

## CHAPTER OVERVIEW

### 2: Understand Your Data

Good data are the basis of any sort of regression model, because we use this data to actually construct the model. If the data is flawed, the model will be flawed. It is the old maxim of garbage in, garbage out. Thus, the first step in regression modeling is to ensure that your data is reliable. There is no universal approach to verifying the quality of your data, unfortunately. If you collect it yourself, you at least have the advantage of knowing its provenance. If you obtain your data from somewhere else, though, you depend on the source to ensure data quality. Your job then becomes verifying your source's reliability and correctness as much as possible.

[2.1: Missing Values](#)

[2.2: Sanity Checking and Data Cleaning](#)

[2.3: The Example Data](#)

[2.4: Data Frames](#)

[2.5: Accessing a Data Frame](#)

---

This page titled [2: Understand Your Data](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [David Lilja \(University of Minnesota Libraries Publishing\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.