STAT 111 – Summer 2017

INFORMATION AND SYLLABUS

Lecturer: Bikram Karmakar

Email: bikramk@wharton.upenn.edu

Office: Room 449 John M. Huntsman Hall, 3730 Walnut Street

Course Title: INTRODUCTORY STATISTICS.

Lectures: There is only one section for this class. Lectures are scheduled five days each week; Monday through Friday 11:00AM – 12:45 PM from 22nd May to 28th June. The lectures are given in Vance Hall room B11 (in the Basement level). Vance Hall is at 3733 Spruce Street.

Each and every lecture is important.

Daily attendance will be taken at the end of each lecture.

Emailing is the most efficient way to get in contact with the instructor outside of the class. When emailing, please start the subject line of the email with "STAT111:".

Email updates and information about various organizational structures of the course are frequently sent to the class. So, please check your email regularly for such messages.

Office Hours: Office hours are open. You are welcome to request appointments for some time between 1:00 pm to 5:00 pm on weekdays. Unless otherwise agreed, the meeting place for all appointments is the office of the instructor.

Textbook: The most relevant materials for the course are the class notes. Depending on the topic being covered in the class on the day, printed notes and slides will be handed out in the class and uploaded on the course webpage (see below). **It is highly recommended that you take detailed note during the class.** There is **no required textbook** for the class. However, the suggested book for the course is "Head First Statistics", Edition 8, by Griffiths, ISBN: 978059652758.

Course Webpage: We are going to be using online system "Canvas". Canvas is available to all Penn students at https://canvas.upenn.edu using pennkey authentication (username and password). Canvas page will have all the class materials and grades throughout the semester.

For getting access to various electronic documents of the course, use the "Files" link on the canvas page. For point score information, use the "Grades" link on the canvas page.

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

Homeworks: The lectures are five days every week. To keep up with the pace of the class, we are going to have **daily homework assignments**. Each homework assignment will be very short and will contain **only about one or two questions**. The homework assignment will be due on the next class day unless otherwise instructed in class.

If you fall behind in the class please consider coming to office hours and also collecting all the notes from a fellow classmates**.**

Assessment:

- Attendance: 20%
- Class participation: 10% + 5% = 15%
- Homeworks: 20%
- Final: 45%

The easiest way to earn points for the class is to attend the lectures. If your class participation is less than 70% you will lose 20% of the grade.

10% of the grade under 'class participation' will be dependent upon your overall performance.

To earn the extra 5% of your grade under 'class participation', you need to ask at least one question during one of the lectures or correct the instructor in case of any wrong calculation in a lecture.

From time to time there will be chances to earn extra credits.

Final exam will be on June 28th (the last day of the classes) in class.

Classroom expectations

- Classes will start and end on time.
- You are expected to arrive at least 2 minutes before scheduled time and participate during the whole time of the lecture.
- It is highly recommended that you take detailed note during the class.
- All cell phones must be turned off in class.
- Daily attendance will be taken at the end of each lecture.

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. The aim of the course is to give you an introduction to the concepts in probability and provide you with a basic idea of statistical inference.

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

If you decide to choose the second option, 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.

You will not be able to use Wharton 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

SYLLABUS

The effective syllabus of the course will be whatever is covered in the class. But the following is the general guideline for the course.

INTRODUCTION

1. Statistics and Probability Theory
   - 1.1 Statistics?
   - 1.2 What Relates Statistics and Probability
   - 1.3 What Differentiates Statistics and Probability

2. Some uses and misuses of Statistics

PROBABILITY THEORY

1. Events
   - 1.1 What are events?
   - 1.2 Notation
   - 1.3 Unions, intersections and complements of events

2. Probability of events
   - 2.1 The three axioms of Probability
   - 2.2 Mutually exclusive events
   - 2.3 Independence of events
   - 2.4 Conditional probability of events
   - 2.5 Examples of probabilities

3. Random variable and its probability distribution
   - 3.1 What is a random variable?
   - 3.2 Definition of a discrete random variable
   - 3.3 The probability distribution of a discrete random variable
   - 3.4 Parameters
     - 3.4.1 Definition
     - 3.4.2 Parametric functions
   - 3.5 Measure of central tendency and dispersion of a random variable

4. The Binomial distribution
   - 4.1 The Binomial formula
   - 4.2 Mean and variance of a binomial distribution

5. Many random variables
   - 5.1 Independently and identically distributed (iid) random variables

5.2 Mean and variance of sum and average of iid random variables
5.3 General formulas

6. Continuous random variables
   6.1 Definition
   6.2 Probability density function
   6.3 Probability and the area under the curve
   6.4 The mean and the variance of a continuous random variable

7. The Normal distribution
   7.1 Normal distribution curve
   7.2 The standardization procedure
   7.3 Five popular numbers related to standard normal distribution
   7.4 The Central Limit Theorem
   7.5 The chi-square distribution

STATISTICS

1. Data and its graphical representation
   1.1 Drawing a histogram
   1.2 Drawing Scatter plots

2. Introduction to Statistical inference
   2.1 What is statistical inference?
   2.2 Three types of statistical inference problems
   2.3 Examples

3. Estimation of a parameter
   3.1 General principles of estimation of a parameter
   3.2 Estimation of the binomial parameter θ
   3.3 Estimation of a mean
   3.4 Estimation of a variance
   3.5 Examples

4. Confidence statements
   4.1 General idea of a confidence interval
   4.2 Level of confidence
   4.3 Confidence interval for a mean μ
   4.4 Examples

5. One sample and two sample problems
   5.1 Examples

6. Simple linear regression (SLR)
   - 6.1 Assumptions of SLR
   - 6.2 Geometry of lines
   - 6.3 Estimation of the parameters
   - 6.4 Examples

7. Testing of Hypothesis
   - 7.1 Introduction
   - 7.2 Examples
   - 7.3 Null and alternative hypothesis
   - 7.4 Two approaches to hypothesis testing
     - 7.4.1 Both approaches - Step 1, Step 2 and Step 3
     - 7.4.2 Approach 1 - Step 4 and Step5
     - 7.4.3 Approach 2 - Step 4 and Step 5
   - 7.5 Type I and Type II error
   - 7.6 Critical value and P value
   - 7.7 The concepts of deduction and induction
   - 7.8 Tests for the equality of two binomial parameters
   - 7.9 Chi-square tests

8. Tests of means
   - 8.1 The one sample t test
   - 8.2 The two sample t test
   - 8.3 The paired two sample t test
   - 8.4 t test for the parameters of SLR
   - 8.5 Examples

9. Non-parametric (= distribution-free) tests
   - 9.1 The Wilcoxon signed-rank test
   - 9.2 The Wilcoxon rank-sum test
   - 9.3 Examples