

Quantitative research - Data analysis - Data Analysis Documentation

Set-up & Conduct- Process & Analyse
data

VERSION

4.0

Aim

To ensure that the analyses can be properly reproduced

Requirements

Clear documentation of the data analysis in a log file (for example SPSS syntax, Do file in STATA, R script or Word file), to be able to reproduce the relevant data analyses.

Documentation

Log file including:

- Specific research questions or purpose of the analysis;
- Databases which are used for the analyses (For example 'get file' statement in SPSS syntax);
- All statistical analyses which are executed.
- Add a 'README' tab in your data files and/or separate descriptive document to ensure you can reproduce the results when needed (e.g. in case of audit or inspection, or journal review) and to promote interoperability of your data files.

Responsibilities

- Executing researcher: To document all steps that are taken throughout the data analysis in a log file.
- Project leaders: To regularly check and discuss the data analysis, by using the documentation in a log file.
- Research assistant: N.a.

How To

It is important in respect of reproducibility and efficiency of data analysis that clear documentation of the data analysis takes place. This may be undertaken by creating a log file for all the relevant analyses. This file needs to start off with the research question to be answered and the date of the analysis, and should end with a(n) (provisional) answer to the question.

A log file (e.g. SPSS syntax) can be used to document your analyses (e.g. for an article) to allow you and others to easily retrieve and reproduce everything. Don't forget to always include the name and location of the datafile (e.g. 'get file' in SPSS), so you know which file is related to your analysis (and where they are stored). Log files should include the code for all statistical tests conducted, to serve as an analysis logbook. Place your code in a logical order and distinguish between variable definitions and analyses (e.g. firstly all variable definitions, then the analyses for table 1, then table 2, etc.). A Dutch example of this can be found [here](#).

Tip: annotate your log files (e.g. by using * followed by text in SPSS syntax). Annotations are an important part of documentation of your data analyses and facilitate reproduction of your results and recycling of your code.

References:

Amsterdam Public Health



Amsterdam UMC
University Medical Centers



VRIJE
UNIVERSITEIT
AMSTERDAM



UNIVERSITY
OF AMSTERDAM

Swaen et al: Responsible Epidemiologic Research Practice: a guideline developed by a working group of the Netherlands Epidemiological Society. J Clin Epidemiol. 2018 Aug;100:111-119.

Audit questions

- Does the data documentation for specific analyses contain the following aspects:
 - Specific research questions or purposes of the analysis;
 - Names of databases which have been used for the analysis (e.g. 'get file');
 - Specific code for statistical tests.

LINKS

Link		

DOCUMENT HISTORY

Version	Status	Date	Name
4.0	Revision	08DEC2021	Dr. Marieke Blom and Laura van Dongen Elize Vlainic
3.0	Revision guideline	26OCT2016	EMGO
2.0	Revision format	12MAY2015	EMGO
1.1	English translation	01JAN2010	EMGO
1.0	Title modified: Documentation instead of Report. Adding details with example of documented syntax	21APR2004	EMGO

DOCUMENT APPROVAL

Role	Name	Date
Project Leader	Dr. Seta Jahfari	12MAY2021