<u>บทที่ 2</u>

<u>ทฤษฎีที่เกี่ยวข้องกับงานวิจัย</u>

2.1 <u>ทฤษฎีที่ใช้ในการวิจัย</u>

ทฤษฎีทางสถิติที่ใช้ในการวิจัยมีดังต่อไปนี้

2.1.1 ถ้าตัวอย่างขนาด n ถูกสุ่มมาจากประชากรที่มีการแจกแจงซึ่งมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ แล้ว ค่าเฉลี่ยตัวอย่างจะมีการแจกแจง ซึ่งมีค่าเฉลี่ยเท่ากับ μ และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\sigma/\sqrt{n}$ สำหรับการสุ่มตัวอย่างแบบแทนที่ (sampling with replacement) และเท่ากับ $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ สำหรับการสุ่มตัวอย่างแบบไม่แทนที่ (sampling without replacement)

2.1.2 ถ้าตัวอย่างขนาด n ถูกสุ่มมาจากประชากรที่มีการแจกแจงซึ่งมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ แล้ว การแจกแจงของค่าเฉลี่ยตัวอย่าง ($\bar{y}$) ประมาณได้ด้วยการแจกแจงปกติที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐานเป็น $\sigma/\sqrt{n}$ เรียกทฤษฎีนี้ว่า "ทฤษฎีแนวโน้มเข้าสู่ส่วนกลาง" (central limit theorem)

2.1.3 กำหนดให้ค่าเฉลี่ยตัวอย่าง $\bar{y} = \sum_{i=1}^{n} y_i/n$ เป็นตัวประมาณที่ไม่เอนเอียง (unbiased estimator) ของค่าเฉลี่ยประชากร $\bar{Y}$ และกำหนดให้การเลือกสุ่มตัวอย่างเป็นวิธีสุ่มตัวอย่างแบบง่าย ความแปรปรวนของ $\bar{y}$ คือ

$$V(\bar{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

ในกรณีที่ไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐาน (σ_y) ของประชากร จะต้องประมาณค่า σ_y ด้วยความคลาดเคลื่อนของตัวอย่าง (s_y) ตัวประมาณที่ไม่เอนเอียงของ $V(\bar{y})$ คือ

$$v(y) = \frac{s_y^2}{n}(1-f)$$

เมื่อ $f = n/N$

2.1.4 กำหนดให้ $p = \frac{a}{n}$ เป็นตัวประมาณที่ไม่เอนเอียงของ $P = \frac{A}{N}$ และกำหนดให้วิธีเลือกตัวอย่างเป็นวิธีสุ่มตัวอย่างแบบง่ายแล้ว ความแปรปรวนของ p คือ

$$V(p) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

ในกรณีที่ไม่ทราบค่า P จะต้องประมาณค่า P ด้วย p ตัวประมาณที่ไม่เอนเอียงของ V(p) คือ

$$v(p) = \frac{pq}{n-1}(1-f)$$

เมื่อ $f = n/N$

2.2 ลักษณะการแจกแจงของประชากรที่เกี่ยวข้องกับการวิจัย

2.2.1 การแจกแจงแบบไฮเปอร์ยีออเมตริก (hypergeometric distribution)

การแจกแจงความน่าจะเป็นของค่าสัดส่วนจากตัวอย่าง (p) หาได้จากการแจกแจงความน่าจะเป็นของจำนวนตัวอย่างซึ่งมีลักษณะที่สนใจ (a = np) ซึ่งพิจารณาได้ดังต่อไปนี้

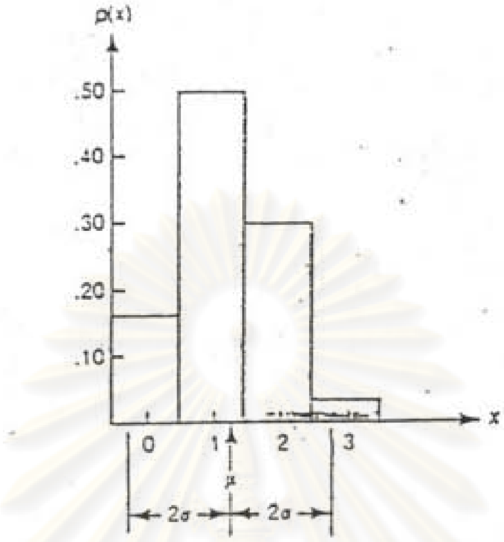
ให้ x แทนจำนวนผลสำเร็จในการสุ่มตัวอย่างขนาด n ที่ไม่ซ้ำกันจากของทั้งหมด N สิ่ง ซึ่งมี k สิ่งที่มีลักษณะซึ่งสนใจหรือเป็นผลสำเร็จ เรียก x ว่าเป็นตัวแปรสุ่มไฮเปอร์ยีออเมตริก (hypergeometric random variable) เขียนแทนด้วย $x \sim H(N,k,n)$ โดยมีฟังก์ชันความหนาแน่นอยู่ในรูป

$$P(X = x) = \frac{\binom{n}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

โดยที่ $x = 0,1,2,\ldots,n$ เมื่อ $n \leqslant k$

และ $x = 0,1,2,\ldots,k$ เมื่อ $n > k$

ลักษณะการกระจายของข้อมูลที่มีการแจกแจงแบบไฮเปอร์ยีออเมตริก แสดงในรูปที่ 2.2.1.1



รูปที่ 2.2.1.1

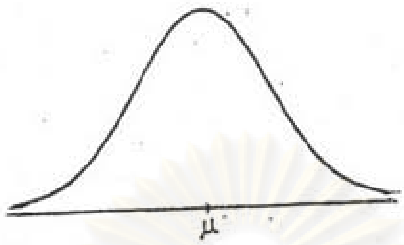
2.2.2 การแจกแจงแบบปกติ (normal distribution)

เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่องที่สำคัญและใช้มากที่สุดในทางสถิติ การแจกแจงแบบปกติมีลักษณะที่เห็นเด่นชัดคือ เส้นโค้งของการแจกแจงเป็นรูปโค้งสมมาตรคล้ายระฆัง (bell-shaped) ซึ่งเรียกว่าโค้งปกติ (normal curve) การแจกแจงแบบนี้ใช้อธิบายลักษณะของข้อมูลที่เกิดขึ้นเองตามธรรมชาติ

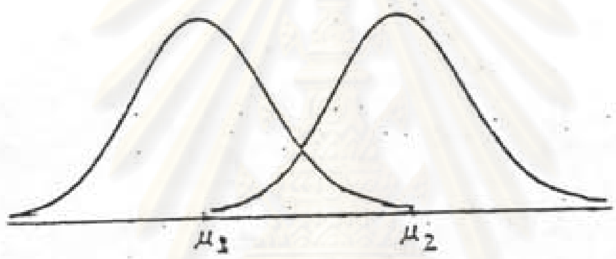
ตัวแปรสุ่ม x ที่มีการแจกแจงแบบปกติ และมีค่าเฉลี่ยและความแปรปรวนเท่ากับ μ และ σ^2 ตามลำดับ เขียนแทนด้วย $x \sim N(\mu, \sigma^2)$ ฟังก์ชันความหนาแน่นของ x อยู่ในรูป

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(1/2)((x-\mu)/\sigma)^2] \quad ; -\infty < x < \infty \ , \ \mu \in R \ , \ \sigma^2 > 0$$

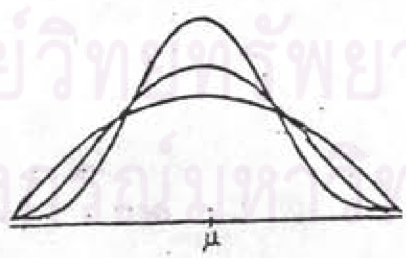
ลักษณะการกระจายของข้อมูลที่มีการแจกแจงแบบปกติ แสดงในรูปต่อไปนี้



รูปที่ 2.2.2.1 แสดงการแจกแจงแบบปกติของประชากร 1 กลุ่ม



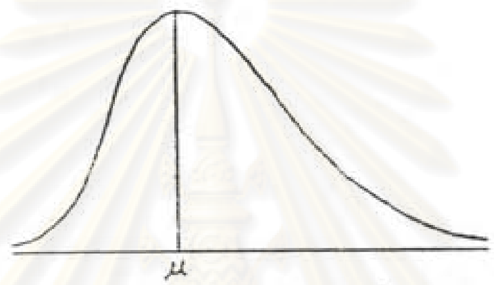
รูปที่ 2.2.2.2 แสดงการแจกแจงแบบปกติของประชากร 2 กลุ่ม ที่มีค่าเฉลี่ยต่างกัน และมีส่วนเบี่ยงเบนมาตรฐานเท่ากัน



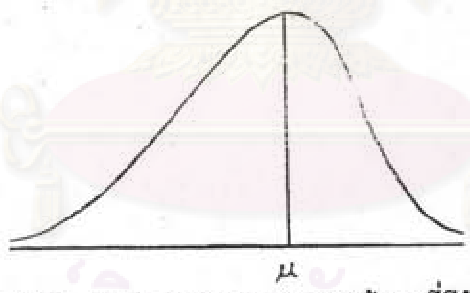
รูปที่ 2.2.2.3 แสดงการแจกแจงแบบปกติของประชากร 3 กลุ่ม ที่มีค่าเฉลี่ยเท่ากัน และมีส่วนเบี่ยงเบนมาตรฐานต่างกัน

2.2.3 การแจกแจงแบบเบ้ (skewed distribution)

การแจกแจงแบบเบ้เป็นการแจกแจงของข้อมูลที่มีลักษณะไม่สมมาตร ข้อมูลส่วนใหญ่จะกระจายไปในทิศทางใดทิศทางหนึ่งจากจุดแทนค่าเฉลี่ยของข้อมูล การแจกแจงจะมีลักษณะเบ้ขวา (possitive skewed) ถ้าข้อมูลส่วนใหญ่ที่พิจารณากระจายไปทางขวาของค่าเฉลี่ย และการแจกแจงจะมีลักษณะเบ้ซ้าย (negative skewed) ถ้าข้อมูลส่วนใหญ่ที่พิจารณากระจายไปทางซ้ายของค่าเฉลี่ย ลักษณะการกระจายของข้อมูลแบบเบ้ขวาและแบบเบ้ซ้ายแสดงดังในรูปที่ 2.2.3.1 และ รูปที่ 2.2.3.2 ตามลำดับ



รูปที่ 2.2.3.1 แสดงการกระจายของข้อมูลที่มีลักษณะเบ้ขวา



รูปที่ 2.2.3.2 แสดงการกระจายของข้อมูลที่มีลักษณะเบ้ซ้าย

ความเบ้ของข้อมูลวัดได้จากการหาสัมประสิทธิ์แห่งความเบ้ เขียนแทนด้วย

$$\alpha_3 = \frac{1}{\sigma^3} E(x-\mu)^3$$
$$= \mu_3/\sigma^3$$

โดยที่ μ แทน ค่าเฉลี่ยประชากร

μ_3 แทน เซ็นทรัลโมเมนต์ที่ 3 = $E(x-\mu)^3$

σ แทน ส่วนเบี่ยงเบนมาตรฐาน

α_3 ใช้เปรียบเทียบความเบ้ระหว่างข้อมูลแต่ละชุด ซึ่งไม่มีหน่วย ถ้า α_3 มีค่าเป็นบวกแสดงว่าข้อมูลมีลักษณะเบ้ขวา และถ้า α_3 มีค่าเป็นลบแสดงว่าข้อมูลมีลักษณะเบ้ซ้าย ความเบ้จะมากหรือน้อยขึ้นอยู่กับค่าของ α_3 ในทางปฏิบัติข้อมูลที่พบโดยส่วนมากมักมีลักษณะเบ้ขวา ในที่นี้จึงขอยกตัวอย่างการแจกแจงบางรูปแบบที่มีลักษณะเบ้ขวาเพียงอย่างเดียว

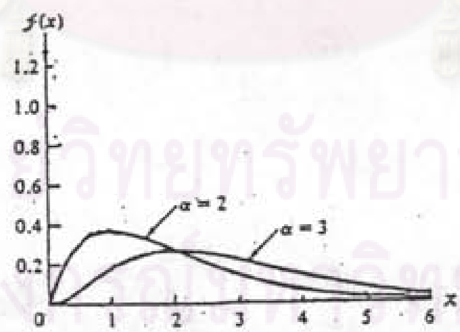
การแจกแจงที่มีลักษณะเบ้ขวาได้แก่

ก. การแจกแจงแบบแกมมา (gamma distribution) มีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} x^{\alpha-1} \dfrac{\exp(-x/\beta)}{\beta^{\alpha}\,\Gamma(\alpha)} & ;\ x>0,\ \alpha>0,\ \beta>0 \\ 0 & ;\ \text{กรณีอื่น ๆ} \end{cases}$$

โดยที่ $\Gamma(k+1) = k!\ ;\ k \in I^{+}$

ข้อมูลจะมีลักษณะเบ้ขวาเมื่อ shape parameter (α) มีค่ามากกว่า 1 และลักษณะเบ้ขวาจะลดน้อยลงเมื่อ α มีค่ามากขึ้น ดังแสดงในรูปที่ 2.2.3.3

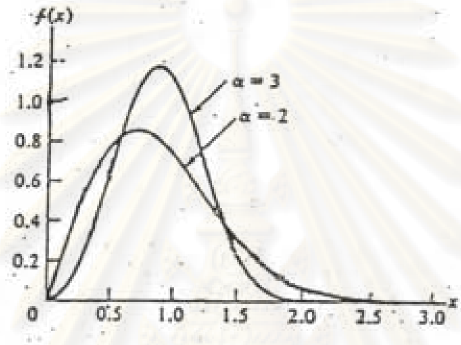


<u>รูปที่ 2.2.3.3</u>

ข. การแจกแจงแบบไวบูลล์ (weibull distribution) มีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} x^{\alpha-1} \dfrac{\exp(-x/\beta)^{\alpha}}{\beta^{\alpha}} & ;\ x>0,\ \alpha>0,\ \beta>0 \\ 0 & ;\ \text{กรณีอื่น ๆ} \end{cases}$$

ข้อมูลจะมีลักษณะเบ้ขวาเมื่อ shape parameter (α) มีค่ามากกว่า 1 และลักษณะเบ้ขวาจะลดน้อยลง เมื่อ α มีค่ามากขึ้น ดังแสดงในรูปที่ 2.2.3.4

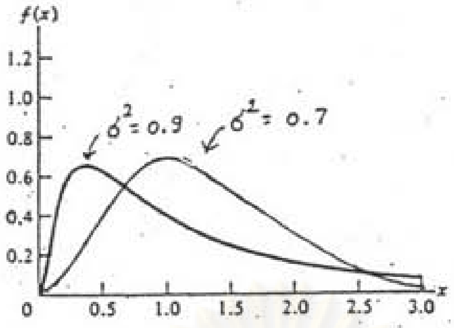


รูปที่ 2.2.3.4

ค. การแจกแจงแบบลอกนอร์มอล (lognormal distribution) มีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \exp[-(\ln x-\mu)^2/2\sigma^2] & ;\ x>0,\ \mu\in R,\ \sigma>0 \\ 0 & ;\ \text{กรณีอื่น ๆ} \end{cases}$$

ลักษณะการกระจายของข้อมูลจะเบ้ขวาไม่เด่นชัด ถ้าค่า shape parameter (σ) มีค่าเล็ก ($\sigma\to 0$) และจุดยอดมีลักษณะค่อนข้างแหลม ลักษณะเบ้ขวาจะเด่นชัดขึ้นเมื่อ σ มีค่ามากขึ้น ดังแสดงในรูปที่ 2.2.3.5



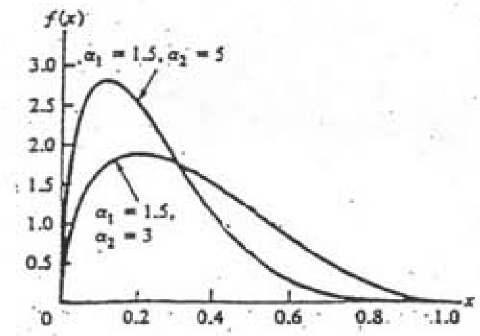
<u>รูปที่ 2.2.3.5</u>

ง. การแจกแจงแบบเบต้า (beta distribution) มีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} \dfrac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} & ; 0<x<1,\ \alpha_1>0,\ \alpha_2>0 \\ 0 & ; \text{กรณีอื่น ๆ} \end{cases}$$

โดยที่ $B(\alpha_1, \alpha_2) = \dfrac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$

ข้อมูลมีลักษณะเบ้ขวาเมื่อ shape parameter (α_1 และ α_2) มีค่ามากกว่า 1 และ α_1 มีค่าน้อยกว่า α_2 ดังแสดงในรูปที่ 2.2.3.6



<u>รูปที่ 2.2.3.6</u>

2.2.4 การแจกแจงแบบที (t-distribution)

$$\text{ให้} \quad x \sim N(0,1)$$

$$y \sim \chi^2_{(n-1)}$$

โดยที่ x และ y เป็นตัวแปรที่ไม่ขึ้นต่อกัน ตัวแปร $t = \dfrac{x}{\sqrt{y/n-1}}$ มีการแจกแจงแบบที และมีองศาความเป็นอิสระเท่ากับ n-1

ในกรณีที่สุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจง $N(\mu, \sigma^2)$ ซึ่งสามารถหาค่าเฉลี่ยและความแปรปรวนจากตัวอย่างได้ ค่าสถิติที หาได้จากความสัมพันธ์ต่อไปนี้

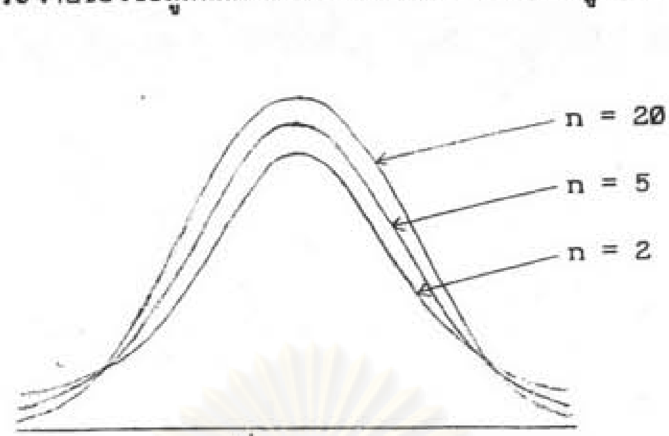
$$\text{จาก} \qquad Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\text{และ} \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\text{ค่าสถิติที (t)} \qquad = \frac{\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)\ s^2/\sigma^2}{n-1}}}$$

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

ลักษณะการกระจายของข้อมูลที่มีการแจกแจงแบบที แสดงในรูปที่ 2.2.4.1



รูปที่ 2.2.4.1

2.3 การประมาณการแจกแจง

ในการประมาณค่าสัดส่วนประชากรแบบช่วง อาจใช้การแจกแจงที่แท้จริงของหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a) หรืออาจประมาณการแจกแจงของ a ด้วยการแจกแจงแบบทวินาม และประมาณการแจกแจงแบบทวินามให้เป็นการแจกแจงแบบปกติได้ตามทฤษฎีต่อไปนี้

2.3.1 การประมาณการแจกแจงแบบไฮเปอร์ยีออเมตริกด้วยการแจกแจงแบบทวินาม

ในทางปฏิบัติเมื่อต้องการให้ความน่าจะเป็นที่จะได้ผลสำเร็จในการทดลองแต่ละครั้งมีค่าคงที่ จึงต้องประมาณการแจกแจงแบบไฮเปอร์ยีออเมตริกด้วยการแจกแจงแบบทวินาม เงื่อนไขในการประมาณคือ จำนวนประชากรมีขนาดใหญ่จนทำให้สิ่งที่สนใจและสิ่งที่ไม่สนใจมีมากพอ และสัดส่วนระหว่างจำนวนตัวอย่างซึ่งมีลักษณะที่สนใจกับจำนวนตัวอย่างทั้งหมดลู่ไปสู่ค่าคงที่ p ทฤษฎีที่ใช้ในการประมาณคือ

2.3.1.1 $$\binom{m}{r} = \frac{m^r}{r!} \quad ; \quad r \to \infty$$

2.3.1.2 $$\lim_{\substack{k\to\infty \\ N-k\to\infty \\ k/N\to p}} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

2.3.2 การประมาณการแจกแจงแบบทวินามด้วยการแจกแจงแบบปกติ

ในกรณีที่การแจกแจงแบบทวินามมี $p \to \frac{1}{2}$ หรือ $q \to \frac{1}{2}$ เมื่อนำมาแสดงด้วยแท่งฮีสโตแกรมจะมีลักษณะคล้ายโค้งปกติ ในกรณีเช่นนี้จึงสามารถประมาณการแจกแจงแบบทวินามด้วย

การแจกแจงแบบปกติได้ การประมาณนี้จะใช้ได้ดีเมื่อขนาดตัวอย่างมีค่ามาก แต่ในกรณีที่ขนาดตัวอย่างไม่มากนัก ถ้าค่า p หรือ q ไม่เข้าใกล้ 0 หรือ 1 มากเกินไป การประมาณดังกล่าวยังคงอนุโลมใช้ได้ สิ่งสำคัญก็คือ เมื่อต้องการประมาณการแจกแจงแบบทวินามซึ่งเป็นการแจกแจงของตัวแปรสุ่มชนิดที่ไม่ต่อเนื่อง (discrete random variable) ด้วยการแจกแจงแบบปกติ ซึ่งเป็นการแจกแจงของตัวแปรสุ่มที่ต่อเนื่อง (continuous random variable) จะต้องปรับค่าตัวแปรสุ่มที่ไม่ต่อเนื่องให้เป็นแบบต่อเนื่อง โดยขยายค่าใหม่ให้คลุมค่าเดิมด้วยการนำ 0.5 ไปบวกเข้าและลบออกตามทฤษฎีต่อไปนี้

ถ้า $x \sim B(n,p)$ จะได้ว่า

$$P(j \leq x \leq k) = \sum_{x=j}^{k} \binom{n}{x} p^{x}(1-p)^{n-x}$$

$$\sim \phi(\beta) - \phi(\alpha)$$

เมื่อ $\phi(\beta) = \dfrac{k-np+0.5}{\sqrt{npq}}$

$$\phi(\alpha) = \frac{j-np-0.5}{\sqrt{npq}}$$

เมื่อ $\phi(\beta)$ และ $\phi(\alpha)$ แทน ฟังก์ชันความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐาน

2.4 การคำนวณขนาดตัวอย่างเมื่อประชากรมีการแจกแจงแบบปกติ

ในกรณีที่ประชากรมีการแจกแจงแบบปกติ และกำหนดให้วิธีการสุ่มตัวอย่างเป็นแบบง่าย ขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยและค่าสัดส่วนของประชากร หาได้ดังต่อไปนี้

2.4.1 ขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยของประชากร

กำหนดให้

$\bar{Y}$ = ค่าเฉลี่ยประชากร

$\bar{y}$ = ค่าเฉลี่ยตัวอย่าง

$\bar{y} - \bar{Y}$ = ขนาดของความผิดพลาดที่ยอมรับได้ในการประมาณค่าเฉลี่ยของประชากร เขียนแทนด้วย e

α = ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นจริง

ความคลาดเคลื่อนในการประมาณค่าเฉลี่ยกำหนดให้อยู่ในรูปของค่าสัมบูรณ์ โดยยอมให้เกิดความผิดพลาดได้ทั้งสองทาง และเนื่องจาก $\bar{y}$ เป็นค่าที่ได้จากตัวอย่าง ดังนั้น $\bar{y}$ จึงเป็นตัวแปรสุ่มที่มีการแจกแจง เมื่อผู้วิเคราะห์กำหนดความคลาดเคลื่อนในการประมาณค่าแบบช่วงเท่ากับ e จึงสามารถแสดงความสัมพันธ์ระหว่างความคลาดเคลื่อนกับระดับนัยสำคัญที่กำหนดให้อยู่ในรูปของความน่าจะเป็นได้ดังนี้

$$\mathrm{Prob}(/\bar{y} - \bar{Y}/ \geq e) = \alpha$$

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

เมื่อแทน σ ด้วย s ที่คำนวณได้จากตัวอย่าง จะได้

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

เทียบความคลาดเคลื่อนกับค่า e ที่กำหนด ทำให้ได้

$$e = ts_{\bar{y}}$$

$$= t\frac{s_y}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

$$n = \frac{t^2 s_y^2}{e^2}\left(\frac{N-n}{N}\right)$$

ถ้าประชากรมีขนาดใหญ่มาก ($N \rightarrow \infty$) ทำให้ $\frac{N-n}{N}$ เข้าใกล้ 1 ดังนั้น

$$n = \frac{t^2 s_y^2}{e^2}$$

$$= \frac{t^2 s_y^2 / \bar{y}^2}{\left(\frac{\bar{y}-\bar{Y}}{\bar{y}}\right)^2}$$

ขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยประชากร ($n_{\bar{y}}$) เมื่อกำหนดระดับนัยสำคัญเท่ากับ α และค่าความคลาดเคลื่อนสัมพัทธ์ในการประมาณค่าเฉลี่ยตัวอย่างเท่ากับ $r_{\bar{y}} = \frac{t^2\,(c.v.(\bar{y}))^2}{r_{\bar{y}}^2}$

เมื่อ $c.v.(\bar{y}) = s_y/\bar{y}$

$$r_{\bar{y}} = \frac{\bar{y}-\bar{Y}}{\bar{y}} \quad ; \quad \bar{y} \neq 0$$

2.4.2 ขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากร

กำหนดให้

a = จำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ

A = จำนวนหน่วยในประชากรซึ่งมีลักษณะที่สนใจ

N = ขนาดประชากร

n = ขนาดตัวอย่าง

P = สัดส่วนประชากร = A/N

p = สัดส่วนจากตัวอย่าง = a/n

$p-P$ = ขนาดของความผิดพลาดที่ยอมรับได้ในการประมาณค่าสัดส่วนประชากร เขียนแทนด้วย e

ในทำนองเดียวกัน จะกำหนดความคลาดเคลื่อนในการประมาณค่าสัดส่วนประชากรแบบช่วงในรูปของค่าสัมบูรณ์ และเนื่องจาก p เป็นตัวแปรสุ่ม ระดับความคลาดเคลื่อน e ที่ผู้วิเคราะห์กำหนดจึงสัมพันธ์กับระดับนัยสำคัญ α ในรูปของความน่าจะเป็นดังนี้

$$\text{Prob}(/p - P/ \geqslant e) = \alpha$$

$$\sigma_p = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}}$$

เมื่อแทน P ด้วย p ที่คำนวณได้จากตัวอย่าง จะได้

$$s_p = \sqrt{\frac{pq}{n-1}} \sqrt{\frac{N-n}{N}}$$

เทียบความคลาดเคลื่อนกับค่า e ที่กำหนด จะได้

$$e = zs_p$$

$$= z \sqrt{\frac{pq}{n-1}} \sqrt{\frac{N-n}{N}}$$

$$n-1 = \frac{z^2pq}{e^2} \left(\frac{N-n}{N}\right)$$

ในกรณีที่ประชากรมีขนาดใหญ่ ($N \longrightarrow \infty$) $\frac{N-n}{N}$ เข้าใกล้ 1

นั่นคือ

$$n = \frac{z^2pq}{e^2} + 1$$

$$= \frac{z^2p(1-p)}{(p-P)^2} + 1$$

$$= \frac{z^2p(1-p) + (p-P)^2}{(p-P)^2}$$

$$= \frac{z^2p(1-p)/p^2 + (p-P/p)^2}{(p-P/p)^2}$$

ขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากร (n_p) เมื่อกำหนดระดับนัยสำคัญเท่ากับ $\propto$ และค่าความคลาดเคลื่อนสัมพัทธ์ในการประมาณค่าสัดส่วนตัวอย่างเท่ากับ $r_p = \frac{z^2(1-p) + pr_p^2}{pr_p^2}$

เมื่อ $r_p = \frac{p - P}{P}$; $p \neq 0$

2.5 <u>การประมาณขนาดตัวอย่างเมื่อประชากรมีการแจกแจงแบบเบ้</u>

ในกรณีที่ประชากรมีการแจกแจงแบบเบ้ การประมาณขนาดตัวอย่างอย่างคร่าว ๆ โดยพิจารณาจากความสัมพันธ์ระหว่างระดับของความเบ้กับขนาดตัวอย่าง หาได้จาก[1]

$$n > 25G^2$$

เมื่อ G แทนระดับความเบ้ของประชากร $= \sum_{i=1}^{N}(yi-\bar{Y})^3/\sigma^3$; $G \neq 0$

และ n แทนขนาดตัวอย่าง

จากความสัมพันธ์ข้างต้นพบว่า การประมาณขนาดตัวอย่างเมื่อประชากรมีการแจกแจง

---

[1]Cochran, W.G. Sampling Techniques. 3d ed. (John-Wiley & Sons Inc, New York : 1977), p.42.

แบบเบ้ขวาหรือแบบเบ้ซ้ายจะให้ขนาดตัวอย่างที่เท่ากันเมื่อแทนค่าของ G ที่ระดับเดียวกัน ดังนั้นค่าที่ใช้ปรับขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบปกติ ให้เป็นขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบเบ้ขวาหรือแบบเบ้ซ้าย จะใช้ค่าปรับที่เท่ากันเมื่อกำหนดระดับของความเบ้ไว้เท่ากัน ในที่นี้จะกล่าวถึงวิธีการหาค่าปรับขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบปกติ ให้เป็นขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบเบ้ขวาเท่านั้น

2.5.1 ขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยประชากร

ให้ $k_{\bar{y}}$ แทน อัตราส่วนระหว่างขีดจำกัดล่างของขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยประชากรที่ได้จากการแจกแจงแบบเบ้ เทียบกับขนาดตัวอย่างที่ได้จากการแจกแจงแบบปกติ

จากขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยประชากร เมื่อประชากรมีการแจกแจงแบบปกติ เมื่อกำหนดค่าระดับนัยสำคัญเท่ากับ α และค่าความคลาดเคลื่อนสัมพัทธ์ในการประมาณค่าเฉลี่ยตัวอย่างเท่ากับ $r_{\bar{y}}$ คือ

$$n_{\bar{y}} = t^2(c.v.(\bar{y}))^2 / r_{\bar{y}}^2$$

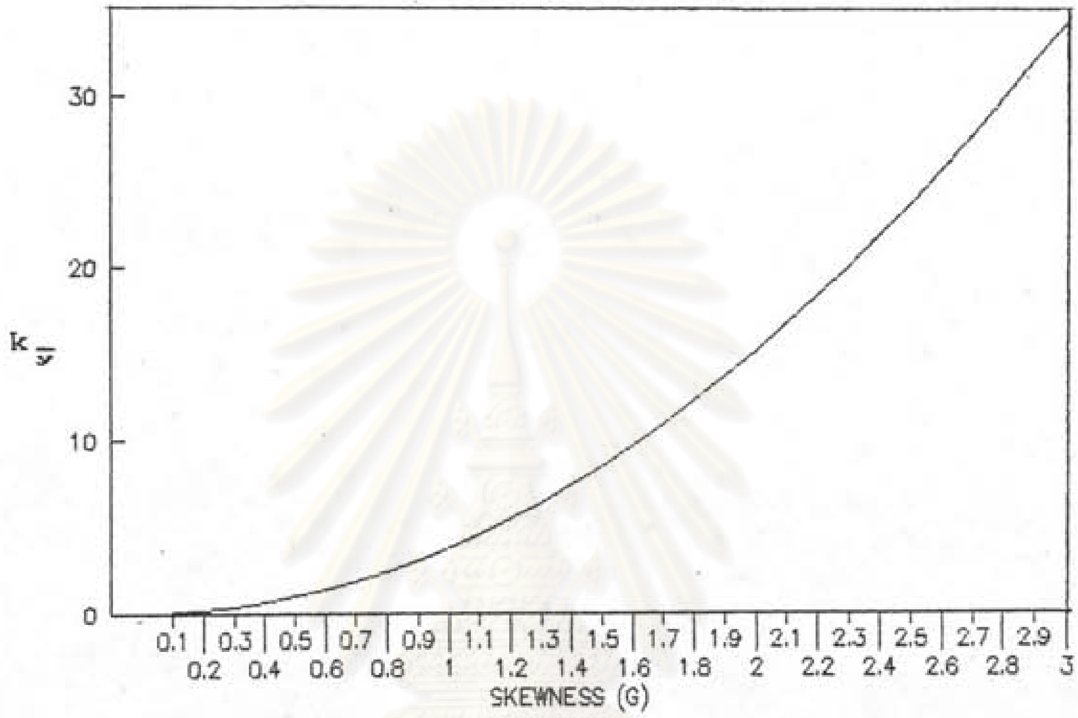
จากความสัมพันธ์ $n > 25G^2$ เมื่อพิจารณาค่าต่ำสุดคือ $n = 25G^2$ ดังนั้นค่าต่ำสุดของ $k_{\bar{y}}$ หาได้จาก

$$k_{\bar{y}} = \frac{r_{\bar{y}}^2\ 25G^2}{t^2\ (c.v.(\bar{y}))^2} \qquad (1)$$

เมื่อแทนค่าของเทอมต่าง ๆ ได้แก่ ค่าวิกฤติจากตารางที (t) ซึ่งได้จากการกำหนดระดับนัยสำคัญ (α) ค่าสัมประสิทธิ์ความแปรผันของค่าเฉลี่ยตัวอย่าง $(c.v.(\bar{y}))$ ค่าความคลาดเคลื่อนสัมพัทธ์จากการประมาณค่าเฉลี่ยประชากร ($r_{\bar{y}}$) และค่าของความเบ้ (G) ตามที่ต้องการลงในสมการที่ (1) จะได้ค่าต่ำสุดของ $k_{\bar{y}}$ ที่ใช้ปรับขนาดตัวอย่างสำหรับประมาณค่าเฉลี่ยประชากร ซึ่งรายละเอียดของค่าต่ำสุดของ $k_{\bar{y}}$ แสดงไว้ในตารางที่ 1 ในภาคผนวก ง. สำหรับบริเวณที่เว้นว่างไว้ในตารางที่ 1 หมายความว่า จะไม่พิจารณาในที่นี้

เมื่อแทนค่า G ตั้งแต่ 0.01 ถึง 3 โดยเพิ่มขึ้นทีละ 0.01 ในสมการที่ (1)

และกำหนดให้ α $r_{\bar{y}}$ และ c.v.($\bar{y}$) มีค่าเท่ากันทุกค่าของ G จะได้ความสัมพันธ์ระหว่างค่าต่ำสุดของ $k_{\bar{y}}$ กับ G แสดงได้ดังรูปที่ 2.5.1.1



รูปที่ 2.5.1.1 แสดงความสัมพันธ์ระหว่างค่าต่ำสุดของ $k_{\bar{y}}$ กับ G เมื่อกำหนดให้ $\alpha = 0.01$ $r_{\bar{y}} = 0.01$ และ c.v.($\bar{y}$) = 0.01

จากรูปที่ 2.5.1.1 จะเห็นได้ว่าเมื่อ G มีค่าเพิ่มขึ้น ค่าต่ำสุดของ $k_{\bar{y}}$ จะเพิ่มขึ้นด้วย และถ้า G มีค่ามากค่าต่ำสุดของของ $k_{\bar{y}}$ จะมีค่าที่สูงมาก เช่น เมื่อ G = 2.7 ค่าต่ำสุดของ $k_{\bar{y}}$ ที่อ่านได้จากกราฟมีค่าประมาณ 28 ซึ่งกล่าวได้ว่าขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบเบ้มากกว่าขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบปกติอย่างน้อยที่สุด 28 เท่า ความแตกต่างที่มากเช่นนี้มักไม่เกิดขึ้นจริงในทางปฏิบัติ ดังนั้นการนำค่า $k_{\bar{y}}$ มาปรับขนาดตัวอย่าง จึงควรพิจารณาในกรณีที่ระดับของความเบ้มีค่าไม่สูงมากนัก

2.5.2 ขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากร

ให้ k_p แทน อัตราส่วนระหว่างขีดจำกัดล่างของขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากรที่ได้จากการแจกแจงแบบเบ้ เทียบกับขนาดตัวอย่างที่ได้จากการแจกแจงแบบปกติ

จากขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากร เมื่อประชากรมีการแจกแจงแบบปกติ เมื่อกำหนดค่าระดับนัยสำคัญเท่ากับ α และค่าความคลาดเคลื่อนสัมพัทธ์ในการประมาณค่าสัดส่วนจากตัวอย่างเท่ากับ r_p คือ

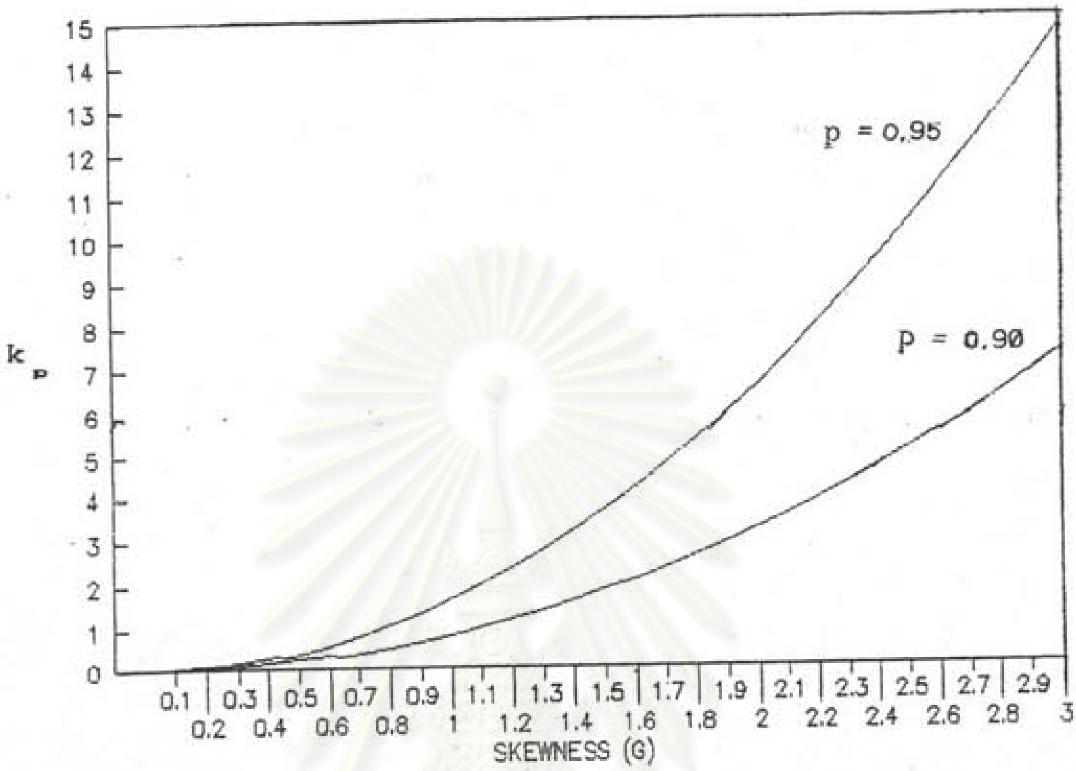
$$n_p = \frac{z^2(1-p) + pr_p^2}{pr_p^2}$$

จากความสัมพันธ์ $n > 25G^2$ เมื่อพิจารณาค่าต่ำสุดคือ $n = 25G^2$ ดังนั้นค่าต่ำสุดของ k_p หาได้จาก

$$k_p = \frac{pr_p^2\ 25G^2}{z^2(1-p)+pr_p^2} \qquad (2)$$

เมื่อแทนค่าของเทอมต่าง ๆ ได้แก่ ค่าวิกฤติจากตารางปกติมาตรฐาน (z) ซึ่งได้จากการกำหนดระดับนัยสำคัญ (α) ค่าสัดส่วนจากตัวอย่าง (p) ค่าความคลาดเคลื่อนสัมพัทธ์จากการประมาณค่าสัดส่วนประชากร (r_p) และค่าของความเบ้ (G) ในสมการที่ (2) จะได้ค่าต่ำสุดของ k_p ที่ใช้ปรับขนาดตัวอย่างสำหรับประมาณค่าสัดส่วนประชากร ซึ่งรายละเอียดของค่าต่ำสุดของ k_p แสดงไว้ในตารางที่ 2 ในภาคผนวก ง. สำหรับบริเวณที่ว่างในตารางที่ 2 หมายความว่า จะไม่พิจารณาในที่นี้

ในทำนองเดียวกัน เมื่อพิจารณาความสัมพันธ์ระหว่างค่าต่ำสุดของ k_p กับ G โดยแทนค่าของ G ตั้งแต่ 0.01 ถึง 3 โดยเพิ่มขึ้นทีละ 0.01 ในสมการที่ (2) พบว่าค่าต่ำสุดของ k_p จะเพิ่มขึ้นเมื่อ G มีค่ามากขึ้น ซึ่งค่าต่ำสุดของ k_p จะเพิ่มขึ้นไม่มากนักในกรณีที่กำหนดให้ α r_p และ p มีค่าคงที่ แต่ค่าต่ำสุดของ k_p จะเพิ่มขึ้นสูงมากเมื่อ G และ p มีค่ามาก ($p \rightarrow 1$) การเปรียบเทียบความสัมพันธ์ระหว่างค่าต่ำสุดของ k_p กับ G เมื่อ p มีค่าเท่ากับ 0.9 และ 0.95 แสดงในรูปที่ 2.5.2.1



รูปที่ 2.5.2.1 แสดงการเปรียบเทียบความสัมพันธ์ระหว่างค่าต่ำสุดของ k_p กับ G เมื่อกำหนดให้ $\alpha=0.10$ $r_p=0.10$ $p=0.9$ กับ $\alpha=0.10$ $r_p=0.10$ $p=0.95$

จากรูปที่ 2.5.2.1 ถ้ากำหนดให้ G เท่ากับ 2.7 ค่าต่ำสุดของ k_p ที่อ่านได้จากกราฟ เมื่อ p เท่ากับ 0.9 มีค่าประมาณ 5.6 และเมื่อ p เท่ากับ 0.95 ค่าต่ำสุดของ k_p ที่อ่านได้จากกราฟมีค่าประมาณ 12 นั่นคือ ความแตกต่างระหว่างขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบเบ้ กับขนาดตัวอย่างสำหรับประชากรที่มีการแจกแจงแบบปกติจะเพิ่มขึ้นอย่างรวดเร็วเมื่อ G มีค่ามาก และ p เข้าใกล้ 1 ดังนั้น การนำค่าต่ำสุดของ k_p ไปปรับขนาดตัวอย่างจึงควรพิจารณาในกรณีที่ค่า G และ p ไม่สูงมากนัก