# Data Science Inference and Modeling

The textbook for the Data Science course series is freely available online.

## Learning Objectives

- The concepts necessary to define estimates and margins of errors of populations, parameters, estimates, and standard errors in order to make predictions about data
- How to use models to aggregate data from different sources
- The very basics of Bayesian statistics and predictive modeling

## Course Overview

### Section 1: Parameters and Estimates

You will learn how to estimate population parameters.

### Section 2: The Central Limit Theorem in Practice

You will apply the central limit theorem to assess how close a sample estimate is to the population parameter of interest.

### Section 3: Confidence Intervals and p-Values

You will learn how to calculate confidence intervals and learn about the relationship between confidence intervals and p-values.

### Section 4: Statistical Models

You will learn about statistical models in the context of election forecasting.

### Section 5: Bayesian Statistics

You will learn about Bayesian statistics through looking at examples from rare disease diagnosis and baseball.

### Section 6: Election Forecasting

You will learn about election forecasting, building on what you've learned in the previous sections about statistical modeling and Bayesian statistics.

## Section 7: Association Tests

You will learn how to use association and chi-squared tests to perform inference for binary, categorical, and ordinal data through an example looking at research funding rates.

## Section 1 Overview

Section 1 introduces you to parameters and estimates.

After completing Section 1, you will be able to: - Understand how to use a sampling model to perform a poll. - Explain the terms population, parameter, and sample as they relate to statistical inference. - Use a sample to estimate the population proportion from the sample average. - Calculate the expected value and standard error of the sample average.

The textbook for this section is available here

## Assessment 1.1: Parameters and Estimates

1. Polling - expected value of S

Suppose you poll a population in which a proportion **p** of voters are Democrats and **1−p** are Republicans. Your sample size is **N=25**. Consider the random variable **S**, which is the total number of Democrats in your sample.

What is the expected value of this random variable **S**?

Possible Answers - [ ] A. E(S)=25(1−p) - [X] B. E(S)=25p - [ ] C. E(S)=√(25p(1−p)) - [ ] D. E(S)=p

2. Polling - standard error of S

Again, consider the random variable S, which is the total number of Democrats in your sample of 25 voters. The variable p describes the proportion of Democrats in the sample, whereas 1−p describes the proportion of Republicans.

What is the standard error of S?

Possible Answers - [ ] A. SE(S)=25p(1−p) - [ ] B. SE(S)=√25p - [ ] C. SE(S)=25(1−p) - [X] D. SE(S)=√(25p(1−p))

3. Polling - expected value of X-bar

Consider the random variable **S/N**, which is equivalent to the sample average that we have been denoting as **X**. The variable **N** represents the sample size and **p** is the proportion of Democrats in the population.

What is the expected value of **X**?

Possible Answers - [X] A. E(X)=p - [ ] B. E(X)=Np - [ ] C. E(X)=N(1−p) - [ ] D. E(X)=1−p

4. Polling - standard error of X-bar

What is the standard error of the sample average, **X**?

The variable **N** represents the sample size and **p** is the proportion of Democrats in the population.

Possible Answers - [ ] A. SE(X)=√(Np(1−p)) - [X] B. SE(X)=√(p(1−p)/N) - [ ] C. SE(X)=√(p(1−p)) - [ ] D. SE(X)=√N