การอนุมานค่าเฉลี่ยของประชากร

208263 : สถิติเบื้องต้น

โดย... ผศ. ดร.สุคนธ์ ประสิทธิ์วัฒนเสรี

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

วัตถุประสงค์

- นักศึกษาสามารถทำการอนุมานค่าเฉลี่ย 1 กลุ่ม และ 2 กลุ่มได้อย่างถูกต้อง เหมาะสม

เนื้อหา

- บทนำ
- การอนุมานค่าเฉลี่ยประชากร 1 กลุ่ม
  - การแจกแจงค่าเฉลี่ยของตัวอย่าง
  - การประมาณค่า
  - การทดสอบสมมติฐาน
- การอนุมานค่าเฉลี่ยประชากร 2 กลุ่ม
  - การแจกแจงค่าเฉลี่ยของตัวอย่าง
  - การประมาณค่า
  - การทดสอบสมมติฐาน

การอนุมานค่าเฉลี่ย 1 กลุ่ม

การประมาณค่า

การทดสอบสมมติฐาน

การอนุมานค่าของประชากร

- สนใจศึกษาความสูงเฉลี่ยของนักศึกษา ม.ช.

Population → **นักศึกษา ม.ช.**

ความสูงเฉลี่ย ?

Parameter

ในกรณีที่มีข้อจำกัดไม่สามารถทำการรวบรวมข้อมูลในนักศึกษา ม.ช. ทั้งหมดได้

จึงเลือกใช้วิธีสุ่มตัวอย่างนักศึกษา 50 คน มาศึกษา

การอนุมานค่าเฉลี่ยประชากร

Sample → **นักศึกษา 50 คน**

ความสูงเฉลี่ย = 168cm

Statistic

การอนุมานค่าของประชากร

Population → **นักศึกษา ม.ช.**

ความสูงเฉลี่ย ?

Parameter?

อยากรู้ว่ามีค่าเป็นเท่าไร?

สงสัยว่ามีค่าเท่ากับ 175cm?

การอนุมานค่าเฉลี่ยปชก. มี 2 แบบ

การประมาณค่า

การทดสอบสมมติฐาน

ความสูงเฉลี่ย 168cm

Sample → **นักศึกษาม.ช. 50 คน**

การประมาณค่าเฉลี่ย 1 กลุ่ม

การประมาณค่าแบบจุด (Point estimate)

การประมาณค่าแบบช่วง (Interval estimate)

การประมาณค่าเฉลี่ยประชากร 1 กลุ่ม

- **การประมาณค่าแบบจุด (Point estimate)**

  ใช้<u>ค่าเฉลี่ยตัวอย่าง</u> - <u>sample mean</u> เป็นค่าประมาณของ <u>ค่าเฉลี่ยประชากร</u> -<u>population mean (μ)</u>

อยากทราบความสูงเฉลี่ยของนศ. มช.
Population mean = 168

จากตัวอย่างนศ. 50 คน
Sample mean = 168

การประมาณค่าแบบจุด

Population
N

Population : 0, 1, 2, 3

Population mean = **1.5**

Population mean = **1.5**

Sample
n = 2

$x_1 = 2$

$x_2 = 1$

Sample mean = (2+1)/2 = 1.5

การประมาณค่าแบบจุด

การประมาณค่าแบบจุด

ไม่สามารถควบคุมความคลาดเคลื่อนนี้ได้

Population : 0, 1, 2, 3

Population mean = 1.5

Population mean = 2.5

เกิดความคลาดเคลื่อนในการประมาณค่า

Population N

Sample n = 2

$x_1 = 2$

$x_2 = 3$

Sample mean = 2.5

การประมาณค่าแบบจุด

การประมาณค่าแบบช่วง (Interval Estimation)

Population
N

Population : 0, 1, 2, 3
Population mean = **1.5**

Population mean **= 2.5 $\pm$ 2(0.7)**
= [1.1 – 3.9]

Sample
n = 2

$x_1 = 2$
$x_2 = 3$

Sample mean = 2.5
Sample sd. = 0.7

การประมาณค่าแบบช่วง

ขนาดของช่วงประมาณนี้ขึ้นอยู่กับ

- ลักษณะการแจกแจงของค่าเฉลี่ยตัวอย่าง
- ระดับความเชื่อมั่น

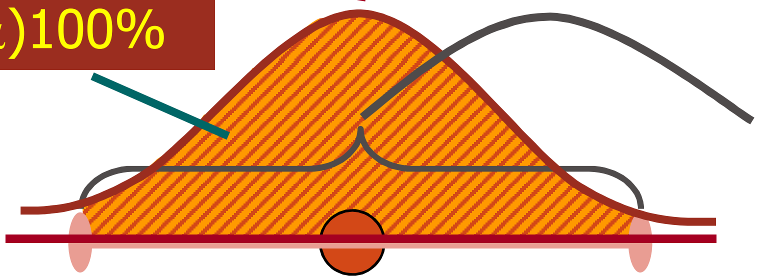
การประมาณค่าเฉลี่ยแบบช่วง

ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของค่าเฉลี่ยประชากร

หมายถึงโอกาสที่ช่วง $[C_1 - C_2]$ จะครอบคลุมค่าเฉลี่ยประชากรเท่ากับ $(1-\alpha)100\%$

การแจกแจงค่าเฉลี่ยตัวอย่าง

สปส.ความเชื่อมั่น $(1 - \alpha)100\%$



Confidence Interval (CI)

ช่วงความเชื่อมั่น

C_1 Sample mean C_2

Sample mean

Population mean = ?

$$c_1 < \mu < c_2$$

C_1 = lower limit (ขอบล่าง)

C_2 = upper limit (ขอบบน)

การแจกแจงค่าเฉลี่ยของตัวอย่าง
(Sampling Distribution of Sample Mean)

Population

μ, σ^2

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

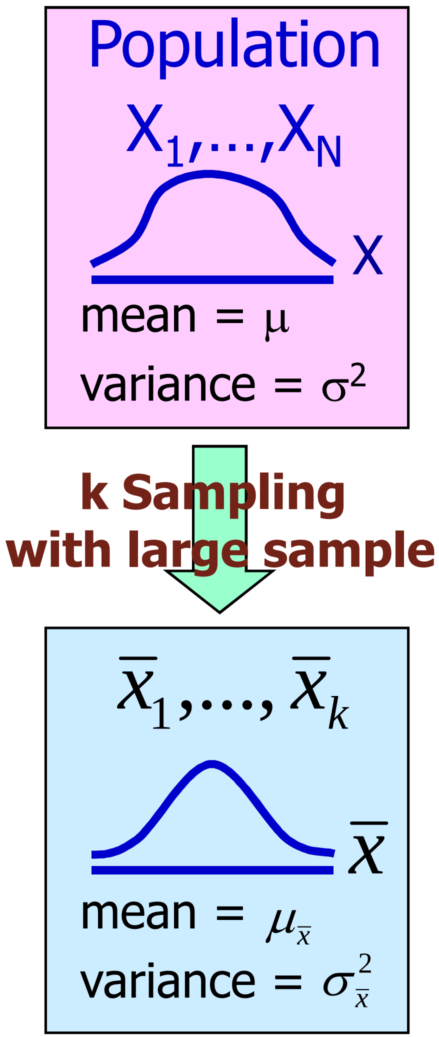
Study 1 .... Study 2 .... Study k

$\bar{x}_1, s_1$ $\qquad$ $\bar{x}_2, s_2$ $\qquad$ $\bar{x}_k, s_k$

พิจารณาลักษณะการแจกแจงของค่าเฉลี่ยตัวอย่าง

$$\begin{array}{c} \bar{x} \\ \bar{x}\ \bar{x}\ \bar{x} \\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x} \\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x} \\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x}\ \bar{x} \end{array}$$

Sampling Distribution of Sample Mean



- การแจกแจงของค่าเฉลี่ยตัวอย่างมีค่าเฉลี่ยเท่ากับค่าเฉลี่ยของประชากร

$$\mu_{\bar{x}} = \mu$$

- การแจกแจงของค่าเฉลี่ยตัวอย่างมีความแปรปรวน, $\sigma_{\bar{x}}^2$, คือ

$$\sigma_{\bar{x}}^2 = \sigma^2 / n$$

sampling with replacement

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \left[\frac{N-n}{N-1}\right]$$

sampling without replacement

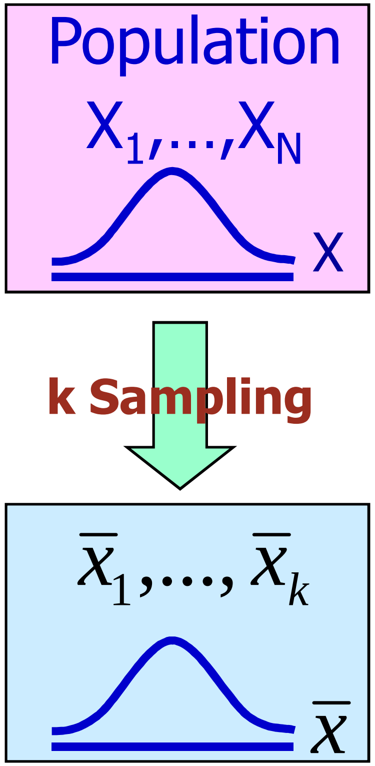
ส่วน $\sigma_{\bar{x}}$ มักรู้จักในชื่อ **standard error**

Sampling Distribution of Sample Mean

Central Limit Theorem

- ถ้าประชากรมีการแจกแจงปกติซึ่งมีความแปรปรวนเท่ากับ σ^2
- สุ่มตัวอย่างขนาด n จำนวน k ชุด จะพบว่า

  การแจกแจงของค่าเฉลี่ยตัวอย่างมีการแจกแจงปกติ

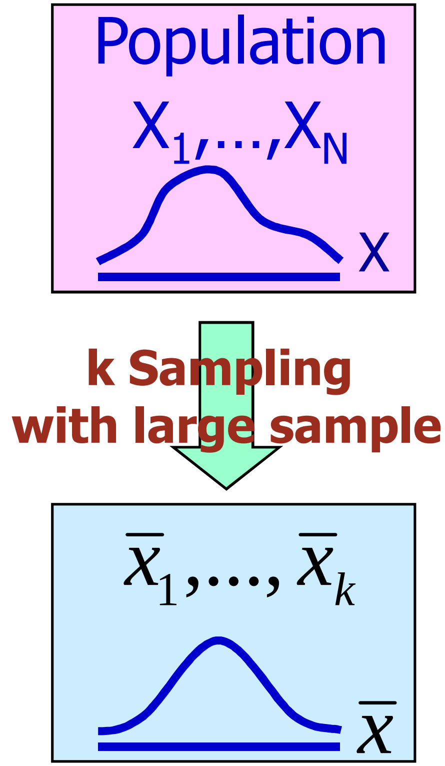


Sampling Distribution of Sample Mean

Central Limit Theorem

- ถ้าประชากรมีการแจกแจงใด ๆ
- สุ่มตัวอย่างขนาด n, โดยที่ $n \geq 30$, จำนวน k ชุด จะพบว่า

  การแจกแจงของค่าเฉลี่ยตัวอย่างมีการแจกแจงเข้าใกล้การแจกแจงปกติ

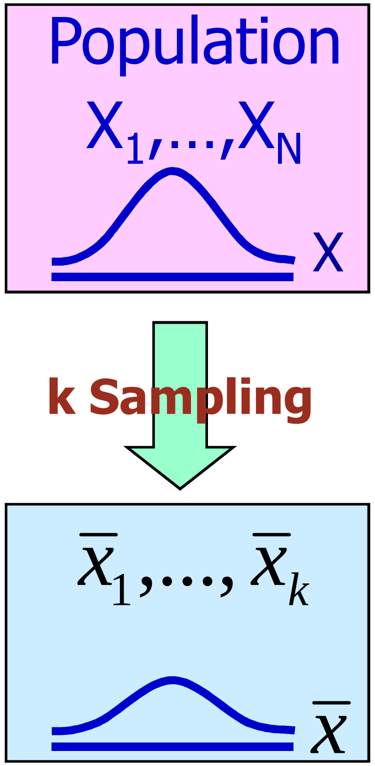


- ขนาดตัวอย่างยิ่งมาก การแจกแจงจะยิ่งเข้าใกล้การแจกแจงปกติยิ่งขึ้น

Sampling Distribution of Sample Mean

- ถ้าประชากรมีการแจกแจงปกติ ซึ่งไม่ทราบค่าความแปรปรวน (σ^2 unknown)
- สุ่มตัวอย่างขนาด n, เมื่อ n < 30, จำนวน k ชุด จะพบว่า

  การแจกแจงของค่าเฉลี่ยตัวอย่างมีการแจกแจงเข้าใกล้การแจกแจงแบบที (t distribution)



การแจกแจงของค่าเฉลี่ยตัวอย่าง

- ลักษณะการแจกแจงค่าเฉลี่ยตัวอย่าง

<u>กรณีที่ 1</u> เมื่อ $X \sim N(\mu, \sigma^2)$ และทราบค่า σ^2

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma^2_{\bar{x}}) \xrightarrow{\text{Transform}} Z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

โดยที่

$$\mu_{\bar{x}} = \mu$$

$$\sigma^2_{\bar{x}} = \sigma^2 / n$$

เมื่อสุ่มตัวอย่างแบบใส่กลับคืน

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \cdot \left[\frac{N-n}{N-1}\right]$$

เมื่อสุ่มตัวอย่างแบบไม่ใส่กลับคืน

$$\approx \frac{\sigma^2}{n}$$

เมื่อ N ขนาดใหญ่

การแจกแจงของค่าเฉลี่ยตัวอย่าง

กรณีที่ 2 เมื่อ n $\geq$ 30 และไม่ทราบ σ^2

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$

Transform

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

โดยที่

$$\sigma_{\bar{x}}^2 = s^2/n$$

เมื่อ

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

การแจกแจงของค่าเฉลี่ยตัวอย่าง

<u>กรณีที่ 3</u> เมื่อประชากรมีการแจกแจงปกติซึ่งไม่ทราบ σ^2 และขนาดตัวอย่าง n < 30

$$\bar{x} \sim t_{v=n-1}$$

Transform

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

โดยมี ค่าเฉลี่ย = $\mu_{\bar{x}}$ ความแปรปรวน = $\sigma_{\bar{x}}^2$

เมื่อ $\sigma_{\bar{x}}^2 = s^2 / n$ $\quad s^2 = \dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}$

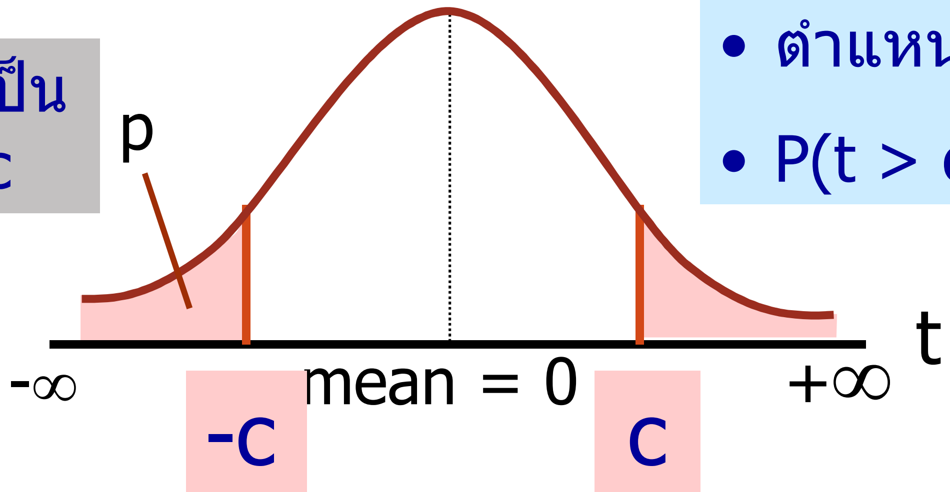
การแจกแจงแบบที (Student’s t distribution)

คุณสมบัติ เมื่อตัวแปรสุ่ม t มีการแจกแจงแบบที

1) t เป็นตัวแปรสุ่มชนิดต่อเนื่อง มีค่าเฉลี่ย = 0

2) โค้งการแจกแจงเป็นแบบสมมาตรที่ t = 0

เขียนได้เป็น
$t_p = -c$



- ตำแหน่ง c กับ –c สมมาตรกัน
- $P(t > c) = P(t < -c)$

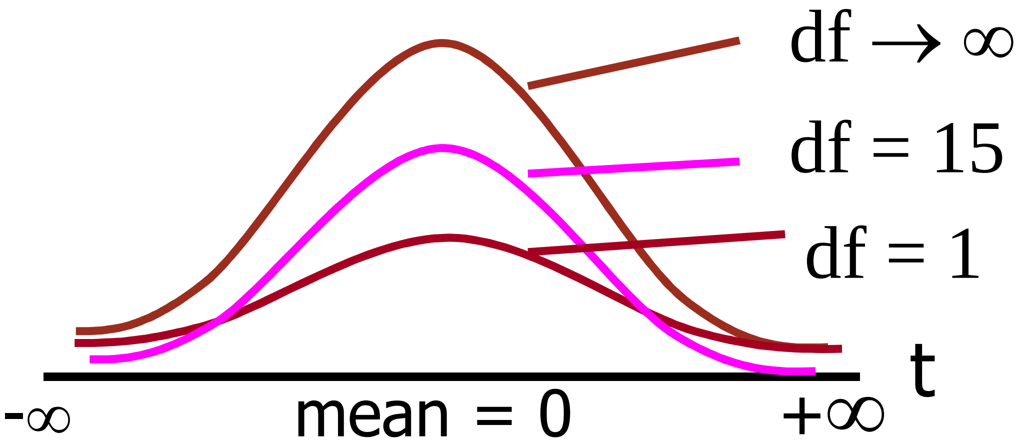
เขียนได้เป็น
$t_{1-p} = c$

3) เมื่อ $n \rightarrow \infty$, t จะมีการแจกแจงเข้าใกล้ปกติ

การแจกแจงแบบ t

ลักษณะโค้งการแจกแจง t

ขึ้นอยู่กับค่าดีกรีความเป็นอิสระ (df. or ν (อ่านว่า nu))



ค่า df คือ จำนวนข้อมูลที่เป็นอิสระไม่ขึ้นกับข้อบังคับใด ๆ

ในกรณีของค่าเฉลี่ยตัวอย่าง df = n – 1

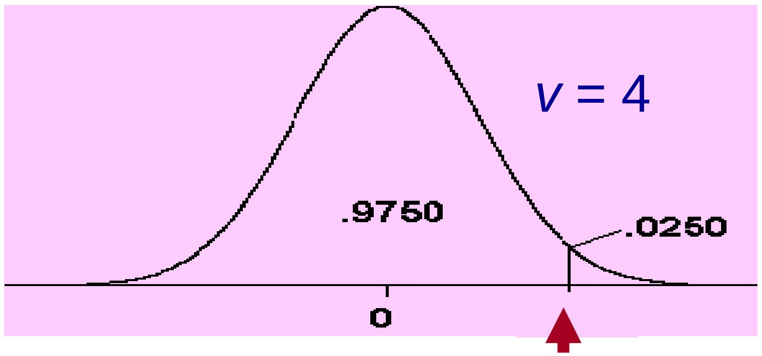
Table Cumulative Student's t Distribution

v \ p	.60	.70	.80	.90	.95	.975	.990	.995	.999	.9995
1	0.000	1.000	1.376	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
…										

$t_{0.975,\, v = 4} = 2.776$

or $P(t < 2.776) = .975$

$t_{0.95,\, v = 8} = ?$



การประมาณค่าแบบช่วง

Population
N

Population mean = ?

$$c_1 < \mu < c_2$$

การประมาณค่าแบบช่วง

Sample

$\overline{x}$ = 2.5

s = 0.7

ขนาดของช่วงประมาณนี้ขึ้นกับ

-ลักษณะการแจกแจงของ $\overline{x}$

- การแจกแจงปกติ (มี 2 แบบ)
- การแจกแจงแบบที

- ระดับความเชื่อมั่น

$(1-\alpha)100\%$

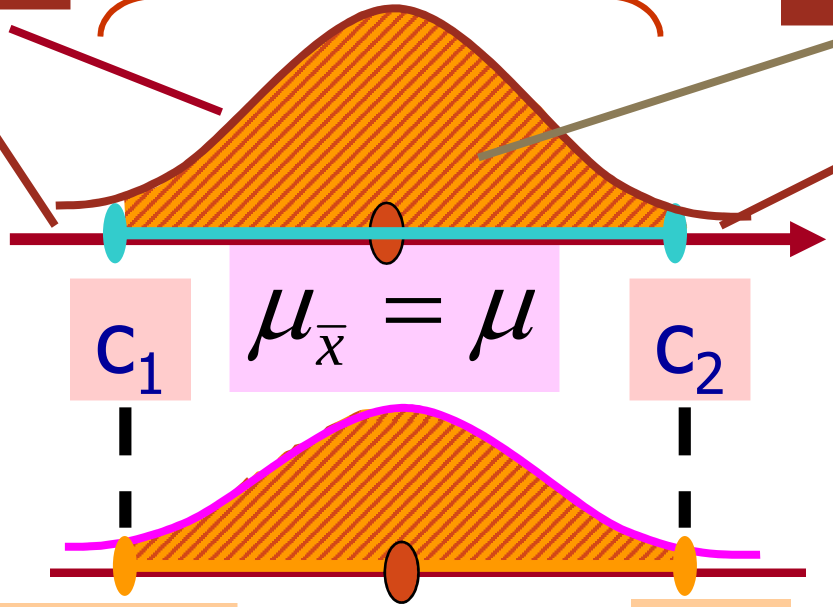
การประมาณค่าเฉลี่ยที่ระดับความเชื่อมั่น (1 - α)

การแจกแจง
ค่าเฉลี่ยตัวอย่าง

Confidence Interval

สปส.ความเชื่อมั่น
(1 - α)

$\alpha/2$ $\alpha/2$



ค่าประมาณแบบช่วง

$c_1 < \mu < c_2$

c_1 $\mu_{\bar{x}} = \mu$ c_2 $\bar{x}$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$-z_{1-\alpha/2}$ or $z_{\alpha/2}$ $z_{1-\alpha/2}$

$$P\left[-z_{1-\frac{\alpha}{2}} < z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$P\left[-z_{1-\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$P\left[\boxed{\bar{x} - z_{1-\frac{\alpha}{2}} . \sigma_{\bar{x}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} . \sigma_{\bar{x}}}\right] = 1-\alpha$$

การประมาณค่าเฉลี่ยแบบช่วง

ค่าประมาณแบบช่วงของค่าเฉลี่ยที่ระดับความเชื่อมั่น $(1 - \alpha)100\%$

$$\bar{x} - z_{1-\frac{\alpha}{2}}.\sigma_{\bar{x}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}.\sigma_{\bar{x}}$$

หรือ $$\mu = \bar{x} \pm z_{1-\frac{\alpha}{2}}.\sigma_{\bar{x}}$$

ลักษณะการแจกแจง

<u>กรณีที่ 1</u> ทราบ σ^2

$$\mu = \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

<u>กรณีที่ 2</u> n $\geq$ 30, ไม่ทราบ σ^2

$$\mu = \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

<u>กรณีที่ 3</u> n < 30, ไม่ทราบ σ^2

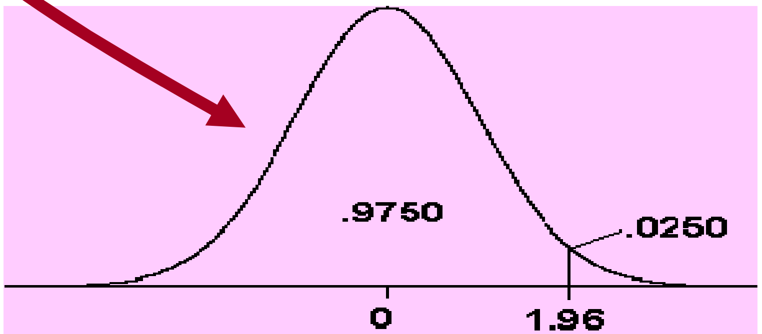
$$\mu = \bar{x} \pm t_{1-\alpha/2,\upsilon=n-1} \cdot \frac{s}{\sqrt{n}}$$

การเปิดค่า Critical value, $z_{1-\alpha/2}$

Table Cumulative Standard Normal Distribution

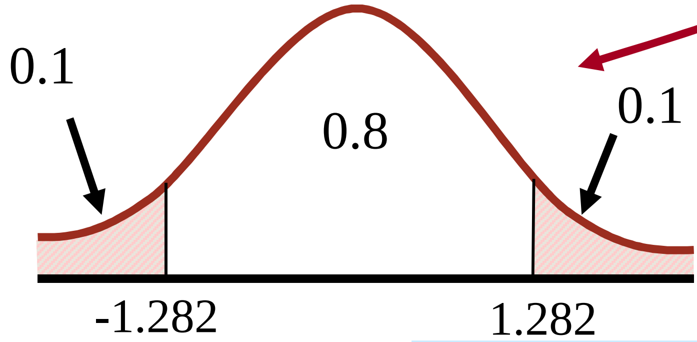
z	1.282	1.645	1.960	2.326	2.576	3.090	...
F(z)	.90	.95	.975	.99	.995	.999	...
2[1-F(z)]	.20	.10	.05	.02	.01	.002	...

เช่น $\alpha = 0.05$, $z_{1-.05/2} = z_{0.975}$



$z_{0.975} = 1.96$

เช่น $\alpha = 0.2$



$-z_{0.9} = -1.282$

$z_{0.9} = 1.282$

Ex : Finding critical value of z and t

- จงหาค่า critical value, z_c ที่ระดับความเชื่อมั่น 99%

- จงหาค่า critical value, t_c ที่ระดับความเชื่อมั่น 95% เมื่อขนาดตัวอย่าง = 15.

การประมาณค่าแบบช่วง μ

การแจกแจงค่าเฉลี่ยตัวอย่าง $\bar{x}$	$(1 - \alpha)100\%$ CI ของ μ
กรณี 1: ทราบค่า σ^2 $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ เมื่อ $\mu_{\bar{x}} = \mu, \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$	$\bar{x} \pm (z_{1-\frac{\alpha}{2}} \cdot \sigma/\sqrt{n})$
กรณี 2: ไม่ทราบค่า σ^2 และ $n \geq 30$ $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ เมื่อ $\sigma_{\bar{x}}^2 = \frac{s^2}{n}$	$\bar{x} \pm (z_{1-\frac{\alpha}{2}} \cdot s/\sqrt{n})$
กรณี 3: ไม่ทราบค่า σ^2 และ $n < 30$ $\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \sim t_{\upsilon = n-1}$ เมื่อ $\sigma_{\bar{x}}^2 = \frac{s^2}{n}$	$\bar{x} \pm (t_{1-\frac{\alpha}{2}} \cdot s/\sqrt{n})$

ตัวอย่าง 3 : การประมาณค่าเฉลี่ย

จากการสำรวจด้วยการสุ่มตัวอย่างเด็กไทยอายุ 15 ปีมา 16 คน พบว่าโดยเฉลี่ยมีน้ำหนัก 38 กก. ด้วยค่าเบี่ยงเบนมาตรฐาน 9 กก. จงหาช่วงความเชื่อมั่น 90% ของน้ำหนักเฉลี่ยของเด็กไทยอายุ 15 ปี

Confidence Interval (CI) ที่ 90%

$n = 16, \bar{x} = 38, s = 9$

$$\mu = \bar{x} \pm t_{1-\frac{\alpha}{2}, v=n-1} \cdot \frac{s}{\sqrt{n}}$$

90% CI => $\alpha = .10$

$t_{1-.10/2} = t_{.95,\ v=15} = 1.753$

$$= 38 \pm 1.753(9/\sqrt{16})$$

$= 38 \pm 3.94$ $= 34.06 - 41.94$

ดังนั้น น้ำหนักเฉลี่ยของเด็กไทยอายุ 15 ปีที่ระดับความเชื่อมั่น 90% มีค่าระหว่าง 34.06 – 41.94 กก.

การทดสอบสมมติฐาน ค่าเฉลี่ย 1 กลุ่ม

การอนุมานค่าของประชากร

Population → **นักศึกษา ม.ช.**

Parameter = ส่วนสูงเฉลี่ย

Unknown?

อยากรู้ว่ามีค่าเป็นเท่าไร?

สงสัยว่ามีค่าเท่ากับ 175cm?

การประมาณค่า

การทดสอบสมมติฐาน

การอนุมานค่าเฉลี่ยปชก.

Sample → **นักศึกษาม.ช. 50 คน**

ส่วนสูงเฉลี่ย 168cm = Statistic

การทดสอบสมมติฐาน (Hypothesis Testing)

- **หลักการทดสอบสมมติฐานทางสถิติ**

ถ่าน AA กระรอกบิน ใช้ได้นานโดยเฉลี่ย 300 นาที (คำโฆษณาของผู้ผลิต)

นาย ก. สงสัยในคำโฆษณาดังกล่าว ???

จึงทำการศึกษาเพื่อตรวจสอบคำโฆษณาดังกล่าว

เลือกถ่านมาอย่างสุ่ม 100 ก้อน ทดลองใช้ แล้วคำนวณอายุการใช้งานเฉลี่ยว่าน้อยกว่าหรือเท่ากับ 300 นาทีหรือไม่?

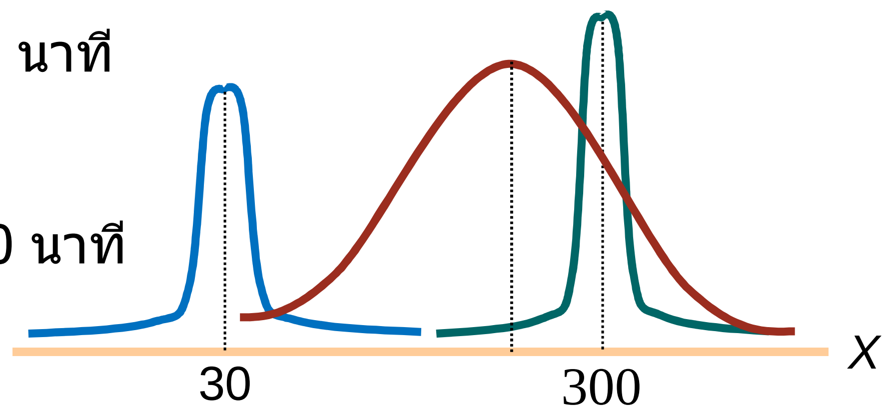
หลักการทดสอบสมมติฐานทางสถิติ

ถ่าน AA กระรอกบิน ใช้ได้นานโดยเฉลี่ย 300 นาที
(คำโฆษณาของผู้ผลิต)

แล้วจะตัดสินใจอย่างไรดี !!!

เมื่อผลการทดสอบถ่าน 100 ก้อน พบว่า

- อายุการใช้งานโดยเฉลี่ย (mean) = 300 นาที และค่า sd = 0.1 นาที
  = **คำโฆษณาน่าเชื่อถือ**
- mean = 30 นาที และค่า sd = 0.5 นาที
  = **คำโฆษณาเกินจริง**
- mean = 294 นาที และค่า sd = 20 นาที



หลักการทดสอบสมมติฐานทางสถิติ

ถ่าน AA กระรอกบิน ใช้ได้นานโดยเฉลี่ย 300 นาที

(คำโฆษณาของผู้ผลิต)

แนวคิดทางสถิติสำหรับการตัดสินใจ

ภายใต้กรณี n = 100, mean = 294 นาที และค่า sd = 20 นาที

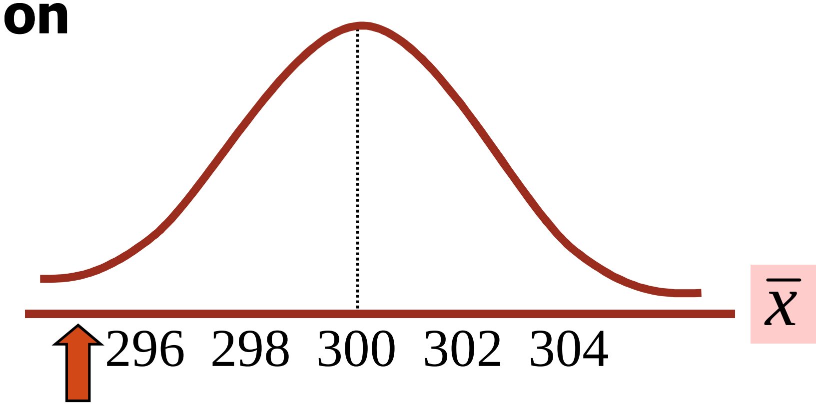
ในทางสถิติ มองว่าผลการศึกษาที่ได้นี้เป็นเพียงค่าที่เป็นไปได้หนึ่งค่า หากทำการศึกษาหลาย ๆ ครั้งก็จะได้ค่า mean ออกมาต่าง ๆ กัน

จึงควรพิจารณา **Sampling Distribution**

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma^2_{\bar{x}})$$

$$\mu_{\bar{x}} = \mu = 300$$

$$\sigma_{\bar{x}} = s/\sqrt{n} = 20/\sqrt{100} = 2$$



$$\bar{x} = 294$$

$\bar{x}$ ห่างจากค่า μ ค่อนข้างมาก จึงไม่ควรเชื่อคำโฆษณานี้

หลักการทดสอบสมมติฐานทางสถิติ

ถ่าน AA กระรอกบิน ใช้ได้นานโดยเฉลี่ย 300 นาที
(คำโฆษณาของผู้ผลิต)

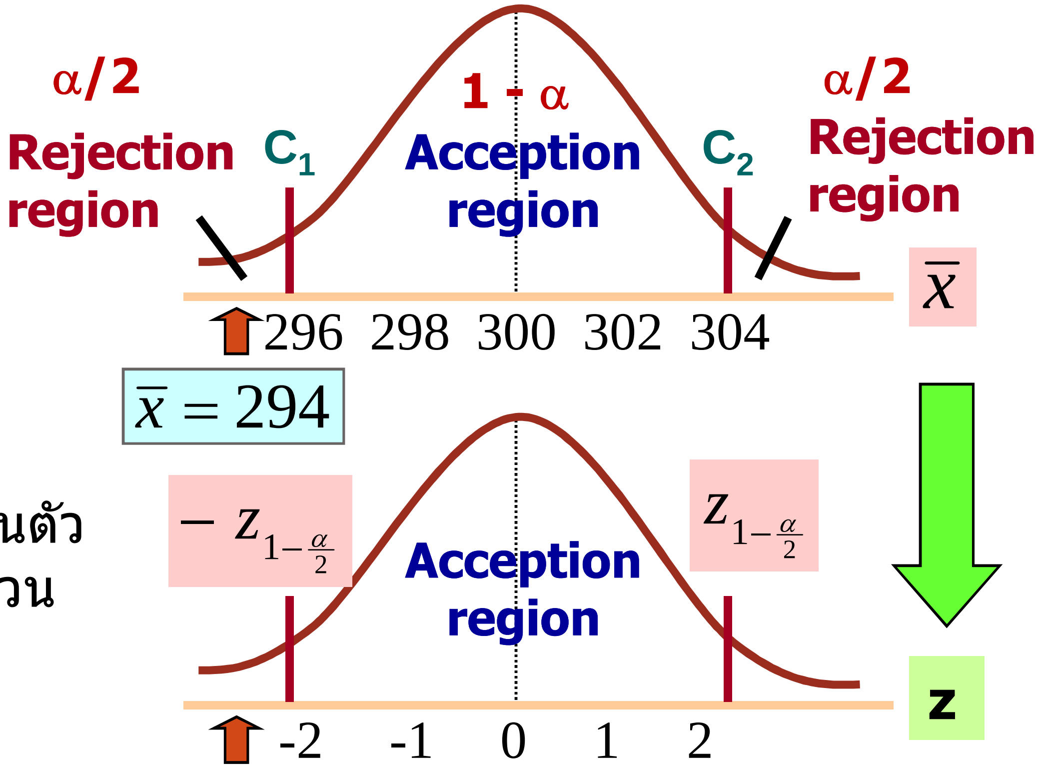
$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$

$$\mu_{\bar{x}} = \mu = 300$$

$$\sigma_{\bar{x}} = s/\sqrt{n} = 20/\sqrt{100} = 2$$

เพื่อให้เกิดมาตรฐานในการตัดสินใจ

ทำการกำหนดค่าวิกฤต (C_1, C_2) เป็นตัวแบ่งอาณาเขตการตัดสินใจเป็น 2 ส่วน



$$\bar{x} = 294$$

$$z = \frac{(294-300)}{2} = -3$$

ส่วนประกอบในการทดสอบสมมติฐาน

- **สมมติฐานทางสถิติ** (Statistical Hypothesis)
- ชนิดของความคลาดเคลื่อน (Type of Error) และระดับนัยสำคัญ (Level of Significance)
- สถิติทดสอบ และอาณาเขตวิกฤต
- การตัดสินใจ และสรุปผล

สมมติฐานทางสถิติ

- คือข้อสันนิษฐานเกี่ยวกับลักษณะของประชากร 1 กลุ่มหรือหลายกลุ่ม ซึ่งอาจจะเป็น**จริง**หรือ**ไม่จริง**ก็ได้

  โดยทั่วไปแล้ว จากข้อสงสัยหรือข้อคำถามที่ต้องการตรวจสอบในทางสถิติจะกำหนดขึ้นเป็นสมมติฐาน 1 คู่ ประกอบด้วย

  - **สมมติฐานหลัก (Null Hypothesis), H_0**
    เป็นสมมติฐานที่ใช้ในการทดสอบพารามิเตอร์ ดังนั้นต้องระบุค่าที่**ชัดเจน**เกี่ยวกับพารามิเตอร์
  - **สมมติฐานทางเลือก (Alternative Hypothesis), H_1**
    เป็นสมมติฐานทางเลือกในกรณีที่ปฏิเสธ H_0

$H_0 : \mu = k$	$H_0 : \mu \geq k$	$H_0 : \mu \leq k$
$H_1 : \mu \neq k$	$H_1 : \mu < k$	$H_1 : \mu > k$

Two-tailed test   One-tailed test

สมมติฐานทางสถิติ

ลองเขียนสมมติฐานทางสถิติจากข้อสงสัยต่อไปนี้

- ถ่าน AA กระรอกบิน ใช้ได้นานเฉลี่ยเท่ากับ 300 นาที จริงหรือไม่

ถ้าให้ μ แทนระยะเวลาเฉลี่ยของการใช้งานถ่าน AA

ข้อสงสัยในที่นี้คือ $\boxed{\mu = 300}$ **?**

สมมติฐานทางสถิติ

ลองเขียนสมมติฐานทางสถิติจากข้อสงสัยต่อไปนี้

- ผู้ผลิต TV กล่าวว่า อายุการใช้งานเฉลี่ยของ TV รุ่นหนึ่งมากกว่า 8 ปี

- รถยนต์ยี่ห้อหนึ่งโฆษณาว่า ช่วยประหยัดน้ำมัน มากกว่า โดยสามารถวิ่งได้ระยะทางเฉลี่ยไม่น้อยกว่า 28 กม./ลิตร

ชนิดของความคลาดเคลื่อน และระดับนัยสำคัญ

ทุก ๆ ครั้งในการทดสอบสมมติฐานทางสถิติ จะเป็นการทดสอบค่าพารามิเตอร์ภายใต้ H_0 ซึ่งข้อสรุปที่เป็นไปได้มี 2 กรณี คือ

- ปฏิเสธ H_0 (Reject H_0)
- ไม่ปฏิเสธ H_0 (No reject H_0)

ในการทดสอบทางสถิติ อาจเกิดการตัดสินใจที่ผิดพลาดได้ ทั้งนี้เพราะอาศัยข้อมูลที่ได้จากกลุ่มตัวอย่างเป็นฐานในการตัดสินใจ

Type I and Type II Errors

Population
10, 20, 30, 40, 50
mean = 30

Sample (n=2)
10, 20
mean = 15

พิจารณากรณีต่อไปนี้

- **กรณีที่ 1**

  สมมติฐาน $H_0 : \mu = 30$ vs. $H_1 : \mu \neq 30$

  การตัดสินใจ ปฏิเสธ H_0

  (ทั้งนี้เพราะภายใต้ H_0 เชื่อว่า $\mu = 30$ แต่ $\bar{x} = 15$ แตกต่างกันเกินไป)

Type I error คือปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นจริง

- **กรณีที่ 2**

  สมมติฐาน $H_0 : \mu = 15$ vs. $H_1 : \mu \neq 15$

  การตัดสินใจ ไม่ปฏิเสธ H_0

  (ทั้งนี้เพราะภายใต้ H_0 เชื่อว่า $\mu = 15$ และ $\bar{x} = 15$ ใกล้เคียงกันมาก)

Type II error คือไม่ปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นเท็จ

ระดับนัยสำคัญ

- **Type I error = การปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นจริง**

  ความน่าจะเป็นที่จะเกิด Type I error เรียกว่า **ระดับนัยสำคัญ (level of significance) ของการทดสอบ** $\rightarrow \alpha$ (alpha)

- **Type II error = การไม่ปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นเท็จ**

  ความน่าจะเป็นที่จะเกิด Type II error $\rightarrow \beta$ (beta)

  ค่า **$(1 - \beta)$** เรียกว่า **กำลังในการทดสอบ (power of test)**

โดยปกติ ในการทดสอบสมมติฐานทางสถิติมักกำหนด**ค่าระดับนัยสำคัญ (α)** ให้มีค่าน้อย ๆ นั่นคือต้องการให้โอกาสที่จะปฏิเสธ H_0 ที่เป็นจริงเกิดขึ้นน้อย ๆ

ที่นิยมใช้คือ α **= 0.10,** α **= 0.05,** α **= 0.01**

สถิติทดสอบ และอาณาเขตวิกฤต

- หลังจากกำหนด H_0 และ H_1 รวมถึงค่า α

พิจารณาตัวสถิติทดสอบ และอาณาเขตวิกฤต จากลักษณะการแจกแจงของค่าเฉลี่ยตัวอย่าง

Sampling distribution | **สถิติทดสอบ**

เมื่อทราบค่า σ^2

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$

$$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

เมื่อไม่ทราบค่า σ^2

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2) \quad (n \geq 30)$$

$$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

$$\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} \sim t \quad (n < 30)$$

$$t = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

สถิติทดสอบ และอาณาเขตวิกฤต

- **อาณาเขตวิกฤต = rejection region**

  เป็นช่วงที่บ่งบอกว่า H_0 ไม่น่าเป็นไปได้ หากค่าสถิติทดสอบตกอยู่ช่วงนี้ จึงควรทำการปฏิเสธ H_0

  โดยกำหนดอาณาเขตวิกฤตมีพื้นที่ = α (เพื่อให้โอกาสในการปฏิเสธ H_0 นั้น จะเป็นการปฏิเสธ H_0 ที่เป็นจริงไม่เกิน α)

  ตามปกติมักกำหนดให้มีค่าน้อย ๆ เช่น **α = 0.10, α = 0.05, α = 0.01**

อาณาเขตวิกฤต

- **ค่าวิกฤต (critical value)**

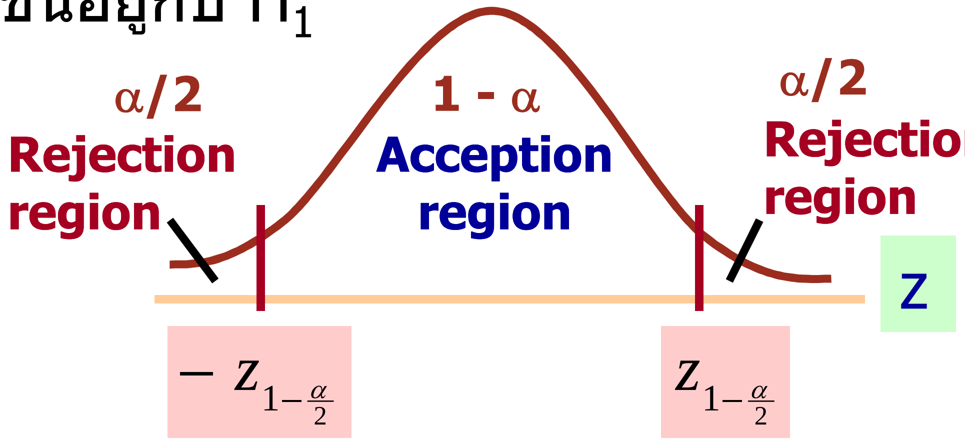
  คือค่าที่แบ่งระหว่างเขตปฏิเสธ H_0 กับเขตไม่ปฏิเสธ H_0

- **รูปแบบอาณาเขตวิกฤต** ขึ้นอยู่กับ H_1

  - การทดสอบแบบ 2 ทาง

$H_0 : \mu = k$

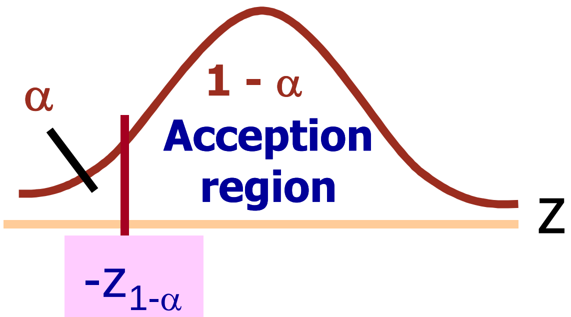
$\boxed{H_1 : \mu \neq k}$



  - การทดสอบแบบทางเดียว

$H_0 : \mu = k$

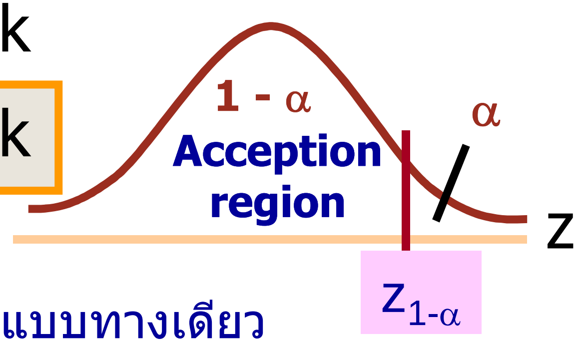
$\boxed{H_1 : \mu < k}$



การทดสอบแบบทางเดียวด้านน้อยหรือซ้ายมือ

$H_0 : \mu = k$

$\boxed{H_1 : \mu > k}$



การทดสอบแบบทางเดียวด้านมากหรือทางขวามือ

การตัดสินใจ และสรุปผล

- **ปฏิเสธ H_0** ถ้าค่าสถิติทดสอบตกอยู่ในเขตปฏิเสธ H_0
- **ไม่ปฏิเสธ H_0** ถ้าค่าสถิติทดสอบไม่ตกอยู่ในเขตปฏิเสธ H_0

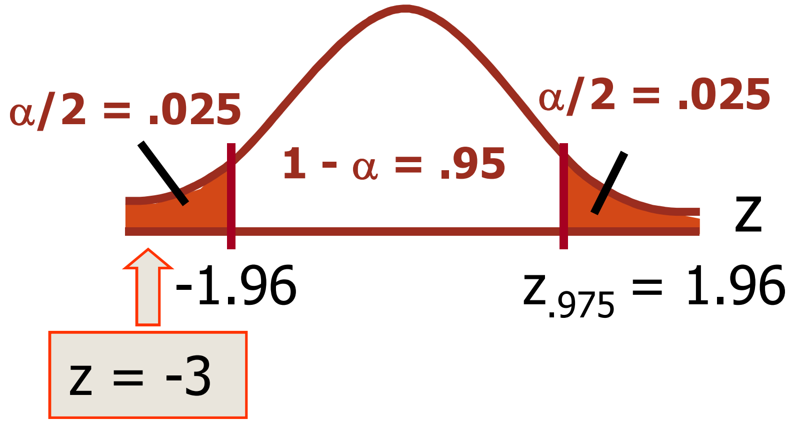
เช่น $H_0 : \mu = 300$ vs. $H_1 : \mu \neq 300$

กรณีที่ 1 : $n = 100, \bar{x} = 294, s = 20$

สถิติทดสอบ $$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu_{\bar{x}}}{s/\sqrt{n}} = \frac{294-300}{20/\sqrt{100}} = -3$$

อาณาเขตวิกฤต : ที่ $\alpha = 0.05$

$z > 1.96$ หรือ $z < -1.96$



การตัดสินใจ :

ปฏิเสธ H_0

การตัดสินใจ และสรุปผล

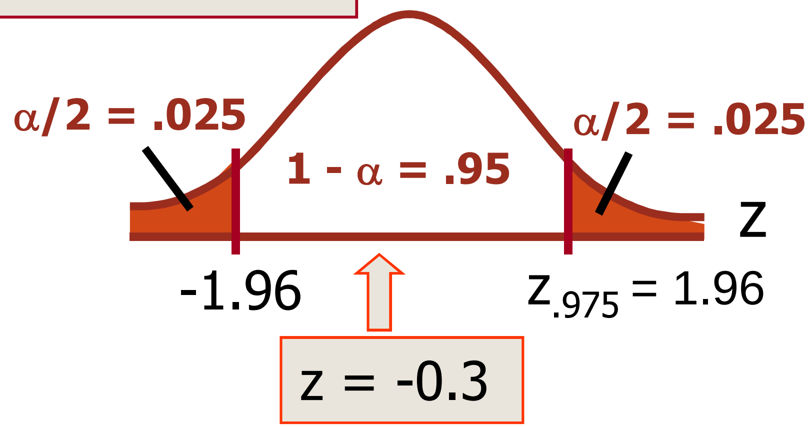
เช่น $H_0 : \mu = 300$ vs. $H_1 : \mu \neq 300$

<u>กรณีที่ 2</u> : $n = 100, \bar{x} = 294, s = 200$

สถิติทดสอบ : $$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{294-300}{200/\sqrt{100}} = -0.3$$

อาณาเขตวิกฤต : ที่ $\alpha = 0.05$

$z > 1.96$ หรือ $z < -1.96$



การตัดสินใจ :

ไม่ปฏิเสธ H_0

สรุปขั้นตอนการทดสอบสมมติฐาน

1. กำหนดสมมติฐานทางสถิติ
2. เลือกตัวสถิติทดสอบ
3. กำหนดอาณาเขตวิกฤต ภายใต้ระดับนัยสำคัญ α
4. คำนวณค่าสถิติ
5. สรุปผลการทดสอบสมมติฐาน

Hypothesis Testing

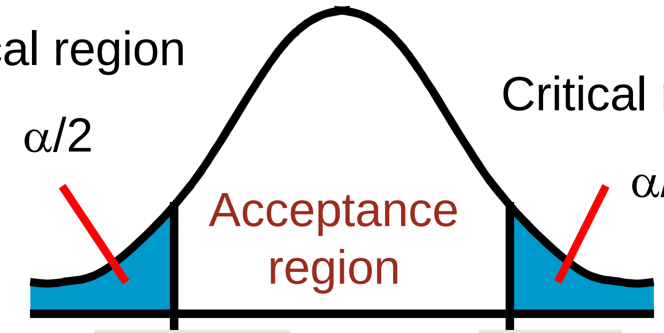
- **สมมติฐานทางสถิติ**

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

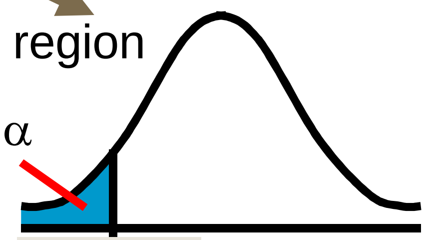
$H_0 : \mu = \mu_0$ or $\mu \geq \mu_0$
$H_1 : \mu < \mu_0$

$H_0 : \mu = \mu_0$ or $\mu \leq \mu_0$
$H_1 : \mu > \mu_0$

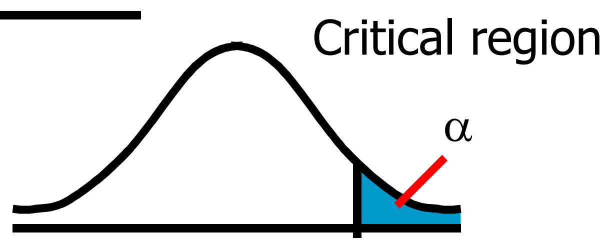
- **เขตวิกฤต ที่ α**



$-Z_{1-\frac{\alpha}{2}}$ $\quad$ $Z_{1-\frac{\alpha}{2}}$



$-Z_{1-\alpha}$



$Z_{1-\alpha}$

- **สถิติทดสอบ**

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}},$$

เมื่อทราบ σ^2

Hypothesis Testing

- **สมมติฐานทางสถิติ**

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

$H_0 : \mu = \mu_0$ or $\mu \geq \mu_0$
$H_1 : \mu < \mu_0$

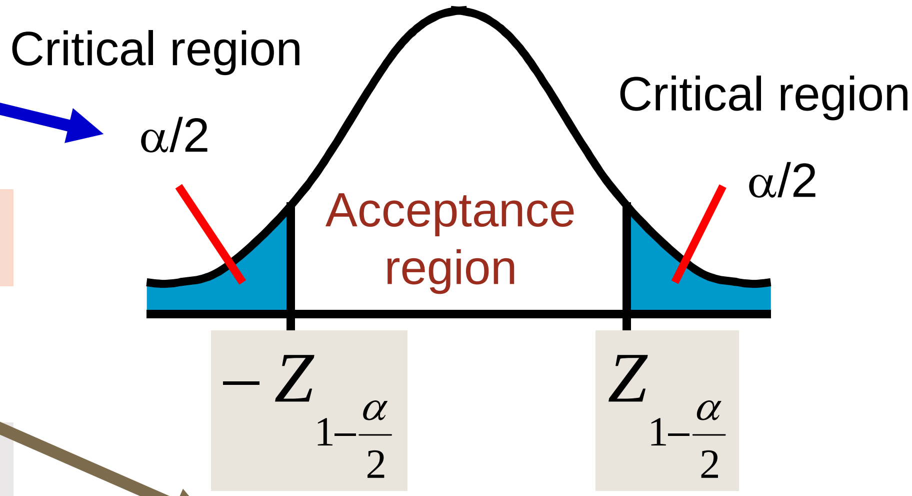
$H_0 : \mu = \mu_0$ or $\mu \leq \mu_0$
$H_1 : \mu > \mu_0$

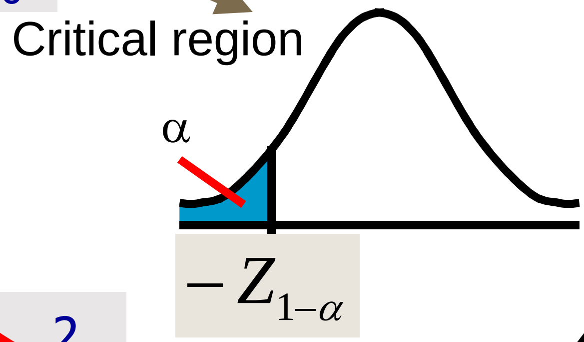
- **สถิติทดสอบ**

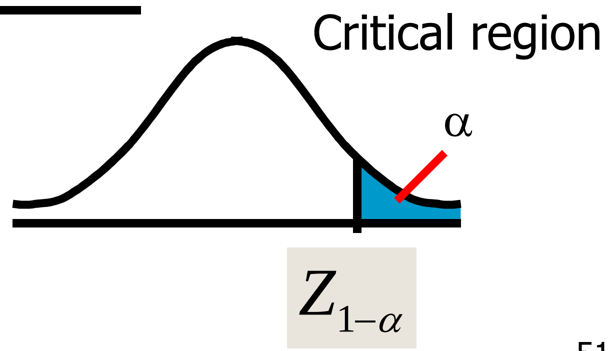
$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

เมื่อไม่ทราบ σ^2,
$n \geq 30$

- **เขตวิกฤต ที่ α**







Hypothesis Testing

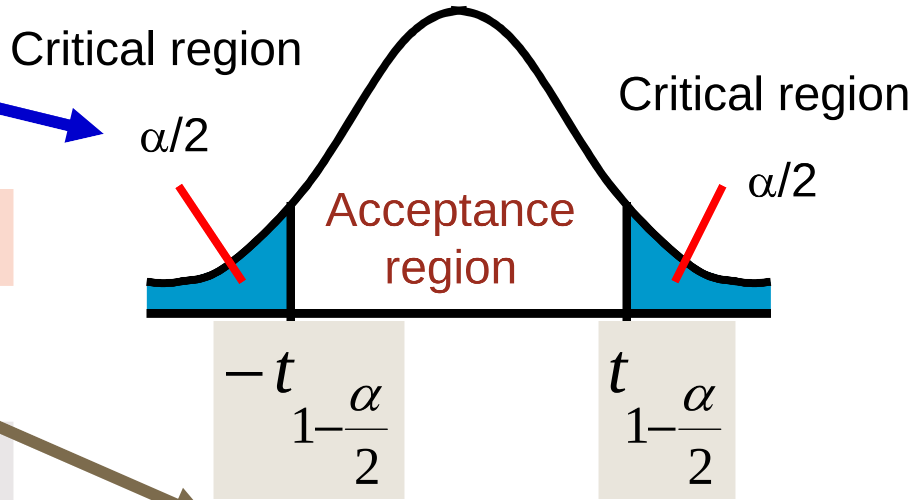
- **สมมติฐานทางสถิติ**

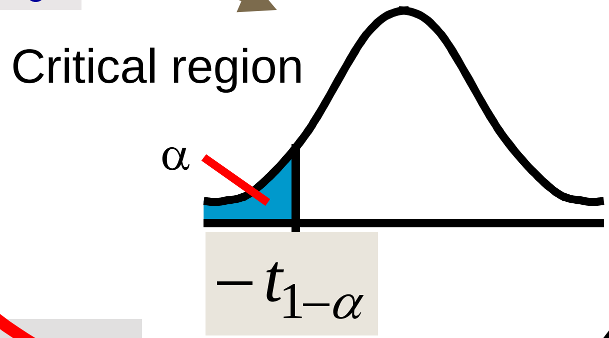
$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

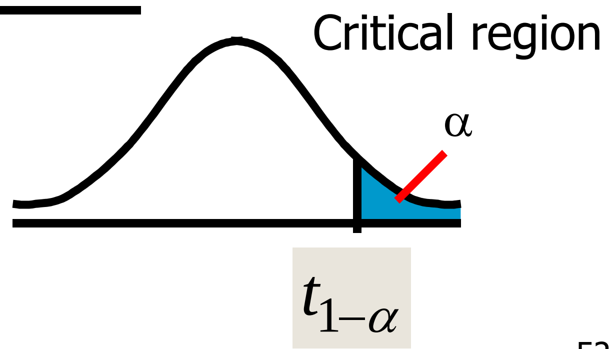
$H_0 : \mu = \mu_0$ or $\mu \geq \mu_0$
$H_1 : \mu < \mu_0$

$H_0 : \mu = \mu_0$ or $\mu \leq \mu_0$
$H_1 : \mu > \mu_0$

- **เขตวิกฤต ที่ α**







- **สถิติทดสอบ**

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}},$$

เมื่อไม่ทราบ σ^2,
$n < 30$

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย

H_0	H_1	ตัวสถิติทดสอบ	เขตวิกฤต
$\mu = \mu_0$	$\mu \neq \mu_0$	กรณี 1 : ทราบค่า σ^2 $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z < -z_{1-\alpha/2}$ or $Z > z_{1-\alpha/2}$
	$\mu > \mu_0$		$Z > z_{1-\alpha}$
	$\mu < \mu_0$	กรณี 2: ไม่ทราบ σ^2, $n \geq 30$ $Z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$Z < -z_{1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	กรณี 3 : ไม่ทราบ σ^2, $n<30$ $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}; \upsilon = n-1$	$t < -t_{1-\alpha/2,\nu}$ or $t > t_{1-\alpha/2,\nu}$
	$\mu > \mu_0$		$t > t_{1-\alpha,\nu}$
	$\mu < \mu_0$		$t < -t_{1-\alpha,\nu}$

ตัวอย่างที่ 4

โรงงานผู้ผลิตเหล็กเส้นต้องการตรวจสอบคุณภาพการผลิต เพื่อที่จะพิจารณาว่าการผลิตได้มาตรฐานหรือไม่ ถ้ากำหนดมาตรฐานความยาวเฉลี่ยเท่ากับ 8.6 นิ้ว และค่าเบี่ยงเบนมาตรฐานของความยาวเหล็กเส้นเท่ากับ 0.3 นิ้ว ถ้าสุ่มตัวอย่างเหล็กเส้นที่ผลิตจากโรงงานนี้ 36 เส้น พบว่า ค่าเฉลี่ยของความยาวเหล็กเส้นตัวอย่างเท่ากับ 8.7 นิ้ว

จะสรุปได้หรือไม่ว่า การผลิตเหล็กเส้นของโรงงานนี้ได้มาตรฐานค่าเฉลี่ยที่ระดับนัยสำคัญ 0.05

ต้องการให้ทำการทดสอบสมมติฐานทางสถิติ

ตัวอย่างที่ 4

Population → ความยาวเหล็กเส้นที่ผลิตโดยโรงงานนี้

Parameter : μ = ความยาวเฉลี่ย = ?

σ = 0.3

สงสัยว่าได้มาตรฐานหรือไม่ (μ = 8.6)

การทดสอบสมมติฐาน

Sample → n = 36

Statistic : ความยาวเฉลี่ยของเหล็กเส้นตัวอย่าง, $\bar{x}$ = 8.7

ตัวอย่างที่ 4

คำถาม การผลิตได้มาตรฐานค่าเฉลี่ยหรือไม่ ?

ให้ μ = ความยาวเฉลี่ยของเหล็กเส้น

ต้องการทดสอบ $\mu = 8.6$?

จากโจทย์ $\sigma = 0.3$, $n = 36$, $\bar{x} = 8.7$

สมมติฐานทางสถิติ $H_0 : \mu = 8.6$

$H_1 : \mu \neq 8.6$

สถิติทดสอบ ทราบ σ^2

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

คำนวณค่าสถิติทดสอบ

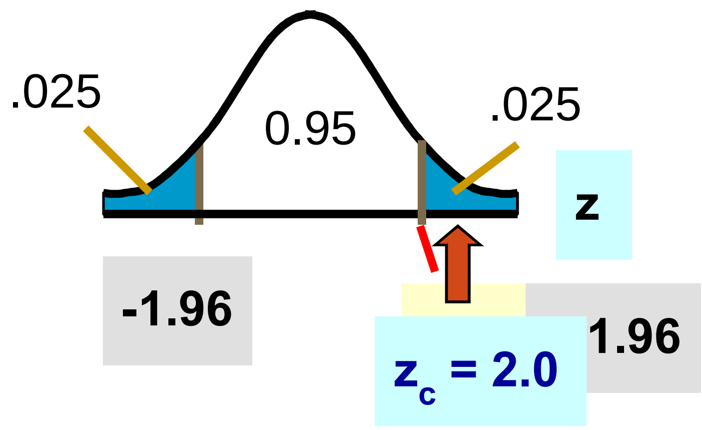
$$z_c = \frac{8.7 - 8.6}{0.3/\sqrt{36}} = 2.0$$

เขตวิกฤต

$\alpha = .05$

z < -1.96 หรือ

z > 1.96



สรุปผล

$z_c > 1.96$ ดังนั้น **ปฏิเสธ H_0**

สรุปได้ว่าความยาวเฉลี่ยของเหล็กเส้นยังไม่ได้มาตรฐาน ที่ $\alpha = 0.05$

ตัวอย่างที่ 5

จากการสำรวจครัวเรือนในเขตเทศบาลนครเชียงใหม่ จำนวน 36 ครัวเรือน พบว่า โดยเฉลี่ยครัวเรือนชมรายการโทรทัศน์ 27 ชั่วโมง/สัปดาห์ ค่าเบี่ยงเบนมาตรฐาน 4 ชั่วโมง/สัปดาห์

ถ้าจำนวนชั่วโมงการชมโทรทัศน์โดยเฉลี่ยของครัวเรือนทั่วประเทศเท่ากับ 25 ชั่วโมง/สัปดาห์

จะกล่าวสรุปได้หรือไม่ว่า จำนวนชั่วโมงการชมรายการโทรทัศน์โดยเฉลี่ยของครัวเรือนในเขตเทศบาลนครเชียงใหม่มากกว่าทั่วประเทศ ที่ระดับนัยสำคัญ 0.01

ต้องการให้ทำการทดสอบสมมติฐานทางสถิติ

ตัวอย่างที่ 5

Population → **ครัวเรือนในเขตเทศบาลเชียงใหม่**

Parameter :

μ = จน.ชั่วโมงเฉลี่ยที่ดู TV = ?

σ = ?

สงสัยว่าจน.ชั่วโมงเฉลี่ยที่ดู TV ของคนเชียงใหม่มากกว่าคนทั่วประเทศ (ซึ่งดู 25 ชม.) หรือไม่ ?

(μ **>** 25) ?

<u>การทดสอบสมมติฐาน</u>

Sample → **n = 36**

Statistic : จน.ชั่วโมงเฉลี่ยที่ดู TV ของตัวอย่าง, $\bar{x}$ = 27

ค่าเบี่ยงเบนมาตรฐาน, s = 4

ตัวอย่างที่ 5

คำถาม ต้องการทดสอบ $\mu > 25$?

จากโจทย์ σ = ?,
$n = 36, \overline{x} = 27, s = 4$

สมมติฐานทางสถิติ $H_0 : \mu = 25$

$H_1 : \mu > 25$

สถิติทดสอบ ไม่ทราบ σ^2, $n \geq 30$

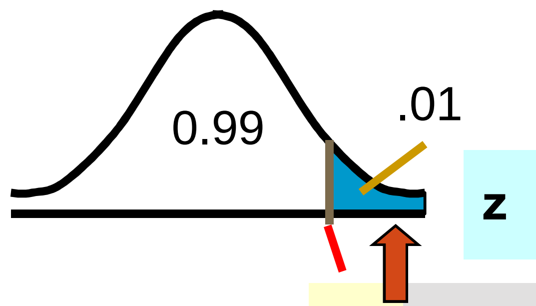
$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

คำนวณค่าสถิติทดสอบ

$$z_c = \frac{27-25}{4/\sqrt{36}} = 3.0$$

เขตวิกฤต **$z > 2.326$**

$\alpha = .01$



$z_c = 3.0$

สรุปผล $z_c > 2.326$ ดังนั้น **ปฏิเสธ H_0**

สรุปได้ว่าจน.ชั่วโมงเฉลี่ยที่ดู TV ของคน
เชียงใหม่มากกว่าคนทั้งประเทศ ที่ $\alpha = 0.01$

ตัวอย่างที่ 6

ระบบการลงทะเบียนเรียนของมหาวิทยาลัยแห่งหนึ่ง โดยเฉลี่ยแล้วน.ศ.คนหนึ่งจะใช้เวลา 3.1 ชม. มหาวิทยาลัยแห่งนั้นจึงจัดระบบการลงทะเบียนใหม่ เพื่อลดระยะเวลาในการลงทะเบียนให้น้อยลง มหาวิทยาลัยจึงทดลองระบบลงทะเบียนเรียนใหม่ โดยจากการทดลองกับนักศึกษา 21 คน บันทึกเวลาลงทะเบียนเรียน (ชั่วโมง) ดังนี้

3.1 2.9 2.7 2.8 3.0 2.6 2.9 2.9 3.0 2.7 2.9 3.0

2.8 3.1 2.9 2.7 3.2 2.8 2.8 3.1 3.0

จากข้อมูลการทดลอง จะสรุปได้หรือไม่ว่า ระบบการลงทะเบียนใหม่ใช้เวลาน้อยลงกว่าเดิม ที่ระดับนัยสำคัญ 0.05

ต้องการให้ทำการทดสอบสมมติฐาน

ระบบการลงทะเบียนเรียนของมหาวิทยาลัยแห่งหนึ่ง โดยเฉลี่ยแล้ว น.ศ.คนหนึ่งจะใช้เวลา 3.1 ชม. มหาวิทยาลัยแห่งนั้นจึงจัดระบบการลงทะเบียนใหม่ เพื่อลดระยะเวลาในการลงทะเบียนให้น้อยลง มหาวิทยาลัยจึงทดลองระบบลงทะเบียนเรียนใหม่ โดยจากการทดลองกับนักศึกษา 21 คน

Population → **นักศึกษาทั้งหมด**

Parameter :

μ = เวลาเฉลี่ยในการลงทะเบียนแบบใหม่= ?

σ = ?

Sample → **n = 21**

3.1 2.9 2.7 2.8 3.0 2.6 2.9 2.9 3.0 2.7 2.9
3.0 2.8 3.1 2.9 2.7 3.2 2.8 2.8 3.1 3.0

ตรวจสอบว่าเวลาเฉลี่ยในการลงทะเบียนแบบใหม่น้อยกว่าเดิม (3.1 ชม.)หรือไม่ ?

(μ < 3.1) ?

<u>การทดสอบสมมติฐาน</u>

Statistic : $\bar{x}$ = ?

s = ?

Sample → **n = 21**

3.1 2.9 2.7 2.8 3.0 2.6 2.9 2.9
3.0 2.7 2.9 3.0 2.8 3.1 2.9 2.7
3.2 2.8 2.8 3.1 3.0

Statistic :

$$\bar{x} = \frac{3.1 + 2.9 + \ldots + 3.0}{21} = 2.9$$

$$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{(3.1^2 + \ldots + 3.0^2) - 21(2.9^2)}{20} = 0.025$$

$$s = \sqrt{0.025} = 0.158$$

ตัวอย่างที่ 6

จากโจทย์ σ = ?,
n = 21, $\bar{x}$ = 2.9,
s = 0.025

คำถาม ต้องการทดสอบ $\mu < 3.1$?

สมมติฐานทางสถิติ $H_0 : \mu = 3.1$

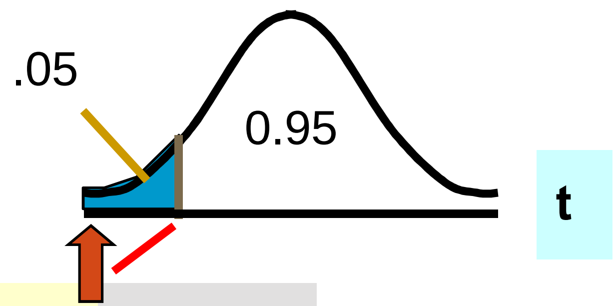
$H_1 : \mu < 3.1$

สถิติทดสอบ ไม่ทราบ σ^2, n < 21

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

เขตวิกฤต t < -1.725

α = .05 $\quad \nu$ = 21-1=20



-1.725

t_c = -5.80

คำนวณค่าสถิติทดสอบ

$$t_c = \frac{2.9-3.1}{0.158/\sqrt{21}} = -5.80$$

สรุปผล t_c < -1.725 ดังนั้น **ปฏิเสธ H_0**

สรุปได้ว่าเวลาเฉลี่ยในการลงทะเบียนระบบใหม่น้อยกว่าระบบเดิม ที่ α = 0.05

การอนุมานค่าเฉลี่ย 2 กลุ่ม

การประมาณค่า

การทดสอบสมมติฐาน

ประชากร 2 กลุ่ม

Population 1 - N_1

Parameter :

μ_1 = ?

σ_1 = ?

Sample 1 - n_1

Statistic :

$\bar{x}_1$

s_1

Population 2 - N_2

Parameter :

μ_2 = ?

σ_2 = ?

Sample 2 - n_2

Statistic :

$\bar{x}_2$

s_2

การอนุมานค่าเฉลี่ยประชากร 2 กลุ่ม

สนใจ : ผลต่างระหว่างค่าเฉลี่ย 2 กลุ่ม ($\mu_1 - \mu_2$)

การอนุมานค่า $\mu_1 - \mu_2$

- **การประมาณค่า**

  การประมาณค่าแบบจุดและแบบช่วง

- **การทดสอบสมมติฐาน**

การประมาณค่า μ_1 - μ_2 (ไม่นิยมใช้)

- การประมาณค่าแบบจุดของ μ_1 - μ_2
  สามารถประมาณได้จากค่า $\boxed{\bar{x}_1 - \bar{x}_2}$

- การประมาณค่าแบบช่วงของ μ_1 - μ_2
  วิธีการประมาณที่ใช้จะขึ้นอยู่กับ
  - ลักษณะการแจกแจงของ $\boxed{\bar{x}_1 - \bar{x}_2}$
  - ระดับความเชื่อมั่น $(1 - \alpha)100\%$

การประมาณค่าแบบช่วง $\mu_1 - \mu_2$

การแจกแจง $\bar{x}_1 - \bar{x}_2$	$(1 - \alpha)100\%$ CI ของ $\mu_1 - \mu_2$
กรณี 1: ทราบค่า σ_1^2, σ_2^2 $\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma^2_{\bar{x}_1 - \bar{x}_2})$ เมื่อ $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2,$ $\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$	$\bar{x}_1 - \bar{x}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
กรณี 2: ไม่ทราบค่า σ_1^2, σ_2^2 และ $n_1, n_2 \geq 30$ $\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma^2_{\bar{x}_1 - \bar{x}_2})$ เมื่อ $\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$	$\bar{x}_1 - \bar{x}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

การประมาณค่าแบบช่วง $\mu_1 - \mu_2$

การแจกแจง $\boxed{\bar{x}_1 - \bar{x}_2}$	$(1 - \alpha)100\%$ CI ของ $\mu_1 - \mu_2$
กรณี 3: ไม่ทราบค่า σ_1^2, σ_2^2 และ (n_1 หรือ n_2) < 30 $\sigma_1^2 = \sigma_2^2$ $\quad \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} \sim t_\upsilon$ เมื่อ $\sigma^2_{\bar{x}_1-\bar{x}_2} = s_p^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)$ $\quad \upsilon = n_1 + n_2 - 2$	$\bar{x}_1 - \bar{x}_2 \pm t_{1-\frac{\alpha}{2},\upsilon}\, s_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ $s_p^2 = \dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
$\sigma_1^2 \neq \sigma_2^2$ $\quad \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} \sim t_\upsilon$ เมื่อ $\sigma^2_{\bar{x}_1-\bar{x}_2} = \dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}$ $\quad \upsilon = \dfrac{\left[\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right]^2}{\dfrac{\left[\dfrac{s_1^2}{n_1}\right]^2}{n_1-1}+\dfrac{\left[\dfrac{s_2^2}{n_2}\right]^2}{n_2-1}}$	$\bar{x}_1 - \bar{x}_2 \pm t_{1-\frac{\alpha}{2},\upsilon}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$

ตัวอย่างที่ 8

การทดลองวัดอายุการใช้งานของหลอดไฟชนิด ก จำนวน 150 หลอด ปรากฏว่า มีอายุการใช้งานเฉลี่ย 1400 ชั่วโมง ค่าเบี่ยงเบนมาตรฐาน 120 ชั่วโมง และจากการทดลองวัดอายุการใช้งานของหลอดไฟชนิด ข จำนวน 200 หลอด พบว่ามีอายุการใช้งานเฉลี่ย 1200 ชั่วโมง และค่าเบี่ยงเบนมาตรฐาน 80 ชั่วโมง

จงสร้างช่วงความเชื่อมั่น 95% สำหรับเป็นค่าประมาณของผลต่างระหว่างอายุการใช้งานโดยเฉลี่ยของหลอดไฟชนิด ก กับชนิด ข

ต้องการให้ทำการประมาณค่าแบบช่วง

ตัวอย่างที่ 8

Population → กลุ่ม 1 : หลอดไฟ ก | กลุ่ม 2 : หลอดไฟ ข

Parameter :

μ_1 = อายุใช้งานเฉลี่ย หลอด ก = ?

σ_1 = ?

μ_2 = อายุใช้งานเฉลี่ย หลอด ข = ?

σ_2 = ?

ค่าประมาณของ $\mu_1 - \mu_2$ = ?

$$\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma^2_{\bar{x}_1 - \bar{x}_2})$$

การประมาณค่าแบบช่วง

Sample → $n_1 = 150$ | $n_2 = 200$

Statistic :

$\bar{x}_1 = 1400, \quad s_1 = 120$

$\bar{x}_2 = 1200, \quad s_2 = 80$

ตัวอย่างที่ 8

คำถาม สร้างช่วงความเชื่อมั่น 95% ของ $\mu_1 - \mu_2$

เนื่องจาก ไม่ทราบค่า σ_1, σ_2 และ n_1, $n_2 > 30$

จากโจทย์

$n_1=150, \bar{x}_1=1400,$
$s_1 = 120$

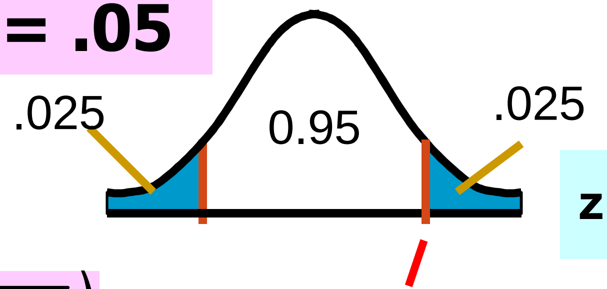
$n_2=200, \bar{x}_2=1200,$
$s_2 = 80$

$$\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1-\bar{x}_2}, \sigma^2_{\bar{x}_1-\bar{x}_2})$$

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

ค่า z

α **= .05**



z

$z_{.975}$ **= 1.96**

$$= (1400 - 1200) \pm 1.96\left(\sqrt{\frac{120^2}{150} + \frac{80^2}{200}}\right)$$

$$= 200 \pm 22.17$$

177.83 < $\mu_1 - \mu_2$ **< 222.17**

สรุปผล ช่วงความเชื่อมั่น 95% ของผลต่างอายุการใช้งานโดยเฉลี่ยของหลอด ก กับหลอด ข คือ 177.83 ถึง 222.17 ชั่วโมง

การทดสอบสมมติฐาน $\mu_1 - \mu_2$

1. กำหนดสมมติฐานทางสถิติ
2. เลือกตัวสถิติทดสอบ
3. กำหนดอาณาเขตวิกฤต ภายใต้ระดับนัยสำคัญ α
4. คำนวณค่าสถิติ
5. สรุปผลการทดสอบสมมติฐาน

การทดสอบสมมติฐาน $\mu_1 - \mu_2$

- สมมติฐานในการทดสอบ

$H_0 : \mu_1 - \mu_2 = d_0$	$H_1 : \mu_1 - \mu_2 \neq d_0$
$H_0 : \mu_1 - \mu_2 = d_0$	$H_1 : \mu_1 - \mu_2 > d_0$
$H_0 : \mu_1 - \mu_2 = d_0$	$H_1 : \mu_1 - \mu_2 < d_0$

ลองตั้งสมมติฐานทางสถิติ

- ต้องการทดสอบว่า ปริมาณน้ำฝนที่ตกโดยเฉลี่ยในเดือนเมษายนของภาคเหนือไม่แตกต่างจากภาคกลาง

ให้ μ_1 = ปริมาณน้ำฝนเฉลี่ยในภาคเหนือ

μ_2 = ปริมาณน้ำฝนเฉลี่ยในภาคกลาง

สงสัย $\mu_1 = \mu_2$?

$\mathbf{H_0}$: $\boxed{\mu_1 = \mu_2}$

$\mathbf{H_1}$: $\mu_1 \neq \mu_2$

การทดสอบสมมติฐาน $\mu_1 - \mu_2$

ลองตั้งสมมติฐานทางสถิติ

- จะกล่าวได้หรือไม่ว่า โดยเฉลี่ยแล้วผู้ป่วยมีระดับคลอเรสเตอรอลสูงกว่าคนปกติเกินกว่า 20 หน่วย

การทดสอบสมมติฐาน $\mu_1 - \mu_2$

H_0	ตัวสถิติทดสอบ	H_1	เขตวิกฤต
$\mu_1 - \mu_2 = d_0$	กรณีทราบค่า σ_1, σ_2 $$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$	$\mu_1 - \mu_2 \neq d_0$	$z < -z_{1-\alpha/2}$ or $z > z_{1-\alpha/2}$
		$\mu_1 - \mu_2 > d_0$	$z > z_{1-\alpha}$
		$\mu_1 - \mu_2 < d_0$	$z < -z_{1-\alpha}$
$\mu_1 - \mu_2 = d_0$	กรณีไม่ทราบ σ_1, σ_2 $$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ เมื่อ $n_1, n_2 \geq 30$	$\mu_1 - \mu_2 \neq d_0$	$z < -z_{1-\alpha/2}$ or $z > z_{1-\alpha/2}$
		$\mu_1 - \mu_2 > d_0$	$z > z_{1-\alpha}$
		$\mu_1 - \mu_2 < d_0$	$z < -z_{1-\alpha}$

การทดสอบสมมติฐาน $\mu_1 - \mu_2$

H_0	ตัวสถิติทดสอบ	H_1	เขตวิกฤต
$\mu_1-\mu_2 = d_0$	กรณีไม่ทราบ σ_1, σ_2 $$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ เมื่อ n_1 or $n_2 < 30$, $\sigma_1 = \sigma_2$	$\mu_1-\mu_2 \neq d_0$	$t < -t_{1-\alpha/2,\nu}$ or $t > t_{1-\alpha/2,\nu}$
		$\mu_1-\mu_2 > d_0$	$t > t_{1-\alpha,\nu}$
		$\mu_1-\mu_2 < d_0$	$t < -t_{1-\alpha,\nu}$
		$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$	$\upsilon = n_1 + n_2 - 2$
$\mu_1-\mu_2 = d_0$	กรณีไม่ทราบ σ_1, σ_2 $$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ เมื่อ n_1 or $n_2 < 30$, $\sigma_1 \neq \sigma_2$	$\mu_1-\mu_2 \neq d_0$	$t < -t_{1-\alpha/2,\nu}$ or $t > t_{1-\alpha/2,\nu}$
		$\mu_1-\mu_2 > d_0$	$t > t_{1-\alpha,\nu}$
		$\mu_1-\mu_2 < d_0$	$t < -t_{1-\alpha,\nu}$

$$\upsilon = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}{\frac{\left[\frac{s_1^2}{n_1}\right]^2}{n_1-1} + \frac{\left[\frac{s_2^2}{n_2}\right]^2}{n_2-1}}$$

ตัวอย่างที่ 11

นักวิจัยท่านหนึ่งต้องการศึกษาเพื่อเปรียบเทียบความถนัดทางช่างระหว่างนศ.ชาย และนศ.หญิง ในคณะวิศวกรรมแห่งหนึ่งว่าจะแตกต่างกันหรือไม่ จึงได้สุ่มตัวอย่างนศ.ชายจำนวน 50 คน และนศ.หญิงจำนวน 50 คน เพื่อทดสอบความถนัดทางช่าง ผลการทดสอบ นศ.ชายได้คะแนนเฉลี่ย 116 คะแนน นศ.หญิงได้คะแนนเฉลี่ย 112 คะแนน ถ้าจากประสบการณ์ที่ผ่านมาพบว่าคะแนนทดสอบทางช่างของนศ.ชายและหญิงมีการแจกแจงปกติ มีส่วนเบี่ยงเบนมาตรฐานเท่ากัน ซึ่งเท่ากับ 15 คะแนน

นักวิจัยผู้นี้จะสรุปผลการศึกษาอย่างไรที่ระดับนัยสำคัญ 0.05

ต้องการให้ทำการทดสอบสมมติฐานทางสถิติ

ตัวอย่างที่ 11

Population → กลุ่ม 1 : นศ.ชาย | กลุ่ม 2 : นศ.หญิง

Parameter :

μ_1 = คะแนนเฉลี่ย นศ.ชาย = ?

σ_1 = 15

μ_2 = คะแนนเฉลี่ย นศ.หญิง = ?

σ_2 = 15

ต้องการเปรียบเทียบว่า $\mu_1 \neq \mu_2$?

$$\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1-\bar{x}_2}, \sigma^2_{\bar{x}_1-\bar{x}_2})$$

การทดสอบสมมติฐาน

Sample →

$n_1 = 50$ | $n_2 = 50$

Statistic :

$\bar{x}_1 = 116$ | $\bar{x}_2 = 112$

ตัวอย่างที่ 11

จากโจทย์ $\sigma_1=\sigma_2=15$,

$n_1 = 50$, $\overline{x_1} = 116$

$n_2 = 50$, $\overline{x_2} = 112$

คำถาม **ต้องการเปรียบเทียบว่า $\mu_1 \neq \mu_2$?**

สมมติฐานทางสถิติ $H_0 : \mu_1 = \mu_2$ $\mu_1-\mu_2=0$

$H_1 : \mu_1 \neq \mu_2$ $\mu_1-\mu_2\neq 0$

สถิติทดสอบ ทราบ σ_1, σ_2

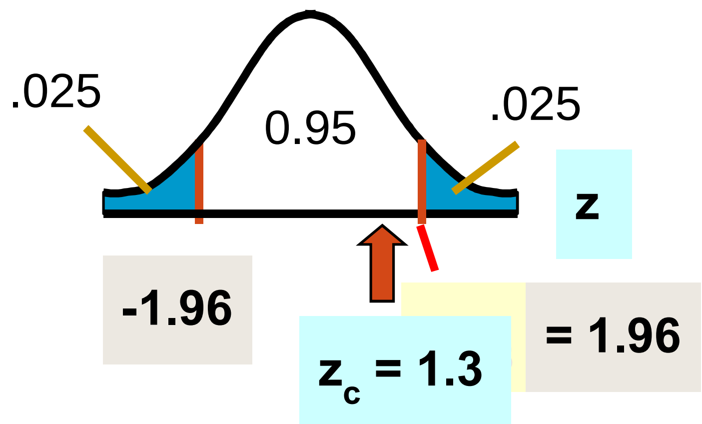
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

เขตวิกฤต

$\alpha = .05$

z < -1.96 หรือ

z > 1.96



คำนวณค่าสถิติทดสอบ

$$z_c = \frac{116-112}{\sqrt{\frac{15^2}{50}+\frac{15^2}{50}}} = 1.33$$

สรุปผล

$z_c < 1.96$ ดังนั้น **ไม่ปฏิเสธ H_0**

นั่นคือยังไม่สามารถกล่าวได้ว่าความถนัดทางช่างระหว่างนศ.ชายและหญิงแตกต่างกัน ที่ $\alpha = 0.05$

ตัวอย่าง 14

หนังสือพิมพ์ฉบับหนึ่งมีจำหน่ายในมหาวิทยาลัย 2 แห่ง ถ้าต้องการศึกษาว่าโดยเฉลี่ยแล้วยอดขายต่อสัปดาห์ของมหาวิทยาลัยทั้งสองจะเท่ากันหรือไม่ จึงสุ่มตัวอย่างยอดขายจากทั้งสองมหาวิทยาลัยได้ข้อมูลดังนี้

มหาวิทยาลัย	จำนวนสัปดาห์	ยอดขายเฉลี่ย	ความแปรปรวน
A	10	123	225
B	6	108	185

จงวิเคราะห์ข้อมูลด้วยระดับนัยสำคัญ 1% ถ้ายอดขายในทั้งสองมหาวิทยาลัยมีความแปรปรวนเท่ากัน

ต้องการให้ทำการทดสอบสมมติฐาน

หนังสือพิมพ์ฉบับหนึ่งมีจำหน่ายในมหาวิทยาลัย 2 แห่ง ถ้าต้องการศึกษาว่าโดยเฉลี่ยแล้วยอดขายต่อสัปดาห์ของมหาวิทยาลัยทั้งสองจะเท่ากันหรือไม่ จึงสุ่มตัวอย่างยอดขายจากทั้งสองมหาวิทยาลัยได้ข้อมูลดังนี้

มหาวิทยาลัย	จำนวนสัปดาห์	ยอดขายเฉลี่ย	ความแปรปรวน
A	10	123	225
B	6	108	185

Population

กลุ่ม 1 : มหาวิทยาลัย A

กลุ่ม 2 : มหาวิทยาลัย B

Parameter

μ_1 **= ยอดขายเฉลี่ย A = ?**

σ_1 **= ?**

$\sigma_1 = \sigma_2$

μ_2 **= ยอดขายเฉลี่ย B = ?**

σ_2 **= ?**

สงสัยว่า $\mu_1 = \mu_2$?

การทดสอบสมมติฐาน

Sample

$n_1 = 10$

$n_2 = 6$

Statistic :

$\bar{x}_1 = 123, \quad s_1^2 = 225$

$\bar{x}_2 = 108, \quad s_2^2 = 185$

ตัวอย่างที่ 14

จากโจทย์ $\sigma_1=\sigma_2$

$n_1 = 10$, $\overline{x}_1 = 123$, $s_1^2 = 225$

$n_2 = 6$, $\overline{x}_2 = 108$, $s_2^2 = 185$

คำถาม ต้องการเปรียบเทียบว่า $\mu_1 = \mu_2$?

สมมติฐานทางสถิติ $H_0 : \mu_1 = \mu_2$ $\quad \mu_1-\mu_2=0$

$H_1 : \mu_1 \neq \mu_2$ $\quad \mu_1-\mu_2\neq 0$

สถิติทดสอบ

ไม่ทราบ σ_1, σ_2

$n_1, n_2 < 30$,

$\sigma_1=\sigma_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

คำนวณค่าสถิติทดสอบ

$$s_p^2 = \frac{(10-1)225+(6-1)185}{(10+6-2)} = 210.71$$

$$s_p = \sqrt{210.71} = 14.516$$

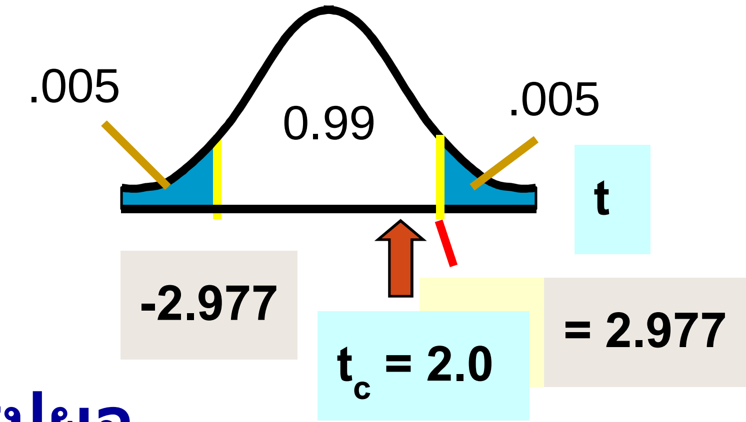
$$t_c = \frac{123-108}{14.516\sqrt{\frac{1}{10}+\frac{1}{6}}} = 2.0$$

เขตวิกฤต

$t < -2.977$

หรือ $t > 2.977$

$\alpha = .01$ $\quad \nu = 10+6-2=14$



สรุปผล

$t_c < 2.977$ ดังนั้น **ไม่ปฏิเสธ H_0**