STAT 1110-910 – Summer 2023 (Updated 05/15/23)

INFORMATION AND SYLLABUS

Lecturer: Katherine Brumberg
Email: kbrum@wharton.upenn.edu
Office location: Room 442D, Wharton Academic Research Building (WARB)
Zoom office hours: M-Th 11:15-11:50AM, F 10:15-10:45AM, or by appointment

TA: Sutton Grossinger
Email: suttgro@sas.upenn.edu
Office hours: TBD

The nature of the course The course will be held fully online via Zoom and Canvas from May 22 through June 28th Mondays through Fridays from 10:15AM to 11:50AM Eastern. There will be no class on Memorial Day, May 29th, and Juneteenth, June 19th.

Course description The content of this course falls into two broad categories, namely probability theory and statistics. A detailed list of the topics covered within these two categories is given in the Syllabus on the following pages. Note that this course is very introductory. If you have previous experience with probability or statistics, this is likely not the course for you. Our goal in this course will be to understand the foundations of statistics, namely probabilities and how they are used to estimate parameters and draw conclusions. We will not cover every possible distribution or hypothesis test, or even touch on many other areas of statistics. However, understanding the material in this course should help you read scientific papers or design experiments or studies with valid statistical analyses.

Structure of the course As this course is only 6 weeks long and meets daily, it is very important not to fall behind. The course is thus structured in a way that should help you succeed in staying focused, up to date, and learning. We will re-evaluate the structure of the course half-way through, and it is possible we will determine a new structure for the final portion of the course to better support your learning.

Monday through Thursday classes

10:15-10:40AM Lecture part I (recorded)

10:40-10:45AM 5 minute break

10:45-11:10AM Lecture part II (recorded)

11:10-11:15AM 5 minute break

11:15-11:50AM Discussion/office hours (not recorded)

Homeworks for Monday through Thursdays These are designed to apply the concepts learned in class. Homeworks will not be graded, but will be the only way for you to practice

the material and stay current with the fast-paced course. Homeworks will be drawn from the problems at the end of the textbook chapters (see the textbook section below). Solutions are available in the "Back matter" of the textbook. Homeworks will also be discussed the following day in the office hour portion of the course.

Friday classes

10:15-10:45AM Discussion/office hours

10:45-11:50AM Open-note quiz on Canvas (please email me if you cannot take the quiz at this time)

Friday homework Optionally, work through another version of the quiz and submit to increase your grade and understanding.

Grading There will be no midterm or final exam. Instead, each of the 5 weekly quizzes will compose 19% of your grade, for a total of 95%. The final 5% will be for active participation in class, through emails, answering questions during lectures, asking questions during lectures or office hours, asking or answering questions on piazza, etc. The grade for the weekly quiz will be the better of: 13% for the original quiz and 6% for the take-home quiz OR 19% for the original quiz. Doing the take-home quiz can only increase your grade.

Academic honesty Homeworks are not graded and can be done using any tools you'd like or together with other students. However, quizzes, both the original canvas quizzes and the take-home quizzes, are to be completed individually. You can use any notes or references, but the work you turn in should be your own. To this end, the questions on the quizzes will be different for each student. You may use calculators and any statistical programs you like on the quizzes, but you cannot use artificial intelligence programs such as ChatGPT. The second quiz attempt, as well as the weekly and open-note nature of the quizzes, is an effort to make the assessments less pressured and more learning oriented.

Disabilities and extenuating circumstances If you require any accommodations for any part of the course, please do not hesitate to reach out.

Textbook The textbook for the course is "Introductory Statistics for Data Analysis" by Ewens and Brumberg, 2023, ISBN 13: 978-3031281884. The hard copy of the textbook will not be published until later in the course, but the online edition is now available. You can access it at `https://link.springer.com/book/10.1007/978-3-031-28189-1`. To access it for free, please click the "Access via your institution" button on the right of the screen, search for "University of Pennsylvania", and log in using your PennID. Once you have done that, you can download the pdf of the entire book on the right hand side of the page or you can scroll down and click on any chapter title to download the PDF of the chapter. Let me know if you experience any issues with this.

If you want to buy a different textbook covering material similar to that discussed in class, Prof. Warren Ewens recommends Downing and Clark, "E-Z Statistics", Barron, 2009, ISBN 13: 978-0-7641-3978-9. This book should be available via the Penn bookstore. However this

book is **not** required, since it is used only as a general guide to the course material and the course is not based on it. References to relevant pages in this book are given below in the Syllabus denoted by "DC". Some material in the course is not covered by DC, sometimes the approach taken in this course differs from that in DC, and sometimes material given in class contradicts (incorrect) material in DC. Because of this, the references to DC are only a broad general guide to the material that will be covered in the lectures. Please note that I myself have not read this textbook.

Calculator You will need some sort of calculator for this course, but do not need a graphing calculator. You may use your phone or computer's calculator.

Programming It is difficult to perform any statistical analyses in the real-world by hand and statistical analyses almost always involve a computer program. These can range from JMP, SAS, SPSS, STATA, Python, R, etc. In social science applications, people tend to use SPSS or STATA. In heavily regulated applications, SAS is often used. Modern mathematicians and scientists likely use Python or R. To this end, I recommend you install R and RStudio on your machine. Instructions will be given for this later. You will not be required to use it. However, I will post a short intro to R for those who would like, and will state the relevant R functions for the material given in class. Using it on homeworks and quizzes will likely save you time and computational errors.

Contacting me You can contact me via canvas or my email address listed above, kbrum@wharton.upenn.edu. You can also post anonymously on Piazza. Please reach out if you'd like to meet individually.

SYLLABUS

Please note that these topics and dates are very tentative and will likely change as we move through the course. However, they should give you a rough idea of the content and pace of the course.

INTRODUCTION

1 Statistics and probability theory (Monday May 22)

1.1 What is Statistics?
1.2 The relation between probability theory and Statistics

PROBABILITY THEORY

2. Events (Monday May 22)

2.1 What are events?
2.2 Notation
2.3 Derived events: complements, unions and intersections of events (DC 34–40)
2.4 Mutually exclusive events

3 Probabilities of events (Tuesday and Wednesday May 23 and 24) (DC 35–40)

3.1 Probabilities of derived events
3.2 Independence of two events (DC 79-80)
3.3 Conditional probabilities
3.4 Conditional probabilities and mutually exclusive events
3.5 Conditional probabilities and independence

4 part I Probability: one discrete random variable (Wednesday and Thursday May 24 and 25)

4.1 Random variables (DC 87–92)
4.2 Random variables and data
4.3 The probability distribution of a discrete random variable (DC 87–106)
4.4 Parameters
4.5 The binomial distribution (DC 107-118)
4.6 The hypergeometric distribution (we will probably skip this)

Friday May 26: quiz 1 on sections 1-4.3

4 part II Probability: one discrete random variable (Tuesday May 30)

4.7 The mean of a discrete random variable (DC 93–95)
4.8 The variance of a discrete random variable (DC 95–99)

5 Many random variables (Wednesday and Thursday May 31 and June 1)

5.1 Introduction
5.2 Notation
5.3 Independently and identically distributed random variables
5.4 The mean and variance of a sum and of an average
5.5 The mean and variance of a difference
5.6 The proportion of successes in n binomial trials

6 part I Continuous random variables (Thursday June 1) (DC 131–140).

6.1 Definition
6.2 The mean and variance of a continuous random variable (DC 138–140).

Friday June 2: quiz 2 on sections 4.4 through 5.6

6 part II Continuous random variables (Monday through Wednesday June 5-7) (DC 131–140).

6.3 The normal distribution (DC 143–155).
6.4 The standardization (z-ing) procedure (DC 147–151).
6.5 Numbers that are see often in Statistics (DC 230)
6.6 Using the Normal distribution chart in reverse
6.7 Sums, averages and differences of independent normal random variables
6.8 The Central Limit Theorem (DC 192-198)
6.9 Approximating discrete random variable probabilities using the Normal distribution (DC 193)

STATISTICS

7 Introduction (Wednesday June 7)

8 Estimation (of a parameter) (Wednesday and Thursday June 7 and 8)

8.1 Introduction
8.2 Estimation of the binomial parameter θ (DC 265–268).

Friday June 9: quiz 3 on section 6

8 Part II Estimation (of a parameter) (Monday through Wednesday June 12-14)

8.3 Estimation of a mean (μ) (DC 205–207, 216-217).
8.4 Estimating the difference between two binomial parameters
8.5 Estimating the difference between two means
8.6 Regression (DC 289–300).

9 part I Testing hypotheses (about the value of a parameter) (Thursday June 15) (DC 227–245)

9.1 Introduction to hypothesis testing (DC 13–15, 231–236)
9.2 Two approaches to hypothesis testing (part 1)

Friday June 16: quiz 4 on section 8

9 part II Testing hypotheses (about the value of a parameter) (Tuesday June 20) (DC 227–245)

9.2 Two approaches to hypothesis testing
9.3 The hypothesis testing procedure and the concepts of deduction and induction
9.4 Power (we will probably skip this)

13 part I Tests on means (Wednesday and Thursday June 21 and 22)

13.1 The one-sample t test (DC 232–233)
13.2 The two-sample t test (DC 236–239)
13.3 The paired two-sample t test (DC 239–240)

Friday June 23: quiz 5 on sections 9 and 13.1-13.2

13 part II Tests on means (Monday June 26)

13.4 t tests in regression (DC 299)
13.5 General notes on t tests
13.6 Exact confidence intervals (we will probably skip this)

10 Testing for the equality of two binomial parameters (DC 240–242) (Tuesday June 27)

11 Chi-square tests (i): Tables bigger than two-by-two (DC 243–245) (Wednesday June 28)

If we are ahead of schedule, we may also cover Chapters 12 and 14, a second use of chi-square tests and non-parametric tests.