แผนบริหารการสอนประจำบทที่ 7

หัวข้อเนื้อหา

1. ข้อมูลที่อยู่ในรูปความถี่
2. หลักการทดสอบไคสแควร์
3. ข้อตกลงเบื้องต้น
4. ประเภทของการทดสอบไคสแควร์
5. สัมประสิทธิ์ตารางการณ์จร

วัตถุประสงค์เชิงพฤติกรรม

1. เพื่อให้ผู้เรียนสามารถบอกความหมาย หลักการ และประโยชน์ของการทดสอบไคสแควร์ได้
2. เพื่อให้ผู้เรียนสามารถระบุขั้นตอนในการทดสอบไคสแควร์ได้
3. เพื่อให้ผู้เรียนเลือกประเภทของการทดสอบไคสแควร์ให้เหมาะสมกับวัตถุประสงค์และประเภทของตารางแจกแจงความถี่ได้
4. เพื่อให้ผู้เรียนสามารถอธิบายผลลัพธ์ของการทดสอบไคสแควร์ และนำไปประยุกต์ใช้ในการตัดสินใจ และวางแผนในทางธุรกิจได้

วิธีสอนและกิจกรรมการเรียนการสอน

1. วิธีสอน

1.1 บรรยาย

1.2 ฝึกปฏิบัติในใบกิจกรรม และกรณีศึกษา

2. กิจกรรมการเรียนการสอน

1. ฝึกการทดสอบสมมติฐานโดยใช้การทดสอบไคสแควร์ จากแบบฝึกหัด
2. กำหนดสมมติฐาน และทดสอบสมมติฐานโดยใช้การทดสอบไคสแควร์ในกรณีศึกษา

สื่อการเรียนการสอน

1. โปรแกรมนำเสนอเรื่องการทดสอบไคสแควร์
2. ตารางสถิติ
3. แบบฝึกหัด
4. กรณีศึกษาตัวอย่าง

การวัดผลและการประเมินผล

1. ความตรงต่อเวลา และความตั้งใจในระหว่างเรียน
2. ความตรงต่อเวลาในการส่งงานหรือแบบฝึกหัด
3. สอบย่อยก่อน หรือหลังเรียน

บทที่ 7
การทดสอบไคสแควร์

การเก็บรวบรวมข้อมูลเชิงปริมาณนั้นในบางครั้งเป็นไปได้ยาก เนื่องจากผู้ให้ข้อมูลไม่ต้องการเปิดเผยข้อมูลส่วนตัว เช่นรายได้ต่อเดือน ผู้ให้ข้อมูลมักไม่ต้องการให้ข้อมูล จึงต้องกำหนดรายได้เป็นช่วงเพื่อให้ผู้ให้ข้อมูลให้ข้อมูลได้สะดวกขึ้น แต่จะทำให้รายละเอียดของข้อมูลน้อยลง หรือในบางครั้งข้อมูลที่เราสนใจอาจปีนข้อมูลเชิงคุณภาพ วิธีการวิเคราะห์ข้อมูลลักษณะนี้คือวิธีการทดสอบไคสแควร์ (Chi-Square Test) ซึ่งเป็นเทคนิคการวิเคราะห์ข้อมูลที่ใช้กับข้อมูลเชิงคุณภาพ หรือข้อมูลเชิงปริมาณที่แบ่งเป็นช่วง และข้อมูลไม่จำเป็นต้องมีการแจกแจงแบบปกติ ตัวเลขที่นำมาวิเคราะห์ในทดสอบไคสแควร์เป็นความถี่ของแต่ละระดับของตัวแปรที่ต้องการศึกษา

ข้อมูลที่อยู่ในรูปความถี่

ข้อมูลที่อยู่ในรูปของความถี่แบ่งออกเป็น 2 ลักษณะ ดังนี้

1. ข้อมูลจำแนกทางเดียว (One-Dimensional Classifical Data)

ข้อมูลจำแนกทางเดียว หรือตารางแจกแจงความถี่แบบทางเดียว (One - Way Frequency Table) เป็นข้อมูลที่จำแนกตามลักษณะใดลักษณะหนึ่งเท่านั้น

ตัวอย่าง 7.1 ตารางแสดงข้อมูลจำแนกทางเดียว

ตารางที่ 7.1 จำนวนนักศึกษาโปรแกรมวิชาคณิตศาสตร์จำแนกตามวันที่ขาดเรียน

วัน	จำนวนนักเรียนที่ขาดเรียน (คน)
จันทร์	12
อังคาร	9
พุธ	10
พฤหัสบดี	7
ศุกร์	12
รวม	50

ตารางที่ 7.2 จำนวนนักศึกษาคณะวิทยาศาสตร์จำแนกตามวิชาที่ชอบมากที่สุด

วิชาที่ชอบมากที่สุด	จำนวนที่ชอบวิชา(คน)
คณิตศาสตร์และสถิติ	210
วิทยาศาสตร์	280
ภาษาอังกฤษ	70
รวม	560

ตารางที่ 7.3 จำนวนลูกค้าจำแนกตามการเลือกยี่ห้อโทรศัพท์มือถือ

ยี่ห้อโทรศัพท์	จำนวนลูกค้าทีใช้โทรศัพท์ยี่ห้อ(คน)
Samsung	7,374
Nokia	6,505
i-phone	6,951
รวม	20,830

2. ข้อมูลจำแนกสองทาง (Two - Way Contingency Table)

ข้อมูลจำแนกสองทาง หรือตารางแจกแจงความถี่จำแนก 2 ทาง (One - Way Frequency Table) หรือตารางการณ์จร (Contingency table) หรือตารางไขว้ (Cross tabulation table) เป็นข้อมูลเรื่องใดเรื่องหนึ่งซึ่ง ถูกจำแนกโดย 2 ลักษณะ

ตัวอย่าง 7.2 ตารางข้อมูลจำแนกสองทาง

ตารางที่ 7.4 จำนวนนักศึกษาคณะวิทยาศาสตร์ จำแนกตามเพศและระดับความสนใจศึกษาต่อ

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

ตารางที่ 7.5 จำนวนนักเรียนชั้นประถมจำแนกตามความผิดปกติของการพูด (speech defect) กับฐานะเศรษฐกิจทางสังคมของพ่อกับแม่

ความผิดปกติของการพูด	ฐานะเศรษฐกิจทางสังคม				รวม
	สูง	สูงปานกลาง	ต่ำปานกลาง	ต่ำ	
ปกติ	8	24	32	27	91
ผิดปกติ	42	121	138	108	409
รวม	50	145	170	135	500

ตารางที่ 7.6 จำนวนผู้บริโภคจำแนกตามรายได้ต่อเดือน กับอาชีพ

รายได้(บาท)	อาชีพ			รวม
	รับราชการ	ทำงานเอกชน	ส่วนตัว	
น้อยกว่า 20,000	893	1,138	400	2,431
20,000 – 59,999	908	927	312	2,147
60,000 – 99,999	763	698	912	2,373
ตั้งแต่ 100,000 บาท	469	1,527	1,016	3,012
รวม	3,033	4,290	2,640	9,963

หลักการของการทดสอบไคสแควร์

หลักการของการทดสอบไคสแควร์ คือเปรียบเทียบความถี่ที่ได้จากตัวอย่าง หรือเรียกว่าความถี่ที่สังเกตได้ (observed frequency) แทนด้วยสัญลักษณ์ O_i กับความถี่ตามทฤษฎี หรือเรียกว่าความถี่คาดหวัง (expected frequency) แทนด้วยสัญลักษณ์ E_i ว่ามีความแตกต่างกันมากเกินไปหรือไม่

สมมติว่าข้อมูลที่ศึกษามี k กลุ่ม เลือกตัวอย่างมาจำนวน n จะได้ความถี่ที่สังเกตได้ O_i และ ความถี่คาดหวัง E_i ดังนี้

ตัวแปร	1	2	3	….	k	รวม
ความถี่ที่สังเกตได้	O_1	O_2	O_3	….	O_k	n
ความถี่คาดหวัง	E_1	E_2	E_3	….	E_k	n

หมายเหตุ 1. การหาความถี่คาดหวัง E_i ต้องทราบสัดส่วนที่ตั้งไว้ในสมมติฐานแล้วคูณกับผลรวมความถี่ทั้งหมด นั่นคือ $E_i = np_I$ หรือ $E_i = np_{io}$

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

ตัวสถิติทดสอบ

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

โดยที่ χ^2 แทนตัวสถิติทดสอบไคสแควร์ที่องศาแห่งความเป็นอิสระ k - 1

k แทนจำนวนกลุ่มข้อมูล

ค่าวิกฤต χ^2 ที่องศาแห่งความเป็นอิสระ k - 1

หมายเหตุ 1. ถ้าความถี่ที่ได้จากการสังเกตใกล้เคียงกับความที่คาดหวัง χ^2 จะมีค่าน้อย

2. ถ้าความถี่ที่ได้จากการสังเกตแตกต่างกับความที่คาดหวังมาก χ^2 จะมีค่ามาก

3. การทดสอบ χ^2 จะมีประสิทธิภาพ n ควรมีขนาดมากพอที่ทำให้ค่าคาดหวังในแต่ละช่องมีค่าไม่ต่ำกว่า 5

4. เพื่อสะดวกในการคำนวณสามารถใช้ $\chi^2 = \Sigma \frac{O^2}{E} - n$ ได้

ข้อตกลงเบื้องต้น

ในการทดสอบไคสแควร์มีข้อตกลงเบื้องต้น ดังนี้

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

ประเภทของการทดสอบไคสแควร์

การทดสอบไคสแควร์มีจุดประสงค์ของการทดสอบตามลักษณะของข้อมูลจำแนกทางเดียว และข้อมูลจำแนกสองทาง

1. การทดสอบไคสแควร์กับข้อมูลจำแนกทางเดียว

การทดสอบไคสแควร์กับข้อมูลจำแนกทางเดียว เรียกว่าการทดสอบภาวะสารูปสนิทดี (Goodness of fit test) แบ่งออกเป็น 2 ลักษณะ ดังนี้

1.1 การทดสอบสัดส่วนของประชากร k กลุ่ม จุดประสงค์ต้องการทดสอบสัดส่วนของความถี่ในแต่ละกลุ่มว่าเท่ากันหรือไม่ หรือทดสอบสัดส่วนของความถี่ในแต่ละกลุ่มว่าเท่ากับค่าคงที่หรือไม่ เช่นสัดส่วนของนักเรียนที่ขาดเรียนในแต่ละวันเท่ากันหรือไม่ สัดส่วนของนักเรียนที่ชอบเรียนในแต่ละรายวิชาเท่ากันหรือไม่ สัดส่วนของลูกค้าที่ใช้โทรศัพท์ยี่ห้อ Samsung:i-phone: Nokia เท่ากับ 0.6:0.3:0.2 หรือไม่ ซึ่งถ้าจะทดสอบความเท่ากันของสัดส่วนของประชากร 2 กลุ่ม ใช้ตัวสถิติทดสอบ Z แต่การทดสอบความเท่ากันของสัดส่วนของประชากรมากกว่า 2 กลุ่ม ใช้ตัวสถิติทดสอบ χ^2 ในการทดสอบ

กำหนด p_i แทนสัดส่วนของประชากรกลุ่มที่ i เมื่อ $i = 1,2,3, \dots , k$

สมมติฐานเชิงสถิติ

$$H_0 : p_1 = p_2 = p_3 = \dots = p_k = \frac{1}{k}$$

$$H_1 : p_I \neq p_j \text{ อย่าง 1 คู่ เมื่อ } i \neq j$$

หรือถ้ามีค่าสัดส่วนที่คาดไว้ (p_0) สมมติฐานเชิงสถิติเป็นดังนี้

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, p_3 = p_{30}, \dots p_k = p_{k0} \text{ หรือ}$$

$$H_0 : p_1 : p_2 : p_3 : \dots p_k = p_{10} : p_{20} : p_{30} : \dots p_{k0}$$

$$H_1 : p_I \neq p_{i0} \text{ อย่างน้อย 1 คู่ เมื่อ } i = 1, 2 \dots, k$$

ตัวสถิติทดสอบ

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

เมื่อ O_i แทนความถี่ที่ได้จากการสังเกต

E_i แทนความถี่คาดหวัง หรือตามทฤษฎี และ $E_i = np_i$

k แทนจำนวนกลุ่ม

n แทนจำนวนตัวอย่าง

ค่าวิกฤต $\chi^2_{1-\alpha, k-1}$

ตัวอย่าง 7.3 ห้างสรรพสินค้า B ได้ผลิตขนมเค้กยี่ห้อ B ออกจำหน่ายตามสาขาต่าง ๆ ของตน โดยจะขายในราคาที่ถูกกว่ายี่ห้ออื่น ๆ ที่มีชื่อเสียงอีก 4 ยี่ห้อ คือ C, D, E และ F ที่ทางห้างรับมาขาย ทางห้างต้องการทราบว่าสัดส่วนของลูกค้าที่ชอบขนมเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบขนมเค้กยี่ห้อ C, D, E และ F หรือไม่ จึงสุ่มลูกค้ามา 100 คน ให้ชิมขนมเค้กทั้ง 5 ยี่ห้อ แล้วให้ลูกค้าบอกว่าชอบยี่ห้อใดมากที่สุด ได้ผลดังนี้

ยี่ห้อ	B	C	D	E	F
จำนวนลูกค้า	17	27	22	15	19

จงทดสอบสิ่งที่ทางห้างต้องการที่ระดับนัยสำคัญ 0.05

วิธีทำ กำหนด p_i แทนสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ i เมื่อ i = B, C, D, E, F

สมมติฐานเชิงสถิติ

$$H_0 : p_B = p_C = p_D = p_E = p_F = \frac{1}{5}$$

$$H_1 : p_i \neq p_j \neq \frac{1}{5} \text{ อย่างน้อย 1 คู่ เมื่อ } i \neq j$$

ยี่ห้อ	**จำนวนลูกค้า (O_i)**	**$E_i=np_i$**	**$O_i - E_i$**	**$(O_i - E_i)^2$**	$\frac{(O_i - E_i)^2}{E_i}$
B	17	20	-3	9	.45
C	27	20	7	49	2.45
D	22	20	2	4	.2
E	15	20	-5	25	1.25
F	19	20	-1	1	.05
รวม	100	100			4.4

ค่าตัวสถิติทดสอบ

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

$$= 4.4$$

ค่าวิกฤติ $\chi^2_{1-\alpha,k-1} = \chi^2_{0.95,4} = 9.49$

เนื่องจากค่าสถิติทดสอบ χ^2 = 4.4 มากกว่า $X^2_{0.95,4}$ = 9.49 อยู่ในบริเวณยอมรับ H_0 หมายความว่าสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้ออื่น ๆ อีก 4 ยี่ห้อ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 7.4 โดยปกติทั่วไปการกระจายของเลือดกลุ่ม O, A, B และ AB มีค่าร้อยละ 44.8, 39.7, 11.3 และ 4.2 ตามลำดับ ถ้าเลือกตัวอย่างผู้หญิงมา 124 คน พบว่ามีเลือดกลุ่ม O, A, B และ AB เท่ากับ 43, 66, 11 และ 4 คน ตามลำดับ อยากทราบว่าผู้หญิงกลุ่มนี้มีการกระจายของกลุ่มเลือดเป็นไปตามปกติทั่วไปหรือไม่ ที่ระดับนัยสำคัญ 0.05 ตารางข้อมูลเป็นดังนี้

กลุ่มเลือด	จำนวนผู้หญิง (O_i)	p_i	$E_i = np_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	43	0.448	55.6	2.86
A	66	0.397	49.2	5.74
B	11	0.113	14.0	0.64
B	4	0.042	5.2	0.28
รวม	124	1.00	124.0	9.52

วิธีทำ สมมติฐานเชิงสถิติ

H_o : กลุ่มเลือด 0:A:B:AB = 44.8 : 39.7 : 11.3 : 4.2

H_1 : กลุ่มเลือด 0:A:B:AB $\neq$ 44.8 : 39.7 : 11.3 : 4.2

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$
$$= \frac{(43-55.6)^2}{55.6} + \frac{(66-49.2)^2}{49.2} + \frac{(11-14.0)^2}{14.0} = \frac{(4-5.2)^2}{5.2}$$
$$= 2.86 + 5.47 + 0.64 + 0.28$$
$$= 9.52$$

ค่าวิกฤติ $\chi^2_{1-\alpha,k-1} = \chi^2_{0.95,3} = 7.18$

เนื่องจากค่าสถิติทดสอบ χ^2 = 9.52 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่าผู้หญิงในชุมชนแห่งนี้มีการกระจายของกลุ่มเลือดไม่เป็นตามปกติที่ระดับนัยสำคัญ 0.05

1.2 การทดสอบการแจกแจงของประชากร ในเรื่องการทดสอบสมมติฐานนั้นตัวสถิติทดสอบที่ใช้ขึ้นอยู่กับการแจกแจงของข้อมูล หากเราทราบว่าข้อมูลมีการแจกแจงลักษณะใดจะทำให้สามารถเลือกใช้ตัวสถิติทดสอบได้ถูกต้องตามข้อจำกัดของวิธีการทดสอบนั้น ๆ ซึ่งการทดสอบ χ^2 เป็นวิธีการหนึ่งที่จะช่วยทำให้ทราบว่าข้อมูลชุดนั้นมีการแจกแจงแบบใด โดยที่

สมมติฐานเชิงสถิติ

H_O : ข้อมูลมีการแจกแจงแบบที่ต้องการ

H_1 : ข้อมูลไม่มีการแจกแจงแบบที่ต้องการ

ตัวสถิติทดสอบ

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

เมื่อ O_i แทนความถี่ที่ได้จากการสังเกต

E_i แทนความถี่คาดหวัง หรือตามทฤษฎี และ $E_i = np_i$

k แทนจำนวนกลุ่ม

ค่าวิกฤต $\chi^2_{1-\alpha,k-1-m}$

เมื่อ m แทนจำนวนพารามิเตอร์ที่ประมาณ (จาก H_0)

ตัวอย่าง 7.5 ในการทดสอบการกระจายของกรดยูริค ของคนไข้ 250 คน ว่ามีการแจกแจงแบบปกติ ด้วยค่าเฉลี่ย 5.74 และ ส่วนเบี่ยงเบนมาตรฐาน 2.01 หรือไม่ ที่ระดับนัยสำคัญ 0.05 ตารางข้อมูลเป็นดังนี้

กรดยูริค	จำนวนคนไข้
< 1	1
1 - 1.99	5
2 - 2.99	15
3 - 3.99	24
4 - 4.99	43
5 - 5.99	50
6 - 6.99	45
7 - 7.99	30
8 - 8.99	29
9 - 9.99	10
≥ 10	5

วิธีทำ กำหนด X เป็นตัวแปรสุ่มแทนปริมาณกรดยูริค

สมมติฐานเชิงสถิติ

H_0 : X มีการแจกแจง N(5.74 , 4.04)

H_1 : X ไม่มีการแจกแจง N(5.74,4.04)

กรดยูริค	ขอบเขตที่แท้จริง	$z = \frac{(x-\mu)}{\sigma}$	P_i	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
$<$ 1	< 0.995	< - 2.36	0.0091	1	2.275	
1 - 1.99	0.995 - 1.995	(-2.36) - (-1.86)	0.0223	5	5.575	0.44
2 - 2.99	1.995 - 2.995	(-1.86) - (-1.37)	0.0539	15	13.475	0.17
3 - 3.99	2.995 - 3.995	(-1.37) - (-0.87)	0.1069	24	26.725	0.28
4 - 4.99	3.995 - 4.995	(-0.87) - (-0.37)	0.1635	43	40.875	0.11
5 - 5.99	4.995 - 5.995	(-0.37) - (0.13)	0.1960	50	49.000	0.02
6 - 6.99	5.995 - 6.995	(0.13) - (0.62)	0.1807	45	45.175	0.006
7 - 7.99	6.995 - 7.995	(0.62) - (1.12)	0.1362	30	35.050	0.48
8 - 8.99	7.995 - 8.995	(1.12) - (1.62)	0.788	22	19.700	0.29
9 - 9.99	8.995 - 9.995	(1.62) - (2.12)	0.356	10	8.900	0.26
$\geq$ 10	> 9.995	> 2.12	0.170	5	4.250	
รวม			1000	250	250	2.03

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= 2.03$$

ค่าวิกฤต $\chi^2_{1-\alpha,k-1-m} = \chi^2_{0.95,8} = 15.5$ (k=9 และ m=0)

เนื่องจากค่าสถิติทดสอบ χ^2 = 2.03 มีค่าน้อยกว่า $\chi^2_{0.95,8}$=15.5 จึงไม่สามารถปฏิเสธสมมติฐาน H_o หมายความว่าการกระจายของกรดยูริคนี้มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 5.44 และส่วนเบี่ยงเบนมาตรฐาน 2.01 ที่ระดับนัยสำคัญ 0.5

หมายเหตุ 1. ในกรณีที่มีความถี่คาดหวัง E_i บางค่าต่ำกว่า 5 อาจจะรวมความถี่ของกลุ่มที่มีค่า E น้อยเข้ากับความถี่ของกลุ่มที่อยู่ติดกัน หรือรวมกลุ่มที่มีลักษณะใกล้เคียงกันเป็นกลุ่มเดียวกัน เพื่อให้ E มีค่ามากกว่า 5 ซึ่งจะทำให้จำนวนกลุ่มลดลง (k ลดลง) จึงมีผลทำให้ df ลดลงด้วย

2. ในกรณีที่ระดับของข้อมูลมี 2 ระดับ (k = 2) องศาเท่าความเป็นอิสระของการทดสอบจะเหลือเพียง 1 จะมีผลทำให้ค่า χ^2 มีจำนวนสูงกว่าที่ควรจะเป็น Frank Yate เสนอให้ปรับค่า χ^2 เป็น ดังนี้

$$\chi^2 = \sum_{i=1}^{2} \frac{(|\ O_i - E_i\ | - .05)^2}{Ei}$$

แต่ถ้าขนาดตัวอย่าง $n \geq 50$ ก็ไม่จำเป็นตัวปรับค่า χ^2

ตัวอย่าง 7.6 ผู้ผลิตรายใหม่รายหนึ่งเชื่อว่าลูกค้าจะชอบนมสดยี่ห้อ A ของเขา ร้อยละ 50 ส่วนอีกร้อยละ 50 ชอบยี่ห้อ B จึงเลือกตัวอย่างลูกค้ามา 45 คน ให้ชิมนมสดยี่ห้อต่าง ๆ ปรากฏว่ามีลูกค้าชอบนมสดยี่ห้อ A 15 คน อยากทราบว่าสิ่งที่ผู้ผลิตรายนี้เชื่อจริงหรือไม่ที่ระดับนัยสำคัญ 0.10

นมสดยี่ห้อ	A	B	รวม
ความถี่ Ei	15	30	45
ความถี่ Oi	22.5	22.5	45

วิธีทำ กำหนด p_A แทนสัดส่วนของลูกค้าที่ชอบนมสดยี่ห้อ A

p_B แทนสัดส่วนของลูกค้าที่ชอบนมสดยี่ห้อ B

สมมติฐานเชิงสถิติ

$$H_O : p_A = p_B = 0.5$$

$$H_1 : p_A \neq p_B = 0.5$$

ตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{2} \frac{(|\ O_i - E_i\ | - .05)^2}{E_i}$$

$$= \frac{(|15 - 22.5| - 0.5)^2}{22.5} + \frac{(|30 - 22.5| - 0.5)^2}{22.5}$$

$$= 2.17 + 2.17$$

$$= 4.34$$

ค่าวิกฤต $\chi^2_{1-\alpha,k-1} = \chi^2_{0.90,1} = 2.71$

เนื่องจากค่าตัวสถิติทดสอบ χ^2 = 4.34 อยู่ในบริเวณปฏิเสธ H_O หมายความว่าสิ่งที่ผู้ผลิตเชื่อว่าลูกค้าชอบนมสดยี่ห้อ A ของเขา ร้อยละ 50 และอีกร้อยละ 50 ชอบยี่ห้อ B นั้น ไม่เป็นความมจริง ที่ระดับนัยสำคัญ 0.10

2. การทดสอบไคสแควร์กับข้อมูลจำแนกสองทาง

การทดสอบไคสแควร์กับข้อมูลจำแนกสองทาง เรียกว่าการทดสอบความเป็นอิสระ (testing of independence) เป็นวิธีการศึกษาความเป็นอิสระของตัวแปรสองตัว หรือความสัมพันธ์ของตัวแปรสองตัวว่าเกี่ยวข้องกันหรือไม่ เช่นสนใจศึกษาความสัมพันธ์ของรายได้กับอายุ ความสัมพันธ์ของประเภทสินค้าที่ขายได้กับรายรับ ความเป็นอิสระของน้ำหนักเด็กแรกเกิดกับปริมาณการสูบบุหรี่ของแม่ขณะตั้งครรภ์ หรือความสัมพันธ์ของยี่ห้อรถยนต์ที่กับอาชีพ เป็นต้น ซึ่งข้อมูลจำแนกสองทางมีลักษณะเป็น ดังนี้

2.1 ลักษณะข้อมูลจำแนกสองทาง ในกรณีที่ข้อมูลเรื่องใดเรื่องหนึ่งถูกจำแนกโดยตัวแปร 2 ตัวแปร ซึ่งโดยทั่วไปข้อมูลที่นำมาทดสอบจะเป็นข้อมูลจำแนก 2 ทาง โดยตัวแปรที่ 1 จะแบ่งเป็น r กลุ่ม (row) และ ตัวแปรที่ 2 จะแบ่งเป็น c กลุ่ม (column) เราจะเรียกตารางนี้ว่า ตารางการณ์จรขนาด r×c (r×c contingency table) โดยข้อมูลที่วิเคราะห์อยู่ในรูปแบบตารางดังนี้

ตัวแปรที่ 1	ตัวแปรที่ 2					รวม
	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
รวม	c_1	c_2	c_3	…	c_c	$\sum r_i=\sum c_j= n$

เมื่อ 0_{ij} แทนความถี่ของตัวแปรที่ 1 กลุ่ม i และตัวแปรที่ 2 กลุ่ม j ; i = 1,…,r, j = 1,…,c

r_i แทนความถี่รวมของตัวแปรที่ 1 กลุ่ม i ; I = 1,…,r

c_j แทนความถี่รวมของตัวแปรที่ 2 กลุ่ม j ; j = 1,…c

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

หมายเหตุ ถ้าตัวแปรที่ 1 มี 2 กลุ่ม และตัวแปรที่ 2 มี 2 กลุ่ม เรียกตารางการณ์จรขนาด 2×2

สมมติฐานเชิงสถิติ

H_0 : ตัวแปรที่ 1 เป็นอิสระกับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 ไม่เป็นอิสระกับตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่มีความสัมพันธ์กับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีความสัมพันธ์กับตัวแปรที่ 2

หรือ

H_o : ตัวแปรที่ 1 ไม่มีผลต่อตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีผลต่อตัวแปรที่ 2

ตัวสถิติทดสอบ

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

โดยที่ χ^2 ~ ไคสแควร์ที่องศาแห่งความเป็นอิสระ (r-1)(c-1)

O_{ij} แทนความถี่ที่ได้จากการสังเกตใน cell(i,j)

E_{ij} แทนความถี่คาดหวังใน cell(i,j)

หมายเหตุ การหา E_{ij} นั้น ต้องหาค่าความน่าจะเป็นที่จะเกิดเหตุการณ์ในแต่ละแถวในสดมภ์ที่ j นั้น คือถ้า H_O เป็นจริงแล้วจะมีค่าเท่ากับ $\frac{c_i}{n}$ และ จำนวนความถี่ที่ได้จากการสังเกตใน cell(i,j) ควรจะมีค่าใกล้เคียงกับขนาดของตัวอย่างชุดที่ i (r_I) คูณกับ $\frac{c_i}{n}$ ดังนั้น

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

เมื่อ r_i แทนความถี่รวมของตัวแปรที่ 1 กลุ่ม i ; i = 1,…,r

c_j แทนความถี่รวมของตัวแปรที่ 2 กลุ่ม j ; j = 1,…c

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

ตัวอย่าง 7.7 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

วิธีทำ สมมติฐานเชิงสถิติ

H_O : 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	$E_{11} = \frac{500 \times 466}{1000} = 233$	$E_{12} = \frac{500 \times 534}{1000} = 267$	500
รวม	466	534	1,000

ตัวสถิติทดสอบ

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

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

$$= 145.07$$

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

เนื่องจากค่าสถิติทดสอบ $\chi^2 =$ 145.07 อยู่ในบริเวณปฏิเสธ H_O หมายความว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 7.8 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	12	110	190
ตั้งแต่ 10,000 ขึ้นไป	48	2	10	60
รวม	116	14	120	250

อยากทราบว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นอยู่กับรายได้หรือไม่ที่ระดับนัยสำคัญ 0.05

วิธีทำ สมมติฐานเชิงสถิติ

H_O : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ไม่ขึ้นกับรายได้

H_1 : ความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นกับรายได้

ความถี่คาดหวัง

$$E_{11} = \frac{190 \times 166}{250} = 88.16 \qquad E_{12} = \frac{190 \times 14}{250} = 10.64 \qquad E_{13} = \frac{190 \times 120}{250} = 91.20$$

$$E_{21} = \frac{60 \times 166}{250} = 27.84 \qquad E_{22} = \frac{60 \times 14}{250} = 3.36 \qquad E_{23} = \frac{60 \times 120}{250} = 28.80$$

ตัวสถิติทดสอบ

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \\ &= \frac{(68-88.16)^2}{88.16} + \frac{(12-10.64)^2}{10.64} + \frac{(110-91.20)^2}{91.20} + \ldots + \frac{(10-28.80)^2}{28.80} \\ &= 36.08\end{aligned}$$

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

เนื่องจากค่าสถิติทดสอบ $\chi^2 =$ 36.08 อยู่ในบริเวณปฏิเสธ H_O หมายความว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นกับรายได้ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 7.9 สถาบันเงินกู้ต้องการศึกษาความสัมพันธ์ระหว่างประเภทของลูกหนี้กับอาชีพเพื่อประกอบการพิจารณาให้เงินกู้ จึงทำการสุ่มตัวอย่างลูกหนี้มาจำนวน 1000 ราย แล้วจำแนกตามประเภทลูกหนี้กับอาชีพ ได้ผลดังแสดงในตาราง

ประเภทลูกหนี้	อาชีพ				รวม
	การเกษตร	รับราชการ	ธุรกิจ	รับจ้าง	
ไม่ขาดการชำระหนี้	126 (107.7)	362 (347.2)	129 (146.6)	78 (93.1)	695
ขาดการชำระหนี้	29 (47.3)	138 (152.5)	82 (64.4)	56 (40.9)	305
รวม	155	500	211	134	1000

จากข้อมูลข้างต้น เราสรุปได้หรือไม่ว่าประเภทลูกหนี้ไม่เป็นอิสระกับอาชีพที่ระดับนัยสำคัญ 0.01 (ตัวเลขในวงเล็บคือความถี่คาดหวัง)

วิธีทำ สมมติฐานเชิงสถิติ

H_0 : ประเภทลูกหนี้เป็นอิสระกับอาชีพ

H_1 : ประเภทลูกหนี้ไม่เป็นอิสระกับอาชีพ

ตัวสถิติทดสอบ

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

$$= \frac{(126-107.7)^2}{107.7} + \frac{(362-347.2)^2}{347.2} + \frac{(129-146.6)^2}{146.6} + \ldots + \frac{(56-40.9)^2}{40.9}$$

$$= 27.14$$

ค่าวิกฤต $\chi^2_{1-\alpha,(r-1)(c-1)} = \chi^2_{0.99,3} = 11.3$

เนื่องจากค่าสถิติทดสอบ $\chi^2 = 27.14$ อยู่ในบริเวณปฏิเสธ H_O หมายความว่าประเภทลูกหนี้ไม่เป็นอิสระกับอาชีพ ที่ระดับนัยสำคัญ 0.05

หมายเหตุ จากตัวอย่างข้างต้นจะเห็นว่าการคำนวณค่าสถิติทดสอบ χ^2 ค่อนข้างยุ่งยากเมื่อตารางการณ์จรมีขนาดใหญ่ขึ้น ดังนั้นเพื่อความสะดวกสามารถคำนวณค่า χ^2 ได้จาก

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

2.2 การทดสอบความเป็นอิสระกรณีตารางการณ์จรขนาด 2×2 ในกรณีที่ตัวแปรที่ 1 และ 2 แบ่งออกเป็นตัวแปรละ 2 กลุ่ม ความถี่นั้นจะอยู่ในรูปตารางการณ์จรขนาด 2×2 ดังนี้

ตัวแปรที่ 1	ตัวแปรที่ 2 1	 2	รวม
1	0_{11}	0_{12}	r_1
2	0_{21}	0_{22}	r_2
รวม	c_1	c_2	n

เมื่อตารางการณ์จร มีขนาด 2×2 ตัวสถิติทดสอบ χ^2 ใช้หลักการเช่นเดิม หรืออาจคำนวณหาค่าตัวสถิติทดสอบ χ^2 ได้สะดวก และง่ายขึ้นโดยใช้สูตร ดังนี้

$$\chi^2 = \frac{n(O_{11}O_{22} - O_{12}O_{21})^2}{r_1 r_2 c_1 c_2}$$

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

หมายเหตุ ข้อจำกัดของการทดสอบ χ^2 สำหรับตารางการณ์จร 2×2

1. $E_{ij} \geq 5$ ทุก ๆ ค่า i,j เมื่อ i = 1,2,….r และ j = 1,…,c
2. เนื่องจากการแจกแจง χ^2 เป็นการแจกแจงของตัวแปรเชิงสุ่มแบบต่อเนื่อง แต่ความถี่ที่ได้มักจะเป็นตัวแปรเชิงสุ่มแบบไม่ต่อเนื่อง อย่างไรก็ดีการแจกแจงแบบ χ^2 เป็นการแจกแจงโดยประมาณที่ดีของค่า χ^2 ที่คำนวณได้ ถ้า df > 1 ในกรณีที่ df = 1 เช่นตารางขนาด 2×2 เราจำเป็นต้องแก้ไขโดยใช้ Yates correction for continuity ดังนี้

$$\chi^2_{\text{corrected}} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(\left|O_{ij} - E_{ij}\right| - 0.5)^2}{E_{ij}}$$

หรือ อาจคำนวณหาโดยใช้สูตรอย่างง่าย ได้ดังนี้

$$\chi^2 = \frac{n(\left|O_{11}O_{22} - O_{12}O_{21}\right| - \frac{1}{2}n)^2}{r_1 r_2\ c_1 c_2}$$

โดยที่ $\chi^2 \sim \chi^2$ df = (r-1) (c-1) ถ้า n $\geq$ 50 ไม่ต้องปรับ χ^2

ตัวอย่าง 7.10 จากตัวอย่าง 7.7 จงคำนวณค่าสถิติทดสอบสำหรับตารางการณ์จรขนาด 2×2

วิธีทำ ตัวสถิติทดสอบ

$$\begin{aligned}\chi^2 &= \frac{n(O_{11}O_{22} - O_{12}O_{21})^2}{r_1 r_2 c_1 c_2} \\ &= \frac{1000\left((328)(362) - (172)(138)\right)^2}{(500)(500)(466)(534)} \\ &= \frac{1000(118736 - 23736)^2}{6.2211\times 10^{10}} \\ &= 145.07\end{aligned}$$

สรุปการทดสอบ χ^2 ของข้อมูลจำแนกทางเดียว และสองทาง

การทดสอบ χ^2

- การทดสอบสมมติฐานสำหรับข้อมูลจำแนกทางเดียว
  - การทดสอบภาวะสารูปสนิทดี (Goodness of fit test)
    - การทดสอบลักษณะต่าง ๆ ของประชากรว่าเป็นไปตามที่คาดไว้หรือไม่
      - การทดสอบความแตกต่างระหว่างสัดส่วนของประชากร k กลุ่ม ($k \geq 3$)
        - $Ho : p_1 = p_2 = p_3 = \ldots = p_k = \frac{1}{k}$
        - $H_1 : p_I \neq p_j$ อย่าง 1 คู่ เมื่อ $i \neq j$
      - การทดสอบสัดส่วนประชากร k กลุ่มว่าเท่ากับที่คาดหรือไม่
        - $Ho : p_1 = p_{10}, p_2 = p_{20}, p_3 = p_{30}, \ldots p_k = p_{k0}$ หรือ
        - $Ho : p_1 : p_2 : p_3 : \ldots p_k = p_{10} : p_{20} : p_{30} : \ldots p_{k0}$
        - $H_1 : p_I \neq p_{i0}$ อย่างน้อย 1 คู่ เมื่อ $i = 1, 2 \ldots, k$
    - การทดสอบการแจกแจงของประชากรว่าเป็นไปตามที่คาดไว้หรือไม่
- การทดสอบสมมติฐานสำหรับข้อมูลจำแนกสองทาง
  - การทดสอบความเป็นอิสระ (Testing of Independence)
    - ตารางการณ์จร ขนาด 2×2
    - ตารางการณ์จร ขนาด r×c

สัมประสิทธิ์ตารางการณ์จร

การทดสอบความเป็นอิสระสำหรับตารางการณ์จรเป็นวิธีการวิเคราะห์ข้อมูลเพื่อตรวจสอบว่าข้อมูลมีความสัมพันธ์กันหรือไม่ ซึ่งในกรณีที่ปฏิเสธสมมติฐาน H_0 ที่ว่าตัวแปรทั้ง 2 เป็นอิสระกัน หมายความว่าตัวแปรทั้ง 2 เกี่ยวข้องกัน หรือสัมพันธ์กัน เราอาจสนใจต่อไปว่าตัวแปรที่สอง สัมพันธ์กันในระดับใด นั้นคือเราจะหาระดับความสัมพันธ์ (degree of dependence) การหาระดับความสัมพันธ์มีหลายวิธี เช่นสัมประสิทธิ์ตารางการณ์จรของเพียรสัน (Pearson's contingency coefficient) สัมประสิทธิ์ตารางการณ์จรของคราเมอร์(Cramer’s ν contingency coefficient) เป็นต้น ในเอกสารฉบับนี้จะหาระดับความสัมพันธ์จากตารางการณ์จรได้โดยใช้สัมประสิทธิ์ตารางการณ์จรของเพียรสัน (Pearson's contingency coefficient) เขียนแทนด้วยสัญลักษณ์ C ได้จาก

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

โดยที่ χ^2 แทนค่าสถิติทดสอบที่คำนวณได้

n แทนจำนวนข้อมูล

ค่าสัมประสิทธิ์ของตารางการณ์จรมีคุณสมบัติ ดังนี้

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

ข้อสังเกต ในทางปฏิบัติค่าต่ำสุดของ C เป็น 0 และค่าสูงสุดของ C ไม่ถึง 1 ขึ้นอยู่กับขนาดของตารางการณ์จร (r กับ c) เราสามารถประมาณค่าสูงสุดของ c ได้จาก

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

โดยที่ q แทนค่า r หรือ c ก็ได้ แล้วแต่ว่าค่าใดมีค่าน้อยกว่า

ตัวอย่าง 7.11 จากตัวอย่าง 7.8 จงหาสัมประสิทธิ์ตารางการณ์จรของเพียรสัน

วิธีทำ จาก

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}}$$
$$= \sqrt{\frac{36.08}{250+36.08}}$$
$$= 0.355$$

และ

$$C_{\max} = \sqrt{\frac{q-1}{q}}$$
$$= \sqrt{\frac{2-1}{2}}$$
$$= 0.707$$

หมายความว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์มีความสัมพันธ์กับรายได้ที่ระดับ 0.355

สรุปท้ายบท

การทดสอบไคแควร์เป็นวิธีการในส่วนของสถิติอ้างอิงที่ใช้ความถี่ของข้อมูลในการวิเคราะห์ โดยใช้หลักการเปรียบเทียบความถี่ที่จากการสังเกตกับความถี่คาดหวังว่าแตกต่างกันเพียงใด ดังนั้นวิธีการนี้จึงเหมาะสมกับข้อมูลเชิงคุณภาพหรือข้อมูลเชิงกลุ่มมากกว่าข้อมูลเชิงปริมาณ เพราะถ้าใช้กับข้อมูลเชิงปริมาณจะต้องจัดข้อมูลเชิงปริมาณให้เป็นกลุ่มเสียก่อน ซึ่งจะสูญเสียรายละเอียดของข้อมูล ประเภทและวัตถุประสงค์ของการทดสอบไคสแควร์ขึ้นอยู่กับความถี่ในตารางแจกแจงความถี่ว่าเป็นตารางแจกแจงความถี่ทางเดียว หรือสองทาง ดังนั้นควรระมัดระวังในการเลือกประเภทการทดสอบไคสแควร์ให้เหมาะสมกับตารางแจกแจงความถี่และวัตถุประสงค์

แบบฝึกหัดท้ายบท

1. เพื่อยืนยันความเชื่อของผู้ผลิตครีมบำรุงผิวยี่ห้อ A ว่าผู้ที่ใช้ครีมบำรุงผิวยี่ห้อ A เป็นผู้ที่มีรายได้สูง (มากกว่า 30,000 บาทต่อเดือน) มีสัดส่วนเป็น 2 เท่าของผู้ใช้ครีมบำรุงผิวยี่ห้อ A ที่มีรายได้ปานกลาง (10,000 –30,000 บาทต่อเดือน) และที่มีรายได้ต่ำ (ต่ำกว่า 10,000 บาทต่อเดือน) จึงเลือกตัวอย่างสตรีที่ใช้ครีมบำรุงผิวยี่ห้อ A มา 200 คน แล้วสอบถามถึงระดับรายได้ว่าอยู่ในช่วงใดได้ข้อมูลดังตาราง จงทดสอบความเชื่อของผู้ผลิตครีมบำรุงผิวยี่ห้อ A ที่ระดับนัยสำคัญ 0.10

ช่วงรายได้ (บาท/เดือน)	จำนวน
ต่ำ : ต่ำกว่า 10,000	48
ปานกลาง : 10,000-30,000	69
สูง : มากกว่า 30,000	83
รวม	200

2. ผู้จัดการโรงงานผลิตตะปูแห่งหนึ่งสุ่มตัวอย่างตะปู 100 ตัว ที่ผลิตได้ในช่วงระยะเวลาหนึ่ง นำมาสร้างตารางแจกแจงความยาวของตะปูได้ ดังนี้

ความยาวของตะปู (เซ็นติเมตร)	จำนวนตะปู
0.5 - 1.5	13
1.5 - 2.5	20
2.5 - 3.5	35
3.5 - 4.5	18
4.5 - 5.5	14

2.1. จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความยาวตะปู

2.2. จงทดสอบว่าความยาวของตะปูที่ผลิตจากโรงงานนี้ทั้งหมด มีการแจกแจงปกติหรือไม่ ที่ระดับนัยสำคัญ 0.05

3. โรงงานแห่งหนึ่งต้องการทดสอบสมมติฐานว่าคุณภาพของผลผลิตในวันหนึ่ง ๆ จะขึ้นกับช่วงเวลาที่ผลิต จากการเก็บข้อมูลสินค้าที่ผลิตทั้งหมด 3,000 ชิ้นในช่วงเวลาต่างๆ พบว่า

ช่วงเวลา 08.00 - 11.00 น. ผลผลิตสินค้าดี 968 ชิ้น เสีย 32 ชิ้น

ช่วงเวลา 13.00 - 16.00 น. ผลผลิตสินค้าดี 972 ชิ้น เสีย 28 ชิ้น

ช่วงเวลา 17.00 - 20.00 น. ผลผลิตสินค้าดี 953 ชิ้น เสีย 47 ชิ้น

จงทดสอบสมมติฐานของโรงงาน ที่ระดับนัยสำคัญ 0.05

4. จากการนำเสนอข้อมูลของหนังสือพิมพ์ฉบับหนึ่ง มีข้อความว่า “ผู้ใช้แรงงานที่ไปทำงานต่างประเทศนั้นมาจากภาคต่าง ๆ คือ เหนือ กลาง อีสาน และใต้ จำนวนแรงงานจากภาคเหนือมีเป็นสองเท่าของภาคกลาง แรงงานจากภาคอีสานมีเป็นสามเท่าของภาคใต้ ส่วนแรงงานจากภาคกลางและภาคใต้มีจำนวนเท่ากัน” การสำรวจผู้ใช้แรงงานในปี พ.ศ. 2538 จำนวน 558 ราย พบว่าผู้ใช้แรงงานมาจากภาคเหนือ 12 คน ภาคกลาง 91 คน ภาคอีสาน 235 คน ที่เหลือมาจากภาคใต้ จงทดสอบสมมุติฐานและสรุปผลว่าเป็นไปตามการนำเสนอข้อมูลของหนังสือพิมพ์นั้นหรือไม่ ที่ระดับนัยสำคัญ 0.05
5. ในการศึกษาเยาวชน ปี 2554 เพื่อดูความสัมพันธ์ของสภาพการอยู่อาศัยของเยาวชนกับลักษณะการติดยาเสพติด พบว่ามีเยาวชนที่ติดยาเสพติดจำนวน 150 คน และไม่ติดยาเสพติด จำนวน 220 คน นำเสนอข้อมูลเป็นร้อยละดังตาราง

	ลักษณะการติดยาเสพติด	
สภาพการอยู่อาศัย	ติดยาเสพติด (%)	ไม่ติดยาเสพติด (%)
อยู่กับบิดามารดา	70	75
อยู่กับญาติพี่น้อง	12	20
อยู่บ้านเช่า/ที่พัก	18	5
รวม	100	100

จงวิเคราะห์และสรุปผลการศึกษานี้ โดยใช้ระดับนัยสำคัญ 0.01

6. ตารางแสดงประเภทลูกค้า และความเสี่ยงของหนี้ ในการกู้ยืมของสถาบันแห่งหนึ่งเป็นดังนี้

	ระดับความเสี่ยง			
ประเภทลูกค้า	ต่ำ	ปานกลาง	สูง	รวม
ส่วนบุคคล	2250	4000	1200	7450
ธุรกิจ	750	1200	600	2550
รวม	3000	5200	1800	10000

อยากทราบว่าระดับความเสี่ยงของลูกค้าขึ้นอยู่กับประเภทของลูกค้าหรือไม่ที่ระดับนัยสำคัญ 0.10

เอกสารอ้างอิง

กัลยา วานิชย์บัญชา.(2551). ***หลักสถิติ.*** พิมพ์ครั้งที่ 9. กรุงเทพ ฯ: ธรรมสาร.

วัชราภรณ์ สุริยาภิวัฒน์. (2552). ***สถิติเบื้องต้นเพื่อธุรกิจ.*** กรุงเทพ ฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สรชัย พิศาลบุตร. (2547). ***สถิติพอเพียง.*** กรุงเทพ ฯ: จูนพับลิชซิ่ง.

____. (2551). ***สถิติธุรกิจ.*** พิมพ์ครั้งที่ 3. กรุงเทพ ฯ: วิทยพัฒน์.

____. (2551). ***การวิจัยทางธุรกิจ.*** กรุงเทพ ฯ: วิทยพัฒน์.

MARILYN K. PELOSI & THERESA M. SANDIFER. (2002). ***Doing Statistics for Business with Excel.*** 2 nd Edition. New York:John WILEY&SONS INC.