



Resources for checking and cleaning data for analysis

Why data checking and cleaning is essential

Data cleaning is the process of detecting and (potentially) correcting corrupt, inaccurate or inconsistent values from a dataset, improving clarity and documenting these steps for better understanding. Data checking and cleaning is essential for many reasons, some of the main ones are:

- ✓ Improves data quality
- ✓ Ensures correct information and allows for reliable data analysis
- ✓ Makes it easier to understand the data
- ✓ Greatly increases productivity avoiding waste of time and resources

Where to find resources on data checking and cleaning

There are many data checking and cleaning resources available, some specific to software such as “MS Excel” or to programming languages such as “R”. The table below provides information on resources.

Author	Resource	Content	Type
UNHCR	Checking and cleaning your data for analysis	Two-part webinar on data cleaning and checking. Part 1: Introduction to data checking and cleaning (Slide deck). Part 2: Case study: Vulnerability assessment of Syrian refugees in Lebanon (2021) (Slide deck).	Webinar
UNHCR	Microdata curation handbook	The handbook is a technical guide for curating personal microdata of people with and for whom UNHCR works, and uploading this data internally on the Raw Internal Data Library (RIDL) and publishing it externally on UNHCR’s Microdata Library (MDL). The section 4.1 “Data check and preparation” provide general tips to identify and correct issues on datasets.	PDF
ACAPS	Data cleaning	This document provides the steps to follow on data cleaning into an Excel spreadsheet for operations collecting primary data during a rapid assessment.	PDF
JIPS	How do we process and analyse the dataset?	Practical steps to follow to process and analyze datasets.	PDF
IMPACT	Data cleaning guidelines for structure data	The minimum standards for data cleaning, and the process of data cleaning.	PDF
IMPACT	Data cleaning minimum standards checklist	Minimum standards that all IMPACT teams (field and HQ) are expected to adhere to in the data cleaning and processing stage of the research cycle.	PDF
IMPACT	Funtion reference	A list of functions that can be utilized to support data cleaning.	Gibtub

