## 4 Testing Hypotheses about a Single Linear Combination of the Parameter

Consider

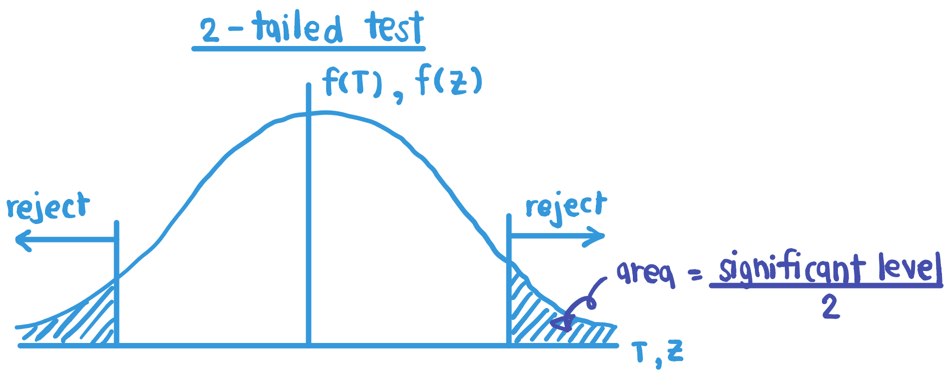
$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 \exp er + u$$

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 year of education at a Junior College is the same as that of the university.

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

agaist

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




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

← we compute this t-statistic and compare with the critical value

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

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

↓

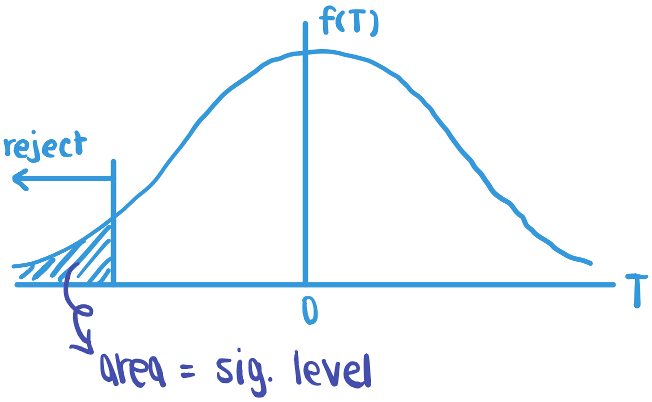
not very straight forward to calculate.

→ We use a variable transformation trick → see notes another possible hypothesis test (one-tailed alternative)

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

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

- It is assumed that $\beta_1$ would not be more than $\beta_2$ (returns to a 2-year college would never be more than returns to university education



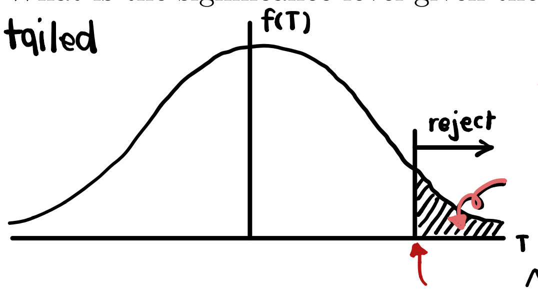

$$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{s.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

* Then, go to the extra note.

## 5 Computing p-Values for t-Tests

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

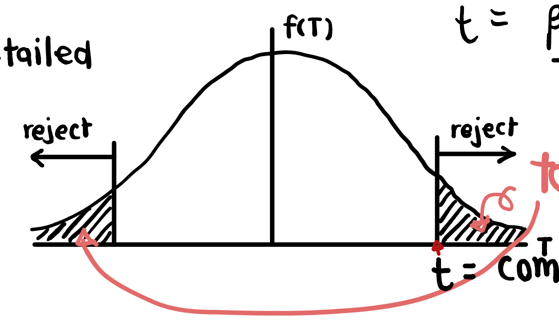
1-tailed




This shaded area in the rejection region is the p-value

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

2-tailed




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.

$\rightsquigarrow$ p-value = probability that a random T value will be greater (in the | | term) than our t in the $H_0$ test.

## 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

1st) write the t-statistic for testing $H_0$

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

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

$$t = \frac{\hat{\theta}_1 - 1}{s.e.(\hat{\theta}_1)} \Rightarrow$$ We need our regression to have $\theta_1$ in it. So, STATA or OLS estimation will automatically give $\hat{\theta}_1$ & s.e. $\hat{\theta}_1$

NOW, $\hat{\beta}_1 = \hat{\theta}_1 + 3\hat{\beta}_2$
or $\beta_1 = \theta_1 + 3\beta_2$

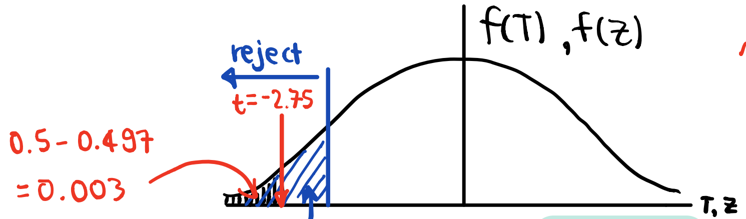
Sub in the main regression and get

$$y = \beta_0 + (\theta_1 + 3\beta_2)X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$
$$= \beta_0 + \theta_1 X_1 + 3\beta_2 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$
$$= \beta_0 + \theta_1 X_1 + \beta_2 (X_2 + 3X_1) + \beta_3 X_3 + u$$

☆ Now, the explanatory variables are going to be $X_1$, $X_2 + 3X_1$, and $X_3$

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

Example 1: $H_0 : \beta_j \# 0$, $H_a : \beta_j < 0$, d.f.= 140. ⇒ z-table




⇒ P-value = what should be the significant level, given the critical value of -2.75 ??

⇒ find the shaded area!

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

0.05

↳ $t_{\hat{\beta}_j} = (\hat{\beta}_j - \beta_j) / s.e.(\hat{\beta}_j)$

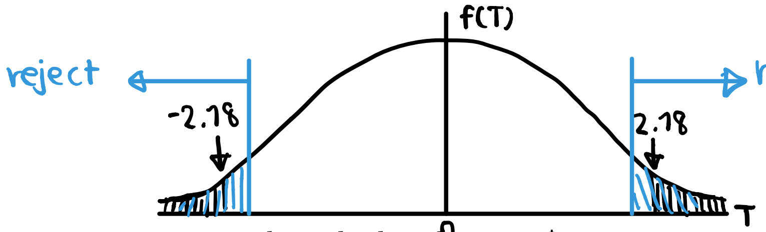
! From the z-table, the value -2.75 corresponds to area = _ _ _ _ _ _ _ _ _ _ _ _ _ _.

! Thus, p-value = 0.003.

! Would we reject $H_0$ if we use the significance level = 5%? Yes

* RULE! we reject $H_0$ if p-value < sig. level

Example 2: $H_0 : \beta_j = a_j$, $H_a : \beta_j \% a_j$, d.f.= 18. ⇒ use 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%?

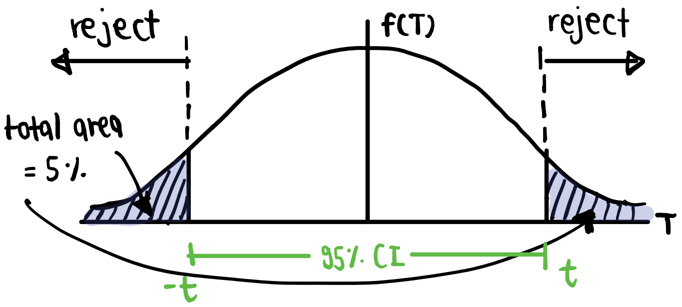
Yes, reject $H_0$ because the area is less than 0.05 or p-value < 0.05

## 6 Confidence Intervals (CI)

! Confidence Intervales for the POPULATION PARAMETER ($\beta_j$)

→ The range of values that would capture the true $\beta_j$ at a 95% chance.

! A 95% CI of $\beta_j$ is given by




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

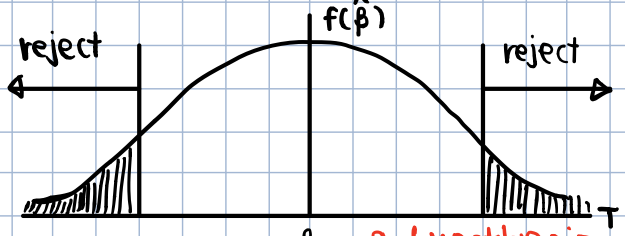
c is the 97.5 percentile in the t-distribution with n-k-1 d.f.

# Inference → Hypothesis testing about "$\beta$" the true parameter.

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

we want to test hypothesis about the true impact ($\beta$) of each x variable (educ, experience) on the dependent variable (Y)

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



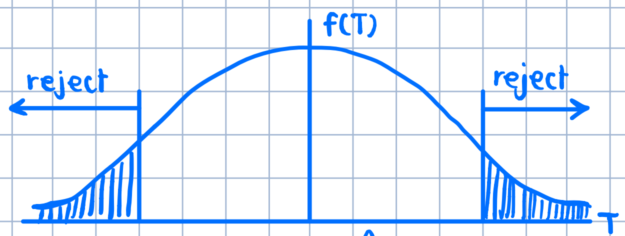

$\beta$ = a hypothesized value
ex. $\beta = 0$ or $\beta = 1$, etc.

1) Test if $\beta$ = 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.

↓




$0 = \dfrac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)}$

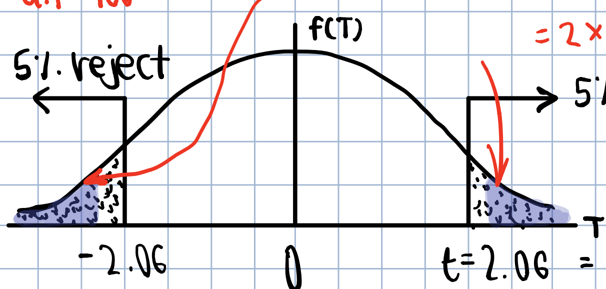
→ t-test ★ → How?

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

**Significant level = total area in the rejection region**

ass. d.f. 100

area = $2 \times (0.5 - 0.4803)$
$= 2 \times 0.197 = 0.0394$ = p-value!



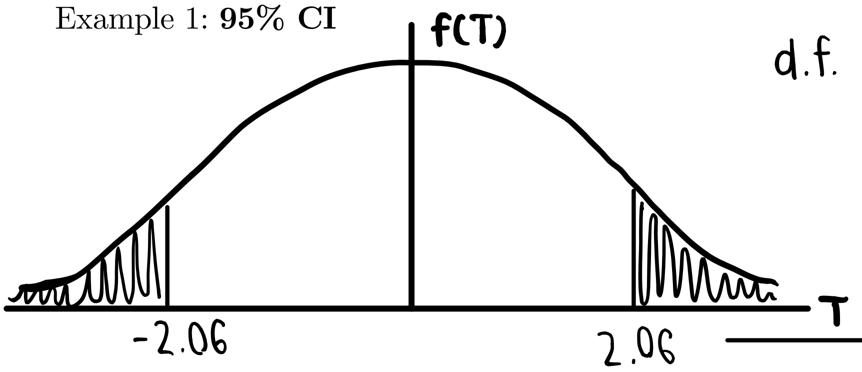

$t = 2.06 = \dfrac{(\hat{\beta}_j - \beta_j)}{s.e.\ \hat{\beta}_j}$

- Suppose, we calculate a t-statistic $= \dfrac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} = 2.06$
- Suppose, we are testing
  $H_0: \beta_j = 0$, $H_a: \beta_j \neq 0$
  ↳ 2 tailed test.
- P-value = total shaded are

P-value = significant level which we will reject the $H_0$
or prob. that we we will reject $H_0$

★ If p-value < significance level ⇒ reject $H_0$

Example 1: **95% CI**




d.f. = 25

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

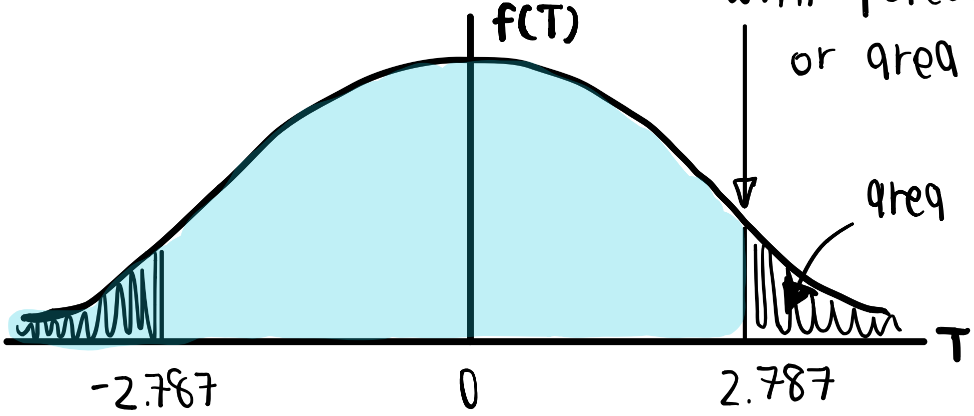
The 95% CI for $\hat{\beta}_j$ = $[\hat{\beta}_j - 2.06 = s.e.(\hat{\beta}_j),\ \hat{\beta}_j + 2.06 = s.e.(\hat{\beta}_j)]$

Example 2: **99% CI**

d.f. = 25

what percentile? 99.5

or area = 1 − 0.005 = 0.995




area = $\frac{1\%}{2} = \frac{0.01}{2} = 0.005$

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