INFERENCEx FOR SIMPLE OLS

Model Assumptions:
("The" Simple Linear Regression Model Version 3)

(Consider x₁, … , xₙ as fixed.)

1. E(Ylx) = η₀ + η₁x (linear mean function)

2. Var(Ylx) = σ² (constant variance)
   (Equivalently, Var(elx) = σ²)

3. y₁, … , yₙ are independent observations. (independence)

4. (NEW) Ylx is normal for each x (normality)

Summarizing (1) + (2) + (4):

Ylx ∼ N(η₀ + η₁x, σ²)

Comments: 1. For some purposes, we need only assume (4) for x = xᵢ’s.
   2. We can sometimes weaken (4) to “n large” and get approximate results. (But how large is large??)

Unless stated otherwise, we will henceforth assume that “The Simple Linear Regression Model” refers to Version 3.

Recall: elx = Ylx - E(Ylx)

So: elx ∼ N(0, σ²)

:. all errors have the same distribution -- so we just say e for elx.

Recall: ˆh₀ and ˆh are linear combinations of the Ylx’s

:. (3) + (4) ⇒ (the sampling distributions of)
   ˆh₀ and ˆh are normally distributed.

Recall:

E( ˆh₀) = η₀    Var( ˆh₀) = σ²

E( ˆh) = η₁    Var( ˆh) = σ²

:. ˆh₀ ~ ˆh₀ ~
Standardize $\hat{\eta}$:

$$\frac{\hat{\eta} - \eta}{\sqrt{\hat{\sigma}^2 / SXX}} \sim N(0,1)$$

$\sigma^2$ is unknown -- so approximate it by $\hat{\sigma}^2$ --

i.e., approximate $\text{Var}(\hat{\eta})$ by

$$\text{Var}(\hat{\eta}) = [\text{s.e.} (\hat{\eta})]^2 = \frac{\hat{\sigma}^2}{SXX}.$$

Problem: $\frac{\hat{\eta} - \eta}{\sqrt{\hat{\sigma}^2 / SXX}}$ isn't normal.

Solution: Rewrite $\frac{\hat{\eta} - \eta}{\sqrt{\hat{\sigma}^2 / SXX}}$ as

$$\left( \frac{\hat{\eta} - \eta}{\sqrt{\hat{\sigma}^2 / SXX}} \right) \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}$$

(*) has (standard) normal numerator.

Facts: (Proofs omitted)

a. $(n-2) \frac{\hat{\sigma}^2}{\sigma^2}$ has a $\chi^2$ distribution with $n-2$ degrees of freedom.

Notation: $(n-2) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

b. $(n-2) \frac{\hat{\sigma}^2}{\sigma^2}$ is independent of $\hat{\eta} - \eta$ (hence independent of the numerator in (*) )

Comments on distributions:

1. Definition of $\chi^2(k)$ distribution:

The distribution of a random variable which is a sum of squares of $k$ independent standard normal random variables.

[Comment: Recall that $\hat{\sigma}^2 = \frac{1}{n-2} \text{RSS}$, so

$$(n-2) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\text{RSS}}{\sigma^2} = \sum \left( \frac{e_i}{\sigma} \right)^2$$

is a sum of $n$ squares; the fact quoted above says that it can also be expressed as a sum of $n-2$ squares of independent standard normal random variables.]
2. Definition of t-distribution with k degrees of freedom:

The distribution of a random variable of the form

\[ Z \frac{U}{\sqrt{U/k}} \]

where

- \( Z \sim N(0,1) \)
- \( U \sim \chi^2(k) \)
- \( Z \) and \( U \) are independent.

Notation: \( t(k) \)

In (*) take

\[ U = (n-2) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2) \]

\[ Z = \frac{\hat{\eta}_1 - \eta_h}{\sqrt{\hat{\sigma}^2 / SXX}} \sim N(0,1) \]

Thus:

\[ \frac{\hat{\eta}_1 - \eta_h}{\sqrt{\hat{\sigma}^2 / SXX}} \sim t(n-2), \]

so we can do inference on \( \eta_1 \), using

\[ t = \frac{\hat{\eta}_1 - \eta_h}{\sqrt{\hat{\sigma}^2 / SXX}} \]

as our test statistic.

**Inference on \( \eta_0 \)**

With the same assumptions, it can be shown in an analogous manner (details omitted) that

\[ \frac{\hat{\eta}_0 - \eta_0}{s.e.(\hat{\eta}_0)} \sim t(n-2), \]

so we can use this statistic to do inference on \( \eta_0 \).