name Tags
Delegates are requested to wear their name tags at all times. Delegates without name tags will not be allowed into the venues and social events.

Emergency Numbers
Campus Security: (012) 420-2310 / 2760
Sonette (official conference): 083 287 3945

3. Social Events
    - **Meet and Greet, Data Summit** Saturday 28 November 2015 17:00: Plant Science Rooftop (building 83)
    - **Conference Opening** Sunday 29 November 2015 17:00: AULA (building 17)
    - **Poster Evening** Monday 30 November 2015 18:00 – 21:00: ENG III foyer (building 82)
    - **Young Statistician's Function** Monday 30 November 2015 21:00 - late : Oom Gert's (building 57)
    - **Gala Dinner** Wednesday 2 December 2015 16:00 - late : Rautenbach Hall (building 17) and the Brooklyn Theatre afterwards for a Christmas Concert and craft beer. There are buses to the theatre, as well as back to the campus afterwards, but you are also welcome to drive yourself.

      Brooklyn Theatre address: Greenlyn Village Centre, C/o Thomas Edison and 13th Streets, Menlo Park.

4. Venues
    - **Conference Venue:** ENG III (building 82)

5. SASA 2015 Organising Committee

Chair: Dr Inger Fabris-Rotelli
Committee Members:
Prof Andriette Bekker
Mr Andre Swanepoel
Mr Johan Ferreira
Official Organiser: Mrs Sonette Olivier sonette@savannaskills.co.za

LOC Competition 2015
The LOC (Local Organising Committee) of SASA2015 has decided to launch a small competition in the run-up to the SASA2015 conference. This competition carries a prize of R 1 500 sponsored by Statomet and only registered delegates may enter.

Question:

What is the probability that you (the entrant) will win this competition?

You are requested to derive/describe $P(win)$ in an innovative way, and you may make any assumptions you deem necessary for the calculation. You can use any reference (textbook/internet/etc.) as long as you reference all sources. Novel/interesting approaches are expected and complete and sufficient descriptions of your approach are required.

Some notes:

- There are approximately 350 delegates for this year's conference;
- Only typed entries will be accepted;
- Delegates are allowed to enter as a group;

- The competition closes on Tuesday 1 December before the first morning session, and submissions to be received by the registration table of the conference before the start of the morning session;
- The winner(s) will be announced at lunch on Tuesday 1 December. Winners will be required to appear in person at this event;
- Members of the LOC will be the judges of this competition. Any member of the LOC is not permitted to take part in this competition. Their decision will be final and they reserve the right not to indicate a winner.

Any additional queries may be directed to Johan Ferreira (johan.ferreira@up.ac.za).

6. Guidelines to Speakers

Please take important note of the following guidelines to all speakers and chairs of sessions. Check carefully where your talk is scheduled and ensure you are aware of your specific guidelines.

Please note there are various types of sessions at SASA this year:

1. **Special sessions:** These are focused sessions with a mini-plenary and a discussion at the end. There are NO questions in between talks.

- A delegate is advised to attend the entire session and participate in the discussion at the end in which questions can be directed at a specific speaker or to the research in general.
- Speakers in a special session must remain in the special session for the duration of the sessions. Speakers should:

- Double check the date and time of your presentation.
- Load your presentation on the computer **before** the start of the session.
- Poster presentations do not have slides. A 5 minute slot in which you talk about your research is allocated. Your poster should be displayed in the foyer for the entire morning of afternoon where the respective special session is scheduled and the presenters should be near his/her poster during the tea slot of that morning or afternoon. Please report to the registration desk for directions to hang your poster. Someone will be there to assist.
- ***Posters should be portrait, A1 size and laminated. Other posters formats cannot be accommodated.***
- Report to the chairperson of the session before the start of the session.
- Keep to the time allocated for your presentation.
- You are not allowed to move your presentation to any other time slot.
- **Once the chair indicates the end of your session, you must stop your presentation immediately.**

2. Open sessions (Statistical methodology and techniques sessions): These are the traditional SASA sessions with questions in between talks.

- Delegates are welcome to move between these sessions but should be aware that there isn't time allocated for this. Speakers should:
  - Double check the date and time of your presentation
  - Load your presentation on the computer **before** the start of the session
  - Report to the chairperson of the session before the start of the session
  - Keep to the time allocated for your presentation
  - You are not allowed to move your presentation to any other time slot
  - **Once the chair indicates the end of your session, you must stop your presentation immediately.**

3. Young Statistician sessions: These are talks by Doctoral and Masters students partaking in the Young Statistician's competition. Every talk in this session will be judged and the prizes awarded at the gala dinner on Wednesday 2 December 2015 (Competition details at: http://sastat.org.za/sasa2015/student-competitions).

Delegates are welcome to move between these sessions but should be aware that there isn't time allocated for this. Speakers should:

- Double check the date and time of your presentation
- Load your presentation on the computer b**efore** the start of the session
- Report to the chairperson of the session before the start of the session
- Keep to the time allocated for your presentation
- You are not allowed to move your presentation to any other time slot
- **Once the chair indicates the end of your session, you must stop your presentation immediately.**

4. Poster Evening: Masters and honours students partake in the poster evening on Monday 30 November 2015. There will be prizes awarded for the top posters at the gala dinner on Wednesday 2 December 2015 (Competition details at: http://sastat.org.za/sasa2015/student-competitions). ***Posters should be portrait, A1 size and laminated. Other posters formats cannot be accommodated.*** Posters should be hung at 2pm on Monday 30 November 2015 (report to the registration desk). Judging will occur before the evening begins.

Chairpersons for all sessions should:

- Double check the date and time of your session.
- **Keep to the scheduled times.**
- No changes are to be made to the programme.
- Check the attendance of all the speakers, and ensure that all presentations have been loaded on the computer before the start of the session.
- Welcome delegates and speakers at the beginning of your session.
- Make the following announcements:
    i. All cell phones to be switched off.
    ii. State the programme for the session.
    iii. Start with the first lecture.
- Warn speakers 5 minutes before the end of their allocated time.
- Thank all speakers and delegates at the end of the session.
- Report to the front desk if a speaker was absent.
- Report shortcomings to the session assistant.

7. Sponsor Demonstrations

Wolfram Mathematica – applications in Data Science

Presenter: Clemens Dempers (Blue Stallion Technologies)

Wolfram Mathematica has a 25 year track record of innovation, integrating numerics, symbolics and graphics with curated data. Some of the later developments added geographic computation, clustering, random process analysis, social media analysis and real time data visualization.

The presentation will include live demonstrations of the technology, including machine learning, semantic data analysis, and spatial data visualization

Programme at a Glance

NOTE: **SURNAME**** indicates a SASA Conference 2015 Proceedings Paper

OPENING – SUNDAY 29 NOVEMBER 2015 Chair: Prof James Allison (SASA President)

18:00 – 18:15 Welcome (Prof Allison)

18:15 – 18:30 Prof Anton Ströh (Vice-Principal, Institutional Planning, University of Pretoria)

18:30 – 18:50 Pali Lahohla, Statistician General and SASA Prize giving (Student Prizes – Prof Delia North) (Sichel Medal – Prof Paul Fatti) (SAS Thought Leader – Prof Paul Mostert) (SASA Honorary Members - Prof Paul Mostert)

18:50 – 19:05 Minister Jeff Radebe, The Presidency, For Planning, Monitoring and Evaluation (Chair: Pali Lahohla)

19:05 – 19:45 Presidential Address - Prof James Allison, SASA President (Chair: Prof Francesca Little)

19:45 – 20:00 Entertainment

20:00 – 20:10 SAS Sponsorship Address

20:10 – 20:50 Plenary Address: Prof Bob Rodriguez, SAS

20:50 – late Cocktail Function in the Aula Foyer

COE-MASS Session: National Doctoral Training Centre for Mathematical and Statistical Sciences (1 December 2015 8am)

The main objective is to put forward a proposal to explore models for postgraduate training in Mathematical and Statistical Sciences in response to some of the recommendations in the report of the Review of Mathematics Research in South Africa in 2009. There are examples both locally and internationally that have been successfully implemented. Amongst other things these would

1. Facilitate poling existing expertise across the South African university sector to provide breadth and depth in postgraduate training;
2. Address the crisis in academic Statistics;
3. Create a pipeline of PhD-ready students and train and graduate cohorts of PhDs;
4. 'Flood' the market with graduates that have sophisticated and strong quantitative skills.

	Saturday 28 November 2015	Sunday 29 November 2015	Sunday 29 November 2015
8:00 - 9:00	Registration, tea and coffee (Plant Science Auditorium foyer)	Registration, tea and coffee (Brown Lab, Informatorium)	Registration, tea and coffee (Plant Science Auditorium foyer)
9:00 - 10:30	Data Science Summit (Plant Science auditorium)	Education Workshop (Brown Lab, Informatorium)	Text Analytics Workshop (Plant Science Auditorium)
10:30 - 11:00	Tea and Coffee (Plant Science Auditorium foyer)	Tea and Coffee (Brown Lab, Informatorium)	Tea and Coffee (Plant Science Auditorium foyer)
11:00 - 12:30	Data Science Summit (Plant Science auditorium)	Education Workshop (Brown Lab, Informatorium)	Text Analytics Workshop (Plant Science Auditorium)
12:30 - 13:30	Lunch (Plant Science Auditorium foyer)	Lunch (Brown Lab, Informatorium)	Lunch (Plant Science Auditorium foyer)
13:30 - 15:00	Data Science Summit (Plant Science auditorium)	Education Workshop (Brown Lab, Informatorium)	13:00 Lunch with LOC, EC, Plenaries and Mini-Plenaries off campus

15:00 - 15:30	**Tea and Coffee (Plant Science Auditorium foyer)**	**Tea and Coffee (Brown Lab, Informatorium)**	**Registration, Tea and Coffee (Aula) 15:00-17:00**
15:30 - 17:00	**Data Science Summit (Plant Science auditorium)**	**Education Workshop (Brown Lab, Informatorium)**	
17:00	**Meet and Greet Cocktail (Light) Function for Data Science Summit Delegates (Plant Science Roof Top)**		**Opening Function (Aula) 17:00 - late: Chair: Prof James Allison (SASA President)**

	Monday 30 November 2015	Monday 30 November 2015	Monday 30 November 2015	Monday 30 November 2015	Monday 30 November 2015	Monday 30 November 2015	Monday 30 November 2015
VENUE	ENG III - 1	ENG III - 2	ENG III - 3	ENG III - 4	ENG III - 5	ENG III - 6	
8:00 - 8:30	Arrival Tea and Coffee (Foyer ENG III)						
8:30 - 10:15	Special Session Part 1 : Special Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa (NORTH, WILD,, KRAAMWINKEL, SCOTT, HAZRA, FLETCHER) Chair: NORTH	Special Session: Offical Statistics (MANKWE, BOTHA, LETSOALO, MASIMULA, SIKHOSANA, NKWINIKA, KGOTHE, MOLATA, CHATINDIARA, MASENYA, SHABANGU, MAZIBUKO, MULIBANA) Chair: NAIDOO	Special Session: Risk theory in finance and actuarial science (RAUBENHEIMER, JOUBERT, VISAGIE, ADEKAMBI**, KONING, CHINHAMU**, KEMDA, KASEKE) Chair: BEYERS	Special Session: Fit In or Fall Out: Statistical Distributions (BALAKRISHNAN, MARQUES, SANTANA**, VAN STADEN, SEKEH, OLUYEDE, LOOTS, MIJBURGH, IIYAMBO, AKARAWAK, ADELEKER, OMACHAR) Chair: MARQUES	Special Session: The Analysis of Data from Clinical Trials (LOMBARD, GROBLER, LEASK, RAMJITH, GUMEDZE) Chair: LITTLE	Special Session: Statistical Image Processing and Robotics (KING, LAU, BIERMAN) Chair: FABRIS-ROTELLI	Young Statistician's Stream (NEMUKULA, DIRIBA**, RABE, DUDENI-TLHONE, RAS**) Chair: DR LEONARD SANTANA
10:15 - 10:45	Tea and Coffee with poster exhibitions (Foyer ENG III)						
10:45 - 11:45	Special Session Part 1 : Special Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa (NORTH, WILD,, KRAAMWINKEL, SCOTT, HAZRA, FLETCHER) Chair: NORTH	Special Session: Offical Statistics (MANKWE, BOTHA, LETSOALO, MASIMULA, SIKHOSANA, NKWINIKA, KGOTHE, MOLATA, CHATINDIARA, MASENYA, SHABANGU, MAZIBUKO, MULIBANA) Chair: NAIDOO	Special Session: Risk theory in finance and actuarial science (RAUBENHEIMER, JOUBERT, VISAGIE, ADEKAMBI**, KONING, CHINHAMU**, KEMDA, KASEKE) Chair: BEYERS	Special Session: Fit In or Fall Out: Statistical Distributions (BALAKRISHNAN, MARQUES, SANTANA**, VAN STADEN, SEKEH, OLUYEDE, LOOTS, MIJBURGH, IIYAMBO, AKARAWAK, ADELEKER, OMACHAR) Chair: MARQUES	Special Session: The Analysis of Data from Clinical Trials (LOMBARD, GROBLER, LEASK, RAMJITH, GUMEDZE) Chair: LITTLE	Young Statistician's Stream (TWABI, MLINDE) Chair: DR ANDREHETTE VERSTER	
11:45 - 12:45	Plenary: Director Rademacher (ENG III-1) Chair: Dr Arul Naidoo						

12:45 - 13:30	Lunch (Rautenbach Hall) (SASA EC Meeting (ENG III-1))						
13:30 - 15:15	Special Session: Managing the global hunger challenge: food security measurement and monitoring in the Sustainable Development Goal era (CAFIERO, NGOMANI, SHABALALA, HENDRIKS) Chair: SCHMIDT	Special Session: Offical Statistics (RADERMACHER, LAHOHLA, NAIDOO, SAPS) Chair: NAIDOO	Special Session: Extreme Value Theory (BEIRLANT, VAN DER MERWE, VERSTER, DIRIBA, MINKAH, MAPOSA, KAOMBE) Chair: KIJKO	Special Session: Bayesian network applications and distributed reasoning systems (PAVLIN, DE WAAL, KOEN, GOODALL, DABROWSKI, CLAESSENS, UDOMBOSO) Chair: DE WAAL	Special Session: Multilevel Modelling (STANCEL-PIATAK, BATIDZIRAI, NEL, STRASHEIM, HOOBLER) Chair: CRAFFORD	Young Statistician's Stream (NUMAPAU GYAMFI, KHENENE, KAMPER, MAGAGULA,) Chair: DR FRANCK ADEKAMBI	
15:15 - 15:45	Tea and Coffee with poster exhibitions (Foyer ENG III)						
15:45 - 16:45	Special Session: Managing the global hunger challenge: food security measurement and monitoring in the Sustainable Development Goal era (CAFIERO, NGOMANI, SHABALALA, HENDRIKS) Chair: SCHMIDT	Special Session: Offical Statistics (RADERMACHER, LAHOHLA, NAIDOO, SAPS) Chair: NAIDOO	Special Session: Extreme Value Theory (BEIRLANT, VAN DER MERWE, VERSTER,DIRIBA**, MINKAH, MAPOSA, KAOMBE) Chair: KIJKO	Special Session: Bayesian network applications and distributed reasoning systems (PAVLIN, DE WAAL, KOEN, GOODALL, DABROWSKI, CLAESSENS, UDOMBOSO) Chair: DE WAAL	Special Session: Multilevel Modelling (STANCEL-PIATAK, BATIDZIRAI, NEL, STRASHEIM, HOOBLER) Chair: CRAFFORD	Young Statistician's Stream (FERREIRA, MAKONI, MAKGAI) Chair: DR RENE EHLERS	
16:45 - 17:45	Plenary: StatsSA: Pali Lahohla (ENG III-1) Chair: Dr Arul Naidoo						
18:00 - 21:00	Poster Evening for young Statisticians (Foyer ENG III)						
21:00 onwards	Young Statistician's Pizza Evening (Oom Gerts) (Speaker: Dr Schalk Human (Nedbank, University of Pretoria, Master of Ceremony: Johan Ferreira))						

	Tuesday 1 December 2015	Tuesday 1 December 2015	Tuesday 1 December 2015	Tuesday 1 December 2015	Tuesday 1 December 2015	Tuesday 1 December 2015	Tuesday 1 December 2015
VENUE	**ENG III - 1**	**ENG III - 2**	**ENG III - 3**	**ENG III - 4**	**ENG III - 5**	**ENG III - 6**	**ENG III - 7**
7:30 - 8:00	**Arrival Tea and Coffee (Foyer ENG III)**						
8:00 - 9:45	**Special Session: Bayesian statistical modelling (LESAFFRE, MOSTERT, BURGER, LOQUIHA, MARTINS) Chair: MOSTERT**	**Special Session: Business Analytics in Data Science (LEE, BURRA, FATTI, HALL) Chair: KANFER**	**Special Session: Experimental Design (HAINES, DEBUSHO**, STEFFENS) Chair: DEBUSHO**	**Special Session: Multivariate Analysis in Economic and Management Sciences (BIESMANS, GUPTA, PILLAY, BOATENG, MANGISA) Chair: LITVINE**	**Young Statistician's Stream (BATIDZIRAI, SMIT, MPHEKGWANA, CLAASSEN, KHUBHEKA) Chair: PROF KIJKO**	**Young Statistician's Stream (VAN NIEKERK, MASOUMI KARAKANI, PAZI, GEMECHU) Chair: PROF ABRIE VAN DER MERWE**	**Statistical methodology and techniques Session: Official Statistics MASEMOLA, MOSOMA, KEKANA, MOTSEPA, PHAKEDI) Chair: DR HERMI BORAINE**
9:45- 10:15	**Tea and Coffee with poster exhibitions (Foyer ENG III)**						
10:15- 11:15	**Special Session: Bayesian statistical modelling (LESAFFRE, MOSTERT, BURGER, LOQUIHA, MARTINS) Chair: MOSTERT**	**Special Session: Business Analytics in Data Science (LEE, BURRA, FATTI, HALL) Chair: KANFER**	**Young Statistician's Stream: Spatial Statistics (RITCHIE, NAIDOO, KRAAMWINKEL) Chair: PROF CHRISTIEN THIART**	**Special Session: Multivariate Analysis in Economic and Management Sciences (BIESMANS, GUPTA, PILLAY, BOATENG, MANGISA) Chair: LITVINE**	**Statistical methodology and techniques Session: Forecasting (OTEKUNRIN, HOLLOWAY, VILJOEN) Chair: PROF GARY SHARP**	**Young Statistician's Stream (KAMBO, WINNAAR, MOTHUPI) Chair: DR MARIEN GRAHAM**	**Statistical methodology and techniques Session: Official Statistics (AYELE, ADEOGUN, MAREMBA) Chair: DR GRETEL CRAFFORD**
11:15- 12:15	**SASA AGM (ENG III-1)**						
12:15 - 13:15	**Lunch (Rautenbach Hall) (Announcement of LOC Competition Winner and Department of Statistics, University of Pretoria Birthday Celebration)**						

13:15 - 14:15	**Plenary: Prof Chen (Eng III-1) Chair: Prof Francesca Little**						
14:15 - 16:00	**Special Session: Biostatistics (CHEN, ZELL, MANDA,MWAMBI, JORDAAN, KABERA) Chair: DEBUSHO**	**Special Session: Business Analytics in Data Science (RODRIGUEZ, ALI, SMITH**; BROMLEY-GANS) Chair: DAS**	**Special Session Part 2 : Special Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa (SJOLANDER, MATIZIROFA, ZONDO, MUTAMBAYI, SWANEPOEL, RAUBENHEIMER, CORBETT) Chair: NORTH**	**Special Session: Bayesian Stream (RUBIN, MALTITZ, VAN NIEKERK, MANJOO, LOUGUE) Chair: RAUBENHEIMER Discussant: MALTITZ**	**SAS honours project winner: DE VILLIERS AND BEZUIDENHOUT ** Statistics SA paper competition winner: JANSE VAN RENSBURG**	**Young Statistician's Stream (MESIKE, SSEKISAKU, JOZI, MIENIE) Chair: DR PAUL VAN STADEN**	**Statistical methodology and techniques Session: Official Statistics (NDLOVU, KONDOWE, MLINDE, PILLAY, CHIFURIRA) Chair: PROF TERTIUS DE WET**
16:00 - 16:30	**Tea and Coffee with poster exhibitions (Foyer ENG III)**						
16:30 - 17:30	**Special Session: Biostatistics (CHEN, ZELL, MANDA,MWAMBI, JORDAAN, KABERA) Chair: DEBUSHO**	**Special Session: Business Analytics in Data Science (RODRIGUEZ, ALI, SMITH**, BROMLEY-GANS) Chair: DAS**	**Special Session Part 2 : Special Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa (SJOLANDER, MATIZIROFA, ZONDO, MUTAMBAYI, SWANEPOEL, RAUBENHEIMER, CORBETT) Chair: NORTH**	**Special Session: Bayesian Stream (RUBIN, MALTITZ, VAN NIEKERK, MANJOO, LOUGUE) Chair: RAUBENHEIMER Discussant: MALTITZ**	**Young Statistician's Stream: Statistical Process Control (MALELA-MAJIKA, CHAKRABORTY, RAMJITH) Chair: DR MARIEN GRAHAM**	**Young Statistician's Stream (KIRKLAND**, KHULUSE-MAHANYA) Chair: DR PRAVESH DEBBA**	**Young Statistician's Stream (PRETORIUS); Statistical methodology and techniques Session: (KIKAWA) Chair: PROF NICO CROWTHER**

17:30 - 18:30	**Plenary: Prof Rubin (Eng III-1) Chair: Dr Lizanne Raubenheimer**
18:30 - 19:30	**Committee Meetings: 1. MDAG (ENG III-1) 2. Biometrics (Eng III-2)**

	Wednesday 2 December 2015	Wednesday 2 December 2015	Wednesday 2 December 2015	Wednesday 2 December 2015	Wednesday 2 December 2015	Wednesday 2 December 2015
VENUE	**ENG III - 1**	**ENG III - 2**	**ENG III - 3**	**ENG III - 4**	**ENG III - 5**	**ENG III - 6**
7:30 - 8:00	**Arrival Tea and Coffee (Foyer ENG III)**					
8:00 - 9:45	**Special Session: Applications of Stochas Processes (LEVITIN, FINKELSTEIN, YADAVALLI, KAOMBE, LITVINE) Chair: FINKELSTEIN**	**Special Session Part 1: Statistical Process Control (QIU, KUMAR, GRAHAM) Chair: CHAKRABORTY**	**Special Session: Statistics in Sport (SWARTZ, DAS, LEMMER, CALDER, JORDAAN) Chair: SHARP**	**Special Session: Complex Sampling (HEERINGA, NEETHLING, RIAZ, MALEPE, CHINOMONA, KISAKU-LWAYO, MAREMBA) Chair: dE WET, Discussant: DE WET, NEETHLING**	**Special Session: Spatial Statistics (STEIN, THIART, DISTLLER, OKANGO, MANDA, NGWENYA, DARIKWA, KHAN) Chair: THIART, Discussant: THIART, FABRIS-ROTELLI**	**COE-MASS Session: National Doctoral Training Centre for Mathematical and Statistical Sciences (Dr Andrew Kaniki, Prof Loyiso Nongxa)**
9:45 - 10:15	**Tea and Coffee with poster exhibitions (Foyer ENG III)**					
10:15 - 11:15	**Special Session: Applications of Stochas Processes (LEVITIN, FINKELSTEIN, YADAVALLI, KAOMBE, LITVINE) Chair: FINKELSTEIN**	**Special Session Part 1: Statistical Process Control (QIU, KUMAR, GRAHAM) Chair: CHAKRABORTY**	**Special Session: Statistics in Sport (SWARTZ, DAS, LEMMER, CALDER, JORDAAN) Chair: SHARP**	**Special Session: Complex Sampling (HEERINGA, NEETHLING, RIAZ, MALEPE, CHINOMONA, KISAKU-LWAYO, MAREMBA) Chair: dE WET, Discussant: DE WET, NEETHLING**	**Special Session: Spatial Statistics (STEIN, THIART, DISTLLER, OKANGO, MANDA, DE KLERK, NGWENYA, DARIKWA, KHAN) Chair: THIART, Discussant: THIART, FABRIS-ROTELLI**	**Statistical methodology and techniques Session: General (RANGANAI, LUBBE, NEMUKULA**) Chair: PROF FRANCESCA LITTLE**
11:15 - 12:15	**Lunch (Rautenbach Hall)**					

12:15 - 14:00	**Special Session: Multivariate data Anlay (VICHI, SWANEPOEL, UYS, NTUSHELO, NIENKEMPER-SWANEPOEL, RABE, MATLWA, KHULE, SALANE) Chair: LUBBE**	**Special Session Part 2: Statistical Process Control (VAN DER MERWE, SHONGWE, CHAKRABORTY, MALELA-MAJIKA, ADEOTI, BADHLYERA) Chair: HUMAN**		**Special Session: Statistics in Nanoscience and Chemical Risk Assessment (GOTTSCHALK, JACOBS, HAYWOOD) Chair: JACOBS**	**Statistical methodology and techniques Session: General (MACDONALD, SANDROCK, RANGANAI, PRETORIUS, SWANEPOEL) Chair: DR LIZANNE RAUBENHEIMER**	**Statistical methodology and techniques Session: Biostatistics (DLAMINI, MUSEKIWA, JUGA, MAJAKWARA, THIEBAUT) Chair: DR CARL LOMBARD**
14:00 - 14:30	**Tea and Coffee with poster exhibitions (Foyer ENG III)**					
14:30 - 15:30	**Special Session: Multivariate data Anlay (VICHI, SWANEPOEL, UYS, NIENKEMPER-SWANEPOEL, RABE, MATLWA, KHULE, SALANE) Chair: LUBBE**	**Special Session Part 2: Statistical Process Control (VAN DER MERWE, SHONGWE, CHAKRABORTY, MALELA-MAJIKA, ADEOTI, BADHLYERA) Chair: HUMAN**		**Statistical methodology and techniques Session: General (SANDROCK, KING, KIFLE) Chair:PROF PAUL MOSTERT**	**Wolfram Mathematica – applications in Data Science**	
16:00 - late	**Gala Dinner (Rautenbach Hall) Chair: Dr Inger Fabris-Rotelli (prize giving for Young Statistician's session and Poster Evening, Thank yous)**					

	Thursday 3 December 2015	Thursday 3 December 2015	Thursday 3 December 2015	Thursday 3 December 2015	Thursday 3 December 2015	Friday 4 December 2015
VENUE	ENG III - 1	ENG III - 2	ENG III - 3	ENG III - 4 (and ENG III - 5, 6, 7 and IT 5-16 for the breakaway)	Brown Lab, Informatorium	Brown Lab, Informatorium
8:00 - 9:00	Arrival Tea and Coffee (Foyer ENG III)				Arrival Tea and Coffee (Brown Lab, Informatorium)	Arrival Tea and Coffee (Brown Lab, Informatorium)
9:00 - 10:30	Biostatistics Workshop (Chen)	Complex Sampling Workshop (Heeringa)	Merging Game Theory and Risk Analysis Optimal Defense of Complex Stochastic Systems (Levitin)	Mentorship Workshop	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression
10:30 - 11:00	Tea and Coffee (Foyer ENG III)				Tea and Coffee (Brown Lab, Informatorium)	Tea and Coffee (Brown Lab, Informatorium)
11:00 - 12:30	Biostatistics Workshop (Chen)	Complex Sampling Workshop (Heeringa)	Merging Game Theory and Risk Analysis Optimal Defense of Complex Stochastic Systems (Levitin)	Mentorship Workshop	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression
12:30 - 13:30	Lunch(Rautenbach Hall)				Lunch (Informatorium)	Lunch (Informatorium)
					Business Statistics using SAS Enterpri	Business Statistics using SAS Enterpri

13:30 - 15:00	Biostatistics Workshop (Chen)	Complex Sampling Workshop (Heeringa)	Merging Game Theory and Risk Analysis Optimal Defense of Complex Stochastic Systems (Levitin)	Mentorship Workshop	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression
15:00 - 15:30	Tea and Coffee (Foyer Eng III)				Tea and Coffee (Brown Lab, Informatorium)	Tea and Coffee (Brown Lab, Informatorium)
15:30 - 17:00	Biostatistics Workshop (Chen)	Complex Sampling Workshop (Heeringa)	Merging Game Theory and Risk Analysis Optimal Defense of Complex Stochastic Systems (Levitin)	Mentorship Workshop	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression	Business Statistics using SAS Enterprise guide ANOVA, Regression, and Logistic Regression

Special Sessions

Special Session Part 1
Monday 30 November 2015 8:30 – 10:15 & 10:45 – 11:45

Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa
Chair: Prof Delia North

- **Prof Delia North**, University of Kwazulu-Natal (co-author: Zewotir, T (School of Mathematics, Statistics and Computer Science, UKZN)) **(20 minutes)**:
  Statistical Capacity Building: Can We Ignore The Online Revolution?

Mini-Plenary:

- **Prof Chris Wild**, University of Auckland, New Zealand **(40 minutes):**
  Data to Insight: Prototyping next-generation introductory statistics

Presentations:

- **Christine Kraamwinkel**, University of Pretoria (co-author: Corbett, AD (Department of Statistics, University of Pretoria)) **(20 minutes):**
  Placing the computer in the students' court
- **Dr Leanne Scott**, University of CapeTown **(20 minutes):**
  Reviewing our blend of online and offline learning at introductory level, UCT
- **Annapurna Hazra**, University of Kwazulu-Natal **(20 minutes)**:
  Simulation-Assisted Teaching for Undergraduates in Statistics
- **Dr Lizelle Fletcher**, University of Pretoria *(co-author: Reyneke, F (Department of Statistics, University of Pretoria))* **(20 minutes):**
  The trials and tribulations of moving towards online teaching

Special Session Part 2
Tuesday 1 December 2015 14:15 – 16:00 & 16:30 – 17:30

Statistics Education Session: Creating on-line Teaching Materials for Teaching Introductory Statistics in South Africa

Presentations:

- **Dr Morné Sjölander (20 minutes):**
  The impact of using multimedia on students' academic achievement in theoretical Mathematical Statistics courses at UFS
- **Lyness Matizirofa**, University of Johannesburg **(20 minutes):**
  Predictors of success and failure in Statistics
- **Nombuso Zondo**, University of Kwazulu-Natal **(20 minutes):**
  Student Attitudes Towards Statistics
- **Ruffin Mutambayi**, University of Fort Hare *(co-authors: Odeyemi , A.O (Department of Statistics, University of Fort Hare), Ndege, J.O (Department of Statistics, University of Fort Hare), Mjoli, Q.T(Department of Industrial Psychology, University of Fort Hare)* **(20 minutes):**
  Statistical analysis of students' attitudes towards statistics: A case study of undergraduate Bachelor of Science students

Short Presentations with a poster:

- **Andre Swanepoel**, University of Pretoria (*co-authors: Engelbrecht, J (Department of Science, Mathematics and Technology Education, University of Pretoria); Harding, A (Department of Mathematics and Applied Mathematics, University of Pretoria)and Fletcher, L (Department of Statistics, University of Pretoria))* **(5 minutes):**
  Which Threshold Concepts exist in First Year Statistics courses at the University of Pretoria?

- **Dr Jacques Raubenheimer (5 minutes):**
  A Comparison Of Rubric Scoring Methods

Panel Discussion (60 minutes):

Prof D. North (UKZN) – CHAIR
Prof. C. Wild (Auckland University, NZ)
Prof. J. Allison (NWU)
Dr. L. Fletcher (UP)
Dr. F. Reyneke (UP)
Dr. Y. Chhana (Wits)
Ms. H. Scott (UCT)

Special Session
Monday 30 November 2015 8:30 – 10:15; 10:45 – 11:45; 13:30 – 15:15 & 15:45 – 16:45

Official Statistics
Chair: Dr Arulsivanathan Naidoo

8:30 – 10:15

Presentations:

- **Sedikoe Godfrey Mankwe,** Statistics South Africa **(15 minutes):**
  Advocacy and importance of official statistics across all spheres of government
- **Vienie Botha,** Statistics South Africa *(Co-author: Mr Kevin Parry)* **(15 minutes):**
  The use of data visualisation techniques and social media channels to increase statistical awareness and literacy
- **Masete Letsoalo,** University of Pretoria *(Co-authors: Dr Boraine H, (University of Pretoria and Department of Planning, Monitoring and Evaluations (DPME)), Swanepoel, A, (University of Pretoria))* **(15 minutes):**
  Analysis of South African household poverty based on Income and Expenditure Survey 2010/11

Short presentations with posters:

- **Sipho Masimula,** Statistics South Africa *(Co-author: Arulsivanathan Naidoo)* **(5 minutes):**
  Determinants of Children School Attendance in South Africa
- **Cleopatra Sikhosana,** Statistics South Africa *(Co-author: Arulsivanathan Naidoo)* **(5 minutes):**
  Does Education Really Disadvantage Women in the Marriage Market?
- **Oupa Nkwinika,** Statistics South Africa **(5 minutes):**
  Gender differentials in housing characteristics and household possessions in South Africa
- **Seipati Kgothe,** Statistics South Africa **(5 minutes):**
  Outcomes of being raised by grandparents as the primary care giver.
- **Ntokozo Molata,** Statistics South Africa *(Co-author: Dr Naidoo, A (Statistics South Africa))* **(5 minutes):**
  A spatial analysis of poverty in South Africa
- **Kenneth Chatindiara,** Statistics South Africa *(Co-author: Naidoo, A (Statistics South Africa))* **(5 minutes):**
  Socio-economic determinants of motor ownership in South Africa
- **Lehlogonolo Masenya,** Statistics South Africa *(Co-author: Dr Arulsivanathan Naidoo)* **(5 minutes):**
  Measuring the efficiency of South African municipalities using Data Envelopment Analysis

Discussion (20 minutes): Dr Arulsivanathan Naidoo

10:45 – 11:45

Presentations:

- **Mzi Shabangu,** Statistics South Africa **(15 minutes):**
  Pro poor public transport: Rea Vaya in the City of Johannesburg
- **Zanele Mazibuko,** Statistics South Africa *(Co-author: Dr Arulsivanathan Naidoo)* **(15 minutes):**
  Spatially variability of men and women determinants of unemployment in Limpopo Province
- **Pinki Mulibana,** Statistics South Africa *(Co-author: Malepe, N (Methodology and Evaluation, Statistics South Africa) and Masemula, M (Methodology and Evaluation, Statistics South Africa))* **(15 minutes):**
  The use of administrative data to derive synthetic estimates for Micro enterprises- in order to reduce response burden and cost

Discussion (15 minutes): Dr Arulsivanathan Naidoo

13:30 – 15:15

Mini-Plenary:

- **Director-General Walter Radermacher,** Eurostat **(40 minutes):**
  On our Way to Sustainable Development - Guidance from Statistics

Presentation:

- **Statistician General Pali Lahohla,** Statistics South Africa **(40 minutes):**
  Community Survey 2016

Discussion (25 minutes): Dr Arulsivanathan Naidoo

15:45 – 16:45

Presentations:

- **Dr Arulsivanathan Naidoo,** Statistics South Africa **(30 minutes):**
  Stats SA dissemination
- **SAPS (30 minutes):**
  Crime statistics

Special Session
Monday 30 November 2015 8:30 – 10:15 & 10:45 – 11:45

Risk theory in finance and actuarial science
Chair: Dr Conrad Beyers

Mini-Plenary:

- **Prof Helgard Raubenheimer,** Centre for BMI, North-West University *(co-authors: PJ de Jongh (Centre for BMI, NWU, South Africa), T de Wet (Centre for BMI, NWU, South Africa) and K Panman (Centre for BMI, NWU, South Africa))* **(40 minutes):**
  A Simulation Comparison of Quantile Approximation Techniques for Compound Distributions popular in Operational Risk

Presentations:

- **Morne Joubert,** North West University **(20 minutes):**
  Estimation technique for deriving the Basel LGD estimate on retail bank mortgage portfolio
- **Dr Jaco Visagie**, North West University **(20 minutes):**
  A generalisation of the mean correcting martingale measure
- **Dr Franck Adekambi**, University of Johannesburg **(20 minutes):**
  A New Approach To Approximating The Distribution Of Aggregate Discounted Claims

- **Knowledge Chinhamu,** University of KwaZulu-Natal (co-author: *Huang, C-K(Department of Statistical Sciences,University of Cape Town) and Chikobvu,D (Department of Mathematical Statistics and Actuarial Science, University of the Free State)* **(20 minutes):**
  Evaluating Risk in Precious Metal Prices With Generalized Hyperbolic And Stable Distributions

Short presentations with posters:

- **Frans Koning,** University of the Free State **(5 minutes):**
  Long Term Care, The South African Outlook, Pricing And Viability
- **Lionel Kemda,** University of Kwazulu-Natal (*co-authors: Chinhamu, K(School of Mathematics, Statistics and Computer Science,University of KwaZulu-Natal) and Huang, C-K (Department of Statistical Sciences, University of Cape Town))* **(5 minutes):**
  Modelling financial data using the Multivariate generalized hyperbolic distribution and Copula.
- **Forbes Kaseke,** University of Kwazulu-Natal **(5 minutes):**
  Modelling Volatility in Stock Returns: Case Study 3 JSE Companies

Discussion (30 minutes): Dr Conrad Beyers

Special Session
Monday 30 November 2015 8:30 – 10:15 & 10:45 – 11:45

Fit In or Fall Out: Statistical Distributions
Chair: Prof Filipe Marques

Mini-Plenary:

- **Prof Narayanaswamy Balakrishnan**, McMaster University, Canada **(40 minutes):**
  Some new attractive families of distributions and associated issues

Presentations:

- **Prof Filipe Marques,** DM, FCT and CMA, Universidade NOVA de Lisboa, Almada, Portugal **(20 minutes):**
  Asymptotic approximations for the sum of independent Gamma random variables and for the product of independent Beta random variables

- **Dr Leonard Santana,** North-West University *(co-authors: Allison, JS (Department of Statistics, North-West University), Visagie, J (Department of Statistics, North-West University), Smit, N (Department of Statistics, North-West University)* **(15 minutes):**
  An objective comparison between various goodness-of-fit tests for exponentiality
- **Dr Paul J. van Staden,** University of Pretoria *(co-author: King, R.A.R. (School of Mathematical and Physical Sciences, University of Newcastle, Australia)*) **(15 minutes):**
  The quantile statistical universe
- **Dr Salimeh Yasaei Sekeh**, Federal University of Sao Carlos (UFSCar), SP, Brazil **(15 minutes):**
  On weighted Gaussian entropy

Short presentations with posters:

- **Olusegun Broderick Oluyede,** Georgia Southern University **(5 minutes):**
  A New Compound Class of Burr Weibull-Poisson Distribution: Properties and Applications
- **Theodor Loots**, University of Pretoria *(co-authors: Bekker, A (Department of Statistics, University of Pretoria) and Balakrishnan, N (Department of Mathematics and Statistics, McMaster University)* **(5 minutes):**
  Arc length estimation of cumulative distribution functions
- **Albert Mijburgh**, University of Pretoria *(co-authors: Bekker, A (Department of Statistics, University of Pretoria) and Human, S (Department of Statistics, University of Pretoria))* **(5 minutes):**
  Generalised Multivariate Beta Type II Distribution
- **Peter Iiyambo,** *(co-author: Robert Schall)* **(5 minutes):**
  Coverage probabilities and average length of generalized confidence intervals for the ratio of scale parameters, difference of location parameters and difference of quantiles of two Weibull distributions.
- **Dr Eno Akarawak,** University of Lagos *(co-authors: Adeleke, I.A. (Department of Actuarial Science and Insurance, University of Lagos) and Okafor, R.O. (Department of Mathematics, University of Lagos))* (5 minutes):
  On the T-X Families of Continuous Distributions
- **I Adeleker,** University of Lagos, Nigeria (*co-authors: Akarawak, E.E.E; (Department of Mathematics, University of Lagos, Nigeria), Olalude, G. A; (Department of*

Statistics, Federal Polytechnic, Ede, Osun State), Okafor, R. O.; (Department of Mathematics, University of Lagos, Nigeria)) **(5 minutes):**
The Four-Parameter Weibull-Logistic Distribution And Its Properties

- **Brenda Omachar,** University of Pretoria *(Co-authors:)* **(5 minutes):**
The skew hyperbolic secant distribution

Discussion (20 minutes): Prof N Balakrishnan, Prof Filipe Marques

Special Session
Monday 30 November 2015 8:30 – 10:15 & 10:45 – 11:45

The Analysis of Data from Clinical Trials
Chair: Prof Francesca Little

Mini-Plenary:

- **Dr Carl Lombard**, Biostatistics Unit, South African Medical Research Council **(40 minutes):**
Analysis of Randomised Controlled Trials – some perspectives

Presentations:

- **Dr Anneke Grobler,** CAPRISA **(20 minutes):**
Adaptive study design to reduce the size of a Phase II clinical trial for HIV prevention

- **J Ramjith,** Division of Biostatistics & Epidemiology, School of Public Health & Family Medicine, University of Cape Town, Cape Town, South Africa **(20 minutes):**
An application of the extensions of the Cox model to model the incidence of pneumonia and repeat episodes of pneumonia in boys & girls in a low-middle income setting in South Africa: The Drakenstein child health study.

- **Dr Freedom Gumedze**, University of Cape Town **(20 minutes):**
Analysis of recurrent hospitalisations and deaths in a tuberculous pericarditis multicentre clinical trial

- **Dr Kerry Leask,** CAPRISA **(20 minutes):**
Design and Analysis of Cluster Randomised Trials

Discussion (40 minutes): Prof Francesca Little

Special Session
Monday 30 November 2015 8:30 – 10:15

Statistical Image Processing and Robotics
Chair: Dr Inger Fabris-Rotelli

Mini-Plenary:

- **Dr Robert King,** University of Newcastl, Australia **(40 minutes):**
  Image analysis in robot soccer

Presentations:

- **Alex Lau,** University of Pretoria *(Co-author: Fabris-Rotelli, I (University of Pretoria) and Bekker, A (University of Pretoria))* **(20 minutes);**
  A Study of Dependence Structures in Image Pixels

- **Dr Surette Bierman,** Stellenbosch University **(20 minutes):**
  Feature selection for kernel models by means of stepwise selection and regularisation: a comparative study

Discussion (25 minutes): Dr I Fabris-Rotelli

Special Session
Monday 30 November 2015 13:30 – 15:15 & 15:45 – 16:45

Managing the global hunger challenge: food security measurement and monitoring in the Sustainable Development Goal era

Chair: Dr Isabel Schmidt

Mini-Plenary:

- **Dr Carlo Cafiero,** Senior Statistician in the FAO Statistics Division, Rome *(Co-authors: Nord, M., Viviani, S.)* **(40 minutes):**
  Constructing, validating, interpreting and presenting household food insecurity measures.

Presentations:

- **Dr Tsakani Ngomani,** DPME **(30 minutes):**
  Data-driven policy making, impact assessment and accountability: The experience of the Department for Planning Evaluation and Monitoring (DPME)

- **Nozipho Shabalala,** Statistics South Africa **(30 minutes):**
  Stats SA's Poverty and Food Security measurements

- **Prof Sheryl Hendriks,** University of Pretoria **(30 minutes):**
  What are we measuring? Comparison of food security indicators from the Eastern Cape

Discussion (30 minutes): Dr Isabel Schmidt

Special Session
Monday 30 November 2015 13:30 – 15:15 & 15:45 – 16:45

Extreme Value Theory
Chair: Prof Andrzej Kijko

Mini-Plenary:

- **Prof Jan Beirlant,** Department of Mathematics, LStat and LRisk KU Leuven, and Department of Mathematical Statistics and Actuarial Science, University of the Free State (Co-authors: Tom Reynkens, Department of Mathematics KU Leuven; Isabel Fraga Alves, Department of Statistics, University of Lisbon; Ivette Gomes, Department of Statistics, University of Lisbon) **(40 minutes):**
  Tail estimation in a bounded world: bounded or unbounded models?

Presentations:

- **Sean van der Merwe,** University of the Free State *(Ntseki, J (Department of Mathematical Statistics and Actuarial Science, University of the Free State) and Teise, C (Department of Mathematical Statistics and Actuarial Science, University of the Free State)* **(20 minutes):**
  Comparison of old and new fit tests for peaks over a known threshold

- **Dr Andréhette Verster,** University of the Free State *(Co-author: Maribe, G (Department of Mathematical Statistics and Actuarial Science, University of the Free State))* **(20 minutes):**
  An Improved unbiased-Bayesian estimation of the Extreme value index for heavy-tailed distributions

- **Tadele Akeba Diriba,** University of Pretoria *(Co-authors: Legesse Kassa Debusho; Joel Botai)* **(20 minutes):**
  Modelling Extreme Daily Temperature using Generalized Pareto Distribution at Port Elizabeth, South Africa

- **Richard Minkah,** Stellenbosch University and University of Ghana (Co-author: Prof. Tertius de Wet, Stellenbosch University) **(20 minutes):**
  Conditional Tail Index and Extreme Quantiles: A Review and Simulation Comparison

Short Presentations with a Poster:

- **Daniel Maposa,** University of Limpopo *(Co-author: Cochran, JJ (Department of Information Systems, Statistics and Management Sciences, University of Alabama, Tuscaloosa, USA) and Lesaoana, M (Department of Statistics and Operations Research, University of Limpopo))* **(5 minutes):**
  Modelling nonstationary extremes in the lower Limpopo River basin of Mozambique

- **Tsirizani Kaombe,** Department Of Mathematical Sciences, Chancellor College, University Of Malawi *(Co-author: Manda, S. O. M. (Biostatistics Unit, South African Medical Research Council, Pretoria, Republic of South Africa))* **(5 minutes):**
  Assessing Influential Observations In Analysis Of Survival Data

Discussion (30 minutes): Prof Andrzej Kijko, Prof Jan Beirlant, Prof Daan de Waal

Special Session
Monday 30 November 2015 13:30 – 15:15 & 15:45 – 16:45

Bayesian network applications and distributed reasoning systems
Chair: Dr Alta de Waal

Mini-Plenary:

- **Dr Gregor Pavlin,** Thales Research & Technology Netherlands/D-CIS Lab **(40 minutes):**
  Situation Assessment Exploiting Correlated Data from Disparate, Spatially Distributed Sources: A Probabilistic Causal Model Approach

Presentations:

- **Dr Alta de Waal,** University of Pretoria **(20 minutes):**
  3D Expert Knowledge Elicitation for Bayesian Networks

- **Hildegarde Koen,** CSIR, University of Pretoria **(20 minutes):**
  A Bayesian Network Approach to Combating Rhino Poaching in the Kruger National Park

- **Dr Victoria Goodall,** Nelson Mandela Metropolitan University *(Co-author: Fatti, L.P. (School of Statistics & Actuarial Science, University of the Witwatersrand) and Owen-Smith, N (School of Animal, Plant & Environmental Sciences, University of the Witwatersrand))* **(20 minutes):**
  Multiple State Allocation for Latent Animal Behavioural States based on Hidden Markov Models

- **Dr Joel Dabrowski,** University of Pretoria *(Co-authors Dr Pieter de Villiers & Dr Conrad Beyers, University of Pretoria))* **(20 minutes):**
  Towards developing early warning systems - behavioural modelling from maritime piracy to banking crises

- **Rik Claessens,** Thales Research & Technology Netherlands/D-CIS Lab, University of Liverpool *(Co-authors: 1) Alta de Waal, University of Pretoria 2) Pieter de Villiers, University of Pretoria & CSIR 3) Ate Penders, Thales Research & Technology Netherlands/D-CIS Lab & Delft University of Technology 4) Gregor Pavlin, Thales Research & Technology Netherlands/D-CIS & University of Amsterdam, 5) Karl Tuyls, University of Liverpool & Delft University of Technology)* **(20 minutes):**
  Multi-Agent Target Tracking using Particle Filters enhanced with Context Data

Short Presentations with a Poster:

- **Dr Chistopher Udomboso,** Department of Statistics, University of Ibadan, Ibadan, Nigeria*) (Co-authors: Dr Chukwu, A U (Department of Statistics, University of Ibadan, Ibadan, Nigeria) and Prof Dontwi I K (Department of Mathematical Sciences, Nkwame Nkrumah University of Science and Technology, Kumasi, Ghana))* **(5 minutes):**
  On Model Selection Criteria in Statistical Neural Network

Discussion (20 minutes): Dr Alta de Waal

Special Session
Monday 30 November 2015 13:30 – 15:15 & 15:45 – 16:45

Multilevel Modeling
Chair: Dr Gretel Crafford

Mini-Plenary:

- **Dr Agnes Stancel-Piątak,** IEA Data Processing and Research Center **(40 minutes):**
  Using Multiple Group Multilevel Latent Models for Cross-Country Comparisons

Presentations:

- **Jesca Batidzirai,** University of KwaZulu- Natal *(Co-authors: Manda, S.O.M (Biostatistics Research Unit, South Africa Medical Research Council, Pretoria)and Mwambi H.G (School of Mathematics, Statistics & Computer Science, University of KwaZulu- Natal))* **(20 minutes):**
  Multilevel Modelling of Event Histories in Family Formation and Dissolution Studies in the sub-Saharan Africa
- **Prof Deon Nel,** University of Pretoria **(20 minutes):**
  Applications of Multilevel Modelling in Brand Value Research
- **Prof Arien Strasheim,** Department of Human Resource Management, University of Pretoria, Faculty of Economic & Management Sciences *(Co-author: Kriel, G (Department of Human Resource Managemen, University of Pretoria, Faculty of Economic & Management Sciencest))* **(20 minutes):**
  Modelling branch-level data in MG SEM
- **Prof Jenny Hoobler,** University of Pretoria, Faculty of Economic & Management Sciences **(20 minutes):**
  Modelling Supervisor-Subordinate Relationship Dyadic Data

Discussion (40 minutes): Prof Jenny Hoobler

Special Session
Tuesday 1 December 2015 8:00 – 9:45 & 10:15 – 11:15

Bayesian statistical modelling
Chair: Prof Paul Mostert

Mini-Plenary:

- **Prof Emmanuel Lesaffre** Leuven Biostatistics and statistical Bioinformatics Centre **(40 minutes):**
  Modeling multivariate multilevel continuous responses with a hierarchical regression model for the mean and covariance matrix applied to a large nursing data set

Presentations:

- **Prof Paul Mostert,** Department of Statistics and Actuarial Science, Stellenbosch University *(Co-author: Van Rooyen, R (Department of Statistics and Actuarial Science, Stellenbosch University))* **(20 minutes):**
  Class of objective priors for a generalised compound Rayleigh model under various loss functions
- **Dr Divan Burger,** University of the Free State and Quintiles, Biostatistics *(Co-author: Prof. Robert Schall)* **(20 minutes):**
  Robust mixed effects regression models with application to colony forming unit count and time to positivity in TB research
- **Osvaldo Loquiha,** Universidade Eduardo Mondlane/Uhasselt *(Co-author: Hens,N (Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Universiteit Hasselt), and Chavane, L (Jhpiego, MCHIP Maternal and Child Health Integrated Program), and Temmerman, M (International Centre for Reproductive Health, Ghent University), and Aerts, M (Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Universiteit Hasselt))* **(20 minutes):**
  Modelling Heterogeneity for Count Data. A Study of Maternal Mortality in Health Facilities in Mozambique
- **Adelino Martins,** Eduardo Mondlane University **(20 minutes):**
  A New Model For Multivariate Current Status Data

Discussion (40 minutes): Prof Paul Mostert

Special Session
Tuesday 1 December 2015 8:00 – 9:45 & 10:15 – 11:15

Business Analytics and Big Data
Chair: Dr Frans Kanfer

Presentations:

- **Prof Gregory Lee,** Wits Business School **(30 minutes):**
  Extrapolating business statistics to financial valuations
- **Pravin Burra,** Customer Insights & Analytics, Standard Bank **(30 minutes):**
  The Business of Counting: From practical considerations to value extraction
- **Prof Paul Fatti,** Wits University **(40 minutes):**
  Big Data, Data Science and Analytics - the end of Statistics?
- **Patrick Hall,** SAS Institute **(40 minutes):**
  An Overview of Machine Learning with SAS Enterprise Miner

Discussion (20 minutes): Dr Frans Kanfer

Special Session
Tuesday 1 December 2015 14:15 – 16:00 & 16:30 – 17:30

Business Analytics and Big Data
Chair: Prof Sonali Das

Mini-plenary:

- **Dr Robert N Rodriguez,** SAS Institute **(40 minutes):**
  Methods, Models, Motivation, and More: Recent Developments in SAS/STAT® Software

Mini-Plenary:

- **Prof Montaz Ali,** School Of Computer Science And Applied Mathematics, University Of The Witwatersrand **(40 minutes):**
  Models and methods for analysing (Big) datasets

Presentations:

- **Peter Smith,** University of Pretoria *(Co-authors: Kanfer, F and Millard, S (Department of Statistics, University of Pretoria))* **(20 minutes):**
  Investment-Policy Surrender Prediction with Random Survival Forests
- **Erin Bromley-Gans,** UTi *(co-authors: Kirshnee Moodley and Calven van der Byl)* **(20 minutes):**
  Demand Forecasting for Inventory Planning

Discussion (30 minutes): Prof Sonali Das

Special Session
Tuesday 1 December 2015 8:00 – 9:45

Experimental Design
Chair: Prof Legesse Debusho

Mini-Plenary:

- **Prof Linda Haines,** University of Cape Town **(40 minutes):**
  Designs for Small Data

Presentations:

- **Prof Legesse Debusho,** University of South Africa *(Co-author: Dibaba Bayisa Gemechu and Linda M. Haines)* **(20 minutes):**
  Properties of A- and D-optimal row-column designs for two-colour cDNA microarray experiments: Robustness against missing arrays
- **Prof Francois Steffens,** University of Pretoria **(20 minutes):**
  A logarithmic logistic regression model

Discussion (25 minutes): Prof Legesse Debusho

Special Session
Tuesday 1 December 2015 8:00 – 9:45 & 10:15 – 11:15

Multivariate Analysis in Economic and Management Sciences
Chair: Prof Igor Litvine

Mini-Plenary:

- **Prof Francis Biesmans,** University of Lorraine, France *(Co-author: Igor Litvine (Centre of Expertise in Forecasting, NMMU, South Africa))* **(40 minutes):**
  The Dynamic Probit Model: A Tool for Forecasting

Presentations:

- **Prof Rangan Gupta,** University of Pretoria *(Co-author: Bejiros, S (European University Institute) and Majumdar, A (Center for Advanced Statistics and Econometrics, Soochow University))* **(20 minutes);**
  Incorporating Economic Policy Uncertainty in US Equity Premium Models: A Nonlinear Predictability Analysis

- **Dr Saragan Pillay,** Statistics South Africa **(20 minutes):**
  Inter-Linkages Between Private Investment, Public Investment And Economic Growth In South Africa

- **Alexander Boateng,** University of Limpopo *(Co-author: Prof Maseka, L (Department of Statistics and Operations Research, University of Limpopo), Prof Gil-Alana, LA(Faculty of Economics, University of Navarra), Prof Hlegani, S (Department of Mathematics and Applied Mathematics, University of Limpopo),Prof Belete, A (Department of Agricultural Economics, University of Limpopo))* **(20 minutes):**
  Consumer Price Index (CPI) inflation rates, Whittle method, Long memory, ARFIMA model

- **Siphumlile Mangisa,** Nelson Mandela Metropolitan University *(Co-author: Das, S (Advanced Mathematical Modelling, Modelling and Digital Science, Council for Scientific and Industrial Research, Pretoria, South Africa; and Department of Statistics, Nelson Mandela Metropolitan University, South Africa) and Sharp, G (Department of Statistics, Nelson Mandela Metropolitan University, South Africa) and Ray, S (School of Mathematics and Statistics, University of Glasgow, UK))* **(20 minutes):**
  A functional data analysis investigation of the relationship between electricity demand and economic indicators in South Africa

Discussion (40 minutes): Prof Igor Litvine

Special Session
Tuesday 1 December 2015 14:00 – 15:15 & 15:45 – 16:45

Biostatistics
Chair: Prof Legesse Debusho

Mini-Plenary:

- **Prof Din Chen,** University of North Carolina at Chapel Hill **(40 minutes):**
  Interval-Censored Time-to-event Data: From Parametric to Nonparametric Survival Data Analysis

Presentations:

- **Elizabeth R Zell,** Stat-Epi Associates Inc.; CDC (retired) **(20 minutes):**
  A Potential Outcomes Approach to Documenting the Public Health Impact of the Introduction of PCV13 for the Prevention of Invasive Pneumococcal Disease

- **Prof Samuel Manda,** South African Medical Research Council **(20 minutes):**
  A Bayesian Modelling Approach for Weighted Survival Data from Non-Proportionally Sampled Strata in Complex Surveys

- **Prof Henry Mwambi,** School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal *(Co-authors: Dr Ali Satty (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal) and Professor Geert Molenberghs (Hasselt University, I-BioStat, 3500 Hasselt, Belgium and KU Leuven - University of Leuven, 3000 Leuven, Belgium))* **(20 minutes):**
  Different Methods for handling incomplete longitudinal binary outcome due missing at random dropout

- **Esmè Jordaan,** Biostatistics unit, MRC **(20 minutes);**
  Applying a Structural Equation Model (SEM) to infer a causal relationship between alcohol use and ART adherence
- **Dr Gaetan Kabera,** South African Medical Research Council *(Co-author: Mr Paul Gatabazi, University of Johannesburg)* **(20 minutes)**:
  A look on additive hazards regression models in survival analysis

Discussion (20 minutes): Prof Legesse Debusho, Prof Din Chen

Special Session
Wednesday 2 December 2015 8:00 – 9:45 & 10:15 – 11:15

Statistics in Sport
Chair: Prof Gary Sharp

Mini-Plenary:

- **Prof Tim Swartz,** Simon Fraser University, Burnaby BC, Canada **(40 minutes):**
  Recent Work in Twenty20 Cricket Analytics

Presentations:

- **Prof Sonali Das,** CSIR *(co-authors: B Ganguli, Univ. of Calcutta; Q Louw, Univ. of Stellenbosch; J Cockcroft (Univ. of Stellenbosch); S Sen Roy (Univ. of Calcutta); N Botha (CSIR, Pretoria))* **(20 minutes):**
  Statistical Analysis of Gait Data
- **Prof Hoffie Lemmer,** University of Johannesburg **(20 minutes):**
  A Measure for the Wicket Taking Ability of Bowlers
- **Arnu Pretorius,** Stellenbosch University *(Co-author: Dr Surette Bierman)* **(20 minutes):**
  Predicting the future of the 2015 Rugby World Cup using Random Forest variants
- **Max Jordaan,** Statistics South Africa **(20 minutes):**
  Spatial Statistical Analysis to determine Cricket Facilities

Short Presentations with a Poster:

- **John Calder,** Nelson Mandela Metropolitan University *(Co-author: Sharp, G (Department of Statistics HoD, Nelson Mandela Metropolitan University))* **(5 minutes):**
  Statistical Methods for Cricket Batting Performance

Discussion (30 minutes): Dr Paul van Staden

Special Session
Tuesday 1 December 2015 14:00 – 15:15 & 15:45 – 16:45

Bayesian Stream
Chair: Dr Lizanne Raubenheimer

Mini-Plenary:

- **Prof Donald B Rubin,** Harvard University **(40 minutes):**
  The Utility of Bayesian Inference in Instrumental Variables Models

Presentations:

- **Dr Michael von Maltitz,** University of the Free State *(Co-author: van der Merwe, AJ (Department of Mathematical Statistics and Actuarial Science, University of the Free State))* **(20 minutes):**
  Investigating the posterior predictive p-value for model evaluation in sequential regression multiple imputation (SRMI)
- **Janet van Niekerk,** University of Pretoria *(Co-authors: A. Bekker*, M. Arashi*' and D.J. de Waal*" *Department of Statistics, Faculty of Natural and Agricultural Sciences,University of Pretoria, Pretoria, South Africa, 'Department of Statistics, School of Mathematical Sciences, University of Shahrood, Shahrood, Iran, **"Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa)* **(20 minutes):**
  Bayesian estimation under the matrix variate elliptical model
- **Raeesa Manjoo,** University of Witwatersrand *(Co-authors: Fitsum Abadi (School of Statistics and Actuarial Science, University of the Witwatersrand, P/Bag 03, Wits 2050, South Africa))* **(20 minutes):**
  A Bayesian capture-recapture model to estimate the survival rate of blue cranes
- **Dr Siaka Lougue,** University of Kwazulu Natal *(Co-author: Ogunsakin Ropo Ebenezer)* **(20 minutes):**
  Medication of people living with cancer in South Africa: A Bayesian approach of statistical analysis

Discussion (40 minutes): Dr Michael von Maltitz

Special Session
Wednesday 2 December 2015 8:00 – 9:45 & 10:15 – 11:15

Applications of Stochastic Processes
Chair: Prof Maxim Finkelstein

Mini-Plenary:

- **Prof Gregory Levitin,** The Israel Electric Corporation *(Co-author: Xing, L (Department of Electrical and Computer Engineering, University of Massachusetts))* **(40 minutes):**
  Stochastic systems with reworking

Presentations:

- **Prof Maxim Finkelstein,** University of the Free State *(co-author: Cha, JH (department of Statistics, Ewha Womans University, Korea))* **(20 minutes):**
  New Shock Models Based on the Generalized Polya Process
- **Prof Sarma Yadavalli,** University of Pretoria *(Co-author: Vaidyanathan S Vaidyanathan, Pondicherry University, Puducherry, India)* **(20 minutes):**
  Estimation of the Modified Traffic intensity of a Markovian Queuing system with Balking
- **Tsirizani Kaombe,** Department of Mathematical Sciences, Chancellor College, University of Malawi *(Co-authors: Samuel O.M. Manda (1Department of Mathematical Sciences, Chancellor College, University of Malawi and South African Medical Research Council, Biostatistics Unit, Pretoria, RSA))* **(20 minutes):**
  Assessing influential observations in analysis of survival data
- **Prof Igor Litvine,** NMMU *(Co-author: Francis Biesmans (Beta, University of Lorraine, France))* **(20 minutes);**
  Dating financial cycles with hierarchical method

Discussion (40 minutes): Prof Maxim Finkelstein, Prof Sarma Yadavalli, Prof Gregory Levitin

Special Session
Wednesday 2 December 2015 8:00 – 9:45 & 10:15 – 11:15

Statistical Process Control
Chair: Prof Subra Chakraborti

Mini-Plenary:

- **Prof Peihua Qiu,** Department of Biostatistics, University of Florida, USA **(40 minutes):**
  Recent Research on Nonparametric Statistical Process Control

Presentations:

- **Dr Nirpeksh Kumar,** MG Kashi Vidyapith, Varanasi, India *(Co-author: Prof. Chakraborti, S. (Department of Information Systems, Statistics and Management Science, University of Alabama, U.S.A.))* **(40 minutes):**
  Bayesian monitoring of times between events: The Shewhart t_r -chart
- **Dr Marien Graham,** University of Pretoria *(co-author: Mukherjee, A (Department of Mathematics, IIT Madras, India), Chakraborti, S (Department of Information Systems, Statistics and Management Science, University of Alabama, USA))* **(40 minutes):**
  Design and Implementation of Distribution-free Phase II EWMA Exceedance Control Charts for Monitoring Unknown Location

Discussion (60 minutes): Prof Subra Chakraborti

Special Session
Wednesday 2 December 2015 12:15 – 14:00 & 14:30 – 15:30

Statistical Process Control
Chair: Dr Schalk Human

Mini-Plenary:

- **Prof Abrie van de Merwe,** University of the Free State *(Co-author: van Zyl, R (Biostatistics, Quintiles) and Groenewald P.C.N (Department of Mathematical Statistics and Actuarial Sciences, University of the Free State))* **(40 minutes):**
  A Bayesian Control Chart for a One-sided Upper Tolerance Limit for the Normal Population

Presentations:

- **Sandile Shongwe,** University of Pretoria *(Co-author: Graham M.A. (Department of Statistics, University of Pretoria))* **(20 minutes):**
  Shewhart-type synthetic and runs-rules charts for monitoring the mean of normally distributed processes

- **Niladri Chakraborty,** University of Pretoria *(co-author: Chakraborti, S (Department of Statistics, University of Pretoria), Human,S.W. (Department of Statistics, University of Pretoria), Balakrishnan, N. (Department of Mathematics and Statistics, McMaster University))* **(20 minutes):**
  A Distribution-Free Generally Weighted Moving Average Control Chart

- **Jean-Claude Malela-Majika,** University of South Africa *(Co-author: E. Rapoo)* **(20 minutes):**
  Distribution-free CUSUM and EWMA Control Charts based on the Wilcoxon Rank-Sum Statistic using Ranked Set Sampling for Monitoring Mean Shifts

- **Dr Olatunde Adeoti,** University of South Africa *(Co-author: Prof John Olaomi (Department of Statistics, University of South Africa))* **(20 minutes):**
  Process capability index based control chart for variables

Short Presentations with a Poster:

- **Oliver Bodhlyera**, University of KwaZulu Natal *(Co-author: Zewotir, T (School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal) and Ramroop, S (School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal)* **(5 minutes):**
  Classification of Timber Genotypes for Chemical Pulping Using Piecewise Regression and Kernel Density based Clustering

Discussion (35 minutes): Dr Schalk Human, Prof Abrie van der Merwe

Special Session
Wednesday 2 December 2015 8:00 – 9:45 & 10:15 – 11:15

Complex Sampling
Chair: Prof Tertius de Wet

Mini-Plenary:

- **Dr Steven Heeringa,** Institute for Social Research, University of Michigan, Ann Arbor, MI *(Co-author: Berglund, P. (Institute for Social Research, University of Michigan), Melipillán, E.R. (Program in Survey Methods, University of Michigan)* **(40 minutes):**
  Survey Sampling and Big Data: Applications to Survey-assisted Modeling for Populations.

Presentations:

- **Dr Ariane Neethling,** Department Mathematical Statistics and Actuarial Science, University of the Free State *(Co-author: Luus, Retha (Department of Statistics and Population Studies, University of the Western Cape) and de Wet, Tertius (Department of Statistics and Actuarial Science, Stellenbosch University))* **(20 minutes):**
  The Role of Weighting in the Analysis of Complex Survey Data

- **Dr Saba Riaz,** Riphah International University Islamabad Pakistan *(co-author: Chakraborti, S (Department of Statistics, University of Pretoria), Human,S.W. (Department of Statistics, University of Pretoria), Balakrishnan, N. (Department of Mathematics and Statistics, McMaster University))* **(20 minutes):**
  A modified class of estimators for estimation of population mean in the presence on non-response

Short Presentations with a Poster:

- **Amos Chinomona,** Rhodes University *(Co-author: Mwambi, H (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal))* **(5 minutes):**
  Hierarchical Logistic Regression for Estimating HIV Prevalence using Survey Data Accounting for Missing Data

- **Maggie Kisaka-Lwayo**, Statistics South Africa *(Co-author: Caiphus Mashaba, Ngoako Mokgerepi, Neo Mashamba)* **(5 minutes):**
  A review of model-based approaches to small area estimation: An exploratory study

- **Thanyani Maremba**, Statistics South Africa) **(5 minutes):**
  Sample design to optimise the estimation of small micro and medium enterprise owners and their characteristics

Discussion (30 minutes): Prof Tertius de Wet, Dr Ariane Neethling

Special Session
Wednesday 2 December 2015 8:00 – 9:45 & 10:15 – 11:15

Spatial Statistics
Chair: Prof Christien Thiart

Mini-Plenary:

- **Prof Alfred Stein,** Twente University, The Netherlands **(40 minutes):**
  Spatial statistics: an overview and some recent developments.

Presentations:

- **Prof Christien Thiart,** Department of Statistical Sciences, University of Cape Town and AEON-ESSRI *(Co-author: Linda Haines (Department of Statistical Sciences, University of Cape Town), Suvira Bodha (Department of Statistical Sciences, University of Cape Town), Divan Stroebel (AEON-ESSRI, Nelson Mandela Metropolitan University) and Maarten de Wit (AEON-ESSRI, Nelson Mandela Metropolitan University))* **(20 minutes):**
  Space-filling designs for finding an optimum sample in order to access the quality of groundwater hydrochemistry of the Karoo

- **Greg Distiller,** University of Cape Town **(20 minutes):**
  Using continuous-time spatial capture-recapture (SCR) models to make inference about animal activity.

- **Elphas Okango,** University of Kwazulu-Natal *(Co-author: Henry Mwambi(1.School of Mathematics, Statistics and Computer Science, University of KwaZulu -Natal, Private Bag X01, 3201 Pietermaritzburg, South Africa), Oscar Ngesa (1, and Mathematics* and Informatics Department, Taita Taveta University College, P.O Box 635-80300, *Voi, Kenya.))* **(20 minutes):**
  Semi-Parametric spatial Joint modeling of HIV and HSV-2 among women in Kenya with spatially varying coefficients

- **Prof Samuel Manda,** South African Medical Research Council **(20 minutes):**
  Multivariate Spatial-Temporal Autocorrelations for Small-Area Multiple Health Outcomes in South Africa

Short Presentations with a Poster:

- **Mzabalazo Ngwenya**, Biometry, Agricultural Research Council (ARC) *(co-authors: Strydom, M. (Centre for Invasion Biology & Department of Conservation Ecology and Entomology, Stellenbosch University), Veldtman, R. (Applied Biodiversity Research, South African National Biodiversity Institute (SANBI)), Esler, K.J. (Centre for Invasion Biology & Department of Conservation Ecology and Entomology, Stellenbosch University)* **(5 minutes):**
  Characterising Australian Acacia seed bank size and its relationship with stand characteristics in the Western Cape

- **Timotheus Darikwa**, Department of Statistics and Operations Research, University of Limpopo *(Co-authors: Manda, S (Biostatistics Research Unit, South African Medical Research Council & School of Mathematics, Statistics and Computer Science, University of Kwazulu-Natal), Leasoana, M (Department of Statistics and Operations Research,University of Limpopo))* **(5 minutes):**
  Investigating Bivariate Spatial Autocorrelations of Cardiovascular Mortality in South Africa: 2011

- **Caiphus Mashaba,** Statistics South Africa *(Co-authors: Mokgerepi, N.; Strauss, M.; Chuene, M.; Ndzhukula, M.; Khan, A)* **(5 minutes):**
  The use of Geo-spatial data for Master Sample Design

Discussion (25 minutes): Prof Christien Thiart, Dr Inger Fabris-Rotelli

Special Session
Wednesday 2 December 2015 12:15 – 14:00 & 14:30 – 15:30

Multivariate Data Analysis
Chair: Prof Sugnet Lubbe

Mini-Plenary:

- **Prof Maurizio Vichi,** Universita di Roma Sapienza **(40 minutes):**
  New Challenges in Clustering and Dimensional Reduction in the Era of Big Data

Presentations:

- **Prof Jan Swanepoel,** North-West University, Potchefstroom **(20 minutes);**
  Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals
- **Prof Danie Uys,** Stellenbosch University **(20 minutes):**
  The histogram and polygon revisited
- **Johané Nienkemper-Swanepoel,** Stellenbosch University *(Co-author: le Roux, NJ (Department of Statistics and Actuarial Science, Stellenbosch University), Lubbe, S (Department of Statistical Sciences, University of Cape Town) and von Maltitz, MJ (Department of Mathematical Statistics and Actuarial Science, University of the Free State))* **(20 minutes):**
  Generalized Orthogonal Procrustes Analysis for the comparison of Multiple Imputed data sets
- **Anasu Rabe,** University of Botswana *(Co-authors: Shangodoyin,D.K.(Department of Statistics, University of Botswana) and Thaga,K.(Department of Statistics, University of Botswana))* **(20 minutes):**
  Cholesky-based Covariance Modeling in Longitudinal Studies

Short Presentations with a Poster:

- **Tshepho Brian Matlwa**, Statistics South Africa **(5 minutes):**
  Is There Hope for Survivalists?| Success In Running a NON-VAT Registered Business In SOUTH AFRICA.
- **Thabo Khule**, Statistics South Africa **(5 minutes):**
  Factors affecting high mortality in Lesotho, 2009
- **Mulalo Salane,** Statistics South Africa **(5 minutes):**
  Influential factors of divorce in South Africa

Discussion (30 minutes): Prof Sugnet Lubbe

Special Session
Wednesday 2 December 2015 12:15 – 14:00

Statistics in Nanoscience and Chemical Risk Assessment
Chair: Rianne Jacobs

Mini-Plenary:

- **Dr Fadri Gottschalk,** ETSS – Environmental, Technical and Scientific Services, Strada, Switzerland *(Co-author: Andrea Sanchini (ETSS – Environmental, Technical and Scientific Services, Strada, Switzerland))* **(40 minutes):**
  Probabilistic environmental exposure, effect and risk assessments in the context of potential chemical/nano risk

Presentations:

- **Rianne Jacobs,** Biometris, Wageningen University and Research Centre *(Co-authors: van der Voet, H (Biometris, Wageningen University and Research Centre) and ter Braak, CJF (Biometris, Wageningen University and Research Centre))* **(20 minutes);**
  Probabilistic methods for the environmental risk assessment of nanoparticles
- **Andries Haywood,** University of Pretoria *(Co-authors: Fabris-Rotelli, I (Department of Statistics, University of Pretoria) and Das, S (Advanced Mathematical Modelling, CSIR Modelling and Digital Science) and Wesley-Smith, J (DST/CSIR National Centre for Nanostructured Materials, CSIR))* **(20 minutes):**
  Bayesian object classification in nanoimages

Discussion (20 minutes): Rianne Jacobs

Statistical Methodology and Techniques (Open) Sessions

Tuesday 1 December 2015 8:00 - 9:45 & 10:15 - 11:15 & 14:15 – 16:00 (Official Statistics)

8:00 - 8:20 **Thabo Masemola**, Statistics South Africa

Long-term trends in living alone among South African adults: Age, gender, and educational differences

8:20 - 8:40 **Rosina Mosoma**, Statistics South Africa
(Co-authors: Dr Naidoo,A (Statistics South Africa))
Patterns of activity and employment in the young adulthood years (18-24) following their exit from the school system

8:40 - 9:00 **Mmanate Kekana**, Statistics South Africa
(Co-authors: Naidoo,A (Statistics South Africa))
Homeownership differentials in South Africa

9:00 - 9:20 **Collen Motsepa**, Statistics South Africa
(Co-authors: Dr Arulsivanathan Naidoo)
Socioeconomic Determinants and Spatial Variation of Fertility in South Africa

9:20 - 9:40 **Gaongalelwe Phakedi**, Statistics South Africa

Spatial variation in disability and poverty – A Case of South Africa

10:15 - 10:35 **Dawit Ayele**, University of KwaZulu-Natal
(Co-authors: Temesgen T. Zewotir
School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal)
Childhood mortality spatial distribution in Ethiopia

10:35 - 10:55 **Adewale Adeogun**, North-West University
(Co-authors: Palamuleni, M. (Department of Population Studies, North-West University)
Palamuleni, L. (School of Environmental & Health Science, North-West University))
Dynamic spatio-temporal analysis of Ebola virus disease: putting in perspective epidemics in Africa

10:55 - 11:15 **Thanyani Maremba**, Statistics South Africa

Sample design to optimise the estimation of small micro and medium enterprise owners and their characteristics

14:15 - 14:35 **Fadzayi Ndlovu**, Department of Statistics and Operations Research, National University of Science and Technology
(Co-authors: Chivafa, A (Department of Statistics and Operations Research, National University of Science and Technology) and Mdlongwa, P (Department of Statistics and Operations Research, National University of Science and Technology))
Modeling Gender Representation: A Case Study of the National University of Science and Technology

14:35 - 14:55 **Fiskani Kondowe**, University of Malawi
(Co-authors: Mwakilama, E (Department of Mathematical Sciences, University of Malawi-Chancellor College))
Assessing The Levels Of Secondary School Dropouts In Relation To Some Socio-Economic Factors: A Case Study Of Khonjeni.

14:55 - 15:15 **Henry Mlinde**, University of Malawi
(Co-authors: Simbeye, J (Department of Mathematics, Chancellor College, University Of Malawi) and Mwakilama, E (Department of Mathematics, Chancellor College, University Of Malawi))
Assessing Factors Affecting Admission Time Of Kaposi Sarcoma Using Survival Analysis, A Case Of Zomba Central Hospital Malawi

15:15 - 15:35 **Xaven Pillay**, StatsSA

Business clustering along the M1-N3-N1 corridor between Johannesburg and Pretoria, South Africa.

15:35 - 15:55 **Retius Chifurira**, University of KwaZulu-Natal
(Co-authors: Chinhamu, K(School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal))
Using Extreme Value Theory To Measure Value-At-Risk For Daily South African Mining Index

Tuesday 1 December 2015 16:30 – 17:30 (General)

16:50 – 17:10 **Cliff Richard Kikawa**, Tshwane University of Technology
(Co-authors: Kloppers, PH (Tshwane University of Technology))
A semi-parametric method for generating time series data: an approach for bootstrapping the residuals

Tuesday 1 December 2015 10:15 - 11:15 (Forecasting)

10:15 - 10:35 **Oluwaseun Otekunrin**, University of Ibadan, Nigeria
(Co-authors: Ariyo,O (Department of Statistics,University of Ibadan))
Modelling Total Electricity Generation in Nigeria: The Response Surface Methodology Approach

10:35 - 10:55 **Jenny Holloway**, CSIR
(Co-authors: Koen, R (CSIR) and Mokilane, P (CSIR))
Comparison of methods for long-term forecasting of electricity load profiles in South Africa

10:55 - 11:15 **Lienki Viljoen**, Stellenbosch University
(Co-authors: Steel, S. J. (Department of Statistics and Actuarial Science, Stellenbosch University))
Identifying a secondary series for Stepwise Common Singular Spectrum Analysis

Wednesday 2 December 2015 10:15 - 11:15 (General)

10:15 - 10:35 **Edmore Ranganai**, University of South Africa

Quality of Fit Measurement in Regression Quantiles: An Elemental Set Method Approach

10:35 - 10:55 **Sugnet Lubbe**, University of Cape Town
(Co-authors: le Roux, NJ (Department of Statistics and Actuarial Science, Stellenbosch University) and Gower, JC (Department of Mathematics and Statistics, The Open University UK))
Fisher Optimal Scores for Visualisation in Categorical Data

10:55 - 11:15 **Murendeni Nemukula**, University Of Limpopo And University Of The Witwatersrand
(Co-Authors: SIGAUKE, C (DEPARTMENT OF STATISTICS, UNIVERSITY OF VENDA) AND (SCHOOL OF STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF THE WITWATERSRAND))
Modelling average minimum daily temperature using extreme value theory with a time varying threshold

Wednesday 2 December 2015 12:15 - 14:00 (General)

12:15 - 12:35 **Iain MacDonald**, Univ of Cape Town

More thoughts on the EM algorithm

12:35 - 12:55 **Trudie Sandrock**, University of Stellenbosch
(Co-authors: Steel, S (Department of Statistics and Actuarial Science, University of Stellenbosch))
Variable selection in multi-label classification using probe variables

12:55 - 13:15 **Edmore Ranganai**, University of South Africa

A Note On Studentized Residuals in the Quantile Regression Framework

13:15 - 13:35 **Charl Pretorius**, Department of Statistics, North-West University, Potchefstroom Campus
(Co-authors: Prof Swanepoel, JWH (Department of Statistics, North-West University, Potchefstroom Campus))
On a new method of constructing bootstrap confidence bounds

13:35 - 13:55 **Cornelia J Swanepoel**, North-West University, Potchefstroom Campus
(Co-authors: Mr. Shawn C. Liebenberg (Statistical Consultation Services, North-West University, Potchefstroom Campus))
Multiple Imputation In The Presence Of A Detection Limit, With Applications: An Empirical Approach

Wednesday 2 December 2015 12:15 – 14:00 (Biostatistics)

12:15 – 12:35 **Welcome Dlamini**, University of KwaZulu-Natal

Statistical Models to Model the Probability of the Under-five Mortality in United Republic of Tanzania

12:35 – 12:55 **Alfred Musekiwa**, University of KwaZulu-Natal (UKZN)
(Co-authors: Manda, S (Biostatistics Unit, South African Medical Research Council) and Mwambi, H (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal))
Meta-analysis of Longitudinal Studies in the Presence of Missing Effect Sizes

12:55 – 13:15 **Adelino Juga**, Eduardo Mondlane University/Uhasselt University
(Co-authors: Niel Hens(Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat),
Hasselt University, Diepenbeek, Belgium) and (Centre for Health Economic Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute)
A Case-Control Study of Tattoo and HIV Infection among Teens in Mozambique

13:15 – 13:35 **Jacob Majakwara**, Wits university
(Co-authors: Suvra P (Department of Mathematics, University of Texas at Arlington, Texas, USA))
Likelihood inference based on EM algorithm for the destructive COM-Poisson cure rate model

13:35 – 13:55 **Nicolene Thiebaut**, Agricultural Research Council, Head-Office, Pretoria
(Co-authors: Dr Andre Nel and Annelie De Beer (Agricultural Research Council, Potchefstroom))
Yield probability as a method for cultivar selection

Wednesday 2 December 2015 14:30 - 15:30 (General)

14:30 - 14:50 **Trudie Sandrock**, University of Stellenbosch

From Bernoulli to Beethoven and Fisher to Pharrell: An Introduction to Music Information Retrieval

14:50 - 15:10 **Robert King**, Department of Statistics, University of Pretoria; School of Mathematical and Physical Sciences, University of Newcastle, Australia
(Co-authors: van Staden, P (Department of Statistics, University of Pretoria))
Mixtures of generalized lambda distributions

15:10 – 15:30 **Yehenew Getachew Kifle**, Department of Statistics & Operations Research, University of Limpopo, South Africa
(Co-authors: Delenasaw Yewhalaw, Delenasaw (Department of Biology, College of Natural sciences, Jimma University, Jimma Ethiopia); Niko Speybroeck (Institute of Health and Society, Universit´e Catholique de Louvain, Brussels, Belgium); Paul Janssen (CenStat, Hasselt)
Assessing the effect of distance from a dam on time to malaria, with distance confounded with the clustering structure.

Young Statistician's Sessions

Monday 30 November 2015 8:30 - 10:15

8:30 - 8:50 **Murendeni Maurel Nemukula**, University of the Witwatersrand
(Co-authors: Sigauke, C (School of Statistics and Actuarial Science, University of the Witwatersrand))
Modelling minimum average daily temperature using extreme value theory with a time varying threshold

8:50 - 9:10 **Tadele Diriba**, University of Pretoria
(Co-authors: Debusho, LK (Department of Statistics, University of South Africa) and Botai, J (Department of Geography, Geo informatics & Meteorology, University of Pretoria).)
Modelling Extreme Daily Temperature Using Generalized Pareto Distribution at Port Elizabeth, South Africa

9:10 - 9:30 **Anasu RABE**, University of Botswana
(Co-authors: Shangodoyin, D.K. (Department of Statistics, University of Botswana) and Thaga,K.(Department of Statistics, University of Botswana))
A New Approach to Covariance Modeling of Longitudinal Data

9:30 - 9:50 **Nontembeko Dudeni-Tlhone**, CSIR

Applicability Of Multilevel Models To Temporal Spectral Data

9:50 - 10:10 **Christiaan Ras**, University of Pretoria

The risk performance of the heteroscedastic preliminary test estimator under different loss functions

10:45 - 11:05 **Halima Twabi**, Chancellor College
(Co-authors: Namangale, J. J (Department of Mathematical Sciences, Chancellor College) and Mukaka, M (Nuffield Department of Medicine, University of Oxford (UK), Mahidol-Oxford Tropical Medicine Research Unit,Faculty of Tropical Medicine, Mahidol University))
Modeling Length of Hospital Stay for Tuberculosis In-Patients at Queen Elizabeth Central Hospital: Applying Competing risks

11:05 - 11:25 **Henry Mlinde**, Department of Mathematics, University of Malawi, Chancellor College, Zomba, Malawi

(Co-authors: J.Simbeye, E.Mwakilama, Department of Mathematics, University of Malawi, Chancellor College, Zomba, Malawi)
Assessing Factors Affecting Admission Time Of Kaposi Sarcoma Using Survival Analysis
A Case Of Zomba Central Hospital

Tuesday 1 December 2015 8:00 - 9:45

8:00 - 8:20 **Jesca Batidzirai**, University of KwaZulu- Natal
(Co-authors: Manda, S.O.M (South Africa Medical Research Council, Pretoria) and Mwambi, H.G (School of Mathematics, Statistics & Computer Science, University of KwaZulu- Natal))
Multilevel Modelling of Event Histories in Family Formation and Dissolution Studies in the sub-Saharan Africa

8:20 - 8:40 **Ansie Smit**, University of Pretoria Natural Hazard Centre, University of Pretoria
(Co-authors: Kijko, A (University of Pretoria Natural Hazard Centre, University of Pretoria) and Fabris-Rotelli, IN (Department of Statistics, University of Pretoria) and Van Staden, PJ (Department of Statistics, University of Pretoria))
New Procedure for Probabilistic Hazard Assessment from Incomplete and Uncertain Data

8:40 - 9:00 **Modupi Peter Mphekgwana**, African Institute for Mathematical Sciences
(Co-authors: Hewson, P ((Department of Statistics, Plymouth University (UK)))
Diagnosis of Zero Inflation

9:00 - 9:20 **Paul Claassen**, Department of Statistics, University of Pretoria
(Co-authors: Fletcher, L (Department of Statistics, University of Pretoria))
The problem of zero-inflated count data: a discussion and application of zero-inflated and hurdle models

9:20 - 9:40 **S Kubheka**, University Of South Africa Department Of Statistics
(Co-authors: E. Ranganai)
Long Memory and Structural Breaks:
An Application to Platinum Price Return Series

Monday 30 November 2015 15:45 - 16:45

15:45 - 16:05 **Johan Ferreira**, University of Pretoria
(Co-authors: Bekker, A (Department of Statistics, University of Pretoria, South Africa) and Arashi, M (Department of Statistics, University of Sharhood, Iran))
Quadratic forms on complex elliptical random variables and its applications

16:05 - 16:25 **Tsitsi Makoni**, University of Pretoria
(Co-authors: van Staden, P.J. (Department of Statistics, University of Pretoria))
Generalized Burr Type II - exponential distribution

16:25 - 16:45 **Seite Littah Makgai**, University of Pretoria
(Co-authors: Prof A. Bekker, Mr J.T. Ferreira (University of Pretoria))
Creating mixtures of Pareto distributions via beta type generators

Monday 30 November 2015 13:30 - 15:15

13:30 - 13:50 **Emmanuel Numapau Gyamfi**, Department Of Statistics, University Of Venda
(Co-authors: Kyei, K.A (Department Of Statistics, University Of Venda) And Gill, R (Department Of Mathematics, University Of Louisville))
Long-Memory In Asset Returns And Volatility: Evidence From West Africa

13:50 - 14:10 **Letlhogonolo Khenene**, Statistics South Africa

The impact of Infrastructure on South Africa's Economic Growth

14:10 - 14:30 **Francois Kamper**, University of Stellenbosch

Marginalization of Multivariate Gaussians with Application in Optimization Problems

14:30 - 14:50 **Sibusiso Magagula**, Nedbank/UNISA

Feasibility in using Greeks...to manage options' risks - The Management Perspective

Tuesday 1 December 2015 10:15 - 11:15

10:15 - 10:35 **Abdalla Kombo**, UKZN
(Co-authors: Satty A (School of Statistics, Mathematics and Computer Science, UKZN) and Mwambi H (School of Statistics, Mathematics and Computer Science, UKZN))
Handling longitudinal continuous outcomes with dropout missing at random: A comparative analysis

10:35 - 10:55 **Lolita Winnaar**, University of the Western Cape
(Co-authors: Prof. Renette Blignaut (University of the Western Cape) and Dr. George Frempong (Human Sciences Research Council))
Using multilevel analysis to determine the learner and school factors associated with mathematics performance

10:55 - 11:15 **Thuto Mothupi**, University Of Botswana
(Co-authors: Arnab,R(Department Of Statistics,University Of Botswana))
A randomized response survey on the risky behaviors of certain University students

Tuesday 1 December 2015 10:15 - 11:15

10:15 - 10:35 **Michaela Ritchie**, Council for Scientific and Industrial Research
A comparison of domain expert classification and unsupervised computer classification techniques: A case study of the Orange Riv

10:35 - 10:55 **Belisha Naidoo**, University of KwaZulu-Natal Westville
Statistical modelling and spatial mapping of crime in South Africa.

10:55 - 11:15 **Christine Kraamwinkel**, University of Pretoria
(Co-authors: Fabris-Rotelli, IN (Department of Statistics, University of Pretoria))
Spatial Sampling

Tuesday 1 December 2015 16:30 - 17:30

16:30 - 16:50 **Lisa-Ann Kirkland**, University of Pretoria
(Co-authors: Kanfer, F (Department of Statistics, University of Pretoria) and Millard, S (Department of Statistics, University of Pretoria))
LASSO Tuning Parameter Selection

16:50 - 17:10 **Sibusisiwe Khuluse-Makhanya**, CSIR
(Co-authors: Stein, A (Faculty of Geo-information Science and Earth Observation, University of Twente) and Debba, P (Built Environment, CSIR))
Sequential regression imputation of air quality data

Tuesday 1 December 2015 14:00 - 15:45

14:00 - 14:20 **Godson Mesike**, university of Lagos, Akoka,Nigeria
(Co-authors: Adeleke, I.A (Department of Actuarial science and Insurance, University of Lagos)
Hamadu, D (Department of Actuarial science and Insurance, University of Lagos))
Industry-Wide Data Governance Model For Credible Rating In Nigeria

14:20 - 14:40 **Farouk Ssekisaka**, Makerere University
(Co-Authors: Shamirah Iga)
Birth Registration In Uganda: Challenges, Opportunities And Lessons.

14:40 - 15:00 **Farouk Ssekisaka**, Makerere University

Islamic Banking as an option for developing Sub-Saharan Africa economies

15:00 - 15:20 **Xolani Jozi**, Statistics South Africa

Modelling Net-Internal Migration in South Africa

15:20 - 15:40 **Barend Mienie**, Nelson Mandela Metropolitan University
(Co-authors: WJ Brettenny
Nelson Mandela Metropolitan University
Department of Statistics
GD Sharp
Nelson Mandela Metropolitan University
Department of Statistics)
Assessing the Productivity of Selective Container Terminals in Africa using DEA

Tuesday 1 December 2015 8:00 - 9:45

8:00 - 8:20 **Janet Van Niekerk**, University of Pretoria
(Co-authors: Bekker, A (Department of Statistics, University of Pretoria) and Arashi, M (Department of Statistics, University of Shahrood, Shahrood, Iran and Department of Statistics, University of Pretoria))
Comparative subjective Bayesian analysis of the normal model

8:20 - 8:40 **Hossein Masoumi Karakani**, University of Pretoria
(Co-authors: Van Niekerk, J (Department of Statistics, University of Pretoria) and Van Staden, P.J (Department of Statistics, University of Pretoria))
The first-order autoregressive process - a Bayesian perspective

8:40 - 9:00 **Sisa Pazi**, Nelson Mandela Metropolitan Univerisity
(Co-authors: Sharp, G.D (Department of Statistics, Nelson Mandela Metropolitan University) and Clohessy C ((Department of Statistics, Nelson Mandela Metropolitan University))
Statistical methods for the detection of non-technical electricity losses: A case study for Nelson Mandela Bay Municipality

9:00 - 9:20 **Dibaba Gemechu**, University of Pretoria
(Co-authors: Debusho, L. K. (Department of Statistics, University of South Africa) and Haines, L. M. (Department of Statistical Sciences, University of Cape Town))
Bayesian optimal block designs for two-colour cDNA microarray experiments

Tuesday 1 December 2015 16:30 - 17:30

16:30 – 16:50 **Arnu Pretorius**, Stellenbosch University
(Co-authors: Dr Surette Bierman)
Predicting the future of the 2015 Rugby World Cup using Random Forest variants

Tuesday 1 December 2015 16:30 - 17:30

16:30 - 16:50 **Jean-Claude Malela-Majika**, UNISA
(Co-authors: Dr Rapoo, E (Department of Statistics, University of South Africa))
Distribution-free CUSUM and EWMA Control Charts based on the Wilcoxon Rank-Sum Statistic using RSS for Monitoring Mean Shifts

16:50 - 17:10 **Niladri Chakraborty**, University of Pretoria
(Co-authors: Chakraborti, S (Department of Statistics, University of Pretoria), Human,S.W. (Department of Statistics, University of Pretoria), Balakrishnan, N. (Department of Mathematics and Statistics, McMaster University))
A Distribution-Free Generally Weighted Moving Average Control Chart

17:10 - 17:30 **Jordache Ramjith**, Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town
(Co-authors: Myer, L (Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town)and Zar, H (Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town) a)
An application of the extensions of the Cox model to model the incidence of pneumonia and repeat episodes of pneumonia in boys &

Tuesday 1 December 2015 14:15 - 16:00

14:15 - 14:35 **Dalene Bezuidenhout**, Stellenbosch University
(Co-authors: de Villiers, Margaret; (Stellenbosch University) and Mostert, Paul J. (Stellenbosch University))
Influence of right-censoring on some kernel-smoothed hazard rates

14:35 - 14:55 **Charl Janse van Rensburg**, University of Pretoria
(Co-authors: Fabris-Rotelli, I (Department of Statistics, University of Pretoria))
Big data, compressed sensing and wavelets

Workshops

Connecting the dots of data science: academia to industry

28 November 2015

Dr Robert N. Rodriguez SAS Institute, Senior Director, Research and Development, SAS and Dr F Kanfer, University of Pretoria

The summit addresses the latest trends in the field including developments in business analytics, data driven solutions, big data and automated data sources, high performance computing and modelling unstructured data.

International experts will discuss and share experiences and novel ideas.

Key industry partners as well as academia are invited to discuss and debate the role of Data Science in industry and how academia (from diverse disciplines, not only statistics) can contribute to develop the necessary skills.

Data Science: Hype and Reality

Patrick Hall, SAS Institute

This talk will disambiguate the buzzwords and bust the myths of data science by answering three basic questions: Is data science a science? Is data science new? What is a data scientist? With this foundation in place, the tools and techniques of data scientists will be explained. Data scientists must be familiar with conventional data warehousing technologies and the newer Hadoop ecosystem. They must understand how to analyze data efficiently, whether on a laptop or a cluster of computers in the cloud.

The crucial importance of statistics and machine learning in data science will also be addressed, along with meaningful similarities and differences between these two disciplines. To conclude, this talk will describe a few emerging trends and new ideas in the field of data science.

The Cloudera Data Science Challenge 2: Finding Anomalies in the United States Medicare Insurance System

Patrick Hall, SAS Institute

The availability of large volumes of data has made it possible to build predictive models that are highly valued in business and scientific applications because they predicts outcomes for customers, patients, and subjects at a detailed and even personalized level. Increasingly, the data are so massive that they must be stored and processed on commodity machines using software frameworks such as Hadoop. This trend is driving the demand for data scientists, and it calls for statisticians to gain an understanding of data infrastructures and acquire tools for large-scale data analysis.

SAS has developed a series of high-performance procedures for statistical modeling and model selection, which are available in SAS/STAT® software. On single machines, these procedures achieve scalability by exploiting all the cores on the machine. In distributed computing environments, these procedures exploit parallel access to the data, along with all the cores and the huge amounts of memory that are available. This presentation explains the architectural concepts, statistical capabilities, and practical benefits of these tools.

An Introduction to High-Performance Statistical Modeling Procedures in SAS

Robert Rodriguez, SAS Institute

The availability of large volumes of data has made it possible to build predictive models that are highly valued in business and scientific applications because they predicts outcomes for customers, patients, and subjects at a detailed and even personalized level. Increasingly, the data are so massive that they must be stored and processed on commodity machines using software frameworks such as Hadoop. This trend is driving the demand for data scientists, and it calls for statisticians to gain an understanding of data infrastructures and acquire tools for large-scale data analysis.

SAS has developed a series of high-performance procedures for statistical modeling and model selection, which are available in SAS/STAT® software. On single machines, these procedures achieve scalability by exploiting all the cores on the machine. In distributed computing environments, these procedures exploit parallel access to the data, along with all the cores and the huge amounts of memory that are available. This presentation explains the architectural concepts, statistical capabilities, and practical benefits of these tools.

A Using data science to make smarter customer engagement decisions Jean Tranter, Head: Analytics, The Foschini Retail Group (TFG)

In today's omni-channel world, customers receive communications and offers from organisations through multiple channels: sms, email, direct mail, telephone calls, social media…

How do your organisation's message or offer stand out? How do you make sure that your offer does not get lost in the clutter? How do you make sure that you target the appropriate customers with the right offer at the right time?

In this presentation, TFG explores how predictive modelling and prescriptive analytics drive customer engagement decisions in a retail environment, with the ultimate objective to enhance the customer experience and increase customer value.

Education Workshop: Online teaching of Statistics – special project.

29 November 2015

Prof Chris Wild, University of Auckland, NZ
and
Prof D North, KZN University

The conference traditionally has an Education Workshop. This year it is called *Statistics in Education – developing a first year online course for countrywide use.* This is the start of a task team to develop a course for use in South Africa to assist all universities with the huge capacity experienced at first year level. The workshop will continue through the conference as a special session with the international expert bringing in novel ideas, and have a follow up discussion workshop on 3 December 2015 (see below). This workshop will be hands on and in a computer lab.

Text Analytics Short Course

29 November 2015

Prof Edward Jones, Texas A&M University

Text analytics started out with simple word count analyses. At present text analytics examines contextual information. Sentiment and opinion analysis, for example, makes it possible to efficiently incorporate the opinions of thousands of customers, rather than just a few. Social media applications are, amongst other, large sources for unstructured text data. The workshop explores common methodology used to analyse large complex text data sets. Although examples are illustrated using SAS Text Miner, the general approach is software independent.

_____

Biostatistics: Applied Meta-analysis using R

3 December 2015

Prof Din Chen, University of North Carolina at Chapel Hill

A workshop on meta-analysis, a very important field of statistics allowing for combining results from various statistical studies, circumventing the need for new data collection.

This workshop is based on the book: "Applied Meta-Analysis Using R (2013)" published by Chapman and Hall/CRC. This workshop provides a most up-to-date development and a thorough presentation of meta-analysis models for clinical trial and biomedical applications with detailed step-by-step illustrations and implementation using R. The examples are compiled from real medical and clinical trial literatures and the analyses are illustrated by a step-by-step fashion using the most appropriate R packages and functions which should enable attendees to follow the logic and gain an understanding of the meta-analysis methods and R implementation so that they may use R to analyze their own data.

Outline

Session 1:

- Brief introduction to R
- Overview to meta-analysis for both fixed-effects and random-effects models in meta-analysis. Real datasets in clinical trials are introduced along with two commonly used R packages of "meta" and "rmeta"
- Meta-analysis models for binary data, such as for risk-ratio, risk difference and odds-ratio
- Meta-analysis models for continuous data, such as for mean difference and standardized mean difference

Session 2:

- Methods to quantify heterogeneity and test the significance of heterogeneity among studies in a meta-analysis and then introduce meta-regression with R package of "metafor".
- Meta-analysis methods for individual-patient data(IPD) analysis and meta-analysis (MA)
- Meta-analysis methods for rare-events which is timely for clinical trials of adverse-events.

· Multivariate meta-analysis and other relevant topics in meta-analysis.

Complex sampling

3 December 2015

Steven G. Heeringa, Institute for Social Research
University of Michigan, Ann Arbor, MI

This one day workshop will provide participants an overview of sampling designs and methods that are fundamental to the practice of data collection, estimation and inference in household, business and agricultural survey programs.

The **morning session** of the workshop will cover the following concepts and techniques that are essential to survey practice including:

- Survey populations, sample frames, sample units and observational units
- Simple techniques for selecting samples of population elements
  - simple random sample selection
  - systematic sampling from list frames
- Design effects—balancing precision of estimates and costs in practical sample designs
- Stratified sampling to improve precision of estimates, support subpopulation estimation
  - Defining strata of elements
  - Stratified random sampling with proportional allocation
  - Disproportional stratified sampling, optimal allocation
- Cluster sampling to make sampling feasible and cost effective
  - Intra-class correlation in clusters of elements
  - Sampling clusters with equal probability
  - Sampling clusters with unequal probability, PPS sampling

The **afternoon session** will focus on more advanced topics in survey sampling practice including design-based techniques for estimation and inference in survey data analysis:

- Multi-stage sample designs
  - Application to surveys of household populations
- Weighting for sample selection, nonresponse and poststratification/calibration
- Analysis and inference from complex sample survey data
  - Design-based estimation and inference
  - Weighted estimation of population statistics
  - Variance estimation for estimates from complex sample surveys

---

Merging game theory and risk analysis in optimal defense of complex stochastic systems

3 December 2015

Dr Gergory Levitin, The Israel Electric Corporation

This workshop will be presented by Dr Gregory Levitin as an extension of the special session on the broader topic of stochastic processes. This workshop will be a satellite workshop allowing non-delegates off campus to also partake via video conferencing in order to allow for a wider audience.

Business Statistics using SAS Enterprise guide
ANOVA, Regression, and Logistic Regression
3 and 4 December 2015

This course is designed for SAS Enterprise Guide users who want to perform statistical analyses. The course is written for SAS Enterprise Guide 7.1 along with SAS 9.4, but students with previous SAS Enterprise Guide versions will also get value from this course. An e-course is also available for SAS Enterprise Guide 5.1 and SAS Enterprise Guide 4.3.

Learn how to:

- generate descriptive statistics and explore data with graphs
- perform analysis of variance
- perform linear regression and assess the assumptions
- use diagnostic statistics to identify potential outliers in multiple regression
- use chi-square statistics to detect associations among categorical variables
- fit a multiple logistic regression model.

_____

Plenary Abstracts

(In Alphabetical Order)

Statistical Meta-Analysis and its Efficiency

Plenary speaker: Din Chen, University of North Carolina at Chapel Hill

It is natural, rather than the exception, that the data collected to address the same/similar scientific question come from diverse sources (such as, multi-center clinical trials, multi-regional intervention studies). The art and science of synthesizing information from diverse sources to draw a more effective inference is generally referred to as systematic reviews or further as meta-analysis. The statistical meta-analysis is to analyse the data quantitatively from the systematic reviews to draw a more powerful statistical inference. This talk will start with an overview to meta-analysis with summary statistics on both fixed-effects and random-effects models to incorporate within/between-study variations and further discuss a research direction on relative efficiency for random-effects meta-analysis model using study-level summary statistics and individual patient-level data.

Data Revolution

Plenary speaker: Pali Lahohlo, Statistics South Africa

Official Statistics 4.0

Plenary speaker: Walter J. Radermacher, Eurostat

Modern democratic societies need reliable and objective statistics to function properly. But statisticians are facing challenges as society is evolving rapidly and becoming more and more data and technology driven.

Therefore, we need to look for appropriate solutions to: stay relevant, to cope with the complexity of the society and its dimensions, to adapt our skills to match new user needs (possibly through new partnerships), to manage the increased amount of basic (big) data and to elaborate it in statistics and indicators.

Is the Statistics Profession Prepared for the World of Big Data?

Plenary speaker: Robert N Rodriguez, SAS Institute, 2012 President, American Statistical Association

The demand for statistical skills has never been greater in areas of business, government, and research where customer value, policy-making, and scientific discovery are increasingly driven by new sources and novel uses of data. According to McKinsey Global Institute, the United States alone will face a shortfall of 140,000 to 190,000 “deep analytical positions” by 2018.

This demand is the result of five trends: the growth of business analytics, the phenomenon of Big Data, the arrival of data science, the power of distributed computing, and the prevalence of unstructured textual data. In response to these trends, we must differentiate the contributions of statisticians from those of data scientists. Training the next generation of statisticians with the technical and leadership skills needed in emerging areas of practice will equip them for unprecedented career opportunities and impact.

On being a sage statistician and the role of conditional calibration

Plenary speaker: Donald B Rubin, Harvard University

The sage statistician tries to adhere, at least approximately, first to principles of good frequentist inference, which entails being calibrated, and second to principles of good Bayesian inference, which entails being conditional on observed values. Attempting to satisfy both desiderata leads the sage statistician to conditional calibration, a rather obvious but apparently somewhat recondite idea.

Special Sessions Abstracts

(In Alphabetical Order)

A New Approach To Approximating The Distribution Of Aggregate Discounted Claims

Presenter: Franck Adekambi, University Of Johannesburg

We illustrate how alternating renewal process can be used for the actuarial modeling of health insurance policies. No previous research has applied the cumulative function and the moment generating function of the discounted value of the aggregate amount of benefit paid out up to the end of the nth sickness period, n = 1;2;3; : : :. But from practical point of view these two expressions are difficult to evaluate. This research thus utilised an approximation of the discounted value of the aggregate amount of benefit paid out up to the end of the sickness period, and for the case of constant force of interest. The approximation will for example be useful to calculate the insurers probability of ruin, which is the probability that the discounted value of the aggregate amount of benefit paid out exceeds the premium received and the insurers initial capital.

Erlang distributions with different parameters are used for both the periods of health and of sickness, and illustrations are presented in tables 1, 2 and 3 for a constant force of interest.

The Four-Parameter Weibull-Logistic Distribution And Its Properties

Presenter: I.A Adeleke, Department of Actuarial Science and Insurance, University of Lagos, Nigeria

Co-author(s): Akarawak, E.E.E, Olalude, G. A., Okafor, R. O. Department of Mathematics, University of Lagos, Nigeria, Department of Statistics, Federal Polytechnic, Ede, Osun State

This study introduced the four-parameter Weibull-Logistic distribution using the Transformed-Transformer framework by combining the Weibull distribution with Logistic distribution. Properties of the resulting convolution are extensively investigated, viz; rth non-central moments, Shanon's entropy, quantiles, survival function and hazard function. Plots have been presented and simulation study was carried out to study the behaviour of the Weibull-Logistic distribution. It is found that the Logistic distribution is a special case of the four-parameter Weibull-Logistic distribution which is unimodal, skewed and normal-type for some values of the shape parameter. The distribution is also found to relate with the Weibull distribution through its quantile function, a general feature of the T-X family. Simulation results show that the shape of the distribution approaches symmetry as the sample size increases. The likelihood functions for estimating the parameters of the distribution are also presented. The applicability of the distribution has been demonstrate.

Process Capability Index Based Control Chart For Variables

Presenter: Olatunde Adeoti, University of South Africa

Co-author(s): Prof John Olaomi (Department of Statistics, University of South Africa)

The capability index, Cp is presented in the literature to form a complementary measure of process performance but does not address the issue of statistical control. In this paper, we introduced a process capability index based control chart for variables using Downton estimator with specified Cp value which is able to address the issue of control and capability. We also provide control chart constant for constructing the process capability index based control chart. Numerical example is presented to explain the application of the proposed chart and the effect of non-normality is discussed. The result shows that the proposed control chart performs better in monitoring and assessing processes and eliminates the usual two-stage procedure in the literature.

On the T-X Families of Continuous Distributions

Presenter: Eno Akarawak, University of Lagos

Co-author(s): Adeleke, I.A. (Department of Actuarial Science and Insurance, University of Lagos) and Okafor, R.O. (Department of Mathematics, University of Lagos)

In this article, the T-X framework is used to obtain families of continuous distributions involving three symmetric distributions: the Normal, Logistic and Cauchy. The cumulative distribution functions (cdf) of the resulting Normal-X, Logistic-X and Cauchy-X families of distributions arise from the logit of any random variable X. In particular, the Logistic-Exponential distribution has been defined, studied and applied. Results show that it can be used to model real life data.

Models and methods for analysing (Big) datasets

Presenter: Montaz Ali, SCHOOL OF COMPUTER SCIENCE AND APPLIED MATHEMATICS, UNIVERSITY OF THE WITWATERSRAND

This talk discusses the current issues such as complexities, existing methodologies, and other variety of inherent difficulties involved in dealing with Big-Data. A number of mathematical tools such as, mathematical models, numerical linear algebra, and optimization used in analysing Big-Data sets are presented. A number of (Big) datasets arose in various applications areas such as telecommunication, mining, bioinformatics, and logistics industries in South Africa, which are of both cross-sectional and longitudinal types, are presented and analyses are shown. Some shortcomings of existing methodologies are discussed.

Some new attractive families of distributions and associated issues

Presenter: Narayanaswamy Balakrishnan, McMaster University, Canada

Distribution theory continues to be an attractive and active area of statistical research, with practical problems and considerations motivating the construction of several new families of distributions and models. In this talk, I will describe some newly introduced families of distributions and explain their attractive features, characteristics and properties. Next, I will describe some associated inferential and model selection issues. Finally, I will present some examples to motivate these distributions. In concluding, I will highlight some of the problems that remain open for further work!

Modelling Volatility in Stock Returns: Case Study 3 JSE Companies

Presenter: Forbes Kaseke, University of KwaZulu- Natal

For investors and policy makers such as governments, the uncertainty of returns on investments is a major problem. The aim of this paper is to study volatility models for financial data for both univariate and multivariate case. The data to be used is monthly and daily asset returns of three different companies. For the univariate case the main focus

is on GARCH models and their subsequent derivatives. Here clearly the GARCH(1,1) outperformed the ARCH and higher order GARCH models. For the Multivariate volatility models all models gave very similar results. Various distributional assumptions such as normal and Student t distributions were assumed for the innovations. Student t and Skewed Student t distributions were more effective because of their ability to capture fat tails of the distributions.

Multilevel Modelling of Event Histories in Family Formation and Dissolution Studies in the sub-Saharan Africa

Presenter: Jesca Batidzirai, University of KwaZulu- Natal

Co-author(s): Manda, S.O.M (Biostatistics Research Unit, South Africa Medical Research Council, Pretoria)and Mwambi H.G (School of Mathematics, Statistics & Computer Science, University of KwaZulu- Natal,)

In family formation and dissolution studies, a subject may experience several events including childbearing, marriage, divorce and new marriage over time yielding event histories. We may be concerned in studying simultaneously the occurrences of two or more of these different events, adjusting for a number of socio- economic factors. In a typical application, the resulting data are in a multilevel structure. Using discrete time survival as a basis, multinomial logistic and competing risks models are used to fit multilevel multistate models to a typical family formation dataset from Sub-Saharan Africa

Tail estimation in a bounded world: bounded or unbounded models?

Presenter: Jan Beirlant, Department of Mathematics, LStat and LRisk KU Leuven, and Department of Mathematical Statistics and Actuarial Science, University of the Free State

Co-author(s): Tom Reynkens, Department of Mathematics KU Leuven, Isabel Fraga Alves, Department of Statistics, University of Lisbon, Ivette Gomes, Department of Statistics, University of Lisbon

In extreme value analysis, natural upper bounds can appear that truncate the probability tail. At other instances ultimately at the largest data, deviations from a Pareto tail become apparent. This matter is especially important when extrapolation outside the sample is required. Given that in practice one does not always know whether the distribution is truncated or not, we consider estimators for extreme quantiles both under truncated and non-truncated distributions.We make use of the estimator of the tail index for the truncated Pareto distribution first proposed in Aban *et al* (2006). We also propose a formal test for truncation in order to help deciding between a truncated and a non-truncated case. In this way we enlarge the possibilities of extreme value modelling using Pareto tails, offering an alternative scenario by adding a truncation point T *that is large with respect to the available data*. Finally a method for reconstructing the underlying non-truncated distribution tail on the basis of truncated data is provided. Truncation can also occur for instance in the Gumbel domain comprising exponential, Weibull or lognormal distributions. We use a pseudo-maximum likelihood approach generalizing the classical Peaks over Threshold approach in order to have a method that does work for Pareto and light tails. This work is motivated using practical examples from different fields such as earthquake modelling, car liability insurance, and modelling of river flows. We provide simulation and asymptotic results.

Feature selection for kernel models by means of stepwise selection and regularisation: a comparative study

Presenter: Surette Bierman, Stellenbosch University

Kernel models have become popular as widely applied tools in a diverse array of disciplines. Their application scope ranges from astronomy to computational biology, combinatorial chemistry, environmental sciences and hyperspectral image classification. In all of these application areas, if well-calibrated, kernel models are known to yield state-of-the-art accuracies (cf. for example Li et al., 2015; and Wang et al., 2015).

Despite kernel models being regularised models, it has frequently been shown that also in their case, post-selection accuracies are generally higher than those based on the full set of available features (see for example Steel et al., 2011). Many proposals regarding feature selection for kernel models can be found in the literature. There are also many papers available where feature selection for kernel models has been applied to solve real-world problems. The most recent examples include Tomar and Argawal (2015), and Chen and Liu (2015).

We distinguish two approaches toward feature selection for kernel models that can be found in the literature, viz. stepwise selection, and selection by means of regularisation. These approaches are discussed, followed by a comparative study based on benchmark- and simulated datasets.

The Dynamic Probit Model: A Tool for Forecasting

Presenter: Francis Biesmans, University of Lorraine, France

Co-author(s): Igor Litvine (Centre of Expertise in Forecasting, NMMU, South Africa)

Static qualitative binary models are well known in the statistical and econometric literature. Their dynamic version was developed much later (see, for example, Zeger and Qaqish (1988)).

Furthermore, these models perform successfully in prediction. Peculiarly, the dynamic probit model has shown its superiority compared to traditional econometric or leading indicators models, if we agree that the test for all forecasting models is their out-of-sample accuracy.

The structure of the contribution is the following. In the first section, we present the dynamic probit models. The second section is devoted to their estimation by maximum likelihood. The next part studies how the model can be used to forecasting aims. Finally, an application to the prediction of recessions is given.

Long Memory and ARFIMA modelling: The case of CPI inflation rates of Ghana and South Africa

Presenter: Alexander Boateng, University of Limpopo

Co-author(s): Prof Maseka, L (Department of Statistics and Operations Research, University of Limpopo), Prof Gil-Alana, LA(Faculty of Economics, University of Navarra), Prof Hlegani, S (Department of Mathematics and Applied Mathematics, University of Limpopo),Prof Bele

Long Memory and ARFIMA Modelling: The case of CPI Inflation rates of Ghana and South Africa

Alexander Boateng *, Luis Alberiko Gil-Alana2, 'Maseka Lesaoana1, Hlengani Siweya3, Abenet Belete4, 1*Department of Statistics and Operations Research, University of Limpopo, South Africa, 2 Faculty of Economics, University of Navarra, Pamplona, Spain, 3 Department of Mathematics and Applied Mathematics, University of Limpopo, South Africa, 4. Department of Agricultural Economics, University of Limpopo, South Africa,

This study examines long memory or long-range dependence in Consumer Price Index (CPI) inflation of Ghana and South Africa using Whittle methods and autoregressive fractionally integrated moving average (ARFIMA) models. Standard I(0)/I(1) methods such as Augmented Dickey-Fuller (ADF), Philips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests were also employed. Our findings indicate that long memory exists in the CPI inflation rates of both countries. After processing fractional differencing and determining the short memory components, the models were specified as ARFIMA (4,0.35,2) and ARFIMA (3,0.49,3) respectively for Ghana and South Africa. Consequently, the CPI inflation rates of both countries are fractionally integrated and mean reverting.

Classification of Timber Genotypes for Chemical Pulping Using Piecewise Regression and Kernel Density based Clustering

Presenter: Oliver Bodhlyera, University of KwaZulu Natal

Co-author(s): Zewotir, T (School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal) and Ramroop, S (School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal)

Chemically bleached wood pulp (dissolving pulp) has a cellulose content of more than 90% and the changes in its chemical properties, over the processing stages, depend on the genotype of the tree being pulped. Raw pulp, which comes after acid bi-sulphite pulping, goes through a number of bleaching processing stages, each with a specific role, to produce dissolving pulp. These processing stages have different effects on the pulp depending on the type of wood genotype that is being processed. The bleaching processing stages can be considered as time points for repeated measurements of the following chemical properties viz., viscosity, lignin, γ-cellulose, α-cellulose, copper number, glucose and xylose. Piecewise regression models were used to compare the changes of the chemical properties of seven pulping tree genotypes throughout the bleaching stages. In order to cut costs on the chemicals used for processing, it is important to identify species/genotypes that have similar chemical properties under the chemical pulping process in order to mix them together for optimised processing. The piecewise regression parameters of the seven genotypes studied were used with kernel density estimation to develop a "mixability index" for the genotypes studied. The process can be adopted for situations were a chemical pulping business has several genotypes feeding into its manufacturing process. Using the index developed in this study, it can be determined which genotypes are optimally mixable for chemical processing.

The use of data visualisation techniques and social media channels to increase statistical awareness and literacy

Presenter: Vienie Botha, Statistics South Africa

Co-author(s): Mr Kevin Parry

National statistical offices are an important source of information for evidence based decision making. However, the standard methodology of releasing statistics makes it difficult for the average citizen to comprehend the importance and value of official statistics to their lives. The use of data visualisation techniques has made statistics more accessible to the person on the street.

This paper outlines a case study on how Stats SA undertook a paradigm shift to lead in the use of data visualisation techniques and social media in the dissemination of official statistics in Africa. In this way, the organisation aims to increase its statistical reach, thereby increasing statistical literacy and awareness in the country. This will enable the delivery of "The South Africa I know, the home I understand".

Demand Forecasting for Inventory Planning

Presenter: Erin Bromley-Gans, UTi

Co-author(s): Moodley, K and van der Byl, C

Inventory planning is essential in ensuring that businesses have the correct products, in the correct quantity, at the correct place and time. Carrying too much inventory has significant cash flow implications and incurs excessive holding costs (warehousing, insurance, etc.). Insufficient inventory levels impose a risk of stocking out – resulting in lost sales and, potentially, lost customers. It is therefore imperative to strike the balance between over- and under-stocking. In many cases, such as manufacturing or overseas suppliers, lead times for obtaining inventory are significant. As a result, reliable and accurate demand forecasts are an essential part of the inventory planning process. This presentation will provide a background to inventory planning and discuss selected issues in forecasting for the purpose of inventory planning.

Robust mixed effects regression models with application to colony forming unit count and time to positivity in TB research

Presenter: Divan Aristo Burger, University of the Free State and Quintiles, Biostatistics

Co-author(s): Prof. Robert Schall

The bactericidal activity of tuberculosis drugs is characterized using regression modeling of colony forming unit (CFU) count and time to culture positivity (TTP) over time. Typically, most CFU and TTP data points deviate little from the fitted regression curve, but gross outliers are occasionally present and can markedly influence estimates of the rate of change in CFU count and TTP which are the parameters of interest. We fitted Bayesian nonlinear mixed effects (NLME) regression models to

data from various trials. Those regression models allowed for heavy-tailed distributions of residuals and random effects in order to provide fits which are robust to outliers and skewness in the data. Model comparison statistics such as the deviance information criterion and compound Laplace-Metropolis Bayes factors were calculated to discriminate between candidate models.

The Business of Counting: From practical considerations to value extraction.

Presenter: Pravin Burra, Customer Insights & Analytics, Standard Bank

Reality is that despite an abundance of smart people many businesses battle to realise the full benefit of analytics within business. There are a couple of simple themes that underlie this value extraction deficit. I will provide a simple framework to help increase the ability of organisations to realise the true opportunity of analytics moving from theory to practice.

Constructing, validating, interpreting and presenting household food insecurity measures.

Presenter: Carlo Cafiero, FAO Statistics Division, Rome

Co-author(s): Nord, M., Viviani, S.

In a quest to provide means to monitor progress in the fight against hunger, researchers from academia, governmental and international institutions have proposed literally hundreds of indicators related to household food security over the last two decades. Unfortunately, most of the proposed indicators yield measures whose reliability is difficult to quantify and that cannot be properly compared across both time and space, for lack of a proper statistic formulation and of a valid standard of reference against which they can be calibrated.

In this paper we argue that a minimal set of requirements for an indicator to be considered a proper measure includes: (a) the definition of a standard of reference, and (b) a statistical method to evaluate the level of uncertainty that surrounds the measures.

When household or individual food insecurity is conceptualized as a latent trait, the Rasch measurement model (Rasch 1960) provides a suitable theoretical basis to define proper measures. In this paper we present methods developed at FAO to apply the Rasch model to data collected through a simple questionnaire on self-reported food insecurity related experiences and to obtain formally correct measures of severity of the food insecurity condition experienced by individuals or households. We also show how to construct a global reference standard for food insecurity measurement against which to calibrate the measures obtained in different countries, so that classifications in food insecurity classes of different severity can be formally compared.

Examples based on different experience based food security scales (FIES, ELCSA, HFSSM) are used to show how to calibrate them against a common standard and how to present results.

Statistical Methods for Cricket Batting Performance

Presenter: John Calder, Nelson Mandela Metropolitan University

Co-author(s): Sharp, G (Department of Statistics HoD, Nelson Mandela Metropolitan University)

This study reviews the literature on batting performance of an international cricket player. The multiplicative measure of performance as suggested by Barr and Kantor (2004) is then revised based on the optimisation of the power term α against the PriceWaterhouseCooper (PwC) ranking currently used to rank players in the three different formats of the modern game.

A Distribution-Free Generally Weighted Moving Average Control Chart

Presenter: Niladri Chakraborty, University of Pretoria

Co-author(s): Chakraborti, S (Department of Statistics, University of Pretoria), Human,S.W. (Department of Statistics, University of Pretoria), Balakrishnan, N. (Department of Mathematics and Statistics, McMaster University)

Control charts are widely used in the manufacturing sector for monitoring and improving the quality of a process. Assuming a specific underlying distribution when a control chart is designed is often very restrictive because it can severely limit the application of the chart. Distribution-free control charts are therefore useful alternatives when information on the process distribution is partially or completely unavailable. In this regard, we propose a distribution-free generally weighted moving average (GWMA) control chart based on the well-known Wilcoxon signed-rank statistic. The performance of the GWMA-SR chart is compared to a number of existing control charts such as (i) the GWMA chart for subgroup averages, (ii) the GWMA chart based on the sign statistic, and (iii) an EWMA chart based on the signed-rank statistic. Results show that the proposed chart performs just as well and in many cases better than the existing charts.

Socioeconomic determinants of motorcar ownership in South Africa

Presenter: Kenneth Chatindiara, Statistics South Africa

Co-author(s): Naidoo, A (Statistics South Africa)

This paper seeks to explore the socio-economic determinants of motor car ownership by South African households. South Africans own motor vehicles irrespective of their socio economic living conditions in violation of Maslow's hierarchy of needs. Census 2011 data was used in the analysis. Ordinary least squares model building and model-free approaches that identify and analyze the complex relationships that may be imbedded in higher order contingency tables common to census data were performed in the analysis. 20 variables from the Census 2011 data were used in the analysis. Ownership of Motor Cars was used as the dependent variable and the other 19 variables as predictors. SAS Enterprise Guide was used to perform forward selection multiple regression and an R2 of 0.971 was found with 12 out of the 19 variables staying in the final model. ArcMap 10.2.1 was used to perform geographically weighted regression and an R2 of 0.9846 was found. Principal component analysis was done as a variable reduction method and 13 out of the 19 predictors formed the first principal component that explained 45.63% of the total variation. Logistic regression

was performed on the data set in SAS Enterprise Guide and 10 out of the 19 maximum likelihood estimates were significant at 95% level of significance. CHAID analysis was performed in SAS Enterprise Miner and the higher order contingency tables showed that 13 out of the 19 predictors had a relationship with the dependent variable. In conclusion, the results showed that ownership of stove, television, refuse removed by municipality, employment status, and tenure status are all negatively related to motor car ownership. On the other hand, household size, household income, marital status and living in formal dwellings are positively related to motor car ownership.

Interval-Censored Time-to-event Data: From Parametric to Nonparametric Survival Data Analysis

Presenter: Din Chen, University of North Carolina at Chapel Hill

Further to the classical right-censored data, interval-censored time-to-event data are more commonly seen in cancer clinical trials, HIV/AIDS and biomedical studies. This presentation provides a review to the recent development on survival analysis in biostatistics for the interval-censored time-to-event data using parametric, semeparametric and nonparametric models. Specifically, we start with the demonstration on biases inherent in the common practice of imputing interval-censored time-to-event data with the classical survival data analysis and then discuss some recent development with software packages in R and SAS to analyze this type of data.

EVALUATING RISK IN PRECIOUS METAL PRICES WITH GENERALIZED HYPERBOLIC AND STABLE DISTRIBUTIONS

Presenter: Knowledge Chinhamu, University of KwaZulu-Natal

Co-author(s): Huang, C-K(Department of Statistical Sciences,University of Cape Town) and Chikobvu,D(Department of Mathematical Statistics and Actuarial Science, University of the Free State)

Risk management tools such as value-at-risk (VaR) are highly dependent on the underlying distributional assumption and identifying a distribution that best captures all aspects of the given financial data may provide vast advantages to both investors and risk managers. In this paper, we investigate this possibility by establishing the best generalized hyperbolic distributions to fit gold and platinum price returns, while comparisons to stable distributions are also drawn. The adequacy of these distributions are assessed through the Anderson-Darling test, the Akaike information criterion, the Bayesian information criterion and backtesting of their respective VaR estimates.

Hierarchical Logistic Regression for Estimating HIV Prevalence using Survey Data Accounting for Missing Data

Presenter: Amos Chinomona, Rhodes University

Co-author(s): Mwambi, H (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal)

Most practical complex survey data exhibit some multilevel or hierarchical structural form brought about by the prominent features of the sampling design and the underlying target population. These data are often obtained using stratified multistage clustered sampling designs and exhibit a 'clustered' or 'nested' effect that usually induces intra-class correlations of units within clusters. Appropriate statistical inference and conclusions based on such data require methods of analysis that take account of the hierarchical and clustered nature of the data. We compute a hierarchical logistic regression model for HIV on demographic, socio-economic and behavioural variables from a generalized linear mixed modelling framework. The hierarchical models are capable of capturing the layered structure of the data and determine how different layers interact and impact a response variable. An additional complexity often associated with survey data is due to the effect of missing data which cannot be ruled out at the planning stage. Most analyses handle missing data by taking a complete-case approach, that is taking a list-wise deletion of all cases with missing values. This is particularly considered on the assumption that missing values are missing completely at random (MCAR). These approaches often result in potential bias in the estimates due to the differences between the observed and the unobserved values, and loss of statistical information especially if the MCAR assumption does not hold. We perform multiple imputations to fill in missing values with plausible ones obtained from predictive distributions for them accounting for the uncertainty introduced by the very process of imputing the missing values. The research used the 2010-11 Zimbabwe Demographic and Health Surveys (2010-11ZDHS) data that depict prominent multi-layering, clustering and are characterized by missing observations. The results obtained from a rectangular dataset with imputed values is presented together with those from a complete-case analysis for comparative purposes. The results show that HIV status is dependent on one's demographic, socio-economic and beahvioural factors and that there is household to household and cluster to cluster (enumeration area) homogeneity. Estimates with improved precision are obtained with the multiple imputations approach.

Multi-Agent Target Tracking using Particle Filters enhanced with Context Data

Presenter: Rik Claessens, Thales Research & Technology Netherlands/D-CIS Lab, University of Liverpool

The proposed framework for Multi-Agent Target Tracking supports i) tracking of objects and ii) search and rescue based on the fusion of very heterogeneous data. The system is based on a novel approach to fusing sensory observations, intelligence and context data (i.e. the data about the environmental conditions relevant for the tracked target). In contrast to the traditional approaches to target tracking (e.g. maritime or aviation domains), the emphasis is on tracking with low quality data sampled at low frequencies from different sensors dispersed throughout a larger area that may be only partially covered. We illustrate a live, real-time target tracking application that uses a Multi-Agent System approach to find and connect relevant information sources.

Towards developing early warning systems - behavioural modelling from maritime piracy to banking crises

Presenter: Joel Dabrowski, University of Pretoria

The problems associated with maritime piracy and banking crises may appear highly unrelated. In the maritime piracy case it is unknown whether a particular vessel, amongst potentially thousands of vessels, is a pirate or not. Similarly, in financial systems it is often realised too late that a given system is entering a crisis. Both of these problems have complex behavioural dynamics that give indications of whether a specific vessel acts like a pirate or whether economic variables jointly

behave to indicate and impending crisis. The hypothesis is that, through modelling behaviour, such extreme events can be identified before they occur. The difficulty however, is in representing a temporal abstract variable such as behaviour. The dynamic Bayesian network (DBN) offers a solution. The DBN models both causal and time dependent relationships between variables within a probabilistic framework. Within this framework, abstract variables may be inferred from observed data. By defining and relating specific variables, the DBN is shown to provide a promising means to develop early warning systems in such diverse areas of application.

Statistical Analysis Of Gait Data

Presenter: Sonali Das, Csir, South Africa

Co-author(s): B Ganguli, Univ. of Calcutta; Q Louw, Univ. of Stellenbosch; J Cockcroft (Univ. of Stellenbosch); S Sen Roy (Univ. of Calcutta); N Botha (CSIR, Pretoria)

Gait relates to movement, and data related to gait generally comprise anthropometric data, spatio-temporal data and functional data. Given a specific disorder of interest, the analysis can involve looking at the data in a number of ways. In this presentation, our focus is on ilio-tibial band syndrome (ITBS) in long distance runners. Some experiences in analysing the ITBS data will be shared.

3D Expert Knowledge Elicitation for Bayesian Networks

Presenter: Alta de Waal, University of Pretoria

Bayesian networks (BNs) are graphical models that consist of nodes and edges. The nodes depict the variables and the edges depict the causal links between them. The edges have direction and there are no cycles in the network. BNs are flexible in handling new information in the sense that expert knowledge, domain knowledge and data can be fused into one network. Although the graphical depiction of a BN provides researchers with a simple way of constructing models, the inference engine of the BN is very powerful and is designed on the basis of conditional independence. Models are usually constructed with stakeholder involvement in one or more of the modelling stages, which is referred to as 'participatory modelling'. When a BN is constructed, the major modelling issues are 1) What are the variables?, 2) What is the graph structure? and 3) What are the parameters?. The last part - What are the parameters? - usually poses the greatest challenge. We propose a cross-tabulation of variables in order to investigate the relationship between two variables with three-dimensional visualisation (probabilities being the third dimension). This greatly alleviates the elicitation task as the parameters are viewed as probabilistic heatmaps rather than flat conditional probability tables. Furthermore It also improves consistency throughout the probability assignment process. We illustrate the process with a few case studies and propose conventions for applying this method.

Properties of A- and D-optimal row-column designs for two-colour cDNA microarray experiments: Robustness against missing arrays

Presenter: Legesse Kassa Debusho, University of South Africa

Co-author(s): Dibaba Bayisa Gemechu and Linda M. Haines

Two-colour complementary deoxyribonucleic acid (cDNA) microarray experiments are the most important experiments that help scientists to study the expression level of thousands of genes simultaneously. If it is assumed that there is gene specific dye effect in a microarray experiment, then there will be two blocking factors, array and dye. In such cases, the microarray experiments can be considered as row-column designs, with dyes as rows and arrays as columns. Furthermore, the experiments can be described using a linear mixed effects model by taking the arrays as random effects, when comparisons of all possible pairs of treatments are of particular interest. One of the important criteria for a good design is its robustness against a missing observation which may occur due to insufficient resolution, image corruption, or scratches on the slide. This may result in disconnectedness of a design which will lead to loss of precision in estimation and/or of possible comparisons between treatments. The main objective of this paper is to investigate robustness properties of the A- and D-optimal row-column designs against one or two missing array(s). The numerical results show that the robustness of optimal designs against missing arrays depends on the unknown parameter, which is a function of the random array variance and the error variance.

Modelling Extreme Daily Temperature using Generalized Pareto Distribution at Port Elizabeth, South Africa

Presenter: Tadele Akeba Diriba, University of Pretoria

Co-author(s): Legesse Kassa Debusho; Joel Botai

The extremes of daily maximum temperature in summer and daily minimum temperature in winter were analysed using the generalized Pareto distribution (GPD) to the Port Elizabeth weather station data, South Africa. Since extremes in minimum and maximum temperatures series do not follow a normal distribution, the non-parametric methods namely, Kendall's tau test and the Sen's slope estimator were used for the trend analysis. A significant positive trend was observed in the extreme annual minimum temperature. However, the inclusion of a linear trend in the the log-scale parameter in the GPD model for the minimum daily winter temperature did not produce an improvement in the precision of parameter estimates. The results from the return level analysis show that by the end of twenty first century the extreme summer maximum temperature could be about 5 oC higher than the current in Port Elizabeth whereas the change in the winter minimum temperature will be less severe because the return level results suggest an increase of about 2 oC.

Big Data, Data Science and Analytics - the end of Statistics?

Presenter: L. Paul Fatti, Wits University

I will discuss the three concepts: Big Data, Data Science and Analytics, which have sprung up in the last few years, apparently without the involvement of mainstream Statistics or statisticians. They

could be seen as a threat to Statistics but also as an exciting opportunity for the discipline. What should we do to ensure that it is the latter?

New Shock Models Based on the Generalized Polya Process

Presenter: Maxim Finkelstein, University of the Free State

Co-author(s): Cha, JH (department of Statistics, Ewha Womans University, Korea)

Various shock models have been extensively studied in the literature, mostly under the assumption of the Poisson process of shocks. In the current paper, we study shock models under the generalized Polya process (GPP) of shocks, which has been recently introduced and characterized in the literature (Cha, 2014). Distinct from the widely used nonhomogeneous Poisson process, the important feature of this process is the dependence of its stochastic intensity on the number of previous shocks. We consider the extreme shock model, where each shock is catastrophic for a system with proba-bility and is harmless with the complementary probability . The corresponding survival and the failure rate functions are derived and analyzed. These results can be used in various applications including engineering, survival analysis, finance, biology and so forth. The cumulative shock model, where each shock results in the increment of wear and a system's failure occurs when the accumulated wear reaches some boundary is also considered. A new general concept describing the de-pendent increments property of a stochastic process is suggested and discussed with respect to the GPP.

The trials and tribulations of moving towards online teaching

Presenter: Lizelle Fletcher, Department of Statistics, University of Pretoria

Co-author(s): Reyneke, F (Department of Statistics, University of Pretoria)

The University of Pretoria has moved to a blended learning system during the past couple of years. The Department of Statistics specifically introduced both an online homework system with immediate feedback, as well as a cloud-based learning platform, which combines a range of students' learning tools: readings, multimedia, activities and assessments, for their largest first year Statistics course.

The practical challenges faced with implementing online material, especially for large groups, are the focus of this talk. For example, it took weeks to get the more than 2000 students registered for the online homework system before any of the assignments could be graded for marks. This was due to various factors. One being the fact that students lack sufficient computer skills despite being known as the technology generation; another problem was vested in the university's firewall policy which prohibited students to register directly on the website of the online homework platform (Cengage); furthermore, students who have already registered on Cengage for a compulsory computer literacy course that was offered to new first year students were prevented from registering again, albeit for another course. The online e-learning platforms were themselves a further source of problems, e.g. cookies must be allowed, the latest version of Flashplayer installed, and pop-up blockers must be disabled, to name a few.

In conclusion, possible solutions to the many problems we encountered - which were hugely time-consuming - will be briefly discussed.

Multiple State Allocation for Latent Animal Behavioural States based on Hidden Markov Models

Presenter: Victoria Goodall, Nelson Mandela Metropolitan University

Co-author(s): Fatti, L.P. (School of Statistics & Actuarial Science, University of the Witwatersrand) and Owen-Smith, N (School of Animal, Plant & Environmental Sciences, University of the Witwatersrand)

Hidden Markov models have become a popular time series method for the analysis of GPS tracked animals. The behavioural state of the animal is inferred from the latent states identified by the model which are determined using the Viterbi algorithm. A statistical challenge is that an animal will not necessarily remain within one behavioural state for the duration of the time period between successive observed locations. It is likely that during the course of the observation period, the animal will engage in a variety of different behaviours. How this time period is proportioned into these various states is of ecological importance although is not supported by the current Hidden Markov model framework. We used the posterior probabilities of state membership obtained from the Viterbi algorithm as a proxy for the proportion of time allocated to each state within each observation period, in order to investigate how the different latent states contribute to the observed displacements. A simulation study was done to investigate the accuracy of the method, and case studies of ungulate movements in the Kruger and Addo Elephant National Parks were used to illustrate the results. This method allowed us to investigate the smaller scale movements, which cannot as yet be teased apart from the GPS tracking locations.

Probabilistic environmental exposure, effect and risk assessments in the context of potential chemical/nano risk

Presenter: Fadri Gottschalk, ETSS – Environmental, Technical and Scientific Services, Strada, Switzerland

Co-author(s): Andrea Sanchini (ETSS – Environmental, Technical and Scientific Services, Strada, Switzerland)

We introduce the general field of chemical/nano risk assessment and move onto some specific methods and case studies used and developed in the context of environmental release of engineered nanomaterial (ENM). Additionally, R-packages as well as some graphical user interface tools are presented for this kind of exposure, effect, and risk assessments developed in the SUN project on "Sustainable Nanotechnologies" of the EU 7th Framework funding. These tools reflect stochastic and probabilistic material flow analysis (PMFA) as well as probabilistic species sensitivity distribution (PSSD) models and risk predictions. They can be extended to Internet-browser based graphical user interface solutions in order to be used by consultants, researchers, and any kind of industrial partners. The software tools are designed in such a way that they could, for example, build the basis for launching a continuously administered environmental risk database for emerging contaminants with scarce and uncertain data. Exemplary cases are presented for some ENM based

research where quantitative evaluation conducted on very little and uncertain data needs expert interpretation.

Design and Implementation of Distribution-free Phase II EWMA Exceedance Control Charts for Monitoring Unknown Location

Presenter: Marien Graham, University of Pretoria

Co-author(s): Mukherjee, A (Department of Mathematics, IIT Madras, India)

Chakraborti, S (Department of Information Systems, Statistics and Management Science, University of Alabama, USA)

Distribution-free (nonparametric) control charts provide a robust alternative to a data analyst when there is lack of knowledge about the underlying distribution. We examine various aspects related to an efficient design and execution of a class of nonparametric Phase II exponentially weighted moving average (denoted by NPEWMA) charts based on exceedance statistics. The choice of the reference (Phase I) sample order statistic used in the design of the control chart is investigated. We use the exact time-varying control limits and the median run-length (MRL) as the performance metric, since the average run-length (ARL) has certain shortcomings. Detailed guidelines and recommendations for selecting the order statistics of the reference (Phase I) sample for practical implementation are provided along with illustrative examples. We conclude with a summary and some remarks.

Adaptive study design to reduce the size of a Phase II clinical trial for HIV prevention

Presenter: Anneke Grobler, CAPRISA

Clinical trials are often large and expensive. In a climate of reduced funding there is a need to design smaller studies. Two adaptive designs are explored to reduce the sample size of a Phase II HIV prevention study.

The two proposed designs are:

Design 1: A group sequential design is followed, while the study is powered for lower efficacy with aggressive early stopping boundaries. A larger study is designed with 540 HIV negative participants. An interim analysis is done at 16 events. If the interim analysis shows that the efficacy lies in a pre-specified range then the study continued to 33 events. However, if the efficacy at the interim analysis is either very large or very low the study stops at 16 events.

Design 2: A smaller study is powered to detect higher efficacy levels and can be increased. The study starts with 300 HIV negative participants. The study will stop early if a predetermined interim analysis at 16 HIV infections shows no efficacy, or the study can expand to continue up to 37 events if there is a sign of efficacy at this interim analysis. This design follows a promising zone design and will continue at the interim analysis if the conditional power lies within a pre-specified range. This design is not as powerful (unconditionally) as a non-adaptive design powered at the lower efficacy level, but is better than non-adaptive designs powered at the higher efficacy level. In a high HIV incidence area, even with a fairly small Phase II study we could find signals of efficacy, if the products being tested are at least 60% effective.

Analysis of recurrent hospitalisations and deaths in a tuberculous pericarditis multicentre clinical trial

Presenter: Freedom Gumedze, University of Cape Town

In heart failure studies patients experience recurrent hospitalisations but often only the first hospitalization is considered in the data analysis. Such analyses of time to first event are not optimal for a chronic disease such as heart failure as recurrent events are ignored. It is important to quantify the effect of treatment and other risk factors on these recurrent events. In this paper repeat hospitalisations from a multicentre clinical trial were analysed using the Anderson-Gill , conditional (Prentice-Williams-Peterson) and Wei-Lin-Weissfeld models. The analysis of recurrent events should also address competing event of death. We used a joint frailty model to analyse the recurrent hospitalisations and time to death simultaneously. The latter model analyses the recurrent events and accounts for the competing event of death. The latter model can be extended to incorporate a model for longitudinal markers such as CD4 counts which are informative of the recurrent event and terminal event (death) processes.

Incorporating Economic Policy Uncertainty in US Equity Premium Models: A Nonlinear Predictability Analysis

Presenter: Rangan Gupta, University of Pretoria

Co-author(s): Bejiros, S (European University Institute) and Majumdar, A (Center for Advanced Statistics and Econometrics, Soochow University)

Information on economic policy uncertainty does matter in predicting the US equity premium, especially when accounting for structural instabilities and omitted nonlinearities in their relationship, via a quantile predictive regression approach over the monthly period 1900:1-2014:2. Unlike as suggested by a linear mean-based predictive model, the extended quantile regression model with the incorporation of the EPU proxy, enhances significantly the out-of-sample stock return predictability. This is observed especially when the market is neutral, exhibits a side or mildly upward trending behavior, yet not when the market appears to turn highly bullish.

Designs for Small Data

Presenter: Linda Haines, University of Cape Town

The aim of this talk is to demonstrate, by means of two case studies, the usefulness and the relevance of the design of experiments within the modern day context. The first study concerns an agricultural application. The selection of high-yielding varieties of a crop such as wheat from a large number of test lines is of great importance in plant breeding. Only a little seed for each of the test lines is usually available and, as a consequence, the lines cannot be replicated in the field. Designs which accommodate unreplicated treatments and which, at the same time, provide precise comparisons of the yields will be discussed. The second study concerns the design of clinical trials. There is much emphasis in the literature on the asymptotic properties of such designs but in practice the trials usually involve only a small number of patients. Designs involving complete and restricted randomisation will be introduced and, in the spirit of the CoE-MaSS 2015 theme, Stochastic

Processes, the use of the theory of random walks in deriving small sample properties of the designs will be explored.

An Overview of Machine Learning with SAS Enterprise Miner

Presenter: Patrick Hall, SAS Institute

SAS software continually incorporates advances in machine learning research in procedures for classification, prediction, and segmentation. SAS Enterprise Miner now includes many proven machine learning algorithms in its new high-performance environment, which takes advantage of leading-edge scalable technologies. This presentation begins with an overview of machine learning. The remainder of the presentation focuses on examples of supervised and unsupervised machine learning that use SAS Enterprise Miner: performing a classification task using a sparse data source, selecting k for k-means clustering, and dimension reduction using deep neural networks. Code and sample data will be made available.

Bayesian object classification in nanoimages

Presenter: Andries Haywood, University of Pretoria

Co-author(s): Fabris-Rotelli, I (Department of Statistics, University of Pretoria) and Das, S (Advanced Mathematical Modelling, CSIR Modelling and Digital Science) and Wesley-Smith, J (DST/CSIR National Centre for Nanostructured Materials, CSIR)

We discuss the importance of having an automated object classification procedure for classifying nanoparticles in nanoscale images (or nanoimages) and provide an overview of such a procedure, proposed by Konomi et al. (2013), with emphasis on applying the procedure to nanoimages of gold nanoparticles and provide a simplified approach to classifying occluded objects when dealing with homogeneously shaped objects.

Nanotechnology is a fast growing research field with its main applications in medical and material sciences and relates to the manipulation of matter at nanoscale (one billionth of a meter). Nanoparticles have increased surface to volume ratio compared to their bulk form, making them more reactive and useful in material manipulation studies (Tiede et al., 2008). In all applications size and size distribution of nanoparticles is of primary concern, however, particle occlusion is most often an unwanted phenomenon occurring during the image analysis. This shortcoming can potentially lead to unwanted results for particle size measurements and subsequent particle size distributions.

We investigate a semi-automated Bayesian technique (Konomi et al., 2013) to combat the problems faced with occlusion in images obtained using Transmission Electron Microscopy (TEM). The technique, as proposed by

(Konomi et al., 2013), can be seen as a two-stage sampler, where a Markov Chain Monte Carlo (MCMC) setting is used to sample the parameters from the pseudo posterior distribution, with an additional Monte Carlo Metropolis Hastings (MCMH) step to account for the unknown normalising constant. The MCMC steps are used to sample the parameters needed to characterise each object, as well as the number of objects. The samplers used include Metropolis-Hastings-within-Gibbs, Independence and Reversible-Jumps MCMC samplers, each with its own set of complications and considerations. A successful implementation of this algorithm in an image analysis software package, such as ImagePro^{\textregistered}, may lead to great gains in better classification of nanoparticles, and subsequently more accurate size measurements and size distributions can be obtained.

Simulation-Assisted Teaching for Undergraduates in Statistics

Presenter: Annapurna Hazra, University of Kwazulu-Natal

In our blended model, core course content is captured on video for students to engage with, and ,in parallel, students continue to have lecturer contact through weekly workshops. These workshops essentially comprise lecturer-driven tutorials where large groups of students receive help and actively engage in problem solving sessions and, in addition, where tutors are available to assist. This is in addition to weekly small-group tutorial sessions and lab based help sessions. Online forums are provided where students can direct academic questions to lecturers and administrative/logistic questions to the course administrator. Early indications are that the approach is an unambiguous success. In this talk, we review the challenges we have tackled and those we foresee ahead.

Survey Sampling and Big Data: Applications to Survey-assisted Modeling for Populations.

Presenter: Steven G Heeringa, Institute for Social Research, University of Michigan, Ann Arbor, MI

Co-author(s): Berglund, P. (Institute for Social Research, University of Michigan)

Melipillán, E.R. (Program in Survey Methods, University of Michigan)

Contemporary advances in large scale data acquisition, compilation and analysis have many statisticians questioning what the future role for traditional sample surveys will be (Groves,2011; Couper, 2014). Relative to the past where existing administrative data sets or other "big data" often served to calibrate complex survey measurements, future surveys may increasingly be designed to fill in information on unmeasured relationships and address the errors in our big data models of social, economic and public health processes and outcomes.

Two case studies conducted over the past 15 years—the Chilean Social Protection Survey (EPS) and the Aging, Dementia and Memory Study (ADAMS)-- will be used to demonstrate the adaptation of special complex probability sample designs and survey data collection to enhance the usefulness of statistical information in existing large scale administrative or survey data programs. Multiple imputation and machine learning techniques for integrating information from complex sample surveys with existing administrative or “big data” systems will be illustrated.

What are we measuring? Comparison of food security indicators from the Eastern Cape

Presenter: Sheryl Hendriks, Institute for Food, Nutrition and Well-being , University ofPretoria

Co-author(s): van der Merwe, C, (School for Information Technology, University of Pretoria), Ngidi MS (Department of Agricultural Economics, Extension and Rural Development, University of Pretoria), Manyamba C, (Department of Agricultural Economics, Extension and Rural

The development of national food security information systems is constrained by a lack of guidance on which indicators to use. This paper compares food security indicators across two seasons (summer and winter) in one of the most deprived areas of the Eastern Cape province of South Africa. The results show that only anthropometric indicators are sensitive enough to differentiate levels of food insecurity. The lack of consistent classification across indicators means that surveys must use a combination of food consumption and experience of hunger measures backed up by anthropometric measures. Targeting interventions is difficult if the measures cannot be relied on. Further investigation is needed to identify a suite of appropriate indicators for a national information and surveillance system.

Modelling Supervisor-Subordinate Relationship Dyadic Data

Presenter: Jenny Hoobler, University of Pretoria, Faculty of Economic & Management Sciences

Many studies in human resource management, organizational behavior, and industrial-organizational psychology utilize research designs with non-unique supervisor-subordinate dyads. An example of this is where supervisors may have provided performance ratings or other perceptual reports on more than one of his or her subordinates. In just the last 5 years, the management research has come a long way in attempting to model this violation of the assumption of independence. This presentation will provide a summary of my publications which have moved from employing within-and-between analysis (WABA), to relying on intraclass coefficient (ICC) values, to Mplus cluster modelling. I will conclude with recommendations for modelling this type of dyadic data.

Coverage probabilities and average length of generalized confidence intervals for the ratio of scale parameters, difference of location parameters and difference of quantiles of two Weibull distributions.

Presenter: Peter Iiyambo

Co-author(s): Robert Schall

Inference for comparing the parameters and quantiles of location-scale and log-location- scale families of distributions is usually based on the maximum likelihood method. However, parameter estimation using maximum likelihood can be difficult and may require extensive programming. This study compares the coverage probabilities and average length of rank-based generalized confidence intervals for the ratio of scale parameters, difference of quantiles of two Weibull distributions, with the coverage probabilities and average length of maximum likelihood-based generalized confidence intervals. Simulation results show that ranked-based methods are very comparable with maximum likelihood- based in terms of relative efficiency of generalized confidence intervals.

Probabilistic methods for the environmental risk assessment of nanoparticles

Presenter: Rianne Jacobs, Biometris, Wageningen University and Research Centre

Co-author(s): van der Voet, H (Biometris, Wageningen University and Research Centre) and ter Braak, CJF (Biometris, Wageningen University and Research Centre)

Engineered nanoparticles (ENPs) are manufactured and used in many products. During manufacturing and product use, these ENPs may leak into the environment. In the environment, the ENPs may pose a potential risk. It is, therefore, important to perform a good environmental risk assessment. There is, however, not much data and knowledge available on the environmental fate or the toxicity of ENPs because of the novelty of the technology. This lack of data and knowledge results in uncertainty in the risk assessment. In the usual deterministic risk assessment, risk assessors make use of worst case scenarios to deal with this uncertainty.

This deterministic worst case method, however, is not good for two reasons. First, it compromises the transparency of the risk assessment. By using worst case scenarios, it is nearly impossible to explicitly quantify how conservative the risk assessment is. In addition, there is the danger of being over conservative, leading to unnecessary limiting regulation on the technology. Second, variability and uncertainty are not separated. Uncertainty in the risk assessment, can, in principle, be reduced, while variability cannot. When variability and uncertainty are mixed up, however, then it is not possible to reduce uncertainty.

Probabilistic methods can provide transparency in the risk assessment and allow for the separation of variability and uncertainty. First, with probabilistic methods, we know the whole distribution instead of a single worst case value. Knowing the whole distribution gives us the complete picture of the risk and allows us to choose how conservative we want to be. We develop methods to accurately estimate low risks when faced with small sample sizes. Second, probabilistic methods allow us to separately quantify variability and uncertainty. It is then possible to find the source of uncertainty, to study the effect of uncertainty on the risk assessment and in doing so, to identify the areas that need more research to reduce uncertainty. Methods such as 2D Monte Carlo and hierarchical Bayesian methods are used to separately quantify variability and uncertainty.

In this talk, I will discuss some of these methods which help to make the environmental risk assessment of ENPs more transparent and to clearly identify uncertainty sources.

Applying a Structural Equation Model (SEM) to infer a causal relationship between alcohol use and ART adherence

Presenter: Esmè Jordaan, Biostatistics unit, MRC

SEMs make it possible to estimate the causal relationships, defined according to a theoretical model, linking two or more latent complex concepts, each measured through a number of observable indicators, usually called manifest variables. Mplus is a dedicated SEM software package that has many new modelling options implemented that facilitates the set up and appropriate model for the problem at hand such as structural equation models with regressions among combinations of continuous latent variables and observed variables and has various estimators including maximum likelihood and weighted least squares estimators. With censored and categorical outcomes, an

alternative weighted least squares estimator is also available. Robust estimators take into account non-normality of outcomes. The flexibility of estimation methods makes it possible to include variables measured on a variety of scales including continuous, categorical-binary or ordinal, censored and count outcomes.

The application looks at a proposed theoretical model on the cross-sectional relationship between Social Support, Depression, Self-efficacy Beliefs, Alcohol use and anti-retroviral therapy (ART) Adherence. It involves a structural equation model with both continuous latent variables as well as discrete observed variables which includes a count variable with an overabundance of zeros (60%). This is handled by a zero-inflated Poisson regression model which is a mixture of two models, a logistic regression component for predicting the always zero outcomes, and a Poisson regression component for predicting the counts. The covariates used in the two components were different.. Proper analysis of these data involves understanding of the two interdependent outcomes and the results give insight into the role of the two components of alcohol use in this complex model.

Spatial Statistical Analysis to determine Cricket Facilities

Presenter: Max Jordaan, StatsSA

Utilising the Geo-spatial Central Place Centre model in conjunction with the statistical spatial planning methodologies of Stats SA to enhance access to cricket facilities to deliver a meaningful athlete development pathway for disadvantaged cricketers in South Africa. CSA delivers 48 facilities in disadvantaged communities as its legacy to the ICC CWC 2003. This paper is a review on the impact such facilities had within these socio-economically deprived communities. The partnership for the delivery of the CL2003 projects consisted of CSA (UCBSA) as co-ordinator, PPC as technical partner and the local community as implementers in the project delivery within the nominated disadvantaged areas. The process got informed by the Basic Need Philosophy in “building with people” and not for people. Census 2011 data is used for the analysis. The Central place theory was applied to determine the location of the Cricket hubs and the cricket centres of excellence and the gaps were identified. The use of these facilities as cricket Hubs following placed CSA in a position to enter into an Operational Agreement with SRSA and DBE to grow its participation numbers cricket in schools. The participation numbers increased significantly and delivered players to provincial and professional teams as well the national team. An understanding of the population demographics would better inform the spatial distribution and function that would be required for the successful provision of sport facilities within disadvantaged communities. This informed multidisciplinary approach of establishing Central Place Cricket Centres, would provide for increased the participation amongst disadvantaged communities and improve the rate for enhanced athlete performance amongst the disadvantaged groups.

An estimation technique for deriving the Basel LGD on a retail bank mortgage portfolio.

Presenter: Morne Joubert, North West University

Credit risk is defined as the risk/probability that a customer will default due to failure to pay its credit obligations in accordance with agreed terms. If this credit risk realizes then an economic loss may be incurred should the bank not recover all monies due. The bank needs to hold a buffer of capital against peak losses. In the Basel II Accord (BCBS, 2006:52), banks adopting the advanced Internal-Rating-Based (IRB) approach, are allowed to model their own estimates for regulatory capital. The risk components that make up regulatory capital include measures of the probability of default (PD), loss given default (LGD), and the exposure at default (EAD). The paper by Leow & Mues (2012:183) describes an approach whereby LGD is calculated by combining two models. The two models are the haircut model and the probability model. The probability model provides an estimate of the probability of each account undergoing a loss event. The haircut model predicts the difference between the forced sale price and the market valuation of the repossessed property.

The presentation will describe the approach followed by Leow & Mues (2012:183) and will further contain a section on how survival analysis, instead of logistic regression, can be used to predict the probability of a loss event occurring.

A look on additive hazards regression models in survival analysis

Presenter: Gaetan Kabera, South African Medical Research Council

Co-author(s): Mr Paul Gatabazi, University of Johannesburg

Regression in survival analysis is generally based on multiplicative risks models. These models include the semi-parametric Cox proportional hazards model, and the parametric proportional and accelerated hazards models. A few additive risks models have been suggested in the statistical literature. We discuss the Aalen additive hazard model and indicate how parameter estimation may be obtained using counting processes and martingales. A real life example is used to illustrate the theory.

Assessing Influential Observations In Analysis Of Survival Data

Presenter: Tsirizani Kaombe, Department Of Mathematical Sciences, Chancellor College, University Of Malawi

Co-author(s): Manda, S. O. M. (Biostatistics Unit, South African Medical Research Council, Pretoria, Republic of South Africa)

The detection of presence of outliers and influential observations to fitted models is well developed for standard linear regression models. Assuming survival censored data and the proportional hazards model, we discuss a set of residuals including Schoenfeld, Cox-Snell, Score and delta-beta, Deviance and martingale for use with survival regression models. These are discussed according to their usefulness in investigating model adequacy; the functional form for the influence of a covariate; accuracy of the model in predicting subject-specific outcome; leverage applied by each subject in the estimated parameters and proportional hazards assumptions. These are illustrated with a real data set.

Assessing influential observations in analysis of survival data

Presenter: Tsirizani Kaombe, Department of Mathematical Sciences, Chancellor College, University of Malawi

Co-author(s): Samuel O.M. Manda (1Department of Mathematical Sciences, Chancellor College, University of Malawi and South African Medical Research Council, Biostatistics Unit, Pretoria, RSA)

The detection of presence of outliers and influential observations to fitted models is well developed for standard linear regression models. Assuming survival censored data and the proportional hazards model, we discuss a set of residuals including Schoenfeld, Cox-Snell, Score and delta-beta, Deviance and martingale for use with survival regression models. These are discussed according to their usefulness in investigating model adequacy; the functional form for the influence of a covariate; accuracy of the model in predicting subject-specific out- come; leverage applied by each subject in the estimated parameters and proportional hazards assumptions. These are illustrated with a real data set.

Modelling financial data using the Multivariate generalized hyperbolic distribution and Copula.

Presenter: Lionel Kemda, University of KwaZulu-Natal

Co-author(s): Chinhamu, K(School of Mathematics, Statistics and Computer Science,University of KwaZulu-Natal) and Huang, C-K (Department of Statistical Sciences, University of Cape Town)

Financial data usually possess some characteristics, such as volatility clustering, asymmetry, heavy and semi-heavy tails thus, making it difficult, if not impossible, to use Normal distributions to model them. As such, we need to use other kind of distributions which can capture these properties. Statistical analyses show that the Generalised hyperbolic distribution is more appropriate for financial returns estimations. However, we extend our analysis to four dimensional returns. Research shows that multivariate affinely transformed versions of this multivariate generalised hyperbolic distribution present more interesting features than the original distribution. In this regard, we investigate the fit of the multivariate generalised hyperbolic distribution as well as the multivariate affine generalised hyperbolic distributions to four financial indices from the Johannesburg Stock Exchange. Based on the kernel smoothing goodness of fit, the multivariate affine normal inverse gaussian distribution provides the best fit for the affine models. On the other hand, the multivariate generalised hyperbolic distribution based on AIC provides the best model for the four returns without any form of affine transformation on the returns. Finally, the positive tail dependencies exhibited between the all share and Gold mining index as well as all share and S&P 500 is best modelled with the Gumbel and Clayton copulas respectively. While the negative dependencies between the other pairwise returns is modelled with the Frank copula.

Outcomes of being raised by grandparents as the primary care giver.

Presenter: Seipati Kgonthe, Statistics South Africa

In life there is wide of variety of reasons why grandparents are called upon to care for their grandchildren. In most cases grandparents do not choose to take on the role of primary caregiver rather a series of events likely result in grandparents performing this unexpected role. The reasons why grandparents are raising their grandchildren might be because of teen pregnancy, abandonment, parental death, and parental imprisonment, difficulties with finances, military deployment, divorce, and unemployment. The aim of this paper is to investigate the outcomes and living standards of grandchildren raised by their disadvantaged grandparents aged 60+.

Ordinary least square model is fitted to find relationship between grandchildren who have not completed secondary and the socio economic condition of elderly population who are raising grandchildren. The determents of the socio economic background are explored for both old age and grandchildren using census 2011. The results indicate that the grandchildren who have not completed secondary are dependent on elderly with less (income, no schooling, black, household headed by female, rural area, parental deaf, elderly person have never worked before), elderly who have difficulties in (writing, reading and calculating). 9 of the 12 independent variable used shows significant p-values. R^2 shows that 47% of the variation is explained by the model.The spatial analysis is used to compare the various municipalities and the results are mapped using ARG_GIS.

Factors affecting high mortality in Lesotho, 2009

Presenter: Thabo Khule, Statistics SA

Background: Lesotho is amongst the countries in the world to experience high infant mortality. Regardless of government interventions infant mortality remains high. Studies show that infant mortality differs depending on educational attainment, source of drinking water, place of residence, type of toilet facility and wealth index. Method:The study was based on secondary data analysis of the Lesotho Demographic Health Survey (LDHS) 2009 infants with a sample size of 3999.Univariate, Bivariate and the Cox Hazard Regression Model was employed to examine both the unadjusted and adjusted effect of infant mortality. Results: The unadjusted results indicated that infant mortality factors including educational attainment, sex of the child and source of drinking water had significant associations with infant mortality. However place of residence, toilet and wealth index did not have significant associations with infant mortality in the region. The multivariate results indicated that infant mortality was associated with some factors which the study focused on. Exclusively secondary education attainment exhibited the lowest risk of mortality. Conclusion: Mothers' educational attainment can be considered as an important variable towards infant mortality

---

Image analysis in robot soccer

Presenter: Robert King, Department of Statistics, University of Pretoria; School of Mathematical and Physical Sciences, University of Newcastle, Australia

Robocup is an international competition for autonomous soccer playing robots. The NUbots, from the University of Newcastle, play in the kid-size humanoid league of the competition. This talk gives an overview of the image analysis problems that arise in this context. Running on a robot platform brings strong resource limitations for the solutions to these problems.

The playing field is colour coded, but is illuminated (artificially and sometimes naturally) with lights of differing colour temperatures, and is subject to shadows. We have used a number of classification techniques to produce colour identification tables. The field contains a number of

lines. Identifying these lines has required boundary identification and line fitting methods. The centre circle and half circles have added the problem of ellipse fitting.

A review of model-based approaches to small area estimation: An exploratory study

Presenter: Maggie Kisaka-Lwayo, Statistics South Africa

Co-author(s): Caiphus Mashaba, Ngoako Mokgerepi, Neo Mashamba

Household based surveys conducted by Statistics South Africa (Stats SA) are generally designed to provide direct estimates up to metro/non metro level. However the demand for statistics at lower levels (for example, municipal level) has necessitated the exploration of specialized methods for estimation in these small areas. Direct estimates of small areas are likely to be highly inefficient and techniques which "borrow strength" across domains may be advantageous. Popular techniques for small area estimation use implicit or explicit statistical models to indirectly estimate the small area parameters of interest. Methodological developments for obtaining small area estimates which have emerged in the past decades, as well as associated estimation challenges are reviewed. This paper seeks to draw from literature on model based approaches to small area estimation with recommendations for use in Stats SA surveys on the basis of data requirements for the proposed model.

A Bayesian Network Approach to Combating Rhino Poaching in the Kruger National Park

Presenter: Hildegarde Koen, CSIR, University of Pretoria

Rhino poaching is a major problem in the world, especially in Southern Africa. We propose a predictive model in the form of a Bayesian network to calculate a posterior distribution over future poaching events, thereby reducing the areas that rangers need to patrol. A “current perspective” Bayesian network was developed as a first order approximate model of the rhino poaching problem. This was used as a template in an expert workshop to refine the model, as oppose to a clean slate approach. The model will be compared to and evaluated against a data-informed model using model likelihoods and cross-validation.

LONG TERM CARE, THE SOUTH AFRICAN OUTLOOK, PRICING AND VIABILITY

Presenter: Frans Koning, University of the Free State

The whole world is facing an increased problem of longevity, where people live longer and longer. In South Africa this is no different, the logical result of advances in medical technology and methods. As a result longer periods of time are spent in long term care (LTC) situations. With aging populations and lower retirement age in South Africa, an LTC event can have dire consequences to a family, depleting lifetime savings, and placing a financial and sociological burden on them.

This research focuses on modeling transition intencities and trends in a caring institution, for the purposes of developing and pricing insurance products to fund LTC in old and frail situations.

Placing the computer in the students' court

Presenter: Christine Kraamwinkel, University of Pretoria

Co-author(s): Corbett, AD (Department of Statistics, University of Pretoria)

Students in the Four Year Programme at the University of Pretoria come from a diverse range of backgrounds, some being highly computer literate whilst others have never had access to a computer. These students are expected to function on the same level when applying statistical concepts taught in class to electronic data sets in Microsoft Excel. A practical guide was therefore developed in order to aid students in developing the necessary skills. Additionally, guidance was provided to solve problems that may have been encountered during the preparation through hands-on demonstration of concepts using self-evaluation exercises given in the guide. This was however highly ineffective since students were mostly arriving unprepared and sessions were spent rushing between students, often not reaching the ones needing help the most. Both students and instructors were left feeling frustrated and unaccomplished.

In an effort to address this issue and equip students with the necessary problem solving skills, the teaching and learning model was changed so that students have to submit their individual solutions to the self-evaluation exercises electronically. The demonstration of key concepts was moved from practical sessions to lectures. The submission of these exercises can be done at any time that suits the student during the week preceding the practical session with a maximum of 3 attempts allowed. To ensure authenticity, variations of the questions given in the practical guide are assigned randomly to students. Students are given a mark after each attempt but cannot see the correct answers or marks obtained for each individual question. Solutions are released once the deadline has passed, giving students the opportunity to review their work and solve any remaining problems before the practical session. The preparation mark contributes to the final practical mark, thereby attaching a tangible value to the exercise and making the effort more rewarding to students.

Although students initially found the preparation challenging, positive feedback on the effectiveness and value became evident. We will show how this blended learning model has culminated in a significant improvement in the pass and distinction rates on the final practical exam.

Bayesian monitoring of times between events: The Shewhart t_r-chart

Presenter: Nirpeksh Kumar, MG Kashi Vidyapith, Varanasi, India

Co-author(s): Prof. Chakraborti, S. (Department of Information Systems, Statistics and Management Science, University of Alabama, U.S.A.)

The traditional (frequentist) t_r-chart is a Shewhart-type chart useful for monitoring times between events (inter-arrival times) following an exponential distribution. This problem often arises in high-yield processes where the defect rate is low and hence the conventional attribute charts such as the c-chart and the u-chart are often ineffective. We consider this problem under the Bayesian framework and propose a Bayesian t_r-chart when the exponential rate parameter is unknown. The Bayesian t_r-chart is also a Shewhart-type chart that incorporates parameter uncertainty via a prior and a posterior distribution, unlike the traditional chart. The control limits are constructed from the predictive distribution of a plotting statistic. The performance of the proposed

chart is evaluated and comparisons are made with the traditional \(t_r\)-chart. The Bayesian chart is seen to be advantageous in certain situations. An illustrative example is given and some conclusions are offered.

A Study of Dependence Structures in Image Pixels

Presenter: Kwok-Ho Lau, University of Pretoria

Co-author(s): Fabris-Rotelli, I (University of Pretoria) and Bekker, A (University of Pretoria)

We investigate the property of *global independence* and *local dependence*. The idea is that for any one pixel, the dependence it exhibits in relation to neighbouring pixels decreases as the distance between that pixel and a neighbouring pixel increases. The assumption is not unsupported. For a group of pixels that dictate a particular object in an image, they should be in close proximity to one another as well as have high correlation with each other in that group. For a pixel x in an image, the following property is demonstrated:

The correlations of x to its neighbours $y \in N(x)$ decreases as the distance between x and y in the image increases, where $N(x)$ is the set of neighbouring pixels of x. That is, at some distance, the correlations are statistically insignificant. Using the results from above, for any arbitrary video, we can then justify the property global independence and local dependence in images.

Design and Analysis of Cluster Randomised Trials

Presenter: Kerry Leask, CAPRISA

Co-author(s): Quarraisha Abdool Karim, Fanelisibonge Ntombela, Natasha Samsunder, Hilton Humphries, Cheryl Baxter, Anneke Grobler, Janet Frohlich, Salim Abdool Karim

Cluster randomised trials are frequently used in health research and differ from randomised controlled trials in that the unit of randomisation is a group of participants (cluster) rather than the individual participants. The rationale for cluster randomisation, choice of clusters, the effect of matching and sample size calculation will be discussed with reference to the intracluster correlation coefficient and design effect. Methods of analysis of data arising from CRTs will be described. These methods can, however, be restricted if the number of clusters is too small and reasons for this, together with alternative methods of analysis, will be investigated in some detail.

Finally, some of the aspects discussed will be illustrated on data arising from the CAPRISA 007 study. This study was an open-label, matched-pair, cluster randomised controlled trial which evaluated the impact of a cash incentivised prevention intervention on HSV-2 and HIV incidence in high school students in rural KwaZulu-Natal.

Extrapolating business statistics to financial valuations:

Presenter: Gregory Lee, Wits Business School

A formal model and agendaIn the universe of business statistics the ability to extrapolate statistical findings to implications for organisational outcomes – notably financial value – is a valuable but underutilised skill. This paper presents a formalisation of the methodology for extrapolating business statistic to financial outcomes such as return on investment (ROI) or economic value added (EVA). It proposes several issues in this process that are not well understood or applied in many fields, such as appropriate cost of capital benchmarks and time-of-effect functions. A research agenda and set of business ideas are considered, including some data science issues.

Community Survey 2016

Presenter: Pali Lehohla, Stats SA

The initial poverty measurement studies were based solely on income. If a person earned an income below a specific value (poverty line) then he was considered to be poor. Other measures such as the poverty gap and the severity of poverty were calculated

But many researchers believe that a person could experience poverty or deprivation on several levels . This resulted in many indices being calculated, for example the world bank's Human Development Index and Oxford University's MPI Many of these studies predetermine the dividing values for each attribute as deprived or non-deprived. Some studies looked at a functioning index by determining what levels of attributes are required to function adequately. None of these studies take into account the relative deprivation of an individual self classifying whether they consider themselves to be deprived or non-deprived on any specific attribute. A farm labourer with five years of schooling does not consider himself educationally deprived as two or three years more schooling is not going to improve his job or salary. On the other hand, a University tutor with an honors degree hold a temporary position because a masters degree is a prerequisite for the permanent position. This study investigates an index of relative deprivation where individuals self-classify whether they are deprived or not on five attributes, income, education, health, access to services, access to household goods

---

A Measure for the Wicket Taking Ability of Bowlers

Presenter: Hoffie Lemmer, University of Johannesburg

After a match or series of matches the bowler who had taken the largest number of wickets is normally called the best bowler. If two or more bowlers had taken the same number of wickets, they are ranked according to the number of runs conceded, with the one who had conceded the smallest number of runs in the top position. Such a ranking reflects the wicket taking performances of the bowlers, but ignores the fact that some bowlers had bowled substantially more overs than others. In this study the number of overs bowled is also taken into account to arrive at a measure for the wicket taking ability of bowlers. The measure is closely related to the strike rate of the bowler, but it is better because it also takes into account the number of runs conceded per wicket taken. It is argued that the traditional method of giving the award (normally a handy amount of cash) to the bowler, who had taken the most wickets, is unfair. The method presented in this study should preferably be used, but otherwise the award should be given to the bowler with the best strike rate.

Modeling multivariate multilevel continuous responses with a hierarchical regression model for the mean and covariance matrix applied to a large nursing data set

Presenter: Emmanuel Lesaffre, Leuven Biostatistics and statistical Bioinformatics Centre

We propose a novel multivariate multilevel model that expresses both the mean and covariance structure as a multivariate mixed effects model. We called this the multilevel covariance regression (MCR) model. Two versions of this model are presented. In the first version the covariance matrix of the multivariate response is allowed to depend on covariates and random effects. In this model the random effects of the covariance part are assumed to be independent of random effects of the mean structure. In the second model this assumption is relaxed by allowing the two types of random effects to be dependent.

The motivating data set is obtained from the RN4CAST (Sermeus et al. 2011) FP7 project which involves 33,731 registered nurses in 2,169 nursing units in 486 hospitals in 12 European countries. As response we have taken the three classical burnout dimensions (Maslach and Jackson, 1981) extracted from a 22-item questionnaire, i.e. emotional exhaustion (EE), depersonalization (DP) and personal accomplishment (PA). There are four levels in the total data set: nurses, nursing units, hospitals and (for the whole data set) countries. The first model is applied to the total data set, while the second model is applied to only the Belgian part of the data. The two models address the following nurse research questions simultaneously: 1) how much variation of burnout could be explained by the level-specific fixed and random effects? 2) do the variances and correlations among burnout stay constant across level-specific characteristics and units at each level? The two models are explored with respect to their statistical properties, but are also compared on the Belgian part of the study.

We opted for the Bayesian approach to estimate the parameters of the model. To this end we made use of the JAGS Markov chain Monte Carlo program through the R package rjags.

Analysis of South African household poverty based on Income and Expenditure Survey 2010/11

Presenter: Masete Letsoalo, University of Pretoria

Co-author(s): Dr Boraine H, (University of Pretoria and Department of Planning, Monitoring and Evaluations (DPME)), Swanepoel, A, (University of Pretoria)

Measures of poverty are usually estimated using data from national household surveys. The sample design of official household surveys is typically complex, involving multi-stage stratified cluster sampling. Complex sampling affects variance estimation and therefore standard error estimation. Accounting for the complexities of sampling is essential for reliable estimation and analysis. In this paper, we present the estimation of different poverty measures and their standard errors in the case of complex multi-stage sampling design, using the Income and Expenditure Survey conducted in 2010-2011 by Statistics South Africa. Multiple comparisons are conducted using z-test and Bonferroni adjusted confidence intervals to test hypothesis of differences in estimated poverty by gender, population group, settlement type and province of the head of household.

Stochastic systems with reworking

Presenter: Gregory Levitin, The Israel Electric Corporation

Co-author(s): Xing, L (Department of Electrical and Computer Engineering, University of Massachusetts)

In a wide class of repairable and standby systems an element resuming the mission execution after a failure must redo some portion of work already performed before the failure. The considered systems are widely used in applications such as computing and communication. To reduce the amount of work that should be redone, data backup procedures are introduced. On one hand these procedures reduce the amount of work lost after the failures, on other hand they increase the total amount of work in the mission by adding the backup actions. This talk describes different backup techniques and some phenomena specific for such systems (non-coherency, specific role of preventive replacements etc.) A numerical algorithm for simultaneous evaluation of the mission success probability, expected completion time, and cost for systems with reworking will be presented. Due to the non-monotonic effect of the backup distribution on the mission performance indices, we will formulate and solve the optimal backup distribution problem considering different combinations of optimization objectives and constraints. In the case of standby systems with non-identical elements, the elements activation order can influence the mission performance significantly. Therefore, we will also consider an optimal element sequencing problem. Finally, the influence of the backup mechanism's failures on the mission performance will be discussed.

Dating financial cycles with hierarchical method

Presenter: Igor Litvine, NMMU, RSA

Co-author(s): Francis Biesmans (Beta, University of Lorraine, France)

Dating financial cycles is important in investment and forecasting in financial markets. A principally new technique is suggested. We compare this technique with traditional ones (e.g. BB and BBQ algorithm) and with techniques based on Computational Intelligence.

Analysis of Randomised Controlled Trials – some perspectives

Presenter: Carl J Lombard, Biostatistics Unit, South African Medical Research Council

In the pursuit of providing evidence on efficacy and effectiveness of therapeutic and non-therapeutic interventions in the medical and health related fields the simple randomised controlled trial (RCT) has seen many innovations over the past three decades. The analysis of clinical trials now often start at the point of enrolment of participants and carries on during the conduct of the trial until the formal post study analysis. The final trial analysis now often have to take into account the adaptive steps taken during the course of the study as well as the major design features. The conduct of the study also has major implications for the trial analysis with elements such as loss to follow-up and compliance with the treatment. We will discuss and highlight the challenges faced by the trial statistician in the primary and secondary analysis of a RCT and illustrate some through real studies that have been conducted.

Arc length estimation of cumulative distribution functions

Presenter: Theodor Loots, University of Pretoria

Co-author(s): Bekker, A (Department of Statistics, University of Pretoria) and Balakrishnan, N (Department of Mathematics and Statistics, McMaster University)

The arc lengths of cumulative distribution functions are easily calculated through numerical approximations, and may be used for fitting sigmoidal-type functions. The residuals resulting from the parameter fit will be illustrated, along with the distributions of the arc length statistic itself. This method will be compared to other conventional methods, such as maximum likelihood and applyed to various situations where sigmoidal functions arise naturally.

Modelling Heterogeneity for Count Data. A Study of Maternal Mortality in Health Facilities in Mozambique

Presenter: Osvaldo Loquiha, Universidade Eduardo Mondlane/Uhasselt

Co-author(s): Hens,N (Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Universiteit Hasselt), and Chavane, L (Jhpiego, MCHIP Maternal and Child Health Integrated Program), and Temmerman, M (International Centre for Reproductive He

Count data are very common in health services research, and very commonly the basic Poisson regression model has to be extended in several ways to accommodate several sources of heterogeneity: i) an excess number of zeros relative to a Poisson distribution, ii) hierarchical structures and correlated data, iii) remaining \unexplained" sources of overdispersion.

We propose hierarchical zero-inflated and overdispersed models with independent, correlated and shared random effects for both components of the mixture model. We show that all different extensions of the Poisson model can be based on the concept of mixture models, and that they can be combined to account for all different sources of heterogeneity. Expressions for the first two moments are derived and discussed. The models are applied to data on maternal deaths and related risk factors within health facilities in Mozambique. The final model shows that the maternal mortality rate mainly depends on the geographical location of the health facility, the percentage of women admitted with HIV and the percentage of referrals from the health facility.

References:

Bohning, D. (1998). Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence. Biometrical Journal. 40(7), pp:833-843.

Dobbie, M.J., and Welsh, A.H. (2002). Modelling correlated zero-inflated count data. Australian and New Zealand Journal of Statistics.43, pp:431-44

Hall, D. B. and Zhang, Z. (2004). Marginal models for zero inflated clustered data. Statistical Modelling. 4, 161-180.

Yau, K.K.W., and Lee, A.H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. Statistics in Medicine.20,2907-20.

Medication of people living with Cancer in South Africa: A Bayesian approach of statistical analysis

Presenter: Siaka Lougue, University of Kwazulu Natal

Co-author(s): Ogunsakin Ropo Ebenezer

Modern days are still challenged by diseases difficult to treat because of the lack of vaccine as well as serum. Cancer is a specific case of deadly diseases for which no real individual prevention strategy can be observed to avoid the sickness. Only early detected cases can be treated efficiently. Up to date, more than 100 types of cancer have been registered. In the past, lit was considered as diseases of the wealthiest. But, the disease is getting momentum and now highly present among the poor as well. In Africa and particularly South Africa, Cancer is challenging citizens, authorities and all researchers in the domain. This study aims to contribute in knowledge about the behavior the patients living with cancer in terms of medicine consumption. In fact, the general household survey included questions about treatment of patient diagnose with cancer. Because of the small number of observations and to improve the quality of statistical results, Bayesian as well as classical statistical techniques are utilized to analyze the medication of people living with cancer in South Africa.

Analyses of this research are based on the General household survey of South Africa (GHS) 2013. However, data from the same survey in 2012 and 2011 are also used in the Bayesian model to build a prior knowledge. The software R is used for classical statistical analysis and the software WinBUGS for the Bayesian analysis. As a requirement of the Bayesian approach several diagnostic tests were perform to check the convergence of the Markov chain Monte Carlo algorithm and the true reflection of the posterior distribution. Diagnostic test were perform in WinBUGS but also in CODA/BOA. Due to the binary nature of the dependent variable and to take into consideration the geographical structure of the issue, a generalized linear mixed model (GLMM) with binary outcome and logistic link function where perform both using classical techniques as well as Bayesian techniques.

Distribution-free CUSUM and EWMA Control Charts based on the Wilcoxon Rank-Sum Statistic using Ranked Set Sampling for Monitoring Mean Shifts

Presenter: JC Malela-Majika, University of South Africa

Co-author(s): E. Rapoo

Whenever a practitioner is not really sure about the underlying process distribution, alternative monitoring schemes that may be used are called nonparametric (NP) charts. NP monitoring schemes have been shown to have some attractive advantages compared to their parametric counterparts e.g. these are more flexible and very robust. A NP scheme mostly used to monitor the difference in the means of two samples is called the Wilcoxon Rank-Sum (WRS). Using extensive Monte-Carlo simulations, in this paper, we show that using the Ranked Set Sampling (RSS) technique rather than the commonly used Simple Random Sampling (SRS) technique results in CUSUM and EWMA WRS schemes with much better out-of-control detection capability. We thoroughly illustrate this phenomenon by using a variety of run-length characteristics and also using the overall performance statistic called the Relative Mean Index. Based on these, the CUSUM and EWMA WRS based on RSS

yields the best performance compared to a number of its competitors and hence makes it a strong contender in many applications where existing WRS schemes are used.

A Bayesian Modelling Approach for Weighted Survival Data from Non-Proportionally Sampled Strata in Complex Surveys

Presenter: Samuel Manda, South African Medical Research Council

Complex health surveys that collect survival data often employ stratified sampling designs where the strata have not been proportionally sampled. The data may contain values of many covariates pertaining to the survival outcome. A Bayesian proportional hazards model analysis is proposed to find the posterior distribution of the overall fixed effects of the covariates.

The non-proportional sampling does not matter when the fixed effect parameters do not vary across the strata. Otherwise, a disaggregated approach is undertaken where the overall fixed effect parameters are the weighted average of the separate strata fixed effect parameters with weights that are the population proportions. Essentially finding the overall fixed effect estimates this way adjusts the weight of each observation on the overall fixed effect estimates after the modelling process. This method can run into problems when the individual stratum sample sizes are fairly small, and the explanatory variables nearly co-linear within a stratum

We investigate look an alternative approach of reweighing the observations before the modelling process. This reshapes the likelihood to a pseudo likelihood having the shape similar to the likelihood that would have been obtained had the strata been sampled proportionally. Our method of finding the posterior distribution could be considered pseudo Bayesian since we use posterior prior ⍰ pseudo likelihood. Simulations are used to illustrate the proposed methodology, and typical complex sampled survival datasets are used for applications.

A functional data analysis investigation of the relationship between electricity demand and economic indicators in South Africa

Presenter: Siphumlile Mangisa, Nelson Mandela Metropolitan University

Co-author(s): Das, S (Advanced Mathematical Modelling, Modelling and Digital Science, Council for Scientific and Industrial Research, Pretoria, South Africa; and Department of Statistics, Nelson Mandela Metropolitan University, South Africa) and Sharp, G (Department of

We investigate the relationship between electricity demand, assuming it to be smooth curve, and other covariates such as, but not limited to, gross domestic product (GDP), unemployment rate and export of goods and services rate. The covariates considered are either scalar or smooth curve types. We use the functional linear regression approach, which is analogous to multiple linear regression, and may be interpreted similarly, though the inferential questions can be challenging. Our investigation focuses on the South African economy, and questions here include whether there is significant relationship between electricity demand and the other economic variables in the functional framework. Another question is how these results compare to those from the traditional regression approach. Preliminary findings from this investigation will be shared and implications discussed.

A Bayesian capture-recapture model to estimate the survival rate of blue cranes

Presenter: Raeesa Manjoo, University of Witwatersrand

Co-author(s): Supervisor:Fitsum Abadi (School of Statistics and Actuarial Science, University of the Witwatersrand)

Modelling population dynamics is important for the conservation and management of a species. Capture-recapture data is one of the types of data that are analysed in population ecology to estimate demographic parameters including survival rate. Capture-recapture data is different from other data due to the fact that one is unable to observe animals throughout their life time. To analyse this kind of data, one needs an appropriate statistical model that accounts for imperfect detection. In this project, we used the Cormack-Jolly-Seber (CJS) model and its modified versions to estimate the survival probability of the blue crane (Anthropoides paradiseus), which is an endangered species in South Africa. We fitted several candidate models taking into account the biology of the species and implemented the models using a Bayesian framework. The deviance information criterion (DIC) was used to select the best model among the candidate models. Based on the best model, the mean detection probability was 0.0939 (95% credible interval (CRI): 0.0022-0.3420) whereas the mean juvenile and adult survival probabilities were 0.3886 (95%CRI: 0.1549-0.6750) and 0.8085 (95%CRI: 0.5752-0.9460), respectively.

Advocacy and importance of official statistics across all spheres of government

Presenter: Sedikoe Godfrey Mankwe, Statistics South Africa

To research on the legislative reform as to find how can statistics south Africa as the National statistics office can take ownership of all government statistics be it civil registration or any other survey that may need to be conducted.

Modelling nonstationary extremes in the lower Limpopo River basin of Mozambique

Presenter: Daniel Maposa, University of Limpopo

Co-author(s): Cochran, JJ (Department of Information Systems, Statistics and Management Sciences, University of Alabama, Tuscaloosa, USA) and

Lesaoana, M (Department of Statistics and Operations Research, University of Limpopo)

In this paper we fit a time-dependent generalised extreme value (GEV) distribution to annual maximum flood heights at three sites: Chokwe, Sicacate and Combomune in the lower Limpopo River basin of Mozambique. A GEV distribution is fitted to six annual maximum time series models at each site, namely: annual daily maximum (AM1), annual 2-day maximum (AM2), annual 5-day maximum (AM5), annual 7-day maximum (AM7), annual 10-day maximum (AM10) and annual 30-day maximum (AM30). Nonstationary time-dependent GEV models with a linear trend in location and scale parameters are considered in this study. The results show lack of sufficient evidence to indicate a linear trend in the location parameter at all the three sites. On the other hand, the findings in this study reveal strong evidence of the existence of a linear trend in the scale parameter at Combomune and Sicacate, while the scale parameter had no significant linear trend at Chokwe.

Further investigation in this study also reveals that the location parameter at Sicacate can be modelled by a nonlinear quadratic trend; however, the complexity of the overall model is not worthwhile in fit over a time-homogeneous model. This study shows the importance of extending the time-homogeneous GEV model to incorporate climate change factors such as trend in the lower Limpopo River basin, particularly in this era of global warming and a changing climate.

Sample design to optimise the estimation of small micro and medium enterprise owners and their characteristics

Presenter: Thanyani Maremba, Statistics South Africa

The small micro and medium enterprises surveys are the main source of information about owners of small, micro, and medium enterprises, as well as self-employed or individual entrepreneurs. The surveys provide information about the characteristics of businesses in the informal sector and to gain an understanding of their operation and access to services. In order to develop effective interventions for the small business sector, it is important to have a comprehensive understanding of the sector; the specific challenges faced by small business owners, and the capacity they have to deal with these challenges. Interventions should be targeted and evidence-based. Availability of reliable and accurate information with regards to the specific needs of specific segments of the small business sector is therefore a key guide to the development of intervention strategies.

Nationally representative surveys are usually carried out to describe the size and scope of the small business sector as well as to segment the small business sector into homogeneous market segments, with the intention of identifying the development and financial needs. Other objectives include to determine the contribution made by businesses which are not registered for VAT towards economic growth, to collect reliable data about people running businesses which are not registered for VAT, to identify the non-income tax paying and income tax paying businesses within the non-VAT paying businesses, to produce comprehensive statistical information about informal sector businesses, at national and provincial levels.

One of the most challenging tasks confronting sampling statisticians is designing an efficient sample for surveying a rare or hidden population and in this case small business owners are considered as rare population. The population of small business owners both informal and non-formal is usually unknown and makes it difficult to design a probability sample. The study will assess other standard sampling methods that include, use of multiple frames, screening and disproportionate sampling. Methods such as multiplicity, snowballing and network sampling that are usually used in sampling rare and hidden population are considered in designing a sample to estimate small business owners and their characteristics.

Asymptotic approximations for the sum of independent Gamma random variables and for the product of independent Beta random variables

Presenter: Filipe marques, DM, FCT and CMA, Universidade NOVA de Lisboa, Almada, Portugal

The authors show that using well known series expansions it is possible to represent a single Gamma distribution, and also the logarithm of a single Beta distribution, as an infinite mixture of Gamma distributions. Then, using these representations, it is possible to derive simple but accurate asymptotic approximations for the distribution of the sum of independent Gamma random variables and for the distribution of the product of independent Beta random variables. These asymptotic approximations are mixtures of Gamma distributions which match a given number of exact moments. The numerical studies developed support the ease of use and accuracy of these new approximations.

A NEW MODEL FOR MULTIVARIATE CURRENT STATUS DATA

Presenter: Adelino Martins, Eduardo Mondlane University

Individual heterogeneity in the acquisition of infectious diseases is recognized as a key concept, which allows improved estimation of important epidemiological parameters. Frailty models allow to represent such heterogeneity. Coull (2006), introduced a computational tractable multivariate random effects model for clustered binary data. The objective of this report was to apply and modify the proposed model, and compare to the shared and correlated gamma frailty models in the context of the analysis of multivariate current status data. The models were applied to the bivariate current status data on Varicella-Zoster Virus and Parvovirus B19 using different baseline hazard functions for the force of infection. The findings revealed that the proposed model which is called in this report as new correlated gamma frailty model is closely related to existing frailty models. The main difference is the way the multivariate gamma is introduced in the model, and the indirect way to specify the baseline hazard function. In terms of construction, a frailty model is typically formulated based on specification of the proportional hazard function, whereas the new correlated gamma frailty model is built using a classical generalized linear mixed model for clustered binary data. Furthermore, in the new model the variances of the frailties are assumed to be identical, whereas in case of the frailty model, the variances can be different or identical and the correlation is constraint by the ratio of the variances.

Measuring the efficiency of South African municipalities using Data Envelopment Analysis

Presenter: Lehlogonolo Masenya, Statistics South Africa

Co-author(s): Arulsivanathan Naidoo

South African municipalities are expected to utilize the funding they receive to provide basic services to the various local communities under their control. This paper seeks to measure the relative efficiency of South Africa's 231 local municipalities using Data Envelopment Analysis (DEA). DEA is a powerful method widely used in the evaluation of performance of Decision Making Units. Constant and variable returns to scale DEA models were applied on the productive efficiency with which municipal councils have delivered basic services by calculating the ratio of inputs to outputs. The inputs are the municipalities' income from assessment rates, trading services (i.e. electricity and water), and equitable share of grants from National Treasury. The outputs are basic services (access to electricity, to piped water, and toilet facilities). The fundamental assumption behind the method is that if one municipality delivers on basic services to households with a specific amount of income, then the other municipalities should be able to produce the same if they were to operate efficiently. In addition, this information is used to rank the municipalities in order of their efficiency. A spatial analysis is also conducted to examine the clustering of municipalities in terms of their efficiency.

Determinants of Children School Attendance in South Africa

Presenter: Sipho Masimula, Stats SA

Co-author(s): Arulsivanathan Naidoo

This study investigates the determinants of school attendance for children aged between 7 to 14 years in Mpumalanga Province at sub place geographical level using data from Census 2011 from Statistics South Africa. Specifically, school attendance is used as the response variable while proportion of employed head of households, gender of head of household, proportion of children aged between 7 and 14 years with access to computer and those with no access, proportion of females aged between 7 and 14 years, proportion of females aged between 7 and 14 years, proportion of head of households who are Black, Indian, White and Coloured are used as predictor variables. Ordinary least squares (OLS) regression model is used to assess global linear relationship between the variables. Due to spatial dependence of our data, stationerity test is conducted to test whether the coefficients of the OLS regression are space-invariant under the hypothesis that coefficients are stationery across space. Moran's' I autocorrelation is employed to conduct the aforementioned test.

The empirical investigation reveals that Proportion of black head of households, proportion of female headed households, proportion of male headed households, and proportion of employed head of households are significant determinants of school attendance rate since the p-values associated with significance test of all variables are all less than 0.05 and the models appears to fit the data well since it has an adjusted R2 of 09.05 implying that 90.05% of variations in school attendance rate can be explained by proportion of black head of households, proportion of female headed households, proportion of male headed households, proportion of females aged between 7 and 14 years, proportion of males aged between 7 and 14, and proportion of employed head of households. However, Moran's I autocorrelation suggests that the data exhibits some clustering and therefore a need for a localized model arises. Geographically weighted regression (GWR) model is constructed to account for spatial dependence of the regression coefficients and it is found that the GWR models outperforms the global regression model since it has the lowest AIC, 2.4 compared to the OLS model with AIC value of 1.9 and the GWR model fits the data well since it had adjusted R2 of 94.2% compared to the OLS model with adjusted R2 of 90.05%.

Predictors of success and failure in Statistics

Presenter: Lyness Matizirofa, University of Johannesburg

The poor performance of students entering South African universities has been well documented in literature. However, there are many factors which have impacted on their study performance and progress. This study identifies factors which lead to students failing statistics.

A cross-sectional study was carried out in three purposively selected study sites. The study settings are Auckland Park Bunting Road campus, Doornfontein campus and Soweto campus at the University of Johannesburg. A simple random sampling technique was used to recruit 100 diploma students majoring in marketing, accounting and engineering programmes at these campuses. In-depth semi-structured interviews were carried out with the students by a trained interviewer administering a validated questionnaire. The questionnaire includes data regarding students' education, demographic information and socio- economic factors. A pilot study was conducted with ten students to ensure validity and reliability of the instrument. The data was analysed by applying descriptive and inferential statistics. Ethical clearance for the study was obtained from the Research Ethics Committee at the University of the Johannesburg. Written informed consent was also obtained from all the participants.

The results of the study revealed that class attendance has a significant effect on performance in statistics. Achievement of students is negatively correlated with low socio-economic status. Time spent on paid work was found to influence academic performance negatively. This study found a significant positive relationship between lecture and tutorial attendance and performance. Doing pure mathematics in high school was significantly associated with good performance in statistics ($\chi^2=56.281$,$p=0.005$).

There are various internal and external factors to the university that contribute to academic performance of students. Identification of predictors of student's performance is useful in understanding the factors that render students vulnerable to failure and hence permit the identification of vulnerable students. Further research is needed to explore the problem on a large sample including a variety of factors. Since class attendance and doing mathematics in high school were significantly associated with performance. It is recommended that either the university offer bridging courses and, remedial tutorials to fill the gaps in student mathematical knowledge. The importance of class attendance has been clearly identified as strong predictive power to students' good performance. Thus a minimum of eighty percept attendance can be made mandatory.

Is There Hope for Survivalists?| Success In Running a NON-VAT Registered Business In SOUTH AFRICA.

Presenter: Tshepho Brian Matlwa, Statistics South Africa

Business failure in S.A is high, with an estimated 40% of new business ventures failing in their first year and 60% by the second year. In this presentation in order to depict what are the major causes of business failure and how can we develop these businesses thereof we use the secondary data from the Survey for Employers and Self Employed of 2013 for age group from 15 and above conducted by Statistics South Africa. We restricted our analysis to a total of 965 964 Non-VAT registered business in South Africa formed by unemployment and the poverty lines of South Africa 2013 produced by Statistics South Africa was the base guidance in drawing a distinction between success and failure of the business thereof. Furthermore the study is based on multivariate analysis. In the process additional considerations are analysed i.e financial literacy, type of records business do keep etc. The study reveals that these businesses do bring a living to many. This shows that if more attention could be drawn to educating entrepreneurs to pursuit post matric studies then business survival in our country will rise and directly giving rise to the economy at large.

Spatially variability of men and women determinants of unemployment in Limpopo Province

Presenter: Zanele Mazibuko, Statistics South Africa

Co-author(s): Naidoo, A (Statistics South Africa)

Limpopo tends to have the highest proportion of rural dwellers in South Africa, hence it is expected that conditions in the province are inferior to the national average; implying higher unemployment rate. Women's unemployment is a much bigger problem especially when women are the bread winners, due to labour migrant system which take men from their homes to other parts of the country. The goal of this study is to investigate whether there is spatially variability in determinants of unemployment in the different parts of Limpopo, and if so, do determinants differ between men and women. Ordinary least squares (OLS) method was employed to evaluate the relationship between the independent variables and unemployment. We explored the spatially variability in determinants of unemployment in the different parts of Limpopo using geographically weighted regression model (GWR) and investigated if there are differences between men and women determinants using spatial model. These analyses were applied to Limpopo at sub-place geographical level using secondary data from Statistics South Africa Census 2011.

Proportion of females' headed-households, proportion of females' with no schooling, Proportion of black female, Proportion of total fertility rate among women and proportion of females who are married where found to be significant determinants of unemployment and adjusted coefficient of determination for the model was found to be 76.4 percent, which suggests that the OLS model is an adequate model for the data. However, the Moran's Index suggested that the data exhibited clustering pattern hence the OLS model failed to capture spatial dependence of the response variable. Therefore the geographically weighted regression (GWR) model was fitted. The GWR outperformed the OLS model since it had a lower AIC value of 254.6 compared to that of OLS, 381.3. Moreover, the GWR model was superior than the OLS model in prediction power since it had coefficient of determination of 89.7 percent and a lower root mean square error (RMSE) of 12.5 compared to that of the OLS model, 26.4,

---

Generalised Multivaraite Beta Type II Distribution

Presenter: Albert Mijburgh, University of Pretoria

Co-author(s): Bekker, A (Department of Statistics, University of Pretoria) and Human, S (Department of Statistics, University of Pretoria)

An exact closed-form expression of the joint probability density function (p.d.f.) of ratios of independent (but not identically distributed) gamma variables is derived. The components of this new multivariate distribution originate from a Statistical Process Control environment when using a change-point formulation to detect a sustained upward step shift in the variance of a normal distribution or the location of an exponential distribution. This new multivariate distribution extends the work of Adamski et al. (2013) and provides an alternative test statistic for detecting a change-point. In this paper we specifically focus on the bi-variate case and do the following: (i) investigate the statistical properties such as the moments and shape of the joint, the marginal and the conditional distributions; (ii) show the relationship between the new distribution and some other well-known bi-variate distributions with bounded and unbounded domain; and (iii) compare the power of the proposed and existing test statistics (used in the change-point setting) using computer simulation.

Conditional Tail Index and Extreme Quantiles: A Review and Simulation Comparison

Presenter: Richard Minkah, Stellenbosch University and University of Ghana

Co-author(s): Prof. Tertius de Wet, Department of Statistics and Actuarial Science, Stellenbosch University, South Africa

Statistics of extremes has many applications in real life including modelling large claims in insurance, Value-at-Risk of firms in finance, heights and levels of sea dikes in hydrology. The estimation of quantiles begin with that of the tail index and these form a central issue in this field. In this paper, we review and use a simulation study to compare several tail index and quantile estimators in the presence of covariate information. The simulation results show three important findings. Firstly, no estimator of the conditional tail index is universally best. However, the exponential regression model estimator appeared competitive in most instances. Secondly, the local polynomial estimators of the conditional quantiles constituted approximately 70% of the estimators that satisfied the bias-variance criterion. Lastly, we find that the accuracy of a conditional tail index estimator does not necessarily lead to a better quantile estimator.

A spatial analysis of poverty in South Africa

Presenter: Ntokozo Molata, Statistics South Africa

Co-author(s): Naidoo, A (Statistics South Africa)

South Africa has the most unequal income distribution in the world. A large proportion of the population lives below the poverty line. This paper looks into the pattern of poverty in the South African context by using the Multi-dimensional Poverty Index (MPI). This allows for the identification of the most deprived households and communities. Census 2011 data is used to measure severe deprivations that each person or household faces with respect to education, health and living standards. A spatial analysis on the poverty rate at small area level was performed and a spatial clustering of poverty was found in South Africa.

Class of objective priors for a generalised compound Rayleigh model under various loss functions

Presenter: Paul Mostert, Department of Statistics and Actuarial Science, Stellenbosch University

Co-author(s): Van Rooyen, R (Department of Statistics and Actuarial Science, Stellenbosch University)

A generalised compound Rayleigh distribution, with its unimodal hazard function, makes it attractive for modelling lifetimes of patients with characteristics of random hazard rate. The Bayes estimators for some lifetime parameters, as well as the parameters of the generalised compound Rayleigh model, are derived for a right censored sample. The estimators for these parameters are obtained, using the squared error loss function and Varian's linear-exponential loss function, as well as some segmented and general entropy loss functions. A few well-known non-informative priors are derived

for the parameters of the model, especially in the presence of vague prior knowledge. This generalised model is somewhat complicated with respect to the number of parameters in the model that had to be estimated, especially if some of these non-informative priors need to be derived. The derivation depends fundamentally on the Fisher information, which in this case is not obtained in closed-form expressions and need to be approximated. Procedures are implemented to simulate the various non-informative priors, hence a simulation study is carried out to assess the performance of the estimators under these loss functions, as well as under the segmented loss function. An example illustrates the proposed estimators for the generalised compound Rayleigh model.

The use of administrative data to derive synthetic estimates for Micro enterprises- in order to reduce response burden and cost

Presenter: Pinki Mulibana, Statistics South Africa

Co-author(s): Malepe,N(Methodology and Evaluation, Statistics South Africa) and Masemula,M(Methodology and Evaluation, Statistics South Africa)

The current practice for conducting business surveys within Statistics South Africa (Stats SA) is that data is collected from all the sampled enterprises regardless of their size. The size of an enterprise is defined in terms of turnover cut-off points as stipulated in the National Small Business Amendment Bill of 2003; whereby the enterprises are grouped into 4 categories i.e. Medium, Small, Very small and Micro enterprises. Medium enterprises are the main contributors to the survey estimates while Micro enterprises are the least contributors. As such, the Medium enterprises are expected to have a large effect on the precision of the estimates, hence they are fully enumerated (sampled with certainty) in all the business statistics surveys. The Small, Very small and Micro enterprises are sampled with some inclusion probability. It is often difficult to attain the desired response rate for the Micro enterprises mainly due to response burden and collection cost. Most of the Micro enterprises do not have proper operational structures (e.g. accountants and bookkeepers) to keep up with the administrative work which include regularly completing surveys questionnaires either telephonically or electronically, hence high non-response rate; they are also unstable in terms of their existence/ contact information, thus resulting in high untraceable rate. Currently in order to ease response burden of the Micro enterprises, about 20% of sampled Micro enterprises are rotated out of the sample on an annual basis and are guaranteed to be kept out of sample for a period of 5 years. This paper intends to look at a different approach, which is the use of administrative data from various administrative sources such as South African Revenue Service (SARS) to derive synthetic estimates for Micro enterprises rather than collecting information from them. The aim of this approach is to reduce data collection costs and response burden. The paper aims to illustrate the method that can be adopted in generating the synthetic estimates using the auxiliary data i.e. the monthly Value Added Tax (VAT) turnover from the SARS.

Statistical analysis of students' attitudes towards statistics: A case study of undergraduate Bachelor of Science students

Presenter: RUFFIN MUTAMBAYI, UNIVERSITY OF FORT HARE

Co-author(s): Odeyemi , A.O (Department of Statistics, University of Fort Hare)

Ndege, J.O (Department of Statistics, University of Fort Hare)

Mjoli, Q.T(Department of Industrial Psychology, University of Fort Hare)

Qin,Y (Department of Statistics,University of Forte H

Different Methods for handling incomplete longitudinal binary outcome due missing at random dropout

Presenter: Henry Mwambi, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal

Co-author(s): Dr Ali Satty (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal) and Professor Geert Molenberghs (Hasselt University, I-BioStat, 3500 Hasselt, Belgium and KU Leuven - University of Leuven, 3000 Leuven, Belgium)

This paper compares the performance of weighted generalized estimating equations (WGEEs), multiple imputation based on generalized estimating equations (MI-GEEs) and generalized linear mixed models (GLMMs) for analyzing incomplete longitudinal binary data when the underlying study is subject to dropout. The paper aims to explore the performance of the above methods in terms of handling dropouts that are missing at random (MAR). The methods are compared on simulated data. The longitudinal binary data are generated from a logistic regression model, under different sample sizes. The incomplete data are created for three different dropout rates. The methods are evaluated in terms of bias, precision and mean square error in case where data are subject to MAR dropout. In conclusion, across the simulations performed, the MIGEE method performed better in both small and large sample sizes. Evidently, this should not be seen as formal and definitive proof, but adds to the body of knowledge about the methods' relative performance. In addition, the methods are compared using data from a randomized clinical trial.

Stats SA dissemintion

Presenter: Arulsivanathan Naidoo, Stats SA

Small area estimates provide a critical source of information used to study local populations. Statistics South Africa regularly collects data from small areas but is prevented from releasing detailed geographical identifiers in public-use data sets due to disclosure concerns. Many National Statistical offices have used small-area maps based on census data enriched by relationships estimated from household surveys that predict variables not covered by the census.

The purpose of this study is to obtain estimates for small areas for which a few observations are available in the survey. The matching of survey data and census data requires a degree of spatial homogeneity which was assumed when the household data from census 2011 was matched with the

2011 QLFS 3rd quarter survey. . The key assumption is that the models estimated from the survey data apply to census observations. Approximately 28 000 households from the QLFS were matched with the census households.

Small area estimation is a mathematical technique for extracting more detailed information from existing data sources by statistical modelling. The Elbers, Lanjouw ,. Lanjouw, (ELL) methodology was used in this study to determine point estimates for each attribute variable.

The method combines census and survey data to produce spatially disaggregated poverty and inequality estimates. To test the method, predicted estimates for a set of target populations are compared with their true values. Estimates are examined along three criteria: accuracy of confidence intervals, bias and correlation with true values. .

The basic approach is straightforward and typically involves a household survey and a population census as data sources. First, the survey data are used to estimate a prediction model. The selection of explanatory variables is restricted to those variables that can be found in the census and survey data. The parameter estimates are then applied to the census data and the predictions are obtained.

Small area estimation (SAE) is a topic of great importance due to the growing demand for reliable small area statistics even when only very small samples are available for these areas.

The Role of Weighting in the Analysis of Complex Survey Data

Presenter: Ariane Neethling, Department Mathematical Statistics and Actuarial Science, University of the Free State

Co-author(s): Luus, Retha (Department of Statistics and Population Studies, University of the Western Cape) and de Wet, Tertius (Department of Statistics and Actuarial Science, Stellenbosch University)

Many large-scale surveys make use of a complex sampling design. Each observation unit is assigned a sampling weight which is developed in different stages. General practise, according to sampling theory, is to firstly assign a design weight to an observation, adjust it to compensate for non-response after which benchmarking is used to ensure that the achieved sample represents the target population as closely as possible. In practice, some researchers directly benchmark the observed data without first assigning the design weights. Is it advisable to "cut out the middle man"?

The use of different sets of weights will be considered through a comparison of the results obtained in the linear modelling of person income from various explanatory variables identified from the Income and Expenditure Survey of 2005/2006. Since it has been observed that benchmarking methods often result in weights having large variability which could affect the precision of any analyses where they are incorporated, a further consideration in the simulation study will be the application of different weight trimming methods to address this phenomenon.

Applications of Multilevel Modelling in Brand Value Research

Presenter: Deon Nel, University of Pretoria

Despite the proliferation of country-of-origin studies, the role of region in global brand value growth has largely been ignored. Drawing on resource-based theory, this study examines how ˜industry" and "firm" effects have an impact on the role of region on global brand value growth. Hypotheses are tested using a multilevel model on a dataset of 1 555 brand value measurement occasions, representing 260 brands nested within 23 industry sectors, across six global regions, for the period 2006 to 2014. Results reveal that region as a predictor of brand value becomes redundant in a world that is increasingly internationalised; and once a brand is taken up in global valuable brands rankings, brand origin from a regional perspective does not matter. The findings support the theoretical reasoning that brands are relatively more important than industry effects, and that the longer the brand appears in the rankings, the more those high persistence brands will outperform low persistence brands.

Data-driven policy making, impact assessment and accountability: The experience of the Department for Planning Evaluation and Monitoring (DPME)

Presenter: Tsakani Ngomani, DPME

Generalized Orthogonal Procrustes Analysis for the comparison of Multiple Imputed data sets

Presenter: Johané Nienkemper-Swanepoel, Stellenbosch University

Co-author(s): le Roux, NJ (Department of Statistics and Actuarial Science, Stellenbosch University), Lubbe, S (Department of Statistical Sciences, University of Cape Town) and von Maltitz, MJ (Department of Mathematical Statistics and Actuarial Science, University of t

In this paper a regularised iterative multiple correspondence analysis (RIMCA) algorithm is used to apply multiple imputation to missing data in simulated categorical data sets. The multiple completed data sets obtained from the imputation process are generally combined using a prescribed set of rules, referred to as Rubin's rules, which enable the use of descriptive statistics for interpretation.

A different approach is proposed to determine the goodness of fit of the imputed data sets. Instead of using Rubin's rules for combining the multiple imputed data sets to obtain estimates, multiple correspondence analysis (MCA) biplots of each data set are constructed. Generalized orthogonal Procrustes analysis (GOPA) allows the comparison of several configurations with a group average configuration. Therefore, GOPA is used to optimally align the MCA biplots so that they can be visually compared with one another as well as with a group average configuration resulting in a detailed description of the consistencies and idiosyncrasies among the various imputed data sets. Finally the group average configuration of the multiple imputations is compared to (a) the MCA plot of the original complete data set to evaluate the accuracy of the imputation and (b) to results obtained using Rubin's rules.

Gender differentials in housing characteristics and household possessions in South Africa

Presenter: Oupa Nkwinika, StatsSA

South Africa has undergone a number of fundamental political, economic and social changes, since 1994 from the past policies of segregation and discrimination that has left a legacy of inequality and poverty. There is a great deal of research evidence on the racial and gender discrimination in South Africa especially on the labor market, and the impact of such discrimination can be mitigated by government policies, redistributing household of resources and by individual efforts. This study focuses on the gender-based segregation and household well-being, our aim is to examine the relationship between the household poverty and gender, focusing on how the head of household gender affects the household wealth or poverty level. We use Household 10% sample data from Census 2011, to develop a measure of household well-being based on household possessions and housing characteristics by constructing a household wealth index for South Africa. We apply multiple linear regression model to estimate the correlation or the relationship between the household wealth index and the gender, age, employment status, and household income of the head of household. We found that 42% female-headed households are more likely to have fewer adults of working age, mostly consisting of children and elderly which contribute significantly to household poverty. Male headed household are better off economically than female headed household, as the female headed household income are, on the average, earning R62 501.9 which is less compared to that of their male counterparts which is on average R111 780.9. This study supports the proposition that female experience gender inequality in South Africa, as a result that women tend to work in the less profitable sectors of the economy and have lower paying jobs with high unemployment rate than male. Even though there has been an increase in an average annual household income in South Africa, the female-headed household still experiences low average annual income as compared with male-headed households. Thus we conclude that those living in a female-headed household are more likely poor than the male-headed household.

Statistical Capacity Building: Can We Ignore The Online Revolution?

Presenter: Delia North, UKZN

Co-author(s): Zewotir, T (School of Mathematics, Statistics and Computer Science, UKZN)

Statistical capacity building has increasingly become critically important for improving the collection, analysis and dissemination of data for effective functioning of government, private enterprises, public institutions and society in general. With the advances in technology of this era however, citizens now live in a very numerate and highly technical world, so that statistical capacity building in this era has to include initiatives that match the need and available resources. The author will share recent experiences and lessons learnt from various statistics capacity building initiatives, with a particular focus on the relevance of on-line teaching materials in the South African context.

A New Compound Class of Burr Weibull-Poisson Distribution: Properties and Applications

Presenter: Olusegun Broderick Oluyede, Georgia Southern University

A new class of distributions called the Burr Weibull-Poisson (BWP) distribution is proposed and its properties are explored. This new distribution is by far a more flexible model for lifetime data. Some statistical properties of the proposed distribution including the expansion of the density function, quantile function, hazard and reverse hazard functions, moments, conditional moments, moment generating function, skewness and kurtosis are presented. Mean deviations, Bonferroni and Lorenz curves, Renyi entropy and distribution of the order statistics are derived. Maximum likelihood estimation technique is used to estimate the model parameters. A simulation study is conducted to examine the bias, mean square error of the maximum likelihood estimators and width of the confidence interval for each parameter. Applications of the model to real data sets are presented to illustrate the usefulness of the proposed class of distributions.

The skew hyperbolic secant distribution

Presenter: Brenda Omachar, Department of Statistics, University of Pretoria

Co-author(s): van Staden, P J (Department of Statistics, University of Pretoria) and King R A R (School of Mathematical and Physical Sciences, University of Newcastle, Australia)

The hyperbolic secant (HS) distribution is a symmetric distribution with heavier tails than the normal and logistic distributions. This paper proposes a skew generalization of this leptokurtic distribution. The properties of the skew hyperbolic secant (SHS) distribution, including its shape characteristics, are presented. We compare the SHS distribution with Hosking's generalized logistic distribution and discuss the relation between the SHS distribution and the half-Cauchy distribution.

Situation Assessment Exploiting Correlated Data from Disparate, Spatially Distributed Sources:

A Probabilistic Causal Model Approach

Presenter: Gregor Pavlin, Thales Research & Technology Netherlands/D-CIS Lab

Contemporary decision making problems require situation assessment based on large quantities of correlated data stemming from heterogeneous types of sources that are often spatially distributed and belong to legacy systems (stove pipes). Examples are tracking in urban environments, search and rescue, threat assessment in security and defence applications and many more. In order to be able to exploit the existing sources of correlated data, however, multiple challenges of computational as well as of engineering nature have to be overcome. In this presentation we address three related topics: the modelling, inference and implementation.

The key to sound decision making is a combination of (i) models that capture non trivial correlations in the physical domain with a sufficient accuracy and (ii) inference methods for correct handling of the data. As the complexity of the modelled domains is often high (many variables and relations) a systematic, theoretically sound approach to modelling and system design is indispensable for the implementation of tractable solutions. In this presentation we will show that, in a relevant class of problems, these challenges can systematically and efficiently be tackled by using Causal Probabilistic Models (CPM). In particular, we will show how CPMs facilitate the development of modular, loosely coupled plug&play inference systems that correctly capture correlations between hypotheses and various types of data sources as they become available during operation. The key feature of CPMs is the explicit and systematic representation of the dependencies between the modelled phenomena. The related theoretical concepts of d-separation and Markov boundaries facilitate a systematic and simple derivation of modelling fragments that allow sound decentralized inference in systems

consisting of loosely coupled processing modules. Moreover, we will show how CPMs were used for the derivation of a theoretically sound tracking approaach that seamlessly combines a particle filtering process with uncertain knowledge about the environment in which the tracked target moves. With the help of examples and experimental results we will illustrate the impact of naïve treatment of dependencies and the improvements with solutions based on CPM.

The resulting modular fusion solutions, however, require non trivial information flows, often established at runtime by discovering the modules providing the right type of data in the right context (e.g. place, time, clearance, credibility, etc.). This is solved with the help of the Dynamic Process Integration Framework (DPIF), a logistic layer on top of an arbitrary communication middleware. DPIF defines processing modules as interoperable services and supports service discovery as well as automated creation of information flows and their maintenance. Special tools allow fast development of interoperable processing modules.

Inter-Linkages Between Private Investment, Public Investment And Economic Growth In South Africa

Presenter: Sagaren Pillay, Statistics South Africa

This paper firstly investigates the link between private and public investment and secondly the link between total expenditure and economic growth in South Africa. The study is undertaken within the theoretical framework of cyclic causality as expounded on by Phillips in the sixties. Many single country studies have shown mixed results as to whether private and public investment is complimentary. This study adds to the literature by examining empirical data within and error correction framework to investigate the inter linkages between private investment, public investment and economic growth. A statistically significant cointegrating relationship is found to exist between both private and public investment on one hand and total investment and economic growth on the other. The results show that there is a complimentary relationship between private and public investment in South Africa both in the long and short run.

Predicting the future of the 2015 Rugby World Cup using Random Forest variants

Presenter: Arnu Pretorius, Stellenbosch University

Co-author(s): Surette Bierman

Random forests (RFs) are known to yield state-of-the-art performance in a wide array of application domains. Examples include astronomical object classification, digital image classification, text classification and genomic data analysis.

Over the past decade, many RF variants have been proposed in the literature. Fawagreh et al. (2014) provide a good overview. Some important aspects in contributions include: limiting the number of trees voting toward predictions, replacing majority voting with more sophisticated dynamic integration techniques, using weighted random sampling to pick features in the face of a large number of uninformative features, extension to on-line RF algorithms, and the use of genetic algorithms to improve RF performances. More recently, contributions focused on modifications to RFs with a view to enhance performance in the face of high-dimensional data. See for example Nguyen et al. (2015) and Xu et al. (2012) in this regard.

We present some of the more important variants, illustrating their application in the prediction of world cup rugby match outcomes. For this purpose, the use of cloud computing services in training online models is also presented.

Recent Research on Nonparametric Statistical Process Control

Presenter: Peihua Qiu, Department of Biostatistics, University of Florida, USA

Statistical process control (SPC) is widely used in practice, ranging from production line monitoring in manufacturing industries to infectious disease surveillance in public health. Conventional SPC charts are designed based on the assumptions that process observations are independent and normally distributed, which are rarely valid in practice. In this talk, I will discuss some recent research on nonparametric SPC charts that do not rely on the normality and certain other conventional assumptions. Specific topics covered by the talk include univariate and multivariate nonparametric SPC, nonparametric profile monitoring, and dynamic screening systems.

Cholesky-based Covariance Modeling in Longitudinal Studies

Presenter: Anasu RABE, University of Botswana

Co-author(s): Shangodoyin,D.K.(Department of Statistics, University of Botswana) and Thaga,K.(Department of Statistics, University of Botswana)

Cholesky-based parameterization have recently become popular and active area of research in modeling covariance structures of longitudinal responses. However, the proposed procedures are diverse in their technical frameworks and inference. As a consequence, there is need for a unified perspective if we are to appreciate the advantages they offer. In this paper, we attempt to bridge this gap for the continuous longitudinal data by drawing analogies between their key aspects: Modeling framework, parameter estimation and inference, model selection, algorithms, efficiency and parsimony/sparsity of parameter estimates. We conclude with a discussion of the key factors and suggesting some directions for further research.

On our Way to Sustainable Development - Guidance from Statistics

Presenter: Walter J. Radermacher, Eurostat

Sustainable development is about meeting the needs of the present generation without compromising the ability of future generations to meet their needs. Societies have to make difficult choices on their way to progress and prosperity. Official Statistics is asked to provide high quality evidence for these choices.

Indicators, accounts and basic statistics should enlighten citizens, entrepreneurs and policy makers and enable them to make informed decisions.

An application of the extensions of the Cox model to model the incidence of pneumonia and repeat episodes of pneumonia in boys & girls in a low-middle income setting in South Africa: The Drakenstein child health study.

Presenter: J Ramjith, Division of Biostatistics & Epidemiology, School of Public Health & Family Medicine, University of Cape Town, Cape Town, South Africa

Co-author(s): L Myer1, H Zar3, F Little2 2Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa, 3Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town, Cape Town, South Africa

Introduction: Pneumonia is one of the leading causes of death in children under the age of five in developing countries. It is uncommon for a proportion of children to experience repeated episodes of pneumonia. Pneumonia incidence literature favours the Cox proportional hazards (CPH) model to assess the effect of risk factors on time to first episode and Poisson regression models the discrete counts of episodes. As a consequence we fail to consider possible correlation between events within infants' follow-up and further overlook the possibility of a temporal effect of covariates. Extensions of the CPH model to understand recurrent pneumonia have been applied within the health sciences.

Aim: We set out to evaluate extensions of the CPH model when investigating the effect of sex and sex adjusted risk factors on the incidence of repeated pneumonia episodes in a cohort of 1008 infants enrolled in the Drakenstein child health study between May 2012 and April 2015.

Methods: Pneumonia was diagnosed according to the WHO clinical case definitions: any infants who presented with cough or difficulty breathing and age-specific tachypnoea (≥ 50 breaths per min for children aged between 2- 12 months) or lower chest wall in-drawing. Repeated events were any events that happened more than 14 days after a previous event. Standard CPH models were used to investigate risk factors on time to first event stratified by sex. Extensions of CPH, the Andersen-Gill model, the Wei, Lin & Weissfeld model and the Prentice, Williams & Peterson's gap-time and total-time models were then applied for repeat episodes.

Discussion & Conclusion: Parameter coefficients and robust standard errors were reported. Scaled Schoenfeld residuals were used to test the PH assumption. Schoenfeld residual plots were used to assess the overall goodness-of-fit of these models. The models were compared on both their performance and interpretability. This type of analysis will provide further insight into the monitoring of children who are at risk of developing repeat pneumonia episodes.

Acknowledgement: This study was funded by the Bill & Melinda Gates Foundation (grant number OPP 1017641). We thank the study staff; the clinical and administrative staff of the Western Cape Government Health Department at Paarl Hospital and at the clinics for support of the study; and the families and children who participated in the study.

A Comparison Of Rubric Scoring Methods

Presenter: Jacques Raubenheimer, University of the Free State

Introduction and aim: Rubrics are a common evaluation method for oral presentations. Most literature on rubrics discusses rubric application. The scant rubric meta-literature that discusses how rubrics should be constructed and used generally covers:

a) How rubric items should be constructed

b) The topic of inter-rater agreement

c) Which contexts are suitable for using rubrics

One topic that is seldom discussed is the actual values used for scoring rubrics, i.e., the scoring scale, and how this scale should be weighted. Even the few examples found always assume that the rubric will use a limited number of categorical scale points.

This study investigated the issue of rubric scoring, not rubric item content or context, specifically whether, given the same items, better inter-rater reliability was obtained by substituting a percentage based scoring system instead of a rating-scale rubric scoring system.

Methodology: Third year students from four departments of the School of Allied Health Sciences, UFS were asked to participate in a descriptive cross-sectional study at the 4th year research presentations. Those consenting (n=111) rated the presentations using the departmental rubric, randomly assigned as using either a categorical- or percentage-based scoring system.

The inter-rater reliabilities of the two scoring systems were compared by calculating the intraclass correlation and the coefficient of concordance.

Results: For two rubrics, the categorical scale showed better interrater reliability than the percentage-based scale, but the reverse held true, and with greater margins, for the remaining two. Modelling of all possible scoring combinations and the variances so obtained for weighted and unweighted scores showed that pre-weighting scores would potentially provide better inter-rater reliability than post-weighted scores.

Conclusion: Shifting to a percentage-based scoring system for rubrics is not a solution that will work for all raters, and thus a categorical scale with the possibility of refined gradings may hold more promise, although this will have to be investigated in a further study.

A Simulation Comparison of Quantile Approximation Techniques for Compound Distributions popular in Operational Risk

Presenter: Helgard Raubenheimer, Centre for BMI, North-West University

Co-author(s): PJ de Jongh (Centre for BMI, NWU, South Africa), T de Wet (Centre for BMI, NWU, South Africa) and K Panman (Centre for BMI, NWU, South Africa)

Many banks currently use the loss distribution approach (LDA) for estimating economic and regulatory capital for operational risk under Basel's Advanced Measurement Approach. The LDA

requires, amongst others, the modelling of the aggregate loss distribution in each operational risk category (ORC). The aggregate loss distribution is a compound distribution resulting from a random sum of losses, where the losses are distributed according to some severity distribution and the number (of losses) distributed according to some frequency distribution. In order to estimate the economic or regulatory capital in a particular ORC, an extreme quantile of the aggregate loss distribution has to be estimated from the fitted severity and frequency distributions. Since a closed form expression for the quantiles of the resulting estimated compound distribution does not exist, the quantile is usually approximated by using brute force Monte Carlo simulation which is very computing intensive. However, a number of numerical approximation techniques have been proposed to lessen the computational burden. Such techniques include Panjer recursion, the fast Fourier transform, and different orders of both the single loss approximation and perturbative approximation. The objective of this paper is to compare these methods in terms of their practical usefulness and potential applicability in an operational risk context. We find that the second order perturbative approximation, a closed-form approximation, performs very well at the extreme quantiles and over a wide range of distributions and very is easy to implement. This approximation can then be used as an input to the recursive fast Fourier algorithm to gain further improvements at the less extreme quantiles.

A modified class of estimators for estimation of population mean in the presence on non-response

Presenter: Saba Riaz, Riphah International University Islamabad Pakistan

In the present paper, the problem of occurrence on non-response is addressed in the variable of interest. A modified class of biased estimators is suggested for estimating the unknown mean of the study variable using information of the auxiliary attributes. Expressions for the asymptotic variance of the proposed class are derived up to the first degree of approximation. Efficiency comparison of the suggested class is acquired with the linear regression estimator theoretically and numerically. It has been shown that the proposed class of estimators is more efficient than the linear regression estimator.

Methods, Models, Motivation, and More: Recent Developments in SAS/STAT® Software

Presenter: Robert N Rodriguez, SAS Institute

SAS/STAT software is expanding in response to emerging statistical needs in areas as diverse as business analytics, government statistics, and clinical trials. This presentation provides an overview of recent enhancements, emphasizing the practical motivation for novel methods and models—the problems they solve and the benefits they offer. New procedures and features are available for predictive model building with generalized linear models, quantile regression, and generalized additive models; Bayesian choice modeling; analysis of missing data; survival analysis with interval-censored data and competing risks; and item response models.

The Utility of Bayesian Inference in Instrumental Variables Models

Presenter: Don Rubin, Harvard University

The use of instrumental variables models estimated by method-of-moments methods has a long history in economics. Although such methods of estimation definitely have their pedagogical advantages, the Bayesian approach can have conceptual, statistical and inferential advantages, for example, by allowing the investigation of models without exclusion restrictions. These issues are illustrated in simple real and simulation examples.

Influential factors of divorce in South Africa

Presenter: Mulalo Salane, Statistics South africa

Despite the evidence that divorce has become more prevalent among weak socio-economic groups, the knowledge about the stratification aspects of divorce in South Africa is lacking. This paper seeks to analyze the variables that contribute to the increase of divorce in South Africa. In 1996 the divorce was 29 percent, 2001 was 35 percent and 2011 is 36 percent.

Our aim is to examine the relationship between social inequality and divorce, focusing on how household income, education, employment stability, relative earnings, household goods and the intersection between them that contribute to the divorce rate in South Africa. The data used is from the years 2001, 2011 from census data of Statistics South Africa and Department of home affairs (DHA).The variables used for the analysis are gender, highest education level, individual income, population group, and employment status.

The methods used were the linear regression analysis with R2 = 20% and P-value<0.005. Multivariate analysis was used for checking the correlation within the variables and logistic regression for the marital dissolution. The couples in lower socio-economic positions had a higher risk of divorce than those classified as in higher socio-economic in South Africa. Higher educational level in general for both spouses in particular showed a decrease in the risk of divorce. The wife's relative earnings had a differential effect on the likelihood of divorce, depending on household income, a wife who earns more than her husband increased the log odds of divorce. This is then implies that there is a positive relationship between the divorce probability and income inequality per couple.

In conclusion the study shows that divorce indeed has a stratified pattern and that the weaker socioeconomic groups experience the highest levels of divorce. Gender inequality within couples has high impact of divorce.

Reviewing our blend of online and offline learning at introductory level, UCT

Presenter: Leanne Scott, UCT

In the second semester of 2014, UCT Statistical Sciences Department presented its first blended model of STA1000, the largest Statistics introductory course, to 1400 students. This was the start of a new era of teaching in both the department and the Science Faculty at UCT, but was also the culmination of many years of collaborative research into teaching approaches for Statistics. Our research had steered the course to becoming increasingly computer-based, using spreadsheets as a platform for teaching statistics, and decreasingly chalk-board-based. The demonstration of core concepts through visual, graphically-based simulation had shifted the focus to understanding through doing-and-interacting and away from a transmission based, copy-and-learn process.

Stats SA's Poverty and Food Security measurements

Presenter: Nozipho Shabalala, Statistics South Africa

Stats SA conducts two surveys, the Living Conditions Survey (LCS) and the Income and Expenditure Survey (IES) that are primarily designed to measure poverty and inequality in South Africa as well as to serve as input towards the updating of the Consumer Price Index (CPI) basket of goods and services. However, these surveys also contain information that can be used for measuring food security. The presentation briefly covers the design and methodology of these two surveys, i.e, IES and LCS; and their contents with special reference to data items related to food security measurements. The discussion also identifies challenges envisaged with the use of LCS and IES as vehicles for the collection of food security information. Other surveys that are currently measuring some aspects related to food security, such as for example the General Household Survey (GHS), are also briefly discussed. The general aim of the presentation is to share information on official statistics that are available related to food security in South Africa and to spark a conversation on how the LCS and IES can be modified for improved food security measurement in future.

Pro poor public transport: Rea Vaya in the City of Johannesburg

Presenter: Mzi Shabangu, Statistics South Africa

This paper looks at the members of the economically active population living in poor residential areas and previously disadvantage communities who suffer the financial burden of higher public transport fares increase as they commute long distances to and from work on a daily basis.The objective is to demographically characterize the Rea Vaya bus stops/stations with poor and very poor catchment areas within the city of Johannesburg, by testing different models of pro poor fares and what they might cost the city of Johannesburg in subsidy. The different models will be used to see the policy-relevant use of the statistical data and spatial referencing. The method is to link the Census 2011 data with the transport survey, using small area layer to compare. The GIS analytical methods used are the proximity-buffer at 0.2km and geometric area calculation. The results show spatially distribution of the rea-vaya line and the bus stops/stations in relation to where the City of Johannesburg can apply subsidies.

Shewhart-type synthetic and runs-rules charts for monitoring the mean of normally distributed processes

Presenter: Sandile Shongwe, University of Pretoria

Co-author(s): Graham M.A. (Department of Statistics, University of Pretoria)

Statistical process control methods combine the power of statistical significance test with time analysis of graphs – which makes it more advantageous than traditional statistical significance methods in quickly detecting process changes. A control chart is the main tool used for this purpose. Here, we build a general framework for synthetic and runs-rules charts for monitoring the mean of a normally distributed process. That is, we conduct an in depth theoretical and empirical zero-state and steady-state study to gain insight into the design of different types or categories of these charts using the Markov chain imbedding technique. More importantly, we show that the synthetic chart with a modified side-sensitive feature, proposed here, has a better overall run-length performance than its Shewhart-type synthetic and runs-rules competitors.

Does Education Really Disadvantage Women in the Marriage Market?

Presenter: Cleopatra Sikhosana, Statistics South Africa

Co-author(s): Arulsivanathan Naidoo

Women empowerment, as a result of South African post apartheid transformation effort has led to profound changes in female career development and labour force participation. This has raised a subject of concern on the conflict that women face between their roles in career and family. One recurring theme is the "success penalty", or the disadvantage career success poses to women in the marriage market. It is argued that women who achieve career success are failing at what they really aspire to – a successful romantic relationship. The purpose of this paper is to investigate the relationship between education and marriage for South African women aged between 30-60 over the period 1996 to 2011, using S.A. Census data. The study uses logistic regression to measure the likelihood of marriage at the lowest and highest levels of education and test whether the probability of getting married decreases as a woman gets more educated. The study also tracked the relationship between education and motherhood by different age groups and race. Spatial autocorrelation and hot spot analysis was used to study the distribution of never married females across South African sub places. The study found a negative relationship between higher education and marriage for women and not for men.

The impact of using multimedia on students' academic achievement in theoretical Mathematical Statistics courses at UFS

Presenter: Morné Sjölander, University of the Free State

In this study, we examine the impact of moving away from the traditional manner of lecturing (using transparencies on an overhead projector and using a black board) to using multimedia (power point slides with animations). We specifically look at the impact of this on students' academic achievement in second year theoretical Mathematical Statistics courses at the University of the Free State. We compare the difference in marks (i.e. the semester marks, exam marks and final marks) of the first semester to the marks of the second semester of two groups of students. Our experimental group was the 2015 second year students, and they were lectured in the traditional manner in the first semester and lectured using multimedia in the second semester. Our control group is the 2014 second year students, and they were lectured in the traditional manner in the first and second semester. We also compare the differences in the results of the course evaluation (which used a Likert scale) of the first semester to the results of the course evaluation of the second semester for the two groups of students. Finally, we look at general feedback (free format self-reporting measures) from students about their experience in being lectured in the traditional manner versus being lectured with multimedia.

Investment-Policy Surrender Prediction with Random Survival Forests

Presenter: Peter Smith, Department of Statistics, University of Pretoria

Co-author(s): Kanfer, F (Department of Statistics, University of Pretoria) and Millard, S (Department of Statistics, University of Pretoria)

In this article we introduce and discuss Random Survival Forests, a modern ensemble method for predicting right-censored survival data, and present an original application of the model in the prediction of surrenders of investment policies. The model's performance is benchmarked against the Cox model - a semi-parametric model that has been the mainstay of survival analysis since its introduction in the early 70s. Predictive performance is measured via an adaptation of the Brier Score for right-censored data using what is known as Inverse Probability of Censoring Weights. In this application the Random Survival Forest is shown to have superior predictive performance to the Cox model.

Using Multiple Group Multilevel Latent Models for Cross-Country Comparisons

Presenter: Agnes Stancel-Piątak, IEA Data Processing and Research Center

The presentation provides an empirical application of Multilevel Structural Equation Modeling with Large Scale Assessment data using an example from educational effectiveness research. Extending MSEM to multiple group analysis (MG-MSEM) a procedure for cross-country comparisons is presented. The analysis considers topics related to the data design of complex samples, for instance weighting and plausible values. Limitations of the method are discussed together with methodological issues, such as inferring causality, validation of latent constructs, linear vs. categorical approaches.

A logarithmic logistic regression model

Presenter: Francois Steffens, University of Pretoria

In an experiment to find optimal combinations of indigenous plant material and certain essential oils for inhibition of a number of bacteria, the design in terms of dosages was, for practical reasons, a logistic design. In terms of log(dosage) the design was a regular factorial design. The response variable was binary (inhibition or not) and thus a logistic regression model was indicated. The choice of a logistic regression model in term of log(dosage) leads to a logistic response model that is not symmetric in the dosage space. The two models (logistic regression and logarithmic logistic regression) are compared and the advantaged and disadvantages of a logarithmic regular grid are discussed.

Modelling branch-level data in MG SEM

Presenter: Arien Strasheim, Department of Human Resource Management, University of Pretoria, Faculty of Economic & Management Sciences

Co-author(s): Kriel, G (Department of Human Resource Managemen, University of Pretoria, Faculty of Economic & Management Sciencest)

This study will use Multiple Group Structural Equation Modelling (MG SEM) to investigate the moderating role of cultural group in a set of attitudinal and behavioural variables within a banking environment. The role of leadership, rolefit and role satisfaction as antecedents of affective commitment is investigated. The findings of ignoring the nested nature of the data will be compared to when the multi-level nature of the data is incorporated in the model, using MPlus cluster modelling.

Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals

Presenter: Jan Swanepoel, North-West University, Potchefstroom

Bernstein estimators attracted considerable attention as smooth nonparametric estimators for distribution functions, densities, copulas and copula densities. In this talk we present a parallel result for the first order derivative of a copula function. We discuss how this result leads to Bernstein estimators for a conditional distribution function and its important functionals, such as the regression and quantile functions. Results of independent interest, such as an almost sure oscillation behavior of the empirical copula process and a Bahadur type almost sure asymptotic representation for the Bernstein estimator of a regression quantile function, are also presented. The outcome of a simulation study demonstrates the good performance of the proposed estimators.

Which Threshold Concepts exist in First Year Statistics courses at the University of Pretoria?

Presenter: Andre Swanepoel, Department of Statistics, University of Pretoria

Co-author(s): Engelbrecht,J (Department of Science, Mathematics and Technology Education, University of Pretoria);Harding,A (Department of Mathematics and Applied Mathematics, University of Pretoria)and Fletcher,L (Department of Statistics, University of Pretoria)

In the teaching of Statistics, certain concepts are experienced as more difficult to comprehend than others. Misconception of such concepts while studying Statistics on the 100 level is problematic since it might prohibit the student from understanding and grasping the core concepts upon which the discipline is developed and will also influence the student`s future studies of the discipline since no proper holistic view of the inner mechanics of the different procedures and techniques nor the interrelatedness of the different procedures and techniques will be present. These concepts are referred to as threshold concepts where a threshold concept is a conceptual gateway that opens up a new and previously inaccessible way of thinking without which you cannot progress in the subject.

The purpose of this research is to identify the threshold concepts in 100 level Statistics at the University of Pretoria in a three year longitudinal study and to also determine their levels of difficulty (which describes how troublesome the concept is to master) and importance (which refers to how

much follow up work is unlocked by mastering the concept). A better understanding of the threshold concepts within Statistics can give insight on difficulties perceived by students which can be indicative to whether education models should be adapted.

The results discussed will be preliminary, based on data gathered in 2014 for the 100 level students of 2013 on whom the longitudinal study will be based. Additional data for the 2011 and 2012 first year students will also be analysed.

Recent Work in Twenty20 Cricket Analytics

Presenter: Tim Swartz, Simon Fraser University, Burnaby BC, Canada

This presentation considers a number of applied problems in Twenty20 cricket. The work is based on the development of a match simulator which takes various factors into account including the batsman, the bowler, the over, the number of wickets, the innings and the target. The parameters used in the simulator are estimated from detailed ball-by-ball data which was obtained through the parsing of match commentaries. With the simulator, we investigate player evaluation, the optimal team lineups and quantify the impact due to fielding.

On Model Selection Criteria in Statistical Neural Network

Presenter: Christopher Udomboso, Department of Statistics, University of Ibadan, Ibadan, Nigeria

Co-author(s): Dr Chukwu, A U (Department of Statistics, University of Ibadan, Ibadan, Nigeria) and Prof Dontwi I K (Department of Mathematical Sciences, Nkwame Nkrumah University of Science and Technology, Kumasi, Ghana)

In any statistical analysis, selection of the best model has been a challenge for a very long time. Many information criteria have been proposed by several authors. Some criteria had been from the viewpoint of both the frequentist and Bayesian. In recent years we have had some also from the viewpoint of artificial intelligence (AI). This paper considers a typical analysis from the statistical neural network (SNN), a branch of AI, and examines a number of selection criteria in determining the best SNN model at different sample sizes and number of hidden neurons. The criteria used include the mean square error, Akaike information criterion, Schwarz information criterion, network information criterion, and adjusted network information criterion. Generally, the values of the criteria increases with increase in sample size, and decreases with increase in number of hidden neuron.

The histogram and polygon revisited

Presenter: Danie Uys, Stellenbosch University

Summarised or grouped data from a frequency table are graphically represented by a histogram. A polygon, consisting of connected line segments, is constructed in addition to the histogram. The coordinates of the polygon are chosen as the midpoint and height of the rectangular block of the corresponding class interval, respectively. An alternative method to determine the coordinates of the polygon, is proposed. Percentiles calculated from this adjusted polygon yield favourable results when compared to population percentiles.

Comparison of old and new fit tests for peaks over a known threshold

Presenter: Sean van der Merwe, University of the Free State

Co-author(s): Ntseki, J (Department of Mathematical Statistics and Actuarial Science, University of the Free State) and Teise, C (Department of Mathematical Statistics and Actuarial Science, University of the Free State)

We do a direct comparison of existing tests for the Generalised Pareto Distribution with a known threshold. In 2001 Choulakian and Stevens explained goodness-of-fit testing for the GPD. Since then many new developments have occurred that could improve testing, but no new direct comparison has been performed. We incorporate both new testing approaches (Villaseñor-Alva and González-Estrada, 2009) and new parameter estimation approaches (Zhang, 2010) to determine under what circumstances they could offer improved accuracy and power.

A Bayesian Control Chart for a One-sided Upper Tolerance Limit for the Normal Population

Presenter: Abrie J van der Merwe, University of the Free State

Co-author(s): van Zyl, R (Biostatistics, Quintiles) and Groenewald P.C.N (Department of Mathematical Statistics and Actuarial Sciences, University of the Free State)

A confidence interval for a quintile is called a tolerance interval. By using air-lead data analysed by Krishnamoorthy and Mathew (2009) a Bayesian procedure is applied to obtain control limits for the upper one-sided tolerance limit. Reference and probability matching priors are derived for the pth quantile of a normal distribution. By simulating the predictive density of a future upper one-sided tolerance limit, “run-lengths” and average “run-lengths” are derived. This talk illustrates the flexibility and unique features of the Bayesian simulation method for obtaining the posterior predictive distribution and control limit of a future one-sided tolerance limit.

Bayesian estimation under the Matrix variate elliptical model

Presenter: Janet van Niekerk, Department of Statistics, Faculty of Natural and Agricultural Sciences,University of Pretoria, Pretoria, South Africa

Co-author(s): A. Bekker*, M. Arashi*' and D.J. de Waal*''

*Department of Statistics, Faculty of Natural and Agricultural Sciences,University of Pretoria, Pretoria, South Africa

'Department of Statistics, School of Mathematical Sciences, University of Shahrood, Shahroo

The problem of estimation within the matrix variate elliptical model is addressed. In this paper a subjective Bayesian approach is followed to derive new estimators for the parameters of the matrix variate elliptical model by assuming the previously intractable normal-Wishart prior. These new estimators are compared to the estimators derived under a normal-inverse Wishart prior as well as the objective Jeffreys' prior which results in the maximum likelihood estimators, using different measures. A valuable contribution is the development of algorithms for the simulation of the posterior distributions of the matrix variate parameters with emphasis on the new proposed

estimators. A simulation study as well as Fisher's Iris data set are used to illustrate the novelty of these new estimators and to investigate the accuracy gained by assuming the normal-Wishart prior.

The quantile statistical universe

Presenter: Paul J. van Staden, Department of Statistics, University of Pretoria

Co-author(s): King, R.A.R. (School of Mathematical and Physical Sciences, University of Newcastle, Australia)

Quantile-based approaches for the construction of generalized families of statistical distributions have become increasing popular in recent years. This paper discusses and compares two of these approaches. Both approaches yield quantile-based families highly flexible in distributional shape. With the first approach, the behavior of each tail of the obtained generalized distribution is uniquely modeled. The second approach generates quantile-based distributions with skewness-invariant measures of kurtosis. Consequently the skewness and kurtosis of these distributions can be identified and analyzed separately.

An Improved unbiased-Bayesian estimation of the Extreme value index for heavy-tailed distributions

Presenter: Andréhette Verster, University of the Free State

Co-author(s): Maribe, G (Department of Mathematical Statistics and Actuarial Science, University of the Free State)

The extended Pareto distribution (EPD) can be used to model excesses above a threshold for distributions in the Frechét domain. The EPD allows for lower thresholds and can thus be fitted to a larger portion of the data, resulting in bias reduced estimates of the Extreme Value index (EVI). Till now the parameters of the EPD was estimated by finding the pseudo-maximum likelihood estimators analytically, although this method shows the improvement of the EPD over the Generalized Pareto distribution (GPD), we further show that by using Bayesian methods as an alternative to estimating parameters of the EPD, we can –to some degree, reduce the bias and improve stability of the EPD, thus resulting in more stable estimates of the EVI. We assess the performance of our estimates by conducting a small scale simulation experiment and a case-study using a real dataset.

New Challenges in Clustering and Dimensional Reduction in the Era of Big Data

Presenter: Maurizio Vichi, Universita di Roma Sapienza

Big Data frequently describe complex economic, social and demographic phenomena that manifest on individuals (units, objects, sites, with a spatial location), by means of a set of variables that show both a diffusion over space and an evolution over time. These data show different relations between, objects (spatial correlation), between variables (cross-sectional correlation) and between times (time series correlation) that need to be analysed. Three or high dimensional arrays (data (iper)-cubes), are used to rearrange the huge number of statistical units (rows) with a spatial location, variables, (columns) and times (tubes). A modelling approach for simultaneous clustering and hierarchical disjoint factorial models is proposed to analyse these data. Special attention is given to two-way data models of clustering and dimensional reduction.

A generalisation of the mean correcting martingale measure

Presenter: Jaco Visagie, North-West University

The mean correcting martingale measure is often used in the calculation of option prices under geometric Lévy models. I propose a generalisation of this measure that can be used to obtain a family of probability measures.

It is well-known that, if a measure change results in a locally equivalent martingale measure, then this measure can be used for the calculation of arbitrage free option prices. The generalised mean correcting martingale measure does not, in general, satisfy the requirement of local equivalence. However, I demonstrate that this measure can be used to calculate an arbitrage free price for a European option under certain geometric Lévy option pricing models.

An objective comparison between various goodness-of-fit tests for exponentiality

Presenter: Leonard Santana, North-West University

Co-author(s): Allison, JS (Department of Statistics, North-West University), Visagie, J (Department of Statistics, North-West University), Smit, N (Department of Statistics, North-West University)

The exponential distribution is a popular model both in practice and in theoretical work. As a result, a multitude of tests have been developed for testing the hypothesis that observed data are realised from this distribution. Many of the recently developed tests contain a tuning parameter, usually appearing in a weight function. These tests are often evaluated over a grid of values for this parameter. However, this method does not lend itself to objective comparisons because the power of the test is highly dependent on the value of the tuning parameter. In this paper we compare the performance of tests that contain a data-dependent choice of the tuning parameter to other classical tests (which do not contain a tuning parameter). It is found that the tests based on the data-dependent choice of the tuning parameter compare favourably to the remaining tests.

Investigating the posterior predictive p-value for model evaluation in sequential regression multiple imputation (SRMI)

Presenter: Michael von Maltitz, University of the Free State

Co-author(s): van der Merwe, AJ (Department of Mathematical Statistics and Actuarial Science, University of the Free State)

In incomplete data analysis, often the diagnostics associated with a study are based on the fit of the overarching analysis model, rather than on the particular imputation model fits. For example, researchers examine the RBIAS and RRMSE of a mean or parameters of a regression model after imposing missingness and after multiple imputation (compared to the true mean or parameter estimates before imposing missingness). However, Cabras, Castellanos and Quirós (2011) have extended an idea by Hjort, Dahl and Steinbakk (2006), to post-process posterior predictive p-(ppp)-values, which are not usually uniform under the null, in order to have a uniform test statistic

to test the Normality of incomplete data during the SRMI process. Cabras et al.'s (2011) methodology is critically reviewed in this paper, and adaptations for assessing the Normality assumption in sequential regression multiple imputation (SRMI) for continuous incomplete data are proposed and tested.

Data to Insight: Prototyping next-generation introductory statistics

Presenter: Chris Wild, University of Auckland, New Zealand

"Data to Insight" is an introduction to statistical data analysis MOOC (massive online open course, but not so massive) first taught on the UK's FutureLearn platform late last year and currently running again. Course completers last year ranged from members of a small high-school physics-honours class to PhD researchers from many areas, and from journalists, linguists and arts administrators to economists, data managers, marketers and scientists. It was pleasing to see a large contingent of high-school mathematics teachers, less so to see so many "data analysts". Completers were extremely enthusiastic about the course.

In addition to being a MOOC introducing its students to statistical data analysis, "Data to Insight" prototyped a much-further-much-faster, more-data-more-quickly Introductory Statistics Course. The most novel acceleration strategies used were: being intensely visual and driving all argument off things you can see supplemented by metaphor; building software solutions (including powerful online visualisation and analysis software) that prevent "how do I get this out of the software?" limiting the speed at which students can encounter new situations and new ideas; and finding some powerful, conceptually-undemanding "extender-capabilities" that immediately open much wider horizons. We will speak about the course, the software, the MOOC environment and educational lessons learned from working in a MOOC environment for both online and classroom teaching. We will highlight some lessons learned from making the course videos with a team of professionals, from the use of online quizzes and from online communication and discussion with and between students.

Estimation of the Modified Traffic intensity of a Markovian Queuing system with Balking

Presenter: Venkata S S Yadavalli, University of Pretoria, Pretoria, South Africa

Co-author(s): Vaidyanathan S Vaidyanathan, Pondicherry University, Puducherry, India; Pichika Chandrasekhar, Loyola College, Chennai, India

By considering a Markovian queueing model with balking, the maximum likelihood and consistent estimators of modified traffic intensity are obtained based on the number of entities present at several sampled time points. Uniform minimum variance unbiased estimator (UMVUE), consistent asymptotically normal (CAN) estimator and an asymptotic confidence interval for the expected number of entities in the system are obtained. Further, Bayes estimators of modified traffic intensity, measures of system performance, minimum posterior risk and minimum Bayes risk associated with these estimators are also derived. The behavior of maximum likelihood and Bayes estimator of modified traffic intensity is illustrated through simulation study.

On weighted Gaussian entropy

Presenter: Salimeh Yasaei Sekeh, Federal University of Sao Carlos (UFSCar), SP, Brazil

Co-author(s): Suhov, Y (DPMMS, University of Cambridge, UK and Math Dept, Penn State University, PA, USA) and Stuhl, I (IMS, University of Sao Paulo, SP, Brazil and Math Dept, University of Denver, CO, USA)

We produce a series of results extending information-theoretical inequalities (discussed by DemboCoverThomas in 1989-1991) involving the Gaussian weighted entropy, they imply a number of new relations for determinants of positive-definite matrices. Furthermore, the standard relative entropy with the weighted case in particular form of Gaussian distributions is compared.

A Potential Outcomes Approach to Documenting the Public Health Impact of the Introduction of PCV13 for the Prevention of Invasive Pneumococcal Disease

Presenter: Elizabeth Zell, Stat-Epi Associates Inc.; CDC (retired)

In March 2010, a new vaccine, PCV13, was introduced for children under five years of age for the prevention of invasive pneumococcal disease (IPD) caused by the bacterium Streptococcus pneumoniae. We used a potential outcomes approach to estimate the cases of IPD prevented in children less than five years of age and the number of cases prevented in adults 65 years and older after PCV13 vaccine introduction. With data from an active, population-based surveillance system, we modeled the monthly IPD case counts for children less than five years of age and for adults 65 years and older between July 2004 and March 2010 to generate a set of parameter estimates and their variance-covariance matrices for the time trend during this period. We then imputed predicted monthly case counts occurring between July 2010 and June 2012 assuming no vaccine had been introduced. To estimate the number of cases prevented by the introduction of PCV13, we compared the predicted cases of IPD in the absence of PCV13 to the actual number of cases observed after PCV13 introduction for both children and adults. This approach to estimating the public health impact of PCV13 will be used to inform national vaccination policy.

Student Attitudes Towards Statistics

Presenter: Nombuso Zondo, UKZN

This study investigates the relationship between student attitudes towards Statistics and their performance in the Statistics course. We adopted the 'SATS-36' survey questionnaire to assess the attitudes of students towards Statistics. We used exploratory factor analysis to group the attitude responses according to factor loadings as was done in other studies using 'SATS-36'. Moreover, we examined whether the attitudes to Statistics locally are related to demographic attributes, field of employment and academic exposure to Statistics.

Statistical Methodology and Techniques Sessions Abstracts

(In Alphabetical Order)

Dynamic spatio-temporal analysis of Ebola virus disease: putting in perspective epidemics in Africa

Presenter: Adewale Adeogun, North-West University

Co-author(s): Palamuleni, M. (Department of Population Studies, North-West University), Palamuleni, L. (School of Environmental & Health Science, North-West University)

Africa is endemic to Ebola virus disease (EVD). The virus discovered three decades ago in DR Congo has resulted in thirteen epidemics on the continent with high human fatalities. Global effort is yet to discover bio-medical solutions hence the prominence of epidemiology remedies. The need for proactive measures against future outbreaks motivated this study, aimed at application of data science to better understand the pattern of EVD epidemics in African countries. In the process the dynamic spatio-temporal analysis tool was used to explore relationship between spatial movements of the disease in time domain. The methodology included a perspective review of past epidemics on a continental scale, and the application of stochastic principles in geostatistics combined with graphic applications. These were used to study the severity of EVD on human populations taking reported human cases (RHCs) as a proxy measure. Variogram and kriging analyses produced interpolated patterns for other African countries beyond known epidemic locations. The result showed that no part of the continent is immune to future EVD outbreaks. Weak healthcare systems, cultural practices and international border commuting are potent means of transmission of the disease from areas more endemic to other parts of the continent rather than risk of infections from primates or fruit-bats. African governments, especially in countries yet to experience EVD outbreak are advised to put in place proactive measures that align with global response mechanisms against future epidemics, while global research efforts should be fast-tracked for a vaccine that will ultimately contain the disease.

Childhood mortality spatial distribution in Ethiopia

Presenter: Dawit Ayele, University of KwaZulu-Natal

Co-author(s): Temesgen T. Zewotir, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal

The risk of a child dying before completing five years of age is highest in Sub-Saharan African countries. But Child mortality rates have shown the substantial decline in Ethiopia. For this study, the 2000, 2005 and 2011 Ethiopian Demographic Survey (EDHS) was used. Generalized linear mixed model with spatial covariance structure was adapted. The model allowed for spatial correlation, and leads to the more realistic estimate for under-five mortality risk factors. The analysis showed that the risk of under-five mortality shows decline in years. But, some regions showed increase in years. The study highlight the need to implement better education for family planning and child care to improve the under-five mortality situation in some administrative areas.

Using Extreme Value Theory To Measure Value-At-Risk For Daily South African Mining Index

Presenter: Retius Chifurira, University of KwaZulu-Natal

Co-author(s): Chinhamu, K(School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal)

Financial data usually possess some characteristics, such as volatility clustering, asymmetry, heavy and semi-heavy tails thus, making it difficult, if not impossible, to use Normal distributions to model them. As such, we need to use other kind of distributions which can capture these properties. Statistical analyses show that the Generalised hyperbolic distribution is more appropriate for financial returns estimations. However, we extend our analysis to four dimensional returns. Research shows that multivariate affinely transformed versions of this multivariate generalised hyperbolic distribution present more interesting features than the original distribution. In this regard, we investigate the fit of the multivariate generalised hyperbolic distribution as well as the multivariate affine generalised hyperbolic distributions to four financial indices from the Johannesburg Stock Exchange. Based on the kernel smoothing goodness of fit, the multivariate affine normal inverse gaussian distribution provides the best fit for the affine models. On the other hand, the multivariate generalised hyperbolic distribution based on AIC provides the best model for the four returns without any form of affine transformation on the returns. Finally, the positive tail dependencies exhibited between the all share and Gold mining index as well as all share and S&P 500 is best modelled with the Gumbel and Clayton copulas respectively. While the negative dependencies between the other pairwise returns is modelled with the Frank copula.

Statistical Models to Model the Probability of the Under-five Mortality in United Republic of Tanzania

Presenter: Welcome Dlamini, University of KwaZulu-Natal

Children are the economic asset of the world and their future development can be affected by factors associated with under-five mortality. The well-being of a child reflects household, community and national involvement on family health. This will have an immense contribution towards the development of a country. Globally, a substantial progress in improving child survival since 1990 has been made. The decline globally in under-five mortality from approximately 12.7 million in 1990 to approximately 6.3 million in 2013 had been observed. However, all regions except Sub-Saharan Africa, Central Asia, Southern Asia and Oceania had reduced the rate by 52% or more in 2013. This study aims to identify factors thats are associated with the under-five mortality in Tanzania. In order to robustly identify these factors, the study utilized different statistical models that accommodate a response which is dichotomous. Models studied include ordinary logistic regression, survey logistic regression, generalized linear mixed model(GLMM) and generalized additive model(GAM). The results show that HIV status of the mother is associated with the under-five mortality. Furthermore, the results shows that mothers age, child birth order, breastfeeding and total number of children alive affects the survival status of the child. This shows that there is a need to intensify child health interventions to reduce the under-five mortality rate and to be inline with the millennium development goal 4(MDG4).

Comparison of methods for long-term forecasting of electricity load profiles in South Africa

Presenter: Jenny Holloway, CSIR

Co-author(s): Koen, R (CSIR) and Mokilane, P (CSIR)

This paper describes the application of three different statistical approaches to the problem of obtaining long-term (20 – 30 years ahead) forecasts of national electricity load profiles for South Africa. These methodologies include: a combination of multilevel modelling and symbolic regression; univariate structural time series models, with separate models fitted to each hour of the day; and ARIMA modelling, which is used as a benchmark for comparison. Particular focus is placed, in this paper, on the suitability of fitting univariate structural time series models and whether this approach could adequately capture the hourly fluctuations evident in the South African electricity load profiles. The accuracy of the forecasts obtained from all three approaches is evaluated and compared using the MAPE for a full year of out-of-sample data. Furthermore, because of the complexity of the patterns within the annual load profile, MAPE values are also compared over periods of the day or year that are of particular importance with respect to the long-term planning of electricity supply requirements.

A Case-Control Study of Tattoo and HIV Infection among Teens in Mozambique

Presenter: Adelino Juga, Eduardo Mondlane University/Uhasselt University

Co-author(s): Niel Hens(Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium) and (Centre for Health Economic Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Antwerp,Belgium), Nafissa Osman(Department of Obstetrics and Gynaecology, Maputo Central Hospital, Maputo, Mozambique) and (Faculty of Medicine, Eduardo Mondlane University, Maputo, Mozambique), Marc Aerts(Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium)

A tattoo is an ink design inserted into the skin, meaning "to strike or mark“. People receive tattoos to: identify themselves with a religious or social group, adorn their bodies, as protective symbols, to cover skin discolorations, etc. Transmission of HIV attributed to tattooing has been suggested and is theoretically possible. In this paper, association between tattoo and HIV transmission among teens in Mozambique was investigated. Cross-sectional data based on national-representative sample of INSIDA survey in Mozambique was used. Several statistical models such as Logistic Regression, Generalized Estimating Equations(GEE), Alternating Logistic Regression(ALR) and Generalized Linear Models Mixed(GLMM) were random effects coming from a conjugate exponential-family distribution (Gaussian, Gamma) were applied motivated by the nature of outcome and by the design of the study. Statistical findings revealed that there is strong association between tattoo and infection of HIV among teens, and this varies from one enumeration area to another

---

Homeownership differentials in South Africa

Presenter: Mmanate Kekana, Statistics South Africa

Co-author(s): Naidoo, A (Statistics South Africa)

Household wealth and income have significant importance to the transition to homeownership. Recent studies argue that homeownership attainment should not only consider individual or household attributes but should also consider spatial location, as the findings show a strong association between spatial location satisfaction and the individual demand for homeownership .This paper focus on the socio economic factors of homeownership in South Africa using Census 2011 data at subplace level. The variables of interest are median household income, age group, employment status, number of workers in a household, gender, family composition and education level which are the independent variables and homeownership as the dependent variable. Principal component analysis(PCA) is used as a global measure to group the variables of interest and then geographically weighted principal component analysis (GWPCA) is applied to the data as a local measure .The factor loadings for each variable are mapped to show the spatial variation in the relative importance of each variable in the component and also to show which variables dominant in locations .The global PCA reveals that three components have eigenvalues over 1, and that they account for about 70 % of the variation in the data. The results of a GWPCA with an adaptive kernel, with 21589 observations suggest that there is considerable variation in social structure. The range of eigenvalues is -2.3 to 3.9 suggesting that locally, more variation is accounted for by the first component that is the case with the global analysis. The map for first component also suggests that areas with the highest local eigenvalues are across all provinces, while for the second component shows areas with the highest local eigenvalues are in Limpopo province.

Assessing the effect of distance from a dam on time to malaria, with distance confounded with the clustering structure.

Presenter: Yehenew Getachew Kifle, Department of Statistics & Operations Research, University of Limpopo, South Africa

Co-author(s): Delenasaw Yewhalaw, Delenasaw (Department of Biology, College of Natural sciences, Jimma University, Jimma Ethiopia); Niko Speybroeck (Institute of Health and Society, Universit´e Catholique de Louvain, Brussels, Belgium); Paul Janssen (CenStat, Hasselt University, Belgium) and Luc Duchateau (Department of Comparative Physiology and Biometrics, Ghent University, Belgium).

Malaria remains an important disease in terms of morbidity and mortality in many developing countries. Around hydro-electric dams, this risk might even increase due to the large water bodies available to the Anopheles mosquito which functions as a vector for the disease. During two years, time to malaria was followed up on a weekly basis around one of the largest hydro-electric dams in Ethiopia, the Gilgel Gibe dam. In each of 2082 households, one child younger than 10 years old was chosen for follow-up. The households are located at different distances from the dam clustered into 16 villages.

Different standard techniques in survival analysis exist to model such clustered survival data, among them the marginal model, the fixed effects model, the stratified model and the frailty model. These time to malaria data have certain characteristics that makes the marginal and conditional

approaches lead to quite diverse effects. Although the interpretation of parameters is different in these two approaches, i.e., a population versus a conditional interpretation, in most real life datasets the two approaches lead to similar parameter estimates. The observed differences in our particular setting are due to the fact that the covariate of interest in the dataset, distance from the dam, is highly confounded with the clustering process, i.e., the village.

Different models that cope with clustering in survival data can lead to contradictory results when the covariate of interest is confounded to a large extent with the clustering mechanism. The marginal model leads to quite different results compared to the other models, especially if the within village distance effect differs from the between village distance effect. In the marginal model, the overall effect of distance is studied, whereas in the fixed and stratified model, rather the within village effect of distance is investigated. The frailty model somehow combines these two approaches, but the way these two estimates are combined depends on factors that are hidden for the data analyst.

The frailty model is often considered the standard model for clustered survival data. In a certain sense, it is the most efficient model under certain assumptions, in that it has the smallest standard error. This increase in efficiency, as compared to the fixed effects model, is obtained by the so-called recovery of between blocks information. The frailty model estimate is a weighted combination of the within and between village estimate of the distance effect. Such a weighted combination, however, makes only sense if the same relationship holds between and within clusters (blocks), i.e., village. This assumption, however, is questionable for the type of dataset that is considered in this study. Therefore, in such situation, we advise to split covariates into two orthogonal covariates, one referring to the covariate effect between clusters, and another referring to the covariate effect within clusters.

A semi-parametric method for generating time series data: an approach for bootstrapping the residuals

Presenter: Cliff Richard Kikawa, Tshwane University of Technology

Co-author: Kloppers, PH (Tshwane University of Technology)

are added to the independent variables as defined in the two models. The idea in the proposed method is to “let data speak for themselves”.

Results and Discussion: Statistical accuracy measures employed on both data sets showed that the proposed method generates bootstrap samples that are more accurate than those from the Efron-Tibishirani method. The RMSE was preferred for the luteinizing hormone levels, since its residuals exhibited a normal distribution. The MAE and the SE was used for the mean monthly levels as the residuals showed a significantly different distribution from normal.

Conclusions: The proposed data generating process produces better bootstrap samples than the Efron-Tibishirani approach. Hence, it is recommended for application in both theoretical and practical problems.

Mixtures of generalized lambda distributions

Presenter: Robert King, Department of Statistics, University of Pretoria; School of Mathematical and Physical Sciences, University of Newcastle, Australia

Co-author(s): van Staden, P (Department of Statistics, University of Pretoria)

Mixtures of distributions have frequently been used to fit univariate data with complex shapes, often without aiming to interpret the components of the mixture. Where users desire interpretable components and the components are skewed and/or have heavy tails, a mixture of flexibly shaped distributions is useful. Here we present a method for a mixture of an arbitrary number of generalized lambda components.

Assessing The Levels Of Secondary School Dropouts In Relation To Some Socio-Economic Factors: A Case Study Of Khonjeni.

Presenter: Fiskani Kondowe, University of Malawi

Co-author(s): Mwakilama, E (Department of Mathematical Sciences, University of Malawi-Chancellor College)

By the year 2009, school drop-out rates were reported high in Sub-Saharan African countries (42%), compared to South and West Asia (33%) and America (17%), with Malawi contributing an overall rate of 24 %. However, so far, little has been done to assess the factors influencing secondary school dropouts. Identifying factors that facilitate frequent secondary school drop-outs creates a starting point to reduce the rates. As such, this study aimed at assessing factors that influence school drop outs in Khonjeni area by identifying any influential socio-economic factors.

This was a cross-sectional study of 100 purposively sampled respondents in capturing primary data while secondary data from selected schools was used. Primary data on socio-demographic characteristics was obtained through questionnaires and interviews while secondary data on dropout occurrences came from the selected schools’ records. Both descriptive and inferential statistical analyses were carried out in SPSS.

Logistic regression and Chi-square analysis revealed that: type of school (p-value=0.022), age (p-value=0.028), gender (p-value=0.014, CI=0.095-0.694), school’s condition and location (p-value=0.006, OR=5.59, CI=1.639-19.069) contribute to secondary school dropouts. Corporal punishments (p-value=0.011) and early marriages (p-value=0.015) are associated with drop-outs. Higher drop-outs were also observed in private schools than public schools (OR=4.323).

Higher figures of drop outs were found among early married pupils and those living far from school locations. Strict policies that regulate school’s conditions and location must therefore be enforced. Further, activities that attract students at schools should be introduced. Further research on higher drop outs in private schools than in public is also recommended.

Fisher Optimal Scores for Visualisation in Categorical Data

Presenter: Sugnet Lubbe, University of Cape Town

Co-author(s): le Roux, NJ (Department of Statistics and Actuarial Science, Stellenbosch University) and Gower, JC (Department of Mathematics and Statistics, The Open University UK)

Multiple Correspondence Analysis (MCA) is well known for visualisation of categorical data. Although already introduced in 1938 by Fisher, the methodology of Fisher Optimal Scores (FOS) is less well known in the statistics literature. In this paper the FOS methodology and underlying matrix algebra will be revisited. With modern computing, extensions of the FOS methodology are possible, which will be discussed and illustrated.

MCA treats all the categorical variables in a similar manner while FOS distinguishes between independent and dependent variables. An overview will be given of the similarities and differences between FOS, MCA and other categorical visualisation methods such as homogeneity analysis and Guttman scores. FOS can also be viewed as the categorical equivalent of biadditive models. We will illustrate how FOS can be extended into the orbit of biadditive (multiplicative) models enabling the construction of biplots to simultaneously visualise main effects and interactions resulting from observing a categorical dependent variable.

More thoughts on the EM algorithm

Presenter: Iain MacDonald, Univ of Cape Town

I have written previously on the somewhat surprising phenomenon that one can easily find published applications of EM that seem unnecessary: unnecessary in the sense that there are available mathematically or computationally simpler methods to

solve the relevant problems. I now add several further examples of this phenomenon. These include the use of EM for penalized maximum-likelihood estimation in a model for survival data, the fitting of a Poisson distribution to zero-truncated counts, the fitting of a generalized gamma distribution to left-truncated, right-censored survival data, and a constrained estimation problem in genetics. I offer some speculations on the reasons for the phenomenon. I end with a brief discussion of recent developments in optimization as applied in Statistics.

Likelihood inference based on EM algorithm for the destructive COM-Poisson cure rate model

Presenter: Jacob Majakwara, Wits university

Co-author(s): Suvra P (Department of Mathematics, University of Texas at Arlington, Texas, USA)

In this talk, we will discuss the destructive COM-Poisson cure rate model that presents a realistic and interesting interpretation of the biological mechanism for the recurrence of tumour in a competing causes scenario. The model assumes the event of interest to undergo a destructive process of the initial risk factors and what is recorded is the undamaged portion of the original number of risk factors. An algorithm is developed for computing estimates for this model's parameters assuming lifetime to follow Weibull distribution and censoring mechanism to be non-informative. The performance of the method of inference developed is examined using a simulation study and a real data set.

Sample design to optimise the estimation of small micro and medium enterprise owners and their characteristics

Presenter: Thanyani Maremba, Statistics South Africa

Long-term trends in living alone among South African adults: Age, gender, and educational differences

Presenter: Thabo Masemola, Statistics South Africa

One of the major demographic changes in South Africa is a rapid increase of one-person households, from 16% in 1996 to 28% in 2011. The increase of one-person households has important implications on the traditional family system. Census 1996, 2001 and 2011 data was used to investigate the long-term trend of the proportion of people living alone, for age and gender differentials.

The paper focuses on two groups, the widowed elderly aged 65 or over and never-married 25 to 34-year-olds. Furthermore, the relationship between education and living alone was examined. Logistic regression was used to predict the odds of living alone by education. Multivariate linear regression analysis was used to examine the factors that related to living alone. Geospatial analysis techniques were also performed to show the spatial variation of one-person households.

The results show that, there is a continuing increase in solo living among South Africans. The rising trend in solo living among elderly widows and never-married men aged 25 to 34, in particular, reveals that the propensity for living alone has increased within specific marital status and age groups. We find that those with higher education were more likely to live alone than their counterparts with lower education. The results also showed that there is a positive relationship (R > 0) between level of education and likelihood of living alone in South Africa. In conclusion, the study reveals that there is a general increase in the proportion of one-person households in South Africa. The one person households are highly correlated to the level of education of the head of households.

ASSESSING FACTORS AFFECTING ADMISSION TIME OF KAPOSI SARCOMA USING SURVIVAL ANALYSIS, A CASE OF ZOMBA CENTRAL HOSPITAL MALAWI

Presenter: Henry Mlinde, University of Malawi

Co-author(s): Simbeye, J (Department of Mathematics, Chancellor College, University Of Malawi) and Mwakilama, E (Department of Mathematics, Chancellor College, University Of Malawi)

Kaposi sarcoma is most HIV-related malignancy worldwide and the most frequently diagnosed cancer in several Africa countries. In Malawi KS accounts up to 34 % of all diagnosed cancer cases. This study aims at assessing factors affecting admission time of Kaposi sarcoma patients using survival analysis.

The study involved patient characteristics that included CD4 count, admission and discharge diagnostics and admission time. Data was collected from patients' hospital registers at Zomba

Central Hospital while entry and analysis was done in SPSS and STATA 12. Descriptive statistics, life table, Kaplan-Meier and Chi-Square were used to measure relationship between patient characteristics and admission time. Cox proportional hazards model was fitted to assess factors that affect length of admission time. Weibull model, Exponential model and Frailty model were fitted to its performance while model diagnosis was done using Schoenfeld residuals, Martingale residuals and Cox-snell residuals.

A total of 163 KS patients had a mean age of 38 years with range of 15 to 92 years and 78 (48%) were males and 85(52%) were females. From all KS patients, 54(33%) were infected with HIV while 21(13%) were not infected while 87(54%) were of unknown HIV status with 44 (29%) on ART. Chi-square results showed association between gender and HIV status (P-Value=0.594) and with ART status (P-Value=0.525). KP showed that females survived better than males (P-Value=0.030) and HIV infected KS patients survived well (P-Value = 0.020). Cox proportional hazards model results showed that admission time was associated with age, gender, HIV status, ART status, CD4 count of KS patients. Exponential model was found to be the best model for original data.

Results of the study indicated that admission time of KS at Zomba Central hospital is greatly affected by gender, HIV status, ART status, CD4 count and age. It is therefore required that bio-data of every KS patient should be known before treatment to reduce admission time.

Patterns of activity and employment in the young adulthood years (18-24) following their exit from the school system

Presenter: Rosina Mosoma, Statistics South Africa

Co-author(s): Naidoo, A (Statistics South Africa)

The high incidence of young people dropping out of school prior to completing secondary schooling remains a nationwide problem in South Africa. While it is commonly assumed that school-leavers will become child workers, in fact little is known about their transition to adulthood. According to Census 2011 data, 4 212 671 were young adults (aged 18–24) in South Africa are not attending school. This paper investigates their patterns of activity and employment following their exit from the school system, the timing and patterns of reaching various markers of adulthood, and their current life situations. 10% sample from Census data was used.

30% of the individuals between the ages of 18–24 were neither working nor attending schools. The likelihood of experiencing idleness was relatively increasing from the age of 18 to 24 and it was highest at age of 24. Among those with early work experience the majority worked in the manufacturing industry, as domestic servants, or as informal traders. Female school-leavers are likely to spend a longer time economically and educationally inactive during their formative years, progress faster to their markers of adulthood, and are less likely to return to school, relative to their male counterparts. Qualitative insights suggest that adolescent dropouts who enter employment early are better off in their young adulthood than those who experience inactivity prior to adulthood.

Logistic regression of being idle from school for individuals between the age 18 and 24 years (both years included) (odds ratios) was performed using STATA 14. Robust sandwich estimation was used to take into account the fact that observations are clustered within individuals. 15 Independent variables were used and only 7 (Marital Status, Employment, Parental Survival, Citizenship, Income, Level of education and Parents' level of education) were significant at $\propto$=0.05 .Arc Map was also used for Spatial analysis to check the auto-correlation as well as the hot-spot or cold-spot of the individuals aged between 18 and 24 who are not attending schools at Sub-place geographical level.

Socioeconomic Determinants and Spatial Variation of Fertility in South Africa

Presenter: Collen Motsepa, Statistics South Africa

Co-author(s): Arulsivanathan Naidoo

The level of fertility in South Africa is among the lowest in the whole of sub-Saharan Africa. South Africa was the First Country in Sub-Saharan Africa to experience decline fertility. The purpose of the paper is to identify the underlying determinants of fertility in South Africa at main place level. The paper is using 2011 Census data. 20 variables from Census data were used to find a relation on fertility using step-wise regression on SAS and Ordinary least squares (OLS) regression was used to determine predictors of fertility at the beginning. The Geographically Weighted Regression (GWR) Model was then applied to modify the traditional regression model and also to reduce the problem of spatial auto-correlation, and the results are compared with that of OLS model. Both OLS regression and GWR were conducted using Esri's ArcMap10.2. Only four independent variables out of the 20 variables were significant at a statistical significance level of 5%. The four variables are: Percentages of females who are unemployed, Percentages of females who have no income, Percentages of females with no schooling background and Percentages of females who are married. The OLS regression R^2 reached 68 percent. The Akaike's information criterion (AIC) is 22141 and Koenker (BP) statistics is significant and also the spatial pattern of the residuals shows that the residuals are exhibiting a significantly positive spatial auto-correlation. The GWR model delivered a significant improvement in the goodness-of-fit and a decline in the Akaike information Criterion (AICc).Both models suggests that there is a strong relationship between fertility and Percentages of females who are unemployed, Percentages of females who have no income, Percentages of females with no schooling background and Percentages of females who are married. However, The Geographically Weighted Regression (GWR) was the better model.

Keywords: Spatial auto-correlation, Fertility, OLS, Geographically Weighted Regression

---

Meta-analysis of Longitudinal Studies in the Presence of Missing Effect Sizes

Presenter: Alfred Musekiwa, University of KwaZulu-Natal (UKZN)

Co-author(s): Manda, S (Biostatistics Unit, South African Medical Research Council) and Mwambi, H (School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal)

Meta-analysis of longitudinal studies combines effect sizes measured at pre-determined time points. Often, a problem arises when different studies report the effect sizes at different time points. The common practice is to conveniently choose specific time points where the data are available across all the studies and performing separate meta-analysis only at these time points. However, this method ignores other reported effect sizes and does not take account of the correlation between longitudinal effect sizes within studies, which may result in the loss of power, imprecise and biased

parameter estimates. This study looks at combining all time points for longitudinal meta-analysis after undertaking missing data techniques, and compares the resulting estimates to those obtained when ignoring missing effect size data. A real practical data set is used to illustrate the application of these methods.

Modeling Gender Representation: A Case Study of the National University of Science and Technology

Presenter: Fadzayi Ndlovu, Department of Statistics and Operations Research, National University of Science and Technology

Co-author(s): Chivafa, A (Department of Statistics and Operations Research, National University of Science and Technology) and Mdlongwa, P (Department of Statistics and Operations Research, National University of Science and Technology)

In many countries, Zimbabwe included, it has been a major concern that females constitute a lesser enrollment in tertiary institutions than males, and they are also less inclined to enroll in science, technology, engineering and mathematics (STEM) fields. This research investigates and analyzes the trends in enrollment according to gender at National University of Science Technology (NUST). Descriptive statistics and chi-square tests were used to compare the proportions of students enrolled according to gender over a period of nine years (2006-2014). Logistic regression was then used to model the enrollment of students into the different STEM and Non-STEM faculties. The results of the research revealed that females still have a lower overall enrollment at the institution; and are more likely to choose a Non-STEM field of study in comparison with males.

Modelling average minimum daily temperature using extreme value theory with a time varying threshold

Presenter: Murendeni Nemukula, University Of Limpopo And University Of The Witwatersrand

Co-Author(S): Sigauke, C (Department Of Statistics, University Of Venda) And (School Of Statistics And Actuarial Science, University Of The Witwatersrand)

In this paper we present an application of the Generalized Pareto Distribution (GPD) in the modelling of average minimum daily temperature in South Africa for the period January 2000 to August 2010. A penalized cubic smoothing spline is used as a time varying threshold as well as to cater for seasonality. We then extract excesses (residuals) above the cubic spline and fit a non-parametric mixture model to get a sufficiently high threshold. The data exhibit evidence of short-range dependence and high seasonality which lead to the declustering of the excesses above the sufficiently high threshold and fit the GPD to cluster maxima. The parameters are estimated using the maximum likelihood method. The estimate of the shape parameter shows that the Weibull family of distributions is appropriate in modelling the upper tail of the distribution of average minimum daily temperature in South Africa. The bootstrap resampling method is used as an assessment tool for uncertainty in the parameter estimation. This study has shown that the use of the penalized cubic smoothing spline as a time varying threshold to time series data which exhibits strong seasonality provides a good fit of the GPD to cluster maxima. This results in accurate estimates of return levels.

Modelling Total Electricity Generation in Nigeria: The Response Surface Methodology Approach

Presenter: Oluwaseun Otekunrin, University of Ibadan, Nigeria

Co-author(s): Ariyo,O (Department of Statistics,University of Ibadan)

This study was aimed at modelling and validating total electricity generation in Nigeria using Response Surface Methodology (RSM). The factors considered were Population/Million (POP), Tariff/Naira (T), Dry Natural Gas Consumption/Billion Kwh (NGC) and Hydroelectric Power Consumption/Billion Kwh (HPC). Total Electricity Generation/Billion Kwh (TG) was the response variable. Secondary data was used and it covered a period of 1985 to 2014. The lack-of-fit of the first order model was significant at 5% level (p = 0.000002695) necessitating the need to move to the second order model. The non-significant lack-of-fit of 0.1001209 at 5% level in the second order model implied that the model was alright for prediction with multiple R^2 of 0.9842.

Using the stationary point technique, the predicted response ($\widehat{TG}$=33.5 Billion kwh) was optimized at levels 31.79 naira, 556.3348 Billion Kwh, 21.2697 Billion Kwh and 174.8507 million people for T, NGC, HPC and POP respectively. This optimum response ($\widehat{TG}$=33.5 Billion kwh) exceeded the current maximum TG in Nigeria (28.4 Billion kwh in year 2013). For model validation, actual values of TG (26.5, 28.4, and 27.78) were relatively close to their predicted values (25.7, 27.97, and 27.42) for years 2012, 2013 and 2014 respectively.

Total electricity generation in Nigeria was modeled using RSM. The developed model has good predictive ability. The optimum value obtained for TG showed that Nigeria has not reached the optimum level for total electricity generation.

Spatial variation in disability and poverty – A Case of South Africa

Presenter: Gaongalelwe Phakedi, Statistics South Africa

In many cases, disability leads to poverty because of barriers to education, economic and social participation. This paper seeks to examine if there is a relationship between disability and poverty, and to show where the poor and disabled people are situated and their socio economic characteristics.

Ten percent sample from Statistics South Africa (Stats SA) Census 2011 was used and the analysis was done using STATA. The shapefiles as produced by Stats SA was used and converted to geodatabase. Multivariate analysis was performed on the data to examine the relationship between disability and socio economic variables. Geographically Weighted Regression (GWR), a local regression technique was also applied to account for spatial variations in the data. Ordinary Least Squares (OLS) was performed using SAS Enterprise Guide to assess the global statistics of proposed model and provide baseline against which to compare the performance of local model using GWR. The two outputs were then performed and the results compared with the one of GWR. ESRI ArcGIS was used for spatial analysis, STATA and SAS Enterprise Guide for statistical analysis.

Analysis on disability and income showed a strong relationship between disability and poverty. The results on socio-economic circumstances indicate that there are associations between school attendance, level of education and disability.

Business clustering along the M1-N3-N1 corridor between Johannesburg and Pretoria, South Africa.

Presenter: Xaven Pillay, StatsSA

As a communication axis between Pretoria and Johannesburg the Old Pretoria Main Road always served as a linear force of attraction. This force was subsequently strengthened, first by the construction of the M1 and N1 motorways and later by the N3. Anecdotal evidence points to these sections of the motorways being some of the fastest-growing development corridors in South Africa. This paper analyses the spatial statistical business clustering along these sections of the corridor from 2001 to 2012 using GIS technology. Of particular importance are the economic geography forces that led to such a spatial clustering of firms and the rationale for locating along these sections of the corridor.

The research aims and objectives in this paper attempts to identify and understand the economic forces that have led to similar businesses clustering together along the N1-M1-N3 corridors of Northern Johannesburg.

The methodology of the paper includes the spatial demarcation of the business clusters along the routes. It puts forward the analysis of the area surveys, the extraction of the building footprints, the calculations of density and the spatial statistical analysis of the geographic directional distribution of the movement of non-residential data using the standard deviational ellipse.

The data sources of the paper include the use of Stats SA data, 2001 and 2012 spatial imagery and non-residential data from Geo-Terra Image.

On a new method of constructing bootstrap confidence bounds

Presenter: Charl Pretorius, Department of Statistics, North-West University, Potchefstroom Campus

Co-author(s): Prof Swanepoel, JWH (Department of Statistics, North-West University, Potchefstroom Campus)

A new method of constructing bootstrap confidence bounds will be discussed. We prove analytically, using Edgeworth and Cornish-Fisher expansions, that these bounds have smaller coverage errors than that of traditional bootstrap confidence bounds, as derived in literature. For a random sample of size~n, our new \emph{percentile} confidence bound has a coverage error of order $O(n^{-1})$, whereas a traditional percentile confidence bound has a coverage error of order $O(n^{-1/2})$. The newly proposed \emph{percentile-t} confidence bound has a coverage error of order $O(n^{-3/2})$, compared to a coverage error of order $O(n^{-1})$ of a traditional percentile-t confidence bound. The talk is concluded with an illustrative example.

A Note On Studentized Residuals in the Quantile Regression Framework

Presenter: Edmore Ranganai, University of South Africa

Regression Quantiles (RQs) are robust to residual outliers. As a consequence few residuals outlier diagnostics exist in the RQ framework. However, RQs are very susceptible to outliers in the predictor space (high leverage points) since their influence functions are bounded in the response variable but unbounded in the predictor space. Therefore in the RQ scenario a residual measure such as the studentized residual that includes leverage information is a more plausible proposal. We therefore propose a studentized residual measure for RQs and derive its distribution.

Quality of Fit Measurement in Regression Quantiles: An Elemental Set Method Approach

Presenter: Edmore Ranganai, University of South Africa

Little attention has been paid to assess the quality of fit in the quantile regression framework (Noh et al., 2013). As a contribution, we propose a coefficient of determination measure and model selection indices based on the elemental set method.

Variable selection in multi-label classification using probe variables

Presenter: Trudie Sandrock, University of Stellenbosch

Co-author(s): Steel, S (Department of Statistics and Actuarial Science, University of Stellenbosch)

Multi-label classification problems arise in scenarios where every data instance can be associated simultaneously with more than one of several available labels. Application areas include music information retrieval, bioacoustics, text and image annotation. Variable selection in a multi-label context is even more challenging than in the single label case, and additional complexity is introduced by the fact that variables which may discriminate well between values of one of the responses will not necessarily do the same for the other responses. In this regard the concepts of local and global relevance of variables are defined. A multi-label variable selection procedure should take cognisance of the possibility that some variables may not be globally relevant, but could be locally relevant for one or more labels.

We propose a multi-label variable selection method, based on a binary relevance problem transformation. Different measures of variable importance (such as correlation, information gain and reliefF) are considered as filters. Probe variables are generated by randomly permuting variable values, and these probes are used to determine the number of variables to be selected.

Empirical results obtained from applying our proposed technique as well as existing techniques (Spolaôr et al, 2013) to benchmark datasets are reported. These results show that our technique performs marginally better, and simultaneously provides output that can be used to ascertain the local and global relevance of variables.

REFERENCES:

Sandrock, T. (2013). Multi-label Feature Selection with Application to Musical Instrument Recognition. Unpublished PhD thesis. University of Stellenbosch, South Africa.

Spolaôr, N., Cherman, E.A., Monard, M.C. and Lee, H.D. (2013). A Comparison of Multi-Label Feature Selection Methods using the Problem Transformation Approach. Electronic Notes in Theoretical Computer Science, 292, 135-151.

Tuv, E., Borisov, A. and Torkkola, K. (2008). Ensemble-Based Variable Selection using Independent Probes. In Computational Methods of Feature Selection. Liu, H., and Motoda, H. (eds). Chapman & Hall/CRC.

From Bernoulli to Beethoven and Fisher to Pharrell: An Introduction to Music Information Retrieval

Presenter: Trudie Sandrock, University of Stellenbosch

Music information retrieval (MIR) is primarily concerned with the reduction of music to a workable data format and then extracting meaningful information from the data. MIR has been a very active field of research in the past decade and it is an interdisciplinary research area, spanning fields such as Music, Mathematics, Statistics, Computer Science, Engineering and Psychology. While Statistics is a field well-suited to dealing with the type of research problems encountered in MIR – and statistical techniques are often used in MIR research – researchers in the field are currently mostly from a computer science (machine learning) background. In this talk, I will aim to give a very brief introduction to the field of MIR and briefly highlight some of the issues encountered in MIR research. I will also touch on the statistical techniques underpinning well-known MIR apps such as Shazam and Soundhound as well as other music recommendation engines.

Multiple Imputation In The Presence Of A Detection Limit, With Applications: An Empirical Approach

Presenter: Cornelia J Swanepoel, North-West University, Potchefstroom Campus

Co-author(s): Mr. Shawn C. Liebenberg (Statistical Consultation Services, North-West University, Potchefstroom Campus)

Missing measurements that are reported to be below a fixed, known detection limit, is a regular occurrence especially in the environmental sciences. Such censored data are often ignored or "guessed" because measurements were made which were incorrectly reported, usually to be zero or to be equal to the detection limit. However, reliable estimates of the population parameters are required to perform statistical analysis. It becomes a complex task to perform when a large number of observations are below this limit. Rigorous robust estimation procedures are then needed.

This study focuses on density estimation in such scenarios by imputing data to replace the censored data below the detection limit in a sensible way. The maximum likelihood procedure of Cohen (1959) and several variants thereof, are then applied to estimate the parameters of the underlying density function. Estimation of this density function is then attempted by using the completed imputed data set. Various boundary kernel density estimators are applied comparatively.

More specifically, in this study three different Log-normal distributions will be considered. The above-mentioned methods are implemented in combination with four new multiple imputation procedures, to assess which of these nonparametric methods are most effective in imputing data to replace the censored values. Several kernel density estimators are fitted to the complete filled-in data set. Comparative measures are applied to establish which combination of strategies are the best to estimate the underlying density function in the presence of a detection limit. The results of a Monte Carlo simulation study are presented and conclusions and recommendations are made.

Yield probability as a method for cultivar selection

Presenter: Nicolene Thiebaut, Agricultural Research Council, Head-Office, Pretoria

Co-author(s): Dr Andre Nel and Annelie De Beer (Agricultural Research Council, Potchefstroom)

The selecting of cultivars in the grain crop industry under different environmental circumstances is very important for seed companies, farmers and industries in optimizing the profit and quality of the product. A few cultivars selection trials for different crops (maize, soya-, dry-beans, wheat and sunflower) are done yearly at different localities. It is very important that a correct classification of the cultivars is done, as well as an understandable and user-friendly presentation of the data for everybody involved. In this presentation the procedure of presenting the yield probability percentage above the mean yield is shown. Certain localities according to the crop criteria i.e. CV out of the ANOVA (coefficient of variation) is used in selecting these localities. For each cultivar a regression line is fitted with the cultivar means at particular locality as y variable versus the overall means for each locality as x variable (Draper and Smith). The yield probability potential for each cultivar for 'n normal curve is then determined and summarized in a table.

Identifying a secondary series for Stepwise Common Singular Spectrum Analysis

Presenter: Lienki Viljoen, Stellenbosch University

Co-author(s): Steel, S. J. (Department of Statistics and Actuarial Science, Stellenbosch University)

Stepwise Common Singular Spectrum Analysis (Stepwise CSSA) is a method to extend Singular Spectrum Analysis (SSA) to two or more time series which share a common manifold (R-flat) by using the stepwise common principal component (CPC) approach of Trendafilov (2010). This technique can be used to forecast a primary time series by using the information from a secondary series. We investigate the possibility of selecting a good secondary time series from a set of available candidates for use. Four procedures were studied reflecting different strategies to select the secondary series. It was based on the residuals obtained by combining the primary series with every candidate secondary series in a pairwise Stepwise CSSA. These procedures were compared versus using SSA when no secondary series is involved. Empirical studies suggest that the proposal performs well.

Young Statistician's Sessions Abstracts

(In Alphabetical Order)

Multilevel Modelling of Event Histories in Family Formation and Dissolution Studies in the sub-Saharan Africa

Presenter: Jesca Batidzirai, University of KwaZulu- Natal

Co-author(s): Manda, S.O.M (South Africa Medical Research Council, Pretoria) and Mwambi, H.G (School of Mathematics, Statistics & Computer Science, University of KwaZulu- Natal)

In family formation and dissolution studies, a subject may experience several events including childbearing, marriage, divorce and new marriage over time yielding event histories. We may be concerned in studying simultaneously the occurrences of two or more of these different events, adjusting for a number of socio- economic factors. In a typical application, the resulting data are in a multilevel structure. Using discrete time survival as a basis, multinomial logistic and competing risks models are used to fit multilevel multistate models to a typical family formation dataset from Sub-Saharan Africa

---

Influence of right-censoring on some kernel-smoothed hazard rates

Presenter: Dalene Bezuidenhout, Stellenbosch University

Co-author(s): de Villiers, Margaret; (Stellenbosch University) and Mostert, Paul J. (Stellenbosch University)

Survival analysis involves the analysis of time to an event of interest and the risk of a subject experiencing the event at a given time, given that the subject has not yet experienced the event. The latter, known as the hazard rate, is an important parameter in survival analysis.

Survival data sets frequently contain incomplete data. Censored samples contain some observations for which only the interval, rather than the exact value of the event time, is known. Right-censored samples contain observations for which it is only known that the

event occurred sometime after a particular time point. The purpose of this project was to investigate the effect of right-censoring on the estimated hazard rate using non-parametric techniques. These non-parametric techniques use the Nelson-Aalen estimator of the

cumulative hazard rate and smoothing with the uniform, Epanechnikov and biweight kernels. Numerous lifetime samples of different sizes with different levels of censoring were generated. The smoothed hazard rate was then estimated, recording the frequency of optimal global bandwidths obtained in each case. The performance of the hazard rate estimator was evaluated by estimating the variance, bias and coverage at pre-selected event times. The method of right-censoring applied in the simulations shortens the range of event times in a sample of lifetimes, thereby also reducing the range over which the hazard rates can be estimated. An increase in the level of censoring also results in shorter bandwidths, which in turn causes greater variance of the estimated hazard rates. The variance of the estimated hazard rates was found to increase steadily with increasing time, as has been found in previous studies, whereas the bias and the coverage did not show any trends at the times tested.

These non-parametric techniques were also applied to a real data set containing survival data on the time to re-infection with the sexually transmitted diseases gonorrhoea and chlamydia. The hazard rates and survival curves of the three risk groups in the data were discussed and compared.

A Distribution-Free Generally Weighted Moving Average Control Chart

Presenter: Niladri Chakraborty, University of Pretoria

Co-author(s): Chakraborti, S (Department of Statistics, University of Pretoria), Human,S.W. (Department of Statistics, University of Pretoria), Balakrishnan, N. (Department of Mathematics and Statistics, McMaster University)

Control charts are widely used in the manufacturing sector for monitoring and improving the quality of a process. Assuming a specific underlying distribution when a control chart is designed is often very restrictive because it can severely limit the application of the chart. Distribution-free control charts are therefore useful alternatives when information on the process distribution is partially or completely unavailable. In this regard, we propose a distribution-free generally weighted moving average (GWMA) control chart based on the well-known Wilcoxon signed-rank statistic. The performance of the GWMA-SR chart is compared to a number of existing control charts such as (i) the GWMA chart for subgroup averages, (ii) the GWMA chart based on the sign statistic, and (iii) an EWMA chart based on the signed-rank statistic. Results show that the proposed chart performs just as well and in many cases better than the existing charts.

The problem of zero-inflated count data: a discussion and application of zero-inflated and hurdle models

Presenter: Paul Claassen, Department of Statistics, University of Pretoria

Co-author(s): Fletcher, L (Department of Statistics, University of Pretoria)

When modelling count data the Poisson regression model is the go-to method. However, the primary assumption of the Poisson distribution is that the mean should be equal to the variance and this is very often not the case in practice. The situation where var(Y)>E(Y) is called overdispersion.

The most common causes of overdispersion are extra variance between subjects that can't be explained by the observed independent variables, also called unobserved heterogeneity, and the presence of excess zeros in the data. Many different models have been developed to handle the existence of overdispersion in count data models. Among them are the negative binomial regression model, zero-inflated and hurdle models.

The negative binomial regression model introduces an additional source of variation in the form of unobserved heterogeneity. This additional effect can be interpreted as either the collective effects of all the possible variables that weren't observed (or captured) and thus not considered in the model, or simply as an extra source of randomness. However, this approach is frequently not very effective when the overdispersion is due to excess zeroes in the dependent variable. This phenomenon occurs in many types of data, for example in health related data where the condition of interest is often not experienced by many of the subjects in the sample.

Zero modified models attempt to account for the excess zeroes by explicitly adjusting the mean structure to allow for the production of zeroes. Both zero-inflated and hurdle models are examples of finite mixture models where the underlying population is assumed to be made up of two unobserved or latent groups that have different experiences in terms of zero and positive counts. The processes that generate the zeros (a binary choice model) and positive counts (a standard count model) are also not required to be the same, as is the case in the Poisson regression model and negative binomial regression model. This allows for improved inference about the population.

As an example the Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, Poisson hurdle and negative binomial hurdle regression models are fitted to a sample dataset from a South African health insurer using SAS procedures as well as R.

Modelling Extreme Daily Temperature Using Generalized Pareto Distribution at Port Elizabeth, South Africa

Presenter: Tadele Diriba, University of Pretoria

Co-author(s): Debusho, LK (Department of Statistics, University of South Africa) and Botai, J (Department of Geography, Geo informatics & Meteorology, University of Pretoria).

The extremes of daily maximum temperature in summer and daily minimum temperature in winter were analysed using the generalized Pareto distribution (GPD) to the Port Elizabeth weather station data, South Africa. Since extremes in minimum and maximum temperatures series do not follow a normal distribution, the non-parametric methods namely, Kendall's tau test and the Sen's slope estimator were used for the trend analysis. A significant positive trend was observed in the extreme annual minimum temperature. However, the inclusion of a linear trend in the the log-scale parameter in the GPD model for the minimum daily winter temperature did not produce an improvement in the precision of parameter estimates. The results from the return level analysis show that by the end of twenty first century the extreme summer maximum temperature could be about 5 oC higher than the current in Port Elizabeth whereas the change in the winter minimum temperature will be less severe because the return level results suggest an increase of about 2 oC.

APPLICABILITY OF MULTILEVEL MODELS TO TEMPORAL SPECTRAL DATA

Presenter: Nontembeko Dudeni-Tlhone, CSIR

This study explored the application of multilevel models (longitudinal growth models, in particular) to analyse temporal spectral measurements collected from the eight tree species of interest. The main focus was to identify relevant models that could be used to answer the key questions concerning chlorophyll variation in time over for the main subjects (leaves nested within trees across species types). Different growth models with varying complexity levels were fitted in order to answer the relevant research question. Some of the key results showed that variation in REP (chlorophyll concentration indicator) was significant from the onset, with an initial average REP exceeding 705nm (standard error=1.85). This variation increased significantly over time (weekly) by about 0.22 units. A suitable model that could be used as input into a discriminatory model for the species was, therefore, identified.

Quadratic forms on complex elliptical random variables and its applications

Presenter: Johan Ferreira, University of Pretoria

Co-author(s): Bekker, A (Department of Statistics, University of Pretoria, South Africa) and Arashi, M (Department of Statistics, University of Sharhood, Iran)

Quadratic form densities of complex random elliptical matrices and their joint eigenvalue densities are derived, where these densities are represented by complex hypergeometric functions of matrix arguments which can be expressed in terms of complex zonal polynomials. An integral representation of this quadratic form is introduced. The connection between these densities and information theory is discussed. Special cases are described and select applications highlighted.

Bayesian optimal block designs for two-colour cDNA microarray experiments

Presenter: Dibaba Gemechu, University of Pretoria

Co-author(s): Debusho, L. K. (Department of Statistics, University of South Africa) and Haines, L. M. (Department of Statistical Sciences, University of Cape Town)

In two-colour complementary deoxyribonucleic acid (cDNA) microarray experiments only two treatments can be co-hybridized simultaneously on a single array and if there are more than two treatments, the problem of design arises, such as, for example, which treatments should be co-hybridized together and which treatments should be labelled with which dye fluorescent. Therefore, carefully designed microarray experiments to obtain efficient and reliable data to ensure the precise estimate of comparisons of interest are required. When the array effects are assumed to be random, the two-colour cDNA microarray experiments can be modelled using the linear mixed effects model. However, the traditional optimality criteria, namely the A- and D-optimality criteria, are functions of an unknown parameter, which is a function of the random array variance and the error variance. In this paper, Bayesian approach is considered in order to calculate optimal or near-optimal designs by introducing a beta distribution as a prior for the unknown parameter. The numerical results show that the Bayesian A- and D-optimal block designs are insensitive to the shape of the prior distributions.

Big data, compressed sensing and wavelets

Presenter: Charl Janse van Rensburg, University of Pretoria

Co-author(s): Fabris-Rotelli, I (Department of Statistics, University of Pretoria)

The aim of our study is to investigate the possibilities of applying the new exciting research area called Compressed Sensing (CS) in the world of big data, with the use of wavelets. CS was developed in the signal processing framework by Candes et al. (Candès, Romberg and Tao, 2006; Candes and Tao, 2006) and Donoho (Donoho, 2006). The aim of CS is to simultaneously acquire and compress a signal $f(t) \in R^N$ which is assumed to be sparse, for instance in the wavelet domain. Instead of sensing all N elements of f(t), we sense, or sample only $M \ll N$ elements from f(t) randomly. The signal f(t) is reconstructed perfectly from the M measurements using convex optimisation. We

argue that images can be seen as big data and hence provide evidence for using CS to solve big data problems utilising wavelets.

Modelling Net-Internal Migration in South Africa

Presenter: Xolani Jozi, Statistics South Africa

The aim of this paper was to model internal migration in South Africa using the 2011 Census data. The net-internal migration was modelled in the district municipalities of South Africa using Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR). The OLS and GWR model explain 71 and 76 percent of the observed net-internal migration at the district municipalities respectively. Additionally, the GWR had a lower AIC, this further indicates that the GWR model performed better than OLS regression in modelling net-internal migration in district municipalities. The model predicts well in the district municipalities of Limpopo. While, it performs poorly in the district municipalities of the Western Cape. The five factors that explains net-internal migration in district municipalities, are population density, proportion of the households that are renting, percentage of the households with no access to services, size of black and white populations. The Monte Carlo significance test results showed that the parameters of the white population vary significantly across space. The results from these models revealed that there was a strong relationship between the net-internal migration and economic variables, as well as living conditions and demographic variables.

Marginalization of Multivariate Gaussians with Application in Optimization Problems

Presenter: Francois Kamper, University of Stellenbosch

We investigate the use of message propagation in solving linear systems of equations without direct matrix inversion. Bickson (2009) shows that solving a linear system is equivalent to finding the mean (mode) vector of a multivariate Gaussian in canonical form and proposes the use of message propagation to perform marginalization. The Gaussian Belief Propagation (GaBP) algorithm requires $O(kp^2)$ computations to complete, where p is the number of equations and k is the number of iterations until convergence. This should be viewed in the context of the $O(p^3)$ computations required by direct matrix inversion. Bickson (2009) successfully applied the GaBP algorithm in fields such as Linear detection, Support Vector Machines (SVMs) and Kalman Filters. We propose further investigation into the behaviour of k and introduce a ridge-type tuning parameter (λ) to lower the computational cost associated with GaBP. Focus will be placed on finding an automatic way of selecting λ by minimizing an upper bound on the number of iterations required for convergence. We propose application of the GaBP algorithm in statistical optimization problems not considered by Bickson (2009). In particular the GaBP algorithm shows promise in the computation of Lasso paths for arbitrary likelihoods through quadratic approximations.

The impact of Infrastructure on South Africa's Economic Growth

Presenter: Letlhogonolo Khenene, Statistics South Africa

Sequential regression imputation of air quality data

Presenter: Sibusisiwe Khuluse-Makhanya, CSIR

Co-author(s): Stein, A (Faculty of Geo-information Science and Earth Observation, University of Twente) and Debba, P (Built Environment, CSIR)

Poor air quality is a public health concern, hence for monitoring, annual statistics such as the number of days an air quality standard is exceeded are of importance. The air quality monitoring network in the Highveld region of South Africa consists of 36 stations whose data is publically available. The main challenge with this data is the high proportion of missing observations. When ignoring the missing data annual air quality statistics have large standard errors. Assuming air quality data to be missing at random, relationships between coarse particulate matter (PM10), nitrogen dioxide (NO2), sulphur dioxide (SO2) and meteorological variables (relative humidity, temperature, wind speed and wind direction) are exploited using sequential regression imputation. A varying coefficients model is chosen to account for temporal and area characteristics; that is seasonality and serial correlation for the former. The results presented are for the Vaal-triangle portion of the network which consists of 6 stations. Using a hold-out sample from two of the six stations, the quality of the imputation is evaluated.

LASSO Tuning Parameter Selection

Presenter: Lisa-Ann Kirkland, University of Pretoria

Co-author(s): Kanfer, F (Department of Statistics, University of Pretoria) and Millard, S (Department of Statistics, University of Pretoria)

The LASSO is a penalized regression method which simultaneously performs shrinkage and variable selection. The output produced by the LASSO consists of a piecewise linear solution path, starting with the null model and ending with the full least squares fit, as the value of a tuning parameter is decreased. The performance of the selected model therefore depends greatly on the choice of this parameter. This paper attempts to provide an overview of methods which are available to select the value of the tuning parameter for either prediction or variable selection purposes. A simulation study provides a comparison of these methods and assesses their performance.

Handling longitudinal continuous outcomes with dropout missing at random: A comparative analysis

Presenter: Abdalla Kombo, UKZN

Co-author(s): Satty A (School of Statistics, Mathematics and Computer Science, UKZN) and Mwambi H (School of Statistics, Mathematics and Computer Science, UKZN)

Dropout is a pervasive problem in longitudinal studies, and it is the result mainly of non-response due to individuals who leave the study and are therefore lost to follow-up. This paper focuses on dropout missing at random (MAR), in the sense that the probability of dropout is dependent on the observed responses. We compare multiple imputation (MI) and inverse probability weighting (IPW) methods to analyze longitudinal data with dropout under different dropout rates and sample sizes. Application will be confined to the continuous outcome case. Based on simulated data, results from IPW are compared with those obtained from MI, in terms of bias and efficiency. The results in general favoured MI over IPW.

Spatial Sampling

Presenter: Christine Kraamwinkel, University of Pretoria

Co-author(s): Fabris-Rotelli, IN (Department of Statistics, University of Pretoria)

Conventional sampling methods often assume that data is independent and identically distributed within the population (or within subpopulations) and that selection probabilities of elements are known. In reality, and specifically in the setting of wildlife research, the data to be collect is usually spatially autocorrelated and heterogeneous with selection probabilities seldomly known. When using conventional sampling designs, this leads to inefficient and non-representative samples with questionable estimation value. We investigate the theory underlying spatial sampling and its possible application to wildlife and animal research.

Long Memory and Structural Breaks: An Application to Platinum Price Return Series

Presenter: S Kubheka, University Of South Africa Department Of Statistics

Co-author(s): E. Ranganai

The platinum sector in South Africa has experienced a lot of setbacks with huge economic impacts which did not only affect South Africa but globally. These events normally introduce jumps and breaks in data which then changes the structure of the underlying information. In this paper, we investigated structural changes in the platinum return series and changes in long range dependence of volatility. Tests that are employed to detect structural changes in returns are the iterative cumulative sum of squares (ICSS) algorithm and multiple structural change models. To test breaks in the long memory of the volatility process, we use methods introduced by Shimotsu (2006) which examine structural changes in the long range dependence of platinum price return series. Visual inspection, Wald statistic and mean differencing methods were used with sub-samples to examine structural changes in long range dependence. To further substantiate the results of the tests done, we used the visual inspection cumulative samples methodology which estimates the long range parameter overtime. All the tests used suggested structural changes in both the return series and long range dependence parameter. This suggests that in modeling of platinum returns, models which take into account different regimes of the series should be considered and compared to standard models to understand whether long memory in the series is true or spurious.

Feasibility in using Greeks...to manage options' risks - The Management Perspective

Presenter: Sibusiso Magagula, Nedbank/UNISA

The 'Greeks' methodology implemented in practise to quantify option's risks is based on the normality assumption. This study investigates in depth the departure from normality assumption when using financial data to calculate 'Greeks' in an option deal which is used to hedge accurately the downward and upward risks of an underlying assets to be purchased (call) or sold (put) in the near future. The relative error methodology developed in the study conclude that stock index call and put options which are 'out-of-the-money' and `in-the-money' respectively, their 'Greeks' have higher model risk due to the normality assumption being violated by the underlying financial data used in the research. This study confirms that hedging options such as stock index options which are short-dated, their risk management tools, that is, the 'Greeks' should not be analysed in isolation but integrated with other risk management tools such as expert judgment and independent oversight provided by other teams in the organisation.

Creating mixtures of Pareto distributions via beta type generators

Presenter: Seite Littah Makgai, University of Pretoria

Co-author(s): A. Bekker, J.T. Ferreira

The beta distribution has been widely used to model a variety of uncertainties as well as probability distributions of variables. The Pareto distribution is known in the modelling and analysis of lifetimes, which forms important aspects of statistical work. The newly proposed class is constructed by taking the Pareto as the parent distribution and the new generalised beta type as the generator distribution. This flexible class includes well known models as the Kumaraswamy-Pareto distribution and the beta type I-Pareto distribution, as well as other new models. By this method, new contributions through mixtures of cumulative distribution functions of the Pareto distribution are proposed and studied. Statistical properties such as moments and Renyi entropy are investigated for each model. A real data set is used to compare the newly derived models with other known distributions, using the method of maximum likelihood estimation to estimate the model parameters.

Generalized Burr Type II - exponential distribution

Presenter: Tsitsi Makoni, University of Pretoria

Co-author(s): van Staden, P.J. (Department of Statistics, University of Pretoria)

In this paper the distributional relationship between the Burr Type II and the generalized exponential distributions is illustrated. Using a quantile-based approach, the generalized Burr Type II - exponential distribution is then developed. Apart from the Burr Type II distribution, the generalized

exponential distribution and their limiting or special cases, the proposed distribution also includes the skew logistic distribution as a special case. The shape characteristics of the distribution are investigated with L-moment ratios, in particular, the L-skewness and L-kurtosis ratios.

Distribution-free CUSUM and EWMA Control Charts based on the Wilcoxon Rank-Sum Statistic using RSS for Monitoring Mean Shifts

Presenter: Jean-Claude Malela-Majika, UNISA

Co-author(s): Rapoo, E (Department of Statistics, University of South Africa)

Whenever a practitioner is not really sure about the underlying process distribution, alternative monitoring schemes that may be used are called nonparametric (NP) charts. NP monitoring schemes have been shown to have some attractive advantages compared to their parametric counterparts e.g. these are more flexible and very robust. A NP scheme mostly used to monitor the difference in the means of two samples is called the Wilcoxon Rank-Sum (WRS). Using extensive Monte-Carlo simulations, in this paper, we show that using the Ranked Set Sampling (RSS) technique rather than the commonly used Simple Random Sampling (SRS) technique results in CUSUM and EWMA WRS schemes with much better out-of-control detection capability. We thoroughly illustrate this phenomenon by using a variety of run-length characteristics and also using the overall performance statistic called the Relative Mean Index. Based on these, the CUSUM and EWMA WRS based on RSS yields the best performance compared to a number of its competitors and hence makes it a strong contender in many applications where existing WRS schemes are used.

The first-order autoregressive process - a Bayesian perspective

Presenter: Hossein Masoumi Karakani, University of Pretoria

Co-author(s): Van Niekerk, J (Department of Statistics, University of Pretoria) and Van Staden, P.J (Department of Statistics, University of Pretoria)

The first-order autoregressive process, AR(1), has been widely used and implemented in time series analysis. Different estimation methods have been developed and proposed in the literature for the autoregressive parameter. This study focusses on subjective Bayesian estimation of the autoregressive parameter as oppose to objective Bayesian estimation. The truncated normal distribution is considered as a prior. The conditional posterior distribution, as well as, the conditional Bayes estimator, are derived. A simulation study is used to investigate the performance of the newly derived estimators using a Markov Chain Monte Carlo sampling scheme as well as the analytical expressions derived. This estimation method is applied to a real dataset.

Industry-Wide Data Governance Model For Credible Rating In Nigeria

Presenter: godson Mesike, university of Lagos, Akoka,Nigeria

Co-author(s): Adeleke, I.A (Department of Actuarial science and Insurance, University of Lagos), Hamadu, D (Department of Actuarial science and Insurance, University of Lagos)

A major data problem facing the insurance companies today is that of relevance, timeliness, completeness and data management deficit pervading the Nigeria insurance sector. The data and information available at various companies are not standardized in their collection, presentation and storage, and raw data are used in many different applications which are put into many formats. This has led to situations in which the same data are reformatted, reproduced and presented to different users for different purposes. Thus, the need to have clearly defined data standards and rule sets that can streamline and keep multiple versions of the data better organized. Accurate and valid data are the lifeline of correct pricing and experience rating. This study proposed an industry-wide data governance in a regulated and competitive Nigeria business environment for credible underwriting and profitability. It allows an organization to consolidate the current data in its disparate and fragmented production systems and combine it with historical values; and also incorporates various initiatives for having in place reliable, up-to-date, efficient and effective statistical system. Rather than relying solely on company-specific claim experience, better estimates may be obtained by incorporating inter-company experience and using industry-wide claims data. The variability of claim costs and the challenge of estimating the cost of insurance at inception of the policy make it necessary for companies to frequently assess the credibility upon which pricing, valuation and other product management decisions are made. It also integrates variation in expected claim costs from insurer to insurer in the industry, variation between expected claim costs from group to group for a given insurer, and variation from insured to insured within a group. Inferences can be made about the industry's average, companies' average and group-specific average. Policy implications and recommendation are discussed.

Assessing the Productivity of Selective Container Terminals in Africa using DEA

Presenter: Barend Mienie, Nelson Mandela Metropolitan University

Co-author(s): WJ Brettenny, Nelson Mandela Metropolitan University, Department of Statistics; GD Sharp, Nelson Mandela Metropolitan University, Department of Statistics

Data envelopment analysis (DEA) is used to assess the efficiency of 15 container terminals in Africa. The Banker, Charnes and Cooper (1984) DEA-BCC model is used to determine and rank the efficiencies of the container terminals for 2013 and 2014. The results show that selected South African container terminals can improve on their operations relative to some of their neighbours to the North. Bootstrapping methods as developed by Simar and Wilson (2000b) are used to investigate and clarify the results. The Malmquist Productivity Index model, as introduced by Färe, Grasskopf, Yaisawarng, Li and Wang (1990), is used to track and explain changes in efficiency over the period 2013 to 2014.

Assessing Factors Affecting Admission Time Of Kaposi Sarcoma Using Survival Analysis: A Case Of Zomba Central Hospital

Presenter: Henry Mlinde, Department of Mathematics, University of Malawi, Chancellor College, Zomba, Malawi

Co-author(s): J.Simbeye, E.Mwakilama, Department of Mathematics, University of Malawi, Chancellor College, Zomba, Malawi

Kaposi is most HIV-related malignancy worldwide and the most frequently diagnosed cancer in several Africa countries. In Malawi KS accounts up to 34 % of all diagnosed cancer cases. This study aims at assessing factors affecting admission time of Kaposi sarcoma patients using survival analysis.

The study was conducted at Zomba Central Hospital one of referral hospital in the southern region of Malawi. Patient characteristics which included CD4 count, admission and discharge diagnostics and admission time were collected from patients register entered in SPSS and analysed using STATA 12. Descriptive statistics, life table, Kaplan-Meier and Chi-Square were used to measure relationship between patient characteristics and admission time. Cox proportional hazards model was fitted to assess factors that affect length of admission time.

A total of 163 KS patients had a mean age of 38 years with range of 15 to 92 years were included in the study of which 78 (48%) were males and 85(52%) females. From all KS patients 54(33%) were infected with HIV while 21(13%) were not infected and 87(54%) were of unknown HIV status with 44 (29%) on ART. Maximum time patients spent in the hospital was 353 days with mean of 195 days and confidence interval of (176.920, 212.410) days. The study showed association between gender and HIV status (P-Value=0.594) and with ART status (P-Value=0.525). There was significant association between admission time and age, gender, HIV status, ART status, CD4 count of KS patients. KP showed that females survived better than males (P-Value=0.030), HIV infected KS patients survived better than both Uninfected and those patients with Unknown status (P-Value = 0.020) and those on ART and Pre-ART survived better than those not on ART and Uninfected (P-Value= 0.013).

Admission time of KS at Zomba Central hospital is greatly affected by gender, HIV status, ART status, CD4 count, admission diagnostics and referrals. It is recommended that HIV status of every KS patient should be known before admission to reduce admission time.

A randomized response survey on the risky behaviors of certain University students

Presenter: Thuto Mothupi, University Of Botswana

Co-author(s): Arnab,R(Department Of Statistics,University Of Botswana)

The HIV/AIDS epidemic continues to ravage Sub Saharan Africa (SSA). HIV/AIDS and related sickness are the leading cause of morbidity and mortality in Botswana. The Botswana AIDS Impact Surveys (BAIS II, BAIS III and BAIS IV) estimated national prevalence rate of 17.1%, 17.6% and 18.5% in the years 2004, 2008 and 2013 respectively. Young people are among the most vulnerable groups; half of new infections in this region in the year 2009 occurred among those in the age range of 15 to 24. The common, risky, sexual practices in this age group include early sexual intercourse, multiple sexual partners, unprotected sexual intercourse, engaging in sex with older partners and non-regular partners such as commercial sex workers. A survey was conducted on the students of a certain university to determine the prevalence of various risky behaviors. The data was collected using direct method (DR) and randomized response (RR) methods using Warners (1965), Greenberg (1969) and Odumande & Singh (2009) for qualitative characteristics and Ericksson (1973) for quantitative characteristics. It is found that RR method yielded higher estimates of the prevalence. The standard errors were determined by using Jacknife method.

Diagnosis of Zero Inflation

Presenter: Modupi Peter Mphekgwana, African Institute for Mathematical Sciences

Co-author(s): Hewson, P ((Department of Statistics, Plymouth University (UK))

The Generalised linear model is one of the most widely used statistical models, where a conventional linear predictor is linked to an error distribution taken from the exponential family. To apply such a model however requires that many assumptions are made, such as the plausibility of the error distribution. A common problem which occurs with discrete error distribution (Poisson, Binomial and Bernoulli) in practice is zero-inflation. In this case, there more zero responses in the data than are predicted by the model. A mixture model, zero-inflated Poisson (ZIP) model has become a popular approach to take into account the excess of zeroes in the data. This project examines the validity of diagnostic procedures used to indicate whether a particular data sets may indeed exhibit zero-inflation. A Vuoung test and a score test have been developed for assessing count data with zero inflation. The power of the test statistics are evaluated by simulation studies. The result shown that the use of Vuong's test for non-nested models as a test of ZIP model is imprecise, and the score test shown to perform satisfactorily under a wide range of conditions. Use of the test is illustrated on road traffic accident in the Limpopo province data.

Statistical modelling and spatial mapping of crime in South Africa.

Presenter: Belisha Naidoo, University of KwaZulu-Natal Westville

This study aims to statistically model and spatially represent the problem of crime in South Africa. We aim to identify the factors affecting crime rates in South Africa, investigate the relationship between perception and outcome of crime and seek to find patterns in occurrence of crime. The data for the study was obtained from the Victims of Crime Survey, conducted by Stats SA (25605 households), as well as aggregated crime data from 1140 police stations.

Modelling minimum average daily temperature using extreme value theory with a time varying threshold

Presenter: Murendeni Maurel Nemukula, University of the Witwatersrand

Co-author(s): Sigauke, C (School of Statistics and Actuarial Science, University of the Witwatersrand)

In this paper we present an application of the Generalised Pareto Distribution (GPD) in the modelling of average minimum daily temperature in South Africa for the period January 2000 to August 2010. A penalized cubic smoothing spline is used as a time varying threshold. A non-parametric extremal

mixture model is then used to obtain a sufficiently high threshold. The data exhibit evidence of short-range dependence and high seasonality. We then decluster the excesses above the sufficiently high threshold and fit the GPD to cluster maxima. The parameters are estimated using the maximum likelihood method. The estimate of the shape parameter shows that the Weibull family of distributions is appropriate in modelling the upper tail of the distribution of average minimum daily temperature in South Africa. The bootstrap resampling method is used as an assessment tool for uncertainty in the parameter estimation. This study has shown that the use of the penalized cubic smoothing spline as a time varying threshold to time series data which exhibits strong seasonality provides a good fit of the GPD to cluster maxima. This results in more accurate estimates of return levels.

Long - memory in Asset Returns and Volatility: Evidence from West Africa

Presenter: Emmanuel Numapau Gyamfi, Department Of Statistics, University Of Venda

Co-author(s): Kyei, K.A (Department Of Statistics, University Of Venda) And Gill, R (Department Of Mathematics, University Of Louisville)

There has been mixed conclusions on market efficiency of stock markets in Africa. This paper measures the degree of long - memory or long - range dependence in asset returns and volatility of stock markets in Ghana and Nigeria. The presence of long - memory opens up opportunities for abnormal returns to be made by analyzing price history of a particular market. We employ the Hurst exponent to measure the degree of long - memory in a given market.The Hurst exponent is used as our efficiency measure which is evaluated by the Detrended Fluctuation Analysis (DFA). Our findings show strong evidence of the presence of long memory in both returns and volatility of stock markets in Ghana and Nigeria. This suggests that none of the markets is weak-form efficient.

Statistical methods for the detection of non-technical electricity losses: A case study for Nelson Mandela Bay Municipality

Presenter: Sisa Pazi, Nelson Mandela Metropolitan Univerisity

Co-author(s): Sharp, G.D (Department of Statistics, Nelson Mandela Metropolitan University) and Clohessy C ((Department of Statistics, Nelson Mandela Metropolitan University)

Electricity losses from source to end user are classified into two categories that is technical and non-technical losses. Technical losses are due to energy dissipated in the conductors, equipment used for transmission and distribution lines. These losses are engineering issues. Non-technical losses are primarily caused by electricity theft, billing errors and illegal connections. There exist several statistical techniques used to identify and detect non-technical losses. The primary purpose of this research is a practical application of statistical assessment to identify and detect electricity fraud. Statistical techniques used include Support Vector Machines (SVM), Naïve Bayes and Hidden Markov Models (HMM). A case study for the Nelson Mandela Bay Municipality (NMBM) will be used and the results of the assessment reported. The research aims to contribute to the sustainability of the energy directorate of NMBM by providing them with a method for electricity theft identification.

Predicting the future of the 2015 Rugby World Cup using Random Forest variants

Presenter: Arnu Pretorius, Stellenbosch University

Co-author(s): Surette Bierman

Random forests (RFs) are known to yield state-of-the-art performance in a wide array of application domains. Examples include astronomical object classification, digital image classification, text classification and genomic data analysis.

Over the past decade, many RF variants have been proposed in the literature. Fawagreh et al. (2014) provide a good overview. Some important aspects in contributions include: limiting the number of trees voting toward predictions, replacing majority voting with more sophisticated dynamic integration techniques, using weighted random sampling to pick features in the face of a large number of uninformative features, extension to on-line RF algorithms, and the use of genetic algorithms to improve RF performances. More recently, contributions focused on modifications to RFs with a view to enhance performance in the face of high-dimensional data. See for example Nguyen et al. (2015) and Xu et al. (2012) in this regard.

We present some of the more important variants, illustrating their application in the prediction of world cup rugby match outcomes. For this purpose, the use of cloud computing services in training online models is also presented.

A New Approach to Covariance Modeling of Longitudinal Data

Presenter: Anasu Rabe, University of Botswana

Co-author(s): Shangodoyin, D.K. (Department of Statistics, University of Botswana) and Thaga,K.(Department of Statistics, University of Botswana)

To date, it has been empirically established in the literature that longitudinal responses tend to exhibit a natural process of growth or decay and we utilize this feature in proposing a cholesky-based joint mean-covariance model for longitudinal data. We establish a direct interpretation of the variance of the cholesky factors to the covariance matrix by exploiting its relationship to eigenvalues. We project a hermitian Eigenvector over Rn and use polar coordinates to obtain unconstrained parameterization of the covariance matrix and develop a joint mean-covariance modeling framework. The efficiency and parsimony of our approach is supported by real data analysis and simulations.

An application of the extensions of the Cox model to model the incidence of pneumonia and repeat episodes of pneumonia in boys &

Presenter: Jordache Ramjith, Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town

Co-author(s): Myer, L (Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town)and Zar, H (Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town) a

Introduction: Pneumonia is one of the leading causes of death in children under the age of five in developing countries. It is uncommon for a proportion of children to experience repeated episodes of pneumonia. Pneumonia incidence literature favours the Cox proportional hazards (CPH) model to assess the effect of risk factors on time to first episode and Poisson regression models the discrete counts of episodes. As a consequence we fail to consider possible correlation between events within infants' follow-up and further overlook the possibility of a temporal effect of covariates. Extensions of the CPH model to understand recurrent pneumonia have been applied within the health sciences.

Aim: We set out to evaluate extensions of the CPH model when investigating the effect of sex and sex adjusted risk factors on the incidence of repeated pneumonia episodes in a cohort of 1008 infants enrolled in the Drakenstein child health study between May 2012 and April 2015.

Methods: Pneumonia was diagnosed according to the WHO clinical case definitions: any infants who presented with cough or difficulty breathing and age-specific tachypnoea (≥ 50 breaths per min for children aged between 2- 12 months) or lower chest wall in-drawing. Repeated events were any events that happened more than 14 days after a previous event. Standard CPH models were used to investigate risk factors on time to first event stratified by sex. Extensions of CPH, the Andersen-Gill model, the Wei, Lin & Weissfeld model and the Prentice, Williams & Peterson's gap-time and total-time models were then applied for repeat episodes.

Discussion & Conclusion: Parameter coefficients and robust standard errors were reported. Scaled Schoenfeld residuals were used to test the PH assumption. Schoenfeld residual plots were used to assess the overall goodness-of-fit of these models. The models were compared on both their performance and interpretability. This type of analysis will provide further insight into the monitoring of children who are at risk of developing repeat pneumonia episodes.

Acknowledgement: This study was funded by the Bill & Melinda Gates Foundation (grant number OPP 1017641). We thank the study staff; the clinical and administrative staff of the Western Cape Government Health Department at Paarl Hospital and at the clinics for support of the study; and the families and children who participated in the study.

The risk performance of the heteroscedastic preliminary test estimator under different loss functions

Presenter: Christiaan Ras, University of Pretoria

The problem of heteroscedasticity is commonly encountered in regression models and it is known that, under heteroscedasticity, the Ordinary Least Squares estimator is relatively inefficient. This presentation focuses on the risk performance of a preliminary test estimator for regression coefficients, after a preliminary test for heteroscedasticity has been performed. The risk performance of these estimators relative to their component estimators, the Ordinary Least Squares and the Two-stage Aitken estimators, has been predominantly investigated under the symmetric Squared Error loss function and the Balanced loss function. However, the use of an unbounded, symmetric loss function can be inappropriate in estimation problems where overestimation and underestimation have different consequences. This presentation sets out the risk performance under different proposed loss functions, namely the symmetric (bounded) Reflected Normal loss functions, as well as the asymmetric (unbounded) Linear Exponential and Bounded Linear Exponential loss functions. The risk for the preliminary test estimator and its component estimators are derived under the different loss functions and numerically evaluated by making use of Monte Carlo simulations. It is shown that, in general, the risk under Linear Exponential loss is higher than the risk under the Reflected Normal loss and Bounded Linear Exponential loss. Also, under a slight asymmetric loss scenario, the risk under the Bounded Linear Exponential loss drops significantly when compared to that of the Reflected Normal loss. An economic application is included and from these results, as well as those from the simulation studies, it is clear that the relative risk gains of the Two-stage Aitken estimator and the preliminary test estimator over the Ordinary Least Squares estimator generally increases with higher loss asymmetry and higher levels of heteroscedasticity.

A comparison of domain expert classification and unsupervised computer classification techniques: A case study of the Orange Riv

Presenter: Michaela Ritchie, Council for Scientific and Industrial Research

The Orange River Estuary is found on the border of South Africa and Namibia and is South Africa's second most important estuary based on 2012 importance scores. Researchers at Nelson Mandela Metropolitan University's Botany department have thus far used visual interpretation of a SPOT-5 satellite image and a field visit to the site along with their domain knowledge to generate a classification map of the Orange River Estuary. Due to inaccessibility and the subsequent lack of sufficient in-situ data and the cost of field visits, unsupervised classification, such as the k-means algorithm, of the same SPOT-5 satellite image has been considered. The unsupervised classification can be used to monitor change in the study area and allow for timeous detection of possible degradation in the study area. Preliminary results from this investigation will be presented.

New Procedure for Probabilistic Hazard Assessment from Incomplete and Uncertain Data

Presenter: Ansie Smit, University of Pretoria Natural Hazard Centre, University of Pretoria

Co-author(s): Kijko, A (University of Pretoria Natural Hazard Centre, University of Pretoria) and Fabris-Rotelli, IN (Department of Statistics, University of Pretoria) and Van Staden, PJ (Department of Statistics, University of Pretoria)

Natural disasters and their impacts are not a new phenomenon. Evidence of these impacts can be seen in remnants of catastrophes through different environmental markers such as geological deposits. Most hazard and risk assessment models were developed in countries where there are extensive catalogues for the different hazards. This is however not the case in third world countries and especially in Africa. In many instances, instrumental data are still not collected on a level which

will allow for the effective use of these modelling tools. The incorporation of additional information into the calculation of hazard and risk models is also necessary to properly calibrate results. This includes information from paleo and historical observations not measured through normal instrumental techniques. Observed extreme natural hazards are normally very few and far between. Many events are known only through investigations of environmental markers and historical narratives, therefore vulnerable areas have long palaeo- and historic records available containing information of the largest and catastrophic occurrences. A technique for the assessment of probabilistic analysis is introduced which permits the assessment of the key distribution parameters in the case when the catalogue consists of the palaeo, historic as well as the most recent, instrumentally recorded ('complete') events. The technique can be applied to different types of natural hazards such as fires, hail, earthquakes and tsunamis. The technique is illustrated through the assessment of probabilistic tsunami hazard assessment for the area of Chile and the probabilistic seismic hazard assessment for Cape Town, South Africa.

Birth Registration In Uganda: Challenges, Opportunities And Lessons

Presenter: Farouk Ssekisaka, Makerere University

Co-author(s): Shamirah Iga

This paper provides a basis to assess the inherent factors affecting birth registration and identify best practices and recommendations to improve systems and procedures for effective birth registration in the sub-Saharan Africa taking Uganda as the case study. The researcher used both primary (key informant interview with a randomly selected sample of 500 individuals and secondary data, using the Uganda Demographic Health Survey 2011 dataset and a descriptive design. A functioning system of birth and civil registration ensures that the country has an up-to-date and reliable database for planning, maintaining education, health and other social services for the community.

Results indicate that, among the 1.5 million babies born each year, only 20% registered under the age of five, awareness of birth registration (45%) was low in the study population. Not only are registration services inaccessible to most Ugandans, but registration fees and other hidden costs (such as transport charges) rendering them too expensive for the majority to afford. In late 2011, UNICEF in partnership with Mulago Hospital and Uganda Telecom launched an electronic birth and death certificate registration process, the Mobile Vital Record System (MobileVRS) in Uganda. This system intended to reduce the cost of producing a certificate, the time it takes to issue a certificate and improve the security and authenticity of the records. Mobile VRS is now operational in all 135 government and missionary hospitals and 36 out of 112 districts across Uganda. This, along with the use of Mobile VRS in health outreach programmes such as Family Health Days, has led to an unprecedented increase in birth registration over the past few years. From 30% in 2011 to approximately 48% today, compared to the nominal increase from 21% in 2006 to 30% in 2011, Uganda has indeed made remarkable progress. Through this, the government has developed a digitalised system that will be used by hospitals and local governments to register births and deaths. However this has not achieved all its intended objectives due to no comprehensive monitoring and evaluation system, poor information technology infrastructure, incompatibility of the data processing and analysis systems, Reliance on volunteers to collect data at the grassroots level and most of all awareness about birthregistration, as most people still do not understand its importance.

The researcher therefore recommended increased awareness of the public in general, women in particular, of the importance of birth, especially as a fundamental right of the child; increased capacity for birthregistration duty-bearers to perform their assigned duties; improve provision of materials, equipment and infrastructure required to administer registration;improved links between birth registration and social services relevant to children (immunization, basic education, special education, orphan care), to improve the automation and computerization of the processes involved in the production of information on birth and death registration for planning purposes and distribution of resources from national to the lower level, and to achieve efficiency and integrity of birth certificates.

Islamic Banking as an option for developing Sub-Saharan Africa economies

Presenter: Farouk Ssekisaka, Makerere University

The objective of this paper is to improve understanding of the market for Islamic banking and finance in the country. The paper sought to assess the future prospects and challenges of Islamic banking system in Sub-Saharan Africa, and to establish whether Islamic banking is viable and practical in the region and the whole African continent. The sample of the study consisted of 100 retail customers who were the holders of accounts in the various banks, 100 Bankers, 100 Economist, 100 Business Entrepreneurs and 100 Business managers. Sample data collected by use of questionnaires administered by the researcher and a research assistant. Data analysis method used is based on the quantitative approach using descriptive statistics and chi-square analysis. The study used both Primary and Secondary Data. The secondary data was collected from recorded materials such as financial reports, journals, research papers and any other written material concerning the above topic using Descriptive, Uni-variate and Bi-variate analysis. The findings of the paper is that Islamic Banking is yet to take up fully as there are still many domestic and regional obstacles to the operations of the system in the region as the necessary legislative amendments have not been made in some countries, limited number of Domestic experts, differences in scholastic interpretation, the industry being mostly demand driven and inadequate political will. It was recommended that Islamic banking is desirable and practicable in Sub-Saharan Africa if the challenges are seriously taken care of by development of local talent ,launching a public awareness campaign to both Muslims and non-Muslims communities, providing the needed infrastructure (i.e. amending all necessary legislations and accounting and prudential frameworks), building capacity at the central bank (especially on supervision), and considering the need to set up an appropriate liquidity management framework and introduce adequate monetary operations instruments.

Modeling Length of Hospital Stay for Tuberculosis In-Patients at Queen Elizabeth Central Hospital: Applying Competing risks

Presenter: Halima Twabi, Chancellor College

Co-author(s): Namangale, J. J (Department of Mathematical Sciences, Chancellor College) and Mukaka, M (Nuffield Department of Medicine, University of Oxford (UK), Mahidol-Oxford Tropical Medicine Research Unit,Faculty of Tropical Medicine, Mahidol University)

A retrospective cohort study was used on adult TB in-patients from Queen Elizabeth Central Hospital (QECH) SPINE database to identify factors explaining time to discharge from hospital while accounting for a competing event; death. The study aimed to apply and compare estimates of competing risk models on TB data that collected patients socio-demographic characteristics and patients medical information. Semi-parametric Cause-specific hazards were used to model the effect of HIV status, ART Status, age, and Sex in relation to death or discharge from hospital. The Fine and Gray regression estimates were compared to the cause-specific estimates. Test for model assumptions and diagnostics were conducted. Findings showed that the Fine and Gray regression explained best the effect of the covariates to the probability of a patient being discharged or dying. Further the main factors affecting length of hospital stay among TB in-patients were age and HIV Status. HIV positive patients were 17.6% less likely to be discharged from hospital compared to HIV negative patients ($p=0.048$) and with an increase in age, the hazard of discharge decreased by 2% ($p < 0.001$). It is important to present results on both the event of interest and the competing risk and use the cumulative Incidence function for calculating probability of an event. Competing risks data should be modeled using both the CSH model and the Fine and Gray model when studying length of hospital stay.

Comparative subjective Bayesian analysis of the normal model

Presenter: Janet Van Niekerk, University of Pretoria

Co-author(s): Bekker, A (Department of Statistics, University of Pretoria) and Arashi, M (Department of Statistics, University of Shahrood, Shahrood, Iran and Department of Statistics, University of Pretoria)

The problem of Bayesian estimation within the univariate normal model is addressed. In this paper a subjective Bayesian approach is followed to derive new estimators for the parameters of the normal model by assuming

the new hypergeometric gamma prior. This prior includes the gamma and the noncentral gamma as special cases. A comparative study is then undergone to evaluate the performance of the estimators for specific cases as well as the

estimators derived under the inverse gamma, gamma priors and the objective Jeffreys' priors, the latter results in the maximum likelihood estimators, using different measures. A simulation study is performed to illustrate the novelty of these new estimators and to investigate the accuracy gained by assuming the hypergeometric gamma prior and using the analytical expressions.

Using multilevel analysis to determine the learner and school factors associated with mathematics performance

Presenter: Lolita Winnaar, University of the Western Cape

Co-author(s): Prof. Renette Blignaut (University of the Western Cape) and Dr. George Frempong (Human Sciences Research Council)

In order for schools to provide quality education it needs to be effective. An effective school as defined by Bennet, Crawford and Cartwright (2003: 176) as a “school in which students’ progress further than might be expected.” The international literature indicates that multilevel modelling provides a better estimate and analysis of school effectiveness, especially when considering the multilevel nature of educational data. Yet, in the South African context, only a few studies have employed multilevel modelling in school effectiveness analysis. Using multilevel analysis the intention is to determine, firstly; the learner home background factors that affect learner mathematics performance (where mathematics is used as a proxy for school performance). Secondly; to select the school level factors that affect learner performance. Research has shown that in addition to the effect of the school on learner performance it is important to note that the learners’ home background also affects the performance of learners. It is thus important to control for the learners’ background in order to determine the factors within and between schools associated with learners’ performance which very often is a proxy for school effectiveness. The results will show that large variation exists between schools in South Africa. At the learner level the background factor that has the largest effect is age; with learners who are age appropriate obtain higher scores than older learners. Learners’ attitudes toward mathematics are extremely important with learners who like and value mathematics obtaining higher scores than those who do not. At the school level the variables found to be significant were school Socio-Economic Status (SES), general infrastructure, teacher working conditions and whether or not a teacher has specialised training in mathematics. Learners in high SES schools obtain higher scores than learners from low SES schools. Learners who are taught by teachers who have specialised in mathematics and are happy with their working conditions perform better than learners taught by teachers who have not specialised in mathematics and who are unhappy with their working conditions. A very important finding is that school SES is still a very strong determinant of mathematics performance but when factors like teacher working conditions, mathematics specialisation and infrastructure is improved in schools then the effect of SES reduces.

Poster Presentations Abstracts

(In Alphabetical Order)

Compressed sensing and Statistical Preprocessing of fMRI data

Presenter: Altus Coetzee, University of Pretoria

Co-author(s): Fabris-Rotelli, I (Department of Statistics, University of Pretoria)

In this research we discuss the field of Functional Magnetic Resonance Imaging (fMRI). The statistical methods used to the adjust the sequences of fMRI images accumulated during such a study are investigated and explained. Sparsity is assumed for these images and compressive sensing applications investigated. Finally an application is done where a limited number of measurements are sampled from such an assumed sparse image and a reconstruction done with enlightening results, which can be implemented with great recommendation in future MRI data.

Nonparametric Bootstrap EWMA Control Chart

Presenter: Evert Coetzee, University of Pretoria

Co-author(s): Graham, M (Department of Statistics, University of Pretoria) and Kanfer, F (Department of Statistics, University of Pretoria)

We examine a bootstrap control limit design for the Exponential Weighted Moving Average (EWMA) control chart, obtained by performing a Phase I (retrospective or design phase) analysis and proceeded in performing the Phase II (the monitoring phase of a process) analysis. The control limits are obtained using the existing bootstrap (resampling methodology) proposed by Efron (1979), the residual bootstrap performed by Bühlmann (1997) and the moving blocks bootstrap proposed by Künsch (1989). The performance of these charts are measured by examining the in-control and out-of-control run-length distributions for several different process distributions. The chart's performance is also compared to the well-known methodology proposed by authors like Jones et al. (2002). We conclude with a summary and some results.

Time series analysis of South African gross domestic product

Presenter: Laruchelle de Almeida, University of Pretoria

Co-author(s): Van Niekerk, J (Department of Statistics, University of Pretoria)

To measure the performance of a country's economy it is preferred to use the gross domestic product (GDP) index. The analysis of GDP is carried out by adopting a relevant time series model. However, the stationarity of this model plays an important role in forecasting. For the purpose of identifying an accurate time series model to analyse the Real GDP of South Africa, we will be testing whether the time series model for the Real GDP is stationary, for the period of 19 years, i.e from 1995 to 2014.

Comparing two different sized photo-bio reactors using growth curves and bootstrapping

Presenter: Kirstie Eastwood, Nelson Mandela Metropolitan University

InnoVenton, a formally registered Research Institute at the Nelson Mandela Metropolitan University, is currently developing operational processes to increase the production of algae as part of their algae-to-energy project. This consists of two major components; converting algae into crude oil and using algae to turn coal dust (duff) into high quality, clean, usable coal. During an experiment, microalgae were grown in two different photo-bio reactors. The objective of this study was to find growth models which give adequate descriptions of the observed bio-densities during the period of measurement for both samples as well as to establish whether different diameter photo-bio reactors cause different growth rates. Four popular growth curve models were fitted to the data, namely the three parameter logistic model, four parameter logistic model, Gompertz model and Richard's Model. In both cases, Gompertz model was deemed the most appropriate. Regression bootstrapping techniques were applied and sampling distributions of the model parameters were approximated. Comparisons of the parameters of the two models were made using relevant statistical tests. The microalgae grown in the smaller diameter photo-bio reactor resulted in a higher growth rate.

A study of the moment generating functions of the generalised $\kappa - \mu$ and $\eta - \mu$ distributions in wireless systems

Presenter: Micaela Giacovazzi, University of Pretoria

Co-author(s): Ferreira, J.T.; Bekker, A., Department of Statistics, University of Pretoria.

In generalised fading models, the $\kappa - \mu$ and $\eta - \mu$ distribution is known for their encompassing nature, having many well-known distributions as special cases. In this study, the $\kappa - \mu$ and $\eta - \mu$ distribution is investigated, taking a particular interest in their moment generating functions (mgf) and the derivation thereof in closed form. The use of the mgf in the calculation of the average bit error rate (a popular performance metric in fading models) is highlighted, with emphasis on the ease of computation with these closed form mgfs.

Bayesian accelerated life testing for the exponential model using the MDI prior

Presenter: Sharkay Izally, Rhodes University

Co-author(s): Raubenheimer, L (Department of Statistics, Rhodes University)

Adekpedjou, A (Department of Mathematics and Statistics, Missouri University of Science and Technology)

Reliability life testing is used for life data analysis in which samples are tested under normal conditions to obtain failure time data for reliability assessment. It can be costly and time consuming to obtain failure time data under normal operating conditions if the mean time to failure of a product is long. The alternative is to use failure time data from an accelerated life test (ALT) to extrapolate the reliability under normal conditions. In ALT, the units are placed under a single higher than normal stress condition such as voltage, current, pressure, temperature, etc., to make the items fail in a shorter period of time. The failure information is then transformed through an accelerated model to predict the reliability under normal operating conditions. In this paper, we will develop a Bayesian inference model under the assumption that the underlying life distribution in the ALT is exponentially distributed. The maximal data

information (MDI) prior will be derived using a commonly used accelerated model known as the power law. The power law model is typically used for non-thermal accelerated stresses. Results

obtained when using the MDI prior will be compared to those obtained when using another non-informative prior. As a result of using a time transformation function, Bayesian inference becomes analytically intractable and so Markov Chain Monte Carlo (MCMC) methods will be used to alleviate this problem.

A rating system for rugby teams from multiple leagues

Presenter: Rion Jansen, University of Pretoria

Co-author(s): van Staden, P; Fabris-Rotelli, I; Vanter, M ((University of Pretoria)

No current system exists to rank rugby teams across multiple leagues. In this report a ranking system to rate rugby teams in multiple rugby leagues, with the intention to measure their relative strength toward each other, was set up. Applying it to past results to get a current rating for the rugby teams can also lead to predicting the winner of a match before the match is played. This system will be applied on past results for teams from three different rugby leagues. An interactive and automated program was developed in SAS/IML for this purpose. A sensitivity analysis was also conducted.

Spatial Econometrics

Presenter: Iketle Maharela, University of Pretoria

Spatial econometrics is a study that merges the field of spatial statistics and econometrics. It provides methods and techniques that acknowledge spatial dependence amongst observations with spatial properties. These methods are necessary as spatial data violates the basic assumption of independence amidst observations. The aim is to demonstrate how spatial models are fitted using different estimation methods.

Tests for Complete Spatial Randomness

Presenter: Francois Meintjes, University of Pretoria

Co-author(s): Co-Author: Fabris-Rotelli, I (Department of Statistics, University of Pretoria)

Spatial statistics is one of the most up and coming areas in statistics which is easier now to consider then years back due to the variety of methods for testing for randomness on some point pattern. In this research report, we explain the main theory and background behind spatial point patterns and discuss the different tests that can be applied to test for spatial randomness. Furthermore, we apply these tests to a certain point pattern obtained from the pulses of the Discrete Pulse Transform and reach a conclusion that our point process is indeed a regular point pattern. Lastly, we will give some conclusion for spatial point patterns in general.

Generalised Multivaraite Beta Type II Distribution

Presenter: Albert Mijburgh, University of Pretoria

Co-author(s): Bekker, A (Department of Statistics, University of Pretoria) and Human, S (Department of Statistics, University of Pretoria)

An exact closed-form expression of the joint probability density function (p.d.f.) of ratios of independent (but not identically distributed) gamma variables is derived. The components of this new multivariate distribution originate from a Statistical Process Control environment when using a change-point formulation to detect a sustained upward step shift in the variance of a normal distribution or the location of an exponential distribution. This new multivariate distribution extends the work of Adamski et al. (2013) and provides an alternative test statistic for detecting a change-point. In this paper we specifically focus on the bi-variate case and do the following: (i) investigate the statistical properties such as the moments and shape of the joint, the marginal and the conditional distributions; (ii) show the relationship between the new distribution and some other well-known bi-variate distributions with bounded and unbounded domain; and (iii) compare the power of the proposed and existing test statistics (used in the change-point setting) using computer simulation.

Detecting and Analysing Financial Cycles

Presenter: Shezad Muttur, NMMU

Co-author(s): Litvine, I (Department of Statistics, Nelson Mandela Metropolitan University)

Financial markets are often thought of as ‘unreal’ because most trades occur electronically without any tangible exchange. However the 2008 financial crisis is a prime example of how financial markets affect the ‘real’ world. The failure to predict the significant recessions has ignited debates about how financial cycles affect business cycles and whether we can use them to forecast booms and busts. This project devises a way to model cyclical behaviour in South African share prices using a dating algorithm called BBQ based on Pagan and Sossunov (2003) with modifications from Harding (2008). The algorithm detects peaks and troughs and forces alternation between the two to form cycles. We then analyse the characteristics of each cycle obtained, namely, Duration (D), Amplitude (A) and Change (C) using time series methods. This analysis reveals auto-correlations and/or non-stationarity in the series which then allows to determine trends and forecast characteristics of future cycles.

Harding, D. (2008). Detecting and forecasting business cycle turning points. University Library of Munich, Germany.

Pagan, A. and Sossounov, K. (2003). A simple framework for analysing bull and bear markets. J. Appl. Econ., 18(1), pp.23-46

Using linear programming to allocate swimmers to relay teams

Presenter: Mbongeni Mzila, University of Pretoria

Co-author(s): Fabris-Rotelli I (Department of Statistics, University of Pretoria), Van Staden P.J (Department of Statistics, University of Pretoria), Venter M (Department of Insurance and Actuarial Science, University of Pretoria)

In Masters swimming constraints such as different age groups, maximum number of events an individual can swim in, and the large number of swimmers available in the squad, leads to complications in selecting optimal relay teams for maximum point scoring. The aim of this research is to provide a solution to allocating swimmers to relay teams using linear programming, thereby maximizing points scored by a squad in a competition. This will be done by minimizing the time differences between the South African swimming records and a local swimming club in Pretoria, South Africa.

Subjective Bayesian analysis of the univariate normal model.

Presenter: Priyanka Nagar, University of Pretoria

Co-author(s): Van Niekerk, J (Department of Statistics, University of Pretoria)

The normal model is widely used in modern statistical modeling and hence the estimation of the parameters are very important. This study produces subjective Bayesian estimators under a normal-inverse gamma prior and a normal-gamma prior and LINEX loss function. It is shown that the normal-gamma prior results in estimators with less error than the well-known inverse gamma prior as well as the MLE's with a simulation study. The analytical expressions of the estimators are used instead of the MCMC sampling.

Modeling Of Road Traffic Fatalities In Namibia: Generalized Linear Model Approach

Presenter: Bertha Nambahu, University of Namibia

Co-author(s): Pazvakawambwa, L. (Department of Statistics and Population Studies, University of Namibia), Neema, I. (Namibia Statistical Agency)

Road traffic networks are key economic drivers in today`s world. They provide a quick, reliable and flexible transportation system, for people, goods and services. Namibia is one of the developing Southern - African countries, where road traffic accidents happen almost every day claiming the lives of many Namibians. Financial implication due to fatalities and injuries caused by these accidents has a tremendous impact on social well-being and socio economic development.Understanding to what extend each factor contributes to the severity of an injury or fatality is one of the effective means to improve road safety. This study assesses and models factors that contribute to road traffic fatalities in Namibia by exploring various count regression models(Poisson, Negative Binomial and Zero Inflated Poisson, Zero Inflated Negative Binomial, Hurdle Poisson and Hurdle Negative Binomial) and adjudicating them on the basis of the MSE and AIC.

A smooth transition autoregressive (STAR) time series model for the South African inflation rate

Presenter: Ané Neethling, Department of Statistics, University of Pretoria

Co-author(s): van Staden, Dr PJ (Department of Statistics, University of Pretoria)

Nonlinear time series models have become popular in the analysis of economic and financial time series, specifically for data with multiple regimes. The smooth transition autoregressive (STAR) time series model is a nonlinear time series model which allows for a smooth transition between two regimes through the use of a logistic or exponential transition function. In order to analyse the business cycles and economic behaviour in this paper, a logistic STAR (LSTAR) model is fitted to the seasonally unadjusted monthly inflation rate for South Africa from January 1969 to July 2015.

Comparison of image metrics for greyscale image segementation

Presenter: Christine Papavarnavas, University of Pretoria

Co-author(s): Fabris-Rotelli, I (Department of Statistics, University of Pretoria)

This report outlines image processing techniques for image comparison which provides effective approximations between the true/original image and a processed image. The development and improvement of quality assessment techniques that attempt to replicate the characteristics of the human visual system is essential for the field of image processing.

Statistical Modelling for Unplanned Capacity Loss in Electricity Generation

Presenter: Emma Plumstead, Nelson Mandela Metropolitan University

Co-author(s): Prof Litvine, I (Department of Statistics, Nelson Mandela Metropolitan University)

The main electricity distributor of South Africa, Eskom, currently faces challenges with capacity loss due to aged equipment and increasing demand for electric power. This study focuses on unplanned capacity loss, which is a result from manually reducing the output, or by the shutting down of a generator when a reading from a SCADA sensor hits a cut-off point. These failures are highly undesirable as they result in a substantial reduction in electricity energy output. The purpose of this study is to investigate the possibility of predicting these approaching failures by modelling the data from previous failures, so that preventative measures are taken before a failure becomes imminent. The utilization of a successful model may reduce the impact of unplanned capacity loss.

Analysis of JSE Stock Prices Using Hurst Exponent

Presenter: Sihle Poswayo, Nelson Mandela Metropolitan University

Co-author(s): Litvine, I (Department of Statistics, Nelson Mandela Metropolitan University)

In this project, we analyse stock prices of the Johannesburg Stock Exchange (JSE) using Hurst Exponent method. Statistical modelling and analysis of financial time series always attracted attention of prominent statisticians and econometricians. Different schools exist which promote various philosophies for such modelling. Particularly, two competing schools are based respectively on the following assumptions: (a) efficient market hypothesis and (b) long memory financial series.

We utilise the Detrended Fluctuation Analysis (DFA) approach on price series of different companies that are listed on the Johannesburg Stock Exchange. We estimate the Hurst exponent using daily closing prices data for the 10-year period from January 2005 to January 2015 and we discuss the surprising results which uncover whether the stock market of JSE is efficient or it follows a long memory process.

Spatial modelling of peak ground acceleration in South Africa

Presenter: Hayley Reynolds, University of Pretoria

Co-author(s): Loots, T; Kijko, A; Smit, A (university of Pretoria)

Spatial statistics involves data whose location plays a significant role in the characteristics of the observations. These observations, which are subject to random influence, have an additional variable, location, which tells the reader exactly where the observation occurred. Geostatistics is most well-known for its application of spatial interpolation in geosciences; predicting values at specific locations for which no observations have been recorded. Emphasis is placed specifically on the spatial interpolation method known as Kriging which calculates estimates and develops graphs to provide more insight into what can be expected at a location based on the values of neighbouring observations. Peak ground acceleration (PGA) is defined as the maximum acceleration amplitude measure of ground motion vibrations of an earthquake. This report uses spatial interpolation to generate a continuous spatial seismic hazard map for South Africa. Following the steps of the Kriging process resulted in a smooth contour plot of point measurements of estimated PGA. From these plots, PGA is expected to be high in the Western Cape, KwaZulu-Natal and the area known as the Witwatersrand Basin. Further research can be done to determine why this is so.

Mixtures of gamma distributions to model the signal-to-noise ratio of wireless channels

Presenter: Brett Rowland, University of Pretoria

Co-author(s): Coauthors: Ferreira, J.T. & Bekker, A., Department of Statistics, University of Pretoria, Pretoria

In the current digital realm, modeling digital communication and wireless channels and investigating the performance thereof is of high importance. A variety of models are available to model wireless channels and some key characteristics thereof - however, some of the characteristics and performance measures associated with these models have clumsy analytical expressions and are cumbersome to compute. In this study, the mixture gamma (MG) distribution is considered as a approximating model for the signal-to-noise ratio of some specific composite wireless channels. A numerical simulation and performance analysis is carried out to identify the accuracy and suitability of the proposed MG models as an approximation of the SNR distributions of the Nakagami-lognormal (NL) and Generalised K (KG) channels, and the advantages of the use of the MG distribution is highlighted.

Quantifying aggregation and zero inflation in faecal egg counts of sheep and goats.

Presenter: Phuti Sebatjane, University of South Africa

Co-author(s): Njuho, P (Department of Statistics, University of South Africa)

In modelling of stochastic variation in count data, the negative binomial distribution is most commonly used as an alternative to the Poisson distribution in the event of extra variation that cannot be accounted for by the latter. In counts of rare species however, the high proportion of zeroes result not only in overdispersion but also possible zero inflation. In this study we characterize both the aggregation and zero inflation for egg counts of 15 most common internal parasites in sheep and goats. To characterize aggregation, two aggregation measures; the variance to mean ratio and index of discrepancy, are computed and compared with the dispersion parameter from the negative binomial and the zero inflated negative binomial distribution. To characterize zero inflation, standard count models are fitted together with zero inflated models to each individual data-set. The zero inflated probability is then estimated under different covariate structures and different distributional assumptions. The index of discrepancy is found to be a better measure of aggregation only in the event of overdispersion. Both the dispersion parameter and the zero inflation probability are found to vary widely with covariate structure and distributional assumptions.

Statistical Robotics

Presenter: Prenil Sewmohan, University of Pretoria

Co-author(s): Fabris-Rotelli, I; Kanfer, F; Millard, S (University of Pretoria)

Abstract This report outlines the key concepts in robotics with respect to statistical theory. It focuses on the importance of stochastic and statistical methods in robot programming, processing and perception. The premise is that integrating statistical methods into programming robotics results in robots which have a higher degree of intelligence. There are various different opinions on what constitutes intelligence in robotics. The Florida Institute for Human and Machine Cognition defines artificial intelligence as "the ability of a system to act appropriately in an uncertain environment where an appropriate action is that which increases the probability of success." It will be with a similar criteria for intelligence that this paper assesses the role of statistical programming in robotics. This will be done with specific reference to state estimation techniques, using information filters and the localization problem. The aim is to set out the basic terminology and theory behind programming a robot statistically, while also programming a robot to perform some basic task.Then finally to grab data from the completion of this task and analyse it with the tools and theory previously examined in an attempt to practically illustrate the theory by improving the initial task.

A structural equation modelling (SEM) analysis of a four factor model with demographic influences

Presenter: Carmen Stindt, Nelson Mandela Metropolitan University

Co-author(s): Clohessy, CM (Department of Statistics, Nelson Mandela Metropolitan University) and Sharp, GD (Department of Statistics, Nelson Mandela Metropolitan University)

The analysis of data generated in the social environment is never easy. In the physical sciences, experimental results coincide with mathematical theory whilst social sciences are influenced by individual personalities. This study analyses data in a social environment using structural equation modelling (SEM) and reports on the results of the analysis. In addition, the researcher will report on the frustrations and confusion experienced having ventured into the social science analytical domain.

The synchronization between stock prices in JSE and related commodities

Presenter: Kylie Tarboton, NMMU

Co-author(s): Litvine, I (Department of Statistics, NMMU)

Both investors and policymakers in South Africa are interested in the relationships between stocks' and commodities' prices. An understanding of this relationship will help to formulate an effective response. Don Harding and Adrian Pagan suggested studying the associations between prices using cycles in the time series. The methods suggested allow revealing if synchronization of cycles is present. Cycles are first identified and then the information in the data is translated to binary variables. Tests for synchronization are performed between JSE listed stocks and commodities that are expected to be related.

Modeling Length of Hospital Stay for Tuberculosis In-Patients at Queen Elizabeth Central Hospital: Applying Competing risks

Presenter: Halima Twabi, Chancellor College

Co-author(s): Dr M Mukaka (Department of Statistics, University of Oxford (UK), Mahidol-Oxford Tropical) , Dr J.J. Namangale (Department of Mathematical Sciences, Chancellor College)

A retrospective cohort study was used on adult TB in-patients from Queen Elizabeth Central Hospital (QECH) SPINE database to identify factors explaining time to discharge from hospital while accounting for a competing event; death. The study aimed to apply and compare estimates of competing risk models on TB data that collected patient's socio-demographic characteristics and patient's medical information. Semi-parametric Cause-specific hazards were used to model the effect of HIV status, ART Status, age, and Sex in relation to death or discharge from hospital. The Fine and Gray regression estimates were compared to the cause-specific estimates. Test for model assumptions and diagnostics were conducted. Findings showed that the Fine and Gray regression explained best the effect of the covariates to the probability of a patient being discharged or dying. Further the main factors affecting length of hospital stay among TB in-patients were age and HIV Status. HIV positive patients were 17.6 \% less likely to be discharged from hospital compared to HIV negative patients ($p=0.048$) and with an increase in age, the hazard of discharge decreased by 2% ($p < 0.001$). It is important to present results on both the event of interest and the competing risk and use the cumulative Incidence function for calculating probability of an event. Competing risks data should be modeled using both the Cause Specific Hazard model and the Fine and Gray model when studying length of hospital stay.

Presenter: Carl van Heerden, North-West University, Potchefstroom Campus

Co-author(s): Jansen van Rensburg, H. (Department of Statistics, North-West University, Potchefstroom Campus)

The 2015 Graduate Destination Survey is the first survey of its kind for the North-West University (NWU), contributing to the development and implementation of a strategy to promote the career prospects of NWU graduates. The purpose of the study was to provide feedback on employment trends of NWU graduates and identify improvement possibilities in the University's education system. Students from all three campuses of the University who completed their degrees in 2014 were identified as the target group for this survey. Various categorical data analysis and modelling techniques were applied to the dataset comprising a total of 1,077 survey responses.

Application of Mixture models for Eland movement in two Eastern Cape National parks

Presenter: Bracken van Niekerk, NMMU

Co-author(s): Goodall, V (Department of Statistics, Nelson Mandela Metropolitan University)

Independent Mixture models and Hidden Markov models have been used to model the movement patterns of a variety of species of animals in many different environments. Unlike the Independent Mixture models, the Hidden Markov models take the serial correlation between successive observations into account. We investigated whether these models can differentiate movement patterns for Eland in two different regions. The models are fitted to Eland in the Nyathi region of the Greater Addo Elephant Park and Mountain Zebra National Park, in the Eastern Cape. Hourly data was used to fit models from which to infer the animals' behavioural states from the latent states of the models. The data spanned for about a year, which covers the four seasons. Results identified different movement strategies of the Eland in the two parks, particular in terms of the most active movement behaviours. The animals' behaviour is strongly influenced by the availability of food and water. These necessities influence the movement patterns and the models were able to identify the different behavioural strategies of the animals in the two parks.

Application of Mixture models for Eland movement in two Eastern Cape National parks

Presenter: Bracken van Niekerk, Nelson Mandela Metropolitan University

Co-author(s): Goodall,V (Department of Statistics, Nelson Mandela Metropolitan University)

Independent Mixture models and Hidden Markov models have been used to model the movement patterns of a variety of species of animals in many different environments. Unlike the Independent Mixture models, the Hidden Markov models take the serial correlation between successive observations into account. We investigated whether these models can differentiate movement patterns for Eland in two different regions. The models are fitted to Eland in the Nyathi region of the

Greater Addo Elephant Park and Mountain Zebra National Park, in the Eastern Cape. Hourly data was used to fit models from which to infer the animals' behavioural states from the latent states of the models. The data spanned for about a year, which covers the four seasons. Results identified different movement strategies of the Eland in the two parks, particular in terms of the most active movement behaviours. The animals' behaviour is strongly influenced by the availability of food and water. These necessities influence the movement patterns and the models were able to identify the different behavioural strategies of the animals in the two parks.

Asymmetric generalizations of the logistic distribution

Presenter: Anika Wessels, University of Pretoria

Co-author(s): Van Staden, P.J (Department of Statistics, University of Pretoria) and Omachar, B.V (Department of Statistics, University of Pretoria)

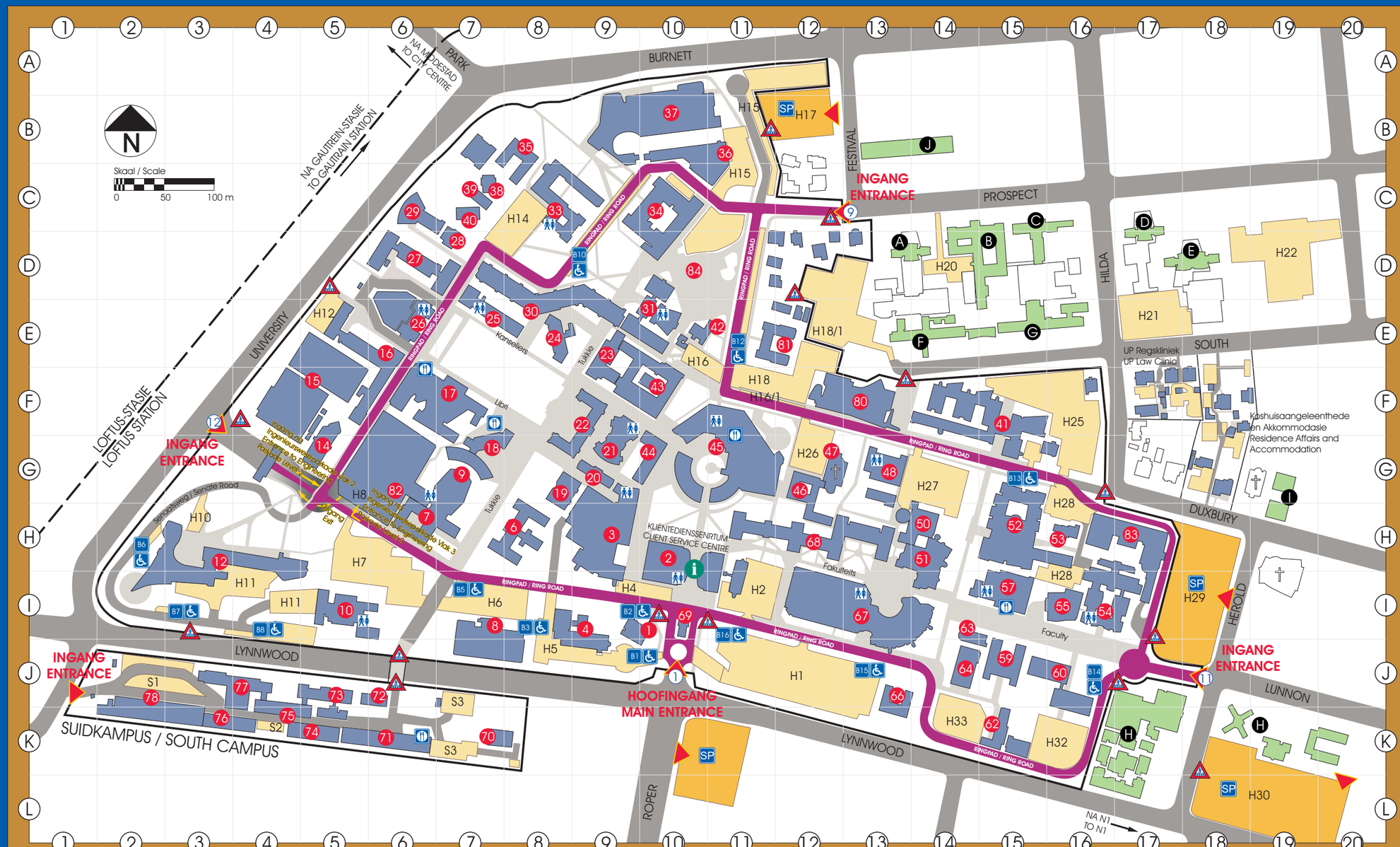
Because the logistic distribution possesses simple expressions for its density, distribution and quantile functions, it has been used extensively in theoretical development and in practical applications. In particular, in distribution theory, various generalizations of the logistic distribution have been developed and proposed in the literature. This paper investigates the flexibility in distributional shape of five asymmetric generalizations, namely the density-based and the quantile-based skew logistic distributions, the Type I and the Type II generalized logistic distributions, and Hosking's generalization of the logistic distribution, which is a reparametrized version of the log-logistic distribution.

Synchronization and conformity in random systems: the hipster effect

Presenter: Keunyoung Yoo, University of Pretoria

In this paper a model of predicting trend that incorporates information delay is investigated as opposed to a Markov chain approach of trend prediction (which does not take information delay into account). This paper will also explain why and how the new model can give us more insight to the problem and possible applications of the model will also be discussed.

Hatfieldkampus / Hatfield Campus / Khamphase ya Hatfield S 25° 45' 21" E 28° 13' 51"



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

- UP-geboue / UP Buildings
- UP-koshuise / Akkommodasie / UP Residences / Accommodation
- Ander Geboue / Other Buildings
- Ringpad / Ring Road
- Voetpaaie / Walkways
- Paaie / Roads
- Parkering / Parking
- Besoekersparkering / Visitors' Parking
- Studenteparkering / Student Parking
- Parkering vir Gestremdes / Parking for Disabled
- Voertuigtoegang / Vehicle Access
- Voertuig-heknommers / Vehicle Gate Numbers
- Voetgangertoegang / Pedestrian Entry
- UP-inligtingstoonbank / UP Information Desk
- Publieke Toilette / Public Toilets
- Kleinhandel- en Eetfasiliteite / Retail and Dining Facilities

Kashuisaangeleenthede en Akkommodasie
Residence Affairs and Accommodation

UP Regskliniek
UP Law Clinic

GEBOUE																							
12	Administrasiegebou	H3	66	Dramagebou	J13	71	Gebou 7 (Suidkampus)	K6	2	Kliëntedienssentrum (KDS)	H10	3	Merensky 2-biblioteek	H9	19	Ou Merenskybiblioteek	G8	37	Tegniese Dienste-gebou	B10	PLEKKE VAN BELANG		
29	AE du Toit-ouditorium en -anneks	C6	80	Eeufeesgebou	F13	72	Gebou 8 (Suidkampus)	J6	48	Kloostersaal	G13	27	Mineraalwetenskappegebou	D6	35	Plantkundegebou	B8	6	Teologiegebou	H8	35	Botaniese Tuin	B8	B	Vergeet-my-nie	D15
9	Amfiteater en Musaion	G7	68	Ekonomiese en Bestuurswetenskappe-gebou	H12	70	Gebou 9 (Suidkampus)	K7	18	Klubsaalgebou	G7	21	Mullersaal (Kanseliersgebou)	G9	83	Plantwetenskappe-kompleks	H17	21	Te Water-saal (Kanseliersgebou)	G9	30	Camera Obscura	E8	C	Asterhof	C15
17	Aula en Rautenbachsaal	F7	54	FABI 1	I16	2	Geesteswetenskappegebou	H10	2	Kommunikasiepatologiegebou	I9	4	Musaion en Amfiteater	G7	84	Thuto-gebou	D10	19	Eduardo Villa-museum	G8	D	Madelief	C17
38	Batemangebou	C7	55	FABI 2	I16	23	Geografiegebou	E9	50	Konferensiesentrum	H14	9	Musiekgebou	H6	53	Toerismehuis	H16	27	Geologiemuseum	D6	E	Magrietjie	D18
20	Bemarkingsdienstegebou	G9	57	Huis en Haard	I15	1	Kya Rosa	I10	52	Nagraadse Sentrum	H15	46	Tukkiewerf	G12	1	Kya Rosa	I10	F	Klaradyn	E14			
69	Besoekersontvangs	I10	26	Ingenieurswese 1-gebou	E6	59	Landbou-anneks	J15	30	Natuurwetenskappe 1-gebou	E8	47	Universiteitskapel	G12	22	Mapungubwe-versameling	F9	G	Jasmyn	E15			
8	Boukundegebou	I7	16	Ingenieurswese 2-gebou	E6	60	Landbouwetenskappegebou	J16	25	Natuurwetenskappe 2-gebou	E7	42	Studentegesondheidsdienste	E11	36	Sci-Enza-sentrum	B11	H	Nerina	K17			
14	CEFIM-gebou	G5	82	Ingenieurswese 3-gebou	G6	51	Lesingsale (E&BW)	H14	31	Noordsaal (Chemiegebou)	E10	21	Van der Bijlsaal (Kanseliersgebou)	G9	47	Universiteitskapel	G12	I	Protea	G19			
31	Chemiegebou	E10	67	Inligtingtegnologiegebou	I13	62	Lier-teater	K15	44	Ou Chemiegebou	G10	28	Van der Graaf-versneller	D7	22	Van Tilburg-versameling	F9	J	Jakaranda	B14			
43	Dierkundegebou	F10	74	Gebou 5 (Suidkampus)	K5	63	JJ Theron-lesingsaal	I14	21	Louwsaal (Kanseliersgebou)	G9	39	Vetmangebou	C7	KOSHUISE / AKKOMMODASIE										
			73	Gebou 6 (Suidkampus)	J5	64	Maskerteater	J14	34	Ou Landbougebou	C10	24	Studentesakegebou	E8	A	Erika	D13							
									22	Ou Letteregebou	F9	45	Studentesentrum	G11									
												31	Suidsaal (Chemiegebou)	E10									
												15	Swaarmasjienlaboratoriums	F5									
												33	Wiskundegebou	C8									

UP-NOODNOMMER: 012 420 2310

BUILDINGS																							
12	Administration Building	H3	75	Building 4 (South Campus)	K4	18	Club Hall	G7	52	Graduate Centre	H15	34	Masker Theatre	J14	22	Old Arts Building	F9	21	Te Water Hall (Chancellor's Building)	G9	PLACES OF INTEREST		
29	AE du Toit Auditorium and Annexe	C6	74	Building 5 (South Campus)	K5	4	Communication Pathology Building	I9	15	Heavy Machinery Laboratories	F5	33	Mathematics Building	C8	44	Old Chemistry Building	G10	6	Theology Building	H8	35	Botanical Garden	B8
59	Agriculture Annexe	J15	73	Building 6 (South Campus)	J5	50	Conference Centre	H14	57	Huis en Haard	I15	27	Mineral Sciences Building	D6	19	Old Merensky Library	G8	84	Thuto Building	D10	30	Camera Obscura	E8
60	Agricultural Sciences Building	J16	71	Building 7 (South Campus)	K6	66	Drama Building	J13	2	Humanities Building	H10	3	Merensky 2 Library	H9	83	Plant Sciences Complex	H17	53	Tourism House	H16	19	Eduardo Villa Museum	G8
9	Amphitheatre and Musaion	G7	70	Building 9 (South Campus)	K7	68	Economic and Management Sciences Building	H12	67	Information Technology Building	I13	48	Monastery Hall	G13	21	Roos Hall (Chancellor's Building)	G9	46	Tukkiewerf	G12	27	Geology Museum	D6
17	Aula and Rautenbach Hall	F7	8	Building Sciences Building	I7	26	Engineering 1 Building	E6	63	JJ Theron Lecture Hall	I14	21	Muller Hall (Chancellor's Building)	G9	81	Roosmaryn	E12	47	University Chapel	G12	1	Kya Rosa	I10
38	Bateman Building	C7	14	CEFIM Building	G5	16	Engineering 2 Building	E6	1	Kya Rosa	I10	9	Musaion and Amphitheatre	G7	31	South Hall (Chemistry Building)	E10	21	Van der Bijl Hall (Chancellor's Building)	G9	22	Mapungubwe Collection	F9
35	Botany Building	B8	80	Centenary Building	F13	82	Engineering 3 Building	G6	41	Law Building	F15	7	Music Building	H6	40	Stoneman Building	C7	28	Van der Graaf Accelerator	D7	36	Sci-Enza Centre	B11
78	Building 1 (South Campus)	J2	21	Chancellor's Building	G9	54	FABI 1	I16	51	Lecture Halls (E&MS)	H14	30	Natural Sciences 1 Building	E8	24	Student Affairs Building	E8	39	Vetman Building	C7	47	University Chapel	G12
77	Building 2 (South Campus)	J4	31	Chemistry Building	E10	55	FABI 2	I16	62	Lier Theatre	K15	25	Natural Sciences 2 Building	E7	45	Student Centre Building	G11	69	Visitors' Reception	I10	22	Van Tilburg Collection	F9
76	Building 3 (South Campus)	K3	2	Client Service Centre (CSC)	H10	23	Geography Building	E9	21	Louw Hall (Chancellor's Building)	G9	31	North Hall (Chemistry Building)	E10	42	Student Health Services	E11	10	Visual Arts Building	I5	RESIDENCES / ACCOMMODATION		
									20	Marketing Services Building	G9	34	Old Agriculture Building	C10	37	Technical Services Building	B10	43	Zoology Building	F10	A	Erika	D13

UP EMERGENCY NUMBER: 012 420 2310

Analytics in Action

Unlock value in your data reservoirs to optimize ROI.

Now more than ever, SAS, the leader in predictive analytics,[1] gives you the power to uncover hidden data insights that will improve your operations and optimize your bottom line.

SAS® Data Management and advanced analytics provide capabilities[2] that span upstream, midstream and downstream segments to convert data into assets that exploit conventional and unconventional resources. Our analytics reduces non-productive time, optimizes return on asset investment as well as forecasts and manages the impact of supply and demand trends on your business.

Learn more
sas.com/oilgas

For more information please contact us.

Website: www.sas.com/sa
Email: intouch@zaf.sas.com
Tel: +27 11 713 3400

Follow us on

@SAS_SouthAfrica SAS Southern Africa

sas
THE POWER TO KNOW.

[1] Magic Quadrant for Business Intelligence and Analytics Platforms.
[2] IDC Worldwide Business Analytics 2014 – 2018 Forecast.