EPIB 607 - Fall 2018
Principles of Inferential Statistics in Medicine
https://sahirbhatnagar.com/EPIB607

4 credits
Mondays 11:30-13:30 and Thursdays, 8:30-10:30
Location: McMed 1034

1 Instructor

Sahir Rai Bhatnagar (https://sahirbhatnagar.com/)
Department of Epidemiology, Biostatistics, and Occupational Health
Department of Diagnostic Radiology
McGill University
sahir.bhatnagar@mcgill.ca
Office hours: Tuesdays in Purvis Hall 37 *by appointment*

TA: Kody Crowell, Guanbo Wang, Himasara Marasinghe
Office hours: TBA

2 Prerequisites

Basic understanding of exponentials, logs, histograms, graphs, mean, median, mode, standard deviation. Enrollment in the Epidemiology or Public Health program at McGill University.

3 Objectives

The aim of this course is to provide students with basic principles of statistical inference so that they can:

- Visualize/Analyze/Interpret data using statistical methods with the R statistical software program.

- Understand the statistical results in a scientific paper.

- Apply statistical methods in their own research.

- Use the methods learned in this course as a foundation for more advanced biostatistics courses.

4 Audience

The principal audience is researchers in the natural and social sciences who haven't had an introductory course in statistics (or did have one a long time ago). This audience accepts that statistics has penetrated the life sciences pervasively and is required knowledge for both doing research and understanding scientific papers.

5 Teaching strategy

This course will follow the Flipped Classroom model: Here, students are expected to have engaged with the material before coming to class (based on very precise pre-class instructions). The students will then be expected to answer a series of conceptual multiple choice questions using the DALITE online platform (https://mydalite.org/, https://www.youtube.com/watch?v=0tJVVy2ay7c). This allows the instructor to delegate the delivery of basic content and definitions to textbooks and videos, and enforces the idea that students cannot be simply passive recipients of information. This approach then allows the professor to focus valuable class time on nurturing efficient discussions surrounding the ideas within the content, guiding interactive exploration of typical misconceptions, and promoting collaborative problem solving with peers.

A focus on computation: Classic introductory statistics textbooks were written during a time when computers were still in their infancy. As such, even the newer editions heavily rely on by-hand computations such as looking up tables for tail probabilities. We take a modern approach and introduce computational methods in statistics with the statistical software program `R`.

6 Tutorials from DataCamp

This class is supported by DataCamp, which will allow you to learn `R` through a combination of short expert videos and hands-on-the-keyboard exercises. You will be asked to complete some of the courses in DataCamp for background reading or for assignments. You can sign up for a free account at this link. Note: you are required to sign up with a @mail.mcgill.ca or @mcgill.ca email address.

7 Content

Course structure will consist of elaborating selected topics from the book Baldi & Moore : "The Practice of Statistics in the Life Sciences", 3rd edition. You will also be asked to watched the accompanying Against all odds video series by Annenberg Learner. We will also cover more advanced material not covered in the textbook, for which class notes will be made available.

7.1 Descriptive Statistics

- Histograms, density plots, measures of center, boxplots, standard deviation
- Data visualization (aesthetics, visual cues, coordinate systems, scales, facets and layers)
- Choosing color palettes: Cynthia Brewer palettes, perceptually uniform palettes, color blind friendly palettes.
- Tidy data

7.2 Sampling Distributions

- Parameters and statistics
- Standard error of the mean
- Normal (Gaussian) distribution
- Central Limit Theorem
- Confidence intervals
- Bootstrap for sampling distributions and confidence intervals

7.3 One Sample Inference

- Inference about a population mean
- P values, power, and sample size considerations
- Inference about a population proportion
- Inference about a population rate

7.4 Regression

- Linear regression for means, difference in means, ratio of means
- Poisson regression for rates, rate differences, rate ratios
- Logistic regression for odds ratios and risk ratios

8 References

8.1 Optional text

Baldi B and Moore D S. *The Practice of Statistics in the Life Sciences*, 3rd edition. Freeman and Company.

8.2   Course notes

These are available as a PDF document on MyCourses to download.

9   Equipment

Hand calculators (with square root, log, and exponential function) are required.

10   Software

The R software package will be introduced and used for in-class illustrations. R is available under GPL (free) at http://cran.r-project.org/.

It is recommended to use the RStudio interactive development environment (IDE) which can be downloaded for free at http://www.rstudio.com/.
Note: to use RStudio, you must first download the R software package at the link provided above.

11   Evaluation

Assignments (submit using MyCourses)	40%
Midterm exam (Formula sheet)	15%
DALITE	15%
Project	10%
Final exam (Formula sheet)	20%

The final grade will consist of a letter grade.

12   Note on academic integrity

McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see http://www.mcgill.ca/integrity/ for more information).