บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนของวรรณคดีที่เกี่ยวข้อง นำเสนอเป็น 3 ตอน คือ

ตอนที่ 1 ประวัติและพัฒนาการของไคสแควร์

ตอนที่ 2 การแจกแจงไคสแควร์

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ประวัติและพัฒนาการของไคสแควร์

คำว่า ไค (CHI) นั้น เป็นคำที่มาจากภาษากรีก ใช้สัญลักษณ์ว่า χ แต่เดิมไม่ได้เรียกว่าไคสแควร์ (Cochran, 1952) เพราะในปีคริสตศักราช 1836 เฮลเมอร์ท (Robert Fricdrich Helmert) ชาวเยอรมันได้ค้นพบการแจกแจงที่เป็นโค้งปกติ และเรียกว่าการแจกแจงแบบเฮลเมอร์ท (Helmert's Distribution) (Rahman 1968) ดังนี้

สมมุติว่า $X_1, X_2, \ldots, X_n$ เป็นกลุ่มที่ได้จากการสังเกต (Observation) ที่เป็นอิสระจากการแจกแจงปกติ (Normal Distribution) มีมัชฌิมเท่ากับ u และค่าความแปรปรวนเท่ากับ σ^2 แล้ว การแจกแจงความน่าจะเป็น (Probability) ของ Xi ใด ๆ คือ

$$\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\frac{(X_i - u)^2}{\sigma}}\, dX_i \qquad \text{ซึ่ง } -\infty < X_i < \infty \quad \text{และ } 0 < \sigma < \infty$$

ถ้าเราแปลง Xi ให้เป็นค่ามาตรฐาน $Z_i = (X_i - u)/\sigma$ แล้ว Zi แต่ละตัวจะเป็นหน่วยที่เป็นอิสระที่มีการแจกแจงความน่าจะเป็น

$$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2} Z_i^2}\, dz_i \qquad -\infty \; Z_i < \infty$$

การแจกแจงนี้เป็นการแจกแจงความน่าจะเป็นแบบปกติ และเขาหยุดงานไว้เพียงนี้ การกระจายที่เดอเมอร์ท ค้นพบนี้ได้ถูกละเลยเป็นเวลา 20 ปีเศษ

ต่อมาในปีคริสตศักราช 1900 เปียร์สัน (Karl Pearson) ได้ค้นพบใหม่และผลงานชิ้นนี้นับว่าเป็นรากฐานสำหรับสถิติแผนใหม่ เขาเริ่มจากการพิสูจน์ว่าถ้ากลุ่มของตัวแปรที่สัมพันธ์กัน Zi (Correlated Variates) n ตัว มีมัชฌิมเท่ากับศูนย์ มีการแจกแจงที่มีเหตุการณ์ที่เกิดขึ้นมีจำนวนมากกว่า 2 และมีการแจกแจงความน่าจะเป็น ดังนี้

$$dF = (2\pi)^{-\frac{n}{2}} \exp^{\frac{Q}{2}} \; dZ1\, dZ2\, dZ3 \ldots\ldots dZn$$

(Charles C. Peter and Walter Van Voorhis 1940) แล้ว กำลังสองของ Q นี้เขาให้ชื่อว่าเป็นการแจกแจง X^2 ที่มีจำนวนชั้นแห่งความเป็นอิสระ n

การพิสูจน์ใช้วิธีทางเรขาคณิต ดังนี้ (Peter and Voorhis 1940 : 404-424)

ถ้า x เป็นค่าใด ๆ ที่เบี่ยงเบนจากมัชฌิมประชากรที่มีการแจกแจงเป็นโค้งปกติแล้ว

$$X^2 = \sum_{i=1}^{n} \frac{Xi^2}{\sigma_x^2} = \sum_{i=1}^{n} Zi^2 \text{ (Rohman 1968 : 375 - 377)}$$

โดยที่ x อาจเป็นมาตรการเดี่ยว (Single Measure) เป็นมัชฌิมของกลุ่มตัวอย่าง เป็นสัดส่วนหรือเป็นค่าสถิติอื่น ๆ ที่มีการแจกแจงเป็นโค้งปกติ

สมมุติว่าพิจารณาค่าของ x หนึ่งค่าในแต่ละครั้ง ค่าของ x^2 แต่ละค่าก็จะเป็น x^2/σ_x^2 เหมือนกัน และถ้าเขียน z_1^2 แทน x^2/σ_x^2 ความน่าจะเป็นของ z_1 หนึ่งค่าที่อยู่ในพิสัย (Range) dZ1 จะเป็น

$$dF = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z1^2} dZ1$$

โอกาสที่จะเป็นไปได้ของการได้ค่า x เท่ากับหรือมากกว่าระดับความมีนัยสำคัญ จะเป็นการหาความน่าจะเป็นของฟังก์ชั่น การแจกแจงโค้งปกติ เช่น

$$P = \int_{Z1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} Z1^2} dZ1$$

ค่านี้เราอาจหาได้จากตารางพื้นที่ใต้โค้งปกติ

ถ้าเราต้องการหาค่าของ x 2 ค่าที่เป็นอิสระต่อกันพร้อมกัน ความน่าจะเป็นที่ตัวแปร 2 ตัวนี้ จะปรากฏพร้อมกันในพื้นที่ที่เพิ่มขึ้น (Elemental Area) $dZ_1 \, dZ_2$ ก็คือ ผลคูณของความน่าจะเป็นของแต่ละกลุ่มตัวอย่าง

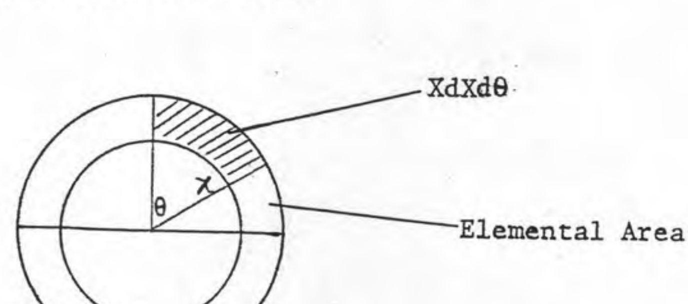
$$dF = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z_1^2} dz_1\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z_2^2} dz_2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-\frac{1}{2}(Z_1^2+ Z_2^2)} \, dZ_1 dZ_2 \ldots\ldots(1)$$

และความน่าจะเป็นของการได้ค่า x ในกลุ่มตัวอย่างเท่ากับหรือมากกว่า 2 ค่านี้ คือ

$$P = \int_{Z_1}^{\infty}\int_{Z_2}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-\frac{1}{2}(Z_1^2+ Z_2^2)} \quad dZ_1 dZ_2$$

ถ้าพิจารณา $x^2= z_1^2+ z_2^2$ และต้องการคำนวณหาความน่าจะเป็น จะได้ค่าของ x^2 ที่คล้ายคลึงกัน ถึงแม้ว่า z_1 และ z_2 อาจแปรไปจากกลุ่มตัวอย่างหนึ่งไปอีกกลุ่มหนึ่ง และสมการนี้จะเป็นสมการวงกลม มีจุดศูนย์กลางที่จุดกำเนิดและรัศมีเท่ากับ x ดังนั้นพื้นที่ใด ๆ ที่ความน่าจะเป็นรวม (Joint Probability) อาจเป็นตำแหน่งใด ๆ ในเขตวงกลมนี้ การหาพื้นที่ (Double Integral) จำเป็นต้องเปลี่ยนให้เป็น โพลาร์ โคออร์ดิเนต (Polar Coordinate) เพื่อให้ง่ายต่อการคำนวณ



ดังนั้นพื้นที่แรก (Elemental Area) ในรูป Polar Coordinate จะเป็น XdXdθ (ดูภาคผนวก ข) และถ้าแทนที่ $z_1^2 + z_2^2$ ด้วย x^2 จะได้

$$dF = \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-\frac{1}{2}X^2} XdXd\theta$$

ถ้าต้องการหาความน่าจะเป็นทั้งหมดของ x ภายในวงกลมจะต้องอินทิเกรต (Integrate) ประพจน์ (Expression) ข้างต้นเทียบกับมุม θ จาก o ถึง 2π มี $\left(\frac{1}{\sqrt{2\pi}}\right)^2 Xe^{-\frac{1}{2}X^2}dx$ เป็นค่าคงที่ เพราะฉะนั้น

$$dF = \left(\frac{1}{\sqrt{2\pi}}\right)^2 Xe^{-\frac{1}{2}X^2} dx \left[\theta\right]_0^{2\pi}$$

$$= \frac{1}{2\pi} Xe^{-\frac{1}{2}X^2} dx(2\pi)$$

$$\therefore dF = Xe^{-\frac{1}{2}X^2} dx \quad \ldots\ldots\ldots\ldots\ldots(2)$$

จากสูตร (2) กำลังของ x นั้นน้อยกว่าตัวแปรอยู่ 1 เสมอ
ถ้าดำเนินวิธีการเดียวกัน สำหรับตัวแปร 3 ตัว จะได้

$$dF = kX^2 e^{-\frac{1}{2}X^2} dX$$

ซึ่งมี k เป็นค่าคงที่ และกำลังของ x นั้นเป็น 2 เมื่อจำนวนตัวแปรที่เป็นอิสระมากกว่า 3 ถึง n จะกำหนดสมการเป็น (Peter and Voorhis, 1940)

$$X^2 = Z_1^2 + Z_2^2 + \text{----}\ Z^2 n$$

และมีความน่าจะเป็นในทรงตัน

$$dF = kx^{n-1} e^{-x^2/2} dx$$

เมื่อ k เป็นค่าคงที่เท่ากับ $\frac{1}{\Gamma n/2} \; 2^{n/2}$

$$\therefore dF = \frac{1}{\Gamma n/2 \; 2^{n/2}} x^{n-1} e^{-\frac{1}{2}x^2}$$

หลังจากนั้นเปียร์สัน (Karl Pearson) ได้เสนอการทดสอบภาวะสารูปสนิทดีที่ข้อมูลจัดแยกออกจากกัน โดยเริ่มจากการทราบความถี่ที่คาดหวัง mi ล่วงหน้า และ xi ใด ๆ ที่ได้จากการสังเกตจะมีการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution) เปียร์สันยอมรับว่า xi นั้น อาจมีการแจกแจงเป็นโค้งปกติ ณ ตรงจุดนี้เขายอมรับว่าค่า Ei นั้นใหญ่พอในทุก ๆ ช่อง ผลงานในตอนแรกของเปียร์สันทำให้ทราบว่าถ้าตัวอย่างขนาดใหญ่ค่า x^2 จะมีการแจกแจงแบบ χ^2 ที่มีค่าความเป็นอิสระ (Degree of Freedom : df) เท่ากับ k - 1

ฟิชเชอร์ (R.A. Fisher) ได้หลีกเลี่ยงความยุ่งยากในงานของเปียร์สัน โดยชี้ให้เห็นว่าถ้ายอมรับว่าสิ่งที่สังเกตได้ xi มีการแจกแจงแบบพัวซอง (Poisson Distribution) คือ

$$\prod_{i=1}^{k} \frac{e^{-mi} \, mi^{xi}}{xi!} = e^{-n} \prod_{i=1}^{k} \frac{mi^{xi}}{xi!} \quad ; \quad m=n$$

มีค่า mi เป็นความถี่ที่คาดหวังหรือมัชฌิมและขนาดของกลุ่มตัวอย่าง n ใหญ่มาก

$$Y = \frac{Xi - mi}{\sqrt{mi}}$$

จะมีการแจกแจงโค้งปกติและการแจกแจงแบบพัวซองมีมัชฌิม mi และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\sqrt{mi}$ ดังนั้นการแจกแจง χ^2 คือ $Y_1^2+Y_2^2+\dots Y_k^2$; Y_i มีการแจกแจงอย่างอิสระแต่มีเงื่อนไขว่า

$$\sum_{i=1}^{k} Yi\sqrt{mi} = \sum_{i=1}^{k} (Xi-mi) = 0$$

เปียร์สันยังคิดถึงกรณีที่ mi ขึ้นอยู่กับพารามิเตอร์ที่ประมาณจากกลุ่มตัวอย่าง ค่าที่คาดหวัง mí ที่ได้จากกลุ่มตัวอย่าง มีความแตกต่าง

$$\chi^2 - \chi'^2 = \sum_{i=1}^{k} \frac{xi^2}{mi} - \sum_{i=1}^{k} \frac{xi'^2}{mi}$$

ที่มีค่าเป็นบวก เมื่อเราสามารถปรับการประมาณค่าพารามิเตอร์ให้เหมาะกับกลุ่มตัวอย่างชนิดต่าง ๆ ค่าความแตกต่างนี้จะมีค่าบวกที่น้อยมาก และจะนับ χ'^2 มีการแจกแจงเหมือน χ^2 ที่มี df = k - 1

ตอนที่ 2 การแจกแจงของไคสแควร์

โดยปกติแล้ว การใช้ไคสแควร์ในการทดสอบจะสัมพันธ์กับการเปรียบเทียบการแจกแจงข้อมูลที่ต่อเนื่อง แต่ในทางปฏิบัติมักจะใช้กับการแจกแจงของข้อมูลที่ไม่ต่อเนื่อง เช่น ข้อมูลที่เป็นความถี่ ทำให้โค้งขาดตอนซึ่งก็ถือว่ามีการแจกแจงแบบไคสแควร์ ลักษณะการแจกแจงของไคสแควร์นั้นขึ้นอยู่กับจำนวนชั้นแห่งความเป็นอิสระ ถ้าจำนวนชั้นแห่งความเป็นอิสระมาก ก็จะมีการแจกแจงที่เบน้อย โดยใช้สูตรในการคำนวณดังนี้

$$\chi^2 = \sum_{i=1}^{k} \frac{(Oi - Ei)^2}{Ei} \quad ....(1)$$

Oi = ความถี่ที่ได้จากการสังเกต (Observation Frequency)

Ei = ความถี่ที่คาดหวัง (Expected Frequency)

แต่สมการส่วนโค้งการกระจายของไคสแควร์ที่แท้จริง เป็นดังนี้

$$Y = \frac{1}{(n/2-1)!\,2^{n/2}} (X^2)^{(n/2-1)} e^{-\frac{1}{2}X^2} \quad ....(2)$$

ทั้งสูตร (1) และ (2) นั้นเมื่อหาค่าความน่าจะเป็นแล้วจะมีค่าใกล้เคียงกันมาก เปียร์สัน (Karl Pearson) ได้แสดงให้เห็นว่าค่าที่ได้จากทั้ง 2 สูตรนั้นมีสหสัมพันธ์กันอยู่ระหว่าง .93-.99 ซึ่งนับว่าสูงมาก (Peter and Voorhis, 1940 : 413)

ดังนั้นจึงได้นำไคสแควร์ไปใช้กันอย่างกว้างขวาง ไม่ว่าข้อมูลเหล่านั้นจะเป็นความถี่ที่ต่อเนื่องหรือไม่ต่อเนื่องก็ได้ แต่หลักใหญ่คือการเปรียบเทียบความแตกต่างของความถี่ที่คาดหวังกับความถี่ที่ได้จากการสังเกตซึ่งในที่นี้ความถี่ที่คาดหวังจะเป็นตัวพารามิเตอร์ และความถี่ที่สังเกตได้จะเป็นค่าประมาณจากพารามิเตอร์ ถ้าทั้งสองค่านี้ใกล้เคียงกันการทดสอบจะไม่มีนัยสำคัญ

เหตุผลสำคัญที่ใช้สูตรไคสแควร์ในรูป $\sum_{i=1}^{k}(Oi - Ei)^2/Ei$ ก็เพราะเป็นผลบวกของค่าประมาณ k ตัว ที่ได้มาจากโค้งปกติซึ่งจะทำให้ค่า (Oi - Ei)/Ei มีการแจกแจงเป็นโค้งปกติมีมัชฌิมเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 เหมือนกับลักษณะการแจกแจงไคสแควร์ที่แท้จริง ดังนั้นจึงสามารถใช้แทนกันได้

เนื่องจากความโค้งของการแจกแจงขึ้นอยู่กับจำนวนชั้นแห่งความเป็นอิสระ ลักษณะการแจกแจงของไคสแควร์จึงแตกต่างกันไปตามจำนวนชั้นแห่งความเป็นอิสระ แต่เท่าที่ปรากฏนั้นจำนวนชั้นแห่งความเป็นอิสระ ยากที่จะปรากฏในการปฏิบัติจริง ดังนั้นเราจะเกี่ยวข้องกับการกระจายที่เป็นไปได้เท่านั้น ส่วนโค้งการแจกแจงของไคสแควร์นั้นจะเกิดขึ้นตามชนิดของจำนวนชั้นแห่งความเป็นอิสระโดยมีค่าของไคสแควร์เป็นแอบซิสซา (Absisa) และให้แกน Y เป็นออร์ดิเนต (Ordinate)

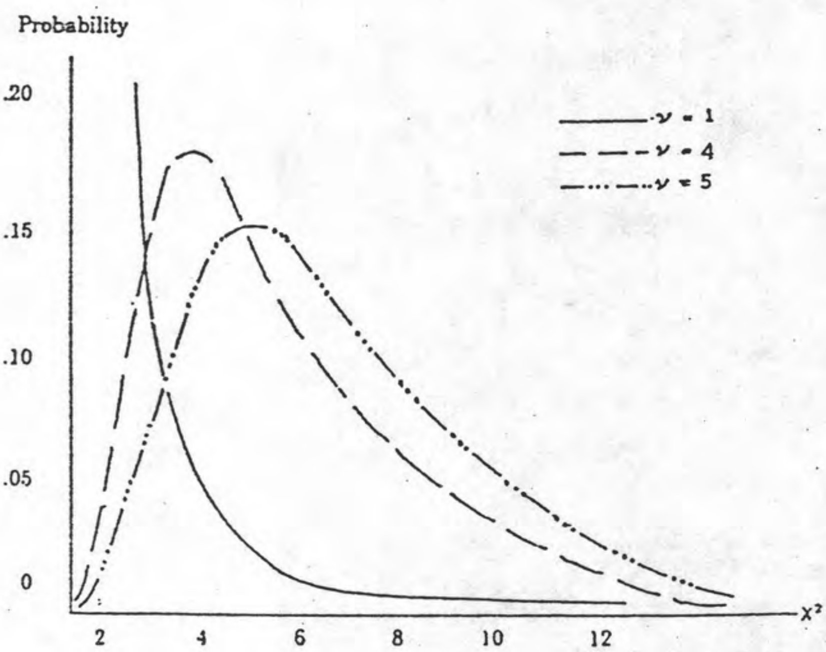
เมื่อจำนวนชั้นแห่งความเป็นอิสระเท่ากับ 1 ส่วนโค้งจะเริ่มจากส่วนสูงสุดค่อย ๆ ลาดเข้าหาออร์ดิเนต (Ordinate) ลดลงไปเรื่อย ๆ จนถึงค่าอนันต์ (Infinity) ความสูงของ Y ที่ x^2= 0.6 เท่ากับ .92 และที่ x^2= .01 ค่าของ Y จะมากกว่า 4 เท่าของ .92 และถ้าค่าของ x^2= 1.0 แล้ว ความสูงจะมีค่าเท่ากับ .242 ส่วนโค้งเมื่อลาดต่ำมาถึง 6.25 ความสูงจะประมาณ .007 โดยไม่คำนึงถึงจำนวนชั้นแห่งความเป็นอิสระ และส่วนโค้งทางด้านขวามือนั้นจะลาดสู่แกน x เรื่อย ๆ แต่ไม่ติดแกน x ถ้าเราให้พื้นที่ใต้โค้งมีค่าเท่ากับ 1 แล้ว พื้นที่ระหว่างออร์ดิเนตกับแอบซิสซา 2 จุดนั้น สามารถคำนวณออกมาเป็นสัดส่วนได้ เช่นเมื่อจำนวนชั้นแห่งความเป็นอิสระเท่ากับ 1 .99 ของพื้นที่จะมากกว่าค่าของไคสแควร์ที่เท่ากับ .000157 และ .05 ของพื้นที่เท่านั้นที่มากกว่า 3.84 หรืออาจ กล่าวได้ว่าความน่าจะเป็นที่ค่าของ x^2จะมากกว่า 3.84 นั้นเท่ากับ .05 ถ้าค่า x^2= 6.63 ความน่าจะเป็นเท่ากับ .01 และเมื่อ x^2 = 10.827 ค่าความน่าจะเป็นจะเท่ากับ .001

เมื่อจำนวนชั้นแห่งความเป็นอิสระเท่ากับ 2 ส่วนที่โค้งจะเริ่มจากความสูง .05 และค่อย ๆ ลดต่ำลงด้วยความชันที่น้อยกว่า ส่วนโค้งที่มีจำนวนชั้นแห่งความเป็นอิสระเท่ากับ 1 และค่าของไคสแควร์จะปรากฏบ่อยครั้งขึ้น สำหรับจำนวนชั้นแห่งความเป็นอิสระเท่ากับ 3

การกระจายของส่วนโค้ง เริ่มต้นจากความสูงเท่ากับศูนย์ และจะมีความสูงสุดอยู่ที่ค่า $x^2 = 1$ จากนั้นส่วนโค้งจะลาดต่ำลงมาเรื่อย ๆ

เมื่อจำนวนชั้นแห่งความเป็นอิสระมีจำนวนมากขึ้น ส่วนโค้งจะเบน้อย และจะมีตำแหน่งทางขวามือมากขึ้น ๆ จนกระทั่งเป็นรูปใกล้เคียงกับโค้งปกติ

การแจกแจงไคสแควร์ จะกำหนดด้วยสัญลักษณ์ x_v^2 ซึ่งกราฟของการแจกแจงไคสแควร์มีลักษณะดังแผนภาพที่ 1



แผนภาพที่ 1 การแจกแจงไคสแควร์ที่มีชั้นแห่งความเป็นอิสระต่างกัน

ก. <u>สมการของการแจกแจงไคสแควร์</u>

การแจกแจงไคสแควร์ เป็นการแจกแจงประเภทหนึ่งของการแจกแจงแกมม่า (Gamma Distribution) ถ้า x เป็นตัวแปรสุ่มแบบต่อเนื่อง และ ν เป็นชั้นแห่งความเป็นอิสระที่มีค่าเป็นจำนวนเต็มบวกแล้ว ฟังก์ชั่นความน่าจะเป็นของตัวแปร x เป็นดังนี้

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

เมื่อ $\Gamma\alpha$ คือ ฟังก์ชั่นแกมม่ามีค่าตามสูตร คือ

$$\Gamma\alpha = \int_0^\infty x^{\alpha-1} e^{-x} dx \qquad ;\alpha>0$$

เรียกตัวแปรสุ่ม x ว่ามีการแจกแจงไคสแควร์ ที่มีชั้นแห่งความเป็นอิสระเท่ากับ ν

ถ้า s^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n ที่สุ่มจากประชากรที่มีการแจกแจงแบบปกติที่มีมัชฌิมเลขคณิตเท่ากับ u และความแปรปรวนเท่ากับ σ^2 จะได้

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

มีการแจกแจงแบบไคสแควร์ ที่มีชั้นแห่งความเป็นอิสระ เท่ากับ n − 1

ข. คุณสมบัติของการแจกแจงไคสแควร์

1. การแจกแจงของไคสแควร์ตามทฤษฎีเป็นการแจกแจงที่ต่อเนื่อง

2. ความโค้งจะเริ่มจาก 0 ถึง ∞

3. ค่าออร์ดิเนตสูงสุดจะอยู่ที่ $x^2_{(n-1)}$ โดยที่ n คือ จำนวนชั้นแห่งความเป็นอิสระ

4. การแจกแจงจะมีลักษณะแตกต่างกันตามค่าของชั้นแห่งความเป็นอิสระ ถ้า $n \geqslant 3$ การแจกแจงจะเบ้บวกหรือเบ้ขวา (Positively Skewed) และเมื่อชั้นแห่งความเป็นอิสระมีค่าเพิ่มขึ้น การแจกแจงจะมีลักษณะใกล้เป็นการแจกแจงแบบปกติมาตรฐาน

5. เป็นส่วนโค้งของเพียร์สันชนิดที่ 3 (Pearson Type III Curve)

6. จุดสูงสุดของโค้งมีเพียงจุดเดียว

7. การแจกแจง เป็นไปตามกฎของการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution)

8. สำหรับการแจกแจงของไคสแควร์ เมื่อชั้นแห่งความเป็นอิสระเท่ากับ n จะมีลักษณะดังนี้

8.1 ค่าเฉลี่ยจะมีค่าเท่ากับ n และค่าความแปรปรวนเท่ากับ 2 n

8.2 ค่ามัธยฐานเท่ากับ n – 2

8.3 ค่าความเบ้ของการแจกแจงเท่ากับ $\sqrt{8/n}$

<u>ตอนที่ 3 งานวิจัย</u>

สแลคเตอร์ (Slakter 1966 : 619–623) ได้สำรวจความถูกต้องของการทดสอบไคสแควร์ที่เป็นแบบภาวะสารูปสนิทดี (Goodness of fit test) โดยใช้กลุ่มตัวอย่างขนาด 10, 25 และ 50 ทำการทดสอบกับข้อมูลที่มีการกระจายแบบยูนิฟอร์ม กำหนดค่าความถี่ที่คาดหวัง ตั้งแต่ .05 (N = 10, k = 20) จนถึงค่าความถี่ที่คาดหวังเท่ากับ 5 (N = 50, k = 10) ในการทดสอบแต่ละกรณีกระทำ 10,000 ครั้ง โดยกำหนดอัตราความคลาดเคลื่อนประเภทที่ 1 (α) ที่ระดับ .01, .05 และ .10 สแลคเตอร์ได้สรุปว่าความคลาดเคลื่อนประเภทที่ 1 ที่ได้จากการทดลองนั้นใกล้เคียงกับอัตราความคลาดเคลื่อนประเภทที่ 1 ที่กำหนด นอกจากนี้การทดสอบไคสแควร์โดยใช้การทดสอบภาวะสารูปสนิทดียังมีความแกร่ง (Robustness)อีกด้วย

ลีวอนตินและเฟลเซนสไตน์ (Lewontin and Felsenstein 1965 : 19–23) ได้ทำการศึกษาโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) ใช้ประชากรที่มีการกระจายแบบยูนิฟอร์ม เพื่อดูความแกร่งของการทดสอบไคสแควร์ที่เป็นแบบอิสระ (Chi - Square test of Independence) โดยใช้ตารางการณ์จรขนาด 2 X k กำหนดความถี่รวมทั้งแถวและคอลัมน์ให้คงที่ โดยศึกษากับตัวอย่างขนาด 4, 8 และ 12 กำหนดจำนวนเซลล์เป็น 5 และ 10 พบว่าการทดสอบด้วยไคสแควร์มีความแกร่งภายใต้ภาวะการณ์ที่กำหนด

ยาร์โนลด์ (Yarnold 1970 : 864–886) ได้ศึกษาเกี่ยวกับการทดสอบภาวะสารูปสนิทดี ในหลาย ๆ สถานการณ์ที่แตกต่างกันได้ผลสรุปว่า การประมาณด้วยไคสแควร์นั้นสามารถจะใช้ค่าความถี่ที่คาดหวังต่ำกว่าที่เคยยอมรับกันมาแต่เดิม โดยได้แนะนำกฎในการหาค่าความถี่ที่คาดหวังว่า ถ้าจำนวนเซลล์ s เป็น 3 หรือมากกว่า และ r เป็นจำนวนค่าความถี่ที่คาดหวังที่ต่ำกว่า 5 แล้ว ค่าความถี่ที่คาดหวังต่ำสุดจะมีค่าเท่ากับ 5 r/ s

รอสโคว์และไบยาร์ (Roscoe and Byars 1971 : 755-759) ได้ทำการศึกษาโดยใช้วิธีซิมูเลทกับการทดสอบไคสแควร์ที่เกี่ยวกับความเป็นอิสระโดยใช้กลุ่มตัวอย่างขนาด 10, 15, 20, 30, 50 และ 100 และใช้วิเคราะห์ตารางการณ์จร ขนาด 2 X 2 ถึง 5 X5 ผลปรากฏว่าความถี่ที่คาดหวังโดยเฉลี่ยแล้วอาจจะต่ำถึง 2 หรือจะสูงถึง 10 ก็ได้ นอกจากนี้ไคสแควร์ยังมีความแกร่งในการทดสอบภาวะสารูปสนิทดีที่มีการกระจายแบบยูนิฟอร์ม

มาร์ช (March, cited by Tate and Hyer 1973) ได้ศึกษาการทดสอบไคสแควร์ที่เป็นตารางการณ์จร ขนาด 2 X 3 โดยวิธีซิมูเลชั่น และได้เสนอแนะว่าความถี่ที่คาดหวังต่ำกว่า 5 หรือ 10 นั้น จะไม่ทำให้การตัดสินใจคลาดเคลื่อนอย่างร้ายแรง แต่เพื่อความมั่นใจควรใช้ค่าความถี่ที่คาดหวังที่สูงกว่านี้

กริซเซิล (Grizzle 1967 : 28-32) ได้ศึกษาความคลาดเคลื่อนประเภทที่ 1 ที่เกิดจากการใช้สถิติทดสอบไคสแควร์กับตารางการณ์จร ขนาด 2 X 2 โดยเปรียบเทียบระหว่างสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้และไม่ใช้ค่าแก้ ทำการศึกษาเป็น 2 กรณี กรณีที่ 1 กำหนดให้ผลรวมของคอลัมน์มีค่าคงที่ ใช้กลุ่มตัวอย่าง ขนาด 10, 20, 30, 40 และ 50 ตามลำดับ ในกรณีที่ 2 กำหนดให้ผลรวมของแถวและคอลัมน์สามารถเปลี่ยนแปลงอย่างอิสระ ใช้กลุ่มตัวอย่าง ขนาด 20, 40, 60, 80, 100, และ 120 ตามลำดับ การศึกษาใช้วิธีซิมูเลทโดยจำลองการทดลอง 500 ครั้งในแต่ละกรณี และกำหนดความคลาดเคลื่อนประเภทที่ 1 ที่ระบุเท่ากับ .05 ผลการศึกษาพบว่าในทุกกรณีสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้มีลักษณะ Conservative ส่วนค่าสถิติทดสอบไคสแควร์ที่ไม่ใช้ค่าแก้นั้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับความคลาดเคลื่อนประเภทที่ 1 ที่ระบุ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก ค่าสถิติทดสอบไคสแควร์ที่ไม่ใช้ค่าแก้จะมีลักษณะ Conservative แต่ค่าสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้จะมีลักษณะ Conservative มากกว่า

สตาร์เมอร์ กริซเซิลและเซน (Starmer, Grizzle and Sen 1974 : 376-378) ได้ศึกษาความคลาดเคลื่อนประเภทที่ 1 ที่เกิดจากการใช้สถิติทดสอบไคสแควร์กับตารางการณ์จร ขนาด 2 x 2 โดยเปรียบเทียบระหว่างสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้ สถิติทดสอบไคสแควร์ที่ไม่ใช้ค่าแก้ ฟิชเชอร์ เอ็กแซก เทส (Fisher's Exact Test) และ แรนดอมไมซ-เทส (Randomized Test) ทำการศึกษา 2 กรณี กรณีที่ 1 กำหนดให้ผลรวมของคอลัมน์มีค่าคงที่ ใช้กลุ่มตัวอย่างขนาด 10, 20, 30, 40, 50, 60, 70, 80, 90, และ 100 ตามลำดับ

ในกรณีที่ 2 กำหนดให้ผลรวมของแถวและคอลัมสามารถเปลี่ยนแปลงอย่างอิสระ ใช้กลุ่มตัวอย่างขนาด 20, 40, 60, 80, 100, 120, 140, 160, 180 และ 200 ตามลำดับ การศึกษาใช้วิธีซิมูเลทโดยจำลองการทดลอง 2,000 ครั้งในแต่ละกรณี ผลการศึกษาพบว่าสถิติทดสอบฟิชเชอร์ เอ็กแซก เทส และสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้มีลักษณะ Conservative เมื่อกลุ่มตัวอย่างมีขนาดใหญ่พอสมควร สถิติทดสอบไคสแควร์ที่ไม่ใช้ค่าแก้มีค่าความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับแรนดอมไมซ์-เทส (ซึ่งใช้เป็นตัวเปรียบเทียบ) มากกว่าค่าสถิติทดสอบไคสแควร์ที่ใช้ค่าแก้ และฟิชเชอร์ เอ็กแซก เทส ที่ใช้กลุ่มตัวอย่างขนาดกลางหรือขนาดใหญ่ แต่สถิติทดสอบไคสแควร์ที่ไม่ใช้ค่าแก้จะมีลักษณะ Conservative เมื่อกลุ่มตัวอย่างมีขนาดเล็ก