บทที่ 5
การชักตัวอย่างและการแจกแจงตัวอย่าง

การชักตัวอย่างเป็นกระบวนการในการเลือกตัวอย่างซึ่งประกอบด้วยหน่วยตัวอย่างจากบางส่วนของประชากร แล้วนำเอาตัวอย่างที่ได้ไปศึกษาหาข้อมูลหรือรายละเอียดต่าง ๆ ที่สนใจ ด้วยวิธีการที่ถูกต้องและเหมาะสม เพื่อนำไปสู่การอ้างอิงถึงประชากรเป็นสิ่งสำคัญ เมื่อเป็นเช่นนั้นคุณภาพของข้อมูลซึ่งเป็นวัตถุดิบในการวิเคราะห์และการประมวลผลจึงมีความสำคัญอย่างยิ่งเช่นกัน การกำหนดวิธีการเลือกตัวอย่างที่มีประสิทธิภาพจึงมีความความจำเป็นอย่างไม่อาจหลีกเลี่ยงได้ การทำความเข้าใจเกี่ยวกับวิธีการชักตัวอย่างจำเป็นต้องมีความรู้ ความเข้าใจทั้งเทคนิควิธีและกระบวนการดำเนินการเพื่อให้ได้ตัวอย่างที่ดี ซึ่งจะทำให้ได้ข้อมูลที่มีคุณภาพ อันจะนำไปสู่การตัดสินใจที่ถูกต้องจากการใช้ข้อมูลเหล่านั้น ในการชักตัวอย่างจะต้องทำความเข้าใจศัพท์ทางสถิติและเทคนิควิธีการที่สำคัญ เช่น ประชากร ตัวอย่าง หน่วยตัวอย่าง เทคนิคการสุ่มตัวอย่างแบบต่าง ๆ

5.1 ประชากร ตัวอย่าง และหน่วยตัวอย่าง

คำว่า “ประชากร (population)” ในทางสถิติ หมายถึง กลุ่มของหน่วยตัวอย่างที่สนใจจะศึกษาเรื่องใดเรื่องหนึ่งจากกลุ่มนั้น เช่น ในการศึกษาเกี่ยวกับวัฒนธรรมทางการเมืองของแม่บ้านในจังหวัดนครศรีธรรมราช ประชากรในที่นี้คือแม่บ้านทุกคนในจังหวัดนครศรีธรรมราช ตัวอย่างจากการศึกษานี้ ผู้วิจัยไม่จำเป็นที่จะต้องเก็บรวบรวมข้อมูลจากแม่บ้านทุกคนในจังหวัดนครศรีธรรมราช เนื่องจากมีสมาชิกเป็นจำนวนมาก ต้องใช้เวลาและค่าใช้จ่ายสูงในการเก็บรวบรวมข้อมูล วิธีที่จะปฏิบัติก็คือใช้ตัวอย่าง ซึ่งเป็นส่วนหนึ่งของประชากรเป็นตัวแทนของประชากร โดยตัวอย่างจะได้จากการใช้เทคนิคการเลือกตัวอย่าง ที่มีอยู่หลายเทคนิคด้วยกัน ในการเลือกใช้เทคนิคแบบใดนั้นขึ้นอยู่กับความเหมาะสม ทั้งทางด้านทฤษฎีและความชัดเจนถูกต้องที่ควรจะเป็นที่ผู้วิจัยได้พิจารณาแล้ว โดยมีวัตถุประสงค์ ที่สำคัญคือการได้ตัวอย่างที่เป็นตัวแทนที่ดีที่สุดนั่นเอง

บทนิยามที่ 5.1 ประชากร คือหน่วยตัวอย่างทั้งหมดซึ่งเกี่ยวข้องกับสิ่งที่สนใจ
บทนิยามที่ 5.2 ตัวอย่าง คือกลุ่มของหน่วยตัวอย่างซึ่งเป็นส่วนหนึ่งของประชากร
บทนิยามที่ 5.3 หน่วยตัวอย่าง คือหน่วยที่ให้รายละเอียดเกี่ยวกับสิ่งที่สนใจศึกษา

หน่วยตัวอย่างทั้งหมดในประชากร คือขนาดของประชากร ตัวอย่างเช่น ในการศึกษาเกี่ยวกับการให้บริการอินเตอร์เน็ตของสำนักวิทยบริการ มหาวิทยาลัยราชภัฏนครศรีธรรมราช ในมหาวิทยาลัยราชภัฏนครศรีธรรมราชมีนักศึกษา 3,500 คน กล่าวได้ว่า ขนาดประชากรเป็น 3,500 คน และจะแทนขนาดประชากรด้วย N หรือ มีแม่บ้านในเขตอำเภอเมือง จังหวัดนครศรีธรรมราช 10,000 คน กล่าวได้ว่า ขนาดประชากรในการศึกษาเกี่ยวกับวัฒนธรรมทางการเมืองของแม่บ้านในเขตอำเภอเมือง จังหวัดนครศรีธรรมราช เป็น 10,000 คน หรือ N = 10,000

ค่าจากแต่ละหน่วยในประชากร เป็นค่าของตัวแปรสุ่ม เช่น ถ้าให้ X เป็นตัวแปรสุ่มจากการโยนลูกเต๋า 2 ลูกไปเรื่อย ๆ ค่าแต่ละค่าที่สังเกตได้จากการโยนลูกเต๋า คือ ค่าของ X ถ้าให้ X แทนผลรวมของตัวเลขบนหน้าลูกเต๋า จะมีค่าที่เป็นไปได้คือ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 หรือ 12 หากลองศึกษาความสัมพันธ์ระหว่างตัวแปรสุ่มและประชากรโดยพิจารณาประชากรต่อไปนี้ซึ่งเป็นขนาดรองเท้าของผู้ชาย 10 คน ได้แก่ 7, 9, 8, 9, 10, 8, 8, 9, 11, 6 โดยให้ X เป็นตัวแปรสุ่มที่แทนขนาดรองเท้าผู้ชายที่ถูกเลือกจากประชากรแล้ว X จะมีค่า ดังนี้ คือ 6, 7, 8, 9, 10, 11 ค่าของ X จะเกี่ยวข้องกับความน่าจะเป็นที่เรียกว่า การแจกแจงความน่าจะเป็นของ X ในที่นี้ความน่าจะเป็นที่เข้าใจก็คือความถี่สัมพัทธ์ (relative frequency) ด้วยเหตุนี้ความน่าจะเป็นของประชากร จึงถูกกำหนดด้วยการแจกแจงความน่าจะเป็น นั่นคือ ทุกหน่วยตัวอย่างในประชากรสามารถแจกแจงได้โดยทราบการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม ที่ถูกกำหนดโดยประชากร

เมื่อกล่าวถึงสมบัติหลายอย่างของตัวแปรสุ่ม มักอ้างไปถึงสมบัติต่าง ๆ ของประชากร เช่น ค่าคาดหมายของตัวแปรสุ่ม X ซึ่งถูกกำหนดเป็นค่าเฉลี่ยของ X เมื่อทำการทดลองซ้ำหลายครั้ง เรียกว่าค่าเฉลี่ยของประชากร ซึ่งมีค่าจากทุกหน่วยตัวอย่างเป็นค่าของตัวแปรสุ่ม X เช่น ถ้า X แทนขนาดของรองเท้าผู้ชายที่ถูกเลือกจากประชากรด้วยความน่าจะเป็นที่เท่า ๆ กัน ค่าคาดหมายของ X คือ

$$\begin{aligned} E(X) &= (6)\left(\frac{1}{10}\right)+(7)\left(\frac{1}{10}\right)+(8)\left(\frac{3}{10}\right)+(9)\left(\frac{3}{10}\right)+(10)\left(\frac{1}{10}\right)+(11)\left(\frac{1}{10}\right) \\ &= 8.5 \end{aligned}$$

เรียก 8.5 ว่าพารามิเตอร์ของการแจกแจงหรือพารามิเตอร์ของประชากรซึ่งคือค่าเฉลี่ยของประชากรนั่นเอง และเขียนแทนด้วย μ ดังนี้

$$\begin{aligned} \mu &= \frac{7+9+8+9+10+8+8+9+11+6}{10} \\ &= 8.5 \end{aligned}$$

บทนิยามที่ 5.4 พารามิเตอร์ คือค่าคงที่ซึ่งเป็นตัวเลขที่ใช้บรรยายลักษณะของประชากร

นิยมใช้อักษรกรีกแทนพารามิเตอร์ เช่น ใช้ μ แทนค่าเฉลี่ยของประชากร σ แทนส่วนเบี่ยงเบนมาตรฐานของประชากร และ σ^2 แทนความแปรปรวนของประชากร

นักสถิติสนใจในการหาข้อสรุปเกี่ยวกับพารามิเตอร์ของประชากร ดังนั้น จำเป็นต้องพิจารณาค่าจากสมบัติของหน่วยตัวอย่าง เพื่อนำไปอนุมาน (inference) พารามิเตอร์ของประชากรซึ่งในเรื่องนี้จะนำไปสู่ทฤษฎีของการสุ่มตัวอย่าง นั่นคือ ถ้าการอนุมานหรือการอ้างอิงต้องการความแม่นยำจำเป็นต้องเข้าใจความสัมพันธ์ของตัวอย่างกับประชากรและแน่นอนที่สุด ตัวอย่างควรเป็นตัวแทนที่ดีของประชากรซึ่งควรมีความไม่เอนเอียงและควรจะเป็นตัวอย่างสุ่ม (random sample) ตัวอย่างที่เลือกได้โดยมีการสุ่มหรือไม่มีการเลือกโดยจงใจนี้ เรียกกันว่าตัวอย่างสุ่ม (ประชุม สุวัตถี, 2527;)

บทนิยามที่ 5.5 ตัวอย่างสุ่มขนาด n คือตัวอย่างซึ่งถูกเลือกมาจากหน่วยตัวอย่าง N หน่วยตัวอย่างของประชากร ด้วยความน่าจะเป็นเท่า ๆ กัน

ในการหาข้อสรุปเกี่ยวกับสัดส่วนของประชากรในเขตหนึ่งของกรุงเทพมหานครเกี่ยวกับความนิยมในการชมข่าวทางโทรทัศน์ เป็นไปไม่ได้ที่จะสอบถามความนิยมจากประชากรทุกคนในกรุงเทพมหานคร แล้วคำนวณค่าพารามิเตอร์ซึ่งเป็นสัดส่วนจริง ๆ เกี่ยวกับความนิยมในการชมข่าวโทรทัศน์ ในทางปฏิบัติจะสุ่มตัวอย่างขนาดใหญ่จากประชากรแล้วคำนวณหาสัดส่วนในการชมข่าวทางโทรทัศน์ ซึ่งค่าสัดส่วนที่ได้จากตัวอย่างจะนำไปอนุมานพารามิเตอร์ของประชากร ค่าที่คำควณได้จากตัวอย่างเรียกว่า ตัวสถิติ (statistic) แต่ในการชักตัวอย่างบางครั้งจะมีตัวอย่างหลาย ๆ ตัวอย่างที่มาจากประชากรเดียวกัน จึงควรพิจารณาตัวสถิติในตัวอย่างที่สนใจศึกษาเท่านั้น

บทนิยามที่ 5.6 ตัวสถิติ คือค่าคงที่ที่คำนวณได้จากตัวอย่างซึ่งเป็นตัวบอกลักษณะของตัวอย่าง

นิยมใช้อักษรลาตินแทนตัวสถิติ เช่น ใช้ $\overline{X}$ แทนค่าเฉลี่ยเลขคณิตของตัวอย่างและใช้ S^2 แทนความแปรปรวนของตัวอย่าง เช่น ถ้าเลือกตัวอย่างสุ่มขนาด 3 จากประชากรของขนาดรองเท้าอาจจะได้ขนาดรองเท้าเป็น 6 8 10 ค่าเฉลี่ยของขนาดรองเท้าเป็นตัวแปรสุ่ม จะแทนด้วย

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

และค่าของตัวแปรสุ่มจะแทนด้วย $\overline{x}$

ดังนั้น

$$\overline{x} = \frac{6+8+10}{3}$$

$$= \frac{24}{3} = 8$$

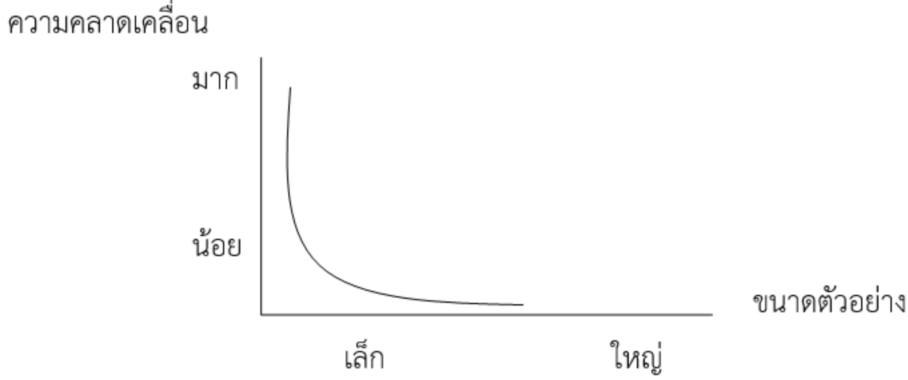
ในทางปฏิบัติ ตัวสถิติจะนำไปใช้ในการประมาณค่าพารามิเตอร์ของประชากร และ จากที่กล่าวมาข้างต้นจะเห็นได้ว่า การเก็บข้อมูลจากทุกหน่วยตัวอย่างในประชากรนั้นทำได้ยากเพราะส่วนใหญ่แล้วประชากรมีขนาดใหญ่ ถ้าทำได้ก็เสียเวลามาก เสียค่าใช้จ่ายสูงอีกทั้งในการเก็บข้อมูลจากทุกหน่วยตัวอย่างซึ่งมีเป็นจำนวนมาก มีโอกาสทำให้เกิดความผิดพลาดได้มาก โดยเฉพาะอย่างยิ่งในกรณีที่ขอบเขตของการเก็บรวบรวมข้อมูลกว้างขวางหรือประชากรมีขนาดใหญ่ ลักษณะของประชากรที่ต้องการเก็บรวบรวมข้อมูลหาได้ยาก อาจส่งผลต่อความเพียงพอของข้อมูลที่จะนำมาประมาณค่าหรือทดสอบสมมติฐานได้ ในการศึกษาวิจัยเกี่ยวกับเรื่องใด ๆ จึงนิยมการสุ่มตัวอย่างจากประชากร ซึ่งใช้เป็นตัวแทนของประชากร ดังนั้น เทคนิคและวิธีการในการเลือกตัวอย่างและขนาดตัวอย่าง จึงมีความสำคัญไม่น้อย นั่นคือ จะต้องทำการเลือกตัวอย่าง อย่างไร ขนาดของตัวอย่าง (sample size) เป็นเท่าใด จึงจะเป็นตัวแทนที่ดีของประชากร นั่นเอง เหตุผลสำคัญที่ต้องเลือกตัวอย่างมี 5 ข้อ ดังนี้ (วินัส พีชวณิชย์ และสมจิต วัฒนาชยากูล, 2537; สรชัย พิศาลบุตร, 2559)

1. ประชากรมีขนาดใหญ่เกินไป ไม่สามารถตรวจสอบได้หมด
2. เป็นการสิ้นเปลืองค่าใช้จ่ายเกินไป ที่จะต้องเก็บข้อมูลจากทุกหน่วยในประชากร
3. เป็นการสิ้นเปลืองเวลา ถ้าศึกษาประชากรโดยตรง
4. ในการเก็บข้อมูลบางครั้งอาจต้องทำลายหน่วยตัวอย่าง เช่น การตรวจสอบคุณภาพของนมสด การตรวจสอบคุณภาพของกระสุน
5. ถ้าทำการเลือกตัวอย่างได้ดีแล้ว การศึกษาจากตัวอย่างจะให้ผลที่เที่ยงตรงมากกว่าการศึกษาจากประชากรโดยตรง

5.2 การกำหนดขนาดตัวอย่าง

การกำหนดขนาดตัวอย่างเป็นการบอกว่าผู้วิจัยใช้ตัวอย่างขนาดเท่าใด หรือใช้หน่วยตัวอย่างเท่าใด หรือใช้หน่วยแจงนับเท่าใด ในการศึกษาวิจัยครั้งหนึ่งต้องกำหนดว่าขนาดตัวอย่างควรเป็นเท่าใด นอกจากจะขึ้นอยู่กับธรรมชาติของประชากร ลักษณะของงานวิจัยแผนของการเลือกตัวอย่างที่กำหนดแล้ว ทรัพยากรด้านงบประมาณ กำลังคน เวลาและเทคนิคการวิเคราะห์ ก็เป็นองค์ประกอบที่จะต้องนำมาพิจารณาด้วย

เคอร์ลิงเจอร์ (Kerlinger, 1973) ได้อธิบายความสัมพันธ์ระหว่างขนาดตัวอย่างกับความคลาดเคลื่อนหรือความผิดพลาดของผลการวิจัย โดยแสดงด้วยกราฟ ดังนี้



ภาพที่ 5.1 ความสัมพันธ์ระหว่างขนาดตัวอย่างกับความคลาดเคลื่อน

5.2.1 เกณฑ์การกำหนดขนาดตัวอย่าง

ก่อนการนำขนาดตัวอย่างไปใช้ในการเลือกตัวอย่างที่เหมาะสม การกำหนดขนาดตัวอย่างให้เกิดความเหมาะสมและสอดคล้องกับทฤษฎีมีความจำเป็นอย่างยิ่ง นักสถิติได้เสนอแนวทางและกำหนดเกณฑ์ในการกำหนดขนาตัวอย่าง ดังนี้ ยุทธ์ ไกยวรรณ์ (2546) ได้เสนอวิธีการกำหนดขนาดตัวอย่างไว้ 4 แบบ ดังนี้

5.2.1 การกำหนดขนาดตัวอย่างโดยใช้เกณฑ์ โดยวิธีนี้ผู้วิจัยต้องทราบจำนวนประชากรที่ค่อนข้างแน่นอนแล้วคำนวณหาขนาดตัวอย่างจากเกณฑ์ต่อไปนี้

5.2.1.1 ใช้ขนาดตัวอย่างเป็น 15-30% ถ้าจำนวนประชากรเป็นหลักร้อย

5.2.1.2 ใช้ขนาดตัวอย่างเป็น 10-15% ถ้าจำนวนประชากรเป็นหลักพัน

5.2.1.3 ใช้ขนาดตัวอย่างเป็น 5-10% ถ้าจำนวนประชากรเป็นหลักหมื่น

5.2.2 การคำนวณหาขนาดตัวอย่างโดยใช้ตารางสำเร็จของคริจซีและมอร์แกน (Krejcie and Morgan) โดยวิธีการนี้ การหาขนาดตัวอย่างคำนวณได้จากสูตร

$$n = \frac{\chi^2 NPQ}{e^2(N-1) + \chi^2 PQ}$$

เมื่อ n ขนาดตัวอย่าง

χ^2 ค่าไคกำลังสองที่ระดับขั้นความเสรีเท่ากับ 1 และระดับความเชื่อมั่น 95%

N ขนาดประชากร

P สัดส่วนของลักษณะที่สนใจในประชากร (P = 0.50)

Q สัดส่วนของลักษณะที่ไม่สนใจในประชากร = 1 – P = 1 – 0.05 = 0.50

e เปอร์เซ็นต์ของความคลาดเคลื่อนที่ยอมรับได้ (e = 0.05)

ตัวอย่างที่ 5.1 สมมติว่าประชากรในการศึกษาสัดส่วนของผู้ไปลงคะแนนเสียงเลือกตั้งสมาชิกสภาผู้แทนราษฎรเป็น 100,000 คน และทราบจากการเลือกตั้งครั้งก่อนว่าสัดส่วนผู้ไปใช้สิทธิ์เป็น 0.40 จงหาขนาดตัวอย่างในการศึกษาสัดส่วนของผู้ไปใช้สิทธิ์เลือกตั้ง ถ้ากำหนดเปอร์เซ็นต์ของความคลาดเคลื่อนเป็น 0.05

วิธีทำ สิ่งที่โจทย์กำหนด

จำนวนประชากร (N) = 100,000

สัดส่วนของประชากร (P) = 0.50

ดังนั้น Q = 1 – P = 0.50

เปอร์เซ็นต์ความคลาดเคลื่อน = 0.05

ค่าไคกำลังสอง (χ^2) ที่ระดับขั้นความเสรี 1 = 3.841

จาก
$$n = \frac{\chi^2 NPQ}{e^2(N-1)+\chi^2 PQ}$$
$$= \frac{(3.841)(100{,}000)(0.50)(0.50)}{(0.05)^2(100{,}000-1)+(3.841)(0.50)(0.50)}$$
$$\approx 382.63$$

ดังนั้น ในการศึกษาสัดส่วนของผู้ไปใช้สิทธิ์เลือกตั้งครั้งนี้ควรใช้ขนาดตัวอย่างประมาณ 383 หน่วยตัวอย่าง

5.2.3 การกำหนดขนาดตัวอย่างในกรณีไม่ทราบขนาดประชากร แบ่งเป็น 2 กรณี

5.2.3.1 กรณีไม่ทราบขนาดประชากรแต่ทราบว่าประชากรมีเป็นจำนวนมาก การกำหนดขนาดตัวอย่างทำได้โดยใช้สูตร (W. G. Cochran, 1953)

$$n = \frac{P(1-P)Z^2}{d^2}$$

เมื่อ n แทนขนาดตัวอย่าง

P แทนสัดส่วนของประชากร (ซึ่งทราบมาก่อน)

Z แทนค่าสถิติ z ที่ระดับนัยสำคัญที่กำหนด 0.05 หรือ 0.01

d แทนสัดส่วนของความคลาดเคลื่อนที่ยอมให้เกิดขึ้นได้

ตัวอย่างที่ 5.2 ถ้าต้องการประมาณสัดส่วนของคนไทยที่เห็นด้วยกับการพยุงราคาน้ำมันของรัฐบาล โดยต้องการให้มีความผิดพลาดไม่เกิน 10% ด้วยระดับความเชื่อมั่น 95% ควรสุ่มตัวอย่างคนไทยมากี่คน ถ้าสมมติว่าสัดส่วนในการศึกษาก่อนหน้านี้เป็น 0.55

วิธีทำ ในที่นี้สัดส่วน $P = 0.55$

$z = 1.96$ ที่ระดับความเชื่อมั่น 95 %

$d = 0.10$

$$n = \frac{P(1-P)Z^2}{d^2}$$

$$= \frac{0.55(1-0.55)(1.96)^2}{(0.10)^2}$$

$$= 95.07$$

$$\approx 96$$

ดังนั้น ควรสุ่มตัวอย่างคนไทยในการศึกษาครั้งนี้ประมาณ 96 คน

5.2.3.2 การกำหนดขนาดตัวอย่างในกรณีไม่ทราบขนาดของประชากร หรือประชากรมีจำนวนไม่แน่นอน การกำหนดขนาดตัวอย่างทำได้โดยใช้สูตร

$$n = \frac{(z)(s)^2}{e}$$

เมื่อ n แทนขนาดตัวอย่าง

Z แทนค่าสถิติ z ที่ระดับนัยสำคัญที่กำหนด 0.05 หรือ 0.01

S ส่วนเบี่ยงเบนมาตรฐาน

e แทนความคลาดเคลื่อนมากที่สุดที่ยอมรับได้

ตัวอย่างที่ 5.3 สำนักวิจัยของบริษัทจำหน่ายรถยนต์แห่งหนึ่ง ต้องการประมาณรายได้เฉลี่ยต่อเดือนของประชาชน เพื่อนำมาตัดสินใจว่าควรผลิตรถรุ่นใหม่ที่ราคาสูงจำหน่ายในประเทศ หรือไม่ เมื่อต้องการให้รายได้เฉลี่ยของตัวอย่างแตกต่างจากรายได้เฉลี่ยของประชากรไม่เกิน 1,000 บาท และส่วนเบี่ยงเบนมาตรฐาน 100 ที่ระดับนัยสำคัญ 0.01 ควรใช้ตัวอย่างขนาดเท่าใด

วิธีทำ เมื่อกำหนดระดับนัยสำคัญ 0.01 แสดงว่าต้องการความเชื่อมั่นถึง 0.99

จะทำให้ได้ $z = z_{0.975} = 1.96$

$$e = 1000$$

$$s = 100$$

$$n = \frac{(z)(s)^2}{e}$$

$$n = \frac{(1.96)(100)^2}{1000}$$

$$= 19.6$$

$$\approx 20$$

ดังนั้น ในการศึกษานี้ ควรใช้ขนาดตัวอย่าง 20

5.2.3.3 การกำหนดขนาดตัวอย่างในกรณีที่ทราบจำนวนประชากรแน่นอน การกำหนดขนาดตัวอย่างในกรณีนี้ทำได้โดยใช้สูตรทาโร ยามาเน (Yamane, 1970) ดังนี้

$$n = \frac{N}{1 + Ne^2}$$

เมื่อ n แทนขนาดตัวอย่าง

N แทนขนาดประชากร

e แทนความคลาดเคลื่อนของการสุ่มตัวอย่าง

ตัวอย่างที่ 5.4 การวางแผนสำรวจภาวะสุขภาพของประชาชนในพื้นที่ที่มีประชากร 5,000 คน เมื่อต้องการให้เกิดความผิดพลาดของการสุ่ม 0.05 ควรใช้ตัวอย่างขนาดเท่าใด

วิธีทำ ขนาดประชากร (N) = 5000

ความผิดพลาดของการสุ่ม (e) = 0.05

$$n = \frac{N}{1 + Ne^2}$$

$$= \frac{5000}{1 + (5000)(0.05)^2}$$

$$= 370.37$$

$$\approx 371$$

ควรใช้ขนาดตัวอย่างประมาณ 371

5.3 เทคนิคการชักตัวอย่าง

วิธีการเลือกหน่วยตัวอย่างจากประชากรที่นิยมใช้กัน แบ่งออกเป็น 2 ประเภท (กัลยา วานิชย์บัญชา, 2540; ยุทธ์ ไกยวรรณ์, 2546; มนตรี สังข์ทอง, 2557) ประกอบด้วยการชักตัวอย่างที่ไม่ใช้ความน่าจะเป็น (non probability sampling) และการชักตัวอย่างที่ใช้ความน่าจะเป็น (probability sampling)

5.3.1 การชักตัวอย่างที่ไม่ใช้ความน่าจะเป็น

การชักตัวอย่างที่ไม่ใช้ความน่าจะเป็น เป็นการเลือกหน่วยตัวอย่างที่ไม่ทราบว่าหน่วยตัวอย่างในประชากรถูกเลือกมาเป็นสมาชิกของตัวอย่างด้วยความน่าจะเป็นเท่าไร การเลือกตัวอย่างแบบนี้จึงไม่ต้องทราบรายชื่อของทุกหน่วยตัวอย่างในประชากร เป็นการประหยัดเวลาและค่าใช้จ่าย วิธีที่นิยมใช้กัน ได้แก่ การชักตัวอย่างแบบตามสะดวก (convenience sampling) การชักตัวอย่างแบบโควต้า (quota sampling) และการชักตัวอย่างแบบใช้วิจารณญาน (judgement sampling)

5.3.1.1 การชักตัวอย่างแบบตามสะดวก เป็นการเลือกหน่วยตัวอย่างที่ไม่มีหลักเกณฑ์ยุ่งยากใด ๆ จะเลือกใครก็ได้ที่สามารถให้ข้อมูลเกี่ยวกับเรื่องที่ต้องการศึกษา ผู้ศึกษาเพียงแต่เลือกหน่วยตัวอย่างตามความสะดวกจนครบจำนวนหน่วยตัวอย่างที่ต้องการ การสุ่มตัวอย่างแบบนี้เป็นการเลือกตัวอย่างมาตามบุญตามกรรม (บุญธรรม กิจปรีดาบริสุทธิ์, 2535) และนิยมใช้อย่างยิ่งในการศึกษาเกี่ยวกับพฤติกรรมของผู้บริโภค (มนตรี สังข์ทอง, 2557) เช่น การประเมินความพึงพอใจของผู้เข้าชมงานในงานสัปดาห์วิทยาศาสตร์ ณ มหาวิทยาลัยราชภัฏนครศรีธรรมราช ผู้ศึกษาหรือผู้ประเมินเพียงแต่เลือกผู้ที่เข้ามาชมงานในวันนั้นแล้วก็สอบถามรายละเอียดจนครบจำนวนที่ต้องการ ก็เพียงพอแล้ว หรือต้องการศึกษาความพึงพอใจของผู้ใช้บริการห้างสรรพสินค้า นักวิจัยก็เพียงนำแบบสอบถามไปสอบถามลูกค้าที่ใช้บริการผู้ใดก็ได้จนครบตามขนาดตัวอย่างที่กำหนดไว้

5.3.1.2 การชักตัวอย่างแบบโควตา เป็นการเลือกตัวอย่างที่ไม่จำเป็น ต้องมีกรอบตัวอย่าง มีแต่เพียงการกำหนดคุณสมบัติหรือลักษณะของหน่วยตัวอย่างที่ต้องการไว้ล่วงหน้าเท่านั้น เช่น ต้องการศึกษาปัญหาเกี่ยวกับการฝึกสหกิจวิชาชีพนักศึกษา มหาวิทยาลัยราชภัฏนครศรีธรรมราช คุณสมบัติของหน่วยตัวอย่างที่ต้องกำหนดก็คือ ต้องเป็นนักศึกษาที่ผ่านการฝึกสหกิจวิชาชีพมาแล้ว ดังนั้น เมื่อพบนักศึกษาที่ผ่านการฝึกฝึกสหกิจวิชาชีพก็สามารถสอบถามรายละเอียดจากหน่วยตัวอย่างได้จนครบตามขนาดตัวอย่างที่ต้องการ หรือกรณีการประเมินความพึงพอใจของผู้เข้าชมงานในงานสัปดาห์วิทยาศาสตร์ ณ มหาวิทยาลัยราชภัฏนครศรีธรรมราช ผู้ประเมินพิจารณาคุณลักษณะของตัวอย่างออกเป็นกลุ่มย่อยโดยใช้ระดับการศึกษาเป็นตัวจำแนกเป็นระดับประถมศึกษา

มัธยมศึกษาตอนต้น มัธยมศึกษาตอนปลายและอุดมศึกษา ในการดำเนินการเก็บรวบรวมข้อมูลก็จัดสรรโควตาให้ตามระดับการศึกษาที่จำแนกไว้

5.3.1.3 การชักตัวอย่างแบบใช้วิจารณญาณ เป็นการเลือกตัวอย่างโดยการใช้ดุลยพินิจของผู้วิจัยเป็นเกณฑ์ในการเลือกตัวอย่าง ตัวอย่างที่สุ่มมานี้เป็นไปตามวัตถุประสงค์ของงานวิจัย เช่น งานวิจัยของ ยุทธ ไกยวรรณ์ เรื่องการสร้างแบบประเมินผลวิชาโครงงานเฉพาะงานช่าง กลุ่มการงานและอาชีพตามหลักสูตรมัธยมศึกษาตอนต้น พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533) เขาได้กำหนดความมุ่งหมายของการวิจัยไว้ว่า “เพื่อสร้างแบบประเมินผลวิชาโครงงานเฉพาะงานช่าง กลุ่มการงานและอาชีพตามหลักสูตรมัธยมศึกษาตอนต้น พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533) ตามความคิดเห็นของผู้เชี่ยวชาญ” ตัวอย่างที่ใช้ในการวิจัย ผู้วิจัยได้เลือกตัวอย่างแบบใช้วิจารณญาณ คือได้กำหนดให้ผู้ให้ข้อมูลคือผู้เชี่ยวชาญ และผู้ทรงคุณวุฒิ (ยุทธ ไกยวรรณ์, 2546) หรือการศึกษาพฤติกรรมของผู้ป่วยโรคไข้เลือดออกผู้ให้ข้อมูลจะต้องเป็นผู้ที่ป่วยหรือเคยป่วยโรคไข้เลือดออก (ศุภวรรณ พรหมเพรา และคณะ, 2559) เท่านั้น ซึ่งการเลือกตัวอย่างแบบนี้ผู้วิจัยจะต้องมีความรู้ความสามารถที่จะยืนยันได้ว่าตัวอย่างที่เลือกมีสมบัติตามที่ระบุไว้จริง

5.3.2 การชักตัวอย่างที่ใช้ความน่าจะเป็น

การชักตัวอย่างที่ใช้ความน่าจะเป็น (probability sampling) เป็นการเลือกหน่วยตัวอย่างจากประชากร โดยที่สามารถบอกความน่าจะเป็นที่แต่ละหน่วยตัวอย่างในประชากรจะถูกเลือกมาเป็นตัวอย่าง การเลือกตัวอย่างแบบนี้จะต้องทราบรายชื่อของทุกหน่วยตัวอย่างและกรอบของประชากรหรือกรอบตัวอย่าง (sampling frame) ซึ่งโดยทั่วไปทำได้ 2 แบบ คือ การชักตัวอย่างแบบคืนที่ (sampling with replacement) และการชักตัวอย่างแบบไม่คืนที่ (sampling without replacement)

5.3.2.1 การชักตัวอย่างแบบคืนที่ เป็นการเลือกตัวอย่างที่ทำโดยการเลือกหน่วยตัวอย่างจากประชากรที่กำหนด เมื่อเก็บข้อมูลหรือรายละเอียดที่ต้องการจากหน่วยตัวอย่างนั้นแล้วให้หน่วยตัวอย่างนั้นกลับคืนสู่ประชากรอีกครั้งก่อนการเลือกหน่วยตัวอย่างครั้งต่อไป ทำอย่างนี้ไปจนครบตามขนาดตัวอย่างที่ต้องการ เช่นการเลือกหน่วยตัวอย่างเป็นวารสารเล่มหนึ่ง สมมติเป็นวารสารหมายเลข 5 มาตรวจสอบคุณภาพตามที่ระบุไว้ แล้วนำวารสารฉบับนั้นกลับคืนสู่ประชากร เมื่อเลือกหน่วยตัวอย่างต่อไปวารสารหมายเลข 5 อาจได้รับการเลือกมาเป็นหน่วยตัวอย่างอีกก็ได้ เมื่อเป็นเช่นนั้นแล้วจะทำให้ขนาดตัวอย่างลดลง ปกติแล้วไม่นิยมการเลือกตัวอย่างแบบนี้ในงานวิจัย

5.3.2.2 การชักตัวอย่างแบบไม่คืนที่ เป็นการเลือกตัวอย่างที่ทำโดยการเลือกหน่วยตัวอย่างใดจากประชากรมาแล้ว จะเก็บข้อมูลหรือรายละเอียดที่ต้องการจากหน่วยตัวอย่างนั้นแล้ว ไม่มี

การนำหน่วยตัวอย่างคืนสู่ประชากร ก่อนการเลือกครั้งต่อไป เช่น การเลือกนักศึกษา 50 คน จากนักศึกษาที่เรียนรายวิชาสถิติวิเคราะห์ 500 คน เพื่อสอบถามถึงความพึงพอใจในวิธีการสอนของอาจารย์ผู้สอน เมื่อเลือกนักศึกษาคนที่ 1 มาสอบถามรายละเอียดหรือข้อมูลที่ต้องการแล้ว ก็จะไม่มีการสอบถามรายละเอียดจากนักศึกษาคนนี้อีก เนื่องจากไม่ได้ส่งนักศึกษาคนนี้เข้าสู่ประชากรก่อนการเลือกหน่วยตัวอย่างต่อไป นั่นคือ นักศึกษา 1 คน มีโอกาส ถูกเลือกมาเป็นหน่วยตัวอย่างเพียงครั้งเดียว ในการศึกษาวิจัยเรื่องหนึ่ง ๆ เท่านั้น ในทางปฏิบัติ จะต้องเลือกตัวอย่างให้ครบตามขนาดของตัวอย่างที่กำหนด ก่อนการสอบถามรายละเอียด มิใช่เลือกมา 1 หน่วยตัวอย่าง แล้วสอบถาม แล้วเลือกหน่วยตัวอย่างที่ 2 แล้วสอบถามรายละเอียด อย่างไรก็ตามการเลือกตัวอย่างที่นิยมใช้ในการทำวิจัยนั้น นิยมใช้การเลือกตัวอย่างตามความน่าจะเป็นและมีเทคนิคการเลือกตัวอย่างที่นิยมใช้ 5 แบบ ดังนี้

1) การชักตัวอย่างอย่างง่าย (simple random sampling)
2) การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ (stratified sampling)
3) การชักตัวอย่างแบบมีระบบ (systematic sampling)
4) การชักตัวอย่างแบบเกาะกลุ่ม (cluster sampling)
5) การชักตัวอย่างแบบหลายขั้น (multi-stage sampling)

5.4 การชักตัวอย่างอย่างง่าย

การชักตัวอย่างอย่างง่าย เป็นการเลือกตัวอย่างจากประชากรที่กำหนดให้หน่วยตัวอย่างแต่ละหน่วยถูกเลือกมาเป็นตัวอย่างด้วยความน่าจะเป็นเท่ากัน การเลือกตัวอย่างแบบนี้ใช้ได้ผลดี ในกรณีที่ประชากรมีความแตกต่างกันไม่มากนักหรือหน่วยตัวอย่างในประชากรคล้ายคลึงกัน เช่น ในการสอบถามนักศึกษาชั้นปีที่ 2 จำนวน 1,000 คน เกี่ยวกับการให้บริการอินเทอร์เน็ตของสำนักวิทยบริการ มหาวิทยาลัยราชภัฏนครศรีธรรมราช ประชากรก็คือ นักศึกษาชั้นปีที่ 2 มหาวิทยาลัยราชภัฏนครศรีธรรมราช สมมติว่าต้องการนักศึกษาชั้นปีที่ 2 จำนวน 100 คน มาเป็นตัวแทนในการศึกษาเรื่องนี้ วิธีการได้ตัวอย่างจากการชักตัวอย่างแบบนี้ทำได้ 2 วิธี คือ การจับฉลาก และการใช้ตารางเลขสุ่ม

5.4.1 การจับฉลาก

การจับฉลากเพื่อเลือกหน่วยตัวอย่างด้วยเทคนิคหรือวิธีการชักตัวอย่างอย่างง่าย ทำได้โดยการทำฉลากให้นักศึกษาชั้นปีที่ 2 ทุกคนซึ่งอาจจะใช้รหัสประจำตัวของนักศึกษาแล้วนำฉลากเหล่านั้น มาคละปนกันอย่างดี จับฉลากโดยสุ่มทีละใบจนครบ 100 ใบ ก็จะได้ตัวแทน 100 คน ตามที่ต้องการ จะเห็นได้ว่าความน่าจะเป็นที่นักศึกษาแต่คนจะถูกเลือกมาเป็นตัวแทนจะเท่ากับ

$\frac{1}{1000}$ ต่อจากนั้นทำการสอบถามรายละเอียดหรือสัมภาษณ์รายละเอียดที่ต้องการจาก 100 หน่วยตัวอย่างซึ่งเป็นตัวแทนจากประชากรในการศึกษาครั้งนี้

5.4.2 การใช้ตารางเลขสุ่ม

การใช้ตารางเลขสุ่มเพื่อเลือกหน่วยตัวอย่างด้วยเทคนิคหรือวิธีการชักตัวอย่างอย่างง่าย ทำได้โดยการกำหนดหมายเลขให้กับทุกหน่วยตัวอย่างในประชากร ในที่นี้มีนักศึกษาชั้นปีที่ 2 จำนวน 1,000 คน ก็กำหนดหมายเลขเป็น 3 หลัก คือ 000, 001, 002, ...,999 แล้วสุ่มเลือกแถวและสดมภ์เริ่มต้นจากตารางเลขสุ่ม โดยทั่วไปตารางเลขสุ่มจะถูกสร้างขึ้นโดยการลำดับตัวเลขจาก 0-9 อย่างสุ่ม ๆ หมายความว่าตัวเลขทุกตัวมีโอกาสปรากฏในตำแหน่งต่าง ๆ ด้วยความน่าจะเป็นเท่ากัน เพื่อป้องกันการลำเอียง อาจใช้การจับฉลากแถวและสดมภ์เริ่มต้นนั้นก็ได้ ตัวอย่างการใช้ตารางเลขสุ่ม ในกรณีสุ่มตัวอย่างนักศึกษาชั้นปีที่ 2 จำนวน 100 คน จากนักศึกษา 1,000 คน ให้หมายเลขกับหน่วยตัวอย่างเป็น 000-999 ดังนั้น ควรใช้ตัวเลข 3 หลักในการเลือกหน่วยตัวอย่างโดยใช้ตารางเลขสุ่มดังตัวอย่างต่อไปนี้

สมมติว่าแถวและสดมภ์ของตารางเลขสุ่มเริ่มต้นที่สุ่มได้คือ แถวที่ 40 สดมภ์ที่ 5 ดังนั้นจะได้หน่วยตัวอย่างที่ 1–100 เป็นหน่วยตัวอย่างที่มีหมายเลข 263, 251, 254, 710, 259, 120, 361, 203, 615, 490, 145, 891, 452, 194, 523,... ไปเรื่อย ๆ จนถึงหมายเลขของหน่วยตัวอย่างที่ 100 กรณีที่หมายเลขที่ได้ซ้ำกันก็จะข้ามไปเลือกหมายเลขถัดไปจนครบ 100 หน่วยตัวอย่าง ต่อจากนั้น ก็ทำการสอบถามรายละเอียดที่ต้องการจากหน่วยตัวอย่างที่ได้

ตารางที่ 5.1 ตัวอย่างส่วนหนึ่งของตารางเลขสุ่ม

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
30	25468	54687	45879	54876	45479	21577	48756	01244	21548	32697
31	72365	93223	63120	21002	89654	96543	46523	69871	79658	59458
32	54652	10425	02154	54879	01226	02145	87542	25463	75423	89631
33	15402	47986	78965	63548	34879	78964	23972	02154	00214	25478
34	15101	20114	84578	79987	97540	58722	40215	78966	56998	96632
35	56897	02154	45213	12457	14587	87546	14254	12025	45678	45462
36	58894	89766	98756	89635	52639	99325	89632	47893	96165	58755
37	62130	32548	32154	21021	36569	13201	58748	30125	60252	64856
38	21364	79645	69782	54789	25423	25478	52369	47962	54620	31323
39	78979	32547	25140	66321	02146	96438	87542	54388	31250	01254
40	20325	26325	12547	10259	12036	12036	15490	14589	14521	94523
41	48793	98745	85369	68742	98756	98754	32697	63205	36978	10231
42	13321	12031	58472	36525	32102	06435	56845	14512	51203	54789
43	55245	25875	15032	03698	36598	78932	60254	69873	21562	45030
44	43687	92564	69875	75412	74521	06597	62301	25016	05987	21558
45	59463	20154	54578	12432	78403	02154	50306	87965	14253	10247
46	21026	78960	63202	15625	26954	87658	98740	03145	69875	89635
47	54789	20359	54598	12547	21547	56423	15036	21032	21548	24852
48	56456	78540	72369	85125	82150	20215	98765	59754	96325	36970
49	01254	21365	87542	90265	36970	48789	54602	69870	41258	36854

5.5 การชักตัวอย่างแบบมีระบบ

การชักตัวอย่างแบบมีระบบ เป็นวิธีการเลือกตัวอย่างที่เหมาะสมกับประชากรที่มีหน่วยตัวอย่างอยู่ในลักษณะที่ลำดับไว้อย่างเรียบร้อยเป็นระเบียบ เช่น การสุ่มร้านอาหารเพื่อสอบถามถึงลูกค้าที่เข้ามารับบริการในแต่ละวัน จากร้านสวัสดิการของโรงพยาบาลแห่งหนึ่ง มีขั้นตอนดังต่อไปนี้

สมมติมีร้านอาหารอยู่ 20 ร้าน (N) ต้องการเลือกร้านอาหารมา 6 ร้าน (n) แต่ละร้านมีหมายเลขประจำร้านเป็น 1, 2, 3, ..., 20

ขั้นตอนที่ 1 เลือกหน่วยตัวอย่างเริ่มต้น (random start) อาจจะใช้การจับฉลาก สมมติว่าจับฉลากได้หน่วยตัวอย่างเริ่มต้นเป็น 5

ขั้นตอนที่ 2 คำนวณหาค่าคงที่ (fixed interval)

จาก $$\frac{N}{n} = \frac{20}{6} = 3.3 \approx 3$$

ร้านจำหน่ายอาหาร 20 ร้าน ตั้งอยู่เรียงกันดังนี้

1	2	3	4	**5**	6	7	**8**	9	10
11	12	13	**14**	15	16	**17**	18	19	**20**

แสดงว่าหน่วยตัวอย่างที่สุ่มได้คือร้านที่

5, 5 + 3, 8 + 3, 11 + 3, 14 + 3, 17 + 3

นั่นคือ จะได้ร้านต่าง ๆ ที่มีหมายเลขดังนี้

5, 8, 11, 14, 17, 20

จะเห็นได้ว่าการเลือกตัวอย่างแบบมีระบบ ทำให้ผู้เลือกเสียเวลาน้อย เพราะเพียงแต่กำหนดหน่วยเริ่มต้นเท่านั้น หน่วยตัวอย่างอื่นก็จะได้มาตามระบบ

5.6 การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ

การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ เป็นการเลือกหน่วยตัวอย่างในกรณีที่ประชากรมีความแตกต่างกันมาก จึงต้องทำการแบ่งประชากรเป็นชั้นภูมิ (stratum) เสียก่อน โดยให้แต่ละชั้นมีความแตกต่างระหว่างชั้นมากที่สุด แต่มีความคล้ายคลึงกันภายในชั้นมากที่สุด หรือมีข้อมูลที่สนใจค่อนข้าง

เหมือนกันในชั้นภูมิเดียวกัน เช่น การแบ่งประชากรตาม ภาคภูมิศาสตร์ ชั้นปี ศาสนา อาชีพ การศึกษา เป็นต้น ตัวอย่างการศึกษาเกี่ยวกับความเห็นของนักศึกษาที่มีต่อการบริหาร กองพัฒนานักศึกษา มหาวิทยาลัยราชภัฏนครศรีธรรมราช

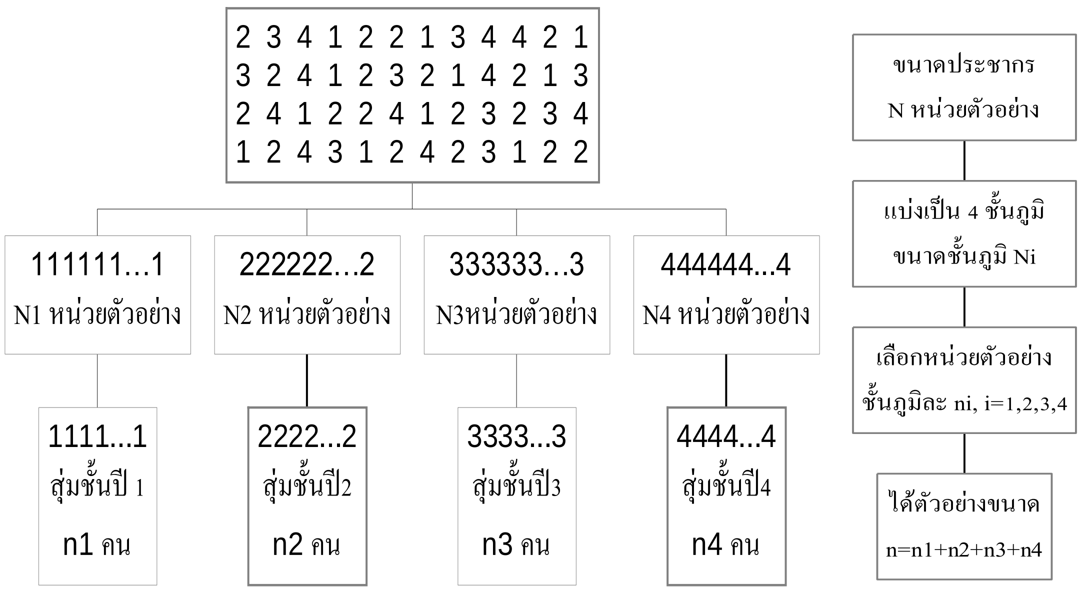
ขั้นตอนที่ 1 แบ่งประชากรนักศึกษาออกเป็น 4 ชั้นปี (4 ชั้นภูมิ) นักศึกษาใน 4 ชั้นปี มีความแตกต่างกันในพื้นฐานของชั้นปี แต่นักศึกษาในชั้นปีเดียวกันมีความแตกต่างกันน้อย

ขั้นตอนที่ 2 ถ้าต้องการตัวอย่างขนาด n จากประชากร N หน่วย โดย $N_1 + N_2 + N_3 + N_4 = N$ เมื่อ N_i คือจำนวนนักศึกษาในแต่ละชั้นปี และ $i = 1, 2, 3, 4$ ก็อาจเลือกตัวอย่างจากแต่ละชั้นมา n_1, n_2, n_3, n_4 ตามลำดับ โดยที่ $n_1 + n_2 + n_3 + n_4 = n$ วิธีการเลือกตัวอย่างจากแต่ละชั้นภูมิอาจจะใช้วิธีการเลือกตัวอย่าง 2 แบบ คือ

1. การชักตัวอย่างอย่างง่ายในชั้นภูมิ (stratified simple random sampling) เป็นการเลือกหน่วยตัวอย่างโดยใช้การสุ่มตัวอย่างอย่างง่ายจากในแต่ละชั้นภูมิ

2. การชักตัวอย่างแบบมีระบบในชั้นภูมิ (stratified systematic random sampling) เป็นการเลือกหน่วยตัวอย่างโดยใช้การสุ่มตัวอย่างแบบมีระบบในแต่ละชั้นภูมิ

เมื่อดำเนินการตามขั้นตอนดังกล่าวทั้ง 2 ขั้นตอนแล้ว จะได้ตัวอย่างขนาด n ตามความต้องการ ขั้นตอนการดำเนินการเป็นไปดังภาพที่ 5.2



ภาพที่ 5.2 แผนผังแสดงการชักตัวอย่างแบบชั้นภูมิขนาด n จากประชากรขนาด N

ส่วนการจัดสรรขนาดตัวอย่าง (allocation of sample size) จากแต่ละชั้นแบ่งได้ 4 วิธี ดังนี้ (ระพีพรรณ พิริยะกุล, 2536;)

1. การจัดสรรแบบตามความใจชอบ (arbitrary allocation) วิธีการนี้จะจัดสรรหน่วยตัวอย่างให้กับแต่ละชั้นภูมิตามใจชอบไม่มีเกณฑ์ใด ๆ ซึ่งวิธีการนี้เหมาะกับผู้ที่มีประสบการณ์ด้านการวิจัยในเรื่องนั้น ๆ มากพอสมควรและคุ้นเคยกับประชากรนั้น

2. การจัดสรรแบบแบ่งให้เท่า ๆ กัน (equal allocation) วิธีการนี้จะจัดสรรหน่วยตัวอย่างให้กับแต่ละชั้นภูมิเท่า ๆ กัน ทำได้ง่าย สะดวก เช่น ถ้าต้องการตัวอย่างขนาด (n) 500 ประชากรถูกแบ่งเป็น 4 ชั้นภูมิ หน่วยตัวอย่างจากในแต่ละชั้นภูมิก็จะเป็น $\frac{n}{4} = \frac{500}{4} = 125$ วิธีการนี้ไม่เหมาะสมในกรณีที่หน่วยตัวอย่างในบางชั้นมีน้อย เช่น ในกรณีตัวอย่างข้างต้น ขนาดตัวอย่างที่ต้องการ 500 ประชากรถูกแบ่งเป็น 4 ชั้น สมมติว่ามีอยู่หนึ่งชั้นภูมิที่มีประชากรอยู่เพียง 100 หน่วยตัวอย่าง แต่ที่ต้องการคือหน่วยตัวอย่างจากชั้นภูมิละ 125 หน่วยตัวอย่าง ทำให้ได้หน่วยตัวอย่างจากชั้นภูมินั้นไม่ครบและทำให้ได้ขนาดตัวอย่างในการศึกษาครั้งนั้น ๆ ไม่ครบตามที่ต้องการ อีกทั้งบางชั้นภูมิที่มีหน่วยตัวอย่างมาก ๆ ก็จะได้หน่วยตัวอย่างมาเป็นตัวแทนน้อยเกินไป ทำให้ได้ข้อมูลไม่เพียงพอที่จะเป็นตัวแทนที่ดีในชั้นภูมินั้น

3. การจัดสรรแบบสัดส่วน (proportional allocation) วิธีการนี้เป็นวิธีการจัดสรรหน่วยตัวอย่างให้กับแต่ละชั้นภูมิที่ช่วยแก้ปัญหา วิธีการที่ 2 ทั้ง 2 กรณี โดยมีหลักเกณฑ์การจัดสรรดังนี้

สมมติมีประชากร คือนักศึกษามหาวิทยาลัยราชภัฏนครศรีธรรมราช ภาคปกติ ซึ่งถูกแบ่งเป็น 5 คณะ กำหนดให้คณะเป็นชั้นภูมิ แสดงว่าประชากรถูกแบ่งเป็น 5 ชั้นภูมิ ให้ n_i คือขนาดตัวอย่างในชั้นภูมิที่ i เมื่อ i = 1, 2, 3, 4, 5 รายละเอียดจำนวนนักศึกษาแต่ละคณะ เป็นดังนี้

ตารางที่ 5.2 จำนวนนักศึกษา ภาคปกติ มหาวิทยาลัยราชภัฏนครศรีธรรมราช ปีการศึกษา 2557

คณะ	จำนวน(คน)	ชั้นภูมิที่
ครุศาสตร์	2,391	1
วิทยาศาสตร์และเทคโนโลยี	1,225	2
มนุษยศาสตร์และสังคมศาสตร์	3,732	3
วิทยาการจัดการ	2,170	4
เทคโนโลยีอุตสาหกรรม	1,141	5
รวม	10,659	

(ที่มา: สำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏนครศรีธรรมราช, 7 พฤศจิกายน 2557)

ถ้าขนาดตัวอย่างที่ต้องการคือ 500 คน (n) จะจัดสรรได้ดังนี้

$$n_1 = 2391 \times \frac{500}{10695} = 112.16 = 112 \text{ คน}$$

$$n_2 = 1225 \times \frac{500}{10695} = 57.46 = 57 \text{ คน}$$

$$n_3 = 3732 \times \frac{500}{10695} = 175.06 = 175 \text{ คน}$$

$$n_4 = 2170 \times \frac{500}{10695} = 101.79 = 102 \text{ คน}$$

$$n_5 = 1141 \times \frac{500}{10695} = 53.52 = 54 \text{ คน}$$

รวม 500 คน

4. การจัดสรรตัวอย่างแบบเหมาะสมที่สุด (optimum allocation) วิธีการนี้จะจัดสรรหน่วยตัวอย่างให้กับแต่ละชั้นภูมิ โดยมีกฎเกณฑ์ ข้อสังเกต ข้อจำกัดเพิ่มขึ้นในเรื่องของค่าใช้จ่ายโดยกำหนดฟังก์ชันค่าใช้จ่าย เป็นดังนี้

$$C = C_o + \sum_{i=1}^{h} c_i n_i$$

เมื่อ C คือ ค่าใช้จ่ายทั้งหมด

C_o คือ ค่าใช้จ่ายคงที่

c_i คือ ค่าใช้จ่ายต่อ 1 หน่วยในชั้นภูมิที่ i

n_i คือ จำนวนหน่วยตัวอย่างในชั้นภูมิที่ i

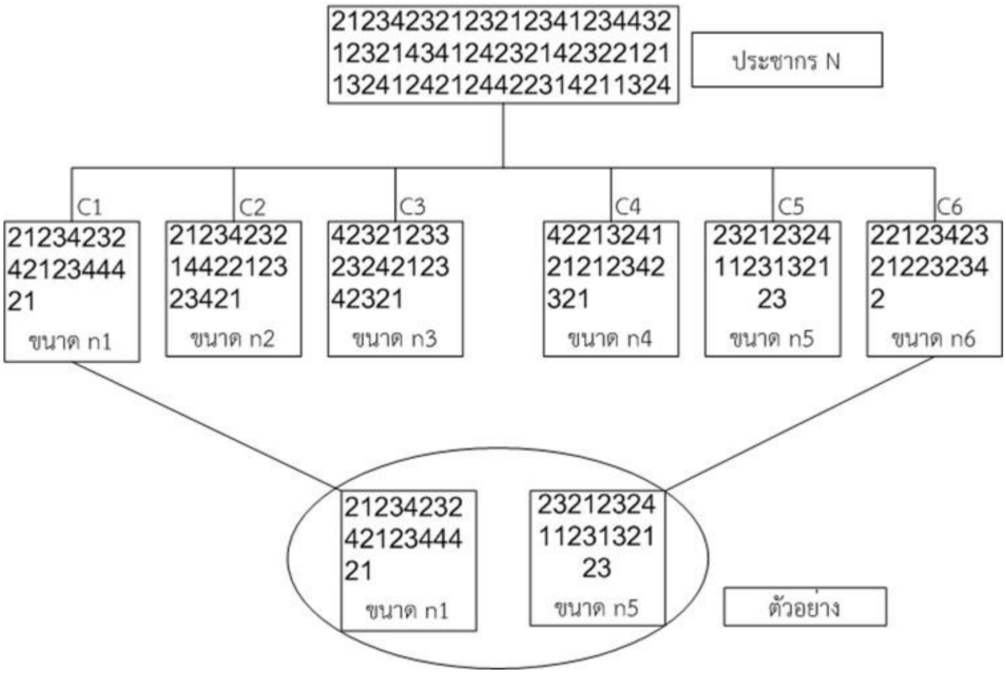
h คือ จำนวนชั้นภูมิ

อย่างไรก็ตาม ในทางปฏิบัตินิยมใช้การจัดสรรตัวอย่างในแต่ละชั้นด้วยการจัดสรร แบบสัดส่วนเพราะทำให้ชั้นที่มีหน่วยตัวอย่างมากน้อยต่างกัน ได้รับการจัดสรรหน่วยตัวอย่างในสัดส่วนที่เหมาะสมมากกว่าแบบอื่น ๆ

5.7 การชักตัวอย่างแบบกลุ่ม

การชักตัวอย่างแบบกลุ่ม เป็นการเลือกตัวอย่างที่เหมาะกับกรณีที่หน่วยตัวอย่าง ในประชากรมีความคล้ายคลึงกัน เมื่อแบ่งประชากรเป็นกลุ่มย่อยตามเกณฑ์ที่กำหนดแล้วประชากรในแต่ละกลุ่มย่อยจะประกอบด้วยหน่วยตัวอย่างมีลักษณะต่าง ๆ ที่สนใจทุกลักษณะคละกัน เรียกแต่ละกลุ่มย่อยนั้นว่ากลุ่มหรือคลัสเตอร์ (cluster) เช่น ผู้วิจัยต้องการศึกษาเกี่ยวกับค่าใช้จ่ายของนักศึกษาขณะศึกษาอยู่ในมหาวิทยาลัยราชภัฏนครศรีธรรมราช แบ่งกลุ่มย่อยโดยใช้สาขาวิชา ในที่นี้สมมติว่าแบ่งเป็น 6 กลุ่มย่อย (สาขาวิชา) ซึ่งพบว่าในแต่ละสาขาวิชาจะมีนักศึกษาคละกันอยู่ทั้งชั้นปี 1, 2, 3

และ 4 มีขนาดคลัสเตอร์เป็น n1 n2 n3 n4 n5 และ n6 ตามลำดับ และจะแสดงการเลือกตัวอย่างได้ดังภาพที่ 5.3



ภาพที่ 5.3 แผนผังแสดงการชักตัวอย่างแบบกลุ่มหรือคลัสเตอร์

ต่อจากนั้นผู้วิจัยก็จะเลือกกลุ่มย่อย (C1, C2, C3...) มาศึกษา จะเป็นกี่กลุ่มขึ้นอยู่กับความเหมาะสม โดยที่ไม่จำเป็นต้องเลือกตัวอย่างมาจากทุกกลุ่ม และอาจจะใช้การเลือกตัวอย่างแบบชักตัวอย่างอย่างง่าย หรือการเลือกตัวอย่างแบบมีระบบในการเลือกกลุ่มก็ได้ แล้วสอบถามรายละเอียดต่าง ๆ ที่ต้องการจากทุกหน่วยตัวอย่างในกลุ่มนั้น

5.8 การชักตัวอย่างแบบหลายขั้น

การชักตัวอย่างแบบหลายขั้น เป็นวิธีการเลือกตัวอย่างที่ทำได้โดยการเลือกตัวอย่างจากกลุ่มใหญ่ไปกลุ่มย่อย ๆ ลงไปเรื่อย ๆ จนได้ตัวอย่างที่ต้องการเก็บรวบรวมข้อมูล เช่น จากการศึกษาเกี่ยวกับค่าใช้จ่ายของนักศึกษาในมหาวิทยาลัยราชภัฏทั่วประเทศ ประชากรที่ทำการศึกษา คือ นักศึกษามหาวิทยาลัยราชภัฏทั้งหมด มีขั้นตอนดำเนินการเลือกตัวอย่าง ดังต่อไปนี้

ขั้นที่ 1 เลือกตัวแทนมหาวิทยาลัยราชภัฏจากมหาวิทยาลัยราชภัฏทุกแห่ง

ขั้นที่ 2 เลือกคณะในมหาวิทยาลัยราชภัฏที่ถูกเลือก

ขั้นที่ 3 เลือกสาขาวิชาในคณะที่ถูกเลือก

ขั้นที่ 4 เลือกนักศึกษาในสาขาวิชาที่ถูกเลือก

ผู้ศึกษาวิจัยก็จะได้นักศึกษาเป็นหน่วยตัวอย่างที่ต้องการข้อมูล การกระทำดังตัวอย่างนี้ก็จะเป็นการเลือกตัวอย่างแบบ 4 ขั้น

(เพิ่มตัวอย่าง)ทำผังการเลือก

ภาพที่ 5.4 แผนผังแสดงการชักตัวอย่างแบบหลายขั้น

พอจะกล่าวโดยสรุปได้ว่า การเลือกตัวอย่างจากประชากรในการศึกษาวิจัยเรื่องใด ๆ ผู้วิจัยจะเลือกใช้วิธีการเลือกตัวอย่างแบบใดนั้นควรคำนึงถึงสิ่งต่าง ๆ ดังนี้

1. ขนาดประชากร ขนาดตัวอย่าง
2. ค่าใช้จ่าย
3. เวลา
4. การเป็นตัวแทนที่ดีของตัวอย่างที่ได้
5. กรอบตัวอย่างและความสมบูรณ์ของกรอบตัวอย่าง

5.9 การแจกแจงตัวอย่าง

การแจกแจงตัวอย่าง (sampling distribution) คือการแจกแจงความน่าจะเป็นของตัวสถิติ เช่น การแจกแจงความน่าจะเป็นของค่าเฉลี่ยของตัวอย่างจะเรียกว่าการแจกการสุ่มตัวอย่างของค่าเฉลี่ย การแจกแจงความน่าจะเป็นของสัดส่วนของตัวอย่างจะเรียกว่าการแจกการสุ่มตัวอย่างของสัดส่วน การแจกแจงการสุ่มตัวอย่างของตัวสถิติขึ้นอยู่กับขนาดของประชากร ขนาดของตัวอย่างและวิธีการเลือกตัวอย่าง (ศุภชัย นาทะพันธ์, 2547) ค่าสถิติที่กล่าวถึง คือค่าคงที่ที่คำนวณได้จากตัวอย่างสุ่ม ตัวอย่างที่สุ่มได้จากประชากรเดียวกันมีความเป็นไปได้หลายชุด เช่น การสุ่มตัวอย่าง 3 ชุดมาจากประชากรเดียวกันและคำนวณค่าเฉลี่ยของตัวอย่างแต่ละชุด จะพบว่าถ้าค่าสถิติที่สนใจคือ

ค่าเฉลี่ยของตัวอย่าง ซึ่งได้แก่ $\overline{X}_1, \overline{X}_2, \overline{X}_3$ ซึ่งแตกต่างกัน การศึกษาถึงการแจกแจงการสุ่มตัวอย่างของค่าเฉลี่ย หรือการศึกษาการแจกแจงค่าความแปรปรวนของตัวอย่าง เป็นการศึกษาว่าถ้ามีการสุ่มตัวอย่างมาหลายชุดจากประชากรเดียวกันแล้ว ค่าสถิติของตัวอย่างเหล่านี้ จะมีการแจกแจงเป็นรูปแบบใดอันจะนำไปสู่การประมาณค่าและการทดสอบสมมติฐาน เพื่อประโยชน์ในการอ้างอิงอันจะนำไปสู่การตัดสินใจ (ชูศรี วงศ์รัตนะ, 2544; สรชัย พิศาลบุตร, 2559)

ตัวสถิติหลายตัวที่ได้จากตัวอย่างที่ดีหลายตัวอย่างจากการชักตัวอย่าง จะทำให้ได้ค่าสถิติที่ใกล้เคียงกับค่าพารามิเตอร์ และจะใช้ค่าสถิติเหล่านี้สรุปไปยังประชากร ตัวอย่างหลาย ๆ ชุดที่แต่ละชุดมีขนาดเท่ากัน ทำให้ได้ค่าสถิติต่าง ๆ กันออกไป ดังนั้น ตัวสถิติจึงเป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็น ค่าคาดหวังและความแปรปรวน การแจกแจงการสุ่มตัวอย่างที่ใช้กันมากในทางสถิติซึ่งจะนำไปเป็นพื้นฐานในการประมาณค่าและการทดสอบสมมติฐาน ได้แก่ การแจกแจงของค่าเฉลี่ย การแจกแจงของผลต่างของค่าเฉลี่ย การแจกแจงแบบที การแจกแจงแบบไคกำลังสอง และการแจกแจงแบบเอฟ

5.10 การแจกแจงของค่าเฉลี่ย

ในการกล่าวถึงการแจกแจงของค่าเฉลี่ย นักสถิติได้สร้างทฤษฎีที่อาศัยการแจกแจงแบบปกติมาตรฐาน (standard normal distribution) โดยทฤษฎีดังกล่าวมีชื่อว่าทฤษฎีขีดจำกัดสู่ส่วนกลาง (central limit theorem: CLT) ซึ่งจะได้กล่าวถึงตามลำดับต่อไป

บทนิยามที่ 5.7 ให้ $X_1, X_2, X_3, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n เรียก $\overline{X}$ ว่า ค่าเฉลี่ยของตัวอย่าง (sample mean) โดยมีค่าเป็น

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

ทฤษฎีบทที่ 5.1 ให้ $X_1, X_2, X_3, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจงหนึ่ง ๆ ที่มีค่าเฉลี่ยของประชากรเป็น μ และความแปรปรวนของประชากรเป็น σ^2 จะได้ค่าเฉลี่ยของ $\overline{X}$ มีค่าเท่ากับค่าเฉลี่ยของประชากร μ และความแปรปรวนของ $\overline{X}$ มีค่าเท่ากับความแปรปรวนของประชากรหารด้วย n นั่นคือ

$$E(\overline{X}) = \mu$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

ในที่นี้จะใช้ $\mu_{\overline{X}}$ แทนค่าเฉลี่ยของ $\overline{X}$ และใช้ $\sigma^2_{\overline{X}}$ แทนความแปรปรวนของ $\overline{X}$ ดังนั้น ค่าคาดหมาย $E(\overline{X})$ และ $Var(\overline{X})$ เขียนได้ดังนี้

$$\mu_{\overline{X}} = E(\overline{X}) = \mu$$

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n}$$

บทนิยามที่ 5.8 ส่วนเบี่ยงเบนมาตรฐานของ $\overline{X}$ เรียกว่าความคลาดเคลื่อนมาตรฐาน (standard error) และใช้สัญลักษณ์ $\sigma_{\overline{X}}$ นั่นคือ

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}}$$

ตัวอย่างที่ 5.5 สมมติว่าประชากรกลุ่มหนึ่งเป็นน้ำหนักของวัสดุที่ใช้ทำตุ๊กตา 4 ตัว ให้ X_1, X_2, X_3, X_4 เป็นน้ำหนักของตุ๊กตา 4 ตัว มีค่าเป็น 200, 250, 125 และ 225 กรัม ถ้าชักตัวอย่างขนาด 2 จากประชากร จงคำนวณหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง แล้วเปรียบเทียบกับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประชากร

วิธีทำ ค่าเฉลี่ยของประชากร (μ)

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

$$= \frac{1}{4}(200 + 250 + 125 + 225)$$

$$= 200$$

ความแปรปรวนของประชากร

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

$$= \frac{(200-200)^2 + (250-200)^2 + (125-200)^2 + (225-200)^2}{4}$$

$$= 2187.5$$

ถ้าชักตัวอย่างขนาด 2 แบบใส่คืนจากประชากร ตัวอย่างที่เป็นไปได้ทั้งหมดเป็น 16 ชุด แสดงในตารางที่ 5.3 1 แสดงข้อมูลในตารางที่ 5.4 และแสดงค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละตัวอย่างในตารางที่ 5.5

ตารางที่ 5.3 ผลการชักตัวอย่างขนาด 2 แบบใส่คืน จากประชากรขนาด 4

	ผลการชักตัวอย่างครั้งที่ 1				
	ตัวอย่างที่	1	2	3	4
ผลการชักตัวอย่างครั้งที่ 2	1	(1,1)	(1,2)	(1,3)	(1,4)
	2	(2,1)	(2,2)	(2,3)	(2,4)
	3	(3,1)	(3,2)	(3,3)	(3,4)
	4	(4,1)	(4,2)	(4,3)	(4,4)

ตารางที่ 5.4 น้ำหนักวัสดุที่ใช้ทำตุ๊กตาในแต่ละตัวอย่าง

	ผลการชักตัวอย่างครั้งที่ 1				
	ตัวอย่างที่	1	2	3	4
ผลการชักตัวอย่างครั้งที่ 2	1	(200,200)	(200,250)	(200,125)	(200,225)
	2	(250,200)	(250,250)	(250,125)	(250,225)
	3	(125,200)	(125,250)	(125,125)	(125,225)
	4	(225,200)	(225,250)	(225,125)	(225,225)

นั่นคือ จะได้ตัวอย่างที่เป็นไปได้ทั้งหมด 16 ชุด มีค่าเฉลี่ยและความแปรปรวน ดังนี้

ตารางที่ 5.5 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละตัวอย่าง

ตัวอย่างที่	น้ำหนักของวัสดุในแต่ละตัวอย่าง	ค่าเฉลี่ย $(\overline{X}_i)$ i = 1, ..., 16	ความแปรปรวน $(\sigma^2_{\overline{X}})$
1	(200,200)	200.0	0
2	(200,250)	225.0	625.00
3	(200,125)	162.5	1406.25
4	(200,225)	212.5	156.25
5	(250,200)	225.0	625.00
6	(250,250)	250.0	0
7	(250,125)	187.5	3906.25

8	(250,225)	237.5	156.25
9	(125,200)	162.5	1406.25
10	(125,250)	187.5	3906.25
11	(125,125)	125.0	0
12	(125,225)	175.0	2500.00
13	(225,200)	212.5	156.25
14	(225,250)	237.5	156.25
15	(225,125)	175.0	2500.00
16	(225,225)	225.0	0

ตารางที่ 5.6 ค่าเฉลี่ย ความถี่ของค่าเฉลี่ย และความถี่สัมพัทธ์ของค่าเฉลี่ย

$(\overline{X}_i)$	ความถี่	ความถี่สัมพัทธ์ $g(\overline{X}_i)$
125.0	1	$\frac{1}{16}$
162.5	2	$\frac{2}{16}$
175.0	2	$\frac{2}{16}$
187.5	2	$\frac{2}{16}$
200.0	1	$\frac{1}{16}$
212.5	2	$\frac{2}{16}$
225.0	3	$\frac{3}{16}$
237.5	2	$\frac{2}{16}$
250.0	1	$\frac{2}{16}$
รวม		1

เมื่อพิจารณาตารางที่ 5.6 พบว่าความถี่สัมพัทธ์ของค่า $(\overline{X})$ แต่ละค่าก็คือความน่าจะเป็นของตัวแปรสุ่ม $(\overline{X})$ แต่ละตัว และตารางที่ 5.6 ก็คือตารางที่ใช้แสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม $(\overline{X})$ และหากหาผลรวมของค่าความถี่สัมพัทธ์ $g(\overline{X})$ จะเท่ากับ 1 โดยค่าเฉลี่ยของ $\overline{X}$ หาได้จาก

$$\mathrm{E}(\overline{X}) = \sum_{i=1}^{9} \overline{X}_i g\,(\overline{X})$$

$$= (125.0)\left(\frac{1}{16}\right) + (162.5)\left(\frac{2}{16}\right) + (175.0)\left(\frac{2}{16}\right) + \ldots + (250.0)\left(\frac{1}{16}\right)$$

$$= 200.0$$

ความแปรปรวนของ $\overline{X}$ หาได้จาก

$$\mathrm{Var}(\overline{X}) = \mathrm{E}(\overline{X} - \mathrm{E}(\overline{X})^2)$$

$$= \sum_{i=1}^{9} (\overline{X}_i - 200.00)^2\, g(\overline{X})$$

$$=(125.0-2000)\left(\frac{1}{16}\right) + (162.5-2000)\left(\frac{2}{16}\right) + (175.0+ 2000)\left(\frac{2}{16}\right) + \ldots+(250.0)\left(\frac{1}{16}\right)$$

$$= 1093.75$$

จะเห็นได้ว่า ค่าเฉลี่ยของตัวอย่าง $\mathrm{E}(\overline{X})$ จะเท่ากับ 200 ซึ่งเป็นค่าเดียวกับค่าเฉลี่ยของประชากร $(\overline{X}_i)$ และความแปรปรวนของตัวอย่าง $\mathrm{Var}(\overline{X})$ มีค่าเท่ากับ 1093.75 ซึ่งเป็นค่าเดียวกับความแปรปรวนของประชากรหารด้วยขนาดตัวอย่าง นั่นเอง

ทฤษฎีบทที่ 5.2 ถ้า $X_1, X_2, X_3, \ldots, X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ ความแปรปรวน σ^2 แล้วค่าเฉลี่ยของตัวอย่างสุ่ม $(\overline{X})$ จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน $\dfrac{\sigma^2}{n}$

ตัวอย่างที่ 5.6 ให้ $X_1, X_2, X_3, \ldots, X_{144}$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 9 ความแปรปรวน 49 จงหา $P(\overline{X} < 10.0)$

วิธีทำ ค่าเฉลี่ยตัวอย่าง $(\overline{X})$ มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย หรือ $E(\overline{X}) = 9$ ความแปรปรวน หรือ $Var(\overline{X}) = \frac{49}{144}$

นั่นคือ $\overline{X} \sim N\left(9, \frac{49}{141}\right)$

$$P(\overline{X} < 10.0) = P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{10.0-9}{\frac{7}{\sqrt{144}}}\right)$$

$$= P(Z < 1.71)$$

$$= 0.9564 \quad \text{(จากตารางในภาคผนวก)}$$

ตัวอย่างที่ 5.7 บริษัทผลิตน้ำดื่มบรรจุขวดแห่งหนึ่งบรรจุน้ำในขวดที่มีฉลากปิดบอกปริมาตรสุทธิ 600 มิลลิลิตร ถ้าปริมาตรสุทธิของน้ำดื่มมีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเป็น 600 มิลลิลิตรและความแปรปรวน 4.2 มิลลิลิตร^2 ในการบรรจุน้ำดื่มลงในแต่ละขวดจะต้องให้ปริมาตรสุทธิใกล้เคียงกับที่บอกไว้บนฉลากมากที่สุด แต่ปริมาตรจริงที่ยอมรับได้เป็น 599 ถึง 601 มิลลิลิตร ถ้าบริษัทชักตัวอย่างน้ำดื่มขนาด 16 ขวด จงหาความน่าจะเป็นที่ค่าเฉลี่ยของปริมาตรสุทธิอยู่ในเกณฑ์ที่ยอมรับได้

วิธีทำ ให้ $X_1, X_2, X_3, \ldots, X_{16}$ เป็นตัวอย่างสุ่มจากการแจกแจงแบบปกติ ด้วยค่าเฉลี่ย 600 มิลลิลิตร ความแปรปรวน 4.2 มิลลิลิตร^2 หรือเขียนเป็น $X \sim N(600, 4.2)$

$$P(599 < (\overline{X}) < 601) = P\left(\frac{599-600}{\sqrt{\frac{4.2}{16}}} < \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{601-600}{\sqrt{\frac{4.2}{16}}}\right)$$

$$= P(-1.95 < Z < 1.95)$$

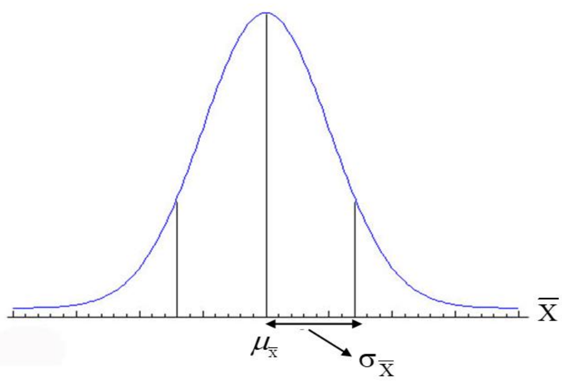
$$= 0.9488 \quad \text{(จากตารางในภาคผนวก)}$$

ทฤษฎีบทที่ 5.3 ให้ตัวอย่างสุ่ม $X_1, X_2, X_3, \ldots, X_n$ มาจากการแจกแจงหนึ่งที่มีค่าเฉลี่ยของประชากรเป็น μ และความแปรปรวนของประชากรเป็น σ^2 ให้ $\overline{X}$ เป็นค่าเฉลี่ยของตัวอย่าง ถ้า $n \to \infty$ แล้วตัวแปรสุ่ม $w = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ จะมีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ยเป็น 0 ความแปรปรวนเป็น 1

หากพิจารณาส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย คือค่า $\frac{\sigma}{\sqrt{n}}$ เมื่อ n คือขนาดตัวอย่าง ค่า $\frac{\sigma}{\sqrt{n}}$ จะลดลงเมื่อตัวอย่างมีขนาดใหญ่ (n มีค่ามาก) และถ้าประชากรมีการแจกแจงปกติ การแจกแจงของค่าเฉลี่ย $(\overline{X})$ จะมีการแจกแจงปกติเสมอไม่ว่าตัวอย่างที่สุ่มมาจะมีขนาดใหญ่หรือเล็กก็ตาม ~~(ชูศรี วงศ์รัตนะ, 2544)~~ ในทางปฏิบัติ หากไม่แน่ใจว่าประชากรมีการแจกแจงปกติ ควรเลือกตัวอย่างขนาดใหญ่มาทำการศึกษา การแจกแจงของค่าเฉลี่ยจึงจะมีการแจกแจงปกติ นอกจากนี้ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย $(\sigma_{\overline{X}})$ ที่หาได้จาก

$$(\sigma_{\overline{X}}) = \frac{\sigma}{\sqrt{n}}$$

เรียกว่า ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (standard error of mean) ซึ่งขึ้นอยู่กับส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) เมื่อส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) มีค่ามาก ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง $(\sigma_{\overline{X}})$ ก็จะมีค่ามากด้วย แต่เมื่อไรก็ตามที่ขนาดตัวอย่างโตขึ้นค่าของ $(\sigma_{\overline{X}})$ ก็จะลดลงทำให้ค่าเฉลี่ยของตัวอย่าง $(\overline{X})$ มีค่าใกล้เคียงค่าเฉลี่ยของประชากร (μ) กราฟแสดงค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย ดังภาพที่ 5.5



ภาพที่ 5.5 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

ตัวอย่างที่ 5.8 ถ้าให้ค่าเฉลี่ยจากตัวอย่างเป็น $(\overline{X})$ โดยที่ตัวอย่างสุ่ม $X_1, X_2, X_3, \ldots, X_{25}$ มาจากการแจกแจงซึ่งมีฟังก์ชันความหนาแน่นเป็น $f(x) = 2x^2, \ -1 < x < 1$ จงหาค่าประมาณของ $P(\overline{X} \le 0.25)$ โดยอาศัยทฤษฎีขีดจำกัดสู่ส่วนกลาง

วิธีทำ ค่าเฉลี่ยและความแปรปรวนของตัวอย่างเป็นดังนี้

$$E(X) = \int_{-1}^{1} x\,(2x^2)dx$$

$$= \int_{-1}^{1} 2x^3 dx$$

$$= 2\left(\frac{x^4}{4}\right)_{-1}^{1}$$

$$= 2\left(\frac{(1)^4}{4} - \frac{(-1)^4}{4}\right)_{-1}^{1}$$

$$= 2(0) = 0$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$= \int_{-1}^{1} x^2\,(2x^2)\,dx - (0)^2$$

$$= \int_{-1}^{1} 2x^4 dx - 0$$

$$= 2\left(\frac{x^5}{5}\right)_{-1}^{1}$$

$$= 2\left(\frac{(1)^5}{5} - \frac{(-1)^5}{5}\right)$$

$$= 2\left(\frac{2}{5}\right) = \frac{4}{5}$$

ดังนั้น ค่าเฉลี่ยของตัวอย่างเป็น 0

ความแปรปรวนของตัวอย่างเป็น $\dfrac{\left(\dfrac{4}{5}\right)}{25} = \dfrac{4}{125}$

$$P(\overline{X} \le 0.25) = P\left(\frac{\overline{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}} \le \frac{0.25-0}{\sqrt{\dfrac{4}{125}}}\right)$$

$$\cong P(Z \le 1.40)$$

$$= 0.9192$$ (จากตารางในภาคผนวก)

เนื่องจาก $\overline{X}$ ไม่ได้มีการแจกแจงแบบปกติ แต่ประมาณค่าความน่าจะเป็นโดยใช้การแจกแจงแบบปกติโดยที่ขนาดตัวอย่างเป็น 25 การประมาณที่ดีนั้นมิได้ขึ้นอยู่กับขนาดตัวอย่างแต่เพียงอย่างเดียว ความชัดเจนของการแจกแจงของตัวอย่างสุ่ม การพิจารณาว่าขนาดตัวอย่างควรเป็นเท่าไร ไม่มีกฎเกณฑ์แน่นอน แต่ประเด็นที่ควรพิจารณาประเด็นหนึ่งก็คือว่า การแจกแจงมีความใกล้เคียงหรือแตกต่างจากการแจกแจงแบบปกติเพียงใด ถ้ามีความใกล้เคียงกับการแจกแจงปกติและเป็นการแจกแจงแบบต่อเนื่อง ขนาดตัวอย่างเป็น 4 หรือ 5 ก็เพียงพอที่จะได้การประมาณค่าที่น่าพอใจ แต่ถ้าการแจกแจงของตัวอย่างสุ่มแตกต่างจากการแจกแจงแบบปกติมาก ก็ควรให้ขนาดตัวอย่างมีค่ามาก โดยทั่วไปขนาดตัวอย่างควรมากกว่า 25 จึงจะได้การประมาณค่าที่ถือว่าดี (ทวีรัตนา ศิวดุลย์, 2539; มนตรี สังข์ทอง, 2557)

5.11 การแจกแจงของผลต่างระหว่างค่าเฉลี่ยของตัวอย่าง

ถ้ากำหนดให้ประชากรกลุ่มที่หนึ่งมีค่าเฉลี่ย μ_1 ความแปรปรวน σ_1^2 ประชากรกลุ่มที่ 2 มีค่าเฉลี่ย μ_2 ความแปรปรวน σ_2^2 และกำหนดให้ตัวแปรสุ่ม $\overline{X}_1$ เป็นค่าเฉลี่ยของตัวอย่างขนาด n_1 ที่สุ่มมาจากประชากรที่ 1 ตัวแปรสุ่ม $\overline{X}_2$ เป็นค่าเฉลี่ยของตัวอย่างขนาด n_2 ที่สุ่มมาจากประชากรที่ 2 อย่างเป็นอิสระกัน ผลต่างระหว่าง $\overline{X}_1$ กับ $\overline{X}_2$ เท่ากับ $\overline{X}_1 - \overline{X}_2$ ก็จะเป็นตัวแปรสุ่มที่มีการแจกแจงตัวอย่างของค่าสถิติ $\overline{X}_1 - \overline{X}_2$

จากทฤษฎีขีดจำกัดสู่ส่วนกลาง จะได้ว่าถ้าตัวอย่างที่เป็นอิสระกันขนาด n_1 และ n_2 ที่สุ่มจากประชากร 2 กลุ่มที่มีค่าเฉลี่ย μ_1 และ μ_2 ส่วนเบี่ยงเบนมาตรฐาน σ_1 และ σ_2 แล้วการแจก

แจงของผลต่างระหว่างค่าเฉลี่ยจะมีการแจกแจงปกติด้วยค่าเฉลี่ย $\mu_{\bar{x}_1-\bar{x}_2}$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\bar{x}_1-\bar{x}_2}$ เมื่อ $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1-\mu_2$ และ $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ เป็นความคลาดเคลื่อนมาตรฐานของผลต่างระหว่างค่าเฉลี่ย (standard error of the difference of means)

5.12 การแจกแจงของสัดส่วนของตัวอย่าง

การแจกแจงของสัดส่วนของตัวอย่าง (sampling distribution of p) ประชากรขนาด N สามาตรแบ่งเป็น 2 ลักษณะ คือลักษณะที่สนใจกับลักษณะที่ไม่สนใจ ถ้าลักษณะที่สนใจในประชากรมีจำนวนหน่วยตัวอย่างเป็น X ลักษณะที่ไม่สนใจมีจำนวนหน่วยตัวอย่างเป็น N-X

ให้ P แทนสัดส่วนของประชากร (population proportion)

P = X/N = จำนวนหน่วยตัวอย่างที่สนใจในประชากร/ขนาดประชากร

ถ้าชักตัวอย่างขนาด n >30 จากประชากรที่มีสัดส่วนของลักษณะที่สนใจเป็น P จะได้ว่าสัดส่วนของตัวอย่าง (sample proportion) แทนด้วย p

p = x/n = จำนวนหน่วยตัวอย่างที่สนใจในตัวอย่าง/ขนาดตัวอย่าง เมื่อ x แทนจำนวนหน่วยตัวอย่างที่สนใจในตัวอย่างสุ่มขนาด n

ดังนั้น p จะมีการแจกแจงแบบปกติโดยประมาณ มีค่าเฉลี่ยเป็น p และความแปรปรวนเป็น pq/n เมื่อ q=1-p

จากทฤษฎีลิมิตเข้าสู่ส่วนกลาง ถ้าสุ่มตัวอย่างขนาดโต (n>30) จากประชากรที่มีการแจกแจงแบบปกติ จะได้ว่าตัวสถิติ p จะมีการแจกแจงแบบปกติด้วยค่าเฉลี่ย p และความแปรปรวน pq/n ทำให้สามารถแปลง p เป็น z ด้วยสูตร

$$z = \frac{p-P}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

ตัวอย่างที่ 5.9 สัดส่วนของครัวเรือนในตำบลเคร็ง อำเภอชะอวด จังหวัดนครศรีธรรมราช ที่พบผู้ป่วยโรคไข้เลือดออกเป็น 0.16 ถ้าสุ่มครัวเรือนมา 159 ครัวเรือน จงหาความน่าจะเป็นที่สัดส่วนของครัวเรือนจะพบผู้ป่วยโรคไข้เลือดออกมีค่าอยู่ระหว่าง 0.15 ถึง 0.20

วิธีทำ ให้ P แทนสัดส่วนของครัวเรือนที่พบผู้ป่วยโรคไข้เลือดออกของประชากร = 0.16

ให้ p แทนสัดส่วนของครัวเรือนที่พบผู้ป่วยโรคไข้เลือดออกของตัวอย่างที่สุ่มมา 159 ครัวเรือน (n>30) ดังนั้น

$$p \sim N\left(P, \frac{PQ}{n}\right)$$

เมื่อ p=0.16 q=1-p = 1-0.16 = 0.64

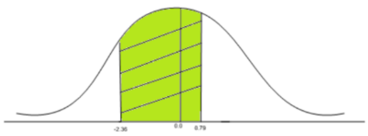
ความน่าจะเป็นที่สัดส่วนของครัวเรือนจะพบผู้ป่วยโรคไข้เลือดออกมีค่าอยู่ระหว่าง 0.15 ถึง 0.20 คือ

$$P(0.15 < p < 0.20) = P\left(\frac{0.15 - 0.16}{\sqrt{\frac{(0.16)(0.64)}{159}}} < \frac{p - P}{\sqrt{\frac{pq}{n}}} < \frac{0.20 - 0.16}{\sqrt{\frac{(0.16)(0.64)}{159}}}\right)$$

$$= P(-2.36 < Z < 0.79)$$

$$= \ 0.7761$$ (จากตารางในภาคผนวก)

เป็นพื้นที่ในภาพ



ตัวอย่างที่ 5.10 ร้านสะดวกซื้อประมาณการว่าลูกค้าที่เข้ามาในร้านจะซื้อสินค้าประมาณ ร้อยละ 40 สุ่มลูกค้ามา 100 คน จงหาความน่าจะเป็นที่ลูกค้าซื้อสินค้าไม่น้อยกว่าร้อยละ 30

วิธีทำ ให้ P แทนสัดส่วนของประชากรลูกค้าที่ซื้อสินค้า = 0.40 จะได้

ให้ p แทนสัดส่วนของตัวอย่างลูกค้าที่ซื้อสินค้าจากที่สุ่มมา 100 คน

P = 0.40 Q = 1-P = 0.60 n = 100

$$P(p \geq 0.30) = P\left(\frac{p - P}{\sqrt{\frac{PQ}{n}}} \geq \frac{0.30 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{100}}}\right)$$

$$\begin{aligned} P(p \geq 0.30) &= P(Z \geq -2.04) \\ &= 1 - \ P(p < 2.04) \\ &= 1 - \ 0.9793 \\ &= 0.0207 \end{aligned}$$

5.13 การแจกแจงของผลต่างของสัดส่วนของตัวอย่าง

การแจกแจงของผลต่างของสัดส่วนของตัวอย่าง (sampling distribution of p_1-p_2) ถ้าชักตัวอย่างขนาด n_1 และ n_2 ที่เป็นอิสระกัน จากประชากรที่ 1 และประชากรที่ 2 ที่มีความน่าจะเป็นของสิ่งที่สนใจเป็น $P_1 = X_1/N_1$ และ $P_2 = X_2/N_2$ ตามลำดับ

ถ้า n_1 และ n_2 มีขนาดโตพอ ค่าสัดส่วนของตัวอย่างสุ่มที่ 1 และ 2 คือ $p_1 = x_1/n_1$ และ $p_2 = x_2/n_2$ ตามลำดับแล้ว ผลต่างของสัดส่วนตัวอย่าง p_1-p_2 จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย P_1-P_2 ความแปรปรวน $P_1Q_1/n_1 + P_2Q_2/n_2$

สัดส่วนของประชากรที่ 1 เป็น $P_1 = X_1/N_1$ สัดส่วนของประชากรที่ 2 เป็น $P_2 = X_2/N_2$ ชักตัวอย่างขนาด $n_1 > 30$ จากประชากรที่ 1 แล้วหาสัดส่วนของตัวอย่าง $p_1 = x_1/n_1$ และชักตัวอย่างขนาด $n_2 > 30$ จากประชากรที่ 2 แล้วหาสัดส่วนของตัวอย่าง $p_2 = x_2/n_2$ ผลต่างของสัดส่วนของตัวอย่าง $p_1\text{-}p_2 \sim N(P_1\text{-}P_2\,,\ P_1Q_1/n_1 + P_2Q_2/n_2)$ สามารถแปลงตัวสถิติ p_1-p_2 เป็น Z ดังนี้

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N(0,1)$$

ตัวอย่างที่ 5.11 ฝรั่ง 100 กรัม เทียบเท่าวิตามินซี 228 กรัม (.......) ถ้าผู้ที่รับประทานฝรั่ง 100 กรัม เป็นประจำ มีความน่าจะเป็นที่จะเป็นหวัด 0.10 ส่วนกลุ่มที่ไม่รับประทานฝรั่งวันละ 100 กรัม เป็นประจำมีโอกาสเป็นหวัด 0.20 ถ้าเลือกตัวอย่างสุ่มมากลุ่มละ 40 คน จงหาความน่าจะเป็นของความแตกต่างของสัดส่วนตัวอย่างของผู้ที่เป็นหวัดทั้งสองกลุ่มว่าจะแตกต่างกันไม่เกิน 0.05

วิธีทำ ให้ P_1 เป็นสัดส่วนของผู้ที่เป็นหวัดในกลุ่มผู้ที่รับประทานฝรั่ง 100 กรัม เป็นประจำ

P_2 เป็นสัดส่วนของผู้ที่เป็นหวัดในกลุ่มผู้ที่ไม่รับประทานฝรั่ง 100 กรัม เป็นประจำ

ให้ p_1 เป็นสัดส่วนของผู้ที่เป็นหวัดในตัวอย่างสุ่ม n_1 จากกลุ่มผู้ที่รับประทานฝรั่งเป็นประจำ

p_2 เป็นสัดส่วนของผู้ที่เป็นหวัดในตัวอย่างสุ่ม n_2 จากกลุ่มผู้ที่ไม่รับประทานฝรั่งเป็นประจำ

เนื่องจากตัวอย่างสุ่มขนาด $n_1 = n_2 = 40$ มีขนาดโต จากทฤษฎีลิมิตเข้าสู่ส่วนกลาง (central limit theorem: CLT) จะได้

$$(p_1 - p_2) \sim N\left(P_1 - P_2, \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}\right)$$

เมื่อ $P_1 = 0.10$, $P_2 = 0.20$ ความน่าจะเป็นของความแตกต่างของสัดส่วนตัวอย่างของผู้ที่เป็นหวัดทั้งสองกลุ่มว่าจะแตกต่างกันไม่เกิน 0.05 คือ

$$P(p_1 - p_2 \le 0.05) = P\left(\frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} < \frac{0.05 - (0.10 - 0.20)}{\sqrt{\frac{(0.10)(0.90)}{50} + \frac{(0.20)(0.80)}{50}}}\right)$$

$$= P\left(Z < \frac{0.05 - (0.10 - 0.20)}{\sqrt{\frac{(0.10)(0.90)}{50} + \frac{(0.20)(0.80)}{50}}}\right)$$

$$= P\left(Z < \frac{0.15}{\sqrt{0.005}}\right)$$

$$= P(Z < 2.12)$$

$$= 0.9830$$

5.14 การแจกแจงของความแปรปรวนของตัวอย่าง

การแจกแจงของความแปรปรวนของตัวอย่าง (sampling distribution of S^2) ซึ่งเป็นพารามิเตอร์ที่สำคัญอีกตัวหนึ่งที่ใช้ประมาณค่าความแปรปรวนของประชากร ความแปรปรวนของตัวอย่างเป็นฟังก์ชันของตัวอย่างสุ่มขนาด n ที่ประกอบด้วยตัวแปรสุ่ม (X_1, X_2, X_3, ..., X_n) ดังนี้

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

เมื่อ n-1 คือ ระดับขั้นความเสรี (degree of freedom) หมายถึงขนาดตัวอย่างที่ถูกหักด้วยจำนวนฟังก์ชันที่สร้างขึ้นจากตัวอย่างสุ่มเพื่อใช้ในการหาค่าตัวสถิติตัวใหม่ เช่นในที่นี้ต้องการหาค่า S^2 แต่ต้องหา $\overline{X}$ ก่อน และ $\overline{X}$ เป็นฟังก์ชันของตัวอย่างสุ่มหนึ่งฟังก์ชัน ระดับขั้นความเสรีของ S^2 จึงเท่ากับ n-1 กล่าวคือ จากตัวอย่างขนาด n ถ้ากำหนดค่า $\overline{X}$ ให้ จะสามารถเลือกค่า X_i ต่าง ๆ ได้อย่างเป็นอิสระเพียง n-1 ตัว เท่านั้น ส่วนตัวสุดท้ายจะไม่มีโอกาสเลือก เนื่องจากต้องเป็นค่าซึ่งเมื่อรวมกับค่าของ X_i ตัวอื่น ๆ อีก n-1 ตัว แล้วหารด้วย n จะต้องได้เท่ากับ $\overline{X}$ ดังนั้น เมื่อกำหนด $\overline{X}$ ให้จึงมี X_i เพียง n-1 ตัว ใน $\sum_{i=1}^{n}(X_i - \overline{X})^2$ ที่สามารถเปลี่ยนค่าได้

บทนิยามที่ 5.9 ระดับขั้นความเสรี คือเลขจำนวนเต็มบวกที่บอกว่ามีค่าของตัวแปรสุ่มที่นำมาสร้างตัวสถิติกี่ค่าที่สามารถเปลี่ยนแปลงหรือกำหนดได้อย่างอิสระเมื่อกำหนดค่าของฟังก์ชันต่าง ๆ ที่ใช้ในการสร้างตัวสถิตินี้แล้ว ระดับขั้นความเสรีจึงมีค่าเท่ากับขนาดตัวอย่างลบด้วย จำนวนฟังก์ชันที่สร้างขึ้นเพื่อประมาณค่าพารามิเตอร์ที่ไม่ทราบค่า

ทฤษฎีบทที่ 5.4 ให้ ($X_1, X_2, X_3, \ldots, X_n$) เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงแบบปกติมีความแปรปรวน σ^2 แล้ว ความแปรปรวนของตัวอย่าง คือ

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

โดยที่

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

มีการแจกแจงไคกำลังสอง (chi-square distribution) ด้วยระดับขั้นความเสรี n-1 หรือเขียนได้ว่า $\chi^2 \sim \chi^2_{n-1}$

ตัวอย่างที่ 5.12 บริษัทผลิตทุเรียนทอดบรรจุกระป๋องแห่งหนึ่ง พบว่าน้ำหนักสุทธิของทุเรียนมีการแจกแจงปกติ มีความแปรปรวน 7 กรัม^2 ถ้าสุ่มทุเรียนทอดมา 15 กระป๋อง จงหาความน่าจะเป็นที่ความแปรปรวนของทุเรียนมากกว่า 9 กรัม^2

วิธีทำ ให้ X เป็นน้ำหนักสุทธิของทุเรียน มีการแจกแจงแบบปกติ ด้วยความแปรปรวนเท่ากับ 7 กรัม^2

S^2 เป็นความแปรปรวนของน้ำหนักทุเรียนที่สุ่มมา 15 กระป๋อง

ความน่าจะเป็นที่ความแปรปรวนของทุเรียนมากกว่า 9 กรัม^2 คือ

$$\begin{aligned} P\left(S^2 > 9\right) &= P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{9(n-1)}{\sigma^2}\right) \\ &= P\left(\chi^2_{15-1} > \frac{9(15-1)}{7}\right) \\ &= P\left(\chi^2_{14} > \frac{126}{7}\right) \\ &= P\left(\chi^2_{14} > 18\right) \end{aligned}$$

จากตารางไคกำลังสองในภาคผนวก

$$P\left(\chi^2_{14} > 21.06\right) < P\left(\chi^2_{14} > 18\right) < P\left(\chi^2_{14} > 6.57\right)$$

จะได้ (โดยประมาณ)

$$0.10 < P\left(\chi^2_{14} > 18\right) < 0.95$$

นั่นคือ

$$0.10 < P\left(S^2 > 9\right) < 0.95$$

5.15 การแจกแจงของอัตราส่วนความแปรปรวนของตัวอย่าง

การแจกแจงของอัตราส่วนความแปรปรวนของตัวอย่าง (sampling distribution of S^2_1/S^2_2)

ทฤษฎีบทที่ 5.5 ถ้า S^2_1 และ S^2_2 เป็นความแปรปรวนของตัวอย่างขนาด n_1 และ n_2 ซึ่งสุ่มมาโดยอิสระกันจาก 2 ประชากรที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ย μ_1 และ μ_2 และมีความแปรปรวน σ^2_1 และ σ^2_2 ตามลำดับแล้ว จะได้

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

มีการแจกแจงไคกำลังสอง (chi-square distribution) ด้วยระดับขั้นความเสรี n-1

$$F = \frac{\frac{\chi_1^2}{(n_1-1)}}{\frac{\chi_2^2}{(n_2-1)}} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}$$

จะมีการแจกแจงแบบเอฟ (F distribution) ด้วยระดับขั้นความเสรี n_1-1 และ n_2-1

ตัวอย่างที่ 5.13 ประชากร 1 และ 2 มีการแจกแจงแบบปกติ ด้วยความแปรปรวน $\sigma^2_1 = 144$ และ $\sigma^2_2 = 81$ สุ่มตัวอย่างขนาด $n_1 = 12$ และขนาด $n_2 = 9$

ตัวอย่างที่ 5.14 สุ่มตัวอย่างขนาด $n_1 = 10$ และขนาด $n_2 = 13$ อย่างเป็นอิสระกัน จากประชากรที่มีการแจกแจงแบบปกติที่มีความแปรปรวน $\sigma^2_1 = 15$ และ $\sigma^2_2 = 8$ จงหาความน่าจะเป็นที่ค่าอัตราส่วนความแปรปรวนของตัวอย่างจะมีค่ามากกว่า .....

บทสรุป

การชักตัวอย่างเป็นวิธีการเลือกตัวอย่างมาเป็นตัวแทนของประชากรโดยให้ทุกหน่วยตัวอย่างมีโอกาสถูกเลือกมาเป็นตัวแทนเท่า ๆ กัน เพื่ออ้างอิงไปยังประชากร นอกจากนี้ขนาดตัวอย่างก็ต้องโตพอจึงจะทำให้ได้ผลการวิจัยที่ถูกต้องตรงตามความเป็นจริง เทคนิคการชักตัวอย่างที่นิยมใช้ได้แก่ การชักตัวอย่างแบบใช้ความน่าจะเป็นหรือการชักตัวอย่างที่เป็นไปตามโอกาสทางสถิติ ทำการเลือกหน่วยตัวอย่างโดยวิธีการจับฉลากหรือใช้ตารางเลขสุ่ม การชักตัวอย่างแบบมีระบบ การชักตัวอย่างแบบแบ่ง

ชั้นภูมิและการชักตัวอย่างแบบแบ่งกลุ่ม ในการเลือกใช้เทคนิคและวิธีการในการชักตัวอย่างแบบใดนั้นยังจะต้องคำนึงถึง ลักษณะโดยธรรมชาติของประชากรด้วยและหากเมื่อใดก็ตามที่ประชากรมีจำนวนน้อยก็สามารถใช้ประชากรในการศึกษาโดยไม่ต้องใช้การชักตัวอย่าง ตัวสถิติที่ได้จากตัวอย่างสุ่มจะถูกนำไปสู่การอ้างอิงไปยังประชากรด้วยการประมาณค่าหรือการทดสอบสมมติฐานการแจกแจงของตัวสถิติจากตัวอย่างที่มีความแตกต่างกันหลายรูปแบบ เช่น การแจกแจงที การแจกแจงปกติ การแจกแจงไคกำลังสอง และการแจกแจงแบบเอฟ มีลักษณะและรูปแบบเฉพาะของข้อมูลแต่ละชุด สิ่งเหล่านี้จะช่วยให้ผู้วิเคราะห์สามารถเลือกสถิติในการวิเคราะห์ข้อมูลได้อย่างเหมาะสมยิ่งขึ้น

แบบฝึกหัด

1. จงให้ความหมายของคำต่อไปนี้

    1.1 หน่วยตัวอย่าง

    1.2 ตัวอย่าง

    1.3 ประชากร

    1.4 ตัวอย่างสุ่ม

    1.5 ตัวแปรสุ่ม

2. จงอธิบายการใช้ตารางเลขสุ่ม เพื่อเลือกตัวอย่างในกรณีที่ต้องการเลือกผู้ตอบแบบสอบถาม จำนวน 100 คน จากนักศึกษาจำนวน 500 คน
3. การชักตัวอย่างแบบชั้นภูมิ (stratified sampling) กับการชักตัวอย่างแบบกลุ่ม (cluster sampling) มีความคล้ายคลึงกันหรือแตกต่างกันอย่างไรบ้าง
4. จงยกตัวอย่างการชักตัวอย่างแบบชั้นภูมิจนได้ตัวอย่างมาโดยละเอียด
5. ในอำเภอหนึ่งมีครัวเรือนจำนวน 200 ครัวเรือน ทำการเลือกตัวอย่างด้วยวิธีการชักตัวอย่างอย่างง่ายแบบไม่แทนที่มาจำนวน 10 ครัวเรือน ครัวเรือนที่สุ่มได้มีจำนวนสมาชิกในครัวเรือน 4, 8, 10, 3, 2, 5, 6, 9, 8, 7 จงหา

    5.1 จำนวนสมาชิกเฉลี่ยของครัวเรือนในอำเภอนี้

    5.2 ความแปรปรวนของจำนวนสมาชิกเฉลี่ย

    5.3 จงคำนวณหาขนาดตัวอย่างของครัวเรือนที่จะต้องสุ่มเพื่อให้ค่าประมาณจำนวนสมาชิกในครัวเรือนเฉลี่ยคลาดเคลื่อน 2 คน ที่ระดับความเชื่อมั่น 90%

6. จงยกตัวอย่างวิธีการชักตัวอย่างแบบไม่ใช้ความน่าจะเป็นมา 3 วิธี อธิบายและยกตัวอย่างประกอบการอธิบาย

7. จงคำนวณหายอดรวมค่าโดยสารรถของนักศึกษาในการเดินทางมายังมหาวิทยาลัยแห่งหนึ่งในแต่ละวัน โดยทำการเลือกนักศึกษาทุก ๆ 50 คน จากรายชื่อนักศึกษา 1,000 คน ได้ข้อมูลค่าโดยสารดังนี้

นักศึกษาคนที่	ค่าโดยสาร (บาท)
1	25
2	30
3	10
4	6
5	25
6	15
7	45
8	30
9	12
10	6
11	20
12	25
13	40
14	12
15	30
16	8
17	25
18	12
19	6
20	30

8. ตัวแปรสุ่ม X มีการแจกแจงแบบ F ที่ระดับขั้นความเสรี เท่ากับ r_1 และ r_2 จงหา

8.1 $P(X \le 3.97)$ ที่ระดับขั้นความเสรี (5,7)

8.2 $P(X \le 14.54)$ ที่ระดับขั้นความเสรี (8,3)

8.3 $P(3.86 \le X \le 5.08)$ ที่ระดับขั้นความเสรี (3,9)

9. จงหาค่าของ

9.1 $F_{0.05, 10,7}$

9.2 $F_{0.005, 12,4}$

9.3 $F_{0.95, 3,10}$

10. ถ้าค่าจ้างแรงงานต่อวันของพนักงานในโรงงานแห่งหนึ่งมีการแจกแจงแบบปกติ มีค่าเฉลี่ย 300 บาท และความแปรปรวน 60 บาท^2 จงหาความน่าจะเป็นที่พนักงานจะได้รับค่าจ้างแรงงานต่อวัน

10.1 มากกว่า 250 บาท

10.2 ตั้งแต่ 190 ถึง 240 บาท

10.3 น้อยกว่า 300 บาท

เอกสารอ้างอิง

Kerlinger, 1973

W. G. Cochran, 1953

ประชุม สุวัตถี, 2527;

(กัลยา วานิชย์บัญชา, 2540; ยุทธ์ ไกยวรรณ์, 2546; มนตรี สังข์ทอง, 2557)

(บุญธรรม กิจปรีดาบริสุทธิ์, 2535)