Japanese Journal of Statistics and Data Science (2022) 5:275–277
https://doi.org/10.1007/s42081-022-00166-y

PREFACE

Special feature: statistics for COVID-19 pandemic data

Koji Kurihara[1]

Published online: 2 June 2022

The eighth special feature of the Japanese Journal of Statistics and Data Science (JJSD) focuses on coronavirus disease 2019 (COVID-19) pandemic data from the viewpoint of statistics and data sciences. The COVID-19 coronavirus is an infectious disease which was first identified in December 2019 and continues to spread worldwide, and as of May 23, 2022, the number of confirmed cases was 526 M, and the number of deaths was 6.28 M. The COVID-19 crisis has generated a tremendous impact on our health and entire economy affecting all areas. New-type research on COVID-19 has been rapidly progressing and related data are being stored, shared, and opened day by day.

A statistical investigative cycle usually consists of five stages (PPDAC) in data analysis: (1) Problem (define problems), (2) Planning (study design), (3) Data (data collection), (4) Analysis (data exploration), and (5) Conclusion (interpretation). Data science is often understood as a task-driven and computationally oriented version of statistics (Carmichael & Marron, 2018). Problem solving and value creation of the target fields are required at the stage of conclusion in data science.

In this special feature, we have the following six original papers presented by statisticians and data scientists. The first four papers are about new statistical methods and algorithms to investigate spatio-temporal data in view of trend components, closeness, clustering of multivariate data and detection of clusters of higher risk of COVID-19 data. The last two papers are mainly concerned with optimal prevention strategy and the relationship between air pollution and COVID-19 hospitalization for problem-solving and value creation regarding COVID19.

Kumar et al. (2021) apply an autoregressive model with a linear spline trend component for COVID-19. The spline function splits series of COVID-19 into different piecewise segments between respective knots in the form of various growth stages. The number of knots was selected by considering the Bayes factor, and the location of knots and other parameters were estimated by the MCMC sampling method. The results advocate that the proposed model appropriately determines the location of knots based on different transmission stages and knowing the current transmission situation of the COVID-19 pandemic in a country.

✉ Koji Kurihara
kurihark@kyoto-wu.ac.jp

[1] Institute of Data Science, Kyoto Women's University, Kyoto, Japan

Matsui et al. (2021) design a practical algorithm for testing the closeness of sequential data by combining distribution testing and Markov chain testing (Wolfer & Kontorovich, 2020). For a given city, the proportion of the population infected with COVID-19 is recorded at specified time-intervals and it is assumed that this time series forms a Markovian process, whose dynamics are governed by some unknown transition kernel. The paper shows whether or not the epidemic evolves in the same way in different cities or in different months or weeks with numerical indicators of acceptance and reject probabilities, and significance levels.

Watanabe (2022) proposes a *k*-means method style clustering algorithm for trends of multivariate time series. The centroid is defined as a common trend and a dissimilarity measure is also introduced for trends. Based on these centroids and dissimilarity measures a *k*-means method style algorithm is proposed for a multivariate trend. In the article, Watanabe (2021) is extended to provide a new definition of a common trend by considering lags for multivariate time series. The proposed method is applied to the time series of COVID-19 cases in each prefecture of Japan. They also discuss that the clustering of original nonstationary time series themselves does not trend briefly using COVID-19 series.

Takemura et al. (2022) propose a new technique to detect regions, called clusters, which pose a significantly higher risk of infection than their surrounding areas, based on spatial scan statistics for the daily number of COVID-19-infected people of each prefecture in Japan. They focused on the hierarchical structure of the data obtained by conducting an echelon analysis (Kurihara et al., 2020; Myers et al., 1997) and applied a space–time cluster detection method based on this structure to enable the capture of changes in a cluster's shape. Furthermore, they visualized the location and period of a cluster's occurrence and considered the cause of the cluster.

Fujita et al. (2022) propose a direction for the utilization of multi-agent simulation (MAS) to consider an optimal prevention strategy for the spread of COVID-19 through a pandemic modeling example in Japan. They construct a spread simulation in a MAS environment that represented a Japanese metropolitan city by considering differences in background information such as age group and household. They present the possibility of offering various recommendations for optimal strategies to suppress a pandemic by combining reinforcement learning (Kompella et al., 2020) with MAS.

Alsaber et al. (2022) investigate the validity of the relationship between air pollution and COVID-19 hospitalization based on time-series approaches. The time series research was carried out in the state of Kuwait. They checked some statistical tests on multivariate time-series using the Vector Error Correction Model (VECM) technique. It is revealed that the concentration rate of some air pollutants influences COVID-19 admitted cases via Granger-cause.

We are particularly grateful to all referees for their efforts and contributions in the reviewing process with useful comments and suggestions based on their knowledge. We would like to thank all authors who have presented original and novel work to this special feature.

References

Alsaber, A. R., Setiya, P., Al-Sultan, A. T., & Pan, J. (2022). Exploring the impact of air pollution on COVID-19 admitted cases, *Japanese Journal of Statistics and Data Science*, *5* (**to appear**).

Carmichael, I., & Marron, J. S. (2018). Data science vs. statistics: two cultures? *Japanese Journal of Statistics and Data Science, 1*, 117–138.

Fujita, S., Kiguchi, R., Yoshida, Y., & Kitanishi, Y. (2022). Determination of optimal prevention strategy for COVID-19 based on multi-agent simulation, *Japanese Journal of Statistics and Data Science*, *5* (to appear).

Kompella, V. et al. (2020). Reinforcement Learning for Optimization of COVID-19 Mitigation policies. arXiv preprint. https://arxiv.org/abs/2010.10560

Kumar, J., Agiwal, V., & Yau, C. Y. (2021). Study of the trend pattern of COVID-19 using spline-based time series model: A Bayesian paradigm. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-021-00127-x

Kurihara, K., Ishioka, F., & Kajinishi, S. (2020). Spatial and temporal clustering based on the echelon scan technique and software analysis. *Japanese Journal of Statistics and Data Science, 3*, 313–332.

Matsui, T., Azzaoui, N., & Murakami, D. (2021). Analysis of COVID-19 evolution based on testing closeness of sequential data. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-021-00144-w

Myers, W. L., Patil, G. P., & Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics, 4*(2), 131–152.

Takemura, Y., Ishioka, F., & Kurihara, K. (2022). Detection of space-time clusters using a topological hierarchy for geospatial data on COVID-19 in Japan. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-022-00159-x

Watanabe, N. (2021). Dissimilarity measures for time series and trend analysis: Application to COVID-19 cases series. *Journal of Mathematics and Systems Sciences*, *11*, 1–12.

Watanabe, N. (2022). A k-means method for trends of time series: an application to time series of COVID-19 cases in Japan. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-022-00148-0

Wolfer, G., & Kontorovich, A. (2020). Minimax testing of identity to a reference ergodic markov chain. In International Conference on Artificial Intelligence and Statistics, 191–201, 2020. http://proceedings.mlr.press/v108/wolfer20a.html