การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression)

ฉัตรศิริ ปิยะพิมลสิทธิ์

ในการวิเคราะห์การถดถอยอย่างง่ายจะเป็นการวิเคราะห์กับตัวแปรตามหรือตัวแปรเกณฑ์ (Y) โดยมีตัวแปรอิสระหรือตัวแปรทำนาย (X) เพียงตัวเดียว อย่างไรก็ตาม ในบางปรากฏการณ์ที่ต้องอธิบายหรือทำนายทางสังคมศาสตร์ หากใช้ตัวแปรอิสระเพียงตัวเดียวจะมีข้อจำกัด ในการอธิบายพฤติกรรมของมนุษย์ซึ่งโดยมากจะมีความซับซ้อน การใช้ตัวแปรทำนายเพียงตัวเดียวจะไม่มีประสิทธิภาพพอที่จะอธิบายตัวแปรเกณฑ์ได้ ในกรณีที่จะพยายามอธิบายสัดส่วนความแปรปรวนของตัวแปรเกณฑ์ได้อย่างมีประสิทธิภาพ จำเป็นต้องมีตัวแปรทำนายมากกว่า 1 ตัว ซึ่งจะนำไปสู่การวิเคราะห์การถดถอยพหุคูณเมื่อมีตัวแปรทำนายตั้งแต่ 2 ตัวขึ้นไปใช้ในการทำนายตัวแปรเกณฑ์ ซึ่งโดยปกติตัวแปรทำนายหรือตัวแปรอิสระจะใช้สัญลักษณ์ X และตัวแปรเกณฑ์หรือตัวแปรตามจะใช้สัญลักษณ์ Y

ตัวอย่างกรณีที่ผู้บริหารสถานศึกษาต้องการใช้คะแนนสอบเข้าศึกษาต่อในระดับบัณฑิตศึกษา (Graduate Record Exam : GRE) ในการทำนายเกรดเฉลี่ยระดับบัณฑิตศึกษา (GPA) ที่จะช่วยผู้บริหารในการตัดสินใจทางการบริหาร แต่เนื่องด้วยมีตัวแปรทำนายอื่น ๆ ที่มีศักยภาพในการทำนายซึ่งอาจจะเป็นเกรดเฉลี่ยในระดับปริญญาตรี ผลการประเมินจากอาจารย์หรือผลการสอบสัมภาษณ์ คำถามการวิจัยเพื่อที่จะพิจารณาว่า GRE, GPA ระดับปริญญาตรี, ผลการประเมินจากอาจารย์ และคะแนนสอบสัมภาษณ์ (ตัวแปรอิสระหรือตัวแปรทำนาย) สามารถทำนายเกรดเฉลี่ยระดับบัณฑิตศึกษาได้หรือไม่ (ตัวแปรตามหรือตัวแปรเกณฑ์) ซึ่งในตัวอย่างที่ยกมานี้สามารถใช้การวิเคราะห์การถดถอยเนื่องจากเหมาะในสถานการณ์ที่มีตัวแปรทำนายหลายตัว

แนวคิดของการวิเคราะห์การถดถอยพหุคูณจะคล้ายกับการวิเคราะห์การถดถอยอย่างง่าย อย่างไรก็ตาม เนื่องจากมีตัวแปรทำนายหลายตัว การคำนวณก็จะยากและซับซ้อนมากขึ้น ซึ่งในเนื้อหานี้จะแสดงตัวแปรการคำนวณในกรณีที่มีตัวแปรทำนายเพียง 2 ตัว การคำนวณกรณีมีตัวแปรทำนายมากกว่า 2 ตัวต้องใช้เมตริกและแคลคูลัสเข้าช่วย ซึ่งจะไม่กล่าวถึงในที่นี้ แต่การคำนวณทั้งหมดนี้จะง่ายขึ้นหากมีการใช้โปรแกรมคอมพิวเตอร์เข้าช่วยในการคำนวณ

ในเนื้อหานี้จะแสดงแนวคิดของสหสัมพันธ์แยกส่วน (partial correlation) สหสัมพันธ์ครึ่งส่วน (semipartial correlation) และสหสัมพันธ์พหุคูณ (multiple correlation) สัมประสิทธิ์การอธิบาย (coefficient of multiple determination) การเพิ่มขึ้นของสัดส่วนความแปรปรวนในการอธิบายเมื่อเพิ่มตัวแปรทำนาย กระบวนการเลือกตัวแปรเข้าสมการทำนาย การตรวจสอบนัยสำคัญ และข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุคูณ

สหสัมพันธ์แยกส่วน ครึ่งส่วนและพหุคูณ

ในการอธิบายเกี่ยวกับการวิเคราะห์การถดถอย ลองพิจารณาแนวคิดเกี่ยวกับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยเฉพาะการอธิบายเกี่ยวกับสหสัมพันธ์แยกส่วนและครึ่งส่วน การวิเคราะห์การถดถอยพหุคูณจะเกี่ยวข้องกับการใช้ตัวแปรทำนายตั้งแต่ 2 ตัวขึ้นไปและตัวแปรเกณฑ์เพียง 1 ตัว ดังนั้นจำนวนที่น้อยที่สุดของตัวแปรคือ 3 ตัวแปรที่จะอยู่ในการวิเคราะห์ ถ้าคิดว่าตัวแปรในการวิเคราะห์นี้จะใช้การหาความสัมพันธ์แบบเพียร์สัน (the Pearson Product-Modment Correlation Coefficient) แล้ว จะเกิดปัญหาเพราะสหสัมพันธ์นี้เป็นการหาความสัมพันธ์ระหว่างตัวแปรได้ครั้งละ 2 ตัวแปรเท่านั้น อย่างไรก็ตามตัวแปรที่เพิ่มเข้ามาในการหาความสัมพันธ์จะทำอย่างไร คำตอบนี้ต้องหาโดยใช้สหสัมพันธ์แยกส่วน สหสัมพันธ์ครึ่งส่วนและสหสัมพันธ์พหุคูณ ดังจะได้อธิบายต่อไปนี้

สหสัมพันธ์แยกส่วน (Partial Correlation)

ในเบื้องต้นจะอธิบายแนวคิดของสหสัมพันธ์แยกส่วนเสียก่อน กรณีที่ง่ายที่สุดจะเกี่ยวข้องกับความสัมพันธ์ระหว่าง 3 ตัวแปร ซึ่งให้ชื่อว่า X_1, X_2 และ X_3 การวิเคราะห์สหสัมพันธ์แยกส่วนระหว่าง X_1 และ X_2 เมื่อ X_3 ถูกควบคุมเอาไว้หรือถูกขจัดออก นั่นคือ อิทธิพลของ X_3 ถูกขจัดออกจากทั้ง X_1 และ X_2 (ตัวแปรทั้งสองถูกปรับแก้ด้วย X_3) ดังนั้นสหสัมพันธ์แยกส่วนจะแสดงความสัมพันธ์เชิงเส้นตรงระหว่าง X_1 และ X_2 ที่เป็นอิสระจากอิทธิพลของ X_3 สหสัมพันธ์แยกส่วนนี้ใช้สัญลักษณ์ว่า $r_{12.3}$ สังเกตว่า ในสัญลักษณ์ไม่แสดง X และมี . เป็นตัวแบ่งระหว่างตัวแปรที่สัมพันธ์กับตัวแปรที่ถูกควบคุม วิธีการคำนวณ $r_{12.3}$ มีสูตรดังนี้

$$r_{12.3} = [r_{12} - r_{13}r_{23}] / \sqrt{[(1-r_{13}^2)(1-r_{23}^2)]}$$

สังเกตว่าในสูตรคำนวณนี้จะใช้สหสัมพันธ์ระหว่างตัวแปร 2 ตัว

ตัวอย่างการหาสหสัมพันธ์แยกส่วน เช่น นักวิจัยสนใจหาความสัมพันธ์ระหว่างส่วนสูง (X_1) และน้ำหนัก (X_2) ซึ่งจะเกี่ยวข้องถึงช่วงอายุของแต่ละคน (X_3) จาก 6 เดือน ถึง 65 ปี สหสัมพันธ์อย่างง่าย ได้ค่า $r_{12} = 0.7$, $r_{13} = 0.1$ และ $r_{23} = 0.6$ สามารถคำนวณ $r_{12.3}$ ได้ว่า

$$\begin{aligned} r_{12.3} &= [r_{12} - r_{13}r_{23}] / \sqrt{[(1-r_{13}^2)(1-r_{23}^2)]} \\ &= [0.7 - 0.1(0.6)] / \sqrt{[(1-0.01)(1-0.36)]} \\ &= 0.80 \end{aligned}$$

สหสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือส่วนสูงและน้ำหนักที่ไม่สนใจอายุ ($r_{12} = 0.7$) จะมีค่าน้อยกว่าสหสัมพันธ์แยกส่วนระหว่างส่วนสูงและน้ำหนักที่ควบคุมอายุเอาไว้ ($r_{12.3} = 0.8$) นั่นคือ สหสัมพันธ์ระหว่างส่วนสูงและน้ำหนักจะสูงขึ้นเมื่อมีการขจัดอิทธิพลของอายุออก แม้ว่าสถิตินี้จะเกี่ยวข้องกับการควบคุมตัวแปรเอาไว้ แต่ในสถานการณ์จริงตัวแปรบางตัวไม่สามารถจะควบคุมให้คงที่ได้

ในบางครั้งอาจจะสนใจผลของสหสัมพันธ์แยกส่วนที่มีโอกาสเกิดขึ้นได้ในกรณีที่มีค่าผิดปกติ ถ้าทั้ง r_{13} และ r_{23} มีค่าเท่ากับ 0 แล้ว $r_{12} = r_{12.3}$ นั่นคือ ถ้าตัวแปรที่ต้องการควบคุมไม่มีความสัมพันธ์กับตัวแปรอีก 2 ตัวที่ต้องการหาความสัมพันธ์กันแล้ว การใช้สหสัมพันธ์แยกส่วนก็จะไม่มีประโยชน์ มีความเป็นไปได้ในกรณีที่สหสัมพันธ์แยกส่วนอาจจะมีความสัมพันธ์มากกว่าหรือน้อยกว่าสหสัมพันธ์ระหว่าง 2 ตัวแปร และบางกรณีที่มีสหสัมพันธ์แยกส่วนเท่ากับ 0 เมื่อสหสัมพันธ์ระหว่าง 2 ตัวแปรไม่เท่ากับ 0

ในการอ้างอิงการเรียกสหสัมพันธ์แยกส่วนนั้น สหสัมพันธ์ระหว่าง 2 ตัวแปรที่มีการควบคุมตัวแปร 1 ตัวแปร จะเรียกว่า first-order partial correlation (เช่น $r_{12.3}$) หากมีตัวแปรที่ควบคุม 2 ตัวแปร จะเรียกว่า second-order partial correlation (เช่น $r_{12.34}$) หรือใช้สัญลักษณ์ทั่วไปได้ว่า $r_{12.w}$ เมื่อ w แสดงถึงจำนวนตัวแปรที่ถูกควบคุมทั้งหมด และการคำนวณสหสัมพันธ์แยกส่วนก็จะซับซ้อนมากกว่า first-order partial correlation

ถ้าสนใจที่จะทดสอบนัยสำคัญของสหสัมพันธ์แยกส่วน จะคล้ายกับการทดสอบนัยสำคัญของสหสัมพันธ์ระหว่าง 2 ตัวแปร ซึ่งสหสัมพันธ์ระหว่าง 2 ตัวแปรจะเขียนสมมติฐานได้ว่า $H_0 : \rho = 0$ มีสูตรว่า

$$z = [z_r - z_0][\sqrt{n-3}]$$

เมื่อ z_r และ z_0 เป็นค่าฟิชเชอร์แปลงรูป และนำค่า z ที่ได้ไปเปรียบเทียบกับค่าวิกฤติที่เปิดจากตารางสถิติ การทดสอบแบบสองทางของค่าวิกฤติ $_{\alpha/2}z$ และ $_{1-\alpha/2}z$ สำหรับการทดสอบแบบทางเดียวค่าวิกฤติจะเป็น $_{\alpha}z$ และ $_{1-\alpha}z$ ขึ้นอยู่กับสมมติฐานอื่น

การทดสอบสหสัมพันธ์แยกส่วน ($H_0 : \rho_{12.w} = 0$) แล้ว

$$z = [z(r_{12.w}) - z_0][\sqrt{n-3-n_w}]$$

เมื่อ n_w คือจำนวนของตัวแปรที่ถูกควบคุม และอ้างอิงค่าวิกฤติจากตารางสถิติ

สหสัมพันธ์ครึ่งส่วน (Semipartial (Part) Correlation)

แนวคิดถัดมาคือสหสัมพันธ์ครึ่งส่วน (Semipartial or Part Correlation) กรณีที่ง่ายคือกรณีที่มีตัวแปรเพียง 3 ตัว ซึ่งให้ชื่อว่า X_1, X_2 และ X_3 การวิเคราะห์สหสัมพันธ์ครึ่งส่วนจะเป็นการหาสหสัมพันธ์ระหว่าง X_1 และ X_2 เมื่อ X_3 ถูกขจัดออกจาก X_2 เท่านั้น นั่นคือ อิทธิพลของ X_3 จะถูกขจัดออกจาก X_2 เท่านั้น ดังนั้น สหสัมพันธ์ครึ่งส่วนจึงเป็นการแสดงถึงความสัมพันธ์เชิงเส้นระหว่าง X_1 และ X_2 หลังจากสัดส่วนความแปรปรวนของ X_2 ที่สามารถทำนายได้จาก X_3 ถูกขจัดออกจาก X_2 สหสัมพันธ์ครึ่งส่วนจะใช้สัญลักษณ์ว่า $r_{1(2.3)}$ เมื่อ X จะไม่แสดงในสัญลักษณ์และจุดที่คั่นหมายถึงการขจัดอิทธิพลของตัวแปร วิธีการคำนวณมีสูตรดังนี้

$$r_{1(2.3)} = (r_{12} - r_{13}r_{23})/\sqrt{(1-r_{23}^2)}$$

สังเกตว่าสูตรคำนวณจะใช้ค่าสหสัมพันธ์ระหว่างตัวแปรเพียง 2 ตัว

ตัวอย่างในกรณีการหาค่าสหสัมพันธ์ครึ่งส่วน เช่น ผู้วิจัยสนใจหาความสัมพันธ์ระหว่าง GPA (X_1) และคะแนน GRE (X_2) และผู้วิจัยต้องการขจัดอิทธิพลของเชาวน์ปัญญา (X_3) ออกจากคะแนน GRE แต่ไม่ต้องการขจัดออกจาก GPA ค่าสหสัมพันธ์ระหว่างตัวแปรได้ค่า r_{12} = 0.5, r_{13} = 0.3 และ r_{23} = 0.7 สามารถคำนวณ $r_{1(2.3)}$ ได้ค่าดังนี้

$$\begin{aligned} r_{1(2.3)} &= (r_{12} - r_{13}r_{23})/\sqrt{(1-r_{23}^2)} \\ &= (0.5 - 0.3(0.7))/\sqrt{(1-0.49)} \\ &= 0.41 \end{aligned}$$

ซึ่งจะกลายเป็นว่าสหสัมพันธ์ระหว่างตัวแปร GPA และ GRE ที่ไม่สนใจตัวแปรเชาวน์ปัญญา (r_{12} = 0.5) มีค่าสูงกว่าสหสัมพันธ์ครึ่งส่วนระหว่าง GPA และ GRE ที่ได้ควบคุมเชาวน์ปัญญาออกจากตัวแปร GRE ($r_{1(2.3)}$= 0.41)

ในการเรียกชื่อสหสัมพันธ์ครึ่งส่วนทำนองเดียวกับสหสัมพันธ์แยกส่วน ถ้ามีเพียงตัวแปรเดียวที่ขจัดออกจาก X_2 จะเรียกว่า the first-order semipartial correlation กรณีที่มีตัวแปร 2 ตัวที่ขจัดออกจาก X_2 ใช้สัญลักษณ์ $r_{1(2.34)}$ จะเรียกว่า the second-order semipartial correlation และ the higher-order semipartial correlation ($r_{1(2.345)}$) หรืออาจจะแสดงสหสัมพันธ์ครึ่งส่วนโดยทั่วไปด้วยสัญลักษณ์ $r_{1(2.w)}$ เมื่อ w คือตัวแปรที่ต้องการขจัดออกจาก X_2 และการคำนวณค่าสหสัมพันธ์ครึ่งส่วนจะซับซ้อนกว่ากรณีมีตัวแปรที่ต้องการขจัดออกเพียงตัวเดียว สุดท้ายคือการทดสอบนัยสำคัญของสหสัมพันธ์ครึ่งส่วนจะมีสูตรคำนวณทำนองเดียวสหสัมพันธ์แยกส่วน

สังเกตว่าในการหาสหสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือมากกว่า (เช่น สหสัมพันธ์แยกส่วน หรือสหสัมพันธ์ครึ่งส่วน) จะใช้สำหรับการตรวจในโมเดลการวิเคราะห์การถดถอยพหุคูณ (multiple regression model) เมื่อมีตัวแปรพยากรณ์ตั้งแต่ 2 ตัวขึ้นไป

การถดถอยพหุคูณเชิงเส้นตรง (Multiple Linear Regression)

ต่อไปเป็นแนวคิดของการถดถอยพหุคูณเชิงเส้นตรง เพื่อรวบรัดจะไม่กล่าวถึงสมการที่อยู่ในรูปของค่าพารามิเตอร์ของกลุ่มประชากร จะกล่าวถึงสมการที่อยู่ในรูปค่าสถิติของกลุ่มตัวอย่าง สมการการถดถอยพหุคุณที่อยู่ในรูปคะแนนมาตรฐานและคะแนนดิบ สัมประสิทธิ์การอธิบาย สหสัมพันธ์พหุคูณ การทดสอบนัยสำคัญ และข้อตกลงเบื้องต้นของสถิติ

สมการถดถอยที่อยู่ในรูปคะแนนดิบ (Unstandardized Regression Equation)

พิจารณาสมการถดถอยเชิงเส้นของกลุ่มตัวอย่างสำหรับการถดถอย Y บน $X_{1,2,...,m}$ คือ

$$Y_i = b_1X_{1i} + b_2X_{2i} + ... + b_mX_{mi} + a + e_i$$

เมื่อ Y คือตัวแปรเกณฑ์ และ X_k คือตัวแปรพยากรณ์ เมื่อ k = 1, ..., m , b_k คือความชัน (partial slope) ของเส้นถดถอยสำหรับ Y ที่ถูกทำนายด้วย X_k, a คือจุดตัดของเส้นถดถอย

สำหรับ Y ที่ถูกทำนายด้วย X_k , e_i คือความคลาดเคลื่อนของการพยากรณ์ของ Y ที่ไม่สามารถพยากรณ์ได้ด้วย X_k และ i คือสัญลักษณ์แทนตัวอย่างที่ i โดยที่ i มีค่าตั้งแต่ 1 จนถึง n เมื่อ n แทนขนาดของกลุ่มตัวอย่าง (โดยปกติเขียน i = 1, ..., n) มีการใช้ความชัน (partial slope) เพราะว่าจะแสดงถึงความชันของ Y บน X_k เฉพาะตัวนั้น ๆ ซึ่งมีการขจัดอิทธิพลของ X_k ตัวอื่น ๆ ออก จึงเป็นเหตุให้มีการใช้สหสัมพันธ์แยกส่วน

สมการทำนายของกลุ่มตัวอย่างคือ

$$Y_i' = b_1X_{1i} + b_2X_{2i} + ... + b_mX_{mi} + a$$

เมื่อ Y' คือค่าที่ถูกทำนายของ Y เมื่อมีการแทนค่า X_k และค่าอื่น ๆ ในสมการ สามารถคำนวณความคลาดเคลื่อน e_i สำหรับแต่ละตัวอย่างจากสมการทำนาย โดยการหาความแตกต่างระหว่าง Y จริงกับ Y ที่ถูกทำนาย จะได้ค่าความคลาดเคลื่อนของตัวอย่างด้วยสมการ

$$e_i = Y_i - Y_i'$$

สำหรับทุกค่า i = 1, ..., n

ในการคำนวณความชันและจุดตัดในการถดถอยพหุคูณกรณีมีตัวแปรทำนาย 2 ตัว โดยปกติจะใช้โปรแกรมคอมพิวเตอร์ช่วยในการคำนวณ

สำหรับกรณีมีตัวแปรทำนาย 2 ตัวแล้ว ความชันและจุดตัดสามารถคำนวณได้ด้วยสูตร

$$b_1 = [(r_{Y1} - r_{Y2}r_{12})s_Y]/[(1 - r_{12}^2)s_1]$$

$$b_2 = [(r_{Y2} - r_{Y1}r_{12})s_Y]/[(1 - r_{12}^2)s_2]$$

และ $$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$$

ความชัน b_1 อ้างอิงว่าเป็น 1) ค่าคาดหวังหรือการเปลี่ยนแปลงใน Y เมื่อ X_1 เปลี่ยนแปลงไป 1 หน่วยโดยที่ X_2 คงที่ 2) อิทธิพลของ X_1 ที่มีต่อ Y เมื่อ X_2 คงที่ และ 3) สัมประสิทธิ์การถดถอยในรูปของคะแนนดิบ สำหรับ b_2 ก็อ้างอิงทำนองเดียวกัน จุดตัดอ้างอิงว่าเป็น 1) ค่าของ Y เมื่อ X_1 และ X_2 เป็น 0 และ 2) ค่าเฉลี่ยของ Y เมื่อ X_1 และ X_2 เป็น 0

อีกวิธีสำหรับการคำนวณความชันจะเกี่ยวข้องกับการใช้สหสัมพันธ์แยกส่วน มีสูตรการคำนวณดังนี้

$$b_1 = r_{Y1.2}\{[s_Y\sqrt{(1 - r_{Y2}^2)}]/[s_1\sqrt{(1 - r_{12}^2)}]\}$$

และ $$b_2 = r_{Y2.1}\{[s_Y\sqrt{(1 - r_{Y1}^2)}]/[s_2\sqrt{(1 - r_{12}^2)}]\}$$

มีเกณฑ์อย่างไรในการใช้ค่าความชันและจุดตัดในสมการการถดถอย มีอยู่หลายวิธีในการคำนวณหาความชันและจุดตัด สำหรับเกณฑ์นั้นโดยปกติจะใช้การวิเคราะห์การถดถอยพหุคูณเชิงเส้น ซึ่งใช้เกณฑ์กำลังสองต่ำสุด เกณฑ์กำลังสองต่ำสุดจะเป็นการคำนวณหาค่าสำหรับความชันและจุดตัดที่เป็นผลรวมของกำลังสองของความคลาดเคลื่อนในการทำนายคือค่าความคลาดเคลื่อนต่ำสุด นั่นคือต้องคำนวณหาสมการถดถอยที่นิยามว่าเป็นชุดของความชันและจุดตัดที่มีผลรวมของกำลังสองของความคลาดเคลื่อนต่ำสุด

พิจารณาการวิเคราะห์กับตัวอย่างจริง ๆ ที่จะนำเสนอในเนื้อหานั้น โดยตัวอย่างจะเป็นคะแนนสอบเข้าระดับบัณฑิตศึกษาฉบับภาษา (GRETOT) และเกรดเฉลี่ยระดับปริญญาตรี (UGPA) ในการทำนายเกรดเฉลี่ยระดับบัณฑิตศึกษา (GGPA) สำหรับ GRETOT มีพิสัยของคะแนนอยู่ระหว่าง 40 ถึง 160 และ GPA มีพิสัยของอยู่ระหว่าง 0.00 ถึง 4.00 เก็บตัวอย่างกับนักเรียน 11 คน แสดงข้อมูลในตาราง 1 จากตัวอย่างนี้ใช้การวิเคราะห์การถดถอยพหุคูณ

ตาราง 1 ข้อมูล GRE-GPA

นักเรียนคนที	GRETOT	UGPA	GGPA
1	145	3.2	4.0
2	120	3.7	3.9
3	125	3.6	3.8
4	130	2.9	3.7
5	110	3.5	3.6
6	100	3.3	3.5
7	95	3.0	3.4
8	115	2.7	3.3
9	105	3.1	3.2
10	90	2.8	3.1
11	105	2.4	3.0

ตัวแปร GRETOT (X_1) ได้ค่า $\bar{X}_1 = 112.73$ และ $s_1^2 = 266.82$ สำหรับ UGPA (X_2) ได้ค่า $\bar{X}_2 = 3.11$ และ $s_2^2 = 0.16$ และ GGPA (Y) ได้ค่า $\bar{Y} = 3.50$ และ $s_Y^2 = 0.11$ นอกจากนี้ยังคำนวณ $r_{Y1} = 0.78$, $r_{Y2} = 0.75$ และ $r_{12} = 0.30$ ความชันและจุดตัดคำนวณได้ดังนี้

$$b_1 = [(r_{Y1} - r_{Y2}r_{12})s_Y]/[(1 - r_{12}^2)s_1]$$
$$= [(0.78 - 0.75(0.30))0.33]/[(1 - 0.30^2)16.33]$$
$$= 0.01$$

$$b_2 = [(r_{Y2} - r_{Y1}r_{12})s_Y]/[(1 - r_{12}^2)s_2]$$
$$= [(0.75 - 0.78(0.30))0.33]/[(1 - 0.30^2)0.40]$$
$$= 0.47$$

และ

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$
$$= 3.5 - 0.01(112.73) - 0.47(3.11)$$
$$= 0.91$$

ในการแปลความหมายความชันและจุดตัด ความชันมีค่า 0.01 สำหรับตัวแปร GRETOT มีความหมายว่า ถ้าคะแนน GRETOT เพิ่มขึ้น 1 หน่วยแล้ว GGPA จะเพิ่มขึ้น 0.01 หน่วย เมื่อควบคุม UPGA ไว้ ทำนองเดียวกัน ความชันมีค่า 0.47 สำหรับตัวแปร UGPA หมายความว่า ถ้า UPGA เพิ่มขึ้น 1 หน่วยแล้ว GGPA จะเพิ่มขึ้น 0.47 หน่วย เมื่อควบคุม GRETOT ไว้ จุดตัดมีค่า 0.91 หมายความว่า ถ้าคะแนน GRETOT และ UGPA มีค่าเป็น 0 แล้ว GGPA จะมีค่า 0.91 อย่างไรก็ตาม มันไม่สำคัญในการอ้างอิง GRETOT เป็น 0 เพราะว่าคะแนนต่ำสุดที่เป็นไปได้คือ 40 คะแนน ทำนองเดียวกัน UGPA เป็น 0 ก็เป็นไปไม่ได้ ไม่งั้นคงไม่สำเร็จการศึกษาระดับปริญญาตรี จากนั้นนำค่าทั้งหมดที่คำนวณได้นำมาใส่สมการการถดถอยพหุคูณ ได้ดังนี้

$$Y_i = b_1X_{1i} + b_2X_{2i} + a + e_i$$

$$Y_i = 0.01X_{1i} + 0.47X_{2i} + 0.91 + e_i$$

ถ้าคะแนน GRETOT มีค่า 130 และ UGPA มีค่า 3.5 แล้ว สามารถทำนายคะแนน GGPA ได้เท่ากับ

$$Y_i = 0.01(130) + 0.47(3.5) + 0.91$$

$$= 3.86$$

ผลของการใช้สมการทำนาย สามารถทำนายเกรดเฉลี่ยระดับบัณฑิตศึกษา GGPA ได้ 3.86

สมการการถดถอยในรูปของคะแนนมาตรฐาน (Standardized Regression Equation)

จากข้างต้นทั้งหมดที่กล่าวมา เป็นการคำนวณการวิเคราะห์การถดถอยพหุคูณเชิงเส้นที่เกี่ยวข้องกับการใช้คะแนนดิบในการคำนวณ ดังนั้นสมการการถดถอยที่ได้จะเรียกว่าเป็นสมการถดถอยในรูปของคะแนนดิบ ความชันจะประมาณค่าได้จากคะแนนดิบ เพราะว่าเป็นการเปลี่ยนแปลงคะแนนดิบใน Y เมื่อค่าคะแนนในตัวแปร X_k เปลี่ยนแปลง 1 หน่วย เมื่อมีการควบคุมตัวแปร X_k อื่น ๆ เอาไว้ มีบางครั้งที่อาจต้องการแสดงการถดถอยในรูปของคะแนนมาตรฐาน z (z-score) มากกว่า ค่าเฉลี่ยและความแปรปรวนของตัวแปรที่อยู่ในรูปของคะแนนมาตรฐาน (z_1 , z_2 และ z_Y) มีค่าเป็น 0 และ 1 ตามลำดับ สมการทำนายเชิงเส้นในรูปของคะแนนมาตรฐานจะกลายเป็น

$$z(Y_i') = \beta_1 z_{1i} + \beta_2 z_{2i} + ... + \beta_m z_{mi}$$

เมื่อ β_k คือความชันในรูปมาตรฐานและเทอมอื่น ๆ ก็ทำนองเดียวกับที่กล่าวไปข้างต้น ในกรณีที่เป็นการถดถอยอย่างง่าย จะไม่มีจุดตัดในสมการทำนายรูปคะแนนมาตรฐานเพราะว่ามีค่าเฉลี่ยของคะแนนมาตรฐานในทุกตัวแปรเป็น 0 ความชันมาตรฐานโดยทั่วไปคำนวณได้ว่า

$$\beta_k = b_k(s_k / s_Y)$$

สำหรับกรณีมีตัวแปรพยากรณ์ 2 ตัว ความชันมาตรฐานสามารถคำนวณได้ว่า

$$\beta_1 = b_1(s_1 / s_Y)$$

หรือ $= (r_{Y1} - r_{Y2}r_{12})/(1 - r_{12}^2)$

และ $\beta_2 = b_2(s_2 / s_Y)$

หรือ $= (r_{Y2} - r_{Y1}r_{12})/(1 - r_{12}^2)$

ถ้า r_{12} = 0 แล้วแสดงว่าตัวแปรพยากรณ์ทั้งสองตัวไม่มีความสัมพันธ์กันแล้ว $\beta_1 = r_{Y1}$ และ $\beta_2 = r_{Y2}$

สำหรับตัวอย่าง คำนวณความชันมาตรฐานได้ค่าเท่ากับ

$$\beta_1 = b_1(s_1 / s_Y)$$

$$= 0.01(16.33/0.33)$$

$$= 0.49$$

และ $\beta_2 = b_2(s_2 / s_Y)$

$$= 0.47(0.40/0.33)$$

$$= 0.57$$

สมการทำนายเขียนได้ว่า

$$z(Y_i') = 0.49z_{1i} + 0.57z_{2i}$$

ความชันมาตรฐานของตัวแปร GRETOT มีค่า 0.49 แปลความหมายเป็นค่าที่ถูกคาดหวังว่าจะเพิ่มขึ้นในตัวแปร GGPA หน่วยคะแนนมาตรฐานเมื่อ GRETOT เพิ่มขึ้น 1 หน่วยคะแนนมาตรฐาน เมื่อควบคุม UGPA ไว้ ทำนองเดียวกันกับตัวแปร UGPA โดยที่ β_k สามารถแปลความหมายเป็นค่าที่ถูกคาดหวังว่า Y จะเปลี่ยนแปลงไปเมื่อมีการเปลี่ยนแปลงใน X_k 1 หน่วยเมื่อมีการควบคุม X_k อื่น ๆ เอาไว้

ในกรณีใดที่ควรใช้การวิเคราะห์การถดถอยในรูปของคะแนนมาตรฐานหรือคะแนนดิบ นักสถิติอย่าง Pedhazur ได้อธิบายถึง β_k จะใช้กับกลุ่มตัวอย่างเฉพาะกลุ่มและจะไม่คงที่หากมีการนำไปใช้กับตัวอย่างต่างกลุ่มกัน อันเนื่องมาจากความแปรปรวนใน X_k ที่เปลี่ยนแปลงไปเมื่อกลุ่มตัวอย่างแตกต่างกัน (เมื่อกลุ่มตัวอย่างมีความแปรปรวนของ X_k เพิ่มขึ้นแล้ว ค่าของ β_k ก็จะเพิ่มขึ้นด้วย) ดังนั้นนักวิจัยโดยมากจะใช้ b_k ในการเปรียบเทียบอิทธิพลของตัวแปรพยากรณ์เมื่อนำไปใช้กับกลุ่มตัวอย่างหรือกลุ่มประชากรที่แตกต่างกัน อย่างไรก็ตาม β_k มีประโยชน์ในการประเมินความสัมพันธ์ระหว่างตัวแปรพยากรณ์แต่ละตัวสำหรับกลุ่มตัวอย่างเฉพาะกลุ่มที่ศึกษา เพราะว่าตัวแปรพยากรณ์แต่ละตัวจะมีสเกลที่แตกต่างกัน ดังนั้น GGPA จะมีความสัมพันธ์กับ GRETOT มากกว่า UGPA ดังแสดงในค่าความชันมาตรฐาน ซึ่งจะผลให้ต้องมีการทดสอบนัยสำคัญของตัวแปรพยากรณ์ทั้งสองตัว

สัมประสิทธิ์การอธิบายและสหสัมพันธ์พหุคูณ (Coefficient of Multiple Determination and Multiple Correlation)

มีคำถามว่า "ตัวแปรพยากรณ์สามารถทำนายตัวแปรเกณฑ์ได้เท่าใด" จากตัวอย่างที่ผ่านมาตัวแปรเกณฑ์คือ GGRE ถูกทำนายด้วย UGRE และ GRETOT ดังนั้นชุดของตัวแปรพยากรณ์มีประโยชน์ในการทำนายตัวแปรเกณฑ์เท่าใด

วิธีการง่าย ๆ สำหรับคำถามนี้คือการแบ่งส่วนความแปรปรวนออกเป็นผลรวมกำลังสองใน Y ซึ่งจะใช้สัญลักษณ์ว่า SS_Y ในการถดถอยพหุคูณเชิงเส้น สามารถคำนวณ SS_Y ได้ว่า

$$SS_Y = [n\Sigma Y^2 - (\Sigma Y)^2]/n$$

หรือ $$= (n - 1)\, s_Y^2$$

เมื่อผลรวมของ Y ตั้งแต่ i = 1, ..., n ถัดมาจะแบ่งส่วน SS_Y ได้เท่ากับ

$$SS_Y = SS_{reg} + SS_{res}$$

หรือ $$\Sigma(Y - \bar{Y})^2 = \Sigma(Y' - \bar{Y})^2 + \Sigma(Y - Y')^2$$

เมื่อ SS_{reg} คือผลรวมกำลังสองอันเนื่องมาจากการถดถอย Y บน X_k (หรือเขียนว่า $SS_{Y'}$) และ SS_{res} คือผลรวมกำลังสองของความคลาดเคลื่อน SS_Y เป็นส่วนที่นำเสนอความแปรปรวนรวมของ Y และ SS_{reg} เป็นส่วนที่นำเสนอความแปรปรวนใน Y ที่ถูกทำนายด้วย X_k และ SS_{res} เป็นที่นำเสนอความแปรปรวนใน Y ที่ไม่สามารถทำนายได้ด้วย X_k

ก่อนหน้านี้ได้อธิบายการคำนวณหา SS_{reg} และ SS_{res} ไปแล้ว ในหัวข้อนี้จะอธิบายถึงสัมประสิทธิ์การอธิบาย ในการวิเคราะห์การถดถอยอย่างง่ายใช้สัญลักษณ์ว่า r_{XY}^2 ในกรณีที่มีตัวแปรพยากรณ์มากกว่าหนึ่งตัว จะใช้สัญลักษณ์ว่า $R_{Y.1,...,m}^2$ ตัวห้อยจะบอกถึงตัวแปรเกณฑ์ Y กับตัวแปรพยากรณ์ที่มีตั้งแต่ X_1, ..., m กระบวนการอย่างง่ายในการคำนวณ R^2 คือ

$$R_{Y.1,...,m}^2 = \beta_1 r_{Y1} + \beta_2 r_{Y2} + ... + \beta_m r_{Ym}$$

สัมประสิทธิ์การอธิบายจะบอกถึงสัดส่วนของความแปรปรวนรวมใน Y ที่สามารถพยากรณ์ได้ด้วยชุดของตัวแปรพยากรณ์ในสมการถดถอยเชิงเส้น อาจเขียนได้ในเทอมของ SS ดังนี้

$$R_{Y.1,...,m}^2 = SS_{reg}/SS_Y$$

ดังนั้นวิธีการหนึ่งในการคำนวณหา SS_{reg} และ SS_{res} จาก R^2 คือ

$$SS_{reg} = R^2 SS_Y$$

และ $$SS_{res} = (1 - R^2)SS_Y$$

$$= SS_Y - SS_{reg}$$

โดยทั่วไป กรณีของการถดถอยเชิงเส้นไม่มีกฎตายตัวว่าควรมีขนาดของสัมประสิทธิ์การอธิบายมากเท่าใดที่จำเป็นในการบอกถึงขนาดของความแปรปรวนที่ถูกทำนายได้อย่างมีความหมาย มีการทดสอบนัยสำคัญอยู่หลายวิธีที่จะอธิบายต่อไป สังเกตว่า $R_{Y.1,...,m}$ อ้างอิงว่าเป็นสัมประสิทธิ์สหสัมพันธ์พหุคูณ

ตัวอย่างข้อมูลในการทำนาย GGPA จาก GRETOT และ UGPA สามารถคำนวณหาค่าของ SS_Y ได้ค่า

$$SS_Y = (n - 1)\, s_Y^2$$
$$SS_Y = (10)0.11$$
$$= 1.1$$

ถ้ดมาคำนวณหา R^2

$$R^2_{Y.12} = \beta_1 r_{Y1} + \beta_2 r_{Y2}$$
$$= 0.49(0.78) + 0.57(0.75)$$
$$= 0.81$$

สามารถใช้ SS_y คำนวณหา SS_{reg} และ SS_{res} ได้

$$SS_{reg} = R^2 SS_Y$$
$$= 0.81(1.1)$$
$$= 0.89$$

และ
$$SS_{res} = (1 - R^2)SS_Y$$
$$= (1 - 0.81)(1.1)$$
$$= 0.21$$

ท้ายสุดสรุปผลสำหรับข้อมูลที่เป็นตัวอย่าง จะพบว่าสัมประสิทธิ์การตัดสินใจจะมีค่าเท่ากับ 0.91 ดังนั้น GRETOT และ UGPA สามารถทำนายความแปรปรวนใน GGPA ได้ 91 เปอร์เซ็นต์ ซึ่งเป็นผลที่ดีที่คะแนนสอบและเกรดเฉลี่ยปริญญาตรีสามารถทำนายความสำเร็จในการศึกษาได้

สังเกตว่า R^2 จะอ่อนไหวต่อขนาดของกลุ่มตัวอย่างและจำนวนของตัวแปรพยากรณ์ ในเทอมของ R จะประมาณค่าความสัมพันธ์พหุคูณของประชากรได้ลำเอียงเนื่องจากความคลาดเคลื่อนในการสุ่มตัวอย่างในความสัมพันธ์ระหว่างสองตัวแปรและในส่วนเบี่ยงเบนมาตรฐานของ X และ Y เพราะว่า R จะประมาณค่าสหสัมพันธ์พหุคูณของประชากรได้สูงเกินความเป็นจริง และสัมประสิทธิ์การอธิบายปรับแก้ (adjusted coefficient of multiple determination) ก็จะถูกนำมาใช้ในการอภิปรายผล ค่า R^2 ปรับแก้สามารถคำนวณได้ด้วยสูตร

$$\text{Adjusted } R^2 = 1 - (1 - R^2)[(n - 1)/(n - m - 1)]$$

ดังนั้นค่า R^2 ปรับแก้จะใช้เมื่อขนาดกลุ่มตัวอย่างมีจำนวน และจำนวนของตัวแปรพยากรณ์มีจำนวนมาก และเปรียบเทียบความเหมาะสมของสมการถดถอยที่ได้ในชุดข้อมูลเดียวกันนี้กับจำนวนตัวแปรพยากรณ์ที่แตกต่างกันและมีข้อมูลของกลุ่มตัวอย่างที่แตกต่างออกไป ความแตกต่างระหว่าง R^2 และ R^2 ปรับแก้จะเรียกว่า shrinkage

เมื่อ n มีจำนวนน้อย จะมีความลำเอียงให้ค่า R^2 มีค่ามาก ในกรณีนี้การปรับแก้จะเข้ามีประโยชน์ โดยใช้ R^2 ปรับแก้ นอกจากนี้กรณีมีกลุ่มตัวอย่างขนาดเล็ก สัมประสิทธิ์การถดถอย

อาจจะประมาณค่าอ้างอิงไปยังประชากรได้ไม่ดี แต่อย่างไรก็ตาม ควรมี n จำนวนมากกว่า m หลายเท่า จะช่วยให้เกิดความลำเอียงน้อยและจะช่วยสรุปอ้างอิงไปยังประชากรได้ดีขึ้น

เมื่อจำนวนตัวแปรพยากรณ์มีมากในการวิเคราะห์การถดถอยพหุคูณ อำนาจการทดสอบจะลดลง และจะไปเพิ่มความคลาดเคลื่อนแบบที่ 1 ในการทดสอบนัยสำคัญ ในกรณีการวิเคราะห์การถดถอยพหุคูณ อำนาจการทดสอบจะเกี่ยวข้องกับขนาดของกลุ่มตัวอย่าง จำนวนของตัวแปรพยากรณ์ ระดับนัยสำคัญและขนาดของอิทธิพลประชากร ไม่มีกฎว่าจะต้องใช้กลุ่มตัวอย่างจำนวนเท่าใดที่จะสัมพันธ์กับจำนวนของตัวแปรพยากรณ์ แต่โดยปกตินักวิจัยจะใช้อัตราส่วนของจำนวน n ต่อ m มาก ๆ

สำหรับข้อมูลจากตัวอย่างนี้ คำนวณหา R^2 ปรับแก้ ได้ค่า

$$\text{Adjusted } R^2 = 1 - (1 - R^2)[(n - 1)/(n - m - 1)]$$
$$= 1 - (1 - 0.81)[(11 - 1)/(11 - 2 - 1)]$$
$$= 0.76$$

บ่งชี้ว่ามีค่าปรับแก้จะมีค่าต่ำกว่าเมื่อเปรียบเทียบกับ R^2

การทดสอบนัยสำคัญ

ในหัวข้อนี้จะอธิบาย 3 วิธีที่ใช้ในการถดถอยพหุคูณ เกี่ยวข้องกับการทดสอบนัยสำคัญของสมการถดถอยทั้งหมด ทดสอบความชันในแต่ละค่า (หรือสัมประสิทธิ์การถดถอย) และการเพิ่มขึ้นของสัดส่วนของความแปรปรวนที่อธิบายได้ด้วยตัวแปรพยากรณ์แต่ละตัว

การทดสอบนัยสำคัญของสมการถดถอยทั้งหมด

การทดสอบแรกเป็นการทดสอบนัยสำคัญของสมการถดถอยทั้งหมดหรือเรียกอีกอย่างว่าการทดสอบนัยสำคัญของสัมประสิทธิ์การอธิบาย เป็นการทดสอบที่จำเป็นในการทดสอบค่า b_k ทั้งหมดที่อยู่ในสมการ สามารถเขียนสมมติฐานศูนย์และสมมติฐานอื่นได้ดังนี้

$$H_0 : \rho^2_{Y.1,\ldots,m} = 0$$
$$H_1 : \rho^2_{Y.1,\ldots,m} \neq 0$$

ถ้า H_0 ถูกปฏิเสธแล้วแสดงว่า มีสัมประสิทธิ์การถดถอยอยู่ 1 ตัวหรือมากกว่า (b_k) อาจจะมีนัยสำคัญทางสถิติแตกต่างจาก 0 อย่างไรก็ตาม เป็นไปได้ที่จะมีนัยสำคัญของ R^2 ทั้งหมดเมื่อไม่มีตัวแปรพยากรณ์แต่ละตัวมีนัยสำคัญ ซึ่งจะบ่งชี้ว่าไม่มีตัวแปรพยากรณ์แต่ละตัวที่มีความแกร่งเหนือตัวแปรอื่น ๆ แต่อย่าลืมว่าตัวแปรพยากรณ์แต่ละตัวอาจจะสัมพันธ์กัน ดังนั้น "ต้องควบคุมตัวแปรพยากรณ์อื่น" ซึ่งเป็นความสำคัญอันดับแรก ถ้า H_0 ไม่ถูกปฏิเสธแต่ไม่มีสัมประสิทธิ์การถดถอยที่มีนัยสำคัญทางสถิติแตกต่างจากศูนย์

การทดสอบอยู่บนพื้นฐานสถิติที่ว่า

$$F = \frac{[R^2/m]}{[(1-R^2)/(n-m-1)]}$$

เมื่อ F คือสถิติทดสอบ F ส่วน R^2 คือสัมประสิทธิ์การอธิบาย (สัดส่วนความแปรปรวนใน Y ที่ถูกทำนายได้ด้วย X_k) ส่วน $1 - R^2$ คือสัมประสิทธิ์การไม่อธิบาย (สัดส่วนความแปรปรวนใน Y ที่ไม่สามารถทำนายได้ด้วย X_k) m คือจำนวนของตัวแปรพยากรณ์ และ n คือขนาดของกลุ่มตัวอย่าง สถิติทดสอบ F จะถูกเปรียบเทียบกับ F วิกฤติที่ได้จากตารางสถิติเป็นแบบทิศทางเดียว โดยมีองศาแห่งความเป็นอิสระเท่ากับ m และ (n - m - 1) ค่าวิกฤติ F จะเท่ากับ $_{(1-\alpha)}F_{m,(n-m-1)}$ สถิติทดสอบสามารถเขียนได้ในรูปหนึ่งว่า

$$F = \frac{(SS_{reg}/df_{reg})}{(SS_{res}/df_{res})}$$

$$= MS_{reg}/MS_{res}$$

เมื่อ $df_{reg} = m$ และ $df_{res} = (n - m - 1)$

สำหรับตัวอย่างข้อมูล สามารถคำนวณสถิติทดสอบได้ค่า

$$F = \frac{[R^2/m]}{[(1-R^2)/(n-m-1)]}$$

$$= \frac{[0.81/2]}{[(1-0.81)/(11-2-1)]}$$

$$= 17$$

หรือ

$$F = \frac{(SS_{reg}/df_{reg})}{(SS_{res}/df_{res})}$$

$$= \frac{(0.89/2)}{(0.21/8)}$$

$$= 17$$

ค่าวิกฤติที่ระดับนัยสำคัญ 0.05 และ $_{0.95}F_{2,8} = 4.46$ สถิติทดสอบมีค่ามากกว่าค่าวิกฤติ จะปฏิเสธ H_0 และสรุปผลว่า ρ^2 ไม่เท่ากับ 0 ที่ระดับนัยสำคัญ 0.05 นั่นคือ GRETOT และ UGPA ร่วมกันทำนายสัดส่วนความแปรปรวนใน GGPA ได้อย่างมีนัยสำคัญทางสถิติ

การทดสอบนัยสำคัญของ b_k

การทดสอบที่สองเป็นการทดสอบนัยสำคัญทางสถิติของความชันและสัมประสิทธิ์การถดถอย b_k ในอีกกรณีหนึ่ง สัมประสิทธิ์การถดถอยที่อยู่ในรูปคะแนนดิบมีนัยสำคัญทางสถิติแตกต่างจากศูนย์หรือไม่ ซึ่งจะเหมือนกับการทดสอบ β_k แต่จะไม่แสดงการทดสอบแยกส่วนออกมา สมมติฐานศูนย์และสมมติฐานอื่นสามารถเขียนได้ดังนี้

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

เมื่อ β_k คือความชันของประชากรในตัวแปร X_k

การถดถอยพหุคูณจำเป็นที่จะต้องคำนวณความคลาดเคลื่อนมาตรฐานสำหรับแต่ละ b_k ซึ่งสามารถคำนวณได้ดังนี้

$$s^2_{res} = SS_{res}/df_{res} = MS_{res}$$

เมื่อ df_{res} = (n - m - 1) องศาแห่งความเป็นอิสระจะสูญหายไปเพราะมีการประมาณค่าความชันของประชากรและจุดตัด นั่นคือ β_k และ α ตามลำดับ จากข้อมูลตัวอย่าง ความคลาดเคลื่อนของความแปรปรวนของการประมาณค่าหาได้จากปริมาณของความแปรปรวนในความคลาดเคลื่อน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าได้โดยการถอดรากที่สองของความความคลาดเคลื่อนของความแปรปรวนของการประมาณค่า และสามารถได้โดยผ่านความเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนของการประมาณค่า เรียกว่า ความคลาดเคลื่อนมาตรฐานของการประมาณค่า ใช้สัญลักษณ์ว่า s_{res}

ท้ายสุดจะคำนวณความคลาดเคลื่อนมาตรฐานของ b_k แต่ละตัว ใช้สัญลักษณ์ความคลาดเคลื่อนมาตรฐานของ b_k ว่า $s(b_k)$ คำนวณได้ดังนี้

$$s(b_k) = s_{res} / \sqrt{[(n-1)s^2_k(1-R^2_k)]}$$

เมื่อ s_k คือความแปรปรวนของกลุ่มตัวอย่างสำหรับตัวแปรทำนาย X_k และ R^2_k คือกำลังสองของสหสัมพันธ์พหุคูณระหว่าง X_k ตัวหนึ่งกับ X_k ที่เหลือ R^2_k จะแสดงถึงการซ้อนทับกันระหว่างตัวแปรพยากรณ์ (X_k) ในกรณีที่มีตัวแปรพยากรณ์ 2 ตัว R^2_k จะเท่ากับ r^2_{12}

ต่อมาเป็นการทดสอบทางสถิติสำหรับการทดสอบนัยสำคัญของ b_k ในการทดสอบทางสถิตินี้จะเป็นอัตราส่วนของค่าสัมประสิทธิ์การถดถอยหารด้วยความคลาดเคลื่อนมาตรฐาน ดังสูตร

$$t = b_k / s(b_k)$$

สถิติทดสอบ t จะนำไปเปรียบเทียบกับค่าวิกฤติแบบสองทางกับระดับนัยสำคัญที่ (n - m - 1) โดยเปิดจากตารางสถิติ ว่าวิกฤติจะเท่ากับ $\pm_{(\alpha/2)}t_{(n-m-1)}$ กรณีแบบสองทาง

ในการหาช่วงความเชื่อมั่นของ b_k โดยการนำ b_k มาบวกและลบออกจากค่าวิกฤติที่เปิดจากตารางแล้วคูณด้วยความคลาดเคลื่อนมาตรฐาน ช่วงความเชื่อมั่นของ b_k คำนวณได้ด้วยสูตร

$$CI(b_k) = b_k \pm {}_{(\alpha/2)}t_{(n-m-1)}s(b_k)$$

จำได้ว่าสมมติฐานศูนย์คือ β_k เท่ากับ 0 ($H_0 : \beta_k = 0$) ดังนั้น ถ้าช่วงความเชื่อมั่นคร่อมศูนย์แล้ว b_k จะไม่มีนัยสำคัญทางสถิติแตกต่างจากศูนย์ที่ระดับนัยสำคัญ นั่นคือแปลความหมายได้ว่า เมื่อเก็บข้อมูลซ้ำกับกลุ่มตัวอย่างหลาย ๆ กลุ่ม จะมีจำนวน (1 - α) เปอร์เซ็นต์ ที่ค่า β_k จะตกอยู่ในช่วงความเชื่อมั่นนี้

รูปแบบโดยทั่วไปของการทดสอบที่สองนี้จะแสดงเป็นสมมติฐานได้ว่า

$$H_0 : \beta_k = \beta_0$$

$$H_1 : \beta_k \neq \beta_0$$

เมื่อ β_0 คือค่าที่ β_k ถูกสมมติให้เท่ากับ เมื่อ $\beta_0 = 0$ แล้ว ก็จะทดสอบเหมือนกับที่กล่าวไว้ข้างต้น รูปแบบการทดสอบโดยทั่วไปของ b_k คือ

$$t = (b_k - \beta_0)/s(b_k)$$

ในรูปแบบโดยมากของช่วงความเชื่อมั่น ก็จะรวม β_0 ไว้ในด้วย

สำหรับตัวอย่างข้อมูล มีสมมติฐานว่า $\beta_k = 0$ และเป็นแบบสองทาง ก่อนดำเนินการทดสอบนัยสำคัญต้องคำนวณความคลาดเคลื่อนของความแปรปรวนเสียก่อน ได้ค่าดังนี้

$$s^2_{res} = SS_{res}/df_{res} = MS_{res}$$
$$= 0.21/8$$
$$= 0.026$$

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า s_{res} จะคำนวณได้เท่ากับ 0.11 ถัดมาคำนวณความคลาดเคลื่อนมาตรฐานของ b_k ได้ค่า

$$s(b_1) = s_{res}/\sqrt{[(n-1)s_1^2(1-R_{12}^2)]}$$
$$= 0.16/\sqrt{[(10)266.82(1-0.30^2)]}$$
$$= 0.003$$

และ
$$s(b_2) = s_{res}/\sqrt{[(n-1)s_2^2(1-R_{12}^2)]}$$
$$= 0.16/\sqrt{[(10)0.16(1-0.30^2)]}$$
$$= 0.1326$$

สุดท้ายคำนวณสถิติทดสอบ t ได้ค่า

$$t_1 = b_1/s(b_1)$$
$$= 0.01/0.003$$
$$= 3.33$$
$$t_2 = b_2/s(b_2)$$
$$= 0.47/0.1326$$
$$= 3.54$$

ประเมินสมมติฐานโดยการนำค่าสถิติทดสอบที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤติ $\pm_{.025}t_8 = \pm 2.306$ สถิติทดสอบ t ทั้งสองค่ามีค่ามากกว่าค่าวิกฤติ แสดงว่าจะปฏิเสธ H_0 ยอมรับ H_1 ในตัวแปรพยากรณ์ทั้งคู่ เราจะสรุปได้ว่าความชันมีนัยสำคัญแตกต่างจาก 0 ที่ระดับนัยสำคัญ 0.05

สุดท้ายคำนวณหาช่วงความเชื่อมั่นสำหรับ b_k ได้ค่า

$$CI(b_1) = b_1 \pm {}_{(\alpha/2)}t_{(n-m-1)}s(b_1)$$
$$= b_1 \pm {}_{.025}t_8\, s(b_1)$$
$$= 0.01 \pm 2.306(0.002)$$

$$= \quad (0.007,\ 0.02)$$

และ $\quad CI(b_2) = b_2 \pm {}_{(\alpha/2)}t_{(n-m-1)}s(b_2)$

$$= \quad b_2 \pm {}_{.025}t_8\ s(b_2)$$

$$= \quad 0.47 \pm 2.306(0.09)$$

$$= \quad (0.25,\ 0.68)$$

สังเกตว่าช่วงความเชื่อมั่นจะไม่คร่อม 0 ซึ่งเป็นค่าเฉพาะในสมมติฐาน H_0 แล้วจะสรุปได้ว่า ทั้งคู่ของ b_k จะมีนัยสำคัญแตกต่างจากศูนย์ที่ระดับ 0.05

การทดสอบการเพิ่มขึ้นของสัดส่วนความแปรปรวนที่สามารถอธิบายได้

การทดสอบที่สาม เป็นการทดสอบการเพิ่มขึ้นของสัดส่วนความแปรปรวนที่สามารถอธิบายได้ได้ด้วยตัวแปรพยากรณ์แต่ละตัว ตัวอย่างที่เป็นโมเดลมีตัวพยากรณ์ 2 ตัว สามารถทดสอบการเพิ่มขึ้นได้ เป็นสัดส่วนของความแปรปรวนที่สามารถอธิบายได้ด้วยตัวแปร 2 ตัว เปรียบเทียบกับตัวแปร 1 ตัว ซึ่งเป็นความจำเป็นในกรณีที่มีโมเดลตัวแปรพยากรณ์ 2 ตัวกับโมเดลทีมีตัวแปรพยากรณ์ 1 ตัว การทดสอบการเพิ่มขึ้นของ X_1 มีสถิติทดสอบคือ

$$F = \frac{[(R^2_{Y.12} - R^2_{Y.2})/(m_2 - m_1)]}{[(1 - R^2_{Y.12})/(n - m_2 - 1)]}$$

เมื่อ $R^2_{Y.2} = r^2_{Y2}$ แล้ว m_2 คือจำนวนของตัวแปรพยากรณ์ในโมเดลที่มีตัวพยากรณ์ 2 ตัว และ m_1 คือจำนวนของตัวแปรพยากรณ์ในโมเดลที่มีตัวพยากรณ์ 1 ตัว สถิติ F จะนำไปเปรียบเทียบกับ F วิกฤติที่เปิดจากตารางแบบทิศทางเดียว ใช้สัญลักษณ์ว่า ${}_{(1-\alpha)}F_{(m2-m1,n-m2-1)}$

ทดสอบการเพิ่มขึ้นของ X_2 ใช้สถิติทดสอบ

$$F = \frac{[(R^2_{Y.12} - R^2_{Y.1})/(m_2 - m_1)]}{[(1 - R^2_{Y.12})/(n - m_2 - 1)]}$$

เมื่อ $R^2_{Y.1} = r^2_{Y1}$ แล้ว m_2 คือจำนวนของตัวแปรพยากรณ์ในโมเดลที่มีตัวพยากรณ์ 2 ตัว และ m_1 คือจำนวนของตัวแปรพยากรณ์ในโมเดลที่มีตัวพยากรณ์ 1 ตัว สถิติ F จะนำไปเปรียบเทียบกับ F วิกฤติที่เปิดจากตารางแบบทิศทางเดียว ใช้สัญลักษณ์ว่า ${}_{(1-\alpha)}F_{(m2-m1,n-m2-1)}$

โดยทั่วไป สามารถเปรียบเทียบโมเดลการถดถอย 2 โมเดลสำหรับกลุ่มตัวอย่างเดียวกัน เมื่อโมเดลเต็มรูปแบบ (full model) นิยามว่ามีตัวแปรพยากรณ์ทุกตัวอยู่ในโมเดล และโมเดลที่ลดรูป (reduced model) นิยามว่าเป็นโมเดลที่มีเพียงบางชุดของตัวแปรพยากรณ์ในโมเดล นั่นคือสำหรับโมเดลลดรูปอาจจะมีตัวแปรพยากรณ์เพียง 1 ตัวหรือมากกว่า ซึ่งเป็นส่วนหนึ่งของโมเดลเต็มรูปแบบที่มีการลดตัวแปรพยากรณ์ลงบางตัว สถิติทดสอบโดยทั่วไปเขียนได้ว่า

$$F = \frac{[(R^2_{full} - R^2_{red})/(m_{full} - m_{red})]}{[(1 - R^2_{full})/(n - m_{full} - 1)]}$$

เมื่อ "full" และ "red" เป็นสัญลักษณ์ของโมเดลเต็มรูปแบบ (full model) และโมเดลลดรูป (reduced model) ส่วน m_{full} คือจำนวนของตัวแปรพยากรณ์ในโมเดลเต็มรูปแบบ และ m_{red} คือจำนวนของตัวแปรพยากรณ์ในโมเดลลดรูป สถิติ F จะนำไปเปรียบเทียบกับ F วิกฤติที่เปิดจากตารางแบบทิศทางเดียว ใช้สัญลักษณ์ว่า $_{(1-\alpha)}F_{(mfull-mred,n-mfull-1)}$

จากตัวอย่างข้อมูล สามารถทดสอบการเพิ่มขึ้นของ X_1 (GRETOT) ได้ดังนี้

$$F = \frac{[(R^2_{Y.12} - R^2_{Y.2})/(m_2 - m_1)]}{[(1 - R^2_{Y.12})/(n - m_2 - 1)]}$$

$$= \frac{[(0.81 - 0.75^2)/(2-1)]}{[(1-0.81)/(11-2-1)]}$$

$$= 10.42$$

สถิติ F จะนำไปเปรียบเทียบกับ F วิกฤติ ซึ่งจะได้ค่า $_{.95}F_{1,8} = 5.32$ การทดสอบการเพิ่มขึ้นของ X_2 (UGPA) ได้ดังนี้

$$F = \frac{[(R^2_{Y.12} - R^2_{Y.1})/(m_2 - m_1)]}{[(1 - R^2_{Y.12})/(n - m_2 - 1)]}$$

$$= \frac{[(0.81 - 0.78^2)/(2-1)]}{[(1-0.81)/(11-2-1)]}$$

$$= 8.488$$

การทดสอบที่สองนี้นำค่า F ที่คำนวณได้ไปเปรียบเทียบค่า F วิกฤติ ดังนั้นเราจะสรุปผลได้ว่า เมื่อรวมตัวแปรพยากรณ์ทั้งสองตัว (GRETOT หรือ UGPA) ถูกเพิ่มเข้าไปอย่างมีนัยสำคัญทางสถิติในการอธิบายความแปรปรวนใน GGPA ได้มากกว่ามีตัวแปรเดียว เพราะว่า $t^2 = F$ ดังนั้นผลการทดสอบนี้สามารถทำไปเปรียบเทียบกับค่า t ที่ทดสอบสัมประสิทธิ์การถดถอยได้

ข้อตกลงเบื้องต้นทางสถิติ

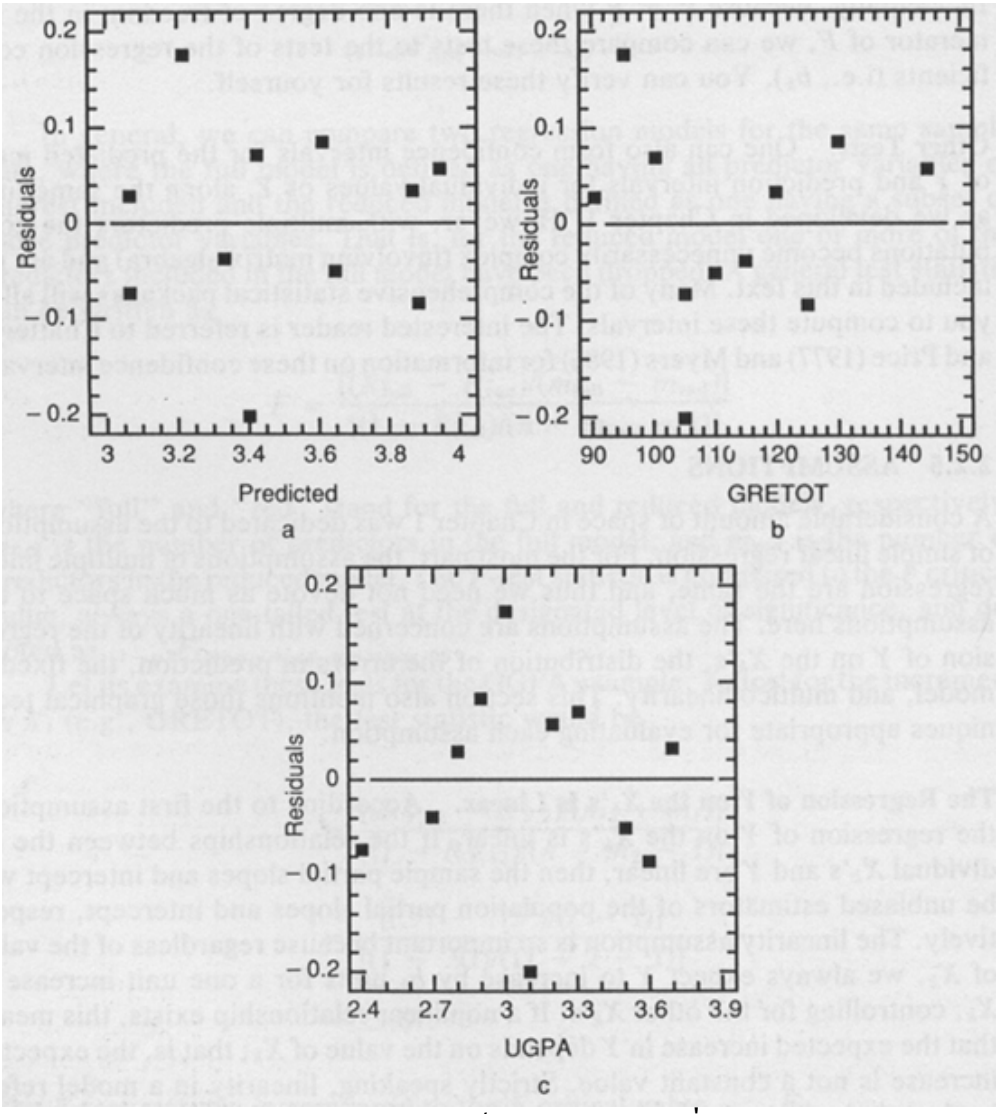
ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุคุณจะเหมือนกับข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยอย่างง่าย แต่มีบางข้อตกลงเบื้องต้นที่แตกต่างออกไป ดังนี้

การถดถอย Y บน X_k เป็นเชิงเส้นตรง

ข้อตกลงเบื้องต้นข้อแรก คือการถดถอย Y บน X_k เป็นเชิงเส้นตรง ถ้าความสัมพันธ์ระหว่าง X_k แต่ละตัวกับ Y เป็นเชิงเส้นตรงแล้ว ความชันและจุดตัดของกลุ่มตัวอย่างจะเป็นตัวประมาณค่าที่ไม่ลำเอียงของความชันและจุดตัดของประชากร ข้อตกลงเบื้องต้นนี้มีความสำคัญเพราะว่าค่าของ X_k โดยปกติจะคาดหวังว่า Y เพิ่มขึ้นเท่ากับ b_k หน่วยเมื่อ X_k เพิ่มขึ้น 1 หน่วย โดยการควบคุมตัวแปร X_k อื่น ๆ ให้คงที่ ถ้าความสัมพันธ์ไม่เป็นเชิงเส้นตรงแล้ว การเพิ่มขึ้นของ Y ที่ขึ้นอยู่กับค่าของ X_k นั้นจะไม่คงที่เสมอไป

การละเมิดข้อตกลงเบื้องต้นสามารถถูกค้นพบผ่านการพล็อตความคลาดเคลื่อนของ e กับ X_k แต่ละตัว และ e กับ Y' (อีกกรณีหนึ่งคือการพล็อต Y กับ X_k และ Y กับ Y') ความคลาดเคลื่อนควรจะมีตำแหน่งภายในแกนที่มีค่าภายใน $\pm 2s_{res}$ ในภาพประกอบ 1 ในภาพ a เมื่อพล็อต e กับ Y' ภาพ b พล็อต e กับ GRETOT และภาพ c พล็อต e กับ UGPA ในแต่ละคู่จะมีกลุ่มตัวอย่างน้อยมาก จะเห็นเป็นแบบแผนอย่างสุ่มของความคลาดเคลื่อนในแต่ละภาพ และจะเห็นว่าแต่ละจุดอยู่ใกล้กับเส้นไม่เกินช่วงความเชื่อมั่นที่กำหนดเป็นไปตามข้อตกลงเบื้องต้นความเป็นเส้นตรงระหว่างตัวแปร

มีอยู่ 2 วิธีที่สามารถจะใช้ได้เมื่อความสัมพันธ์ไม่เป็นเชิงเส้นตรง (nonlinearity) โดยใช้การแปลงรูปและโมเดลเชิงเส้นโค้ง



ภาพประกอบ 1 พล็อตความคลาดเคลื่อน

การแจกแจงของความคลาดเคลื่อนในการทำนาย

ข้อตกลงเบื้องต้นประการที่สองจะเกี่ยวกับรูปแบบของความคลาดเคลื่อนในการทำนาย หรือก็คือ e_i มีอยู่ 4 ประการที่จะอธิบายในที่นี้ ***ประการแรก*** ความคลาดเคลื่อนในการทำนายจะถูกสมมติว่าเป็นไปอย่างสุ่มและเป็นอิสระจากกัน นั่นคือไม่มีความคลาดเคลื่อนอย่างเป็นระบบเกิดขึ้นและความคลาดเคลื่อนแต่ละค่าที่เกิดขึ้นจะเป็นอิสระจากความคลาดเคลื่อนอื่น

วิธีการที่ง่ายที่สุดสำหรับการประเมินความเป็นอิสระของความคลาดเคลื่อนคือการนำมาพล็อตเป็นแผนภาพ ถ้าข้อตกลงเบื้องต้นของความเป็นอิสระเป็นจริงแล้ว ความคลาดเคลื่อนควรจะตกลงอย่างสุ่มในแผนภาพ ถ้าข้อตกลงเบื้องต้นถูกละเมิดแล้วความคลาดเคลื่อนจะตกลงเป็นกระจุกอยู่รวมกัน สถิติ Durbin-Watson สามารถใช้ในการตรวจสอบ autocorrelation ได้ การละเมิดความเป็นอิสระของความคลาดเคลื่อนเกิดขึ้นได้ 3 กรณีคือ ข้อมูลเป็นอนุกรมเวลา ค่าสังเกตถูกจัดเป็นบล็อก และมีการวัดซ้ำ ความไม่เป็นอิสระจะมีอิทธิพลต่อความคลาดเคลื่อนมาตรฐานของโมเดลการถดถอย วิธีการแก้ไขความไม่เป็นอิสระของความคลาดเคลื่อนคือใช้วิธี generalized หรือ weighted least squares ในการประมาณค่าพารามิเตอร์ หรือใช้วิธีการแปลงรูป

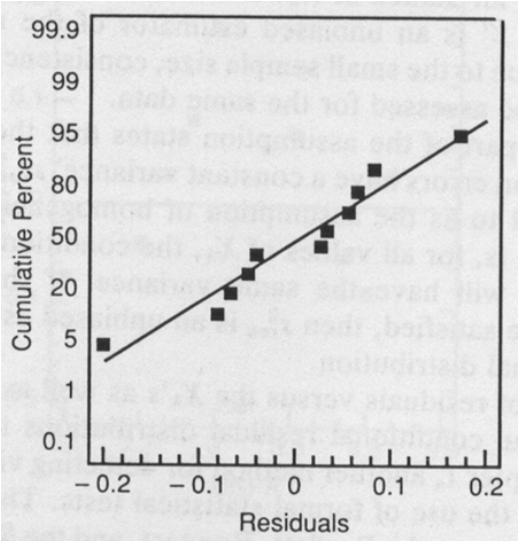
ประการที่สอง การแจกแจงอย่างมีเงื่อนไขของความคลาดเคลื่อนของการทำนายในทุกค่าของ X_k จะมีค่าเฉลี่ยเป็นศูนย์ ถ้าข้อตกลงเบื้องต้น 2 ประการแรกนี้เป็นจริงแล้ว Y' จะเป็นตัวประมาณค่าที่ไม่ลำเอียงของค่าเฉลี่ย

ประการที่สาม การแจกแจงอย่างมีเงื่อนไขของความคลาดเคลื่อนของการทำนายมีความแปรปรวนคงที่ s^2_{res} สำหรับทุกค่าของ X_k ซึ่งข้อตกลงเบื้องต้นนี้มักจะเรียกอีกอย่างว่าความเป็นเอกพันธ์ของความแปรปรวน นั่นคือทุกค่าของ X_k การแจกแจงแบบมีเงื่อนไขของความคลาดเคลื่อนของการทำนายจะมีความแปรปรวนเท่าเทียมกัน ถ้าข้อตกลงเบื้องต้นสามประการแรกนี้เป็นจริงแล้ว s^2_{res} จะเป็นตัวประมาณค่าที่ไม่ลำเอียงของความแปรปรวนสำหรับการแจกแจงแบบมีเงื่อนไข

ในการพล็อตความคลาดเคลื่อนกับ X_k และกับ Y' ความสอดคล้องของความแปรปรวนของการแจกแจงแบบมีเงื่อนไขของความคลาดเคลื่อนอาจจะถูกตรวจสอบได้ มีอีกวิธีหนึ่งในการค้นหาการละเมิดความเป็นเอกพันธ์ของความแปรปรวนโดยการใช้สถิติทดสอบ ก็มีทั้ง Hartley's F_{max} หรือ Cochran's test หรือ Bartlett-Box test และอื่น ๆ การละเมิดความเป็นเอกพันธ์ของความแปรปรวนอาจจะนำไปสู่การมากเกินไปของความคลาดเคลื่อนมาตรฐาน หรือการแจกแจงแบบมีเงื่อนไขไม่เป็นโค้งปกติ วิธีการแก้ไขนั้นก็ทำนองเดียวกันคือใช้การแปลงรูปข้อมูล (เช่น $\sqrt{Y}$ หรือ Log Y) อาจใช้วิธีการประมาณค่าแบบ generalized หรือ weighted least squares

ประการที่สี่ คือการแจกแจงแบบมีเงื่อนไขของความคลาดเคลื่อนของการทำนายมีการแจกแจงแบบปกติ สำหรับทุกค่าของ X_k ความคลาดเคลื่อนในการทำนายจะมีการแจกแจงเป็นโค้งปกติ การละเมิดข้อตกลงเบื้องต้นนี้อาจจะมีผลมาจากการเกิดค่าผิดปกติ (outliers) อาจจะค้นหา

โดยการแจกแจงความถี่ พล็อตกราฟ หรือคำนวณการวัดการกระจายของข้อมูล ซึ่งตัวอย่างข้างต้นสามารถนำมาแสดงได้ดังภาพประกอบ 2



ภาพประกอบ 2 พล็อตความน่าจะเป็นปกติ

ตัวแปร X เป็นโมเดลกำหนด

ข้อตกลงเบื้องต้นที่สามคือค่าของ X_k เป็นแบบกำหนด นั่นคือ X_k จะเป็นตัวแปรกำหนดไม่ใช่ตัวแปรสุ่ม ผลนี้ในโมเดลการถดถอยจะเที่ยงตรงเฉพาะสำหรับค่าของ X_k ที่ถูกสังเกตและใช้ในการวิเคราะห์ ดังนั้นค่าเดียวกันของ X_k ควรจะถูกใช้ในกลุ่มตัวอย่างที่ซ้ำกัน นั่นหมายความว่าการวิจัยเชิงทดลอง ผู้วิจัยควรจะกำหนดค่าของ X_k ในการออกแบบและการวิจัยที่ไม่ใช่เชิงทดลอง ผู้วิจัยควรจะเลือกค่าของ X_k มาจากประชากรค่าของ X_k และควรจะเลือกมาอย่างเจาะจงสำหรับการศึกษา แนวคิดทำนองเดียวกันคือจะแสดงในการวิเคราะห์อิทธิพลสุ่ม (random effects) และอิทธิพลกำหนด (fixed effects) ในเรื่องการวิเคราะห์ความแปรปรวน

Multicollinearity

ข้อตกลงเบื้องต้นประการสุดท้าย คือ Multicollinearity เป็นความสัมพันธ์เชิงเส้นระหว่างตัวแปรพยากรณ์ตั้งแต่ 2 ตัวขึ้นไป ซึ่งการมีอยู่ของความสัมพันธ์ระหว่างตัวแปรพยากรณ์จะมีปัญหาในหลายลักษณะ ประการแรก จะนำไปสู่การไม่คงที่ของสัมประสิทธิ์การถดถอย ทั้งขนาดและเครื่องหมาย (อาจเป็นได้ทั้งบวกและลบ) ทั้งนี้เพราะความคลาดเคลื่อนมารฐานของสัมประสิทธิ์การถดถอยมีมาก ดังนั้นจะนำไปสู่ความยากในการมีนัยสำคัญทางสถิติ นอกจากนี้ผลอาจจะเกี่ยวข้องกับการมีนัยสำคัญของ R^2 รวม แต่ไม่มีผลกับตัวแปรพยากรณ์แต่ละตัวที่มีนัยสำคัญแตกต่างจาก 0 ความแปรปรวนของสัมประสิทธิ์การถดถอยมีแนวโน้มจะมีค่ามาก ซึ่ง Multicollinearity จะมีผลโดยทั่วไปของการประมาณค่าในโมเดลการถดถอย

จำได้ว่าสัมประสิทธิ์การถดถอยแบบแยกส่วน จะเป็นการควบคุมตัวแปรพยากรณ์อื่น ๆ ให้คงที่ ในกรณีที่เกิด Multicollinearity ตัวแปรอื่น ๆ ไม่สามารถจะคงที่ได้เพราะว่ามีความสัมพันธ์กันสูงมาก ๆ ซึ่ง Chatterjee และ Price (1977) อธิบายว่า Multicollinearity อาจจะบ่งชี้ถึงการเปลี่ยนแปลงอย่างมากในสัมประสิทธิ์ที่ถูกประมาณค่า เนื่องมาจาก 1) ตัวแปรถูกเพิ่มหรือลด และ/หรือ 2) ค่าสังเกตถูกเพิ่มหรือลด Multicollinearity จะเกิดขึ้นได้เมื่อตัวแปรพยากรณ์เป็นตัวแปรที่แสดงถึงองค์ประกอบของตัวแปรพยากรณ์อื่น ๆ (เช่น GRETOT ประกอบด้วย GREQ และ GREV)

ในการค้นหาการละเมิดข้อตกลงเบื้องต้นนี้ วิธีการที่ง่ายที่สุดคือใช้ชุดของการวิเคราะห์การถดถอย อย่างในตัวอย่างข้างต้นนี้มีตัวแปรพยากรณ์ 3 ตัว ต้องวิเคราะห์การถดถอย 3 ครั้ง คือ 1) ถดถอยตัวแปรพยากรณ์ตัวแรก X_1 บนตัวแปรพยากรณ์อีก 2 ตัวที่เหลือ (X_2 และ X_3) 2) ถดถอยตัวแปรพยากรณ์ตัวที่สอง X_2 บนตัวแปรพยากรณ์อีก 2 ตัวที่เหลือ (X_1 และ X_3) และ 3) ถดถอยตัวแปรพยากรณ์ตัวที่สาม X_3 บนตัวแปรพยากรณ์อีก 2 ตัวที่เหลือ(X_1 และ X_2) ถ้าผลของ R_k^2 มีค่าเข้าใกล้ 1 (เกณฑ์คือมากกว่าหรือเท่ากับ 0.9) แล้วอาจจะเกิดปัญหา Multicollinearity อย่างไรก็ตาม ค่า R^2 ที่มากเนื่องมาจากกลุ่มตัวอย่างขนาดเล็กก็เป็นได้ การเก็บรวบรวมข้อมูลให้มากขึ้นก็จะมีประโยชน์ สำหรับตัวอย่างข้อมูลข้างต้น R_{12}^2 = 0.09 และไม่เกิด Multicollinearity

ตาราง 2 ข้อตกลงเบื้องต้นและผลของการละเมิดข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอยพหุคูณ

ข้อตกลงเบื้องต้น	***ผลการละเมิดข้อตกลงเบื้องต้น***
1. การถดถอย Y บน X_k เป็นเชิงเส้น	เกิดความลำเอียงในความชันและจุดตัด
2. ความคลาดเคลื่อนเป็นอิสระจากกัน	มีอิทธิพลของความคลาดเคลื่อนมาตรฐานของโมเดล
3. ความคลาดเคลื่อนมีค่าเฉลี่ยเป็น 0	มีความลำเอียงใน Y'
4. ความเป็นเอกพันธ์ของความแปรปรวนของความคลาดเคลื่อน	มีความลำเอียงใน s_{res}^2
5. ความคลาดเคลื่อนมีการแจกแจงเป็นโค้งปกติ	มีความเที่ยงตรงต่ำในค่าของความชันและสัมประสิทธิ์การอธิบาย
6. ค่าของ X_k เป็นแบบกำหนด	ความคลาดเคลื่อนในการทำนายมีค่าสูง มีความลำเอียงในความชันและจุดตัด
7. ไม่เกิด Multicollinearity ระหว่าง X_k	สัมประสิทธิ์การถดถอยไม่คงที่ R^2 รวม อาจมีนัยสำคัญในขณะที่ตัวแปรพยากรณ์แต่ละตัวอาจไม่มีนัยสำคัญ มีข้อจำกัดในการสรุปอ้างอิงของโมเดล

มีอีกสถิติหนึ่งสำหรับการค้นหาการเกิด Multicollinearity คือการใช้ variance inflation factor (VIF) สำหรับตัวแปรพยากรณ์แต่ละตัว หรือมีค่าเท่ากับ $1/(1 - R_k^2)$ ค่า VIF นี้จะเป็นค่าที่บ่งบอกถึงโอกาสของสัมประสิทธิ์การถดถอยแต่ละค่าที่จะสัมพันธ์กันเอง Wetherill (1986) แนะนำว่า ค่าที่มากที่สุดของ VIF ควรจะไม่เกิน 10 จึงถือว่าไม่เกิด Multicollinearity

วิธีการแก้ไขปัญหาเมื่อมีการละเมิดข้อตกลงเบื้องต้นนี้คือ ประการแรก ขจัดตัวแปรพยากรณ์ที่สัมพันธ์กันออกตัวหนึ่ง ประการที่สอง ใช้เทคนิคการถดถอย ridge ประการที่สาม นำตัวแปรพยากรณ์มาวิเคราะห์องค์ประกอบ ประการที่สี่ ใช้การแปลงรูปตัวแปร และสุดท้ายเป็นตัวเลือกที่ควรเลือกสุดท้ายคือใช้การถดถอยอย่างง่ายกับตัวแปรพยากรณ์ทีละตัว

กระบวนการเลือกตัวแปรเข้าสมการ

โมเดลที่มีตัวแปรพยากรณ์หลายตัวเราต้องมีการพิจารณาว่าตัวแปรใดที่สามารถเข้าไปทำนายในสมการได้ นั่นคือตัวแปรพยากรณ์ทั้งหมดถูกเลือกเข้าในสมการ พารามิเตอร์ทั้งหมดถูกประมาณค่า ชุดของตัวแปรพยากรณ์ที่ถูกเลือกเอาไว้แล้วจะเป็นโมเดลที่มีตัวแปรพยากรณ์เข้าสมการ (entered) หรือถูกเลือกเข้าสมการ (selected) แต่มีบางกรณีที่มีตัวแปรพยากรณ์ที่ไม่ได้ถูกเลือกเข้าสมการ ซึ่งโมเดลที่มีตัวแปรเข้าสมการหรือไม่มีตัวแปรเข้าสมการจะเรียกว่ากระบวนการเลือกตัวแปร (variable selection procedures) ซึ่งมีอยู่หลายวิธีในการเลือกตัวแปรทั้งแบบ backward elimination, forward selection, stepwise selection และ ชุดย่อยของตัวแปรพยากรณ์ (all possible subsets regression) ซึ่งกระบวนการทั้งหมดนี้จะเกี่ยวโยงไปถึงข้อตกลงเบื้องต้นของการเกิด Multicollinearity ด้วย

วิธีแรก จะอธิบายวิธี backward elimination ซึ่งเป็นกรณีที่ตัวแปรถูกขจัดออกจากโมเดลโดยเลือกตัวแปรที่ทำนายตัวแปรเกณฑ์ได้น้อยที่สุด ในขั้นตอนแรกของการวิเคราะห์ ตัวแปรพยากรณ์ทั้งหมดจะถูกรวมอยู่ในโมเดล ในขั้นตอนที่สอง ตัวแปรพยากรณ์จะถูกลบออกจากโมเดล โดยเลือกตัวแปรที่อธิบายความแปรปรวนในตัวแปรเกณฑ์ได้น้อยที่สุด ซึ่งการขจัดออกนี้อาจจะเลือกตัวแปรที่มีค่า t หรือ F ต่ำสุดที่ไม่มีนัยสำคัญ หรือคือตัวแปรพยากรณ์ที่ถูกขจัดออกจะเป็นตัวแปรที่พยากรณ์ Y ได้น้อยที่สุดและไม่มีนัยสำคัญ และวิเคราะห์เช่นนี้ต่อเนื่องจนกระทั่งเหลือตัวแปรพยากรณ์ในโมเดลที่สามารถทำนายตัวแปรเกณฑ์ได้อย่างมีนัยสำคัญ อาจเปรียบเทียบค่าสถิติ t หรือ F ของตัวแปรพยากรณ์แต่ละตัว ในโปรแกรมคอมพิวเตอร์บางโปรแกรมสามารถเลือกเกณฑ์การขจัดออกได้โดยกำหนดเป็นค่า F สูงสุดในการขจัดออก โดยโปรแกรมจะเลือกตัวแปรพยากรณ์ที่มีค่า F น้อยกว่าค่า F ที่เป็นเกณฑ์ในการขจัดออก โดยตัวแปรพยากรณ์ที่เหลืออยู่ในโมเดลจะมีค่า F สูงกว่าค่า F ที่เป็นเกณฑ์

วิธีถัดมาคือ forward selection ตัวแปรจะถูกเพิ่มหรือถูกเลือกเข้าโมเดลบนพื้นฐานของความสามารถในการพยากรณ์ตัวแปรเกณฑ์ได้สูงสุด ในขั้นแรกของการวิเคราะห์จะไม่มีตัวแปรพยากรณ์ใดอยู่ในโมเดล ในขั้นตอนที่สอง ตัวแปรพยากรณ์จะถูกเพิ่มเข้าไปในโมเดลโดยเลือกตัว

แปรพยากรณ์ที่สามารถอธิบายตัวแปรเกณฑ์ได้สูงที่สุด คือตัวแปรที่มีค่าสถิติ t หรือ F สูงที่สุด และมีนัยสำคัญทางสถิติ หรือคือตัวแปรพยากรณ์ที่ถูกเลือกถัดมาจะสามารถพยากรณ์ Y ได้สูงสุด วิเคราะห์ต่อเนื่องจนกระทั่งตัวแปรพยากรณ์แต่ละตัวที่ถูกเลือกเข้าในโมเดลจะมีนัยสำคัญทางสถิติในการพยากรณ์ Y โมเดลคอมพิวเตอร์บางโปรแกรมจะใช้เกณฑ์ค่า F ต่ำสุด โปรแกรมจะเลือกตัวแปรพยากรณ์ที่มีค่า F มากกว่าค่า F ที่เป็นเกณฑ์เข้าในโมเดลทีละตัว

ถ้ามีการอ้างถึงชุดของข้อมูลเดียวกันและมีระดับของนัยสำคัญเหมือนกันแล้ว การใช้วิธี backward elimination และ forward selection ไม่จำเป็นว่าจะได้โมเดลที่เหมือนกัน เนื่องจากความแตกต่างของตัวแปรที่ถูกเลือก วิธี backward elimination จะถูกใช้มากกว่าวิธี forward selection เพราะง่ายที่จะดูอิทธิพลโดยรวมทั้งหมดและพิจารณาในแต่ละขั้นของการประมาณค่าพารามิเตอร์และความคลาดเคลื่อนมาตรฐาน

วิธี stepwise selection เป็นการประยุกต์ใช้วิธี forward selection กับความแตกต่างที่สำคัญประการหนึ่ง คือตัวแปรทำนายจะถูกเลือกเข้าในโมเดลสามารถจะถูกขจัดออกจากโมเดลได้ ดังนั้นก็จะเป็นการใช้แนวคิดที่เกี่ยวกับวิธี backward elimination ในกรณีนี้มีโอกาสที่ตัวแปรพยากรณ์เมื่อเป็นตัวแปรสำคัญในขั้นตอนที่นำเข้าสมการแล้ว อาจจะไม่มีนัยสำคัญเมื่ออยู่ในโมเดล นั่นคือตัวแปรพยากรณ์ก็อาจจะถูกขจัดออกและมีการเพิ่มตัวแปรพยากรณ์ใหม่เข้าในสมการ

วิธี stepwise selection ในขั้นแรกของการวิเคราะห์จะไม่มีตัวแปรพยากรณ์ใด ๆ ในโมเดล ในขั้นตอนที่สอง ตัวแปรพยากรณ์จะถูกเพิ่มเข้าในโมเดล ซึ่งจะเป็นตัวแปรที่สามารถอธิบายตัวแปรเกณฑ์ได้สูงที่สุด หรืออาจพิจารณาเลือกจากค่า t หรือ F ที่สูงที่สุดซึ่งมีนัยสำคัญทางสถิติ หรือคือตัวแปรพยากรณ์ที่ถูกเลือกจะสามารถพยากรณ์ Y ได้สูงสุด ตัวแปรพยากรณ์ที่ถูกนำเข้าแล้วก็จะถูกตรวจสอบ ถ้าตัวแปรพยากรณ์ที่อยู่ในสมการมีนัยสำคัญก็จะอยู่ในโมเดลต่อไป แต่ถ้าไม่มีนัยสำคัญทางสถิติ ตัวแปรพยากรณ์ที่อยู่ในสมการนั้นก็จะถูกขจัดออกจากโมเดล การวิเคราะห์ดำเนินการต่อเนื่องไปจนกระทั่งตัวแปรพยากรณ์แต่ละตัวที่อยู่ในโมเดลสามารถพยากรณ์ Y ได้อย่างมีนัยสำคัญทางสถิติ ในขณะที่ไม่มีตัวแปรพยากรณ์อื่นที่มีนัยสำคัญทางสถิติอาจจะเปรียบเทียบสถิติ t หรือ F สำหรับตัวแปรพยากรณ์แต่ละตัวกับค่าวิกฤติ โปรแกรมคอมพิวเตอร์อาจจะใช้เกณฑ์ค่า F ในการเลือกตัวแปรเข้าหรือขจัดตัวแปรออก ถ้าหากมีการอ้างถึงข้อมูลชุดเดียวกันและระดับนัยสำคัญเท่ากันแล้ว ผลที่ได้จากวิธี blackward elimination, forward selection และ stepwise selection อาจจะไม่ได้โมเดลเหมือนกัน เนื่องจากความแตกต่างในการเลือกตัวแปรนำเข้า

วิธีการเลือกตัวแปรวิธีสุดท้ายคือ ใช้ชุดย่อยของตัวแปรพยากรณ์ที่เป็นไปได้ (all possible subsets regression) สมมติว่ามีตัวแปรพยากรณ์ 5 ตัว ในวิธีคัดเลือกด้วยวิธีนี้อาจจะมีโมเดลที่มีตัวแปรพยากรณ์ 1, 2, 3 และ 4 ตัวแปรที่ถูกวิเคราะห์ (ตัวแปรพยากรณ์ 5 ตัวจะมีได้เพียงโมเดลเดียว) ดังนั้นอาจจะได้โมเดลตัวแปรพยากรณ์เดียว 5 ตัวแปร หรือโมเดลที่มีตัวแปร

พยากรณ์ 2 ตัวแปรถึง 10 โมเดล หรือมีตัวแปรพยากรณ์ที่มี 3 ตัวแปร 10 โมเดล และมีตัวแปรพยากรณ์ที่มี 4 ตัวแปรถึง 5 โมเดล โมเดลที่มีตัวแปรพยากรณ์ k ตัวแปรสามารถจะถูกเลือกเข้าโมเดลโดยพิจารณาจาก R^2 สูงสุด ตัวอย่างเช่น โมเดลที่มีตัวแปรพยากรณ์ 3 ตัวจะต้องประมาณค่า R^2 ทั้ง 10 โมเดล และโมเดลที่ถูกเลือกควรเป็นโมเดลที่ดีที่สุดหรือโมเดลที่มีค่า R^2 ปรับแก้สูงสุด นั่นคือโมเดลที่มีค่า R^2 สูงสุดสำหรับจำนวนตัวแปรพยากรณ์ที่มีจำนวนน้อย อย่างไรก็ตาม นักวิจัยไม่แนะนำให้ใช้วิธีนี้ หากสามารถใช้วิธีการเลือกตัวแปรด้วยวิธีอื่น เมื่อจำนวนของตัวแปรพยากรณ์มีจำนวนมาก สำหรับวิธีการนี้ ผู้วิจัยอาจจะได้โมเดลที่เป็นขยะ (Gabagein gabage out : GIGO) นั่นคือจำนวนของโมเดลที่ได้จะเท่ากับ 2^m สำหรับกรณีที่มีตัวแปรพยากรณ์ 10 ตัวจะมีชุดของโมเดลย่อย ๆ ได้ 1,024 โมเดล

ในวิธีการเลือกตัวแปร มีวิธีอื่น ๆ อีก 2 วิธีที่จะกล่าวถึงอย่างคร่าว ๆ คือ 1) การวิเคราะห์แบบเชิงชั้น (Hierarchical regression) ผู้วิจัยจะมีลำดับเฉพาะในการนำเข้าตัวแปร ดังนั้น การวิเคราะห์จะเป็นแบบ forward selection (หรือ backward elimination) เป็นวิธีที่จะนำเข้าตัวแปรตามลำดับเชิงทฤษฎี 2) การวิเคราะห์ setwise regression (หรือ blockwise, chunkwise, forced stepwise regression) ผู้วิจัยจะมีลำดับสำหรับชุดของตัวแปร ซึ่งชุดของตัวแปรจะถูกกำหนดโดยผู้วิจัย เช่น เป็นชุดตัวแปรเชิงทฤษฎี (เช่น ชุดตัวแปรภูมิหลัง ประกอบด้วย ความถนัดทางการเรียน, ความสนใจใฝ่รู้ และอื่น ๆ) ตัวแปรที่อยู่ภายในชุดจะถูกเลือกตามลำดับด้วยวิธีการเลือกตัวแปร (เช่น backward elimination, forward selection, stepwise selection) ตัวแปรที่ถูกเลือกในแต่ละชุดที่นำเข้าสมการจะเข้าตามลำดับชุดที่มีความสำคัญก่อนหลังเชิงทฤษฎี

ลองพิจารณาด้วยตัวอย่างใหม่ที่แสดงถึงวิธีการคัดเลือกตัวแปร โดยข้อมูลใหม่นี้จะมีตัวแปรพยากรณ์ 4 ตัว และมีกลุ่มตัวอย่าง 20 คน แสดงในตาราง 3 ตัวแปรเกณฑ์คือความเข้าใจในการอ่าน (Reading comprehension) ตัวแปรอิสระคือ การระบุอักษร (Letter identification : X_1) ความรู้ในคำศัพท์ (Word knowledge : X_2) ทักษะการถอดรหัส (Decoding skill : X_3) และ อัตราการอ่าน (Reading rate : X_4) วิธีการแรกคือการเลือกตัวแปรแบบ forward selection แสดงอยู่ในตาราง 4 ในขั้นตอน 0 สังเกตว่าไม่มีตัวแปรพยากรณ์อยู่ในโมเดล แต่ทักษะการถอดรหัสจะเป็นตัวแปรแรกที่ถูกเลือกเข้าในโมเดล ทักษะการถอดรหัสจะถูกเลือกโมเดลในขั้นตอน 1 ค่า R^2 ปรับแก้มีค่า 0.47 และการระบุอักษรจะถูกเลือกเข้าโมเดลเป็นตัวแปรถัดไป ในขั้นตอน 2 การระบุอักษรจะถูกเลือกเข้าโมเดล มีค่า R^2 ปรับแก้เพิ่มขึ้นเป็น 0.57 ที่ระดับนัยสำคัญ 0.05 ไม่มีตัวแปรพยากรณ์ใด ๆ ถูกเลือกข้าโมเดล (ค่าสถิติ F สูงสุดคือตัวแปร X_2 มีค่า 3.06 ไม่มีนัยสำคัญทางสถิติที่ระดับ 0.05) และโดยปกติจะหยุดการวิเคราะห์ อย่างไรก็ตาม หากลองดำเนินการเลือกตัวแปรอย่างต่อเนื่องเข้าสมการทุกตัวแปรแล้ว ในขั้น 3 ความรู้ในคำศัพท์จะถูกเพิ่มเข้าในโมเดล ค่า R^2 ปรับแก้จะมีค่าเพิ่มเป็น 0.61 และอัตราการอ่านจะถูกเลือกเข้าโมเดลถัดไป ในขั้นสุดท้ายอัตราการอ่านจะถูกเพิ่มเข้าในโมเดลและ R^2 ปรับแก้จะมีค่าลดลงเหลือ 0.61 และตัวแปรทุกตัวพร้อมอยู่ในโมเดล

ตาราง 3 ตัวอย่างข้อมูลความเข้าใจในการอ่าน

คนที่	ความเข้าใจในการอ่าน	การระบุอักษร	ความรู้ในคำศัพท์	ทักษะการถอดรหัส	อัตราการอ่าน
1	54	2	26	16	3
2	53	10	29	15	4
3	40	4	6	5	3
4	53	5	20	16	5
5	49	9	10	8	4
6	53	10	25	11	6
7	53	7	21	11	5
8	54	9	25	13	4
9	50	8	6	13	4
10	40	4	11	6	5
11	51	5	19	7	3
12	47	5	9	10	3
13	46	9	6	5	2
14	53	9	26	16	3
15	53	7	34	16	6
16	53	10	20	16	5
17	52	10	14	5	4
18	52	9	23	15	6
19	48	6	28	10	4
20	52	7	10	16	5

ตาราง 4 ผลการวิเคราะห์การถดถอยคัดเลือกตัวแปรแบบ forward selection

		ตัวแปรในโมเดล	สัมประสิทธิ์	F	ตัวแปรที่ไม่อยู่ในโมเดล	F
ขั้นตอน 0	R^2 = 0.00	ไม่มี	Adj R^2 = 0.00		X_1 X_2 X_3 X_4	3.64 12.29 18.05 1.95
ขั้นตอน 1	R^2 = 0.50	X_3	0.70 Adj R^2 = 0.47	18.05	X_1 X_2 X_4	4.92 3.07 0.01
ขั้นตอน 2	R^2 = 0.61	X_1 X_3	0.58 0.66 Adj R^2 = 0.57	4.92 19.52	X_2 X_4	3.06 0.13
ขั้นตอน 3	R^2 = 0.67	X_1 X_2 X_3	0.55 0.15 0.49 Adj R^2 = 0.61	4.81 3.06 7.90	X_4	0.52
ขั้นตอน 4	R^2 = 0.69	X_1 X_2 X_3 X_4	0.59 0.16 0.52 -0.44 Adj R^2 = 0.60	5.12 3.33 8.19 0.52	ไม่มี	

ข้อมูลชุดเดียวกันนี้ ผลของการใช้ backward elimination และ stepwise selection จะให้ผลคล้ายกันกับ forward selection นั่นคือ การระบุอักษรและทักษะการถอดรหัสจะถูกเลือกเข้าในโมเดลที่ระดับนัยสำคัญ 0.05 แต่ก็ไม่ใช่เสมอไปที่ผลที่ได้จากการเลือกตัวแปรแต่ละวิธีจะให้ผลเหมือนกัน ผลการเลือกตัวแปรแบบ backward elimination และ stepwise selection จะไม่นำเสนอในที่นี้

สุดท้ายจะใช้ข้อมูลเดียวกันนี้วิเคราะห์แบบวิธี all possible subsets regression ซึ่งสรุปผลการวิเคราะห์ดังตาราง 5 สังเกตว่า โมเดลที่มีตัวแปรพยากรณ์เดียว ทักษะการถดถอดรหัสจะให้ผลดีที่สุด (ค่า R^2 ปรับแก้ มีค่า 0.47) โมเดล 3 ตัวแปรพยากรณ์ที่ดีที่สุดคือ การระบุอักษร

ความรู้ในคำศัพท์ และทักษะการถอดรหัส (ค่า R^2 ปรับแก้ มีค่า 0.61) และสุดท้ายโมเดลที่มีตัวแปรพยากรณ์ 4 ตัว (ค่า R^2 ปรับแก้ มีค่า 0.60) บนพื้นฐานของความมีนัยสำคัญทางสถิติที่ระดับ 0.05 โมเดลที่มีตัวแปรพยากรณ์ 2 ตัวแปร (คือ การระบุอักษรและทักษะการถอดรหัส) จะเป็นโมเดลที่ดีที่สุดในบรรดาโมเดลทั้งหมด และการเพิ่มขึ้นของตัวแปรที่เหลือไม่ได้ช่วยให้เพิ่มความสามารถในการพยากรณ์ความเข้าใจในการอ่านได้อย่างมีนัยสำคัญทางสถิติ

ตาราง 5 ผลการเลือกตัวแปรแบบ all possible subsets regression

	ตัวแปรในโมเดล	R^2	R^2 ปรับแก้
โมเดลตัวแปรเดียว	X_1	0.71	0.12
	X_2	0.41	0.37
	X_3	0.50	0.47
	X_4	0.10	0.50
โมเดล 2 ตัวแปร	X_1, X_2	0.51	0.46
	X_1, X_3	0.61	0.57
	X_1, X_4	0.22	0.12
	X_2, X_3	0.58	0.53
	X_2, X_4	0.41	0.34
	X_3, X_4	0.50	0.44
โมเดล 3 ตัวแปร	X_1, X_2, X_3	0.67	0.61
	X_1, X_2, X_4	0.51	0.42
	X_1, X_3, X_4	0.62	0.54
	X_2, X_3, X_4	0.58	0.50
โมเดล 4 ตัวแปร	X_1, X_2, X_3, X_4	0.69	0.60

ตัวอย่างกรณีมีกลุ่มตัวอย่าง 100 คน และมีตัวแปรพยากรณ์เพียง 2 ตัวในโมเดล มีค่า R^2 เท่ากับ 0.05 กับกรณีมีตัวแปรพยากรณ์ 90 ตัวแปร กับมีค่า R^2 เท่ากับ 0.90 ถ้าคำนวณค่า R^2 ปรับแก้ในตัวอย่างนี้เราจะพบว่ากรณีตัวแปรพยากรณ์ 2 ตัวมีค่า R^2 ปรับแก้ เท่ากับ 0.03 เปรียบเทียบกับกรณี 90 ตัวแปรมีค่า R^2 ปรับแก้เท่ากับ 0.10 เมื่อ ซึ่งเป็นที่น่าสงสัยถึงความแตกต่างของ R^2 ที่มีความความแตกต่างกันมาก กรณีมีตัวแปรพยากรณ์จำนวนมาก ในกรณีของจำนวนตัวแปรและค่าของ R^2 มีความสำคัญในการประเมินผลประโยชน์ของโมเดลการถดถอยที่ได้ สำหรับเกณฑ์ที่ประเมินว่าโมเดลการถดถอยมีประโยชน์มากน้อยเพียงใดคือควรจะเป็นโมเดลที่มีค่า MS_{res} ต่ำสุด หรือมีความคลาดเคลื่อนมาตรฐานของ b_k น้อยที่สุด

แปลและเรียบเรียงจาก

Lomax, Richard G. (1992). **Statistical Concepts : A Second Course for Education and the Behavioral Sciences**. London : Lawrence Erlbaum Associates, Inc.