# บทที่ 7

# การทดสอบภาวะสารูปสนิทดี

การทดสอบภาวะสารูปสนิทดี เป็นวิธีการทางสถิติที่ใช้สำหรับวิเคราะห์ข้อมูลที่ได้มาจากการนับ (counts) โดยสถิติที่ใช้ในการทดสอบจะใช้สถิติทดสอบไคสแควร์ (Chi – Square Test) ในการทดสอบ สำหรับการเก็บรวบรวมข้อมูลจะถูกเก็บอยู่ในรูปของตารางการแจกแจงแบบทางเดียวโดยจัดเป็นข้อมูลกลุ่มหรือข้อมูลจำแนกประเภท (categorical data) เช่น

**ตัวอย่างที่ 1** ตารางแสดงจำนวนอาจารย์ในสังกัดคณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ประจำภาคการศึกษาที่ 1/2557 จำแนกตามสาขาวิชาที่สังกัด

| สาขาวิชา | จำนวนอาจารย์ |
|---|---|
| คณิตศาสตร์ | 7 |
| เคมี | 13 |
| ชีววิทยา/ชีววิทยาประยุกต์ | 10 |
| ฟิสิกส์ | 9 |
| วิทยาศาสตร์การอาหาร | 3 |
| เทคโนโลยีสารสนเทศ | 16 |
| วิทยาการคอมพิวเตอร์ | 12 |
| วิทยาศาสตร์การกีฬา | 3 |
| วิทยาศาสตร์สิ่งแวดล้อม | 7 |
| สถิติประยุกต์ | 6 |
| สิ่งทอ/วิทยาศาสตร์สิ่งทอ | 4 |
| สาธารณสุขชุมชน | 10 |

**ตัวอย่างที่ 2** แสดงจำนวนถั่วเมล็ดเรียบกับถั่วเมล็ดขรุขระที่ได้จากการผสมพันธุ์

| ลักษณะของถั่ว | จำนวนที่สังเกตได้ (ความถี่) |
|---|---|
| เมล็ดเรียบ | 69 |
| เมล็ดขรุขระ | 31 |
| รวม | 100 |

คำถามที่มักเกิดขึ้นจากตารางการแจกแจงแบบทางเดียวก็คือ “จำนวนหรือความถี่ของข้อมูลที่สังเกตได้ (Observed frequency) จากกลุ่มข้อมูลย่อยของตัวแปรที่สนใจตัวหนึ่งจะมีค่าแตกต่างจากจำนวนหรือความถี่ของข้อมูลที่คาดหมายตามทฤษฎีของกลุ่มข้อมูลย่อยในตัวแปรนั้น (Expected frequency) หรือไม่ หรืออาจเกิดข้อสงสัยเกี่ยวกับแจกแจงของข้อมูลที่เก็บรวบรวมมาว่ามีการแจกแจงแบบใด ซึ่งในการตอบข้อสงสัยจากคำถามเหล่านี้เราจะใช้การทดสอบที่เรียกว่าการทดสอบภาวะสารูปสนิทดี

# 7.1 ขั้นตอนการทดสอบภาวะสารูปสนิทดี

1. กำหนดสมมติฐานเพื่อการทดสอบ
2. คำนวณค่าความถี่ที่คาดหมายทางทฤษฎี ($E_i$)

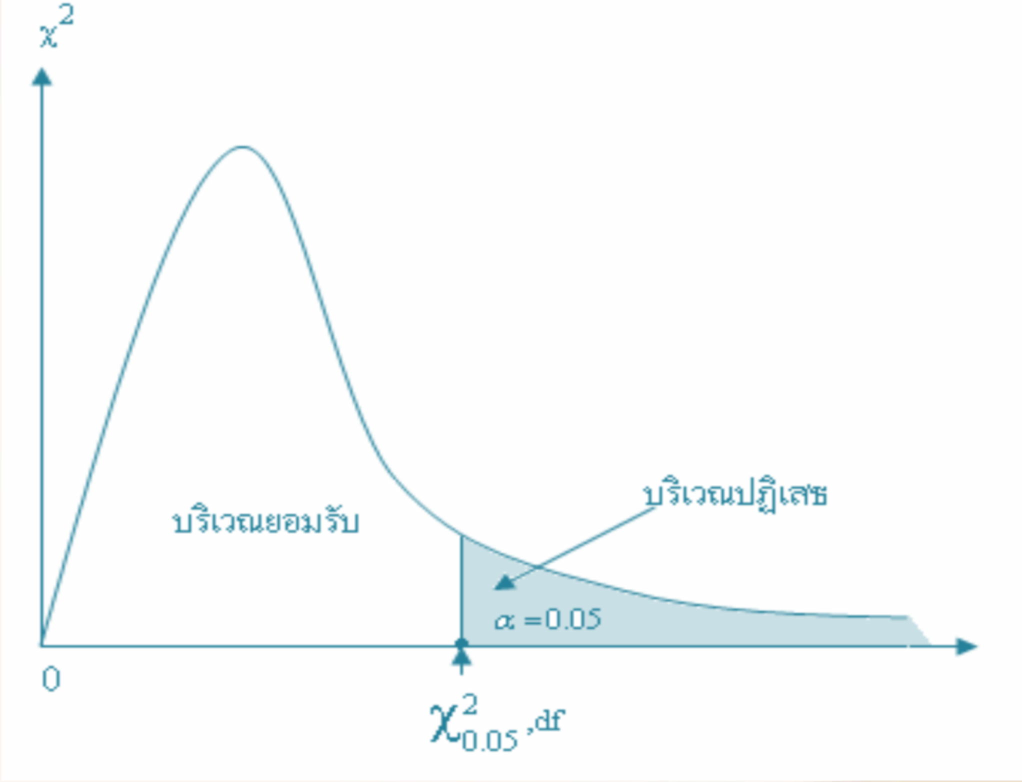
โดยที่ $E_i = Np_i \dfrac{N \cdot c_i}{\sum_{i=1}^{k} c_i}$ และ $N = \sum_{i=1}^{k} O_i$

3. กำหนดระดับนัยสำคัญของการทดสอบ ($\alpha$)
4. คำนวณค่าสถิติไค – สแควร์

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

5. หาบริเวณอาณาเขตวิกฤต

เปิดตาราง หาค่าบริเวณอาณาเขตวิกฤตที่ระดับนัยสำคัญ $\alpha$ ซึ่งการเปิดตารางจะขึ้นอยู่กับระดับความเป็นอิสระของชั้นหรือกลุ่มข้อมูลย่อย (degree of freedom: df) โดยนับจำนวนชั้นหรือกลุ่มข้อมูลย่อย ที่มีอิสระต่อการลงข้อมูลและยังคงทำให้ผลรวมความถี่ข้อมูลที่สังเกตได้ยังคงเหมือนเดิม



6. สรุปผล

ถ้าค่า $\chi^2$ ที่คำนวณได้ มีค่ามากกว่าหรือเท่ากับ $\chi^2$ ที่เปิดจากตาราง จะปฏิเสธสมมุติฐานหลัก ($H_0$) และยอมรับสมมุติฐานรอง ($H_1$)

# 7.2 การทดสอบภาวะสารูปสนิทดี

## 7.2.1 การทดสอบอัตราส่วน

ในประชากรหนึ่งถ้ามีเหตุการณ์ที่สนใจ k เหตุการณ์ (k>2) ดังตาราง

**ตารางแสดงลักษณะข้อมูลความถี่ลักษณะที่สนใจเกิดขึ้นจริง ($O_i$) กับค่าคาดหมาย ($E_i$)**

| กลุ่มหรือระดับ | ค่าความถี่จากการสังเกตที่เกิดขึ้นจริง | ค่าความถี่คาดหวัง |
|---|---|---|
| 1 | $O_1$ | $E_1$ |
| 2 | $O_2$ | $E_2$ |
| 3 | $O_3$ | $E_3$ |
| . | . | . |
| . | . | . |
| . | . | . |
| k | $O_k$ | $E_k$ |
| | N | N |

ถ้าอยากทราบว่าการเกิดขึ้นของเหตุการณ์ต่างๆในที่นี้ให้เป็น $A_1, A_2, A_3, \ldots, A_k$ เป็นไปตามอัตราส่วน $c_1 : c_2 : c_3 \ldots : c_k$ หรือไม่ สามารถสรุปโดยการทดสอบสมมติฐานด้วยสถิติทดสอบไค – สแควร์ ซึ่งมีองศาความเป็นอิสระ k – 1 โดยมีขั้นตอนการทดสอบสมมติฐานดังนี้

<u>ขั้นที่ 1</u> ตั้งสมมติฐาน $H_0 : A_1 : A_2 : A_3 : \ldots : A_k = c_1 : c_2 : c_3 \ldots : c_k$

$H_1 : A_1 : A_2 : A_3 : \ldots : A_k \neq c_1 : c_2 : c_3 \ldots : c_k$

<u>ขั้นที่ 2</u> คำนวณค่าสถิติไค – สแควร์

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

โดยที่ $E_i = Np_i \frac{N \cdot c_i}{\sum_{i=1}^{k} c_i}$ และ $N = \sum_{i=1}^{k} O_i$

ขั้นที่ 3 อาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$ พิจารณาอาณาเขตวิกฤต

$$\chi^2 \geq \chi^2_{\alpha,k-1}$$

ขั้นที่ 4 สรุปผล ถ้า $\chi^2_{cal}$ ตกอยู่ในอาณาเขตวิกฤต สามารถสรุปได้ว่าข้อมูลที่รวบรวมได้มาจากประชากรซึ่งการเกิดเหตุการณ์ต่างๆไม่เป็นไปตามอัตราส่วนตามสมมติฐาน $H_0$

**ตัวอย่างที่ 1** บริษัทผู้ผลิตรถยนต์ยี่ห้อหนึ่ง ต้องการวางแผนการผลิตโดยคาดการณ์ว่าประเภทรถยนต์สี่ประเภทที่ผลิตมีผู้นิยมดังนี้ รถเก๋ง 20% รถกระบะ 2 ตอน 35% รถกระบะตอนเดียว 27% และรถตู้ 18% สุ่มตัวอย่างลูกค้ามา 200 ราย พบว่าลูกค้าซื้อรถดังนี้

| ประเภทของรถยนต์ | รถเก๋ง | รถกระบะ 2 ตอน | รถกระบะตอนเดียว | รถตู้ |
|---|---|---|---|---|
| ความถี่ที่สังเกตได้ | 43 | 65 | 52 | 40 |

จงทดสอบสมมติฐานที่บริษัทผู้ผลิตคาดการณ์ความนิยมไว้ถูกต้องหรือไม่ที่ระดับนัยสำคัญ 0.01

**วิธีทำ** <u>ขั้นที่ 1</u> ตั้งสมมติฐาน

$H_0$ : รถเก๋ง : รถกระบะ 2 ตอน : รถกระบะตอนเดียว : รถตู้ = 0.2 : 0.35 : 0.27 : 0.18

$H_1$ : รถเก๋ง : รถกระบะ 2 ตอน : รถกระบะตอนเดียว : รถตู้ ≠ 0.2 : 0.35 : 0.27 : 0.18

<u>ขั้นที่ 2</u> คำนวณค่าสถิติไค – สแควร์

| ประเภทของรถยนต์ | $O_i$ | $P_i$ | $E_i=NP_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|---|---|
| รถเก๋ง | 43 | 0.2 | 40 | 3 | 9 | 0.225 |
| รถกระบะ 2 ตอน | 65 | 0.35 | 70 | -5 | 25 | 0.357 |
| รถกระบะตอนเดียว | 52 | 0.27 | 54 | -2 | 4 | 0.074 |
| รถตู้ | 40 | 0.18 | 36 | 4 | 16 | 0.444 |
| ผลรวม | | | | | | $\chi^2 = 1.1$ |

ขั้นที่ 3 อาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$ พิจารณาอาณาเขตวิกฤต

$$\chi^2 \geq \chi^2_{0.01,4-1} = 11.34$$

<u>ขั้นที่ 4</u> สรุปผล ค่า $\chi^2$ ที่คำนวณได้มีค่าเท่ากับ 1.1 ซึ่งน้อยกว่า 11.34 จึงไม่ตกอยู่ในบริเวณอาณาเขตวิกฤต ดังนั้นจะยอมรับสมมติฐาน $H_0$ นั่นคือสามารถสรุปได้ว่าที่ระดับนัยสำคัญ 0.01 บริษัทผู้ผลิตคาดการณ์ความนิยมไว้ถูกต้อง

**ตัวอย่างที่ 2** บริษัทผู้ผลิตคอมพิวเตอร์แห่งหนึ่งต้องการวางแผนการผลิตคอมพิวเตอร์ว่าในอีก 2 ปีข้างหน้าควรจะทำการผลิตคอมพิวเตอร์แต่ละชนิดในปริมาณเท่าใด จากการคาดการณ์ของฝ่ายการตลาดคาดว่าสัดส่วนของการใช้คอมพิวเตอร์คือ คอมพิวเตอร์แบบตั้งโต๊ะ โน๊ตบุ๊คคอมพิวเตอร์ แล็ปท็อปคอมพิวเตอร์ และ ปาล์มท็อปคอมพิวเตอร์ จะเป็น 0.1 0.3 0.4 และ 0.2 ตามลำดับ ฝ่ายวิจัยของบริษัทจึงทำการสำรวจจากตัวอย่างคนในวัยทำงานจำนวน 200 คน ปรากฏว่าได้ข้อมูลดังนี้

| ชนิดของคอมพิวเตอร์ | จำนวนผู้ใช้ |
|---|---|
| คอมพิวเตอร์แบบตั้งโต๊ะ | 25 |
| โน๊ตบุ๊คคอมพิวเตอร์ | 55 |
| แล็ปท็อปคอมพิวเตอร์ | 85 |
| ปาล์มท็อปคอมพิวเตอร์ | 35 |
| รวม | 200 |

จงทดสอบสมมติฐานที่ฝ่ายการตลาดของบริษัทคาดการณ์ไว้ถูกต้องหรือไม่ที่ระดับนัยสำคัญ 0.10

**วิธีทำ** ขั้นที่ 1 ตั้งสมมติฐาน

$H_0$ : คอมพิวเตอร์แบบตั้งโต๊ะ : โน๊ตบุ๊คคอมพิวเตอร์ : แล็ปท็อปคอมพิวเตอร์ : ปาล์มท็อปคอมพิวเตอร์ = 0.1 : 0.3 : 0.4 : 0.2

$H_1$ : คอมพิวเตอร์แบบตั้งโต๊ะ : โน๊ตบุ๊คคอมพิวเตอร์ : แล็ปท็อปคอมพิวเตอร์ : ปาล์มท็อปคอมพิวเตอร์ ≠ 0.1 : 0.3 : 0.4 : 0.2

## ขั้นที่ 2 คำนวณค่าสถิติไค – สแควร์

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

| ชนิดของคอมพิวเตอร์ | $O_i$ | $P_i$ | $E_i=NP_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|---|---|
| คอมพิวเตอร์แบบตั้งโต๊ะ | 25 | 0.1 | 20 | 5 | 25 | 1.25 |
| โน๊ตบุ๊คคอมพิวเตอร์ | 55 | 0.3 | 60 | -5 | 25 | 0.417 |
| แล็ปท็อปคอมพิวเตอร์ | 85 | 0.4 | 80 | 5 | 25 | 0.3125 |
| ปาล์มท็อปคอมพิวเตอร์ | 35 | 0.2 | 40 | -5 | 25 | 0.625 |
| รวม | 200 | | | | | $\chi^2$ = 2.6045 |

ขั้นที่ 3 อาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$ พิจารณาอาณาเขตวิกฤต

$$\chi^2 \geq \chi^2_{0.10,4-1} = 6.25$$

ขั้นที่ 4 สรุปผล ค่า $\chi^2$ ที่คำนวณได้มีค่าเท่ากับ 2.6045 ซึ่งมีค่าน้อยกว่า 6.25จึงไม่ตกอยู่ในบริเวณอาณาเขตวิกฤต ดังนั้นจะยอมรับสมมติฐาน $H_0$ นั่นคือสามารถสรุปได้ว่าที่ระดับนัยสำคัญ 0.01 ฝ่ายการตลาดของบริษัทคอมพิวเตอร์คาดการณ์ไว้ถูกต้อง

### 7.2.2 การทดสอบการแจกแจง

ในการศึกษาตัวแปรหนึ่ง ถ้ามีความสงสัยเกี่ยวกับการแจกแจงของตัวแปรนั้น สามารถทดสอบได้ด้วยสถิติทดสอบไค – สแควร์ โดยมีขั้นตอนการทดสอบสมมติฐานดังนี้

ขั้นที่ 1 ตั้งสมมติฐาน

$H_0$ : ตัวแปรมีการแจกแจงตามข้อสงสัย

$H_1$ : ตัวแปรไม่มีการแจกแจงตามข้อสงสัย

ขั้นที่ 2 คำนวณค่าสถิติ

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

ขั้นที่ 3 อาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$ พิจารณาอาณาเขตวิกฤต

$$\chi^2 \geq \chi^2_{\alpha,\, k-r-1}$$

ขั้นที่ 4 สรุปผล ถ้า $\chi^2_{cal}$ ตกอยู่ในอาณาเขตวิกฤต ไม่สามารถสรุปได้ว่าตัวแปรมีการแจกแจงตามข้อสงสัย

**ตัวอย่างที่ 1** ข้อมูลที่กำหนดให้ต่อไปนี้เป็นความยาวของทารกแรกเกิด จำนวน 125 คน

| ความยาว (ซม.) | 45 – 46.9 | 47 – 48.9 | 49 – 50.9 | 51 – 52.9 | 53 – 54.9 |
|---|---|---|---|---|---|
| จำนวน | 28 | 32 | 35 | 20 | 10 |

สรุปได้หรือไม่ว่าความยาวของทารกแรกเกิดมีการแจกแจงแบบปกติ ที่ระดับนัยสำคัญ 0.05

**วิธีทำ** ขั้นที่ 1 ตั้งสมมติฐาน

$H_0$ : ความยาวของทารกแรกเกิดมีการแจกแจงแบบปกติ

$H_1$ : ความยาวของทารกแรกเกิดไม่ได้มีการแจกแจงแบบปกติ

ขั้นที่ 2 คำนวณค่าสถิติ

| ความยาว (ซม.) | จำนวน ($f_i$) | จุดกึ่งกลางชั้น ($X_i$) | $X_i^2$ | $f_iX_i$ | $f_iX_i^2$ |
|---|---|---|---|---|---|
| 45 – 46.9 | 28 | 45.95 | 2111.4025 | 1286.6 | 59,119.27 |
| 47 – 48.9 | 32 | 47.95 | 2299.2095 | 1534.4 | 73,574.70 |
| 49 – 50.9 | 35 | 49.95 | 2495.0025 | 1748.25 | 87,325.09 |
| 51 – 52.9 | 20 | 51.95 | 2698.8025 | 1039 | 53,976.05 |
| 53 – 54.9 | 10 | 53.95 | 2910.6025 | 539.5 | 29,106.03 |
| รวม | 125 | | | 6147.75 | 303,101.14 |

จะได้ค่า

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} = \frac{6,147.75}{125} = 49.18$$ ซึ่งเป็นตัวประมาณของ $\mu$

$$s^2 = \frac{\sum_{i=1}^{R} f_i X_i^2 - \frac{\left(\sum_{i=1}^{R} f_i X_i\right)^2}{n}}{n-1} = \frac{303,101.14 - \frac{(6147)^2}{125}}{125-1}$$

$$= \frac{303,101.14 - 302,284.87}{124} = \frac{816.27}{124} = 6.58$$

$$s = \sqrt{6.58} = 2.57$$

| ขอบเขตชั้น | จำนวน ($O_i$) | ค่ามาตรฐาน ($z = \frac{x-\mu}{\sigma}$) | P(Z) | E = N P(Z) |
|---|---|---|---|---|
| ไม่เกิน 44.95 | 0 | ไม่เกิน (-1.65) | 0.0495 | 6.1875 |
| 44.95 – 46.95 | 28 | (-1.65) – (-0.87) | 0.147 | 18.375 |
| 46.95 – 48.95 | 32 | (-0.87) – (-0.09) | 0.2719 | 33.9875 |
| 48.95 – 50.95 | 35 | (-0.09) – (0.69) | 0.2908 | 36.35 |
| 50.95 – 52.95 | 20 | (0.69) – (1.47) | 0.1743 | 21.7875 |
| 52.95 - 54.95 | 10 | (1.47) – (2.25) | 0.0586 | 7.325 } 8.85 |
| 54.95 เป็นต้นไป | 0 | 2.25 เป็นต้นไป | 0.0122 | 1.525 |
| รวม | 125 | | | |

$$\chi^2_{cal} = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(0-6.1875)^2}{6.1875} + \frac{(28-18.375)^2}{18.375} + \frac{(32-33.9875)^2}{33.9875} +$$

$$\frac{(35-36.35)^2}{36.35} + \frac{(20-21.7875)^2}{21.7875} + \frac{(10-8.85)^2}{8.85}$$

$$= 6.188 + 5.042 + 0.116 + 0.05 + 0.147 + 0.149$$

$$= 11.692$$

<u>ขั้นที่ 3</u> พิจารณาอาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$=0.05 จะได้บริเวณอาณาเขตวิกฤตที่

$$\chi^2_{cal} \geq \chi^2_{0.05,6-2-1} = \chi^2_{0.05,3} = 7.8$$

<u>ขั้นที่ 4</u> สรุปผล เนื่องจากค่า $\chi^2_{cal}$ ที่คำนวณได้มีค่ามากกว่า 7.8 จึงตกอยู่ในบริเวณอาณาเขตวิกฤต ดังนั้นจึงปฏิเสธสมมติฐาน $H_0$ นั่นคือ ที่ระดับนัยสำคัญ 0.05 ความยาวของทารกแรกเกิดไม่ได้มีการแจกแจงแบบปกติ

**ตัวอย่างที่ 2** คาร์แคร์แห่งหนึ่งเปิดบริการล้างรถยนต์ด้วยเครื่องล้างอัตโนมัติในการคิดราคาค่าบริการขึ้นอยู่กับการเลือกใช้บริการของลูกค้า เจ้าของคาร์แคร์ต้องการทราบว่าจำนวนรถที่เข้ามาใช้บริการในช่วงเวลาเย็นวันทำงาน (16.30 – 18.30 น.) มีการแจกแจงแบบปัวส์ซองหรือไม่จึงทำการรวบรวมข้อมูลจำนวนรถที่มาใช้บริการในช่วงเวลาดังกล่าวเป็นเวลา 180 วัน ได้ข้อมูลดังนี้

| จำนวนรถที่เข้ามาใช้บริการล้างรถ | ความถี่ที่สังเกตได้ |
|---|---|
| 0 | 15 |
| 1 | 55 |
| 2 | 60 |
| มากกว่า 2 คัน | 50 |
| รวม | 180 |

จงทดสอบที่ระดับนัยสำคัญ 0.01 ถ้าจากประสบการณ์ที่ผ่านมาทราบว่าจำนวนรถยนต์ที่เข้ามาใช้บริการในช่วงเวลาดังกล่าวเฉลี่ยแล้วจะเป็น 2 คัน

**วิธีทำ** <u>ขั้นที่ 1</u> ตั้งสมมติฐาน

$H_0$ : จำนวนรถที่เข้ามาใช้บริการในช่วงเวลาเย็นวันทำงาน (16.30 – 18.30 น.) มีการแจกแจงแบบปัวส์ซอง

$H_1$ : จำนวนรถที่เข้ามาใช้บริการในช่วงเวลาเย็นวันทำงาน (16.30 – 18.30 น.) ไม่ได้มีการแจกแจงแบบปัวส์ซอง

<u>ขั้นที่ 2</u> คำนวณค่าสถิติ

จากโจทย์ทราบว่า $\lambda = 2$ คำนวณค่าสถิติดังตาราง

| จำนวนรถที่เข้ามาใช้บริการล้างรถ | ความถี่ที่สังเกตได้ ($O_i$) | P(X=x) | E = N P P(X=x) |
|---|---|---|---|
| 0 | 15 | 0.1353 | 24.354 |
| 1 | 55 | 0.2707 | 48.726 |
| 2 | 60 | 0.2707 | 48.726 |
| มากกว่า 2 คัน | 50 | 0.3233 | 58.194 |
| รวม | 180 | 1.0000 | 180 |

จากสูตร

$$\chi^2_{cal} = \sum_{i=1}^{6} \frac{\left(O_i - E_i\right)^2}{E_i}$$

$$= \frac{\left(15-24.354\right)^2}{24.354} + \frac{\left(55-48.726\right)^2}{48.726} + \frac{\left(60-48.726\right)^2}{48.726}$$

$$+ \frac{\left(50-58.194\right)^2}{58.194}$$

$$= 3.593 + 0.808 + 2.609 + 1.154$$

<u>ขั้นที่ 3</u> พิจารณาอาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ $\alpha$=0.01 จะได้บริเวณอาณาเขตวิกฤตที่ $\chi^2_{cal} \geq \chi^2_{0.01,4-1-1} = \chi^2_{0.01,2} = 9.21$

<u>ขั้นที่ 4</u> สรุปผล เนื่องจากค่า $\chi^2_{cal}$ ที่คำนวณได้มีค่าน้อยกว่า 9.21 จึงไม่ตกอยู่ในบริเวณอาณาเขตวิกฤต ดังนั้นจึงยอมรับสมมติฐาน $H_0$ นั่นคือที่ระดับนัยสำคัญ 0.01 จำนวนรถที่เข้ามาใช้บริการในช่วงเวลาเย็นวันทำงาน (16.30 – 18.30 น.) มีการแจกแจงแบบปัวส์ซอง

## 7.3 บทสรุป

การทดสอบภาวะสารูปสนิทดี เป็นการทดสอบสมมติฐานในกรณีที่ข้อมูลถูกเก็บรวบรวมมาอยู่ในรูปของความถี่ (frequency data) หรือข้อมูลจำแนกประเภท (categorical data) โดยที่ข้อมูลถูกจำแนกประเภทแบบทางเดียว (one dimensional data) หรือเป็นตารางความถี่แบบทางเดียว (One-Way frequency table) โดยในการทดสอบสมมติฐานจะแบ่งเป็น 2 ลักษณะ คือ การทดสอบอัตราส่วนของประชากรเป็นไปตามที่คาดไว้หรือไม่ และทดสอบการแจกแจงของประชากรเป็นไปตามที่คาดไว้หรือไม่ สำหรับสถิติที่ใช้ในการทดสอบคือสถิติทดสอบไคสแควร์ ($\chi^2$ – test) ซึ่งเหมาะสมกับตัวอย่างขนาดใหญ่ ถ้าขนาดของตัวอย่างเป็น 4 หรือ 5 เท่าของจำนวนเหตุการณ์ ค่าไคสแควร์จะประมาณได้ค่อนข้างดี แม้ว่าค่าความถี่คาดหวังจะน้อย แต่ค่าความถี่คาดหวังก็ไม่ควรน้อยกว่า 5 ถ้าค่าความถี่คาดหวังของเหตุการณ์ใดน้อยกว่า 5 จะต้องทำการปรับแก้โดยรวมความถี่ของเหตุการณ์นั้นเข้ากับความถี่ของเหตุการณ์อยู่ใกล้เคียง ซึ่งจะทำให้องศาความเป็นอิสรลดลง หรือแนวทางปรับแก้ที่ดีก็คือ ควรเก็บข้อมูลเพิ่มเติมให้ตัวอย่างมีขนาดใหญ่ขึ้น