

MATH 189: Exploratory Data Analysis and Inference

Spring 2020

Structure of a Data Analysis Report

A data analysis report is somewhat different from other types of professional writing that you may have done or seen, or will learn about in the future. It is related to but not the same as:

- A typical psych/social science paper organized around “intro/methods/analysis/results/discussion” sections.
- A research article in an academic journal.
- An essay.
- A lab report in a science class.

The overall structure of a data analysis report is simple:

1. Introduction
2. Body
3. Conclusion(s)/Discussion
4. Appendix/Appendices

The data analysis report is written for several different audiences at the same time:

- **Primary audience: A primary collaborator or client.** Reads the *Introduction* and perhaps the *Conclusion* to find out what you did and what your conclusions were, and then perhaps fishes/skims through the *Body*, stopping only for some additional details on the parts that he/she thought were interesting or eye-catching. *Organize the paper around an agenda for a conversation you want to have with this person about what you’ve learned about their data: e.g., from most general to most specific, or from most important to least important, etc. Provide the main evidence from your analysis (tabular, graphical, or otherwise) in the Body to support each point or conclusion you reach, but save more detailed evidence, and other ancillary material, for the Appendix.*
- **Secondary Audience: An executive person.** Probably only skims the *Introduction* and perhaps the *Conclusion* to find out what you did and what your conclusions are. *Leave signposts in the Introduction, Body and Conclusion to make it easy for this person to swoop in, find the “headlines” of your work and conclusions, and swoop back out.*
- **Secondary Audience: A technical supervisor.** Reads the *Body* and then examines the *Appendix* for quality control: How good a job did you do in (raising and) answering the interesting questions? How efficient were you? Did you reach reasonable conclusions by defensible statistical methods? Etc. *Make specific cross-references between the Body and specific parts of the Appendix so that this person can easily find supporting and ancillary material related to each main analysis you report in the Body. Add text to the technical material in the Appendix so that this person sees how and why you carried out the more detailed work shown in the Appendix.*

The data analysis report has two very important features:

- It is organized in a way that makes it easy for different audiences to skim/fish through it to find the topics and the level of detail that are of interest to them.
- The writing is as invisible/unremarkable as possible, so that the content of the analysis is what the reader remembers, not distracting quirks or tics in the writing. Examples of distractions include:

- Extra sentences, overly formal or flowery prose, or at the other extreme overly casual or overly brief prose.
- Grammatical and spelling errors.
- Placing the data analysis in too broad or too narrow a context for the questions of interest to your primary audience.
- Focusing on process rather than reporting procedures and outcomes.
- Getting bogged down in technical details, rather than presenting what is necessary to properly understand your conclusions on substantive questions of interest to the primary audience.

It is less important to worry about the latter two items in the Appendix which is expected to be more detailed and process-oriented. However, there should be enough text annotating the technical material in the Appendix so that the reader can see how and why you carried out the more detailed work shown there.

The data analysis report isn't quite like a research paper or term paper in a class, nor like a research article in a journal. It is meant, primarily, to start an organized conversation between you and your client/collaborator. In that sense it is a kind of "internal" communication, sort of like an extended memo. On the other hand it also has an "external" life, informing a boss or supervisor what you've been doing.

Now let's consider the basic outline of the data analysis report in more detail:

1. **Introduction.** Good features for the Introduction include:

- Summary of the study and data, as well as any relevant substantive context, background, or framing issues.
- The "big questions" answered by your data analyses, and summaries of your conclusions about these questions.
- Brief outline of remainder of paper.

The above is a pretty good order to present this material in as well.

2. **Body.** The body can be organized in several ways. Here are two that often work well:

- *Traditional.* Divide the body up into several sections at the same level as the Introduction, with names like:
 - Data
 - Methods
 - Analysis
 - Results

This format is very familiar to those who have written psych research papers. It often works well for a data analysis paper as well, though one problem with it is that the Methods section often sounds like a bit of a stretch: In a psych research paper the Methods section describes what you did to get your data. In a data analysis paper, you should describe the analyses that you performed. Without the results as well, this can be pretty sterile sounding, so I often merge these "methods" pieces into the "Analysis" section when I write.

- *Question-oriented.* In this format there is a single Body section, usually called "Analysis", and then there is a subsection for each question raised in the introduction, usually taken in the same order as in the introduction (general to specific, decreasing order of importance, etc.). Within each subsection, statistical method, analyses, and conclusion would be described (for each question). For example:

2. Analysis

2.1 Success Rate

Methods

Analysis

Conclusions

2.2 Time to Relapse

Methods

Analysis

Conclusions

2.3 Effect of Gender

Methods

Analysis

Conclusions

2.4 Hospital Effects

Methods

Analysis

Conclusions

Etc. . .

Other organizational formats are possible too. Whatever the format, it is useful to provide one or two well-chosen tables or graphs per question *in the body of the report*, for two reasons: First, graphical and tabular displays can convey your points more efficiently than words; and second, your “skimming” audiences will be more likely to have their eye caught by an interesting graph or table than by running text. However, too much graphical/tabular material will break up the flow of the text and become distracting; so extras should be moved to the Appendix.

3. **Conclusion(s)/Discussion.** The conclusion should reprise the questions and conclusions of the introduction, perhaps augmented by some additional observations or details gleaned from the analysis section. New questions, future work, etc., can also be raised here.
4. **Appendix/Appendices.** One or more appendices are the place to put details and ancillary materials. These might include such items as
 - Technical descriptions of (unusual) statistical procedures
 - Detailed tables or computer output
 - Figures that were not central to the arguments presented in the body of the report
 - Computer code used to obtain results.

In all cases, and especially in the case of computer code, it is a good idea to add some text sentences as comments or annotations, to make it easier for the uninitiated reader to follow what you are doing.

It is often difficult to find the right balance between what to put in the appendix and what to put in the body of the paper. Generally you should put just enough in the body to make the point, and refer the reader to specific sections or page numbers in the appendix for additional graphs, tables and other details.

Evaluation criteria

Each homework assignment/project will deal with a real data set.

Projects are evaluated based on all of the following five criteria:

- 1. Composition and Presentation**
(literature review, scientific writing and presentation of logic arguments)
- 2. Basic analysis**
(mentioned during lectures and discussions)
- 3. Graphs and tables**
(must be of the best scientific quality)
- 4. Advanced analysis**
(additional materials prepared outside of topics mentioned during lectures)
- 5. Appendix for technical material**
(all statistics **must** be supported with mathematical/statistical arguments for its usage - consistency, convergence and expectations)