MAJOR IN STATISTICS

The major in Statistics is designed to meet three goals: to introduce some of the central ideas of information and data science, to develop problem-solving ability by teaching students to combine creative thinking with rigorous reasoning, and to develop interdisciplinary skills by applying statistics to an application area of interest.

REQUIREMENTS (10 courses plus colloquium)

The major in Statistics consists of ten courses plus a colloquium requirement. The major includes courses in mathematics, computer science and statistics.

Mathematics (2 courses)

MATH 150 or 151 Multivariable Calculus or equivalent high school course
MATH 250 Linear Algebra

Except in unusual circumstances, students planning to major in statistics should complete the calculus sequence (MATH 130, 140, 150/151) before the spring of the sophomore year, at the latest. MATH 150 is a prerequisite for STAT 201 and MATH 250 is a prerequisite for STAT 346.

Computer Science (1 course)

CSCI 134 Intro to Computer Science or CSCI 135 Diving into the Deluge of Data or CSCI 136 Data Structures and Advanced Programming or some other course in the Computer Science Department with prior approval of the Math/Stat department.

Core Courses (4 courses)

STAT 201 Statistics and Data Analysis or STAT 202 Introduction to Statistical Modeling
STAT 346 Regression and Forecasting
STAT 341 Probability
STAT 360 Inferential Statistics

Continuation (2 courses)

Any two courses among the 300 or 400 level courses in the department with a STAT prefix.

Capstone Course (1 course)

The capstone course is a 400-level STAT course taken in the senior year. Although the specific methodological emphasis of the course may vary from year to year, an in-depth project with both a written report and an oral presentation is typically part of the capstone course.

Colloquium Requirement

Participation in the Department Colloquium series, in which all senior majors present and attend talks on statistical topics of their choice. Majors have to attend at least 20 colloquia in their senior year and present one themselves.

PLACEMENT

Students with an AP Stat score of 5 or 4 are placed in the advanced introductory course Stat 202.

NOTES

Substitutions, Study Abroad, and Transfer Credit: In some cases, and with prior permission of the Mathematics and Statistics Department,
appropriate courses from other institutions may be substituted for the application and continuation requirements, but at least eight courses must be taken from the Department of Mathematics and Statistics at Williams.

These can, with prior permission, include courses taken away. Students with transfer credit should contact the department about special arrangements.

**Double Counting:** No course may count towards two different majors.

**Early Senior Capstone Course:** In exceptional circumstances, with the prior permission of the department, a student may be allowed to satisfy the Senior Capstone Course requirement in the junior year, provided that the student has completed at least three 300-level statistics courses before enrolling in the capstone course.

**Planning Courses:** Core courses are normally offered every year. Other 300 and 400 level statistics courses are offered on an irregular basis. Students should check with the department before planning far into the future.

**Course Admission:** Courses are normally open to all students meeting the prerequisites, subject to any course caps. Students with questions about the level at which courses are conducted are invited to consult department faculty.

**FAQ**

Students MUST contact departments/programs BEFORE assuming study away credit will be granted toward the major or concentration.

*Can your department or program typically pre-approve courses for major/concentration credit?*

Yes, in many cases, though students should be sure to contact the department.

*What criteria will typically be used/required to determine whether a student may receive major/concentration credit for a course taken while on study away?*

Course title and description, and complete syllabus including readings/assignments.

*Does your department/program place restrictions on the number of major/concentration credits that a student might earn through study away?*

No.

*Does your department/program place restrictions on the types of courses that can be awarded credit towards your major?*

Yes. They have to be approved MATH/STAT courses.

*Are there specific major requirements that cannot be fulfilled while on study away?*

Yes. Colloquium requirement, Senior Seminar requirement.

*Are there specific major requirements in your department/program that students should be particularly aware of when weighing study away options? (Some examples might include a required course that is always taught in one semester, laboratory requirements.)*

Yes. The highly cumulative structure of the major.

*Give examples in which students thought or assumed that courses taken away would count toward the major or concentration and then learned they wouldn’t:*

None to date.

**THE DEGREE WITH HONORS IN STATISTICS**

The degree with honors in Statistics is awarded to the student who has demonstrated outstanding intellectual achievement in a program of study which extends beyond the requirements of the major. The principal considerations for recommending a student for the degree with honors will be: Mastery of core material and skills, breadth and, particularly, depth of knowledge beyond the core material, ability to pursue independent study of statistics, originality in methods of investigation, and, where appropriate, creativity in research.

An honors program normally consists of two semesters (STAT 493 and 494) and a winter study (WSP 031) of independent research, culminating in a thesis and a presentation. One of STAT 493 or STAT 494 can count as a continuation course, but not both. Neither counts as the 400-level senior capstone course.

An honors program in actuarial studies requires significant achievement on four appropriate examinations of the Society of Actuaries.

Highest honors will be reserved for the rare student who has displayed exceptional ability, achievement or originality. Such a student usually will have written a thesis or pursued actuarial honors. In all cases, the award of honors and highest honors is the decision of the Department.

---

**STAT 101 (F)** Elementary Statistics and Data Analysis (QFR)
It is impossible to be an informed citizen in the world today without an understanding of data and information. Whether opinion polls, unemployment rates, salary differences between men and women, the efficacy of vaccines or consumer webdata, we need to be able to separate the signal from the noise. We will learn the statistical methods used to analyze and interpret data from a wide variety of sources. The goal of the course is to help reach conclusions and make informed decisions based on data.

**Class Format:** lecture

**Requirements/Evaluation:** evaluation will be based primarily on performances on quizzes and exams

**Extra Info:** may not be taken on a pass/fail basis; not available for the fifth course option

**Prerequisites:** MATH 102 (or demonstrated proficiency on a diagnostic test)

**Enrollment Limit:** 50

**Expected Class Size:** 40

**Department Notes:** Students with calculus background and social science interest should consider STAT 161. Students with MATH 150 should enroll in STAT 201. Students with a 5 on AP Stats should enroll in STAT 202. Students with a 4 on AP Stat should consult the department.

**Distributions:** (D3) (QFR)

**Attributes:** BGNP Recommended Courses; COGS Related Courses; PHLH Statistics Courses;

---

**STAT 161 (F) Introductory Statistics for Social Science**  (QFR)

This course will cover the basics of modern statistical analysis with a view toward applications in the social sciences and sciences. Topics include exploratory data analysis, elements of probability theory, basic statistical inference, and introduction to statistical modeling. The course focuses on the application of statistics tools to solve problems, to make decisions, and the use of statistical thinking to understand the world.

**Class Format:** lecture

**Extra Info:** may not be taken on a pass/fail basis; not available for the fifth course option

**Prerequisites:** MATH 130 (or equivalent). Not open to students who have completed STAT 101 or equivalent.

**Enrollment Limit:** 40

**Enrollment Preferences:** Economics majors, sophomores

**Expected Class Size:** 40

**Department Notes:** Students with MATH 150 should consider STAT 201. Students with a 5 on AP Stats should enroll in STAT 202. Students with a 4 on AP Stats should consult the department. Students without any calculus background should consider STAT 101.

**Distributions:** (D3) (QFR)

**Distribution Notes:** QFR: It is a quantitative course

---

**STAT 201 (F) Statistics and Data Analysis**  (QFR)

Statistics can be viewed as the art and science of turning data into information. Real world decision-making, whether in business or science is often based on data and the perceived information it contains. Sherlock Holmes, when prematurely asked the merits of a case by Dr. Watson, snapped back, “Data, data, data! I can't make bricks without clay.” In this course, we will study the basic methods by which statisticians attempt to extract information from data. These will include many of the standard tools of statistical inference such as hypothesis testing, confidence intervals, and linear
regression as well as exploratory and graphical data analysis techniques. This is an accelerated introductory statistics course that involves computational programming and incorporates modern statistical techniques.

**Class Format:** lecture

**Requirements/Evaluation:** evaluation will be based primarily on performance on quizzes and exams

**Prerequisites:** MATH 150 or equivalent. Not open to students who have completed STAT 101 or STAT 161 or equivalent.

**Enrollment Limit:** 40

**Expected Class Size:** 40

**Department Notes:** Students with a 5 on AP Stats should enroll in STAT 202. Students with a 4 on AP Stats should consult the department. Students with MATH 130/140 background should consider STAT 161. Students with no calc. should consider STAT 101.

**Distributions:** (D3) (QFR)

**Attributes:** BGNP Recommended Courses; COGS Related Courses; EVST Methods Courses; PHLH Statistics Courses;

---

### STAT 202 (F) Introduction to Statistical Modeling (QFR)

Data come from a variety of sources sometimes from planned experiments or designed surveys, but also arise by much less organized means. In this course we’ll explore the kinds of models and predictions that we can make from both kinds of data as well as design aspects of collecting data. We’ll focus on model building, especially multiple regression, and talk about its potential as well as its limits to answer questions about the world. We’ll emphasize applications over theory and analyze real data sets throughout the course.

**Class Format:** lecture

**Requirements/Evaluation:** evaluation will be based on homework, exams and projects

**Extra Info:** may not be taken on a pass/fail basis; not available for the fifth course option

**Prerequisites:** AP Statistics 5 or STAT 101, 161 or 201 or permission of instructor

**Enrollment Limit:** 25

**Expected Class Size:** 20

**Department Notes:** Students with a 4 on the AP Stats exam should contact the department for proper placement.

**Distributions:** (D3) (QFR)

**Attributes:** EVST Methods Courses; PHLH Statistics Courses;

---

### STAT 341 (F) Probability (QFR)

Crosslistings: STAT341 / MATH341

**Secondary Crosslisting**

While probability began with a study of games, it has grown to become a discipline with numerous applications throughout mathematics and the sciences. Drawing on gaming examples for motivation, this course will present axiomatic and mathematical aspects of probability. Included will be discussions of random variables, expectation, independence, laws of large numbers, and the Central Limit Theorem. Many interesting and important
applications will also be presented, potentially including some from coding theory, number theory and nuclear physics.

**Class Format:** lecture

**Requirements/Evaluation:** evaluation will be based primarily on homework, classwork, and exams

**Prerequisites:** MATH 250 or permission of the instructor

**Enrollment Limit:** 40

**Expected Class Size:** 20

**Distributions:** (D3) (QFR)

---

Fall 2018

LEC Section: 01    MWF 11:00 am - 11:50 am    Steven J. Miller

Spring 2019

LEC Section: 01    MWF 9:00 am - 9:50 am    Thomas A. Garrity

---

**STAT 344 (F) Statistical Design of Experiments (QFR)**

What does statistics have to do with designing and carrying out experiments? The answer is, surprisingly perhaps, a great deal. In this course, we will study how to design experiments with the fewest number of observations possible that are still capable of understanding which factors influence the results. After reviewing basic statistical theory and two sample comparisons, we cover one and two-way ANOVA and (fractional) factorial designs extensively. The culmination of the course will be a project where each student designs, carries out, analyzes, and presents an experiment of interest to him or her. Throughout the course, we will use both the statistics program R and the package JMP to carry out the statistical analyses.

**Class Format:** lecture

**Requirements/Evaluation:** problem sets, midterm, final exam, project

**Extra Info:** may not be taken on a pass/fail basis; not available for the fifth course option

**Prerequisites:** STAT 201, 202, or equivalent

**Enrollment Limit:** 20

**Enrollment Preferences:** Statistics majors, seniors

**Expected Class Size:** 15

**Distributions:** (D3) (QFR)

**Attributes:** COGS Related Courses;

---

Fall 2018

LEC Section: 01    TF 2:35 pm - 3:50 pm    Richard D. De Veaux

---

**STAT 346 (F) Regression and Forecasting (QFR)**

This course focuses on the building of empirical models through data in order to predict, explain, and interpret scientific phenomena. Regression modeling is the standard method for analyzing continuous response data and their relationship with explanatory variables. This course provides both theoretical and practical training in statistical modeling with particular emphasis on simple linear and multiple regression, using R to develop and diagnose models. The course covers the theory of multiple regression and diagnostics from a linear algebra perspective with emphasis on the practical application of the methods to real data sets. The data sets will be taken from a wide variety of disciplines.

**Class Format:** lecture

**Requirements/Evaluation:** evaluation will be based primarily on performance on exams, homework, and a project

**Extra Info:** may not be taken on a pass/fail basis; not available for the fifth course option

**Prerequisites:** STAT 201 or 202, and MATH 150 and 250; or permission of instructor

**Enrollment Limit:** 22

**Expected Class Size:** 15

**Distributions:** (D3) (QFR)
STAT 355 (S) Multivariate Statistical Analysis (QFR)
To better understand complex processes, we study how variables are related to one another, and how they work in combination. Therefore, we want to make inferences about more than one variable at a time? Elementary statistical methods might not apply. In this course, we study the tools and the intuition that are necessary to analyze and describe such data sets. Topics covered will include data visualization techniques for high dimensional data sets, parametric and non-parametric techniques to estimate joint distributions, techniques for combining variables, as well as classification and clustering algorithms.

Class Format: lecture
Requirements/Evaluation: evaluation will be based on homework and exams
Prerequisites: STAT 201 and MATH 250
Enrollment Limit: 25
Expected Class Size: 10
Distributions: (D3) (QFR)

Spring 2019
LEC Section: 01    MWF 8:30 am - 9:45 am     Xizhen Cai

STAT 356 (S) Time Series Analysis (QFR)
Time series--data collected over time--crop up in applications from economics to engineering to transit. But because the observations are generally not independent, we need special methods to investigate them. This course will include exploratory methods and modeling for time series, including smoothing, ARIMA and state space models, and a foray into the frequency domain. We will emphasize applications to a variety of real data.

Class Format: lecture
Requirements/Evaluation: evaluation will be based primarily on projects, homework, and exams
Prerequisites: STAT 346 (may be taken concurrently) or permission of instructor
Enrollment Limit: 30
Expected Class Size: 20
Distributions: (D3) (QFR)
Not offered current academic year

STAT 359 (S) Statistical Computing (QFR)
This course introduces a variety of computational and data-centric topics of applied statistics, which are broadly useful for acquiring, manipulating, visualizing, and analyzing data. We begin with the R language, which will be used extensively throughout the course. Then we'll introduce a variety of other useful tools, including the UNIX environment, scripting analyses using bash, databases and the SQL language, alternative data formats, techniques for visualizing high-dimensional data, and text manipulation using regular expressions. We'll also cover some modern statistical techniques along the way, which are made possible thanks to advances in computational power. This course is strongly computer oriented, and assignments will be project-based.

Class Format: lecture
Requirements/Evaluation: based primarily on projects, homework, and exams
Prerequisites: STAT 201 or 202 and CSCI 134, 135, or 136
Enrollment Limit: 30
Enrollment Preferences: juniors and seniors, Statistics majors
Expected Class Size: 15
Distributions: (D3) (QFR)
Not offered current academic year

STAT 360 (S) Statistical Inference  (QFR)
How do we estimate unknown parameters and express the uncertainty we have in our estimate? Is there an estimator that works best? Many topics from introductory statistics such as random variables, the central limit theorem, point and interval estimation and hypotheses testing will be revisited and put on a more rigorous mathematical footing. The focus is on maximum likelihood estimators and their properties. Bayesian and computer intensive resampling techniques (e.g., the bootstrap) will also be considered.

Class Format: lecture
Requirements/Evaluation: evaluation will be based on problem sets and exams
Extra Info: may not be taken on a pass/fail basis
Prerequisites: MATH 250, STAT 201 or 202, STAT 341
Enrollment Limit: 30
Enrollment Preferences: Statistics majors
Expected Class Size: 30
Distributions: (D3) (QFR)

Spring 2019
LEC Section: 01    TR 9:55 am - 11:10 am     Daniel B. Turek

STAT 362 (F) Design of Experiments  (QFR)
How do you get informative research results? By doing the right experiment in the first place. We'll look at the techniques used to plan experiments that are both efficient and statistically sound, the analysis of the resulting data, and the conclusions we can draw from that analysis. Using a framework of optimal design, we'll examine the theory both of classical designs and of alternatives when those designs aren't appropriate. On the applied side, we'll make extensive use of R to work with real-world data.

Class Format: lecture
Requirements/Evaluation: based primarily on projects, homework, and exams
Extra Info: may not be taken on a pass/fail basis; not available for the fifth course option
Prerequisites: STAT 346
Enrollment Limit: 30
Enrollment Preferences: seniors/juniors and Statistics majors
Expected Class Size: 20
Distributions: (D3) (QFR)
Not offered current academic year

STAT 365 (F) Bayesian Statistics  (QFR)
The Bayesian approach to statistical inference represents a reversal of traditional (or frequentist) inference, in which data are viewed as being fixed and model parameters as unknown quantities. Interest and application of Bayesian methods have exploded in recent decades, being facilitated by recent advances in computational power. We begin with an introduction to Bayes' Theorem, the theoretical underpinning of Bayesian statistics which dates back to the 1700's, and the concepts of prior and posterior distributions, conjugacy, and closed-form Bayesian inference. Building on this, we introduce modern computational approaches to Bayesian inference, including Markov chain Monte Carlo (MCMC), Metropolis-Hastings sampling, and the theory underlying these simple and powerful methods. Students will become comfortable with modern software tools for MCMC using a variety of applied hierarchical modeling examples, and will use R for all statistical computing.

Class Format: lecture
STAT 372 (S) Longitudinal Data Analysis: Modeling Change over Time (QFR)

This course explores modern statistical methods for drawing scientific inferences from longitudinal data, i.e., data collected repeatedly on experimental units over time. The independence assumption made for most classical statistical methods does not hold with this data structure because we have multiple measurements on each individual. Topics will include linear and generalized linear models for correlated data, including marginal and random effect models, as well as computational issues and methods for fitting these models. We will consider many applications in the social and biological sciences.

Class Format: lecture
Requirements/Evaluation: evaluation will be based primarily on performance on exams, homework, and a project
Extra Info: may not be taken on a pass/fail basis; not available for the fifth course option
Prerequisites: STAT 201 and STAT 346
Enrollment Limit: 30
Enrollment Preferences: junior and senior Statistics majors
Expected Class Size: 20
Distributions: (D3) (QFR)
Attributes: PHLH Statistics Courses;

Spring 2019
LEC Section: 01    TF 1:10 pm - 2:25 pm    Anna M. Plantinga

STAT 377 (F) Operations Research (WI) (QFR)
Crosslistings: MATH377 / STAT377
Secondary Crosslisting

In the first N math classes of your career, you can be misled as to what the world is truly like. How? You're given exact problems and told to find exact solutions. The real world is sadly far more complicated. Frequently we cannot exactly solve problems; moreover, the problems we try to solve are themselves merely approximations to the world! We are forced to develop techniques to approximate not just solutions, but even the statement of the problem. Additionally, we often need the solutions quickly. Operations Research, which was born as a discipline during the tumultuous events of World War II, deals with efficiently finding optimal solutions. In this course we build analytic and programming techniques to efficiently tackle many problems. We will review many algorithms from earlier in your mathematical or CS career, with special attention now given to analyzing their run-time and seeing how they can be improved; students will be implementing many of these algorithms on computer systems of their choice. The culmination of the course is a development of linear programming and an exploration of what it can do and what are its limitations. For those wishing to take this as a Stats course, the final project must have a substantial implementation computation (respectively, statistics) component approved by the instructor.
Class Format: lecture
Requirements/Evaluation: evaluation will be based primarily on homework, classwork, projects, presentations and exams; at least 20 pages of writing
Prerequisites: MATH 350 or 351 and permission of instructor
Enrollment Limit: 40
Enrollment Preferences: Computer Science, Mathematics and Statistics majors
Expected Class Size: 25
Department Notes: http://web.williams.edu/Mathematics/sjmiller/public_html/317/
Distributions: (D3) (WI) (QFR)
Not offered current academic year

STAT 397 (F) Independent Study: Statistics
Directed independent study in Statistics.
Class Format: independent study
Prerequisites: permission of department
Distributions: (D3)

Fall 2018
IND Section: 01   TBA   Susan R. Loepp

STAT 398 (S) Independent Study: Statistics
Directed independent study in Statistics.
Class Format: independent study
Prerequisites: permission of department
Distributions: (D3)

Spring 2019
IND Section: 01   TBA   Susan R. Loepp

STAT 440 (F) Categorical Data Analysis (QFR)
This course focuses on methods for analyzing categorical response data. In contrast to continuous data, categorical data consist of observations classified into two or more categories. Traditional tools of statistical data analysis are not designed to handle such data and pose inappropriate assumptions. We will develop methods specifically designed to address the discrete nature of the observations and consider many applications in the social and biological sciences as well as in medicine, engineering and economics. All methods can be viewed as extensions of traditional regression models and ANOVA.
Class Format: lecture
Requirements/Evaluation: evaluation will be based primarily on performance on exams, homework, and a project
Prerequisites: STAT 346 and STAT 360
Enrollment Limit: 14
Enrollment Preferences: seniors and Statistics Majors
Expected Class Size: 12
Distributions: (D3) (QFR)
Attributes: PHLH Statistics Courses;
Not offered current academic year

STAT 442 (S) Statistical Learning and Data Mining (QFR)
In both science and industry today, the ability to collect and store data can outpace our ability to analyze it. Traditional techniques in statistics are often unable to cope with the size and complexity of today's data bases and data warehouses. New methodologies in Statistics have recently been developed, designed to address these inadequacies, emphasizing visualization, exploration and empirical model building at the expense of traditional hypothesis testing. In this course we will examine these new techniques and apply them to a variety of real data sets.
STAT 458 (F) Spatio-Temporal Data (QFR)

Everything happens somewhere and sometime. But the study of data collected over multiple times and locations requires special methods, due to the dependence structure that relates different observations. In this course, we'll look at exploring, analyzing, and modeling this kind of information--introducing standard methods for purely time-series and purely spatial data, and moving on to methods that incorporate space and time together. Topics will include autocovariance structures, empirical orthogonal functions, and an introduction to Bayesian hierarchical modeling. We'll use R to apply these techniques to real-world datasets.

Class Format: lecture
Requirements/Evaluation: project work, homework, exams, and contribution to discussion.
Extra Info: may not be taken on a pass/fail basis; not available for the fifth course option
Prerequisites: STAT 346, or permission of instructor
Enrollment Limit: 14
Enrollment Preferences: Seniors and Statistics majors
Expected Class Size: 10
Distributions: (D3) (QFR)

Spring 2019
LEC Section: 01  TF 2:35 pm - 3:50 pm  Richard D. De Veaux

STAT 493 (F) Senior Thesis: Statistics

Each student carries out an individual research project under the direction of a faculty member that culminates in a thesis. See description under The Degree with Honors in Statistics.

Class Format: independent study
Distributions: (D3)

Fall 2018
HON Section: 01  TBA  Susan R. Loepp

STAT 494 (S) Senior Thesis: Statistics

Each student carries out an individual research project under the direction of a faculty member that culminates in a thesis. See description under The Degree with Honors in Statistics.

Class Format: independent study
STAT 497 (F) Independent Study: Statistics

Directed independent study in Statistics.

Class Format: independent study

Prerequisites: permission of department

Distributions: (D3)

STAT 498 (S) Independent Study: Statistics

Directed independent study in Statistics.

Class Format: independent study

Prerequisites: permission of department

Distributions: (D3)

STAT 499 (F) Statistics Colloquium

Statistics senior colloquium. Meets every week for an hour both fall and spring. Senior statistics majors must participate. This colloquium is in addition to the regular four semester-courses taken by all students.

Class Format: seminar

Requirements/Evaluation: delivering a passing talk and participation throughout the year

Prerequisites: Statistics majors must take the colloquium in their senior year

Enrollment Limit: none

Enrollment Preferences: none

Expected Class Size: 25

Distributions: (D3)