บทที่ 12
ไค-สแควร์
(CHI-SQUARE)

วัตถุประสงค์

เมื่อท่านศึกษาเนื้อหาบทที่ 12 โดยละเอียดแล้ว ควรมีความสามารถ ดังนี้

1. บอกลักษณะการกระจายของไค-สแควร์ได้
2. ใช้ไค-สแควร์ทดสอบความสอดคล้องของความถี่ได้
3. ใช้ไค-สแควร์ทดสอบความเป็นอิสระของตัวแปรได้

เนื้อหา

12.1 การกระจายของไค-สแควร์

12.2 การทดสอบความสอดคล้องของความถี่ (Test of Goodness of Fit)

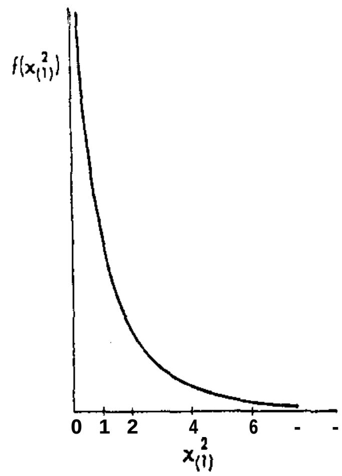
12.3 การทดสอบความเป็นอิสระของตัวแปร (Test of Independent)

เนื้อหาที่ 12.1 การกระจายของไค-สแควร์

ถ้าตัวแปร X มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 เราสามารถเปลี่ยนตัวแปร X ให้อยู่ในรูปของหน่วยมาตรฐาน Z ซึ่งมีค่าเฉลี่ย เท่ากับ 0 และความแปรปรวนเท่ากับ 1 ได้ และกำลังสองของหน่วยมาตรฐาน Z จะมีการกระจายแบบไค-สแควร์ ซึ่งมีชั้นแห่งความอิสระ (df) เท่ากับ 1 เขียนเป็นสัญลักษณ์ได้ดังนี้

$$Z_1^2 = \chi_1^2$$

การกระจายของไค-สแควร์ ซึ่งมีชั้นแห่งความอิสระเท่ากับ 1 (n = 2) เขียนกราฟได้ดังรูป 12.1

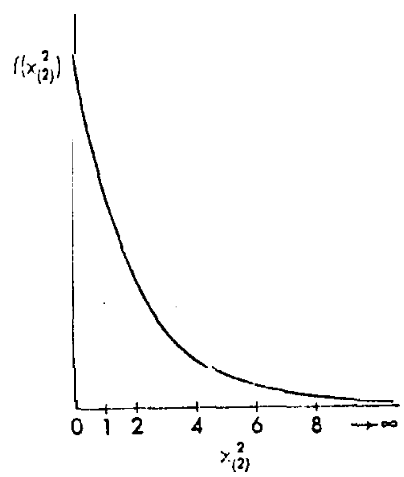


รูป 12.1 *การกระจายของไค-สแควร์เมื่อ* df = 1

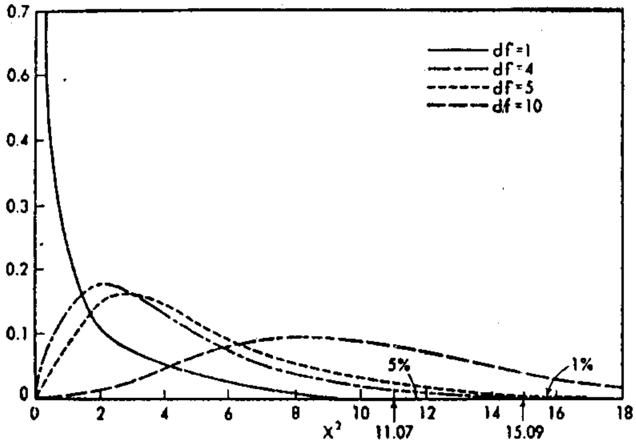
โดยทั่ว ๆ ไป ถ้ามีตัวแปรอิสระ n ตัวแล้ว

$$Z_1^2 + Z_2^2 + \text{----------------} + Z_n^2 = \chi_n^2$$

ซึ่งถ้า n มีค่าน้อย การกระจายของไค-สแควร์จะมีลักษณะเบ้ไปทางบวก (positively skewed) และถ้า n มีค่าเพิ่มขึ้นโค้งจะลดความเบ้ลงและจะมีลักษณะสมมาตร (symmetry) มากขึ้น ดังรูป 12.2 และ 12.3



รูป 12.2 *การกระจายของไค-สแควร์เมื่อ* df = 2



รูป 12.3 การกระจายของไค-สแควร์ เมื่อ df มีค่าต่าง ๆ

เนื้อหาที่ 12.2 การทดสอบความสอดคล้องของความถี่

เป็นการใช้ไค-สแควร์ทดสอบดูว่าความถี่ หรือสิ่งที่เกิดขึ้นซึ่งเราสังเกตเห็นได้ (observed frequency) แตกต่างจากความถี่ที่จะเกิดขึ้นตามทฤษฎี หรือตามที่คาดหวัง (expected frequency) หรือไม่ ในกรณีนี้ไค-สแควร์มีค่าชั้นแห่งความอิสระ (df) เท่ากับ n-1

ข้อตกลงเบื้องต้น

ตัวแปร อยู่ในมาตรานามบัญญัติ โดยนับเป็นความถี่ หรือจำนวนการแจกแจงของข้อมูล

สถิติที่ใช้ในการทดสอบสมมุติฐาน

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \ df = n-1 \qquad ....(12.1)$$

เมื่อ O = ความถี่ที่สังเกตเห็นได้

E = ความถี่ที่คาดว่าจะเป็น หรือความถี่ที่คาดว่าจะเกิดขึ้นตามทฤษฎี

n = จำนวนกลุ่ม หรือประเภทของกลุ่มตัวอย่าง

สมมุติฐาน

H_O : ความถี่ที่สังเกตเห็นเท่ากับความถี่ที่คาดหวัง

H_1 : ความถี่ที่สังเกตเห็นไม่เท่ากับความถี่ที่คาดหวัง

ขอบเขตวิกฤต

จะปฏิเสธสมมุติฐานกลาง ถ้าค่า χ^2 ที่คำนวณได้มากกว่าค่า $\chi^2_{\infty;\, df = n-1}$

ตัวอย่าง 12.1 ในการโยนลูกเต๋า 1 ลูก 600 ครั้ง ปรากฏผลดังนี้

หน้าของลูกเต๋า	1	2	3	4	5	6
ความถี่ที่สังเกต	89	113	98	104	117	79

จากข้อมูลข้างบนนี้ อยากทราบว่า มีการถ่วงลูกเต๋าหรือไม่ที่ระดับนัยสำคัญทางสถิติ 0.05

วิธีทำ

1) คำนวณหาความถี่ที่คาดหวัง (E)

เนื่องจากในการทอดลูกเต๋า 1 ลูก 1 ครั้ง โอกาสที่จะขึ้นหน้าต่าง ๆ ของลูกเต๋าเท่ากับ $\frac{1}{6}$

ดังนั้น ถ้าทอดลูกเต๋า 600 ครั้ง ในทางทฤษฎีแล้ว โอกาสที่จะขึ้นหน้าต่าง ๆ ของลูกเต๋าเท่ากับ $\frac{1}{6} \times 600 = 100 = E$

นำค่าความถี่ที่คาดหวัง (E) ไปเขียนลงในตารางเปรียบเทียบกับค่าความถี่ที่สังเกตเห็น (0) ได้ดังนี้

หน้าของลูกเต๋า	1	2	3	4	5	6
ความถี่ที่สังเกต (0)	89	113	98	104	117	79
ความถี่ที่คาดหวัง (E)	100	100	100	100	100	100

2) สมมุติฐาน

H_0 : ความถี่ที่สังเกตเห็นเท่ากับความถี่ที่คาดหวัง

H_1 : ความถี่ที่สังเกตเห็นไม่เท่ากับความถี่ที่คาดหวัง

3) สถิติที่ใช้ในการทดสอบจากสูตร 12.1

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

ในที่นี้ $n = 6$ ดังนั้น $df = 5$

แทนค่าในสูตร 12.1

$$\chi^2 = \frac{(89-100)^2}{100} + \frac{(113-100)^2}{100} + \frac{(98-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(117-100)^2}{100} + \frac{(79-100)^2}{100}$$

$$\chi^2 = \frac{(89-100)^2 + (113-100)^2 + (98-100)^2 + (104-100)^2 + 117-100)^2 + (79-100)^2}{100}$$

$$= \frac{121 + 169 + 4 + 16 + 289 + 441}{100}$$

$$= \frac{1040}{100} = 10.40$$

4) จะปฏิเสธสมมุติฐานกลาง ถ้า χ^2 ที่คำนวณได้มีค่ามากกว่า $\chi^2\,(.05\,;\,5)$

จากการเปิดตาราง χ^2 ที่ $\infty = .05$, $df = 5$ ได้ค่า $\chi^2 = 11.07$

5) เนื่องจาก χ^2 ที่คำนวณได้มีค่าน้อยกว่า χ^2 ที่ได้จากการเปิดตาราง ฉะนั้นเราจึงไม่ปฏิเสธสมมุติฐานกลาง แสดงว่า ความถี่ที่สังเกตได้เท่ากับความถี่ที่คาดหวัง จึงอาจสรุปได้ว่า ลูกเต๋าที่ใช้ทดลองเป็นลูกเต๋าที่มีสมดุล ไม่ถูกถ่วงให้เอียงไปหน้าใดหน้าหนึ่ง

เนื้อหาที่ 12.3 การทดสอบความเป็นอิสระของตัวแปร

เป็นการทดสอบความเป็นอิสระของตัวแปร 2 ตัว ซึ่งเป็นตัวแปรแบบไม่ต่อเนื่อง ในกรณีนี้ χ^2 จะมีชั้นแห่งความอิสระ : (df) เท่ากับ (R-1) (C-1) เมื่อ R คือจำนวนแถว และ C คือจำนวนสดมภ์ ข้อมูลจะมีลักษณะดังตาราง 2 ทาง ข้างล่างนี้

ตัวแปร A

ตัวแปร B	1	2	3	——	k	
1	n_{11}	n_{12}	n_{13}	----------	n_{1k}	R_1
2	n_{21}	n_{22}	n_{23}	---------	n_{2k}	R_2
3	n_{31}	n_{32}	n_{33}	----------	n_{3k}	R_3
–	–		–		–	–
–	–	–	–		–	–
–	–	–			–	–
–	–			----------		–
j	n_{j1}	n_{j2}	n_{j3}	----------	n_{jk}	R_j
	C_1	C_2	C_3	----------	C_k	

เราสามารถคำนวณหาความถี่ที่คาดหวัง (E) ในทุก ๆ เซลล์ได้จากสูตร

$$E_{jk} = \frac{R_j \times C_k}{N}$$

เมื่อ R_j = ผลรวมในแถว j

C_k = ผลรวมในสดมภ์ k

N = จำนวนความถี่ทั้งหมด

ข้อตกลงเบื้องต้นสำหรับการใช้ไค-สแควร์ทดสอบความเป็นอิสระของตัวแปร 2 ตัว

ตัวแปร 2 ตัวเป็นตัวแปรแบบไม่ต่อเนื่อง

<u>สถิติที่ใช้ในการทดสอบสมมุติฐาน</u>

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \text{ df} = (R-1)(C-1) \quad ....(12.2)$$

เมื่อ O = ความถี่ที่สังเกตเห็น

E = ความถี่ที่คาดหวัง

n = จำนวน 1 เซลล์ในตาราง 2 ทาง

<u>สมมุติฐาน</u>

H_0 : ตัวแปร A และตัวแปร B เป็นอิสระจากกัน

H_1 : ตัวแปร A และตัวแปร B ไม่เป็นอิสระจากกัน

<u>ขอบเขตวิกฤต</u>

จะปฏิเสธสมมุติฐานกลาง ถ้า χ^2 ที่คำนวณได้มีค่ามากกว่าค่า $\chi^2_{\infty, \text{ df} = (R-1)(C-1)}$

<u>ตัวอย่าง 12.2</u> จากการสำรวจสถิติการอ่านหนังสือพิมพ์สยามรัฐกับบ้านเมือง ของผู้ที่สำเร็จการศึกษาชั้นปริญญาตรีขึ้นไป กับผู้ที่ไม่ได้ปริญญา ปรากฏผลดังตารางข้างล่าง

	สยามรัฐ	บ้านเมือง	
ได้ปริญญา	117	74	191 = R_1
ไม่ได้ปริญญา	83	126	209 = R_2
	200 = C_1	200 = C_2	400 = N

อยากทราบว่าการเลือกอ่านหนังสือพิมพ์กับระดับการศึกษามีความสัมพันธ์เกี่ยวข้องกันหรือไม่ ที่ระดับนัยสำคัญทางสถิติ .05

วิธีทำ

คำนวณหาความถี่ที่คาดหวังของแต่ละเซลล์คือ

$$E_{11} = \frac{R_1C_1}{N} = \frac{(191)(200)}{400} = 95.5$$

$$E_{12} = \frac{R_1C_2}{N} = \frac{(191)(200)}{400} = 95.5$$

$$E_{21} = \frac{R_2C_1}{N} = \frac{(209)(200)}{400} = 104.5$$

$$E_{22} = \frac{R_2C_2}{N} = \frac{(209)(200)}{400} = 104.5$$

สมมุติฐาน

H_0 : การเลือกอ่านหนังสือพิมพ์กับระดับการศึกษาเป็นอิสระจากกัน

H_1 : การเลือกอ่านหนังสือพิมพ์กับระดับการศึกษาไม่เป็นอิสระจากกัน

สถิติที่ใช้ในการทดสอบสมมุติฐาน

จากสูตร 12.2 $\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$, df = (R−1) (C−1)

$$df = (2-1)(2-1) = 1$$

$$\chi^2 = \frac{(117-95.5)^2}{95.5} + \frac{(74-95.5)^2}{95.5} + \frac{(83-104.5)^2}{104.5} + \frac{(126-104.5)^2}{104.5}$$

$$= 18.53$$

ขอบเขตวิกฤต

จะปฏิเสธสมมุติฐานกลางถ้า χ^2 ที่คำนวณได้มีค่ามากกว่า $\chi^2_{.05;\,(2-1)\,(2-1)}$

จากการเปิดตาราง χ^2 ที่ $\infty = .05$ df $= 1$ มีค่า 3.841

สรุปผล

เนื่องจาก χ^2 ที่คำนวณได้มีค่ามากกว่า χ^2 ที่ได้จากการเปิดตาราง ฉะนั้นจึงปฏิเสธสมมุติฐานกลาง แสดงว่าการเลือกอ่านหนังสือพิมพ์กับระดับการศึกษามีความสัมพันธ์เกี่ยวข้องกัน

สรุปเนื้อหาบทที่ 12

1. ถ้าตัวแปร X มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับมิว (μ) และความแปรปรวนเท่ากับซิกมากำลังสอง (σ^2) เราสามารถเปลี่ยนตัวแปร X ให้อยู่ในรูปของหน่วยมาตรฐาน ซี (Z) ซึ่งมีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับ 1 ได้ และกำลังสองของหน่วยมาตรฐาน ซี จะมีการกระจายแบบไค-สแคว์ ซึ่งมีชั้นแห่งความอิสระ (df) เท่ากับ 1

2. ถ้าตัวแปรอิสระ (n) มีค่าน้อย การกระจายของไค-สแคว์จะมีลักษณะเบ้ไปทางบวก และถ้าตัวแปรอิสระ (n) มีค่าเพิ่มขึ้นโค้งจะลดความเบ้ลง และจะมีลักษณะสมมาตรมากขึ้น

3. สูตร $\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$, df = n-1 ใช้สำหรับทดสอบดูว่าความถี่ที่เกิดขึ้นจากการทดสอบซึ่งสามารถสังเกตเห็นได้ แตกต่างไปจากความถี่ที่คาดว่าจะเกิดขึ้นโดยทฤษฎีหรือไม่

4. สูตร $\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$, df = (R-1) (G-1) ใช้สำหรับทดสอบความเป็นอิสระของตัวแปร 2 ตัว เมื่อข้อมูลอยู่ในมาตรานามบัญญัติ

คำถามท้ายบทที่ 12

1. จากการสำรวจสภาพการใช้ถนนแห่งหนึ่งในกรุงเทพฯ ในช่วงเวลา 6.00 น.-10.00 น. ปรากฏว่าจำนวนการใช้ถนนแต่ละเลนของรถที่ขับผ่านมีดังนี้

ถนนเลนที่	1	2	3	4
จำนวนรถที่ขับผ่าน	294	276	238	192

   อยากทราบว่า การใช้ถนนในแต่ละเลนเป็นไปตามทฤษฎีความน่าจะเป็นหรือไม่

2. ในการศึกษาประสิทธิภาพของการใช้ยากับคนไข้ประเภทหนึ่ง โดยมีกลุ่มทดลองใช้ยากับกลุ่มควบคุม ได้ผลดังนี้

	กลุ่มทดลอง	กลุ่มควบคุม
อาการดีขึ้น	117	74
อาการไม่ดีขึ้นเลย	83	126

   อยากทราบว่าข้อมูลข้างบนมีประสิทธิภาพพอที่จะระบุว่าการใช้ยาชนิดนี้ทำให้อาการของคนไข้ดีขึ้นที่ระดับนัยสำคัญทางสถิติ .05

3. ในการสอบถามความคิดเห็นเกี่ยวกับการขึ้นราคาค่ารถเมล์กับกลุ่มตัวอย่าง 48 คน ปรากฏว่ามีผู้ไม่เห็นด้วย 24 คน เห็นด้วย 12 คน เฉย ๆ 12 คน อยากทราบว่า ความเห็นนี้เป็นไปตามทฤษฎีความน่าจะเป็นหรือไม่ ที่ระดับนัยสำคัญทางสถิติ .05

4. ผลจากการวัดเจตคติเกี่ยวกับวิธีการเลือกตั้งกับนักศึกษาจำนวน 100 คน ได้ผลดังข้างล่าง อยากทราบว่าคำตอบนี้มีการแจกแจงตรงข้ามกับการแจกแจงที่คาดหวังหรือไม่ ถ้าเราคาดว่าไม่มีความแตกต่างระหว่างกลุ่ม (∞ .05) ผลที่ได้จากสอบถามมีดังนี้

เห็นด้วยอย่างยิ่ง	23 คน
เห็นด้วย	18 คน
เฉย ๆ	24 คน
ไม่เห็นด้วย	17 คน
ไม่เห็นด้วยอย่างยิ่ง	18 คน

5. ผลจากการสำรวจความถนัดในการใช้มือ และการใช้ตาของนักเรียน 413 คนเป็นดังนี้

	ถนัดตาซ้าย	ถนัดทั้งสองตา	ถนัดตาขวา	รวม
ถนัดมือซ้าย	34	62	28	124
ถนัดทั้ง 2 มือ	27	28	20	75
ถนัดมือขวา	57	105	52	214
รวม	118	195	100	413

อยากทราบว่าความถนัดในการใช้มือกับความถนัดในการใช้ตาเป็นอิสระจากกันหรือไม่ ที่ระดับนัยสำคัญทางสถิติ .05

6. ในการตัดเกรดวิชาหนึ่งมีผู้เรียน 100 คน ครูตัดเกรดเป็น 5 เกรด ดังนี้ เกรด เอ ให้ 2 คน เกรด บี ให้ 20 คน เกรด ซี ให้ 60 คน เกรด ดี ให้ 15 คน และ เกรด อี ให้ 3 คน แต่จากการตัดเกรดโดยยึดหลัก normal curve นั้นควรให้เกรดดังนี้

คะแนนที่ตกอยู่ระหว่าง ± 1 SD	จากคะแนนเฉลี่ยให้ เกรด ซี
คะแนนที่ตกอยู่ระหว่าง + 1 SD	ถึง + 2 SD ให้ เกรด บี
คะแนนที่มีค่ามากกว่า + 2 SD	ให้เกรด เอ
คะแนนที่ตกอยู่ระหว่าง − 1 SD	ถึง − 2 SD ให้เกรด ดี
คะแนนที่มีค่าน้อยกว่า − 2 SD	ให้เกรด อี

อยากทราบว่าการตัดเกรดเป็นไปตามหลัก normal curve หรือไม่ ที่ระดับนัยสำคัญทางสถิติ .05 ในการตอบคำถามข้อนี้ ให้ท่านแสดงรายละเอียดในเรื่องเหล่านี้

ก. สมมุติฐาน
ข. สถิติที่จะใช้ในการทดสอบสมมุติฐาน
ค. แทนค่าในสูตรและแสดงให้เห็นว่าได้ตัวเลขเหล่านั้นมาอย่างไร
ง. degree of freedom
จ. สรุปผล