# บทที่ 4
# แผนสำรวจแบบ Systematic

## 4.1 เหตุผลและความจำเป็น

ในกรณีที่กลุ่มประชากรมีขนาดใหญ่มาก ๆ ปัญหาเท่าที่ปรากฏในทางปฏิบัติก็คือความยุ่งยากในการสร้างกรอบตัวอย่างซึ่งจะต้องกำหนดระหัสประจำตัวให้แก่หน่วยสำรวจพร้อมทั้งตำบลที่อยู่ที่ทันสมัยสามารถติดต่อหรือเข้าถึงได้ ปัญหานี้นับว่าเป็นปัญหาที่สำคัญที่สุดและสร้างความยากลำบากให้แก่นักวิจัยเป็นอย่างยิ่งเพราะทำให้สิ้นเปลืองงบประมาณและเสียเวลาไปเป็นจำนวนมิใช่น้อยอีกประการหนึ่งแม้ว่าจะสามารถปรับปรุงหรือสร้างกรอบตัวอย่างที่ถูกต้องเหมาะสมและทันสมัยได้ก็มิได้ประกันว่ากลุ่มตัวอย่างที่เลือกได้ไม่ว่าจะเลือกโดยอาศัยตารางเลขสุ่ม Random Number Generator หรือวิธีการอื่น ๆ ที่ทดแทนกันได้นั้นจะเป็นตัวแทนที่ดีของกลุ่มประชากรเพราะถ้ากลุ่มประชากรมีขนาดใหญ่เสียแล้วเหตุการณ์ที่เกิดขึ้นอยู่เป็นปกติก็คือกลุ่มตัวอย่างกระจายไปไม่ทั่วกลุ่มประชากร หรือนัยหนึ่ง กลุ่มตัวอย่างมิได้เป็นตัวแทนที่มาจากทุกส่วนทั่วกลุ่มประชากร อาจมาจากเพียงบางส่วนของประชากรเท่านั้น ตัวอย่างเช่นการเลือกกลุ่มตัวอย่างครัวเรือนในเขตกทม. มาเป็นตัวอย่าง 500 ครัวเรือน ครัวเรือนตัวอย่างเหล่านี้อาจได้รับเลือกมาจากเฉพาะบางเขต เช่น ส่วนใหญ่มาจากเขตมีนบุรี เขตพญาไท เขตบางกอกน้อย ที่เหลือนอกนั้นอาจมาจากเขตอื่น ๆ เขตละเพียงเล็กน้อย บางเขตอาจไม่ได้รับเลือกเลย และที่สำคัญก็คือเราเองมีกำลังพอที่จะรวบรวมรายชื่อหัวหน้าครัวเรือน พร้อมทั้งกำหนดหมายเลขระหัส 1 ถึง N รวมทั้งตำบลที่อยู่ของครัวเรือนเหล่านี้ในเขตกทม. ได้ครบถ้วนหรือ? หรืออีกตัวอย่างหนึ่งคือสมมุติว่าท่านต้องการทราบว่าห้างสรรพสินค้าที่ท่านทำงานอยู่นั้นจัดบริการให้แก่ลูกค้า

ได้ดีหรือไม่เพียงใดโดยมุ่งสอบถามความคิดเห็นและข้อแนะนำจากลูกค้า ถ้าท่านจะใช้วิธี SRS ท่านจะทำได้หรือไม่? ถ้าจะใช้วิธี SRS สิ่งแรกที่ต้องทำก็คือการสร้างกรอบตัวอย่างของลูกค้าที่เข้าซื้อหาสินค้าและบริการจากห้าง กรณีเช่นนี้ท่านจะทราบได้อย่างไรว่าใครบ้างเคยเข้าซื้อหาและรับบริการจากห้าง? แม้แต่ท่านจะลดขนาดประชากรลงโดยมุ่งหมายเอาเฉพาะลูกค้าที่กำลังเดินเลือกสินค้าอยู่ในห้างขณะนั้นท่านก็คงจะสร้างกรอบตัวอย่างไม่ทันเพราะลูกค้าจะเข้า-ออกตลอดเวลา

จากปัญหาทั้ง 3 ประการดังกล่าวข้างต้นคือ ปัญหาการสร้างกรอบตัวอย่าง ปัญหาที่ตัวอย่างไม่กระจายไปทั่วกลุ่มประชากร และปัญหาที่หน่วยสำรวจมีการเคลื่อนไหวหรือเคลื่อนที่สับเปลี่ยนตลอดเวลาเหล่านี้ล้วนนำไปสู่การพัฒนาเทคนิคการสำรวจแบบอื่น เทคนิคที่จะกล่าวถึงก็คือ Systematic Sampling Plan ซึ่งเหมาะสมกับกลุ่มประชากรขนาดใหญ่หรือกลุ่มประชากรขนาดไม่ใหญ่นักแต่หน่วยสำรวจมีการเคลื่อนไหวอยู่ตลอดเวลา

## 4.2 ความหมายและวิธีการของ Systematic Sampling Plan

แผนสำรวจแบบ Systematic คือแผนสำรวจที่เลือกตัวอย่างมีระบบ โดยจะทำการเลือกหน่วยตัวอย่างหน่วยแรกขึ้นมาโดยสุ่ม (Random Start) และเลือกหน่วยตัวอย่างต่อ ๆ ไปจากทุก ๆ $k$ หน่วย หรือเลือกมา 1 หน่วยเว้นไป $k-1$ หน่วย ดังนี้เรื่อย ๆ ไปจนครบขนาดตัวอย่างที่ต้องการ[1] เช่นการสำรวจปัญหาทางสังคมวิทยาในหมู่บ้านจัดสรร นักวิจัยจะเลือกบ้านเริ่มต้นมาเป็น Random Start แล้วเลือกบ้านต่อไป 1 บ้านทุก ๆ 10 บ้าน หรือเลือก 1 บ้านเว้น 9 บ้าน หรือการศึกษาภาวะเศรษฐกิจของทหารผ่านศึกในนิคมทหารผ่านศึกชานุมาน อำเภอชานุมาน จังหวัดอุบลราชธานี การสำรวจดำเนินการโดยเลือกบ้านเริ่มต้นมาเป็น Random Start แล้วเลือกบ้านต่อไป 1 บ้านเว้น 2 บ้านจนครบขนาดตัวอย่าง 133 ครัวเรือน หรือการศึกษาภูมิหลังของนักศึกษาที่สำเร็จการศึกษาไปแล้ว การวิจัยจะดำเนินการเลือกตัวอย่างโดยใช้ทะเบียนประวัติที่บันทึกลงบัตรโดยเลือกมาเป็นตัวอย่าง

---

[1] มีความหมายว่าหน่วยเริ่มต้นเป็นตัวกำหนดตัวอย่างทั้งชุด

1 บัตร ทุก ๆ บัตรที่เรียงต่อกัน 1 นิ้ว หมายความว่า เลือกบัตรเริ่มต้นมา 1 บัตรแล้วเลือกบัตรต่อ ๆ ไปโดยใช้ไม้บันทัดทาบไปบนกลุ่มบัตรที่เรียงติดต่อกันโดยดึงมา 1 บัตรทุก ๆ ระยะความยาว 1 นิ้ว ดังนี้เป็นต้น

อนึ่งการเลือกหน่วยเริ่มต้น (Random Start) นั้นจำเป็นต้องเลือกจากหน่วยใดหน่วยหนึ่งโดยสุ่มจากชั้นภูมิแรก หน่วยตัวอย่างต่อ ๆ ไปจะปรากฏอยู่ในชั้นภูมิถัดไป โดยจะเลือกตัวอย่าง 1 ตัวอย่างจากทุก ๆ ชั้นภูมิจนครบขนาดตัวอย่างที่ต้องการ โดยนัยนี้ขนาดตัวอย่าง n ก็คือจำนวนชั้นภูมินั่นเอง การกำหนดขนาดตัวอย่างให้ยึดถือวิธีการคำนวณตามวิธี SRS โดยอนุโลม

หรือจะคำนวณขึ้นเองจากความสัมพันธ์ $d_0^2 = \hat{V}(\hat{\theta})Z^2_{1-\alpha/2}$ หรือจะกำหนดขนาดตัวอย่างโดยอาศัยตารางสำหรับกำหนดขนาดตัวอย่างในภาคผนวกท้ายเล่มก็ได้ ส่วนค่าของ k ซึ่งแสดงลักษณะการเว้นระยะ (Sampling Interval) ระหว่างหน่วยสำรวจอย่างมีระบบนั้นเราสามารถคำนวณหาได้โดยง่ายโดยอาศัยความสัมพันธ์ $k = N/n$ เมื่อ $N$ = ขนาดประชากร และ $n$ = ขนาดตัวอย่างตามที่กำหนดขึ้น

โดยทั่วไปเราควรกำหนดค่า n ก่อนแล้วค่อยหาค่า k จากความสัมพันธ์ $k = \frac{N}{n}$ แต่เราจะกำหนดค่า k ขึ้นก่อนแล้วหาค่า n จากความสัมพันธ์เดิมก็ได้เพราะมีผลเสมอกัน เช่นในกรณีตัวอย่าง เมื่อ $N = 120$ ถ้าต้องการสุ่มตัวอย่าง 1 หน่วยทุก ๆ 5 หน่วยคือ $k = 5$ จะพบว่า $n = \frac{N}{k} = \frac{120}{5} = 24$ ขณะเดียวกันถ้ากำหนดว่าจะสุ่มตัวอย่างมา 24 หน่วยคือ $n = 24$ จะพบว่า $k = \frac{N}{n} = \frac{120}{24} = 5$ หรือกำหนดว่าจะสุ่มตัวอย่างมา 20% แสดงว่า $n = (.20) \times 120 = 24$ ขณะเดียวกันจะพบว่า $20\% = \frac{20}{100} = \frac{1}{5}$ หรือ 1 ต่อ 5 ซึ่งแสดงว่าสุ่มตัวอย่างมา 1 หน่วยทุก ๆ 5 หน่วย

ปัญหาที่พบก็คือกรณีการหารไม่ลงตัวคือ k หาร N ไม่ลงตัว หรือ n หาร N ไม่ลงตัว ซึ่งอาจสร้างความสงสัยให้แก่นักวิจัยขึ้นได้ว่าในกรณีนี้ควรจะกำหนด k หรือว่ากำหนด n ขึ้นมาก่อนกันแน่ สำหรับปัญหานี้ขอชี้แจงว่า เรายังคงใช้หลักเกณฑ์เดิมคือจะกำหนดใครก่อนกันก็ได้ ผลประการหนึ่งที่ติดตามมาคือสำหรับกรณี $nk \neq N$ ก็คือขนาดตัวอย่างที่ได้มีค่าไม่แน่นอน เช่น $N = 120$ และกำหนด $k = 7$ จะพบว่าขนาดตัวอย่างคือ $n = \frac{120}{7}$

$= 17\frac{1}{7}$ แสดงว่าขนาดตัวอย่างอาจมีค่าเท่ากับ 17 หรือ 18 ก็ได้ ขณะเดียวกันถ้ากำหนดให้ $n = 50$ จะพบว่า $k = \frac{120}{50} = 2.4$ ซึ่งแสดงว่า k อาจมีค่าเป็น 2 หรือ 3 ก็ได้ การแก้ปัญหาทั้งสองประการนี้มิใช่เรื่องยากเพราะเราสามารถใช้ Circular Systematic Sampling แก้ปัญหาของกรณีที่ k มีค่าไม่แน่นอนได้ หรือจะใช้ Fractional Sampling Interval แก้ปัญหาของกรณีที่ k เป็นเลขทศนิยมก็ได้ ส่วนกรณีที่ n มีค่าไม่แน่นอนในทางทฤษฎีเราจะถือว่า ถ้าขนาดตัวอย่าง n ที่คำนวณได้มีค่ามากกว่า 50 คือ $n > 50$ เราจะไม่สนใจปัญหาทั้งมวลไปโดยอัตโนมัติ หรือแม้แต่จะมีขนาดเล็กกว่า 50 ก็ตาม

อย่างไรก็ตาม ในกรณีที่เรากำหนดค่า n ขึ้นก่อนเพื่อกำหนด k ในภายหลังนั้นในทางปฏิบัติ ถ้าพบว่า n หาร N ไม่ลงตัวให้ปัดเศษ คือถ้าเศษเกิน 0.5 ให้ปัดเป็น 1 เช่น $k = 4.6$ ให้ถือว่า $k = 5$ ถ้า $k = 4.5$ ให้ถือว่า $k = 5$ ถ้า $k = 4.1$ ให้ถือว่า $k = 4$ นั่นคือให้ใช้

$k = [\frac{N}{n}]$ ถ้า $r \leqslant n/2$ และใช้ $k = [\frac{N}{n}] + 1$ ถ้า $r > n/2$ เมื่อ r คือเศษเหลือ สัญลักษณ์ $[\frac{N}{n}]$ หมายความว่าคิดเฉพาะจำนวนเต็ม (Integer) ของการหาร

อนึ่งเมื่อพิจารณาค่า k จะพบว่านอกจาก k จะแสดงลักษณะการเว้นระยะระหว่างหน่วยสำรวจอย่างมีระบบแล้ว k ยังแสดงขนาดของชั้นภูมิและจำนวนกลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมด (All Possible Sample) อีกด้วย

พิจารณาตัวอย่างต่อไปนี้

สมมุติว่ากลุ่มประชากรประกอบไปด้วยหน่วยสำรวจ 12 หน่วย $(N = 12)$ คือ

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, \ldots, X_{12}$$

และเราต้องการเลือกหน่วยสำรวจมาเป็นตัวอย่าง 4 หน่วย $(n = 4)$ ดังนั้น โดยคำอธิบายข้างต้นจะพบว่าเราต้องการจัดจำแนกประชากรออกเป็น 4 ชั้นภูมิ (n = ขนาดตัวอย่าง = จำนวนชั้นภูมิ) ชั้นภูมิหนึ่ง ๆ มีขนาดเท่ากับ $k = N/n = 12/4 = 3$ หรือมีความหมายในทางปฏิบัติว่าเราจะดำเนินการเลือกหน่วยตัวอย่างมาทุก ๆ 3 $(k = 3)$ หน่วยจากหน่วยเริ่มต้น หรือทำการสำรวจ 1 หน่วยเว้น 2 หน่วยนับจากหน่วยเริ่มต้น ลักษณะการแบ่งชั้นภูมิจึงปรากฏดังนี้ เส้นดิ่ง " | " คือเครื่องหมายแสดงการแบ่งประชากรออกเป็นชั้นภูมิ

$x_1 , x_2 , x_3$ | $x_4 , x_5 , x_6$ | $x_7 , x_8 , x_9$ | $x_{10} , x_{11} , x_{12}$

| ชั้นภูมิที่ 1 | ชั้นภูมิที่ 2 | ชั้นภูมิที่ 3 | ชั้นภูมิที่ 4 |
|---|---|---|---|
| หรือ | หรือ | หรือ | หรือ |
| Cluster ที่ 1 | Cluster ที่ 2 | Cluster ที่ 3 | Cluster ที่ 4 |

การเลือกตัวอย่างให้สุ่มเลือกหน่วยเริ่มต้นมาจากชั้นภูมิที่ 1 แล้วเลือกหน่วยต่อไปทุก ๆ $k = 3$ หน่วยหรือสุ่ม 1 หน่วยเว้น $k-1 = 3-1 = 2$ หน่วยเรื่อยไปจนครบขนาดตัวอย่าง $n=4$ หรือสุ่มไปเรื่อย ๆ จนครบทุกชั้นภูมิ กลุ่มตัวอย่างที่เป็นไปได้ทั้งหมดจะมีได้ทั้งสิ้น $k = 3$ ชุด ตามจำนวนหน่วยเริ่มต้นในชั้นภูมิที่ 1 ซึ่งเป็นไปได้ทั้งสิ้น $k = 3$ หน่วย ดังนี้

ถ้าหน่วยเริ่มต้นคือ $x_1$ กลุ่มตัวอย่างที่ได้คือ $(x_1,x_{1+3},x_{1+6},x_{1+9})$ ซึ่งก็คือ $(x_1,x_4,x_7,x_{10})$

ถ้าหน่วยเริ่มต้นคือ $x_2$ กลุ่มตัวอย่างที่ได้คือ $(x_2,x_{2+3},x_{2+6},x_{2+9})$ ซึ่งก็คือ $(x_2,x_5,x_8,x_{11})$

ถ้าหน่วยเริ่มต้นคือ $x_3$ กลุ่มตัวอย่างที่ได้คือ $(x_3,x_{3+3},x_{3+6},x_{3+9})$ ซึ่งก็คือ $(x_3,x_6,x_9,x_{12})$

หรือสรุปกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมดดังนี้

ตัวอย่างที่ 1 $x_1,x_4,x_7,x_{10}$

ตัวอย่างที่ 2 $x_2,x_5,x_8,x_{11}$

ตัวอย่างที่ 3 $x_3,x_6,x_9,x_{12}$

**ข้อสังเกต**

(1) จำนวนตัวอย่างที่พึงเป็นไปได้ทั้งหมดคือ $k = 3$ ตัวอย่าง = จำนวนระยะเว้นห่างระหว่างหน่วย = ขนาดชั้นภูมิ แต่ละตัวอย่างมีขนาด $= n = 4 =$ จำนวนชั้นภูมิ

(2) การเลือกตัวอย่างมีลักษณะเช่นเดียวกับแผนสำรวจแบบชั้นภูมิที่ชั้นภูมิหนึ่ง ๆ สุ่มตัวอย่างมา 1 หน่วย ต่างกันเล็กน้อยที่ถ้าเป็นแผนสำรวจแบบแบ่งชั้นภูมินั้นที่ตั้งหรือตำแหน่งของหน่วยตัวอย่างจะปรากฏ ณ. ตำแหน่งต่าง ๆ ในชั้นภูมิโดยสุ่ม อาจอยู่ต้นชั้นภูมิกลางชั้นภูมิ หรือปลายชั้นภูมิ แต่แผนสำรวจแบบ Systematic นี้ตำแหน่งของหน่วยตัวอย่างจะคงที่อยู่ ณ. ตำแหน่งเดียวกันในทุกชั้นภูมิ ทั้งนี้ขึ้นอยู่กับตำแหน่งของหน่วยเริ่มต้นในชั้นภูมิแรก ซึ่งในทางปฏิบัติแล้วแผนสำรวจแบบ Systematic จะปฏิบัติได้ง่ายกว่าเพราะเรามีความ

จำเป็นต้อง Randomize เพียงครั้งเดียวในชั้นภูมิแรกเพื่อหาหน่วยที่เริ่มต้น หน่วยต่อไปไม่จำเป็นต้อง Randomize เพราะเราจะสุ่ม 1 หน่วย เว้นระยะ $k-1$ หน่วยเรื่อย ๆ ไปจนครบทุกชั้นภูมิ ซึ่งเมื่อเป็นเช่นนี้ใคร ๆ ก็สุ่มตัวอย่างได้ แม้แต่กับบุคคลที่ไม่ทราบหรือไม่มีความรู้ทางทฤษฎีการสำรวจเลย และที่สำคัญโดยเฉพาะในทางปฏิบัติก็คือวิธีนี้ไม่จำเป็นต้องให้ความสนใจต่อการสร้างหรือปรับปรุงกรอบตัวอย่างมากนัก เพียงแต่สำรวจแผนผังที่ตั้งของหน่วยสำรวจเพียงคร่าว ๆ ครั้งหนึ่งก็สามารถปฏิบัติสนามได้ทันที ไม่มีความจำเป็นต้องออกเลขระหัสให้แก่หน่วยสำรวจ จึงมีประโยชน์มากในการปฏิบัติงานสนาม แต่ถ้านักวิจัยต้องการปฏิบัติงานอย่างเข้มงวดกวดขันก็ควรจะสร้างกรอบตัวอย่างที่สมบูรณ์แล้วกำหนดหน่วยสำรวจไว้ให้เสร็จสิ้นไปจากสำนักงานเลย ผู้เขียนเคยปฏิบัติทั้งสองวิธีคือการสร้างกรอบตัวอย่างที่สมบูรณ์แล้วเลือกตัวอย่างให้เสร็จสิ้นไปจากสำนักงานก่อนปฏิบัติงานสนามจริงวิธีหนึ่ง กับการสำรวจที่ตั้งหน่วยสำรวจอย่างคร่าว ๆ และวาดแผนผังไว้คร่าว ๆ ไว้โดยให้ความสนใจเฉพาะหน่วยสำรวจที่กระจายห่างออกไปจากกลุ่มอีกวิธีหนึ่ง พบว่าวิธีแรกสิ้นเปลืองแรงงานและเวลามากกว่า แต่มีประสิทธิภาพสูงและสะดวกรวดเร็วกว่ามากในขณะปฏิบัติงานสนาม ขณะที่วิธีที่สองสิ้นเปลืองเวลาน้อยกว่าแต่พนักงานสัมภาษณ์มักเข้าสัมภาษณ์หน่วยตัวอย่างผิดลำดับ แต่ถ้าหากหน่วยสำรวจกระจายตัวกันอย่างเป็นระเบียบเช่น บ้านจัดสรร นิคมเกษตรกรรม หมู่บ้านในชนบทบางแห่ง วิธีที่สองจะมีประสิทธิภาพสูงใกล้เคียงกับวิธีแรก แต่ถ้าหน่วยสำรวจกระจายตัวอย่างไร้ระเบียบแบบแผน เช่นย่านเสื่อมโทรมดินแดง หรือคลองเตย วิธีที่สองจะประสบความล้มเหลวอย่างที่สุด สำหรับแผนสำรวจแบบแบ่งชั้นภูมินักวิจัยไม่มีหนทางเลือกเป็นอย่างอื่น จำเป็นต้องสร้างกรอบตัวอย่างที่สมบูรณ์ของทุกชั้นภูมิและต้อง Randomize หน่วยสำรวจทุกชั้นภูมิ

(3) หน่วยสำรวจหนึ่ง ๆ สามารถปรากฏในกลุ่มตัวอย่างได้เพียง 1 ตัวอย่าง และถ้านำหน่วยสำรวจจากทุก ๆ ตัวอย่างมารวมกัน ผลลัพธ์ก็คือกลุ่มประชากร ตามตัวอย่างข้างต้น จะพบว่ากลุ่มตัวอย่างทั้งสามคือ $(x_1,x_4,x_7,x_{10})$, $(x_2,x_5,x_8,x_{11})$ และ $(x_3,x_6,x_9,x_{12})$ ถ้ากระจายทุกตัวอย่างออกก็จะกลายเป็นกลุ่มประชากรคือ $x_1,x_2,\ldots,x_{12}$ ข้อสังเกตนี้มีความสำคัญมาก ขอให้นักศึกษาให้ความสนใจไว้มาก ๆ เพราะจะเป็นเครื่องช่วยชี้ให้เห็นเหตุผลของความล้มเหลวประการที่สำคัญของแผนสำรวจแบบ Systematic ซึ่งจะได้กล่าวถึงต่อไป

(4) ความน่าจะเป็นที่จะได้ตัวอย่างใด ๆ หรือความน่าจะเป็นที่กลุ่มตัวอย่างหนึ่งกลุ่มตัวอย่างใดจะได้รับเลือก $= \frac{1}{3} = \frac{1}{k}$ = 1/(จำนวนกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมด)

## 4.3 การเลือกตัวอย่าง

การเลือกตัวอย่างซึ่งจะกล่าวถึงในลำดับต่อไปนี้จะยึดถือแนวคิดขั้นต้นว่าการเลือกตัวอย่างนั้นนักวิจัยจะต้องกำหนดค่า k ขึ้นมาก่อน (อ่านคำอธิบายท้ายตอน 4.2) และขอให้ทำความเข้าใจร่วมกันไว้ในที่นี้ว่าในทางปฏิบัตินั้นเราจะกำหนดค่า k ขึ้นมาก่อน หรือกำหนดค่า n ขึ้นมาก่อนก็ได้เพราะมีผลเสมอกันและสืบเนื่องถึงกัน ทั้งนี้เนื่องมาจากค่า k เกิดขึ้นจากการกำหนดขนาดตัวอย่างแบบ Relative Size ซึ่งมีผลให้ทราบค่า n ได้ โดยอาศัยความสัมพันธ์

$$n = N/k$$

ขณะเดียวกัน ถ้ากำหนดค่า n ขึ้นก่อนเราย่อมสามารถคำนวณหาค่า k ได้เช่นกันโดยอาศัยความสัมพันธ์เดิมคือ

$$k = N/n$$

อย่างไรก็ตาม ถ้า $nk \neq N$ คือ k หาร N ไม่ลงตัวหรือ n หาร N ไม่ลงตัว ขนาดตัวอย่าง n จะมีได้ 2 ขนาดทำนองเดียวกันกับ k มีค่าได้ 2 ค่า ซึ่งค่า n และ k ย่อมสอดคล้องสืบเนื่องถึงกันอยู่เช่นเดิม

การเลือกตัวอย่างกระทำได้หลายวิธีแต่จะเสนอไว้ในที่นี้เพียง 3 วิธีดังนี้

**วิธีที่ 1**[1] Linear Systematic Sampling เลือกหน่วยเริ่มต้นมาจากชั้นภูมิแรกโดยสุ่มหรือเลือกเลขสุ่มที่มีค่าอยู่ระหว่าง 1 ถึง k มา 1 ค่าเพื่อกำหนดหน่วยเริ่มต้น แล้วเลือกหน่วยต่อ ๆ ไปทุก ๆ k หน่วย ดังนั้นกลุ่มตัวอย่างที่ต้องการคือกลุ่มตัวอย่างที่มีหมายเลขประจำตัวหน่วยสำรวจดังนี้คือ

---

1 $E(\bar{x}_{sy}) = \bar{X}$ เฉพาะเมื่อ $nk = N$ ถ้า $nk \neq N$ แล้ว $E(\bar{x}_{sy}) \neq \bar{X}$

$$j, j+k, j+2k, j+3k, \ldots, j+(n-1)k \ ; \ j = 1,2,\ldots,k$$

เมื่อ j คือเลขที่ของหน่วยเริ่มต้นที่เลือกจากชั้นภูมิแรก

ทั้งนี้นักวิจัยจะต้องสร้างกรอบตัวอย่างที่สมบูรณ์ขึ้นมาก่อนโดยกำหนดเลขระหัสที่ตั้งแต่ 1 ถึง k ให้แก่หน่วยสำรวจในกลุ่มประชากร แต่ในทางปฏิบัติเรามีทางเลี่ยงปัญหาการสร้างกรอบตัวอย่างที่สมบูรณ์ได้โดยเพียงแต่สำรวจและวาดแผนผังที่ตั้งหน่วยสำรวจไว้คร่าว ๆ แล้วเลือกหน่วยเริ่มต้น j เมื่อ $1 \leqslant j \leqslant k$ แล้วสำรวจต่อไป 1 หน่วยเว้นไว้ $k-1$ หน่วยเรื่อยไปจนครบขนาดตัวอย่าง n ตามต้องการ

พิจารณาตัวอย่างต่อไปนี้

กลุ่มประชากรประกอบไปด้วยหน่วยสำรวจ 15 หน่วย (N = 15) ซึ่งจัดกรอบตัวอย่างแล้วปรากฏรอบตัวอย่างดังนี้คือ $x_1, x_2, \ldots, x_{15}$ ถ้าต้องการเลือกหน่วยสำรวจมา 1 หน่วยจากทุก ๆ 3 หน่วย (k = 3) จะพบว่า

ขนาดตัวอย่าง $n = \frac{N}{k} = \frac{15}{3} = 5 =$ จำนวนชั้นภูมิ

k = 3 แสดงว่าชั้นภูมิหนึ่ง ๆ มีขนาดเท่ากับ 3 ดังนั้นหน่วยเริ่มต้นจึงเป็นไปได้ 3 หน่วย คือหน่วยที่ 1, 2 และ 3 และแสดงให้เห็นว่าจะมีตัวอย่างที่เป็นไปได้ทั้งหมด 3 ตัวอย่าง แต่ละตัวอย่างมีโอกาสได้รับเลือกเท่า ๆ กันเท่ากับ $\frac{1}{3} = \frac{1}{k}$

ให้ j เป็นหน่วยเริ่มต้นใด ๆ กลุ่มตัวอย่างที่ต้องการคือกลุ่มตัวอย่างที่มีระหัสประจำตัวดังนี้

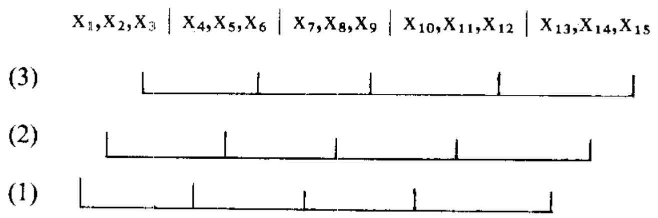
$$j, j+3, j+2(3), j+3(3), j+4(3) \quad ; \quad j = 1,2,3$$

ถ้า $j=1$ กลุ่มตัวอย่างคือ $x_1, x_4, x_7, x_{10}, x_{13}$

ถ้า $j=2$ กลุ่มตัวอย่างคือ $x_2, x_5, x_8, x_{11}, x_{14}$

ถ้า $j=3$ กลุ่มตัวอย่างคือ $x_3, x_6, x_9, x_{12}, x_{15}$

หรือจัดจำแนกกลุ่มประชากรเป็น 5 ชั้นภูมิได้ดังนี้คือ




กลุ่มตัวอย่างที่เป็นไปได้ปรากฏดังไดอะแกรมข้างต้น ขอให้สังเกตปลายลูกศรของเส้นตรงแต่ละเส้น เส้นที่ (1) คือตัวอย่างที่อาจเป็นไปได้ (Possible Sample) ชุดที่ 1 เส้นที่ (2) คือตัวอย่างที่อาจเป็นไปได้ ชุดที่ 2 เส้นที่ (3) คือตัวอย่างที่อาจเป็นไปได้ชุดที่ 3

ในกรณีที่ $n \neq \frac{N}{k}$ หรือ $nk \neq N$ หรือนัยหนึ่ง k หาร N ไม่ลงตัว กรณีเช่นนี้ จะมีผลให้ได้ขนาดตัวอย่างที่ไม่เท่ากันทุกตัวอย่าง แต่โอกาสที่ตัวอย่างหนึ่ง ๆ จะได้รับเลือกไปใช้ประโยชน์ในทางปฏิบัติงานสำรวจจะยังคงเท่า ๆ กันเท่ากับ $\frac{1}{k}$ เช่นเดิม

พิจารณาตัวอย่างข้างต้นคือ N = 15 ถ้ากำหนดว่าจะเลือกตัวอย่างมา 1 หน่วยทุก ๆ 4 หน่วย (k = 4) จะพบว่า

$$n = \frac{N}{k} = \frac{15}{4} = 3\frac{3}{4} = 3 \text{ หรือ } 4$$

ดังนี้หมายความว่าขนาดตัวอย่างอาจมีค่าเท่ากับ 3 หรือ 4 ก็ได้แต่เรายังคงได้จำนวนกลุ่มตัวอย่างทั้งสิ้น 4 กลุ่ม (k = 4) ตามเดิมและโอกาสที่ตัวอย่างหนึ่ง ๆ จะได้รับการเลือกยังคงมีค่าเท่ากับ $\frac{1}{4} = \frac{1}{k}$ ไม่เปลี่ยนแปลง

กลุ่มตัวอย่างที่เป็นไปได้คือกลุ่มที่ใช้เลขระหัสดังนี้

$$j, j+4, j+2(4), j+3(4) \qquad ; \ j = 1,2,3$$

และ $$j, j+4, j+2(4) \qquad ; \ j = 4$$

ซึ่งปรากฏดังตาราง

กลุ่มตัวอย่างสำหรับกรณีที่ $N = 15$ และ $k = 4$

| ตัวอย่างที่ | | ขนาดตัวอย่าง | โอกาสที่ได้รับเลือกใช้สำรวจ |
|---|---|---|---|
| 1 | $x_1, x_5, x_9, x_{13}$ | 4 | 1/4 |
| 2 | $x_2, x_6, x_{10}, x_{14}$ | 4 | 1/4 |
| 3 | $x_3, x_7, x_{11}, x_{15}$ | 4 | 1/4 |
| 4 | $x_4, x_8, x_{12}$ | 3 | 1/4 |

**วิธีที่ 2**[1] เลือกเลขสุ่ม j ใด ๆ ที่มีค่าอยู่ระหว่าง 1 – N มา 1 ตัวเพื่อใช้เป็นระหัสของหน่วยเริ่มต้น แต่ต้องปรับค่าเลขสุ่ม j นั้นให้มีค่าไม่เกินขนาดชั้นภูมิ (k) เสียก่อนโดยอาศัยเทคนิคการปรับเช่นเดียวกับที่กล่าวถึงแล้วในตอน 2.2 คืออาจนำ k ไปหารเลขสุ่ม j หรือนำพหุคูณของ k ไปลบออกจาก j ก็ได้ เศษที่เหลือของการปรับค่าคือระหัสประจำตัวของหน่วยเริ่มต้น นั่นก็คือ เลือกเลขสุ่ม j โดยที่ $1 \leqslant j \leqslant N$

หน่วยเริ่มต้นคือหน่วยที่ r เมื่อ $\frac{j}{k} = a + r$ โดยที่ a = ผลหาร และ r = เศษเหลือ ทั้งนี้หรือ $1 \leqslant r \leqslant k$[2] หน่วยเริ่มต้นคือหน่วยที่ r เมื่อ $j - ak = r$ โดยที่ a = พหุคูณของ k และ r = เศษเหลือทั้งนี้ $1 \leqslant r \leqslant k$

ดังนั้นกลุ่มตัวอย่างที่ต้องการคือกลุ่มตัวอย่างที่มีระหัสประจำตัวดังนี้

$$r, r+k, r+2k, r+3k, \ldots, r+(n-1)k \;;\; 1 \leqslant r \leqslant k$$

[1] วิธีนี้ $E(\bar{x}_{sy}) = \bar{X}$ เสมอไม่ว่า $nk = N$ หรือ $nk \neq N$

[2] หรือจะใช้ $0 \leqslant r \leqslant k-1$ ก็ได้ ถ้าถือว่า $r = 0$ หมายถึงการหารลงตัว หน่วยเริ่มต้นสำหรับกรณี $r = 0$ คือหน่วยที่ k นั่นคือ ถ้า $r = 0$ หน่วยเริ่มต้นคือ $x_k$ ถ้า $r = 1$ หน่วยเริ่มต้นคือ $x_1, \ldots,$ ถ้า $r = k-1$ หน่วยเริ่มต้นคือ $x_{k-1}$ แต่ถ้าใช้ $1 \leqslant r \leqslant k$ จะไม่มีปัญหาเรื่อง mapping

ตามตัวอย่างข้างต้นซึ่งกรอบตัวอย่างประกอบไปด้วย $x_1, x_2, \ldots, x_{15}$ การเลือกตัวอย่างให้เลือกเลขสุ่ม j ระหว่าง 1 - 15 มา 1 ตัว เพื่อใช้เป็นหน่วยเริ่มต้นโดยปรับค่า j ให้ไม่เกิน k เสียก่อน จะพบว่า

ก. กรณี $nk = N$

กำหนดให้ $k = 3$ จะพบว่ากลุ่มตัวอย่างปรากฏดังนี้

กลุ่มตัวอย่างสำหรับกรณีที่ $N = 15$ และ $k = 3$

| ตัวอย่างที่ | หน่วยตัวอย่าง | ขนาดตัวอย่าง | เลขสุ่ม | ความน่าจะเป็นที่ได้รับเลือก |
|---|---|---|---|---|
| 1 | $x_1, x_4, x_7, x_{10}, x_{13}$ | 5 | 1,4,7,10,13 | $\frac{1}{3} = \frac{5}{15} = \frac{n}{N}$ |
| 2 | $x_2, x_5, x_8, x_{11}, x_{14}$ | 5 | 2,5,8,11,14 | $\frac{1}{3} = \frac{5}{15} = \frac{n}{N}$ |
| 3 | $x_3, x_6, x_9, x_{12}, x_{15}$ | 5 | 3,6,9,12,15 | $\frac{1}{3} = \frac{5}{15} = \frac{n}{N}$ |

การเลือกตัวอย่างแบบ Systematic วิธีที่ 1 และวิธีที่ 2 นั้นแม้ในทางปฏิบัติจะให้ผลตรงกันคือให้ขนาดตัวอย่างเดียวกัน แต่มีวิธีปฏิบัติต่างกันเล็กน้อยและความน่าจะเป็นที่กลุ่มตัวอย่างใด ๆ จาก all possible sample จะได้รับเลือกจะต่างกันหรือสรุปได้ดังนี้

1. การเลือกตัวอย่างให้เลือกเลขสุ่ม j ใด ๆ จากตัวเลขที่อยู่ระหว่าง 1 – N ในที่นี้คือ 1 – 15 แล้วปรับค่าด้วย k เพื่อให้หน่วยที่เริ่มต้นเป็นหน่วยที่มีระหัสไม่เกิน k

เช่น เลขสุ่ม $j = 1$ จะพบว่า $\frac{j}{k} = \frac{1}{3}$ เศษ $r = 1$ หน่วยเริ่มต้นคือ $x_1$

$j = 9$ จะพบว่า $\frac{j}{k} = \frac{9}{3} = 2\frac{3}{3}$ เศษ $r = 3$ หน่วยเริ่มต้นคือ $x_3$

$j = 10$ จะพบว่า $\frac{j}{k} = \frac{10}{3} = 3\frac{1}{3}$ เศษ $r = 1$ หน่วยเริ่มต้นคือ $x_1$

$j=14$ จะพบว่า $\frac{j}{k} = \frac{14}{3} = 4\frac{2}{3}$ เศษ $r=2$ หน่วยเริ่มต้นคือ $x_2$

ดังนี้จึงพบว่า ถ้า $j = 1,4,7,10,13$ เศษ r จะเท่ากับ 1 หน่วยเริ่มต้นคือ $x_1$ หรือถ้า- $j = 1,4,7,10,13$ กลุ่มตัวอย่างที่ต้องการจะเป็นกลุ่มตัวอย่างเดียวกัน

ถ้า $j = 2,5,8,11,14$ เศษ r คือ 2 หน่วยเริ่มต้นคือ $x_2$ เดียวกัน

ถ้า $j = 3,6,9,12,15$ เศษ r คือ 3 หน่วยเริ่มต้นคือ $x_3$ เดียวกัน

ดังนั้น จำนวนกลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมดจึงยังคงเท่ากับ k กลุ่มเช่นเดียวกับการเลือกโดยวิธีที่ 1

2. จากคำอธิบายในข้อ 1 ชี้ให้เห็นว่าเราไม่มีความจำเป็นจะต้องปรับค่าของ j ก็ได้ การปรับค่า j ให้มีค่าไม่เกิน k มุ่งหมายเพียงเพื่อให้การเลือกตัวอย่างเริ่มจากซ้ายไปขวาจากหน่วยเริ่มต้น เมื่อสุดกรอบตัวอย่างก็หยุดเลือกตัวอย่างดังไดอะแกรมสำหรับกรณี $x_1$ เป็นหน่วยเริ่มต้นดังนี้

$$x_1,x_2,x_3 \mid x_4,x_5,x_6 \mid x_7,x_8,x_9 \mid x_{10},x_{11},x_{12} \mid x_{13},x_{14},x_{15}$$

จะเห็นว่าเราเริ่มที่ $x_1$ แล้ว “เดินหน้า” เรื่อย ๆ โดยเลือก 1 หน่วยเว้น 2 หน่วย จนถึง $x_{13}$ ก็จะหยุดสุ่มเพราะสุดกรอบตัวอย่างแล้ว ถ้าไม่หยุดหน่วยต่อไปจะต้องเป็น $x_{16}$ ซึ่งไม่มีในกรอบตัวอย่าง

แต่ถ้าไม่ปรับค่า j กลุ่มตัวอย่างจะยังคงเป็นชุดเดิมแต่จะต้องยอมให้การเลือกตัวอย่างได้ทั้ง “ไปและกลับ” หรือทั้ง “ซ้ายและขวาของหน่วยเริ่มต้น” โดยการเลือกจะเลือกไปจนครบขนาดตัวอย่าง n หรือครบทุกชั้นภูมิ ดังนี้กลุ่มตัวอย่างที่ต้องการคือกลุ่มตัวอย่างที่มีระหัสประจำตัวดังนี้

$$j, j \pm k, j \pm 2k, \ldots \quad ; \quad 1 \leqslant \leqslant j \leqslant N$$

หมายความว่านับจากหน่วยเริ่มต้นเราจะนับไปทางขวาทุก ๆ k หน่วย และนับไปทางซ้ายของหน่วยเริ่มต้นทุก ๆ k หน่วย โดยจะหยุดเลือกตัวอย่างถ้าพบว่า หน่วยสำรวจที่เลือก

ได้นั้นอยู่นอกกรอบตัวอย่าง วิธีนี้เรียกว่า Modified Systematic Sampling (MSS) ซึ่งเสนอโดย Cochran

เช่นเมื่อ $j = 9$ จะพบว่าหน่วยที่เริ่มต้นคือหน่วยที่ 9 หน่วยต่อไปคือหน่วยที่ $j \pm 3$ คือหน่วยที่ $9+3 = 12$ และ $9-3 = 6$ หน่วยต่อไปคือหน่วยที่ $j \pm 2(3) = 9 \pm 6$ คือหน่วยที่ 15 และหน่วยที่ 3 ไม่มีหน่วยอื่น ๆ นอกจากนี้เพราะอยู่นอกกรอบตัวอย่างทั้งหมด

ดังนั้น กลุ่มตัวอย่าง $(x_9, x_{12}, x_6, x_{15}, x_3)$ ก็คือ $(x_3, x_6, x_9, x_{12}, x_{15})$ นั่นเอง

จะเห็นว่าแผนการเลือกตัวอย่างเลือกจากทั้งหน่วยทางซ้ายและขวาของหน่วยเริ่มต้น ดังไดอะแกรม

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}$$

3. เนื่องจากเลขสุ่มระหว่าง 1 – N มีโอกาสได้รับการเลือกเท่ากันเท่ากับ $\frac{1}{N}$ ซึ่งมีผลให้หน่วยสำรวจใด ๆ มีโอกาสได้รับการเลือกเท่ากับ $\frac{1}{N}$ และจากตารางข้างต้นจะพบว่าถ้า $j = 1,4,7,10,13$ จะได้ตัวอย่างที่ 1 ถ้า $j = 2,5,8,11,14$ จะได้ตัวอย่างที่ 2 และถ้า $j = 3,6,9,12,15$ จะได้ตัวอย่างที่ 3 และเนื่องจากเลขสุ่ม j ใด ๆ มีโอกาสได้รับเลือกเท่ากับ-$\frac{1}{15}$ ดังนั้นตัวอย่างที่ 1,2,3 จึงมีโอกาสได้รับเลือกเท่ากับ $\frac{5}{15}$ , $\frac{5}{15}$ และ $\frac{5}{15}$ ตามลำดับ[1]

หรือในกรณีทั่วไปสำหรับกลุ่มตัวอย่างขนาด n กลุ่มตัวอย่างดังกล่าวจะมีโอกาสได้รับเลือกเท่ากับ $\frac{n}{N}$ (ดูการพิสูจน์ในบทที่ 2 ตอน 2.1)

ข. กรณี $nk \neq N$

สำหรับกรณี $nk \neq N$ ก็ยังคงปฏิบัติเช่นเดียวกับข้อ ก. แต่ขนาดตัวอย่างไม่เท่ากันในทุกกลุ่ม ส่วนความน่าจะเป็นที่ตัวอย่างชุดใด ๆ จะได้รับเลือกยังคงเท่ากับ $\frac{n}{N}$

ตามตัวอย่างข้างต้น ถ้า $N = 15$ และ $k = 4$ จะพบว่า $n = \frac{N}{k} = \frac{15}{4} = 3\frac{3}{4} = 3$ หรือ 4 และปรากฏกลุ่มตัวอย่างดังนี้

---

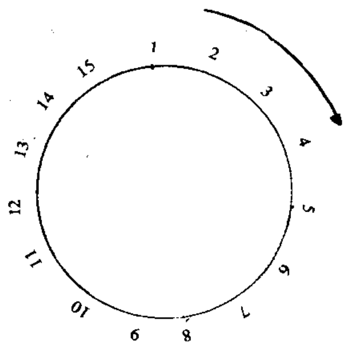
1 ค่าเท่ากับกับกรณีการเลือกตัวอย่างวิธีที่ 1

กลุ่มตัวอย่างสำหรับกรณีที่ N = 15 และ k = 4

| ตัวอย่างที่ | หน่วยตัวอย่าง | ขนาดตัวอย่าง n | เลขสุ่ม | ความน่าจะเป็นที่จะได้รับเลือก |
|---|---|---|---|---|
| 1 | $x_1,x_5,x_9,x_{13}$ | 4 | 1,5,9,13 | 4/15 |
| 2 | $x_2,x_6,x_{10},x_{14}$ | 4 | 2,6,10,14 | 4/15 |
| 3 | $x_3,x_7,x_{11},x_{15}$ | 4 | 3,7,11,15 | 4/15 |
| 4 | $x_4,x_8,x_{12}$ | 3 | 4,8,12 | 3/15 |

**วิธีที่ 3** Circular Systematic Sampling (CSS)[1]

CSS คือแผนการเลือกตัวอย่างที่จัดเรียงระหัสประจำตัวของหน่วยสำรวจไว้ในรูปวงกลม เจตนา สำคัญมุ่งแก้ปัญหาเมื่อ $nk \neq N$ [2] เพราะเมื่อ n หาร N ไม่ลงตัวหรือ k หาร N ไม่ลงตัว ขนาดตัวอย่างที่ได้รับจะไม่คงที่เท่ากัน ดังตัวอย่างชุดที่ 4 ของตารางข้างบนซึ่งมีขนาดตัวอย่าง n = 3 ไม่อาจมีขนาดเท่ากับ 4 เหมือน 3 ตัวอย่างแรกที่เป็นเช่นนี้เพราะถ้าจะให้ได้ขนาดตัวอย่างครบ n = 4 หน่วยตัวอย่างต่อไปจะเป็น $x_{16}$ ซึ่งพ้นออกจากกรอบตัวอย่าง แต่ถ้าหากนำระหัสในกรอบตัวอย่างเรียงต่อเป็นวงจรในรูปวงกลม จะพบว่าหน่วยตัวอย่างต่อไปจะไม่หลุดพ้นนอกกรอบตัวอย่าง และพบว่าหน่วยตัวอย่างต่อไปก็คือ $x_1$ [3] ดังภาพ




[1] การนำ CSS มาใช้มีผลทำให้เรียกแผนการเลือกตัวอย่าง 2 วิธีแรกว่า Linear Systematic Sampling (LSS)

[2] ถ้า nk = N แผนการเลือกตัวอย่าง CSS จะให้ผลลัพธ์ทุกประการตรงกันกับ LSS เมื่อ nk = N

[3] ปรับ $x_{16}$ ให้มีระหัสไม่เกิน N = 15 ด้วยการหักลบด้วย 15 นั่นคือหน่วยที่ 4 คือ $x_{16-15} = x_1$

จึงเห็นได้ว่าเมื่อใช้แผนการเลือกตัวอย่างแบบ CSS ขนาดตัวอย่างจะคงที่เท่ากันเสมอ

วิธีการของ CSS ปรากฏดังนี้คือ

1. เลือกเลขสุ่ม j ใด ๆ ที่ $1 \leqslant j \leqslant N$ เพื่อใช้เป็นหน่วยเริ่มต้น

2. หน่วยต่อไปคือหน่วยที่ $j+ik$ ถ้า $j+ik \leqslant N$ และ $j+ik-N$ ถ้า $j+ik > N$

เมื่อ $i = 0,1,2,\ldots, n-1$

**หมายเหตุ**

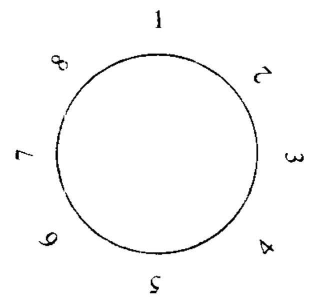
1. $j+ik-N$ ถ้า $j+ik > N$ หมายความว่าถ้าระหัสของหน่วยสำรวจจากกรอบตัวอย่างที่คำนวณได้จากสูตร $j+ik$ มีค่าเกิน N ให้ปรับแก้เสียใหม่ให้ไม่เกิน N ด้วยการหักลบด้วย N

2. การกำหนดว่า $i = 0,1,2,\ldots,n-1$ มีความหมายเช่นเดียวกันกับการกำหนดกลุ่มตัวอย่างเป็น $j,j+k,j+2k,\ldots,j+(n-1)k$ ของ LSS

3. CSS กับ MSS มีลักษณะใกล้เคียงกันมาก

พิจารณาตัวอย่างต่อไปนี้

กำหนดให้กรอบตัวอย่างประกอบด้วยหน่วยสำรวจที่มีเลขระหัส $1-8$ $(N=8)$ และต้องการเลือกตัวอย่างมา 1 หน่วยทุก ๆ 3 หน่วย $(k=3)$ จะพบว่า $n = 8/3 = 2.67 \simeq 3$ ดังนั้น กลุ่มตัวอย่างที่เป็นไปได้ทั้งหมดจะปรากฏดังนี้ (ดูไดอะแกรม)




$1 \leqslant j \leqslant 8$

CSS สำหรับกรณีที่ N = 8 และ k = 3

| ตัวอย่างที่ | หน่วยตัวอย่าง | ขนาดตัวอย่าง | ความน่าจะเป็นที่ได้รับเลือก[1] |
|---|---|---|---|
| 1 | $x_1,x_4,x_7$ | 3 | 1/8 |
| 2 | $x_2,x_5,x_8$ | 3 | 1/8 |
| 3 | $x_3,x_6,x_1$ | 3 | 1/8 |
| 4 | $x_4,x_7,x_2$ | 3 | 1/8 |
| 5 | $x_5,x_8,x_3$ | 3 | 1/8 |
| 6 | $x_6,x_1,x_4$ | 3 | 1/8 |
| 7 | $x_7,x_2,x_5$ | 3 | 1/8 |
| 8 | $x_8,x_3,x_6$ | 3 | 1/8 |

ถ้าต้องการเลือกตัวอย่างมา 1 หน่วยทุก ๆ 4 หน่วย (k = 4) จะพบว่า $n = 8/4 = 2$ ดังนั้น กลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมดจะปรากฏดังนี้

CSS สำหรับกรณี N = 8 และ k = 4

| ตัวอย่างที่ | หน่วยตัวอย่าง | ขนาดตัวอย่าง | ความน่าจะเป็นที่ได้รับเลือก |
|---|---|---|---|
| 1 | $x_1,x_5$ | 2 | 1/4 |
| 2 | $x_2,x_6$ | 2 | 1/4 |
| 3 | $x_3,x_7$ | 2 | 1/4 |
| 4 | $x_4,x_8$ | 2 | 1/4 |

[1] มี 8 possible sample แต่ละตัวอย่างจึงมีโอกาสได้รับเลือกเท่ากับ 1/8

จากตารางข้างต้นมีข้อที่น่าสังเกตและสามารถสรุปคุณลักษณะของ CSS ได้ดังนี้

1. ถ้า $nk = N$ ผลลัพธ์จะตรงกับ LSS ทุกประการ คือมีตัวอย่างที่เป็นไปได้ทั้งสิ้น k ชุด ชุดหนึ่ง ๆ มีโอกาสได้รับเลือกเท่ากับ $n/N = 1/k$

2. ถ้า $nk \neq N$ จะพบว่า

ก. จำนวนกลุ่มตัวอย่างที่พึงเป็นไปได้มีอยู่ทั้งสิ้น N ชุด

ข. หน่วยสำรวจแต่ละหน่วยในกลุ่มประชากรสามารถรวมอยู่ในกลุ่มตัวอย่างได้ทั้งสิ้น n ชุด

ค. หน่วยสำรวจแต่ละหน่วยมีโอกาสได้รับเลือกเท่ากับ $1/N$

ง. กลุ่มตัวอย่างแต่ละกลุ่มมีโอกาสได้รับเลือกเท่ากับ $1/N$

จ. ไม่มีข้อจำกัดใดว่าหน่วยเริ่มต้นจะต้องเริ่มจากเลข $1-k$ หรือนัยหนึ่งเราจะใช้หน่วยใดเป็นหน่วยเริ่มต้นก็ได้คือ $1 \leqslant j \leqslant N$ แล้วเลือกหน่วยต่าง ๆ ไปจนครบขนาดตัวอย่างที่กำหนด

ฉ. $E(\bar{x}_{sy}) = \bar{X}$

ในทางปฏิบัติเราสามารถเลือกใช้แผนการเลือกตัวอย่างวิธีใดก็ได้และไม่จำเป็นต้องคำนึงว่า $nk = N$ หรือ $nk \neq N$ มากนัก เพราะเราสามารถมองข้ามประเด็นนี้ไปได้ ไม่ว่าขนาดตัวอย่าง n จะมากกว่า 50 หรือไม่ก็ตามเพราะเชื่อว่าจะไม่ส่งผลเสียหายต่อค่าประมาณมากนัก[1]

สำหรับการศึกษาในลำดับต่อไปนี้ เราจะถือว่า $nk = N$ เสมอ ทั้งนี้ก็เพื่อความสะดวกและเชื่อว่าแม้ $nk \neq N$ ก็คงไม่ส่งผลสะท้อนต่อค่าประมาณมากนักจนสามารถมองข้ามไปได้

## 4.4 การกะประมาณค่าเฉลี่ยประชากร

### 4.4.1 การประมาณค่ายอดรวม T ค่าเฉลี่ยประชากร $\bar{X}$ และสัดส่วนประชากร P

---

[1] William G. Cochran "Sampling Technique" p.207

**ทฤษฎี 4.1** ในการสุ่มตัวอย่างโดยวิธี Systematic โดยเลือกตัวอย่างมาทุก ๆ k หน่วย

ก. ตัวประมาณค่าที่ปราศจากอคติของยอดรวมประชากร T คือ

$$\hat{T}_{sy} = k \sum_{j}^{n} x_{ij} = kT_i ; i = 1,2,\ldots,k$$

ข. ตัวประมาณค่าที่ปราศจากอคติของค่าเฉลี่ยประชากรคือ

$$\hat{\bar{X}} \doteq \bar{x}_{sy} = \hat{T}_{sy}/N = \frac{k}{N} \sum_{j}^{n} x_{ij} ; i = 1,2,\ldots,k$$

ค. ตัวประมาณค่าที่ปราศจากอคติของสัดส่วนประชากร P คือ

$$\hat{P} = p_{sy} = \frac{k}{N} \sum_{j}^{n} x_{ij} ; i = 1,2,\ldots,k \text{ เมื่อ } x_{ij} = 0,1$$

ทั้งนี้มิได้มีข้อจำกัดว่าขนาดตัวอย่างจะต้องคงที่เท่ากันเสมอหรือนัยหนึ่งก็คือไม่มีข้อจำกัดว่า $nk = N$ หรือ $nk \neq N$ [1]

**ข้อสังเกต**

ดังที่กล่าวแล้วในตอนต้นว่า Systematic Sampling ก็คือกรณีหนึ่งของการสำรวจแบบแบ่งชั้นภูมิที่ชั้นภูมิหนึ่ง ๆ มีขนาดเท่ากันเท่ากับ $k(N_h = k)$ และเลือกตัวอย่างจากแต่ละชั้นภูมิซึ่งมีอยู่ทั้งสิ้น $n(L = n)$ ชั้นภูมิ มาชั้นภูมิละ 1 หน่วย $(n_h = 1)$

ดังนั้น โดยนัยของแผนสำรวจแบบแบ่งชั้นภูมิจะพบว่า

$$\hat{T}_{st} = \sum_{h}^{L} N_h\bar{x}_h \text{ และ } \bar{x}_{st} = \frac{1}{N} \sum_{h}^{L} N_h\bar{x}_h$$

เมื่อเปรียบเทียบกับกรณีเฉพาะนี้จะพบว่า $N_h = k, n_h = 1, L = n, N = \sum_{h}^{L} N_h = \sum_{h}^{n} k = nk = N$ และเปลี่ยน subscript h เป็น i และ i เป็น j (เพื่อประโยชน์ในการเปรียบเทียบ)

[1] Dej Raj, p.44

จะพบว่า $\bar{x}_h = \frac{1}{n_h}\sum_{i}^{n_h} x_{hi} = x_{ij}$

ดังนั้น $\hat{T}_{st} = \sum_{h}^{L} N_h\bar{x}_h = \sum_{j}^{n} kx_{ij} = k\sum_{j}^{n} x_{ij} = \hat{T}_{sy}$

$$\bar{x}_{st} = \frac{1}{N}\sum_{h}^{L} N_h\bar{x}_h = \frac{k}{N}\sum_{j}^{n} x_{ij} = \bar{x}_{sy}$$

**พิสูจน์**

$$E(\hat{T}_{sy}) = E(k\sum_{j}^{n} x_{ij}) = kE(\sum_{j}^{n} x_{ij})$$

$$= k\sum_{i}^{k}(\sum_{j}^{n} x_{ij})_i \frac{1}{k} \quad ^{1.}$$

$$= \sum_{i}^{k}\sum_{j}^{n} x_{ij} = T \quad ^{2.}$$

ผลลัพธ์ที่ตามมาก็คือ $E(\bar{x}_{sy}) = E(\hat{T}_{sy}/N) = T/N = \bar{X}$ และ $E(p_{sy}) = P$

**ทฤษฎี 4.2** เมื่อเลือกตัวอย่างโดยใช้ CSS ตัวประมาณค่าที่ปราศจากอคติของ $\bar{X}$ คือ $\bar{x}_{sy} = \frac{1}{n}\sum_{j}^{n} x_{ij}$

**พิสูจน์**

$$E(\bar{x}_{sy}) = \frac{1}{n}E(\sum_{j}^{n} x_{ij})$$

$$= \frac{1}{n}\sum_{i}^{N}(\sum_{j}^{n} x_{ij})_i \cdot \frac{1}{N} \quad ^{3}$$

$$= \frac{1}{n}\cdot\frac{1}{N}\sum_{i}^{N}\sum_{j}^{n} x_{ij}$$

$$= \frac{1}{n}\cdot\frac{1}{N}n(x_1 + x_2 + \ldots + x_N) \quad ^{4}$$

---

[1] ความน่าจะเป็นที่ตัวอย่างกลุ่มใด ๆ จะได้รับเลือกค่าเท่ากับ 1/k

[2] ดูข้อสังเกตที่ (4) ตอน 4.2

[3] ใน CSS นั้นกลุ่มตัวอย่างที่ i ใด ๆ มีโอกาสได้รับเลือกเท่ากับ 1/N

[4] ใน CSS หน่วยสำรวจหนึ่ง ๆ จะปรากฏในกลุ่มตัวอย่างได้ n ชุด

$$= \frac{1}{N}(x_1 + x_2 + \ldots + x_N)$$

$$= \bar{X}$$

**หมายเหตุ** ในที่นี้แม้จะไม่สนใจว่า n จะหาร N ลงตัวหรือไม่เพราะยึดถือเอาความง่ายและความสะดวกเป็นประมาณและถือว่าแม้ $nk \neq N$ ก็มิได้มีผลต่อการประมาณค่ามากนัก แต่ก็ไม่อาจมองข้ามกรณีที่น่าสนใจของ CSS ไปได้เพราะถ้าพบกรณีที่ไม่อาจละเลยปัญหา $nk \neq N$ เช่น $n << 50$ ก็จำเป็นจะต้องสนใจ CSS เพราะเหมาะสมกว่าขอให้สังเกตว่าโครงสร้างของ $\bar{x}_{sy}$ ใน CSS มีโครงสร้างเดียวกันกับ $\bar{x}$ ใน SRS

## 4.5 ความแปรปรวนของค่าประมาณ $\bar{x}_{sy}$

การศึกษาในลำดับต่อไปนี้จะเป็นการพิจารณาความแปรปรวนของการประมาณค่า $\bar{x}_{sy}$ แต่สิ่งหนึ่งที่ต้องการชี้ให้เห็นเป็นข้อสังเกตไว้ก่อนก็คือสูตร $V(\bar{x}_{sy})$ จะเป็นโครงสร้างที่เกี่ยวข้องกับความแปรปรวนของประชากรทั้งกลุ่ม ($S^2$) และความผันแปรระหว่างหน่วยสำรวจในทุกกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมด ($S^2_{wsy}$) ซึ่งจะไม่ให้ประโยชน์ในการนำไปใช้ประโยชน์ในทางปฏิบัติเลย โดยเฉพาะ $S^2_{wsy}$ จะต้องใช้ข้อมูลจากทุกกลุ่มตัวอย่างทั้ง k กลุ่ม แต่ในทางปฏิบัตินั้นเราจะสุ่มตัวอย่างมาใช้เพียงกลุ่มเดียว ซึ่งเมื่อเป็นเช่นนี้ในทางปฏิบัติเราจึงไม่อาจหาค่า $S^2_{wsy}$ ได้ อย่างไรก็ตามสูตร $V(\bar{x}_{sy})$ แม้ไม่อาจนำไปใช้ในทางปฏิบัติได้ แต่ก็สามารถช่วยชี้ให้เห็นพัฒนาการขั้นต่อไปของแผนสำรวจแบบนี้ เช่นลักษณะหรือธรรมชาติของกลุ่มประชากร เช่นการเรียงลำดับของหน่วยสำรวจ และการพัฒนาโครงสร้างสร้าง $V(\bar{x}_{sy})$ ในรูปอื่น ๆ ต่อไป

**ทฤษฎี 4.3** ในการสำรวจโดยใช้แผนสำรวจแบบ Systematic ความแปรปรวนของค่าประมาณ $\bar{x}_{sy}$ คือ

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{1}{N} \sum_{i}^{k} \sum_{j}^{n} (x_{ij} - \bar{x}_i)^2$$

หรือ $V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy}$

เมื่อ $S^2_{wsy} = \frac{1}{k(n-1)} \sum_i^k \sum_j^n (x_{ij} - \bar{x}_i)^2$ = ความแปรปรวนระหว่างหน่วยสำรวจในกลุ่มตัวอย่างชุดเดียวกัน[1]

**พิสูจน์** จากนิยาม $V(\bar{x}) = E(\bar{x}_i - \bar{X})^2 = \sum_{\text{all } i} (\bar{x}_i - \bar{X})^2 \cdot Pr(\bar{x}_i)$

ดังนั้น $V(\bar{x}_{sy}) = \frac{1}{k} \sum_i^k (\bar{x}_i - \bar{X})^2$ [2] ; $Pr(\bar{x}_i) = \frac{1}{k}$

พิจารณา $SST = (N-1)S^2 = \sum_i^k \sum_j^n (x_{ij} - \bar{X})^2$

โดยอาศัยหลักเกณฑ์การจำแนกความแปรปรวนใน ANOVA [3]

$$=> (N-1)S^2 = \sum_i^k \sum_j^n (x_{ij} - \bar{X})^2 = \sum_i^k \sum_j^n (x_{ij} - \bar{x}_i + \bar{x}_i - \bar{X})^2 \text{ [4]}$$

$$= \sum_i^k \sum_j^n (x_{ij} - \bar{x}_i)^2 + \sum_i^k \sum_j^n (\bar{x}_i - \bar{X})^2 \text{ [5]}$$

---

[1] มีความหมายเดียวกันกับ Mean Square Within Group (MSW) ใน ANOVA ขอให้สังเกตว่า

$\sum_i^k \sum_j^n (x_{ij} - \bar{x}_i)^2$ ก็คือ SSW

[2] มีความหมายเดียวกันกับ Mean Square Between Group (MSB)

[3] ถือว่าเรามี Treatment อยู่ k Treatment ซึ่ง $\bar{x}_i$ คือ Treatment Mean

[4] $\bar{X}_i = \frac{1}{n} \sum_j^n x_{ij}$ = ตัวแปรสุ่มแสดงค่าเฉลี่ยของกลุ่มตัวอย่างที่ $i = 1,2,\ldots,k$ หรือจะนิยามให้สอดคล้องกับที่นิยามในเรื่อง ANOVA ก็ได้ คือใช้ $X_{i.} = \sum_j^n x_{ij}$ และ $\bar{X}_{i.} = \frac{1}{n} X_{i.}$ ค่าที่เป็นตัวเลขของตัวแปรสุ่ม $\bar{X}_i$ คือ $\bar{x}_i$ หรือ $\bar{x}_{i.}$ แต่ในที่นี้ใช้ $\bar{x}_i$ เพราะถือว่า $\bar{X}_i$ คือพารามิเตอร์

[5] เทอมไขว้มีค่าเท่ากับ 0 เพราะ

$$\sum_i^k \sum_j^n (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{X}) = \sum_j^n (x_{ij} - \bar{x}_i) \sum_i^k (\bar{x}_i - \bar{X}) = (\sum_j^n x_{ij} - n\bar{x}_i) \sum_i^k (\bar{x}_i - \bar{X})$$

$$= (\frac{n}{n} \sum_j^n x_{ij} - n\bar{x}_i) \sum_i^k (\bar{x}_i - \bar{X}) = (n\bar{x}_i - n\bar{x}_i) \sum_i^k (\bar{x}_i - \bar{X}) = 0$$

$$= \sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2 + n\sum_{i}^{k}(\bar{x}_i-\bar{X})^2$$

หารตลอดด้วย nk

$$=> \frac{(N-1)S^2}{nk} = \frac{1}{nk}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2 + \frac{1}{k}\sum_{i}^{k}(\bar{x}_i-\bar{X})^2$$

แต่เนื่องจาก nk = N และ $V(\bar{x}_{sy}) = \frac{1}{k}\sum_{i}^{k}(\bar{x}_i-\bar{X})^2$

ดังนั้น $\frac{(N-1)S^2}{N} = \frac{1}{N}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2 + V(\bar{x}_{sy})$

$$=> \quad V(\bar{x}_{sy}) = \frac{N-1}{N}S^2 - \frac{1}{N}\sum_{k}^{k}\sum_{n}^{n}(x_{ij}-\bar{x}_i)^2$$

และเมื่อนิยามว่า $S^2_{wsy} = \frac{1}{k(n-1)}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2$ แสดงถึงความแปรปรวนภายในกลุ่มตัวอย่างเดียวกัน[1] ดังนั้นเราจึงเสนอสูตร $V(\bar{x}_{sy})$ ได้อีกรูปหนึ่งคือ

$$V(\bar{x}_{sy}) = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S^2_{wsy}$$

**ข้อสังเกต**

พิจารณาโครงสร้าง $V(\bar{x}_{sy})$ ข้างต้นจะพบว่า

1. $V(\bar{x}_{sy})$ ประกอบไปด้วยความแปรปรวน 2 ส่วนคือ $S^2$ และ $S^2_{wsy}$ โดยที่ $S^2$ คือความแปรปรวนภายในกลุ่มประชากรทั้งกลุ่มคือ

$$S^2 = \frac{1}{N-1}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{X})^2 = \frac{1}{N-1}\left\{(x_1-\bar{X})^2+(x_2-\bar{X})^2+\ldots+(x_n-\bar{X})^2\right\}$$

และ $S^2_{wsy} = \frac{1}{k(n-1)}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2$

จะพบว่าในทางปฏิบัตินั้นเราจะเลือกตัวอย่างมาใช้เพียงชุดเดียวมิได้นำทุก Possible

---

[1] wsy ย่อมาจาก within same systematic sample