STAT 111– Summer 2016

SYLLABUS AND GENERAL INFORMATION

Lecturer: Zijian Guo
Email: zijguo@wharton.upenn.edu
Office: JMHH 476, 3730 Walnut Street

Course title INTRODUCTORY STATISTICS.

Lectures Monday - Friday at 11:00 am to 12:45 pm. The class starts from May 23, 2016 and lasts until June 29, 2016.

Office hours 4-5pm Monday and Thursday, 3-5pm Sunday.

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.)

Homeworks homeworks will normally be handed out on Fridays in classes and the answers should be handed in one week after. When handing in any homework, 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.

Homework 1 will be a short homework. It will be handed out in class on May 23 and should be handed in on Friday May 27.

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

Quizzes: There will be in-class quizzes on every Monday of classes.
The first quizz is on Tuesday May 31 since that Monday is Memorial Day.

Final Exam The final exam will be given **Thursday 30 June, 3 - 5 pm**. More information as to the location of this exam will be given later, when it becomes available.

Assessment

- Homeworks: 10%;
- Attendance: 10%;
- Quizzes: 40%;
- Final: 40%.

Some of the questions on the quizzes and final exam 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 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. 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 firmly based on it. (It also contains some errors.)

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.

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 pasword) to use Canvas.

For a copy of the class notes, and also for the syllabus/information document 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** Also you can contact the instructor for problems concerning Canvas.

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 (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.

If you have any questions about JMP, contact the instructor.

Disabilities If you are registered through the Weingarten Center for special arrangements for exams etc., please contact the instructor as soon as possible.

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
1.3 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
. 9.2.10 Approach 2, Step 5, the coin example
9.3 The hypothesis testing procedure and the concepts of deduction and induction
9.4 Tests for the equality of two binomial parameters (DC 240–242)
9.5 Tables bigger than two-by-two (DC 243–245)
9.6 Another use of chi-square: testing for a specified probability distribution (DC 246–247)

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 Non-parametric (= distribution-free) tests (DC 277)
. 10.5.1 Introduction
. 10.5.2 The non-parametric alternative to the one-sample t test: the Wilcoxon
. signed-rank test (DC 282–284)
. 10.5.3 The non-parametric alternative to the two-sample t test: the Wilcoxon
. rank-sum test (DC 280–281)