Miss Natalie CHUNG

Statistician, Census and Statistics Department

Mr Ian NG

Statistician, Census and Statistics Department

Mr Benjamin CHAN

Research Manager, Census and Statistics Department

Natalie, Ian and Benjamin are young professional staff joining the Census and Statistics Department as Statisticians/Research Managers in the past few years and have been working in the area of trade statistics since then. With strong academic background in the fields of statistics and data science, they have been actively engaged in the research on applying big data analytics in processing data from trade declarations so as to improve the efficiency and quality of the trade statistics system of Hong Kong.

Anomaly Detection in Merchandise Trade Data: From Rule-based to Deep Learning Approach

Merchandise trade statistics of Hong Kong, China are compiled by the Census and Statistics Department (C&SD) based on trade declarations lodged by importers/exporters within 14 days after the shipment. While the post-shipment declaration arrangement facilitates trade activities in general, the clarity and accuracy of data declared in relevant trade documents might not be as high as that of similar documents in other places serving customs clearance purposes. To safeguard the quality of trade statistics, C&SD has been adopting for many years a rule-based computer validation system based mainly on numeric data to identify trade declarations that are most susceptible to having reporting errors. This talk presents the results of C&SD's exploration in the application of big data analytics in this quality checking mechanism. It has been shown that the use of deep learning models based on both unstructured textual information and numeric data could bring about impressive improvements in terms of both accuracy and efficiency.

Seminar in commemoration of
55th Anniversary of the Census and Statistics Department and
45th Anniversary of the Hong Kong Statistical Society

Perspectives on uncertainty and error of statistics

Distinguished Speakers

Professor Ian MCKEAGUE

Head and Chair Professor, Department of Biostatistics,
City University of Hong Kong

Statistical Uncertainty and the Humor of Groucho Marx

Professor Tarani CHANDOLA

Professor of Medical Sociology, Director of the Methods Hub,
Faculty of Social Sciences, The University of Hong Kong

Non-response and Missing Data Methods in Census and Survey Data

Miss Natalie CHUNG, Mr Ian NG, Mr Benjamin CHAN

Statistician and Research Manager, Census and Statistics Department

Anomaly Detection in Merchandise Trade Data: From Rule-based to Deep Learning Approach

Date: 11 November 2022 (Friday)
Time: 2:30 p.m. - 4:30 p.m. (Registration starts at 2:00 p.m.)
Venue: Chiang Chen Studio Theatre
The Hong Kong Polytechnic University, Kowloon
(Physical seminar with live streaming)
Language: English

For registration:

For enquiry, please contact Ms Helen LEE (Tel: 3903 7381, Email: 55A_SEMINAR@censtatd.gov.hk)

政府統計處
Census and Statistics Department

香港統計學會
Hong Kong Statistical Society

Professor Ian MCKEAGUE

Head and Chair Professor,
Department of Biostatistics,
City University of Hong Kong

Professor Ian MCKEAGUE has been a Professor of Biostatistics at Columbia University since 2004. In 2021, he joined City University of Hong Kong as Head and Chair Professor in the Department of Biostatistics. His research interests include post-selection inference, empirical likelihood, order-restricted inference, non-standard asymptotics, statistical methods in physical oceanography, functional data analysis, inference for stochastic processes, survival analysis, competing risks models for HIV/AIDS data, Markov chain Monte Carlo and Bayesian methods, efficient estimation for semiparametric models, missing data, counting processes and spatial point processes. Professor McKeague is currently serving on the editorial boards of the Journal of the American Statistical Association, Statistical Science, Statistical Inference for Stochastic Processes, and the International Journal of Biostatistics. During 2020-2023 he is serving as Co-Editor of the Journal of the American Statistical Association. He is a fellow of the Institute of Mathematical Statistics and a fellow of the American Statistical Association.

Statistical Uncertainty and the Humor of Groucho Marx

This is non-technical talk for a general audience. I will start by discussing some aspects of statistical science seen through the lens of Groucho Marx's inimitable brand of humor. More broadly, I will examine the interaction between cultural ideas and statistical ideas in the era of big data. In the sense that "there are no new ideas under the sun," Groucho made subtle use of uncertainty in his art. Uncertainty quantification has of course always been a core challenge for statistics, but our tools are limited. Much of modern data science concerns itself with "point estimates" and single shot reconstructions or predictions, which give the mistaken impression that uncertainty has been vanquished. Groucho knew better.

Professor Tarani CHANDOLA

Professor of Medical Sociology,
Director of the Methods Hub,
Faculty of Social Sciences,
The University of Hong Kong

Tarani is a Professor of Medical Sociology. He is the director of the Methods Hub in the Faculty of the Social Sciences at the University of Hong Kong. He joined the Department of Sociology in August 2021 and was formerly the Head of Department of Social Statistics at the University of Manchester. He is a co-director of the ESRC International Centre for Lifecourse Studies in Society and Health http://www.ucl.ac.uk/icls/, a member of the ESRC Strategic Advisory Network Strategic Advisory Network - Economic and Social Research Council (ukri.org), a fellow of the Academy of Social Sciences https://www.acss.org.uk/ and the Royal Statistical Society https://rss.org.uk/. He obtained his DPhil in Sociology from Nuffield College, University of Oxford in 1998. His research is primarily on the social determinants of health, focusing on health inequalities and psychosocial factors, and the analysis of longitudinal cohort studies.

Non-response and Missing Data Methods in Census and Survey data

Non-response in surveys is one of the biggest concerns for statisticians, potentially leading to less efficient estimates and biased associations. This talk illustrates the different mechanisms that result in missing data in census and survey data and some methods to compensate for the bias generated by such missing data. The use of auxiliary data such as from the census in methods such as weighting for non-response will be discussed. Different methods for compensating for missing data in surveys will be compared, using health as the dependent variable.