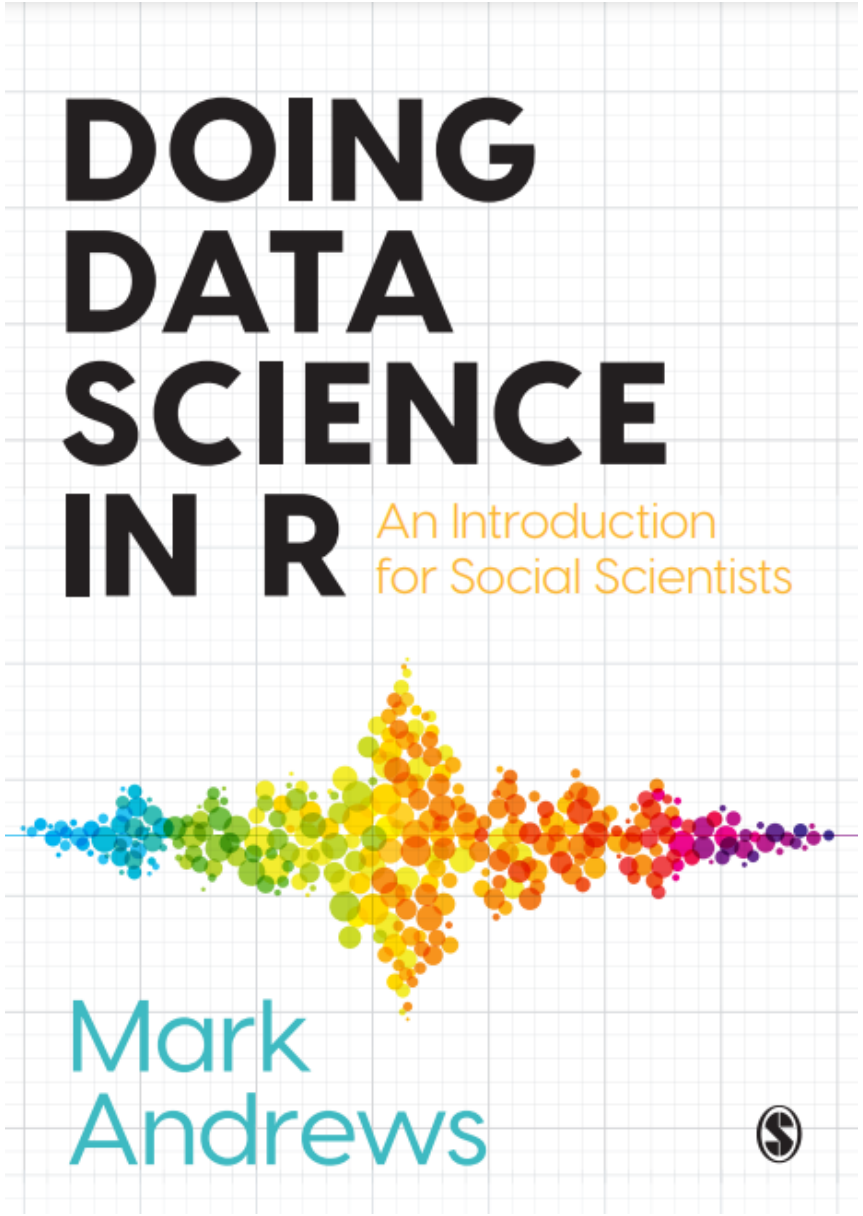

Doing Data Science in R: An Introduction for Social Scientists

Chapter 9

Normal Linear Models

Introduction

Normal linear models play a foundational role in statistical modelling.

In a sense, they can be seen as the backbone of most statistical modelling techniques.

The univariate normal linear model

In these models, we assume we have n independent observations, which can be represented as n pairs as follows:

$$(y_1, \vec{x}_1), (y_2, \vec{x}_2), \ldots, (y_i, \vec{x}_i), \ldots, (y_n, \vec{x}_n).$$

Classical approaches to normal linear models

Given observed values of an outcome variable y1,…,yn, and given n corresponding vectors of K predictor variables □ □ x x 1 n ,..., , and if we model y1,…,yn using the normal linear model

$$y_i \sim \mathrm{N}\left(\mu_i, \sigma^2\right), \quad \mu_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}, \quad \text{for } i \in 1,\ldots,n,$$

Maximum likelihood estimation using lm

The main command for doing normal linear modelling in R is lm. This is probably the most widely used statistical modelling command in R.

$$y_i \sim \mathrm{N}\left(\mu_i, \sigma^2\right), \quad \mu_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}, \quad \text{for } i \in 1, \ldots, n,$$

Bayesian approaches to normal linear models

In the Bayesian approach to normal linear models, our starting point is identical to that of the classical approach.

Specifically, we assume we have n independent observations that can be represented as

$$(y_1, \vec{x}_1), (y_2, \vec{x}_2), \ldots, (y_i, \vec{x}_i), \ldots, (y_n, \vec{x}_n),$$

Monte Carlo approaches

As mentioned, in situations where a closed-form expression for the posterior distribution is not available, we may use Monte Carlo methods to draw samples from this distribution

Categorical predictor variables

The important feature of these variables is that they are defined on a metric space

As a simple example, again using weight as our outcome variable, we could have a single explanatory variable gender, which takes on two categorically distinct values: male, female.