Course Information

Stat 200: Statistical Analysis (3 credits)

Course Description

Statistics provides the tools to sort through data to make objective decisions. Stat 200 provides an accelerated introduction to the basic tools for quantitively oriented students, with particular emphasis on understanding which tools are appropriate for which problems. We cover experimental design (including basics of causal inference from observations), basic probability (both frequentist and Bayes), descriptive statistics, inference from samples (including null hypothesis significance tests), analysis of variance, multiple regression, logistic regression, and non-parametric methods. We include both an easy-to-use data analysis program and exercises in the R programming language.

For more details see http://courses.atlas.illinois.edu/fall2020/STAT200/

Prerequisite: Completion of or simultaneous registration in Calculus I is recommended. High school students interested in taking this course have to submit proof of satisfying this prerequisite at the time of registration.

Course Objectives

Large stores of information have become readily available on crucial issues, e.g. Covid-19. You can choose either to ignore the information or to make sense of it, which means learning statistics. Statistics is a collection of real tools. The key is to understand which one to use when and why, not just to memorize some complicated rules. We teach

1. How to use a conceptual, intuitive approach to understand a set of complex statistical methods. We build a unifying framework for general predictive models.
2. How to determine whether predictors are also causes. Would changing X change Y? We examine real data sets to sort out such causal effects from confounders.
3. How to use statistical software both to help us understand what the statistical methods are doing and to do the calculations for us. For those who wish to become more fluent in R programming, we also offer a free non-credit R course .

Course Outline

Study Design - observational studies vs. randomized experiments, why randomized controls are key, identifying possible confounders in observational studies.

Descriptive Statistics - mean, median, SD, histograms, box plots, normal curve, etc.

Probability - multiplication rule, addition rule, conditional probability, Bayes rule

Diagnostic Tests - Type I and II errors, sensitivity and specificity.

Statistics for Random Variables - expected value and standard error of chance processes, probability histograms and convergence to normal curve. Focus is on developing simple chance models (box models- drawing numbers at random from a box) that more complicated sampling processes can be translated into.

Sampling and Statistical Inference - using sample means and percents to estimate population means and proportions, and attaching margins of errors to our estimates by computing confidence intervals. Why randomized sampling is key.

Significance Tests - one sample and two sample Z-tests and t-tests and chi-square tests for goodness of fit and independence. Focus is on understanding how these tests depend on chance models.

Simple Linear Regression - correlation coefficient, regression equation, etc.

Inference for Simple Linear Regression - Understanding the Simple Linear Model and Assumptions, Confidence Intervals and Significance tests for the Slope, Analysis of Variance for regression, etc.

Binary Variables in Multiple Linear Regression - Causal inference, adjusting for likely confounders by including them as covariates in the regression model. Interactions between binary and quantitative variables. Models with 2 binary predictors.

Multiple Regression with Quantitative X's - 3-D scatter plots and interpreting slopes graphically, Interactions, F-tests for overall regression effect and t-tests for slopes.

Re-randomization Methods - Randomization tests to calculate p-values for ANOVA and regression.

ANOVA for Comparing Group Means

Transformation of Variables- Fitting a linear model to non-linear data, log and square root transformations

Logistic Regression -The log odds equation, making predictions and interpreting the slopes, the odds ratio, multiple logistic regression, maximum likelihood methods to estimate the slopes.

Non-Parametric Statistics- Transform data into ranks to compute p-values using Wilcoxon Mann-Whitney test (rank sum or U stat), Kruskal-Wallis test and Spearman's rank-order correlation coefficient

Format

- This course is entirely online. All homework, surveys and bonus work are submitted and graded on Lon-Capa.

- Required Text: **Stat 200 Incomplete Notes Workbook**. Fireman, Laska, and Marden Fall 2020 edition. Publisher: Cengage. PDFs of the first few chapters will be posted until all students are able to get hard copies from the Illini Union Bookstore (IUB) or electronic versions from Cengage. This is the workbook you'll fill in while watching the lectures and submit for extra credit points. Only Fall 2020 edition will be accepted for extra credit.
    - Hard copies of the workbook may be purchased from IUB here. In the drop-down menu, select U OF ILLINOIS - FALL 2020 for **Campus Term**, STAT-STATISTICS for **Department**, 200 for **Course**, and L1-Fireman for **Section**.
    - Students located outside the contiguous United States (both international and domestic) should email Florence.Januzik@cengage.com to purchase a digital copy of the Workbook.
    - In special circumstances students may contact the Stat 200 instructor Ellen Fireman fireman@illinois.edu

- This course has fixed due dates and deadlines for homework and exams. You'll turn in homework twice weekly and R assignments once a week. All homework is submitted on Lon-Capa and you'll get immediate feedback on each problem as soon as you submit it. You also get multiple tries on all the problems. No late HW will be accepted.

- This course has three exams administered via an online platform. Additional information about exams will be provided within your course LMS.