APPLIED STATISTICS BOOTCAMP

MICHAELMAS TERM 2023

CONVENORS: JAVIER FUENZALIDA and FLAVIA GALVANI

3 October 2023

"In an age in which data plays an ever-more prominent role in society, the ability to spot ways in which numbers can be misused, and to be able to deconstruct claims based on statistics, should be a standard civic skill."

David Spiegelhalter, The Art of Statistics

Aims and Objectives

The primary aim of this introductory course is to enable students to acquire and apply basic statistical concepts, developing their ability to understand and communicate findings based on data and statistical analysis.

After completing the course, students should:

- Develop a better understanding of key statistical concepts, including describing data and testing relationships.
- Become better equipped to understand statistically-based results and facts and the way in which statistical analysis can help inform effective policymaking.

The course will introduce students to the basic language and the fundamental ideas of statistics, with an emphasis on the use and interpretation of statistics in policy questions. The course will look at:

- How to describe data using averages, variance, standard deviations, and how to interpret basic graphs.
- The meaning and use of p-values, effect sizes, and confidence intervals.

- How to interpret the results of statistical tests and regressions.

Teaching Team

There are 6 members in the core teaching team:

- Javier Fuenzalida, course co-convenor and lecturer
- Flavia Galvani, course co-convenor
- Aung Hein, seminar leader
- Ashley Wright, seminar leader
- Filipe Recch, seminar leader
- Kritika Ray, seminar leader

You are welcome to contact Javier Fuenzalida (javier.fuenzalida@bsg.ox.ac.uk) or Flavia Galvani (flavia.galvani@bsg.ox.ac.uk) if you have any questions or concerns related to the course.

Organisation of Teaching

The course has four 3-hour sessions: the first two in Week 5, the third in Week 7 and the fourth in Week 9. Each session consists of a two-hour lecture, followed by a one-hour practice session.

Preparation tasks

Before every session, students are expected to read and/or watch materials that have been specifically curated to introduce key statistical concepts that will be further explored in the lectures and practice sessions.

Watching and reading the required materials for each session should take no more than one hour. Please ensure that you use just the links provided and review the

materials in the order they are shown in this syllabus. The lecturer will assume that students have read and watched all the required materials for each class, so please do your best to ensure that this is the case.

Lectures

These two-hour sessions will build on the content of the pre-reading/watching materials to further develop an understanding of key statistical concepts and techniques and their application in public policy.

Practice sessions

In these one-hour sessions, students will apply statistical concepts to real world policy problems. The practice sessions will have Excel and non-Excel learning streams, but both streams will work on the same exercise. You are free to choose which stream to join depending on your learning preferences, but you must be proficient in the use of Excel to join the Excel learning stream[1].

Assessment

This course will not be formally assessed.

Sessions overview

Week 5 – Part 1: Describing data
Lecture: 7 November 9 AM – 11 AM
Practice session: 7 November 2 PM – 3 PM

In this first session, we will discuss what is (and what is not) a variable and look at how to describe and summarise data.

Specific contents	Pre-readings/pre-watchings

---

[1] Please note that the University offers an online Excel course for students

Motivation of this course	• [Citizens need to know numbers](#)
Variables	• [What is a variable and types of variables](#)
Describing the data with a single value: measures of central tendency	• [The notion of central tendency and the three most common measures of central tendency](#) • Calculating the [mean](#), [median,](#) and [mode](#)
Describing the spread of a data distribution: measures of dispersion	• [The notion of data dispersion and common measures of dispersion](#) • Calculating the [range](#), and the [variance and standard deviation](#)
Frequencies, distributions, and density curves	• [What is a frequency distribution and examples](#) (please click and <u>skim</u> the visualizations of frequency distributions at the end) • [The idea behind density curves](#)
Normal distribution and z-scores	• [Introduction to the normal distribution](#) (please <u>skim</u> the section "How To Check Data") • [Calculation and interpretation of z-scores](#)

Week 5 – Part 2: Using samples to estimate characteristics of a population
Lecture: 9 November 9 AM – 11 AM
Practice session: 9 November 3:30 PM – 4:30 PM

In this second session, we will introduce the notion of sampling, that is, extracting a sample from a population. In addition, we will discuss how we can use the characteristics of the sample (statistic) to estimate the characteristics of the population (parameter) and the uncertainty involved in these inferences.

Specific contents	**Pre-readings/pre-watchings**
Basics of population and sampling	• [Parameters, statistics, sample size and sampling distribution](#)

	The Central Limit Theorem and its importance
Statistical uncertainty when estimating parameters	Introduction to inferences, and understanding uncertainty Estimation and interpretation of confidence intervals (please skip the section "Properties of 'Good' Estimators")

Week 7: Using samples to assess statements of population characteristics
Lecture: 23 November 9 AM – 11 AM
Practice session: 23 November 12:45 PM – 1:45 PM

In this third session, we will discuss how to use information from a sample to judge claims about the characteristics of a population. We will examine how to interpret the results of hypothesis tests to make comparisons using sample statistics.

Specific contents	**Pre-readings/pre-watchings**
Statistical uncertainty and hypothesis testing	Introduction to hypothesis testing, and key steps for hypothesis testing
Differences in means	Hypothesis test for difference of means
Differences in proportions	Hypotheses test for difference in proportions (please skip the "Theorem" and just focus on "Example 9-5")

Week 9: Modelling relationships between variables
Lecture: 4 December 1:30 PM – 3:30 PM
Practice session: 4 December 4 PM – 5 PM

In this last session, we will start by reviewing the relationships between variables and the notion of correlation. We will then take this idea further to introduce the use of regression models to examine associations between two or multiple variables.

Specific contents	**Pre-readings/pre-watchings**

Relationship between two variables	• [Bivariate linear relationships, strength, and direction](#) • [Correlation coefficient intuition](#)
Introduction to least-squares regression	• [Introduction to lest-squares regression](#) • [Using least-squares regression output](#) • [Reading a regression table](#)