

EQUIVALENCE WITH THE TEST ABOUT THE SIGNIFICANCE OF β_2

We are considering the simple linear model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ and the system of hypotheses

$$\begin{cases} H_0: \text{the model does not help to explain the variability of } Y \\ H_1: \text{the model helps to explain the variability of } Y \end{cases}$$

which can be expressed in terms of the coefficient R^2 as

$$\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{cases}$$

We have seen that we can use the test statistic $(n-2) R^2 / (1-R^2)$, which, under H_0 , has an $F_{2, n-2}$ distribution.

$$\begin{aligned} F &= \frac{R^2}{1-R^2} \cdot (n-2) = \frac{SSR}{SSE} (n-2) = \\ &= \left(\frac{SST}{SSE} - 1 \right) (n-2) = \\ &= \left(\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} - 1 \right) (n-2) = \\ &= \frac{\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}}{\sum_{i=1}^n Y_i^2} (n-2) \stackrel{H_0}{\sim} F_{2, n-2} \end{aligned}$$

In the case of the simple linear model, this test is equivalent to a test about the significance of β_2 , i.e., $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$.

Preliminary result:

If $T \sim t_n$, and $V = T^2$ then $V \sim F_{1, n}$

PROOF:

$$\text{Let's start from } \frac{SST}{SSE} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} = \frac{\sum_{i=1}^n \tilde{\varepsilon}_i^2}{\sum_{i=1}^n \varepsilon_i^2} \quad \text{with } \tilde{\varepsilon}_i = Y_i - \bar{Y}, \quad \varepsilon_i = Y_i - \hat{Y}_i$$

Now, notice that we can write

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \sum_{i=1}^n (Y_i - \bar{Y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)^2 = \\ &= \sum_{i=1}^n [(Y_i - \bar{Y}) - \hat{\beta}_1 (x_i - \bar{x})]^2 = \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 + \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2 \hat{\beta}_1 \sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x}) \\ &= \sum_{i=1}^n \tilde{\varepsilon}_i^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

$$\Rightarrow \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \tilde{\varepsilon}_i^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\Rightarrow \sum_{i=1}^n \tilde{\varepsilon}_i^2 = \sum_{i=1}^n \varepsilon_i^2 + \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

Moreover recall that

$$V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad ; \quad \hat{V}(\hat{\beta}_1) = \frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad ; \quad \frac{(n-2) S^2}{\sigma^2} \sim \chi_{n-2}^2$$

Going now back to the test statistic

$$\begin{aligned} \frac{R^2}{1-R^2} (n-2) &= \left(\frac{\sum_{i=1}^n \tilde{\varepsilon}_i^2}{\sum_{i=1}^n \varepsilon_i^2} - 1 \right) (n-2) = \left(\frac{\sum_{i=1}^n \tilde{\varepsilon}_i^2 + \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n \varepsilon_i^2} - 1 \right) (n-2) = \\ &= \left(\cancel{1} + \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n \varepsilon_i^2} - \cancel{1} \right) (n-2) = \\ &= \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(n-2) S^2} (n-2) \quad \begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = n \hat{\sigma}^2 = (n-2) S^2 \end{aligned} \\ &= \frac{\hat{\beta}_1^2 \cdot \frac{1}{\sigma^2}}{\left(\frac{(n-2) S^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \cdot \frac{1}{\sigma^2}} (n-2) = \\ &= \frac{\hat{\beta}_1^2}{\left(\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} \cdot \frac{1}{\frac{(n-2) S^2}{\sigma^2}} = \frac{\hat{\beta}_1^2}{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \cdot \frac{1}{\frac{(n-2) S^2}{\sigma^2}} = \\ &= \frac{\left(\frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \right)^2}{\frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)}} \stackrel{H_0}{\sim} \frac{\chi_{n-2}^2}{n-2} = T^2 \end{aligned}$$

$$\text{Hence, } \frac{R^2}{1-R^2} (n-2) = T^2$$

$$\text{where } T = \frac{\frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}}}{\sqrt{\frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)}}} = \frac{\frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}}{\sqrt{\frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)}}} = \frac{\hat{\beta}_1 \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{S^2}} = \frac{\hat{\beta}_1}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \stackrel{H_0}{\sim} t_{n-2}$$

$$\Rightarrow F = \frac{R^2}{1-R^2} \cdot (n-2) = T^2 \stackrel{H_0}{\sim} F_{2, n-2}$$

So, we have derived the distribution of the test statistic (in the case of simple lin).

Remark: connection with the p-value of the test $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$

$$\begin{aligned} P_{H_0}(F \geq f^{obs}) &= P_{H_0}(T^2 \geq (t^{obs})^2) \\ &= P_{H_0}(|T| \geq |t^{obs}|) = \\ &= 2 P_{H_0}(T \geq |t^{obs}|) \quad T \sim t_{n-2} \end{aligned}$$



where T is exactly the test statistic we derived to test β_2

