Week 1 Tutorial Overview

Admin

- See the tutorial slides at **https://www.harrymayne.com/oxford** → I will be uploading everything through this page. The tutorial slides introduce the TA sessions, set expectations and contain the deadlines. TA sessions will be deepdives into theory + extension questions.
- In general, weekly assignments should be uploaded to **Canvas** by Wednesday 11:59PM the week after the TA session.

Overview view of course and summative

- Course topics
    1. Introduction to analytical statistics
    2. Sampling and statistical significance
    3. Linear regression I
    4. Linear regression II
    5. Logistic regression
    6. Multi-level modelling
    7. Topics
    8. Topics
- Summative discussion
    - A maximum 5,000 word project report
    - What are important skills to do a good job at this?
    - **Friday of Week 4, Michaelmas term** deadline for summative ideas.
    - This summative is the most similar to the overall thesis so a good chance to get that kind of writing experience. IMO the biggest gap MSc students have is being overconfident in their writing abilities (!)
    - How important are coding skills? Discussion.

Key points from this week's lecture

- Introduction to statistics
- Population statistics vs sample statistics
- *Descriptive* statistics (or EDA) and *inferential* statistics. Critically inferential statistics are drawing conclusions about a population based on a sample, i.e. ***inference*** from the sample to the population. Adam will drive the point hard that it your summative must be inferential statistics.
- How does inferential statistics work? Hypotheses and testing them.

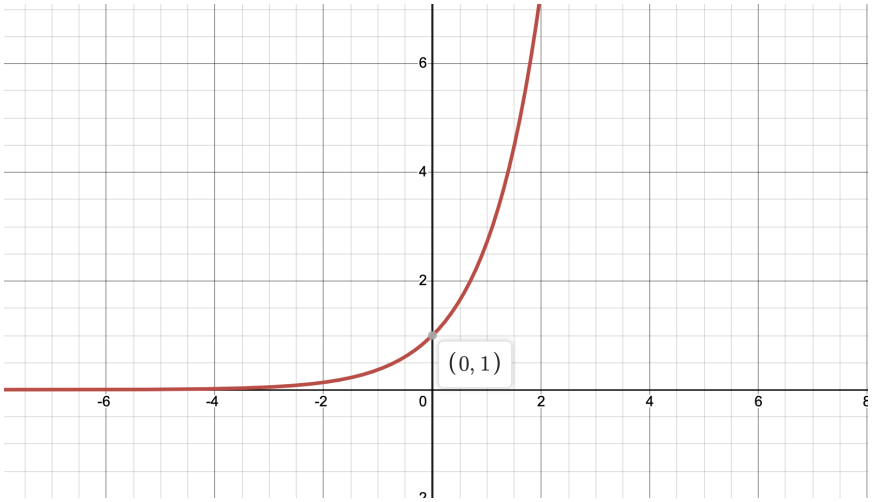
- What is machine learning? How is it different?

Course maths refresher

- This course assumes nothing other above GCSE maths as a starting point (high-school). Things can look complex but nothing is too bad.

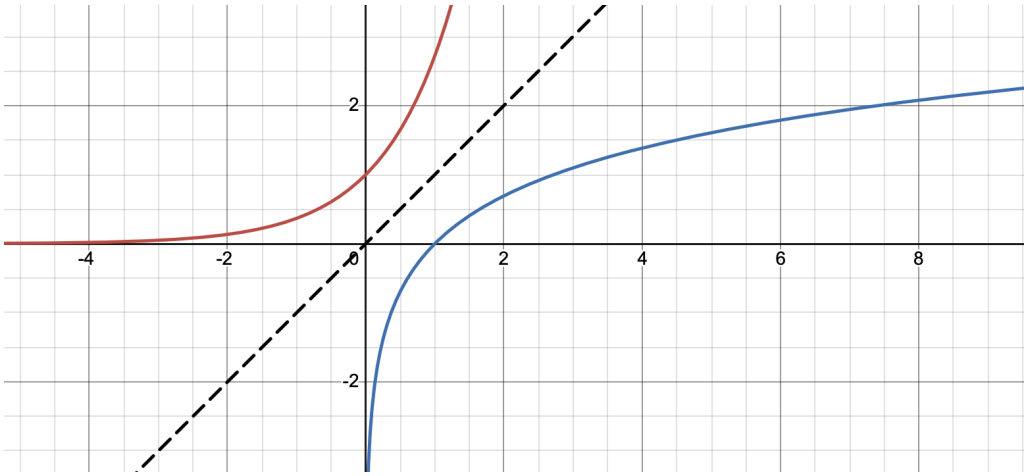
Things you should revise if you're not comfortable with them:

- Basic algebra
- Sums and products of the types: $\sum_{i=1}^{n} x_i$ and $\prod_{i=1}^{n} x_i$
- Functions e.g. $f(x) = a + bx^2$
- Simple differentiation e.g. $f'(x) = 2bx$
- Indices: x^2
- Basic probability (e.g. discrete random variables). $p(X = 6) = \frac{1}{6}$
- Exponentials, $f(x) = a^x$ for $a > 0, a \neq 1$ → $y = e^x$



$y = e^x$ is special because $y = \frac{dy}{dx}$...etc.

- Logarithms as the inverse of exponentials, especially the $\ln(x)$ function. Logarithm laws e.g. $\log(ab) = \log(a) + \log(b)$ and $\log(a^k) = k\log(a)$



$y = \ln(x)$ → This is the inverse of $y = e^x$. An inverse is the reflection in $y = x$

Resources basic maths recap

- The Fox textbook is **bad for this** → no nice maths appendix
- Wooldridge econometrics is a great textbook. ***MATH REFRESHER A, Basic Mathematical Tools*** has basically everything you need and is only ~pages. See a PDF download here.
- UK A-level textbooks → PDFs will be online.
- Loads of online resources for basic maths recaps

What the assignments look like

- See the first assignment at https://www.harrymayne.com/oxford

Intro to key terms (if we get time...)

An intro to some key terms in this course. A lot of this stuff will be covered in Adam's lecture next week + we'll cover in the next TA session (no need to revise this stuff this week)

[A lot of this is in ***MATH REFRESHER B*** and ***MATH REFRESHER C*** in Wooldridge if you want to get ahead]

- Random variables (discrete vs continuous random variables). A random variables is just an object that can take different values based on randomness. i.e. a probability distribution over the possible values it can take. Probabilities must sum to 1. $\sum_{i=1}^{n} p(X = x_i) = 1$
- We test to distinguish different variables with $Y, X, X_1, X_2, ..., X_k$. Note why this is confusing given we also use subscript to denote the individual: $X_{1,i}$
- Expected value and variance of random variables e.g. outcome of a dice: $E[X]$ and

$$Var(X) = E[(X - \mu)^2]$$

$$
\begin{aligned}
&= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

- Population level statistics (capital letters). X, **Population mean** $E[X] = \mu$, **Population variance** $Var(X) = \sigma^2$. Note can be μ_X and σ^2_X
- Sample level statistics (lower case letters) x_i is a specific instance of the random variable X. x_i where $i = 1, ..., n$. $x_i \bar{x}$ and s^2. Or we can represent these as our estimates of the population statistics $\hat{\mu}$ and $\hat{\sigma}^2$
- Parameters usually gives by lower case greek letters e.g. α, β, γ
- Estimated parameters given by hats e.g. $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$
- What is a distribution? $X \sim D(\mu, \sigma^2)$ e.g. the normal distribution $X \sim N(\mu, \sigma^2)$