

Common Sources of Data Errors and Error-Checking Techniques

Sources of Data Errors

The table below displays the evaluation stages at which you might find different sources of data errors. You may use the error-checking techniques and helpful hints in the next section to avoid or correct these data errors as they occur.

Evaluation Stage	Sources of Errors	
	Common Mistakes	Example
Data collection	<ul style="list-style-type: none"> • Missing questions • Unanswered questions • Incongruent or extra responses to a single question • Wrong box checked • Response is not readable • Writing error • Response is out of expected range 	A survey was administered in both paper and electronic formats. However, the paper survey was an older version that included questions with seven possible response options, whereas the electronic survey had questions with five possible response items.
Data entry and cleaning	<ul style="list-style-type: none"> • Data incorrectly transferred from the instrument • Values entered in the wrong field • Inadvertent deletion or duplication during database handling • Outliers and inconsistencies carried over from the instrument • Values incorrectly entered • Values incorrectly changed during previous data cleaning 	A survey was administered with a five-point rating scale, as well as a “not applicable” option. When these survey responses were entered into a database, the scaled responses were correctly coded as 1 through 5, but “not applicable” was coded as 6.
Data analysis	<ul style="list-style-type: none"> • Data incorrectly extracted from the database • Data incorrectly extracted or coded • Inadvertent deletion or duplication during analysis • Outliers and inconsistencies carried over from the database • Data incorrectly extracted or coded • Sorting errors (spreadsheets) • Data cleaning errors 	A survey was analyzed using a statistical software program. Several duplicate responses were identified. In an attempt to remove the duplicates, the duplicate as well as the original responses were deleted.

Error-Checking Techniques

Descriptive analysis: Calculate descriptive statistics, including the mean and the range, and check that they are sensible. Does the mean seem reasonable? Is the range of values inside the range of theoretically possible examples?

Double entry: Arrange for two or more people to enter the same data and then check for discrepancies.

Logic check: Carefully review the electronically entered data to make sure that the answers to the different questions make sense. For example, if teachers indicated that they were not evaluated during the school year in one survey question, it would be illogical for them to rate their satisfaction with evaluator feedback in their response to another question.

Spot-checking: Randomly select several participants and check their raw data against the data entered in a spreadsheet, document, or database. If you find any errors, randomly select another group and check their data in a similar manner. Examine the overall pattern of the data for data-entry or coding mistakes. For example, in spot-checking the age variable for several participants, you notice the value for age is 100. You know that this is an error because your sample consists of high school students.

Helpful Tips

- Always keep a copy of original files. If an original file is modified, save it with a new name in a different folder.
- Train data entry or data management staff.
- Develop instructions for data entry and data manipulation and establish data decision rules.
- Keep a log of the data errors found and the changes made. The log should include information about who found and corrected the error, and it should be easily accessible to anyone working with the data.
- Always triple-check everything and screen data for errors frequently.

Note. Adapted from the following sources:

The Pell Institute. (n.d.). *Enter, organize, & clean data.* <http://toolkit.pellinstitute.org/evaluation-guide/analyze/enter-organize-clean-data/>

United Nations High Commission for Refugees. (n.d.). Cleaning data. In *Coordination toolkit.* <http://www.coordinationtoolkit.org/wp-content/uploads/130813-Data-cleaning.pdf>

Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R. & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLOS Medicine*, 2(10), Article e267. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1198040/#!po=27.7778>

This handout was prepared under Contract ED-IES-17-C-0005 by Regional Educational Laboratory Central, administered by Marzano Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.