1 Least Square Estimators in Simple Linear Regression

Consider the least square estimators:

$$\hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{x}_n \qquad \hat{\beta} = \frac{\sum_1^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{SXX}$$

where $SXX = \sum_1^n (x_i - \bar{x}_n)^2$. Since $\sum_1^n (x_i - \bar{x}_n) = 0$,we can rewrite $\hat{\beta}$ as:

$$\hat{\beta} = \frac{\sum_1^n (x_i - \bar{x}_n)Y_i - \sum_1^n (x_i - \bar{x}_n)\bar{Y}_n}{SXX} = \frac{\sum_1^n (x_i - \bar{x}_n)Y_i}{SXX}$$

1.1 Expectation

$\hat{\beta}$ is an unbiased estimator:

$$\begin{aligned} E[\hat{\beta}] &= \frac{\sum_1^n (x_i - \bar{x}_n)E[Y_i]}{SXX} \\ &= \frac{\sum_1^n (x_i - \bar{x}_n)(\alpha + \beta x_i)}{SXX} \\ &= \frac{\beta \sum_1^n (x_i - \bar{x}_n)x_i}{SXX} = \beta \end{aligned}$$

where the last step follows since $\sum_1^n (x_i - \bar{x}_n)x_i = \sum_1^n (x_i - \bar{x}_n)x_i - \sum_1^n (x_i - \bar{x}_n)\bar{x} = SXX$. See the textbook [1, page 331] for a proof that $\hat{\alpha}$ is also unbiased, and [1, Exercise 22.12] for a different proof for $\hat{\beta}$.

1.2 Variance and Standard Errors of the Coefficients

We calculate:

$$Var(\hat{\beta}) = \frac{\sum_1^n (x_i - \bar{x}_n)^2 Var(Y_i)}{SXX^2} = \sigma^2 \frac{\sum_1^n (x_i - \bar{x}_n)^2}{SXX^2} = \frac{\sigma^2}{SXX} \tag{1}$$

and:

$$\begin{aligned} Var(\hat{\alpha}) &= Var(\bar{Y}_n - \hat{\beta}\bar{x}_n) \\ &= Var(\bar{Y}_n) + \bar{x}_n^2 Var(\hat{\beta}) - 2\bar{x}_n Cov(\bar{Y}_n, \hat{\beta}) \\ &= \frac{\sigma^2}{n} + \bar{x}_n^2 \frac{\sigma^2}{SXX} - 0 = \sigma^2 (\frac{1}{n} + \frac{\bar{x}_n^2}{SXX}) \end{aligned} \tag{2}$$

The covariance in the formula is zero because (recall that $Y_1, \dots, Y_n$ are independent):

$$\begin{aligned} Cov(\bar{Y}_n, \hat{\beta}) &= Cov(\frac{1}{n}\sum_1^n Y_i, \frac{\sum_1^n (x_i - \bar{x}_n)Y_i}{SXX}) \\ &= \frac{1}{nSXX} Cov(\sum_1^n Y_i, \sum_1^n (x_i - \bar{x}_n)Y_i) \\ &= \frac{1}{nSXX} \sum_1^n Cov(Y_i, (x_i - \bar{x}_n)Y_i) \\ &= \frac{1}{nSXX} \sum_1^n (x_i - \bar{x}_n) Var(Y_i) = \frac{\sigma^2}{n} \frac{\sum_1^n (x_i - \bar{x}_n)}{SXX} = 0 \end{aligned}$$

The *standard errors* of the coefficient estimators are defined as the estimates of the standard deviations (see (1) and (2)):

$$se(\hat{\alpha}) = \hat{\sigma}\sqrt{(\frac{1}{n} + \frac{\bar{x}_n^2}{SXX})} \qquad se(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{SXX}}$$

where:

$$\hat{\sigma} = \sqrt{\frac{1}{n-2}\sum_1^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}$$

is the estimate of σ (see [1, page 332]).

1.3 Variance and Standard Errors of Fitted Values

For a given value of the explanatory variable, say x_0, the estimator $Y = \hat{\alpha} + \hat{\beta}x_0$ has expectation $E[Y] = \alpha + \beta x_0$, which is then the best estimate for the fitted value. We can compute the variance of Y as:

$$\begin{aligned} Var(Y) &= Var(\hat{\alpha} + \hat{\beta}x_0) \\ &= Var(\hat{\alpha}) + x_0^2 Var(\hat{\beta}) + 2x_0 Cov(\hat{\alpha}, \hat{\beta}) \\ &= Var(\hat{\alpha}) + (x_0^2 - 2x_0\bar{x}_n)Var(\hat{\beta}) \\ &= \sigma^2(\frac{1}{n} + \frac{\bar{x}_n^2}{SXX}) + \frac{(x_0^2 - 2x_0\bar{x}_n)\sigma^2}{SXX} \\ &= \sigma^2(\frac{1}{n} + \frac{(\bar{x}_n - x_0)^2}{SXX}) \end{aligned}$$

because:

$$\begin{aligned} Cov(\hat{\alpha}, \hat{\beta}) &= Cov(\bar{Y}_n - \hat{\beta}\bar{x}_n, \hat{\beta}) \\ &= Cov(\bar{Y}_n, \hat{\beta}) - \bar{x}_n Cov(\hat{\beta}, \hat{\beta}) \\ &= -\bar{x}_n Var(\hat{\beta}) \end{aligned}$$

The *standard error* of the fitted value is then the estimate:

$$se(\hat{y}) = \hat{\sigma}^2(\frac{1}{n} + \frac{(\bar{x}_n - x_0)^2}{SXX})$$

References

[1] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester. *A Modern Introduction to Probability and Statistics*. Springer, 2005.