Inference: t-test (test ~~whether~~ an parameter equals to a constant) usually zero.

(central limit theorem) CLT

MLR.6: $u_i \sim N(0, \sigma^2)$ independently of $x_{i1}, x_{i2}, \dots, x_{ik}$.

$$\Rightarrow y \mid x \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \sigma^2).$$

MLR.1 to MLR.6 is called classical linear model assumption. (CLM)

Theorem 4.1: Uneder assumptions MLR.1 – MLR.6

$$\hat{\beta}_j \sim N(\beta_j, Var(\hat{\beta}_j)), \quad \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1)$$

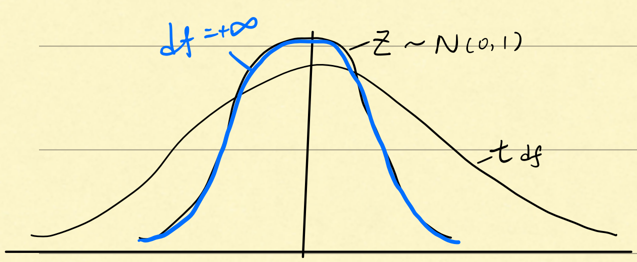
Theorem 4.2: $\dfrac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$

$t_{df} = \dfrac{X}{\sqrt{\frac{Y}{df}}}$

$X \sim N(0,1)$ $Y \sim \chi^2(df)$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \div \frac{se(\hat{\beta}_j)}{sd(\hat{\beta}_j)} \qquad \left(\frac{se(\hat{\beta}_j)}{sd(\hat{\beta}_j)}\right)^2 = \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{n} \frac{\hat{u}_i^2}{\sigma^2}}{n-k-1} \overset{\text{不证}}{\sim} \frac{\chi^2(n-k-1)}{n-k-1}$$

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1)$$



as $df \uparrow$, $t_{df} \to Z$

$E(t_{df}) = 0$, $df > 1$.

$Var(t_{df}) = \dfrac{df}{df - 2}$, $df > 2$

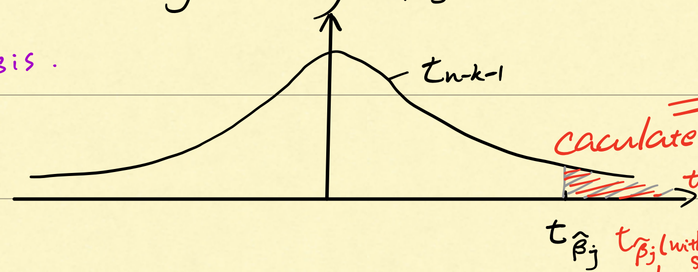
test **Null hypothesis**: $H_0: \beta_j = 0$.

$\Rightarrow$ t statistic / t ratio: $t_{\hat{\beta}_j} \equiv \dfrac{\hat{\beta}_j}{se(\hat{\beta}_j)}$

if $|t_{\hat{\beta}_j}|$ is large, we reject H_0.

set $H_1: \beta_j > 0$, (ignoring $\beta_j < 0$).

(one-sided)
alternative hypothesis.



p-value.
caculate the prob that another $t_{\hat{\beta}_j}$ (with other samplings) is bigger than the $t_{\hat{\beta}_j}$

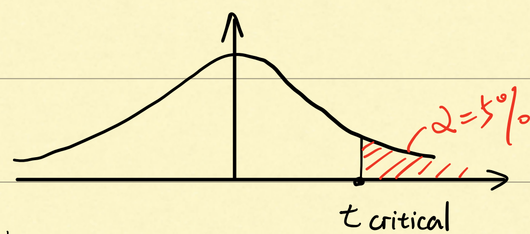
弃真
type I error: H_0 is true, but reject H_0

存伪
type II error: H_0 is false, but accept H_0

α = prob of type I error. (we can stand)

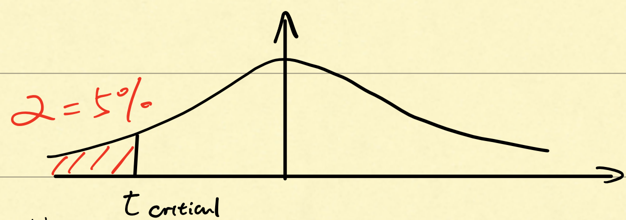
(given) = significance level size of the test.

$\alpha = 0.05$ (usually)



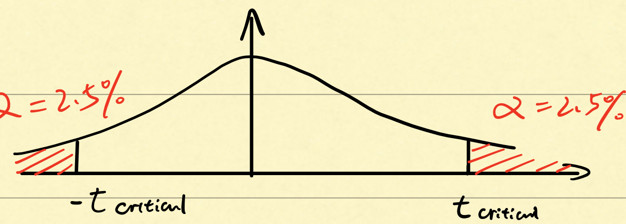
if $t_{\hat{\beta}_j} > t_{critical}$ reject H_0.

$t_{\hat{\beta}_j} \leq t_{critical}$ accept H_0.



if $t_{\hat{\beta}_j} < t_{critical}$ reject H_0.

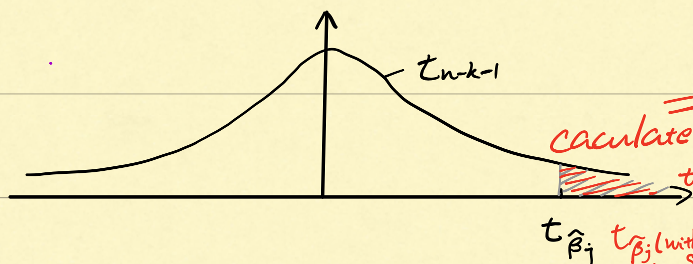
$t_{\hat{\beta}_j} \geq t_{critical}$ accept H_0.



if $|t_{\hat{\beta}_j}| > t_{critical}$ reject H_0.

$|t_{\hat{\beta}_j}| \leq t_{critical}$ accept H_0.

p-value



p-value.
caculate the prob that another $t_{\hat{\beta}_j}$ (with other samplings) is bigger than the $t_{\hat{\beta}_j}$

one-sided

$$\text{p-value} = P(T > t_{\hat{\beta}_j}) \text{ or } P(T < t_{\hat{\beta}_j})$$

twosided.

$$\text{p-value} = P(|T| > t_{\hat{\beta}_j})$$

if p-value $> \alpha$ accept H_0

p-value $< \alpha$ reject H_0

Statistical significance (depends on $t_{\hat{\beta}_j}$)

not necessarily mean economically (or practically)

Significant (depends on $\hat{\beta}_j$).

Confidence intervals (two-sided).

by theorem 4.2 : $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$

$$P\left(-c < \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} < c\right) = 1-\alpha$$

(c_α)

c is the $(1-\frac{\alpha}{2})^{th}$ percentile in a t_{n-k-1} distribution.

$$\hat{\beta}_j - c \cdot se(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + c \cdot se(\hat{\beta}_j)$$

$1-\alpha$ confidence intervals : $\beta_j \in [\hat{\beta}_j - c \cdot se(\hat{\beta}_j), \hat{\beta}_j + c \cdot se(\hat{\beta}_j)]$

Test hypotheses about a linear combination of the parameters.

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u.$$

$H_0 : \beta_1 = \beta_2$. $H_1 : \beta_1 < \beta_2$.

$\Rightarrow \beta_1 - \beta_2 = 0$

$$E(\hat{\beta}_1 - \hat{\beta}_2) = \beta_1 - \beta_2 ,$$

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

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

$$= \sqrt{[se(\hat{\beta}_1)]^2 + [se(\hat{\beta}_2)]^2 - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$

Method 1.

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)} \sim t_{n-k-1}$$

Method 2.

Let $\theta = \beta_1 - \beta_2 \Rightarrow H_0: \theta = 0,\ H_1: \theta < 0$

$$\beta_1 = \theta + \beta_2$$

$$\log(wage) = \beta_0 + (\theta + \beta_2)jc + \beta_2\, univ + \beta_3\, exper + u.$$

$$= \beta_0 + \theta\, jc + \beta_2(jc + univ) + \beta_3\, exper + u.$$

$$t = \frac{\hat{\theta}}{se(\hat{\theta})} \qquad se(\hat{\theta}) = se(\hat{\beta}_1 - \hat{\beta}_2)$$

R code: w=jc+univ , lm(y ~ jc + w +exper)

(alternatively) lm (y ~ jc + I(jc+univ) + exper)