บทที่ 11. แบบแผนการสุ่มตัวอย่าง

Our knowledge, Our attitudes,
and our actions are based to
a very large extent upon samples.

W. G. Cochran.

การสุ่มตัวอย่าง (Sampling) เป็นกระบวนการของการเลือกตัวแทนซึ่งเป็นส่วนหนึ่งของประชากร และเป็นกระบวนการตรงกันข้ามกับการแจงนับทุกหน่วยในประชากรที่เรียกว่าการสำมะโน (Census) การสำรวจด้วยตัวอย่างได้ใช้แทนการแจงนับทุกหน่วยด้วยเหตุผลที่สำคัญดังนี้

(1) การสุ่มตัวอย่างทำได้รวดเร็วและถูกกว่า (Faster and Cheaper) เพราะตัวอย่างน้อยกว่าประชากร การเก็บข้อมูลและการสรุปผลจึงรวดเร็วกว่า และเสียเงินน้อยกว่า

(2) ตัวอย่างสุ่มสามารถให้ข่าวสารได้กว้างขวางกว่า (More Comprehensive) เพราะตัวอย่างน้อย ๆ สามารถสำรวจหรือสืบสวนได้ละเอียด สำหรับประชากรโต ๆ ค่าใช้จ่ายในการสำรวจจะมาก และเสียเวลามากด้วย

(3) ข้อมูลจากตัวอย่างมีความถูกต้องมากกว่า (More Accurate) เพราะสามารถกลั่นกรอง อบรม และควบคุมพนักงานที่ออกไปเก็บข้อมูลให้มีประสิทธิภาพได้

(4) เมื่อมีงบประมาณและเวลาจำกัด การสุ่มตัวอย่างจะให้ข่าวสารที่กว้างขวางกว่า และสามารถทำได้พร้อมกันหลาย ๆ ประชากร

11.1 กรอบตัวอย่าง (Sampling frames)

ในการสำรวจด้วยตัวอย่างนั้นมีประชากรที่ต้องพิจารณา 2 ชนิด คือ (1) ประชากรเป้าหมาย (Target Population) ซึ่งเป็นประชากรที่เราต้องการทราบรายละเอียดเกี่ยวกับข่าวสารต่าง ๆ และ (2) ประชากรตัวอย่าง (Sampling Population) ซึ่งเป็นประชากรที่ตัวอย่างได้รับการสุ่มจริง ๆ โดยพิจารณาจากกรอบตัวอย่าง กรอบตัวอย่าง (Sampling Frame) หมายถึงรายการของหน่วยตัวอย่าง (Sampling units) ซึ่งแทนประชากร

ถ้าเราต้องการประมาณรายจ่ายต่อเดือนของนักเรียน ม. 6 ในโรงเรียนแห่งหนึ่ง แล้ว**ประชากรเป้าหมาย** จะเป็นนักเรียน ม. 6 ทั้งหมดในโรงเรียนนั้นที่ลงทะเบียนไว้ตั้ง

แต่ต้นปี แต่ในระหว่างกลางปีอาจมีนักเรียนลาออกหรือย้ายไปเรียนที่อื่นได้ นักเรียนที่เหลือจะเป็น**ประชากรตัวอย่าง** รายการหรือรายชื่อนักเรียน ม. 6 ที่เหลือก็คือ **กรอบตัวอย่าง** และนักเรียนแต่ละคนที่มีชื่อในรายบัญชีรายชื่อก็จะเป็น**หน่วยตัวอย่าง**

11.2 การเลือกหน่วยตัวอย่าง

ในประชากรหนึ่งๆจะประกอบด้วยหน่วยแจงนับหรือหน่วยเบื้องต้น (Elementary Unit) แต่กรอบตัวอย่างสร้างจากหน่วยตัวอย่างบ่อยครั้งที่หน่วยแจงนับและหน่วยตัวอย่างเป็นหน่วยเดียวกัน แต่ก็ไม่จำเป็นเสมอไป ตัวอย่างต่อไปนี้จะแสดงความแตกต่างระหว่างประชากร เป้าหมาย และกรอบตัวอย่าง กับระหว่างหน่วยแจงนับและหน่วยตัวอย่าง และยังแสดงถึงแนวความคิดในการเลือกหน่วยตัวอย่าง

ตัวอย่าง (1) ถ้าต้องการประมาณคะแนนเฉลี่ยนักเรียน ม. 6 ของโรงเรียนรัฐบาลทั้งหมดในเขต กทม. แล้วหน่วยแจงนับจะเป็นนักเรียนแต่ละคน นักเรียน ม. 6 ทั้งหมดในโรงเรียนรัฐบาลที่อยู่ใน กทม. จะเป็นประชากรเป้าหมาย ถ้าไม่สามารถจัดเตรียมรายชื่อนักเรียนดังกล่าวได้ทั้งหมด แต่มีรายชื่อโรงเรียนรัฐบาลใน กทม. ทั้งหมด แล้ววิธีหนึ่งที่จะสุ่มได้คือการใช้รายชื่อโรงเรียนรัฐบาลเป็นกรอบตัวอย่าง และสุ่มโรงเรียนจากรายชื่อเหล่านั้น ดังนั้นโรงเรียนจะเป็นหน่วยตัวอย่าง

(2) ถ้าต้องการศึกษาเกี่ยวกับผู้หญิงในสถานกักกัน แล้วหน่วยแจงนับก็คือผู้หญิงคนหนึ่ง ๆ ในสถานกักกัน ผู้หญิงที่ถูกกักกันทั้งหมดเป็นประชากรเป้าหมาย ถ้ามีรายชื่อผู้หญิงเหล่านั้นทั้งหมด รายชื่อนั้นก็คือกรอบตัวอย่าง และสุ่มผู้หญิงตัวอย่างจากกรอบตัวอย่างนั้น ดังนั้นหน่วยตัวอย่างก็คือผู้หญิงหนึ่งคน

(3) ถ้าต้องการสำรวจทัศนคติของแม่บ้านใน กทม. เกี่ยวกับสินค้าชนิดหนึ่ง แล้วครัวเรือนหนึ่ง ๆ จะเป็นตัวแจงนับ รายชื่อครัวเรือนทั้งหมดใน กทม. เป็นกรอบตัวอย่าง ครัวเรือนทั้งหมดใน กทม. จะเป็นประชากรเป้าหมาย สมมติว่าไม่มีรายชื่อครัวเรือนทั้งหมดแต่มีแผนที่ของ กทม. ซึ่งมีรายละเอียดโดยการแบ่งครัวเรือนเป็นกลุ่ม ๆ ที่เรียกว่าชุมรุมอาคาร (Block) ต้องใช้ชุมรุมอาคารเป็นหน่วยตัวอย่าง และรายชื่อชุมรุมอาคารทั้งหมดจะเป็นกรอบตัวอย่าง

(4) ถ้าต้องการประมาณรายได้เฉลี่ยต่อปีของชาวนาในจังหวัดหนึ่ง หนวยแจง

นับก็คือชาวนาหนึ่งคน ประชากรเป้าหมายคือชาวนาทั้งหมดในจังหวัดนั้น ถ้ามีรายชื่อชาวนาทั้งหมดนั้น แต่เป็นบัญชีที่ล้าสมัยหรือไม่สมบูรณ์ เมื่อจะทำบัญชีใหม่ต้องใช้เงินและเวลามาก ดังนั้นจึงใช้วิธีสุ่มหน่วยตัวอย่างที่ใหญ่ขึ้น เช่น หมู่บ้าน รายชื่อหมู่บ้านทั้งหมดจะเป็นกรอบตัวอย่าง และหมู่บ้านจะเป็นหน่วยตัวอย่าง

11.3 เทคนิคการสุ่มตัวอย่าง (Sampling Techniques)

ต่อไปเราจะอธิบายถึงวิธีการสุ่มตัวอย่างทั้งวิธีโดยอาศัยความน่าจะเป็น (Probability Sampling) และไม่อาศัยความน่าจะเป็น (Nonprobability Sampling) แต่ละวิธีจะอธิบายวิธีการเลือกหน่วยตัวอย่าง และการประมาณคุณลักษณะประชากรที่สนใจในการศึกษา

11.3.1 การสุ่มตัวอย่างแบบธรรมดา (Simple Random Sampling)

การสุ่มตัวอย่างแบบธรรมดาเป็นวิธีการสุ่มตัวอย่างโดยใช้ความน่าจะเป็นที่ง่ายที่สุด วิธีนี้จะไม่ได้ใช้มากนักในเวลาทำงานสำรวจด้วยตัวอย่างจริง ๆ แต่ก็ช่วยให้เข้าใจวิธีการสุ่มตัวอย่างแบบอื่น ๆ ได้ดีขึ้น

การสุ่มตัวอย่างแบบธรรมดา เป็นกระบวนการเลือกตัวอย่างหรือส่วนย่อยจากกลุ่มของหน่วยตัวอย่างทั้งหมดโดยให้หน่วยตัวอย่างทุกหน่วย ที่อยู่ในกรอบตัวอย่างมีโอกาสถูกเลือกเป็นตัวอย่างเท่า ๆ กัน หรือเป็นกระบวนการเลือกตัวอย่างโดยที่ทุก ๆ ตัวอย่างที่เป็นไปได้มีโอกาสถูกเลือกเท่า ๆ กันหมด นั่นคือในการสุ่มตัวอย่างขนาด n จากประชากรขนาด N แล้วโอกาสที่หน่วยตัวอย่างใด ๆ จะมีโอกาสถูกเลือกเป็น $1/N$ หรือแต่ละตัวอย่างที่เป็นไปได้จำนวน $\binom{N}{n}$ นั้นมีโอกาสที่จะถูกเลือกเป็นตัวแทนของประชากรเท่า ๆ กันหมด นั่นคือเท่ากับ $1/\binom{N}{n}$

(1) วิธีการเลือกตัวอย่าง การเลือกหน่วยตัวอย่างเพื่อนำมาใช้เป็นตัวอย่างหรือตัวแทนประชากร โดยให้แต่ละหน่วยมีโอกาสถูกเลือกเท่า ๆ กันนั้น เรามีวิธีที่จะทำได้ดังนี้

(ก) วิธีจับฉลาก (Lottery Method) เป็นวิธีที่เหมาะกับการเลือกหน่วยตัวอย่างในประชากรตัวอย่างที่มีขนาดไม่โตนัก เช่นการเลือกครัวเรือนตัวอย่างในตำบลหนึ่ง การ

จับฉลากทำได้ 2 วิธีคือ โดยการแทนที่กับไม่แทนที่ (With and Without Replacement) ในทางปฏิบัติจริง ๆ ไม่นิยมใช้การแทนที่ เพราะไม่มีประโยชน์อันใดที่จะไปเก็บข้อมูลจากหน่วยตัวอย่างเดียวกันมากกว่าครั้งหนึ่ง

(ข) วิธีอาศัยตารางเลขสุ่ม (Random Numbers) ในกรณีที่ประชากรตัวอย่างมีหน่วยตัวอย่างมาก ๆ เราไม่สามารถเลือกโดยวิธีจับฉลากได้ เพราะต้องทำฉลากเป็นจำนวนมาก ดังนั้นจึงจำเป็นต้องใช้เครื่องมืออย่างหนึ่งช่วยซึ่งเรียกว่า ตารางเลขสุ่ม (เลขสุ่มคือ เลขที่ถูกสร้างขึ้นมาโดยกระบวนการที่ทำให้เลขเหล่านั้นมีโอกาสที่จะเกิดขึ้นเท่า ๆ กัน) ตารางเลขสุ่มมีผู้สร้างไว้หลายชุด การใช้ตารางเลขสุ่มมีวิธีการดังนี้

สมมติว่ามีประชากรตัวอย่างที่ประกอบด้วยหน่วยตัวอย่าง 800 หน่วย เราได้หมายเลขเรียงลำดับแก่หน่วยประชากรดังนี้

หน่วย	U_1	U_2	U_{800}
หมายเลข	001	002	800

ถ้าต้องการเลือกตัวอย่างขนาด 40 จากประเทศ 800 หน่วย เราก็ใช้ตารางเลขสุ่ม 3 หลักมีค่าไม่เกิน 800 แล้วทำการเลือกสุ่มในช่วง 001 – 800 ขึ้นมาทีละ 3 ตัวโดยไม่เจาะจง เช่นถ้าได้ 315 หน่วย หน่วยที่ตกเป็นตัวอย่างคือ U_{315} ต่อไปถ้าได้ 043 เราก็ได้ U_{043} และเลือกต่อ ๆ ไปจนกระทั่งครบจำนวน 40 หมายเลข ซึ่งเป็นจำนวนตัวอย่างทั้งหมดที่ต้องการ

(2) การประมาณค่า เมื่อได้หน่วยตัวอย่างที่ต้องการแล้ว ผู้สำรวจก็เริ่มสอบถามรายละเอียดที่ต้องการศึกษาได้จากหน่วยตัวอย่างเหล่านี้ ข้อมูลที่ได้มาก็จะใช้ประมาณค่าต่าง ๆ ของประชากร ดังนี้

ก. ยอดรวมประชากร (T) ให้ X_i เป็นข้อมูลที่แทนลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ i ของประชากรขนาด N แล้วยอมรวมประชากรจะเป็น

$$\tau = \sum_{1}^{N} X_i$$

เมื่อใช้ตัวอย่างขนาด n เพื่อประมาณยอมรวมประชากร แล้วตัวประมาณค่ายอมรวมประชากร จะเป็น

$$\hat{\tau} = N\bar{X} = N(\sum_{i}^{n} X_i/n)$$

ในเมื่อ $\bar{X}$ เป็นค่าเฉลี่ยตัวอย่าง

ความคลาดเคลื่อนมาตรฐานของยอดรวม $\hat{\tau}$ จะประมาณได้จาก

$$S_{\hat{\tau}} = \sqrt{N^2(1-n/N)S^2/n}$$

ในเมื่อ S^2 เป็นความแปรปรวนตัวอย่างซึ่งกำหนดไว้ว่า $S^2 = \Sigma(X-\overline{X})^2/(n-1)$ และเทอม $(1-n/N)$ นี้เรียกว่าการแก้ไขความต่อเนื่อง (Finite Population Correction, f.p.c.) ในทางปฏิบัติจะไม่ใช้ fpc เมื่อ $1-n/N \geq .95$ หรือ $n/N \leq .05$

ข. ค่าเฉลี่ยประชากร (μ) ถ้า X_i เป็นข้อมูลจากหน่วยตัวอย่างที่ i ของประชากรขนาด N แล้วค่าเฉลี่ยประชากรจะเป็น

$$\mu = \sum_{i}^{N} X_i / N = \tau/N$$

ตัวประมาณค่าของค่าเฉลี่ยประชากรโดยอาศัยตัวอย่างขนาด n จะเป็น

$$\hat{\mu} = \overline{X} = \sum_{i}^{n} X_i/n$$

ซึ่งมีความคลาดเคลื่อนมาตรฐานที่ประมาณได้เป็น

$$S_{\overline{X}} = \sqrt{(1-n/N)S^2/n}$$

ตัวอย่าง ในการศึกษาค่าใช้จ่ายอาหารกลางวันของนักเรียนชั้นหนึ่งมี 15 คน โดยอาศัยตัวแทนนักเรียน 6 คน ได้ข้อมูลมาดังนี้

10, 6, 8, 12, 5, 7

นักเรียนชั้นนี้ใช้จ่ายอาหารกลางวันโดยเฉลียวันละ $\overline{X} = 8$ บาท

$$\overline{X} = (10+6+8+12+5+7)/6 = 8$$

และนักเรียนทั้งชั้นใช้จ่ายค่าอาหารกลางวันวันละ $\hat{\tau} = 15(8) = 120$ บาท

ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยและยอมรวมตัวอย่าง ประมาณได้จากข้อมูลดังนี้

$$S^2 = ((10-8)^2+(6-8)^2+\ldots+(7-8)^2/(6) = 6.8$$

$$S_{\overline{X}} = \sqrt{(1-6/15)(6.8/6)} = \sqrt{0.68} = 0.82$$

ไม่ตัด f.p.c ทิ้งเพราะ $1-6/15 = .6$ ซึ่งไม่มากกว่า .95

$$S_{\hat{\tau}} = \sqrt{15^2(1-6/15)(6.8/6)} = 12.37$$

ค. สัดส่วนหรือร้อยละในประชากร (π. 100 π) ถ้าให้ $X_i = 1$ ถ้าหน่วยตัวอย่างที่ i มีลักษณะที่สนใจ และ $X_i = 0$ ถ้าไม่มีลักษณะที่สนใจ แล้วสัดส่วนประชากรของลักษณะที่สนใจจะเป็น

$$\pi = \sum_{i}^{N} X_i/N; \quad X_i = 0,1$$
$$= A/N$$

ในเมื่อ A เป็นยอดรวมของหน่วยตัวอย่างที่มีลักษณะที่สนใจ

ตัวประมาณค่าของสัดส่วนประชากรโดยอาศัยตัวอย่างขนาด n นั้นกำหนดไว้เป็น

$$p = \hat{\pi} = \sum_{i}^{n} X_i/n \;, \quad X_i = 0,1$$
$$= a/n$$

ในเมื่อ a. เป็นจำนวนหน่วยตัวอย่างที่มีลักษณะที่สนใจในตัวอย่างขนาด n

สำหรับยอดรวมประชากรประมาณได้จาก

$$\hat{A} = NP$$

ความคลาดเคลื่อนมาตรฐานของสัดส่วนและยอดรวมตัวอย่างจะประมาณได้โดยอาศัยตัวอย่างดังนี้

$$S_p = \sqrt{(1-n/N)P(1-P)/(n-1)}$$
$$S_a = NS_p$$

ตัวอย่าง ถ้าสนใจสัดส่วนของนักเรียนที่อยู่ต่างจังหวัด จากตัวอย่างนักเรียน 6 คนแล้ว พบว่าอยู่ต่างจังหวัด 2 คน ดังนั้นสัดส่วนของนักเรียนต่างจังหวัดจะเป็น

$$P = 2/6 = 1/3$$

โดยมีความคลาดเคลื่อนมาตรฐานเป็น

$$S_p = \sqrt{(1-6/15)(1/3)(1-1/3)/(6-1)} = 0.163$$

ง. อัตราส่วน (R) ถ้า X_i เป็นลักษณะอย่างหนึ่งของหน่วย i ในประชากร และ Y_i เป็นอีกลักษณะหนึ่งของหน่วย i แล้วอัตราส่วนระหว่างสองลักษณะ จะเป็น

$$R = \tau_y/\tau_x = \mu_y/\mu_x$$

ตัวประมาณค่าของ R โดยอาศัยตัวอย่างขนาด n จะเป็น

$$\hat{R} = \hat{\tau}_y/\hat{\tau}_x = \bar{Y}/\bar{X}$$

ซึ่งมีความคลาดเคลื่อนมาตรฐานที่ประมาณได้เป็น

$$S_{\hat{R}} = \sqrt{\{\sum Y_i^2 - 2\hat{R}\sum X_i Y_i + \hat{R}^2 \sum X_i^2\}/n(n-1)\bar{X}^2}\sqrt{(1-n/N)}$$

(3) การกำหนดขนาดตัวอย่าง จากตัวอย่างที่ผ่านมานั้นเราได้กำหนดขนาดตัวอย่างตามใจชอบ โดยไม่ทราบว่าขนาดตัวอย่างที่กำหนดไปนั้นอาจจะเล็กเกินไปหรือโตเกินไป วิธีการกำหนดขนาดตัวอย่างนั้นมีกรณีที่กำหนดได้ดังนี้

ก. กรณีที่ทราบค่าความแปรปรวน σ^2 หรือสัดส่วน π ก่อนที่จะกำหนดขนาดตัวอย่างนั้นมีสิ่งที่จำเป็นต้องทราบหรือตัดสินใจดังนี้

— จำนวนหน่วยตัวอย่างทั้งหมดในประชากร N

— ความคลาดเคลื่อนสูงสุดที่จะยอมรับได้ α

— ตัวแปรปกติมาตรฐาน Z ซึ่งจะให้ค่าความเชื่อมั่นตามระดับที่ต้องการ ค่าของตัวแปรปกติมาตรฐานที่ใช้กันเสมอมีดังนี้

เปอร์เซนความเชื่อมั่น	80 %	90 %	95 %	100 %
Z	1.290	1.645	1.960	3.00 (ประมาณ)

— ความรู้เกี่ยวกับ σ^2 หรือ π ค่าเหล่านี้สามารถประมาณการ, หรือเดาได้ก่อนการสำรวจโดยวิธีต่าง ๆ ดังนี้

(ก) ใช้ผลของการสำรวจเบื้องต้น (Pilot Survey)

(ข) สุ่มตัวอย่างเป็น 2 ขั้นตอน คือ ขั้นที่หนึ่ง เลือกตัวอย่างขนาดเล็ก n_1 เพื่อประมาณ σ^2 หรือ π และขั้นที่สอง ใช้ค่าที่ประมาณได้จากขั้นแรกมาคำนวณตัวอย่างคือ n_2 แล้วสุ่มตัวอย่างเพิ่มเติมจนครบ $(n_2 - n_1)$

(ค) ใช้ผลการสำรวจก่อน ๆ ในประชากรเดียวกัน

(ข) ใช้วิธีเดา (Guesswork) หลังจากปรึกษาหารือกับผู้ที่มีความรู้เกี่ยวกับประชากรที่จะสำรวจนั้น

ขนาดตัวอย่างสำหรับการประมาณค่าเฉลี่ยและสัดส่วนกำหนดไว้ดังนี้

$$n = NZ^2S^2/(Nd^2+Z^2S^2)$$
$$n = NZ^2P(1-P)/\{Nd^2+Z^2P(1-P)\}$$

ในเมื่อ Z เป็นค่าของตัวแปรเชิงสุ่มตามระดับความเชื่อมั่น, S^2 เป็นความแปรปรวน d เป็นความคลาดเคลื่อน, และ P เป็นสัดส่วน

ในกรณีที่ไม่มีความรู้เกี่ยวกับ π เลย เราจะใช้ $P(1-P)$ ที่ใหญ่สุด คือเมื่อ $P = 0.50$ หรือ $P(1-P) = 0.25 = 1/4$ ดังนั้นขนาดตัวอย่างซึ่งโตสุดจะเป็น

$$n \leq NZ^2/(4Nd^2+Z^2)$$

ตัวอย่าง ต้องการประมาณขนาดครัวเรือนของหมู่บ้านแห่งหนึ่งซึ่งมีทั้งหมด 150 ครัวเรือน โดยให้ค่าเฉลี่ยขนาดครัวเรือนมีความคลาดเคลื่อนจากค่าเฉลี่ยจริงไม่เกิน 1 หน่วย (นั่นคือ ถ้า $\mu = 6.0$ แล้ว $\bar{X}$ ต้องอยู่ระหว่าง 5.0 ถึง 7.0) ไม่ทราบค่า σ^2 แต่จากการสำรวจครั้งก่อนในหมู่บ้านนั้นพบว่า $\sigma^2 = 9.0$ เมื่อต้องการความเชื่อมั่น 90% จะต้องใช้ขนาดตัวอย่างเป็น

$$\begin{aligned} n &= NZ^2S^2/(Nd^2+Z^2S^2) = 150(1.645)^2(9.0)/\{150(1)^2+(1.645)^2(9.0)\} \\ &= 20.9 \approx 21 \end{aligned}$$

ตัวอย่าง จากตัวอย่างที่แล้วถ้าต้องการประมาณสัดส่วนครัวเรือนที่มีเครื่องรับวิทยุหรือโทรทัศน์ โดยให้มีความคลาดเคลื่อนไม่เกิน 0.10 ด้วยความเชื่อมั่น 100% แล้วจะต้องใช้ขนาดตัวอย่างเท่าใด ถ้าเดาว่าเปอร์เซนต์ครัวเรือนที่มีวิทยุหรือโทรทัศน์มีประมาณ 40%?

$$\begin{aligned} n &= NZ^2P(1-P)/\{Nd^2+Z^2P(1-P)\} \\ &= 150(3.0)^2(0.40)(1-0.40)/\{150(0.10)^2+(3.0)^2(0.40)(1-0.40)\} \\ &= 88.5 \cong 89 \end{aligned}$$

ถ้าไม่มีความรู้เกี่ยวกับประชากรนี้เลย เราก็เดาค่า P ไม่ได้เราก็จะได้ขนาดตัวอย่างโตที่สุดเป็น

$$n \leq 150(3)^2/\{4(150)(0.10)^2+(3.0)^2\} = 90$$

ข. กรณีที่ต้องการศึกษาคุณลักษณะมากกว่าหนึ่งอย่าง ในกรณีแรกนั้นขนาดตัวอย่างพิจารณาโดยพื้นฐานของคุณลักษณะเดียว เช่นขนาดครัวเรือนเท่านั้น หรือสัดส่วนเท่านั้น ถ้าเราต้องการศึกษามากกว่าหนึ่งคุณลักษณะ เช่นขนาดครัวเรือน สัดส่วน และรายได้ แล้วขั้นตอนในการกำหนดขนาดตัวอย่างจะเป็นดังนี้

- เลือกลักษณะทั้งหมดที่เกี่ยวข้องกับการศึกษาและสำคัญต่อการสำรวจ
- ประมาณขนาดตัวอย่างสำหรับแต่ละลักษณะที่เลือก
- พิจารณาว่าขนาดตัวอย่างสำหรับแต่ละลักษณะที่ศึกษาใกล้เคียงกันหรือไม่?

ถ้าใกล้เคียงกันให้ n ที่โตสุด หรือเฉลี่ยค่า n ต่าง ๆ แต่ก็ขึ้นกับงบประมาณด้วย ถ้าไม่ใกล้เคียงกันและความเที่ยงตรงของค่าประมาณ (Z และ d) ไม่สนใจ แล้วก็เลือกค่าของ n ที่น้อยกว่า แต่อย่างไรก็ตามก็ขึ้นอยู่กับข้อจำกัดในเรื่องงบประมาณ

ค. กรณีที่ไม่ทราบความแปรปรวน σ^2 หรือสัดส่วน π เมื่อไม่ทราบค่า π เราสามารถใช้สูตรที่กำหนดขนาดตัวอย่างโตสุดดังกล่าวแล้วในกรณี ก. และถ้าไม่ทราบค่า σ^2 เลยก็ให้กำหนดเป็นเปอร์เซนต์ของประชากร เช่น 2, 5, 10, 20, หรือ 50% ของ N แต่มีข้อควรพิจารณาในการกำหนดเปอร์เซนต์ดังนี้

– เมื่อขนาดประชากรโตมาก ให้ใช้เปอร์เซนต์น้อย ๆ และขนาดประชากรน้อยให้ใช้เปอร์เซนต์สูง

– ขนาดตัวอย่างไม่ควรต่ำกว่า 30

– ขนาดตัวอย่างควรจะมากเท่าที่งบประมาณและเวลาจะเอื้ออำนวยให้

(4) ข้อได้เปรียบและเสียเปรียบ การสุ่มตัวอย่างแบบธรรมดานั้นมีข้อได้เปรียบและเสียเปรียบดังนี้

– ทฤษฎีที่เกี่ยวข้องง่ายต่อการเข้าใจ

– เมื่อหน่วยตัวอย่างในประชากรมีความแปรปรวนสูง ความเที่ยงตรงจะน้อยเนื่องจากตัวอย่างที่สุ่มมาได้นั้นไม่ทำหน้าที่ตัวแทนที่ดี.

– การเลือกตัวอย่างใช้วิธีเลือกจากบัญชีรายชื่อหน่วยตัวอย่างโดยการกำหนดหมายเลขแก่หน่วยตัวอย่างนั้น บางครั้งในทางปฏิบัติไม่สะดวก

– มีปัญหาทางด้านภูมิศาสตร์ เช่น ถ้าจะสุ่มชาวนาไทยทั่วประเทศ ใช้วิธีสุ่มแต่ละภาค หรือแต่ละจังหวัดจะสะดวกกว่า

11.3.2 การสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Random Sampling)

บ่อยครั้งที่ประชากรประกอบด้วยหน่วยตัวอย่างที่ไม่เหมือนกันทางด้านคุณลักษณะที่ศึกษา จึงควรแยกประชากรออกเป็นประชากรย่อย (Subpopulation) โดยให้หน่วยต่าง ๆ ที่เหมือนกันอยู่ในประชากรย่อยเดียวกัน ประชากรย่อยนั้นเรียกว่าชั้นภูมิ (Strata)

แล้วจึงทำการสุ่มตัวอย่างจากแต่ละชั้นภูมินั้น

(1) วิธีการเลือกตัวอย่าง ในการเลือกตัวอย่างนั้นมีขั้นตอนในการเลือก 2 ขั้น คือ ก. แยกประชากรออกเป็นชั้นภูมิโดยให้หน่วยตัวอย่างในชั้นภูมิมีคุณลักษณะคล้าย ๆ กัน และ ข. สุ่มตัวอย่างจากแต่ละชั้นภูมินั้น

(2) การประมาณค่า ให้ประชากรขนาด N แบ่งออกเป็น l ชั้นภูมิโดยมีขนาดเป็น $N_1 N_2,\ldots, N_l$ และตัวอย่างจากแต่ละชั้นภูมิมีขนาดเป็น $n_1\, n_2,\ldots, n_l$ โดยมี $n_1 + n_2 + \ldots + n_l = n$ การแบ่งขนาดตัวอย่าง n ออกเป็นส่วน ๆ นี้มีผลกระทบกระเทือนถึงการประมาณค่าพารามิเตอร์ที่สนใจด้วย จึงต้องพิจารณาให้ดี สำหรับพารามิเตอร์ที่จะประมาณค่าโดยทั่วไปจะเป็นดังนี้

ก. ค่าเฉลี่ยประชากร ให้ $\overline{X}_{st}$ เป็นตัวประมาณค่าของค่าเฉลี่ยประชากรในเมื่อประชากรแบ่งเป็นชั้นภูมิก่อนการสุ่มตัวอย่าง และได้ $\overline{X}_h$ เป็นตัวประมาณค่าของค่าเฉลี่ยในชั้นภูมิ h จากชั้นภูมิทั้งหมด l ชั้นภูมิ แล้ว $\overline{X}_{st}$ จะกำหนดไว้ดังนี้

$$\overline{X}_{st} = (1/N)\sum_h^l N_h\overline{X}_h = \sum_h^l (N_h/N)\overline{X}_h = \sum_h^l W_h\overline{X}_h$$

ซึ่งมีความแปรปรวนหรือกำลังสองของความคลาดเคลื่อนมาตรฐานประมาณได้เป็น

$$V(\overline{X}_{st}) = \sum_h^l \frac{N_h^2}{N^2}\left((N_h-n_h)/N_h\right)\left(S_h^2/n_h\right) = (1/N^2)\sum_h^l N_h(N_h-n_h)\, S_h^2/n_h$$

ในเมื่อ S_h^2 เป็นความแปรปรวนตัวอย่างในชั้นภูมิ h นั่นคือ

$$S_h^2 = \sum_i^{n_h} (X_{hi}-\overline{X}_h)^2/(n_h-1)$$

N_h และ n_h เป็นขนาดประชากรและตัวอย่างในชั้นภูมิ h, $\overline{X}_h$ เป็นค่าเฉลี่ยตัวอย่างในชั้น

ภูมิ h และ x_{hi} เป็นข้อมูลที่ i ในชั้นภูมิ h.

เมื่อขนาดของชั้นภูมิเป็นขนาดใหญ่ นั่นคือ f. p. c. $= (1 - n_h / N_h) \geq 0.95$ สำหรับทุกชั้นภูมิ แล้ว $V(\bar{X}_{st})$ จะเป็น

$$V(\bar{X}_{st}) = \frac{1}{N^2} \Sigma N_h^2 S_h^2 / n_h$$

ข. ยอดรวมประชากร ตัวประมาณค่ายอดรวมประชากรกำหนดไว้ดังนี้

$$\hat{\tau}_{st} = \sum_{h}^{1} N_h \bar{X}_n$$

โดยมีความแปรปรวนที่ประมาณได้เป็น

$$V(\hat{\tau}_{st}) = \sum_{h}^{1} N_h(N_h - n_h) S_h^2 / n_h$$

ตัวอย่าง ในการสำรวจขนาดครัวเรือนและยอดรวมพลเมืองในเมืองหนึ่งได้แบ่งเขตสำรวจเป็น 5 ชั้นภูมิ ตามลักษณะภูมิศาสตร์ แล้วจึงสุ่มแบบธรรมดาจากแต่ละชั้นภูมิ ได้ข้อมูลสรุปดังนี้

h	N_2	n_o	$\bar{X}_h$	S_h^2
1	448	81	6.49	6.65
2	131	31	6.77	12.11
3	81	14	6.50	4.58
4	108	20	7.25	7.57
5	100	17	6.76	5.19

$N = 868 \quad n = 163$

$$\bar{X}_{st} = (1/868) \left\{ 448\,(6.49) + 131\,(6.77) + \ldots + 100\,(6.76) \right\}$$

$$= 6.66 \approx 7.0$$

นั้นคือขนาดครัวเรือนในเมืองนี้ประมาณครัวเรือนละ 7 คน

ค่าประมาณความแปรปรวนของ $\bar{X}_{st}$ จะเป็น

$$V(\bar{X}_{st}) = (1/868^2) \left\{ 448\,(448 - 81)\,(6.65)/81 + \ldots\ldots \right.$$

$$........ + 100\ (100-17)\ (6.76)/17 \}$$

$$= 0.0352$$

สำหรับยอดรวมพลเมืองจะเป็น

$$\hat{T}_{st} = 448\ (6.49) + 131\ (6.77) + ... + 100\ (6.76)$$

$$= 5779.89 \simeq 5780.$$

นั่นคือพลเมืองในเมืองนั้นมีประมาณ 5780 คน

ค. สัดส่วนประชากร ในการประมาณสัดส่วนประชากรมีวิธีการประมาณเหมือนกับค่าเฉลี่ยทุกประการ แต่ค่าของ X_{hi} เป็น 0 หรือ 1 เท่านั้น นั่นคือตัวประมาณค่าของ π_{st} จะเป็น

$$P_{st} = (1/N) \sum_{h}^{1} (N_h/n_h) \sum_{i}^{n_h} X_{hi} \quad ; \quad X_{hi} = 0,1$$
$$= (1/N) \sum_{h}^{1} N_h \quad P_h$$

ในเมื่อ P_h เป็นสัดส่วนตัวอย่างในชั้นภูมิ h และ $P_h = \sum_{i}^{n_h} X_{hi}/n_h$

ค่าประมาณความแปรปรวนของสัดส่วนตัวอย่างจะกำหนดไว้ดังนี้

$$V(P_{st}) = \sum_{h}^{1} (N_h^2/N^2)(1-n_h/N_h) \ P_h(1-P_h)/(n_h-1)$$
$$= (1/N^2) \sum_{h}^{1} N_h(N_h-n_h) \ P_h(1-P_h)/(n_h-1)$$

สำหรับยอดรวมประชากรจากสัดส่วนประมาณได้จาก

$$\hat{X}_{st} = \sum_{h}^{1} N_h P_h = NP_{st}$$

ตัวอย่าง ในการประมาณสัดส่วนของหัวหน้าครัวเรือนที่เห็นด้วยกับการประหยัดน้ำมันของรัฐบาลโดยการแบ่งเป็นชั้นภูมิ ได้ข้อมูลสรุปดังนี้

h	N_h	n_h	P_h	$1-P_h$
1	448	74	0.730	0.270
2	131	15	0.800	0.200
3	81	26	0.923	0.077

4	108	17	0.882	0.118
5	100	31	0.774	0.226

$N = 868 \quad n = 163$

$$P_{st} = (1/868)\left\{448\ (0.730) + \ldots\ldots + 100\ (0.774)\right\}$$

$$= 679.26 / 868 = 0.780$$

$$V(P_{st}) = (1/868^2)\left\{448\ (448-74)\ (0.730)\ (0.270)/(74-1) + \ldots\ldots\ldots + 100\ (100-31)\ (0.774)\ (0.226)/(31-1)\right\}$$

$$= 742.982 / 868^2 = 0.00099$$

สำหรับยอดรวมหรือจำนวนหัวหน้าครัวเรือนที่เห็นด้วยกับการประหยัดน้ำมันจะประมาณได้เป็น

$$\hat{X} = 448\ (0.730) + 131\ (0.800) + \ldots\ldots + 100\ (0.774)$$

$$= 679.26 \simeq 680.$$

ง. **อัตราส่วน** การประมาณอัตราส่วนก็ทำได้เช่นเดียวกับการสุ่มตัวอย่างแบบธรรมดา นั่นคือ

$$\hat{R} = \bar{Y}_{st}/\bar{X}_{st} = \hat{Y}_{st}/\hat{X}_{st}$$

$$= \sum_{h}^{I} N_h\bar{Y}_h \Big/ \sum_{h}^{1} N_h\bar{X}_h$$

(3) **การกำหนดและการจัดสรรขนาดตัวอย่าง** ในการสุ่มตัวอย่างแบบแบ่งชั้นภูมิเรามีการกำหนดและจัดสรรขนาดตัวอย่างเป็น 2 ขั้นตอน ดังนี้

— กำหนดขนาดตัวอย่างทั้งหมด (n)

— จัดสรรตัวอย่างให้แก่ชั้นภูมิต่าง ๆ

สำหรับวิธีการกำหนดและจัดสรรนั้นเราใช้กันอยู่เสมอมี 4 วิธี ดังนี้

ก. วิธีจัดสรรแบบเท่ากัน (Equal Allocation Method) วิธีนี้ควรใช้เมื่อ

(ก. 1) จำนวนหน่วยตัวอย่างในแต่ละชั้นภูมิ (N_h) มีจำนวนใกล้เคียงกัน

(ก. 2) ความแปรปรวนในชั้นภูมิ (S_h^2) และค่าใช้จ่ายต่อหน่วยตัวอย่าง (C_h) ไม่แตกต่างกันมากนัก และ (ก. 3) ไม่ทราบค่าที่แท้จริงของ σ_h^2 หรือ π_h และ C_h

สำหรับขนาดตัวอย่างทั้งหมด และขนาดตัวอย่างในแต่ละชั้นภูมิกำหนดไว้ดังนี้

$$n = 1 \sum_h^1 N_h^2 S_h^2 \Big/ \left(N^2 d^2/Z^2 + \sum_h^1 N_h S_h^2\right), \qquad n_h = n/1$$

ในเมื่อ d_1, Z_1 และ S_h^2 เป็นความคลาดเคลื่อน, ค่าของตัวแปรปกติมาตรฐาน, และความแปรปรวนในชั้นภูมิ

ข. วิธีจัดสรรแบบสัดส่วน (Proportional Allocation Method) วิธีนี้ควรใช้เมื่อ (ข. 1) จำนวนหน่วยตัวอย่างในชั้นภูมิ (N_h) แตกต่างกันมาก (ข. 2) และ (ข. 3) เหมือนกับวิธีจัดสรร ก. สำหรับขนาดตัวอย่างทั้งหมด และขนาดตัวอย่างในแต่ละชั้นภูมิกำหนดไว้ดังนี้

$$n = N \sum_h^1 N_h^2 S_h^2 \Big/ \left(N^2 d^2/Z^2 + \sum_h^i N_h S_h^2\right), \quad n_h = n(N_h/N)$$

ค. วิธีจัดสรรแบบเนย์แมน (Neyman Allocation Method) วิธีควรใช้เมื่อคาดว่าความแปรปรวนหรือสัดส่วนในชั้นภูมิ (σ_h^2 หรือ π_h) จะแตกต่างกันมาก และกำหนดไว้ดังนี้

$$n = \left(\sum_h^1 N_h S_h\right)^2 \Big/ \left(N^2 d^2/Z^2 + \sum_h^1 N_h S_h^2\right), \quad n_h = n(N_h S_h \Big/ \sum_h^1 N_h S_h)$$

ง. วิธีจัดสรรแบบดีที่สุด (Optimum Allocation Method) วิธีควรใช้เมื่อคาดว่าความแปรปรวน (σ_h^2) หรือสัดส่วน (π_h) และค่าใช้จ่ายต่อหน่วยต่ออย่าง (C_h) จะแตกต่างกันระหว่างชั้นภูมิต่าง ๆ สำหรับการจัดสรรกำหนดไว้ดังนี้

$$n = \left(\sum_n^1 N_h S_h \sqrt{C_h}\right)\left(\sum_h^1 N_h S_h / \sqrt{C_h}\right) \Big/ \left(N^2 d^2/Z^2 + \sum_h^1 N_h S_h^2\right),$$

$$n_h = n N_h S_h / \sqrt{C_h} \Big/ \left(\sum_h^1 N_h S_h / \sqrt{C_h}\right)$$

ตัวอย่าง ต้องการประมาณขนาดครัวเรือนเฉลี่ยโดยให้มีความคลาดเคลื่อนไม่เกิน 0.50 หน่วยจากค่าที่แท้จริง ด้วยความเชื่อมั่น 90% จะต้องใช้ขนาดตัวอย่างเท่าใด และจะต้องสุ่มจากชั้นภูมิต่าง ๆ ด้วยจำนวนเท่าใด ถ้ามีความรู้เกี่ยวกับองค์ประกอบต่าง ๆ ดังตารางต่อไปนี้

h	N_h	S_h^2	C_h	$N_h S_h^2$	$N_h S_h$	$N_h^2 S_h^2$	$N_h S_h / \sqrt{C_h}$	$N_h S_h \sqrt{C_h}$
1	448	6	2	2688	1097.6	1204224	776.2	1552.0

2	131	10	2	1310	414.2	171610	292.9	585.7
3	81	5	3	405	181.1	32805	104.6	313.7
4	108	7	3	756	285.8	81648	165.0	495.0
5	100	5	1	500	223.6	50000	223.6	223.6
	$N = 868$			5659	2202.3	1540287	1562.3	3170.0

ก. วิธีจัดสรรแบบแบ่งเท่ากัน

$$n = \frac{5(1540287)}{(868)^2(0.50)^2/11.645)^2 + 5659} = 102.7 \simeq 103$$

$$n_h = 103/5 = 20.6 \simeq 21$$

ข. วิธีจัดสรรแบบสัดส่วน

$$n = \frac{868(5659)}{(868)^2(0.50)^2/(1.645)^2 + 5659} = 65.5 \simeq 66$$

$$n_1 = 66(448/868) = 34, \quad n_2 = 66(131/868) = 9.9 \simeq 10$$

$$n_3 = 66(81/868) = 6.2 \simeq 7, \quad n_4 = 66(108/868) = 8.2 \simeq 9$$

$$n_5 = 66(100/868) = 7.6 \simeq 8$$

ค. วิธีจัดสรรแบบเนย์แมน

$$n = \frac{(2202.3)^2}{(868)^2(0.50)^2/(1.645)^2 + 5659} = 64.7 \simeq 65$$

$$n_1 = 65(1097.6/2202.3) = 31.8 \simeq 32$$

$$n_2 = 65(414.2/2202.3) = 12$$

$$n_3 = 65(181.1/2202.3) = 5.3 \simeq 6$$

$$n_4 = 65(285.8/2202.3) = 8.3 \simeq 9$$

$$n_5 = 65(223.6/2202.3) = 6.5 \simeq 7$$

ง. วิธีจัดสรรแบบดีที่สุด

$$n = \frac{3170(1562.3)}{(868)^2(0.50)^2/(1.645)^2 + 5659} = 65.9 \simeq 66$$

$$n_1 = 69(776.2)/(1562.3) = 32.8 \simeq 33$$

$$n_2 = 66(292.9)/(1562.3) = 12.4 \simeq 13$$

$$n_3 = 66(104.6)/(1562.3) = 4.4 \simeq 5$$

$$n_4 = 66(165.0)/(1562.3) = 6.9 \simeq 7$$

$$n_5 = 66(223.6)/(1562.3) = 9.4 \simeq 10$$

ตัวอย่าง จากตัวอย่างที่แล้วนี้ ถ้าต้องการประมาณสัดส่วนของหัวหน้าครัวเรือนที่เห็นด้วยกับนโยบายเศรษฐกิจของรัฐบาล โดยให้มีความคลาดเคลื่อนไม่เกิน 0.05 ด้วยความเชื่อมั่น 90% แล้วเราสามารถกำหนดตัวอย่างได้เช่นเดียวกับวิธีที่ผ่านมาแต่ใช้ $P_h(1-P_h)$ แทน S_h^2 และอาศัยตารางต่อไปนี้

h	P_h	$N_hP_hQ_h$	$N_h^2P_hQ_h$	$N_h\sqrt{P_hQ_h}$	$N_h\sqrt{P_hQ_h}\sqrt{C_h}$	$N_h\sqrt{P_hQ_h}/\sqrt{C_h}$
1	0.7	94.1	42147	205.2	290.1	145.1
2	0.7	27.5	3603	60.0	84.8	42.4
3	0.7	17.0	1378	37.1	64.3	21.4
4	0.7	11.7	1449	49.5	85.7	28.6
5	0.7	21.0	2100	45.8	45.8	45.8
		182.3	51677	397.6	570.7	283.3

$Q_h = 1 - P_h$

ก. วิธีจัดสรรแบบแบ่งเท่ากัน

$$n = 1\sum N_h^2P_hQ_h/(N^2d^2/Z^2 + \sum N_hP_hQ_h)$$
$$= 5(51677)/\{(868)^2(0.05)^2/(1.645)^2+182.3\}$$
$$= 296$$

$$n_h = n/1 = 296/5 \approx 60$$

ข. วิธีจัดสรรแบบสัดส่วน

$$n = N\sum N_hP_hQ_h/(N^2d^2/Z^2+\sum N_hP_hQ_h)$$
$$= 868(182.3)/\{(868)^2(0.05)^2/(1.645)^2+182.3\} \approx 181$$

$n_h = n(N_h/N)$, $n_1 = 181(448/868) \approx 94$

$n_2 = 181(131/868) \cong 28$, $n \simeq 181(81/868) = 17$

$n_4 \; 181(108/868) \simeq 23$, $n_4 \simeq 181(100/868) = 21$

ค. วิธีจัดสรรแบบเนย์แมน

$$n = (\Sigma N_h \sqrt{P_h Q_h})^2 / (N^2 d^2 / Z^2 + \Sigma N_h P_h Q_h)$$

$$= (397.6)^2 / \{(868)^2 (0.05)^2 / (1.645)^2 + 182.3\} \simeq 181$$

$$= \left[N_h \sqrt{P_h Q_h} / (\Sigma N_h \sqrt{P_h Q_h}) \right] n$$

$n_1 = (205.2/397.6)\,181 \simeq 94$, $n_2 = (60.0/397.6)\,181 \simeq 28$

$n_3 = (37.1/397.6)\,181 \simeq 17$, $n_4 = (49.5/397.6)\,181 \simeq 23$

$n_5 = (45.8/397.6)\,181 \simeq 21$

ง. วิธีจัดสรรแบบดีที่สุด

$$n = (\Sigma N_h \sqrt{P_h Q_h} \sqrt{C_h})(\Sigma N_h \sqrt{P_h Q_h} / \sqrt{C_h}) / (N^2 d^2 / Z^2 + \Sigma N_h P_h Q_h)$$

$$= 570.7(283.3) / \{(868)^2 (0.05)^2 / (1.645)^2 + 182.3\} = 185$$

$$n_h = \left[N_h \sqrt{P_h Q_h} / \sqrt{C_h} / (\Sigma N_h \sqrt{P_h Q_h} / \sqrt{C_h}) \right] n$$

$n_1 = (145.1/283.3)185 \simeq 95$, $n_2 = (42.4/283.3)185 \simeq 28$

$n_3 = (21.4/283.3)185 \simeq 14$, $n_4 = (28.6/283.3)185 \simeq 19$

$n_5 = (45.8/283.3)185 \simeq 30$

(4) **ข้อได้เปรียบและเสียเปรียบ** การสุ่มตัวอย่างแบบแบ่งชั้นภูมิมีข้อได้เปรียบและเสียเปรียบดังต่อไปนี้

- มีประสิทธิภาพสูงกว่าแบบธรรมดา
- ได้รายละเอียดแยกเป็นรายชั้นภูมิ
- ควบคุมด้านบริหารสะดวกกว่าแบบธรรมดา
- ต้องจัดเตรียมข่าวสารเกี่ยวกับประชากรล่วงหน้าเพื่อจัดแบ่งเป็นชั้นภูมิ
- ต้องใช้กรอบตัวอย่างสำหรับแต่ละชั้นภูมิ

11.3.3 การสุ่มตัวอย่างแบบระบบ (Systematic Sampling)

การสุ่มตัวอย่างแบบระบบที่มีจุดเริ่มสุ่ม (Random Start) เป็นการเลือกตัวอย่าง โดยการเลือกทุกหน่วยที่ k จากประชากรที่เรียงลำดับไว้ และหน่วยแรกได้รับเลือกแบบสุ่ม สำหรับ k นั้นเรียกว่าช่วงสุ่ม (Sampling interval) และส่วนกลับ 1/k เรียกว่า ส่วนการสุ่ม (Sampling fraction)

(1) วิธีการเลือกตัวอย่าง ต้องมีรายชื่อหน่วยตัวอย่างหรือแผนที่ หรือในกรณีตรวจสอบสินค้าจากที่ผลิตต้องกำหนดเวลาเป็นช่วง ๆ แล้วเราสามารถเลือกตัวอย่างได้ดังนี้

วิธีที่ 1 (ก.) ให้หมายเลขแก่หน่วยตัวอย่างจาก 1 ถึง N (ข.) เลือกสุ่มจำนวนเลขระหว่าง 1 ถึง N สมมติว่าเป็น A แล้ว A จะอยู่ในตัวอย่าง (ค) จากหน่วยตัวอย่างหมายเลข A เราเลือกหน่วยตัวอย่างทุก k นั่นคือ A+k, A+2k, ... สำหรับรายชื่อหน่วยตัวอย่างนั้นจะเป็นรายชื่อแบบวงกลม (Circular list)

ตัวอย่าง ต้องการประมาณจำนวนผู้อยู่อาศัยโดยเฉลี่ยต่อชุมรุมอาคาร 24 แห่ง โดยมีแผนที่ชุมรุมอาคารเหล่านั้น และจะใช้ตัวอย่างชุมรุมอาคาร 6 แห่ง

กรอบตัวอย่างจะเป็นแผนที่ชุมรุมอาคารซึ่งบ่งไว้ชัดแจ้ง หรือเป็นรายชื่อของชุมรุมอาคารในพื้นที่ซึ่งเรียงตามลำดับที่อยู่ สำหรับหน่วยตัวอย่างจะเป็นชุมรุมอาคาร และมีช่วงสุ่มเป็น 24/6 = 4 วิธีการสุ่มตัวอย่างตามวิธีที่ 1 ทำได้ดังนี้ (ก) ให้หมายเลขชุมรุมอาคารจาก 1 ถึง 24 (ข) เลือกหมายเลขระหว่าง 1 และ 24 สมมติว่าได้เป็น 15 แล้วชุมรุมอาคารหมายเลข 15 จะอยู่ในตัวอย่าง จากหมายเลข 15 เลือกตัวอย่างทุก ๆ 4 ชุมรุมอาคาร ซึ่งจะเป็นหมายเลข 19, 23, 23+4 หรือ 3, 23+8 หรือ 7, และ 11 นั่นคือชุมรุมอาคารที่ตกเป็นตัวอย่างจะเป็น 3, 7, 11, 15, 19, 23

วิธีที่ 2 (ก) เลือกหมายแบบสุ่มระหว่าง 1 และ k (ข) หมายเลขที่เลือกได้นั้นจะเป็นหน่วยแรงที่อยู่ในตัวอย่าง แล้วจึงเลือกหน่วยต่อไปทุกหน่วยที่ k จากตัวอย่างนั้นถ้าเลือกหมายเลขระหว่าง 1 และ 4 ได้เป็น 2 แล้วหน่วยต่อไปจะเป็น 1, 10, 14, 18, 22 ดังนั้นตัวอย่างจะประกอบด้วยหน่วยตัวอย่างหมายเลข 2, 6, 10, 14, 18, 22

ถ้าขนาดประชากรเป็น $N = nk$ พอดี แล้วช่วงสุ่มจะเป็น $k = N/n$ พอดี แต่ถ้า $k \neq N/n$ แล้วต้องใช้ k นี้เป็นจำนวนเต็มใกล้กับ N/n สำหรับ $N > nk$ แล้วขนาดตัวอย่างจะเป็น n หรือ $n+1$ และเมื่อ $N \angle nk$ แล้วขนาดตัวอย่างจะเป็น n หรือ $n-1$

วิธีที่ 1 จะดีกว่าวิธีที่ 2 เพราะทุกตัวอย่างในประชากรมีโอกาสถูกเลือกเป็นหน่วยแรกเท่า ๆ กัน

(2) การประมาณค่า ถ้ารายชื่อประชากรเรียงกันแบบสุ่ม นั่นคือคุณลักษณะของหน่วยที่จะวัดไม่มีความสัมพันธ์กับการให้อันดับที่แล้วให้ใช้สูตรการประมาณค่าเช่นเดียวกับการสุ่มแบบธรรมดา

ตัวอย่าง ในการศึกษาจำนวนผู้อยู่อาศัยในชุมรุมอาคารดังตัวอย่างที่แล้วมา สมมติว่าได้ข้อมูลมาดังนี้

ชุมรุมอาคาร	2	6	10	14	18	22
จำนวนผู้อาศัย	7	15	5	8	3	10

ก. ค่าประมาณของค่าเฉลี่ยประชากร

$$\bar{X}_{sy} = \Sigma X_i / n$$
$$= (7+15+5+8+3+10)/6 = 8$$

ซึ่งค่าประมาณของความแปรปรวนเป็น

$$V(\bar{X}_{sy}) = (1-n/N)S^2/n = \{(N-n)/nN\}(n\Sigma X^2 - (\Sigma X)^2)/n(n-1)$$
$$= [(24-6)/6(24)]\{(6(472)-(48)^2\}/6(6-1) = 2.2$$

ในเมื่อ $\Sigma X = 7+15+...+3+10 = 48$

$$\Sigma X^2 = 7^2+15^2+...+3^2+10^2 = 472$$

ข. ค่าประมาณของสัดส่วนประชากร

$$P_{sy} = Q/n$$

ซึ่งมีความแปรปรวนโดยประมาณดังนี้

$$V(P_{sy}) = (1-n/N)\,P(1-P)/(n-1)$$

(3) ข้อได้เปรียบและเสียเปรียบ ในการสุ่มตัวอย่างแบบระบบเรามีข้อได้

เปรียบและเสียเปรียบดังนี้

— การเลือกตัวอย่างทำได้สะดวก รวดเร็ว และถูกกว่าวิธีสุ่มธรรมดา

— บางครั้งไม่ต้องใช้กรอบตัวอย่าง เช่นสุ่มสินค้ามาตรวจทุกชั่วโมง เป็นต้น

— เหมาะกับงานสำรวจที่ต้องการให้หน่วยตัวอย่างกระจายไปทั่วประชากร โดยเฉพาะกระจายตามพื้นที่หรือภูมิศาสตร์

— เหมาะกับการสุ่มตัวอย่างจากประชากรที่เรียงกันเป็นแฟ้มหรือเป็นบัตรรายการ

— ถ้าประชากรไม่เรียงเป็นแบบสุ่ม แล้วจะไม่สามารถหาค่าประมาณของ $V(\bar{X}_{sy})$ จากตัวอย่างเดียวได้

— ถ้าในบัญชีรายการเป็นคาบ (Periodic regularities) แล้วตัวอย่างที่สุ่มได้จะประกอบด้วยพวกที่เหมือนกัน เช่นการสุ่มข้าราชการจากบัญชีที่เรียงตามลำดับอาวุโส อาจสุ่มได้แต่ระดับเดียวกันจากทุก ๆ กอง

11.3.4 การสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)

การสุ่มตัวอย่างแบบแบ่งกลุ่ม เป็นวิธีการเลือกตัวอย่างของกลุ่มหรือชุมรุม (Cluster) ต่าง ๆ ของหน่วยที่เล็กกว่า ซึ่งเรียกว่าสมาชิก (Element) ชุมรุมตัวอย่างอาจจะเลือกโดยวิธีสุ่มแบบธรรมดาหรือแบบระบบก็ได้ ลักษณะของชุมรุมคล้ายกับชั้นภูมิ คือเป็นประชากรย่อยที่ไม่ซ้ำซ้อนกันและรวมกันเป็นประชากรทั้งหมด แต่ต่างกับชั้นภูมิโดยที่ชุมรุมประกอบด้วยหน่วยตัวอย่างที่มีลักษณะต่าง ๆ กัน ส่วนชั้นภูมิจะประกอบด้วยหน่วยตัวอย่างที่มีลักษณะคล้ายคลึงกัน แต่ละชุมรุมจะรวมลักษณะทั้งหลายของประชากรไว้ด้วยกัน

จำนวนสมาชิก (M) ในชุมรุมหนึ่งจะเรียกว่าขนาดชุมรุม จำนวนชุมรุม (N) ในประชากรเรียกว่าขนาดประชากรของชุมรุม ชุมรุมอาจจะมีขนาดเท่ากันหรือไม่ก็ได้

(1) วิธีการเลือกตัวอย่าง กรอบตัวอย่างจะเป็นรายชื่อชุมรุมต่าง ๆ โดยมีหน่วยตัวอย่างเป็นชุมรุม วิธีการเลือกตัวอย่างจะเป็น 2 ขั้นตอนคือ ให้หมายเลขแก่ชุมรุมต่าง ๆ จาก 1 ถึง N แล้วสุ่มหมายเลขจากตารางเลขสุ่มมา n จำนวน (หมายเลขเหล่านั้นจะไม่เกิน N) และหมายเลขที่ได้จะเป็นชุมรุมที่เราต้องการสำรวจทุกสมาชิกในชุมรุมนั้น

ตัวอย่าง ถ้าต้องการส่งใบสอบถามให้พนักงาน 50 คน จากแผนกต่าง ๆ 20 แผนก แต่ละแผนกมีคนงาน 10 คน ถ้าใช้วิธีสุ่มแบบธรรมดาก็จะต้องเตรียมรายชื่อพนักงานทั้ง 200 คน แล้วจึงเลือกสุ่มจากรายชื่อนั้น แต่วิธีสุ่มแบบแบ่งกลุ่มจะไม่ต้องทำรายชื่อพนักงานทั้งหมด เพียงแต่ทำบัญชีรายชื่อแผนกต่าง ๆ เพียง 20 แผนก เมื่อสุ่มมา 5 แผนกจาก 20 แผนก แล้วถามรายละเอียดจากพนักงาน 5 แผนกนั้นก็จะได้ครบ 50 คน โดยไม่ต้องทำบัญชีรายชื่อพนักงานทั้ง 200 คน

(2) **การประมาณค่า** ในการประมาณคุณลักษณะประชากรหรือพารามิเตอร์ที่น่าสนใจ เราทำได้ดังนี้

ก. ค่าเฉลี่ยประชากร ตัวประมาณค่าจะเป็นค่าเฉลี่ยตัวอย่างซึ่งกำหนดไว้ดังนี้

$$\bar{X}_{cl} = \sum_{i}^{n} \sum_{j}^{M} X_{ij} / nM$$

ในเมื่อ X_{ij} เป็นค่าสังเกตสำหรับสมาชิกที่ j ในชุมรุมที่ i

สำหรับความแปรปรวนที่ประมาณได้ของค่าเฉลี่ยตัวอย่างจะเป็น

$$\begin{aligned} V(\bar{X}_{c1}) &= \{(N-n)/nNM^2\}\left(\sum_{i}^{n} (X_i - \bar{X})^2 / (n-1)\right) \\ &= \{(N-n)/nNM^2\}\left(n \sum_{i}^{n} X_i^2 - \left(\sum_{i}^{n} \sum_{j}^{M} X_{ij}\right)^2\right) / n(n-1) \end{aligned}$$

ในเมื่อ X_i เป็นผลรวมของค่าสังเกตในชุมรุม i นั่นคือ $X_i = \sum_{j}^{M} X_{ij}$ และ $\bar{X}$ เป็นค่าเฉลี่ยต่อชุมรุม หรือ $\bar{X} = \sum_{i}^{n} X_i$

ข. สัดส่วนประชากร เมื่อ a_i เป็นจำนวนสมาชิกในชุมรุม i ซึ่งมีลักษณะที่สนใจ แล้วค่าประมาณของสัดส่วนประชากรจะเป็น

$$P = \sum_{i}^{n} a_i / nM$$

โดยมีค่าประมาณของความแปรปรวนเป็น

$$\begin{aligned} V(P) &= (N-n)/Nn \left(\sum_{i}^{n} (P_i - P)^2 / (n-1)\right) \\ &= (N-n)/Nn \left(n \sum_{i}^{n} P_i^2 - \left(\sum_{i}^{n} P_i\right)^2\right) / n(n-1) \end{aligned}$$

ในเมื่อ $P_i = a_i/M$ เป็นสัดส่วนของสมาชิกในชุมรุม i ซึ่งมีลักษณะที่น่าสนใจ

ตัวอย่าง ในการศึกษาค่าใช้จ่ายประเภทเครื่องดื่มต่อวันของพนักงาน และสัดส่วนพนักงานชาย ปรากฏว่าได้ข้อมูลดังนี้

กลุ่ม (i)		1	2	3	4	5
หน่วย (j)	1	ช 5.50	ช 2.50	ญ 0.90	ช 3.20	ญ 4.80
	2	ช 1.50	ญ 1.70	ช 1.00	ช 1.20	ช 3.00
	3	ญ 0.50	ญ 2.00	ช 3.60	ช 1.00	ญ 4.70
	4	ช 4.10	ช 1.60	ช 4.20	ญ 2.00	ญ 3.20
	5	ช 2.80	ญ 0.90	ญ 2.80	ญ 1.30	ช 1.00
	6	ญ 1.75	ช 0.40	ช 1.70	ช 0.50	ช 1.80
	7	ญ 4.00	ช 1.30	ญ 1.60	ญ 4.00	ญ 2.00
	8	ญ 3.40	ช 4.50	ญ 1.00	ช 3.90	ช 1.00
	9	ช 0.60	ญ 1.30	ช 4.00	ญ 2.15	ญ 2.50
	10	ญ 4.75	ญ 3.40	ช 3.30	ช 2.20	ช 3.60

ΣX_i	28.90	19.60	24.10	21.45	27.60	121.65
ΣX_i^2	35.21	384.16	580.81	460.1025	761.76	3022.0425
a_i	5	5	6	6	5	27
P_i	0.50	0.50	0.60	0.60	0.50	2.10
P_i^2	0.25	0.25	0.36	0.36	0.25	1.47

ค่าประมาณของค่าเฉลี่ยประชากร

$$\bar{\bar{X}}_{cl} = 121.65/5(10) = 2.433$$

$$V(\bar{\bar{X}}_{cl}) = \frac{20-5}{5(20)(10)^2}\left\{\frac{5(3022.0425)-(121.65)^2}{5(5-1)}\right\} = 0.023$$

ค่าประมาณของสัดส่วนประชากร

$$P = 27/5(10) = 0.54$$

$$V(P) = \frac{20-5}{5(20)}\left\{\frac{5(1.47)-(2.10)^2}{5(5-1)}\right\} = 0.029$$

(3) **ข้อได้เปรียบและเสียเปรียบ** ในการสุ่มตัวอย่างแบบแบ่งกลุ่มนี้มีข้อได้เปรียบและข้อเสียเปรียบดังนี้

– ไม่จำเป็นต้องทำบัญชีรายชื่อสมาชิกในประชากรเหมือนการสุ่มแบบธรรมดาและแบบชั้นภูมิ กรอบตัวอย่างของการสุ่มแบบแบ่งกลุ่มจะเป็นบัญชีรายชื่อชุมรุม

– แม้จะมีบัญชีรายชื่ออยู่พร้อมแล้วโดยไม่ต้องเตรียมใหม่ การสุ่มตัวอย่างแบบแบ่งกลุ่มก็ยังประหยัดกว่าเมื่อปฏิบัติงานสนาม กล่าวคือหน่วยที่ตกเป็นตัวอย่างจะอยู่ใกล้เคียงกันมากกว่าวิธีอื่น

– มีประสิทธิภาพด้อยกว่าการสุ่มแบบธรรมดาและแบบชั้นภูมิ เนื่องจากตัวอย่างสุ่มเช่นครัวเรือนที่อยู่ใกล้เคียงกันมักมีคุณลักษณะคล้าย ๆ กันมากกว่าครัวเรือนที่อยู่ห่างไกล จึงเป็นการกระทบกระเทือน "ความเป็นตัวแทน" ของตัวอย่าง และผลอันนี้จะทำให้ความคลาดเคลื่อนเพิ่มขึ้น

อย่างไรก็ตามการสุ่มตัวอย่างแบบแบ่งกลุ่มเหมาะสมกับงานสำรวจที่ไม่สามารถหาบัญชีรายชื่อหน่วยตัวอย่างที่สมบูรณ์ได้ หรือสำหรับงานที่งบประมาณไม่มากจึงไม่สามารถเน้นหนักด้านความถูกต้องแม่นยำ

11.3.5 การสุ่มตัวอย่างแบบหลายขั้นตอน (Multistage Sampling)

ในการสุ่มตัวอย่างแบบหลายขั้นตอนนั้นการเลือกตัวอย่างจะต้องทำตั้งแต่สองขั้นตอนขึ้นไป โดยในขั้นแรกจะต้องแบ่งประชากรออกเป็นหน่วยขั้นแรก (First–Stage unit or Primary Sampling Unit, PSU) แล้วจึงสุ่มตัวแทนมาจำนวนหนึ่ง ขั้นต่อไปก็แบ่งตัวแทนที่สุ่มมาได้นั้นออกเป็นกลุ่มย่อยๆ อีก แล้วสุ่มมาอีกจำนวนหนึ่งจะทำขั้นต่อไปอีกก็ทำได้ ตัวอย่างเช่น ในการสำรวจครัวเรือนของเมืองหนึ่งซึ่งมี $N = 10$ ชุมรุมอาคาร แต่ละชุมรุมอาคารประกอบด้วย $M = 10$ ครัวเรือน (แต่ละครัวเรือนไม่จำเป็นต้องมีสมาชิกจำนวนเท่ากัน) ถ้าต้องการตัวอย่างเพียง 30 ครัวเรือน โดยสุ่มชุมรุมตัวอย่างมา $n = 5$ ชุมรุม แต่ละชุมรุมสุ่มครัวเรือนตัวอย่างมาชุมรุมละ $m = 6$ ครัวเรือน จะเห็นได้ว่าต้องใช้การสุ่มแบบสองขั้นตอน

(1) **วิธีการเรียกตัวอย่าง** จากตัวอย่างของการสำรวจครัวเรือนโดยอาศัยตัว

อย่าง 30 ครัวเรือนนั้นเรามีกรอบตัวอย่างและหน่วยตัวอย่างดังนี้

กรอบตัวอย่าง ขั้น 1 รายชื่อชุมรุมอาคารทั้งหมด

ขั้น 2 รายชื่อครัวเรือนทั้งหมดในชุมรุมที่ถูกเลือก

หน่วยตัวอย่าง ขั้น 1 ชุมรุมอาคาร (PSU)

ขั้น 2 ครัวเรือนซึ่งเป็นหน่วยขั้นสอง (Secondary Sampling Unit, SSU)

ในการเลือกตัวอย่างจะทำได้ดังนี้

– ให้หมายเลขแก่ชุมรุมอาคารในกรอบตัวอย่างคือ 1 ถึง N = 10

– ใช้ตารางเลขสุ่มทำการสุ่มตัวเลข 1 หลักมา 5 จำนวน สมมติว่าได้ 3, 8, 1, 4, 5 ดังนั้นชุมรุมที่มีหมายเลขดังกล่าวจะตกเป็นตัวอย่าง

– ให้หมายเลขแก่ครัวเรือนจาก 1 ถึง M = 10 จากกรอบตัวอย่างของแต่ละชุมรุมอาคารที่เลือกได้ขั้นแรก

– ใช้ตารางเลขสุ่ม เลือกหมายเลขมา 5 ชุด ๆ ละ 6 จำนวน โดยมีแต่ละจำนวนจะน้อยกว่าหรือเท่ากับ M = 10 สมมติว่าได้เป็น

ชุด 1	6	2	7	5	3	8
2	1	5	6	7	2	4
3	8	2	5	3	1	6
4	1	4	3	8	2	9
5	10	9	7	3	4	2

– เลือกครัวเรือนหมายเลข 6, 2, 7, 5, 3, 8, ในชุมรุมอาคาร 3 ครัวเรือนหมายเลข 1, 5, 6, 7, 2, 4 ในชุมรุมอาคาร 8 และต่อ ๆ ไปเป็นตัวแทนที่จะศึกษาจำนวนคนที่อยู่ในครัวเรือนนั้น ๆ

(2) **การประมาณค่า** ในการสุ่มตัวอย่างแบบหลายขั้นตอนนั้น เรามีการประมาณค่าของพารามีเตอร์ประชากรดังนี้

ก. ค่าเฉลี่ยประชากร ค่าเฉลี่ยตัวอย่างซึ่งเป็นค่าประมาณของค่าเฉลี่ยประชากรจะเป็น

$$\bar{X}_{ms} = \sum_{i}^{n} \sum_{j}^{m} X_{ij}/nm$$

ในเมื่อ X_{ij} เป็นค่าสังเกตุสำหรับสมาชิกขั้นที่สอง j ในหน่วยขั้นแรก i

สำหรับค่าประมาณของความแปรปรวนของค่าเฉลี่ยตัวอย่างจะเป็น

$$V(\bar{X}_{ms}) = (N-n)/nN \sum_{i}^{n} (\bar{X}_i-\bar{X}_{ms})^2/(n-1)+1/n(M-m)/mM \sum_{i}^{n} \sum_{j}^{m}(X_{ij}-\bar{X}_i)^2/n(m-1)$$

$$=(N-n)/nN\left\{(n \sum_{i}^{n}X^2_{i} - (\sum_{i}^{n} \sum_{j}^{m} X_{ij})^2 /nm^2(n-1)\right\}$$

$$+1/N(M-m)/mM\left\{n \sum_{i}^{n} \sum_{j}^{m} X_{ij}^2 - \sum_{i}^{n} X^2_{i/nm(m-1)}\right\}$$

ในเมื่อ $X_i = \sum_j X_{ij}$ เป็นผลรวมของค่าสังเกตุในหน่วย และ $\bar{X}_i = X_i/m$ เป็นค่าเฉลี่ยต่อหน่วย i

ข. สัดส่วนประชากร ในเมื่อ a_i เป็นจำนวนสมาชิกซึ่งมีลักษณะที่สนใจในหน่วยแรก i แล้วค่าประมาณของสัดส่วนประชากรจะเป็น

$$\bar{P} = \sum_{i}^{n}a_i/mn$$

โดยมีค่าประมาณของความแปรปรวนดังนี้

$$V(\bar{P}) = (N-n)/n(n-1)N \sum_{i}^{n}(P_i-\bar{P})^2+1/nN\ (M-m)/(m-1)M\sum_{i}^{n} P_i(1-P_i)$$

$$=(N-n)/nN\ (n\sum P_i^2-(\sum P_i)^2)/n(n-1)+1/nN\ (M-m)/(m-1)M \sum_{i}^{n}P_i(1-P_i)$$

ในเมื่อ $P_i = a_i/m$ เป็นสัดส่วนของสมาชิกซึ่งมีลักษณะที่สนใจในชุมรุม i

ตัวอย่าง ในการสำรวจค่าเครื่องดื่มต่อวันของครัวเรือนได้ข้อมูลมาดังนี้

ชุมรุมอาคาร (i)		1	2	3	4	5
ครัวเรือน (j)	1	15	5	8	7	11
	2	5	4	10	16	5
	3	20	20	12	14	15
	4	18	15	10	5	20
	5	6	6	5	10	10

6	2	10	5	10	5	
X_i	66	60	50	62	66	304
X_i^2	4356	3600	2500	3844	4356	18656
a_i	3	3	5	4	3	18
P_i	0.50	0.50	0.83	0.67	0.50	3
P_i^2	0.25	0.25	0.69	0.44	0.25	1.88
$P_i(1-P_i)$	0.25	0.25	0.14	0.22	0.25	1.11

ค. จำนวนครัวเรือนในหน่วย i ที่มีรายจ่ายค่าเครื่องดื่มน้อยกว่าหรือเท่ากับ 10

$$\sum \sum X_{ij}^2 = 15^2+5^2+...+10^2+5^2 = 3896$$

ค่าประมาณของค่าเฉลี่ยประชากร

$$\overline{X}_{ms} = 304/5(6) = 10.133$$

$$V(\overline{X}_{ms}) = (10-5)/5(10)\{(5(18656)-(304)^2/5(6^2)(5-1)\}+1/10(10-6)/6(10)$$

$$\{(6(3896)-18656)/5(6)(6-1)\}$$

$$= 0.120+0.021 = 0.141$$

ค่าประมาณของสัดส่วนประชากร

$$\overline{P} = 18/5(6) = 0.60$$

$$V(\overline{P}) = (10-5)/5(10)\{(5(1.88)-3^2)/5(5-1)\}+1/5(10)$$

$$\{(10-6)/(6-1)10(1.11)\}$$

$$= 0.033+0.002 = 0.035$$

(3) ข้อได้เปรียบและเสียเปรียบ การสุ่มตัวอย่างแบบหลายขั้นตอนมีข้อได้เปรียบและเสียเปรียบดังนี้

- มีประสิทธิภาพและยืดหยุ่นมากกว่าการสุ่มแบบขั้นตอนเดียว
- การจัดทำบัญชีรายชื่อไม่ต้องทำทั้งหมดนอกจากขั้นแรก
- ประหยัดค่าใช้จ่ายในการเดินทางไปเก็บข้อมูลโดยเฉพาะเมื่อหน่วยตัวอย่างในขั้นแรกถูกจำแนกตามภูมิศาสตร์
- วิธีการสลับซับซ้อนยากแก่ความเข้าใจ โดยเฉพาะวิธีการประมาณค่าซึ่งเป็น

เรื่องที่ทำความลำบากใจแก่ผู้ที่ไม่ใช่นักสถิติ

อย่างไรก็ตามควรใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอนสำหรับประชากรที่อยู่กระจัดกระจายทางภูมิศาสตร์อย่างกว้างขวาง และไม่มีรายชื่อสำหรับเป็นกรอบตัวอย่างให้เลือกสุ่มตัวอย่าง.