

03/0039/32

บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล
บทความ

ข้อพิจารณาการใช้สถิติการถดถอยพหุสำหรับงานวิจัย Considerations in Using Multiple Regression for Research

ธวัชชัย วรพงษ์ศร*

Abstract : Thavatchai Vorapongsathon. 1989. Considerations in using multiple regression for research
Thai J Hlth Resch 3(1) : 55-62

Multiple regression is a statistical method for analysis a casual relationship between several independent variables and a single dependent variable. Whether the results from using multiple regression analysis are valid and reliable depend on the important factors in which the researcher must logically concern. These relevant factors are an appropriate multiple regression model, logical selection of independent variables, the characteristics of data as follow the assumption of a multiple regression model, adequate sample size and awareness of the weakpoints in interpretation of a multiple correlation coefficient.

บทคัดย่อ : ธวัชชัย วรพงษ์ศร. 2532. ข้อพิจารณาการใช้สถิติการถดถอยพหุสำหรับงานวิจัย. วารสารวิจัยวิทยาศาสตร์-
การแพทย์. 3 (1) : 55-62

การวิเคราะห์การถดถอยพหุ เป็นสถิติที่ใช้วิเคราะห์ความสัมพันธ์ในเชิงเหตุและผล ระหว่างตัวแปรอิสระหลายตัว กับตัวแปรตามตัวเดียว ผลการวิเคราะห์การถดถอยพหุ จะถูกต้องและเป็นที่น่าเชื่อถือได้มากน้อยแค่ไหน ขึ้นอยู่กับปัจจัยสำคัญที่ผู้วิจัยจะต้องคำนึงถึง ซึ่งปัจจัยเหล่านี้ได้แก่รูปแบบที่เหมาะสมของสถิติการถดถอยพหุ การเลือกตัวแปรอิสระอย่างมีเหตุมีผล ข้อมูลมีคุณลักษณะเป็นไปตามข้อตกลงของรูปแบบการถดถอยพหุ มีจำนวนตัวอย่างพอเพียง และควรระวังในจุดอ่อนของการแปลผลค่าสัมประสิทธิ์สหสัมพันธ์

* ภาควิชาชีวสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล กรุงเทพมหานคร 10400
Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok 10400.

การถดถอยพหุเป็นเครื่องมือทางสถิติ ที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตามหนึ่งตัว ซึ่งช่วยอธิบายเหตุการณ์หรือสะท้อนภาพความเป็นจริง ให้เข้าใจได้มากกว่าการวิเคราะห์การถดถอยแบบง่าย ซึ่งใช้ตัวแปรอิสระเพียงตัวเดียว ตัวอย่างเช่น การใช้ความแตกต่างของอายุ (X_1) เพียงปัจจัยเดียว ช่วยอธิบายระดับของคอเลสเตอรอล (Y) ย่อมจะไม่ชัดเจนเท่ากับการเพิ่มปัจจัยอื่น ๆ เข้าไปอีกหลายปัจจัย เช่น น้ำหนัก (X_2) เพศ (X_3) การดื่มสุรา (X_4) การสูบบุหรี่ (X_5) การรับประทานอาหารที่มีไขมันมาก (X_6) และการออกกำลังกาย (X_7) เป็นต้น ปัจจัยเหล่านี้ย่อมจะช่วยอธิบายการเปลี่ยนแปลงของระดับคอเลสเตอรอล ได้ชัดเจนมากกว่าการใช้ปัจจัยเพียงตัวใดตัวหนึ่งเท่านั้น ด้วยเหตุนี้การวิเคราะห์การถดถอยพหุจึงเป็นสถิติที่นิยมใช้กันอย่างกว้างขวางทั้งในงานวิจัยแบบสำรวจและงานวิจัยแบบทดลอง

บทความนี้ต้องการเน้นถึงการใช้อนุสัจการถดถอยพหุอย่างถูกต้อง ซึ่งผู้วิจัยจำเป็นต้องตระหนักถึงปัจจัยสำคัญที่เกี่ยวข้องหลายประการด้วยกัน ซึ่งได้แก่ รูปแบบของสถิติการถดถอยพหุ (multiple regression model) การเลือกตัวแปรอิสระ ข้อตกลง (assumptions) ของรูปแบบการถดถอยพหุ จำนวนตัวอย่าง และการแปลผลของค่าสัมประสิทธิ์สหสัมพันธ์ (multiple correlation coefficient : R^2)

รูปแบบของสถิติการถดถอยพหุ

การวิเคราะห์ด้วยสมการถดถอยพหุ ต้องมีการกำหนดรูปแบบของฟังก์ชันแสดงความสัมพันธ์ ระหว่างตัวแปรอิสระและตัวแปรตามในลักษณะของรูปเชิงเส้น (linear) ซึ่งสามารถแยกออกได้เป็นสองลักษณะ คือรูปแบบเชิงเส้นตรง (straight-line model) และรูปแบบเชิงเส้นโค้ง (polynomial or curvilinear model) ทั้งสองรูปแบบนี้จะใช้สมการถดถอยพหุที่แตกต่างกันออกไป ดังนั้นการวิเคราะห์การถดถอยพหุ ผู้วิจัยจึงจำเป็นต้องมีการทดสอบรูปแบบเสียก่อนว่า เป็นไปตามรูปแบบที่คาดหวังไว้หรือไม่ ถ้าเป็นไปตามรูปแบบที่คาดหวังไว้ จึงสามารถดำเนินการทดสอบทางสถิติในขั้นต่อไปได้ แต่ถ้าผลการทดสอบไม่ได้เป็นไปตามรูปแบบที่คาดหวังไว้ ก็จำเป็นต้องเลือกรูปแบบใหม่ หรือทำการปรับข้อมูลใหม่ให้เหมาะสมกับรูปแบบที่คาดหวังไว้ก็ได้ ซึ่งทั้งนี้ขึ้นอยู่กับว่าผู้วิจัยนั้นทราบรายละเอียดเกี่ยวกับการกระจาย และระดับการวัดของข้อมูลนั้น ๆ มากน้อยแค่ไหน

ในทางปฏิบัติมักพบว่า งานวิจัยส่วนใหญ่มีการใช้รูปแบบเชิงเส้นตรงมากกว่ารูปแบบเชิงเส้นโค้งโดยอัตโนมัติ ทั้ง ๆ ที่ไม่ได้มีการทดสอบอย่างถูกต้องตามหลักการทางสถิติเสียก่อน ซึ่งโดยแท้ที่จริงแล้วลักษณะความสัมพันธ์ของข้อมูลในงานวิจัยบางเรื่อง ไม่เหมาะสมกับรูปแบบเชิงเส้นตรง แต่ถ้าใช้รูปแบบเชิงเส้นโค้งจะสามารถอธิบายผลของความสัมพันธ์ ระหว่างตัวแปรได้ชัดเจนมากกว่าด้วยซ้ำ ด้วยเหตุนี้การเลือกรูปแบบการถดถอยพหุ ที่ถูกต้องและเหมาะสม จึงถือเป็นด่านแรกที่สำคัญ ที่สามารถมีอิทธิพลต่อผลการวิเคราะห์ข้อมูลในขั้นสุดท้ายได้

การเลือกตัวแปรอิสระ

การเลือกตัวแปรอิสระในรูปแบบการถดถอยพหุ ผู้วิจัยควรคำนึงหลักที่สำคัญสองประการคือ ประการแรก ตัวแปรอิสระทุกตัวในสมการถดถอยพหุ ต้องมีความสัมพันธ์กับตัวแปรตามในเชิงเหตุและผล (cause-effect) เท่านั้น โดยเฉพาะงานวิจัยที่มีการเก็บข้อมูลในช่วงเวลาเดียว (cross-sectional data) ต้องระวังปัญหานี้ให้มาก เพราะการระบุความสัมพันธ์ในเชิงเหตุและผล ระหว่างตัวแปรอิสระและตัวแปรตามนั้น อย่างน้อยในแง่ของตรรกวิทยา ตัวแปรอิสระย่อมต้องเกิดขึ้นก่อนตัวแปรตามเสมอ จะเกิดขึ้นภายหลังตัวแปรตามไม่ได้ เช่น การสูบบุหรี่ (ตัวแปรอิสระ หรือ ปัจจัยเหตุ) ย่อมเกิดขึ้นก่อนการเป็นโรคมะเร็งในปอด (ตัวแปรตาม หรือ ปัจจัยผล) จะเกิดขึ้นภายหลังการเกิดโรคมะเร็งในปอดแล้วไม่ได้ การระบุความสัมพันธ์ระหว่างตัวแปรเหตุและตัวแปรผล นอกจากจะพิจารณาในแง่ของตรรกวิทยาแล้ว อาจต้องใช้หลักฐานอื่น ๆ มาประกอบ เช่น หลักฐานทางทฤษฎี ทางวิชาการ งานวิจัย หรือ

ข้อคิดเห็นจากผู้เชี่ยวชาญในเรื่องนั้น ๆ มาช่วยสนับสนุนด้วย ก็จะทำให้ลักษณะความสัมพันธ์ในเชิงเหตุและผลมีน้ำหนักน่าเชื่อถือได้มากยิ่งขึ้น

ประการที่สอง ตัวแปรอิสระทุกตัวในสมการถดถอย ควรเป็นตัวแปรที่คิดว่ามีอิทธิพลต่อตัวแปรตามอย่างมาก เพราะในความเป็นจริงแล้ว ตัวแปรที่มีอิทธิพลต่อตัวแปรตามมีอยู่มากมาย ผู้วิจัยจึงต้องคัดเลือกเฉพาะตัวแปรที่มีความสำคัญจริง ๆ มาวิเคราะห์เท่านั้น นอกจากนี้การพิจารณาตัวแปรอิสระทุกตัว ที่คัดเลือกเข้ามาใส่ในสมการถดถอยพหุ ตัวแปรอิสระเหล่านั้นต้องไม่มีความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง หรืออาจมีความสัมพันธ์กันได้บ้างเล็กน้อย ถ้าตัวแปรอิสระมีความสัมพันธ์กันเอง จะเกิดปัญหาที่เรียกว่า multicollinearity ทำให้การประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอย (regression coefficient) เชื่อถือไม่ได้ การเกิดปัญหา multicollinearity ผู้วิจัยสามารถสังเกตได้จากลักษณะต่าง ๆ ทางสถิติที่เกิดขึ้น ในลักษณะเดียวหรือรวมกันหลายลักษณะ ในขณะที่กำลังวิเคราะห์ข้อมูลดังต่อไปนี้

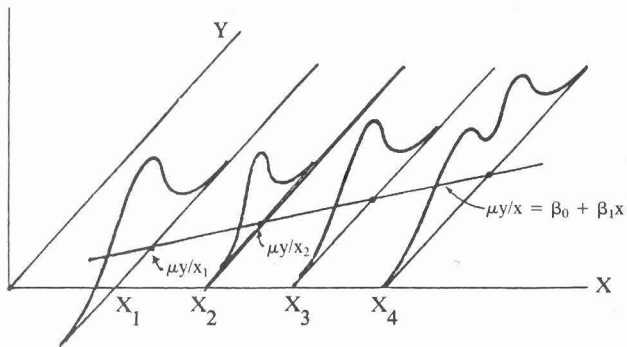
1. การเพิ่มหรือลดตัวแปรอิสระในสมการถดถอยพหุ จะทำให้ค่าสัมประสิทธิ์การถดถอยเกิดการเปลี่ยนแปลงอย่างมาก
2. ค่าผลรวมยกกำลังสอง (sum of square) ที่เพิ่มขึ้น จากการใส่หรือลดตัวแปรอิสระในสมการถดถอยพหุ จะแตกต่างกันอย่างมาก ซึ่งขึ้นอยู่กับตัวแปรอิสระที่ใส่เข้าไปก่อนหน้านี้ในสมการถดถอยพหุ
3. ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัว อาจไม่มีนัยสำคัญทางสถิติ ทั้ง ๆ ที่ค่าสถิติ F ที่ทดสอบผลรวมของตัวแปรอิสระทั้งหมด มีนัยสำคัญทางสถิติ
4. ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละคู่ (simple correlation) มีค่าสูง ซึ่งอาจมีค่ามากกว่า 0.5 หรือ 0.6 ขึ้นไป
5. ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระที่สำคัญ มีขอบเขตความเชื่อมั่น (confidence interval) กว้างมาก
6. ค่าความคลาดเคลื่อนมาตรฐาน (standard error) ของตัวแปรอิสระที่สำคัญ มีค่าสูงมาก

การแก้ปัญหา multicollinearity ในทางปฏิบัติที่ง่ายที่สุด คือพิจารณาคัดเลือกตัวแปรอิสระที่มีความสำคัญน้อยกว่า ในระหว่างตัวแปรอิสระแต่ละคู่ที่มีความสัมพันธ์กันสูงออกจากสมการถดถอยพหุ

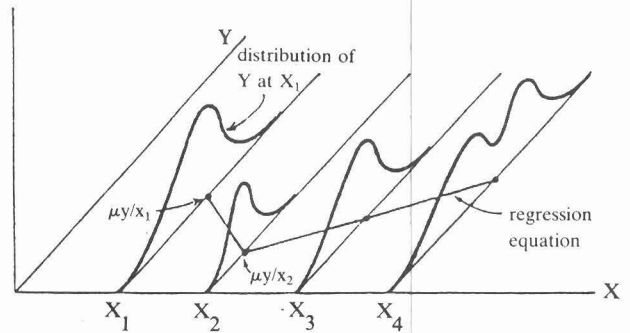
ข้อตกลงสำหรับรูปแบบการถดถอยพหุ

นักสถิติที่สร้างวิธีการวิเคราะห์การถดถอยพหุ ได้กำหนดข้อตกลงเบื้องต้นสำหรับคุณลักษณะของข้อมูลที่เหมาะสมกับเทคนิคของสถิตินี้ไว้อย่างชัดเจน ดังนั้นผลการวิเคราะห์การถดถอยพหุ จะเชื่อถือได้มากน้อยแค่ไหน ส่วนหนึ่งจะขึ้นอยู่กับคุณลักษณะของข้อมูลที่มีลักษณะเป็นไปตามข้อตกลง ซึ่งผู้วิจัยจำเป็นต้องระมัดระวังไว้ด้วย สำหรับข้อตกลงที่จะกล่าวในที่นี้ จะขอเน้นเฉพาะข้อตกลงที่สำคัญ ที่เกี่ยวข้องกับรูปแบบการถดถอยพหุเชิงเส้นตรงเท่านั้น ซึ่งมีข้อตกลงที่ผู้ใช้ควรทราบ 5 ประเด็นด้วยกัน (Kleinbaum and Kupper, 1978) คือ

ข้อตกลงที่ 1 ค่าเฉลี่ยของตัวแปรตามสำหรับทุก ๆ ค่าของผลรวมของตัวแปรอิสระทั้งหมด จะมีความสัมพันธ์กันแบบเชิงเส้นตรงดังภาพ ก. ถ้ามีการละเมิดข้อตกลงข้อที่ 1 นี้ ความสัมพันธ์ระหว่างตัวแปรอิสระ (X) กับตัวแปรตาม (Y) จะไม่เป็นแบบเส้นตรงดังภาพ ข. การละเมิดข้อตกลงนี้ ถือว่าเป็นข้อผิดพลาดอย่างมาก เพราะเป็นการใช้รูปแบบที่ผิดจากที่ตั้งไว้ ซึ่งทำให้ผลวิเคราะห์ผิดไปและมีผลกระทบทำให้การแปลผลผิดพลาดไปด้วย

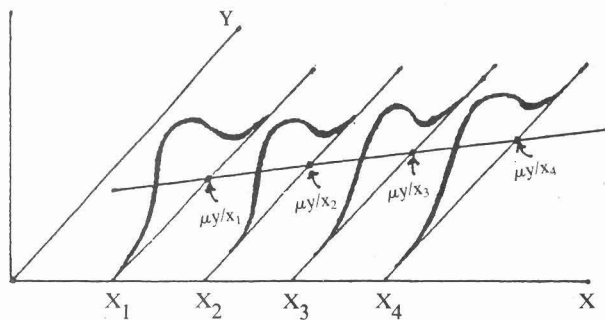


ภาพ ก. ลักษณะความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X ที่เป็นแบบเส้นตรง

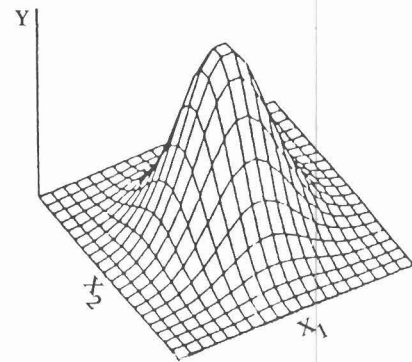


ภาพ ข. ลักษณะความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X ที่ไม่เป็นแบบเส้นตรง

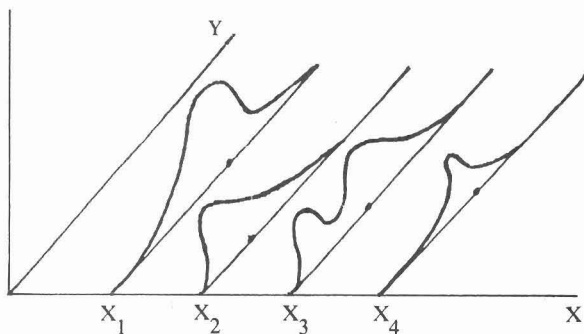
ข้อตกลงที่ 2 ตรรกะรวมของตัวแปรอิสระทั้งหมดแต่ละค่า จะมีค่าของตัวแปรตามมีลักษณะเป็นตัวแปรสุ่ม และมีการกระจายแบบโค้งปกติ ดังภาพ ก. และ ง. ส่วนภาพ จ. และภาพ ฉ. เป็นภาพที่ลักษณะการกระจายของตัวแปรตามไม่เป็นไปตามข้อตกลง จุดที่ต้องการเน้นในที่นี้ก็คือ ลักษณะการกระจายของตัวแปรตามต้องไม่ละเมิดข้อตกลงมากเกินไป ในกรณีที่ตัวแปรตามมีลักษณะการกระจายไม่เบ้ (skew) ไปทางหนึ่งทางใดมากนัก ผู้วิจัยสามารถปรับข้อมูล (transformation) ของตัวแปรตามนั้นเสียใหม่ เพื่อให้มีลักษณะการกระจายใกล้เคียงกับโค้งปกติได้



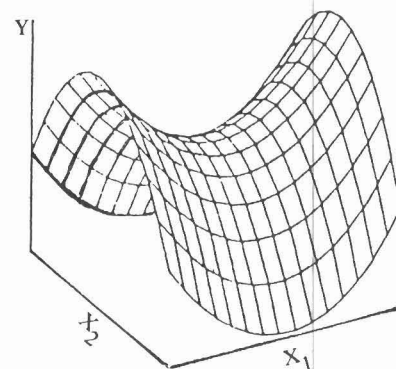
ภาพ ค. ลักษณะการกระจายแบบโค้งปกติของตัวแปรตาม Y ตามค่าที่แตกต่างกันของตัวแปรอิสระ X



ภาพ ง. ลักษณะการกระจายแบบโค้งปกติของตัวแปรตาม Y ตามค่าต่าง ๆ ของตัวแปรอิสระ X_1 และ X_2



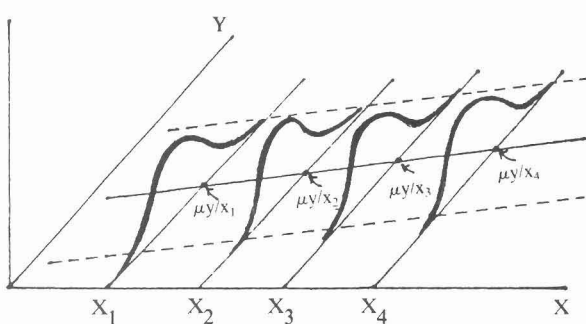
ภาพ จ. ลักษณะการกระจายของตัวแปรตาม Y ที่ไม่เป็นแบบโค้งปกติตามค่าที่แตกต่างกันของตัวแปรอิสระ X



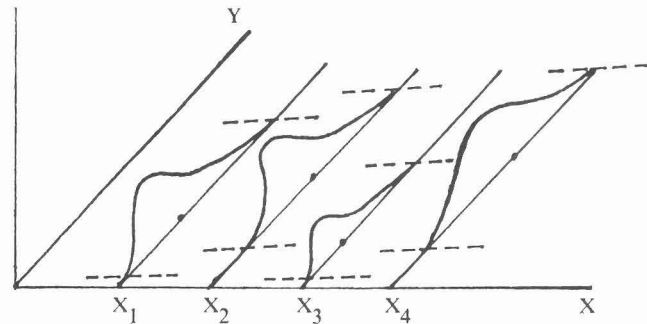
ภาพ ฉ. ลักษณะการกระจายของตัวแปรตาม Y ที่ไม่เป็นแบบโค้งปกติ ตามค่าต่าง ๆ ของตัวแปรอิสระ X_1 และ X_2

ข้อตกลงที่ 3 ค่าของตัวแปรตามแต่ละค่า (หรือค่าของ Y จากตัวอย่างแต่ละราย) ต้องเป็นอิสระต่อกัน ในบางครั้งมีการละเมิดข้อตกลงนี้โดยที่ผู้วิจัยไม่ได้พิจารณาอย่างรอบคอบ โดยเฉพาะการวัดค่าตัวแปรตามจากตัวอย่างคนเดียวกัน ในเวลาต่างกัน เช่น คำนวณน้ำหนักตัวที่วัดได้ครั้งแรก จะมีความสัมพันธ์กับค่าน้ำหนักตัวที่วัดในครั้งหลัง หรือ ระดับภูมิคุ้มกันที่วัดได้ในระยะหลัง จะมีส่วนสัมพันธ์กับระดับภูมิคุ้มกันที่วัดได้ในระยะแรก เป็นต้น

ข้อตกลงที่ 4 ค่าความแปรปรวน (variance) ของตัวแปรตาม สำหรับทุก ๆ ค่าของผลรวมของตัวแปรอิสระทั้งหมดจะต้องเท่ากัน (homoscedasticity) ดังแสดงในภาพ ข. โดยปกติถ้าลักษณะของข้อมูลมีการบิดเบือนไปจากข้อตกลงนี้ไม่มากนัก ก็พออนุโลมให้ใช้การวิเคราะห์การถดถอยพหุได้ โดยที่ยังคงให้ผลการวิเคราะห์ที่อยู่ในระดับที่เชื่อถือได้ แต่ถ้าลักษณะของข้อมูลมีการละเมิดข้อตกลงอย่างมาก ดังภาพ ข. จะทำให้การประมาณค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนสูงผิดปกติ จึงทำให้การทำนายโดยใช้สมการการถดถอยพหุมีโอกาสผิดพลาดมาก

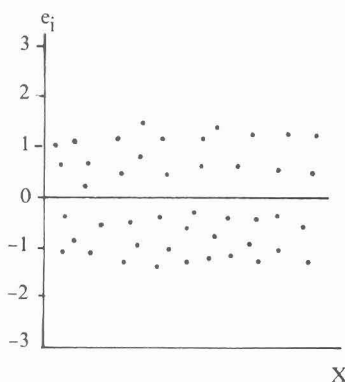


ภาพ ข. ลักษณะค่าความแปรปรวนที่เท่ากันของตัวแปรตาม Y สำหรับทุกค่าของตัวแปรอิสระ X

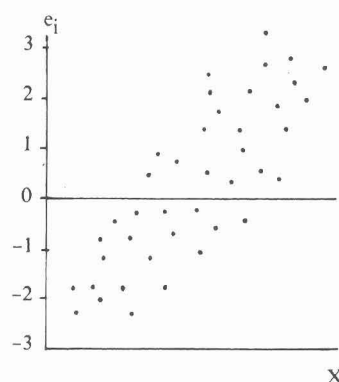


ภาพ ข. ลักษณะค่าความแปรปรวนที่แตกต่างกันของตัวแปรตาม Y สำหรับทุกค่าของตัวแปรอิสระ X

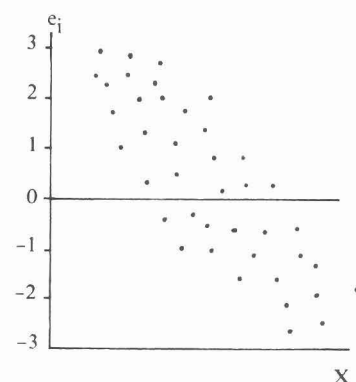
ข้อตกลงที่ 5 คุณลักษณะของค่าความคลาดเคลื่อนจากการทำนาย (residuals) โดยใช้สมการถดถอยพหุ ต้องมีการกระจายแบบโค้งปกติ ไม่มีความสัมพันธ์กันเอง (ภาพ ฉ.) และไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ ด้วย ปัญหาส่วนใหญ่ที่เกิดขึ้นมากที่สุดคือ ค่าความคลาดเคลื่อนมีความสัมพันธ์กันเอง (ภาพ ฉ. และภาพ จ.) ทำให้มีผลกระทบในการประมาณค่าสัมประสิทธิ์การถดถอย มีความผิดพลาดมากอีกเช่นกัน



ภาพ ฉ. ลักษณะของค่าความคลาดเคลื่อนที่ไม่มี ความสัมพันธ์กัน



ภาพ จ. ลักษณะของค่าความคลาดเคลื่อนที่มีความสัมพันธ์ในเชิงบวก



ภาพ จ. ลักษณะของค่าความคลาดเคลื่อนที่มีความสัมพันธ์ในเชิงลบ

สรุปแล้ว ผู้วิจัยควรตรวจสอบลักษณะข้อมูล ไม่ให้บิดเบือนจากข้อตกลงข้อหนึ่งข้อใดมากเกินไป มิฉะนั้นจะทำให้มีผลกระทบต่อความเชื่อถือได้ของผลการวิเคราะห์นั่นเอง

จำนวนตัวอย่าง

ปัจจัยที่มีส่วนเกี่ยวข้องกับการใช้สถิติการถดถอยพหุด้วย คือจำนวนตัวอย่าง ตัวอย่างที่น้อยเกินไป จะทำให้มีปัญหาในเรื่องความไม่แน่นอนของสมการถดถอยพหุ ในการทำนายการเปลี่ยนแปลงของตัวแปรตามเนื่องจากตัวแปรอิสระ ปัญหานี้มีให้เห็นอยู่บ่อย ๆ โดยเฉพาะงานวิจัยทางคลินิกที่ใช้ตัวอย่างน้อย เนื่องจากมีข้อจำกัดในการเลือกตัวอย่าง

โดยหลักการแล้ว ยังไม่มีข้อกำหนดที่แน่นอนว่า การวิเคราะห์การถดถอยพหุจะต้องใช้จำนวนตัวอย่างที่น้อยที่สุดเท่าไร ? จำนวนตัวอย่างจึงขึ้นอยู่กับปัจจัยที่สำคัญคือ ความแปรปรวนของตัวแปรที่วิเคราะห์ และระดับของความถูกต้องในการประมาณค่าพารามิเตอร์ ซึ่งพิจารณาได้จากค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอย (Levey and Lemeshow, 1980)

อย่างไรก็ตาม มีนักวิจัยและนักสถิติได้เสนอแนะในเรื่องจำนวนตัวอย่าง สำหรับการวิเคราะห์การถดถอยพหุไว้หลายท่าน เช่น Kerlinger and Pedhazur (1973) และ Thorndike (1978) Thorndike ได้เสนอแนะการคำนวณตัวอย่างโดยใช้หลักเกณฑ์ง่าย ๆ คือ ให้ใช้จำนวนตัวอย่าง 10 ราย ต่อตัวแปรอิสระ 1 ตัว แล้วเพิ่มตัวอย่างเข้าไปในขั้นสุดท้ายอีก 50 ราย ตัวอย่างเช่น ถ้าใช้ตัวแปรอิสระ 5 ตัวก็จะใช้จำนวนตัวอย่างที่น้อยที่สุด 100 ราย สำหรับ Kerlinger and Pedhazur ได้เสนอแนะเกี่ยวกับจำนวนตัวอย่างโดยพิจารณาจากสูตร $n/k \geq 30$ n คือจำนวนตัวอย่าง และ k คือ จำนวนตัวแปรอิสระ ตัวอย่างเช่น ในสมการถดถอยพหุที่ใช้จำนวนตัวแปรอิสระ 5 ตัว จะใช้จำนวนตัวอย่างประมาณ $n = 30 \times 5 = 150$ ซึ่งมากกว่าจำนวนตัวอย่างที่คำนวณได้จากข้อเสนอแนะของ Thorndike

ในทางปฏิบัติจริง ๆ การใช้จำนวนตัวอย่างตามข้อเสนอแนะที่กล่าวมาแล้วนั้น เป็นเรื่องที่ยากพอสมควร เพราะต้องใช้จำนวนตัวอย่างค่อนข้างมาก โดยเฉพาะในงานวิจัยทางคลินิกมักจะใช้จำนวนตัวอย่างน้อยกว่าข้อเสนอแนะทั้งสิ้น (Wilson and Reynold, 1982) ในที่นี้ผู้เขียนขอเสนอแนะทางสายกลางที่พอยืดหยุ่นสามารถปฏิบัติได้ โดยที่ผลของการวิเคราะห์ยังคงอยู่ในระดับที่เชื่อถือได้ คือ ให้ใช้เกณฑ์จำนวนตัวอย่าง 10 ราย ต่อ ตัวแปรอิสระ 1 ตัว ตามข้อเสนอแนะของ Thorndike แต่ไม่ต้องเพิ่มจำนวนตัวอย่างในขั้นสุดท้ายอีก 50 ราย ดังนั้นในกรณีที่ใช้ตัวแปรอิสระ 5 ตัว จึงใช้จำนวนตัวอย่างเพียง 50 รายเท่านั้นแทนที่จะเป็น 100 ราย เหมือนในตอนแรก อย่างไรก็ตามข้อเสนอแนะนี้เหมาะสำหรับงานวิจัย ที่มีปัญหาหรือมีข้อจำกัดอย่างมากในการเลือกตัวอย่าง และยอมรับผลการวิจัยให้มีความคลาดเคลื่อนได้พอสมควร ในกรณีที่งานวิจัยบางเรื่อง ไม่ยินยอมให้มีความคลาดเคลื่อนหรือยอมให้มีได้เพียงเล็กน้อย เนื่องจากเกรงว่าถ้านำผลงานวิจัยนั้นไปใช้ จะมีผลกระทบซึ่งอาจเป็นอันตรายต่อสิ่งมีชีวิตได้ งานวิจัยนั้นก็ควรใช้ตัวอย่างจำนวนมากกว่าข้อเสนอแนะของผู้เขียน และทางที่ดีก็ควรใช้เกณฑ์ของ Kerlinger and Pedhazur (1973) หรือ Thorndike (1978) จะทำให้ผลการวิจัยที่ใช้สถิติการถดถอยพหุ เป็นที่น่าเชื่อถือได้มากยิ่งขึ้น

การแปลผลค่าสัมประสิทธิ์สหสัมพันธ์ (R^2)

ค่าสถิติที่ใช้แสดงความเหมาะสมของรูปแบบสมการการถดถอยพหุ คือค่าสัมประสิทธิ์สหสัมพันธ์ (R^2) ซึ่ง “เป็นค่าที่แสดงถึง สัดส่วนของความแปรปรวนในตัวแปรตาม ซึ่งอธิบายโดยตัวแปรอิสระทั้งหมดในรูปแบบสมการการถดถอยพหุนั้น” หรือกล่าวเป็นภาษาง่าย ๆ คือ “ตัวแปรอิสระทั้งหมดในรูปแบบสมการ สามารถมีอิทธิพลให้เกิดการเปลี่ยนแปลงในตัวแปรตามได้กี่เปอร์เซ็นต์” โดยปกติค่า R^2 มีค่าอยู่ระหว่าง 0 ถึง 1 ค่าที่

ใกล้ 1 แสดงว่ารูปแบบสมการการถดถอยพหุนั้น สามารถอธิบายความแปรปรวนในตัวแปรตามได้เป็นส่วนใหญ่ เช่น ค่า R^2 เท่ากับ 0.85 ก็หมายความว่า ตัวแปรอิสระทั้งหมดในรูปแบบสมการ สามารถมีอิทธิพลทำให้เกิดการเปลี่ยนแปลงในตัวแปรตามได้ 85 เปอร์เซ็นต์ และในทางตรงกันข้าม ค่า R^2 ที่น้อย (ต่ำกว่า 0.50 ลงมา) ก็หมายถึง ตัวแปรอิสระทั้งหมดในรูปแบบสมการ ไม่ค่อยมีอิทธิพลทำให้เกิดการเปลี่ยนแปลงในตัวแปรตามเท่าไรนัก

อย่างไรก็ตามในทางปฏิบัติ การแปลความหมายค่า R^2 ผู้วิจัยควรระมัดระวังจุดอ่อนในประเด็นปลีกย่อยเหล่านี้ไว้ด้วยคือ ค่า R^2 ที่สูงก็ไม่จำเป็นต้องแสดงว่ารูปแบบสมการการถดถอยพหุดังกล่าวดีเสมอไป สาเหตุนี้อาจเกิดขึ้นได้เนื่องจากสองกรณี กรณีแรกอาจเกิดจากตัวแปรตามมีค่าความแปรปรวนน้อย และกรณีที่สองอาจเกิดจากตัวแปรอิสระบางตัวที่ใส่เข้าไปในสมการ มีความสัมพันธ์กับตัวแปรตามในลักษณะที่ไม่เป็นไปในเชิงเหตุและผลก็ได้ (ถ้าผู้วิจัยไม่ได้เลือกตัวแปรอิสระอย่างถูกต้อง โดยใช้หลักความเป็นเหตุเป็นผล) และในทางกลับกัน ค่า R^2 ที่ต่ำก็ไม่ได้หมายความว่า รูปแบบสมการการถดถอยพหุนั้นไม่ดีเสมอไป ทั้งนี้เพราะค่า R^2 ส่วนหนึ่งขึ้นอยู่กับลักษณะข้อมูลของตัวแปรในสมการด้วย ในกรณีที่รูปแบบสมการดีเหมาะสม แต่ถ้าตัวแปรส่วนใหญ่ในสมการมีค่าความแปรปรวนผิดปกติ ก็สามารถทำให้ค่า R^2 ต่ำได้เช่นกัน

สรุป

การวิเคราะห์การถดถอยพหุ เป็นสถิติที่ใช้มากในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัว และตัวแปรตามที่มีระดับการวัดเป็นแบบช่วงมาตรา (interval scale) โดยเน้นในลักษณะความสัมพันธ์ของตัวแปรอิสระ ที่มีผลกระทบหรือมีอิทธิพลต่อตัวแปรตามเป็นพื้นฐานสำคัญ การวิเคราะห์การถดถอยพหุเป็นวิธีอธิบายความสัมพันธ์ ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม ซึ่งสะท้อนให้เห็นความเป็นจริงตามธรรมชาติ มากกว่าการวิเคราะห์การถดถอยแบบง่าย ที่ใช้ตัวแปรอิสระเพียงตัวเดียว การใช้สถิติการวิเคราะห์การถดถอยพหุ ผู้วิจัยควรระมัดระวังจุดอ่อนที่สำคัญไว้ดังนี้

1. การเลือกตัวแปรอิสระเข้าในสมการถดถอย ควรพิจารณาถึงความเป็นเหตุเป็นผลให้มาก ไม่ใช่ปล่อยให้คอมพิวเตอร์ช่วยเลือกให้อย่างเดียว โดยเฉพาะในปัจจุบันมีโปรแกรมคอมพิวเตอร์สำเร็จรูปทางสถิติมากมาย ที่ใช้ช่วยวิเคราะห์การถดถอยพหุได้อย่างสะดวก เพียงแต่ผู้วิจัยตั้งจำนวนตัวแปรอิสระทั้งหมดเข้าไปในสมการ แล้วให้คอมพิวเตอร์ช่วยคัดเลือกให้อีกที คอมพิวเตอร์จะคัดเลือกตัวแปรไว้ตามหลักเกณฑ์ทางสถิติเท่านั้น แต่คอมพิวเตอร์ไม่สามารถบอกได้ว่าตัวแปรที่คัดเลือกนั้น มีความสำคัญที่แสดงความสัมพันธ์กับตัวแปรตามในเชิงเหตุและผลได้ ซึ่งเรื่องนี้เป็นหน้าที่ของผู้วิจัย ที่จะต้องพิจารณากันกรองตั้งแต่แรกก่อนที่จะใส่ตัวแปรเข้าไปในสมการ ถ้าผู้วิจัยไม่คำนึงถึงเรื่องนี้ ตัวแปรอิสระที่ใช้ในสมการถึงแม้ว่าคอมพิวเตอร์จะคัดเลือกแล้วว่ามีค่าสำคัญทางสถิติ ก็อาจไม่มีความหมายในเชิงตรรกวิทยาก็ได้ และทำให้การแปลผลการวิเคราะห์ในขั้นสุดท้ายไม่มีความหมายแต่อย่างใด

2. ผลการวิเคราะห์ข้อมูลที่ใช้สถิติการถดถอยพหุ จะเชื่อถือได้มากน้อยแค่ไหน ขึ้นอยู่กับว่ารูปแบบการถดถอยพหุเหมาะสมหรือไม่ และคุณลักษณะของข้อมูลเป็นไปตามข้อตกลงของสถิติที่กำหนดไว้หรือไม่ ซึ่งโดยปกติการวิเคราะห์การถดถอยพหุ จะยินยอมให้ลักษณะของข้อมูลละเมิดข้อตกลงได้บ้างในระดับเล็กน้อย จนถึงระดับปานกลาง ถ้าเป็นการละเมิดข้อตกลงในระดับปานกลาง ผู้วิจัยอาจต้องทำการปรับแก้ไขข้อมูลเสียใหม่ ให้มีคุณลักษณะใกล้เคียงกับข้อตกลงนั้นเสียก่อน สำหรับการละเมิดข้อตกลงอย่างรุนแรง ก็ไม่สมควรที่จะใช้สถิติวิธีนี้ ควรหันไปพิจารณาสถิติวิธีอื่นที่เหมาะสมต่อไป อนึ่งเท่าที่สังเกตงานวิจัยส่วนใหญ่ในปัจจุบัน นักวิจัยมักจะมองข้ามการทดสอบคุณลักษณะของข้อมูลตามข้อตกลงของสถิตินี้ไป ความจริงแล้วตรงนี้เป็นจุดอ่อนที่สำคัญมาก เพราะถ้าข้อมูลมีลักษณะบิดเบือนจากข้อตกลงอย่างมากแล้ว ถึงแม้ว่างานวิจัยนั้นมีขบวนการทำมาอย่างดีเพียงไร ผลการวิจัยก็เชื่อถือไม่ได้ด้วยนั่นเอง

เอกสารอ้างอิง

- Kerlinger F N, and Pedhazur E J. 1973. *Multiple regression in behavioral research*. New York: Holt, Rinehart and Wilson, Inc.
- Kleinbaum D G, and Kupper L L. 1978. *Applied regression analysis and other multivariable methods*. North Scituate, Massachusetts: Duxbury Press, pp.136-137.
- Levy P, and Lemeshow S. 1980. *Sampling for health professionals*. Belmont, California: Wadsworth, Inc.
- Thorndike R. 1978. *Correlational procedures for research*. New York: Gardner Press, Inc. p.184.
- Wilson V L, and Reynold C R. 1982. *Methodological and statical problems in determining membership in clinical populations*. *Clinical Neuropsychology* 4: 134-138.