8

Integrating project work in an applied statistics course

Christian Pipper

Department of Basic Sciences and Environment, LIFE, University of Copenhagen

Introduction

As the new course director of the PhD course in applied statistics ”Mixed Linear Models” at LIFE I decided on a major restructuring of teaching material and teaching activities to get a better alignment to the intended learning outcomes (ILOs).

One of my ideas was to actively use that this is a project course. That is, the course is intended – and expected by the students – to give them the competencies to justify, generate, and report statistical statements and conclusions so that they can answer relevant research questions on their own data. As such the students are for the most part highly intrinsically motivated (Biggs & Tang; 2007, p. 34) when focus is on their projects. I wanted to construct TLAs that promoted this motivation into deep learning in a *theory Y climate* as described in (Biggs & Tang; 2007, chapters 2-3).

To specifically channel student enthusiasm about their own data into in depth reflection on course material and ensure an optimal foundation for their project work in terms of gaining the above competencies, I constructed a self-reflection exercise for each of the 5 course days.

The self-reflection exercises consisted of questions helping the students to gain operational skills with their data, prompting them to reflect on important parts of the teaching material in relation to their own data, and guiding them through the process of writing a project synopsis that could function as a recipe of how they were to produce their project. As a built in part of the exercise the students were to present and test their ideas and discuss their progress and problems in pairs.

The student opinions on these exercises were evaluated by a questionnaire scoring the content, structure, and gain of the exercises.

Course description and specific TLAs

Course contents

The course covers basic theory and applications of mixed models, that is models having both fixed effects (as in ordinary linear regression or ANOVA models) and random effects. In addition these models may also have serial correlation structures, which are often useful in the context of repeated measurements/longitudinal data. The main focus is models for continuous response, but mixed models for categorical response are also briefly introduced. Operational skills for analyzing mixed models is based on the statistical programming language **R**.

Overall structure and some facts

The teaching sessions of the course take place on 5 whole days and are followed by a one month supervised project period. The project period is completed by a final whole day session where the students in turn present and discuss their projects.

In 2010 12 PhD students from various areas of bio-sciences at the University of Copenhagen attended the course. They were given 6 ECTS credit for completing the course.

Teaching material

The course material has been structured so that all concepts and methods have been illuminated and materialized by data-examples from research projects within agriculture/veterinary science/epidemiology. This goes for exercises as well as they all evolve around aspects of the statistical analysis of a concrete data-example. The examples have been chosen to cover research areas of typical students and thereby hopefully appear relevant to the students.

Specific TLAs

- **Summaries from previous day:** Student teacher discussion of concepts from the previous day. The product is a white-board overview of important concepts
- **Student presentation of exercises:** Students present a computer exercise and report their findings. Followed by questions from other students and student teacher discussion.
- **Student seminar sessions:** Based on a subject specific teacher written essay the students in pairs of two lecture over a subject. Each essay contains a number of open questions that form the basis of a classroom discussion.
- **Lectures:** Power-point presentations of general theory and concepts accompanied by student teacher discussion.
- **Computer exercises:** Analysis of data-examples by means of the statistical programming language **R** as well as interpretation and reporting of results. Supervised by teacher.
- **Self-reflection exercises:** The students relate course contents to their own data and discuss their findings with other students and teacher.
- **Project supervision:** Each student is entitled to two 1 hour supervision sessions with the teacher during the project period. Before each session the student sends an email with concrete questions to deal with at the session.
- **Project presentation:** 15 min oral student presentation of project followed by 5-10 min discussion with teacher and other students.
- **Project feedback:** Written evaluation/feedback from teacher on final project.

More information about the course

Additional material can be found on the course home-page:
http://www.matfys.kvl.dk/stat/phdcourses/mixed/

The structure and content of the self-reflection exercises

This section focuses on describing the actual self-reflection exercises in terms of their ingredients and progression. In this context extracts of the

actual exercises are included in italic to exemplify concepts. The exercises in their full length can be found in Appendix A.

Each of the four first self-reflection exercises begins with an introduction advocating the purpose and gain of the exercise:

"This exercise is intended to give you time to reflect on and discuss the structure of your own data. Such exercises will also be present on the other course days and are intended to guide you through the process of deciding on an analysis strategy for your own data."

This is followed by guide-lines on how much time to use for the questions and student discussion. For instance the guidelines for the first course day were:

"You should use approx 15 min. to consider the questions below and then taking turns use approx 10 min. each to discuss your answers and considerations with the course-participant next to you."

The actual content of the exercise is then presented in the form of 1 to 3 questions directly related to student data. For instance on day 1 the questions were:

1 *Make a short description of your data.*
2 *Concisely state the one most important research question related to your data.*
3 *Consider and justify ways of stating this research question in terms of outcome and explanatory variables.*

Closely linked to the progression of the course I tried to induce the following flow in the four first self reflection exercises

- **Day 1:** Make a concrete link between research question and data. Exercise duration 25 min.
- **Day 2:** Make a strategy for getting data into **R**. Describe data in terms of quantitative/qualitative /random variables as a first step towards a statistical model. Exercise duration 45 min.
- **Day 3:** Get data into **R** based on strategy from day 2. Attempt to formalize research question and data design into a first shot at a mixed linear model. Exercise duration 75 min.
- **Day 4:** Develop an analysis strategy for the chosen model with emphasis on the consequences of the strategy in terms of results and reporting. Exercise duration 75 min.

The self-reflection exercise on the 5th day of the course consists of tying together the previous 4 self-reflection exercises into a project-synopsis that

may form the basis of the project work to come. Two hours are allocated for working on this exercise. The project synopsis is motivated and described in the exercise as follows:

A synopsis or analysis plan is a much used tool in statistical project work. It's purpose is three-fold

- *It is usually the product of a combined effort of several parties constituting an agreement on the justification and purpose of the specific statistical procedures to be used in the project.*
- *It is a recipe to follow for the person to do the actual analyses and as such it provides overview and induces efficiency in the analysis process.*
- *Several parts of the synopsis can be reused in the process of reporting the results in terms of for instance a paper.*

Usual contents of a synopsis

- *A concise data description*
- *Research objectives/questions*
- *Description of chosen statistical methods.*

Evaluation of self-reflection exercises

Student evaluation

The students were asked to anonymously score the following questions from 1 to 5 with 1 meaning "I don't agree at all", 3 meaning "I agree", and 5 meaning that "I agree totally".

- **Content:** The self-reflection exercises provided a good opportunity for me to relate the course material to my own data
- **Structure:** The self-reflection exercises provided a good forum for discussion of my own data
- **Gain:** The self-reflection exercises helped me in the process of producing a project synopsis

The questions are part of the questionnaire used for overall evaluation of the course which is presented in Appendix B along with a graphical overview of the results. In the resulting student evaluations, question specific average scores were analyzed by a robust GEE approach (Liang & Zeger; 1989) taking into account within person correlation. All analyses were made in **R** version 2.10.1 (www.r-project.org).

The questionnaires were completed by 8 of 12 students. The average scores for the above 3 questions relating to the content, structure, and gain were: 4.1 (3.5-4.7), 4.3 (3.6-4.9), and 3.9 (3.1-4.6). There was no significant difference in average scores due to specific questions (p-value=0.14 (Robust Wald test)). The overall average score for the self-reflection exercises was 4.1 (3.4-4.7).

Additional student comments about the self-reflection exercises were

- Did the self-reflection exercises at home. Didn't think there was time during 13-16
- Very good course, great to work with own projects. Plenty of time for questions.

The evaluations indicate that the students are in general very satisfied with all aspects of the self-reflection exercises.

My evaluation

It is my observation that the students use the self-reflection exercises very differently. In this respect the students may be grouped into 3 different categories

- The time allocated for the exercise is mostly used to look at aspects of own data, largely disregarding the structure and content of the specific self-reflection exercise.
- The time allocated for the exercises is mostly used to look at aspects of own data, largely respecting the structure and content of the specific self-reflection exercise.
- The time allocated for the exercise is largely used to look at the computer exercises instead.

For the first two categories these exercises are very motivational and I sense that for many of these students they constitute the high-light of the day. I have also noted that the students obeying the structure and content of the exercises gain a lot from this. In addition to the student pair discussions I had very constructive short discussions with most of the students about their data during the time allocated for the self-reflection exercises. The students neglecting the structure and contents of the exercise prefer to discuss their data with me which leads to minor outbursts of frustration as I can not engage in the 10-15 minute discussion they are intended to have with another student.

A few students focus solely on digesting the course material of the day in terms of using all the time allocated for exercises on the computer exercises. I find this to be a constructive and natural choice as they were in fact struggling with the concepts in the course material and in seen this light I can respect that they have no additional capacity for the self-reflection exercises.

Concluding remarks

Based on my own observations and student evaluations I would have to conclude that the self-reflection exercises have been a very positive addition to the course in terms of project integration and achievement of course ILOs.

This of course does not mean that self-reflection exercises are a context-free recipe for success as they require student data of a high enough quality and complexity to make the process of completing the self-reflection exercises – and learning complicated statistical methods – worth-while for the student. However, it is my experience that the vast majority of PhD students attending these types of Applied Statistics Courses have data-examples and research questions fulfilling these requirements.

Also the success of the self-reflection exercises ideally requires the course material to cover a wide range of data-examples, so that a sound analysis of student data is exemplified in the course material. This approach was successfully adopted in the present course where students on average scored their agreement with the statement *"The examples were relevant to my field of research"* to 3.5 (2.6-4.4) (see Appendix B).

Another problematic aspect of the self-reflection exercises in the specific context of this course was that a substantial fraction of the students neglected the actual structure and content of the exercises and thus attacked their data in a rather unstructured manner. I find this a pity as I noted how the students who followed exercises more stringently benefited from it. During the course this problem was discussed with my pedagogical supervisor, who suggested that I emphasize the importance of structuring statistical project-work – which is really what the self-reflection exercises are all about – by including it as an ILO. I shall definitely adopt this suggestion next time I give the course. Yet another possible solution to this problem, that would be worthwhile trying out, is to formally pair the stu-

dents for these exercises, perhaps even on the basis of similarities in their projects.

A Appendix: The 5 self-reflection exercises

Day 1

This exercise is intended to give you time to reflect on and discuss the structure of your own data. Such exercises will also be present on the following course days and are intended to guide you through the process of deciding on an analysis strategy for your own data.

You should use approx 15 min. to consider the questions below and then taking turns use approx 10 min. each to discuss your answers and considerations with the course-participant next to you.

1 Make a short description of your data.
2 Concisely state the one most important research question related to your data.
3 Consider and justify ways of stating this research question in terms of outcome and explanatory variables.

Day 2

This exercise is intended to give you time to reflect on and discuss the structure of your own data. Such exercises will also be present on the other course days and are intended to guide you through the process of deciding on an analysis strategy for your own data.

You should use approx 30 min. to consider the questions below and then taking turns use approx 15 min. each to discuss your answers and considerations with the course-participant next to you.

1 Develop a strategy for getting your data into **R** and if possible try it out. Identify the problems you encounter and think of a solution.
2 Determine which of your explanatory variables are qualitative and which are quantitative. Would **R** agree with you?
3 Which of your variables would you consider random? Why?

Day 3

This exercise is intended to give you time to reflect on and discuss the structure of your own data. Such exercises will also be present on the other course days and are intended to guide you through the process of deciding on an analysis strategy for your own data.

You should use approx 60 min. to consider the assignments below and then taking turns use approx 15 min. each to discuss your answers and considerations with the course-participant next to you.

1 Work further on getting your data into **R** and customize them to your needs within **R**
2 Prepare a small presentation of your data and your research question
3 Consider the possibility of assessing your research question in terms of a linear mixed model. What model would you use? How would you shed light on your research question within this model? Do you foresee any problems with the model fit?

Day 4

This exercise is intended to give you time to reflect on and discuss the structure of your own data. Such exercises will also be present on the other course days and are intended to guide you through the process of deciding on an analysis strategy for your own data.

You should use approx 60 min. to consider the assignment below and then taking turns use approx 15 min. each to discuss your considerations with the course-participant next to you.

- Elaborate further on a statistical strategy for answering your research questions in terms of models, hypotheses, and quantification. You may be inspired by the analysis of the Hibiscus data at the final lecture today.

Day 5

In this exercise you are intended to tie together your work from the previous 4 self reflection exercises in writing a synopsis on your project as described below. You should use approximately 2 hours for making the synopsis.

The idea of the synopsis

A synopsis or analysis plan is a much used tool in statistical project work. It's purpose is three-fold:

- It is usually the product of a combined effort of several parties constituting an agreement on the justification and purpose of the specific statistical procedures to be used in the project.
- It is a recipe to follow for the person to do the actual analyses and as such it provides overview and induces efficiency in the analysis process.
- Several parts of the synopsis can be reused in the process of reporting the results in terms of for instance a paper.

Usual contents of a synopsis

- A concise data description
- Research objectives/questions
- Description of chosen statistical methods.

B Appendix: Questionnaire for student-evaluation

About the evaluation

This questionnaire is meant to provide feedback on the teaching activities and course material used during the course in terms of your perception of their quality and relevance. The feedback will be used to optimize both teaching material and teaching activities in terms of providing the best possible circumstances for guiding future students through the process of producing statistically founded quality statements and conclusions for their projects. As such your honest and critical feedback is much appreciated by me.

You should anonymously rate your agreement on each of the statements in the questionnaire from 1 to 5 with 1 meaning ”I don’t agree at all” and 5 meaning that ”I agree totally”. Also further comments on your part are most welcome and there is made room for these as well.

	Not at all		Agree		Totally
	1	2	3	4	5
Course material					
The methods and concepts in the course material (slides, essays, exercises) were justified by examples					
The focus on examples in the course material aided my understanding of methods and concepts					
The examples were relevant to my field of research					
The examples were interesting in general					
The general quality of course material was high					
Comments on course material:					
Teaching activities					
The quality of the lectures was high					
There was a good forum for discussion and questioning at the lectures					
The quality of the student seminar sessions was high					
There was a good forum for discussion and questioning at the student seminar sessions					
My own student seminar presentation helped me get a better understanding of the material I presented					
The quality of the computer exercises was high					
I got appropriate and constructive feedback at the computer exercises					
The self-reflection exercises provided a good opportunity for me to relate the course material to my own data					
The self-reflection exercises provided a good forum for discussion of my own data					
The self-reflection exercises helped me in the process of producing a project synopsis					

	Not at all		Agree		Totally
	1	2	3	4	5
There was a good balance between lectures, student seminar sessions, computer exercises, and self-reflection exercises (Please comment this statement if you don't agree)					
Comments on teaching activities:					
The course in general					
I feel confident that I have obtained the necessary competencies to justify, generate, and report statistical statements and conclusions in relation to my course project					
I feel confident that I have obtained the necessary competencies to justify, generate, and report statistical statements and conclusions in relation to other projects I might encounter					
General comments on the course:					

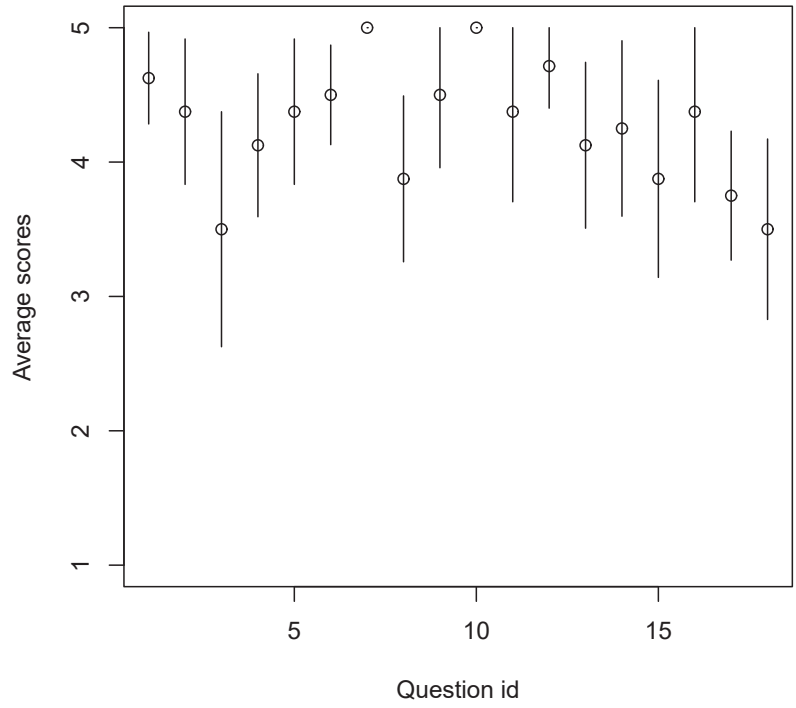

Fig. 8.1. Question specific average scores of agreement (1=not at all, 3=agree, 5=agree totally). Vertical lines correspond to 95% confidence intervals.

All contributions to this volume can be found at:

http://www.ind.ku.dk/publikationer/up_projekter/2009-2-1/

The bibliography can be found at:

http://www.ind.ku.dk/publikationer/up_projekter/kapitler/2009_vol2_nr1_bibliography.pdf/