Properties of Least Square Estimates

DS 301

Iowa State University

Today's Agenda

- Implementation of multiple linear regression in `R`

- Properties of least square estimates

Multiple Linear Regression (Recap)

$$Y = f(X) + \varepsilon$$

- If we assume a linear relationship btwn X & Y

$X \downarrow X_1, X_2, \ldots, X_p$

$$f(X) = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_p X_p.$$

↳ $Y = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_p X_p + \varepsilon$

(true population regression line)

$B_0, B_1, B_2, \ldots, B_p$ are unknown parameters.

$\Longrightarrow$ estimate them from (training) data

$\hat{B}_0, \hat{B}_1, \hat{B}_2, \ldots, \hat{B}_p$

(least squares criterion) $\left\{ \sum_{i=1}^{n} (y_i - \underline{\hat{y}_i})^2 : \text{RSS} \right.$

$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + \cdots + \hat{B}_p X_{ip}$

↑ predicted value / fitted value

let's find $\hat{y}_i$ that minimizes RSS

Interpretation of least squares coefficients

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_{i1} + \cdots + \hat{B}_p X_{ip}$$

is an estimate for $E(Y)$, not Y.

$\hat{\beta}_j$ can be interpreted as the average change in Y associated with a 1 unit change in X_j, *holding all other predictors constant*.

Assumptions

(1) There is a linear relationship btwn $X_1, X_2, \ldots, X_p$ and Y

(2) $E(\varepsilon_i) = 0$

(3) $Var(\varepsilon_i) = \sigma^2$

(4) ε_i's are uncorrelated

Implementation in `R`

See R script `IntroMLR.R`

Properties of least square estimators

- Remember in real applications, the true parameters $\beta_0, \beta_1, \dots, \beta_p$ are unknown to us.

- Ideally, we hope our least square estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ are close to the true values of $\beta_0, \beta_1, \dots, \beta_p$.

- We can quantify how 'close' our estimates are to the truth using the following concepts:
    - Bias
    - Standard error

Properties of least square estimators

How good are our estimates?

The least sq. estimates $\hat{B}_0, \hat{B}_1, \ldots, \hat{B}_p$ are unbiased estimates of $B_0, B_1, \ldots, B_p$ (respectively).

conceptually, the property of unbiasedness says that if we took the average of $\hat{B}_0, \hat{B}_1, \ldots, \hat{B}_p$ obtained over a huge number of datasets, then these averages would exactly equal the true parameters $B_0, B_1, \ldots, B_p$

formal definition: $E(\hat{B}_0) = B_0$, $E(\hat{B}_1) = B_1$, $\ldots$ $E(\hat{B}_p) = B_p$

Properties of least square estimators

$\hat{B_0}, \hat{B_1}, \ldots, \hat{B_p} \Rightarrow \hat{Y}$

$\hat{Y}$ is an unbiased estimate for $E(Y)$.

$$E(\hat{Y}) = E(\hat{B_0} + \hat{B_1} X_1 + \cdots + \hat{B_p} X_p)$$
$$= B_0 + B_1 X_1 + \cdots + B_p X_p = E(Y).$$

$$E(\hat{Y}) = E(Y)$$

↳ on average, our predictions $(\hat{Y})$ will equal the true mean of Y: $E(Y)$

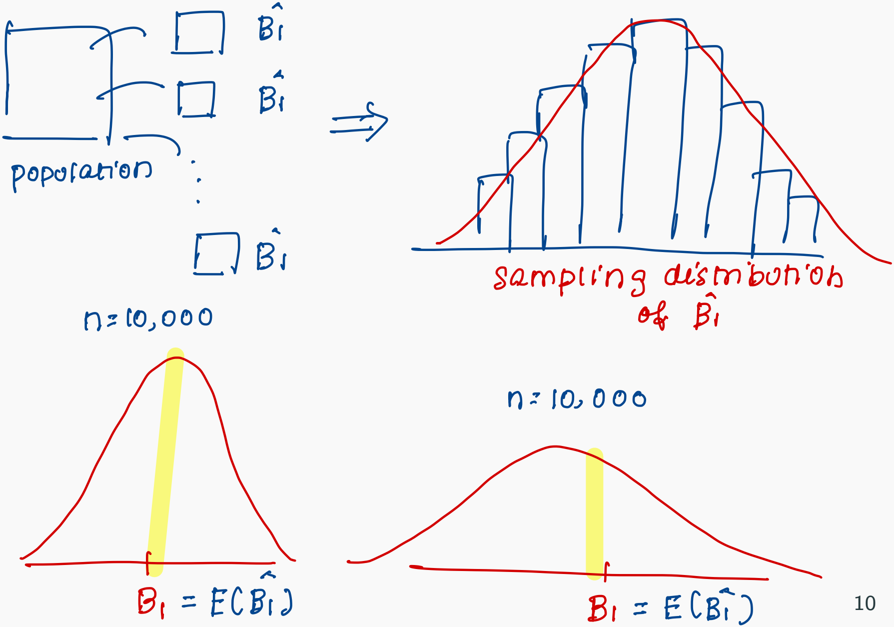
↳ $f(X)$

$\Rightarrow$ This gives us some sense that we can trust our model.

Accuracy of $\hat{\beta}$

- The unbiasedness property tell us that the average of our estimates from many many datasets will be very close to the true population parameter $\boldsymbol{\beta}$.
- But we don't have access to many many datasets. For a particular data set, the single estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ may be a substantial underestimate or overestimate of the true $\beta_0, \beta_1, \ldots, \beta_p$.
- How far off will our single estimate $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ be?

Accuracy of $\hat{\beta}$ (standard error)



Accuracy of $\hat{\beta}$: How far off is a single estimate?

$\Rightarrow$ formalize this with the standard error of $\hat{B}_0, \hat{B}_1, \ldots, \hat{B}_p$.

$$se(\hat{B}_0) = \sqrt{\sigma^2 / \Sigma(x_i - \bar{x})^2}$$

$$se(\hat{B}_1) = \sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{\Sigma(x_i-\bar{x})^2}\right)}$$

$\vdots$ $\hat{B}_2$ $\vdots$ $\hat{B}_p$

σ^2 is a parameter. It represents the variance of Y.

$$Y = B_0 + B_1 X_1 + \varepsilon$$

$$\text{var}(Y) = \text{var}(B_0 + B_1 X_1 + \varepsilon)$$

$$= \underbrace{\text{var}(B_0)}_{0} + \underbrace{\text{var}(B_1 X_1)}_{0} + \underbrace{\text{var}(\varepsilon)}_{\sigma^2}$$

$\hookrightarrow$ σ^2 needs to be estimated from data.