6 การประยุกต์ไคสแควร์
CHI-SQUARE APPLICATIONS

กระบวนวิชา 208263 : สถิติเบื้องต้น

โดย... ผศ. ดร. สุคนธ์ ประสิทธิ์วัฒนเสรี
ภาควิชาสถิติ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่

เนื้อหา

- การทดสอบภาวะรูปดีโดยใช้ χ^2 - test
  - การทดสอบอัตราส่วน
  - การทดสอบการแจกแจงของข้อมูล
- การทดสอบความเป็นอิสระโดยใช้ χ^2 - test

วัตถุประสงค์ :

นศ. เข้าใจหลักการทดสอบไคสแควร์ และประยุกต์ใช้อย่างถูกต้องเหมาะสม

2

Chi-square test

- **การทดสอบภาวะรูปดี (goodness of fit test)**
  ใช้วิเคราะห์ข้อมูลกับตัวแปรคุณภาพ 1 ตัว
  - การทดสอบอัตราส่วน
  - การทดสอบการแจกแจงของข้อมูล
- **การทดสอบความเป็นอิสระ (test for independence)**
  ใช้วิเคราะห์ข้อมูลเมื่อมีตัวแปรคุณภาพ 2 ตัว

เป็นการทดสอบสมมุติฐานเท่านั้น

3

ขั้นตอนการทดสอบสมมติฐานทางสถิติ

- ตั้งสมมุติฐานทางสถิติ
- เลือกตัวสถิติทดสอบ + คำนวณค่าสถิติ
- กำหนดเขตวิกฤต
- สรุปผล

4

การทดสอบภาวะรูปดี
- การทดสอบอัตราส่วน
- การทดสอบการแจกแจง

5

การทดสอบอัตราส่วน

ลักษณะข้อมูล : ข้อมูลคุณภาพ k กลุ่ม
ค่าสังเกตที่สนใจ : ความถี่ในแต่ละกลุ่ม (O_i)

เช่น หมู่เลือดของตัวอย่างนักศึกษา 10 คน

คนที่	Group
1	O
2	A
3	A
4	O
5	A

คนที่	Group
6	B
7	B
8	AB
9	O
10	B

Group	ความถี่	O_i
O	3	O_1
A	3	O_2
B	3	O_3
AB	1	O_4

6

การทดสอบอัตราส่วน

สมมุติฐานทางสถิติ

H_0	H_1
อัตราส่วนค่าสังเกต**เท่ากับ**อัตราส่วนที่ต้องการทดสอบ	อัตราส่วนค่าสังเกต**ไม่เท่ากับ**อัตราส่วนที่ต้องการทดสอบ

เช่น

H_0 **:** อัตราส่วนนศ.ที่มีหมู่เลือด O:A:B:AB = 2:1:1:1

H_1 **:** อัตราส่วนนศ.ที่มีหมู่เลือด O:A:B:AB $\neq$ 2:1:1:1

การทดสอบอัตราส่วน

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \quad \upsilon = k - 1 \qquad \chi^2 = \sum_{i=1}^{k} \frac{O_i^2}{E_i} - N$$

เมื่อ O_i = ความถี่ที่สังเกตได้ (observed frequency) ของกลุ่ม i (i = 1, …, k)

E_i = ความถี่คาดหวัง (expected frequency) ของกลุ่ม i = Np_i

p_i = โอกาสที่จะเกิดหรือพบกลุ่ม i ภายใต้ H_0

N = ขนาดตัวอย่างที่ใช้

เงื่อนไข $\sum O_i = \sum E_i = N$ $\quad \sum p_i = 1$ $\quad E_i \geq 5$

การทดสอบอัตราส่วน

การคำนวณค่า $E_i = Np_i$

เช่น สนใจศึกษาหมู่เลือดของนักศึกษา

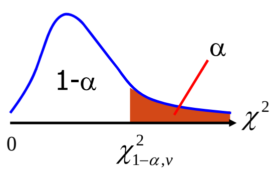
H_0 : อัตราส่วนของหมู่เลือด O:A:B:AB = 2:1:1:1

H_1 : อัตราส่วนของหมู่เลือด O:A:B:AB $\neq$ 2:1:1:1

Group	ความถี่ (O_i)	p_i	$E_i = Np_i$
O	3	p_1=2/5	= 10(2/5) = 4
A	3	p_2=1/5	= 10(1/5) = 2
B	3	p_3=1/5	= 10(1/5) = 2
AB	1	p_4=1/5	= 10-4-2-2 = 2
รวม (N)	10	$\sum p_i = 1$	$\sum E_i = \sum O_i = 10$

การทดสอบอัตราส่วน

เขตวิกฤต (Critical region) ที่ระดับนัยสำคัญ α



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

ข้อสังเกต
เขตวิกฤตใน Chi-square Test จะมีด้านขวามือเพียงด้านเดียวเท่านั้น

ทดลองทำ

- หากต้องการตรวจสอบว่าอัตราส่วนของการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง ไม่เท่ากับ 4 : 2 : 1 หรือไม่ ให้ทำการตั้งสมมุติฐานทางสถิติที่ใช้ทดสอบ

a) H_0: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง = 4:2:1
   H_1: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง $\neq$ 4:2:1

b) H_0: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง $\neq$ 4:2:1
   H_1: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง = 4:2:1

ทดลองทำ

- หากต้องการตรวจสอบว่ามีผู้เสียชีวิตจากอุบัติเหตุเป็น 4 เท่าของผู้เสียชีวิตจากโรคมะเร็ง และมีผู้เสียชีวิตจากโรคหัวใจไม่แตกต่างจากผู้เสียชีวิตจากโรคมะเร็ง ใช่หรือไม่
  ให้ทำการตั้งสมมุติฐานทางสถิติที่ใช้ทดสอบ

a) H_0: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคมะเร็ง = 4:1 และ หัวใจ : มะเร็ง = 1:1
   H_1: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคมะเร็ง $\neq$ 4:1 และ หัวใจ : มะเร็ง $\neq$ 1:1

b) H_0: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง = 4:1:1
   H_1: อัตราส่วนการเสียชีวิตจากอุบัติเหตุ : โรคหัวใจ : โรคมะเร็ง $\neq$ 4:1:1

ตัวอย่าง 1

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

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

13

ตัวอย่าง 1

- **สมมุติฐานทางสถิติ**

H_0 : อัตราส่วน A : B : AB : O = 2 : 2 : 1 : 3

H_1 : อัตราส่วน A : B : AB : O ≠ 2 : 2 : 1 : 3

คำนวณค่า $E_i=Np_i$

- **สถิติทดสอบ**

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

$$= \frac{(20-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(10-12.5)^2}{12.5} + \frac{(45-37.5)^2}{37.5}$$

= 3.0

กรุ๊ปเลือด	O_i	p_i	$E_i=Np_i$
A	20	2/8	25
B	25	2/8	25
AB	10	1/8	12.5
O	45	3/8	37.5
รวม	100		

100 – 25 – 25 -12.5

$$\chi^2 = \sum_{i=1}^{k} \frac{O_i^2}{E_i} - N = [\frac{(20)^2}{25} + \frac{(25)^2}{25} + \frac{(10)^2}{12.5} + \frac{(45)^2}{37.5}] - 100 = 3.0$$

14

ตัวอย่าง 1

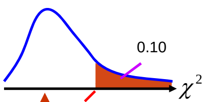
H_0: อัตราส่วน A:B:AB:O = 2:2:1:3

X_c^2 = 3.0

เขตวิกฤต ระดับนัยสำคัญ 10% **X^2 > 6.25**

v = k-1

= 3



$\chi^2_{0.90}$ = 6.25

สรุปผล

ไม่ปฏิเสธ H_0

นั่นคือ ที่ระดับนัยสำคัญ 0.10 อัตราส่วนของกรุ๊ปเลือดเป็นไปตามที่นักสถิติสงสัย

15

การทดสอบภาวะรูปดี

- การทดสอบอัตราส่วน

- การทดสอบการแจกแจง

16

การทดสอบภาวะรูปดี

การทดสอบการแจกแจงของข้อมูล

เป็นการทดสอบดูว่า ข้อมูลที่สนใจมีการแจกแจงความน่าจะเป็นในรูปแบบใด

ในที่นี้สนใจทดสอบ 3 การแจกแจง คือ

- การแจกแจงทวินาม (Binomial Distribution)
- การแจกแจงปัวซง (Poisson Distribution)
- การแจกแจงปกติ (Normal Distribution)

17

หลักการพิจารณาลักษณะการแจกแจงของข้อมูล

- **การแจกแจงทวินาม**
  - เป็นการทดลองสุ่มซ้ำ ๆ กัน n ครั้ง
  - แต่ละครั้งมีผลลัพธ์เกิดขึ้น 2 อย่าง คือ Success กับ Failure
  - ความน่าจะเป็นที่เกิด Success ในแต่ละครั้งเท่ากับ p
  - สนใจ X = จำนวนครั้งที่เกิด Success โดยมีค่าที่เป็นไปได้ 0, 1, …, n

จะได้ว่า X ~ b(x; n, p) โดย

$P(x) = {}^nC_x . p^x . q^{n-x}$ เมื่อ x = 0, 1, …, n

$E(X) = \mu = np, V(X) = npq$

18

ลองคำนวณค่าความน่าจะเป็นของตัวแปรสุ่มทวินาม

- X ~ b(x; n = 5, p = 0.4) จงคำนวณ P(X = 2)

  $P(X = 2) = {}^{5}C_{2}.(0.4)^{2}.(0.6)^{5-2}$
  หรือใช้ตารางทวินาม ที่ p = 0.4, n = 5 และ x = 2 จะได้
  P(X = 2) = 0.3456

- X ~ b(x; n = 10, p = 0.7) จงคำนวณ P(X = 6)

หลักการพิจารณาลักษณะการแจกแจงของข้อมูล

- **การแจกแจงปัวซง**
  - เป็นการทดลองสุ่มซ้ำ ๆ กัน ภายใต้ขอบเขตที่กำหนด (ช่วงเวลา, พื้นที่, ขนาด, ...)
  - สนใจ X = จำนวนครั้งที่เกิดเหตุการณ์ที่สนใจ โดยมีค่าที่เป็นไปได้ 0, 1, 2, ...

จะได้ว่า X ~ p(x; λ) โดย

$$P(x) = \frac{e^{-\lambda}.\lambda^{x}}{x!}$$

เมื่อ x = 0, 1, …, n
λ = ค่าเฉลี่ยของจำนวนครั้งที่เกิดเหตุการณ์สนใจ

E(X) = μ = λ, V(X) = λ

หลักการพิจารณาลักษณะการแจกแจงของข้อมูล

- **การแจกแจงปกติ**
  - เงื่อนไขที่สำคัญ : X เป็นตัวแปรปริมาณที่ต่อเนื่อง

เมื่อ X ~ N(μ = 10, σ = 2) หากต้องการประมาณค่าความน่าจะเป็นของ X ที่น้อยกว่า 12

หากต้องการประมาณค่าความน่าจะเป็นของ X ที่น้อยกว่า 8

การทดสอบลักษณะการแจกแจงของตัวแปร

ลักษณะข้อมูล : ข้อมูลคุณภาพ k กลุ่ม (ที่ได้จากตารางแจกแจงความถี่ข้อมูลปริมาณ)

ค่าสังเกตที่สนใจ : ความถี่ในแต่ละกลุ่ม (O_i)

เช่น สนใจศึกษาจำนวนเมล็ดถั่วเขียวที่ไม่งอก จากการทดลองปลูกในแปลงตัวอย่าง 10 แปลง ๆ ละ 5 เมล็ด

แปลง	จน.เมล็ด
1	1
2	2
3	3
4	4
5	1

แปลง	จน.เมล็ด
6	3
7	1
8	4
9	2
10	2

จน.เมล็ด	ความถี่	O_i
1	3	O_1
2	3	O_2
3	2	O_3
4	2	O_4

การทดสอบลักษณะการแจกแจงของตัวแปร

เช่น สนใจศึกษาอายุของตัวอย่างผู้ป่วย 10 คน

คนที่	อายุ
1	10
2	12
3	30
4	35
5	19

คนที่	อายุ
6	15
7	18
8	40
9	25
10	21

กลุ่มอายุ	ความถี่	O_i
< 15	2	O_1
15 – 20	3	O_2
20 – 25	2	O_3
> 25	3	O_4

การทดสอบการแจกแจงของตัวแปร

สมมติฐานทางสถิติ
H_0 : ข้อมูลมีการแจกแจงแบบที่สงสัย
H_1 : ข้อมูลไม่มีการแจกแจงแบบที่สงสัย

สถิติทดสอบ

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

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

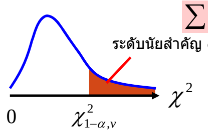
v = k – m – 1

m = จำนวนพารามิเตอร์ที่ต้องประมาณค่า

เงื่อนไขการวิเคราะห์: ค่า $E_i \geq 5$ (หาก $E_i<5$ ให้รวมกลุ่มที่มีลักษณะใกล้เคียง)

$\sum O_i = \sum E_i = N$ $\quad \sum p_i = 1$

เขตวิกฤต



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

ตัวอย่าง 3

- ในการปลูกเมล็ดพันธุ์ผัก 450 เมล็ด โดยปลูกเป็นแถว ๆ ละ 5 เมล็ด จำนวน 90 แถว พบว่าเมล็ดที่งอกในแต่ละแถวเป็นดังนี้

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

จากผลการทดลองนี้จะพิจารณาว่าข้อมูลดังกล่าวควรมีการแจกแจงความน่าจะเป็นแบบใดและทดสอบสมมติฐาน โดยใช้ระดับนัยสำคัญ 5%

X **ตัวแปรที่สนใจ :** จำนวนเมล็ดที่งอกในแต่ละแถว

: 0, 1, 2, 3, 4, 5

X ~ b(x, 5, p) ?

ตัวอย่าง 3

สมมติฐานทางสถิติ H_0 : ข้อมูลมีการแจกแจงแบบทวินาม

H_1 : ข้อมูลไม่มีการแจกแจงแบบทวินาม

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

คำนวณค่า $E_i=Np_i$

X	O_i	p_i	p_i
			P(X=0)
1	1	P(X=1)	P(X=1)
2	11	P(X=2)	P(X=2)
3	30	P(X=3)	P(X=3)
4	38	P(X=4)	P(X=4)
5	10	P(X=5)	P(X=5)
รวม	90	1.00?	1.00

$P(X = 1) = {}^5C_1.p^1q^4$

p = Prob.ที่เมล็ดจะงอก

= 3.5 / 5 = 0.7

จาก X ~ b(x;n,p)

E(X) = np = 5p → μ

ประมาณด้วย $\bar{X} = \sum fx / \sum f$

$= \frac{(1)(1)+...+(10)(5)}{90} = 3.5$

คำนวณค่า $E_i=Np_i$

ตรวจสอบเงื่อนไข $E_i \geq 5$

X	O_i	$p_i = {}^5C_x(.7)^x(.3)^{5-x}$		$E_i=Np_i$
0	-	P(X=0)	.0024	0.216
1	1	P(X=1)	.0284	2.556
2	11	P(X=2)	.1323	11.907
3	30	P(X=3)	.3087	27.783
4	38	P(X=4)	.3602	32.418
5	10	P(X=5)	.1681	15.12
รวม	90		1.00	

12

2.772

14.679

90 – (0.216 + ... + 32.418)

$$\chi_c^2 = \frac{(12-14.679)^2}{14.679} + \frac{(30-27.783)^2}{27.783} + \frac{(38-32.418)^2}{32.418} + \frac{(10-15.12)^2}{15.12} = 3.36$$

$$\chi_c^2 = [\frac{(12)^2}{14.679} + \frac{(30)^2}{27.783} + \frac{(38)^2}{32.418} + \frac{(10)^2}{15.12}] - 90 = 3.36$$

H_0 : ข้อมูลมีการแจกแจงแบบทวินาม

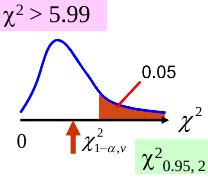
H_1 : ข้อมูลไม่มีการแจกแจงแบบทวินาม

χ^2_c = 3.36

เขตวิกฤต $\chi^2 > 5.99$

α= 0.05,

v = k – m – 1 = 2



$\chi^2_{0.95,\,2} = 5.99$

สรุปผล ไม่ปฏิเสธ H_0

ที่ α = 0.05 ข้อมูลจำนวนเมล็ดที่งอกมีการแจกแจงทวินาม

ตัวอย่าง 4

ในการศึกษาความชุกชุมของยุงลาย โดยทำการทดลองนำกระป๋องขนาดใหญ่บรรจุน้ำวางกระจัดกระจายในท้องที่หนึ่งเป็นเวลา 1 สัปดาห์ แล้วสำรวจจำนวนลูกน้ำของยุงลายในแต่ละกระป๋อง บันทึกข้อมูลได้ดังนี้

จำนวนลูกน้ำ	0	1	2	3	4	5	6	7
จำนวนกระป๋อง	14	18	29	18	10	7	3	1

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่าจำนวนลูกน้ำในกระป๋องมีการแจกแจงแบบปัวส์ซองหรือไม่

X **ตัวแปรที่สนใจ :** จำนวนลูกน้ำในกระป๋อง

X ~ p(x; λ) ? : 0, 1, 2, 3, ...

สมมติฐานทางสถิติ H_0 : ข้อมูลมีการแจกแจงแบบปัวส์ซง

H_1 : ข้อมูลไม่มีการแจกแจงแบบปัวส์ซง

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

คำนวณค่า $E_i=Np_i$

X_i	0	1	2	3	4	5	6	7	รวม
O_i	14	18	29	18	10	7	3	1	100
p_i	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X=6)	P(X=7)	1.00?

$P(X = 0) = (e^{-\lambda}.\lambda^x)/x!$

λ = จำนวนลูกน้ำเฉลี่ยที่พบ = ?

ดังนั้นประมาณค่า λ ด้วย 2.3

λ = ? → จากคุณลักษณะของตัวแปร X ~ p(x; λ)

E(X) = λ → $\bar{X}$

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{(14)(0)+\ldots+(1)(7)}{100} = 2.3$$

คำนวณค่า $E_i = Np_i$

X_i	0	1	2	3	4	5	6	7	>7	รวม
O_i	14	18	29	18	10	7	3	1	-	100
p_i	P(X=0) .1003	P(X=1) .2306	P(X=2) .2652	P(X=3) .2033	P(X=4) .1169	P(X=5) .0538	P(X=6) .0206	P(X=7) .0068	P(X>7) .0025	1.0
E_i	10.03	23.06	26.52	20.33	11.69	5.38	2.06	0.68	0.25	100

100 – (10.03 + ... + 0.68)

31

คำนวณค่า χ^2

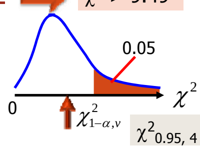
X_i	0	1	2	3	4	5	6	7	>7	รวม
O_i	14	18	29	18	10	7	3	1 (11)	-	100
E_i	10.03	23.06	26.52	20.33	11.69	5.38 (8.37)	2.06 (2.99)	0.68 (0.93)	0.25	100

≥ 5

$$\chi_c^2 = \frac{(14-10.03)^2}{10.03} + \ldots + \frac{(11-8.37)^2}{8.37} = 4.25$$

เขตวิกฤต → $\chi^2 > 9.49$

$\alpha = 0.05$,
v = k–m–1
= 4



$\chi^2_{0.95,\,4} = 9.49$

สรุปผล Accept H_0

ที่ α = 0.05 ข้อมูลจำนวนลูกน้ำมีการแจกแจงปัวส์ซง

32

ตัวอย่าง 5

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

สรุปที่ระดับนัยสำคัญ 5% ได้หรือไม่ว่า ความยาวของทารกแรกเกิดมีการแจกแจงปกติ

X ตัวแปรที่สนใจ : ความยาวของทารก

X ~ N(μ,σ) ? : > 0 ซม.

33

ตัวอย่าง 5

สมมติฐานทางสถิติ H_0 : ค.ยาวทารกมีการแจกแจงปกติ

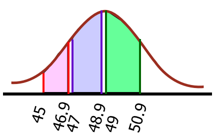
H_1 : ค.ยาวทารกไม่มีการแจกแจงปกติ

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

คำนวณค่า $E_i = Np_i$

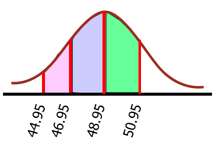
ความยาว - X	O_i	p_i = P(X)
45 - 46.9	28	P(45 < X < 46.9)
47 - 48.9	32	P(47 < X < 48.9)
49 - 50.9	35	P(49 < X < 50.9)
51 - 52.9	20	P(51 < X < 52.9)
53 - 54.9	10	P(53 < X < 54.9)
		1.00?

ต้องทำการขยายเขตของแต่ละชั้นให้ต่อเนื่องกัน



34

ตัวอย่าง 5



คำนวณค่า $E_i = Np_i$

ความยาว - X	O_i	X	p_i = P(X)
	0	< 44.95	P(X < 44.95)
45 - 46.9	28	44.95 – 46.95	P(44.95 < X < 46.95)
47 - 48.9	32	46.95 – 48.95	P(46.95 < X < 48.95)
49 - 50.9	35	48.95 – 50.95	P(48.95 < X < 50.95)
51 - 52.9	20	50.95 – 52.95	P(50.95 < X < 52.95)
53 - 54.9	10	52.95 – 54.95	P(52.95 < X < 54.95)
	0	> 54.95	P(X > 54.95)
			1.00

35

ตัวอย่าง 5

แปลง X เป็น Z โดย
Z = (x - μ)/σ

แต่ μ, σ unknown
→ $\mu \sim \bar{x}$, $\sigma \sim s$

คำนวณค่า $E_i = Np_i$

x_i	ความยาว - X	O_i	p_i = P(X)	p_i = P(Z)
-	<44.95	0	P(X < 44.95)	P(Z < -1.73)
45.95	44.95 - 46.95	28	P(44.95 < X < 46.95)	P(-1.73 < Z < -0.91)
47.95	46.95 - 48.95	32	P(46.95 < X < 48.95)	P(-0.91 < Z < -0.09)
49.95	48.95 - 50.95	35	P(48.95 < X < 50.95)	P(-0.09 < Z < 0.72)
51.95	50.95 - 52.95	20	P(50.95 < X < 52.95)	P(0.72 < Z < 1.54)
53.95	52.95 - 54.95	10	P(52.95 < X < 54.95)	P(1.54 < Z < 2.36)
-	>54.95	0	P(X > 54.95)	P(Z > 2.36)

$$\bar{x} = \frac{\sum f_i x_i}{N} = 49.18 \qquad s^2 = \frac{\sum f_i x_i^2 - N\bar{x}^2}{N-1} = 5.99 \Rightarrow s = 2.45$$

36

ตัวอย่าง 5

คำนวณค่า $E_i = Np_i$

$p_i = P(Z)$	F_i	p_i
$P(Z < -1.73)$	$P(Z < -1.73) = 1-.9582 = .0418$	.0418
$P(-1.73 < Z < -0.91)$	$P(Z < -0.91) = 1-.8186 = .1814$	=.1814-.0418
$P(-0.91 < Z < -0.09)$	$P(Z < -0.09) = 1-.5359 = .4641$	= .4641-.1814
$P(-0.09 < Z < 0.72)$	$P(Z < 0.72) = .7642$	= .7642-.4641
$P(0.72 < Z < 1.54)$	$P(Z < 1.54) = .9382$	= .9382-.7642
$P(1.54 < Z < 2.36)$	$P(Z < 2.36) = .9909$	= .9909-.9382
$P(Z > 2.36)$	$P(Z > 2.36) = 1-.9909 = .0091$	.0091

37

ตัวอย่าง 5

ตรวจสอบเงื่อนไขการวิเคราะห์: $E_i \geq 5$

คำนวณค่า $E_i = Np_i$

ความยาว - X	O_i	P_i	E_i
<44.95	0	.0418	5.23
44.95 - 46.95	28	.1396	17.45
46.95 - 48.95	32	.2827	35.34
48.95 - 50.95	35	.3001	37.51
50.95 - 52.95	20	.1740	21.75
52.95 - 54.95	10	.0527	6.59
>54.95	0	.0091	1.13
	125	1.0000	

$\left.\begin{matrix} 6.59 \\ 1.13 \end{matrix}\right\} 7.72$

$v = k - m - 1$
$= 6 - 2 - 1 = 3$

$m = 2$ เนื่องจากไม่รู้ค่า μ, σ

$$\chi_c^2 = \frac{(0-5.23)^2}{5.23} + \frac{(28-17.45)^2}{17.45} + \frac{(32-35.34)^2}{35.34} + \ldots + \frac{(10-7.72)^2}{7.72} = 12.91$$

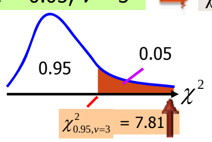
$$\chi_c^2 = [\frac{(0)^2}{5.23} + \frac{(28)^2}{17.45} + \frac{(32)^2}{35.34} + \frac{(35)^2}{37.51} + \frac{(20)^2}{21.75} + \frac{(10)^2}{7.72}] - 125 = 12.91$$

38

ตัวอย่าง 5

H_0: ค.ยาวทารกมีการแจกแจงปกติ
$X_c^2 = 12.91$

- **เขตวิกฤต** $\alpha = 0.05, v = 3$ ➡ $\chi^2 > 7.81$



- **สรุปผล**

เนื่องจาก $X_c^2 > 7.81$ จึงปฏิเสธ H_0
นั่นคือ ที่ระดับนัยสำคัญ 0.05 สรุปว่าความยาวทารกไม่มีการแจกแจงเป็นปกติ

39

การทดสอบความเป็นอิสระ
TEST FOR INDEPENDENCE

40

การทดสอบความเป็นอิสระ

- เป็นการทดสอบที่สนใจดูความสัมพันธ์ระหว่างตัวแปร(คุณภาพ) 2 ตัว เช่น
  - เพศมีความสัมพันธ์กับการสูบบุหรี่หรือไม่
  - ความสัมพันธ์ระหว่างสาขาวิชาที่ศึกษากับอาชีพการงาน

การทดสอบความเป็นอิสระ

ลักษณะข้อมูลในการศึกษา

ในการศึกษาความสัมพันธ์ระหว่างเพศกับการสูบบุหรี่ของนศ. 20 คน

คนที่	เพศ	สูบบุหรี่
1	ชาย	สูบ
2	หญิง	ไม่สูบ
3	หญิง	สูบ
…	…	…
20	ชาย	สูบ

ตารางแจกแจง 2 ทาง มิติ 2x2

	สูบบุหรี่	ไม่สูบ	รวม
ชาย	6 O_{11}	4 O_{12}	10 R_1
หญิง	5 O_{21}	5 O_{22}	10 R_2
รวม	11 C_1	9 C_2	20 N

ค่าความถี่ที่ได้จากการสังเกต O_{ij}

42

การทดสอบความเป็นอิสระ

สมมติฐานทางสถิติ

H_0 : ตัวแปร X และตัวแปร Y **เป็นอิสระกัน**
H_1 : ตัวแปร X และตัวแปร Y **ไม่เป็นอิสระกัน**

H_0 : ตัวแปร X **ไม่ขึ้นกับ**ตัวแปร Y
H_1 : ตัวแปร X **ขึ้นกับ**ตัวแปร Y

H_0 : ตัวแปร X และ Y **ไม่มีความสัมพันธ์กัน**
H_1 : ตัวแปร X และ Y **มีความสัมพันธ์กัน**

43

ทดลองทำ

- ต้องการวิเคราะห์เพื่อศึกษาว่า เพศ (ชาย-หญิง) มีความสัมพันธ์กับการสูบบุหรี่ (สูบ-ไม่สูบ) หรือไม่
  ลองตั้งสมมติฐานทางสถิติเพื่อทดสอบ

ก. H_0: เพศมีความสัมพันธ์กับการสูบบุหรี่
   H_1: เพศไม่มีความสัมพันธ์กับการสูบบุหรี่
ข. H_0: เพศไม่มีความสัมพันธ์กับการสูบบุหรี่
   H_1: เพศมีความสัมพันธ์กับการสูบบุหรี่
ค. H_0: เพศกับการสูบบุหรี่เป็นอิสระกัน
   H_1: เพศกับการสูบบุหรี่ไม่เป็นอิสระกัน

44

ทดลองทำ

- ต้องการศึกษาว่า รูปแบบการได้รับยา (แบบพ่น, แบบทาน, แบบฉีด) มีผลต่อผลการรักษาผู้ป่วย (ดีขึ้น, เหมือนเดิม, แย่ลง) หรือไม่
  ลองตั้งสมมติฐานทางสถิติเพื่อทดสอบ

ก. H_0: รูปแบบการได้รับยาไม่มีผลต่อผลการรักษา
   H_1: รูปแบบการได้รับยามีผลต่อผลการรักษา
ข. H_0: รูปแบบการรับยาและผลการรักษาเป็นอิสระกัน
   H_1: รูปแบบการรับยาและผลการรักษาไม่เป็นอิสระกัน
ค. H_0: รูปแบบการได้รับยามีผลต่อผลการรักษา
   H_1: รูปแบบการได้รับยาไม่มีผลต่อผลการรักษา

45

การทดสอบความเป็นอิสระ

สถิติทดสอบ

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

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

- $E_{ij} = Np_{ij}$, r = จำนวนกลุ่มตัวแปร X, c = จำนวนกลุ่มตัวแปร Y
- p_{ij} = ความน่าจะเป็นที่จะเกิดเหตุการณ์ i และ j ภายใต้ H_0
- N = ขนาดตัวอย่างที่ใช้ในการศึกษา

กรณีนี้ H_0 : ตัวแปร X กับ Y เป็นอิสระกัน
จากคุณสมบัติของความน่าจะเป็นของเหตุการณ์ที่เป็นอิสระกัน จะได้ P(XY) = P(X).P(Y)

46

การทดสอบความเป็นอิสระ

คำนวณค่า E_{ij}

X \ Y	Y_1	Y_2	Total
X_1	O_{11} E_{11}	O_{12} E_{12}	R_1
X_2	O_{21} E_{21}	O_{22} E_{22}	R_2
X_3	O_{31} E_{31}	O_{32} E_{32}	R_3
Total	C_1	C_2	N

$$\sum\sum O_{ij} = \sum\sum E_{ij} = N$$

$\Rightarrow E_{11} = E(X_1Y_1) = N.p_{11}$
ภายใต้กฎของเหตุการณ์อิสระ
$p_{11} = p(X_1).p(Y_1)$
เมื่อ $p(X_1) = R_1/N$, $p(Y_1) = C_1/N$
ดังนั้น $p_{11} = (R_1/N)(C_1/N)$

$\mathbf{E_{11} = (R_1.C_1)/N}$

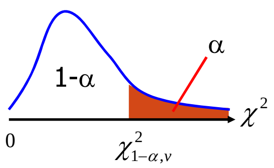
$\boxed{\mathbf{E_{ij} = (R_i.C_j)/N}}$

$\Rightarrow E_{21} = N.p_{21} = (R_2.C_1)/N$
แต่สำหรับแถว & คอลัมน์สุดท้าย
เช่น $\Rightarrow E_{12} = R_1 - E_{11}$
$\Rightarrow E_{31} = C_1 - E_{11} - E_{21}$

47

การทดสอบความเป็นอิสระ

เขตวิกฤต (Critical region) ที่ระดับนัยสำคัญ α



เขตวิกฤต คือ

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

จะปฏิเสธ H_0 เมื่อค่า χ^2_c ที่คำนวณได้ตกอยู่ในเขตวิกฤต

ในกรณีที่ปฏิเสธ H_0 จะสรุปได้ว่าตัวแปรทั้งสองมีความสัมพันธ์กัน สามารถคำนวณขนาดความสัมพันธ์ที่เกิดขึ้นจากค่า Cramer's V

$$V = \sqrt{\frac{\chi^2}{N.\min(r-1, c-1)}}$$

โดย $0 \le V \le 1$
เมื่อ χ^2 = ค่าตัวสถิติไคสแควร์ที่คำนวณได้

48

ตัวอย่าง 6

ข้อมูลที่กำหนดให้ในตารางข้างล่างนี้ได้จากการตรวจสอบ เพื่อศึกษาความสัมพันธ์ระหว่างเชื้อชาติและชนิดของกรุ๊ปเลือดของคน จงทดสอบความเป็นอิสระด้วยระดับนัยสำคัญ 1%

เชื้อชาติ	กรุ๊ปเลือด O	A	B	AB	รวม
1	176	148	96	72	492
2	78	50	45	12	185
3	15	19	8	7	49
รวม	269	217	149	91	726

49

ตัวอย่าง 6

สมมติฐานทางสถิติ

H_0 : กรุ๊ปเลือดและเชื้อชาติไม่มีความสัมพันธ์กัน

H_1 : กรุ๊ปเลือดและเชื้อชาติมีความสัมพันธ์กัน

สถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{O_{ij}^2}{E_{ij}} - N$$

50

คำนวณค่า E_{ij}

ตาราง O_{ij}

176	148	96	72	492
78	50	45	12	185
15	19	8	7	49
269	217	149	91	726

182.30	147.06	100.98	61.66	492
68.55	55.30	37.97	23.18	185
18.15	14.64	10.05	6.16	49
269	217	149	91	726

E_{11} = (492)(269)/726

E12 = (492)(217)/726

E13 = (492)(149)/726

E14 = 492 – (182.30+147.06+100.98)

E21 = (185)(269)/726

E22 = (185)(217)/726

E23 = (185)(149)/726

E24 = 185 – (68.55+55.30+37.97)

คำนวณค่า χ^2

$$\chi_c^2 = \frac{(176-182.30)^2}{182.30} + \frac{(148-147.06)^2}{147.06} + \ldots + \frac{(7-6.16)^2}{6.16} = 13.09$$

$$\chi_c^2 = \left[\frac{(176)^2}{182.30} + \frac{(148)^2}{147.06} + \frac{(96)^2}{100.98} + \frac{(72)^2}{61.66} + \ldots + \frac{(7)^2}{6.16}\right] - 726 = 13.09$$

เขตวิกฤต

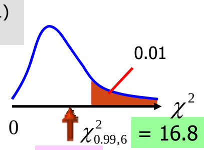
H_0 : กรุ๊ปเลือดและเชื้อชาติไม่มีความสัมพันธ์กัน

H_1 : กรุ๊ปเลือดและเชื้อชาติมีความสัมพันธ์กัน

χ^2_c = 13.09

α= 0.01,

v = (r–1)(c–1)
= 6



$\chi^2_{0.99,6}$ = 16.8

13.09

สรุปผล Accept H_0

ที่ α = 0.01 กรุ๊ปเลือดและเชื้อชาติไม่มีความสัมพันธ์กัน

ลองคำนวณค่า Cramer's V

$$V = \sqrt{\frac{\chi^2}{N.\min(r-1, c-1)}} = \sqrt{\frac{13.09}{726.\min(3,2)}} = \sqrt{\frac{13.09}{726\times 2}} = 0.095$$

52

ข้อจำกัดของการวิเคราะห์ไคสแควร์

- เหมาะสำหรับข้อมูลที่มีขนาดตัวอย่างใหญ่ โดยที่ค่า $E_i \geq 5$ สำหรับทุก ๆ กลุ่ม
- หากพบว่าค่า E_i ของกลุ่มใดมีค่าน้อยกว่า 5 มีแนวทางในการแก้ปัญหาดังนี้
  - ทำการเก็บรวบรวมข้อมูลเพิ่มเติม
  - ทำการรวมกลุ่มข้อมูลที่มีค่าใกล้เคียงกัน

53

ข้อจำกัดของการวิเคราะห์ไคสแควร์

- กรณี df = 1
  ควรใช้ Corrected χ^2 ในการวิเคราะห์ โดย

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

54

เหตุใดจึงมีเพียงเขตวิกฤตทางขวามือ

- พิจารณาจากสูตร Chi-square ที่ใช้

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

- ค่า χ^2 จะมีค่ามาก ถ้าค่า O_i กับ E_i แตกต่างกันมากๆ
- ค่า χ^2 จะมีค่าน้อย ถ้าค่า O_i กับ E_i ใกล้เคียงกัน

ดังนั้นหากกำหนดเขตวิกฤตทางซ้ายมือ
จะทำให้สามารถปฏิเสธ H_0 ได้ในกรณีที่ค่า χ^2 มีค่าน้อย และสรุปว่า O_i กับ E_i แตกต่างกัน **ซึ่งเป็นเท็จ**

55

ตัวอย่างการคำนวณ

H_0 : G1:G2:G3 = 1:1:1

H_1 : G1:G2:G3 $\neq$ 1:1:1

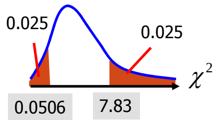
คำนวณค่าสถิติ

$\chi_c^2 = 0.0$

กลุ่ม	O_i	p_i	$E_i=Np_i$
1	20	1/3	20
2	20	1/3	20
3	20	1/3	20
Tol	60	1.00	60

เขตวิกฤต

$\alpha = 0.05$, $\nu = 2$



สรุปผล ปฏิเสธ H_0

ด้วยเหตุนี้ จึงกำหนดเขตวิกฤตเพียงด้านขวามือเท่านั้น

56