

บทที่ 7

การวิเคราะห์ความสัมพันธ์สำหรับตาราง 3 ทาง

1. ความสัมพันธ์ของตัวแปร 2 ตัว เมื่อควบคุมตัวแปรอื่น ๆ

ตัวแปรร่วม (Confounding variable) คือ ตัวแปรอื่น ๆ ที่อาจมีผลต่อความสัมพันธ์ระหว่างตัวแปร 2 ตัวที่สนใจศึกษา ดังนั้นในการหาความสัมพันธ์ระหว่างตัวแปรทั้ง 2 ตัวนั้นจึงควรมีการควบคุมตัวแปรร่วม มิฉะนั้นอาจจะทำให้ได้ผลสรุปที่ไม่ถูกต้อง ตัวอย่างเช่น การศึกษาผลของการสูบบุหรี่ในอดีตว่ามีความสัมพันธ์กับการเป็นโรคมะเร็งปอดหรือไม่ การทำการศึกษา ณ จุดเวลาใดเวลาหนึ่ง (cross – section study) ทำได้โดยการเปรียบเทียบอัตราส่วนของผู้ที่เป็นมะเร็งปอดระหว่างกลุ่มผู้ที่ไม่สูบบุหรี่ที่มีคู่สมรสสูบบุหรี่กับกลุ่มผู้ที่ไม่สูบบุหรี่ที่มีคู่สมรสไม่สูบบุหรี่ เราควรจะควบคุมตัวแปรอื่น ๆ เช่น อายุ ฐานะทางเศรษฐกิจ หรือปัจจัยอื่น ๆ ที่อาจมีผลต่อทั้งคนที่มีคู่สมรสสูบบุหรี่และไม่สูบบุหรี่ และคนที่เป็นและไม่เป็นโรคมะเร็งปอด

1.1 การหาความสัมพันธ์ระหว่างตัวแปร X และ Y เมื่อควบคุมตัวแปร Z เพียงตัวเดียว

ซึ่งตัวแปรทั้ง 3 ตัวนี้เป็นแบบแบ่งประเภท ข้อมูลเป็นจำนวนความถี่อยู่ในตารางการแจกแจงความถี่แบบ 3 ทาง (three – way contingency tables) การพิจารณาเฉพาะความสัมพันธ์ของ X และ Y เมื่อกำหนดระดับของ Z เป็นการแสดงอิทธิพลของ X ต่อ Y ขณะที่ควบคุม Z เป็นการขจัดอิทธิพลของ Z โดยทำให้เป็นค่าคงที่ ดังนั้นจึงเป็นการพิจารณาตาราง 2 ทาง ของตาราง 3 ทาง เรียกว่าตารางบางส่วน (partial tables)

สำหรับตาราง 2 ทาง ที่ได้จากการรวมตารางบางส่วน เรียกว่า ตารางมาจิ้นัล (marginal table) ของ X – Y และจำนวนความถี่ของแต่ละช่องในตารางมาจิ้นัล คือผลรวมจากช่องที่อยู่ตำแหน่งเดียวกันในตารางบางส่วน ตารางมาจิ้นัลไม่มีข้อมูลเกี่ยวกับตัวแปร Z

ความสัมพันธ์ในตารางบางส่วน เรียกว่า ความสัมพันธ์แบบมีเงื่อนไข (Conditional association) เพราะมาจากอิทธิพลของ X ต่อ Y โดยมีเงื่อนไขการกำหนด Z ที่บางระดับ

ตัวอย่างการศึกษาเกี่ยวกับอิทธิพลของคุณลักษณะทางเชื้อชาติของจำเลยต่อการตัดสินประหารชีวิตนักโทษแต่ละคนที่มีความผิดฐานฆ่าคนตาย กลุ่มตัวอย่างคือ จำเลยทั้งหลายที่ต้องคดีเกี่ยวกับฆาตกรรม ในฟลอริดาระหว่างปี ค.ศ. 1976 ถึง 1987 ตัวแปรตามคือ ผลการตัดสินประหารชีวิต (Y) ที่มี 2 ระดับ คือ 1 = ไซ้ และ 2 = ไม่ไซ้ ตัวแปรอิสระคือ เชื้อชาติของจำเลย (X) ที่มี 2 ระดับคือ 1 = ขาว และ 2 = ดำ และตัวแปรร่วมคือ เชื้อชาติของเหยื่อ (Z) ที่มี 2 ระดับ คือ 1 = ขาว และ 2 = ดำ ได้ข้อมูลในตารางบางส่วน 2×2 ในตาราง 3 ทาง ดังนี้

ตารางที่ 7.1 จำนวนนับของผลการตัดสินประหารชีวิตจำเลยที่มีเชื้อชาติแตกต่างกันและเชื้อชาติของเหยื่อแตกต่างกัน

เชื้อชาติของเหยื่อ (Z)	เชื้อชาติของจำเลย (X)	ผลการตัดสินประหารชีวิต (Y)		รวม	เปอร์เซ็นต์ ไซ้
		ไซ้	ไม่ไซ้		
ขาว	ขาว	53	414	467	11.3
	ดำ	11	37		48
ดำ	ขาว	0	16	16	0.0
	ดำ	4	139	143	2.8
รวม	ขาว	53	430		11.0
	ดำ	15	176		7.9

แหล่งที่มา : M.L Radelet and G.L. Pierce , Florida law Rev. 43 : 1 - 34 (1991) อ้างถึงใน A. Agresti , An Introduction to Categorical Data Analysis p. 54 (1996)

การหาความสัมพันธ์ระหว่างตัวแปรเชิงชาติของจำเลยและตัวแปรผลการตัดสินใจประหารชีวิต เมื่อควบคุมตัวแปรร่วมคือเชื้อชาติของเหยื่อ สามารถทำได้โดยการหาค่า **odds ratios** สำหรับเหยื่อผิวขาว และ **odds ratios** ของเหยื่อผิวดำ

การคำนวณค่า **odds ratios** สำหรับเหยื่อผิวขาวและผิวดำ ทำได้ดังนี้

การคำนวณค่า **odds ratio** เพื่อหาความสัมพันธ์ระหว่างเชื้อชาติของจำเลยกับผลการตัดสินใจประหารชีวิตสำหรับเหยื่อผิวขาว คือ

$$\theta_{XY(1)}^{\wedge} = \frac{(53)(37)}{(414)(11)} = 0.43$$

เรียกว่า **conditional odds ratio** ระหว่าง **X** และ **Y** ที่ระดับของ **Z = 1** หมายความว่าจำเลยผิวขาวได้รับการตัดสินประหารชีวิต เทียบกับจำเลยผิวดำคือเท่ากับ 43% ของจำเลยผิวดำ

การคำนวณค่า **odds ratio** เพื่อหาความสัมพันธ์ระหว่างเชื้อชาติของจำเลยกับผลการตัดสินใจประหารชีวิตสำหรับเหยื่อผิวดำ คือ

$$\theta_{XY(2)}^{\wedge} = \frac{(0)(139)}{(16)(4)} = 0.0$$

เรียกว่า **conditional odds ratio** ระหว่าง **X** และ **Y** ที่ระดับของ **Z = 2** ได้ค่า **odds ratio** เท่ากับ 0 เนื่องจากไม่มีจำเลยผิวขาวที่ถูกตัดสินประหารชีวิต

การคำนวณค่า **odds ratio** ถ้าไม่คำนึงถึงเชื้อชาติของเหยื่อ จะได้ความสัมพันธ์ระหว่างเชื้อชาติของจำเลยกับผลการตัดสินใจประหารชีวิตสำหรับเหยื่อทั้ง 2 ผิว คือ

$$\theta_{XY}^{\wedge} = \frac{(53)(176)}{(430)(15)} = 1.45$$

เรียกว่า **marginal odds ratio** หมายความว่าจำเลยผิวขาวถูกตัดสินประหารชีวิตเทียบกับจำเลยผิวดำ คือสูงกว่าจำเลยผิวดำ 45% จะเห็นว่าค่ากลับกันกับค่า **odds ratio** ที่คิดแยก

ตามเชื่อกาติของเหยื่อ เหตุผลคือ มีผลของตัวแปรร่วม (Z) กับตัวแปรที่สนใจศึกษา ความสัมพันธ์ทั้ง 2 ตัวนั้น (X - Y)

ถ้า X และ Y เป็นอิสระกันที่แต่ละระดับของ Z เรียกว่าเป็นอิสระแบบมีเงื่อนไข คือ เมื่อกำหนด Z แล้ว conditional odds ratios ระหว่าง X และ Y จะเท่ากับ 1 แต่ marginal odds ratio อาจแตกต่างจาก 1 ได้ ตัวอย่างเช่น ให้ X คือ ทริทเมนต์ยา (A, B) ให้ Y คือ ผลลัพธ์ (สำเร็จ, ล้มเหลว) และให้ Z คือ คลินิก (1, 2) ได้ข้อมูลดังตารางที่ 7.2

ตารางที่ 7.2 ข้อมูลจำนวนนับของผลลัพธ์ของคนไข้ที่ได้รับทริทเมนต์ยาแตกต่างกันในคลินิก 2 แห่ง

คลินิก (Z)	ทริทเมนต์ (X)	ผลลัพธ์ (Y)	
		สำเร็จ	ล้มเหลว
1	A	18	12
	B	12	8
2	A	2	8
	B	8	32
ผลรวม	A	20	20
	B	20	40

การคำนวณค่า odds ratio เพื่อหาความสัมพันธ์ของ X และ Y ที่แต่ละระดับของ Z คือ

$$\theta_{XY(1)}^{\hat{}} = \frac{(18)(8)}{(12)(12)} = 1.0$$

$$\theta_{XY(2)}^{\hat{}} = \frac{(2)(32)}{(8)(8)} = 1.0$$

หมายความว่า X และ Y เป็นอิสระกันที่แต่ละระดับของ Z

การคำนวณค่า odds ratio เพื่อหาความสัมพันธ์ของ X และ Y โดยไม่คำนึงถึงระดับของ Z คือ

$$\theta_{XY}^{\wedge} = \frac{(20)(40)}{(20)(20)} = 2.0$$

หมายความว่า X และ Y ไม่เป็นอิสระกัน คือ odds ของสำเร็จสำหรับทริทเมนต์ A เป็น 2 เท่าของทริทเมนต์ B ถ้าไม่คำนึงถึงระดับของ Z ซึ่งผลที่ได้นี้ไม่เหมือนกับผลที่ได้จาก conditional odds ratio เหตุผลคือ Z มีผลต่อทั้ง X และ Y

1.2 ความสัมพันธ์แบบโฮโมจีเนียส (Homogeneous Association)

ความสัมพันธ์ระหว่าง X และ Y ที่เป็นโฮโมจีเนียสคือ ค่า odds ratio ของ X และ Y จะเท่ากันที่ทุกระดับของ Z นั่นคือ อิทธิพลของ X บน Y จะเหมือนกันที่ทุกระดับของ Z และจะทำให้ค่า odds ratio แบบมีเงื่อนไขของตัวแปรคู่อื่น ๆ มีคุณสมบัติเช่นเดียวกันนี้ด้วย ตัวอย่างเช่น ค่า odds ratio แบบมีเงื่อนไขของ X และ Z จะเหมือนกันที่ทุกระดับของ Y เมื่อเป็นโฮโมจีเนียส หมายถึงไม่มีปฏิสัมพันธ์ระหว่างตัวแปร 2 ตัวนั้นกับตัวแปรตัวที่สาม และถ้าไม่เป็นโฮโมจีเนียส หมายถึง ค่า odds ratio แบบมีเงื่อนไขของตัวแปรคู่ใด ๆ ที่แต่ละระดับของตัวแปรตัวที่สามจะไม่เท่ากัน ตัวอย่างเช่น กำหนดให้ X แทนการสูบบุหรี่ (สูบ , ไม่สูบ) Y แทนการเป็นโรคมะเร็งปอด (เป็น , ไม่เป็น) และ Z แทนอายุ (1 คือ <45 , 2 คือ $45-65$, 3 คือ >65) สมมติว่า $\theta_{XY(1)} = 1.2$, $\theta_{XY(2)} = 2.8$, และ $\theta_{XY(3)} = 6.2$ หมายความว่า การสูบบุหรี่มีอิทธิพลน้อยต่อการเป็นโรคมะเร็งปอดในกลุ่มคนอายุน้อย แต่มีอิทธิพลมากขึ้นในกลุ่มอายุที่มากขึ้น

2. การทดสอบความเป็นอิสระแบบมีเงื่อนไข และการทดสอบความสัมพันธ์แบบโฮโมจีเนียส

ตัวอย่างเช่น การศึกษาความสัมพันธ์ระหว่างการสูบบุหรี่และการเป็นโรคมะเร็งปอดในเมืองต่าง ๆ 8 เมือง ของประเทศจีน กำหนดให้ X แทนการสูบบุหรี่แบ่งเป็น 2 ระดับ คือ สูบ และไม่สูบ ให้ Y แทนการเป็นโรคมะเร็งปอดแบ่งเป็น 2 ระดับคือ เป็น และไม่เป็น และให้ Z แทนตัวแปรร่วมแบ่งเป็นระดับต่าง ๆ คือ เมืองต่าง ๆ ที่ทำการศึกษา กลุ่มตัวอย่างที่นำมาศึกษาในเมือง ต่าง ๆ อาจมีคุณลักษณะแตกต่างกันตามเมืองที่อยู่ เช่น สถานภาพทางเศรษฐกิจและสังคม ซึ่งอาจมีผลต่ออัตราการสูบบุหรี่และอัตราการเป็นโรคมะเร็งปอด ดังนั้นจึงต้องการศึกษาความสัมพันธ์ระหว่าง X และ Y เมื่อควบคุมตัวแปร Z ข้อมูลอยู่ในตารางที่ 7.3

ข้อตกลงเบื้องต้นของการทดสอบซึ่งเกี่ยวข้องกับลักษณะของข้อมูลความถี่ในช่องของตาราง คือ

- (1) มีการแจกแจงแบบบิวซองและเป็นอิสระกัน หรือ
- (2) มีการแจกแจงแบบมัลติโนเมียล โดยมีการกำหนดขนาดของกลุ่มตัวอย่างทั้งหมด หรือ
- (3) มีการแจกแจงแบบมัลติโนเมียล โดยมีการกำหนดขนาดของกลุ่มตัวอย่างที่แต่ละระดับของตัวแปรร่วม และเป็นอิสระกัน หรือ
- (4) มีการแจกแจงแบบไบโนเมียล โดยมีการกำหนดผลรวมทางแถว หรือคอลัมน์ภายในตาราง $X - Y$ ที่แต่ละระดับของตัวแปร Z

ตารางที่ 7.3 ข้อมูลการสูบบุหรี่และการเป็นโรคมะเร็งปอดของชาวจีนในเมืองใหญ่ 8 เมือง กับผล การทดสอบ Cochran-Mantel-Haenszel Test

City	Smoking	Lung Cancer		Odds Ratio	μ_{11k}	Var(n_{11k})
		Yes	No			
Beijing	Smokers	126	100	2.20	113.0	16.9
	Nonsmokers	35	61			
Shanghai	Smokers	908	688	2.14	733.2	179.3
	Nonsmokers	497	807			
Shenyang	Smokers	913	747	2.18	799.3	149.3
	Nonsmokers	336	598			
Nanjing	Smokers	235	172	2.85	203.5	31.1

Harbin	Nonsmokers	58	121	2.32	355.0	57.1
	Smokers	402	308			
Zhengzhou	Nonsmokers	121	215	1.59	169.0	28.3
	Smokers	182	156			
Taiyuan	Nonsmokers	72	98	2.37	53.0	9.0
	Smokers	60	99			
Nanchang	Nonsmokers	11	43	2.00	96.5	11.0
	Smokers	104	89			
	Nonsmokers	21	36			

แหล่งที่มา : Based on data in Z. Liu, Smoking and lung cancer in China, Intern. J. Epidemiol., 21 : 197-201 (1992). อ้างถึงใน A. Agresti, An Introduction to Categorical Data Analysis p.60 (1996)

บันทึกข้อมูลลงในแฟ้มข้อมูล CMH1.sav กำหนดให้ เมืองใหญ่ 8 เมือง แทนด้วยตัวแปร city (1 = Beijing , 2 = Shanghai , 3 = Shenyang , 4 = Nanjing , 5 = Harbin , 6 = Zhengzhou , 7 = Taiyuan , 8 = Nanchang) การสูบบุหรี่ มี 2 ระดับ แทนด้วยตัวแปร smoke (1 = สูบ , 2 = ไม่สูบ) และการเป็นโรคมะเร็งปอด มี 2 ระดับ แทนด้วยตัวแปร cancer (1 = เป็น , 2 = ไม่เป็น) ให้ตัวแปร count แทนจำนวนนับ ได้รูปแบบของข้อมูลดังนี้

ตารางที่ 7.4 รูปแบบของการบันทึกข้อมูลการสูบบุหรี่และการเป็นโรคมะเร็งปอดของชาวจีนแยกตามเมือง

cit	smo	canc	cou	cit	smo	can	cou	cit	smo	can	cou
-----	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----

y	ke	er	nt	y	ke	cer	nt	y	ke	cer	nt
1	1	1	126	4	1	1	235	7	1	1	60
1	1	2	100	4	1	2	172	7	1	2	99
1	2	1	35	4	2	1	58	7	2	1	11
1	2	2	61	4	2	2	121	7	2	2	43
2	1	1	908	5	1	1	402	8	1	1	104
2	1	2	688	5	1	2	308	8	1	2	89
2	2	1	497	5	2	1	121	8	2	1	21
2	2	2	807	5	2	2	215	8	2	2	36
3	1	1	913	6	1	1	182				
3	1	2	747	6	1	2	156				
3	2	1	336	6	2	1	72				
3	2	2	598	6	2	2	98				

2.1 การทดสอบความเป็นอิสระแบบมีเงื่อนไข

สมมติฐานทางสถิติที่ต้องการทดสอบ คือ $H_0 : X$ และ Y เป็นอิสระกัน เมื่อกำหนด Z หมายความว่า odds ratio $\theta_{XY(Z)} = 1$ ที่ทุกระดับของ Z

สถิติทดสอบความเป็นอิสระแบบมีเงื่อนไข คือ สถิติ Cochran - Mantel - Haenszel (CMH) มีสูตรคือ

$$CMH = \frac{\left[\sum_k (n_{11k} - \mu_{11k}) \right]^2}{\sum_k \text{Var}(n_{11k})}$$

โดยที่
$$\mu_{11k} = E(n_{11k}) = \frac{n_{1.k} n_{.1k}}{n_{.k}}$$

และ
$$\text{Var}(n_{11k}) = \frac{n_{1.k} n_{2.k} n_{.1k} n_{.2k}}{n_{.k}^2 (n_{.k} - 1)}$$

สถิติ CMH มีการแจกแจงแบบไคสแคว และจำนวนชั้นอิสระเท่ากับ 1 ค่าสถิติ CMH จะมีค่ามากเมื่อ $(n_{11k} - \mu_{11k})$ เป็นค่าบวกที่ทุกระดับ k หรือเป็นค่าลบที่ทุกระดับ k ค่าสถิตินี้ใช้ได้ดีเมื่อความสัมพันธ์ระหว่าง X และ Y เหมือนกันในแต่ละระดับ k และค่าสถิตินี้ไม่เหมาะสมสำหรับในกรณีที่ความสัมพันธ์ระหว่าง X และ Y แตกต่างกันในแต่ละระดับ k

2.2 การทดสอบความสัมพันธ์แบบโฮโมจีเนียส

สมมติฐานทางสถิติที่ต้องการทดสอบเกี่ยวกับ odds ratio ระหว่าง X และ Y เท่ากันที่ทุกระดับของ Z เขียนเป็นสัญลักษณ์ได้คือ $H_0 : \theta_{XY(1)} = \dots = \theta_{XY(k)}$

สมมติว่า $\theta_{XY(1)} = \dots = \theta_{XY(k)}$ สามารถประมาณค่าตามวิธีของ Mantel-Haenszel ได้ดังนี้

$$\hat{\theta}_{MH} = \frac{\sum_k (n_{11k} n_{22k} / n_{.k})}{\sum_k (n_{12k} n_{21k} / n_{.k})}$$

สถิติทดสอบความสัมพันธ์แบบโฮโมจีเนียส คือ สถิติ Breslow-Day มีสูตรคือ

$$\sum_{ijk} \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

ซึ่งมีการแจกแจงโดยประมาณแบบไคสแควที่มีจำนวนชั้นอิสระ df เท่ากับ $k - 1$ โดยที่ขนาดของตัวอย่างในแต่ละช่องของตารางควรจะใหญ่พอ คือ $\mu_{ijk} \geq 5$ อย่างน้อยที่สุด 80% ของจำนวนช่องทั้งหมดของตาราง

3. การใช้คำสั่ง Crosstabs

วิธีการศึกษาเกี่ยวกับความสัมพันธ์ระหว่างการสูบบุหรี่และการเป็นโรคมะเร็งปอดในเมืองต่าง ๆ 8 เมือง ของประเทศจีน ทำโดยการจับคู่ผู้ที่เป็โรคมะเร็งปอดกับผู้ที่ไม่เป็นโรคมะเร็งปอด แล้วบันทึกข้อมูลของแต่ละตัวอย่างว่าเคยสูบบุหรี่หรือไม่ ข้อมูลที่ได้ในแต่ละเมืองเราถือว่าจำนวนความถี่ในแต่ละคอลัมน์มีการแจกแจงแบบไบนอมียลที่กำหนดจำนวนรวมทั้งหมดทางคอลัมน์ สมมติฐานทางสถิติที่ต้องการทดสอบคือ H_0 : การสูบบุหรี่และการเป็นมะเร็งปอดเป็นอิสระกัน หมายความว่า odds ratio เท่ากับ 1.0 ทุกเมืองที่ทำการศึกษา

ขั้นตอนการใช้คำสั่งในการทดสอบสมมติฐานคือ

1. ไปที่เมนูบาร์ คลิกที่ Data , Weight Cases จะได้นหน้าต่าง Weight Cases คลิกที่ O Weight cases by แล้วคลิกที่ตัวแปร count ให้ย้ายเข้าไปอยู่ในช่อง Frequency Variable: แล้วคลิกที่ปุ่ม OK หน้าต่างนี้จะถูกปิดไป

2. ไปที่เมนูบาร์ คลิกที่ Analyze , Descriptive Statistics , Crosstabs จะได้นหน้าต่าง Crosstabs

3. ในหน้าต่าง Crosstabs คลิกที่ตัวแปร smoke ให้ย้ายเข้าไปอยู่ในช่อง Row(s): แล้วคลิกที่ตัวแปร cancer ให้ย้ายเข้าไปอยู่ในช่อง Column(s): แล้วคลิกที่ตัวแปร city ให้ย้ายเข้าไปอยู่ในช่อง Layer 1 of 1

คลิกที่ปุ่ม Statistics ... จะได้นหน้าต่าง Crosstabs : Statistics

4. ในหน้าต่าง Crosstabs : Statistics

คลิกที่ Risk เพื่อจะได้ค่า odds ratio

คลิกที่ Cochran's and Mantel - Haenszel statistics เพื่อจะได้ค่าสถิติ CMH และในช่อง Test common odds ratio equals พิมพ์เลข 1 ลงในช่องได้เป็น 1 โดยทั่วไปโปรแกรมกำหนดให้อยู่แล้ว

คลิกที่ปุ่ม Continue หน้าต่างนี้จะถูกปิดไป

5. ในหน้าต่าง Crosstabs คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 7.1

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
smoke * cancer * city	8419	100.0%	0	.0%	8419	100.0%

smoke * cancer * city Crosstabulation

Count			cancer		Total
city	smoke		เป็น	ไม่เป็น	
Beijing	smoke	สูบ	126	100	226
		ไม่สูบ	35	61	96
	Total		161	161	322
Shanghai	smoke	สูบ	908	688	1596
		ไม่สูบ	497	807	1304
	Total		1405	1495	2900
Shenyang	smoke	สูบ	913	747	1660
		ไม่สูบ	336	598	934
	Total		1249	1345	2594
Nanjing	smoke	สูบ	235	172	407
		ไม่สูบ	58	121	179
	Total		293	293	586
Harbin	smoke	สูบ	402	308	710
		ไม่สูบ	121	215	336
	Total		523	523	1046
Zhengzhou	smoke	สูบ	182	156	338
		ไม่สูบ	72	98	170
	Total		254	254	508
Taiyuan	smoke	สูบ	60	99	159
		ไม่สูบ	11	43	54
	Total		71	142	213
Nanchang	smoke	สูบ	104	89	193
		ไม่สูบ	21	36	57
	Total		125	125	250

Risk Estimate

city		Value	95% Confidence Interval	
			Lower	Upper
Beijing	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.196	1.343	3.590
	For cohort cancer = เป็น	1.529	1.146	2.041
	For cohort cancer = ไม่เป็น	.696	.564	.860
	N of Valid Cases	322		
Shanghai	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.143	1.846	2.488
	For cohort cancer = เป็น	1.493	1.376	1.619
	For cohort cancer = ไม่เป็น	.697	.649	.748
	N of Valid Cases	2900		
Shenyang	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.175	1.844	2.565
	For cohort cancer = เป็น	1.529	1.389	1.683
	For cohort cancer = ไม่เป็น	.703	.654	.755
	N of Valid Cases	2594		
Nanjing	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.850	1.969	4.125
	For cohort cancer = เป็น	1.782	1.420	2.237
	For cohort cancer = ไม่เป็น	.625	.537	.728
	N of Valid Cases	586		
Harbin	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.319	1.775	3.031
	For cohort cancer = เป็น	1.572	1.345	1.838
	For cohort cancer = ไม่เป็น	.678	.604	.761
	N of Valid Cases	1046		
Zhengzhou	Odds Ratio for smoke (สูบ / ไม่สูบ)	1.588	1.095	2.303
	For cohort cancer = เป็น	1.271	1.040	1.555
	For cohort cancer = ไม่เป็น	.801	.674	.952
	N of Valid Cases	508		
Taiyuan	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.369	1.135	4.945
	For cohort cancer = เป็น	1.852	1.054	3.256
	For cohort cancer = ไม่เป็น	.782	.652	.937
	N of Valid Cases	213		
Nanchang	Odds Ratio for smoke (สูบ / ไม่สูบ)	2.003	1.091	3.680
	For cohort cancer = เป็น	1.463	1.016	2.105
	For cohort cancer = ไม่เป็น	.730	.569	.938
	N of Valid Cases	250		

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	5.200	7	.636
Tarone's	5.200	7	.636

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	280.383	1	.000
Mantel-Haenszel	279.376	1	.000

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			2.174
ln(Estimate)			.777
Std. Error of ln(Estimate)			.047
Asymp. Sig. (2-sided)			.000
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	1.984
		Upper Bound	2.383
	ln(Common Odds Ratio)	Lower Bound	.685
		Upper Bound	.868

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

ภาพที่ 7.1

ผลลัพธ์ที่ได้คือ

1. สรุปจำนวนข้อมูลทั้งหมดที่นำมาวิเคราะห์ คูที่ตาราง Case Processing Summary ได้จำนวน N เท่ากับ 8419 คิดเป็น 100%

2. ตารางการแจกแจงความถี่ของข้อมูลขนาด 2×2 ของตัวแปร smok และ cancer ที่แต่ละระดับของตัวแปร city ซึ่งมีทั้งหมด 8 ระดับ หรือ 8 เมือง ดังนั้นจึงมีตาราง 2×2 ทั้งหมด 8 ตาราง คูในตาราง smoke * cancer * city Crosstabulation

3. ค่าประมาณของค่า odds ratio ของกลุ่มตัวอย่างทั้ง 8 เมือง คูในตาราง Risk Estimate ตัวอย่างเช่น ที่ city = Beijing ได้ค่า odds ratio for smoke (สูบ/ไม่สูบ) เท่ากับ 2.196 และ 95% confidence interval เท่ากับ (1.343 , 3.590)

4. ผลการทดสอบโฮโมจีเนียสของค่า odds ratio คูในตาราง Tests of Homogeneity of the Odds Ratio และผลการทดสอบความเป็นอิสระแบบมีเงื่อนไข คูในตาราง Test of Conditional Independence

ค่าสถิติของการทดสอบความสัมพันธ์แบบโฮโมจีเนียสของตัวแปร smoke และ cancer ที่เมืองต่าง ๆ 8 เมือง สถิติทดสอบคือ Breslow - Day มีค่าเท่ากับ 5.200 มีจำนวนชั้นอิสระ df เท่ากับ 7 มีค่า Asymp. Sig. (2-sided) เท่ากับ .636 ซึ่งมากกว่าระดับนัยสำคัญที่กำหนด ($\alpha = .05$)สรุปได้ว่ายอมรับ H_0 นั่นคือ odds ratio ของการมีปัจจัยเสี่ยง (สูบบุหรี่) ของทั้ง 8 เมือง เท่ากันหมด

ค่าสถิติของการทดสอบความเป็นอิสระแบบมีเงื่อนไขระหว่างตัวแปร smok และ cancer สถิติทดสอบคือ Cochran's Mantel Haenszel อยู่ในช่อง Chi-Square มีค่าเท่ากับ 280.383 มีจำนวนชั้นอิสระ df เท่ากับ 1 มีค่า Asymp. Sig. (2-sided) เท่ากับ .000 ซึ่งน้อยกว่าระดับนัยสำคัญที่กำหนด ($\alpha = .05$) สรุปได้ว่าปฏิเสธ H_0 นั่นคือ การสูบบุหรี่และการเป็นโรคมะเร็งปอดไม่เป็นอิสระกัน

5. ผลการประมาณค่า common odds ratio ตามวิธีของ Mantel - Haenszel คูในตาราง Mantel - Haenszel Common Odds Ratio Estimate ได้ค่าประมาณ common odds ratio เท่ากับ 2.174 หมายความว่า ผู้ที่สูบบุหรี่จะเป็นโรคมะเร็งปอดเป็น 2 เท่าของผู้ที่ไม่สูบบุหรี่ สามารถคิดเป็นค่า ln ของค่าประมาณ คือ $\ln(2.174)$ เท่ากับ .777 ที่มีความคลาดเคลื่อนมาตรฐานเท่ากับ .047

และได้ค่า 95% confidence interval ของค่า common odds ratio เท่ากับ (1.984 , 2.383) และ 95% confidence interval ของค่า $\ln(\text{common odds ratio})$ เท่ากับ (.685 , .868)