บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในการศึกษาประชากรอันตะ (Finite Population) ซึ่งเป็นประชากรที่สามารถนับจำนวนหน่วยประชากร (Population Unit) ทั้งหมดได้จริงนั้น โดยทั่วไปลักษณะสำคัญของประชากร (Population Characteristic) ที่ใช้นำมาอธิบายสภาพความเป็นไปเกี่ยวกับประชากรภายใต้กรอบความสนใจมี 4 ประเภท คือ ยอดรวมจำนวนประชากร ค่าเฉลี่ยประชากร สัดส่วนประชากร และอัตราส่วนประชากร สำหรับสัดส่วนประชากร (P) คือค่าที่สามารถคำนวณได้โดยทำการนับจำนวนหน่วยประชากรที่มีคุณลักษณะที่สนใจ (A) แล้วนำมาเปรียบเทียบกับจำนวนหน่วยประชากรทั้งหมด (N) ภายใต้การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling ; SRS) ที่ไม่ใส่คืน ตัวประมาณที่ไม่เอนเอียงสำหรับ P คือ $p = a/n$ หรือสัดส่วนตัวอย่าง โดยที่ a คือจำนวนรวมของหน่วยตัวอย่างซึ่งมีคุณลักษณะในกลุ่มที่สนใจ

สำหรับในส่วนของทฤษฎีการอนุมานทางสถิติแบบคลาสสิค (Classical Statistical Inference) ความน่าจะเป็นของการเกิดผลสำเร็จสำหรับการทดลองสุ่มแบบเบอร์นูลี (Probability of Success for a Bernoulli Trial : π) ถือเป็นค่าพารามิเตอร์ที่สำคัญอย่างหนึ่งที่ได้มีการพัฒนาทฤษฎีทางสถิติที่เกี่ยวข้องมากมาย ตัวประมาณแบบจุดสำหรับพารามิเตอร์ดังกล่าว คือ $\hat{\pi} = \sum_{i=1}^{n} x_i / n$ โดยที่ $\sum_{i=1}^{n} x_i$ คือจำนวนรวมของหน่วยตัวอย่างซึ่งมีคุณลักษณะในกลุ่มที่สนใจ (Success) ซึ่งเป็นตัวประมาณที่เหมือนกับ $p = a/n$ จากการประมาณสัดส่วนประชากร ภายใต้ SRS

จะเห็นได้ว่า ตัวประมาณทั้งสองมีลักษณะที่เหมือนกัน คือ การรวมยอดของหน่วยตัวอย่างที่มีคุณลักษณะที่สนใจเทียบกับจำนวนหน่วยตัวอย่างทั้งหมด แต่ทั้งนี้ตัวอย่างจากกรณีของ SRS ไม่เป็นตัวอย่างสุ่ม (Independent Identically Distributed Random Sample) ดังกรณีของทฤษฎีการอนุมานทางสถิติ ในทฤษฎีสถิติฟังก์ชันความน่าจะเป็น (Probability Function) พื้นฐานของตัวสถิติ $\sum_{i=1}^{n} x_i$ เป็นการแจกแจงแบบทวินาม ในขณะที่ a เป็นการแจกแจงแบบไฮเปอร์จิออเมตริก

จากที่กล่าวมาแสดงให้เห็นได้อย่างชัดเจนว่า ตัวประมาณสำหรับ P ภายใต้การสุ่มตัวอย่างแบบ SRS ที่ไม่ใส่คืน และตัวประมาณ $\hat{\pi}$ สำหรับ π มีรูปแบบที่เหมือนกันแต่รากฐานการประมาณทั้งสองมีที่มาแตกต่างกันโดยสิ้นเชิง ทั้งนี้ในทางปฏิบัติพบว่าการสำรวจตัวอย่างเพื่อศึกษาสัดส่วนประชากรเมื่อพูดถึงการประมาณก็ยังไม่มีความชัดเจน เนื่องจากในตำราสถิติทั่วไปได้เขียนเกี่ยวกับ

สองเรื่องนี้ในลักษณะเป็นเรื่องเดียวกัน ทำให้การวิเคราะห์ข้อมูลขาดหลักการทางทฤษฎีที่ถูกต้องเข้ามารองรับ และสิ่งที่ตามมาอาจมีผลต่อความคลาดเคลื่อนหรือคุณภาพของตัวสถิติที่นำมาใช้ประมาณ

การศึกษาถึงรากฐานแนวคิดของการประมาณแบบจุดและการประมาณแบบช่วงสำหรับประมาณสัดส่วนประชากร (P) ภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี (π) ให้เกิดความกระจ่างชัดถึงการพัฒนากระบวนการทางสถิติที่เกี่ยวข้องในอันที่จะนำไปใช้ประโยชน์ได้อย่างถูกต้องตามทฤษฎีสถิติเป็นวัตถุประสงค์หลักของการศึกษาในครั้งนี้

1.2 วัตถุประสงค์ของการวิจัย

- เพื่ออธิบายแนวคิดและความแตกต่างระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี (Probability of Success in Bernoulli Distribution) อย่างละเอียดชัดเจน
- เพื่อศึกษาวิธีการและระบุความแตกต่างของการประมาณค่าแบบจุดสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และพารามิเตอร์ π ของประชากรที่มีการแจกแจงแบบเบอร์นูลี (Parameter for Bernoulli Distribution)
- เพื่อศึกษาเปรียบเทียบวิธีการระหว่างการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน โดยใช้การประมาณด้วยแจกแจงแบบปกติ และการประมาณด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ ของ M. E. Thompson (1997)

1.3 ขอบเขตของการวิจัย

- ศึกษาแนวคิดและความแตกต่างระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี
- ศึกษาวิธีการและความแตกต่างของการประมาณค่าแบบจุดระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างสุ่มแบบง่ายที่ไม่ใส่คืน และพารามิเตอร์ π ของประชากรที่มีการแจกแจงแบบเบอร์นูลี
- ศึกษาวิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน โดยใช้การประมาณด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ ของ M. E.Thompson (1997)

- ศึกษาวิธีการและทำการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน โดยใช้การประมาณด้วยการแจกแจงแบบปกติ
- ศึกษาเปรียบเทียบวิธีการประมาณค่าสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย

1.4 วิธีดำเนินการวิจัย

- ทำการอธิบายแนวคิดและความแตกต่างระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี จากตำราสถิติทั่วไปทั้งที่เป็นภาษาไทยและภาษาอังกฤษ
- ทำการอธิบายวิธีการและระบุความแตกต่างของการประมาณค่าแบบจุดระหว่างสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และพารามิเตอร์ π ของประชากรที่มีการแจกแจงแบบเบอร์นูลี
- การทดสอบคุณสมบัติการแจกแจงแบบปกติของจำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a) โดยทำการจำลองค่าตัวแปรสุ่ม a ที่มีการแจกแจงแบบไฮเปอร์จีออเมตริก ด้วยโปรแกรม SPSS พารามิเตอร์เท่ากับ (N, n, A) ซึ่งขอบเขตของการศึกษามีดังนี้

ขนาดประชากร (N) ที่ใช้ศึกษา กำหนดให้มีขนาดเท่ากับ 50 100 500 1,000 2,000 2,500 5,000 และ 10,000

กำหนดค่า $P = \dfrac{A}{N}$ มีค่าตั้งแต่ 0.05 ถึง 0.95 โดยทำการเพิ่มค่าขึ้นทีละ 0.05

กำหนดขนาดตัวอย่างในรูปร้อยละของประชากร 4 ระดับ คือ 5 10 15 และ 20

จำนวนรอบของการทำซ้ำซึ่งถือว่าเป็นการกำหนดขนาดตัวอย่างสำหรับการประมาณค่าสัดส่วนประชากร (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 $\alpha = 0.01$

- ทำการเปรียบเทียบวิธีการระหว่างการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน โดยใช้การประมาณด้วยการแจกแจงแบบปกติ และการประมาณด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ ของ M. E. Thompson (1997) โดยทำการจำลองข้อมูลตัวแปรสุ่ม a ให้มีขอบเขตการศึกษาเหมือนกับการทดสอบคุณสมบัติการแจกแจงแบบปกติของจำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a) จากนั้นคำนวณสัดส่วนตัวอย่าง $(p = a / n)$ แล้วคำนวณช่วงความเชื่อมั่น จากการประมาณ 2 วิธี สำหรับการกำหนดจำนวนรอบของการจำลองข้อมูลนี้ถือว่าเป็นการกำหนดขนาดตัวอย่างสำหรับการประมาณค่าพารามิเตอร์ π^* ซึ่งในการศึกษานี้คือระดับความเชื่อมั่น หมายถึงความน่าจะเป็นที่ค่า P ตกอยู่ในช่วงที่สร้างขึ้น ทั้งนี้ในการคำนวณจำนวนรอบกำหนดค่าพารามิเตอร์ π^* 3 ระดับ คือ 0.90 0.95 และ 0.99 ซึ่งมีสูตรการคำนวณดังนี้

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

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

- เปรียบเทียบวิธีการประมาณค่าสัดส่วนประชากรภายใต้แผนการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย โดยใช้ตัวประมาณ 2 ประเภทคือ ตัวประมาณแบบการสุ่มตัวอย่างแบบง่าย ตัวประมาณแบบการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย ซึ่งใช้หลักเกณฑ์การพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error) การกำหนดพารามิเตอร์ในการศึกษานี้ แบ่งออกเป็น 2 กรณี คือ

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

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

สร้างตัวแปรสุ่ม a ที่มีการแจกแจงแบบไฮเปอร์จีออเมตริก ด้วยโปรแกรม SPSS โดยให้มีขอบเขตดังนี้

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

จำนวนรอบของการทำซ้ำมีการกำหนดเงื่อนไขและวิธีการเหมือนกับหัวข้อทดสอบการแจกแจงแบบปกติของจำนวนหน่วยตัวอย่างซึ่งมีลักษณะที่สนใจ (a)

1.5 คำจำกัดความ

สัดส่วนประชากร (Population Proportion) หมายถึง จำนวนหน่วยในประชากรที่มีลักษณะที่สนใจ เทียบกับจำนวนหน่วยทั้งหมดในประชากร

ความน่าจะเป็นของการเกิดผลสำเร็จ (Probability of Success) หมายถึง โอกาสที่จะเกิดเหตุการณ์ที่สนใจในการทดลองแต่ละครั้ง

ตัวอย่างสุ่ม (Random Sample) หมายถึง เซทของตัวแปรสุ่ม n ตัว ที่มีการแจกแจงเดียวกันและเป็นอิสระกัน

การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) หมายถึง วิธีการสุ่มตัวอย่างที่กำหนดให้ตัวอย่างในขนาดที่กำหนดทุกตัวอย่างที่อาจเป็นไปได้มีโอกาสเกิดขึ้นเท่า ๆ กัน

การสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย (Stratified Random Sampling) หมายถึง การสุ่มตัวอย่างแบบมีชั้นภูมิและสุ่มตัวอย่างในทุกชั้นภูมิด้วยวิธีการสุ่มตัวอย่างแบบง่าย

1.6 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

- เพื่อเกิดความเข้าใจที่ชัดเจนในทฤษฎีการสำรวจตัวอย่างเกี่ยวกับการประมาณค่าสัดส่วนประชากรภายใต้การสุ่มตัวอย่างแบบง่ายที่ไม่ใส่คืน และทฤษฎีการอนุมานทางสถิติเกี่ยวกับความน่าจะเป็นของการเกิดผลสำเร็จในการแจกแจงแบบเบอร์นูลี
- ทำให้สามารถนำหลักการทางทฤษฎีไปใช้ในทางปฏิบัติได้ถูกต้อง
- เพื่อเป็นแนวทางในการศึกษาทฤษฎีทางสถิติอื่น ๆ ต่อไป