

# **Applied Statistics Lab 1: Developing a Statistical Analysis**

Introduction to Statistical Thinking and Data Analysis  
MSc in Epidemiology / Health Data Analytics  
Autumn 2022

10 October 2022

# Today

- Introduction to Applied Statistics Lab (15 minutes)
- Getting to know each other (30 minutes)
- Steps in developing a statistical analysis (15 minutes)
- Group exercise: developing a statistical analysis plan (60 minutes)
  - Hypertension case study
  - Based on CDC Field Epidemiology Training Program (FETP): NCD Training Module on *Creating an Analysis Plan*  
[https://www.cdc.gov/globalhealth/healthprotection/fetp/training\\_modules/9/Creating-Analysis-Plan\\_PW\\_Final\\_09242013.pdf](https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/9/Creating-Analysis-Plan_PW_Final_09242013.pdf)

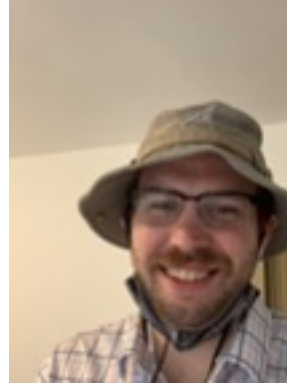
# Applied Stats Lab tutors

## Room 1

Katherine  
Davis



Thomas  
Wright



Haowei  
Wang



## Room 2

Bethan  
Cracknell  
Daniels



Lucas  
Cheng



Lanre  
Edun



# Applied statistical analysis

- Statistics is the science of:
  - collecting, summarizing, presenting, and interpreting data,
  - estimating the magnitude of associations,
  - testing hypotheses.
- Developing, conducting, and communicating statistics involves each of these components.
- Both written and oral communication are important.
  - Written: journal articles, reports, *mini-project assessments*, *MSc theses*.
  - Oral: Meetings, lectures and seminars, conferences, policy presentations.

# Communicating statistical analyses

- Communication of statistical findings almost *never* involves computer code or screen output from statistical programs.
- The exact steps and processes of statistical analysis (R commands run, assumptions checked, etc.) are usually not the focus reporting on an analysis.
- But it is important that the reporting of results communicates that statistical steps have been done properly.
  - This will also be the standard for assessment in this course.

# Applied Statistics Lab

- Practice *doing* statistics:
  - Data preparation and exploratory analysis.
  - Developing an analysis plan.
  - Conducting analysis and interpreting results.
  - Presenting findings.
- Randomly allocated into groups of 4-5 students.
- Three group projects analysing a dataset to address a research question.
  - Continuous outcomes and linear regression
  - Binary data and logistic regression, and
  - Longitudinal data and survival analysis.
- Culminating in 10-minute group presentation of findings (weeks 4, 7, 10).

10 Oct	Developing a statistical analysis plan
17 Oct	Project 1: Analysis plan and exploratory analysis
24 Oct	Project 1: Inferential analysis and interpretation of results
<b>31 Nov</b>	<b>Project 1: Communicating findings (presentation)</b>
7 Nov	Project 2: Analysis plan and exploratory analysis
14 Nov	Project 2: Inferential analysis and interpretation of results
<b>21 Nov</b>	<b>Project 2: Communicating findings (presentation)</b>
28 Nov	Project 3: Analysis plan and exploratory analysis
5 Dec	Project 3: Inferential analysis and interpretation of results
<b>12 Dec</b>	<b>Project 3: Communicating findings (presentation)</b>

# Applied Stats Lab: G64

**Tutors: Katherine Davis, Thomas Wright, Haowei Wang**

G64			
Group 1	Jingxian Huang Elena Venero Garcia Michaelis Vasiliadis Jian Chen Seth Howes	Group 4	Yuchen Xie Elin Rowlands Daniel Adams Lea Maria Khoueiry Vaishnavi Shridar
Group 2	Yiyang Shi Emmanuelle Kern Anu Bode-Favours Ka Ki Lui Siwei Wu	Group 5	Bing Chen Oliver Simmons Daniel Huntley Marina Berger Wenjia Zhang
Group 3	Shuhui Li Pin-Chun Wang Cameron Appel Kheerthiharan Saravanan Sreenidhi Venkatesh	Group 6	Mi Ma Harrison Goldspink David Ensor Megan Pete Wenqi Cho

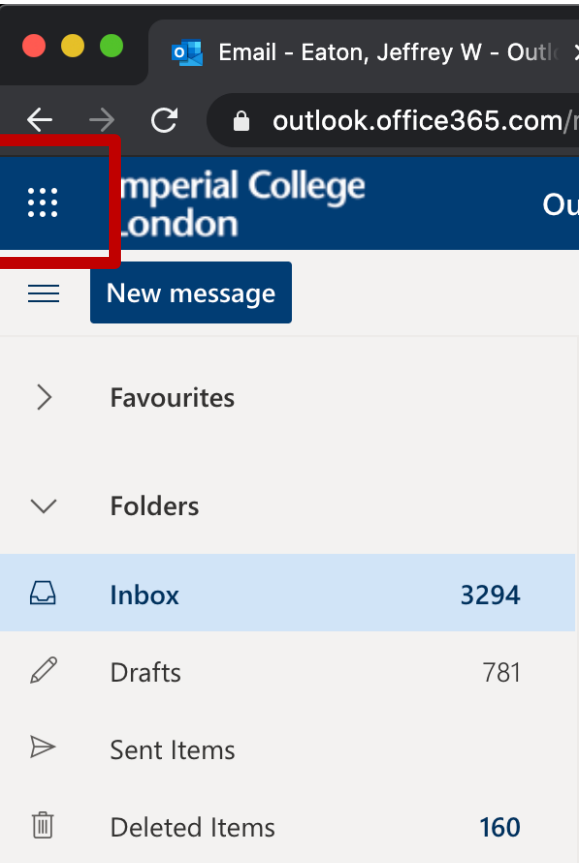
# Applied Stats Lab: G65

**Tutors: Bethan Cracknell-Daniels, Lucas Cheng, Lanre Edun**

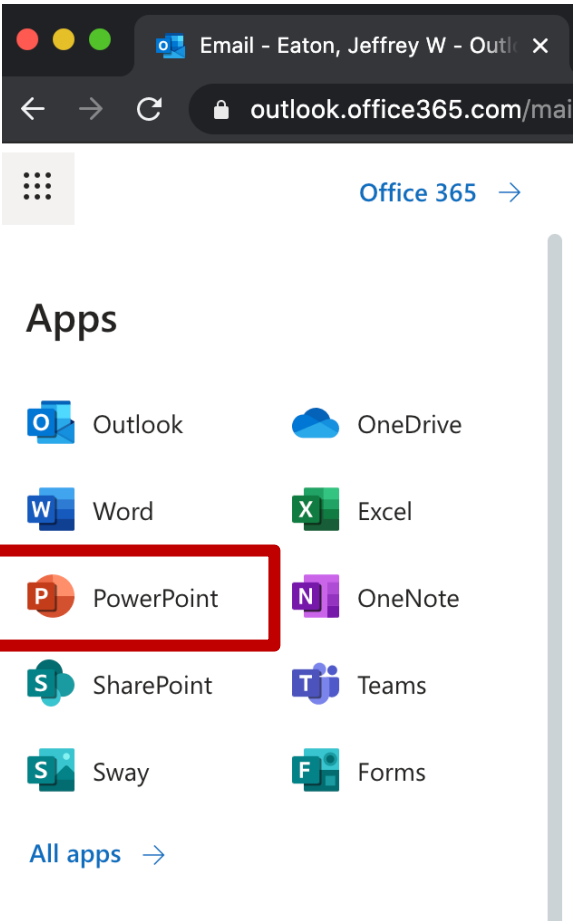
G65			
Group 7	Xihao Cao Thomas Allwright Emily Knight Mehak Gurnani Xheni Prebibaj	Group 10	Ciara Hamilton Alia Rafiq Huike Cheng Ria Sachdeva Helena Bicanic-Popovic
Group 8	Chiara Pligersdorffer Angela Aumonier Fiona Rice Nicole Cizauskas Yang Shen	Group 11	Sandra Gudziunaite Aditya Ramani Jaidip Gill Robert Campbell
Group 9	Mathias Brugel Abdul-Hakeem Khan Gabrielle Provost Omar Eweis Yuju Ahn	Group 12	Juliet Arukwe Gillian Sigle-Hall James Tait Samuel Quill



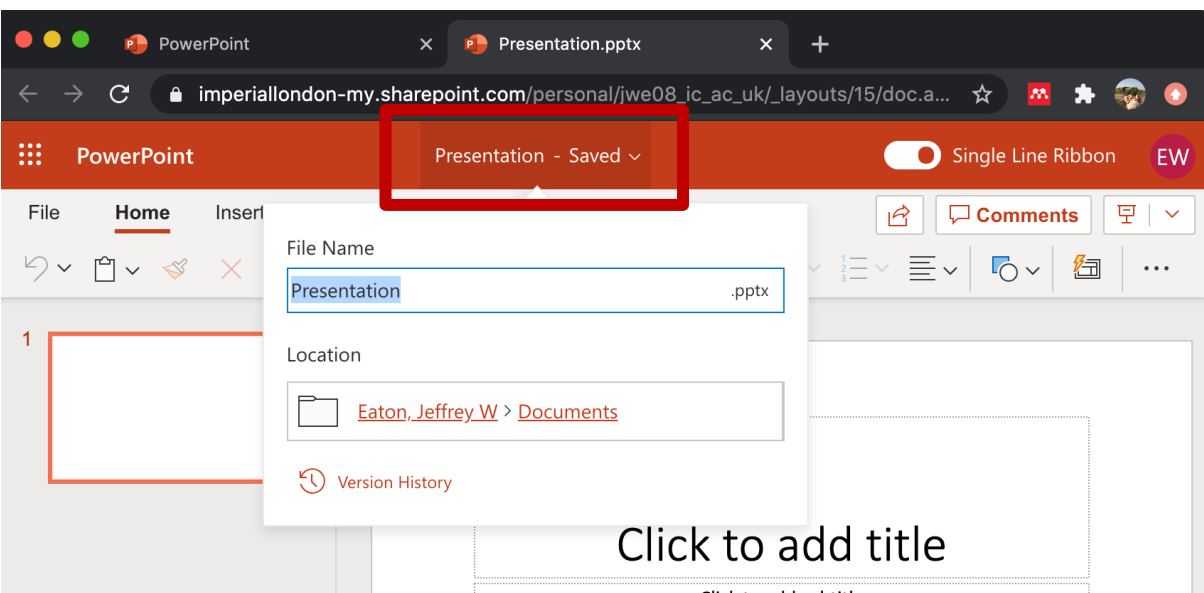
1)



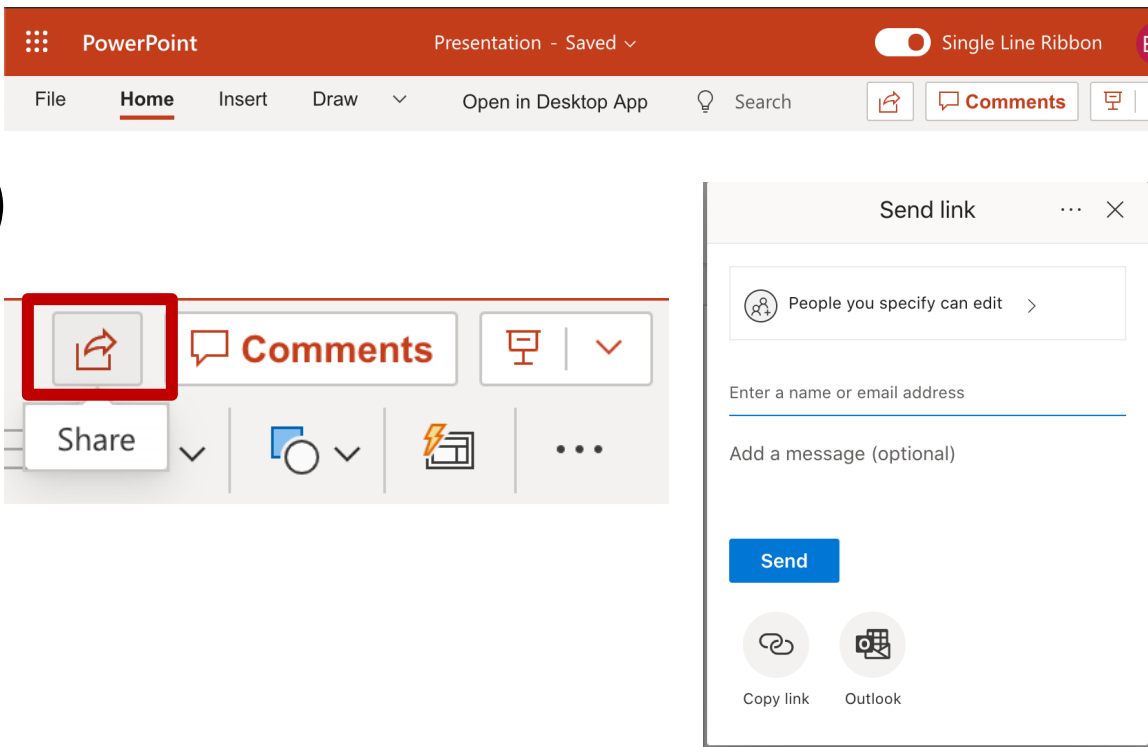
2)



3)



4)



# Group introductions

1. What are your goals for pursuing MSc in Epidemiology / Health Data Analytics?
  2. What is your previous experience with medical statistics?
  3. *Desert Island Statisticians*\*:
    - You're being castaway to work from home during a global pandemic.
    - Everyone gets a copy of Kirkwood & Sterne and a web archive of Stack Overflow.
    - On your way out of the office, allow you to select one more textbook and an essential office item.
    - What's your (1) textbook of choice and (2) essential office item?
- **Instructions:** Spend 15 minutes in your groups addressing 3 questions.
    - Record your answers to the (1) textbook of choice, and (2) essential office item in a table on 1 PowerPoint slide.
    - Return to your classroom and present your name, textbook, and essential office item.

\* If you are not familiar with *Desert Island Disks* on BBC Radio 4, I especially recommend this episode with Bernardine Evaristo: <https://www.bbc.co.uk/programmes/m000mrb1>

# **Part 2: Developing a Statistical Analysis Plan**

# Steps in a statistical analysis

1. Identify the research question and hypothesis
2. Define target population
3. Select the dataset and define inclusion/exclusion criteria.
4. Identify variables to be used in main analysis (outcome, main exposure, stratifying variables)
5. Define key shell tables

# Steps in a statistical analysis

1. Conduct exploratory data analysis: understand your dataset.
2. Conduct inferential statistics: answer your research question
3. Check assumptions and sensitivity analysis: are the findings robust
4. Interpret the results: state the conclusion to the research question
  - Should be as close to a 'yes'/'no' answer to your hypothesis as possible.
5. Contextualize the findings: how do the findings compare to existing evidence.
6. Implications for policy and practice

# Two types of statistics

## **Descriptive statistics: *Exploratory analysis***

- Describe and summarise key features of the dataset
- *Understanding the characteristics of a particular sample.*

## **Inferential statistics: *Hypothesis driven***

- Use data from a *sample* to draw conclusions about the *population*.
- Generalizable scientific conclusions that add to general body of knowledge.

# Reading a paper

- Scientific papers in medicine, epidemiology, and public health follow a formulaic format that map to the steps of a statistical analysis.
- Introduction:
  - What is the research question?
  - Why is it important?
  - What is the target population?
- Methods
  - Define the dataset
  - Inclusion and exclusion criteria
  - Statistical methods and assumptions

# Reading a paper

- Results
  - ‘Table 1’: descriptive statistics
  - Statistical inference results: answers to the research questions
- Conclusions
  - Answer primary research question
  - Contextualize findings, compare with other literature
  - Implications for policy, practice, future research



# Exercise: Hypertension Case Study

- Read the Hypertension Case Study in the handout.
- In your group address and record the following components of the analysis plan:
  - Research question(s) and/or hypotheses.
  - Dataset(s) to be used.
  - Inclusion/exclusion criteria.
  - Variables to be used in the main analysis (List 3-4 outcome and exposure variables).
  - Statistical methods and software to be used.

# Hypertension Case Study

- Nutrition transition → ‘double disease burden’
- Non-communicable diseases (NCDs) responsible for 60% of deaths worldwide.
  - 80% of NCD deaths occur in low- and middle-income countries (LMIC)
  - 30% occur below age 60 in LMIC vs. <20% in HIC
- **Cardiovascular diseases (CVD)** are largest cause of death, 30% of all deaths; 50% of NCDs.
  - Behavioural risk factors: tobacco use, physical inactivity, unhealthy diet, harmful use of alcohol
- Effective prevention exists.
  - E.g. reducing sodium consumption

# Hypertension Case Study

- Country X—rapidly modernizing nation of 10 million people, growing middle-class.
- Increasing overall life expectancy, but a surprisingly high rate of CVD deaths, especially stroke.
  - Increased use of renal dialysis reported in two southern provinces.
- Traditional diet is generally rich in fresh vegetables, whole grains, and healthy oils.
- Young and middle-aged have increased consumption of meals out of the home.
  - Higher in fat, salt, and processed foods.
- Tremendous expansion of restaurants that serve meals not typical of the traditional diet in the country.

# Hypertension Case Study: task

- Health recommendations: particular concern about lack of basic information on **hypertension**, potential underlying risk factor.
- You are the leader of the Ministry of Health's Chronic Disease Surveillance Unit.
  - MOH request to analyse national health survey data on **hypertension**, collected every two years for past decade.
  - Provide the report to national and provincial decision makers.
- Objectives:
  1. Better understand magnitude and burden of hypertension.
  2. Key determinants and underlying factors of this public health burden.

Criteria	Marks (Total 40)
<u>Introduction</u>	
Research question and population of interest identified.	3
Clear motivation for why research is important and rationale for study.	3
Specific objectives and testable hypotheses articulated.	3
<u>Methods</u>	
Dataset, inclusion criteria, and data processing clearly described and suitable to address question.	3
Clear description of analysis plan to adjudicate hypotheses, including clear identification of primary outcome and exposure variables.	3
Clear description of choice of statistical model to address the hypothesis and why chosen.	3

Criteria	Marks
<u>Results</u>	
Appropriate presentation of descriptive statistics; communicated understanding of key relevant features of the study population.	3
Presentation of results inferential statistical analyses proves clear and concise answers to study hypotheses	3
Clear and focused figures and tables used effectively to aid communication of results narrative.	3
Accurate and elegant presentation and interpretation of statistical outcomes and sensitivity analyses.	3
<u>Conclusions</u>	
Clear statement and interpretation of results linked to overall study aims and hypotheses	3
Findings are contextualized and implications for public health practice and research articulated	3
Presentation in allotted time and well paced	2
Overall quality and clarity of presentation slides	2