

Properties of the least squares estimates

2018-01-18

Warmup

Let a and b be scalar constants, and X be a scalar random variable.

Fill in the blanks

$$\begin{aligned} E(aX + b) &= \underline{\hspace{2cm}} \\ \text{Var}(aX + b) &= \underline{\hspace{2cm}} \end{aligned}$$

Solution:

$$\begin{aligned} E(aX + b) &= aE(X) + b \\ \text{Var}(aX + b) &= a^2\text{Var}(X) \end{aligned}$$

Goal

Recall that the least squares estimates are:

$$\hat{\beta}_{p \times 1} = (X^T X)^{-1} X^T y$$

Our goal today is to learn about the statistical properties of these estimates, in particular their expectation and variance.

Random Vectors

$\hat{\beta}$ is a vector-valued random variable, so we first need to cover a little more background.

Let U_1, \dots, U_n be scalar random variables. Then the vector

$$\mathbf{U} = (U_1, \dots, U_n)^T$$

is a vector valued random variable, a.k.a. a **random vector**.

The expectation of \mathbf{U} is the vector of expectations,

$$E(\mathbf{U}) = (E(U_1), \dots, E(U_n))^T$$

And the variance-covariance matrix of \mathbf{U} is

$$\text{Var}(\mathbf{U}) = \text{Cov}(\mathbf{U}) = \begin{pmatrix} \text{Var}(U_1) & \text{Cov}(U_1, U_2) & \cdots & \text{Cov}(U_1, U_n) \\ \text{Cov}(U_2, U_1) & \text{Var}(U_2) & \cdots & \text{Cov}(U_2, U_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(U_n, U_1) & \text{Cov}(U_n, U_2) & \cdots & \text{Var}(U_n) \end{pmatrix}_{n \times n}$$

For example, the errors in multiple linear regression, ϵ_i , $i = 1, \dots, n$ are independent with mean 0 and variance σ^2 .

Then,

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad E(\epsilon) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}, \quad \text{Var}(\epsilon) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

Properties of Expectation and Variance for random vectors

Let:

- $\mathbf{U}_{n \times 1}$ be a random vector
- $A_{m \times n}$ be a constant matrix
- $b_{m \times 1}$ be a constant vector

Then:

$$E(A\mathbf{U} + b) = AE(\mathbf{U}) + b$$

$$\text{Var}(A\mathbf{U} + b) = A\text{Var}(\mathbf{U})A^T$$

These are the vector analogs of the properties you wrote down in the warmup.

Find $E(y)$ and $\text{Var}(y)$ where $y_{n \times 1}$ satisfies the multiple linear regression equation:

$$y = X\beta + \epsilon$$

and $E(\epsilon) = \mathbf{0}$, $\text{Var}(\epsilon) = \sigma^2 I$

Solution:

$$E(y) = E(X\beta + \epsilon) = X\beta + E(\epsilon) = X\beta$$

$$\text{Var}(y) = \text{Var}(X\beta + \epsilon) = \text{Var}(\epsilon) = \sigma^2 I$$

Expectation of the least squares estimates

Assume the regression set up (with the usual dimensions):

$$y = X\beta + \epsilon$$

where X is fixed with rank p , $E(\epsilon) = 0$, and $\text{Var}(\epsilon) = \sigma^2 I_n$.

Fill in the blanks to show the least squares estimates are unbiased

$$\begin{aligned}
 E(\hat{\beta}) &= E\left((X^T X)^{-1} X^T y\right) \\
 &= E\left((X^T X)^{-1} X^T (\quad)\right) \quad \text{plug in the regression equation for } y \\
 &= E\left(\quad + \quad\right) \quad \text{expand} \\
 &= E\left(\quad + \quad\right) \quad \text{simplify term on the left } A^{-1}A = I \\
 &= \quad + \quad E(\quad) \quad \text{property of expectation} \\
 &= \quad \quad \quad \text{regression assumptions} \\
 &= \beta
 \end{aligned}$$

Solution:

$$\begin{aligned}
 E(\hat{\beta}) &= E\left((X^T X)^{-1} X^T y\right) \\
 &= E\left((X^T X)^{-1} X^T (X\beta + \epsilon)\right) \quad \text{plug in the regression equation for } y \\
 &= E\left((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \epsilon\right) \quad \text{expand} \\
 &= E\left(\beta + (X^T X)^{-1} X^T \epsilon\right) \quad \text{simplify term on the left } A^{-1}A = I \\
 &= \beta + (X^T X)^{-1} X^T E(\epsilon) \quad \text{property of expectation} \\
 &= \beta + (X^T X)^{-1} X^T \mathbf{0} \quad \text{regression assumptions} \\
 &= \beta
 \end{aligned}$$

Variance-covariance matrix of the least square estimates

Fill in the blanks to find the variance covariance matrix of least squares estimates

$$\begin{aligned}
 \text{Var}(\hat{\beta}) &= \text{Var}\left((X^T X)^{-1} X^T y\right) \\
 &= \text{Var}\left((X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon\right) \quad \text{plug in reg. eqn. and expand} \\
 &= 0 + \left(\quad \right) \text{Var}(\epsilon) \left(\quad \right)^T \quad \text{property of Var} \\
 &= \left(\quad \right) \left(\quad \right)^T \quad \text{regression assumption} \\
 &= \left(\quad \right) \left(\quad \right)^T \quad \text{move scalar to front} \\
 &= \sigma^2 \left(\quad \right) \quad \text{distribute transpose} \\
 &= \sigma^2 \quad \text{since } A^{-1} A = I \\
 &= \sigma^2 (X^T X)^{-1}
 \end{aligned}$$

Solution:

$$\begin{aligned}
 \text{Var}(\hat{\beta}) &= \text{Var}\left((X^T X)^{-1} X^T y\right) \\
 &= \text{Var}\left((X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon\right) \quad \text{plug in reg. eqn. and expand} \\
 &= 0 + (X^T X)^{-1} X^T \text{Var}(\epsilon) \left((X^T X)^{-1} X^T\right)^T \quad \text{property of Var} \\
 &= (X^T X)^{-1} X^T \sigma^2 I \left((X^T X)^{-1} X^T\right)^T \quad \text{regression assumption} \\
 &= \sigma^2 (X^T X)^{-1} X^T \left((X^T X)^{-1} X^T\right)^T \quad \text{move scalar to front} \\
 &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \quad \text{distribute transpose} \\
 &= \sigma^2 (X^T X)^{-1} \quad \text{since } A^{-1} A = I \\
 &= \sigma^2 (X^T X)^{-1}
 \end{aligned}$$

We can pull out the variance of a particular parameter estimate, say $\hat{\beta}_i$, from the diagonal of the matrix:

$$\text{Var}(\beta_i) = \sigma^2 (X^T X)^{-1}_{i+1i+1}$$

where A_{ij} indicates the element in the i 'th row and j 'th column of the matrix A .

Why $i + 1$?

The off diagonal terms tell us about the covariance between parameter estimates.

Estimating σ

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

An unbiased estimate is:

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

The denominator $n - p$ is known as the model **degrees of freedom**.

Standard errors of particular parameters

The standard error of a particular parameter is then the squareroot of the variance replacing σ^2 with its estimate:

$$\text{SE}(\beta_i) = \hat{\sigma}^2 \sqrt{(X^T X)^{-1}_{i+1i+1}}$$

Gauss Markov Theorem

You might wonder if we can find estimates with better properties. The Gauss-Markov theorem says the least squares estimates are BLUE (**B**est **L**inear **U**nbiased **E**stimator).

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

Of course if you are willing to have a non-linear estimate and/or, a biased estimate you might be able to find an estimate with smaller variance.

Proof, see Section 2.8 in Faraday

Summary

For the linear regression model

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

where $E(\epsilon) = 0_{n \times 1}$, $\text{Var}(\epsilon) = \sigma^2 I_n$, 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

$$E(\hat{\beta}) = \beta, \quad \text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$
$$E(\hat{\sigma}^2) = E\left(\frac{1}{n-p} \sum_{i=1}^n e_i^2\right) = \sigma^2$$

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