STAT 501: Statistical Consulting

Ivy Liu, Nokuthaba Sibanda, Lisa Woods

School of Mathematics and Statistics

Victoria University of Wellington

Part I

“Good communication skills are required to be an effective consultant.”

- Cabrera and McDougall

Statistical consulting process

There are two common elements in the statistical consulting process:

1. Verbal interaction with the client(s) which continues until the project ends – covered in next lecture.
2. Preparation of technical summaries, report writing and presentation of results.

These are key communication skills that are often not covered during a statistician's training, but left as something to be picked up later.

Writing consulting reports

In our contacts with leading researchers in industry, the most prevalent complaint we hear is that statistics graduates have great difficulty in writing reports. Indeed, there are cases where industry scientists involved in a project routinely reject the inclusion of a junior statistician who does not possess good written communication skills.

- Cabrera and McDougall

Report outline

- Title page
- Introduction
- Materials & Methods
- Results
- Conclusion
- References
- Appendices

Title page

- Information on the title page should give the reader a clear idea of the content and important points of your report.
- Structure
  - **Project title** – choose an informative or interesting title
  - **Authors**
  - **Date**
  - **Executive summary** – must be very short and to the point. The executive summary contains a brief account of your conclusion. Do not describe the problem or discuss the methods used, simply state the results and conclusion.

The Stork Reality

by

N. Author
A. N. Other

1 July 2020

Executive Summary

This report contains evidence of a strong correlation between the population of Oldtown and the number of storks observed over a period of seven years. Clinical tests conducted during this same period however, account for 95% of all known births in Oldtown. A causal effect is **not** suspected for the other 5%.

Example title page

Introduction

- The Introduction should state the problem, as in the executive summary, but with more background information.
- You can indicate here your basic method of approach to the problem, stating in plain language what you plan to share with the client.
- **Do NOT simply copy the problem description from the client**: restate in your own words, highlighting the statistical issues (e.g. design structure, nesting).
- The introduction should state briefly, at the end, a very brief description of each section of the report.

Materials & Methods

- Here you identify the aspects of the study that are key to understanding the design.
- If the study is an experiment, explain briefly why the design involves relevant features (e.g., subsampling, blocking or repeated measures).
- If the study is an observational design, explain the inclusion criteria, what exclusions were made and why.
- Write in plain language using the client's terminology where possible.
- Questions and concerns about the nature of experimental units, random assignment and the way measurements were taken should be addressed in this section.

Results

- This section should speak directly to the client. Avoid patronizing phrases like `obviously', `it is clear that'.
- Write in plain English with enough depth to let the client understand the spirit of what you did, but you can refer any essential technical details to an Appendix.
- Figures and tables should be well-annotated, self-contained, and tell a major portion of the story.
- This section should state what you found and how you found it (sometimes it is useful to do this in two sections).

Results

- The section should be organized to tell the story you finally uncovered, **not** the circuitous path you took to get there.
  - For the **case study report** I would like to see the exploratory analyses you do to decide on the method of analyses.
  - For the **client reports**, you can just report the results of your analyses with brief statements of why you used the method you used.

- Some results tables may or may not be important (e.g., ANOVA tables). Think hard about what to include and what can be left out. **Keep it simple.**

Conclusion

- The concluding section should restate the results in a concise fashion, in different words and more depth than the summary.
- Refer back to the key questions posed by the client and relate these to the results.
- Put the conclusions (which should be substantiated by the results) in the larger context of the industry. For example, depending on the industry, the language of monetary costs, efficiency gains or lives saved would be extremely important in reports to senior management.

Conclusion

- The conclusion looks back over the study and analysis, and can raise questions and suggestions about future investigations.
- The client may not be able to collect more data, or more precise data, due to costs and other considerations; keep this in mind when you make recommendations.
- Remember to respect the client throughout, but especially in the conclusions.

References and Appendix

References

- This gives the list of books, articles, software and other documents that you cited in the text.
- Do not include irrelevant references that are not specifically related to the project.

Appendix

- This should be used for computer output, computer code, graphs or any other reference documentation that need not appear in the main text.
- **Do not** use the Appendix as a dumping ground! It should be presented in an organised and appropriate format for the reader (e.g., use different sections for figures, tables and computer code).

Final word on consulting reports

- There is no right answer to most real problems; however, your report should have enough detail that **another statistical consultant** could figure out what you did and (largely) reproduce it.
- Remember that as a consultant one of your roles is being an educator. Teach the client good organization and reasoning skills (where needed); these are far more important than complex statistical analyses.
- You are also an artist, and will learn best by practice and apprenticeship. Take charge of your learning; find out about the resources at your disposal, including your lecturers, the computer and the library. Your own mistakes are also a fundamental part of the learning process.

Part II

GROUP DISCUSSIONS AND FEEDBACK

Group discussion sample reports

Group discussion on reports

In your groups

- Discuss the three sample reports and agree on a ranking for each of the reports
- For the last ranked report, discuss and create a list of improvements you might make to the report, taking into account:
  - Clarity – is the report clear and easy to understand?
  - Content – were all the key elements (Introduction, Methods, Results, Conclusion) included in the report?
  - Is there enough detail for you to figure out what was done and to largely repeat the analysis yourself?

Group feedback – main points

Group A

Group feedback – main points

Group B

Case Study Assignment

PISA Science scores

The dataset contains science scores by country from the Programme for International Student Assessment (PISA) 2006, along with GNI per capita (Purchasing Power Parity, 2005 dollars), Educational Index, Health Index, and Human Development Index from UN data. The key variables are as follows (variable names in bold):

Overall Science Score (average score for 15 year olds)

Interest in science

Support for scientific inquiry

Income Index

Health Index

Education Index

Human **D**evelopment **I**ndex (composed of the Income index, Health Index, and Education Index)

PISA Science scores

- Education index is measured by mean of years of schooling for adults aged 25 years and expected years of schooling for children of school entering age.
- Health index is based on life expectancy at birth.
- The wealth component is based on the gross national income per capita.
- The HDI sets a minimum and a maximum for each dimension, and values indicate where each country stands in relation to these endpoints, expressed as a value between 0 and 1. More information on the HDI measures can be found here.

Case study report

You've been tasked as a consultant to summarise the data as follows:

- Identify the top ten countries in overall sciences scores.
- Determine whether wealthier (higher income) countries tend to provide more support.
- Determine whether students from countries with a higher HDI show more interest in science.
- Determine whether the level of support and interest have an effect on overall science scores.

Analysis and reporting

Things to consider/include:

- Exploratory data analysis
- Dealing with missing values
- Non-linear relationships
- Correlation vs regression

For helpful advice on planning your report, see the Cengage Appendix on Statistical Reporting.

Your report

- Up to 10 pages long, excluding the title page, references and appendices
- Will be assessed more on the content and clarity of your report than on the analysis itself – there is more than one way to answer the questions asked.
- Marking criteria (**35 marks**)
  - ❑ Introduction [6]
  - ❑ EDA [4]
  - ❑ Methods [6]
  - ❑ Results [6]
  - ❑ Discussion [6]
  - ❑ Presentation [5]
  - ❑ Depth and breadth of report [2]

Part III

VERBAL INTERACTION – PREPARING FOR THE FIRST CLIENT MEETING

The first client meeting

Five stages of the consulting process

Kirk (1991) – Section 2

1. Establishing rapport
2. Identifying the research problem
3. Setting goals
4. Agreeing on division of responsibility
5. Summing up

Consulting role play (for Lecture 4)

The purpose of the role play is to practice how to plan and run the first client meeting.

Read Section 2 of Kirk (1991)

Prepare your allocated part of the consulting process

Client and meeting details

You have been approached by a leading debt charity, StepChange, who are looking to conduct research into the possible relationship between savings and debt.

You have the following information about the client prior to the "meeting":

- Dr Mary Smith, Research Leader, StepChange
- PhD in Social Policy
- Well-published in social policy literature on dealing with problem debt

Client and meeting details

You are also told the following about the data collection:

- Detailed longitudinal national survey of private households conducted using face-to-face interviews
- Includes data on:
  - level, distribution, nature and type of assets (including savings possessed) of households and individuals
  - debts of all types
  - Attitudes to financial planning, saving and financial advice

Project details

You have been asked to investigate the potential statistical link between cash savings and problem debt. In particular, you will be asked to estimate:

- Level of savings necessary to help households stay out of problem debt
- Number of households without this adequate level of saving
- How much savings levels need to be boosted, on average, in order to prevent or minimise problem debt.