

บทที่ 7

แบบจำลองถดถอยและตัวแปรหุ่น
Regression Model and Dummy Variables

ในบทนี้จะเป็นการศึกษาวิธีการประมาณการแบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพ (qualitative variable) โดยที่ตัวแปรเชิงคุณภาพจะเป็นตัวแปรที่บอกถึงคุณลักษณะของข้อมูล เช่น เพศ ฤดูกาล หรือ ช่วงเวลา ก่อนหรือหลังการเปลี่ยนแปลงระบบอัตราแลกเปลี่ยน จึงไม่สามารถที่จะเขียนตัวแปรเชิงคุณภาพให้อยู่ในรูปของตัวแปรเชิงปริมาณ (quantitative variable) ได้ ซึ่งโดยทั่วไปถ้าต้องการสร้างแบบจำลองถดถอยที่มีตัวแปรเชิงคุณภาพมาทำหน้าที่เป็นตัวแปรภายนอกจะต้องทำการเปลี่ยนตัวแปรเชิงคุณภาพให้เป็นตัวแปรหุ่น (dummy variable) เสียก่อน แล้วจึงนำเอาตัวแปรหุ่นเข้าไปทำหน้าที่เป็นตัวแปรภายนอกในแบบจำลองถดถอย โดยที่แบบจำลองถดถอยที่มีตัวแปรเชิงคุณภาพเป็นตัวแปรภายนอกที่จะกล่าวถึงในบทนี้จะกล่าวถึงแบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพทั้งหมด

การสร้างแบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพทั้งหมด สามารถจำแนกออกเป็น 2 ลักษณะ คือ แบบจำลองที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพเพียงตัวแปรเดียว และแบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพมากกว่า 1 ตัว ดังนี้

7.1 แบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพตัวแปรเดียว

แบบจำลองที่จะกล่าวถึงในหัวข้อนี้มีลักษณะเช่นเดียวกับแบบจำลองถดถอยอย่างง่ายสมการ (2.1.1) แต่จะเป็นแบบจำลองที่มีตัวแปรภายนอกเป็นตัวแปรหุ่นที่สร้างขึ้นจากคุณสมบัติของตัวแปรเชิงคุณภาพ ดังนี้

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i \quad (7.1.1)$$

โดยที่ D_i คือ ตัวแปรหุ่นที่สร้างขึ้นจากคุณสมบัติของตัวแปรเชิงคุณภาพ

การสร้างตัวแปรหุ่นจากคุณสมบัติของตัวแปรเชิงคุณภาพ จำแนกออกเป็น 2 วิธีตามจำนวนคุณสมบัติของตัวแปรเชิงคุณภาพ โดยการสร้างตัวแปรหุ่นกรณีแรกจะเป็นการสร้างตัวแปรหุ่นในกรณีที่ตัวแปรเชิงคุณภาพสามารถแบ่งคุณสมบัติของข้อมูลออกเป็น 2 กลุ่ม เช่น เพศ (ชายหรือหญิง) ความเห็น (เห็นด้วยหรือไม่เห็นด้วย) ระบบอัตราแลกเปลี่ยน (ระบบอัตราแลกเปลี่ยนแบบคงที่หรือระบบอัตราแลกเปลี่ยนแบบลอยตัว) หรือ ระดับการศึกษา (สำเร็จการศึกษาระดับมหาวิทยาลัยหรือไม่สำเร็จการศึกษาระดับมหาวิทยาลัย เป็นต้น) ซึ่งการสร้างตัวแปรหุ่นในกรณีนี้ทำได้โดยการกำหนดค่าให้กับคุณสมบัติหนึ่งเท่ากับศูนย์ และกำหนดให้ค่าของอีกคุณสมบัติหนึ่งที่เหลือมีค่าเท่ากับหนึ่ง เช่น กำหนดให้ D_i เป็นตัวแปรหุ่นของตัวแปรเชิงคุณภาพเพศ โดยกำหนดให้ $D_i = 0$ เมื่อตัวอย่างเป็นเพศชาย และ $D_i = 1$ เมื่อตัวอย่างเป็นเพศหญิง หรือกำหนดให้ X_i เป็นตัวแปรหุ่นของตัวแปรเชิงคุณภาพของระดับการศึกษา โดยที่ $X_i = 0$ เมื่อตัวอย่างไม่สำเร็จการศึกษาระดับมหาวิทยาลัย และ $X_i = 1$ เมื่อตัวอย่างสำเร็จการศึกษาระดับมหาวิทยาลัย

การสร้างตัวแปรหุ่นในกรณีที่สอง จะใช้เมื่อตัวแปรเชิงคุณภาพที่ต้องการศึกษามีการแบ่งคุณสมบัติของข้อมูลมากกว่า 2 คุณสมบัติ เช่น ระดับการศึกษา (ปริญญาตรี ปริญญาโท หรือปริญญาเอก) หรือ ไตรมาส (ไตรมาสที่หนึ่ง ไตรมาสที่สอง ไตรมาสที่สาม หรือไตรมาสที่สี่) เป็นต้น ในกรณีนี้จำเป็นต้องมีการสร้างตัวแปรหุ่นมากกว่าหนึ่งตัว โดยจำนวนของตัวแปรหุ่นที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนคุณสมบัติของตัวแปรเชิงคุณภาพ ตัวอย่างเช่น ระดับการศึกษา สามารถนำมาสร้างเป็นตัวแปรหุ่นทั้งสิ้น 3 ตัว หรือ ไตรมาส สามารถนำมาสร้างเป็นตัวแปรหุ่นทั้งสิ้น 4 ตัว โดยในกรณีของตัวแปรระดับการศึกษา จะทำการสร้างตัวแปรหุ่น 3 ตัว คือ D_{i1} , D_{i2} และ D_{i3} โดยที่ $D_{i1} = 1$ ถ้าตัวอย่างสำเร็จการศึกษาสูงสุดปริญญาตรี, $D_{i1} = 0$ ถ้าตัวอย่างไม่ได้สำเร็จการศึกษาสูงสุดปริญญาตรี, $D_{i2} = 1$ ถ้าตัวอย่างสำเร็จการศึกษาสูงสุดปริญญาโท, $D_{i2} = 0$ ถ้าตัวอย่างไม่ได้สำเร็จการศึกษาสูงสุดปริญญาโท, $D_{i3} = 1$ ถ้าตัวอย่างสำเร็จการศึกษาสูงสุดปริญญาเอก และ $D_{i3} = 0$ ถ้าตัวอย่างไม่ได้สำเร็จการศึกษาสูงสุดปริญญาเอก ในทำนองเดียวกันก็สามารถที่จะสร้างตัวแปรหุ่นจำนวน 4 ตัวให้กับตัวแปรเชิงคุณภาพไตรมาส

แบบจำลองถดถอยในลักษณะแรกที่จะพิจารณา คือ แบบจำลองถดถอยสมการ (7.1.1) โดยที่ D_i สร้างขึ้นจากตัวแปรเชิงคุณภาพที่แบ่งคุณสมบัติของตัวอย่างออกเป็น 2 กลุ่ม โดยที่ D_i ในสมการ (7.1.1) จะมีค่าเท่ากับ 0 ถ้าตัวอย่างมีคุณสมบัติที่หนึ่ง และ D_i ในสมการ (7.1.1) จะมีค่าเท่ากับ 1 ถ้าตัวอย่างมีคุณสมบัติที่สอง

หาค่าคาดหวังหรือค่าเฉลี่ยของสมการ (7.1.1) จะได้

$$E[Y_i] = \beta_0 + \beta_1 D_i \quad (7.1.2)$$

ค่าคาดหวังอย่างมีเงื่อนไข (conditional expected value) ดังนี้

$$E[Y_i | D_i = 0] = \beta_0 \quad (7.1.3)$$

$$E[Y_i | D_i = 1] = \beta_0 + \beta_1 \quad (7.1.4)$$

สมการ (7.1.3) คือ ค่าเฉลี่ยของตัวแปรภายในเมื่อตัวอย่างมีคุณสมบัติที่หนึ่ง และสมการ (7.1.4) คือ ค่าเฉลี่ยของตัวแปรภายในเมื่อตัวอย่างมีคุณสมบัติที่สอง ซึ่งจะเห็นได้ว่า ความหมายของพารามิเตอร์ในแบบจำลองสมการ (7.1.1) มีความแตกต่างไปจากความหมายของพารามิเตอร์ในแบบจำลองถดถอยอย่างง่ายสมการ (2.1.1) โดยในกรณีของสมการ (7.1.1) ค่า β_0 เมื่อพิจารณาจากสมการ (7.1.3) ระบุได้ว่าเป็นค่าเฉลี่ยของตัวแปรภายในกรณีที่ตัวแปรเชิงคุณภาพมีคุณสมบัติที่หนึ่ง หรือ $D_i = 0$ และจากสมการ (7.1.4) ที่ระบุว่า $\beta_0 + \beta_1$ คือ ค่าเฉลี่ยของตัวแปรภายในกรณีที่ตัวแปรเชิงคุณภาพมีคุณสมบัติที่สอง หรือ $D_i = 1$ ดังนั้น ค่าของ β_1 ตีความได้ว่าเป็นพารามิเตอร์ที่บอกถึงความแตกต่างของค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติที่สองและคุณสมบัติที่หนึ่ง

ภายหลังจากการนำแบบจำลองดังสมการ (7.1.1) ไปประมาณการจะได้ตัวประมาณการของค่าเฉลี่ยของตัวแปรภายในกรณีที่ตัวอย่างมีคุณสมบัติที่หนึ่ง หรือ $\hat{\beta}_0$ และตัวประมาณการของความ

แตกต่างของค่าเฉลี่ยตัวแปรภายในระหว่างตัวอย่างที่มีคุณสมบัติที่สองและตัวอย่างที่มีคุณสมบัติที่หนึ่ง หรือ β_1 จึงอาจกล่าวได้ว่า ผลการประมาณการสมการ (7.1.1) จะทำให้ทราบถึงค่าเฉลี่ยของตัวแปรภายในกรณีที่มีตัวอย่างที่มีคุณสมบัติของตัวแปรเชิงคุณภาพที่แตกต่างกัน

การทดสอบผลการประมาณการที่ได้จะเป็นการทดสอบสมมติฐานของพารามิเตอร์ เพื่อให้ทราบว่าค่าเฉลี่ยของตัวแปรภายในมีความแตกต่างกันระหว่างตัวอย่างที่มีคุณสมบัติของตัวแปรเชิงคุณภาพแตกต่างกันหรือไม่ ซึ่งก็คือ การทดสอบสมมติฐานเกี่ยวกับ β_1 โดยมีสมมติฐานการทดสอบ ดังนี้

$$\begin{aligned} H_0 : \beta_1 = 0 & \text{ หรือ } \text{ไม่มีความแตกต่างของค่าเฉลี่ยของตัวแปรภายใน} \\ & \text{ระหว่างตัวอย่างที่มีคุณสมบัติของตัวแปรเชิงคุณภาพแตกต่างกัน} \\ H_a : \beta_1 \neq 0 & \text{ หรือมีความแตกต่างของค่าเฉลี่ยของตัวแปรภายใน} \\ & \text{ระหว่างตัวอย่างที่มีคุณสมบัติของตัวแปรเชิงคุณภาพแตกต่างกัน} \end{aligned}$$

การทดสอบสมมติฐานข้างต้น จะอาศัยค่าสถิติ t ในการทดสอบ โดยที่ ค่าสถิติ t สามารถคำนวณได้ดังนี้

$$t_{n-2} = \frac{\hat{\beta}_1}{\sqrt{\text{Var}[\hat{\beta}_1]}} \quad (7.1.5)$$

โดยที่ $n-2$ คือ ลำดับของความเป็นอิสระ

การสร้างแบบจำลองถดถอยสำหรับกรณีที่มีตัวแปรเชิงคุณภาพแบ่งคุณสมบัติของตัวอย่างออกเป็นกลุ่มมากกว่า 2 กลุ่ม เช่น ระดับการศึกษา หรือไตรมาส จำเป็นต้องมีการสร้างตัวแปรหุ่นจำนวนเท่ากับจำนวนคุณสมบัติของตัวแปรเชิงคุณภาพ ดังนั้น ถ้าตัวแปรเชิงคุณภาพมีคุณสมบัติทั้งสิ้น m คุณสมบัติ จะต้องสร้างตัวแปรหุ่นทั้งสิ้น m ตัว และเขียนแบบจำลองถดถอย ได้ดังนี้

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_m D_{im} + \varepsilon_i \quad (7.1.6)$$

โดยที่ D_{im} คือ ตัวแปรหุ่นสำหรับคุณสมบัติที่ m โดยที่ $D_{im} = 1$ ถ้าตัวอย่างที่ i มีคุณสมบัติที่ m และ $D_{im} = 0$ ถ้าตัวอย่างที่ i ไม่มีคุณสมบัติที่ m

แบบจำลองถดถอยสมการ (7.1.6) ไม่สามารถที่จะนำไปประมาณการด้วยวิธีกำลังสองน้อยที่สุดได้ เนื่องจากตัวแปรหุ่นทั้งหมด m ตัวที่สร้างขึ้นนำมาซึ่งปัญหาความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์ระหว่างตัวแปรภายในนอก (perfect multicollinearity) เนื่องจาก $D_{i1} + D_{i2} + \dots + D_{im} = 1$ ปัญหาดังกล่าวเรียกว่า กับดักของตัวแปรหุ่น (dummy trap)

ในการแก้ไขปัญหากับดักของตัวแปรหุ่นที่เกิดขึ้นในแบบจำลองถดถอยสมการ (7.1.6) สามารถทำได้ 2 วิธี คือ การตัดค่าคงที่ออกจากแบบจำลอง (drop constant term) หรือการตัดตัวแปรหุ่นตัวใดตัวหนึ่งออกจากแบบจำลอง (drop dummy term) โดยการตัดค่าคงที่ออกจากแบบจำลอง หรือการตัด β_0 ออกจากแบบจำลอง จะได้สมการถดถอยสมการใหม่ ดังนี้

$$Y_i = \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_m D_{im} + \varepsilon_i \quad (7.1.7)$$

ภายหลังจากการตัดค่าคงที่ออกจากแบบจำลองตั้งสมการ (7.1.7) ส่งผลให้เกิดความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์ระหว่างตัวแปรภายในหมดไป และหาค่าคาดหวังของตัวแปรภายในได้ดังนี้

$$E[Y_i] = \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_m D_{im} \quad (7.1.8)$$

และค่าเฉลี่ยของตัวแปรภายในเมื่อกำหนดคุณสมบัติของตัวอย่างมาให้ สามารถเขียนได้ดังนี้

$$E[Y_i | D_{i1} = 1, D_{i2} = 0, \dots, D_{im} = 0] = \beta_1 \quad (7.1.9a)$$

$$E[Y_i | D_{i1} = 0, D_{i2} = 1, \dots, D_{im} = 0] = \beta_2 \quad (7.1.9b)$$

$$\vdots$$

$$E[Y_i | D_{i1} = 0, D_{i2} = 0, \dots, D_{im} = 1] = \beta_m \quad (7.1.9m)$$

จากสมการ (7.1.9) แสดงให้เห็นว่า พารามิเตอร์ β_m คือ ค่าเฉลี่ยหรือค่าคาดหวังของตัวแปรภายในเมื่อตัวอย่างมีคุณสมบัติที่ m

ในด้านการแก้ปัญหาเกี่ยวกับดักของตัวแปรหุ่น โดยการตัดตัวแปรหุ่นตัวใดตัวหนึ่งออกจากแบบจำลอง ทำให้จำนวนตัวแปรหุ่นลดลงไป 1 ตัว และส่งผลให้เกิดความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์ระหว่างตัวแปรภายในหมดไป ซึ่งจะสามารถหาตัวประมาณการด้วยวิธีกำลังสองน้อยที่สุดได้ ตัวอย่างเช่น การตัดตัวแปรหุ่นที่ m หรือ D_{im} ออกจากแบบจำลองถดถอยสมการ (7.1.6) จะได้แบบจำลองใหม่ ดังนี้

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_{m-1} D_{i,m-1} + \varepsilon_i \quad (7.1.10)$$

สมการ (7.1.10) หาค่าเฉลี่ยของตัวแปรภายในและค่าเฉลี่ยของตัวแปรภายในเมื่อกำหนดคุณสมบัติของตัวอย่างมาให้ เขียนได้ดังนี้

$$E[Y_i] = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_{m-1} D_{i,m-1} \quad (7.1.11)$$

$$E[Y_i | D_{i1} = D_{i2} = \dots = D_{i,m-1} = 0] = \beta_0 \quad (7.1.12a)$$

$$E[Y_i | D_{i1} = 1, D_{i2} = 0, \dots, D_{i,m-1} = 0] = \beta_0 + \beta_1 \quad (7.1.12b)$$

$$E[Y_i | D_{i1} = 0, D_{i2} = 1, \dots, D_{i,m-1} = 0] = \beta_0 + \beta_2 \quad (7.1.12c)$$

$$\vdots$$

$$E[Y_i | D_{i1} = 0, D_{i2} = 0, \dots, D_{i,m-1} = 1] = \beta_0 + \beta_{m-1} \quad (7.1.12m)$$

สมการ (7.1.12) ทั้งหมด แสดงให้เห็นว่า ผลจากการตัดตัวแปรหุ่นที่ m หรือ D_{im} ออกจากแบบจำลองถดถอยสมการ (7.1.6) ทำให้ β_0 เป็นพารามิเตอร์ที่บอกถึงค่าเฉลี่ยของตัวแปรภายใน

เมื่อตัวอย่างมีคุณสมบัติที่ m ในขณะที่พารามิเตอร์อื่นๆในสมการ (7.1.10) จะบอกถึงความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติที่ m กับค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติใดๆ เช่น ค่า β_1 คือ ความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติที่ m กับค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติที่ 1

การทดสอบสมมติฐานของพารามิเตอร์ในแบบจำลองถดถอยสมการ (7.1.10) จะทำการทดสอบพารามิเตอร์ทีละตัว (partial test) เพื่อเป็นการระบุว่า มีความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรภายในที่มีคุณสมบัติแตกต่างกันหรือไม่ โดยอาศัยค่าสถิติ t เช่นเดียวกับสมการ (6.1.5)

ในกรณีที่ตัวแปรเชิงคุณภาพเชิงคุณภาพจำแนกคุณสมบัติของกลุ่มตัวอย่างมากกว่า 2 คุณสมบัติแต่ไม่มีการสร้างตัวแปรหุ่นให้เท่ากับจำนวนคุณสมบัติ แต่ทำการสร้างตัวแปรหุ่นเพียงตัวเดียวและกำหนดค่าของตัวแปรหุ่นให้มีความแตกต่างกันไปตามคุณสมบัติ เช่น ตัวแปรเชิงคุณภาพที่มี m คุณสมบัติ ทำการสร้างตัวแปรหุ่นเพียงตัวเดียว คือ D_i โดยที่ $D_i = 0$ ถ้าตัวอย่างมีคุณสมบัติที่ 1, $D_i = 1$ ถ้าตัวอย่างมีคุณสมบัติที่ 1, ..., $D_i = m-1$ ถ้าตัวอย่างมีคุณสมบัติที่ m ทำให้แบบจำลองมีลักษณะเช่นเดียวกับสมการถดถอย (7.1.1) และมีค่าเฉลี่ยของตัวแปรภายในตามสมการ (7.1.2) แต่จะทำให้ค่าคาดหวังหรือค่าเฉลี่ยของตัวแปรภายในเมื่อกำหนดคุณสมบัติของตัวอย่างมาให้มีความแตกต่างจากสมการ (7.1.3) และสมการ (7.1.4) โดยในกรณีที่มีการสร้างตัวแปรหุ่นเพียงตัวเดียวให้กับตัวแปรเชิงคุณภาพที่มี m คุณสมบัติ จะมีค่าความคาดหวังหรือค่าเฉลี่ยอย่างมีเงื่อนไข ดังนี้

$$E[Y_i | D_i = 0] = \beta_0 \quad (7.1.13a)$$

$$E[Y_i | D_i = 1] = \beta_0 + \beta_1 \quad (7.1.13b)$$

$$E[Y_i | D_i = 2] = \beta_0 + 2\beta_1 \quad (7.1.13c)$$

$$E[Y_i | D_i = 3] = \beta_0 + 3\beta_1 \quad (7.1.13d)$$

$$\vdots$$

$$E[Y_i = m] = \beta_0 + [m-1]\beta_1 \quad (7.1.13m)$$

จากสมการ (7.1.13) เห็นได้ว่า β_0 คือ พารามิเตอร์ที่แสดงถึงค่าเฉลี่ยของตัวแปรภายในเมื่อตัวอย่างมีคุณสมบัติที่ 1 และจากความแตกต่างกันของค่าเฉลี่ยของตัวแปรภายในจะเห็นได้ว่า ความแตกต่างของค่าเฉลี่ยในแต่ละคุณสมบัติจะมีขนาดของความแตกต่างที่คงที่เท่ากับ β_1 ซึ่งเรียกว่า ข้อสมมติที่ต้องดำรงไว้ (maintain hypothesis) ที่ไม่อาจเกิดขึ้นได้ในโลกแห่งความเป็นจริง จึงอาจสรุปได้ว่า กรณีที่ตัวแปรเชิงคุณภาพจำแนกกลุ่มตัวอย่างออกเป็น m คุณสมบัติ จำเป็นต้องมีการสร้างตัวแปรหุ่นทั้งสิ้น m ตัวในแบบจำลองสมการถดถอย และตัวแปรหุ่นแต่ละตัวจะต้องมีเพียง 2 ค่าเท่านั้น คือ 0 และ 1

7.2 แบบจำลองถดถอยที่มีตัวแปรภายนอกเป็นตัวแปรเชิงคุณภาพมากกว่า 1 ตัวแปร

แบบจำลองถดถอยที่จะกล่าวถึงในหัวข้อนี้จะเป็นแบบจำลองถดถอยลักษณะเช่นเดียวกับสมการถดถอยเชิงพหุสมการ (3.1.1) ที่มีจำนวนตัวแปรภายนอกมากกว่า 1 ตัว แต่ในส่วนนี้จะพิจารณาในกรณีที่แบบจำลองมีตัวแปรภายนอกเป็นตัวแปรหุ่นที่สร้างขึ้นจากตัวแปรเชิงคุณภาพ หรือ

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_m D_{im} + \theta_1 R_{i1} + \theta_2 R_{i2} + \dots + \theta_p R_{ip} + \varepsilon_i \quad (7.2.1)$$

โดยที่ $D_{i1}, D_{i2}, \dots, D_{im}$ คือ ตัวแปรหุ่นของตัวแปรเชิงคุณภาพที่จำแนกคุณสมบัติของตัวอย่างออกเป็น m คุณสมบัติ และ $D_{im} = 1$ เมื่อตัวอย่างที่ i มีคุณสมบัติที่ m และ $D_{im} = 0$ เมื่อตัวอย่างที่ i ไม่มีคุณสมบัติที่ m และ $R_{i1}, R_{i2}, \dots, R_{ip}$ คือ ตัวแปรหุ่นของตัวแปรเชิงคุณภาพที่จำแนกคุณสมบัติของตัวอย่างออกเป็น p คุณสมบัติ และ $R_{ip} = 1$ เมื่อตัวอย่างที่ i มีคุณสมบัติที่ p และ $R_{ip} = 0$ เมื่อตัวอย่างที่ i ไม่มีคุณสมบัติที่ p

แบบจำลองสมการ (7.2.1) ไม่สามารถนำไปประมาณการด้วยวิธีกำลังสองน้อยที่สุด เพราะเกิดความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์ระหว่างตัวแปรภายนอกในหลากหลายรูปแบบ เช่น $D_{i1} + D_{i2} + \dots + D_{im} = 1$ หรือ $R_{i1} + R_{i2} + \dots + R_{ip} = 1$ แม้ว่าจะมีการแก้ปัญหาโดยการตัดค่าคงที่ออกจากแบบจำลองก็ไม่สามารถที่จะแก้ปัญหาที่เกิดขึ้น เนื่องจากจะยังคงเหลือความสัมพันธ์ เช่น $D_{i1} = D_{i2} + \dots + D_{im} - R_{i1} - R_{i2} - \dots - R_{ip}$

การตัดตัวแปรหุ่นของตัวแปรเชิงคุณภาพแต่ละตัวออกจากแบบจำลองชุดละหนึ่งตัว เช่น ตัด D_{im} และ R_{ip} ออกจากแบบจำลองสมการ (7.2.1) จะได้

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_{m-1} D_{i,m-1} + \theta_1 R_{i1} + \theta_2 R_{i2} + \dots + \theta_{p-1} R_{i,p-1} + \varepsilon_i \quad (7.2.2)$$

ผลจากการตัดตัวแปรหุ่นออกชุดละหนึ่งตัวจะส่งผลให้ความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์ระหว่างตัวแปรภายนอกหมดไป สามารถที่จะนำไปประมาณการด้วยวิธีกำลังสองน้อยที่สุดได้ แต่อย่างไรก็ตาม ผลจากการประมาณการสมการ (7.1.15) จะก่อให้เกิดข้อสมมติที่ต้องดำรงไว้

จากข้างต้นจะเห็นได้ว่า การประมาณการแบบจำลองถดถอยตามสมการ (7.1.14) ไม่สามารถทำได้ แม้ว่าจะมีการแก้ไขปัญหาดังกล่าวด้วยการตัดค่าคงที่หรือตัดตัวแปรหุ่นออกจากแบบจำลอง ดังนั้น ในกรณีที่ต้องการประมาณการสมการถดถอยที่มีลักษณะเช่นเดียวกับสมการ (7.1.14) จึงต้องมีการแก้ไขปัญหาโดยเริ่มต้นจากการตัดตัวแปรหุ่นออกชุดละหนึ่งตัวตั้งสมการ (7.1.15) จากนั้นจึงทำการสร้างตัวแปรภายนอกขึ้นมา เรียกว่า **interaction term** ที่เกิดขึ้นจากผลคูณระหว่างตัวแปรหุ่นและตัวแปรเชิงปริมาณ เช่น $D_{i1}R_{i1}$ เป็น **interaction term** ระหว่าง D_{i1} และ R_{i1} ซึ่งจะทำให้สมการ (7.1.15) เขียนใหม่ได้ดังนี้

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \dots + \beta_{m-1} D_{i,m-1} + \theta_1 R_{i1} + \theta_2 R_{i2} + \dots + \theta_{p-1} R_{i,p-1} + \delta_{11} D_{i1} R_{i1} + \delta_{12} D_{i1} R_{i2} + \dots + \delta_{1,p-1} D_{i1} R_{i,p-1}$$

$$+ \delta_{m-1,1} \mathbf{D}_{i,m-1} \mathbf{R}_{i1} + \dots + \delta_{m-1,p-1} \mathbf{D}_{i,m-1} \mathbf{R}_{i,p-1} + \varepsilon_i \quad (7.2.3)$$

แบบฝึกหัดบทที่ 7

แบบฝึกหัดที่ 7.1: จงพิสูจน์ว่าแบบจำลองถดถอยสมการ (7.1.6) ไม่สามารถนำไปหาค่าตัวประมาณการด้วยวิธีกำลังสองน้อยที่สุดได้

แบบฝึกหัดที่ 7.2: จงพิสูจน์ว่าการตัวค่าที่หรือการตัดตัวแปรหุ่นตัวใดตัวหนึ่งออกจากแบบจำลองถดถอยสมการ (7.1.6) สามารถแก้ไขปัญหากับดักของตัวแปรหุ่นได้

แบบฝึกหัดที่ 7.3: จงพิสูจน์ว่าการประมาณการแบบจำลองถดถอยสมการ (7.1.15) จะก่อให้เกิดข้อสมมติที่ต้องดำรงไว้ (maintained hypothesis)

แบบฝึกหัดที่ 7.4: จงพิสูจน์ว่าการประมาณการแบบจำลองถดถอยสมการ (7.2.3) ไม่ก่อให้เกิดข้อสมมติที่ต้องดำรงไว้