

Data Sheet for Systolic Blood Pressure (SBP) Dataset*

INF312 Worlds Become Data: Tutorial 10

Mingjia Chen

March 19, 2024

For the purpose of this tutorial, the hypothetical datasheet is constructed based on the textbook by Wickham et al. (2019) and journal article by Gebru et al. (2021), using open source statistically programming language R (R Core Team 2023). The dataset includes systolic blood pressure and other health factor, and is provided by University of Toronto Scarborough course called STAC67 Regression Analysis. I got to obtain and analyze the dataset which is based on the medical scientific journal by Haji-Maghsoudi et al. (2021) for my last degree ¹.

1 Motivation

2 Composition

3 Collection Process

4 Recommended Uses

*Code and data are available at: <https://github.com/MjChen120/INF312Tutorial10.git>. Thanks to Catherine Catherine Punnoose for feedbacks.

¹https://github.com/MjChen120/Past_Research/tree/main/R/STAC67_Research_Analysis_Paper

References

- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. “Datasheets for Datasets.” *Communications of the ACM* 64 (12): 86–92.
- Haji-Maghsoudi, Saiedeh, Azadeh Mozayani Monfared, Majid Sadeghifar, Ghodratollah Roshanaei, and Hossein Mahjub. 2021. “Factors Affecting Systolic Blood Pressure Trajectory in Low and High Activity Conditions.” *Medical Journal of The Islamic Republic of Iran*, April. <https://doi.org/10.47176/mjiri.35.95>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.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>.