บทที่ 8

ทฤษฎีการเลือกกลุ่มตัวอย่าง

(Sampling Theory)

8.1 นิยามศัพท์เฉพาะ

ทฤษฎีการเลือกกลุ่มตัวอย่าง เป็นทฤษฎีที่ว่าด้วยความสัมพันธ์ระหว่างประชากร (Population) กับกลุ่มตัวอย่าง (Samples) โดยมีจุดมุ่งหมายที่จะให้ข้อมูลที่ได้มาจากกลุ่มตัวอย่าง สามารถประมาณค่าพารามิเตอร์ต่าง ๆ ของประชากรได้อย่างถูกต้องเชื่อถือได้

การสุ่มกลุ่มตัวอย่าง (Sampling) หมายถึงวิธีการในการให้ได้มาซึ่งกลุ่มตัวอย่างที่เป็นตัวแทนที่ดี (representative) ของประชากร

ประชากร (Population) หมายถึงเซ็ตของค่าวัด (Measure) ทั้งหมดที่เป็นไปตามเงื่อนไขที่เราต้องการศึกษา เช่น ถ้าต้องการศึกษาความสูงของนักศึกษารามคำแหง ประชากรก็คือนักศึกษามหาวิทยาลัยรามคำแหงทุกคน หรือถ้าต้องการศึกษารายได้เฉลี่ยของคนในกรุงเทพมหานคร ประชากรก็คือคนในกรุงเทพมหานครทั้งหมด ประชากรอาจเป็นคน สัตว์ สิ่งของ ฯลฯ ประชากรอาจแบ่งได้ 2 ลักษณะคือ

1. ประชากรที่มีขนาดจำกัด (Finite population) หมายถึงประชากรที่ประกอบด้วยหน่วย (unit) ที่สามารถนับจำนวนได้ครบถ้วนแน่นอน เช่นจำนวนนักศึกษามหาวิทยาลัยรามคำแหง จำนวนโรงภาพยนตร์ในกรุงเทพมหานคร

2. ประชากรที่มีขนาดไม่จำกัด (Infinite population) หมายถึงประชากรที่ไม่สามารถนับจำนวนได้ หรือนับจำนวนได้ไม่แน่นอน เช่นจำนวนหัวที่เกิดจากการโยนเหรียญ 1 เหรียญ จำนวนเม็ดทรายที่ชายหาดบางแสน

กลุ่มตัวอย่าง (Sample) หมายถึงสับเซ็ตของประชากรที่ถูกสุ่มขึ้นมาเพื่อใช้ในการศึกษาแทนประชากร กลุ่มตัวอย่างควรมีลักษณะเหมือนกับประชากรทุกอย่าง ต่างกันที่จำนวนเท่านั้น หากได้กลุ่มตัวอย่างที่เป็นตัวแทนของประชากร จะทำให้ผู้วิจัยสามารถสรุปอ้างอิงผลที่ได้ไปยังประชากรได้

พารามิเตอร์ (Parameter) เป็นค่าที่แสดงลักษณะของประชากร ในการหาค่าพารามิเตอร์นั้นต้องคำนวณจากข้อมูลประชากร (Population data) ค่าพารามิเตอร์ถือว่าเป็นค่าอันแท้จริง (True value) แต่โดยทั่วไปมักจะไม่ทราบค่า ดังนั้นจึงต้องอาศัยการประมาณค่าทางสถิติ และนิยมเขียนแทนค่าพารามิเตอร์ด้วยอักษรกรีก เช่น

μ แทนค่าเฉลี่ยของประชากร

σ^2 แทนความแปรปรวนของประชากร

สถิติ (Statistic) เป็นค่าที่แสดงลักษณะของกลุ่มตัวอย่าง ในการหาค่าสถิตินั้นต้องคำนวณจากข้อมูลกลุ่มตัวอย่าง (Sample data) และนิยมเขียนแทนค่าสถิติด้วยอักษรโรมัน เช่น

$\overline{X}$ แทนค่าเฉลี่ยของกลุ่มตัวอย่าง

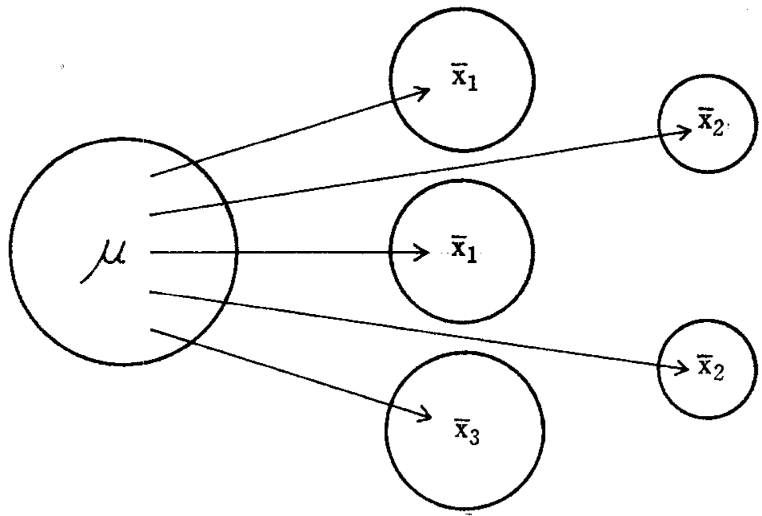
S^2 แทนความแปรปรวนของกลุ่มตัวอย่าง

8.2 การแจกแจงของกลุ่มตัวอย่าง (Sampling distribution)

หมายถึงการแจกแจงความน่าจะเป็น (Probability distribution) ของสถิติ ตัวอย่างเช่น การแจกแจงความน่าจะเป็นของค่าเฉลี่ย ($\overline{x}$) การแจกแจงความน่าจะเป็นของความแปรปรวน (S^2) การแจกแจงความน่าจะเป็นของผลต่างของค่าเฉลี่ย ($\overline{x}_1 - \overline{x}_2$) เป็นต้น

8.2.1 การแจกแจงของ $\overline{x}$

ถ้าเรามีประชากรอันหนึ่ง และเราสุ่มกลุ่มตัวอย่างจากประชากรได้หลายกลุ่ม ในประชากรเราจะมีค่าเฉลี่ย (μ) ในกลุ่มตัวอย่างแต่ละกลุ่ม เราก็จะมีค่าเฉลี่ย ($\overline{x}$) ซึ่ง $\overline{x}$ เหล่านี้บางกลุ่มอาจจะเท่ากัน ดังนั้นก็จะทำให้เกิดมีความน่าจะเป็นของ $\overline{x}$ แต่ละค่าขึ้น



ตัวอย่างเช่น ถ้ามีกลุ่มตัวอย่าง 100 กลุ่ม ใน 100 กลุ่มนี้ มี 20 กลุ่มที่มี $\bar{x} = \bar{x}_1 = 120$ มี 30 กลุ่มที่มี $\bar{x} = \bar{x}_2 = 115$ มี 10 กลุ่ม ที่มี $\bar{x} = \bar{x}_3 = 121$ มี 40 กลุ่มที่มี $\bar{x} = \bar{x}_4 = 146$ ดังนั้น

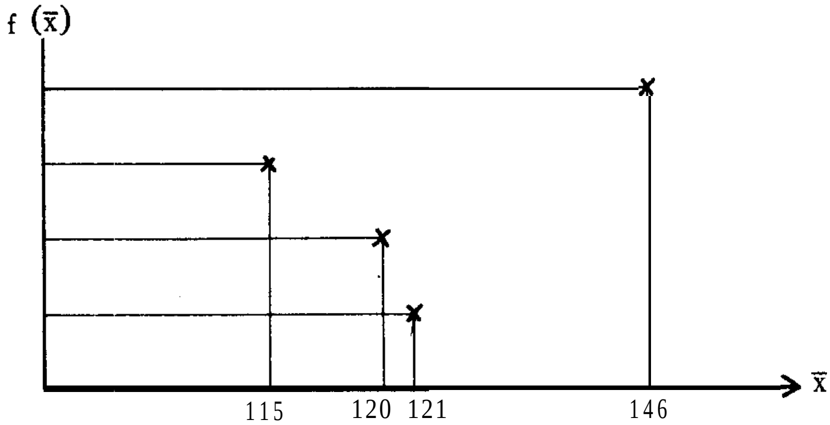
ความน่าจะเป็นของ $\bar{x}_1 = f(\bar{x}_1) = f(120) = \frac{20}{100}$

ความน่าจะเป็นของ $\bar{x}_2 = f(\bar{x}_2) = f(115) = \frac{30}{100}$

ความน่าจะเป็นของ $\bar{x}_3 = f(\bar{x}_3) = f(121) = \frac{10}{100}$

ความน่าจะเป็นของ $\bar{x}_4 = f(\bar{x}_4) = f(146) = \frac{40}{100}$

ถ้านำค่า $\bar{x}$ และ $f(\bar{x})$ มาเขียนกราฟ จะได้ดังนี้



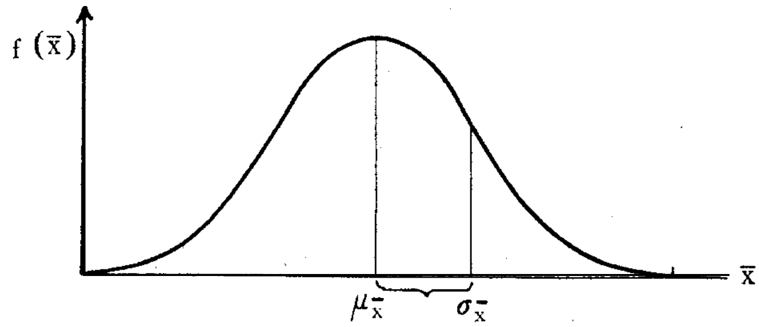
การแจกแจงดังกราฟในรูปนี้ก็จะเป็นการแจกแจงความน่าจะเป็นของ $\bar{x}$ และถ้าจำนวนกลุ่มตัวอย่างมีมากถึง infinite การแจกแจงความน่าจะเป็นของ $\bar{x}$ ก็จะเป็นฟังก์ชั่นที่ต่อเนื่องหรือเป็น curve

การแจกแจงของสถิติอื่น ๆ เช่น s^2, s, $\bar{x}_1-\bar{x}_2$ ก็จะมีความหมายในทำนองเดียวกัน

8.2.2 **ทฤษฎีลิมิตกลาง** (Central Limit Theorem) ทฤษฎีนี้กล่าวว่า ถ้าสุ่มกลุ่มตัวอย่างจากประชากรที่มีค่าเฉลี่ย และความแปรปรวนเป็น μ และ σ^2 ตามลำดับ และถ้าจำนวนกลุ่มตัวอย่างที่สุ่มมามีขนาดใหญ่ ($n > 30$) แล้วค่าเฉลี่ย ($\bar{x}$) ของกลุ่มตัวอย่างจะมีการแจกแจงปกติ (Normal distribution) ไม่ว่าการแจกแจงของประชากรเดิมจะเป็นอย่างไรก็ตาม

นั่นคือ ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ ($n > 30$) แล้ว การแจกแจงปกติก็จะเป็นค่าประมาณที่ดีของการแจกแจงค่าสถิติ $\bar{x}$

ถ้าการแจกแจงของ $\bar{X}$ เป็นการแจกแจงปกติ จุดกึ่งกลางของโค้งการแจกแจงก็จะเป็น $\mu_{\bar{x}}$ และความเบี่ยงเบนมาตรฐานเป็น $\sigma_{\bar{x}}$ ดังรูป



ถ้าให้ N เป็นขนาดของประชากร และ n เป็นขนาดของกลุ่มตัวอย่าง

เราจะได้ว่า $\mu_{\bar{x}} = \mu$

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

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

ทฤษฎี 1 $\mu_{\bar{x}} = \mu$

พิสูจน์
$$\begin{aligned}
\mu_{\bar{x}} &= E(\bar{X}) \\
&= E\left(\frac{\Sigma X}{n}\right) \\
&= \frac{1}{n} E(\Sigma X) \\
&= \frac{1}{n} \Sigma (E(X)) \\
&= \frac{1}{n} \Sigma \mu \\
&= \frac{1}{n} (n\mu) \\
&= \mu
\end{aligned}$$

ทฤษฎี 2 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

พิสูจน์
$$\begin{aligned}
\sigma_{\bar{x}} &= \sqrt{\sigma^2_{\bar{x}}} \\
&= \sqrt{V(\bar{x})} \\
&= \sqrt{V\left(\frac{\Sigma X}{n}\right)} \\
&= \sqrt{\frac{1}{n^2} V(\Sigma X)} \\
&= \sqrt{\frac{1}{n^2} \Sigma (V(X))} \\
&= \sqrt{\frac{1}{n^2} \Sigma \sigma^2} \\
&= \sqrt{\frac{1}{n^2} (n\sigma^2)} \\
&= \sqrt{\frac{\sigma^2}{n}} \\
&= \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

ในทำนองเดียวกัน ถ้าเป็นการแจกแจงของ $\bar{x}_1 - \bar{x}_2$ เราก็จะได้ว่า

$$\mu_{\bar{x}_1 \pm \bar{x}_2} = \mu_{\bar{x}_1} \pm \mu_{\bar{x}_2} = \mu_1 \pm \mu_2$$

$$\sigma_{\bar{x}_1 \pm \bar{x}_2} = \sqrt{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

8.2.3 การแจกแจงของสัดส่วน (Sampling distribution of proportions)

ถ้าหากประชากรมีการแจกแจงเป็น Binomial มีขนาดเท่ากับ n มีความน่าจะเป็นของผลที่คาดหวัง (Success) เป็น p ดังนั้น $E(X) = \mu = np$ และ $V(X) = \sigma^2 = npq$

เราจะได้ว่าในการแจกแจงความน่าจะเป็นของ p นั้น

$$\mu_p = P$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$$

ทฤษฎี 3 $\mu_p = P$

พิสูจน์

$$\begin{aligned}\mu_p &= E(p) \\ &= E\left(\frac{X}{n}\right) \\ &= \frac{1}{n} E(X) \\ &= \frac{1}{n}(np) \\ &= P\end{aligned}$$

ทฤษฎี 4 $\sigma_p = \sqrt{\frac{pq}{n}}$

พิสูจน์

$$\begin{aligned}\sigma_p &= \sqrt{V(p)} \\ &= \sqrt{V\left(\frac{X}{n}\right)} \\ &= \sqrt{\frac{1}{n^2} V(X)} \\ &= \sqrt{\frac{1}{n^2}(npq)} \\ &= \sqrt{\frac{pq}{n}}\end{aligned}$$

8.3 ความคลาดเคลื่อนมาตรฐาน (Standard error)

หมายถึงความเบี่ยงเบน มาตรฐานของการแจกแจงของกลุ่มตัวอย่าง หรือหมายถึงความเบี่ยงเบนมาตรฐานของสถิตินั่นเอง ซึ่งในตอนที่ผ่านมาเราจะเห็นว่า

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

ความคลาดเคลื่อนมาตรฐานของ $p = \sigma_p = \sqrt{\frac{pq}{n}}$

ความคลาดเคลื่อนมาตรฐานของ $\bar{x}_1 - \bar{x}_2 = \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

นอกจากนี้ยังมีความคลาดเคลื่อนมาตรฐานของสถิติอื่น ๆ เช่น s, s^2 ฯลฯ ซึ่งจะไม่กล่าวถึงในที่นี้

ให้สังเกตว่าสถิติเมื่อบวกลบ คูณ หาร กับพารามิเตอร์หรือสถิติ ก็จะกลายเป็นสถิติตัวใหม่เช่น $\bar{x}_1 - \bar{x}_2$ หรือสถิติที่จะพบในบทต่อ ๆ ไปเช่น $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ หรือ $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ เป็นต้น

8.4 หลักการเลือกกลุ่มตัวอย่าง (Sampling Principle)

หลักในการเลือกกลุ่มตัวอย่าง มีหลักใหญ่ ๆ ที่ควรพิจารณา 2 ประการคือ

1. การเลือกกลุ่มตัวอย่างแบบใส่คืน (Sampling with replacement) หมายถึงการเลือกกลุ่มตัวอย่างโดยใส่หน่วยที่ถูกเลือกขึ้นมาแล้วกลับคืนเข้าไปอีก ก่อนที่จะทำการเลือกหน่วยอื่น ๆ การเลือกแบบนี้มีโอกาสที่จะได้กลุ่มตัวอย่างที่ซ้ำกับที่เคยเลือกมาแล้ว นั่นคือแต่ละหน่วยย่อย ๆ มีโอกาสถูกเลือกมากกว่า 1 ครั้งนั่นเอง

2. การเลือกกลุ่มตัวอย่างแบบไม่ใส่คืน (Sampling without replacement) หมายถึงการเลือกกลุ่มตัวอย่างโดยการแยกหน่วยที่เคยถูกเลือกมาแล้วออกไปเลย ไม่ต้องใส่หน่วยที่เคยถูกเลือกคืนลงไป แล้วทำการเลือกหน่วยอื่น ๆ ต่อไป การเลือกแบบนี้ไม่มีโอกาสที่จะได้กลุ่มตัวอย่างที่ซ้ำกับที่เคยเลือกมาเลย ทั้งนี้เพราะแต่ละหน่วยย่อย ๆ มีโอกาสถูกเลือกเพียงครั้งเดียว

ในทางปฏิบัติ นิยมเลือกกลุ่มตัวอย่างแบบไม่ใส่คืน ทั้งนี้เพราะการเก็บข้อมูลเรื่องเดียวกัน จากหน่วยเดียวกันหลาย ๆ ครั้ง ไม่มีประโยชน์อะไรมากนัก

8.5 วิธีการเลือกกลุ่มตัวอย่าง (Sampling Methods)

วิธีการเลือกกลุ่มตัวอย่าง ก็คือวิธีการเลือกตัวแทนของประชากรมาศึกษานั่นเอง ซึ่งสามารถแบ่งเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท คือ

1. การเลือกกลุ่มตัวอย่างที่มีโอกาสถูกเลือกเท่า ๆ กัน (Probability Sampling)
2. การเลือกกลุ่มตัวอย่างที่มีโอกาสถูกเลือกไม่เท่ากัน (Non — Probability Sampling)

1. **การเลือกกลุ่มตัวอย่างที่มีโอกาสถูกเลือกเท่า ๆ กัน** (Probability Sampling) คือการเลือกกลุ่มตัวอย่างโดยที่หน่วยของประชากรมีโอกาสที่จะถูกเลือกเท่า ๆ กันหมดทุกหน่วย วิธีการเลือกแบบนี้ต้องอาศัยเทคนิคการสุ่มกลุ่มตัวอย่าง (Random Sampling Technique) ซึ่งการสุ่ม (Random) ก็คือการให้โอกาสทุก ๆ หน่วยมีโอกาสถูกเลือกเท่า ๆ กัน การเลือกกลุ่มตัวอย่างประเภทนี้ บางครั้งเรียกว่า การสุ่มกลุ่มตัวอย่าง ซึ่งสามารถแยกย่อยออกได้อีก 4 วิธี คือ

1.1 การสุ่มกลุ่มตัวอย่างอย่างง่าย (Simple Random Sampling) เป็นการสุ่มกลุ่มตัวอย่างที่ให้สมาชิกทุกตัวของประชากรที่จะศึกษามีโอกาสที่จะถูกเลือกเป็นตัวแทนเท่า ๆ กัน การสุ่มแบบนี้ทำได้หลายวิธี เช่น เขียนชื่อสมาชิกของประชากรทุกตัวลงในสลาก แล้วจับสลากตามจำนวนกลุ่มตัวอย่างที่ต้องการ หรือใช้วิธีสุ่มจากตารางเลขสุ่ม (Table of Random Number) ซึ่งตารางนี้มีอยู่ในหนังสือเทคนิคการสุ่มกลุ่มตัวอย่างหรืออาจจะหาได้ทั่วไปในหนังสือสถิติ วิธีการสุ่มจากตารางเลขสุ่มนี้ทำได้โดยการกำหนดตัวเลขแทนสมาชิกทุกตัวของประชากรเสียก่อน จากนั้นก็ใช้วิธีการสุ่มลงไปในตารางเลขสุ่ม เพื่อหาสมาชิกตัวแรกก่อน เมื่อได้ตัวเลขตัวแรกแล้ว ก็ใช้วิธีกำหนดทิศทางต่อไปว่าจะสุ่มไปทางไหนของตาราง เมื่อกำหนดอย่างไรก็ต้องทำอย่างนั้นตลอดบรรทัด แล้วขึ้นบรรทัดใหม่ต่อไปนับจำนวนไปเรื่อย ๆ จนกว่าจะได้จำนวนตามต้องการ และถ้าได้ตัวเลขซ้ำเดิมก็สุ่มใหม่จนครบ เช่น สมาชิกของประชากรมี 100 ตัว ตัวที่เริ่มต้นก็ควรเป็นเลข 2 หลัก แล้วนับเรียงเลขต่อไป ตัวละ 2 หลัก โดยให้ 00 แทนสมาชิกตัวที่ 100 นับไปจนกว่าจะครบตามจำนวนของกลุ่มตัวอย่างที่ต้องการ

1.2 การสุ่มกลุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) วิธีการสุ่มแบบนี้ใช้ได้ดีเมื่อประชากรมีการจัดเรียงลำดับเป็นระบบอยู่แล้ว เช่น การเรียงประชากรตามตัวอักษรในสมุดโทรศัพท์ เป็นต้น วิธีการสุ่มก็ทำได้โดยการกำหนดช่วงของกลุ่มตัวอย่างเสียก่อนว่า แต่ละตัวจะให้ห่างกันเท่าไร แล้วจึงเริ่มสุ่มตัวแรก แล้วนับช่วงห่างไปเรื่อย ๆ จนครบตามจำนวนที่ต้องการ เช่น ประชากร 100 คน ต้องการกลุ่มตัวอย่าง 10 คน อาจจะกำหนดช่วงละ 10 คนก็ได้ เช่น 10 20 30 .... ถึง 100 จะได้ 10 คนพอดี หรือกำหนดช่วงละ 6 คน เริ่มสุ่มคนแรกได้คนที่ 7 คนต่อไปก็คือคนที่ 7+6 คือ 13 คนที่ 3 คือ 13+6 ซึ่งก็คือ 19 นับไปจนกว่าจะครบ 10 คน เป็นต้น

1.3 การสุ่มกลุ่มตัวอย่างแบบแบ่งชั้นหรือแบ่งพวก (Stratified Random Sampling) การสุ่มแบบนี้เป็นการสุ่มจากประชากรที่มีลักษณะแตกต่างกัน หลายพวกหลายกลุ่ม ซึ่งถ้าใช้วิธีการสุ่ม 2 วิธีแรกจะไม่เหมาะสม วิธีนี้ผู้วิจัยต้องแบ่งประชากรที่จะศึกษาออกเป็นชั้น ๆ (Strata) หรือเป็นกลุ่มเสียก่อน แล้วจึงใช้การสุ่มกลุ่มตัวอย่างอย่างง่ายภายหลัง ชั้นหรือกลุ่มที่แบ่งนี้ เป็นการจัดประชากรที่มีลักษณะเป็นเอกพันธ์เข้าชั้นหรือกลุ่มเดียวกัน โดยพยายามที่จะทำให้เกิดความแตกต่างภายในชั้นน้อยที่สุด และเพิ่มความแตกต่างระหว่างชั้นให้มากขึ้น เช่น การแบ่งชั้นตามขนาดของโรงเรียน เป็น ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก แบ่งนักเรียนตามระดับอายุ 13 ปี 14 ปี และ 15 ปี เป็นต้น เมื่อแบ่งชั้นได้แล้ว จึงทำการสุ่มอย่างง่ายในแต่ละชั้นอีกครั้งหนึ่ง ถ้าจำนวนแต่ละชั้นต่างกันมาก ก็ควรใช้วิธีสุ่มอย่างง่ายเทียบตามสัดส่วนของแต่ละชั้นด้วย

1.4 การสุ่มกลุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling) การสุ่มแบบนี้เป็นการสุ่มจากประชากรที่ใช้สถานที่หรือสภาพทางภูมิประเทศเป็นหลักในการแบ่งกลุ่ม บางครั้งก็เรียกว่า การสุ่มตามพื้นที่ (Area Sampling) การสุ่มแบบนี้พยายามที่จะทำให้เกิดความแตกต่างระหว่างกลุ่มน้อยที่สุด คือ กลุ่มที่แบ่งตามภูมิประเทศหรือพื้นที่แต่ละแห่งไม่ค่อยแตกต่างกัน และเพิ่มความแตกต่างภายในกลุ่มแต่ละกลุ่มให้มากขึ้น ซึ่งตรงข้ามกับการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) วิธีการสุ่มแบบนี้ก็จะใช้การสุ่มอย่างง่ายสุ่มเป็นกลุ่มก่อน แล้วจึงสุ่มภายในกลุ่มอีกครั้งหนึ่ง และการสุ่มครั้งหลังนี้จะใช้วิธีสุ่มอย่างง่ายหรืออย่างมีระบบก็ได้แล้วแต่ลักษณะของสมาชิกในแต่ละกลุ่มที่สุ่มมาได้ การสุ่มแบบนี้เป็นที่นิยมเพราะประหยัดเวลาและค่าใช้จ่ายได้มาก

ในการสุ่มกลุ่มตัวอย่างถ้าใช้วิธีการสุ่มหลาย ๆ ครั้ง หลายวิธี อาจเรียกวิธีการเช่นนี้ว่า การสุ่มแบบหลายขั้นตอน (Multistage Sampling) ซึ่งวิธีการนี้มีลักษณะคล้ายกับการสุ่มกลุ่มตัวอย่างแบบแบ่งกลุ่ม เช่น การสุ่มนักเรียนชั้น ป. 6 ทั่วประเทศ วิธีการสุ่มอาจจะใช้การสุ่มแบบหลายขั้นตอนได้ คือ สุ่มเขตการศึกษามาก่อนแล้วสุ่มจังหวัด สุ่มอำเภอ สุ่มโรงเรียน และสุ่มห้องในที่สุด

2. การเลือกกลุ่มตัวอย่างที่มีโอกาสถูกเลือกไม่เท่ากัน (Non Probability Sampling) การเลือกกลุ่มตัวอย่างวิธีนี้ เป็นการเลือกที่หน่วยของประชากรแต่ละหน่วยมีโอกาสถูกเลือกไม่เท่ากัน บางหน่วยอาจจะไม่มีโอกาสถูกเลือกเลยก็ได้ และกลุ่มตัวอย่างที่ได้มีโอกาสถูกเลือกก็ไม่ได้เป็นตัวแทนที่ดีของประชากรทั้งหมด วิธีการเลือกกลุ่มตัวอย่างประเภทนี้ สามารถแยกได้อีก 3 วิธี คือ

2.1 การเลือกกลุ่มตัวอย่างแบบบังเอิญ (Accidental Sampling) คือการเลือกกลุ่มตัวอย่างเท่าที่ผู้วิจัยจะรวบรวมกลุ่มตัวอย่างได้ หรือพยายามให้กลุ่มตัวอย่างครบตามจำนวนที่ต้องการโดยไม่มีหลักเกณฑ์อะไรเลย พบใครที่ให้ความร่วมมือในการรวบรวมข้อมูลก็เลือกเป็นกลุ่มตัวอย่างได้เลย กลุ่มตัวอย่างที่ได้ก็ไม่ได้เป็นตัวแทนที่ดีของประชากรทั้งหมด ผลที่ได้ก็ไม่สามารถสรุปอ้างอิงได้ แต่ก็เป็นประโยชน์บ้างที่ทำให้ผู้วิจัยเห็นแนวทางว่าควรจะศึกษาปัญหานั้นต่อไปหรือไม่ ถ้าควร ควรจะศึกษาในลักษณะใด เช่น การสำรวจประชามติจากประชาชนตามป้ายรถเมล์ เป็นต้น

2.2 การเลือกกลุ่มตัวอย่างโดยอาศัยสัดส่วนขององค์ประกอบ (Quota Sampling) การเลือกแบบนี้เป็นการเลือกกลุ่มตัวอย่างจากประชากรที่มีลักษณะหลายอย่างหลายกลุ่มแตกต่างกันไป ถ้าใช้วิธีแบบบังเอิญ ก็อาจจะได้ไม่ครบทุกลักษณะหรือทุกกลุ่ม จึงต้องอาศัยสัดส่วนของประชากรแต่ละกลุ่มต่อประชากรทั้งหมด แล้วจึงเลือกกลุ่มตัวอย่างในแต่ละกลุ่มโดยวิธีเลือกแบบบังเอิญอีกครั้ง เช่น ประชากรประกอบด้วยบุคคล 5 อาชีพ ผู้วิจัยก็เลือกแบบบังเอิญแต่ละอาชีพตามสัดส่วนของประชากรแต่ละอาชีพต่อประชากรรวมทั้งหมด เป็นต้น

2.3 การเลือกกลุ่มตัวอย่างแบบเจาะจง (Purposive Sampling) การเลือกกลุ่มตัวอย่างแบบนี้เป็นการเลือกโดยที่ผู้วิจัยได้กำหนดเงื่อนไขหรือคุณลักษณะของประชากรที่จะศึกษาไว้ล่วงหน้าแล้ว เพื่อให้เกิดความเหมาะสมกับเรื่องที่จะทำวิจัย และเมื่อพบกับประชากร

ลักษณะเช่นที่กำหนดก็จะเลือกเป็นกลุ่มตัวอย่างทันที การเลือกกลุ่มตัวอย่างแบบนี้ มักจะเกิดการลำเอียง ทำให้กลุ่มตัวอย่างไม่เป็นตัวแทนที่ดีของประชากร

8.6 ขนาดของกลุ่มตัวอย่าง (Sample Size)

ขนาดของกลุ่มตัวอย่างก็คือจำนวนของกลุ่มตัวอย่างที่ใช้ในการวิจัยนั่นเอง ซึ่งจะมากหรือน้อยขึ้นอยู่กับเรื่องที่ทำวิจัย และลักษณะของประชากร ถ้ากลุ่มตัวอย่างยิ่งมาก ความคลาดเคลื่อนในการสุ่มก็จะน้อยลง แต่บางครั้งประชากรมีจำนวนมาก และผู้วิจัยก็ไม่สามารถจะใช้ขนาดของกลุ่มตัวอย่างมาก ๆ ได้เนื่องจากข้อจำกัดบางประการ เช่นเงิน เวลา และแรงงาน ผู้วิจัยจึงมีความจำเป็นที่จะต้องพยายามสุ่มกลุ่มตัวอย่างให้มีขนาดเล็กลง แต่ยังเป็นตัวแทนของประชากรได้ จึงเกิดปัญหาว่า ควรจะใช้กลุ่มตัวอย่างสักเท่าไรจึงจะเพียงพอสำหรับการศึกษาวิจัย ในการกำหนดขนาดของกลุ่มตัวอย่างนั้นควรคำนึงถึงสิ่งต่อไปนี้

1. ขนาดของประชากร (Size of Population) โดยทั่วไปถ้าประชากรมีขนาดใหญ่ กลุ่มตัวอย่างก็ควรจะมีขนาดใหญ่ตามไปด้วย

2. ความแปรปรวนของประชากร (Variation of Population) โดยทั่วไปถ้าประชากรมีลักษณะคล้าย ๆ กัน (Homogeneous) ขนาดของกลุ่มตัวอย่างก็จะเล็กลง แต่ถ้าประชากรมีลักษณะแตกต่างกันมาก (Heterogeneous) ขนาดของกลุ่มตัวอย่างก็จะใหญ่ขึ้น

3. ระดับของความแม่นยำ (Degree of Precision) ถ้าต้องการความแม่นยำในการประมาณค่าพารามิเตอร์มาก ขนาดของกลุ่มตัวอย่างก็ควรให้มากขึ้นด้วย ความแม่นยำก็คือระดับความแตกต่างระหว่างค่าสถิติกับค่าพารามิเตอร์ ถ้าผลการวิจัยมีระดับความแม่นยำมาก ก็แสดงว่าค่าสถิติกับค่าพารามิเตอร์มีค่าใกล้เคียงกันมาก

4. ขอบเขตของความคลาดเคลื่อน (α — error) ผู้วิจัยจะต้องตัดสินใจว่าจะยอมให้เกิดความคลาดเคลื่อนระหว่างค่าสถิติกับค่าพารามิเตอร์ได้มากน้อยเพียงไร ถ้ายอมให้เกิดความคลาดเคลื่อนมากก็จะใช้ขนาดของกลุ่มตัวอย่างน้อยลง แต่ถ้ากำหนดขอบเขตของความคลาดเคลื่อนไว้ต่ำก็จำเป็นต้องกำหนดขนาดของกลุ่มตัวอย่างเพิ่มขึ้น ความคลาด

เคลื่อนจะมีความสัมพันธ์ทางลบกับระดับความแม่นยำ กล่าวคือถ้าความคลาดเคลื่อนน้อย ระดับความแม่นยำก็จะมากขึ้น โดยทั่วไปผู้วิจัยมักจะยอมให้เกิดความคลาดเคลื่อน 1% , 5% และอย่างมากไม่เกิน 10%

8.7 การคำนวณหาขนาดของกลุ่มตัวอย่าง

ปัญหาที่สำคัญอย่างหนึ่งของการวิจัยก็คือปัญหาเกี่ยวกับขนาดของกลุ่มตัวอย่างว่าจำนวนกลุ่มตัวอย่างควรจะเป็นเท่าใด จึงจะเพียงพอ การตอบคำถามเกี่ยวกับปัญหานี้ควรระมัดระวังให้ถูกต้องตามหลักวิชาและหลักประหยัด เพราะหากตอบโดยไม่มีกฎเกณฑ์ก็จะมีข้อเสียคือกลุ่มตัวอย่างอาจมากเกินไป ทำให้เสียค่าใช้จ่ายมาก หรือบางทีกลุ่มตัวอย่างอาจน้อยเกินไปซึ่งจะทำให้ผลการวิจัยมีความคลาดเคลื่อนสูง และเพื่อไม่ให้เกิดข้อเสียดังกล่าวจึงได้มีผู้คิดหาสูตรในการคำนวณขนาดของกลุ่มตัวอย่างไว้มากมายในที่นี้ขอยกมากล่าวเพียงบางสูตรเท่านั้น ดังตัวอย่าง

1. ถ้าต้องการประมาณค่าเฉลี่ย (μ) ของประชากร สูตรที่ใช้ในการคำนวณขนาดของกลุ่มตัวอย่างจะเป็นดังนี้

$$n = \frac{N\ k^2\sigma^2}{N\ E^2+k^2\sigma^2} \qquad \text{.....................(8–1)}$$

เมื่อ n แทน จำนวนคนในกลุ่มตัวอย่าง

N แทน จำนวนคนในประชากร

σ^2 แทน ความแปรปรวนของประชากร

E แทน ขนาดของความคลาดเคลื่อนที่ยอมให้เกิด

k แทน ค่าคงที่โดยผู้วิจัยจะเป็นผู้กำหนดขึ้น โดย

ถ้า $k = 1$ จะตรงกับ $\alpha = .32$ จะมีความเชื่อมั่นในการประมาณ 68 %

$k = 2$ จะตรงกับ $\alpha = .05$ จะมีความเชื่อมั่นในการประมาณ 95 %

$k = 3$ จะตรงกับ $\alpha = .01$ จะมีความเชื่อมั่นในการประมาณ 99 %

โดยปกติค่าความแปรปรวนของประชากร (σ^2) ผู้วิจัยมักจะไม่ทราบค่าที่แท้จริง ผู้วิจัยจำเป็นจะต้องประมาณจากการสำรวจล่วงหน้า (Pilot study) หรืออาศัยจากข้อมูลในอดีตแทนค่าพารามิเตอร์ดังกล่าว

จากสูตร (8–1) สามารถทำให้อยู่ในรูปอย่างง่าย ๆ โดยที่ค่าที่คำนวณได้ไม่แตกต่างไปมากนัก ดังนี้

$$n = \left(\frac{ZS}{E}\right)^2 \qquad \text{.....................(8–2)}$$

เมื่อ Z แทนค่าอัตราส่วนวิกฤต Z หาได้จากการเปิดตารางตำแหน่งที่ $\frac{\alpha}{2}$

S แทน ความเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

2. ถ้าต้องการประมาณสัดส่วน สูตรที่ใช้ในการคำนวณขนาดของกลุ่มตัวอย่างจะเป็นดังนี้

$$n = \frac{Nk^2\, p\,(1-p)}{NE^2 + k^2 p\,(1-p)} \qquad \text{.....................(8–3)}$$

และสูตร (8–3) สามารถเขียนเป็นสูตรอย่างง่ายในการคำนวณ ได้ดังนี้

$$n = \frac{1}{4}\left(\frac{Z}{E}\right)^2 \qquad \text{.....................(8–4)}$$

3. ในกรณีที่กำหนดสัดส่วนของประชากร (Population proportion) เป็น π สูตรที่ใช้ในการคำนวณ จะเป็นดังนี้

$$n = \frac{Z^2\pi\,(1-\pi)\,N}{Z^2\pi\,(1-\pi) + Ne^2} \qquad \text{.....................(8–5)}$$

ถ้าให้ $\pi = 0.5$, $Z = 2$ จะได้

$$n = \frac{2^2\,(0.5)^2\,N}{2^2\,(0.5)^2 + Ne^2}$$

$$n = \frac{N}{1 + Ne^2} \qquad \text{.....................(8–6)}$$

เมื่อ e แทน α – error

จากสูตร (8–6) สามารถสรุปเป็นตารางแสดงขนาดของกลุ่มตัวอย่างได้ดังนี้

ตาราง 8—1 แสดงขนาดของกลุ่มตัวอย่าง เมื่อประชากรมีจำนวนตั้งแต่ 500 คน ขึ้นไป

Size of Population (N)	Sample Size (n) for Precision (e) of ±1%	±2%	±3%	±4%	±5%	±10%
500	b	b	b	b	222	83
1,000	b	b	b	385	286	91
1,500	b	b	638	441	316	94
2,000	b	b	714	476	333	95
2,500	b	1,250	769	500	345	96
3,000	b	1,364	811	517	353	**97**
3,500	b	1,458	843	530	359	97
4,000	b	1,538	870	541	364	98
4,500	b	1,607	891	549	367	98
5,000	b	1,667	909	**556**	370	98
6,000	b	1,765	938	566	375	98
7,000	b	1,842	959	574	378	99
8,000	b	1,905	976	580	381	99
9,000	b	1,957	989	584	383	99
10,000	5,000	2,000	1,000	588	385	99
15,000	6,000	2,143	1,034	600	390	99
20,000	6,667	2,222	1,053	606	392	100
25,000	7,143	2,273	1,064	610	394	100
50,000	8,333	2,381	1,087	**617**	397	100
100,000	9,091	2,439	1,099	621	398	100
$\rightarrow \infty$	10,000	2,500	1,111	625	400	100

ตัวอย่าง 1 จงคำนวณหาขนาดของกลุ่มตัวอย่างเพื่อใช้ในการสำรวจอัตราการบริโภคข้าวของไทยซึ่งมีประชากร 46,000,000 คน ความแปรปรวนของประชากร (σ^2) = 900 และในการประมาณอัตราการบริโภคข้าวครั้งนี้ยอมให้เกิดความคลาดเคลื่อนได้ไม่เกิน 5 ก.ก.

จากสูตร

$$n = \frac{N k^2 \sigma^2}{NE^2 + k^2 \sigma^2}$$

ถ้าให้ k = 3 หมายความว่าจะมีความเชื่อมั่นในการประมาณ 99 %

แทนค่าในสูตร จะได้

$$n = \frac{46,000,000 (3)^2 900}{46,000,000 (5)^2 + (3)^2 (900)}$$

$$= \frac{372,600,000,000}{1,150,000,000 + 8100}$$

$$= \frac{372,600,000,000}{1,150,008,100}$$

$$\simeq 324$$

ตัวอย่าง 2 จะต้องสุ่มกลุ่มตัวอย่างจำนวนเท่าใดจากประชากร 10000 คน โดยให้มีความเชื่อมั่นในการประมาณค่าเป็น 95%

จากสูตร $n = \frac{N}{1 + Ne^2}$

$$= \frac{10000}{1 + 10000(.05)^2}$$

$$= \frac{10000}{1 + 10000(.0025)}$$

$$= \frac{10030}{1 + 25}$$

$$= \frac{10000}{26}$$

$$= 384.615$$

$$\simeq 385$$

แบบฝึกหัด 8

1. จงบอกคุณสมบัติของ Random sample
2. ถ้าประชากรมีขนาด N มีความแปรปรวน σ^2 และให้ $\bar{X}$ เป็นค่าเฉลี่ยของ Random sample ซึ่งมีขนาด n จงพิสูจน์ว่าถ้า N มีจำนวนจำกัด หรือ Random Sample ขนาด n แบบใส่คืนแล้ว $\sigma^2_{\bar{x}} = \sigma^2 \frac{N-n}{n(N-1)}$
3. กำหนดให้ประชากรเป็นเซ็ตของ 1, 3, 5, 7, 7, 13, 19, 23 ถ้าสุ่มตัวอย่างแบบใส่คืนโดยให้มีขนาด $n = 40$ จงหาค่าเฉลี่ย และความแปรปรวนของการแจกแจงกลุ่มตัวอย่างของ $\bar{X}$
4. กำหนดให้ประชากรเป็นเซ็ตของ 13, 15, 17, 18, 24, 35 ถ้าสุ่มตัวอย่างแบบใส่คืน โดยให้มีขนาด $n = 35$ จงหาความน่าจะเป็นที่ค่าเฉลี่ยของกลุ่มตัวอย่างจะมีค่าระหว่าง 14 กับ 16
5. นักศึกษารามคำแหง 200,000 คน มีความสูงเฉลี่ย 150 ซ.ม. ความเบี่ยงเบนมาตรฐาน 8 ซ.ม. สุ่มเข้าชั้นแบบใส่คืน 200 ชั้น ๆ ละ 1000 คน จงหาความแปรปรวนของการแจกแจงกลุ่มตัวอย่างของ $\bar{X}$
6. จงยกตัวอย่างเพื่อแสดงว่าไม่ว่าการแจกแจงของประชากรจะเป็นแบบใดก็ตาม ถ้าสุ่มตัวอย่างแบบใส่คืน โดยที่ $n > 30$ แล้ว การแจกแจงของกลุ่มตัวอย่างของ $\bar{X}$ จะใกล้การแจกแจงปกติ
7. จงพิสูจน์ว่า $\mu_{\bar{x}_1 + \bar{x}_2} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2}$
8. จงพิสูจน์ว่า ถ้า $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots \bar{X}_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันแล้ว

$$\sigma^2_{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_n} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} + \ldots + \frac{\sigma_n^2}{n_n}$$

9. จงพิสูจน์ว่า $\sigma^2_{\bar{x}_1 \pm \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \pm \frac{2\sigma_1\sigma_2}{\sqrt{n_1 n_2}} r_{\bar{X}_1 \bar{X}_2}$
10. จงแสดงว่าถ้าประชากรมีการแจกแจงไบโนเมียล และ $\mu_{\hat{p}} = p$, $\sigma^2_{\hat{p}} = \frac{pq}{n}$

แล้ว ตัวแปรสุ่ม $Z = \frac{\cancel{(\hat{p} - p)\sqrt{n}}}{\sqrt{pq}}$ จะมีการแจกแจงปกติ