**บทที่ 6**

**การวิเคราะห์ข้อมูลที่อยู่ในรูปความถี่**

การแจกแจงแบบไคสแควร์ที่จะนำมากล่าวในบทนี้ เป็นการนำเอาการแจกแจงแบบไคสแควร์มาประยุกต์ เพื่อเปรียบเทียบความถี่ที่สังเกต (Observed frequency) กับความถี่ที่คาดหวัง (Expected frequency) ซึ่งสอดคล้องกับความน่าจะเป็นของเหตุการณ์ที่ข้อมูลถูกจัดแบ่งตามกลุ่ม การแจกแจงแบบไคสแควร์ประยุกต์ จะใช้เพื่อการทดสอบเท่านั้น และการทดสอบที่จะกล่าวถึงในบทนี้ได้แก่ การทดสอบภาวะรูปดี (test for goodness of fit) และการทดสอบความเป็นอิสระ (test for independence)

ถ้าให้ k คือ จำนวนเหตุการณ์หรือกลุ่มของข้อมูลที่ได้จัดแบ่งทั้งหมด

Oi คือ ความถี่ที่สังเกตของเหตุการณ์ที่ i

Ei คือ ความถี่คาดหวังของเหตุการณ์ที่ i โดยที่ Ei = Npi

N คือ ผลรวมของความถี่ทั้งหมด

Pi คือ ความน่าจะเป็นของเหตุการณ์ที่ i

และถ้าแปลง Oi และ Ei เป็นตัวแปร χ2 ดังนี้



แล้วตัวแปร χ2 จะมีการแจกแจงแบบไคสแควร์

การทดสอบแบบไคสแควร์ประยุกต์เป็นการเปรียบเทียบความถี่ที่สังเกต (Oi) กับความถี่ที่คาดว่าจะเป็นตามทฤษฎีหรือข้อกำหนดที่ต้องการทดสอบ (Ei) ถ้าความถี่ที่สังเกตสอดคล้องกับความถี่ตามทฤษฎีหรือข้อกำหนด ค่า Oi กับค่า Ei จะใกล้เคียงกันหรือเท่ากัน เมื่อคำนวณค่า χ2 จะได้ค่าใกล้เคียงหรือเท่ากับศูนย์ แต่ถ้าความถี่ที่สังเกตไม่สอดคล้องกับความถี่ตามทฤษฎีหรือข้อกำหนด ค่า Oi กับค่า Ei จะแตกต่างกัน ค่า χ2 ที่คำนวณได้จะมีค่ามาก ดังนั้นในการทดสอบสมมติฐานว่าความถี่ที่สังเกตสอดคล้องกับความถี่ตามทฤษฎีหรือข้อกำหนด โดยใช้ไคสแควร์แบบประยุกต์ จะกำหนดอาณาเขตทางขวามือเท่านั้น กล่าวคือพิจารณาว่า Oi และ Ei จะแตกต่างกันมากจนทำให้ค่า  อยู่ในอาณาเขตหรือไม่ สำหรับดีกรีความเป็นอิสระ (υ) จะกล่าวต่อไปในการทดสอบแต่ละอย่างที่นำไปประยุกต์



อาณาเขตวิกฤตคือ χ2 > 

**ข้อจำกัดของการใช้ไคสแควร์แบบประยุกต์**

1. การทดสอบด้วยไคสแควร์แบบประยุกต์นี้เหมาะกับตัวอย่างขนาดใหญ่ ถ้าขนาดตัวอย่างเป็น 4 หรือ 5 เท่าของจำนวนเหตุการณ์ ค่าไคสแควร์จะประมาณได้ค่อนข้างดี แม้ว่าความถี่คาดหวังจะน้อย อย่างไรก็ดีไม่ควรน้อยกว่า 5 ถ้าความถี่คาดหวังของเหตุการณ์ใดน้อยกว่า 5 จะปรับแก้โดยรวมความถี่ของเหตุการณ์นั้นเข้ากับความถี่ของเหตุการณ์ที่อยู่ใกล้เคียง ซึ่งจะทำให้ดีกรีความเป็นอิสระลดลง แนวทางปรับแก้ที่ดีก็คือ เก็บข้อมูลเพิ่มเติมให้ตัวอย่างมีขนาดใหญ่ขึ้น

2. การหาดีกรีความเป็นอิสระ สำรับข้อมูลที่แบ่งเป็นเป็น k เหตุการณ์ υ = k – r – 1 โดยที่ r คือ จำนวนพารามิเตอร์ที่ต้องประมาณจากข้อมูลเพื่อใช้ในการหาค่าความถี่คาดหวังลบด้วย 1 เพราะว่าสูญเสียความเป็นอิสระในการประมาณความถี่ที่คาดหวัง ทั้งนี้เนื่องจากข้อจำกัดที่ว่า  นั่นคือความถี่ที่คาดหวังของเหตุการณ์สุดท้ายอาจหาจาก N - 

3. กรณีดีกรีความเป็นอิสระเท่ากับ 1 การประยุกต์ไคสแควร์จะใช้ได้ไม่ค่อยดี Frank Yate ได้ปรับแก้ตัวสถิติไคสแควร์ เพื่อให้การประยุกต์ไคสแควร์ได้ดีขึ้น ดังนี้



**การทดสอบภาวะรูปดี (Goodness of fit)**

การบันทึกข้อมูลของเหตุการณ์ต่างๆอาจถูกจัดขึ้นด้วยลักษณะหรือรูปแบบที่แตกต่างกัน เช่น

1. ข้อมูลการขายเครื่องซักผ้าชนิดหนึ่ง จำนวน 300 เครื่อง จำแนกตามสีที่ผู้นิยมซื้อ

|  |  |
| --- | --- |
| สี | น้ำตาล แดง ฟ้า ขาว |
| จำนวน | 65 52 40 55 |

1. ข้อมูลการผสมไม้ดอกชนิดหนึ่ง ซึ่งสามารถแยกเป็นสีแดง สีชมพู และสีขาว เมื่อเพาะไม้ดอก 120 ต้น ได้ผลดังนี้

|  |  |
| --- | --- |
| สี | แดง ชมพู ขาว |
| จำนวน | 24 80 16 |

จากตัวอย่างข้างต้นจะเห็นว่าข้อมูลถูกจำแนกเป็นเหตุการณ์ตามกลุ่ม โดยไม่ทราบความน่าจะเป็นของแต่ละกลุ่ม จุดประสงค์สำคัญที่ต้องการทราบก็คือ รูปแบบของความน่าจะเป็นแบบไหนที่สอดคล้องกับข้อมูลที่สังเกตได้เหล่านี้ ถ้าเห็นว่าน่าจะใช้รูปแบบของความน่าจะเป็นชนิดใดกับข้อมูลเหล่านั้นแล้ว จะต้องทดสอบว่าถูกต้องหรือไม่ เรียกการทดสอบนี้ว่า การทดสอบภาวะรูปดี อาจสรุปได้ว่า การทดสอบลักษณะนี้เป็นการทดสอบว่า ความน่าจะเป็น หรือการแจกแจงความน่าจะเป็นเป็นไปตามทฤษฎีหรือไม่

ในที่นี้จะจำแนกการทดสอบภาวะรูปดีออกเป็น การทดสอบอัตราส่วน และการทดสอบการแจกแจง ดังนี้

1. **การทดสอบอัตราส่วน**

ในประชากรหนึ่งถ้ามีเหตุการณ์ที่สนใจ k เหตุการณ์ , k ≥ 2 อยากทราบว่าการเกิดขึ้นของ

เหตุการณ์ต่างๆ ในที่นี้ให้เป็น A1 , A2 , A3 ,…,Ak เป็นไปตามอัตราส่วน c1 : c2 : c3 , … : ck หรือไม่ สามารถสรุปโดยการทดสอบสมมติฐานด้วยตัวสถิติไคสแควร์ ซึ่งมีดีกรีความเป็นอิสระเท่ากับ k – 1 ทั้งนี้เนื่องจากภายใต้ความถี่ที่ได้จากการสังเกตทั้งหมดของการทดลองจะมีความเป็นอิสระในการหาความถี่คาดหวัง จำนวน k – 1 เหตุการณ์เท่านั้น โดยมีขั้นตอนการทดสอบสมมติฐานดังนี้

ขั้นที่ 1 กำหนดสมมติฐาน

H0 : A1 , A2 , A3 ,…,Ak = c1 : c2 : c3 , … : ck

H1 : A1 , A2 , A3 ,…,Ak ≠ c1 : c2 : c3 , … : ck

ขั้นที่ 2 คำนวณค่าสถิติไคสแควร์



โดยที่ Ei = Npi =  และ 

ขั้นที่ 3 อาณาเขตวิกฤต ภายใต้ระดับนัยสำคัญ α พิจาณาอาณาเขตวิกฤต

χ2 ≥ 

ขั้นที่ 4 สรุปผล ถ้า  จะตกอยู่ในอาณาเขตวิกฤต สามารถสรุปได้ว่าข้อมูลที่รวบรวมได้มาจากประชากร ซึ่งการเกิดขึ้นของเหตุการณ์ต่างๆไม่เป็นไปตามอัตราส่วนของสมมติฐาน H0

**ตัวอย่าง** นักสถิติของธนาคารเลือดมีความสงสัยว่ากรุ๊ปเลือด A , B, AB และ O เป็นไปตามอัตราส่วน

2 : 2: 1: 3 หรือไม่ จึงรวบรวมข้อมูลเพื่อทำการทดสอบ โดยสุ่มตัวอย่างคนทั่วไปจำนวน 100 คน เจาะเลือดเพื่อตรวจสอบกรุ๊ปเลือด ปรากฏว่าเป็นเลือดกรุ๊ป A , B, AB และ O จำนวน 20, 25, 10 และ 45 คน ตามลำดับ ที่ระดับนัยสำคัญ 10% จะสรุปข้อสงสัยข้างต้นได้หรือไม่

**วิธีทำ** 1. กำหนดสมมติฐานเพื่อการทดสอบ

H0 : A : B : AB : O = 2 : 2: 1: 3

H1 : A : B : AB : O ≠ 2 : 2: 1: 3

2. คำนวณค่าสถิติ 

|  |  |  |  |
| --- | --- | --- | --- |
| เหตุการณ์ | O | E |  |
| A  B  AB  O | 20  25  10  45 |  | 1  0  0.5  1.5 |
| รวม | 100 | 100 | 3.0 |

3. ระดับนัยสำคัญ α = 0.10 ,  อาณาเขตวิกฤต 

4. สรุปผล เนื่องจาก  จึงไม่ตกอยู่ในอาณาเขตวิกฤต ดังนั้นจะยอมรับสมมติฐาน H0 สรุปได้ว่าอัตราส่วนของกรุ๊ปเลือดของคนทั่วไป เป็นไปตามอัตราส่วนที่นักสถิติของธนาคารเลือดคาดไว้ คือ A : B : AB : O = 2 : 2: 1: 3

**2. การทดสอบการแจกแจง**

ในการศึกษาตัวแปรหนึ่ง ถ้ามีความสงสัยเกี่ยวกับการแจกแจงของตัวแปรนั้น สามารถทดสอบได้ด้วยตัวสถิติไคสแควร์ ในที่นี้จะสนใจเฉพาะการทดสอบการแจกแจงพื้นฐาน คือ การแจกแจงแบบ

ทวินาม การแจกแจงแบบปัวส์ซอง และการแจกแจงแบบปกติ ซึ่งการทดสอบการแจกแจงนี้มีความสำคัญค่อนข้างมาก โดยเฉพาะอย่างยิ่ง การทดสอบการแจกแจงแบบปกติ ทั้งนี้เนื่องจากเงื่อนไขหรือข้อสมมติของสถิติบางอย่างจะต้องเป็นตัวแปรที่มีการแจกแจงแบบปกติ โดยมีขั้นตอนการทดสอบสมมติฐานดังนี้

ขั้นที่ 1 กำหนดสมมติฐาน

H0 : ตัวแปรมีการแจกแจงตามข้อสงสัย

H1 : ตัวแปรไม่มีการแจกแจงตามข้อสงสัย

ขั้นที่ 2 คำนวณค่าสถิติ



ขั้นที่ 3 พิจารณาอาณาเขตวิกฤตภายใต้ระดับนัยสำคัญ α

 เมื่อ r เป็นจำนวนพารามิเตอร์ที่ต้องประมาณ

k เป็นจำนวนเหตุการณ์ที่ปรับค่าคาดหวังแล้ว

ขั้นที่ 4 สรุปผล ถ้า  ตกอยู่ในอาณาเขตวิกฤต ไม่สามารถสรุปได้ว่าตัวแปรมีการแจกแจงตามข้อสงสัย

**ตัวอย่าง** ในการปลูกเมล็ดพันธุ์ผัก 450 เมล็ด โดยปลูกเป็นแถวๆละ 5 เมล็ด จำนวน 90 แถว เมล็ดที่งอกในแต่ละแถวเป็นดังนี้

|  |  |
| --- | --- |
| จำนวนเมล็ดพันธุ์ที่งอก /แถว | 1 2 3 4 5 |
| จำนวนแถว | 1 11 30 38 10 |

จากผลการทดลองนี้จะพิจารณาว่าข้อมูลดังกล่าวควรมีการแจกแจงความน่าจะเป็นแบบใดและทดสอบสมมติฐาน โดยใช้ระดับความมีนัยสำคัญ 5%

**วิธีทำ** พิจารณาข้อมูลตามค่าสังเกต ที่จำแนกเป็นการงอกของเมล็ดในแต่ละแถว โดยปลูกแถวละ 5 เมล็ด ถ้านิยามตัวแปรสุ่ม X ดังนี้

X : จำนวนเมล็ดที่งอกในแต่ละแถว x = 0 , 1 , 2 , 3 , 4 , 5

จะเห็นว่าตัวแปร X มีคุณสมบัติเบื้องต้นเป็นการแจกแจงแบบทวินาม

1. กำหนดสมมติฐาน

H0 : ข้อมูลสอดคล้องกับการแจกแจงแบบทวินาม

H1 : ข้อมูลไม่สอดคล้องกับการแจกแจงแบบทวินาม

1. คำนวณค่าสถิติ



การคำนวณค่าความถี่คาดหวังต้องใช้ความน่าจะเป็นของการแจกแจงแบบทวินามตาม H0 ซึ่งจะเห็นว่าไม่ทราบค่า p (ความน่าจะเป็นที่เมล็ดแต่ละเมล็ดจะงอก) จึงต้องประมาณจากค่าสังเกตดังนี้





เนื่องจากค่าเฉลี่ยของการแจกแจงแบบทวินาม = np

ดังนั้น np = 3.5 กรณีนี้ n = 5

นั่นคือ p = 

จะหา P(X=x) จากตารางทวินาม b(x ; 5,0.7)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | O | P(X=x) | E = NP(X=x) |  |
| 0  1  2  3  4  5 | 0  1  11  30  38  10 | 0.0024  0.0284  0.1323  0.3087  0.3602  0.1681 | 0.216  =14.679  2.556  11.907  27.783  32.418  15.129 | 0.489  0.177  0.961  1.739 |
| รวม | 90 | 1.0000 | 90 | 3.366 |



1. การหาดีกรีความเป็นอิสระ เพื่อหาอาณาเขตวิกฤต เมื่อ α = 0.05 จากตารางคำนวณข้างต้น

จะเห็นว่า ความถี่คาดหวังใน 2 กลุ่มแรกน้อยกว่า 5 ดังนั้น จึงต้องรวมกับกลุ่มที่ 3 เป็น 14.679 ทำให้กลุ่มของข้อมูลเหลือเพียง 4 กลุ่ม นั่นคือ k = 4 และในกรณีนี้ต้องประมาณพารามิเตอร์ p จากค่าสังเกตทำให้เสียดีกรีความเป็นอิสระไป 1

ดังนั้น υ = 4 – 1 – 1 = 2



เนื่องจาก  ไม่อยู่ในอาณาเขตวิกฤตจึงยอมรับสมมติฐาน H0 ที่ว่าข้อมูลการทดลองสอดคล้องกับการแจกแจงแบบทวินาม ด้วยระดับนัยสำคัญ 5%

**3. การทดสอบความเป็นอิสระ**

ตารางการแจกแจงแบบสองทาง ในการจัดเตรียมข้อมูลเบื้องต้น เพื่อใช้ในการวิเคราะห์จะเห็นว่าการจัดในรูปตารางไขว้เป็นวิธีที่นิยมใช้เสมอ คือ จัดเป็นแกนนอน (row) และแกนตั้ง (column) ซึ่งเรียกว่าการแจกแจงแบบสองทาง ทั้งนี้เพื่ออธิบายลักษณะของข้อมูลเป็น 2 ลักษณะ อาจเรียก ลักษณะทั้งสองว่าตัวแปร ซึ่งเป็นตัวแปรคุณภาพ ถ้าแทนด้วยตัวแปร X จัดให้อยู่ตามแนวนอนและตัวแปร Y จัดให้อยู่ตามแนวตั้ง เช่น

X : วิชาเอกที่ศึกษา Y : อาชีพ

X : เพศ Y : ความนิยมอ่านหนังสือพิมพ์คอลัมน์ต่างๆ

X : ระดับชั้นปีของนักศึกษา Y : ความคิดเห็นเกี่ยวกับการอยู่ในหอพักของมหาวิทยาลัย

ตารางแจกแจงแบบสองทางที่สร้างขึ้นเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปร X และ Y ถ้าจัดตัวแปร X ตามแกนนอนจำนวน r แถว และจัดตัวแปร Y ตามแกนตั้งจำนวน c สดมภ์ ตารางนี้จะเรียกว่า ตารางแจกแจง 2 ทางมิติ r ×c ซึ่งจะใช้ทดสอบความเป็นอิสระของตัวแปร X และตัวแปร Y

**การตั้งสมมติฐาน**

สมมติฐานสำหรับการทดสอบอาจเขียนอย่างใดอย่างหนึ่ง ดังนี้

H0 : ตัวแปร X และตัวแปร Y เป็นอิสระต่อกัน

H0 : ตัวแปร X ไม่ขึ้นอยู่กับตัวแปร Y

H0 : ตัวแปร X และตัวแปร Y ไม่มีความสัมพันธ์กัน

เช่น H0 : ลักษณะอาชีพไม่ขึ้นอยู่กับสาขาวิชาเอกที่สำเร็จการศึกษา

H0 : ความเห็นของนักศึกษาเกี่ยวกับการอยู่หอพักของมหาวิทยาลัยไม่มีความสัมพันธ์กับชั้นปีของนักศึกษาสังกัดอยู่

H0 : ความนิยมอ่านหนังสือพิมพ์คอลัมน์ต่างๆไม่ขึ้นอยู่กับเพศ

**การหาค่าความถี่คาดหวัง**

การทดสอบสมมติฐานเกี่ยวกับความเป็นอิสระของตัวแปรทั้งสองนี้ เมื่อจะหาค่าคาดหวัง Ei = Npi จะต้องหาค่า pi ที่สอดคล้องกับสมมติฐาน กล่าวคือ สอดคล้องกับคุณสมบัติของความน่าจะเป็นของเหตุการณ์ที่เป็นอิสระกัน P(AB) = P(A)P(B)

พิจารณาตารางแจกแจงสองทางมิติ 4×3 ต่อไปนี้

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| คุณลักษณะ X | คุณลักษณะ Y | | | รวม |
| Y1 | Y2 | Y3 |
| X1 | O11 | O12 | O13 | R1 |
| X2 | O21 | O22 | O23 | R2 |
| X3 | O31 | O32 | O33 | R3 |
| X4 | O41 | O42 | O43 | R4 |
| รวม | C1 | C2 | C3 | N |

ความน่าจะเป็นที่จะเกิด X1 ; P(X1) = 

ความน่าจะเป็นที่จะเกิด Y1 ; P(Y1) = 

พิจารณาเหตุการณ์ X1 Y1 จะหาความน่าจะเป็นที่เกิด X1 Y1 เพื่อจะหาเทียบกับ O11 ตามสมมติฐานที่ว่าตัวแปร X และตัวแปร Y เป็นอิสระกันจะได้ว่า

P(X1 Y1) = P(X1)P(Y1) = 

E11= NP(X1 Y1) = N×=

ทำนองเดียวกัน การหาความน่าจะเป็นของเหตุการณ์ร่วมอื่นๆก็ทำได้อย่างเดียวกัน

เช่น , เป็นต้น

แล้วจะได้ตารางความถี่คาดหวัง ดังนี้

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| คุณลักษณะ X | คุณลักษณะ Y | | | รวม |
| Y1 | Y2 | Y3 |
| X1 | E11 | E12 | E13 | R1 |
| X2 | E21 | E22 | E23 | R2 |
| X3 | E31 | E32 | E33 | R3 |
| X4 | E41 | E42 | E43 | R4 |
| รวม | C1 | C2 | C3 | N |

**ดีกรีความเป็นอิสระ** ในการพิจารณาค่าดีกรีความเป็นอิสระของการทดสอบด้วย ตัวสถิติไคสแควร์แบบนี้จะพิจารณาจากข้อกำหนด คือ  แต่เนื่องจากลักษณะของข้อมูลเป็นตารางแจกแจงสองทางซึ่งหมายความว่า  ดังนั้นจะต้องทำให้ผลรวมของทุกแถวและทุกสดมน์ของตารางความถี่คาดหวังเท่ากับตารางความถี่ของการสังเกต นั่นคือ ค่าความถี่คาดหวังในแถวสุดท้ายและสดมภ์สุดท้าย ไม่จำเป็นต้องหาจากค่าความน่าจะเป็นดังแสดงข้างต้น แต่สามารถหาได้จากการลบผลรวมของแถวหรือสดมภ์ด้วยค่าที่คาดหวังที่หาจากความน่าจะเป็นอื่นๆนั่นคือ จะมีจำนวนแถวและสดมภ์ที่เป็นอิสระเท่ากับ (r – 1) และ (c – 1) ตามลำดับ ดังนั้นค่าระดับความเป็นอิสระจึงมีค่าเท่ากับ (r – 1) (c – 1)

**ตัวอย่าง** จากตารางการแจกแจงความคิดเห็นของคนงานที่มีต่อการหยุดงานในวันเสาร์จำแนกตามเพศ

เป็นดังนี้

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| เพศ | ความคิดเห็น | | | รวม |
| เห็นด้วย | ไม่เห็นด้วย | ไม่มีความคิดเห็น |
| ชาย  หญิง | 42  74 | 100  20 | 8  6 | 150  100 |
| รวม | 116 | 120 | 14 | 250 |

จงทดสอบว่าที่ระดับนัยสำคัญ 0.05 เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานมีความสัมพันธ์กันหรือไม่

**วิธีทำ**

**สมมติฐาน**

H0: เพศกับความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานเป็นอิสระ

หรือไม่มีความสัมพันธ์

H1: เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานไม่เป็นอิสระ

หรือมีความสัมพันธ์กัน

เนื่องจาก 











#### สถิติทดสอบ



 

= 54.63

ค่า  ที่เปิดจากตารางที่องศาความเป็นอิสระ (2 – 1)(3 – 1) = 2 และระดับนัยสำคัญ 0.05 เท่ากับ 5.99

**สรุป** ปฏิเสธสมมติฐาน H0 เนื่องจาก นั่นคือ เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานไม่เป็นอิสระหรือมีความสัมพันธ์กัน

ดังข้อมูลจากตาราง จะพบว่า ระดับความคิดเห็น เพศชายหญิงไม่ไปในทิศทางเดียวกัน เช่น

ชาย เห็นด้วย 42 ไม่เห็นด้วย 100 (น้อย , มาก)

หญิง เห็นด้วย 74 ไม่เห็นด้วย 20 (มาก , น้อย)

**การวัดระดับความสัมพันธ์**

จาก 



**แบบฝึกหัด**

**บทที่ 6 การวิเคราะห์ข้อมูลที่อยู่ในรูปความถี่**

1. ตารางที่กำหนดให้ต่อไปนี้ได้จากการสำรวจการประสบความสำเร็จในการประกอบอาชีพของคนงาน 432 คน จำแนกตามระดับความสามารถทางสมอง (I.Q.)

|  |  |  |  |
| --- | --- | --- | --- |
| ระดับ I.Q. | ความสำเร็จในการทำงาน | | |
| ดีมาก | พอใช้ | ไม่ดี |
| ระดับสูง  ระดับกลาง  ระดับต่ำ | 41  84  50 | 12  61  49 | 26  48  61 |

จงวิเคราะห์ข้อมูลดังกล่าวด้วยระดับนัยสำคัญ 5 % ว่าความสำเร็จในการทำงานขึ้นอยู่กับระดับ I.Q. หรือไม่

1. จากการสังเกตลูกกระต่ายที่มีครอกละ 3 ตัว จำนวน 80 ครอก นับจำนวนตัวผู้ได้ดังตารางต่อไปนี้

|  |  |
| --- | --- |
| จำนวนตัวผู้ในครอก | 0 1 2 3 |
| จำนวนครอก | 19 32 22 7 |

จากข้อมูลดังกล่าวจะสรุปได้หรือไม่ว่าการแจกแจงของจำนวนตัวผู้เป็นการแจกแจงทวินามโดยมีความน่าจะเป็นของการเกิดเป็นตัวผู้เท่ากับ 0.5 ใช้ระดับนัยสำคัญ 5%

1. บริษัทประกันภัยแห่งหนึ่ง ต้องการศึกษาเกี่ยวกับความเชื่อมั่นว่า อุบัติเหตุจะมากหรือน้อยขึ้นอยู่กับอายุของผู้ขับขี่ จึงได้รวบรวมจำนวนอุบัติเหตุที่เกิดขึ้นในรอบปีที่ผ่านมา และอายุของคนขับรถ จากรถที่มาขอขอรับประกันได้ข้อมูล ดังตารางต่อไปนี้

|  |  |  |  |
| --- | --- | --- | --- |
| อายุคนขับ  (ปี) | จำนวนครั้งของการเกิดอุบัติเหตุ | | |
| 0 | 1 | 2 |
| 18 – 25  26 – 35  36 – 60 | 51  173  203 | 183  104  154 | 146  57  79 |

ที่ระดับนัยสำคัญ 0.01 จงทดสอบว่าอายุของคนขับรถไม่ได้มีผลต่อจำนวนอุบัติเหตุที่เกิดขึ้น

1. ตารางข้างล่างนี้เป็นข้อมูลบันทึกการเกิดของทารกในรอบปี ซึ่งแบ่งเป็น 4 คาบ

|  |  |
| --- | --- |
| คาบ | ม.ค. – มี.ค. เม.ย. – มิ.ย. ก.ค. – ก.ย. ต.ค. – ธ.ค. |
| จำนวนทารก | 110 57 53 80 |

ด้วยระดับนัยสำคัญ 1% จะกล่าวได้หรือไม่ว่าในช่วง ม.ค. – มี.ค. มีทารกเกิดเป็น 2 เท่าของช่วงอื่นๆ

1. ตารางที่กำหนดให้ต่อไปนี้แสดงน้ำหนัก (กิโลกรัม) ของเด็กนักเรียนประถมต้นกลุ่มหนึ่ง

|  |  |
| --- | --- |
| น้ำหนัก | จำนวน |
| 9 หรือต่ำกว่า  10 – 14  15 – 19  20 – 24  25 – 29  30 – 34  35 หรือมากกว่า | 1  10  37  36  13  2  1 |

ด้วยระดับนัยสำคัญ 5% กล่าวได้หรือไม่ว่าการแจกแจงของน้ำหนักข้างต้นเป็นการแจกแจงแบบปกติที่มีค่าเฉลี่ย 20 กิโลกรัม และส่วนเบี่ยงเบนมาตรฐาน 5 กิโลกรัม