บทที่ 5
ระเบียบวิธีการสุ่มตัวอย่าง

เราทราบดีว่าการอนุมานเชิงสถิติ คือการตัดสินใจหรือให้ข้อสรุปกับประชากรโดยใช้ข่าวสารจากส่วนหนึ่งของประชากร คือตัวอย่าง เช่นในการแนะนำสินค้าใหม่จะใช้วิธีแจกตัวอย่างให้ลองใช้ก่อน แม่ครัวจะชิมรสจากน้ำแกงเพียง 1 ช้อน เวลาเราซื้อส้ม จะขอลองชิมสักชิ้นหนึ่งก่อน หรือเวลาเปิดทีวี เราจะลองชมรายการนั้นสักครู่ ก่อนตัดสินใจว่าจะชมต่อไปหรือจะหมุนไปช่องอื่น เหล่านี้เป็นเรื่องการ "ลองดู" หรือการสุ่มตัวอย่างทั้งสิ้น แต่ตัวอย่างทางสถิติจะต้องมีระเบียบวิธีการที่รัดกุมขึ้น และต้องใช้ทฤษฎีความน่าจะเป็น เพื่อจะได้อนุมานค่าของประชากรได้ ตัวอย่างมี 2 ชนิด คือตัวอย่างสุ่ม (random sampling) และตัวอย่างไม่สุ่ม (nonrandom random) เราจะกล่าวถึงตัวอย่างแบบสุ่ม

1. ประชากร (Population)

ประชากรคือ กลุ่มของสมาชิกที่อยู่ในปัญหาที่สนใจ บางครั้งเรียกว่า Universe

ตัวอย่างเช่น

1. ผู้จัดการบริษัทรถเช่า สนใจศึกษาการใช้น้ำมันของรถให้เช่าของบริษัท ดังนี้ ประชากร คือ รถทุกคันที่ให้เช่า ของบริษัท และสมาชิกของประชากรคือรถแต่ละคัน

2. ผู้จัดการฝ่ายควบคุมคุณภาพต้องการทราบระดับคุณภาพของระบบการผลิตชิ้นส่วนอีเล็คโทรนิค เพื่อใช้กับเครื่องคอมพิวเตอร์ ดังนี้ ประชากรคือชิ้นส่วนอิเล็คโทรนิคทั้งหมดที่ผลิตจากระบบดังกล่าว และสมาชิกหรือหน่วยของประชากรคือชิ้นส่วนแต่ละชิ้น

ในตัวอย่างที่ 1 ประชากรเป็นแบบจำกัด (finite population) แต่ในตัวอย่างที่ 2 ประชากรเป็นแบบไม่จำกัด (infinite population)

ประชากรแบบจำกัด

ประชากรแบบจำกัด จะประกอบด้วยสมาชิกที่มีจำนวนจำกัด หรือนับถ้วน เช่น จำนวนรถของบริษัทให้เช่ารถ จำนวนผู้มีสิทธิออกเสียงเลือกผู้บริหาร จำนวนประชากรของประเทศ

ถ้าตัวแปรที่เราสนใจเป็นเชิงปริมาณ เช่น รายได้ต่อครัวเรือน อายุของบุคคล เรานับแทนตัวแปรซึ่งมีทั้งหมด N หน่วยในประชากร ด้วย $x_1, x_2, \ldots, x_N$ ส่วนค่าเฉลี่ยของ N จำนวน คือ μ หรือค่าเฉลี่ยของประชากร และ σ^2 คือความแปรปรวนของประชากร กล่าวโดยสรุป

ถ้า $x_1, x_2, \ldots, x_N$ คือค่าจากหน่วยของ N ประชากร ค่าเฉลี่ยของประชากร คือ

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

ความแปรปรวนของประชากร คือ

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x - \mu)^2}{N}$$

และ ค่าเบี่ยงเบนมาตรฐานของประชากร คือ

$$\sigma = \sqrt{\sigma^2}$$

ค่า μ, σ, σ^2 เรียกว่า **ค่าพารามิเตอร์ของประชากร**

ตัวอย่างที่ 1

คะแนนของนักศึกษาระดับปริญญาเอกซึ่งเพิ่งเข้าศึกษาในคณะสถิติประยุกต์ 5 คน

มีดังนี้

$$x_1 = 3.74 \quad x_2 = 3.89 \quad x_3 = 4.00 \quad x_4 = 3.68 \quad x_5 = 3.69$$

$$\mu = \frac{3.74 + 3.89 + 4.00 + 3.68 + 3.69}{5} = 3.80$$

$$\sigma^2 = \frac{(3.74 - 3.80)^2 + \ldots + (3.69 - 3.80)^2}{5} = .01564$$

$$\sigma = \sqrt{.01564} = .125$$

ข้อสังเกต

1. การหาค่าพารามิเตอร์ของประชากรแบบจำกัดโดยการแจงนับสมาชิกทุก ๆ หน่วย เช่นในตัวอย่างที่ 1 เป็นวิธีหนึ่ง
2. อีกวิธีหนึ่งสำหรับหาค่าพารามิเตอร์ของประชากร คือเมื่อประชากรมีขนาดใหญ่ (แต่จำกัด) จะใช้ทฤษฎีความน่าจะเป็น เช่น การสุ่มครัวเรือน จาก N ครัวเรือน โดยให้ทุกครัวเรือนมีโอกาสถูกเลือกเป็นตัวอย่างเท่ากัน คือ 1/N และให้ตัวแปรเชิงสุ่ม x แทนรายได้ของครัวเรือนที่สุ่มมา ดังนั้น x จะมีค่าที่เป็นได้ N ค่า คือ $x_1, x_2, \ldots, x_N$ ด้วยความน่าจะเป็นเท่ากันหมดทุกหน่วย คือ 1/N และจากนิยามการหาค่าคาดหมายของตัวแปรเชิงสุ่มแบบไม่ต่อเนื่อง จะได้ $E(x) = \sum_{i=1}^{N} x_i \left(\frac{1}{N}\right) = \sum_{i=1}^{N} X_i/N = \mu$

และความแปรปรวนจากนิยาม เมื่อ x เป็นตัวแปรเชิงสุ่มแบบไม่ต่อเนื่อง คือ

$$\sigma_X^2 = E(x - \mu)^2 = \Sigma(x - \mu)^2 f(x)$$
$$= \sum_{i=1}^{N}(x - \mu)^2 \left(\frac{1}{N}\right)$$
$$= \sum_{i=1}^{N}(x - \mu)^2/N = \sigma^2$$

นั่นคือ ในประชากรแบบจำกัด μ และ σ^2 คือ ค่าคาดหมาย และความแปรปรวนของตัวแปรเชิงสุ่ม ซึ่งมีความน่าจะเป็นเท่ากันที่จะถูกเลือกจากประชากร

ประชากรแบบไม่จำกัดจำนวน

ประชากรแบบไม่จำกัดจำนวน (infinite population) จะประกอบด้วยสมาชิกไม่จำกัดจำนวน โดยทั่ว ๆ ไป มักหมายถึงกระบวนการที่ประกอบด้วยสมาชิกที่เป็นผลจากการดำเนินการโดยไม่จำกัด แต่อยู่ภายใต้สภาวะที่คงที่ เช่นการผลิตชิ้นส่วนอีเล็คโทรนิค เครื่องจักรจะผลิตออกมาเรื่อย ๆ ดังนั้นการบรรยายลักษณะของประชากรแบบไม่จำกัด จึงต้องอยู่ในรูปการแจกแจงความน่าจะเป็น เช่น ในกระบวนการผลิตยางรถยนต์ ให้ X แทนจำนวนยางที่ต่ำกว่าคุณภาพ (ชำรุด) และสมมติว่ามีการแจกแจงความน่าจะเป็น ดังนี้

x	f (x)
0	.75
1	.10
2	.15
รวม	1.00

นั่นคือ มีโอกาสที่จะไม่ชำรุดเลย .75% หมายถึงถ้าผลิตภายใต้สภาวะเดิมนี้ต่อไปเรื่อย ๆ (กระทำซ้ำ ๆ กัน) จะมียางไม่ชำรุดอยู่ 75%

การหาค่าพารามิเตอร์ของประชากรแบบอนันต์ คือ

ค่าเฉลี่ย : $\mu = E(x)$

ความแปรปรวน : $\sigma^2 = \sigma^2(x) \quad \Sigma(x - \mu)^2 f(x)$

ค่าเบี่ยงเบนมาตรฐาน $\sigma = \sqrt{\sigma^2}$

ตัวอย่างที่ 2

จากเรื่องการผลิตยางรถ และ X คือจำนวนสินค้าชำรุดจากการหยิบมาแบบสุ่ม 2 เส้น และมีการแจกแจงความน่าจะเป็น ดังนี้

x	0	1	2
f (x)	.75	.10	.15

$$\mu = E(X) = \sum_{i=1}^{3} x\, f(x)$$

$$= 0\,(.75) + 1\,(.10) + 2\,(.15) = .40$$

$$\sigma^2 = \Sigma(x - \mu)^2\, f(x)$$

$$= (0 - .40)^2\,(.75) + (1 - .40)^2\,(.10) + (2 - .40)^2\,(.15)$$

$$= .54$$

ดังนั้น $= .73$

ตัวอย่างที่ 3 ให้ประชากรแบบไม่จำกัดจำนวนที่สนใจ คือ น้ำหนักของลิ่มโลหะจากเครื่องหลอม และให้แทนด้วยตัวแปร X สมมุติ X มีการแจกแจงความน่าจะเป็นแบบ N(520, 121) ดังนั้น ประชากรที่เราสนใจ คือ ประชากร แบบปกติ (N = Normal) ซึ่งมีค่าเฉลี่ยของประชากร μ = 520 ปอนด์ และความแปรปรวน σ^2 = 121 ปอนด์

ตัวอย่างที่ 4 ให้ประชากรแบบไม่จำกัดจำนวนที่อยู่ในความสนใจ คือ เวลาตอบสนอง X (เป็นนาที) ที่กองกำลังตำรวจ (191 หรือ 123) จะมาถึงสถานที่เกิดเหตุนับจากได้รับแจ้งเหตุร้าย และสมมุติว่า X มีการแจกแจงแบบเอ็กซโพเน็นเชียล ด้วย $\lambda = .2$ นั่นคือ ประชากรที่อยู่ในความสนใจคือประชากรแบบเอ็กซ์โพเน็นเชียล ซึ่งมีค่าเฉลี่ย $\mu = 5$ นาที และความแปรปรวน $\sigma^2 = 25$

หมายเหตุ ถ้า X มีการแจกแจงแบบเอ็กซ์โพเน็นเชียล

$$E(X) = \frac{1}{\lambda} \quad \text{และ} \quad \sigma^2(x) = \frac{1}{\lambda^2}$$

2. การสำมะโนและการสุ่มตัวอย่าง (Censuses and Samples)

วิธีการสำมะโน (Census)

วิธีการสำมะโนของประชากรแบบจำกัดก็คือการสำรวจ (survey) **ซึ่งรวมหน่วยสมาชิกทุกหน่วยของประชากรนั้น**

ในบางกรณี เป็นการง่ายที่จะใช้วิธีสำมะโนประชากร ดังกรณีต่อไปนี้

1. สถานที่ทำงานแห่งหนึ่งมีพนักงานทั้งหมด 50 คน ถ้าฝ่ายบริหารต้องการศึกษาความพอใจของพนักงานต่อระบบค่าตอบแทนระบบใหม่ เนื่องจากเป็นประชากรขนาดเล็ก จึงเป็นการง่ายที่จะรับฟังข้อคิดเห็นจากพนักงานทุกคน ดังนั้น วิธีสำมะโนจึงเหมาะสม

2. องค์การของรัฐต้องการทราบจำนวนนักศึกษาที่กำลังศึกษาอยู่ในมหาวิทยาลัย ในปี 2526 เนื่องจากข้อมูลนี้มีอยู่พร้อมแล้ว และอยู่ในความควบคุมของทบวงมหาวิทยาลัย ซึ่งทบวงมหาวิทยาลัย จะได้ข้อมูลโดยการกำหนดให้ทุกมหาวิทยาลัยในสังกัดส่งจำนวนนักศึกษาของตนตามกำหนดระยะเวลา เช่นทุกภาคการศึกษา วิธีนี้ควรใช้วิธีสำมะโน เพราะไม่เป็นการยากที่จะสำรวจสมาชิก (มหาวิทยาลัย) ทุก ๆ หน่วย และตัวเลขที่ได้จะมีความแม่นยำ เหมาะสมกับการใช้อ้างอิง ในราชการส่วนอื่น ๆ ของรัฐ

ข้อสังเกต

การพยายามใช้วิธีสำมะโนประชากร โดยพยายามรวมสมาชิกทุกหน่วยในประชากร อาจไม่ประสบผลสำเร็จในทุกกรณี แม้การสำรวจจำนวนประชากรของสหรัฐอเมริกา และคานาดา ก็ได้พบภายหลังว่า มีจำนวนข้อผิดพลาดไม่น้อย แม้ว่าจะได้ใช้ความพยายามอย่างสูงที่จะครอบคลุมประชากรทุกหน่วย

การสำรวจด้วยตัวอย่าง

ถ้าประชากรแบบจำกัดมีขนาดโตเกินไป การใช้วิธีสำมะโนย่อมไม่สะดวก เพราะจะต้องเสียค่าใช้จ่ายและเวลามาก ดังนั้นจึงนิยมใช้วิธีการสำรวจจากตัวอย่าง

ตัวอย่าง คือ ส่วนหนึ่งของประชากรที่อยู่ในความสนใจและถูกคัดเลือกมาเพื่อหาข่าวสารของประชากรนั้น

รัฐบาลนิยมใช้การสำรวจจากตัวอย่างมาก เช่นการสำรวจสภาวะว่างงาน การสำรวจพื้นที่เพื่อการเกษตร และอุตสาหกรรม การสำรวจการท่องเที่ยว เป็นต้น ทางด้านธุรกิจก็นิยมใช้การสุ่มตัวอย่างเพื่อควบคุมคุณภาพสินค้าที่ผลิตได้ และการตรวจรับวัสดุที่ซื้อมาเพื่อการผลิต การสำรวจความนิยมของสินค้าของผู้ผลิต การสำรวจทัศนคติของพนักงาน และแม้แต่ในชีวิต

ประจำวันของเราเอง ก็ต้องใช้วิธีสุ่มตัวอย่าง เช่นการลองชิมลิ้นจี่ ก่อนตกลงซื้อ การลองใช้สินค้า 1 หน่วยก่อนตกลงซื้อจำนวนมาก การพลิกอ่านเพียงบางหน้าก่อนตกลงใจว่าจะอ่านทั้งเล่มหรือไม่ เป็นต้น

เหตุที่ต้องทำการสุ่มตัวอย่าง

1. การสุ่มตัวอย่างทำให้ได้ข้อมูลที่เชื่อถือได้ และมีประโยชน์ ในอัตราค่าใช้จ่ายต่ำกว่าการสำมะโน เช่นถ้าต้องการทราบอุปนิสัยการดูรายการโทรทัศน์ของพลเมืองในเมืองหนึ่ง ถ้าใช้สำมะโนโดยการแจงนับประชากรทุกหน่วย (พลเมืองทั้งหมดในเมืองนั้น) จะเสียค่าใช้จ่ายสูงมาก ข้อมูลจากตัวอย่างก็น่าจะเชื่อถือได้ และใช้ค่าใช้จ่ายต่ำกว่ามาก

2. ประหยัดเวลากว่าวิธีสำมะโน เพราะมีข้อมูลน้อยกว่า จึงเหมาะกับจุดประสงค์ถ้าต้องการข่าวสารโดยด่วน

3. บ่อยครั้งที่ข้อมูลจากตัวอย่างมีความถูกต้องแม่นยำกว่าการสำมะโน เนื่องจากขอบข่ายงานมีขนาดเล็กกว่า จึงควบคุมการวัดความผิดพลาดได้ดีกว่า และมีโอกาสที่จะเลือกพนักงานที่มีคุณภาพ เพราะไม่ต้องใช้พนักงานจำนวนมากเท่าวิธีสำมะโน

4. ถ้าต้องการข่าวสารโดยละเอียดจากผู้ตอบคำถาม วิธีสำรวจด้วยตัวอย่างจะเหมาะสมกว่า เพราะใช้ขนาดตัวอย่างเล็กกว่า เนื่องจากค่าใช้จ่ายต่อหน่วยสูง จึงใช้วิธีสำมะโนไม่ได้

5. การทดสอบสิ่งของบางอย่างเป็นการทำลายสิ่งของนั้น เช่นการศึกษาอายุการใช้งานของแบตเตอรี่ หรือหลอดไฟ เมื่อได้ข้อมูลแล้วก็ต้องทิ้งวัสดุนั้นด้วย จึงทำการสำรวจทุกหน่วยไม่ได้

สำหรับประชากรแบบอนันต์ เราจะได้ข่าวสารของประชากรแบบนี้ก็โดยข่าวสารจากตัวอย่างเท่านั้น

ชนิดของตัวอย่าง

ตัวอย่างมี 3 ชนิด คือ

1. ตัวอย่างที่มาจากการใช้ทฤษฎีความน่าจะเป็น เรียกว่า probability samples
2. ตัวอย่างจากดุลยพินิจของผู้สุ่มตัวอย่าง เรียกว่า judgement samples

3. ตัวอย่างที่ได้มาโดยถือความสะดวก เรียกว่า convenience samples

Probability Sample

คือตัวอย่างซึ่งเกิดจากการคัดเลือก จากประชากรโดยวิธีความน่าจะเป็น

การเลือกตัวอย่างจากประชากรโดยใช้ความน่าจะเป็น คือการเลือกตัวอย่างโดยให้สมาชิกทุกหน่วยมีสิทธิ หรือโอกาสเท่าเทียมกันที่จะปรากฏเป็นตัวอย่าง วิธีนี้มีข้อดี 2 ข้อ คือ

1. ข้อมูลที่ได้จากตัวอย่าง สามารถใช้วิธีการทางสถิติกำหนดความคลาดเคลื่อนจากการสุ่มตัวอย่างได้
2. เป็นการหลีกเลี่ยงความอคติ (biases) อันเกิดจากการใช้ดุลยพินิจการตัดสินใจของผู้สุ่มตัวอย่าง ตัวอย่างโดยใช้ความน่าจะเป็นมีประโยชน์ และสำคัญที่สุดในการใช้อ้างอิงเชิงสถิติ

Judgement Sample

เป็นตัวอย่างที่มีลักษณะตรงข้ามกับตัวอย่างความน่าจะเป็น **ตัวอย่างแบบนี้จะได้มาจากการตัดสินใจ หรือดุลยพินิจในการถูกเลือก "เป็นตัวแทน" ของสมาชิกของประชากรนั้น**

ตัวอย่างเช่นการกำหนด "ความเป็นตัวแทน" ของประชากร ของตลาดผงซักฟอกชนิดหนึ่ง ผู้จัดการอาจกำหนด จังหวัดที่มีลักษณะ "พื้นฐาน" (typical) เพียงไม่กี่จังหวัดเพื่อทดสอบผงซักฟอกสูตรใหม่ เช่น อาจเลือก กรุงเทพ เชียงใหม่ อุดรธานี นครศรีธรรมราช ชลบุรี พิษณุโลก สำหรับเป็นตลาดทดลอง ข้อมูลที่ได้จากตัวอย่างประเภทนี้จะใช้วิธีการทางสถิติหาความคลาดเคลื่อนจากการสุ่มตัวอย่างไม่ได้ บางครั้งการสุ่มตัวอย่างโดยวิธีนี้ให้ผลลัพธ์แย่มาก ถ้าหน่วยที่ตกเป็นตัวอย่าง ไม่มี ลักษณะ "ความเป็นตัวแทน" ที่ดีของประชากรนั้น

การสุ่มตัวอย่างแบบโควต้า (quota sample) เป็นลักษณะตัวอย่างชนิดหนึ่งของ judgement sample ด้วยวิธีกำหนดจำนวนตัวอย่างให้ผู้สัมภาษณ์ โดยกำหนดลักษณะร่วมของตัวอย่าง เช่น เพศ อายุ ระดับความรู้ รายได้ และท้องที่อยู่อาศัย เป็นต้น ส่วนการเลือกตัวอย่างให้เป็นหน้าที่ของผู้สัมภาษณ์ ขอแต่เพียงให้ได้ลักษณะที่กำหนด ครบตามโควต้าที่กำหนดให้ เท่านั้น ทั้งนี้โดยหวังว่า ผู้สัมภาษณ์จะใช้ดุลยพินิจที่ดีในการคัดเลือกตัวอย่างที่เหมาะสม แต่ในทางปฏิบัติ

ส่วนใหญ่พบว่าผู้สัมภาษณ์มักเลือกตัวอย่างจากสิ่งใกล้ตัว คือญาติพี่น้อง หรือจากสิ่งที่พอจะหาได้ (available) ในเวลาที่ออกสัมภาษณ์ เช่น พบแต่แม่บ้านซึ่งไม่ต้องทำงานนอกบ้าน ตัวอย่างแบบนี้จึงมีความอคติหรือความเอียงเฉ ซึ่งถ้าใช้วิธีความน่าจะเป็นจะไม่มีโอกาสเกิดขึ้น เช่น ถ้าประชากรที่อยู่ในความสนใจ คือผู้ซื้อรถรุ่นใหม่ของบริษัท บริษัทได้ส่งแบบสอบถามไปให้ลูกค้าที่ซื้อรถรุ่นใหม่ทั้งหมด แต่สมมุติว่ามีผู้ตอบรับหรือส่งแบบตอบถามกลับคืนมาเพียง 50% ถ้าผู้วิจัยตกลง (สรุป) ว่า ผู้ตอบรับ 50% นี้ มีลักษณะพื้นฐานของผู้ซื้อรถใหม่ทั้งหมด เช่น ความเป็นตัวแทนด้านอายุ รายได้ ที่อยู่อาศัย เป็นต้น ตัวอย่างนี้จึงเป็น "judgement sample" แม้ว่าเริ่มแรกจะใช้วิธีความน่าจะเป็น แต่เมื่อผู้วิจัยใช้เหตุผลตัดสินใจ เลือกผู้ซื้อ 50% นี้เป็น "ตัวแทน" ของผู้ซื้อทั้งหมด จึงกลายเป็นตัวอย่างแบบ judgement

Convenience Sample

ตัวอย่างแบบ convenience ไม่ใช่ตัวอย่างแบบความน่าจะเป็น และไม่เหมือนตัวอย่างแบบ judgement เพราะไม่มีความพยายามที่จะให้เกิด "ความเป็นตัวแทน" บางครั้งเรียกว่า "Chunk" คือ**แทนกลุ่มหนึ่ง จากประชากรซึ่งได้มาโดยความสะดวก** เช่นครูเลือกนักเรียนในชั้นเรียนของตนเพื่อศึกษาอิทธิพลของแรงจูงใจ ชั้นเรียนที่ครูสอนเป็นตัวอย่างแบบสะดวก (convenience sample) จะเห็นว่าครูไม่ได้คำนึงถึง "ความเป็นตัวแทน" ของประชากรที่อยู่ในความสนใจ อีกตัวอย่างคือ การเลือกตัวอย่างโดยอาศัยความสะดวก โดยเลือกจากญาติพี่น้องเพื่อนฝูงที่อยู่ใกล้ตัว จะเห็นว่าลักษณะตัวอย่างแบบ judgement และแบบ convenience คล้ายกันมาก ข้อแตกต่างคือ ตัวอย่างแบบ convenience ใช้ได้ สำหรับจุดประสงค์ที่จำกัด เพราะไม่มั่นใจว่าเป็นตัวแทนของประชากรที่อยู่ในความสนใจ

3. การสุ่มตัวอย่างแบบง่ายจากประชากรแบบจำกัด

(Simple Random Sampling From a Finite Population)

การสุ่มวิธีนี้เป็นวิธีหลักของตัวอย่างความน่าจะเป็น และมักเรียกสั้น ๆ ว่าตัวอย่างสุ่ม (random sample)

นิยาม

ตัวอย่างสุ่มแบบง่ายจากประชากรแบบจำกัด คือตัวอย่างที่เกิดจากวิธีการเลือกสรรโดยให้กลุ่มตัวอย่างที่เป็นไปได้ทุก ๆ กลุ่มในประชากรนั้นมีโอกาสโดยเท่ากัน ที่จะได้รับเลือกเป็นตัวอย่าง

เช่นจากตัวอย่างที่ 1 มีนักศึกษาปริญญาเอก 5 คน คือ A, B, C, D, E ถ้าจะเลือกมาเป็นตัวอย่าง 2 คน จะมีกลุ่มตัวอย่างขนาด 2 คนที่เป็นไปได้ทั้งหมดในประชากรนั้นอยู่ 10 กลุ่มตัวอย่าง คือ

A, B	A, D	B, C	B, E	C, E
A, C	A, E	B, D	C, D	D, E

ถ้าจะใช้การสุ่มตัวอย่างแบบง่ายเพื่อเลือกนักเรียนตัวอย่างมา 2 คน จะต้องใช้วิธีการที่ทุกกลุ่มตัวอย่างมีโอกาสเท่ากัน คือ .10 ที่จะถูกเลือกเป็นตัวอย่าง

ข้อสังเกต

การสุ่มตัวอย่างนักเรียน 2 คน จากประชากร 5 คน ในตัวอย่างนั้น เป็นการสุ่ม**แบบไม่มีการแทนที่** ซึ่งเปิดโอกาสให้สมาชิกแต่ละหน่วยตกเป็นตัวอย่างได้เพียงครั้งเดียว ส่วนการสุ่ม**แบบมีการแทนที่** จะเปิดโอกาสให้สมาชิกหน่วยเดิมมีโอกาสตกเป็นตัวอย่างได้มากกว่า 1 ครั้ง ซึ่งการสุ่มแบบมีการแทนที่นี้ไม่ค่อยนิยมใช้กับประชากรแบบจำกัด

การให้ได้มาซึ่งตัวอย่างแบบง่าย (การเลือกหรือหยิบตัวอย่าง)

การเลือกสมาชิก n หน่วย โดยการสุ่มแบบง่าย และไม่มีการแทนที่ จากประชากรแบบจำกัด N หน่วย จะมีขั้นตอน ดังนี้

1. เลือกหน่วยที่ 1 โดยให้ทุกหน่วยจาก N หน่วยมีโอกาสถูกเลือกเท่ากัน นั่นคือ มีโอกาส 1/N ที่จะถูกเลือกเป็นตัวอย่าง
2. เลือกหน่วยที่ 2 โดยให้ทุกหน่วยที่เหลือ คือ N - 1 หน่วยมีโอกาสเท่ากันที่จะถูกเลือก นั่นคือมีโอกาส 1/(N-1) ที่จะถูกเลือกเป็นตัวอย่าง

3. ทำซ้ำเช่นนี้กับหน่วยต่อไปเรื่อย ๆ จนได้ครบ n ตัวอย่าง

สิ่งสำคัญอีกอันหนึ่งของการสุ่มตัวอย่างแบบง่ายคือ ต้องมี **กรอบตัวอย่าง**

นิยาม

กรอบตัวอย่าง คือรายชื่อสมาชิกทุกหน่วยของประชากร

เช่นต้องการสุ่มนักศึกษาตัวอย่าง กรอบตัวอย่างคือรายชื่อนักศึกษารหัสต่าง ๆ ทุกรหัส ที่ฝ่ายทะเบียน (โดยเครื่องคอมพิวเตอร์) รวบรวมไว้

ถ้ากรอบตัวอย่างไม่สมบูรณ์ เช่นต้องการรายชื่อทนายความทั่วประเทศ ถ้าเราใช้รายชื่อจากสมาคมทนายความอาจไม่ครอบคลุมทนายทั่วประเทศ แต่เนื่องจากเป็นแหล่งข้อมูลที่ดีที่สุด จึงต้องใช้รายชื่อนี้ เช่นนี้จะเกิดประชากร 2 อย่างคือ **ประชากรเป้าหมาย** (target population) ในที่นี้คือ ทนายความทั่วประเทศ และ**ประชากรตัวอย่างสุ่ม** (sampled population) ในที่นี้คือทนายความที่เป็นสมาชิกของสมาคมทนายความ ในกรณีที่กรอบตัวอย่างของประชากร 2 แบบนี้แตกต่างกัน ควรใช้ความพยายามอย่างมากที่สุด ให้กรอบตัวอย่าง 2 อันนี้ ซ้ำซ้อนกันมากที่สุด จะได้ไม่มีข้อผิดพลาดมาก

การใช้ตารางเลขสุ่ม

ตาราง B-8 เป็นตัวอย่างของตารางเลขสุ่ม เพื่อใช้สำหรับเลือกตัวอย่างสุ่มแบบง่าย ตารางเลขสุ่มประกอบด้วยเลข 0, 1, 2, ..., 9 ซึ่งประกอบด้วยความน่าจะเป็นเท่ากัน คือ .10 จึงเปิดโอกาสให้เลือกใช้เลขจาก 00 ถึง 99 หรือ 000 - 999 หรือที่จำนวนมากกว่านี้ โดยเลขทุกจำนวนมีโอกาสปรากฏด้วยความน่าจะเป็นเท่ากัน เช่นต้องการสุ่มทนายความจากบัญชีรายชื่อซึ่งได้จากสมาคมทนายความ สมมุติมีสมาชิกอยู่ 950 คน รายชื่อ 950 คนนี้เป็นกรอบตัวอย่าง เราจะให้เลขที่ 001 จนถึง 950 เป็นเลข 3 หลัก แล้วเราจะเลือกตัวเลข 3 หลัก จากตารางเลขสุ่มคือตาราง B-8 สมมุติเราเริ่มต้นจาก บรรทัดที่ 101 เอาเลขจากมุมซ้าย 3 ตัว แล้วต่อมาบรรทัดที่ 2 มุมซ้ายอีก 3 ตัว และบรรทัดต่อไปเรื่อย ๆ ถ้าเลขใดซ้ำให้ตัดทิ้งไปจนครบตามต้องการ เลขที่ได้แทนหมายเลขของทนายความในบัญชีรายชื่อที่ใช้เป็นกรอบตัวอย่าง ดังนั้นทนายความ

ที่ตกเป็นตัวอย่างคือ หมายเลข 132, 212, 001, 605, 912 เรื่อย ๆ ไปจนครบจำนวนตามต้องการ

การใช้ตารางเลขสุ่ม เลือกตัวอย่าง

บรรทัด	(1) - (5)	ทนายความ	
101	13284	132	
102	21224	212	
103	99052	~~990~~	→ ตัดทิ้งไปเพราะในรายชื่อมีเพียง 950
104	00199	001	
105	60578	605	คน
106	91240	912	
107	97458	~~974~~	→ ตัดทิ้งไปเพราะในรายชื่อมีเพียง 950
108	35249	352	
109	38980	389	คน
110	10750	107	

Table B–8 Table of random digits

Line	*(1)–(5)*	*(6)–(10)*	*(11)–(15)*	*(16)–(20)*	*(21)–(25)*	*(26)–(30)*	*(31)–(35)*
101	13284	16834	74151	92027	24670	36665	00770
102	21224	00370	30420	03883	94648	89428	41583
103	99052	47887	81085	64933	66279	80432	65793
104	00199	50993	98603	38452	87890	94624	69721
105	60578	06483	28733	37867	07936	98710	98539
106	91240	18312	17441	01929	18163	69201	31211
107	97458	14229	12063	59611	32249	90466	33216
108	35249	38646	34475	72417	60514	69257	12489
109	38980	46600	11759	11900	46743	27860	77940
110	10750	52745	38749	87365	58959	53731	89295
111	36247	27850	73958	20673	37800	63835	71051
112	70994	66986	99744	72438	01174	42159	11392
113	99638	94702	11463	18148	81386	80431	90628
114	72055	15774	43857	99805	10419	76939	25993
115	24038	65541	85788	55835	38835	59399	13790
116	74976	14631	35908	28221	39470	91548	12854
117	35553	71628	70189	26436	63407	91178	90348
118	35676	12797	51434	82976	42010	26344	92920
119	74815	67523	72985	23183	02446	63594	98924
120	45246	88048	65173	50989	91060	89894	36036

Table B–8 Table of random digits

Line	*(1)–(5)*	*(6)–(10)*	*(11)–(15)*	*(16)–(20)*	*(21)–(25)*	*(26)–(30)*	*(31)–(35)*
121	76509	47069	86378	41797	11910	49672	88575
122	19689	90332	04315	21358	97248	11188	39062
123	42751	35318	97513	61537	54955	08159	00337
124	11946	22681	45045	13964	57517	59419	58045
125	96518	48688	20996	11090	48396	57177	83867
126	35726	58643	76869	84622	39098	36083	72505
127	39737	42750	48968	70536	84864	64952	38404
128	97025	66492	56177	04049	80312	48028	26408
129	62814	08075	09788	56350	76787	51591	54509
130	25578	22950	15227	83291	41737	59599	96191
131	68763	69576	88991	49662	46704	63362	56625
132	17900	00813	64361	60725	88974	61005	99709
133	71944	60227	63551	71109	05624	43836	58254
134	54684	93691	85132	64399	29182	44324	14491
135	25946	27623	11258	65204	52832	50880	22273
136	01353	39318	44961	44972	91766	90262	56073
137	99083	88191	27662	99113	57174	35571	99884
138	52021	45406	37945	75234	24327	86978	22644
139	78755	47744	43776	83098	03225	14281	83637
140	25282	69106	59180	16257	22810	43609	12224
141	11959	94202	02743	86847	79725	51811	12998
142	11644	13792	98190	01424	30078	28197	55583
143	06307	97912	68110	59812	95448	43244	31262
144	76285	75714	89585	99296	52640	46518	55486
145	55322	07598	39600	60866	63007	20007	66819
146	78017	90928	90220	92503	83375	26986	74399
147	44768	43342	20696	26331	43140	69744	82928
148	25100	19336	14605	86603	51680	97678	24261
149	83612	46623	62876	85197	07824	91392	58317
150	41347	81666	82961	60413	71020	83658	02415

SOURCE: Excerpt from *Table of 105,000 Random Decimal Digits*. Interstate Commerce Commission, Bureau of Transport Economics and Statistics, May 1949.

TEXT REFERENCE: This table is discussed on pp. 185–187.

ข้อสังเกต

1. มีหลายวิธีสำหรับเลือกตัวเลขจากตารางเลขสุ่ม โดยมีหลักการว่าต้องกำหนดหลักการล่วงหน้าว่าจะเริ่มต้นตรงไหน และการหยิบตัวเลขต้องทำแบบมี**ระบบ**เดียวกัน เช่นต้องการเลข 3 ตัว จะอ่านตามบรรทัดจากซ้ายไปขวาจนจบบรรทัดแล้วขึ้นบรรทัดใหม่ เหมือนอ่านหนังสือ แล้ว

ตัดตัวเลขทีละ 3 ตัว ไปเรื่อย ๆ หรือจะอ่านลงมาในคอลัมน์เดียวกัน แบบใดก็ได้ถ้าทำแบบเดียวกันตลอด

2. บางครั้งสมาชิกหรือหน่วยของประชากรมีเลขที่อยู่แล้ว เช่น ใบเสร็จรับเงิน ใบรับ-ส่งสินค้า รหัสนักศึกษา เราใช้ตัวเลขเหล่านี้แทนหน่วยตัวอย่างได้เลย ถ้าเลขเหล่านั้นไม่มีการซ้ำซ้อนกัน

3. บางครั้งอาจไม่ต้องให้เลขที่กับรายชื่อในกรอบตัวอย่าง เช่นต้องการสุ่มพนักงานตัวอย่าง และมีประวัติการทำงานเก็บเข้าแฟ้มไว้อย่างดี สมมุติมีพนักงาน 8,500 คน แฟ้มหนึ่งเก็บบันทึก 100 คน จึงต้องใช้ 85 แฟ้ม ถ้าได้เลขสุ่ม 0117 หมายความว่าอยู่ในแฟ้มที่ 1 และคนที่ 17 ถ้าได้ 0417 แสดงว่าต้องอยู่ในแฟ้มที่ 4 คนที่ 17 เป็นต้น

4. การสุ่มตัวอย่างแบบง่ายจากประชากรแบบไม่จำกัดจำนวน (Simple Random Sampling From Infinite Population)

เมื่อประชากรเป็นแบบไม่จำกัดจำนวน ก็มีวิธีเดียวที่จะทราบข่าวสารจากประชากร คือจากตัวอย่าง เช่นสินค้าที่ผลิตจากเครื่องจักรมีจำนวนมากนับไม่ถ้วน เราไม่ทราบว่า N เป็นเท่าใด เพราะเครื่องทำการผลิตทุกวัน ดังนั้น สินค้า (ผลิตภัณฑ์) จากกระบวนการผลิตก็คือตัวอย่างสุ่มนั่นเอง

นิยาม

ตัวแปร $x_1, x_2, \ldots, x_n$ คือตัวแปรเชิงสุ่ม ซึ่งเกิดจากกระบวนการเพื่อให้ได้ตัวอย่างสุ่มแบบง่ายจากประชากรแบบไม่จำกัดจำนวน ถ้า

1. ตัวแปร $x_1, x_2, \ldots, x_n$ มีการแจกแจงความน่าจะเป็นแบบเดียวกัน
2. ตัวแปร $x_1, x_2, \ldots, x_n$ เป็นอิสระ

ข้อสังเกต

บางครั้งตัวอย่างจากประชากรแบบไม่จำกัด คือ **ประชากร** เช่นชิ้นส่วนอีเล็กโทรนิค**สำหรับ**เครื่องคอมพิวเตอร์ที่ผลิตได้ใน 1 สัปดาห์ จะ**แทนตัวอย่าง**จากประชากรแบบไม่จำกัด

ซึ่งเกี่ยวกับกระบวนการผลิต แต่โรงงานส่งชิ้นส่วนนี้ไปให้ผู้ผลิตเครื่องคอมพิวเตอร์ ชิ้นส่วนเหล่านี้จะกลายเป็น**ประชากร**ของผู้ผลิตคอมพิวเตอร์ และผู้ผลิตจะต้องมีแผนตรวจรับสินค้าโดยการสุ่มตัวอย่างจำนวนหนึ่งมาตรวจสอบคุณภาพ

5. ค่าสถิติที่ได้จากตัวอย่าง (Sample Statistics)

เมื่อเลือกตัวอย่างแบบสุ่มมา n จำนวนแล้ว เราจะวัดหรือหาค่าลักษณะที่สนใจและแทนด้วยตัวแปร $x_1, x_2, \ldots, x_n$ จากข้อมูลนี้เราจะได้ค่าสถิติหลายอย่างจากตัวอย่าง นั่นคือเราเรียกค่าที่วัดจากตัวอย่างว่าค่าสถิติตัวอย่าง หรือเรียกสั้น ๆ ว่าค่าสถิติ เพื่อให้แตกต่างกับค่าที่วัดจากประชากรซึ่งเราเรียกว่า ค่าพารามิเตอร์ ค่าสถิติเบื้องต้นที่ได้จากตัวอย่าง คือ

ค่าเฉลี่ยจากตัวอย่าง :

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

ความแปรปรวนของตัวอย่าง :

$$S^2 = \frac{\sum_{i=1}^{N} (x-\bar{x})^2}{n-1}$$

และ

ค่าเบี่ยงเบนมาตรฐาน :

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x-\bar{x})^2}{n-1}} = \sqrt{S^2}$$

นั่นคือ เราใช้ข่าวสารจากตัวอย่าง ประมาณค่าประชากร ดังนี้

ประมาณค่า μ ด้วย $\bar{x}$

ประมาณค่า σ^2 ด้วย S^2

และ ประมาณค่า σ ด้วย S

ดังนั้นช่วงเชื่อมั่นของ μ คือ

$$\bar{X} \pm Z_{\alpha/2}\, S/\sqrt{n}$$

ในบทนี้เราจะพูดเฉพาะกรณีที่ $n \geq 30$ และ N เป็นค่าโตมาก แต่เป็นประชากรแบบจำกัด และ n เป็นค่าเล็กมากเมื่อเทียบกับ N ดังนั้น สัดส่วนของตัวอย่างสุ่ม คือ n/N จึงเป็นค่าเล็ก และไม่ควรสูงกว่า 5% มิฉะนั้นจะต้องปรับค่าด้วย finite population correction factor (fpc = $\frac{N-n}{N-1}$) และการสุ่มตัวอย่างในบทนี้ทั้งบทจะยกตัวอย่างเฉพาะการสุ่มแบบไม่มีการแทนที่ เพราะในทางปฏิบัติ การสุ่มแบบนี้เป็นที่นิยม เพราะปิดโอกาสไม่ให้หน่วยที่ซ้ำกันตกเป็นตัวอย่างซึ่งจะทำให้ได้ข่าวสารซ้ำกับค่าเดิม จึงขาดประโยชน์

นอกจากนี้ยังมี พารามิเตอร์อีกตัวหนึ่งที่น่าสนใจ คือ π ซึ่งคือสัดส่วนของลักษณะที่อยู่ในความสนใจซึ่งปรากฏในประชากร

$$\pi = \frac{y_1 + y_2 \ldots + y_N}{N}\ ; \quad \begin{matrix} y_1 = 0 \text{ ถ้าเป็น F} \\ y = 1 \text{ ถ้าเป็น S} \end{matrix}$$

และค่าประมาณของ π คือ

$$p = \frac{y_1 + y_2 + \ldots + y_n}{n}\ ; \quad \begin{matrix} y = 0 \text{ ถ้าเกิด F} \\ y = 1 \text{ ถ้าเกิด S} \end{matrix}$$

และช่วงเชื่อมั่นของ π คือ

$$p \pm Z_{2/2} \sqrt{\frac{pq}{n}}$$

นอกจากการประมาณค่าเฉลี่ย μ และสัดส่วน π แล้ว บางครั้งเรายังสนใจ ประมาณ **ยอดรวม**

ให้ T แทนยอดรวม ถ้าทราบ $\bar{x}$

$T = N\bar{x}$ ในเมื่อ x เป็นตัวแปรแบบต่อเนื่องจากประชากรแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ($X \sim N(\mu, \sigma^2)$

แต่ถ้าทราบ p, $p = \frac{\Sigma y_i}{n}$; y = 0 ถ้าเป็น F, y = 1 ถ้าเป็น S

$T = Np$ ในเมื่อ y มีการแจกแจงแบบทวินาม ด้วยค่าเฉลี่ย $n\pi$ และความแปรปรวน $n\pi(1-\pi)$

$$(y \sim b\,(n\,p\,,\ n\,p\,(1-\pi)$$

ดังนั้นช่วงเชื่อมั่นของ T เมื่อ $x \sim n\,(\mu, \sigma^2)$ คือ

$$N \times (\bar{x} \pm Z_{\alpha/2}\, S/\sqrt{n})$$

$$= N\bar{x} \pm NZ_{\alpha/2}\, S/\sqrt{n}$$

และ ช่วงเชื่อมั่นของ T เมื่อ $y \sim b\,(n\,\pi,\ n\pi\,(1-\pi)$ คือ

$$N \times (p \pm Z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

$$= Np \pm NZ_{\alpha/2} \sqrt{\frac{pq}{n}}$$

การสุ่มตัวอย่างแบบมีระบบ (Systematic random sampling)

การสุ่มตัวอย่างแบบมีระบบเป็นลักษณะหนึ่งของ convenient sample design มีวิธีการดังนี้

1. ต้องให้เลขที่กับหน่วยตัวอย่างในกรอบตัวอย่าง จาก 1, 2,..., N
2. ถ้าต้องการสุ่มตัวอย่างขนาด n ต้องหาค่า sampling interval = K = N/n
3. หาหน่วยเริ่มต้นแบบสุ่ม เรียกว่า random start หรือ R $R \leq k$
4. ดังนั้น หน่วยแรกที่ใช้เป็นตัวอย่างคือ R

   หน่วยที่ 2 คือ R + I

   หน่วยที่ 3 คือ R + 2I

   .

   .

   .

   หน่วยสุดท้ายคือ n คือ R + (n-1) I

   ด้วยวิธีนี้จะได้ตัวอย่างจากบัญชีรายชื่อเป็นหน่วยที่ห่างกันด้วยระยะ k หน่วยตลอด จำนวน n หน่วย

ผลของการสุ่มแบบนี้ ไม่ใช่ simple random เนื่องจากหน่วยต่าง ๆ ที่ได้รับเลือกเป็นตัวอย่าง

ไม่เป็นอิสระกัน และมีเพียงหน่วยกลุ่มแรกคือหน่วยที่ 1 ถึง k เท่านั้น ที่มีโอกาสถูกเลือกแบบสุ่มตอนหาค่า R แต่เมื่อ R ถูกกำหนดแล้ว ตัวอย่างที่เหลือจะถูกกำหนดทันทีว่า จะต้องห่างกันด้วยช่วง k หน่วย

ถ้าหน่วยตัวอย่างในกรอบตัวอย่างมีลักษณะแบบเรียงลำดับ หรือมีลักษณะเป็นวัฏจักร ไม่ควรใช้การสุ่มตัวอย่างแบบมีระบบ เพราะอาจได้ตัวอย่างลักษณะเดียวซ้ำ ๆ กันหมด เมื่อ k ไปพ้องกับรอบของวัฏจักร เช่น การเก็บบัตรประวัติพนักงาน ถ้าใบที่ 10, 20, 30 แทนหัวหน้ากลุ่ม และถ้าเราได้ k = 10 ถ้า R = 0 หรือ 10, เราจะได้ตัวอย่างทั้งหมดเป็นหัวหน้ากลุ่ม ถ้า $1 \le R \le 9$ เราจะได้กลุ่มตัวอย่างที่ไม่มีหัวหน้ากลุ่มเลย ซึ่งตัวอย่างทั้ง 2 ชนิดนี้ ไม่ดี

ข้อดีของการสุ่มแบบมีระบบ คือความสะดวก และควรใช้เมื่อแน่ใจว่าไม่มีลักษณะการเปลี่ยนแปลงแบบมีระบบในประชากร (กรอบตัวอย่าง) ผลที่ได้จากการสุ่มแบบมีระบบจะใกล้เคียงกับการสุ่มแบบง่าย จึงใช้สูตรคำนวณค่าประมาณต่าง ๆ ของการสุ่มแบบง่าย

แบบฝึกหัด

5.1 สมมุติว่าสุ่มผู้มีอายุเกิน 18 ปี ในจังหวัดกรุงเทพฯ มา 400 คน คำถามหนึ่งที่สัมภาษณ์คือ "ท่านทำงานสัปดาห์ละ 35 ชั่วโมงขึ้นไป ในปีที่แล้วเป็นจำนวนกี่สัปดาห์?
สมมุติ ได้ $\bar{X} = 30$, $S = 10.5$, ให้ประชากรผู้ใหญ่จังหวัดกรุงเทพมี 5 ล้านคน

ก) จงประมาณค่าเฉลี่ยของตัวแปร x สำหรับประชากรทั้งจังหวัด และสร้างช่วงเชื่อมั่น 95%

ข) จงประมาณจำนวนรวมของสัปดาห์ที่ประชากรจังหวัดกรุงเทพฯ ทำงานเกินสัปดาห์ละ 35 ชั่วโมง และสร้างช่วงเชื่อมั่น 95%

ค) จงอธิบายความหมายของผลที่ได้ในข้อ (ก) และ (ข)

5.2 จากข้อ 5.1 ถ้าประชากรผู้ใหญ่ในจังหวัดกรุงเทพมี 5 ล้านคน และจากตัวอย่างที่สุ่มมา 400 คน เป็นชาย 170 คน และหญิง 230 คน (สุ่มแบบง่าย) และมีข้อมูลสรุป ดังนี้

	$\bar{X}$	S
ชาย	40	2.7
หญิง	22.6	7.6

ก) จงสร้างช่วงเชื่อมั่น 95% ของจำนวนสัปดาห์การทำงานโดยเฉลี่ยของประชากรชายในกรุงเทพ ในปีที่แล้ว

ข) ทำเหมือนข้อ (ก) แต่หาช่วงเชื่อมั่นของประชากรหญิง (ที่มีงานทำ)

ค) เปรียบเทียบช่วงเชื่อมั่นระหว่างข้อ (ก) และ (ข) กับข้อ 5.1 (ก) เหตุใดช่วงเชื่อมั่นจึงมีขนาดที่ต่างกัน

5.3 ในการศึกษาการทำงานของพนักงานผู้หนึ่ง ถ้าใช้วิธีสังเกตการทำงานของเขาในช่วงเวลาที่ห่างเท่า ๆ กัน เช่นทุก 15 นาที โดยใช้การถ่ายภาพ หรือ TV หรือให้คนสังเกตการณ์ วิธีนี้คือการสุ่มตัวอย่างแบบมีระบบ ท่านคิดว่าเหมาะสมหรือไม่

5.4 ถ้าสินค้าที่พนักงานผู้นั้นผลิต จะได้รับการบรรจุใส่หีบห่อเก็บไว้ในห้องเก็บสินค้า และมีการตรวจสภาพสินค้าคงคลังเป็นประจำทุกสัปดาห์ เพื่อคัดสินค้าที่เน่าเสียทิ้งไป ท่านจะใช้วิธีคัดเลือกแต่ละหน่วยด้วยการสุ่มแบบมีระบบไหม? เช่นการตรวจทุก ๆ กล่องที่ 50 โดยนับจากซ้ายไปขวา หรือจากข้างล่างขึ้นข้างบน หรือจากด้านหน้าไปด้านหลัง ของห้องเก็บสินค้าคงคลัง? เพราะเหตุใด?

6. แผนงานสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ

ยังมีแผนงานสุ่มตัวอย่างแบบอื่น ๆ นอกเหนือจากการสุ่มตัวอย่างแบบง่าย และเป็นแผนงานที่มีประสิทธิภาพ ที่ควรกล่าวถึง คือ

6.1 แผนงานสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ (Stratified Random Sampling)

6.2 แผนงานสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling)

6.1 แผนงานสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ

ถ้าเราสามารถจำแนกประชากรตามคุณลักษณะบางอย่าง เช่น เราจำแนกบุคคลตามอายุ เพศ การศึกษา สถานภาพสมรส เป็นต้น เราอาจจำแนกธุรกิจตามขนาดของทุนและสินทรัพย์ จำนวนขาย จำนวนพนักงาน ลักษณะที่เราจำแนกนี้ เรียกว่า ชั้นภูมิ (strata) และวิธีการแบ่งประชากรออกเป็นชั้นภูมิ เรียกว่า statratification นั่นคือเราสร้างชั้นภูมิจากคุณลักษณะที่เราทราบ การแบ่งชั้นภูมิของประชากรมีหลักการ 2 ข้อ คือ

1. ต้องให้ความผันแปร**ภายในชั้น**ภูมิเดียวกันมีค่าน้อยที่สุด นั่นคือต้องมีความเป็นเอกภาพมากที่สุด
2. ควรให้มีความผันแปร**ระหว่างชั้น**ภูมิต่าง ๆ มากที่สุด

ปัญหาที่ตามมาคือ

1. จะใช้คุณลักษณะใดเป็นหลักในการแบ่งชั้นภูมิ?
2. จะแบ่งสรร n ตัวอย่าง ให้แต่ละชั้นภูมิเท่าใด?
3. เมื่อได้ข้อมูลจากชั้นภูมิต่าง ๆ แล้ว จะวิเคราะห์อย่างไร?

คำตอบคือ

1. คุณลักษณะที่สมควรเลือกใช้เป็นเกณฑ์แบ่งชั้นภูมิคือ คุณลักษณะที่เราทราบว่า **ทำให้ข้อมูลภายในชั้นภูมิเดียวกันมีความผันแปรน้อยที่สุด และข้อมูลจากชั้นภูมิที่ต่างกัน มีความผันแปรมากที่สุด** เช่นการศึกษาความพอใจของพนักงาน ต่อระบบสวัสดิการต่าง ๆ เช่น เงินตอบแทน, บำนาญ, ค่ารักษาพยาบาล, การประกันชีวิต,

การให้หยุดพักผ่อน ความเห็นของพนักงานอาจแตกต่างกันระหว่างเพศ และอายุของพนักงาน ปกติจะแบ่งประชากรเป็นชั้นภูมิ ตามเพศ และอายุ ก่อนการสุ่มตัวอย่าง

2. จะจัดสรรขนาดตัวอย่างให้แต่ละชั้นภูมิอย่างไร? วิธีที่ควรพิจารณาใช้มี 2 วิธี คือ

2.1 **การจัดสรรแบบสัดส่วน** (Proportional allocation)

2.2 **การจัดสรรแบบ Optimum** (Optimum allocation)

ตัวอย่าง

สมมุติ กรุงเทพฯ มีประชากรผู้ใหญ่ 5 ล้านคน เป็นชาย 2 ล้าน และหญิง 3 ล้าน ถ้าเราใช้หน่วยเป็นล้านคน เราจะมี

$$N = 5,\ N_{ช} = 2,\ N_{ญ} = 3$$

รายได้เฉลี่ยของชาย : $\mu_{ช} = 1600$ บาท , $\sigma_{ช} = 406$

รายได้เฉลี่ยของหญิง : $\mu_{ญ} = 600$ บาท , $\sigma_{ญ} = 300$

ดังนั้นค่าเฉลี่ยรวบยอด :

$$\mu = \frac{N_{ช}\mu_{ช} + N_{ญ}\mu_{ญ}}{N_1 + N_2}$$

$$= \frac{2(1600) + 3(600)}{2 + 5}$$

$$= 1{,}000$$

และความแปรปรวนของรายได้ คือ

$$\sigma^2 = \frac{N_{ช}(\mu_{ช} - \mu)^2 + N_{ญ}(\mu_{ญ} - \mu)^2 + N_{ช}\sigma_{ช}^2 + N_{ญ}\sigma_{ญ}^2}{N_{ช} + N_{ญ}}$$

$$= \frac{2(1600 - 1000)^2 + 3(600 - 1000)^2 + 2(406)^2 + 3(300)^2}{2 + 3}$$

$$= 360{,}000$$

และ $\sigma = \sqrt{360{,}000} = 600$

และ 95% ช่วงเชื่อมั่นของ μ คือ

$$= 1000 \pm 1.96 (600/\sqrt{400})$$

$$= 1000 \pm 58.8$$

$$= 941.2, 1058.8$$

ถ้าแบ่งประชากรตามเพศ จะได้ 2 ชั้นภูมิ คือ ชาย และหญิง

$$N_ช = 2, \ N_ญ = 3$$

ถ้าจะจัดสรรแบบเป็นสัดส่วน นั่นคือต้องสุ่มในอัตราส่วน ช : ญ = 2 : 3 หรือ 40% กับ 60% เช่นต้องการตัวอย่าง n = 400 $n_ช$ = .40 (400) = 160 คน และ $n_ญ$ = .60 (400) = 240 คน จากตัวอย่างจะได้ค่าสถิติอีก 4 ตัว คือ $\bar{X}_ช, \bar{X}_ญ$ และ $S_ช, S_ญ$

ดังนั้น ถ้าไม่ทราบค่าแท้จริงของ μ เราจะใช้ข่าวสารจากตัวอย่าง 2 ชั้นภูมินี้ประมาณ จะต้องถ่วงน้ำหนักด้วยขนาดของชั้นภูมิก่อน ให้ $\bar{X}_S$ คือค่าเฉลี่ยรวบยอด (ไม่คำนึงถึงเพศ) ซึ่งหาได้โดยวิธีการสุ่มแบบมีชั้นภูมิ ($\bar{X}_S$ = stratified mean)

$$\bar{\bar{X}}_S = \frac{N_ช\bar{x}_ช + N_ญ\bar{x}_ญ}{N_ช + N_ญ = N}$$

$\bar{X}_S$ จะมีความแปรปรวน ดังนี้

$$V(\bar{X}_S) = V\left(\frac{N_ช\bar{x}_ช}{N}\right) + V\left(\frac{N_ญ\bar{x}_ญ}{N}\right)$$

$$= \left(\frac{N_ช}{N}\right)^2 V(\bar{x}_ช) + \left(\frac{N_ญ}{N}\right)^2 V(\bar{x}_ญ)$$

$$= \left(\frac{N_ช}{N}\right)^2 \frac{\sigma_ช^2}{n_ช} + \left(\frac{N_ญ}{N}\right)^2 \frac{\sigma_ญ^2}{n_ญ}$$

และต้องประมาณค่า $\sigma_{ช}^2$ และ $\sigma_{ญ}^2$ ด้วย $S_{ช}^2$ และ $S_{ญ}^2$

ดังนั้น $V(\bar{x}_S) = (\frac{2}{5})^2 \frac{(406)^2}{160} + (\frac{3}{5})^2 \frac{(300)^2}{240}$

$= 300$

$\sigma_{\bar{x}_S} = \sqrt{300} = 17.32$

ดังนั้น 95% ช่วงเชื่อมั่นของ μ จากการสุ่มแบบแบ่งเป็นชั้นภูมิ คือ

$\bar{\bar{x}}_S \pm 1.96\, S_{\bar{x}_S}$

$= \bar{\bar{x}}_S \pm 1.96\,(17.32)$

การจัดสรรแบบ optimum

วิธีจัดสรรแบบนี้ให้ความแม่นยำสูงที่สุด เพราะได้พิจารณาทั้งขนาดและความผันแปรภายในชั้นภูมิต่าง ๆ นั่นคือขนาดตัวอย่างของแต่ละชั้นภูมิจะเป็นสัดส่วนโดยตรงกับค่าเบี่ยงเบนมาตรฐาน คือ $\sigma_{ช}$, $\sigma_{ญ}$ และขนาดของชั้นภูมิ คือ $N_{ช}$ และ $N_{ญ}$ นั่นคือ

$$\frac{n_{ช}}{n_{ญ}} = \frac{N_{ช}\sigma_{ช}}{N_{ญ}\sigma_{ญ}} = \frac{2 \times 406}{3 \times 300} \approx \frac{190}{210}$$

นั่นคือเมื่อกำหนดขนาดตัวอย่าง n = 400 จะสุ่มชายมา 190 คน และหญิง 210 คน เมื่อเปรียบเทียบกับการจัดสรรวิธีแรก คือแบบเป็นสัดส่วน (2 : 3 = 160 : 240) จะเห็นว่าขนาดตัวอย่างของชายต้องเพิ่มขึ้น เพราะ $\sigma_{ช} > \sigma_{ญ}$

ค่าเฉลี่ยรวบยอด $\bar{x}_S$ ที่ได้จากตัวอย่างแบบ optimum จะมีความแปรปรวน ดังนี้

$$V(\bar{x}_S) = \left(\frac{N_{ช}}{N}\right)^2 \frac{\sigma_{ช}^2}{n_{ช}} + \left(\frac{N_{ญ}}{N}\right)^2 \frac{\sigma_{ญ}^2}{n_{ญ}}$$

$$= \left(\frac{2}{5}\right)^2 \frac{(406)^2}{190} + \left(\frac{3}{5}\right)^2 \frac{(300)^2}{210}$$

$$= 293$$

และ $S_{\bar{x}_S} = \sqrt{293} = 17.13$

ดังนั้น 95% ช่วงเชื่อมั่นของ μ จากตัวอย่างแบบ optimum คือ

$$\bar{x}_S \pm 1.96\ (17.13)$$

หากค่าใช้จ่ายในการสัมภาษณ์ต่อหน่วยตัวอย่างไม่เท่ากัน บางชั้นภูมิต่ำ บางชั้นภูมิสูง เช่นนี้ จะต้องนำค่าใช้จ่ายต่อหน่วยเข้าร่วมพิจารณาด้วย ด้วยมีหลักว่า ชั้นภูมิที่เสียค่าใช้จ่ายต่อหน่วยสูง (ต้องเดินทางไกล ฯลฯ) จะใช้ขนาดตัวอย่างน้อยกว่าชั้นภูมิที่มีค่าใช้จ่ายต่อหน่วยต่ำ นั่นคือ ขนาดตัวอย่างของชั้นภูมิจะเป็นสัดส่วนแบบผกผันกับค่าใช้จ่ายต่อหน่วย นั่นคือ

$$\frac{n_ช}{n_ญ} = \frac{N_ช\sigma_ช / \sqrt{C_ช}}{N_ญ\sigma_ญ / \sqrt{C_ญ}}$$

สมมุติว่ามีค่าใช้จ่ายสำหรับการสัมภาษณ์ 4,000 บาท ค่าใช้จ่ายในการสัมภาษณ์ชาย หน่วยละ 16 บาท หญิงหน่วยละ 9 บาท เราจะได้อัตราส่วนดังนี้

$$\frac{n_ช}{n_ญ} = \frac{2(406)/\sqrt{16}}{3(300)/\sqrt{9}} = \frac{19}{28}$$

และจะต้องเลือก n ให้สอดคล้องกับงบประมาณ 4,000 บาทที่มีอยู่ นั่นคือ

$$\frac{19}{(19+28)}\ n \times 16 + \frac{28}{19+28}\ n \times 9 = 4000$$

จะได้ n = 338 จึงจัดสรรในอัตรา 19 : 28 ดังนี้

$$n_ช = \frac{19}{19+28}(338) = 136$$

$$n_ญ = \frac{28}{19+28}(338) = 202$$

และเมื่อคำนวณ $\sigma_{\bar{x}_S} = 18.83$ ซึ่งโตกว่าการไม่คิดค่าใช้จ่ายต่อหน่วยและงบประมาณ แสดงให้

เห็นว่าความแม่นยำในการประมาณค่า ลดลง เพราะลดขนาดตัวอย่างในกลุ่มชายซึ่งมีความแปรปรวนสูง

ในทางปฏิบัติ เรามักไม่ทราบค่า $\sigma_{ช}$, $\sigma_{ญ}$ และถ้าไม่สามารถหาค่าประมาณที่เหมาะสมได้ ก็ควรใช้การจัดสรรแบบสัดส่วน

สำหรับ พารามิเตอร์ π เมื่อต้องการประมาณค่า π และใช้การสุ่มแบบชั้นภูมิ เราจะประมาณ π ด้วย p_s ในเมื่อ

$$p_s = \frac{N_{ช}p_{ช} + N_{ญ}p_{ญ}}{N_{ช} + N_{ญ}}$$

และ p_s จะมีความแปรปรวน ดังนี้

$$V(p_s) = (\frac{N_{ช}}{N})^2 V(p_{ช}) + (\frac{N_{ญ}}{N})^2 V(p_{ญ})$$

ในเมื่อ $V(p_{ช}) = \frac{\pi_{ช}(1 - \pi_{ช})}{n_{ช}}$

$$V(p_{ญ}) = \frac{\pi_{ญ}(1 - \pi_{ญ})}{n_{ญ}}$$

และถ้าไม่ทราบว่า $\pi_{ช}$, $\pi_{ญ}$ ให้ประมาณด้วย $p_{ช}$, $p_{ญ}$ ตามลำดับ

โดยทั่ว ๆ ไป การสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ มิได้มีเพียง 2 ชั้นภูมิจะมีทั้งหมด L ชั้นภูมิ และจะต้องจัดสรร n ตัวอย่าง สำหรับ L ชั้นภูมิ ให้ n_h แทนขนาดตัวอย่างของชั้นภูมิที่ h และ

$$n_1 + n_2 + \ldots + n_h = n$$

ถ้าใช้ optimum allocation และคำนึงถึงค่าใช้จ่ายต่อหน่วยคือ C_h

$$n_h = \frac{N_h\sigma_h/\sqrt{C_h}}{\Sigma N_h\sigma_h/\sqrt{C_h}} \times n$$

ถ้า C_h ไม่ต่างกันระหว่างชั้นภูมิ

$$n_h = \frac{N_h \sigma_h}{\Sigma N_h \sigma_h} \times (n)$$

และถ้า σ_h ของแต่ละชั้นภูมิ ไม่ต่างกันมาก

$$n_h = \frac{N_h}{N} \times (n)$$

แบบฝึกหัด

5.5 บริษัทหนึ่งต้องการสุ่มพนักงานตัวอย่างมาจำนวน 500 คน และมั่นใจว่าตัวแปรที่มีอิทธิพลต่อข้อมูลที่เก็บมาคืออายุการทำงานกับบริษัท ดังนั้นจึงจะแบ่งเป็นชั้นภูมิตามอายุการทำงาน ซึ่งบริษัทมีข้อมูลดังนี้

อายุการทำงาน	h	N_h	σ_h
น้อยกว่า 2 ปี	1	2,000	.7
2 - 5 ปี	2	1,000	1.4
5 ปี ขึ้นไป	3	1,000	2.8

ก) จงหาขนาดตัวอย่าง n_h ของแต่ละชั้นภูมิ โดยวิธีจัดสรรแบบเป็นสัดส่วน

ข) จงจัดสรรแบบ optimum

ค) เหตุใดผลที่ได้ในข้อ (ก) และ (ข) จึงต่างกัน

5.6 จากข้อ 5.5 ถ้าได้ข้อมูลจากตัวอย่าง ดังนี้

(x = มูลค่าการถือหุ้นของบริษัท เป็นบาท)

อายุการทำงาน	h	$\bar{x}_h$	S_h
น้อยกว่า 2 ปี	1	60	25
2 - 5 ปี	2	200	60
มากกว่า 5 ปี	3	2,500	300

ก) สมมุติว่าข้อมูลมาจากการจัดสรรแบบ**เป็นสัดส่วน**กับขนาดของประชากร จงหา $\bar{X}_s$ และ $V(\bar{x}_s)$

ข) สมมุติว่าเป็นการจัดสรรแบบ optimum จงหา $\bar{x}_s$ และ $V(\bar{x}_s)$

ค) จงเปรียบเทียบ $\bar{x}_s$ จาก (ก) และ (ข) และวิจารณ์ในกรณีที่ต่างกัน

ง) จงเปรียบเทียบ $V(\bar{x}_s)$ จาก (ก) และ (ข) และวิจารณ์ในกรณีที่ต่างกัน

5.7 จากข้อ 5.5 สมมุติได้ข้อมูลจากตัวอย่าง ดังนี้

(x = จำนวนปีที่สำเร็จการศึกษา)

อายุการทำงาน	h	$\bar{x}_h$	S_h
ต่ำกว่า 2 ปี	1	13	3
2 - 5 ปี	2	12	2.5
เกิน 5 ปี	3	10	2

ก) สมมุติใช้การจัดสรรแบบเป็นสัดส่วน จงประมาณจำนวนปีที่จบการศึกษาโดยถัวเฉลี่ย $\bar{x}_s$ และ $V(\bar{x}_s)$

ข) ถ้าใช้การจัดสรรแบบ optimum จงหาค่า $\bar{x}_s$ และ $V(\bar{x}_s)$

ค) จงเปรียบเทียบ $V(\bar{x}_s)$ ในข้อ (ก) และ (ข) และวิจารณ์

5.8 งานสำรวจชิ้นหนึ่ง ต้องทำการสำรวจจากครอบครัวตัวอย่างซึ่งมีรายได้ต่ำ ซึ่งอาศัยอยู่ในชุมชนแออัด ครอบครัวในท้องที่ตัวอย่างแบ่งเป็น 2 ประเภท คือ มีการศึกษา และไม่มีการศึกษา (ยึดหัวหน้าครอบครัวเป็นหลัก) สำหรับครอบครัวที่ไม่มีการศึกษาต้องใช้พนักงานสำรวจที่มีประสิทธิภาพสูง เพื่อการสื่อความหมายที่ดีและได้ข้อมูลที่เชื่อได้ จึงต้องใช้พนักงานที่มีความรู้พิเศษ และจะต้องเสียค่าใช้จ่ายสูงกว่าปกติ ข้อมูลของการสำรวจมีดังนี้

การศึกษา	จำนวนครัวเรือน	ค่าใช้จ่ายต่อหน่วย	S_h
มีการศึกษา	6,000	9	500
ไม่มีการศึกษา	4,000	16	800

ก) จงจัดสรรตัวอย่างขนาด 165 ครัวเรือนให้กับชั้นภูมิทั้ง 2 โดยวิธีสัดส่วน

ข) จงใช้วิธี optimum จัดสรร

ค) ถ้ามีงบประมาณสำหรับสัมภาษณ์ 2,000 บาท จงจัดสรรให้แก่ 2 ชั้นภูมิ

ง) จะต้องเสียค่าใช้จ่ายในการสัมภาษณ์ แต่ละชั้นภูมิเป็นเงินเท่าใด?

จ) จงหา $V(\bar{x}_s)$ ของการจัดสรรแต่ละประเภท

5.9 คลังเก็บสินค้าของตัวแทนจำหน่ายผลิตภัณฑ์ไฟฟ้าเสียหายเล็กน้อย เนื่องจากพายุไต้ฝุ่น กิจการต้องการเปิดกล่องสินค้าเพื่อสำรวจความเสียหายเนื่องจากโดนน้ำจำนวนหนึ่ง ผลความเสียหายส่วนใหญ่เกิดกับชั้นล่างของคลังเก็บสินค้า ดังนั้นจึงจะแบ่งโดยใช้ชั้นต่าง ๆ เป็นชั้นภูมิ ผลการสำรวจ ได้ข้อมูล ดังนี้

ชั้น	จำนวนรวม	จำนวนกล่องที่เปิดตรวจ	จำนวนกล่องที่เสีย
1	10,000	60	30
2	12,000	50	15
3	8,000	40	8

ก) จงประมาณเปอร์เซนต์ของเครื่องไฟฟ้า ที่เสียหายเพราะน้ำท่วม

ข) จงหา $V(p_s)$

ค) ถ้าต้นทุนเฉลี่ยของสินค้ากล่องละ 25 บาท จงหา 95.45% ช่วงเชื่อมั่นของมูลค่าสินค้าที่เสียหายเนื่องจากน้ำท่วม

7. การประมาณค่าโดยวิธีอัตราส่วน (Ratio Estimation)

วิธีนี้ใช้หลักการของความถดถอยและสหสัมพันธ์ คือการประมาณค่าตัวแปร y โดยอาศัยความสัมพันธ์กับตัวแปร x เราจะเลือกตัวแปรที่ให้ค่า R^2 สูงสุดเป็นตัวแปร x ดังนั้นแม้เราจะไม่ทราบค่า μ_y แต่เราทราบค่า μ_x ถ้าเราสุ่มตัวอย่าง n จำนวน และได้ค่าสังเกต $x_1, x_2, \ldots, x_n$ และ $y_1, y_2, \ldots, y_n$ เราจะได้ r คืออัตราส่วนของค่าเฉลี่ย และอัตราส่วนของผลรวม ดังนี้ (r = ratio)

$$r = \frac{\bar{y}}{\bar{x}} = \frac{\Sigma y_i}{\Sigma x_i}$$

r เป็นค่าประมาณของ ρ = อัตราส่วนค่าเฉลี่ยของประชากร

$$\rho = \frac{\mu_y}{\mu_x}$$

ตัวอย่างที่แสดงการใช้ประโยชน์ของวิธีการประมาณค่าแบบอัตราส่วนเกิดขึ้นในประเทศสหรัฐ-อเมริกา เมื่อประมาณเกือบ 20 ปีก่อน ซึ่งในอเมริกายังมีชาวพื้นเมืองเผ่าต่าง ๆ อยู่กระจายทั่วประเทศ และในความพยายามที่จะสำมะโนจำนวนชาวพื้นเมืองในรัฐหนึ่ง มีวิธีเดียวคือการสัมภาษณ์หัวหน้าเผ่า และให้หัวหน้าเผ่าประมาณจำนวนลูกน้อง หรือพลเมืองของตน ให้ x_i เป็นค่าประมาณที่ได้จากหัวหน้าเผ่า i เมื่อรวมค่า x_i จากทุก ๆ เผ่า ปรากฏว่า ให้ยอดรวมสูงกว่าค่าที่ควรจะเป็นของค่าประชากรอย่างมาก จึงทำการสุ่มตัวอย่างชาวพื้นเมืองมาจำนวนหนึ่ง แล้วนับจำนวนสมาชิกโดยเจ้าหน้าที่แห่งรัฐ ให้ y_i คือ จำนวนที่เจ้าหน้าที่นับได้ของกลุ่มที่ i และเพื่อจะหาค่าประมาณที่ดีของ N กลุ่ม จะใช้ความสัมพันธ์ $r = \bar{y}/\bar{x}$ ซึ่งเป็นมาตราที่วัด

"ความโม้" ของหัวหน้าเผ่า เนื่องจากหัวหน้าเผ่ามีความหยิ่ง และรู้สึกโก้ที่จะบอกว่ามีสมาชิกจำนวนสูง และ "ความโม้" ของหัวหน้าเผ่าทั้งหลายเผอิญเกิดในระดับ (degree) ใกล้เคียงกัน ดังนั้น ค่า y_i/x_i ของแต่ละเผ่า จึงไม่ต่างกันมาก จึงนับว่าเป็นความโชคดีของนักสถิติ จำนวนประชากร (ชาวพื้นเมือง) ในรัฐนั้น จะประมาณได้ โดย $Nr\,\mu_x$.

ตัวอย่างทางธุรกิจของวิธี ratio estimate คือ กรณีตัวอย่างของบริษัท Jiffy โดยผู้จัดการฝ่ายการตลาดต้องการประมาณยอดรวมล่วงหน้าของผลิตภัณฑ์ชิ้นใหม่ของบริษัทสำหรับปีหน้า เขาจึงใช้สถิติการขาย และยอดรวมจำนวนขายสำหรับลูกค้าแต่ละรายของปีที่แล้วของสินค้าซึ่งมีคุณภาพทัดเทียมกัน ให้ x_i คือจำนวนขายสำหรับลูกค้ารายที่ i ของสินค้าเปรียบเทียบในปีก่อน และ y_i คือจำนวนการสั่งซื้อล่วงหน้าจากลูกค้า i สำหรับผลิตภัณฑ์ออกใหม่ ได้ข้อมูลจากลูกค้าทั้งหมดของบริษัท จำนวน 20 ราย ดังนี้

ถ้าต้องการประมาณค่าเฉลี่ยประชากร μ_y จากความสัมพันธ์ จะได้

$$\mu_y = r\mu_x$$

และค่าประมาณของยอดรวมของประชากร $N\mu_y$ คือ

$$N\mu_y = Nr\mu_x$$

และความแปรปรวนของ r จะหาได้**โดยประมาณ** ดังนี้

$$V(r) = \frac{\Sigma d_i^2}{n(n-1)\mu_x^2}$$

ในเมื่อ $d_i = (y_i - r\mu_x)$

จากสูตร V(r) จะเห็นว่า

ถ้า $r = y_i/x_i$, $v(r) = 0$

แต่ไม่ค่อยพบ V(r) = 0 นอกจากใช้เป็นหลักเกณฑ์ว่า ถ้าอัตราส่วน y_i/x_i มีค่าใกล้เคียงกัน (Uniform) จะทำให้ d_i มีค่าเล็ก และจะมีผลให้ V(r) เป็นค่าเล็กด้วย

ส่วนความแปรปรวนของค่า ratio estimate ของ μ_y คือ

$$V(\text{ratio estimate } \mu_y) = V(r\mu_x) = \mu_x^2 V(r) = \frac{Cd,'}{n(n-1)}$$

จำนวนการซื้อปีก่อน และการสั่งซื้อล่วงหน้า
สำหรับผลิตภัณฑ์ใหม่ จากลูกค้าของ
บริษัท Jiffy จำนวน 20 ราย

ลูกค้า	x_i = จำนวนซื้อปีก่อน	y_i = การสั่งซื้อล่วงหน้า
1	5,071	325
2	7,230	400
3	1,325	67
•		
•		
20	3,207	146
รวม	100,000	5,000

ผู้จัดการฝ่ายการตลาดพบว่า อัตราส่วน y_i/x_i ของลูกค้า 20 ราย คล้าย ๆ กัน

และ $$r = \frac{\bar{y}}{\bar{x}} = \frac{5,000/20}{100,000/20} = .05$$

ถ้า Jiffy มีลูกค้าทั้งสิ้น 750 ราย = N และทราบว่าลูกค้าซื้อสินค้าเปรียบเทียบในปีก่อนโดยเฉลี่ยคนละ 4,500 บาท (μ_x = 4,500) จะประมาณยอดขายเฉลี่ยของสินค้าใหม่ในปีหน้า ต่อคน (μ_y)

ดังนี้

Ratio estimate μ_y = $r\mu_x$ = (.05) (4,500) = 225 บาท

และประมาณยอดรวม $N\mu_y$

Ratio estimate $N\mu_y$ = $Nr\mu_x$ = 750 (.05) (4,500) = 168, 750 บาท

ลองประมาณค่าเฉลี่ยแบบ ratio กับแบบการสุ่มแบบง่าย โดยไม่ใช่ข่าวสารของปีกลาย (x_i)

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{5,000}{20} = 250 \text{ บาท}$$

$$N\bar{y} = (750)\ (250) = 187,500 \text{ บาท}$$

ค่าประมาณแบบอัตราส่วนดีกว่า เนื่องจากตัวอย่างลูกค้า 20 รายมีการซื้อค่อนข้างสูงกว่าลูกค้าทั่ว ๆ ไป เมื่อใช้แบบอัตราส่วน คือ จำนวนซื้อของปีก่อนมาคิดด้วย ค่าประมาณแบบอัตราส่วนจึงปรับแก้ไขโดยอัตโนมัติ

นอกจากนั้น ค่าประมาณแบบอัตราส่วนยังมีความคลาดเคลื่อนมาตรฐานต่ำกว่าด้วย

$$S(\text{ratio estimate } \mu_y) = \sqrt{\mu_x^2\, V(r)} = \sqrt{\frac{\Sigma d_i^2}{n(n-1)}} = 12.61 \text{ บาท}$$

ส่วนค่าประมาณจากการสุ่มแบบง่าย มีความคลาดเคลื่อนมาตรฐาน

$$S_{\bar{y}} = \frac{S_y}{\sqrt{n}} = \sqrt{\frac{\Sigma (y - \bar{y})^2}{n(n-1)}} = 37.61$$

แบบฝึกหัด

5.10 บริษัทเคมีภัณฑ์ได้ค้นพบสารใหม่ ซึ่งมีคุณสมบัติ ลดกรด จากเศษขยะที่ทิ้งในแม่น้ำและจากโรงงาน ปริมาณสารลดกรดที่โรงงานต่าง ๆ ควรใช้ จะเป็นสัดส่วนโดยตรงกับปริมาณสินค้าที่แต่ละโรงงานผลิต (เป็นต้น) ถ้าในปีก่อน ๆ จากสถิติ ผลิตภัณฑ์รายปี