T-test for β_i

Recall that for normally distributed residuals,

1. $V\hat{\beta}_i = c_{ii}\sigma^2$
2. $\hat{\beta}_i$ is normal
3. $S^2 = SSE/(n-2)$ estimates σ^2
4. $(n-2)S^2/\sigma^2 = SSE/\sigma^2 \sim \chi^2(n-2)$.

We can use this information to construct hypothesis tests for the coefficients β_i.

First, we get the usual large sample Wald test using a test statistic

$$Z = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}} \underset{H_0}{\sim} \mathcal{N}(0,1)$$

with a two-tailed rejection region $\{|z| > z_{\alpha/2}\}$ where $z_\alpha = F^{-1}_{\mathcal{N}(0,1)}(1-\alpha)$.

But we can also notice that Z is normal, $W=(n-2)S^2/\sigma^2$ is $\chi^2(n-2)$ and independent, so

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}} \Bigg/ \sqrt{\frac{(n-2)S^2}{(n-2)\sigma^2}} =$$

$$= \frac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}} \cdot \frac{\sigma}{S} = \frac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$$ is $t(n-2)$ distributed.

This gives us T-tests and T-based CIs for β_i.

Inferences on $a_0\beta_0 + a_1\beta_1$

We may have reason to make inferences on linear functions of β_0 and β_1 — for instance if we want to make inferences about $\beta_0 + \beta_1 x^*$ for some possibly new x^*.

We can create an estimator $\hat{\vartheta} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1$. By linearity of $\mathbb{E}$, this is unbiased.
Using our calculations for $\mathbb{V}\hat{\beta}_0, \mathbb{V}\hat{\beta}_1, cov(\hat{\beta}_0, \hat{\beta}_1)$, we can calculate $\mathbb{V}\hat{\vartheta}$:

$$\mathbb{V}\hat{\vartheta} = a_0^2\,\mathbb{V}\hat{\beta}_0 + a_1^2\,\mathbb{V}\beta_1 + 2a_0a_1\,cov(\hat{\beta}_0, \hat{\beta}_1) =$$

$$= \sigma^2\left(a_0^2 c_{00} + a_1^2 c_{11} + 2a_0a_1 c_{01}\right)$$

where

$$c_{00} = \sum x_i^2 / nS_{xx}, \quad c_{11} = 1/S_{xx} \quad \text{and} \quad c_{01} = -\bar{x}/S_{xx}.$$

Knowing the variance of an unbiased estimator, we can construct both Wald tests and t-tests.

The case $\beta_0 + \beta_1 x^*$

We can use these tests for estimating the expected value over a new value x^* for the predictors.

In this case, $a_0 = 1$, $a_1 = x^*$

$$\sigma^2 \left(a_0^2 c_{00} + a_1^2 c_{11} + 2 a_0 a_1 c_{01} \right) =$$

$$= \frac{\sigma^2}{S_{xx}} \left(a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2 a_0 a_1 \bar{x} \right) = \quad \text{insert } a_0 = 1,\ a_1 = x^*.$$

$$= \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2}{n} + x^{*2} - 2 x^* \bar{x} \right) = \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2}{n} - \frac{n\bar{x}^2}{n} + \bar{x}^2 + x^{*2} - 2 x^* \bar{x} \right) =$$

$$= \frac{\sigma^2}{S_{xx}} \left(S_{xx}/n + (x^* - \bar{x})^2 \right) = \sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}} \right).$$

A confidence interval of

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2}\, s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$ follows.

The width here shrinks when:

* $(x^* - \bar{x})^2$ shrinks. Estimation is easier near $\bar{x}$.
* S_{xx} grows. Estimation is easier when the x_i are spread out more.

Prediction Intervals

Often, predicting a range for the next observation is more valuable than estimating a range for the long-term average value.

For a new value $Y^* = \beta_0 + \beta_1 x^* + \varepsilon$, we would want an estimator, for instance

$$\hat{Y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

This quantity is simultaneously a predictor of Y^* and an estimator of $\mathbb{E}Y$.

We want to minimize our prediction error $Y^* - \hat{Y}^*$.

$$\mathbb{E}\,\text{error} = \mathbb{E}\left[Y^* - \hat{Y}^*\right] = \mathbb{E}Y^* - \mathbb{E}\hat{Y}^* = (\beta_0 + \beta_1 x^*) - (\beta_0 + \beta_1 x^*) = 0.$$

$$\mathbb{V}\,\text{error} = \mathbb{V}Y^* + \mathbb{V}\hat{Y}^* - 2\,\text{cov}(Y^*, \hat{Y}^*).$$

Since we are predicting, the two are independent, so

$$\mathbb{V}\,\text{error} = \mathbb{V}Y^* + \mathbb{V}\hat{Y}^* = \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right) = \sigma^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right).$$

If $\varepsilon \sim N(0, \sigma^2)$, then the error is also normal. Hence we get a Wald-like statistic

$$Z = \frac{Y^* - \hat{Y}^*}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}}$$ and its t-statistic $$T = \frac{Y^* - \hat{Y}^*}{S\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}} \sim t(n-2).$$

With T as a pivot, we get our prediction intervals

$$Y^* \in \hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}}$$

It is worth noticing that the prediction interval width has a lower bound of $2t_{\alpha/2}S$. Prediction <u>cannot</u> get better than this no matter how much data is collected.