Process to convert SEMCOG 2015 Demographic data to ABM test inputs

Conversion script

<https://github.com/SEMCOG/SEMCOG_popsim/ABM>\_inputs/

ABM tests need several variables "HHT", "pemploy", "pstudent", "ptype" that are not part of demographic inputs for SEMCOG Regional Forecast Model. Therefore current 2015 SEMCOG model (UrbanSim) base year (2015) demographic data needs to be further processed to aquire those addtional information. Some of the variables, such as "HHT" use original PUMS values while others are derived from multiple PUMS attributes.

Notes from RSG: Logics to convert PUMS data to ABM inputs

Household attributes:

1. Hworkers: Yes, this is number of workers in the household based on each member’s Employment Status Recode (ESR). ESR is defined in PUMS as follows:

Members with ESR as 1, 2, 4 and 5 are counted as workers

2. HHT: Yes, this is the original PUMS field – Household/family type

Person Type (ptype):

1. The ptype code is defined using the following PUMS person-level variables:

a. ESR: Employment Status Recode (ESR)

b. WKHP: Usual hours worked per week past 12 months

c. WKW: Weeks worked during past 12 months

d. SCHG: Grade level attending

e. AGEP: Age

2. The person type is derived from person’s age, employment status (pemploy) and student status (pstudent).

3. The employment status is derived from ESR, WKHP, WKW and Age

4. The student status is derived from SCHG, Age and employment status

As long as we have ESR, WKHP, WKW, SCHG and AGEP in the person file, employment status, student status and person type can be derived.

We have documented the person type coding process for ODOT. Please follow this link for a detailed description of person type coding logic: https://github.com/RSGInc/SOABM/wiki/Person-Type-Coding-in-SOABM

----------------------------------------------------------------------------

Additional notes by SEMCOG:

Workers: ABM definition is different from REMI (based on labor force)

ABM worker: ESR category 1,2,4,5

REMI worker: ESR category 1,2,3,4,5

SCHG: PUMS defintion changes over year. RSG version is different from 2015, See below

SCHG https://github.com/RSGInc/SOABM/wiki/Person-Type-Coding-in-SOABM

Grade level attending

b .N/A (not attending school)

1 .Nursery school/preschool

2 .Kindergarten

3 .Grade 1 to grade 4

4 .Grade 5 to grade 8

5 .Grade 9 to grade 12

6 .College undergraduate

7 .Graduate or professional school

SCHG (2015 PUMS variable codes, SEMCOG model base year data)

Grade level attending

bb .N/A (not attending school)

01 .Nursery school/preschool

02 .Kindergarten

03 .Grade 1

04 .Grade 2

05 .Grade 3

06 .Grade 4

07 .Grade 5

08 .Grade 6

09 .Grade 7

10 .Grade 8

11 .Grade 9

12 .Grade 10

13 .Grade 11

14 .Grade 12

15 .College undergraduate years (freshman to senior)

16 .Graduate or professional school beyond a bachelor's degree

Procedure:

SEMCOG final 2015 RDF model inputs ("all\_semcog\_data\_02-02-18.h5") doesn't contain the original 2015 PUMS sample id ("SERIALNO"). Therefore, SEMCOG has to use the household id and person id from synthesized households and persons ("starter6\_20171019-1526.h5) to link back to original PUMS samples. In that way, additional PUMS attributes desired by ABM can be acquired. However, due to the modeling need in regional forecast, the model input data was modified with additional households and persons (around 5% of the synthesized population). Re-linking those different households and persons is not straightforward as those data doesn’t hold matching IDs from synthesized population. This script uses matching household or person attributes plus geographic locations, such as PUMA or county, to identify or sample original PUMS records. Then extract needs variables and convert them to ABM inputs.