EQUIVALENCE WITH THE TEST ABOUT THE SIGNIFICANCE OF β_2

We are considering the simple linear model $\quad Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \,, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

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 $F_{1,n-2}$ distribution.

$$F = \frac{R^2}{1-R^2} \cdot (n-2) = \frac{SSR}{SSE}\,(n-2) =$$

$$= \left(\frac{SST}{SSE} - 1\right)(n-2) =$$

$$= \left(\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} - 1 \right)(n-2) =$$

$$= \frac{\hat{\Sigma}^2 - \hat{\hat{\Sigma}}^2}{\hat{\Sigma}^2}\,(n-2) \overset{H_0}{\sim} F_{1,n-2}$$

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:

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

Now, notice that we can write

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 = \sum_{i=1}^{n} (Y_i - \bar{Y} + \hat{\beta}_2 \bar{x} - \hat{\beta}_2 x_i)^2 =$$

$$= \sum_{i=1}^{n} \left[(Y_i - \bar{Y}) - \hat{\beta}_2 (x_i - \bar{x}) \right]^2 =$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\hat{\beta}_2 \sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})$$

$$\hat{\beta}_2 \underbrace{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}_{= \hat{\beta}_2} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$-2\hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} \underbrace{(Y_i - \bar{Y})^2}_{\varepsilon_i^{*2}} - \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\Rightarrow \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \varepsilon_i^{*2} - \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\Rightarrow \sum_{i=1}^{n} \varepsilon_i^{*2} = \sum_{i=1}^{n} \varepsilon_i^2 + \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Moreover, recall that

$$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad ; \quad \hat{V}(\hat{\beta}_2) = \frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad ; \quad \frac{(n-2) S^2}{\sigma^2} \sim \chi^2_{n-2}$$

Going now back to the test statistic

$$\frac{R^2}{1-R^2}(n-2) = \left(\frac{\sum_{i=1}^{n} \varepsilon_i^{*2}}{\sum_{i=1}^{n} \varepsilon_i^2} - 1 \right)(n-2) = \left(\frac{\sum_{i=1}^{n} \varepsilon_i^2 + \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} \varepsilon_i^2} - 1 \right)(n-2) =$$

$$= \left(\cancel{1} + \frac{\hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} \varepsilon_i^2} - \cancel{1} \right)(n-2) =$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = n \hat{\Sigma}^2 = (n-2) S^2$$

$$= \frac{\hat{\beta}_2^2 \sum_{i=1}^{n} (x - \bar{x})^2}{(n-2) S^2}\,(n-2)$$

$$= \frac{\hat{\beta}_2^2 \cdot \frac{1}{\sigma^2}}{\left((n-2) \frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) \cdot \frac{1}{\sigma^2}}\,(n-2) =$$

$$\frac{\hat{\beta}_2}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}} \overset{H_0}{\sim} N(0,1)$$

$$= \frac{\frac{\hat{\beta}_2^2}{\left(\frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)}}{\frac{(n-2) S^2}{\sigma^2}} \cdot \frac{1}{\frac{1}{(n-2)}} =$$

$$\frac{(n-2) S^2}{\sigma^2} \overset{H_0}{\sim} \chi^2_{n-2}$$

$$= \frac{\left(\frac{\hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}} \right)^2}{\frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)}} = T^2$$

$$\left(\frac{\hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}} \right)^2 \overset{H_0}{\sim} N(0,1)^2 \qquad \frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)} \overset{H_0}{\sim} \frac{\chi^2_{n-2}}{n-2}$$

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

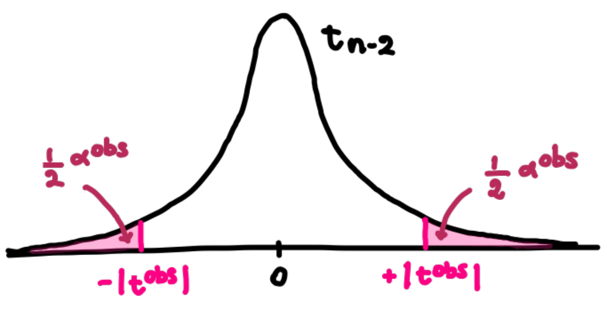
$$\text{where} \quad T = \frac{\frac{\hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}}}{\sqrt{\frac{(n-2) S^2}{\sigma^2} \cdot \frac{1}{(n-2)}}} = \frac{\frac{\hat{\beta}_2}{\sqrt{\frac{\cancel{\sigma^2}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}}}{\sqrt{\frac{\cancel{(n-2)} S^2}{\cancel{\sigma^2}} \cdot \frac{1}{\cancel{(n-2)}}}} = \frac{\hat{\beta}_2 \sqrt{\sum_{i=1}^{n} (x - \bar{x})^2}}{\sqrt{S^2}} = \frac{\hat{\beta}_2}{\sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_2}{\sqrt{\hat{\text{var}}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{n-2}$$

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

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

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

$$\mathbb{P}_{H_0}(F \geq f^{obs}) = \mathbb{P}_{H_0}(T^2 \geq (t^{obs})^2)$$
$$= \mathbb{P}_{H_0}(|T| \geq |t^{obs}|) =$$
$$= 2\, \mathbb{P}_{H_0}(T \geq |t^{obs}|) \qquad T \overset{H_0}{\sim} t_{n-2}$$



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