AUSTRALIAN DIGITAL INCLUSION INDEX

Small Area Estimates: Method

2021

While the Australian Internet Usage Survey enables a national Index score for a range of socio-demographics, its sample size does not collect large numbers of responses for every regional area. One way to adjust for this is to use a robust technique called Small Area Estimates (SAE) to provide a 'probable' Index score based on a range of demographic characteristics or attributes within a particular location and relating these to the national Index measure derived by the ADII team.

SAEs combine survey data with additional data from sources with broader coverage, such as the Australian Census. They are a valuable approach for enhancing the value and detail of survey results from a finite sample of participants.

In modelling SAEs for the Index, analysts at the Social Research Centre combine national AIUS survey data with the ABS Census to estimate Index scores at the State/Territory and Local Government Area levels.

Method

The SAEs used for the ADII are derived from a nested error linear regression model (Battese et al., 1988; Rao and Molina, 2015). This entailed the following steps:

1. Use the AIUS data to build a linear mixed-effects model for the total Index and each of the three dimensions, conditional on the available demographic characteristics. This yielded coefficients for each characteristic ("fixed effects" in model parlance), along with an intercept term for each observed LGA ("random effects").

2. The model coefficients and intercepts were then applied to Census 2016 distributions (ABS, 2016) for the corresponding characteristics, within each LGA. There was no random effect for LGAs present in the Census data but not in the survey data. This yielded a mean value for each LGA and index.

Models were fit using the sae package (Molina and Marhuenda, 2015) in R (R Core Team, 2021).

Characteristics used for modelling

The characteristics used for modelling were asked of AIUS respondents, and their LGA distributions exported from Census 2016 TableBuilder.

1. Accessibility and Remoteness Index of Australia (3 categories: Major Cities, Inner Regional, Rest of Australia);

2. Age group (6 categories: 18-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, 75+ years)

3. Dwelling tenure (3 categories: Own outright, Own with a mortgage, Other);

4.   Gender (2 categories: Female, Male);

5.   Greater Capital City Statistical Area (2 categories: Capital city, Rest of state);

6.   Highest education (5 categories: Less than Year 12, Year 12 or equivalent, Undergraduate certificate or Diploma, Bachelor degree, Postgraduate degree or diploma);

7.   Household income (9 categories);

8.   Household structure (4 categories: Person living alone, Couple living alone, Couple family with dependent children, Other);

9.   Language spoken at home (2 categories: English, Other); and

10.   Quintile for Index of Relative Socio-economic Disadvantage (5 categories).

As is standard for the treatment of categorical variables in regression models, each characteristic was represented as a set of dummy indicator variables.

A note on LGAs with low population counts

Australian LGAs range in size from fewer than 100 adults through to almost 900,000 adults. ABS TableBuilder applies a perturbation algorithm to protect the confidentiality of its data (ABS, 2020), and this may have a noticeable impact on SAEs when the underlying count of cases in an LGA is low. As a result, SAEs for LGAs with populations of fewer than 500 adults are not displayed within the ADII dashboards.

AUSTRALIAN DIGITAL INCLUSION INDEX

References

Australian Bureau of Statistics (2016). LGA by various characteristics [Census TableBuilder]. Accessed 26 April 2021.

Australian Bureau of Statistics (2020). TableBuilder User Guide. Accessed 27 May 2021. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1406.0.55.005.

Battese, G.E., R.M Harter and W.A. Fuller (1988). An error-components model for prediction of county crop areas using survey and satellite data, Journal of the American Statistical Association 83 (401), 28-36. DOI: 10.1080/01621459.1988.10478561.

Gonzalez-Manteiga, W., M. Lombardia, I. Molina, D. Morales and L. Santamaria (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. Computational Statistics and Data Analysis 52 (12), 5242-5252. DOI:10.1016/j.csda.2008.04.031

Molina, I. and Y. Marhuenda Y (2015). sae: An R package for small area estimation. The R Journal 7 (1), 81-98.

Molina, I. and J.N.K Rao (2010), Small area estimation of poverty indicators. Canadian Journal of Statistics 38, 369-385. DOI:10.1002/cjs.10051.

Nakagawa, S., P.C.D. Johnson, and H. Schielzeth (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. Journal of The Royal Society Interface 14 (134), 20170213. DOI: 10.1098/rsif.2017.0213.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rao, J.N.K. and I. Molina (2015). Small Area Estimation (2nd edition). New York, NY: John Wiley and Sons. DOI: 10.1002/9781118735855.