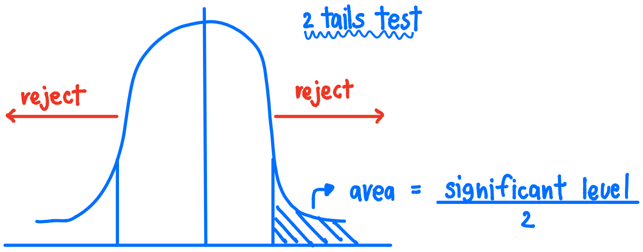
4 Testing Hypotheses about a Single Linear Combination of the Parameter

Consider

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 \exp er + u$$

where jc = number of years attending a two-year college
$univ$ = number of years at a four-year college
$\exp er$ = months in the workforce.
We want to test whether $\beta_1 = \beta_2$. → if the returns from 1 more years of education at a Junior College is the same as that of the university

$H_0: \beta_1 = \beta_2 \Rightarrow H_0: \beta_1 - \beta_2 = 0$
against
$H_a: \beta_1 \neq \beta_2 \Rightarrow H_a: \beta_1 - \beta_2 \neq 0$



$t = \dfrac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{s.e.(\hat{\beta}_1 - \hat{\beta}_2)}$ → we compute this t-statistic and compair with the critical value

where $s.e.(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{var}(\hat{\beta}_1 - \hat{\beta}_2)}$

$= \sqrt{var(\hat{\beta}_1) + var(\hat{\beta}_2) - 2cov(\hat{\beta}_1, \hat{\beta}_2)}$

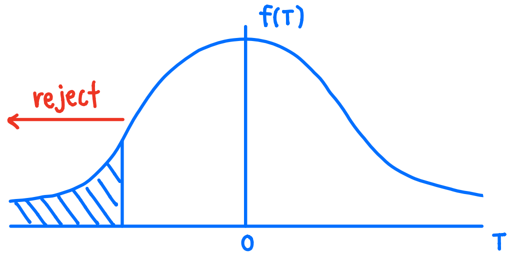
not very straight forward to calculate
↝ we use a variable tranformation trick ↝ see notes.

another possible hypothesis test (one-tailed alternative) going to university give more return than going to college

$H_0: \beta_1 = \beta_2 \Rightarrow H_0: \beta_1 - \beta_2 = 0$

$H_a: \beta_1 < \beta_2 \Rightarrow H_a: \beta_1 - \beta_2 < 0$

- It is assume that β_1 would not be more than β_2 (return to a 2-year college would never be more than return to university education)

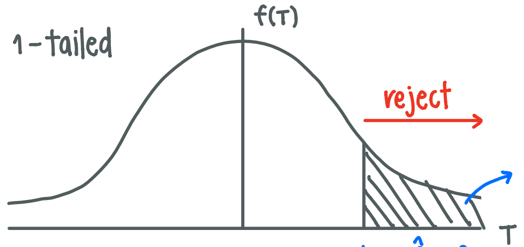


$$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{s.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

※ Then, go to the extra note

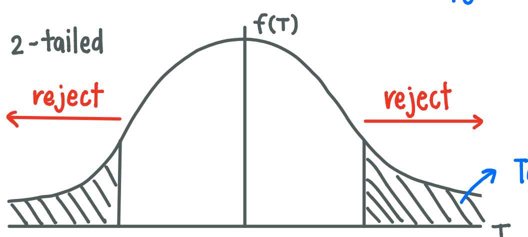
5 Computing p-Values for t-Tests

- What is the significance level given the computed t-statistics?



This shaded area in the rejection is the p-value

$$t = \frac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j}$$



Total area from 2 sides

$t =$ computed, $t = \dfrac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j}$

- p-value : $P(|T| > |t|)$

T = t-distributed random variable with d.f. $= n-k-1$

t = computed t-statistic

↝ p-value = probability that a random T value will be greater (in the | | term) than our t-statistic in the H_0 testing

In class exercise

Consider the multiple regression model, assume MLR 1-6 are satisfied.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

You would like to test the $H_0: \beta_1 - 3\beta_2 = 1$

$H_a:$ Otherwise is true

1) write the T-statistic for testing H_0

$$t = \frac{(\hat{\beta}_1 - 3\hat{\beta}_2) - 1}{\text{s.e.}(\hat{\beta}_1 - 3\hat{\beta}_2)}$$

2) Define $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2 \Rightarrow H_0: \theta_1 = 1$

$H_a: \theta_1 \neq 1$

$t = \frac{\hat{\theta}_1 - 1}{\text{s.e.}(\hat{\theta}_1)} \longrightarrow$ we need our regression to have θ_1 in it

So, STATA or OLS estimator will automatically give $\hat{\theta}_1$ and s.e. $\hat{\theta}_1$

Now, $\hat{\beta}_1 = \hat{\theta}_1 + 3\hat{\beta}_2$

or $\beta_1 = \theta_1 + 3\beta_2$

Substitute in the main regression and get

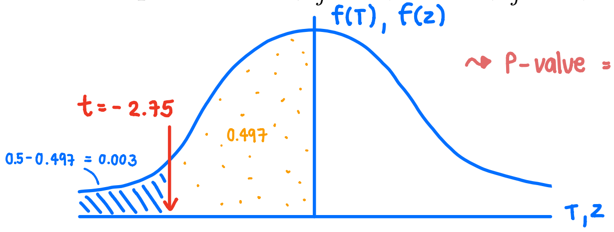
$$\begin{aligned} y &= \beta_0 + (\theta_1 + 3\beta_2)x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\ &= \beta_0 + \theta_1 x_1 + 3\beta_2 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \\ &= \beta_0 + \theta_1 x_1 + \beta_2(x_2 + 3x_1) + \beta_3 x_3 + u \end{aligned}$$

★ Now, the explanatory variables are going to be x_1, $x_2 + 3x_1$, and x_3

- we can calculate $t = \frac{\hat{\theta}_1 - 1}{\text{s.e.}\ \hat{\theta}_1}$

Example 1: $H_0 : \beta_j \geq 0$, $H_a : \beta_j < 0$, d.f.$= 140$.

z-table



P-value = what shaded be the significant level given the critical value of -2.75
⇒ find the shaded area

suppose the calculated $t_{\hat{\beta}_j} = -2.75$

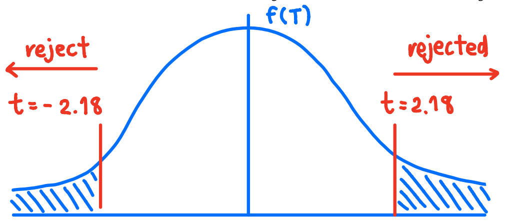
$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j}$

- From the z-table, the value -2.75 corresponds to area = _ _0.003_ _ _ _ _ _ _ _ _.
- Thus, p-value = _ _ _0.003_ _ _ _ _.
- Would we reject H_0 if we use the significance level = 5%? Yes

RULE : we reject Ho if p-value < significant level

Example 2: $H_0 : \beta_j = a_j$, $H_a : \beta_j \neq a_j$, d.f.$= 18$.

t-table

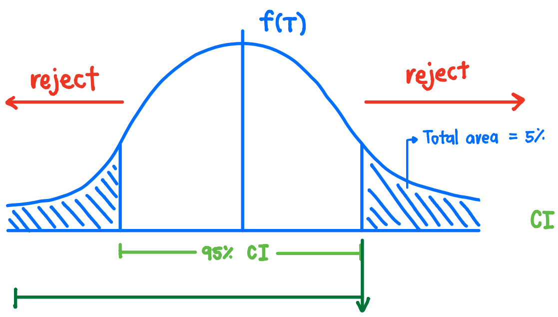


suppose the calculated $t_{\hat{\beta}_j} = -2.18$

- From the t-table, the value -2.18 corresponds to area = _0.02 to 0.05 _ _ _ _ _.
- Thus, p-value = is between 0.02 – 0.05 _.
- Would we reject H_0 if we use the significance level = 5%?
  Yes, reject Ho because the area is less than 0.05
  or p-value < 0.05

6 Confidence Intervals (CI)

- Confidence Intervales for the POPULATION PARAMETER (β_j)
  → The range of value that would capture the true β_j at a 95% change
- A 95% CI of β_j is given by



CI ⇒ $\hat{\beta}_j \pm C \times s.e.(\hat{\beta}_j)$

C is the 97.5 percentile in the t-distribution with n-k-1 df.

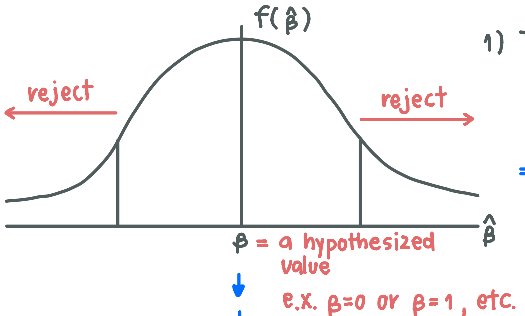
want to know exactly here
SO C = 97.5%

Inference → Hypothesis testing about "β" the true parameter

wage = $\beta_0 + \beta_1 educ + \beta_2 exper + \dots + u$

we want to test hypotheses about the true impact (β) of each x variable (education, experience) on the dependent variable (Y)

BUT. We don't know what the true β are. So, we use $\hat{\beta}$ (estimator) and s.e. ($\hat{\beta}$) to test the hypotheses



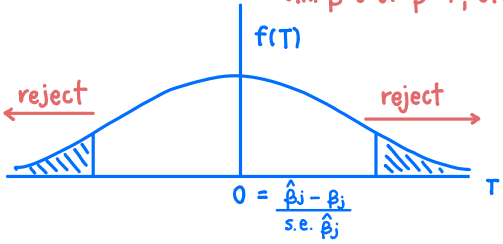
1) Test is β = same number

e.g. $\beta_j = 0$ → x_j has no impact on Y

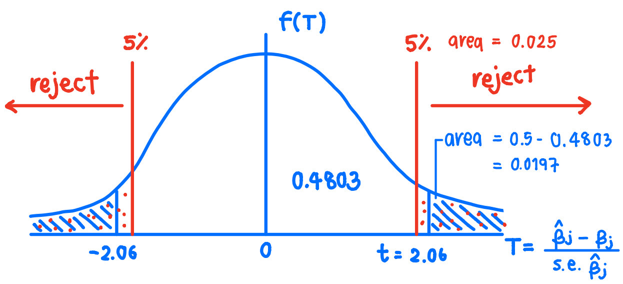
$\beta_j = 1$ → 1 unit ↑ in x_j correspond to 1 unit ↑ in Y

⇒ t-test

$$\frac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j} \sim t_{d.f.}$$



Significant level = total area in the rejection region



- Suppose, we calculate ∝ ass. d.f. 100

  t-statistic = $\frac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j}$ = 2.06

- Suppose we are testing, $H_0: \beta_j = 0$

  $H_a: \beta_j \neq 0$ 2-tailed test

- p-value = total shaded area

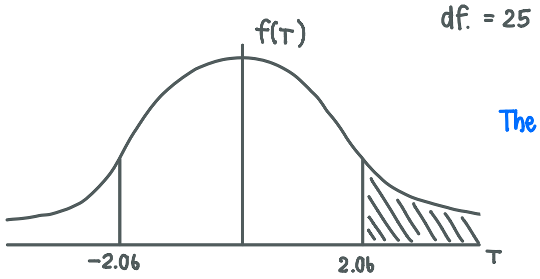
P-value = significant level when we will reject the Ho or prob that we will reject Ho

= 2(0.5 - 0.4803)

= 0.0394

* If p-value < significant level ⇒ reject Ho

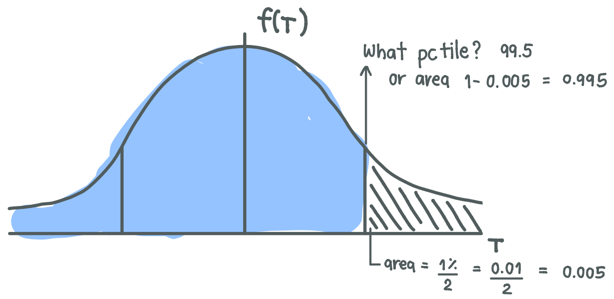
Example 1: **95% CI**



The 95% CI for $\hat{\beta}_j = [\hat{\beta}_j - 2.06 \cdot s.e.(\hat{\beta}_j), \hat{\beta}_j + 2.06 \cdot s.e.(\hat{\beta}_j)]$

value of 97.5th percentile
in the t_{25} distribution

Example 2: **99% CI**



The 99% CI for $\hat{\beta}_j = [\hat{\beta}_j - 2.787 \cdot s.e.(\hat{\beta}_j), \hat{\beta}_j + 2.787 \cdot s.e.(\hat{\beta}_j)]$

F-test motivation

⇒ We want to the test the significance of a group of hypotheses (multiple hypotheses)

$Grade_{325} = \beta_0 + \beta_1 * times\_front + \beta_2 * times\_back + \beta_3\, hours\_study + \beta_4\, past\_GPA + \beta_5\, gender + u$

H_0: seat position doesn't have impact on GPA

$\beta_1 = 0$ and $\beta_2 = 0 \Rightarrow \beta_1 = \beta_2 = 0$

H_a: seat position matters

$\beta_1 \neq 0$ and $\beta_2 \neq 0$ or $\beta_1 \neq 0$ and $\beta_2 = 0$, $\beta_1 = 0$ and $\beta_2 \neq 0$ } at least one of the $\beta_1, \beta_2 \neq 0$