Peer Code Review Summary - FY25 Q3

DIME Analytics (March 2025)
Research Assistant: [redacted]
Principal Investigator: [redacted]
Software used: Stata
Project: [redacted]

Project Diagnostic

WAS REVIEWER ABLE TO RUN THE CODE? (This is only applicable if de-identified data was submitted for review. N/A implies no data was submitted.)	 CODE RUNS AFTER ADJUSTMENTS
DID OUTPUTS REPRODUCE? (N/A implies code did not create any outputs)	 NOT APPLICABLE
WAS REVIEWER ABLE TO UNDERSTAND THE CODE?	 YES, PROJECT MATERIAL WAS ENOUGH
HOW EASY IS IT TO MAINTAIN THE CODE?	 CODE IS EASY TO MAINTAIN
HOW MANY DAYS WOULD IT TAKE FOR REVIEWER TO UNDERSTAND THE CODE WELL ENOUGH TO CONTRIBUTE TO IT?	1

Code Feedback

Strengths:
(1) The structure is the code is good and it is easy to read and follow. (2) The creation of new variables is explained clearly and the use of functions is correct. Names and labels of new variables are self-explanatory (but be careful, because some names are a bit long!). (3) The use of snapshots to save and restore data is correct and efficient.

Areas for improvement:
(1) Cleaning and constructed is not properly separated. I would suggest to encoding and applying the extended missing values before starting all the construct. In this case, not all variables were encoded and the negative values were cleaned in the construct dofile. (2) . (3) Adding additional comments to the code.

For example, some variables are drop but there is no explanation of why. Also, in some merges, there is no explanation about the data that we are merging and why.

Additional feedback:
(1) In do-file hps_lab, in the reshape the season variable is created as a string, but later it is converted to numeric. Same as variable plot, which is later converted to numeric, without doing nothing in between in both cases. So, to make code more efficient, season could be created as a numeric and plot should not be converted to string.

Level of support needed from RA to begin contributing:
I would have needed the questionnaire. The names of some variables were not self explanatory, so it is difficult to understand a few lines of codes without this info. Also, the names of the datasets were not self explanatory either, so without providing any info about which data contains is dataset, is difficult to follow and understand all the merges and the reasoning behind them.

Adoption of Best Practices