

มุมมองวิจัย

Methodology Corner

การเลือกใช้การวิเคราะห์ข้อมูลด้วย Poisson Regression Model และ Negative Binomial Regression

ชนิษฐา กู้ศรีสกุล วท.ม*, กนิษฐา บุญธรรมเจริญ Ph.D.*, ระพีพงศ์ สุพรรณไชยมาตย์ Ph.D.**

* สำนักงานพัฒนานโยบายสุขภาพระหว่างประเทศ; ** กองระบาดวิทยา กรมควบคุมโรค

ตัวแบบการถดถอยแบบพัวซอง (Poisson regression model) และตัวแบบการถดถอยทวินามเชิงลบ (negative binomial regression) เป็นตัวแบบการวิเคราะห์การถดถอยที่นิยมใช้ในการวิเคราะห์ข้อมูลที่มีลักษณะเป็นจำนวนนับ (discrete count data) เช่น จำนวนผู้ป่วยที่เข้ามารักษาในโรงพยาบาลในแต่ละวัน จำนวนผู้ที่เสียชีวิต เป็นต้น ซึ่งการแจกแจงแบบพัวซอง คือการอธิบายจำนวนครั้งของเหตุการณ์หรือจำนวนสิ่งที่น่าสนใจที่เกิดขึ้นในช่วงเวลาที่กำหนด โดยมีข้อสมมติฐานข้อหนึ่งที่ว่าค่าเฉลี่ย (mean) ของตัวแปรสุ่ม (variables) จะต้องเท่ากับค่าความแปรปรวน (variance) อย่างไรก็ตามข้อมูลจำนวนนับมักเกิดปัญหาคือข้อมูลมีค่าความแปรปรวนสูงกว่าค่าเฉลี่ย (over dispersion) ซึ่งอาจไม่เป็นไปตามข้อสมมติฐานของการใช้ตัวแบบพัวซอง ดังนั้นจึงมีตัวแบบทวินามเชิงลบ (negative binomial regression) ที่แก้ปัญหาค่า over dispersion ดังสมการที่ 1 และ 2 ซึ่งแสดงสมการการแจกแจงแบบพัวซองและแบบทวินามเชิงลบ เห็นได้ว่าการแจกแจงแบบพัวซองมีพารามิเตอร์ตัวเดียวคือ λ ซึ่งหมายถึงค่าเฉลี่ยหรือค่าคาดหวังและค่าความแปรปรวน (ค่าเท่ากัน) ส่วนการกระจายแบบทวินามเชิงลบจะมีพารามิเตอร์ 2 ตัวคือ λ และ α โดย λ หมายถึงค่าเฉลี่ยหรือค่าคาดหวัง ส่วน α หมายถึงค่า over dispersion จากทั้ง 2 สมการกล่าวได้ว่า การกระจายแบบทวินามเชิงลบก็คือการกระจายแบบพัวซองนั่นเอง ต่างกัน

ที่มีพารามิเตอร์ที่แสดงค่า over dispersion ที่แก้ข้อจำกัดตามสมมติฐานของการกระจายแบบพัวซองที่ว่าค่าเฉลี่ยและค่าความแปรปรวนต้องเท่ากัน

Poisson distribution⁽¹⁾

สมการที่ 1

$$\Pr(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} \text{ for } y = 0,1,2, \dots$$

การกระจายแบบพัวซองมีพารามิเตอร์ 1 ตัวคือ λ ซึ่งหมายถึง mean or expected และค่า variance

Negative binomial distribution⁽¹⁾

สมการที่ 2

$$\Pr(Y = y|\lambda, \alpha) = \frac{\Gamma(y+\alpha-1)}{y!\Gamma(\alpha-1)} \left(\frac{\alpha-1}{\alpha-1+\lambda}\right)^{\alpha-1} \left(\frac{\lambda}{\alpha-1+\lambda}\right)^y \text{ for } y = 0,1,2, \dots$$

การกระจายแบบทวินามเชิงลบมีพารามิเตอร์ 2 ตัวคือ λ และ α ซึ่ง λ หมายถึง mean or expected และ α เป็นค่า over dispersion

เกณฑ์พิจารณาการเลือกใช้⁽²⁾

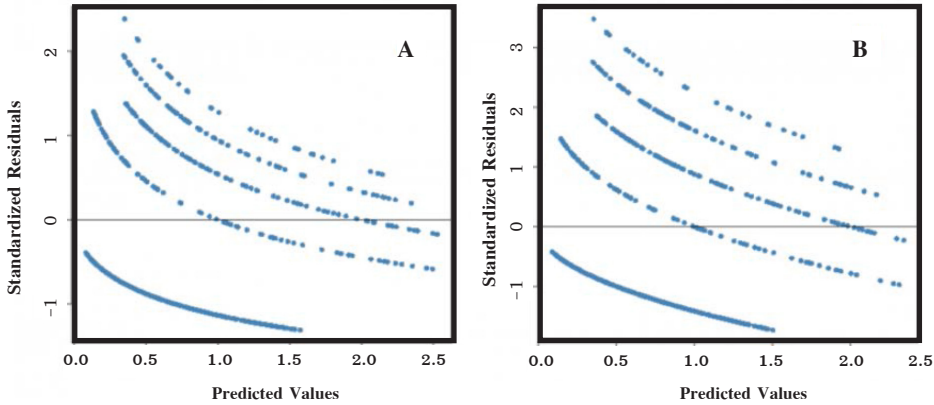
การวิเคราะห์การถดถอยด้วยตัวแบบพัวซองและตัวแบบทวินามเชิงลบอาจมีได้หลายวิธี แต่มี 2 วิธีที่นิยมใช้คือ

1) พิจารณาจากค่า residual โดยสร้างกราฟดูการกระจายของ standardized residuals เทียบกับค่า predicted values จากตัวแบบการถดถอย หาก standardized residuals อยู่ในช่วง -2 ถึง 2 ตัวแบบพัวซองน่าจะเหมาะสม

แต่หากค่าอยู่นอกช่วง -2 ถึง 2 ตัวแบบทวินามเชิงลบอาจ จะเหมาะสมกว่า จากภาพที่ 1 B เห็นได้ว่าค่า standard-

ized residuals มีบางส่วนเกิน 3 ฉะนั้นตัวแบบทวินามเชิง ลบน่าจะเหมาะสมกว่าตัวแบบพัชชอง

ภาพที่ 1 เปรียบเทียบการกระจายของ standardized residuals เทียบกับค่า predicted values (A) standardized residuals อยู่ ในช่วง -2 ถึง 2 (B) standardized residuals อยู่นอกช่วง -2 ถึง 2



2) สร้างตัวแบบพัชชองและตัวแบบทวินามเชิงลบจาก ข้อมูลชุดเดียวกัน แล้วนำมาทดสอบด้วย Likelihood Ratio Test และพิจารณาจากค่า p-value หากค่าต่ำกว่า 0.05 ตัวแบบทวินามเชิงลบน่าจะเหมาะสมกว่าตัวแบบ พัชชอง

เชิงลบจะเหมาะสมกว่าตัวแบบพัชชอง ซึ่งค่า over dispersion อาจพิจารณาจากค่า residual และการทดสอบ ด้วย likelihood ratio test อย่างไรก็ตามการจะเลือกใช้วิธี ไต อาจขึ้นอยู่กับความเหมาะสมตามที่นักวิจัยเห็นสมควร

อย่างไรก็ตามจากเกณฑ์พิจารณาทั้ง 2 วิธีเมื่อสร้างตัว แบบพัชชองหรือตัวแบบทวินามเชิงลบด้วยโปรแกรม ต่างๆ เช่น R⁽³⁾ Stata SAS เป็นต้น ค่าเหล่านี้จะแสดง ผลออกมาพร้อมตัวแบบอยู่แล้วก็สามารถนำค่าที่ได้นั้น มาใช้พิจารณาในการเลือกตัวแบบที่เหมาะสมได้เลย

เอกสารอ้างอิง

1. Fernandez GA, Vatcheva KP. A comparison of statistical methods for modeling count data with an application to hospital length of stay. BMC Med Res Methodol 2022;22(1):211. <https://doi.org/10.1186/s12874-022-01685-8>.
2. Statology. Negative binomial vs. Poisson: how to choose a regression model [Internet]. [cited 2022 Nov 1]. Available from: <https://www.statology.org/negative-binomial-vs-poisson/>
3. Zeileis A, Kleiber C, Jackman S. (2008). Regression Models for Count Data in R. Journal of Statistical Software 2008;27(8):1-25.

สรุป

การเลือกใช้การวิเคราะห์ด้วยตัวแบบการถดถอยแบบ พัชชองหรือตัวแบบการถดถอยแบบทวินามเชิงลบ ประการแรกข้อมูลต้องมีลักษณะเป็นจำนวนนับ ประการ ที่สอง การจะเลือกใช้ตัวแบบพัชชองหรือตัวแบบทวินาม เชิงลบให้พิจารณาจากค่าความแปรปรวน หากข้อมูลมี ลักษณะของ over dispersion การเลือกใช้ตัวแบบทวินาม