

Course titles in orange count for BBS credit

To search for all courses:

https://courses.my.harvard.edu/psp/courses/EMPLOYEE/EMPL/h/?tab=HU_CLA_SS_SEARCH

Biostatistics, Programming, and Computer Science
- BBS

2016

Biomedical Informatics (BMI) 713 / Genetics 229 Computational Statistics for Biomedical Sciences

Fall

Course Description: Analysis of large datasets has become an integral part of biological and biomedical sciences. This course will provide a practical introduction to data analysis, with high-throughput sequencing data as the main source of examples. In the first half, it will cover basic statistical concepts and techniques, including hypothesis testing, nonparametric methods, principal component analysis, correlation analysis, and linear regression. In the second half, it will cover several advanced topics, focusing on issues that one encounters in the literature but are seldom covered in introductory statistics courses. To carry out statistical tests and visualize data, students will learn R, a powerful programming language for statistical computing and graphics. The class will be a combination of lecture and computer lab. We will use recent literature to motivate the statistical methods, and assignments will frequently include attempts to reproduce published findings.

Prerequisites: No previous knowledge in statistics or programming is assumed. However, those with little or no programming experience may need to devote extra time. Additional sessions will also be provided for those interested in learning Python, a widely used programming language.

BST 282 Introduction to Computational Biology and Bioinformatics (formerly Bio 512)

Spring

Basic problems, technology platforms, algorithms and data analysis approaches in computational biology. Algorithms covered include dynamic programming, hidden Markov model, Gibbs sampler, clustering and classification methods.

This course is targeted at students with some statistics and computer programming background who have an interest in exploring genomic data analysis and algorithm development as a potential future direction.

Course restricted: Biostatistics students only (or instructor permission). If you are not a BIO student but took STAT110 and CS50 (FAS courses), please contact the Registrar's Office for an override.

STAT 110 Introduction to Probability

Fall

A comprehensive introduction to probability. Basics: sample spaces and events, conditional probability, and Bayes' Theorem. Univariate distributions: density functions, expectation and variance, Normal, t, Binomial, Negative Binomial, Poisson, Beta, and Gamma distributions. Multivariate distributions: joint and conditional distributions, independence, transformations, and Multivariate Normal. Limit laws: law of large numbers, central limit theorem. Markov chains: transition probabilities, stationary distributions, convergence.

STAT 111 Introduction to Theoretical Statistics

Spring

Basic concepts of statistical inference from frequentist and Bayesian perspectives. Topics include maximum likelihood methods, confidence and Bayesian interval estimation, hypothesis testing, least squares methods and categorical data analysis.

STAT 115/215 Introduction to Computational Biology and Bioinformatics

Spring

The course will cover basic technology platforms, data analysis problems and algorithms in computational biology. Topics include sequence alignment and search, high throughput experiments for gene expression, transcription factor binding and epigenetic profiling, motif finding, RNA/protein structure prediction, proteomics and genome-wide association studies. Computational algorithms covered include hidden Markov model, Gibbs sampler, clustering and classification methods.

Good quantitative skills, strong interest in biology, willingness and diligence to learn programming.

215 meets with 115 class, but graduate students are required to do more coding, complete a research project and submit a written report during reading period in addition to completing all work assigned for Statistics 115.

ES150 Introduction to Probability with Engineering Applications

Spring

This course introduces students to probability theory and statistics, and their applications to physical, biological and information systems. Topics include: random variables, distributions and densities, conditional expectations, Bayes' rules, laws of large numbers, central limit theorems, Markov chains, Bayesian statistical inferences and parameter estimations. The goal of this course is to prepare students with adequate knowledge of probability theory and statistical methods, which will be useful in the study of several

advanced undergraduate/graduate courses and in formulating and solving practical engineering problems.

BST 281 Genomic Data Manipulation

Spring

Introduction to genomic data, computational methods for interpreting these data, and a survey of current functional genomics research. Covers biological data processing, programming for large datasets, high-throughput data (sequencing, proteomics, expression, etc.), and related publications.

This course is targeted at students in experimental biology programs with an interest in understanding how available genomic techniques and resources can be applied in their research.

BST 210 Applied Regression Analysis

Fall

Topics include model interpretation, model building, and model assessment for linear regression with continuous outcomes, logistic regression with binary outcomes, and proportional hazards regression with survival time outcomes. Specific topics include regression diagnostics, confounding and effect modification, goodness of fit, data transformations, splines and additive models, ordinal, multinomial, and conditional logistic regression, generalized linear models, overdispersion, Poisson regression for rate outcomes, hazard functions, and missing data. The course will provide students with the skills necessary to perform regression analyses and to critically interpret statistical issues related to regression applications in the public health literature.

Math 19B Linear Algebra, Probability and Statistics for the Life Sciences

Spring

Probability, statistics and linear algebra with applications to life sciences, chemistry, and environmental life sciences. Linear algebra includes matrices, eigenvalues, eigenvectors, determinants, and applications to probability, statistics, dynamical systems. Basic probability and statistics are introduced, as are standard models, techniques, and their uses including the central limit theorem, Markov chains, curve fitting, regression, and pattern analysis.

Math 21B Linear Algebra and Differential Equations

Spring

Matrices provide the algebraic structure for solving myriad problems across the sciences. We study matrices and related topics such as linear transformations and linear spaces, determinants, eigenvalues, and eigen vectors. Applications include dynamical systems, ordinary and partial differential equations, and an introduction to Fourier series.

MIT 6.047/878 Computational Biology

Fall

Covers the algorithmic and machine learning foundations of computational biology combining theory with practice. We cover both foundational topics in computational biology, and current research frontiers. We study fundamental techniques, recent advances in the field, and work directly with current large-scale biological datasets.

Genomes:

Biological sequence analysis, hidden Markov models, gene finding, comparative genomics, RNA structure, sequence alignment, hashing

Networks:

Gene expression, clustering/classification, EM/Gibbs sampling, motifs, Bayesian networks, microRNAs, regulatory genomics, epigenomics

Evolution:

Gene/species trees, phylogenomics, coalescent, personal genomics, population genomics, human ancestry, recent selection, disease mapping

In addition to the technical material in the course, the term project provides practical experience: (1) writing an NIH-style research proposal, (2) reviewing peer proposals, (3) planning and carrying out independent research, (4) presenting research results orally in a conference setting, and (5) writing results in a journal-style scientific paper. You will work on a project of your choice with regular feedback and advice from a mentor, your peers, and the teaching staff.

Stat 121a Data Science

Fall

Learning from data in order to gain useful predictions and insights. This course introduces methods for five key facets of an investigation: data wrangling, cleaning, and sampling to get a suitable data set; data management to be able to access big data quickly and reliably; exploratory data analysis to generate hypotheses and intuition; prediction based on statistical methods such as regression and classification; and communication of results through visualization, stories, and interpretable summaries. Built around three modules: prediction and elections, recommendation and business analytics, and clustering and text analysis.

Stat 139 Statistical Sleuthing through Linear Models

Spring

A serious introduction to statistical inference with linear models and related methods. Topics include t-tools and nonparametric alternatives (including bootstrapping and permutation-based methods), multiple-group comparisons, analysis of variance, linear regression, model checking and refinement, and causation versus correlation. Emphasis on thinking statistically, evaluating assumptions, and developing tools for real-life applications.

Stat 149 Statistical Sleuthing through Generalized Linear Models

Fall/Spring

Sequel to Statistics 139, emphasizing common methods for analyzing continuous non-normal and categorical data. Topics include logistic regression, log-linear models, multinomial logit models, proportional odds models for ordinal data, Gamma and inverse-Gaussian models, over-dispersion, analysis of deviance, model selection and criticism, model diagnostics, and an introduction to non-parametric regression methods.

SCRB 152 Asking Cells Who They Are: Computational Transcriptomics Using RNA-Seq

Fall

This course is a hands-on introduction to computational analysis of RNA sequencing data as a measure of genome-wide transcription. We will cover methods spanning the spectrum of RNA-Seq analysis: starting from raw sequencing reads, obtaining gene expression measures, and interpreting biological significance by differential expression analyses, clustering, and visualization. Coursework will consist of programming assignments in Python exploring real datasets. The course will emphasize skills applicable to independent biological research.

Systems Biology 200 Dynamic and Stochastic Processes in Cells

Fall

Rigorous introduction to (i) dynamical systems theory as a tool to understand molecular and cellular biology (ii) stochastic processes in single cells, using tools from statistical physics and information theory.

HST 508/Biophysics 170

Fall

This course provides a foundation in the following four areas: evolutionary and population genetics; comparative genomics; structural genomics and proteomics; and functional genomics and regulation.

SEAS AC209a Intro to Data Science

<http://cs109.github.io/2015/>

SEAS AC209b Advanced topics in Data Science

<http://cs109.github.io/2015/>

CS 50 Introduction to Computer Science

<https://cs50.harvard.edu/>

MIT 6.00 Introduction to Computer Science and Programming

Fall

<http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-00sc-introduction-to-computer-science-and-programming-spring-2011/>

Cell Biology 302qc. Advanced Experimental Design for Biologists

Neurobiology 206qc. Bootcamp in Quantitative Methods

<http://springerlab.org/qmbc/>

Genetics 303qc. Current Tools for Gene Analysis

<http://bioinformatics.hms.harvard.edu/>

Online:

Harvard Chan Bioinformatics Core Workshops

<http://bioinformatics.hms.harvard.edu/>

HarvardX/EdX Statistics and R for Life Sciences

Introduction to Statistics

Free online course from udacity.com

Course does not have a start date; students start class on their own time

Course is self-paced

<https://www.udacity.com/course/st101>

Mathematical Biostatistics

<https://www.coursera.org/learn/biostatistics>

7 week course, 3-5 hours a week of work

Course includes use of R statistical programming language

(This course was taken by multiple students from the previous year)

Hopkins – JHUSPH Open Courseware: Introduction to Biostatistics

4 week, 10 lecture series with practice problem sets associated with each lecture

<http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/IntroBiostats/coursePage/schedule/>

Statistics: Making Sense of Data Free online course from Coursera Pending start dates
<https://www.coursera.org/course/introstats>

Computing for Data Analysis

4 weeks, 3-5 hours a week of work Course is free
<https://www.coursera.org/course/compdat>

Learn to Program: The Fundamentals

No courses currently planned, but future sessions can be added to a watch list <https://www.coursera.org/course/programming1>

Learn to Program: Crafting Quality Code

Focus on writing quality Python code that runs correctly and efficiently. No courses currently planned, but future sessions can be added to a watch list <https://www.coursera.org/course/programming2>

Introduction to Computer Science - Programming Methodology

Stanford
12 weeks, 2 hours a week of work
<http://see.stanford.edu/see/courseinfo.aspx?coll=824a47e1-135f-4508-a5aa-866adcae1111>

Khan Academy Linear Algebra

<https://www.khanacademy.org/math/linear-algebra>

Institute for Quantitative Social Science

Courses in R and Python
<http://www.iq.harvard.edu/data-science-resources>

Python:

MIT 6.0001 Introduction to Computer Science and Programming using Python

https://courses.edx.org/courses/course-v1:MITx+6.00.1x_7+3T2015/info

MIT 6.0002 Introduction to Computational Thinking and Data Science

<https://www.eecs.mit.edu/academics-admissions/academic-information/subject-updates-ft-2014/600022>

SEAS January ComputeFest Workshop

Free workshops
Covers basics of computer science, R as well as Python
<http://computefest.seas.harvard.edu/workshops-2015>

Intro to Computer Science

Free course & paid course available

Covers basics of computer science as well as Python

<https://www.udacity.com/course/cs101>

Google's Python Class

All this material makes up an intensive 2-day class

The videos are organized as the day-1 and day-2 sections Class is free

<https://developers.google.com/edu/python/>

The Hitchhiker's Guide to Python

<https://python-guide.readthedocs.org/en/latest/>

Python for Data Analysis

Available for purchase on Amazon, \$23.99

<http://www.amazon.com/Python-Data-Analysis-Wes-McKinney/dp/1449319793>

Codecademy

Learn the fundamentals of programming to build web apps and manipulate data Course is free

<http://www.codecademy.com/tracks/python>

Dataquest Python

Online python courseware

<https://www.dataquest.io/subject/learning-python?>

R-Programming:

Intro to Data Science with R

Offers a 2 day or 3 day course options

<https://www.rstudio.com/training/curriculum/intro-to-data-science.htm>

Roger Peng, Introducing R and basic programming concepts. Computing for Data Analysis

Week 1: <http://www.youtube.com/playlist?list=PLjTlxb-wKvXNSDfcKPFH2gzHGyjpeCZmI>

Week 2: <http://www.youtube.com/playlist?list=PLjTlxb-wKvXNnjUTX4C8IelhPBjPkng6B>

Week 3: http://www.youtube.com/playlist?list=PLjTlxb-wKvXOzl2h0F2_rYZHIXz8GWBop

Week 4:

http://www.youtube.com/playlist?list=PLjTlxbwKvXOdzysAE6qrEBN_aSBC0LZS

John Hopkins

4 weeks, 3-5 hours a week of work

Class is free

<https://www.coursera.org/course/rprog>

The R Book by Michael Crawly

Available for purchase on Amazon, \$73.72 + free shipping

<http://www.amazon.com/The-Book-Michael-J-Crawley/dp/0470973927>

The Art of R Programming by Norman Matloff

Available for purchase on Amazon, \$25.35 + free shipping

<http://www.amazon.com/The-Art-Programming-Statistical-Software/dp/1593273843>

Google's R Style Guide

<http://google-styleguide.googlecode.com/svn/trunk/Rguide.xml>

R Programming

<http://dist.stat.tamu.edu/pub/rvideos/>

R Programming – Research Technology Consulting

<http://projects.iq.harvard.edu/rtc/r-prog>

Codeschool

Course is free

Highly recommended by past BIRT fellows

<https://www.codeschool.com/courses/try-r>

Coursera Introduction to R

<https://www.coursera.org/learn/r-programming>

Database:

Introduction to Databases

Introduction to Databases is being launched on the new edX-based platform in June, but can still be accessed through the link provided

<http://class2go.stanford.edu/>

SQL:

SQL Tutorial

SQL tutorial will teach you how to use SQL to access and manipulate data in: MySQL, SQL Server, Access, Oracle, Sybase, DB2, and other database systems

<http://www.w3schools.com/sql/default.asp>

MySQL Crash Course

http://www.amazon.com/MySQL-Crash-Course-Ben-Forta/dp/0672327120/ref=sr_1_3?ie=UTF8&qid=1343427063&sr=8-3&keywords=mysql Available for purchase on Amazon, \$24.66 + free shipping