Statistical Analysis Plan: Statistical Patterns of Sense of School-Belonging

Mariana Nold

July 31, 2024

1 Use of regression: modeling process

To evaluate the strength of the descriptive model precisely means to assess the capability of the model to give a meaningful characterization of subpopulations. It is therefore the validity of the model to answer the descriptive research question. The model smooths complex dependencies not easily captured in a simple formula. The strength of the descriptive model is given when it is clear how the subgroups of interest are defined and how the functional form of the model is justified. Following Shmueli (2010, p. 305), the modeling process is divided into three dimensions. The use of regressions defines precisely how each of the three dimensions is addressed, these dimensions are:

- Explanatory power
- Predictive power
- Strength of the descriptive model

Based on Borsboom et al. (2021) the empirical phenomenon is understood as stable and general features of the world that scientists aim to explain and the data is understood as a specific empirical pattern. The goal of regressions analysis is to describe associations thus a particular empirical pattern[1]. The selection of variables is justified by the fact that these are important to define the subgroups of students based on the withdrawal circle.

One statistical focus is goodness of fit: Bayesian model criticism, in form of PPC, provides a tool that is very helpful to define the strength of the descriptive model because PPC is a mean of assessing the predictive quality of those elements of the model that are important to the research question. In concrete terms, it can be used here to investigate the strength of the descriptive model:

[1] The purpose of the description is not theory construction as in Borsboom et al. (2021), but just to provide an informative summary that can be used as an empirical basis to discuss the phenomenon of interest. If patterns are found in the data, explanations can be identified and the quality of these explanations can be compared or discussed. This analysis only uses data from Austria. In order to see whether the statistical patterns is stable, the analysis could be carried out analogously for other countries.

We can us PPC within levels of grouping variables to assess in how far the model predictions are fine for specific subpopulations. Moreover, marginal predictive checks are helpful to find out if the model is well-calibrated (Gelman et al. 2013, pp. 152-153). In addition, randomized quantile residuals (Feng et al. 2017) can be computed (in Bayes and frequentist approach) to measure goodness of fit. The three relevant dimensions of the modeling process are addressed as follows in the analysis plan:

- Explanatory power: Definition of subgroups of interest to adequately describe the phenomenon of interest. Descriptive model does not directly address the construct level, here one needs to be clear about the definition of the scales used.
- Predictive power: Report LOOIC
- Strength of the descriptive model: Report kernel-density-PPC, posterior predictive summary (based on median absolute error, scaled median absolute error and corresponding intervals (Dogucu et al. 2021)) randomized quantile residuals and LOO-PIT (Gelman et al. 2013, p. 153) (Gabry et al. 2019)

2 Structure of the statistical analysis plan

The structure of this workflow is motivated by the workflow given in (Kaplan 2023, chap. 12).

1. **Specify the outcome and set of predictors of interest**
   - The outcome is sense of belonging, the corresponding variable is denoted as `belong`. This is a discrete responses.
   - To derive the estimand, focal predictors are `bull` and `ATT4`. The variable `bull` is binary, it is 1 if a student has severe bullying experience and 0 otherwise. `ATT4` is a ordinal variable that indicates the corresponding quartile of the level of truancy of a school. Based on the `ATT4` and `bull` variables, eight groups of students can be distinguished. The mean value of the outcome variable in these groups is the focus of the research question.
   - Non-focal predictors are given in table 1 in the task description. These variables are used later in step 7 to derive the estimand for fictive representative students that are of interest to describe the phenomenon of school-belonging.
2. **Specify the functional form of the relationship between the outcome and the predictors**
   - The discrete outcome is modeled based on a multilevel normal model, thus a model that assumes a continuous outcome. Given that the

estimand is based on averages, the model should be appropriate for this purpose.

- Iteration 1: Start with a simple multilevel model where each variable is a summand in the linear predictor including the interaction of the focal predictors `ATT4` and `bull`.
- PPC is used in step 6 and will assess the predictive quality of the simple multilevel model, use a log-transformation (Iteration 2) if the simple multilevel model does not pass the check. Decide on the exact transformation of the outcome variable based on PPC.

3. **Take note of the complexities of the data structure**
   - Use multilevel modeling because of the clustered data structure, students are nested in schools.
   - A simple random sample was assumed in the task, so that the issue of weighting is omitted. Thus the estimand is valid if ignorability holds[2].

4. **Decide on the prior distributions for all parameters in the model.**
   - Use weakly informative priors, namely the default prior of `rstanarm`.

5. **Check the convergence criteria**
   - Detailed information about these MCMC diagnostics can be found in Stan Development Team (2021).
   - If no convergence is reached, increase the number of chains or iterations and consider alternatively tuning parameters of the samplers to improve mixing.

6. **Refinement of model building: Use PPC to decide about possible transformations of the outcome**
   - Use PPC, e. g. kernel-density-PPC LOO-PIT and group-specific PPCs for focal predictors, to identify deficiencies of the iteration 1 model. If the iteration 1 model does not pass this checks improve the functional form, decide about the transformation of the outcome by understanding the model shortcomings detected by PPC. Report all tried improvements. If no appropriate improvement due to transformation of the outcome can be found, stop and document the results. In this case the selected variables are not suitable to predict the outcome. If an appropriate improved model is derived compute LOOIC and document it.

[2]This assumption implies that the model for sense of belonging is conditional on all variables that affect the probability of inclusion. Whether this assumption is plausible and how the analysis could be implemented if not is excluded from this workflow.

- **Sensitivity analysis:** Use a series of plausible changes in functional form to create a list of (nested) candidate models.
- **Model selection or model combination:** Is there a candidate model that significantly improves the prediction quality? LOOIC is used to see if the more flexible form improves predictive accuracy. If continuous model expansion becomes too complex, you should use discrete model expansion, e.g. Bayesian stacking[3]. If model combination is used one should compute PPCs for each model in the ensemble.
- **Conclusion of model building:** A model or model ensemble that has withstood criticisms, in the form of PPCs is appropriate for the description of the phenomena of interest OR selected variables are not suitable to predict the outcome. If a model or model ensemble is found go to step 7.

7. **Full description of the posterior distributions**

   - Compute prediction for the eight groups mentioned in the research question: The posterior distribution is summarized based on numerical and graphical tools. Summarize the results based on appropriate adjusted predictions (on the original scale) similar to Nold et al. (2024, figure 3). The non-focal predictors are set to quantiles based on substantive considerations. In this sense we compute between-unit-comparisons for groups of fictive representative students:
     - Compute predictions for hypothetical representative students to summarize the posterior knowledge.
     - To define representative students, the scales of the non-focal predictors are poled so that high values, with regard to literature or based on plausibility, are associated with a high sense of school-belonging.
     - For each of the eighth groups based on the levels of the focal grouping variables, compute three representative students, namely at the 25 %, the 50 % and the 75 % quantile of the non-focal predictors.
   - Does the data reveal a certain pattern with regard to bullying experience and the truancy level of schools?

[3]See e. g. Bayesian notes

References

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A. & Haig, B. D. (2021), 'Theory construction methodology: A practical framework for building theories in psychology', *Perspectives on Psychological Science* **16**(4), 756–766. PMID: 33593167.
URL: *https://doi.org/10.1177/1745691620969647*

Dogucu, M., Johnson, A. & Ott, M. (2021), *bayesrules: Datasets and Supplemental Functions from Bayes Rules! Book*. R package version 0.0.2.9000.
URL: *https://github.com/bayes-rules/bayesrules*

Feng, C. X., Sadeghpour, A. & Li, L. (2017), Randomized quantile residuals: an omnibus model diagnostic tool with unified reference distribution.
URL: *https://api.semanticscholar.org/CorpusID:126143700*

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. (2019), 'Visualization in bayesian workflow', *Journal of the Royal Statistical Society Series A: Statistics in Society* **182**(2), 389–402.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. & Rubin, D. (2013), *Bayesian Data Analysis*, 3rd ed edn, CreateSpace, United States.

Kaplan, D. (2023), *Bayesian statistics for the social sciences*, 2nd edn, Guilford Press, New York.

Nold, M., Meinfelder, F. & Kaplan, D. (2024), 'On the improvement of predictive modeling using bayesian stacking and posterior predictive checking'.

Shmueli, G. (2010), 'To Explain or to Predict?', *Statistical Science* **25**(3), 289–310.
URL: *DOI: 10.1214/10-STS330*

Stan Development Team (2021), *Stan Reference Manual,Version 2.30.* ISBN 3-900051-07-0.
URL: *https://mc-stan.org/docs/reference-manual/index.html*