What is IDA?

IDA (=Initial Data Analysis) aims to provide reliable knowledge about the data to enable valid statistical analyses and interpretation.

1

Know your data

How was the data collected?

What is missing? (missing values, patterns of missingness)

Univariable descriptions of variables in the planned models

Multivariable descriptions

Timing of measurements or follow-up period

2

IDA results -oops!

Suspicious values (outliers, inconsistent time differences)

Suspicious rows (eligibility)

Missing data

Unexpected distributions of variables

Unexpected heterogeneity of the analysis population

Highly correlated variables

Update analysis plan!

3

Avoid sneak peaks

Stick to the research aims: IDA does not generate hypotheses ✓

Change the analysis plan based on data properties ✓

Explore associations with the outcome ⊘

Data driven selection of analyses ⊘

4

Make IDA reproducible

Keep track of changes (data and program scripts) from **ALL** team members!

Document metadata

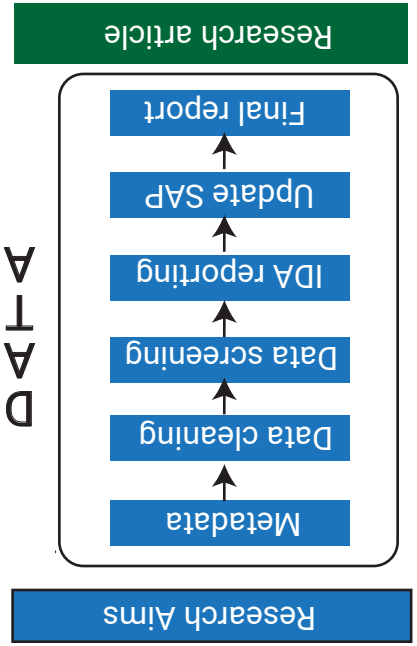
Make data cleaning code repeatable

IDA takes time

It saves more time later!!

5

IDA Framework



6

Marianne Huebner
2025-06-02

Initial Data Analysis (IDA) What? Why?

STATS DOC

CC-BY-NC

CSTAT Center for Statistical Training & Consulting

cstat.msu.edu