INFERENCE FOR SIMPLE OLS

Model Assumptions *("The" Simple Linear Regression Model Version 3):*
(We consider $x_1, \dots, x_n$ as fixed.)

1. $E(Y|x) = \eta_0 + \eta_1 x$ (linear mean function)
2. $Var(Y|x) = \sigma^2$ (Equivalently, $Var(e|x) = \sigma^2$) (constant variance)
3. $y_1, \dots, y_n$ are independent observations. (independence)
4. (*NEW)* $Y|x$ is normal for each x (normality)

(1) + (2) + (4) can be summarized as:

$$Y|x \sim N(\eta_0 + \eta_1 x, \sigma^2)$$

Comments: 1. For some purposes, we need only assume (4) for $x = x_i$'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: $e|x = Y|x - E(Y|x)$

So: $e|x \sim N(0, \sigma^2)$

i.e., all errors have the same distribution -- so we just say e instead of e|x .

Since $\hat{\eta}_0$ and $\hat{\eta}_1$ are linear combinations of the $Y|x_i$'s, (3) + (4) imply that $\hat{\eta}_0$ and $\hat{\eta}_1$ are normally distributed random variables (that is, their sampling distributions are normal). Recalling that

$$E(\hat{\eta}_1) = \eta_1 \qquad Var((\hat{\eta}_1) = \frac{\sigma^2}{SXX} \qquad E(\hat{\eta}_0) = \eta_0 \qquad Var\,(\hat{\eta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right),$$

We have

$$\hat{\eta}_1 \sim \qquad\qquad\qquad \hat{\eta}_0 \sim$$

Look more at $\hat{\eta}_1$: We can standardize to get

$$\frac{\hat{\eta}_1 - \eta_1}{\sqrt{\sigma^2 / SXX}} \sim N(0,1)$$

But we don't know σ^2, so need to approximate it by $\hat{\sigma}^2$ -- in other words approximate Var($\hat{\eta}_1$) by $V\hat{a}r(\hat{\eta}_1)$= [s.e. ($\hat{\eta}_1$)]^2 $= \frac{\hat{\sigma}^2}{SXX}$. Thus we want to use $\frac{\hat{\eta}_1 - \eta_1}{\sqrt{\hat{\sigma}^2 / SXX}}$. But we can't expect this to be normal, too. However,

$$\frac{\hat{\eta}_1 - \eta_1}{\sqrt{\hat{\sigma}^2 / SXX}} =$$

$$(*) \qquad \left(\frac{\hat{\eta}_1 - \eta_1}{\sqrt{\sigma^2 / SXX}} \right) \Big/ \sqrt{\hat{\sigma}^2 / \sigma^2}$$

The numerator of (*) *is* normal (in fact, standard normal), as noted above.

Facts: (Proofs omitted)

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

Notation: $\quad$ (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 $\hat{\eta}_1 - \eta_1$ (hence independent of the numerator in (*))

Comments on distributions:

1. A χ^2(k) distribution is defined as the distribution of a random variable which is a sum of squares of k independent standard normal random variables.
[Comment: Recall that $\hat{\sigma}^2 = \frac{1}{n-2} RSS$, so (n-2)$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{RSS}{\sigma^2} = \sum \left(\frac{\hat{e}_i}{\sigma} \right)^2$ *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. A t-distribution with k degrees of freedom is defined as the distribution of a random variable of the form $\frac{Z}{\sqrt{U / k}}$ where

- Z~N(0,1)
- U~ χ^2(k)
- Z and U are independent.

In the fraction (*) above, take

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

$$Z = \frac{\hat{\eta}_1 - \eta_1}{\sqrt{\sigma^2/SXX}} \sim N(0,1)$$

Thus: $$\frac{\hat{\eta}_1 - \eta_1}{\sqrt{\hat{\sigma}^2/SXX}} \sim t(n-2),$$

so we can do inference on η_1, using $t = \frac{\hat{\eta}_1 - \eta_1}{\sqrt{\hat{\sigma}^2/SXX}}$ as our test statistic.

Inference on η_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 η_0.