

Analysis Data Construction Guide

Guidance for Original Data Sources:

In many cases analysis data are constructed from multiple original data sources, these original data can come from a variety of sources: other researchers, online databases, government agencies, and even hand-copied from tables on webpages. Authors are required to provide code and/or instructions for obtaining and constructing their analysis data as part of the data verification policy. If your particular use case is not described below, please contact the editors with questions.

For analysis data generated by code:

If you have code files that generate the analysis data files, please provide the code and (if possible) the original data files used by the code. This allows verifiers to run the code and ensure that it does, in fact, create the analysis data files used to generate the manuscript results.

Example:

[Citation] Gill, Jeff, 2020, "Replication Data for: Measuring Constituency Ideology Using Bayesian Universal Kriging", <https://doi.org/10.15139/S3/7NNASB>, UNC Dataverse, V2, UNF:6:g4eRjAslUYrkiMSLD4V5yg== [fileUNF]

For original data downloaded via APIs:

If you used an API to acquire the original data used in your analysis, a full data citation of the original data source must be included and must have a stable URL or DOI leading directly to the data source. Complete instructions for using the API to acquire the exact data used must be included in the documentation (either the README or Codebook). If you are able, please provide the script used that calls the API and collects the data.

Example:

[Citation] Swift, Clint; VanderMolen, Kathryn, 2020, "Replication Data for: Marginalization and Mobilization: The Roots of Female Legislators' Collaborative Advantage in the States", <https://doi.org/10.15139/S3/WVINT7>, UNC Dataverse, V1, UNF:6:u7vdgosus/giRHYAayhLaw== [fileUNF]

For original data pulled from online databases:

If you have used data from online databases to construct your analysis data, please provide full data citations to the online database. In addition to the citation, detailed instructions containing key terms and queries used to navigate the database are required. This information can be provided in the README or codebook, whichever makes the most sense. The more information provided, the easier it will be for secondary users to acquire the original data used.

Example:

[Citation] Swift, Clint; VanderMolen, Kathryn, 2020, "Replication Data for: Marginalization and Mobilization: The Roots of Female Legislators' Collaborative Advantage in the States", <https://doi.org/10.15139/S3/WVINT7>, UNC Dataverse, V1, UNF:6:u7vdg0sus/giRHYAayhLaw== [fileUNF]

For original data hand copied from websites:

There are instances where you may have hand copied data from a site. In order to provide the most information on how these data were obtained, you must include a full data citation for the original data source. Direct links to the page(s) where the data were obtained must be provided with instructions for the specific variable(s) copied into your analysis data.

Example:

[Citation] Manento, Cory; Schenk, Marie, 2020, "Replication Data for: Role Models or Partisan Models? The Effect of Prominent Women Officeholders", <https://doi.org/10.15139/S3/HUN6XO>, UNC Dataverse, V1, UNF:6:WPyeqBPYQ4isactKJoabQQ== [fileUNF]

For original data that cannot be shared due to licensing, access restrictions, or ethical concerns:

In some cases the main hurdle may be sharing the original data files. Please review the terms of use and possible restrictions for all original data sources. If there are no restrictions, you must share these data within the journal Dataverse and provide a full data citation for the original data source. Or, if possible, have the code call the data from the original source site.

If there are restrictions, authors will need to contact the original data producer to see if it is possible to share the original data files with editors and Odum Institute for verification purposes only. The data will be deleted upon successful verification.

If the data still cannot be shared due to restrictions, inform the editors of the restrictions and provide detailed instructions within the documentation on requesting the specific data files used to construct their analysis data. In some cases, Odum may be able to request access to the data and get permissions; however, if there are costs involved, the final decision will be made by the editors as to how access to original data sources are to be handled.