บทที่ 13
ไคสแควร์
(Chi Square)

13.1 การแจกแจงของไคสแควร์

ถ้าให้ X เป็นตัวแปรสุ่ม ที่มาจากประชากรที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 เราสามารถเปลี่ยนตัวแปร X ให้อยู่ในรูปของคะแนนมาตรฐาน Z โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ได้จากสูตร

$$Z = \frac{X - \bar{X}}{S}$$

ถ้ายกกำลังสองของคะแนนมาตรฐาน Z จะได้การแจกแจงแบบไคสแควร์ ซึ่งมี 1 degree of freedom ซึ่งเขียนเป็นสัญญลักษณ์ได้ดังนี้

$$\chi_1^2 = Z^2$$

ถ้าเลือกตัวอย่างมา 2 ตัว ซึ่งมีความเป็นอิสระต่อกันจะได้

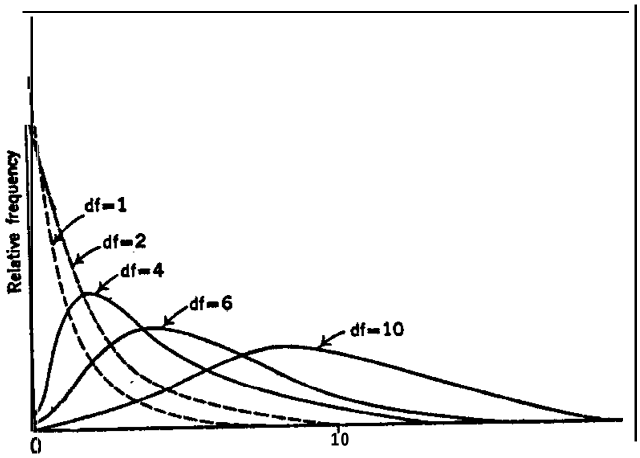
$$\chi_2^2 = Z_1^2 + Z_2^2$$

มีการแจกแจงแบบไคสแควร์ ซึ่งมี 2 degree of freedom

โดยทั่ว ๆ ไปถ้าเลือกตัวอย่างจำนวน N ซึ่งเป็นอิสระต่อกันแล้ว ผลบวกของกำลังสองของคะแนนมาตรฐานเหล่านั้น จะมีการแจกแจงแบบไคสแควร์ โดยมี N degree of freedom

$$\chi_n^2 = Z_1^2 + Z_2^2 + \dots + Z_n^2 \qquad \text{..................(13–1)}$$

โดยที่ χ^2 จะมีค่าเฉลี่ยเท่ากับ n และความแปรปรวนเท่ากับ 2 n การแจกแจงของไคสแควร์นั้นจะขึ้นอยู่กับ degree of freedom ดังรูป



จากภาพจะเห็นได้ว่า ถ้ามี degree of freedom ค่าหนึ่งก็จะให้โค้งการแจกแจงของ $\underline{\chi}^2$ โค้งหนึ่ง เช่นเดียวกับการแจกแจงของ t แต่ต่างกันที่การแจกแจงของ χ^2 จะมีลักษณะไม่สมมาตร

13.2 คุณสมบัติของการแจกแจงไคสแคว

1. $\underline{\chi}^2$ มีค่าเฉลี่ยเท่ากับ n ความเบี่ยงเบนมาตรฐานเท่ากับ $\sqrt{2n}$
2. $\underline{\chi}^2$ มีค่ามัธยฐานเท่ากับ $n-1$ และมีค่าความเบ้ (skewness) เท่ากับ $\sqrt{\frac{8}{n}}$
3. ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่มาก ๆ การแจกแจง χ^2 จะมีลักษณะเข้าใกล้การแจกแจงปกติ

13.3 ข้อตกลงเบื้องต้นเกี่ยวกับการทดสอบโดยใช้ ไคสแคว

1. กลุ่มตัวอย่างต้องถูกเลือกมาแบบสุ่ม
2. กลุ่มตัวอย่างต้องมีขนาดใหญ่พอ
3. ข้อมูลที่ได้จากการวัดอยู่ในมาตรานามบัญญัติ โดยนับเป็นความถี่

13.4 The **Goodness** of Fit **Test**

เป็นการใช้ไคสแควเพื่อทดสอบดูว่าสิ่งที่เกิดขึ้นนั้นแตกต่างจากสิ่งที่จะเกิดตามทฤษฎี (หรือตามที่คาดหวัง) หรือไม่ โดยทั่วไปจะใช้ χ^2 ทดสอบในกรณีที่ข้อมูลที่ได้อยู่ในมาตรานามบัญญัติ (Nominal Scale) กล่าวคือข้อมูลจะออกมาเป็นความถี่ ถ้าผลการทดสอบโดยใช้ χ^2 แล้วปรากฏว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (significance) แสดงว่าการเกิดนั้นแตกต่างจากทฤษฎี ค่า χ^2 คำนวณได้จากสูตร

$$\underline{\chi}^2 = \Sigma \frac{(O-E)^2}{E} \qquad \text{.....................(13—2)}$$

เมื่อ O แทน ความถี่ที่ได้จากการสังเกต (Observed frequency)

E แทน ความถี่ที่ได้จากทฤษฎี (Expected frequency)

χ^2 มี degree of freedom เท่ากับ n—1

13.5 การใช้ x^2 เพื่อทดสอบความแตกต่างของ O กับ E ซึ่งมีลักษณะการเกิดแบบ equal probability

ตัวอย่าง 1 ในการทอดลูกเต๋า 300 ครั้ง ปรากฏผลดังนี้

หน้า 1 ออก 43 ครั้ง
หน้า 2 ออก 55 ครั้ง
หน้า 3 ออก 39 ครั้ง
หน้า 4 ออก 56 ครั้ง
หน้า 5 ออก 63 ครั้ง
หน้า 6 ออก 44 ครั้ง

จงหาว่าลูกเต๋านี้เที่ยงตรงตามทฤษฎีไหม (ที่ระดับ .05)

X	0	E	(O–E)	$(O-E)^2$
1	43	50	– 7	49
2	55	50	5	25
3	39	50	– 11	121
4	56	50	6	36
5	63	50	13	169
6	44	50	– 6	36

จากสูตร $x^2 = \Sigma \frac{(O-E)^2}{E}$

$$= \frac{49}{50} + \frac{25}{50} + \underline{150} + \underline{36} + \underline{160} + \frac{36}{50}$$

$$= 8.72$$

จากตารางที่ α = .05, df = 5, χ^2 = 11.07

แต่ค่า χ^2 ที่ได้จากการคำนวณมีค่าเท่ากับ 8.72 ซึ่งน้อยกว่าค่า χ^2 จากตาราง

$\therefore$ χ^2 ไม่มีนัยสำคัญทางสถิติที่ระดับ .05 หมายความว่าลูกเต๋านี้อยู่ในเกณฑ์ใช้ได้ กล่าวคือการเกิดแต้มต่าง ๆ ไม่แตกต่างไปจากทฤษฎี

13.6 การใช้ χ^2 เพื่อทดสอบความแตกต่างของ O กับ E ซึ่งมีลักษณะการเกิดแบบ unequal probability

ตัวอย่าง 2 ครูคนหนึ่งให้เกรดนักเรียนจำนวน 200 คน ในวิชาสถิติ โดยตัด 4 เกรด ดังนี้ เกรด A ให้ 30 คน เกรด B ให้ 74 คน เกรด C ให้ 84 คน และ เกรด D ให้ 12 คน อยากทราบว่าครูคนนี้ให้เกรดเป็นไปตามทฤษฎีไหม (ที่ระดับ .05)

Grade	O	E	O - E	$(O-E)^2$
A	30	13	17	289
B	74	87	-13	169
C	84	87	- 3	9
D	12	13	- 1	1

ความถี่ที่คาดหวัง (E) คำนวณตามทฤษฎีการตัดเกรดตามโค้งปรกติ ซึ่งจะไม่แสดงวิธีคำนวณในที่นี้ ค่าความถี่ที่คาดหวัง (E) คำนวณแล้วได้ผลดังตารางข้างบน

$$\text{จากสูตร} \quad \chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{289}{13} + \frac{169}{87} + \frac{9}{87} + \frac{1}{13}$$

$$= 24.354$$

จากตารางที่ α = .05, df = 3, χ^2 = 7.815

แต่ค่า χ^2 ที่ได้จากการคำนวณมีค่าเท่ากับ 24.354 ซึ่งมากกว่าค่า χ^2 จากตาราง

$\therefore$ χ^2 มีนัยสำคัญทางสถิติที่ระดับ .05 หมายความว่าการตัดเกรดนี้ไม่เป็นไปตามทฤษฎี

13.7 การใช้ χ^2 เพื่อทดสอบความแตกต่างของ O กับ E เมื่อ E มีค่าต่ำกว่า 5 หรือมี 2 กลุ่ม (df = 1) ให้ใช้ Yate's correction ดังนี้

$$\chi^2 = \Sigma \frac{(|O-E|-0.5)^2}{E} \qquad \text{.....................(13–3)}$$

ที่ใช้ Yate's correction ก็เพื่อให้ค่าของ χ^2 ถูกต้องมากขึ้น

ตัวอย่าง 3 ข้อสอบถูกผิดจำนวน 100 ข้อ มีเด็กคนหนึ่งทำได้ถูก 60 ข้อ จงหาว่าเด็กคนนี้มีความสามารถแตกต่างไปจากลิงหรือไม่ (ลิงในที่นี้มีความสามารถเป็นไปตามทฤษฎี)

คำตอบ	O	E	$\lvert O-E \rvert$	$(\lvert O-E \rvert-.5)^2$
ถูก	60	50	10	$(9.5)^2$
ผิด	40	50	10	$(9.5)^2$

$$\chi^2 = \Sigma \frac{(|O-E|-.5)^2}{E}$$

$$= \frac{(9.5)^2}{50} + \frac{(9.5)^2}{50}$$

$$= 3.60$$

จากตารางที่ $\alpha = .05$, df = 1 , $\chi^2 = 3.84$

แต่ค่า χ^2 ที่ได้จากการคำนวณมีค่าเท่ากับ 3.60 ซึ่งน้อยกว่าค่า χ^2 จากตาราง

$\therefore$ χ^2 ไม่มีนัยสำคัญทางสถิติที่ระดับ .05 หมายความว่าเราไม่พบว่าเด็กคนนี้มีความสามารถแตกต่างไปจากลิง (ทฤษฎี)

13.8 Test of Independent

เป็นการทดสอบว่าตัวแปร 2 ตัวนั้นเป็นอิสระต่อกันหรือไม่ซึ่งตัวแปรทั้งสองนั้นเป็นตัวแปรแบบไม่ต่อเนื่อง

จากการทดสอบ χ^2 หากปรากฏว่าค่า χ^2 มีนัยสำคัญทางสถิติ (significance) หมายความว่าตัวแปรทั้ง 2 นั้นขึ้นต่อกัน

หากผลจากการทดสอบ χ^2 ปรากฏว่าค่า χ^2 ไม่มีนัยสำคัญทางสถิติ หมายความว่าจากข้อมูลเราไม่พบว่าตัวแปรทั้ง 2 นั้นขึ้นต่อกัน

χ^2 แบบนี้มี degree of freedom เท่ากับ $(r-1)\ (c-1)$

ตัวอย่าง 4 จากการสำรวจนักเรียน 170 คน ดูความถนัดในการใช้มือกับระดับความสามารถในการเรียน ผลปรากฏดังตาราง

ระดับ / ความถนัด	เก่ง	ปานกลาง	อ่อน	รวม
2 มือ ซ้าย	15	30	11	56
2 มือ	10	11	8	29
ขวา	25	39	21	85
รวม	50	80	40	170

จงทดสอบว่าความสามารถในการเรียน ขึ้นอยู่กับความถนัดของมือหรือไม่ (ที่ระดับ .05)

ก่อนอื่นต้องคำนวณหา ความถี่ที่คาดหวัง (E) ในทุก ๆ ช่อง (cell) จากสูตร

$$E_{ij} = \frac{R_i \times C_j}{N} \qquad \text{.............................(13—4)}$$

เมื่อ R คือผลรวมในแต่ละแถว

C คือผลรวมในแต่ละคอลัมน์

N คือจำนวนคนในกลุ่มตัวอย่างทั้งหมด

ระดับ / ความถนัด	เก่ง	ปานกลาง	อ่อน	รวม
ซ้าย	15 (16.47)	30 (26.35)	11 (13.18)	56
2 มือ	10 (8.53)	11 (13.65)	8 (6.82)	29
ขวา	25 (25.00)	39 (40.00)	21 (20.00)	85
รวม	50	80	40	170

ตัวเลขในวงเล็บ คือค่าความถี่ที่คาดหวัง (E) ในแต่ละช่อง ซึ่งคำนวณจากสูตรข้างบนดังนี้

ถ้า O = 15 แล้ว $E = \frac{56 \times 50}{170} = 16.47$

O = 30 แล้ว $E = \frac{56 \times 80}{170} = 26.33$

O = 11 แล้ว $E = \frac{56 \times 40}{170} = 13.18$

O = 10 แล้ว $E = \frac{29 \times 50}{170} = 8.53$

$O = 11$ แล้ว $E = \frac{29 \times 80}{170} = 13.65$

$O = 8$ แล้ว $E = \frac{29 \times 40}{170} = 6.82$

$O = 25$ แล้ว $E = \frac{85 \times 50}{170} = 25.00$

$O = 39$ แล้ว $E = \frac{85 \times 80}{170} = 40.00$

$O = 21$ แล้ว $E = \frac{85 \times 40}{170} = 20.00$

O	E	O - E	$(O-E)^2$
15	16.47	-1.47	2.16
30	26.35	3.65	13.32
11	13.18	-2.18	4.75
10	8.53	1.47	2.16
11	13.65	-2.65	7.02
8	6.82	1.18	1.39
25	25.00	0	0
39	40.00	- 1	1
21	20.00	1	1

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{2.16}{16.47} + \frac{13.32}{26.35} + \frac{4.75}{13.78} + \frac{2.16}{0.53} + \frac{7.02}{13.65}$$

$$+ \frac{1.39}{6.82} + \frac{0}{25} + \frac{1}{40} + \frac{1}{20}$$

$$= 2.04$$

จากตารางที่ $\alpha = .05$, $df = (3-1)(3-1) = 4$, $\chi^2 = 9.88$

แต่ค่า χ^2 ที่ได้จากการคำนวณ มีค่าเท่ากับ 2.04 ซึ่งน้อยกว่าค่า χ^2 จากตาราง

$\therefore \chi^2$ ไม่มีนัยสำคัญทางสถิติ หมายความว่า จากข้อมูลชุดนี้ เราไม่พบว่าความเก่งอ่อนขึ้นอยู่กับความถนัดของมือ

การทดสอบ Independent นั้นผู้วิจัยไม่ควรหยุดวิเคราะห์เพียงแค่นี้ แล้วสรุปแต่เพียงว่าตัวแปร 2 ตัวนี้มีความสัมพันธ์กัน สิ่งที่ผู้วิจัยควรจะทำต่อไป เมื่อรู้ว่าตัวแปร 2 ตัวนั้นมีความสัมพันธ์กันก็คือจะต้องระบุปริมาณความสัมพันธ์ได้

สำหรับกรณีที่ค่า χ^2 อยู่ในลักษณะตาราง 2×2 อาจหาความสัมพันธ์ได้จากสูตร

$$\varnothing = \sqrt{\frac{\chi^2}{N}} \quad \text{.................................}(13\text{—}5)$$

การให้ความหมายของ $\varnothing$ จะมีลักษณะเช่นเดียวกับค่าสัมประสิทธิ์สหสัมพันธ์ (r_{xy}) แบบ Pearson Product Moment Correlation

ถ้าหากค่า χ^2 จากการคำนวณไม่อยู่ในลักษณะตาราง 2×2 ก็อาจหาความสัมพันธ์ระหว่างตัวแปรทั้งสองโดยใช้สูตร

$$\varnothing = \sqrt{\frac{\chi^2}{N+\chi^2}} \quad \text{..........................}(13\text{—}6)$$

13.9 Chi Square Test of Homogeneity of Proportions

ในบางครั้งนักวิจัยอาจต้องการทดสอบสมมุติฐานที่ระบุความเท่ากันของสัดส่วนของประชากร เช่น ต้องการดูว่านักศึกษาชายที่สูบบุหรี่กับไม่สูบบุหรี่ในคณะศึกษาศาสตร์ วิทยาศาสตร์ และรัฐศาสตร์ โดยผู้วิจัยได้ทำการสุ่มตัวอย่างจากนักศึกษาคณะต่าง ๆ ผลการรวบรวมข้อมูลปรากฏดังนี้

ตัวอย่าง 5 จงทดสอบสัดส่วนของนักศึกษาที่สูบบุหรี่ในคณะต่าง ๆ ว่าแตกต่างกันหรือไม่

คณะ / การสูบบุหรี่	ศึกษาศาสตร์	วิทยาศาสตร์	รัฐศาสตร์	รวม
สูบบุหรี่	152	72	176	400
ไม่สูบบุหรี่	128	88	84	300
รวม	280	160	260	700

วิธีทำ $H_0 : p_1 = p_2 = p_3$

$H_1 : H_0$ ไม่จริง

ถ้าให้ p_1 แทนสัดส่วนของนักศึกษาคณะศึกษาศาสตร์ที่สูบบุหรี่

p_2 แทนสัดส่วนของนักศึกษาคณะวิทยาศาสตร์ที่สูบบุหรี่

p_3 แทนสัดส่วนของนักศึกษาคณะรัฐศาสตร์ที่สูบบุหรี่

$$p_1 = \frac{152}{280} = .54$$

$$p_2 = \frac{12}{160} = .45$$

$$p_3 = \frac{176}{260} = .68$$

ค่า p_1, p_2, p_3 นี้เป็นค่าสัดส่วนที่ได้จากการสังเกต (O)

ส่วนการคำนวณความถี่ที่คาดหวัง (Expected frequency) เมื่อคาดว่าสัดส่วนเท่ากัน สามารถคำนวณโดยการเทียบบัญญัติไตรยางค์ ดังนี้

นักศึกษา 700 คน จะเป็นผู้สูบบุหรี่ 400 คน

นักศึกษา 280 คน น่าจะสูบบุหรี่ $\frac{400}{700} \times 280 = 160$ คน

ค่า Expected frequency ในแต่ละเซล จะเป็นดังนี้

	ศึกษาศาสตร์	วิทยาศาสตร์	รัฐศาสตร์	รวม
สูบบุหรี่	160.00	91.43	148.57	400
ไม่สูบบุหรี่	120.00	68.57	111.43	300
รวม	280.00	160.00	260.00	700

จากสูตร $\chi^2 = \Sigma \frac{(O-E)^2}{E}$

แทนค่าจากสูตร

$$\chi^2 = \frac{(152-160)^2}{160} + \frac{(72-91.43)^2}{91.43} + \frac{(176-149.57)^2}{148.57}$$

$$+\frac{(128-120)^2}{120} + \frac{(88-68.57)^2}{69.57} + \frac{(84-111.43)^2}{111.43}$$

$$= .4 + 4.13 + 5.06 + 0.53 + 5.51 + 6.77$$

$$= 22.40$$

จากตารางที่ $\alpha = .05$, df $= 2$, $\chi^2 = 5.99$

แต่ค่า χ^2 ที่ได้จากการคำนวณมีค่าเท่ากับ 22.40 ซึ่งมากกว่าค่า χ^2 จากตาราง ดังนั้นเราจึงปฏิเสธสมมุติฐานเป็นกลาง (H_0) ที่ว่าสัดส่วนของนักศึกษาที่สูบบุหรี่ในคณะต่าง ๆ เท่ากัน

แบบฝึกหัด 13

1. จากตารางการแจกแจงแบบ Chi–square จงหาค่า

ก. χ^2 ที่ $\alpha = .025$, df = 15

ข. χ^2 ที่ $\alpha = .01$, df = 7

ค. χ^2 ที่ $\alpha = .05$, df = 24

2. ในการทดลองเกี่ยวกับความดันโลหิตสูง ว่าขึ้นอยู่กับการสูบบุหรี่หรือไม่ ผลปรากฏดังนี้

นิสัยการสูบบุหรี่ / ความดันโลหิต	ไม่เคยสูบบุหรี่	สูบบุหรี่ปานกลาง	สูบบุหรี่หนัก
ความดันโลหิตสูง	21	36	30
ไม่เป็นความดันโลหิตสูง	48	26	19

จงทดสอบดูว่าการเป็นหรือไม่เป็นความดันโลหิตสูง ขึ้นอยู่กับนิสัยของการสูบบุหรี่หรือไม่ ($\alpha = .05$)

3. ในการสำรวจนักศึกษามหาวิทยาลัยรามคำแหงจำนวน 200 คน ปรากฏผลดังตาราง

ระดับชั้น / นิสัย	ปีที่ 1	ปีที่ 2	ปีที่ 3	ปีที่ 4
ดื่มเหล้า	21	33	25	20
ไม่ดื่มเหล้า	47	26	19	9

จงทดสอบดูว่าการดื่มเหล้าขึ้นอยู่กับปีที่ศึกษาหรือไม่ ($\alpha = .05$)

4. ในการสำรวจพ่อบ้าน 200 คน ปรากฏผลดังนี้

ระดับการศึกษา	จำนวนบุตร		
	0-1	2-3	มากกว่า 3
ประถมศึกษา	14	37	32
มัธยมศึกษา	19	42	17
อุดมศึกษา	12	17	10

จงทดสอบว่าจำนวนบุตรขึ้นอยู่กับระดับการศึกษาของบิดาหรือไม่ ($\alpha = .05$)

5. ในการสำรวจเด็ก 90 คน เกี่ยวกับจำนวนชั่วโมงที่ดูโทรทัศน์ใน 1 เดือน ผลปรากฏดังนี้

จำนวนชั่วโมงดูโทรทัศน์ \ เพศ	ชาย	หญิง
ดูมากกว่า 25 ชั่วโมง	15	29
ดูน้อยกว่า 25 ชั่วโมง	27	19

จงทดสอบดูว่าจำนวนชั่วโมงที่ดูโทรทัศน์ขึ้นอยู่กับเพศหรือไม่ ($\alpha = .01$)