หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

1.1 ความสำคัญและที่มาของปัญหา

ปัจจุบันนี้วิชาสถิติมีความสำคัญมากในการวิจัยสาขาวิชาต่าง ๆ โดยเฉพาะสาขาวิทยาศาสตร์และสังคมศาสตร์ เพราะฉะนั้น ผู้วิจัยจะต้องมีความรู้ทางด้านสถิติพอสมควร โดยเฉพาะอย่างยิ่งในขั้นตอนที่เกี่ยวกับการวิเคราะห์และสรุปผล แต่อย่างไรก็ตามแม้นักวิจัยจะมีความรู้ทางสถิติ แต่ก็ไม่ได้หมายความว่า จะต้องเป็นผู้เชี่ยวชาญทางด้านสถิติถึงขั้นเป็นนักสถิติอย่างน้อยที่สุดนักวิจัย ควรรู้วิธีการทางสถิติที่จำเป็น และที่จะนำไปใช้ในการวิเคราะห์ กล่าวคือรู้ถึงวัตถุประสงค์ของวิธีการทางสถิติ วิธีการคำนวณทางสถิติ ข้อกำจัดและความเหมาะสมของแต่ละวิธีที่จะนำมาใช้กับข้อมูล ตลอดจนความสอดคล้องระหว่างสถิติที่ใช้กับเป้าหมายของการวิเคราะห์

เมื่อพิจารณาในแง่ของการวิเคราะห์และสรุปผล อาจแบ่งวิธีการทางสถิติได้เป็น 2 ประเภท ตามลักษณะของข้อมูลคือ

- ประเภทที่ไม่ใช้พารามิเตอร์ หรือไม่ทราบรูปแบบการแจกแจงของข้อมูล (Nonparametric Statistics or Distribution-free Methods)
- ประเภทที่ใช้พารามิเตอร์ (Parametric Statistics) คือทราบรูปแบบการแจกแจงของข้อมูล

สถิติประเภทหลังมีการใช้กันอย่างกว้างขวางในการวิจัย เพราะข้อมูลที่เก็บรวบรวมได้มักจะมีมากพอ สำหรับที่ใช้ในการวิเคราะห์และสรุปผล แต่ในทางปฏิบัติบางครั้งข้อมูลที่วัดหรือเก็บรวบรวมมานั้นมีน้อยเกินไป ที่จะวิเคราะห์โดยใช้สถิติที่ใช้พารามิเตอร์ จึงต้องหันมาใช้วิธีการที่ไม่ใช้พารามิเตอร์แทน

สำหรับสถิติประเภทที่ต้องการทราบรูปแบบการแจกแจง การทดสอบทางสถิติหรือการประมาณค่าพารามิเตอร์ของประชากรจำเป็นจะต้องทราบข้อตกลงเบื้องต้นบางประการเกี่ยวกับรูปแบบของการแจกแจงหรือลักษณะของประชากรที่มีอยู่จริง ทั้งนี้เพราะหลักเกณฑ์ต่าง ๆ เกี่ยวกับวิธีทางสถิติเหล่านี้ได้สร้างขึ้นมาจากข้อสมมติฐานเกี่ยวกับการแจกแจงของประชากรนั้น ๆ เช่น การทดสอบค่าเฉลี่ยของประชากร 2 กลุ่มประชากร ทั้ง 2 กลุ่มจะต้องมีการแจกแจงแบบปกติ และความแปรปรวนของประชากรทั้ง 2 กลุ่มต้องเท่ากัน เป็นต้น

ในกรณีที่ทราบการแจกแจงของข้อมูล จะสามารถทำการวิเคราะห์เชิงสถิติได้ง่าย เนื่องจากการวิเคราะห์เชิงสถิติแบบทราบรูปแบบการแจกแจงได้รับการพัฒนาเป็นอย่างดีแล้ว แต่ในทางปฏิบัติ นักวิจัยส่วนใหญ่มักจะไม่ทดสอบว่าการแจกแจงของข้อมูลที่ศึกษามีลักษณะตามข้อตกลงเบื้องต้นหรือไม่ ซึ่งนับเป็นผลเสีย ข้อตกลงเบื้องต้นนั้นอาจจะเป็นจริงหรือไม่ก็ได้ ในกรณีที่ไม่เป็นจริงจะทำให้ผลที่ได้จากการวิเคราะห์ไม่ถูกต้อง หรือคลาดเคลื่อนจากที่ควรจะเป็นจริงได้มาก

การแจกแจงของข้อมูลจะเป็นชนิดใด อาจบอกได้โดยตรงจากศึกษาการแจกแจงข้อมูลของประชากร แต่โดยปกตินักวิจัยมักจะไม่ได้ศึกษาข้อมูลของประชากรทั้งหมด ทั้งนี้เพราะขนาดของประชากรที่สนใจมักมีขนาดใหญ่มาก ในแต่ละโครงการวิจัย เวลา ทุนทรัพย์และจำนวนพนักงานที่ใช้เก็บข้อมูลมีจำกัด ในทางปฏิบัติจึงมีการสุ่มตัวอย่างจากประชากรที่สนใจมาเพียงบางส่วน และทำการศึกษาข้อมูลจากตัวอย่างเพื่ออนุมานคุณสมบัติของประชากร จากคุณสมบัติของตัวอย่างที่ได้ความถูกต้องและความเชื่อถือได้ของการคาดคะเนคุณสมบัติของประชากรขึ้นอยู่กับการวิเคราะห์ทางสถิติ รวมทั้งข้อกำหนดอื่น ๆ ดังนั้นปัญหาสำคัญคือทำอย่างไร จึงจะทราบว่าข้อมูลของประชากรมีการแจกแจงอย่างไรเมื่อข้อมูลที่มีอยู่เป็นข้อมูลจากตัวอย่างเท่านั้น

มีการทดสอบทางสถิติหลายวิธีที่สามารถบอกได้ว่าข้อมูลจากตัวอย่างที่ได้จะมาจากประชากรมีการแจกแจงคล้ายคลึงกับการแจกแจงมาตรฐานชนิดใด การทดสอบที่ใช้กันอยู่คือ การทดสอบแบบไคสแควร์ (Chi-Square Test) การทดสอบแบบโคลโมโกรอฟ-สเมย์นอฟ (Kolmogorov-Smirnav One Sample Test) และการทดสอบแบบลิลลายฟอรัส (Lillie Forus Test)

ในการทดสอบแต่ละวิธี จะต้องทราบฟังก์ชั่นการแจกแจงของข้อมูลตามทฤษฎีเพื่อใช้คำนวณค่าสถิติที่ใช้ในการทดสอบ ปัญหาที่ตามมาคือ จะหาค่าของพารามิเตอร์ ซึ่งอยู่ในฟังก์ชั่นของการแจกแจงได้อย่างไร โดยปกติผู้ทดสอบจะประมาณค่าพารามิเตอร์จากข้อมูลตัวอย่างที่ได้รับนั่นเอง เช่น ถ้าฟังก์ชั่นการแจกแจงเป็นฟังก์ชั่นของค่าเฉลี่ยจะคำนวณค่าเฉลี่ยจากตัวอย่างมาใช้แทนค่าเฉลี่ยของประชากร หรือถ้าฟังก์ชั่นการแจกแจงขึ้นอยู่กับฟังก์ชั่นของค่าเฉลี่ยและความแปรปรวนก็จะคำนวณค่าทั้งสองจากตัวอย่างมาแทนพารามิเตอร์ที่ต้องการ เมื่อกำหนดค่าพารามิเตอร์แล้วจึงหาค่าสถิติทดสอบ กำหนดขอบเขตของการยอมรับและปฏิเสธ (Acceptance and Rejection Region) สมมติฐานโดยนำค่าสถิติทดสอบจากตารางมาเป็นตัวกำหนดแล้วพิจารณาดูว่าสถิติทดสอบที่คำนวณได้ตกอยู่ในขอบเขตใด ถ้าค่าสถิติทดสอบตกอยู่ในเขตยอมรับ จะสรุปว่าข้อมูลชุดนี้มีการแจกแจงตามสมมติฐานที่ตั้งไว้ แต่ถ้าตกอยู่ในเขตที่จะปฏิเสธสมมติฐานจะสรุปว่าข้อมูลชุดนี้ไม่มีการแจกแจงตามสมมติฐานที่ตั้งไว้

สำหรับในการวิจัยนี้ใช้วิธีการทดสอบแบบไคสแควร์ เป็นที่ทราบกันโดยทั่วไปว่า ถ้าค่าสถิติทดสอบไคสแควร์ต่ำลงมาก ๆ จนเข้าใกล้ศูนย์แล้วแสดงว่าข้อมูลที่ศึกษามีการแจกแจงเข้าใกล้กับฟังค์ชั่นการแจกแจงที่กำหนดมากขึ้น

อย่างไรก็ดียังไม่มีหลักฐานยืนยันได้แน่ชัดว่า ค่าของตัวสถิติทดสอบไคสแควร์ที่คำนวณได้นั้นเป็นค่าที่ดีที่สุด ค่าที่ดีที่สุดของตัวสถิติไคสแควร์ในการวิจัยนี้หมายความว่าเป็นค่าของตัวสถิติทดสอบไคสแควร์ที่ต่ำที่สุดในบรรดาค่าของตัวสถิติทดสอบไคสแควร์ค่าอื่น ๆ ที่ใช้ทดสอบกับข้อมูลชุดเดียวกัน และค่าของตัวสถิติทดสอบเหล่านี้ต่างกันก็เพราะใช้ตัวประมาณค่าของพารามิเตอร์ที่ต่างค่ากัน ทั้งนี้เพราะในการทดสอบการแจกแจงของข้อมูลโดยทั่วไปถ้าผลการทดสอบปฏิเสธสมมติฐานที่ตั้งเอาไว้ เราจะหาค่าของตัวประมาณค่าของพารามิเตอร์ใหม่ แล้วทำการทดสอบการแจกแจงของข้อมูลตามวิธีการเดิม โดยได้ค่าของสถิติทดสอบไคสแควร์ตัวใหม่ และจะทำเช่นนี้ไปเรื่อย ๆ จนกว่าผลการทดสอบยอมรับสมมติฐานใด สมมติฐานหนึ่ง ซึ่งเรียกวิธีการอย่างนี้ว่า CUT AND TRY จะเห็นว่าการยอมรับหรือปฏิเสธสมมติฐานที่ตั้งไว้ ขึ้นอยู่กับค่าของตัวสถิติทดสอบไคสแควร์ ที่ได้จากการกำหนดค่าประมาณของตัวพารามิเตอร์สำหรับข้อมูลที่ทำการวิจัย ถ้าค่าของตัวสถิติทดสอบไคสแควร์ที่ได้ เป็นค่าที่ต่ำที่สุด ซึ่งได้จากการกำหนดค่าประมาณของพารามิเตอร์ และไม่มีค่าประมาณของพารามิเตอร์ค่าอื่น ๆ ที่ทำให้ค่าของตัวสถิติทดสอบไคสแควร์ต่ำกว่านี้อีกแล้วจะทำให้มีความมั่นใจ ในผลการทดสอบที่ยอมรับหรือปฏิเสธสมมติฐานที่ตั้งไว้

ดังนั้นการวิจัยนี้จึงทำการทดลองว่าจะมีเครื่องมือที่ดีพอหรือไม่ ที่จะทำให้เราแน่ใจได้ว่าค่าของตัวสถิติทดสอบไคสแควร์ที่ได้จากการกำหนดค่าประมาณของพารามิเตอร์แบบใดเป็นค่าที่ต่ำที่สุด และค่าประมาณของพารามิเตอร์ที่ได้จากการทำให้ค่าของตัวสถิติทดสอบไคสแควร์ต่ำที่สุดนั้น ยังคงเป็นค่าประมาณของพารามิเตอร์ที่ดีพอสำหรับนำไปใช้งานในขั้นต่อ ๆ ไปทางสถิติได้หรือไม่

1.2 วัตถุประสงค์

1. ศึกษาวิธีการประมาณค่าพารามิเตอร์จาก
   - 1.1 การประมาณค่าจากข้อมูลโดยตรง
   - 1.2 การค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ใช้อนุพันธ์ (Powell's Univariate Search without Derivative)
   - 1.3 การค้นหาตัวแปรแบบ Hooke-Jeeves Pattern Search

2. เปรียบเทียบค่าสถิติทดสอบไคสแควร์ที่ได้จากการประมาณค่าพารามิเตอร์จากวิธีการประมาณค่าจากข้อมูลโดยตรง การค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ใช้อนุพันธ์และการค้นหาตัวแปรแบบ HOOKE-JEEVES PATTERN-SEARCH เมื่อกำหนดขนาดของช่วงข้อมูล (Observed Interval) คงที่

3. เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธีเมื่อขนาดช่วงของข้อมูลเปลี่ยนแปลงไปในหลาย ๆ รูปแบบ

1.3 สมมติฐานของการวิจัย

วิธีการประมาณค่าพารามิเตอร์วิธี HOOKE-JEEVES PATTERN SEARCH จะให้ค่าสถิติทดสอบไคสแควร์ต่ำที่สุด รองลงมาคือวิธีค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ใช้อนุพันธ์ และวิธีประมาณค่าจากข้อมูลโดยตรง ตามลำดับไม่ว่าขนาดช่วงของข้อมูลจะเปลี่ยนแปลงไปในรูปแบบใดก็ตาม

1.4 ขอบเขตการวิจัย

1. ในการศึกษาครั้งนี้จะทำการศึกษาเปรียบเทียบการประมาณค่าพารามิเตอร์โดยวิธีประมาณค่าจากข้อมูลโดยตรง การค้นหาตัวแปรเดียวของพาวเวลล์ และการค้นหาตัวแบบ HOOKE-JEEVES PATTERN SEARCH

2. การแจกแจงของประชากรที่นำมาศึกษานั้นได้แก่ การแจกแจงแบบเอ็กโปเนนเชียล ทวินามและปัวซอง

3. ข้อมูลที่ใช้ในการวิจัยจำลองขึ้นโดยใช้เครื่องคอมพิวเตอร์และใช้เทคนิค MONTE CARLO SIMULATION

4. สุ่มตัวอย่างจากประชากรโดยขนาด 50 100 และ 200

5. สุ่มตัวอย่างจำนวน 1000 ครั้งในแต่ละขนาดตัวอย่าง

6. กำหนดค่าพารามิเตอร์สำหรับการแจกแจงแบบเอ็กโปเนนเชียล (β) เป็น 0.35 0.75 1.50 10 20 30 และ 50

7. กำหนดค่าพารามิเตอร์สำหรับการแจกแจงแบบปัวซอง (λ) เป็น 0.5 1 2 3 20

8. กำหนดค่าพารามิเตอร์สำหรับการแจกแจงแบบทวินาม (n) เป็น 50 100 200

และ P = 0.10 0.25 0.5 0.75

9. กำหนดขนาดช่วงของข้อมูล (INTERVAL SIZE) สำหรับการแจกแจงแบบเอ็กโปเนนเชียล เป็น 0.20 0.80 1.50 2.00 5.00 10 20

1.5 ข้อตกลงเบื้องต้น

1. สถิติทดสอบไคสแควร์ที่ใช้ทดสอบการแจกแจงของประชากรคือ

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

ซึ่งจะเป็น APPROXIMATION CHI-SQUARE DISTRIBUTION

2. การแจกแจงแบบ EXPONENTIAL มีฟังก์ชั่นความหนาแน่นเป็น

$$f(X) = \frac{1}{\beta} e^{\frac{-X}{\beta}}, \quad X \geq 0 \quad \beta > 0$$

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นแนวทางในการหาค่าประมาณของพารามิเตอร์ในการทดสอบภาวะสารูปสนิทดี (Goodness of fit test) ซึ่งตัวสถิติที่ใช้ทดสอบภาวะสารูปสนิทดีไม่จำเป็นต้องใช้ตัวสถิติทดสอบไคสแควร์ อาจจะใช้วิธีการอื่น ๆ ก็ได้

2. ตัวประมาณค่าของพารามิเตอร์ที่ได้จากการใช้เทคนิคการค้นหาค่าอุตมะ (Optimal) ด้วยวิธีการค้นหาตัวแปรเดียวของพาวเวลล์โดยไม่ใช้อนุพันธ์ และวิธีการค้นหาตัวแปรแบบ HOOKE-JEEVES PATTERN SEARCH สามารถนำเอาไปใช้ในการแก้ปัญหาเชิงสถิติได้ กล่าวคือ การนำตัวประมาณค่าของพารามิเตอร์ใหม่นี้ไปใช้ทดสอบจะให้ผลการทดสอบที่มีความเชื่อถือได้มากกว่า ตัวประมาณค่าของพารามิเตอร์ที่หาได้โดยอาศัยวิธีมาตรฐานทั่วไป Conventional Methods)