

# 1 Least Square Estimators in Simple Linear Regression

Consider the least square estimators:

$$\hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{x}_n \quad \hat{\beta} = \frac{\sum_1^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{SXX} \quad (1)$$

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:

$$\hat{\beta} = \frac{\sum_1^n (x_i - \bar{x}_n)Y_i - \sum_1^n (x_i - \bar{x}_n)\bar{Y}_n}{SXX} = \frac{\sum_1^n (x_i - \bar{x}_n)Y_i}{SXX} \quad (2)$$

## 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:

$$Var(\hat{\beta}) = \frac{\sum_1^n (x_i - \bar{x}_n)^2 Var(Y_i)}{SXX^2} = \sigma^2 \frac{\sum_1^n (x_i - \bar{x}_n)^2}{SXX^2} = \frac{\sigma^2}{SXX} \quad (3)$$

and:

$$\begin{aligned} Var(\hat{\alpha}) &= Var(\bar{Y}_n - \hat{\beta}\bar{x}_n) \\ &= Var(\bar{Y}_n) + \bar{x}_n^2 Var(\hat{\beta}) - 2\bar{x}_n Cov(\bar{Y}_n, \hat{\beta}) \\ &= \frac{\sigma^2}{n} + \bar{x}_n^2 \frac{\sigma^2}{SXX} - 0 = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}_n^2}{SXX} \right) \end{aligned} \quad (4)$$

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

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

The *standard errors* of the coefficient estimators are defined as the estimates of the standard deviations (see (3) and (4)):

$$se(\hat{\alpha}) = \hat{\sigma} \sqrt{\left(\frac{1}{n} + \frac{\bar{x}_n^2}{SXX}\right)} \quad se(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{SXX}} \quad (5)$$

where:

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

is the estimate of  $\sigma^2$  (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:

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

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:

$$se(\hat{y}) = \hat{\sigma} \sqrt{\left(\frac{1}{n} + \frac{(\bar{x}_n - x_0)^2}{SXX}\right)} \quad (6)$$

## 2 Confidence Intervals

In this section, we make the *normality assumption* that  $U_i \sim \mathcal{N}(0, \sigma^2)$  in the simple linear regression model [1, page 257]:

$$Y_i = \alpha + \beta x_i + U_i$$

A fortiori, we have  $Y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2)$ .

### 2.1 Confidence Intervals of the Coefficients

By (2), the estimator  $\hat{\beta}$  is a linear combination of the  $Y_i$ 's, hence it has normal distribution as well. By Sections 1.1 and 1.2, it must be that:

$$\hat{\beta} \sim \mathcal{N}(\beta, \text{Var}(\hat{\beta}))$$

where the variance  $\text{Var}(\hat{\beta})$  is unknown. The studentized statistics:

$$\frac{\hat{\beta} - \beta}{\sqrt{\text{Var}(\hat{\beta})}} \sim t(n-2) \quad (7)$$

has a t-student distribution with  $n-2$  degrees of freedom ( $n-2$  because 2 parameters are already estimated). The proof of this fact can be found in [2, pages 45]. Hence:

$$P\left(-t_{n-2,0.025} \leq \frac{\hat{\beta} - \beta}{\sqrt{\text{Var}(\hat{\beta})}} \leq t_{n-2,0.025}\right) = 0.95$$

where  $t_{n-2,0.025}$  is the critical value of  $t(n-2)$  at 0.025. Hence, a 95% confidence interval is:

$$\hat{\beta} \pm t_{n-2,0.025} se(\hat{\beta})$$

where  $se(\hat{\beta})$  is the standard error from (5). By following the same reasoning, we obtain the confidence intervals for  $\alpha$ :

$$\hat{\alpha} \pm t_{n-2,0.025} se(\hat{\alpha})$$

### 2.2 Confidence Intervals of the Fitted Values

Analogously to the previous subsection, for a fitted value  $\hat{y} = \hat{\alpha} + \beta \hat{x}_0$ :

$$\hat{y} \pm t_{n-2,0.025} se(\hat{y})$$

where  $se(\hat{y})$  is from (6). In particular, this interval concerns *the mean of fitted values* at  $x_0$ . For example, we could conclude that the predicted value at  $x_0$  is on average between  $\hat{y} + t_{n-2,0.025} se(\hat{y})$  and  $\hat{y} - t_{n-2,0.025} se(\hat{y})$ . For a given single prediction, we must also account for the variance of the error term  $U$  in:

$$Y = \hat{\alpha} + \beta \hat{x}_0 + U$$

Let us assume that  $U \sim \mathcal{N}(0, \sigma^2)$ . By reasoning as in Section 1.3, it can be shown that  $\text{Var}(Y) = \sigma^2(1 + \frac{1}{n} + \frac{(\bar{x}_n - x_0)^2}{SXX})$ , and then by defining:

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

we have that the prediction interval is:

$$\hat{y} \pm t_{n-2,0.025} se'(\hat{y})$$

In this case, we could conclude that the specific predicted value at  $x_0$  is on between  $\hat{y} + t_{n-2,0.025} se'(\hat{y})$  and  $\hat{y} - t_{n-2,0.025} se'(\hat{y})$ .

### 2.3 Hypothesis Testing

Consider now the following test of hypothesis:

$$H_0 = \beta = 0 \quad H_1 = \beta \neq 0$$

The p-value of observing  $|\hat{\beta}|$  or a greater value under the null hypothesis, can be calculated from (7) as:

$$p = P(|T| > |t|) = 2 \cdot P\left(T > \left| \frac{\hat{\beta} - 0}{se(\hat{\beta})} \right|\right)$$

for  $T \sim t(n-2)$ . Hence,  $H_0$  can be rejected in favor of  $H_1$  at significance level of 0.05, i.e.  $p < 0.05$ , if  $|t| > t_{n-2,0.025}$ . A similar approach applies to the intercept.

### References

- [1] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester. *A Modern Introduction to Probability and Statistics*. Springer, 2005.
- [2] M. H. Kutner, C. J. Nachtsheim, J. Neter, and Li W. *Applied Linear Statistical Models*. 5th edition, 2005.