

Viewing: **STAT 207 : Data Science Exploration**

Last approved: 04/14/21 3:38 am

Last edit: 04/13/21 11:09 am

Catalog Pages

referencing this course

[Business + Data Science, BS](#)

[CS - Computer Science](#)

[Finance + Data Science, BS](#)

[IS - Information Sciences](#)

[School of Information Sciences](#)

Programs referencing

this course

[0329: Statistics Minor, UG](#)

[10KM6076BS: Finance + Data Science, BS](#)

[10KM6077BS: Accountancy + Data Science, BS](#)

[10KM6146BS: Business + Data Science, BS](#)

History

1. Oct 13, 2020 by Joseph Zarnsy (zarnsy)
2. Dec 8, 2020 by Beth McKown (bmckown1)
3. Apr 14, 2021 by Deb Forgacs (dforgacs)

General Information

Effective Term:

College: Liberal Arts & Sciences

Department/Unit: Statistics (1583)

Name (ORG Code):

Course Subject: Statistics (STAT)

Course Number: 207

Course Title:

Abbreviated Title: Data Science Exploration

Course

Description:

Explores the data science pipeline from hypothesis formulation, to data collection and management, to analysis and reporting. Topics include data collection, preprocessing and checking for missing data, data summary and visualization, random sampling and probability models, estimating parameters, uncertainty quantification, hypothesis testing, multiple linear and logistic regression modeling, classification, and machine

learning approaches for high dimensional data analysis. Students will learn how to implement the methods using Python programming and Git version control.

Justification

Justification for change:

Please Note: a syllabus is required for General Education review:

Course Information

Course Credit

Course credit:

Undergraduate: 4

Graduate:

Professional:

Registrar Use Only:

Banner Credit: 0 OR 4

Billable Hours: 0 OR 4

Grading Type

Grading type: Letter Grade

Alternate Grading Type (optional):

Available for DFR: No

Repeatability

May this course be repeated? No

Credit Restrictions

Credit

Restrictions:

Advisory Statements

Prerequisites:

STAT 107 or consent of instructor.

Concurrent

Enrollment

Statement:

Restricted

Audience

Statement:

Cross-listing

Cross Listed

Courses:

Class Schedule Information

Class Schedule

Information:

Fees

Is a fee requested No
for this course?

Course Description in the Catalog Entry

This is how the above information will be represented in the Catalog:

Explores the data science pipeline from hypothesis formulation, to data collection and management, to analysis and reporting. Topics include data collection, preprocessing and checking for missing data, data summary and visualization, random sampling and probability models, estimating parameters, uncertainty quantification, hypothesis testing, multiple linear and logistic regression modeling, classification, and machine learning approaches for high dimensional data analysis. Students will learn how to implement the methods using Python programming and Git version control. Course Information: Prerequisite: STAT 107 or consent of instructor.

Additional Course Notes

Enter any other
course
information

details to be included in the catalog:

Course Detail

Frequency of course:

Every Fall
Every Spring

Duration of the course Full

Anticipated Enrollment: 300

Expected distribution of student registration:	Freshman: 50 %	Sophomore: 40 %	Junior: 10 %	Senior: 0 %
--	----------------	-----------------	--------------	-------------

General Education

General Education Category Quantitative Reasoning II

General Information

Is the course required for a major concentration? No

Is the course part of a sequence? No

What is the frequency with which the course will be offered?:
(For Example: every semester, once a year)

Both fall and spring semester annually

Briefly describe how the course fulfills the General Education objectives:

This course is intended for freshman and sophomores to give them exposure to the concepts and methods of data science including how to find or collect relevant data, understanding the information that data can convey, visualizing and analyzing the data, evaluating hypothesis informed by the data and drawing conclusions supported by the data. The 107/207 course sequence is intended to be accessible to all students at the university in that no prerequisites to the sequence are required, and no prior programming experience is assumed, and mathematical ideas will be introduced as needed in the context of data processing and analysis.

Scholarship on women and gender issues will be integrated through discussion of the potential for and mitigation of gender bias in questionnaires and machine learning training algorithms. The impact of women in data science will be modeled by inclusion of women authors of technical works to the extent that these works are accessible as case studies for a 200-level audience of students. These topics are in Week 14, Lectures 36-37 of the syllabus.

Describe the instructional format and provide special justification, if necessary:

The instructional format is 3 hours lecture/discussion and 2 hours lab/discussion. Students also use online platforms such as Piazza to discuss questions raised by the course content and assignments.

Describe the means by which the Communication Skills goal will be achieved:

Students are encouraged to discuss together the lab homework assignments during weekly lab meetings. Weekly assignments are organized around dynamic documents that the students create and submit. These documents (e.g. Jupyter notebooks) include all analytics and descriptions of their methods and conclusions. Projects are more extensive reports requiring team cooperation, development and execution of a project plan and communication of the outcomes of the study via written and oral reports.

Describe how evaluation and adherence to General Education guidelines will be monitored:

Please indicate the timeline for such evaluations

The department curriculum committee annually reviews course syllabus and enrollment information. The department will ensure that sufficient lab sections are available with teaching assistant support and periodically review course products and student course evaluations. With an estimated 300 enrolled per semester, and 10 lab sections of 30 each, the instructors are expected to include 2 faculty plus 4 teaching assistants for lectures plus lab/discussion sections.

Indicate those who will teach the course and describe procedures for training & supervising teaching assistants:

The course will be led by faculty members such as Douglas Simpson, Kelly Findley, and Victoria Ellison. Teaching assistants for the course work closely with the faculty to develop lab exercises and discussion experiences. Teaching assistants also receive specific departmental training in teaching statistics and data science.

Quantitative Reasoning II

Which type of course is this?

Probability or Statistics

Show how at least 25 percent of the course content and graded material incorporates the relationships between the assumptions of the probabilistic and statistical models discussed and the conclusions drawn?

Many real data examples and case studies are used throughout the course to help students practice and gain an understanding of the role and limitations of modeling assumptions. They also learn how to use computational tools such as Monte Carlo simulation and cross-validation to evaluate the quality of statistical models and analytics.

The course integrates probability with sampling distributions and variation in data. Conclusions drawn are based on the probability foundations of confidence intervals, hypothesis tests and train/test resampling methods for predictive model assessment.

Additional Course Information

Does this course replace an existing course? No

Does this course impact other courses? No

Does the addition of this course impact the departmental curriculum? Yes

Specify the curriculum and explain:

STAT 207 will be required in forthcoming X + Data Science proposals, including one from Statistics, and will be a cognate to STAT 200 in the Statistics major and Statistics and Computer Science major.

Has this course been offered as a special topics or other type of experimental course? Yes

Please indicate the
Banner subject,
course number,
section ID, term
and enrollment for
each offering:

Some of the approaches in the course were tested in selected
sections of STAT 200, with Python replacing R as the
programming language.

STAT 200, Section DS

Fall 2019: 58 enrolled

Spring 2020: 40 enrolled

(Enrollment data from DMI as of 12/13/2019)

Will this course be
offered on-line?

Face-to-Face

Faculty members
who will teach
this course:

Victoria Ellison Kelly Findley Douglas Simpson

Course ID: 1012153

Comments to
Reviewers:

Form to request Gen Ed credit (QR2) attached. Crosslist request removed.

Course Edits
Proposed by:

Key: 11845