Writing About Operations Research – The Data Section

1   Overview

- Your project report must explain your approach to solving your problem in detail.

- This includes describing

  - the **data** you used to solve your problem,

  - the **wrangling** and/or **analysis** of the data that you needed to perform in order to input the data into your **model**.

- The way you need to write about this will depend on the exact nature of your project.

- Below are some general guidelines.

2   Describing the data

- Give your reader a sense of the data you used to solve your problem.

- Start by describing the sources of your data.

- Next, describe in words the contents of your data.

  - If you have tabular data, what tables do you have? In each table, what do the rows and columns represent?

  - If your data is made up of sets and parameters, what do they represent?

- Finally, describe the size of your data.

  - If you have tabular data, in each table, how many rows do you have?

  - If your data is made up of sets and parameters, how big are the sets?

3   Describing the wrangling and analysis

Data wrangling

- 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, <u>briefly</u> 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 <u>not</u> explain your computations using a specific programming language, such as Python or R. Instead, describe your computations in words, pseudocode, or mathematical formulas.

Fitting distributions

- If you are fitting distributions to your input data, you need to describe your input data analysis.

- For each set of input data that you fit to a distribution:

    - State the candidate distributions that you fit to the data, and discuss why you chose these distributions.

        - ◇ Support your choices using histograms, characteristics of the data itself, and context.

    - Evaluate the fit for each candidate distribution. This includes:

        - ◇ A graph comparing a histogram of the data with plots of the candidate distribution densities.

        - ◇ (Optional, encouraged if useful) A QQ-plot and PP-plot of the data for each candidate distribution.

        - ◇ Goodness-of-fit statistics for each candidate distribution, presented in a table.

    - Conclude by stating your choice of distribution, and why you made that choice.