บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

ในการทดสอบสมมติฐานสำหรับงานวิจัยด้านต่าง ๆ นั้น นอกจากผู้วิจัยจะต้องมีความรู้ความเข้าใจในเรื่องที่จะศึกษาเป็นอย่างดี เพื่อกำหนดแผนแบบการทดลอง หรือเทคนิคการสุ่มตัวอย่างที่เหมาะสมแล้ว ผู้วิจัยจะต้องมีความรู้ในการเลือกวิธีการวิเคราะห์ หรือเลือกสถิติทดสอบที่เหมาะสมกับลักษณะข้อมูล ซึ่งสถิติทดสอบแต่ละวิธีต่างก็มีข้อตกลงเบื้องต้น (Assumptions) เกี่ยวกับลักษณะของข้อมูลที่จะนำมาวิเคราะห์ ดังนั้น การเลือกใช้สถิติทดสอบที่มีข้อตกลงเบื้องต้นสอดคล้องกับลักษณะข้อมูล จะมีผลทำให้การสรุปผลของการวิจัยมีความถูกต้องและเชื่อถือได้มากยิ่งขึ้น

การทดสอบสมมติฐานทางสถิติ เป็นวิธีการหนึ่งที่จะตอบปัญหาที่เกี่ยวข้องกับค่าพารามิเตอร์ของประชากร ไม่ว่าพารามิเตอร์ที่สนใจจะเป็นค่าความแปรปรวน (σ^2) ค่าเฉลี่ย (μ) หรือ ค่าสัดส่วนของสิ่งที่สนใจในประชากรก็ตามต่างก็มีวิธีการทดสอบเพื่อตอบข้อสงสัยของผู้วิจัย ทั้งนี้ต้องอาศัยข้อมูลจากการสำรวจตัวอย่าง เพื่อสรุปอ้างอิงถึงพารามิเตอร์ของประชากรที่เราสนใจ ในที่นี้จะพิจารณาการทดสอบสมมติฐานที่เกี่ยวข้องกับค่าเฉลี่ยของประชากร ซึ่งอาจจะสนใจทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากรเพียง 1 ประชากร หรือ 2 ประชากร โดยสถิติทดสอบที่นำมาใช้ก็คือสถิติทดสอบที (t-test) หรือสถิติทดสอบ Z(Z-test) ซึ่งการเลือกใช้สถิติทดสอบใดนั้นจะต้องพิจารณาลักษณะข้อมูลที่มีอยู่ให้สอดคล้องกับข้อตกลงเบื้องต้นของการทดสอบ จึงจะทำให้ผลการทดสอบที่ได้น่าเชื่อถือมากขึ้น

สำหรับการทดสอบสมมติฐานที่เกี่ยวกับความเท่ากันของค่าเฉลี่ยประชากรในกรณีที่จำนวนประชากรที่สนใจศึกษามีมากกว่า 2 ประชากร และแผนการทดลองเป็นแบบสุ่มสมบูรณ์ (Completely Randomized Design : CRD) นั้นผู้วิจัยมักเลือกใช้วิธีการวิเคราะห์ความแปรปรวนทางเดียว (One-Way Analysis of Variance : ANOVA F-Test) เพื่อทดสอบสมมติฐานดังกล่าว ซึ่งวิธีการดังกล่าวจะให้ผลสรุปที่ถูกต้องและน่าเชื่อถือได้ก็ต่อเมื่อลักษณะของข้อมูลที่จะนำมาวิเคราะห์ จะต้องอยู่ภายใต้ข้อตกลงเบื้องต้นที่ว่า ประชากรต่าง ๆ ที่นำมาทดสอบจะต้องเป็นอิสระต่อกันมีการแจกแจงเป็นแบบปกติและมีค่าความแปรปรวนเท่ากันคือ σ^2 ซึ่งใน

สภาพการณ์โดยทั่วไปลักษณะของข้อมูลที่นำมาวิเคราะห์อาจไม่เป็นไปตามข้อตกลงเบื้องต้นบางประการ กล่าวคือในทางปฏิบัติบ่อยครั้งที่การแจกแจงของประชากรที่ต้องการศึกษาเบี่ยงเบนไปจากการแจกแจงปกติ โดยเฉพาะอย่างยิ่งการแจกแจงของประชากรทางด้านชีววิทยาโดยที่ปลายหางของการแจกแจงมีลักษณะยาวกว่าปลายหางของการแจกแจงแบบปกติหรือเป็นการแจกแจงผิดลอง-เทลด์ (Long-Tailed Distribution) (Bessel 1818 ; Newcomb; Student 1927 ; Jeffreys 1932, อ้างถึงใน Huber 1972 : 1043-1044) ลักษณะตัวอย่างสุ่มที่ได้จากประชากรที่มีการแจกแจงแบบนี้จะมีค่าสังเกตที่มีค่ามากและ/หรือน้อยผิดปกติกว่าค่าสังเกตอื่น ๆ ในชุดตัวอย่างนั้น ซึ่งตัวอย่างสุ่มที่มีลักษณะดังกล่าวนี้ จะเรียกว่าข้อมูลเอาท์ลายเออร์ (Outliers data) เมื่อเป็นเช่นนี้แล้ว ถ้าผู้วิจัยยังคงใช้วิธีการวิเคราะห์ความแปรปรวน ผลสรุปที่ได้อาจขัดแย้งกับความเป็นจริง ทั้งนี้เนื่องจาก ANOVA F-Test จะมีความไว (Sensitive) ต่อลักษณะการแจกแจงของประชากรที่ไม่เป็นไปตามข้อตกลงเบื้องต้น กล่าวคือ ในการทดสอบสมมติฐานเกี่ยวกับความเท่ากันของค่าเฉลี่ยของประชากรนั้น เรามุ่งที่จะตอบข้อสงสัยเกี่ยวกับค่าเฉลี่ยประชากร โดยคาดหวังว่าลักษณะการแจกแจงของประชากรไม่มีผลต่อการสรุปผลของการทดสอบ สมมติว่าในการวิจัยครั้งนั้นสรุปผลได้ว่า ปฏิเสธสมมติฐานหลัก H_o เมื่อข้อมูลมาจากประชากรที่มีการแจกแจงแบบปกติ แต่เมื่อลองเปลี่ยนการแจกแจงของข้อมูลจากการแจกแจงปกติไปเป็นการแจกแจงแบบอื่นแล้ว ผลของการทดสอบกลับยอมรับสมมติฐานหลัก H_o ซึ่งลักษณะเช่นนี้กล่าวได้ว่า วิธีการทดสอบที่นำมาใช้นั้น มีความไวต่อการแจกแจงของประชากรที่ไม่เป็นตามข้อตกลงเบื้องต้น อันจะมีผลทำให้การสรุปผลผิดพลาดไปจากความเป็นจริง

จากปัญหาที่เกิดขึ้นนี้เอง ได้มีนักสถิติหลายท่านพัฒนาวิธีการทดสอบที่มีคุณสมบัติที่ดีและเหมาะสม กล่าวคือ เมื่อลักษณะของข้อมูลที่นำมาทดสอบมีข้อตกลงเบื้องต้นผิดพลาดไปบ้าง ก็จะมีผลกระทบเพียงเล็กน้อยต่อการแจกแจงของตัวอย่างของสถิติทดสอบ (Sampling Distribution of the test) ลักษณะเช่นนี้เราจะเรียกสถิติทดสอบนั้นว่ามีความแกร่ง (Robust) Box(1953) ได้อธิบายความหมายของสถิติทดสอบที่มีความแกร่งไว้ว่า ในกรณีที่ข้อตกลงเบื้องต้นผิดพลาดไปบ้างนั้น สถิติทดสอบที่มีความแกร่งจะต้องไม่ทำให้ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (Type I error) มากขึ้น ดังนั้นคุณสมบัติที่จะชี้ให้เห็นว่า สถิติทดสอบนั้นมีความแกร่งหรือไม่นั้น น่าจะพิจารณาได้จากความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ (Power of the test) นั่นเอง

จากการศึกษาของ Brown and Forsyth (1974) ค้นพบว่าในกรณีข้อมูลที่นำมาวิเคราะห์ไม่เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนแล้ว ผู้วิจัยไม่ควรใช้ ANOVA F-Test เพื่อทดสอบสมมติฐานดังกล่าว ทั้งนี้เพราะว่า ความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 จะมีค่าสูง ซึ่งหมายความว่า ผู้วิจัยจะมีความเสี่ยงสูงต่อการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักถูก ในกรณีนี้ผู้วิจัยอาจจะพิจารณาการทดสอบอื่นที่แกร่งกว่า ANOVA F-Test ซึ่งได้มีนักสถิติหลายท่านทำการศึกษาหรือพัฒนาสถิติทดสอบที่มีความแกร่งหลายตัวด้วยกัน เช่น การทดสอบที่ดัดแปลงมาจากการวิเคราะห์ความแปรปรวน (Modified F-Test) การทดสอบที่ดัดแปลงมาจากสถิติทดสอบของ Welch (Modified Welch Statistic) นอกจากนี้ยังได้มีการพัฒนาสถิติทดสอบ โดยพิจารณาการตัดค่าสังเกตที่ปลายหางทั้งสองด้านของการแจกแจงในกรณีที่ข้อมูลนั้นมาจากการแจกแจงแบบสมมาตรหางยาว (Long-Tailed Distribution; Heavy-Tail Distribution) เช่นสถิติทดสอบแบบ Trimmed F และสถิติทดสอบแบบ Trimmed W ดังนั้นจึงน่าที่จะศึกษาหาข้อสรุปว่า ในกรณีที่ลักษณะการแจกแจงของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนแล้วสถิติทดสอบวิธีใดจะมีความแกร่งมากกว่ากัน โดยพิจารณาระดับการตัดค่าสังเกตที่ปลายหางทั้งสองด้านของการแจกแจงด้วย และด้วยเหตุที่ว่า การคำนวณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบเมื่อลักษณะการแจกแจงไม่เป็นไปตามข้อตกลงเบื้องต้นโดยตรงนั้นกระทำได้ยาก ดังนั้นผู้วิจัยจะทำการศึกษาโดยวิธีมอนติคาร์โล (Monte Carlo Method) ซึ่งเป็นวิธีที่ศึกษาในรูปของการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่มเทียม (Psudo Random Numbers) ซึ่งสามารถระบุขนาดของกลุ่มตัวอย่าง ค่าความแปรปรวน ค่าเฉลี่ยและลักษณะการแจกแจงของประชากรได้ตามที่ผู้วิจัยต้องการศึกษา

1.2 วัตถุประสงค์ของการวิจัย

เพื่อค้นหาสถิติทดสอบที่มีความแกร่ง (Robustness) สำหรับทดสอบสมมติฐานเกี่ยวกับการเท่ากันของค่าเฉลี่ยประชากร เมื่อลักษณะการแจกแจงของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวน โดยใช้สถิติทดสอบดังนี้

1. การวิเคราะห์ความแปรปรวนทางเดียว (ANOVA F-TEST)
2. สถิติทดสอบแบบ Trimmed W (Trimmed W - Statistics)
3. สถิติทดสอบแบบ Trimmed F (Trimmed F - Statistics)

1.3 ขอบเขตของการวิจัย

สำหรับการศึกษาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบของสถิติทดสอบแบบ ANOVA F-TEST สถิติทดสอบแบบ Trimmed W และสถิติทดสอบแบบ Trimmed F เมื่อประชากรทั้ง 3 ชุด มีการแจกแจงแบบเดียวกัน คือ การแจกแจงแบบปกติและการแจกแจงแบบปกติปลอมปน นั้นจำเป็นจะต้องกำหนดปัจจัยที่คาดว่าจะมีผลต่อการศึกษาภายใต้แต่ละการแจกแจงดังต่อไปนี้

1. กำหนดจำนวนประชากรศึกษา 3 ประชากร
2. กำหนดลักษณะการแจกแจงของประชากร

2.1 การแจกแจงแบบปกติ (Normal Distribution)

2.2 การแจกแจงแบบปกติปลอมปน (Contaminated Normal Distribution) โดยที่มีเปอร์เซ็นต์การปลอมปน (p) เท่ากับ 10 และ 20% สเกลแฟคเตอร์ (r) 2 ระดับคือ 5 และ 10 ดังนั้นจะได้การแจกแจงแบบปกติปลอมปน 4 รูปแบบ คือ

2.2.1 การแจกแจงแบบปกติปลอมปนที่ $p = 10\%, r = 5$

2.2.2 การแจกแจงแบบปกติปลอมปนที่ $p = 10\%, r = 10$

2.2.3 การแจกแจงแบบปกติปลอมปนที่ $p = 20\%, r = 5$

2.2.4 การแจกแจงแบบปกติปลอมปนที่ $p = 20\%, r = 10$

3. กำหนดขนาดตัวอย่าง

3.1 ศึกษากรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน 2 ระดับคือ (10, 10, 10) และ (50,50,50)

3.2 ศึกษากรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน 2 ระดับคือ (5,10,15) และ (30,40,50)

4. กำหนดสัดส่วนของส่วนเบี่ยงเบนมาตรฐานของประชากร

4.1 กลุ่มตัวอย่างที่มีขนาดเท่ากัน มีสัดส่วนของส่วนเบี่ยงเบนมาตรฐาน ($\sigma_1 : \sigma_2 : \sigma_3$) เป็น 1:1:1 และ 1:2:3

4.2 กลุ่มตัวอย่างที่มีขนาดไม่เท่ากัน มีสัดส่วนของส่วนเบี่ยงเบนมาตรฐาน ($\sigma_1 : \sigma_2 : \sigma_3$) เป็น 1:1:1, 1:2:3 และ 3:2:1

5. กำหนดสัดส่วนของค่าเฉลี่ยของประชากร

5.1 ในกรณีที่ $\sigma_1 : \sigma_2 : \sigma_3$ เท่ากับ 1 : 1 : 1 จะกำหนดให้มีสัดส่วนของค่าเฉลี่ยประชากร ($\mu_1 : \mu_2 : \mu_3$) เป็น 1:1:1, 1:1:2 และ 1:2:3

5.2 ในกรณีที่ $\sigma_1 : \sigma_2 : \sigma_3$ เท่ากับ 1:2:3 หรือ 3:2:1 จะกำหนดให้มีสัดส่วนของค่าเฉลี่ยประชากร ($\mu_1 : \mu_2 : \mu_3$) เป็น 1:1:1, 1:1:2 และ 2:1:1

6. สำหรับสถิติทดสอบแบบ Trimmed F และ Trimmed W จะกำหนดร้อยละของการตัดค่าสังเกตที่ปลายทั้ง 2 ด้านของการแจกแจงด้านละ 5,10,15 และ 20% ตามลำดับ

7. กำหนดระดับนัยสำคัญของการทดสอบ 2 ระดับคือ .05 และ .01

8. สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามที่ต้องการศึกษาด้วยวิธีมอนติคาร์โล (Monte Carlo Technique) โดยเขียนโปรแกรมภาษาฟอร์แทรน 77 (FORTRAN 77) และประมวลผลด้วยเครื่องคอมพิวเตอร์IBM 370/3031 โดยการทดสอบจะกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ที่ศึกษา

ด้วยปัจจัยต่าง ๆ ที่กล่าวมานี้ จึงเป็นที่น่าสนใจศึกษาเพื่อหาข้อสรุปว่าระหว่างสถิติทดสอบที่ไม่ได้มีการพิจารณาระดับการตัดค่าสังเกต กับสถิติทดสอบที่มีการพิจารณาระดับการตัดค่าสังเกตที่ปลายหางทั้ง 2 ด้าน วิธีการใดจะมีความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจการทดสอบสูงกว่าภายใต้สถานการณ์ที่พิจารณา

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.4 สมมติฐานของการวิจัย

ภายใต้ลักษณะการแจกแจงที่ไม่เป็นแบบปกติ แต่ยังคงมีลักษณะเป็นแบบสมมาตรหางยาวนั้น เมื่อกำหนดให้สัดส่วนของส่วนเบี่ยงเบนมาตรฐานและสัดส่วนของค่าเฉลี่ยของประชากรมีรูปแบบต่าง ๆ กันแล้ว สถิติทดสอบที่มีการกำหนดร้อยละของการตัดค่าสังเกตที่ปลายทั้ง 2 ด้านของการแจกแจง (Trimmed W, Trimmed F) จะมี

1. ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ดีกว่าสถิติทดสอบที่ไม่ได้พิจารณาระดับการตัดค่าสังเกตที่ปลายทั้ง 2 ด้านของการแจกแจง

2. อำนาจการทดสอบสูงกว่า สถิติทดสอบที่ไม่พิจารณาระดับการตัดค่าสังเกตที่ปลายทั้ง 2 ด้านของการแจกแจง

1.5 ข้อกำหนดของการวิจัย

การวิจัยครั้งนี้ถือว่า ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบเป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการพิจารณาเลือกสถิติทดสอบที่เหมาะสมสำหรับแต่ละสถานการณ์ที่กำหนดในขอบเขตของการวิจัย

1.6 คำจำกัดความ

ความแกร่ง (Robustness) ของการทดสอบหมายถึงคุณสมบัติของการทดสอบที่ไม่แสดงความไวต่อการเปลี่ยนแปลงปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การเบี่ยงเบนจากข้อตกลงเบื้องต้นของการทดสอบอันจะมีผลต่อความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

อำนาจของการทดสอบ (Power of the test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง (Null Hypothesis) เมื่อสมมติฐานว่างนั้นผิดซึ่งจะมีค่าเท่ากับ 1-β เมื่อ β คือความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 (Kirk 1969 : 555, Minium 1978: 364)

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นถูก

1.7 ประโยชน์ของการวิจัย

1. ได้สถิติทดสอบที่มีความแกร่งและเหมาะสมกับลักษณะของข้อมูล เพื่อที่ผู้ทำวิจัยในสาขาต่าง ๆ จะได้เลือกนำไปใช้ทดสอบสมมติฐานเกี่ยวกับการเท่ากันของค่าเฉลี่ยประชากร สำหรับกรณีที่ลักษณะการแจกแจงของประชากรที่นำมาทดสอบนั้นมีลักษณะการแจกแจงเป็นแบบปกติและแบบปกติปลอมปน

2. เป็นแนวทางในการวิจัยเพื่อค้นหาสถิติทดสอบใหม่ต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย