

Introduction to Model-based Statistics

Frequentist and Bayesian Reasoning with Simulation

Dr. Cheng-Han Yu

2025-12-07

Table of contents

Welcome	3
Three ways to think about inference	3
What is different about this book	4
Who this book is for	5
Using this book in different courses	5
1 Introduction	7
2 Summary	8
References	9

Welcome

This book offers an introduction to statistics built around a model based view of data and uncertainty. Rather than treating descriptive statistics, probability, and inference as separate topics, we will return again and again to a simple organizing idea:

$$Y_i = f(x_i) + \varepsilon_i.$$

In words, we assume that each observation Y_i can be described by

- information about the unit x_i (such as group membership or a predictor),
- a systematic part $f(x_i)$ that links x_i to the typical value of Y_i ,
- a random error term ε_i that captures variability we do not explain.

Even the basic problem of estimating a single population mean fits into this framework: it is the case where $f(x_i)$ is a constant. More complex tasks such as comparing groups or fitting a regression line simply enrich the form of $f(x)$ and the information available in x .

This model based perspective is the backbone of the entire book. It allows us to present classical frequentist methods, simulation based methods, and Bayesian methods as three complementary ways to answer the same inferential questions.

Three ways to think about inference

For almost every inferential problem in this book, you will see three approaches side by side.

- **Distribution-based methods**

We treat the parameters in the model as fixed but unknown quantities and study the long run behavior of procedures. This leads to familiar ideas such as standard errors, confidence intervals, hypothesis tests, and p values. Distribution based methods use mathematical results about sampling distributions, such as the Central Limit Theorem and the t distribution.

- **Simulation-based methods**

We use the computer to mimic what could have happened under repeated sampling. Bootstrap methods approximate the sampling distribution of statistics directly from the data. Permutation and randomization tests approximate the distribution of test statistics under a null hypothesis. Simulation provides intuition for frequentist ideas and is often easier to extend to new situations than purely algebraic formulas.

- **Bayesian methods**

We treat the unknown parameters in the model as random quantities and specify prior distributions that represent our initial beliefs or information. The data update these priors through the likelihood to produce posterior distributions. From the posterior we obtain point estimates, credible intervals, and probability statements about parameters and predictions that have a direct and intuitive interpretation.

The same model $Y_i = f(x_i) + \varepsilon_i$ underlies all three perspectives. What differs is how we treat the unknown parts of the model and how we quantify uncertainty.

What is different about this book

Many introductory statistics texts focus almost entirely on distribution based frequentist methods and mention simulation or Bayesian ideas only briefly, if at all. This book aims to be different in several ways.

- **Model based from the start**

One sample problems, two sample comparisons, analysis of variance, and regression are all presented as special cases of a common model based framework. This makes it easier to see connections between topics and to transition to more advanced modeling.

- **Frequentist and Bayesian reasoning at the introductory level**

Bayesian thinking is introduced early in simple settings, such as proportions and means, using graphical and computational tools rather than heavy algebra. Students see both confidence intervals and credible intervals, and they learn how to interpret each clearly.

- **Simulation as a first class tool**

Bootstrap intervals and permutation tests are not add ons at the end of the course. They are used throughout to build intuition for sampling variability, to check the behavior of classical methods, and to extend inference to situations where standard formulas do not apply.

- **Comparisons across methods**

For key inferential tasks, such as inference for one mean, differences in means, proportions, and simple regression, the book presents distribution based, simulation based, and Bayesian approaches in a common structure. Each method is described in terms of its assumptions, interpretation, and strengths and limitations. Students see how the answers are similar, how they differ, and why.

Who this book is for

This book is designed for

- students in a first course in statistics or data analysis who want a conceptually rich and modern introduction,
- instructors who wish to integrate simulation and Bayesian ideas into an introductory course without giving up core frequentist content,
- readers in fields such as the social sciences, health sciences, and data science who want a unified model based view of statistical inference.

The mathematical level assumes familiarity with high school algebra. Calculus is not required, although some optional sections sketch connections for interested readers. Programming experience is not required too.

Using this book in different courses

The book is intentionally more comprehensive than a typical one term course. It is written to support several different paths:

- a short quarter course that emphasizes data, visualization, and a gentle introduction to inference,
- a one semester course that focuses mainly on frequentist methods, supported by simulation and brief Bayesian examples,
- a one semester course that uses frequentist methods as a starting point and emphasizes Bayesian reasoning,
- a two semester sequence in which a first term covers frequentist and simulation based inference and a second term develops Bayesian modeling more deeply.

In the next section, “Recommended paths,” we outline concrete chapter selections for these different uses and indicate which sections are core, which are extensions, and which are more advanced.

Throughout the book, recurring visual conventions and section labels will help you recognize which material is frequentist, which is simulation based, which is Bayesian, and which is more advanced. This is intended to make the text flexible for instructors while giving students a coherent and connected experience.

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

add new words here

2 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```

References

- Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.