บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

สถิติเป็นเครื่องมือที่สำคัญของงานวิจัยในสาขาวิชาต่าง ๆ ไม่ว่าจะเป็นทางสังคมศาสตร์ ศึกษาศาสตร์ หรือวิทยาศาสตร์ก็ตาม โดยทั่วไปใช้การอนุมานจากกลุ่มตัวอย่างไปยังประชากร ดังคำกล่าวที่ว่า วัตถุประสงค์ของวิธีการทางสถิติเพื่ออนุมานเกี่ยวกับประชากร โดยอาศัยข้อมูลที่รวบรวมจากกลุ่มตัวอย่างที่นำมาศึกษา ในการอนุมานอาจเป็นการทำนายหรือการตัดสินใจเกี่ยวกับลักษณะบางอย่างของประชากรตามที่ผู้วิจัยสนใจ ดังนั้นอาจกล่าวโดยสรุปได้ว่า สถิติเป็นวิชาหนึ่งในสาขาวิทยาศาสตร์ซึ่งเกี่ยวข้องกับการวางแผนการทดลอง หรือวิธีการเลือกตัวอย่าง การวิเคราะห์ข้อมูล และการอนุมานเกี่ยวกับประชากรจากข้อมูลหรือรายละเอียดที่รวบรวมได้จากตัวอย่าง[1] การอนุมานทางสถิติต้องอาศัยทฤษฎีสถิติ และในทฤษฎีสถิติเกี่ยวกับปัญหาการอนุมานนั้น ลักษณะของประชากรที่ต้องการศึกษามักจะอยู่ในรูปของการแจกแจงของตัวแปรสุ่ม หรือพารามิเตอร์ของการแจกแจงนั้น[2]

เมื่อลักษณะของประชากรอยู่ในรูปของการแจกแจงของตัวแปรสุ่ม ซึ่งการแจกแจงของตัวแปรสุ่มจะเป็นแบบใดแบบหนึ่งต้องเป็นไปตามเงื่อนไขการแจกแจงของตัวแปรสุ่มนั้น การได้ทราบการแจกแจงของตัวแปรสุ่มหรือทราบลักษณะของประชากรนับว่าเป็นประโยชน์อย่างยิ่งต่อการอนุมานทางสถิติ ปัญหาจึงอยู่ที่ว่าเมื่อผู้วิจัยได้วางแผนการวิจัยจนถึงขั้นที่ได้

---

[1] William Menhall, Introduction to Probability and Statistics (California : Duxbury Press, 1971), p. 5

[2] สุชาดา กีระนันทน์, การอนุมานเชิงสถิติ : ทฤษฎีขั้นต้น (กรุงเทพมหานคร : ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2525), หน้า 26

ข้อมูลตัวอย่างมา 1 ตัวอย่าง ผู้วิจัยจะแน่ใจได้อย่างไรว่า ประชากรของกลุ่มตัวอย่างที่นำมาศึกษามีลักษณะอย่างไร โดยทั่วไปนิยมใช้การทดสอบภาวะสารูปดี ซึ่งเป็นการทดสอบสมมติฐานว่าประชากรของกลุ่มตัวอย่างมีการแจกแจงอย่างใดอย่างหนึ่งหรือไม่ การที่จะตั้งข้อสมมติว่าประชากรของกลุ่มตัวอย่างมีการแจกแจงแบบใดจึงเป็นปัญหา ดังนั้น ถ้าผู้วิจัยสามารถประมาณฟังก์ชันการแจกแจงของประชากรจากตัวอย่างก่อน จากนั้นจึงทำการทดสอบสมมติฐานก็จะช่วยลดปัญหาที่เกิดขึ้นอีกทั้งทำให้เกิดความมั่นใจได้ยิ่งขึ้น

ตามที่กล่าวมาแล้วว่าลักษณะของประชากรอยู่ในรูปการแจกแจงของตัวแปรสุ่มนั้น ในกรณีที่ตัวแปรสุ่มหนึ่งมีลักษณะดังนี้ คือ[3]

1. จำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาหนึ่งหรือในอาณาบริเวณหนึ่ง เป็นอิสระกับจำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาอื่น หรืออาณาบริเวณอื่น ซึ่งไม่มีส่วนใดคาบเกี่ยวกัน

2. ความน่าจะเป็นของความสำเร็จ 1 ครั้งที่เกิดขึ้นในช่วงเวลาที่สั้นมาก หรือในอาณาบริเวณที่เล็กมาก เป็นสัดส่วนกับความยาวของช่วงเวลา หรือขนาดของอาณาบริเวณ และจะไม่ขึ้นกับจำนวนครั้งของความสำเร็จที่เกิดขึ้นภายนอกช่วงเวลา หรืออาณาบริเวณดังกล่าว

3. ความน่าจะเป็นของการเกิดความสำเร็จมากกว่า 1 ครั้งในช่วงเวลาสั้น ๆ หรืออาณาบริเวณเล็ก ๆ มีค่าน้อยมาก ถือว่าเป็น 0

ตัวแปรสุ่มที่มีลักษณะตามข้อ 1 2 และ 3 ดังกล่าวเป็นตัวแปรสุ่มที่มีการแจกแจงปัวส์ซงพารามิเตอร์เท่ากับ λ เมื่อ λ เป็นจำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาหนึ่ง หรือในอาณาบริเวณหนึ่งที่เราสนใจ ถ้าให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงปัวส์ซงพารามิเตอร์ λ การแจกแจงของ X เป็นดังนี้

---

[3]Ronald Walpale E. and Raymond H. Myers , <u>Probability and Statistics for Engineers and Scientists</u> (Macmillan Publishing, 1978), p. 99

$$P(X = x) = (e^{-\lambda}\lambda^{x}) / x! \quad , \quad x = 0, 1, 2, \ldots$$
$$= 0 \quad , \quad \text{ที่อื่น ๆ}$$

นอกจากนี้ยังอาจแสดงให้เห็นว่า การแจกแจงปัวส์ซงประมาณมาจากการแจกแจงทวินาม ในกรณีที่การแจกแจงทวินามนั้นมีพารามิเตอร์ n ที่มีค่ามาก ซึ่งแสดงได้โดย[4]

กำหนดตัวแปรสุ่ม X แทนจำนวนครั้งของความสำเร็จ (หรือจำนวนครั้งที่เกิดเหตุการณ์ A) ในช่วงเวลาหนึ่ง

แบ่งช่วงเวลา n ออกเป็นช่วงสั้น ๆ คือ ทำให้ n มีค่ามาก ๆ

ความน่าจะเป็นที่จะเกิดความสำเร็จ 1 ครั้ง ในช่วงเวลา 1/n คือ λ (1/n) (ตามข้อ 2) หรืออาจเขียนได้ว่า

$$p = \lambda\,(1/n) = \lambda/n$$
$$\text{และ} \quad q = 1 - (\lambda/n)$$

ถ้าให้ตัวแปรสุ่ม X มีการแจกแจงทวินามพารามิเตอร์เท่ากับ n และ p การแจกแจงของ X คือ

$$P(X = x) = \binom{n}{x} p^{x}(1 - p)^{n-x} \quad , \quad x = 0, 1, 2, \ldots, n$$
$$= [n! / x!(n - x)!]\, p^{x}(1 - p)^{n-x}$$

---

[4]Robert V. Hogg and Elliot A. Tanis , <u>Probability and Statistical Inference</u> (Macmillan Publishing, 1983), pp. 87-88

จากสมการดังกล่าวเราสามารถแสดงได้ว่า เมื่อ n มีค่ามาก ($n \rightarrow \infty$) สมการดังกล่าวเป็นดังนี้

$$P(X = x) = (\lambda^{x} e^{-\lambda}) / x! \quad , x = 0, 1, 2, \ldots$$

ดังนั้นการแจกแจงของประชากรจะเป็นการแจกแจงปัวส์ซงพารามิเตอร์เท่ากับ λ เมื่อตัวแปรสุ่ม X ต้องมีลักษณะตามข้อ 1 2 และ 3 หรือเป็นตัวแปรสุ่ม X ที่มีการแจกแจงทวินาม โดย n มีค่ามาก ๆ แต่ในทางปฏิบัติโดยเฉพาะงานด้านวิจัย ผู้วิจัยมักจะไม่ได้ทำการตรวจสอบลักษณะต่าง ๆ ดังกล่าว หรือตรวจสอบไม่ได้ วิธีที่นิยมใช้คือการทดสอบภาวะสารูปดี ดังนั้นถ้ามีวิธีการประมาณฟังก์ชันการแจกแจงของประชากรจากข้อมูลตัวอย่างขึ้นมาก่อน จากนั้นจึงทำการทดสอบภาวะสารูปดี ก็จะทำให้ผู้วิจัยสามารถสรุปผลเกี่ยวกับการแจกแจงของประชากรได้อย่างมั่นใจยิ่งขึ้น

ในการวิจัยครั้งนี้ ผู้วิจัยสนใจการประมาณฟังก์ชันการแจกแจงของประชากรที่มีการแจกแจงปัวส์ซง โดยประมาณจากข้อมูลตัวอย่างเพื่อศึกษาว่า รูปแบบของการประมาณฟังก์ชันการแจกแจงของประชากรควรจะเป็นอย่างไร

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อประมาณฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซงด้วยวิธีเคอร์เนล

1.2.2 เพื่อหารูปแบบการประมาณฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซงให้เหมาะสมกับขนาดของกลุ่มตัวอย่าง และลักษณะของข้อมูลตัวอย่าง

1.2.3 เพื่อประมาณฟังก์ชันความน่าจะเป็นของจำนวนเซลล์ที่มีสารอัลคาลอยด์ในใบหญ้าเกล็ดปลา

1.3 <u>ขอบเขตของการวิจัย</u>

1.3.1 ศึกษาการประมาณฟังก์ชันความน่าจะเป็นของประชากรที่มีการแจกแจงปัวส์ซง พารามิเตอร์ λ เท่ากับ 1 2 3 4 และ 5 ตามลำดับ โดยวิธีเคอร์เนล ซึ่งต้องขึ้นกับฟังก์ชันเคอร์เนล และความกว้างของช่วงขนาดต่าง ๆ กัน ฟังก์ชันเคอร์เนลที่ศึกษาคือ

$$K_1(t) = 1 - \left| t \right| \quad , \text{ ถ้า } \left| t \right| < 1$$
$$= 0 \quad , \text{ ค่า } t \text{ อื่น ๆ}$$

และ $$K_2(t) = (1/\sqrt{2\pi})e^{-(1/2)t^2} \quad , \quad -\infty < t < \infty$$

ความกว้างของช่วงที่ใช้ศึกษาคือ

$$h_1 = 1.06\,\sigma n^{-(1/5)}$$

$$h_2 = 0.79\,Rn^{-(1/5)}$$

และ $$h_3 = 0.5$$

1.3.2 คำนวณหาค่าผลรวมความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Sum Square Error) ของฟังก์ชันความน่าจะเป็นที่ประมาณได้จากแต่ละแบบในข้อ 1.3.1 เพื่อศึกษาว่ารูปแบบการประมาณแบบใดจะให้ค่าผลรวมความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

1.3.3 ขนาดตัวอย่างที่ใช้ศึกษาคือ 30 50 และ 100 ตามลำดับ

1.3.4 จำลองข้อมูลตัวอย่างในแต่ละกรณี จำนวน 200 ครั้ง โดยใช้เครื่องคอมพิวเตอร์

1.4 <u>ประโยชน์ที่คาดว่าจะได้รับ</u>

ทำให้ผู้วิจัยได้รูปแบบการประมาณฟังก์ชันความน่าจะเป็นที่เหมาะสมกับข้อมูลตัวอย่าง โดยเฉพาะกับข้อมูลตัวอย่างที่มีขนาด 30 50 และ 100 จากประชากรที่มีการแจกแจงปัวส์ซง พารามิเตอร์ λ เท่ากับ 1 2 3 4 และ 5