บทที่ 2

การสุ่มตัวอย่างอย่างง่าย

Simpe Random Sampling Plan หรือ Unrestricted Random Sampling Plan

2.1 ความหมาย

การสุ่มตัวอย่างคือการเลือกตัวแทนของประชากรมาเพียงบางหน่วยเพื่อศึกษาคุณลักษณะทางประชากร (Characteristics) ของประชากรกลุ่มนั้น โดยถือว่าประชากรคือกลุ่มของหน่วยสำรวจหรือหน่วยขั้นต้นซึ่งมีจำนวนที่นับได้ คือ N หน่วย และเราสุ่มตัวอย่างหน่วยสำรวจมาจากกลุ่มประชากรกลุ่มนี้มาเพียง n หน่วย ในที่นี้ $n < N$

การสุ่มตัวอย่างนั้น โดยปกติจะใช้วิธีสุ่มแล้วไม่ใส่คืน (Sampling Without Replacement) ทั้งนี้เพราะถือว่าการสุ่มแบบไม่ใส่คืนจะทำให้ได้หน่วยตัวอย่างที่แตกต่างกัน กล่าวคือในกลุ่มตัวอย่างชุดเดียวกันจะไม่มีหน่วยสำรวจซ้ำกัน ส่วนการสุ่มแบบใส่คืนจะทำให้เกิดหน่วยซ้ำกันได้ในกลุ่มตัวอย่างชุดเดียวกันซึ่งในกรณีเช่นนี้จะไม่ทำให้ได้รับข้อสนเทศอะไรเพิ่มขึ้นมาจากหน่วยที่ซ้ำกันนั้น ดังนั้น จากกลุ่มตัวอย่างขนาด n ที่สุ่มมาจากประชากร N หน่วย โดยใช้วิธีการสุ่มแล้วไม่ใส่คืนจะทำให้ได้กลุ่มตัวอย่างที่แตกต่างกัน (All Possible Sample) ทั้งสิ้น

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \quad \text{ชุด}$$

ซึ่งแต่ละชุดจะประกอบไปด้วยหน่วยตัวอย่างที่ไม่ซ้ำกัน n หน่วย ดังนั้นแผนการสำรวจอย่างง่าย (Simple Random Sampling; SRS) คือแผนสำรวจที่ถือว่า การเลือกตัวอย่าง

ก็คือการเลือก (โดยสุ่ม) กลุ่มตัวอย่างจากที่มีทั้งสิ้น $\binom{N}{n}$ ชุดมาเพียง 1 ชุด ซึ่งแต่ละชุดมีโอกาสได้รับเลือกเท่ากัน เท่ากับ $\frac{1}{\binom{N}{n}}$ หรือ SRS คือแผนสำรวจที่หน่วยสำรวจแต่ละหน่วยมีโอกาสได้รับเลือกอย่างเท่าเทียมกัน หมายความว่าถ้าจะพิจารณาให้ลึกลงไปถึงรากฐานเดิมของเรื่องแล้วจำเป็นจะต้องพิจารณาถึงโอกาสที่แต่ละหน่วยจะได้รับการคัดเลือกจากกลุ่มประชากรปัจจุบันขณะนั้น[1] ในกรณีของ SRS เราต้องให้โอกาสแก่หน่วยสำรวจทุกหน่วยในอันที่จะได้รับการเลือกอย่างเท่าเทียมกัน มิเช่นนั้นจะเกิดอคติขึ้นได้ และเมื่อตั้งเป้าหมายว่าการคัดเลือกครั้งนี้เราต้องการหน่วยสำรวจมาเพียง n หน่วย โอกาสที่หน่วยสำรวจแต่ละหน่วยในกลุ่มประชากรปัจจุบันขณะนั้นจะถูกคัดเลือกเข้าไว้ในตัวอย่างขนาด n ซึ่งเท่ากับ $\frac{n}{N}, \frac{n-1}{N-1}, \ldots, \frac{1}{N-n+1}$ จะประกอบกันเป็นโอกาสที่จะได้กลุ่มตัวอย่างใด ๆ ที่มีขนาด n

ความน่าจะเป็นที่จะได้จากกลุ่มตัวอย่างขนาด n นั้นสามารถคำนวณให้เห็นได้ 2 วิธี ในที่นี้จะแสดงให้เห็นโดยใช้กลุ่มตัวอย่างขนาดเล็ก คือ n = 3 ทั้งนี้เพื่อป้องกันความสับสนแล้วค่อยขยายความ (Generalization) สู่กรณีกลุ่มตัวอย่างขนาด n ซึ่งเป็นกรณีทั่วไปในภายหลัง

วิธีที่ 1 กลุ่มตัวอย่างที่ต้องการประกอบด้วยหน่วยสำรวจ 3 หน่วย จากกลุ่มประชากรขนาด N

หน่วยที่ 1 มีโอกาสได้รับเลือก $= \frac{1}{N}$

หน่วยที่ 2 มีโอกาสได้รับเลือก $= \frac{1}{N-1}$

---

[1] คำว่าขนาดประชากรปัจจุบัน หมายถึง ขนาดประชากรที่มีอยู่หรือที่เหลืออยู่ในขณะที่กำลังจะเลือก เช่น ขนาดประชากร N = 100 เมื่อจะเลือกหน่วยตัวอย่างที่ 1 ขนาดประชากรปัจจุบันจะเท่ากับ 100 ขณะจะเลือกตัวอย่างที่ 2 ขนาดประชากรปัจจุบันจะเท่ากับ 99 ทั้งนี้เพราะขนาดประชากรลดลง 1 หน่วยเนื่องจากการเลือกหน่วยตัวอย่างที่ 1 ดังนี้เป็นต้น

หน่วยที่ 3 มีโอกาสได้รับเลือก $= \frac{1}{N-2}$

ดังนั้น โอกาสที่จะได้กลุ่มตัวอย่างขนาด $n = 3$ ชุดหนึ่งชุดใดจากกลุ่มประชากรขนาด N

$$= \frac{1}{N} \cdot \frac{1}{N-1} \cdot \frac{1}{N-2}$$

แต่กลุ่มตัวอย่างที่ประกอบด้วยหน่วยสำรวจ 3 หน่วยนั้นมีโอกาสปรากฏขึ้นได้ 3! วิธี ทั้งนี้เพราะในบรรดาหน่วยสำรวจทั้ง 3 นั้น หน่วยหนึ่งหน่วยใดอาจได้รับเลือกก่อนหลังสลับกันได้ 3! วิธี

$\therefore$ โอกาสที่จะได้กลุ่มตัวอย่างขนาด $n = 3$

$$= \frac{1}{N} \cdot \frac{1}{N-1} \cdot \frac{1}{N-2} \cdot 3!$$

$$= \frac{3!}{N(N-1)(N-2)}$$

$$= \frac{3!}{\frac{N!}{(N-3)!}}$$

$$= \frac{1}{\frac{N!}{(N-3)!\,3!}}$$

$$= \frac{1}{\binom{N}{3}}$$

ดังนั้นโอกาสที่จะได้รับกลุ่มตัวอย่างขนาด n

$$= \frac{1}{N} \cdot \frac{1}{N-1} \cdot \frac{1}{N-2} \cdots \frac{1}{N-(N-1)} \cdot n!$$

$$= \frac{n!}{N(N-1)(N-2)\ldots(N-(n-1))}$$

$$= \frac{n!}{\frac{N!}{(N-n)!}}$$

$$= \frac{1}{\frac{N!}{(N-n)!\,n!}}$$

$$= \frac{1}{\binom{N}{n}}$$

ข้อสังเกต

$\frac{1}{N}$ คือความน่าจะเป็นที่หน่วยสำรวจหน่วยหนึ่งหน่วยใดในกลุ่มประชากรขนาด N จะได้รับเลือกเมื่อเลือกไปแล้ว 1 หน่วย ขนาดประชากรปัจจุบันที่เหลืออยู่คือ $N - 1$ ดังนั้นโอกาสที่หน่วยสำรวจหน่วยหนึ่งหน่วยใดในกลุ่มประชากรขนาด $N - 1$ จะได้รับเลือกจึงเท่ากันกับ $\frac{1}{N-1}$ ขนาดประชากรจะลดลงคราวละ 1 หน่วยทุกครั้งที่มีการเลือกตัวอย่างออกไป 1 หน่วย ดังนั้นความน่าจะเป็นที่จะได้หน่วยตัวอย่างลำดับต่อ ๆ ไปจึงมีค่าเท่ากับ $\frac{1}{N-2}, \frac{1}{N-3}, \frac{1}{N-4}, \ldots$ เรื่อยไป จนกระทั่งหน่วยที่ n ซึ่งมีโอกาสได้รับเลือกเท่ากับ $\frac{1}{N-(n-1)}$ ทั้งนี้เพราะได้มีการเลือกหน่วยสำรวจออกไปก่อนหน้านั้นแล้ว $n - 1$ หน่วย จึงมีหน่วยสำรวจเหลือให้เลือกเพียง $N - (n - 1)$ หน่วย

วิธีที่ 2 คำนวณหาความน่าจะเป็นที่หน่วยสำรวจหนึ่งใดจะถูกเลือกเข้ามาในกลุ่มตัวอย่าง ขนาด n หลังจากนั้น คำนวณหาค่าความน่าจะเป็นที่จะได้ตัวอย่างขนาด n ตามต้องการ

ในที่นี้จะกำหนดให้ $n = 3$

ก. หาความน่าจะเป็นที่หน่วยสำรวจใด ๆ จะถูกเลือกเข้ามาในกลุ่มตัวอย่างขนาด $n = 3$

สมมุติว่าหน่วยที่ต้องการคือ A

(1) Pr (A จะได้รับเลือกในการเลือกครั้งแรก) $= \frac{1}{N}$

(2) Pr (A จะได้รับเลือกในการเลือกครั้งที่ 2) = Pr (A ไม่ได้รับเลือกในการเลือกครั้งแรกหรือว่าครั้งแรกเลือกได้หน่วยอื่น แต่กลับได้รับเลือกในการเลือกครั้งที่ 2)

= Pr (A ไม่ได้รับเลือกในการเลือกครั้งแรก) $\cdot$ Pr (A ได้รับเลือกในการเลือกครั้งที่ 2)

$$= \frac{N-1}{N} \cdot \frac{1}{N-1}$$

(3) Pr (A ได้รับเลือกในการเลือกครั้งที่ 3) = Pr (A ไม่ได้รับเลือกในการเลือกครั้งแรกและครั้งที่ 2 แต่ได้รับเลือกในการเลือกครั้งที่ 3)

= Pr (A ไม่ได้รับเลือกในการเลือกครั้งแรก) $\cdot$ Pr (A ไม่ได้รับเลือกในการเลือกครั้งที่ 2) . Pr (A ได้รับเลือกในการเลือกครั้งที่ 3)

$$= \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{1}{N-2}$$

ดังนั้น Pr (A ได้รับเลือกเข้ามาเป็นตัวอย่างในกลุ่มตัวอย่างขนาด $n = 3$)

= Pr (A อาจได้รับเลือกในการเลือกครั้งแรก หรือ ได้รับเลือกในการเลือกครั้งที่ 2 หรือ ได้รับเลือกในการเลือกครั้งที่ 3 ก็ได้)

= Pr (A ได้รับเลือกในการเลือกครั้งแรก) + Pr (A ได้รับเลือกในการเลือกครั้งที่ 2) + Pr (A ได้รับเลือกในการเลือกครั้งที่ 3)

$$= \frac{1}{N} + \frac{N-1}{N} \cdot \frac{1}{N-1} + \frac{N-1}{N} \, \frac{N-2}{N-1} \cdot \frac{1}{N-2}$$

$$= \frac{1}{N} + \frac{1}{N} + \frac{1}{N}$$

$$= \frac{3}{N}$$

นั่นคือ Pr (หน่วยสำรวจหนึ่งใดจะได้รับเลือกเข้ามาเป็นตัวอย่างในกลุ่มตัวอย่างขนาด $n = 3$) $= \frac{3}{N}$

จากผลการคำนวณในกรณีของกลุ่มตัวอย่างขนาด $n = 3$ นี้ทำให้สามารถขยายความสู่กรณีทั่วไปของกลุ่มตัวอย่างขนาด n ได้ดังนี้

Pr (หน่วยสำรวจหนึ่งใดจะได้รับเลือกเข้ามาเป็นตัวอย่างในกลุ่มตัวอย่างขนาด n)

$$= \frac{1}{N} + \frac{N-1}{N} \cdot \frac{1}{N-1} + \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{1}{N-2} + \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{N-3}{N-2}$$

$$\frac{1}{N-3} + \ldots + \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{N-3}{N-2} \cdot \frac{N-4}{N-3} \cdots \frac{N-(n-1)}{N-(n-2)} \cdot \frac{1}{N-(n-1)}$$

$$= \frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N}$$

$$= \frac{n}{N}$$

ข) ความน่าจะเป็นที่จะได้กลุ่มตัวอย่างใด ๆ ขนาด n

ในที่นี้จะทดลองคำนวณกับกรณีขนาดตัวอย่าง $n = 3$ สมมุติกลุ่มตัวอย่างคือ $S = (A,B,C)$

$$\text{Pr (A จะได้รับเลือกเข้ามาในกลุ่มตัวอย่าง S)} = \frac{3}{N}$$

$$\text{Pr (B จะได้รับเลือกเข้ามาในกลุ่มตัวอย่าง S)} = \frac{2}{N-1}$$

$$\text{Pr (C จะได้รับเลือกเข้ามาในกลุ่มตัวอย่าง S)} = \frac{1}{N-2}$$

ดังนั้น

$$\text{Pr (กลุ่มตัวอย่าง S จะได้รับเลือก)} = \frac{3}{N} \cdot \frac{2}{N-1} \cdot \frac{1}{N-2}$$

$$= \frac{6}{N(N-1)(N-2)}$$

$$= \frac{3!}{N(N-1)(N-2)}$$

$$= \frac{3!}{N!/(N-3)!} = \frac{1}{\binom{N}{3}}$$

เมื่อขยายความสู่กรณีของกลุ่มตัวอย่างขนาด n จะพบว่า

$$\text{Pr (ได้กลุ่มตัวอย่างใด ๆ ขนาด n)} = \frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \cdots \frac{1}{N-(n-1)}$$

$$= \frac{n!}{N(N-1)(N-2)\ldots(N-(n-1))}$$

$$= \frac{n!}{N!(N-n)!} = \frac{1}{\binom{N}{n}}$$

2.2 การเลือกตัวอย่าง

ในทางปฏิบัตินั้นไม่มีเหตุจำเป็นใด ๆ ที่จะต้องแจกแจงว่าตัวอย่างที่พึงเป็นไปได้ทั้งหมดซึ่งมีอยู่ทั้งสิ้น $\binom{N}{n}$ ชุดนั้นคือใคร เพราะเป็นเรื่องที่ฟุ่มเฟือยเกินไป เวลาปฏิบัติจริงให้ยึดถือเอาความเข้าใจว่า ถ้าใช้วิธี SRS แล้วจะมีตัวอย่างที่เป็นไปได้อยู่ $\binom{N}{n}$ ชุด ซึ่งแตกต่างไม่ซ้ำกันและหน่วยตัวอย่างในแต่ละกลุ่มก็ไม่ซ้ำกันหรือหน่วยสำรวจแต่ละหน่วยมีโอกาสได้รับเลือกเท่ากัน การเลือกตัวอย่างจึงกระทำโดยสร้างกรอบตัวอย่างขึ้นมาก่อน โดยให้เลขที่หรือเลขลำดับแก่หน่วยสำรวจทุกหน่วยจาก 1 ถึง N พร้อมทั้งชื่อและที่อยู่หรือ ID (Identification) อื่น ๆ ที่จะเป็น จากนั้นใช้ตารางเลขสุ่มเลือกตัวอย่างขึ้นมา n จำนวน หลังจากปรับแก้เลขสุ่มทั้ง n จำนวนนั้นให้เหมาะสมแล้วตัวเลขสุ่มดังกล่าวจะถูกใช้แสดงกลุ่มตัวอย่างตามวิธี SRS เลขสุ่มใด ตรงกับลำดับที่ของหน่วยสำรวจใด ก็ถือว่าหน่วยสำรวจนั้นคือหน่วยตัวอย่าง แต่ถ้าไม่ต้องการใช้ตารางเลขสุ่ม เราอาจใช้ Random Number Generator ซึ่งจะสร้างขึ้นโดยวิธีการต่าง ๆ มากมายหลายวิธีด้วยเครื่องคอมพิวเตอร์ก็ได้ การเลือกตัวอย่างโดยวิธีนี้หน่วยสำรวจแต่ละหน่วยจะมีโอกาสได้รับเลือกเท่าเทียมกัน

การใช้ตารางเลขสุ่มและการปรับแก้ค่าของเลขสุ่ม

ก. การใช้ตารางเลขสุ่ม

การใช้ตารางเลขสุ่มให้ปฏิบัติดังนี้คือ

(1) ใช้ดินสอหรือวัตถุปลายแหลมชี้ลงที่เลขใดเลขหนึ่งในตารางเลขสุ่มเพื่อหาตัวเลขที่ใช้แทนเลขหน้าของตารางเลขสุ่มในกรณีที่มีตารางเลขสุ่มหลายหน้า การชี้ต้องเป็นไปโดยสุ่ม อาจทำโดยการหลับตาหรือหันหน้าไปทางอื่นแล้วจึงชี้ปลายแหลมของดินสอลงไปบนตาราง ตัวเลขที่ได้จะแสดงหน้าของตารางเลขสุ่ม เช่น ถ้าเสนอเลขสุ่มไว้ 6 หน้า ถ้าเลขที่ได้เป็นเลข 3 แปลว่าเราจะใช้ตารางเลขสุ่มหน้า 3 ถ้าได้เลข 6 เราจะใช้ตารางเลขสุ่มหน้า 6 ดังนี้เป็นต้น แต่ถ้าเลขที่ได้เป็นเลขที่เกินหน้าที่มีอยู่ ให้เอาจำนวนหน้าหักลบออกเสียก่อนผลต่างที่ได้คือเลขหน้าที่ต้องการ เช่นชี้ได้เลข 8 หน้าที่ต้องการคือหน้า $8 - 6 = 2$ หรืออีกวิธีหนึ่งคือใช้จำนวนหน้าไปหารเลขที่สุ่มได้เศษเหลือของผลหารก็คือเลขหน้าที่ต้องการ เช่นชี้ได้ 8 ผลหาร คือ $\frac{8}{6}$ เศษเท่ากับ 2 หน้าที่ 2 คือหน้าของเลขสุ่มที่ต้องการ สำหรับกรณีที่มีตารางเลขสุ่มเพียงหน้าเดียวไม่จำเป็นต้องดำเนินการขั้นนี้ให้ดำเนินการขั้นที่ 2 ต่อไป

(2) ใช้ดินสอชี้เลข 2 จำนวนเพื่อหา Random Row และ Random Column ในขั้นนี้เราจะใช้ดินสอชี้เลข 2 จำนวน จำนวนแรกจะแสดงที่ตั้งของแถว จำนวนที่ 2 แสดงที่ตั้งของสดมภ์ ตัวเลขในตารางเลขสุ่ม ณ. ตำแหน่งแถวและสดมภ์ที่ระบุจะถูกใช้เป็นตัว Random Start ต่อไป

เช่นชี้ครั้งแรกเป็นเลข 9 ชี้ครั้งที่ 2 ได้เลข 12 แสดงว่าเลขสุ่มที่ได้อยู่ ณ. ตำแหน่ง (9, 12) หรือนัยหนึ่งตัวเลขที่อยู่ในตำแหน่งที่แถว 9 กับสดมภ์ 12 ตัดกัน คือตัว Random Start หรือตัวเริ่มต้น

(3) ดำเนินการเลือกตัวเลขในตารางเลขสุ่มมา n ชุด โดยเริ่มต้นจากตัว Random-Start แต่ก่อนอื่นจำเป็นต้องทราบเสียก่อนว่าจะเลือกเลขกี่หลัก การเลือกเลขสุ่มเราจะเลือกหลักใดขึ้นอยู่กับหลักเกณฑ์ของกลุ่มประชากร ถ้าขนาดของประชากรเป็นเลข 2 หลัก เลขสุ่มที่เลือกมาใช้ก็จะต้องเป็นเลข 2 หลัก ถ้าขนาดของประชากรเป็นเลข 3 หลัก เลขสุ่ม

ที่ใช้ก็ต้องเป็นเลข 3 หลัก เช่น N = 1200 เลขสุ่มที่จะเลือกต้องเป็นเลข 4 หลัก ดังนี้เป็นต้น การเลือกให้เริ่มจากตัว Random Start โดยเลือกมาทีละจำนวนแต่ละจำนวนเป็นเลขหลักเดียวกับหลักเลขของขนาดประชากร จนกระทั่งครบ n ชุด โดยที่ผู้เลือกจะเลือกตัวเลขโดยไล่ไปทางซ้าย ขวา บน หรือล่างตัว Random Start ก็ได้ เมื่อสุดแถวหรือสุดสดมภ์จะเลี้ยวโดยวนไปทางซ้ายหรือขวาก็ได้ ซึ่งในทางปฏิบัตินิยมเลือกเลขสุ่มไว้มากกว่า n ชุด ทั้งนี้เป็นการเลือกเผื่อไว้สำหรับกรณีที่พบว่ามีเลขสุ่มซ้ำชุดกัน

ข. การปรับค่าของเลขสุ่ม

เมื่อได้เลขสุ่มครบตามจำนวนที่ต้องการ ขั้นต่อไปก็คือการปรับตัวเลขให้เหมาะสม ความเหมาะสมที่กล่าวถึงก็คือค่าของเลขสุ่มจะต้องมีค่าไม่เกินค่าของ N ทั้งนี้เพราะ N ในที่นี้ก็คือลำดับที่ลำดับสุดท้ายของหน่วยสำรวจในกรอบตัวอย่าง ค่าของเลขสุ่มที่เกิน N จึงเป็นลำดับที่ของหน่วยที่อยู่นอกขอบเขตของกรอบตัวอย่าง เช่น ประชากรประกอบด้วยหน่วยสำรวจ 400 หน่วย (N = 400) คือหน่วยสำรวจที่ 1, 2, 3, ..., 400 ในกรณีนี้ 400 เป็นเลข 3 หลักเลขสุ่มที่เลือกขึ้นมาจะเป็นเลข 3 หลักเช่นกัน อาจเป็น 022, 004, 277, 946, 999, 877, 426 ,... จะเห็นว่าเลขสุ่ม 946, 999, 877, 426 มีค่าเกิน 400 ถ้าพูดตามความหมายของกรอบตัวอย่างและการสุ่มตัวอย่างก็หมายความว่าการสุ่มตัวอย่างครั้งนี้เราจะใช้หน่วยสำรวจลำดับที่ 22, 4, 277, 946, 999, 877, 426 ,... เป็นตัวอย่าง แต่หน่วยที่ 946, 999, 877, 426 ไม่มีอยู่ในกรอบตัวอย่าง ตัวเลขเหล่านี้จึงเป็นตัวเลขที่ไม่เหมาะสม เราจำเป็นต้องแก้ไขหรือปรับค่าเสียใหม่ให้สอดคล้องกับเงื่อนไขของกรอบตัวอย่างเสียก่อน นั่นก็คือปรับค่าของตัวเลขเหล่านี้ให้มีค่าไม่เกิน 400

วิธีปรับค่าของเลขสุ่มกระทำได้ 2 วิธีคือ

วิธีที่ 1 ใช้วิธีหักลบ โดยนำ N หรือพหุคูณของ N ไปหักลบออกจากค่าของเลขสุ่ม ผลต่างที่ได้รับคือ ลำดับที่ในกรอบตัวอย่างที่ต้องการ นั่นก็คือ หน่วยตัวอย่าง = ค่าของเลขสุ่ม − k เท่าของ N

เช่น เลขสุ่มคือ 946 เลขสุ่มที่ปรับค่าแล้วคือ 946 − 2(400) = 146

เลขสุ่มคือ 999 เลขสุ่มที่ปรับค่าแล้วคือ 999 − 2(400) = 199

เลขสุ่มคือ 877 เลขสุ่มที่ปรับค่าแล้วคือ 877 − 2(400) = 77

เลขสุ่มคือ 426 เลขสุ่มที่ปรับค่าแล้วคือ 426 − 1(400) = 26

ดังนั้น ในกรณีนี้ เลขสุ่มก็คือ 022, 004, 146, 199, 077, 026 ,... หมายความว่าเราจะใช้หน่วยสำรวจในลำดับที่ 22, 4, 146, 199, 77, 26 ,... ของกรอบตัวอย่างเป็นกลุ่มตัวอย่าง

วิธีที่ 2 ให้วิธีหาเศษเหลือ โดยนำ N ไปหารค่าของเลขสุ่ม เศษของการหารก็คือลำดับที่ในกรอบตัวอย่างตามต้องการ นั่นคือ

$$\text{หน่วยตัวอย่าง} = \frac{\text{ค่าของเลขสุ่ม}}{N} = a + r \text{ เมื่อ } a \text{ คือผลหาร } r \text{ คือเศษเหลือ}$$

เช่น เลขสุ่มคือ 946 ดังนั้น 946/400 = 2 เศษ 146 เลขสุ่มที่ปรับค่าแล้วก็คือ 146

เลขสุ่มคือ 999 ดังนั้น 999/400 = 2 เศษ 199 เลขสุ่มที่ปรับค่าแล้วคือ 199

เลขสุ่มคือ 877 ดังนั้น 877/400 = 2 เศษ 77 เลขสุ่มที่ปรับค่าแล้วก็คือ 77

เลขสุ่มคือ 426 ดังนั้น 426/400 = 1 เศษ 26 เลขสุ่มที่ปรับค่าแล้วก็คือ 26

ถ้าจะสังเกตให้ดีจะพบว่าวิธีการทั้งสองคือวิธีหักลบและวิธีการใช้เศษเหลือ (Remainder) ก็คือวิธีเดียวกัน

ในกรณีที่เลขสุ่มหรือเลขสุ่มที่ปรับค่าแล้วเป็นเลขจำนวนเดียวกัน จำเป็นต้องตัดเลขที่ซ้ำกันนั้นทิ้งไป แล้วใช้เลขสุ่มค่าอื่นซึ่งเลือกเผื่อไว้แทน ชุดของเลขสุ่มที่ปรับค่าแล้วทั้ง n ชุด จะเป็นชุดของตัวเลขที่แสดงลำดับของหน่วยสำรวจในกรอบตัวอย่าง ซึ่งจะใช้เป็นตัวอย่างตามวิธี SRS ต่อไป

อนึ่งวิธีการหาเลขสุ่มโดยวิธีการเลือกจากตารางเลขสุ่มอาจล่าช้าสิ้นเปลืองกำลังแรงงานมากเกินไปและไม่ทันสมัย โดยเฉพาะในรายที่เกี่ยวข้องกับกลุ่มประชากรขนาดใหญ่ซึ่งทำให้ต้องปรับค่าเลขที่สุ่มที่เป็นเลขหลายหลัก วิธีการที่นิยมใช้กันมากก็คือการสร้างเลขสุ่มโดย Random Number Generator โดยใช้โปรแกรมคอมพิวเตอร์ซึ่งเป็นโปรแกรมสั้น ๆ วิธีนี้สะดวกรวดเร็วและไม่จำเป็นต้องเสียเวลาในการปรับค่าแต่ประการใด

นิยามและสัญลักษณ์

1. x_i ; = i = 1, 2, ..., N = ค่าของหน่วยสำรวจในกลุ่มประชากร (Value of Population Characteristics)

2. x_i ; i = 1, 2, ..., n = ค่าของหน่วยสำรวจในกลุ่มตัวอย่าง

ขอให้สังเกตว่าในที่นี้ใช้ x ตัวพิมพ์เล็กในความหมายของทั้งค่าของหน่วยสำรวจทั้งในกลุ่มประชากรและในกลุ่มตัวอย่าง ที่เป็นเช่นนี้เพราะหน่วยสำรวจในตัวอย่างก็คือหน่วยสำรวจในกลุ่มประชากรที่ถูกเลือกขึ้นมา ค่าของคุณลักษณะของประชากร (Characteristics) จึงเป็นค่าเดียวกัน ดัชนีที่ช่วยชี้ให้เห็นว่าค่าใดคือค่าของกลุ่มประชากรและค่าใดคือค่าของกลุ่มตัวอย่าง คือ N และ n โดยที่ n คือขนาดตัวอย่าง และ N คือขนาดของประชากร (Population Size)

3. $\bar{x} = \frac{1}{n} \sum_{i}^{N} x_i$ = มัชฌิมเลขคณิตของกลุ่มประชากร (Population Mean)

4. $\bar{x} = \frac{1}{n} \sum_{i}^{n} x_i$ = มัชฌิมเลขคณิตของกลุ่มตัวอย่าง (Sample Mean)

5. $T = \sum_{i}^{N} x_i$ = ยอดรวมของ (ที่ได้รับจาก) กลุ่มประชากร (Population Total)

6. $R = \frac{\bar{Y}}{\bar{X}} = \sum_{i}^{N} y_i \,/\, \sum_{i}^{N} x_i$ = อัตราส่วนของ (ที่ได้รับจาก) กลุ่มประชากร (Population Ratio) โดยที่ y_i ; i = 1, 2, 3, ..., N คือ ค่าของคุณลักษณะทางประชากรลักษณะที่ 1 x_i ; i = 1, 2, ..., N คือค่าของคุณลักษณะทางประชากรลักษณะที่ 2

7. $r = \frac{\bar{y}}{\bar{x}} = \sum_{i}^{n} y_i \,/\, \sum_{j}^{n} x_i$ = อัตราส่วนของ (ที่ได้รับจาก) กลุ่มตัวอย่าง (Sample Ratio)

8. $P = \sum_{i}^{N} \frac{x_i}{N}$ โดยที่ x_i = 0, 1; i = 1, 2, ..., N

= สัดส่วนของ (ที่ได้รับจาก) กลุ่มประชากร (Population Proportion)

9. $p = \sum_{i}^{n} \frac{x_i}{n}$ โดยที่ $x_i = 0, 1; i = 1, 2, \ldots, n$

= สัดส่วนของ (ที่ได้รับจาก) กลุ่มตัวอย่าง (Sample Proportion)

10. $S^2 = \frac{1}{N-1} \sum_{i}^{N} (x_i - \overline{X})^2$ ค่าความแปรปรวนของ (ที่ได้รับจาก) กลุ่มประชากร (Population Variance)

11. $s^2 = \frac{1}{n-1} \sum_{i}^{n} (x_i - \bar{x})^2$ = ค่าความแปรปรวนของ (ที่ได้รับจาก) กลุ่มตัวอย่าง (Sample Variance)

2.4 การประมาณค่าของพารามิเตอร์

พารามิเตอร์ (Parameter) คือค่าของคุณลักษณะทางประชากร (Characteristics) ของกลุ่มประชากรซึ่งเราให้ความสนใจและต้องการทราบ ทั้งนี้เพื่อประโยชน์แก่การศึกษาวิจัยและการบริหารงานของฝ่ายรัฐและเอกชน แต่โดยปกติพารามิเตอร์เป็นค่าที่ไม่อาจทราบได้ เว้นแต่จะมีการสำมะโน พารามิเตอร์อาจหมายถึงอะไรก็ได้แล้วแต่ความสนใจของบุคคลหรือหน่วยงานซึ่งอาจมีได้แตกต่างกันไป เช่น รัฐบาลต้องการทราบจำนวนครอบครัวที่มีรายได้ต่ำกว่าเดือนละ 1,000 บาท จำนวนบุคคลผู้ว่างงาน สัดส่วนของบุคคลวัยทำงาน อัตราส่วนของบุคคลผู้เป็นภาระ (Dependent Ratio) ขนาดของครอบครัวโดยเฉลี่ย จำนวนโคกระบือในประเทศ ฯลฯ ห้างหุ้นส่วนหรือบริษัทเอกชนอาจต้องการทราบปริมาณความต้องการสินค้าและผลิตภัณฑ์ของตน รสนิยมของผู้บริโภค ทัศนคติของผู้บริโภคต่อคุณภาพและราคาของผลิตภัณฑ์ ฯลฯ สิ่งเหล่านี้เรียกว่าคุณลักษณะทางประชากร ค่าที่เป็นตัวเลขของคุณลักษณะทางประชากรเรียกว่าพารามิเตอร์ ซึ่งเราจะทราบได้ก็ต่อเมื่อมีการสำมะโนเท่านั้น แต่การทำสำมะโนเป็นเรื่องที่ต้องพิจารณาให้รอบคอบเพราะเสียค่าใช้จ่ายสูงและเสียเวลามาก ข้อมูลที่ได้รับอาจล่าช้าไม่ทันกันกับความต้องการใช้ นอกจากนี้การที่จะจัดทำสำมะโนนั้น ควรจะได้พิจารณาถึงสิ่งต่อไปนี้ด้วยคือ

1. เป้าหมายของงาน หมายความว่างานนั้นเราต้องการอะไร ข้อเท็จจริงหรือผลสรุป หรือว่าต้องการทราบรายละเอียดจากทุกหน่วยของประชากร

2. งานสำมะโนจะให้ข้อสนเทศที่ชัดแจ้งและถูกต้องสมบูรณ์หรือไม่ แน่ใจหรือไม่ว่าจะควบคุมมิให้มีความบกพร่องเกิดขึ้นในกระบวนการสำมะโนและความถูกต้องของข้อมูล

3. งานนั้นพอจะมีหนทางให้ใช้วิธีสำรวจด้วยตัวอย่างขึ้นใช้แทนหรือไม่

เมื่อพิจารณาโดยรอบคอบแล้วจะพบว่าวิธีสำมะโนเป็นสิ่งที่ทำได้ยาก อุปสรรคที่สำคัญก็คือวิธีสำมะโนสิ้นเปลืองงบประมาณมากและเสียเวลารวบรวมข้อมูลและประมวลผลนานเกินกว่าที่จะรอได้ หมายความว่าวิธีสำมะโนทำให้ได้รับข้อมูลไม่ทันกับความต้องการ หรือกว่าจะได้รับข้อมูลมาใช้ประโยชน์ สถานการณ์อาจเปลี่ยนแปลงไปทำให้ข้อมูลที่ได้รับมาแล้วล้าสมัย

ดังนั้นวิธีสำรวจด้วยตัวอย่างจึงเป็นวิธีที่ได้รับความนิยมอย่างกว้างขวางในปัจจุบัน เพราะนอกจากจะสะดวกรวดเร็วและสิ้นเปลืองงบประมาณน้อยกว่าแล้ว วิธีสำรวจด้วยตัวอย่างยังเป็นวิธีที่ทำให้ได้รับข้อมูลที่ลึกซึ้งหลายแง่หลายมุมและควบคุมมิให้ตกหล่นหรือมีความผิดพลาดได้ดีกว่าวิธีสำมะโนเพราะเกี่ยวข้องกับหน่วยสำรวจจำนวนน้อยกว่า การควบคุมงานสนามกระทำได้ใกล้ชิดกว่าข้อมูลที่ได้รับจะประมวลกันเข้าเป็นตัวประมาณค่า (Estimates) ใช้สำหรับประมาณค่าพารามิเตอร์ที่สนใจ ปัญหาที่หลายคนสงสัยก็คือตัวประมาณค่าเหล่านี้ ใช้ประมาณค่าพารามิเตอร์ได้จริงหรือไม่ ถูกต้องแม่นยำเพียงใด?

สิ่งที่ทุกคนจะต้องยอมรับกันไว้ในขั้นต้นนี้ก็คือผลลัพธ์ของงานประมวลผลจากกลุ่มตัวอย่างนั้นมิใช่ค่าจริงของพารามิเตอร์หากเป็นเพียงค่าประมาณ ซึ่งเราถือว่านั่นเป็นวิธีที่ดีที่สุดแล้วที่จะทำให้ทราบข้อเท็จจริงเกี่ยวกับพารามิเตอร์แม้จะไม่ถูกต้องร้อยเปอร์เซ็นต์ ทั้งนี้เนื่องจากการสำมะโน เป็นสิ่งที่ปฏิบัติได้ยากโดยเฉพาะในกลุ่มประชากรขนาดใหญ่ ตัวประมาณค่าที่จะได้ผลใกล้เคียงกับค่าจริงของพารามิเตอร์เพียงใดขึ้นอยู่กับวิธีการสำรวจด้วยตัวอย่าง ถ้าพยายามเลือกตัวอย่างด้วยความระมัดระวังโดยใช้กลุ่มตัวอย่างกระจายไปทั่วกลุ่มประชากรได้และเป็นการเลือกโดยสุ่มมิได้เจาะจงเลือกหน่วยสำรวจเองตามอัธยาศัย กลุ่มตัวอย่างก็จะเป็นตัวแทนที่ดีของกลุ่มประชากร ค่าประมาณที่ได้ก็จะมีค่าใกล้เคียงกับ

ค่าจริงของพารามิเตอร์ ในทางทฤษฎีเราสามารถพิสูจน์ให้เห็นความจริงข้อนี้ได้และสามารถวัดระดับความถูกต้องแม่นยำของตัวประมาณค่าดังกล่าวได้ด้วย และอย่างน้อยแม้ค่าประมาณอาจคลาดเคลื่อนไปจากค่าจริงได้ เราก็สามารถบอกได้ว่าค่าจริงนั้น ๆ ควรปรากฏอยู่ในช่วงใดด้วยความมั่นใจเพียงใด

พารามิเตอร์ที่น่าสนใจ โดยทั่วไปมี 4 ตัวคือ

1. มัชฌิมเลขคณิตหรือค่าเฉลี่ย (Population Mean)

โดยปกติใช้สัญลักษณ์ μ แต่ในที่นี้ใช้ $\overline{X}$ โดยที่ $\overline{X} = \frac{1}{N} \sum_{i}^{N} x_i$ คือค่าเฉลี่ยแสดงอัตราถัวเฉลี่ยต่อหน่วยสำรวจ หน่วยสำรวจอาจหมายถึงบุคคล ครัวเรือน ห้องเรียน ชุมรุมครัวเรือน หมู่บ้าน ตำบล อำเภอ จังหวัด ฯลฯ แตกต่างกันไปตามวัตถุประสงค์ของงาน ดังนั้นค่าเฉลี่ยจึงอาจเป็นอัตราถัวเฉลี่ยต่อบุคคล ต่อครอบครัว ต่อห้องเรียน ต่อชุมรุมครัวเรือน ฯลฯ เช่น รายได้ต่อบุคคล จำนวนบุตรต่อ 1 ครอบครัว จำนวนคนต่อ 1 ชุมรุมครัวเรือน จำนวนนักเรียนต่อ 1 หมู่บ้านเป็นต้น

2. ยอดรวม (Population Total) ใช้สัญลักษณ์ T หรือ X ในที่นี้ใช้ T โดยที่ $T = \sum_{i}^{N} x_i$ ยอดรวมเป็นค่าแสดงผลรวมของคุณลักษณะทางประชากรที่เราสนใจ เช่นยอดรวม-ของจำนวนโคนมที่เลี้ยงในท้องที่สำรวจ ยอดรวมจำนวนผู้ว่างงาน ยอดรวมผลผลิตใบยาสูบของท้องที่เพาะปลูก ยอดรวมจำนวนบ้านที่ไม่มีเลขที่ ฯลฯ

3. อัตราส่วน (Ratio) ใช้สัญลักษณ์ R โดยที่ $R = \sum_{i}^{N} x_i / \sum_{i}^{N} y_i = \frac{\overline{X}}{\overline{Y}}$ โดยที่ $\sum_{i}^{N} x_i$ และ $\sum_{i}^{N} y_i$ แสดงผลยอดรวมของคุณลักษณะของประชากร 2 ลักษณะ ในที่นี้ทั้ง x และ y ต่างก็เป็นตัวแปรที่เราสนใจทั้งคู่ ในทางปฏิบัติเราอาจใช้ $R = \frac{\overline{Y}}{\overline{X}}$ หรือ $R = \frac{\overline{X}}{\overline{Y}}$ ก็ได้ ทั้งนี้ขึ้นอยู่กับความสนใจเป็นสำคัญ เช่น R = อัตราส่วนของครูต่อนักเรียน เราอาจใช้ R = อัตราส่วนของนักเรียนต่อครูก็ได้ และหน่วยที่ใช้จะใช้เป็นต่อหนึ่ง ต่อร้อย ต่อพัน ฯลฯ ก็ได้ ทั้งนี้เพ่งเล็งที่ความหมายของจุดทศนิยมเป็นสำคัญ ถ้าตัวแปรที่สนใจเป็นหน่วยที่ต้องแสดงด้วยเลขจำนวนเต็ม เช่น คน สัตว์ สิ่งของบางชนิด จุดทศนิยมจะแสดง

ความหมายที่ไม่แจ่มชัดเราจำเป็นต้องแปลงให้เป็นเลขเต็มหลัก อาจเป็นหลักร้อย หลักพัน หรืออื่น ๆ ตัวอย่างเช่น ในตำบลหนึ่งมีประชากร 15,000 คน ในจำนวนนี้เป็นเพศชาย 6,200 คน เป็นเพศหญิง 8,800 คน อัตราส่วนระหว่างเพศ = จำนวนประชากรเป็นเพศชาย/จำนวนประชากรที่เป็นเพศหญิง = $\frac{6200}{8800}$ = .705 = $\frac{.705}{1}$ หมายความว่าในตำบลดังกล่าวมีประชากรชาย .705 คนต่อหญิง 1 คน เพื่อให้การตีความหมายชัดเจนขึ้นเนื่องจากทศนิยมในที่นี้ไม่มีความหมายที่แจ้งชัด เราจึงคูณด้วย 1,000 (การคูณหมายถึงคูณทั้งเศษและส่วนด้วย 1000) ค่า R จึงกลายเป็น $\frac{705}{1000}$ ซึ่งหมายความว่าในตำบลดังกล่าวจะมีประชากรเพศชาย 705 คนต่อประชากรหญิง 1000 คน ดังนี้เป็นต้น

ค่าอัตราส่วนที่น่าสนใจอาจเป็นเรื่องเกี่ยวกับ อัตราส่วนระหว่างเพศ อัตราส่วนระหว่างเด็กต่อสตรี อัตราส่วนที่เป็นภาระ (จำนวนผู้ที่ทำงานเลี้ยงชีพเองไม่ได้ต่อจำนวนผู้ที่ทำงานได้) อัตราส่วนค่าใช้จ่ายด้านอาหารต่อเด็กและคนชรา อัตราส่วนพื้นที่ทำกินต่อสมาชิกครอบครัว อัตราส่วนรายได้พิเศษต่อรายได้ประจำ อัตราส่วนผู้คัดค้านต่อผู้สนับสนุน ฯลฯ

ขอให้สังเกตว่าอัตราส่วนคือค่าที่แสดงการเปรียบเทียบระหว่างคุณลักษณะทางประชากร 2 ลักษณะ อัตราส่วนจึงเป็นเรื่องเปรียบเทียบกันในระหว่างตัวแปรคู่ที่เราสนใจคู่หนึ่ง โดยที่ตัวแปรทั้งสองเป็นตัวแปรที่แสดงคุณลักษณะทางประชากรของกลุ่มประชากรกลุ่มเดียวกัน

4. สัดส่วนหรือร้อยละ (Proportion หรือ Percentage)

โดยปกติใช้สัญลักษณ์ P หรือ 100P% โดยที่ $P = \frac{1}{N}\sum_{i}^{N} x_i$ สัดส่วนเป็นค่าแสดงอัตราเปรียบเทียบของตัวแปรตัวใดตัวหนึ่งต่อยอดรวมทั้งหมด เช่น สัดส่วนของเด็กต่อประชากรทั้งหมด สัดส่วนของรายได้จากการทำนาต่อรายได้ทั้งหมด สัดส่วนของเงินลงทุนเพื่อการอุตสาหกรรมต่อเงินทุนทั้งหมด ฯลฯ ขอให้เป็นที่สังเกตไว้ว่า คำว่าสัดส่วน ก็คือการจำแนกกลุ่มประชากรออกเป็นส่วน ๆ ตามลักษณะของตัวแปรที่เราสนใจ แล้วคำนวณดูว่าแต่ละส่วนคิดเป็นร้อยละเท่าไร (ถ้าคิดเทียบเป็นร้อย) หรือคิดเป็นสัดส่วนเท่าไรต่อทั้งหมด โดยนัยนี้กลุ่มประชากรจึงสามารถจำแนกได้เป็นอย่างน้อย 2 ส่วน ถ้าประชากรจำแนก

เป็น 2 ส่วนการคำนวณให้ยึดถือวิธีการของการกระจายแบบทวินาม (Binomial Distribution) ในลักษณะเห็นด้วย-ไม่เห็นด้วย ใช่-ไม่ใช่ ถูก-ผิด ยอมรับ-ไม่ยอมรับ หรือที่ใช้กันโดยทั่วไปว่า Success-Failure (S-F) นั่นเอง แต่ถ้าประชากรจำแนกออกเป็นส่วน ๆ ตามจำนวนตัวแปร มากกว่า 2 ส่วน การคำนวณให้ยึดถือวิธีการของการกระจายแบบพหุนาม (Multinomial Distribution) เรื่องเหล่านี้นักศึกษาจะได้พบในลำดับต่อไป อย่างไรก็ตามทางการศึกษา เรื่องสัดส่วนมิได้หมายความว่าข้อมูลที่ใช้จะต้องเป็นข้อมูลเชิงคุณภาพเสมอไปอาจเป็น ข้อมูลเชิงปริมาณก็ได้ ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของงานเป็นสำคัญ

ความหมายของสัดส่วนและร้อยละแตกต่างกันเพียงเฉพาะหน่วยที่ใช้เทียบกล่าวคือ สัดส่วนใช้หน่วยเทียบเป็น 1 เช่น สัดส่วนของนักศึกษาสายตาสั้นเท่ากับ .02 หมายความว่า ในบรรดานักศึกษา 1 คน จะมีนักศึกษาสายตาสั้น .02 คน และด้วยเหตุที่การตีความหมาย โดยนัยนี้ไม่แจ้งชัด เราจึงใช้หน่วยเป็น 100 คือคูณตลอดด้วย 100 กลายเป็น $P = \frac{2}{100} = 2\%$ หรือตีความใหม่ได้ชัดเจนขึ้นว่า ในบรรดานักศึกษา 100 คน จะมีนักศึกษาสายตาสั้น ปะปนอยู่ 2 คน ดังนี้ ถ้านักศึกษาไม่ใช้ความสังเกตให้ดีอาจนำไปสู่การสับสนกับเรื่อง อัตราส่วนโดยเฉพาะในกรณีใช้หน่วยเทียบต่อ 100 เช่นกัน เพื่อป้องกันความสับสนจึงขอย้ำ ไว้ในที่นี้อีกครั้งว่า อัตราส่วนเป็นเรื่องของการเปรียบเทียบกันระหว่างตัวแปรที่เราสนใจ 2 ตัว สัดส่วนเป็นเรื่องของการเปรียบเทียบตัวหนึ่งตัวใดเพียงตัวเดียวกับประชากรทั้งกลุ่ม

2.4.1 การประมาณค่าเฉลี่ยและยอดรวมของกลุ่มประชากร (Estimation of Population Mean and Population Total)

ทฤษฎี 2.1 ถ้าดำเนินการสุ่มตัวอย่างโดยเปิดโอกาสให้หน่วยสำรวจทุกหน่วย มีโอกาสได้รับเลือกมาเป็นตัวอย่างโดยเท่าเทียมกัน (SRS) แล้ว ค่าเฉลี่ย (มัชฌิมเลขคณิต) ของกลุ่มตัวอย่างจะเป็นตัวประมาณค่าที่ปราศจากอคติ (Unbiased Estimator) ของค่าเฉลี่ย จริงของกลุ่มประชากร (Population Mean)

นั่นคือ $E(\bar{x}) = \bar{X}$

พิสูจน์ $E(\bar{x}) = \sum_{s}^{\binom{N}{n}} \bar{x}_s \cdot Pr(\bar{x}) = \sum_{s}^{\binom{N}{n}} \bar{x}_s \cdot \frac{1}{\binom{N}{n}}$

$$= \frac{1}{\binom{N}{n}} \cdot \sum_{s}^{\binom{N}{n}} \{(x_1+x_2+\ldots+x_n)/n\}_s$$

$$= \frac{1}{n} \cdot \frac{1}{\binom{N}{n}} \cdot \sum_{s}^{\binom{N}{n}} (x_1+x_2+\ldots+x_n)_s$$

แต่เนื่องจากหน่วยสำรวจใด ๆ สามารถปรากฏในกลุ่มตัวอย่างต่าง ๆ ได้ $\binom{N-1}{n-1}$ กลุ่ม[1] ดังนั้น

$$\sum_{s}^{\binom{N}{n}} (x_1+x_2+\ldots+x_n)_s = \binom{N-1}{n-1} x_1 + \binom{N-1}{n-1} x_2 + \ldots + \binom{N-1}{n-1} x_N$$

$$\therefore \; E(\bar{x}) = \frac{1}{n} \cdot \frac{1}{\binom{N}{n}} \cdot \binom{N-1}{n-1} \cdot (x_1+x_2+\ldots+x_N)$$

$$= \frac{1}{n} \cdot \frac{(N-n)!\, n!}{N!} \cdot \frac{(N-1)!}{(N-n)!\,(n-1)!} \sum_{i}^{N} x_i$$

$$= \frac{1}{N} \sum_{i}^{N} x_i$$

$$= \bar{X}$$

---

[1] อ่านคำอธิบายเพิ่มเติม 1

คำอธิบายเพิ่มเติม 1

ให้ประชากรมีขนาด $N = 4$ ประกอบไปด้วย (x_1, x_2, x_3, x_4) เราต้องการเลือกตัวแทนมาเพียง $n = 3$ หน่วย

ดังนั้นจำนวนกลุ่มตัวอย่างที่เป็นไปได้จึงมีอยู่ทั้งสิ้น $\binom{N}{n} = \binom{4}{3} = 4$ ชุด คือ

$S_1 = (x_1, x_2, x_3)$, $S_2 = (x_1, x_2, x_4)$, $S_3 = (x_1, x_3, x_4)$ และ $S_4 = (x_2, x_3, x_4)$

1. จะเห็นได้ว่าโอกาสที่จะเลือกใช้ตัวอย่างชุดใดชุดหนึ่ง (หรือโอกาสที่จะได้ค่าเฉลี่ยของตัวอย่างชุดใดชุดหนึ่ง) จึงเท่ากับ $\dfrac{1}{\binom{N}{n}} = \dfrac{1}{4}$

2. เมื่อพิจารณาหน่วยสำรวจแต่ละหน่วยจะเห็นว่าแต่ละหน่วยปรากฏในกลุ่มตัวอย่างต่าง ๆ ได้ $= 3 = \binom{3}{2} = \binom{4-1}{3-1}$ ชุด คือ

x_1 ปรากฏได้ทั้งใน S_1, S_2 และ S_3

x_2 ปรากฏได้ทั้งใน S_1, S_2 และ S_4

x_3 ปรากฏได้ทั้งใน S_1, S_3 และ S_4

x_4 ปรากฏได้ทั้งใน S_2, S_3 และ S_4

ดังนั้น ในกรณีกลุ่มประชากรขนาด N และกลุ่มตัวอย่างขนาด n หน่วยสำรวจหนึ่ง ๆ ย่อมมีสิทธิปรากฏขึ้นในกลุ่มตัวอย่างถึง $\binom{N-1}{n-1}$ ชุด

3. พิจารณา $\sum\limits_{s}^{\binom{N}{n}} (x_1 + x_2 + \ldots + x_n)_s$ สำหรับกรณีข้างต้นจะพบว่า

$$\sum_{s}^{4} \left(\sum_{i}^{3} x_i \right)_s = (x_1 + x_2 + x_3) + (x_1 + x_2 + x_4) + (x_1 + x_3 + x_4) + (x_2 + x_3 + x_4)$$

$$= 3x_1 + 3x_2 + 3x_3 + 3x_4$$

$$= \binom{4-1}{3-1}x_1 + \binom{4-1}{3-1}x_2 + \binom{4-1}{3-1}x_3 + \binom{4-1}{3-1}x_4$$

$$= \binom{4-1}{3-1}(x_1 + x_2 + x_3 + x_4)$$

ดังนั้นในกรณีของกลุ่มประชากรขนาด N และกลุ่มตัวอย่างขนาด n จึงสามารถสรุปได้ว่า

$$\sum_{s}^{\binom{N}{n}} (x_1 + x_2 + \ldots + x_n)_s = \binom{N-1}{n-1}(x_1 + x_2 + \ldots + x_N)$$

ทฤษฎีนี้ชี้ให้เห็นว่าถ้ามีการเลือกกลุ่มตัวอย่างมาใช้อย่างถูกต้อง คือ ทุกหน่วยได้รับการเลือกมาโดยสุ่มหรือมีโอกาสได้รับการเลือกอย่างเท่าเทียมกัน มิใช่เลือกอย่างมีอคติ คือจงใจเลือกหน่วยสำรวจโดยยึดถือเอาความสะดวก เช่น สัมภาษณ์เฉพาะคนที่รู้จักหรืออยู่ใกล้บ้าน หรือไม่สร้างกรอบตัวอย่างที่สมบูรณ์ขึ้นมาก่อน (การไม่สร้างกรอบตัวอย่างที่ถูกต้องสมบูรณ์จะทำให้หน่วยสำรวจบางหน่วยตกหล่น สูญเสียโอกาสที่จะได้รับการเลือก) แล้วค่าเฉลี่ยของกลุ่มตัวอย่างจะใช้เป็นตัวแทนของค่าเฉลี่ยของกลุ่มประชากรได้

ปัญหาต่อไปก็คือ จะเชื่อได้มากน้อยเพียงใดว่าค่าเฉลี่ย $\bar{x}$ สามารถประมาณค่า $\bar{X}$ ได้แม่นยำ เกี่ยวกับปัญหานี้เรามีเครื่องมือวัดความแม่นยำของ $\bar{x}$ ได้โดยใช้เครื่องมือที่เรียกว่า Standard Error หรือ Error Variance คือ $\sqrt{V(\bar{x})}$ ซึ่งนักศึกษาจะได้ศึกษาในลำดับต่อไป ในชั้นนี้เราควรจะได้ศึกษาวิธีประมาณค่ายอดรวมของกลุ่มประชากรเสียก่อน เพราะเป็นเรื่องที่ต่อเนื่องกันกับค่าเฉลี่ย ในทางปฏิบัตินอกจากเราจะสนใจค่าเฉลี่ยแล้วเราอาจสนใจยอดรวมด้วย เช่น ในการสำรวจภาวะเศรษฐกิจของชาวไร่ยาสูบอำเภอสันทราย จังหวัดเชียงใหม่ นอกจากเราจะสนใจว่าชาวไร่ผลิตใบยาสูบในฤดูการผลิตหนึ่ง ๆ เฉลี่ยครอบครัวละกี่กิโลกรัมแล้ว เรายังสนใจว่าผลผลิตยาสูบรวมของจังหวัดเชียงใหม่นั้นเป็น

ผลผลิตที่มาจากในเขตท้องที่อำเภอสันทรายกี่กิโลกรัม (ตัน) หรือคิดเป็นร้อยละเท่าไร หรือในการพิสูจน์อักษรจากต้นฉบับ เราอาจต้องการทราบว่าโดยถัวเฉลี่ยแล้วหน้าหนึ่ง ๆ มีคำผิดกี่คำ และทั้งเล่มมีคำผิดรวมทั้งสิ้นกี่คำ ดังนี้เป็นต้น

บทแทรก 2.1 เมื่อดำเนินการสุ่มตัวอย่างโดยใช้แผนสำรวจแบบ SRS แล้ว $\hat{T} = N\bar{x}$ จะเป็นตัวประมาณที่ปราศจากอคติของยอดรวมประชากร T โดยที่ $T = \sum_{i}^{N} x_i$

นั่นคือ $E(\hat{T}) = T$

พิสูจน์ จากทฤษฎี 1.1 เราพิสูจน์แล้วว่า $E(\bar{x}) = \bar{X}$

$$\text{ดังนั้น } E(\hat{T}) = E(N\bar{x}) = N\{E(\bar{x})\} = N\bar{X}$$

$$= N \cdot \sum_{i}^{N} x_i/N = \sum_{i}^{N} x_i = T$$

จะเห็นได้ว่าการประมาณค่ายอดรวมของกลุ่มประชากรนั้นกระทำได้โดยง่ายคือ คำนวณหาค่าเฉลี่ยของคุณลักษณะที่สนใจจากกลุ่มตัวอย่างเสียก่อน จากนั้นนำค่าเฉลี่ยดังกล่าวมาคูณกับ N ซึ่งเป็นขนาดของประชากร ผลลัพธ์ที่ได้จะเป็นค่าประมาณของค่ายอดรวมของคุณลักษณะทางประชากรของกลุ่มประชากรนั้น เช่นต้นฉบับ 200 หน้า สุ่มตัวอย่างมา 30 หน้าพบคำผิดถัวเฉลี่ยหน้าละ 10 คำ แสดงว่าต้นฉบับทั้งเล่มมีคำผิดทั้งสิ้น $N\bar{x} = 200 \times 10 = 2{,}000$ คำ เป็นต้น

ขอให้สังเกตว่ายอดรวมจากกลุ่มตัวอย่างคือ $Sn = (x_1 + x_2 + \ldots + x_n)$ มิใช่ตัวประมาณค่าที่ปราศจากอคติของ T ความจริงข้อนี้สามารถพิสูจน์ให้เห็นโดยง่ายดังนี้

$$E(Sn) = \sum_{j=1}^{\binom{N}{n}} (Sn)_j \Pr(Sn)_j$$

$$= \frac{1}{\binom{N}{n}} \sum_{j}^{\binom{N}{n}} (S_n)_j$$

$$= \frac{1}{\binom{N}{n}} \sum_{j}^{\binom{N}{n}} (x_1 + x_2 + \ldots + x_n)_j$$

$$= \frac{1}{\binom{N}{n}} \cdot \binom{N-1}{n-1} (x_1 + x_2 + \ldots + x_N)$$

$$= \frac{(N-n)!n! \ . \ (N-1)!}{N! \ (N-n)! \ (n-1)!} \sum_{i}^{N} x_i$$

$$= \frac{n}{N} \sum_{i}^{N} x_i \neq T$$

แสดงว่า $E(Sn) \neq T$ หรือ $(x_1 + x_2 + \ldots + x_n)$ มิใช่ตัวประมาณค่าที่ปราศจากอคติของ $T = (x_1 + x_2 + \ldots + x_N)$

แต่จากผลการพิสูจน์

$$E(Sn) = \frac{n}{N} \sum_{i}^{N} x_i$$

ถ้านำ $\frac{N}{n}$ คูณตลอดจะเห็นได้ว่า

$$\frac{N}{n} E(Sn) = \frac{N}{n} \cdot \frac{n}{N} \sum_{i}^{N} x_i$$

นั่นคือ $\frac{N}{n} E(Sn) = E\left(\frac{NSn}{n}\right) = \sum_{i}^{N} x_i = T$

แต่ $\frac{NSn}{n} = N(x_1 + x_2 + \ldots + x_n)/n = N\bar{x}$

ดังนั้น

$$\frac{N}{n} \cdot E(Sn) = E\left(\frac{NSn}{n}\right) = E(N\bar{x}) = T$$

หรือตัวประมาณค่าที่ควรใช้ประมาณค่ายอดรวมของกลุ่มประชากร

$$T = \sum_{i}^{N} x_i \text{ คือ } \hat{T} = N\bar{x}$$

ทฤษฎี 2.2 ถ้าใช้แผนสำรวจ SRS ในการเลือกกลุ่มตัวอย่างแล้ว ความแปรปรวนของตัวประมาณค่า (Variance of Estimate หรือ Error Variance) $\bar{x}$ และ $\hat{T}$ คือ

$$1.\ V(\bar{x}) = \frac{N-n}{N} \cdot \frac{S^2}{n} \text{ และ } 2.\ V(\hat{T}) = N^2 \frac{N-n}{N} \cdot \frac{S^2}{n}$$

โดยที่ $S^2 = \frac{1}{N-1} \sum_{i}^{N} (x_i - \bar{X})^2$, $\bar{X} = \frac{1}{N} \sum_{i}^{N} x_i$

พิสูจน์ $V(\bar{x}) = E(\bar{x} - \bar{X})^2$

$$= E\left\{ \frac{n}{n} (\bar{x} - \bar{X}) \right\}^2$$

$$= \frac{1}{n^2} E(n\bar{x} - n\bar{X})^2$$

$$= \frac{1}{n^2} E\left\{ n\left(\frac{x_1 + x_2 + \ldots + x_n}{n} \right) - \underbrace{(\bar{X} + \bar{X} + \ldots + \bar{X})}_{n \text{ ครั้ง}} \right\}^2$$

$$= \frac{1}{n^2} E\left\{ (x_1 + x_2 + \ldots + x_n) - (\bar{X} + \bar{X} + \ldots + \bar{X}) \right\}^2$$

$$= \frac{1}{n^2} E\left\{ (x_1 - X) + (x_2 - \bar{X}) + \ldots + (x_n - \bar{X}) \right\}^2$$

$$= \frac{1}{n^2} E\left\{ \sum_{i}^{n} (x_i - \bar{X})^2 + \sum_{i \neq j}^{n} \sum^{n} (x_i - \bar{X})(x_j - \bar{X}) \right\}^1$$

---

[1] ดูคำอธิบายเพิ่มเติม 2

$$= \frac{1}{n^2} \sum_{i}^{n} E(x_i - \overline{X})^2 + \frac{1}{n^2} \sum_{i \neq j}^{n} \sum^{n} E(x_i - \overline{X})\ (x_j - \overline{X}) \qquad \text{..........(1)}$$

พิจารณา $\frac{1}{n^2} \sum_{i}^{n} E(x_i - \overline{X})^2$ จะพบว่า

$$\frac{1}{n^2} \sum_{i}^{n} E(x_i - \overline{X})^2 = \frac{1}{n^2} \sum_{i}^{n} \left\{ \sum_{i}^{N} (x_i - \overline{X})^2 \frac{1}{N} \right\}$$

$$= \frac{1}{n^2} \sum_{i}^{n} \left\{ (N-1)S^2 \cdot \frac{1}{N} \right\}$$

$$= \frac{1}{n^2} \cdot \frac{N-1}{N} \sum_{i}^{n} S^2$$

$$= \frac{1}{n^2} \frac{N-1}{N} \cdot nS^2$$

$$= \frac{1}{n} \frac{N-1}{N} \cdot S^2 \qquad \text{..........(2)}$$

พิจารณา $\frac{1}{n^2} \sum_{i \neq j}^{n} \sum^{n} E(x_i - \overline{X})(x_j - \overline{X})$ จะพบว่า

$$\frac{1}{n^2} \sum_{i \neq j}^{n} \sum^{n} E(x_i - \overline{X})(x_j - \overline{X}) = \frac{1}{n^2} \sum_{i \neq j}^{n} \sum^{n} \left\{ \sum_{i \neq j}^{N} \sum^{N} (x_i - \overline{X})(x_j - \overline{X}) \cdot \frac{1}{N} \cdot \frac{1}{N-1} \right\}$$ [1]

[1] ดูคำอธิบายเพิ่มเติม 2

และพิจารณา $\{\sum_{i}^{N}(x_i-\overline{X})\}^2$ จะพบว่า

$$\{\sum_{i}^{N}(x_i-\overline{X})\}^2 = \{(x_1-\overline{X})+(x_2-\overline{X})+\ldots+(x_N-\overline{X})\}^2$$

$$= \sum_{i}^{N}(x_i-\overline{X})^2 + \sum_{i\neq j}^{N}\sum^{N}(x_i-\overline{X})(x_j-\overline{X})$$

ดังนั้น $\sum_{i\neq j}^{N}\sum^{N}(x_i-\overline{X})(x_j-\overline{X}) = \{\sum_{i}^{N}(x_i-\overline{X})\}^2 - \sum_{i}^{N}(x_i-\overline{X})^2$ [1]

$$= 0 - \sum_{i}^{N}(x_i-\overline{X})^2$$

$$= -(N-1)S^2$$

ดังนั้น $\frac{1}{n^2}\sum_{i\neq j}^{n}\sum^{n}E(x_i-\overline{X})(x_j-\overline{X}) = \frac{1}{n^2}\sum_{i\neq j}^{n}\sum^{n}\{\sum_{i\neq j}^{N}\sum^{N}(x_i-\overline{X})(x_j-\overline{X})\cdot\frac{1}{N}\cdot\frac{1}{N-1}\}$

$$= \frac{1}{n^2}\cdot\frac{1}{N}\cdot\frac{1}{N-1}\sum_{i\neq j}^{n}\sum^{n}\{-(N-1)S^2\}$$

$$= \frac{-(N-1)}{n^2N(N-1)}\sum_{i\neq j}^{n}\sum^{n}S^2$$

$$= \frac{-(N-1)n.(n-1)}{n^2N(N-1)}\cdot S^2 \text{ [2]} = \frac{n-1}{nN}\cdot S^2$$

---

1 $\sum_{i}^{N}(x_i-\overline{X}) = \sum_{i}^{N}x_i - N\overline{X} = \frac{N}{N}\sum_{i}^{N}x_i - N\overline{X} = N\overline{X} - N\overline{X} = 0$

2 ดูคำอธิบายเพิ่มเติม 2