CHAPTER 7

Inference in Ordinary Least Squares

7.0 What We Need to Know When We Finish This Chapter

This chapter addresses the question of how accurately we can estimate the values of β and α from b and a. Regression produces an estimate of the standard deviation of ε_i. This, in turn, serves as the basis for estimates of the standard deviations of b and a. With these, we can construct confidence intervals for β and α and test hypotheses about their values. Here are the essentials.

1. **Section 7.1:** We assume that the true distributions of b and a are the normal. Because we have to estimate their variances in order to standardize them, however, we have to treat their standardized versions as having t-distributions if our samples are small.
2. **Section 7.2:** *Degrees of freedom* count the number of independent observations that remain in the sample after accounting for the sample statistics that we've already calculated.
3. **Equation (7.3), section 7.2:** The ordinary least squares (OLS) estimator for σ^2, the variance of ε_i, is

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

4. **Equation (7.9), section 7.3:** The $(1-\alpha)$% confidence interval for β is

$$1-\alpha = \mathrm{P}\left(b - t_{\alpha/2}^{(n-2)} \sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} < \beta < b + t_{\alpha/2}^{(n-2)} \sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \right).$$

5. **Section 7.3:** Larger samples and greater variation in x_i yield narrower confidence intervals. So does smaller σ^2, but we can't control that.
6. **Equation (7.16), section 7.4:** The two-tailed hypothesis test for $H_0 : \beta = \beta_0$ is

$$1-\alpha = \mathrm{P}\left(\beta_0 - t_{\alpha/2}^{(n-2)} \left(\sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \right) < b < \beta_0 + t_{\alpha/2}^{(n-2)} \left(\sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \right) \right).$$

   The alternative hypothesis is $H_1 : \beta \neq \beta_0$.
7. **Section 7.4:** The test of the null hypothesis $H_0 : \beta = 0$ is always interesting because, if true, it means that x_i doesn't affect y_i.

8. **Equation (7.26), section 7.4:** The upper-tailed hypothesis test for $H_0 : \beta = \beta_0$ is

$$1-\alpha = \mathrm{P}\left(b < \beta_0 + t_{\alpha}^{(n-2)} \sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \right).$$

9. **Section 7.5:** The best linear unbiased estimator of E(y_0) is $\hat{y}_0 = a + bx_0$.
10. **Section 7.5:** Predictions are usually more reliable if they are based on larger samples and made for values of the explanatory variable that are similar to those that appear in the sample.