香港統計學會
Hong Kong Statistical Society
c/o Department of Statistics & Actuarial Science,
The University of Hong Kong, Pokfulam Road, Hong Kong
http://www.hkss.org.hk

Bulletin
Volume 43 No.1
April 2021

Editor's Foreword

Dear Members,

Welcome to the 2021 issue of the HKSS Bulletin.

It's my honour to serve the HKSS as the Publication Secretary. I would like to express my gratitude to the predecessor, Professor Ben CHAN, for his support and help in preparing this bulletin. I would also want to thank the members of the Editorial Board, Dr Billy LI and Ms Carly LAI for their valuable contributions to the Bulletin.

In this issue, Dr Kent LIU of The Hang Seng University of Hong Kong shares with us the applications and challenges of functional data analysis. Dr Wilson KWAN briefs us the techniques for data visualisation through R. Ms Olivia OR of Census and Statistics Department shares with us the implementation of online questionnaires in the General Household Survey. The Organising Committee of the 2019/20 Statistical Project Competition briefs us the successful completion of the Competition despite the challenge of COVID-19.

I am very sorry to tell you that our life member, Professor Lai Kow CHAN passed away in December 2020. His family members wrote a brief biography of Professor CHAN. A memoir written by Professor Geoffrey TSO of City University of Hong Kong shared with us about Professor CHAN.

Benson LAM

		Phone	Fax	Email
Editor	: Dr LAM, Benson Shu-yan, HSUHK	3963 5450	-	bensonlam@hsu.edu.hk
Secretary	: Dr LI, Billy Yeuk-goat, C&SD	2867 1068	2537 2575	-
Member	: Ms LAI, Carly Yuk-ling, C&SD	2582 4914	2802 1101	yllai@censtatd.gov.hk

CONTENTS

(Vol. 43/No.1, April 2021)

President's Forum

Mr Leo YU Chun-keung

It is a very sad news to our Society and the statistical community that our life member, Professor Lai Kow CHAN, passed away on December 23, 2020. Professor CHAN had contributed much to the Society in the past few decades and he had always actively participated in the Society's events. I trust all HKSS Members would join me in expressing our most sincere condolences to his family. A brief biography of Professor CHAN and a memoir are included in this Bulletin (p.16-19).

To commemorate the third World Statistics Day (20 October 2020) promulgated by the United Nations, HKSS had joined hands with the Education University of Hong Kong (EduHK), the Census and Statistics Department (C&SD) and the Education Bureau (EDB) of the Hong Kong Government to organise an online seminar with the theme "Connecting the World with Data We can Trust" on the special day. The seminar invited three distinguished speakers to share their views on the topic. Professor Philip YU shared the importance of data veracity and the negative impact of fake data. Ms Eva TSUI presented how extreme gradient boosting, a machine learning technique, had been used to develop prognostic models that could effectively predict COVID-19 patients' clinical outcome of hospital admission in Hong Kong. Mr Jimmy LIN introduced various sources of official statistics compiled by C&SD. He also gave an introduction on how the website of C&SD would soon be enhanced with interactivity for disseminating open data.

HKSS and EduHK had also jointly invited Professor Paul EMBRECHTS from ETH Zurich, Switzerland, a renowned researcher in the areas of actuarial/financial mathematics and quantitative risk management, to deliver an online seminar on 19 November 2020. Professor EMBRECHTS explained the importance of rare events in various fields including insurance, finance, medicine, etc. He gave a brief overview of Extreme Value Theory (EVT) for analysing rare events.

Albeit under the unanticipated challenge of COVID-19, the 2019/2020 Statistical Project Competition (SPC) had been completed successfully. Immediately following the outbreak of the disease in early 2020, secondary schools had to suspend their face-to-face teaching. We had to extend the deadline for submission of entries from mid-February to May. The panel interviews for determining the winning teams had to be conducted online for the first time. The Prize Presentation Ceremony had also adopted a different format, in which only the winning teams were invited to attend the Ceremony in person while the entire event was broadcast through live streaming. The successful conduct of the SPC greatly relied on the concerted effort of patrons, sponsors, the Adjudication Panel and the Organising Committee. I would like to express my heartfelt thanks to them.

Last but not least, I wish everyone healthy and safe!

Applications and Challenges of Functional Data Analysis

Dr Kent LIU
The Hang Seng University of Hong Kong

Advances in information technology enable us to collect and process densely observed data over some temporal or spatial domains. The resulting data are coined functional data to differentiate them from the traditional, scalar data. Put simply, functional data are data expressed as functions, such as blood pressure over years, the location of a basketball players during a match, and the stock price during a trading period.

With the availability of functional data, we can answer questions that may not be possible with the scalar variables only. One of the recent and perhaps perplexing issue is whether social distancing could help mitigate the transmission of COVID-19 disease. The mobility data from COVID-19 Community Mobility Report from Google, an example of functional data, help governments or authorities ascertain the association between social interaction and the transmission of the disease, and hence, they may find effective measures to suppress the transmission. By applying a generalised additive model for studying the relationship in COVID-19 mortality rates over time and functional mobility data, scientists discovered links between the mobility and the COVID-19 related mortality (McGough et. al. 2020). They found that declines in mobility were associated with up to 15% lower mortality rates 39 relative to pre-social distancing levels of mobility, but effects were lagged between 25-30 days.

Another application of functional data is the prediction of patient health on the basis of the Sequential Organ Failure Assessment (SOFA) functional data. In the Improving Care of Acute Lung Injury Patients study, scientists investigated the long-term complications of patients who suffered from acute lung injury/acute respiratory distress syndrome (Needham et al. 2006). In the study, data were recorded once the patients were admitted in the ICU, and then daily during hospitalisation. The SOFA scores are daily measurements which measure the overall organ function status of a patient, with higher score suggesting inferior health. Hao et al. (2020) treated the history of each subject's SOFA scores, in the first week, as a functional covariate, where is the number of days since the admission to the ICU and they considered the following Cox's proportional hazards model:

$$h(t) = h_0(t)\exp\left(\theta^T Z + \int_1 X(s)\beta(s)ds\right),$$

where $h_0(\cdot)$ is an unspecified baseline hazard function, Z is a covariate vector, and θ is an unknown parameter, and $\beta(\cdot)$ is an unknown coefficient function. This study suggests that the SOFA scores in last two days in the first week of ICU stay could be used as an indicator of patients' health in the future.

Functional data arise from myriad of areas, but not limited to public health and clinical science. Wherever functional data arises, statistical methods can be adapted to incorporate this kind of data and this necessitates the theoretical studies of functional data analysis. Recently, Crambes, Kneip, and Sarda (2009), Yuan and Cai (2010), Cai and Yuan (2012), Cheng and Shang (2015), and Shang and Cheng (2015), among others, proposed roughness regularisation methods to control the model complexity in a continuous manner. This overcomes the imprecise control on the model complexity due to the truncation parameter in the functional principal component analysis FPCA based approaches, as pointed out by Ramsay and Silverman (2005).

When time-to-event data and functional data are available, the proportional hazards model (Cox 1972) can be modified to integrate these data. Since the proposal of functional Cox model Chen et al. (2011), efforts have been put to the theoretical studies. Recently, Kong et al. (2018) established the rate of convergence of the maximum approximate partial likelihood estimator and conducted a score test for testing the nullity of the slope function related to functional predictors. Qu, Wang, and Wang (2016) studied the asymptotic properties of the maximum partial likelihood estimator under the framework of reproducing kernel Hilbert space and established the asymptotic normality and the efficiency of the estimator of the scalar parameter. Hao et. al. (2020) further studied the asymptotic distribution of the maximum partial likelihood estimator of the slope function and the partial likelihood ratio test.

However, there are many challenges remained to be addressed. The theoretical works of other useful functional models such as functional accelerated failure time models and functional additive hazards models with right-censored data, where a partial likelihood is unavailable, have not been explored yet.

For further exploration of functional data analysis, the interested readers may refer to Ramsay et al. (2009) and Kokoszka and Reimherr (2017) for introductory texts; and Febrero-Bande and de la Fuente (2012) and Goldsmith et al. (2019) for free software packages.

References

Cai, T., and Yuan, M. (2012), "Minimax and Adaptive Prediction for Functional Linear Regression," *Journal of the American Statistical Association*, 107, 1201–1216.

Chen, K., Chen, K., Müller, H., and Wang, J. L. (2011), "Stringing High-Dimensional Data for Functional Analysis," *Journal of the American Statistical Association*, 106, 275–284.

Cheng, G., and Shang, Z. (2015), "Joint Asymptotics for Semi-nonparametric Regression Models with Partially Linear Structure," *The Annals of Statistics*, 43, 1351–1390.

Cox, D. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

Crambes, C.,Kneip,A., and Sarda, P. (2009), "Smoothing Splines Estimators for Functional Linear Regression," *The Annals of Statistics*, 37, 35–72.

Febrero-Bande, M., & de la Fuente, M. (2012). "Statistical Computing in Functional Data Analysis: The R Package fda.usc". *Journal of Statistical Software*, 51(4), 1 - 28.

Goldsmith, J., F. Scheipl, L. Huang, J. Wrobel, C. Di, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, and P. T. Reiss (2019). refund: Regression with Functional Data. *R package* version 0.1-21.

Hao, M., Liu, K., Xu, W., and Zhao, X. (2020). "Semiparametric Inference for the Functional Cox Model." *Journal of the American Statistical Association*, 1-11.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.

Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018), "FLCRM: Functional Linear Cox Regression Model," *Biometrics*, 74, 109–117.

McGough, S. F., Gan, R. W., Tibshirani, R., and Meyer, A. (2020), "Modeling COVID19 Mortality in the US: Community Context and Mobility Matter,", *medRxiv*.

Needham, D. M., Dennison, C. R., Dowdy, D. W., Mendez-Tellez, P. A., Ciesla, N., Desai, S. V., Sevransky, J., Shanholtz, C., Scharfstein, D., Herridge, M. S., and Pronovost, P. J. (2006), "Study Protocol: The Improving Care of Acute Lung Injury Patients (ICAP) study," *Critical Care*, 10, R9.

Qu, S., Wang, J. L., and Wang, X. (2016), "Optimal Estimation for the Functional Cox Model," *The Annals of Statistics*, 44, 1708–1738.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. New York: Springer.

Ramsay, J., and Silverman, B.W. (2005), *Functional Data Analysis*, Springer Series in Statistics (2nd ed.), New York: Springer.

Shang, Z., and Cheng, G. (2015), "Nonparametric Inference in Generalized Functional Linear Models," *The Annals of Statistics*, 43, 1742–1773.

Yuan, M. T., and Cai, T. (2010), "A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression," *The Annals of Statistics*, 38, 3412–3444.

Implementation of Online Questionnaires in the General Household Survey [1]

Ms Olivia OR Oi-yan
Census and Statistics Department

Introduction

The General Household Survey (GHS) conducted by the Census and Statistics Department (C&SD) is a monthly household survey on the characteristics of the labour force and household income. In each month, a random sample of about 9,000 households is enumerated, of which about half are new cases sampled for the first time and half are repeated cases that had been sampled 3 months ago. In the past, data collection for the GHS dealt with repeated cases without telephone numbers (including cases which could not be enumerated and cases which were enumerated but the respondents concerned were unwilling to provide telephone numbers in the first round) and new cases (all without telephone numbers) via face-to-face interviews through field visits, while repeated cases with telephone numbers were enumerated through computer-assisted telephone interviews (CATI).

With the wider penetration and advancement of information technology, adopting online questionnaires (OQ) in statistical surveys has become more feasible in recent years. Meanwhile, households are increasingly difficult to be reached by field visits as people are returning home later from their daily activities and many prefer not to be disturbed at home. Tighter security control in many residential buildings also increases the cost and difficulty in reaching households by field visits. These changes had prompted C&SD to consider introducing OQ as a new self-reporting channel in the GHS to cater for the living styles and preferences of modern households.

This article summarises C&SD's implementation experience of the OQ, with discussions on the review of literatures and precedent relevant activities undertaken prior to the implementation, the approach to implementation and the results of implementation.

Literature review

There are many literatures discussing methodologies and precedent cases of OQ implementation in National Statistical Offices (NSOs), particularly in labour force surveys. In the course of planning for the OQ implementation in the GHS, desktop research had been conducted and references had been drawn from the experiences of other NSOs on the stages of data collection methodology and design, system and questionnaire design, and implementation approach.

---

(1) OR, Olivia Oi-yan. "Implementation of Online Questionnaires in the General Household Survey in Hong Kong", Statistical Journal of the IAOS, vol. 36, no. 4, pp. 987-995, 2020. This is an abridged version and the full publication is available at IOS Press through http://dx.doi.org/10.3233/SJI-200746.

On data collection, as there were already two data collection modes in place in the GHS, C&SD adopted a sequential mixed-mode design in which the OQ is provided to respondents as the first option introduced in the invitation letter that is mailed to the sampled households. If the respondents do not respond to the OQ by a certain date, then the department's field officers will follow up with field visits or CATI (for repeated cases with telephone number available). Such design has several advantages. Firstly, similar designs have been widely studied and adopted internationally and proved to be able to bring improvement to the response rate. Secondly, such design would increase the likelihood of receiving responses from those households that could not be normally reached or interviewed by field visits. Thirdly, starting off with OQ as the first data collection option which is relatively less costly than face-to-face interviews, data collection costs can be reduced with increased participation from sampled households, in line with other preceding cases. Fourthly, studies have suggested that allowing the respondents to choose the mode of responding could bring the benefits of higher data quality with less rounded numerical responses and higher completion.

On system and questionnaire design, references were drawn to similar implementations in other NSOs including the United Kingdom Office for National Statistics, Statistics Canada, the Federal Statistical Office of Germany and Statistics Netherlands to ensure that the computer system interface is user-friendly and visually appealing, and the question flow, question wordings and instructions are suitably tuned to the perspectives of the respondents. In particular, in order not to overburden the respondents to the extent that they would drop out in the middle of completing the OQ, only essential routing and edit checks that are less sophisticated than interviewer-administered modes are included in the OQ for checking on the spot. All other validation checks and follow-ups are deferred to after the completed OQ is submitted.

Implementation approach

The GHS is a monthly survey with data collection workload shared out by nearly 200 field officers. Statistics of moving 3-month periods compiled from the GHS, including size of labour force, unemployment rate, underemployment rate and household income, are vital statistical indicators and are published monthly. It is thus imperative to ensure that data collection remains uninterrupted such that these data series will continue to be available as scheduled without affecting its continuity. On the other hand, adding a new data collection mode to existing ones is operationally and technically complex. If a full-blown implementation approach was adopted, it would take a long time to get all testing completed, operational procedures well tuned and all staff trained up at the same time, and it would be prone to service disruption if unexpected problems with the new system or operations occurred. In particular, in view of the short data collection period of the survey, any hiccups in the new operations might pose high risks to the timely dissemination of the survey results.

In view of the above, it was decided to roll out the OQ on a trial basis first, starting with a small number of cases and then gradually increasing the scale of the trial month by month. The trial was divided into 2 stages. The first stage focused on testing the adequacy of the OQ computer systems and operational arrangements, the effectiveness of the OQ questionnaire design, and the receptiveness of GHS respondents. The second stage involved a gradual rollout of the OQ to representative sub-samples of the GHS sampled cases of increasing size month by month, focusing on studying the take-up rate and impacts on key survey estimates and manpower resources requirement.

The first stage of trial implementation lasted from July 2017 to April 2018 for 10 months. It began with only 80 cases (less than 1% of all sampled cases in a month) in the first month of trial and increased in each subsequent month to 360 cases in the 10th month (about 4% of all sampled cases in a month). The selection of cases was mainly based on the ability of field officers in handling the new operational procedures involving the OQ. Only cases assigned to well-trained field officers were selected at first, while training was extended to the rest of the field officers progressively. A small scale in the first stage also allowed prompt support to be provided to fix any technical or operational problems encountered. Based on the experience and feedback received from both field officers and respondents, operational procedures and system functionalities were refined, such as simplifying the OQ activation page, modifying the question flow, and deriving some data fields automatically based on answers to other questions rather than asking directly for the data.

The second stage of trial implementation commenced in May 2018 and lasted for 8 months up to December 2018. The monthly sample of the GHS consists of 18 random replicates of quarters, each being a representative sample of quarters in Hong Kong. In the first month, one replicate was invited to use the OQ. Introducing the OQ to one out of 18 replicates allowed expanding OQ adoption in a controlled manner. In case something went wrong, the remaining 17 replicates could still be used to compile representative survey results. This replicate-by-replicate approach also allowed comparison of the OQ replicates and non-OQ replicates to understand the impacts of introducing the OQ on key survey estimates. The number of replicates invited to use the OQ then gradually increased every month, until all 18 replicates were invited in the last month of the trial in December 2018.

To ensure that the data collected via this new data collection mode is of as high quality as those collected via the usual modes (i.e. face-to-face interview and CATI), all OQ cases were selected for quality checking during the first stage of trial implementation to identify potential areas for improvement. Towards the second stage of trial implementation, a sampling plan was formulated to select a certain proportion of the OQ cases for quality checking. The plan took into consideration the experience of the first stage of the trial implementation, such as which data items in the submitted OQs are subject to higher risk of reporting errors and thus which categories of the submitted OQs should be subject to a higher rate of checking.

Results of implementation

Error rate and enquiries

The error rate in the OQ data, as measured by the number of OQ cases with errors identified in quality checks over the total number of cases responded by the OQ (including partially responded cases), was in general low (below 5%). They were largely attributed to misinterpretation of the survey concepts and were later remedied by fine-tuning the wording of questions and explanatory notes in the OQ.

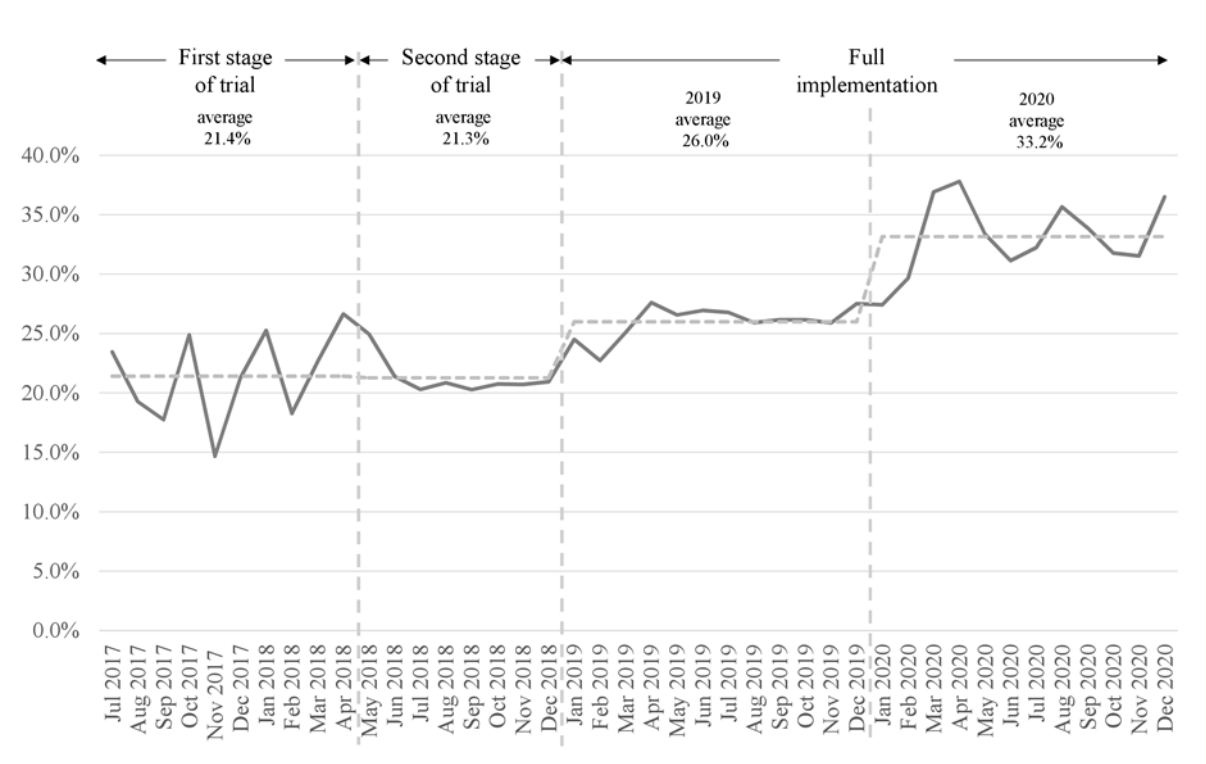
Enquiries received during the trial implementation period also provided valuable information on users' experiences. Throughout the trial implementation, the number of OQ-related enquiries (as a percentage of cases invited to use OQ) had been dropping, from some 3% to a very low level of below 1%. Detailed comments were requested from these enquirers, so that enhancements could be made promptly on the questions, system functionalities and performance.

Take-up of OQ and response rate

The take-up of the OQ was quite satisfactory from the beginning of the trial. Figure 1 shows the overall take-up rate of OQ (in terms of number of cases responded to OQ over all cases invited to use OQ). In the first stage of trial implementation, OQ cases were selected based on field officers, and since each officer handled assignments of a different geographical area, the geographical coverage of cases varied month by month. Therefore, the OQ take-up rate showed fluctuation as some geographical areas with higher elderly population might not use the OQ as much as other areas with more younger households. The average take-up rate was 21.4%, notably higher than the planned level.

The OQ take-up rate remained above 20% towards the second stage of trial implementation, with an average of 21.3%. This was very encouraging in light of the GHS being a voluntary survey with a short OQ reporting period of about 10 days only. The take-up rate was also comparable to those of 20-25% in the Labour Force Surveys of Australia and Canada, which are both mandatory surveys. After full implementation, the OQ take-up rate was further boosted with better communications with sampled households. The average OQ take-up rate in the first year of full implementation reached 26.0%.

Figure 1: OQ take-up rate in GHS during trial implementation and after full implementation (July 2017 – December 2020)



It was noticed during the trial period that the average response rate of OQ invited cases was consistently higher than that of non-OQ cases by 2-3 percentage points. Apparently, the OQ has the added benefit of facilitating data reporting of some sampled households who have difficulties to respond to the GHS under conventional interviewing modes, such as persons being at home only at a late time of the night or weekends (and cannot be reached by field officers), persons not willing to disclose their personal information to strangers face-to-face, and persons feeling intruded when being called or visited by strangers. The OQ helped to improve the overall response rate and resultantly the overall data quality of the GHS.

Impact on GHS statistics

In general, there were no significant differences observed in the demographic profiles of persons and households enumerated between the OQ replicates and the non-OQ replicates. Nevertheless, it is noticed from the responded cases during the 3-month period of August to October 2018 (with similar numbers of sampled cases in the OQ replicates and non-OQ replicates) that respondents' education level in the OQ replicates was significantly (statistically) higher than that in the non-OQ replicates, and households in the OQ replicates had a significantly smaller household size than that in the non-OQ replicates. Although not at a statistically significant level, respondents in the OQ replicates were also observed to have slightly higher personal income than those in the non-OQ replicates. This confirmed that introducing the OQ enabled the C&SD to reach those households who were difficult to be contacted at home and more privacy-concerned, who generally had better education, higher income, and smaller household size. As to households who could be reached by traditional methods, their response pattern would not be affected by the introduction of OQ since the former data collection channels are still in place and households who are not receptive to OQ, such as the elderly, would continue to be reached and enumerated by field visits or CATI.

On the aspect of data quality, literatures suggested that respondents would be more likely to report more accurate answers in self-responding modes than in interviewer-administered modes due to social desirability bias. This bias reduction in responses happens in the GHS as evidenced in, for example, the more precise reporting of income figures in OQ (e.g. HK$17,300 in OQ as compared to HK$17,000 in telephone or face-to-face interview). Moreover, as after the OQ is launched, more respondents who were more receptive to OQ, such as working persons with higher education and persons in one-person household, could be enumerated, the sampling error is lowered for compiled statistics concerning these sub-groups.

Another important aspect is mode effect. As mode effects can affect various types of errors which tend to be confounded with one another, dedicated experiments are required to accurately measure the "pure" mode effects on measurement errors. Due to resources and operational considerations, no embedded experiments were set up in the trial implementation for testing the mode effect. A method similar to the "matching" method proposed by de Leeuw et al. and similar testing done in Statistics Canada was used instead, in which the key statistics compiled from the OQ replicates and the non-OQ replicates were compared to assess if the introduction of the new OQ mode have any impact on the key statistics. Given that these "newly included" respondents who could not be enumerated via field visits prior to OQ introduction are small in number and only account for a small proportion in the total number of responded cases, the overall unemployment rates and underemployment rates compiled with the OQ replicates and the non-OQ replicates in all the moving 3-month periods in the second stage of trial implementation showed no notable differences upon statistical testing.

Concluding remarks

The use of OQ effectively provides an additional convenient and round-the-clock channel for the sampled households to respond to the GHS. With a phased and gradual implementation approach, meticulously designed computer system and operational procedures, attention to data quality, and clear communications with both internal staff and the public, the implementation of OQ in the GHS has been smooth and successful, with no disruption to monthly survey operations and the compilation of statistics.

It is noteworthy that during the recent COVID-19 outbreak in 2020, face-to-face enumeration had to be suspended in the few months when the epidemic situation was worse. With additional efforts put in place to encourage respondents to make use of the OQ to respond to the GHS, including issuing press releases to the public and appeal letters/reminder letters/SMSs to sampled cases with clear and targeted message, the average OQ take-up rate further increased to 33.2% in 2020, which was more than 7% points higher than the average for 2019 (Figure 1). As a result, the overall response rate for the GHS could still be maintained at an acceptable level in face of fieldwork suspension to enable the compilation and dissemination of labour force statistics to continue as scheduled without any interruption.

References

Bethlehem J, Bifflgnandi S. Handbook of web surveys. Hoboken, NJ: John Wiley and Sons. 2012.

Dillman D, Smith J, Christian L. Internet, Mail and Mixed-Mode Surveys – The tailored design method. Hoboken, NJ: John Wiley and Sons. 2009.

Betts P, Lound C. The application of alternative modes of data collection on UK Government social surveys: Literature review and consultation with National Statistical Institutes. London: Office for National Statistics. 2010.

de Leeuw ED, Hox JJ, Dillman DA. International handbook of survey methodology. New York: Taylor & Francis Group, Psychology Press. 2008.

Dillman DA. Phelps G, Tortora R, Swift K, Kohrell J, Berck J, et al. Response rate and measurement differences in mix-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Social Science Research. 2009; 38(a): 1-18.

Roberts C. Mixing modes of data collection in surveys: A methodological review. London: ESRC National Centre for Research Methods. 2007.

Conrad FG, Schober MF, Antoun C, Yan HY, Hupp AL, Johnston M, et al. Respondent mode choice in a smartphone survey. Public Opinion Quarterly. 2017;81(S1):307–37.

Betts P, Cubbon B. Adapting Labour Force Survey questions from interviewer-administered modes for web self-completion in a mixed-mode design. Statistical Journal of the IAOS. 2015; 31 (4): 629 – 44.

Aubrey-Smith SA, Blanke K, Gravem DF, Jarvensivu M, Meertens V, Runz P. Report WP II "Web Data Collection": Testing web questionnaires for the Labour Force Survey in five countries. ESSnet DCSS, Collaboration in Research and Methodology for Official Statistics, European Commission. 2014.

Tourangeau R, Conrad FG, Couper MP. The science of web surveys. Oxford University Press: New York; 2013.

Couper MP. Designing effective web surveys. Cambridge: Cambridge University Press. 2008.

Karaganis M, St-Denis M. Electronic Questionnaire Collection at Statistics Canada. United Nations Economic Commission for Europe. Conference of European Statisticians, Seminar on New Frontiers for Statistical Data Collection. 2012 October 31 – November 2; Geneva, Switzerland.

Betts P, Lound C. The application of alternative modes of data collection on UK Government social surveys: A report for the Government Statistical Service. London: Office for National Statistics. 2010.

References (cont'd)

Jäckle A, Roberts C, Lynn P. Assessing the effect of data collection mode on measurement. International Statistical Review. 2010; 78(1): 3-20.

Hsu, JW, Mcfall BH. Mode effects in mixed-mode economic surveys: Insights from a randomized experiment. Finance and Economics Discussion Series 2015-008. Washington: Board of Governors of the Federal Reserve System. 2015.

Francis J, Laflamme G. Evaluating Web Data Collection in the Canadian Labour Force Survey. Paper presented at the Federal Committee on Statistical Methodology Research Conference. Ottawa. 2015. Available on: https://nces.ed.gov/fcsm/pdf/H2_Francis_2015FCSM.pdf

Australian Bureau of Statistics. Labour Force, Australia, Detailed - Electronic Delivery, Feb 2016: Online Collection in the Labour Force Survey. Australian Bureau of Statistics. 2016. Available on: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6291.0.55.001Main%20Features5Feb%202016

2019/20 SPC for Secondary School Students Successfully Completed Despite COVID-19

Organising Committee* of the 2019/20 Statistical Project Competition

The year 2020 was a challenging year as the local community was severely disrupted by the COVID-19 epidemic. Despite the disruption, the 2019/20 Statistical Project Competition (SPC) for Secondary School Students was successfully completed.

The 34th round of SPC

The 2019/20 SPC was the 34th round since 1986/87. It was jointly organised by the Hong Kong Statistical Society (HKSS) and the Education Bureau, and sponsored by the Hang Seng Indexes Company Limited. The primary objective of HKSS is to encourage secondary school students to understand the local community in a scientific and objective manner through the proper use of statistics, thereby promoting their social awareness and sense of civic responsibilities.

The SPC is divided into two Sections, namely Junior Section for Secondary 1 to 3 students and Senior Section for Secondary 4 to 6 students. Junior Section participants are required to submit their projects in the form of a poster on one of the following themes: population, economic development or transport and housing, while Senior Section participants in the form of a report with their own choices of themes. In addition to the First, Second, Third and Distinguished Prizes, each Section of the Competition also offers the Hang Seng Indexes Company Limited Prize for the Best Index Application and the Department of Management Sciences, City University of Hong Kong Prize for the Best Graphical Presentation of Statistics.

To help interested participants prepare for the Competition, a briefing seminar and an exhibition of past winning projects was held on 26 October 2019 at City University of Hong Kong. The winners of the last round were also invited to share their experiences.

Extension of submission deadline

Affected by COVID-19, the face-to-face lessons of secondary schools had been suspended after the Lunar New Year holiday in January 2020 and only resumed by phase from May 2020, posting great difficulties for school teachers and participating students in discussing and preparing of their statistical projects. In view of the unanticipated situation, the Organising Committee extended the deadline for submission of entries for the Competition from 14 February 2020 to 8 May 2020. Finally, 33 entries and 27 entries were received for the Junior Section and the Senior Section respectively from 32 secondary schools. Demonstrating the diversity of topics, the entries covered various socio-economic issues of Hong Kong.

Adjudication panel led by Dr Geoffrey TSO

An adjudication panel, led by Dr Geoffrey TSO of City University of Hong Kong, who had served as the Chief Adjudicator since the 2015/16 round SPC, and comprised some 24 academics from local tertiary institutions and statisticians working in the Government, was set up for the Competition. Panel members scrutinised all the received projects stringently, shortlisted the more outstanding entries, and interviewed students of the shortlisted projects before determining the winning teams of the various awards. To reduce the risk of infection, this round of panel interviews were conducted through Zoom meetings.

Prize Presentation Ceremony went live streaming the first time

While the winning teams had been determined in July 2020, the Organising Committee decided to postpone the prize presentation ceremony in view of continual threats of COVID-19 and that many secondary schools had already entered into their summer holidays. A Prize Presentation Ceremony of SPC for 2019/2020 cum Briefing Seminar for SPC for 2020/2021 was subsequently held on 26 October 2020 at the Education Bureau Kowloon Tong Education Services Centre. To uphold the social distancing measure to reduce the risk of infecting the COVID-19, only the winning teams were invited to the venue. Simultaneously, the event was broadcast through the Internet for views by other school teachers and students. This was the first time for the SPC prize presentation and briefing seminar went live streaming. Mr Leo YU, President of HKSS and Deputy Commissioner of Census and Statistics Department, was invited to present the prizes to the winning teams.

Regarding the results of the Competition, students of Stewards Pooi Kei College, who used official statistics to study Hong Kong's housing issues, won the First Prize, the Prize for the Best Index Application, and the Best Graphical Presentation of Statistics of the Junior Section. Students of St. Stephen's College won the Second Prize, while students of St. Catherine's School for Girls won the Third Prize.

As for the Senior Section, the statistical report from students of Stewards Pooi Kei College was appraised as the best amongst all the projects. Their winning project studied the retirement of Hong Kong citizens. Students of Good Hope School won the Second Prize, while students of CCC Chuen Yuen College won the Third Prize. Students of HKUGA College won the Prize for the Best Index Application, while students of Stewards Pooi Kei College also won the Prize for the Best Graphical Presentation of Statistics.

► Mr Leo YU had a fruitful exchange with school principals, teachers and participants of the Competition.

Applause, gratitude, and special thanks

The Organising Committee was proud to be able to successfully complete the 2019/20 SPC and would like to give an applause to all the participants of SPC and their teachers, who had demonstrated strong motivation and perseverance in preparing the statistical projects during the hard time of year 2020. We also express sincere gratitude to the patrons of the Competition, Ms Marion CHAN, Commissioner for Census and Statistics, and Mrs HONG CHAN Tsui-wah, Deputy Secretary for Education; the sponsor of the Competition, the Hang Seng Indexes Company Limited; as well as the Department of Management Sciences of City University of Hong Kong for their support to the event. Special thanks to helpers and members of the Adjudication Panel, particularly Dr Geoffrey TSO, who served as the Chief Adjudicator of SPC for five consecutive rounds.

**Organising Committee for the 2019/20 SPC:*

Mr Raymond TSE	Census and Statistics Department
Mr CHAN Sau-tang	Education Bureau
Mr Alex LI	Census and Statistics Department
Miss Carmen LO	Census and Statistics Department
Mr Noel TSANG	Census and Statistics Department
Mr Oliver HO	Census and Statistics Department
Mr Proton NG	Census and Statistics Department
Mr Ronald CHEUNG	Census and Statistics Department

Mr Leo YU, Deputy Commissioner for Census and Statistics and President of HKSS, delivered a speech and presented the prizes to the winning team in the Prize Presentation Ceremony.

In Memory of Lai Kow CHAN

2020 年 **12** 月 **23** 日，陳乃九 ，我們的親愛丈夫、父親、祖父和叔叔，在上帝的懷抱中平靜地去世。他生長在香港，畢業予香港浸會大學，結交一生一直與他親近朋友。他繼續在加拿大西安大略大學獲得統計學博士學位。在畢業期間 ,他的博士生導師讓他去火車站接一個新研究生，研究生便是鳳儀。結婚之後，陳貝、陳鐵 、陳莉按序來臨。

他任教予西安大略大學後，成為馬尼托巴大學統計系主管，在中國世界各地演說，並與本系和國際教授合作。因為熱愛香港，成為香港城市大學管理科學系主任、商業學院院長、澳門科技大學副校長和研究主任。

他結合專業知識、外交、常識和正直人品，贏得了同事、同學們的尊敬。作為 **W.Edwards Deming**（著名工程統計學家）門徒，他在工作和研究過程中一絲不苟，同事和研究生知道他會每事詳情詢問。他把學術研究應用於現實服務，以教學、研究、諮詢服務作為三角框架，指導他的職業原則。

在研究、諮詢服務中 ,他的團隊們的成就包括: 馬尼托巴大學統計質量控制、過程改進講習班 ,收益支持教育研究和研究生; 香港城市大學香港中心城市指數; 澳門科技大學澳門消費者信心指數，海峽兩岸四地成為中國消費者信心指數一部分。

他的人關係很好， 他培養多個博士生，他關心和理解年輕人。在醫院他把醫師，護士和支助人員作為朋友，他，一個教育家，嘗試利用自己的病人經驗改善醫院流程。 多年來，很多人懷念他在他們的專業和個人層面的愛護和關心。

在晚年，他與鳳儀周遊世界，經常與子女、五個孫子女同行。他喜歡海鮮、生果同甜點。他寵愛孫子女，允許他們在早餐吃餡餅，和特別鼓勵他們做好自己。

他是這樣老實、善良、體貼，他的妻子、孩子女、孫子女、大家庭、朋友、 同事都會深深懷念。

請隨時發布一個故事，致敬或照片，這些都會在此站點上共享: **https://www.forevermissed.com/professor-laikow-chan**。

要發送私人慰問，請發送電子郵件至 **rrkrystal@live.com**。

In Memory of Lai Kow CHAN

On December 23, 2020, Lai Kow CHAN, loving husband, father, grandfather, and uncle, passed away peacefully into the arms of God. He was born and raised in Hong Kong and received his undergraduate degree at Hong Kong Baptist University, where he made many friends who would continue to be close to him throughout his life. His studies continued at the University of Western Ontario, where he received his Ph.D. in statistics. Upon graduating, his Ph.D. supervisor asked Lai to pick up a new graduate student at the train station. The graduate student was Fung-Yee. They married, and their children Bertha, David, and Leo came thereafter.

Lai was a popular and successful professor at the University of Western Ontario and later became the Head of the Statistics Department at the University of Manitoba. He also presented at conferences in China and all over the world, collaborating with his department and international colleagues. He loved Hong Kong and returned to his homeland when he became the Head of the Department of Management Sciences, Dean of Faculty of Business at City University of Hong Kong, and later, Vice Rector and Research Director at Macau University of Science and Technology.

With his unique combination of expertise, diplomacy, common sense, and integrity, Lai earned the respect of his colleagues and his students. As a disciple of W. Edwards Deming (famous engineering statistician), he was meticulous in his work and research process to the chagrin of his colleagues and graduate students, who knew to expect many questions from him. Seeking to apply academic study to real-life situations, Lai developed the research, teaching, and consulting triangle framework, which served as the guiding principle for his professional life.

In research and consulting, his teams' accomplishments include the Statistical Quality Control and Process Improvement consulting workshops at the University of Manitoba, where the proceeds supported research and graduate students; the development of the Hong Kong Centa-City Index at City of University of Hong Kong; the development of the Macao Consumer Confidence Index, which is part of the Chinese Consumer Confidence Index: Four Places across Strait, while at Macau University of Science and Technology.

Lai was a people person throughout his life. He nurtured many Ph.D. students. He cared deeply about young people and understood them. Even in the hospital, he befriended doctors, nurses, and support staff. From his bedside, he continued his role as an educator, attempting to use his own experience as a patient to improve hospital processes. Many who encountered Lai over the years remember how he genuinely cared about them and helped them on a professional and personal level.

Throughout life and in his later years, Lai travelled the world with Fung-Yee by his side, and often with his children and five grandchildren. He loved cruises and enjoyed eating, especially seafood, fruit, and desserts. He liked spoiling his grandchildren, letting them eat pie for breakfast, and notably encouraging them to be themselves.

He was so honest, kind, and thoughtful, he will be deeply missed by his wife, children, grandchildren, extended family, and friends and colleagues.

Please feel free to share a tribute, memory, or story about Lai : https://www.forevermissed.com/professor-laikow-chan.

To send private condolences, please email rrkrystal@live.com.

In memory of Professor Lai Kow CHAN

By Geoffrey TSO of City University of Hong Kong

It is sad to learn that Professor Lai Kow CHAN passed away on December 23, 2020. Professor CHAN is a life member of the Hong Kong Statistical Society. He had participated in many Society's events in the past few decades. Professor CHAN was born in in Hong Kong in 1940; he received his undergraduate degree at Hong Kong Baptist University, then his master and PhD degrees at the University of Western Ontario and stayed there as faculty member. He was a popular and successful professor at the University of Western Ontario, and later became the Head of the Statistics Department at the University of Manitoba. In 1994, Professor CHAN returned to Hong Kong and joined City University of Hong Kong as Chair Professor. He served as the Head of the Department of Management Sciences and Dean of the College of Business. He retired from the CityU in 2006 and joined the Macau University of Science and Technology as Director/Professor of the Institute for Sustainable Development and Vice Rector. Professor CHAN is a Fellow of many prestige organisations including American Statistical Association, American Society for Quality Control, American Association for the Advancement of Science, and Institute of Mathematical Statistics. He is also an Elected Member of International Statistical Institute.

Professor CHAN always encouraged his colleagues to work on all three areas of Teaching, Research, and Knowledge Transfer, and used them as guiding principle for staff development. He established the Statistical Quality Control and Process Improvement consulting workshops at the University of Manitoba. Under his leadership, the Department of Management Sciences has developed the Centa-City Index with Centaline Property. He maintained good connection and collaboration with many universities in China, and first produced the Macao Consumer Confidence Index, as part of the Chinese Consumer Confidence Index: Four Places (mainland China, Macau, Hong Kong, Taiwan) across Strait. He genuinely cared about the students' growth in both professional and personal level, and so become a life-long mentor for many graduated students. He often referred to the Deming cycle to stimulate and encourage students for their continuous improvement. He is a thoughtful educator in his whole life. He will be deeply missed by his family, friends, students and colleagues.

LATEST NEWS

News Section

◆ Personnel Changes (New Appointments, Promotions and Retirements)

※ Dr Michael ZHANG joined the Department of Statistics and Actuarial Science of The University of Hong Kong as Assistant Professor in January 2021 .

※ Dr HAN Ruijian and Dr ZHU Huichen joined the Department of Statistics of The Chinese University of Hong Kong (CUHK) as Research Assistant Professor with effect from August 2020.

※ Dr CHAN Chun-man and Dr CHEUNG King-chau joined the Department of Statistics of CUHK as Lecturer with effect from January 2020 and September 2020 respectively.

※ Dr Jingyu HE and Dr Simon TRIMBORN joined the Department of Management Science (Business Statistics) of City University of Hong Kong (CityU) as Assistant Professor with effect from August 2020.

※ Professor Alan WAN has been appointed as Head of the Department of Management Science (Business Statistics) of CityU with effect from November 2020.

※ Professor Philip YU joined the Department of Mathematics and Information Technology of the Education University of Hong Kong (EdUHK) as Professor with effect from September 2020.

※ Dr FU Hong and Dr LO Chung-kwan joined the Department of Mathematics and Information Technology of EdUHK as Assistant Professor with effect from September 2020.

※ Dr LING Man-ho Alpha, Dr POON Kin-man Leonard, and Dr SONG Yanjie of Department of Mathematics and Information Technology of EdUHK advanced to Associate Professor with effect from July 2020.

※ Dr Becky LEE of Department of Mathematics, Statistics and Insurance of The Hang Seng University of Hong Kong advanced to Associate Professor in July 2020.

LATEST NEWS

◆ Seminar on "The Statistical Analysis of Extreme Events"

The seminar was held on 19 November 2020 and organised by The Education University of Hong Kong and Hong Kong Statistical Society. Professor Dr Paul EMBRECHTS explained the importance of small probability outcomes in various insurance finance, the environment, engineering, and medicine.

◆ Seminar on "Connecting The World With Data We Can Trust" on World Statistics Day

The United Nations General Assembly celebrates World Statistics Day (WSD) every five years and 20 October 2020 was designated as the third WSD. The theme of WSD 2020 was "Connecting the World with Data We can Trust" to reflect on the importance of trust, authoritative data, innovation and the public good in national statistical systems.

To celebrate WSD 2020, the Hong Kong Statistical Society (HKSS) jointly organised a public virtual lecture on "Connecting the World with Data We can Trust" with Faculty of Liberal Arts and Social Sciences of the Education University of Hong Kong, Census and Statistics Department (C&SD), and Education Bureau of the Hong Kong Government. The public virtual lecture was delivered via Zoom webinar and received favourable responses with 356 participants in total.

Mr Leo YU, the President of HKSS and Deputy Commissioner of C&SD, gave an opening speech at the public virtual lecture and introduced the three invited speakers.

Mr Leo YU welcoming guests and participants in the public virtual lecture.

Three talks were presented: (1) “Data Veracity”, delivered by Professor Philip YU, Professor at the Department of Mathematics and Information Technology of the Education University of Hong Kong; (2) “Development and Application of a Data-driven COVID-19 Prognostication Tool in Hong Kong”, delivered by Ms Eva TSUI, the former Chief Manager (Statistics and Data Science Department) of Hospital Authority (HA), and (3) “Latest Development in Official Statistics Dissemination Services”, delivered by Mr Jimmy LIN, a Senior Statistician in C&SD.

Professor Philip YU (Left), Ms Eva TSUI (Middle), and Mr Jimmy LIN (Right) giving their talks in the public virtual lecture.

Group photo of the guests of honour and speakers (from left):
Mr Leo YU, Professor May WONG, Mr Jimmy LIN, Ms Eva TSUI, Professor Philip YU, Ms Marion CHAN and Mr Thomas CHEUNG

To recapitulate the public lecture again, please visit the website: https://hkss.org.hk/index.php/events/list/academic/event/83-seminar-on-connecting-the-world-with-data-we-can-trust-on-world-statistics-day.

Workshop on "Exploring Data Visualisation Through R"

Dr Wilson KWAN Chun-kit
The Hong Kong Polytechnic University - Hong Kong Community College

Data visualisation has been widely used for helping decision making in various domains such as major revenue of an industrial company. The statistical software R provides a range of data visualisation tools including boxplots, probability density plot, correlation plots...

To promote this statistical software R to secondary school, a workshop namely "Exploring Data Visualisation through R" was successfully co-arranged with EDB on 8 January 2021 via online platform Zoom. This workshop aims to promote visualising statistical information in various dimensions and formats to local secondary school teachers. Over 80 secondary school teachers participated the online workshop.

In the workshop, the built-in "iris" flower dataset was adopted to illustrate the visualisation ability of R. This dataset is very famous as it is widely used as an essential introductory dataset in machine learning and statistics. The dataset contains 150 observations of iris flowers. There are four attributes of measurements (sepal length, sepal width, petal length and petal width) for iris flowers in centimeters. The fifth attribute is the species of the flower observed. There are only three species (setosa, versicolor and virginica) in the dataset. All observed flowers belong to one of these three species.

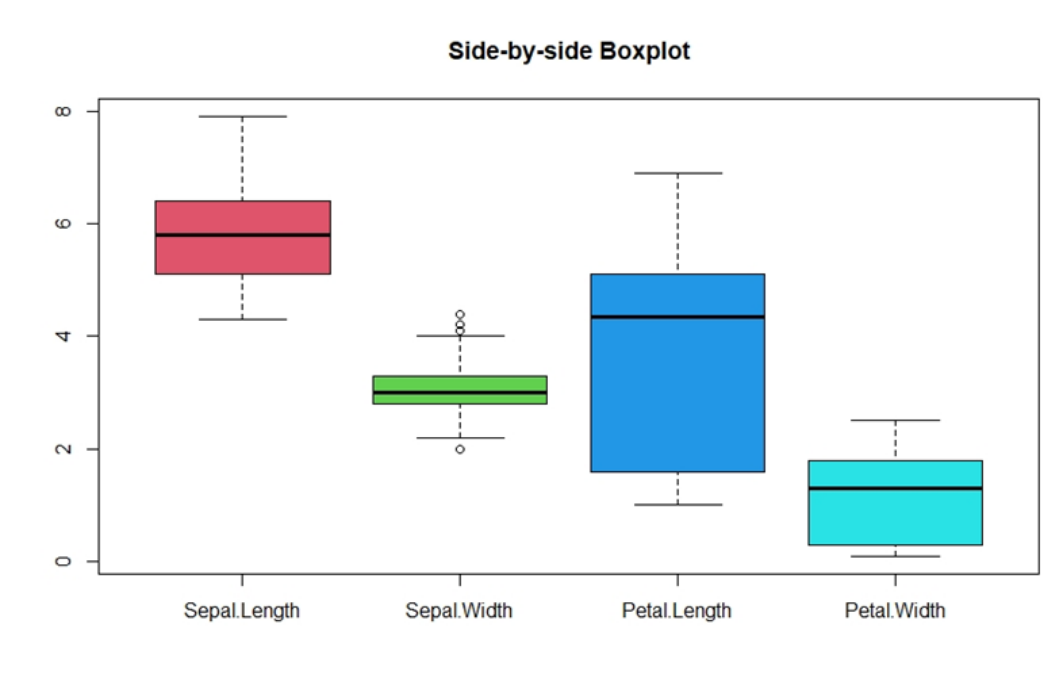


Figure 1

Firstly, a box and whisker plot for each attribute is plotted by using R for visualising the distribution of different measurements of the dataset. The box plot is shown in figure 1. Based on the boxplots in the figure, it can be observed that each measurement has a different distribution. To have a more depth understanding of the distribution for each species, anther boxplot with different attributes among different species is plotted in figure 2.

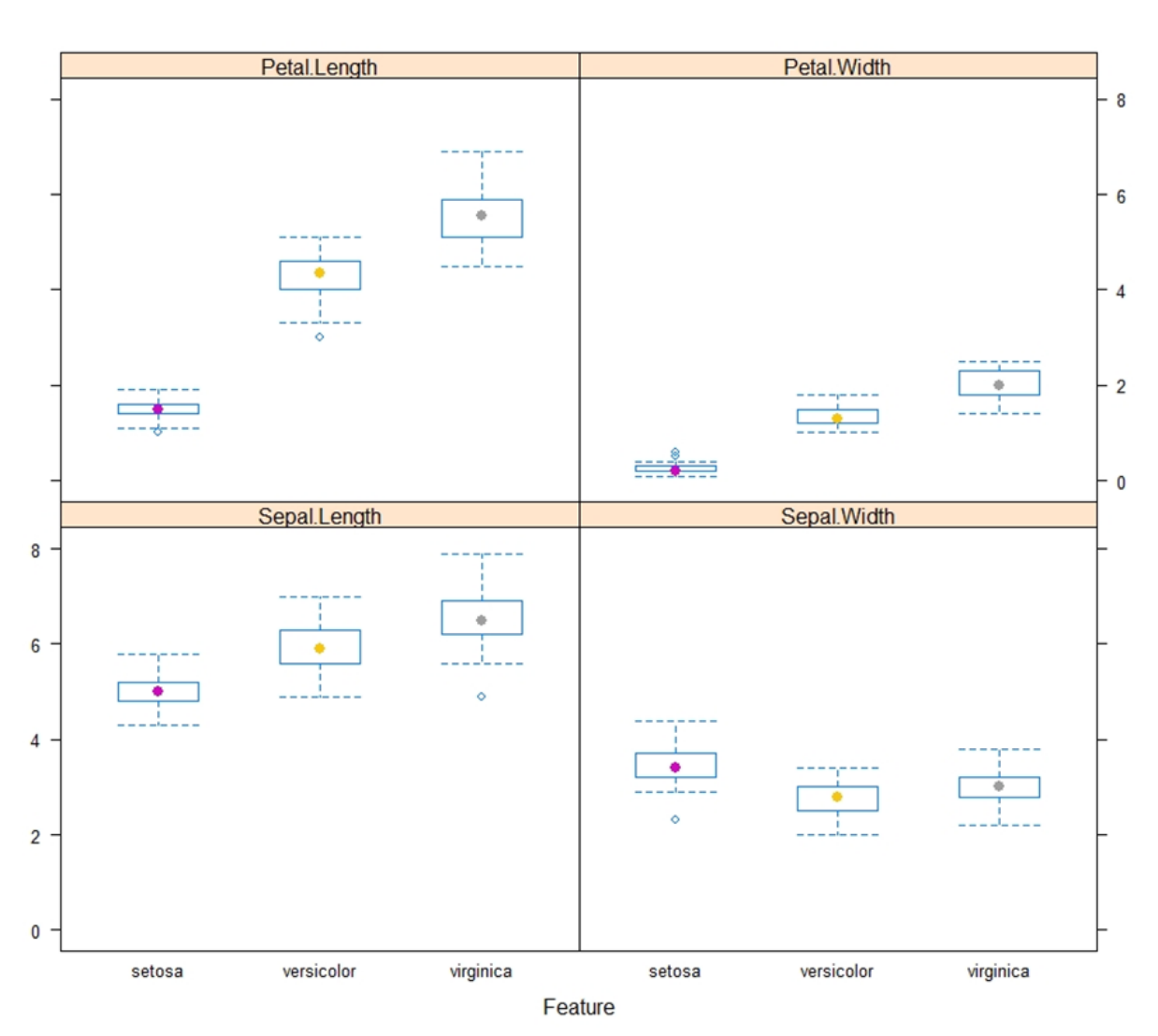


Figure 2

We can also use probability density plot to explore the idea of distribution for each attribute. A probability density plot is illustrated in Figure 3. The colors of the density line for “setosa”, “versicolor” and “virginica” are blue, purple and green respectively.

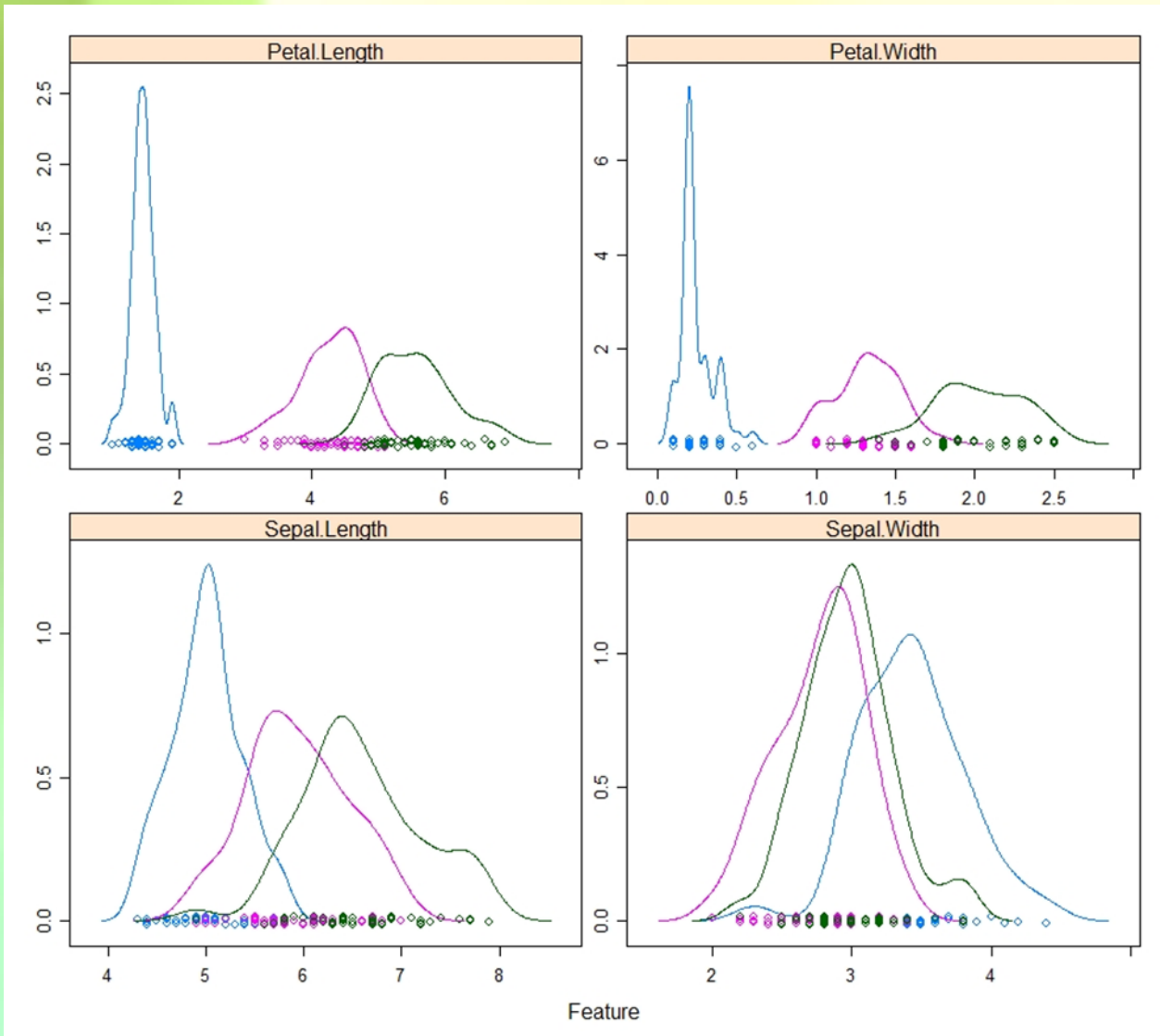


Figure 3

Based on the figure 3, different species are having a different distribution each attribute. Figure 4 illustrates that petal length and petal width among those three species “setosa”, “versicolor” and “virginica” in general has a positive correlation in its values.

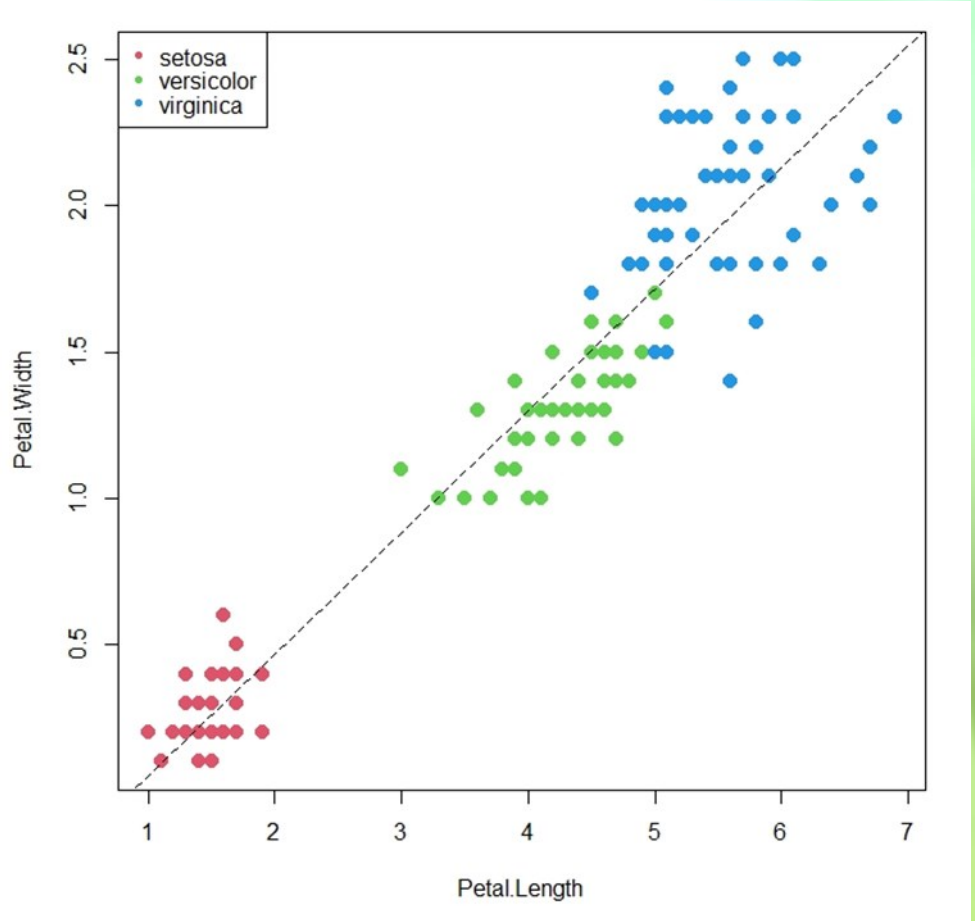


Figure 4

Figure 5 demonstrates the correlation between any two attributes of measurements among all three species of iris flowers. Each species is grouped in an ellipse with the same color in the correlation plot. The colors of ellipse for “setosa”, “versicolor” and “virginica” are blue, purple and green respectively. It can be observed that each species is clustered in the correlation plots.

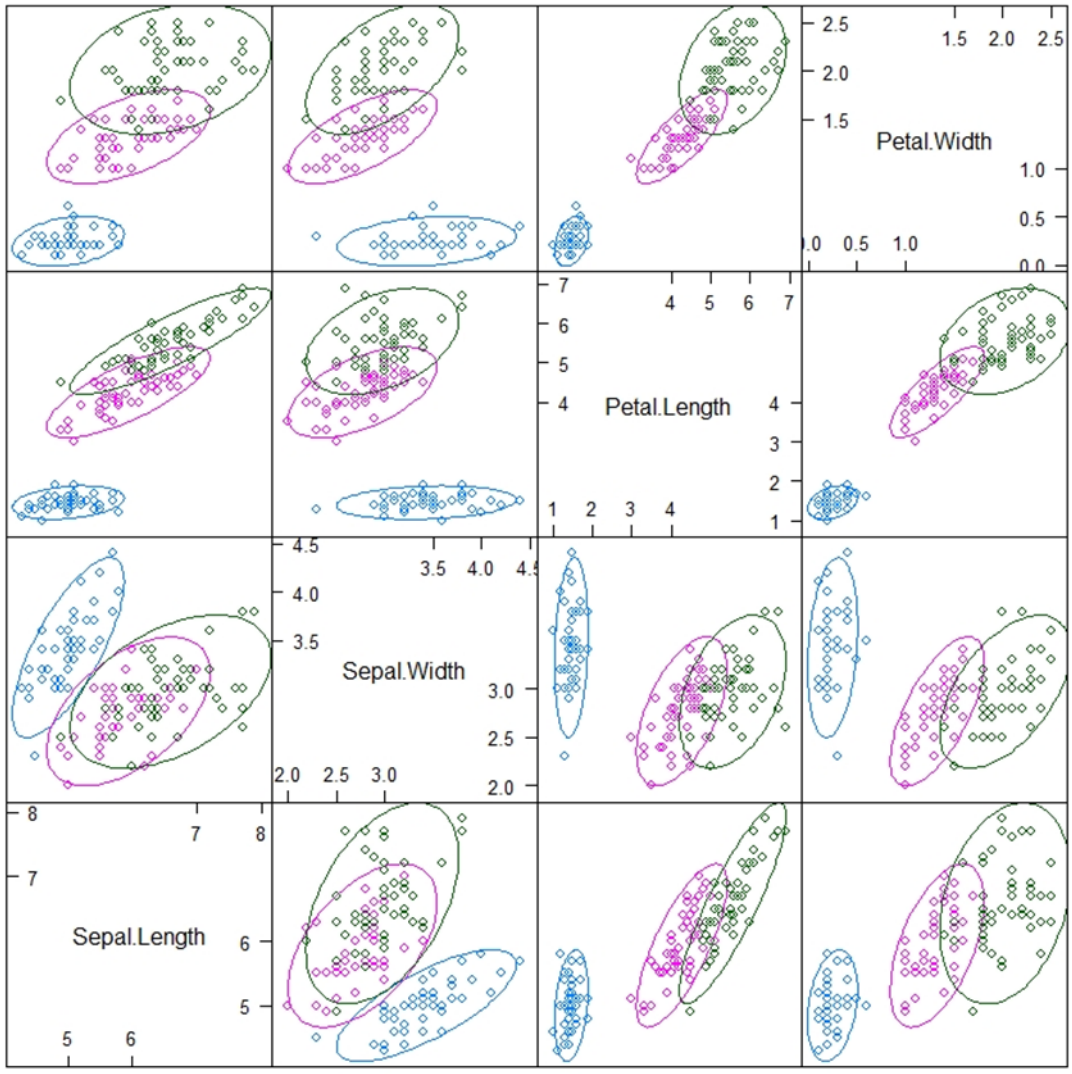


Figure 5

In the workshop, participants had a chance to gain hands-on experience in exploring the iris data set through various graphs of presentation. The feedback from participants were positive and indicated that the workshop could inspire them for initiative and stylised presentation of data in a visualised manner.