`hopit`: An R Package for Analysis of Reporting Behavior Using Generalized Ordered Probit Models

Maciej J. Dańko[1]

1 Max Planck Institute for Demographic Research, Rostock, Germany

Summary

`hopit` is an open source software library written in the R (R-Core-Team, 2018) and C++ (Bates & Eddelbuettel, 2013; Eddelbuettel & François, 2011) programming languages. The `hopit` package provides versatile methods for fitting and analyzing ordered response data in the context of heterogeneity in self reporting behavior.

The ordered response data classify a measure of interest into ordered categories collected during a survey. For example, if the dependent variable is a happiness rating, then a respondent typically answers a question such as: "Taking all things together, would you say you are … ?; and then selects from response options such as:"very happy","pretty happy","not too happy", and"very unhappy" (Liao, Fu, & Yi, 2005). Similarly, if interviewees are asked to evaluate their health in general (e.g., "Would you say your health is … ?"), they may choose among several categories, such as "very good", "good", "fair", "bad", and "very bad" (Jürges, 2007; King, Murray, Salomon, & Tandon, 2004; Oksuzyan, Dańko, Caputo, Jasilionis, & Shkolnikov, 2019; Rebelo & Pereira, 2014). In political science, a respondent may be asked for an opinion about recent legislation (e.g. "Rate your feelings about the proposed legislation.") and asked to choose among categories like "strongly oppose", "mildly oppose", "indifferent", "mildly support", and "strongly support" (Greene & Hensher, 2010). It is easy to imagine other multi-level ordinal variables that might be used during a survey and to which the methodology described below could be applied.

In practice, it is assumed that when responding to a survey question about their general happiness, health, feelings, attitudes or other status, participants are assessing their true value of this unobserved continuous variable, and project it onto the discrete scale provided. The thresholds that individuals use to categorize their true status by selecting a specific response option may be affected by the reference group chosen, their earlier life experiences, and cross-cultural differences in using scales. Thus, the responses of individuals may differ depending on their gender, age, cultural background, education, and personality traits; among other factors.

From the perspective of reporting behavior modeling, one of the main tasks researchers face is to compute this continuous estimate of the underlying, latent measures of individuals based on several specific characteristics of the responses considered (e.g., health variables or happiness variables), and to account for variations in reporting across socio-demographic and cultural groups. More specifically, to build a latent, underlying measure, a generalized hierarchical ordered threshold model is fitted that regresses the reported status/attitude/feeling on two sets of independent variables (Boes & Winkelmann, 2006; Greene, Harris, Hollingsworth, & Weterings, 2014). When the dependent reported ordered variable is self-rated health status, then the first set of variables – i.e., health variables – assess specific aspects of individuals' health, such as measures of chronic conditions, mobility, difficulties with a range of daily activities, grip strength, anthropometric characteristics, and lifestyle behaviors. Using the second set of independent variables (threshold variables), the model also adjusts for differences across socio-demographic and cultural groups, such as differences in cultural background,

gender, age, and education (Jürges, 2007; King et al., 2004; Oksuzyan et al., 2019; but see Rebelo & Pereira, 2014).

The `hopit` package delivers functions and methods to fit (`hopit`), summarize (e.g., `summary`), check (e.g., `profile`), and compare (e.g., `AIC` and `anova`) fitted models. The latent and threshold formulas are defined separately. The interactions can be specified both within and between these formulas. Depending on how an interactions between latent and threshold variables is interpreted, it can be added to either the latent or the threshold formula. The package has also an option to include a survey design using the `survey` package (Lumley, 2004, 2019).

Once the model is fitted, the model estimates are used to determine reporting behavior; i.e., how the continuous latent measure is projected onto the categorical response. In practice, this is done by comparing actual categorical ordered responses with theoretical responses that are adjusted for heterogeneity in reporting behaviors, and are, therefore, more comparable across individuals.
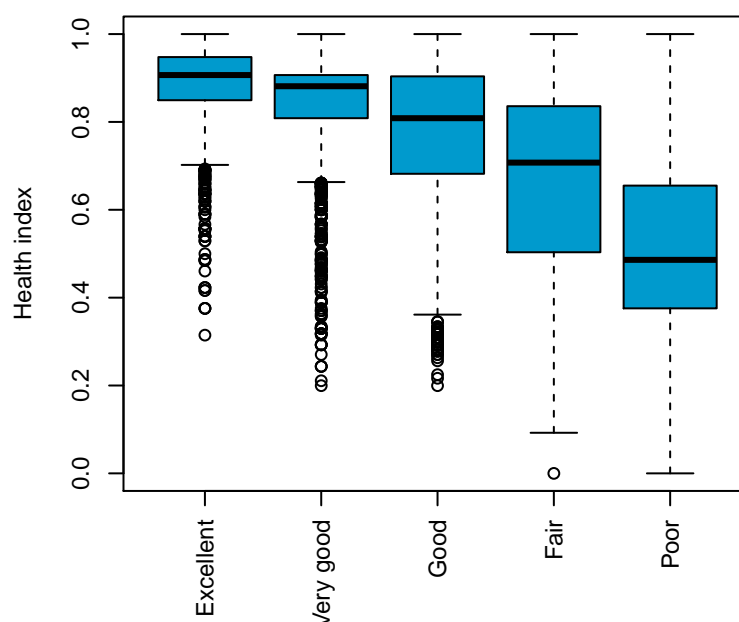

Figure 1: Health index vs. self-reported health for a model fitted to an exemplary data.

One of the first steps of the analysis is the standardization of the latent measure to obtain the latent index. In the self-rated health example the latent health index is a proxy for the true underlying health of an individual, and varies from 0, representing the (model-based) worst health state in the sample, to 1, representing the (model-based) best health state in the sample (Fig. 1).

The predicted latent measure obtained from the model is also used to standardize the latent variable coefficients. In the self-rated health example the standardized coefficients are called disability weights (Jürges, 2007; Oksuzyan et al., 2019) and are calculated for each health variable to provide information about the impact of a specific health measure on the latent index (Fig. 2). The disability weight for a health variable is equal to the ratio of the corresponding health coefficient and the difference between the lowest and the highest values of predicted latent health. In other words, the disability weight reduces health index by some given amount or percentage; i.e. the health index of every individual is reduced by the same amount if the person had a heart attack or other heart problems)(Jürges, 2007; Oksuzyan et al., 2019). While the latent index is intended to reflect the underlying health, happiness

or other status across individuals, the standardized coefficients (disability weights for health case), are computed for an average individual in the study population.
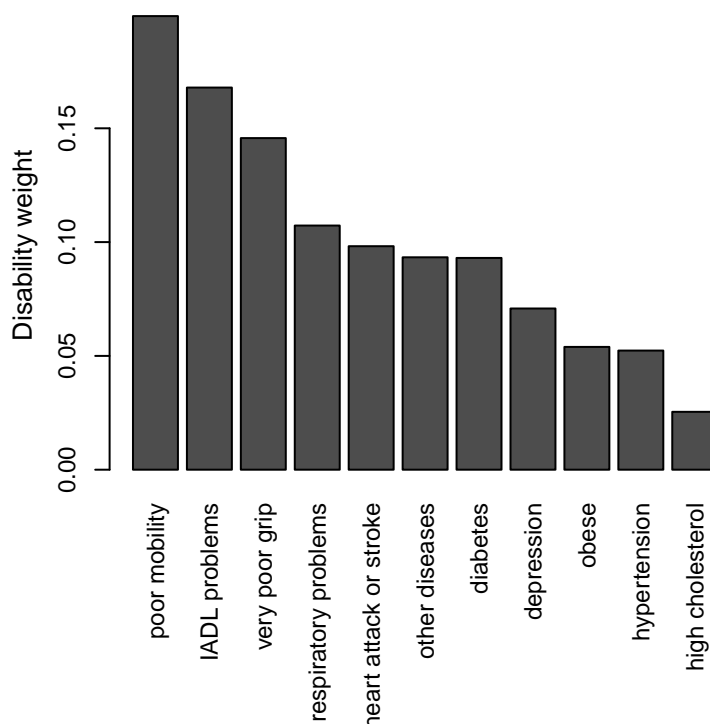

Figure 2: Disability weights for the fitted model.

Reporting behavior analysis is based on the reclassification of individuals into new response categories. There are two methods of reclassification: (1) Jürges (2007) percentile method (see also Rebelo & Pereira, 2014; Oksuzyan et al., 2019) and (2) reclassification directly based on model-estimated thresholds. In the first method, the classification is based on the calculated latent index which is adjusted for inter-individual differences in reporting behavior. This method is based on the original distribution of the categorical response variable (see also Oksuzyan et al. (2019)).
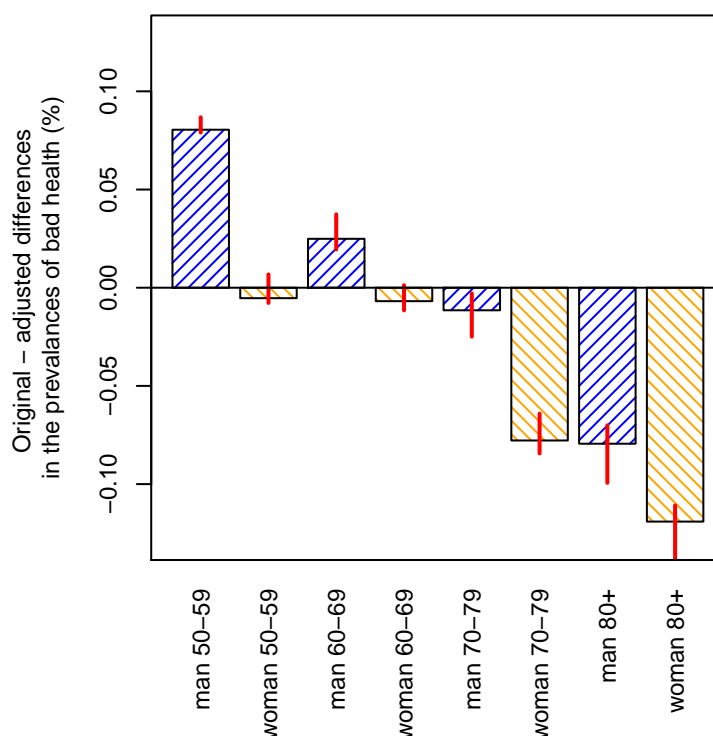
Figure 3: Differences between original and adjusted prevalences of bad health for the fitted model.

The package offers functions `boot_hopit` and `percentile_CI` for calculating the confidence intervals for any measure derived from the model using parametric bootstrap methods. In each of the bootstrap repetitions, a set of new model coefficients is drawn from the multivariate normal distribution, assuming the originally estimated model coefficients as a mean and using the model estimated variance-covariance matrix. The drawn coefficients are then used to calculate the measure of interest via a user defined function. In the example presented in Fig. 3, the confidence intervals of the difference between the original and the adjusted frequencies of bad health are calculated. The bad health is determined by the presence of "Poor" or "Fair" self-rated health categories.

The results (Fig.3) show that men tend to over-report bad health at ages (50,60] and (50,70], whereas women tend to over-report bad health at ages [70,80]; and that both sexes at ages above 80 tend to under-report bad health. See also Oksuzyan et al. (2019) for similar analyses done on true SHARE data.

Acknowledgements

In all examples presented above I use `healthsurvey`, which is a completely artificial data set that is simulated using the distributions of some major health and socio-demographic characteristics. The distributions and the data structure is roughly based on the WAVE1 SHARE database (DOIs: 10.6103/SHARE.w1.600) see Börsch-Supan et al. (2013) for technical details about SHARE database. None of the records represent any part of the true data.

References

Bates, D., & Eddelbuettel, D. (2013). Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software*, *52*(5), 1–24. doi:10.18637/jss.v052.i05

Boes, S., & Winkelmann, R. (2006). Ordered response models. *Allgemeines Statistisches Archiv*, *90*(1), 167–181. doi:10.1007/s10182-006-0228-y

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., et al. (2013). Data resource profile: The survey of health, ageing and retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001. doi:10.1093/ije/dyt088

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. doi:10.18637/jss.v040.i08

Greene, W., Harris, M. N., Hollingsworth, B., & Weterings, T. A. (2014). Heterogeneity in ordered choice models: A review with applications to self-assessed health. *Journal of Economic Surveys*, *28*(1), 109–133. doi:10.1111/joes.12002

Greene, W., & Hensher, D. (2010). *Modeling ordered choices.* Cambridge University Press.

Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics*, *16*(2), 163–178. doi:10.1002/hec.1134

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207. doi:10.1017/S000305540400108X

Liao, P.-S., Fu, Y.-C., & Yi, C.-C. (2005). Perceived quality of life in Taiwan and Hong Kong: An intra-culture comparison. *Journal of Happiness Studies*, *6*(1), 43–67. doi:10.1007/s10902-004-1753-6

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(1), 1–19. doi:10.18637/jss.v009.i08

Lumley, T. (2019). Survey: Analysis of complex survey samples. *R package version 3.35-1.* Retrieved from https://cran.r-project.org/web/packages/survey/index.html

Oksuzyan, A., Dańko, M. J., Caputo, J., Jasilionis, D., & Shkolnikov, V. M. (2019). Is the story about sensitive women and stoical men true? Gender differences in health after adjustment for reporting behavior. *Social Science & Medicine*, *228*, 41–50. doi:10.1016/j.socscimed.2019.03.002

R-Core-Team. (2018). R: A language and environment for statistical computing [internet]. Vienna, austria. R version 3.4.4 (2018-03-15). Retrieved from https://www.r-project.org

Rebelo, L. P., & Pereira, N. S. (2014). Assessing health endowment, access and choice determinants: Impact on retired europeans' (in)activity and quality of life. *Social Indicators Research*, *119*(3), 1411–1446. doi:10.1007/s11205-013-0542-1