Fall 2020 Course Offerings

Undergraduate Stat Courses	Course Title	Instructors	Prerequisite	Course Explorer Description	Instructor Description
STAT 100	Statistics	K. Flanagan	Math 112	First course in probability and statistics at a precalculus level; emphasizes basic concepts, including descriptive statistics, elementary probability, estimation, and hypothesis testing in both nonparametric and normal models.	Stat 100 is a first course in statistics and probability at a precalculus level designed for non-STEM related majors. This course emphasizes basic concepts, including experimental design, descriptive statistics, elementary probability, 2 variable data, sampling and estimating, and hypothesis testing.
STAT 107	Data Science Discovery	W. Fagen K. Flanagan		Data Science Discovery is the intersection of statistics, computation, and real-world relevance. As a project-driven course, students perform hands-on-analysis of real-world datasets to analyze and discover the impact of the data. Throughout each experience, students reflect on the social issues surrounding data analysis such as privacy and design.	Flanagan: Data Science Discovery is the intersection of statistics, computation, and real-world relevance. As a project-driven course, students perform hands-on-analysis of real-world datasets to analyze and discover the impact of the data. Throughout each experience, students reflect on the social issues surrounding data analysis such as privacy and design.
STAT 200	Statistical Analysis	E. Fireman	One year of calculus experience	Survey of statistical concepts, data analysis, designed and observational studies and statistical models. Statistical computing using a statistical package such as R or a spreadsheet. Topics to be covered include data summary and visualization, study design, elementary probability, categorical data, comparative experiments, multiple linear regression, analysis of variance, statistical inferences and model diagnostics. May be taken as a first statistics course for quantitatively oriented students, or as a second course to follow a basic concepts course.	Most people think statistics is boring and difficult. Statistics is to data what grammar is to words. And like grammar, it's only interesting if it's used to understand something interesting. In Stat 200, we use statistics to research a topic we're all interested in - ourselves. We collect data on ourselves through anonymous surveys, largely on the sort of social questions on which students have shown intense interest. Having real questions that we want to answer motivates real understanding, not just memorizing some complicated rules. Statistics is a collection of real tools- the key is to understand which one to use when and why.
STAT 207 (Currently DS section of 200)	Data Science Exploration	D. Simpson		The DS Section of Stat 200 (Future Stat 207) is an accelerated introductory course in statistics. This section will use Python for data exploration and analysis	This course surveys basic statistical concepts, statistical modeling and train/test analytics to lay a foundation for further studies in statistics and data science. Concepts are developed in context of real data and simulation studies. We employ computational tools of data science such as Python programming with Jupyter notebooks and GitHub, to enable students to gain understanding of data workflows and reproducible analysis. Topics covered include preprocessing and checking for missing data, data summary and visualization, random sampling and probability models, estimating parameters, computing standard errors and confidence intervals, hypothesis testing, multiple linear and logistic regression modeling, classification and train/test evaluation of sensitive and specificity.
STAT 212	Biostatistics	K. Findley		Application of statistical reasoning and statistical methodology to biology. Topics include descriptive statistics, graphical methods, experimental design, probability, statistical inference and regression. In addition, techniques of statistical computing are covered.	This course will provide a serious introduction to statistics for biology and biomedical applications. The basic statistics will follow a similar structure as in Stat 200, including a major emphasis on experimental design and causation, along with customary topics such as descriptive statistics, probability, random variables, and sampling and inference (significance tests, power, confidence intervals) and data visualization and statistical computing in R. There will be somewhat less mathematical depth on general linear models (simple and multiple regression, confidence intervals for coefficients, F and chi-sq tests, ANOVA, randomization tests for p-values ...) and advanced methods (transformation of variables, logistic regression, non-parametric tests) . Instead, we will introduce techniques needed for readers of biomedical papers, e.g. Kaplan-Meier survival plots, hazard ratios, and Cox proportional hazard models.
STAT 385	Statistics Programming Methods	C. Kinson	Stat 200 or 212; Knowledge of R	Statisticians must be savvy in programming methods useful to the wide variety of analysis that they will be expected to perform. This course provides the foundation for writing and packaging statistical algorithms through the creation of functions and object oriented programming. Fundamental programming techniques and considerations will be emphasized. Students will also create dynamic reports that encapsulate their implemented algorithms. Students must have access to a computer on which they can install software.	This course is a serious programming course for students who are fairly new to R. Its intended audience is students who recently completed STAT 200/212 but not for students who have strong knowledge and familiarity in R. We cover everything from a object-oriented design beginning with simple objects such as vectors and matrices. Then we build on what we can do with those objects, such as matrix algebra, loop structures, and applying functions. Then we move onto more complicated structures, such as data frames, tibbles, functions, and expressions. At the same we cover some aspects of statistics such as calculating certain point estimates, data visualization, and creating algorithms. We do all of this with a reproducible documentation sometimes called notebooks (Rmarkdown). The final project might be to apply course concepts into an interactive web-based application (Shiny).
STAT 400	Statistics and Probability I	A. Yu	Math 241	Introduction to mathematical statistics that develops probability as needed; includes the calculus of probability, random variables, expectation, distribution functions, central limit theorem, point estimation, confidence intervals, and hypothesis testing. Offers a basic one-term introduction to statistics and also prepares students for STAT 410.	Stat 400 starts with probability and uses it to introduce some fundamental concepts in statistics such as estimation and testing. The course helps students develop a good intuition for distributions and random variables, which are ideas that will be used in all of the higher level courses. Non-majors who finish this course should be able to talk about probability and statistics in their fields with more mathematically precise language. Students should be fairly comfortable with Calculus 3 and algebra before taking this course.
STAT 410	Statistics and Probability II	A. Stepanov S. Chatterjee	Stat 400	Continuation of STAT 400. Includes moment-generating functions, transformations of random variables, normal sampling theory, sufficiency, best estimators, maximum likelihood estimators, confidence intervals, most powerful tests, unbiased tests, and chi-square tests.	
STAT 420	Methods of Applied Statistics	U. Ravat M. Wang	Stat 400	Systematic, calculus-based coverage of the more widely used methods of applied statistics, including simple and multiple regression, correlation, analysis of variance and covariance, multiple comparisons, goodness of fit tests, contingency tables, nonparametric procedures, and power of tests; emphasizes when and why various tests are appropriate and how they are used.	Ravat: In this course we focus on modeling using linear models. We cover analysing data using linear regression covering simple and multiple regression, anova, logistc regression including methods for fitting models, verifying assumptions and assessing various model diagnostics and selecting the 'best' model. We also cover briefly some extensions of linear models but those are not the focus of this course.

Fall 2020 Course Offerings

Undergraduate Stat Courses	Course Title	Instructors	Prerequisite	Course Explorer Description	Instructor Description
STAT 425	Applied Regression and Design	X. Shao N. Narisetty L. Bravo	Stat 410	Explores linear regression, least squares estimates, F-tests, analysis of residuals, regression diagnostics, transformations, model building, factorial designs, randomized complete block designs, Latin squares, split plot designs. Computer work is an integral part of the course.	**Xiaofeng**: Requires knowledge of statistical tools and concepts, as provided by STAT 410. Basic background in linear algebra is also beneficial. Covers both linear regression and basic design and analysis of experiments. Most of the methods covered are fundamental to classical data analysis. Other topics frequently covered are mixed models, binary regression, and Poisson regression. Traditional lecture course, with grading based on homework and exams. Implementation of methods using the R statistical language.
STAT 426 (Pending)	Sampling and Categorical Data		Stat 410	Sampling: simple random, stratified, systematic, cluster, and multi-stage sampling. Categorical data: multiway contingency tables, maximum likelihood estimation, goodness-of-fit tests, model selection, logistic regression. Computer work is an integral part of the course.	
STAT 428	Statistical Computing	U. Ravat	Stat 410; and knowledge of a programming language	Examines statistical packages, numerical analysis for linear and nonlinear models, graphics, and random number generation and Monte Carlo methods.	In addition to what is listed in the course description, this course will also develop students' skills in Monte Carlo simulations and designing algorithms to use computational power to do statistics. That is, we learn how we can use computational methods for model estimation where mathematical methods may fall short. We write programs in R to develop these simulation skills.
STAT 429	Time Series Analysis	H. Lee X. Chen	Stat 410	Studies theory and data analysis for time series; examines auto-regressive moving average model building and statistical techniques; and discusses spectral model building and statistical analysis using windowed periodograms and Fast Fourier Transformations.	**Lee**: In addition to what is listed in the course description, real life problems and examples will be used to demonstrate the methods for time series analysis and forecasting. An introduction to the R Software and R tools for time series analysis will be an important part of the class.
STAT 430 (IS 407)	Introduction to Data Science	J. He	Junior/Senior standing per IS	This course introduces students to data science approaches that have emerged from recent advances in programming and computing technology. They will learn to collect and use data from a variety of sources, including the web, in a modern statistical inference and visualization paradigm. The course will be based in the programming language R, but will also use HTML, regular expressions, basic unix tools, XML, and SQL. Supervised and unsupervised statistical learning techniques made possible by recent advances in computing power will also be covered.	
Stat 430	Nonparametric Statistics				This course considers nonparametric methods of statistical analysis. Topics include smoothing and spline methods for estimation of probability density and regression functions, as well as resampling techniques for inference. Prerequisites: STAT 410 and STAT 425.
STAT 430	Data Science Programming Methods	D. Eddelbuettel	Stat 410, Stat 420, and Stat 425 or consent of instructor.	This course provides the computational foundation for rigorous data science work, both applied and in research. Starting from key foundations (the shell, git, Markdown and SQL), we focus on a solid introduction to programming in R. Next we discuss keys to reproducible computing (R packages, Docker) as well as some computational and algorithmic foundations. Finally, we examine in some detail extensions for better performance, notably using C++ with R.	Statistics and Data Science are focused on making sense of data---and face an ever-increasing demand for their work. At the same time, data sets increase in size, scope, and complexity. Proper tooling is essential to meet these challenges, and as applied work in data analysis is in effect applied computational work, we will learn the computational tools and programming methods to meet these data science---and data engineering-challenges. This goal of this class is nothing less that to make you proficient at the Unix shell, familiar with git for version control, give you a basic understanding of SQL for interacting with data bases, and first and foremost lets you gain actual expertise in R programming (as well as in communicating via (R)Markdown, visualizations and dashboards). We will use RStudio Cloud instances so students are not required to install and maintain all required components. You can expect actual hands-on programming exercises. Prior programming experience (in R or another language) will be helpful, but is not a requirement for taking the course.
STAT 431	Applied Bayesian Analysis	S. Culpepper	Stat 410; knowledge of R	Introduction to the concepts and methodology of Bayesian statistics, for students with fundamental knowledge of mathematical statistics. Topics include Bayes' rule, prior and posterior distributions, conjugacy, Bayesian point estimates and intervals, Bayesian hypothesis testing, noninformative priors, practical Markov chain Monte Carlo, hierarchical models and model graphs, and more advanced topics as time permits. Implementations in R and specialized simulation software.	
STAT 432	Basics of Statistical Learning	D. Dalpiaz	Stat 410, and one of Stat 420 or Stat 425	Topics in supervised and unsupervised learning are covered, including logistic regression, support vector machines, classification trees and nonparametric regression. Model building and feature selection are discussed for these techniques, with a focus on regularization methods, such as lasso and ridge regression, as well as methods for model selection and assessment using cross validation. Cluster analysis and principal components analysis are introduced as examples of unsupervised learning.	STAT 432 provides a broad overview of machine learning, through the eyes of a statistician. As a first course in machine learning, core ideas are stressed, and specific details are de-emphasized. After completing the course, students should be able to train and evaluate statistical models. While we will not discuss an exhaust list of methods, given the framework developed throughout the course, students should feel comfortable exploring new methods and models on their own. Previous experience with R programming is necessary for success in the course as students will be tested on their ability to use the methods discussed through the use of a statistical computing environment.
STAT 433	Stochastic Processes	A. Stepanov	Stat 400; Stat 410 preferred; Math 225 (or equivalent) is highly recommended	A stochastic process is a random process that represents the evolution of some system over time. Topics may include discrete-time and continuous-time Markov chains, birth-and-death chains, branching chains, stationary distributions, random walks, Markov pure jump processes, birth-and-death processes, renewal processes, Poisson process, queues, second order processes, Brownian motion (Wiener process), and Ito's lemma.	

Fall 2020 Course Offerings

Undergraduate Stat Courses	Course Title	Instructors	Prerequisite	Course Explorer Description	Instructor Description
STAT 434	Survival Analysis	S. Zhao	Stat 410; Stat 420; knowledge of R	Introduction to the analysis of time-to-event outcomes. Topics center around three main procedures: the Kaplan-Meier estimator, the log-rank test, and Cox regression. Emphasis on big-picture concepts, basic methodological understanding, and practical implementation in R.	In addition to the three main topics addressed in the catalog description, we will also cover clinical trials design. Students who finish this course should be able to conduct basic survival analysis procedures commonly used in industry, in particular, at large pharmaceutical companies. The course is fairly conceptual and mathematical but will also discuss implementation in R.
STAT 440	Statistical Data Management	D. Unger	Stat 400 or Stat 409	The critical elements of data storage, data cleaning, and data extractions that ultimately lead to data analysis are presented. Includes basic theory and methods of databases, auditing and querying databases, as well as data management and data preparation using standard large-scale statistical software. Students will gain competency in the skills required in storing, cleaning, and managing data, all of which are required prior to data analysis.	In most statistics courses, your data is ready to go. Open the textbook, and there it is -- sitting on a silver platter for you. Now, calculate this. Model that. The data is already clean so that you can focus on the methods of statistical analysis. But in reality, there can be a lot to do between collecting your data and having it ready for all the cool modeling techniques we use as data analysts. Read the official course description. Though in many more words, it basically alludes to a two-pronged set of goals for this course: 1. Gain experience utilizing strategies and methods for preparing and managing data from collection to statistical analysis. 2. Gain experience using SAS and SQL (the "large-scale software") for that purpose.
STAT 443	Professional Statistics	D. Glosemeyer	Stat 420 or consent of instructor	This project-based course emphasizes written, visual, and oral communication of statistical results and conclusions. An introduction to statistical consulting is also provided. Additional topics include introductions to statistical methodologies in industry and aspects of careers in statistics.	This project-based course emphasizes written, visual, and oral communication of statistical results and conclusions. An introduction to statistical consulting is also provided. Additional topics include introductions to statistical methodologies in industry and aspects of careers in statistics. This course should provide students with the skills needed to effectively communicate results of real world statistical analyses via hands-on group projects, presentation of project results, and critical evaluation of data-analytic presentations given by others in and outside of the classroom.
STAT 448	Advanced Data Analysis	L. Bravo C. Kinson	Stat 400 or Stat 409; credit or concurrent enrollment with Stat 410	Several of the most widely used techniques of data analysis are discussed with an emphasis on statistical computing. Topics include linear regression, analysis of variance, generalized linear models, and analysis of categorical data. In addition, an introduction to data mining is provided considering classification, model building, decision trees, and cluster analysis.	**Bravo**: The use of an appropriate data analysis technique requires knowledge about a wide spectrum of sound statistical methods. The purpose of this class is to introduce the student to the main statistical inference techniques and modelling approaches for several types of data variables, including categorical data, experimental design and multivariate data. The topics in the course include applications of descriptive statistics and visualization, hypothesis tests for population locations and distributional goodness of fit tests, basic introduction to categorical data analysis, ANOVA (balanced and unbalanced), simple and multiple linear regression, logistic regression, generalized linear models, PCA, hierarchical cluster analysis, and discriminant analysis (LDA and QDA). The class setting uses a computational environment for the use of statistical packages.