Applied Analytical Statistics

Week 1: Introduction

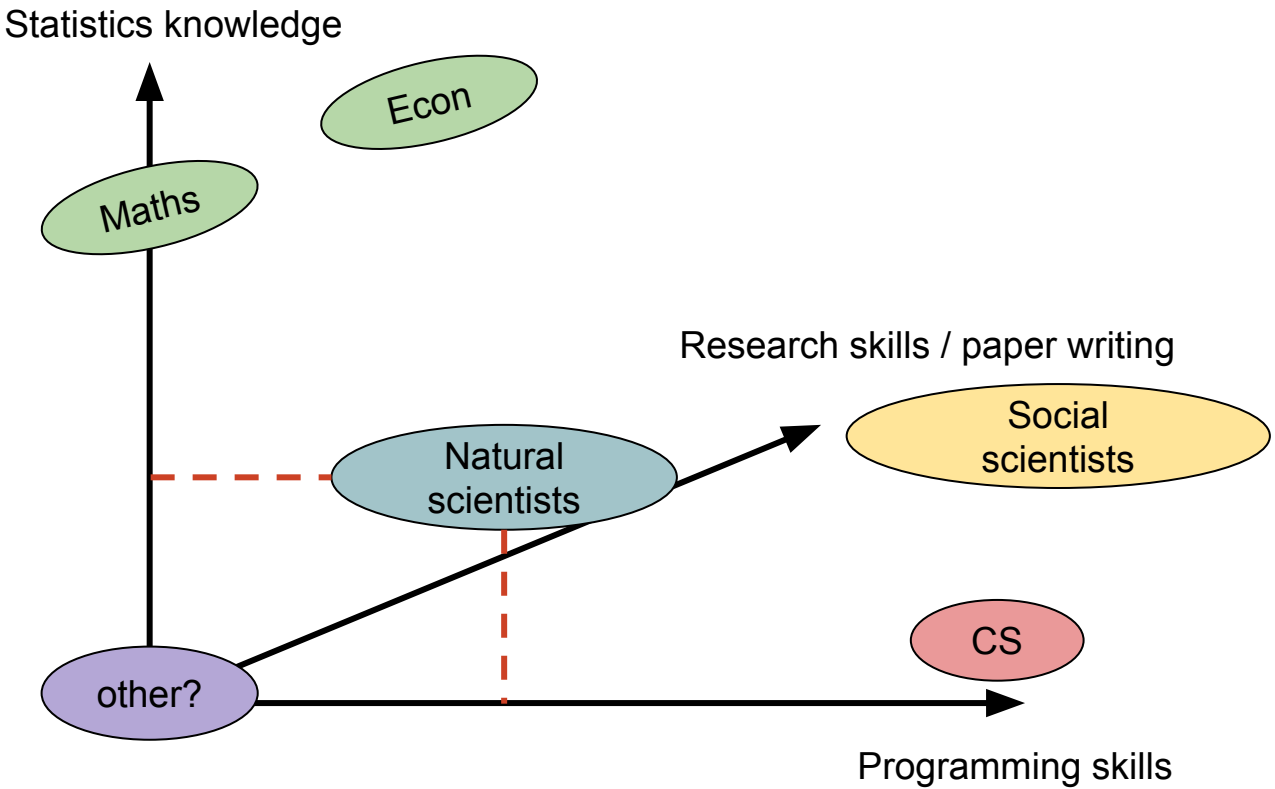
Harry Mayne

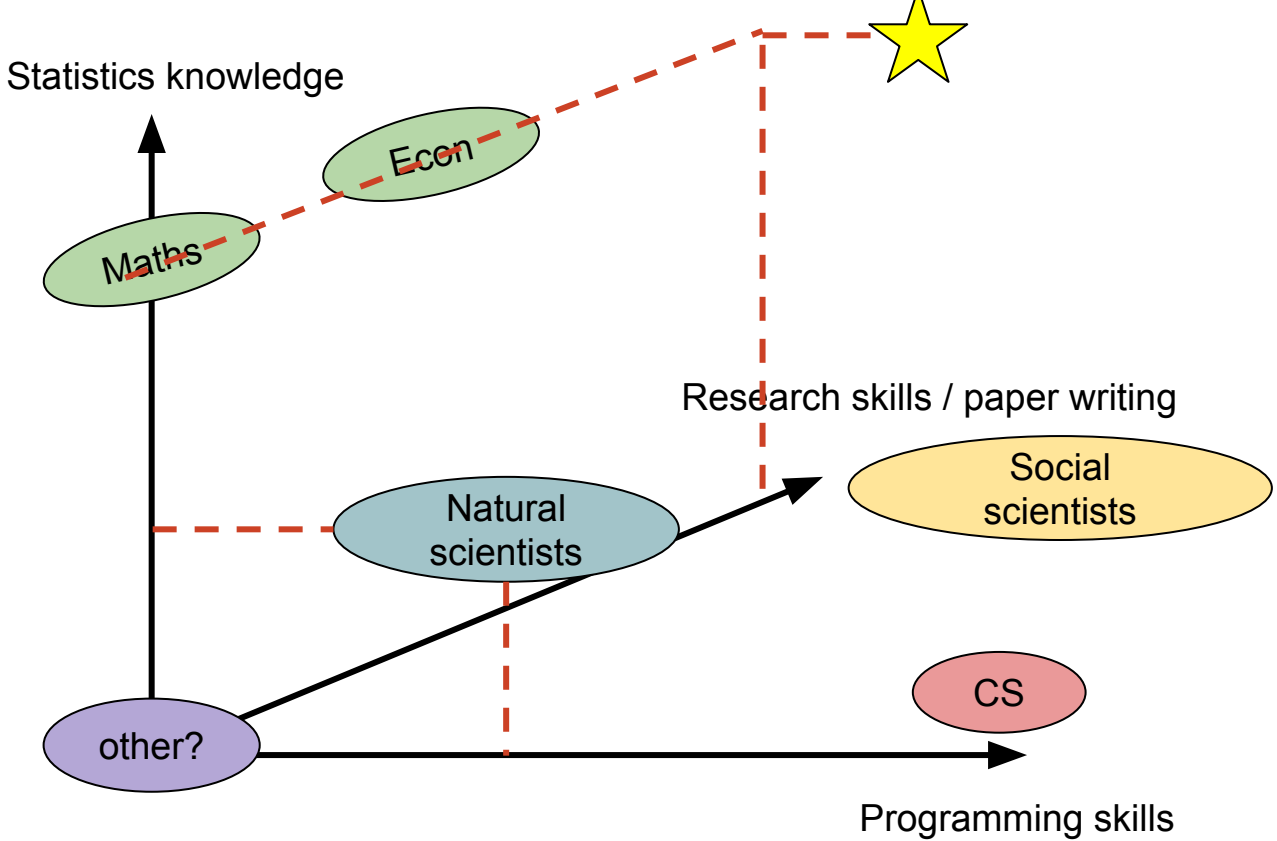
harry.mayne@oii.ox.ac.uk

- 2nd-year SDS DPhil working in Adam's lab
- I research mechanistic interpretability/ representation engineering for LLMs
- How to read and control the cognition process of LLMs
- Also work on benchmarks as part of the LingOly team

Course aims

1. Learn statistical methods to use **data** to test **hypotheses**
2. Learn how to apply methods in **Python**
3. Learn how to **write** a research paper around those methods


Statistics knowledge
Econ
Maths
Research skills / paper writing
Social scientists
Natural scientists
CS
other?
Programming skills

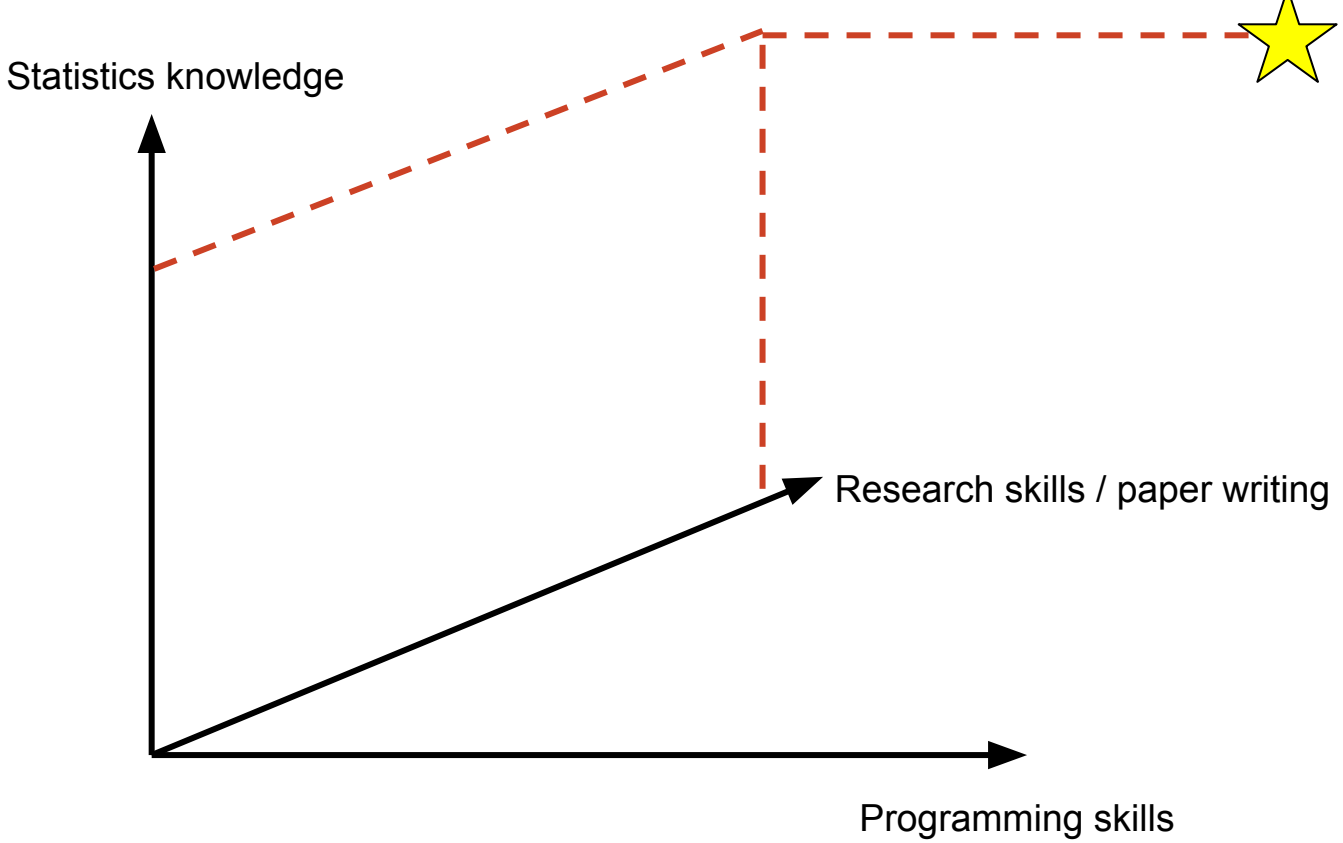

Statistics knowledge
Econ
Maths
Research skills / paper writing
Social scientists
Natural scientists
CS
other?
Programming skills

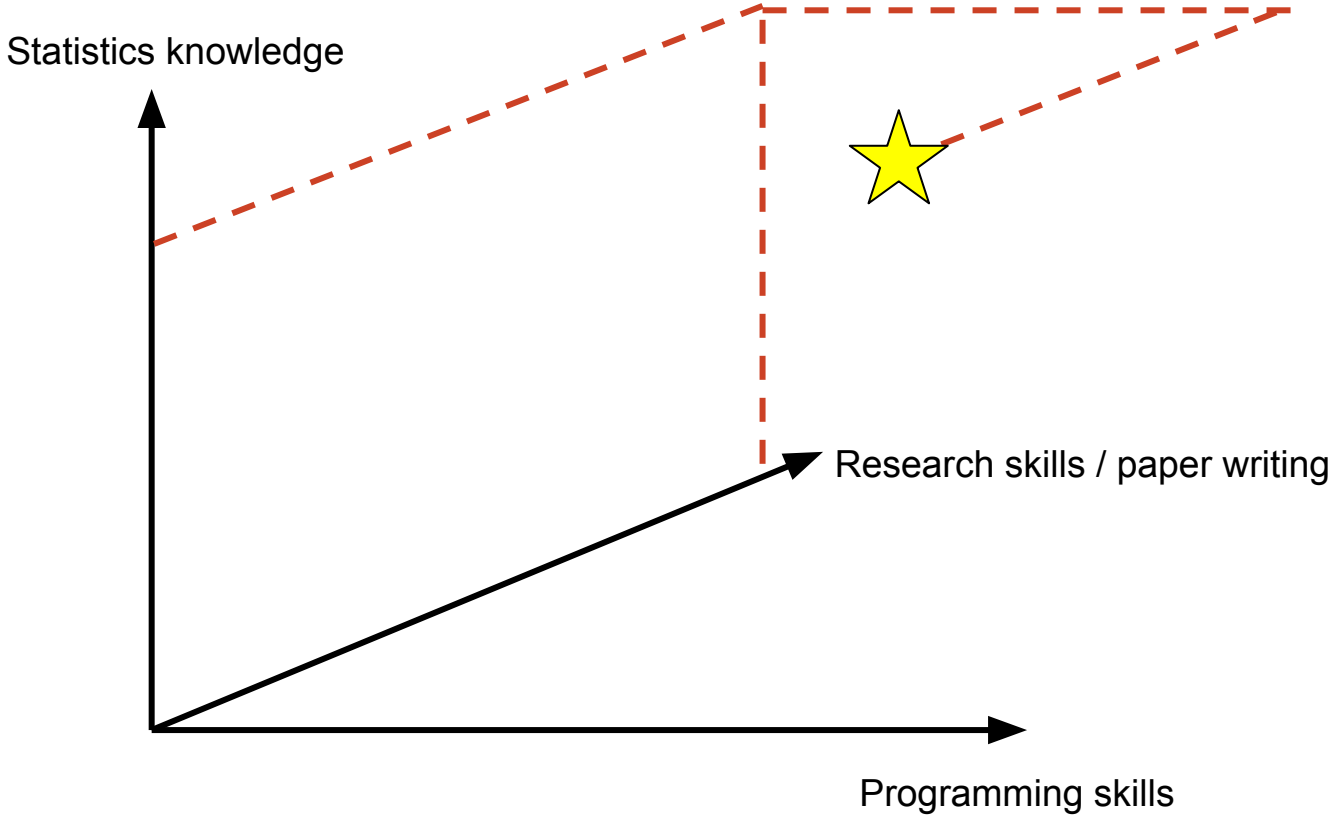
What's the point in the TA sessions?

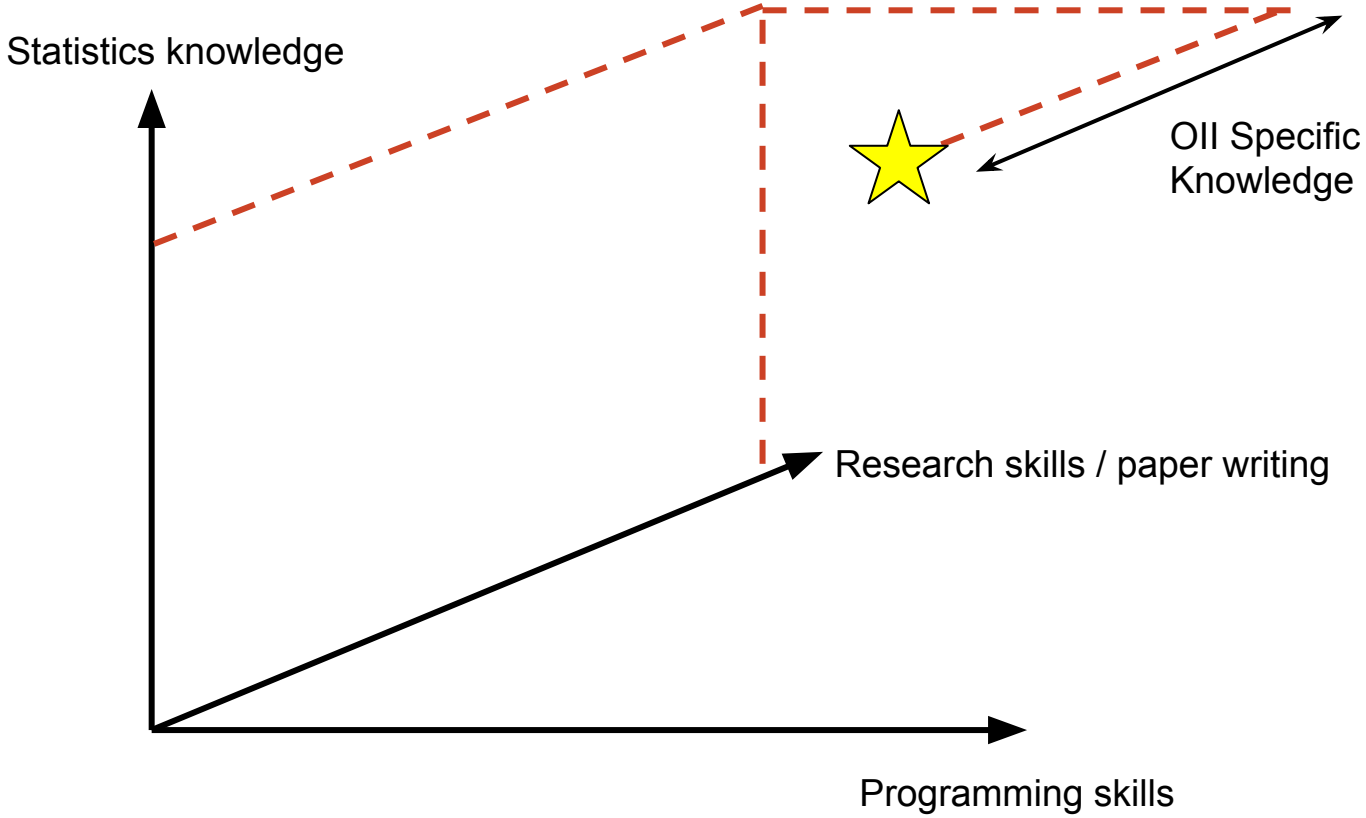
1. **Recap theory** and address questions
2. Learn how to apply the methods in Python
3. Learn how to write the paper (in the OII style)

- TA sessions are going to be theory deep dives
- Explaining the main concept from the lecture
- ~1/3rd answering your questions on theory
- Assignments done in own time

What if I’m here already?


Statistics knowledge
Research skills / paper writing
Programming skills


Statistics knowledge
Research skills / paper writing
Programming skills


Statistics knowledge
OII Specific Knowledge
Research skills / paper writing
Programming skills

How you will be assessed

- **One research paper**
- 5,000 words
- Done over Christmas

TIP: Start thinking about this as early as possible

Marked on the quality of your whole research paper NOT JUST the quality of the statistics in your paper

How you will be assessed

- Pretty flexible research project
- The most similar paper in style to your final thesis (in terms of format)
- We'll discuss more in future weeks. We'll have long discussions about exactly what you need to do, go through writing tips and look at model papers.

Python Notebooks

- One each week
- Released **Friday 11:30AM**
- Solutions released alongside work
- Due the following **Wednesday 11:59PM**
- Submit on Canvas

Problem Sheets

- Two problem sheets (probability and regression)
- Released with answers to self-mark
- Due **Week 5** and **Week 7 (Wednesday 11:59PM)**.

Summative Idea

- Short description of your plan for the summative (no more than ½ page)
- Research question, datasets and potentially methods
- Due **Week 4**: **Friday 8th November 11:59PM**
- Get Yes/No approval after

Week 1	
Week 2	• Week 1 notebook
Week 3	• Week 2 notebook
Week 4	• Week 3 notebook • **Summative plan (Friday 11:59 PM)**
Week 5	• Week 4 notebook • Problem sheet 1(Probability, CLT)
Week 6	• Week 5 notebook
Week 7	• Week 6 notebook • Problem sheet 2 (Regression)
Week 8	• Week 7 notebook
Week 9	• Week 8 notebook

All deadlines Wednesdays 11:59 PM

(except summative plan)

Python notebooks

- Notebooks, data and solutions available at www.harrymayne.com
- I recommend use Google Colab to complete
- Notebooks are **.ipynb** files. Please submit as .ipynb files on Canvas under the assignment section
- I provide you with the solutions straightaway
- Submission is compulsory and we record student submissions

Python notebooks Q&A doc

Live questions and answers page

Questions?

Today

Overview of key notation/definitions/maths…etc