Lecture 16

Econometrics-I

26.03.2021.

Inference

H_0 : Null Hypothesis ; H_a : Alternative Hypothesis

Truth (based on population character).

		H_0 is true	H_0 is false
Inference or Decision (Corresponds to a dataset/sample)	Reject H_0	WRONG/Error Type-I error	✓ CORRECT.
	Fail to reject H_0 (Accept H_0)	✓ Correct	WRONG/Error Type-II error

$$Pr(\text{type-I error}) = \alpha \rightarrow \text{significance level of the test.}$$

α = 0.05, 0.01, 0.10

Example

Test: $\underline{H_0 : \mu = 0}$ or $\mu_0 = 5$
constant ✓

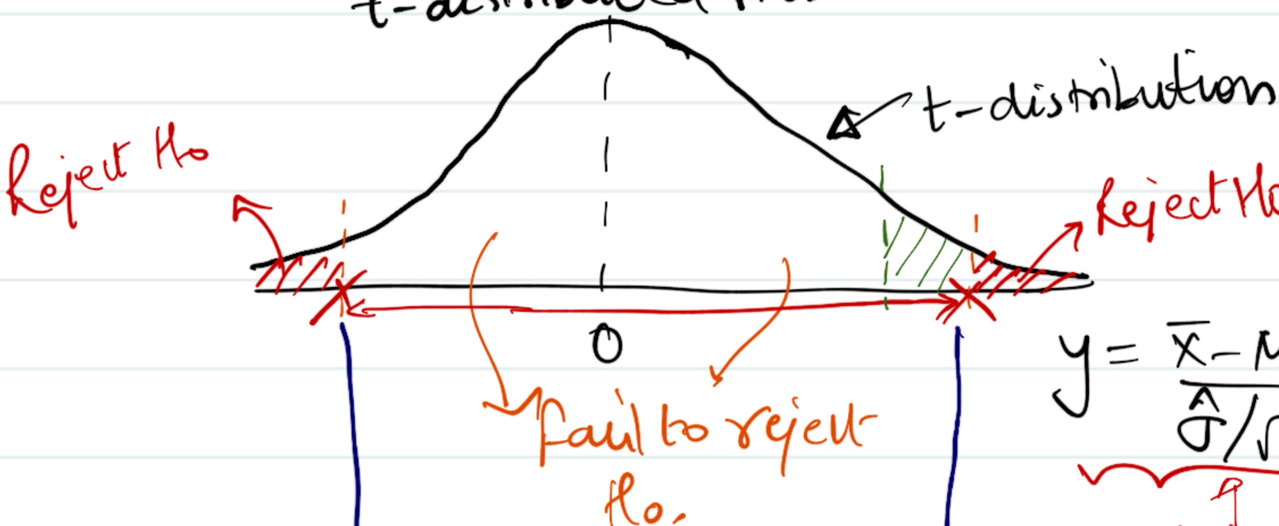
$H_a : \mu \neq 0$

t-distribution for conducting the above test:-

$N = 22$, $\bar{X} = 2.73$, $\hat{\sigma}^2 = 0.57$

we know,

$$y = \underbrace{\left(\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}\right)}_{\text{t-distributed r.v.}} \sim t_{N-1}, \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}$$



$$\underbrace{y = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}}_{\text{Under the null } H_0.} \quad \mu = 0 \;\; (=5)$$

critical values.

$-t_{N-1,\,0.025} = -2.080$ $t_{N-1,\,0.025} = 2.080$

Pr(shaded $\boxed{///}$ region) = 0.05
type I error

Inference: $\frac{\bar{X}}{\hat{\sigma}/\sqrt{N}} = t > 2.080$ then reject Ho

$\frac{\bar{X}}{\hat{\sigma}/\sqrt{N}} = t \le 2.080$ then fail to reject the Ho

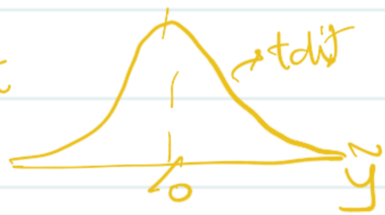
$$t = \frac{\overset{=\bar{X}}{2.73}}{\sqrt{\frac{0.57}{22}}} = \frac{2.73}{0.16} = 17.06$$

Reject the Null hypothesis Ho: $\mu = 0$

Decision/Inference → μ is statistically different from Zero.

Under the Ho: $\mu = 5$ $\quad \bar{X} \sim N\left(5, \frac{\sigma^2}{N}\right)$

$\tilde{y} = \frac{\bar{X} - 5}{\sqrt{\hat{\sigma}^2/N}} = \tilde{t} \sim t$



$\bar{X} - 5 \sim N(0, \sigma^2/N)$

Bring the idea of statistical inference to regression analysis:

Ⓐ Simple linear regression model :-

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad i = 1, 2, \ldots, N$$

$$\hat{\beta}_{1,OLS} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$$u_i \sim N(\times, \times)$$

$$y_i \sim N(\times\times, \times\times)$$ ($\because$ u_i and y_i are linearly related).

Given x_i's ; $\hat{\beta}_1 \sim N\left(E(\hat{\beta}_1), V(\hat{\beta}_1)\right)$,

$$\hat{\beta}_1 \sim N\left(E(\hat{\beta}_1), V(\hat{\beta}_1)\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right) \checkmark$$

Now, let's say we want to test

$H_0 : \beta_1 = 0$ | $H_0 : \beta_1 = \beta_0$ (const) $\approx$ 5.2

$H_a : \beta_1 \neq 0$ | $H_a : \beta_1 \neq \beta_0$ $\approx$ 5.2

$\gtrless$

example:

$$wage_i = \beta_0 + \beta_1 \, train_i + u_i$$

$H_0 : \beta = 0$; $H_a : \beta > 0$

If an econometrician fails to reject H_0 for our example, it has real consequences on allocation of budget.

$$\hat{\beta}_{1,OLS} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)$$

Under the Null $H_0 : \beta = \beta_0 = 0$

$$\hat{\beta}_1 \sim N\Bigg(0, \underbrace{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}_{=V(\hat{\beta}_1)}\Bigg)$$

then, $$\frac{\hat{\beta}_1 - \overset{(=\beta_0)}{0}}{\sqrt{V(\hat{\beta}_1)}} \sim N(0,1)$$

$$\Rightarrow \frac{\hat{\beta}_1}{\sqrt{\sigma^2 / \underbrace{\sum_{i=1}^{N}(x_i-\bar{x})^2}_{\checkmark}}} \sim N(0,1)$$

do not have from the data

Now, we consider:

$$\hat{u}_i \sim N(0, \sigma^2)$$

$$\frac{\hat{u}_i}{\sigma} \sim N(0,1)$$

$$\underbrace{\sum_{i=1}^{N} \left(\frac{\hat{u}_i}{\sigma}\right)^2}_{\text{sum of } N \text{ standar normals.}} \sim \chi^2_{N-2}$$

why N-2?

Bcoz:

$E(\hat{u}_i) = 0$

$E(\hat{u}_i x_i) = 0$

We have:

$$\frac{\hat{\beta}_1 - 0}{\sigma / \sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \sim N(0,1)$$

$$\sum_{i=1}^{N} \frac{\hat{u}_i^2}{\sigma^2} \sim \chi^2_{N-2}$$

Use the defn. of t-distribution:

$$\frac{\dfrac{\hat{\beta}_1 - 0}{\cancel{\sigma} / \sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}}{\sqrt{\underbrace{\sum_{i=1}^{N} \frac{\hat{u}_i^2}{\cancel{\sigma^2}} \Big/ N-2}_{=\hat{\sigma}^2}}} \sim t_{N-2}$$

Therefore, in order to test $H_0: \beta_1 = 0$; $H_a: \beta_1 \neq 0$

we use the following t-statistic:

$$\frac{\hat{\beta}_1 - 0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}} \sim t_{N-2}$$

* is known as the standard error of $\hat{\beta}_{1,OLS}$

* Notice the s.e.$(\hat{\beta}_{1,OLS})$ is different from s.d.$(\hat{\beta}_{1,OLS}) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$

Inference:

1. If $t = \hat{\beta}_1 / s.e(\hat{\beta}_1) > t^*_{N-2,\alpha/2}$ then reject H_0.
2. If $t = \hat{\beta}_1 / s.e.(\hat{\beta}_1) \leq t^*_{N-2,\alpha/2}$ then fail to reject H_0.

Inference (alt.)

① β_1 is statistically different from zero.

② β_1 is NOT statistically different from zero.

---

when $H_0: \beta_1 = 0$

$H_a: \beta_1 > 0$ } one-sided test.

then my inference changes slightly. How?

if $\boxed{t = \dfrac{\hat{\beta}_1 - 0}{\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{N}(x_i - \bar{x})^2}}} > t^*_{N-2,\alpha}$ then reject H_0.

if t —〃— $\leq$ —〃— (α d.f.) then fail to reject H_0.

Apply the idea of statistical inference for

(B) Multiple Linear Regression models

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + u_i$$

$i = 1, 2, \ldots, N$

Test: $H_0: \beta_2 = 0$

$H_a: \beta_2 \neq 0$ ($<$, $>$) } two-sided test.

$$t = \frac{\hat{\beta}_2 - 0}{s.e.(\hat{\beta}_2)} \sim t_{N-(K+1)}$$

Inference:

If $t > t^*_{N-K-1, \alpha/2}$ then reject H_0.

If $t \leq t^*_{N-K-1, \alpha/2}$ then fail to reject H_0.

in case of SLRM $K = 1$ and hence t had $N-2$ d.f.

Next lecture:

Advance to testing linear restriction for the MLRM

Example: $H_0: \beta_1 - 2\beta_3 = 6$

$H_a: \beta_1 - 2\beta_3 \neq 6$