Connecting Students with Statistical Research

Paul Northrop
p.northrop@ucl.ac.uk

Department of Statistical Science
University College London

RSS 3 April 2019

With thanks to Nick Grindle, Elinor Jones and colleagues who have generously contributed their time

Outline

- Research-based education; UCL Connected Curriculum
- ‘Meet your Professor’ assessment in Statistical Science
  - How it started
  - Implementation
  - Feedback from students (and staff)
  - Reflection
  - The future

Research-based eduation

Russell group paper on research-intensive learning

- “ensure that all the students have access to eminent academics early in their university careers”

Connected Curriculum at UCL

- “engaging all UCL students in research and enquiry”
- “dialogue between staff and students”
- “make strong connections across disciplines”

In Mathematical Sciences

- A pro: Statistical research is often inter-disciplinary
- A (perceived) con: undergraduates don’t know enough to understand current mathematical research

Activities in UCL Statistical Science

- Statistical projects (finalists / MSc)
- Selected Topics in Statistics module (finalists / MSc)
- Post-exams project (all)

Specific research-based assessments

1st year Introduction to Probability and Statistics

- write a lay summary of a research paper
- support from an interview with the author

2nd year Introduction to Applied Probability (Elinor Jones)

- answer questions about a research paper
- support from a fortnightly Journal Club

3rd year Stochastic Systems (Simon Harden)

In the beginning . . .

UCL Arena Centre for Research-based Education has sent information about a Meet Your Professor scheme for first year students, according to which groups of students are given the task of finding out about the research of a professor, presenting the result back to this person. This is motivated by the Provost's research-based teaching agenda. There was some discussion about whether we should offer something like this. XX proposed to do this without assessment, but it was felt that making it an assessment as part of STAT0002 would be better. This was decided. (Action: next STAT0002 lecturer, which probably will be PN.)

DTC minutes 4/6/2014

STAT0002: "exposing the students to real-world investigations."

Meet your Professor

An induction activity proposed by UCL Arena Centre for Research-based Education

- Groups of 6-8 students
- Investigate the research of an academic
- Meet to share findings
- Interview the academic
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 6-8 students
- Investigate the research of an academic
- Meet to share findings
- Interview the academic
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Investigate the research of an academic
- Meet to share findings
- Interview the academic
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to share findings
- Interview the academic
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the academic
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the author of the paper
- Group communicates the key themes of their research, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the author of the paper
- Group communicates the key themes of the paper, aimed at a non-specialist audience
- Powerpoint representation, leaflet, . . .
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the author of the paper
- Group communicates the key themes of the paper, aimed at a non-specialist audience
- A short (1 side of A4 paper) structured report
- The academic to give feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the author of the paper
- Group communicates the key themes of the paper, aimed at a non-specialist audience
- A short (1 side of A4 paper) structured report
- One person (me) marks the reports and gives feedback

Meet your Professor

An in-course assessment for the 1st year undergraduate module STAT0002 Introduction to Probability and Statistics

- Groups of 5-6 students (assigned by me)
- Read a group-specific research paper
- Meet to discuss the paper
- Interview the author of the paper
- Group communicates the key themes of the paper, aimed at a non-specialist audience
- A short (1 side of A4 paper) structured report
- One person (me) marks the reports and gives feedback

Time allowed: 6 weeks

The first assessment Oct-Dec 2014

I was hopeful but apprehensive about some things

- Support from colleagues (this wasn't a problem)
- Making the task clear, providing support to students
- Students working in groups
- Some papers too challenging ?
- Wide range of experience and degree programmes:

  Stats, Stats and Econ/Finance/Management/Language, Maths and Stats, Natural Sciences, Arts and Sciences, . . .

- Marking
  - judging the summaries fairly
  - (my) workload : each group has a different paper

Structure of the summary

1. **Background and aims**
   General context, why important, question(s) addressed, main application

2. **Data, statistical ideas and methodology**
   Describe the data, key ideas, assumptions, methods, model, inference. Some reading may be necessary

3. **Results and findings**
   Main results, relate the findings to the aims

4. **Concluding remarks**
   Emphasize main implication, limitations, future

Emphasis on clear communication

Structure of the summary

1. **Background and aims**
   General context, why important, question(s) addressed, main application

2. **Data, statistical ideas and methodology**
   Describe the data, key ideas, assumptions, methods, model, inference. Some reading may be necessary

3. **Results and findings**
   Main results, relate the findings to the aims

4. **Concluding remarks**
   Emphasize main implication, limitations, future

Emphasis on clear communication

"Don't try to make yourself look clever: it won't work!"

Marking criteria

1. Clear overview of the main contribution of the paper, linking results to the aims
2. Understandable by non-specialists ?
3. Describe, in general terms, the statistical methodology used and ideas related to those encountered in STAT0002
4. Avoid making errors in describing the statistical aspects, particularly errors that reveal a lack of understanding

A+ grade ($\geq$ 80%) only if at least satisfactory on all 4 criteria

The papers

- Colleagues offer their 'friendliest' papers
- Application areas: medicine, environment, economics and finance, genetics, energy markets, social networks, sport, food, transport, insurance, epidemics, ...

Students' perception of papers

- 'Hard': long, lots of maths, the main scientific purpose/application is difficult to appreciate initially
- 'Easy': short, not much maths (if any), the paper is driven by a specific application

BUT

- much of a long paper can be ignored (for their purposes)
- an account of methodology is required: students need to do some reading

Example paper 1

Ultrasound Obstet Gynecol 2014; 44: 503–514
Published online 12 October 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/uog.13437

Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis

N. NUNES*, G. AMBLER†, X. FOO*, J. NAFTALIN*, M. WIDSCHWENDTER‡ and D. JURKOVIC*

- no maths
- main scientific aims and the results are clear
- statistical methods not described in detail
- students need to do some work to understand and explain:
  - meta-analysis
  - sensitivity and specificity
  - confidence interval
  - hypothesis testing and *p*-values

Example paper 2

Bayesian Analysis (2010) 5, Number 3, pp. 429–450

Selection Sampling from Large Data Sets for Targeted Inference in Mixture Modeling

Ioanna Manolopoulou*, Cliburn Chan† and Mike West‡

- Rather more Maths
- Bayesian inference for a Gaussian mixture model
- Dirichlet process prior for the mixing probabilities
- Markov Chain Monte Carlo
- Application: distinguish different types of blood cells
- Challenge: explain model and inference in simple terms

Student support

- Lecture: explain the task and give advice
- Worked example:
  - a paper discussed during a lecture (Challenger Space Shuttle Disaster)
  - deconstruct the paper, assign key points to sections 1-4
  - example summary
- A detailed 3-page assignment sheet
  - "Lets be honest: you will probably not understand many of the more technical aspects of the paper . . . "
  - Impossible to cover everything, focus on the main idea(s)
- General questions via a Moodle discussion forum
- "Look at general feedback given in previous years"

Resources

- Goodwill of colleagues
  - they propose 1-3 papers each
  - each paper: 1-hour interview
- Marking and feedback: $\approx$ 4 days

Feedback to students

- General feedback
- Specific feedback on each group's report
  - Overall summary
  - Summary for each of the 4 sections
  - Highlighting specific issues
  - Giving praise

Feedback on a (relatively) weak summary

Volatility prediction based on scheduled macroeconomics announcements

Overall: You have made a good effort to try to explain some tricky mathematical details. One problem is that you have relied on many unexplained technical words and phrases, things that your target reader will not understand.

Section 2: You have included in your text an explanation of the meaning of the notation that appears in the model equation. This is good. However, I would have been more impressed with a clear and simple general description of the structure of the model, in particular description that related the error term in the model, the ϵ_t, to volatility and explained that in the model this volatility is affected by both the macroeconomic announcements and the recent volatility of the realized exchange rates. More generally, I admire the fact that you have tried to explain a lot of the detail contained in the paper, but I think that you have been over-ambitious: it would have been better to explain fewer things and therefore have the space to be able to explain them more carefully and more simply.

Feedback on a good summary

A Nonparametric Approach to the Removal of Documented Inhomogeneities in Climate Time Series

Overall: This is a very clear summary. You might perhaps have said something about how the models were fitted, and perhaps even that the GAM requires the choice of a smoothing parameter, but I am happy with the choices that you made. It is better that you used the limited space available to make the main ideas clear than to risk clarity by explaining more.

Section 2: What you have written here is very clear. If I had to find a criticism I would say that you could have explained better the role of the thousand simulated datasets : the idea is that these series have known jumps and that the task of the models is to recover the magnitudes of the jumps. However, this becomes absolutely clear in Section 3, so it is not a problem.

Feedback from staff

Informally

- Between-group variation
  - Some groups well-prepared, targeted questions . . .
  - Some groups cope well with organising the interview . . .
- Benefit from explaining research in simple terms
- Nice when when students take a genuine interest
- Students better informed when they choose projects

Feedback from students

Nick Grindle's survey: Dec 2016. An open, optional question

> *Say you are meeting up with a friend following the 'Meet the Researcher' activity.*
>
> *What would you tell them about it?*
>
> *Feel free to mention anything at all, for example what you learned, what you enjoyed or didn't enjoy, what was easy or hard, what was well organised or wasn't, or anything else.*

213 students (of 219) gave a response

Learning: 136 students

Of these students

- knowledge (facts and theories) 73%
- skills (applying knowledge, . . .) 27%
- attributes (mindset required for research) 9%

". . . I could feel [Dr Fs] enthusiasm . . . and his deep insight into statistical theories such as Bayesian framework. He explained his idea . . . in a understandable way and list a number of examples . . . '

". . . I developed new skills in communication, e.g. writing formal emails and asking concise questions to gain effective answering"

"This activity does help to improve my team work ability ."

". . . enabled me not only to gain a deeper understanding of statistics, but also the importance of coherence and communication ability in conveying the statistical information . . . ."

Peers vs Profs: 63 students

- 63 students report learning from the prof (dialogue)
- 51 (also) report learning from peers (collaboration)

"...it is really useful for us to meet the professor. ...she has solved the questions we prepared. And also, we have known the aim of this paper much more deeply."

"The professor was really good to us. He answered all the questions and explained every detail. I have almost understood his paper."

"...everyone was active and shared the ideas with each other. Overall, I learned a lot from this activity, not only the theoretical knowledge, but also how to well cooperate and communicate with others."

Difficulty / challenge: 76 students

"It was an enjoyable process to understand a research paper that is beyond my current statistical knowledge "

"I felt intimidated about the idea of interviewing the professor, but when you're working with group it makes it much easier "

"It was an interesting experience to talk to such an intelligent person, however, the difference of the knowledge between us and the prof was too big to understand too much "

"The hardest part of the task was deciding which parts to leave out of the summary as a large paper had to be condensed into one page"

"It was challenging to coordinate the group and everyone had different opinions, but that too [was] an opportunity to develop skills "

"[The papers difficulty] only made the feeling of success after understanding it with the help of my group even greater."

Reflection

Run on 5 occasions

2014: 216 students (36 groups)
2018: 288 students (48 groups)

Issues

- Non-engagement
- Within-group disagreements
- Writing, especially for students whose first language is not English
- Scientific writing: clarity, use of mathematical notation

Recurring feedback themes

- Lack of cohesion between sections. . . . group members writing independently
- Less can be better. Don't try to describe 6 different models. Explain the most important one well.
- What are "the data"?
- Give numbers a context. What does a reference value, e.g. 1, mean?
- Equations. IF you include an equation then you must explain the notation. Some of the best summaries avoid maths.
- The first sentence(s). Simple non-technical, set the scene

The first sentence

Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis

1. "The two main aims of this paper were to validate the International Ovarian Tumor Analysis (IOTA) 'simple rules' method and perform a meta-analysis comparing a recent research with six previous studies." (2018)
2. "One general aim of diagnostics is to be less invasive while retaining validity." (2016)

Of these I prefer 2., but perhaps

Ovarian cancer is . . . . It is important to diagnose this disease accurately because . . .

Positive aspects

- The keenest and most able students are stretched

  "Thank you, thank you so much! I am loving reading the paper!" (2018)

  [Paper: Bayesian hierarchical model for the prediction of football results]

- Students understand better what we do
- Teamwork, learning from each other
- Interviewing
- Writing something (o/w Stats students may not do this)

The future

- Alternative outputs: blog, vlog, press release, podcast, wiki, tweet(s), . . . ?
- Allow students to choose the papers and/or their group members ??
- Measures to ensure that all students contribute
- Increase the credit? Currently only 10% of final mark

Thank you for your attention