DIME Analytics
Peer Code Review - Sampling and Random Treatment Assignment

Reviewer Details

Reviewer Name: Coder Name:

Note: Please complete this checklist **only if** the submission includes **sampling and/or randomization** tasks.

Sampling and Randomization Tasks

This checklist highlights key aspects to review in your partner's **sampling and randomization** code.

- **Sampling** refers to defining the sample frame, i.e., selecting which units (e.g., individuals, households) will be included in the study.
- **Randomization** is the process of assigning treatment and control groups within the sampled units.

Once completed, please submit it as an attachment along with this form.

Sampling Checklist

This section focuses on reviewing the code that performs **sampling**.

Setting Parameters

The script explicitly sets the software version (e.g., `version` in Stata).

A random seed is set for reproducibility (e.g., `set seed` in Stata or `set.seed()` in R).

If yes, the seed values are set using an externally generated, unique seed (e.g., from random.org)

The dataset is sorted by a unique identifier.

Sampling Methdology

The sampling strategy is clearly documented, including inclusion/exclusion criteria.

The sampling method (e.g., stratified, clustered, or simple random sampling) is explicitly stated.

Handling of unequal cluster sizes is clearly specified and justified.

Probability weights, if used, are calculated and stored.

Defining the Sampling Frame

The dataset contains necessary variables for sampling, such as clusters and strata.

Advanced check: Clusters, strata, and IDs do not contain any personally identifiable information (PII).

Advanced check: The resulting dataset is checked for stability using commands like `datasignature`, file hashing, `assert`, and/or `isid, sort`.

Sampling Execution

The script outputs a dataset containing a categorical variable that clearly marks sampled and non-sampled groups.

Final sampling results are stable and reproducible across multiple runs.

If the sample is finalized for field use, a logic switch prevents accidental overwriting.

All outputs include codebooks, value labels, and documentation.

Randomization Checklist

This section focuses on reviewing the code that performs **random treatment assignment**.

Setting Parameters

Verify that the coding environment is properly configured to ensure reproducibility.

The script specifies the software version (e.g., `version` in Stata or package version control in R).

A random seed is set for reproducibility (e.g., `set seed` in Stata or `set.seed()` in R).

If yes, the seed values are set using an externally generated, unique seed (e.g., from random.org)

The dataset is sorted by a unique identifier before randomization to maintain consistency across runs.

Random Treatment Assignment

Treatment assignment is conducted using a script (not manual methods).

The randomization method is well-documented (e.g., simple, stratified, or block randomization).

A categorical variable is created to assign treatment and control groups, with appropriate labels.

No new observations are created during the treatment assignment process.

Randomization is reproducible and produces consistent results across multiple runs.

The output includes a dataset with treatment assignments and relevant documentation.

If the treatment assignment is finalized for use, a logic switch prevents accidental overwriting.