ปัญหาเกี่ยวกับค่าเฉลี่ย

เรือนที่มุงดี ฝนรั่วไหลเข้าไปไม่ได้
ใจที่อบรมดีแล้ว ราคะก็รั่วไหลเข้าครอบงำไม่ได้เช่นกัน

พุทธวจนะ ธรรมบท

พารามิเตอร์ค่าเฉลี่ย หรือค่ากลาง (Average or Location Parameter) ที่เราสนใจกันบ่อยๆ ก็คือค่าเฉลี่ยเลขคณิต (μ) และมัธยฐาน (M) สำหรับค่าเฉลี่ยเลขคณิตนั้นเป็นมาตราวัดแนวโน้มสู่ส่วนกลาง (Meaure of Central Tendency) ที่กระบวนการอ้างอิงเชิงพารามิเตอร์จะเกี่ยวข้องด้วย ส่วนกระบวนการแบบไร้พารามิเตอร์ที่อ้างอิงเกี่ยวกับมาตราวัดแนวโน้มสู่ส่วนกลาง หรือพารามิเตอร์ค่ากลางนั้นจะเกี่ยวข้องกับมัธยฐานมากกว่าค่าเฉลี่ยเลขคณิต แต่อย่างไรก็ตาม ถ้าประชากรสมมาตร (Symmetric) แล้วการสรุปผลเกี่ยวกับมัธยฐานจะใช้ได้กับค่าเฉลี่ยเลขคณิตด้วย ทั้งนี้เพราะประชากรที่สมมาตรนั้นจะมีค่าเฉลี่ยและมัธยฐานเป็นค่าเดียวกัน ในการอ้างอิงหรืออนุมานเกี่ยวกับพารามิเตอร์ค่าเฉลี่ยของประชากร เราจะขอกล่าวเป็น 5 กรณี ดังรายละเอียดต่อไปนี้

3.1 กรณีตัวอย่างเดียว (One-Sample Case)

การอ้างอิงเกี่ยวกับค่าเฉลี่ยซึ่งเป็นการทดสอบ และประมาณค่าเฉลี่ยนั้นเรามีแบบทดสอบและประมาณค่าดังต่อไปนี้ สำหรับกระบวนการในเชิงพารามิเตอร์ เราก็อาศัยการแจกแจงแบบที (Student t)

3.1.1 แบบทดสอบทวินาม (Binomial Test)

แบบทดสอบทวินามนั้นนอกจากจะใช้ทดสอบสมมติฐานเกี่ยวกับสัดส่วนเปอร์เซ็นต์ หรือความน่าจะเป็น แล้วเรายังใช้ทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของตัวแปรเชิงสุ่มทวินามได้ ทั้งนี้ก็เพราะสมมติฐานหลักที่เกี่ยวกับสัดส่วน หรือ

$$Ho : \pi = \pi_0$$

นั้นสมนัยกับสมมติฐานหลักที่เกี่ยวกับค่าเฉลี่ย คือ

$$Ho : n\,\pi = n\,\pi_0 \text{ หรือ } Ho : \mu = \mu_0$$

ตัวสถิติทดสอบสำหรับสมมติฐานหลักดังกล่าวก็คือ ตัวสถิติ T ซึ่งเป็นจำนวนหน่วยตัวอย่างที่มีลักษณะดังที่สนใจจากตัวอย่างขนาด n นั่นเอง ตัวสถิติ T นี้มีการแจกแจงทวินามที่มีค่าเฉลี่ย และความแปรปรวนเป็นดังนี้ ภายใต้สมมติฐานหลัก Ho เป็นจริง

$$E(T) = \mu_0 = n\,\pi_0$$
$$V(T) = \sigma_0^2 = n\,\pi_0\,(1-\pi_0)$$

ดังนั้นในการกำหนดเขตวิกฤต เราจึงอาศัยการแจกแจงทวินาม ที่มีค่าเฉลี่ย และความแปรปรวนดังกล่าว

สำหรับตัวอย่างขนาดโต ($n \geq 20$) เราใช้ตัวสถิติทดสอบ Z

$$Z = \frac{T - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$$

ซึ่งมีการแจกแจงปกติมาตรฐาน N (0, 1) นั่นคือในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย โดยอาศัยตัวอย่างขนาดโต เราก็ใช้ตัวสถิติ Z นี้นั่นเอง

ในการประมาณค่าเฉลี่ยแบบช่วง เราก็อาศัยช่วงเชื่อมั่น $100(1-\alpha)\%$ สำหรับสัดส่วนแล้วคูณด้วยขนาดตัวอย่าง n ซึ่งจะได้ช่วงเชื่อมั่น $100(1-\alpha)\%$ สำหรับค่าเฉลี่ยตามต้องการ

ตัวอย่าง ในการทดสอบคำกล่าวที่ว่า มัธยฐานรายได้ต่อครัวเรือนของกรรมกรในโรงงานทอผ้าจะประมาณ 2,000 บาท ได้อาศัยตัวอย่างของครัวเรือนกรรมกร 100 ราย ปรากฏว่ามีรายได้สูงกว่ามัธยฐานดังกล่าวจำนวน 23 ราย คำกล่าวน่าจะเชื่อถือได้หรือไม่ ?

$Ho : \pi = 0.5$ หรือ $Ho : \mu = 100\ (0.5) = 50$

$Ha : \pi \neq 0.5$ หรือ $Ha : \mu = \neq 50$

เมื่อ T = 23 เราได้ค่าของตัวสถิติทดสอบ Z เป็น

$$Z = \frac{23-100(0.5)}{\sqrt{100(0.5)\ (1-0.5)}} = -5.4$$

จากตารางปกติเราได้ $-Z_{.025} = -1.96$ จึงปฏิเสธ Ho นั่นคือคำกล่าวไม่น่าเชื่อถือ จากข้อมูลนี้สามารถสรุปได้ว่า มัธยฐานรายได้ต่อครัวเรือนน่าจะน้อยกว่า 2,000 บาท

3.1.2 แบบทดสอบเครื่องหมายชนิดตัวอย่างเดียว (One Sample Sign Test, Fisher 1925)

แบบทดสอบเครื่องหมายเป็นกระบวนการไร้พารามิเตอร์ที่เก่าแก่มาก ได้ใช้กันมาแต่ปี 1925 อันที่จริงแบบทดสอบเครื่องหมายนี้ก็เป็นแบบทดสอบทวินามที่มี $\pi = 0.50$ นั่นเอง การที่เรียกว่า แบบทดสอบเครื่องหมายก็เพราะว่าข้อมูลที่จะวิเคราะห์นั้นสามารถแปลงให้อยู่ในอนุกรมของเครื่องหมาย (+) และ (-) ได้ ตัวสถิติทดสอบจึงขึ้นอยู่กับเครื่องหมายเหล่านั้นนั่นเอง

ในการทดสอบเกี่ยวกับมัธยฐานที่กำหนดไว้ (M_0) นั้น เรามีแบบฟอร์มของสมมติฐานที่จะทดสอบเป็นแบบหนึ่งใน 3 แบบ ดังนี้

(ก) $Ho : M = M_0$; $Ha : M \neq M_0$

(ข) $Ho : M \leq M_0$; $Ha : M > M_0$

(ค) $Ho : M \geq M_0$; $Ha : M < M_0$

การทดสอบสมมติฐานดังกล่าวนี้เราก็อาศัยตัวอย่างขนาด n จากประชากรที่ไม่ทราบมัธยฐาน (M) ของตัวแปรที่สนใจ ค่าสังเกตของตัวอย่างนี้อย่างน้อยต้องเป็นแบบอันดับ และเป็นแบบต่อเนื่อง

ให้ค่าสังเกตจากตัวอย่างเป็น $X_1, X_2, .., X_n$ กำหนดตัวแปร Z_i ซึ่งเป็นผลต่างระหว่าง X_i กับ M_o นั่นคือ

$$Z_i = X_i - M_o \text{ , } i = 1, 2, .., n$$

และกำหนดตัวแปรดัชนี (Indicator Variables) ไว้เป็น φ_i

φ_i = 0 ถ้า Z_i เป็นลบ ($Z_i < 0$)

= 1 ถ้า Z_i เป็นบวก ($Z_i > 0$)

แล้วตัวสถิติทดสอบสำหรับสมมติฐานหลัก Ho จะเป็น r

$$T = \sum_{i}^{n} \varphi_i$$

ตัวสถิติทดสอบ T ก็คือจำนวนเครื่องหมาย + นั่นเอง

กำหนด n_o เป็นจำนวนค่า X_i ที่ไม่เท่ากับ M_o หรือ n_o เป็นจำนวนเครื่องหมาย + และ – ถ้า X_i เท่ากับ M_o หรือ $Z_i = 0$ ก็ให้ตัดค่า X_i นั้นจากการวิเคราะห์

ภายใต้สมมติฐานหลัก Ho ตัวสถิติทดสอบ T จะมีการแจกแจงทวินาม ที่มีพารามิเตอร์ n_o และ $\pi_o = 0.50$ นั่นคือตัวสถิติ T มีค่าเฉลี่ย และความแปรปรวนเป็น

$$E(T) = n_o (0.50) = n_o / 2$$

$$V(T) = n_o (0.50)(0.50) = n_o / 4$$

เกณฑ์ตัดสินใจสำหรับสมมติฐานต่าง ๆ จึงกำหนดไว้ว่า

(ก) สำหรับ Ho : $M = M_o$ และ Ha : $M \neq M_o$ นั้น ถ้าตัวสถิติ T มีค่ามากหรือน้อยเกินไป จะแสดงว่า Ho ไม่น่าเป็นจริง นั่นคือ Ha : $M \neq M_o$ น่าจะเป็นจริง ดังนั้นจึงปฏิเสธ Ha ณ ระดับนัยสำคัญ $\alpha = 2\alpha_1$ ถ้า $T \leq t$ หรือ $T \geq n_o - t$ ในเมื่อ t เป็นค่าของตัวแปรเชิงสุ่มทวินาม Y ที่มีพารามิเตอร์ n_o และ $\pi_o = 0.50$ และทำให้ $P(Y \leq t) = \alpha_1 \approx \alpha/2$

(ข) สำหรับ Ho : $M \leq M_o$ และ Ha : $M > M_o$ นั้น ถ้าตัวสถิติ T มีค่ามากเกินไปก็จะแสดงว่า Ho ไม่น่าจะเป็นจริง นั่นคือ Ha น่าจะเป็นจริง ดังนั้นจึงปฏิเสธ Ho ณ ระดับนัยสำคัญ $\alpha = \alpha_1$ ถ้า $T \geq n_o - t$ ในเมื่อ t เป็นค่าจากตารางทวินามที่มี n_o และ $\pi_o = 0.50$ และทำให้ $P(Y \leq t) = \alpha_1 \approx \alpha$

(ค) สำหรับ Ho : $M \geq M_o$ และ Ha : $M < M_o$ นั้น ถ้าตัวสถิติ T มีค่าน้อยเกินไปก็จะแสดงว่า Ho ไม่น่าเป็นจริง นั่นคือ Ha น่าจะเป็นจริง ดังนั้นจึงปฏิเสธ Ho ณ ระดับนัยสำคัญ $\alpha = \alpha_1$ ถ้า $T \leq t$ ในเมื่อ t เป็นค่าจากตารางทวินามที่มี n_o และ $\pi_o = 0.50$ และทำให้ $P(Y \leq t) = \alpha_1 = \alpha$

สำหรับตัวอย่างขนาดโต ($N_o \geq 20$) เราใช้ตัวสถิติทดสอบ Z ซึ่งมีการแจกแจงปกติมาตรฐาน N(0,1)

$$Z = \frac{(T \pm 0.50) - n_o / 2}{\sqrt{n_o / 4}}$$

ตัวอย่าง ในการทดสอบสมมติฐานที่ว่า มัธยฐานสำหรับระดับการศึกษาของหัวหน้าครอบครัวในหมู่บ้านรามนิเวศน์จะมากกว่า 11.8 ปี จากการศึกษาโดยอาศัยตัวอย่างขนาด 50 ครอบครัว ได้ข้อมูลมาดังนี้

จำนวนปีที่ใช้ศึกษา	<8	8	9	10	11	12	13	14	15	16	>16
จำนวน	2	5	2	6	5	7	2	6	1	11	3

สมมติฐานที่จะทดสอบ จะเป็น

$$Ho : M \leq 11.8;\ Ha : M > 11.8$$

ค่าของตัวสถิติทดสอบ T จะเป็น

$$T = 7 + 2 + 6 + 1 + 11 + 3 = 30$$

เมื่อ Ho เป็นจริง เราได้

$$E(T) = n_o / 2 = 50/2 = 25$$

$$V(T) = n_o / 4 = 50/4 = 12.5$$

$$Z = \frac{(30-0.5)-25}{\sqrt{12.5}} = 1.27$$

เนื่องจาก $Z_{.05} = 1.645$ เราจึงปฏิเสธ Ho ไม่ได้ นั่นคือระดับการศึกษาของหัวหน้าครอบครัวในหมู่บ้านรามนิเวศน์จะไม่มากกว่า 11.8 ปี

แบบทดสอบเครื่องหมายนี้ยังใช้ทดสอบการเท่ากันของสองมัธยฐานที่มีสหสัมพันธ์กัน ทดสอบแนวโน้ม และทดสอบสหสัมพันธ์อีกด้วย ซึ่งจะได้กล่าวต่อไป

3.1.3 ช่วงเชื่อมั่นของมัธยฐานประชากร ที่ขึ้นอยู่กับแบบทดสอบเครื่องหมาย (Confidence Interval for Population Median based on Sign Test)

ธอมป์สัน (Thompson, 1936) และเซฝัว (Savua, 1937) ได้เสนอวิธีการประมาณช่วงเชื่อมั่นสำหรับมัธยฐานประชากรไว้ดังนี้

จากตัวอย่างขนาด n ซึ่งได้ค่าสังเกตที่เรียงตามขนาดแล้วเป็น $X_{(1)} \leq X_{(2)} \leq .. \leq X_{(n)}$ นั้น เราจะได้มัธยฐานตัวอย่าง ($\hat{M}$) ดังนี้

$$\hat{M} = \begin{cases} X_{(k+1)}, & k = (n-1)/2 \text{ ถ้า } n \text{ เป็นเลขคี่} \\ (X_{(k)} + X_{(k+1)})/2, & k = n/2 \text{ ถ้า } n \text{ เป็นเลขคู่} \end{cases}$$

ในการประมาณมัธยฐานประชากรแบบช่วง เราพัฒนาจากการทดสอบสมมติฐานสองทาง นั่นคือสมมติฐานที่จะทดสอบเป็น

$$Ho : M = M_0 ;\ Ha : M \neq M_0$$

โดยมีเกณฑ์ตัดสินใจว่า จะปฏิเสธ Ho ถ้า $T \leq t$ หรือ $T \geq n-t$ ในเมื่อ t และ n–t เป็นจุดสมมาตร

ของการแจกแจงทวินามที่มีพารามิเตอร์ n และ $\pi = 0.50$ ช่วงเชื่อมั่นสำหรับมัธยฐานประชากรที่สร้างขึ้นนั้นก็อาศัยค่า t และ n–t นั่นเอง ซึ่งเราจะได้ช่วงเชื่อมั่นสำหรับมัธยฐานประชากรเป็นดังนี้

$$X_{(t+1)} \le M \le X_{(n-t)}$$

ช่วงนี้จะให้สัมประสิทธิ์เชื่อมั่นเท่ากับหรือมากกว่า $1-\alpha$ นั่นคือ

$$P\{X_{(t+1)} \le M \le X_{(n-t)}\} = 1-\alpha$$

กรณีตัวอย่างขนาดโต เราได้ค่าประมาณของ t และ n–t ดังนี้

$$t = (n/2 - 1/2) - Z_{\alpha/2}\sqrt{n/4}$$

$$n-t = (n/2 + 1/2) + Z_{\alpha/2}\sqrt{n/4}$$

ตัวอย่าง ในการประมาณมัธยฐานประชากรโดยอาศัยตัวอย่างขนาด 25 เราได้ข้อมูลซึ่งเรียงตามขนาดแล้วเป็นดังนี้

	8.7	8.9	11.0	11.1	11.3	14.3	14.8	16.5
16.8	17.4	20.5	21.5	21.8	22.2	23.0	24.3	24.4
25.2	26.8	30.6	30.7	31.6	33.2	40.1	41.2	

สำหรับ $\alpha = 0.05$ หรือ $1 - \alpha = 0.95$ เราได้ค่า t และ n–t จากตารางทวินามที่มี n = 25 และ π = 0.05 เป็น t = 7 และ n–t = 25–7 = 18 ดังนั้นช่วงเชื่อมั่น 95% สำหรับมัธยฐานประชากร M จะเป็น

$$X_{(8)} \le M \le X_{(18)} \text{ หรือ } 16.5 \le M \le 25.2$$

ในเมื่อ $X_{(8)} = 16.5$ และ $X_{(18)} = 25.2$

เมื่อใช้การประมาณค่าโดยอาศัยตัวอย่างขนาดโต เราได้

$$t = (25/2-1/2) - 1.96\sqrt{25/4} = 7.1 = 7$$

$$n-t = (25/2+1/2) + 1.96\sqrt{25/4} = 179 = 18$$

แล้วเราจะได้ช่วงเชื่อมั่น 95% สำหรับ M เช่นเดียวกันดังนี้

$$X_{(8)} \le M \le X_{(18)} \text{ หรือ } 16.5 \le M \le 25.2$$

3.1.4 แบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย (One Sample Wilcoxon Signed-Rank Test)

ในการทดสอบสมมติฐานเกี่ยวกับมัธยฐานประชากรที่ระบุไว้ (M_0) นั้น เรายังมีแบบทดสอบที่วิลคอกซัน (Wilcoxon, 1945) เสนอไว้ซึ่งอาศัยอันดับที่ (Rank) แบบทดสอบนี้ใช้ข้อมูลข่าวสารมากกว่าแบบทดสอบเครื่องหมาย จึงมีอำนาจทดสอบมากกว่า แบบทดสอบวิลคอกซันนี้มีข้อกำหนดเกี่ยวกับประชากรว่าจะต้องเป็นแบบสมมาตร ดังนั้นการสรุปผลเกี่ยวกับมัธยฐานจึงใช้ได้กับค่าเฉลี่ยเลขคณิต (μ) ด้วย

สมมติฐานเกี่ยวกับมัธยฐานที่จะทดสอบจะเป็นแบบหนึ่งแบบใดใน 3 แบบต่อไปนี้

$Ho : M = M_0$; $Ha : M \neq M_0$

$$Ho : M \geq M_0 \; ; \; Ha : M < M_0$$
$$Ho : M \leq M_0 \; ; \; Ha : M > M_0$$

ตัวอย่างสุ่มจากประชากรที่ใช้สรุปผลเกี่ยวกับสมมติฐานเหล่านี้จะต้องให้ค่าสังเกตที่เป็นอิสระกัน สเกลการวัดในตัวแปรที่สนใจซึ่งเป็นแบบต่อเนื่องนั้นอย่างน้อยต้องเป็นแบบช่วง และตัวอย่างนั้นจะสุ่มมาจากประชากรแบบสมมาตรที่ไม่ทราบมัธยฐาน (M)

จากตัวอย่างสุ่ม $X_1, X_2, .., X_n$ นั้นเรากำหนดผลต่าง Z_i เป็นดังนี้

$$Z_i = X_i - M_o \; , \; i = 1, 2, .., n$$

(ถ้า $Z = 0$ หรือ $X_i = M_o$ ให้ตัดทิ้ง X_i นั้นเสีย) เรียงลำดับค่าสัมบูรณ์ Z_i จากน้อยไปมาก และให้ R_i แทนอันดับที่ของ Z_i นั้น กำหนดตัวแปรดัชนี φ_i ไว้ดังนี้

$$\varphi_i = 0 \text{ ถ้า } Z_i < 0$$
$$= 1 \text{ ถ้า } Z_i > 0$$

แล้วตัวสถิติทดสอบจะกำหนดให้เป็น T

$$T = \Sigma R_i \varphi_i$$

ซึ่งตัวสถิติ T นี้จะเป็นผลรวมของอันดับที่เครื่องหมายบวก (Positive Signed Ranks) ของ Z_i นั่นเอง ตัวสถิติ T นี้มีค่าเฉลี่ย และความแปรปรวนดังนี้

$$E(T) = n_o (n_o + 1) / 4$$
$$V(T) = n_0 (n_o + 1) (2n_o + 1) / 24$$

เมื่อใช้ตัวอย่างขนาดโต เราก็อาศัยตัวสถิติ Z ซึ่งมีการแจกแจงปกติมาตรฐาน N(0,1)

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

กฎตัดสินใจสำหรับสมมติฐานหลัก Ho จะกำหนดไว้ตามสมมติฐานรอง Ha ดังนี้

(ก) สำหรับ $Ha : M \neq M_o$ จะปฏิเสธ Ho ถ้าตัวสถิติ T มีค่ามากหรือน้อยเกินไป ดังนั้นจะปฏิเสธ Ho ณ ระดับนัยสำคัญ α ถ้า $T > W_{1-\alpha/2}$ หรือ $T < W_{\alpha/2}$ ในเมื่อ W_p เป็นควอนไลท์ที่ p ของตัวสถิติ T ซึ่งพิจารณาได้จากตาราง 10

(ข) สำหรับ $Ha : M < M_o$ จะปฏิเสธ $Ha : M \geq M_o$ ถ้าตัวสถิติ T มีค่าน้อยเกินไป ดังนั้น ณ ระดับนัยสำคัญ α จะปฏิเสธ Ho ถ้า $T < W_\alpha$

(ค) สำหรับ $Ha : M > M_o$ จะปฏิเสธ $Ho : M \leq M_o$ ถ้าตัวสถิติ T มีค่ามากเกินไป นั่นคือ ณ ระดับนัยสำคัญ α จะปฏิเสธ Ho ถ้า $T > W_{1-\alpha}$

สำหรับกรณีตัวอย่างขนาดโต เราประมาณ W_p ได้จาก

$$W_p = E(T) + Z_p\sqrt{V(T)}$$

ในเมื่อ Z_p เป็นควอนไทล์ที่ p ของการแจกแจงปกติมาตรฐาน N(0,1)

ในกรณีที่ $|Z_i|$ มีค่าเท่ากัน จะต้องกำหนดอันดับที่เฉลี่ยให้แก่ค่าที่เท่ากันเหล่านั้น และความแปรปรวนของตัวสถิติ T จะเป็น

$$V(T) = n_o(n_o+1)(2n_o+)/24 + (\Sigma g^3 - \Sigma g)/48$$

ในเมื่อ g เป็นจำนวนของ $|Z_i|$ ที่เท่ากัน

ตัวอย่าง ในการศึกษาเพื่อทดสอบคำกล่าวที่ว่า มัธยฐานของจำนวนคำที่ตอบถูกจากคำที่กำหนดให้ 30 คำ จะประมาณ 15 คำ นั้นได้อาศัยตัวอย่างของนักศึกษา 10 คน ได้ข้อมูลซึ่งเป็นจำนวนคำที่ตอบถูก ดังนี้

13 14 20 28 21 24 23 11 29 22

จากข้อมูลเหล่านี้เราได้ $Z_i = Y_i - 15$ และอันดับที่ของ $|Z_i|$ เป็นดังนี้

Z_i	−2	−1	5	13	6	9	8	−4	14	7
ψ_i	0	0	1	1	1	1	1	0	1	1
R_i	2	1	4	9	5	8	7	3	10	6

ดังนั้นค่าของตัวสถิติทดสอบ T จะเป็น

$$T = 2(0) + 1(10) + 4(1) + 9(1) + \ldots + 6(1)$$
$$= 4 + 9 + 5 + 8 + 7 + 10 + 6 = 49$$

สมมติฐานที่ทดสอบจะเป็น

$$Ho : M = 15 \text{ , } Ha : M \neq 15$$

เกณฑ์ตัดสินใจกำหนดไว้ว่า จะปฏิเสธ Ho ณ ระดับนัยสำคัญ 0.05 ถ้า $T < W_{.025} = 9$ และ $T > W_{.975} = 10(10+1)/2 - W_{.025} = 55-9 = 46$

เนื่องจากค่าของตัวสถิติทดสอบ T เป็น 49 ซึ่งตกอยู่ในเขตวิกฤต จึงปฏิเสธ Ho แสดงว่ามัธยฐานของจำนวนคำที่ตอบถูกจะมากกว่า 15 คำ

3.1.5 ช่วงเชื่อมั่นของมัธยฐานประชากรที่ขึ้นอยู่กับแบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย

ฮอดจ์-เลฮ์แมน (Hodges-Lehman, 1963) ได้เสนอวิธีประมาณมัธยฐานประชากร (M) ไว้ดังนี้

(1) จากค่าสังเกตตัวอย่าง $Y_1, Y_2, .., Y_n$ เราหาค่าเฉลี่ยของสองค่า X_i และ X_j ใด ๆ ซึ่งจะให้เป็น U_{ij} นั่นคือ U_{ij} จะกำหนดไว้เป็น

$$U_{ij} = (Y_i + Y_j)/2, \quad 1 \le i \le j \le n$$

(2) เรียงลำดับ U_{ij} จากน้อยไปมาก ซึ่งจะให้เป็น $U_{(1)} \le U_{(2)} \le \ldots \le U_{(k)}$; $k = n(n-1)/2 + n = n(n+1)/2$

ค่าประมาณแบบจุดของมัธยฐานประชากรจะเป็นมัธยฐานของ U_{ij} นั่นเอง ดังนั้น

$$\hat{M} = \text{มัธยฐาน}\{(Y_i - Y_j)/2 \; ; \quad 1 \le i \le j \le n\}$$

ทูกี (Tukey, 1949) ได้เสนอวิธีหาช่วงเชื่อมั่นสำหรับมัธยฐานประชากรจากค่า U_{ij} ต่างๆ ไว้ดังนี้ - จากตาราง 10 ค่าเฉลี่ย U_{ij} ที่น้อยสุดอันดับที่ $W_{\alpha/2}$ จะเป็นค่าจำกัดต่ำสุดของช่วงเชื่อมั่น ดังนั้นช่วงเชื่อมั่น $100(1-\alpha)\%$ และค่าเฉลี่ยมากที่สุดที่ $W_{\alpha/2}$ จะเป็นค่าจำกัดสูงสุดของช่วงเชื่อมั่น $100(1-\alpha)\%$สำหรับมัธยฐานประชากร M จะเป็น

ในเมื่อ $a = W_{\alpha/2}$ และ $b = k + 1 - W_{\alpha/2}$; $k = n(n+1)/2$

ตัวอย่าง จากตัวอย่างเกี่ยวกับจำนวนคำที่ตอบถูกในตัวอย่างที่แล้ว เราได้ค่า U_{ij} ที่เป็นไปได้ดังนี้

	Y_j	11	13	14	20	1	22	23	24	28	29
Y_i	11	11	12	12.5	15.5	16	16.5	17	17.5	19.5	20
	13		13	13.5	16.5	17	17.5	18	18.5	20.5	21
	14			14	17	17.5	18	18.5	19	21	21.5
	20				20	20.5	21	21.5	22	24	24.5
	21					21	21.5	22	22.5	24.5	25
	22						22	22.5	23	25	25.5
	23							23	23.5	25.5	26
	24								24	26	26.5
	28									28	28.5
	29										29

จากตาราง 10 เรามีขนาดตัวอย่าง n = 10 จึงได้ $W_{\alpha/2}$ สำหรับ $\alpha = 0.05$ เป็น $W_{.025} = 9$ ดังนั้นตัวน้อยที่สุดและมากที่สุดอันดับที่ 9 จะให้ช่วงเชื่อมั่น 95% สำหรับมัธยฐานประชากร นั่นคือช่วงเชื่อมั่น 95% สำหรับมัธยฐาน M จะเป็น

$$U_{(9)} \leq M \leq U_{(47)}$$

$$16.5 \leq M \leq 25$$

ในเมื่อ $k = 10\,(10+1)/2 = 5$ และ $b=k+1-W_{.025} = 47$

3.1.6 แบบทดสอบควอนไทล์หรือแฟรคไทล์ (Quantile or Fractile Test)

แบบทดสอบทวินามนั้นยังใช้ทดสอบสมมติฐานเกี่ยวกับควอนไทล์ของประชากรหรือตัวแปรเชิงสุ่มได้ แต่จะเรียกว่า แบบทดสอบควอนไทล์ ควอนไทล์เป็นค่าเฉลี่ยอย่างหนึ่ง แต่เป็นค่าเฉลี่ยที่บอกตำแหน่ง (Position Average) ควอร์ไทล์ เดไซล์ เป็นเปอร์เซ็นต์ไทล์ก็เป็นควอนไทล์แบบหนึ่งนั่นเอง ควอนไทล์ของประชากร X กำหนดไว้ว่า

X_π จะเป็นควอนไทล์ที่ π $(0 < \pi < 1)$ ของประชากร X ถ้า

$P(X < X_\pi) \leq \pi$ และ $P(X > X_\pi) = 1 - \pi$

หรือ $P(X \leq X_\pi) \geq \pi$ และ $P(X \geq X_\pi) = 1 - \pi$

กรณีที่ประชากร X เป็นแบบต่อเนื่อง แล้วเราจะใช้การเท่ากัน (=) แทนการไม่เท่ากัน (≤ , ≥) นั่นคือ

$$P(X \le X_\pi) = \pi \text{ และ } P(X \ge X_\pi) = 1-\pi$$

ให้ $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มีค่าสังเกตเป็นแบบอันดับหรือสูงกว่า และให้ X_o เป็นค่าที่ระบุไว้ของควอนไทล์ที่ π_0 $(0 < \pi_0 < 1)$ แล้วสมมติฐานที่จะทดสอบจะเป็นดังนี้

(ก) Ho : ควอนไทล์ที่ π_0 ของประชากร เป็น x_o

Ha : ควอนไทล์ที่ π_0 ของประชากรไม่ใช่ x_o

หรือ Ho : $P(X \le X_o) \ge \pi_0$ และ $P(X < X_o) \le 1-\pi_0$

Ha : $P(X \le X_o) < \pi_0$ และ $P(X < X_o) > 1-\pi_0$

(ข) Ho : ควอนไทล์ที่ π_0 ของประชากรอย่างน้อยเท่ากับ x_o

Ha : ควอนไทล์ที่ π_0 ของประชากรน้อยกว่า x_o

หรือ Ho : $P(X < X_o) \le \pi_0$ และ Ha : $P(X < X_o) > \pi_0$

(ค) Ho : ควอนไทล์ที่ π_0 ของประชากรไม่มากกว่า x_o

Ha : ควอนไทล์ที่ π_0 ของประชากรมากกว่า x_o

หรือ Ho : $P(X \le X_o) \ge \pi_0$ และ Ha : $P(X \le X_o) < \pi_0$

ตัวสถิติทดสอบสำหรับสมมติฐาน กำหนดไว้เป็น T_1 หรือ T_2

T_1 = จำนวนค่าสังเกตที่น้อยกว่าหรือเท่ากับ X_o

T_2 = จำนวนค่าสังเกตที่น้อยกว่า X_o

โดยทั่วไป T_1 มากกว่า T_2 แต่จะเท่ากัน ถ้าจำนวนค่าสังเกตที่เท่ากับ X_o ไม่มีเลย

ถ้าสมมติฐานหลัก Ho เป็นจริง ตัวสถิติทดสอบ T_1 หรือ T_2 จะมีการแจกแจงทวินามที่มีพารามิเตอร์ n และ π_0 ดังนั้นกฎตัดสินใจจึงกำหนดไว้ดังนี้

(1) สำหรับสมมติฐาน (ก) ถ้า T_2 โตเกินไป จะแสดงว่า $P(X < X_o) > \pi_0$ หรือค่าของ T_1 น้อยเกินไป จะแสดงว่า $P(X \le X_o) < \pi_0$ ดังนั้นสมมติฐานหลัก Ho จะได้รับการปฏิเสธ ถ้า $T_1 \le t_1$ หรือ $T_2 > t_2$ ณ ระดับนัยสำคัญ $\alpha_1 + \alpha_2 \cong \alpha$ ในเมื่อ α_1 และ α_2 กำหนดไว้ว่า

$$P(Y \le t_1) = \alpha_1 \cong \alpha/2$$

$$P(Y > t_2) = \alpha_2 \cong \alpha/2 \text{ หรือ } P(Y \le t_2) = 1-\alpha_2$$

โดยที่ Y เป็นตัวแปรเชิงสุ่มทวินาม ที่มีพารามิเตอร์ n และ π_0

(2) สำหรับสมมติฐาน (ข) ถ้าค่าของตัวสถิติ T_2 มีค่ามาก จะแสดงว่า Ho เป็นเท็จ ค่าวิกฤต t_2 จะพิจารณาจากตารางทวินาม ที่มีพารามิเตอร์ n และ π_0 ดังนี้

$$P(Y > t_2) = \alpha \text{ หรือ } P(Y \le t_2) = 1-\alpha$$

ดังนั้นจะปฏิเสธ Ho ถ้า $T_2 > t_2$ ณ ระดับนัยสำคัญ

(3) สำหรับสมมติฐาน (ค) ถ้าค่าของตัวสถิติ T_1 มีค่าน้อยไป จะแสดงว่า Ho ไม่เป็นจริง จากตารางทวินามที่มีพารามิเตอร์ n และ π_0 ตามที่ระบุไว้ เราได้ค่าวิกฤต t_1 จาก

$$P(Y \le t_1) = \alpha$$

ดังนั้นจึงปฏิเสธ Ho ถ้า $T_1 \le t_1$ ณ ระดับนัยสำคัญ α

ในกรณีตัวอย่างขนาดโต เราพิจารณาค่า t_1 และ t_2 ได้โดยอาศัยการแจกแจงปกติมาตรฐาน N (0, 1) หรือการแจกแจงปัวซองได้

ตัวอย่าง คนงานที่มีประสิทธิภาพจะผลิตสินค้าต่อวันได้ในระดับควอร์ไทล์สูง (Q_3) เท่ากับ 193 ถ้าจะทดสอบว่าคนงานคนหนึ่งมีประสิทธิภาพหรือไม่นั้นจะทำได้โดยการสุ่มตัวอย่างสินค้าที่เขาผลิตต่อวันมา สมมติว่าสุ่มมา 15 วัน ได้ผลผลิตต่อวัน ดังนี้

189 233 195 160 212 176 231 185 199 213 202 193

174 166 248

ผลสรุปจะเป็นอย่างไร ?

Ho : ควอร์ไทล์สูงของการผลิตเท่ากับ 193

Ha : ควอร์ไทล์สูงของการผลิตไม่เท่ากับ 193

ตัวสถิติทดสอบ T_1 และ T_2 นี้ภายใต้ Ho เป็นจริง จะมีการแจกแจงทวินามที่มีพารามิเตอร์ n และ π_0 จากตารางทวินามที่มี n = 15 และ $\pi_0 = 0.75$ เราจะได้ว่า

$P(Y \le 7) = 0.0173$ และ $P(Y \le 14) = 0.9866 = 1-0.0134$

ดังนั้นเราจะปฏิเสธ Ho ถ้า $T_1 \le 7$ หรือ $T_2 > 14$ ณ ระดับนัยสำคัญ $\alpha = 0.0173-0.0134 = 0.0307$

จากข้อมูลเราได้ค่าตัวสถิติ $T_1 = 7$ และ $T_2 = 6$ ซึ่งจะเห็นได้ว่า T_1 น้อยไป จึงปฏิเสธ Ho แสดงว่าควอร์ไทล์สูงของการผลิตไม่เป็น 193

ในกรณีที่ $\pi = 0.50$ ก็จะเป็นการทดสอบเกี่ยวกับมัธยฐานนั่นเอง นั่นคือทดสอบสมมติฐานหลัก Ho ที่ว่า

Ho : ควอนไทล์ที่ 0.50 ของประชากรเป็น x_0

หรือ Ha : มัธยฐานของประชากรเป็น x_0

3.1.7 ช่วงเชื่อมั่นสำหรับควอนไทล์ (Confidence Interval for Quantile)

วิธีการหาช่วงเชื่อมั่นสำหรับความน่าจะเป็น (π) ที่กล่าวมาในบทก่อนนั้น สามารถใช้หาช่วงเชื่อมั่นสำหรับฟังก์ชั่นแจกแจง หรือความน่าจะเป็นสะสมที่ x_0, $F(x_0)$ ได้ นั่นก็คือ เมื่อกำหนด x_0 ให้ เราสามารถหาช่วงเชื่อมั่นสำหรับความน่าจะเป็น $F(x_0)$ หรือ $P(X \le x_0)$ ได้ อย่างไรก็ตามเมื่อกำหนดความน่าจะเป็น π_0 ให้ เราก็สามารถหาช่วงเชื่อมั่นสำหรับควอนไทล์ x_0 ได้ ช่วงเชื่อมั่นแบบนี้ก็คือช่วงเชื่อมั่นสำหรับควอนไทล์นั่นเอง

สำหรับ $0 \le \pi_0 \le 1$ เราจะได้ช่วงเชื่อมั่น $100(1-\alpha)$ % สำหรับควอนไทล์ x_0 จะอยู่ในรูปฟอร์มดังนี้

$$X_{(r)} \le x_0 \le X_{(s)}$$

ในเมื่อ $X_{(r)}$ และ $X_{(s)}$ เป็นตัวสถิติอันดับ (Order Statistics) ที่ r และ s ค่าของ r และ s สามารถพิจารณาได้จากขนาดตัวอย่าง และระดับความเชื่อมั่น $(1-\alpha)$

วิธีการสร้างช่วงเชื่อมั่นสำหรับควอนไทล์ที่ π_0 ทำได้ดังนี้ – เมื่อตัวอย่างขนาดเล็ก ($n \le 20$) เราอาศัยตารางทวินาม เพื่อหาค่า r และ s นั่นคือจาก n และ π_0 ที่กำหนดให้ เราจะได้ค่าความน่าจะเป็นตามแนวตั้งที่เกือบเท่ากับ $\alpha/2$ และให้ $\alpha_1 = \alpha/2$ ค่า y ที่สมนัยกับความน่าจะเป็นนี้จะให้เป็น s-1 แล้วอ่านต่อไปจนพบค่าที่ให้ความน่าจะเป็นเกือบเท่ากับ $1-\alpha/2$ ซึ่งจะให้ $1-\alpha_2$ ค่า y ที่สมนัยนี้จะให้เป็น s-1 ดังนั้นเราจะได้ α_1, α_2, r, s และ $1-\alpha_1-\alpha_2$ ตัวประมาณค่าแบบช่วงนั้นจะเป็นช่วงระหว่าง $X_{(r)}$ กับ $X_{(s)}$ และค่าทั้งสองนี้พิจารณาได้จากข้อมูลตัวอย่างนั่นเอง ดังนั้นช่วง

$$X_{(r)} \le x_0 \le X_{(s)}$$

จะเป็นช่วงเชื่อมั่น $100(1-\alpha_1-\alpha_2)$ % สำหรับควอนไทล์ π_0 หรือ x_0

เมื่อใช้ตัวอย่างขนาดโต ($n \ge 20$) เราหาค่า r และ s โดยอาศัยการแจกแจงปกติมาตรฐาน ดังนี้

$$r = n_0 - Z_{\alpha/2}\sqrt{n_0(1-\pi_0)}$$

และ $$s = n_0 + Z_{\alpha/2}\sqrt{n_0(1-\pi_0)}$$

โดยทั่วไป r และ s ไม่เป็นจำนวนเต็ม จึงต้องทำให้เป็นจำนวนเต็มโดยการเพิ่มขึ้น เช่น r = 10.15 ก็เป็น r = 11 เป็นต้น

นอกจากนี้เรายังสามารถหาช่วงเชื่อมั่นทางเดียวสำหรับควอนไทล์ได้โดยการหาค่า r หรือ s เพียงค่าเดียวเท่านั้น

ถ้าฟังก์ชันแจกแจงเป็นแบบต่อเนื่อง แล้วช่วงเชื่อมั่นทางเดียวสำหรับ x_0 จะอยู่ในรูปฟอร์ม

$$P(X_{(r)} \le x_0) = 1-\alpha_1 \text{ และ } P(x_0 \le X_{(s)}) = 1-\alpha_2$$

ตัวอย่าง ผู้ควบคุมฝ่ายผลิต ต้องการประมาณช่วงเชื่อมั่น 90% สำหรับควอนไทล์สูง ($x_{.75}$) ของอายุใช้งานของหลอดวิทยุ จึงสุ่มหลอดวิทยุมา 16 หลอด จากที่ผลิตไว้ในสัปดาห์หนึ่ง เมื่อทำการทดสอบหาอายุใช้งาน และเรียงตามขนาดแล้ว จะเป็น

46.9	47.2	49.1	56.5	56.8	59.2	59.9	63.2
63.3	63.4	63.7	64.1	67.1	67.7	73.3	78.5

จากตารางทวินามที่มี n = 16 และ π =.75 เมื่ออ่านค่าความน่าจะเป็นตามแนวตั้งในช่อง $\pi = 0.75$ ปรากฏว่าความน่าจะเป็น 0.0271 ใกล้กับ $\alpha/2 = 0.10/2 = 0.05$ ค่านี้สมนัยกับ y = 8 ดังนั้น r = 9 และความน่าจะเป็นที่ใกล้ $1-\alpha/2$ =0.95 มากที่สุดคือ $0.9365 = 1-\alpha_2$ (นั่นคือ $\alpha_2 = 0.0635$) ค่านี้สมนัยกับ r = 14 ฉะนั้น s = 15

ดังนั้นระดับความเชื่อมั่นจะเป็น 1 - 0.0271 - 0.0635 = 0.9094 และช่วงเชื่อมั่น 90.94% ของควอร์ไทล์สูงจะเป็น

$$X_{(9)} \le x_{.75} \le X_{(15)}$$

(อายุใช้งานของหลอดวิทยุ ถือว่าเป็นตัวแปรเชิงสุ่มแบบต่อเนื่องได้)

จากข้อมูลเราได้ว่า $X_{(9)} = 63.3$ และ $X_{(15)} = 73.3$ เพราะฉะนั้นช่วงเชื่อมั่น 90.94% สำหรับควอร์ไทล์สูงจะเป็น

$$63.3 \leq x_{.75} \leq 73.3$$

ถ้าใช้การประมาณค่าโดยอาศัยการแจกแจงปกติมาตรฐาน แล้วค่า r และ s จะเป็น

$$r = 16(0.75) - 1.645\sqrt{16(.75)(.25)}$$
$$= 12 - 2.86 = 9.14$$

และ $$s = 12 + 2.86 = 14.86$$

เพราะฉะนั้น r = 10 และ s = 15 ซึ่งจะให้ช่วงเชื่อมั่น 90% เป็น (63.4, 73.3)

3.2 กรณีสองตัวอย่างที่เป็นอิสระกัน (Two Independent Samples)

ในกรณีสองตัวอย่างนี้เราจะเกี่ยวข้องกับประชากรสองประชากรที่เป็นอิสระกัน โดยทำการอ้างอิงถึงผลต่างของค่าเฉลี่ยหรือมัธยฐานจากประชากรดังกล่าวนั้น แต่สำหรับการทดสอบสมมติฐานเรามักจะสนใจเปรียบเทียบมัธยฐาน นั่นคือสนใจว่าผลต่างของมัธยฐานแตกต่างจากศูนย์หรือไม่นั่นเอง แบบทดสอบ และวิธีการประมาณในเชิงพารามิเตอร์เราก็อาศัยการแจกแจงแบบที (t) แต่ในกระบวนการไร้พารามิเตอร์นั้นเรามีแบบทดสอบและวิธีการประมาณค่า ดังต่อไปนี้

3.2.1 แบบทดสอบมัธยฐาน (Median Test)

แบบทดสอบมัธยฐานเป็นแบบทดสอบที่ง่ายๆ แต่ใช้กันบ่อยๆ สำหรับทดสอบสมมติฐานหลัก Ho ที่ว่า สองตัวอย่างสุ่มเลือกมาจากสองประชากรที่มีมัธยฐานเท่ากัน ดังนั้นสมมติฐานสถิติที่จะทดสอบจึงเป็นดังนี้

Ho : สองประชากรเหมือนกัน

Ha : สองประชากรมีมัธยฐานแตกต่างกัน

ในการทดสอบสมมติฐานนี้ก็อาศัยตัวอย่างสุ่มที่เป็นอิสระจากสองประชากรที่สนใจ ให้ X_1, X_2, .., X_n และ Y_1, Y_2, .., Y_m เป็นสองตัวอย่างสุ่มขนาด n และ m จากประชากรทั้งสองนั้น ข้อมูลจากตัวอย่างจะต้องเป็นค่าของตัวแปรแบบต่อเนื่องที่เราสนใจ โดยมีสเกลการวัดอย่างน้อยเป็นแบบอันดับ ข้อมูลจากตัวอย่างทั้งสองนั้นนำมาหามัธยฐานร่วม (Common Median) แล้วแยกข้อมูลในแต่ละตัวอย่างออกเป็น 2 พวก คือพวกหนึ่งอยู่เหนือมัธยฐาน และอีกพวกหนึ่งอยู่ใต้มัธยฐาน แล้วจะได้ข้อมูลสรุปดังตารางต่อไปนี้

ตัวอย่าง	1	2	รวม
เหนือมัธยฐานร่วม	A	B	A+B
ใต้มัธยฐานร่วม	C	D	C+D
รวม	n=A+C	m=B+D	n+m

ในเมื่อ A และ B เป็นจำนวนค่าสังเกต หรือข้อมูลจากตัวอย่าง 1 และ 2 ที่อยู่เหนือมัธยฐานร่วม ส่วน C และ D เป็นจำนวนข้อมูลที่อยู่ใต้มัธยฐานร่วม

ตัวสถิติทดสอบสำหรับสมมติฐานหลักนี้อยู่กับ A และ B ถ้าสองประชากรมีมัธยฐานเดียวกันจริง เราก็หวังว่าครึ่งหนึ่งของค่าสังเกตในแต่ละตัวอย่างจะอยู่เหนือมัธยฐานร่วม และอีกครึ่งหนึ่งอยู่ใต้มัธยฐานร่วม นั่นคือเราจะได้ว่า A หรือ C จะประมาณ n/2 และ B หรือ D จะประมาณ m/2 และจากข้อมูลตัวอย่างนั้นแบบทดสอบมัธยฐานจะสรุปได้ว่า "สมมติฐานหลัก Ho จะเป็นเท็จก็ต่อเมื่อขนาดของความแตกต่าง (Discrepancy) ระหว่างสัดส่วนที่ค่าสังเกตจะอยู่เหนือและใต้มัธยฐานร่วมมีมากเพียงพอ" ถ้าสมมติฐานหลักเป็นจริงเราจะได้ตัวประมาณค่าของมัธยฐานร่วมประชากร (Population Common Median) ซึ่งก็คือมัธยฐานร่วมตัวอย่างนั่นเอง

มูด (Mood, 1950) ได้แสดงว่า การแจกแจงตัวอย่างของ A และ B ภายใต้สมมติฐานหลักจะเป็นแบบไฮเปอร์จีโอเมทริค ดังนี้

$$P(A, B) = \binom{n}{A}\binom{m}{B} / \binom{n+m}{a}$$

ในเมื่อ a = (n+m+1)/2 ถ้า n+m เป็นเลขคี่ หรือ a = (n+m)/2 ถ้า n+m เป็นเลขคู่ ดังนั้นในการทดสอบสมมติฐานหลัก เราจึงใช้การแจกแจงนี้ช่วยตัดสินใจ

ถ้าตัวอย่างขนาดโต (n+m) > 10 จะไม่สะดวกในการคำนวณ จึงจำเป็นต้องใช้วิธีประมาณค่าแบบตัวอย่างขนาดโต เพราะภายใต้เงื่อนไขนั้นเราสามารถจะประมาณการแจกแจงไฮเปอร์จีโอเมทริคด้วยการแจกแจงทวินาม และประมาณการแจกแจงทวินามด้วยการแจกแจงปกติ นั่นคือเราจะได้ตัวสถิติทดสอบ Z ดังนี้

$$Z = \frac{A/n - B/m}{\sqrt{P(1-P)(1/n + 1/m)}}\ ;\ P = (A+B)/(n+m)$$

ตัวสถิตินี้จะมีการแจกแจงปกติมาตรฐาน หรือจะใช้ตัวสถิติทดสอบ

$$\chi^2 = \frac{(m+n)\ (AD-BC)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

ซึ่งมีการแจกแจงไคสแควร์ ด้วยองศาความเป็นอิสระ 1

การเท่ากันของค่าสังเกต (Tied Observations) อาจจะเกิดขึ้นได้ ถึงแม้จะมีข้อกำหนดเกี่ยวกับความต่อเนื่องของตัวแปรที่สนใจ นั่นคืออาจจะมีค่าสังเกตตั้งแต่หนึ่งค่าขึ้นไปเท่ากับมัธยฐานร่วมพอดี เมื่อเกิดกรณีเท่ากันขึ้น เราทำได้ 2 วิธีดังนี้ (1) ถ้า n+m โตพอ และมีค่าสังเกตน้อยค่าที่เท่ากับมัธยฐานร่วม แล้วให้ตัดทิ้งค่าเหล่านั้นออกไปจากการวิเคราะห์ และ (2) ให้แยกประเภทข้อมูลจากตัวอย่างทั้งสองใหม่เป็น อยู่เหนือมัธยฐานร่วม (หรือมากกว่ามัธยฐานร่วม) และไม่อยู่เหนือมัธยฐานร่วม (หรือมากกว่าหรือเท่ากับมัธยฐานร่วม)

ตัวอย่าง ในการศึกษาเด็ก 4 ขวบ ซึ่งเป็นชาย 32 คน และหญิง 16 คน เพื่อจะดูการก้าวร้าวจากการให้เด็กทั้งหมดเล่นด้วยกัน แล้วสังเกตเด็กแต่ละคน และบันทึกขนาดของความก้าวร้าวเป็นคะแนนปรากฏว่าได้เป็นดังนี้

เด็กชาย 25 13 19 46 25 30 17 20 17 20 37 25 26 23 20 17 18
26 11 36 30 12 32 48 24 20 16 18 21 37 31 26

เด็กหญิง 31 43 21 42 38 30 19 20 38 29 13 50 32 41 28 30

มัธยฐานของความก้าวร้าวระหว่างเด็กชายและเด็กหญิงแตกต่างกันหรือไม่ ?

Ho: มัธยฐานของความก้าวร้าวในเด็กชายและหญิงไม่แตกต่างกัน

Ha: ความก้าวร้าวของเด็กชายมีความโน้มเอียงที่จะมากกว่าเด็กหญิง หรือเด็กหญิงมีความโน้มเอียงที่จะมากกว่าเด็กชาย

จากคะแนนความก้าวร้าว 48 ค่า เราหามัธยฐานร่วมได้เป็น (25+26) /2 = 25.5 แล้วเราแยกคะแนนในแต่ละตัวอย่างที่อยู่เหนือหรือใต้มัธยฐานร่วม ได้เป็นดังนี้

กลุ่ม	เด็กชาย	เด็กหญิง	รวม
เหนือมัธยฐาน	12	12	24
ใต้มัธยฐาน	20	4	24
รวม	32	16	48

เราได้ P = (12+12)/48 = 0.50 และ

$$Z = \frac{(12/32-12/16)}{\sqrt{0.50(1-0.50)\ (1/32+1/16)}} = -2.45$$

สำหรับ $\alpha = 0.05$ ค่าวิกฤตของตัวสถิติทดสอบ Z จะเป็น ± 1.96 เนื่องจากค่าของตัวสถิติ Z เป็น -2.45 ซึ่งน้อยกว่า -1.96 และอยู่ในเขตวิกฤต จึงปฏิเสธ Ho และสรุปได้ว่า มัธยฐานของความก้าวร้าวในเด็กทั้งสองเพศแตกต่างกัน (เด็กหญิงมีความโน้มเอียงที่จะมากกว่าเด็กชาย)

3.2.2 แบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์ (Wilcoxon-Mann-Whitney Test)

วิลคอกซัน (Wilcoxon, 1945) ได้เสนอแบบทดสอบสมมติฐานเกี่ยวกับการเท่ากันของพารามิเตอร์ค่ากลาง (Location Parameter) ในสองประชากร โดยศึกษาเฉพาะกรณีที่ขนาดตัวอย่างเท่ากัน และใช้ผลรวมอันดับที่ (Rank Sum) เป็นตัวสถิติทดสอบ แมนน์และวิทนีย์ (Mann-Whitney, 1947) ได้เสนอกรณีที่ขนาดตัวอย่างไม่เท่ากัน และยังชี้ให้เห็นถึงความสัมพันธ์ระหว่างตัวสถิติของเขากับตัวสถิติวิลคอกซันด้วย

เนื่องจากแบบทดสอบนี้อาศัยตัวอย่างที่ประกอบด้วยข้อมูลแบบอันดับ ดังนั้นความแตกต่างหรือการเท่ากันของค่ากลางที่น่าสนใจก็คือ ประชากรทั้งสองมีมัธยฐานเท่ากันหรือไม่ นั่นคือสมมติฐานหลัก จะเป็น

$$Ho : M_1 = M_2 \text{ หรือ } Ho : M_1 - M_2 = 0$$

การทดสอบสมมติฐานหลักนี้ก็อาศัยตัวอย่างสุ่มที่เป็นอิสระกันจากประชากรทั้งสองโดยที่ตัวแปรที่สนใจนั้นเป็นแบบต่อเนื่องและมีสเกลการวัดอย่างน้อยแบบอันดับ และการแจกแจงของสองประชากรนั้นถ้าจะแตกต่างกัน ก็จะแตกต่างกันเฉพาะค่ากลางเท่านั้น

ให้ $X_1, X_2, .., X_n$ และ $Y_1, Y_2, .., Y_m$ เป็นตัวอย่างสุ่มขนาด n และ m ($n \leq m$) จากสองประชากรที่สนใจ และตัวอย่างทั้งสองนั้นจะต้องเป็นอิสระกัน ให้รวมค่าสังเกตเข้าเป็นกลุ่มเดียวกันและเรียงลำดับจากน้อยไปมาก และให้อันดับ 1 แก่ข้อมูลน้อยสุด อันดับ 2 แก่ข้อมูลน้อยสุดรองลงมาและต่อๆ ไปจนถึงมากสุด ซึ่งเป็นอันดับ n+m ถ้าข้อมูลหลายๆ ค่าเท่ากันก็กำหนดอันดับเฉลี่ยให้แก่ข้อมูลที่เท่ากันเหล่านั้น

ถ้าให้ $R(X_i)$ และ $R(Y_j)$ เป็นอันดับที่ของค่าสังเกตซึ่งกำหนดให้แก่ X_i และ Y_j แล้วตัวสถิติทดสอบวิลคอกซันจะเป็นผลรวมอันดับของ X_i ดังนี้

$$S = \Sigma R(X_i)$$

แต่ตัวสถิติทดสอบแมนน์และวิทนีย์ กำหนดไว้ดังนี้

$$T = \sum_{i,j}^{n,m} \phi(X_i, Y_j)$$

ในเมื่อ $\phi(X_i, Y_j) = 1$ ถ้า $X_i > Y_j$ และ $= 0$ ถ้าเป็นอย่างอื่น

ตัวสถิติทดสอบ T นี้คำนวณได้ดังนี้ – แต่ละคู่ของ X_i และ Y_j ถ้า $X_i > Y_j$ ก็ให้เป็น 1 สำหรับคู่นั้นแต่ถ้า $X_i < Y_j$ ก็ให้เป็น 0 แล้วรวมค่า 0 กับ 1 ที่ได้เหล่านั้น ก็จะเป็นค่าของตัวสถิติ T นั่นเอง

ตัวสถิติวิลคอกซัน S และตัวสถิติแมนน์–วิทนีย์ T มีความสัมพันธ์กันดังนี้

$$T = S - n(n+1)/2$$

ค่าเฉลี่ย และความแปรปรวนของตัวสถิติ T จะเป็นดังนี้

$$E(T) = nm/2, \quad V(T) = nm(n+m+1)/12$$

ถ้าตัวสถิติทดสอบ S ที่เป็นผลรวมอันดับ ซึ่งกำหนดให้แก่ค่าที่มาจากประชากร X นั้นมีค่าน้อยเกินไป (หรือมากเกินไป) ก็แสดงว่าค่าจากประชากร X จะมีความโน้มเอียงที่จะน้อยกว่า (หรือมากกว่า) ประชากรอื่น Y นั่นคือตัวสถิติ T ซึ่งสัมพันธ์กับตัวสถิติ S ก็จะสรุปผลเกี่ยวกับสมมติฐานหลักได้เช่นเดียวกัน นั่นคือเกณฑ์ตัดสินใจสำหรับตัวสถิติ T กำหนดไว้ตามสมมติฐานดังนี้

(1) สำหรับ $Ho : M_1 = M_2$ และ $Ha : M_1 \neq M_2$ ถ้า T มีค่ามากหรือน้อยเกินไปก็จะแสดงว่า Ho ไม่เป็นจริง ดังนั้นจึงปฏิเสธ Ho ณ ระดับนัยสำคัญ ถ้า $T < W_{\alpha/2}$ หรือ $T > W_{1-\alpha/2}$ ในเมื่อ $W_{1-\alpha/2}$ เป็นควอนไทล์ที่ $\alpha/2$ จากตารางแมนน์–วิทนีย์ (ตาราง 11) ส่วน $W_{1-\alpha/2}$ กำหนดได้จาก

$$W_{1-\alpha/2} = nm - W_{\alpha/2}$$

(2) สำหรับ $Ho : M_1 \geq M_2$ และ $Ha : M_1 < M_2$ ถ้า T มีค่าน้อยเกินไป จะแสดงว่า Ha เป็นจริง ดังนั้นจึงปฏิเสธ Ho ณ ระดับนัยสำคัญ ถ้า $T < W_\alpha$

(3) สำหรับ Ho : $M_1 \leq M_2$ และ Ha : $M_1 > M_2$ ถ้า T มีค่ามากเกินไปจะแสดงว่า Ha เป็นจริง ดังนั้นจึงปฏิเสธ Ho ณ ระดับนัยสำคัญ α ถ้า $T > W_{1-\alpha}$

ในกรณีตัวอย่างขนาดโต เราก็อาศัยการแจกแจงปกติมาตรฐาน นั่นคือเราใช้ตัวสถิติ

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

ถ้าค่าสังเกตภายในกลุ่มเดียวกันเท่ากัน (Ties) ก็จะไม่มีผลกระทบต่อตัวสถิติทดสอบ แต่ถ้าต่างกลุ่มกันก็จะมีผลกระทบ โนเธอร์ (Noether) ชี้ให้เห็นว่าถ้าค่าสังเกตที่ไม่เท่ากันมีมากพอ การเท่ากันจะมีผลกระทบต่อตัวสถิติทดสอบเล็กน้อยเท่านั้น สำหรับตัวอย่างขนาดโต และมีการเท่ากันเกิดขึ้น เราปรับปรุง V(T) ด้วยแฟคเตอร์ f ต่อไปนี้

$$f = \frac{nm(\Sigma g^3 - \Sigma g)}{12(n+m)\ (n+m-1)}$$

ในเมื่อ g เป็นจำนวนของการเท่ากันในอันดับที่หนึ่ง ๆ ดังนั้น V(T) จึงเป็นดังนี้

$$V'(T) = nm(n+m+1)/12 - f$$

ข้อสังเกต (ก) เมื่อตัวอย่างขนาดโตมาก แบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์ จะมีอำนาจทดสอบสูง

(ข) เมื่อตัวอย่างขนาดเล็ก ให้ใช้แบบทดสอบเชิงสุ่มของฟิชเชอร์ (Fisher Randomization Test)

(ค) ถ้าสมมติฐานหลัก Ho เป็น Ho : $M_1 - M_2 = D_0$ ในเมื่อ D_0 เป็นค่าที่ระบุไว้ ค่าของ X_i จะต้องแปลงเป็น $X_1 = X_i - D_0$ ก่อน แล้วจึงนำมาจัดอันดับที่ร่วมกับ Y_j

(ง) ถ้า S' เป็นตัวสถิติทดสอบที่กำหนดจากผลรวมอันดับที่ของ Y นั่นคือ $S' = \sum_j^m R(Y_j)$ แล้ว S กับ S' จะมีความสัมพันธ์กันดังนี้

$$S' = (n+m)\ (n+m+1)/2 - S$$

ตัวอย่าง มีผู้กล่าวกันว่า เด็กชนบทมีร่างกายสมบูรณ์กว่าเด็กในเมือง ในการทดสอบคำกล่าวนี้ได้สุ่มเด็กนักเรียนที่มีอายุเท่ากันมาจากชนบทและในเมืองมาจำนวนหนึ่ง แล้วทำการทดสอบความสมบูรณ์ของร่างกาย โดยให้เป็นคะแนน (คะแนนสูงก็แสดงว่าสมบูรณ์มาก) ดังนี้

เด็กชนบท (X_i)	14.8	7.3	5.6	6.3	9.0	4.2	10.6	12.5	2.7	12.9	16.1	11.4
เด็กในเมือง (Y_j)	12.7	16.9	7.6	2.4	6.2	9.9	14.2	7.9	2.1	11.3	6.4	6.1
	10.6	15.3	14.8	12.6	16.0	8.3	9.1	10.6	6.7	6.7	10.6	5.0
	17.7	5.6	3.6	18.6	1.8	2.6	11.8	5.6	1.0	3.2	5.9	4.0

สมมติฐานที่จะทดสอบ จะเป็นดังนี้

Ho : เด็กชนบทมีร่างกายสมบูรณ์พอ ๆ กับเด็กในเมือง ($M_1 \neq M_2$)

Ha : เด็กชนบทสมบูรณ์กว่าเด็กในเมือง ($M_1 > M_2$)

คะแนน X_i และ Y_i ต่าง ๆ นั้นเมื่อนำมาเรียงจากน้อยไปมาก และให้อันดับที่จาก 1 ไปถึง n + m = 12 + 36 = 48 สำหรับคะแนน X_i ได้อันดับที่ ดังนี้

41.5 30.5 22 36 13 39 18 45 26 34 10 6

ตัวสถิติทดสอบ S จึงมีค่าเป็นดังนี้

$$S = 41.5 + 30.5 + ... + 10 + 6 = 321$$

ตัวสถิติทดสอบ T จึงมีค่าเป็น

$$T = S - n(n+1)/2$$

$$= 321 - 12(12+1)/1 = 321 - 78 = 243$$

เนื่องจากตัวอย่างขนาดโต เราจึงใช้ตัวสถิติทดสอบ Z ดังนั้นจึงหาค่า E(T) และ V(T) ได้เป็น

$$E(T) = 12(36)/2 = 216$$

$$V(T) = 12(36)(12 + 36 + 1)/12 = 1764$$

โดยที่ค่าสังเกตซ้ำกัน เราจึงหาแฟคเตอร์ f ได้ดังนี้

ค่าสังเกตที่เท่ากัน

X	Y			g	g^3
5.6	5.6	5.6		3	27
	6.7	6.7		2	8
10.6	10.6	10.6	10.6	4	64
14.8	14.8			2	8
				11	107

$$f = \frac{12(36)(107 - 11)}{12(12+36)(12+36-1)} = 1.53$$

ดังนั้น V(T) ที่ปรับปรุง จะเป็น V'(T)

$$V'(T) = V(T) - f$$

$$= 1764 - 1.53 = 1762.47$$

เราจึงได้ค่า Z เป็นดังนี้

$$Z = \frac{243 - 216}{\sqrt{1762.47}} = 0.64$$

ค่าวิกฤต ณ ระดับนัยสำคัญ $\alpha = 0.05$ ของตัวสถิติทดสอบ Z จะเป็น $Z_{.05} = 1.645$ ซึ่งจะปฏิเสธ Ho ไม่ได้ แสดงว่าเด็กชนบทมีร่างกายสมบูรณ์พอ ๆ กับ เด็กในเมือง

3.2.3 การประมาณผลต่างมัธยฐานโดยอาศัยตัวสถิติวิลคอกซัน-แมนน์-วิทนีย์ (Estimation for Median Difference)

ให้ $X_1, X_2, .., X_n$ และ $Y_1, Y_2, .., Y_m$ เป็นตัวอย่างที่เป็นอิสระกันจากประชากร 1 และ 2 ตามลำดับ และประชากรทั้งสองนี้จะเหมือนกันยกเว้นแต่ค่ากลาง (มัธยฐาน) การประมาณผลต่างมัธยฐาน $M_1 - M_2 = M_D$ ทำได้ดังนี้

(1) เรียงค่าสังเกต X และ Y ตามขนาด แล้วหาค่าผลต่างที่เป็นไปได้จำวน nm ค่า ซึ่งจะแทนด้วย

$$U_{ij} = X_i - Y_j, \ (i = 1, 2 .., n ; j = 1, 2, .., m)$$

(2) เรียงค่าผลต่าง U_{ij} จากน้อยไปมาก ซึ่งจะให้เป็น $U_{(1)}, U_{(2)}, ..., U_{(nm)}$

ตัวประมาณค่าแบบจุดของผลต่างมัธยฐาน M_D นั้น ฮอดจ์-เลฮ์แมน (Hodges – Lehman) ได้เสนอไว้เป็น $\hat{M}_D$

$$\hat{M}_D = \text{มัธยฐาน} \left\{ U_{ij} = X_i - Y_j, \quad \begin{matrix} i = 1, 2 .., n \\ j = 1, 2, .., m \end{matrix} \right\}$$

$$= \begin{cases} U_{(k)} & \text{ถ้า nm เป็นเลขคี่} \\ (U_{(k)} + U_{(k+1)})/2 & \text{ถ้า nm เป็นเลขคู่} \end{cases}$$

ในเมื่อ $k = (nm + 1)/2$ หรือ $k = nm/2$ แล้วแต่ว่า nm เป็นเลขคี่หรือเลขคู่

สำหรับการประมาณค่าแบบช่วงนั้นเราหาขีดจำกัดบน และขีดจำกัดล่าง (Upper and Lower Limit) ของช่วงเชื่อมั่นสำหรับผลต่างมัธยฐานได้ดังนี้ – หาค่า $W_{\alpha/2}$ จากตาราง 11 ในเมื่อ $1-\alpha$ เป็นระดับความเชื่อมั่นที่ต้องการ แล้วผลต่าง U_{ij} ที่น้อยสุดอันดับที่ $W_{\alpha/2}$ จะเป็นขีดจำกัดล่าง ($\hat{\theta}_L$) และผลต่างที่มากสุดอันดับที่ $W_{\alpha/2}$ จะเป็นขีดจำกัดบน ($\hat{\theta}_U$) ดังนั้นช่วงเชื่อมั่น (อย่างน้อย) 100 $(1-\alpha)$% สำหรับผลต่างมัธยฐานประชากร M_D จะเป็น

$$\hat{\theta}_L < M_D < \hat{\theta}_U$$

ในเมื่อ $\hat{\theta}_L = U_{(W_{\alpha/2})}$ และ $\hat{\theta}_U = U_{(nm+1-W_{\alpha/2})}$

ถ้าตัวอย่างขนาดโต เราประมาณ $W_{\alpha/2}$ ได้ดังนี้

$$W_{\alpha/2} \cong nm/2 \pm Z_{\alpha/2} \sqrt{nm(n+m+1)/12}$$

โดยปกติ $W_{\alpha/2}$ ที่ประมาณได้นี้มักจะไม่เป็นจำนวนเต็ม จำเป็นต้องทำให้เป็นจำนวนเต็มที่ใกล้สุด

ตัวอย่าง ในการศึกษาอายุใช้งานของหลอดไฟ 2 ยี่ห้อ (ก, ข) ได้อายุใช้งานเป็นเดือนมาดังตารางข้างล่างนี้ โดยมี X เป็นอายุใช้งานของ ก และ Y เป็นของ ข เมื่อเรียงค่า X และ Y ตามขนาดแล้ว เราจะได้ผลต่าง $U_{ij} = X_i - Y_j$ ที่เป็นไปได้ $nm = 10(11) = 110$ ค่า ดังนี้

	X_i	77	82	85	86	86	86	89	91	92	93	100
Y_j	65	12	17	20	21	21	21	24	26	27	28	35
	65	12	17	20	21	21	21	24	26	27	28	35
	73	4	9	12	13	13	13	16	18	19	20	27
	75	2	7	10	11	11	11	14	16	17	18	25
	77	0	5	8	9	9	9	12	14	15	16	23
	78	−1	4	7	8	8	8	11	13	14	15	22
	83	−6	−1	2	3	3	3	6	8	9	10	17
	85	−8	−3	0	1	1	1	4	6	7	8	15
	90	−13	−8	−5	−4	−4	−4	−1	1	2	3	10
	97	−20	−15	−12	−11	−11	−11	−8	−6	−5	−4	3

สำหรับระดับความไม่เชื่อมั่น α = .05 และ n = 11, m = 10 เราได้ค่า $W_{\alpha/2}$ จากตาราง 11 เป็น $W_{\alpha/2}$ = 27 แล้วเราจะได้ค่าน้อยสุด และมากสุดอันดับที่ 1 และ 17 นั้นคือ $U_{(27)} = 1$ และ $U_{(110+1-27)} = 17$

ดังนั้นช่วงเชื่อมั่น 95% สำหรับผลต่างมัธยฐาน $M_1 - M_2 = M_D$ คือ

$$1 \leq M_1 - M_2 \leq 17$$

ตัวประมาณค่าแบบจุดของผลต่างมัธยฐาน M_D จะเป็น

M_D = มัธยฐาน { U_{ij}, i = 1, 2 .., 10 ; j = 1, 2, .., 11}

$= (U_{(55)} + U_{(56-2)})/2 = (8 + 9)/2$

$= 8.50$

ในการประมาณค่าแบบช่วงนั้น ถ้า n และ m โต เราต้องคำนวณผลต่าง $U_{ij} = X_i - Y_j$ เป็นจำนวนมากมาย และยังต้องเรียงค่า U_{ij} ตามขนาดอีกด้วย จะเห็นว่าเป็นงานที่น่าเบื่อ แต่ถ้ามีคอมพิวเตอร์ช่วยก็ไม่เป็นไร กรณีที่ไม่ใช้คอมพิวเตอร์ช่วยก็อาศัยวิธีกราฟ (Graphical Method) ซึ่งโมเสส (Moses, 1965) ได้เสนอไว้ดังนี้

(1) บวกหรือลบค่าสังเกต n + m ค่านั้นด้วยค่าคงที่ เพื่อให้ทุกค่าเป็นบวกและใกล้ๆ ศูนย์ แล้วเขียนจุดของคู่ (X_i, Y_j) จำนวน nm จุด บนกราฟที่ใช้สเกลเดียวกันทั้งสองแกน

(2) ลากเส้น 45° (L_1) จากซ้ายไปขวา โดยให้มีจุด (X_i, Y_j) จำนวน $W_{\alpha/2}$ จุด อยู่บนเส้นหรือเหนือเส้น L_1 นี้ จุดตัดแกน X ของเส้น L_1 นี้จะเป็นขีดจำกัดสูงของผลต่างมัธยฐานประชากร $\hat{\theta}_U = U_{(nm+1-a)}$; $a = W_{\alpha/2}$

ในทำนองเดียวกัน ลากเส้น 45° (L_2) จากซ้ายไปขวา โดยให้มีจุดอยู่บนเส้นหรือต่ำกว่าเส้น L_2 นี้จำนวน $W_{\alpha/2}$ จุด จุดตัดแกน X ของเส้นนี้จะเป็นขีดจำกัดต่ำของผลต่างมัธยฐาน $\hat{\theta}_L = U_{(W\alpha/2)}$

จุดกลางของช่วง หรือ $(\hat{\theta}_L + \hat{\theta}_U)/2$ นี้จะใช้เป็นตัวประมาณค่าแบบจุดของ M_D

3.2.4 แบบทดสอบชนิดคะแนนปกติของเทอร์รี่-โฮฟฟ์ดิง (Terry - Hoeffding Normal - Scores Test)

ในแบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์นั้นค่าสังเกตเดิมจะทำให้อยู่ในแบบอันดับ แต่ในแบบทดสอบชนิดคะแนนปกตินี้ค่าสังเกตเดิมจะทำให้อยู่ในแบบคะแนนปกติ ซึ่งตัวแปรที่แปลงนั้นจะเกี่ยวข้องกับคุณลักษณะต่างๆ ของการแจกแจงปกติ สำหรับคะแนนมาตรฐานที่แทนคะแนนเดิมนั้นจะเป็นตัวสถิติอันดับแบบปกติคาดหวัง (Expected Normal Order Statistic) หรือ $V^{(i)}$ ที่สอดคล้องกับอันดับที่ของคะแนนนั้น แบบทดสอบที่อาศัยคะแนนปกตินี้จะมีประสิทธิภาพมากกว่าแบบทดสอบที่อาศัยอันดับที่

ถ้าให้ (i) เป็นคะแนนอันดับที่ i ในตัวอย่างรวมขนาด $n = n_1 - n_2$ แล้ว $V^{(i)}$ จะเป็นค่าคาดหวังของคะแนนในตำแหน่งที่ i สำหรับ $V^{(i)}$ นั้นจะทำหน้าที่เป็นมาตราวัดระยะทางเช่นเดียวกับที่คะแนนมาตรฐาน Z แสดงระยะทางสัมพัทธ์ในการแจกแจงปกติ อันดับที่สูงๆ จะมี $V^{(i)}$ ที่เป็นบวกมากๆ และอันดับที่ต่ำๆ จะมี $V^{(i)}$ เป็นลบมากๆ อันดับที่กลางๆ ของการแจกแจงจะมี $V^{(i)}$ ใกล้ๆ 0 สำหรับตัวอย่างขนาด n นั้นผลรวมของ $V^{(i)}$ จะเป็นศูนย์

ค่าของ $V^{(i)}$ ซึ่งสมมติว่ามีโอกาสเกิดขึ้นเท่าๆ กันภายใต้สมมติฐานหลัก Ho ที่ว่าสองตัวอย่างมีค่าเฉลี่ยไม่แตกต่างกัน นั้นจะมีค่าคาดหวัง และความแปรปรวนดังนี้

$$E(V^{(i)}) = (1/n)\Sigma V^{(i)} = 0$$

$$V(V^{(i)}) = (1/n)\Sigma\left\{V^{(i)} - E\,(V^{(i)})\right\}^2$$

$$= (1/n)\Sigma(V^{(i)})^2$$

ตัวสถิติทดสอบสำหรับสมมติฐานหลัก จะกำหนดไว้ว่า

$$T_1 = \sum^{n} V^{(i)}$$

เราจะได้ค่าเฉลี่ยและความแปรปรวนดังนี้

$$E(T_1) = n_1 E(V^{(i)}) = 0$$

$$V(T_1) = n_1 V(V^{(i)})\,(n-n_1)/(n-1)$$

$$= \frac{n_1 n_2}{n-1}(1/n)\sum^{n}(V^{(i)})^2$$

ในเมื่อ $n_1 + n_2 = n$ เป็นตัวอย่างจากประชากรแบบจำกัด (Finite Population)

ถ้าขนาดตัวอย่างในแต่ละตัวอย่างมากกว่าหรือเท่ากับ 8 เราใช้การประมาณแบบปกติ ซึ่งจะได้เป็น

$$Z = \frac{T_1 - E(T_1)}{\sqrt{V(T_1)}}$$

ตัวสถิติ Z นี้จะมีการแจกแจงแบบปกติมาตรฐาน N(0, 1)

ตัวอย่าง ในการศึกษาเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มทดลองกับกลุ่มควบคุม ได้ข้อมูลมาดังนี้

กลุ่มควบคุม	36	42	17	40		
กลุ่มทดลอง	39	55	48	61	54	64

โดยอาศัยตาราง 12 เราแปลงค่าสังเกตจากกลุ่มทั้งสองที่รวมกันเป็น ค่า $V^{(i)}$ ได้ดังนี้

กลุ่มควบคุม			กลุ่มทดลอง		
คะแนน	อันดับที่	$V^{(i)}$	คะแนน	อันดับที่	$V^{(i)}$
36	2	−1.00	39	3	−0.66
42	5	−0.12	55	6	0.12
17	1	−1.54	48	7	0.38
40	4	−0.38	61	8	0.66
			54	9	1.00
			64	10	1.54
$T_1 = -3.04$			$T_2 = 3.04$		

$$T_1 = (-1.54) + (-1.00) + (-0.38) + (-0.12) = -3.04$$

$$T_2 = (-0.66) + 0.12 + 0.38 + 0.66 + 1.00 + 1.54 = 3.04$$

เราจะเห็นได้ว่า $T_1 + T_2 = 0$ นั่นคือ $T_1 = -T_2$ ภายใต้สมมติฐานหลัก

$$Ho : M_1 = M_2 \quad \text{หรือ } Ho : E(T_1) = E(T_2) = 0$$

นั้น ถ้าไม่มีความแตกต่างในค่าเฉลี่ยระหว่างสองประชากรที่สุ่มตัวอย่างมา เราก็หวังว่าความแตกต่างระหว่าง $\bar{T}_1$ กับ $\bar{T}_2$ จะใกล้ศูนย์ เมื่อ $T_1 = -T_2$ เราจึงใช้เพียง T_1

$$T_1 = \sum^{n} V^{(i)}$$

เท่านั้นเป็นตัวสถิติทดสอบ ในเมื่อ T_1 เกี่ยวข้องกับกลุ่มที่มีจำนวนค่าสังเกตน้อยกว่า

กฎตัดสินใจสำหรับตัวอย่างไม่โต พิจารณาได้ดังนี้ – ภายใต้สมมติฐานหลัก (Ho) จำนวนวิธีกำหนด $V^{(i)}$ ที่มีโอกาสเท่าๆ กันให้แก่ค่าสังเกต 4 ค่า ในตัวอย่างแรก และ 6 ค่า ในตัวอย่างหลังจะเป็น

$$T = \binom{10}{4} = \binom{10}{6} = 210$$

สำหรับการทดสอบสองทางที่มี = 0.05 เราได้จำนวนค่า T_1 ชนิดสุดเหวี่ยงเป็น

$$K = \alpha\binom{n_1 + n_2}{n_1} = \alpha\binom{n_1 + n_2}{n_1}$$

$$= 0.05(210) = 10.5 \cong 10$$

ค่าสุดเหวี่ยง 10 ค่า ซึ่งเป็นเขตปฏิเสธสำหรับตัวสถิติเทอร์รี่-โฮฟฟ์ดิง แสดงได้ดังตารางต่อไปนี้

ค่า $V^{(i)}$ ที่กำหนดแก่ตัวอย่างที่มีจำนวนน้อยกว่า				ค่าของ T_1
-1.54	-1.00	-0.66	-0.38	-3.58
-1.54	-1.00	0.66	-0.12	-3.32
-1.54	-1.00	-0.66	-0.12	-3.08
-1.54	-1.00	-0.38	-0.12	-3.04
-1.54	-1.00	-0.66	-0.38	-2.84
1.54	1.00	0.66	-0.38	2.82
1.54	1.00	0.38	-0.12	3.04
1.54	1.00	0.66	-0.12	3.08
1.54	1.00	0.66	-0.12	3.32
1.54	1.00	0.66	-0.38	3.58

เนื่องจากค่าสังเกตของ T เป็น -3.04 ซึ่งอยู่ในเซทปฏิเสธ จึงปฏิเสธสมมติฐานหลัก Ho นั่นคือ กลุ่มทดลองมีค่าเฉลี่ยมากกว่ากลุ่มควบคุม

3.2.5 แบบทดสอบชนิดคะแนนปกติผกผันของแวนเดอร์แวร์เดน (Van der Waerden Inverse Normal-Scores Test)

แบบทดสอบนี้ใช้ประยุกต์แบบเดียวกับแบบทดสอบเทอร์รี่-โฮฟฟ์ดิง แต่ใช้คะแนนปกติผกผันแทนค่าสังเกตเดิม คะแนนปกติผกผันหาได้ดังนี้

ให้ $r_1, r_2, .., r_n$ แทนอันดับที่ของค่าสังเกตที่เรียงอันดับ และให้ p_i เป็นอันดับเปอร์เซนไทล์ที่เกี่ยวพันกับค่าสังเกตแบบปกติชนิดอันดับที่ i นั่นคือ

$$p_i = r_i / (n + 1) = \phi(Z^{(i)}),\ i = 1, 2, .., n = n_1+n_2$$

และคะแนนปกติผกผัน $Z^{(i)}$ จะเป็น

$$Z^{(i)} = \phi^{-1}(p_i) = \phi^{-1}\{r_i/(n + 1)\}$$

ซึ่งพิจารณาได้โดยอาศัยตาราง 13 หรือตารางปกติมาตรฐาน

ตัวสถิติทดสอบจะเป็นผลรวมของคะแนนปกติผกผันที่กำหนดให้แก่ตัวอย่างที่มีค่าสังเกตจำนวนน้อยกว่า และจะให้เป็น T_1

$$T_1 = \sum^{n_1} Z^{(i)}$$

เช่นเดียวกับแบบทดสอบเทอร์รี่-โฮฟฟ์ดิง เราจะได้ว่า $T_1+T_2 = 0$ ค่าคาดหวังและความแปรปรวนของตัวสถิติ T_1 จะเป็นดังนี้

$$E(T_1) = 0$$
$$V(T_1) = \frac{n_1 n_2}{n-1}(1/n)\Sigma(Z^{(i)})^2$$

ถ้าขนาดตัวอย่าง n_1 และ n_2 ต่างก็มากกว่าหรือเท่ากับ 8 แล้วเราจะได้ตัวสถิติ Z ซึ่งมีการแจกแจงปกติมาตรฐาน N(0,1)

$$Z = \frac{T_1 - E(T_1)}{\sqrt{V(T_1)}}$$

นักจิตวิทยา และนักวิจัยการศึกษามักจะใช้คะแนนมาตรฐานในตัวแบบของการวัด (Measurement Models) และคะแนนแบบ แวน เดอร์ แวร์เดน นั้นสามารถแปลงเป็นคะแนน T ที่มีค่าเฉลี่ย 50 และส่วนเบี่ยงเบนมาตรฐาน 10 ได้ทันที ดังนั้นแบบทดสอบชนิดคะแนนปกติจึงเป็นที่สนใจของนักจิตวิทยาและนักวิจัยการศึกษา

ตัวอย่าง จากตัวอย่างในแบบทดสอบเทอร์รี-โฮฟฟ์ดิง เราแปลงค่าสังเกตเป็นคะแนนปกติผกผันได้ดังนี้

กลุ่มควบคุม

คะแนน	r_i	P_i	$Z^{(i)}$
36	2	2/11	-0.91
42	5	5/11	-0.11
17	1	1/11	-1.34
40	4	4/11	-0.35

$T_1 = -2.71$

กลุ่มทดลอง

คะแนน	r_i	P_i	$Z^{(i)}$
39	3	3/11	-0.60
48	6	6/11	0.11
54	7	7/11	0.35
55	8	8/11	0.60
61	9	9/11	0.91
64	10	10/11	1.34

$T_2 = 2.71$

ค่าสุดเหวี่ยง 10 ค่า ของตัวสถิติ T_1 ซึ่งเป็นเซทของค่าวิกฤต จะแสดงได้ดังตารางต่อไปนี้

ค่า $Z^{(i)}$ ที่กำหนดแก่ตัวอย่างที่มีจำนวนน้อยกว่า				ค่าของ T_1
-1.34	-0.91	-0.60	-0.35	-3.20
-1.34	-0.91	-0.60	-0.11	-2.96
-1.34	-0.91	-0.60	-0.11	-2.74
-1.34	-0.91	-0.35	-0.11	-2.71
-1.34	-0.91	-0.35	0.11	-2.49
1.34	0.91	0.35	-0.11	2.49
1.34	0.91	0.35	0.11	2.71
1.34	0.91	0.60	-0.11	2.74
1.34	0.91	0.60	0.11	2.96
1.34	0.91	0.60	0.35	3.20

เนื่องจากตัวสถิติ T_1 มีค่าสังเกตเป็น -2.71 ซึ่งอยู่ในเขตปฏิเสธ จึงปฏิเสธ Ho : $M_1 = M_2$ หรือ Ha : $E(T_1) = E(T_2) = 0$

จากคะแนนปกติผกผัน $Z^{(i)}$ ในกลุ่มควบคุมนั้น เมื่อแปลงให้อยู่ในรูปของคะแนน T

$$T = 50 + 10\,(Z^{(i)})$$

จะได้เป็น 40.9, 48.9, 36.6, 46.5 ซึ่งมีค่าเฉลี่ยเป็น

$$T_1 = (1/4)\,(40.9 + 48.9 + 36.6 + 46.5) = 43.2$$

และคะแนน T สำหรับกลุ่มทดลองเป็น 44.0, 51.1, 53.5, 56.0, 59.1, 63.4 ซึ่งมีค่าเฉลี่ยเป็น 54.5

ความแตกต่าง $\hat{\Delta} = \bar{T}_1 - \bar{T}_2 = 43.2 - 54.5 = -10.2$ ซึ่งแตกต่างอย่างมีนัยสำคัญ และความแตกต่างนี้เป็น 1.02 หน่วยเบี่ยงเบนมาตรฐาน (Standard Deviation Units)

3.2.6 แบบทดสอบชนิดคะแนนปกติของเบลล์-ดอคซัม (Bell - Doksum Norman - Scores Test)

แบบทดสอบนี้ใช้เปรียบเทียบค่าเฉลี่ยของสองประชากรเช่นเดียวกับแบบทดสอบเทอร์รี-โฮฟฟ์ดิง และแบบทดสอบแวน เดอร์ แวร์เดน แต่ใช้ส่วนเบี่ยงเบนปกติเชิงสุ่ม (Random Normal Deviates) แทนคะแนนเดิม ส่วนเบี่ยงเบนปกติเชิงสุ่มที่กำหนดให้แก่คะแนนเดิมนั้นต้องมีอันดับที่สอดคล้องกับอันดับที่ในคะแนนเดิม แล้วก็ดำเนินการต่อไปเช่นเดียวกับแบบทดสอบชนิดคะแนนปกติทั้งสองที่กล่าวมาแล้วนั่นเอง นั่นคือตัวสถิติทดสอบ T_1 จะเป็น

$$T_1 = \sum^{n_1} Z_{(i)}$$

ในเมื่อ $Z_{(i)}$ เป็นส่วนเบี่ยงเบนปกติเชิงสุ่ม ซึ่งพิจารณาได้จากตาราง 13 ที่มี $n = n_1 + n_2$

ถ้าตัวอย่างขนาดไม่โต เราพิจารณาเขตปฏิเสธได้เช่นเดียวกับแบบทดสอบชนิดคะแนนปกติทั้งสองที่กล่าวมาแล้วนั้น แต่ถ้าตัวอย่างมีขนาดโต เราก็อาศัยตัวสถิติทดสอบ Z

$$Z = \frac{\bar{T}_1 - (\bar{T}_2)}{\sqrt{1/n_1 + 1/n_2}}$$

ซึ่งมีการแจกแจงแบบปกติมาตรฐาน N(0,1) ทั้งนี้ก็เพราะว่า ภายใต้สมมติฐานหลัก Ho เราได้ $E(Z_{(i)}) = 0$ และ $V(Z_{(i)}) = 1$ กับ $E(\bar{T}_1 - \bar{T}_2) = T_1/n_1 + T_2/n_2$ และ $V(\bar{T}_1 - \bar{T}_2) = 1/n_1 + 1/n_2$

เนื่องจาก $T_1 + T_2 \neq 0$ และเนื่องจากส่วนเบี่ยงเบนปกติเชิงสุ่มนั้นผันแปรไปตามการสุ่มอย่างที่ซ้ำ ๆ กัน จึงทำให้นักวิจัยสองคนสรุปข้อมูลเดิมแตกต่างกันได้ ทั้งนี้ก็เพราะว่าได้ค่าต่างกันนั่นเอง

ตัวอย่าง (1) จากตัวอย่างในแบบทดสอบของเทอร์รี-โฮฟฟ์ดิง เราหาค่า $Z_{(i)}$ ได้โดยการสุ่มค่าเบี่ยงเบนปกติเชิงสุ่มขนาด $n = n_1 + n_2 = 10$ จากตารางส่วนเบี่ยงเบนปกติเชิงสุ่ม สมมติค่าที่ได้เมื่อเรียงลำดับแล้วจะเป็น -1.94, -0.89, -0.63, -0.54, -0.44 -0.24, 0.40, 0.49, 0.53, 0.66 เมื่อกำหนดค่าเหล่านี้ให้แก่ค่าสังเกตเดิมตามขนาดจากน้อยไปมาก จะได้ดังนี้

	กลุ่มควบคุม			กลุ่มทดลอง	
คะแนน	อันดับที่	$Z_{(i)}$	คะแนน	อันดับที่	$Z_{(i)}$
36	2	−0.89	39	3	−0.63
42	5	−0.44	48	6	−0.24
17	1	−1.94	54	7	0.40
40	4	−0.54	55	8	0.49
			61	9	0.53
			64	10	0.66
	$T_1 = -3.81$			$T_2 = 1.21$	

เซทของค่าสุดเหวี่ยง 10 ค่า หาได้ดังนี้

ค่าของ $Z_{(i)}$	ที่กำหนดแก่ตัวอย่างที่น้อยกว่า			ค่าของ T_1
−1.94	−0.89	−0.63	0.54	4.00
1.94	0.89	0.63	0.44	3.90
1.94	0.89	0.54	0.44	3.81
1.94	0.89	0.63	0.24	3.70
1.94	0.89	0.54	0.24	3.61
0.53	0.49	0.40	0.24	1.18
0.66	0.53	0.49	0.24	1.24
0.66	0.49	0.40	0.24	1.31
0.66	0.53	0.40	0.24	1.35
0.66	0.53	0.49	0.24	1.44
0.66	0.53	0.49	0.40	2.08

เนื่องจากค่าสังเกตของ $T_1 = -3.81$ อยู่ในเขตปฏิเสธ จึงปฏิเสธ Ho

ตัวอย่าง (2) นักวัดผลการศึกษาต้องการทราบว่าข้อทดสอบที่เรียงตามความยากง่ายนั้นจะมีผลต่อคะแนนเฉลี่ยหรือไม่ จากการทดลองโดยอาศัยนักเรียนที่มีความสามารถใกล้เคียงกัน 21 คน ได้คะแนนสอบดังนี้

กลุ่ม	คะแนน	อันดับที่	$V^{(i)}$	p_i	$\phi^{-1}(p_i)$	$Z_{(i)}$
เรียงความ	15	19	1.16	0.86	1.10	1.04
ยากง่าย	10	8	−0.36	0.36	−0.35	−0.88
(X)	7	3	−1.16	0.14	−1.10	−1.36
	9	5	−0.78	0.23	−0.75	−0.98
	11	9	−0.24	0.41	−0.23	−0.85

กลุ่ม	คะแนน	อันดับที่	$V^{(i)}$	p_i	$\phi^{-1}(p_i)$	$Z_{(i)}$
	19	21	1.89	0.96	1.69	1.54
	9	7	-0.49	0.32	-0.47	-0.89
	12	13	0.24	0.59	0.23	-0.35
	14	17	0.78	0.77	0.75	0.49
	11	12	0.12	0.54	0.11	-0.37
	17	20	1.43	0.91	1.34	1.14
		134	2.59		2.32	-1.47
ไม่ได้เรียง	3	1	-1.89	0.04	-1.69	-1.66
ความยากง่าย	11	10	-0.12	0.46	-0.11	-0.72
(Y)	11	11	0.00	0.50	0.00	-0.41
	13	16	0.63	0.73	0.60	0.42
	15	18	0.95	0.82	0.91	0.62
	8	4	-0.95	0.18	-0.91	-1.15
	13	14	0.36	0.64	0.35	-0.07
	6	2	-1.43	0.09	-1.34	-1.28
	13	15	0.49	0.68	0.47	0.17
	9	6	-0.63	0.27	-0.60	-0.93
		97	-2.59		-2.32	-5.01

ถ้าข้อกำหนดเกี่ยวกับแบบทดสอบที (T test) ที่ว่า (1) ภายในกลุ่มเป็นอิสระกัน (2) ระหว่างกลุ่มเป็นอิสระกัน (3) ตัวอย่างสุ่มมาจากประชากรที่แจกแจงปกติ และ (3) ทั้งสองประชากรที่สนใจมีความแปรปรวนเท่ากัน แล้วเราจะได้ว่า

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{1/n_1 + 1/n_2}} = \frac{12.8 - 10.20}{3.69\sqrt{1/11 + 1/10}} = 1.22$$

เมื่อ $\alpha = .05$ เราได้ $t^{(19)}_{.05} = 1.729$ จึงสรุปว่าทั้งสองกลุ่มมีคะแนนเฉลี่ยไม่แตกต่างกัน

สำหรับแบบทดสอบวิลคอกซัน เราได้ค่าของตัวสถิติเป็น

$$S = \Sigma R(X_i) = 134$$

$$E(S) = n_1 (n + 1) / 2 = 11(21 + 1) / 2 = 121$$

$$V(S) = n_1 n_2 (n + 1) / 12 = 11 (10 (21 + 1) = 201.67$$

$$Z = \frac{S - E(S)}{\sqrt{V(S)}} = \frac{134 - 121}{\sqrt{201.67}} = 0.91$$

เมื่อ α = .05 เราได้ $Z_{.05}$ = 1.645 จึงสรุปได้เช่นเดียวกับแบบทดสอบที

สำหรับแบบทดสอบเทอร์รี-โฮฟฟ์ดิง เราได้ว่า

$$T_1 = \sum^{n} V^{(i)} = 2.59$$

$$E(T_1) = 0$$

$$V(T_1) = \frac{n_1 n_2}{n-1} (1/n) \sum^{n} (V^{(i)})^2$$

$$= \frac{11(10)}{21-1} (1/21) \{ (1.16)^2 + ... + (-.63)^2 \}$$

$$= 4.8778$$

$$Z = \frac{2.59 - 0}{\sqrt{4.8778}} = 1.17$$

เมื่อ α = .05 เราก็ปฏิเสธ Ho ไม่ได้เช่นเดียวกัน

สำหรับแบบทดสอบ แวน เดอร์ แวร์เดน เราได้ค่าของตัวสถิติทดสอบเป็น

$$T_1 = \sum^{n_1} Z^{(i)} = \sum^{n_1} (Z^{(i)}) = 2.32$$

$$E(T_1) = 0$$

$$V(T_1) = \frac{n_1 n_2}{n-1} (1/n) \sum^{n} (Z^{(i)})^2$$

$$= \frac{11(10)}{21-1} (1/21) \{ (1.10)^2 + ... + (-.60)^2 \} = 4.1916$$

$$Z = \frac{2.32 - 0}{\sqrt{4.1916}} = 1.13$$

เมื่อ α = .05 เราก็ปฏิเสธ Ho ไม่ได้อีก

สำหรับแบบทดสอบเบลล์-ดอคซัม เราได้

$$T_1 = \sum^{n_1} Z_{(i)} = -1.47$$

$$= -5.01$$

$$Z = \frac{T_1 - T_2}{\sqrt{1/n_1 - 1/n_2}} = \frac{(-1.47/11) - (-5.02/10)}{\sqrt{1/11 + 1/10}} = 0.84$$

เมื่อ α = .05 เราก็สรุปได้เช่นเดียวกับแบบทดสอบอื่นๆ ที่กล่าวมานั่นเอง

ข้อสังเกต แบบทดสอบชนิดคะแนนปกติที่กล่าวมานี้จะดีที่สุด ถ้าคะแนนหรือข้อมูลจากตัวอย่างนั้นมาจากประชากรที่มีส่วนหางน้อยๆ (Sharp tails) เช่นประชากรแบบปกติ เอกซ์โพเนนเชียล เป็นต้น แต่แบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์ จะดีกว่า ถ้าคะแนนหรือข้อมูลของตัวอย่างมาจากประชากรที่มีส่วนหางมากๆ (Heavy tails) เช่นประชากรยูนิฟอร์ม โลยิสติค ดับเบิล-เอกซ์ โพเนนเชียล หรือคอชี เป็นต้น

3.2.7 แบบทดสอบวิลคอกซันชนิดชั้นภูมิ (Wilcoxon's Stratified Test)

แบบทดสอบนี้เป็นแบบหนึ่งของแบบทดสอบวิลคอกซันชนิดผลรวมอันดับที่ หรือแบบทดสอบแมนน์-วิทนีย์ ซึ่งใช้เปรียบเทียบสองตัวอย่างสุ่มแบบชั้นภูมิที่เป็นอิสระกัน จำนวนค่าสังเกตในแต่ละชั้นภูมิของสองตัวอย่างจะต้องเท่ากัน แบบทดสอบนี้ยังใช้เปรียบเทียบผลกระทบของสองกรรมวิธีทดลอง โดยการทดลอง ณ ระดับต่างๆ หรือเมื่อทั้งสองกรรมวิธีประยุกต์กับกลุ่มตัวอย่างที่เป็นอิสระกันตั้งแต่ 2 กลุ่มขึ้นไป โดยที่กลุ่มตัวอย่างนี้ใช้เป็นชั้นภูมินั่นเอง

สมมติฐานหลักที่จะทดสอบก็จะเป็นเช่นเดียวกับแบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์ นั่นคือ

Ho : สองประชากรมีการแจกแจงเหมือนกัน

หรือ Ho : สองตัวอย่างมาจากประชากรเดียวกัน

ข้อมูลจากตัวอย่างที่ใช้ในการตัดสินใจเกี่ยวกับสมมติฐานหลัก สรุปได้ดังนี้

ตัวอย่าง		1	2
ชั้นภูมิ	1	$X_{111}, X_{112}, \ldots, X_{11n1}$	$X_{121}, X_{122}, \ldots, X_{12n1}$
	2	$X_{211}, X_{212}, \ldots, X_{21n2}$	$X_{221}, X_{222}, \ldots, X_{22n2}$
	:		
	b	$X_{b11}, X_{b12}, \ldots, X_{b1n_b}$	$X_{b21}, X_{b22}, \ldots, X_{b2n_b}$

ในเมื่อ b เป็นจำนวนชั้นภูมิ หรือกลุ่มในแต่ละตัวอย่าง $n_1, n_2, \ldots, n_b$ เป็นจำนวนค่าสังเกตในแต่ละชั้นภูมิของแต่ละตัวอย่าง X_{ijk} เป็นค่าสังเกตที่ k ของตัวอย่างที่ j ในชั้นภูมิที่ i (i = 1, 2, .., b ; j = 1, 2 ; k = 1, 2, .., n_i) เช่น X_{125} เป็นค่าสังเกตตัวที่ 5 ของตัวอย่าง 2 ที่อยู่ในชั้นภูมิ 1 เป็นต้น

ตัวสถิติทดสอบขึ้นอยู่กับอันดับที่ของค่าสังเกต สำหรับอันดับที่ของค่าสังเกตหาได้ดังนี้ - รวมค่าสังเกตในแต่ละชั้นภูมิ แล้วเรียงค่าสังเกตจากน้อยไปมาก ให้อันดับ 1 แก่ค่าสังเกตที่น้อยสุด อันดับ 2 แก่ค่าที่น้อยสุดรองลงมา และต่อๆ ไป จนถึงอันดับ $2n_i$ แก่ค่าที่มากสุด ถ้าค่าสังเกตเท่ากันก็ให้อันดับที่เฉลี่ย

ผลรวมของอันดับที่ในตัวอย่าง 1 และตัวอย่าง 2 จะให้เป็น R_1 และ R_2 ตามลำดับ แล้วตัวสถิติทดสอบจะกำหนดไว้เป็น

$$R = \min (R_1 , R_2)$$

ถ้าค่าของตัวสถิติทดสอบ R น้อยมาก จะแสดงว่าสมมติฐานหลักไม่น่าจะเป็นจริง นั่นคือสองตัวอย่างหรือสองกรรมวิธีน่าจะไม่แตกต่างกัน ค่าวิกฤตของแบบทดสอบนี้กำหนดไว้ดังตาราง 15 ตารางนี้กำหนดให้จำนวนค่าสังเกตในแต่ละชั้นภูมิเท่ากัน กรณีที่จำนวนค่าสังเกตในแต่ละชั้นภูมิหรือจำนวนชั้นภูมิมากกว่า 7 หรือเมื่อจำนวนค่าสังเกตในแต่ละชั้นภูมิไม่เท่ากัน ก็ใช้การประมาณค่าแบบตัวอย่างขนาดโต นั่นคือใช้ตัวสถิติทดสอบ Z ซึ่งมีการแจกแจงแบบปกติมาตรฐาน N(0,1)

$$Z = \frac{\Sigma ni(2n_i - 1) - 2R}{\sqrt{\Sigma n_i^2(2n_i + 1)/3}}$$

ถ้า $n_1 = n_2 = \ldots = n_b = n$ แล้วตัวสถิติทดสอบจะเป็น

$$Z = \frac{bn(2n + 1) - 2R}{\sqrt{bn^2(2n + 1)/3}}$$

ตัวอย่าง นักกีฏวิทยาต้องการเปรียบเทียบยาฆ่าแมลง 2 แบบ ที่ผลิตขึ้นมาโดยทำการทดลองกับระดับความเข้มข้นของ ดี ดี ที 3 ระดับ ได้ข้อมูลซึ่งเป็นเปอร์เซ็นต์ที่แมลงถูกฆ่าตาย ดังนี้

ยาฆ่าแมลง		1					2				
	10%	18	30	26	50	53	34	52	42	63	67
ความเข้มข้น	20%	33	44	42	44		60	66	62	80	
	30%	44	56	50			74	84	77		

สมมติฐานที่จะทดสอบจะเป็นดังนี้

Ho : ยาฆ่าแมลง 2 แบบ ไม่มีความแตกต่างกัน

Ha : ยาฆ่าแมลง 2 แบบ แตกต่างกัน

อันดับที่ของค่าสังเกตในแต่ละชั้นภูมิ จะเป็นดังนี้

10%	1	3	2	6	8	4	7	5	9	10
20%	1	3.5	2	3.5		5	7	6	8	
30%	1	3	2			4	6	5		
	$R_1 = 36$					$R_2 = 76$				

$\Sigma n_i(2n_i-1) = 5(10-1)-4(8-1)-3(6-1) = 112$

$\Sigma n_i^2(2n_i-1) = 5^2(10-1)-4^2(8-1)-3^2(6-1) = 482$

$$Z = \frac{112 - 2(36)}{\sqrt{482/3}} = 3.16$$

ดังนั้นจึงปฏิเสธสมมติฐานหลักที่ว่ายาฆ่าแมลง2แบบไม่แตกต่างกันนั่นคือยาฆ่าแมลงแบบ 2 จะฆ่าแมลงได้ดีกว่าแบบ 1

3.3 กรณีสองตัวอย่างที่สัมพันธ์กัน (Two Related Sample Case)

การอ้างอิงเกี่ยวกับสองตัวอย่างที่สัมพันธ์กันนั้นทางสถิติพารามิเตอร์ก็อาศัยแบบทดสอบที (T-test) สำหรับข้อมูลจับคู่กัน แต่ถ้าข้อกำหนดเกี่ยวกับแบบทดสอบทีนั้นไม่เป็นจริง หรือถ้าสเกลการวัดต่ำกว่าแบบช่วง หรือต้องการผลอย่างรีบด่วน แล้วเราก็อาศัยวิธีการไร้พารามิเตอร์ ซึ่งจะมีวิธีการหรือกระบวนการต่างๆ กันดังต่อไปนี้

3.3.1 แบบทดสอบเครื่องหมายชนิดสองตัวอย่างที่สัมพันธ์กัน (Sign Test for Two Related Sample)

แบบทดสอบเครื่องหมายชนิดสองตัวอย่างที่สัมพันธ์กันนี้ใช้เปรียบเทียบมัธยฐานของสองประชากรโดยอาศัยเครื่องหมายบวกและลบ (+, –) เป็นหลักในการวิเคราะห์ สมมติฐานที่จะทดสอบจะเป็นดังนี้

ก. $Ho : M_1 = M_2$; $Ha : M_1 \neq M_2$

ข. $Ho : M_1 \leq M_2$; $Ha : M_1 > M_2$

ค. $Ho : M_1 \geq M_2$; $Ha : M_1 < M_2$

เนื่องจากแบบทดสอบเครื่องหมายนี้เป็นกรณีพิเศษของแบบทดสอบทวินามที่มี $\pi = 0.50$ เราจึงกล่าวสมมติฐานในเชิงความน่าจะเป็นของเครื่องหมายบวก (+) และลบ (–) ได้ดังนี้

ก. $Ho : P(+) = P(-) = 0.50$, $Ha : P(+) \neq P(-) = 0.50$

ข. $Ho : P(+) \geq P(-)$, $Ha : P(+) < P(-)$

ค. $Ho : P(+) \leq P(-)$, $Ha : P(+) > P(-)$

ในการทดสอบสมมติฐานเหล่านี้ก็อาศัยตัวอย่างขนาด n คู่ จากประชากรที่สนใจ และมีค่าสังเกตเป็น $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$, ในเมื่อแต่ละคู่ของค่าสังเกตได้จากหน่วยเดียวกันหรือหน่วยที่จับคู่กันโดยเกณฑ์อย่างหนึ่ง ค่าสังเกต n คู่นี้จะต้องเป็นอิสระกัน และเป็นค่าของตัวแปรแบบต่อเนื่อง สเกลการวัดอย่างน้อยต้องเป็นแบบอันดับ เพื่อใช้พิจารณาว่าสมาชิกในแต่ละคู่นั้นสมาชิกไหนมากกว่า

กำหนดตัวแปร Z_i ซึ่งเป็นผลต่างระหว่าง X_i กับ Y_i ดังนี้

$$Z_i = X_i - Y_i \; ; i = 1, 2, .., n$$

และกำหนดตัวแปรดัชนี φ_i ไว้ว่า $\varphi_i = 0$ ถ้า $Z_i < 0$ และ $\varphi_i = 1$ ถ้า $Z_i > 0$

ดังนั้นตัวสถิติทดสอบจึงกำหนดไว้ว่า

$$T = \sum_{i}^{n} \varphi_i$$

ตัวสถิติ T นี้มีการแจกแจงทวินามที่มีพารามิเตอร์ n_o และ $\pi = 0.50$ ถ้า Ho เป็นจริง ในเมื่อ n_o เป็นจำนวนคู่ที่ $X_i \neq Y_i$ (ส่วนมาก $n_o = n$)

เกณฑ์ตัดสินใจจะกำหนดไว้ตามสมมติฐานรอง Ha ดังนี้

ก. สำหรับ $Ha : M_1 \neq M_2$ จะปฏิเสธ Ho ถ้า $T \leq t$ หรือ $T \geq n_o - t$ ณ ระดับนัยสำคัญ $\alpha = 2\alpha_1$ ในเมื่อ $P(Y \leq t) = \alpha_1 \cong \alpha/2$ โดยที่ t เป็นค่าจากตารางทวินามที่มีพารามิเตอร์ n_o และ $\pi = 0.50$

ข. สำหรับ $Ha : M_1 > M_2$ จะปฏิเสธ Ho ถ้า $T \leq n_o - t$ ณ ระดับนัยสำคัญ α

ค. สำหรับ $Ha : M_1 < M_2$ จะปฏิเสธ Ho ถ้า $T \leq t$ ณ ระดับนัยสำคัญ $\alpha_1 \cong \alpha$

กรณีตัวอย่างขนาดโต ($n_o \geq 20$) เราใช้ตัวสถิติทดสอบ Z ซึ่งมีการแจกแจงปกติมาตรฐาน ดังนี้

$$Z = \frac{(T \pm 0.50) - n_o/2}{\sqrt{n_o/4}}$$

บางครั้งเราต้องการทดสอบสมมติฐานหลักที่ว่า

$$Ho : M_1 - M_2 = M_0$$

เราจะอาศัยค่า $Z'_i = Z_i - M_c$; $i = 1, 2, .., n$ และตัวสถิติทดสอบ T ก็ใช้ค่า Z'_i เหล่านี้ แทน Z_i แล้วดำเนินการเช่นเดียวกับที่กล่าวมาแล้ว

ตัวอย่าง สินค้าชนิดหนึ่งผลิตโดยวิธีการเดิม ต่อมาวิธีการผลิตใหม่ ผู้ผลิตประสงค์จะศึกษาว่าผู้บริโภคที่ใช้วิธีการผลิตใหม่กว่าที่ผลิตโดยวิธีการเดิมหรือไม่ จากการสุ่มผู้บริโภค 10 คน และให้สินค้าที่ผลิตจากวิธีการใหม่และเก่าไปใช้ในระยะหนึ่ง แล้วก็สอบถามว่าชอบแบบไหน ปรากฏว่าชอบสินค้าที่ผลิตจากวิธีการใหม่ 8 ราย ชอบพอๆ กัน 1 ราย และชอบวิธีการเดิม 1 ราย

สมมติฐานที่จะทดสอบจะเป็น

Ho : สินค้าที่ผลิตโดยวิธีการใหม่กับวิธีการเก่าได้รับความนิยมเท่าๆ กัน

Ha : สินค้าที่ผลิตโดยวิธีการใหม่ได้รับความนิยมมากกว่า

ให้ + แทนเหตุการณ์ที่วิธีการผลิตใหม่ดีกว่าเก่า และ – แทนเหตุการณ์ที่วิธีการเก่าดีกว่าใหม่ แล้วตัวสถิติทดสอบ T จะเป็นจำนวนเครื่องหมาย + จากการทดลองได้ + เป็น 8 และ – เป็น 1 ดังนั้น $n_o = 8 + 1 = 9$

เขตวิกฤตขนาด $\alpha_1 = 0.0195$ จะสมนัยกับ $t = 1$ หรือ $T \geq n_o - t = 9-1 = 8$

ค่าของตัวสถิติทดสอบ T เท่ากับ 8 จะอยู่ในเขตวิกฤต จึงปฏิเสธ Ho ณ ระดับนัยสำคัญ 0.0195 แสดงว่าสินค้าที่ผลิตโดยวิธีการใหม่จะได้รับความนิยมจากลูกค้ามากกว่าที่ผลิตจากวิธีการเดิม

3.3.2 ช่วงเชื่อมั่นสำหรับผลต่างมัธยฐานที่ขึ้นอยู่กับแบบทดสอบเครื่องหมาย (Confidence Interval for Median Difference)

กระบวนการในการสร้างช่วงเชื่อมั่น $100(1-\alpha)\%$ ชนิดสองทางที่สมมาตรกันสำหรับผลต่างมัธยฐาน (M_1-M_2) จะเหมือนกับกระบวนการซึ่งใช้สร้างช่วงเชื่อมั่นสำหรับมัธยฐานที่กล่าวมาแล้วนั่นเอง กระบวนการสรุปได้ดังนี้

(1) จากแต่ละคู่ของค่าสังเกต พิจารณา $D_i = X_i - Y_i$; $i = 1, 2, ..., n$

(2) จากตารางทวินามที่มีพารามิเตอร์ n และ $\pi = 0.50$ เราหาค่า t ที่ทำให้ $P(Y \leq t) \leq \alpha/2$

(3) เรียงค่า D_i ตามขนาดได้เป็น $D_{(1)} \leq D_{(2)} \leq .. \leq D_{(n)}$

ตัวประมาณค่าแบบจุดของ $\theta = M_1 - M_2$ กำหนดไว้ว่า

$\hat{\theta}$ = มัธยฐาน $\{D_i ; 1 \le i \le n\}$

(4) ช่วงเชื่อมั่นที่มีสัมประสิทธิ์เชื่อมั่นเท่ากับหรือมากกว่า 1- α จะกำหนดไว้เป็น

$$D_{(t+1)} \le M_1 - M_2 \le D_{(n-t)}$$

ตัวอย่าง ในการศึกษาผลต่างระหว่างมัธยฐานของกลุ่มที่ควบคุม กับกลุ่มที่ใช้กรรมวิธีทดลอง โดยอาศัยการจับคู่หน่วยทดลองตัวอย่าง 12 คู่ เมื่อทำการวัดหลังจากทดลอง ได้ข้อมูลมาดังนี้

กลุ่มทดลอง X	7.9	9.1	9.2	8.1	4.2	7.2	5.4	4.9	6.6	4.7	5.2	7.3
กลุ่มควบคุม Y	10.2	10.2	11.5	8.0	6.6	7.4	7.7	7.2	8.2	6.2	6.0	8.7

ค่าของผลต่าง D_i ที่เรียงตามขนาดแล้วจะเป็นดังนี้ -0.1 0.2 0.8 1.1 1.4 1.5 1.6 2.3 2.3 2.3 2.3 2.4

ตามตารางทวินามที่มี n = 12 และ π = 0.50 เราได้ $P(Y \le 2) = 0.0193$ ดังนั้นช่วงเชื่อมั่น 100(1-2(0.0193)) = 96.14% สำหรับผลต่างมัธยฐาน (M_1-M_2) เป็น

$D_{(3)} \le M_1 - M_2 \le D_{(10)}$

$0.8 \le M_1 - M_2 \le 2.3$

ตัวประมาณค่าแบบจุดของ $\theta = M_1-M_2$ ก็คือ

$\hat{\theta}$ = มัธยฐาน $\{D_i , 1 \le i \le 12\}$

$= (D_{(6)} + D_{(7)})/2 = (1.5 + 1.6)/2$

$= 1.55$

3.3.3 แบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย (Wilcoxon Matched-Pairs Signed-Ranks Test)

ในการทดสอบผลต่างมัธยฐานที่ใช้แบบทดสอบเครื่องหมายนั้นเราใช้กับค่าสังเกตที่มีสเกลการวัดแบบอันดับ ถ้าค่าสังเกตที่วัดได้มีข้อมูลข่าวสารมากกว่า หรือมีสเกลการวัดสูงกว่าแบบอันดับ นั่นคือมีสเกลการวัดเป็นแบบช่วง หรือแบบอัตราส่วน แล้วแบบทดสอบเครื่องหมายก็ไม่เป็นแบบทดสอบที่ดี แบบทดสอบที่ดีกว่าก็คือ แบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย นั่นเอง

สมมติฐานเกี่ยวกับผลต่างมัธยฐานที่จะทดสอบ จะเป็นดังนี้

(ก) $Ho : M_1 = M_2$ หรือ $Ho : M_D = 0$

$Ha : M_1 \ne M_2$ หรือ $Ha : M_D \ne 0$

(ข) $Ho : M_1 \le M_2$; $Ha : M_1 > M_2$

(ค) $Ho : M_1 \ge M_2$; $Ha : M_1 < M_2$

ข้อมูลที่ใช้วิเคราะห์จะประกอบด้วย n ค่า ของผลต่าง $Z_i = X_i - Y_i$ จากแต่ละคู่ของค่าสังเกต (X_i, Y_i) ที่วัดจากหน่วยเดียวกัน หรือจากหน่วยที่จับคู่กันโดยเกณฑ์อย่างน้อยหนึ่งอย่าง ผลต่าง Z_i นั้นจะแทนค่าสังเกตที่เป็นอิสระกันของตัวแปรเชิงสุ่มแบบต่อเนื่อง การแจกแจงในประชากรของผลต่างจะสมมาตร และสเกลการวัดสำหรับผลต่างอย่างน้อยเป็นแบบช่วง

ให้เรียงลำดับจากน้อยไปมากของ $|Z_i|$; i = 1, 2, .., n และให้ R_i เป็นอันดับที่ของ $|Z_i|$ เมื่อกำหนดตัวแปรดัชนี

$$\varphi_i = 1 \text{ ถ้า } Z_i > 0$$
$$= 0 \text{ ถ้า } Z_i < 0$$

แล้วสถิติทดสอบจะเป็น T

$$T = \Sigma R_i \varphi_i$$

ซึ่งเป็นผลรวมของอันดับที่สำหรับเครื่องหมายบวกนั่นเอง และ T นี้จะมีค่าเฉลี่ย และความแปรปรวนดังนี้

$$E(T) = n_o(n_o + 1)/4$$
$$V(T) = n_o(n_o + 1)(2n_o + 1)/24$$

ในเมื่อ n_o เป็นจำนวนคู่ที่ $X_i \neq Y_i$

เมื่อตัวอย่างขนาดโต เราจะได้ว่า

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

มีการแจกแจงปกติมาตรฐาน และถ้าข้อมูลมีค่าเท่ากันมาก ๆ เราต้องปรับปรุง V(T) ดังนี้

$$V'(T) = n(n + 1)(2n + 1) / 24 - (1/48) \Sigma (g_i^3 - g_i)$$

ในเมื่อ g_i เป็นจำนวนที่เท่ากันของข้อมูลกลุ่มหนึ่ง ๆ และ n เป็นขนาดตัวอย่าง สำหรับ E(T) นั้นจะเท่ากับ n(n + 1)/4 นั่นคือใช้ n แทน n_o

เกณฑ์ตัดสินใจ กำหนดไว้ตามสมมติฐานรอง Ha ดังนี้

(ก) สำหรับ Ha : $M_1 \neq M_2$ จะปฏิเสธ Ho ณ ระดับนัยสำคัญ α ถ้า $T > W_{1-\alpha/2}$ หรือ $T < W_{\alpha/2}$ ในเมื่อ W_p เป็นควอนไทล์ที่ p ของตัวสถิติ T ซึ่งพิจารณาได้จากตาราง

(ข) สำหรับ Ha : $M_1 > M_2$ จะปฏิเสธ Ho ณ ระดับนัยสำคัญ α ถ้า $T > W_{1-\alpha}$

(ค) สำหรับ Ha : $M_1 < M_2$ จะปฏิเสธ Ho ณ ระดับนัยสำคัญ α ถ้า $T < W_\alpha$

สำหรับตัวอย่างขนาดโต เราประมาณ W_p ได้จากสมการ

$$W_p = E(T) + Z_p\sqrt{V(T)}$$

ในเมื่อ W_p เป็นควอนไทล์ที่ p ของการแจกแจงปกติมาตรฐาน N(0,1)

ในบางครั้งเราต้องการทดสอบสมมติฐานหลักที่ว่า

$$Ho : M_1 = M_2 = M_0$$

ในเมื่อ M_0 เป็นค่าผลต่างที่ระบุไว้ แล้วเราจะต้องปรับปรุงค่า Z_i เป็น $Z'_i = Z_i - M_0$ และทำการวิเคราะห์ค่า Z'_i นี้แบบค่า Z_i เดิมนั่นเอง

ตัวอย่าง นักจิตวิทยาได้ทดสอบทางจิตวิทยาแก่ฝาแฝด 12 คู่ เพื่อพิจารณาว่าเด็กเกิดก่อน จะก้าวร้าวกว่าเด็กที่เกิดทีหลังหรือไม่ ผลจากการทดลองได้คะแนนที่แสดงถึงความก้าวร้าว ดังนี้

ฝาแฝดคู่ที่	1	2	3	4	5	6	7	8	9	10	11	12
เกิดก่อน (Y)	86	71	77	68	91	72	77	91	70	71	88	87
เกิดหลัง (X)	88	77	76	64	96	72	65	90	65	80	81	72
Z_i	2	6	−1	−4	5	0	−12	−1	−5	9	−7	−15
อันดับที่ $\|Z_i\|$	3	7	1.5	4	5.5	–	10	1.5	5.5	9	8	11
$R_i\varphi_i$	3	7	0	0	5.5	–	0	0	0	9	0	0

สมมติฐาน

Ho : ฝาแฝดที่เกิดก่อนจะไม่ก้าวร้าวกว่าที่เกิดทีหลัง ($M_x \geq M_y$)

Ha : ฝาแฝดที่เกิดทีหลังจะก้าวร้าวน้อยกว่าที่เกิดก่อน ($M_x < M_y$)

เกณฑ์ตัดสินใจ กำหนดไว้ดังนี้ จากตารางวิลคอกซันที่มี $n_0 = 11$ และ $\alpha = .05$ เราได้จุดวิกฤตของ T เป็น 14 ดังนั้นจะปฏิเสธ Ho ณ ระดับนัยสำคัญ .05 ถ้า T < 14

เนื่องจากค่าของตัวสถิติทดสอบ T คำนวณได้เป็น

$$T = \Sigma R_i\varphi_i = 3 + 7 + 5.5 + 9 = 24.5$$

ซึ่งตกอยู่ในเขตยอมรับ จึงปฏิเสธ Ho ไม่ได้ แสดงว่าฝาแฝดที่เกิดก่อน และเกิดหลังจะมีความก้าวร้าวพอ ๆ กัน

ในการประมาณ α นั้นจะอาศัยตารางวิลคอกซัน เนื่องจาก T = 24.5 นั้นอยู่ระหว่าง 23 และ 27 ซึ่งเป็นควอนไทล์ที่ 20 หรือ $W_{.20}$ และ 30 หรือ $W_{.30}$ เราจึงได้ α เป็น

$$\hat{\alpha} \cong 20 + (.30 - .20)(24.5 - 23) / (27 - 23)$$

$$= 0.2375$$

ข้อมูลในแบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมายนี้สามารถแปลงให้เป็นคะแนนปกติที่เป็นบวก (Positive Normal Scores) เช่นเดียวกับอันดับที่ R_i ได้ เราจะขอกล่าวเฉพาะคะแนนปกติแบบ แวน เดอร์ แวร์เดน แล้วตัวสถิติทดสอบจะเป็น

$$T = \Sigma Z^{(i)}\varphi_i$$

ในเมื่อ $Z^{(i)}$ เป็นคะแนนปกติสัมบูรณ์ (Absolute Normal Scores) แบบ แวน เดอร์ แวร์เดน

เมื่อตัวอย่างขนาดโต เราใช้การแจกแจงปกติประมาณค่า นั่นคือจะได้ตัวสถิติ Z ซึ่งการแจกแจงปกติมาตรฐาน ดังนี้

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

ในเมื่อ E(T) และ V(T) กำหนดไว้เป็น

$$E(T) = (1/2)\Sigma Z_{(i)} \quad \text{และ} \quad V(T) = (1/4)\Sigma(Z^{(i)})^2$$

ตัวอย่าง ในการเปรียบเทียบวิธีการผลิตสินค้าแบบใหม่กับแบบเก่า โดยอาศัยตัวอย่างคนคุมเครื่องจักร 7 คน ได้ข้อมูลซึ่งเป็นผลผลิตต่อนาที ดังนี้

วิธีการผลิต	วิธีใหม่	วิธีเก่า	Z_i	R_i	P_i	$Z^{(i)}$
คนคุม ก	31	36	-5	3	.375	.49
ข	42	38	4	2	.250	.32
ค	44	33	11	6	.750	1.15
ง	48	36	12	7	.875	1.53
จ	51	53	-2	1	.125	.16
a	57	49	8	4	.500	.67
ช	62	52	10	5	.625	.89
						5.21

คะแนนปกติที่เป็นบวกนั้นหาได้โดยอาศัยตาราง 13 ที่มีพารามิเตอร์ 2n +1 = 2(7)+1 = 15 นั่นคือค่าเบี่ยงเบนปกติที่เป็นบวก 7 ค่าจะเป็น 0.16, 0.32, .., 1.15, 1.53

ดังนั้น T = 0.16(0)+ 0.32(1) . . + 1.53(1)

= 4.56

สำหรับ $\alpha = 0.05$ และทดสอบทางเดียว ค่าสุดเหวี่ยงจะมีจำนวน $K = 2^n(\alpha) = 2^7(0.05) = 6$ ค่า ดังนี้

$Z^{(i)}$	0.16	0.32	0.49	0.67	0.89	1.15	1.53	ค่า T
วิธีกำหนด	+	+	+	+	+	+	+	5.21
เครื่องหมาย	-	+	+	+	+	+	+	5.05
+ และ	-	-	+	+	+	+	+	4.73
ให้แก่คา $Z^{(i)}$	-	+		+	+	+	+	4.56

เราจะเห็นได้ว่า T = 4.56 อยู่ในเซตปฏิเสธ แสดงว่าวิธีการผลิตแบบใหม่จะดีกว่าแบบเก่า

เมื่อใช้การประมาณค่าแบบตัวอย่างขนาดโต เราได้ E(T) และ V(T) เป็น

$$E(T) = (1/2)\Sigma Z^{(i)} = (1/2)(5.21) = 2.605$$

$$V(T) = (1/4)\Sigma(Z^{(i)})^2 = (1/4)\{0.16^2+0.32^2+..+1.53^2\}$$

$$= 1.3181$$

ดังนั้น

$$Z = \frac{4.56 - 2.605}{\sqrt{1.3181}} = 1.70$$

สำหรับ α = .05 เราจะปฏิเสธ Ho เพราะ $Z < Z_{.05} = 1.645$

3.3.4 ตัวประมาณค่า และช่วงเชื่อมั่นของผลต่างมัธยฐานที่ขึ้นอยู่กับแบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย

ในการประมาณค่าของ $\theta = M_1 - M_2$ แบบจุด และแบบช่วง โดยอาศัยแบบทดสอบวิลคอกซันชนิดอันดับที่เครื่องหมาย เรามีกระบวนการประมาณค่าดังนี้

(1) จากข้อมูล (X_i, Y_i) ; i = 1, 2, .., n เราหาค่า

$$Z_i = X_i - Y_i \; ; \; i = 1, 2, .., n$$

(2) หาค่าเฉลี่ยของ Z_i กับ Z_j , $1 \le i \le j \le n$ จำนวน k = n(n+1)/2 ค่า นั่นคือ

$$U_{ij} = (Z_i + Z_j) / 2 \; ; \; 1 \le i \le j \le n$$

ตัวประมาณค่าแบบจุดของ θ จะเป็น $\hat{\theta}$

$$\hat{\theta} = \text{มัธยฐาน } \{U_{ij} \; ; \; i \le j\}$$

(3) เรียงลำดับ U_{ij} จากน้อยไปมาก ซึ่งจะให้เป็น $W_{(1)} \le W_{(2)} \le .. \le W_{(k)}$

(4) จากตาราง 10 เราพิจารณา $W_{\alpha/2}$ สำหรับ n และ α ที่กำหนดให้ แล้วช่วงเชื่อมั่น 100(1-α)% สำหรับ $\theta = M_1 - M_2$ จะเป็น

$$W_{(W\alpha/2)} \le \theta \le W_{(k+1-W\alpha/2)}$$

ตัวอย่าง ในการศึกษาเพื่อประมาณผลต่างมัธยฐาน $(M_1 - M_2)$ โดยอาศัยตัวอย่าง ได้ข้อมูลมาดังนี้

X:	0.878	0.647	0.598	2.05	1.06	1.29	1.06	3.14	1.29
Y:	1.83	0.50	1.62	2.48	1.68	1.88	1.55	3.06	1.30

เราได้ค่า Z_i เป็นดังนี้

-.952 1.47 -1.022 -.430 -.620 -.590 -.490 .080 -.010

และได้ค่า U_{ij} จำนวน 9(9+1)/2 = 45 ค่า ซึ่งเรียงตามลำดับดังนี้

-1.022	-.987	-.952	-.821	-.806	-.786	-.771	-.756
-.726	-.721	-.691	-.620	-.605	-.590	-.555	-.540
-.525	-.516	-.510	-.490	-.481	-.471	-.460	-.4375
-.436	-.430	-.4025	-.315	-.300	-.270	-.255	-.250
-.2365	-.2215	-.220	-.205	-.175	-.1715	-.1415	-.010
.035	.0685	.080	.1135	.147			

ดังนั้น $\hat{\theta} = W_{(23)} = -.460$

จากตาราง สำหรับ α = .05 และ n = 9 เราได้ $W_{\alpha/2} = 6$ และ $k + 1 - W_{\alpha/2} = 45 + 1 - 6 = 40$ นั่นคือช่วงเชื่อมั่น 95% สำหรับ $\theta = M_1 - M_2$ จะเป็น

$$W_{(6)} \le \theta \le W_{(40)}$$

$$-.786 \text{ s } \theta \le -.100$$

3.3.5 แบบทดสอบวอลช์ (Walsh Test)

แบบทดสอบนี้ใช้เปรียบเทียบค่าเฉลี่ย หรือมัธยฐานของสองประชากรที่เป็นแบบต่อเนื่องและสมมาตร ตัวอย่างที่สุ่มมาจากประชากรนั้นจะมีสหสัมพันธ์กัน และมีสเกลการวัดอย่างน้อยเป็นแบบช่วง แบบทดสอบนี้จะมีอำนาจทดสอบสูง สำหรับสมมติฐานหลักที่จะทดสอบจึงเป็น

Ho : ค่าเฉลี่ยหรือมัธยฐานของสองประชากรเท่ากัน

ให้ผลต่างของค่าสังเกตจากสองตัวอย่างที่มีขนาด n นั้นเป็น $d_i = X_i - Y_i$; i = 1, 2, .., n เรียงลำดับผลต่าง d_i จากน้อยไปมาก ซึ่งจะได้เป็น $d_{(1)}, d_{(2)}, .., d_{(k)}$ แล้วสมมติฐานหลักจะเขียนได้เป็น

Ho : ผลต่างของค่าสังเกตสุ่มมาจากประชากรที่มีมัธยฐานเป็นศูนย์

หรือ Ho : $M_d = 0$

ตัวสถิติทดสอบจะเป็นผลต่างของค่าสังเกต $d_{(i)}$ นั่นเอง สำหรับเกณฑ์ตัดสินใจตามขนาดตัวอย่าง n กำหนดได้โดยอาศัยตาราง 10 เช่นเมื่อตัวอย่างขนาด 5 ถ้าสมมติฐานรองเป็น Ha : $M_1 \neq M_2$ หรือ Ha : $M_d \neq 0$ แล้วเกณฑ์ตัดสินใจจะเป็นดังนี้ "ปฏิเสธ Ho ณ ระดับนัยสำคัญ $\alpha = .125$ ถ้า $(1/2)\{d_{(4)} - d_{(5)}\} < 0$ และ ณ ระดับนัยสำคัญ $\alpha = .062$ ถ้า $d_{(5)} = 0$" และเมื่อตัวอย่างขนาด 10 สมมติฐานรองเป็น Ha : $M_1 > M_2$ หรือ $M_d > 0$ แล้วเกณฑ์ตัดสินใจจะเป็น "ปฏิเสธ Ho ณ ระดับนัยสำคัญ $\alpha = .048$ ถ้า $\min\{d_{(5)}, \frac{1}{2}(d_{(1)} + d_{(8)})\} > 0$ และ ณ ระดับนัยสำคัญ $\alpha = .011$ ถ้า $\min\{\frac{1}{2}(d_{(1)} + d_{(6)}), \frac{1}{2}(d_{(3)} + d_{(4)})\} > 0$" เป็นต้น

ตัวอย่าง ในการศึกษาเพื่อทดสอบสมมติฐานที่ว่า "เด็กจะจำคำที่มีความหมายดีกว่าคำที่ไม่มีความหมาย" ได้อาศัยตัวอย่างของเด็ก 15 ราย โดยให้เด็กแต่ละคนเรียนคำ 10 คำ เป็นคำที่มีความหมาย 5 คำ และไม่มีความหมาย 5 คำ เมื่อเว้นระยะห่างไว้ 1 วัน ก็ให้เด็กระลึกคำต่างๆ เหล่านั้น ปรากฏว่าได้จำนวนคำที่ระลึกได้ในแต่ละประเภท ดังนี้

นักเรียน	คำที่มีความหมาย (X)	คำที่ไม่มีความหมาย (Y)	ผลต่าง (d)
1	5	2	3
2	4	2	2
3	3	0	3
4	5	3	2
5	2	3	−1
6	4	2	2
7	2	3	−1
8	2	1	1
9	4	1	3

นักเรียน	คำที่มีความหมาย (X)	คำที่ไม่มีความหมาย (Y)	ผลต่าง (d)
10	4	3	1
11	3	4	-1
12	1	2	-1
13	5	2	3
14	3	4	-1
15	1	0	1

Ho : เด็กจะจำคำที่มีความหมายได้ไม่แตกต่างจากคำที่ไม่มีความหมาย

Ha : เด็กจะจำคำที่มีความหมายได้ดีกว่าคำที่ไม่มีความหมาย

เมื่อเรียงค่า d_i จากน้อยไปมาก จะได้เป็น

$$-1 \quad -1 \quad -1 \quad -1 \quad -1 \quad 1 \quad 1 \quad \ldots \quad 3 \quad 3 \quad 3 \quad 3$$

เกณฑ์ตัดสินใจสำหรับ n = 15 และ $\alpha = .047$ ก็คือ ปฏิเสธ Ho ถ้า $\min\{\frac{1}{2}(d_{(1)} + d_{(12)}), (d_{(2)} + d_{(11)})\} > 0$

เนื่องจาก $d_{(1)} = -1$, $d_{(12)} = 3$, $d_{(2)} = -1$, $d_{(11)} = 2$ เราจึงได้ $\min\{\frac{1}{2}(-1+2), \frac{1}{2}(-1+3)\} = 1/2 > 0$ จึงปฏิเสธ Ho แสดงว่าเด็กจะจำคำที่มีความหมายได้ดีกว่าคำที่ไม่มีความหมาย

3.3.6 การรวมแบบทดสอบวิลคอกซันสำหรับข้อมูลชนิดบลอค (Combined Wilcoxon Test for Blocked Data)

ในการเปรียบเทียบกลุ่มหรือกรรมวิธีโดยอาศัยการวางแผนทดลองชนิดแบ่งบลอคสมบูรณ์ (Randomized Block Design) นั้นก็อาศัยแบบทดสอบฟรีดแมน ดังจะได้กล่าวต่อไป สำหรับกรณีที่มีกรรมวิธีที่จะเปรียบเทียบกันเพียงสองกรรมวิธีเท่านั้น จะสามารถใช้แบบทดสอบวิลคอกซันโดยรวมกันได้ ซึ่งมีวิธีการดังนี้

เรียงลำดับข้อมูลภายในแต่ละบลอค และให้อันดับที่แล้วคำนวณตัวสถิติวิลคอกซัน S_i , i = 1, 2, .., n ของแต่ละบลอค นั่นคือ

$$S_i = \sum_{j=1}^{n_i} R_{ij}$$

ในเมื่อ R_{ij} เป็นอันดับที่ของข้อมูลในกรรมวิธีแรกของบลอค i และ n_i เป็นจำนวนข้อมูลในบลอค i ที่อยู่ในกรรมวิธีแรก

เนื่องจากอันดับที่ 1, 2, .., N_i (N_i เป็นจำนวนข้อมูลทั้งหมดในบลอค i) นั้นแจกแจงแบบเดียวกันภายในบลอค i เมื่อสมมติฐานหลักเป็นจริง เราจึงได้ว่า

$$E(S_i) = E\{\Sigma R_{ij}\} = n_i E(R_{ij})$$
$$= n_i\{N_i+1)/2\}$$

และ

$$V(S_i) = V(\Sigma R_{ij}) = n_i V(R_{ij})(N_i-n_i)/(N_i-1)$$
$$= n_i\{(N_i^2 - 1)/12\}(N_i-n_i)/(N_i-1)$$
$$= n_i(N_i-n_i)\ (N_i+1)/12$$

ความจริงผลลัพธ์สุดท้ายนี้ขึ้นอยู่กับการกำหนดอันดับที่ให้แก่ตัวอย่างขนาด n_i ที่เลือกสุ่มจากประชากรขนาดจำกัด N_i ซึ่งจะมีการแก้ไขประชากรชนิดจำกัด (Finite Population Correction, fpc) เป็น $(N_i - n_i) / (N_i - 1)$

ดังนั้นเมื่อเรารวมระหว่างบลอค แล้วจะได้

$$S = \sum_{i=1}^{b} S_i$$

$$E(S) = \Sigma E(S_i) = \Sigma n_i(N_i+1)/2$$

$$V(S) = \Sigma V(S_i) = \Sigma n_i(N_i-n_i)(N_i+1)/12$$

เมื่อขนาดตัวอย่างทั้งหมดเท่ากัน นั่นคือ $n_i = n$ และ $N_i = N$ เราจะได้

$$E(S) = bn(N+1)/2$$

$$V(S) = bn(N+1)(N-n)/12$$

ในทั้งสองกรณีจะได้ Z โดยไม่แก้ไขความต่อเนื่อง ดังนี้

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

ซึ่งมีการแจกแจงแบบปกติมาตรฐาน N(0,1) ถ้าสมมติฐานหลัก Ho เป็นจริง

ตัวอย่าง ในการเปรียบเทียบวิธีสอนสะกดคำ 2 วิธี กับเด็กที่แบ่งตามระดับสติปัญญา ได้ข้อมูลซึ่งเป็นจำนวนคำที่สะกดถูกจากคำที่กำหนดให้ 70 คำ ดังนี้

วิธีสอน	วิธี 1	วิธี 2	ค่าเฉลี่ย
ระดับสติปัญญา			
-95	20 (2)	39 (5)	
	23 (3)	38 (4)	30.33
	17 (1)	45 (6)	
96-105	31 (3)	40 (4)	
	28 (2)	47 (5)	37.00
	24 (1)	52 (6)	
106 -	45 (3)	53 (4)	
	29 (1)	59 (5)	
	35 (2)	66 (6)	

(ตัวเลขในวงเล็บเป็นอันดับที่ของข้อมูล)

เนื่องจาก $B = 3$, $N_b = 6$, $n_b = 3$ และ $S_1 = 2+3+1$, $S_2 = 3+2+1$, $S_3 = 2+1+3$ จึงได้

$$S = S_1 + S_2 + S_3 = (2+3+1)+(3+2+1)+(2+1+3)$$

$$= 18$$

$E(S) = bn(N+1)/2 = 3(3)(6+1)/2 = 31.5$

$V(S) = bn(N+1)(N-n)/12 = 3(3)(6+1)(6-3)/12 = 15.75$

$$Z = \frac{(S+1/2)-E(S)}{\sqrt{V(S)}} \quad S < E(S)$$

$$= (18.5 - 31.5)/\sqrt{15.75} = -3.27$$

เมื่อ **a** = .05 จึงปฏิเสธสมมติฐานหลัก Ho ($Z < -1.96$)

THE MEANS MUST JUSTIFY THE END.

M. Gandhi

3.3.7 แบบทดสอบฮอดจ์-เลฮ์แมนชนิดสองกรรมวิธีสำหรับค่าสังเกตที่ปรับปรุง (Two - Treatment Hodges - Lehman Test for Aligned Observations)

แบบทดสอบฮอดจ์-เลฮ์แมนนี้ได้ใช้ประโยชน์ของการเปรียบเทียบระหว่างบลอค (Interblock) ซึ่งแบบทดสอบวิลคอกซันไม่ได้สนใจ และยังใช้ประโยชน์ของการเปรียบเทียบภายในบลอค (Intrablock) ซึ่งพิจารณาจากแบบทดสอบวิลคอกซันชนิดรวมกัน (Combined Wilcoxon Test) ถ้าค่าสังเกตแตกต่างกันเนื่องจากบลอค และถ้าความแตกต่างของบลอคเหล่านี้ไม่ได้พิจารณาก่อนให้อันดับที่ แล้วผลกระทบเนื่องจากกรรมวิธีและบลอคจะปะปนกัน (Confounded) ในอันดับที่ซึ่งกำหนดให้แก่ค่าสังเกต ตัวสถิติฮอดจ์-เลฮ์แมนจะขจัดปัญหาเหล่านี้โดยการปรับปรุงค่าสังเกตในค่าเฉลี่ยหรือค่ากลางของบลอคก่อนที่จะให้อันดับที่

สำหรับตัวแบบการวิเคราะห์ความแปรปรวนที่ไม่มีผลร่วมระหว่างกรรมวิธีและบลอค จะเขียนค่าสังเกตได้เป็น

$$X_{ijl} = \mu + \alpha_{.j} + \beta_{i..} + \varepsilon_{ijl}$$

ในเมื่อ $i = 1, 2, ..., b$; $j = 1, 2,$ และ $l = 1, 2, .., n_{ij}$

ภายใต้ตัวแบบนี้ค่าสังเกตที่ปรับปรุงจะเป็น

$$X^{*}_{ijl} = X_{ijl} - \beta_{i..} = \mu + \alpha_{.j} + \varepsilon_{ijl}$$

ซึ่งเราจะเห็นว่าเป็นอิสระจากผลกระทบของบลอค สำหรับตัวอย่าง เราประมาณค่า $\beta_{i..}$ จาก

$$\hat{\beta}_{i..} = \bar{X}_{i..} - \bar{X}_{...}$$

ดังนั้น X^{*}_{ijl} ประมาณได้จาก

$$\hat{X}^{*}_{ijl} = X_{ijl} - \bar{X}_{i..} + \bar{X}_{...} = X_{ijl} - \hat{\beta}_{i..}$$

แบบทดสอบที่แท้จริงของสมมติฐานที่ว่า สองกรรมวิธีไม่แตกต่างกัน นั้นสามารถหาได้โดยอาศัยหลักการสุ่ม (Randomization Basis) เช่นเดียวกับแบบทดสอบอื่นๆ เช่นแบบทดสอบ

การสุ่มของฟิชเชอร์ เป็นต้น อย่างไรก็ตามค่าสังเกตถึงแม้จะมีจำนวนไม่มาก แต่จะต้องคำนวณค่าวิกฤตของตัวสถิติ W มากมาย จึงจำเป็นต้องใช้การประมาณค่าแบบปกติ ซึ่งทำได้อย่างรวดเร็ว

สำหรับการประมาณค่าด้วยตัวอย่างขนาดโตนั้น ความเป็นอิสระของค่าสังเกตระหว่างบลอคสามารถใช้พัฒนาการแจกแจงหลัก (Null Distribution) สำหรับตัวสถิติ W ได้ ถ้าแต่ละ W_i แทนผลรวมอันดับที่ในกรรมวิธี 1 ของบลอค i แล้ว

$$W = W_1 + W_2 + \dots + W_b$$

เป็นผลรวมของ b ตัวแปรเชิงสุ่มที่เป็นอิสระกัน ถ้า R_{ij} แทนอันดับที่ของค่าสังเกตปรับปรุงที่ I ในบลอค i (I = 1, 2, .., N_i) และ n_i แทนจำนวนของอันดับที่ในบลอค i ซึ่งอยู่ในกรรมวิธี 1 เมื่อการกำหนดอันดับที่ทั้งหมดภายในบลอคนั้นมีโอกาสเท่าๆ กัน แล้วเราจะได้

(1) $E(R_{ij}) = (1/N_i)\sum_{j}^{N_i} R_{ij} = \bar{R}_{i}$

(2) $V(R_{ij}) = (1/N_i)\sum(R_{ij} - \bar{R}_{i.})^2 = \{N_i\sum R^2_{ij} - (\sum R_{ij})^2\}/N^2$

(3) $E(W_i) = n_i E(R_{ij}) = n_i\bar{R}_{i.}$

(4) $V(W_i) = nV(R_{ij}) = n_i\{\dfrac{N_i\sum R^2_{ij} - (\sum R_{ij})^2}{N_i^2}\}\dfrac{N_i - n_i}{N_i - 1}$

(5) $E(W) = \sum E(W_i) = \sum n_i R_{i.}$

(6) $V(W) = \sum V(W_i) = V(W_1) + V(W_2) + \dots + V(W_b)$

เมื่อ b มีขนาดโต การแจกแจงหลักของ

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

จะมีการแจกแจงปกติมาตรฐาน ในเมื่อ W อยู่ในสเกลไม่ต่อเนื่อง จึงแก้ไขความต่อเนื่องด้วย ± .5 นั่นคือถ้า W<E(W) ก็ใช้ .5 แต่ถ้า W>E(W) ก็ใช้ –.5 ฮอดจ์ และเลย์แมน (1962) ได้แสดงว่าความคลาดเคลื่อนจากการใช้การประมาณแบบปกติจะผันแปรอยู่ระหว่าง .001 และ .0001 เมื่อมี 5 บลอค

ตัวอย่าง สำหรับข้อมูลในแบบทดสอบวิลคอกซันที่รวมกันนั้น เมื่อเราปรับปรุงด้วยค่าเฉลี่ยของบลอค และกำหนดอันดับที่ให้ แล้วจะเป็นดังนี้

	กรรมวิธี 1			**กรรมวิธี 2**		
บลอค 1	28.06	31.06	25.06	47.06	46.06	53.06
	(5)	(7)	(2)	(13)	(12)	(16)
บลอค 2	32.39	29.39	25.39	41.39	48.39	53.39
	(8)	(6)	(3)	(10)	(14)	(17)
บลอค 3	35.56	19.56	25.56	43.56	49.56	56.56
	(9)	(1)	(4)	(11)	(15)	(18)

ในเมื่อ $\bar{X}_{1..} = 30.33$, $\bar{X}_{2..} = 37.00$, $\bar{X}_{3..} = 47.83$ สำหรับผลกระทบกรรมวิธีที่ประมาณได้ จะเป็นดังนี้

$\hat{\beta}_{1..} = 30.33 - 38.39 = -8.06$

$\hat{\beta}_{2..} = 37.00 - 38.39 = -1.39$

$\hat{\beta}_{3..} = 47.83 - 38.39 = 9.44$

ค่าสังเกตที่ปรับปรุงสำหรับคะแนนแรกในตาราง จะเป็น $\hat{X}_{111}$

$\hat{X}_{111} = 20 - (-8.06) = 28.06$

ส่วนคะแนนอื่นจะหาได้ทำนองเดียวกัน ซึ่งสรุปได้ดังตารางที่แล้วมานี้

เนื่องจากความแตกต่างได้จำกัดออกไปแล้ว จึงกำหนดอันดับที่ให้แก่สองตัวอย่าง (รวมกัน) ดังแบบทดสอบวิลคอกซัน จากอันดับที่เหล่านี้ จะได้

$$W = \Sigma W_i = (5+7+2) + (8+6+3) + (9+1+4) = 45$$

ตัวสถิติที่จำเป็นสำหรับการประมาณค่าแบบปกติ สรุปได้ดังตารางต่อไปนี้

บลอค	N_i	n_i	R_{ij}	$E(W_i)$	$V(W_i)$
1	6	3	2, 5, 7, 12, 13, 16	27.50	42.85
2	6	3	3, 6, 8, 10, 14, 17	29.00	40.00
3	6	3	1, 4, 9, 11, 15, 18	29.00	62.20

ตัวสถิติเหล่านี้ขึ้นอยู่กับอันดับที่ของค่าสังเกตที่ปรับปรุงจากตารางที่ผ่านมานั้น

ในบลอค 1 $E(R_{1i}) = (2+5+...+16)/6 = 55/6 = 9.17$

$V(R_{1i}) = (1/6^2)\{(2^2+5^2+...+16^2)-(2+5+...+16)^2\}$

$= 23.81$

$E(W_1) = 3(9.17) = 27.50$

$V(W_1) = 3(23.81)\,(6-3)/(6-1) = 42.85$

สำหรับบลอคอื่นๆ ก็หาได้เช่นเดียวกัน แล้วจะได้

$$E(W) = \Sigma E(W_i) = 27.50 + 29.00 + 29.00 = 85.50$$

$$V(W) = \Sigma V(W_i) = 42.85 + 40.00 + 62.20 = 145.05$$

$$Z = (45+0.5 - 85.50) / \sqrt{145.05} = -3.32$$

เมื่อ a =.05 จะได้ค่าวิกฤตแบบสองทางเป็น ± 1.96 จึงปฏิเสธสมมติฐานหลัก Ho เช่นเดียวกับแบบทดสอบวิลคอกซันชนิดรวมกัน

การทดสอบเกี่ยวกับการเท่ากันของ k การแจกแจงโดยอาศัยอันดับที่นั้นจะเป็นแบบทดสอบที่มีอำนาจเมื่อกราฟแท่งตัวอย่าง (Sample Histogram) ของความคลาดเคลื่อนเป็นแบบที่แบนกว่าปกติ (Platykurtic) ถ้าการแจกแจงของความคลาดเคลื่อนเป็นแบบที่สูงกว่าปกติหรือเบ้ (Leptokurtic or Skewed) แล้วประสิทธิภาพของแบบทดสอบนี้เมื่อเทียบกับแบบทดสอบเอฟ ในการวิเคราะห์ความแปรปรวน (ANOVA) สามารถจะปรับปรุงได้โดยการแทนอันดับที่ด้วยคะแนนปกติ

ฮอดจ์ และ เลฮ์แมน ได้แสดงว่า สำหรับทุกการแจกแจง ประสิทธิภาพของรูปฟอร์มอันดับที่ในแบบทดสอบฮอดจ์-เลฮ์แมน นั้นจะไม่น้อยกว่า E = 0.864 การแจกแจงส่วนมากในการวิจัยพฤติกรรมศาสตร์นั้นประสิทธิภาพของแบบทดสอบฮอดจ์-เลฮ์แมนในรูปของค่าสังเกตที่ปรับปรุงหนึ่ง เมื่อเปรียบเทียบกับแบบทดสอบเอฟ ในกรณีที่ k = 2 แล้ว E = 0.955 เมื่อเปรียบเทียบกับแบบทดสอบเอฟ ถ้าข้อสมมติถูกต้อง ยิ่งกว่านั้นประสิทธิภาพสัมพันธ์ (ARE, Asymptotic Relative Efficiency) ของแบบทดสอบฮอดจ์-เลฮ์แมนจะเพิ่มขึ้นเมื่อจำนวนกรรมวิธีหรือเงื่อนไขเพิ่มขึ้น

แบบทดสอบฮอดจ์-เลฮ์แมนชนิดค่าสังเกตที่ปรับปรุงนี้สามารถจะปรับปรุงได้โดยอาศัยการแทนอันดับที่ด้วยคะแนนปกติ ถ้าเป็นที่เชื่อแน่ว่าตัวแปรที่สนใจนั้นมีการแจกแจงชนิดหางยาว (long tails)

เราจะแทนอันดับที่ด้วยคะแนนปกติของ แวน เดอร์ แวร์เดน ดังนี้

บลอค	อันดับที่	คะแนนปกติ	$E(W_i)$	$V(W_i)$
1	2, 5, 7,	−1.25, −0.63, −0.34	−0.2000	1.0183
	12, 13, 16	0.34, 0.48, 1.00		
2	3, 6, 8,	−1.00, −0.48, −0.20	0.1350	0.9668
	10, 14, 17	0.07, 0.63, 1.25		
3	1, 4, 9,	−1.62, −0.80, −0.07	0.0650	1.9713
	11, 15, 18	0.20, 0.80, 1.62		

ตัวสถิติที่จำเป็น หาได้ดังนี้

$$W = (-0.63 - 0.34 - 1.25) + (-0.20 - 0.84 - 1.00) + (-0.07 - 1.62 - 0.80)$$
$$= -6.39$$

$$E(W) = -0.2000 + 0.1350 + 0.0650 = 0.0000$$

$$V(W) = 1.0183 + 0.9668 + 1.9713 = 3.9564$$

$$Z = (-6.39 - 0.0000)/\sqrt{3.9546} = -3.21$$

เนื่องจาก Z < −1.96 จึงปฏิเสธ Ho ณ α =0.05 การตัดสินใจจะเป็นเช่นเดียวกับแบบทดสอบวิลคอกซันชนิดรวมกัน และแบบทดสอบฮอดจ์-เลฮ์แมนชนิดปรับปรุงอันดับที่

จงสังเกตว่าไม่มีข้อสมมติเกี่ยวกับแต่ละ ε_{ijl} ที่ว่าแจกแจงแบบปกติ แต่สมมติว่าทั้งหมดของ ε_{ijl} เป็นอิสระกันและแจกแจงเหมือนกัน สมมติฐานที่ทดสอบนั้นจะกล่าวว่าการแจกแจงทั้งหมดจะเหมือนกัน ซึ่งอาจจะเป็นแบบปกติ สมมาตร เบ้ สูงกว่าปกติ ต่ำกว่าปกติ หรืออื่นๆ ก็ได้ สมมติฐานรองที่แบบทดสอบจะฉับไวมากก็คือการแจกแจงที่มีรูปร่างเหมือนกัน แต่ค่ากลางแตกต่างกัน ดังนั้นถ้าปฏิเสธสมมติฐานหลัก Ho ได้ แล้วการสรุปที่มีเหตุผลก็คือค่ากลางแตกต่างกัน ข้อสมมติและการสรุปผลดังกล่าวนั้นจะเป็นจริงสำหรับแบบทดสอบวิลคอกซันชนิดรวมกัน และแบบทดสอบฮอดจ์-เลฮ์แมนชนิดปรับปรุงด้วย

ข้อสังเกต ในกระบวนการปรับปรุงนั้น สามารถปรับปรุงโดยอาศัยค่าเฉลี่ยเลขคณิต มัธยฐาน แต่ถ้าข้อมูลบางตัวสุดเหวี่ยงเกินไป ก็อาศัยค่าเฉลี่ยแบบวินเซอร์ไรส์ (Winsorized Means) หรือค่าเฉลี่ย (Trimmed Means) ค่าเฉลี่ยสองแบบหลังนี้อธิบายได้ดังนี้ ถ้าค่าสังเกตที่มากสุดหรือน้อยสุดในตัวอย่างนั้นไม่น่าไว้ใจว่ามาจากประชากรที่สนใจ แล้วสามารถ

(1) แทนค่าเหล่านั้นด้วยค่าใกล้สุดกับค่าเหล่านั้น

(2) ทิ้งค่าเหล่านั้นเสีย

เมื่อใช้ตัวแบบ (1) ค่าเฉลี่ยนั้นเรียกว่า ค่าเฉลี่ยแบบวินเซอไรส์ แต่ถ้าใช้ตัวแบบ (2) ค่าเฉลี่ยนั้นจะเรียกว่า ค่าเฉลี่ยปรับปรุง

พิจารณาข้อมูลตัวอย่างที่เรียงอันดับ ต่อไปนี้

8 12 23 26 27 27 28 30 55 69

ค่าเฉลี่ยแบบวินเซอไรส์ ซึ่งแทน 8 ด้วย 12 และ 69 ด้วย 55 จะเป็น

$$\bar{X}_W = (12+12+23+...+30+55+55)/10 = 29.5$$

และค่าเฉลี่ยปรับปรุง จะเป็น

$$\bar{X}_T = (12+23+26+...+28+30+55)/8 = 28.5$$

ถ้าแทนค่าสุดเหวี่ยง 4 ค่า คือ 8 12 55 69 แล้วจะได้

$$\bar{X}_W = (23+23+23+26+...+28+30+30+30)/10 = 26.7$$

$$\bar{X}_T = (23+26+27+27+28+30)/6 = 26.8$$

ค่าเฉลี่ยทั้งสองแบบนี้จึงใช้เมื่อค่าสุดเหวี่ยงนั้นมากหรือน้อยเกินไป หรือกรณีที่ค่าสังเกตหายไป

3.4 กรณีหลายตัวอย่างที่เป็นอิสระกัน (Several Independent Samples)

ในกรณีหลายตัวอย่างที่เป็นอิสระกันนี้ก็เป็นรูปทั่วไปของกรณีสองตัวอย่างที่เป็นอิสระกัน จะได้เสนอเทคนิคไร้พารามิเตอร์เพื่อวิเคราะห์ข้อมูลจากหลายตัวอย่างที่เป็นอิสระกัน ซึ่งส่วนมากจะสนใจการทดสอบสมมติฐานที่ว่า ตัวอย่างต่าง ๆ นั้นสุ่มมาจากประชากรเดียวกัน หรือจากหลายประชากรที่มีพารามิเตอร์เท่ากัน วิธีการทดสอบเชิงพารามิเตอร์นั้นจะใช้การวิเคราะห์ความแปรปรวนทางเดียว หรือแบบทดสอบเอฟ แต่การวิเคราะห์ความแปรปรวนทางเดียวนั้นมีข้อสมมติว่า ตัวอย่างต่าง ๆ นั้นเลือกแบบสุ่มและอิสระกันจากประชากรแบบปกติที่มีความแปรปรวนเท่ากัน กระบวนการไร้พารามิเตอร์ที่ใช้แทนการวิเคราะห์ความแปรปรวนทางเดียวนั้นจะได้กล่าวต่อไปนี้ ความไว้ใจได้ของมันจะไม่ขึ้นอยู่กับข้อกำหนดเหล่านั้น

3.4.1 แบบทดสอบมัธยฐานทั่วไป (Extension of Median Test)

แบบทดสอบนี้ก็ทำหน้าที่เช่นเดียวกับแบบทดสอบมัธยฐาน แต่ใช้ประยุกต์กับตัวอย่างสุ่มอย่างน้อยสองตัวอย่าง นั่นคือแบบทดสอบมัธยฐานทั่วไปนี้จะใช้ตรวจสอบว่า หลาย ๆ ตัวอย่างมาจากประชากรที่มีมัธยฐานเท่ากันหรือไม่ ดังนั้นสมมติฐานที่จะทดสอบ จึงเป็น

Ho : ประชากรทั้งหมดมีมัธยฐานเท่ากัน

Ha : มีอย่างน้อยสองประชากรที่มีมัธยฐานต่างกัน

ในการทดสอบว่าหลายประชากร หรือหลายตัวอย่างมีมัธยฐานเหมือนกันก็โดยการสุ่มตัวอย่างจากแต่ละประชากร สมมติว่ามีประชากรที่สนใจ k ประชากร ($k \geq 2$) และตัวแปรที่สนใจจะมีสเกลการวัดอย่างน้อยแบบอันดับ แล้วสามารถสร้างตารางจรณ์ 2 x k ซึ่งมีด้านหนึ่งเป็นตัวอย่าง และอีกด้านหนึ่งเป็นประเภทของค่าสังเกตที่อยู่เหนือและใต้มัธยฐานร่วม สำหรับมัธยฐานร่วมนั้นหาได้จากค่าสังเกตทั้งหมดรวมกัน ตารางจรณ์ 2 x k ของข้อมูลจะเป็นดังนี้

ตัวอย่าง	1	2	3	...	k	รวม
เหนือมัธยฐานร่วม	0_{11}	0_{12}	0_{13}		0_{1k}	a
ใต้มัธยฐานร่วม	0_{21}	0_{22}	0_{23}		0_{2k}	b
รวม	n_1	n_2	n_3		n_k	n

ในเมื่อ $n_1, n_2, .., n_k$ เป็นขนาดตัวอย่างจากประชากร 1, 2, .., k ตามลำดับ 0_{1j}, 0_{2j} เป็นจำนวนค่าสังเกตในตัวอย่าง j ที่อยู่เหนือและใต้มัธยฐานร่วมตามลำดับ (j = 1, 2, .., k) และ a กับ b เป็นจำนวนค่าสังเกตทั้งหมดที่อยู่เหนือและใต้มัธยฐานร่วม

ตัวสถิติทดสอบ จะกำหนดไว้ดังนี้

$$T = \sum^{2,k} (0_{ij} - E_{ij})^2 / E_{ij}$$

ในเมื่อ $E_{ij} = n_j p = n_j(a/n)$ เป็นจำนวนค่าสังเกตที่คาดหวังภายใต้ Ho เป็นจริง ตัวสถิติทดสอบนี้จะมีการแจกแจงไคสแควร์ ด้วยองศาความเป็นอิสระ k-1 เนื่องจาก i = 1, 2 ตัวสถิตทดสอบ T จะเขียนใหม่ได้เป็น

$$T = (n^2/ab)\Sigma(0^2_{1j} - n_j p)^2/n_j$$
$$= (n^2/ab)\Sigma 0^2_{1j} / n_j - na/b$$

ถ้า a เกือบเท่ากับ b หรือ $a \cong b$ และไม่มีค่าสังเกตเป็นจำนวนมากที่เท่ากับมัธยฐานร่วม แล้วตัวสถิติทดสอบจะเป็น

$$T \cong \Sigma(0_{1j} - 0_{2j})^2/n_j$$

ถ้าขนาดตัวอย่าง (n_j) บางตัวอย่างเล็กเกินไป แล้วการประมาณจะไม่ถูกต้องมาก ในทางปฏิบัตินั้นการประมาณจะใช้ไม่ได้ถ้าจำนวนตัวอย่าง (k) มีขนาดตัวอย่างน้อยกว่า 10 เป็นจำนวนมากกว่า 20% แต่ถ้าขนาดตัวอย่างต่าง ๆ (n_j) ส่วนมากเกือบเท่ากันก็ใช้ได้

ตัวอย่าง นักเกษตรต้องการเปรียบเทียบวิธีการปลูกข้าวโพด 4 แบบ ว่าให้ผลผลิตแตกต่างกันหรือไม่ จากการทดลองได้ข้อมูล (ต่อไร่) มาดังนี้

วิธีปลูก 1	83	91	94	89	89	96	91	92	90	
2	91	90	81	83	84	83	88	91	89	84
3	101	100	91	93	96	95	94			
4	78	82	77	79	81	80	81	81		

ในการพิจารณาว่าผลผลิตจะแตกต่างกันเพราะวิธีปลูกหรือไม่นั้น จะใช้แบบทดสอบมัธยฐานวิเคราะห์ สมมติฐานที่จะทดสอบจึงเป็น

Ho : วิธีปลูกทั้งสี่วิธีจะให้ผลผลิตต่อไร่ไม่แตกต่างกัน

Ha : วิธีปลูกทั้งสี่วิธีจะให้ผลผลิตต่อไร่แตกต่างกัน

ข้อมูลทั้งหมด 34 จำนวน ดังนั้นมัธยฐานร่วมจะเป็นค่าเฉลี่ยของข้อมูลตัวที่ 17 และ 18 ซึ่งเรียงจากน้อยไปมากแล้ว นั่นคือมัธยฐานร่วมจะเป็น (89+89)/2 = 89 สำหรับจำนวนค่าสังเกตในแต่ละตัวอย่างที่อยู่เหนือและใต้มัธยฐานร่วม (89) จะเป็นดังตารางจรณ์ 2 x 4 ต่อไปนี้

วิธีปลูก	1	2	3	4	รวม
มากกว่า 89	6	3	7	0	16
น้อยกว่า 89	3	7	0	8	18
ขนาดตัวอย่าง	9	10	7	8	34

จะเห็นได้ว่าขนาดตัวอย่างไม่โต แต่ทว่าเกือบเท่ากัน (ไม่สนใจกรณีที่ว่ามากกว่า 20% ของ n_j ต่าง ๆน้อยกว่า 10) จึงใช้ตัวสถิติไคสแควร์ T และโดยที่ a ≅ b ค่าของตัวสถิติทดสอบจึงเป็น

$$T = (6-3)^2/9 + (3-7)^2/10 + (7-0)^2/7 + (0-8)^2/8$$
$$= 17.6$$

แต่จากตารางไคสแควร์ ด้วยองศาความเป็นอิสระ 4-1 = 3 จะได้ $\chi^{2(3)}_{.05} = 7.815$ ดังนั้นจึงปฏิเสธ Ho เพราะ T = 17.6 > 7.815 แสดงว่าวิธีปลูกทั้งสี่วิธีให้ผลผลิตต่อไร่ไม่เท่ากันหมด

ถ้าจากการทดสอบด้วยแบบทดสอบมัธยฐานทั่วไปนี้ได้ผลว่า ปฏิเสธ Ho และต้องการจะทราบต่อไปว่าประชากรไหนบ้างที่มีมัธยฐานแตกต่างกัน ทำได้โดยทำการทดสอบตัวอย่างทีละสองหรือมากกว่าด้วยแบบทดสอบมัธยฐานนี้ จนกระทั่งแยกได้ว่าประชากรไหนแตกต่างกัน อย่างไรก็ตามการกระทำเช่นนั้นจะทำให้ระดับนัยสำคัญของการทดสอบเปลี่ยนไปจากการทดสอบด้วยทุกตัวอย่าง

แบบทดสอบมัธยฐานนี้สามารถขยายให้เป็นแบบทดสอบควอนไทล์ เพื่อใช้ทดสอบสมมติฐานที่ว่า ประชากรต่าง ๆ มีควอนไทล์เท่ากันได้โดยการหาจำนวนข้อมูลที่อยู่เหนือและใต้ควอนไทล์ร่วม (Grand Quantile) และก็ใช้แบบทดสอบมัธยฐานวิเคราะห์เช่นเดียวกัน

แบบทดสอบมัธยฐานนี้ยังสามารถขยายให้ทดสอบกรรมวิธี หรือประชากรต่าง ๆ ว่าแตกต่างกันหรือไม่ โดยการแบ่งค่าสังเกตจากตัวอย่างตามช่วงต่าง ๆ ช่วงนั้นก็กำหนดได้จากควอร์ไทล์ร่วม เดไซล์ร่วม หรือเปอร์เซนไทล์ร่วม ก็ได้ แล้วใช้ตัวสถิติไคสแควร์ประยุกต์

นอกจากนี้แบบทดสอบมัธยฐานยังสามารถขยายให้วิเคราะห์การทดลองที่สลับซับซ้อนได้อีก ซึ่งจะได้กล่าวต่อไป

3.4.2 แบบทดสอบครัสคัล-วอลลิส (Kruskal-Wallis Test)

แบบทดสอบวิลคอกซัน-แมนน์-วิทนีย์ ซึ่งใช้ทดสอบสองตัวอย่างที่เป็นอิสระกันนั้น สามารถขยายออกไปใช้กับการวิเคราะห์ตัวอย่างสุ่มที่เป็นอิสระกันได้หลายตัวอย่าง ครัสคัล และ วอลลิส (1952) ได้เสนอแบบทดสอบกรณีนี้ไว้ แบบทดสอบครัสคัล-วอลลิส นี้ยังใช้วิเคราะห์ข้อมูลที่ได้จากการวางแผนทดลองชนิดสุ่มสมบูรณ์ (CRD, Completely Randomized Design) อีกด้วย ดังนั้นแบบทดสอบนี้จึงได้ชื่อว่า การวิเคราะห์ความแปรปรวนทางเดียวชนิดอันดับที่

สำหรับข้อมูลที่ได้จากการวางแผนทดลองชนิดสุ่มสมบูรณ์นั้นสามารถวิเคราะห์ได้โดยอาศัยแบบทดสอบมัธยฐาน แต่แบบทดสอบครัสคัล-วอลลิสได้ใช้ข้อมูลข่าวสารที่อยู่ในค่าสังเกตมากกว่านั่นคือตัวสถิติครัสครัล-วอลลิสนั้นเป็นฟังก์ชันของอันดับที่ของค่าสังเกตในตัวอย่างรวม แต่ตัวสถิติทดสอบมัธยฐานเพียงแต่ขึ้นอยู่กับความรู้ที่ว่าค่าสังเกตนั้นสูงหรือต่ำกว่ามัธยฐานรวมหรือไม่ ด้วยเหตุผลนี้เอง ตัวสถิติทดสอบครัสคัล-วอลลิสจึงมีอำนาจทดสอบมากกว่าตัวสถิติทดสอบมัธยฐาน แต่การคำนวณตัวสถิติจะยุ่งยากกว่า เพราะตัวสถิติจะเกี่ยวข้องกับการให้อันดับที่แก่ค่าสังเกตทั้งหมด ถ้าอันดับที่สามารถกำหนดให้แก่ค่าสังเกตได้หลายวิธี (เนื่องจากค่าสังเกตเท่ากันมาก ๆ) แล้ว ควรใช้แบบทดสอบมัธยฐานดีกว่า ทั้ง ๆ ที่สูญเสียอำนาจทดสอบ เพราะในสถานการณ์เช่นนั้นแบบทดสอบครัสคัล-วอลลิส จะให้ระดับนัยสำคัญ (ที่ประมาณได้) ต่างจากระดับนัยจริง ๆ มาก อย่างไรก็ตามถ้าปัญหาเกี่ยวกับการเท่ากันของค่าสังเกตไม่เป็นปัญหาใหญ่ก็ไม่ควรจะใช้แบบทดสอบมัธยฐานตามคำแนะนำนั้น

สมมติฐานหลัก Ho เกี่ยวกับค่าเฉลี่ยจะเป็นแบบหนึ่งแบบใด ดังนี้

(ก) Ho : ฟังก์ชันแจกแจงทั้ง k ประชากรจะเหมือนกัน

(ข) Ho : $M_1 = M_2 = \ldots = M_k$

หรือ (ค) Ho : $\tau_1 = \tau_2 = \ldots = \tau_k$

ในเมื่อ $M_1, M_2, \ldots, M_k$ เป็นค่าเฉลี่ยของประชากรต่าง ๆ และ $\tau_1, \tau_2, \ldots, \tau_k$ เป็นผลกระทบกรรมวิธี (Treatment Effect) ของกรรมวิธีทดลองต่าง ๆ โดยที่ $\Sigma\tau_j = 0$ และ $\tau_1 = M_j - M$ ซึ่ง M เป็นค่าเฉลี่ยรวม (Overall Mean)

ในการทดสอบสมมติฐานหลักก็อาศัยตัวอย่างที่เป็นอิสระขนาด $n_1, n_2, .., n_k$ จาก k ประชากร ซึ่งมีตัวแปรที่สนใจเป็นแบบต่อเนื่อง และมีสเกลการวัดอย่างน้อยเป็นแบบอันดับ และประชากรที่สนใจนั้นจะเหมือนกันทั้งหมด ยกเว้นแต่ค่าเฉลี่ยซึ่งอาจจะแตกต่างกันได้ ให้ตัวอย่างที่ j ขนาด n_j เป็น $X_1, X_2, .., X_{n_j}$; j = 1, 2, .., k แล้วข้อมูลสามารถจัดให้อยู่ในตารางทางเดียวได้เป็น

ตัวอย่าง	1	2	...	j	...	k
	X_{11}	X_{12}		X_{1j}		X_{1k}
	X_{21}	X_{22}		X_{2j}		X_{2k}
			:		:	
	$X_{n_1 1}$	$X_{n_2 2}$		$X_{n_j j}$		$X_{n_k k}$

ค่าสังเกตทั้งหมดจะเป็น $n = n_1 + n_2 + ... + n_k$ ค่า ซึ่งจะกำหนดอันดับที่ให้แก่ค่าสังเกตเหล่านั้จากน้อยไปมาก นั่นคือกำหนดอันดับ 1 ให้แก่ค่าสังเกตที่น้อยสุด อันดับ 2 แก่ค่าที่น้อยที่สุดถัดไป และต่อๆ ไป จนถึงค่าที่มากสุด คืออันดับ ถ้าอันดับที่สามารถกำหนดได้หลายวิธีเนื่องจากการเท่ากันของค่าสังเกต แล้วให้กำหนดอันดับที่เฉลี่ยแก่ค่าสังเกตที่เท่ากันนั้น

ให้ $R(X_{ij})$ แทนอันดับที่ของค่าสังกต X_{ij} และให้ผลรวมของอันดับที่ในตัวอย่างที่ j เป็น R_j นั่นคือ

$$R_j = \Sigma R(X_{ij})\ ;\ j = 1, 2, .., k$$

ตัวสถิติ R_j นี้จะใช้เป็นตัวสถิติทดสอบ สำหรับค่าเฉลี่ย R_j และค่าเฉลี่ยรวม R จะเป็นดังนี้

$$\bar{R}_j = R_j/n_j,\quad \bar{R} = \Sigma R_j/n = (n + 1) / 2$$

ดังนั้นตัวสถิติทดสอบสำหรับสมมติฐานหลัก จึงกำหนดไว้ดังนี้

$$H = \frac{12}{n(n+1)} \Sigma n_j(\bar{R}_j - \bar{R})^2$$
$$= \frac{12}{n(n+1)} \Sigma R_j^2/n_j - 3(n + 1)$$

เมื่อมีการเท่ากันของค่าสังเกต จึงต้องปรับปรุงด้วยแฟคเตอร์

$$f = 1 - \Sigma G/(n^3 - n)$$

ในเมื่อ $G = g^3 - g$ และ g เป็นจำนวนค่าสังเกตที่เท่ากันในกลุ่มหนึ่งๆ ดังนั้นตัวสถิติทดสอบที่ปรับปรุงแล้วจะเป็น

$$H_c = H/f$$

แฟคเตอร์ f นี้จะทำให้ตัวสถิติทดสอบ H มีค่ามากขึ้น ดังนั้นถ้าค่าของตัวสถิติทดสอบที่ไม่ปรับปรุงนั้นสามารถปฏิเสธสมมติฐานหลักได้ใน ณ ระดับนัยสำคัญ α แล้วก็ไม่จำเป็นต้องปรับปรุงตัวสถิติทดสอบ H นั้น

เมื่อตัวอย่างขนาดโต ตัวสถิติทดสอบ H จะมีการแจกแจงไคสแควร์ ด้วยองศาความเป็นอิสระ k-1 ดังนั้นเกณฑ์ตัดสินใจจึงอาศัยการแจกแจงนี้ แต่ถ้าขนาดตัวอย่างไม่เกิน 5 และมีกลุ่มตัวอย่างเพียง 3 กลุ่ม ก็จะอาศัยตาราง 17 ช่วยกำหนดเกณฑ์ตัดสินใจ

ตัวอย่าง ในการศึกษาวิธีปลูกข้าวโพด 4 วิธี ซึ่งใช้การวางแผนทดลองชนิดสุ่มสมบูรณ์ ได้ข้อมูลซึ่งเป็นผลผลิตต่อไร่ ดังนี้

วิธีปลูก	1		2		3		4	
	83	(11)	91	(23)	101	(34)	78	(2)
	91	(23)	90	(19.5)	100	(33)	82	(9)
	94	(28.5)	81	(6.5)	91	(23)	81	(6.5)
	89	(17)	83	(11)	93	(27)	77	(1)
	89	(17)	84	(13.5)	96	(31.5)	79	(3)
	96	(31.5)	83	(11)	95	(30)	81	(6.5)
	91	(23)	88	(15)	94	(28.5)	80	(4)
	90	(19.5)	89	(17)			81	(6.5)
			84	(13.5)				
R_j	196.5		153.0		207.0		38.5	
n_j	9		10		7		8	

สมมติฐานที่จะทดสอบจะเป็น

Ho : วิธีการปลูกทั้งสี่วิธีจะให้ผลผลิตข้าวโพดต่อไร่ไม่แตกต่างกัน

Ha : วิธีปลูกข้าวโพดบางวิธีจะให้ผลผลิตต่อไร่ต่างจากวิธีอื่น

ค่าสังเกตทั้งหมด 34 ค่า เมื่อนำมาเรียงจากน้อยไปมาก และกำหนดอันดับที่จากอันดับ 1 ถึง 34 ค่าที่เท่ากันก็ให้อันดับที่เฉลี่ย แล้วจะได้ดังตารางข้างบนนี้ ค่าของตัวสถิติจะเป็น

$$H = \frac{12}{34(34-1)}\{(196.5)^2/9 + (153.0)^2/10 + (207.0)^2/7 + (38.5)^2/8\} - 3(34+1)$$
$$= 25.46$$

สำหรับระดับนัยสำคัญ .05 เราได้ค่าวิกฤตเป็น $\chi^{2(4-1)} = 7.815$ จึงปฏิเสธ Ho นั่นคือมีวิธีปลูกข้าวโพดบางวิธีจะให้ผลผลิตต่อไร่ต่างจากวิธีอื่น

เมื่อการวิเคราะห์นำไปสู่การปฏิเสธสมมติฐานหลัก แล้วจะสามารถสร้างช่วงเชื่อมั่นพหุคูณ (Multiple Confidence Interval) สำหรับความแผกกัน (Contrast) ψ เพื่อดูความแตกต่างของกลุ่มตัวอย่างต่าง ๆ ได้ ช่วงเชื่อมั่นพหุคูณ กำหนดไว้ดังนี้

$$\varphi = \hat{\varphi} \pm \sqrt{\chi_{\alpha}^{2(k-1)} V(\hat{\varphi})}$$

ในเมื่อ $\varphi = a_1M_1 + a_2M_2 + \ldots + a_kM_k$ และ $\hat{\varphi} = a_1\bar{R}_1 + a_2\bar{R}_2 + \ldots + a_k\bar{R}_k$ โดยที่ $\Sigma a_j = 0$ ส่วน $V(\hat{\varphi}) = \frac{n(n+1)}{12}\Sigma a_j^2/n_j$

ถ้าทำการเปรียบเทียบเป็นคู่ (Pairwise Comparisons) ก็จะได้ช่วงเชื่อมั่นของ $M_u - M_v$, $u < v$ ดังนี้

$$M_u - M_v = (\bar{R}_u - \bar{R}_v) \pm \sqrt{\chi_{\alpha}^{2(k-1)}}\sqrt{\frac{n(n+1)}{12}(1/n_u + 1/n_v)}$$

แต่ถ้าจะเปรียบเทียบคู่ ๆ เท่านั้น อาจจะใช้วิธีการของทูกี้ (Tukey's Method) ซึ่งต้องมีขนาดตัวอย่างเท่ากัน นั่นคือช่วงเชื่อมั่นจะเป็น

$$M_u - M_v = (\bar{R}_u - \bar{R}_v) \pm q(\alpha,k,\infty)\sqrt{\frac{n(n+1)}{12}(1/n_u)}$$

ในเมื่อ $n_u = n_v$ และ $q(\alpha,k,\infty)$ เป็นค่าพิสัยสตูเดนท์ (Studentized Range) จากตาราง 12 นั่นเอง

กรณีที่มีค่าสังเกตเท่ากันมาก ๆ ความแปรปรวนของความแผกกัน $\hat{\varphi}$ จะกำหนดไว้ดังนี้

$$V(\hat{\varphi}) = f \ \frac{n(n+1)}{12} (\Sigma a_j^2/n_j)$$

ในเมื่อ f เป็นแฟคเตอร์ที่ปรับปรุงในตัวสถิติครัสคัล-วอลลิส นั่นเอง

ในการเปรียบเทียบพหุคูณ จะได้กล่าวในรายละเอียดต่อไปนี้

3.4.3 การเปรียบเทียบพหุคูณที่อาศัยผลรวมอันดับที่ของครัสคัล-วอลลิส

(1) เปรียบเทียบทุกกรรมวิธี (All Treatment Comparisons) ในกรณีที่ตัวอย่างไม่เท่ากันนั้น ยังไม่มีตารางของค่าวิกฤต จึงจำเป็นต้องกำหนดให้ขนาดเท่ากัน

เมื่อกำหนดอัตราความคลาดเคลื่อนตามการทดลอง (Experimentwise Error Rate) ไว้เป็น α แล้วเกณฑ์ตัดสินใจเกี่ยวกับความแตกต่างของ M_u และ M_v, u<v จะเป็นดังนี้

$$M_u \neq M_v \text{ ถ้า } |R_u - R_v| \geq y(\alpha,k,n)$$

ในเมื่อ R_u และ R_v เป็นผลรวมของอันดับที่ในกรรมวิธีทดลองหรือกลุ่มตัวอย่างที่ u และ v ตามลำดับ และ $y(\alpha,k,n)$ เป็นค่าวิกฤตที่สอดคล้องสมการต่อไปนี้ ซึ่งพิจารณาได้จากตาราง 19

$$P\{|R_u - R_v| < y(\alpha,k,n) \ ; \ u, v = 1, 2, .., k \ ; \ u<v\} = 1 - \alpha$$

เมื่อขนาดตัวอย่างไม่เท่ากัน ก็ใช้วิธีการประมาณอย่างต่ำ (Conservative Procedure) นั่นคือเมื่ออัตราความคลาดเคลื่อนอย่างสูงเป็น α แล้วเกณฑ์ตัดสินใจจะเป็น

$$M_u \neq M_v \text{ ถ้า } |\bar{R}_u - \bar{R}_v| \geq \sqrt{h(\alpha,k;n_1,n_2,..,n_k)\, V(\bar{R}_u - \bar{R}_v)}$$

ในเมื่อ $\bar{R}_u$ และ $\bar{R}_v$ เป็นค่าเฉลี่ยของอันดับที่ในกรรมวิธี u และ v ส่วน $h(\alpha,k;n_1n_2,..,n_k)$ เป็นค่าวิกฤตของตัวสถิติครัสคัล-วอลลิส จากตาราง 17 และ$V(R_u - R_v)$ กำหนดไว้เป็น

$$V(\bar{R}_u - \bar{R}_v) = \frac{n(n+1)}{12} (1/n_u + 1/n_v) \ ; \ n = n_1+n_2+...+n_k$$

เมื่อตัวอย่างขนาดโต ใช้วิธีการประมาณค่าดังนี้

ก. เมื่อขนาดตัวอย่างเท่ากัน มิลเลอร์ (Miller, 1966) ได้เสนอเกณฑ์ตัดสินใจไว้ว่า

$$M_u \neq M_v \text{ ถ้า } |\bar{R}_u - \bar{R}_v| \geq q(\alpha,k,\infty) \sqrt{k(kn + 1)/12}$$

ในเมื่อ $q(\alpha,k,\infty)$ เป็นค่าคงที่จากตาราง 18 และ $n_1 = n_2 = ... = n_k = n$

ข. เมื่อขนาดตัวอย่างเท่ากันหรือไม่เท่ากันก็ได้ ดันน์ (Dunn, 1964) ได้เสนอเกณฑ์ตัดสินใจไว้ดังนี้

$$M_u \neq M_v \text{ ถ้า } |\bar{R}_u - \bar{R}_v| \geq z_{\alpha_0}\sqrt{\frac{n(n+1)}{12}(1/n_u + 1/n_v)}$$

ในเมื่อ $n = \Sigma n_j$ และ $\alpha_0 = \alpha/k(k-1)$

ตัวอย่าง ในการศึกษาวิธีการเก็บสินค้า 5 แบบ เมื่อเก็บตัวอย่างจากวิธีการเก็บแต่ละแบบ แล้ววัดความชื้นได้เป็นดังนี้

วิธีการเก็บ	1	2	3	4	5
	8.3 (10.5)	7.1 (3.5)	8.1 (9)	10.0 (14)	7.1 (3.5)
	7.8 (7)	5.4 (1)	6.4 (2)	7.9 (8)	
	7.6 (6)	7.4 (5)		9.5 (13)	
	8.4 (12)				
	8.3 (10.5)				
R_i	46	9.5	11	35	3.5

โดยอาศัยวิธีการของดันน์ เมื่อ α = .15 หรือ $z_{.0075}$ = 2.43 จะได้การเปรียบเทียบพหุคูณ ดังต่อไปนี้

(u,v)	$\lvert\bar{R}_u - \bar{R}_v\rvert$		$Z_{.0075}\sqrt{\frac{n(n-1)}{12}}(1/n_u + 1/n_v)$
(1,2)	6.03	<	7.42
(1,3)	3.70	<	8.51
(1,4)	2.47	<	7.42
(1,5)	5.70	<	11.14
(2,3)	2.33	<	9.28
(2,4)	8.50	<	8.30
(2,5)	0.33	<	11.74
(3,4)	6.17	<	9.28
(3,5)	2.00	<	12.45
(4,5)	8.17	<	11.74

จะเห็นได้ว่า วิธีการเก็บ 2 กับ 4 เท่านั้นที่แตกต่างกัน

(2) เปรียบเทียบกรรมวิธีทดลองกับกรรมวิธีควบคุม (Treatment Versus Control) ให้กรรมวิธี 1 เป็นกรรมวิธีควบคุม กรณีตัวอย่างขนาดเล็ก จะสมมติว่าขนาดตัวอย่างเท่าๆ กัน นั่นคือ $n_i = n$ สำหรับตัวอย่างขนาดโตไม่จำเป็นต้องให้ขนาดตัวอย่างเท่ากันก็ได้

เมื่อกำหนดอัตราความคลาดเคลื่อนเป็น α เกณฑ์ตัดสินใจทางเดียว และสองทาง จะกำหนดไว้ดังนี้

$M_u > M_1$ ถ้า $(R_u - R_1) \geq y^*(\alpha,k-1,n)$; $u = 2, 3, .., k$

ในเมื่อ $y^*(\alpha,k-1,n)$ เป็นค่าคงที่จากตาราง 20 ซึ่งสอดคล้องกับสมการ

$$P\{(R_u - R_1) < y^*(\alpha,k-1,n)\ ;\ u = 2, 3 ., k\} = 1 - \alpha$$

และ $M_u \neq M_1$ ถ้า $\lvert R_u - R_1\rvert \geq y^{**}(\alpha,k-1,n)$; $u = 2, 3, ..., k$

ในเมื่อ $y^{**}(\alpha,k-1,n)$เป็นค่าคงที่จากตาราง 21 ซึ่งสอดคล้องกับสมการ

$$P\{(R_u - R_1) < y^{**}(\alpha,k-1,n)\ ;\ u = 2, 3, ..., k\} = 1 - \alpha$$

สำหรับตัวอย่างขนาดโต และ $n_1 = b,\ n_2 = n_3 = n_4 = .. = n_k = n$ นั้น มิลเลอร์ได้เสนอวิธีการประมาณค่าของเกณฑ์ตัดสินใจไว้ดังนี้

$$M_u > M_1 \text{ ถ้า } \bar{R}_u - \bar{R}_1 \geq m(\alpha,k-1,n/(b+n))\sqrt{\frac{N(N+1)}{12}(1/b+1/n)}$$

ในเมื่อ $u = 2, 3, .., k\ ;\ N = n_1 + n_2 + .. + n_k$ และ $m(\alpha,k-1,\rho)$ เป็นค่าคงที่จากตาราง 22

$$\text{และ } M_u \neq M_1 \text{ ถ้า } \bar{R}_u - \bar{R}_1 \geq |m|\ (\alpha,k-1,n/(b-n))\sqrt{\frac{N(N+1)}{12}(1/b+1/n)}$$

ในเมื่อ $|\,m\,|\ (\alpha,k-1,\rho)$ เป็นค่าคงที่จากตาราง 23

ดันน์ (1964) ได้เสนอวิธีการประมาณค่าของเกณฑ์ตัดสินใจไว้ และไม่ได้จำกัดการเท่ากันของขนาดตัวอย่างไว้ ดังนี้

$$M_u > M_1 \text{ ถ้า } (\bar{R}_u - \bar{R}_1) \geq Z_{\alpha/(k-1)}\sqrt{\frac{N(N+1)}{12}(1/n_u+1/n_1)}$$

$$\text{และ } M_u \neq M_1 \text{ ถ้า } |\bar{R}_u - \bar{R}_1| \geq Z_{\alpha/2(k-1)}\sqrt{\frac{N(N+1)}{12}(1/n_u+1/n_1)}$$

ในเมื่อ $u = 2, 3, .., k$ และ $N = n_1 + n_2 +...+ n_k$

ตัวอย่าง ในการศึกษาเพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่าง 3 กลุ่ม โดยที่กลุ่ม ก เป็นกลุ่มควบคุม ส่วน ข และ ค เป็นกลุ่มทดลอง ได้ข้อมูลจากการศึกษาดังนี้

กลุ่มตัวอย่าง	กลุ่มควบคุม (ก)	กลุ่ม ข	กลุ่ม ค
	40 (5.5)	38 (2.5)	48 (18)
	35 (1)	40 (5.5)	40 (5.5)
	38 (2.5)	47 (17)	45 (15)
	43 (10.5)	44 (13)	43 (10.5)
	44 (13)	40 (5.5)	46 (16)
	41 (8)	42 (9)	44 (13)
	40.5	52.5	78

จากตาราง 20 เมื่ออัตราความคลาดเคลื่อน $\alpha = 0.049$ จะได้ $y^*(.049,2,6) = 36$ จากข้อมูลข้างบน เมื่อคำนวณผลต่างระหว่างผลรวมของอันดับที่ของกลุ่มควบคุม และกลุ่มอื่นๆ จะได้เป็นดังนี้

$$R_2 - R_1 = 52.5 - 40.5 = 12$$

$$R_3 - R_1 = 78 - 40.5 = 37.5$$

จะเห็นได้ว่า $R_3 - R_1 > 36$ ซึ่งแสดงว่า กลุ่มทดลอง ค แตกต่างจากกลุ่มควบคุม ณ ระดับนัยสำคัญ $\alpha = 0.049$

3.4.4 แบบทดสอบสมมติฐานรองชนิดอันดับของจองเคียร์-เทอร์พสตรา (Jonckheere-Terpstra Test for Ordered Alternatives)

แบบทดสอบนี้ใช้ทดสอบสมมติฐานหลัก Ho เช่นเดียวกับแบบทดสอบครัสคัล-วอลลิส แต่มีสมมติฐานรอง Ha เป็นแบบอันดับ นั่นคือสมมติฐานที่ทดสอบ จะเป็น