หอสมุดกลาง จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

แผนการทดลองแบบสุ่มภายในบล็อคสมบูรณ์ (Randomized Complete Block Design) เป็นแผนการทดลองเพื่อทดสอบความแตกต่างระหว่างอิทธิพลของทรีทเมนต์ (Treatment Effect) เมื่อหน่วยทดลอง (Experimental Unit) มีความแตกต่างกัน 2 ทาง แผนการทดลองแบบนี้เป็นแผนการทดลองที่นิยมใช้กันทั่วไปในทางปฏิบัติ

ในการทดสอบความแตกต่างระหว่างอิทธิพลของทรีทเมนต์ เมื่อมีจำนวนทรีทเมนท์มากกว่าสองทรีทเมนต์ขึ้นไป ทดสอบโดยการวิเคราะห์ความแปรปรวน (Analysis of Variance) ในการทดสอบถ้าผลของการวิเคราะห์ไม่มีนัยสำคัญทางสถิติ หมายความว่า ค่าเฉลี่ยของทุกประชากรเท่ากัน แต่ถ้าผลการวิเคราะห์มีนัยสำคัญทางสถิติ นั่นมิได้หมายความว่า ค่าเฉลี่ยทุกประชากรแตกต่างกัน แต่หมายความเพียงว่ามีอย่างน้อยค่าเฉลี่ยของประชากรหนึ่งแตกต่างจากค่าเฉลี่ยของประชากรอื่น ในกรณีเช่นนี้ ถ้าเราต้องการวิเคราะห์ต่อไปว่าประชากรใดมีค่าเฉลี่ยแตกต่างกันบ้างนั้น จะใช้วิธีการเปรียบเทียบที่เรียกว่า การเปรียบเทียบเชิงพหุ (Multiple Comparison)

ในการวิจัยผู้วิจัยมักต้องการผลสรุปที่ถูกต้องและเชื่อถือได้ ซึ่งจะเป็นไปได้ก็โดยที่การวิจัยนั้นจะต้องออกแบบการทดลองที่เหมาะสม และเลือกใช้สถิติทดสอบที่เหมาะสมกับลักษณะของข้อมูลที่ต้องการวิเคราะห์นั้น โดยจะต้องคำนึงถึงข้อตกลงเบื้องต้นของสถิติทดสอบแต่ละวิธี แต่ในการวิจัยโดยทั่ว ๆ ไป ปัญหาที่ผู้วิจัยพบมากก็คือ ลักษณะการแจกแจงของความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น เช่น การแจกแจงของความคลาดเคลื่อนอาจจะเบ้หรือโด่งไปจากการแจกแจงปกติ ซึ่งจำเป็นต้องมีการแก้ไขด้วยการแปลงข้อมูล (Transformation of Data) เหล่านั้นเสียก่อน จึงนำข้อมูลที่แปลงแล้วมาวิเคราะห์ จึงจะได้ผลถูกต้องตามความเป็นจริง การทดสอบแบบนี้เรียกว่า การทดสอบแบบพาราเมตริก (Parametric Test) ซึ่งจะเห็นได้ว่า ขั้นตอนการทดสอบนั้นค่อนข้างยุ่งยาก ในกรณีเช่นนี้ ผู้วิจัยอาจจะเลือกใช้

สถิติทดสอบแบบนอนพาราเมตริก (Nonparametric Test) โดยที่สถิติทดสอบแบบนอนพารา-เมตริกนี้ จะช่วยแก้ปัญหาเกี่ยวกับข้อสมมุติต่าง ๆ และการแปลงข้อมูล เพราะว่าเป็นสถิติทดสอบที่ไม่มีข้อจำกัดเกี่ยวกับการแจกแจงของประชากร และยังสามารถคำนวณได้รวดเร็ว ทำความเข้าใจได้ง่าย นอกจากนี้นยังใช้ได้กับข้อมูลที่มีสเกลการวัดแบบนามสเกล (Nominal Scale) หรือแบบจัดเรียงลำดับ (Ordinal Scale)

สำหรับการเปรียบเทียบเชิงพหุที่ใช้อยู่ในปัจจุบันมีด้วยกันหลายวิธี คือ วิธีของทูกี้ (Tukey Test) วิธีของนิวแมน-คูลส์ (Newman-Keuls Test) วิธีของเชฟเฟย์ (Scheffe´ Test) วิธีเปรียบเทียบแบบช่วงพหุคูณ (Duncan's new Multiple Range Test) และวิธีเปรียบเทียบแบบความแตกต่างอย่างมีนัยสำคัญน้อยที่สุด (Least Significant Difference = LSD) ซึ่งวิธีที่กล่าวมาทั้งหมดนี้เป็นสถิติทดสอบแบบพาราเมตริก แต่ในการวิจัยครั้งนี้จะเลือกศึกษาเฉพาะ 3 วิธีแรกที่กล่าว เนื่องจากวิธีเปรียบเทียบแบบช่วงพหุคูณ ไม่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้[1] และวิธีเปรียบเทียบแบบความแตกต่างอย่างมีนัยสำคัญน้อยที่สุด ซึ่งเป็นวิธีที่ฟิชเชอร์นำมาใช้นั้น จะไม่นำมาศึกษาเนื่องจากวิธีนี้มีความผิดพลาดมากเพราะจะเพิ่มโอกาสให้ปฏิเสธสมมติฐานที่เป็นจริงมากขึ้น ทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 (Type I error) ได้มาก (Roscoe; 1975. P. 313) ส่วนการเปรียบเทียบเชิงพหุแบบนอนพารามเตริกนั้น วิธีที่ศึกษาในการวิจัยครั้งนี้ คือ วิธีของฟรีดแมน (Friedman's Test) และวิธีของด็อกซัม (Doksum's Test) ซึ่งทั้ง 2 วิธีนี้ยังไม่มีผู้ศึกษาในเรื่องของความคลาดเคลื่อนประเภทที่ 1 และกำลังการทดสอบเลย

การพิจารณาเปรียบเทียบกำลังการทดสอบนั้นเป็นหนทางนำไปสู่การตัดสินใจเลือกวิธีทดสอบ ทั้งนี้จะพิจารณาตามคำกล่าวของเนย์แมน (. Neyman 1950: 265 อ้างโดย Direk Srisukho 1970: 38) ซึ่งกล่าวว่า "เมื่อต้องการที่จะเลือกใช้สถิติทดสอบ เราต้องเริ่มพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 (Type I error) ก่อน

---

[1] Lewis F. Petrinovich and Curtis D. Hardyck: Error rates for Multiple Comparison Methods; Psychological Bulletin 1969 Vol. 71 No. 1, 43-54.

โดยยอมให้ความน่าจะเป็นที่เกิดความคลาดเคลื่อนประเภทที่ 1 ไม่เกินขอบเขตของเกณฑ์ที่กำหนดไว้" นั่นคือสถิติทดสอบนั้นต้องมีความแกร่ง (Robustness) หรือสถิติทดสอบนั้นจะต้องมีความไว (sensitive) ต่อการเปลี่ยนแปลงของปัจจัยที่ต้องการทดสอบ และจะต้องไม่มีความไว (Insensitive) ต่อการเปลี่ยนแปลงของสิ่งอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ

ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการคำนวณความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบทั้ง 5 วิธี เมื่อประชากรมีการแจกแจงแบบปกติ โลจิสติก ดับเบิ้ลเอ็กซ์โปเนนเชียล ปกติปลอมปน และแบบเบ้ ซึ่งกระทำได้ยากมาก ดังนั้นการวิจัยครั้งนี้จะศึกษาโดยใช้วิธีมอนติคาร์โล (Monte Carlo method) ซึ่งสามารถกำหนดขนาดการทดลอง ลักษณะการแจกแจงของประชากร ค่าเฉลี่ย และความแปรปรวนได้

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบกำลังการทดสอบแบบพาราเมตริก และนอนพาราเมตริก ของการเปรียบเทียบเชิงพหุ ของสถิติทดสอบแบบต่าง ๆ คือ

1.2.1 สถิติทดสอบแบบพาราเมตริก

1.2.1.1 วิธีของทูกี้

1.2.1.2 วิธีของนิวแมน-คูลส์

1.2.1.3 วิธีของเชฟเฟ่ย์

1.2.2 สถิติทดสอบแบบนอนพาราเมตริก

1.2.2.1 วิธีของฟรีดแมน

1.2.2.2 วิธีของด็อกซั่ม

1.3 สมมติฐานของการวิจัย

เมื่อข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของการเปรียบเทียบเชิงพหุ กำลังการทดสอบแบบพาราเมตริก กับแบบนอนพาราเมตริก จะแตกต่างกัน

1.4 ข้อตกลงเบื้องต้น

1.4.1 แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ เป็นกรณีที่ไม่มีหน่วยตัวอย่างย่อย และจำนวนหน่วยตัวอย่างในแต่ละทรีทเมนต์เท่ากัน

1.4.2 เป็นแบบ Fixed Effect

1.4.3 การทดลองของฟรีดเมนต์และด็อกซั่มจะใช้ในกรณีไม่มีข้อมูลซ้ำ

1.4.4 ถือว่ากำลังการทดสอบ เป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการเลือกสถิติทดสอบ

1.5 ขอบเขตการวิจัย

1.5.1 ศึกษาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และกำลังการทดสอบเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ โลจิสติส ดับเบิ้ลเอ็กซ์โปเนนเชียล ปกติปลอมปน และแบบเบ้

1.5.2 กำหนดระดับนัยสำคัญของการทดสอบ เท่ากับ 0.05 และ 0.01

1.5.3 ประชากรที่ศึกษา เท่ากับ 3 และ 5 ประชากรจำนวนบล็อกที่ศึกษา เท่ากับ 5 10 และ 15 บล็อค

1.5.4 ลักษณะการแจกแจงในกลุ่มประชากรทั้งหมดมีรูปแบบเดียวกัน

1.5.5 เมื่อพิจารณาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ใช้ความแปรปรวน และค่าเฉลี่ยของประชากรเป็น 100 100 ตามลำดับ

1.5.6 คำนวณค่ากำลังการทดสอบเป็นจุด ๆ ที่เดลต้ามากที่สุด ($\delta_{max} = |\mu_i - \mu_j|$) มีค่าต่าง ๆ เช่น 1.0σ, 1.5σ, 2.0σ, 2.5σ

1.5.7 ในกรณีการแจกแจงแบบปกติปลอมปน จะศึกษาที่เปอร์เซ็นต์การปลอมปนเป็น 10 % และ 25 % สำหรับสเกลเฟคเตอร์* (Scale factor) ที่ 2 ระดับ คือ 10 และ 30

---

* สเกลเฟคเตอร์ หมายถึง ค่าที่ปรับให้ข้อมูลมีการกระจายมากขึ้น โดยสเกลเฟคเตอร์ที่มีค่าสูงจะทำให้เกิดค่าผิดปกติในข้อมูล และจากการศึกษาที่ผ่านมาพบว่า สเกลเฟคเตอร์ที่มีค่าน้อยกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลเฟคเตอร์ที่มีค่ามากกว่า 30 จะมีโอกาสของการเกิดค่าผิดปกติมาก จึงทำการศึกษาเพียง 2 ระดับดังกล่าว

1.5.8 ในกรณีการแจกแจงแบบเบ้ จะศึกษาเมื่อ

ก. กำหนดความเบ้ เท่ากับ 0 และความโด่งเท่ากับ 2.0 และ 2.4

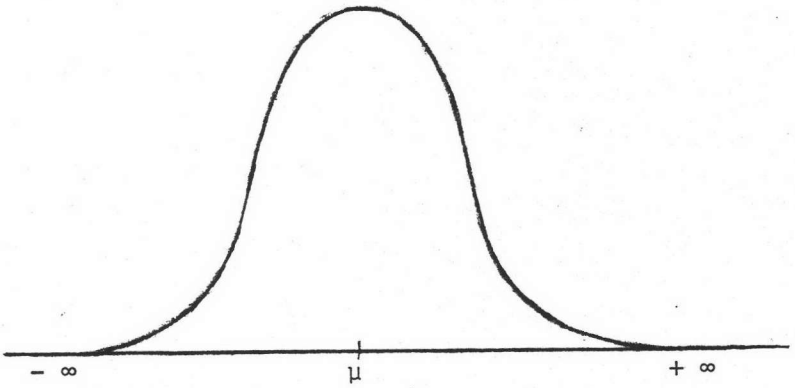
ข. กำหนดความเบ้ เท่ากับ .25 และ .50 และความโด่งเท่ากับ 2.4 3.0 และ 4.2

1.5.9 ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 100*

1.5.10 การวิจัยครั้งนี้จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โล จากเครื่อง IBM 370/3031 ซึ่งจะศึกษาเมื่อความคลาดเคลื่อน มีการแจกแจงแบบปกติ (Normal distribution) โลจิสติค (Logistic distribution) ดับเบิ้ลเอ็กซ์โปเนนเชียล (Double exponential distribution) ปกติปลอมปน (Scale-Contaminated normal distribution) และแบบเบ้ (Skewed distribution) โดยมีฟังก์ชั่นความน่าจะเป็นค่าคาดหวังและความแปรปรวนของการแจกแจงแต่ละรูปแบบเป็นดังนี้

การแจกแจงปกติ

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < +\infty$$



รูปที่ 1.1 เส้นโค้งของการแจกแจงแบบปกติ

---

* การวิจัยในครั้งนี้กำหนดค่าความแปรปรวนเป็นค่าดังกล่าว เนื่องจากการทดลองกระทำ พบว่า ค่าความแปรปรวนไม่มีผลต่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$E(x) = \mu$

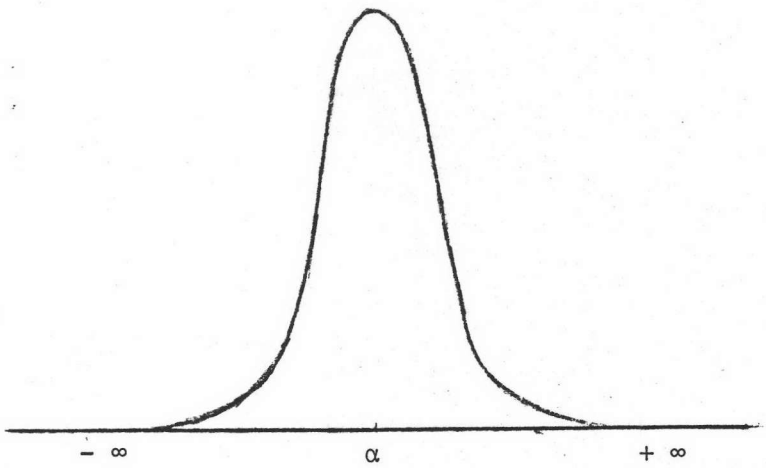
$V(x) = \sigma^2$

Skewness = 0.

Kurtosis = 3.0

<u>การแจกแจงแบบโลจิสติค</u>

$$f(x) = \frac{1}{\beta} \frac{\ell^{-(\frac{x-\alpha}{\beta})}}{\left[1 + \ell^{-(\frac{x-\alpha}{\beta})}\right]^2}, \quad -\infty < x < +\infty, \quad \alpha, \beta > 0$$



รูปที่ 1.2 เส้นโค้งของการแจกแจงแบบโลจิสติค

$E(x) = \alpha$

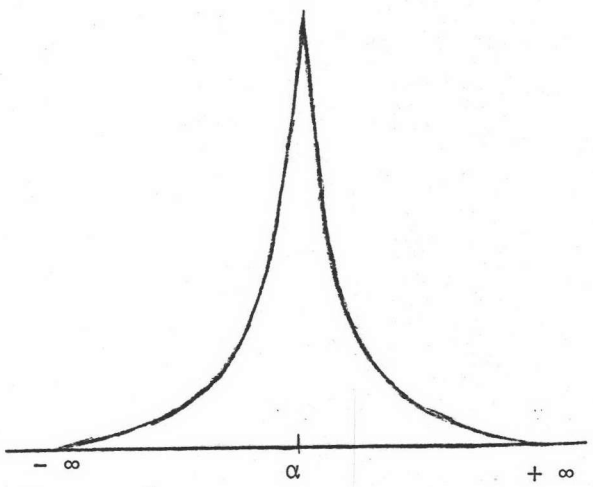
$V(x) = \frac{1}{3} \pi^2 \beta^2$

Skewness = 0.

Kurtosis = 4.2

การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

$$f(x) = \frac{1}{2\beta} \ell^{-\left|\frac{x-\alpha}{\beta}\right|}, \quad -\infty < x < \infty, \quad -\infty < \alpha < \infty, \quad \beta > 0$$



รูปที่ 1.3 เส้นโค้งของการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

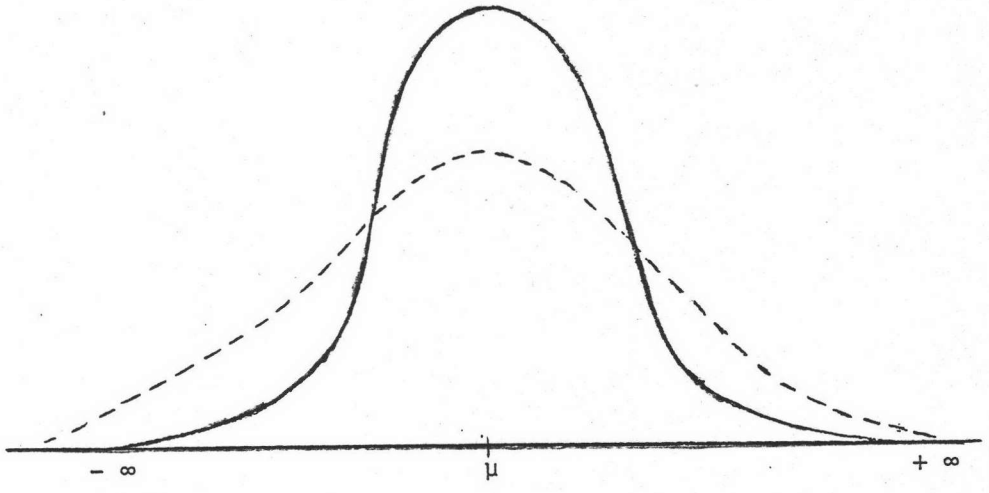
$E(x) = \alpha$

$V(x) = 2\beta^2$

Skewness = 0

Kurtosis = 6.0

การแจกแจงแบบปกติปลอมปน



รูปที่ 1.4 เส้นโค้งของการแจกแจงแบบปกติปลอมปน

ลักษณะการแจกแจงแบบปกติปลอมปนที่พิจารณาในที่นี้เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันการแปลงเป็นดังนี้

$$X = (1 - P)N(\mu, \sigma^2) + PN(\mu, c^2\sigma^2) \quad , \quad c > 0$$

หมายความว่า X มาจากการแจกแจง $N(\mu, \sigma^2)$ ด้วยความน่าจะเป็น $(1 - P)$ และมาจากการแจกแจง $N(\mu, c^2\sigma^2)$ ด้วยความน่าจะเป็น P

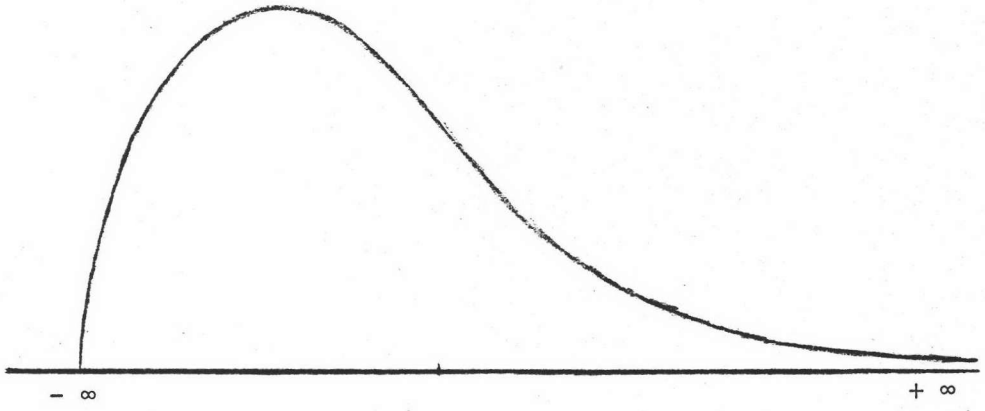
ในที่นี้ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของความคลาดเคลื่อน

P และ c เป็นค่าที่กำหนดสัดส่วนการปลอมปนและสเกลแฟคเตอร์ ซึ่งจะทำให้เกิดค่าผิดปกติ*

---

*ค่าผิดปกติ หมายถึง ค่าสังเกตที่มีค่ามากหรือน้อยกว่าค่าสังเกตอื่น ๆ อย่างผิดปกติ

การแจกแจงแบบเบ้



รูปที่ 1.5 เส้นโค้งของการแจกแจงแบบเบ้

ลักษณะการแจกแจงแบบเบ้ที่พิจารณาในที่นี้ เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบยูนิฟอร์ม ซึ่งเรียกว่า Generalized Lambda Distribution (GLD) ซึ่งมีฟังก์ชั่นการแจกแจงเป็นดังนี้

$$X = \lambda_1 + (P^{\lambda_3} - (1-P)^{\lambda_4}) / \lambda_2 \quad , \quad 0 < P < 1$$

P มีการแจกแจงแบบยูนิฟอร์มในช่วง (0, 1)

λ_1 และ λ_2 เป็นพารามิเตอร์ที่กำหนดตำแหน่ง (Location parameter) และพารามิเตอร์ที่กำหนดสเกล (Scale parameter) ตามลำดับ

λ_3 และ λ_4 เป็นพารามิเตอร์ที่กำหนดรูปแบบ (Shape) ของการแจกแจง คือ ความเบ้ (Skewness) และความโด่ง (Kurtosis) ตามลำดับ

สำหรับค่าคาดหวัง (E(x)) ความแปรปรวน (V(x)) ความเบ้ และความโด่งของ X แสดงไว้ในภาคผนวก

1.5.10 การจำลองการทดลองจะกระทำซ้ำกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

1.6.1 กำลังการทดสอบ (Power of the test) หมายถึงความน่าจะเป็นที่จะปฏิเสธ (reject) สมมติฐานว่าง (null Hypothesis) เมื่อสมมติฐานว่างนั้นไม่จริง มีค่าเท่ากับ 1 - β เมื่อ β คือ ความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 2 (Probability of type II error)

1.6.2 ความคลาดเคลื่อนประเภทที่ I (Type I error) คือการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นจริง

1.6.3 ความคลาดเคลื่อนประเภทที่ II (Type II error) คือ การยอมรับ (Accept) สมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง

1.6.4 ความแกร่ง (Robustness) ของการทดสอบ หมายถึง คุณสมบัติของการทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น ซึ่งสิ่งที่ใช้พิจารณาความแกร่งของการทดสอบคือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

1.7 ประโยชน์ของการวิจัย

เพื่อให้ผู้ใช้สถิติมีข้อสรุป และหลักฐานเกี่ยวกับสถิติที่ใช้ในการเปรียบเทียบเชิงพหุ และสามารถเลือกใช้สถิติทดสอบได้อย่างถูกต้องและเหมาะสมกับการแจกแจงของประชากร