

Field Property Values: A Tool to Identify and Adjust Missing Data from 'Relational' Extraction

By

Mohammad Mushtaq and April Beaulé

University of Michigan

Outline of the Presentation

1. About the PSID/Survey and the Use of Blaise 4/5
2. The Issue with data-out from the PSID Mixed Mode Pilot 2019
3. The Identification/Debug
4. The Evidence/Solution

About the PSID/Survey and the Use of Blaise 4/5

1. Nationally representative survey since 1968
2. Blaise 4 since 2003
3. Blaise 5 since 2016, for pilot surveys
4. Used Blaise 5 for TAS 2019 production
5. Data extraction from Blaise at block level, aka ASCII-relational

The Issue with the PSID Mixed Mode Pilot (MMP) 2019

1. Analyze Field Properties Values (FPS) file from “wide” and “relational” data extraction
2. Cell level indicator of visited fields plus F2 notes/remarks
3. 55,638 rows from “wide” (W) extraction
4. 50,993 rows from “relational” (R) extraction
5. 179 vs.174 F2 notes from W vs. R extraction
6. Note the difference between W and R data extraction - that’s the issue

The Identification/Debug

1. Create cell level linkable dataset from the survey data
 1. 139 tables (6,279 variables) and 106 tables (5,176 variables) with data
 2. 1,187,455 rows in the long format
2. FPS data from “wide” extraction
 1. Subset where property=IsVisited, value=1
 2. Data at .bmix level, adjust for multi mention (set of) to match with data-out
 3. Parse the path values; create variable name and table instance to match with data-out
3. Left join FPS (#2) and data-out long format (#1)

The Evidence/Solution

1. 74,536 values flagged as visited
 1. 55,458 found in the data-out tables
 2. 18,887 on-route/missing
 3. 191 new rows/values into tables
2. Alternative approach: rollup data-out at .bmix level, “set of” variables are counted once
 1. 4,481 on-route/missing
3. The data are extracted into SAS files, then treated for on-route/missing values (.V) and finally converted to .R (refused) values

Conclusions

1. Identification of on-route/missing data values is important for data processing and release of the PSID data
2. The FPS from “wide” extraction is used to identify/adjust on-route/missing data from the “relational” data extraction
3. The issue should be further investigated in collaboration with the PSID staff and the Blaise development team at the Statistics Netherlands
4. Questions
5. Thank you!