ตัวแบบถดถอยลอจิสติกพหุคูณกับการคำนวณค่าอัตราส่วนออดส์ปรับค่า

อรุณ จิรวัฒน์กุล*

การเกิดโรคแต่ละโรคปรกติจะมีปัจจัยเสี่ยงหลายปัจจัย ค่าอัตราส่วนสัมพัทธ์ (relative ratio; RR) หรืออัตราส่วนออดส์ (odds ratio; OR) ที่คำนวณได้จากข้อมูลของแต่ละปัจจัยจะเป็นขนาดความเสี่ยงหยาบ ซึ่งขนาดความเสี่ยงที่คำนวณได้เกิดจากอิทธิพลของปัจจัยดังกล่าวรวมกับอิทธิพลจากปัจจัยเสี่ยงของตัวแปรอื่น ๆ ร่วมอยู่ด้วย. ถ้านักวิจัยต้องการทราบขนาดความเสี่ยงจากปัจจัยใดปัจจัยหนึ่งเพียงปัจจัยเดียว จะต้องคำนวณอัตราส่วนออดส์ที่ปรับแล้ว (adjusted OR) หรืออัตราส่วนสัมพัทธ์ที่ปรับค่า (adjusted RR) ด้วยวิธีวิเคราะห์แบบแบ่งชั้น (stratify analysis) หรือตัวแปรถดถอยลอจิสติกพหุคูณ (multiple logistic regression).

ตัวแบบถดถอยลอจิสติกพหุคูณ

ตัวแบบถดถอยลอจิสติกพหุคูณเป็นการหาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัว ($X_1, X_2, X_3, ..., X_n$) กับตัวแปรตาม Y ที่มีการแจกแจงปรกติ โดยมีตัวแบบดังนี้ $y = a + b_1X_1 + b_2X_2 + b_3X_3$

ในกรณีที่ตัวแปรตาม Y เป็นตัวแปรที่จัดค่าได้ ๒ กลุ่ม (dichotomy) เช่น เป็นโรค / ไม่เป็นโรค จะมีการแจกแจงแบบทวินามไม่ใช่แบบปรกติ. ตัวแบบที่ใช้แสดงความสัมพันธ์ในกรณีที่ตัวแปรตาม Y ที่เป็น ๒ กลุ่ม คือตัวแบบถดถอยลอจิสติกพหุคูณมีตัวแบบดังนี้

$$\text{Pn}\left(\frac{P}{1-P}\right) = a + b_1x_1 + b_2x_2 + ...+b_kx_k$$

โดยที่ P เป็นความน่าจะเป็นของการเกิดโรค ค่า $\frac{P}{1-P}$ คือค่าออดส์จะได้ว่า ℓn(odds) ของการเกิดโรคมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ (ปัจจัยเสี่ยง) ทั้งหลาย.

จากข้อมูลตัวอย่างที่ศึกษานำมาคำนวณค่าสัมประสิทธิ์ของตัวแบบ a, b_1, b_2,...b_k ด้วยวิธีหาค่าประมาณเสมือนสูงสุด (maximum like estimate; MLE) ถ้าผลการทดสอบค่าสัมประสิทธิ์ด้วยสถิติ LR หรือสถิติ Wald มีนัยสำคัญ แสดงว่าตัวแปรดังกล่าวมีอิทธิพลต่อการเปลี่ยนค่าตัวแปรตาม Y. ตัวแปรนั้นจะถูกเลือกให้อยู่ในตัวแบบ.

ดังนั้นการที่ตัวแปรใดถูกเลือกให้อยู่ในตัวแบบจะขึ้นอยู่กับความสัมพันธ์ของตัวแปรปัจจัยกับตัวแปรตาม Y ของข้อมูลในตัวอย่าง ในกรณีที่ขนาดตัวอย่างไม่ใหญ่พอ. เมื่อมีการศึกษาด้วยตัวอย่างชุดใหม่ ตัวแปรที่ถูกเลือกจากข้อมูลตัวอย่างต่างชุดกัน อาจมีปัจจัยเสี่ยงในตัวแบบไม่เหมือนกันทุกตัว.

กรณีที่นักวิจัยต้องการให้ตัวแปรที่มีทฤษฎีสนับสนุนว่ามีความสัมพันธ์ทุกตัวให้อยู่ในตัวแบบก็สามารถทำได้. การเลือกตัวแปรเข้าในตัวแบบด้วยวิธีนี้ อาจมีตัวแปรปัจจัยบางตัวที่ผลการทดสอบค่าสัมประสิทธิ์ไม่มีนัยสำคัญอยู่ในตัวแบบ.

ตัวแปรอิสระในตัวแปรถดถอยลอจิสติกพหุคูณ จะเป็นตัวแปรประเภทใดก็ได้ ไม่ว่าจะเป็นตัวแปรต่อเนื่อง เช่น อายุ, ตัวแปรกลุ่ม เช่น เพศ, หรือตัวแปรอันดับ เช่นระดับความ

*ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

รุนแรงของโรค.

การคำนวณอัตราส่วนออดส์ปรับค่า (adjusted OR)

ในการศึกษาเรื่องการได้รับอุบัติเหตุฯจากการขับขี่รถจักรยานยนต์ พบว่ามีความสัมพันธ์เชิงบวกกับการดื่มสุรา. ถ้านักวิจัยคิดว่าการได้รับอุบัติเหตุน่าจะมีความสัมพันธ์กับกลุ่มอายุของผู้ขับขี่ด้วย และอัตราส่วนออดส์ที่ได้จากการวิเคราะห์เฉพาะการดื่มสุรากับการได้รับอุบัติเหตุฯ จะไม่สามารถแสดงอิทธิพลที่แท้ของการดื่มสุรา เพราะจะมีอิทธิพลของตัวแปรกลุ่มอายุร่วมอยู่ด้วย. กลุ่มอายุอาจเป็น ปัจจัยตัวกวน (confounding factor) หรือปัจจัยตัวปรับแต่งผล (effect modifier factor). ในทางกลับกัน ถ้าจะดูปัจจัยเสี่ยงของกลุ่มอายุ ก็มีอิทธิพลของการดื่มสุราร่วมอยู่ด้วย.

ในการคำนวณอัตราส่วนออดส์ปรับค่า แสดงระดับความสัมพันธ์ของการดื่มสุรากับการได้รับอุบัติเหตุฯ โดยปรับอิทธิพลของกลุ่มอายุ ด้วยตัวแปรแบบถดถอยลอจิสติกพหุคูณที่มีตัวแปรอิสระ ๒ ตัวคือการดื่มสุรา (X_1) และกลุ่มอายุ (X_2) อยู่ในตัวแบบดังนี้

$$\ell n\ (\mathrm{Odds}) = b_0 + b_1X_1 + b_2X_2$$

ในการคำนวณค่าอิทธิพลของการดื่มสุรา เมื่อกำหนดให้กลุ่มอายุเท่ากัน เช่น เมื่อกำหนดให้กลุ่มอายุ (X_2) มีค่าเท่ากับ ๑ สามารถคำนวณค่าออดส์ของการดื่มสุรา ($X_1 = 1$) และ ไม่ดื่มสุรา ($X_1 = 0$) ได้ดังนี้

$$\ell n(Odds_{x_1=1}) = a + 1.b_1 + 1.b_2 = a + b_1 + b_2$$
$$\ell n(Odds_{x_1=0}) = a + 0.b_1 + 1.b_2 = a + b_2$$
$$\ell n(Odds_{x_1=1}) - \ell n(Odds_{x_1 0}) = a + b_1 + b_2 - a - b_2$$
$$\ell n(Odds_{x_1=1}) - \ell n(Odds_{x_1 0}) = b_1$$
$$\ell n\left(\frac{Odds_{x_1=1}}{Odds_{x_1=0}}\right) = b_1$$
$$\ell n(OR) = b_1$$
$$OR = e^{b_1}$$

ค่าสัดส่วนออดส์ของการดื่มสุราที่คำนวณจากตัวแบบนี้จะมีเฉพาะอิทธิพลของการดื่มสุราเพียงอย่างเดียว โดยการปรับอิทธิพลของกลุ่มอายุ (ให้มีค่าเท่ากัน) ออกแล้ว.

ตัวอย่างผลการวิเคราะห์ปัจจัยเสี่ยงของการได้รับอุบัติเหตุจากการขับขี่รถจักรยานยนต์กับการดื่มสุรา และกลุ่มอายุแสดงในตารางที่ ๑ ข้างล่างนี้.

ตารางที่ ๑

ปัจจัยเสี่ยง	การได้รับอุบัติเหตุ		OR
	ไม่เคย	เคย	
กลุ่มอายุ			
≥ ๒๕ ปี	๒๗๙ (๙๑.๒%)	๒๗ (๘.๘%)	๘.๔๒
< ๒๕ ปี	๑๖๒ (๕๕.๑%)	๑๓๒ (๔๔.๙%)	
การดื่มสุรา			
ไม่ดื่ม	๓๖๓ (๘๙.๖%)	๔๒ (๑๐.๔%)	๑๒.๙๖
ดื่ม	๗๘ (๔๐.๐%)	๑๑๗ (๖๐.๐%)	

ตารางที่ ๒ ผลการวิเคราะห์ตัวแบบถดถอยลอจิสติกพหุคูณของการเกิดอุบัติเหตุฯ กับกลุ่มอายุ และการดื่มสุรา

ตัวแปร	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI ของ EXP(B)	
							Lower	Upper
อายุ	๑.๐๑๗	๐.๒๘๕	๑๒.๗๖๙	๑	.๐๐๐	๒.๗๖๕	๑.๕๓๘	๔.๘๒๙
แอลกอฮอล์	๒.๐๐๐	๐.๒๕๘	๖๐.๑๕๓		.๐๐๐	๗.๓๙๑	๔.๔๕๘	๑๒.๒๕๒
ค่าคงที่	-๒.๕๓๑	๐.๒๑๐	๑๔๔.๖๕๓	๑	.๐๐๐	.๐๘๐		

จากตารางที่ ๒ ค่าสัมประสิทธิ์ (b_1) ของการดื่มสุราเท่ากับ ๒.๐๐๐ มีค่าพีของสถิติ Wald <๐.๐๐๐๑ แสดงว่าการดื่มสุรามีความสัมพันธ์กับการได้รับอุบัติเหตุ. ค่า Exp(B) คือ อัตราส่วนออดส์ของการดื่มสุราเท่ากับ ๗.๓๙ มีค่าช่วงความเชื่อมั่นร้อยละ ๙๕ อยู่ระหว่าง ๔.๔๖ กับ ๑๒.๒๕ ซึ่งมีค่ามากกว่า ๑ แสดงว่าการดื่มสุราเป็นปัจจัยเสี่ยงของการเกิดอุบัติเหตุฯ. อัตราส่วนออดส์ = ๗.๓๙ ที่ได้เป็นอัตราส่วนออดส์ที่มีการปรับอิทธิพลของตัวแปรกลุ่มอายุแล้ว (adjusted OR) ซึ่งจะไม่เท่ากับอัตราส่วนออดส์หยาบ (crude OR) = ๑๒.๙๖ จากตารางที่ ๑ ที่ยังไม่มีการปรับอิทธิพลของตัวแปรกลุ่มอายุ.

ตัวอย่างที่นำเสนอนี้คำนวณจากข้อมูลที่ได้จากการศึกษาแบบวิเคราะห์ภาคตัดขวาง ซึ่งตัวอย่างที่สุ่มเข้ามาศึกษาแต่ละตัวมีความเป็นอิสระต่อกัน, ค่าสัมประสิทธิ์ของตัวแบบคำนวณด้วยวิธีตัวแบบถดถอยลอจิสติกไร้เงื่อนไข (unconditional logistic regression). แต่ถ้าข้อมูลที่ได้จากแบบงานวิจัยแบบการศึกษาที่มีตัวควบคุมที่เข้าคู่กัน (match case-control study) ตัวอย่างที่สุ่มมาศึกษาไม่เป็นอิสระต่อกัน, ค่าสัมประสิทธิ์ของตัวแบบจะคำนวณด้วยวิธีตัวแบบถดถอยลอจิสติกมีเงื่อนไข (conditional logistic regression).

สรุป

ตัวแบบถดถอยลอจิสติกพหุคูณ ใช้สำหรับตัวแบบที่ตัวแปรตามที่จัดค่าได้สองกลุ่ม (dichotomy). ส่วนตัวแปรอิสระจะเป็นตัวแปรประเภทใดก็ได้. อัตราส่วนออดส์ที่คำนวณได้จากค่าสัมประสิทธิ์ของตัวแปรแต่ละตัวจะเป็นค่าอัตราส่วนออดส์ที่ปรับอิทธิพลของตัวแปรอื่น ๆ แล้ว (adjusted OR). การสร้างตัวแบบจะต้องพิจารณาแบบงานวิจัยด้วยว่าตัวอย่างที่นำมาศึกษาสุ่มมาอย่างเป็นอิสระต่อกันหรือไม่ และจะต้องเลือกใช้การคำนวณตัวแบบให้สอดคล้องกับความเป็นอิสระของตัวอย่าง.

เอกสารประกอบการเรียบเรียง

๑. Kleinbaum DG, Klein M. Logistic regression. A self-learning text,. New York : Springer; 2002.

๒. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research principles and quantitative methods,. New York : Van Nostrand Reinhold; 1982.