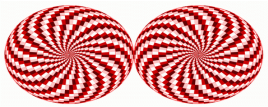
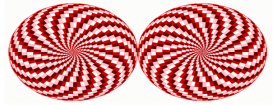
การทดสอบไคสแควร์

Chi-Square Test



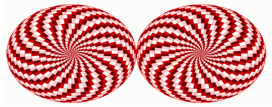
การทดสอบไคสแควร์ (Chi-Square Test)

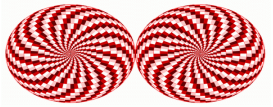


เป็นเทคนิคการวิเคราะห์ข้อมูลที่ใช้กับ

ข้อมูลเชิงคุณภาพ(Qualitative Data)

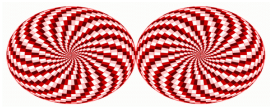
ข้อมูลจำแนกประเภท(Categorical Data)





ตัวเลขที่นำมาวิเคราะห์เป็นความถี่

ของแต่ละกลุ่มของข้อมูล(ตัวแปร)ที่ต้องการศึกษา



ข้อมูลในรูปของความถี่

1. ข้อมูลจำแนกทางเดียว (One-way table) หรือตารางแจกแจงความถี่แบบทางเดียว (One - Way Frequency Table) เป็นตารางที่จำแนกข้อมูลเป็นกลุ่ม ๆ เพียงด้านเดียว เช่น

ตาราง จำนวนนักศึกษาจำแนกตามวิชาที่ชอบ

วิชา	จำนวน
คณิตศาสตร์ + สถิติ	15
ภาษาอังกฤษ	20
วิทยาศาสตร์	12
รวม	47

ตาราง จำนวนผู้หญิงจำแนกตามกลุ่มเลือด

กลุ่มเลือด	จำนวนคน
O	43
A	66
B	11
AB	4
รวม	124

ตาราง จำนวนอุบัติเหตุทางรถยนต์จำแนกตามวัน

วัน	จันทร์	อังคาร	พุธ	พฤหัสฯ	ศุกร์	เสาร์	อาทิตย์	รวม
จำนวนอุบัติเหตุทางรถยนต์	100	75	85	113	49	62	105	589

ตาราง จำนวนผู้เล่นดนตรีจำแนกตามชนิดของดนตรี

ชนิดของดนตรี	จำนวนผู้เล่นดนตรี
Classical	8
Country	210
Gospel	72
Oldies	10
Pop	75
Rock	125
รวม	500

2. ข้อมูลจำแนกสองทาง (Two - Way Table) หรือตารางแจกแจงความถี่จำแนกสองทางหรือ ตารางการณ์จร (Two - Way Contingency Table) หรือตารางไข้ว (Crosstab table) เป็นตารางที่จำแนกข้อมูลเป็นกลุ่ม ๆ โดย 2 ลักษณะ เช่น

ตาราง จำนวนนักศึกษาจำแนกตามเพศและระดับความสนใจศึกษาต่อ

เพศ	ความสนใจศึกษาต่อ			รวม
	สนใจมาก	สนใจ	ไม่สนใจ	
ชาย	80	120	20	220
หญิง	110	175	55	340
รวม	190	295	75	560

ตาราง จำนวนนักเรียนชั้นประถมจำแนกตามความผิดปกติของการพูดกับฐานะทางเศรษฐกิจสังคมของพ่อกับแม่

ความผิดปกติของการพูด (speech defect)	ฐานะเศรษฐกิจสังคม				รวม
	สูง	สูงปานกลาง	ต่ำปานกลาง	ต่ำ	
ปกติ	8	24	32	27	91
ผิดปกติ	42	121	138	108	409
รวม	50	145	170	135	500

ข้อตกลงเบื้องต้น

1. ตัวอย่างแต่ละชุดเป็นตัวอย่างที่ได้จากการสุ่ม
2. ผลที่เกิด(outcomes) ของตัวอย่างเป็นอิสระต่อกัน (mutually independence)
3. ค่าสังเกตแต่ละต้องถูกจัดให้อยู่ในชั้นใดชั้นหนึ่งเท่านั้น

ตาราง จำนวนผู้ใช้คอมพิวเตอร์จำแนกตามพฤติกรรมการเล่น Facebook และช่วงอายุ

พฤติกรรมการเล่น Facebook	ช่วงอายุ			รวม
	ไม่เกิน 20 ปี	21 – 30 ปี	มากกว่า 30 ปี	
เล่น	80	120	20	220
ไม่เล่น	110	175	55	340
รวม	190	295	75	560

การทดสอบไคสแควร์

จึงขึ้นอยู่กับลักษณะการจำแนกข้อมูล

การทดสอบไคสแควร์

ข้อมูลจำแนกทางเดียว

เรียกว่า

- การทดสอบภาวะสารูปสนิทดี (Goodness of fit test)
- การทดสอบลักษณะต่าง ๆ ของประชากรว่าเป็นไปตามที่คาดไว้หรือไม่
- การทดสอบการแจกแจงของข้อมูลว่าเป็นไปตามที่คาดไว้หรือไม่

ข้อมูลจำแนกสองทาง

เรียกว่า

การทดสอบความเป็นอิสระ (Testing of Independence)

ตัวแปร 2 ตัว มีความสัมพันธ์กันหรือไม่

หลักการทดสอบไคสแควร์

ความถี่คาดว่าจะเกิดขึ้นหรือเป็นไปตามทฤษฎีเรียกว่าความถี่คาดหวัง (Expected Frequency) แทนด้วยสัญลักษณ์ E_i	เปรียบเทียบ	ความถี่ที่สังเกตได้จากตัวอย่างเรียกว่าความถี่จากการสังเกต (Observed Frequency) แทนด้วยสัญลักษณ์ O_i

ว่ามีความแตกต่างกันมากเกินไปหรือไม่

ข้อมูลกลุ่มที่	1	2	3	....	k	รวม
ความถี่จากการสังเกต	O_1	O_2	O_3	....	O_k	n
ความถี่คาดหวัง	E_1	E_2	E_3	....	E_k	n

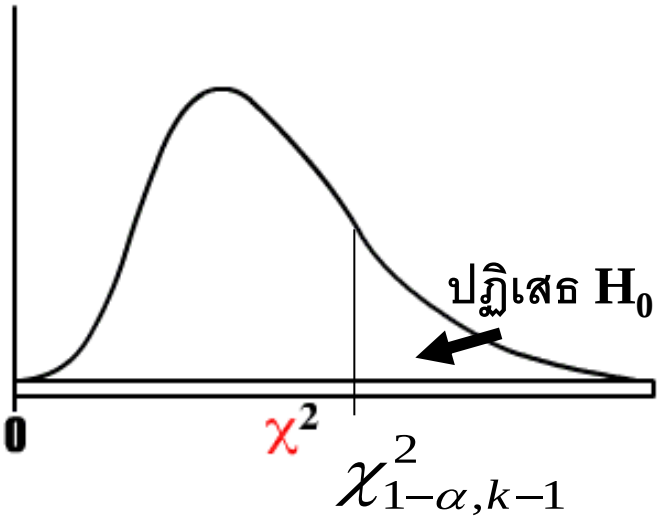
$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = n$$

ข้อมูลกลุ่มที่	1	2	3	...	k
ความถี่ที่สังเกตได้	O_1	O_2	O_3	...	O_k
ความถี่คาดหวัง	E_1	E_2	E_3	...	E_k
$(O_i - E_i)$	$(O_1 - E_1)$	$(O_2 - E_2)$	$(O_3 - E_3)$	...	$(O_k - E_k)$
$(O_i - E_i)^2$	$(O_1 - E_1)^2$	$(O_2 - E_2)^2$	$(O_3 - E_3)^2$	...	$(O_k - E_k)^2$
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(O_1 - E_1)^2}{E_1}$	$\frac{(O_2 - E_2)^2}{E_2}$	$\frac{(O_3 - E_3)^2}{E_3}$	...	$\frac{(O_k - E_k)^2}{E_k}$

ตัวสถิติทดสอบ

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

ค่าวิกฤต $\chi^2_{1-\alpha,k-1}$



หมายเหตุ

การทดสอบ χ^2 จะมีประสิทธิภาพ ความถี่คาดหวังในแต่ละช่องควรมีค่าไม่ต่ำกว่า 5

ลักษณะข้อมูลจำแนก 2 ทาง

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	0_{11}	0_{12}	0_{13}	…	0_{1c}	r_1
2	0_{21}	0_{22}	0_{23}	…	0_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	0_{r1}	0_{r2}	0_{r3}	…	0_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

โดยที่ 0_{ij} คือ ความถี่ของตัวแปรที่ 1 ระดับ i และตัวแปรที่ 2 ระดับ j

เมื่อ i = 1,…,r j = 1,…,c

r_i คือ ความถี่รวมของตัวแปรที่ 1 ระดับ i ; i = 1,…,r

c_j คือ ความถี่รวมของตัวแปรที่ 2 ระดับ j ; j = 1,…c

ความถี่รวมทั้งหมด = $n = \sum r_i = \sum c_j = \sum\sum O_{ij}$

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	O_{11}	O_{12}	O_{13}	…	O_{1c}	r_1
2	O_{21}	O_{22}	O_{23}	…	O_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	O_{r1}	O_{r2}	O_{r3}	…	O_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	E_{11}	E_{12}	E_{13}	…	E_{1c}	r_1
2	E_{21}	E_{22}	E_{23}	…	E_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	E_{r1}	E_{r2}	E_{r3}	…	E_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

ตัวสถิติทดสอบ

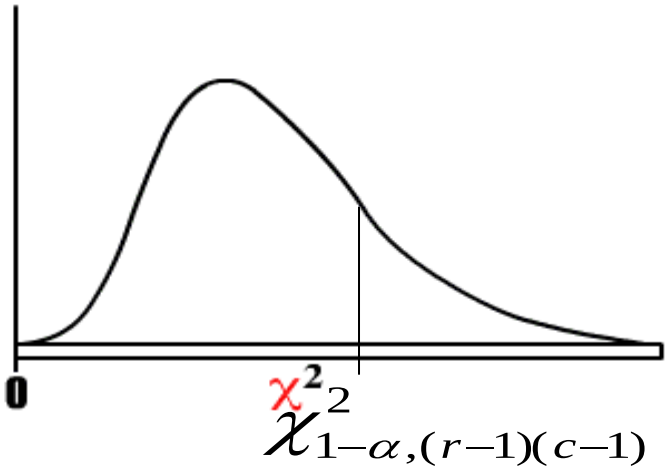
$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

O_{ij} ความถี่สังเกตใน cell(i,j)

E_{ij} ความถี่คาดหวังใน cell(i,j)

เมื่อ $i = 1,\ldots,r \quad j = 1,\ldots,c$

ค่าวิกฤต ⇒ $\chi^2_{1-\alpha,(r-1)(c-1)}$



การทดสอบความเป็นอิสระ (Testing of Independence)

เป็นการศึกษาว่า ตัวแปร 2 ตัวเป็นอิสระกัน

ตัวแปร 2 ตัวมีความสัมพันธ์กัน

น้ำหนักของเด็กแรกเกิดกับปริมาณการสูบบุหรี่ของแม่ขณะตั้งครรภ์มีความสัมพันธ์กัน

ความชอบเรียนคณิตศาสตร์ขึ้นอยู่กับเพศ

ความรุนแรงของโรคกระเพาะอาหารมีความสัมพันธ์กับกลุ่มเลือด

สถานะทางครอบครัวมีผลต่อพฤติกรรมทางเพศของเด็กวัยรุ่น

สมมติฐานเชิงสถิติ

H_0 : ตัวแปรที่ 1 ไม่มีความสัมพันธ์กับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีความสัมพันธ์กับตัวแปรที่ 2

ตัวอย่าง

H_0 : ความชอบเรียนคณิตศาสตร์เป็นอิสระกับเพศ

H_1 : ความชอบเรียนคณิตศาสตร์ไม่เป็นอิสระกับเพศ

H_0 : ความรุนแรงของโรคกระเพาะอาหารไม่มีความสัมพันธ์กับกลุ่มเลือด

H_1 : ความรุนแรงของโรคกระเพาะอาหารมีความสัมพันธ์กับกลุ่มเลือด

H_0 : สถานะทางครอบครัวไม่เกี่ยวข้องกับพฤติกรรมทางเพศของเด็กวัยรุ่น

H_1 : สถานะทางครอบครัวเกี่ยวข้องกับพฤติกรรมทางเพศของเด็กวัยรุ่น

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	0_{11}	0_{12}	0_{13}	…	0_{1c}	r_1
2	0_{21}	0_{22}	0_{23}	…	0_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	0_{r1}	0_{r2}	0_{r3}	…	0_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	E_{11}	E_{12}	E_{13}	…	E_{1c}	r_1
2	E_{21}	E_{22}	E_{23}	…	E_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	E_{r1}	E_{r2}	E_{r3}	…	E_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

ตัวสถิติทดสอบ

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

O_{ij} ความถี่สังเกตใน cell(i,j)

E_{ij} ความถี่คาดหวังใน cell(i,j)

เมื่อ $i = 1,\ldots,r \quad j = 1,\ldots,c$

ค่าวิกฤต ⇒ $\chi^2_{1-\alpha,(r-1)(c-1)}$

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	E_{11}	E_{12}	E_{13}	…	E_{1c}	r_1
2	E_{21}	E_{22}	E_{23}	…	E_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	E_{r1}	E_{r2}	E_{r3}	…	E_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

$$E_{ij} = \frac{r_i c_j}{n}$$

ตัวแปรที่ 1	ตัวแปรที่ 2					ผลรวม
	1	2	3	…	c	
1	0_{11}	0_{12}	0_{13}	…	0_{1c}	r_1
2	0_{21}	0_{22}	0_{23}	…	0_{2c}	r_2
⋮	⋮	⋮	⋮	…	⋮	⋮
r	0_{r1}	0_{r2}	0_{r3}	…	0_{rc}	r_r
ผลรวม	c_1	c_2	c_3	…	c_c	$\sum r_i = \sum c_j = n$

A

B

$$P(A_1) = \frac{r_1}{n} \qquad P(B_1) = \frac{c_1}{n}$$

$$P(A_1 \cap B_1) = P(A_1)P(B_1) = \frac{r_1}{n}\frac{c_1}{n} \Rightarrow P_{11} \Rightarrow E_{11} = np_{11} = n\frac{r_1}{n}\frac{c_1}{n} = \frac{r_1 c_1}{n}$$

$$P(A_2) = \frac{r_2}{n} \qquad P(B_2) = \frac{c_2}{n}$$

$$P(A_2 \cap B_2) = P(A_2)P(B_2) = \frac{r_2}{n}\frac{c_2}{n} \Rightarrow P_{22} \Rightarrow E_{22} = np_{22} = n\frac{r_2}{n}\frac{c_2}{n} = \frac{r_2 c_2}{n}$$

$$P(A_i) = \frac{r_i}{n} \qquad P(B_j) = \frac{c_j}{n}$$

$$P(A_i \cap B_j) = P(A_i)P(B_j) = \frac{r_i}{n}\frac{c_j}{n} \Rightarrow P_{ij} \Rightarrow E_{ij} = np_{ij} = n\frac{r_i}{n}\frac{c_j}{n} = \frac{r_i c_j}{n}$$

ตัวอย่าง

จงหาความถี่คาดหวังสำหรับแต่ละเซลในตารางการณ์จร จำนวนผู้ป่วย จำแนกตามการเป็นโรคติดสุราและอายุ

การเป็นโรคติดสุรา	อายุ					
	ต่ำกว่า 39	40-49	50-59	60-69	มากกว่า 70	รวม
เป็น	42	69	108	60	21	300
ไม่เป็น	5	18	85	120	22	250
รวม	47	87	193	180	43	550

ความถี่คาดหวัง $E_{ij} = \frac{r_i c_j}{n}$

$$E_{11} = \frac{300 \times 47}{550} \approx 25.64 \quad E_{12} = \frac{300 \times 87}{550} \approx 47.45 \quad E_{13} = \frac{300 \times 193}{550} \approx 105.27$$

$$E_{14} = \frac{300 \times 180}{550} \approx 98.18 \quad E_{15} = \frac{300 \times 43}{550} \approx 23.45 \quad E_{21} = \frac{250 \times 47}{550} \approx 21.36$$

$$E_{22} = \frac{250 \times 87}{550} \approx 39.55 \quad E_{23} = \frac{250 \times 193}{550} \approx 87.73 \quad E_{24} = \frac{250 \times 180}{550} \approx 81.82$$

$$E_{25} = \frac{250 \times 43}{550} \approx 19.55$$

การเป็นโรคติดสุรา	อายุ					
	ต่ำกว่า 39	40-49	50-59	60-69	มากกว่า 70	รวม
เป็น	25.64	47.45	105.27	98.18	23.45	300
ไม่เป็น	21.36	39.55	87.73	81.82	19.55	250
รวม	47	87	193	180	43	550

ตัวอย่าง บริษัทผลิตมือถือแห่งหนึ่งต้องการทราบว่าความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ จึงทำการเก็บรวบรวมข้อมูลจากผู้ใช้โทรศัพท์มือถือจำนวน 1,000 คน เป็นเพศชาย 500 คน และ หญิง 500 คน ได้ผลดังนี้

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	328	172	500
หญิง	138	362	500
รวม	466	534	1,000

อยากทราบว่าความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ ที่ระดับนัยสำคัญ 0.05

สมมติฐานเชิงสถิติ

H_0 : ความชอบสีของโทรศัพท์มือถือไม่มีความสัมพันธ์กับเพศของผู้ใช้

H_1 : ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้

ค่าความถี่คาดหวัง(E_{ij})

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	$E_{11} = \frac{500 \times 466}{1000} = 233$	$E_{12} = \frac{500 \times 534}{1000} = 267$	500
หญิง	$E_{21} = \frac{500 \times 466}{1000} = 233$	$E_{22} = \frac{500 \times 534}{1000} = 267$	500
รวม	466	534	1,000

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	328	172	500
หญิง	138	362	500
รวม	466	534	1,000

O_{ij}

E_{ij}

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	233	267	500
หญิง	233	267	500
รวม	466	534	1,000

ตัวสถิติทดสอบ

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

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)} = \chi^2_{0.95,1} = 3.84$

$$= \frac{(328-233)^2}{233} + \frac{(172-267)^2}{267} + \frac{(138-233)^2}{233} + \frac{(362-267)^2}{267}$$

$$= 145.07$$

เนื่องจาก $\chi^2 = 145.07$ อยู่ในบริเวณปฏิเสธ H_o หมายความว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง บริษัทแห่งหนึ่งต้องการเปิดทำงานในช่วงวันเสาร์และอาทิตย์ โดยคิดว่าพนักงานต้องการมีรายได้เสริม เพื่อการตัดสินใจที่ถูกต้องจึงทำการเก็บรวบรวมข้อมูลจากคนงานในบริษัทจำนวน 250 คน โดยแบ่งรายได้ออกเป็น 2 ช่วงคือ รายได้ต่ำกว่า 10,000 บาทและตั้งแต่ 10,000 บาทขึ้นไป และความคิดเห็นเป็น 3 ระดับ ได้ผลดังนี้

รายได้	ความคิดเห็น			รวม
	เห็นด้วย	ไม่มีความเห็น	ไม่เห็นด้วย	
ต่ำกว่า 10,000	68 (88.16)	12 (10.64)	110 (91.2)	190
ตั้งแต่ 10,000 ขึ้นไป	48 (27.84)	2 (3.36)	10 (28.8)	60
รวม	116	14	120	250

สมมติฐานเชิงสถิติ

H_O : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ไม่ขึ้นกับรายได้

H_1 : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นกับรายได้

ตัวสถิติทดสอบ

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

$$= \frac{(68-88.16)^2}{88.16} + \frac{(12-10.64)^2}{10.64} + \frac{(110-91.20)^2}{91.20} + \ldots + \frac{(10-28.80)^2}{28.80}$$

$$= 36.08$$

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)} = \chi^2_{0.95,2} = 5.99$

เนื่องจาก $\chi^2 = 36.08$ อยู่ในบริเวณปฏิเสธ H_0 หมายความว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นกับรายได้ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง เพื่อที่จะตรวจสอบว่าผลสัมฤทธิ์ของพนักงานในขณะที่เข้าโครงการฝึกอบรมกับผลสัมฤทธิ์ของพนักงานในการปฏิบัติหน้าที่ในองค์กรมีความสัมพันธ์กัน จึงได้มีการเลือกตัวอย่างพนักงานมา 400 คน แล้วเก็บข้อมูลได้ดังนี้

ผลสัมฤทธิ์ในโครงการฝึกอบรม

ผลสัมฤทธิ์ในหน้าที่การงาน	ต่ำกว่ามาตรฐาน	ปานกลาง	สูงกว่ามาตรฐาน	รวม
พอใช้	23	60	29	112
ดี	28	79	60	167
ดีมาก	9	49	63	121
รวม	60	188	152	400

จงทดสอบสมมติฐานที่ว่าผลสัมฤทธิ์ในโครงการฝึกอบรมกับผลสัมฤทธิ์ในหน้าที่การงานนั้นมีความสัมพันธ์กันที่ระดับนัยสำคัญ 0.01

สมมติฐานเชิงสถิติ

H_0 : ผลสัมฤทธิ์ในโครงการฝึกอบรมและในหน้าที่การงานเป็นอิสระต่อกัน

H_1 : ผลสัมฤทธิ์ในโครงการฝึกอบรมและในหน้าที่การงานไม่เป็นอิสระต่อกัน

ผลสัมฤทธิ์ในหน้าที่การงาน	ผลสัมฤทธิ์ในโครงการฝึกอบรม			
	ต่ำกว่ามาตรฐาน	ปานกลาง	สูงกว่ามาตรฐาน	รวม
พอใช้	23(16.80)	60(52.64)	29(42.56)	112
ดี	28(25.05)	79(78.49)	60(63.46)	167
ดีมาก	9(18.15)	49(56.87)	63(45.98)	121
รวม	60	188	152	400

ตัวสถิติทดสอบ

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

$$= \frac{(23-16.80)^2}{16.80} + \frac{(60-52.64)^2}{52.64} + \frac{(29-45.56)^2}{45.56} + \ldots + \frac{(63-45.98)^2}{45.98}$$

$$= 20.179$$

ค่าวิกฤต

$$\chi^2_{\alpha,\,(r-1)(c-1)} = \chi^2_{0.01,\,(3-1)(3-1)} = \chi^2_{0.01,\,4} = 13.30$$

จากผลการวิเคราะห์ข้อมูล พบว่าปฏิเสธ H_0 หมายความว่า ผลสัมฤทธิ์ในโครงการฝึกอบรมกับผลสัมฤทธิ์ในหน้าที่การงานนั้น มีความสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ 0.01

สัมประสิทธิ์ตารางการณ์จร ของเพียรสั้น (Pearson's contingency coefficient)

เขียนแทนด้วยสัญลักษณ์ C

เป็นตัวเลขที่บอกระดับความสัมพันธ์
(degree of dependence)
ของตัวแปร 2 ตัว จากตารางการณ์จร

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

โดยที่ χ^2 คือ ค่าสถิติที่คำนวณได้
n คือ จำนวนข้อมูล

คุณสมบัติ

1. $0 \leq C \leq 1$
2. C = 0 หมายถึงไม่มีความสัมพันธ์
3. C ยิ่งมากขนาดของความสัมพันธ์ก็ยิ่งมีการขึ้นด้วย

เราสามารถประมาณค่าสูงสุดของ c ได้จาก ⇒ $C_{max} = \sqrt{\frac{q-1}{q}}$

โดยที่ q คือค่า r หรือ c ก็ได้ แล้วแต่ว่าค่าใดมีค่าน้อยกว่า

จากตัวอย่าง

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$= \sqrt{\frac{36.08}{250 + 36.08}}$$

$$= 0.355$$

$$C_{max} = \sqrt{\frac{q-1}{q}}$$

$$= \sqrt{\frac{2-1}{2}}$$

$$= 0.707$$

ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์มีความสัมพันธ์กับรายได้ในระดับ 0.355

สัมประสิทธิ์ตารางการณ์จร ของคราเมอร์(Cramer's ν contingency coefficient) เขียนแทนด้วยสัญลักษณ์ ν

$$\nu = \sqrt{\frac{\chi^2}{n \times \min(r-1, c-1)}}$$

โดยที่ χ^2 คือ ค่าสถิติที่คำนวณได้
n คือ จำนวนข้อมูล

คุณสมบัติ

1. $0 \leq \nu \leq 1$
2. $\nu = 0$ หมายถึงไม่มีความสัมพันธ์
3. ν ยิ่งมากขนาดของความสัมพันธ์ก็ยิ่งมีการขึ้นด้วย

จากตัวอย่าง

$$\nu = \sqrt{\frac{\chi^2}{n \times \min(r-1, c-1)}}$$

$$= \sqrt{\frac{36.08}{250 \times \min(2-1, 3-1)}}$$

$$= \sqrt{\frac{36.08}{250 \times 1}}$$

$$= \sqrt{0.14432}$$

$$= 0.38$$

ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์มีความสัมพันธ์กับรายได้ในระดับ 0.38

สถิติสำหรับงานวิจัย

Statistics for Research

การทดสอบไคสแควร์ด้วย SPSS

คำสั่ง

Analyze→ Descriptive Statistics→Crosstab…

ตัวอย่าง file example15.sav

สมมติฐาน: การนับถือศาสนากับถิ่นที่อยู่อาศัยมีความสัมพันธ์กัน

Area * Regionalism Crosstabulation

			Regionalism			
			Protestant	Christ	Jew	Total
Area	East	Count	182	215	203	600
		Expected Count	201.6	210.6	187.8	600.0
		% within Area	30.3%	35.8%	33.8%	100.0%
		% within Regionalism	54.2%	61.3%	64.9%	60.0%
		% of Total	18.2%	21.5%	20.3%	60.0%
	West	Count	154	136	110	400
		Expected Count	134.4	140.4	125.2	400.0
		% within Area	38.5%	34.0%	27.5%	100.0%
		% within Regionalism	45.8%	38.7%	35.1%	40.0%
		% of Total	15.4%	13.6%	11.0%	40.0%
Total		Count	336	351	313	1000
		Expected Count	336.0	351.0	313.0	1000.0
		% within Area	33.6%	35.1%	31.3%	100.0%
		% within Regionalism	100.0%	100.0%	100.0%	100.0%
		% of Total	33.6%	35.1%	31.3%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.069[a]	2	.018
Likelihood Ratio	8.053	2	.018
Linear-by-Linear Association	7.774	1	.005
N of Valid Cases	1000		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 125.20.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.090	.018
	Cramer's V	.090	.018
N of Valid Cases		1000	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

สมมติฐานเชิงสถิติ

H_O : การนับถือศาสนากับถิ่นที่อยู่อาศัยไม่มีความสัมพันธ์กัน

H_1 : การนับถือศาสนากับถิ่นที่อยู่อาศัยมีความสัมพันธ์กัน

จากตารางการทดสอบไคสแควร์ พบว่าค่าตัวสถิติทดสอบไคสแควร์เท่ากับ 8.069 และ ค่า Sig < 0.05 หมายความว่าการนับถือศาสนากับถิ่นที่อยู่อาศัยมีความสัมพันธ์กันในระดับ 0.09 (Cramer,s V = 0.09) อย่างมีนัยสำคัญที่ระดับ 0.05

Home Work

1. ทดสอบสมมติฐานที่ว่าชั้นปีกับความสำเร็จในการทำงานที่ได้รับมอบหมายมีความสัมพันธ์กัน
2. ทดสอบสมมติฐานที่ว่าชั้นปีกับระดับความสำคัญของปัญหาและอุปสรรคในการทำงานที่ได้รับมอบหมายรวมมีความสัมพันธ์กัน

ชั้นปี	การทำงานของนักศึกษา				รวม
	กลุ่ม 1 ทำงานเสร็จทุกครั้ง	กลุ่ม 2 ทำงานเสร็จเป็นส่วนใหญ่	กลุ่ม 3 ทำงานเสร็จเป็นส่วนน้อย	กลุ่ม 4 ทำงานไม่เสร็จเลย	
ชั้นปี 2					
ชั้นปี 3					
ชั้นปี 4					
รวม					

$$\chi^2 = \qquad , p-value = \qquad , Cramer's V =$$