บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ข้อมูลสถิติเป็นขั้นตอนที่สำคัญของการวิจัย เพราะจะได้ข้อสรุปของการวิจัย การวิจัยเชิงทดลองจัดอยู่ในการวิจัยประเภทหนึ่ง โดยทั่วไปมุ่งศึกษาเปรียบเทียบความแตกต่างระหว่างทรีทเมนต์ (Treatment) ที่ให้กับหน่วยทดลอง (Experimental Unit) ว่ามีอิทธิพลแตกต่างกันหรือไม่ ซึ่งผู้วิจัยจะต้องมีความเข้าใจในเรื่องที่จะศึกษาและออกแบบการทดลองได้เป็นอย่างดี นอกจากนี้ยังขึ้นอยู่กับเทคนิคในการวางแผน และวิธีวิเคราะห์ข้อมูลทางสถิติ เพื่อทดสอบสมมติฐานสำหรับงานวิจัยนั้นด้วย การเลือกใช้สถิติวิเคราะห์ข้อมูลได้อย่างเหมาะสม จะทำให้ผลสรุปของการวิจัยมีความถูกต้องและสมเหตุผล เพราะแต่ละแผนแบบการทดลองจะมีข้อตกลงเบื้องต้น (Assumptions) ในทางทฤษฎีกำหนดไว้ โดยเฉพาะอย่างยิ่งข้อกำหนดเกี่ยวกับประชากรซึ่งต้องมีการแจกแจงแบบปกติ แต่ในทางปฏิบัติลักษณะข้อมูลที่ใช้มักจะมีการแจกแจงเบี่ยงเบนไปจากการแจกแจงแบบปกติ ทำให้ไม่สอดคล้องกับข้อตกลงจึงเป็นปัญหาต่อผู้วิจัย ในบางครั้งเกิดมีข้อมูลผิดปกติที่มีค่าสูงหรือค่าต่ำมากถูกรวบรวมมาด้วย ส่วนมากจะพบกรณีเช่นนี้ในการศึกษาวิจัยทางด้านสิ่งมีชีวิต เคมี และการแพทย์ ตามที่ W.J. Dixon และ John W. Tukey (1968) ได้กล่าวไว้ อันจะส่งผลต่อการสรุปผลที่คลาดเคลื่อนจากความเป็นจริงได้ ผู้วิจัยจึงควรตรวจสอบข้อมูลเบื้องต้น หากพบข้อมูลที่ผิดปกติจะได้แก้ปัญหาก่อนวิเคราะห์ข้อมูล การตรวจสอบข้อมูลผิดปกติในงานวิจัยเชิงทดลองกระทำได้โดยใช้เทคนิคการตรวจสอบข้อกำหนดของแผนแบบการทดลองประกอบกับวิธีการทางกราฟ Full Normal Plot ของ Tukey (1962) และ Half Normal Plot ของ Daniel (1959) ซึ่งทั้งสองวิธีนี้ Gentleman และ Wilk (1975) ได้ตรวจสอบและยืนยันว่าวิธีกราฟนี้ใช้ได้ดีมีประโยชน์เมื่อเกิดค่าผิดปกติเพียงค่าเดียว

การแก้ปัญหาข้อมูลผิดปกติจึงเป็นสิ่งจำเป็นอย่างยิ่งสำหรับผู้วิจัย ทั้งนี้เพื่อขจัดผลของข้อมูลผิดปกติที่จะเกิดขึ้น การพิจารณาปัญหาข้อมูลผิดปกติเลือกได้หลายวิธี เช่น การไม่ยอมรับ

การยอมรับ หรือการปรับปรุงข้อมูลนั้น (Barnett และ Lewis 1979 : 287) การประมาณค่าแทนข้อมูลผิดปกติโดยสมมติว่าเป็นข้อมูลสูญหาย นอกจากนั้นการใช้วิธีวิเคราะห์แบบนอนพาราเมติก (Nonparametric Test) ซึ่งไม่คำนึงถึงข้อกำหนดของการแจกแจงของประชากรเหมือนกับการทดสอบแบบพาราเมตริก (Parametric Test) และกล่าวได้ว่าเป็น Distribution - Free (Siegel 1956 : 3) ก็สามารถปฏิบัติได้เช่นเดียวกัน

การเลือกใช้เทคนิคการแก้ปัญหาข้อมูลผิดปกติต่าง ๆ เหล่านี้ ขึ้นอยู่กับดุลยพินิจของผู้วิจัยเป็นสำคัญ

ดังนั้น จึงควรมีการศึกษาเปรียบเทียบการแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากงานวางแผนการทดลองโดยใช้อำนาจการทดสอบ เพื่อจะได้นำมาใช้ในทางปฏิบัติได้อย่างเหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบการแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากงานวางแผนการทดลอง โดยการตัดข้อมูลผิดปกติ การยอมรับ การประมาณค่าข้อมูลผิดปกติ และการวิเคราะห์ข้อมูลแบบนอนพาราเมตริก โดยใช้การพิจารณาจากอำนาจการทดสอบ

1.3 สมมติฐานของการวิจัย

การแก้ปัญหาข้อมูลผิดปกติโดยใช้วิธีวิเคราะห์แบบนอนพาราเมตริก ซึ่งไม่คำนึงถึงการแจกแจงของประชากร จะให้อำนาจการทดสอบดีกว่าวิธีการแก้ปัญหาด้วยวิธีอื่น

1.4 ข้อตกลงเบื้องต้น

1. การวิจัยครั้งนี้ถือว่า อำนาจการทดสอบเป็นดัชนีสำคัญในการใช้เป็นเกณฑ์วินิจฉัยเลือกการแก้ปัญหาข้อมูลผิดปกติ

2. ศึกษาเฉพาะกรณีเกิดข้อมูลผิดปกติค่าเดียว (Single Outlier) โดยทำให้เป็นค่าสูงสุด โดยกำหนดเกณฑ์ค่าสังเกตที่มีค่ามากกว่าค่าขอบเขตบน (Upper Confidence Interval) ของช่วงเชื่อมั่น 99% ของค่าเฉลี่ยเลขคณิตของตัวอย่าง (มนตรี พิริยะกุล 2524 : 15) ซึ่งจะเห็นได้ว่าค่าข้อมูลที่อยู่ในช่วงนี้มีจำนวนน้อย ดังนั้น จึงจัดได้ว่าเป็นค่าสูงมาก และทำให้ผลการทดสอบข้อกำหนดของแผนแบบการทดลองไม่เป็นไปตามที่กำหนดไว้

3. การวิเคราะห์ข้อมูลแบบนอนพาราเมตริก จะใช้ในกรณีไม่มีข้อมูลซ้ำ

1.5 ขอบเขตของการวิจัย

1. เปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของผลการวิเคราะห์ข้อมูลเมื่อได้แก้ปัญหาข้อมูลผิดปกติ เฉพาะภายใต้การแจกแจงแบบต่าง ๆ คือ การแจกแจงแบบโลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และสเกลคอนทามิเนทน์อร์มอลที่เปอร์เซนต์คอนทามิเนทต์ (Percent Contaminate) และสเกลแฟคเตอร์ (Scale Factor) ต่าง ๆ ดังนี้

1.1 เปอร์เซนต์คอนทามิเนทต์ 1 และสเกลแฟคเตอร์ 3

1.2 เปอร์เซนต์คอนทามิเนทต์ 5 และสเกลแฟคเตอร์ 3

1.3 เปอร์เซนต์คอนทามิเนทต์ 10 และสเกลแฟคเตอร์ 3

1.4 เปอร์เซนต์คอนทามิเนทต์ 25 และสเกลแฟคเตอร์ 3

1.5 เปอร์เซนต์คอนทามิเนทต์ 1 และสเกลแฟคเตอร์ 10

1.6 เปอร์เซนต์คอนทามิเนทต์ 5 และสเกลแฟคเตอร์ 10

1.7 เปอร์เซนต์คอนทามิเนทต์ 10 และสเกลแฟคเตอร์ 10

1.8 เปอร์เซนต์คอนทามิเนทต์ 25 และสเกลแฟคเตอร์ 10

2. แผนแบบการทดลองที่จะศึกษา

2.1 การเปรียบเทียบความแตกต่างระหว่าง 2 ประชากร

2.1.1 จำนวนซ้ำ (Replication) เท่ากัน เท่ากับ 5 15 และ 20

2.1.2 การแก้ปัญหาข้อมูลผิดปกติ

2.1.2.1 การยอมรับข้อมูลผิดปกติ

2.1.2.2 การตัดข้อมูลผิดปกติ

2.1.2.3 การประมาณค่าข้อมูลผิดปกติ

2.1.2.3.1 ใช้ค่าเฉลี่ยเลขคณิตของข้อมูลเมื่อตัดข้อมูลผิดปกติออก

2.1.2.3.2 ใช้ค่ามัธยฐาน

2.1.2.3.3 ใช้ค่าใกล้เคียงกับข้อมูลผิดปกติ

2.1.2.4 การวิเคราะห์ข้อมูลแบบนอนพาราเมตริกด้วยวิธี Mann - Whitney U Test

2.2 แผนการทดลองแบบสุ่มตลอด

2.2.1 จำนวนทรีทเมนต์ เท่ากับ 3

2.2.2 จำนวนซ้ำ (Replication) เท่ากัน เท่ากับ 3 4 7 และ 12

2.2.3 การแก้ปัญหาข้อมูลผิดปกติ

2.2.3.1 การยอมรับข้อมูลผิดปกติ

2.2.3.2 การตัดข้อมูลผิดปกติ

2.2.3.3 การประมาณค่าข้อมูลผิดปกติ

2.2.3.3.1 ใช้ค่าเฉลี่ยเลขคณิตของข้อมูลเมื่อตัดข้อมูลผิดปกติออก

2.2.3.3.2 ใช้ค่ามัธยฐาน

2.2.3.3.3 ใช้ค่าใกล้เคียงกับข้อมูลผิดปกติ

2.2.3.4 การวิเคราะห์ข้อมูลแบบนอนพาราเมตริกด้วยวิธี Kruskal Wallis One - Way Analysis Of Variance

2.3 แผนการทดลองแบบสุ่มในบล็อก

2.3.1 จำนวนทรีทเมนต์ เท่ากับ 3

2.3.2 จำนวนบล็อกเป็น 3 4 และ 5

2.3.3 การแก้ปัญหาข้อมูลผิดปกติ

2.3.3.1 การยอมรับข้อมูลผิดปกติ

2.3.3.2 การคำนวณค่าข้อมูลสูญหาย (Missing Value)

2.3.3.3 การวิเคราะห์ข้อมูลแบบนอนพาราเมตริกด้วยวิธี Friedman Two - Way Analysis Of Variance

3. ใช้สัมประสิทธิ์ของความผันแปรของประชากร (Coefficient Of Variation ; C.V.) เท่ากับ 10% 20% และ 30%

4. เปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ เฉพาะการทดสอบแบบสองด้าน (Two - Tailed Test) ณ ระดับนัยสำคัญ .01 .05 และ .10

5. ในการพิจารณาความคลาดเคลื่อนประเภทที่ 1 กำหนดค่าเฉลี่ยเลขคณิตของประชากร เท่ากับ 100

6. ในการพิจารณาอำนาจการทดสอบ ใช้สัดส่วนความแตกต่างของค่าเฉลี่ยเลขคณิตของประชากร ดังนี้

6.1 กรณี 2 ประชากร กำหนดสัดส่วน 1:1.25

6.2 กรณี 3 ประชากร กำหนดสัดส่วน 1:1.25:1.50

7. การวิจัยครั้งนี้จำลองการทดลองโดยใช้เครื่องคอมพิวเตอร์ด้วยเทคนิคการจำลองแบบ (Simulation Technique) โดยทำการทดลองซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) คือความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นจริง ใช้แทนด้วยสัญญลักษณ์ α

ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) คือความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง ใช้แทนด้วยสัญญลักษณ์ β

อำนาจการทดสอบ (Pewer Of The Test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง ซึ่งมีค่าเท่ากับ $1 - \beta$

การแจกแจงแบบสเกลคอนทามิเนทด์นอร์มอล ที่เปอร์เซนต์คอนทามิเนทด์ p และสเกลแฟคเตอร์ c คือการแจกแจงของประชากรที่สร้างขึ้นเพื่อให้มีลักษณะเป็นการแจกแจงชนิดลองเทลด์ โดยส่วนหนึ่งของประชากรมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ ความแปรปรวน σ^2 ด้วยความน่าจะเป็น $1-(p/100)$ และอีกส่วนหนึ่งมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน $c^2\sigma^2$ ด้วยความน่าจะเป็น $p/100$

ค่าใกล้เคียงข้อมูลผิดปกติ คือ ค่าข้อมูลตัวอย่างที่ใกล้เคียงกับค่าข้อมูลผิดปกติมากที่สุด

1.7 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

เพื่อให้ผู้วิจัยใช้เป็นแนวทางในการเลือกใช้วิธีการแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากงานวางแผนการทดลองได้อย่างเหมาะสม