บทที่ 6

[illegible]

ในการศึกษาสถิติอ้างอิงหรือสถิติอนุมาน (Inference statistics) นั้น จุดมุ่งหมายที่สำคัญก็คือ การนำเอาข้อมูลที่ได้จากตัวอย่างไปอ้างอิง หรือช่วยในการตัดสินใจ หรือสรุปผลเกี่ยวกับคุณลักษณะของประชากรที่เราไม่ทราบค่า เพราะว่าเราไม่สามารถที่จะศึกษาจากทุก ๆ หน่วยของประชากรได้ เนื่องจากมีทรัพยากรจำกัด

การแจกแจงของตัวสถิติเป็นการศึกษาความสัมพันธ์ระหว่างค่าที่ได้จากตัวอย่างกับค่าของประชากร ดังนั้น ถ้าเราสามารถเลือกตัวอย่างที่เป็นตัวแทนที่เหมาะสมและแสดงคุณลักษณะต่าง ๆ ของประชากรได้อย่างแท้จริง เราก็สามารถนำเอาค่าที่ได้จากการศึกษาตัวอย่างไปใช้สรุปผลเกี่ยวกับค่าของประชากรได้

ซึ่งในการศึกษาสถิติอ้างอิงในบทนี้จะกล่าวถึงการแจกแจงของตัวสถิติ การประมาณค่า การทดสอบสมมติฐาน และการศึกษาความสัมพันธ์ระหว่างตัวแปร โดยจะขอเริ่มที่การแจกแจงของตัวสถิติก่อนดังนี้

6.1 การแจกแจงของตัวสถิติ

การแจกแจงของตัวสถิติหรือเรียกอีกอย่างว่าการแจกแจงของการสุ่มตัวอย่าง (Sampling Distribution) ในการเลือกตัวอย่างจากประชากรนั้น ถือว่าเป็นการทดลองเชิงสุ่มอย่างหนึ่ง ซึ่งมี Sample Space ที่ประกอบด้วยตัวอย่างที่จะเป็นไปได้ทั้งหมดตามขนาดที่กำหนดไว้ โดยเลือกมาจากประชากรที่เราสนใจ ตัวสถิติที่ได้มาจากตัวอย่างแต่ละชุดก็จะเป็นตัวแปรเชิงสุ่มตามตัวอย่างด้วย และมีค่าตามตัวอย่างหรือผลการทดลองที่กำหนด ดังนั้น ตัวสถิติจึงมีการแจกแจงอย่างหนึ่ง และการแจกแจงของตัวสถิตินี้เราเรียกว่าการแจกแจงของการสุ่มตัวอย่าง (Sampling Distribution) ซึ่งในวิชา SC 101 จะกล่าวถึงการแจกแจงของการสุ่มตัวอย่างของ

ค่าเฉลี่ย (Sampling Distribution of the Mean) การแจกแจงของการสุ่มตัวอย่างของสัดส่วน (Sampling Distribution of Proportion) และการแจกแจงของการสุ่มตัวอย่างของความแปรปรวน (Sampling Distribution of Sample Variance)

6.1.1 การแจกแจงของการสุ่มตัวอย่างของค่าเฉลี่ย (Sampling Distribution of the Mean)

จากประชากรขนาด N ซึ่งมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ สุ่มตัวอย่างมาขนาด n ซึ่งอาจจะสุ่มแบบแทนที่ (Sampling with replacement) หรือสุ่มแบบไม่แทนที่ (Sampling without replacement) ก็ได้จากตัวอย่าง แต่ละแบบที่สุ่มได้นำมาคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานได้ ซึ่งจะแทนด้วย $\bar{x}$ และ s ตามลำดับ

สมมติว่าเลือกตัวอย่างได้ k แบบ แต่ละแบบก็คำนวณหา $\bar{x}$ ออกมา ดังนั้น จะมี $\bar{x}$ อยู่ k ตัว คือ $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ ซึ่ง $\bar{x}$ เหล่านี้เราเรียกว่า Sample Mean การศึกษาการแจกแจงของ $\bar{x}$ ก็คือ ดูว่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ $\bar{x}$ เหล่านี้ เป็นอย่างไร มีความสัมพันธ์กับค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของประชากรอย่างไรบ้าง โดยใช้ $\mu_{\bar{x}}$ แทนค่าเฉลี่ยของ $\bar{x}$ และ $\sigma_{\bar{x}}$ แทน ความคลาดเคลื่อนมาตรฐานของ $\bar{x}$ ซึ่งการคำนวณหา $\mu_{\bar{x}}$ และ $\sigma_{\bar{x}}$ หาได้ ดังนี้

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_k}{k}$$

หรือ
$$\boxed{\mu_{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}} \quad (1)$$

และ
$$\sigma_{\bar{x}} = \sqrt{\frac{(\bar{x}_1 - \mu_{\bar{x}})^2 + (\bar{x}_2 - \mu_{\bar{x}})^2 + \ldots + (\bar{x}_k - \mu_{\bar{x}})^2}{k}}$$

หรือ
$$\boxed{\sigma_{\bar{x}} = \sqrt{\frac{\Sigma (\bar{x}_i - \mu_{\bar{x}})^2}{k}}} \quad (2)$$

การหาค่าของ $\mu_{\bar{x}}$ และ $\sigma_{\bar{x}}$ เป็นงานที่ยุ่งยาก ซึ่งในที่นี้จะไม่พิสูจน์ให้ดู แต่จะแสดงให้เห็นด้วยตัวอย่างดังต่อไปนี้ จาก ① จะได้ว่า $\mu_{\bar{x}} = \mu$ ซึ่งหมายความว่าค่าเฉลี่ยของ $\bar{x}$ จะมีค่าเท่ากับค่าเฉลี่ยของประชากร และ จาก ② จะได้ว่า

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$ ในกรณีที่ประชากรเป็นแบบจำกัด และการเลือกตัวอย่างเป็นแบบไม่แทนที่

และ $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$ ในกรณีที่ประชากรเป็นแบบไม่จำกัด หรือเป็นแบบจำกัดและการเลือกตัวอย่างเป็นแบบแทนที่

ตัวอย่าง 6.1 มีเลขอยู่ 4 จำนวน คือ 4, 8, 13 และ 11 ต้องการสุ่มอย่างขนาด $n = 2$ แบบแทนที่ และแบบไม่แทนที่ จงแสดงว่า

1. $\mu_{\bar{x}} = \mu$
2. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ เมื่อเลือกตัวอย่างแบบไม่แทนที่
3. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ เมื่อเลือกตัวอย่างแบบแทนที่

วิธีทำ

จากประชากร 4, 8, 13, 11

หาค่าเฉลี่ย $\bar{x}$ และความแปรปรวน σ^2 ของประชากรได้ดังนี้

$$\mu = \frac{4+8+13+11}{4} = \frac{36}{4} = 9$$

$$\sigma^2 = \frac{(4-9)^2+(8-9)^2+(13-9)^2+(11-9)^2}{4} = \frac{25+1+16+4}{4} = \frac{46}{4} = \frac{23}{2}$$

1. ต้องการแสดงให้เห็นว่า $\mu_{\bar{x}} = \mu$

ก. เลือกตัวอย่างแบบไม่แทนที่ เราจะได้จำนวนตัวอย่างทั้งหมด ${}^{N}C_{n}$ ซึ่งเท่ากับ ${}^{4}C_{2} = \frac{4!}{2!2!}$ $= 6$ ตัวอย่างซึ่งได้แก่ (4,8), (4,13), (4,11), (8,13), (8,11) และ (13,11) เราต้องหาค่าเฉลี่ยของตัวอย่างแต่ละตัวอย่างซึ่งจะได้ค่าเฉลี่ยของตัวอย่างทั้งหมด 6 ค่าด้วยกัน ดังนี้

ตัวอย่าง (4,8) ได้ $\bar{x}_1 = \frac{4+8}{2} = 6$

ตัวอย่าง (4,13) ได้ $\bar{x}_2 = \frac{4+13}{2} = 8.5$

ตัวอย่าง (4,11) ได้ $\bar{x}_3 = \frac{4+11}{2} = 7.5$

ตัวอย่าง (8,13) ได้ $\bar{x}_4 = \frac{8+13}{2} = 10$

ตัวอย่าง (8,11) ได้ $\bar{x}_5 = \frac{8+11}{2} = 9.5$

ตัวอย่าง (13,11) ได้ $\bar{x}_6 = \frac{13+11}{2} = 12$

$$\therefore \quad \mu_{\bar{x}} = \frac{\bar{x}_1+\bar{x}_2+\dots\dots+\bar{x}_6}{6}$$

$$= \frac{6+8.5+7.5+10.5+9.5+12}{6}$$

$$= \frac{54}{6} = 9$$

$$= \mu$$

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

ข. เลือกตัวอย่างแบบแทนที่ จำนวนตัวอย่างที่ได้จะเท่ากับ N^{n} ตัวอย่างซึ่งเท่ากับ $4^{2} = 16$ ตัวอย่างซึ่งได้แก่

(4,4), (4,8), (4,13), (4,11)

(8,4), (8,8), (8,13), (8,11)

(13,4), (13,8), (13,13), (13,11)

(11,4), (11,8), (11,13), (11,11)

$\therefore$ ค่าเฉลี่ยของแต่ละตัวอย่างคือ

$$\bar{x}_1 = \frac{4+4}{2} = 4$$

$$\bar{x}_2 = \frac{4+8}{2} = 6$$

$$\bar{x}_3 = \frac{4+13}{2} = 8.5$$

$$\bar{x}_4 = \frac{4+11}{2} = 7.5$$

....

....

$$\bar{x}_{16} = \frac{11+11}{2} = 11$$

$$\therefore \mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \ldots\ldots + \bar{x}_{16}}{16}$$

$$= \frac{4+6+8.5+7.5+ \ldots\ldots + 11}{16}$$

$$= \frac{144}{16} = 9 = \mu$$

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

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

2. ต้องการแสดงให้เห็นว่า $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ เมื่อเลือกตัวอย่างแบบไม่แทนที่ ดังนั้น ตัวอย่างที่เลือกได้เท่ากับ ${}^{N}C_{n}$ ซึ่งเท่ากับ ${}^{4}C_{2} = 6$ แบบ

$$\therefore \sigma^2_{\bar{x}} = \frac{(6-9)^2+(8.5-9)^2+(7.5-9)^2+(10.5-9)^2+(9.5-9)^2+(12-9)^2}{6}$$

$$= \frac{(-3)^2+(-.5)^2+(-1.5)^2+(1.5)^2+(.5)^2+(3)^2}{6}$$

$$= \frac{9+.25+2.25+2.25+.25+9}{6}$$

$$= \frac{23}{6}$$

ซึ่ง ถ้าพิจารณา $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ หรือ $\frac{\sigma^2}{\sqrt{n}}[\frac{N-n}{N-1}]$ จะได้

$$\frac{\sigma^2}{n}[\frac{N-n}{N-1}] = \frac{23}{2} \times \frac{1}{2} [\frac{4-2}{4-1}]$$

$$= \frac{23}{4} \times \frac{2}{3} = \frac{23}{6} = \sigma^2_{\bar{x}}$$

แสดงว่า $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

3. ต้องการแสดงให้เห็นว่า $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ เมื่อเลือกตัวอย่างแบบแทนที่คำนวณตัวอย่างที่เลือกได้เท่ากับ $N^n = 4^2 = 16$ ตัวอย่าง

$$\therefore \quad \sigma^2_{\bar{x}} = \frac{1}{16} \left[(4-9)^2 + (6-9)^2 + \ldots\ldots + (11-9)^2 \right]$$

$$= \frac{1}{16} \left[25 + 9 + \ldots\ldots + 4 \right] = \frac{92}{16} = \frac{23}{4}$$

ซึ่งถ้าพิจารณา $\frac{\sigma}{\sqrt{n}}$ หรือ $\frac{\sigma^2}{n}$ จะได้

$$\frac{\sigma^2}{n} = \frac{23}{2} \times \frac{1}{2} = \frac{23}{4}$$

แสดงว่า $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

จากตัวอย่างข้างต้นนี้จะได้ทฤษฎีดังต่อไปนี้

ทฤษฎี 6.1

ถ้า $\bar{X}$ คือ ค่าเฉลี่ยของตัวอย่างสุ่มขนาด n จากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้ว เมื่อ n มีขนาดโต $(n \to \infty)$ ตัวสถิติ $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ จะมีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1

ทฤษฎี 6.2

ถ้า X มีการแจกแจงที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ และถ้าเลือกตัวอย่างขนาด n จากประชากรนี้ จะได้ $\bar{X}$ มีการแจกแจงเข้าใกล้การแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเป็น $\mu_{\bar{x}}$ และความคลาดเคลื่อนมาตรฐานเป็น $\sigma_{\bar{x}}$ นั่นคือ $Z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ จะเป็นตัวแปรเชิงสุ่มแบบปกติมาตรฐาน N (0,1) โดยที่ $\mu_{\bar{x}} = \mu$ และ

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad \text{หรือ} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

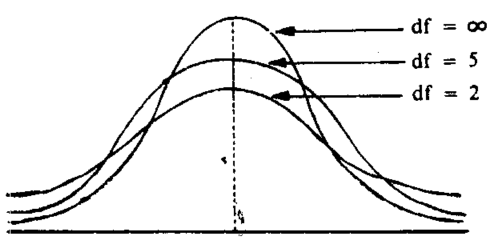
สำหรับการหาความน่าจะเป็นที่ตัวแปรเชิงสุ่มจะมีค่าต่าง ๆ นั้น ก็ใช้วิธีการคำนวณหาแบบเดียวกับการหาความน่าจะเป็นของการแจกแจงแบบปกติ

โดยปกติแล้วมักจะไม่ทราบความแปรปรวนของประชากรที่เราเลือกตัวอย่างสุ่มมา สำหรับตัวอย่างที่มีขนาดโต ($n \geq 30$) จะใช้ความแปรปรวนของตัวอย่าง (s^2) มาเป็นตัวประมาณค่า σ^2

$$\therefore \quad \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

ซึ่ง $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ ก็ยังคงมีการแจกแจงแบบปกติมาตรฐาน แต่ถ้าตัวอย่างที่สุ่มมามีขนาดเล็ก ($n < 30$) $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ จะมีการแจกแจงแบบที (Student's t distribution) ซึ่งการแจกแจงของ $t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ นี้จะคล้ายกับการแจกแจงของ z คือลักษณะของโค้งจะเป็นรูประฆังคว่ำ และมีสมมาตรรอบค่าเฉลี่ยซึ่งมีค่าเป็นศูนย์ แต่ t มีความแปรปรวนมากกว่า ซึ่งทำให้โค้งของ t แบนกว่า และค่าของ t ขึ้นอยู่กับค่าเฉลี่ยของตัวอย่าง ($\bar{X}$) และความแปรปรวนของตัวอย่าง (s^2) แต่ค่าของ z ขึ้นอยู่กับเฉลี่ยของตัวอย่างเพียงค่าเดียวเท่านั้น แต่อย่างไรก็ตาม เมื่อตัวอย่างมีขนาดโต ($n \geq 30$) การแจกแจงของ z และ t จะไม่แตกต่างกันเลย

จากความแปรปรวนของตัวอย่าง $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ ค่า $n-1$ นี้คือองศาความเป็นอิสระ (Degree of freedom) มักใช้ตัวย่อว่า df หรือ v ดังนั้น การแจกแจงของ t จะมีรูปร่างต่างกันถ้าองศาความเป็นอิสระต่างกัน ดังรูป 6.1



รูป 6.1

รูปแสดงการแจกแจงเป็น t เมื่อค่า df ต่างกัน

ทฤษฎี 6.3

ถ้า $\bar{x}$ และ s^2 เป็นค่าเฉลี่ยและความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ เมื่อไม่ทราบ σ^2 ดังนั้น

$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ จะมีการแจกแจงแบบที ที่มีองศาความเป็นอิสระ (df) เท่ากับ n-1

6.1.2 การแจกแจงของการสุ่มตัวอย่างของสัดส่วน (Sampling Distribution of Proportion)

จากการทดลองแบบทวินามที่ประกอบด้วย n การทดลองที่เป็นอิสระกันโดยมีความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ π และความน่าจะเป็นที่จะเกิดความไม่สำเร็จเท่ากับ $1-\pi$

ถ้า x เป็นจำนวนครั้งที่เกิดความสำเร็จในการทดลอง n ครั้ง x จะเป็นตัวแปรเชิงสุ่มที่มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติที่มีค่าเฉลี่ย และความแปรปรวนเป็นดังนี้ คือ

$$\mu = n\pi \quad \text{และ} \quad \sigma^2 = n\pi(1-\pi)$$

ซึ่งถ้าพิจารณาสัดส่วนที่จะเกิดความสำเร็จในการทดลอง n ครั้ง โดยให้ $p = \frac{x}{n}$ (สัดส่วนที่สำเร็จ) และจะได้ P เป็นตัวแปรเชิงสุ่มที่มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติโดยมีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$\mu_p = \pi \quad , \quad \sigma^2_p = \frac{\pi(1-\pi)}{n}$$

$$\therefore \quad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

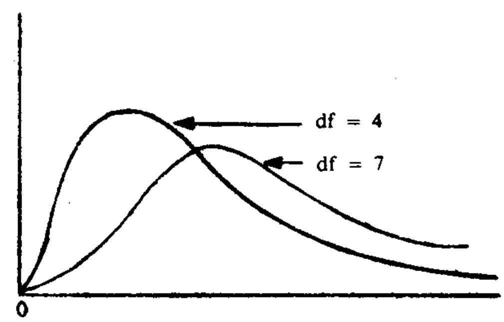
ดังนั้น $Z = \frac{P-\pi}{\sigma_p} = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ จะมีการแจกแจงแบบปกติมาตรฐาน N(0,1)

ทฤษฎี 6.4

ตัวอย่างสุ่มขนาด n สุ่มมาจากประชากรที่มีการแจกแจงแบบทวินามที่มีค่าเฉลี่ย $n\pi$ และความแปรปรวน $n\pi(1-\pi)$ สัดส่วนที่เกิดความสำเร็จ $(P) = \frac{x}{n}$ จะมีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติที่มีค่าเฉลี่ย π และความแปรปรวน $\frac{\pi(1-\pi)}{n}$ ดังนั้น $Z = \dfrac{P - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$ จะมีการแจกแจงแบบปกติมาตรฐาน $N(0,1)$

6.1.3 การแจกแจงการสุ่มตัวอย่างของความแปรปรวน (Sampling Distribution of Sample Variance)

ถ้า S^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากรแบบปกติที่มีความแปรปรวน σ^2 แล้วตัวแปรเชิงสุ่ม $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ จะมีการแจกแจงแบบไคสแควร์ (Chi-square Distribution) ที่มีองศาความเป็นอิสระ $n-1$ การแจกแจงของ χ^2 จะเบ้ไปทางขวา ดังนั้นค่าของ χ^2 จะเป็นบวกเสมอและการแจกแจงของ χ^2 จะไม่สมมาตรกันรอบค่าศูนย์ ค่าของ χ^2 จะมีค่าตั้งแต่ 0 ถึง ∞ และรูปร่างของ χ^2 จะขึ้นอยู่กับขนาดของตัวอย่างหรือ df เช่นเดียวกับการแจกแจงของ t ดังรูป 6.2



รูป 6.2
รูปแสดงการแจกแจงแบบ χ^2 ที่มี df ต่างกัน

6.2 การประมาณค่า (Estimation)

การประมาณค่า คือกระบวนการใช้ข้อมูลจากตัวอย่างไปประมาณค่าพารามิเตอร์ประชากรที่ไม่ทราบค่า เช่น ใช้ข้อมูลที่ได้จากตัวอย่างไปประมาณค่าเฉลี่ยของประชากร หรืออัตราส่วนของประชากรซึ่งเราไม่ทราบค่า แต่เราสามารถประมาณค่าเหล่านี้ได้ โดยการสุ่มตัวอย่างมาจำนวนหนึ่ง แล้วใช้ค่าจากตัวอย่างสุ่มไปประมาณค่าของประชากรโดยอาศัยทฤษฎีการแจกแจงของการสุ่มตัวอย่าง

พารามิเตอร์ (Parameter) หมายถึง ค่าคงที่ที่แสดงคุณลักษณะของประชากรซึ่งเราไม่ทราบค่าที่แท้จริง ซึ่งเขียนแทนด้วย θ

ตัวประมาณค่า (Estimator) เป็นฟังก์ชันของค่าสังเกตในตัวอย่าง ซึ่งใช้เป็นตัวประมาณค่าพารามิเตอร์ เช่น $\bar{X}$, P, S^2 เป็นต้น

ค่าประมาณ (Estimate) เป็นค่าที่จะเป็นไปได้ของตัวประมาณค่าเขียนแทนด้วย $\hat{\theta}$ และแบ่งเป็น 2 ชนิด คือ

ก. ค่าประมาณแบบจุด (Point estimate) เป็นค่าประมาณของประชากรซึ่งได้จากตัวอย่างและเป็นเลขจำนวนเดียว

ข. ค่าประมาณแบบช่วง (Interval estimate) เป็นพิสัยหรือช่วงที่สร้างขึ้นรอบ ๆ ค่าประมาณแบบจุดด้วยความเชื่อมั่นที่เรากำหนดขึ้นไว้

6.2.1 การประมาณค่าแบบจุด (Point estimation) เป็นการใช้ ค่า 1 ค่า จากตัวอย่างเป็นค่าประมาณพารามิเตอร์ ซึ่งการประมาณแบบนี้มีโอกาสผิดพลาดได้มาก

ในการเลือกตัวประมาณค่าแบบจุดมีหลักเกณฑ์ในการเลือกดังนี้ คือ

ก. ต้องเป็นตัวประมาณค่าที่ไม่มีความเอียงเฉ (Unbiased estimator) ถ้าให้ $\hat{\theta}$ เป็นตัวประมาณค่าของพารามิเตอร์ θ $\hat{\theta}$ จะเป็นตัวประมาณค่าที่ไม่มีความเอียงเฉ ถ้าค่าคาดหวังของ $\hat{\theta}$ ที่คำนวณได้จากตัวอย่างสุ่มมีค่าเท่ากับค่าพารามิเตอร์ θ นั่นคือ

$$E(\hat{\theta}) = \theta$$

ตัวอย่างเช่น $E(\bar{X}) = \mu$ ดังนั้น $\bar{X}$ จึงเป็น Unbiased estimator ของ μ

ข. ต้องเป็นตัวประมาณที่มีประสิทธิภาพ (Efficient estimator) ถ้าให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าที่ไม่มีความเอียงเฉของพารามิเตอร์ θ และความแปรปรวนของ $\hat{\theta}_1$ น้อยกว่าความแปรปรวนของ $\hat{\theta}_2$ เราจะได้ $\hat{\theta}_1$ เป็นตัวประมาณค่าที่ดีที่สุดของพารามิเตอร์ θ เพราะว่า $\hat{\theta}_1$ มีความแปรปรวนต่ำที่สุด (Minimum Variance)

ดังนั้น ในบรรดาตัวประมาณค่าที่ไม่มีความเอียงเฉทั้งหลายของพารามิเตอร์ θ ตัวประมาณค่าที่มีความแปรปรวนต่ำสุดจะเป็นตัวประมาณค่าที่ดีที่สุด

6.2.2 การประมาณค่าแบบช่วง (Interval Estimation) ถ้าให้ $(1-\alpha)$ เป็นค่าของความน่าจะเป็นที่มีค่าสูง L และ U เป็นฟังก์ชันของค่าสังเกต $x_1, x_2, \ldots, x_n$ ซึ่งทำให้ $P[L<\theta<U] = (1-\alpha)$ แล้วเราจะเรียก interval (L,U) ว่า 100 $(1-\alpha)$% ช่วงของความเชื่อมั่น (Confidence interval) ของพารามิเตอร์ θ และเรียก $(1-\alpha)$ ว่าเป็นระดับของความเชื่อมั่นของ interval นี้ นั่นคือ $P[L<\theta<U] = 1-\alpha$ คือความน่าจะเป็นที่ random interval (L,U) จะรวมค่าของพารามิเตอร์ θ อยู่ด้วยเป็น $(1-\alpha)$

ดังนั้น การประมาณค่าแบบนี้จะให้ค่าประมาณเป็นช่วงและสามารถบอกความน่าจะเป็นที่ช่วงของการประมาณจะคลุมค่าของพารามิเตอร์ด้วย ซึ่งก็ยังต้องอาศัยค่าของการประมาณแบบจุดเป็นหลัก การประมาณแบบนี้จะผิดพลาดน้อยกว่าแบบแรกและการประมาณค่าแบบช่วงนี้เราใช้ความรู้เกี่ยวกับการแจกแจงการสุ่มตัวอย่างเข้ามาช่วยในการประมาณค่า สำหรับในวิชา SC 101 นี้จะกล่าวถึงการประมาณค่าแบบช่วงของค่าเฉลี่ยประชากร การประมาณค่าแบบช่วงของสัดส่วนประชากร และการประมาณค่าแบบช่วงของความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานประชากร ดังนี้

ก. การประมาณค่าแบบช่วงของค่าเฉลี่ยประชากร เราใช้ $\bar{x}$ เป็นตัวประมาณค่าเฉลี่ยของประชากร (μ) เนื่องจาก $\bar{x}$ เป็น unbiased estimator และมีความแปรปรวนต่ำสุด การประมาณค่าเฉลี่ยของประชากรทำได้ 2 กรณี ดังนี้

1. ถ้าประชากรมีการแจกแจงแบบปกติ และทราบความแปรปรวนของประชากร (σ^2) หรือเมื่อไม่ทราบความแปรปรวนของประชากร แต่ตัวอย่างที่สุ่มมามีขนาดโต $(n \geq 30)$ ช่วงเชื่อมั่น $100(1-\alpha)$% ของค่าเฉลี่ยประชากร (μ) คือ

$$\boxed{\bar{X} - Z_{\alpha/2} \cdot \sigma/\sqrt{n} \quad \leqslant \mu \leqslant \quad \bar{X} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}}$$ (เมื่อทราบ σ^2)

$$\boxed{\bar{X} - Z_{\alpha/2} \cdot S/\sqrt{n} \quad \leqslant \mu \leqslant \quad \bar{X} + Z_{\alpha/2} \cdot S/\sqrt{n}}$$ (เมื่อไม่ทราบ σ^2 แต่ $n \geqslant 30$)

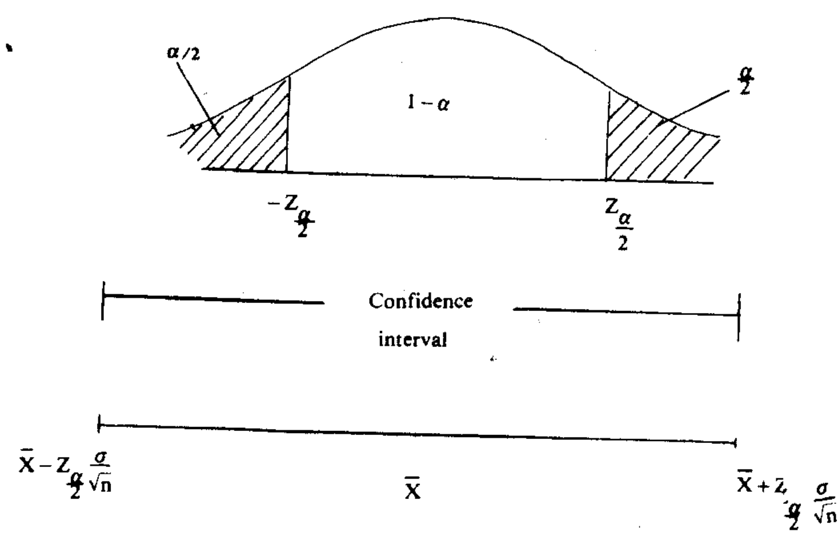
เมื่อ $\bar{x}$ เป็นค่าเฉลี่ยของตัวอย่าง

σ เป็นส่วนเบี่ยงเบนมาตรฐานของประชากร

S เป็นส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

n เป็นขนาดตัวอย่าง

$Z_{\alpha/2}$ เป็นค่าจากตารางปกติที่ทำให้เกิดพื้นที่ทางขวามือเป็น $\alpha/2$



รูป 6.3

ตัวอย่าง 6.2 จะหา Confidence interval ของ μ เมื่อทราบส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 3 ค่าเฉลี่ยของตัวอย่าง ($\bar{X}$) เท่ากับ 24.2 และขนาดตัวอย่าง (n) เท่ากับ 36

ระดับความเชื่อมั่นที่ต้องการ	Z	สูตร	คำนวณ	ช่วงที่ได้
90%	1.65	$\bar{X} \pm 1.65\ \sigma/\sqrt{n}$	$24.2 \pm 1.65 \frac{3}{\sqrt{36}}$ $= 24.2 \pm .825$	23.375 ถึง 25.025
95%	1.96	$\bar{X} \pm 1.96\ \sigma/\sqrt{n}$	$24.2 \pm 1.96 . \frac{3}{\sqrt{36}}$ $= 24.2 \pm .980$	23.220 ถึง 25.180
99%	2.58	$\bar{X} \pm 2.58\ \sigma/\sqrt{n}$	$24.2 \pm 2.58 . \frac{3}{\sqrt{36}}$ $= 24.2 \pm 1.290$	23.110 ถึง 25.690

จากตัวอย่างข้างต้นจะเห็นได้ว่า ถ้าสัมประสิทธิ์แห่งความเชื่อมั่นสูง ช่วงแห่งความเชื่อมั่นจะกว้าง แต่ถ้าสัมประสิทธิ์แห่งความเชื่อมั่นต่ำ ช่วงแห่งความเชื่อมั่นจะแคบ นอกจากนี้ยังมีตัวประกอบ (factors) อื่น ๆ อีกที่มีผลทำให้ช่วงแห่งความเชื่อมั่นแคบหรือกว้าง ซึ่งได้แก่ขนาดตัวอย่างและส่วนเบี่ยงเบนมาตรฐานของประชากร ดังนี้

ก. ผลของ Confidence coefficient

Confidence	Z	Width of interval
68%	1.00	
95%	1.96	
99%	2.58	

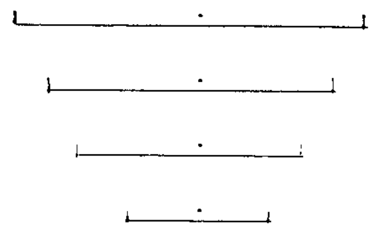
ข. ผลของขนาดตัวอย่าง ขนาดตัวอย่าง Width of interval

8

16

32

64



ค. ผลของส่วนเบี่ยงเบน

มาตรฐานของประชากร

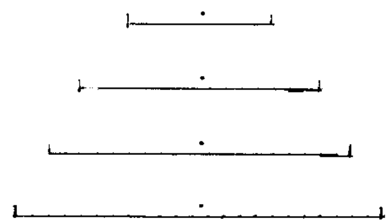
(σ) σ Width of interval

5

10

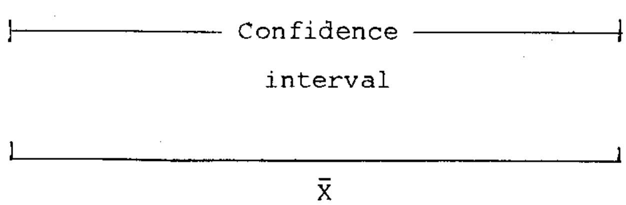
15

20



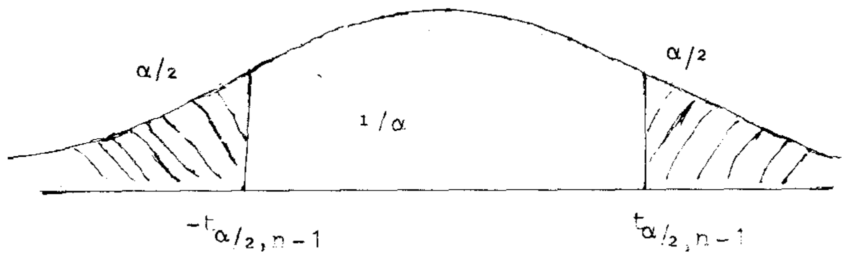
2. ถ้าประชากรมีการแจกแจงแบบปกติไม่ทราบความแปรปรวนของประชากร (σ^2) แต่ตัวอย่างที่สุ่มมามีขนาดเล็ก ($n < 30$) ช่วงเชื่อมั่น $100(1-\alpha)\%$ ของค่าเฉลี่ยประชากร (μ) คือ

$$\boxed{\bar{X} - t_{\alpha/2,n-1} \, S/\sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2,n-1} \, S/\sqrt{n}}$$



$\bar{X} - t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}$ $\bar{X} + t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}$

รูป 6.4



รูป 6.5

เมื่อ $\bar{x}$, s และ n เป็นค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและขนาดตัวอย่างของตัวอย่าง ตามลำดับ $t_{\alpha/2},{n-1}$ เป็นค่าจากตาราง ที่มีองศาความเป็นอิสระ (df) = n-1 ที่ทำให้พื้นที่ทางขวามือเป็น $\alpha/2$

ตัวอย่าง 6.3 ต้องการหา Confidence interval ของ μ โดยใช้ t-distribution เมื่อกำหนดให้ sample mean ($\bar{x}$) = 20 sample standard deviation (S) = 1.5 และ sample size (n) = 25

$$\therefore \ df = 25-1 = 24$$

ดังนั้น ช่วงเชื่อมั่นของ μ ตามระดับความเชื่อมั่นต่าง ๆ จะเป็นดังนี้

ระดับความเชื่อมั่นที่ต้องการ	$t_{\alpha/2},{n-1}$	สูตร	คำนวณ	ช่วงที่ได้
90%	1.711	$\bar{x} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$	$20 \pm 1.711\frac{15}{\sqrt{25}}$ = 20~0.5133	19.49 ถึง 20.51
95%	2.064	$\bar{x} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$	$20 \pm 2.064\frac{1.5}{\sqrt{25}}$ = 20 ± 0.6192	19.38 ถึง 20.62
99%	2.797	$\bar{x} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$	$20 \pm 2.797\frac{1.5}{\sqrt{25}}$ = 20f0.8391	19.16 ถึง 20.84

ข. การประมาณค่าแบบช่วงของสัดส่วนประชากร (π) สำหรับตัวอย่างขนาดโต

ตัวประมาณค่าสัดส่วนของประชากร (π) คือ p โดยที่

$\pi = \frac{A}{N}$ เมื่อ A คือจำนวนหน่วยที่สนใจในประชากร N

และ $p = \frac{a}{n}$ เมื่อ a คือจำนวนหน่วยที่สนใจในตัวอย่าง n

ถ้าสุ่มตัวอย่างขนาด $n(n \geq 30)$ จากประชากรที่ไม่ทราบสัดส่วน (π) ที่แท้จริง แล้ว $100(1-\alpha)\%$ ช่วงเชื่อมั่นของ π คือ

$$\boxed{P - Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} \leq \pi \leq P + Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}}$$

เมื่อ $Z_{\alpha/2}$ คือค่าที่ได้จากตาราง Z ที่ทำให้พื้นที่ทางขวามือเท่ากับ $\alpha/2$

ตัวอย่าง 6.4 ต้องการประมาณช่วงเชื่อมั่น 95% สำหรับสัดส่วนที่แท้จริง (π) ของครอบครัวที่ใช้รถ Honda ถ้าสุ่มตัวอย่างครอบครัวที่ใช้รถยนต์ในเขตกรุงเทพมหานครมา 500 ครอบครัว พบว่ามี 160 ครอบครัวที่ใช้รถ Honda

$$\therefore P = \frac{160}{500} = 0.32$$

ดังนั้น ช่วงเชื่อมั่น 95% สำหรับ π คือ

$$P - Z_{.025}\sqrt{\frac{P(1-P)}{n}} \leq \pi \leq P + Z_{.025}\sqrt{\frac{P(1-P)}{n}}$$

$$0.32 - 1.96\sqrt{\frac{(.32)(.68)}{500}} \leq \pi \leq 0.32 + 1.96\sqrt{\frac{(.32)(.68)}{500}}$$

$$0.32 - (1.96 \times .02) \leq \pi \leq 0.32 + (1.96 \times .02)$$

$$0.32 - 0.04 \leq \pi \leq 0.32 + 0.04$$

$$0.28 \leq \pi \leq 0.36$$

$\therefore$ 95% ช่วงเชื่อมั่นของสัดส่วนของครอบครัวที่ใช้รถ Honda คือ 0.28 ถึง 0.36

ค. การประมาณค่าแบบช่วงของความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานประชากร (σ^2 และ σ)

ตัวประมาณค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของประชากร คือ s^2 และ s ตามลำดับ

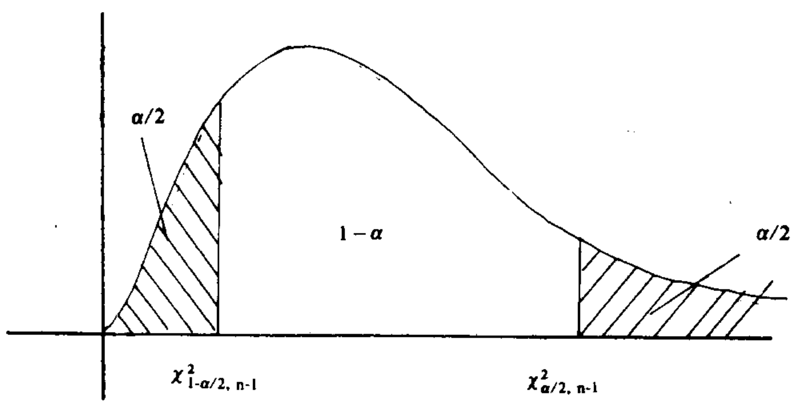
ถ้าสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบปกติ ช่วงเชื่อมั่น $100(1-\alpha)\%$ ของความแปรปรวนของประชากรและส่วนเบี่ยงเบนมาตรฐานของประชากรคือ

$$\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,\,n-1}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

และ

$$\sqrt{\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,\,n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,\,n-1}}}$$

โดยที่ $\chi^2_{1-\alpha/2,\,n-1}$ และ $\chi^2_{\alpha/2,\,n-1}$ คือค่าที่ได้จากตาราง χ^2 ที่มี $df = n-1$ ดังรูป



รูป 6.6

ตัวอย่าง 6.5 ต้องการหา 95% ช่วงเชื่อมั่นของความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานที่แท้จริงของน้ำระกำ โดยการสุ่มตัวอย่างน้ำระกำมา 25 กระป๋อง ได้น้ำหนักโดยเฉลี่ยเท่ากับ 30 กรัม และความแปรปรวนเท่ากับ 0.27 กรัม ดังนั้น 95% ของ σ^2 คือ

$$\frac{(n-1)\,s^2}{\chi^2_{.025,\,24}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{.975,\,24}}$$

$$\frac{(24)\ (.27)}{39.364} \leq \sigma^2 \leq \frac{(24)\ (.27)}{12.401}$$

$$\frac{6.48}{39.364} \leq \sigma^2 \leq \frac{6.48}{12.401}$$

$$0.165 \leq \sigma^2 \leq 0.523$$

และ 95% ของ σ คือ

$$\sqrt{0.165} \leq \sigma \leq 0.523$$

$$0.41 \leq \sigma \leq 0.72$$

6.3 การทดสอบสมมติฐาน (Test Hypothesis)

6.3.1 สมมติฐานทางสถิติ (Statistical hypothesis) นั้นหมายถึงคำกล่าวหรือข้อเสนอที่เกี่ยวกับคุณลักษณะของประชากรที่ต้องการจะศึกษา หรือสำรวจ โดยนำเอาข้อมูลที่ได้จากตัวอย่างที่สุ่มมาจากประชากรมาใช้ในการพิสูจน์ คำกล่าว หรือข้อเสนอนั้น ๆ

วิธีการทดสอบสมมติฐานทางสถิตินั้นทำได้โดยการตั้งสมมติฐานขึ้นมาแล้วพยายามหาข้อเท็จจริงที่รวบรวมได้จากตัวอย่างมาเป็นเครื่องมือช่วยในการตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมติฐานที่ตั้งขึ้น ในการทดสอบสมมติฐานนั้น สมมติฐานที่ตั้งขึ้นมามีอยู่ 2 สมมติฐานด้วยกันคือ

ก. สมมติฐานหลัก (Null hypothesis) เขียนแทนด้วย H_o เป็นสมมติฐานที่กำหนดขึ้นในลักษณะที่ต้องการจะไม่ยอมรับ ซึ่งมักจะเป็นข้อความที่ตรงข้ามกับคำกล่าวอ้างของผู้ที่ทำการทดสอบหรือผู้กล่าวข้อความ

ข. สมมติฐานรอง (Alternative hypothesis) เขียนแทนด้วย H_a เป็นสมมติฐานที่มักจะเป็นข้อความตามคำกล่าวอ้างของผู้ที่ทำการทดสอบ หรือผู้กล่าวข้อความ

ตัวอย่าง 6.6

1. ในการผลิตของอย่างหนึ่งจะมีของเสียเป็นสัดส่วน 20% หลังจากปรับปรุงกระบวนการผลิตแล้วได้สุ่มตัวอย่างมาจำนวนหนึ่ง แล้วนับจำนวนของเสียต้องการจะทดสอบดูว่าการปรับปรุงกระบวนการผลิตจะลดจำนวนของเสียลงหรือไม่ สมมติฐานที่ตั้งขึ้น คือ

$$H_o : \pi = 0.20$$

$$H_a : \pi < 0.20$$

2. โรงพิมพ์แห่งหนึ่งเชื่อว่าแท่นพิมพ์ขนาดใหญ่มีอายุการใช้งานเฉลี่ยเครื่องละ 13,000 และ σ = 2,000 ชั่วโมง จากตัวอย่างสุ่มที่สุ่มมาจำนวนหนึ่ง จะสรุปได้ไหมว่า อายุการใช้งานน้อยกว่า 13,000 ชั่วโมง สมมติฐานที่ตั้งขึ้น คือ

$$H_o : \mu = 13,000 \text{ ชั่วโมง}$$

$$H_a : \mu < 13,000 \text{ ชั่วโมง}$$

3. พ่อค้าขายลำไยตั้งเกณฑ์ในการรับซื้อลำไยจากชาวสวนว่าจะรับซื้อลำไย ถ้ามีลำไยเสียปนอยู่ไม่เกิน 10% สมมติฐานที่ตั้งขึ้น คือ

$$H_o : \pi = 0.10$$

$$H_a : \pi > 0.10$$

4. ในการสำรวจเพศของลูกฝาแฝดจำนวน 300 คู่ ปรากฏว่ามี 84 คู่ที่เป็นชายทั้ง 2 คน มี 126 คู่ที่เป็นชาย 1 คน และหญิง 1 คน และอีก 90 คู่เป็นหญิงทั้ง 2 คน ต้องการทดสอบดูว่าจำนวนคู่ฝาแฝดจะอยู่ในอัตราส่วน

ชช : ชญ : ญญ = 1 : 2 : 1 หรือไม่

H_o : อัตราส่วนของคู่แฝด ชช : ชญ : ญญ เป็น 1 : 2 : 1

H_a : อัตราส่วนของคู่แฝด ชช : ชญ : ญญ ไม่เป็น 1 : 2 : 1

5. ต้องการจะทดสอบดูว่ามีความสัมพันธ์ระหว่างเพศกับการชอบสีหรือไม่ ที่ $\alpha = 0.05$ โดยสุ่มตัวอย่างคนมา 100 คน

H_o : เพศและการชอบสีไม่มีความสัมพันธ์ต่อกัน

H_a : เพศและการชอบสีมีความสัมพันธ์กัน

6. ตัวแทนจำหน่ายน้ำยาซักผ้าขาวชนิดหนึ่ง ต้องการทดสอบดูว่าความนิยมน้ำยาซักผ้าขาวชนิดนี้ในแต่ละภาคมีพอ ๆ กันหรือไม่ โดยสุ่มตัวอย่างมาจาก 4 ภาค

H_o : ความนิยมน้ำยาซักผ้าขาวชนิดนี้ในแต่ละภาคมีพอ ๆ กัน

H_a : H_o ไม่จริง

จากตัวอย่างของการตั้งสมมติฐานที่กล่าวมานั้นจะได้ข้อสังเกตดังนี้

1. สมมติฐานที่ตั้งขึ้นจะต้องเกี่ยวกับลักษณะของประชากรหรือพารามิเตอร์ของประชากร เช่น μ, σ^2 , π เป็นต้น นอกจากนี้อาจจะเกี่ยวกับการแจกแจงของประชากรหรือเกี่ยวกับความสัมพันธ์ของตัวแปร เป็นต้น

2. การตั้งสมมติฐานหลัก (H_o) นั้น ข้อความหรือคำกล่าวใน H_o มักจะมีคำว่า "เท่ากับ" รวมอยู่ด้วยเสมอ ส่วนสมมติฐานรอง (H_a) ข้อความหรือคำกล่าวใน H_a มักจะเป็นคำกล่าวของผู้ที่ทำการทดสอบ หรือผู้กล่าวข้อความ นอกเสียจากว่าถ้าในคำกล่าวนั้นมีคำว่า "เท่ากับ" รวมอยู่ด้วย คำกล่าวนั้นจะนำไปตั้งในสมมติฐาน H_o แทน

3. สำหรับสมมติฐานรอง (H_a) นั้นมี 2 แบบ คือ

ก. สมมติฐานรองทางเดียว

ถ้าให้ θ เป็นพารามิเตอร์ของประชากร

θ_0 เป็นค่าที่ระบุไว้ของพารามิเตอร์ θ

H_a ที่ตั้งขึ้นอาจจะเป็นรูปใดรูปหนึ่งใน 2 แบบนี้ คือ

H_a : $\theta > \theta_0$ หรือ

H_a : $\theta < \theta_0$

ข. สมมติฐานรองสองทาง

H_a ที่ตั้งขึ้นจะเป็นดังนี้ คือ

H_a : $\theta \neq \theta_0$

ในการทดสอบสมมติฐานนั้นเนื่องจากเราต้องอาศัยตัวอย่างที่สุ่มมาจากประชากรมาช่วยตัดสินใจว่าจะยอมรับ หรือปฏิเสธสมมติฐานหลัก (H_o) ที่เราตั้งขึ้นมา ดังนั้น ในการตัดสินใจอาจจะมีการผิดพลาดได้ ซึ่งความผิดพลาดนี้แบ่งออกได้เป็น 2 ประเภท คือ

1. ความผิดพลาดประเภทที่ 1 (Type one error) เป็นความผิดพลาดที่เกิดจากการที่เราปฏิเสธสมมติฐานหลัก (H_o) ทั้ง ๆ ที่สมมติฐานหลัก (H_o) นั้นเป็นจริง และความน่าจะเป็นหรือโอกาสที่จะเกิดความผิดพลาดประเภทนี้เราเรียกว่าการเสี่ยงแบบ 1 (Alpha risk) หรือระดับนัยสำคัญ (Level of significance) และเขียนแทนด้วย α

$$\therefore \alpha = P[\text{เกิดความผิดพลาดประเภทที่ 1}]$$

$$= P[\text{ปฏิเสธ } H_o \mid H_o \text{ จริง}]$$

2. ความผิดพลาดประเภทที่ 2 (Type two error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐานหลัก (H_o) ทั้ง ๆ ที่สมมติฐานหลัก (H_o) นั้นไม่เป็นจริง และความน่าจะเป็นหรือโอกาสที่จะเกิดความผิดพลาดประเภทนี้เราเรียกว่าการเสี่ยงแบบ 2 (Beta risk) และเขียนแทนด้วย β

$$\therefore \beta = P[\text{เกิดความผิดพลาดประเภทที่ 2}]$$

$$= P[\text{ยอมรับ } H_o \mid H_o \text{ ไม่จริง}]$$

ในการทดสอบสมมติฐานนั้น ต้องการให้โอกาสที่จะเกิดความผิดพลาดทั้ง 2 ชนิดนี้มีค่าน้อย ๆ แต่เนื่องจากว่าไม่สามารถที่จะทำให้ α และ β มีค่าน้อยพร้อมกันได้ ดังนั้น ในทางปฏิบัติมักจะกำหนดค่า α ให้มีค่าน้อยแล้วพยายามหาวิธีการทดสอบที่ทำให้ β มีค่าน้อย ซึ่งโดยทั่วไปแล้วมักจะกำหนดให้ $\alpha = 0.05$ หรือ $\alpha = 0.01$

6.3.2 ขั้นตอนในการทดสอบสมมติฐาน

การทดสอบสมมติฐานไม่ว่าจะเป็นการทดสอบเกี่ยวกับพารามิเตอร์ตัวใดก็ตามจะมีลำดับขั้นตอนในการทดสอบเหมือนกันหมด ซึ่งมีทั้งหมด 6 ขั้นตอนด้วยกัน คือ

1. ตั้งสมมติฐาน (Hypothesis formulation) สมมติฐานที่จะตั้งขึ้นนั้นมีสมมติฐานหลัก (H_o) และสมมติฐานรอง (H_a) ซึ่งมี 3 แบบด้วยกันที่จะเป็นไปได้ ดังนี้

ถ้าให้ θ เป็นพารามิเตอร์ของประชากร

θ_0 เป็นค่าที่ระบุไว้ของพารามิเตอร์ θ

แบบที่ 1

H_o : $\theta = \theta_0$ (รวมค่า $\theta < \theta_0$ ไว้ด้วย)

H_a : $\theta > \theta_0$ (เป็นสมมติฐานรองแบบทางเดียว)

แบบที่ 2

H_o : $\theta = \theta_0$ (รวมค่าที่ $\theta > \theta_0$ ไว้ด้วย)

H_a : $\theta < \theta_0$ (เป็นสมมติฐานรองแบบทางเดียว)

แบบที่ 3

H_o : $\theta = \theta_0$

H_a : $\theta \neq \theta_0$ (เป็นสมมติฐานแบบสองทาง)

2. กำหนดระดับนัยสำคัญ (Level of significance หรือ α) และขนาดตัวอย่าง (n)

3. กำหนดตัวสถิติที่ใช้ทดสอบ (Test statistic) โดยที่ถ้า H_o จริงแล้ว เราต้องทราบการแจกแจงของการสุ่มตัวอย่างของตัวสถิตินั้น

4. กำหนดเขตวิกฤต หรือเขตปฏิเสธ H_o (Critical region หรือเขียนแทนด้วย CR) โดยดูจาก

ก. สมมติฐานรองที่ตั้งขึ้นว่าเป็นแบบทางเดียวหรือ 2 ทาง เพื่อจะได้ทราบทิศทางของเขตปฏิเสธ H_o ว่าจะอยู่ทางด้านไหน

ข. ระดับนัยสำคัญ (α) เพื่อที่จะได้ทราบขนาดของเขตปฏิเสธ H_o

ค. การแจกแจงของตัวสถิติที่ใช้ทดสอบ เพื่อที่จะได้ทราบจุดแบ่งเขตปฏิเสธ H_o และเขตยอมรับ H_o ซึ่งเขตวิกฤตมีแบบต่าง ๆ ตาม H_a ดังนี้

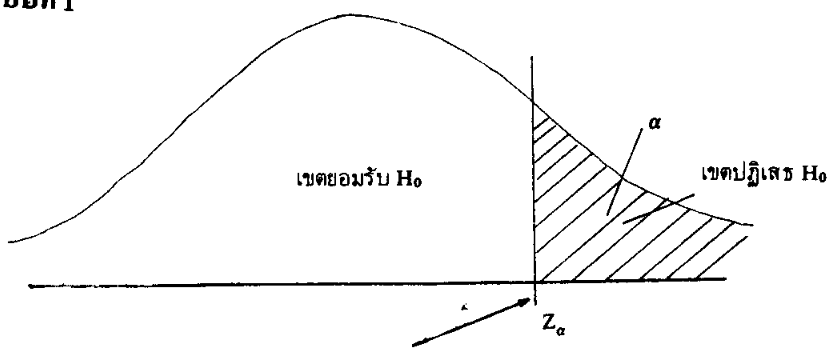
แบบที่ 1

H_o : $\theta = \theta_0$

H_a : $\theta > \theta_0$

สมมติว่าการแจกแจงของตัวสถิติที่ใช้ทดสอบคือการแจกแจงแบบปกติ

แบบที่ 1

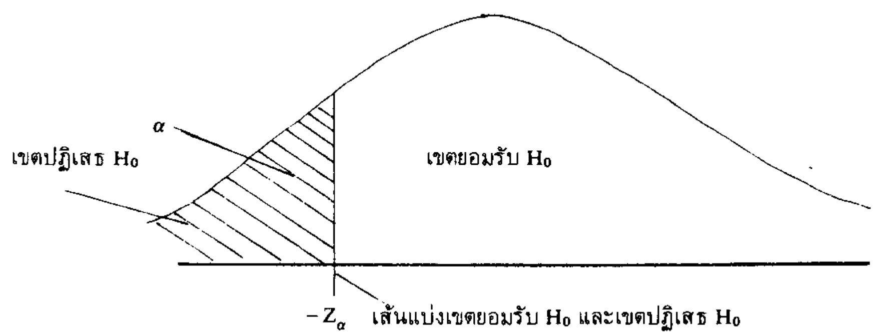


เส้นแบ่งเขตยอมรับ H_0 และเขตปฏิเสธ H_0 ซึ่งจะมีค่าเท่าใดนั้น
เราเปิดจากตาราง Z หาค่า Z_α ว่าจะมีค่าเท่าใด

แบบที่ 2

$H_0 : \theta = \theta_0$

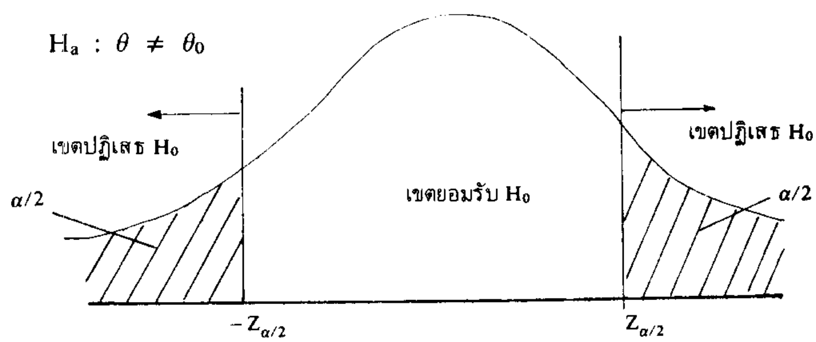
$H_a : \theta < \theta_0$



แบบที่ 3

$H_0 : \theta = \theta_0$

$H_a : \theta \neq \theta_0$



รูป 6.7

5. ทำการสุ่มตัวอย่างแล้วคำนวณค่าของตัวสถิติทดสอบโดยใช้ข้อมูลจากตัวอย่างที่สุ่มได้

6. สรุปผล ถ้าค่าที่คำนวณได้ในขั้นที่ 5 ตกอยู่ในเขตวิกฤต (CR) ก็จะปฏิเสธ H_o ที่ระดับนัยสำคัญ = α แต่ถ้าค่าที่คำนวณได้ในขั้นที่ 5 ตกอยู่ในเขตยอมรับ H_o หรืออยู่นอกเขตวิกฤตก็จะยอมรับ H_o ที่ระดับนัยสำคัญ = α

สิ่งสำคัญที่สุดในการสรุปผลนั้น จะต้องสรุปผลให้สอดคล้องกับคำกล่าวของผู้ทำการทดสอบ หรือสมมติฐานนั้น ๆ ด้วยเสมอ ไม่ใช่บอกแต่เพียงว่าปฏิเสธ H_o หรือยอมรับ H_o

การทดสอบสมมติฐานที่จะกล่าวถึงในวิชา SC 101 นั้นจะขอกล่าวถึงการทดสอบเกี่ยวกับค่าเฉลี่ย การทดสอบเกี่ยวกับสัดส่วน การทดสอบเกี่ยวกับความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเป็นอิสระ และการทดสอบความเป็นเอกภาพ ดังนี้

6.3.3 การทดสอบเกี่ยวกับค่าเฉลี่ยของประชากร (μ) ที่มีการแจกแจงแบบปกติ

เป็นการทดสอบสมมติฐานที่กำหนดว่า ค่าเฉลี่ยของประชากรมีค่าเท่ากับค่าใดค่าหนึ่งที่กำหนดไว้ (μ_0) แล้วนำข้อมูลที่รวบรวมได้จากตัวอย่างมาคำนวณหาค่าสถิติ เพื่อใช้เป็นเครื่องมือในการตัดสินใจว่าจะยอมรับ หรือปฏิเสธสมมติฐานหลักที่ตั้งไว้ ซึ่งเราแบ่งได้เป็น 2 กรณีด้วยกัน ดังนี้

กรณีที่ 1 การทดสอบค่าเฉลี่ยของประชากร เมื่อทราบความแปรปรวนของประชากร หรือเมื่อไม่ทราบความแปรปรวนของประชากร แต่ใช้ตัวอย่างขนาดโต ($n \geq 30$) ขั้นตอนในการทดสอบ มีดังนี้

1. ตั้งสมมติฐาน

H_o : $\mu = \mu_0$

H_a : 1. $\mu > \mu_0$ หรือ

2. $\mu < \mu_0$ หรือ

3. $\mu \neq \mu_0$

2. กำหนด α และขนาดตัวอย่าง n

3. ตัวสถิติที่ใช้ในการทดสอบ คือ

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \quad (\text{เมื่อทราบ } \sigma^2)$$

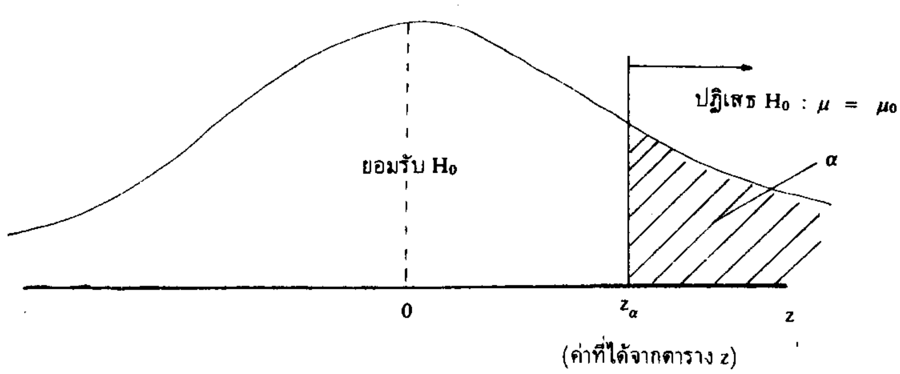
หรือ

$$Z = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \quad (\text{เมื่อไม่ทราบ } \sigma^2 \text{ แต่ } n \geq 30)$$

4. เขตวิกฤต (CR.) โดยดูจาก H_a แบบต่าง ๆ กันทั้ง 3 แบบดังนี้

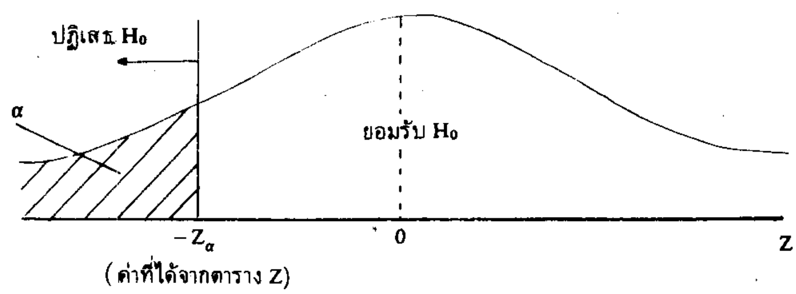
1. ถ้า $H_a: \mu > \mu_0$

CR. : ปฏิเสธ H_0 ถ้า $Z_c > Z_\alpha$ ซึ่งเขียนรูปแสดงได้ดังนี้



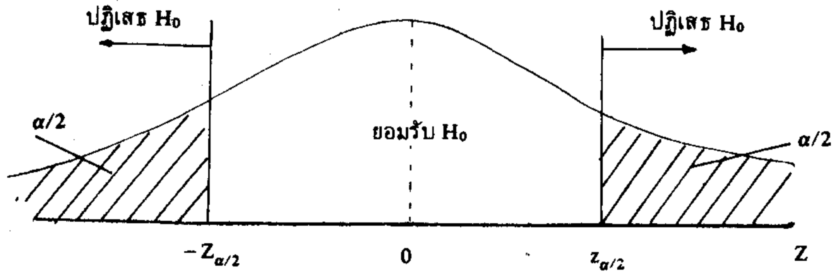
2. ถ้า $H_a : \mu < \mu_0$

CR : ปฏิเสธ H_0 ถ้า $Z_c < -Z_\alpha$ ดังรูป



3. ถ้า $H_a : \mu \neq \mu_0$

CR : ปฏิเสธ H_0 ถ้า $Z_c < -Z_{\alpha/2}$ หรือ $Z_c > Z_{\alpha/2}$ ดังรูป



รูป 6.8

5. จากตัวอย่างที่สุ่มมาขนาด n คำนวณหาค่า Z_c จาก

$$Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad (\text{เมื่อทราบ } \sigma^2)$$

หรือ

$$Z_c = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (\text{เมื่อไม่ทราบ } \sigma^2 \text{แต่ } n \geqslant 30)$$

เมื่อ $S^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]$

6. สรุปผล คือ

   1. ถ้า Z_c ตกอยู่ใน CR. จะปฏิเสธ H_o นั่นคือยอมรับ H_a
   2. ถ้า Z_c ตกอยู่นอก CR. จะยอมรับ H_o

ตัวอย่าง 6.7

โรงงานผลิตของเด็กเล่นแห่งหนึ่ง ทราบว่าคนงานคนหนึ่งสามารถผลิตของเล่นได้โดยเฉลี่ยชั่วโมงละ 15 ชิ้น และค่าเบี่ยงเบนมาตรฐาน 5 ชิ้น มีผู้เสนอให้ใช้วิธีใหม่ซึ่งจะช่วยทำให้คนงานสามารถผลิตของเล่นเด็กได้มากขึ้น เมื่อได้ทดลองวิธีใหม่ในเวลา 100 ชั่วโมงดู ปรากฏว่าคนงานผลิตของเล่นเด็กได้โดยเฉลี่ยชั่วโมงละ 16 ชิ้น ถ้าให้ $\alpha = 0.01$ มีเหตุผลเพียงพอที่จะเชื่อว่า วิธีการใหม่จะช่วยให้คนงานผลิตของเล่นได้มากขึ้นหรือไม่

1. H_o : $= 15$

H_a : > 15

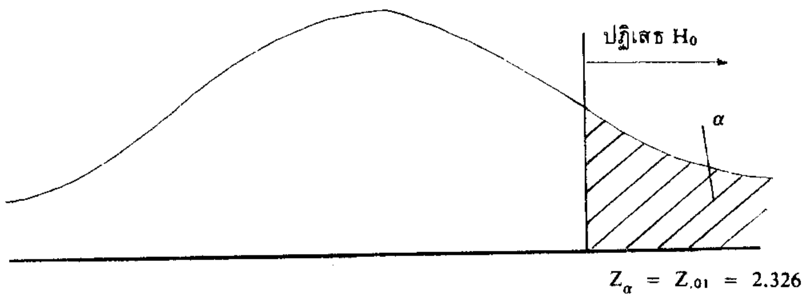
2. $\alpha = 0.01$, $n = 100$ ชั่วโมง

3. ตัวสถิติที่ใช้ทดสอบคือ

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

เราทราบ $\sigma^2 = 5^2 = 25$

4. CR. : ปฏิเสธ H_o ถ้า $z_c > z_\alpha$



รูป 6.9

$\therefore$ CR. : จะปฏิเสธ H_o ถ้า $z_c > z_{.01}$

(เปิดตาราง z จะได้ $z_{.01} = 2.326$)

นั่นคือ CR.: ปฏิเสธ H_o ถ้า $z_c > 2.326$

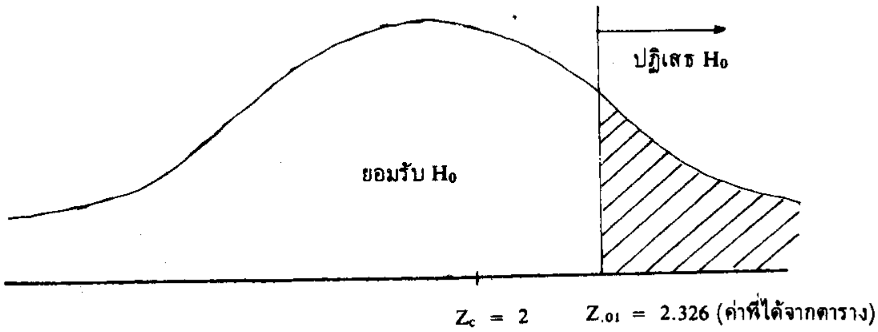
5. คำนวณค่า z_c จาก

$$z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

$$= \frac{16 - 15}{S/\sqrt{100}} = 2$$

6. สรุปผล

จะเห็นได้ว่า $z_c < 2.326$ ดังนั้น ค่า z_c ตกอยู่นอก CR. ดังรูป 6.10



รูป 6.10

∴ เรายอมรับ H_o ที่ว่า $\mu = 15$ แสดงว่าไม่มีเหตุผลเพียงพอที่จะเชื่อว่าวิธีการใหม่จะช่วยให้คนงานผลิตของเล่นได้มากขึ้น

6.3.4 การทดสอบเกี่ยวกับสัดส่วนของประชากร

ขั้นตอนในการทดสอบมีดังนี้

1. ตั้งสมมติฐาน

   H_o : $\pi = \pi_0$

   H_a : 1. $\pi > \pi_0$ หรือ

   2. $\pi < \pi_0$ หรือ

   3. $\pi \neq \pi_0$

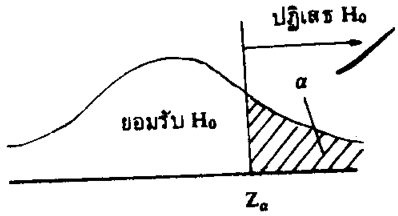
2. กำหนด α และขนาดตัวอย่าง n

3. ตัวสถิติที่ใช้ทดสอบ คือ

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

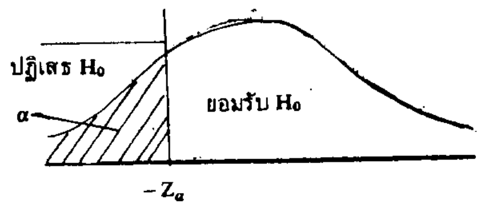
4. CR.: 1. ถ้า $H_a : \pi > \pi_0$

จะปฏิเสธ H_o ถ้า $Z_c > Z_\alpha$



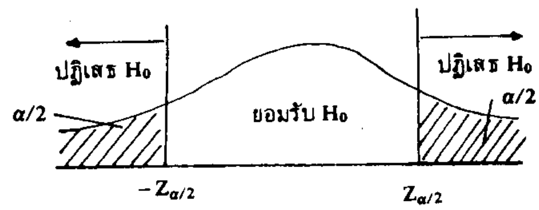
2. ถ้า $H_a : \pi < \pi_0$

จะปฏิเสธ H_o ถ้า $Z_c < -Z_\alpha$



3. ถ้า $H_a : \pi \neq \pi_0$

จะปฏิเสธ H_o ถ้า $Z_c < -Z_{\alpha/2}$ หรือ $Z_c > Z_{\alpha/2}$



รูป 6.11

5. คำนวณหา Z_c จากสูตร

$$Z_c = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

6. สรุปผล : 1. ถ้า Z_c ตกอยู่ใน CR.จะปฏิเสธ H_o และยอมรับ H_a

2. ถ้า Z_c ตกอยู่นอก CR. จะยอมรับ H_o

ตัวอย่างที่ 6.8

บริษัทผู้ผลิตอ้างว่า กล้องถ่ายรูปที่ผลิตออกมาจะมีกล้องที่ชำรุดไม่เกิน 10% จากการสุ่มตัวอย่างกล้อง 200 อัน ปรากฏว่า มีกล้องที่ชำรุด 4 อัน จงทดสอบคำกล่าวอ้างนี้ที่ $\alpha = 0.01$

1. ตั้งสมมติฐาน

$H_o : \pi = 0.01$

$H_a : \pi > 0.01$

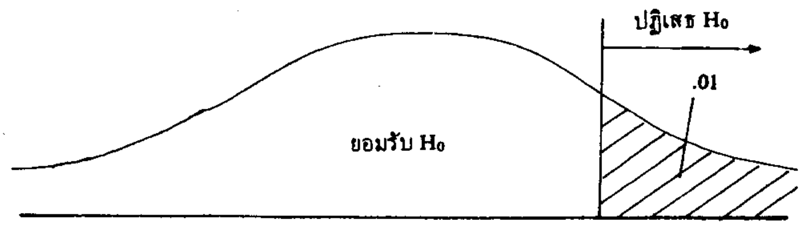
2. $\alpha = 0.01, \quad n = 200$

3. ตัวสถิติที่ใช้ทดสอบ คือ

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

4. CR. : ปฏิเสธ H_o ถ้า $Z_c > Z_{.01}$

คือ $Z_c > 2.33$



รูป 6.12

5. $P = \frac{4}{200} = \frac{1}{50} = 0.02$

$\pi_0 = 0.01, \quad n = 200$

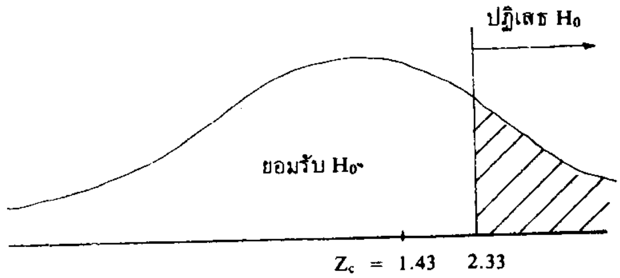
$$\therefore \quad Z_c = \frac{.02 - .01}{\sqrt{\frac{(.01)(.99)}{200}}}$$

$$= \frac{.02 - .01}{.007} = 1.43$$

6. $\because$ $Z_c < 2.33$ ดังรูป 6.13

จะเห็นได้ว่า Z_c ตกอยู่นอก CR.

$\therefore$ เรายอมรับ H_o แสดงว่ากล้องถ่ายรูปที่ผลิตจากบริษัทนี้จะมีของชำรุดไม่เกิน 1% จริง



รูป 6.13

6.3.5 การทดสอบเกี่ยวกับความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของประชากรที่มีการแจกแจงแบบปกติ

สุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบปกติ ซึ่งมีความแปรปรวนเป็น σ^2 ขั้นตอนในการทดสอบมีดังนี้

1. ตั้งสมมติฐาน

สำหรับความแปรปรวน

$H_o : \sigma^2 = \sigma_0^2$

H_a : 1. $\sigma^2 > \sigma_0^2$ หรือ

2. $\sigma^2 < \sigma_0^2$ หรือ

3. $\sigma^2 \neq \sigma_0^2$

สำหรับส่วนเบี่ยงเบนมาตรฐาน

$H_o : \sigma = \sigma_0$

H_a : 1. $\sigma > \sigma_0$ หรือ

2. $\sigma < \sigma_0$ หรือ

3. $\sigma \neq \sigma_0$

2. กำหนด α และขนาดตัวอย่าง n

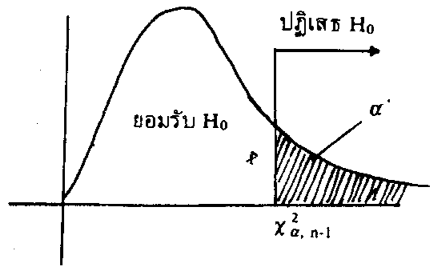
3. ตัวสถิติที่ใช้ทดสอบ คือ

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

มี df = n-1

4. CR. : 1. ถ้า H_a : $\sigma^2 > \sigma_0^2$ หรือ $\sigma > \sigma_0$

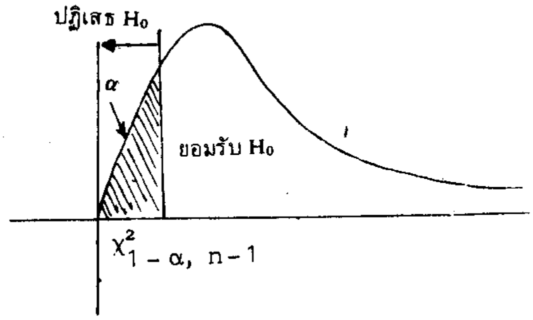
จะปฏิเสธ H_o ถ้า $\chi_c^2 > \chi_{\alpha,n-1}^2$



รูป 6.14

2. ถ้า H_a : $\sigma^2 < \sigma_0^2$ หรือ $\sigma < \sigma_0$

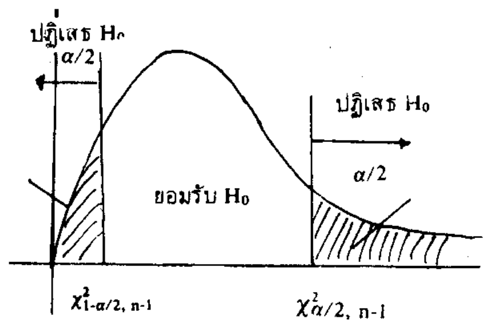
จะปฏิเสธ H_o ถ้า $\chi_c^2 < \chi_{1-\alpha,\ n-1}^2$



รูป 6.15

3. ถ้า H_a : $\sigma^2 \neq \sigma_0^2$ หรือ $\sigma \neq \sigma_0$

จะปฏิเสธ H_o ถ้า $\chi_c^2 < \chi_{1-\alpha/2,\ n-1}^2$ หรือ $\chi_c^2 > \chi_{\alpha/2,\ n-1}^2$



รูป 6.16

5. คำนวณหาค่า χ_c^2 จาก

$$\chi_c^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

เมื่อ $$S^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]$$

6. สรุปผล

1. ถ้า χ_c^2 ตกอยู่ใน CR. จะปฏิเสธ H_o และยอมรับ H_a
2. ถ้า χ_c^2 ตกอยู่นอก CR. จะยอมรับ H_o

ตัวอย่าง 6.9

บริษัทผลิตยางรถยนต์โฆษณาว่า อายุการใช้งานของยางรถยนต์ มีความแปรปรวนเท่ากับ 0.81 ปี ถ้าสุ่มตัวอย่างยางรถยนต์มา 10 เส้นแล้ว คำนวณส่วนเบี่ยงเบนมาตรฐานได้ 1.2 ปี จงทดสอบว่า $\sigma^2 > 0.81$ ปี โดยใช้ $\alpha = 0.05$

1. $H_o : \sigma^2 = 0.81$ ปี

   $H_a : \sigma^2 > 0.81$ ปี

2. $\alpha = 0.05, \quad n = 10$

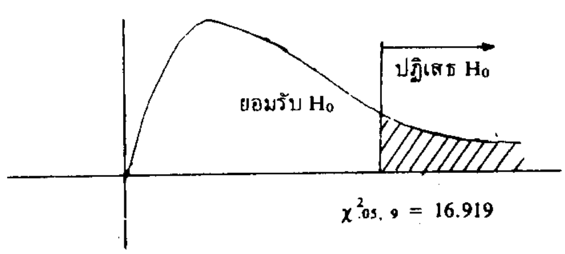
3. ตัวสถิติที่ใช้ทดสอบ คือ

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$df = 10-1 = 9$$

4. CR. : ปฏิเสธ H_o ถ้า $\chi_c^2 > \chi^2_{.05,9}$

   คือ $\chi_c^2 > 16.919$



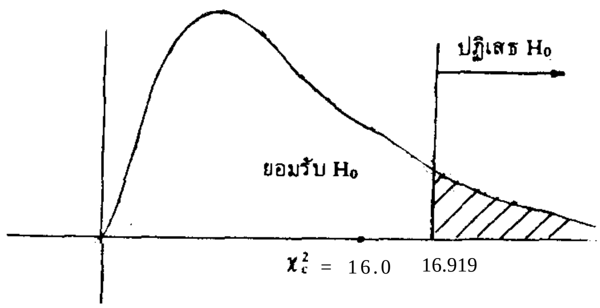
รูป 6.17

5. คำนวณค่า χ_c^2 จาก

$$\chi_c^2 = \frac{(n-1)s^2}{\sigma_0^2}$$
$$= \frac{(9)(1.2)^2}{0.81}$$
$$= \frac{(9)(1.44)}{0.81} = \frac{12.36}{0.81}$$
$$= 16$$

6. สรุปผล $\because$ χ_c^2 ตกอยู่นอก CR.

$\therefore$ ยอมรับ H_o ที่ว่า $\sigma^2 = 0.81$ แสดงว่าอายุการใช้งานของยางรถยนต์ที่ผลิตจากบริษัทนี้มีความแปรปรวนเท่ากับ 0.81 ปี



รูป 6.18

6.3.6 การทดสอบความเป็นอิสระต่อกันของลักษณะ 2 ลักษณะ

ในการศึกษาลักษณะ 2 ลักษณะในประชากร หรือกลุ่มที่เราสนใจว่าจะมีความสัมพันธ์กันหรือไม่นั้น จะจัดจำแนกข้อมูลในตัวอย่างขนาด n ออกเป็นกลุ่ม ๆ โดยใช้ลักษณะทั้ง 2 เป็นตัวจัดจำแนก ถ้าให้ลักษณะหนึ่งเป็น A โดย A แบ่งได้เป็น r ส่วนย่อย คือ $A_1, A_2, \ldots, A_r$ และอีกลักษณะหนึ่งเป็น B โดย B แบ่งได้เป็น c ส่วนย่อยคือ $B_1, B_2, \ldots, B_c$

เมื่อทำการสุ่มตัวอย่างขนาด n จะได้ ตารางการณ์จร (contigency table) ขนาด r × c ของค่าสังเกตหรือความถี่ดังนี้

	O_{ij}	ลักษณะ B B_1	B_2	....	B_c	ผลรวมทาง row
ลักษณะ A	A_1	O_{11}	O_{12}	....	O_{1c}	R_1
	A_2	O_{21}	O_{22}	....	O_{2c}	R_2
	A_3	O_{31}	O_{32}	....	O_{3c}	R_3
	A_r	O_{r1}	O_{r2}	....	O_{rc}	R_r
ผลรวมทาง column		C_1	C_2	....	C_c	n

เมื่อ O_{ij} = ความถี่ที่มีลักษณะ A_i และ B_j

R_i = ผลรวมของ row ที่ i

C_j = ผลรวมของ column ที่ j

ในการทดสอบว่าลักษณะ A กับ ลักษณะ B เป็นอิสระต่อกันนั้นจะต้องหาค่า E_{ij} (ความถี่ที่คาดหมายตามทฤษฎี)

$$E_{ij} = \frac{R_i \times C_j}{n}$$

และองศาความเป็นอิสระ (df) หาได้จากสูตร

$$df = (r-1)(c-1)$$

ขั้นตอนในการทดสอบความเป็นอิสระมีดังนี้

1. ตั้งสมมติฐาน

H_o : ตัวแปรทั้ง 2 เป็นอิสระต่อกัน (ไม่มีความสัมพันธ์กัน)

H_a : ตัวแปรทั้ง 2 ไม่เป็นอิสระต่อกัน (มีความสัมพันธ์ต่อกัน)

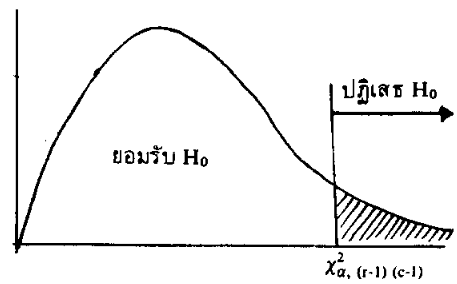
2. กำหนด α และขนาดตัวอย่าง n

3. ตัวสถิติที่ใช้ทดสอบ คือ

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

4. CR. : ปฏิเสธ H_o ถ้า

$$\chi^2_c > \chi^2_{\alpha,(r-1)(c-1)}$$



รูป 6.19

5. คำนวณหาค่า χ^2_c จาก

$$\chi^2_c = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

โดยที่ $E_{ij} = \frac{R_i \times C_j}{n}$

6. สรุปผล :

1. ถ้า χ^2_c ตกอยู่ใน CR. จะปฏิเสธ H_o และยอมรับ H_a
2. ถ้า χ^2_c ตกอยู่นอก CR. จะยอมรับ H_o

ตัวอย่าง 6.10

ต้องการทดสอบดูว่ามีความสัมพันธ์ระหว่างเพศกับการชอบสีหรือไม่ ที่ $\alpha = 0.05$ สมมติว่าเลือกสุ่มตัวอย่างคนมา 100 คน ปรากฏผลดังนี้

สี \ เพศ	ชาย	หญิง	รวม
แดง	10	20	30
เขียว	20	10	30
เหลือง	30	10	40
รวม	60	40	100

1. ตั้งสมมติฐาน

   H_o : เพศและการชอบสีเป็นอิสระต่อกัน

   H_a : H_o ไม่จริง

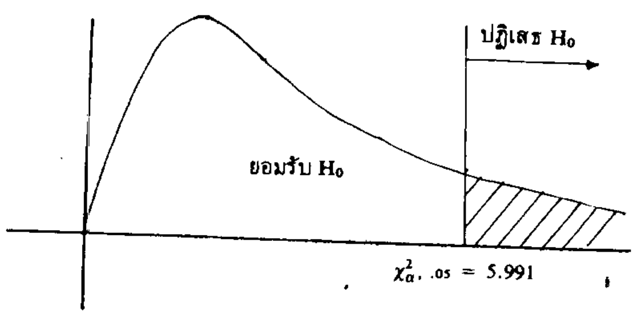
2. $\alpha = 0.05$, $n = 100$

3. ตั้งสถิติที่ใช้ทดสอบ คือ

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

4. CR.: จะปฏิเสธ H_o ถ้า $\chi_c^2 > \chi^2_{.05,\ (3-1)(2-1)}$

   คือ $\chi_c^2 > 5.991$



รูป 6.20

5. คำนวณค่า χ_c^2 ซึ่งต้องหาค่า E_{ij} ก่อนจาก

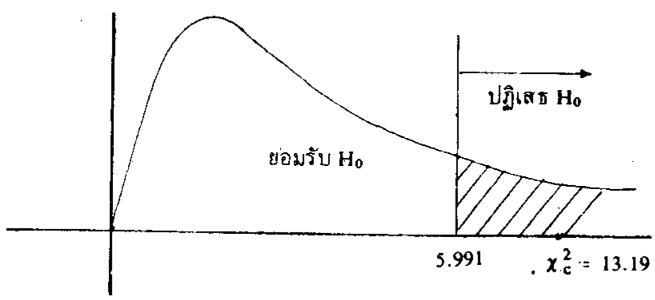
$$E_{ij} = \frac{R_i \times C_j}{n}$$

ซึ่งจะได้ค่าดังนี้

O_{ij}	10	20	30	20	10	10
E_{ij}	$\frac{30 \times 60}{100} = 18$	$\frac{30 \times 60}{100} = 18$	$\frac{40 \times 60}{100} = 24$	$\frac{30 \times 40}{100} = 12$	$\frac{30 \times 40}{100} = 12$	$\frac{40 \times 40}{100} = 16$

$$\therefore \quad \chi_c^2 = \frac{(10-18)^2}{18} + \frac{(20-18)^2}{18} + \ldots\ldots + \frac{(10-16)^2}{16}$$

$$= 13.19$$



รูป 6.21

6. สรุปผล $\because$ χ_c^2 ตกอยู่ใน CR.

$\therefore$ เราปฏิเสธ H_o ที่ว่าเพศและความชอบสีเป็นอิสระต่อกัน

6.3.7 การทดสอบความเป็นเอกภาพ

เป็นการทดสอบดูว่าประชากรหรือกลุ่มต่าง ๆ เป็นเอกภาพ หรือเหมือนกันหรือไม่ โดยการสุ่มตัวอย่างขนาด $R_1, R_2, \ldots, R_r$ จาก r ประชากรซึ่ง คือ $A_1, A_2, \ldots, A_r$ แล้วจัดจำแนก R_i ตามลักษณะของ B ซึ่งมีอยู่ c ลักษณะย่อย คือ $B_1, B_2, B_3, \ldots, B_c$ ซึ่งจะได้ตาราง $r \times c$ ดังนี้

		ลักษณะ B				ผลรวมทาง
	O_{ij}	B_1	B_2	....	B_c	row
	A_1	O_{11}	O_{12}	....	O_{1c}	R_1
ประชากร	A_2	O_{21}	O_{22}	....	O_{2c}	R_2
	¦	¦	¦		¦	¦
	A_r	O_{r1}	O_{r2}	....	O_{rc}	R_r
ผลรวมของ column		C_1	C_2	....	C_c	n

ขั้นตอนในการทดสอบ

1. ตั้งสมมติฐาน

   H_o : ประชากรต่าง ๆ เป็นเอกภาพกัน

   H_a : H_o ไม่จริง

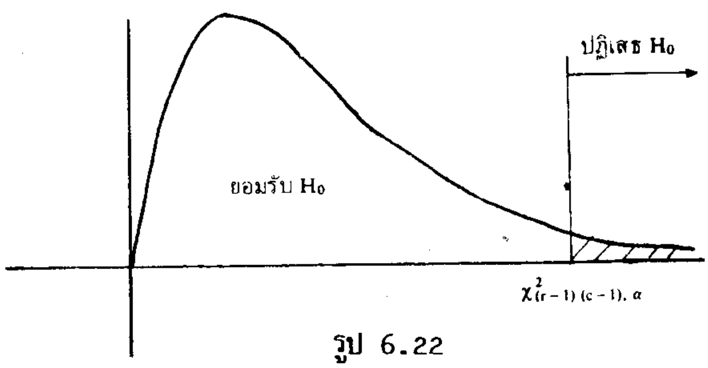
2. กำหนด α และ n

3. ตัวสถิติที่ใช้ทดสอบ คือ

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$df = (r-1)(c-1)$$

4. CR.: ปฏิเสธ H_o ถ้า $\chi^2_c > \chi^2_{\alpha,\ (r-1)(c-1)}$



รูป 6.22

5. คำนวณค่า χ^2_c จาก

$$\chi^2_c = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

โดยที่ $E_{ij} = \frac{R_i \times C_j}{n}$

6. สรุปผล 1. ถ้า χ^2_c ตกอยู่ใน CR. จะปฏิเสธ H_o และยอมรับ H_a

2. ถ้า χ^2_c ตกอยู่นอก CR. จะยอมรับ H_o

ตัวอย่าง 6.11

ตัวแทนจำหน่ายน้ำยาปรับผ้านุ่มชนิดหนึ่ง กล่าวว่า ความนิยมน้ำยาปรับผ้านุ่มชนิดนี้ ในแต่ละภาคมีพอ ๆ กัน เพื่อยืนยันคำกล่าวนี้ เขาจึงสุ่มตัวอย่างมาจาก 4 ภาค ได้ผลดังนี้

ภาค	จำนวนคนนิยม	จำนวนคนไม่นิยม	รวม
เหนือ	120(140)	80(60)	200
กลาง	200(175)	50(75)	250
ตะวันออกเฉียงเหนือ	200(210)	100(90)	300
ใต้	180(175)	70(75)	250
รวม	700	300	1000

จงทดสอบที่ $\alpha = .01$ ว่าคำกล่าวของตัวแทนจำหน่ายน้ำยาปรับผ้านุ่มชนิดนี้จะเชื่อถือได้หรือไม่

1. H_o: ความนิยมน้ำยาปรับผ้านุ่มชนิดนี้ในแต่ละภาคมีพอ ๆ กัน

   H_a: H_o ไม่จริง

2. $\alpha = .01$, $n = 1000$

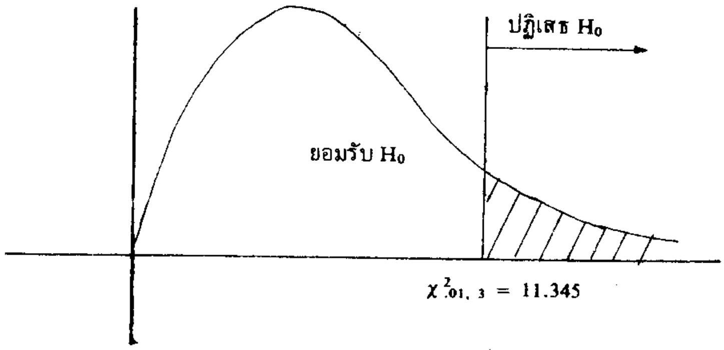
3. ตัวสถิติที่ใช้ทดสอบคือ

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

df = (4-1) (2-1) = 3

4. CR.: จะปฏิเสธ Ho ถ้า

$\chi^2_c > \chi^2_{.01,\ 3}$ คือ $\chi^2_c > 11.345$



5. คำนวณหาค่า χ^2_c โดยหา E_{ij} ก่อน โดยค่าที่ได้จะอยู่ในตารางที่โจทย์ คือ ค่าในวงเล็บ

$$\therefore\ \chi^2_c = \frac{(120-140)^2}{140} + \frac{(200-175)^2}{175} + \ldots\ldots\ldots\ldots\ldots\ldots + \frac{(70-75)^2}{75}$$

$$= 23.49$$

6. สรุปผล $\because\ \chi^2_c > 11.345$

$\therefore$ ปฏิเสธ H_o และยอมรับ H_a แสดงว่าความนิยมน้ำยาปรับผ้านุ่มชนิดนี้ในแต่ละภาคไม่เท่ากัน

6.4 ความสัมพันธ์ระหว่างตัวแปร

ในหัวข้อนี้จะศึกษาตัวแปร 2 ตัวให้เป็นตัวแปร X และตัวแปร Y โดยจะดูว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันหรือไม่ และถ้าตัวแปรทั้ง 2 มีความสัมพันธ์กัน จะมีความสัมพันธ์กันอย่างไร ซึ่งเป็นเรื่องของการศึกษาสหสัมพันธ์ เช่น จะศึกษาดูว่าน้ำหนักของคนจะมีความสัมพันธ์กับส่วนสูงหรือไม่ หรือจะศึกษาดูว่าคนที่เรียนวิชาภาษาอังกฤษเก่งจะเรียนวิชาภาษาฝรั่งเศสเก่งหรือไม่ เป็นต้น

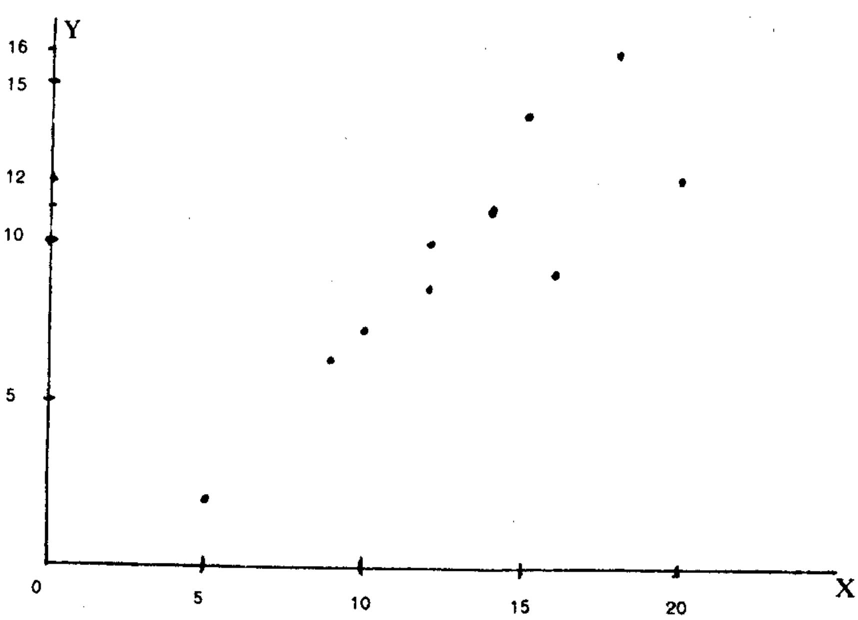
ในการที่จะตรวจดูว่าตัวแปรทั้ง 2 นั้นมีความสัมพันธ์กันหรือไม่นั้น อาจจะทำอย่างคร่าว ๆ ได้ โดยใช้กราฟแสดงการกระจายของจุดที่เรียกว่า สะแกตเตอร์แกรม (Scattergram) ตัวอย่างเช่น คะแนนสอบวิชาภาษาอังกฤษ และวิชาภาษาฝรั่งเศสของนักศึกษาคณะมนุษยศาสตร์ จำนวน 10 คน เป็นดังนี้

ให้ X เป็นคะแนนสอบวิชาภาษาอังกฤษ

Y เป็นคะแนนสอบวิชาภาษาฝรั่งเศส

X	20	18	16	15	14	12	12	10	8	5
Y	12	16	10	14	12	10	9	8	7	2

ซึ่งจะดูความสัมพันธ์ระหว่างตัวแปรทั้ง 2 อย่างคร่าว ๆ โดยเขียนจุดบนกราฟตามค่าพิกัด (X, Y) ของค่าสังเกตแต่ละคู่ ดังนี้



Scatter diagram แสดงความสัมพันธ์ระหว่างตัวแปร X และ Y

รูป 6.23

6.4.1 สหสัมพันธ์ (Correlation)

ในการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร เราเรียกว่า การวิเคราะห์สหสัมพันธ์ (correlation analysis) จุดมุ่งหมายในการศึกษาสหสัมพันธ์ คือ จะพิจารณาดูว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันหรือไม่ และถ้ามีความสัมพันธ์กันขนาดความสัมพันธ์จะมีมากน้อยแค่ไหน ซึ่งการวัดขนาดของความสัมพันธ์ระหว่างตัวแปรทั้ง 2 เราใช้สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เป็นตัววัด สหสัมพันธ์ระหว่างตัวแปรอาจจะเป็นเชิงเส้นตรง หรืออาจจะไม่เป็นเส้นตรง ในที่นี้จะศึกษาเฉพาะกรณีเชิงเส้นตรงเท่านั้น

6.4.2 สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient)

ใช้สัญลักษณ์ ρ แทนสัมประสิทธิ์สหสัมพันธ์ของประชากรและสัญลักษณ์ r แทนสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง โดยที่ค่าของ ρ และ r จะมีค่าตั้งแต่ -1 ถึง $+1$ การพิจารณาสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r) จะดู 2 อย่างด้วยกัน คือ

1. ดูเครื่องหมายของ r ว่าเป็นเครื่องหมายบวกหรือลบ ถ้ามีเครื่องหมายบวกแสดงว่ามีความสัมพันธ์ไปในทิศทางเดียวกัน คือถ้าค่าของตัวแปรตัวหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรก็จะเพิ่มตามด้วย และถ้ามีเครื่องหมายลบ แสดงถึงความสัมพันธ์ในทิศทางตรงกันข้าม คือถ้าตัวแปรตัวหนึ่งเพิ่มค่าของอีกตัวแปรหนึ่งจะลดลง

2. ดูขนาดของตัวเลข ถ้า $r \geqslant 0.8$ ถือว่าขนาดความสัมพันธ์มีมาก ถ้า r มีค่าประมาณ 0.5 ถือว่ามีความสัมพันธ์ปานกลาง ถ้า $r \leqslant 0.3$ ถือว่ามีความสัมพันธ์ระดับต่ำหรือเกือบไม่มีเลย

สัมประสิทธิ์สหสัมพันธ์ (r) มีคุณสมบัติดังนี้

1. ค่าของ r จะอยู่ในพิสัย -1.00 ถึง $+ 1.00$ คือ

$$-1.00 \leqslant r \leqslant + 1.00$$

2. ความสัมพันธ์ทางบวกระหว่างตัวแปรทั้ง 2 หมายถึง ถ้าค่าของตัวแปรตัวหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรหนึ่งก็จะเพิ่มขึ้นด้วย หรือถ้าค่าของตัวแปรตัวหนึ่งลดลงค่าของอีกตัวแปรหนึ่งก็จะลดลงด้วย

3. ความสัมพันธ์ทางลบระหว่างตัวแปรทั้ง 2 หมายถึง ถ้าค่าของตัวแปรตัวหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรหนึ่งจะลดลง

4. ถ้า $r = 0$ หมายถึงไม่มีความสัมพันธ์ระหว่างตัวแปรทั้ง 2

การคำนวณหาค่า r นั้น คำนวณหาได้จากสูตรดังต่อไปนี้

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2)-(\Sigma x)^2][n(\Sigma y^2)-(\Sigma y)^2]}}$$

เมื่อ r = สัมประสิทธิ์สหสัมพันธ์

x และ y เป็นตัวแปร 2 ตัวที่เราต้องการวัดสหสัมพันธ์ n เป็นจำนวนคู่ค่าสังเกต

ตัวอย่าง 6.12

จากข้อมูลเกี่ยวกับคะแนนสอบวิชาภาษาอังกฤษ (x) และคะแนนสอบวิชาภาษาฝรั่งเศส (y) ของนักศึกษา 5 คน เป็นดังนี้

คนที่	คะแนนสอบวิชาภาษาอังกฤษ(x)	คะแนนสอบวิชาภาษาฝรั่งเศส(y)	x^2	y^2	xy
1	2	1	4	1	2
2	4	2	16	4	8
3	5	4	25	16	20
4	7	5	49	25	35
5	8	5	64	25	40
รวม	26	17	158	71	105

หาค่า r ได้จากสูตร

$$r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{[n\,\Sigma x^2 - (\Sigma x)^2]\,[n\,\Sigma y^2 - (\Sigma y)^2]}}$$

แทนค่า $\Sigma x = 26$, $\Sigma y = 17$, $\Sigma x^2 = 158$, $\Sigma y^2 = 71$, $\Sigma xy = 105$ และ $n = 5$ ลงในสูตรจะได้

$$r = \frac{5(105) - (26)(17)}{\sqrt{[5(158) - (26)^2][5(71) - (17)^2]}}$$

$$= \frac{525 - 442}{\sqrt{(790 - 676)(355 - 289)}}$$

$$= \frac{83}{\sqrt{7524}} = 0.96$$

จากค่า $r = 0.96$ แสดงให้เห็นความสัมพันธ์ระหว่างคะแนนสอบวิชาภาษาอังกฤษ กับคะแนนสอบวิชาภาษาฝรั่งเศสว่ามีความสัมพันธ์ในเชิงบวก คือความสัมพันธ์ในทิศทางเดียวกัน และขนาดของความสัมพันธ์อยู่ในระดับสูง

6.4.3 การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์

นอกจากการคำนวณหาค่า r แล้วยังมีการทดสอบดูว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันจริงหรือไม่ โดยทำการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์ วิธีการทดสอบมีขั้นตอนเหมือนหัวข้อ 6.3 ที่กล่าวมาแล้ว ดังนี้

1. ตั้งสมมติฐาน

   $H_o: \rho = 0$

   $H_a: \rho \neq 0$

2. กำหนด α และ n

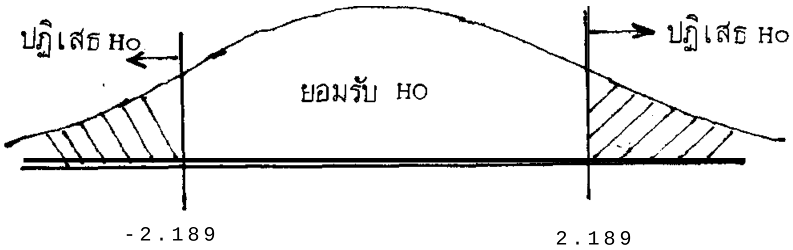
3. ตัวสถิติที่ใช้ทดสอบ คือ

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

   $df = n-2$

4. CR.: ปฏิเสธ H_o ถ้า $t_c < -t_{\alpha/2,\, n-2}$ หรือ $t_c > t_{\alpha/2,\, n-2}$

   หรือ $|t_c| > t_{\alpha/2,\, n-2}$



รูป 6.24

5. คำนวณหาค่า t_c จากสูตร

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

$$= \frac{0.50\sqrt{24-2}}{\sqrt{1-(.50)^2}}$$

$$= \frac{(.50)(4.69)}{\text{A x -}}$$

$$= 2.71$$

6. สรุปผล : $\because$ t_c ตกอยู่นอก CR.

$\therefore$ ยอมรับ H_o: $\rho = 0$

แสดงว่าไม่มีสหสัมพันธ์ในประชากรทั้ง 2

แบบฝึกหัดบทที่ 6

1. จงให้ความหมายของคำต่อไปนี้

   ก. การประมาณค่า

   ข. ค่าประมาณ

   ค. พารามิเตอร์

   ง. ตัวประมาณค่า

2. การประมาณค่าพารามิเตอร์มีกี่วิธี อะไรบ้าง จงอธิบาย

3. ตัวประมาณค่าที่ดีจะต้องมีคุณสมบัติอย่างไร

4. ช่วงเชื่อมั่นคืออะไร

5. โรงงานผลิตเครื่องใช้ไฟฟ้าผลิตหลอดไฟฟ้าซึ่งมีอายุการใช้งานที่มีการแจกแจงแบบปกติและมีส่วนเบี่ยงเบนมาตรฐาน 40 ชั่วโมง สุ่มตัวอย่างหลอดไฟฟ้ามา 30 หลอด คำนวณได้อายุการใช้งานโดยเฉลี่ย 780 ชั่วโมง จงหาช่วงเชื่อมั่น 95% ของค่าเฉลี่ยที่แท้จริงของหลอดไฟฟ้าที่ผลิตจากโรงงานนี้

6. สุ่มตัวอย่างผู้มีสิทธิออกเสียงเลือกตั้ง 100 คน จากผู้มีสิทธิออกเสียงทั้งหมดของหมู่บ้านแห่งหนึ่ง ปรากฏว่ามี 55% ที่ชอบผู้สมัคร จงหาช่วงเชื่อมั่น 99% ของสัดส่วนที่แท้จริง

7. สุ่มตัวอย่างพนักงานจากโรงงานแห่งหนึ่ง พบว่า พนักงานจากโรงงาน 9 คน มีส่วนเบี่ยงเบนมาตรฐานของรายได้เท่ากับ 6,000.- บาทต่อปี

   ก. จงหาช่วงเชื่อมั่น 95% ของส่วนเบี่ยงเบนมาตรฐานที่แท้จริงของรายได้ของพนักงาน

   ข. จงหาช่วงเชื่อมั่น 99% ของความแปรปรวนของรายได้ที่แท้จริงของพนักงาน

8. จงนิยามความหมายของคำต่อไปนี้

   ก. สมมติฐาน

   ข. ระดับนัยสำคัญ

   ค. ความคลาดเคลื่อนแบบ 1

   ง. ความคลาดเคลื่อนแบบ 2

   จ. เขตวิกฤต

9. สมมติฐานหลักและสมมติฐานรอง คืออะไร

10. การทดสอบแบบทางเดียว และการทดสอบแบบ 2 ทาง คืออะไร

11. จงบอกลำดับขั้นตอนในการทดสอบสมมติฐาน

12. การหาเขตวิกฤตจะต้องอาศัยอะไรบ้าง

13. ตัวสถิติทดสอบ คืออะไร

14. จงพิจารณาสมมติฐานต่อไปนี้ว่าเป็นการทดสอบแบบทางเดียวหรือแบบสองทาง

ก. $H_1 : \mu \neq 4.10$

ข. $H_1 : \mu < 4.10$

ค. $H_1 : \mu > 81$

ง. $H_1 : \mu > 0.66$

จ. $H_1 : \mu \neq 1.90$

ฉ. $H_1 : \mu < 3$

15. จงตั้งสมมติฐานหลักและสมมติฐานรองจากข้อความต่อไปนี้

ก. ค่าใช้จ่ายในการซื้อรองเท้าต่อปีโดยเฉลี่ยของนักศึกษาหญิงในมหาวิทยาลัยแห่งหนึ่งมากกว่า 500.- บาท

ข. นักศึกษาคณะมนุษยศาสตร์ มหาวิทยาลัยรามคำแหง เห็นด้วยกับนโยบายของรัฐบาลในการแก้ปัญหาเศรษฐกิจมากกว่าครึ่งหนึ่ง

ค. ความนิยมของผงซักฟอกบรีสมีพอ ๆ กันในทุก ๆ ภาค

ง. จำนวนบุตรมีความสัมพันธ์กับระดับรายได้ของครัวเรือน

จ. การสูบบุหรี่มีความสัมพันธ์กับโรคปอด

ฉ. เจ้าหน้าที่แผนกลงทะเบียนของมหาวิทยาลัยแห่งหนึ่งกล่าวว่าการลงทะเบียนเรียนภาคหนึ่งใช้เวลาโดยเฉลี่ยไม่น้อยกว่า 50 นาที

16. จงเปิดตารางหาค่า $t_{.025,9}$, $t_{.05,9}$, $z_{.01}$, $z_{.05}$, $\chi^2_{.01,9}$, $\chi^2_{.99,9}$

17. ในการศึกษาความสัมพันธ์ระหว่างการสูบบุหรี่และการเป็นโรคปอด โดยอาศัยตัวอย่าง 200 คน ได้ข้อมูลดังนี้

	เป็นโรคปอด	ไม่เป็นโรคปอด	รวม
สูบบุหรี่	40	80	120
ไม่สูบบุหรี่	10	70	80
รวม	50	150	200

จากข้อมูลดังกล่าวจะสรุปได้ไหมว่า การเป็นโรคปอดขึ้นอยู่กับการสูบบุหรี่ ที่ $\alpha = .05$

18. ในการศึกษาเพื่อดูว่าอัตราการสูบบุหรี่ของนักศึกษารามคำแหงปีต่าง ๆ แตกต่างกันหรือไม่ ได้สุ่มนักศึกษาปีที่ 1 มา 50 คน ปีที่ 2 มา 40 คน ปีที่ 3 มา 60 คน และปีที่ 4 มา 50 คน จัดจำแนกตามพิสัยการสูบบุหรี่ได้ผลดังนี้

	นิสัยการสูบบุหรี่			รวม
	น้อยมาก	ปานกลาง	มาก	
ปีที่ 1	21	12	17	50
ปีที่ 2	13	8	19	40
ปีที่ 3	13	18	29	60
ปีที่ 4	3	22	25	50

จงทดสอบที่ $\alpha = .05$ ว่าอัตราการสูบบุหรี่ของนักศึกษารามคำแหงทั้ง 4 ปี ไม่ต่างกัน

19. สหสัมพันธ์คืออะไร

20. สัมประสิทธิ์สหสัมพันธ์ คืออะไร มีคุณสมบัติอย่างไร และคำนวณหาสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างได้อย่างไร

21. จงคำนวณหาสัมประสิทธิ์สหสัมพันธ์จากข้อมูลแต่ละชุด ดังนี้

ก.

X	Y
34	21
30	22
40	25
34	28
39	15
35	24
42	24
45	22
43	17

$\bar{X} = 38$, $\bar{Y} = 22$

$s_X = 5$, $s_Y = 4$

ข.

X	Y
3.9	46
4.6	46
6.0	52
2.8	50
3.7	48
3.4	40
4.2	42
4.0	44

$\bar{X} = 4$, $\bar{Y} = 46$

$s_X = 1$, $s_Y = 4$.

22. จากตัวอย่าง 42 คู่ ของค่าสังเกต X และ Y คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (r) ได้เท่ากับ 0.22 จงทดสอบสมมติฐานที่ว่า $H_0: \rho = 0$, $H_1: \rho \neq 0$ เมื่อ $\alpha = 0.05$