Chapter 11.1 Introduction

Jim Albert and Monika Hu

Chapter 11 Simple Linear Regression

Review

- Continuous response variables
  - Roger Federer's time-to-serve data in Chapter 8
  - snowfall amounts in Buffalo, New York in Chapter 9
- Normal sampling models have been applied
  - observations are identically and independently distributed (i.i.d.) according to a Normal density

$$Y_i \overset{i.i.d.}{\sim} \mathrm{Normal}(\mu, \sigma) \tag{1}$$

- What if μ_i is different for each record i?

Adding a predictor variable

- It is common that other variables are recorded that may be associated with the primary response measure
- The tennis example: the rally lengthof the previous point
- The Buffalo snowfall example: the average temperature in winter season

Adding a predictor variable cont'd

- A Normal curve for modeling the snowfalls $Y_1, ..., Y_n$ for n winters (Chapter 9)

$$Y_i \mid \mu, \sigma \overset{i.i.d.}{\sim} \mathrm{Normal}(\mu, \sigma), \;\; i = 1, \cdots, n \tag{2}$$

- Additional information: average temperature in winter i, x_i
- Weather the snowfall amount Y_i can be explained by the average temperature x_i in the same winter?
- x_i: a predictor variable

An observation-specific mean

- To intorudce a new variable in the sampling model: the common mean μ is replaced by a winter specific mean μ_i

$$Y_i \mid \mu_i, \sigma \overset{ind}{\sim} \mathrm{Normal}(\mu_i, \sigma), \ i = 1, \cdots, n \tag{3}$$

- The observations $Y_1, ..., Y_n$ are no longer identically distributed since they have different means, but the observations are still independent

Linear relationship between the mean and the predictor

- One basic approach for relating a predictor x_i and the response Y_i: assume that the mean of Y_i, μ_i, is a linear function of x_i
$$\mu_i = \beta_0 + \beta_1 x_i, \tag{4}$$
for $i = 1, \ldots, n$
- Each x_i is a known constant (that is why a small letter is used for x)
- β_0 and β_1 are unknown parameters (Bayesian approach: prior + data $\rightarrow$ posterior)

Linear relationship between the mean and the predictor cont'd

- The linear function $\beta_0 + \beta_1 x_i$ is interpreted as the **expected** snowfall amount when the average temperature is equal to x_i
- The intercept β_0 represents the expected snowfall when the winter temperature is $x_i = 0$
- The slope parameter β_1 gives the increase in the expected snowfall when the temperature x_i increases by one degree

Linear relationship between the mean and the predictor cont'd

- The linear function $\beta_0 + \beta_1 x_i$ is interpreted as the **expected** snowfall amount when the average temperature is equal to x_i
- The intercept β_0 represents the expected snowfall when the winter temperature is $x_i = 0$
- The slope parameter β_1 gives the increase in the expected snowfall when the temperature x_i increases by one degree
- This linear relationship is a statement about the **expected** or average snowfall amount μ_i, not the **actual** snowfall amount Y_i

Linear regression model

- One expression:

$$Y_i \mid \beta_0, \beta_1, \sigma \overset{ind}{\sim} \mathrm{Normal}(\beta_0 + \beta_1 x_i, \sigma), \;\; i = 1, \cdots, n \quad (5)$$

  - Y_i independently follow a normal density with observation specific mean $\beta_0 + \beta_1 x_i$ and common standard deviation σ
  - also known as simple linear regression (one predictor)

Linear regression model

- One expression:

$$Y_i \mid \beta_0, \beta_1, \sigma \overset{ind}{\sim} \mathrm{Normal}(\beta_0 + \beta_1 x_i, \sigma), \;\; i = 1, \cdots, n \quad (5)$$

  - Y_i independently follow a normal density with observation specific mean $\beta_0 + \beta_1 x_i$ and common standard deviation σ
  - also known as simple linear regression (one predictor)

- Another expression:

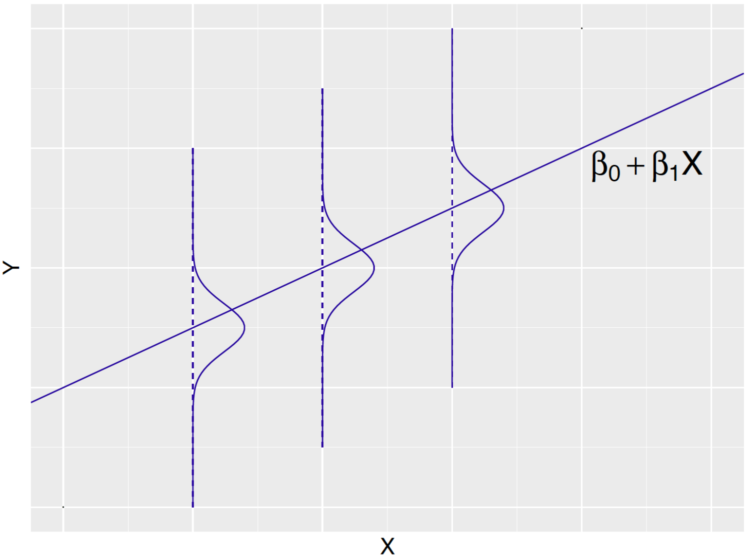
$$Y_i = \mu_i + \epsilon_i, i = 1, \cdots, n \quad (6)$$

  - the mean response $\mu_i = \beta_0 + \beta_1 x_i$ and the residuals $\epsilon_1, ..., \epsilon_n$ are $i.i.d.$ from a normal distribution with mean 0 and standard deviation σ

Linear regression model cont'd

- Our model: the snowfall for a particular season Y_i is a linear function of the average season temperature x_i plus a random error ϵ_i that is normal with mean 0 and standard deviation σ

Linear regression model cont'd



Summary

- The observation Y_i is random
- The predictor x_i is a fixed constant
- The unknown parameters are β_0, β_1, and σ

Summary

- The observation Y_i is random
- The predictor x_i is a fixed constant
- The unknown parameters are β_0, β_1, and σ
- The Bayesian paradigm:
  - a joint prior for $(\beta_0, \beta_1, \sigma)$
  - after the response values $Y_i = y_i, i = 1, ..., n$ are observed
  - MCMC estimation for $(\beta_0, \beta_1, \sigma)$ to get posterior
  - posterior summarization for inferences

Summary cont'd

- Some inference questions:
  - learning about the relationship between the average temperature and the mean snowfall that is described by the linear model $\mu = \beta_0 + \beta_1 x$
  - the posterior probability of $\beta_1 < 0$: what can we learn?
  - predicting future snowfall amount