Properties of the LS estimates

ST552 Lecture 6

Charlotte Wickham

January 15, 2016

The least squares estimates are unbiased

Go over last time's fill in the blanks.

The least squares estimates have variance-covariance matrix

$$\text{Var}\left(\hat{\beta}\right) =$$

We can pull out the variance of individual parameters from the diagonal, and covariances from the off diagonals:

$$\text{Var}\left(\hat{\beta}_j\right) = \sigma^2 \left((X^T X)^{-1}\right)_{j,j}$$
$$\text{Cov}\left(\hat{\beta}_j, \hat{\beta}_k\right) = \sigma^2 \left((X^T X)^{-1}\right)_{j,k}$$

(Careful: the way I did the notation, the row and column indexing starts at 0 (i.e. for β_0))

Gauss Markov Theorem

You might wonder if we can find estimates with better properties. The Guass-Markov theorem says the least squares estimates are BLUE.

> *Of all linear, unbiased estimates, the least squares estimates have the smallest variance.*

Of course if you are willing to let go of linear and/or, unbiasedness you might be able to find an estimate with smaller variance.

Gauss Markov Theorem (proof)

Estimating σ

To make use of the variance-covariance results we need to be able to estimate σ^2.

An unbiased estimate is:

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2 = \frac{||e||^2}{n-p}$$

Proof:

Standard errors

$$\sqrt{\widehat{\mathsf{Var}}\left(\hat{\beta}_j\right)} = \mathsf{SE}\left(\hat{\beta}_j\right) = \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{j,j}}$$

Summary

For the linear regression model

$$Y = X\beta + \epsilon$$

where $\mathsf{E}\left(\epsilon\right) = 0_{n\times 1}$, $\mathsf{Var}\left(\epsilon\right) = \sigma^2 I_n$, and the matrix $X_{n\times p}$ is fixed with rank p.

The least squares estimates are

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Furthermore, the least squares estimates are BLUE, and

$$\mathsf{E}\left(\hat{\beta}\right) = \beta, \qquad \mathsf{Var}\left(\hat{\beta}\right) = \sigma^2 (X^T X)^{-1}$$

$$\mathsf{E}\left(\hat{\sigma}^2\right) = \mathsf{E}\left(\frac{1}{n-p}\sum_{i=1}^{n} e_i^2\right) = \sigma^2$$

We have not used any Normality assumptions to show these properties.