Assessing Attitudes Towards Wood in the Context of Family Habits: Potential for Reuse of Raw Data from a Questionnaire Survey

ISTVÁN FEKETE (iD)

JUTKA NMARNÉ KENDÖL (iD)

*Author affiliations can be found in the back matter of this article

ABSTRACT

The dataset covers a statistically representative quantitative questionnaire survey conducted with 430 children with the aim of assessing their knowledge, habits and attitudes towards wood and trees. The goal of the study was to examine the variables that affect environmental awareness and the use of wood in the future. The dataset contains 430 rows (participants) and 49 columns (variables). Data has a reuse potential in the fields of childhood pedagogy and environmental education research but also opens new ways for demonstrating the formula of statistical representativeness or machine-learning analyses such as clustering techniques, associative learning, or decision trees.

CORRESPONDING AUTHOR:

István Fekete

Faculty of Culture, Media and Psychology, Macromedia University, Cologne, DE

istvfekete@gmail.com

(1) OVERVIEW

REPOSITORY LOCATION

Link to the dataset: https://doi.org/10.6084/m9.figshare.20055500.

CONTEXT

The dataset was produced at the University of Sopron, as part of the doctoral project of Jutka Nmarné Kendöl at the Roth Gyula Doctoral School of Forestry and Wildlife Management, and has not been used in another paper yet.

Findings of the survey study are highly relevant in the field of environmental education research, as they investigate the primary predictors of future use of wood. Such predictors included variables such as knowledge about trees and wood, attitude towards trees and wood, knowledge about traditions related to trees, extracurricular activities, habits related to trees and wood in both school and family contexts.

(2) METHOD

STEPS

1. **Pilot phase:** The questions of the survey were first piloted on five respondents. The questions and the length of the questionnaire were tailored to the level of the age groups under investigation.
2. **Software:** Questionnaires were administered online using *Google Docs* during normal classes.
3. **Variables:** After the sociodemographic variables in the questionnaire, questions followed related to the respondents' habits in school and family, traditions, feelings, and willingness to use wood in the future.
4. **Randomisation of items:** The questions were not (pseudo-)randomised across respondents because we did not expect any order-effects (e.g., the tendency of some questions affecting response behaviour with questions appearing later in the questionnaire).
5. **Debriefing:** Participants did not report any inconsistencies in the questionnaire in debriefings after completing the questionnaire.
6. **Statistical analysis:** *Google Docs* generated an *Excel* file which was submitted to statistical analyses employing the R software (R Core Team, 2021).

SAMPLING STRATEGY

Our questionnaire was completed by 230 male and 200 female students using non-random sampling to ensure that the relevant sociodemographic variables (gender, age, school type, and size of settlement) were counterbalanced using stratification weighting. Given the high number of participants and counterbalancing, the county-wide questionnaire survey is statistically representative of Győr-Moson-Sopron County. Based on data from the Central Statistical Office in Hungary, in Győr-Moson-Sopron County there were 33.996 pupils at primary schools and 7.507 at secondary schools enrolled in the year of 2014 (Központi Statisztikai Hivatal, 2015). We relied on the data from 2014 and counted with a target population of around 41.500, a margin error of 5%, and a confidence interval of 95%. Given these conditions, relying on the formula for statistical representativeness, 381 participants are needed for the study to be statistically representative (Daniel, 1999).

QUALITY CONTROL

Two raters performed a post-hoc plausibility check. An implausible questionnaire sample would be, for example, one that contains considerable missing data, or one completed by a participant without serious interest, a sign of which could be, for instance, the same response pattern across the questions. No such respondent was identified. Implausible values (e.g., age of 36) were removed with negligible data loss due to implausibility.

(3) DATASET DESCRIPTION

OBJECT NAME

The dataset is called "Survey data of children's attitudes towards trees and the use of wood". The dataset can be cited as follows: Fekete, István; Nmarné Kendöl, Jutka (2022): Survey data of children's attitudes towards trees and the use of wood. figshare. Dataset. https://doi.org/10.6084/m9.figshare.20055500.

FORMAT NAMES AND VERSIONS

The dataset is available in .csv

CREATION DATES

The survey was carried out in April, May, and June in 2021.

DATASET CREATORS

The doctoral candidate Jutka Nmarné Kendöl created the questionnaire and collected the data. Colleagues of the Roth Gyula Doctoral School were consulted in methodological questions.

LANGUAGE

The names of the variables as well as the levels of the factor variables have been translated into English. Given that the variable names are abbreviated, there is a list of the variable names on Figshare with the descriptive statistics under the name "Descriptive statistics".

LICENSE

Data has been deposited under CC BY 4.0 license.

REPOSITORY NAME

The data that support the findings of this study are openly available in the Figshare repository at https://figshare.com/projects/Assessing_attitudes_towards_wood_in_the_context_of_family_habits_a_large-scale_quantitative_study_in_Hungary/132230. The owner of the dataset is Jutka Nmarné Kendöl.

PUBLICATION DATE

The dataset was published in the Figshare repository on 10 February 2022.

(4) REUSE POTENTIAL

Given the high number of 49 variables, the sample size of 430 participants, and the statistically representativeness of our survey, the data allows further advanced statistical analysis:

1. Data can be analysed using data-mining approaches such as associative learning algorithms (e.g., market-basket analysis) to reveal participant groupings based on the variables (Patwary, Eshan, Debnath & Sattar 2021). Association chains can be extracted, for example, between sociodemographic variables, habits, behaviour, and attitudes (for the variables, see APPENDIX). Such an analytical framework could reveal hidden associations between variables and could help generate further research questions (e.g., what is the intimate relationship between family, school, knowledge, habits, and various aspects of attitude towards wood and trees?).

2. Data can be used as an example dataset in statistics to demonstrate multiple cluster analysis techniques. Cluster analyses can be run either on the variables or on the participants to explore how (i) the variables and (ii) the participants group together, respectively. For instance, via clustering, groups of outliers (participants) can be identified as well as the reasons of being an outlier. Thereby, interventions in environmental education can be proposed. Further, it can be asked which variables group together (e.g., do variables related to family and school cluster together?).

3. Given the high number of variables, data can be submitted to decision tree models such as conditional inference trees or random forests (e.g., Hothorn, Hornik & Zeileis 2006; Katuwal, Suganthan & Zhang 2020). The idea is to select an outcome variable (e.g., "the importance of wood and trees") and use a high number of independent variables to explain or predict the outcome variable.
4. Dataset is suitable for teaching statistical representativeness.
5. Dataset allows for further analyses mainly in the fields of childhood pedagogy and environmental education. Specifically, new aspects of environmental pedagogy, environmental education, sustainable development, climate protection, sylviculture, environmental awareness of families, adult environmental education, and education policies can also be investigated from the perspective of environmental awareness.

For instance, research into climate protection can benefit from further analyses on the dataset to gain insights about the amount of wood and trees used in families. This issue is highly relevant, as the use of wood as a raw material can bind coal for decades and even centuries. Second, adult environmental education can profit from further analysis by examining the level of environmental awareness about wood and trees in adults. In light of the findings, new steps in environmental education can be implemented via advertisement, media, and environmental programs to raise the environmental awareness of adults, and to spread or increase the use of wood. Third, information about the amount of wood and trees used in families could be useful for sylviculture, as the amount of trees to be planted has to be planned in advance.

One of the limitations can be that socioeconomic status was not verified. However, we claim that given the relative high number of participants, the possible effect of this confounder has been partialled out.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **APPENDIX.** The APPENDIX contains descriptive tables illustrating the factor and the numeric variables in the survey. Numbers indicate frequencies per level of the factors. DOI: https://doi.org/10.5334/johd.82.s1

- **A summary analysis of all the ordinal variables in the dataset.** The number of responses are represented by "n". SD designates standard deviation of the mean. DOI: https://doi.org/10.5334/johd.82.s2

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

István Fekete**:** Formal analysis, Software, Visualization, Data curation, Methodology, Validation, Writing – original draft, Writing – review & editing

Jutka Nmarné Kendöl: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing

AUTHOR AFFILIATIONS

István Fekete orcid.org/0000-0002-6734-3660
Faculty of Culture, Media and Psychology, Macromedia University, Cologne, DE

Jutka Nmarné Kendöl orcid.org/0000-0003-1478-4312
Roth Gyula Doctoral School of Forestry and Wildlife Management, Faculty of Forestry, University of Sopron, Sopron, HU

REFERENCES

Daniel, W. W. (1999). Biostatistics: A foundation for analysis in the health sciences. John Wiley & Sons: Hoboken. DOI: https://doi.org/10.1002/1097-0258(20010130)20:2<324::AID-SIM635>3.0.CO;2-O

Hothorn, T., Hornik, K., & **Zeileis, A.** (2006). Unbiased recursive partitioning: A Conditional inference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651–674. DOI: https://doi.org/10.1198/106186006X133933

Katuwal, R., Suganthan, P. N., & **Zhang, L.** (2020). Heterogeneous oblique random Forest. *Pattern Recognition, 99*. DOI: https://doi.org/10.1016/j.patcog.2019.107078

Patwary, A. H., Eshan, P., Debnath, T., & **Sattar, A.** (2021). Market Basket Analysis Approach to Machine Learning, *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. pp. 1–9. DOI: https://doi.org/10.1109/ICCCNT51525.2021.9580175

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

Statisztikai Hivatal [Hungarian Central Statistical Office]. (2015). GYŐR-MOSON-SOPRON MEGYE SZÁMOKBAN. Available from: https://www.ksh.hu/docs/hun/xftp/idoszaki/regiok/mesz/08_gy_14.pdf (last accessed: 31 March 2022).