บทที่ 2

สถิติทดสอบและผลงานวิจัยที่เกี่ยวข้อง

ในกรณีที่ผู้วิจัยต้องการทดสอบสมมติฐานเกี่ยวกับความแปรปรวนระหว่างประชากรสองกลุ่มนั้น สถิติทดสอบแบบพาราเมตริก (Parametric) ที่ใช้ทดสอบสมมติฐานนี้คือการทดสอบเอฟ (F test) สำหรับสถิติทดสอบแบบนอนพาราเมตริก (Nonparametric) มีอยู่หลายวิธีเช่น การทดสอบคล้อทซ์ (Klotz test) การทดสอบมู้ด (Mood test) การทดสอบทูกี้-ซีเกล (Tukey-Siegel test) และการทดสอบสแควร์แรงค์ (Squared ranks test) เป็นต้น ซึ่งสถิติทดสอบแต่ละวิธีนั้นอาจจะมีข้อตกลงเบื้องต้นแตกต่างกันบ้าง

ซึ่งในบทนี้จะกล่าวถึงรายละเอียดของการทดสอบแต่ละวิธี สำหรับกรณีของการทดสอบเอฟ จะกล่าวถึงการแจกแจงเอฟ (F distribution) ด้วย ทั้งนี้เพราะว่าตัวสถิติของการทดสอบเอฟมีการแจกแจงแบบเอฟ ส่วนตอนท้ายของบทนี้จะนำเสนอผลงานวิจัยที่เกี่ยวข้อง ซึ่งรายละเอียดต่าง ๆ เป็นดังนี้

<u>การทดสอบเอฟ</u>

ในการทดสอบสมมติฐานการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่มนั้น การทดสอบเอฟเป็นการทดสอบแบบพาราเมตริกที่ใช้ทดสอบสมมติฐานนี้ ตัวสถิติของการทดสอบเอฟคือ $\frac{S_1^2}{S_2^2}$ เมื่อ S_1^2 และ S_2^2 เป็นความแปรปรวนของตัวอย่างจากประชากรกลุ่มที่ 1 และจากประชากรกลุ่มที่ 2 ตามลำดับ ซึ่งตัวสถิตินี้มีการแจกแจงแบบเอฟ เนื่องจากตัวสถิติของการทดสอบเอฟมีการแจกแจงแบบเอฟ ดังนั้นในส่วนนี้จะกล่าวถึงการแจกแจงเอฟก่อนแล้วจึงจะกล่าวถึงรายละเอียดของการทดสอบเอฟ

1. การแจกแจงเอฟ (F-distribution)

Sir Ronald A. Fisher เป็นบุคคลแรกที่เสนอการแจกแจงเอฟ ในปี ค.ศ.1924 เดิมใช้ชื่อว่าการแจกแจง ซี (Z-distribution) โดยมีสูตร $Z = \frac{1}{2} \ln \frac{S_1^2}{S_2^2}$ ซึ่งค่าลอกการิทึม (logarithm) ทำให้ยุ่งยากในการคำนวณ W.Snedecor จึงปรับปรุงใหม่ โดยใช้สูตร $F = \frac{S_1^2}{S_2^2}$ ค่า S_1^2 และ S_2^2 คือค่าความแปรปรวนของกลุ่มตัวอย่างจากประชากรที่ 1 และกลุ่มตัวอย่างจากประชากรที่ 2 ตามลำดับ Snedecor ได้ตั้งชื่อการ

แจกแจงใหม่นี้ว่าการแจกแจงเอฟ (F distribution) เพื่อเป็นเกียรติแก่ Ronald A. Fisher แต่ในบางครั้งจะเรียกว่าการแจกแจงเอฟของสเนคดีเคอร์ (Snedecor's F distribution) เพื่อเป็นเกียรติแก่ W. Snedecor

1.1 ฟังก์ชั่นความน่าจะเป็นของการแจกแจงเอฟ

ตัวแปรสุ่มที่มีการแจกแจงแบบเอฟจะได้จากอัตราส่วนของตัวแปรสุ่มแบบไคสแควร์ (Chi Square) 2 ตัวที่เป็นอิสระต่อกัน ซึ่งแต่ละตัวหารด้วยจำนวนชั้นแห่งความเป็นอิสระ (degree of fredom) ดังนั้นฟังก์ชั่นความน่าจะเป็นของการแจกแจงเอฟแสดงได้ดังนี้

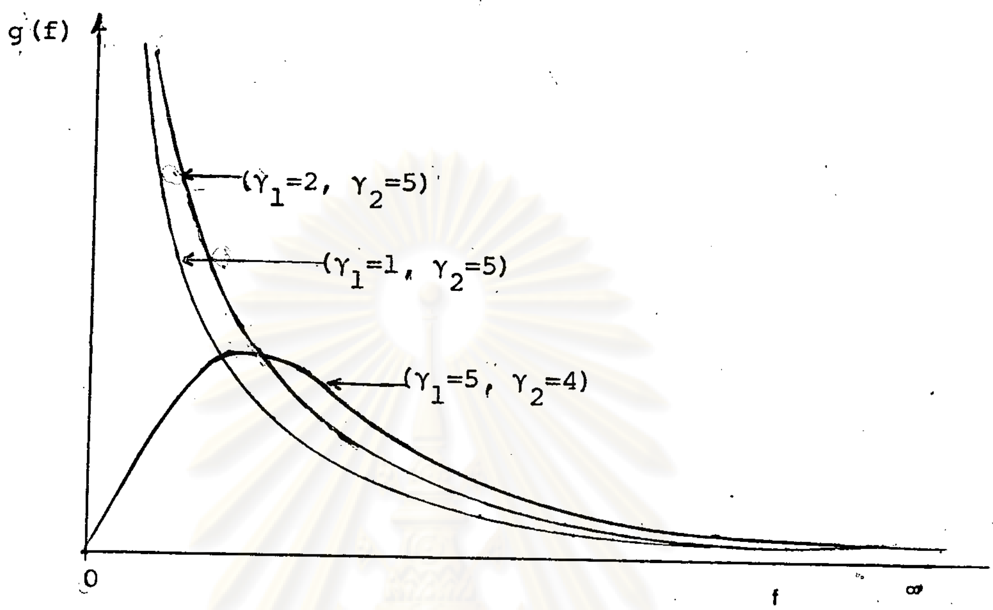
$$g(f) = \frac{\Gamma\left(\frac{1}{2}(\gamma_1+\gamma_2)\right)}{\Gamma(\gamma_{1/2})\,\Gamma(\gamma_{2/2})} \left(\frac{\gamma_1}{\gamma_2}\right)^{\frac{\gamma_1}{2}} f^{\frac{\gamma_1}{2}-1} \left(1+\frac{\gamma_1}{\gamma_2}f\right)^{-\frac{1}{2}(\gamma_1+\gamma_2)}, \quad f > 0$$

ค่า γ_1 และ γ_2 เป็นจำนวนชั้นแห่งความเป็นอิสระของตัวแปรสุ่มแบบไคสแควร์ที่เป็นเศษและเป็นส่วนตามลำดับ

จากฟังก์ชั่นความน่าจะเป็นของการแจกแจงเอฟ จะเห็นว่าการแจกแจงเอฟมีค่าพารามิเตอร์ 2 ตัว คือ γ_1 และ γ_2 จะเป็นค่าที่กำหนดรูปแบบ (Shape) ของการแจกแจง ซึ่งลักษณะของการแจกแจงเอฟไม่สมมาตร (Asymmetric) โดยมีลักษณะเบ้ไปทางขวา (Positive Skewness) เช่นเมื่อ $\gamma_1 = 1, 2$ ลักษณะการแจกแจงจะเป็นรูปกลับของอักษร J และถ้า γ_1 มากกว่า 3 การแจกแจงจะเบ้ไปทางขวา ซึ่งแสดงได้ดังรูป 2.1 ในกรณีที่ γ_1 และ γ_2 มีค่าใกล้อนันต์ (infinity) การแจกแจงเอฟจะมีลักษณะสมมาตร

รูป 2.1 แสดงรูปแบบของการแจกแจงเอฟ เมื่อ

($\gamma_1 = 1, \gamma_2 = 5$), ($\gamma_1 = 2, \gamma_2 = 5$) และ ($\gamma_1 = 5, \gamma_2 = 4$)



การแจกแจงเอฟมีค่าคาดหวัง (Expected Value) = γ_2/γ_2-2 และความแปรปรวน (Variance) มีค่าเป็น $\frac{2\gamma_2^2(\gamma_1+\gamma_2-2)}{\gamma_1(\gamma_2-2)^2(\gamma_2-4)}$ ในกรณีที่ $\gamma_2 \leq 4$ การแจกแจงเอฟจะหาค่าความแปรปรวนไม่ได้

1.2 คุณสมบัติของการแจกแจงเอฟ

ถ้า $F(\gamma_1, \gamma_2)$ มีการแจกแจงเอฟ ซึ่งมีขั้นแห่งความเป็นอิสระ γ_1 และ γ_2 แล้ว $F(\gamma_2, \gamma_1) = \frac{1}{F(\gamma_1, \gamma_2)}$ จะมีการแจกแจงเอฟด้วยขั้นแห่งความเป็นอิสระ γ_2 และ γ_1

1.3 ค่าวิกฤติของการแจกแจงเอฟ

ในปี ค.ศ. 1946 Snedecor ได้สร้างตารางเอฟเพื่อกำหนดค่าวิกฤต (Critical Value) ที่ระดับนัยสำคัญ .05 และ Fisher และ Yates ได้สร้างตารางเพื่อกำหนดค่าวิกฤตที่ระดับนัยสำคัญ .01 ในปี ค.ศ. 1953 Menington และ Thompson สร้างตารางเอฟที่ระดับนัยสำคัญ .25, .10, .025 และ .005 ต่อมา Novton ได้สร้างตารางเพิ่มเติมที่ระดับนัยสำคัญ .20 และ Calcord และ Deming ได้คำนวณค่าวิกฤตที่ระดับนัยสำคัญ .001

ดังนั้นในปัจจุบันนี้ตารางกำหนดค่าวิกฤติของการทดสอบเอฟจึงมีที่ระดับนัยสำคัญต่าง ๆ กันดังนี้คือ .001, .005, .01, .025, .05, .10, .20 และ .25

1.4 ความสำคัญของการแจกแจงเอฟ

การแจกแจงเอฟมีความสำคัญในงานวิจัยต่าง ๆ เป็นอันมาก เช่น ใช้ในการประมาณช่วง (Interval estimate) และการทดสอบสมมติฐานทางสถิติ ตัวอย่างเช่น ใช้ทดสอบสมมติฐานเกี่ยวกับการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากรหลาย ๆ กลุ่ม เป็นต้น

2. การทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม

สมมติฐานที่ทดสอบ

A : การทดสอบแบบสองทาง (Two-Sided Test)

H_o: ประชากรทั้งสองกลุ่มมีความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

H_A: ประชากรทั้งสองกลุ่มมีความแปรปรวนไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$)

B : การทดสอบแบบทางเดียว (One-Sided Test)

H_o: ประชากรทั้งสองกลุ่มมีความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

H_A: ประชากรที่ 1 มีความแปรปรวนน้อยกว่าประชากรที่ 2 ($\sigma_1^2 < \sigma_2^2$)

C : การทดสอบแบบทางเดียว (One-Sided Test)

H_o: ประชากรทั้งสองกลุ่มมีความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

H_A: ประชากรที่ 1 มีความแปรปรวนมากกว่าประชากรที่ 2 ($\sigma_1^2 > \sigma_2^2$)

2.1 ข้อตกลงเบื้องต้น

2.1.1 ตัวอย่างสุ่มทั้งสองกลุ่มเป็นอิสระต่อกัน

2.1.2 ประชากรที่ศึกษาจะต้องมีการแจกแจงแบบปกติ

2.2 สถิติทดสอบ

ตัวสถิติที่ใช้ในการทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม คือ ตัวสถิติเอฟ (F-Statistic) ซึ่งมีสูตรในการคำนวณดังนี้

$$F = \frac{S_1^2}{S_2^2}$$

ตัวสถิติเอฟ มีการแจกแจงแบบเอฟ ซึ่งมีจำนวนชั้นแห่งความเป็นอิสระ n_1-1 และ n_2-1 (Hogg and Craig 1978:266) โดยที่ s_1^2 และ s_2^2 เป็นค่าความแปรปรวน

ของกลุ่มตัวอย่างจากประชากรที่ 1 และกลุ่มตัวอย่างจากประชากรที่ 2 ตามลำดับ และมีสูตรในการคำนวณดังนี้

$$s_1^2 = \sum_{i=1}^{n_1} \frac{(x_i-\bar{x})^2}{(n_1-1)}$$

$$s_2^2 = \sum_{i=1}^{n_2} \frac{(y_i-\bar{y})^2}{n_2-1}$$

2.3 เกณฑ์การตัดสินใจ

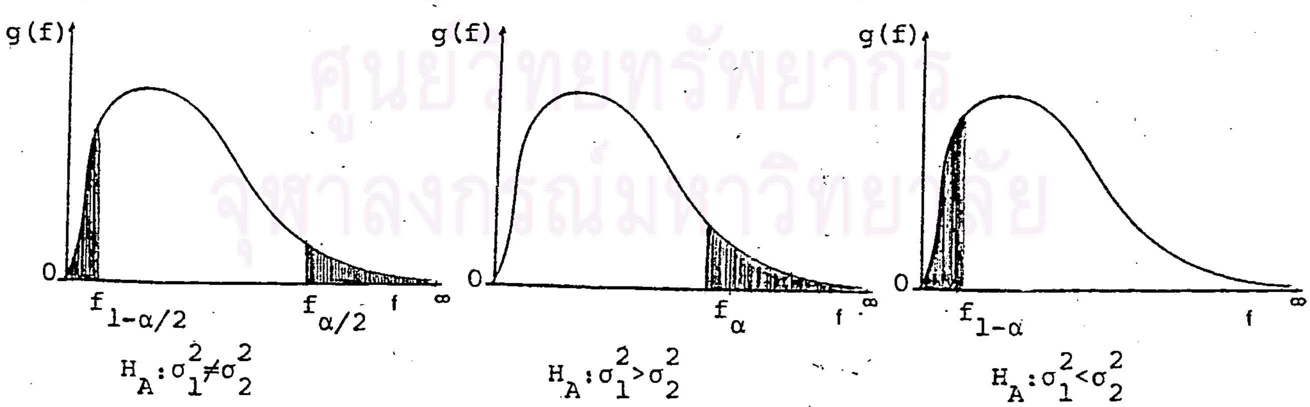
สมมติฐานแบบ A จะปฏิเสธ H_o เมื่อ $F \geq f_{\alpha/2}\ (\gamma_1, \gamma_2)$ หรือ $F \leq f_{1-\alpha/2}\ (\gamma_1, \gamma_2)$

สมมติฐานแบบ B จะปฏิเสธ H_o เมื่อ $F \leq f_{1-\alpha}\ (\gamma_1, \gamma_2)$

สมมติฐานแบบ C จะปฏิเสธ H_o เมื่อ $F \geq f_{\alpha}\ (\gamma_1, \gamma_2)$

ค่าวิกฤติสำหรับสมมติฐานแย้ง (Alternative Hypothesis) แบบต่าง ๆ แสดงได้ดังรูป 2.2

รูป 2.2 แสดงตำแหน่งของค่าวิกฤติของการทดสอบเอฟ สำหรับการทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม เมื่อตั้ง H_A แบบต่าง ๆ



▥ - คือขอบเขตของการปฏิเสธ H_o

f - เป็นค่าวิกฤติที่จะปฏิเสธ H_o

สถิติทดสอบนอนพาราเมตริกสำหรับทดสอบความแปรปรวนระหว่างประชากรสองกลุ่ม

ในการทดสอบสมมติฐานเกี่ยวกับการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม สถิติทดสอบนอนพาราเมตริกที่ใช้ทดสอบความแปรปรวนเช่นเดียวกับการทดสอบเอฟมีอยู่หลายวิธี แต่โดยทั่ว ๆ ไปแล้วตัวสถิติที่นำมาทดสอบจะอยู่ในรูปของสมการ $Y = \sum_{i=1}^{N} a_i z_i$ โดยที่ a_i จะเป็นค่าถ่วงน้ำหนัก (Weight) N เป็นผลรวมของขนาดตัวอย่างทั้ง 2 กลุ่ม และ z_i เป็นตัวอินดิเคเตอร์ (Indicator) ซึ่งแต่ละสถิติทดสอบจะมีค่าถ่วงน้ำหนักที่แตกต่างกัน เช่น ค่าถ่วงน้ำหนักเป็นกำลังสองของอันดับ (rank) กำลังสองของอันดับที่แตกต่างจากค่าเฉลี่ย หรือกำลังสองของค่าอินเวิสนอร์มอลสกอร์ (Inverse Normal score) เป็นต้น

สถิติทดสอบนอนพาราเมตริกที่ใช้ทดสอบความแปรปรวนระหว่างประชากรสองกลุ่ม ซึ่งใช้ในการวิจัยครั้งนี้ คือ การทดสอบคล้อทซ์ (Klotz test) การทดสอบมู้ด (Mood test) การทดสอบทูกี้-ซีเกล (Tukey-Siegel test) และการทดสอบสแควร์แรงค์ (squared ranks test) ซึ่งจำแนกตามข้อตกลงเบื้องต้นเป็น 2 กลุ่มคือ กลุ่มแรกได้แก่การทดสอบคล้อทซ์ การทดสอบมู้ด และการทดสอบทูกี้-ซีเกล และกลุ่มที่สองได้แก่ การทดสอบสแควร์แรงค์ ซึ่งข้อตกลงเบื้องต้นของสถิติทดสอบแต่ละกลุ่มเป็นดังนี้

ข้อตกลงเบื้องต้นของการทดสอบคล้อทซ์ การทดสอบมู้ด และการทดสอบทูกี้-ซีเกล (Daniel 1978:93 , Marascuilo and Sweeney 1977:289, 291)

1. สุ่มตัวอย่าง $x_1, x_2, x_3, \ldots, x_m$ จากประชากร X และ $y_1, y_2, \ldots, y_n$ จากประชากร y โดยที่ $m \leqslant n$

2. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน

3. ประชากรที่สนใจศึกษามีการแจกแจงแบบต่อเนื่อง

4. เสกลที่ใช้วัดข้อมูลอย่างน้อยเป็นเสกลแบบจัดเรียงอันดับ (Ordinal Scale)

5. ประชากรที่สนใจศึกษามีรูปแบบการแจกแจงแบบเดียวกัน (Identical Distribution) มีค่ามีเดียนเท่ากัน

ข้อตกลงเบื้องต้นของการทดสอบสแควร์แรงค์ (Conover 1980:240)

1. ข้อมูลประกอบด้วยกลุ่มตัวอย่างสองกลุ่ม คือ $x_1, x_2, \ldots, x_m$ จากประชากร X และ $y_1, y_2, \ldots, y_n$ จากประชากร y

2. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน

3. เสกลที่ใช้วัดข้อมูลอย่างน้อยเป็นเสกลแบบช่วง (Interval Scale)

สำหรับรายละเอียดของการทดสอบแต่ละวิธีเป็นดังนี้

1. การทดสอบมู้ด (Mood Test)

Mood เป็นผู้เสนอการทดสอบนี้ขึ้นในปี ค.ศ.1954 โดยกำหนดค่าสถิติ ซึ่งมีค่าถ่วงน้ำหนักเป็นกำลังสองของอันดับที่เบี่ยงเบนไปจากค่าเฉลี่ยของอันดับทั้งหมดแล้วนำค่าถ่วงน้ำหนักของกลุ่มที่มีขนาดตัวอย่างเล็กกว่ามารวมกัน ดังนั้นสถิติทดสอบของ Mood แสดงได้ดังนี้

1.1 สถิติทดสอบ

$$M = \sum_{i=1}^{N} \left(r_i - \frac{N+1}{2}\right)^2 Z_i$$

โดย Z_i = 1 เมื่อตัวอย่างที่ i มาจากกลุ่มของตัวอย่างที่มีขนาดเล็กกว่า

0 นอกเหนือจากนี้

และ r_i คือ อันดับของค่าสังเกตที่ i จากการเรียงอันดับร่วมกันระหว่างกลุ่มตัวอย่างทั้งสองกลุ่ม

ในกรณีที่ M มีค่ามาก แสดงว่าความแปรปรวนของประชากรที่สุ่มตัวอย่างมาจำนวนน้อยกว่า มีค่าสูงกว่าความแปรปรวนของประชากรอีกกลุ่มหนึ่ง

1.2 ค่าวิกฤติของการทดสอบมู้ด

ตารางกำหนดค่าวิกฤตของการทดสอบ Mood นั้น Laubocher, Steffens และ Deloge ได้สร้างขึ้นใน ค.ศ.1967 สำหรับกรณีที่ N ≤ 20 ณ ระดับนัยสำคัญ .005, .01, .025, .05, .10 สำหรับการทดสอบทางเดียว และที่ระดับนัยสำคัญ .01, .02, .05, .10 และ .20 สำหรับการทดสอบแบบสองทาง

ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ (N ≥ 30) Laubsch แนะนำให้ประมาณด้วยการแจกแจงแบบปกติ ดังนี้

$$Z = \frac{M - E(M)}{\sqrt{V(M)}}$$

$$E(M) = \frac{m}{12}(N-1)(N+1) \qquad , N = m+n$$

$$V(M) = \frac{mn}{180}(N+1)(N-2)(N+2)$$

แต่เมื่อขนาดของกลุ่มตัวอย่างน้อยกว่า 30 การประมาณด้วยการแจกแจงแบบปกติ ควรจะปรับค่า Z ด้วย $\frac{1}{2}$ V(M)

ในกรณีที่มีค่าซ้ำ (Tied observation) Daniel (1978:96) กล่าวว่า เมื่อ m มีค่าน้อย ค่าซ้ำจะมีผลต่อค่า M มาก แต่ถ้า m และ N มีค่ามาก ๆ และจำนวนซ้ำที่เกิดขึ้นมีน้อย การคำนวณไม่ต้องคำนึงถึงการเกิดค่าซ้ำกัน แต่อย่างไรก็ตาม Mielke (1968: 312-314) ได้กำหนดค่าความแปรปรวนของ M ในกรณีที่เกิดค่าซ้ำ สำหรับการคำนวณเมื่อประมาณด้วยการแจกแจงแบบปกติ ดังนี้

$$V(M) = \frac{mn}{180}(N+1)(N^2-4) - \frac{mn}{180}(N(N-1))\sum_{j=1}^{k} t_j(t_j^2-1)\left[(t_j^2-4+15(N-s_j-s_{j-1})^2\right]$$

โดยที่ t_j = จำนวนค่าซ้ำในกลุ่มที่ j

และ $S_j = \sum_{h=1}^{j} t_h$

2. การทดสอบทูกี้-ซีเกล (Tukey-Siegel Test)

Tukey และ Siegel เป็นผู้เสนอการทดสอบนี้ในปี ค.ศ.1960 โดยอาศัยแนวคิดพื้นฐานจากการทดสอบวิลค๊อกซัน (Wilcoxon Rank Sum Test) แต่ลักษณะการกำหนดค่าถ่วงน้ำหนักแตกต่างกัน โดยที่การทดสอบทูกี้-ซีเกล จะกำหนดค่าถ่วงน้ำหนัก 1 ให้กับค่าต่ำสุด 2 ให้ค่ากับสูงสุด 3 ให้กับค่ามากเป็นอันดับสอง 4 ให้กับค่าน้อยเป็นอันดับที่สอง 5 ให้กับค่าน้อยเป็นอันดับ 3 6 ให้กับค่ามากเป็นอันดับ 3 7 ให้กับค่ามากเป็นอันดับ 4 สลับค่าการกำหนดน้ำหนักในลักษณะนี้จนกระทั่งครบทุกค่าสังเกต (Marascuilo Mc Sweeney 1977:289 และ Lehmann 1972:29) ดังรูป 2.3

รูป 2.3 แสดงการกำหนดค่าถ่วงน้ำหนักของการทดสอบทูกี้-ซีเกล

อันดับ	1	2	3	4	5	6	7	....	N-7	N-6	N-5	N-4	N-3	N-2	N-1	N
ค่าถ่วงน้ำหนัก	1	4	5	8	9	12	13		15	14	11	10	7	6	3	2

2.1 สถิติทดสอบ

$$S = \sum_{i=1}^{N} a_i z_i$$

โดยที่ a_i = เป็นค่าถ่วงน้ำหนักของค่าสังเกตที่ i

และ z_i = 1 เมื่อค่าสังเกตที่ i มาจากกลุ่มตัวขนาดที่มีขนาดน้อยกว่า

0 นอกเหนือจากนี้

2.2 ค่าวิกฤตของการทดสอบทูกี้-ซีเกล

ตารางกำหนดค่าวิกฤตสำหรับการทดสอบทูกี้-ซีเกลนั้น สามารถที่จะใช้ตารางของการทดสอบวิลคอกซัน แทนได้ เพราะว่าลักษณะการแจกแจงของ S และ W ของการทดสอบวิกค๊อกซันเหมือนกัน หรืออาจจะใช้ตารางที่ Tukey และ Siegel กำหนดไว้ ซึ่งตารางที่ Tukey และ Siegel กำหนดไว้นั้นใช้สำหรับการทดสอบทางเดียวที่ระดับนัยสำคัญ .001 .005 .01 .025 .05 และ .10 และการทดสอบสองทางที่ระดับนัยสำคัญ .002 .01 .02 .05 .10 และ .20 โดยที่ขนาดของกลุ่มตัวอย่างที่กำหนดไว้ในตาราง มีค่า m น้อยกว่าหรือเท่ากับ 20 และ n น้อยกว่าหรือเท่ากับ 20 ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ จะประมาณการแจกแจง S ด้วยการแจกแจงแบบปกติดังนี้

$$Z = \frac{S - E(S) \pm 1/_2}{\sqrt{V(S)}}$$

$$E(S) = \frac{m}{2}(N+1)$$

$$V(S) = \frac{mn}{12}(N+1)$$

ในการประมาณด้วยการแจกแจงแบบปกติ กรณีที่ขนาดของกลุ่มตัวอย่างทั้งสองกลุ่มต่างกันมาก เช่น (m, n) = (2, 20) หรือ (3, 20) Tukey และ Siegel ได้แนะนำให้ปรับค่า Z โดยวิธีของ Cornish-Fisher ซึ่งใช้ Edgeworth expansion ดังสูตร

$$Z_{cf} = Z + \frac{r_2}{24}(Z^3 - 3Z)$$

$$\text{โดยที่ } r_2 = \frac{6}{5} \, \frac{n_2^2 (n_2+1) - n_1^2 (n_1 + 1)}{n_2 n_1 (n_2 + n_1 + 1)(n_2 - n_1)}$$

อีกวิธีหนึ่งที่ Tukey และ Siegel แนะนำคือให้ปรับค่า z ดังสูตร

$$z_e = z + \left(\frac{1}{10\,n_1} - \frac{1}{10\,n_2}\right)(z^3 - 3z)$$

ในกรณีที่มีค่าสังเกตซ้ำ Tukey และ Siegel แนะนำให้ปรับสูตรความแปรปรวนของ S ด้วยค่า $\frac{S_1 - S_2}{N}$ ดังนั้นความแปรปรวนของ S ที่ปรับแล้วจะเป็นดังนี้

$$V(S) = \frac{1}{12}(N^2 - 1) - \frac{S_1 - S_2}{N}$$

โดยที่ S_1 = ผลรวมกำลังสองของค่า a_i ที่ไม่ใช่ค่าเฉลี่ย

S_2 = ผลรวมกำลังสองของค่า a_i ที่เกิดจากการเฉลี่ยของอันดับ (Mid rank)

3. การทดสอบคล้อทซ์ (Klotz Test)

Klotz เป็นผู้เสนอสถิติทดสอบนี้ในปี ค.ศ.1962 โดยกำหนดค่าถ่วงน้ำหนักเป็นกำลังสองของค่าอินเวิสนอร์มอลสกอร์ (Inverse Normal Score) ดังนั้นสถิติทดสอบของคล้อทซ์เป็นดังนี้

3.1 สถิติทดสอบ

$$K = \sum_{i=1}^{N} a_i z_i$$

โดยที่ $a_i = \left[\phi^{-1}\left(\frac{r_i}{N+1}\right)\right]^2$

z_i = 1 ถ้าค่าสังเกตที่ i อยู่ในกลุ่มตัวอย่างที่มีขนาดเล็กกว่า

= 0 นอกเหนือจากนี้

และ r_i คือ อันดับของค่าสังเกตที่ i ในการเรียงอันดับของกลุ่มตัวอย่างทั้งสองกลุ่มร่วมกัน

ϕ (z) คือ ฟังก์ชั่นการแจกแจงสะสม (Cumulative Distribution Function) ของการแจกแจงแบบปกติมาตรฐาน

ค่าอินเวิสนอร์มอลสกอร์อาจจะหาได้จากตารางแสดงค่าฟังก์ชั่นการแจกแจงสะสม (F(z)) ของการแจกแจงแบบปกติมาตรฐาน โดยพิจารณาจากค่า $\frac{r_i}{N+1}$ ซึ่ง

เป็นค่าแสดงความน่าจะเป็นที่ Z จะน้อยกว่า z โดยค่า z จะเป็นค่าอินเวิสนอร์มอลสกอร์ หรืออาจจะใช้ค่าซึ่ง Van Der Waerden กำหนดไว้สำหรับกรณีที่ N น้อยกว่าหรือเท่ากับ 40 แต่อย่างไรก็ตาม Klotz ก็ได้กำหนดค่า $\left[\phi^{-1}\left(\frac{r_i}{N+1}\right)\right]^2$ ไว้ เมื่อ N น้อยกว่าหรือเท่ากับ 20

3.2 ค่าวิกฤติของการทดสอบคลอัทซ์

ตารางกำหนดค่าวิกฤตสำหรับการทดสอบคลอัทซ์กำหนดไว้ในตาราง A-18 ในหนังสือ Nonparametric And Distribution-Free for Social Science ของ Marascuilo และ Sweeney, Mc (1977) หรือตารางที่ Klotz เสนอไว้ ณ ระดับนัยสำคัญ .005 .001 .025 .05 และ .10 สำหรับการทดสอบแบบทางเดียว และ .01 .02 .05 .10 และ .20 สำหรับการทดสอบสองทาง โดยขนาดของกลุ่มตัวอย่างที่กำหนดไว้ในตารางมีค่า N น้อยกว่าหรือเท่ากับ 20 และ m น้อยกว่าหรือเท่ากับ 10 ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ ให้ประมาณด้วยการแจกแจงแบบปกติ ดังนี้

$$Z = \frac{K - E(K)}{\sqrt{V(K)}}$$

โดยที่ $$E(K) = \frac{m}{N}\sum_{i=1}^{N}\left[\phi^{-1}\left(\frac{i}{N+1}\right)\right]^2$$

$$V(K) = \frac{mn}{N(N-1)}\sum_{i=1}^{N}\left[\phi^{-1}\left(\frac{i}{N+1}\right)\right]^4 - \frac{n}{m(N-1)}\left[E(K)\right]^2$$

4. การทดสอบสแควร์แรงค์ (Squared Ranks Test)

Taha (1964) เป็นผู้เสนอการทดสอบนี้ โดยแนะนำให้ใช้ผลบวกกำลังสองของอันดับ ซึ่งเกิดจากการเรียงลำดับค่าสังเกตของกลุ่มตัวอย่างทั้งสองกลุ่มร่วมกันแทนที่จะใช้ผลรวมของอันดับเหมือนการทดสอบแบบอื่น ๆ ดังนั้นสถิติทดสอบของ Taha กำหนดได้ดังนี้

$T = \sum_{i=1}^{N} R_i^2 Z_i$ ค่าประสิทธิภาพของการทดสอบสแควร์แรงค์ Jam (1967) Duman และ Mielke (1968) และ Mielke (1972) ได้ทำการศึกษาซึ่งได้ข้อสรุปว่า เมื่อ f(x) อยู่ในกลุ่มของแกมมา (gamma family) แล้ว การทดสอบสแควร์แรงค์จะมีค่า Asymtotic Relative Efficiency (ARE) สูง โดยเฉพาะเมื่อประชากรมีการแจกแจงแบบ Truncated t การทดสอบสแควร์แรงค์จะมีค่า ARE สูงที่สุด

อย่างไรก็ตามการทดสอบสแควร์แรงค์ ได้มีการเปลี่ยนแปลงโดย Duran (1971) โดยกำหนดค่า r_i เป็นอันดับของค่า $|X_i|$ แทนที่จะเป็นอันดับของ x_i ซึ่ง Duran ยังแสดงได้ว่าค่า ARE ของการทดสอบสแควร์แรงค์มีค่าเท่ากับการทดสอบมู้ด เมื่อประชากรมีการแจกแจงแบบสมมาตร ต่อมาการทดสอบสแควร์แรงค์ ได้มีการเปลี่ยนแปลงโดย Filgner และ Killeen (1976) และ Conover และ Iman (1978) ซึ่งในการวิจัยครั้งนี้ศึกษาเฉพาะกรณีของ Conover และ Iman

4.1 การทดสอบสแควร์แรงค์ ซึ่งเสนอโดย Conover และ Iman

Conover และ Iman เสนอการทดสอบสแควร์แรงค์ โดยอาศัยแนวคิดจากการทดสอบค่าเฉลี่ยของประชากร ซึ่งมีสมมติฐานหลัก คือ $H_o : E(X) = E(Y)$ และตัวสถิติที่ใช้ทดสอบเป็นฟังก์ชั่นเมื่อเส้นตรงของอันดับ ซึ่งเกิดจากการเรียงลำดับร่วมกันของกลุ่มตัวอย่างทั้งสองกลุ่ม พิจารณาความแปรปรวนของประชากรซึ่งเป็นค่าคาดหวังของ $(X-\mu)^2$ โดย μ เป็นค่าเฉลี่ยของประชากร ดังนั้นในการทดสอบสมมติฐาน $H_o:E(X-\mu_x)^2=E(y-\mu_y)^2$ ก็จะอาศัยหลักการเกี่ยวกับการทดสอบสมมติฐาน $H_o:E(X)=E(Y)$ แต่จะพิจารณาอันดับของ $(x_i-\mu_x)^2$ และ $(y_j-\mu_y)^2$ และตัวสถิติที่ใช้ทดสอบก็ควรจะอยู่ในรูปของผลรวมของอันดับ เช่นเดียวกับการทดสอบค่าเฉลี่ย แต่จากการศึกษาของ Duman และ Mielke (1968) พบว่าการศึกษาเกี่ยวกับความแปรปรวนของประชากร ตัวสถิติทดสอบที่ใช้ผลรวมกำลังสองของอันดับจะให้ประสิทธิภาพสูงกว่าสถิติทดสอบที่ใช้ผลรวมของอันดับ

4.2 สถิติทดสอบ

$$T = \sum_{i=1}^{N} a_i Z_i$$

โดยที่ $a_i = r_i^2$

$$Z_i = \begin{cases} 1 & \text{เมื่อค่าสังเกตที่ } i \text{ มาจากกลุ่มตัวอย่างที่น้อยกว่า} \\ 0 & \text{นอกเหนือจากนั้น} \end{cases}$$

ค่า r_i เป็นอันดับของค่าผลต่างกำลังสองของค่าสังเกตุที่ i กับค่าเฉลี่ยของประชากรจากการจัดเรียงผลต่างกำลังสองของค่าสังเกตุกับค่าเฉลี่ยของประชากรระหว่างกลุ่มตัวอย่างทั้งสองกลุ่มร่วมกัน ในกรณีที่ไม่ทราบค่าเฉลี่ยของประชากรให้ใช้ค่าเฉลี่ยของกลุ่มตัวอย่างแทน (Conover 1980:240) เพราะว่าอันดับของการจัดเรียงค่าผลต่างกำลังสองร่วมกันเท่ากับอันดับของการจัดเรียงค่าสัมบูรณ์ร่วมกัน ดังนั้นในการคำนวณจะใช้กรณีไหนก็ได้

4.3 ตารางกำหนดค่าวิกฤติของการทดสอบสแควแรงค์

Conover และ Iman (1978) สร้างตารางกำหนดค่าวิกฤติของการทดสอบสแควแรงค์ เมื่อ $3 \leq m \leq 10$ และ $N \leq 25$ สำหรับระดับนัยสำคัญ .01, .025, .05 และ .10 สำหรับการทดสอบทางเดียว และ .02, .05, .10 และ .20 สำหรับการทดสอบแบบสองทาง

ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ ให้ประมาณด้วยการแจกแจงแบบปกติ คือ

$$Z = \frac{T - E(T)}{\sqrt{V(T)}}$$

$$\text{โดยที่ } E(T) = \frac{m(N+1)\ (2N+1)}{6}$$

$$V(T) = \frac{mn\ (N+1)\ (2N+1)\ (8N+11)}{180}$$

ถ้าหากว่ามีค่าสังเกตซ้ำกัน (Tied observation) ก็ใช้ค่าเฉลี่ยของอันดับ (Mid Rank) แล้วใช้วิธี J-Approximation ซึ่งมีสูตรในการคำนวณดังนี้

$$J = \frac{Z}{2}\left(1 + \sqrt{\frac{N-2}{N-1-Z^2}}\right)$$

เมื่อ $Z = \frac{T - E(T)}{\sqrt{V(T)}}$ และค่า J ที่คำนวณได้ให้นำไปเปรียบเทียบค่าจากตาราง J ซึ่งกำหนดโดย Iman (1976) หรืออาจจะคำนวณจากสูตร $J_\alpha = (Z_\alpha + t_\alpha)/2$ โดย Z มีการแจกแจงแบบปกติมาตรฐาน และ t มีการแจกแจงแบบที (t distribution)

ผลงานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่เกี่ยวข้องนั้น ปรากฏว่ามีนักวิจัยเป็นจำนวนมากได้ศึกษาค่าแอซซิมโทติครีเลทีฟเอฟฟิชเชียนซี (Asymptatic Relative Efficiency) ระหว่างการทดสอบเอฟ (F-test) และการทดสอบแบบนอนพาราเมตริก (Nonparametric) หรือระหว่างการทดสอบนอนพาราเมตริกด้วยกัน แต่ก็มีผู้วิจัยบางท่านที่ใช้วิธีมอนติคาร์โล (Monte Carlo Method) เพื่อศึกษาอำนาจการทดสอบเอฟและการทดสอบแบบนอนพาราเมตริก ซึ่งในส่วนนี้จะเสนอเฉพาะผลงานวิจัยที่สำคัญเท่านั้น

Klotz (1962:498-508) ศึกษาเปรียบเทียบประสิทธิภาพของการทดสอบทูกี้-ซีเกล และการทดสอบมู้ดเทียบกับการทดสอบคล้อท และได้ศึกษาประสิทธิภาพของการทดสอบทูกี้-ซีเกล เทียบกับการทดสอบมู้ด โดยใช้ limiting Pitman efficiency ซึ่งเป็น ARE ที่กล่าวใน หน้า 8 บทที่ 1 ดังตาราง 2.1

ตาราง 2.1 แสดงค่า Limiting Pitman Efficiency ภายใต้ลักษณะการแจกแจงแบบต่าง ๆ ของการทดสอบทูกี้-ซีเกล (T-S) การทดสอบมู้ด (M) และการทดสอบนอร์มอลสกอร์ (N-S) หรือการทดสอบคล้อทซ์

	T-S/N-S	M/N-S	T-S/M
Exponential	0.	0.	.600
Rectangular	0.	0.	.806
Normal	.608	.760	.800
Logistic	.750	.896	.837
Double Exponential	.774	.900	.860
Cauchy	1.783	1.67	1.068

ซึ่งจากค่า Limiting Pitman Efficiency แสดงว่า การทดสอบคล้อทซ์ หรือการทดสอบนอร์มอลสกอร์มีประสิทธิภาพสูงที่สุด เทียบกับการทดสอบมู้ดและการทดสอบทูกี้-ซีเกล ยกเว้นเมื่อประชากรมีการแจกแจงแบบ cauchy เท่านั้นที่การทดสอบมู้ดและการทดสอบทูกี้-ซีเกล มีประสิทธิภาพสูงกว่า ในขณะเดียวกันเมื่อเปรียบเทียบประสิทธิภาพระหว่างการทดสอบมู้ดและการทดสอบทูกี้-ซีเกล ปรากฏว่าการทดสอบมู้ด โดยทั่ว ๆ ไปแล้ว การทดสอบมู้ดมีประสิทธิภาพสูงกว่าการทดสอบทูกี้-ซีเกล นอกจากในกรณีที่ประชากรมีการแจกแจงแบบ cauchy เท่านั้นที่การทดสอบทูกี้-ซีเกล มีประสิทธิภาพสูงกว่าเล็กน้อย

Duran และ Mielke (1968:338-344) เปรียบเทียบค่า ARE ระหว่างการทดสอบสแควร์แรงค์ การทดสอบวิลค๊อกซัน (Wilcoxon) และการทดสอบ Locally most powerfull rank-test (LMPRT) สำหรับการแจกแจงแบบต่าง ๆ กัน ซึ่งมีค่าฟังก์ชันความน่าจะเป็น (probality density function) ดังตาราง 2.2

ตาราง 2.2 แสดงลักษณะการแจกแจงของประชากรและค่าฟังก์ชั่นความน่าจะเป็น ซึ่ง Duran และ Mielke ใช้เปรียบเทียบค่า ARE ของการทดสอบสแควร์แรงค์ การทดสอบวิลคอกซัน และการทดสอบ LMPRT

การแจกแจง	ฟังก์ชั่นความน่าจะเป็น
Gamma	$f(x) = \frac{x^{p-1} e^{-x}}{\Gamma p}, \quad x \geq 0, \quad p > 0$
truncated t	$f(x) = \frac{2\Gamma(r+1)/2}{(r\pi)^{1/2}\Gamma r/2 \; (1+\frac{x^2}{r})^{(r+1)/2}}, \quad x \geqslant 0$
truncated Logistic	$f(x) = \frac{2e^x}{(1+e^x)^2}, \quad x \geqslant 0$
Lognormal	$f(x) = x^{-1} (2\pi)^{-\frac{1}{2}} e^{-(\ln x)^2/2}, \quad x \geqslant 0$

ซึ่งผลปรากฏว่าเมื่อประชากรมีการแจกแจงแบบแกมมา (gamma) ค่า ARE ของการทดสอบสแควร์แรงค์เมื่อเทียบกับการทดสอบ LMPRT สูงที่สุด มีค่า .9372 เมื่อค่าพารามิเตอร์เป็น 5 และ 6 และการทดสอบสแควร์แรงค์มีค่า ARE เป็น 1 เมื่อเทียบกับการทดสอบ LMPRT เมื่อประชากรมีการแจกแจงแบบ truncated t ซึ่งมีค่าจำนวนชั้นแห่งความเป็นอิสระเป็น 2 สำหรับค่า ARE ของการทดสอบวิลคอกซันเทียบกับการทดสอบสแควร์แรงค์ ปรากฏว่าโดยส่วนมากแล้วค่า ARE มีค่าน้อยกว่า 1

Miller (1968:567-582) ใช้วิธีการมอนติคาร์โล ศึกษาความแกร่ง (Robustness) และอำนาจของการทดสอบระหว่างการทดสอบเอฟ การทดสอบบ๊อกซ์ และแอนเดอร์สัน (Box and Anderson-test) การทดสอบแจคไนฟ์ (Jacknife-test) การทดสอบลีเวนเอส (Levene -s-test) การทดสอบบ๊อกซ์ (Box-test) และการทดสอบโมเสส (Moses-test) เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน 2 รูปแบบ คือ 10 และ 25 ภายใต้ลักษณะการแจกแจงแบบปกติ ยูนิฟอร์ม ดับเบิลเอ็กโพเนนเชียล (Double Exponential) สคิวเนสเอ็กโพเนนเชียล (Skewness Exponential) และการแจกแจงแบบ Sixth Power ซึ่งมีฟังก์ชั่นความน่าจะเป็น คือ $f(x) = \frac{5}{2}(1+|x|)^{-6}$ ได้ข้อสรุปดังนี้

1. การทดสอบเอฟไม่แกร่ง (Nonrobust) เมื่อลักษณะการแจกแจงประชากรไม่เป็นแบบปกติ

2. เมื่อขนาดของกลุ่มตัวอย่างเป็น 25 การทดสอบบ๊อกซ์และแอนเดอสันกับการทดสอบแจคไนฟ์ (k=1) มีอำนาจการทดสอบพอ ๆ กันที่ α = .05 แต่อำนาจของการทดสอบแจคไนฟ์สูงกว่าเล็กน้อยที่ α = .01

3. การทดสอบแจคไนฟ์ เมื่อ k = 5 นั้นมีอำนาจการทดสอบน้อยกว่า เมื่อ k = 1 แต่ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 เมื่อ k = 5 ดีกว่า

4. เมื่อขนาดของกลุ่มตัวอย่างเป็น 10 การทดสอบบ๊อกซ์และแอนเดอร์สัน และการทดสอบแจคไนฟ์ เมื่อ k = 1 มีค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ต่างจากค่า α มาก

5. การทดสอบ ลีเวนเอส เป็นการทดสอบที่แกร่ง (Robust) แต่เป็นการทดสอบที่มีอำนาจของการทดสอบน้อยเมื่อเทียบกับการทดสอบแจคไนฟ์และการทดสอบบ๊อกซ์และแอนเดอร์สัน

Pearson และ Please (1975:223-241) ศึกษาความสัมพันธ์ระหว่างรูปแบบของการแจกแจงของประชากรกับความแกร่งของการทดสอบ 4 วิธี คือการทดสอบ ที สำหรับกลุ่มตัวอย่าง 1 กลุ่ม การทดสอบ ที สำหรับกลุ่มตัวอย่าง 2 กลุ่ม การทดสอบไคสแควร์ และการทดสอบเอฟที่ใช้ทดสอบความแปรปรวนของประชากรสองกลุ่ม เมื่อความเบ้ของประชากรมีค่าอยู่ระหว่าง [0,0.8] และความโด่งมีค่าในช่วง [2.2,4.4] ขนาดของกลุ่มตัวอย่างเท่ากัน คือ 10 และ 25 ได้ผลสรุปว่า

1. ความแกร่งของการทดสอบที มีอิทธิพลเนื่องจากความเบ้

2. ความแกร่งของการทดสอบไคสแควร์มีผลกระทบจากความโด่ง

3. ความแกร่งของการทดสอบเอฟมีผลกระทบจากความโด่ง

4. ความแกร่งของการทดสอบทีสำหรับกลุ่มตัวอย่างสองกลุ่ม ไม่มีอิทธิพลจากความเบ้และความโด่ง