HANDLING MISSING DATA

Real-world datasets often contain missing values, a problem not always avoidable, even in well-designed research. Missing data should be handled carefully; otherwise, they may skew your analysis and compromise your results.
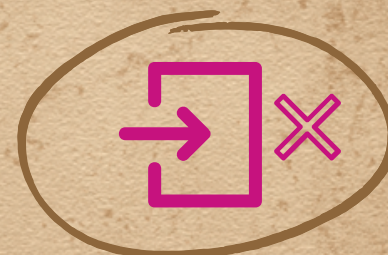
COMMON SOURCES OF MISSING VALUES

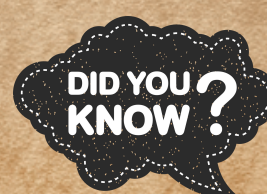| Some people skip survey questions	Some data can be hard to obtain	Secondary sources may have missing data points	Data entry errors	Equipment malfunction

RECORDING MISSING VALUES

Researchers may follow different approaches to note missing data points on datasets. Below are some considerations about these techniques:

Never replace missing values with a zero (0). They are indistinguishable from a true zero. Zeros represent data to a computer, not its absence, and will distort results.

Avoid -999 or 999. These are not recognized as missing by many programs without user input and can inadvertently be computed into the calculation.

Avoid hyphens and other symbols. They can cause problems with data types.

NA and NULL are compatible with most software and can be used with proper annotation.

Leave missing values as blank cells so it won't affect the analysis and produce incompatibility issues. Make sure to include that information on your README.txt file.

MISSINGNESS ASSUMPTIONS & POSSIBLE REMEDIES

Researchers should run a missing value analysis to see what the missingness patterns may say about the remaining observed data. Results will guide assumptions and help identify the best treatment to mitigate problems before the data analysis.

Missing Completely at Random (MCAR)	Missingness does not depend on the missing or the observed data	Deletion of cases/rows (listwise), missing data points (pairwise) or variables (columns).
Missing at Random (MAR)	Missingness depends only on the observed data	Imputation/replacement.
Missing Not at at Random (MNAR)	Missingness depends on unobserved data	Find data and augment the dataset.

DID YOU KNOW? The UCSB DataLab offers stats consultation to campus affiliates.
Learn more: datascience.ucsb.edu/consulting