

Como implementar algunos modelos de imputación múltiple para datos de panel

Anónimo

Abstract Los datos faltantes son todo un reto en los análisis estadísticos porque los resultados que arrojan tienen limitaciones. La imputación, entendida como el proceso de reemplazar los datos faltantes con un valor estimado, es un problema regular en los proyectos de investigación. Existen muchos modelos y paqueterías destinadas para este proceso, sin embargo, la selección del modelo de imputación adecuado al tipo de datos disponibles es trascendental para la fiabilidad del resultado. En este estudio se trabaja con una tabla de datos cruzada que involucran series de tiempo (datos panel) para 33 países y 17 variables (Índice de Gini anual para período 2000-2016), con un 24 % de datos faltantes. Con el objetivo de imputar estos datos, se utilizó un modelo de imputación múltiple propuesto por Honaker y King (2010) y se agregaron algunas restricciones al sistema. Los principales resultados obtenidos, conducen a la siguiente interrogante: ¿Se puede confiar en la imputación? Todos los archivos necesarios para reproducir los resultados presentados están disponibles en: <https://gitlab.com/iesta.fcea.udelar/como-implementar-algunos-modelos-de-imputacion-multiple-para-datos-de-panel>.

Palabras clave: Datos faltantes - Datos panel - Imputación

Referencias

- Arellano, Manuel, and Olympia Bover. 1990. "La Econometría de Datos Panel." *Investigaciones Económicas XIV* (1): 3–45. <https://www.fundacionsepi.es/investigacion/revistas/paperArchive/Ene1990/v14i1a1.pdf>.
- Bell, Melanie L, Mallorie Fiero, Nicholas J Horton, and Chiu-Hsieh Hsu. 2014. "Handling Missing Data in RCTs; a Review of the Top Medical Journals," 8.
- Honaker, H., G. King, and M. Blackwell. 2018. "AMELIA II."
- Honaker, James, and Gary King. 2010. "What to Do About Missing Values in TimeSeries CrossSection Data." *American Journal of Political Science* 54 (2): 561–81. <https://doi.org/https://doi.org/10.1111/j.1540-5907.2010.00447.x>.
- Kossen, Tabea, Michelle Livne, Vince I Madai, Ivana Galinovic, Dietmar Frey, and Jochen B Fiebach. 2019. "A Framework for Testing Different Imputation Methods for Tabular Datasets." *bioRxiv*, January, 773762. <https://doi.org/10.1101/773762>.
- Leite, W., and N. Beretvas. 2010. "The Performance of Multiple Imputation for Likert-Type Items with Missing Data." *Journal of Modern Applied Statistical Methods* 9: 64–74. <https://doi.org/10.22237/jmasm/1272686820>.
- Madley-Dowd, P., R. Hughes, K. Tilling, and J. Heron. 2019. "The Proportion of Missing Data Should Not Be Used to Guide Decisions on Multiple Imputation." *Journal of Clinical Epidemiology* 110: 63–73. <https://doi.org/https://doi.org/10.1016/j.jclinepi.2019.02.016>.
- Medina, F., and M. Galván. 2007. "Imputación de Datos: Teoría Y Práctica." Publicación de las Naciones Unidas: CEPAL. https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf.
- Muñoz-Rosas, Juan Francisco, and Encarnación Álvarez-Verdejo. 2009. "Métodos de Imputación Para El Tratamiento de Datos Faltantes: Aplicación Mediante R/Splus." *Revista de Métodos Cuantitativos Para La Economía Y La Empresa*, 3–30. <http://www.upo.es/RevMetCuant/art25.pdf>.
- Murray, Jared S. 2018. "Multiple Imputation: A Review of Practical and Theoretical Findings." *Statistical Science* 33 (2): 142–59. <https://doi.org/10.1214/18-STS644>.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. USA: John Wiley & Sons, Inc. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316696.fmatter>.
- Schafer, Joseph L., and John W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7 (2): 147–77.
- Takahashi, M. 2017. "Multiple Ratio Imputation by the EMB Algorithm: Theory and Simulation." *Journal of Modern*

Applied Statistical Methods 16 (1): 630–56. <https://doi.org/10.22237/jmasm/1493598840>.

Van Buuren, S., and K Groothuis-Oudshoorn. 2011. “Mice: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software, Articles* 45 (3): 1–67. <https://doi.org/10.18637/jss.v045.i03>.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. Segunda. Florida, USA: Taylor & Francis. <https://stefvanbuuren.name/fimd/>.

Wood, Angela M, Ian R White, and Simon G Thompson. 2004. “Are Missing Outcome Data Adequately Handled? A Review of Published Randomized Controlled Trials in Major Medical Journals.” *Clinical Trials: Journal of the Society for Clinical Trials* 1 (4): 368–76. <https://doi.org/10.1191/1740774504cn032oa>.

Yamaguchi, Yusuke, Mai Ueno, Kazushi Maruo, and Masahiko Goshio. 2020. “Multiple Imputation for Longitudinal Data in the Presence of Heteroscedasticity Between Treatment Groups.” *Journal of Biopharmaceutical Statistics* 30 (1): 178–96. <https://doi.org/10.1080/10543406.2019.1632878>.

Zhang, Zhongheng. 2016. “Missing Data Imputation: Focusing on Single Imputation.” *Annals of Translational Medicine* 4 (1): 9. <https://doi.org/10.3978/j.issn.2305-5839.2015.12.38>.