General Feedback:
Graphical Models MSc Practical (HT2007)

Anjali Mazumder*
University of Oxford

May 3, 2007

1   General Comments

This practical was tricky for three main reasons. First, you were simply given a dataset and expected to apply a graphical models approach to analyse the data. This is a somewhat "backward" way to analysing data. In the real world, you would be given a real data problem or research question and have to determine the appropriate statistical method to apply. In this case, you needed to develop an appropriate research question that could be answered by the graphical model approach. Second, you needed to leave the world of regression behind. Many of you interpreted this analysis as a regression problem even though you performed a graphical models analysis. There was no response variable and no explanatory variables. There were simply four variables and the interest was in the dependence structure between these four variables. Third, model selection is a difficult process to explain clearly. There are many procedures and you have to determine the most logical steps that will lead you to the simplest and best fitting/suitable model. If everything is properly accounted for (i.e. type of variable, asymptotics versus Monte Carlo, etc), you should get to a reasonable model.

In general, the statistical analysis was performed well but the interpretation of the results was relatively poor and the overall writing and presentation of the report requires significant improvement. The marking scheme for this report was divided into five sections: Introduction (2), Data (2), Analysis (8), Interpretation and Conclusion (4) and Overall Presentation (4). I have summarised some of the key mistakes made on this practical according to these sections. I have also provided some (hopefully) useful tips on writing and structuring your practical report.

---

*Reviewed by Professor S. Lauritzen.

2 Introduction

The introduction section was of variable quality. I realise that no formal question was layed out for you; however, you should be able to develop and formulate an appropriate "question" to investigate based on the dataset given and analytical approach to be used. The problem needs to be put into context i.e. explain the problem and its importance. Many of you, in lieu of a proper introduction, provided a description of the variables. Background on possible relationships between factors (or variables) is a reasonable discussion; however, a description of the variables available in the dataset is not an introduction. There should not be any tables in the introduction.

3 Data

Most people presented the data well - explaining where it came from and provided some form of exploratory data analysis. Some of you presented contingency tables and/or mosaic plots but did not provide any textual explanation or discussion regarding the tables or figures. Often they were not labelled properly. Most of them looked like they were copied from another document and were missing a header for 'Age'.

4 Analysis

4.1 General

The analysis, in general, was performed well. Some common problems related to explaining and presenting the analysis were:

1. **Understanding and Explaining the Analytical Process:** Several of you followed the steps presented in Professor Lauritzen's example on Danish households. This was only a guide. These steps are not set in stone. Each dataset brings its own issues to the analysis process. Make sure you understand the steps involved and why a particular step is being taken (i.e. a particular search selection option, which marginals are being tested and against what). Don't assume that the steps you are taking are obvious. Provide justifications for your approach.

2. **Writing about Model Selection Steps:** There is a chronological process involved in model selection. Try and break up the methods and results into clear subsections. For example, it could like this: Fitting the Saturated Model; Model Selection Procedures; Examining Higher Order Interactions; Marginal Tests; Residual Analysis; Summary of Key Findings and the Final Model. Grouping analysis steps into a logical order will help you to explain the results in a clearer manner and help the reader follow your analytical process.

3. **R Commands and Output:**

- Statements such as "I used the 'u' procedure" does not explain what you did. Do not assume that your reader knows the software you are using - explain the procedure (i.e. what it is and why you are using it). Please do not include R commands in your text. R source code can be annotated and put in an appendix for those who want to know the technical details of your work.

- Please do not copy and paste R output. Summarize your results into nicely formatted and labelled tables. Try to summarize lots of similar/relevant output into one table. For instance, the results from the asymptotic, Monte Carlo and unrestricted stepwise selection procedures could be summarized into one table. Results should still be explained chronologically but should refer to a single table demonstrating the similarity or differences between these approaches and results. This provides a simple but clear presentation of the results from a few complicated steps.

4.2 Ordinality

Most of you stated that age was ordinal and some of you also argued that opinion may be considered ordinal or nominal. However, several of you failed to provide any arguments regarding ordinality and simply stated or inferred that it was obvious. Justify your reasoning. Do not assume that such things are obvious.

4.3 Monte Carlo Test

Many of you performed afirst stepwise selection using the Monte Carlo approach since you were "worried about the validity of the asymptotics". Good! However, all of you also concluded that the results were exactly the same as the asymptotic tests and hence you chose to continue with asymptotic tests. This was not true. Although the conclusions were the same in the sense that small p-values remained small and large remained large, but the p-values differed, sometimes at the most significant digit. Hence, it would be advisable to continue with the Monte Carlo approach as the results may have yielded different conclusions as further analysis was done.

4.4 Unrestricted Search Selection

Many performed an unrestricted search selection and concluded that the model chosen was the best model, ignoring the results of the ordinality tests. As a result, you would not have performed marginal tests on the age:race:opinion clique. The unrestricted search selection procedure would cycle through both decomposable and non-decomposable graphs. Some of

you provided a discussion around this and hence were able to justify the final model or argue against it.

4.5 Residual Analysis

In general, residual analyses were well performed. However, it is also useful to examine the relationship between the residuals and the variables of interest. This may help to explain sources of variation.

4.6 Additional Analyses Performed

Several of you chose to perform additional analyses such as fitting a multinomial (regression) model to the data and performing odds ratios. There seemed to be a tendancy to perform anything and everything possible you had learned in Further Statistical Methods. The purpose of this analysis was not regression or any other methods that seemed plausible to apply. Although many of these approaches were not incorrect, they were inappropriate for the purposes of this practical. Be careful not to anaylse the data to death - choose an appropriate method that answers your data question.

5 Interpretation

Most of you found interpreting a graphical model difficult. The tendancy was either to explain things in terms of regression (e.g. the effect of age on opinion was strongest) or contingency tables (e.g. the proportion of people who were for opinion were also young). These types of statements were incorrect based on the analysis performed. Several of you commented on age profiles and what that could tell you about levels of opinion. Interpreting conditional independence was tricky. A reasonable statement could have been that 'Age is conditionally independent of Race and Sex given Opinion'. This could be elaborated to: 'for a given level of opinion, there is no (obvious/direct) relationship between age and race or age and sex. Several of you tried to elaborate further and lost the meaning. Tip: Quit while you are ahead! Keep the interpretation simple.

6 Overall Presentation

There were some basic problems related to writing and overall presentation, in general.

6.1 Writing Issues

1. **Typos and Grammar:** The report should be free of any typos or grammatical errors. Several word processors, such as Microsoft Word,

provide grammar and spell check tools. Although these tools are helpful, do not depend on them. They will not correct words or phrases that are already correct. You may have a correctly spelled word but not the right word. For example, because of a typo, introduction may become induction! Check the spelling and grammar yourself.

2. **Repetitiveness and Verbosity:** Do not repeat the same information and/or sentences in the Introduction, Data, Analysis, Conclusion, and Summary. Each section has its own purpose and relevant content. Repetitiveness is sometimes useful to emphasize something, e.g. a property or result but it can also make the report verbose and/or convoluted. Do not write everything and anything you can think of. Be analytical and critical. Read over sentences and paragraphs and make sure that they add value to your explanation, not just space. Cutting out unnecessary words can make the report much clearer.

3. **Language and Content:** The content and language of your report should reflect what you have done. If you use background information from other resources to explain the problem, please provide references. Do not copy and claim it as your own. The language should also be consistent. Try not to change tenses or different words to describe the same thing. And again, please do not use R language. R is not a colloquial or written language; it is a statistical programming language.

6.2   Visual Presentation

1. **Layout and Font:** Choose suitable sections that best explain your work. There are standard sections such as Introduction, Methods (which includes Data and Analysis), Results, and Discussion. Use subsections to provide a clearer delineation of the work. These are especially useful for methods and results sections. Please do not change font size or font type either within or between sections. Text should all be in one size, usually 11pt or 12pt. Only use other font types such bold or italics to emphasize words, not whole sentences or paragraphs.

2. **Stating Models and Equations:** State and number models/equations explictly. Find a notation that is simple and clear. The reader should be able to know what the equation is expressing and not have to figure it out by referring to other figures or large amounts of text.

3. **Tables and Figures:** Tables and Figures are used to summarize lots of information in a clear manner. Tables and Figures are distinct. A Table provides a numerical summary of the data while a Figure presents a graphical summary of the data. Refer to them accordingly.

- Tables and Figures should be numbered and labelled sequentially as they are introduced.
- Tables should be numbered and labelled above while Figures are numbered and labelled below.
- Tables and Figures should have descriptive titles. Provide captions with complete information. Several of you wrote Figure x: interaction graph. What is it an interaction graph of? What is it showing? Be clear about what the Table is displaying or the Figure is showing.

A Table or Figure should never be found floating in your document. Text should refer directly to and explain what the Table is displaying or Figure is showing. Don't leave this to the reader to do. Pick and choose Tables and Figures that add value to your text (whatever it is that you are trying to explain) not just space. As they are used to summarize lots of information, they can be used as a guide to your written text, highlingting or comparing results from lots of analysis. Refer to them directly in the text.

4. **Supporting Material:** An Appendix provides additional details or information that would otherwise clutter the main body of the document. Some things that you may put in the Appendix are:

   - Tables and Figures that are very large (i.e. multiple page table);
   - R code or other software output; or
   - mathematical derivations that are too detailed.

   Keep the main body of the document clear and precise. Use the appendix to add useful informtion that supports your work.

7 Some useful tips for writing a scientific/practical report

If the reader cannot understand what you are doing, why you are doing it, how you are doing it, what the results are or what they mean, then you have failed to communicate your work effectively and not written a scientific report of any merit. It is very difficult to assess work that has not been clearly explained or presented. It is to your disadvantage as you may have done everything "right" or appropriate, but not been given credit because it was not clearly presented. You've done the work, now all you have to do is show it!

Please read over your work before submitting it and make sure it (1) is free of typos and grammatical mistakes (to the best of your abilities), (2) is layed out in a reasonable manner with appropriate sections, (3) includes meaningful tables and figures that are numbered and labelled, and (4) answers all of the questions above.

There are several examples and guidelines for writing a scientific report available on-line. These can be tailored to writing a scientific report for your practical. Much of the format and content is the same regardless of the discipline. Here are a few guidelines and tips. I hope you find them useful.

7.1 Layout: Key Sections of a Scientific Report

7.1.1 Title Page

The title page should include:

1. **Title:** Use an appropriate and descriptive title that clearly states what the report is about.

2. **Author:** Write your name.

3. **Date:** Today's date should be written.

Other relevant details may be included, e.g. subtitle: MSc Practical in Further Statistical Methods.

7.1.2 Abstract

This should be on a seperate page. It should only be one paragraph in length (approximately 200-300 words). The abstract/summary is precisely what is says - a summary. The best tip to writing this is to take one or two sentences that best describe each of your main sections - introduction, methods, results, discussion and conclusion. If the abstract is badly written, no one will want to read the report.

7.1.3 Introduction

Introduce the problem by providing relevant background, explaining the problem it adn it's importance. End with stating the purpose of the report/study. Try to answer the questions: what are you doing and why are you doing it.

7.1.4 Methods

This is usually made up of two main subsections: Data and Analysis.

1. **Data:** There are two main goals. The first is to state where the data came from and describe any relevant features of the dataset (e.g. types of variables). The second is to provide an exploratoey data analysis that helps to explain the structure of the dataset that is distinct from your main analysis.

2. **Methods:** Explain the steps involved in studying your problem. This should be written logically and in a chronological manner. There may be subsections that group together analysis steps. There should be enough detail so that a reader can reproduce your results by following your methods and given the same dataset. Try to answer the question: how are you solving the problem?

7.1.5 Results

Explain the actual findings. Use similar subheadings to those used to explain the methods. Use only Tables and Figures that are relevant to the problem being studied. Link all Figures and Tables directly to the text. Try to answer the question: what are the results?

7.1.6 Discussion

Discuss the main findings with regard to this importance and implications. Relate it back to the introduction and what it was you were trying to investigate. Try to answer the questions: what does these results mean? have I answered the questions from the introduction? what were the limitations? what are the implications?

7.1.7 Conclusion

Restate the purpose and summarize your findings. The reader should be left with a clear message.

7.1.8 Other Sections

Some of the above sections can be combined slightly. Sometimes the Discussion section is combined with either the Results or Conclusion. However it is set out, it should be clear and easy to follow. Depending on the report, it may be appropriate to include references and/or an appendix.

7.2 General Tips

I won't repeat myself and simply suggest you read over section 6.

7.3   Final Comment

Happy scientific report writing and good luck on the next one!

7.4   References

Comrie, A. C. *Scientific Report Writing.* http://geog.arizona.edu/ comrie/geog230/report.htm