

Further explanations on GitHub: Nonresponse and imputation

In the related non-response analysis, particular attention was paid to employment status, income, gender, number of persons in a household, household type, educational level, migration background, and whether a person works in a systemically important occupation, as well as the Covid-19 incidence at NUTS-3 regional level (on the day of the interview). Post-stratification was based on distributions taken from the German Microcensus 2018 for various regional and socio-economic characteristics, including age, gender, household size, citizenship, size of municipality, and federal state. The derivation of the respective survey weights is described in Siegers et al. (2021).

The proportion of complete cases is less than 95%. Little's (1988) test shows that the missingness mechanism is not missing completely at random. To counteract selection bias and at the same time increase the statistical power, we therefore multiply imputed missing values. For this purpose, we used the multivariate imputation by chained equations (mice) algorithm by van Buuren and Groothuis-Oudshoorn (2011), applying classification and regression trees (CART) as the imputation routine. To improve the predictive power of the imputation routine, we used several auxiliary variables in addition to the focal variables of this study (namely, migration background, educational attainment, household type, federal state). As suggested by Kim et al. (2006), we entered survey weights into the corresponding imputation models as explanatory variables. To maintain their autocorrelation and serial correlation structure, the data were imputed in wide format. We imputed $m=20$ data sets with 20 iteration steps in the Gibbs sampler of mice. We checked the convergence and meaningfulness of the estimated imputation models by means of the associated mice diagnostics (e.g., contrasting distributions of observed and imputed data).

References

- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Kim, J. K., Michael Brick, J., Fuller, W. A. & Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3), 509-521.
- Siegers, R., Steinhauer, H.W., & Zinn, S. (2021). Weighting the SOEP-CoV study 2020. SOEP Survey Papers 989888: SOEP Survey Papers: Series C - Data Documentations. https://www.diw.de/documents/publikationen/73/diw_01.c.820260.de/diw_ssp0989.pdf (accessed July 2021)
- Little, R.J.A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 83, 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>