1 Least Square Estimators in Simple Linear Regression

Consider the least square estimators:

\[
\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} \quad \hat{\beta} = \frac{\sum^n_1(x_i - \bar{x})(Y_i - \bar{Y})}{SXX}
\]  

(1)

where \( SXX = \sum^n_1(x_i - \bar{x})^2 \). Since \( \sum^n_1(x_i - \bar{x}) = 0 \), we can rewrite \( \hat{\beta} \) as:

\[
\hat{\beta} = \frac{\sum^n_1(x_i - \bar{x})Y_i - \sum^n_1(x_i - \bar{x})\bar{Y}_n}{SXX} = \frac{\sum^n_1(x_i - \bar{x})Y_i}{SXX}
\]

(2)

1.1 Expectation

\( \hat{\beta} \) is an unbiased estimator:

\[
E[\hat{\beta}] = \frac{\sum^n_1(x_i - \bar{x})E[Y_i]}{SXX} = \frac{\sum^n_1(x_i - \bar{x})(\alpha + \beta x_i)}{SXX} = \beta \frac{\sum^n_1(x_i - \bar{x})x_i}{SXX} = \beta
\]

where the last step follows since \( \sum^n_1(x_i - \bar{x})x_i = \sum^n_1(x_i - \bar{x})x_i - \sum^n_1(x_i - \bar{x})\bar{x} = SXX \). 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:

\[
Var(\hat{\beta}) = \frac{\sum^n_1(x_i - \bar{x})^2 Var(Y_i)}{SXX^2} = \sigma^2 \frac{\sum^n_1(x_i - \bar{x})^2}{SXX} = \frac{\sigma^2}{SXX}
\]

(3)

and:

\[
Var(\hat{\alpha}) = Var(\bar{Y}_n - \hat{\beta}\bar{x}_n) = Var(\bar{Y}_n) + \bar{x}_n^2 Var(\hat{\beta}) - 2\bar{x}_n Cov(\bar{Y}_n, \hat{\beta}) = \frac{\sigma^2}{n} + \bar{x}_n^2 \frac{\sigma^2}{SXX} - 0 = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_n^2}{SXX}\right)
\]

(4)

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

\[
Cov(\bar{Y}_n, \hat{\beta}) = Cov\left(\frac{1}{n} \sum^n_1 Y_i, \frac{\sum^n_1(x_i - \bar{x}_n)Y_i}{SXX}\right)
\]

\[
= \frac{1}{nSXX} Cov\left(\sum^n_1 Y_i, \sum^n_1(x_i - \bar{x}_n)Y_i\right)
\]

\[
= \frac{1}{nSXX} \sum^n_1 Cov(Y_i, (x_i - \bar{x}_n)Y_i)
\]

\[
= \frac{1}{nSXX} \sum^n_1 (x_i - \bar{x}_n)Var(Y_i) = \frac{\sigma^2}{n} \frac{\sum^n_1(x_i - \bar{x}_n)}{SXX} = 0
\]
The standard errors of the coefficient estimators are defined as the estimates of the standard deviations (see (3) and (4)):

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

(5)

where:

\[ \hat{\sigma} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}x_i)^2} \]

is the estimate of \( \sigma \) (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:

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

because:

\[
Cov(\hat{\alpha}, \hat{\beta}) = Cov(\bar{Y}_n - \hat{\beta}\bar{x}_n, \hat{\beta}) \\
= Cov(\bar{Y}_n, \hat{\beta}) - \bar{x}_n Cov(\hat{\beta}, \hat{\beta}) \\
= -\bar{x}_n Var(\hat{\beta})
\]

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

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

(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 \[ \text{page 257} \]:

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

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

2.1 Confidence Intervals of the Coefficients

By \[ \text{2} \], the estimator $\hat{\beta}$ is a linear combination of 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:

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

has a t-student distribution with $n - 2$ degrees of freedom ($n - 2$ because 2 parameters are already estimated). The proof is this fact can be found in \[ \text{2} \] 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:

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

where $\text{se}(\hat{\beta})$ is the standard error from \[ \text{6} \]. By following the same reasoning, we obtain the confidence intervals for $\alpha$:

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

2.2 Confidence Intervals of the Fitted Values

Analogously to the previous subsection, for a fitted value $\hat{y} = \hat{\alpha} + \hat{\beta}x_0$:

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

where $\text{se}(\hat{y})$ is from \[ \text{6} \]. 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 $\hat{y} + t_{n - 2, 0.025} \text{se}(\hat{y})$ and $\hat{y} - t_{n - 2, 0.025} \text{se}(\hat{y})$. For a given single prediction, we must also account for the variance of the error term $U$ in:

$$Y = \hat{\alpha} + \hat{\beta}x_0 + U$$

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

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

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

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
