Introduction to Statistics and Data Science: Instructor's Guide

Danielle Sass

2024-03-27

Table of contents

Preface

This book is meant to be a resource for instructors to provide a complete guide for teaching Introduction to Statistics and Data Science.

All activities and lecture slides can be downloaded from the Posit Cloud ISDS Course Content

Reading Tutorials are completed using RStudio. Visit the isdsTutorials website for information on how to install the package and run the tutorials.

```
r
fontawesome::fa("star",
fill =
"steelblue",
prefer_type
= "solid")
```

This icon indicates a tip/suggestion!

Sample syllabus

Course info

	Day	Time	Location
Lectures	MWF	9:00 am - 9:50 am	Room 107

Prerequisite: High School Algebra

Instructional Team

	Title	Email	Office Hours
Prof. Name	Professor		By Appointment
TA Name	TA		Tue/Thur 4-5pm

Learning objectives

By the end of the quarter, you will be able to…

1. Use statistical software to manage and process data.
2. Use statistical software to perform exploratory data analyses. That is, explore data numerically and visually to gain understanding through data and generate hypotheses and inferences to later test.

"Numbers don't lie. That's where we come in."

3. Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
4. Build a conceptual understanding of the unified nature of statistical inference.
5. Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
6. Model numerical response variables using a single or multiple explanatory variables.
7. Interpret results in context without relying on statistical jargon.
8. Critique and evaluate data-based claims and decisions.

Course Structure

This class will follow an active learning design. Meaning the majority of each lecture will be dedicated to working on activities. A lot of what you do in this course will involve writing code, and coding is a skill that is best learned by doing. A typical class will devote 10-15 minutes to discussion/lecture with the remainder of the class devoted to working on activities where students will either work by themselves or in groups. Throughout the class we will discuss and review the work on the activities. In many cases we may come together to work on parts of an activity as a class.

Textbooks

We will be using Introduction to Statistics and Data Science which is a free online book that we have been developing for this course.

Software

We will be using/introducing the free statistical software Posit Cloud.

Hardware

Students will need a laptop or Chromebook to be able to follow lectures and to work with Posit Cloud to complete activities. If access to a laptop is an issue, then please contact the course instructor and we will work to find an accommodation.

Assessment

Assessment for the course is comprised of the following components: reading tutorials, activities, 3 exams, and a final project.

Participation

Reading Tutorials

Reading tutorials will be completed using Posit Cloud/Rstudio and uploaded to the course Canvas page. Each reading tutorial will be scaled to be worth 10 points. Reading tutorials will be accepted up to 3 days after the due date with a 10% late penalty.

The lowest reading tutorial grade will be dropped at the end of the semester.

Activities

Daily activities will be worth 10 points. Activities will be accepted up to 3 days after the due date with a 10% late penalty.

The lowest activity grade will be dropped at the end of the semester.

Exams

There will be 3 in-class exams; they will be structured very similarly to your reading tutorials. Roughly half of it will focus on conceptual knowledge and roughly half will focus on practical applications. Students will be allowed one 8.5 x 11 inch cheat sheet (front & back) on each in-class exams. The exams are not cumulative.

Project

The final project will be completed in groups of 3-6 people and allow you to explore a dataset of your choice. More information will be provided later in the quarter.

Exam Improvement Policy

We have worked to develop a policy geared towards a growth mindset. That is, we want a policy where students clearly demonstrate that they have used the exam as a diagnostic tool to learn from and improve their understanding of statistics. There is NO final cumulative exam during the designated final exam time, instead you may choose to retake 1 exam during the exam time. This exam will replace your old score — only in cases where it is an improvement.

Missed Exam Policy

There are no make-up exams. If you miss an exam due to illness, travel, etc., you will need to take the exam during the final exam period.

Grading

The final course grade will be calculated as follows:

Category	Percentage
Reading Tutorials	15%
Activities	10%
Exam 1	20%
Exam 2	20%
Exam 3	20%
Project	15%

The final letter grade will be determined based on the following thresholds:

Letter Grade	Final Grade
A	>= 93
A-	90 - 92.9
B+	87 - 89.9
B	83 - 86.9
B-	80 - 82.9

Letter Grade	Final Grade
C+	77 - 79.9
C	73 - 76.9
C-	70 - 72.9
D	60 - 69.9
F	< 59.9

Tips for success

- Dedicate yourself to being an active and engaged learner.
- Prepare for class by *reading and working* through code *before* class.
- Work in groups to learn and complete activities.
- Ask questions! Ask them during class, office hours, or on Campuswire.
- Contribute to a welcoming and inclusive learning environment.
- Don't be afraid to make mistakes, you learn from mistakes.

Asking Questions & Course Communication

This term we will be using Campuswire ("Enrollment Code: XXXX") as our preferred platform for questions about activities, reading tutorials, and general course questions. The system is highly catered to getting you help quickly and efficiently from classmates and the instructional team. Rather than emailing questions to the instructional team, you should post your questions on Campuswire.

The instructional team will check Campuswire periodically and answer questions, but we strongly encourage students to answer each other's questions. To this end, student will be able to earn bonus points — see Canvas for details.

Please do not expect answers during weekends and evenings.

School Policies

Add your standard school policies here.

Tentative schedule

11 week Schedule

	Topic	Textbook	Agenda
Day 1	Syllabus Day		Syllabus Day
Day 2	Intro to R	Preface and Chapter 1	Activity 01
Day 3	Data Visualization	Sections 2.0-2.3	Activity 02
Day 4	Data Visualization	Sections 2.4-2.6	Activity 03
Day 5	Data Visualization	Sections 2.7-2.9	Activity 04
Day 6	Data Wrangling	Sections 3.0-3.3	Activity 05
Day 7	Data Wrangling	Sections 3.4-3.9	Activity 06
Day 8	Tidy Data	Chapter 4	Activity 07
Day 9	Exam 1		
Day 10	Regression	Sections 5.0-5.1	Activity 08
Day 11	Regression	Sections 5.2-5.4	Activity 09
Day 12	Regression	Sections 6.0-6.1	Activity 10
Day 13	Regression	Sections 6.2-6.4	Activity 11
Day 14	Randomization + Causality	Chapter 7	Activity 12
Day 15	Populations + Generalizability	Chapter 8	Activity 13
Day 16	Exam 2		
Day 17	Distributions	Sections 9.0-9.1	Activity 14
Day 18	Repeated Sampling	Sections 9.2-9.4	Activity 15
Day 19	CLT	Sections 9.5-9.7	Activity 16
Day 20	Confidence Intervals	Chapter 10	Activity 17
Day 21	Confidence Intervals	Chapter 10	Activity 17
Day 22	P-values	Chapter 11	Activity 18
Day 23	Hypothesis Testing	Chapter 12	Activity 19
Day 24	Hypothesis Testing	Chapter 12	Activity 20
Day 25	Hypothesis Testing	Chapter 12	Activity 20
Day 26	Review	Chapter 13	
Day 27	Exam 3 Final Project		

Day 01

Welcome to class! Today is all about getting setup with the needed resources and reviewing the syllabus.

Agenda

time	agenda
15 min	Syllabus and expectations
20 min	Welcome and get students setup with all of the needed resources
15 min	Students take survey and get to know their neighbors

Today's tasks

I recommend using Posit Cloud if possible for ease of getting started. Otherwise you will need to dedicate more time towards downloading R and RStudio.
See the sample survey for some examples of questions to ask your students! We use these survey results on the exams.

Make sure you have completed the following agenda items by the end of class!

1. Review the syllabus
2. Visit the course's Campuswire page using your school email. Enrollment code: XXXX
3. We will be using AHA slides for participation.
4. Gain access to the Posit Cloud Class Workspace or install R and RStudio
5. Set up and test out the Reading Tutorials System. Install the required package: `remotes::install_github("NUstat/isdsTutorials", dependencies = TRUE)`
   Try running the first tutorial: `learnr::run_tutorial("01_intro", package = "isdsTutorials")`
6. Login/create your Northwestern Zoom Account if you do not have one. This will be used for office hours or requested appointments.
7. Complete the google survey - we will use this data later.

Homework

- Read Preface and Chapter 1 of the book
- Complete RT 01 by running `learnr::run_tutorial("01_intro", package = "isdsTutorials")` in your Console and submitting the downloaded html to a learning management system

References