Notes on reconciling Python and R model results

8/5/2020 Chris Zhang

1. low\_wage in R ACS are decimal (wage?). It should be binary. See data dictionary for post-sim ACS.
2. R ACS has variable names like ‘married.1’ likely duplicated column from merging in R code.
3. First column of R ACS has no column label – please verify what this column is or if it’s needed at all.
4. R ACS has empsize simulated as exact numbers. Need to convert to code 1-6 as CPS. See data dictionary for empsize definition.
5. [fixed 08/05/2020] ~~Python model cannot replace results using same seed. Replication is fine with R model in GUI.~~
6. [fixed 08/05/2020 – updated cps\_clean\_[year].csv, 3 files] ~~In Python, empsize predicted to be 0 (CPS code NOT IN UNIVERSE) and will be excluded from eligible workers (empsize>=1). Correct this in CPS/ACS imputation.~~
7. R does not store intermediate cleaned FMLA/ACS for sim
8. take/need\_types in ACS are different between Py/R. NT check
   1. cleaned FMLA input data
   2. cleaned ACS input data
   3. NA handling
   4. classifier (logit GLM)