บทที่ 2

ตัวสถิติทดสอบและผลงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ศึกษาเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติทดสอบต่าง ๆ ภายใต้ลักษณะการแจกแจงของประชากรชนิดลอง-เทลด์ ตัวสถิติทดสอบที่นำมาศึกษาเปรียบเทียบคือ ตัวสถิติทดสอบทริมด์ ที ตัวสถิติทดสอบแมน-วิทนีย์ ยู และตัวสถิติทดสอบ แวน เดอ แวร์เดน และสำหรับตัวสถิติทดสอบทริมด์ ที นั้นศึกษาเปรียบเทียบวิธีการเลือกระดับการทริมด์ 3 วิธี ได้แก่ วิธีค่าประมาณความแปรปรวนน้อยที่สุด วิธีคอมบายด์ คิว และวิธีเฉลี่ย คิว ซึ่งในบทนี้จะกล่าวถึงรายละเอียดสำหรับตัวสถิติทดสอบทริมด์ ที วิธีการเลือกระดับการทริมด์ของตัวสถิติทดสอบทริมด์ ที ทั้ง 3 วิธี ตัวสถิติทดสอบแมน-วิทนีย์ ยู และตัวสถิติทดสอบแวน เดอ แวร์เดน พร้อมทั้งเสนอผลงานวิจัยที่เกี่ยวข้องในส่วนท้ายบท

2.1 ตัวสถิติทดสอบทริมด์ ที

Yuen และ Dixon (1973: 369-374) ได้เสนอตัวสถิติทดสอบการเท่ากันของค่าเฉลี่ยประชากร 2 ชุด สำหรับประชากรที่มีลักษณะการแจกแจงชนิดลอง-เทลด์ที่สมมาตร และมีความแปรปรวนเท่ากัน ทั้งในกรณีที่ขนาดตัวอย่างทั้ง 2 ชุดเท่ากันและไม่เท่ากัน การคำนวณค่าสถิติทดสอบสำหรับตัวสถิติทดสอบทริมด์ ที นี้มีรูปแบบเดียวกับตัวสถิติทดสอบที แต่ตัวสถิติทดสอบทริมด์ ที ใช้ค่าเฉลี่ยทริมด์ (Trimmed mean) เป็นตัวประมาณค่าเฉลี่ยของประชากรแทนค่าเฉลี่ยมัชฌิมเลขคณิตในตัวสถิติทดสอบที และใช้ตัวประมาณความแปรปรวนของความแตกต่างระหว่างตัวประมาณค่าเฉลี่ยประชากรที่คำนวณจากส่วนเบี่ยงเบนกำลังสองแบบวินเซอร์ไรซ์แทนตัวประมาณความแปรปรวนของความแตกต่างระหว่างตัวประมาณค่าเฉลี่ยประชากรที่ใช้ในตัวสถิติทดสอบที

กำหนดให้ n แทนขนาดตัวอย่างที่เลือกมาจากแต่ละประชากร และ $x_1,\ldots,x_n$ เป็นสถิติอันดับ สูตรคำนวณค่าสถิติทดสอบทริมด์ ที ที่ใช้ทดสอบสมมติฐานว่าง $H_0: \mu_1 = \mu_2$ คือ

$$t_{t/w} = \frac{\bar{x}_{1tg} - \bar{x}_{2tg}}{\left[\dfrac{SSD_{1wg} + SSD_{2wg}}{h(h-1)}\right]^{1/2}} \quad ; \; h = n-2g$$

เมื่อ $\bar{x}_{tg} = \frac{1}{n-2g}(x_{g+1} + \ldots + x_{n-g})$ คือ ค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ g (g-times trimmed mean)

$$\bar{x}_{wg} = \frac{1}{n}\left[(g+1)x_{g+1} + x_{g+2} + \ldots, + x_{n-g-1} + (g+1)x_{n-g}\right]$$ คือ

ค่าเฉลี่ยวินเซอร์ไรซ์ที่ระดับ g (g-times Winsorized mean)

$$SSD_{wg} = (g+1)(x_{g+1} - \bar{x}_{wg})^2 + (x_{g+2} - \bar{x}_{wg})^2 + \ldots\ldots + (x_{n-g-1} - \bar{x}_{wg})^2 + (g+1)(x_{n-g} - \bar{x}_{wg})^2$$ คือผลรวมส่วนเบี่ยงเบนกำลังสองแบบวินเซอร์ไรซ์

g คือระดับการทริมด์ที่เป็นจำนวนเต็มโดยที่ $0 \leq g \leq \frac{1}{4} n$

จากการใช้เทคนิคมอนติคาร์โลซิมูเลชั่นพบว่าสามารถประมาณการแจกแจงของตัวสถิติทดสอบ $t_{t/w}$ ได้ด้วยการแจกแจงแบบที ที่องศาความเป็นอิสระ 2(h-1) เมื่อการแจกแจงของประชากรทั้ง 2 ชุดเป็นแบบปกติ โดยมีความแปรปรวนเท่ากัน และตัวอย่างสุ่มจากประชากรทั้ง 2 ชุด ซึ่งมีขนาดตัวอย่างเท่ากัน เป็นอิสระต่อกัน และมีขนาดใหญ่กว่า 6

สำหรับกรณีที่ขนาดตัวอย่างที่เลือกมาจากประชากรทั้ง 2 ไม่เท่ากัน ถ้าขนาดตัวอย่างที่เลือกมาจากประชากรแต่ละชุดเท่ากับ n_1 และ n_2 สูตรคำนวณค่าสถิติทดสอบของตัวสถิติทดสอบทริมด์ ที เป็นดังนี้

$$t_{t/w} = \frac{\bar{x}_{1tg} - \bar{x}_{2tg}}{\left[\left(\frac{SSD_{1wg} + SSD_{2wg}}{h_1 + h_2 - 2}\right)\left(\frac{1}{h_1} + \frac{1}{h_2}\right)\right]^{\frac{1}{2}}} \qquad h_i = n_i - 2g_i, \quad i = 1,2$$

$t_{t/w}$ จะมีการแจกแจงแบบที โดยประมาณที่องศาความเป็นอิสระ h_1+h_2-2 เมื่อการแจกแจงของประชากรทั้ง 2 ชุดเป็นแบบปกติ ซึ่งมีความแปรปรวนเท่ากัน และตัวอย่างสุ่มจากประชากรทั้ง 2 ชุดเป็นอิสระต่อกัน โดยที่ระดับการทริมด์ของตัวอย่างทั้ง 2 ชุดคือ g_1 และ g_2 เป็นอัตราส่วนเดียวกับอัตราส่วนของ n_1 และ n_2

อนึ่งสูตรคำนวณค่าสถิติทดสอบทริมด์ ที กรณีขนาดตัวอย่างไม่เท่ากัน จะเป็นสูตรเดียวกับกรณีขนาดตัวอย่างเท่ากัน คือเมื่อ $n_1 = n_2 = n$ และหากไม่มีการทริมด์ คือเมื่อ $g = 0$ ตัวสถิติทดสอบทริมด์ ที จะเป็นตัวสถิติทดสอบ ที นั่นเอง

อย่างไรก็ตาม Yuen และ Dixon (1973: 369-374) มิได้กล่าวถึงการเลือกระดับการทริมด์ในทางปฏิบัติเมื่อได้ค่าตัวอย่างมาแล้ว ดังนั้นผู้วิจัยจึงได้ศึกษารวบรวมวิธีการเลือกระดับการทริมด์สำหรับตัวสถิติทดสอบทริมด์ ที ซึ่งสามารถสรุปได้ 3 วิธี โดยแต่ละวิธีมีรายละเอียดดังนี้

1. วิธีค่าประมาณความแปรปรวนน้อยที่สุด (Minimum estimate variance)

เนื่องจากตัวสถิติทดสอบทริมด์ ที สำหรับกรณีประชากร 2 ชุดนี้ เป็นตัวสถิติทดสอบที่ Yuen และ Dixon ดัดแปลงจากตัวสถิติทดสอบทริมด์ ที สำหรับกรณีประชากร 1 ชุด (one-sample trimmed t) ของ Tukey และ Mc. Laughlin (1963: 331-352, อ้างถึงใน Yuen และ Dixon 1973: 369) ซึ่งใช้การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร คือทดสอบสมมติฐานว่าง $H_0 : \mu = \mu_0$ โดยมีสูตรคำนวณค่าสถิติทดสอบดังนี้

$$t_{t/g} = \frac{\bar{x}_{tg} - \mu_0}{\left[\frac{SSD_{wg}}{h(h-1)}\right]^{1/2}} \quad ; \quad h = n - 2g$$

โดย $t_{t/g}$ มีการแจกแจงแบบที โดยประมาณ ที่องศาความเป็นอิสระ h-1 เมื่อการแจกแจงของประชากรเป็นแบบปกติ

$\left[\frac{SSD_{wg}}{h(h-1)}\right]^{1/2}$ เป็นค่าประมาณส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ g ซึ่ง Tukey และ Mc. Laughlin (Hogg 1974: 911) เลือกระดับการทริมด์ g ที่ทำให้ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยทริมด์มีค่าน้อยที่สุด และโดยที่

$$\left[\frac{SSD_{1wg} + SSD_{2wg}}{h(h-1)}\right]^{1/2} \quad \text{และ} \quad \left[\left(\frac{SSD_{1wg} + SSD_{2wg}}{h_1+h_2-2}\right)\left(\frac{1}{h_1} + \frac{1}{h_2}\right)\right]^{1/2}$$

คือค่าประมาณส่วนเบี่ยงเบนมาตรฐานของผลต่างค่าเฉลี่ยทริมด์ของตัวอย่าง 2 ชุด คือ $\bar{x}_{1tg} - \bar{x}_{2tg}$ เมื่อขนาดตัวอย่างเท่ากัน และไม่เท่ากันตามลำดับ (Hogg 1975: 659) ดังนั้นการวิจัยครั้งนี้จึงศึกษาวิธีการเลือกระดับการทริมด์สำหรับค่าสถิติทดสอบทริมด์ ที โดยเลือกระดับการทริมด์ g ที่เป็นจำนวนเต็มในช่วง $0 \leq g \leq \frac{1}{4} n$ ที่ให้ค่า $\left[\frac{SSD_{1wg} + SSD_{2wg}}{h(h-1)}\right]^{1/2}$ มีค่าน้อยที่สุดเมื่อขนาดตัวอย่างเท่ากัน และเลือกระดับการทริมด์ g_1 ที่เป็นจำนวนเต็มในช่วง $0 \leq g_1 \leq \frac{1}{4}n_1$ และ $g_2 = g_1 n_2/n_1$ ที่ทำให้ค่า $\left[\left(\frac{SSD_{1wg} + SSD_{2wg}}{h_1+h_2-2}\right)\left(\frac{1}{h_1} + \frac{1}{h_2}\right)\right]^{1/2}$ มีค่าน้อยที่สุดเมื่อขนาดตัวอย่างไม่เท่ากัน

2. วิธีคอมบายด์ คิว

โดยทั่วไปนักสถิติมักจะใช้ค่าความโด่งจากตัวอย่าง (Sample Kurtosis) และค่าความเบ้จากตัวอย่าง (Sample Skewness) ในการคาดลักษณะการแจกแจงของประชากร โดยเฉพาะอย่างยิ่งใช้พิจารณาถึงระดับความมากน้อยของลักษณะลอง-เทลด์ของการแจกแจง

(Ferguson 1961, อ้างถึงใน Huber 1972: 1044) มีนักสถิติหลายท่านได้ใช้ค่าความโด่งจากตัวอย่างเป็นเกณฑ์ในการเลือกระดับการทริมด์สำหรับค่าเฉลี่ยทริมด์ในการประมาณค่าเฉลี่ยของประชากรที่มิได้มีลักษณะการแจกแจงเป็นแบบปกติ ต่อมา Hogg และ Randle ได้เสนอค่าสถิติ Q_1 เป็นดัชนีวัดระดับความเป็นลอง-เทลด์ หรือช่อร์ท-เทลด์ (Long-tailed or Short-tailed) ของการแจกแจงคือ

$$Q_1 = [\bar{U}(.05)-\bar{L}(.05)]/[\bar{U}(.5)-\bar{L}(.5)]$$

โดยที่ $\bar{U}(\beta)$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลที่มากที่สุดจำนวน $n\beta$ และ $\bar{L}(\beta)$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลที่น้อยที่สุดจำนวน $n\beta$ เมื่อ n คือจำนวนข้อมูลทั้งหมด ในกรณีที่ $n\beta$ ไม่เป็นจำนวนเต็มให้คำนวณค่า $\bar{U}$ หรือ $\bar{L}$ ตามสัดส่วน เช่น $n = 30$ จะได้ $(.05)\cdot n = 1.5$ ดังนั้น $\bar{L}(.05) = (x_1 + .5x_2)/1.5$

Hogg(1974: 913) กล่าวว่าเนื่องจาก Q_1 เป็นอัตราส่วนของฟังก์ชั่นเส้นตรงของสถิติอันดับ ดังนั้นจึงมีคุณสมบัติการลู่เข้า (Convergence) ดีกว่าความโด่ง ค่า Q_1 ของประชากรจะมีค่าประมาณ 1.95 2.58 และ 3.30 สำหรับการแจกแจงแบบยูนิฟอร์ม แบบปกติ และแบบดับเบิลเอกซ์โปเนนเชียล ตามลำดับ และในกรณีที่ศึกษาเฉพาะการแจกแจงของประชากรชนิดลอง-เทลด์เท่านั้น Hogg ได้ดัดแปลงตัวสถิติ Q_1 เป็นตัวสถิติ Q คือ

$$Q = [\bar{U}(.2)-\bar{L}(.2)]/[\bar{U}(.5)-\bar{L}(.5)]$$

โดยมีเหตุผลว่า ค่า $\bar{U}(\beta) - \bar{L}(\beta)$ เมื่อ β มีค่าประมาณ $\frac{1}{6}$ หรือ $\frac{1}{5}$ จะเป็นค่าประมาณที่ดีสำหรับส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปกติ และค่า $\bar{U}(.5)-\bar{L}(.5)$ ซึ่งเท่ากับ $\sum_{i=1}^{n} |x_i\text{-Median}|/n$ จะเป็นค่าประมาณที่ดีสำหรับส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบดับเบิลเอกซ์โปเนนเชียล เมื่อการแจกแจงของประชากรเป็นแบบปกติ Q ของประชากรจะมีค่าประมาณ 1.75 และจะมีค่าประมาณ 1.93 สำหรับการแจกแจงแบบดับเบิลเอกซ์โปเนนเชียล ส่วนการแจกแจงแบบโลจิสติคจะได้ค่า Q ของประชากรเป็น 1.81 สถิติ Q ได้ถูกนำไปใช้สร้างตัวสถิติทดสอบสมมติฐานเกี่ยวกับระดับความเป็นลอง-เทลด์ของประชากร (Hogg 1972: 422-424) Q และ Q_1 เป็นสถิติที่ไม่ขึ้นกับตำแหน่งของการแจกแจงของประชากร (Location free statistics) ดังนั้นจึงไม่มีความสัมพันธ์กับค่าเฉลี่ยทริมด์ นั่นคือไม่มีผลกระทบต่อคุณสมบัติการเป็นตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยทริมด์ในกรณีที่การแจกแจงของประชากรสมมาตร

ในการหาค่าประมาณของค่าเฉลี่ยประชากรที่มีการแจกแจงชนิดลอง-เทลด์นั้น Hogg ได้แนะนำให้ใช้ค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ g ต่าง ๆ ในช่วง 0 ถึง $\frac{3}{8}$ เมื่อ n คือขนาดตัวอย่าง เป็นตัวประมาณค่าเฉลี่ยของประชากร ซึ่งการเลือกระดับการทริมด์ g จะขึ้นอยู่กับค่าสถิติ Q และขนาดตัวอย่าง ดังนี้คือ

เมื่อ $n \leq 10$ เลือกระดับการทริมด์ $g = \left[\frac{3}{16}\right]$ สำหรับทุกค่าของ Q

เมื่อ $10 < n \leq 20$ เลือกระดับการทริมด์ $g = \left[\frac{1}{8}n\right]$ ถ้า $Q \leq 1.84$

เลือกระดับการทริมด์ $g = \left[\frac{2}{8}n\right]$ ถ้า $1.84 < Q$

เมื่อ $20 < n \leq 30$ เลือกระดับการทริมด์ $g = \left[\frac{3}{32}n\right]$ ถ้า $Q \leq 1.81$

เลือกระดับการทริมด์ $g = \left[\frac{6}{32}n\right]$ ถ้า $1.81 < Q \leq 1.87$

เลือกระดับการทริมด์ $g = \left[\frac{9}{32}n\right]$ ถ้า $1.87 < Q$

และเลือกในทำนองเดียวกันเมื่อขนาดตัวอย่างมากกว่า 30 โดยที่ [a] คือจำนวนเต็มที่มากที่สุดที่น้อยกว่าหรือเท่ากับ a

Randles และ Hogg (1973: 337-356, อ้างถึงใน Hogg 1974: 909-923) ได้เสนอตัวสถิติทดสอบอแดพทีฟ ดิสตริบิวชั่น-ฟรี (Adaptive distribution-free test) กล่าวคือเป็นตัวสถิติทดสอบที่เลือกตัวสถิติทดสอบแบบดิสตริบิวชั่น-ฟรีต่าง ๆ ในการใช้ทดสอบตัวอย่างสุ่มที่ได้ให้เหมาะสมกับลักษณะของข้อมูลตัวอย่างนั้น เช่น ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างค่าเฉลี่ยประชากรแบบต่อเนื่อง 2 ชุด ที่เป็นอิสระต่อกัน อาจเลือกใช้ตัวสถิติทดสอบมีเดียน (Median test) ตัวสถิติทดสอบวิลค๊อกสัน (Wilcoxon test) หรือตัวสถิติทดสอบอย่างเร็วของตูกี้ (Tukey's quick test) เมื่อกำหนดระดับนัยสำคัญในการทดสอบของแต่ละตัวสถิติทดสอบเท่ากัน คือ α และหากใช้เกณฑ์ในการเลือกตัวสถิติทดสอบที่เป็นอิสระกับตัวสถิติทดสอบแบบดิสตริบิวชั่น-ฟรี เหล่านั้นแล้ว ระดับนัยสำคัญทั้งหมดของตัวสถิติทดสอบอแดพทีฟ ดิสตริบิวชั่น-ฟรีจะเท่ากับ α ด้วย กล่าวคือ ถ้าสมมติให้ R_1 R_2 และ R_3 แทนเขตวิกฤติของตัวสถิติทดสอบทั้ง 3 ดังกล่าวข้างต้น และ C_1 C_2 และ C_3 แทนเหตุการณ์ที่จะเลือกตัวสถิติทดสอบแต่ละตัว ดังนั้นระดับนัยสำคัญทั้งหมดของตัวสถิติทดสอบอแดพทีฟ ดิสตริบิวชั่น-ฟรี จะเท่ากับ $Pr.(C_1 \cap R_1) + Pr.(C_2 \cap R_2) + Pr.(C_3 \cap R_3)$ ถ้าเกณฑ์ในการเลือก

ใช้ตัวสถิติทดสอบเป็นอิสระกับตัวสถิติทดสอบแล้ว จะได้ระดับนัยสำคัญทั้งหมดเป็น $Pr.(C_1)\cdot Pr.(R_1) + Pr.(C_2)\cdot Pr.(R_2) + Pr.(C_3)\cdot Pr.(R_3)$ เมื่อกำหนดระดับนัยสำคัญ α สำหรับทุกตัวสถิติทดสอบจะได้ $[Pr.(C_1)+Pr.(C_2)+Pr.(C_3)]\cdot\alpha$ นั่นคือระดับนัยสำคัญทั้งหมดเท่ากับ α ซึ่งทำให้ตัวสถิติทดสอบอแดพทีฟ ดิสตริบิวชั่น-ฟรี เป็นตัวสถิติทดสอบที่มีคุณสมบัติดิสตริบิวชั่น-ฟรี

เกณฑ์ที่ใช้ในการเลือกตัวสถิติทดสอบซึ่งเป็นอิสระกับตัวสถิติทดสอบแบบดิสตริบิวชั่น-ฟรีสำหรับกรณีตัวอย่างสุ่ม 2 ชุดนั้น Hogg ได้เสนอให้ใช้สถิติอันดับที่ได้จากการจัดลำดับร่วมกันของตัวอย่างสุ่มทั้ง 2 ชุดและตัวสถิตินี้จะเป็นสถิติที่พอเพียง (Complete sufficient statistics) สำหรับทุกการแจกแจงของประชากรแบบต่อเนื่องเมื่อสมมติฐานว่างเป็นจริง จึงทำให้เป็นตัวสถิติที่เป็นอิสระกับตัวสถิติทดสอบที่จะเลือกใช้ ซึ่งตัวสถิติ Q มีคุณสมบัติดังกล่าว แม้ว่าในกรณีที่สมมติฐานแย้งเป็นจริง ตัวสถิตินี้จะไม่เป็นอิสระกับตัวสถิติทดสอบต่าง ๆ และทำให้ลดความถูกต้องในการเลือกใช้ตัวสถิติทดสอบที่เหมาะสมกับลักษณะของข้อมูล แต่ Hogg กล่าวว่าอย่างไรก็ตาม หากมีความแตกต่างระหว่างค่าเฉลี่ยที่แท้จริงของประชากรทั้ง 2 ชุดไม่มากนัก จะไม่มีผลเปลี่ยนแปลงในการเลือกใช้ตัวสถิติทดสอบ แต่ถ้ามีความแตกต่างกันมากไม่ว่าจะเลือกตัวสถิติทดสอบใดโดยปกติจะมีโอกาสมากอยู่แล้วที่จะได้ผลการทดสอบเป็นการปฏิเสธสมมติฐานว่างจากตัวสถิติทดสอบนั้น

ดังนั้นการเลือกระดับการทริมด์โดยใช้วิธีคอมบายด์ คิว สำหรับการวิจัยนี้จึงใช้ค่าสถิติ Q ที่คำนวณจากการนำตัวอย่างสุ่มทั้ง 2 ชุดมาจัดอันดับร่วมกันและนำค่าสถิติ Q ที่ได้เป็นตัวกำหนดระดับการทริมด์ g ตามเกณฑ์ที่ Hogg กำหนด ดังได้กล่าวแล้วข้างต้น สำหรับกรณีขนาดตัวอย่างไม่เท่ากันนั้นใช้ขนาดตัวอย่างที่น้อยกว่าเป็นขนาดตัวอย่างในการพิจารณาตามเกณฑ์ของ Hogg

3. วิธีเฉลี่ย คิว

Fisher (Hogg 1974: 920) ได้ศึกษาวิธีการเลือกประชากรที่มีค่าเฉลี่ยประชากรมากที่สุด 1 ประชากรจากประชากร k ชุดที่เป็นอิสระต่อกัน ซึ่งมีการแจกแจงแบบเดียวกัน และมีความแปรปรวนเท่ากัน วิธีการเลือกวิธีการหนึ่งที่ Fisher ศึกษาได้แก่วิธีการเลือกแบบอแดพทีฟ คือเลือกประชากรที่ให้ค่าของตัวสถิติตัวใดตัวหนึ่งซึ่งมีค่ามากที่สุด ส่วนจะเลือกประชากรที่มีค่าเฉลี่ยประชากรมากที่สุดโดยพิจารณาจากตัวสถิติตัวใดนั้น Fisher ใช้ค่าสถิติ Q_1 และ Q_2 เป็นเกณฑ์ในการพิจารณา โดย Q_1 เป็นดัชนีบอกระดับความเป็น

ลอง-เทลด์ หรือ ช่อร์ท-เทลด์ และ Q_2 เป็นดัชนีบอกระดับความเบ้ ซึ่ง $Q_2 = |\bar{U}(.05)-m(.25)| / |m(.25)-\bar{L}(.05)|$ เมื่อ m(.25) คือค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ g เท่ากับ $\frac{1}{4}$ ของขนาดตัวอย่าง เนื่องจากเป็นการเปรียบเทียบประชากรจำนวน k ชุด ดังนั้นจึงมีค่า Q_1 และ Q_2 อย่างละจำนวน k ค่า Fisherได้ใช้ค่าเฉลี่ยของ Q_1 และ Q_2 คือ $\bar{Q}_1$ และ $\bar{Q}_2$ เป็นเกณฑ์ในการเลือกตัวสถิติที่จะใช้ในการเปรียบเทียบเพื่อเลือกประชากรที่มีค่าเฉลี่ยมากที่สุด

ดังนั้นในการวิจัยครั้งนี้จึงทดลองศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติทดสอบทริมด์ ที โดยใช้วิธีการเลือกระดับทริมด์แบบวิธีเฉลี่ย คิว ด้วย กล่าวคือใช้ค่าเฉลี่ยของ Q จากตัวอย่างทั้ง 2 ชุดในการพิจารณาเลือกระดับการทริมด์โดยใช้เกณฑ์ของ Hogg ดังกล่าวแล้ว ในวิธีคอมบายด์ คิว สำหรับกรณีที่ขนาดตัวอย่างจากประชากรทั้งสองไม่เท่ากันจะใช้ค่าเฉลี่ย Q แบบถ่วงน้ำหนักด้วยขนาดตัวอย่างที่ใช้

2.2 <u>ตัวสถิติทดสอบแมน-วิทนีย์ ยู</u>

ตัวสถิติทดสอบแมน-วิทนีย์ ยู เป็นตัวสถิติทดสอบแบบนอนพาราเมตริกและเป็นตัวสถิติทดสอบที่ไม่ขึ้นกับการแจกแจงของประชากรคือเป็นตัวสถิติทดสอบแบบดิสตริบิวชั่น-ฟรี Mann และ Whitney เป็นผู้เสนอตัวสถิติทดสอบนี้ในปี ค.ศ. 1947 เพื่อใช้เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากร 2 ชุดที่เป็นอิสระต่อกัน โดยพิจารณารูปแบบของตัวอย่างสุ่มที่ได้จากการนำตัวอย่างสุ่มจากประชากรทั้ง 2 ชุดมาจัดเรียงอันดับร่วมกัน และหาผลรวมของจำนวนครั้งที่ค่าตัวอย่างสุ่มจากประชากรชุดหนึ่งมากกว่าค่าตัวอย่างสุ่มจากประชากรอีกชุดหนึ่ง

ข้อตกลงเบื้องต้นของตัวสถิติทดสอบแมน-วิทนีย์ ยู สำหรับการทดสอบสมมติฐานการเท่ากันของค่าเฉลี่ยประชากร 2 ชุด (Gibbon 1971: 140-149) มีดังนี้

1. ข้อมูลประกอบด้วยตัวอย่างสุ่ม $x_1, x_2, ..., x_m$ จากประชากร X และ $y_1, y_2, ..., y_n$ จากประชากร Y
2. ตัวอย่างสุ่มทั้ง 2 ชุดเป็นอิสระต่อกัน
3. ประชากรที่ศึกษามีการแจกแจงแบบต่อเนื่อง
4. สเกลที่ใช้วัดข้อมูลอย่างน้อยเป็นสเกลแบบจัดเรียงอันดับ (Ordinal Scale)
5. ประชากรที่ศึกษามีรูปแบบการแจกแจงแบบเดียวกันและมีการกระจายเท่ากัน

ค่าสถิติทดสอบของตัวสถิติทดสอบแมน-วิทนีย์ ยู คือ U และ U' ซึ่งคำนวณได้ดังนี้

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

$$U' = mn - U = \sum_{i=1}^{m} \sum_{j=1}^{n} (1 - D_{ij})$$

เมื่อ $D_{ij} = \begin{cases} 1 \text{ ถ้า } y_i < x_i & \text{สำหรับทุกค่าของ } i = 1, 2, \ldots, m \\ 0 \text{ ถ้า } y_j > x_i & \text{สำหรับทุกค่าของ } j = 1, 2, \ldots, n \end{cases}$

ถ้า m หรือ n มีขนาดใหญ่การคำนวณค่า U และ U' ดังวิธีข้างต้นอาจไม่สะดวก อาจคำนวณค่า U และ U' ได้อีกวิธีหนึ่งดังนี้

$$U = mn + \frac{m(m+1)}{2} - R_X$$

$$U' = mn + \frac{n(n+1)}{2} - R_Y$$

โดยที่ R_X = ผลรวมของค่าอันดับของ X จากการจัดอันดับร่วมกันของ X และ Y

R_Y = ผลรวมของค่าอันดับของ Y จากการจัดอันดับร่วมกันของ X และ Y

ถ้าค่า U หรือ U' ที่คำนวณได้มีค่าน้อยกว่า $C_{\alpha/2}$ จะปฏิเสธสมมติฐานว่างสำหรับกรณีการทดสอบแบบ 2 ด้าน และสำหรับการทดสอบแบบด้านเดียว (One-sided test) จะปฏิเสธสมมติฐานว่างเมื่อ $U < c_\alpha$ สำหรับสมมติฐานแย้งประชากร X มากกว่าประชากร Y และปฏิเสธสมมติฐานว่างเมื่อ $U' < c_\alpha$ สำหรับสมมติฐานแย้งประชากร Y มากกว่าประชากร X โดยค่าวิกฤต (Critical value) c นี้ สามารถหาได้จากตาราง Mann-Whitney ทั่วไป

ในกรณีที่ตัวอย่างสุ่มมีขนาดใหญ่ ค่าสถิติ U จะมีการแจกแจงโดยประมาณเป็นแบบปกติ โดยมีค่าคาดหวังและความแปรปรวนเป็นดังนี้

$$E(U) = mn/2$$

$$Var(U) = mn(m+n+1)/12$$

ดังนั้น $Z = \frac{U - (mn/2)}{[mn(m+n+1)/12]^{1/2}}$

จะมีการแจกแจงโดยประมาณเป็นแบบปกติมาตรฐาน ซึ่งสามารถหาค่าวิกฤติได้จากตารางพื้นที่ใต้โค้งปกติและการประมาณนี้สามารถใช้ได้ดีแม้ขนาดตัวอย่างทั้ง 2 ชุด มีขนาดเพียง 6 เท่านั้น (Gibbon 1971: 145)

2.3 <u>ตัวสถิติทดสอบแวน เดอ แวร์เดน</u>

แวน เดอ แวร์เดน เป็นผู้สร้างตัวสถิติทดสอบนี้ขึ้นมาเมื่อ ค.ศ. 1953 มีวัตถุประสงค์เพื่อให้ได้ตัวสถิติทดสอบที่ให้ค่าแอสซิมโทติครีเลทีฟเอฟฟิเชียนซี สูงขึ้นกว่าตัวสถิติทดสอบแบบนอนพาราเมตริกที่ใช้อันดับแทนค่าสังเกต โดยแวน เดอ แวร์เดน ได้แปลงค่าอันดับให้เป็นค่าอินเวอร์ส-นอร์มอล สกอร์ (Inverse-normal scores) ก่อน แล้วจึงนำไปคำนวณค่าสถิติทดสอบ

ข้อตกลงเบื้องต้นของตัวสถิติทดสอบแวน เดอ แวร์เดน สำหรับการทดสอบสมมติฐานการเท่ากันของค่าเฉลี่ยประชากร 2 ชุดเป็นเช่นเดียวกับข้อตกลงเบื้องต้นของตัวสถิติทดสอบแมน-วิทนีย์ ยู ดังได้กล่าวแล้ว

ค่าสถิติทดสอบของตัวสถิติทดสอบแวน เดอ แวร์เดน คือ V ซึ่งมีขั้นตอนในการคำนวณดังนี้

ให้ $r_1, r_2, \ldots, r_N$ เป็นค่าอันดับของค่าสังเกตที่ได้จากการจัดอันดับร่วมกันของค่าตัวอย่างสุ่มทั้ง 2 ชุด

และให้ p_i เป็นเปอร์เซ็นไทล์แรงค์ (Percentile rank) ซึ่งสัมพันธ์กับนอร์มอลไลซ์ออบเสิร์ฟเวชั่น (Normalized observation) คือ

$$p_i = \frac{r_i}{N+1}, \quad i = 1, 2, \ldots, N$$

$$; N = m+n$$

เมื่อให้ $\Phi(z^{(i)})$ คือ Cumulative standard normal distribution ของ $z^{(i)}$ จะได้ว่า

$$p_i = \Phi(z^{(i)})$$

จะได้อินเวอร์ส-นอร์มอลสกอร์ คือ $z^{(i)} = \Phi^{-1}(p_i) = \Phi^{-1}\left(\frac{r_i}{N+1}\right)$

ซึ่งอาจจะหาค่าของ $z^{(i)}$ ได้จากการใช้สูตรข้างต้นนี้จากตารางพื้นที่ใต้โค้งปกติ หรืออาจได้ค่า $z^{(i)}$ จากตารางที่ได้คำนวณค่า $z^{(i)}$ ไว้เรียบร้อยแล้ว เมื่อทราบค่า p_i (Marascuilo 1977: 485) หรือเมื่อทราบค่าของ r_i

ค่าสถิติทดสอบคือ $V = |V_1| = |V_2|$

เมื่อ V_1 คือ ผลรวมของค่าอินเวอร์ส-นอร์มอลสกอร์ของประชากร X

V_2 คือ ผลรวมของค่าอินเวอร์ส-นอร์มอลสกอร์ของประชากร Y

นำค่า V ที่ได้นี้ไปเปรียบเทียบกับค่าวิกฤติที่ได้จากตารางที่แสดงค่าวิกฤติสำหรับตัวสถิติทดสอบแวน เดอ แวร์เดน ซึ่ง Waerden และ Nievergelt ได้คำนวณไว้ (Gibbon 1971: 169) การคำนวณค่าวิฤติที่ระดับนัยสำคัญ α มีวิธีการคำนวณดังนี้คือ นำค่า $z^{(i)}$ ทั้งหมดมาเลือกครั้งละ n ตัว แล้วหาผลรวม จะได้ค่าผลรวมทั้งหมด $\binom{m+n}{n}$ ค่า เรียงลำดับค่าผลรวมที่ได้จากน้อยไปมาก ตัดค่าผลรวมที่จัดอันดับแล้ว $\alpha\binom{m+n}{n}$ ค่าสำหรับการทดสอบแบบด้านเดียว หรือ $\alpha\binom{m+n}{n}/2$ ค่าสำหรับการทดสอบแบบสองด้าน ถ้าค่า V ตกอยู่ในเขตวิกฤติจะปฏิเสธสมมติฐานว่าง

เมื่อตัวอย่างมีขนาดใหญ่ ค่าสถิติ V จะมีการแจกแจงแบบปกติโดยประมาณ โดยมีค่าคาดหวังและความแปรปรวนเป็นดังนี้

$$E(V) = 0, \quad Var(V) = \frac{m+n}{m+n+1} \cdot \sum_{i=1}^{N} \left(\frac{Z^{(i)}}{N}\right)^2$$

ดังนั้น
$$Z = \frac{V}{\left[var(V)\right]^{1/2}}$$

จะมีการแจกแจงโดยประมาณเป็นแบบปกติมาตรฐาน ซึ่งสามารถหาค่าวิกฤติได้จากตารางพื้นที่ใต้โค้งปกติ และการประมาณนี้ใช้ได้ดีเมื่อขนาดตัวอย่างทั้งสองมีค่ามากกว่า 7 ขึ้นไป (Marascuilo 1977: 283)

2.4 ผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่สำคัญซึ่งเกี่ยวข้องกับการวิจัยในครั้งนี้ มีดังนี้

Yuen และ Dixon (1973: 369-374) กล่าวว่าตัวสถิติทดสอบที สำหรับกรณีประชากร 1 ชุด ยังให้ผลไม่เป็นที่น่าพอใจนักในกรณีที่การแจกแจงของประชากรไม่เป็นแบบปกติ มีนักสถิติหลายท่านได้พยายามหาตัวสถิติทดสอบสำหรับค่าเฉลี่ยประชากรที่มีประสิทธิภาพสูงและมีความแกร่ง (Robust) สำหรับการแจกแจงที่ไม่เป็นแบบปกติ เช่น ตัวสถิติทดสอบทริมด์ ที ตัวสถิติทดสอบวินเซอร์ไรซ์ ที ซึ่งต่างเป็นตัวสถิติทดสอบที่สร้างขึ้นเพื่อใช้สำหรับการแจกแจงของประชากรชนิดลอง-เทลด์ Yuen และ Dixon ได้เสนอตัวสถิติทดสอบทริมด์ ที สำหรับทดสอบความแตกต่างของค่าเฉลี่ยประชากรที่เป็นอิสระต่อกันสองชุด โดยดัดแปลงจากตัวสถิติทดสอบทริมด์ ที กรณีประชากรชุดเดียว เหตุที่ดัดแปลงจากตัวสถิติทดสอบทริมด์ ที เนื่องจากการศึกษาของกลุ่มนักสถิติจาก Princeton ที่ศึกษาเกี่ยวกับความแกร่งของตัวประมาณตำแหน่งของประชากร (Robust estimates of population location) พบว่า ค่าเฉลี่ยทริมด์มีประสิทธิภาพสูงสำหรับการแจกแจงในหลายรูปแบบ และที่การแจกแจงแบบปกติประสิทธิภาพของตัวประมาณแบบค่าเฉลี่ยทริมด์เมื่อเทียบกับค่าเฉลี่ยเลขคณิต มีค่าค่อนข้างสูง คือมีเศษส่วนการสูญเสีย (fractional loss) เพียงประมาณ $\frac{2}{3}$ g/n เท่านั้น ในขณะที่ตัวประมาณแบบอื่น เช่น มัธยฐาน มีเศษส่วนการสูญเสียถึงหนึ่งในสาม และนอกจากนั้นยังทราบลักษณะการแจกแจงโดยประมาณของตัวสถิติทดสอบทริมด์ ที กรณีประชากรชุดเดียว ซึ่งเป็นแนวทางในการประมาณการแจกแจงสำหรับตัวสถิติทดสอบทริมด์ ที กรณีประชากร 2 ชุด

ในการประมาณการแจกแจงของตัวสถิติทดสอบทริมด์ ที กรณีประชากร 2 ชุด Yuen และ Dixon ได้ใช้เทคนิคมอนติคาร์โลซิมูเลชั่น โดยกำหนดการแจกแจงของประชากรเป็นแบบปกติ และทดลองศึกษาที่ขนาดตัวอย่างเป็น 5 7 10 และ 20 ใช้ระดับการทริมด์ตั้งแต่ 0 ถึง $\frac{1}{4}$ ของขนาดตัวอย่าง ผลปรากฏว่าตัวสถิติทดสอบทริมด์ ที สามารถประมาณได้ด้วยการแจกแจงแบบที ที่องศาอิสระ 2(h-1) เมื่อขนาดตัวอย่างมากกว่า 6 และผลโดยทั่วไปการแจกแจงของตัวสถิติทดสอบทริมด์ ที จะมีความน่าจะเป็นที่บริเวณปลายหางสูงกว่าการแจกแจงแบบทีเล็กน้อย โดยเฉพาะที่ระดับการทริมด์มากขึ้น ยกเว้นที่ขนาดตัวอย่างน้อย ๆ คือ 5 และระดับการทริมด์เท่ากับ 1 จะประมาณได้ดีมาก แต่อย่างไรก็ตามสำหรับกรณีตัวอย่างมีขนาดเล็กคือน้อยกว่า 7 Yuen และ Dixon แนะนำให้ใช้ค่าวิกฤติจากตารางที่คำนวณได้จากการทำซิมูเลชั่นจะเหมาะสมกว่าใช้ค่าวิกฤติที่ได้จากตารางการแจกแจงแบบที นอกจากนั้นยังได้ทำ

ซิมูเลชั่นเพื่อศึกษาอำนาจการทดสอบและหาค่ารีเลทีฟเอฟฟิเชียนซี โดยใช้การทดสอบสมมติฐานแบบสองทางที่ระดับนัยสำคัญ .05 ซึ่งคาดว่าที่ระดับนัยสำคัญอื่น ๆ จะให้ผลทำนองเดียวกัน เมื่อกำหนดขนาดตัวอย่างเป็น 5 7 10 และ 20 สำหรับการแจกแจงแบบปกติ ให้ผลการทดลองคือค่าอำนาจการทดสอบของตัวสถิติทดสอบทริมด์ ที น้อยกว่าตัวสถิติทดสอบทีเสมอ ทั้งนี้เนื่องจากตัวสถิติทดสอบทีเป็นตัวสถิติที่ดีที่สุด สำหรับการแจกแจงแบบปกติ และที่ตัวอย่างขนาดเล็กเช่นขนาดตัวอย่างเท่ากับ 5 ปรากฏว่ารีเลทีฟเอฟฟิเชียนซี ของตัวสถิติทดสอบทริมด์ ที เทียบกับตัวสถิติทดสอบ ที จะมีค่าไม่สูงนัก แต่ถ้าขนาดตัวอย่างมากกว่า 6 จะได้ค่ารีเลทีฟเอฟฟิเชียนซีสูง ซึ่งมีค่าโดยประมาณเท่ากับ $\{1-(2g)/(3n)\}\times 100$ %

สำหรับที่ภายใต้การแจกแจงแบบสเกลคอนทามิเนทด์นอร์มอล ได้ศึกษาที่เปอร์เซ็นต์คอนทามิเนทด์ 20 สำหรับสเกลแฟคเตอร์ 3 5 และ 7 และใช้ระดับการทริมด์ g ที่ 0 1, 0 1 2 และ 0 1 3 5 สำหรับขนาดตัวอย่าง 5, 10 และ 20 ตามลำดับ ได้ผลสรุปดังนี้

1. ตัวสถิติทดสอบทริมด์ ที และตัวสถิติทดสอบที มีความคลาดเคลื่อนประเภทที่ 1 ลดลงกว่าการแจกแจงแบบปกติ

2. ตัวสถิติทดสอบที ให้ความคลาดเคลื่อนประเภทที่ 1 แตกต่างจากระดับนัยสำคัญที่กำหนดมากกว่าตัวสถิติทดสอบทริมด์ ที นอกจากที่ขนาดตัวอย่างเท่ากับ 20 ที่ระดับการทริมด์ g เท่ากับ 1

3. เมื่อตัวอย่างมีขนาดเล็กและการแจกแจงไม่มีลักษณะเป็นลอง-เทลด์มากนัก ตัวสถิติทดสอบทริมด์ ที ให้ค่าอำนาจการทดสอบน้อยกว่าตัวสถิติทดสอบที แต่ที่ขนาดตัวอย่างมากกว่าหรือเท่ากับ 10 ตัวสถิติทดสอบทริมด์ ที จะให้ค่าอำนาจการทดสอบสูงกว่า

4. โดยทั่วไปการใช้ตัวสถิติทดสอบทริมด์ ที ทำให้ค่าอำนาจการทดสอบเพิ่มขึ้นกว่าการใช้ตัวสถิติทดสอบที ส่วนค่าอำนาจการทดสอบที่เพิ่มขึ้นนี้จะมากหรือน้อยขึ้นอยู่กับระดับของคอนทามิเนชั่น ขนาดตัวอย่าง ระดับการทริมด์ และความแตกต่างระหว่างค่าเฉลี่ยของประชากรทั้ง 2 ชุด

5. ภายใต้การแจกแจงแบบสเกลคอนทามิเนทด์นอร์มอล ที่สเกลแฟคเตอร์เท่ากับ 7 และขนาดตัวอย่างเป็น 20 ระดับการทริมด์ g เท่ากับ 5 ตัวสถิติทดสอบทริมด์ ที ให้ค่ารีเลทีฟเอฟฟิเชียนซี เมื่อเทียบกับตัวสถิติทดสอบที สูงถึง 5

ในกรณีที่ขนาดตัวอย่างจากประชากรทั้ง 2 ชุดไม่เท่ากัน Yuen และ Dixon ได้เสนอสูตรคำนวณค่าสถิติทดสอบทริมด์ ทีด้วย และได้ใช้วิธีการซิมูเลชั่น ศึกษาการแจกแจงของตัวสถิติที่ขนาดตัวอย่าง $n_1 = 10, n_2 = 20$ และ $n_1 = 5, n_2 = 20$ โดยใช้ระดับการทริมด์ g_1 และ g_2 เป็นอัตราส่วนเดียวกับอัตราส่วนของขนาดตัวอย่างทั้ง 2 ชุด เช่นใช้ $g_1 = 1$, $g_2 = 2$ สำหรับขนาดตัวอย่าง $n_1 = 10, n_2 = 20$ ผลจากการทำซิมูเลชั่นสรุปได้ว่าตัวสถิติทดสอบทริมด์ ที มีการแจกแจงโดยประมาณแบบที ที่องศาความเป็นอิสระ h_1+h_2-2 โดยกรณีที่ $n_1 = 10, n_2 = 20$ และที่ $n_1 = 5, n_2 = 20$ ตัวสถิติทดสอบทริมด์ ที มีลักษณะการแจกแจงที่เบี่ยงเบนจากการแจกแจงแบบทีมากที่สุดเพียง 0.259% และ 0.218% ตามลำดับเท่านั้น

Hogg(1974: 911) กล่าวว่าการเลือกระดับการทริมด์ ขึ้นกับแต่ละสถานการณ์ คือ ขึ้นกับลักษณะความมากหรือน้อยของเอาท์ลายเออร์ ตัวอย่างเช่น ภายใต้การแจกแจงที่มีระดับความเป็นลอง-เทลด์ระหว่างการแจกแจงแบบปกติและการแจกแจงแบบคอชี (Cauchy Distribution) Crow และ Siddiqui ใช้ระดับการทริมด์ $\frac{1}{4}$ หรือ $\frac{1}{3}$ ของขนาดตัวอย่าง และภายใต้การแจกแจงที่มีระดับความเป็นลอง-เทลด์ระหว่างการแจกแจงแบบปกติและการแจกแจงแบบดับเบิ้ลเอกซ์โปเนนเชียล ใช้ระดับการทริมด์ประมาณ $\frac{1}{5}$ ซึ่งการศึกษาทั้งสองกรณีนี้ผลปรากฏว่าค่าเฉลี่ยทริมด์เป็นตัวประมาณที่มีความแกร่ง Hogg ได้กล่าวต่อไปว่าการใช้วิธีการข้างต้นนี้ระดับการทริมด์จะถูกกำหนดคงที่โดยการพิจารณาล่วงหน้าก่อนที่จะทำการสุ่มตัวอย่าง ซึ่งมีนักสถิติหลายท่านเห็นว่าไม่เหมาะสม เนื่องจากในทางปฏิบัติเราไม่สามารถทราบลักษณะการแจกแจงของข้อมูลที่ได้แน่ชัด และได้แนะนำให้เลือกระดับการทริมด์ภายหลังจากได้ทำการสุ่มตัวอย่างแล้ว โดยใช้ข้อมูลตัวอย่างประกอบการพิจารณา ดังเช่น Tukey และ Mc.Laughlin ตลอดจน Jaeckel ได้ใช้วิธีเลือกระดับการทริมด์ที่ทำให้ค่าประมาณความแปรปรวนของค่าเฉลี่ยทริมด์มีค่าน้อยที่สุด และกล่าวว่าค่าเฉลี่ยทริมด์ที่ได้โดยวิธีนี้เป็นตัวประมาณที่มีความแกร่งด้วยเช่นกัน

เมื่อกำหนดระดับการทริมด์ $g = [n\alpha]$ และให้ $\hat{g}$ และ $\hat{\alpha}$ เป็นค่าที่ทำให้ค่าประมาณความแปรปรวนของค่าเฉลี่ยทริมด์มีค่าน้อยที่สุด ค่าเฉลี่ยทริมด์ที่ได้โดยวิธีนี้ $(m(\hat{\alpha}))$ เรียกว่าตัวประมาณของ Jaeckel หรือค่าเฉลี่ยออพติมอล ทริมด์ (Jaeckel's estimator or optimal trimmed mean) $m(\hat{\alpha})$ เป็นสถิติที่ขึ้นกับข้อมูลตัวอย่างและเป็นตัวสถิติอแดพทีฟ (Adaptive Statistic) จะเห็นได้ว่า $\hat{\alpha}$ ขึ้นกับค่าส่วนเบี่ยงเบนกำลังสองแบบวินเซอร์ไรซ์ (Winsorized sum of squared deviation) เป็นอย่างมาก ซึ่งทำให้

ตัวประมาณความแปรปรวนของ $m(\hat{\alpha})$ คือ $S_m^2(\hat{\alpha})$ มีค่าต่ำกว่าความเป็นจริง ผลอันนี้อาจจะเป็นข้อเสียในการใช้ $m(\hat{\alpha})$ แต่อย่างไรก็ตาม Jaeckel ได้ทำซิมูเลชั่นที่ขนาดตัวอย่างเท่ากับ 20 ที่การแจกแจงแบบปกติ แบบโลจิสติค และแบบคอซี และกำหนดช่วงของ α เป็น $0 \leq \alpha \leq [\frac{1}{4} n]$ หาค่าความแปรปรวนของ $m(\hat{\alpha})$ ที่ได้จากการทดลอง และเปรียบเทียบกับ $S_m^2(\alpha)$ คือ ค่าความแปรปรวนของค่าเฉลี่ยทริมด์ แบบที่กำหนด ระดับการทริมด์คงที่ $(m(\alpha))$ ซึ่งเป็นระดับการทริมด์ที่เหมาะที่สุดสำหรับแต่ละการแจกแจง ผลปรากฏว่าได้ค่าความแปรปรวนใกล้เคียงกัน ทั้งนี้อาจเนื่องมาจาก $S_m^2(\alpha)$ เป็นตัวประมาณความแปรปรวนของ $m(\alpha)$ จึงอาจมีผลทำให้ $S_m^2(\hat{\alpha})$ สามารถใช้เป็นตัวประมาณความแปรปรวนของ $m(\hat{\alpha})$ ได้เช่นกัน อย่างไรก็ตาม Hogg แนะนำว่าควรมีการศึกษาทั้งในเชิงทดลองและเชิงทฤษฎีเกี่ยวกับเรื่องนี้โดยละเอียดต่อไป

Hogg Fisher และ Randles (1975: 656-661) ได้ศึกษาเปรียบเทียบตัวสถิติทดสอบต่าง ๆ ในการทดสอบสมมติฐานการเท่ากันของค่าเฉลี่ยประชากร 2 ชุด ภายใต้การแจกแจงของประชากรทั้ง 2 ชุดเป็นแบบยูนิฟอร์ม แบบปกติ แบบคอซี และการแจกแจงที่มีความโด่งและความเบ้ชนิดเบ้ขวาในระดับต่าง ๆ กัน ตัวสถิติทดสอบที่ใช้ในการศึกษาได้แก่ ตัวสถิติทดสอบที ตัวสถิติทดสอบวิลค๊อกสัน ตัวสถิติทดสอบทริมด์ ที ที่ระดับการทริมด์ $\frac{1}{5}$ ของขนาดตัวอย่าง และตัวสถิติทดสอบแดพทีฟ ดิสตริบิวชั่น-ฟรี ตัวสถิติทดสอบที่ตัวสถิติทดสอบอแดพทีฟใช้ได้แก่ ตัวสถิติทดสอบวิลค๊อกสัน ตัวสถิติทดสอบแรงค์แอล (Rank L test) ตัวสถิติทดสอบแรงค์ เอส (Rank S test) และตัวสถิติทดสอบมีเดียน (Median test) สำหรับการพิจารณาว่าจะเลือกใช้ตัวสถิติทดสอบใด ได้ใช้ค่าสถิติ Q_1 และ Q_2 ซึ่งคำนวณจากข้อมูลตัวอย่างทั้ง 2 ชุดที่นำมาจัดอันดับร่วมกันเป็นตัวพิจารณาโดยค่าสถิติ Q_1 เป็นดัชนีที่ใช้วัดระดับความเป็นลอง-เทลด์หรือชอร์ท-เทลด์ และ Q_2 ใช้วัดความเบ้ชนิดเบ้ขวา

ตัวสถิติทดสอบอแดพทีฟ ดิสตริบิวชั่น-ฟรี นี้จะเลือกใช้ตัวสถิติทดสอบมีเดียนเมื่อตัวอย่างสุ่มมีลักษณะเป็นลอง-เทลด์มาก ๆ คือ $Q_2 > 7$ และสำหรับทุกค่าของ Q_1 ใช้ตัวสถิติทดสอบวิลค๊อกสัน เมื่อมีลักษณะสมมาตรและเป็นลอง-เทลด์ในระดับกลาง คือ Q_1 อยู่ระหว่าง 0 ถึง 2 และ Q_2 อยู่ระหว่าง 2 ถึง 7 และที่ชอร์ท-เทลด์แบบสมมาตร คือ Q_1 อยู่ระหว่าง 0 ถึง 2 และ Q_2 น้อยกว่าหรือเท่ากับ 2 ใช้ตัวสถิติทดสอบแรงค์ แอล สำหรับที่ลักษณะของตัวอย่างสุ่มเป็นแบบเบ้ขวา คือ $Q_1 > 2$ และ $Q_2 \leq 7$ ใช้ตัวสถิติทดสอบแรงค์ เอส

Hogg and others ได้ใช้เทคนิคมอนติคาร์โลซิมูเลชั่นในการศึกษาเปรียบเทียบตัวสถิติทดสอบดังกล่าวข้างต้น ที่ระดับนัยสำคัญ .05 โดยที่แต่ละสถานการณ์ทำซ้ำจำนวน 5,000 ครั้ง กำหนดขนาดตัวอย่างทั้ง 2 ชุดเท่ากัน คือ 15 ได้ผลสรุปเป็นดังนี้คือ

1) ตัวสถิติทดสอบทุกตัวให้ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นจริง ใกล้เคียงกับระดับนัยสำคัญที่กำหนด ยกเว้นตัวสถิติทดสอบทีและตัวสถิติทดสอบทริมด์ ที ที่การแจกแจงของประชากรเป็นแบบคอชีเท่านั้น

2) ตัวสถิติทดสอบอแดพทีฟ ให้ค่าอำนาจการทดสอบต่ำกว่าตัวสถิติทดสอบ ที เพียงเล็กน้อยที่การแจกแจงของประชากรแบบปกติ แต่จะสูงกว่าที่ทุกการแจกแจงแบบอื่นที่ศึกษา โดยเฉพาะอย่างยิ่งที่การแจกแจงชนิดลอง-เทลด์และชนิดเบ้

3) ตัวสถิติทดสอบอแดพทีฟ ให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบทริมด์ ที ที่การแจกแจงแบบสมมาตรชนิดชอร์ท-เทลด์ และที่การแจกแจงชนิดเบ้ แต่สำหรับที่การแจกแจงแบบสมมาตรชนิดลอง-เทลด์ ตัวสถิติทริมด์ ที ให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบอแดพทีฟ

4) ตัวสถิติทดสอบอแดพทีฟให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบวิลค๊อกสันมาก นอกจากที่การแจกแจงแบบสมมาตรชนิดลอง-เทลด์ ตัวสถิติทดสอบแดพทีฟจะดีกว่าเล็กน้อย

นอกจากนั้นยังได้ตั้งข้อสังเกตว่า จากการที่ตัวสถิติทดสอบทริมด์ ที ที่ระดับการทริมด์ $\frac{1}{5}$ ของขนาดตัวอย่างให้ค่าอำนาจการทดสอบที่สูง นอกจากที่การแจกแจงแบบสมมาตรชนิดชอร์ท-เทลด์ และที่การแจกแจงชนิดเบ้นั้น บางทีตัวสถิติทดสอบอแดพทีฟที่ใช้ตัวสถิติทดสอบทริมด์ ที โดยเลือกใช้ระดับการทริมด์ต่าง ๆ กัน ให้เหมาะสมกับการแจกแจงของประชากรทั้งสองลักษณะนี้ด้วย อาจทำให้ได้ตัวสถิติทดสอบอแดพทีฟ ที่ให้ค่าอำนาจการทดสอบสูงในทุกลักษณะการแจกแจงของประชากร

Fisher (Hogg 1974: 920) ได้ศึกษาเปรียบเทียบวิธีการต่าง ๆ ในการเลือกประชากรที่มีค่าเฉลี่ยของประชากรมากที่สุด จากประชากร k ชุดที่เป็นอิสระต่อกัน ซึ่งมีการแจกแจงแบบเดียวกันและมีความแปรปรวนเท่ากัน โดยการแจกแจงที่ศึกษาคือ การแจกแจงแบบยูนิฟอร์ม แบบปกติ แบบดับเบิ้ลเอกซ์โปเนนเชียล และการแจกแจงที่มีความโด่งและความเบ้ชนิดเบ้ขวาในระดับต่าง ๆ กัน Fisher ได้เลือกประชากรที่มีค่าเฉลี่ยมากที่สุดโดยเลือกประชากรชุดที่ให้ค่าสถิติจากตัวอย่างมากที่สุด ค่าสถิติที่ใช้คือ ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ $\frac{1}{4}$ ของขนาดตัวอย่าง และค่าสถิติอแดพทีฟซึ่งจะเลือกใช้ค่าเฉลี่ยทริมด์ที่ระดับการ

ทริมด์ต่าง ๆ กัน ตามแต่ลักษณะของข้อมูลตัวอย่าง

การเลือกใช้ระดับการทริมด์ของค่าเฉลี่ยทริมด์สำหรับค่าสถิติอแดพทีฟนั้น พิจารณาจากค่าสถิติ 2 ตัว คือ Q_1 และ Q_2 และเนื่องจากเป็นการเปรียบเทียบประชากรจำนวน k ชุด ดังนั้นจึงมีค่า Q_1 และ Q_2 อย่างละ จำนวน k ค่า Fisher ได้ใช้ค่าเฉลี่ยของ Q_1 และ Q_2 คือ $\bar{Q}_1$ และ $\bar{Q}_2$ เป็นเกณฑ์ในการพิจารณาเลือกระดับการทริมด์ให้เหมาะสมกับลักษณะของข้อมูล กล่าวคือ ถ้าหากว่าค่า $\bar{Q}_1$ และ $\bar{Q}_2$ แสดงว่าข้อมูลมีลักษณะเบ้ขวา จะทริมด์เฉพาะค่าตัวอย่างที่มีค่ามาก ถ้าข้อมูลมีลักษณะสมมาตรแบบช่อร์ท-เทลด์ จะใช้ค่าสถิติ m^c (.05) คือค่าเฉลี่ยที่ได้จากการนำเฉพาะ 5% ของข้อมูลที่มีค่าน้อยและมากที่สุดมาหาค่าเฉลี่ย ถ้าข้อมูลเป็นแบบปกติจะใช้ค่าเฉลี่ยเลขคณิต และหากค่าสถิติ $\bar{Q}_1$ และ $\bar{Q}_2$ แสดงว่าข้อมูลมีลักษณะสมมาตรชนิดลอง-เทลด์ จะใช้ค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ $\frac{1}{4}$ ของขนาดตัวอย่าง เป็นตัวพิจารณาเลือกประชากรที่มีค่าเฉลี่ยประชากรมากที่สุด

จากการศึกษาเปรียบเทียบการใช้ค่าสถิติ 3 ตัว คือ ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยทริมด์ที่ระดับการทริมด์ $\frac{1}{4}$ ของขนาดตัวอย่าง และค่าสถิติอแดพทีฟ ได้ผลโดยสรุป คือ ค่าสถิติอแดพทีฟให้ผลการเลือกประชากรได้อย่างถูกต้องมากกว่าค่าสถิติอีก 2 ตัวมาก ในทุกการแจกแจงของประชากร นอกจากที่การแจกแจงแบบปกติค่าสถิติอแดทีฟจะให้ความถูกต้องในการเลือกประชากรน้อยกว่าค่าเฉลี่ยเลขคณิตเล็กน้อย

Gibbon (1971: 149) ได้กล่าวถึงตัวสถิติทดสอบแมน-วิทนีย์ ยู ว่า เป็นตัวสถิติทดสอบที่มีความไว (Sensitivity) ต่อการทดสอบความแตกต่างระหว่างตำแหน่งของประชากร 2 ชุดที่เป็นอิสระต่อกัน ในกรณีที่มีข้อตกลงเบื้องต้นว่า รูปแบบการแจกแจงและการกระจายของประชากรทั้ง 2 ชุดเหมือนกันนั้น ตัวสถิติทดสอบแบบแมน-วิทนีย์ ยู สามารถใช้แทนตัวสถิติทดสอบที ได้ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างของค่าเฉลี่ยประชากร 2 ชุด หากการแจกแจงของประชากรไม่เป็นแบบปกติ ค่าแอสซิมโทติครีเลทีฟเอฟฟิเซียนซี ของตัวสถิติทดสอบแมน-วิทนีย์ ยู เมื่อเทียบกับตัวสถิติทดสอบที ไม่ว่าภายใต้การแจกแจงของประชากรแบบใด ๆ จะไม่ต่ำกว่า 0.864 และสามารถมีค่าสูงได้ถึงอนันต์ และ Gibbon (1971: 168) ได้กล่าวถึงตัวสถิติทดสอบแวน เดอ แวร์เดน ว่าเป็นตัวสถิติทดสอบที่มีค่าแอสซิมโทติครีเลทีฟเอฟฟิเซียนซี มากกว่าหรือเท่ากับ 1 เสมอ