บทที่ 8
การแจกแจงของการสุ่มตัวอย่าง (Sampling Distribution)

ในสถิติอนุมาน (Inference statistics) จุดประสงค์ที่สำคัญก็คือ การนำเอาข้อมูลที่ได้จากตัวอย่างไปอ้างอิง หรือช่วยในการตัดสินใจ หรือสรุปผลเกี่ยวกับคุณลักษณะของประชากรที่เราไม่ทราบค่า เพราะว่าเราไม่สามารถที่จะศึกษาทุก ๆ หน่วยของประชากรได้ เนื่องจากมีทรัพยากรจำกัด

การแจกแจงของการสุ่มตัวอย่างเป็นการศึกษาถึงความสัมพันธ์ระหว่างค่าที่ได้จากตัวอย่างกับค่าของประชากร ดังนั้น ถ้าเราสามารถเลือกตัวอย่างที่เป็นตัวแทนที่เหมาะสม และแสดงคุณลักษณะต่าง ๆ ของประชากรได้อย่างแท้จริง เราก็สามารถนำเอาค่าที่ได้จากการศึกษาตัวอย่างไปใช้สรุปผลเกี่ยวกับค่าของประชากรได้

8.1 การสุ่มตัวอย่าง โดยทั่ว ๆ ไปทำได้ 2 วิธี คือ

1. **การสุ่มแบบไม่แทนที่** (Sampling without replacement)

เป็นการสุ่มตัวอย่างโดยไม่เปิดโอกาสให้สมาชิกของประชากรที่ได้รับเลือกเป็นตัวอย่างแล้วได้รับเลือกซ้ำอีก เช่น การหยิบลูกบอลสีและขนาดต่าง ๆ ในกล่องใบหนึ่ง เมื่อหยิบได้เป็นสีอะไร และลูกใดแล้ว ก็จะเอาลูกนั้นออกมาจากกล่องเลย ไม่เอาใส่กลับคืนเข้าไปใหม่ ดังนั้น ค่าสังเกตที่ได้จะไม่ซ้ำกัน

2. **การสุ่มแบบแทนที่** (Sampling with replacement)

เป็นการสุ่มตัวอย่างที่เปิดโอกาสให้สมาชิกของประชากรที่ได้รับเลือกให้เป็นตัวอย่างแล้วได้รับเลือกซ้ำอีก เช่น ในการหยิบลูกบอลสีและขนาดต่าง ๆ จากกล่องใบหนึ่ง เมื่อหยิบได้เป็นสีอะไร ลูกใดแล้ว ก็จะนำเอาลูกบอลที่หยิบได้นั้นใส่ลงไปในกล่องใหม่ ก่อนที่จะหยิบลูกต่อไป ดังนั้น ค่าสังเกตที่ได้จะมีโอกาสซ้ำกันได้

8.2 การแจกแจงของการสุ่มตัวอย่าง (Sampling Distribution)

การเลือกตัวอย่างจากประชากร เป็นการทดลองเชิงสุ่มอย่างหนึ่งซึ่งมี Sample Space ที่ประกอบด้วยตัวอย่างที่จะเป็นไปได้ทั้งหมดตามขนาดที่กำหนดไว้ โดยเลือกมาจากประชากรที่เราสนใจ ตัวสถิติที่ได้มาจากตัวอย่างแต่ละชุด ก็จะเป็นตัวแปรเชิงสุ่มตามตัวอย่างด้วย และมีค่าตามตัวอย่าง หรือผลการทดลองที่กำหนด ดังนั้นตัวสถิติจึงมีการแจกแจงอย่างหนึ่งและการแจกแจงของตัวสถิตินี้เรียกว่า การแจกแจงของการสุ่มตัวอย่าง (Sampling distribution) สำหรับส่วนเบี่ยงเบนมาตรฐานของตัวสถิติจะเรียกว่า ความคลาดเคลื่อนมาตรฐานของตัวสถิติ (Standard error of the statistics) และการแจกแจงของค่าเฉลี่ยของตัวอย่าง ($\overline{X}$) เรียกว่า การแจกแจงของการสุ่มตัวอย่างของค่าเฉลี่ย

8.3 การแจกแจงของการสุ่มตัวอย่างของค่าเฉลี่ย (Sampling distribution of the Mean)

จากประชากรขนาด N ซึ่งมีค่าเฉลี่ย = μ และส่วนเบี่ยงเบนมาตรฐาน = σ สุ่มตัวอย่างมาขนาด n ซึ่งอาจจะสุ่มแบบแทนที่ (Sampling with replacement) หรือสุ่มแบบไม่แทนที่ (Sampling without replacement) ก็ได้ จากตัวอย่างแต่ละแบบที่สุ่มได้นำมาคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานได้ ซึ่งจะแทนด้วย $\overline{X}$ และ S ตามลำดับ

สมมติว่าเลือกตัวอย่างได้ k แบบ แต่ละแบบก็คำนวณหา $\overline{X}$ ออกมา ดังนั้นจะมี $\overline{X}$ อยู่ k ตัว คือ $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k$ ซึ่ง $\overline{X}$ เหล่านี้ เราเรียกว่า Sample Mean การศึกษาการแจกแจงของ $\overline{X}$ ก็คือ ดูว่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ $\overline{X}$ เหล่านี้เป็นอย่างไร มีความสัมพันธ์กับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประชากรอย่างไรบ้าง โดยใช้ $\mu_{\bar{x}}$ แทนค่าเฉลี่ยของ $\overline{X}$ และ $\sigma_{\bar{x}}$ แทนความคลาดเคลื่อนมาตรฐานของ $\overline{X}$ ซึ่งการคำนวณหา $\mu_{\bar{x}}$ และ $\sigma_{\bar{x}}$ หาได้โดยวิธีเดียวกับที่ได้กล่าวมาแล้ว ดังนี้

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

$$= \frac{\sum_{i=1}^{k} \overline{X}_i}{k} \qquad \ldots\ldots\ldots\ldots(1)$$

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

$$= \sqrt{\frac{\sum_{i=1}^{k}(\bar{X}_i-\mu_{\bar{x}})^2}{k}} \qquad ....(2)$$

การหาค่าของ $\mu_{\bar{x}}$ และ $\sigma_{\bar{x}}$ เป็นงานที่ยุ่งยากมาก ในที่นี้จะไม่พิสูจน์ให้ดูแต่จะแสดงให้เห็นด้วยตัวอย่างซึ่งจะกล่าวต่อไป

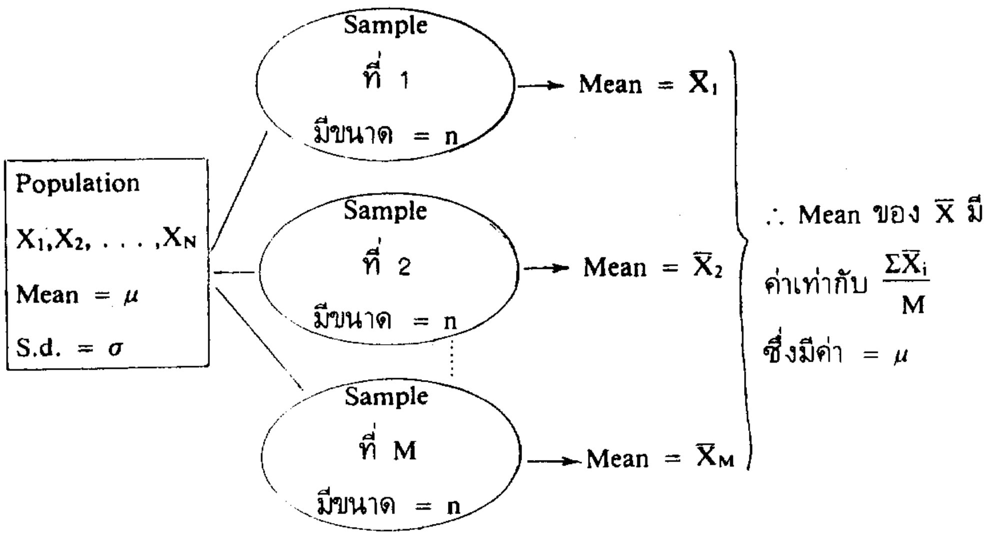
จาก (1) จะได้ว่า $\mu_{\bar{x}} = \mu$

ซึ่งหมายความว่าค่าเฉลี่ยของ $\bar{X}$ มีค่าเท่ากับค่าเฉลี่ยของประชากรและจาก (2) จะได้ว่า

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

และ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ ในกรณีที่ประชากรเป็นแบบไม่จำกัด (Infinite population) หรือเป็นแบบจำกัดและการเลือกตัวอย่างเป็นแบบแทนที่

รูปแสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของตัวอย่างและประชากร



จากตัวอย่างต่อไปนี้จะแสดงให้เห็นว่า $\mu_{\bar{x}} = \mu$ และ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ หรือ

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

ตัวอย่าง

มีเลขอยู่ 4 จำนวนคือ 4, 8, 13 และ 11 ต้องการสุ่มตัวอย่างขนาด $n=2$ แบบแทนที่และแบบไม่แทนที่ 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

ได้ค่าเฉลี่ย และความแปรปรวนของประชากรดังนี้

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

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{(4-9)^2(8-9)^2+(13-9)^2+(11-9)^2}{4}$$

$$= \frac{25+1+16+4}{4} = \frac{46}{4}$$

$$= \frac{23}{2}$$

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

ก) ถ้าเลือกตัวอย่างแบบไม่แทนที่ (without replacement) เราจะได้จำนวนตัวอย่างทั้งหมด ${}^{N}C_{n}$ ซึ่งในที่นี้จะได้เท่ากับ ${}^{4}C_{2} = \frac{4!}{2!2!} = \frac{4\times 3}{2} = 6$ ตัวอย่าง ซึ่งได้แก่ (4,8),(4,13),(4,11),(8,13),(8,11),(13,11) เราต้องหาค่าเฉลี่ยของตัวอย่างแต่ละตัวอย่างซึ่งจะได้ค่าเฉลี่ยของตัวอย่างทั้งหมด 6 ค่าด้วยกันดังนี้

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

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

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

สำหรับตัวอย่าง (8,13) ได้ $\bar{X}_4 = \frac{8+13}{2} = 10.5$

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

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

$$\therefore \mu_{\bar{x}} = \frac{\sum_{i=1}^{6} \bar{X}_i}{6} = \frac{\bar{X}_1+\bar{X}_2+\bar{X}_3+\bar{X}_4+\bar{X}_5+\bar{X}_6}{6}$$

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

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

ดังนั้น $\mu_{\bar{x}} = \mu$

ข. ถ้าเลือกตัวอย่างแบบแทนที่ (With replacement) เราจะได้จำนวนตัวอย่างทั้งหมด = N^n ตัวอย่าง ซึ่งในที่นี้จำนวนตัวอย่างที่ได้เท่ากับ $4^2 = 16$ ตัวอย่าง ซึ่งได้แก่

(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)

ดังนั้น ค่าเฉลี่ยของตัวอย่างจะมีทั้งหมด 16 ค่าด้วยกันดังนี้

$\bar{X}_1 = \frac{4+4}{2} = 4$; $\bar{X}_2 = \frac{4+8}{2} = 6$

$\bar{X}_3 = \frac{4+13}{2} = 8.5$; $\bar{X}_4 = \frac{4+11}{2} = 7.5$

$\bar{X}_5 = \frac{8+14}{2} = 6$; $\bar{X}_6 = \frac{8+8}{2} = 8$

$\bar{X}_7 = \frac{8+13}{2} = 10.5$; $\bar{X}_8 = \frac{8+11}{2} = 9.5$

$\bar{X}_9 = \frac{13+4}{2} = 8.5$; $\bar{X}_{10} = \frac{13+8}{2} = 10.5$

$\bar{X}_{11} = \frac{13+13}{2} = 13$; $\bar{X}_{12} = \frac{13+11}{2} = 12$

$\bar{X}_{13} = \frac{11+4}{2} = 7.5$; $\bar{X}_{14} = \frac{11+8}{2} = 9.5$

$\bar{X}_{15} = \frac{11+13}{2} = 12$; $\bar{X}_{16} = \frac{11+11}{2} = 11$

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

$$= \frac{4+6+8.5+\ldots+12+11}{16}$$

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

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

แสดงว่าไม่ว่าจะเลือกตัวอย่างแบบแทนที่หรือแบบไม่แทนที่จะได้

$\mu_{\bar{X}} = \mu$ เสมอ

2. ต้องการแสดงให้เห็นว่า

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ เมื่อเลือกตัวอย่างแบบไม่แทนที่ (without replacement)

ตัวอย่างที่เลือกได้เท่ากับ ${}^{N}C_{n}$ ซึ่งเท่ากับ ${}^{4}C_{2} = 6$ แบบ

$$\text{จาก } \sigma^2_{\bar{X}} = \frac{\sum_{i=1}^{6} (\bar{X}_i - \mu_{\bar{X}})^2}{6}$$

$$\mu_{\bar{X}} = 9$$

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

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

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

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

ซึ่งถ้าพิจารณา $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ หรือ $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ จะได้

$$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{23}{2}\times\frac{1}{2}\left(\frac{4-2}{4-1}\right) = \left(\frac{23}{4}\right)\left(\frac{2}{3}\right)$$

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

$$\therefore \quad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

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

3. ต้องการแสดงให้เห็นว่า

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ เมื่อเลือกตัวอย่างแบบแทนที่ (with replacement) ซึ่งตัวอย่างที่เลือกได้มีจำนวน $N^n = 4^2 = 16$ ตัวอย่าง

จาก $\quad \sigma^2_{\bar{x}} = \frac{\sum_{i=1}^{6}(\bar{X}_i-\mu_{\bar{x}})^2}{16}$

เมื่อ $\quad \mu_{\bar{x}} = \mu = 9$

$$\therefore \quad \sigma^2_{\bar{x}} = \frac{1}{16}[(4-9)^2+(6-9)^2+(8.5-9)^2+\ldots+(12-9)^2+(11-9)^2]$$

$$= \frac{1}{16}[25+9+0.25+\ldots+9+4]$$

$$= \frac{92}{16} = \frac{23}{4}$$

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

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

$$\therefore \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

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

จากตัวอย่างข้างต้นเราได้ทฤษฎีดังนี้

ทฤษฎี ถ้า $\bar{X}$ คือค่าเฉลี่ยของตัวอย่างสุ่มขนาด n จากประชากรที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้วเมื่อ n มีขนาดโต $(n \to \infty)$ ตัวสถิติ $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ จะมีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ย = 0 และความแปรปรวน = 1

ทฤษฎี ถ้า X มีการแจกแจงซึ่งมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ และถ้าเลือกตัวอย่างขนาด n จากประชากรนี้จะได้ $\bar{X}$ มีการแจกแจงเข้าใกล้การแจกแจงแบบปกติซึ่งมีค่าเฉลี่ยเป็น $\mu_{\bar{x}}$ 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 $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

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

$$\therefore \quad Z = \frac{\bar{X}-\mu}{\sigma_{\bar{x}}}$$

สำหรับการหาความน่าจะเป็นที่ตัวแปรเชิงสุ่มจะมีค่าต่าง ๆ นั้น ก็ใช้วิธีการคำนวณหาแบบเดียวกับการหาความน่าจะเป็นของการแจกแจงแบบปกติ

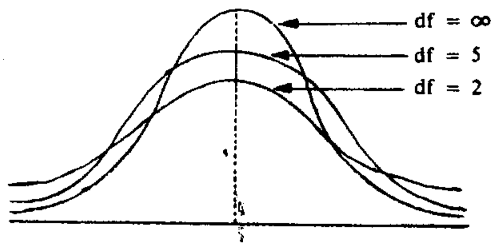
โดยปกติแล้วเรามักจะไม่ทราบความแปรปรวนของประชากร σ^2 ที่เราเลือกตัวอย่างสุ่มมา สำหรับตัวอย่างที่มีขนาดโต $(n \geqslant 30)$ แล้วเราจะใช้ความแปรปรวนของตัวอย่าง (S^2) มาเป็นตัวประมาณค่า σ^2

ดังนั้น

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$$

ซึ่ง $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ ก็ยังคงมีการแจกแจงแบบปกติมาตรฐาน แต่ถ้าตัวอย่างที่สุ่มมามีขนาดเล็ก ($n < 30$) $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ จะไม่มีการแจกแจงแบบปกติมาตรฐานแต่ $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ จะมีการแจกแจงที่เรียกว่าการแจกแจงแบบที (Student's t distribution) ซึ่งการแจกแจงของ $t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ จะคล้ายกับการแจกแจงของ Z คือโค้งเป็นรูประฆังคว่ำ และสมมาตรกันรอบค่าเฉลี่ยซึ่งมีค่าเป็นศูนย์ แต่ t มีความแปรปรวนมากกว่าซึ่งทำให้โค้งของ t แบนกว่า และค่าของ t ขึ้นอยู่กับค่าเฉลี่ยของตัวอย่าง ($\bar{X}$) และความแปรปรวนของตัวอย่าง (S^2) แต่ค่าของ Z ขึ้นอยู่กับค่าเฉลี่ยของตัวอย่าง ($\bar{X}$) เพียงค่าเดียวเท่านั้น แต่อย่างไรก็ตามเมื่อตัวอย่างมีขนาดโต ($n > 30$) 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 ν ดังนั้นการแจกแจงของ t จะมีรูปร่างต่างกัน ถ้าองศาความเป็นอิสระต่างกัน ดังรูป



รูปแสดงการแจกแจงแบบ t เมื่อค่า d.f. ต่างกัน

ทฤษฎี

ถ้า $\overline{X}$ และ S^2 เป็นค่าเฉลี่ยและความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ ถ้าไม่ทราบ σ^2 ดังนั้น

$$t = \frac{\overline{X}-\mu}{S/\sqrt{n}} \text{ จะมีการแจกแจงแบบ t}$$

ที่มีองศาความเป็นอิสระ $(df) = n-1$

8.4 การแจกแจงการสุ่มตัวอย่างของผลต่างของค่าเฉลี่ย 2 ค่า

(Sampling distribution of the differences of two means)

สมมติว่าเรามีประชากร 2 กลุ่ม ประชากรกลุ่มแรกมีค่าเฉลี่ย μ_1 และความแปรปรวน σ_1^2 ประชากรกลุ่มที่ 2 มีค่าเฉลี่ย μ_2 และความแปรปรวน σ^2_2 ถ้าให้ตัวแปร $\overline{X}_1$ แทนค่าเฉลี่ยของตัวอย่างสุ่มขนาด n_1 ที่เลือกมาจากประชากรกลุ่มแรก และให้ตัวแปร $\overline{X}_2$ แทนค่าเฉลี่ยของตัวอย่างสุ่มขนาด n_2 ที่เลือกมาจากประชากรกลุ่มที่ 2 โดยที่ตัวอย่างเหล่านี้เป็นอิสระกัน ค่าเฉลี่ยของผลต่างระหว่างค่าเฉลี่ยของตัวอย่างทั้งสอง $\mu_{\bar{x}_1-\bar{x}_2}$ จะมีค่าเท่ากับผลต่างของค่าเฉลี่ยของประชากรทั้งสอง $\mu_1-\mu_2$ และความแปรปรวนของผลต่างของค่าเฉลี่ยของตัวอย่าง $\sigma^2_{\bar{x}_1-\bar{x}_2}$ จะมีค่าเท่ากับผลบวกของค่าความแปรปรวนของตัวอย่างทั้งสองคือ $\sigma^2_{\bar{x}_1}-\sigma^2_{\bar{x}_2}$ และนอกจากนี้รูปร่างของการแจกแจงของ $\overline{X}_1-\overline{X}_2$ ยังมีลักษณะใกล้เคียงกับการแจกแจงปกติซึ่งถ้า n_1 และ n_2 มีขนาดโต การแจกแจงของ $\overline{X}_1-\overline{X}_2$ จะเป็นแบบปกติ ดังนั้น

$$Z = \frac{(\overline{X}_1-\overline{X}_2)-\mu_{\bar{x}_1-\bar{x}_2}}{\sigma_{\bar{x}_1-\bar{x}_2}} \text{ จะมีการแจกแจงแบบปกติมาตรฐาน } N(0, 1)$$

เมื่อ n_1 และ n_2 มีขนาดโต $(n_1, n_2 \geqslant 30)$

ตามปกติแล้วค่า σ_1^2 และ σ_2^2 เรามักไม่ค่อยทราบค่า ดังนั้นถ้าเราไม่ทราบค่า σ_1^2 และ σ_2^2 แต่ทราบว่าเท่ากัน นั่นคือ $\sigma_1^2 = \sigma_2^2 = \sigma^2$ เราจะได้

$$Z = \frac{(\overline{X}_1-\overline{X}_2)-\mu_{\bar{x}_1-\bar{x}_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

หรือ $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

ซึ่ง σ^2 เราสามารถประมาณได้จากการรวมความแปรปรวนของตัวอย่างทั้งสอง ซึ่งคือ S_1^2 และ S_2^2 และความแปรปรวนรวมที่ประมาณได้ เราให้เป็น S_p^2 โดยที่

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

เมื่อแทนค่า σ^2 และ S_p^2 แล้วเราจะได้ตัวสถิติ T ซึ่งจะมีการแจกแจงแบบ t ที่มีองศาความเป็นอิสระ $n_1 + n_2 - 2$ ดังทฤษฎีต่อไปนี้

ทฤษฎี

ถ้า $\bar{X}_1$, $\bar{X}_2$ และ S_1^2, S_2^2 เป็นค่าเฉลี่ย และความแปรปรวนของตัวอย่างสุ่มที่เป็นอิสระกันขนาด n_1 และ n_2 จากประชากรที่มีค่าเฉลี่ย μ_1, μ_2 และความแปรปรวนที่ไม่ทราบค่า σ_1^2, σ_2^2 แต่ทราบว่าความแปรปรวนทั้ง 2 เท่ากันคือ $\sigma_1^2 = \sigma_2^2 = \sigma^2$ แล้ว

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ จะมีการแจกแจงแบบ t ที่มีองศาความเป็นอิสระ $n_1 + n_2 - 2$

ในกรณีที่เราไม่ทราบความแปรปรวนของประชากรทั้ง 2 (σ_1^2 และ σ_2^2) และทราบว่าไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$) ตัวสถิติ T′ คือ

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n_2}}}$$

จะมีการแจกแจงแบบ t (โดยประมาณ) และมีองศาความเป็นอิสระดังนี้

$$d.f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n-1)} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2-1)}}$$

ทั้ง 3 กรณีที่กล่าวมาข้างต้น เป็นกรณีที่ตัวอย่างที่สุ่มมาเป็นอิสระกันแต่ถ้าตัวอย่างที่สุ่มมาไม่เป็นอิสระกัน หรือในกรณีที่ข้อมูลจากตัวอย่างจับคู่กัน (Paired observation)

ถ้าให้ $(X_{11}, X_{12}), (X_{21}, X_{22}) \ldots (X_{n1}, X_{n2})$ เป็นค่าสังเกต n คู่ ของตัวอย่างที่ไม่เป็นอิสระกันขนาด n ซึ่งเลือกมาจากประชากร (ซึ่งความแปรปรวนของประชากรจะเท่ากันหรือไม่ก็ได้)

และให้ d_i เป็นผลต่างระหว่างค่าสังเกตในคู่ที่ i ดังนั้น

$$d_i = X_{i1} - X_{i2} \; ; \; i = 1, 2, \ldots, n$$

ถ้าเราสมมติว่า d_i ต่าง ๆ ประกอบเป็นตัวอย่างสุ่มจากประชากรแบบปกติที่มีค่าเฉลี่ย $\mu_d = \mu_1 - \mu_2$ แล้วตัวสถิติ T คือ

$$T = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}}$$ จะมีการแจกแจงแบบ t ที่มีองศาความเป็นอิสระเท่ากับ $n-1$ เมื่อ $\bar{d}$ เป็นค่าเฉลี่ยของ d_i และ S_d^2 เป็นความแปรปรวนของ d_i ดังนี้

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

$$S_d^2 = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right]$$

8.5 การแจกแจงของการสุ่มตัวอย่างของสัดส่วน

(Sampling distribution of Proportion)

จากการทดลองแบบทวินามที่ประกอบด้วย n การทดลองที่เป็นอิสระต่อกัน โดยมีความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ π และความน่าจะเป็นที่จะเกิดความไม่สำเร็จเท่ากับ $1-\pi$

ถ้า X เป็นจำนวนครั้งที่เกิดความสำเร็จในการทดลอง n ครั้ง X จะเป็นตัวแปรเชิงสุ่มที่มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยและความแปรปรวนเป็นดังนี้คือ

$$\mu = n\pi$$

และ $$\sigma^2 = n\pi(1-\pi)$$

ซึ่งถ้าเราพิจารณาสัดส่วนที่จะเกิดความสำเร็จในการทดลอง n ครั้ง โดยให้

$P = \frac{x}{n}$ (คือสัดส่วนที่สำเร็จ) และเราจะได้ P เป็นตัวแปรสุ่มที่มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติโดยมีค่าเฉลี่ย และความแปรปรวนดังนี้

$$\mu_p = \pi$$

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n} \text{ ซึ่งจะได้ } \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

ดังนั้น $$Z = \frac{P-\pi}{\sigma_p}$$

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

จะมีการแจกแจงแบบปกติมาตรฐาน N(0, 1)

ทฤษฎี

ตัวอย่างสุ่มขนาด n สุ่มมาจากประชากรที่มีการแจกแจงแบบทวินามที่มีค่าเฉลี่ย $\mu = n\pi$ และความแปรปรวน $= n\pi(1-\pi)$ สัดส่วนที่เกิดความสำเร็จ $P = \frac{x}{n}$ จะมีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติโดยมีค่าเฉลี่ย $\mu_p = \pi$ และความแปรปรวน $\sigma_p^2 = \frac{\pi(1-\pi)}{n}$ ดังนั้น $Z = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ จะมีการแจกแจงแบบปกติมาตรฐาน N(0, 1)

8.6 การแจกแจงการสุ่มตัวอย่างของผลต่างของสองสัดส่วน

(Sampling distribution of difference of two proportion)

พิจารณาการสุ่มตัวอย่างขนาด n_1 และ n_2 ซึ่งเป็นอิสระกัน จากประชากรแบบทวินาม 2 ประชากร ที่มีค่าเฉลี่ย $\mu_1 = n_1 \pi_1$ และ $\mu_2 = n_2 \pi_2$ และมีความแปรปรวน $\sigma_1^2 = n_1 \pi_1 (1-\pi_1)$ และ $\sigma_2^2 = n_2 \pi_2(1-\pi_2)$ ตามลำดับ การแจกแจงของ $P_1 - P_2$ จะมีการแจกแจงใกล้เคียงกับการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติที่มีค่าเฉลี่ยและความแปรปรวนดังนี้

$$\mu_{p_1-p_2} = \pi_1 - \pi_2$$

$$\sigma^2_{p_1-p_2} = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

ดังนั้น $$Z = \frac{(P_1 - P_2) - \mu_{p_1-p_2}}{\sigma_{p_1-p_2}}$$ หรือ

$$Z = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

จะมีการแจกแจงแบบปกติมาตรฐาน N(0, 1)

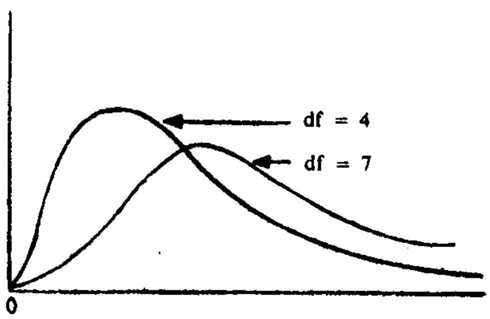
ซึ่งถ้าขนาดของตัวอย่างโต (n_1 และ $n_2 > 30$) การแจกแจงของ $P_1 - P_2$ จะใกล้การแจกแจงปกติมาตรฐาน

8.7 การแจกแจงการสุ่มตัวอย่างของความแปรปรวน

(Sampling distribution of sample variance)

ถ้า S^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากร แบบปกติที่มีความแปรปรวน σ^2 แล้วตัวแปรเชิงสุ่ม $\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$ จะมีการแจกแจงแบบไคส-แควร์ (Chi-square distribution) ที่มีองศาความเป็นอิสระ $n-1$ การแจกแจงของ χ^2 จะเบ้ไปทางขวา ดังนั้นค่าของ χ^2 จะเป็นบวกเสมอ และการแจกแจงของ χ^2 จะไม่สมมาตรกันรอบค่าศูนย์ χ^2 จะมีค่าตั้งแต่ 0 ถึง ∞

และรูปร่างของ χ^2 จะขึ้นอยู่กับขนาดของตัวอย่างเช่นเดียวกับการแจกแจงของ t ดังรูป



รูปแสดงการแจกแจงแบบ χ^2 ที่มี df ต่างกัน

แบบฝึกหัด

1. การสุ่มตัวอย่างแบบแทนที่ และการสุ่มตัวอย่างแบบไม่แทนที่คืออะไร
2. จงให้ความหมายของคำว่า Sampling distributron
3. การแจกแจงแบบทีมีลักษณะอย่างไร
4. การแจกแจงแบบไคสแควร์มีลักษณะอย่างไร
5. โรงงานผลิตหลอดไฟฟ้าแห่งหนึ่ง ถ้าอายุการใช้งานของหลอดไฟมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 800 ชั่วโมง และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 40 ชั่วโมง จงหาความน่าจะเป็นที่หลอดไฟที่สุ่มได้ 16 หลอดจะมีอายุการใช้งานเฉลี่ยน้อยกว่า 775 ชั่วโมง
6. แบตเตอรี่ ที่ผลิตจากบริษัท ก. มีอายุการใช้งานเฉลี่ยเท่ากับ 6.5 ปี ส่วนเบี่ยงเบนมาตรฐาน 0.9 ปี ของบริษัท ข. มีอายุการใช้งานเฉลี่ย 6.0 ปี และส่วนเบี่ยงเบนมาตรฐาน 0.8 ปี จงหาความน่าจะเป็นที่ตัวอย่างแบตเตอรี่ที่สุ่มได้จากบริษัท ก. จำนวน 36 อันจะมีอายุการใช้งานเฉลี่ยมากกว่าบริษัท ข. ซึ่งใช้ตัวอย่างที่สุ่มมา 49 อัน อย่างน้อยที่สุด 1 ปี
7. ให้ความน่าจะเป็นที่เมล็ดพืชชนิดหนึ่งจะงอกเท่ากับ 60% จงหาความน่าจะเป็นที่เมล็ดพืช 100 เมล็ดจะงอกมากกว่า 75 %
8. ในการทดลองฉีดวัคซีนป้องกันโรคชนิดหนึ่งให้กับสัตว์ทดลองพบว่าสัตว์ที่ได้รับการฉีดวัคซีนจะตายด้วยโรค 5% และสัตว์ที่ไม่ได้ฉีดวัคซีนจะตายด้วยโรค 12% จงหาความน่าจะเป็นที่สัตว์ที่ฉีดวัคซีนจะตายน้อยกว่าสัตว์ที่ไม่ได้ฉีดวัคซีนมากกว่า 10% ถ้าตัวอย่างสัตว์ที่สุ่มมามีจำนวนเท่า ๆ กันคือ 150 ตัว
9. สุ่มตัวอย่างขนาด $n_1 = 10$ จากประชากรที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ย $\bar{x}_1 = 20$ และส่วนเบี่ยงเบนมาตรฐาน $s_1 = 5$ อีกตัวอย่างหนึ่งสุ่มมา $n_2 = 12$ จากประชากรที่มีการแจกแจงแบบปกติอีกประชากรหนึ่งได้ค่าเฉลี่ย $\bar{x}_2 = 24$ และส่วนเบี่ยงเบนมาตรฐาน $s_2 = 6$ ถ้า $\mu_1 = 22$ และ $\mu_2 = 19$ และไม่ทราบค่า σ_1^2, σ_2^2 แต่ทราบว่า $\sigma_1^2 = \sigma_2^2$ จงหาค่า

10. x เป็นตัวแปรเชิงสุ่มที่มีการแจกแจงแบบไคสแควร์ ซึ่งมีองศาความเป็นอิสระเท่ากับ 23 จงหาค่าของ

ก) $P\left[10.20 < x < 35.17\right]$

ข) a และ b ที่ทำให้ $P\left[a < x < b\right] = 0.95$

และ $P\left[x < a\right] = 0.025$

11. อายุต่าง ๆ ของเด็กในครอบครัวที่มีเด็ก 4 คนเป็นดังนี้

$x_1 = 2$ ปี

$x_2 = 4$ ปี

$x_3 = 6$ ปี

$x_4 = 8$ ปี

ก) จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอายุเด็ก 4 คน ดังกล่าว

ข) จงหาตัวอย่างของอายุของเด็กดังกล่าวที่จะเป็นไปได้ทั้งหมด โดยเลือกเด็กมาครั้งละ 2 คน และให้หาค่าเฉลี่ยของแต่ละตัวอย่างด้วย

ค) จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวอย่างทั้ง 6 ตัวอย่าง ในข้อ (ข)

ง) จงแสดงให้เห็นว่า $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$