EC508: Econometrics
Standard Errors for $\hat{\beta}_1$

Jean-Jacques Forneron

Spring, 2023

Boston University

Central Limit Theorem

- Let $Z_1, \ldots, Z_n$ be iid with $\mathbb{E}(Z_i^2) = \sigma_Z^2 < \infty$
- Then:

$$t_n = \sqrt{n}\frac{\bar{Z}_n - \mathbb{E}(Z_i)}{\sigma_Z} \xrightarrow{d} \mathcal{N}(0,1)$$

- $\xrightarrow{d}$ is the convergence in distribution:

$$\mathbb{P}(t_n \in [a,b]) \to \mathbb{P}(t \in [a,b]), \text{ as } n \to \infty \text{ with } t \sim \mathcal{N}(0,1)$$

- In R, you can compute $\mathbb{P}(t \leq b)$ using *pnorm(b)*

Slutsky's Theorem

- Let Z_n, W_n be random variables such that:

$$Z_n \xrightarrow{d} Z \text{ (r.v.)}$$
$$W_n \xrightarrow{p} c \text{ (constant)}$$

- Then:

  i. $Z_n + W_n \xrightarrow{d} Z + c$
  ii. $Z_n W_n \xrightarrow{d} Z \times c$
  iii. $Z_n / W_n \xrightarrow{d} Z/c$ if $c \neq 0$

Central Limit Theorem for $\hat{\beta}_1$

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

- LLN: $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{p} \sigma_X^2$
- Let $v_i = (X_i - \mathbb{E}(X_i))u_i$, then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)u_i = \frac{1}{n}\sum_{i=1}^{n} v_i + \frac{1}{n}\sum_{i=1}^{n}(\bar{X}_n - \mathbb{E}(X_i))u_i$$

- CLT+WLLN+Slutsky:
  $\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(\bar{X}_n - \mathbb{E}(X_i))u_i = \sqrt{n}(\bar{X}_n - \mathbb{E}(X_i))\bar{u}_n \xrightarrow{d} 0$
- CLT: $\sqrt{n}\bar{v}_n \xrightarrow{d} \mathcal{N}(0, \sigma_v^2)$

Central Limit Theorem for $\hat{\beta}_1$, cont'd

Putting everything together:

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) &= \frac{\sqrt{n}\bar{v}_n}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} + \frac{\sqrt{n}(\bar{X}_n - \mathbb{E}(X_i))\bar{u}_n}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} \\ &\xrightarrow{d} \frac{1}{\sigma_X^2}\mathcal{N}(0, \sigma_v^2)\end{aligned}$$

$\hat{\beta}_1$ is asymptotically normal with asymptotic variance given by $\sigma_v^2/[\sigma_X^2]^2$, this allows us to compute the standard errors:

$$se(\hat{\beta}_1) = \frac{1}{\sqrt{n}}\frac{\sigma_v}{\sigma_X^2}$$

Such that

$$\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \xrightarrow{d} \mathcal{N}(0, 1)$$

Hypothesis Testing and the Standard Error of $\hat{\beta}_1$ (SW 5.1)

- The objective is to test a hypothesis, like $\beta_1 = 0$, using data – to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.
- **General setup**:
  - Null hypothesis and two-sided alternative:

    $$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 \neq \beta_{1,0}$$

    where $\beta_{1,0}$ is the hypothesized value under the null.
  - Null hypothesis and one-sided alternative:

    $$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 < \beta_{1,0}$$

General approach: construct t-statistic, and compute p-value (or compare to the $\mathcal{N}(0,1)$ critical value)

- In general:

$$t = \frac{\text{estimator - hypothesized value}}{\text{standard error of the estimator}}$$

  where the SE of the estimator is the square root of an estimator of the variance of the estimator.

- For testing the mean of Y:

$$t = \sqrt{n}\frac{\bar{Y} - \mu_Y}{\sigma_Y}$$

- For testing β_1

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \xrightarrow{d} \mathcal{N}(0,1)$$

  under H_0, where $se(\hat{\beta}_1) =$ the square root of an estimator of the asymptotic variance of the sampling distribution of $\hat{\beta}_1$

Computing $se(\hat{\beta}_1)$

- Recall the expression for the asymptotic variance of $\hat{\beta}_1$ (large n):

$$\sigma^2_{\beta} = \frac{\sigma^2_v}{[\sigma^2_X]^2}, \text{ where } v_i = (X_i - \mu_X)u_i$$

- The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_v and σ_X by estimators constructed from the data:

$$\hat{\sigma}^2_{\beta} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{v}_i^2}{[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]^2}$$

where $\hat{v}_i = (X_i - \bar{X}_n)\hat{u}_i$; $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$

Computing $se(\hat{\beta}_1)$

$$\hat{\sigma}_{\beta}^2 = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{v}_i^2}{[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]^2}$$

- $se(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\beta}^2 / n}$
- This is a bit nasty, but:
  - It is less complicated than it seems. The numerator estimates $var(v_i)$, the denominator estimates $[var(X)]^2$.
  - Why the degrees-of-freedom adjustment n–2? Because two coefficients have been estimated $(\hat{\beta}_0, \hat{\beta}_1)$.
  - $SE(\hat{\beta}_1)$ is computed by regression software
  - R has memorized this formula so you don't need to.

Application to the California Test Score in R

```r
# packages to compute standard errors
library(sandwich)
library(lmtest)

library(foreign)
data = read.dta('caschool.dta')
data$score = 0.5*(data$math_scr + data$read_scr)
linear_model = lm(score~str,data=data)

# compute standard errors, t-statistics
coeftest(linear_model, vcov. = vcovHC)
```

Table 1: Coefficients, Standard Errors, t-statistics and p-values

	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	698.93295	10.46054	66.8162	$<$ 2.2e-16 ***
str	-2.27981	0.52436	-4.3478	1.729e-05 ***