The OLS Estimator (2)

STA 211: The Mathematics of Regression

Yue Jiang

January 17, 2023

Review: the OLS estimate

In the linear regression model $\mathbf{y} = \mathbf{X}\beta + \epsilon$, the ordinary least squares (OLS) estimate $\widehat{\beta}$ minimizes the mean squared error $(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$ between the predicted outcomes and the observed outcomes. The estimate is given by:

$$\widehat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

- What actually *is* a "linear model" anyway?

Linear regression models

Which of the following (if any) depict a relationship that can be considered a "linear regression model"?

1. $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3 x_{i2} + \epsilon_i$
2. $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$
3. $y_i = \beta_0 + \frac{\beta_1 x_{i1}}{\beta_2 x_{i2} + \beta_3 x_{i3}} + \epsilon_i$
4. $y_i = \beta_0 + \beta_1 x_{i1}^{(x_{i2} + x_{i3})} + \epsilon_i$
5. $y_i = \beta_0 + \beta_1 \sin(x_{i1} + \beta_2 x_{i2}) + \beta_3 x_{i3} + \epsilon_i$
6. $y_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}) + \epsilon_i$
7. $y_i = \beta_1 x_{i1}^{(x_{i2} + \beta_2 x_{i3})} + \epsilon_i$
8. $y_i = \beta_0 + \beta_1 \cos(x_{i1}) + \beta_2 \sin(x_{i2}) + \beta_3 x_{i3}^{1/2} + \epsilon_i$
9. $y_i = \beta_1 e^{x_{i1}} + \beta_2 e^{x_{i2}} + \epsilon_i$
10. $y_i = \beta_0 + \beta_1 e^{\beta_2 x_{i1}} + \beta_3 x_{i2} + \epsilon_i$

Linear regression models

Linear regression models are linear *in the parameters*. That is, for a given observation Y_i:

$$Y_i = \beta_0 + \beta_1 f_1(X_{i1}) + \beta_2 f_2(X_{i2}) + \cdots + \beta_p f_p(X_{ip}) + \epsilon_i$$

The functions $f_1, \cdots f_p$ may themselves be non-linear, but as long as the $\boldsymbol{\beta}$ are linear in $\mathbf{y}$, we have a linear regression model.

- Why would we want to use any function such that $f_k(u) \neq u$?
- What about $y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$?

Transforming predictors

Still technically a "linear regression model":

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \epsilon_i$$

Example: linear model interpretations

Let's consider some various regression functions (most of them linear). What happens when x_1 changes in some various models?

$$\frac{\partial}{\partial x_1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) = \beta_1$$

$$\frac{\partial}{\partial x_1}\left(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2\right) = \beta_1 + 2\beta_2 x_1$$

$$\frac{\partial}{\partial x_1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2) = \beta_1 + \beta_3 x_2$$

$$\frac{\partial}{\partial x_2}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2) = \beta_2 + \beta_3 x_1$$

$$\frac{\partial}{\partial x_1}\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) = ...?$$

That other model

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$$
$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Although the RHS is a linear function of β, we do not have a linear model *for* y; it is linear in $\log(y)$.

This type of model (i.e., a linear relationship between $\boldsymbol{\beta}$ and an invertible function of y) is known as a *generalized* linear model (well, technically with the conditional expectation of Y, but more on that later), and we will study this class of model in the second half of the semester

Interlude

A quick break to visit the other class I'm teaching this semester...

STA 210 stuff...

Interlude

Back to our regularly scheduled programming.

By the way, if you're very curious about this type of thing, in a few stats courses from now (perhaps in graduate school?), you might want to check out the notion of **M-estimators** *(and then help clean up the not-so-well-written Wikipedia page for them!).*

Example: linear model interpretations

Let's focus on the "interaction model"

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i3} + \epsilon_i$$

- What might we expect from a change in either of the predictor variables?
- How might we use this intuition to interpret a model with these so-called "interaction terms"?

Unpacking the design matrix

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

- How might you deal with a *categorical* predictor, say with k levels?

Dummy variables

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

R, RStudio, etc.

Matrix operations in R

- `as.matrix()` function sets an object as a matrix object in R
- `%*%` is the matrix multiplication operation (e.g., `A %*% B` for two matrices A and B)
- `t()` function takes the transpose of a matrix
- `solve()` function inverts a matrix

Matrix operations in R

- `install.packages("palmerpenguins")`
- `library(palmerpenguins)`
- `head(penguins)`

A basic regression model

$$(\text{Body mass})_i = \beta_0 + \beta_1(\text{Flipper length})_i + \beta_2(\text{Bill length})_i + \epsilon_i$$

```
> summary(lm(body_mass_g ~ flipper_length_mm + bill_length_mm, data = penguins))

Call:
lm(formula = body_mass_g ~ flipper_length_mm + bill_length_mm,
    data = penguins)

Residuals:
    Min      1Q  Median      3Q     Max
-1090.5  -285.7   -32.1   244.2  1287.5

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -5736.897    307.959 -18.629   <2e-16 ***
flipper_length_mm    48.145      2.011  23.939   <2e-16 ***
bill_length_mm        6.047      5.180   1.168    0.244
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 394.1 on 339 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:   0.76,     Adjusted R-squared:  0.7585
F-statistic: 536.6 on 2 and 339 DF,  p-value: < 2.2e-16
```

- $\hat{\beta}_0$ = -5736.897; $\hat{\beta}_1 = 48.145$; $\hat{\beta}_2 = 6.047$
- Can you recover these estimates *from the dataset directly* using matrix operations?

Homework 2: Due Jan. 31

1. Must there always be a linear relationship between some predictor x_k and the outcome y in a linear regression model? If yes, provide a proof; if no, provide a counterexample of such a model and clearly demonstrate a non-linear relationship between the two.
2. Create a linear model in R that predicts the body mass of a penguin based on its flipper length, bill length, *and* which species it is. Provide your design matrix, and clearly label what each column corresponds to. Display the estimated regression coefficients from the `lm` function, and recover these same estimates from the dataset directly using matrix operations on the underlying data.