HADS custom variable definitions

* OWNRENT (comes from Tenure)
  + 0 – Rental
  + 1 – Owner
* COST06, COST08, COST12, COSTMED
  + Calculated cost to owner, this may be what we want to predict
* UTILITY
  + Is imputed for vacant units using monthly rent, structure type, region and tenure
* BURDEN
  + -1 for houses with no income, otherwise housing cost divided by monthly income
* Assisted housing
  + Overreported, use with caution

Variables that might be useful for prediction

ZINC2 – Household Income

IPOV – Poverty level income (according to the area)

ABLMED, ABL30, ABL50, ABL80 – Median income adjusted for bedrooms

ASSISTED – 0, 1, -9 – Are they receiving some kind of government assistance

FMR – Fair market rent – calculated by HUD as a factor of many things for a given county

LMED, L80, L50, L30 – Income levels of area

APLMED – Income adjusted for persons in household (AHS underestimates sometimes)

age1 – Age of head of household

BEDRMS

BUILT

OWNRENT

METRO3

REGION – Census Region

STATUS

STRUCTURETYPE / TYPE

ZADEQ – Adequacy of unit

ZSMHC – Monthly housing costs – we may have to delete this because it includes mortgage

NUNITS - # units of building

OTHERCOST – ZSMHC may be inconsistent, OTHERCOST is sum of insurance, land rent (not rent), and fees.

PER (# Of persons in household)

ROOMS (# Of Rooms in Unit)

TENURE (Owner/renter status of unit)

TOTSAL – Sum of salary income over all members of household

UTILITY (Monthly utility cost)

VALUE – Current market value (what we are trying to predict)

Discard Variables:

CONTROL – The AHS control number. I believe this is just a private key

COSTX – anything cost related, because we can directly compute the fair market value with that

GL30, GL50, GL80, GLMED – Doesn’t pertain to our dataset

Anything FMR – these are all just recalculated percentages of FMR

ADEQ – Not in model, we use ZADEQ

ZHMHC