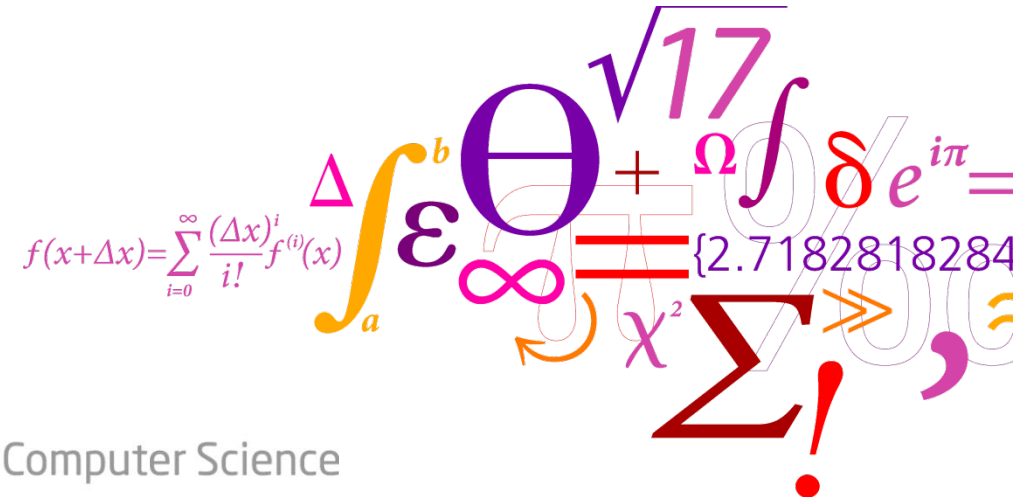
Writing Statistical Reports

January 9, 2025

Anders Stockmarr
Section for Statistics and Data Analysis, DTU.
anst@dtu.dk



DTU Compute
Department of Applied Mathematics and Computer Science

Why write statistical reports?

- At University: Perhaps because teachers says so.

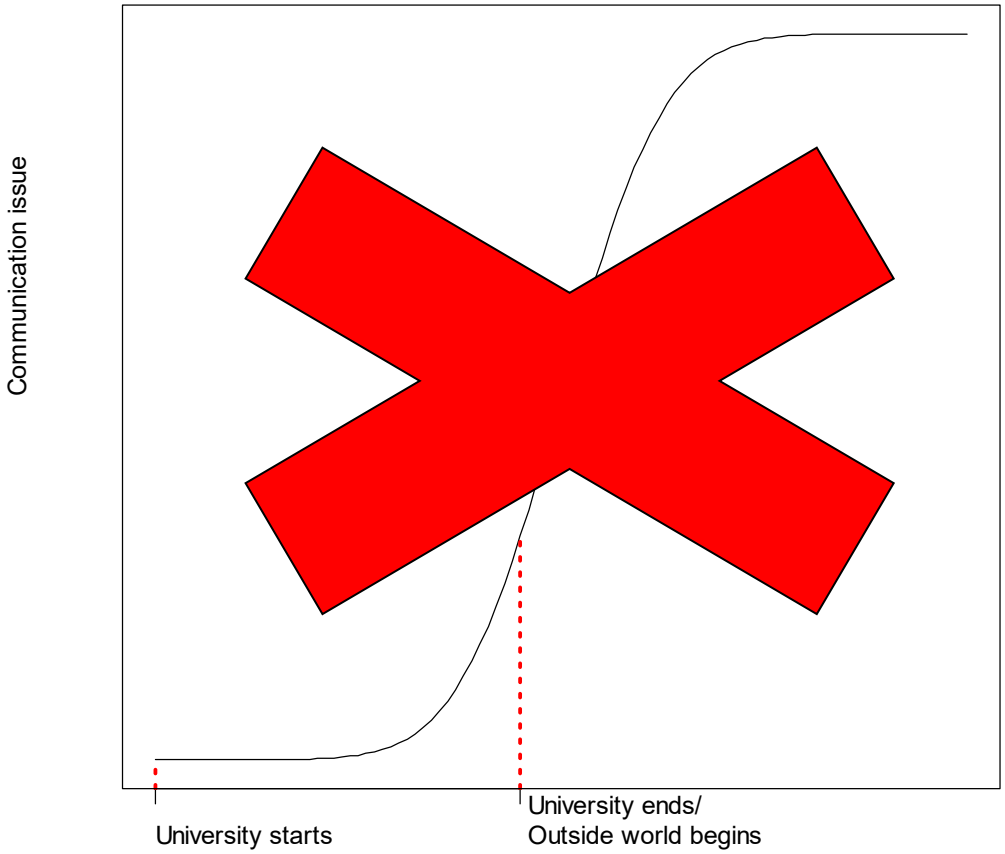
- Other reasons cold be:

- To document what you have done in a study?

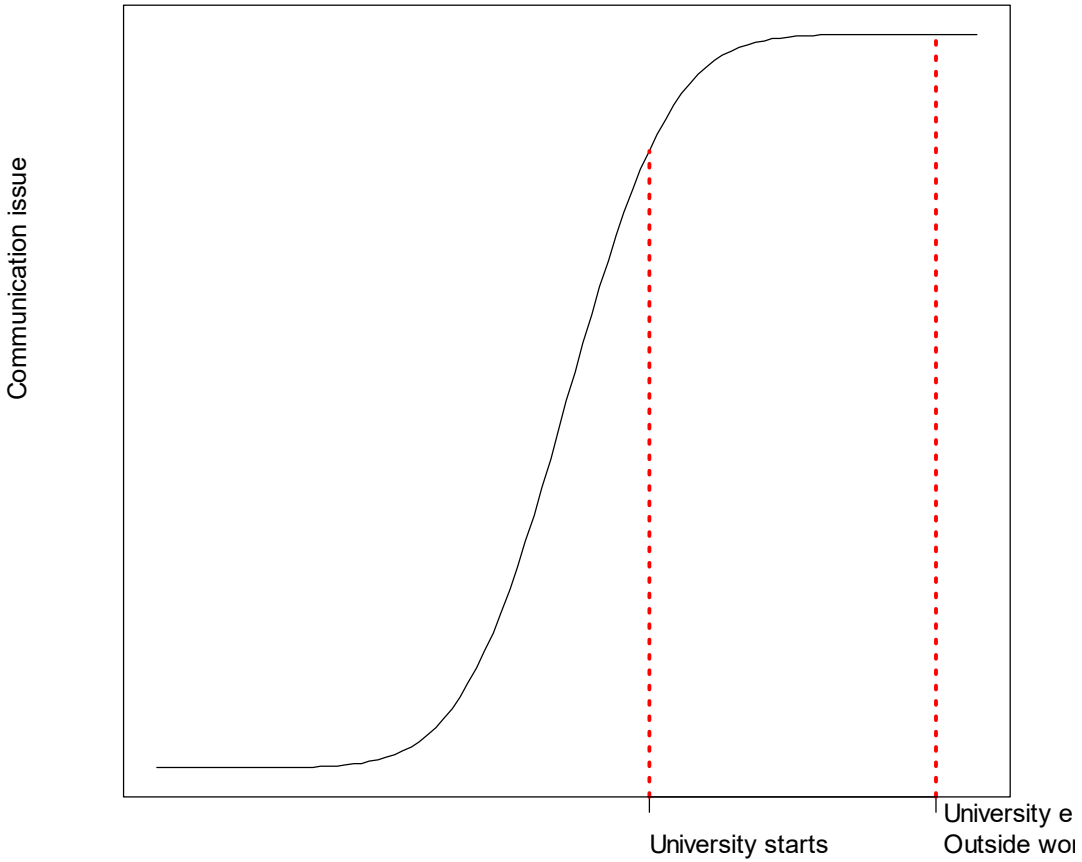
- Yes, always, but that is not the whole issue when it comes to **statistical** reports.

- Statistics is a science;
- applied statistics is the **application of this science within another discipline**;
- the reader of the report is therefore **not necessarily equipped** to put the results in context;
- there is therefore, also, an issue of ***communication***.

Statistical reports at university and elsewhere



Statistical reports at university and elsewhere



Statistical Reports

- Issues:
  - Document your methods;
    - **matter-of-fact, listing, specifying.**
  - Tell your story;
    - **Why, what, how.**
  - Communicate your story;
    - **Context, examples, discussion.**

- Portrait **the Client** – the expected reader/end-user (this person may or may not exist IRL).

Statistical Reports – the Client

- Portrait **the Client** – the expected reader/end-user.

- **The Client** could be an external person;

- Or **the Client** could be **yourself**, in 12 months time when you have forgotten most about the analysis.

Statistical Reports – the Client

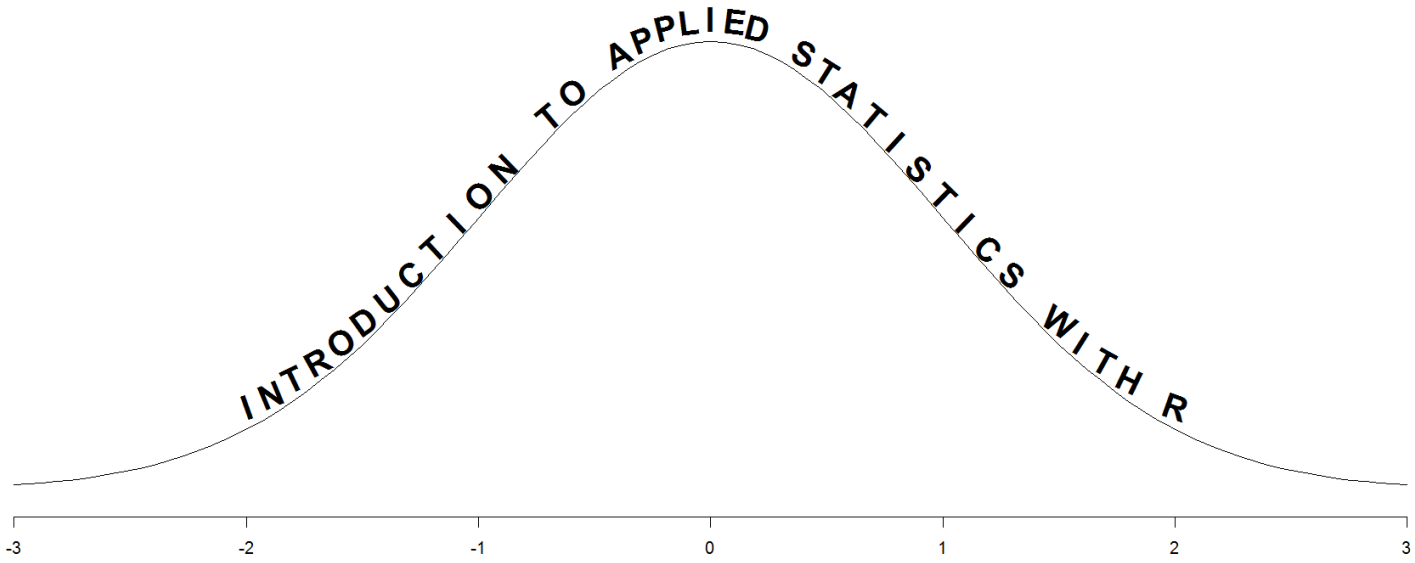
- Portrait **the Client** – the expected reader/end-user.
  - What does **the client** already know? **(basic/advanced science on the subject, statistical methods, project circumstances)**
  - What does **the client** not know? **(basic/advanced science on the subject, statistical methods, project circumstances)**
  - What is the interest of **the Client**? **(research question, p-values, effect parameters, issues with data handling)**
  - What is NOT the interest of **the Client**? **(R code, issues with data handling, intermediate analyses)**
- Adapt the contents and structure (not the results though ☺) to fit the knowledge and interests of **the Client**.

Report Structure – Contents of the Good Report

- Front page.
- A summary (less than one page).
- A table of contents.
- Introduction.
- Description of data.
- Statistical analyses, results and discussion.
- Conclusion.
- Appendices.

Front page

- Your **name**.
- Your **Affiliation** (Here: student number)
- The **title** of your report
- Perhaps a nice **picture** to display, to give an idea of the contents:



Summary

- Should be short.
- Summary of questions posed, the main results and conclusions.
- **The Client will read this section**: Include what you want **the Client** to know about your study; and what **the Client** should look for details about in the report.
- Other readers than the Client may never read more than this! Be sure to include the main findings here.

Introduction

- Should set the context of the report, give a background description and a formulation of the main reason for the work.
- Should contain one or more specific research question(s) that the work is (was) supposed to answer.
- Pay attention to the **Client** portrait here. What level of information is necessary to set the context properly?

Description of Data

- Describe the data!
- What are the different data types available? which are outcomes, which are explanatory variables etc.
- Possible means of describing data:

  - verbal means;
  - summary tables; means, medians, quantiles etc.
  - graphical means; scatter plots, histograms etc.

Statistical Analyses, Results and Discussion

- Three different sections!

Statistical analyses:

- For each research question, specify the statistical methods used, and why you use exactly these.

- Adress if the methods are proper. Are the any assumptions behind the methods (normality, independence etc.) that you needed to verify, and how did that go?

Results:

- Present results in a matter-of-fact manner.

To communicate results, graphs are nice, but tables are needed as well. Only in special situations should tables not be supplied.

  - Example: For investments on the stock market, you need a value that is not read off from a graph.

Statistical analyses, Results and Discussion

Discussion:

- **Interpret** the results in the context, in order to communicate the results to the Client.

- Example: Result: ”$\beta<0$, $p=0.0001$”. Interpretation: ”The coefficient is significantly negative, so people who were given treatment A have a higher one-year expected survival rate”.

- **Discuss** the results in the context, and relate it to the interests of the Client:

- Example: ”The sample was chosen representatively, so the results applies at the population level. However, the data were self-reported so the estimate may be biased downwards”.

- In particular, determine the level of reliability of the results. Did the data pass the model control? And similar issues.

Conclusion

- Summarize the analyses.
- Make sure that you adress the research question(s) formulated in the introduction.
- Make the conclusions with the weight that your analysis indicate that you can (ie. for example, reservations for things like self-reported data).
- The Conclusion is **the second section that you can count on the Client reading**. Therefore, the interests of **the Client** (ie. questions /context of interest) should be addressed.

Appendices

- This is for issues of documentation.
- Necessary information, which is not of immediate interest.
- Examples: R code, unproblematic model control charts, graphs which are not of major interest.

Specifics – Level of Information

- Pay attention to that you **supply enough information** so that the demands for documentation are satisfied.

- **Replication principle**: With access to data, **the Client** should be able to replicate your analysis, based on facts from the report.

- Pay attention to that you do not **supply redundant information**: The most common communication error is oversized reports on relatively simple problems. **Oversized reports** will simply **not be read in full** by **the Client**.

- For each piece of information, consider if it is necessary for either:

- Documentation/the replication principle;
- Results;
- Communication.

If a piece of infomation is not neccessary, **LEAVE IT OUT**.

Specifics – the Client

- The Client is not stupid, avoid patronizing **the Client**.
- On the other hand, the **Client** portrait may reveal lack of knowledge in specific areas: Adress these areas.
- In particular, **the Client**, as the intended reader, may be a non-statistician, and not trained in statistical methodology. If this is the case, consider this when communicating the results.
- The **Client** interests are typically the results and their context, and not how you got there. IE: The R code is typically **NOT** the interest of the Client. Put the R code in an appendix, the justification of the R code is often only for documentational purposes.

Specifics – Confidence Intervals etc.

- An estimator is usually worthless unless accompagnied by an assessment of the level of statistical uncertainty of it. Supply this uncertainty, always.
- Statistical uncertainty can be represented in many forms; but usually a **standard confidence interval** is the right choice.
- Consider this for predicted values as well, also in graphical presentations.
- Include the information in results and discussion.
- Include the full relevant information in summary statements. Example:
  - Simple statement: ”Treatment A was significantly better than Treatment B.”
  - Detailed statement: ”The odds of success were increased with 80% for treatment A when compared to treatment B (p=0.002), with a 95% confidence interval of 65% to 87%”.
- The better detailed statement both includes **effect size**, **significance level** and **statistical uncertainty**.

The Client and Your Report.... Make the Proper Reservations

"'BE CAREFUL'! ALL YOU CAN TELL ME IS 'BE CAREFUL'?"