Revision May 21, 2021
Draft Calendar--Subject to change

Statistics for Data Analysis

Instructor Information

Instructor: John Harrington
Email: jharrington@jhu.edu
Office Hours: by Appointment
Website: https://sais.jhu.edu/users/jharrin1

Teaching Assistant Information

TBD

Course Description

This course is designed to furnish students with the fundamental tools of statistical analysis, including analysis of descriptive statistics, probability and probability distributions, statistical inference via confidence intervals and significance tests, and correlation and simple/multiple linear regression analysis. Moreover, an introduction to big data and Machine Learning is provided at the end of the course. Aim of the course is to introduce the basic statistical tools required to conduct and evaluate empirical research in economics and the social sciences. Special attention will be given to the application of these statistical tools to the analysis of real phenomena using computer to solve problems and to reinforce statistical concepts.

Course Objectives

By the end of this course, students will be able to:

- Use statistics to describe and picture data
- Apply the language of probability to statistics

- Illustrate discrete random variables including Binomial and Poisson
- Describe a continuous random variable including uniform and normal
- Solve binomial distribution problems using the normal approximation to the binomial
- Describe sampling distributions for sample means and sample proportions
- Apply statistical inference to confidence intervals and hypothesis testing
- Apply regression analysis in two variable and multiple variable models
- Understand the difference between explanatory models and predictive models

Required Textbooks & Materials

Note: There may be a change in the textbook. All textbook readings will be available at no additional cost electronically.

Statistics for Business and Economics, Anderson, Sweeney, Williams, Camm and Cochran, South Western Cengage Learning, 13e (2018) 12e (2014), (note: 10e or 11e should work as well). (Hereafter ASWCC) (Optional: eBook is available to rent or buy directly from Cengage or through Amazon)

Calculator. You will need a calculator. A cell phone calculator will be fine. If you purchase a calculator, I suggest you purchase a business/financial calculator such as TI BA II Plus.

Required Software

The software required are STATA and Excel. Both are available from JHU.

There are two ways to obtain STATA

1. Access remotely through JHU Cloud Apps Access.
2. *OPTIONAL Purchase a STATA Student single-user license. For this course, the 6-month or annual Stata/IC license is sufficient.*

Microsoft Excel (Microsoft Office 365 is available through the MyJHU SAIS Portal > Technology > Office 365 Portal or the professional version is available from jhu.onthehub.com)

Technology & Skills Requirements

- Reliable high-speed internet service
- Recommend using current version of Firefox browser (Blackboard's supported web browsers)
- Navigate and use Blackboard Learn (Blackboard help)
- Create and save MS Word, Excel and/or PowerPoint documents (Office Help & Training)
- Send, receive, and manage email

Assignments and Grading Policy

This course will be graded based on the following assignments.

Assignments	Points possible	Due Date
Graded Problem Sets	20%	End of Module Week
Midterm	30%	July 24-27
Final	50%	August 17-18
TOTAL	**100%**	

This course uses the following grade scale:

Grade	Description	Percent	Pre-Term Grade for Transcript
A	Outstanding	95% to 100%	HP
A-	Excellent	90% and less than 95%	HP
B+	Very good	87% and less than 90%	P
B	Good	83% and less than 87%	P
B-	Pass	80% and less than 82%	P
C+	Low pass	75% and less than 80%	P
C	Minimal pass	70% and less than 75%	P
D	Failure	Less than 70%	Course not on transcript

Assignment General Guidelines

Students are required to adhere to the following guidelines when submitting written assignments, unless otherwise noted in the assignment.

- Handwritten assignments should be written neatly, scanned into a single PDF document and submitted in Blackboard.
- Submit all assignments in Blackboard
- Important Note: Some of the solutions to the problems to be turned in appear to be available. Those solutions involve some “mistakes” that do not match the current problem set. Anyone found copying these answers will end up with a -40 for that problem set which will likely reduce your final grade by one or two grade levels. If this happens more than once, an Honor Code violation will be brought before an Honor Board
- If typing an assignment, use a 12-point font (e.g. Arial, Calibri, Times New Roman), double space and 1-inch margin.

Late Policy

Please submit your assignments by the deadlines outlined in the course syllabus and Blackboard. If you are not able to meet an assignment deadline (including the grace period) contact your instructor in advance of the deadline. Each problem set is due at the end of the week. You should make every effort

to post the solutions on time. You will be given a two day “grace” period after which you will have .2 points per day deducted from your grade (out of 10 points).

Assignment Feedback

The instructor will return assignments to you within 5-7 days following the due date. Your grade and feedback will be in the My Grades area of Blackboard.

Course Structure

This course is divided into 14 modules. The first seven will be on-line. The final modules will be in-class.

Module 1 – Descriptive Statistics and Data Visualization
This module will cover some basic measures of central tendency and dispersion, including the mean, median, mean deviation (mean absolute error), and standard deviation. For stationary series, the class will cover grouping the data and drawing a picture of the grouped data called a histogram. Other examples of data visualization will be introduced: pie chart, scatterplot, and stem-and-leaf display. The concept of a population and a sample will be covered as well as the summation notation. We will also look at the relationships between two variables including measuring this relationship with the correlation coefficient.

Module 2 – Probability
Probability is a very broad subject, but we will limit our discussion to one class and focus only on those topics that help one to understand the concept of probability. The concept of probability is fundamental to the understanding of statistics. Topics will include the definition of probability; events, mutually exclusive events, the "or" rule; conditional probability; statistical dependence, statistical independence, and the "and" rule. Statistical dependence arising from sampling without replacement, and statistical independence, arising from sampling with replacement will be covered. Finally, all of the above rules will be combined to develop Bayes Theorem.

Module 3 – Random Variable
The concept of a random variable is a core concept in statistics. In this class, the concept of a discrete random variable associated with a discrete probability distribution is covered. The definition of the expected value (mean) and the standard deviation of a discrete probability distribution will be presented. Calculating the covariance between two joint probability distributions will be covered and an application in financial analysis will be presented.

Module 4 – Binomial and Poisson Probability Distributions
Many types of applications of statistics are based on experiments with only two outcomes. For example, a citizen will either vote for or against a particular candidate. The type of experiment that leads to a binomial probability distribution will be discussed as well as how to derive and apply the binomial probability distribution. Some outcomes occur randomly over time or space. These outcomes can be described by the Poisson Probability Distribution.

Module 5 – Uniform and Normal Probability Distributions
The idea of a continuous random variable will be presented with a special case of the uniform probability distribution. The general uniform probability distribution will be presented along with the cumulative uniform probability distribution. The normal distribution arises from the concept of sampling

as we will see in the next class. In this module you will learn the definition of the normal probability distribution, how to use the normal probability distribution table, and how to solve normal distribution problems, including how to use the normal distribution to approximate the binomial probability distribution.

Module 6 –Review and Mid-Term Exam
Exam will cover material from the first five modules.

Module 7 – Sampling and Sampling Distributions
In this module we will discuss the concept of a random and a probability sample. The concept of a random sample will be carried over to the Central Limit Theorem. From this we will develop concept of distribution of sample mean and distribution of sample proportion.

Module 8 – Confidence Intervals or Interval Estimation
In this module, you will learn how to make probability statements about population parameters based on sample information. We will cover the distribution of sample means and the distribution of sample proportions. We will develop these ideas for large samples (normal distribution) and for small sample (t distribution).

Module 9 – Hypothesis Testing
In this module, there will be no new theory, but the theory that was used to develop confidence intervals will be used to look at the problem of making statements about a population parameter. These statements are called hypotheses, and we will learn how to make these statements, the types of errors one can make (type I and type II errors) and how we can apply hypothesis testing using the distribution of sample means and the distribution of sample proportions.

Module 10 – Applications of Confidence Intervals and Hypothesis Testing to Parameters of Two Populations; Introduction to Non-parametric Statistics; Contingency Table
Further examples of hypothesis testing will be explored covering the distribution of the difference of sample means - with large and small samples; difference between sample means for matched samples; and difference between two population proportions. The Mann-Whitney test for the difference of sample means with small samples will be discussed as an example of a non-parametric or distribution free technique in statistics. Policy analysis often involves the use of surveys. This module will cover an important application of the Chi-squared distribution as it applies to the examination of the results of surveys by testing the independence of two categorical variables.

Module 11 – Linear Regression - Basic Concepts and Functional Forms
The mathematical technique for describing the relationships between variables is the function. The technique of function estimation is called regression. In this module we will develop the basic concepts for two variable regression (function fitting with one independent variable), as well as measures for how well a function fits the original data. This discussion will include decomposition of the sum of squares leading to the coefficient of determination. Functions involving logarithms will be covered permitting estimating certain non-linear functions. There is a combined problem set for Modules 11 and 12.

Module 12 – Linear Regression - Statistical Analysis; Introduction to Multiple Regression

This module will use a simulation to explain how a slope coefficient can be thought of as drawn from a sampling distribution of all possible slope coefficients. Using the assumptions made when regressions are run, this class will then develop the concept of standard error of a regression coefficient, confidence interval for the true slope, and hypothesis testing. Use of regression with two or more independent variables will be covered briefly, but it will not be included in the final exam. There is a combined problem set for Modules 11 and 12.

Module 13 – Explanatory Models and Predictive Models
In this module we will compare the explanatory objectives of regression analysis with an introduction to predictive models that use Machine Learning. We will introduce concepts such as Artificial Intelligence, the different types of Machine Learning algorithms, Deep Learning, and a framework for building Machine Learning models. A concrete example of how Machine Learning works will be provided. This module will not be included in the final exam.

Module 14 – Review and Final Exam
Exam will cover material from the Module 7 to Module 12 although understanding the first five modules is essential to understand Modules 7-12.

Changes to the Course

Changes to the course will be posted in the Announcements.

Honor Code Statement

Enrollment at SAIS obligates each student to conduct all activities in accordance with the rules and spirit of the school's Honor Code located in The Red Book: SAIS Student and Academic Handbook. The Honor Code governs student conduct at SAIS. It covers all activities in which students present information as their own, including written papers, examinations, oral presentations and materials submitted to potential employers or other educational institutions. It requires that students be truthful and exercise integrity and honesty in their dealings with others, both inside SAIS and in the larger community.

While the Honor code goes well beyond plagiarism, it is important that each student understand what is and is not plagiarism. Plagiarism will result in failure of the paper or exam and may result in failing the course depending on the judgment of the professor. Note: All papers submitted for this course will be automatically processed by an anti- plagiarism system to ensure the integrity of work.

University Policies

The Johns Hopkins SAIS Student and Academic Handbook, also known as "The Red Book," is a compilation of policies, regulations and procedures for students. Its purpose is two-fold: to communicate the standards of The Johns Hopkins University that support and guide life at Johns Hopkins SAIS as part of the greater JHU community and to describe the academic policies and procedures that form a framework for conducting the school's teaching mission. Of particular importance is the Honor Code, which sets out the behavioral standards expected of all Johns Hopkins SAIS students.

THE RED BOOK

The information contained in this handbook is not available in any other school publication, and students are responsible for familiarizing themselves with its contents. Questions related to the manual should be directed to Academic Affairs.

The policies and procedures detailed in The Red Book are subject to revision at any time, and changes are communicated to students only through their assigned JHU e-mail addresses. It is imperative that you activate and monitor this account so as not to miss these and other important announcements and messages throughout the year.

Disability Services

Johns Hopkins University is committed to providing an accessible and welcoming learning environment for students with disabilities under the Americans with Disabilities Act of 1990 and its 2008 Amendments, as well as Section 504 of the Rehabilitation Act of 1973. The Johns Hopkins University Disability Services collaborates with students, faculty, and staff to provide equitable, inclusive, and sustainable learning environments that promote academic success for all. Johns Hopkins University is committed to making academic programs, support services and facilities accessible.

Students seeking accommodations must submit the Student Request for Accommodation of Disability Form (available on the Insider Portal). Documentation must be from a qualified professional, such as a physician. It is the student's responsibility to provide or pay for the cost of this documentation. The documentation, depending on the type of disability, must be recent and no more than three years old. Please consult the JHU Documentation Guidelines for Individuals with Disabilities or contact the Office of Student Life for further specification. Johns Hopkins University reserves the right to request or require more current or updated documentation. Documentation may be submitted to us at any time; however, students should leave a margin of at least three weeks prior to the intended start of the accommodation in order to provide adequate time for review and processing of the request.

Student Life will inform the student of the status of his or her request within five business days from the intended beginning of the accommodation. Accommodations take effect upon approval and apply to the remainder of the time for which a student is registered and enrolled.

The Johns Hopkins University Executive Director of Student Disability Services reviews student documentation and reserves the right to determine the most effective and timely accommodations after consultation with the student. There are detailed procedures for use of the services and accommodations.

Title IX

The Sexual Misconduct Policy and Procedures (“SMPP”) apply to cases of sexual misconduct, which includes sexual harassment, sexual assault, relationship violence, and stalking. Complaints of sexual misconduct are processed pursuant to The Johns Hopkins University Sexual Misconduct Policy and Procedures. Questions regarding this Policy and these Procedures and any questions concerning Title IX should be referred to the University’s Title IX Coordinator: Joy Gaslevic, The Johns Hopkins University, Office of 4 Institutional Equity, Wyman Park Building, Suite 515, 3400 North Charles Street, Baltimore, MD 21218, Telephone: 410.516.8075, TTY: Dial 711, email titleixcoordinator@jhu.edu.

Student Code of Conduct

Becoming a member of the Johns Hopkins University community is an honor and privilege. Acceptance of membership in the University community carries with it an obligation on the part of each individual to respect the rights of others, to protect the University as a forum for the free expression of ideas, and to obey the law. Students are required to know and abide by the University Student Conduct Code. It is important that you take a few minutes to read, review and know the Code before arriving on campus as your academic success is enhanced when you are member of a respectful, safe and healthy community.

Complaints asserting Conduct Code violations may be initiated by: (1) The Assistant Dean for Student Affairs or designee; (2) a student; or (3) a member of the faculty or staff. The Assistant Dean for Student Affairs or designee has responsibility for administering matters initiated under the Conduct Code.

We urge individuals who have experienced or witnessed incidents that may violate this code to report them to campus security, the appropriate Director of Student Life or the Assistant Dean for Student Affairs. The university will not permit retaliation against anyone who, in good faith, brings a complaint or serves as a witness in the investigation of a complaint.

Course Schedule

This schedule is subject to change with fair notice. Any changes will be posted in the Course Announcements.

Module	Date	Module Title	Readings	Assignments (Due Dates)
1	July 5 – July 7	Descriptive Statistics	ASWCC: Chapters 1-3 *(most of the lecture/discussion material will be from Chapter 3; Chapters 1 and 2 should be read)*	Practice problem set Graded problem set (July 11)
2	July 8 – July 10	Probability	ASWCC: Chapter 4	Practice problem set Graded problem set (July 11)
3	July 12 – July 14	Discrete Random Variables	ASWCC: Chapter 5.1-5.4	Practice problem set Group Graded problem set (July 18)
4	July 15 – July 17	Examples of Discrete Probability Distributions: Binomial Probability Distribution and Poisson Probability Distribution	ASWCC: Chapter 5.5 - 5.6	Practice problem set Graded problem set (July 18)
5	July 19 – July 21	Uniform and Normal Probability Distributions	ASWCC: Chapter 6.1-6.3	Practice problem set Group Graded problem set (July 23)
6	July 22– July 26	Review and Midterm Exam		Midterm (July 25-26)

7	July 27 – July 29	Sampling Distributions	ASWCC: Chapter 7	Practice problem set Graded problem set (Oct 31)
8	August 2	Statistical Inference I (In class lecture)	ASWCC: Chapter 8	Practice problem set Graded problem set (August 8)
9	August 4	Hypothesis Testing I	ASWCC: Chapter 9	Practice problem set Group Graded problem set (August 8)
10	August 6	Statistical Inference and Hypothesis Testing II Applications of Chi-squared distribution: Contingency Tables; Example of Non-parametric (Distribution Free) statistical test	ASWCC: Chapter 10; Chapter 12.2 and Chapter 18.2	Practice problem set Graded problem set (August 8)
11	August 9	Two Variable Linear Regression – Estimation and Sampling Distribution	ASWCC: Chapter 14.1-14.3	Practice problem set Class 11-12 Graded problem set (August 14)
12	August 11	Two Variable linear Regression – Confidence Intervals and Hypothesis Testing; Introduction to Multiple Regression	ASWCC: Chapter 14.4-14.9	Practice problem set Class 11-12 Graded problem set (August 14)
13	August 13	Explanatory Models and Predictive Models		No new graded problem set
14	August 16-August 18	Review and Final Exam		Final Exam (August 17-18)