

Syllabus

Course Title

Statistics for Genomic Data Science

Course Instructor

Jeff Leek, Ph.D

Course Description

isn't enough to just know how to use the tools. Doing genomic data science well means understanding the statistical principles at work. This class will provide an introduction to the statistics behind the most popular genomic data science projects. This will help you ask better questions, plan better research, and interpret the results more accurately.

Course Content

Introduction

Welcome

What is statistics?

Finding statistics you can trust

Getting help

What is data?

Representing data

Week 1

Week 1 introduction

Reproducible Research

Achieving reproducible research

R markdown

The three tables in genomics

The three tables in genomics (in R)

Experimental Design: variability, replication, and power

Experimental Design: confounding and randomization

Exploratory Analysis

Exploratory Analysis in R

Data transforms

Clustering

Clustering in R

Week 2

Week 2 Introduction

Dimension reduction

Dimension reduction (in R)

Pre-processing and normalization

Quantile normalization (in R)

The linear model

Linear models with categorical covariates

Adjusting for covariates

Linear regression in R
Many regressions at once
Many regression in R
Batch effects and confounders
Batch effects in R

Week 3

Week 3 introduction
Logistic regression
Regression for counts
GLMs in R
Inference
Null and alternative hypotheses
Calculating statistics
Comparing models
Calculating statistics in R
Permutation
Permutation in R
P-values
Multiple testing
P-values and multiple testing in R

Week 4

Week 4 introduction
Gene set analysis
More enrichment
Gene set analysis in R
The process for RNA-seq
The process for Chip-Seq
The process for DNA methylation
The process for GWAS/WGS
Combining data types (eQTL)
eQTL in R
Researcher degrees of freedom
Inference vs. prediction
Knowing when to get help
Course Wrap-Up

Quiz Scoring

You may attempt each quiz up to 3 times in 8 hours. The score from your most successful attempt will count toward your grade.

Grading policy

You must receive a final grade of 70% or better on each assignment (quizzes and project) to pass the course.

Your final grade will be calculated as follows:

Quiz 1 = 25%

Quiz 2 = 25%

Quiz 3 = 25%

Quiz 4 = 25%

Differences of opinion

Keep in mind that currently data analysis is as much art as it is science - so we may have a difference of opinion - and that is ok! Please refrain from angry, sarcastic, or abusive comments on the message boards. Our goal is to create a supportive community that helps the learning of all students, from the most advanced to those who are just seeing this material for the first time.

Plagiarism

Johns Hopkins University defines plagiarism as "...taking for one's own use the words, ideas, concepts or data of another without proper attribution. Plagiarism includes both direct use or paraphrasing of the words, thoughts, or concepts of another without proper attribution." We take plagiarism very seriously, as does Johns Hopkins University.

We recognize that many students may not have a clear understanding of what plagiarism is or why it is wrong. Please see the following guide for more information on plagiarism:

🔗🔗🔗🔗🔗🔗

<http://www.jhsph.edu/academics/degree-programs/master-of-public-health/current-students/JHSPH-ReferencingHandbook.pdf> 🔗

It is critically important that you give people/sources credit when you use their words or ideas. If you do not give proper credit -- particularly when quoting directly from a source -- you violate the trust of your fellow students.

The Coursera Honor code includes an explicit statement about plagiarism:

I will register for only one account. My answers to homework, quizzes and exams will be my own work (except for assignments that explicitly permit collaboration). I will not make solutions to homework, quizzes or exams available to anyone else. This includes both solutions written by me, as well as any official solutions provided by the course staff. I will not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.

Go to next

item

