DIME Analytics
Peer Code Review - Construction Checklist

Reviewer Details

Reviewer Name: Coder Name:

NOTE: Please complete this checklist **only if** your partner's submission includes **indicator construction** tasks.

Indicator Construction Tasks

This checklist highlights key aspects to review in your partner's **construction scripts/code**. Once completed, please submit it as an attachment along with this form.

Variable Construction Checks

Each variable's purpose and construction logic aligns with its documented definition (e.g., codebook, project documentation).

Correct functions are used and properly explained (e.g., transformations, normalizations).

Categorical variables are properly encoded (e.g., labeled factors in Stata/R).

Merge Checks

If any observations are dropped, a clear justification is provided in the code.

Any mismatches between datasets are explicitly explained in the code.

m:m merge is NOT perfomed

Collapse and group-wise calculations Checks

Missing values are handled appropriately and documented.

If sorting affects results, the data is being sorted on a unique or a combination of unique IDs.

Aggregation functions (sum, mean, etc.) are correctly applied and documented.

Winzorization and Outlier handling

The choice of winsorization or other outlier-handling techniques is clearly justified.

Documentation explains how cutoff percentiles were chosen and why one/both tails were modified.

Constructed Dataset Checks

The dataset follows a **tidy** structure: each row represents an observation, and each variable is a column.

Variable names are informative and follow a consistent naming convention.

Variable labels provide clear descriptions of their contents.

Value labels are informative and consistent (e.g., avoiding cases where varA: 1 = yes, 0 = no, but varB: 1 = yes, 2 = no).

All labels are grammatically correct and do not contain special characters.

Documentation (variable dictionary, variable labels, value labels, comments) is complete and consistent (e.g. codebook constructed using iecodebook).

Each row has a unique ID (or valid combination of unique keys).

The constructed dataset is saved only once throughout the script and is not overwritten multiple times.