บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยต้องการศึกษาเปรียบเทียบการประมาณฟังก์ชันความน่าจะเป็นของประชากรที่มีการแจกแจงปัวส์ซงพารามิเตอร์ λ เท่ากับ 1 2 3 4 และ 5 โดยวิธีเคอร์เนล ขั้นตอนของการวิจัยเป็นดังนี้

1. การประมาณฟังก์ชัน

1.1 ประมาณฟังก์ชัน $\hat{f}$ ซึ่งเป็นฟังก์ชันความหนาแน่นของตัวแปรสุ่มต่อเนื่องจากสูตร

$$\hat{f}(x) = (1/nh)\sum_{i=1}^{n} K[(x-X_i) / h]$$

โดยใช้ฟังก์ชัน K เป็น K_1 และ K_2 ส่วนความกว้างของช่วงใช้ความกว้างเป็น h_1 h_2 และ h_3 ตามลำดับ ซึ่งจะทำให้ได้ฟังก์ชันความหนาแน่น $\hat{f}$ โดยประมาณจาก K_1h_1 K_1h_2 K_1h_3 K_2h_1 K_2h_2 และ K_2h_3 เป็น 6 รูปแบบ

1.2 จากฟังก์ชันความหนาแน่นที่ประมาณได้แต่ละแบบในข้อ 1.1 ผู้วิจัยทำการประมาณความน่าจะเป็นที่แต่ละ X_i (X_i เป็นค่าสังเกตหรือค่าของข้อมูลตัวอย่าง) จะอยู่ระหว่าง $X_i - 0.5$ ถึง $X_i + 0.5$ และเขียนแทนด้วย $\hat{f}(0)$, $\hat{f}(1)$, $\hat{f}(2)$,...

2. ข้อมูลที่ใช้ในการศึกษา เป็นข้อมูลที่ได้จากการจำลองโดยใช้เครื่องคอมพิวเตอร์ โดยจำลองจากประชากรที่มีการแจกแจงปัวส์ซงพารามิเตอร์ λ เป็น 1 2 3 4 และ 5 และใช้ขนาดกลุ่มตัวอย่างเท่ากับ 30 50 และ 100 โดยแต่ละกรณีจะกระทำซ้ำ ๆ กัน 200 ครั้ง การจำลองข้อมูลขึ้นมาศึกษาแต่ละครั้งจะทำการทดสอบภาวะสารูปดี ใช้ $\alpha = 0.05$ โดยสูตร

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

O_i เป็นความถี่ของการเกิดค่า X_i ตามที่จำลองได้

E_i เป็นความถี่ของการเกิดค่า X_i ตามทฤษฎี

X^2 เป็นค่าสถิติทดสอบ มีการแจกแจงโดยประมาณเป็นการแจกแจงไคกำลังสอง องศาอิสระ k - 1

ครั้งใดที่ค่า X_i ที่จำลองได้ ทำให้ค่า X^2 (คำนวณ) มากกว่า X^2 (ตาราง) องศาอิสระ k - 1 แล้วจะไม่ใช้ข้อมูลตัวอย่างชุดนั้น แต่จะทำการจำลองข้อมูลขึ้นมาใหม่

3. คำนวณค่าคาดหวังของ $\hat{f}(x)$ เมื่อ x = 0, 1, 2, ...

จาก $$E\hat{f}(x) = (1/200)\sum_{i=1}^{200} \hat{f}_i(x), \qquad x = 0, 1, 2, \ldots$$

$\hat{f}_i(x)$ คือ ค่าของ $\hat{f}$ ที่ประมาณได้จากข้อมูลตัวอย่างที่จำลองได้ในครั้งที่ i เมื่อ i มีค่า 1, 2, 3, ..., 200

4. คำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squre Error) จากสูตร[19]

$$MSE_x(\hat{f}) = E(\hat{f}(x) - f(x))^2$$

เช่นเมื่อ x = 0 ได้

$$MSE_0(\hat{f}) = (1/200)\,[(\hat{f}_1(0) - f(0))^2 + (\hat{f}_2(0) - f(0))^2 + \ldots$$
$$\ldots(\hat{f}_{200}(0) - f(0))^2]$$

5. คำนวณค่าผลรวมความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Sum Squre Error) ของ $\hat{f}$ เขียนแทนด้วย MSSE($\hat{f}$) ค่า MSSE($\hat{f}$) ประมาณจากค่าของ MISE (Mean Integrated Square Error) ของ $\hat{f}$ ซึ่งหาจากสูตร[20]

---

[19] B.W. Silverman, <u>Density Estimation for Statistics and Data Analysis</u> (London : Chapman and Hall, 1986), p. 45.

[20] Ibid., p. 35.

$$\mathrm{MISE}(\hat{f}) = E \int [\hat{f}(x) - f(x)]^2\, dx$$

ดังนั้น

$$\begin{aligned}\mathrm{MSSE}(\hat{f}) &= E \sum_{x} [\hat{f}(x) - f(x)]^2 \\ &= \sum_{x} E\,[\hat{f}(x) - f(x)]^2 \\ &= \sum_{x} \mathrm{MSE}_x(\hat{f})\end{aligned}$$

6. คำนวณค่าความแปรปรวนของ $\hat{f}(x)$ ที่ได้ในแต่ละครั้งของการจำลองข้อมูล โดยใช้สูตร

$$S^2_x(\hat{f}) = (1/199) \sum_{i=1}^{200} [\hat{f}_i(x) - E\hat{f}(x)]^2$$

การศึกษาครั้งนี้ ประมาณ

$$S^2_x(\hat{f}) = \mathrm{Var}\, \hat{f}(x) = 1/200 \sum_{i=1}^{200} [\hat{f}_i(x) - E\hat{f}(x)]^2$$

ซึ่งได้จากสูตรการหา $\mathrm{MSE}_x(\hat{f})$ ตามข้อ 4 กล่าวคือ

$$\begin{aligned}\mathrm{MSE}_x(\hat{f}) &= E\,[\hat{f}(x) - f(x)]^2 \\ &= [E\hat{f}(x) - f(x)]^2 + \mathrm{Var}\, \hat{f}(x) \\ &= (\mathrm{Bias})^2 + \mathrm{Var}\, \hat{f}(x)\end{aligned}$$

7. ผู้วิจัยเลือกรูปแบบการประมาณฟังก์ชันความน่าจะเป็นของประชากรที่มีการแจกแจงปัวส์ซงพารามิเตอร์ λ เป็น 1 2 3 4 และ 5 เมื่อขนาดของกลุ่มตัวอย่างเป็น 30 50 และ 100 โดยเลือกจากรูปแบบการประมาณที่ทำให้ได้ค่า $\mathrm{MSSE}(\hat{f})$ น้อยที่สุด

8. ใช้รูปแบบการประมาณฟังก์ชันความน่าจะเป็นของประชากรที่เหมาะสม ประมาณฟังก์ชันความน่าจะเป็นของจำนวนเซลล์ที่มีสารอัลคาลอยด์ในใบหญ้าเกล็ดปลา ซึ่งข้อมูลตัวอย่างได้จากการนับจำนวนเซลล์ที่มีสารอัลคาลอยด์ที่มีอยู่ในชิ้นส่วนของใบหญ้าเกล็ดปลาตามที่ส่องพบโดยกล้องจุลทรรศน์ โดยสุ่มใบหญ้าเกล็ดปลามาตัดขวางเป็นชิ้นบาง ๆ ให้ได้กลุ่มตัวอย่างขนาด 30 50 และ 100