

Writing About Operations Research – Data

1 The data section

- The purpose of the **data** section is to describe
 - the **data** you used in your study, and
 - the data **wrangling** and/or **analysis** that you performed in order to input the data into your **model**.
- Suggested outline:
 1. **Basic description of data.** (1 paragraph per data source)
 2. **Summary statistics and visualizations.** (1-3 tables or visualizations per data source)
 3. **Data wrangling and analysis.** (1-3 paragraphs.)
- See below for more details on each part of the outline.

1.1 Basic description of data

- For each data source:
 - a. Describe where the data came from.
 - b. In words, describe the contents of your data.
 - For tabular data: What does each row represent? For each row, what do the columns represent?
 - c. Describe the size of your data.
 - For tabular data: How many rows do you have?

1.2 Summary statistics and visualizations

- Identify a few key variables in your data, and give your reader a sense of these variables:
 - a. Report their summary statistics (e.g., mean, standard deviation, min, max) in a table.
 - b. Provide histograms to show the distributions of the variables.
 - c. Provide other visualizations (e.g., scatter plots) to show the relationships between the variables.
- Some general guidelines on presenting tables and graphs:

- You should introduce every table or figure to your reader with some text, e.g.

The histogram in Figure 5 shows the distribution of car prices in the data.

- Number each table or figure.
- Give a descriptive caption for each table or figure.
- Use a descriptive title for each table column.
 - ◊ e.g., replace raw variable names such as “beer_consumption_us” with reader-friendly names like “US Beer Consumption (in millions of gallons)”
- Label the axes of each graph.

1.3 Data wrangling and analysis

- Sketch your data wrangling process.
 - Give enough detail so that someone with a computing background can replicate your work, but...
 - Do not describe every step of your data wrangling in painstaking detail.
 - For example, suppose your data wrangling outputs a “final” table to be input into a statistical model.
 - ◊ In this case, simply describe which tables were combined and the layout of the final table, like this:

We manipulated and merged Tables A and C described above to create a table suitable for analysis. This table consists of the following columns: Q, R, S, T. Each row of this table represents one X; this table has a total of Z rows.
 - As another example, suppose your data wrangling outputs sets and parameters to be input into an optimization or simulation model.
 - ◊ In this case, briefly explain how these sets and parameters were computed, like this:

We took the average of the X values in Table A described above for each group i to obtain the parameters v_i for each item $i \in S$.
 - ◊ Do not explain your computations using a specific programming language, such as Python or R. Instead, describe your computations in words, pseudocode, or mathematical formulas.
- Describe other data analyses, if applicable.
 - We will discuss this in more detail later in the semester.