Inaugural Young-ISA Meeting

Friday 4th October 2019
Glenroyal Hotel
Maynooth, Ireland

Contact us young.istat@gmail.com

Follow us @Young_Stats_ISA

Stay tuned young-istat.github.io

10:00 – 10:30	**Registration**
10:30 – 10:50	**Welcome Address**
10:50 – 11:40	**Statisticians in Academia** *Norma Bargary – University of Limerick*
11:40 – 12:00	**Cofeee Break**
12:00 – 13:00	**Research Presentations – Round 1** • Anomaly detection for streaming advanced manufacturing data *Caitríona Ryan – Maynooth University* • Clustering of accelerometer profiles *Richard Weedle – University College Cork* • Introducing network meta analysis *Laura Ngouafo – University of Limerick*
13:00 – 14:00	**Lunch**
14:00 – 14:50	**Statisticians in Industry -Statistician or Data Scientist?** *Aidan Boland – Edge by Ascential*
14:50 – 15:20	**Round Table** *Norma Bargary and Aidan Boland*
15:20 – 15:50	**Coffee Break**
15:50 – 16:50	**Research Presentations – Round 2** • A network approach to votes exchange in the Eurovision Song Contest *Silvia D'Angelo – University College Dublin* • Modelling menstrual cycle length using state space models *Thiago de Paula Oliveira – NUI Galway* • Aesthetic image analysis *Koustav Ghosal – Trinity College Dublin*
16:50 – 19:00	**Wine Reception and Closing Address**

Statisticians in Academia

Norma Bargary
University of Limerick

This talk will outline my career in academia to date. It will discuss the route I took to becoming a statistician, and my decision to work in academia. It will highlight the advantages and disadvantages of being a young statistician in academia in the current climate, and will discuss some of the main things I have learned in the time from my PhD to now.

Anomaly detection for streaming advanced manufacturing data

Caitríona Ryan, Andrew Parnell
Maynooth University and I-Form

This talk presents some current research on statistical anomaly detection in streaming manufacturing data as part of I-Form, the SFI Research Centre for Advanced Manufacturing. Research goals include (i) decision support for machine operators, (ii) defect detection in build parts and (iii) predicting machine health. With the advent of modern dynamometers, IoT sensors, microscopy, and ethernet-connected machine tools, we are now in a position to leverage new artificial intelligence techniques to quantify the performance and behaviour of the manufacturing process. I will also present a novel recursive extreme studentised deviate algorithm (R-ESD) to detect anomalies in streaming time series data via statistical learning. We adapt the generalised ESD test (Rosner, 1983) to streaming time series data by using time series decomposition and a sliding window approach. This is made computationally feasible by recursive updates of the ESD test statistic (Grubbs, 1950). Our method is statistically principled and it outperforms the *AnomalyDetection* software package, recently released by Twitter Inc. (Twitter) and used by multiple teams at Twitter as their state of the art on a daily basis (Vallis, Hochenbaum and Kejariwal, 2014).

Clustering of accelerometer profiles

Richard Weedle, Kathleen O'Sullivan, Tony Fitzgerald
University College Cork

Physical activity has a major impact on health. Questionnaires are the most common method of physical activity assessment. While cost effective, these are subjective and are known to correlate poorly with actual activity levels. Accelerometers have gained popularity given their accuracy, objectivity and ability to capture large amounts of data. The study is based on a cohort of individuals who wore accelerometers in a free living environment for one week. This paper aims to identify and profile subgroups within a cohort. To achieve this clustering and functional principal component analysis (FPCA) were utilised.

Introducing network meta analysis

Laure Ngouanfo[1], Joy Leahy[2], Ali Sheikhi[1], Cathal Walsh[2]
[1]*Health Research Institute, MACSI, University of Limerick*
[2]*National Centre for Pharmacoeconomics, St James Hospital*

Public healthcare entities, as well as community members rely on the best available information to base their health decisions on. Statisticians, in their perpetual quest to solve contemporary research questions, have gradually developed methods such as Network Meta Analysis (NMA), to optimize the quality of evidence in healthcare. NMA is an extension of the traditional pairwise meta analysis that allows the comparison of more than two interventions even if some have not been compared in a head to head trial. Thus NMA has been used to produce up-to-date guidelines informing on the ever-growing number of treatment options and to derive efficient conclusions on the choice of first-line regimens. We walk through the initial approaches that includes indirect comparison, then the evolution to Bayesian NMA synthesizing results from multi arm trials, the diffusion of the method and the fundamental assumptions. We examine the performance of using an applied example and compare it to more simple syntheses. As seen in the applied work of Noone et al. and Welton et al, when many of the treatment interactions have not been observed it can be difficult to distinguish between the models. However, through our theoretical example we see that an increase in the amount of data can be of benefit. We do this using bootstrap replication. NMA when used appropriately, indeed has the potential to influence healthcare delivery.

Statisticians in Industry - Statistician or Data Scientist?

Aidan Boland
Edge by Ascential

In 2017, The Economist published an article saying that the worlds most valuable resource is no longer oil but data. While this is not a perfect analogy it does highlight the importance of data. Data, however, is only useful if analysed appropriately. As statisticians, we are more than familiar with the importance of this. The demand for knowledge in statistics is constantly growing in industry. This may not be immediately evident as the terminology evolves and broadens; a statistician using linear regression is often referred to as a data scientist using machine learning. The title data scientist tends to be very broad and can vary substantially across different organisations. Thus, data scientists can have different skill sets. There is a focus on not only being able to apply statistical techniques in once off analysis, but also to implement algorithms inside large production workflows. The focus on presentation and accessibility is also important, and tools such as RMarkdown and Shiny have enabled statisticians to easily share work. It is now reasonably straight forward to build a user interface or to create an interactive presentation both of which can be used to share analysis. My talk will focus on my journey in industry, what I have learned along the way, and where I see the current role of statisticians in the world of data. I will also present a few examples of data science in action.

A network approach to votes exchange in the Eurovision Song Contest

Silvia D'Angelo[1], Marco Alfò[2], Thomas Brendan Murphy[1]
[1]*University College Dublin*
[2]*Sapienza University of Rome*

The Eurovision Song Contest is a popular TV singing competition held annually among country members of the European Broadcasting Union. In this competition, each member can be both contestant and jury, as it can participate with a song and/or vote for other countries tunes. During the years, the voting system has repeatedly been accused of being biased by tactical voting; votes would represent strategic interests rather than actual musical preferences of the voting countries. In this work, we develop a latent space model to investigate the presence of a latent structure underlying the exchange of votes. Focusing on the period from 1998 to 2015, we represent the vote exchange as a multivariate network: each edition is a network, where countries are the nodes and two countries are linked by an edge if one voted for the other. The different networks are taken to be independent replicates of a conditional Bernoulli distribution, with success probability specified as a function of a common latent space capturing the overall relationships among the countries. Proximity denotes similarity, and countries close in the latent space are more likely to exchange votes. If the exchange of votes depends on the similarity between countries, the quality of the competing songs might not be a relevant factor in the determination of the voting preferences, and this would suggest the presence of some bias. A Bayesian hierarchical modelling approach is employed to estimate the parameters, where the probability of a connection between any two countries is a function of their distance in the latent space, network- specific parameters and edge-specific covariates. The estimated latent space is found to be relevant in the determination of edge probabilities, however, the positions of the countries in such space only partially correspond to their actual geographical positions.

Modelling menstrual cycle length using state space models

Thiago de Paula Oliveira, John Newell
NUI Galway

Times are changing. At an elite level, female athletes and coaches across the globe are now starting to work with the menstrual cycle to gain a performance edge. By tracking the menstrual cycle, and knowing how, why and when hormone fluctuations affect female physiology, an athletes training, nutrition and recovery can be tailored to their cycle to sustain peak performance at the highest level. Such individualization is possible if cycle length can be predicted to a high degree of accuracy. To achieve this, we built a hybrid predictive model using data on 3396 cycles collected from a sample of 561 women (mean age 35.16 years, range 13.18 - 53.34, number of menstrual cycles range 4 27). In order to capture the within subject temporal correlation a mixed- effect state space model was fitted incorporating a Bayesian approach for process forecasting to predict the duration (in days) of the next menstrual cycle as well as its forecasting interval. The modelling procedure was split into three steps (i) a time trend component using a random walk or linear mixed effects model as appropriate, (ii) an autocorrelation component using a mixed effect autoregressive moving-average (ARMA) model and (iii) a linear predictor to account for covariates (e.g. injury, stomach cramps, training intensity). To assess the performance and prediction accuracy of the model each womans last observation was used as test data. The Root Mean Square Error (RMSE), Concordance Correlation Coefficient (CCC) and Pearson correlation coefficient (r) between the observed and predicted values were calculated. The model had an RMSE of 1.06 days, a precision of 0.82 and overall accuracy

of 0.99. The ability to predict cycle length to a high degree of precision enables recommender systems to be built to help female athletes track their period and tailor their training and nutrition correspondingly, knowing when to push harder, when to prioritize recovery and how to minimise the impact of menstrual symptoms on performance. In conclusion, hybrid models are a useful approach for predicting menstrual cycle length which in turn can be used to support female athlete wellness and optimize performance at all levels.

Aesthetic image analysis

Koustav Ghosal
Trinity College Dublin

Aesthetic Image Analysis (AIA) refers to the task of understanding the properties that make an image look pleasing or dull. It is a complex but comparatively less studied problem in digital image processing research. Traditionally, in image processing research, the focus is on understanding the physical properties of an image such as objects, scene type, etc. While early approaches relied on handcrafted features, with the advent of big data and neural networks, the current trend is to learn the features directly from raw data. By learning from large image databases using deep and complex neural networks, in recent years, there has been a rapid development in this area both in terms of quality of results and quantity of work with applications such as traffic behaviour analysis, face detection, medical image analysis, etc.With the maturation of the traditional problems, more complex and subtle areas of image understanding have begun to be explored such as image style transfer or AIA. Aesthetic properties such as “too dark”, “distracting background”, “subject out of focus” are partly subjective and partly objective ambiguous concepts and hence more difficult to learn than the physical properties such as category, count, etc. In this work, we focus on two different problems in AIA. In the first part of the talk, we will talk about an application that can predict the dominant aesthetic attributes of an image such as “motion blur”, “leading lines”, etc. In the second part, we will talk about a system that can generate textual feedback for a photograph such as “I like the textures of the foreground but the background is a bit too dar”. The final part of the talk will involve discussing the ongoing and potential problems in the area.