12.2: Estimating Model Parameters

Taylor

University of Virginia

Introduction

With an SLR model $Y = \beta_0 + \beta_1 x + \epsilon$, we don't know $\beta_0, \beta_1, \sigma^2$. We have to use the data to estimate these. That's what this chapter is about.

We don't say anything about how to fit logistic regression models here.

Motivation

Last class we showed that $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, for $i = 1, \ldots, n$. So

$$f(y_1, \ldots, y_n; \beta_0, \beta_1, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left[-\frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

is our likelihood that we want to maximize.

$$\log f(y_1, \ldots, y_n; \beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\sigma^2) - \frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$$

even though the book doesn't say this, we estimate using maximum likelihood (sort of).

Principle of Least Squares

We can estimate the βs and σ^2 separately. To estimate β_0 and β_1 we minimize this expression:

$$f(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$$

The $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize this are called the **least squares estimates** (they're also the same as the MLE estimates).

The **estimated regression line** is then

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Principle of Least Squares

$$
\begin{aligned}
\frac{\partial}{\partial \beta_0} f(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} (y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} f(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i)
\end{aligned}
$$

Principle of Least Squares

Setting both of these equal to 0 gives us the **normal equations**:

$$n\beta_0 + \left(\sum x_i\right)\beta_1 = \sum y_i$$

$$\left(\sum x_i\right)\beta_0 + \left(\sum x_i^2\right)\beta_1 = \sum x_i y_i$$

Then we solve for β_0 and β_1

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

so SLR admits a **closed-form** expression for the estimated coefficients.

Estimating σ^2

First a definition:

The **fitted (predicted) values**, denoted $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$, are what you get when you plug in $x_1, \ldots, x_n$ into the estimated regression line $\hat{\beta}_0 + \hat{\beta}_1 x$.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

for $i = 1, \ldots, n$.

Estimating σ^2

The **error (residual) sum of squares** is

$$\text{SSE} = \sum_i (y_i - \hat{y}_i)^2$$

and our estimate for the variance of ϵ is

$$\hat{\sigma}^2 = s^2 = \frac{\text{SSE}}{n-2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$$

Note: this ISN'T the MLE estimate of σ^2. This one is unbiased, though, so we use this one instead.

Coefficient of Determination

Think of SSE as the variability in Y that isn't explained by X. The total variability in Y is the **total sum of squares**

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The coefficient of determination R^2 is the proportion of the total variability that is explained (higher is better)

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

with this interpretation it should be easy to remember

$$0 \leq R^2 \leq 1$$

Coefficient of Determination

$SST - SSE$ has a name. It is called the **regression sum of squares** (SSR).

$$SSR = \sum(\hat{y}_i - \bar{Y})^2.$$

A homework exercise will be to prove the following very important identity:

$$SST = SSR + SSE;$$

so total variation can be broken down into good and bad variation (SSR and SSE, respectively).

If we plug this into the last slide's formula, we get another expression for R^2

$$R^2 = \frac{SSR}{SST}$$

Coefficient of Determination

The book calls the coefficient of determination r^2. I'm calling it R^2. Typically R^2 denotes exactly what we defined $\frac{SSR}{SST}$, whereas r^2 denotes the sample correlation, squared $\left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}\right)^2$.

In the case of SLR, when we have one predictor, $r^2 = R^2$. However, it is not always clear to talk about r when we talk about multiple linear regression (MLR). This is a regression model that has more than one predictor/input/covariates. You can't calculate the correlation between Y and more than one set of Xs, right?