การทดสอบสำหรับตัวอย่างชุดเดียว (II)

208348 : สถิตินอนพาราเมตริก

โดย ผศ. ดร. สุคนธ์ ประสิทธิ์วัฒนเสรี
ภาควิชาสถิติ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่

เนื้อหา

- การทดสอบซึ่งใช้การแจกแจงทวินามเป็นหลัก
- **ขีดจำกัดที่ยอมรับได้**
- การทดสอบแบบไคสแควร์
- การทดสอบซึ่งอาศัยหลักสถิติของโคลโมโกรอฟ-สเมอร์นอฟ
- การทดสอบซึ่งอาศัยหลักของรันส์

ขีดจำกัดที่ยอมรับได้ (Tolerance Limits)

- **ความแตกต่างระหว่าง C.I. และ T.L.**
  - **$(1 - \alpha)100\%$ C.I. ของ P** $\Rightarrow P_L < P < P_U$
    หมายความว่า โอกาสที่ช่วง $[P_L : P_U]$ จะครอบคลุมค่าพารามิเตอร์ P เท่ากับ $1 - \alpha$
  - **$(1 - \alpha)100\%$ T.L. ของค่าตัวแปรใด ๆ**
    หมายความว่า จะมีประชากรอย่างน้อย q ส่วนที่มีค่าอยู่ในช่วงดังกล่าวด้วยความน่าจะเป็นเท่ากับ $1 - \alpha$

ขีดจำกัดที่ยอมรับได้ (Tolerance Limits)

- **$(1 - \alpha)100\%$ T.L. ของค่าตัวแปร X**

  มักนำไปประยุกต์ใช้ประมาณขนาดตัวอย่างที่เหมาะสม เช่น ต้องการสุ่มตัวอย่างขนาด n โดย $X_1, ..., X_n$ เป็นตัวแปรที่สนใจศึกษา

  หากกำหนดให้ขีดจำกัดที่ยอมรับได้ $(1 - \alpha)100\%$ T.L. ของ X คือต้องมีประชากรอย่างน้อย **q** ส่วน มีค่าระหว่าง $\mathbf{X^{(r)}}$ ถึง $\mathbf{X^{(n+1-m)}}$

  สามารถทำการคำนวณค่าของ **n** ที่จะทำให้เงื่อนไขดังกล่าวเป็นจริงได้

ขีดจำกัดที่ยอมรับได้แบบข้างเดียว (One-sided Tolerance Limits)

- **ด้านน้อย : ขีดจำกัด $[X^{(r)}, \infty]$ ($X^{(n+1-m)} = \infty$; เมื่อ m = 0)**

  อย่างน้อย q ส่วนของประชากรมีค่าน้อยกว่า $X^{(r)}$ ด้วยความน่าจะเป็น α

  หรือ อย่างน้อย q’ ส่วนของประชากรมีค่ามากกว่า $X^{(r)}$ ด้วยความน่าจะเป็น 1 - α

- **ด้านมาก : ขีดจำกัด $[-\infty, X^{(n+1-m)},]$ ($X^{(r)} = -\infty$; เมื่อ r = 0)**

  อย่างน้อย q ส่วนของประชากรมีค่ามากกว่า $X^{(n+1-m)}$ ด้วยความน่าจะเป็น α

  หรือ อย่างน้อย q’ ส่วนของประชากรมีค่าน้อยกว่า $X^{(n+1-m)}$ ด้วยความน่าจะเป็น 1 - α

Tolerance Limits

- **ข้อมูล (Data)**

  กำหนดค่า
  - ช่วงความเชื่อมั่น 1 - α
  - ค่า r (เลขจำนวนเต็มบวก)
  - ค่า m (เลขจำนวนเต็มบวก)
  - ค่า q ($0 \leq q \leq 1$)

Tolerance Limits

- **วิธีการประมาณ n**
  - **กรณีที่ 1** : เมื่อ r + m = 1 (r หรือ m เท่ากับ 0)
    ใช้**ตารางที่ 5** ประมาณ n เมื่อทราบค่า 1-α และ q
  - **กรณีที่ 2** : เมื่อ r + m = 2
    ใช้**ตารางที่ 6** ประมาณ n เมื่อทราบค่า 1-α และ q
  - **กรณีที่ 3** : นอกเหนือจากทั้ง 2 กรณีข้างต้น

$$n \approx \tfrac{1}{4}\chi^2_{1-\alpha, df=2(r+m)}\left(\tfrac{1+q}{1-q}\right) + \tfrac{1}{2}(r+m-1)$$

เมื่อ $\chi^2_{1-\alpha}$ เปิดจากตารางไคสแควร์ที่ df = 2(r+m)

Table 5 : Sample sizes for one-sided nonparametric tolerance limits

1-α	q = .500	.700	.750	...	.990
.500	1	2	3	...	69
.700	2	4	5	...	120
.750	2	4	5	...	138
...					
.999	10	20	25	...	688

n

Table 6 : Sample sizes for nonparametric tolerance limits when r+m = 2

1-α	q = .500	.700	.750	...	.990
.500	3	6	7	...	168
.700	5	8	10	...	244
.750	5	9	10	...	269
...					
.999	14	27	33	...	920

n

ตัวอย่าง 1

สินค้าที่ผลิตจากโรงงานแห่งหนึ่ง แต่ละชิ้นมีความยาวไม่เท่ากัน ดังนั้นในการสร้างกล่องบรรจุที่มีขนาดเหมาะสมเพื่อที่จะส่งสินค้าทางเรือ จึงต้องกำหนดขีดจำกัดบนและล่างไว้ (ค่าสูงสุดและค่าต่ำสุดของความยาวสินค้าที่ผลิตได้) ซึ่งภายในช่วงนี้จะต้องมั่นใจได้ 90% ว่าจะมีอย่างน้อย 80% ของสินค้าที่ผลิตมาจะบรรจุในกล่องได้ จงหาว่า n ควรเป็นเท่าใดเพื่อใช้ประมาณความยาวของสินค้า

- พิจารณาข้อมูล
  - ต้องการประมาณ n
  - กำหนดขีดจำกัดที่ยอมรับได้คือ ค่าต่ำสุด ถึง ค่าสูงสุด
    $\Rightarrow X^{(1)}$ ถึง $X^{(n)}$
    ดังนั้น r = 1 และ m = 1
  - กำหนด 1 - α = .90 และ q = .80

– กำหนดขีดจำกัดที่ยอมรับได้คือ $X^{(1)}$ ถึง $X^{(n)}$
  ดังนั้น r = 1 และ m = 1
– กำหนด 1 - α = .90 และ q = .80

- **ประมาณ n**

จากตารางที่ 6,
จะได้ n = 18

จากตารางไคสแควร์
$\chi^2_{.90,df=4} = 7.78$

ลองประมาณ n จากสูตร

$$n = \frac{1}{4}\chi^2_{1-\alpha,df=2(r+m)}\left(\frac{1+q}{1-q}\right) + \frac{1}{2}(r+m-1)$$

= ¼(7.78)(1+.80)/(1-.80) + ½(1+1-1)
= 18.003

ตัวอย่าง 2

โรงงานแห่งหนึ่งผลิตแท่งเหล็กจำนวนมาก และแยกไว้เป็นกอง ใหญ่หลาย ๆ กอง โรงงานจะต้องกำหนดค่ารับรองหรือค่าการันตี ของโรงงาน (manufacturer guarantees) เพื่อรับรองว่าอย่างน้อย 90% ของแท่งเหล็กจะมีจุดแตกหัก (breaking point) เหนือค่าค่า หนึ่งสำหรับเหล็กแต่ละกอง เนื่องจากสภาวะการณ์ในการผลิตไม่ แน่นอน ดังนั้นการกำหนดค่าการันตีสำหรับแท่งเหล็กแต่ละกองก็ ทำเป็นกอง ๆ ไป โดยการลองหักแท่งเหล็กที่สุ่มออกมาจากเหล็ก กองใหญ่ และจะกำหนดค่าการันตีเท่ากับค่าต่ำสุดของจุดแตกหักที่ ได้ในตัวอย่างแต่ละชุด จงหาว่าจะต้องเลือกตัวอย่างขนาดเท่าใด ถึงจะทำให้เชื่อมั่นได้ถึง 95% ว่าค่าการันตีที่กำหนดขึ้นนั้นถูกต้อง

- พิจารณาลักษณะข้อมูล
  – ต้องการประมาณ n
  – กำหนดขีดจำกัดที่ยอมรับได้คือ ค่าต่ำสุดของจุดแตกหัก ถึง ∞, $X^{(1)} - \infty$
  – ค่า 1 - α = .95 และ q = .90

– กำหนดขีดจำกัดที่ยอมรับได้คือ $X^{(1)}$ ถึง ∞
ดังนั้น $r = 1$ และ $m = 0$
– กำหนด $1 - \alpha = .95$ และ $q = .90$

- **ประมาณ n**

จากตารางที่ 5,
จะได้ $n = 29$

ดังนั้นจะต้องเลือกตัวอย่างเหล็กจำนวน 29 ชิ้น จะทำให้มีอย่างน้อย 90% ของเหล็กเส้นมีคุณภาพดี(อยู่ภายใต้ขีดจำกัดที่ยอมรับได้) ด้วยความน่าจะเป็น 95%

เนื้อหา

- การทดสอบซึ่งใช้การแจกแจงทวินามเป็นหลัก
- ขีดจำกัดที่ยอมรับได้
- **การทดสอบแบบไคสแควร์**
- การทดสอบซึ่งอาศัยหลักสถิติของโคลโมโกรอฟ-สเมอร์นอฟ
- การทดสอบซึ่งอาศัยหลักของรันส์

การทดสอบแบบไคสแควร์ (Chi-square test)

- ในกรณีที่สนใจศึกษาตัวแปรคุณภาพในมาตรนามบัญญัติจำนวน **c** กลุ่ม
- ต้องการเปรียบเทียบความถี่หรือจำนวนในแต่ละกลุ่มว่าเป็นไปตามที่สงสัยหรือคาดหวังไว้หรือไม่
- เรียกรูปแบบของการทดสอบลักษณะนี้ว่า **goodness of fit test** โดยทำการเปรียบเทียบระหว่าง Observed frequency และ Expected frequency ด้วย**การทดสอบไคสแควร์ (Chi-square test)**

Chi-square test

- **Data**
  - X : X_1 X_2 ... X_N
    โดยค่าของ X = {กลุ่ม 1, กลุ่ม 2, ..., กลุ่ม c}
  - การเตรียมข้อมูล
    - สร้างตารางความถี่

	กลุ่ม 1	กลุ่ม 2	...	กลุ่ม c	รวม
จำนวนค่าสังเกตหรือความถี่	O_1	O_2	...	O_c	N

โดย O_i = จำนวนค่าสังเกตหรือความถี่ของกลุ่มที่ i

Chi-square test

- **Assumptions**
  - ตัวอย่างเป็นตัวอย่างที่ได้จากการสุ่ม
  - ข้อมูลต้องอยู่ในมาตรนามบัญญัติ
- **Statistical Hypothesis**

H_0 : ไม่มีความแตกต่างระหว่างจำนวนค่าสังเกตจริง และจำนวนค่าที่คาดหวัง

H_1 : มีความแตกต่างระหว่างจำนวนค่าสังเกตจริง และจำนวนค่าที่คาดหวัง

Chi-square test

- **Statistical Test**

$$\boxed{T = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}}$$

เมื่อ O_i = จำนวนค่าสังเกตในกลุ่ม i

E_i = จำนวนค่าคาดหวังในกลุ่ม i

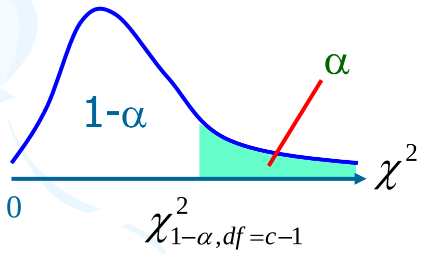
= ($N \times p_j$) (ต้องมีค่า $\geq$ 5)

p_i = ความน่าจะเป็นที่ค่าสังเกตจะอยู่ในกลุ่ม i

Chi-square test

- **เขตวิกฤต (Critical region)**

ที่ระดับนัยสำคัญ α



เขตวิกฤต คือ

$$T > \chi^2_{1-\alpha, df=c-1}$$

จะปฏิเสธ H_0 เมื่อค่า T ที่คำนวณได้ตกอยู่ในเขตวิกฤต

ตัวอย่างที่ 1

แฟนม้าแข่งมักจะมีความเชื่อว่า มีการได้เปรียบเสียเปรียบกันในระหว่างลู่วิ่งของม้า กล่าวคือ ม้าลู่ในจะได้เปรียบม้าลู่นอก หากจัดลู่วิ่งโดย ลู่ที่ 1 เป็นลู่ในสุด และลู่ที่ 8 เป็นลู่นอกสุด ที่ระดับนัยสำคัญ 0.01 จงทดสอบดูว่า ลู่วิ่งมีผลต่อการแข่งขันหรือไม่ โดยการวิเคราะห์ผลการแข่งขันที่ผ่านมา 144 เที่ยว ดังนี้

ลู่วิ่งที่	1	2	3	4	5	6	7	8	รวม
จำนวนที่ชนะ (เที่ยว)	29	19	18	25	17	10	15	11	144

- **พิจารณาข้อมูล**
  - เป็นข้อมูลนามบัญญัติจำนวน 8 กลุ่ม
  - ต้องการทดสอบลู่วิ่งมีผลต่อการแข่งขันหรือไม่

    หาก**ลู่วิ่งมีผล** ย่อมทำให้จำนวนคาดหวังของแต่ละลู่จะชนะแตกต่างกัน

    หาก**ลู่วิ่งไม่มีผล** จำนวนคาดหวังแต่ละลู่จะชนะไม่แตกต่างกัน

ตัวอย่างที่ 1

	ลู่วิ่งที่	1	2	3	4	5	6	7	8	รวม
O_i	จำนวนที่ชนะ (เที่ยว)	29	19	18	25	17	10	15	11	144
	p_i	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1
	$E_i = Np_i$	18	18	18	18	18	18	18	18	144

- **สมมติฐาน**

  H_0 : Observed freq. = Expected freq.

  H_1 : Observed freq. $\neq$ Expected freq.

- **สถิติทดสอบ** $T = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$ = [(29-18)²/18 +…+ (11-18)²/18]

  = 16.33

- **เขตวิกฤต** ที่ α = 0.01, df = 8-1 =7 **•สรุปผล** ไม่ปฏิเสธ H_0

  $\chi^2_{.99,df=7}$ = 18.48 => T > 18.48

การทดสอบแบบไคสแควร์ (Chi-square test)

- ในกรณีที่ต้องการทดสอบเกี่ยวกับฟังก์ชันการแจกแจง (Distribution function) หรือ ฟังก์ชันการแจกแจงสะสม (Cumulative distribution function) ของตัวแปร (F(x)) ว่าเป็นไปตามที่สงสัยหรือคาดหวังไว้ (F*(x)) หรือไม่
- เรียกรูปแบบของการทดสอบลักษณะนี้ว่า **goodness of fit test**

Chi-square test

- **Statistical Hypothesis**

  $H_0 : F(x) = F^*(x)$

  $H_1 : F(x) \neq F^*(x)$

  เมื่อ $F^*(x)$ = ฟังก์ชันการแจกแจงที่ต้องการทดสอบ

- **Statistical Test**

$$\boxed{T = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}}$$

เมื่อ O_i = จำนวนค่าสังเกตในกลุ่ม i
E_i = จำนวนค่าคาดหวังในกลุ่ม i
= $(N \times p_i)$ (ต้องมีค่า ≥ 5)
p_i = ความน่าจะเป็นที่ค่าสังเกตจะอยู่ในกลุ่ม i คำนวณภายใต้ $F^*(x)$

ตัวอย่างที่ 2

ข้อมูลต่อไปนี้เป็นข้อมูลที่ได้จากการสุ่มตัวอย่างมา 20 จำนวน

16.7	18.8	24.0	35.1	39.8	17.4	19.3	24.7	22.5	27.0
35.8	42.1	18.1	22.4	25.7	36.5	43.2	18.2	37.6	46.2

จงทดสอบว่าตัวอย่างนี้ถูกสุ่มมาจากประชากรที่มีการแจกแจงปกติซึ่งมีค่าเฉลี่ยเท่ากับ 30 และความแปรปรวนเท่ากับ 100 หรือไม่ ที่ระดับนัยสำคัญ 0.05

พิจารณาข้อมูล

- ต้องการทดสอบลักษณะการแจกแจงของข้อมูลว่ามีการแจกแจงปกติหรือไม่

  $X \sim N(\text{mean} = 30, \text{variance} = 100)$?

- ต้องจัดข้อมูลให้เป็นข้อมูลกลุ่ม

ตัวอย่างที่ 2

- **สมมติฐาน**

  H_0 : ข้อมูลมีการแจกแจงปกติ หรือ $F(x) = F^*(x)$

  H_1 : ข้อมูลไม่มีการแจกแจงปกติ หรือ $F(x) \neq F^*(x)$

  เมื่อ $F^*(x)$ = Normal distribution function

- **สถิติทดสอบ**

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

เริ่มต้นด้วยการสร้างตารางแจกแจงความถี่ โดยจะทำการแบ่งข้อมูลออกเป็นกลุ่ม

ตัวอย่างที่ 2

แนวทางการจัดกลุ่มข้อมูลเพื่อตรวจสอบการแจกแจงปกติ

<u>แนวทางที่ 1</u>

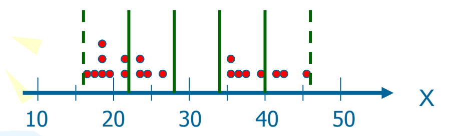
- แบ่งข้อมูลที่สังเกตได้เป็น k ชั้น
- นับความถี่ข้อมูลแต่ละชั้น – O_i
- คำนวณ P_i แต่ละชั้นภายใต้ $X \sim N(\mu=30, \sigma=10)$
- คำนวณ $E_i = NP_i$

<u>แนวทางที่ 2</u>

- แบ่ง $X \sim N(\mu=30, \sigma=10)$ ออกเป็น k ชั้นซึ่งมีพื้นที่แต่ละชั้นเท่า ๆ กัน ($P_i = 1/k$)
- คำนวณ $E_i = NP_i$
- คำนวณค่า X_p ซึ่งเป็นจุดแบ่งหรือขอบเขตของแต่ละชั้น
- ภายใต้ค่า X_p นับความถี่ข้อมูลแต่ละชั้น – O_i

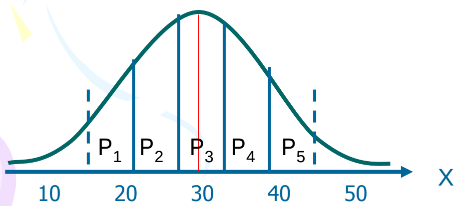
ตัวอย่างที่ 2

- พิจารณาลักษณะข้อมูล



- การแจกแจงของ X ภายใต้ H_0

X ~ N(mean = 30, variance = 100)



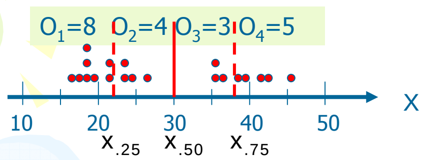
แนวทางที่ 1

- แบ่งข้อมูลที่สังเกตได้เป็น k ชั้น
- นับความถี่ข้อมูลแต่ละชั้น – O_i
- คำนวณ P_i แต่ละชั้นภายใต้ $X \sim N(\mu=30, \sigma=10)$
- คำนวณ $E_i = NP_i$

ขอบเขตชั้น	O_i	P_i	E_i
16.65 – 22.65	8	P_1	E_1
22.65 – 28.65	4	P_2	E_2
28.65 – 34.65	0	P_3	E_3
34.65 – 40.65	5	P_4	E_4
40.65 – 46.65	3	P_5	E_5
รวม	20		

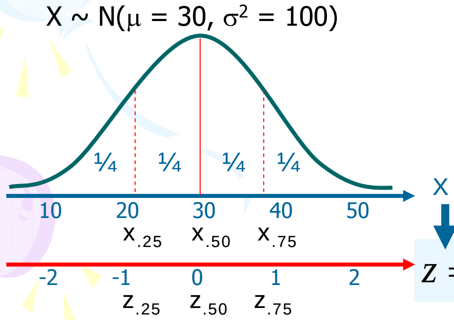
ตัวอย่างที่ 2

- พิจารณาลักษณะข้อมูล



- การแจกแจงของ X ภายใต้ H_0

$X \sim N(\mu = 30, \sigma^2 = 100)$



$$Z = \frac{X-\mu}{\sigma}$$

แนวทางที่ 2

- แบ่ง $X \sim N(\mu=30, \sigma=10)$ ออกเป็น k ชั้นซึ่งมีพื้นที่แต่ละชั้นเท่า ๆ กัน ($P_i = 1/k$)
- คำนวณ $E_i = NP_i$
- คำนวณค่า X_p ซึ่งเป็นจุดแบ่งหรือขอบเขตของแต่ละชั้น
- ภายใต้ค่า X_p นับความถี่ข้อมูลแต่ละชั้น – O_i

แบ่งข้อมูลแบบควอไทล์ (Quartile)

- $E_1 = \ldots = E_4 = 20*1/4 = 5$
- คำนวณค่า X_p
  - $z_{.75} = 0.674$
    - $\Rightarrow x_{.75} = 30 + 10(0.674) = 36.74$
  - $z_{.50} = 0$
    - $\Rightarrow x_{.50} = 30$
  - $z_{.25} = -0.674$
    - $\Rightarrow x_{.25} = 30 + 10(-0.674) = 23.26$

ตัวอย่างที่ 2

ขอบเขตชั้น	O_i	E_i
< 23.26	8	5
23.26 – 30.00	4	5
30.00 – 36.74	3	5
> 36.74	5	5

- **สถิติทดสอบ**

  $T = [(8-5)^2/5 + (4-5)^2/5 + (3-5)^2/5 + (5-5)^2/5] = 2.8$

- **เขตวิกฤต** ที่ $\alpha = 0.05$, df = 4-1 =3

  $\chi^2_{.95,df=3} = 7.8 \Rightarrow T > 7.8$

- **สรุปผล** ไม่ปฏิเสธ H_0

  ข้อมูลมีการแจกแจงปกติ

ตัวอย่างที่ 2 (ลองแนวทางที่ 1)

- **คำนวณค่า O_i และ E_i**

ขีดจำกัดชั้น	ขอบเขตชั้น	O_i
16.7 – 22.6	16.65 – 22.65	8
22.7 – 28.6	22.65 – 28.65	4
28.7 – 34.6	28.65 – 34.65	0
34.7 – 40.6	34.65 – 40.65	5
40.7 – 46.6	40.65 – 46.65	3
	รวม	20

ขยายขอบเขตชั้น	p_i
<16.65	P(X<16.65)
16.65 – 22.65	P(16.65<X<22.65)
22.65 – 28.65	P(22.65<X<28.65)
28.65 – 34.65	P(28.65<X<34.65)
34.65 – 40.65	P(34.65<X<40.65)
40.65 – 46.65	P(40.65<X<46.65)
>46.65	P(X>46.65)

ตัวอย่างที่ 2

- คำนวณค่า O_i และ E_i

$Z = (X - 30)/10$

ขยายขอบเขตชั้น	O_i	$p_i = P(x_i)$	$p_i = P(z_i)$	p_i
<16.65	-	P(X<16.65)	P(Z<-1.335)	.09
16.65 – 22.65	8	P(16.65<X<22.65)	P(-1.335<X<-0.735)	.14
22.65 – 28.65	4	P(22.65<X<28.65)	P(-0.735<X<-0.135)	.22
28.65 – 34.65	0	P(28.65<X<34.65)	P(-0.135<X<0.465)	.23
34.65 – 40.65	5	P(34.65<X<40.65)	P(0.465<X<1.065)	.18
40.65 – 46.65	3	P(40.65<X<46.65)	P(1.065<X<1.665)	.09
>46.65	-	P(X>46.65)	P(X>1.665)	.05

$P(X < 16.65) = P(Z < (16.65 - 30)/10) = P(Z < -1.335)$

ตัวอย่างที่ 2

- คำนวณค่า O_i และ E_i

ขยายขอบเขตชั้น	O_i		p_i	$E_i=Np_i$	$E_i=Np_i$
<16.65	-		.09	1.82	
16.65 – 22.65	8	12	.14	2.80	8.92
22.65 – 28.65	4		.22	4.30	
28.65 – 34.65	0		.23	4.65	4.65
34.65 – 40.65	5		.18	3.55	
40.65 – 46.65	3	8	.09	1.91	6.42
>46.65	-		.05	0.96	

- **สถิติทดสอบ**

$T = [(12-8.92)^2/8.92 + (0-4.65)^2/4.65 + (8-6.42)^2/6.42] = 6.10$

- **เขตวิกฤต** ที่ $\alpha = 0.05$, df = 3-1 =2

$\chi^2_{.95,df=2} = 6.0 \Rightarrow T > 6.0$

- **สรุปผล** ปฏิเสธ H_0

ข้อมูลไม่มีการแจกแจงปกติ

ตัวอย่างที่ 3

โปรแกรมคอมพิวเตอร์หนึ่งใช้สั่งพิมพ์ตัวเลขซึ่งเป็นเลขสุ่ม (random digits) 0 – 9 ถ้าโปรแกรมนี้เขียนถูกต้องตามจุดประสงค์แล้ว จะต้องพิมพ์ตัวเลขออกมาอย่างสุ่ม โดยตัวเลขที่ได้ต้องเป็นอิสระต่อกันและมีการแจกแจงเหมือนกัน แต่ละตัวเลข 0, 1, ..., 9 จะมีโอกาสเกิดขึ้นเท่า ๆ กัน (p = 0.1) การทดสอบทำได้โดยแบ่งตัวเลขซึ่งพิมพ์ออกมาเป็นตัวเลขยาว ๆ ติดกันออกเป็นกลุ่ม ๆ ละ 10 ตัว และนับจำนวนเลขคู่ในแต่ละกลุ่ม จำนวนเลขคู่แต่ละกลุ่มควรมีการแจกแจงแบบทวินาม ด้วย n = 10 และ p = 0.5

สมมติตัวเลขที่พิมพ์ออกมาเป็นดังนี้

1578748416	4705188926	6936349612
4563813213	0282868892	...

จงทดสอบว่าตัวเลขที่ปรากฏออกมาเป็นตัวเลขสุ่มที่ระดับนัยสำคัญ .05

พิจารณาข้อมูล

- ต้องการทดสอบตัวเลขที่ปรากฏเป็นเลขสุ่มหรือไม่
  เมื่อให้ X = จำนวนเลขคู่ที่ปรากฏในแต่ละกลุ่ม
  หากตรวจสอบพบว่า X ~ b(n=10, p=0.5) แสดงว่าตัวเลขที่ได้เป็นเลขสุ่ม

ตัวอย่างที่ 3

- **สมมติฐาน**

  H_0 : ตัวเลขเป็นเลขสุ่ม หรือ X มีการแจกแจงทวินาม

  H_1 : ตัวเลขไม่เป็นเลขสุ่ม หรือ X ไม่มีการแจกแจงทวินาม

- **สถิติทดสอบ**

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

เริ่มต้นด้วยการสร้างตารางแจกแจงความถี่ของค่า X

ตัวอย่างที่ 3

- **คำนวณค่า O_i**

แบ่งกลุ่มตัวเลขสุ่มเป็น 30 กลุ่ม ๆ ละ 10 ตัว

และจาก X = จำนวนเลขคู่ในแต่ละกลุ่ม, x = 0, 1, 2, ..., 10

X	0	1	2	3	4	5	6	7	8	9	10	รวม
O_i	0	4	1	1	3	14	5	1	0	1	0	30

- นับจำนวนเลขคู่ในแต่ละกลุ่ม (x)

1578748416 **x=5** 4705188926 **x=6** 6936349612 **x=5**
4563813213 **x=4** 0282868892 **x=9** 3928057043 **x=5**
5101259393 **x=2** 9837006785 **x=5** 3011679938 **x=3**
7122863085 **x=6** 6528271107 **x=5** 2956427027 **x=6**
...
4659773922 **x=4** 9246724287 **x=7** 8326143939 **x=4**

ตัวอย่างที่ 3

- **คำนวณค่า $E_i = NP_i$**

จาก X ~ b(n=10, p=.5) คำนวณค่า $P_i = P(X=x_i)$

X	0	1	2	3	4	5	6	7	8	9	10	รวม
O_i	0	4	1	1	3	14	5	1	0	1	0	30
P_i	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010	1.00
E_i	.030	.294	1.317	3.516	6.153	7.383	6.153	3.516	1.317	.294	.030	30

6 2

5.157 5.157

เช่น P(X = 0) = .0010 P(X = 1) = .0098 ... P(X = 10) = .0010

ตัวอย่างที่ 3

X	O_i	E_i
≤ 3	6	5.157
4	3	6.153
5	14	7.838
6	5	6.153
≥ 7	2	5.157

- **สถิติทดสอบ**

  $T = [(6-5.157)^2/5.157 + ... + (2-5.157)^2/5.157] = 9.84$

- **เขตวิกฤต** ที่ $\alpha = 0.05$, df = 5-1 =4

  $\chi^2_{.95,df=4} = 9.49 \Rightarrow T > 9.49$

- **สรุปผล** ปฏิเสธ H_0

  X ไม่มีการแจกแจงทวินาม หรือตัวเลขที่ได้ไม่เป็นเลขสุ่ม

ตัวอย่างที่ 4

ในการสุ่มเลข 2 หลัก (two digit) จากสมุดรายนามผู้ใช้โทรศัพท์ และจะใช้การทดสอบไคสแควร์เพื่อทดสอบดูว่า ตัวเลขเหล่านี้เป็นค่าสังเกตที่ได้มาจากตัวแปรสุ่มที่มีการแจกแจงปกติหรือไม่ ที่ระดับนัยสำคัญ .05

หลังจากเรียงตัวเลขจากน้อยไปมากได้เป็นดังนี้

23	23	24	27	29	31	32	33	33	35
36	37	40	42	43	43	44	45	48	48
54	54	56	57	57	58	58	58	58	59
61	61	62	63	64	65	66	68	68	70
73	73	74	75	77	81	87	89	93	97

พิจารณาข้อมูล

- ต้องการทดสอบตัวเลขที่ปรากฏมีการแจกแจงปกติหรือไม่ นั่นคือ ตรวจสอบว่า $X \sim N(\mu,\sigma)$

ตัวอย่างที่ 4

- **สมมติฐาน**

  H_0 : ข้อมูลมีการแจกแจงปกติ หรือ F(x) = F*(x)

  H_1 : ข้อมูลไม่มีการแจกแจงปกติ หรือ $F(x) \neq F^*(x)$

  เมื่อ F*(x) = Normal distribution function

- **สถิติทดสอบ**

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

เริ่มต้นด้วยการสร้างตารางแจกแจงความถี่ โดยจะทำการแบ่งข้อมูลออกเป็นกลุ่ม

ตัวอย่างที่ 4

แนวทางที่ 1

- แบ่งข้อมูลที่สังเกตได้เป็น 4 ชั้น
- นับความถี่ข้อมูลแต่ละชั้น – O_i
- คำนวณ P_i แต่ละชั้นภายใต้ $X \sim N(\mu=?, \sigma=?)$
- คำนวณ $E_i = NP_i$

- **คำนวณค่า O_i และ E_i**

ขอบเขตชั้น (x)	O_i
[20 – 40)	14
[40 – 60)	18
[60 – 80)	15
[80 – 100)	5
รวม	50

ขยายขอบเขตชั้น	p_i
<20	P(X<20)
20 – 40	P(20<X<40)
40 – 60	P(40<X<60)
60 – 80	P(60<X<80)
80 – 100	P(80<X<100)
>100	P(X>100)

จะแปลง X -> Z

ต้องทราบ μ, σ

- ประมาณ μ ด้วย $\bar{x}$
- ประมาณ σ ด้วย s

$$\bar{x} = \frac{23+\ldots+97}{50} = 55.04$$

$$s = \sqrt{\frac{(23^2+\ldots+97^2)-50(55.04^2)}{49}} = 19.00$$

ตัวอย่างที่ 4

- **คำนวณค่า O_i และ E_i**

ขยายขอบเขตชั้น	O_i	$p_i = P(x_i)$	$p_i = P(z_i)$	$F(z_i)$	P_i
<20	-	P(X<20)	P(Z<-1.84)	.0329	.0329
20 – 40	12	P(20<X<40)	P(-1.84<X<-0.79)	.2148	.1819
40 – 60	18	P(40<X<60)	P(-0.79<X<0.26)	.6026	.3878
60 – 80	15	P(60<X<80)	P(0.26<X<1.31)	.9049	.3023
80 – 100	5	P(80<X<100)	P(1.31<X<2.37)	.9911	.0862
>100	-	P(X>100)	P(X>2.37)	1.00	.0089

$P(X < 20) = P(Z < (20 - 55.04)/19) = P(Z < -1.84)$

ตัวอย่างที่ 4

- **คำนวณค่า O_i และ E_i**

ขยายขอบเขตชั้น	O_i	p_i	$E_i=Np_i$	
<20	-	.0329	1.645	10.74
20 – 40	12	.1819	9.095	
40 – 60	18	.3878	19.39	
60 – 80	15	.3023	15.115	
80 – 100	5	.0862	4.31	4.755
>100	-	.0089	0.445	

- **สถิติทดสอบ**

$$T = [(12-10.74)^2/10.74 + \ldots + (5-4.755)^2/4.755] = 0.261$$

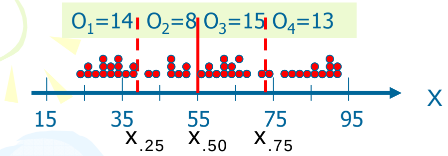
- **เขตวิกฤต** ที่ $\alpha = 0.05$, $df = c - m - 1 = 4-2-1 = 1$

$\chi^2_{.95,df=1} = 3.84 \Rightarrow T > 3.84$

- **สรุปผล** ไม่ปฏิเสธ H_0 นั่นคือข้อมูลมีการแจกแจงปกติ

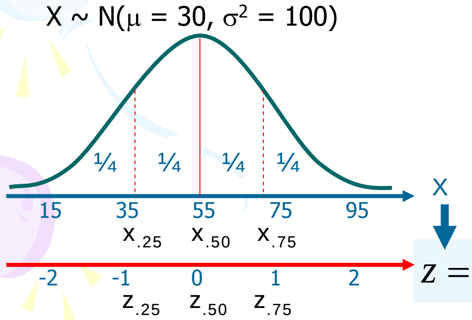
ตัวอย่างที่ 4

- **พิจารณาลักษณะข้อมูล**



- **การแจกแจงของ X ภายใต้ H_0**

$X \sim N(\mu = 30, \sigma^2 = 100)$



$$Z = \frac{X-\mu}{\sigma}$$

<u>แนวทางที่ 2</u>

- แบ่ง $X \sim N(\mu \sim 55.04, \sigma \sim 19)$ ออกเป็น 4 ชั้น ซึ่งมีพื้นที่แต่ละชั้นเท่า ๆ กัน ($P_i = 1/4$)
- คำนวณ $E_i = NP_i = 12.5$
- คำนวณค่า X_p ซึ่งเป็นจุดแบ่งหรือขอบเขตของแต่ละชั้น

  - $z_{.75} = 0.674$

    $\Rightarrow x_{.75} = 55.04 + 19(0.674) = 67.846$

  - $z_{.50} = 0$

    $\Rightarrow x_{.50} = 55.04$

  - $z_{.25} = -0.674$

    $\Rightarrow x_{.25} = 55.04 + 19(-0.674) = 42.234$

- ภายใต้ค่า X_p นับความถี่ข้อมูลแต่ละชั้น – O_i

ตัวอย่างที่ 4

ขอบเขตชั้น	O_i	E_i
< 23.26	14	12.5
23.26 – 30.00	8	12.5
30.00 – 36.74	15	12.5
> 36.74	13	12.5

- **สถิติทดสอบ**

$T = [(14-12.5)^2/12.5 + \ldots + (13-12.5)^2/12.5] = 2.32$

- **เขตวิกฤต** ที่ $\alpha = 0.05$, df = 4-2-1 =1

$\chi^2_{.95,df=1} = 3.84 \Rightarrow T > 3.84$

- **สรุปผล** ไม่ปฏิเสธ H_0

ข้อมูลมีการแจกแจงปกติ

Chi-square test for contingency table

- **ตารางการจรณ์ (Contingency table)**

	X	Y	Z	
A	O_{11}	O_{12}	O_{13}	r_1
B	O_{21}	O_{22}	O_{23}	r_2
	c_1	c_2	c_3	

ตารางขนาด 2x3

O_{ij} = ความถี่สังเกตได้ (observed freq.)

$\sum\sum O_{ij} = \sum c_j = \sum r_i = N$

Chi-square test of Homogeneity

- **Data**
  - r independent samples (X_i, Y_i, …)

X	X_1	X_2	…	X_{n1}	โดย X = {-, 0, +}
Y	Y_1	Y_2	…	Y_{n2}	Y = {-, 0, +}
…					

  - การเตรียมข้อมูล
    - สร้างตารางการจรณ์

	-	0	+	
X	O_{11}	O_{12}	O_{13}	n_1
Y	O_{21}	O_{22}	O_{23}	n_2
	c_1	c_2	c_3	$N = n_1 + n_2$

O_{ij} = ความถี่สังเกตได้ (observed freq.)

Chi-square test

- **Assumptions**
  - ตัวอย่างแต่ละชุดเป็นตัวอย่างสุ่ม
  - ตัวอย่างแต่ละชุดมีความเป็นอิสระกัน
  - ค่าสังเกตแต่ละค่าอาจถูกจัดให้อยู่ในชั้นใดชั้นหนึ่งใน c ชั้นนั้น

Chi-square test

- **Statistical Hypothesis**

  H_0 : ความน่าจะเป็นหรือสัดส่วนทั้งหลายในสดมภ์เดียวกันมีค่าเท่า ๆ กัน

  H_1 : ความน่าจะเป็นอย่างน้อย 2 ค่า ในสดมภ์เดียวกันมีค่าไม่เท่ากัน

Chi-square test

- **Statistical Test**

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

เมื่อ E_{ij} = (n_i x c_j)/N (ต้องมีค่า $\geq$ 5)

r = จำนวนแถว

c = จำนวนคอลัมน์

กรณีตาราง 2x2

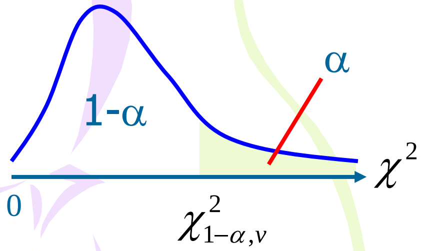
$$T=\frac{N(O_{11}O_{22}-O_{12}O_{21})^2}{n_1 n_2 c_1 c_2}$$

	-	+	
X	O_{11}	O_{12}	n_1
Y	O_{21}	O_{22}	n_2
	c_1	c_2	

Chi-square test

- **เขตวิกฤต (Critical region)**

ที่ระดับนัยสำคัญ α



เขตวิกฤต คือ

$$\chi^2 > \chi^2_{1-\alpha,v}$$

จะปฏิเสธ H_0 เมื่อค่า T ที่คำนวณได้ตกอยู่ในเขตวิกฤต

ตัวอย่างที่ 5

ในการรายงานเกี่ยวกับการมีและไม่มีภาวะการหายใจลำบาก (RDS) ของเด็กเล็ก 2 กลุ่ม โดยกลุ่มที่ 1 เป็นเด็กที่ถุงน้ำคร่ำของแม่แตกก่อนคลอด 24 ชั่วโมง หรือน้อยกว่านั้น ซึ่งมีจำนวน 42 คน และกลุ่มที่ 2 เป็นเด็กที่ถุงน้ำคร่ำของแม่แตกก่อนคลอดมากกว่า 24 ชั่วโมง ซึ่งมีจำนวน 22 คน ดังตาราง

	มี	ไม่มี	รวม
กลุ่ม 1	27	15	42
กลุ่ม 2	7	15	22
รวม	34	30	64

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่า ประชากรทั้ง 2 กลุ่มไม่แตกต่างกัน ในด้านการมี RDS

ตัวอย่างที่ 5

	มี	ไม่มี	รวม
X	27	15	42
Y	7	15	22
รวม	34	30	64

- **สมมติฐาน**

  H_0 : $P_X(RDS) = P_Y(RDS)$

  H_1 : $P_X(RDS) \neq P_Y(RDS)$

หรือ H_0 : สัดส่วน RDS ในเด็กทั้งสองกลุ่มเท่ากัน

H_1 : สัดส่วน RDS ในเด็กทั้งสองกลุ่มแตกต่างกัน

- **สถิติทดสอบ**

$$T = \frac{64(27\times15-15\times7)^2}{(34)(30)(42)(22)} = 6.112$$

- **เขตวิกฤต** ที่ α = 0.05, df=1

  T > 3.8

- **สรุปผล** ปฏิเสธ H_0

ตัวอย่าง 6

ในการวิจัยครั้งหนึ่ง ผู้วิจัยต้องการทราบว่าเพศหญิงกับเพศชายมีความเชื่อโชคลางต่างกันหรือไม่ ผลการสำรวจปรากฏดังตาราง

เพศ	ระดับความเชื่อ					รวม
	มากที่สุด	มาก	เฉย ๆ	ไม่เชื่อ	ไม่เชื่อเลย	
ชาย	13	27	25	47	8	120
หญิง	50	63	11	16	10	150
รวม	63	90	36	63	18	270

จงทดสอบสมมุติฐานที่ระดับนัยสำคัญ 0.01

พิจารณาลักษณะข้อมูล:

- ประชากร 2 กลุ่ม : ชาย และ หญิง
- ต้องการเปรียบเทียบระดับความเชื่อในเพศชายกับหญิง

ตัวอย่าง 6

สมมติฐานทางสถิติ

H_0 : ระดับความเชื่อในเพศชายและหญิงไม่แตกต่างกัน

H_1 : ระดับความเชื่อในเพศชายและหญิงแตกต่างกัน

สถิติทดสอบ

$$T = \sum_{i=1}^{2}\sum_{j=1}^{5}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

ตัวอย่าง 6

ตาราง O_{ij}

13	27	25	47	8	120
50	63	11	16	10	150
63	90	36	63	18	270

คำนวณค่า E_{ij}

E_{11} = (120x63)/270

E_{12} = (120x90)/270

E_{15} = 120–28–40-16-28

28	**40**	**16**	**28**	**8**	**120**
35	**50**	**20**	**35**	**10**	**150**
63	**90**	**36**	**63**	**18**	**270**

E_{21} = 63-28

E_{22} = 90-40

คำนวณค่า T

$$T = \frac{(13-28)^2}{28} + \ldots + \frac{(10-10)^2}{10} = 54.39$$

ตัวอย่าง 6

H_0: ระดับความเชื่อในชายและหญิงไม่แตกต่าง

T = 54.39

เขตวิกฤต

α= 0.01, ν = (r–1)(c–1) = 4

$\chi^2_{0.99,\,4} = 13.03$

ดังนั้น T > 13.03

สรุปผล **ปฏิเสธ H_0**

ที่ α = 0.01 สรุปได้ว่าระดับความเชื่อระหว่างเพศชายและหญิงแตกต่างกัน

Chi-square Test for Independent

- **Data**
  - 1 sample with 2 variables (X, Y)

X	X_1	X_2	…	X_{n1}	โดย X = {a1, a2, a3}
Y	Y_1	Y_2	…	Y_{n2}	Y = {-, 0, +}

…

  - การเตรียมข้อมูล
    - สร้างตารางการจรณ์

	-	0	+	
a1	O_{11}	O_{12}	O_{13}	R_1
a2	O_{21}	O_{22}	O_{23}	R_2
a3	O_{31}	O_{32}	O_{33}	R_3
	C_1	C_2	C_3	N

Chi-square test

- **Assumptions**
  - ตัวอย่างแต่ละชุดเป็นตัวอย่างสุ่ม
  - ตัวอย่างแต่ละชุดมีความเป็นอิสระกัน
  - ค่าสังเกตแต่ละค่าจะถูกจัดกลุ่มภายใต้ตัวแปร 2 ตัว โดย

    อาจถูกจัดให้อยู่ในชั้นใดชั้นหนึ่งใน r ชั้นของตัวแปร1

    อาจถูกจัดให้อยู่ในชั้นใดชั้นหนึ่งใน c ชั้นของตัวแปร2

Chi-square test

- **Statistical Hypothesis**

  H_0 : ตัวแปรทั้งสอง หรือลักษณะทั้งสองเป็นอิสระกัน

  H_1 : ตัวแปรทั้งสอง หรือลักษณะทั้งสองไม่เป็นอิสระ

หรือ

H_0 : ตัวแปรทั้งสอง หรือลักษณะทั้งสองไม่มีความสัมพันธ์กัน

H_1 : ตัวแปรทั้งสอง หรือลักษณะทั้งสองมีความสัมพันธ์กัน

Chi-square test

- **Statistical Test**

$$\boxed{T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}, \quad \upsilon = (r-1)(c-1)}$$

เมื่อ $E_{ij} = (n_i \text{ x } c_j)/N$ (ต้องมีค่า ≥ 5)

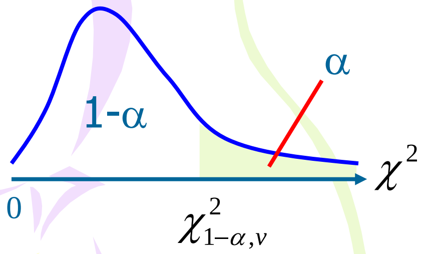
r = จำนวนแถว

c = จำนวนคอลัมน์

Chi-square test

- **เขตวิกฤต (Critical region)**

ที่ระดับนัยสำคัญ α



เขตวิกฤต คือ

$$\chi^2 > \chi^2_{1-\alpha,v}$$

จะปฏิเสธ H_0 เมื่อค่า T ที่คำนวณได้ตกอยู่ในเขตวิกฤต

ตัวอย่าง 8

นักวิทยาศาสตร์การแพทย์มีข้อสงสัยว่า การบริโภคกาแฟมีส่วนทำให้เป็นโรคหัวใจหรือไม่ จึงได้สุ่มตัวอย่างประชาชนมา 1040 ราย จำแนกตามจำนวนกาแฟที่บริโภคต่อวัน และการเป็นและไม่เป็นโรคหัวใจ ได้ดังตาราง

โรคหัวใจ	จำนวนที่บริโภคต่อวัน (ถ้วย)				รวม
	0	1 – 2	3 – 4	≥ 5	
เป็น	4	17	17	9	47
ไม่เป็น	185	457	226	31	993
รวม	189	474	243	134	1040

จงทดสอบสมมุติฐานที่ระดับนัยสำคัญ 0.05 ว่าการบริโภคกาแฟมีส่วนทำให้เป็นโรคหัวใจหรือไม่

พิจารณาลักษณะข้อมูล:

- ตัวแปร 2 ตัว : ปริมาณการบริโภคกาแฟ และการป่วยเป็นโรคหัวใจ
- ต้องการทดสอบว่า การบริโภคกาแฟมีส่วนทำให้เป็นโรคหัวใจ

ตัวอย่าง 8

สมมติฐานทางสถิติ

H_0 : การบริโภคกาแฟไม่มีผลต่อการป่วยเป็นโรคหัวใจ

H_1 : การบริโภคกาแฟมีผลต่อการป่วยเป็นโรคหัวใจ

สถิติทดสอบ

$$T = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

T = (4-8.54)²/8.54 +...+(31-127.94)²/127.94 = 8.44

ตาราง O_{ij}

4	17	17	9	47
185	457	226	31	993
189	474	243	134	1040

ตาราง E_{ij}

8.54	21.42	10.98	6.06	47
180.46	452.58	232.02	127.94	993
189	474	243	134	1040

ตัวอย่าง 8

H_0: การบริโภคกาแฟไม่มีผลต่อการเป็นโรคหัวใจ

T = 8.44

เขตวิกฤต α = 0.05, v = (r–1)(c–1) = 3

$\chi^2_{0.95,\,3} = 7.8$ ดังนั้น T > 7.8

สรุปผล ปฏิเสธ H_0

ที่ α = 0.05 สรุปได้ว่าการบริโภคกาแฟมีผลต่อการป่วยเป็นโรคหัวใจ

การวัดความไม่เป็นอิสระ (Measure of Dependence)

- **เป้าหมาย**

  ทำการวิเคราะห์ขนาดความไม่เป็นอิสระหรือการขึ้นต่อกัน (dependence) ระหว่างตัวแปรคุณภาพ 2 ตัว

  โดยการประมาณค่า **degree of dependence**

Degree of Dependence

- **วิธีการประมาณ**

  อาศัยค่าตัวสถิติ $T = \sum\sum(O - E)^2/E$ ช่วยในการประมาณค่า degree of dependence ด้วยการคำนวณค่า

  - **Chi-square probability**
  - **Contingency coefficient**

Degree of Dependence

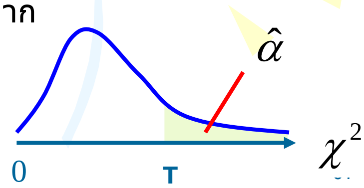
- **Chi-square probability**

จากค่าตัวสถิติ $T = \sum\sum(O - E)^2/E$ ที่ df = (r-1)(c-1) คำนวณค่า p-value หรือ ระดับวิกฤต (critical level)

$$\boxed{P(\chi^2 > T \mid df=(r-1)(c-1)) = \alpha\text{^ or p-value}}$$

ถ้าค่า α^ มีค่าน้อย ⇒ ขึ้นต่อกันมาก

ถ้าค่า α^ มีค่ามาก ⇒ อิสระต่อกันมาก



Degree of Dependence

- **Chi-square probability**

จากตัวอย่างที่ 8 ค่าสถิติ T = 8.44, df = 3 อาศัยตารางไคสแควร์

$\chi^2_{.95} = 7.8$ และ $\chi^2_{.975} = 9.5$

$P(\chi^2 > 8.44 \mid df = 3) \approx 0.04$

แสดงว่าการบริโภคกาแฟกับการเป็นโรคหัวใจขึ้นต่อกันมาก

Degree of Dependence

- **Contingency coefficient**

  จากค่าตัวสถิติ $T = \sum\sum(O - E)^2/E$ คำนวณค่า Contingency coefficient (R) ได้ 3 แบบ

  - **Cramer's contingency coefficient, R_1**

    ทำการเทียบค่า T กับค่า T_{max} โดย $T_{max} = N(q - 1)$ เมื่อ $q = \min(r, c)$

    $R_1 = T/N(q - 1)$

    โดย $0 \leq R_1 \leq 1$ ถ้า $R_1 \to 0$ แสดงว่า dependence น้อย

    $R_1 \to 1$ แสดงว่า dependence มาก

  - **Pearson's contingency coefficient, R_2**

    $R_2 = \sqrt{\frac{T}{N+T}}$ โดย $0 \leq R_2 \leq 1$

  - **Pearson's Mean-Square contingency coefficient. R_3**

    $R_3 = T/N$ โดย $0 \leq R_3 \leq q - 1$

Degree of Dependence

- **Contingency coefficient, R**

  จากตัวอย่างที่ 8 ค่าสถิติ T = 8.44, N = 1040,

  r = 2, c = 4

  คำนวณค่า R

  $R_1 = 8.44/1040(2\text{-}1) = .0081$

  $$R_2 = \sqrt{\frac{8.44}{1040+8.44}} = .0897$$

  $R_3 = 8.44/1040 = .0081$

  แสดงว่าการบริโภคกาแฟกับการเป็นโรคหัวใจขึ้นต่อกันน้อย