Notes on the least squares estimator

Daniel Sorensen[*]

July 11, 2023

The argument leading to the asymptotic variance of the least squares estimator (6.20) may seem obscure. This file elaborates on the topic.

Consider the simple linear regression model

$$y = 1\alpha + \widetilde{X}b + e \tag{1}$$

where y is a vector of observations with n elements, 1 is a vector of $1's$ with n elements, the scalar α is the intercept, $\widetilde{X}$ is the observed $n \times 1$ full rank matrix containing the values of the covariate across observations, b is the unknown regression parameter (a scalar) and e is the vector of n residuals, independent of $\widetilde{X}$, with mean 0 and variance $I\sigma^2$. Write (1) as

$$y = X\beta + e \tag{2}$$

where $X = \{X_i\}_{i=1}^n$ has the appended column of $1's$ in its first column and $\beta = (\alpha, b)$. The least squares estimator is

$$\hat{\beta} = (X'X)^{-1} X'y \tag{3}$$

and its variance is

$$Var\left(\hat{\beta}|, X, \sigma^2\right) = (X'X)^{-1}\sigma^2. \tag{4}$$

It is easy to check that $X'X$ has the following form,

$$\begin{aligned} X'X &= \begin{bmatrix} n & \sum_i X_i \\ \sum_i X_i & \sum_i X_i^2 \end{bmatrix} \\ &= n \begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \frac{1}{n}\sum_i X_i^2 \end{bmatrix} \end{aligned} \tag{5}$$

where $\overline{X}$ is $\sum_i X_i/n$, the average value of X. The determinant of (5) is

$$\begin{aligned} \det(X'X) &= n\left(\sum_i X_i^2 - n\overline{X}^2\right) \\ &= n\sum_i \left(X_i - \overline{X}\right)^2. \end{aligned}$$

[*]Center for Quantitative Genetics and Genomics, Aarhus University, C F Møllers Alle 3, bygning 1130, 8000 Aarhus, Denmark

The inverse matrix $(X'X)^{-1}$ is therefore

$$(X'X)^{-1} = \frac{1}{\sum_i \left(X_i - \overline{X}\right)^2} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}. \tag{6}$$

The least squares estimator (3) can be expressed as

$$\begin{bmatrix} \hat{\alpha} \\ \hat{b} \end{bmatrix} = \frac{1}{\sum_i \left(X_i - \overline{X}\right)^2} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix} \begin{bmatrix} \sum_i y_i \\ \sum_i X_i y_i \end{bmatrix},$$

and

$$\hat{b} = \frac{\sum_i X_i y_i - n\overline{X}\,\overline{y}}{\sum_i \left(X_i - \overline{X}\right)^2} = \frac{\sum_i \left(X_i - \overline{X}\right)(y_i - \overline{y})}{\sum_i \left(X_i - \overline{X}\right)^2}. \tag{7}$$

From (4) and (6), multiplying and dividing by n,

$$Var\left(\hat{b}|X, \sigma^2\right) = \frac{\sigma^2}{n}\left[\frac{1}{n}\sum_i \left(X_i - \overline{X}\right)^2\right]^{-1}, \tag{8}$$

where $\frac{1}{n}\sum_i \left(X_i - \overline{X}\right)^2$ is the sampling variance of X.

The unconditional (with respect to X) variance of the least squares estimator is

$$\begin{aligned} Var\left(\hat{b}|\sigma^2\right) &= E_X\left[Var\left(\hat{b}|X, \sigma^2\right)\right] + Var_X\left[E\left(\hat{b}|X, \sigma^2\right)\right] \\ &= E_X\left[Var\left(\hat{b}|X, \sigma^2\right)\right] \\ &= \frac{\sigma^2}{n} E\left\{\left[\frac{1}{n}\sum_i \left(X_i - \overline{X}\right)^2\right]^{-1}\right\}. \end{aligned} \tag{9}$$

As n increases, the sampling variance of X converges to the true variance of X, $Var(X)$, the inverse of the sampling variance converges to the inverse of the true variance, and

$$Var\left(\hat{b}|\sigma^2\right) \to \frac{\sigma^2}{n}[Var(X)]^{-1}. \tag{10}$$

Another line of argument followed in the book that leads to (10) is as follows. Instead of fitting model (1) consider fitting the model to the original data y excluding the intercept and using centred covariates $x_i = \left(X_i - \overline{X}\right)$. The equation for the mean of y given x is

$$E(y|x) = xb. \tag{11}$$

The least squares estimator is now

$$\begin{aligned} \hat{b} &= (x'x)^{-1}x'y \\ &= \left[\sum_i \left(X_i - \overline{X}\right)^2\right]^{-1} \sum_i \left(X_i - \overline{X}\right) y_i \\ &= \frac{\sum_i \left(X_i - \overline{X}\right)(y_i - \overline{y})}{\sum_i \left(X_i - \overline{X}\right)^2} \end{aligned}$$

as in (7) with sampling variance

$$\begin{aligned} Var\left(\hat{b}|x,\sigma^2\right) &= (x'x)^{-1}\sigma^2 \\ &= \frac{\sigma^2}{\sum_i \left(X_i - \overline{X}\right)^2} \end{aligned} \tag{12}$$

as in (8). Arguing as before, the same asymptotic unconditional variance (10) is obtained in the case of (12).

Availability of $\hat{b}$ leads to the estimator of α

$$\hat{\alpha} = \overline{y} - \hat{b}\overline{X}. \tag{13}$$

These results obtained using a single covariate regression model as an example, extend to a model based on an arbitrary number p of (possibly) correlated covariates, provided $p < n$ and matrix X is of full rank.

References