บทที่ 3

วิธีดำเนินการศึกษา

3.1 ความนำ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อทำการแสดงให้เห็นถึงรากฐานแนวคิดของการประมาณสัดส่วนประชากรและความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี โดยแบ่งการศึกษาออกเป็นสองส่วน คือ การวิเคราะห์แนวคิดพื้นฐานจากกรณีศึกษาและการจำลองข้อมูล ซึ่งในการจำลองข้อมูลจะทำการสร้างข้อมูลให้มีลักษณะตามสถานการณ์ที่กำหนดในขอบเขตการศึกษา ด้วยโปรแกรมสำเร็จรูป SPSS จากนั้นนำข้อมูลมาทำการศึกษาตามขั้นตอนที่กำหนดไว้

ทั้งนี้เพื่อความสะดวกในการศึกษาจึงกำหนดสัญลักษณ์ต่าง ๆ ดังนี้

P หมายถึง สัดส่วนประชากร

P_h หมายถึง สัดส่วนที่มีลักษณะที่ต้องการในชั้นภูมิ h

p หมายถึง สัดส่วนตัวอย่างภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน

π หมายถึง ความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี

N หมายถึง ขนาดประชากร

N_h หมายถึง ขนาดชั้นภูมิ h

n หมายถึง ขนาดตัวอย่าง

L หมายถึง จำนวนชั้นภูมิ

3.2 วิเคราะห์แนวคิดพื้นฐานจากกรณีศึกษา

1. อธิบายแนวคิดพื้นฐานสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี

2. อธิบายวิธีการประมาณค่าแบบจุดระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และพารามิเตอร์ π ของประชากรที่มีการแจกแจงแบบเบอร์นูลี

3.3 การจำลองข้อมูล

จากที่ได้กล่าวมาแล้วในบทที่ 2 ว่าภายใต้การสุ่มตัวอย่างสุ่มแบบง่ายที่ไม่ใส่คืน ตัวประมาณ $p = a/n$ เป็นตัวประมาณที่ใช้สำหรับการประมาณค่าสัดส่วนประชากร ในส่วนนี้จึงได้ทำการศึกษาโดยการจำลองข้อมูลออกเป็น 3 ส่วนคือ การทดสอบคุณสมบัติการแจกแจงแบบปกติ

ของจำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a) การตรวจสอบค่าความน่าจะเป็นของการครอบคลุมของช่วงความเชื่อมั่นที่ได้จากตัวประมาณแบบช่วง และการศึกษาเปรียบเทียบคุณสมบัติการเป็นตัวประมาณที่ดีภายใต้แผนการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย

3.3.1 การทดสอบคุณสมบัติการแจกแจงแบบปกติของจำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a)

ศึกษารูปแบบการแจกแจงของตัวแปรสุ่ม a และ Asymptotic Distribution ว่าเข้าสู่การแจกแจงแบบปกติเมื่อใด โดยขั้นแรกทำการจำลองข้อมูลตัวแปรสุ่ม a ภายใต้การแจกแจงแบบไฮเปอร์จิออเมตริก ด้วยโปรแกรม SPSS ให้มีขอบเขตของการศึกษาตามที่กำหนดต่อไปนี้

- ขนาดประชากร (N) ที่ใช้ศึกษา กำหนดให้มีขนาดเท่ากับ 50 100 500 1,000 2,000 2,500 5,000 และ 10,000
- กำหนดค่า $P = \frac{A}{N}$ มีค่าตั้งแต่ 0.05 ถึง 0.95 โดยทำการเพิ่มค่าขึ้นทีละ 0.05
- กำหนดขนาดตัวอย่างในรูปร้อยละของประชากร 4 ระดับ คือ 5 10 15 และ 20

โดยจำนวนรอบของการทำซ้ำซึ่งถือว่าเป็นการกำหนดขนาดตัวอย่างสำหรับการประมาณค่า P มีหลักการดังนี้

$$P(|p-P| \geq d) = \alpha$$

$$n^{*} = \frac{z^{2}_{\alpha/2}(1-P)P}{d^{2}}$$

โดยที่ d เป็นความผิดพลาดที่ยอมให้เกิดขึ้น การศึกษานี้กำหนดให้ $P = 0.50$ $d = 0.05$ และ $\alpha = 0.01$ ดังนั้นจำนวนรอบของการทำซ้ำคือ 663 รอบ

เมื่อสร้างข้อมูลได้ตามที่กำหนดแล้ว ในขั้นตอนแรกทำการตรวจสอบข้อมูลเบื้องต้นจาก 3 วิธี คือ Stem-and Leaf Normal Q-Q Plot และ Descriptive Statistics จากนั้นทดสอบการแจกแจงของ a ที่สร้างขึ้นว่ามีการแจกแจงแบบปกติหรือไม่ โดยใช้สถิติทดสอบ χ^2 ซึ่งมีรายละเอียดดังนี้

สมมติฐานการทดสอบ คือ

H_0 : a มีรูปแบบการแจกแจงแบบปกติ

H_1 : a ไม่ได้มีรูปแบบการแจกแจงแบบปกติ

สถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

โดยที่ O_i คือ ความถี่ที่เกิดในระดับ/ลักษณะที่ i ; i=1,2,...,k

E_i คือ ความถี่ที่คาดว่าจะเกิดในระดับ/ลักษณะที่ i ภายใต้สมมติฐาน H_0

หรือ $E_i = n \times$ ความน่าจะเป็นที่เกิดในระดับ/ลักษณะที่ i ของการแจกแจงที่คาดไว้

ขอบเขตวิกฤต

จะปฏิเสธ H_0 เมื่อ $\chi^2 > \chi^2_{1-\alpha;k-1-m}$

โดยที่ k = จำนวนระดับ/ลักษณะ

m = จำนวนพารามิเตอร์ที่ต้องการประมาณค่า

หมายเหตุ : ในกรณีนี้ m = 2 เนื่องจากต้องประมาณค่า μ และ σ^2

3.3.2 การตรวจสอบค่าความน่าจะเป็นของการครอบคลุมของช่วงความเชื่อมั่นที่ได้จากตัวประมาณแบบช่วง

เป็นการศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืนจากการประมาณ 2 วิธีคือ การประมาณด้วยการแจกแจงแบบปกติและการประมาณด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ ของ M.E. Thompson (1997) โดยทำการจำลองข้อมูลตัวแปรสุ่ม a ให้มีขอบเขตการศึกษาเหมือนกับหัวข้อ 3.3.1 จากนั้นคำนวณสัดส่วนตัวอย่าง ($p = a / n$) แล้วคำนวณช่วงความเชื่อมั่นจากวิธีการประมาณทั้งสองวิธี จากนั้นเปรียบเทียบค่าความน่าจะเป็นของการครอบคลุมที่ได้กับระดับความเชื่อมั่นที่กำหนด การศึกษานี้กำหนดระดับความเชื่อมั่น 3 ระดับคือ 0.90 0.95 และ 0.99 ซึ่งจะได้จำนวนรอบของการทำซ้ำเท่ากับ 663 350 และ 73 ตามลำดับ

ในการกำหนดจำนวนรอบของการจำลองข้อมูลถือว่าเป็นการกำหนดขนาดตัวอย่างสำหรับการประมาณค่าพารามิเตอร์ π^* ซึ่งในการศึกษานี้คือระดับความเชื่อมั่น หมายถึงความน่าจะเป็นที่ค่า P ตกอยู่ในช่วงที่สร้างขึ้น ทั้งนี้ในการคำนวณจำนวนรอบกำหนดค่าพารามิเตอร์ π^* 3 ระดับ คือ 0.90 0.95 และ 0.99 ซึ่งมีสูตรการคำนวณดังนี้

$$n^{**} = \frac{z^2_{\alpha/2}(1-\pi^*)\pi^*}{d^{*2}}$$

โดยที่ $d^* = 0.03$ และ $\alpha = 0.01$

3.3.3 การศึกษาเปรียบเทียบคุณสมบัติการเป็นตัวประมาณที่ดีภายใต้แผนการสุ่มตัวอย่าง แบบมีชั้นภูมิอย่างง่าย

เพื่อแสดงให้เห็นการใช้ตัวประมาณที่ถูกต้องตามแผนการสุ่มตัวอย่าง ซึ่งจากการศึกษาภายใต้การสุ่มตัวอย่างแบบง่ายแล้ว ได้เพิ่มเติมการศึกษาภายใต้การสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย เพื่อเป็นการพัฒนาความเข้าใจในขั้นที่สูงขึ้น การศึกษานี้ทำการจำลองข้อมูลเป็นลักษณะการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย แล้วประมาณสัดส่วนประชากรโดยใช้ตัวประมาณ 2 ประเภท คือ ตัวประมาณแบบการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling ; SRS) และตัวประมาณแบบการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย (Stratified Random Sampling)

จากนั้นทำการเปรียบเทียบตัวประมาณสัดส่วนประชากรทั้งสอง โดยใช้หลักเกณฑ์การพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error) รายละเอียดเป็นดังนี้

สำหรับการกำหนดพารามิเตอร์ ในการศึกษานี้ แบ่งออกเป็น 2 กรณี คือ

จำนวนชั้นภูมิ (L)	L=2	L=5
กรณีที่ 1	$N_1 = N_2$ $P_1 = \frac{1}{2}P_2$	$N_1 = N_2 = N_3 = N_4 = N_5$ $P_1 = \frac{1}{2}P_2 = P_3 = P_4 = P_5$
กรณีที่ 2	$N_1 = \frac{1}{2}N_2$ $P_1 = \frac{1}{2}P_2$	$N_1 = \frac{1}{2}N_2 = N_3 = N_4 = N_5$ $P_1 = \frac{1}{2}P_2 = P_3 = P_4 = P_5$

การคำนวณสัดส่วนประชากรจริงมาจากสูตร $P = \frac{1}{N}\sum N_h P_h$

สร้างตัวแปรสุ่ม a ที่มีการแจกแจงแบบไฮเปอร์จิออเมตริก ด้วยโปรแกรม SPSS โดยให้มีสถานการณ์ดังนี้

- กำหนดให้จำนวนชั้นภูมิ (L) เป็น 2 ระดับ คือ 2 และ 5

- ขนาดประชากร (N)

  กรณีที่ 1 เท่ากับ 1,000 และ 5,000

  กรณีที่ 2 เท่ากับ 1,500 และ 7,500

- กำหนดค่า $P_1 = A_1 / N_1$ มีค่าตั้งแต่ 0.05 ถึง 0.5 โดยทำการเพิ่มค่าทีละ 0.05
- กำหนดขนาดตัวอย่างในรูปร้อยละของประชากร 4 ระดับ คือ 5 10 15 และ 20
- ขนาดตัวอย่างของชั้นภูมิแบ่งเป็น 2 ลักษณะ คือ เป็นสัดส่วนโดยตรงกับขนาดของชั้นภูมิ และขนาดเท่ากันทุกชั้นภูมิ
- จำนวนรอบของการทำซ้ำเป็น 663 รอบ โดยกำหนดเงื่อนไขและวิธีการคำนวณเหมือนกับหัวข้อ 3.3.1

จากนั้นประมาณสัดส่วนประชากรโดยใช้การประมาณ 2 วิธีคือ

ตัวประมาณแบบ SRS

$$p = \frac{1}{n} \sum_{h=1}^{L} \sum_{i=1}^{n_h} y_{hi}$$

กำหนดให้ $y_{hi} = 1$ ถ้าหน่วย i มีลักษณะที่ต้องการ

0 ในกรณี อื่น ๆ

ตัวประมาณแบบ Stratified Random Sampling

$$p_{st} = \sum_{h=1}^{L} W_h p_h$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เปรียบเทียบการประมาณทั้งสองโดยใช้เกณฑ์การพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error) ซึ่งเท่ากับ $\frac{\sum_{i=1}^{n} |p - P|}{n^*}$