Hypothesis Testing Two Means Large Independent Samples

↳ all the hypothesis testing we've done in the past has always been involving either one mean or one standard deviation of a population.

↳ some time you really have 2 groups of people and you want to compare and you want to see, for instance, if they're how they're (population) related to each other statistically

↳ 2 ตัวเลือกที่ต้องรู้ความแตกต่าง

↳ independent sample : two sets of sample data that are not connected or related at all.

dependent sample : study the same people but the study is done pre and post test
หรือ pre or post event

↳ Hypothesis Testing - 2 Means

↳ Large ⇒ $n > 30$

Independent - 2 Group are not connect

2 means

→ หลักการเหมือนเดิม

↳ ใช้ p-value หรือ normal, reject region ตั้งสมมุติฐาน

↳ ex μ_1 คือ ความสูงของผู้หญิงทุกคนในเมือง

μ_2 คือ ความสูงของผู้ชายทุกคนในเมือง

เป็นอะไรก็ได้ตามโจทย์ต้องการ

โดย $H_0: \mu_1 \leq \mu_2$ — mean$_1$ is less than or equal to mean$_2$

$H_a: \mu_1 > \mu_2$ — mean$_1$ is greater than mean$_2$

ปกติเขาจะเขียนอยู่ในรูปแบบของ

$H_0: \mu_1 - \mu_2 \leq 0$

$H_a: \mu_1 - \mu_2 > 0$

right tail test

แรกๆ การเขียนแบบนี้อาจจะงงๆ เดี๋ยวก็ชิน ^^

ถ้ากรณี left tail test

$\mu_1 - \mu_2 \geq 0$

$\mu_1 - \mu_2 < 0$

ถ้ากรณี two tail test

$\mu_1 - \mu_2 = 0$

$\mu_1 - \mu_2 \neq 0$

↳ test statistic → ดูว่าเส้นไปตกที่ไหน → reject or fail to reject → everything depend on test statistic

↳ $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ลองแทนสูตร จะมีค่าออกมาตัวนึง กลับไปดูกับของเดิม

↳ แล้วก็ไปดูว่าจะตกใน rejection region หรือเปล่า

↳ ex A researcher thinks that grocery shoppers spend more when they haven't eaten. To test, he sample 41 shoppers who didn't eat breakfast. These people spent on average 72.27\$ with standard deviation of 8.05\$. 52 shoppers who did eat breakfast spend on average 69.43\$ with standard deviation 9.22\$. Test with 95% level of confidence

↳ 2 mean, large sample → normal dis → use p-value ✓

↳ let μ_1 → people didnot eat breakfast

μ_2 → people did eat breakfast

① โจทย์ให้อะไรมาบ้าง

$\bar{X}_1 = 72.27$	$\bar{X}_2 = 69.43$	$\alpha = 0.05$
$n_1 = 41$	$n_2 = 52$	
$S_1 = 8.05$	$S_2 = 9.22$	

② เขียน hypothesis

challenge hypo

$\mu_1 > \mu_2$; อยากพิสูจน์อะไรก่อนเลย → เป็น alternative hypothesis → "researcher thinks..."

$H_a: \mu_1 - \mu_2 > 0$

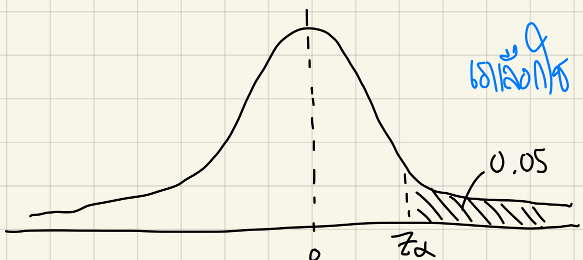
เมื่อได้ H_a ก็ไปหา H_0 ได้เพราะจะเป็นตรงข้ามกันเสมอ

$H_0: \mu_1 - \mu_2 \leq 0$; $\mu_1 \leq \mu_2$

$H_a: \mu_1 - \mu_2 > 0$; $\mu_1 > \mu_2$

→ มองแบบเก่า

↳ right tail test



เราเลือกใช้ p-value

↳ ง่ายดี ถถถ

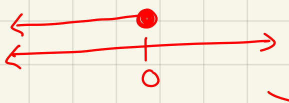
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{(72.27 - 69.43) - (\mu_1 - \mu_2)}{\sqrt{\frac{(8.05)^2}{41} + \frac{(9.22)^2}{52}}}$$

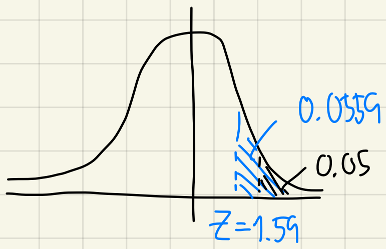
$H_0: \underline{\mu_1 - \mu_2} \le 0$ เป็นที่ๆเดียวที่เราเจอ μ_n how the fuck we can get it from?

เหมือนกับตรง

ย้อนกลับไปที่ null hypothesis



→ ค่ามันเป็นอะไรก็ได้ เราเลยเลือก 0 $(\mu_1 - \mu_2) = 0$



$z = 1.59$

p-value $= P(Z < -1.59)$

$= 0.0559 > \alpha = 0.05$

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

↳ ex a new cholesterol drug claims to lower cholesterol by over 20 pts. Group A of 55 ppl exercise and take drug and lower cholesterol by avg of 44.7 pts with standard deviation of 6.8 pts. Group B is 55 different people who exercise but don't take the drug and lowwer cholesterol by avg 23.1 pts with standard deviation 5.3 pts. Test at 0.01 los

↳ Group A (μ_A)
$n_A = 55$
$\bar{x}_A = 44.7$ pts
$S_A = 6.8$ pts

$\alpha = 0.01$

Group B (μ_B)
$n_B = 55$
$\bar{x}_B = 23.1$
$S_B = 5.3$

$n > 30 \rightarrow$ normal dis

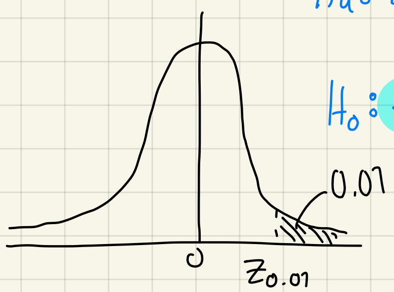
ไอที่อยากพิสูจน์ของการ claim...

hypothesis ที่เค้าโฆษณาคือ $\mu_A > \mu_B + 20 \rightarrow$ นี่คือ alternative hypotesis

$H_a: \mu_A - \mu_B > 20$; เขียนไปเลยแบบนี้เลย → right tail test

ตรงข้าม

$H_0: \mu_A - \mu_B \le 20$ null hypo



จะหาจากไหนหล่ะเนี่ย = ?

null hypo

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

$$= \frac{(44.7 - 23.1) - (20)}{\sqrt{\frac{(6.8)^2}{55} + \frac{(5.3)^2}{55}}}$$

$= 1.38 \rightarrow$ หาขั้นต่อ

p-value $= P(Z < -1.38)$

$= 0.0838 > \alpha = 0.01$

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

ex It is always accepted that people from 2 near by cities exercised the same amount. A researcher proposes that the 2 cities don't execise the same. City A asks 36 people who answer that they exercise on avg 2.4 hrs/wk with standard deviation 1.1 hrs. City B asks 38 people who answer they exercise 2.7 hrs/week with standard deviation 1.0 hrs. Test the claim at 0.05 level of significance.

City A (μ_1)	$\alpha = 0.05$	City B (μ_2)
$n_1 = 36$		$n_2 = 38$
$\bar{x}_1 = 2.4$		$\bar{x}_2 = 2.7$
$S_1 = 1.1$		$S_2 = 1.0$

$\mu_1 = \mu_2$; จากบรรทัดแรกเป็น null hypo → it is alway accepted

$H_0 : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 \neq 0$; don't exercise the same

miror ทั้ง 2 ข้าง

two tail test

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{(2.4 - 2.7) - (0)}{\sqrt{\frac{1.1^2}{36} + \frac{1^2}{38}}}$$

$$= 0.82$$

$p\text{-value} = P(Z < -0.82) \times 2$

$= (0.2061) \times 2$

$= 0.4122 > \alpha = 0.05$

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

Hypothesis Testing Two Means Small Independent Samples

- different → we use t-dist
- Pooled problem → assume: variance of 2 populations are assumed equal ($\sigma_1^2 = \sigma_2^2$)
  - test statistic → $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$; small sample t-dist

    $$dof = n_1 + n_2 - 2$$
- not pooled problem → assume: variance of 2 populations are assumed not equal
- how do we know what type of problem we're dealing with
  - depending on your problem you can make an educated guess and figure out what kind of class of problem you might be in.
  - test statistic → $$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

    $dof = n_1 - 1$ or $n_2 - 1$ which is smaller
- จุดที่เริ่มเลยคือ rejection region → small sample size → not normal dis
- ex Its claimed that people who go to home improvement classed finish project in less time. Group A is 10 people who attend class. On average, they finish projects in 14.1 hr with standard deviation of 2.3 hr. Group B is 10 people who don't attend the class and finish project on average 15.0 hrs with standard deviation 2.4 hrs. Test the claim at 0.01 level of significance.

① decide pooled or not pooled?

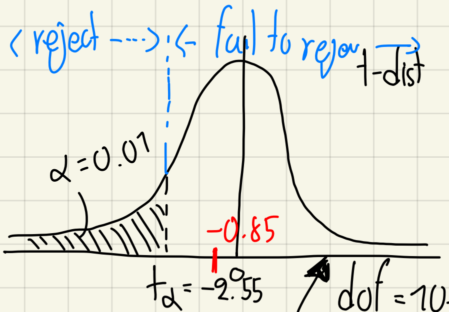
- pooled: เพราะว่ากลุ่มคนที่เราเลือกมาสำรวจอยู่ในประชากรบริเวณเดียวกัน เช่น โรงเรียน, ห้อง, ตำบล

data A (μ_1) took class		data B (μ_2) not took class
$n_1 = 10$	$\alpha = 0.01$	$n_2 = 10$
$\bar{x}_1 = 14.1$ hrs	99% LOC	$\bar{x}_2 = 15.0$ hrs
$s_1 = 2.3$ hrs		$s_2 = 2.4$ hrs

"claimed..." $\mu_1 < \mu_2$ → left tail test

$H_a : \mu_1 - \mu_2 < 0$

$H_0 : \mu_1 - \mu_2 \geq 0$ คือ null hypo



dof = 10+10−2 = 18

$t_{0.01,18} = -(2.55) = -2.55$

มองสะท้อนไปทางขวา

ค่า ⊕ จะรู้ว่าอยู่ขวา

t => area to the right

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(14.1 - 15.0) - (0)}{\sqrt{\frac{9(2.3)^2 + 9(2.4)^2}{18}}\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$= -0.85$$

∴ fail to reject H_0 with $\alpha = 0.01$

↳ **ex** A test company claim that their class increase test score. In City A, 15 students took the class and got an average score of 942 points with standard deviation 103 points. In City B 18 students did not take the class and got average score of 898 with standard deviation 95. Test the claim at 0.95 level of significance. Because students are in different cities, assume variances are not equal → not pooled relation

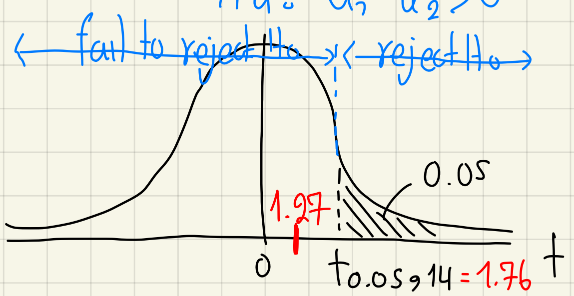
City A (μ_1) take class		City B (μ_2) don't take class
$n_1 = 15$	$c = 0.95$	$n_2 = 18$
$\bar{x}_1 = 942$	$\alpha = 0.05$	$\bar{x}_2 = 898$
$s_1 = 103$		$s_2 = 95$

small sample $n < 30$

claim that $\mu_1 > \mu_2$

$H_0 : \mu_1 - \mu_2 \leq 0$ ← ตรงข้ามเสมอ

$H_a : \mu_1 - \mu_2 > 0$



dof = min(n_1−1, n_2−1) = 14

$t_{0.05,14} = 1.76$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(942 - 898) - (0)}{\sqrt{\frac{103^2}{15} + \frac{95^2}{18}}}$$

$$= 1.27$$

∴ fail to reject H_0 with $\alpha = 0.05$

↳ ในที่นี้ H_0 เป็นที่ยอมรับ → evidence is not strong enough at this level of confidence

↳ ก็ว่าคะแนนเฉลี่ยดูเหมือนจะสูงกว่า แต่ก็ยังไม่เพียงพอ

Hypothesis Testing Two Proportions Large <u>Independent</u> Samples

↳ large sample

↳ ไม่เหมือนใน population mean ที่เอา $n \geq 30$

ใช้ normal distribution ได้

↳ constraint : $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$ → ค่าไม่ต้องเยอะมากก็ได้ → ยืนยันว่า large enough

↳ จะแยกปรับใช้ $n_1\hat{p}_1 \geq 5$ and $n_1(1-\hat{p}_1) \geq 5$ | $n_2\hat{p}_2 \geq 5$ and $n_2(1-\hat{p}_2) \geq 5$

↳ test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

average of something

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

number of sample

; x เป็นคนที่ตอบตามที่เราสนใจ

↳ ใช้ rejection region และ p-value ก็ได้ เพราะเป็น normal distribution

↳ **ex** A reporter think that the president's approval rating has improved following an article he published. Before the article, 480 of 1200 citizens approved of the president. After, 550 of 1180 citizens approved of his work. Test the claim at 5% level of significance.

↳ ดูข้อกำหนดก่อน $n_1\hat{p}_1 \geq 5$ and $n_1(1-\hat{p}_1) \geq 5$

$n_2\hat{p}_2 \geq 5$ and $n_2(1-\hat{p}_2) \geq 5$ ✓ จบ(เช็คแล้ว)

Before	After	
$x_1 = 480$	$x_2 = 550$	$\alpha = 0.05$
$n_1 = 1200$	$n_2 = 1180$	

$\hat{p}_1 = \frac{x_1}{n_1} = \frac{480}{1200} = 0.4$

$\hat{p}_2 = \frac{556}{1180} = 0.47$

claim $p_1 < p_2$ จากโจทย์

$H_0 : p_1 - p_2 \geq 0$

$H_a : p_1 - p_2 < 0$ ↳ left tail test

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

เหมือนเดิม

ตามของมัน

$$= \frac{(0.4 - 0.47) - (0)}{\sqrt{0.43(1-0.43)\left[\frac{1}{1200} + \frac{1}{1180}\right]}}$$

$= -3.49$

เฉลี่ย average proportion

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{480 + 556}{1200 + 1180} = 0.43$$

p-value $= P(Z < -3.49)$

$= 0.0002 < \alpha = 0.05$

∴ reject H_0 with $\alpha = 0.05$

↳ แปลง alternative hypo → ตามข้อมูลที่บอก article

↳ ex a dealer claims that more cars are purchased by single women than single men. When studying single women, 100 out of 500 purchased were made. In men's study 72 out of 500 single men bought cars. Test the claim at 0.01 level of significance

Women → 1	$\alpha = 0.01$	Men → 2
$x_1 = 100$		$x_2 = 72$
$n_1 = 500$		$n_2 = 500$

large แน่นอน จบเลยแล้ว

claim that $p_1 > p_2$

$H_0 : P_1 - P_2 \le 0$

$H_a : P_1 - P_2 > 0$; right tail test

สัดส่วนประชากร $\hat{p}_1 = \frac{100}{500} = 0.2$

$\hat{p}_2 = \frac{72}{500} = 0.144$

$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{100+72}{500+500} = 0.172$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$= \frac{(0.2 - 0.144) - (0)}{\sqrt{0.172(1-0.172)\left[\frac{1}{500} + \frac{1}{500}\right]}}$$

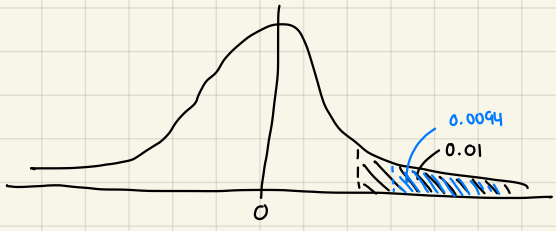
$$= 2.35$$

p-value $= P(Z < -2.35)$

$= 0.0094 < \alpha = 0.01$

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

ต้องลองวาดรูปทีหลัง



พื้นที่น้อยกว่าต้องอยู่ใน rejection regeion

$\therefore$ the claim that single buy more car is true at this level of significance

Hypothesis Testing Two Means Dependent Sample

↳ dependent sample : study the same person → pre and post test → "paired" (paired data)
↳ assume: the data is drawn from the normal population

one person
↳ เมื่อเจอ paired data → what important: different between before and after (before, after)

↳ d = after - before ; ทุกๆ paired data

↳ $\bar{d}$; ผลต่างเฉลี่ยคืออะไร → the mean of paired different of sample

$$\bar{d} = \frac{\Sigma d}{n}$$

↳ μ_d ; ของจริง → mean of paire different of population

↳ test statistic → $t = \dfrac{\bar{d} - \mu_d}{\left(\frac{S_d}{\sqrt{n}}\right)}$ → เป็น t-dist ใช้ rejection region แน่นอน

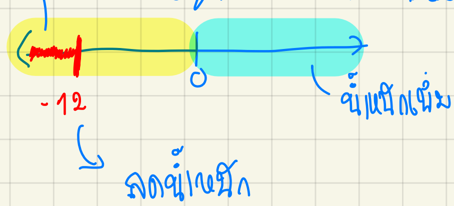
S_d ก็เหมือนด้านบน คิดเองได้แล้ว

$dof = n-1$

↳ ex A diet pill claim to make a person lose more than 12 lbs in the first month. To test, 15 people take the pill and are weighted at the beginning and end of month. State H_0 and H_a.

population 1 = Before

population 2 = After



claim $\mu_d < -12$; ลดมากกว่า 12 ใน 1 เดือนแรก จากโจทย์

H_0 $\mu_d \geq 12$

H_a : $\mu_d < -12$

4) ex A memory class claim that after taking class you will lose key less often. To test, 12 people are interviewed before and after class. Test the claim at 0.05 level of significance

data

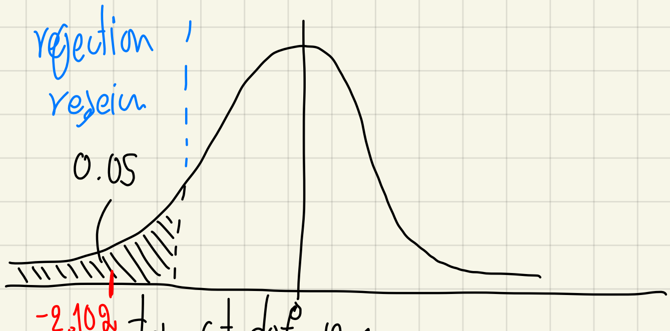
	1	2	3	4	5	6	7	8	9	10	11	12
Before	8	10	6	7	4	11	12	5	6	3	6	4
after	6	5	6	6	5	9	4	5	4	4	5	4
d	-2	-5	0	-1	1	-2	-8	1	-2	1	-1	0

$\bar{d} = \frac{\Sigma d}{n} = -1.583$; ที่ level of confidence เลยมีนัยสำคัญพอจะยอมรับที่ค่าของตัวอย่างดี

claim ลืมน้อยลง $\mu_d < 0$; after - before ต้องติดลบถึงจะดี

$H_0 : \mu_d \geq 0$

$H_a : \mu_d < 0$



t_α at dof = 12-1

= 11

$t_{0.05} = 1.7976$ (ดู t, area to the right)

$t_{0.05} = -1.7976$ (ฝั่งซ้าย=ซ้อน)

$t = \frac{\bar{d} - \mu_d}{\left(\frac{S_d}{\sqrt{n}}\right)}$ → จาก null hypo

$= \frac{-1.583 - 0}{\left(\frac{2.610}{\sqrt{12}}\right)}$

$= -2.102$

ตกใน rejection regeion

∴ reject H_0 with $\alpha = 0.05$