Professional Data Scientists:
Who are they and how do we train them?

by Riaan de Jongh
Centre for Business Mathematics and Informatics®, North-West University.

What is Data Science?

Data science is a multi-disciplinary field consisting of a number of disciplines (e.g. applied mathematics, statistics, machine learning, operations research, artificial intelligence). It is used to solve problems in various application areas, for example health sciences, astrophysics, agriculture, telecommunications and finance. Its primary aim is to extract insight from data in various forms, both structured and unstructured. At the core is so-called “big data” that are stored in various ways and exhibit complex many-to-many relationships, which is made more challenging by the ever-increasing requirement to process these in real time to support “instantaneous” decision-making. The rise of data science can largely be attributed to advances in computer technology and processing speed, low cost storage of data, and the massive availability of data from the Internet and other sources. The access to big data and the advances in computer technology make possible the renewed application of machine learning and statistical techniques on problems, reporting huge successes in a wide range of applications. One of the subfields of data science is statistics, a branch of mathematics dealing with the collection, analysis and interpretation of data. Statistics have established itself firmly as an academic discipline and has been in existence since the eighteenth century. Because of the big data explosion, a number of the classical statistical approaches that perform reasonably well for small datasets fail when dealing with huge datasets. Despite this, many more recently developed statistical techniques are used successfully in a big data context. Examples are logistic regression, cluster analysis, and decision trees. Machine learning and artificial intelligence are relatively new subfields of data science and concentrate on brute force computer power and complex optimisation algorithms to solve real-time prediction problems. Examples are the successful applications of neural networks and deep learning in the area of speech recognition and language processing (e.g. Siri and Google Assistant). It should be noted that machine learning is frequently concerned with prediction tasks and models in this context (e.g. recommender systems and factorisation machines). Unlike statistics, machine learning is not concerned with traditional aspects of statistical inference (e.g. about the significance of the estimates of model parameters). Although statistics and machine learning are different disciplines, there is some overlap, for example a technique like random forests are frequently quoted in both fields. A reader of the literature in both fields will quickly realise a difference in the terminology used for similar concepts. It is
interesting to note that in a recent paper, a prominent researcher at Harvard University (Meng, 2018) warned about the big data paradox, i.e. he emphasised that data quality plays an enormous role and that having more data will fool us when making population inferences in a big data context.

Another subfield of data science is operations research (OR), which became popular in the early eighties. Spurned by the advent and wider availability of personal computers, OR, like data science now, was all about using the mathematical and computing sciences to solve real-world problems in a multi- and interdisciplinary way.

**What is a data scientist?**

Because data science is so wide in scope, many professionals may claim that they are data scientists, e.g. statisticians, operations researchers, engineers, computer scientists, actuaries, physicists and machine learners. From my own practical experience, it is clear that when solving data science problems, you need a range of people of which some can work in depth on theory and others can tend to application. It is a way to attempt to cover the whole spectrum. When solving complex problems in data science, one person cannot handle all aspects, but it could possibly be achieved with a group of people. Currently the main focus of data scientists is to use innovative techniques emanating from the subfields to solve problems in a particular application area of interest. It should be noted that the application areas and therefore the type of problems encountered are very different, frequently necessitating a deep knowledge of the particular subject matter. For example, consider astrophysics and the squared kilometre array. Apparently these telescopes will receive data at one terabyte per second and researchers are typically interested in detecting tiny signals engulfed in white noise. On the other hand, in finance, amongst others, researchers exploit large data bases to learn more about the credit behaviour of customers.

**What can we learn from the established subfields of data science?**

As stated before, statistics and operations research are two of the oldest subfields in data science and many practicing statisticians and operations researchers consider themselves data scientists. What can we learn from these fields that could help us in training the data scientists of the future? Universities all over the world have largely failed to deliver professionally trained graduates in the fields of OR and statistics. Although well trained academically, many newly appointed graduates find it difficult to immediately add value at their place of employment. Typically they lack subject matter knowledge of the application field (e.g. finance or physics) and struggle with real-world problem solving abilities, such as the formulation of messy problems, meaningful interaction with clients, interrelationships with team members and business communication. Some students also lack numerical and data handling programming skills that are not addressed adequately in many curricula.

Some of the lessons I learnt in the many industry projects I have been involved in, include:

- Always focus on the business value throughout the course of the project.
- Involve all role-players and instil trust and confidence about your ability as a consultant.
- Manage the client’s expectations, communicate clearly and pay attention to fostering good interpersonal relationship skills.
- Test the client’s understanding of his/her own problem and educate the client when necessary.
- Be sure that the problem to be solved is well formulated, because you do not want to solve the wrong problem.
- Always be cognisant of the importance of simplicity and when your solution is very complicated, seek a simpler solution, if possible.
- Do not be fixated on new untested technologies.
- Solve the critical aspects that will determine eventual success, first.
- Always revisit the scope and risks of the project and plan properly.

How do we, however, train students to ensure that they become professional data scientists? This is not easy and will be addressed in the last section.

**What should we teach aspiring data scientists?**

From the above it should be clear that a training programme should include training in the following:

- **The mathematical and computational sciences**: Topics could include courses in statistical and probability theory, artificial intelligence, machine learning, operations research, and computer science.
- **Programming skills**: Numerical programming skills in languages such as SAS, R and Python.
• **Data management skills:** Topics should include data bases and warehousing that concentrate data manipulation and merging skills in languages such as SQL, SAS, R and Python.

• **Subject matter knowledge** in selected fields of application.

• **Professional problem-solving skills.**

Assuming a sound knowledge of undergraduate training in the mathematical and computer sciences, one could include the following topics in a graduate programme: generalised additive models; regularisation (lasso and elastic nets); model selection; time series analysis; multi-variate statistics; cluster analysis; optimisation; neural networks and deep learning; support vector and factorisation machines; event stream processing; text analytics; database handling and extraction. All of these courses should have a practical element, where the techniques are programmed in one of the above-mentioned programming languages and applied to data and problems in a relevant application. Depending on the application areas, suitable courses on the important concepts in these fields should be included. For example, in astrophysics, it might be necessary to include courses such as signal processing and pattern recognition and basic concepts in astrophysics. Similarly, if the application area is finance, courses could include scorecard model building, risk management and other important financial concepts (e.g. value-at-risk). It is of course, not practical to cater for all the fields and possible topics, if not impossible. At my university we have spread the programme over two years, where all the technical courses are covered in an honours degree and half of the masters’ degree. The remainder of the masters’ programme addresses the professional training aspects.

**Adding professionalism to the training programme**

Teaching students the problem-solving skills necessary for the industry is a real challenge. The instructor should facilitate a mind set change among students to ensure they focus on the importance of solving the business problem and not a statistical or mathematical sub-problem. More importantly, these courses should be taught by people with the necessary experience in solving problems in the particular application area (see e.g. Coetzer & de Jongh, 2016). This suggests that data science programmes comprising only academics with no experience in solving industrial or business problems will make it extremely difficult to equip data scientists with the requisite skills to function effectively in industry.

In our Masters programme we follow an integrated hands-on training approach in solving problems in the area of application. This is done in the form of on-site (at the client company) internships where a student is assigned to a specific problem posed by industry. The student has to complete the project over a six-month period with the assistance of an academic supervisor (responsible for academic quality) and a client project officer (responsible for business value add). Formal on-site project meetings are scheduled where all role players should be present to discuss progress. In this way the academic supervisors also gain industry experience and get a feel for the problems being faced by industry. This often leads to industry directed research projects by the supervisor for the company. The company has the benefit of screening the student for employment and potential problem-solving value add at relatively low cost. The demand for these students has increased dramatically, supported by the fact that project proposals outnumber available students 2:1. As a fringe benefit the programme has spurred a number of research imperatives between academia and industry and many papers have already been submitted, which have been co-authored by academics and practitioners. It should be noted that although student projects are classified confidential and although students have signed non-disclosure agreements with the assigned company, it is amazing how quickly client project officers can share sensitive information when they are offered co-authorship of a paper. Interestingly, but not surprisingly, alumni of this programme become future client project leaders and research collaborators.

Please see the references for more information about this programme.

**REFERENCES**


De Jongh, P.J.  2018. University-industry engagement in data science. [Video].

Editorial

Dear SASA member

Best wishes for 2019! Welcome to all new members!

We are all inspired after the successful 2018 SASA Conference at Florida Campus, Roodepoort hosted by the University of South Africa. Prof Riaan de Jongh of the BMI at the North West University addressed the opening session with a talk titled “Professional Statisticians or Data Scientists: Who are they and how do we train them?”

See photos from the conference in this issue. The Sichel medal was awarded, the 2018 Thought Leader is Prof Tertius de Wet, and the student prize winners received certificates. Congratulations to all members and students receiving awards and prizes. One of the overseas visitors, Prof Louise Ryan from Australia, shared her experience of the conference and visit to South Africa.

We are looking forward to the 2019 SASA Conference to be hosted by the Department of Statistics of the Nelson Mandela University in Port Elizabeth from 26-29 November 2019.

We welcome Maseka Lesaoana as the president of SASA, with Danie Uys the past president. Congratulations to Trudie Sandrock who will be the future Newsletter Editor, and to Humphrey Brydon who will take over the Young Statisticians portfolio from Jaco Visagie.

In this issue you can read about the World Statistics Day event hosted by the CSIR on 19 October 2018. The Education News and the seminar program for the first semester of 2019 from the Stellenbosch University is also included.

Proff Dawie Stoker and Michael Brown sadly passed away at the end of last year. Articles about these two statisticians are to be found in this issue.

Thank you for all the contributions received from you during the seven years that I was in the portfolio of Newsletter Editor! It was a great pleasure doing it.

Keep well.

Mardi

Welcome!

New members of SASA

Mr J Boana-Danquah, Tutor, Limpopo Department of Education, jerrykd1979@gmail.com
Prof DG Chen, Professor University of Pretoria/University of North Carolina at Chapel Hill, din.chen@up.ac.za
Dr M Naderi, Postdoctoral student, University of Pretoria, mehrdad.naderi@ymail.com
Dr E Nortey, Lecturer, University of Ghana, ennnortey@ug.edu.gh
Dr RE Ogunsakin, Postdoctoral student, University of Pretoria, oreropo@gmail.com
Ms H Pretorius, Senior Analyst, EOH Information Services, hedre.pretorius@gmail.com
Mr WJ Scholtz, Lecturer Akademia, willemj.scholtz@gmail.com
Ms L Uzoma, Consultant, University of the Witwatersrand, nnajilovelyn@gmail.com

URGENT DEADLINE

Call for nominations for visitors for the 2019 conference

Any SASA member wishing to nominate an overseas visitor for the 2019 SASA Conference, please contact the secretary of the Executive Committee with supporting information by 28 February 2019. The contact details for the secretary are:

The Secretary: SASA
P O Box 3341
Matieland
7602
email: Chantelle.Clohessy@mandela.ac.za

Announcement

The host of the SASA 2019 Conference will be the Nelson Mandela University. It will take place in Port Elizabeth from 26 - 29 November 2019.
The 60th conference SASA Conference was hosted by the University of South Africa from 26 - 29 November 2018 at the Florida Campus, Johannesburg. As always, the annual SASA conference brought together statisticians and researchers in relates fields from around the country and abroad to share their research, discuss new ideas and to meet and establish collaborations.

A special welcome was extended to postgraduate students who could take advantage of preferential registrations fees.
Poster competition

Best honours poster: Claire Bezuidenhout (UP) with “GARCH models for time series exhibiting volatility”.

Best masters poster: Christopher Dunderdale (NMU) with “Image classification in photovoltaic systems using machine learning techniques”.

Oral competition

Best talk (joint first prize): Aviwe Gquaka (NMU) with “Evaluation of selection tools for the inefficiency distribution in stochastic frontier models” and Andre Zitzke from SAS (photo left).

Best talk (joint first prize): Shawn Liebenberg (NWU) with “A goodness-of-fit test for the Raleigh distribution based on a lesser-known characterisation” and Andre Zitzke from SAS (photo far left).
This past November, I had the privilege to travel to Johannesburg to attend the annual meetings of the South African Statistical Association (SASA - https://sastat.org.za). This piece provides some of my reflections of the conference, mostly abstracted from my President’s Corner contribution to our latest Biometric Bulletin.

Being President of an International Society is exciting and challenging, but can be tiring due to the extensive travel involved. With South Africa being in the same hemisphere as Australia, I thought this might be an easier trip than usual. However, it still took me 14 hours to fly with Qantas who have a direct flight from Sydney to Johannesburg. Interestingly, it only took 11 hours to return home .... Perhaps we went via a convenient time warp? More likely it had to do with trade winds. I had been to South Africa once before, way back in 1998 when the International Biometric Conference was held at University of Cape Town. The world has changed a lot in twenty years so I was curious to return. To tell the truth, I was slightly nervous since one often hears that Johannesburg is not the safest of cities. But I had a great time.

I was very impressed with the conference in terms of how well it was organized, the quality of all the presentations and posters, and of the overall atmosphere. The conference was hosted by the Department of Statistics at the University of South Africa, under the very capable leadership of their Chair, Eeva Rapoo. It really was a lovely venue, with plenty of room for smaller and larger sessions and a fabulous open area for morning and afternoon teas and poster sessions. The photo above shows the main lecture theatre used for the plenary sessions – this was actually the very final session where student prizes were also awarded. As you can see, even in this not very good photo taken on my iPhone, the diversity of the group was particularly notable, with lots of young women, a wide range of ethnic and racial backgrounds and many different countries represented. At a time when so many places in the world seem embroiled in racial, ethnic and religious tensions, it was refreshing and in fact quite inspiring to see that such differences could be transcended through a shared interest in statistics and biometry. Just by throwing ourselves enthusiastically into the work of the IBS, we are all making a difference.

I was pretty busy at the conference, delivering a short-course (spatial statistics for health research) and a keynote address. I chose to talk about my work with the Knowledge Integration Project, funded through the Bill and Melinda Gates Foundation (http://kiglobalhealth.org), where my colleagues and I compare different approaches to longitudinal growth modelling based on a very rich data repository of data from more than twenty cohorts. I was pleased that my presentation seemed to be well received. It has been a challenging but exciting project. Anyone interested can read more in our paper that will shortly appear in Statistics in Medicine (you can see the online version at https://onlinelibrary.wiley.com/doi/full/10.1002/sim.7693). In response to inquiries from several of you, I have been
speaking to my colleagues at the Foundation about how people can get involved in the project.

I was pleased to catch up with lots of IBS members and officers from the various African regions, both informally as well as at a special IBS meeting held during the SASA conference. At that meeting, we had a chance to talk about some of the latest activities going on in IBS, especially efforts around revamping of the website. I reminded participants about the availability of funds to attend conferences in other regions (see https://www.biometricsociety.org/2018/11/ibs-travel-grant-application/) and encouraged people to consider applying. I also just tried to listen to local concerns and have some open dialogue.

I enjoyed the opportunity to talk with quite a few young people at the poster sessions. I particularly enjoyed talking to Lizalise Mngcele about his poster. As part of his work with the African Earth Observatory Network (AEON) in the Department of Geoscience at Nelson Mandela University, Lizalise used drone technology to capture data on the locations and sizes of anthills in the Karoo region of South Africa. He then used spatial statistics techniques to assess whether or not the anthills were located randomly. His hypothesis was that factors such as altitude and water affected the spatial patterns. It seemed to me that he was asking great questions and ones that could easily lead into some interesting methodological research in spatial statistics.

The conference offered two interesting excursion options in the middle. I chose to visit Soweto, a township just 20 minutes to the south of the University. We spent some time at the famous Hector Pieterson memorial that commemorates the role of South Africa’s young students in the struggle against apartheid back in the 1970s. It was a poignant experience, especially seeing so many stark images and reading about the death of the twelve year Hector who was shot by police whilst watching the older students march. Above is a picture of Freedom Gumede and myself just down the road outside Mandela House, which is where the Mandela family lived for many years. Freedom is a faculty member in the Department of Statistical Sciences at University of Cape Town and one of our IBS Executive Board Members. As you will have read elsewhere, Freedom was recently re-elected for an additional term. Congratulations Freedom!
Awards

THE HERBERT SICHEL AWARD FOR 2018

This year’s award of the Herbert Sichel Medal represents the twenty first since it was instituted by the Association in 1997 in memory of Professor Herbert Sichel, one of South Africa’s statistical pioneers, who made seminal contributions to Statistics in fields ranging from mining to statistical linguistics, and including a distribution named after him. Except for one year (2001), the medal has been awarded annually to a member (or members) of the Association whose paper appeared (in print or online) during the previous calendar year and was judged best by a panel of judges appointed by the Association. A list of Winners of the Award since its inception in 1997 appears on the SASA website.

Six papers were submitted for this year’s award and the panel of judges, comprising Maksim Finkelshtein (University of the Free State), Linda Haines (University of Cape Town), Tertius de Wet (University of Stellenbosch) and Paul Fatti (Wits University), used the four criteria of Impact, Innovation, Relevance and Quality of Writing to rank them. The high quality of all the submissions made the judges’ task difficult, but while the judges differed in their rankings, there was unanimous agreement about the winning paper. We would like to congratulate Ansie Smit, Andrzej Kijko and Alfred Stein on their excellent paper entitled: “Probabilistic Tsunami Hazard Assessment from Incomplete and Uncertain Historical Catalogues with Applications to Tsunamigenic Regions in the Pacific Ocean”, which appeared in the Journal of Pure and Applied Geophysics.

Congratulations to Prof Hoffie Lemmer and Prof Niel le Roux for being elected as Honorary members of SASA.

Fellow Membership awarded to Dr Roelof Coetzer (left)

SAS Thought Leader award presented to Prof Tertius de Wet (left)

CONGRATULATIONS!

Ansie Smit receiving the Herbert Sichel award from Paul Fatti.
David Johannes Stoker was born at Potchefstroom in April, 1929. On 1 February 1955 Dawie joined the Department of Statistics at the University of Pretoria as senior lecturer. In December 1955 he married Wilmien Naude (from Lichtenburg). After the death of Prof. B de Loor in February 1962 he was promoted to Professor and Head of the Department. He served in this capacity until 31 December 1980. In 1956 the two-year major course in Mathematical Statistics was introduced at UP and in 1967 the three-year major course for the BSc- and BA-degrees.

Dawie was awarded overseas bursaries for doctoral study in Mathematical Statistics. On the advice of Prof. B de Loor (UP) and Prof. S. Pretorius (US) he decided to enroll at the University of Amsterdam, The Netherlands, under Prof. David van Dantzig. At the same time he acquired practical experience in the application of Statistics at the Mathematical Center in Amsterdam. Prof. van Dantzig advised him to specialize in non-parametric Statistics and indicating as starting point to study the famous Hoeffding paper on U-statistics published in 1948 in the Annals of Mathematical Statistics. This paper then became the focus of Dawie’s doctoral study. The Dr. Math. et Phys. degree was awarded to him on 19 January 1955 after defending his doctorate thesis in public, as was and still is, required in The Netherlands.

Due to enormous pressure from the HSRC’s President, Dr J C Garbers and the Vice-President, Dr A J van Rooy, he joined the HSRC as Director of the Institute of Statistical Research on 1 January 1981. Due to lack of experience on the application of statistical methods in the social sciences, he had to study the literature in this field. He identified log-linear modeling, logistic regression and survey sampling as the main areas. Although he had presented post-graduate courses in sampling, based on Cochran: Sampling Techniques, at UP, he immediately realized that he had no real knowledge of the application of sampling techniques. As a consequence he got permission from the HSRC’s President to visit various social research institutes, Census Bureau’s and leading statisticians in the indicated fields. In March 1962 he visited, amongst others, in the USA the Institute of Social Research at the University of Michigan, Ann Arbor (Prof Leslie Kish and others), the Triangle Research Institute, National Opinion Research Center (NORC), the US Bureau of the Census, Prof Goodman at the University of Chicago, Prof Stephen Fienberg at the University of Pittsburg, Prof Gary Koch and Prof James Grizzle at the University of North Carolina and in England Dr Denise Lievesley at the City University of London and the British Bureau of Statistics, also in London. The visit to Prof Kish in 1982 was followed up by attending, as a guest, in July/August 1984 the advanced courses in survey sampling and the application of survey sampling during the summer school of the Institute of Social Statistics at the University of Michigan, Ann Arbor.

In July 1982 Dawie was promoted to one of the four vice-president posts of the HSRC. Prof. Nico Crowther from the University of Port-Elizabeth was then appointed as Director of the Institute of Statistical Research at the HSRC. Notwithstanding a high managerial work load as vice-president, Dawie preferred to continue with his work as sampling specialist in the Institute of Statistical Research, a request that was granted. During the years 1981 to 1990 various university personnel were assisting the Institute of Statistical Research in specialized areas, amongst others Prof Michael Browne, Dr Stephen du
In Memoriam

Toit, Dr Fanie Steyn, Dr Hennie Groeneveldt and Dr Kotie Roux.

While being at UP and the HSRC he supervised 10 D-students under whom Hoffie Lemmer, Gustav Reinach, Herman Schoeman, Chris Smit, Roger Markham and Mathilda du Toit (née Barnard). He was appointed as honorary professor in the Department of Statistics (UP) on an annual basis from 1991 till 2008 (he became seriously ill in 2009).

At the HSRC Dawie was involved in many training courses of HSRC’s personnel and other interested parties in Pretoria as well as at various statistical offices and universities. He also headed, amongst others, two important large-scale projects, namely Investigation into differential entrance requirements for tertiary educational institutions, and Sample based population estimates in 88 inaccessible areas of the RSA as part of the population census in 1991

Immediately after his retirement he accepted appointment as a chief research specialist at the HSRC and from 1 January 1991 as director of the Institute for Statistical Research as a consequence of Prof Nico Crowther’s appointment as Prof. and Head of the Department of Statistics at UP. On the 1 July 1992 he left the HSRC to take up the position of director of STATOMET in the Department of Statistics at UP. STATOMET was founded by Prof. Crowther in the department with the view to assist the private sector with the solving of research problems. Dawie was succeeded by Deon Herbst as director to enable him to work only on a part-time basis. He finally resigned at the end of 1996 and from January 1, 1997 he became a private consultant working from his home.

As a consultant his main clients were the Central Statistics Service (CSS, headed by Dr M Orkin), – later on the name changed to Statistics South Africa or StatsSA), the HSRC and private market research organizations. As soon as it became known that Dawie was no longer at STATOMET, Dr Orkin requested him to assist the CSS with training of the relevant personnel in appropriate statistical methods and with the design of samples. He initially worked three days per week at the CSS, later on reduced to two days per week. This involvement continued until the end of 2004. Dr Orkin was succeeded by Mr Pali Lehohla (a previous Statistician General) whom Dawie met for the first time in 1986 at one of the HSRC's training courses. As a consultant for StatsSA Dawie was, amongst others, responsible for the estimation of the undercount in the 1996 population census and co-responsible for the estimation of the undercount in the 2001 population census and the community survey in 2007. One of his main areas of interest was the estimation of the standard error of estimates in multi-stage cluster (i.e. complex) sampling and related issues.

In 1966 he was requested by the EC of SASA to launch the SA Statistical Journal. He acted as editor and managerial editor for the first issue of the journal in the first half of 1967 but due to the amount of work involved Dr Cas Crouse was appointed as editor from the second issue in 1967 and onwards while Dawie continued as managerial editor. In 1967 and for many years thereafter the journal was typed using a “ball”-type IBM-typewriter, which involved heavy monitoring. During his retirement he did consulting mostly in the field of sampling, and he continued with his work as consultant at StatsSA.

Dawie served on the Statistics Council for several terms. He was also a member, fellow, past-president, honorary member and honorary president of SASA.

Dawie passed away on 21 November 2018 – he would have turned 90 next year.

The statistical community honor a big name in science in South Africa and specifically in Statistics. His legacy is enjoyed by many people in South Africa, ranging from students, researchers in many disciplines to practitioners in a vast spectrum of sciences.

We salute him and pray for his family who will dearly miss him!

Chris Smit, Mossel Bay, 21 November 2018

“The statistical community honor a big name in science in South Africa and specifically in Statistics. His legacy is enjoyed by many people in South Africa...”
Professor Emeritus Michael Browne Passed Away

Michael Browne, Emeritus Professor of Psychology and Statistics, passed away on October 10, 2018. He received his BA and M.Sc. degrees from the University of the Witwatersrand and his Ph. D. from the University of South Africa.

After serving at the South African National Institute and University of South Africa, Dr. Browne joined the Ohio State University in 1991 and worked here until his retirement in 2011. During those years he received lifetime career achievement awards from the Society for Multivariate Experimental Psychology and from the Psychometric Society, in addition to receiving the Fred Brown Research Award from Department of Psychology and serving as the Roger E. Kirk Scholar.

Dr. Browne was a brilliant man, remarkably modest about his accomplishments, and an engaging conversationalist. It is very sad to think that he is no longer with us, but he leaves a legacy of colleagues and students who benefited greatly from his expertise and his warmth.

Memoriam

Professor Emeritus Michael Browne

SEMINAR PROGRAM: SECOND SEMESTER 2018

Stellenbosch University, Department of Statistics and Actuarial Science

<table>
<thead>
<tr>
<th>Date</th>
<th>Speaker</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 February</td>
<td>Peter Filzmoser</td>
<td>Robust and sparse classification in high dimensions</td>
</tr>
<tr>
<td>22 February</td>
<td>Adriaan Rowan</td>
<td>Using and understanding non-linear predictive models</td>
</tr>
<tr>
<td>8 March</td>
<td>Etienne Roussouw</td>
<td>The rapidly developing world of (re)insurance</td>
</tr>
<tr>
<td>12 April</td>
<td>Piet Maree</td>
<td>Incentives, physical activity and absenteeism: investigating the impact of Vitality on employee wellbeing and how an Apple Watch can prevent absenteeism</td>
</tr>
<tr>
<td>26 April</td>
<td>Mark Nasila</td>
<td>Why Statisticians are key in harnessing the ‘Fourth Industrial Revolution’</td>
</tr>
<tr>
<td>10 May</td>
<td>Loamie Kotze</td>
<td>Markov modelling of disease progression in the presence of missing covariates</td>
</tr>
</tbody>
</table>

Lectures start at 13:00 and are held in room 2048 of the Van der Sterr Building, c/o Victoria and Bosman streets, Stellenbosch.

Enquires: Danie Uys  Telephone: 021 808 3879  Fax:021 808 3830  e-mail: dwu@sun.ac.za

Stay informed: visit the SASA website: https://sastat.org.za/

Expert list: https://sastat.org.za/statistical-consultants-expert-list
Competition and bursary information: https://sastat.org.za/competitions-bursaries-and-scholarships
We would also like to congratulate the following Postgraduate Paper Competition winners. We received six papers from four universities this year. Thank you to the University of Cape Town (UCT), the University of Pretoria (UP), the University of South Africa (Unisa) and the University of Stellenbosch (US) for their entries.

First Prize

JC Malela-Majika (UNISA)
“Modified side-sensitive synthetic double sampling monitoring scheme for simultaneously monitoring the process mean and variability”

Second Prize

I Singini (UCT)
“Diagnostics for joint models for longitudinal and survival data”

Third Prize

A Oosthuysen (UP)
“Two new tests for exponentiality based on the waiting times of the Poisson process”

The SASA Education Committee would like to thank the following judges for their time and effort: Prof Jacky Galpin, Dr Sibusiswe Khuluse (CSIR), Prof Paul Mostert (US) and Dr Khangalani Zuma (HSRC). We appreciate that it is a huge task to adjudicate these papers.

We would like to encourage everyone to go like the South African Statistical Association facebook page for announcements on competitions, bursary or other related statistics education news.
Honours Project Competition Sponsored by SAS®

It is the SASA Education Committee’s pleasure to announce the winners of the 2017/2018 SASA Honours Project Competition. We received nineteen honours projects from eight universities this year. Thank you to Nelson Mandela University (NMU), Stellenbosch University (SU), University of Cape Town (UCT), University of Fort Hare (UFH), University of KwaZulu-Natal (UKZN), University of Pretoria (UP), University of Venda (UV) and University of the Witwatersrand (Wits) for their entries. Congratulations to the following entries:

First Prize
James Lloyd (UCT)
“The Classification of Galaxy Images Using Convolutional Neural Networks”
The prize included sponsorship to attend and present the project at SASA2018.

Second Prize
Sibusiso Thabene (UCT)
“Sampling from Distributions Using Genetic Algorithms”

Third Prize + Special prize for best use of SAS
Mark de Lancey (UP)
“CLARA Algorithm for Image Clustering”
The SASA Education Committee is grateful to the following judges for their time and effort: Prof Legesse Debusho (UNISA), Prof Carl Lombard (MRC), Dr Shaun Ramroop (UKZN), Dr Victoria Goodall (NMU), Dr Vincent Micali (Stats4Buz) and Dr Justin Harvey (SU). As always, ranking the top projects was a difficult task, in particular due to the variety of standards and subject areas, and the number of projects that had to be adjudicated. We’d also like to extend a special word of thanks to SAS for their continued support of this competition.

2018/2019 Honours Project Competition

Honours Project Competition Sponsored by SAS®
The SASA education committee invites all South African Universities to submit their top two honours projects in Statistics from 2018 for the 2018/2019 Honours Project Competition. The head of the Department/School at each University may submit up to two projects and group work is allowed. The national winner will be sponsored to present his/her project at the annual SASA conference. The closing date for entries is 29 March 2019. Please contact Nombuso should you require further details.

SASA Education Committee
Contact person: Nombuso Zondo
E-mail: zondon@ukzn.ac.za
President 
Prof. M Lesaona (Maseka)  
Department of Statistics and Operations Research  
University of Limpopo  
lesaona.maseka@ul.ac.za

Past President 
Prof. DW Uys (Danie)  
Department of Statistics and Actuarial Science  
University of Stellenbosch  
dwu@sun.ac.za

Secretary 
Dr. C Clohessy (Chantelle)  
Department of Statistics  
Nelson Mandela University  
Chantelle.Clohessy@mandela.ac.za

Treasurer 
Prof. L Raubenheimer (Lizanne)  
School of Mathematical and Statistical Sciences  
North-West University  
lizanne.raubenheimer@nwu.ac.za

Managing Editor of Journal 
Dr. C Pretorius (Charl)  
School of Mathematical & Statistical Sciences  
North-West University  
charl.pretorius@nwu.ac.za

Editor of Conference Proceedings 
Prof. L Santana (Leonard)  
School of Mathematical & Statistical Sciences  
North-West University  
leonard.santana@nwu.ac.za

Newsletter Editor 
Dr. T Sandrock (Trudie)  
Department of Statistics and Actuarial Science  
University of Stellenbosch  
trudies@sun.ac.za

Funding 
Prof. F Gumedze (Freedom)  
Department of Statistical Sciences  
University of Cape Town  
freedom.gumedze@uct.ac.za

Sponsorships 
Prof. S Das (Sonali)  
CSIR, SDas@csir.co.za

Chairpersons of Committees 
Education Committee  
Dr. L Fletcher (Lizelle)  
Department of Statistics, University of Pretoria  
lizelle.fletcher@up.ac.za

Youngs Statisticians’ Network  
Dr. H Brydon (Humphrey)  
Department of Statistics and Population Studies, University of the Western Cape  
hbrydon@uwc.ac.za

Additional Members 
ICCSSA Board Chair person  
Mrs. Y Chhana (Yoko)  
School of Statistics and Actuarial Science  
University of the Witwatersrand  
Yoko.Chhana@wits.ac.za

https://sastat.org.za/executive

Contact details for the newsletter editor:  
Dr. Trudie Sandrock  
Department of Statistics and Actuarial Science  
University of Stellenbosch  
Private Bag X1  
Matieland 7602  
Tel: 021-808-3244 (w)  
Fax: 021-808-3830  
trudies@sun.ac.za

Please send all Newsletter matters to the Editor at the above address, and all other SASA matters to  
The Secretary,  
R.O. Box 3341, Matieland,  
7602, South Africa  
email: Chantelle.Clohessy@mandela.ac.za

The Newsletter is an informal collection of news items and short articles of interest.  
The Newsletter will appear in April, September, December/January

Submission deadlines:  
• May newsletter: 1 April  
• September newsletter: 1 August  
• December/January newsletter: 1 December

The views expressed in this Newsletter are those of the contributors and do not necessarily reflect the policies of the SA Statistical Association or the editor of the Newsletter.

Advertising in the Newsletter  
Advertise the Statistics programmes offered by your institution in the quarterly issued newsletter of SASA.

The costs for advertising in the newsletter are:  
• Half A4 page Black and White R1000+vat  
• Full A4 page Black and White R2000+vat  
• Half A4 page Colour R2500+vat  
• Full A4 page Colour R5000+vat  
Contact the newsletter editor, Dr Trudie Sandrock  
trudies@sun.ac.za if you are interested.
Statistics has entered almost every aspect of human endeavour from sport to weather, health, population growth rate, agriculture, housing, schooling, unemployment and the stock market.

At the centre of statistics lies data. This entails designing ways to collect, summarise, visualise, present and draw inferences from data. This data is then used for better planning, more efficient delivery of services and increased productivity.

All these concepts were unpacked at the World Statistics Day open day recently hosted by the CSIR and jointly organised with the South African Statistical Association, the Institute of Certificated and Chartered Statisticians of South Africa, the Operations Research Society of South Africa, the Department of Science and Technology-National Research Foundation Centre of Excellence for Mathematics and Statistical Sciences and the Statistical Analysis System.

Speaking at the event, Dr Daniel Visser, CSIR Research and Development Strategy Manager, said statistics is an extremely important field because it is used every single day, whether it is census data, crime statistics or financial statistics.

The event was attended by representatives from industry, academia and the public sector.

“From a government point of view, we rely on statistics and data to plan better for the country,” says Daniel. He warned that the advent of concepts such as big data, data analytics and artificial intelligence will have an enormous impact in the context of a rapidly evolving technological world and on the field of statistics.

“We are gaining the ability to work with extremely huge data sets now, make sense of them and define trends out of these data sets, effectively helping us to make decisions faster and sometimes autonomously,” he says.

Prof Pravesh Debba, CSIR’s manager of spatial planning and systems competence area, spoke about the future of statistics. “The flood of data is now accompanied by a flood of questions, perhaps thousands of them, many of which are complex. This flood of data led to a big data revolution. Interest in the discipline of statistics and the analysis of data is booming. It is estimated that 90% of the world’s data have been created in the last two years.”

Pravesh added that initiatives such as the National Development Plan, Data Revolution, the Fourth Industrial Revolution and the Sustainable Development Goals are creating the need for people with skills in handling and analysing data. This would create further demand for such people in both industry and government departments.
“The role of higher education in producing data analytics skills needed for the market is vital and this would also lead to the development of new statistics curricula at universities,” says Pravesh.

Statisticians in Pravesh’s group use statistical methods to understand how phenomena such as predicting election results, urban growth and electricity demand forecasting could influence policies and lead to interventions going forward.

Kelly Lu, from Statistical Analysis System, touched on the importance of data and analytics as key drivers to success outside of the business world. Since analytics had an impact in transforming the way people conduct their business, it has the potential to make a positive impact on the environment and people’s lives in the same way.

Dr Christine Khoza gave a status update on the planning for the South African Census 2021, emphasising the objectives and new strategies for data collection and highlighted lessons learnt from previous censuses.

CSIR senior researcher Nontembeko Dudeni-Tlhone spoke about the use of census data as key building blocks in segmenting households into clusters of homogeneous groups that are expected to behave in a similar way at the neighbourhood level. This provides a framework for predictive modelling at the neighbourhood scale, which has the potential to enhance ongoing long-term planning efforts and decisions.

Quintin van Heerden, a CSIR senior researcher gave an overview of the different uses of census data within the urban growth modelling and simulation endeavours of the CSIR.

Dr Mathetha Mokonyama, CSIR’s manager for transport systems and operations competence area made a presentation on how large-scale household surveys, including census, are critical for this specific type of analysis; and further makes a case for improved integration of survey instruments in order to optimally address socioeconomic challenges that are unique to South Africa.

CSIR senior researcher in the spatial planning and systems area, Dr Sibusisiwe Makhanya, demonstrated how census data were used in mapping particulate matter concentrations and in calculating indicator 11.6.2 of Sustainable Development Goal 11.
Get trained.
Get certified.
Add value.

SAS® Certifications available in big data, advanced analytics and data science.

Discover more at
sas.com/academy-data-science
Transform the way we do banking

Apply for our graduate programmes

For more information on our graduate programmes visit graduate.standardbank.com or email Gregory.Newnham@standardbank.co.za

graduate.standardbank.com

Standard Bank Moving Forward™