



Data-out Experience and Challenges in Blaise 5

Mohammad Mushtaq, April Beaulé
The 17th International Blaise Users Conference
The Hague, Netherlands
October 2016

Outline

- Study and Background
- Data Extraction/Export
- Transformation – wide to long
- Documentation/Release

About the PSID

- Nationally representative sample of 9,000 U.S. families
- Panel survey conducted since 1968
- Main CATI instrument programmed in Blaise 4.84 and C#.net
- Self-administered web pilot programmed in Blaise 5

Background

- Main survey instrument reprogrammed in Blaise in 2001 and been using Blaise very successfully since then
- Family level interview which collects data on all members of the family via one respondent
- Size and complexity of instrument requires data be extracted by sections (aka ASCII relational format)
- Opt-in web (limited) launched in the Spring of 2015
- Pilot web version launched in 2016

Blaise 5 Data Extraction

- Was it a challenge?
- It's relative to the size of survey (R's&Q's) and data-out format
- Tool used since 2001, Manipula, was not available for Opt-in panel
- ASCII relational output is still not available
- Using wide format, had to reverse engineer data blocks, it was a challenge!

Blaise 5 Data Export - Wide

- From Blaise 5 Control Center
- Home tab: ETL Export Tools be used with caution
- Instead, use Browse Database, Data Viewer tab, and use Export Data
- Demo of data export from Home tab and Data Viewer

Transformation: Wide to Long

- Wide file has 11664 variables
- Import into a SAS table
- Create blocks and sub-blocks
- SAS/Macros to split blocks and sub-blocks by instance
- Stack them all to create long files for sub-blocks
- Sample SAS code

Identify Blocks and Sub-blocks

- Know your study instrument
- Section_A.A3
- Section_A.A5CKPT
- Section_A.A37FOR[1].A37B – loop -1
- Section_A.A37FOR[1].A37C
- Section_A.A37FOR[2].A37B – loop -2
- Section_A.A37FOR[2].A37C

SAS code: foreclosure loop1 & 2

```
data A37FOR_01 ;
  length SampleId $24 A37FORInstance 3 ;
  set V1.psid15 (keep=
    Sampleid
    Section_A_A37FOR_1__A37B
    Section_A_A37FOR_1__A37C
    rename=(
      Section_A_A37FOR_1__A37B = A37B
      Section_A_A37FOR_1__A37C = A37C )
  ) ;
  A37FORInstance = 01 ;
run ;
```

```
data A37FOR_02 ;
  length SampleId $24 A37FORInstance 3 ;
  set V1.psid15 (keep=
    Sampleid
    Section_A_A37FOR_2__A37B
    Section_A_A37FOR_2__A37C
    rename=(
      Section_A_A37FOR_2__A37B = A37B
      Section_A_A37FOR_2__A37C = A37C )
  ) ;
  A37FORInstance = 02 ;
run ;
```

SAS Code for sub-blocks

```
data A37FOR (drop=rowsum) ;  
  length SampleId $24 A37FORInstance 3  
    A37B 8  
    A37C 8  
  ;  
  set |  
    A37FOR_01  
    A37FOR_02  
  ;  
  rowsum = sum(A37B,A37C) ;  
  if rowsum > . ;  
  attrib _all_ label=" " ;  
run ;
```

- Wide table: 1489 rows and 11664 columns
- 52 relational tables, 2554 variables
- Indeed, it was a challenge!

Metadata

- Need for data documentation
- Variable type and length
- Questions text, code values and labels
- Available as XML export
- Here comes another challenge;
- Code values are missing from enumerations and sets

Determining a partial interview

- This is a straight-forward task with interviewer administered instrument
- Know when a web interview form has been submitted – but it's hard to know if they are 'complete'
- Processing team works closely with researchers to determine acceptable the level of item non-response
- Knowing the number of variables 'on-route' is critical to determine if this case meets the acceptable partial threshold

Knowing On-Route vs Off-Route

ER53680		F6AB WTR REC FREE BRKFT/LUNCH LAST YR	
F6AB. And during 2012, did (CHILD NAME/any child in your family between 5 and 18 years old) receive free or reduced-cost breakfasts or lunches at school? If "Yes": Was that breakfast, lunch, or both?			
Count	%	Value/Range	Text
49	.54	1	Yes, breakfast only
217	2.39	2	Yes, lunch only
1,193	13.16	3	Yes, both
1,618	17.85	5	No, neither
17	.19	8	DK
26	.29	9	NA; refused
5,943	65.57	0	Inap.: all FU members younger than 5 or older than 18; all children with current age 5-18 not in the FU in 2012 (ER53679=5)
Years Available: [13]ER53680			
Index Summary: Family Public Data Index 01>NON-CASH ASSISTANCE 02>Public 03>subset selectors 04>school meals programs, received last year: 05>type, whether:			

Using a 'Box and Arrow' Questionnaire

F7CKPT. CAI CHECKPOINT: WHETHER ANY CHILDREN AGE 5-18 YEARS OLD WERE IN THE FU IN PYEAR
(FUP1YEAR.AGEIWDATE=5-18)

1. CHILD(REN) AGE 5-18 IN PYEAR FU 5. ALL OTHERS → GO TO F6CCKPT



F6AB. **And during [PYEAR],** did [ONE CHILD IN FU PYEAR: [CYNAMF CYNAML] / ALL OTHERS: any child in [your / the] family between 5 and 18 years old] receive free or reduced-cost breakfasts or lunches at school? If "Yes": Was that breakfast, lunch, or both?

1. Yes, breakfast only 2. Yes, lunch only 3. Yes, both 5. No, neither

F6CCKPT. CAI CHECKPOINT: WHETHER ANY CHILDREN AGE 15 YEARS OLD OR YOUNGER WERE IN THE FU IN PYEAR
(FUP1YEAR.AGEIWDATE<16)

1. CHILD(REN) AGE < 16 IN PYEAR FU 5. ALL OTHERS → GO TO F7B2CHPT



F7 (F6D). How much did [you [and your family living there] / they / the family living there] pay for child care in [PYEAR]?
• ENTER amount here, then SELECT unit of time on next screen (Week, Two weeks, Month, Year)

\$ 0 → GO TO F7A \$ 1 - 999,997 DK/RF → GO TO F7MO



Write out all skips in SAS

```
/*section J*/
/*A1a rule--mother in house*/

  /*to fix 2 zeros to NA*/
    if J1A='0' and A1A ne '9' and A1A ne '7' then J1A='9';
  /*to fix 3 zeros to NA*/
    if J2A='0' and A1A ne '9' and A1A ne '7' then J2A='9';
  /*to fix 4 zeros to NA*/
    if J3A='0' and A1A ne '9' and A1A ne '7' then J3A='9';
  /*to fix 3 zeros to NA*/
    if J4A='0' and A1A ne '9' and A1A ne '7' then J4A='9';
  /*to fix 3 zeros to NA*/
    if J5A='0' and A1A ne '9' and A1A ne '7' then J5A='9';
  /*to fix 3 zeros to NA*/
    if J6A='0' and A1A ne '9' and A1A ne '7' then J6A='9';
  /*to fix 5 zeros to NA*/
    if J7A='0' and A1A ne '9' and A1A ne '7' then J7A='9';
  /*to fix 3 zeros to NA*/
    if J8A='0' and A1A ne '9' and A1A ne '7' then J8A='9';
  /*to fix 4 zeros to NA*/
    if J9A='0' and A1A ne '9' and A1A ne '7' then J9A='9';
  /*to fix 4 zeros to NA*/
    if J10A='0' and A1A ne '9' and A1A ne '7' then J10A='9';
  /*to fix 5 zeros to NA*/
    if J11A='0' and A1A ne '9' and A1A ne '7' then J11A='9';
  /*to fix 4 zeros to NA*/
    if J12A='0' and A1A ne '9' and A1A ne '7' then J12A='9';
  /*to fix 5 zeros to NA*/
    if J13A='0' and A1A ne '9' and A1A ne '7' then J13A='9';

/*A1b rule--father in house*/

  /*to fix 8 zeros to NA*/
    if J1B='0' and A1B ne '9' and A1B ne '7' then J1B='9';
  /*to fix 8 zeros to NA*/
    if J2B='0' and A1B ne '9' and A1B ne '7' then J2B='9';
  /*to fix 8 zeros to NA*/
    if J3B='0' and A1B ne '9' and A1B ne '7' then J3B='9';
  /*to fix 8 zeros to NA*/
    if J4B='0' and A1B ne '9' and A1B ne '7' then J4B='9';
  /*to fix 9 zeros to NA*/
    if J5B='0' and A1B ne '9' and A1B ne '7' then J5B='9';
  /*to fix 9 zeros to NA*/
```

Conclusion

- At time of web pilot, ASCII-relational extraction was not available
- Intimate knowledge of PSID questionnaire allowed reverse engineering from Blaise 5 output to Blaise 4.84 output
- Significant effort required for extraction and coding of off-route/on-route skips
- Continue to explore self-administered options for large surveys to reduce costs
- We hope the similar extraction and assignment of 'skipped' options in 4.84 will be offered in Blaise 5 soon

Thank You

Mohammad Mushtaq

mushtaq@umich.edu

April Beaulé

abeaule@umich.edu