UNSW RESEARCH INFRASTRUCTURE
Mark Wainwright Analytical Centre

Stats Central
The Statistical Consulting and Collaboration Unit

Book a consultation

Questions or comments?

Email us at
stats.central@unsw.edu.au

In this issue:

Stats Central now supports Data Science!
Upcoming Events
Going off Main Campus
Stats Central Team
Recent Activities
Highlights
Statistics and Ethics
Current Partners and Projects

Welcome to the Spring newsletter for Stats Central — the UNSW Statistical Consulting and Collaboration Unit. Below you will find recent news and what's coming up next. In particular, note we now offer consulting on Data Science, and have been busy off main campus!

David Warton and Nancy Briggs Stats Central leadership

Stats Central now supporting Data Science!

Data science is a multi-disciplinary field at the interface between statistics and computer science that goes beyond study design and analysis, into data management and big data. A/Prof Raymond Wong (School of Computer Science & Engineering) has joined the Stats Central team to add depth to our Data Science expertise, and he leads our new Data Science initiatives, including:
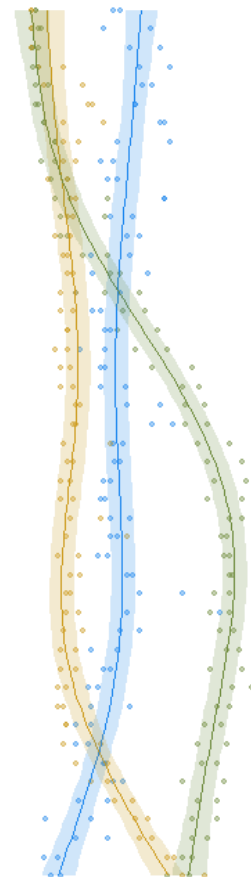- Data science consulting (especially on analytics, data mining, machine learning).
- Data Science courses and seminars, including, "*Introduction to Python for Data Science*"
- Contract work on Data Science projects, including with ANSTO (Australian Nuclear Science and Technology Organisation).

Book a consultation now!

Upcoming Events

Seminar

14 November Seminar: 2019 in Graphs, Professor David Warton (Stats Central and School of Mathematics and Statistics)
As we come towards the end of 2019, we will look back on the year that was – politics, trade wars, climate change and S25 – through the lens of data visualisation!

Short courses

Sample Size and Power Calculations: November 26, Mark Donoghoe
Power calculations and sample size determination are essential parts of planning a scientific study. In this one-day course we will introduce the basic principles of precision-based and power-based sample size calculations.

Introduction to Python for Data Science: December 3, Raymond Wong
Python is a widely used programming language to manipulate, analyze, and visualize data. It is one of the most popular languages for Data Science, especially when dealing with complex, uncurated or text datasets.

For further details visit our website at *statscentral.unsw.edu.au*

Going off Main Campus

While Kensington Campus is a main focus for UNSW activities, there are dozens of other UNSW sites in Sydney and beyond, such as teaching hospitals, research institutes and UNSW Canberra. These campuses have been an increasing focus of our activities over the last year, including the following highlights:

- Partnerships with four units based off the Kensington campus, with one of our consultants physically based in an off-campus unit for part of their working week.
- Seminar series on introductory statistics at St Vincent's Clinical School and St George Clinical School.
- 45% of our consultations over the last year have been with researchers off Kensington campus!
- Consultations with UNSW researchers based across over 20 campuses.

If you want to book a consultation and are a bit of a drive away, we are happy to video conference.
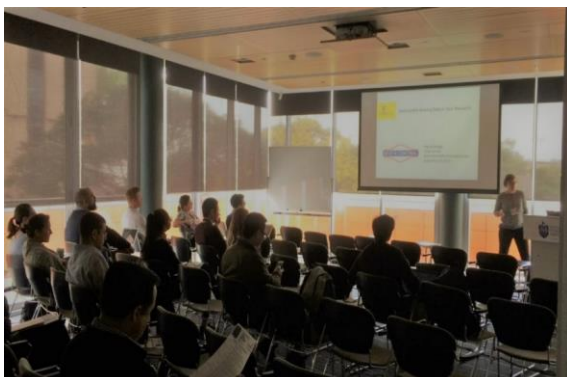
Seminar series on Introductory Statistics at St Vincent's Clinical School – July 2019

Stats Central Team

We are delighted to announce Associate Professor Raymond Wong joined Stats Central from Computer Science and Engineering as a Data Science Consultant at the start of 2019. He has over 20 years experience in academic, research, startups and industry consulting experience in the areas of data management, data mining and analytics. He has previously worked with various interdisciplinary teams. He enjoys identifying problems, decomposing complex problems into smaller pieces, and solving these problems using techniques from data science. Data science is a multi-disciplinary field that uses mathematics, algorithms and systems to extract knowledge, insights and trends from data.

He oversees our data science consulting, course and seminar offerings and has started contract work with ANSTO. To date he has given a well-attended seminar on 12th March "*What can Data Science do for you?*", and ran a short course on 29th August "*Introduction to Python for Data Science*".

Ben Maslen joined Stats Central in October 2018, with expertise in study design and multivariate analysis.

Gordana Popovic is back on the team and working with Taronga Zoo, UNSW Kirby, and developing on-line chemometrics tools in collaboration with the School of Chemistry.

Above is a photo of the Stats Central team, left-to-right:
Peter Geelan-Small, Mark Donoghoe, Eve Slavich, Sue Wilson, Nancy Briggs, Gordana Popovic, Zhixin Liu, David Warton
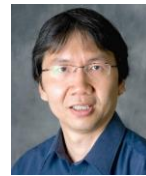
Raymond Wong
Data Science
Consultant

Ben Maslen

Recent Activities

Consulting: We offer free design and analysis advice to UNSW HDR students and free design advice to staff, and can be contracted to collaborate on larger projects.

In 2018, we supported 415 UNSW research projects from 40 schools across eight Faculties, we taught seven short courses to over 190 participants, and presented 7 seminars to over 560 attendees.

The Faculties with highest usage are Medicine and Science, where all of our partner schools and units are currently based, including several based off campus. In fact, over the last year *45%* of consultations were with researchers not based on the main Kensington campus (see separate news item).

Short courses and seminars: In 2019 we have delivered 7 short courses at heavily discounted rates to over 195 attendees from eight Faculties, UNSW affiliation and external organisations. This included our new course "*Introduction to Python for Data Science*", presented by Raymond Wong, our Data Science Consultant. We can also assist in teaching, for example, Professor Susan Wilson (Senior Statistical Consultant) presented guest lectures on BINF3010/9010 Applied Bioinformatics to masters students in School of Computer Science and Engineering.

We organise a monthly seminar series, and have given 10 seminars to over 930 attendees, plus guest lectures within UNSW partners and schools.

Recently we have developed an *Introductory Statistics Seminar Series:* which has been taught by contract at St. Vincent's Clinical School and St George and Sutherland Clinical School.

Highlights

Most popular Seminar June 11: Visualising data - making sure your graph is worth 1,000 words - Hosted by Peter Geelan-Small
Over 200 attendees from UNSW staff, researchers, students and external organisations.

Research Trade Expo for Medicine, Science, and Engineering - July 25
Stats Central team participated in the event and engaged with various UNSW Faculties, external suppliers and research industries.

Grant review: Ask us to review your next grant!

There is evidence from a number of sources that the leading reason a grant application does not get funded a poor design and analysis plan – even more important than track record! Stats Central offers statistical review of competitive grant applications.

To maximise benefit from our grant review service it is important to approach us as early as you can – now is not too early to start talking with us about your 2020 plans.

For NHMRC and ARC grant application deadlines see Grant Management Office (GMO) websites:
NHMRC
https://research.unsw.edu.au/national-health-medical-research-council-nhmrc-0
ARC
https://research.unsw.edu.au/australian-research-council-arc

Use by Faculty Sep 2018 - Aug 2019

Top Schools by Usage Sep 2018 - Aug 2019

Statistics and Ethics

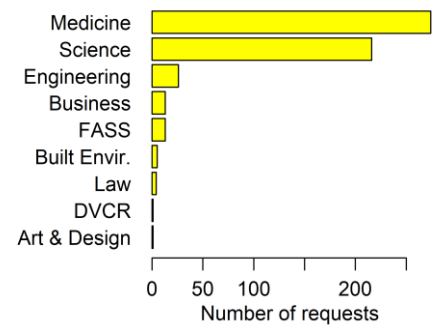*Does your study have ethical implications? Well so does your use of statistics!*

If the study is too large then it is involving extra subjects, hence putting extra people at risk (as well as unnecessarily increasing costs). If the study is too small, it will not be able to detect biologically, or clinically, important effects. Such a study is unethical in its use of subjects, as well as other resources.

Any ethics application needs to justify why the proposed sample size is appropriate. This is more than a simple sample size calculation – it requires thinking about what the overriding consideration in the study is, particularly the primary research objective, which guides how data will be collected and analysed. How the data will be analysed in turn drives the sample size calculation(s), which may be very complex, depending on the data and analysis plan. Often there is a lot of interplay between the data to be collected, the analyses and the sample size needed.
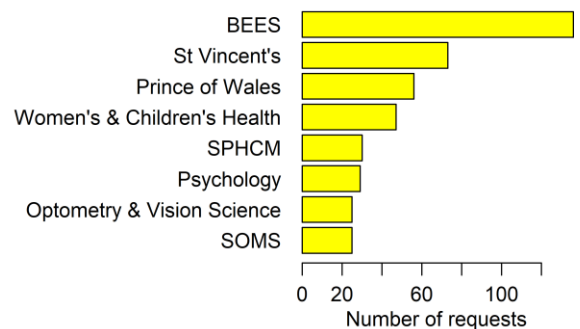
Two tips for the planning stage:
(i) Your objectives should be divided into primary and secondary objectives, as relevant to the study. The primary objective drives the sample size deliberations. If you have more than a single primary objective, a multiplicity correction needs to be incorporated.

(ii) The sample size calculation needs to be in accord with the method of analysis. For example if your primary objective is to evaluate changes over time, or involves repeated measurements, then not only does the statistical analysis need to be relevant to such data, but sample size estimation needs to be relevant to the analysis. Basing an estimate on just the initial and final measurements is not appropriate for a repeated measures study.

Two common misconceptions:
(i) Sample size calculations can only be done when there are clearly specified hypotheses. Very often, precision of the estimate(s) is important, and this is used to determine the required sample size.
(ii) A sample size calculation is not needed because a pilot study is being planned. For example, the primary purpose of such a study may be to obtain an estimate of measurement(s) that in turn will drive the planning for future research. How precise this estimate should be will determine the sample size.

Finally, don't hide behind "statistical" significance. What is fundamentally important is the biological (or clinical, or medical) significance, which needs to be justified within the overall context!

For further questions contact us or visit our website at *statscentral.unsw.edu.au*

Current partners and projects

Stats Central is funded by the Research Division, with support from the School of Mathematics and Statistics, and a suite of Partner Schools.

Schools and units can partner with Stats Central, buying out a portion of a consultant's time to dedicate to the partner. The consultant can be physically based in the partner school or unit for that time to increase access, facilitate specialisation and deeper collaborations. Basically, to enable best-practice consultancy and to tailor what we offer based on the partner's needs — consulting, teaching assistance, workshops, grants advice, support for Honours or ILP students.

Stats Central uses funding from partners to grow our unit in response to demand, so there is also broader institutional benefit to partnering with Stats Central. 2019 partners:

o School of Mathematics and Statistics
o Prince of Wales Clinical School
o St Vincent's Clinical School
o School of Biological, Earth and Environmental Sciences (BEES)
o School of Biotechnology and Biomolecular Sciences (BABS)
o School of Optometry and Vision Sciences
o Kids Cancer Centre
o Neuroscience Research Australia (NeuRA)

We also collaborate with research groups on projects requiring specialised expertise, under similar funding arrangements, and have had considerable recent growth in this area. Current projects:

o Maintain Your Brain (Dementia Collaborative Research Centre and Centre for Healthy Brain Aging)
o Opendatafit: Data analysis and storage for drug activity, kinetics and flow cytometry (Chemistry)
o Latent Emotion Trajectories Among Blood Donors (Psychology)
o Early Hospitalisation for RSV as a Risk Factor for Later Respiratory Problems (Paediatrics)
o Health Outcomes Amongst Urban and Rural Residents (Kirby)
o Antibiotic Prophylaxis in Prevention of Catheter-Related Infection (St. George)
o ANSTO: Water chemistry and microbial analysis
o Muscle Growth in the Lower Extremity: The MUGgLE study (NeuRA)
o NPS MedicineWise: Impact of the NPS Medicinewise program on adherence to metformin using a 10% PBS sample
o Taronga Zoo: Reviewing the Regent Honeyeater breeding program
o School of Public Health and Community Medicine (the career pathways project)

Consider writing Stats Central funding into your grant applications and joining this list!