Estimating σ^2

- We now know a lot about estimating $\boldsymbol{\beta}$.
- There's another parameter in the regression model - σ^2 – the variance of each ε_i .
- Note that $\sigma^2 = var.(\varepsilon_i) = E[(\varepsilon_i - E(\varepsilon_i))^2] = E(\varepsilon_i^2)$.
- The *sample* counterpart to this *population* parameter is the *sample* average of the "residuals": $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\boldsymbol{e}'\boldsymbol{e}$.
- However, there is a *distortion* in this estimator of σ^2 .
- Although mean of e_i's is zero (if intercept in model), not all of e_i's are independent of each other – only $(n - k)$ of them are.

It can be shown that:

$$E(\hat{\sigma}^2) = E(\frac{1}{n}\boldsymbol{e}'\boldsymbol{e}) = \frac{1}{n}(n-k)\sigma^2 < \sigma^2 \quad ; \quad \textbf{BIASED}$$

Easy to convert this to an ***Unbiased estimator –***

$$s^2 = \frac{1}{(n-k)}\boldsymbol{e}'\boldsymbol{e}$$

- "$(n - k)$" is the "*degrees of freedom*" – number of independent sources of information in the "*n*" residuals (e_i's).
- We can use "*s*" as an estimator of σ, but it is a *biased estimator*.
- Call "*s*" the "*standard error of the regression*", or the "*standard error of estimate*".
- s^2 is a *statistic* – has its own sampling distribution, *etc*. More on this to come.
- Let's see one immediate *application* of s^2 and *s*.
- Recall sampling distribution for LS estimator, ***b***:

$$\boldsymbol{b} \sim N[\boldsymbol{\beta}\ ,\ \sigma^2(X'X)^{-1}]$$

- So, $var.(b_i) = \sigma^2[(X'X)^{-1}]_{ii}$; σ^2 is *unobservable*.
- If we want to report variability associated with b_i as an estimator of β_i, we need to use an estimator of σ^2 .

- $est.var.(b_i) = s^2[(X'X)^{-1}]_{ii}$.
- $\sqrt{est.var.(b_i)} = \widehat{s.d.}(b_i) = s\{[(X'X)^{-1}]_{ii}\}^{1/2}$.
- We call this the "*standard error*" of b_i.
- This quantity will be very important when it comes to constructing *interval estimates* of our regression coefficients; and when we construct *tests of hypotheses* about these coefficients.

Confidence Intervals & Hypothesis Testing

- So far, we've concentrated on "*point*" estimation.
- We will make use of the assumption of *Normally distributed* errors.
- Recall that:

$$\boldsymbol{b} \sim N[\boldsymbol{\beta}\,, \sigma^2(X'X)^{-1}]$$

$$b_i \sim N[\beta_i\,, \sigma^2((X'X)^{-1})_{ii}] \quad ; \quad \text{why still } \textit{Normal}?$$

- So, we can *standardize*:

$$z_i = (b_i - \beta_i)/\sqrt{\sigma^2[(X'X)^{-1}]_{ii}}$$

- But σ^2 is *unknown*, so we can't use z_i directly to draw inferences about b_i. We must replace σ^2 with an estimator, e.g. s^2

When we replace σ^2 with s^2 in the formula for z_i, z_i is no longer Normally distributed. Instead, the statistic follows a Student-t distribution, and we call it a *t* statistic. That is:

$$\left[\frac{b_i - \beta_i}{s.e.(b_i)}\right] \sim t_{(n-k)}$$

We can use this to construct *confidence intervals* and *test hypotheses* about β_i .

Note: This result used all of our assumptions about the linear regression model – including the assumption of *Normality for the errors*.

Note: The t-distribution becomes the Normal distribution as the sample size grows.

Try running the following R code one line at a time:

```
par(lwd = 3)
#”dt” is the density function for the t-distribution
curve(dt(x,5), from=-3, to=3, col = "red", ylab = "", ylim = c(0, 0.4))
par(new=TRUE)
curve(dt(x,100), from=-3, to=3, col = "blue", ylab = "", ylim = c(0, 0.4))
par(new=TRUE)
curve(dt(x,1000), from=-3, to=3, col = "green", ylab = "", ylim = c(0, 0.4))
par(new=TRUE)
#”dnorm” is the density function for the Normal distribution
curve(dnorm(x), from=-3, to=3, col = "pink", ylab = "", ylim = c(0, 0.4))
```

Degrees of freedom for the t-distribution (n-k)

```
#Area under the curve, to the left of "-2", t-distribution for various d.o.f.
pt(-2, 5)
pt(-2, 100)
pt(-2, 1000)
#Area under the curve, to the left of "-2", standard Normal distribution
pnorm(-2)

#To get the 10th percentile from a t-distribution with 15 d.o.f., for example
qt(0.1, 15)
```

Example 1:

$$\hat{y} = 1.4 + 0.2x_2 + 0.6x_3$$
$$(0.7) \quad (0.05) \quad (1.4)$$

$$H_0\!: \ \beta_2 = 0 \qquad vs. \qquad H_A\!: \ \beta_2 > 0$$

$$t = \left[\frac{b_2 - \beta_2}{s.e.(b_2)}\right] = \left[\frac{0.2-0}{0.05}\right] = 4$$; suppose $n = 20$

$$t_c(5\%) = 1.74 \ ; \ t_c(1\%) = 2.567$$; d.o.f. = 17

$$t > t_c \ \Rightarrow \ Reject \ H_0 \ .$$

Degrees of Freedom	90th Percentile	95th Percentile	97.5th Percentile	99th Percentile	99.5th Percentile
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
:	:	:	:	:	:
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	**1.333**	**1.740**	**2.110**	**2.567**	**2.898**

Example 2:

$$\hat{y} = 1.4 + 0.2x_2 + 0.6x_3$$
$$(0.7) \quad (0.05) \quad (1.4)$$

$H_0\!: \ \beta_1 = 1.5 \quad vs. \quad H_A\!: \ \beta_1 \neq 1.5$

$t = \left[\frac{b_1 - \beta_1}{s.e.(b_1)}\right] = \left[\frac{1.4-1.5}{0.7}\right] = -0.1429$; d.o.f. = 17

$t_c(5\%) = \pm 2.11$

$|t| < t_c \quad \Rightarrow \quad Do\ Not\ Reject\ H_0$

(Against H_A , at the 5% significance level.)

Example 3:

$$\hat{y} = 1.4 + 0.2x_2 + 0.6x_3$$
$$(0.7) \quad (0.05) \quad (1.4)$$

$H_0\!: \ \beta_1 = 1.5 \quad vs. \quad H_A\!: \ \beta_1 < 1.5$

$t = \left[\frac{b_1 - \beta_1}{s.e.(b_1)}\right] = \left[\frac{1.4-1.5}{0.7}\right] = -0.1429$; d.o.f. = 17

$p - value = Pr.[t < -0.1429\ | H_0\ is\ True]$

in R: `pt(-0.1429,17)`

***p* = 0.444**

What do you conclude?

Some Properties of Tests:

Null Hypothesis (H_0) Alternative Hypothesis (H_A)

Classical hypothesis testing –

- Assume that H_0 is *TRUE*
- Compute value of test statistic using random sample of data
- Determine *distribution* of the test statistic (*when* H_0 *is true*)
- Check of observed value of test statistic is likely to occur, *if* H_0 *is true*
- If this event is sufficiently *unlikely*, then **REJECT H_0** (in favour of H_A)

Note:

1. Can never **accept** H_0. Why not?
2. What constitutes "*unlikely*" – subjective?
3. Two types of errors we might incur with this process

   Type I Error: **Reject** H_0 when in fact it is **True**

   Type II Error: **Do Not Reject** H_0 when in fact it is **False**

   - Pr.[I] = α = Significance level of test = "size" of test
   - Pr.[II] = β ; say
   - Value of β will depend on <u>*how*</u> H_0 is **False**. Usually, many ways.

Definition:

The "**Power**" of a test is Pr.[**Reject** H_0 when it is **False**].

So, Power = 1 – Pr.[**Do Not Reject** H_0 | H_0 is **False**] = 1 – β.

- As β typically changes, depending on the *way* that H_0 is false, we usually have a **Power Curve**.
- For a fixed value of α, this curve plots Power against parameter value(s).
- We want our tests to have *high power*.
- We want the power of our tests to *increase* as H_0 becomes *increasingly false*.

Confidence Intervals

We can also use our t-statistic to construct a confidence interval for β_i.

$$Pr.[-t_c \leq t \leq t_c] = (1 - \alpha)$$

$$\Rightarrow \quad Pr.\left[-t_c \leq \left[\frac{b_i - \beta_i}{s.e.(b_i)}\right] \leq t_c\right] = (1 - \alpha)$$

$$\Rightarrow \quad Pr.[-t_c\, s.e.(b_i) \leq (b_i - \beta_i) \leq t_c\, s.e.(b_i)] = (1 - \alpha)$$

$$\Rightarrow \quad Pr.[-b_i - t_c\, s.e.(b_i) \leq (-\beta_i) \leq -b_i + t_c\, s.e.(b_i)]$$

$$= (1 - \alpha)$$

$$\Rightarrow \quad Pr.[b_i + t_c\, s.e.(b_i) \geq \beta_i \geq b_i - t_c\, s.e.(b_i)] = (1 - \alpha)$$

$$\Rightarrow \quad Pr.[b_i - t_c\, s.e.(b_i) \leq \beta_i \leq b_i + t_c\, s.e.(b_i)] = (1 - \alpha)$$

Interpretation –

The interval, $[b_i - t_c\, s.e.(b_i)\ ,\ b_i + t_c\, s.e.(b_i)]$ is *random*.

The parameter, β_i, is *fixed* (but unknown).

> If we were to take a sample of *n* observations, and construct such an interval, and then repeat this exercise many, many, times, then $100(1 - \alpha)\%$ of such intervals would cover the true value of β_i.

If we just construct an interval, for our *given* sample of data, we'll never know if *this particular* interval covers β_i, or not.

Example 1

$$\hat{y} = 0.3 - 1.4x_2 + 0.7x_3$$

(0.1) (1.1) (0.2)

Construct a 95% confidence interval for β_1 when n = 30.

d.o.f. = $(n - k)$ = 27 ; (α/2) = 0.025

$t_c = \pm 2.052$; $b_1 = 0.3$; $s.e.(b_1) = 0.1$

The 95% Confidence Interval is:

$$[b_1 - t_c\, s.e.(b_1) \;, \qquad b_1 + t_c\; s.e.(b_1)]$$

⇒ [0.3 – (2.052)(0.1) , 0.3 + (2.052)(0.1)]

⇒ [0.0948 , 0.5052]

Don't forget the units of measurement!

Example 2

$$\hat{y} = 0.3 - 1.4x_2 + 0.7x_3$$

(0.1) (1.1) (0.2)

Construct a 90% confidence interval for β_2 when n = 16.

d.o.f. = $(n - k)$ = 13 ; (α/2) = 0.05

$t_c = \pm 1.771$; $b_2 = -1.4$; $s.e.(b_2) = 1.1$

The 95% Confidence Interval is:

$$[b_2 - t_c\, s.e.(b_2) \;, \qquad b_2 + t_c\; s.e.(b_2)]$$

⇒ [-1.4 – (1.771)(1.1) , -1.4 + (1.771)(1.1)]

⇒ [-3.3481 , 0.5481]

Questions:

- Why do we construct the interval *symmetrically* about point estimate, b_i?
- How can we use a Confidence Interval to test hypotheses?
- For instance, in the last Example, can we reject H_0: $\beta_2 = 0$, against a 2-sided alternative hypothesis?