**Statistical Rethinking 2023 – Summary version**

Table des matières

[Chapter 1: The Golem of Prague 2](#_Toc204240401)

[Summary 2](#_Toc204240402)

[Chapter 2: Small Worlds and Large Worlds 2](#_Toc204240403)

[Summary 2](#_Toc204240404)

[Practice 3](#_Toc204240405)

[Easy exercises 3](#_Toc204240406)

[Medium exercises 3](#_Toc204240407)

[Hard exercises 3](#_Toc204240408)

[Chapter 3: Sampling the Imaginary 4](#_Toc204240409)

[Summary 4](#_Toc204240410)

[Practice 4](#_Toc204240411)

[Chapter 4: Linear Models 4](#_Toc204240412)

[Summary 4](#_Toc204240413)

[Practice 4](#_Toc204240414)

[Chapter 8: Markov Chain Monte Carlo (MCMC) 5](#_Toc204240415)

[Summary 5](#_Toc204240416)

[Practice 5](#_Toc204240417)

# Chapter 1: The Golem of Prague

## Summary

This chapter lays the foundation for the book's approach. The author compares statistical models to "golems"—powerful creatures from Jewish mythology that faithfully execute commands but lack wisdom and judgment. In other words, a statistical model is a very powerful tool for analyzing data, but it cannot think for itself. It will do exactly what you tell it to do, even if it's not what you truly intended or if it leads to absurd conclusions.

The central idea is that statistics is not just a science of discovery, but also a branch of engineering. Instead of simply choosing a "statistical test" from a manual (like picking a wrench from a toolbox), the author encourages us to learn how to *build* specific procedures and refine them to answer our questions. This means moving away from the traditional approach of "testing the null hypothesis" (the idea that there is no effect or relationship) and instead focusing on building and analyzing *multiple* different models that represent various possibilities.

The chapter also introduces three fundamental tools that will be explored in depth later:

1. **Bayesian inference**: an approach to statistics that allows us to update our beliefs about phenomena as we receive new data.
2. **Multilevel models**: models that allow us to handle data structured hierarchically (e.g., students in different classes, classes in different schools).
3. **Information-theoretic model comparison**: a method for evaluating and comparing different models to determine which one best explains the data, without being limited to rejecting or accepting a single hypothesis.

# Chapter 2: Small Worlds and Large Worlds

## Summary

This crucial chapter addresses the fundamental distinction between theory and reality. The author uses the analogy of Christopher Columbus who, in seeking a route to Asia, underestimated the size of the Earth. The "small world" is the statistical model itself: it is a logical, self-contained, and simplified system of reality. In this small world, we define rules, assumptions, and relationships. It is the controlled and predictable environment we create to analyze our data.

The "large world," on the other hand, is complex and unpredictable reality. This is where unexpected things happen, where our assumptions might be wrong, and where our models can fail. A model is never an exact replica of reality; it is always a simplification. Consequently, any statistical model will inevitably make mistakes or fail to capture all the nuances of the real world. The challenge for the statistician is to understand this distinction and to know how to navigate between the simplified logic of the model and the unpredictable complexity of the real world, in order to avoid drawing erroneous conclusions.

## Practice

### Easy exercises

E1: The probability of rain on Monday = Pr(rain|Monday) (2)

E2: Pr(Monday|rain) = the probability that it is Monday, given that is raining (3)

E3: the probability that is Monday, given that is raining = Pr(Monday|rain) (1)

E4: I don’t understand the question

### Medium exercises

M1:

1. W, W, W:
2. W, W, W, L:
3. L, W, W, W, L, W, W:

M2:

M3:

We want .

We know:

We use Bayes formula:

We use the total probability formula to get the probability of land :

Finally:

M4:

M5:

M6:

M7:

### Hard exercises

H1:

H2:

H3:

H4:

# Chapter 3: Sampling the Imaginary

## Summary

This chapter delves into the practice of Bayesian inference. It explains how to work with "samples" from the "posterior distribution." Simply put, the posterior distribution is how our beliefs about a phenomenon change after observing data. Imagine you have an idea of the average height of a population before measuring anyone; after measuring a few people, your initial idea is updated and refined—this updated idea is the posterior distribution.

The Bayesian approach uses simulations to "sample" this posterior distribution. This means that instead of solving complex mathematical equations, we can simply generate many examples ("samples") of what reality might look like given our data and initial beliefs. This method transforms complex calculation problems into much simpler and more intuitive data summarization problems. Even for simple models, the author emphasizes the importance of using this sampling approach. This allows scientists to ask more sophisticated questions and explore their models without needing an advanced math degree.

## Practice

### Easy exercises

E1:

E2:

E3:

E4:

E5:

E6:

E7:

### Medium exercises

M1:

M2:

M3:

M4:

M5:

### Hard exercises

H1:

H2:

H3:

H4:

H5:

# Chapter 4: Linear Models

## Summary

This chapter introduces linear models, a cornerstone of statistics. A linear model is a tool that helps us understand how one variable (called the dependent variable or response) changes in relation to one or more other variables (called independent variables or predictors). The idea is to find a line (or a plane in multiple dimensions) that best describes the relationship between these variables.

The chapter explains why the normal distribution (the famous bell curve) is so often used in statistics. It then shows how to build and describe statistical models using a specific language. You will learn to create a Gaussian model (based on the normal distribution) to predict, for example, individuals' height. Then, the chapter guides you on how to add predictors to this model (for instance, how height is affected by age or sex) and how to use polynomial regression, which allows modeling relationships that are not strictly straight (e.g., height might increase rapidly at first, then slow down).

## Practice

# Chapter 8: Markov Chain Monte Carlo (MCMC)

## Summary

This chapter introduces a fundamental and powerful technique for Bayesian inference: Markov Chain Monte Carlo (MCMC). When statistical models become complex, it becomes impossible to directly calculate the posterior distribution (our updated beliefs). This is where MCMC comes in.

MCMC is a method that allows us to "jump" and "explore" the posterior distribution randomly, but intelligently, to draw representative samples from it. The major advantage is that there's no need to make assumptions about the shape of this posterior distribution (e.g., that it is Gaussian). MCMC can work even for very complex distributions. The chapter highlights the efficiency of certain MCMC variants, particularly Hamiltonian Monte Carlo (HMC), which is especially powerful for models with many parameters. It explains that software tools like Stan automate much of the complex tuning of HMC, making this method accessible. Finally, the chapter provides practical advice on how to run Markov chains and check if they have converged properly (i.e., if they have explored the distribution enough to give reliable results).

These chapters will guide you step by step in understanding Bayesian statistics and modeling, giving you the tools to build and interpret your own statistical "golems" with greater wisdom.

## Practice

Meeting with Conor:

p(a|b) = p(b|a)p(a) / p(b)

p(b|a) is the model!

p(a) is the prior where a is the model parameters

p(a|b) is the parameters given the data

b is the data

p(b)?

p(b) = integral\_A da p(b,a)

p(b,a) = p(b|a)p(a)

p(b) = integral\_A p(b|a) p(a) da

MCMC allows us to sample p(a|b) if we know p(b|a) and p(a)!

NUTS

HMC

MCMC - the idea of sampling

Metropolis Hastings : the first idea about how to propose samples.

HMC - Hamiltonian / Hybrid Monte Carlo

a chain is lots of a's satisfying p(aZb)

in the space of possible a's it jumps around in a way that matches the distribution p(a|b)

it does that a bit like a ball rolling around on hills that look like -log(p(a|b))

r-hat

say one of the parameters in what I'm calling a is a mean, say mu; mu\_t and mu\_n

mu\_t> mu\_n say that's your hypothesis

mu\_t - mu\_n for all the 1000 or 4000 values of mu's in your chain and you count how often is this positive and how often is it negative

say if is bigger than zero more than 950 times out of 1000 then you'd say that the difference is significant.

email with some suggested times for Thursday?