การทดสอบด้วยค่าสถิติไคว์สแควร์(Chi-square test)

ประโยชน์ในเชิงสถิติ: ใช้ในการทดสอบว่าปัจจัยทั้งสองปัจจัยมีความสัมพันธ์กันหรือไม่ หรือใช้ในการทดสอบgoodness-of-fit (ในเอกสารวิชาการนี้มุ่งเน้นแสดงให้เห็นการทดสอบว่าปัจจัยทั้งสองมีความสัมพันธ์กันหรือไม่เท่านั้น)

ตารางสถิติที่ใช้ประกอบ: ตารางค่า χ^2 (***ไม่จำเป็นต้องใช้ตารางหากวิเคราะห์โดยโปรแกรมสำเร็จรูป***)

ลักษณะของการทดสอบทางสถิติ: เป็นการทดสอบแบบone-tailed และ right-tailed

ข้อสมมติฐานที่ใช้ทดสอบทางสถิติ: H_o : ปัจจัยทั้งสองมีความเป็นอิสระต่อกันและกัน(หรือ ปัจจัยที่หนึ่งไม่มีความสัมพันธ์กับปัจจัยที่สอง)

H_a : ปัจจัยทั้งสองไม่มีความเป็นอิสระต่อกันและกัน (หรือ ปัจจัยที่หนึ่งมีความสัมพันธ์กับปัจจัยที่สอง)

ประโยชน์ในการประยุกต์ใช้งาน : ใช้ในการทดสอบข้อมูลที่มีโครงสร้างเป็นตารางที่เรียกว่า contingency table

รูปแบบ: ถ้าปัจจัยทั้งสองมีความเป็นอิสระต่อกันและกัน สามารถคำนวณหาค่าความถี่ทางทฤษฎีหรือที่คาดว่าจะเป็น แล้วนำไปเปรียบเทียบกับค่าความถี่ที่สังเกตเห็นได้ โดยคำนวณหาค่าสถิติไคว์สแควร์ซึ่งมีสูตรดังนี้

$$\chi^2 = \sum (o_i - e_i)^2 / e_i$$

โดยที่ χ^2 = ค่าสถิติไคว์สแควร์

o_i = ค่าความถี่ที่สังเกตเห็นได้ (observed frequency)

e_i =ค่าความถี่ทางทฤษฎีหรือคาดว่าจะเป็น (expected frequency)

หลังจากนั้นเปรียบเทียบค่า χ^2 ที่คำนวณได้กับค่าวิกฤติ χ^2 ที่ degree of freedom=(จำนวนแถว-1)X(จำนวนสดมภ์-1)

ตัวอย่างที่ 1:ในการสำรวจความคิดเห็นของลูกค้าที่ใช้บริการที่สาขาธนาคารแห่งหนึ่ง ปรากฏว่าลูกค้าที่เป็นสุภาพบุรุษและสตรีต่างให้คะแนนแตกต่างกัน ปรากฏผลดังตารางต่อไปนี้

	คุณภาพการบริการ			
เพศ	สูงกว่ามาตรฐาน	เท่ากับมาตรฐาน	ต่ำกว่ามาตรฐาน	รวม
หญิง	7(O1)	24(O2)	28(O3)	59 (T1)
ชาย	8(O4)	26(O5)	7(O6)	41(T2)
รวม	15(T3)	50(T4)	35(T5)	100(T6)

ตารางข้างต้นแสดงค่าความถี่ที่สังเกตเห็นได้ ส่วนตารางค่าความถี่ที่คาดว่าจะเป็นสามารถสร้างได้ ทั้งนี้ต้องมีข้อสมมติฐานว่าปัจจัยทั้งสอง (เพศและคุณภาพการบริการ)มีความเป็นอิสระต่อกันและกัน

หากปัจจัยทั้งสองประการข้างต้นมีความเป็นอิสระต่อกันและกันจริง เราสามารถหาค่าความถี่ในแต่ละcell ได้ดังนี้

	คุณภาพการบริการ			
เพศ	สูงกว่ามาตรฐาน	เท่ากับมาตรฐาน	ต่ำกว่ามาตรฐาน	รวม
หญิง	T6XT3XT1/(T6XT6)	T6XT4XT1/(T6XT6)	T6XT5XT1/(T6XT6)	T1
ชาย	T6XT3XT2/(T6XT6)	T6XT4XT2/(T6XT6)	T6XT5XT2/(T6XT6)	T2
รวม	T3	T4	T5	T6

ซึ่งจะได้ตารางแสดงค่าความถี่ที่คาดว่าจะเป็นดังนี้

	คุณภาพการบริการ			
เพศ	สูงกว่ามาตรฐาน	เท่ากับมาตรฐาน	ต่ำกว่ามาตรฐาน	รวม
หญิง	8.85	29.5	20.65	59
ชาย	6.15	20.5	14.35	41
รวม	15	50	35	

ดังนั้นค่าสถิติไคว์สแควร์ $\chi^2 = (7-8.85)^2/8.85 + (24-29.5)^2/29.5 + \ldots\ldots\ldots\ldots + (7-14.35)^2/14.35$

$= 9.824997$

ถ้ากำหนดว่า α =0.05 และที่ degree of freedom=2(จำนวนแถว-1คูณกับจำนวนสดมภ์ -1) ค่าวิกฤติไคว์สแควร์เท่ากับ 5.991 แสดงว่าเพศและความคิดเห็นไม่มีความเป็นอิสระต่อกันและกัน หรือ ความคิดเห็นมีความสัมพันธ์กับเพศของกลุ่มตัวอย่าง

ตัวอย่างที่ 2: ธนาคารแห่งหนึ่งต้องการศึกษาว่า สถานภาพของการทำงานของภรรยาในช่วงที่ลูกค้าผู้เป็นสามีทำเรื่องขอกู้เงินจากธนาคารจะมีผลต่อโอกาสที่หนี้เงินกู้จะเป็นหนี้ผิดนัดในภายหลัง ตารางต่อไปนี้แสดงข้อมูลที่เก็บรวบรวมข้อมูลได้

	สถานภาพของภรรยาในช่วงขอกู้		
ประเภทของหนี้เงินกู้	ทำงาน	ไม่ได้ทำงาน	รวม
ผิดนัด	8	15	23
ไม่ได้ผิดนัด	75	60	135
รวม	83	75	158

ซึ่งถ้าปัจจัยทั้งสองมีความเป็นอิสระซึ่งกันและกัน เราสามารถสร้างตารางแสดงความถี่ที่คาดว่าจะเป็นได้ดังนี้

	สถานภาพของภรรยาในช่วงขอกู้		
ประเภทของหนี้เงินกู้	ทำงาน	ไม่ได้ทำงาน	รวม
ผิดนัด	12.08	10.92	23
ไม่ได้ผิดนัด	70.92	64.08	135
รวม	83	75	158

ดังนั้นค่าสถิติไคว์สแควร์ $\chi^2 = (8-12.08)^2/12.08 + (15-10.92)^2/10.92 + (75-70.92)^2/70.92 + (60-64.08)^2/64.08$

$= 3.400755$

ถ้ากำหนดว่า $\alpha = 0.05$ และที่ degree of freedom=1(จำนวนแถว-1คูณกับจำนวนสดมภ์ -1) ค่าวิกฤติไคว์สแควร์เท่ากับ 3.841 แสดงว่าโอกาสที่หนี้จะผิดนัดหรือไม่นั้นไม่ขึ้นอยู่กับสถานภาพการทำงานของภรรยาในช่วงขอกู้ หรือ โอกาสที่หนี้จะผิดนัดไม่มีความสัมพันธ์กับสถานภาพการทำงานของภรรยาในช่วงขอกู้

ข้อควรระวังในการใช้เทคนิคสถิตินี้ : ถ้าค่าความถี่ที่สังเกตเห็นได้ในแต่ละcell มีค่าน้อย จะมีผลทำให้ค่าสถิติไคว์สแควร์ที่คำนวณได้มีค่าสูง ดังนั้นในทางปฏิบัติ นักวิจัยต้องพยายามให้ค่าความถี่ที่คำนวณได้ในแต่ละเซลล์มีค่ามากกว่า 5 ถ้าค่าความถี่น้อยกว่า 5 นักวิจัยอาจจำเป็นต้องยุบรวมแถวหรือสดมภ์ที่ใกล้เคียงกัน จนในที่สุดค่าความถี่ที่สังเกตเห็นได้ในแต่ละเซลล์มีค่าไม่ต่ำกว่า 5

ตัวอย่างที่ 3: สมมติค่าความถี่ที่คาดว่าจะเป็นมีปรากฏแสดงได้ดังตารางต่อไปนี้

	ปัจจัยที่ 1		
ปัจจัยที่ 2	ระดับ 1	ระดับ2	ระดับ3
ระดับ1	1	7	8
ระดับ2	4	8	9
ระดับ3	4	6	8
ระดับ4	3	5	5

จะต้องปรับให้เป็น

	ปัจจัยที่ 1		
ปัจจัยที่ 2	ระดับ 1	ระดับ2	ระดับ3
ระดับ1-2	5	15	17
ระดับ3-4	7	11	13

Contribution this issue: ดร.ดนัย ปัตตพงศ์*

* ผู้แต่ง/เจ้าของ Statistics Talks(70+ issues) เอกสารวิชาการด้านการวิจัยและสถิติประยุกต์ที่นำเสนอเครื่องมือทางสถิติเพื่อการวิจัยในระดับปริญญาโท เอก ตลอดจนตัวอย่างของการนำไปประยุกต์ใช้ในชีวิตจริง

* ผู้พัฒนา FINALYSIS... a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์(บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯได้เห็นตัวเลขก่อนโครงการเกิดหลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง (กำหนด DEBUT 1 เมษายน 2569)

* ผู้แต่งหนังสือ" การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน " ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม