

Writing about Data

A Handout from the EcoTeach Center, Duke University
Prepared by Paul Dudenhefer, Writing Tutor



In empirical economics papers, it is customary to describe the data one uses, usually in a section of its own headed “Data.” The best way to learn about writing a data section is to read several data sections in the literature on your topic and pay attention to the kinds of information they contain.

What you tell your readers about your data will depend in large part on the kind of analysis you are conducting. Generally speaking, however, your data section should do at least the following:

- Identify the data source (e.g., “This study uses data from the 1999 wave of the Panel Study of Income Dynamics.”).
- Provide a general description of the data source (number of observations, population groups sampled, time period during which the data were collected, method of data collection, etc.).
- State the strengths and weaknesses of the data source, especially how they relate to other data sources used in the literature (does the data source provide more observations, and/or more recent observations, than other sources? was the data collected in a more reliable manner? etc.).
- Note any features of the data that may affect your results (were certain populations overrepresented or underrepresented? is there attrition bias or selection bias?)
- Explain how you may have computed certain kinds of data that the source did not give directly (did you have to add/subtract/multiply/divide two given pieces of data to get a third?)
- Describe how you selected your sample (did you have to eliminate certain kinds of observations?)

Data sections often contain a table of descriptive statistics, statistics of relevance about the sample. These statistics usually include the mean (e.g., mean income, mean age, mean years of schooling, etc.) and standard deviation. For categorical data (like race), however, you do not report a mean; instead, you report the percentage of the observations in each group.

Again, the nature of your project will determine how best to describe your data. And it bears repeating that the best way to learn how to write a data section is to read several data sections in the literature and pay attention to the kinds of information they contain.

