EC508: Econometrics
Testing Hypotheses about β_1

Jean-Jacques Forneron

Spring, 2023

Boston University

Two Sided Test $H_0 : \beta_1 = \beta_{1,0}$ vs. $\beta_1 \neq \beta_{1,0}$

- Construct the t-statistic

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

- Reject at 5% significance level if $|t^{act}| > 1.96$
- The p-value is $p = \mathbb{P}(|t| > |t^{act}|) =$ probability in tails of a standard normal distribution outside $|t^{act}|$; you reject at the 5% significance level if the p-value is $< 5\%$.
- This procedure relies on the large-n approximation that is normally distributed; typically n $= 50$ is large enough for the approximation to be excellent.

One Sided Test $H_0 : \beta_1 = \beta_{1,0}$ vs. $\beta_1 > \beta_{1,0}$

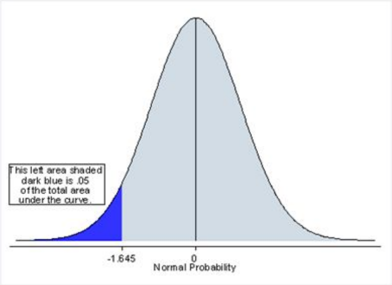
- Construct the t-statistic

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

- Reject at 5% significance level if $t^{act} > 1.64$
- The p-value is $p = \mathbb{P}(t > t^{act}) =$ probability in tails of a standard normal distribution over t^{act}; you reject at the 5% significance level if the p-value is $< 5\%$.

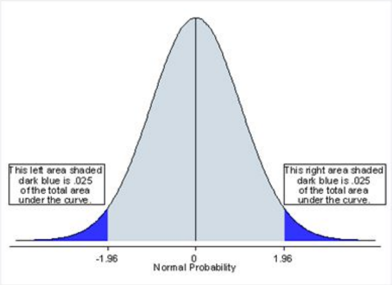
One-tailed vs two-tailed t-test

One-tailed t-test



A one-tailed test will test either if the mean is significantly greater than x or if the mean is significantly less than x, but not both. The one-tailed test provides more power to detect an effect in one direction by not testing the effect in the other direction.

Two-tailed t-test



A two-tailed test will test both if the mean is significantly greater than x and if the mean significantly less than x. The mean is considered significantly different from x if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05.

Example: Test Scores and STR, California data

- Estimated regression line:

$$\text{TEST SCORE} = 698.9\text{–}2.28 \times \text{STR}$$

- R reports the standard errors:

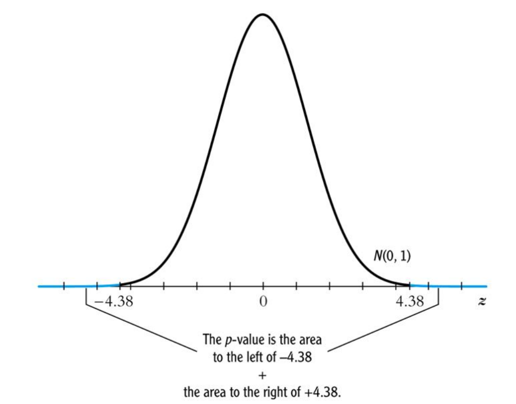
$$SE(\hat{\beta}_0) = 10.4, \quad SE(slope) = 0.52$$

- t-statistic testing $\beta_{1,0} = 0$

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{-2.28 - 0}{0.52} = \text{–}4.38$$

- The 1% 2-sided significance level is 2.58, so we reject the null at the 1% significance level.
- Alternatively, we can compute the p-value. In R, simply type *2*(1-pnorm(4.38))* which yields 1.2×10^{-5}.

Example: Test Scores and STR, California data



The p-value based on the large-n standard normal approximation to the t-statistic is 0.00001 (10^{-5})

Confidence Intervals for β_1 (SW 5.2)

- Recall that a 95% confidence interval is, equivalently:
  - The set of points that cannot be rejected at the 5% significance level;
  - A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

- Because the t-statistic for β_1 is N(0,1) in large samples, construction of a 95% confidence for β_1 is just like the case of the sample mean:

$$\text{95\% confidence interval for } \beta_1 = \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$$

Confidence interval example: Test Scores and STR Estimated regression line: Test Score = 698.9 – 2.28×STR

$$SE(\hat{\beta}_0) = 10.4, \quad SE(slope) = 0.52$$

- 95% Confidence Interval for $\hat{\beta}_1$:

  $$\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} = \{-2.28 \pm 1.96 \times 0.52\} = \{-3.30, -1.26\}$$

- The following two statements are equivalent (why?)
  - The 95% confidence interval does not include zero;
  - The hypothesis $\beta_1 = 0$ is rejected at the 5% level

A concise (and conventional) way to report regressions

Put standard errors in parentheses below, or after, the estimated coefficients to which they apply.

$\widehat{\text{Test Score}} = 698.9\,(10.4) – 2.28 \times STR\,(0.52), R^2 = .05, SER = 18.6$

This expression gives a lot of information:

- The estimated regression line is

$$\widehat{\text{Test Score}} = 698.9 – 2.28 \times STR$$

- The standard error of $\hat{\beta}_0$ is 10.4
- The standard error of $\hat{\beta}_1$ is 0.52
- The R^2 is .05; the standard error of the regression is 18.6

Application to the California Test Score in 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	$Pr(>\lvert t \rvert)$
(Intercept)	698.93295	10.46054	66.8162	$<$ 2.2e-16 ***
str	-2.27981	0.52436	-4.3478	1.729e-05 ***

Constructing a 95% Confidence Interval in R

```r
library(foreign)
library(sandwich)
library(lmtest)

data = read.dta('caschool.dta')

linear_model = lm(score~str,data=data)

se = sqrt(vcovHC(linear_model)[2,2])
CI = linear_model$coef[2] + 1.96*se*c
    (-1,1)
print(CI,digits=3) # Print CI with 2
    decimal points
```

95% Confidence Interval

The 95% Confidence Interval for β_1 is $[-3.31, -1.25]$

Summary of statistical inference about β_0 and β_1

- **Estimation**:
  - OLS estimators $\hat{\beta}_0$ and and $\hat{\beta}_1$ have approximately normal distribution in large samples
- **Testing**:
  - $H_0 : \beta_1 = \beta_{1,0}$ vs. $\beta_1 \neq \beta_{1,0}$ ($\beta_{1,0}$ is the value of β_1 under H_0)
  - $t = (\hat{\beta}_1 - \beta_{1,0})/SE(\hat{\beta}_1)$
  - p-value = area under standard normal outside t^{act} (large n)
- **Confidence Intervals**:
  - 95% confidence interval for β_1 is $\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$
  - This is the set of β_1 that is not rejected at the 5% level
  - The 95% CI contains the true β_1 in 95% of all samples.