Properties of the LS estimates

ST552 Lecture 6

Charlotte Wickham

January 15, 2015

The least squares estimates are unbiased

Go over last time's fill in the blanks.

Property of $\hat{\beta}$: <u>The least squares estimates of β are unbiased</u>

$$E(\hat{\beta}) = E\left((X^TX)^{-1}X^TY\right)$$

$$= E\left((X^TX)^{-1}X^T(X\beta + \varepsilon)\right) \qquad \text{plug in regression equation for } Y$$

$$= E\left(\underbrace{\cancel{(X^TX)^{-1}}}\,\underbrace{\cancel{X^TX}}\,\beta + (X^TX)^{-1}X^T\varepsilon\right) \qquad \text{expand}$$

$$(*) \qquad = E\left(\beta + (X^TX)^{-1}X^T\varepsilon\right) \qquad \text{simplify term on left } A^{-1}A = I$$

$$= \beta + (X^TX)^{-1}X^T E(\varepsilon) \qquad \text{linearity of expectation}$$

$$= \beta \qquad \text{regression assumptions } E(\varepsilon) = 0$$

The least squares estimates have variance-covariance matrix

$$\text{Var}\left(\hat{\beta}\right) = \sigma^2(X^TX)^{-1}$$

$$\underset{p\times 1}{\text{Var}(\hat{\beta})} = \text{Var}(\beta + (X^TX)^{-1}X^T\varepsilon) \quad \text{from } (*)$$

$$= (X^TX)^{-1}X^T\,\text{Var}(\varepsilon)\,X(X^TX)^{-1}$$

$$= (X^TX)^{-1}X^T\,\sigma^2 I\,X(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1}_{p\times p}$$

$\text{Var}(AU+b)$

$= A\,\text{Var}(U)A^T$

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

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

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

$\text{Var}(a\hat{\beta})$

$\text{Var}(\hat{\beta}_1 - \hat{\beta}_2)$

$= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2)$

$- 2\,\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$

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)

Let $Y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 I$

$\hat{\beta} = \underbrace{(X^TX)^{-1}X^T}_{}Y$
$= u^T Y$

If $\psi = \underset{p\times 1}{c^T} \underset{p\times 1}{\beta}$ is estimable, then $\hat{\psi} = c^T\hat{\beta}$ where $\hat{\beta}$ are the least squares estimates, has smallest variance in the class of all linear unbiased estimates and is unique.

Outline: Set up an arbitrary linear unbiased estimator a^Ty,

show $Var(a^Ty) = Var(c^T\hat{\beta}) + var(\cdot\ )$

Let a^Ty be an unbiased estimator of $\psi = c^T\beta$

$$E(a^Ty) = E(a^T(X\beta + \varepsilon))$$

$$= a^TX\beta = c^T\beta \quad \text{by unbiasedness}$$

$$\Rightarrow \quad a^TX = c^T$$

$$c^T\hat{\beta} = \underbrace{a^T X (X^TX)^{-1}} X^T y$$

$$= \lambda^T X^T y$$

$$\mathrm{var}(a^T y) = \mathrm{var}\big(\underbrace{a^T y - c^T\hat{\beta}} + \underbrace{c^T\hat{\beta}}\big)$$

$$= \mathrm{var}(a^T y - c^T\hat{\beta}) + \mathrm{var}(c^T\hat{\beta})$$

$$+ 2\,\mathrm{cov}(a^T y - c^T\hat{\beta},\ c^T\hat{\beta})$$

$$\mathrm{Cov}(a^T y - c^T\hat{\beta},\ c^T\hat{\beta}) = \mathrm{Cov}(a^T y - \lambda^T X^T y,\ \lambda^T X^T y)$$

$$= \mathrm{Cov}\big((a^T - \lambda^T X^T)y,\ \lambda^T X^T y\big)$$

$$= (a^T - \lambda^T X^T)\,\sigma^2 I\,(X\lambda)$$

$$= (a^T X - \lambda^T X^T X)\,\sigma^2 I \lambda$$

$$= (c^T - c^T)\,\sigma^2 I \lambda$$

$$= 0$$

Aside:

$E(aU+b)$

$\mathrm{Var}(aU+b)$

$\mathrm{Cov}(aU, bV)$

$\mathrm{Cov}(Ay, By)$

$A\,\underset{\mathrm{Var}(y)}{\mathrm{Cov}(y)}\,B^T$

$$\lambda^T X^T X = c^T$$

$$a^T X \cancel{(X^TX)^{-1} X^T} X = a^T X$$

$$= c^T$$

$$\underbrace{\mathrm{var}(a^T y)}_{\text{arbitrary linear unbiased estimator}} = \underbrace{\mathrm{var}(c^T \hat{\beta})}_{\text{least squares estimator}} + \underbrace{\mathrm{var}(a^T y - c^T \hat{\beta})}_{\geq 0}$$

$$\mathrm{var}(a^T y) \geq \mathrm{var}(c^T \hat{\beta})$$

Done!

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:

The trick: express e a vector of ~~cor~~ correlated variables as a function of uncorrelated variables, ε.