บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนของวรรณคดีที่เกี่ยวข้องนี้ นำเสนอเป็น 4 ตอน ดังนี้คือ

ตอนที่ 1 มาตรการวัด

ตอนที่ 2 สหสัมพันธ์ที่ศึกษาในงานวิจัยนี้

ตอนที่ 3 การแจกแจงที่เกี่ยวข้องกับงานวิจัยนี้

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มาตราการวัด (Measurement Scales)

ในการใช้สถิติเพื่อวิเคราะห์ข้อมูลในการวิจัยนั้น ผู้วิจัยจำเป็นต้องทราบลักษณะของข้อมูลที่ใช้ในการวิจัยเสียก่อน เพื่อเป็นเครื่องวินิจฉัยว่าจะใช้วิธีการทางสถิติแบบใดในการวิเคราะห์จึงจะถูกต้องเหมาะสม ดังนั้นผู้วิจัยจึงควรทราบถึงมาตราการวัดของข้อมูลก่อน โดย Stevens (1946) ได้จำแนกมาตราการวัดออกเป็น 4 ระดับ ดังนี้

1. มาตรานามบัญญัติ (Nominal Scales) เป็นมาตราการวัดในระดับต่ำสุด เป็นการวัดแบบง่ายๆ โดยการจำแนกหรือแยกประเภท (classification) ตามคุณลักษณะที่ไม่เหมือนกัน อาจเป็นเพียงการเรียกชื่อ ดังนั้นการวัดในระดับนี้บางทีจึงไม่เป็นที่ยอมรับว่าเป็น " การวัด " เพราะไม่สามารถบอกปริมาณมากน้อยได้ เป็นแต่เพียงแสดงให้เห็นความแตกต่างของสิ่งต่างๆ เท่านั้น ตัวอย่างการวัดในระดับนี้เช่น การจำแนกคนตามเพศ เป็นเพศชาย - หญิง หรือจำแนกคนตามศาสนาที่นับถือ เป็นคนที่นับถือศาสนาพุทธ คริสต์ อิสลาม เป็นต้น

การกำหนดตัวเลขให้กับสิ่งต่าง ๆ ในมาตรานามบัญญัติจะเป็นเพียงตัวแทนประเภทของสิ่งที่ถูกวัด หรือเพื่อใช้ในการสื่อความหมาย ให้จำได้ง่ายหรือสะดวกในการนับ โดยที่ตัวเลขดังกล่าวไม่มีความหมายในเชิงปริมาณแต่ประการใด ดังนั้นจึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หรือหารกันได้ นอกจากกระทำได้แต่เพียงการนับจำนวนหรือ

หาความถี่ของลักษณะที่มีเหมือนกันเท่านั้น เช่น กำหนดตัวเลข 1 แทนเพศชาย และเลข 2 แทนเพศหญิง เป็นต้น

2. มาตราจัดอันดับ (Ordinal Scales) การวัดตามมาตรานี้จะมีระดับการวัดสูงกว่ามาตรานามบัญญัติ ลักษณะการวัดเป็นการจัดอันดับ (rank order) ให้กับคุณสมบัติบางอย่างว่ามากกว่าหรือน้อยกว่าข้อมูลอื่นๆ ซึ่งสามารถจัดอันดับข้อมูลตามตำแหน่งได้จากมากที่สุดไปหาน้อยที่สุด หรือจากน้อยที่สุดไปหามากที่สุดได้ เช่น ผลการเรียนของนักเรียนตามลำดับ ดีมาก ดี ปานกลาง พอใช้ ต้องปรับปรุง หรือ นักเรียนที่สอบได้ที่ 1,2,3,... เป็นต้น ผลการวัดในมาตรานี้ช่วงห่างระหว่างอันดับจะไม่เท่ากัน และไม่ทราบว่าห่างกันเป็นปริมาณเท่าใด เช่น จะบอกว่านักเรียนที่สอบได้ที่ 1 เก่งกว่านักเรียนที่สอบได้ที่ 2 เท่ากับ นักเรียนที่สอบได้ที่ 3 เก่งกว่านักเรียนที่สอบได้ที่ 4 ไม่ได้ เป็นต้น ดังนั้นข้อมูลที่ปรากฏออกมาเป็นตัวเลขในระดับนี้ไม่สามารถนำมาบวก ลบ คูณ หรือหารกันได้ จะบอกได้แต่เพียงว่าต่างกันไปในทางไหน หรือทำให้ทราบทิศทางเท่านั้น

3. มาตราอันตรภาค (Interval Scales) มาตราการวัดในมาตรานี้มีลักษณะเหมือนมาตราจัดอันดับทุกอย่าง แต่มีคุณสมบัติดีกว่าตรงที่แต่ละหน่วยของการวัดมีระยะห่างเท่าๆ กัน จึงสามารถเปรียบเทียบกันได้ว่ามากน้อยกว่ากันเท่าใด แต่ไม่สามารถบอกได้ว่าเป็นกี่เท่าของกันและกัน เพราะมาตรานี้ไม่มีศูนย์แท้หรือศูนย์สมบูรณ์ (true zero or absolute zero) เช่น ผู้ที่สอบได้ศูนย์คะแนนไม่ได้หมายความว่าไม่มีความรู้ในวิชานั้น เป็นแต่เพียงผู้สอบคนนั้นทำข้อสอบฉบับนั้นไม่ได้เลย เป็นต้น ด้วยเหตุนี้ข้อมูลที่วัดได้ในมาตรานี้จึงไม่สามารถนำมาคูณ หรือหารกันได้ จะใช้ได้เฉพาะบวกหรือลบกันเท่านั้น

4. มาตราอัตราส่วน (Ratio Scales) เป็นมาตราการวัดในระดับสูงที่สุด กล่าวคือครอบคลุมคุณสมบัติทุกอย่างของมาตราอันตรภาค แต่ดีกว่ามาตราอันตรภาคตรงที่มีศูนย์แท้ ข้อมูลในระดับนี้สามารถบวก ลบ คูณ หรือหารกันได้ มาตราการวัดที่จัดอยู่ในระดับนี้มักจะเป็นการวัดทางฟิสิกส์ เช่น ความสูง ความยาว น้ำหนัก เวลา เป็นต้น

ตอนที่ 2 สหสัมพันธ์ที่ศึกษาในงานวิจัยนี้

สหสัมพันธ์ของสเปียร์แมน (Spearman's Rank Correlation)

โดยในปี ค.ศ. 1904 Carl Spearman ได้พัฒนาวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ขึ้นมาวิธีหนึ่ง ซึ่งดัดแปลงมาจากวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เพื่อใช้หาสหสัมพันธ์ของข้อมูลที่อยู่ในมาตราจัดอันดับ เพราะการวัดตัวแปรบางประเภทไม่สามารถวัดค่าออกมาในรูปของมาตราอันตรภาคหรือมาตราอัตราส่วนได้โดยตรง วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนนี้เป็นวิธีหาความสัมพันธ์ระหว่างตัวแปรสองตัวที่นิยมใช้กันอย่างแพร่หลาย เพราะเป็นการหาความสัมพันธ์อย่างง่ายๆ สามารถคำนวณได้อย่างรวดเร็ว เหมาะสำหรับข้อมูลที่มีจำนวนไม่มาก ขนาดของความสัมพันธ์ที่วัดได้และผลของการทดสอบนัยสำคัญได้ผลใกล้เคียงกับวิธีของเพียร์สัน แต่ค่าที่ได้อาจแตกต่างกันไปบ้าง สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนคือ r

เมื่อข้อมูลอยู่ในมาตราจัดอันดับ สูตรที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน สามารถกำหนดใช้ได้ตามเงื่อนไขใน 3 กรณี ต่อไปนี้

1. กรณีข้อมูลไม่มีอันดับซ้ำกัน

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เมื่อข้อมูลไม่มีอันดับซ้ำกัน มีสูตรในการคำนวณดังนี้

$$r = 1 - \frac{6\Sigma d^2}{n^3 - n}$$

เมื่อ d หมายถึง ผลต่างของอันดับ

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ที่มาของสูตร จะเสนอดังต่อไปนี้

เนื่องจาก X และ Y เป็นอันดับที่ ดังนั้นตัวแปร X และ Y จะมีค่าเป็น 1,2,...,n ซึ่งอาจจะสลับที่กันอย่างไรก็แล้วแต่ข้อมูล

$$\bar{Y} = \bar{X} = \Sigma \frac{X}{N} = \frac{1 + 2 + \ldots + n}{n} = \frac{n + 1}{n}$$

และ $$\Sigma Y^2 = \Sigma X^2 = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

ดังนั้น $$\Sigma x^2 = \Sigma(X - \bar{X})^2 = \Sigma X^2 - \bar{X}\Sigma X = \Sigma Y^2 - \bar{Y}\Sigma Y$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n+1}{2} \cdot \frac{n(n+1)}{2}$$

$$= \frac{n(n^2-1)}{12}$$

ถ้าให้ d = อันดับที่ของ X − อันดับที่ของ Y ในแต่ละคู่

$$= X - Y$$

$$= (X - \bar{X}) - (Y - \bar{Y}) \qquad \because \bar{X} = \bar{Y}$$

$$= x - y$$

$$\Sigma d^2 = \Sigma(x - y)^2$$

$$= \Sigma x^2 + \Sigma y^2 - 2\Sigma xy$$

หรือ $$\Sigma xy = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d^2}{2}$$

ในการหาค่าสัมประสิทธิ์สหสัมพันธ์โดยทั่วไปใช้สูตร

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}}$$

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d^2}{2\sqrt{\Sigma x^2 . \Sigma y^2}}$$

เนื่องจาก $\Sigma x^2 = \Sigma y^2$

ดังนั้น $$r = \frac{2\Sigma x^2 - \Sigma d^2}{2\Sigma x^2} = 1 - \frac{\Sigma d^2}{2\Sigma x^2} = 1 - \frac{\Sigma d^2}{\frac{2n(n^2-1)}{12}}$$

$$= 1 - \frac{6\Sigma d^2}{n^3-n}$$

2. กรณีข้อมูลมีอันดับซ้ำกัน

กรณีที่ข้อมูลมีอันดับซ้ำกัน อาจมีข้อมูลซ้ำได้ในกรณีต่อไปนี้

1. มีข้อมูลซ้ำเฉพาะข้อมูลในตัวแปร X หรือ
2. มีข้อมูลซ้ำเฉพาะข้อมูลในตัวแปร Y หรือ
3. มีข้อมูลซ้ำทั้งข้อมูลในตัวแปร X และ Y

กรณีที่มีอันดับซ้ำกันในกรณีใดกรณีหนึ่งดังกล่าว การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน ต้องมีการแก้ไขความคลาดเคลื่อนให้ถูกต้อง ดังนั้นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนจึงต้องใช้สูตรใหม่ดังนี้ (Kendall 1975 : 38)

$$r_a = 1 - \frac{6\left\{\Sigma d^2 + T_x + T_y\right\}}{n^3 - n}$$

$$r_b = \frac{\frac{1}{6}(n^3 - n) - \Sigma d^2 - T_x - T_y}{\sqrt{\frac{1}{6}(n^3-n)-2T_x}\sqrt{\frac{1}{6}(n^3-n)-2T_y}}$$

เมื่อ T_x หมายถึง ค่าแก้เมื่อตัวแปร X มีค่าซ้ำกัน

โดยที่ $T_x = \frac{1}{12}\Sigma(t_x^3 - t_x)$

T_y หมายถึง ค่าแก้เมื่อตัวแปร Y มีค่าซ้ำกัน

โดยที่ $T_y = \frac{1}{12}\Sigma(t_y^3 - t_y)$

t_x หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร X

t_y หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร Y

ถึงแม้จะมีสูตรในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนในกรณีนี้ถึง 2 สูตรก็ตาม แต่ส่วนใหญ่มักนิยมใช้ค่าสัมประสิทธิ์ r_b มากกว่า

3. กรณีข้อมูลมีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร

นอกจากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน ใน 2 กรณีดังกล่าวแล้ว ในปี ค.ศ.1960 Alan Stuart ได้พัฒนาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนมาประยุกต์ใช้กับข้อมูลที่มีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร r X c (r หมายถึง แถว และ c หมายถึง สดมภ์) ซึ่งตามปกติวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวจะใช้กับข้อมูลที่มีจำนวนมาก การจัดอันดับของข้อมูลลงในตารางการณ์จรย่อมเกิดการซ้ำของอันดับโดยอัตโนมัติ และการซ้ำของอันดับดังกล่าวจะเกิดการซ้ำเป็นจำนวนมาก การหาค่าสัมประสิทธิ์ r_a และ r_b จึงไม่เหมาะที่จะนำมาใช้กับข้อมูลในลักษณะนี้ สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์วิธีนี้คือ r_c

การคำนวณหาค่าสัมประสิทธิ์ r_c ตามวิธีของStuart นั้นได้เสนอวิธีการหาค่าต่างๆ เพื่อนำมาใช้ในสูตรการคำนวณดังแสดงในแผนภาพตารางต่อไปนี้

	แยกประเภทตามสดมภ์				รวมความถี่ (แถว)	ค่าเฉลี่ยของอันดับ (แถว)
	1 (ค่าต่ำสุด)	2	....	c (ค่าสูงสุด)		
แยกประเภทตามแถว 1 (ค่าต่ำสุด)	n_{11}	n_{12}	....	n_{1c}	$n_{1.}$	$R_{1.} = \frac{n_{1.}+1}{2}$
2	n_{21}	n_{22}	....	n_{2c}	$n_{2.}$	$R_{2.} = n_{1.} + \frac{n_{2.}+1}{2}$
3	n_{31}	n_{32}	....	n_{3c}	$n_{3.}$	$R_{3.} = \sum_{i=1}^{2} n_{i.} + \frac{n_{3.}+1}{2}$
⋮	⋮	⋮		⋮	⋮	⋮
r (ค่าสูงสุด)	n_{r1}	n_{r2}	....	n_{rc}	$n_{r.}$	$R_{r.} = \sum_{i=1}^{r-1} n_{i.} + \frac{n_{r.}+1}{2}$
รวมความถี่ (สดมภ์)	$n_{.1}$	$n_{.2}$	....	$n_{.c}$	N (ความถี่รวมทั้งหมด)	
ค่าเฉลี่ยของอันดับ (สดมภ์)	$R_{.1} = \frac{n_{.1}+1}{2}$	$R_{.2} = n_{.1} + \frac{n_{.2}+1}{2}$	.....	$R_{.c} = \sum_{j=1}^{c-1} n_{.j} + \frac{n_{.c}+1}{2}$		

ดังนั้นการคำนวณหาค่าสัมประสิทธิ์ r_c สามารถคำนวณได้จากสูตรต่อไปนี้

$$r_c = 1 - \frac{6m^2\Sigma d^2}{n^3(m^2 - 1)}$$

เมื่อ m หมายถึง ค่าต่ำสุดของจำนวนแถวหรือจำนวนสดมภ์

<u>ข้อตกลงเบื้องต้นของสัมประสิทธิ์</u> r_c

1. มีความเป็นอิสระในการรวบรวมข้อมูล
2. ข้อมูลจัดอยู่ในรูปตารางการณ์จร r X c
3. ตัวแปรทั้งสองมีระดับการวัดอยู่ในมาตราจัดอันดับ
4. ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นความสัมพันธ์เชิงเส้นตรง
5. ผลรวมความถี่ของแถวและสดมภ์ใดๆ ต้องไม่เท่ากับศูนย์

<u>การทดสอบนัยสำคัญของสัมประสิทธิ์</u> r_c

เพื่อให้เกิดความเชื่อมั่นว่าค่าสัมประสิทธิ์ r_c ที่คำนวณได้นั้นจะมีความสัมพันธ์กันจริงหรือไม่ จึงต้องมีการทดสอบนัยสำคัญของค่าสัมประสิทธิ์ r_c ด้วย ในการทดสอบนัยสำคัญจะตั้งสมมติฐานว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กัน หรือความสัมพันธ์มีค่าเท่ากับ 0 นั่นคือ $H_o: \rho = 0$ และตั้งสมมติฐานรองว่า ความสัมพันธ์ไม่เท่ากับ 0 หรือ $H_1 : \rho \neq 0$

เนื่องจากค่าสัมประสิทธิ์ r_c มีลักษณะการแจกแจงเช่นเดียวกับการแจกแจงที (Student's t - distribution) ที่ชั้นแห่งความเป็นอิสระ(degree of freedom) เท่ากับ n - 2 ดังนั้นการทดสอบนัยสำคัญของค่าสัมประสิทธิ์ r_c จึงใช้สูตรการทดสอบที (t - test) ในการคำนวณ ซึ่งมีสูตรดังนี้

$$t = r_c\sqrt{(n-2)}/\sqrt{(1-r_c^2)} \qquad ; \quad \nu = n - 2$$

เมื่อ r_c หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ν หมายถึง ชั้นแห่งความเป็นอิสระ

<u>สหสัมพันธ์ของเคนดอล</u> (Kendall's Rank Correlation)

การหาค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลที่เป็นอันดับ นอกจากจะใช้วิธีของสเปียร์แมน ซึ่งเป็นการพิจารณาความสัมพันธ์โดยใช้ผลต่างระหว่างอันดับของตัวแปรแล้ว อาจทำได้อีกวิธีหนึ่ง โดยการพิจารณาถึงการเรียงอันดับของข้อมูลในตัวแปร ทั้งนี้เพราะถ้าตัวแปรทั้งคู่มีความสอดคล้องกันดี หรือมีความสัมพันธ์กันมาก การเรียงอันดับควรจะเป็นไปในลักษณะเดียวกัน หรือเป็นไปในลักษณะธรรมชาติ (natural order) ถ้าการเรียงอันดับของข้อมูลเป็นไปอย่างไม่มีระบบ แสดงว่าข้อมูลเหล่านั้นอาจไม่มีความสัมพันธ์กัน หรือมีความสัมพันธ์กันน้อยมาก ผู้คิดวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลในลักษณะดังกล่าวคือ Maurice G. Kendall ที่คิดขึ้นในปี ค.ศ.1938 ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้จากวิธีของเคนดอลแทนด้วยสัญลักษณ์ τ (Tau อ่านว่า เทา) ซึ่งเคนดอลได้นิยามค่าสัมประสิทธิ์ τ ไว้ดังนี้ (Kendall 1938 : 82)

$$\tau = \frac{\text{actual score}}{\text{maximum possible score}}$$

โดยที่ actual score คือผลรวมของจำนวนคู่ของตัวแปร Y ที่เกิดขึ้นจริง ซึ่งจะใช้สัญลักษณ์ S แทน และ maximum possible score คือผลรวมสูงสุดของจำนวนคู่ทั้งหมดของตัวแปร Y ที่เป็นไปได้ตามธรรมชาติ

สูตรที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์ τ สามารถกำหนดใช้ได้ตามเงื่อนไขใน 3 กรณีต่อไปนี้

1. กรณีข้อมูลมีอันดับไม่ซ้ำกัน

ถ้าข้อมูลมีทั้งหมด n ตัว จำนวนคู่ทั้งหมดที่เป็นไปได้ของตัวแปร Y คือ $\binom{n}{2}$ หรือ ${}^{n}C_2 = \frac{n(n-1)}{2}$ ถ้าทุกคู่มีการเรียงอันดับเป็นไปตามธรรมชาติทั้งหมด โดยให้แต่ละคู่มีค่าเท่ากับ 1 จะได้ผลรวมสูงสุดของจำนวนคู่ของตัวแปร Y เป็น $\frac{n(n-1)}{2}$ ด้วย ดังนั้น

$$\tau_a = \frac{S}{\frac{n(n-1)}{2}}$$

เมื่อ τ_a หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลกรณีข้อมูลมีอันดับไม่ซ้ำกัน

S หมายถึง ผลต่างของความสอดคล้องและความไม่สอดคล้องของการจัดอันดับ

n หมายถึง ขนาดของกลุ่มตัวอย่าง

2. กรณีข้อมูลมีอันดับซ้ำกัน

กรณีที่ข้อมูลมีอันดับซ้ำกันชุดใดชุดหนึ่งหรือทั้งสองชุด การกำหนดอันดับที่ซ้ำกันโดยการเฉลี่ย จะมีผลทำให้ค่าผลรวมสูงสุดของจำนวนคู่ทั้งหมดของตัวแปร (maximum possible score) มีค่าน้อยกว่า $\frac{n(n-1)}{2}$ ทั้งนี้เพราะคู่ที่มีอันดับเท่ากันจะมีค่าเป็น 0 จึงต้องปรับค่าของผลรวมสูงสุดของจำนวนคู่ทั้งหมดของตัวแปรใหม่ ดังนั้นสูตรที่ใช้สำหรับคำนวณค่าสัมประสิทธิ์ τ ในกรณีนี้ คือ

$$\tau_b = \frac{S}{\sqrt{\frac{1}{2}n(n-1)-T_x}\sqrt{\frac{1}{2}n(n-1)-T_y}}$$

เมื่อ τ_b หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลกรณีข้อมูลมีอันดับซ้ำกัน

T_x หมายถึง ค่าแก้เมื่อตัวแปร X มีค่าซ้ำกัน

โดยที่ $T_x = \frac{1}{2}\Sigma t_x(t_x-1)$

T_y หมายถึง ค่าแก้เมื่อตัวแปร Y มีค่าซ้ำกัน

โดยที่ $T_y = \frac{1}{2}\Sigma t_y(t_y-1)$

t_x หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร X

t_y หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร Y

3. กรณีข้อมูลมีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร

นอกจากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอล ใน 2 กรณีดังกล่าวแล้ว ในปี ค.ศ.1953 Alan Stuart ได้พัฒนาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลมาประยุกต์ใช้กับข้อมูลที่มีการแยกประเภท และจัดอันดับในรูปตารางการณ์จร r X c

สามารถใช้ได้กับทุก ๆ ขนาดของ r และ c และสามารถทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าสูงสุดถึง ± 1 ซึ่งค่าของ τ_b สามารถจะให้ค่าสูงสุดถึง ± 1 ได้ก็ต่อเมื่อ r = c เท่านั้น (Kendall and Stuart 1960 : 563) วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ตามที่ Stuart ได้พัฒนาขึ้นมานี้จะใช้กับข้อมูลที่มีจำนวนมาก สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์ด้วยวิธีนี้คือ τ_c ซึ่งบางทีเรียกว่า Stuart's τ_c เพื่อให้เกียรติแก่ Stuart ที่พัฒนาวิธีนี้ขึ้นมา

การคำนวณหาค่าสัมประสิทธิ์ τ_c ยังคงใช้สัดส่วนระหว่าง actual score (S) กับ maximum possible score แต่สำหรับ maximum possible score ได้ดัดแปลงสูตรเป็น $\frac{1}{2}n^2\ (m-1)/m$ เพื่อให้สะดวกในการคิดคำนวณ ดังนั้นสูตรที่ใช้คำนวณค่าสัมประสิทธิ์ τ_c จึงมีสูตรในการคำนวณดังนี้

$$\tau_c = \frac{S}{\frac{1}{2}n^2\left[(m-1)/m\right]}$$

เมื่อ m หมายถึง ค่าต่ำสุดของจำนวนแถวหรือจำนวนสดมภ์

สำหรับการแจกแจงค่าสัมประสิทธิ์ τ_c จะมีลักษณะเป็นแบบปกติที่มีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$E(\tau_c) = \frac{(n-1)}{n}\ \frac{N}{(N-1)}\ \rho$$

$$Var(\tau_c) \leqslant \frac{2}{n}\left\{\left[\frac{(n-1)}{n}\ \frac{m}{(m-1)}\right]^2 - \left[\frac{(n-1)}{n}\ \frac{N}{(N-1)}\ \rho\right]^2\right\}$$

แต่ถ้าประชากร N และกลุ่มตัวอย่าง n มีขนาดใหญ่ จะสามารถประมาณค่าเฉลี่ยและความแปรปรวนได้ดังนี้

$$E(\tau_c) = \rho$$

$$Var(\tau_c) \leqslant \frac{2}{n}\left[\left(\frac{m}{m-1}\right)^2 - \rho\right]^2$$

<u>ข้อตกลงเบื้องต้นของสัมประสิทธิ์ τ_c</u>

1. มีความเป็นอิสระในการรวบรวมข้อมูล
2. ข้อมูลจัดอยู่ในรูปตารางการณ์จร r X c
3. ตัวแปรทั้งสองมีระดับการวัดอยู่ในมาตราจัดอันดับ
4. ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นความสัมพันธ์เชิงเส้นตรง
5. ผลรวมความถี่ของแถวและสดมภ์ใด ๆ ต้องไม่เท่ากับ 0

<u>การทดสอบนัยสำคัญของสัมประสิทธิ์ τ_c</u>

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์ τ_c ที่คำนวณได้นั้น จะทำให้รู้ว่าตัวแปรทั้งสองมีความสัมพันธ์กันจริงหรือไม่ ในการทดสอบนัยสำคัญของค่าสัมประสิทธิ์ τ_c จะตั้งสมมติฐานว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กัน หรือความสัมพันธ์มีค่าเท่ากับ 0 นั่นคือ $H_o: \rho = 0$ และตั้งสมมติฐานรองว่า ความสัมพันธ์ไม่เท่ากับ 0 หรือ $H_1 : \rho \neq 0$

เนื่องจากค่า τ_c มีลักษณะการแจกแจงเช่นเดียวกับการแจกแจงแบบปกติ (Normal Distribution) ดังนั้นการทดสอบนัยสำคัญของค่า τ_c จึงใช้สูตรการทดสอบซี (Z - test) ในการคำนวณ ซึ่งมีสูตรดังนี้

$$Z = \frac{S - 1}{\sigma}$$

เมื่อ S หมายถึง ผลต่างของความสอดคล้องและความไม่สอดคล้องของการจัดอันดับ

σ หมายถึง ค่าเบี่ยงเบนมาตรฐาน

โดยที่
$$\sigma = \frac{1}{8}\left\{n(n-1)(2n+5)-\Sigma t_x(t_x-1)(2t_x-5)-\Sigma t_y(t_y-1)(2t_y-5)\right\} + \frac{1}{9n(n-1)(n-2)}\left\{\Sigma t_x(t_x-1)(t_x-2)\right\}\left\{\Sigma t_y(t_y-1)(t_y-2)\right\} + \frac{1}{2n(n-1)}\left\{\Sigma t_x(t_x-1)\right\}\left\{\Sigma t_y(t_y-1)\right\}$$

ในที่นี้ n หมายถึง ขนาดของกลุ่มตัวอย่าง

t_x หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร X

t_y หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร Y

สหสัมพันธ์ของเครมเมอร์วี (Cramer's V Correlation)

ในการหาความสัมพันธ์ระหว่างตัวแปรสองตัว จากข้อมูลที่อยู่ในรูปตารางการณ์จร สามารถนำสถิติที่ใช้ไคสแควร์เป็นพื้นฐานมาหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทั้งสองได้หลายวิธี ได้แก่ สัมประสิทธิ์ฟี (ϕ) สัมประสิทธิ์ C (Contingency Coefficient) สัมประสิทธิ์ของ Tschuprow (T) และสัมประสิทธิ์ของเครมเมอร์ (V) ซึ่งแต่ละวิธีมีข้อจำกัดและมีความเหมาะสมกับสถานการณ์แตกต่างกัน กล่าวคือ

สัมประสิทธิ์ฟี จะมีค่าเท่ากับ 0 เมื่อตัวแปรไม่มีความสัมพันธ์กันเลย และจะมีค่าเท่ากับ 1 เมื่อตัวแปรทั้งสองมีความสัมพันธ์กันอย่างสมบูรณ์ แต่จะเท่ากับ 1 เฉพาะกรณีที่เป็นตารางการณ์จร 2 X 2 เท่านั้น ถ้าจำนวนแถวและจำนวนสดมภ์มากกว่านี้ ค่าสัมประสิทธิ์ฟีอาจจะมีค่ามากกว่า 1 ได้ (Jacobson 1976 : 437)

สำหรับสัมประสิทธิ์ C จะมีค่าเท่ากับ 0 เมื่อตัวแปรไม่มีความสัมพันธ์กันเลย แต่ไม่มีโอกาสที่จะมีค่าสูงถึง 1 ได้เลย เมื่อจำนวนแถวและสดมภ์มีค่าเท่ากัน ค่าสูงสุดของสัมประสิทธิ์ C จะมีค่าเท่ากับ $\sqrt{(r-1)/r}$ (r หมายถึง จำนวนแถวที่มีค่าเท่ากับจำนวนสดมภ์) เช่น ถ้าตารางการณ์จร 5 X 5 ค่าสัมประสิทธิ์ C จะมีค่าสูงสุดเท่ากับ 0.894 เป็นต้น

ส่วนสัมประสิทธิ์ T ระหว่างสองตัวแปรจะมีค่าเท่ากับ 1 เสมอ หากตัวแปรทั้งสองมีความสัมพันธ์กันอย่างสมบูรณ์ แต่มีข้อแม้ว่าจำนวนแถวและจำนวนสดมภ์จะต้องเท่ากัน ถ้าหากจำนวนแถวและจำนวนสดมภ์ไม่เท่ากัน ค่าสัมประสิทธิ์ T จะมีค่าต่ำกว่า 1

ด้วยข้อจำกัดต่าง ๆ ของการหาค่าสัมประสิทธิ์สหสัมพันธ์ในรูปตารางการณ์จร r X c ดังได้กล่าวแล้ว ในปี ค.ศ.1946 เครมเมอร์ (Harald Cramer) เป็นผู้คิดวิธีปรับปรุงแก้ไข โดยทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ไม่ขึ้นอยู่กับเงื่อนไขดังกล่าว การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของเครมเมอร์ มีสูตรที่ใช้ในการคำนวณดังนี้

$$V = \sqrt{\chi^2/n(m-1)}$$

โดยที่ $$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad ; \quad \nu = (r-1)(c-1)$$

เมื่อ V หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเครมเมอร์

X^2 หมายถึง ค่าไคสแควร์ (Pearson's X^2 Statistic)

n หมายถึง ขนาดของกลุ่มตัวอย่าง

m หมายถึง ค่าต่ำสุดของจำนวนแถวหรือจำนวนสดมภ์

O_{ij} หมายถึง ค่าความถี่ที่สังเกตได้ (Observed Frequency) ในแถวที่ i สดมภ์ที่ j

E_{ij} หมายถึง ค่าความถี่ที่คาดหวัง (Expected Frequency) ในแถวที่ i สดมภ์ที่ j

โดยที่ค่า E_{ij} สามารถคำนวณได้จากสูตร

$$E_{ij} = \frac{R_i C_j}{n}$$

เมื่อ R_i หมายถึง ความถี่รวมในแถวที่ i

C_j หมายถึง ความถี่รวมในสดมภ์ที่ j

<u>ข้อตกลงเบื้องต้นของสัมประสิทธิ์</u> V

การคำนวณค่าสัมประสิทธิ์ V นั้น นิยมคำนวณค่าไคสแควร์ก่อน ดังนั้นข้อมูลที่นำมาคำนวณค่าสัมประสิทธิ์ V นี้ จะต้องมีความสอดคล้องกับข้อตกลงเบื้องต้นของไคสแควร์ด้วย จึงจะทำให้ค่าสัมประสิทธิ์ V ที่คำนวณได้มีความเหมาะสมถูกต้อง ข้อตกลงเบื้องต้นของสัมประสิทธิ์ V มีดังต่อไปนี้ (Marascuilo 1971 : 410)

1. มีความเป็นอิสระในการรวบรวมข้อมูล
2. ตัวแปรทั้งสองมีลักษณะเป็น multinomial
3. ค่าความถี่ที่คาดหวังภายในเซลของตารางต้องมีค่ามากกว่า 5
4. มีกลุ่มตัวอย่างเพียงกลุ่มเดียว

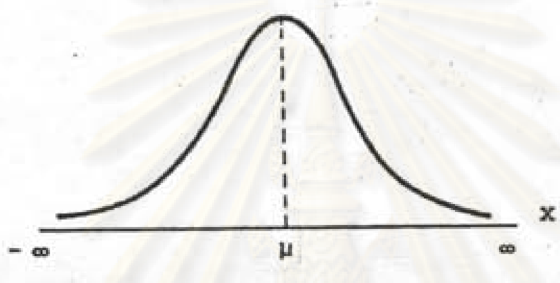
<u>การทดสอบนัยสำคัญของสัมประสิทธิ์ V</u>

ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ เช่น r_c และ τ_c เป็นต้น เมื่อคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แล้ว จึงนำไปทดสอบนัยสำคัญเพื่อต้องการทราบว่าค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีความเชื่อมั่นอย่างมีนัยสำคัญทางสถิติหรือไม่ โดยนำค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ไปทดสอบด้วยวิธีการทางสถิติที่เหมาะสม แต่สำหรับค่าสัมประสิทธิ์ V นั้นถือได้ว่าเป็นกรณีพิเศษ เพราะค่าสัมประสิทธิ์ V ที่ได้นี้ต้องคำนวณหาค่าไคสแควร์มาก่อน ซึ่งค่าไคสแควร์ที่ได้นี้ความจริงเป็นการทดสอบความเป็นอิสระของตัวแปรทั้งสองว่ามีความสัมพันธ์กันหรือไม่ โดยเมื่อคำนวณค่าไคสแควร์ได้แล้ว ก็นำค่าที่ได้ไปเปรียบเทียบกับค่าไคสแควร์ในตารางมาตรฐาน ที่ชั้นแห่งความเป็นอิสระเท่ากับ $(r-1)(c-1)$ ปกติเมื่อค่าไคสแควร์ที่คำนวณได้มีนัยสำคัญทางสถิติแล้วจะเป็นการบอกแต่เพียงว่าตัวแปรทั้งสองมีความสัมพันธ์กันเท่านั้น ยังไม่รู้ว่าตัวแปรทั้งสองมีความสัมพันธ์กันในระดับใด ซึ่งสามารถรู้ได้จากการคำนวณหาค่าสัมประสิทธิ์ต่อไปในภายหลัง

ตอนที่ 3 การแจกแจงที่เกี่ยวข้องกับงานวิจัยนี้

การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงที่สำคัญ และใช้ประโยชน์มากในวิชาสถิติทั้งทางด้านทฤษฎีและด้านปฏิบัติ คือการแจกแจงแบบปกติ ซึ่งเป็นการแจกแจงแบบต่อเนื่อง (continuous distribution) กราฟของการแจกแจงแบบปกติเรียกว่า โค้งปกติ (normal curve) มีลักษณะคล้ายระฆังคว่ำ ดังแผนภาพที่ 2



แผนภาพที่ 2 ลักษณะของโค้งปกติ

การแจกแจงแบบปกติค้นพบในปี ค.ศ.1733 โดย Abraham De Moire นักคณิตศาสตร์ชาวฝรั่งเศส แต่ไม่ได้นำมาใช้อย่างแพร่หลายจนกระทั่ง Pierre Laplace และ Karl Fredrich Gauss ได้พบการแจกแจงแบบปกติโดยแต่ละฝ่ายไม่ทราบผลงานของ De Moire มาก่อน บุคคลทั้งสองได้ทำการศึกษาการแจกแจงของความคลาดเคลื่อนในการวัดทางวิทยาศาสตร์กายภาพ ด้วยการวัดซ้ำ ๆ กัน และพบว่าผลของการแจกแจงเป็นการแจกแจงแบบปกติซึ่งเรียกว่า โค้งปกติของความคลาดเคลื่อน (The normal curve of error) และเพื่อเป็นเกียรติแก่ Laplace และ Gauss บางครั้งจึงเรียกการแจกแจงแบบปกติว่า การแจกแจงของลาปลาส (Laplacian Distribution) หรือ การแจกแจงของเกาส์ (Gaussian Distribution)

สมการของการแจกแจงแบบปกติ

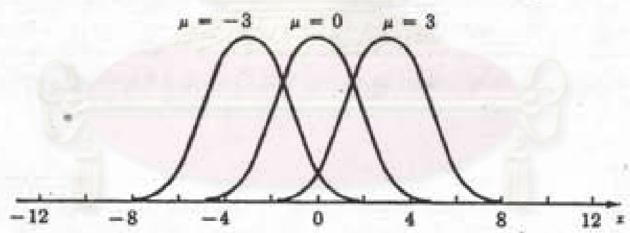
ตัวแปรสุ่ม X มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 ฟังก์ชั่นความน่าจะเป็น (Probability density function) ของ X กำหนดได้ดังนี้คือ

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

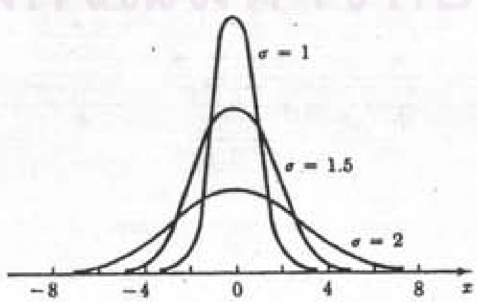
เมื่อ μ และ σ เป็นพารามิเตอร์ โดยที่ $-\infty < \mu < \infty$ และ $\sigma > 0$

$\pi = 3.14159...$, $e = 2.71828...$

เนื่องจากสมการของการแจกแจงแบบปกติขึ้นอยู่กับพารามิเตอร์ 2 ตัวคือ μ และ σ เราอาจจะเขียนฟังก์ชั่นความน่าจะเป็นของ X แทนด้วย $N(\mu,\sigma^2)$ ซึ่งค่าของ μ และ σ จะเป็นตัวกำหนดตำแหน่งและรูปร่างของโค้งปกติ ดังแผนภาพที่ 3 และ 4



แผนภาพที่ 3 การแจกแจงแบบปกติที่มีค่าเฉลี่ยแตกต่างกัน แต่มีส่วนเบี่ยงเบนมาตรฐานเท่ากัน ($\sigma = 1.5$)



แผนภาพที่ 4 การแจกแจงแบบปกติที่มีส่วนเบี่ยงเบนมาตรฐานแตกต่างกัน แต่มีค่าเฉลี่ยเท่ากัน ($\mu = 0$)

คุณสมบัติของการแจกแจงแบบปกติ

1. มีลักษณะเป็นรูประฆังคว่ำ ส่วนสูงของโค้งขึ้นอยู่กับส่วนเบี่ยงเบนมาตรฐาน ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่ามาก ส่วนสูงของโค้งจะแบนลาดลง แต่ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าน้อย ส่วนสูงของโค้งจะมีค่ามาก และมีส่วนโค้งสูงสุดอยู่ที่ $X = \mu$

2. มีลักษณะสมมาตร

3. เส้นโค้งมีจุดเปลี่ยนโค้ง (point of influction) อยู่ที่ $X = \mu \pm \sigma$

4. ค่าเฉลี่ย มัธยฐาน และฐานนิยม อยู่ที่จุดเดียวกัน

5. พื้นที่ทั้งหมดที่อยู่ภายใต้เส้นโค้ง มีค่าเท่ากับ 1

6. ปลายโค้งปกติทั้งสองด้านจะลาดลงต่ำเข้าหาแกน X เมื่อ X มีค่าห่างจาก μ ไปทุกที แต่จะไม่สัมผัสกับแกน X โดยปลายโค้งจะมีขอบเขต $-\infty$ ถึง ∞

การแจกแจงแบบปกติที่มีค่า $\mu = 0$ และ $\sigma = 1$ เรียกว่าการแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) ใช้สัญลักษณ์ N(0,1) โดยปกติแล้วมักกำหนดให้ Z เป็นตัวแปรสุ่มของการแจกแจงแบบปกติมาตรฐาน ดังนั้น Z มีการแจกแจงแบบปกติมาตรฐาน บางทีเรียกว่า คะแนนมาตรฐาน ซึ่งมีฟังก์ชั่นความน่าจะเป็น ดังนี้

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad ; \quad -\infty < z < \infty$$

ค่าเฉลี่ยจะอยู่ที่จุด Z = 0 จุดเปลี่ยนโค้งอยู่ที่จุด Z = -1 และ Z = 1 เนื่องจากการแจกแจงแบบปกติมาตรฐานเป็นกรณีเฉพาะของการแจกแจงแบบปกติ ดังนั้น คุณสมบัติต่าง ๆ ก็เป็นเช่นเดียวกับคุณสมบัติของการแจกแจงแบบปกติ

การแจกแจงไคสแควร์ (Chi-Square Distribution)

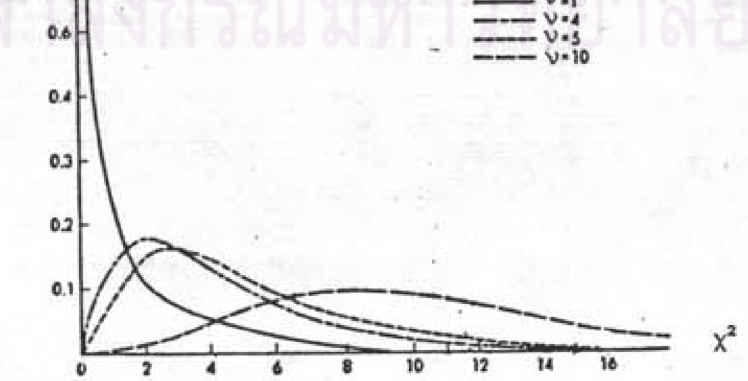
การแจกแจงไคสแควร์เป็นการแจกแจงอีกแบบหนึ่งที่มีความสำคัญมากซึ่ง F.R. Helmert นักฟิสิกส์ชาวเยอรมัน ได้เป็นผู้ค้นคิดเป็นคนแรกในปี ค.ศ.1876 ต่อมาในปี ค.ศ.1900 Karl Pearson ได้พัฒนาและเผยแพร่จนเป็นที่รู้จักกันอย่างแพร่หลาย

สมการของการแจกแจงไคสแควร์

การแจกแจงไคสแควร์ เป็นการแจกแจงประเภทหนึ่งของการแจกแจงแกมม่า (gamma distribution) ดังนั้นถ้า X เป็นตัวแปรสุ่มแบบต่อเนื่อง และมี ν เป็นชั้นแห่งความเป็นอิสระที่มีค่าเป็นจำนวนเต็มบวกแล้ว ฟังก์ชั่นความน่าจะเป็นของตัวแปร X เป็นดังนี้

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} e^{-x/2} & ; 0 < x < \infty \\ 0 & ; x \text{ มีค่าอื่น ๆ} \end{cases}$$

เมื่อ $\Gamma(p)$ คือฟังก์ชั่นแกมม่ามีค่าตามสูตร $= \int_0^\infty x^{p-1} e^{-x} dx \quad ; p > 0$ เรียกตัวแปรสุ่ม X ว่ามีการแจกแจงไคสแควร์ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν การแจกแจงไคสแควร์จะกำหนดด้วยสัญลักษณ์ χ^2_ν ซึ่งกราฟของการแจกแจงไคสแควร์มีลักษณะดังแผนภาพที่ 5



แผนภาพที่ 5 การแจกแจงไคสแควร์ที่มีชั้นแห่งความเป็นอิสระต่างกัน

ถ้า s^2 เป็นความแปรปรวนของตัวอย่างเชิงสุ่มขนาด n ที่สุ่มจากประชากรที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 ดังนั้น

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

จะมีค่าการแจกแจงแบบไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ n - 1

<u>คุณสมบัติของการแจกแจงไคสแควร์</u>

1. ลักษณะการแจกแจงไม่สมมาตร โดย χ^2 จะมีค่าตั้งแต่ 0 ถึง ∞

2. การแจกแจงจะมีลักษณะแตกต่างกันตามค่าของชั้นแห่งความเป็นอิสระ ถ้า $\nu \geqslant 3$ การแจกแจงจะเบ้บวกหรือเบ้ขวา (positively skewed) และเมื่อชั้นแห่งความเป็นอิสระมีค่าเพิ่มขึ้น การแจกแจงไคสแควร์จะมีลักษณะใกล้เป็นการแจกแจงแบบปกติมาตรฐาน

3. ค่าออร์ดิเนตสูงสุด จะอยู่ที่ $\chi^2_{\nu-1}$

4. จุดสูงสุดของโค้งมีเพียงจุดเดียว (unimodal)

5. การแจกแจงเป็นไปตามกฎของการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution)

6. สำหรับการแจกแจงของไคสแควร์ เมื่อชั้นแห่งความเป็นอิสระเท่ากับ ν จะมีลักษณะดังนี้

6.1 ค่าเฉลี่ยจะมีค่าเท่ากับ ν และค่าความแปรปรวนเท่ากับ 2ν

6.2 ค่ามัธยฐานเท่ากับ $\nu - 2$

6.3 ค่าความเบ้ของการแจกแจงเท่ากับ $\sqrt{8/\nu}$

<u>การแจกแจงที</u> (Student's t - Distribution)

การแจกแจงที่มีการนำมาประยุกต์ใช้กันมากอีกอย่างหนึ่งคือ การแจกแจงที ค้นคิดโดย William Sealy Gosset ตีพิมพ์ครั้งแรกในปี ค.ศ.1908 ซึ่งใช้นามปากกาว่า "Student" ดังนั้นการแจกแจงทีจึงมีชื่อว่า Student's t-Distribution หรือบางทีเรียกว่า t - distribution

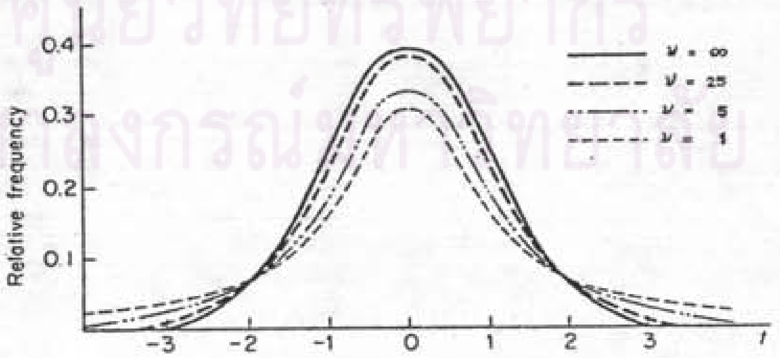
<u>สมการของการแจกแจงที</u>

ถ้า T เป็นตัวแปรสุ่มแบบต่อเนื่อง ฟังก์ชั่นความน่าจะเป็นของตัวแปร T เป็นดังนี้

$$f(t) = \frac{\Gamma(\nu+1)/2}{\sqrt{\nu\pi}\ \Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad ; -\infty < t < \infty$$

เมื่อ $\Gamma(p)$ คือฟังก์ชั่นแกมม่ามีค่าตามสูตร $= \int_0^\infty x^{p-1} e^{-x}\, dx \quad ; p > 0$

เรียกตัวแปรสุ่ม T ว่ามีการแจกแจงที ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν ตัวอย่างกราฟของการแจกแจงที แสดงดังแผนภาพที่ 6



แผนภาพที่ 6 การแจกแจงที เมื่อมีชั้นแห่งความเป็นอิสระต่างกัน

ถ้า Z เป็นตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐาน และ V เป็นตัวแปรสุ่มที่มีการแจกแจงไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν และถ้า Z และ V เป็นตัวแปรสุ่มอิสระแล้ว ตัวแปรสุ่ม T ซึ่งมีสมการดังนี้คือ

$$T = \frac{Z}{\sqrt{V/\nu}}$$

ตัวแปรสุ่ม T ดังกล่าวจะมีการแจกแจงเป็นการแจกแจงที ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν

<u>คุณสมบัติของการแจกแจงที</u>

1. มีลักษณะสมมาตร เมื่อเทียบกับแกน $t = 0$ จุดนี้จะเป็นค่าเฉลี่ย มัธยฐาน และฐานนิยม

2. ในกรณีที่ $\nu = 1$ การแจกแจงจะไม่มีค่าเฉลี่ย แต่ถ้า $\nu = 2,3,...$ ค่าเฉลี่ยจะมีค่าเท่ากับ 0

3. เมื่อ $\nu = 1,2$ การแจกแจงจะไม่มีค่าความแปรปรวน สำหรับค่า $\nu = 3,4,...$ ค่าความแปรปรวนจะเท่ากับ $\nu / \nu - 2$

4. ถ้า ν มีค่ามาก t จะมีการแจกแจงใกล้เป็นการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 1

<u>การแจกแจงแบบปกติสองตัวแปร</u> (Bivariate Normal Distribution)

ถ้า X และ Y เป็นตัวแปรสุ่มแบบต่อเนื่อง มีการแจกแจงแบบปกติสองตัวแปร ฟังก์ชั่นความน่าจะเป็นทั้งสอง เขียนเป็นสมการได้ดังนี้

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-{}^{2})}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

เมื่อ $-\infty < x < \infty$, $-\infty < y < \infty$

การแจกแจงแบบปกติสองตัวแปรตามสมการดังกล่าว จะประกอบด้วย พารามิเตอร์ $\mu_1, \mu_2, \sigma_1, \sigma_2$ และ ρ โดยที่ $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $\sigma_1 > 0$, $\sigma_2 > 0$ และ $-1 < \rho < 1$

ตัวแปร X และ Y เป็นตัวแปรที่มีการแจกแจงแบบปกติสองตัวแปร ที่มี

ค่าเฉลี่ย $\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

และเมตริกความแปรปรวนร่วม $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

โดยที่ ρ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และ Y

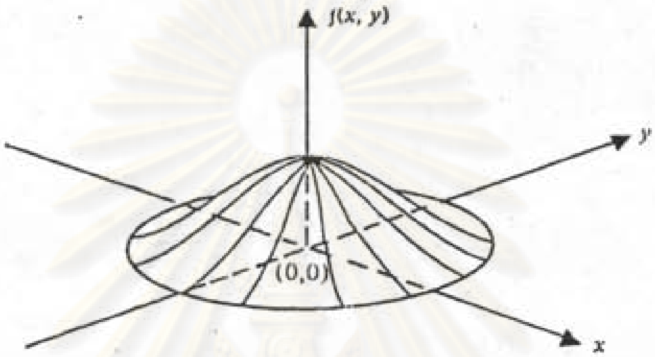
μ_1 หมายถึง ค่าเฉลี่ยของตัวแปร X

μ_2 หมายถึง ค่าเฉลี่ยของตัวแปร Y

σ_1^2 หมายถึง ค่าความแปรปรวนของตัวแปร X

σ_2^2 หมายถึง ค่าความแปรปรวนของตัวแปร Y

ในกรณีที่การแจกแจงแบบปกติสองตัวแปร มีค่า $\mu_1 = \mu_2 = 0$ และ $\sigma_1^2 = \sigma_2^2 = 1$ เรียกการแจกแจงแบบปกติสองตัวแปรนี้ว่า การแจกแจงแบบปกติมาตรฐานสองตัวแปร (Standard Bivariate Normal Distribution) ฟังก์ชั่นความน่าจะเป็นของตัวแปร X และ Y มีลักษณะเป็นรูประฆังคว่ำที่มีฐานอยู่บนระนาบ XY ดังแผนภาพที่ 7



แผนภาพที่ 7 การแจกแจงแบบปกติมาตรฐานสองตัวแปร

การแจกแจงแบบปกติมาตรฐานสองตัวแปร จะมีค่าเฉลี่ย $\underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ และเมตริกความแปรปรวนร่วม $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

<u>การสร้างประชากรที่มีการแจกแจงแบบปกติมาตรฐานสองตัวแปร</u>

ในการวิจัยครั้งนี้ ผู้วิจัยได้สร้างประชากรที่มีการแจกแจงแบบปกติมาตรฐานสองตัวแปรจากตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐานสองตัวคือ X และ Y ซึ่ง Y ที่ได้นี้แท้จริงแล้วเป็นค่าตัวแปร Y ที่ขึ้นอยู่กับค่าตัวแปร X ดังสมการ

$$Y = \rho X + \sqrt{1 - \rho^2} \cdot W$$

เมื่อ W คือค่าตัวแปรอิสระจากค่าตัวแปร X

สมการดังกล่าวนี้เป็นคุณสมบัติประการหนึ่งของการแจกแจงแบบปกติสองตัวแปร (Brownlee 1965 : 404) และการแจกแจงแบบมีเงื่อนไข (Conditional Distribution) กล่าวคือ ถ้าตัวแปรสุ่ม X และ Y มีการแจกแจงแบบปกติ

สองตัวแปร การแจกแจงแบบมีเงื่อนไขของตัวแปร X เมื่อกำหนดค่า $Y = y$ จะมีลักษณะเป็นการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ $\mu_1 + (\rho\sigma_1/\sigma_2)(y-\mu_2)$ และค่าความแปรปรวนเท่ากับ $\sigma_1^2\ (1-\rho^2)$

ส่วนการแจกแจงแบบมีเงื่อนไขของตัวแปร Y เมื่อกำหนดค่า $X = x$ จะมีลักษณะเป็นการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ $\mu_2 + (\rho\sigma_2/\sigma_1)(x-\mu_1)$ และค่าความแปรปรวนเท่ากับ $\sigma_2^2\ (1-\rho^2)$

การวิจัยครั้งนี้ได้ทำการศึกษาเฉพาะกรณีที่ $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ และ $\rho = \mathrm{cov}(X,Y)$ ดังนั้นถ้าจะหาการแจกแจงของตัวแปร $Y|X=x$ สามารถหาได้จากสมการ

$$\mu_{Y|X=x} = \mu_2 + (\rho\sigma_2/\sigma_1)(x-\mu_1)$$

$$\mu_{Y|X=x} = \rho x \qquad (\text{เมื่อ } \mu_1 = \mu_2 = 0 \text{ และ } \sigma_1^2 = \sigma_2^2 = 1)$$

และ

$$\sigma^2_{Y|X=x} = \sigma_2^2\ (1-\rho^2)$$

$$\sigma^2_{Y|X=x} = 1-\rho^2 \quad (\text{เมื่อ } \sigma_2^2 = 1)$$

$$\therefore \quad Y \sim N(\rho x\ ,\ 1-\rho^2)$$

นั่นคือ ตัวแปร Y จะมีลักษณะการแจกแจงเป็นการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ ρx และค่าความแปรปรวนเท่ากับ $1-\rho^2$

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

Chow, Miller, and Dickinson (1974:189-195) ได้ศึกษาเปรียบเทียบคุณสมบัติบางประการของค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของสเปียร์แมน และของเคนดอล โดยใช้วิธีมอนติคาร์โล ซิมูเลชั่น เมื่อกลุ่มตัวอย่างขนาดเล็กจำนวน 4 กลุ่ม ขนาด 5,10, 15 และ 20 และค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร ρ = 0.0,0.1,...,0.9 ผลปรากฏว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน จะประมาณค่า ρ ได้ใกล้เคียงกว่าค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอล ที่ทุก ๆ ค่าของ ρ ที่ทำการศึกษา แต่การประมาณค่า ρ ของค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 2 วิธี ก็ยังต่ำกว่าค่าของ ρ สำหรับอำนาจการทดสอบของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ในการปฏิเสธสมมติฐานสูญ $H_o : \rho = 0$ ของค่าสัมประสิทธิ์สหสัมพันธ์ทั้งสองจะไม่แตกต่างกันที่กลุ่มตัวอย่างขนาด 5 แต่สำหรับกลุ่มตัวอย่างขนาด 10,15 และ 20 นั้น ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน มีอำนาจในการปฏิเสธสมมติฐานสูญสูงกว่าค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอล ทั้งที่ระดับ $\alpha = .05$ และ .01 แต่ค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอล มีค่าความแปรปรวนของการแจกแจงน้อยกว่าค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน ในการเลือกใช้ค่าสัมประสิทธิ์สหสัมพันธ์วิธีใดนั้น คณะผู้ทำการศึกษาแนะนำให้เลือกใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนในการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบมีอันดับ

Agresti (1976 : 49-55) ได้ศึกษาเปรียบเทียบความมีเสถียรภาพของการวัด (Stability of the Measure) ความสัมพันธ์แบบอันดับในรูปของตารางจำแนกประเภท (cross classification tables) วิธีต่าง ๆ ได้แก่ วิธีหาค่าสัมประสิทธิ์แกมม่า(γ), τ_b, τ_c, r_b, r_c และ r_{xy} เมื่อข้อมูลมีลักษณะการแจกแจงแบบปกติสองตัวแปร และศึกษาที่ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร ρ = 0.2,0.5 และ 0.8 โดยมีการเปลี่ยนค่าของสิ่งที่จะทำการศึกษาบางอย่าง คือจำนวนแถวและจำนวนสดมภ์และผลรวมของความถี่ตามแนวแถวและสดมภ์ ผลปรากฏว่าเมื่อข้อมูลไม่ได้จัดกลุ่ม (ungroup data) ค่าสัมประสิทธิ์ τ_b มีเสถียรภาพดีที่สุด โดยพิจารณาจากค่า mean square error แต่ค่าสัมประสิทธิ์ τ_b ก็มีข้อบกพร่องบางประการเกี่ยวกับเรื่องประสิทธิภาพในการทดสอบสมมติฐาน และมีความผันแปรไปตามสัดส่วนของข้อมูลที่มีการซ้ำกันอีกด้วย

Brown and Benedetti (1977 : 309-315) ได้พัฒนาสูตรการคิดคำนวณค่า asymptotic standard error (ASE) ของค่าสัมประสิทธิ์สหสัมพันธ์ γ, τ_b, τ_c, r และ r_{xy} ขึ้นมาอีกลักษณะหนึ่งเพื่อนำมาใช้ในการทดสอบสมมติฐานสูญที่ว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กัน หรือความสัมพันธ์มีค่าเท่ากับ 0 ($H_o : \rho = 0$) โดยจะใช้สถิติทดสอบซี (Z - test) ในการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละวิธี แล้วเปรียบเทียบผลที่ได้กับการทดสอบสมมติฐานสูญ โดยใช้ค่า ASE ตามสูตรเดิมซึ่งในการศึกษาครั้งนี้ คณะผู้วิจัยได้ทำการทดลองซ้ำในแต่ละกรณีจำนวน 1,000 ครั้ง เมื่อข้อมูลอยู่ในรูปตารางการณ์จรขนาด 4 X 4 และ 8 X 8 โดยตารางการณ์จร 4 X 4 จะศึกษากับกลุ่มตัวอย่างขนาด 25,50,100 และ 200 ส่วนตารางการณ์จร 8 X 8 ศึกษากับกลุ่มตัวอย่างขนาด 25,50,100,200 และ 400 ที่ระดับนัยสำคัญทางสถิติเท่ากับ 0.20,0.10,0.05,0.025 และ 0.01 ผลปรากฏว่าการทดสอบซีที่ใช้ในการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ค่า ASE ที่คณะผู้วิจัยได้พัฒนาขึ้นมาใหม่มีความสามารถในการทดสอบสมมติฐานสูญได้ดีกว่าการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ค่า ASE แบบเดิม

Särndal (1974 : 165-187) ได้ใช้วิธีการซิมูเลชั่นเพื่อทำการศึกษาเปรียบเทียบวิธีการวัดค่าความสัมพันธ์แบบต่าง ๆ จำนวน 15 วิธี ทั้งที่เป็นวิธีการวัดค่าความสัมพันธ์แบบเดิมและที่ผู้วิจัยได้พัฒนาขึ้นมาใหม่ โดยต้องการศึกษาเปรียบเทียบมโนทัศน์พื้นฐาน และคุณสมบัติของความสัมพันธ์แบบต่าง ๆ เมื่อตัวแปร X อยู่ในมาตรานามบัญญัติ ที่จำแนกประเภทได้ตั้งแต่ 2 ประเภทขึ้นไป และตัวแปร Y อยู่ในมาตรานามบัญญัติ จัดอันดับ และอันตรภาค ขนาดของตัวอย่างที่ศึกษาจำนวน 400 และ ρ = 0.0, 0.1,...,1.0 จากการศึกษาเปรียบเทียบวิธีการวัดค่าความสัมพันธ์ของเคนดอลเทา(τ_c) กับของเครมเมอร์วี (V) ซึ่งผู้วิจัยเปรียบเทียบเมื่อตัวแปรทั้งสองอยู่ในมาตรานามบัญญัติ และบันทึกข้อมูลลงในตารางการณ์จร ผลปรากฏว่าเมื่อ ρ มีค่าเพิ่มขึ้นจาก 0 ถึง 1 ทั้งสัมประสิทธิ์ τ_c และ V จะมีค่าเพิ่มขึ้นอย่างช้า ๆ จนมีค่าถึง 1 และเมื่อใช้ค่าแก้ในการวัดค่าความสัมพันธ์ ที่ $\rho = 0$ ทั้งสัมประสิทธิ์ τ_c และ V มีค่าใกล้เคียงกับค่าของ ρ แต่โดยทั่วไปแล้วการใช้ค่าแก้ในการวัดความสัมพันธ์ระหว่างตัวแปรทั้งสอง สัมประสิทธิ์ τ_c สามารถใช้ได้ผลดีกว่าสัมประสิทธิ์ V

วีณา เตชะพนาคร (2528 : 89-92) ได้ศึกษาความสัมพันธ์ระหว่างสัมประสิทธิ์ C และสัมประสิทธิ์ของเครมเมอร์วี กับค่าทดสอบไคสแควร์ โดยพิจารณาถึงอิทธิพลของขนาดตัวอย่าง ขนาดตาราง และการจัดแบ่งกลุ่มข้อมูล ซึ่งศึกษาจากข้อมูลเชิงสุ่มชนิดตัวแปรปกติสองตัวแปร เมื่อกลุ่มตัวอย่างขนาด 20,30,40,50,75 และ 100 ขนาดตารางระดับ 2 X 2, 2 X 3, 2 X 4, 2 X 5, 3 X 3, 3 X 4, 3 X 5, 4 X 4, 4 X 5 และ 5 X 5 และตัวแปรปกติสองตัวแปรมีสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ 0.00 ถึง 0.98 โดยแบ่งเป็นช่วงดังนี้ 0.00 ถึง 0.40 จะมีความห่าง 0.02 0.40 ถึง 0.60 จะมีความห่าง0.01 และ 0.60 ถึง 0.98 จะมีความห่าง 0.02 ผลปรากฏว่า เมื่อพิจารณาในแง่ของการทดสอบสมมติฐาน ขนาดตัวอย่างและขนาดตารางจะทำให้ระดับนัยสำคัญจากการจำลอง สูงกว่าระดับนัยสำคัญจากทฤษฎี ในแต่ละขนาดของตัวอย่างจะได้ค่าวิกฤตจากการจำลองแตกต่างกัน ที่ระดับนัยสำคัญและชั้นแห่งความเป็นอิสระเดียวกัน ถ้าพิจารณาในแง่ของความสัมพันธ์ระหว่างค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์กับค่าทดสอบไคสแควร์ พบว่าเมื่อขนาดของตัวอย่างและขนาดตารางเพิ่มขึ้นค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์จะลดลง ส่วนการจัดกลุ่มข้อมูลที่แตกต่างกันจะทำให้ค่าเฉลี่ยและความแปรปรวนของค่าไคสแควร์แตกต่างกันในแต่ละกลุ่ม