**Intro to Biostatistics: Data Science Methods and Applications in Public Health and Biomedicine**

### Basic information

|  |  |
| --- | --- |
| Course Title | Intro to Biostatistics |
| Instructor | Professor of [Preventive Medicine (Biostatistics)](https://www.preventivemedicine.northwestern.edu/index.html), [Feinberg School of Medicine](https://www.feinberg.northwestern.edu/), [Northwestern University](https://www.northwestern.edu/) |
| Prerequisites | Basic understanding of Probability and Linear Algebra |
| Required Text & Tools | All material will be provided in class in the form of PowerPoints.  Recommended textbook: Introduction to Biostatistics for Biomedical Research, by Frank Harrell and James Slaughter.  Programing language: R |
| Grading Criteria | TBD |
| Course Key Words | Biostatistics, data science, public health, biomedical and clinical research, R. |

### Course Description

The course will include introduction to background, development of a sequence of statistical techniques (background, methodology, application), their realizations using R (which is arguably the most popular open-source statistical software), as well as data analysis and reporting. Topics to be covered include but are not limited to: development of statistical data science, applications in public health and biomedicine, statistical analysis using R (a free statistical software), exploratory and summary statistics (concept and application), linear regression (concept, theory, and application), classification and clustering, time series analysis (concept, theory, and application), brief introduction of infectious disease modeling, other advanced topics (e.g. survival analysis, neural network, design of clinical trials), and statistical report writing.

### Schedule

|  |  |
| --- | --- |
| No. | Topics |
| Lecture 1 | Background and development of data science; applications in public health and biomedicine. |
| Lecture 2 | Applications in Public Health and Biomedicine |
| Lecture 3 | Exploratory Data Analysis and Summary Statistics:  why (the objectives of analysis), what (the list of analysis tasks), and how (analysis using excel and R) |
| Lecture 4 | Hypothesis Testing and Confidence Interval |
| Lecture 5 | Linear Regression:  why (the objectives), what (statistical concept, mathematical formulation), and how (estimation and inference) |
| Lecture 6 | Model comparison and evaluation |
| Lecture 7 | Time-varying; Time series regression analysis: why (the objectives), what (statistical concept, mathematical formulation), and how (estimation and inference) |
| Lecture 8 | Tree-Based Methods; classification tree, regression tree; tree trim; random forest |
| Lecture 9 | Classification and Clustering: why (the objectives), what (statistical concept, mathematical formulation), and how (estimation and inference) Classification: logistic regression Clustering: K-means and hierarchical clustering |
| Lecture 10 | Other Advanced Topics (e.g. survival analysis, neural network, design of clinical trials) |