STAT 111 – Summer 2015

GENERAL INFORMATION AND SYLLABUS

Lecturer: Ville Satopää

satopaa@wharton.upenn.edu

Office: Room 450 Huntsman Hall.

Contact

- The main points of contact with the instructor are after class and during the highly recommended office hours.
- Your primary source for asking questions and finding answers will be the Canvas Discussions board (see below).
- Canvas has an e-mail feature which you reach by clicking Inbox. Therefore, never use regular e-mail; always send messages from Canvas Inbox to Ville Satopaa.

Office Hours

2:00-3:00pm Monday - Thursday in JMHH 441 (Conference Room in Statistics Department)
Except on Tuesday, June 16 and Thursday, June 18.
7:00-8:30pm Sunday in JMHH 441 (Conference Room in Statistics Department)

Class time

The course runs from May 26 - July 1, 2015.
The lectures are Monday - Friday 11:00am - 12:45pm in 105 SH-DH.
Review sessions on Monday, June 29 and Tuesday, June 30.
Final Exam on Wednesday, July 1.

Course Website

STAT 111 uses the online system "**Canvas**". Searching "Wharton Canvas" in Google will get you to the login page. All materials for this course will be managed via this website, and you will be able to monitor your grade entries throughout the semester.

An important part of Canvas is the "**Discussions**" feature where everybody can place questions and comments. I will be using Canvas Discussions extensively for answering your questions about homeworks, exams and scheduling. If you have questions, you **must** go there first to see whether your question has already been asked and answered and, if not,

to place your question so it can be answered once for everybody. I also encourage you to answer other students' questions if you know the answer. If you give a wrong answer, it will not be held against you; someone else will place the correct answer and you can delete your own post. Contributing to Canvas Discussions will be equivalent to class participation and office hour attendance in its impact on your grade.

The Canvas Discussions settings will allow you to add, edit and delete your own posts, create discussion topics, and also attach files to your posts. Please, read the initial post by the instructor regarding purpose and etiquette.

Note for non-Wharton students: If you do not have a Wharton computing account, you will need to establish one to access the website. The account also provides access to the computing labs in Wharton and to the intranet. To get an account, on or after the first day of classes, go to `https://app.wharton.upenn.edu/accounts/`. After you have obtained your account, allow up to 12 hours for activation. Wharton students and students who have recently taken a Wharton course have existing accounts.

Course Material

- **Textbook:** There is no required textbook for this course. Printed notes will be available at no cost to you, and these can serve as a textbook. If however you do want to buy a textbook you should get Downing and Clark, "E-Z Statistics", Barron, 2009, ISBN 13: 978-0-7641-3978-9. However this book is **not** required, since it is used only as a general guide to the course material and the course is not firmly based on it. (It also contains some errors.)

- **JMP:** The course will in part be given in association with use of 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 Wharton computers that have it installed. You will not be able to use these computers until you have created a Wharton account. If you are a non-Wharton student please create a class account at:

  `https://whartonstudentsupport.zendesk.com/hc/en-us/articles/202127736-Creating-a-Wharton-CLASS-Account`

  If you decide to go this route, it might be helpful to check out the Wharton Virtual Lab feature which allows you to use the Wharton computing environment remotely from your laptop:

  `http://supportcenteronline.com/link/portal/632/655/Article/5657/5a-Virtual-Lab-for-Laptops`

  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.

Exams

- There will be 5 quizzes, one each Monday in the beginning of class.
- The final exam will take place on **Wednesday, July 1** during class time (room TBA).

During examinations strict rules will be in effect with regard to honor code. There is a possibility of contesting the results of exams as follows:

- Hand a written note to the instructor with the questions you are contesting and the reasoning as concisely as possible.
- However, it is recommended that you carefully check your reasoning before sending your official inquiry because if your reasoning reveals a misunderstanding of concepts taught in class, you may incur a deduction from your exam score.

Homework

- There will be 4 long and 1 short homework assignments.
- The first assignment is short and will be handed out during the first day of class. This is due during the class on Friday, May 29. After this, homeworks will be assigned on Fridays and will be due the following Friday (in class).
- **Late homeworks are typically not accepted.** If you have a valid reason for late homework submission (e.g., computer broken, health problems,...), send a Canvas Inbox email to the instructor. It is possible to receive extensions for valid reasons.
- Your solutions must show on the cover page your name (**as it appears on Canvas**)
- Homework is designed to teach, and you are encouraged to seek help from the instructor. You may also work with and help each other, in person and above all on Canvas Discussions.
- **You must, however, submit your own solutions, with your own write-up and in your own words. Verbatim copying and working off someone elses file is a violation of the honor code. Printed homework solutions and homework files must not be exchanged!**
- **Missing** homeworks receive a score of zero. **It is your responsibility to turn your homework in on time during class.**

Grading Policy

- Your course grade will be calculated as 10% homework, 50% quizzes, and 40% final exam.

- The lowest quiz score is dropped.
- Homeworks are graded only based on completion. Some of the questions on the quizzes and final exam will be questions previously set in homeworks. Thus the homeworks actually carry a higher percentage value of the overall score than is suggested by the above.
- Participation in class and on Canvas Discussion as well as office hour attendance will also figure into the final grade as follows: students who contribute exceptionally in at least one of these three ways will experience a rise of their grade by one step (e.g., from B- to B).

Classroom Expectations

- Class starts and ends on time.
- There will be a 5 min break about halfway through each lecture.
- Late entry or reentry is a disruption that must be explained to the instructor during the break or after class.
- All phones, laptops and other electronic devices must be turned off. The instructor may allow temporary laptop use for specific software demonstrations.

The general principles of conduct in this class are:

Fairness, Respect and Consideration.

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 tentative list of the topics covered within these two categories is given in the syllabus on the next two pages. **The goal is to cover as much material as possible without rushing. References to corresponding material in the textbook by Downing and Clark for these topics are given in parentheses (....), as 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
1. An example

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 An unfair die

4 Probability: one discrete random variable
4.1 Random variables and data
4.2 Definition: one discrete random variable (DC 87–92)
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 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 normal distribution 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 General principles of the estimation of a parameter
8.3 Estimation of the binomial parameter θ (DC 265–268).
8.4 Estimation of a mean (μ) (DC 205–207).
8.5 The 95% confidence interval for a mean μ (DC 216–217)
8.6 Estimation of a variance
8.7 Notes on the above example
8.8 Estimating the difference between two binomial parameters
8.9 Estimating the difference between two means
8.10 Regression. (DC 289–300).

9 Hypothesis testing (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