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

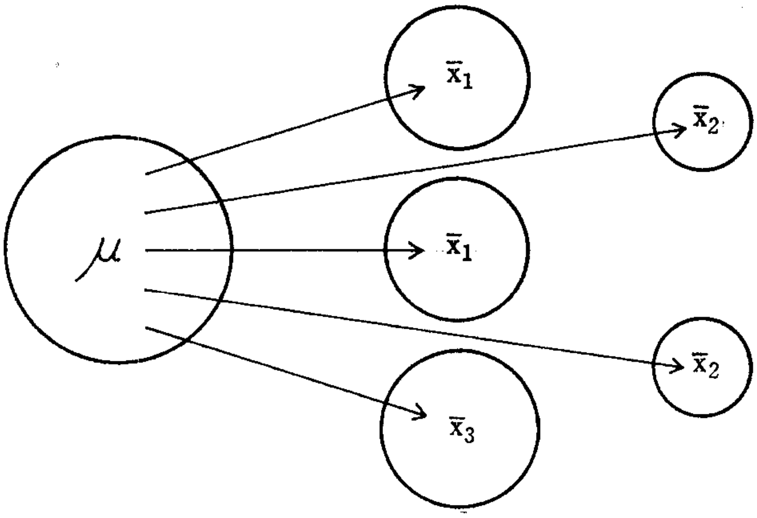
S^2 แทนความแปรปรวนของกลุ่มตัวอย่าง

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

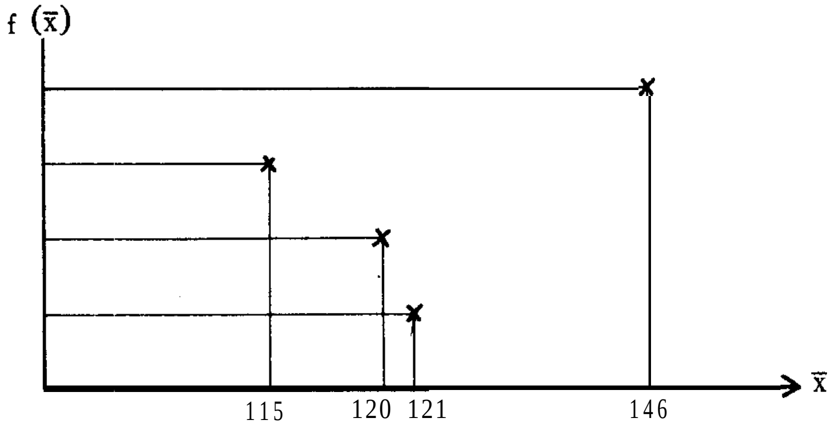
ความน่าจะเป็นของ $\bar{x}_1 = f(\bar{x}_1) = f(120) = \frac{20}{100}$

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



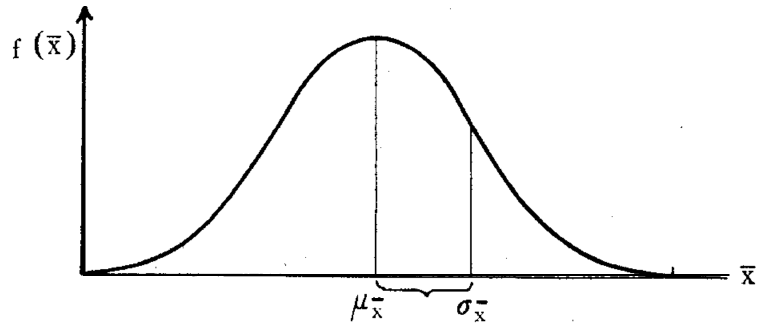
การแจกแจงดังกราฟในรูปนี้ก็จะเป็นการแจกแจงความน่าจะเป็นของ $\bar{x}$ และถ้าจำนวนกลุ่มตัวอย่างมีมากถึง infinite การแจกแจงความน่าจะเป็นของ $\bar{x}$ ก็จะเป็นฟังก์ชั่นที่ต่อเนื่องหรือเป็น curve

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

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

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

1.3 การสุ่มกลุ่มตัวอย่างแบบแบ่งชั้นหรือแบ่งพวก (Stratified Random Sampling) การสุ่มแบบนี้เป็นการสุ่มจากประชากรที่มีลักษณะแตกต่างกันหลายพวกหลายกลุ่ม 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