การประมาณค่าและการทดสอบสมมติฐาน
(Estimation and Testing Hypothesis)

- การประมาณค่า คือ การประมาณสิ่งที่เราสนใจในประชากร โดยที่เราไม่ทราบว่าค่าในประชากร (ค่าพารามิเตอร์) มีค่าเป็นเท่าใด
- การประมาณสามารถประมาณโดยใช้ค่าเพียงค่าเดียวในการประมาณซึ่งเรียกว่าการประมาณแบบจุด
- ประมาณแบบช่วงซึ่งได้จากการสร้างช่วงมาหนึ่งช่วง
- การประมาณเป็นการประมาณค่าพารามิเตอร์ เช่น ค่าเฉลี่ย ค่าสัดส่วน และ ค่าส่วนเบี่ยงเบนมาตรฐาน เป็นต้น

- ***การประมาณค่าแบบจุด***

การประมาณค่าแบบจุด คือ การใช้ค่าจากตัวอย่าง เพียงค่าเดียว ประมาณ ค่าพารามิเตอร์ของประชากรเช่น

ใช้ค่าเฉลี่ยตัวอย่างแทนด้วย $\overline{X}$ ประมาณ ค่าเฉลี่ยของประชากร แทนด้วย μ

ใช้ค่าสัดส่วนตัวอย่างแทนด้วย p ประมาณ ค่าสัดส่วนของประชากร แทนด้วย π

ใช้ค่าแปรปรวนตัวอย่างแทนด้วย S^2 ประมาณค่าแปรปรวนของประชากรแทนด้วย σ^2

- **<u>ตัวอย่างที่ 1</u>**

- สุ่มตัวอย่างนิสิต 10 คนเพื่อทำการประมาณค่าใช้จ่ายเฉลี่ยต่อวันของนิสิตสำนักวิชาเศรษฐศาสตร์ได้ ข้อมูลดังนี้
- 120 145 185 100 95 65 120 100 80 90
- จากข้อมูลดังกล่าว โจทย์ต้องการประมาค่าเฉลี่ยคือ μ
- สามารถใช้ค่าเฉลี่ยตัวอย่างคือ ทำการประมาณค่า μ

- n= 10 = $\bar{X} = \dfrac{\sum_{i=1}^{10} X_i}{n} = \dfrac{120 + 145 + \dots + 90}{10}$ = 110

- $\therefore$ ด้วยการประมาณค่าแบบจุด สามารถประมาณค่าเฉลี่ยของประชากรเกี่ยวกับค่าใช้จ่ายเฉลี่ยต่อวันของนิสิตเป็น 110 บาท

- การประมาณแบบช่วง คือการสร้างช่วงของตัวเลขช่วงหนึ่ง โดยอาศัยค่าที่ได้จากตัวอย่าง ช่วงเชื่อมั่นของการประมาณประกอบด้วย
- ขีดจำกัดของการประมาณ สร้างโดยอาศัยการแจกแจงของค่าสถิติที่ได้จากตัวอย่าง เช่น การประมาณค่าเฉลี่ยของประชากรเราสามารถสร้างโดยพิจารณาจาก
- การแจกแจงของค่าเฉลี่ยตัวอย่างสุ่ม ซึ่งในการประมาณค่าพารามิเตอร์ จะต้องอาศัยการแจกแจงค่าสถิติของตัวอย่างเสมอ การประมาณค่าจะบอกถึงขีดจำกัดของการประมาณ ซึ่งประกอบด้วยขีดจำกัดล่างและขีดจำกัดบน
- ระดับความเชื่อมั่น การสร้างช่วงเชื่อมั่นต้องบอกว่าช่วงที่สร้างมีโอกาสของความถูกต้องร้อยละเท่าใด กำหนดโดยใช้โอกาสของความถูกต้องเป็น $1-\alpha$ และมีโอกาสผิดพลาดเป็น α ความผิดพลาด α เรียกว่าระดับนัยสำคัญ
-

ขั้นตอนของการประมาณค่า

- กำหนดพารามิเตอร์ที่ต้องการประมาณว่าเป็นค่าใดและจะใช้ค่าสถิติอะไรทำการประมาณ
- กำหนดโอกาสของการผิดพลาดหรือ ระดับนัยสำคัญ α
- พิจารณาการแจกแจงค่าสถิติว่ามีการแจกแจงแบบใด ในบทการแจกแจงค่าที่ได้จากตัวอย่าง และเขียนช่วงเชื่อมั่น โดยอาศัยข้อมูลจากตัวอย่าง

ตารางสรุปการแจกแจงของค่าสถิติและช่วงของการประมาณค่า

ค่าพารามิเตอร์	ค่าสถิติ	กรณีและการแจกแจงค่าสถิติ	ขีดจำกัดของการประมาณ
μ	$\overline{X}$	1. สุ่มตัวอย่างจากประชากรที่มีการแจกแจงปกติ หรือ 2. สุ่มตัวอย่างจากประชากรที่ไม่ทราบการแจกแจงแต่มีขนาดตัวอย่างใหญ่ ($n \geq 30$) **การแจกแจง** $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$\mu = \overline{X} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
		3. การประมาณค่าเฉลี่ยกรณีไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) **การแจกแจง** $t = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$	$\mu = \overline{X} \pm t_{\frac{\alpha}{2}, n-1} \dfrac{S}{\sqrt{n}}$
π	p	**การแจกแจง** $Z = \dfrac{p - \mu_p}{\sigma_p}$	$\pi = p \pm z_{\frac{\alpha}{2}} \sqrt{\dfrac{p(1-p)}{n}}$

- **<u>ตัวอย่างที่2</u>**
- ค่าใช้จ่ายเฉลี่ยต่อเดือนของนิสิตสำนักฯมีการแจกแจงปกติ ด้วยค่าส่วนเบี่ยงเบนมาตรฐาน 1600 บาท สุ่มตัวอย่างจำนวนนักศึกษา 1000 คน จากมศว ค่าใช้จ่ายโดยเฉลี่ยต่อเดือน 3000 บาท จงประมาณค่าใช้จ่ายเฉลี่ยต่อเดือนของ นิสิตกลุ่มนี้ด้วยความเชื่อมั่น 95 %
- **วิธีทำ** โจทย์ต้องการประมาณค่าเฉลี่ย

กำหนดให้ μ แทนค่าใช้จ่ายเฉลี่ยต่อเดือนของนิสิตกลุ่มนี้

แทนค่าใช้จ่ายเฉลี่ยต่อเดือนของนิสิตในตัวอย่าง

n= 1000 = 3000 σ =1600 α = 0.05

$z_{0.025}$= 1.96

- ขีดจำกัดของการประมาณสำหรับค่า μ

-
- = $$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
- $$= 3000 \pm 1.96\frac{1600}{\sqrt{1000}}$$
- = (2900.83,3099.17)
- $\therefore$95% ช่วงเชื่อมั่นของการประมาณสำหรับค่า μ คือ
- 2900.83$< \mu <$ 3099.17
- ค่าใช้จ่ายโดยเฉลี่ยของนิสิตกลุ่มนี้มีค่าอยู่ระหว่าง 2900.83 บาท และ 3099.17 บาท ด้วยความเชื่อมั่น 95 %

- **ตัวอย่างที่3**
- ผู้วิจัยต้องการทราบว่านิสิตเพศชายมีสัดส่วนเท่าใดในมหาวิทยาลัยเอกชน จึงสุ่มตัวอย่างนิสิตในมหาวิทยาลัยเอกชนจำนวน 423 คน มีนิสิตชาย 196 คน จงประมาณสัดส่วนของนิสิตเพศชายในมหาวิทยาลัยเอกชนด้วยความเชื่อมั่น 95%
- **วิธีทำ** กำหนด p แทนสัดส่วนนิสิตเพศชายในมหาวิทยาลัยเอกชนในตัวอย่าง
- π แทนสัดส่วนนิสิตเพศชายในมหาวิทยาลัยเอกชนในประชากร

$$n = 423 \quad p = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 196/423 = 0.463$$

$$\alpha = 0.05 \quad z_{0.025} = 1.96$$

ขีดจำกัดของการประมาณสำหรับค่า π คือ $p \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$

$$= 0.463 \pm 1.96\sqrt{\frac{0.463(1-0.463)}{423}}$$

$$= (0.415, 0.511)$$

$\therefore$ 95% ช่วงเชื่อมั่นของการประมาณสำหรับค่า π คือ $0.415 < \pi < 0.511$

สัดส่วนของนิสิตเพศชายในมหาวิทยาลัยเอกชน
มีค่าระหว่าง 0.415 และ 0.511 ด้วยความเชื่อมั่น 95 %

ค่าวิกฤต(CRITICAL VALUE) คือค่าที่แบ่งบริเวณที่ใช้ทดสอบออกสองส่วนคือบริเวณการปฏิเสธสมมติฐานหลักกับบริเวณการยอมรับสมมติฐานหลัก

- **การทดสอบแบบด้านเดียวและการทดสอบแบบสองด้าน**
- ให้ θ เป็นพารามิเตอร์ และ k เป็นค่าคงที่ใดๆ
- θ หมายถึง ค่าเฉลี่ย (μ) ค่าสัดส่วน(π) และ ค่าแปรปรวน(σ^2)

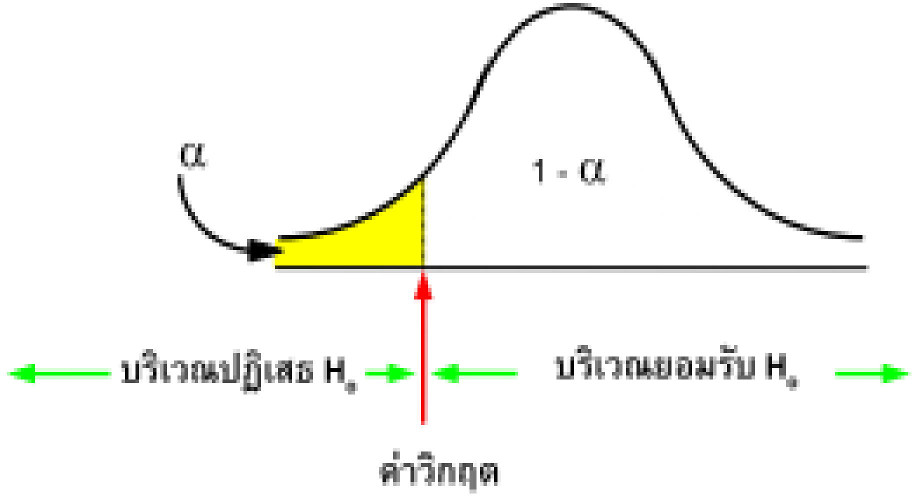
- การทดสอบแบบด้านเดียว

การทดสอบแบบด้านเดียวเกิดขึ้นเมื่อทำการทดสอบสมมติฐานการทดสอบสนใจด้านมากหรือด้านน้อยต่อไปนี้

ก. $H_0 : \theta \geq k$ หรือ $\theta = k$

$H_1 : \theta < k$

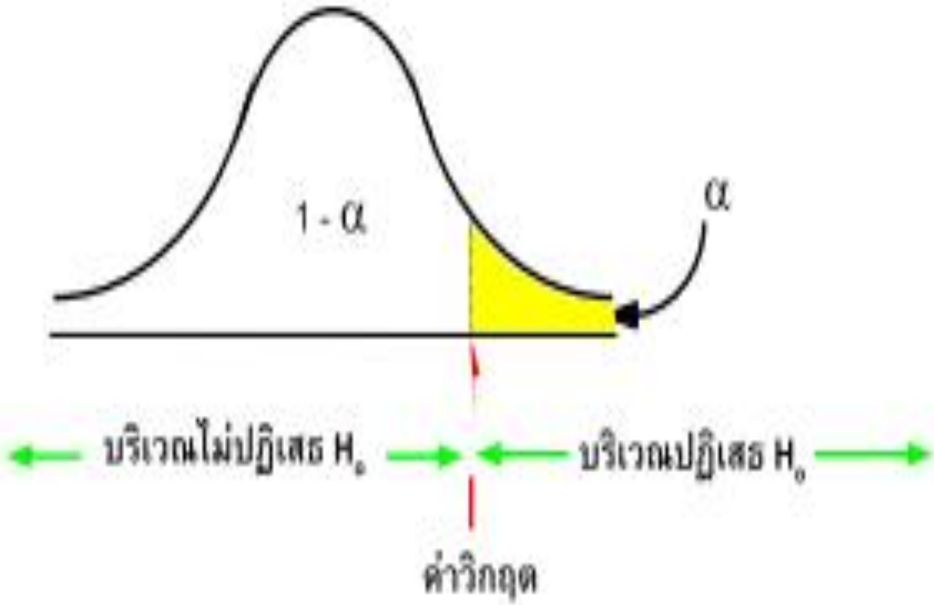
บริเวณวิกฤตหรือบริเวณที่จะปฏิเสธ H_0 จะอยู่ปลายหางด้านซ้ายมือ ซึ่งเรียกว่าเป็นการทดสอบด้านเดียวด้านน้อยดังรูป



ข. $H_0 : \theta \le k$ หรือ $\theta = k$

$H_1 : \theta > k$

บริเวณวิกฤตหรือบริเวณที่จะปฏิเสธ H_0 จะอยู่ปลายหางด้านขวามือ ซึ่งเรียกว่าเป็น



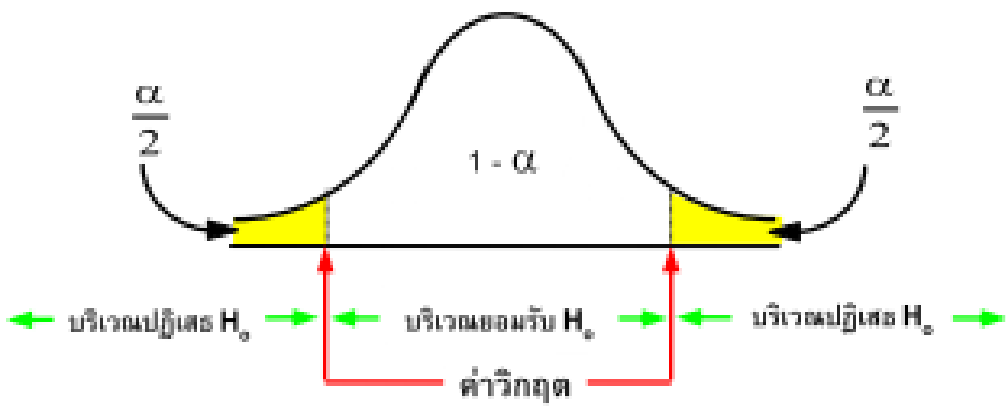
การทดสอบด้านเดียวด้านมาก(Upper-tailed test) ดังรูป

- การทดสอบแบบสองด้าน (two-tailed test)

การทดสอบแบบสองด้าน เป็นการทดสอบสมมติฐานที่สนใจทั้งสองด้านพร้อมกันสมมติฐานคือ

$$H_0 \;:\; \theta = k$$
$$H_1 \;:\; \theta \neq k$$

บริเวณวิกฤตหรือบริเวณที่จะปฏิเสธ H_0 จะอยู่ปลายหางทั้ง 2 ด้านดังรูป



ในการสรุปผลการทดสอบเมื่อคำนวณค่าสถิติที่เก็บจากตัวอย่างให้นำค่าที่ได้มาเปรียบเทียบกับค่าวิกฤตในรูป หากค่าที่ได้ตกบริเวณปฏิเสธ ตัดสินใจปฏิเสธ H_0 หากตกบริเวณยอมรับตัดสินใจยอมรับ H_0

ขั้นตอนการทดสอบสมมติฐาน

1. ตั้งสมมติฐานเราต้องพิจารณาว่าพารามิเตอร์ที่ต้องการทดสอบคือค่าใดเช่นถ้าเป็น ค่าเฉลี่ยแทนด้วย μ ค่าสัดส่วนแทนด้วยπ และ ค่าแปรปรวนแทนด้วยσ^2จากนั้นถามตัวเองว่า ข้อความที่ต้องการทดสอบมีความหมายบ่งบอกการเท่ากันหรือไม่ หากมีความหมายบ่งบอกการเท่ากัน ข้อความที่ต้องการทดสอบต้องตั้งที่ H_0 ส่วนที่เหลือจะเป็นข้อความที่ H_1 หากข้อความนั้น ไม่มีความหมายบ่งบอกการเท่ากันให้ตั้งที่ H_1 ส่วนที่เหลือจะเป็นข้อความที่ H_0

2. กำหนดระดับนัยสำคัญ (α) โดยทั่วไปกำหนดเป็น 0.01,0.05 และ 0.10

3. กำหนดสถิติที่ใช้ทดสอบ การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยและสัดส่วนของประชากรเดียวตัวสถิติที่ใช้ทดสอบคือ Z ,T,χ^2 และ F ซึ่งการที่จะกำหนดตัวสถิตินี้ จะต้องสอดคล้องกับพารามิเตอร์ที่สมมติฐานตั้งขึ้นเพื่อทำการทดสอบและลักษณะของประชากรที่เก็บรวบรวมข้อมูลมา

4. ดูบริเวณการยอมรับและปฏิเสธสมมติฐาน โดยใช้ค่าวิกฤต ค่าวิกฤตที่กำหนดขึ้นมาจะต้องขึ้นกับตัวสถิติที่ใช้ทดสอบ เช่น ตัวสถิติที่ใช้ทดสอบคือ Z ค่าวิกฤตจะได้จากการเปิดตาราง Z และถ้าตัวสถิติที่ใช้ทดสอบคือ t ค่าวิกฤตก็จะได้จากการเปิดตาราง t ค่าวิกฤตจะมีค่าเดียวหรือ 2 ค่าก็ขึ้นอยู่กับการทดสอบว่าเป็นแบบด้านเดียวหรือแบบ 2 ด้าน ซึ่งจะทำให้บริเวณวิกฤตมีด้านเดียวหรือ 2 ด้าน

5. ตัดสินใจและสรุปผล ถ้าค่าสถิติที่คำนวณได้จากตัวอย่างตกอยู่ในบริเวณปฏิเสธจะตัดสินใจว่า ปฏิเสธ H_0 แต่ถ้าค่าสถิติคำนวณได้จากตัวอย่างตกอยู่ในบริเวณยอมรับตัดสินใจว่ายอมรับ H_0

ในการพิจารณานอกเหนือจากการหาค่าวิกฤต โดยใช้ค่าจากตาราง ซึ่งอาจจะเป็นการแจกแจงของ Z, t, χ^2, และการแจกแจงของ F การสรุปผลสามารถใช้ค่า p-value ได้

หากค่าจากตัวอย่างให้ค่าความน่าจะเป็นน้อยกว่าค่า α แสดงว่าค่าจากตัวอย่างยังอยู่ในบริเวณของการปฏิเสธจึงตัดสินใจปฏิเสธ H_0

หากค่าจากตัวอย่างให้ค่าความน่าจะเป็นมากกว่าค่า α แสดงว่าค่าจากตัวอย่างอยู่ในบริเวณของการยอมรับจะตัดสินใจยอมรับ H_0

เกณฑ์การตัดสินใจ ถ้า p- value $< \alpha$ ผลสรุปคือ ปฏิเสธ H_0

ถ้า p- value $\geq \alpha$ ผลสรุปคือ ไม่ปฏิเสธ H_0

การทดสอบสมมติฐานประชากรเดียว

พารามิเตอร์	กรณี	ค่าสถิติทดสอบ
μ	• สุ่มตัวอย่างสุ่มมาจากการแจกแจงแบบปกติและทราบค่าความแปรปรวนของประชากร	$Z = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$
	• ตัวอย่าง สุ่มมาจากประชากรที่มีการแจกแจงแบบใด ๆที่ไม่ทราบค่าความแปรปรวนของประชากรแต่ขนาดตัวอย่างโตพอ(n≥30)	$Z = \dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ โดย ที่ $S^2 = \dfrac{\Sigma(X_i-\overline{X})^2}{n-1}$
	• กรณีที่ตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ ไม่ทราบค่าความแปรปรวนของประชากรและขนาดตัวอย่างที่สุ่มน้อยกว่า 30	$t = \dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ โดยที่ $df = n-1$
π	• กรณีสุ่มตัวอย่างขนาดใหญ่	$Z = \dfrac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$

- **ตัวอย่างที่4**
- เชื่อว่า70% ของผู้ชมที่เป็นชายเห็นด้วยกับ กรรมการตัดสินการประกวดนางงาม ที่ให้ สาวงามจากประเทศไทยได้รับรางวัลชนะเลิศประเภทแต่งกายประจำชาติ จากการสุ่มตัวอย่างชายจำนวน 500 คน มีคนเห็นด้วย 300 คน จงทดสอบสมมติฐานดังกล่าวว่าเป็นจริงหรือไม่ ที่ระดับนัยสำคัญ 0.05
- **วิธีทำ**
- กำหนด p แทน สัดส่วนของผู้ชายที่เห็นด้วยกับกรรมการตัดสินในตัวอย่าง
- π แทน สัดส่วนของผู้ชายที่เห็นด้วยกับกรรมการตัดสินในประชากร

1. ตั้งสมมติฐาน

$H_0 : \quad \pi = 0.7$

$H_1 : \quad \pi \neq 0.7$

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.05$

3. กำหนดตัวสถิติทดสอบและคำนวณค่า

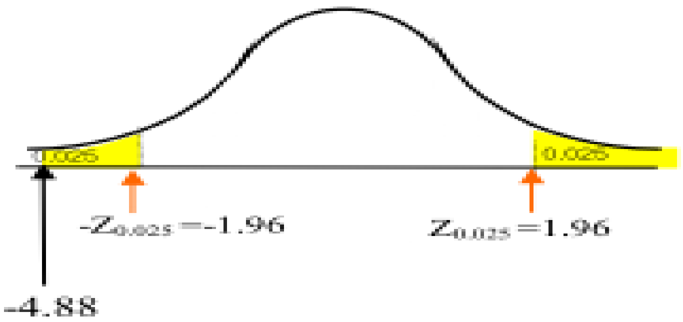
ตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงปกติ ตัวสถิติที่ใช้ในการทดสอบคือ

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

แทนค่า p = 300/500= 0.6

$$Z = \frac{0.6 - 0.7}{\sqrt{\frac{0.7(1-0.7)}{500}}} \qquad = -4.88$$

4. ดูบริเวณการยอมรับและปฏิเสธ H_0 ตั้งสมมติฐานสองด้าน α /2= 0.025



ตัดสินใจปฏิเสธ H_0 สรุปว่าสัดส่วนสัดส่วนของผู้ชายที่เห็นด้วยกับกรรมการตัดสินไม่เป็น70% ที่ระดับนัยสำคัญ 0.05

การทดสอบค่าผลต่างของประชากร 2 กลุ่ม ในการทดสอบสามารถแบ่งได้เป็น 2 กรณีคือ

ประชากรเป็นอิสระต่อกัน

ค่าสถิติ	กรณี	การแจกแจง	หมายเหตุ
$\bar{x}_1 - \bar{x}_2$	1.ประชากรมีการแจกแจงปกติ	$Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$	
	2.ประชากรมีการแจกแจงปกติ แต่ไม่ทราบค่าแปรปรวน 2.1 $\sigma_1^2 \neq \sigma_2^2$	$t = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim t_{(v)}$	$v = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$
	2.2 $\sigma_1^2 = \sigma_2^2$	$t = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} \sim t_{(n_1+n_2-2)}$	$S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
	3.ขนาดตัวอย่างทั้งสอง มีค่าตั้งแต่30	$Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0,1)$	

p_1-p_2	ขนาดตัวอย่างใหญ่	$Z = \dfrac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\dfrac{\pi_1(1-\pi_1)}{n_1} + \dfrac{\pi_2(1-\pi_2)}{n_2}}} \sim N(0,1)$	ถ้าสมมติฐานตั้งว่า $H_0 \quad : \pi_1 - \pi_2 = 0$ หมายถึง $\pi_1 = \pi_2$ และถ้าไม่ทราบค่าπ_1และπ_2ประมาณค่าค่าπ_1และπ_2 ด้วย p_1 และ p_2 p_1 และ p_2 ประมาณด้วย p $= \dfrac{X_1 + X_2}{n_1 + n_2} = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ ค่าสถิติที่ใช้คือ $Z = \dfrac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\dfrac{p(1-p)}{n_1} + \dfrac{p(1-p)}{n_2}}}$

การทดสอบค่าผลต่างของประชากร 2 กลุ่ม

ประชากรเป็นอิสระต่อกัน

ค่าสถิติ	กรณี	การแจกแจง	หมายเหตุ
$\bar{x}_1 - \bar{x}_2$	1.ประชากรมีการแจกแจงปกติ	$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	
	2.ประชากรมีการแจกแจงปกติ แต่ไม่ทราบค่าแปรปรวน 2.1 $\sigma_1^2 \neq \sigma_2^2$	$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(v)}$	$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$
	2.2 $\sigma_1^2 = \sigma_2^2$	$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \sim t_{(n_1+n_2-2)}$	$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
	3.ขนาดตัวอย่างทั้งสอง มีค่าตั้งแต่30	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$	

p_1-p_2	ขนาดตัวอย่างใหญ่	$Z = \dfrac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\dfrac{\pi_1(1-\pi_1)}{n_1} + \dfrac{\pi_2(1-\pi_2)}{n_2}}} \sim N(0,1)$	ถ้าสมมติฐานตั้งว่า $H_0 \quad : \pi_1 - \pi_2 = 0$ หมายถึง $\pi_1 = \pi_2$ และถ้าไม่ทราบค่าπ_1และπ_2ประมาณค่าค่าπ_1และπ_2 ด้วย p_1 และ p_2 p_1 และ p_2 ประมาณด้วย p $= \dfrac{X_1 + X_2}{n_1 + n_2} = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ ค่าสถิติที่ใช้คือ $Z = \dfrac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\dfrac{p(1-p)}{n_1} + \dfrac{p(1-p)}{n_2}}}$

ขั้นตอนในการทดสอบเหมือนกับการทดสอบประชากรเดียว สำหรับตัวสถิติ

บริษัทต้องการทดสอบประสิทธิภาพการใช้เครื่องมือชนิด Aและ B โดยทราบว่าค่าแปรปรวนของประชากรไม่เท่ากัน ทำการสุ่มตัวอย่างดังนี้

$n_1 = 16$ $\bar{x}_1 = 0.41$ $S_1^2 = 0.0025$

$n_2 = 12$ $\bar{x}_2 = 0.38$ $S_2^2 = 0.0004$

จงทดสอบสมมติฐานว่า ประสิทธิภาพของเครื่องมือทั้งสองยี่ห้อแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ 0.05

<u>วิธีทำ</u>

กำหนดให้ μ_1 แทนประสิทธิภาพของเครื่องมือยี่ห้อA

μ_2 แทนประสิทธิภาพของเครื่องมือยี่ห้อ**B**

1. ตั้งสมมติฐาน

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

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

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.05$$

3. กำหนดตัวสถิติทดสอบและคำนวณค่า ตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงปกติ ไม่ทราบค่าแปรปรวนและไม่เท่ากันขนาดตัวอย่างเล็ก ตัวสถิติที่ใช้ในการทดสอบคือ

ทดสอบคือ

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(v)}$$

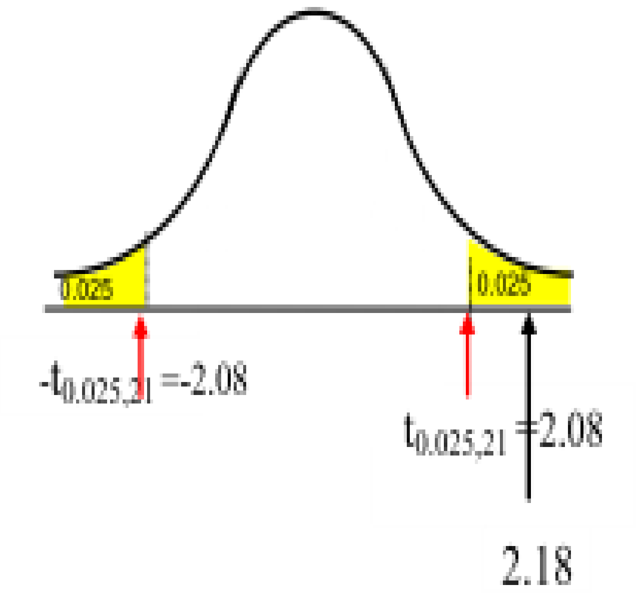
มีการแจกแจงแบบทีด้วยองศาแห่งความเป็นอิสระ v โดยที่

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = 20.8 \cong 21$$

แทนค่า

$$t = \frac{0.41 - 0.38}{\sqrt{\frac{0.0025}{16} + \frac{0.0004}{12}}} = 2.18$$

4. ดูบริเวณการยอมรับและปฏิเสธ H_0 ตั้งสมมติฐานแบบสองด้าน



5.ตัดสินใจปฏิเสธ H_0 สรุปว่าประสิทธิภาพการใช้เครื่องมือยี่ห้อ A และ B แตกต่างกันที่ระดับนัยสำคัญ 0.05

ประชากรไม่เป็นอิสระต่อกัน

ในกรณีที่ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน เช่นทดสอบการทำงานของคู่แฝด การทดลองยาลดความอ้วน ประสิทธิภาพการใช้งานก่อนและหลังการอบรมการใช้ จะสังเกตุเห็นว่าประชากรเป็นประชากร ไม่เป็นอิสระต่อกัน เราเรียกข้อมูลลักษณะนี้ว่าเป็นข้อมูลคู่ (pair data) ลักษณะข้อมูลเป็นดังนี้

เก็บข้อมูลจากประชากรที่มีการแจกแจงปกติมาเป็นคู่ n คู่

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ หาค่าผลต่าง $D_1, D_2, \ldots, D_n$ หาค่า $\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n}$ และ

$$S_D = \sqrt{\frac{\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2}{n-1}} = S_D = \sqrt{\frac{\sum_{i=1}^{n} D_i^2 - n\overline{D}^2}{n-1}}$$

ได้ว่า

ตัวสถิติ $t = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$ จะมีการแจกแจงแบบทีด้วยองศาแห่งความเป็นอิสระ n-1

หากจะประมาณค่าตัวสถิติ μ_D สามารถทำได้เหมือนตัวสถิติอื่นๆในบทที่ 7 โดยเขียนขีดจำกัดการประมาณของ μ_D คือ $\overline{D} \pm t_{\frac{\alpha}{2},n-1} \frac{S_D}{\sqrt{n}}$

$\therefore$ (1-α) 100 % ช่วงของการประมาณค่า μ_D คือ

$$\overline{D} - t_{\frac{\alpha}{2},n-1} \frac{S_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\frac{\alpha}{2},n-1} \frac{S_D}{\sqrt{n}}$$

ถ้าข้อมูลที่สุ่มมาไม่ได้มาจากประชากรที่มีการแจกแจงปกติ และขนาดตัวอย่างมีค่าตั้งแต่ 30 สามารถใช้การแจกแจงปกติประมาณได้

การทดสอบผลต่างค่าสัดส่วนของประชากร

เมื่อต้องการทดสอบว่า สัดส่วนของประชากร 2 กลุ่ม แตกต่างกันหรือไม่ เช่น สัดส่วนของคนที่ชอบใช้บัตรเครดิต ในอาชีพแม่บ้านกับพ่อบ้านแตกต่างกันหรือไม่ เราสามารถศึกษาความแตกต่างของสัดส่วนของประชากร 2 กลุ่ม $(\pi_1 - \pi_2)$ ได้เช่นเดียวกับความแตกต่างของค่าเฉลี่ยประชากร 2 กลุ่ม

ให้ p_1 และ p_2 แทนสัดส่วนของสิ่งที่สนใจศึกษาจากตัวอย่างขนาด n_1 และ n_2 ซึ่งสุ่มมาจากประชากรที่ 1 และ 2 ตามลำดับ

ถ้า π_1 และ π_2 เป็นสัดส่วนของประชากรที่ 1 และที่ 2 ตามลำดับ

ขั้นตอนในการทดสอบสมมติฐาน เหมือนการทดสอบค่าเฉลี่ยสองประชากร

ตัวอย่าง(ลองทำ)

จากการศึกษาอัตราการเสี่ยงชีวิตของยา 2 ชนิดคือ Aspirin และ Placebo สำหรับบุคคลผู้เป็นโรคหัวใจได้ผลตามตารางดังนี้

ผล	Aspirin	Placebo
ตาย	104	189
ขนาด	11037	11034

จงทดสอบว่าสัดส่วนการเสี่ยงของผู้เป็นโรคหัวใจที่ใช้ยา Aspirin น้อยกว่า Placebo ที่ระดับนัยสำคัญ 0.05

กำหนด π_1 เป็นสัดส่วนการตายเนื่องจากใช้ยา Aspirin

กำหนด π_2 เป็นสัดส่วนการตายเนื่องจากใช้ยา Placebo

1.ตั้งสมมติฐาน

$H_0 \quad : \pi_1 - \pi_2 = 0$

$H_1 \quad : \pi_1 - \pi_2 < 0$

2 . กำหนดระดับนัยสำคัญ $\alpha = 0.05$

3 . กำหนดตัวสถิติทดสอบ และคำนวณค่าตัวสถิติทดสอบนั้น

$$p_1 = \frac{104}{11037} = 0.0094 \text{ และ } p_2 = \frac{189}{11034} = 0.0171$$

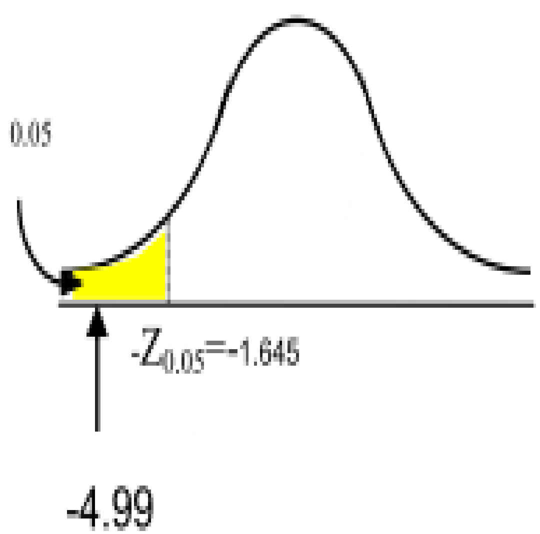
$$\text{ใช้ } p = \frac{104+189}{11037+11034} = 0.0133$$

ตัวสถิติทดสอบ

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}} = \frac{(0.0094 - 0.0171) - 0}{\sqrt{\frac{0.0133(1-0.0133)}{11037} + \frac{0.0133(1-0.0133)}{11037}}}$$

$= -4.99$

4. ดูบริเวณการยอมรับและปฏิเสธ H_0



5. ตัดสินใจ ปฏิเสธ H_0

สรุปว่าสัดส่วนการเสี่ยงของผู้เป็นโรคหัวใจที่ใช้ยา Aspirin น้อยกว่า Placebo ที่ระดับนัยสำคัญ 0.05