

Risk of mortality from infecting COVID-19 is age sensitive and is likely to be gender sensitive*

Ziyu Jiang

2022/4/3

Abstract

This report presents an analysis of the City of Toronto's data on the outcomes of the cases of infection of Toronto residents. The results indicate that patients' age and gender affect their risk of mortality from infecting COVID-19. The analysis is conducted with logistic regression model created with the statistical programming language R. The results reinforce our understanding of the probable variables related to the consequences of COVID-19 infections, which provides new probable ideas of preventing COVID-19 infections from the social perspective.

Keywords: COVID-19, COVID-19 cases Toronto, COVID-19 risk of mortality, COVID-19 gender impacts, COVID-19 age impacts, toronto residents, open data toronto

Contents

1 Introduction	2
2 Data	2
3 Methodology	2
4 Result	2
5 Discussion	2
References	3

*Code and data are available in this GitHub repository: [Owlll11/Analysis-about-Covid-19-cases](https://github.com/Owlll11/Analysis-about-Covid-19-cases).

1 Introduction

During the two years of COVID-19 pandemic, people are getting more and more worried about the consequences of infecting COVID-19. It was being said that the COVID-19 pandemic cases happened on children with a markedly low proportion than people in other age groups (Davies et al. 2020). At the same time, it was said that there are increased number of cases and higher risk of severe disease for people with higher ages for infecting COVID-19 (Davies et al. 2020). These suggests that people with higher age should pay more attention on preventing COVID-19 infection. Now a similar question has occurred after the conclusion of age sensitive for COVID-19: Would COVID-19 also be gender sensitive? Gender-related factors may cause differences for an individual's likelihood of exposure to COVID-19, but they may also influence whether an individual tries to acquire a test and whether they are given one (Tadiri et al. 2020). Analyzing this topic would help people have some ideas about society impacts on COVID-19 infection while they tried to have prevention from it.

After analyzing the data in the certain period of time, we can see that the patients' age and gender would have some effects on the symptoms and final results from infecting COVID-19. In this report, we will use the COVID-19 Cases in Toronto data from Toronto Open Data Portal (Gelfand 2020) to investigate the relationship between the final results of infecting COVID-19 and other independent variables, such as age and gender. We will discuss the limitation of the data and the potential bias from the data, apply the logistic regression model on the data. At last, we will discuss the distributions of the independent variables in interests, the resulting influences of the final model, the weakness of our researching processes and the lessons learned from the research and the possible improvement in future.

R statistical programming language(R Core Team 2021) is used for analyzing, and the package `tidyverse`(Wickham et al. 2019) is used for data visualizing and data manipulating in this project. The package `broom`(Robinson, Hayes, and Couch 2022) is also used in this project to convert the statistical objects into tidy tibbles, the package `knitr`(Xie 2014) is used to knit the R markdown file to pdf form and to create the tables, and the package `ggplot2`(Wickham 2016) is used for creating graphs. The packages `dplyr`(Wickham et al. 2022) and `janitor`(Firke 2021) are used to proceed the data cleaning processes. Finally, the package `car`(Fox and Weisberg 2019) is used to creating regression models of the variables.

2 Data

3 Methodology

4 Result

5 Discussion

References

- Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, and Rosalind M. Eggo. 2020. *Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics*. <https://www.nature.com/articles/s41591-020-0962-9>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robinson, David, Alex Hayes, and Simon Couch. 2022. *Broom: Convert Statistical Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.
- Tadiri, Christina P., Teresa Gisinger, Alexandra Kautzky-Willer, Karolina Kublickiene, Maria Trinidad Herrero, Valeria Raparelli, Louise Pilote, and Colleen M. Norris. 2020. *The Influence of Sex and Gender Domains on COVID-19 Cases and Mortality*. <https://www.cmaj.ca/content/192/36/E1041>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.