หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความสำคัญและความเป็นมาของปัญหา

ในการทดสอบสมมติฐานซึ่งเกี่ยวข้องกับประชากรหนึ่งกลุ่มหรือมากกว่าหนึ่งกลุ่มนั้น การเลือกใช้สถิติทดสอบที่เหมาะสมกับลักษณะของข้อมูลที่ต้องการวิเคราะห์นั้นจะต้องคำนึงถึงข้อตกลงเบื้องต้นของสถิติทดสอบแต่ละวิธี ในการวิจัยโดยทั่ว ๆ ไป ปัญหาที่ผู้วิจัยพบมากก็คือ ลักษณะการแจกแจงของประชากรไม่เป็นไปตามข้อตกลงเบื้องต้น เช่น การแจกแจงของประชากรอาจจะเบ้หรือโด่งไปจากการแจกแจงแบบปกติ ในกรณีเช่นนี้ผู้วิจัยอาจจะเลือกใช้สถิติทดสอบแบบนอนพาราเมตริก (Nonparametric) ซึ่งเป็นสถิติทดสอบที่ไม่มีข้อจำกัดเกี่ยวกับลักษณะการแจกแจงของประชากร สามารถคำนวณได้รวดเร็ว ทำความเข้าใจได้ง่าย และยังใช้ได้กับข้อมูลที่มีเสกลการวัดแบบนามเสกล (nominal Scale) หรือแบบจัดเรียงอันดับ (ordinal scale) (Daniel 1978 : 16, Bradley 1968 : 17-19, Conover 1980 : 3, Marascuilo & Mc Sweeny (1977 : 5)) ซึ่งก็เป็นสถิติทดสอบอีกรูปแบบหนึ่งที่ผู้วิจัยสามารถเลือกใช้ได้

สำหรับการทดสอบสมมติฐานเกี่ยวกับการเท่ากันระหว่างความแปรปรวนของประชากรสองกลุ่มที่เป็นอิสระต่อกัน ผู้วิจัยมักจะใช้การทดสอบเอฟ (F test) ซึ่งเป็นสถิติทดสอบแบบพาราเมตริก (Parametric) และมีข้อตกลงเบื้องต้นเกี่ยวกับลักษณะการแจกแจงของประชากรกล่าวคือ ประชากรทั้งสองกลุ่มจะต้องมีการแจกแจงแบบปกติ แต่โดยสภาพการณ์ทั่ว ๆ ไปแล้ว ลักษณะการแจกแจงของประชากรอาจจะไม่เป็นแบบปกติ ดังนั้นผู้วิจัยอาจจะเลือกใช้สถิติทดสอบแบบนอนพาราเมตริก ซึ่งใช้ทดสอบเกี่ยวกับความแปรปรวนของประชากร เช่นเดียวกับการทดสอบเอฟ เช่น การทดสอบมูด (Mood test) การทดสอบคลอทซ์ (Klotz test) การทดสอบทูกี้-ซีเกล (Tukey-Siegel test) หรือการทดสอบสแควร์แรงค์ (squared ranks test) ซึ่งอาจจะเป็นสถิติทดสอบที่เหมาะสมกว่า

การพิจารณาความเหมาะสมของสถิติทดสอบนั้น สิ่งที่ควรจะพิจารณาคือ อำนาจของการทดสอบ (Power of test) และความแกร่ง (Robustness) ดังคำกล่าวของ Box

และ Anderson (อ้างโดย Gibbon 1971 : 5) ว่า "เพื่อที่จะให้ความต้องการของผู้ทดสอบบรรลุตามเป้าหมายในการทดลองนั้น เกณฑ์ที่สำคัญในการเลือกใช้สถิติทดสอบคือ

1. สถิติทดสอบนั้นจะต้องมีความไว (Sensitive) ต่อการเปลี่ยนแปลงของปัจจัยที่ต้องการทดสอบ

2. สถิติทดสอบนั้นจะต้องไม่มีความไว (Insensitive) ต่อการเปลี่ยนแปลงของสิ่งอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ "ซึ่งคุณสมบัติ 2 ข้อนี้ คือ อำนาจของการทดสอบและความแกร่ง ตามลำดับ หรือพิจารณาได้จากคำกล่าวของ Neyman (1950 : 265 อ้างโดย Direk Srisukho 1974:38) ซึ่งกล่าวว่า "เมื่อต้องการที่จะเลือกใช้สถิติทดสอบ เราต้องเริ่มพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ก่อน แล้วจึงจะพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 2 โดยมีขั้นตอนดังนี้คือ ให้ความน่าจะเป็นที่ยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน α ที่กำหนดไว้ และเมื่อเป็นไปตามเงื่อนไขดังกล่าวแล้ว สิ่งที่จะใช้ในการพิจารณาเลือกสถิติทดสอบอีกก็คือเลือกสถิติทดสอบที่มีโอกาสน้อยที่สุดที่จะยอมรับสมมติฐาน H_o เมื่อสมมติฐาน H_o นั้นผิด ซึ่งหมายความว่าให้อำนาจของการทดสอบสูงที่สุด

สำหรับความแกร่งของการทดสอบเอฟ มีผู้ศึกษาเรื่องนี้หลายท่าน เช่น การศึกษาของ Bradley (1968 : 35) Brown และ Forsythe (1974 : 364-366) และ Miller (1968 : 567-587) ต่างก็มีข้อสรุปว่า การทดสอบเอฟไม่แกร่ง (Nonrobust) เมื่อลักษณะการแจกแจงของประชากรไม่เป็นแบบปกติ แต่อย่างไรก็ตาม Pearson และ Please (1975 : 223-241) ศึกษาความสัมพันธ์ระหว่างรูปแบบการแจกแจงของประชากรกับความแกร่งของการทดสอบเอฟ การทดสอบไคว์สแควร์ (Chi Square test) และการทดสอบที (t test) พบว่า ความแกร่งของการทดสอบเอฟ มีผลกระทบจากความโด่ง (Kurtosis) ของรูปแบบการแจกแจงของประชากร ซึ่งเบี่ยงเบนไปจากความโด่งของการแจกแจงแบบปกติ

การศึกษาอำนาจของการทดสอบเอฟหรือการทดสอบแบบนอนพาราเมตริกนั้น ได้มีผู้ศึกษาในรูปของค่าแอย์ยิมโทติครีเลทีฟเอฟฟิเชียนซี (Asymptotic Relative efficiency) หรือค่า ARE ของการทดสอบแบบนอนพาราเมตริกเมื่อเทียบกับการทดสอบเอฟหรือเทียบกับการทดสอบนอนพาราเมตริกด้วยกัน เช่น การศึกษาของ Mood (1954 : 514-522) Klotz (1962 : 498-512) Duran และMielke(1968 : 335-344) และ Hwang และ Klotz (1975 : 947-954) โดยเฉพาะการศึกษาของ Klotz และ Hwang และ Klotz พบว่าการ

ทดสอบคลอทซ์มีค่า ARE สูงที่สุด แต่อย่างไรก็ตามค่า ARE เป็นเพียงค่าลิมิตของอัตราส่วนของขนาดตัวอย่างระหว่างสถิติทดสอบ 2 วิธี เมื่อลิมิตของอำนาจของการทดสอบทั้ง 2 วิธีเท่ากัน (Gibbon 1971 : 280) เท่านั้น ดังนั้นค่า ARE อาจจะเป็นเพียงดัชนีที่จะบ่งชี้ถึงค่าอำนาจของการทดสอบ

จากแนวความคิดเกี่ยวกับการเลือกใช้สถิติทดสอบ สรุปได้ว่าการเลือกใช้สถิติทดสอบนั้นควรจะพิจารณาถึงอำนาจของการทดสอบและความแกร่ง และจากผลงานวิจัยที่ผ่านมา การศึกษาเกี่ยวกับการทดสอบเอฟ การทดสอบมูด การทดสอบคลอทซ์ การทดสอบทูกี้-ซีเกล และการทดสอบสแควร์แรงค์ โดยส่วนมากแล้วเป็นการศึกษาในลักษณะของค่า ARE เท่านั้น ดังนั้นอำนาจของการทดสอบต่าง ๆ ดังกล่าวข้างต้น เมื่อลักษณะการแจกแจงของประชากรมีรูปแบบต่าง ๆ กัน จึงเป็นสิ่งที่น่าศึกษาเพื่อจะได้มีข้อสรุปของอำนาจการทดสอบ และเป็นแนวทางในการเลือกใช้สถิติทดสอบที่เหมาะสมสำหรับการทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม ซึ่งการคำนวณค่าอำนาจของการทดสอบแบบนอนพาราเมตริก หรืออำนาจของการทดสอบแบบพาราเมตริก เมื่อลักษณะการแจกแจงไม่เป็นตามข้อตกลงเบื้องต้นกระทำได้ยากมาก นอกจากจะศึกษาโดยวิธีมอนติคาร์โล (Monte Carlo method) (Gibbons 1971:6 Miller 1968 : 574) ซึ่งเป็นวิธีที่ศึกษาในรูปของการจำลอง (Simulate) โดยใช้ตัวเลขสุ่ม สามารถระบุขนาดของกลุ่มตัวอย่าง ความแปรปรวน ค่าเฉลี่ย และลักษณะการแจกแจงของประชากรได้

<u>วัตถุประสงค์ของการวิจัย</u>

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบสถิติที่ใช้ทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม 5 วิธี จำแนกเป็นสถิติทดสอบแบบพาราเมตริก (Parametric) คือการทดสอบเอฟ (F test) และการทดสอบแบบนอนพาราเมตริก (Nonparametric) คือการทดสอบคลอทซ์(Klotz test) การทดสอบมูด (Mood test) การทดสอบสแควร์แรงค์(squared ranks test) และการทดสอบทูกี้-ซีเกล(Tukey-Siegel test) เพื่อหาข้อสรุปเกี่ยวกับอำนาจการทดสอบดังกล่าว ภายใต้ลักษณะการแจกแจงของประชากรเป็นแบบปกติ (Normal Distribution) แบบยูนิฟอร์ม (Uniform Distribution) แบบโลจิสติค (Logistic Distribution) และรูปแบบการแจกแจงที่เบ้ (Skewed Distribution)

สมมติฐานของการวิจัย

1. ลักษณะการแจกแจงของประชากรที่มีรูปแบบต่าง ๆ กัน จะมีผลให้ค่าอำนาจของการทดสอบแตกต่างกัน

2. ค่าแอซิมโทติครีเลทีฟเอฟฟิเชียนซี (Asymptotic Relative Efficiency) จะเป็นดัชนีที่จะบ่งชี้ถึงค่าอำนาจการทดสอบ

ข้อตกลงเบื้องต้น

การวิจัยครั้งนี้ถือว่า อำนาจของการทดสอบ และความแกร่ง (Robustness) เป็นดัชนีที่สำคัญที่นักวิจัยจะใช้เป็นเกณฑ์ในการเลือกสถิติทดสอบ

ขอบเขตของการวิจัย

1. ศึกษาอำนาจการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบเอฟ การทดสอบมัด การทดสอบคล้อทซ์ การทดสอบทูกี้-ซีเกล และการทดสอบสแควร์แรงค์ เมื่อประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติค สำหรับกรณีที่ประชากรมีการแจกแจงแบบเบ้ จะศึกษาเฉพาะการทดสอบเอฟและการทดสอบสแควร์แรงค์

2. ในกรณีการแจกแจงแบบเบ้ จะศึกษาเฉพาะการทดสอบสแควร์แรงค์ ซึ่งค่าสถิติคำนวณจากอันดับของค่าสมบูรณ์ของค่าสังเกต ซึ่งแตกต่างจากค่าเฉลี่ยของกลุ่มตัวอย่างเท่านั้น

3. ในการวิจัยครั้งนี้จะศึกษาเฉพาะกรณีที่ความเบ้ของประชากรอยู่ในช่วง [.25, 1.0] และความโด่งอยู่ในช่วง [2.8, 4.2]

4. กำหนดค่าเฉลี่ยของประชากรทั้งสองกลุ่มเท่ากัน

5. ศึกษาอำนาจของการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อค่าเฉลี่ยของประชากรเป็น 100 และอัตราส่วนความแปรปรวนของประชากร $(\frac{\sigma_y^2}{\sigma_x^2})$ มีค่าเป็น $\frac{1}{8}$, $\frac{1}{7}$, $\frac{1}{6}$, $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, 2, 3, 4, 5, 6, 7 และ 8

6. การศึกษาอำนาจการทดสอบของการทดสอบสแควร์แรงค์ เมื่อประชากรมีการแจกแจงแบบปกติ ยูนิฟอร์มและแบบโลจิสติค จะคำนวณค่าอำนาจการทดสอบเมื่อการทดสอบดังกล่าวสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ และในกรณีที่ประชากรมีการแจกแจง

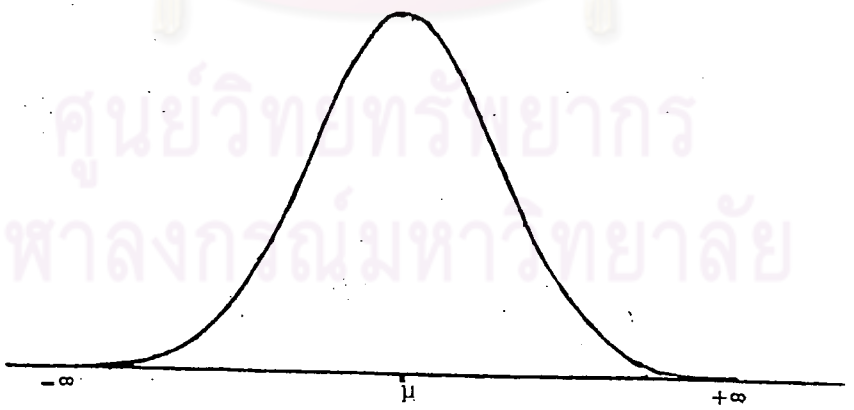
แบบเบ้ จะคำนวณค่าอำนาจการทดสอบก็ต่อเมื่อการทดสอบเอฟและการทดสอบสแควร์แรงค์สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เท่านั้น

7. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้แบ่งเป็น 2 กรณี คือ ขนาดของตัวอย่างทั้งสองกลุ่มเท่ากัน และไม่เท่ากัน กรณีที่ขนาดของกลุ่มตัวอย่างเท่ากันจะศึกษาเมื่อขนาดของตัวอย่างเป็น 10, 15 และ 20 ในกรณีที่ขนาดของตัวอย่างไม่เท่ากันจะศึกษาเฉพาะกรณีที่อัตราส่วนของขนาดตัวอย่างทั้งสองกลุ่มเป็น (5:10), (5:15), และ (5:20)

8. ในการวิจัยครั้งนี้จำลองการทดลองในเครื่องคอมพิวเตอร์ (Computer) โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation Technique) ซึ่งจะศึกษาเมื่อประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม แบบโลจิสติค และแบบเบ้ โดยมีค่าฟังก์ชั่นความน่าจะเป็น (Probability density function) ค่าคาดหวัง (Expected value) ความแปรปรวน (Variance) ความเบ้ (Skewness) และความโด่ง (Kurtosis) ของการแจกแจงแต่ละรูปแบบเป็นดังนี้

8.1 การแจกแจงแบบปกติ

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



$E(X) = \mu$

$V(X) = \sigma^2$

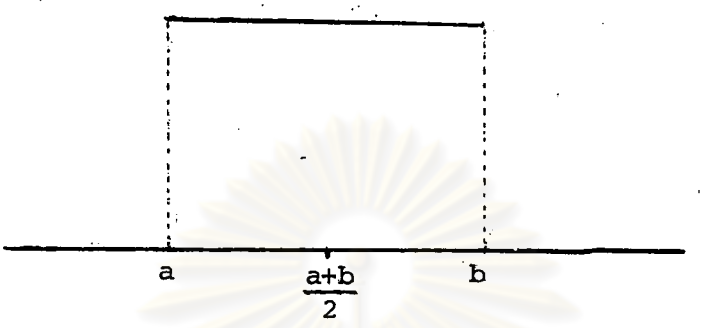
Skewness = 0.

Kurtosis = 3.0

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

8.2 การแจกแจงแบบยูนิฟอร์ม

$$f(x) = \frac{1}{b-a} \quad , a < x < b$$



$$E(X) = \left(\frac{a+b}{2}\right)$$

$$V(X) = \frac{(b-a)^2}{12}$$

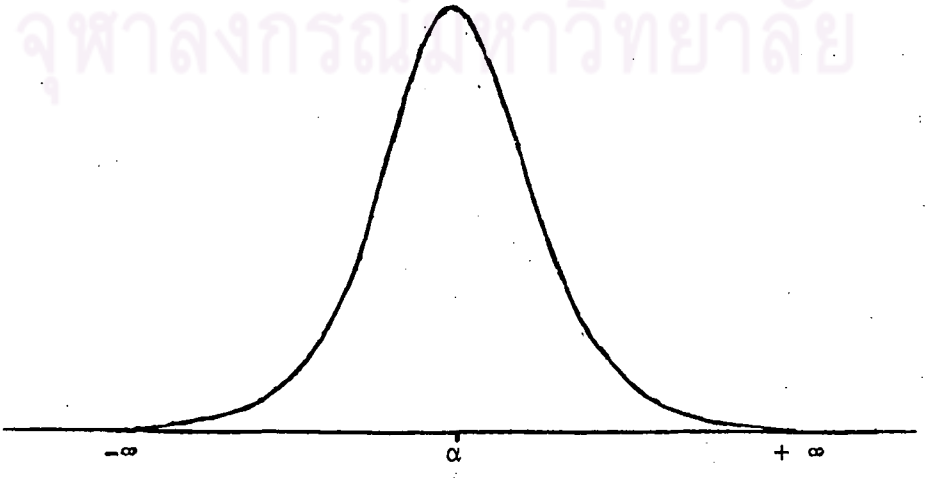
Skewness = 0.

Kurtosis = 1.8

8.3 การแจกแจงแบบโลจิสติค

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

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



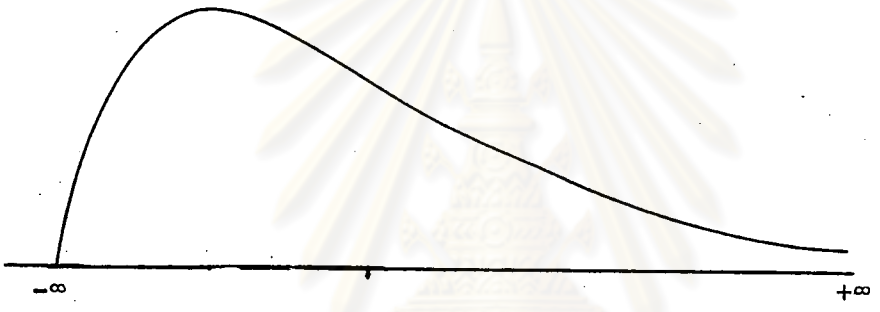
$E(X) = \alpha$

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

Skewness = 0.

Kurtosis = 4.2

8.4 การแจกแจงที่เบ้ (Skewed Distribution)



ลักษณะการแจกแจงที่เบ้ที่พิจารณาในวิทยานิพนธ์นี้เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบยูนิฟอร์ม ซึ่งเรียกว่า Generalized Lambda Distribution (GLD) ซึ่งมีฟังก์ชันการแปลงเป็นดังนี้

$$X = \lambda_1 + (P^{\lambda 3} - (1-P)^{\lambda 4})/\lambda_2 , \qquad 0 < P < 1$$

P มีการแจกแจงแบบยูนิฟอร์มในช่วง (0, 1)

λ_1 และ λ_2 เป็นพารามิเตอร์ที่กำหนดตำแหน่ง (Location parameter) และพารามิเตอร์ที่กำหนดเล่กล (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

9. ศึกษาอำนาจของการทดสอบเฉพาะการทดสอบแบบสองทาง (Two-sided Test) ณ ระดับนัยสำคัญ .01, .05 และ .10

10. การจำลองการทดลองจะกระทำซ้ำกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

คำจำกัดความ

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธ H_0 เมื่อสมมติฐาน H_0 ถูก

ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐาน H_o เมื่อสมมติฐาน H_o ผิด

อำนาจของการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐาน H_o เมื่อ สมมติฐาน H_o ผิด

Relative Efficiency ของการทดสอบP เมื่อเทียบกับการทดสอบ Q มีค่า $\frac{m}{n}$ หมายความว่า เมื่อตัวทดสอบ P และการทดสอบ Q ใช้ทดสอบสมมติฐานอย่างเดียวกัน $H_o: (\theta = \theta_o)$ ณ จุดที่อำนาจการทดสอบของการทดสอบยังอ่อนวิธีเท่ากัน และระดับนัยสำคัญเดียวกัน การทดสอบ P จะต้องใช้ขนาดตัวอย่างเป็น n ในขณะที่การทดสอบ Q จะต้องใช้ขนาดตัวอย่างเป็น m (Swat Pratomraj 1970:3)

Asymptotic Relative Efficiency (ARE) ของการทดสอบ P เมื่อเทียบกับการทดสอบ Q หมายถึง ค่าลิมิตของค่า Relative Efficiency ของการทดสอบ P เมื่อเทียบกับการทดสอบ Q เมื่อ n มีค่าเข้าใกล้อนันต์ (Infinity) หรืออาจจะพิจารณาจากสูตร

$$ARE\ (P\ ,\ Q) = \lim_{n\to\infty} \frac{e(Q_n)}{e(P_n)}$$

โดยที่ $e\ (P_n)$ คือค่า efficacy ของสถิติทดสอบ P เมื่อใช้ทดสอบสมมติฐาน

$H_o : \theta = \theta_o$ และ $e\ (P_n)$ มีค่า $= \dfrac{[d(E\ (P_n))/\ d\theta]^2}{\sigma^2\ (P_n)\ |_{\theta=\theta_o}}$ (Gibbons 1971:281)

ความแกร่ง (Robustness) ของการทดสอบ หมายถึง คุณสมบัติของการทดสอบที่ไม่แสดงถึง ความไว (sensitive) ต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น ซึ่งสิ่งที่ใช้พิจารณาความแกร่งของการทดสอบ คือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

การทดสอบสัดส่วนแรงค้า($\bar{X}$) หมายถึงการทดสอบสัดส่วนแรงค้า ซึ่งคำนวณค่าสถิติจากอันดับของค่าสัมบูรณ์ (Absolute Value) ของค่าสังเกตซึ่งแตกต่างจากค่าเฉลี่ยของกลุ่มตัวอย่าง

การทดสอบสัดส่วนแรงค้า (μ) หมายถึงการทดสอบสัดส่วนแรงค้าซึ่งคำนวรค่าสถิติจากอันดับของค่าสัมบูรณ์ (Absolute Value) ของค่าสังเกตซึ่งแตกต่างจากค่าเฉลี่ยของประชากร

<u>ประโยชน์ของการวิจัย</u>

เพื่อให้ผู้ใช้สถิติมีข้อสรุปและหลักฐานเกี่ยวกับสถิติที่ใช้ทดสอบความแปรปรวนระหว่างประชากรสองกลุ่ม และสามารถเลือกใช้สถิติทดสอบความแปรปรวนได้อย่างถูกต้องและเหมาะสมกับลักษณะการแจกแจงของประชากร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย