InfSyll Fall 2017

STAT 111 – Fall 2017

INFORMATION AND SYLLABUS

Lecturer: Dr Warren J Ewens

wewens@sas.upenn.edu

Dr Ewens' office hours Dr Ewens' office hours are open. Do not hesitate to make an appointment for any time 9-5 Mon-Fri via the email address above.

Office location Because of renovations in Huntsman Hall my office will initally be in room **307 Leidy Labs.** Leidy Labs is the "Zoology" building at the corner of 38th St and Hamilton Walk. Unless otherwise agreed, and until further notice, the meeting place for all appointments will be that room.

Contacting Dr Ewens Email is the ONLY way that you should contact Dr Ewens. NOTE: do NOT send emails via the "reply" or "reply all" mechanism to a message which was sent to the entire class (and thus headed "folks"). Doing so might send your message to the entire class. Instead, initiate a new separate email in reply to any message sent to the entire class.

Email messages Messages are often sent to the class by email, so check your email at least twice a day for messages relating to this class.

Lectures Lectures are given through two sections, Sections 1 and 2. Section 1 meets Tuesday-Thursday at 11:00 am - 11:50 am and Section 2 meets Tuesday-Thursday at 2 pm-2:50 pm. When you register for this course you also register for one of these two sections. All lectures are given in Steinberg-Dietrich Hall room 351. The first lecture for the semester will be held on Tuesday August 29. On any given day the lectures in the two sections are identical, so that if on some day you cannot go to your normally scheduled section's lecture you can go to the other section's lecture meeting on the same day.

Announcements Important announcements will often be given in class. If for some reason you miss a class it is up to you to find out from a friend if any important announcements were made in the class that you missed. So far as is possible these announcements will also be posted on "Canvas". (For more on "Canvas" see **Web Resource: "Canvas"** below.)

Mid-term break arrangements The Fall Mid-term break is October 5 - 8 inclusive. Arrangements concerning homeworks due around that time will be announced later.

Recitation classes Recitation classes are held on Fridays. The first recitation class will be on **Friday September8.** Except for Homework 1, which will be handed out in class

Thursday August 31, homeworks will be handed out to you in recitation classes and your answers are to be handed in at the recitation class one week later. (Homeworks will also be posted on Canvas - see below). When you register for this course you also register for the recitation class you plan to attend. Students attending Section 1 lectures must register for either recitation section 201, 202, 203 or 204. Students attending Section 2 lectures must register for either recitation section 205, 206, 207 or 208.

Recitation classes are given by teaching assistants (TAs) and will be held in ground-floor rooms in Huntsman Hall. Each recitation class lasts for 50 minutes. The times and places for these are as follows:

Recitation class 201 is at 11 am, room 105 SH-DH
Recitation class 202 is at 12 noon, room 105 SH-DH
Recitation class 203 is at 1 pm, room 105 SH-DH
Recitation class 204 is at 2 pm, room 105 SH-DH
Recitation class 205 is at 11 am, room 213 SH-DH
Recitation class 206 is at 12 noon, room 213 SH-DH
Recitation class 207 is at 1 pm, room 213 SH-DH
Recitation class 208 is at 2 pm, room 213 SH-DH

SH-Dh is Steinberg Deitrich Hall, on Locust Walk.

The TAs for these classes are as follows:

Recitation classes 201 and 202: Shaokun Li
Recitation classes 203 and 204: Dan Sample
Recitation classes 205 and 206: Linjun Zhang
Recitation classes 207 and 208: Arun Kuchibhotia

The TAs are there to help you. Please remember the name of your TA and the recitation class that you are in.

Homeworks As noted above, homeworks will normally be handed out on Fridays in recitation classes, and answers should be handed in at recitation class one week later. (Special arrangements will be made later for the Fall break and the week of Thanksgiving.) There are many students in this class with similar names and it is easy to confuse homeworks unless you indicate your name clearly on each homework. To assist with this, please write your first name first and in full (i.e. not just initials), and then your family name written last and in CAPS, for example Mary SMITH on each homework. If you have a hypenated family name, please write in CAPS the name (or names) in Penn records (that is, either Mary SMITH-Jones, Mary Smith-JONES or Mary SMITH-JONES, whichever is appropri-

ate. **Asian students**: Please follow the Western convention of writing your family name last and in CAPS. Also, please use your correct Asian first name and not an anglicised first name/nickname.

Homework and exam point scores will be posted on "Canvas" - see more on Canvas below. You should check regularly that all your homework scores are entered correctly.

Unless there is a reason for handing a homework in late, (such as a medical reason), there will be a point score penalty for late homeworks.

Homeworks that are not picked up in recitation classes will be placed in a box labelled "STAT 111" in the lobby of the Statistics Department (4th floor, Huntsman Hall). It is important that any homework not picked up in a recitation class be picked up from that box as soon as possible.

Homework point scores and mid-term exam point scores as given on canvas on December 14 will be taken as final.

Exams There will be one mid-term exam and one final exam. Details are as follows.

The **Mid-term exam** will be given **6- 7:30 pm Wednesday October11.** Although that might not be a convenient day and time for some people, there are problems with other days due to availability of room space, religious holidays, and so on. If this day and time is not possible for you, please let Dr Ewens know immediately at the email address given above. The lecture on October 10 will be a pre-exam review. Also, Dr Ewens will be available all day at Penn on October 11 on a walk-in basis in room 307 Leidy Labs to answer any pre-exam questions that you might have, and will also be available by email on that day. Note: depending on the Huntsman Hall renovation schedule, this walk-in location might be changed to Huntsman Hall.

The location of this exam is still being finalized, and you will be told the location several days before the exam. The mid-term exam will be of 1 1/2 hours duration.

The **Final exam** will be held 3 - 5 pm Thursday December 14. This exam will be of 2 hours duration. The timing of this exam is set by the university and cannot be changed. If this time is **impossible** for you, contact Dr Ewens immediately. More information as to the location of this exam will be given later, when it becomes finalized.

Assessment The assessment in this course is by homeworks (10%), the Mid-term exam (35%) and the final exam (55%). Some of the questions on the mid-term and final exams will be questions previously set in homeworks. Thus the homeworks in effect carry a higher

percentage value of the overall score than is suggested by the above.

Textbook Printed notes will be sent to you via email, and will also be placed on canvas. These notes are in effect the textbook for the course. It is likely that not all topics in these notes will be covered in class. **Only those topics discussed in class are examinable.**

If you do want to buy a published textbook you should get Downing and Clark, "E-Z Statistics", Barron, 2009, ISBN 13: 978-0-7641-3978-9. This book should be available in 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. (It also contains some errors.) References to relevant pages in this book are given below.

Calculator You will need a hand calculator for this course. All that is needed is a calculator doing the elementary operations of addition, subtraction, multiplication and division as well as taking square roots. You do not need a graphing calculator or one doing operations like taking logarithms. **You will need your calculator for both the mid-term and the final exams.**

Web resource: "Canvas" The web resource in this course is "Canvas". This is available to all Penn students at **https://canvas.upenn.edu** You will need your pennkey authentification (username and password) to use Canvas. You will be able to see your homework scores on canvas, as well as the homework problems, the course notes and other material.

For a copy of the class notes, for this information/ syllabus document, for announcements and for weekly homeworks, use the "Files" link in canvas. For point score information, use the "Grades" link.

For questions about using Canvas you can contact the Wharton Computing Student support office at 215 898 8600 or at **https://spike.wharton.upenn.edu/support**.

JMP The course will be given in part in association with the statistical package JMP. You should either buy and install this package on your computer or (much better, since buying JMP is expensive) use the (free) Wharton computers that have it installed.

Note for non-Wharton students. If you do not have a Wharton computing account you will need to establish one to be able to access Wharton computers. To create an account, go to https://app.wharton.upenn.edu/accounts/ .

Alternatively, Penn students can get a JMP license through e-academy at
http://www.onthehub.com/jmp/ for $30 for a 6 month license or $50 for a year license.

It is also possible that you can carry out JMP operations by using Wharton virtual lap from your laptop. Instructions about this are in this link:
https://whartonstudentsupport.zendesk.com/hc/en-us/articles/202151436-Virtual-Lab-for-Laptops-

If you have any further questions about JMP, contact Dr Ewens.

Disabilities If you are registered through the Weingarten Center for special arrangements for exams etc., **please contact Dr Ewens as soon as possible and let him know this**. Also, please forward to Dr Ewens all messages that you send to, or receive from, the Weingarten Center.

Course description The content of this course falls into two broad categories, namely probability theory and Statistics. The reason why we discuss probability theory will be given in the first lecture. A more detailed list of the topics covered within these two categories is given in the syllabus below. **References to the pages for the corresponding material in the textbook by Downing and Clark for these topics are given in parentheses, for example (DC107-118).** Note that some material in the course is not covered by Downing and Clark, that sometimes the approach taken in class to some topics differs from that in Downing and Clark, and that sometimes material given in class contradicts (incorrect) material in Downing and Clark. Therefore the references to Downing and Clark are only a general guide to the material that will be covered in class.

SYLLABUS

INTRODUCTION

1 Statistics and probability theory

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

PROBABILITY THEORY

2. Events (DC 32–34)
2.1 What are events?
2.2 Notation
2.3 Unions, intersections and complements of events (DC 34–40).

3 Probabilities of events (DC 35–40)

3.1 Probabilities of derived events
3.2 Mutually exclusive events
3.3 Independence of events. (DC 79-80).
3.4 Examples of probability calculations involving unions and intersections
3.5 Conditional probabilities of events (DC 75–86).
3.6 Example using a die.

4 Probability: one discrete random variable

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

5 Many random variables

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 Two generalizations
5.6 The proportion of successes in n binomial trials
5.7 The standard deviation and the standard error
5.8 Means and averages

6 Continuous random variables (DC 131–140).

6.1 Definition
6.2 The mean and variance of a continuous random variable (DC 138–140).
6.3 The normal distribution (DC 143–155).
6.4 The standardization (z-ing) procedure (DC 147–151).
6.5 Numbers that you will see often (DC 230)
6.6 Sums, averages and differences of independent normal random variables
6.7 The Central Limit Theorem (DC 192-198)
6.8 The Central Limit Theorem and the binomial distribution (DC 193)
6.9 The chi-square distribution (DC 161–164).

STATISTICS

7 Introduction

8 Estimation (of a parameter)

8.1 Introduction
8.2 Estimation of the binomial parameter θ (DC 265–268).
8.3 Estimation of a mean (μ) (DC 205–207, 216-217).
8.4 Estimation of a variance
8.5 Notes on the above example
8.6 Estimating the difference between two binomial parameters
8.7 Estimating the difference between two means
8.8 Regression. (DC 289–300).

9 Testing hypotheses (DC 227–245)

9.1 Introduction (DC 13–15, 231–236)
9.2 Two approaches to hypothesis testing
. 9.2.1 Both approaches, Step 1
. 9.2.2 Both approaches, Step 2
. 9.2.3 Both approaches, Step 3
. 9.2.4 Approach 1, Step 4, the medicine example
. 9.2.5 Approach 1, Step 5, the medicine example
. 9.2.6 Approach 1, Step 4, the coin example
. 9.2.7 Approach 1, Step 5, the coin example
. 9.2.8 Approach 2, Step 4, the medicine and the coin examples
. 9.2.9 Approach 2, Step 5, the medicine example
. 9.2.10 Approach 2, Step 5, the coin example
9.3 The hypothesis testing procedure and the concepts of deduction and induction

10 Tests on means

10.1 The one-sample t test (DC 232–233)
10.2 The two-sample t test (DC 236–239)
10.3 The paired two-sample t test (DC 239–240)
10.4 t tests in regression (DC 299)
10.5 General notes on t tests

11 Testing for the equality of two binomial parameters (DC 240–242)
11.1 two-by-two tables
11.2 Tables bigger than two-by-two (DC 243–245)

11.3 Another use of chi-square: testing for a specified probability distribution (DC 246–247)

12 The nature of the tests of hypothesis considered so far

13 Non-parametric (= distribution-free) tests (DC 277)

. 13.1 Introduction

. 13.2 The non-parametric alternative to the one-sample t test: the Wilcoxon

. signed-rank test (DC 282–284)

. 13.3 The non-parametric alternative to the two-sample t test: the Wilcoxon

. rank-sum test (DC 280–281)

. 13.4 Testing for randomness of events in space or in time: the" runs" test