Process to convert SEMCOG 2015 Demographic data to ABM test inputs

**SEMCOG 2015 Demographic Data**

Residential Households and Persons:

SEMCOG Regional Forecast 2045 base year (2015) demographic data was initially developed using Synthpop (<https://github.com/SEMCOG/synthpop>). SEMCOG applied 8 household controls ( “size”, “race of head”, “Hispanic\_head”, “children”, “income”, “workers”, “cars”, “tenure\_agehead”) and 3 person controls (“age”,”race” and “sex”) at block group level in synthesis process. Then based on the modeling needs to recode or categorized PUMS variables to model variables. A particular case is the “worker” variables. Although it was defined by “ESR” code 1,2,4,5 in synthesis process, it was post-processed to include ESR 3 (unemployment) as the worker controls from REMI was based on the definition of labor force. All block group level synthesized households and persons were further placed into individual housing units based on matching PUMS housing attributes such as property values, year built etc to the housing stocks (building data collected by SEMCOG) within each block group in 2015. The final model residential variables can be seen as follows:

**Households:**

*'building\_id', 'cars', 'workers', 'persons', 'race\_id', 'income', 'age\_of\_head', 'children'*

**Persons:**

*'relate', 'age', 'worker', 'sex', 'race\_id', 'member\_id', 'household\_id'*

During the forecast, SEMCOG tested the demographic data and observed some sharp growths or declines at sub-county level forecast, especially in early forecast period. The reason was largely due to some significant differences between REMI demographic controls and Census 2015 numbers. Thus, SEMCOG tried to make some additional updates to the demographic inputs. Since REMI controls had population distributions by particular attributes, such as age and race, it will be difficult and error-prone to arbitrarily allocate REMI county level distributions to Census block groups. Therefore, instead of developing new household and person controls for block groups, SEMCOG focused on the MCD level reviews and used post-process methods to make up the difference between REMI and Census. That result additional household and person record, which are not directly available from PUMS records. The total amount of records were less than 0.5% of the total households and persons.

Group Quarter population:

2015 group quarter (GQ) population was developed separately. SEMCOG developed own building level GQ controls based on local data and knowledge and sampled 2015 5-year ACS records to generate full GQ population. The attributes available in GQ population were age, race and GQ type. Since GQ forecast was not part of UrbanSim simulation process. SEMCOG only developed base year GQ population but not GQ households. The final GQ table:

**Group Quarters:**

*'age', 'type', 'gq\_code', 'building\_id', 'race\_id'*

**SEMCOG ABM Input Requirement**

Based on SEMCOG ABM design, following variables are needed for model test:

**Households (include GQ households):**

|  |  |
| --- | --- |
| **HHID** | Unique household ID number |
| **TAZ** | Transportation analysis zone of home location |
| **TYPE** | Type of unit |
| **HINCP \*** | Household income |
| **ADJINC \*** | adjustment factor for dollar amounts |
| **NP** | Number of persons in the household |
| **HHT \*** | Household/family type |
| **VEH** | Number of vehicles available |

**Persons (include GQ population)**

|  |  |
| --- | --- |
| **HHID** | Unique household ID number |
| **PERID** | Unique person ID number |
| **AGEP** | Person's age in years |
| **SEX** | Gender |
| **ESR \*** | Employment status recode |
| **WKHP\*** | Usual hours worked per week past 12 months |
| **WKW\*** | Weeks worked during past 12 month |
| **SCHG\*** | Grade level attending |
| **MIL\*** | Military service |

\*marked are new variables needed by ABM

As shown in the tables, both household and person tables will need additional PUMS variables. It would be easy if all individual records has original serial numbers from PUMS. Since SEMCOG has been keeping the original synthesized dataset with PUMS SERIALNO, 99.5% of model demographic records can find original PUMS samples reliablly. However there are still around 0.5% households or persons need to assign new values for those variables. For GQ, there are less personal attributes available than residential population, meaning more attributes (+ SEX) has to be generated and the whole GQ household has to be built.

**Method to generate new variables for non-synthesized demographic data:**

To assign reasonable variables to non-synthesized households and persons, SEMCOG used a “matching and dropping” process described as follows:

1. Households: The process used a predetermined set of households attributes, including demographic and geographic information ( 'income', 'race\_id',  'age\_of\_head','cars', 'county\_id', 'PUMA5CE') to match PUMS records to each non-synthesized household. A python program will query PUMS samples using above attributes from each target household. If more than one records were found, only one household will be sampled and its SERIALNO will be attached to the household. If no available households, the process will drop the last attribute and redo the matching process. This procedure repeats until a valid PUMS sample is found. When all household attached valid SERIALNO, needed PUMS variables will be joined to households.
2. Persons: Similar procedure was applied and predetermined attributes were ('county\_id',  'worker',  'sex',   'race\_id',  'age') .
3. Group quarter person: Matching data set are model GQ persons and PUMS GQ samples. Matching attributes are ('TYPE', 'race\_id', 'age', 'PUMA')
4. Group quarter households: Since matched GQ persons already contain valid SERIALNO, GQ households use either existing attributes such as age, race etc or attach new information from PUMS.
5. Households and GQ households, persons and GQ persons are combined to make the final demographic data.

The full conversion python script can be located at

<https://github.com/SEMCOG/SEMCOG_popsim/tree/master/urbansim_to_abm>

**Post-process for ABM (Optional)**

Final section of the program converts PUMS variable to direct input variables in ABM. Here are some special notes helpful in the process

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