บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัย

ในการศึกษาเปรียบเทียบการทดสอบความเป็นอิสระระหว่างตัวแปร ตั้งแต่ 2 ตัวแปรขึ้นไป ซึ่งมีข้อมูลที่ใช้ทดสอบเป็นข้อมูลจำนวนนับ อยู่ในรูปตารางการณ์จรจะใช้วิธีการวิเคราะห์เพื่อทดสอบความเป็นอิสระระหว่างตัวแปร 2 วิธีด้วยกันคือ

1. การทดสอบโดยใช้ตัวแบบลอกการิทึมเชิงเส้นตรง
2. การทดสอบแบบไคสแควร์

2.1 ตัวแบบลอกการิทึมเชิงเส้นตรง

การวิเคราะห์ข้อมูล โดยใช้ตัวแบบลอกการิทึมเชิงเส้นตรง เป็นการวิเคราะห์ข้อมูลจำนวนนับ หรือข้อมูลที่อยู่ในรูปความถี่ ซึ่งได้นำมาจัดลงในตารางการณ์จรได้ การวิเคราะห์โดยวิธีนี้ นอกจากจะเป็นการวิเคราะห์เพื่อทดสอบความเป็นอิสระระหว่างตัวแปร และใช้อธิบายถึงความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) หรือตัวแปรอธิบาย (Explanatory Variable) กับตัวแปรตาม (Dependent Variables) หรือตัวแปรตอบสนอง (Response Variables) ในรูปของตัวแบบเชิงเส้นตรง (Linear Model) คล้าย ๆ กับการวิเคราะห์ความแปรปรวน (Analysis of Variance) และการวิเคราะห์ความถดถอย (Regression Analysis) แล้วยังใช้เป็นการวิเคราะห์เพื่อประมาณค่าและทดสอบค่าของพารามิเตอร์ภายใต้ตัวแบบที่สร้างขึ้น ซึ่งจะทดสอบความเหมาะสมของตัวแบบนั้นโดยใช้ตัวสถิติไคสแควร์ (Chi-Square Statistic χ^2) หรือตัวสถิติไลลิฮูด (Likelihood Statistic G^2) เพื่อดูว่าตัวแบบใดบ้างที่สอดคล้องกับค่าสังเกต (observed values) ในตารางการณ์จรที่นำมาทดสอบทั้งหมด

2.1.1 ที่มาของตัวแบบลอกการิทึมเชิงเส้นตรง

ในการทดสอบเพื่อหาความเป็นอิสระระหว่างตัวแปร 2 ตัวจากตารางการณ์จร ก็คือ การหาว่าตัวแปรทั้งสองเป็นอิสระซึ่งกันและกัน (Mutually Independent) โดยการทดสอบว่า

$$p_{ij} = p_{i.}p_{.j} , \quad i = 1, 2, \ldots, r \quad \ldots\ldots\ldots\ldots\ldots(1)$$
$$j = 1, 2, \ldots, c$$

โดยที่ $p_{i.}$ หมายถึง ผลรวมของความน่าจะเป็น (Marginal probability) ในแถวที่ i ของตารางการณ์จร

$p_{.j}$ หมายถึง ผลรวมของความน่าจะเป็นในสดมภ์ที่ j ของตารางการณ์จร

และ p_{ij} หมายถึง ความน่าจะเป็นในเซลที่ (i, j)

แต่เนื่องจากผลบวกของความน่าจะเป็นทั้งหมดของทุกเซลในตารางการณ์จรรวมกันเท่ากับ 1 ดังนั้น สำหรับตารางการณ์จรขนาด 2 x 2 จำเป็นต้องทราบค่าของ $p_{1.}$ หรือ $p_{2.}$ และ $p_{.1}$ หรือ $p_{.2}$ เพื่อจะได้หาค่าของความน่าจะเป็นในแต่ละเซลของตารางนั้น ๆ ได้ แต่ในความเป็นจริงแล้ว ไม่สามารถทราบค่าความน่าจะเป็นของประชากร (population probability) ที่แท้จริงได้เลย ดังนั้น เราจะใช้ความถี่ของค่าสังเกต (observed cell frequency) ในแต่ละเซล มาประมาณค่าของมัน นั่นคือ จากสมการ (1) ซึ่งเป็นตัวแบบของความน่าจะเป็นของค่าสังเกตในเซล (i, j) ของตารางการณ์จรอยู่ในรูปผลคูณของผลรวมของความน่าจะเป็นของแถวที่ i กับผลรวมของความน่าจะเป็นของแถวที่ j ของตารางการณ์จรนั้น ดังนั้น ถ้าจัดรูปสมการเสียใหม่ให้อยู่ในรูปของผลบวกโดยการใส่ลอกการิทึมธรรมชาติ (Natural logarithm) ลงไปทั้งสองข้างของสมการดังกล่าว จะได้ตัวแบบใหม่เป็นดังนี้

$$\log_e p_{ij} = \log_e p_{i.} + \log_e p_{.j} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

หรือเขียนได้ใหม่ในเทอมของความถี่ทางทฤษฎี (Theoritical frequency) เป็น

$$\log_e F_{ij} = \log_e F_{i.} + \log_e F_{.j} - \log_e N \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

เมื่อ $F_{ij} = Np_{ij}$, $F_{i.} = Np_{i.}$, $F_{.j} = Np_{.j}$

และ N = จำนวนข้อมูลหรือความถี่ทั้งหมดในตารางนั้น

โดยการบวกทั้งสองข้างของสมการ (3) ตั้งแต่แถวที่ 1 ถึงแถวที่ r จะได้

$$\sum_{i=1}^{r} \log_e F_{ij} = \sum_{i=1}^{r} \log_e F_{i.} + r\log_e F_{.j} - r \log_e N \quad \ldots\ldots\ldots\ldots\ldots(4)$$

โดยการบวกทั้งสองข้างของสมการ (3) ตั้งแต่สดมภ์ที่ 1 ถึงสดมภ์ที่ c จะได้

$$\sum_{j=1}^{c} \log_e F_{ij} = c \log_e F_{i.} + \sum_{j=1}^{c} \log_e F_{.j} - c \log_e N \quad \ldots\ldots\ldots\ldots\ldots(5)$$

โดยการบวกทั้งสองข้างของสมการ (3) ตั้งแต่แถวที่ 1 ถึงแถวที่ r และสดมภ์ที่ 1 ถึงสดมภ์ที่ c จะได้

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \log_e F_{ij} = c \sum_{i=1}^{r} \log_e F_{i.} + r \sum_{j=1}^{c} \log_e F_{.j} - rc \log_e N \quad \ldots\ldots(6)$$

ดังนั้น สมการ (3) จะเขียนได้ใหม่เป็น

$$\log_e F_{ij} = u + u_{1(i)} + u_{2(j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

โดยที่ $u = \dfrac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \log_e F_{ij}$

$$u_{1(i)} = \frac{1}{c} \sum_{j=1}^{c} \log_e F_{ij} - \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \log_e F_{ij}$$

และ $u_2(j) = \dfrac{1}{r} \sum_{i=1}^{r} \log F_{ij} - \dfrac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \log_e F_{ij}$

จากสมการ (7) ถ้าให้ $\log_e F_{ij} = v_{ij}$ จะเขียนได้ใหม่เป็น

$$v_{ij} = u + u_{1(i)} + u_{2(j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

โดยที่ $u = \dfrac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} v_{ij} = v_{..}$ $\quad \ldots\ldots\ldots\ldots\ldots(9)$

$$u_{1(i)} = \frac{1}{c} \sum_{j=1}^{c} v_{ij} - \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} v_{ij} = \frac{1}{c} \sum_{j=1}^{c} v_{ij} - v_{..}$$

ดังนั้น $u_{1(i)} = v_{i.} - v_{..}$ $\quad \ldots\ldots\ldots\ldots\ldots(10)$

และ $u_{2(j)} = \dfrac{1}{r} \sum_{i=1}^{r} v_{ij} - \dfrac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} v_{ij} = \dfrac{1}{r} \sum_{i=1}^{r} v_{ij} - v_{..}$

ดังนั้น $u_{2(j)} = v_{.j} - v_{..}$ $\quad \ldots\ldots\ldots\ldots\ldots(11)$

สมการ (8) จะอยู่ในรูปของตัวแบบเชิงเส้นตรงของลอกการิทึม (Linear Model of Logarithm) ของความถี่หรือค่าสังเกตซึ่งจะเรียกชื่อได้อีกอย่างหนึ่งว่า ตัวแบบลอกการิ-ทึมเชิงเส้นตรง ซึ่งถ้าพิจารณาเฉพาะตัวแบบโดยไม่พิจารณาถึงวิธีการวิเคราะห์ จะมีรูปแบบคล้าย ๆ กับตัวแบบที่ใช้กันอยู่ ในการวิเคราะห์ความแปรปรวนแบบสองทาง (Two Way ANOVA หรือ Randomized Blocks Design, RBD) ซึ่งมีตัวแบบเป็นดังนี้

$$X_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, 2, \ldots, t$$
$$j = 1, 2, \ldots, r$$

โดยที่ X_{ij} หมายถึง ค่าสังเกตจากหน่วยทดลองในบล็อกที่ j ได้รับทรีทเมนท์ที่ i

μ หมายถึง ค่าเฉลี่ยของประชากร

τ_i หมายถึง อิทธิพลเนื่องจากทรีทเมนท์ที่ i

β_j หมายถึง อิทธิพลเนื่องจากบล็อกที่ j

ε_{ij} หมายถึง ค่าความคลาดเคลื่อนของการทดลองจากหน่วยทดลองที่ (i, j)

นอกจากนี้ยังมีรูปแบบคล้ายกับตัวแบบถดถอย ซึ่งมีลักษณะดังนี้

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

เป็นการถดถอยพหุคูณ (Multiple Regression) ที่มีตัวแปรอิสระ 2 ตัว

โดยที่ Y แทน ตัวแปรตาม

X แทน ตัวแปรอิสระ

α แทน ระยะตัดแกน Y (Y - intercept)

β แทน ความชัน (slope) ของเส้นถดถอย

และ ε แทน ความคลาดเคลื่อนระหว่างค่า Y ที่แท้จริงกับค่าประมาณ $\hat{Y}$

โดยที่ $\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$

a แทน ค่าประมาณของพารามิเตอร์ α

b แทน ค่าประมาณของพารามิเตอร์ β

ซึ่งตัวแบบถดถอยข้างต้นเป็นตัวแบบที่ใช้ทำนายค่า Y เมื่อทราบค่า X เช่นเดียวกับสมการ (7) และ (8) ซึ่งใช้ทำนายค่าของ $\log_e F_{ij}$ หรือ v_{ij} เมื่อทราบค่า u, $u_{1(i)}$ และ $u_{2(j)}$

โดยที่ u หมายถึง ค่าเฉลี่ยทั้งหมดของลอกการิทึมของความถี่ (overall mean effect)

$u_{1(i)}$ หมายถึง อิทธิพลสำคัญของระดับที่ i (main effect of i^{th} category) ของตัวแปรที่ 1

$u_{2(j)}$ หมายถึง อิทธิพลสำคัญของระดับที่ j (main effect of j^{th} category) ของตัวแปรที่ 2

และจากสมการ (10) และ (11) จะเห็นว่า $u_{1(i)}$, $u_{2(j)}$ เป็นค่าเบี่ยงเบนจากค่าเฉลี่ย u ของแต่ละแถว และแต่ละสดมภ์

หรือ $$u + u_{1(i)} = \frac{1}{c}\sum_{j=1}^{c}\log_e F_{ij}$$

= ค่าเฉลี่ยของลอกการิทึมของความถี่ของ c เซล ในระดับที่ i ของตัวแปรตัวที่ 1

และ $$u + u_{2(j)} = \frac{1}{r}\sum_{i=1}^{r}\log_e F_{ij}$$

= ค่าเฉลี่ยของลอกการิทึมของความถี่ของ r เซล ในระดับที่ j ของตัวแปรตัวที่ 2

ดังนั้น ผลที่ได้ตามมาคือ

$$* \quad \sum_{i=1}^{r} u_{1(i)} = \sum_{j=1}^{c} u_{2(j)} = 0$$

จากสมการ (7) จะเห็นว่าค่าต่าง ๆ ในตัวแบบที่ได้เป็นค่าที่อยู่ในเทอมของความถี่ทางทฤษฎี ซึ่งในทางปฏิบัติไม่สามารถหาค่าพารามิเตอร์ u, $u_{1(i)}$ และ $u_{2(j)}$ ได้โดยตรง เพราะไม่ทราบค่าของ p_{ij} หรือค่าของ F_{ij} ดังนั้น จึงต้องประมาณค่าของมันด้วย $\hat{u}$, $\hat{u}_{1(i)}$ และ $\hat{u}_{2(j)}$ โดยการใช้ค่าคาดหวังของค่าสังเกตในแต่ละเซลภายใต้ตัวแบบที่กำหนด และประมาณค่า v ด้วย y ในสมการ (8) ข้างต้น

$y_{i.} - y_{..} = \frac{y_{11}+y_{12}}{2} - \frac{y_{11}+y_{12}+y_{21}+y_{22}}{4} = \frac{2y_{11}+2y_{12}-y_{11}-y_{12}-y_{21}-y_{22}}{4}$

$= \frac{y_{11}+y_{12}-y_{21}-y_{22}}{4}$

ดังนั้น สำหรับตารางการณ์จรขนาด 2 x 2 จะได้เป็นดังนี้

$$\left.\begin{aligned} \hat{u} &= y_{..} = \frac{y_{11}+y_{12}+y_{21}+y_{22}}{4} \\ \hat{u}_{1(i)} &= y_{i.} - y_{..} = \frac{y_{11}+y_{12}-y_{21}-y_{22}}{4} \\ \text{และ}\quad \hat{u}_{2(j)} &= y_{.j} - y_{..} = \frac{y_{11}-y_{12}+y_{21}-y_{22}}{4} \end{aligned}\right\} \quad \ldots\ldots\ldots(12)$$

$y_{ij} = v_{ij} = \log_e F_{ij}$

ในเมื่อ $y_{ij} = \log_e \hat{m}_{ij}$

และ m_{ij} = ค่าคาดหวังของค่าสังเกตในเซล (i, j)

นั่นคือ ตัวแบบที่ใช้ในการทำนายเป็นดังนี้

$$y_{ij} = \hat{u} + \hat{u}_{1(i)} + \hat{u}_{2(j)} \quad \ldots\ldots\ldots\ldots\ldots(13)$$

$v_{ij} = u + u_{1(i)} + u_{2(j)}$

แต่ถ้าความสัมพันธ์ระหว่าง 2 ตัวแปรนั้นไม่เป็นอิสระกัน ก็จะมีปฏิกริยาร่วม (interaction) ระหว่างสองตัวแปรเข้ามาเกี่ยวข้องกับตัวแบบในสมการ (7) ข้างต้นด้วย คือ

$$\log_e F_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad \ldots\ldots\ldots\ldots(14)$$

หรือ $v_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad \ldots\ldots\ldots\ldots(15)$

โดยที่ $u_{12(ij)}$ หมายถึง ปฏิกริยาร่วมระหว่างระดับที่ i และ j ของตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2

และ $\sum_{i=1}^{r} u_{12(ij)} = \sum_{j=1}^{c} u_{12(ij)} = 0, \quad i = 1,2,\ldots,c$

$j = 1,2,\ldots,r$

ซึ่งสามารถแทนค่าของ v ด้วย y ดังนั้น จะได้ค่าประมาณ

$$\left.\begin{aligned} \hat{u} &= y_{..} \\ \hat{u}_{1(i)} &= y_{i.} - y_{..} \\ \hat{u}_{2(j)} &= y_{.j} - y_{..} \\ \text{และ}\quad \hat{u}_{12(ij)} &= y_{ij} - y_{i.} - y_{.j} + y_{..} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(16)$$

จากสมการ (15) ถ้าพิจารณาดูจะเห็นว่าตัวแบบดังกล่าวคล้ายกับตัวแบบของการวิเคราะห์ความแปรปรวนที่อยู่ในรูป

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

โดย α_i = อิทธิพลของระดับที่ i ของแฟคเตอร์ที่ 1

β_j = อิทธิพลของระดับที่ j ของแฟคเตอร์ที่ 2

$(\alpha\beta)_{ij}$ = อิทธิพลร่วมของระดับที่ i ของแฟคเตอร์ที่ 1 กับระดับที่ j ของแฟคเตอร์ที่ 2

ε_{ijk} = ความคลาดเคลื่อนของข้อมูลในหน่วยทดลองที่ (i, j) ซ้ำที่ k ที่มีลักษณะการแจกแจงแบบโค้งปกติมีค่าเฉลี่ยเป็นศูนย์ และค่าความแปรปรวนร่วมเป็น σ^2

ภายใต้ตัวแบบซึ่งมีค่าพารามิเตอร์ที่แตกต่างกัน จะเห็นว่า ตัวแบบในสมการ (14) หรือ (15) เป็นตัวแบบที่เหมาะสม (fit) กับข้อมูลได้อย่างสมบูรณ์ เนื่องจากค่าคาดหวังภายใต้ตัวแบบนี้ มีค่าเท่ากับความถี่ของค่าสังเกต ดังนั้น ตัวแบบที่มีจำนวนพารามิเตอร์ในตัวแบบเท่ากับจำนวนเซลในตารางการณ์จร ซึ่งเป็นตัวแบบที่มีพารามิเตอร์มากที่สุดซึ่งจะเรียกว่าเป็น ตัวแบบอิ่มตัว (Saturated Model)

$u_{1(1)}, u_{2(1)}$
$u_{1(2)}, u_{2(\ldots)}$

ดังนั้น ตัวแบบอิ่มตัวสำหรับตารางขนาด 2 x 2 ซึ่งมีตัวแปร 2 ตัว แต่ละตัวแปรแยกออกได้เป็น 2 ระดับ จะมีพารามิเตอร์ทั้งหมด 4 ตัวคือ u, $u_{1(1)}$, $u_{2(1)}$, $u_{12(11)}$ หรือ u, $u_{1(1)}$, $u_{2(2)}$, $u_{12(12)}$ หรือ u, $u_{1(2)}$, $u_{2(1)}$, $u_{12(21)}$ หรือ u, $u_{1(2)}$, $u_{2(2)}$, $u_{12(22)}$ โดยมีข้อจำกัดของค่าของ u ดังนี้

จาก $\sum_i u_{1(i)} = \sum_j u_{2(j)} = 0$

และ $\sum_i u_{12(ij)} = \sum_j u_{12(ij)} = 0$

จะได้ $u_{1(2)} = -u_{1(1)}$

$u_{2(2)} = -u_{2(1)}$

$u_{12(22)} = -u_{12(12)} = -u_{12(21)} = u_{12(11)}$

ส่วนตัวแบบอื่น ๆ ของตารางขนาด 2 x 2 ได้แก่

$= \log p_{ij}$ $y_{ij} = \log m_{ij}$

1) $v_{ij} = u + u_{1(i)}$ ....(17)

เมื่อมีสมมติฐานว่า ค่าความน่าจะเป็นของแต่ละระดับของตัวแปรตัวที่ 2 เท่ากัน (equally probable) นั่นคือ ไม่มีอิทธิพลเนื่องจากตัวแปรตัวที่ 2

2) $v_{ij} = u + u_{2(j)}$ ....(18)

เมื่อมีสมมติฐานว่า ค่าความน่าจะเป็นของแต่ละระดับของตัวแปรตัวที่ 1 เท่ากัน นั่นคือ ไม่มีอิทธิพลเนื่องจากตัวแปรตัวที่ 1

3) $v_{ij} = u$ ....(19)

เมื่อมีสมมติฐานว่า ความน่าจะเป็นของแต่ละระดับของตัวแปรทั้งสองเท่ากัน (equiprobable) นั่นคือ ไม่มีอิทธิพลเนื่องจากตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2

2.1.2 การประมาณค่าคาดหวังของความถี่ภายใต้ตัวแบบต่าง ๆ ของตารางขนาด 2 มิติ (Two-Dimensional Tables)

1) ตัวแบบอิ่มตัว ประมาณค่าคาดหวังของความถี่ได้จาก

$$\hat{m}_{ij} = x_{ij} , \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, m$$

โดยที่ x_{ij} หมายถึง ความถี่ที่สังเกตมาได้จากเซล (i, j)

n หมายถึง จำนวนแถวของตารางการณ์จร

และ m หมายถึง จำนวนสดมภ์ของตารางการณ์จร

2) ตัวแบบแห่งความเป็นอิสระ (Independence Model) หรือตัวแบบที่ไม่มีอิทธิพลร่วม หรือปฏิกริยาร่วม (No treatment effect model) นั่นคือ ตัวแบบซึ่ง $u_{12(ij)} = 0$ หรือตัวแบบซึ่งมีสมมติฐานหลักที่ต้องการทดสอบคือ

$H_0 : u_{12(ij)} = 0$ หรือเขียนสั้น ๆ ว่า $H_0 : u_{12} = 0$

ประมาณค่าคาดหวังของความถี่ได้จาก

$$\hat{m}_{ij} = \frac{x_{i.}\ x_{.j}}{N}, \quad i = 1, 2, \ldots, n$$
$$j = 1, 2, \ldots, m$$

โดยที่ $x_{i.}$ หมายถึง จำนวนความถี่ที่สังเกตได้ทั้งหมดในแถวที่ i

$x_{.j}$ หมายถึง จำนวนความถี่ที่สังเกตได้ทั้งหมดในสดมภ์ที่ j

และ N หมายถึง จำนวนข้อมูลทั้งหมด

3) ตัวแบบซึ่งไม่มีอิทธิพลเนื่องจากสดมภ์ (No Column Effect Model) เป็นตัวแบบซึ่งไม่มีอิทธิพลเนื่องจากตัวแปรทางสดมภ์หรือตัวแปรตัวที่ 2 นั่นคือ ตัวแบบซึ่ง $u_{2(j)} = 0$ หรือตัวแบบซึ่งมีสมมติฐานหลักที่ต้องการทดสอบ คือ

$H_0 : u_{2(j)} = 0$ หรือเขียนสั้น ๆ ว่า $H_0 : u_2 = 0$

ประมาณค่าคาดหวังของความถี่ได้จาก

$$\begin{array}{l}
\hat{m}_{11} = \hat{m}_{12} = \ldots = \hat{m}_{1n} = \hat{m}_{1.} = \frac{x_{1.}}{c} \\
\hat{m}_{21} = \hat{m}_{22} = \ldots = \hat{m}_{2n} = \hat{m}_{2.} = \frac{x_{2.}}{c} \\
\vdots \qquad \vdots \qquad \qquad \vdots \qquad \vdots \\
\hat{m}_{i1} = \hat{m}_{i2} = \ldots = \hat{m}_{in} = \hat{m}_{i.} = \frac{x_{i.}}{c} \qquad \therefore \hat{m}_{ij} = \frac{x_{i.}}{c} \\
\vdots \qquad \vdots \qquad \qquad \vdots \qquad \vdots \\
\hat{m}_{m1} = \hat{m}_{m2} = \ldots = \hat{m}_{mn} = \hat{m}_{m.} = \frac{x_{m.}}{c}
\end{array}$$

โดยที่ $\hat{m}_{1.}, \hat{m}_{2.}, \ldots, \hat{m}_{i.}, \ldots, \hat{m}_{m.}$ เป็นค่าเฉลี่ยของค่าสังเกตในแต่ละแถว

4) ตัวแบบซึ่งไม่มีอิทธิพลเนื่องจากแถว (No Row Effect Model) เป็นตัวแบบซึ่งไม่มีอิทธิพลเนื่องจากตัวแปรทางแถวหรือตัวแปรตัวที่ 1 นั่นคือ เป็นตัวแบบซึ่ง $u_{1(i)} = 0$ หรือตัวแบบซึ่งมีสมมติฐานหลักที่ต้องการทดสอบคือ $H_0 : u_{1(i)} = 0$ เขียนสั้น ๆ ว่า $H_o : u_1 = 0$

ประมาณค่าคาดหวังของความถี่ได้จาก

$$
\begin{array}{l}
\hat{m}_{11} = \hat{m}_{21} = \ldots = \hat{m}_{j1} = \ldots = \hat{m}_{m1} = \hat{m}_{.1} = \frac{x_{.1}}{r} \\
\hat{m}_{12} = \hat{m}_{22} = \ldots = \hat{m}_{j2} = \ldots = \hat{m}_{m2} = \hat{m}_{.2} = \frac{x_{.2}}{r} \\
\vdots \\
\hat{m}_{1i} = \hat{m}_{2i} = \ldots = \hat{m}_{ji} = \ldots = \hat{m}_{mi} = \hat{m}_{.i} = \frac{x_{.3}}{r} \qquad \hat{m}_{ij} = \frac{x_{.j}}{r} \\
\vdots \\
\hat{m}_{1n} = \hat{m}_{2n} = \ldots = \hat{m}_{.i} = \ldots = \hat{m}_{mn} = \hat{m}_{.n} = \frac{x_{.4}}{r}
\end{array}
$$

โดยที่ $\hat{m}_{.1}, \hat{m}_{.2}, \ldots, \hat{m}_{.i}, \ldots, \hat{m}_{.n}$ เป็นค่าเฉลี่ยของค่าสังเกตหรือความถี่ในแต่ละสดมภ์

5) ตัวแบบซึ่งไม่มีอิทธิพลใด ๆ (No Effect Model) หรือตัวแบบที่มีความน่าจะเป็นเท่ากันเป็นตัวแบบซึ่งไม่มีอิทธิพลเนื่องจากตัวแปรตัวที่ 1 ตัวแปรตัวที่ 2 และปฏิกริยาร่วมระหว่างตัวแปรทั้งสอง นั่นคือ ตัวแบบซึ่ง $u_{1(i)} = u_{2(j)} = u_{12(ij)} = 0$ หรือตัวแบบซึ่งมีสมมติฐานหลักที่ต้องการทดสอบคือ $H_0 : u_1 = u_2 = u_{12} = 0$

ประมาณค่าคาดหวังของความถี่ได้จาก

$$\hat{m}_{ij} = \frac{N}{mn} = \frac{\text{จำนวนข้อมูลทั้งหมด}}{\text{จำนวนเซลทั้งหมดในตาราง}}$$

2.1.3 การทดสอบความเป็นอิสระโดยใช้ตัวแบบลอกการิทึมเชิงเส้นตรงในตารางการณ์จรขนาด 2 มิติ

ในการทดสอบว่าข้อมูลแต่ละชุดในตารางการณ์จรขนาด i x j มีลักษณะเป็นไปตามข้อสมมติที่ว่า ตัวแปร 2 ตัวเป็นอิสระต่อกันหรือไม่มีอิทธิพลเนื่องจาก

ตัวแปรทั้งสองนั้นทดสอบได้โดย

1) คำนวณค่าประมาณของค่าคาดหวังของความถี่ภายใต้ตัวแบบ

แห่งความเป็นอิสระ $v_{ij} = u + u_{1(i)} + u_{2(j)}$

โดย $\hat{m}_{ij} = \frac{x_{i.}x_{.j}}{N}$, $i = 1, 2, \ldots, n$

$j = 1, 2, \ldots, m$

2) คำนวณค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2

โดยที่ $$\chi^2 = \sum_{i,j} \frac{(x_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

และ $$G^2 = 2 \sum_{i,j} x_{ij}(\log_e x_{ij} - \log_e \hat{m}_{ij})$$

ซึ่งทั้ง χ^2 และ G^2 ต่างก็มีการแจกแจงเข้าสู่การแจกแจงแบบไคสแควร์ (asymptotically distributed as χ^2) ซึ่งมีองศาแห่งความเป็นอิสระเท่ากับ $(n-1)(m-1)$

3) นำค่า χ^2 หรือ G^2 ที่คำนวณได้จากข้อ 2) ซึ่งมีองศาแห่งความเป็นอิสระเท่ากับ $(n-1)(m-1)$ มาเปรียบเทียบกับค่า χ^2 จากตารางการแจกแจงแบบไคสแควร์ ที่องศาแห่งความเป็นอิสระเดียวกัน

สมมติฐานหลักที่ต้องการทดสอบคือ

H_0 : ตัวแปรที่ 1 กับตัวแปรที่ 2 เป็นอิสระต่อกัน

หรือ $H_0 : u_{12(ij)} = 0$ ซึ่งเขียนสั้น ๆ เป็น $H_0 : u_{12} = 0$

ผลการทดสอบจะปฏิเสธสมมติฐานหลัก เมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่าค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ $(n-1)(m-1)$ ณ ระดับนัยสำคัญที่กำหนดไว้ คือ 0.200, 0.100, 0.050, 0.010 และ 0.005 ตามลำดับ ซึ่งแสดงว่า ตัวแบบดังกล่าวข้างต้น ซึ่งเป็นตัวแบบที่ทดสอบความเป็นอิสระของตัวแปรก็จะไม่สอดคล้องหรือเหมาะสมกับข้อมูลในตารางชุดนั้น ซึ่งโอกาสจะเป็นไปได้

ก็มีเพียงทางเดียวคือตัวแบบที่มีอิทธิพลเนื่องจากตัวแปรทั้งสองเข้ามาเกี่ยวข้องหรือสัมพันธ์กัน หรือที่เรียกว่าตัวแบบอิ่มตัว ซึ่งประมาณค่าพารามิเตอร์ได้ดังนี้

$$\hat{u} = y_{..}$$

$$\hat{u}_{1(i)} = y_{i.} - y_{..}$$

$$\hat{u}_{2(j)} = y_{.j} - y_{..}$$

$$\hat{u}_{12(ij)} = y_{ij} - y_{i.} - y_{.j} + y_{..}$$

โดยที่ $y_{ij} = \log_e \hat{m}_{ij}$

แต่ถ้าค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่าน้อยกว่าค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระ $(n-1)(m-1)$ ผลการทดสอบก็จะไม่สามารถปฏิเสธสมมติฐานหลักได้ นั่นคือ ตัวแบบดังกล่าวก็จะสอดคล้องหรือเหมาะสมกับข้อมูลชุดนั้น ซึ่งประมาณค่าพารามิเตอร์ได้ดังนี้

$$\hat{u} = y_{..}$$

$$\hat{u}_{1(i)} = y_{i.} - y_{..}$$

และ $$\hat{u}_{2(j)} = y_{.j} - y_{..}$$

โดยที่ $y_{ij} = \log_e \hat{m}_{ij}$

2.1.4 ตัวแบบลอกการิทึมเชิงเส้นตรงสำหรับตาราง 3 มิติ (I x J x K) ที่สำคัญมีดังนี้

1) ตัวแบบอิ่มตัว มีรูปแบบดังนี้

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

หรือ $$v_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

โดยที่ u หมายถึง ค่าเฉลี่ยทั้งหมด

$u_{1(i)}, u_{2(j)}, u_{3(k)}$ หมายถึง อิทธิพลสำคัญเนื่องจากตัวแปรที่ 1, 2 และ 3 ตามลำดับ

$u_{12(ij)}, u_{13(ik)}, u_{23(jk)}$ หมายถึง อิทธิพลร่วมของ 2 ตัวแปร (Interaction between two variables หรือ First order interaction effect หรือ Two-factor effect) 12, 13 และ 23 ตามลำดับ

และ $u_{123(ijk)}$ หมายถึง อิทธิพลร่วมของ 3 ตัวแปร (Second-order effect between the three variables หรือ Three-factor effect)

ซึ่งมีข้อจำกัดว่า

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \cdots = \sum_i u_{12(ij)} = \sum_j u_{12(ij)} = \cdots$$

$$\cdots = \sum_k u_{123(ijk)} = 0$$

ดังนั้น $$v_{\cdots} = \sum_i \sum_j \sum_k \frac{v_{ijk}}{IJK}$$ log P_{ijk}

โดยที่ $v_{\cdots}$ หมายถึง ค่าเฉลี่ยของลอกการิทึมของความน่าจะเป็นทั้งหมด

ซึ่งสามารถประมาณได้ด้วย $y_{\cdots} = \sum_i \sum_j \sum_k \frac{y_{ijk}}{IJK}$; $y_{ijk} = \log_e \hat{m}_{ijk}$

$$v_{i\cdot\cdot} = \sum_j \sum_k \frac{v_{ijk}}{JK}$$

โดยที่ $v_{i\cdot\cdot}$ หมายถึงค่าเฉลี่ยของลอกการิทึมของความน่าจะเป็นของระดับที่ i ของตัวแปรตัวที่ 1 ทั้งหมด

สามารถประมาณได้ด้วย $y_{i\cdot\cdot} = \sum_j \sum_k \frac{y_{ijk}}{JK}$

$u_{1(i)} = v_{i..} - v_{...}$ ประมาณได้ด้วย $\hat{u}_{1(i)} = y_{i..} - y_{...}$

$u_{2(j)} = v_{.j.} - v_{...}$ ประมาณได้ด้วย $\hat{u}_{2(j)} = y_{.j.} - y_{...}$

$u_{3(k)} = v_{..k} - v_{...}$ ประมาณได้ด้วย $\hat{u}_{3(k)} = y_{..k} - y_{...}$

ในทำนองเดียวกัน สามารถหาค่าของ u อื่น ๆ ได้ เช่นเดียวกัน เช่น

ถ้า $v_{ij.} = \sum_k \frac{v_{ijk}}{K}$, $v_{i.k} = \sum_j \frac{v_{ijk}}{J}$

ดังนั้น

$u_{12(ij)} = v_{ij.} - v_{i..} - v_{.j.} + v_{...}$ ประมาณได้ด้วย

$\hat{u}_{12(ij)} = y_{ij.} - y_{i..} - y_{.j.} + y_{...}$

$u_{123(ijk)} = v_{ijk} - v_{ij.} - v_{i.k} - v_{.jk} + v_{i..} + v_{.j.} + v_{..k} - v_{...}$

สามารถประมาณได้ด้วย

$\hat{u}_{123(ijk)} = y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...}$

และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก

$\hat{m}_{ijk} = x_{ijk}$

ซึ่งตัวแบบนี้จะสอดคล้องหรือ<u>เหมาะสมกับข้อมูลในตาราง 3 มิติทุกจุดที่องค่าแห่งความเป็นอิสระเท่ากับศูนย์</u> .

2) ตัวแบบซึ่งทั้ง 3 ตัวแปรเป็นอิสระต่อกัน (Model of complete independence หรือ Mutual independence of variable in a three way table) ซึ่งเป็นตัวแบบที่ไม่มีความสัมพันธ์ระหว่าง 3 ตัวแปรเลย มีรูปแบบดังนี้

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$$

หรือ $$v_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \quad \ldots\ldots\ldots(20)$$

และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก

$$\hat{m}_{ijk} = \left(\frac{x_{i..}}{N}\right)\left(\frac{x_{.j.}}{N}\right)\left(\frac{x_{..k}}{N}\right)N$$

$$= \frac{x_{i..}x_{.j.}\ x_{..k}}{N^2}$$

โดยมีสมมติฐานหลักที่ต้องการทดสอบคือ

$H_0 : u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = 0$; สำหรับทุก ๆ ค่า

ของ i, j และ k หรือเขียนสั้น ๆ ว่า

$H_0 : u_{12} = u_{13} = u_{23} = 0$

ผลการทดสอบจะปฏิเสธสมมติฐานหลักเมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่าค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ (IJK-I-J-K+2) ณ ระดับนัยสำคัญที่กำหนดไว้

แต่ถ้าค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่าน้อยกว่าหรือเท่ากับค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ (IJK-I-J-K+2) ณ ระดับนัยสำคัญที่กำหนดไว้ ก็จะไม่สามารถปฏิเสธสมมติฐานหลักนั้นได้ นั่นคือ ตัวแบบที่สอดคล้องหรือเหมาะสมกับข้อมูลชุดดังกล่าว ซึ่งก็คือยอมรับสมมติฐานหลักนั้น และสามารถประมาณค่าพารามิเตอร์ในตัวแบบนั้นได้ในลักษณะเดียวกับตัวแบบในข้อ 1) ข้างต้น

3) ตัวแบบซึ่งมีความเป็นอิสระอย่างมีเงื่อนไข (Conditional independence model) ซึ่งมีหลายรูปแบบ เช่น

ตัวแบบซึ่งมีตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 เป็นอิสระต่อกันเมื่อกำหนดตัวแปรตัวที่ 3 ให้คงที่ในระดับต่าง ๆ (conditional independence model of variable 1 and 2 given the level of variable 3) มีรูปแบบดังนี้ .

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$$

$$\text{หรือ}\ v_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(21)$$

และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก

$$\hat{m}_{ijk} = \frac{x_{i.k}x_{.jk}}{x_{..k}}$$

โดยมีสมมติฐานหลักที่ต้องการทดสอบคือ

$H_0 : u_{12(ij)} = 0,\ u_{123(ijk)} = 0$; สำหรับทุก ๆ ค่าของ i, j และ k หรือเขียนสั้น ๆ ว่า

$$H_0 : u_{12} = 0,\ u_{123} = 0$$

ผลการทดสอบจะปฏิเสธสมมติฐานหลัก เมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ $K(I-1)(J-1)$ ณ ระดับนัยสำคัญที่กำหนด
KIJ-JK-IK+K
ไว้

แต่ถ้าค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่าน้อยกว่าหรือเท่ากับ χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระ $K(I-1)(J-1)$ ณ ระดับนัยสำคัญที่กำหนด ก็จะไม่สามารถปฏิเสธสมมติฐานหลักได้ แสดงว่า ตัวแบบดังกล่าว เหมาะสมกับข้อมูลอยู่นั้นแล้ว ก็สามารถประมาณค่าพารามิเตอร์ในตัวแบบนั้นได้

4) ตัวแบบซึ่งมีเพียง 1 ผลของปฏิกิริยาร่วมของ 2 ตัวแปร (1 Two-factor effect) หรือตัวแบบแห่งความเป็นอิสระเพียงบางส่วน (Partial independence model) เป็นตัวแบบที่มีตัวแปรตัวหนึ่งเป็นอิสระจากอีก 2 ตัวแปร มีรูปแบบดังนี้

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$$

หรือ $v_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$ ....(22)

และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก

$$\hat{m}_{ijk} = \frac{x_{ij.}\,x_{..k}}{x_{...}}$$

$$= \frac{x_{ij.}\,x_{..k}}{N}$$

โดยมีสมมติฐานหลักที่ต้องการทดสอบคือ

$H_0 : u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0$ สำหรับทุก ๆ ค่าของ i, j และ k หรือเขียนสั้น ๆ ว่า

$$H_0 : u_{13} = u_{23} = u_{123} = 0$$

ผลการทดสอบจะปฏิเสธสมมติฐานหลักเมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ $(K-1)(IJ-1)$ ณ ระดับนัยสำคัญที่กำหนดไว้

IJK − IJ − K + 1

5) ตัวแบบภายใต้ข้อสมมติที่ว่า ไม่มีปฏิกริยาร่วมอันดับสอง (No second-order interaction) หรือไม่มีอิทธิพลร่วมของทั้ง 3 ตัวแปร มีรูปแบบดังนี้

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

หรือ $v_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$ ....(23)

และค่าประมาณของค่าคาดหวังของความถี่ไม่สามารถเขียนออกมาเป็นสูตรสำเร็จได้ แต่สามารถหาได้ โดยใช้วิธีทำซ้ำของค่าคาดหวัง (Iterative computation of expected values) ดังนี้

<u>ขั้นที่ 1</u> ให้ $\hat{m}_{ijk}^{(0)} = 1$ สำหรับทุก ๆค่าของ i,j,k ....(24)

เมื่อ $\nu = 0$ คำนวณ

<u>ขั้นที่ 2</u>

$$\hat{m}_{ijk}^{(3\nu+1)} = \frac{x_{ij.}}{\hat{m}_{ij.}^{(3\nu)}} \hat{m}_{ijk}^{(3\nu)} \qquad \ldots\ldots\ldots\ldots\ldots(25)$$

<u>ขั้นที่ 3</u>

$$\hat{m}_{ijk}^{(3\nu+2)} = \frac{x_{i.k}}{\hat{m}_{i.k}^{(3\nu+1)}} \hat{m}_{ijk}^{(3\nu+1)} \qquad \ldots\ldots\ldots\ldots\ldots(26)$$

<u>ขั้นที่ 4</u>

$$\hat{m}_{ijk}^{3(\nu+1)} = \frac{x_{.jk}}{\hat{m}_{.jk}^{(3\nu+2)}} \hat{m}_{ijk}^{(3\nu+2)} \qquad \ldots\ldots\ldots\ldots\ldots(27)$$

เมื่อคำนวณถึงขั้นที่ 4 แล้ว ก็เป็นอันว่าเสร็จรอบที่ 1 ของการทำซ้ำ (first cycle of the iteration) ของการคำนวณหาค่าประมาณของค่าคาดหวัง ซึ่งจะเห็นว่าในขั้นที่ 2 คำนวณได้ค่าของ $\hat{m}_{ijk}^{(1)}$

ขั้นที่ 3 คำนวณได้ค่าของ $\hat{m}_{ijk}^{(2)}$

ขั้นที่ 4 คำนวณได้ค่าของ $\hat{m}_{ijk}^{(3)}$

ขั้นต่อไปทำซ้ำจากขั้นที่ 2 ถึงขั้นที่ 4 ใหม่อีก โดยให้ $\nu = 1,2,\ldots$ จนกระทั่งค่า $\hat{m}_{ijk}^{(s)} = \hat{m}_{ijk}^{(s-1)}$ ในเมื่อ $s = 1, 2, \ldots$

นั่นคือ ค่า $\hat{m}_{ijk}^{(s)}$ มีค่าไม่เปลี่ยนแปลงต่อไปอีก หรือถ้าจะเปลี่ยนแปลงก็น้อยมาก <u>คือน้อยกว่า 0.01</u> ดังนั้น ค่าประมาณของค่าคาดหวังของความถี่สำหรับตัวแบบนี้ ก็คือ $\hat{m}_{ijk}^{(s)}$ นั่นเอง

สมมติฐานหลักที่ต้องการทดสอบคือ

$H_0 : u_{123(ijk)} = 0$ สำหรับทุก ๆ ค่าของ i, j, k

หรือเขียนสั้น ๆ ว่า $H_0 : u_{123} = 0$

ผลการทดสอบจะปฏิเสธสมมติฐานหลักเมื่อค่าสถิติไคสแควร์ X^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่า X^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ $(I-1)(J-1)(K-1)$ ณ ระดับนัยสำคัญที่กำหนดไว้

<u>หมายเหตุ</u> การใช้วิธีทำซ้ำนี้ จะใช้ได้กับตัวแบบต่าง ๆ ที่กล่าวมาแล้วก็ได้

2.1.5 ตัวแบบแสดงอันดับ (Hierarchical Model)

ตัวแบบใดจะเรียกว่าเป็น ตัวแบบแสดงอันดับ ก็ต่อเมื่อถ้ามีอิทธิพลเนื่องจากตัวแปรที่มีอันดับสูงกว่า (higher order effect) อยู่ในตัวแบบนั้นก็ จะต้องมีอิทธิพลเนื่องจากตัวแปรที่มีอันดับต่ำกว่า (lower order effect) ปรากฏอยู่ด้วยเสมอ

เช่น ถ้า $u_{123(ijk)}$ อยู่ในตัวแบบหนึ่ง จะต้องมีพจน์ของ $u_{12(ij)}$, $u_{13(ik)}$, $u_{23(jk)}$, $u_{1(i)}$, $u_{2(j)}$ และ $u_{3(k)}$ อยู่ในตัวแบบนั้นด้วย หรือ $u_{123(ijk)}$ ไม่สามารถอยู่ในตัวแบบนั้นได้ ถ้าหากไม่มี $u_{12(ij)}$, $u_{13(ik)}$ และ $u_{23(jk)}$ อยู่ในตัวแบบนั้น

ตัวอย่างเช่น $\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{123(ijk)}$ ไม่เป็นตัวแบบแสดงอันดับ

ดังนั้น ในการกล่าวถึงตัวแบบในตารางการณ์จร 4 มิติ และมากกว่าขึ้นไป ต่อไปนี้จะพิจารณาถึงตัวแบบแสดงอันดับด้วย

2.1.6 ตัวแบบลอกการิทึมเชิงเส้นตรงสำหรับตารางการณ์จร ขนาด 4 มิติ และมากกว่า

สำหรับตารางการณ์จรขนาด 4 มิติ (IxJxKxL Table) จะพิจารณาเพียงบางตัวแบบต่อไปนี้ คือ

1) ตัวแบบง่าย ๆ สำหรับตาราง 4 มิติ ก็คือ ตัวแบบที่ตัวแปรทั้ง 4 เป็นอิสระต่อกัน ซึ่งมีรูปแบบดังนี้

$$\log_e F_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)}$$

หรือ $$v_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} \quad \ldots\ldots(28)$$

โดยมีข้อจำกัดว่า

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_k u_{3(k)} = \sum_l u_{4(l)} = 0$$

ซึ่งสัญลักษณ์ย่อสำหรับตัวแบบนี้คือ [1] [2] [3] [4] และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก

$$\hat{m}_{ijkl} = \frac{x_{i\ldots}\, x_{.j..}\, x_{..k.}\, x_{...l}}{N^3} \qquad \left(\frac{x_{i\ldots}}{N}\,\frac{x_{.j..}}{N}\,\frac{x_{..k.}}{N}\,\frac{x_{...l}}{N}\right)N$$

โดยมีสมมติฐานหลัก ที่ต้องการทดสอบคือ

$$H_0 : u_{12(ij)} = u_{23(jk)} = u_{34(kl)} = u_{13(ik)} = u_{14(il)} = u_{24(jl)} = 0$$

สำหรับทุก ๆ ค่าของ i, j, k และ l

หรือเขียนสั้น ๆ ว่า $H_0 : u_{12} = u_{23} = u_{34} = u_{13} = u_{14} = u_{24} = 0$

ผลการทดสอบจะปฏิเสธสมมติฐานหลัก เมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่าค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ (IJKL-I-J-K-L+3) ณ ระดับนัยสำคัญที่กำหนดไว้

2) ตัวแบบที่มีตัวแปรเป็นอิสระต่อกันอย่างมีเงื่อนไขมีได้หลายรูปแบบมากมาย (โดยอาศัยข้อจำกัดของตัวแบบแสดงอันดับ เข้าช่วยในการพิจารณาด้วย) ซึ่งสำหรับตาราง 4 มิติจะมีได้ถึง 113 ตัวแบบแสดงอันดับที่แตกต่างกันและทุกตัวแบบประกอบด้วยอิทธิพลสำคัญทั้ง 4 ด้วย[1]

---

[1] Stephen E. Fienberg, The Analysis of Cross-Classified Categorical Data, 2nd ed. (combridge, Mass.: The MIT Press, 1980), p. 72.

พิจารณาตัวแบบ

$$\log_e F_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{23(jk)} + u_{34(kl)} + u_{14(il)} + u_{13(ik)} + u_{24(jl)} + u_{234(jkl)} + u_{134(ikl)} \quad \ldots\ldots\ldots\ldots\ldots(29)$$

ซึ่งเป็นตัวแบบที่มี $u_{12(ij)} = 0$ สำหรับทุก ๆ ค่าของ i และ j ดังนั้น จะได้ว่า $u_{123(ijk)} = u_{124(ijl)} = u_{1234(ijkl)} = 0$ ด้วย โดยมีข้อจำกัดว่า

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_k u_{3(k)} = \sum_l u_{4(l)} = 0$$

และ $\sum_j u_{23(jk)} = \sum_k u_{23(jk)} = \sum_i u_{14(il)} = \ldots = \sum_l u_{ikl(134)} = 0$

ซึ่งสัญลักษณ์ย่อ สำหรับตัวแบบนี้คือ [234][134] และค่าประมาณของค่าคาดหวังของความถี่หาได้จาก
minimal sufficient stat (Configuration)

$$\hat{m}_{ijkl} = \frac{x_{i.kl}\,x_{.jkl}}{x_{..kl}}$$

โดยมีสมมติฐานหลัก ที่ต้องการทดสอบคือ

$$H_0 : u_{12} = 0$$

ผลการทดสอบจะปฏิเสธสมมติฐานหลักเมื่อค่าสถิติไคสแควร์ χ^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่าค่า χ^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ KL (I-1)(J-1) ณ ระดับนัยสำคัญที่กำหนดไว้

3) ตัวแบบซึ่งประกอบด้วยอิทธิพลเนื่องจากตัวแปร 2 ตัว ซึ่งเป็นอิสระกันทั้งหมด โดยไม่มีอิทธิพลเนื่องจากตัวแปรที่มีระดับสูงกว่าเลย ซึ่งมีรูปแบบดังนี้

$$\log_e F_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} \quad \ldots\ldots(30)$$

โดยมีข้อจำกัดว่า

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_k u_{3(k)} = \sum_l u_{4(l)} = 0$$

และ $$\sum_i u_{12(ij)} = \sum_j u_{12(ij)} = \sum_i u_{13(ik)} = \cdots = \sum_l u_{34(kl)} = 0$$

ซึ่งสัญญลักษณ์ย่อสำหรับตัวแบบนี้คือ [12] [13] [14] [23] [24] [34] หรือ "all two way" และค่าประมาณของค่าคาดหวังของความถี่หาได้โดยใช้วิธีการทำซ้ำของค่าคาดหวัง โดยแต่ละรอบของการทำซ้ำจะมี 6 ขั้นตอนต่าง ๆ กัน สำหรับค่าของผลรวม (Marginal Totals) 6 ค่าที่แตกต่างกัน คือ $x_{ij..}$, $x_{i.k.}$, $x_{i..l}$, $x_{.jl.}$, $x_{.j.k}$ และ $x_{..kl}$ โดยมีสมมติฐานหลักที่ต้องการทดสอบคือ

$$H_0 : u_{123} = u_{134} = u_{234} = u_{124} = u_{1234} = 0$$

ผลการทดสอบจะปฏิเสธ สมมติฐานหลักเมื่อค่าสถิติไคสแควร์ X^2 หรือค่าสถิติไลลิฮูด G^2 จากการคำนวณมีค่ามากกว่าค่า X^2 จากตารางการแจกแจงไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ IJKL-IJ-IK-IL-JK-JL-KL + 2(I+J+K+L) -3 ณ ระดับนัยสำคัญที่กำหนดไว้

ในการวิจัยนี้จะแสดงเฉพาะโปรแกรมที่เขียนขึ้นเองเพื่อใช้ทดสอบข้อมูลภายใต้ตัวแบบนี้เท่านั้น (ดูในภาคผนวก ก) แต่ไม่ได้นำไปใช้ทดสอบกับข้อมูลตัวอย่างที่สมมติขึ้นทั้งหมด เพื่อเปรียบเทียบกับการทดสอบแบบไคสแควร์ เพราะในการทดสอบแบบไคสแควร์ไม่สามารถทดสอบความเป็นอิสระระหว่างตัวแปรหลาย ๆ ตัว พร้อมกันได้ โดยไม่ควบคุมให้ตัวแปรที่ 3 และตัวแปรที่เหลืออื่น ๆ คงที่

นอกจากนี้ ยังมีตัวแบบอื่น ๆ อีกมากมายหลายตัวแบบซึ่งแสดงไว้ในตารางที่ 2.1 เพียงบางตัวแบบเท่านั้น โดยแยกเป็น 2 ประเภท คือ ตัวแบบชนิดที่สามารถประมาณค่าคาดหวังของความถี่ได้โดยตรง และตัวแบบชนิดประมาณค่าคาดหวังของความถี่โดยอ้อม โดยวิธีการทำซ้ำของค่าคาดหวัง

ตารางที่ 2.1 แสดงตัวแบบลอกการิทึมเชิงเส้นตรงแบบต่าง ๆ โดยใช้สัญญลักษณ์ย่อ (Model Abbreviation) พร้อมทั้งองศาแห่งความเป็นอิสระ

ตัวแบบโดยใช้สัญญลักษณ์ย่อ	องศาแห่งความเป็นอิสระ
1. ตัวแบบซึ่งประมาณค่าคาดหวังโดยตรง	
$v_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)}$	
(1) [1][2][3][4]	IJKL-I-J-K-L+3
(2) [12][3][4]	IJKL-IJ-K-L+2
(3) [12][34]	(IJ-1)(KL-1)
(4) [12][23][4]	J(IKL-I-K+1)-L+1
(5) [12][23][34]	IJKL-IJ-KL-JK+J+K
(6) [12][13][14]	I(JKL-J-K-L+2)
(7) [123][4]	(IJK-1)(L-1)
(8) [123][34]	K(IJ-1)(L-1)
(9) [123][234]	JK(I-1)(L-1)
2. ตัวแบบซึ่งประมาณค่าคาดหวังโดยอ้อม	
(10) [12][13][23][4]	IJKL-IJ-JK-IK-L+I+J+K
(11) [12][13][23][34]	IJKL-IJ-JK-IK-KL+I+J+2K-1
(12) [12][13][23][24][34]	IJKL-IJ-JK-IK-JL-KL+I+2J+2K+L-2
(13) all two way	IJKL-IJ-JK-IK-IL-JK-JL-KL+2(I+J+K+L)-3
(14) [123][24][34]	IJKL-IJK-JL-KL+J+K+L-1
(15) [123][14][24][34]	IJKL-IJK-IL-JL-KL+I+J+K+2L-2
(16) [123][124][34]	(IJ-1)(K-1)(L-1)
(17) [123][124][234]	(IJ-J+1)(K-1)(L-1)
(18) all three way	(I-1)(J-1)(K-1)(L-1)

สำหรับตารางการณ์จรที่มีขนาดมากกว่า 4 มิติขึ้นไป ก็จะมีตัวแบบลอกการิทึมเชิงเส้นตรงในลักษณะที่คล้ายคลึงกับตัวแบบลอกการิทึมเชิงเส้นตรงขนาด 4 มิติ ซึ่งได้กล่าวมาแล้ว เพียงแต่ตัวแบบมีลักษณะที่ขยายใหญ่ขึ้น ประกอบด้วยจำนวนพารามิเตอร์มากขึ้น แต่การวิเคราะห์ก็ยังคงเป็นไปในลักษณะเดียวกันเพียงแต่ยุ่งยากซับซ้อนขึ้นกว่าเดิมเท่านั้น

2.2 การทดสอบแบบไคสแควร์

การทดสอบแบบไคสแควร์ เป็นการทดสอบสมมติฐาน โดยข้อมูลที่นำมาทดสอบอยู่ในรูปของข้อมูลจำนวนนับ หรือข้อมูลที่อยู่ในรูปของความถี่ซึ่งได้มาจากตัวแปรเชิงคุณลักษณะหรือแบ่งประเภท โดยสามารถนำไปจัดลงในตารางการณ์จรได้ การทดสอบโดยวิธีนี้ นอกจากจะเป็นการทดสอบความเป็นอิสระระหว่างตัวแปรแล้ว ยังใช้ในการทดสอบว่าการแจกแจงของประชากรที่สนใจศึกษาเป็นไปตามลักษณะ การแจกแจงที่คาดไว้หรือไม่ ซึ่งเรียกว่า การทดสอบความเหมาะสมของตัวแบบหรือการทดสอบภาวะสารูปสนิทดี

2.2.1 การทดสอบความเป็นอิสระระหว่างตัวแปร 2 ตัว

ในการทดสอบเพื่อความเป็นอิสระระหว่างตัวแปร 2 ตัวจากตารางการณ์จร ก็คือการหาว่าตัวแปรทั้งสองเป็นอิสระซึ่งกันและกัน โดยการตั้งสมมติฐานหลักเพื่อการทดสอบว่า

H_0 : ตัวแปรทั้งสองเป็นอิสระต่อกัน

H_1 : ตัวแปรทั้งสองไม่เป็นอิสระต่อกัน

ตัวสถิติที่ใช้ในการทดสอบคือ

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

เมื่อ O_{ij} = ความถี่ของข้อมูลที่ได้จากการสังเกตหรือทดลองของระดับที่ i ของตัวแปรตัวที่ 1 และระดับที่ j ของตัวแปรตัวที่ 2

E_{ij} = ความถี่ที่คาดหวังของระดับที่ i ของตัวแปรตัวที่ 1 และระดับที่ j ของตัวแปรตัวที่ 2 ภายใต้ $H_0 : E_{ij} = \frac{X_{i.} X_{.j}}{X_{..}}$

$X_{i.}$ = ความถี่รวมของระดับที่ i ของตัวแปรตัวที่ 1 และทุกระดับของตัวแปรตัวที่ 2

$X_{.j}$ = ความถี่รวมของระดับที่ j ของตัวแปรตัวที่ 2 และทุกระดับของตัวแปรตัวที่ 1

$X_{..}$ = ความถี่รวมทั้งหมด หรือจำนวนข้อมูลทั้งหมด

r = จำนวนแถวของตัวแปรตัวที่ 1

c = จำนวนสดมภ์ของตัวแปรตัวที่ 2

ปฏิเสธสมมติฐานหลัก H_0 ถ้าค่า χ^2 ที่ได้จากการคำนวณมีค่ามากกว่า χ^2 จากตารางการแจกแจงแบบไคสแควร์ที่องศาแห่งความเป็นอิสระเท่ากับ (r-1)(c-1) ณ ระดับนัยสำคัญที่กำหนดไว้

ในกรณีที่แต่ละตัวแปรที่นำมาทดสอบเป็นตัวแปรชนิดที่เป็นแบบที่แบ่งเป็น 2 ลักษณะเท่านั้น (Dichotomous Variable) นั่นคือ ข้อมูลอยู่ในรูปตารางการณ์จรขนาด 2 x 2 และค่าความถี่ที่คาดหวังในแต่ละเซล มีบางตัวมีความถี่ไม่ถึง 20 จะใช้วิธีการคำนวณค่า χ^2 ด้วยสูตรของการปรับแก้ของเยทส์ (Yates' Correction)

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$$

ปฏิเสธสมมติฐานหลัก H_0 เมื่อค่า χ^2 ที่คำนวณได้มากกว่าค่า χ^2 จากตารางการแจกแจงแบบไคสแควร์ ที่องศาแห่งความเป็นอิสระเท่ากับ 1 ณ ระดับนัยสำคัญที่กำหนดไว้

2.2.2 การทดสอบความเป็นอิสระระหว่างตัวแปร 3 ตัวและมากกว่า

ในการทดสอบเพื่อหาความเป็นอิสระระหว่างตัวแปรหลาย ๆ ตัว คือตั้งแต่ 3 ตัวแปรขึ้นไปนั้น จะทำการทดสอบความเป็นอิสระระหว่างตัวแปรทีละคู่ โดยควบคุมให้ตัวแปรที่เหลือคงที่ ณ ที่ระดับต่าง ๆ ทั้งหมดของตัวแปรนั้น ตัวอย่าง เช่น จะทดสอบว่าตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 เป็นอิสระต่อกัน เมื่อกำหนดตัวแปรตัวที่ 3 ให้คงที่ ที่ระดับต่าง ๆ ก็จะทำการทดสอบโดยทดสอบความเป็นอิสระระหว่างตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 โดยการทดสอบไคสแควร์ทีละระดับ หรือกลุ่มย่อยของตัวแปรตัวที่ 3 ทำไปจนครบทุกระดับของตัวแปรตัวที่ 3

ผลการทดสอบจะปฏิเสธสมมติฐานหลักที่ว่า ตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 เป็นอิสระต่อกัน เมื่อกำหนดตัวแปรตัวที่ 3 ให้คงที่ในระดับต่าง ๆ เมื่อมีการปฏิเสธสมมติฐานของการทดสอบสมมติฐาน ที่ว่า ตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 เป็นอิสระต่อกันที่ระดับใดระดับหนึ่งของตัวแปรตัวที่ 3 (เพียงระดับเดียวเท่านั้น) และ ผลการทดสอบจะไม่สามารถปฏิเสธสมมติฐานหลัก หรือยอมรับสมมติฐานหลัก เมื่อทดสอบสมมติฐานแล้วได้ว่าตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 เป็นอิสระต่อกันที่ทุก ๆ ระดับของตัวแปรตัวที่ 3

สำหรับการทดสอบความเป็นอิสระระหว่างตัวแปรที่มากกว่า 3 ตัวแปรขึ้นไป ก็ทดสอบได้ในทำนองเดียวกัน โดยทดสอบความเป็นอิสระระหว่างตัวแปรทีละคู่ เมื่อควบคุมให้ตัวแปรที่เหลืออยู่คงที่ ณ ระดับต่าง ๆ ทั้งหมดของตัวแปรที่เหลืออยู่นั้น

2.3 การทดสอบสัดส่วนของผลการทดสอบที่ตรงกัน

หลังจากที่ได้ทดสอบความเป็นอิสระระหว่างตัวแปร โดยใช้วิธีการทดสอบด้วยตัวแบบลอกการิทึมเชิงเส้นตรงและการทดสอบแบบไคสแควร์แล้ว นำผลจากการทดสอบมาเปรียบเทียบกัน หาผลการทดสอบที่ตรงกันและไม่ตรงกัน แล้วนำผลดังกล่าวมาทดสอบสัดส่วนของผลการทดสอบที่ตรงกันของการทดสอบทั้งสองวิธีนี้ ว่า ให้ผลการทดสอบที่ตรงกันด้วยสัดส่วนเท่าใด

การทดสอบสัดส่วนของผลการทดสอบที่ตรงกันนี้ใช้ทดสอบโดยการทดสอบแบบไคสแควร์ ดังนี้

ให้ P แทน สัดส่วนของผลการทดสอบที่ตรงกันของประชากร เมื่อใช้วิธีทดสอบโดยใช้ตัวแบบลอกการิทึมเชิงเส้นตรงและการทดสอบแบบไคสแควร์

และ P_0 แทน สัดส่วนของผลการทดสอบที่คาดว่าตรงกัน เมื่อใช้วิธีทดสอบโดยใช้ตัวแบบลอกการิทึมเชิงเส้นตรงและการทดสอบแบบไคสแควร์

ทดสอบสมมติฐานหลัก $H_0 : P = P_0$

$H_1 : P \neq P_0$

ตัวสถิติในการทดสอบ คือ

$$\chi^2 = \sum_{i=1}^{2} \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

โดยที่ O_i = จำนวนผลการทดสอบ เมื่อใช้วิธีทั้งสองทดสอบโดยที่

O_1 แทน จำนวนผลการทดสอบที่ตรงกัน และ

O_2 แทน จำนวนผลการทดสอบที่ไม่ตรงกัน

E_i = จำนวนผลการทดสอบที่คาดว่าจะเป็น เมื่อใช้วิธีทั้งสองทดสอบ

โดยที่ E_1 แทน จำนวนผลการทดสอบที่คาดว่าจะตรงกัน

และ E_2 แทนจำนวนผลการทดสอบที่คาดว่าจะไม่ตรงกัน

ปฏิเสธสมมติฐานหลัก H_0 เมื่อค่า χ^2 ที่คำนวณได้มากกว่าค่า χ^2 จากตารางการแจกแจงแบบไคสแควร์ที่องศาแห่งความเป็นอิสระ เท่ากับ 1 ณ ระดับนัยสำคัญ 0.20, 0.10 และ 0.05 ตามลำดับ