Probability Density Function (PDF)

- ตัวแปรในการสุ่มจาก ~~single interval / disjoint interval.~~
- $P(x=c)=0$ (จุดใดๆ), total Area under graph $=1$

PDF w/ continuous variable

- $P(a \le x \le b) = \int_a^b f(x)\,dx$ (density function.)
  - $f(x) \ge 0$
  - $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- Uniform Distribution → ใช้ก็ต่อเมื่อ Prob ใน interval มีค่าเท่ากัน.

Ex: $f(x) = 0.15\left(e^{-0.15x-(-0.15)(0.5)}\right)$

$\rightarrow f(x) = 0.15e^{0.75} \cdot e^{-0.15x}$

; $\int_{0.5}^{\infty} f(x)\,dx = \int_{0.5}^{\infty} 0.15e^{0.75} \cdot e^{-0.15x}\,dx.$

; $0.15e^{0.75}\int_{0.5}^{\infty} e^{-0.15x}\,dx$

; $0.15e^{0.75}\int \frac{e^{u}}{-0.15}\,du$

let $u=-0.15x \rightarrow du = \frac{du}{dx}\cdot dx = -0.15\,dx$

$\therefore du = -0.15\,dx$

$\therefore \left[dx = \frac{du}{-0.15}\right]$

$\int e^{u}\,du = e^{u}+C$

; $0.15e^{0.75}\cdot\left(-\frac{1}{0.15}e^{u}\right)$ ($u=-0.15x$)

$\hookrightarrow 0.15e^{0.75}\cdot\left(-\frac{1}{0.15}e^{-0.15x}\right)\Big|_{0.5}^{\infty}$

; $\cancel{0.15}e^{0.75}\cdot \cancel{-}\frac{1}{\cancel{0.15}}\left[\cancel{e^{-0.15\cdot\infty}}^{\,0} \cancel{-} e^{-0.75}\right]$

; $\cancel{0.15}\cancel{e^{0.75}}\cdot\frac{1}{\cancel{0.15}}\cdot\frac{1}{\cancel{e^{0.75}}} = 1$ #

Confidence Interval (CI)

ช่วงของค่าประมาณที่ประกอบไปด้วย min, max ที่คำนวณขึ้นมา
ช่วงดังกล่าวจะครอบคลุม parameter และ prob ตามที่กำหนด.

($\bar{x}$)
median (ค่าเฉลี่ย) → $\frac{\Sigma x}{n}$ — ผลรวมของ Population, — จน. Population.

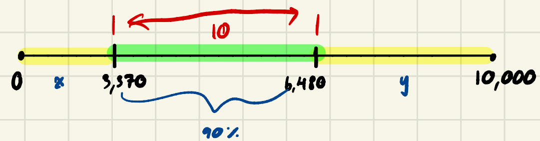
$$\boxed{\bar{x} - \frac{z \cdot (S.D.)}{\sqrt{n}} < u < \bar{x} + \frac{z \cdot (S.D.)}{\sqrt{n}}}$$

$\bar{x}$ - ค่าเฉลี่ย.
z - ค่าที่ได้จากตาราง
S.D. - ส่วนเบี่ยงเบนมาตรฐาน
n - จำนวนครั้งที่ทำการสุ่ม.
u - ค่าที่สุ่มมาและอยู่ในช่วงพารามิเตอร์.

Confidence level	z.
50%	0.68
75%	1.15
80%	1.28
90%	1.65
95%	1.96
97%	2.17
99%	3.29

Ex: Confidence level : 90% ของค่าใช้จ่ายโดยเฉลี่ย. อยู่ระหว่าง 3,370 - 6,480 baht.
↳ หมายความว่า: มีความมั่นใจที่ 90% ที่ช่วงประมาณค่าใช้จ่ายโดยเฉลี่ยที่ได้ (3,370 - 6,480 baht) จะครอบคลุมค่าใช้จ่ายที่เป็นค่าเฉลี่ยจริง.

สมมุติค่าใช้จ่ายในช่วง : 0 - 10,000 baht.



ช่วง: x+y = 10%.

$\bar{x}$
z
S.D.
$n = ?$

$$\left(\bar{x} + \frac{z \cdot (S.D.)}{\sqrt{n}}\right) - \left(\bar{x} - \frac{z \cdot (S.D.)}{\sqrt{n}}\right) = 10$$

$n = ?$

Hypothesis testing

สมมุติฐาน - ลักษณะของประชากร, สิ่งที่เกี่ยวกับ parameter. Ex: population mean.

└ 2 แบบ
- 1. สมมุติฐานหลัก (Null hypothesis) H_0
  - → ใช้ในการทดสอบ.
  - → มีเครื่องหมาย $=, \geq, \leq$
- 2. สมมุติฐานแย้ง (Alternative hypothesis) H_1
  - 2 กรณี
    - → One side Alternative Hypothesis. (เมื่อ Ho ใช้ $<, >, \leq, \geq$)
    - → Two side Alternative Hypothesis. (เมื่อ Ho ใช้ $=, \neq$)

Ex:

$H_0 : \mu \geq 20$	$H_0 : \mu = 20$
$H_1 : \mu < 20$	$H_1 : \mu \neq 20$

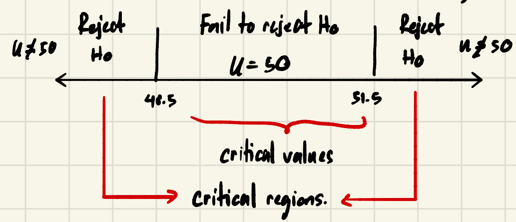
Possible hypothesis test outcome.

Decision	State of Nature	
	Ho true	Ho false
Accept Ho	No ERR $(1-\alpha)$	Type II ERR (β)
Reject Ho	Type I ERR (α)	No ERR $(1-\beta)$

Note! ERR type one จะมีความรุนแรง มากกว่า ERR type two (ส่วนใหญ่)

• test statistic → ค่าต่อยอดในการทดสอบ. (ถ้าเก็บหลายๆ รอบ จะเอา CI มาช่วย)

Ex:



Sensei didn't teach So don't worry.

• Probability Value (P-value) → เป็นการกำหนดค่าระดับนัยสำคัญ (α (type I), β (type II)) เพื่อที่จะ accept, reject Ho

↳ 2 กรณี.
- → ถ้า P-value $> \alpha$ ∴ accept Ho
- → ถ้า P-value $< \alpha$ ∴ reject Ho

Ex: บริษัทแห่งหนึ่งผลิตบุหรี่ มีสาร Nicotine ไม่เกิน 1.5 mg

[∴ Null hypothesis (H_0) = 1.5 mg   Alternate hypothesis (H_1) > 1.5 mg.]

given random sample = 32 cigarettes

H_0 — true ⟷ $E(\bar{x}) = 1.5$

H_0 — false ⟷ $E(\bar{x}) > 1.5$ → rejection region.

Ex: รถมีกันชน ไปได้รับ ค.เสียหาย 25%

บ. ออกแบบกันชนใหม่ → $\boxed{R_8 = \{8, 9, 10, \ldots, 20\}}$ → reject ho if $x \geq 8$ → rejection region

rejection Region

→ ∴ $H_0 : p = 0.25$

$H_1 : p > 0.25$

$n = 20$ (ครั้ง) no damage + no crash.

α = P(type one error) = P(H_0 is rejected, but H_0 is true)

↳ $P(x \geq 8) = 1 - \underbrace{P(x < 8)}_{\alpha}$ → $B(p; n, x) = \overset{0.112}{B(0.25; 20, 7)} + \overset{0.168}{B(0.25; 20, 6)} + \overset{0.202}{B(0.25; 20, 5)} + \overset{0.189}{B(0.25; 20, 4)}$

$+ \underset{0.133}{B(0.25; 20, 3)} + \underset{0.066}{B(0.25; 20, 2)} + B(0.25; 20, 1)$

↳ $1 - 0.87 = 0.13$

∴ 10% → H_0 is rejected, but H_0 is true

Preposition.

แปรผกผัน.

$\alpha \downarrow \ \beta \uparrow$

แปรผกผัน.

$p \uparrow \ \beta \downarrow$

Ex: Type II error จากข้อที่แล้ว.

→ H_0 is not rejected, but H_0 is false

∴ ให้ค่า p ต่าง เราก็จะได้ค่า β ต่าง

given $\beta(0.3)$ = P(type II error when $p = 0.3$)

$\beta = P(x \leq 7) \rightarrow B(p; n, x) = B(0.3; 20, 7) + B(0.3; 20, 6) + B(0.3; 20, 5) + B(0.3; 20, 4) + B(0.3; 20, 3) + B(0.3; 20, 2) + B(0.3; 20, 1)$

$\beta = 0.772 = 77\%$ → H_0 is not rejected, but H_0 is false

CH:5 Distribution of sample mean.

1. $E(\bar{x}) = u_{\bar{x}} = u$
2. $V(\bar{x}) = S.D._{\bar{x}}^{2} = \frac{S.D.^2}{n} \rightarrow S.D._{\bar{x}} = \frac{S.D.}{\sqrt{n}}$
3. standardize $z = \frac{x-u}{S.D.} \rightarrow z = \frac{\bar{x}-u_{\bar{x}}}{S.D._{\bar{x}}} \rightarrow z = \frac{\bar{x}-u_{\bar{x}}}{\frac{S.D.}{\sqrt{n}}}$
4. $S.D. = \sqrt{\frac{(x-\bar{x})^2}{n-1}} = \sqrt{\frac{n\Sigma x^2-(\Sigma x)^2}{n(n-1)}}$

Example of level α test.

ค่าเฉลี่ยบุหรี่ u. (สาร nicotine)

objective $\rightarrow$ $H_0: u = 1.5$, $H_1: u > 1.5$, upper

Sample(n) $\rightarrow x_1, x_2, x_3, \dots, x_{32}$ $S.D. = 0.20$

$S.D._{\bar{x}} = \frac{S.D.}{\sqrt{n}} = \frac{0.20}{\sqrt{32}} = 0.0354$

$\therefore z = \frac{\bar{x}-u}{\frac{S.D.}{\sqrt{n}}} = \frac{\bar{x}-1.5}{0.0354}$

$\rightarrow$ ค่า 'z' บอกระยะห่างของ $\bar{x}$, expected value (u) เมื่อ Ho $\rightarrow$ True

$1-0.05=0.95$ (ไม่มา)

c

$\alpha = 0.05$, type I

$z \geq c$

Ho rejected, when Ho = True.

$\alpha = P(\text{type I error})$

$= P(z \geq c \text{ when } z \sim N(0,1))$

$= 1 - P(z < c)$

$= 1 - 0.05 = 0.95$

เปิดตาราง $\Rightarrow c = \frac{1.64+1.65}{2} = 1.645$ #

rejection region.

α

$c = 1.645$

หา $\bar{x}$ ย้อน $z = \frac{\bar{x}-u_{\bar{x}}}{S.D._{\bar{x}}}$

$\therefore \bar{x} = z(S.D._{\bar{x}}) + u_{\bar{x}}$

$= (1.645)(0.0354) + 1.5$

$\bar{x} = 1.56$

α $\bar{x} \geq 1.56 \rightarrow$ rejection region, Ho is true

β $\bar{x} < 1.56 \rightarrow$ accept, Ho is false #

Test Procedure.

└ **3 cases!** → Normal Pop + รู้ S.D.

→ Large Sample test $(n > 40)$ } ไม่รู้ S.D.

→ Normal Pop. Dis $(n < 40)$

└ t-test distribution.

} z-test (standard normal)

3 case

→ $H_0 : u = u_0$, $H_1 : u \neq u_0$ — 2 sided test.

→ $H_0 : u = u_0$, $H_1 : u > u_0$ — upper tailed test

→ $H_0 : u = u_0$, $H_1 : u < u_0$ — lower tailed test

Case I:

- กำหนด Null hypothesis (Ho)
- กำหนด $u_0 \to H_0 : u = u_0$
- sample mean $\bar{x}$, ได้ expected value $u_{\bar{x}} = u$ $\left[S.D._{\bar{x}} = \frac{S.D.}{\sqrt{n}} \right]$
- เมื่อ Ho is true $\left[z = \frac{\bar{x} - u_0}{S.D./\sqrt{n}} \right]$

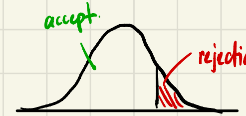
Ex: given. $H_0 : u = 100$

$S.D. = 10$, $n = 25$, $\bar{x} = 103$

$\therefore z = \frac{103 - 100}{10/5} = \frac{3}{2} = 1.5$ #

└ ค่าระยะห่างจาก $\bar{x}$, กับ expected value (u)

Ex: upper tailed test : $z \geq z_\alpha$



given $\alpha = 0.05$ (5%)

└ c (critical value) = 1.645

; $\alpha = P(\text{Type I Error})$

; $\alpha = P(\text{Ho is rejected, Ho is true})$

; $\alpha = P(z \geq c)$

; $[1 - \alpha = P(z < c)]$ → $\therefore P(z < c) = 1 - 0.05 = \underline{0.95}$ ↱ เปิดตาราง z_α (critical value)

$z_\alpha \approx \frac{1.64 + 1.65}{2} = 1.645$

Ex: lower tailed test : $z \leq z_\alpha$

given $\alpha = 0.10$ (10%)

└ critical value = −1.285

; $\alpha = P(\text{Type I Error})$

; $\alpha = P(\text{Ho is rejected, Ho is true})$

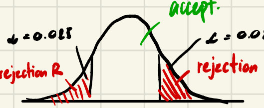
; $\alpha = P(z \leq c)$

; $[1 - \alpha = P(z > c)]$ → $\therefore P(z > c) = 1 - 0.10 = \underline{0.90}$ — table

$z_\alpha \approx \frac{-1.28 + (-1.29)}{2} = -1.285$

Ex: two tailed test: $z \leq z_{\alpha/2}$

$z \geq z_{\alpha/2}$



given $\alpha = 0.05$

; $\alpha = P(\text{Type I Error})$

; $\alpha = P(\text{Ho is rejected, Ho is true})$

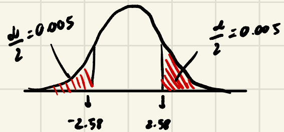
; $\alpha = P(z \leq -c) + P(z \geq c) = 2P(z \geq c)$ ← เพราะ 2 ฝั่ง ค.น่าจะเป็นเท่ากัน

ต่อ ; $2(1 - P(z < c)) = \alpha$

$\therefore \frac{\alpha}{2} = (1 - P(z < c))$

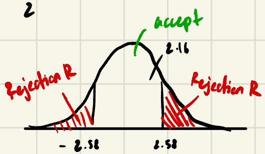
Ex: given $H_0: \mu = 130$, $H_1: \mu \neq 130$

$\bar{x} = 131.08$, $n = 9$, $S.D. = 1.5$, $\alpha = 0.01$



→ find $z = \dfrac{\bar{x}-\mu}{S.D./\sqrt{n}} = \dfrac{131.08-130}{1.5/3} = \dfrac{1.08}{0.5} = 2.16$

$\therefore \dfrac{\alpha}{2} = 0.005$ table $\therefore$ c(critical) = $2.58, -2.58$

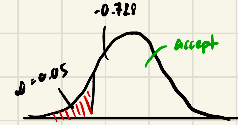


Ans: Ho จะถูกยอมรับ , เมื่อระดับความเชื่อมั่นเป็น 0.01

#

Case: II

- ไม่รู้ S.D.
- n > 40 ขึ้นไป.
- $z = \dfrac{\bar{x}-\mu_0}{S.P./\sqrt{n}}$



$C_{critical} = \dfrac{-1.64 + (-1.65)}{2} = -1.645$

Ex:

14.1	14.5	15.5	16.0	16.0	16.7	16.9	17.1	17.5	17.8
17.8	18.1	18.2	18.3	18.3	19.0	19.2	19.4	20.0	20.0
20.8	20.8	21.0	21.5	23.5	27.5	27.5	28.0	28.3	30.0
30.0	31.6	31.7	31.7	32.5	33.5	33.9	35.0	35.0	35.0
36.7	40.0	40.0	41.3	41.7	47.5	50.0	51.0	51.8	54.4
55.0	57.0								

→ find $\bar{x}$, S.D.

Calculated $\bar{x} = 28.7615$, $S.D. = 12.2147$, $n = 52$

Hypothesis $H_0: \mu = 30$, $H_1: \mu < 30$

$\therefore z = \dfrac{28.7615 - 30}{12.2147/\sqrt{52}} = -0.728$ given $\alpha = 0.05$

Ans: Ho จะถูกยอมรับ , เมื่อ $\alpha = 0.05$

#

Case III.

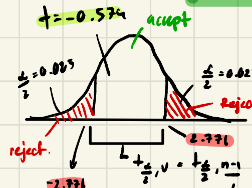
- ขนาดเล็ก n<40
- ไม่รู้ S.D.
- t-test value
- ค่า $v = n-1$ (ตาราง)

Ex: $S = \{2.67, 4.62, 4.14, 3.81, 3.85\}$ $\therefore n = 5$

$H_0: \mu = 4$, $H_1: \mu \neq 4$ $\alpha = 0.05$

Calculated mean = 3.814 , S.D. = 0.7185

$\left[T = \dfrac{\bar{x}-\mu}{S.D./\sqrt{n}} \right] = \dfrac{3.814 - 4}{0.7185/\sqrt{5}} = -0.579$



$t_{\frac{\alpha}{2}, v} = t_{\frac{\alpha}{2}, n-1} = t_{0.025, 5-1} = 2.776$

$n-1 = 5-1 = 4$

$\therefore$ Ho is accept when $\alpha = 0.05$

Formula : Procedure Test (แบบทั่วไป)

① Mean $= \dfrac{\Sigma\ \text{Sample}}{\text{no. Sample}}$

Case II, III

② S.D. , σ $= \sqrt{\dfrac{(x-\bar{x})^2}{n-1}} = \sqrt{\dfrac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$

case I

Case I, II   Case III

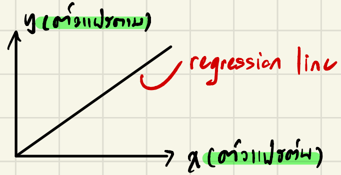
③ z-test, t-test $= \dfrac{\bar{x}-\mu_0}{S.D./\sqrt{n}}$

$S.D._{\bar{x}} = \dfrac{S.D}{\sqrt{n}}$

Linear Regression

- regression analysis → หาค่าแนวโน้มของตัวแปร — ตัวแปรตาม / ตัวแปรอิสระ
- linear correlation → correlation coefficient → สปส. สหสัมพันธ์
  - ↳ ค.สัมพันธ์ของตัวแปรทั้ง 2 เป็นอย่างไร?

Structure



Deterministic Model

Equation: $y = a + bx$

- slope: b
- x — independent v.
- a — y-intercept
- y — dependent v.

Simple regression → 2 variables
- independent — ต้น. — เหตุ — input
- dependent — ตาม. — ผล. — output.

Complete Equation (model)

$y = a + bx + \varepsilon$ → Random Error Term.

a, b — population parameters

Regression Line → เลือกเส้นที่ดีที่สุดในการอธิบายตัวแปร 2 ตัว โดยใช้ 'Least square method'

↳ Least Square Regression Line → Estimate regression model

random error term

$$\varepsilon = y - \hat{y}$$

y — real, $\hat{y}$ — estimate

SSE (Sum of Square Error)

$$SSE = \Sigma \varepsilon^2 = \Sigma (y-\hat{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

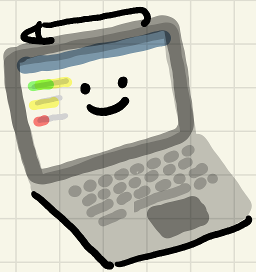
$$b = \frac{SS_{xy}}{SS_{xx}}$$

$$a = \bar{y} - b\bar{x}$$

$\bar{y}$ = mean of y

$\bar{x}$ = mean of x.

$$\hat{y} = a + bx$$



Ex:

Income (x)	Food Expenditure (y) (Real)
55	14
83	24
38	13
61	16
33	9
49	15
67	17

$n = 7$

random.

Find line of regression

$\rightarrow \Sigma x = 386$

$\Sigma y = 108$

$\Sigma xy = 6403$

$\Sigma x^2 = 23{,}058$

$S_{xy} = 6{,}403 - \dfrac{(386)(108)}{7} = 447.5714$

$S_{xx} = 23{,}058 - \dfrac{(386)^2}{7} = 1772.8571$

$$\boxed{S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}$$

$$\boxed{SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

$\bar{x} = \dfrac{386}{7} = 55.1429$, $\bar{y} = \dfrac{108}{7} = 15.4286$

$\therefore b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{447.5714}{1772.8571} = 0.2525$

$a = \bar{y} - b\bar{x} = 1.5050$

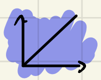
$\therefore$ Estimated Model $\rightarrow \hat{y} = 1.5050 + 0.2525x$ #

predict.

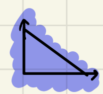
$\hat{y} = 1.5050 + (.2525)(61) = 1690.75$ \$

$\therefore \varepsilon =$ real $-$ predict. $= 1600 - 1690.75 = -90.75$ \$ #

Relationship : $b \rightarrow$ positive



, $b \rightarrow$ negative



Assumption of Regression Line

1. has mean equal to `0' for each x
2. error เป็นอิสระต่อกัน
3. normal distribution.
4. S.D. เป็นค่าคงที่ (σ_ε)

S.D. of Error

$$\boxed{S_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}}}$$

S_{xx} , S_{xy} , b ใช้สูตรเดิม.

$$\boxed{S_{yy} = \sum_{i=1}^{n} y^2 - \frac{\left(\sum_{i=1}^{n} y\right)^2}{n}}$$

Coefficient of determination

$\rightarrow$ หาว่า Regression model ดี, ไม่ดี ?

sum of square total (SST) $\rightarrow$ $\boxed{SST = \Sigma y^2 - \dfrac{(\Sigma y)^2}{n}}$

sum of square regression (SSR) $\rightarrow$ SST $-$ SSE

$$\boxed{SSE = \Sigma \varepsilon^2 = \Sigma (y - \hat{y})^2}$$

Note! SST $\geq$ SSE

จาก SST $\rightarrow$ SSE

Error ลดลงน้อย

จาก SST $\rightarrow$ SSE

Error ลดลงเยอะมาก

Note! $0 \leq r^2 \leq 1$

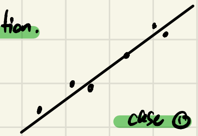
$$\boxed{r^2 = \frac{bSS_{xy}}{SS_{yy}}}$$

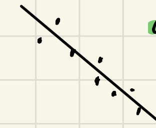
Example population.

$$\boxed{r^2 = \rho^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}}$$

Linear Correlation. (สหสัมพันธ์เชิงเส้น)

- Correlation Coefficient. → ดูความ (scatter point) กระจายไกล / ใกล้กับ RL. แค่ไหน?
  - ค่าอยู่ระหว่าง -1 , 1 → $-1 \le \rho \le 1$ (population) , $-1 \le r \le 1$ (example.)
    - 3 cases ① $r = 1$ → Perfect positive
      - ใกล้ 1 มาก = strong
      - ไกล 1 ออกมา = weak
    - ② $r = -1$ → Perfect negative
      - ใกล้ -1 มาก = strong
      - ไกล -1 ออกมา = weak
    - ③ $r = 0$ → No correlation.





Formula:

$$\boxed{r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}}$$

- Hypothesis testing about Linear Correlation Coefficient → ใช้ ρ ประมาณ.
  - t- distribution. $n < 40$. , $H_0: \rho = 0$
  - formula:

$$\boxed{t = r\sqrt{\frac{n-2}{1-r^2}}}$$

Ex: 1% level of significance

linear correlation is positive → $\therefore \rho > 0$, upper tailed test.

$n = 7$, $r = 0.95$

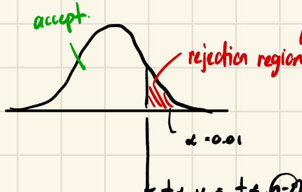
→ $H_0: \rho = \rho_0$ → $H_0: \rho = 0$

$H_1: \rho > 0$

find $t = r\sqrt{\frac{n-2}{1-r^2}}$

$t = 0.95\sqrt{\frac{7-2}{1-(0.95)^2}}$

$\therefore t = 6.803$



→ $n-2 = 7-2 = 5$ เปิดตาราง $= 3.365$

Ans: reject H_0