บทที่ 2

สถิติทดสอบและผลงานวิจัยที่เกี่ยวข้อง

ในการศึกษาเปรียบเทียบการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวนั้น เมื่อข้อมูลที่เก็บรวบรวมมาได้อยู่ในรูปของข้อมูลเชิงปริมาณ จะสามารถหาค่าความสัมพันธ์ระหว่างตัวแปร 2 ตัวได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน และทำการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวโดยการทดสอบนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์ ถ้าหากนำข้อมูลเชิงปริมาณมาจัดเสนอข้อมูลใหม่ในรูปตารางการณ์จร จะได้ค่าที่ปรากฏในตารางเป็นความถี่ของค่าสังเกตที่เก็บรวบรวมมา แล้วสามารถทำการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยวิธีการทดสอบแบบไคสแควร์ ซึ่งผลของการทดสอบไคสแควร์ขึ้นอยู่กับจำนวนร้อยละของขนาดความถี่คาดหวังน้อยกว่า 5 ด้วย ดังนั้น เพื่อทำการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว จึงสนใจผลการทดสอบทั้ง 2 วิธีการ ซึ่งวิธีการทดสอบทั้ง 2 วิธีที่สนใจคือ

1. การทดสอบสัมประสิทธิ์สหสัมพันธ์
2. การทดสอบไคสแควร์

2.1 <u>สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน</u>

การทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยการทดสอบสัมประสิทธิ์สหสัมพันธ์ต้องคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ก่อนแล้วนำค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้มาทำการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยการทดสอบนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์ นอกจากนี้ อาจใช้สมการความถดถอยเชิงเส้นเพื่อพยากรณ์ค่าตัวแปรตัวหนึ่ง เมื่อทราบค่าตัวแปรอีกตัวหนึ่ง ด้วยวิธีการประมาณค่าตัวแปรตาม เมื่อทราบค่าตัวแปรอิสระ

วิธีการทางสถิติที่ใช้ในการวัดขนาดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ของข้อมูลเชิงปริมาณที่นิยมใช้กันมาก คือ สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน (Pearson's Product Moment Coefficient of Correlation) อาจเรียกชื่อสั้น ๆ ว่า สัมประสิทธิ์สหสัมพันธ์ (Correlation of Coefficient) หรือบางครั้งอาจมีชื่อเรียกต่างกัน เช่น สหสัมพันธ์ของเพียร์สัน (Pearson's Correlation) สหสัมพันธ์อันดับศูนย์ (Zero Order Correlation) สหสัมพันธ์เชิงเส้น (Linear Correlation) หรือสหสัมพันธ์แบบง่าย (Simple Correlation)

เป็นต้น หากเป็นการวัดขนาดความสัมพันธ์ระหว่างตัวแปรมากกว่า 2 ตัวขึ้นไป จะเรียกว่าสหสัมพันธ์ร่วม (Multiple Correlation) โดยทั่วไปนิยมใช้ r เป็นสัญลักษณ์แทนสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน ซึ่งค่าของสัมประสิทธิ์สหสัมพันธ์ (r) จะเชื่อถือได้มากหรือน้อยนั้น ยังขึ้นอยู่กับข้อตกลงเบื้องต้นของสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน

2.1.1 ข้อตกลงเบื้องต้นของสัมประสิทธิ์สหสัมพันธ์

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันมีข้อตกลงเบื้องต้นดังนี้คือ

1. ข้อมูลได้มาจากตัวแปรสุ่มที่มีความสัมพันธ์กันแบบเส้นตรง (Linearity) หรือข้อมูลที่ได้ในแต่ละคู่เป็นอิสระต่อกันไม่มีความเกี่ยวข้องกับคู่อื่น

2. การแจกแจงของตัวแปรสุ่มเหล่านี้ต้องมีลักษณะการแจกแจงแบบปกติ (Normal Distribution) กล่าวคือ ตัวแปรสุ่ม x มีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเป็น μ_x และค่าความแปรปรวนเป็น σ_x^2 $\left[x \sim N(\mu_x , \sigma_x^2) \right]$ ตัวแปรสุ่ม y มีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเป็น μ_y และค่าความแปรปรวนเป็น σ_y^2 $\left[y \sim N(\mu_y , \sigma_y^2) \right]$

3. ตัวแปรสุ่ม 2 ตัวนี้ต้องมีความแปรปรวน (Variance) เท่ากัน (Homosecdasticity) หรือมีลักษณะการกระจายเหมือนกันคือ $E\left[e_x \; e_y \right] = \sigma^2$ เมื่อ $x = y$ โดยที่ $x = 1, 2, \ldots\ldots, n$, $y = 1, 2, \ldots\ldots\ldots, n$

ภายใต้ข้อตกลงเบื้องต้น จะได้ตัวแบบ (Model) ของสัมประสิทธิ์สหสัมพันธ์เป็นดังนี้ สมมติให้ x และ y แปรไปตามกันหรือมีการแจกแจงร่วมกัน (Joint Distribution) ถ้าตัวแบบของการแจกแจงร่วมเป็นการแจกแจงแบบปกติ (Normal Distribution) จะเรียกตัวแบบนี้ว่าการแจกแจงแบบปกติของตัวแปรร่วม 2 ตัวแปร (Bivariate Normal Distribution) ซึ่งมีฟังก์ชัน (Function) การแจกแจงความหนาแน่นดังนี้

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\frac{(x-\mu_x)^2}{\sigma_x} + \frac{(y-\mu_y)^2}{\sigma_y} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right\}\right]$$

ซึ่งในฟังก์ชันมีพารามิเตอร์อยู่ 5 ตัวคือ ค่าเบี่ยงเบนมาตรฐานของตัวแปร x (σ_x) ค่าเบี่ยงเบนมาตรฐานของตัวแปร y (σ_y) ค่าเฉลี่ยของตัวแปร x (μ_x) ค่าเฉลี่ยของตัวแปร y (μ_y) และค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร x กับ ตัวแปร y (ρ) ใช้วัดขนาดความสัมพันธ์เชิงเส้นระหว่างตัวแปร x กับตัวแปร y ถ้าหากไม่ทราบหรือไม่แน่ใจว่าการแจกแจงร่วมของตัวแปรที่เราสนใจเป็นการแจกแจงแบบปกติ อาจทำให้ผลลัพธ์ในการคำนวณไม่ดีเท่าที่ควรจะเป็น (วัชราภรณ์ สุริยาภิวัฒน์ 2527 : 415-418)

2.1.2 การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์

1. วิธีคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน ซึ่งสัมประสิทธิ์สหสัมพันธ์ชนิดมาตรฐานและใช้กันทั่วไปคือ สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน และมีสูตรพื้นฐานในการคำนวณค่าความสัมพันธ์ระหว่างตัวแปร 2 ตัวดังนี้ (จี.พี. กิลฟอร์ด และเบนจามิน ฟรัชเตอร์ 2526 : 111-118) คือ

$$\rho_{XY} = \frac{\Sigma xy}{N\sigma_X\sigma_Y} \quad ....(2.1)$$

เมื่อ ρ_{XY} คือ สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน

$x = (X-\bar{X})$ คือค่าของ X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X

$y = (Y-\bar{Y})$ คือค่าของ Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y

$\Sigma xy = \Sigma(X-\bar{X})(Y-\bar{Y})$ คือผลบวกของผลคูณระหว่างค่า X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X กับค่า Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y ในลำดับที่เดียวกัน

$\sigma_X = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$ คือส่วนเบี่ยงเบนมาตรฐานของตัวแปร X

$\sigma_Y = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{N}}$ คือส่วนเบี่ยงเบนมาตรฐานของตัวแปร Y

และ N คือจำนวนคู่ลำดับของตัวแปร X กับ Y

2. วิธีลัดในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันคือ ไม่ต้องใช้ค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปร X (σ_X) และค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปร Y (σ_Y) เมื่อไม่ต้องการใช้ค่า σ_X และ σ_Y ในการคำนวณหรือวิเคราะห์ข้อมูลอื่น อาจหาค่าความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยใช้สูตรวิธีลัดในการคำนวณดังนี้คือ

$$\rho_{XY} = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} \qquad \text{.......................(2.2)}$$

เมื่อ ρ_{XY} คือ สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน

$x = (X-\bar{X})$ คือ ค่าของ X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X

$y = (Y-\bar{Y})$ คือ ค่าของ Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y

$\Sigma xy = \Sigma(X-\bar{X})(Y-\bar{Y})$ คือ ผลบวกของผลคูณระหว่างค่า X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X กับค่า Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y ในลำดับที่เดียวกัน

$x^2 = (X-\bar{X})^2$ คือ ค่าของ X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X ยกกำลังสอง

$y^2 = (Y-\bar{Y})^2$ คือ ค่าของ Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y ยกกำลังสอง

$\Sigma x^2 = \Sigma(X-\bar{X})^2$ คือ ผลบวกของค่า X ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร X ยกกำลังสอง

$\Sigma y^2 = \Sigma(Y-\bar{Y})^2$ คือ ผลบวกของค่า Y ใด ๆ ที่เบี่ยงเบนจากค่าเฉลี่ยของตัวแปร Y ยกกำลังสอง

การแสดงวิธีการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน จากสูตรพื้นฐานและสูตรวิธีลัด ตามสมการที่ (2.1) และสมการที่ (2.2) ดังนี้

ตารางที่ 2.1 สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันระหว่างตัวแปร 2 ตัวที่ได้จากข้อมูลชุดเดียวกันแบบไม่ได้จัดกลุ่ม

ลำดับที่	ตัวแปร X	ตัวแปร Y	x $X-\bar{X}$	y $Y-\bar{Y}$	x^2 $(X-\bar{X})^2$	y^2 $(Y-\bar{Y})^2$	xy
1	13	11	+ 5.5	+ 3	30.25	9	+ 16.5
2	12	14	+ 4.5	+ 6	20.25	36	+ 27.5
3	10	11	+ 2.5	+ 3	6.25	9	+ 7.5
4	10	7	+ 2.5	- 1	6.25	1	- 2.5
5	8	9	+ 0.5	+ 1	0.25	1	+ 0.5
6	6	11	- 1.5	+ 3	2.25	9	- 4.5
7	6	3	- 1.5	- 5	2.25	25	+ 7.5
8	5	7	- 2.5	- 1	6.25	1	+ 2.5
9	3	6	- 4.5	- 2	20.25	4	+ 9.0
10	2	1	- 5.5	- 7	30.25	49	+ 38.5
Σ=รวม	75	80	0.0	0	124.50	144	102.0
เฉลี่ย	7.5	8.0					

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันจากสูตรพื้นฐาน ตามสมการที่ (2.1)

$$\sigma_X = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{\Sigma (X-\bar{X})^2}{N}} = \sqrt{\frac{124.50}{10}} = 3.528$$

$$\sigma_Y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{\Sigma (Y-\bar{Y})^2}{N}} = \sqrt{\frac{144}{10}} = 3.795$$

แทนค่าสมการที่ (2.1) จะได้ว่า

$$\rho_{XY} = \frac{\Sigma xy}{N\,\sigma_X\,\sigma_Y} = \frac{102.0}{(10)(3.528)(3.795)} = .76$$

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันด้วยวิธีลัด ซึ่งไม่ต้องคำนวณค่า σ_X และ σ_Y

แทนค่าสมการที่ (2.2) จะได้ค่า

$$\rho_{XY} = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{102.0}{\sqrt{(124.50)(144)}} = .76$$

3. การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันจากข้อมูลเดิมของตัวแปร 2 ตัว ซึ่งเป็นสูตรที่นิยมใช้กันโดยทั่วไป และง่ายแก่การคิดคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ พิจารณาวิธีการแปลงสูตรพื้นฐานเพื่อให้อยู่ในรูปข้อมูลเดิมและไม่ต้องใช้ค่าเบี่ยงเบนมาตรฐานของข้อมูล จะได้สูตรดังนี้

$$r_{XY} = \frac{n\ \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n\Sigma X^2 - (\Sigma X)^2\right]\left[n\ \Sigma Y^2 - (\Sigma Y)^2\right]}} \quad \text{.........(2.3)}$$

เมื่อ r_{XY} คือ สัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันของตัวอย่าง

X คือ ค่าของ X ใด ๆ จากตัวแปร X

Y คือ ค่าของ Y ใด ๆ จากตัวแปร Y

n คือ จำนวนคู่ลำดับของตัวแปร X กับ Y

ขั้นตอนการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันเมื่อ X กับ Y คือค่าเดิมของตัวแปร X กับ Y ดังสมการที่ 2.3

ขั้นตอนที่ 1 ยกกำลังสองของค่า X และ Y ทุกค่า แล้วกรอกไว้ในช่องสดมภ์ถัดไป

ขั้นตอนที่ 2 หาผลคูณของค่า X กับ Y ในทุกคู่ ได้ค่า XY แล้วกรอกไว้ในช่องสดมภ์ถัดไป

ขั้นตอนที่ 3 หาผลรวมของค่า X, Y, X^2 , Y^2 และ XY จะได้ค่า ΣX, ΣY , ΣX^2 , ΣY^2 และ ΣXY

ขั้นตอนที่ 4 แทนค่าต่าง ๆ จากการคำนวณทั้ง 3 ขั้นตอนลงในสูตรข้อมูลเดิมของสมการที่ 2.3 ผลการคำนวณจะได้ค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน

การแสดงวิธีการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันตามขั้นตอนทั้ง 4 ขั้น ดังตารางที่ 2.2 นี้

ตารางที่ 2.2 แสดงขั้นตอนทั้ง 3 ขั้นตอนในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันระหว่างตัวแปร 2 ตัวที่ได้จากข้อมูลเดิมชุดเดียวกัน

ลำดับที่	ตัวแปร X	ตัวแปร Y	X^2	Y^2	XY
1	13	7	169	49	91
2	12	11	144	121	132
3	10	3	100	9	30
4	8	7	64	49	56
5	7	2	49	4	14
6	6	12	36	144	72
7	6	6	36	36	36
8	4	2	16	4	8
9	3	9	9	81	27
10	1	6	1	36	6
Σ = รวม	70	65	624	533	472
สัญลักษณ์	ΣX	ΣY	ΣX^2	ΣY^2	ΣXY

แทนค่าสมการที่ 2.3 จะได้ว่า

$$r_{XY} = \frac{n\,\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n\Sigma X^2 - (\Sigma X)^2\right]\left[n\,\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$= \frac{10(472) - (70)(65)}{\sqrt{\left[10(624) - (70)^2\right]\left[10(533) - (65)^2\right]}}$$

$$= .14$$

2.1.3 ความหมายของสัมประสิทธิ์สหสัมพันธ์

จากการคำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันตามสูตรทั้ง 3 สมการข้างต้น จะได้ว่าสัมประสิทธิ์สหสัมพันธ์มีขอบเขตตั้งแต่ 0 ถึง $\pm$ 1 ($-1 < r < 1$) ในการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ว่าอยู่ในระดับใด ให้พิจารณาเฉพาะตัวเลขอย่างเดียว เครื่องหมายหน้าตัวเลขจะแสดงลักษณะความสัมพันธ์ระหว่างข้อมูลเท่านั้น คือ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก แสดงว่าข้อมูลทั้งสองมีลักษณะตามกัน ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบแสดงว่าข้อมูลทั้งสองมีลักษณะตรงกันข้าม การตัดสินค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลคู่ใด ๆ ว่าอยู่ในระดับสูง กลาง หรือ ต่ำ ต้องพิจารณาถึงสภาพแห่งลักษณะของข้อมูลทั้งสองที่สัมพันธ์กันด้วย อาจแบ่งระดับของค่าสัมประสิทธิ์สหสัมพันธ์ได้อย่างกว้าง ๆ ดังนี้

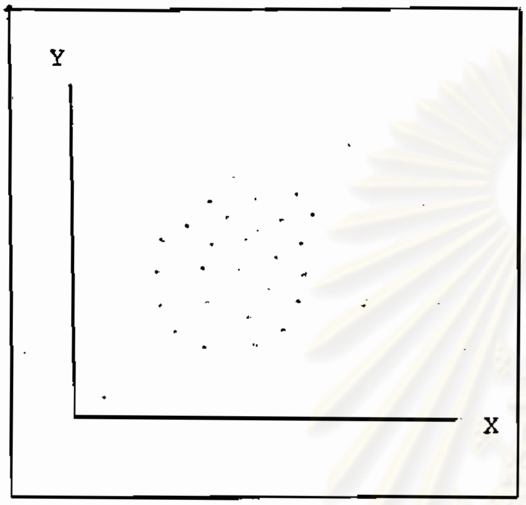
1. เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1.00 คืออยู่ระหว่าง .70 ถึง .90 โดยทั่วไปถือว่าตัวแปรทั้งสองสัมพันธ์กันในระดับสูง ถ้าสูงกว่า .90 ถือว่ามีความสัมพันธ์กันอยู่ในระดับสูงมาก

2. เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ .50 คืออยู่ระหว่าง .30 ถึง .70 โดยทั่วไปถือว่าตัวแปรทั้งสองมีความสัมพันธ์กันในระดับกลาง

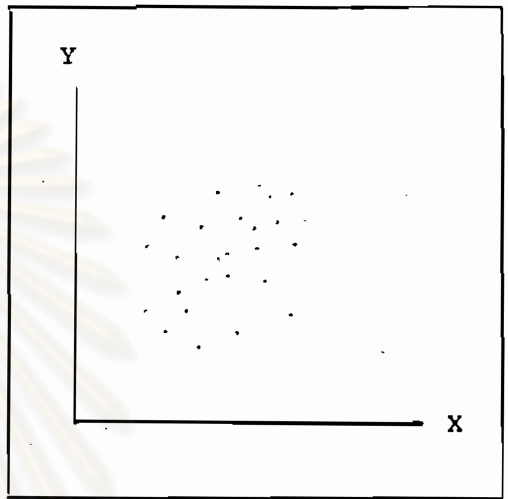
3. เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 0 คือประมาณ .30 และต่ำกว่า โดยทั่วไปถือว่าตัวแปรทั้งสองมีความสัมพันธ์กันในระดับต่ำ

ในกรณีที่ข้อมูลทั้ง 2 ตัวแปรมีความสัมพันธ์กัน ทั้งในทางตรงกันหรือในทางตรงกันข้าม จะเรียกว่าข้อมูลทั้ง 2 ตัวแปรมีความสัมพันธ์กันอย่างสมบูรณ์ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้เข้าใกล้หรือเท่ากับ + 1 หรือ - 1 แต่ในทางปฏิบัติจะหาไม่ได้เลยสำหรับค่า $r = \pm 1$

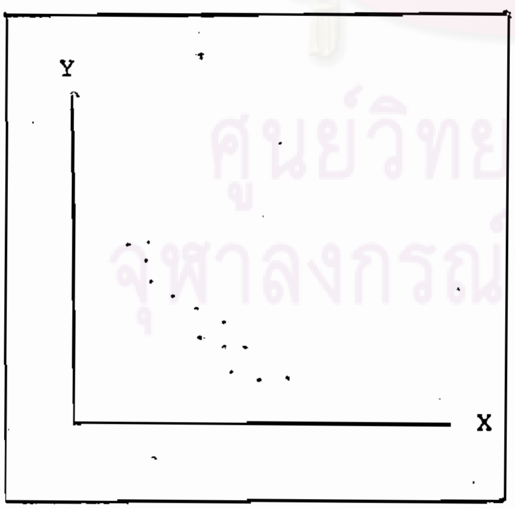
<u>รูปที่ 2.1</u> แสดงค่าสัมประสิทธิ์สหสัมพันธ์ในระดับความสัมพันธ์ต่าง ๆ



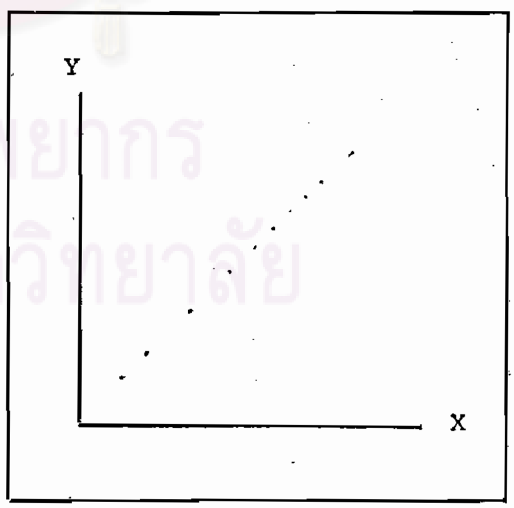
ก. เมื่อ r = 0



ข. เมื่อ r = 0.45



ค. เมื่อ r = - 0.85



ง. เมื่อ r = 1

2.1.4 การทดสอบสัมประสิทธิ์สหสัมพันธ์

การทดสอบสัมประสิทธิ์สหสัมพันธ์เป็นการทดสอบเพื่อต้องการทราบว่า ตัวแปร 2 ตัวเป็นอิสระต่อกันหรือไม่ เมื่อคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันระหว่างตัวแปร 2 ตัวแล้ว อาจทำการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยใช้การทดสอบนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์ ซึ่งอาจดำเนินการเพื่อการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวตามขั้นตอนดังนี้ คือ

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวดังนี้ คือ

สมมติฐานว่าง (Null Hypothesis) Ho : $\rho = 0$

(ตัวแปร 2 ตัวเป็นอิสระกัน)

สมมติฐานแย้ง (Alternative Hypothesis) Ha : $\rho \neq 0$

(ตัวแปร 2 ตัว ไม่เป็นอิสระต่อกันหรือตัวแปร 2 ตัวมีความสัมพันธ์กัน)

ขั้นตอนที่ 2 ตัวสถิติที่ใช้เพื่อการทดสอบ เมื่อต้องการทดสอบสมมติฐานที่ว่าสัมประสิทธิ์สหสัมพันธ์เท่ากับศูนย์ ($\rho = 0$) จะใช้ตัวสถิติ t ดังนั้นจะได้ว่า

$$t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{(n-2)} \qquad \text{........(2.4)}$$

เมื่อ t มีการแจกแจงแบบทีหรือการแจกแจงแบบ สติวเดนท์-ที (t-Distribution or Student's t-Distribution) ที่มีองศาความเป็นอิสระ (Degree of Freedom - d.f) เท่ากับ $n - 2$

r คือ ค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สันระหว่างตัวแปร X กับ Y และ n คือ จำนวนคู่ลำดับของตัวแปร X กับ Y หรือจำนวนตัวอย่างที่ใช้ในการวิเคราะห์

ขั้นตอนที่ 3 กำหนดระดับนัยสำคัญ (Level of Significance) ของการทดสอบกำหนดให้ระดับนัยสำคัญของการทดสอบเท่ากับ α

ขั้นตอนที่ 4 สรุปผลการทดสอบสมมติฐาน

เมื่อคำนวณค่า t จากสูตร $t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{(n-2)}$ จากสมการที่ (2.4)ได้แล้วนำค่า t ที่คำนวณได้มาเทียบกับค่า $t_{\alpha/2,(n-2)}$ ที่มีระดับนัยสำคัญเท่ากับ $\frac{\alpha}{2}$ และมีองศาความเป็นอิสระเท่ากับ n - 2 ถ้า $t_{\alpha/2,(n-2)}$ จากตารางการแจกแจงแบบ t จะยอมรับสมมติฐานว่าง Ho : ที่ว่า $\rho = 0$ นั่นคือ ตัวแปร 2 ตัวเป็นอิสระต่อกัน ถ้าหาก $t < t_{\alpha/2,(n-2)}$ จากตารางการแจกแจงแบบ t จะปฏิเสธสมมติฐานว่าง Ho และยอมรับสมมติฐานแย้ง Ha ที่ว่า $\rho = 0$ นั่นคือตัวแปร 2 ตัวไม่เป็นอิสระต่อกันหรือตัวแปรทั้ง 2 ตัวมีความสัมพันธ์กัน

<u>ตัวอย่าง</u> แสดงการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยการทดสอบสมมติฐานทางสถิติของสัมประสิทธิ์สหสัมพันธ์

เมื่อคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบผลคูณของเพียร์สัน (r) ได้เท่ากับ .14 จากตารางที่ 2.2 ดำเนินการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว ดังขั้นตอนข้างต้น ดังนี้ คือ

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว ดังนี้คือ

Ho : $\rho = 0$ (ตัวแปร 2 ตัวเป็นอิสระต่อกัน)

Ha : $\rho \neq 0$ (ตัวแปร 2 ตัวไม่เป็นอิสระต่อกัน)

ขั้นตอนที่ 2 ตัวสถิติที่ใช้เพื่อการทดสอบ จากสมการที่ (2.4)

ใช้ตัวสถิติ
$$t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{(n-2)}$$

$$= .14\sqrt{\frac{10-2}{1-(.14)^2}}$$

$$= .3999$$

ขั้นตอนที่ 3 กำหนดระดับนัยสำคัญของการทดสอบ

กำหนดให้ระดับนัยสำคัญในการทดสอบครั้งนี้เท่ากับ .05 เปิดตารางการแจกแจงแบบ t ที่ระดับนัยสำคัญ $\frac{\alpha}{2}$ คือ $\frac{.05}{2} = .025$

และองศาความเป็นอิสระ n-2 $= 10 - 2 = 8$

ได้ค่า $t_{.025,(8)} = 2.306$

ขั้นตอนที่ 4 สรุปผลการทดสอบสมมติฐานเปรียบเทียบค่า t จากการคำนวณกับค่า $t_{\alpha/2,(n-2)}$ จากตารางการแจกแจงแบบ t จะได้ว่า $t < t_{.025,(8)}$ ดังนั้นยอมรับสมมติฐานว่าง Ho : ที่ว่า $\rho = 0$ นั่นคือตัวแปร 2 ตัวเป็นอิสระต่อกัน

2.1.5 ความถดถอยเชิงเส้น

การวิเคราะห์ข้อมูลโดยการหาความถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เพื่อใช้พิจารณาตัวแบบที่เป็นไปได้ของความสัมพันธ์ระหว่างตัวแปร 2 ตัว และเป็นประโยชน์ในการพยากรณ์ (Predict) หรือคาดประมาณ (Estimate) ค่า ๆ หนึ่งซึ่งสัมพันธ์กับค่าที่กำหนดให้อีกค่าหนึ่ง ดังนั้นจึงควรทราบข้อสมมติเบื้องต้นของความถดถอยเชิงเส้น เพื่อใช้ประกอบในการพิจารณาผลลัพธ์ที่ได้จากการหาความถดถอยเชิงเส้นอย่างง่าย

1. ข้อสมมติเบื้องต้นของความถดถอยเชิงเส้นอย่างง่าย

เมื่อต้องการหาตัวแบบ (Model) แสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยให้ X และ Y เป็นตัวแปรในตัวแบบความถดถอยเชิงเส้น ซึ่งจะเรียก X ว่าตัวแปรอิสระ (Independent Variable) หรือตัวแปรอธิบาย (Explanatory Variable) และเรียก Y ว่าตัวแปรตาม (Dependent Variable) หรือตัวแปรตอบสนอง (Response Variable) เพราะจาก X ในแต่ละค่าจะได้ค่า Y ที่สัมพันธ์กันในแต่ละค่าของ X ด้วย ซึ่งจะต้องมีข้อสมมติเบื้องต้นดังนี้

ก. ค่าของตัวแปรอิสระ X จะต้องคงที่ กล่าวคือ ผู้ทำการทดลองจะเลือกค่า X ไว้ล่วงหน้า และค่า X เหล่านี้จะไม่แปรเปลี่ยนไปตามลักษณะการทดลอง อย่างไรก็ตาม การวิเคราะห์ความถดถอยยังสามารถใช้กับข้อมูลซึ่ง X เป็นตัวแปรสุ่มได้เช่นกัน

ข. ตัวแปร X ต้องได้รับการวัดอย่างถูกต้อง เนื่องจากการเก็บข้อมูลเพื่อการวิเคราะห์ใด ๆ มักจะไม่สมบูรณ์หรือถูกต้องครบถ้วนทั้งหมด ดังนั้น ในกรณีนี้จึงไม่สนใจความผิดพลาดที่เกิดจากการวัดค่า X

ค. แต่ละค่า Y จะเป็นอิสระต่อกัน นั่นคือ เมื่อสุ่มตัวอย่างข้อมูล สมมติว่าค่า Y ซึ่งเลือกขึ้น ณ ค่า X ค่าใดค่าหนึ่งจะไม่ขึ้นกับค่า X ที่เลือกจากค่า X ตัวอื่น ๆ

2. ตัวแบบความถดถอยเชิงเส้น

การดำเนินงานวิจัยโดยทั่วไป จะเริ่มต้นจากการเก็บข้อมูลตัวอย่างจากประชากร แล้วทำการวิเคราะห์ข้อมูลเพื่อนำผลลัพธ์ไปอ้างอิงถึงประชากรทั้งหมดต่อไป ดังนั้น จึงจำเป็นที่จะต้องเข้าใจลักษณะของประชากรที่กำลังศึกษาก่อน ในการวิเคราะห์ข้อมูลด้วยความถดถอยเชิงเส้น ก็ต้องมั่นใจว่าตัวแบบความถดถอย (The Regression Model) ที่เลือกใช้นี้เหมาะสมกับข้อมูลที่มีอยู่ มิฉะนั้นผลลัพธ์ที่ได้จะไม่ถูกต้องและนำไปใช้ประโยชน์มิได้

จากข้อสมมติฐาน 3 ข้อข้างต้น สามารถสร้างตัวแบบความถดถอยเชิงเส้นได้ดังนี้ คือ

$$Y = \alpha + \beta X + E \quad \text{..............................} (2.5)$$

เมื่อ Y คือ ตัวแปรตาม ซึ่งเป็นค่าเฉพาะค่าใดค่าหนึ่งของ X ตัวหนึ่ง

X คือ ตัวแปรอิสระ ซึ่งเป็นค่าที่เลือกไว้ล่วงหน้า

α คือ จุดตัดบนแกน Y

β คือ สัมประสิทธิ์ความถดถอย

E คือ ค่าความคลาดเคลื่อน

3. การหาความถดถอยเชิงเส้น

ในการวิเคราะห์ความถดถอยเชิงเส้นนั้น สิ่งที่ผู้วิจัยสนใจคือ สมการความถดถอยของประชากร (The Population Regression Equation) ซึ่งเป็นสมการแสดงความสัมพันธ์ที่แท้จริงระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X เพื่อให้บรรลุถึงสิ่งที่ต้องการ

นักวิจัยจะเลือกตัวอย่างจากประชากรที่ศึกษา แล้วคำนวณสมการความถดถอยจากตัวอย่าง (The Sample Regression Equation) จากสมการความถดถอยจากตัวอย่างนี้ สามารถใช้อ้างอิงผลไปสู่ประชากรทั้งหมดได้

วิธีการหนึ่งที่นิยมใช้กันทั่วไปเพื่อหาความถดถอยเชิงเส้น ซึ่งเป็นตัวแทนที่ดีของจุด X และ Y คือวิธีกำลังสองน้อยที่สุด (Least Square Method) เส้นตรงที่ได้จากวิธีการนี้เรียกว่า เส้นตรงกำลังสองน้อยที่สุด (Least Square Line) เหตุผลที่ถือว่าเส้นตรงที่ได้จากวิธีกำลังสองน้อยที่สุดเป็นเส้นตรงที่ดีที่สุด เพราะว่าผลบวกกำลังสองของความแตกต่างระหว่างค่าที่สังเกตมา (Y) กับค่าประมาณสำหรับ Y บนเส้นตรงกำลังสองน้อยที่สุด ณ จุด Y เดียวกันมีค่าน้อยกว่าผลบวกกำลังสองของความแตกต่างระหว่างค่าที่สังเกตมากับค่าประมาณสำหรับ Y บนเส้นตรงเส้นอื่น ๆ ณ จุด X เดียวกัน

การคำนวณเส้นตรงกำลังสองน้อยที่สุด โดยใช้ตัวแบบสมการเส้นตรงโดยทั่วไปดังนี้ คือ

$$\hat{Y} = a + bX \quad \text{..............................} (2.6)$$

เมื่อ $\hat{Y}$ คือ ค่าตามแกนตั้ง

X คือ ค่าตามแกนนอน

a คือ จุดซึ่งตัดแกน X เป็นค่าคงที่

b คือ ปริมาณที่เส้นตรงจะเปลี่ยนไปเมื่อ X เปลี่ยนไปหนึ่งหน่วย หรือเป็นค่าความชันของเส้นตรง

ในการลากเส้นตรงใด ๆ จะต้องทราบค่า a และ b ก่อน จากนั้นก็แทนค่าต่าง ๆ ของ X ในสมการจะได้ค่า Y สำหรับแต่ละค่าของ X เมื่อได้คู่ลำดับ (X,Y) สองคู่ใด ๆ สามารถลากเส้นตรงเชื่อมจุดสองจุดได้ทันที ดังนั้นจึงสามารถหาค่า a และ b ได้จากการแก้สมการ 2 สมการที่เรียกว่าสมการปกติ (Normal Equation) เพื่อทำให้อยู่ในรูปสูตรสำเร็จของการหาค่า a และ b คือ

ก. เมื่อ a และ b อยู่ในรูปของสมการปกติ ซึ่งประกอบด้วยสมการ 2 สมการคือ

$$\Sigma Y_i \quad = \quad na + b\Sigma Y_i \qquad ....(2.7)$$

$$\Sigma X_i Y_i \quad = \quad a\Sigma X_i + b\Sigma X_i^2 \qquad .... (2.8)$$

ข. เมื่อ a และ b อยู่ในรูปสูตรสำเร็จ จากการแก้สมการ (2.7) และ(2.8) จะได้ว่า

$$b \quad = \quad \frac{n\ \Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} \qquad ....(2.9)$$

$$a \quad = \quad \frac{\Sigma Y - b\Sigma X}{n} \qquad ....(2.10)$$

หรือ $$a \quad = \quad \bar{Y} - b\bar{X} \qquad .... (2.11)$$

เมื่อ a คือ จุดตัดแกน Y

b คือ ค่าความชันของเส้นตรง

X คือ ค่าสังเกตจากตัวแปร X

Y คือ ค่าสังเกตจากตัวแปร Y

$\bar{X}$ คือ ค่าเฉลี่ยของตัวแปร X

$\bar{Y}$ คือ ค่าเฉลี่ยของตัวแปร Y

n คือ จำนวนคู่ลำดับของตัวแปร X กับ Y

แสดงการคำนวณค่า a และ b โดยใช้ข้อมูลจากตารางที่ 2.1 ซึ่งมี $\Sigma X = 75$ $\Sigma Y = 80$ $\Sigma XY = 702$ $\Sigma X^2 = 687$

$\bar{X} = 7.5$ $\bar{Y} = 8.0$ $n = 10$

แทนค่าสมการที่ (2.9) จะได้ว่า

$$b = \frac{n\ \Sigma XY - (\Sigma X)(\Sigma Y)}{n\ \Sigma X^2 - (\Sigma X)^2} = \frac{10(702) - (75)(80)}{10(687) - (75)^2}$$

$$= 0.819$$

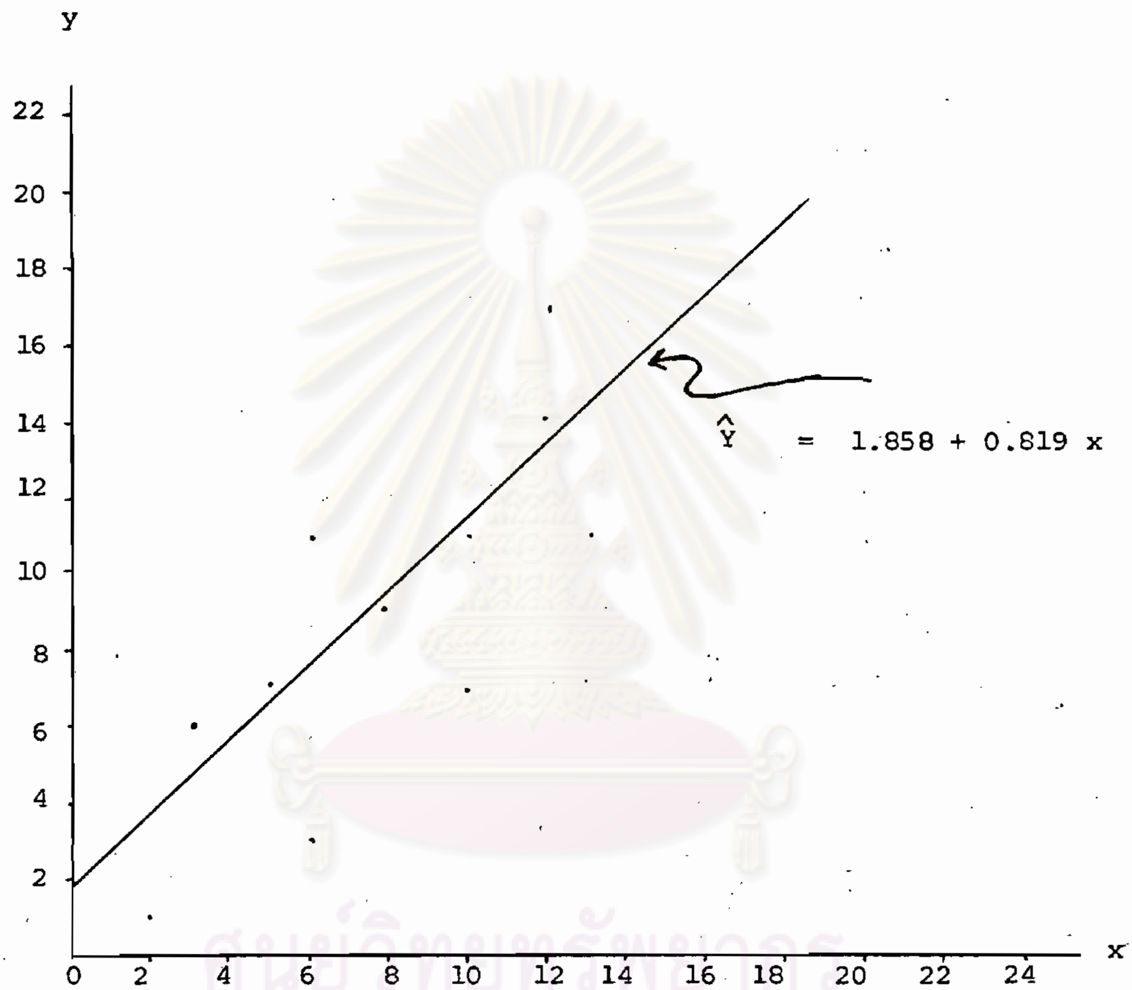
แทนค่าสมการที่ (2.11) จะได้ว่า

$$a = \bar{Y} - b\bar{X} = 8.0 - (0.819)(7.5) = 1.858$$

สมการความถดถอยเชิงเส้นของข้อมูลชุดนี้ ซึ่งคำนวณได้โดยใช้วิธีที่กล่าวแล้วข้างต้นคือ

$$\hat{Y} = 1.858 + 0.819\ X$$

รูปที่ 2.2 แสดงข้อมูลของตัวแปร X และ Y ตามตารางที่ 2.1 และความถดถอยเชิงเส้นที่คำนวณได้



4. สัมประสิทธิ์ของการตัดสินใจ

เมื่อคำนวณได้สมการความถดถอยเชิงเส้นจากข้อมูลตัวอย่างแล้ว ต่อไปคือการประเมินสมการความถดถอยเชิงเส้น เพื่อต้องการทราบว่าสมการที่ได้สามารถใช้อธิบายถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัวที่กำลังพิจารณาอยู่หรือไม่ จึงหาตัวที่ใช้วัดซึ่งเรียกว่า สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination)

พิจารณาที่จุดค่าสังเกตค่าใดค่าหนึ่ง (Y_i) และวัดระยะตามแนวดิ่ง เรียกระยะนี้ว่าค่าเบี่ยงเบนทั้งหมด (Total Deviation) เขียนแทนด้วย ($Y_i - \bar{Y}$) จากเส้นค่าเบี่ยงเบนทั้งหมด ($Y_i - \bar{Y}$) สามารถแบ่งออกเป็น 2 ส่วน คือ ความเบี่ยงเบนที่อธิบายได้ (Explained Deviation) เขียนแทนด้วย ($\hat{Y} - \bar{Y}$) เพราะเป็นส่วนที่บอกว่าค่าเบี่ยงเบนทั้งหมดจะเปลี่ยนแปลงเท่าใด เมื่อใช้เส้นตรงถดถอยแทนจุดที่สังเกตมา และอีกส่วนหนึ่งคือค่าวัดความแปรปรวนของ Y รอบเส้นตรงถดถอย (Residual Sum of Squares) เขียนแทนด้วย ($Y_i - \bar{Y}$) ค่าความแปรปรวนของ Y นี้จะลดลงเมื่อใช้เส้นถดถอยที่คำนวณจากวิธีกำลังสองน้อยที่สุด กล่าวคือ

$$SS\,(Total) = SS(Explained) + SS(Unexplained)$$

$$SS(Total) = \Sigma(Y_i - \bar{Y})^2$$

$$= \Sigma Y_i^2 - \frac{(\Sigma Y_i)^2}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.12)$$

$$SS(Explained) = \Sigma\,(\hat{Y} - \bar{Y})^2$$

$$= b^2\,\Sigma(X_i - \bar{X})^2$$

$$= b^2\left[\Sigma X_i^2 - (\Sigma X_i)^2/\,n\right] \ldots\ldots\ldots\ldots\ldots\ldots (2.13)$$

การที่สมการถดถอยจะใช้ได้ดีในการแสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัว SS(Explained) จะต้องรวมเอาส่วนของ SS(Total) ไว้มากที่สุด นั่นคือ จะต้องพิจารณาอัตราส่วนระหว่าง SS(Explained) กับ SS(Total) อัตราส่วนนี้จะบอกให้ทราบถึงปริมาณของ SS(Explained) เมื่อเทียบกับ SS(Total) ซึ่งเป็นประโยชน์ในการประเมินสมการถดถอย และเรียกอัตราส่วนที่คำนวณจากข้อมูลนี้ว่า "สัมประสิทธิ์ของการตัดสินใจ" เขียนแทนด้วยสัญลักษณ์ r^2 จะได้ว่า

$$r^2 = \frac{SS(Explained)}{SS(Total)} \qquad \text{.....................(2.18)}$$

$$r^2 = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2} \qquad \text{.....................(2.19)}$$

$$r^2 = \frac{b^2\left[\Sigma X_i^2 - (\Sigma X_i)^2/n\right]}{\Sigma Y_i^2 - (\Sigma Y_i)^2/n} \qquad \text{.....................(2.20)}$$

เมื่อ r^2 คือ สัมประสิทธิ์ของการตัดสินใจ

X คือ ค่าสังเกตจากตัวแปร X

Y คือ ค่าสังเกตจากตัวแปร Y

n คือ จำนวนคู่ลำดับของตัวแปร X และ Y

b คือ ค่าความชันของเส้นตรง

เมื่อ r^2 ใช้วัดความเหมาะสมของสมการความถดถอยที่จะใช้อธิบายข้อมูล ดังนั้นถ้าจุดของข้อมูลจริงยิ่งมีความใกล้ชิดหรืออยู่บนเส้นถดถอยมากเพียงใด สมการความถดถอยที่ได้ถือว่าเป็นสมการที่ดีเพียงนั้น สิ่งที่ตามมาคือ r^2 จะมีค่ามากตามไปด้วย เช่น r^2 = .986 หมายความว่า 98.6% ของความแปรปรวนทั้งหมดถูกอธิบายโดยเส้นถดถอย เมื่อ r^2 = .50 หมายความว่าน้อยกว่า 50% ของความแปรปรวนทั้งหมดถูกอธิบายโดยเส้นถดถอย r^2 จะมีค่าสูงสุดเป็น 1 หมายความว่าจุดข้อมูลที่ได้จากค่าสังเกตจะอยู่บนเส้นตรงถดถอยที่คำนวณได้ และ r^2 จะมีค่าต่ำสุดเป็น 0

5. การพยากรณ์ค่าจากสมการความถดถอยเชิงเส้น

การพยากรณ์หมายถึงการคำนวณต่อไปข้างหน้า เป็นการอ้างอิงโดยใช้ข้อมูลในอดีต

ในการพยากรณ์เชิงสถิติ อาจจำแนกเป็น 3 ประเภทใหญ่ ๆ ดังนี้ คือ

ก. การพยากรณ์ ณ ช่วงเวลาใดเวลาหนึ่ง ได้แก่การสำรวจความคิดเห็นสหสัมพันธ์ (Correlation) และความถดถอย (Regression)

ข. การพยากรณ์เชิงอนุกรมเวลา ได้แก่การพยากรณ์เกี่ยวกับความกดดันอากาศ ความก้าวหน้าทางเศรษฐกิจ ดัชนี เป็นต้น

ค. การพยากรณ์โดยใช้เศรษฐมิติ

โดยทั่วไปแล้ว การพยากรณ์สามารถจำแนกออกเป็น 6 ขั้นตอนดังนี้คือ

ขั้นตอนที่ 1 การเตรียมข้อมูล ก่อนอื่นต้องรวบรวมข้อมูลและเมื่อได้ข้อมูลมาแล้ว ปัญหาของการพยากรณ์มีอยู่ว่าจะคำนวณหาตัวแบบของข้อมูลชุดนั้นอย่างไร ข้อมูลอาจอยู่ในตัวแบบต่าง ๆ กัน เช่น มีค่าคงที่ตลอดเวลา เป็นอนุกรมเลขคณิต อนุกรมเลขคณิตกำลังสอง หรือไม่แน่นอน

ขั้นตอนที่ 2 ตัวแบบ (Model) คือชุดของสมการแสดงความสัมพันธ์ระหว่างตัวแปรต่าง ๆ แต่ละสมการประกอบด้วยตัวแปรและสัมประสิทธิ์ การสร้างสมการดังกล่าวอาศัยข้อเท็จจริงประกอบกับทฤษฎี แล้วใช้ข้อมูลและวิธีการทางสถิติประเมินความถูกต้องแน่นอน

ขั้นตอนที่ 3 การปรับปรุง เป็นการแก้ไขปรับปรุงการกะประมาณค่าสัมประสิทธิ์ในตัวแบบ ในการนี้จะต้องยึดความถูกต้องและประสิทธิภาพของการคำนวณ

ขั้นตอนที่ 4 การพยากรณ์ เมื่อประมาณค่าของสัมประสิทธิ์ในตัวแบบได้แล้ว ก็สามารถคำนวณค่าตัวแปรต่าง ๆ ได้ บางครั้งการพยากรณ์อาจใช้ฟังก์ชันหนึ่งสำหรับข้อมูลในอดีตและอีกฟังก์ชันหนึ่งสำหรับการพยากรณ์ในอนาคต

ขั้นตอนที่ 5 การวัดความผิดพลาด เนื่องจากไม่มีการพยากรณ์หรือคาดคะเนใด ๆ ที่ให้ผลถูกต้องสมบูรณ์ จึงต้องตรวจสอบดูว่าค่าที่ได้จากการพยากรณ์หรือการคาดคะเนแตกต่างจากความเป็นจริงมากน้อยเพียงไร

ขั้นตอนที่ 6 การตัดสินใจภายใต้ความไม่แน่นอน ซึ่งเป็นขั้นตอนสุดท้ายในการพยากรณ์ค่าจากสมการความถดถอยเชิงเส้น

2.2 การทดสอบไคสแควร์

การทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยการทดสอบไคสแควร์ จะต้องนำข้อมูลมาจัดเสนอในรูปของตารางการณ์จร แล้วคำนวณค่าไคสแควร์ เพื่อนำค่าไคสแควร์ที่คำนวณได้มาเปรียบเทียบกับค่าไคสแควร์จากตารางการแจกแจงไคสแควร์ ณ ระดับนัยสำคัญที่กำหนดให้ ซึ่งควรทราบรายละเอียดเกี่ยวกับการทดสอบไคสแควร์ ดังนี้

2.2.1 การแจกแจงไคสแควร์

2.2.2 การใช้การทดสอบไคสแควร์

2.2.1 การแจกแจงไคสแควร์

การแจกแจงไคสแควร์ (Chi-Square Distribution) หรือ Central Chi-Square Distribution ซึ่งมีฟังก์ชันการแจกแจงความหนาแน่น ดังนี้

$$f(X) = \frac{1}{(\frac{k}{2} - 1)!\; 2^{\frac{k}{2}}} X^{\frac{k}{2} - 1} e^{\frac{-X}{2}} \quad \text{เมื่อ} \quad X > 0$$

เมื่อ e เป็นจำนวนอตักยะ (Irrational) มีค่าประมาณ 2.71828 และ k คือองศาความเป็นอิสระ

หรืออาจเขียนฟังก์ชันการแจกแจงความหนาแน่นไคสแควร์ได้ในรูปของ

$$f(X) = \frac{1}{2^{k/2}\,\Gamma(\frac{k}{2})} X^{\frac{k}{2} - 1} e^{-\frac{X}{2}} \quad \text{เมื่อ} \quad X > 0$$

ซึ่งการแจกแจงความหนาแน่นไคสแควร์ เป็นการแจกแจงความหนาแน่นในกรณีพิเศษ (Special Case) ของฟังก์ชันการแจกแจงความหนาแน่นแกมมา เมื่อให้ $a = \frac{1}{2}$ และ $n = \frac{k}{2}$ โดย k เป็นจำนวนเต็มบวกใด ๆ ซึ่งฟังก์ชันการแจกแจงความหนาแน่นแกมมา เป็นดังนี้

$$f(X) = \frac{a}{\Gamma(n)} (aX)^{n-1} e^{-aX} \quad \text{เมื่อ } X > 0 \text{ และ } n > 1$$

ตัวแปร X สามารถเขียนแทนด้วยอักษรกรีก Chi (χ) ได้ ดังนั้น จึงเรียกการแจกแจงจาก f(X) ว่าการแจกแจงไคส์แควร์ ซึ่งมีค่าเฉลี่ยและความแปรปรวนคือ k และ 2k ตามลำดับ ฐานนิยมมีค่าเท่ากับ (k - 2) และจะเป็น 0 เมื่อ k = 1

1. คุณสมบัติที่สำคัญของการแจกแจงไคส์แควร์

การแจกแจงไคส์แควร์มีคุณสมบัติสำคัญ ที่ควรทราบดังนี้

ก. ในกรณีที่ X มีการแจกแจงปกติมาตรฐาน (Standard Normal Distribution) โดยมีค่า $Y = X^2$ จะได้ว่า Y มีการแจกแจงแบบไคส์แควร์ (χ^2) ซึ่งมีองศาความเป็นอิสระเท่ากับ 1 หรือ $\chi^2_{(1)}$ - Distribution

ข. ถ้า X_i เป็นตัวแปรสุ่มอิสระที่มีการแจกแจงปกติมาตรฐาน เมื่อ $i = 1, 2, 3, \ldots, k$ และ Y คือผลบวกกำลังสองของ X_i หรือ $Y = X_1^2 + X_2^2 + X_3^2 + \ldots\ldots + X_k^2$ แล้ว Y จะมีการแจกแจงแบบไคส์แควร์ (χ^2) ซึ่งมีองศาความเป็นอิสระเท่ากับ k หรือ $\chi^2_{(k)}$ - Distribution

ค. ถ้าการแจกแจงความน่าจะเป็นของข้อมูลชุดใด มีค่าเฉลี่ย (mean) เท่ากับ 0 ความแปรปรวนเท่ากับ 1 และเป็นการแจกแจงแบบ Z แล้ว จะได้ Z^2 คือ Chi-Square Distribution ที่มีองศาความเป็นอิสระเท่ากับ 1 หรือ $\chi^2_{(1)}$ - Distribution

จากคุณสมบัติของการแจกแจงไคส์แควร์ทั้ง 3 ข้อข้างต้น สามารถอธิบายได้ว่า การแจกแจงไคส์แควร์สามารถหามาจากการแจกแจงแบบปกติได้ สมมติให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (Normal Distribution) มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 เมื่อสุ่มตัวอย่างขนาด k = 1 หลาย ๆ ตัวอย่าง โดยที่แต่ละกลุ่มตัวอย่างเป็นอิสระต่อกัน แต่ละค่าที่เลือกมา สามารถแปลงเป็นตัวแปรสุ่มปกติมาตรฐาน Z ได้ โดยใช้สูตร

$$Z = \frac{Y_i - \mu}{\sigma}$$

นำแต่ละค่าของ Z ยกกำลังสองได้ Z^2 จะได้ว่าการแจกแจงของกลุ่มตัวอย่างของ Z^2 เป็นการแจกแจงไคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ 1 นั่นคือ

$$Z^2 = \left(\frac{Y - \mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$$

ต่อไปสมมติว่าสุ่มตัวอย่างขนาด $k = 2$ จากประชากรที่มีการแจกแจงปกติที่มีค่าเป็น Y โดยที่ตัวอย่างแต่ละค่าต่างเป็นอิสระต่อกันสามารถแปลงค่า Y ให้เป็นตัวแปรสุ่มปกติมาตรฐาน Z ได้ จากนั้นยกกำลังสอง นำค่า Z^2 สำหรับแต่ละตัวอย่างมารวมกัน ซึ่งเขียนผลรวมดังกล่าวได้ดังนี้

$$Z_1^2 + Z_2^2 = \left(\frac{Y_1 - \mu}{\sigma}\right)^2 + \left(\frac{Y_2 - \mu}{\sigma}\right)^2 \sim \chi^2_{(2)}$$

ซึ่ง $\chi^2_{(2)}$ จะมีการแจกแจงแบบไคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ 2 ซึ่งเท่ากับจำนวนเทอมของตัวแปรกำลังสองที่นำมาบวกกัน

โดยวิธีการทำนองเดียวกัน ถ้าสุ่มตัวอย่างขนาด k จะได้ว่าผลรวมของ Z^2 จะมีการแจกแจงแบบไคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ k หรือ

$$Z_1^2 + Z_2^2 + Z_3^2 + \ldots\ldots + Z_k^2 \sim \chi^2_{(k)}$$

2. การหาค่าไคสแควร์จากตารางการแจกแจงไคสแควร์

การหาค่าไคสแควร์จากตารางการแจกแจงไคสแควร์ ถ้าองศาความเป็นอิสระน้อยกว่าหรือเท่ากับ 30 ($d.f. \leq 30$) สามารถหาค่าได้จากตารางการแจกแจงไคสแควร์ท้ายเล่มของหนังสือสถิติโดยทั่วไป ซึ่งการแจกแจงของค่าไคสแควร์ (χ^2) จะสมมาตรยิ่งขึ้นเมื่อองศาความเป็นอิสระเพิ่มขึ้น การแจกแจงของ $\sqrt{\chi^2}$ หรือ χ จะสมมาตรกว่าการแจกแจงไคสแควร์ (χ^2) เมื่อองศาความเป็นอิสระมากกว่า 30 ($d.f. > 30$) การแจกแจงของ χ จะเป็นปกติ โดยประมาณและมีค่าเฉลี่ย (Mean) เป็น $\sqrt{d.f. - 0.5}$ และความเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็น $\sqrt{0.5}$ ฉะนั้นจึงอาจใช้ตารางพื้นที่ใต้โค้งปกติช่วยในการ

ทดสอบความมีนัยสำคัญของไคสแควร์ (χ^2) เมื่อมีองศาความเป็นอิสระมากกว่า 30 ได้ โดยการแปลงค่า χ ให้อยู่ในรูปของ Z ดังนี้

$$Z = \frac{\chi - \mu_\chi}{\sigma_\chi}$$

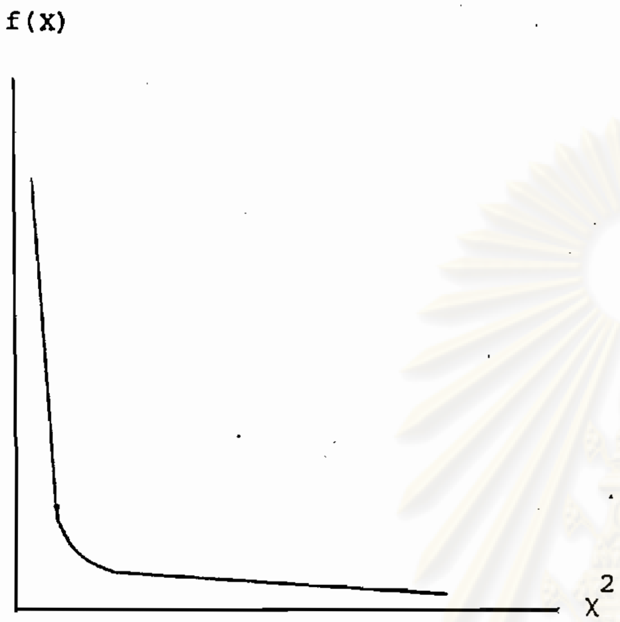
$$Z = \frac{\chi - \sqrt{d.f. - 0.5}}{\sqrt{0.5}}$$

$$Z = \chi\sqrt{2} - \sqrt{2\ d.f. - 1}$$

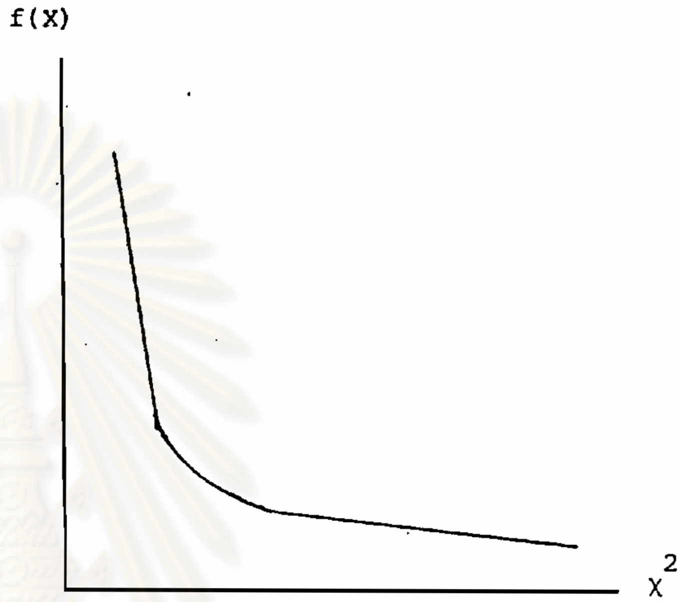
$$Z = \sqrt{2\chi^2} - \sqrt{2\ d.f. - 1}$$

ดังนั้นจึงอาจใช้ตารางการแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution = Z) แทนการแจกแจงไคสแควร์ได้เมื่อมีองศาความเป็นอิสระมากกว่า 30

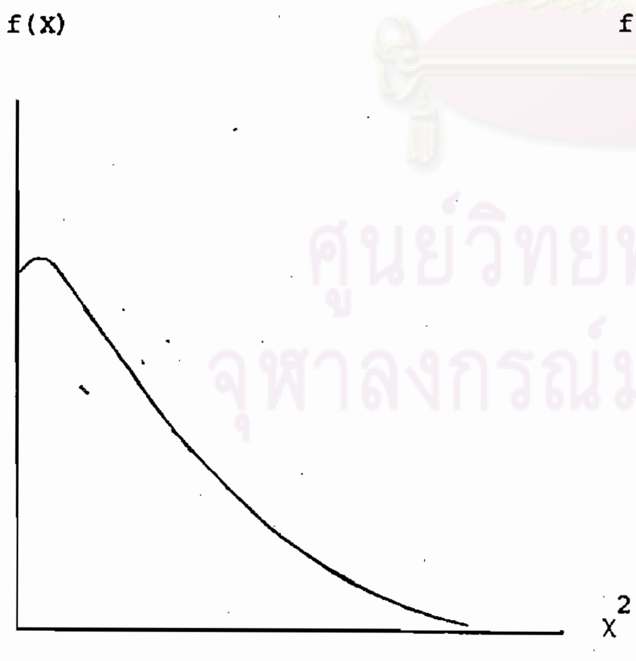
รูปที่ 2.3 แสดงการแจกแจงไคสแควร์เมื่อมีจำนวนองศาความเป็นอิสระแตกต่างกัน



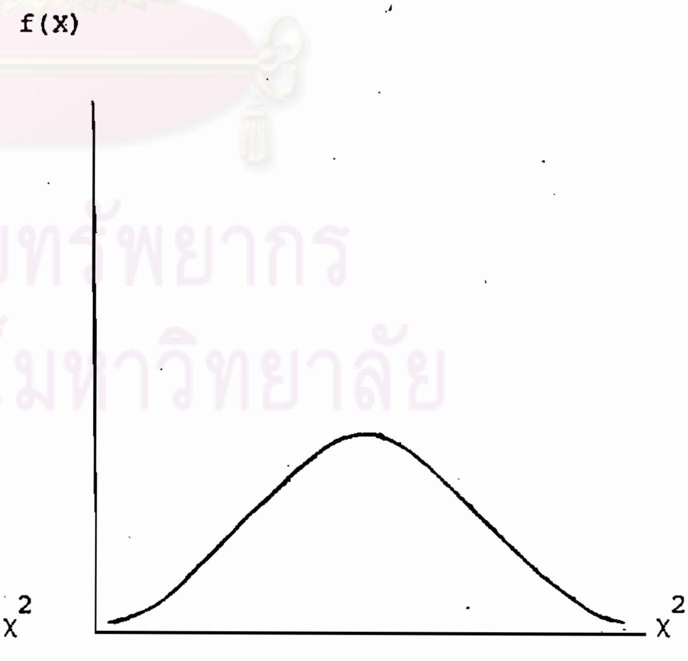
เมื่อองศาความเป็นอิสระเท่ากับ 1



เมื่อองศาความเป็นอิสระเท่ากับ 2



เมื่อองศาความเป็นอิสระเท่ากับ 3



เมื่อองศาความเป็นอิสระเท่ากับ 10

2.2.2 การใช้การทดสอบไคสแควร์

การทดสอบไคสแควร์สามารถใช้ประโยชน์ในการทดสอบสมมติฐานของข้อมูลที่อยู่ในรูปความถี่หรือเป็นข้อมูลจำนวนนับ หรืออยู่ในรูปตารางการณ์จรลักษณะต่าง ๆ เพื่อทดสอบความเป็นอิสระต่อกันของตัวแปรชุดที่สนใจศึกษา (Test of Independence) ทดสอบความเหมือนกันของข้อมูลที่สนใจศึกษา (Test of Homogencity) ทดสอบการแจกแจงของประชากรที่สนใจศึกษาว่าเป็นไปตามลักษณะการแจกแจงที่คาดไว้หรือไม่ ซึ่งเรียกว่าการทดสอบความเหมาะสมของตัวแบบหรือการทดสอบภาวะสารูปสนิทดี (Goodness of Fit)

ในที่นี้สนใจเฉพาะการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัวของข้อมูลเชิงปริมาณ โดยการทดสอบไคสแควร์ เมื่อนำข้อมูลมาจัดเสนอใหม่ในรูปของตารางการณ์จร

การทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว โดยการทดสอบไคสแควร์จากตารางการณ์จร (Contingency Table) สามารถดำเนินการได้เป็น 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว คือ

สมมติฐานว่าง Ho : ตัวแปร 2 ตัวเป็นอิสระซึ่งกันและกัน

สมมติฐานแย้ง Ha : ตัวแปร 2 ตัวขึ้นอยู่กับกันและกัน

หรือ Ho : ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน

Ha : ตัวแปรทั้งสองมีความสัมพันธ์กัน

ขั้นตอนที่ 2 ตัวสถิติที่ใช้ในการทดสอบ คือ

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2 \qquad \text{.............(2.21)}$$

$$\text{d.f.} = (R - 1)(C - 1)$$

เมื่อ O_{ij} คือ ความถี่ของข้อมูลที่ได้จากการสังเกตหรือทดลองในแถว (Row) ที่ i ของตัวแปรตัวที่ 1 และหลัก (Column) ที่ j ของตัวแปรตัวที่ 2

E_{ij} คือ ความถี่คาดหวังในแถวที่ i ของตัวแปรตัวที่ 1 และหลักที่ j ของตัวแปรตัวที่ 2 ภายใต้สมมติฐาน H_o ดังนั้นจะได้ว่า $E_{ij} = \frac{[X_{i.}][X_{.j}]}{X_{..}}$

$X_{i.}$ คือ ผลรวมของความถี่ในแถวที่ i ของตัวแปรตัวที่ 1

$X_{.j}$ คือ ผลรวมของความถี่ในหลักที่ j ของตัวแปรตัวที่ 2

$X_{..}$ คือ ผลรวมของความถี่ทั้งหมดหรือจำนวนข้อมูลทั้งหมด

R คือ จำนวนแถว (Row) ของตัวแปรที่ 1

C คือ จำนวนหลัก (Column) ของตัวแปรที่ 2

d.f. คือ องศาความเป็นอิสระ (Degree of Freedom) ซึ่งเท่ากับผลคูณของจำนวนแถวลบด้วยหนึ่ง (R - 1) กับจำนวนหลักลบด้วยหนึ่ง (C - 1)

ขั้นตอนที่ 3 คำนวณค่าไคสแควร์จากสูตรสมการที่ (2.21) และองศาความเป็นอิสระ

ขั้นตอนที่ 4 กำหนดระดับนัยสำคัญ (Level of Significance) สมมติให้เท่ากับ α

ขั้นตอดที่ 5 เปรียบเทียบค่า X^2 ที่คำนวณได้กับค่าไคสแควร์ $\left[X^2_{(R-1)(C-1),\ \alpha}\right]$ จากตารางการแจกแจงไคสแควร์ ณ องศาความเป็นอิสระเท่ากับ (R-1)(C-1) และระดับนัยสำคัญ α

ขั้นตอนที่ 6 สรุปผลการทดสอบ ถ้า $X^2 < X^2_{(R-1)(C-1),\alpha}$ จะยอมรับสมมติฐานว่าง Ho : ที่ว่าตัวแปร 2 ตัวเป็นอิสระซึ่งกันและกัน และถ้า $X^2 > X^2_{(R-1)(C-1),\alpha}$ จะปฏิเสธสมมติฐานว่าง Ho และยอมรับสมมติฐานแย้ง Ha : ที่ว่าตัวแปร 2 ตัวขึ้นอยู่กับกันและกัน

ในกรณีที่แต่ละตัวแปรที่นำมาทดลอบ เป็นตัวแปรชนิดที่แบ่งเป็น 2 ลักษณะเท่านั้น (Dichotomous Variable) นั่นคือ ข้อมูลอยู่ในตารางการณ์จรขนาด 2 x 2 ซึ่งมีองศาความเป็นอิสระเท่ากับ 1 และมีค่าขนาดความถี่คาดหวังบางช่องมีค่าไม่ถึง 20 จะใช้วิธีการคำนวณค่าไคสแควร์ ด้วยสูตรการปรับแก้ของเยทส์ (Yate's Correction) ดังนี้

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(\left| O_{ij} - E_{ij} \right| - 0.5)^2}{E_{ij}} \sim \chi^2_{(1)} \text{ ......(2.22)}$$

ในกรณีของตารางการณ์จรแบบ 2 x 2 ค่าไคสแควร์อาจคำนวณได้โดยใช้สูตรลัดดังนี้

$$X^2 = \frac{n(ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)} \sim \chi^2_{(1)} \text{ ....(2.23)}$$

ในกรณีที่ความถี่คาดหวังและจำนวนตัวอย่างทั้งหมดมีขนาดเล็ก ๆ จะต้องหาทางแก้ไขโดยใช้สูตรการปรับแก้ของเยทส์ (Yate's Correction for Continuity) เมื่อข้อมูลอยู่ในรูปตารางการณ์จรขนาด 2 x 2 ซึ่งมีองศาความเป็นอิสระเท่ากับ 1 จะได้สูตรการปรับแก้ของเยทส์เป็นดังนี้

$$X^2_{Corrected} = \frac{n(ad - bc - .5n)^2}{(a+c)(b+d)(a+b)(c+d)} \sim \chi^2_{(1)} \text{ .... (2.24)}$$

เมื่อ n คือ ขนาดตัวอย่าง

a คือ ความถี่ของค่าสังเกตในแถวที่ 1 และสดมภ์ที่ 1

b คือ ความถี่ของค่าสังเกตในแถวที่ 1 และสดมภ์ที่ 2

c คือ ความถี่ของค่าสังเกตในแถวที่ 2 และสดมภ์ที่ 1

และ d คือ ความถี่ของค่าสังเกตในแถวที่ 2 และสดมภ์ที่ 2

ตารางที่ 2.3 แสดงลักษณะของข้อมูลในตารางการณ์จรขนาด 2 x 2

ตัวแปรที่ 2 / ตัวแปรที่ 1	สดมภ์ที่ 1	สดมภ์ที่ 2	รวม
แถวที่ 1	a	b	a+b
แถวที่ 2	c	d	c+d
รวม	a+c	b+d	n

2.3 ผลงานที่เกี่ยวข้องกับการวิจัย

จากสูตรที่ใช้ในการคำนวณไคสแควร์ $\left[X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim X^2_{(R-1)(C-1)} \right]$ จะได้ว่า ค่าไคสแควร์ได้จากผลรวมกำลังสองของผลต่างระหว่างความถี่ของค่าสังเกตและความถี่ของค่าคาดหวังแล้วหารด้วยความถี่ของค่าคาดหวัง ดังนั้น หากมีการรวมกันมากครั้งเท่าใด ค่าของไคสแควร์จะมากขึ้นตามไปด้วย ทั้งที่ระดับความสัมพันธ์ระหว่างตัวแปรยังเหมือนเดิม นอกจากนี้ ค่าของผลต่างระหว่างความถี่ของค่าสังเกตและความถี่ของค่าคาดหวังยังขึ้นอยู่กับจำนวนความถี่ของค่าสังเกต ยิ่งจำนวนความถี่ของค่าสังเกตมีมากขึ้นเท่าใด ค่าไคสแควร์จะมีค่ามากขึ้นด้วย ทั้งที่ระดับความสัมพันธ์ระหว่างตัวแปรยังเหมือนเดิม ดังนั้น ค่าของไคสแควร์ที่สูงมากไม่สามารถจะบอกได้ว่า ตัวแปรอิสระก่อให้เกิดความแตกต่างต่อตัวแปรตามมากด้วย กล่าวคือ ไม่สามารถใช้บ่งชี้ได้ว่า ตัวแปรทั้งสองตัวมีความสัมพันธ์กันมาก ซึ่งจะเห็นว่าค่าของไคสแควร์มีค่าขึ้นอยู่กับจำนวนช่องในตาราง (องศาความเป็นอิสระ) และจำนวนความถี่ จึงเป็นจุดอ่อนของการทดสอบความสัมพันธ์ระหว่างตัวแปรโดยใช้การทดสอบไคสแควร์ ซึ่งควรระมัดระวังไว้ (สุชาติ ประสิทธิ์รัฐสินธุ์ และพิสิฐ ศุภรัยพงศ์ 2525 : 27-30)

การใช้การทดสอบไคสแควร์เพื่อทดสอบความสัมพันธ์ระหว่างตัวแปรยังมีข้อพึงระมัดระวังอื่น ๆ อีก คือ

1. ค่าความถี่คาดหวังควรจะมากกว่า 5 หากจะให้เชื่อถือได้มากขึ้น ค่าความถี่คาดหวังควรจะมากกว่า 10

2. ตัวแปร 2 ตัวที่นำมาทดสอบความสัมพันธ์กัน จะต้องไม่มีความสัมพันธ์กันตั้งแต่การเก็บข้อมูล เช่น น้ำหนักของบุคคลย่อมมีความสัมพันธ์กับส่วนสูง ในกรณีเช่นนี้ไม่ควรใช้การทดสอบไคสแควร์ เพราะค่าของไคสแควร์ได้มาจากการสมมติว่า ตัวแปรทั้งสองตัวเป็นอิสระต่อกัน

3. ผู้ใช้ควรระมัดระวังการแบ่งกลุ่มชั้นของแต่ละตัวแปรไม่ให้เกิดปัญหา หรือมีข้อสงสัยว่าเป็นการแบ่งชั้นที่ไม่ถูกต้อง ทั้งนี้เพราะการแบ่งกลุ่มชั้นของแต่ละตัวแปร อาจสร้างความสัมพันธ์ระหว่างตัวแปรได้ โดยที่ตัวแปรทั้งสองไม่มีความสัมพันธ์กันตั้งแต่แรก

ในการวิเคราะห์ข้อมูลจากตารางการณ์จร (Contingency Table) มักพบปัญหาเกี่ยวกับความถี่คาดหวังมักมีขนาดเล็กหรือความถี่คาดหวังมีค่าต่ำเกินไปคือ มีค่าต่ำกว่า 5 แม้ว่ายังไม่มีกฎเกณฑ์ตายตัวในการจัดการกับปัญหานี้ แต่นิยมกระทำกันดังนี้คือ ในกรณีที่ทดสอบตารางการณ์จรที่มีองศาความเป็นอิสระมากกว่า 1 ค่าของขนาดความถี่คาดหวังของแต่ละช่อง (Cell) ที่ต่ำกว่า 5 ควรมีอยู่ไม่เกินร้อยละ 20 ของช่องข้อมูลทั้งหมด ดังนั้น ในบางครั้งจึงมักรวมข้อมูลที่สังเกตมาในช่องต่าง ๆ ที่อยู่ติดต่อกัน เพื่อให้ค่าของขนาดความถี่คาดหวังมีค่ามากกว่า 5 หรือหากรวมกลุ่มไม่ได้ ควรตัดกลุ่มที่น้อยและไม่อาจรวมในกลุ่มอื่นได้ออกจากการวิเคราะห์ไป แม้ว่าจะเป็นวิธีการที่ไม่นิยมใช้กัน ในกรณีของตารางการณ์จรขนาด 2 x 2 เมื่อมีค่าความถี่คาดหวังน้อยเกินไป คือต่ำกว่า 5 จะต้องหาทางแก้ไขโดยใช้สูตรการปรับแก้ของเยทส์ (Yate's Correction for Continuity) ซึ่งการแก้ไขนี้ควรทำแม้แต่ในกรณีที่ค่าคาดหวังมากกว่า 5 แต่น้อยกว่า 10 (วัชราภรณ์ สุริยาภิวัฒน์ 2527 : 464-467)