STAT 100 - Introduction to Quantitative Methods
for the Social Sciences and Humanities
FALL 2017 Syllabus

Instructor: Kerrie Nelson
                        Office: Science Center Room 300E
                        Email: knelson@hsph.harvard.edu
                        Office Hours: Mon and Wed from 10:30am-11:30am

Lectures: Twice a week. Mon and Wed 1:05pm – 2:30pm in Science Center Hall E

Sections: Announced during first week of class

Teaching Fellow: Jaquelyn Jahn   *Email*: jaquelyn_jahn@g.harvard.edu

Course Assistant: Emma Dwight   *Email*: edwight@college.harvard.edu

Course web site: https://canvas.harvard.edu/courses/27830

Learning Objectives and Prerequisites:

Our objectives of this course are to:
  · explain basic concepts of statistics and probability
  · develop skills to enable students to apply basic statistical techniques using statistical software,
  · help students become informed and critical readers of academic and other literature

Some familiarity with basic algebraic notation at high school level is assumed, but we do not assume a strong math background at all. The class will emphasize a more applied approach rather than theoretical. Graduate students who wish to tailor the course to relate to their research are invited to discuss this with the instructor.

Lectures:

Statistics 100 has two lectures each week (1 1/2 hours each) and a discussion section. Lecture notes will be posted on our canvas course website, and paper copies brought to each class.   You may also view the notes on your laptop during class. You should attend as many lectures as possible as important material not included on the prepared course notes will be discussed during class.

If you miss a class for any reason it is <u>your responsibility</u> to obtain class notes and materials from our canvas website and from fellow students. We will often use clickers in class - these can be signed out for the entire semester at the Science Cabot library on the ground floor of the Science Center.

Weekly Section:

The one-hour weekly section begins in Week 2 (details to be announced). We will offer a few choices of dates and times each week, and it is suggested that you attend one per week. While attendance is optional, section provides an excellent opportunity to work through examples and to review difficult lecture material.   Our teaching fellow and course assistants are a great resource for answers to questions, help with problem sets and course projects.

Stat 100 vs Stat 104: Stat 100, Stat 101, Stat 102 and Stat 104 are the primary introduction to statistics courses offered by the Harvard Department of Statistics. Stat 101 is designed specifically for students intending to concentrate in psychology. Stat 102 emphasizes applications to biology and medicine. Stat 100 and 104 are designed for more general audiences. Both courses offer similar exposure to basic data analysis, probability, statistical inference and building models from data. Stat 100 focuses more on social sciences and humanities whereas Stat 104 focuses more on economics and policy analysis. However, about half the students in either course are not concentrators in these fields, so both courses also cover a variety of datasets from other fields.   Stat 104 covers more material and so progresses a little quicker than Stat 100. Stat 100 is more suitable for those students not as comfortable with mathematical notation and equations.

Outline of topics:

The following is an outline of material covered in the course:

1. Introduction and describing data (2 lectures)
2. Basic probability and random variables (2-3 lectures)
3. Binomial and Normal random variables, Central Limit Theorem (3 lectures)
4. Elementary statistical inference from one sample (2 lectures)
5. Statistical inference from two samples (2 lectures)
6. Analysis of Variance (ANOVA) – statistical inference from many samples (1 lecture)
7. Paired data (1 lecture)
8. Inference for categorical data (1 lecture)
9. Simple and multiple linear regression, intro to logistic regression (5 lectures)

Textbook and Case Study:

Textbook: Moore, David S., McCabe, George P., and Craig, Bruce A. (2014). *Introduction to the Practice of Statistics*, 8th edition W. H. Freeman and Company, New York.

Purchase of this text is optional. Either 7th or 8th edition (new or second-hand) will provide a very useful reference. Copies are also available on deskcopy in the Statistics Library. A further free introductory text is *OpenIntro Statistics* by Diez et al, 3rd Edition, (https://www.openintro.org/stat/ textbook.php).

Clicker: You will need to sign out a Turning Point clicker at the Science Cabot library, for use in class.

Case Study: This is a required one-off purchase of approximately $8, partway through the semester. Details will be provided nearer the time.

Laptop and Cell Phone Policy:

Responsible use of laptops is permitted for viewing class notes in class. Please be considerate and silence and do not use cell or smartphones during class.

Homework (20%):

Approximately ten homework assignments will be assigned. These will be available at our canvas course website. Homeworks are to be submitted online each Tuesday by 5p.m. at the latest at our canvas website as either a word or pdf document. *No late homework will be accepted for any reason*.   You may drop one homework during the semester for any reason with no penalty.   Homework is not only a

fairly substantial portion of your grade, but it is vital to your success in this class. Working with other students on homework is allowed and encouraged (and with the instructor, TF and course assistants), but you must hand in your own work, and do not simply copy someone else's work. <u>Solutions prepared "in committee" or by copying/pasting or paraphrasing someone else's work will not be accepted</u>. All computer output you submit must come from work that you have done yourself. Solutions will be posted on the course website.

Homeworks are graded on the clarity and correctness of your presentation of the solutions. Homeworks that are generally clear and correct will score higher scores while sloppy and/or incomplete homeworks will receive a low score. Any clerical errors will be corrected without any hassle. Other regrade requests must be submitted in writing within a week of the item's return, where the entire item will be subject to regrading.

Exams (60%):

There will be one in-class 1 1/2 hour midterm exam during the semester, and a three-hour final exam. The dates for the exams are:

Midterm Exam (20%):   Wednesday October 18, 2017 in class
Final Exam (40%):   Monday December 11, 2017 2pm - 5pm

All exams will be closed book exams, though you will be allowed to bring two double-sided 8.5x11 inch page of your own notes to the midterm, and four double-sided 8.5x11 inch pages of your own notes to the final exam, and a multifunction calculator (smartphones may not be used during class or exams). Queries regarding grading of the midterm and final exam must be made to the instructor by Friday October 27th 2017 for the midterm, and December 31st 2017 for the final exam.

Missed Exams: Harvard's policy on missed exams may be found here :
http://static.fas.harvard.edu/registrar/ugrad_handbook/current/chapter2/attendance_absences_etc.html
Students who miss an exam due to a religious observance are entitled to a make-up exam. For any other excused absence (medical, sports or certain club activities) the remaining exam percentages will be reweighted and a make-up will not be offered.

Computing:

You will need a calculator with basic statistical functions including log, exponential and square-root functions. All course handouts, including lecture notes, homeworks, etc., will be available online at our Canvas website. Some homeworks will include computer assignments using a statistical software package, R and R Studio, which runs on both PCs and Macs. The software is free and available online through www.r-project.org. You will be taught everything you need to know how to run R. When you run analyses using R, you should cut and paste relevant output only into a word document containing your written text for the homework. Do not include lots of pages of computer output with your homework.

Course project (10%):

A project is required of all students. The projects will be carried out in pairs of two students, and provide students with valuable experience in analyzing a dataset from scratch and working as part of a team. We will provide the project description later in the semester. The project will be due on a date to be announced later in the semester with plenty of advance notice. Late projects will not be accepted after the due date and time.

Course help:

There are several avenues of help available if you need help with the course material during the semester:

- For questions or clarifications, you can send an e-mail message to the TF or instructor
- You are always welcome to visit our weekly office hours (or by making an appointment)
- Attending weekly sections run by the TF and Course Assistants
- Tutoring and problem set help will be also offered by the Study Network, which is held several evenings per week for the most weeks during semester. Details to be announced.

If you feel like are getting behind with the course material and need extra individual mentoring time beyond course office hours, you may want to consider contacting the Bureau of Study Counsel very early in the semester (http://bsc.harvard.edu). They employ former Statistics 100 students who can serve as private tutors at a nominal hourly fee.

Grades:

Course grades will be determined by the following components, with the weights shown:

Problem Sets	20%
Midterm Exam	20%
Final Exam	40%
Course Project	10%
Class participation	10%
	100%

Student Responsibilities and Academic Integrity:

Harvard College is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. Cheating on exams and quizzes, plagiarism and copying others' work on homework assignments and projects, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, and will not be tolerated. Such incidences will result in a grade of zero for all parties involved as well as being reported to the Administration Board. Please review Harvard's Academic Dishonesty policies at the following two links: http://handbook.fas.harvard.edu/book/academic-integrity#onehttp://static.fas.harvard.edu/ registrar/ugrad_handbook/current/chapter2/academic_dishonesty.html

Some Helpful R Resources:

All the commands you will need in R will be demonstrated in lecture or section, but you might find the following resources useful too:

25 Recipes for Getting started with R (GSR) by P. Teetor.
The R Cookbook (TRC), by P. Teetor is a longer version of similar material.
Getting started with R Studio (GSRStudio), by J. Verzani.

The Harvard Library System has free access to these as eBooks (links on course website). Other useful resources will also be mentioned as the semester progresses.

Students with Disabilities:
Students with disabilities who believe they may need accommodations in this class are encouraged to contact the Accessible Education Office as soon as possible to ensure that such accommodations can be made. The instructor must be informed in writing of any required AEO accommodations <u>at least two weeks before the midterm</u>.

Course Schedule: (subject to change during semester)

Class 1 (Wednesday August 30th): Introduction and Describing Data, Part 1
- Review important class information
- General introduction
- Use of R
- Graphical and numerical summaries of data

Monday 4th September: Labor Day Holiday – no class

Class 2 (Wednesday September 6): Describing Data, Part 2
- Graphical and numerical summaries of data

Class 3 (Monday September 11): Introduction to probability
- What is probability?
- Outcomes and events in an experiment
- Different ways to calculate probabilities

Class 4 (Wednesday September 13): Topics in probability
- Using tables to find probabilities
- The Independence Rule
- The multiplication Rule

Class 5 (Monday September 18): Further topics in probability Case Study
- Conditional probability

Class 6 (Wednesday September 20): Discrete Distributions
- What is a random variable?
- Discrete versus continuous random variables
- The Binomial distribution
- The Poisson distribution

Class 7 (Monday September 25): Continuous Distributions
- The normal distribution
- Using the normal distribution to find probabilities
- The uniform distribution

Class 8 (Wednesday September 27): The Central Limit Theorem and Sums of Random Variables
- Defining the Central Limit Theorem
- Uses of the Central Limit Theorem
- Sums of random variables

Class 9 (Monday October 2): Introduction to Hypothesis Testing, Part I
- A hypothesis test for a single mean
- What is a p-value?
- The t-distribution
- Conducting a hypothesis test in R

Class 10 (Wednesday October 4): Introduction to Hypothesis Testing, Part II
- A hypothesis test for comparing two means
- Type I and II errors
- Sample size versus population size
- Practical versus statistical significance

Monday October 9 - Columbus Day Holiday – no class

Class 11 (Wednesday October 11): Confidence intervals, Part I
- The single proportion
- What does a confidence interval tell us?
- A confidence intervals for a single mean and a single proportion

Class 12 (Monday October 16): Midterm Review

Class 13 (Wednesday October 18): Midterm Exam (in-class) - covers Classes 1 to 12

Class 14 (Monday October 23): Confidence intervals, Part II
- A hypothesis test and confidence interval for comparing two proportions
- A confidence interval for comparing two means
- How big a sample should I take?

Class 15 (Wednesday October 25): Paired Data
- How to deal with dependencies in data
- McNemar's Test

Class 16 (Monday October 30): One-way Analysis of Variance (ANOVA)
- Comparing several groups at once
- The F-test
- Correcting for multiple comparisons

Class 17 (Wednesday November 1): Categorical Data
- Two-by-two tables
- The Chi-square test
- Large tables

Class 18 (Monday November 6): Relationships between two continuous variables
- Correlation

Class 19 (Wednesday November 8): Simple Linear Regression, Part 1
- When we use simple linear regression
- How to find the best-fitting line
- Using the regression line

Class 20 (Monday November 13): Simple Linear Regression, Part 2
- Checking assumptions of simple linear regression
- Analysis of the residuals

Class 21 (Wednesday November 15): Multiple Regression, Part 1
- The multiple regression model
- Confidence intervals and hypothesis tests for the regression coefficients

Class 22 (Monday November 20): Multiple Regression, Part 2
- Checking assumptions of multiple regression

Wednesday November 22 - Thanksgiving Break, no class

Class 23 (Monday November 27): Logistic Regression
- An introduction to logistic regression
- Differences to linear regression

Class 24 (Wednesday November 29): Final Exam Review

Final Exam: Monday 11th December 2017 2pm to 5pm (tentative)