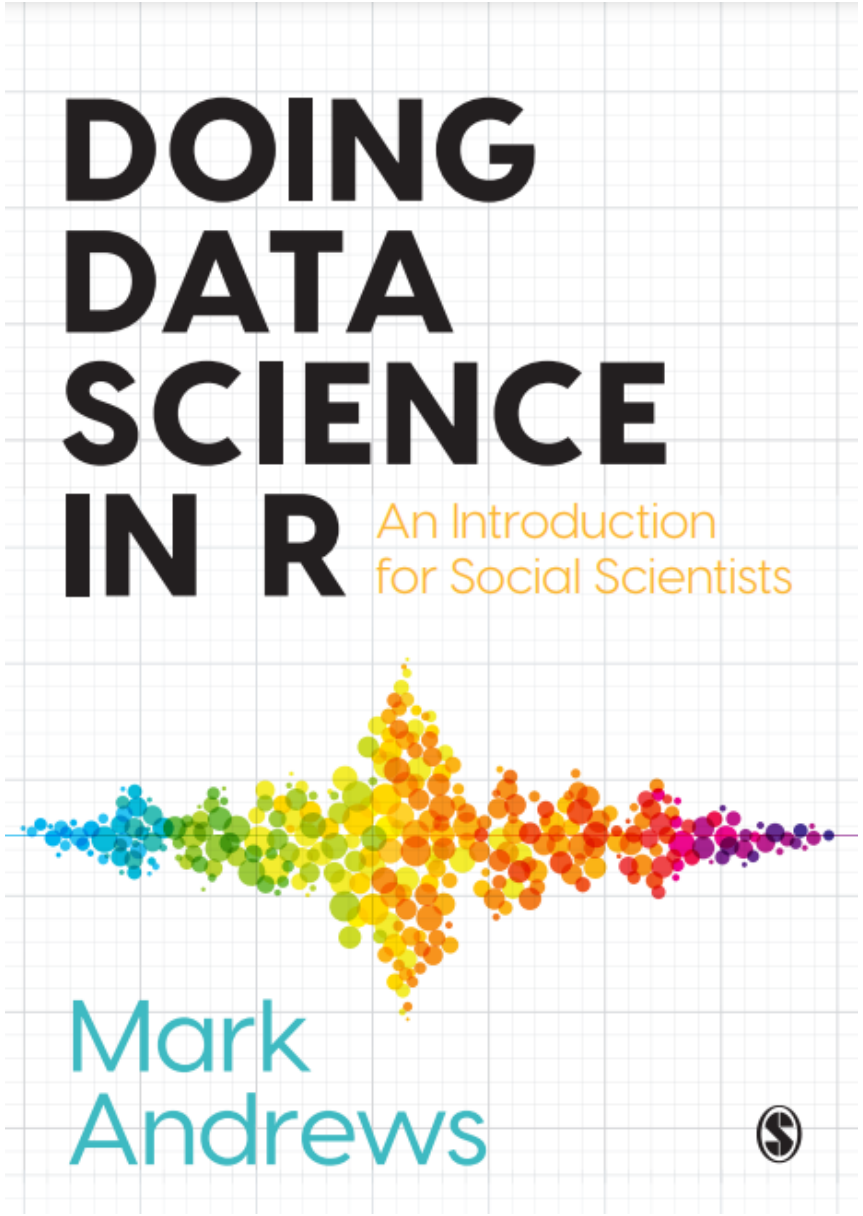

Doing Data Science in R: An Introduction for Social Scientists

Chapter 8

Statistical Models and Statistical Inference

Introduction

Exploratory data analysis aims to describe and visualize the data and to identify possible trends and patterns in it.

Statistical models, by contrast, are mathematical models of the population from which the data originated.

Statistical inference

The first approach is variously referred to as the classical, frequentist, or sampling theory based approach.

The second is the Bayesian approach. The classical approach is still the dominant one in practice, particularly for the more introductory or medium-level topics

Classical statistical inference

Classical statistical inference begins with an estimator of the value of θ, denoted by $\hat{\theta}$, and then considers the sampling distribution of $\hat{\theta}$ for any hypothetical true value of θ.

Informally speaking, we can see the estimator $\hat{\theta}$ as an educated guess of what the true value of θ is. T

Maximum likelihood estimation

The maximum likelihood estimator of θ is the value of θ that maximizes the likelihood function.

Sampling distribution of θ

The maximum likelihood estimator can be seen as a random variable.

It is a deterministic function of the observed data m, but m would vary were we repeat the experiment even under identical circumstances

p-values

Informally speaking, a p-value tells us whether the value of an estimator, is consistent with some hypothetical value of the unknown variable.

If the p-value is low, the estimator’s value is not consistent with the hypothesized value.

Bayesian statistical inference

Bayesian approaches to statistical inference ultimately aim to solve the same problem as classical approaches

While the classical approach is based on calculation of estimators and their sampling distributions, Bayesian approaches rely on an eighteenth-century mathematical result known as Bayes' rule or Bayes' theorem