<u>Chi-Squared Test</u>

1. ตั้งสมมติฐาน

    H_0 : ไม่มีความสัมพันธ์กันระหว่างกลุ่มข้อมูล

    H_1 : มีความสัมพันธ์กันระหว่างกลุ่มข้อมูล

2. คำนวณ expected frequencies

$$E_{ij} = \frac{(\text{sum of row } i) \cdot (\text{sum of column } j)}{\text{sum of all sample}}$$

    โดยที่ m, n คือจำนวนประเภทของข้อมูลในกลุ่มข้อมูลนั้นๆ

3. คำนวณค่าทางสถิติของ chi-squared test

$$\mathcal{X}^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

    โดยที่ m, n คือจำนวนประเภทของข้อมูลในกลุ่มข้อมูลนั้นๆ

4. คำนวณค่า degree of freedom

$$\text{Degrees of Freedom} = (m-1) \times (n-1)$$

5. คำนวณค่าทางสถิติของ chi-squared test ที่จุด critical point (α= significant level)
6. สรุปผลโดยพิจารณาว่า
    - ถ้าค่าทางสถิติของ chi-squared test จากข้อ 3 $>$ ค่าทางสถิติของ chi-squared test ที่จุด critical point จากข้อ 4: ปฏิเสธ H0 นั่นหมายความว่า มีความสัมพันธ์กันระหว่างกลุ่มข้อมูล
    - ถ้าค่าทางสถิติของ chi-squared test จากข้อ 3 $\leq$ ค่าทางสถิติของ chi-squared test ที่จุด critical point จากข้อ 4: ไม่ปฏิเสธ H0 นั่นหมายความว่า ไม่มีความสัมพันธ์กันระหว่างกลุ่มข้อมูล

ตัวอย่างการคำนวณ

กลุ่มข้อมูล 1: กลุ่มของอายุ (Children, Adult)

กลุ่มข้อมูล 2: รสชาติของไอศกรีม (Chocolate, Vanilla, Strawberry, Mint)

	Chocolate	Vanilla	Strawberry	Mint	total
Children	30	20	10	10	70
Adult	15	25	20	10	70
total	45	45	30	20	140

1. ตั้งสมมติฐาน

    H_0 : ไม่มีความสัมพันธ์กันระหว่างกลุ่มของอายุและรสชาติไอศกรีม

    H_1 : มีความสัมพันธ์กันระหว่างกลุ่มของอายุและรสชาติไอศกรีม

2. คำนวณ expected frequencies

$$E_{11} = \frac{(\text{sum of row 1})\cdot(\text{sum of column 1})}{\text{sum of all sample}} = \frac{70\cdot45}{140} = 22.5$$

$$E_{12} = \frac{(\text{sum of row 1})\cdot(\text{sum of column 2})}{\text{sum of all sample}} = \frac{70\cdot45}{140} = 22.5$$

$$E_{13} = \frac{(\text{sum of row 1})\cdot(\text{sum of column 3})}{\text{sum of all sample}} = \frac{70\cdot30}{140} = 15$$

$$E_{14} = \frac{(\text{sum of row 1})\cdot(\text{sum of column 4})}{\text{sum of all sample}} = \frac{70\cdot20}{140} = 10$$

$$E_{21} = \frac{(\text{sum of row 2})\cdot(\text{sum of column 1})}{\text{sum of all sample}} = \frac{70\cdot45}{140} = 22.5$$

$$E_{22} = \frac{(\text{sum of row 2})\cdot(\text{sum of column 2})}{\text{sum of all sample}} = \frac{70\cdot45}{140} = 22.5$$

$$E_{23} = \frac{(\text{sum of row 2})\cdot(\text{sum of column 3})}{\text{sum of all sample}} = \frac{70\cdot30}{140} = 15$$

$$E_{24} = \frac{(\text{sum of row 2})\cdot(\text{sum of column 4})}{\text{sum of all sample}} = \frac{70\cdot20}{140} = 10$$

3. คำนวณค่าทางสถิติของ chi-squared test

$$\mathcal{X}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

$$= \frac{(O_{11}-E_{11})^2}{E_{11}} + \frac{(O_{12}-E_{12})^2}{E_{12}} + \cdots + \frac{(O_{24}-E_{24})^2}{E_{24}}$$

$$= \frac{(30-22.5)^2}{22.5} + \frac{(20-22.5)^2}{22.5} + \cdots + \frac{(10-10)^2}{10}$$

$$= 2.5 + 0.278 + 1.667 + 0 + 2.5 + 0.278 + 1.667 + 0$$

$$= 8.89$$

4. คำนวณค่า degree of freedom

$$\text{Degrees of Freedom(df)} = (2-1)\times(4-1) = 3$$

5. คำนวณค่าทางสถิติของ chi-squared test ที่จุด critical point (α= significant level)

กำหนดให้ $\alpha = 0.05$ และค่า $\text{df} = 3$

<u>ดังนั้น</u> ค่าทางสถิติของ chi-squared test ที่จุด critical point: $\mathcal{X}_c^2 = 7.815$

6. สรุปผลได้ว่า

$$\mathcal{X}^2 = 8.89 > 7.815 = \mathcal{X}_c^2$$

ซึ่งหมายความว่าเราจะปฏิเสธ H0 นั่นหมายความว่า มีความสัมพันธ์กันระหว่างกลุ่มข้อมูล