

Ex Post Harmonization And Data Integration of Cultural Access and Participation Surveys

Early Draft.

Daniel Antal, CFA

1/27/2022

Introduction

In this article, we will create a harmonized Cultural Access and Participation Dataset for live music and active participation (playing music and singing.) The standardized CAP survey is the most important measurement of the topic. It is a standardized, hierarchical survey. The Eurobarometer files contain the first and sometimes second priority questions, whereas the Hungary and Slovakia CAP contains third and fourth priority (granular) questions for music. Our harmonization and data integration will allow us to compare pan-European and national surveys, and the possibility to make inference from the far more detailed Hungary and Slovakia CAP to the less detailed Eurobarometer CAP which in turn has a far larger geographical coverage. Our paper is structured as follows. First, we introduce the concept of Cultural Access & Participation Surveys, and define a music-relevant subsets from 5 (hopefully 7 with Slovakia 2015,2020) CAP surveys. In the second part, we will show how we can create a unified longitudinal panel datasets for about 30 countries and 7 years with a uniform coding, and a unified codebook using reproducible ex post survey harmonization techniques, and the retroharmonize open source program library developed for this purpose. In the third part, we will introduce data integration or data fusion, when we will investigate the quality limitation of the unified dataset, and make potential statistical improvements for their joint use, for example, with creating an adjusted post-stratification weight for each observation. In the conclusions, we will show what kind of picture one can get from a national, a single-year pan-European CAP survey and how we can improve our knowledge about access to live music and participation in music with data integration.

NOTE: if the University of Bologna or the University of Turku has an agreement with Eurostat for microdata access (or you can send one) we can add three further CAP surveys, which are not in Eurobarometer (which is open access), but in AES, EU-SILC.

Introduction to Cultural Access & Participation Measurement

- **visit_concert, visit_cinema, visit_church:** These are the standard CAP variables about visiting frequencies. We will use cinema as a comparator to concerts. Church will be important to compare with singing, because in many parts of Europe mainly women sing and mainly in church ceremonies. The variables are differently coded in the Eurobarometer CAP 2007, Eurobarometer CAP 2013, and the Hungary CAP 2017, 2018, 2019 CAP surveys. Eurobarometer have them in categories (0, 1-2, 3-5, more) and the Hungary CAP has them as a number. We will use these variables with different coding. For visit_church, when available, we will convert it into a binary variable.
- **played_music:** A variable of the frequency of playing music as an amateur or professional practice, in some surveys we have only a binary variable (played in the last 12 months or not) in other we have frequencies. We will convert this into a binary variable.
- **sung:** A variable of the frequency of singing as an amateur or professional practice, like played_music. Because of the liturgical use of singing, and the differences of religious practices among countries and gender, this is a significantly different variable from played_music.
- **age:** The respondent's age as an integer number. **sex:** The respondent's sex as a binary variable.

- **country**: an ISO country code
- **geo**: an ISO code that separates Germany to the former East and West Germany, and the United Kingdom to Great Britain and Northern Ireland, and Cyprus to Cyprus and the Turkish Cypriot community.[we may leave Turkish Cyprus out for practical reasons.]
- **cooperation**: the score of the cooperation level of the respondent, whenever available.
- **age_education**: This is a harmonized education proxy. Because we work with the data of more than 30 countries, education levels are difficult to harmonize, and we use the Eurobarometer standard proxy, age of leaving education. It is a specially coded variable, and we will re-code them into two variables, **age_education** and **is_student**.
- **is_student**: is a dummy variable for the special coding in **age_education** for “still studying,” i.e. the person does not have yet a school leaving age. It would be tempting to impute **age** in this case to **age_education**, but we will show why this is not a good strategy.
- **type_of_community**: this is a subjective urbanization level variable, which is used in Eurobarometer to make urbanization of small countries like Malta and Estonia comparable with large federal states like Germany.
- **occupation, occupation_recat_4**: a 14-level standardized occupation code, and a 4-level grouping. We will create dummy variables from both the 14 occupations and the 4 groups.
- **region**: A NUTS2 region of the respondent. We will show some problems with this—but we will not go in this publication to small area statistics issues.
- **difficulty_bills**: A subjective income variable that makes income comparable among many countries, it is a three value categorical variable describing the frequency of having difficulties paying the respondent’s bills (never, from time to time, always or almost always.)
- **w, wex**: post-stratification weights and projected weights for the 15+ years old population of each country (treating Northern Ireland, Great Britain, the United Kingdom, the former GDR, the former West Germany, and Germany as six countries.)

Retrospective survey harmonization

We will refer here to the other small article introducing `retroharmonize` as a software.

Data integration

Findings

References