บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การทดสอบความแตกต่างกันของค่ากลางของ 2 ประชากรที่เป็นอิสระกันด้วยวิธีของ Wilcoxon – Mann – Whitney (WMW)

บางครั้งเรียกว่า Mann- Whitney U Test หรือ Mann-Whitney-Wilcoxon Test โดย Wilcoxon ได้ศึกษากรณีใช้ผลรวมลำดับที่ (rank sum) เป็นตัวสถิติทดสอบ โดยที่ Mann และ Whitney ได้ชี้ถึงความสัมพันธ์ระหว่างตัวสถิติทดสอบที่เขาตั้งขึ้นกับของ Wilcoxon การทดสอบนี้นับได้ว่าเป็นการทดสอบที่มีประสิทธิภาพมากที่สุด มักนิยมใช้เพื่อเลี่ยงการใช้การทดสอบแบบที ในสถิติที่ใช้พารามิเตอร์ หรือเมื่อข้อมูลมีมาตราวัดต่ำกว่าแบบอันตรภาค

ข้อกำหนดเบื้องต้น

1. ข้อมูลประกอบด้วยตัวอย่างสุ่ม ด้วยค่า $X_1, X_2 \ldots X_{n_1}$ จากประชากรที่ 1 และตัวอย่างสุ่มอีก 1 ชุด ด้วยค่าสังเกต $Y_1, Y_2 \ldots Y_{n_2}$ จากประชากรที่ 2 ซึ่งเป็นอิสระกัน
2. ตัวอย่าง 2 ชุดนี้เป็นอิสระกัน
3. ค่าตัวแปรสุ่มมีค่าต่อเนื่อง (continuous)
4. มาตรวัดอย่างน้อยเป็นแบบเรียงลำดับ (ordinal scale)
5. ฟังก์ชันการแจกแจง ของ 2 ประชากร ต่างกันเฉพาะค่ากลาง (ซึ่งนิยมวัดด้วยมัธยฐาน, M_X, M_Y) นั่นคือประชากรทั้ง 2 ต้องมีการแจกแจงที่เหมือนกัน ต่างกันเฉพาะค่ากลางเท่ากัน

หมายเหตุ: ในทางปฏิบัติไม่จำเป็นต้องทราบว่ามีการแจกแจงแบบใด

สมมติฐาน ถ้าให้ M_X และ M_Y แทนค่ามัธยฐานของประชากรที่ 1 และ 2 ตามลำดับ อาจทำการทดสอบสองหางหรือหางเดียว ได้ดังนี้

H_0: $M_X = M_Y$

H_1: $M_X \neq M_Y$

หรือ H_0: $M_X \geq M_Y$

H_1: $M_X < M_Y$

หรือ H_0 : $M_X \leq M_Y$

H_1 : $M_X > M_Y$

สถิติที่ใช้ทดสอบ ในที่นี้จะเสนอวิธีการของ Wilcoxon (1945) และ Mann, Whitney (1947) ซึ่งต่างก็เสนอวิธีการทดสอบของตนเอง และในที่สุดสามารถหาความสัมพันธ์ของทั้ง 2 วิธีดังต่อไปนี้

2.1.1.1 วิธีการของ Wilcoxon

ได้ใช้แนวคิดคล้ายการทดสอบของ Wilcoxon Signed Rank Test คือใช้ผลรวมของลำดับที่ (sum of ranks or rank sum) ของตัวอย่างชุดหนึ่ง ในข้อมูลรวมทั้งหมด ($n_1 + n_2$ จำนวน) ที่ได้เรียงลำดับจากน้อยไปหามาก โดยคาดว่าถ้า H_0 เป็นจริง ในข้อมูลรวมทั้งหมดนั้นค่าลำดับที่ของตัวอย่างชุดหนึ่งควรจะมีคละกันไปทั้งค่าน้อย ปานกลาง และมาก ซึ่งจะทำให้ได้ผลรวมลำดับที่ ค่าหนึ่งที่ไม่มากเกินไปหรือน้อยเกินไป แต่ถ้า H_1 เป็นจริง ค่าผลรวมของลำดับที่จากตัวอย่างชุดหนึ่งจะมีค่ามาก หรือน้อยเกินไปดังตัวอย่างต่อไปนี้

ถ้ามีตัวอย่างสุ่มขนาด 4 ด้วยค่าตัวแปรสุ่ม X และอีกชุดหนึ่งด้วยขนาด 5 ด้วยตัวแปรสุ่ม Y ปรากฏว่าเมื่อนำทั้ง 9 จำนวนมารวมกัน และเรียงลำดับ

และให้ S = ผลรวมของลำดับที่ของข้อมูล X ในข้อมูลรวมทั้งหมด

$$= \sum_{i=1}^{n_1} \text{Rank}(X_i)$$

ถ้าข้อมูลรวมทั้งหมด เมื่อนำมาเรียงลำดับแล้วได้ลำดับที่ดังนี้

ชุดที่ 1	YYYYYXXXX	กรณีนี้จะได้ค่า $S = 6+7+8+9$	$=30$
หรือ ชุดที่ 2	XXXXYYYYY	กรณีนี้จะได้ค่า $S = 1+2+3+4$	$=10$
หรือ ชุดที่ 3	XYXYXYXYY	กรณีนี้จะได้ค่า $S = 1+3+5+7$	$=16$

จะพบว่าในตัวอย่างรวมชุดที่ 1 ตัวแปร X อยู่ในตอนท้ายได้ค่า $S = 30$ มีค่าใหญ่มากในตัวอย่างชุดนี้ น่าจะคาดเดาว่า ประชากรกลุ่ม X มีแนวโน้มที่จะมีค่ามากกว่ากลุ่ม Y

ในตัวอย่างรวมชุดที่ 2 ตัวแปร X อยู่ในตอนต้น ได้ค่า $S = 10$ มีค่าน้อยดังนั้น น่าจะทำให้ยอมรับ H_1 : ประชากร X มีแนวโน้มที่จะมีค่าน้อยกว่า Y

และในตัวอย่างรวมชุดที่ 3 ตัวแปร X อยู่ในลักษณะผสม (mix) กันอย่างดีกับ Y ทำให้มีลำดับที่ ทั้งค่าน้อย ปานกลาง และมาก ได้ค่า $S = 16$ ซึ่งมีค่าปานกลางจากตัวอย่างนี้ น่าจะทำให้เรายอมรับ H_0 : ประชากร X และ Y มีค่ามัธยฐานไม่ต่างกัน

Wilcoxon ได้สร้างตารางแสดงค่าความน่าจะเป็นของค่า S ที่น้อยหรือมากเกินไปนี้ ซึ่งสามารถใช้ตารางดังกล่าวหาค่า p – value เพื่อตัดสินใจยอมรับ H_0 หรือปฏิเสธ H_0 ได้ แต่เนื่องจากค่า S ที่เล็กที่สุด จะแตกต่างกันไปตามขนาดตัวอย่างที่สุ่มมา จึงทำให้การสร้างตารางยากขึ้น และค่อนข้างใหญ่

ทำให้ไม่สะดวกในการใช้ ในที่นี้ไม่เสนอวิธีการของ Wilcoxon โดยตรงนี้ แต่จะปรับสูตรสถิติที่ใช้ทดสอบให้สัมพันธ์กับค่า S นี้ และสอดคล้องกับวิธีของ Mann, Whitney ซึ่งจะได้เสนอในลำดับต่อไป

ในกรณีตัวอย่างขนาดใหญ่สามารถประมาณการแจกแจงของ S ด้วยการแจกแจงปกติ ดังสูตร

$$Z = \frac{(S \pm 0.5) - n_1(n_1 + n_2 + 1)/2}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

2.1.1.2 วิธีการของ Mann, Whitney

มักเรียกชื่อการทดสอบของเขาทั้งสองว่า Mann – Whitney U test ซึ่งกำหนดให้ตัวสถิติ U คือ การนับจำนวนค่าสังเกตในตัวอย่างชุดหนึ่งที่นำหน้า (exceeding) แต่ละค่าสังเกตในตัวอย่างอีกชุดหนึ่งในข้อมูลที่นำมารวมกันและเรียงลำดับ การคำนวณหาค่า U สามารถทำได้ง่ายไม่จำเป็นต้องใช้คอมพิวเตอร์ และวิธีการนี้ยังเป็นพื้นฐานในการหาช่วงความเชื่อมั่นของผลต่างค่ามัธยฐานใน 2 ประชากรด้วย

ให้ $U = \sum_{i=1}^{n_1} U_i$

= ผลรวม (จำนวนค่า Y ที่น้อยกว่า หรือนำหน้า X_i ในข้อมูลรวมทั้งหมดที่เรียงลำดับแล้ว)

เช่น มีข้อมูลรวม	YYYYYXXXX จะได้	U	=	5 + 5 + 5 + 5	= 20
หรือ	XXXXYYYYY จะได้	U	=	0	
หรือ	XYXYXYXYY จะได้	U	=	0 + 1 + 2 + 3	= 6

จะเห็นว่าค่า U ที่ใหญ่เกินไปหรือน้อยเกินไป ทำให้น่าจะเชื่อว่า H_1 เป็นจริง ในขณะที่ U ที่มีค่าปานกลางจะทำให้เชื่อว่า H_0 เป็นจริง ซึ่งจะสอดคล้องกับค่า S ของ Wilcoxon

นอกจากการนับจำนวนเพื่อหาค่า U แล้ว อาจใช้สูตรหาค่า U ดังนี้

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - S_2$$

เมื่อ S_2 = ผลรวมลำดับที่ของตัวแปร Y จากตัวอย่างขนาด n_2

ลองพิจารณาตัวอย่างต่อไปนี้

ถ้ามีตัวอย่างสุ่มชุดที่ 1 และ 2 ด้วยค่าสังเกต ดังนี้

ข้อมูล X :	110	70	53	51	
ข้อมูล Y :	78	64	75	45	82

รวมทั้งหมดเข้าด้วยกัน และเรียงลำดับจะได้

45	51	53	64	70	75	78	82	110
Y	X	X	Y	X	Y	Y	Y	X

หาค่า $U = 1+1+2+5 = 9$

ถ้าใช้สูตร $U = 4\times5+\frac{5(5+1)}{2}-(1+4+6+7+8)$

$= 9$

Mann-Whitney ได้สร้างตารางค่าความน่าจะเป็นเมื่อ U มีค่าต่าง ๆ ที่ค่า n_1, n_2 ต่าง ๆ กัน แต่การใช้ตารางจำเป็นต้องเลือกใช้ค่า U ที่มีค่าน้อยที่สุด เพราะค่าความน่าจะเป็นที่คำนวณในตารางเป็นความน่าจะเป็นด้านซ้ายของโค้งการแจกแจง

การเลือกใช้ค่า U ที่น้อยที่สุด ให้ใช้ความสัมพันธ์

$U' = n_1n_2 - U$ แล้วเลือกค่า U' หรือ U ที่เล็กที่สุด

เช่นตัวอย่างข้างต้น $U' = 4x5-9 = 11$

ดังนั้นเลือกใช้ U = 9 เป็นสถิติที่ใช้ทดสอบ แต่การสร้างตารางการแจกแจงของค่า U จะง่ายขึ้น เนื่องจากค่าต่ำสุดของ U = 0 เสมอ

ในกรณีตัวอย่างใหญ่สามารถประมาณการแจกแจงค่า U ด้วยการแจกแจงปกติมาตรฐานได้ค่า

$$Z = \frac{U - n_1n_2/2}{\sqrt{n_1n_2(n_1+n_2+1)/12}}$$

2.1.1.3 วิธีการของ Wilcoxon และ Mann-Whitney

Mann-Whitney ได้แสดงความสัมพันธ์ระหว่างสถิติที่ใช้ทดสอบของเขากับของ Wilcoxon พบว่า ถ้าให้ $T = S-\frac{n_1(n_1+1)}{2}$ แล้วค่า T ที่ได้จะมีค่าเท่ากับค่า U นั่นเอง (สามารถทดลองจากตัวอย่างข้างต้น จะพบว่าค่า T จะเท่ากับ U ในทุกกรณี) หลักในการหาอาณาเขตวิกฤตยังคงคล้ายการพิจารณาค่า S เนื่องจากค่า T มีความสัมพันธ์ในเชิงเส้นตรงกับค่า S ดังนั้นค่า T ที่มากเกินไปหรือน้อยเกินไปจะทำให้ปฏิเสธ H_0 เพื่อยอมรับ H_1 แต่การสร้างตารางการแจกแจงความน่าจะเป็นของค่า T จะง่ายขึ้น เนื่องจากค่าเล็กที่สุดของ T = 0 เสมอ

ดังนั้น สถิติที่ใช้ทดสอบคือ $T = S-\frac{n_1(n_1+1)}{2}$

เมื่อ S = ผลรวมลำดับที่ของตัวอย่างขนาด n_1 ในข้อมูลรวมทั้งหมดที่เรียงลำดับแล้ว

และ $T' = n_1n_2 - T$

การตัดสินใจ ใช้ตารางแสดงค่าวิกฤตของสถิติที่ใช้ทดสอบ T

ในกรณีการทดสอบสองหางจะปฏิเสธ H_0 ถ้าพบว่าค่า T น้อยเกินไป หรือใหญ่เกินไป

อาณาเขตวิกฤต คือ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$

เมื่อ $W_{1-\frac{\alpha}{2}} = n_1 n_2 - W_{\frac{\alpha}{2}}$

เมื่อเป็นการทดสอบหางเดียว ด้านน้อยกว่า คือ $H_1 : M_X < M_Y$ จะปฏิเสธ H_0 เมื่อ พบว่าค่า T น้อยเกินไป อาณาเขตวิกฤต คือ $T < W_{\alpha}$

เมื่อเป็นการทดสอบหางเดียว ด้านมากกว่า $H_1 : M_X > M_Y$ จะปฏิเสธ H_0 เมื่อพบ ว่า T ใหญ่เกินไป อาณาเขตวิกฤต คือ $T > W_{1-\alpha}$

เมื่อ $W_{1-\alpha} = n_1 n_2 - W_{\alpha}$

2.1.1.4 การทดสอบของ Wilcoxon- Mann- Whitney แบบมีซ้ำมาก (The Wilcoxon – Mann – Whitney Test with many ties)

ถ้าค่าข้อมูลที่เป็นลำดับที่ในกลุ่มใดๆ 2 กลุ่มนั้นมีค่า ซ้ำกันมาก อาจจัดข้อมูลใหม่เป็น ตารางแจกแจง 2 ทาง โดยแถวนอนมี 2 แถว หมายถึง 2 กลุ่มตัวอย่างที่อิสระกัน และแถวตั้งจัดเป็นค่าลำดับที่ และนับความถี่ลงเซลต่างๆ ซึ่งตารางนี้จะมีลักษณะแบบเรียงลำดับทางแถวตั้ง อาจเรียกชื่อว่าเป็น Singly ordered contingency table เช่นมีข้อมูลดังต่อไปนี้

สุ่มตัวอย่างคนไข้ 4 คน ให้ใช้ยา A และคนไข้ 3 คนใช้ยา B หลังจากเวลาผ่านไประยะหนึ่ง ประเมินผลการรักษา โดยให้ แพทย์ประเมินในระดับความพอใจ จาก 1 – 3 (1 น้อยที่สุด) ได้ข้อมูลดังนี้

ยา A	1 1 2 2	$n_A = 4$ จำนวน
ยา B	2 3 3	$n_B = 3$ จำนวน

จะพบว่าค่าข้อมูลซึ่งมีค่า 1 – 3 มีค่าซ้ำกันมาก การให้ลำดับที่ต้องใช้ค่าเฉลี่ย ของลำดับที่ โดย

ค่า 1 มีสองค่า ได้ลำดับที่เฉลี่ยของลำดับ 1 และ 2 คือ 1.5

ค่า 2 มีสามค่า ได้ลำดับที่เฉลี่ยของลำดับ 3, 4 และ 5 คือ 4

ค่า 3 มีสองค่า ได้ลำดับที่เฉลี่ยของลำดับ 6 และ 7 คือ 6.5

จะวิเคราะห์ข้อมูลแบบตารางที่เรียกว่า Singly ordered contingency table ด้วยสถิติ Wilcoxon – Mann – Whitney แบบปรับ ties ดังนี้

จัดข้อมูลลงตารางการณ์จร 2 ทาง และบันทึกความถี่ จะได้ตารางการณ์จรดังนี้

	ผลการรักษาตามลำดับความพอใจ			ผลรวม
	1	2	3	
ยา A	2	2	0	4 $= n_A$
ยา B	0	1	2	3 $= n_B$
				$n = 7$

ถ้าพิจารณาตารางใหม่นี้ จะพบว่าสามารถใช้สถิติ χ^2 วิเคราะห์ได้ เพราะเป็นตารางการณ์จร แต่เนื่องจากขนาดตัวอย่างเล็กมาก รวมทั้งความถี่คาดหวังจะมีค่าน้อยกว่า 5 มากเกิน 20 % ของจำนวนเซลทั้งทั้งหมด การใช้ χ^2 จึงไม่เหมาะสม (ได้ค่า p – value = 0.31 ซึ่งก็ไม่พบนัยสำคัญระหว่างยาทั้งสอง)

เราสามารถใช้สถิติทดสอบของ Wilcoxon – Mann – Whitney แบบมี ties มากๆ ดังนี้

จากสูตร $$T = S - \frac{n_1(n_1+1)}{2}$$

เมื่อ S = ผลรวมลำดับที่ของยา A = 1.5+1.5+4+4 = 11

ดังนั้น $$T = 11 - \frac{4(5)}{2} = 1$$ (หรือ T = 11 เมื่อคิดจากกลุ่ม B)

ซึ่งได้ค่า p – value = 0.0857 พบว่ามีนัยสำคัญที่ระดับ $\alpha = 0.10$ ซึ่งดีกว่าใช้ χ^2

เมื่อ H_0 : ยา A และ ยา B ให้การรักษา ไม่ต่างกัน

H_1 : ยา A และ ยา B ให้การรักษา ต่างกัน

หรือ H_0 : ไม่มีความสัมพันธ์กันระหว่างยาที่ใช้กับผลการรักษา

H_1 : มีความสัมพันธ์กันระหว่างยาที่ใช้กับผลการรักษา

ตารางที่ปรับเป็นตารางการณ์จรข้างต้น จะมีเพียงทางเดียวที่เรียงลำดับ (อาจจะเป็นแนวนอน หรือ แถวตั้งก็ได้) จึงเรียกว่า Singly ordered contingency table และอาจจะขยายเป็น 3 แถวนอน (หรือแถวตั้ง) ในแถวที่ไม่ได้เรียงลำดับ เช่นเป็น กรณี 3 กลุ่มตัวอย่างหรือ 4 กลุ่มตัวอย่าง และสามารถใช้สถิติทดสอบที่ใช้กับมากกว่า 2 ประชากร ได้เช่นกัน โดยใช้หลักการคำนวณ เช่นเดียวกันนี้

2.1.1.5 การทดสอบไคสแควร์สำหรับตัวอย่าง 2 กลุ่ม ที่เป็นอิสระกัน หรือการทดสอบไคสแควร์เกี่ยวกับความเป็นเอกภาพของสัดส่วน (The Chi-Square Test for 2 Independent Samples or The Chi-Square Test for homogeneity of Proportions)

การทดสอบไคสแควร์ สามารถใช้กับข้อมูลที่มีมาตรวัดแบบนามบัญญัติ เมื่อข้อมูลมาจากประชากร 2 กลุ่ม ที่เป็นอิสระกัน และภายในแต่ละกลุ่มมีลักษณะย่อย เก็บข้อมูลเป็นความถี่จากตัวอย่าง 2 กลุ่มนั้นในแต่ละลักษณะย่อย เช่น สอบถามความคิดเห็นเรื่องกฎหมายการทำแท้งของนักการเมือง 2 พรรค ว่ามีความเห็นด้วยหรือไม่เห็นด้วย หรือไม่มีความเห็น การทดสอบด้วยไคสแควร์นี้จะช่วยบอกให้ทราบว่า มีความแตกต่างกันในสัดส่วนของความคิดเห็นต่อเรื่องนี้หรือไม่ ระหว่างนักการเมือง 2 กลุ่มนี้ หรือ ถ้าสัดส่วนมีค่าเท่ากันใน 2 กลุ่มก็คือ พรรคการเมืองไม่มีความสัมพันธ์กับความคิดเห็น คือเป็นอิสระกันระหว่างพรรคการเมืองที่สังกัดกับความคิดเห็น

วิธีการทดสอบ จะใช้หลักการเปรียบเทียบความถี่ที่ได้จากการสังเกตกับความถี่ที่คาดหวังไว้ของแต่ละลักษณะย่อยใน 2 กลุ่มตัวอย่างนั้น

ข้อมูล ประกอบด้วย ความถี่ที่สังเกตได้จากข้อมูลตัวอย่าง 2 ชุด ขนาด n_1และ n_2 ตามลำดับ จำนวนทั้งหมด = $N = n_1 + n_2$ และจัดความถี่นี้ลงในตารางการจรณ์ (contingency table)

ถ้าในแต่ละประชากรนั้น มีลักษณะย่อยที่น่าสนใจ k ประเภท และให้ค่า O_{ij} แทนความถี่หรือจำนวนค่าสังเกตจากลักษณะย่อยที่ j ของตัวอย่างชุดที่ i (i =1, 2) จะได้ตารางแสดงความถี่ที่สังเกตได้ดังนี้

ลักษณะย่อย	1	2	3	......j	.......k	
ตัวอย่าง 1	O_{11}	O_{12}	O_{13}	......O_{1j}	....O_{1k}	n_1
ตัวอย่าง 2	O_{21}	O_{22}	O_{23}	......O_{2j}	.....O_{2k}	n_2
	C_1	C_2	C_3	......C_j	....C_k	N

และในกรณีที่มีลักษณะย่อยเพียง 2 แบบ ตารางที่ได้จะเป็น **ตารางชนิด 2 x 2**

สมมติฐาน จะทดสอบว่า มีความแตกต่างของสัดส่วนในลักษณะย่อยต่างๆ ระหว่างประชากร 2 กลุ่มนี้หรือไม่ หรือมีความเป็นเอกภาพของสัดส่วนในแต่ละลักษณะย่อยของสองประชากรนี้หรือไม่

นั่นคือ $H_0 : P_{11} = P_{21}, P_{12} = P_{22}, P_{13} = P_{23}, \ldots, P_{1k} = P_{2k}$

หรือ $H_0 : P_{1j} = P_{2j}$ (j =1,2,...., k)

หรือ H_0 : **ไม่มีความสัมพันธ์กันระหว่างตัวแปรทางแถวนอนและแถวตั้ง**

สถิติที่ใช้ทดสอบ หลักการของการทดสอบนี้ คือเปรียบเทียบความถี่จากการสังเกตได้กับความถี่คาดหวังตามทฤษฎี ซึ่งความน่าจะเป็นที่จะเกิดความถี่คาดหวังนั้นเป็นไปตาม H_0
สามารถใช้การทดสอบไคสแควร์ได้ ดังนี้

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{ด้วย df.} = (2-1)(k-1)$$

เมื่อ O_{ij} = ความถี่ที่สังเกตได้จากแถวนอนที่ i และแถวตั้งที่ j ของตารางการณ์จร

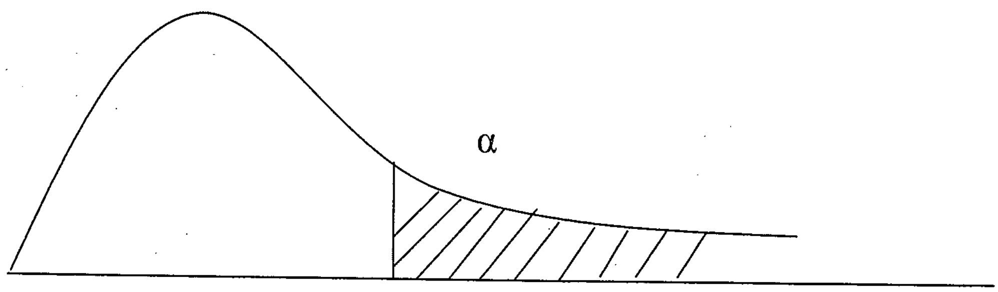
E_{ij} = ความถี่คาดหวังภายใต้ H_0 จากแถวนอนที่ i และแถวตั้งที่ j ของตารางการณ์จร

$$\sum_i \sum_j O_{ij} = \sum_i \sum_j E_{ij} = N$$

df. = (2 – 1)(k – 1) เมื่อจำนวนแถวนอนในตารางการณ์จรมี 2 แถวนอน และ k แถวตั้ง
สำหรับการหาค่า E_{ij} ใช้สูตรดังนี้

$$E_{ij} = \frac{(\text{ผลรวมแถวนอนที่ i})\ (\text{ผลรวมของแถวตั้งที่ j})}{\text{จำนวนความถี่ทั้งหมด}}$$

การตัดสินใจ ใช้ค่า χ^2 ที่คำนวณได้จากข้อมูลตัวอย่างเปรียบเทียบกับค่า χ^2 จากการแจกแจงแบบไคสแควร์ที่ df. =(k – 1)



$\chi^2_{\alpha,(k-1)}$

ถ้าค่า χ^2 ที่คำนวณได้มีค่ามากกว่า $\chi^2_{\alpha,(k-1)}$ จะปฏิเสธ H_0 ที่ระดับนัยสำคัญ α

หมายเหตุ

1. สำหรับตารางการจรณ์ ชนิด 2 x 2 ถ้าให้ความถี่ที่สังเกตได้จากตาราง 2 x 2 คือ

A	B
C	**D**

ผลรวม = N

หาค่า χ^2 ตามสูตรข้างต้น จะพิสูจน์ได้ว่า

$$\chi^2 = \frac{N\left[\,|AD - BC\,| - \frac{N}{2}\right]^2}{(A+B)(C+D)(A+C)(B+D)} \text{ ด้วย d.f.} = 1$$

2. การทดสอบนี้ จะคล้ายการทดสอบความเป็นอิสระ (χ^2–test for Independence) จะพบว่าตัวสถิติทดสอบ และการตัดสินใจจะเหมือนกันทุกประการ แต่มีที่ต่างกัน 2 กรณี คือ
   ก. การสุ่มตัวอย่างขนาด n_1และ n_2 จาก 2 ประชากรที่เป็นอิสระกัน ดังนั้นจำนวน n_1 และ n_2 นั้นจะมีค่าคงที่ และทราบล่วงหน้า ในขณะที่การทดสอบความเป็นอิสระเราเพียงสุ่มตัวอย่างขนาด n จากประชากรหนึ่ง แล้วจึงมาจัดความถี่ลงตารางการณ์จร ผลรวมทางแนวนอนหรือแนวตั้ง จึงไม่ใช่ค่าคงที่และไม่ทราบล่วงหน้า
   ข. การคำนวณค่าความถี่คาดหวัง ใช้หลักที่ว่า $E_{ij} = n_i C_j /N$ ซึ่งต่างไปจากการทดสอบความเป็นอิสระที่ใช้ P(AB) = P(A).P(B) แต่เมื่อกระจายสูตรแล้วได้ผลลัพธ์เหมือนกัน

ตัวอย่าง การทดลองหนึ่ง สุ่มตัวอย่างนักเรียนชาย 12 คน และหญิง 12 คน ให้มาปฏิบัติงานชิ้นหนึ่ง เพื่อดูสมรรถภาพการทำงานภายในระยะเวลาที่จำกัด โดยจำแนกดูว่า แต่ละคนทำงานเสร็จหรือไม่เสร็จ ได้ข้อมูลดังนี้

	งานเสร็จ	ไม่เสร็จ	
นักเรียนชาย	10	2	
นักเรียนหญิง	1	11	
	11	13	24

เขียนสมมติฐานได้ดังนี้

H_0: สัดส่วนของนักเรียนที่ทำงานเสร็จ (หรือไม่เสร็จ) เท่ากัน จากนักเรียน ชาย และหญิง

H_1: สัดส่วนของนักเรียนที่ทำงานเสร็จ (หรือไม่เสร็จ) ไม่เท่ากัน จากนักเรียน ชายและหญิง

หรือ H_0: นักเรียน ชายและหญิงกับการทำงานสำเร็จหรือไม่สำเร็จ เป็นอิสระต่อกัน

การคำนวณความถี่คาดหวัง

ค่า H_0 เป็นจริงนั่นคือสัดส่วนของนักเรียนที่ทำงานเสร็จ จะเท่ากันใน 2 กลุ่ม

โดยมีค่าสัดส่วนหรือความน่าจะเป็น = 11/24

ดังนั้น ความถี่คาดหวังของ นักเรียนชาย ที่ทำงานเสร็จ

= จำนวนนักเรียนชาย × prob. (ทำงานเสร็จ)

$$= 12 \times \frac{11}{24} = 5.5$$

เมื่อหาความถี่ที่คาดหวังได้แล้ว เราสามารถวิเคราะห์ข้อมูล

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(10-5.5)^2}{5.5} + \frac{(2-6.5)^2}{6.5} + \frac{(1-5.5)^2}{5.5} + \frac{(11-6.5)^2}{6.5}$$

$$= \frac{(4.5)^2}{5.5} + \frac{(-4.5)^2}{6.5} + \frac{(-4.5)^2}{5.5} + \frac{(4.5)^2}{6.5}$$

$$= 13.594$$

ที่ระดับนัยสำคัญ 0.05 ได้ค่าวิกฤต คือ $\chi^2_{0.05,\,1} = 3.84$ ค่า χ^2 จากตัวอย่าง มีค่ามากกว่าค่าวิกฤต จึงปฏิเสธ H_0

สรุป นักเรียนชายและหญิง มีสัดส่วนในทำงานเสร็จ (หรือไม่เสร็จ) แตกต่างกันที่ระดับนัยสำคัญ 0.05 หรือ เพศของนักเรียนกับความสำเร็จของงานที่ทำมีความสัมพันธ์กัน

จากตัวอย่างนี้เป็นตารางชนิด 2 x 2 สามารถใช้สูตร χ^2 สำหรับตาราง 2 x 2 ได้ดังนี้

จาก $$\chi^2 = \frac{N\left[|AD-BC| - \frac{N}{2}\right]^2}{(A+B)(C+D)(A+C)(B+D)}$$

$$= \frac{24\left[|10\times 11-2\times 1|-\frac{24}{2}\right]^2}{12\times 12\times 11\times 13}$$

$$= \frac{16\times 96}{11\times 13}$$

$$= 10.741$$

เช่นเดียวกัน ที่ระดับนัยสำคัญ 0.05 คือปฏิเสธ H_0 เหมือนตอนต้น

หมายเหตุ จากการใช้สูตร χ^2 ได้ค่าต่างกันจาก 2 สูตร เนื่องจากในสูตรแรก ไม่มีการปรับค่าต่อเนื่อง (Yate's correlation for continuity) ซึ่งควรจะใช้เมื่อ χ^2 มีd.f. = 1 ในขณะที่สูตรที่ 2 ใช้ค่าปรับแล้ว

2.2 งานวิจัยที่ที่เกี่ยวข้อง

Gray Simon (10)ได้เสนอทางเลือกให้นักวิจัยเพื่อวิเคราะห์ข้อมูลจากตารางการณ์จรแบบเรียงลำดับได้หนึ่งทาง ด้วยสถิติแบบ Log Linear Model หรือ แบบ Accumulated Logits within Rows ซึ่งเป็นสถิติชั้นสูงที่ต้องใช้ความรู้ความเข้าใจในทฤษฎีสถิติเกี่ยวกับ การวิเคราะห์ข้อมูลเชิงกลุ่ม (Categorical data analysis) แต่ยังไม่เป็นที่รู้จักแพร่หลายในหมู่นักวิจัยในสาขาต่างๆ

Emerson and Moses(11) ได้นำเสนอว่าข้อมูลทางด้านชีววิทยา และการแพทย์ มักใช้ข้อมูลจากตารางการณ์จรแบบเรียงลำดับได้ และแนะนำว่าสถิติแบบ WMW แบบมี ties มาก เหมาะสมที่จะใช้วิเคราะห์ข้อมูลประเภทนี้ โดยใช้การประมาณด้วยการแจกแจงปกติจากโปรแกรมสำเร็จรูปที่คำนึงถึงการแจกแจงที่แท้จริงของสถิติแบบ WMW (exact WMW distribution)

Klotz(12) ได้กล่าวถึงข้อมูลจากตารางการณ์จรแบบเรียงลำดับได้ ว่าควรใช้สถิติแบบ WMW แบบมี ties มาก แต่การแจกแจงของสถิตินี้ภายใต้สมมติฐานเบื้องต้น(null distribution)จะขึ้นอยู่กับรูปแบบของค่าซ้ำ และยากที่จะคำนวณ จากการพยายามหาวิธีการประมาณโดยใช้โมเมนต์ที่สี่ พบว่าสามารถใช้การแจกแจงแบบปกติในกรณีตัวอย่างมีขนาดใหญ่เท่านั้น

ในขณะที่ Moses, Emerson and Hosseini (13)แนะนำว่าข้อมูลประเภทนี้ ไม่ควรวิเคราะห์ด้วยการทดสอบไคสแควร์ เนื่องจากจะสูญเสียสาระข้อมูล(ไม่ได้ใช้ประโยชน์จากการทราบลำดับที่ เพราะการทดสอบไคสแควร์ใช้ข้อมูลแบบกลุ่มเท่านั้น) และไม่ไว (insensitive)ต่อความแตกต่างในเทอมลำดับที่ แต่เสนอให้ใช้การทดสอบแบบที โดยเปลี่ยนค่าลำดับที่ให้เป็นคะแนน

จากการศึกษาของนักวิจัยดังกล่าวข้างต้น จะพบว่า ส่วนใหญ่แนะนำให้ใช้สถิติแบบ WMW แบบมี ties มาก แต่มีข้อจำกัดเรื่องการแจกแจงภายใต้สมมติฐานเบื้องต้น และคำนวณได้ยาก

และในตำราสถิติเมื่อกล่าวถึงสถิติแบบ WMW แบบมี ties มาก มักจะเปรียบเทียบผลการวิเคราะห์กับการทดสอบแบบไคสแควร์ แต่ยังไม่มีงานวิจัยใดที่ให้ผลสรุปถึงการเปรียบเทียบนี้ จึงเป็นที่มาของงานวิจัยนี้