

Math 425E (Introduction to Statistical Genomics)

Instructor: Nusrat Jahan

Lecture Topics:

Week 1

Population, Sample, Data collection methods
Descriptive statistics and visualizing data
Sampling techniques, Experimental designs

Week 2

Molecular genetics: DNA, RNA, Protein
High throughput -omic data
High throughput data pre-processing

Randomness and probability distributions (Normal & t)

Week 3

Sampling distributions, Central Limit Theorem
Inference: Estimation & Test of Hypothesis
Nonparametric tests

Week 4

Multiple testing problems
Family Wise Error Rate
False discovery rate, q value

Week 5

Accessing genomic data from public data repositories
Review R packages related to genomic data analysis
Data pre-processing

Week 6

Genomic data analysis:
 Identifying differential expressions
 Controlling for false discovery rate
 Detecting biological significance of the differential expressions

Week 7

Detecting biological significance of the differential expressions (from week 6)
Multivariate analysis:
 Principal component analysis
 Cluster analysis

Week 8

Discriminant analysis
Exam 1

Week 9 *Spring Break!***Week 10**

Gene selection and grouping using multivariate techniques
Heatmap

Week 11

Biological Networks

Gene regulatory networks

Metabolism networks, protein-protein interaction networks

Discrete probability distributions:

Binomial, Poisson, & Negative Binomial, & Multinomial

Week 12

RNA sequence data

Analysis: Goals & Objectives

Different RNA-seq analysis methods

Likelihood ratio test

Week 13

Sequencing mapping (for known genome)

SNP calling, Variation detection (for known genome)

De novo assembly (for unknown genome)

Gene transcription analysis (for RNA Seq)

Week 14

Discovery of novel splices & isoforms

Classification of high throughput-omic data using *R*:

Decision Tree

K-Nearest Neighbors

Penalized Logistic Regression

Week 15

Boosting

Bagging

Exam 2

Week 16 *Catching up with materials & Review!*