Introduction to Probability and Statistics

BMI XXXXX

Instructors

* Edgar Javier Hernandez, Ph.D.: [edgarjavi@gmail.com](mailto:edgarjavi@gmail.com)

Prerequisite

BMI Introduction to Programming (Python)

Course Description

This course offers an introduction to an extensive array of methods for mathematical data analysis with emphasis on three major topics (probability analysis, statistical inference, and the basic concepts of statistical pattern recognition through machine learning), with a clear emphasis in the biomedical field. In here we will cover basic probability concepts such as recognizing the importance of the analysis of random events in real life applications, using probabilities axioms and rules. This course will present descriptive and inferential data methods for predictive analysis on samples and populations. This introductory course lays the foundation for more advance classes offered at the Biomedical Informatics Department. As an additional component of this class will be the extensive use of the statistical software R which is one of the most used statistical packages in most fields.

***Text Books:***

* Introduction to Probability and Statistics Using R (2010 - <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf> ), Kerns, Jay.
* Think Bayes (2012 - <http://www.greenteapress.com/thinkbayes/thinkbayes.pdf> ), Downey, Allen. (both freely available online).
* R for data Science (2017), Grolemund & Wickham - <http://r4ds.had.co.nz/index.html>
* Also a suggested text book is
  + Think Stats: Exploratory Data Analysis in Python (2014 - <http://greenteapress.com/thinkstats2/thinkstats2.pdf> ) Downey, Allen.
  + Introduction to Biostatistics 2nd ed. (2009): Sokal & Rohld

The University of Utah's subscription to the [Safari Technical Books Online](http://proquest.safaribooksonline.com.ezproxy.lib.utah.edu/) has a large selection of books that can be read online. For example.

* Introduction to Probability (2015) Blizstein & Hwang
* Probability and Statistics for Computer Scientists (2015) Baron
* Probability and Statistics (2015) Rukmangadachari & Reddy

# Learning Objectives:

1. By the end of this course, you will understand basic probability concepts, use probability rules, distinguish between discrete and continuous variables and the methods for their analysis, and solve problems related to random events.
2. By the end of the course, you will be able to describe the main methods for data exploration and analysis. You will use descriptive statistics to understand the nature of your data, and you will use hypothesis testing to predict the behavior of variables from samples and populations.
3. By the end of the course, you will understand the concepts of machine learning and pattern recognition, such as cluster recognition, data classification, dimensionality reduction, and temporal patterns.
4. By the end of the course, you will have experience with of the most widely use statistical software in the sciences, R. You will be able to use many of the statistical and graphical packages available for this software.

The skills we will be focusing on include:

* Identifying and explaining the characteristics of discrete and continuous variables.
* Distinguishing the various methods used to predict random events.
* Comparing inferential statistical methods, existing tools, and terminological resources.
* Proposing adequate statistical methods for specific clinical cases.
* Programming statistical analysis using R

# Course Schedule:

1. Introduction to the course
2. Working with R
3. Probability fundamentals
4. Exploratory Data Analysis
5. Bayes’ rule
6. Probability functions
7. Probability Distributions
8. Hypothesis testing
9. Visualization tools for hypothesis testing
10. Linear regression and correlation
11. Discrete estimation, contingency tables, exact test
12. Clustering
13. Boostrap
14. Survival Analysis
15. Bayesian inference
16. Information Theory
17. Introduction to Machine learning
18. Case studies