Estimate Population Proportions

↳ we want to understand the mean of populations, we cannot survey entire population, we take sample, how can we use that information and estimate what the population mean is, the number should be pretty close to population, there's margin of error.

↳ อันนี้มี ๒ แบบ large sample $(n > 30)$, small sample $(n < 30)$

↳ normal distribution ↳ t-distribution

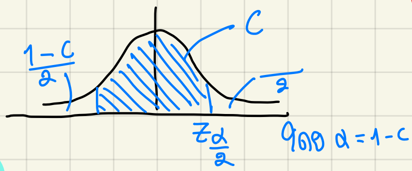
↳ propotion(p) : fraction of population

sample proportion($\hat{p}$) : prop for a sample of size n

margin of error (E) : $E = Z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

critical value

Z_c หาด้านเดียวของอะ อีกด้าน ก็ตรงข้ามตามคุณสมบัติ



confidence interval(CI) $\hat{p} - E < p < \hat{p} + E$

เหมือนของเดิมเป๊ะ

↳ **ex.** 13 out of 147 teachers know sign language. Find the margin of error for 90% CI for proportion of teacher who know sign language → เรารู้ $\hat{p} = \frac{13}{147}$

$E = Z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$= 1.645\sqrt{\frac{\frac{13}{147}(1-\frac{13}{147})}{\sqrt{147}}}$

$Z_c = Z_{\frac{1-c}{2}}$

$= Z_{0.05}$ ไปหา z ที่มีพื้นที่ 0.05

$= 1.645$

$= 0.038$ หรือ 3.8%

↳ ก็เลยได้ช่วงนี้ $5\% < p < 12.6\%$

↳ **ex** 190 adults are surveyed and 71 say that they eat out regulaly. Construct 95% CI for the proportion of adults who eat regularly.

$n = 190$, $\hat{p} = \frac{71}{190}$, $c = 0.95$, $\alpha = 0.05$

② $E = Z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

① $Z_c = Z_{0.025}$

$= 1.96$

$= 1.96\sqrt{\frac{0.37(1-0.37)}{190}}$

$= 0.069$

③ $\hat{p} + E = 0.37 + 0.069 = 0.439$

$\hat{p} - E = 0.37 - 0.069 = 0.301$

$0.301 < p < 0.439$

↳ ex you want to find a 98% confidence interval for the proportion of college students who get a student loan. You read an estimate that states that around 78% of students to get loan. you want your max error to be 5%. how many sample you need

$c = 0.98$, $\hat{p} = 0.78$, $E_{max} = 0.05$ หา n ?

$$E = Z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$E^2 = \frac{Z_c^2 \hat{p}(1-\hat{p})}{n}$$

$$\boxed{n = \left(\frac{Z_c}{E}\right)^2 (\hat{p})(1-\hat{p})}$$

$$Z_c = Z_{\frac{1-\alpha}{2}}$$

$$= Z_{\frac{0.02}{2}}$$

$$= Z_{0.01}$$

$$= 2.33$$

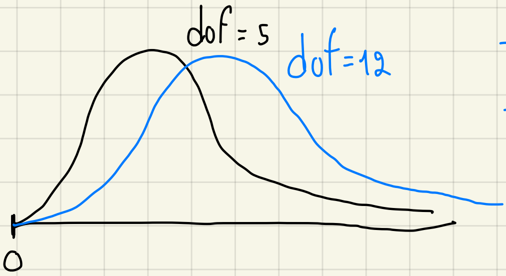
$$\rightarrow n = \left(\frac{2.33}{0.05}\right)^2 (0.78)(1-0.78)$$

$$n = \frac{0.932}{0.0025} = 372.6$$

$$= 373$$ ตัวอย่าง

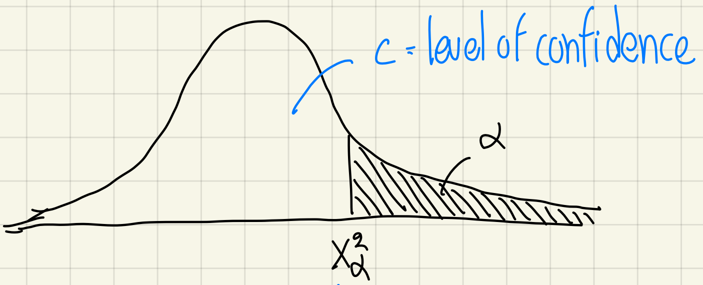
The Chi-Square Distribution: χ^2

- we use this distribution when we end up talking about confidence interval that involve variance and standard deviation
  - จะหาว่า population standard deviation or variance อยู่ในส่วนไหนของช่วงชั้น



- chi-square not symmetric about 0
- Asymmetrical
- ยิ่ง dof ↑ ยิ่งสมมาตร
  - แต่ไม่ใช่ normal เพราะตรงกลางไม่ใช่ 0
    - เลยไม่สามารถคำนวณ math กับมันได้แบบ normal and t

- การอ่านตารางก็เหมือน t เลย



หาค่าของ chi-square ที่ทำให้เกิดพื้นที่ α

$\alpha = 1 - C$; มีผลโดยตรงต่อกัน

"level of confidence"

"level of significant"

dof	$\chi^2_{0.990}$	$\chi^2_{0.975}$ ...	$\chi^2_{0.025}$	$\chi^2_{0.01}$
↓	the value of χ^2 from chart			

Estimating Population Variance

↳ ตอนก่อนๆ หน้านี้เราเอา E ที่ได้ไป ± กับ sample point estimate

↳ เราทำได้ (±) เพราะว่ามัน symmetrical แต่ไม่สามารถกระจายไป ± ได้ เราก็เลยไม่มี margin of error

↳ confidence interval - variance and standard deviation

↳ Procedure: → from the samples, find s^2 (variance)

→ from the confidence level, find $\alpha = 1 - C$

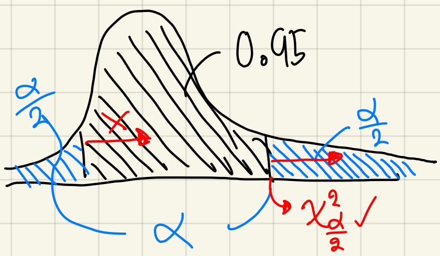
→ find $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$

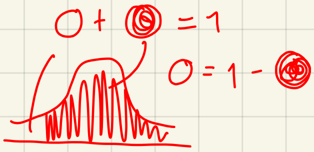
→ from the table (χ^2), find $\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{1-\frac{\alpha}{2}}$

→ CI: $\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})}}$

ไม่เข้าใจก็จำตามรูปนี้ไป

↳ ดูกฎ





chi-square ไม่ใช่ทรงระฆัง(คว่ำ)ขึ้น

แสดงว่าค่าตรงนี้คือ $\chi^2_{(1-\frac{\alpha}{2})}$

↳ ex a pencil factory is testing the variance in the mass of pencil produced. Sampling 15 pencils randomly show variance 3.4 g. What is 95% confidence interval for variance of all pencil produced?

↳ ① find s^2: $s^2 = 3.4$ $\quad df = 14$

② find α: $\alpha = 1 - 0.95$
$= 0.05$

③ $\frac{\alpha}{2} = 0.025$, $1 - \frac{\alpha}{2} = 0.975$

④ $\chi^2_{0.025} = 26.119$, $\chi^2_{0.975} = 5.629$

⑤ $\dfrac{(15-1)(3.4)}{26.119} < \sigma^2 < \dfrac{(15-1)(3.4)}{5.629}$

$1.82 < \sigma^2 < 8.46$

95% CI for variance

ex On a factory assembly line for softdrinks, 40 are sampled and the amount of liquid inside has standard deviation of 7.7 ml. Find 98% confidence interval for the standard deviation of liquid in the drinks produced by this factory.

① $s^2 = (7.7)^2$

$= 59.29$

② $\alpha = 1 - 0.98$

$= 0.02$

③ $\frac{\alpha}{2} = 0.01$, $1 - \frac{\alpha}{2} = 0.99$

④ $df = 39$

$\chi^2_{0.01} = 63.691$, $\chi^2_{0.99} = 22.164$

⑤ $\sqrt{\frac{(n-1)s^2}{\chi^2_{0.01}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{0.99}}}$

2 อย่างที่ส่งผล CI

$\sqrt{\frac{(40-1)(59.29)}{63.691}} < \sigma < \sqrt{\frac{(40-1)(59.29)}{22.164}}$

$6.0 < \sigma < 10.2$ ml

CI at 98% standard dev of liquid in cans

Introduction to Hypothesis Testing

↳ hypothesis : a premise or claim that we want to test

null hypothesis : denote H_0, currently accepted value for a parameter

alternative hypothesis : denote H_a, involves the claim to the test

เหมือนการโดนกล่าวหาใน USA คนโดนกล่าวหาถูก assume ว่าถูกอยู่เสมอ จนกว่าจะเจอหลักฐาน

มาหักล้าง null hypo

↳ เช่น รายได้เฉลี่ยของผู้คนในปีนี้สูงกว่าในอดีต → เราต้องทำการ test ก่อน

↳ ex It is believed that a candy machine makes chocolates bars that are on average 5g. A worker claims that the machine after maintenance no longer make 5g bars

↳ จาก $H_0 : \mu = 5g$ ← assume this is true

$H_a : \mu \neq 5g$ ← no longer make 5g anymore

↳ basically alternative hypo will be whatever the null hypothesis is not

↳ we will assume the null hypothesis is TRUE

↳ Possible outcome of this test :

↳ reject null hypothesis

↳ fail to reject null hypothesis

} การพิสูจน์ว่า this is false and this is not false

↳ มันยากมากที่จะพิสูจน์ความจริงร้อยเปอร์เซ็นต์

↳ แล้วจะ test ยังไง → ใช้ test statistic

↳ calculate from sample data and used to decide

↳ ค่าที่ได้จาก sample มาคำนวณทางสถิติ

↳ statistically significant : where do we draw the line to make decision?

↳ เหมือนหยิบตัวอย่างขนมมาชั่ง avg ใหม่ได้ 5.12 มันก็ยังใกล้ 5 อยู่นะ ถือว่าผิดไหม?

↳ ลองเก็บ sample ไปเรื่อยๆ อาจจะทำให้ reject ก็ได้ → ถ้าเสียสุดๆ avg ได้ 7g งี้

↳ แล้วแบบอะไรเป็นเกณฑ์ในการตัดสินใจ

statistics isnot about just your whim and how do you think it should be

we have to have a concrete way of looking at this null hypothesis, collecting data

to decide when we reject the null hypothesis and when we leave it there

↳ this is what hypothesis test does

hypothesis test $\xrightarrow{\text{output}}$ number → เอาไปใช้ในการตัดสินใจ

↳ collect data and put it in equation

↳ level of confidence (c)

↳ how confidence are we in our decision

fail to reject the null hypothesis

↳ don't forget we are doing an test and we are deciding to reject the null hypothesis

↳ c is how sure we are that we make a right thing

↳ level of significance (α)

↳ $\alpha = 1 - c$

Writing the Null and Alternative Hypothesis

↳ **ex.** a company has start their straw machine make that are 4 mm diameter. A worker believe the machine no longer makes straw of this size and sample 100 straw to perform a hypothesis test with 99% confidence.

$H_0: \mu = 4\text{mm}$

$H_a: \mu \neq 4\text{mm}$

ถ้าเป็นสั้นกว่าก็ทำได้และ

$H_0: \mu = 4\text{mm}$

$H_a: \mu < 4\text{mm}$

↳ **ex.** the school board claims that at least 60% of students bring a phone to school. A teacher believe this number is too high and randomly sample 25 students to test at a level of significance of 0.02.

↳ this is proportion, the same concept

$H_0: p \geq 0.6$

$H_a: p < 0.6$

challenge

the way we will do test

$n = 25$

$\alpha = 0.02$

Rejection Regions in Hypothesis Testing.

↳ how to draw a line that if the value jump over that we will reject it.

↳ $n < 30$: t , $n > 30$ normal , {σ^2, σ} chi-square → choose it to do test

↳ assume → we don't know σ of population
→ the population is normal distribution

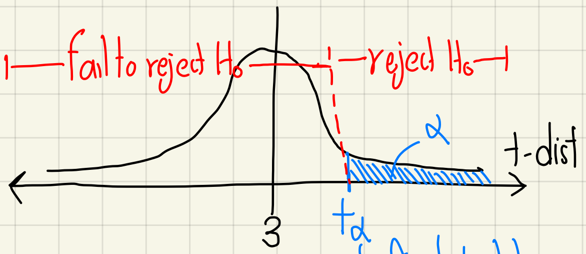
↳ hypothesis test : small sample ($n < 30$) use t-distribution
↳ how do we perfrom test
↳ look at Ha → > " right tail test "
→ < " left tail test "
→ ≠ " two tail test "

↳ right tail test

$H_0 : \mu \leq 3$ cm

$H_1 : \mu > 3$ cm → ได้ $\underbrace{3.1, 3.5, 3.2}_{\text{อันไหนก็อาจไม่ ok}}$, 5 → แต่เราสุ่มแบบ small sample นะ

↳ เราต้องหา threshould ถ้ามากกว่าเท่านี้ๆจะเชื่ออันใหม่เลยนะ



เช่นตัวอย่างของ right tail test

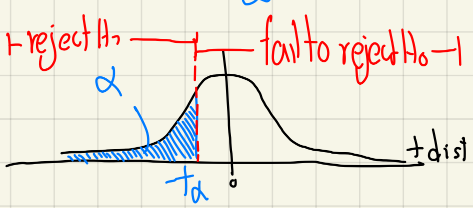
↳ ดูใน t table เลย

ถ้าได้ t ของ data ที่ sample ออกมาอยู่ในช่วงสีฟ้า trigger → reject null H_0

↳ left tail test

↳ $H_0 : \mu \geq 2g$
$H_a : \mu < 2g$ challenge

ผลิตภัณฑ์ guruntree มีน้ำหนักน้อยสุดเท่านั้นนะ → illegal

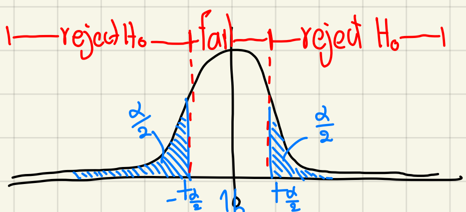


เหมือนเดิมไม่อยากให้ความรู้สึก → หา threshold

จากที่ดูยิ่ง C มากค่าใน α น้อยจะยกลงฝั่งนั้น ออก → ทำให้ detect ยากขึ้น

↳ two tails test

↳ $H_0 : \mu = 16$ inch
$H_a : \mu \neq 16$ inch

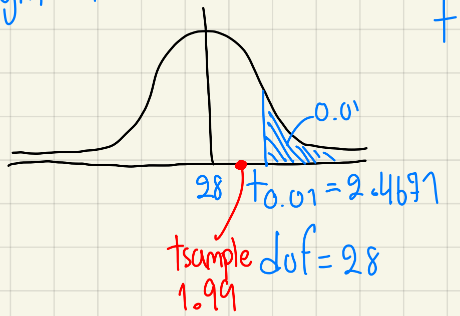


Hypothesis Testing for Means (Small Samples) $n<30 \Rightarrow$ t-dist

↳ ที่ต้องรู้ dof, α

↳ ex $H_0: \mu \le 28$ | $n=29$ | tsample $= 1.99$ → เราต้องไปหาเองในตาราง
$H_a: \mu > 28$ | $\alpha = 0.01$

↳ right tail test



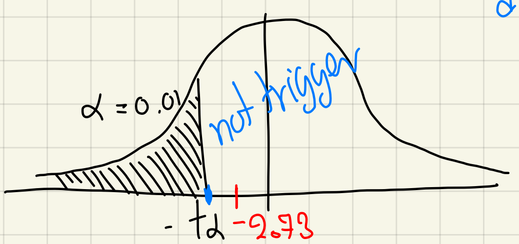
$t_{0.01} = 2.4671 > \text{tsample} = 1.99$

$\therefore$ fail to reject H_0 at this α

↳ ไม่เกิน threshold

↳ ex $H_0: \mu \ge 98$ | $n=10$ | tsample $= -2.73$ ๆๆ–ๆๆ
$H_a: \mu < 98$ | $\alpha = 0.01$

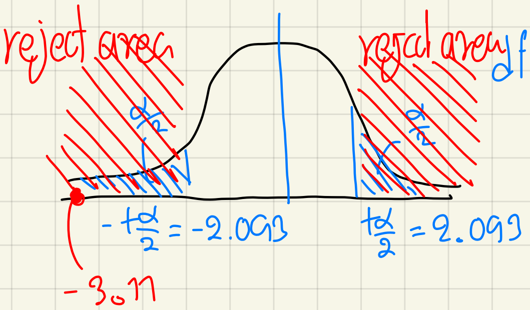
df = 9



$\therefore$ fail to reject H_0 at $\alpha = 0.01$

$-t_{0.01} = -(t_{0.01})$ at df 9
$= -(2.82)$
$= -2.82$

↳ ex $H_0: \mu = 195$ $n=20$ tsample $= -3.11$ ๆๆ–ๆๆ
$H_1: \mu \ne 195$ $\alpha = 0.05$



df = 19 และ $t_{\frac{\alpha}{2}} = 2.093$

$\therefore$ reject H_0 at α 0.05

↳ หาค่า t จาก (test statistics)

$$\boxed{t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}}$$

→ สมมติฐาน
→ standard deviation of sample

ตกอย่างของตัวอย่างที่ไม่มีค่าของ เพื่อให้น้องดูหมาเป็นเลขเดียว
↳ ทั้ง 6, n มีผลหมด

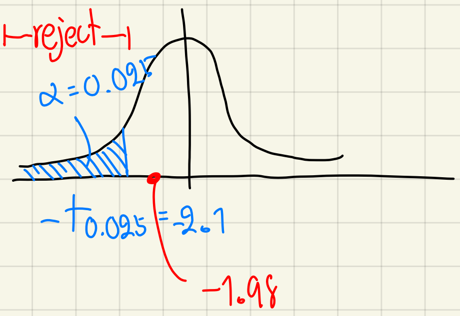
↳ ex doctor at hospital believe they see on average at least 8 patients per day. Management claim that they don't see this many consistently. they sample 19 doctors who report a sample mean of 7.5 patients seen per day with standard deviation of 1.1 patients. Test management's claim at 0.025 level of significance.

mean ✓
n<30 ✓
σ unknow ✓
pop normal ✓
use t dist

↳ $H_0 : \mu \geq 8$
$H_a : \mu < 8$

left tail test

$\bar{x} = 7.5$ $n = 19$ ใช้ t-dist
$s = 1.1$ $\alpha = 0.025$



$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{7.5 - 8}{\frac{1.1}{\sqrt{19}}}$$

จากสูตรไปผล

$$= \frac{-0.5}{\left(\frac{1.1}{4.36}\right)}$$

ลองฝึกเล่นๆดู

$$= \frac{-0.5}{0.252}$$

$$= -1.98$$

∴ fail to reject H_0 at this α

สังเกตผลลองตอนเดียวก็พูดได้ ไม่เป็นไรไม่เยอะมาก เพราะขึ้นกับ level of confidence

↳ ex a gracey store assume that the average shopper spends no more than \$100 in store. The new store manager belive that they spend more. He choose 27 shoppers randomly and averay is 104.93\$ with a standard deviation 9.07\$. Test the manager's clam at the 0.05 significance level

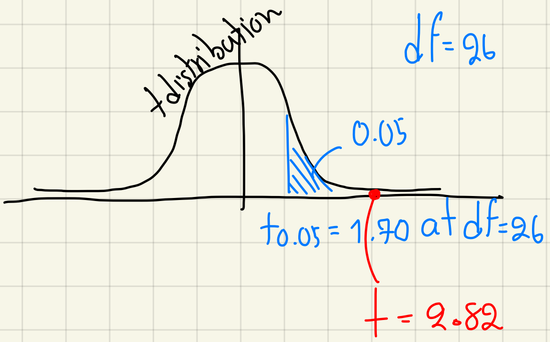
↳ $H_0 : \mu \leq 100$ $\alpha = 0.05$ $n = 27 \rightarrow$ t-dist
$H_1 : \mu > 100$ $\bar{x} = 104.93$ $s = 9.07$ → ขอแก้ตัวแปรหน่อยเพื่อดึงของจริงข้อมูลด้วย นอก outliner

df = 26

ผลต่างของ sample $t = \frac{104.93 - 100}{\left(\frac{9.07}{\sqrt{27}}\right)}$

n มาก t มาก
s มาก t น้อย

$t = 2.82$



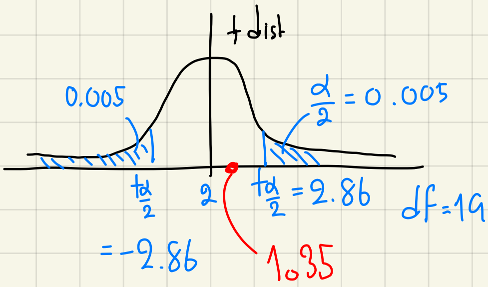
∴ reject H_0 at $\alpha = 0.05$

↳ นักสถิติเปลี่ยนผลข้อมูลได้เสมอ

↳ ถ้า α ไม่เพิ่มขึ้นและทำให้ t ลดลง → ก็ข้อมูลมีการกระจายเยอะ

↳ **ex** $H_0: \mu = 2$ $\quad n = 20$ $\quad S = 0.33$

$H_1: \mu \neq 2$ $\quad \bar{x} = 2.10$ $\quad \alpha = 0.01$ $\quad C = 0.99 \rightarrow$ อยากได้แบบแม่นยำมาก



$df = 19$

$$t = \frac{\bar{x} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}$$

$$= \frac{2.10 - 2}{\left(\frac{0.33}{\sqrt{20}}\right)}$$

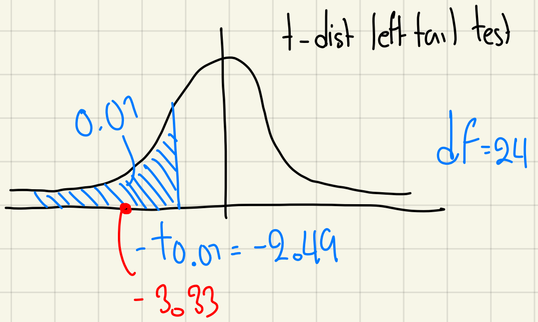
$$= 1.35$$

ปัญหาแบบเก่าแล้วยอมคนไปไม่ได้

$\therefore$ fail to reject H_0 at $\alpha = 0.01$

↳ **ex.** $H_0: \mu \geq 4$ $\quad \bar{x} = 3.4$ $\quad S = 0.9$

$H_a: \mu < 4$ $\quad n = 25$ $\quad C = 0.99$ $\quad \alpha = 0.01$



$df = 24$

$$t = \frac{\bar{x} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}$$

$$= \frac{3.4 - 4}{\left(\frac{0.9}{\sqrt{25}}\right)}$$

$$= -3.33$$

$\therefore$ reject H_0 at $\alpha = 0.01$

↳ **ex** a pizza chain advertises they will deliver your pizza in no more than 20 mins. You don't believe this and decide to test the claim at 95% level of confidence. You sample 7 friends who report an average of 22.7 minute delivery with standard deviation of 4.3 min. Test the claim.

↳ $n = 7$, $\bar{x} = 22.7$, $s = 4.3$

$H_0: \mu \leq 20$

$H_a: \mu > 20$ $\quad$ right tail test

$t_\alpha = t_{0.05, 6}$

≈ 1.94

$$t = \frac{22.7 - 20}{\left(\frac{4.3}{\sqrt{7}}\right)}$$

$$= 1.67$$

$t_{\text{ข้อมูล}} < t_\alpha$

$\therefore$ fail to reject H_0 at $\alpha = 0.05$

↳ ลองวิเคราะห์ว่าทำไมค่าที่คำนวณออกมาเป็น fail ทั้งๆที่ $\bar{x} > 20$

↳ (1) n น้อยมาก

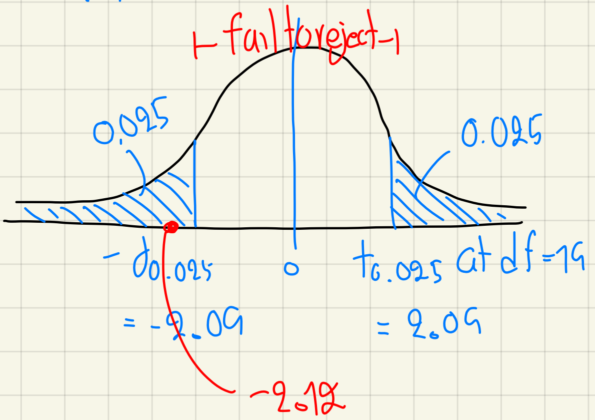
(2) s มีการกระจาย 18.4 —— 27 นาที

↳ ข้อมูลไม่เกาะกลุ่ม

↳ ex A computer company think it takes customers on average 15 min to set up their comp. A newly redigned model has been launched, and manager want to know if it takes user different from 15 min to set up. They sample 20 peoples who report an averge of 14.1 min with standard deviation of 1.9 min. Test the clam at 0.05 level of significance.

↳ $H_0 : \mu = 15$ min
$H_1 : \mu \neq 15$ min } 2 tail test

$n = 20$ $\bar{x} = 14.1$ $\alpha = 0.05$
$df = 19$ $s = 1.9$



$$t = \frac{14.1 - 15}{\frac{1.9}{\sqrt{20}}}$$

$= -2.12$ ⟶ ถ้าเป็น p-value พื้นที่น้อยกว่า)
ก็ reject
ค่าแบบ p-value
อาจจะง่ายกว่า
(อยู่ vol5)

∴ reject H_0 with $\alpha = 0.05$