บทที่ 3
การวิเคราะห์เส้นถดถอยเชิงซ้อน
(Multiple Regression Analysis)

ในบทก่อน เราได้กล่าวถึงเส้นถดถอยอย่างง่ายซึ่งเป็นความสัมพันธ์ในเชิงเศรษฐกิจระหว่าง 2 ตัวแปร $Y = f(x)$ ในกรณีตัวอย่างเส้น demand ซึ่งมีตัวแบบ $Q_d = a + bP + e$ เราได้สร้างเทคนิคในการประมาณค่า a, b, ค่าสถิติต่าง ๆ และการทดสอบเกี่ยวกับความสัมพันธ์ของมัน อย่างไรก็ดี ความสัมพันธ์ทางเศรษฐกิจเกี่ยวข้องยุ่งยากมากกว่านี้ จำนวน demand มักจะไม่ขึ้นอยู่กับตัวแปรตัวเดียวแต่ขึ้นอยู่กับกลุ่มของตัวแปรอิสระทั้งกลุ่ม เช่น ราคาสินค้า X ราคาสินค้า Y ราคาสินค้า Z รายได้ และรสนิยม เป็นต้น ซึ่งจะเขียนเป็นฟังชั่นได้ว่า $Q_d = f(P_x, P_y, P_z, Y, T)$ ทั้งนี้เนื่องจากค่าตัวแปรในทางเศรษฐศาสตร์เป็นค่าสังเกตุที่มิใช่เกิดจากการทดลองจึงไม่สามารถควบคุมได้การศึกษาต้องศึกษาทั้งกลุ่ม ดังนั้น จึงต้องการเทคนิคการวิเคราะห์แบบ Multiple Regression

<u>ตัวแบบเส้นถดถอยเชิงซ้อน</u> (Multiple Regression Model)

ตัวแบบทั่วไปเกี่ยวกับประชากร : $\mu = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$

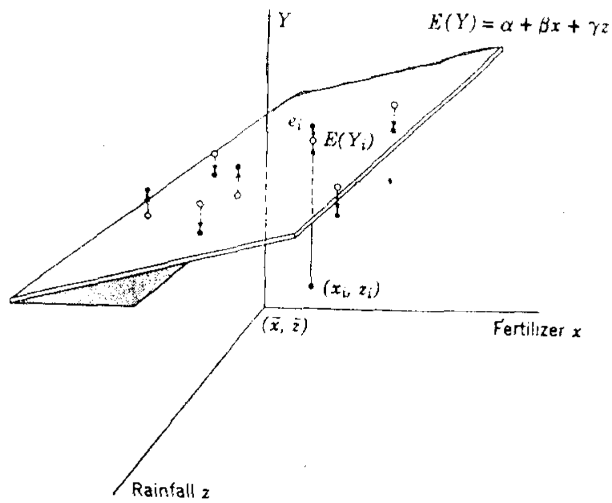
ตัวแบบที่ใช้ในตัวอย่าง : $Y = a + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki} + e_i \quad (i = 1,2,..n)$

u_i เป็นตัวแปรอิสระสุ่มซึ่ง u มี mean = 0 และ Variance = σ_u^2, X อาจจะเป็นค่ากำหนดให้ที่ไม่สุ่ม (nonrandom) หรือไม่ก็เป็นตัวแปรสุ่ม (random) ซึ่งมีการแจกแจงของ u เป็นอิสระ ข้อสมมุติของเส้นถดถอยเชิงซ้อนก็เหมือนกับข้อสมมุติของเส้นถดถอยอย่างง่ายนอกจากแทนที่จะมีตัวแปรอิสระหรือตัวแปรภายนอกเพียงตัวเดียวเท่านั้น

ก็มีตัวแปรอิสระถึง k ตัว

สมมติความสัมพันธ์ของตัวแปร 3 ตัว ถ้านำมาเขียนกราฟ 3 มิติจะได้

ดังรูปที่ 3.1



รูปที่ 3.1 ภาพกระจายของค่าสังเกตุบนพื้นระนาบ regression ที่แท้จริง

<u>ตัวประมาณค่า a, b, c โดยวิธีกำลังสองน้อยที่สุด</u>

เพื่อที่จะหาตัวประมาณค่าของพารามิเตอร์ในสมการถดถอยเชิงซ้อน โดยวิธีกำลังสองน้อยที่สุด สมมุติว่าพิจารณาจากสมการถดถอย 3 ตัวแปร คือ ตัวแปรภายใน Y ขึ้นอยู่กับตัวแปรภายนอก X Z อย่างเป็นเส้นตรง

$$Y_i = a + bX_i + cZ_I + e_i \quad (i = 1, 2,....n)$$

e_i เป็นส่วนคลาดเคลื่อนที่เหลือ (residual error) หรือความเบียงเบนของค่าที่ประมาณได้ $\hat{Y}$ จากค่า Y ที่แท้จริง

โดยวิธีกำลังสองน้อยที่สุด ค่า a, b, c สามารถหาได้จาก normal equation ที่เกิดจากการ minimize Σe_i^2 ดังนี้

$$\emptyset = \Sigma_1^n e_i^2 = \Sigma_1^n (Y_i - a - bX_i - cZ_i)^2$$

ใช้แคลคูลัส โดย differentiate $\emptyset$ มุ่งต่อ a, b, c และตั้ง derivative แต่ละอันเท่ากับ 0 จะได้สมการ 3 สมการ ซึ่งเรียกว่า normal equation

$$\sum_1^n (Y_i - a - bX_i - cZ_i) = 0$$

$$\sum_1^n (Y_i - a - bX_i - cZ_i) X_i = 0$$

$$\sum_1^n (Y_i - a - bX_i - cZ_1) Z_i = 0$$

เพราะว่า $Y_i - a - bX_i - cZ_i = e_i$ ดังนั้นสมการข้างบนจะเขียนได้เป็น

$$\Sigma_1^n e_i = 0, \quad \Sigma_1^n e_iX_i = 0, \quad \Sigma_1^n e_iZ_i = 0$$

เพื่อหาค่า a, b, c ในเทอมของ Y_i, X_i และ Z_i เราจึงเขียนสมการ normal equation เสียใหม่ดังนี้

$$an + b\Sigma_1^n X_i + c\Sigma_1^n Z_i = \Sigma_1^n Y_i \qquad (1)$$

$$a\Sigma_1^n X_i + b\Sigma_1^n X_i^2 + c\Sigma_1^n X_i Z_i = \Sigma_1^n Y_i X_i \qquad (2)$$

$$a\Sigma_1^n Z_i + b\Sigma_1^n X_i Z_i + c\Sigma_1^n Z_i^2 = \Sigma_1^n Y_i Z_i \qquad (3)$$

สมการ (1) จะให้เราทราบว่าพื้นระนาบ $Y = a + bX + cZ$ ซึ่งฟิตกับข้อมูลที่ที่สุด ต้องผ่านจุด $\bar{Y}, \bar{X}, \bar{Z}$ จะเห็นได้จากเอา n หารสมการ (1) ตลอด จะได้ :-

$$a + b\bar{X} + c\bar{Z} = \bar{Y} \qquad (1')$$

* สูตร $a = \bar{Y} - b\bar{X} - c\bar{Z}$

เราสามารถจะหา b และ c โดยขจัด a จากสมการ (2)และ (3) ซึ่งทำได้โดยคูณสมการ (1) ด้วย $\bar{X}$ แล้วนำไปหักจากสมการ (2), คูณสมการ (2)ด้วย $\bar{Z}$ แล้วนำไปหักออกจาก (3)

เพราะว่า $\Sigma_1^n X_i - n\bar{X} = 0$ $\Sigma_1^n Z_i - n\bar{Z} = 0$ จะได้ :-

$$b\,(\Sigma_1^n X_i^2 - \bar{X}\Sigma_1^n X_i) + c\,(\Sigma_1^n X_i Z - \bar{X}\Sigma_1^n Z_i) = \Sigma_1^n Y_i X_i - \bar{X}\Sigma_1^n Y_i$$

และ

$$b\,(\Sigma_1^n X_i Z_i - \bar{Z}\Sigma_1^n X_i) + c\,(\Sigma_1^n Z_i^2 - \bar{Z}\Sigma_1^n Z_i) = \Sigma_1^n Y_i Z_i - \bar{Z}\Sigma_1^n Y_i$$

เพราะว่า

$$\Sigma_1^n X_i^2 - \bar{X}\,\Sigma_1^n X_i = \Sigma_1^n (X_i - \bar{X})^2,$$

$$\Sigma_1^n X_i Z_i - \bar{X}\,\Sigma_1^n Z_i = \Sigma_1^n (X_i - \bar{X})(Z_i - \bar{Z}_i) ,\quad \text{เป็นต้น}$$

สมการข้างต้นนี้จะเท่ากับสมการข้างล่างคือ

$$b\,\Sigma_1^n (X_i - \bar{X})^2 + c\,\Sigma_1^n (X_i - \bar{X})(Z_i - \bar{Z}_i) = \Sigma_1^n (X_i - \bar{X})(Y_i - \bar{Y}), \quad (2')$$

$$b\,\Sigma_1^n (X_i - \bar{X})(Z_i - \bar{Z}) + c\,\Sigma_1^n (Z_i - \bar{Z})^2 = \Sigma_1^n (Z_i - \bar{Z})(Y_i - \bar{Y}) \quad (3')$$

จากสมการ $(2')$ และ $(3')$ จะได้ค่า b และ c และเพื่อให้จำง่าย ให้
$x_i = X_i - \bar{X}$ $z_i = Z_i - \bar{Z}$ และ $y_i = Y_i - \bar{Y}$

แล้ว $\Sigma_1^n (X_i - \bar{X})^2 = \Sigma_1^n x_i^2$, $\Sigma_1^n (X_i - \bar{X})(Z_i - \bar{Z}) = \Sigma x_i z_i$ เป็นต้น

ค่า b และ c ในเทอมของ x_i , y_i และ z_i จะเป็นดังนี้

$$* \text{ สูตร} \quad \hat{b} = \frac{\Sigma z_i^2 \Sigma y_i x_i - \Sigma x_i z_i \Sigma y_i z_i}{\Sigma x_i^2\, \Sigma z_i^2 - (\Sigma x_i z_i)^2}$$

$$\hat{c} = \frac{\Sigma x_i^2\, \Sigma y_i z_i - \Sigma x_i z_i\, \Sigma y_i x_i}{\Sigma x_i^2\, \Sigma z_i^2 - (\Sigma x_i z_i)^2}$$

ตัวประมาณค่า a , b และ c โดยวิธีกำลังสองน้อยที่สุดนี้สามารถคำนวณด้วยมือได้ แต่ละตัวในสมการ b และcจะหาได้โดยตรงหรือจะหาตามวิธีข้างล่าง ซึ่งจะรวดเร็วและง่ายกว่าก็ได้ แล้วไปแทนค่าในสมการ b และ c เราจะได้ค่าประมาณของ b และ c

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$\sum z^2 = \sum Z^2 - \frac{(\sum X)^2}{n}$$

$$\sum yx = \sum XY - \frac{(\sum X)(\sum Y)}{n}$$

$$\sum yz = \sum YZ - \frac{(\sum Y)(\sum Z)}{n}$$

$$\sum xz = \sum XZ - \frac{(\sum X)(\sum Z)}{n}$$

เมื่อได้ค่า a, b และ c แล้ว นำไปแทนในสมการถดถอย เราจะได้สมการถดถอยเชิงซ้อนที่มีความหมายและอธิบายได้ดังนี้

<u>ตัวอย่างที่ 3.1</u>

สมมติคำนวณสมการถดถอยเชิงซ้อนของการเสนอซื้อส้มขึ้นอยู่กับราคาส้มและรายได้ของผู้บริโภค คือ

$$Y = 15 - 2X + Z$$

โดย Y เป็นการเสนอซื้อส้ม(ก.ก.) X เป็นราคาส้ม (10 บาท/ ก.ก.) Z เป็นรายได้ของผู้บริโภค,(100 บาท) สมการนี้สามารถที่จะใช้พยากรณ์การเสนอซื้อส้มได้ ถ้ารู้ราคาส้ม(X) และรายได้ของผู้ซื้อ (Z) โดยแทนค่า X และ Zลงในสมการ

ค่า a, b และ c สามารถที่จะอธิบายได้ดังนี้

a เรียกว่า ตัวคงที่ (Constant) หมายถึง เมื่อราคาส้มและรายได้ของผู้บริโภคเป็น 0 การเสนอซื้อส้มจะเท่ากับ 15 กิโลกรัม (ค่า a ไม่ใคร่มีความสำคัญ บางครั้งใช้ไม่ได้)

b เรียกว่า สัมประสิทธิ์ถดถอย หรือสัมประสิทธิ์ถดถอยบางส่วน (Partial Regression Coefficient) หมายถึง แต่ละการเพิ่มในราคาส้ม 1 หน่วยหรือ 10 บาทต่อ ก.ก. เป็นผลให้การเสนอซื้อส้มลดลง 2 กิโลกรัม โดยให้รายได้ของผู้ซื้ออยู่คงที่

c เรียกเช่นเดียวกับค่า b, c หมายถึง แต่ละการเพิ่มในรายได้ 1 หน่วยหรือ 100บาทเป็นผลให้การเสนอซื้อเพิ่มขึ้น1 กิโลกรัม โดยให้ราคาส้มอยู่คงที่

<u>ค่าสถิติของสมการถดถอยเชิงซ้อน</u>

1. <u>ความคลาดเคลื่อนมาตรฐานของการกะประมาณ</u> (Standard Error of Estimate : S E E, $S_{y.12}$)

จากบทที่ 2 เรารู้ว่า ถ้าการกระจายของข้อมูลห่างจากเส้นถดถอยน้อย (อยู่ใกล้เส้นถดถอย) ความคลาดเคลื่อนของการพยากรณ์ก็น้อย ความคลาดเคลื่อนหรือความแปรปรวนใน Y จะแสดงออกในรูปStandard Error of Estimate สำหรับค่า Y ในสมการถดถอยเชิงซ้อนเช่นเดียวกัน ความคลาดเคลื่อนของการ กะประมาณสำหรับค่าของ Y บนพื้นราบถดถอย อธิบายได้ดังนี้

$$\text{SEE หรือ } S_{y.12} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-3}} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

อาจจะใช้สูตรที่เป็นค่าเบี่ยงเบนจาก means (อักษรตัวเล็ก)

$$\text{SEE} = \sqrt{\frac{\sum y^2 - b\sum xy - c\sum yz}{n-3}}$$

หรืออีกสูตรหนึ่ง

$$\text{SEE} = \sqrt{\frac{\sum y^2 - \sum(bx+cx)^2}{n-3}}$$

ถ้าหากใช้เครื่องคำนวณ สูตรข้างล่างจะนิยมใช้มากกว่า เพราะสูตรนี้ไม่ต้องใช้ทั้งค่า $\hat{Y}$ และค่าเบี่ยงเบนจาก means ซึ่งค่อนข้างล่าช้าสำหรับ ค่าสังเกต (n) จำนวนมาก

$$\text{SEE} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY - c\sum YZ}{n-3}}$$

<u>2. สัมประสิทธิ์การตัดสินใจ</u> (Multiple Coefficient of Determination : R^2)

R^2 ของสมการถดถอยเชิงซ้อนมีความหมายเช่นเดียวกับของสมการถดถอยอย่างง่าย คือ การวัดอัตราส่วนของความแปรปรวนของตัวแปรตามที่ถูกอธิบายโดยสมการถดถอย ถ้าอธิบายในเชิงรูปภาพเรขาคณิต (รูปที่ 3.1) R^2 ของสมการถดถอยเชิงซ้อน คือ การวัดว่า พื้นที่ราบซึ่งแทนสมการถดถอยมีความเหมาะสม (fit)กับจุดต่าง ๆ ของข้อมูล (ค่าสังเกต)บนพื้นราบได้ดีเพียงไร สูตรของ R^2 คือ

$$R^2 = \frac{\text{ความแปรปรวนที่อธิบายได้}}{\text{ความแปรปรวนทั้งหมด}}$$

$$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

อาจจะใช้สูตรที่เป็นค่าเบี่ยงเบนจาก means

$$R^2 = \frac{b\sum xy + c\sum yz}{\sum y^2}$$

$$R^2 = \frac{\sum (bx+cz)^2}{\sum y^2}$$

ถ้ามีตัวแปรอิสระมากกว่า 2 ตัวขึ้นไป การคำนวณสมการถดถอยเชิงซ้อนส่วนใหญ่จะใช้โปรแกรมสำเร็จรูปทางคอมพิวเตอร์

ตัวอย่างที่3.2

จากตัวอย่างสมการถดถอยของการเสนอซื้อส้มขึ้นอยู่กับราคาส้มและรายได้ของผู้บริโภคสมมติคำนวณได้ $R^2 = 0.92$ มีความหมายว่า ประมาณ 92 เปอร์เซนต์ของความแปรปรวนในการเสนอซื้อส้มถูกอธิบายโดยสมการถดถอยอีก 8 เปอร์เซ็นต์(100-92%) ของความแปรปรวนของการเสนอซื้อส้มมาจากสาเหตุอื่นหรือตัวแปรอื่นที่ไม่ได้นำเข้ามาในสมการ นั่นคือ ราคาส้มและรายได้ของผู้ซื้อมีอิทธิพลต่อการเสนอซื้อส้ม 92 เปอร์เซ็นต์ (ค่า R^2 ของสมการถดถอยที่มีตัวแปรอิสระมากกว่า 1 ตัวจะสูงกว่าสมการที่มีตัวแปรอิสระเพียงตัวเดียว แสดงว่าในความเป็นจริงตัวแปรมีความสัมพันธ์กันมากกว่า 1ตัว)

สัมประสิทธิ์การตัดสินใจปรับปรุง $(\bar{R}^2)$

ในสมการถดถอยเชิงซ้อน ตัวแปรอิสระจะมีมากกว่าในสมการถดถอยอย่างง่าย ถ้ายิ่งตัวแปรอิสระเพิ่มมากขึ้นเท่าไร R^2 ยิ่งมีค่าสูงขึ้น หรือ " ใกล้ 1 "เข้าไปเท่านั้น เพราะว่าค่า $b\sum xy + c\sum yz$ ซึ่งเป็น "ความแปรปรวนที่อธิบายได้" ค่าจะเพิ่มขึ้นตามตัวแปรอิสระที่เพิ่มขึ้นแต่ $\sum y^2$ ซึ่งเป็น"ความแปรปรวนทั้งหมด"เป็นตัวหารยังมีค่าคงเดิม($R^2 = \frac{b_1\sum x_1 y + b_2\sum x_2 y + \ldots + b_k\sum x_k y}{\sum y^2}$)การที่ R^2 มี ค่าใกล้ 1 ไม่แสดงว่าเส้นถดถอยที่คำนวณได้ฟิตกับข้อมูลที่สุด การแก้จะต้องปรับ R^2 เสียใหม่ด้วย degree of freedom ของสมการถดถอยเชิงซ้อนคือ

$$\bar{R}^2 = 1 - \frac{\sum e_i^2}{\sum y^2}\frac{(n-1)}{(n-k-1)}$$

อย่างไรก็ดี ถ้าหากว่าจำนวนค่าสังเกต (n) ที่ใช้น้อย การสร้างสมการต้องไม่ให้มีตัวแปรอิสระมากเกินไป หรืออาจจะเพิ่ม n ให้มากขึ้น แต่ก็ควรคำนวณ $\bar{R}^2$ อีกครั้ง เพื่อความแน่ใจในค่า R^2

4. <u>ความคลาดเคลื่อนมาตรฐานของ a , b และ c</u>

เนื่องจากค่าประมาณต่างๆเป็นค่าสถิติจากตัวอย่างซึ่งมักจะมีความความคลาดเคลื่อนอันเนื่องมาจากการสุ่มตัวอย่าง(sampling error) ดังนั้นนอกจากจะคำนวณตัวประมาณค่า a, b , c แล้วเราควรต้องหาความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเหล่านี้คือ

$$S_a = \sqrt{S_a^2} = \sqrt{\frac{S^2\sum X^2\sum Z^2-(\sum XZ)^2}{n\left[\sum x^2\sum z^2-(\sum xz)^2\right]}}$$

$$S_B = \sqrt{S_b^2} = \sqrt{\frac{S^2\sum z^2}{\sum x^2\sum z^2-(\sum xz)^2}}$$

$$S_c = \sqrt{S_c^2} = \sqrt{\frac{S^2\sum x^2}{\sum x^2\sum z^2-(\sum xz)^2}}$$

<u>ช่วงความเชื่อมั่นของการประมาณค่า α และ β_1</u>

วัตถุประสงค์ของการวิเคราะห์ถดถอยเชิงซ้อนอย่างหนึ่งคือต้องการทราบช่วงความเชื่อมั่นของสัมประสิทธิ์ถดถอยที่แท้จริง (*true regression coefficient* : $\beta_{1,}$) เพราะ β_1 เป็นตัวที่วัดผลกระทบของ X_1 ที่เพิ่มขึ้น 1 หน่วยต่อค่า Y เมื่อ X_2 อยู่คงที่ และ β_2 วัดผลกระทบของ X_2 ที่เพิ่มขึ้น 1 หน่วยต่อค่า Y เมื่อ X_1 อยู่คงที่ เมื่อเราทราบความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์bถดถอยจากตัวอย่าง S_a ,S_b, S_c ถ้าเราตั้งความเชื่อมั่นอยู่ที่ 100 (1 - α) เปอร์เซ็นต์ ช่วงความเชื่อมั่นของค่าคงที่ (α) และสัมประสิทธิ์ถดถอยที่แท้จริงจะหาได้ดังนี้

ช่วงความเชื่อมั่นของค่าคงที่ที่แท้จริง $\alpha = a \pm t\, t_{\alpha/2} S_a$ ที่ d.f. n-k-1

ช่วงความเชื่อมั่นของสัมประสิทธิ์ถดถอยที่แท้จริง $\beta_1 = b_1 \pm t_{\alpha/2} S_{b1}$ ที่ d.f.n-k-1

ช่วงความเชื่อมั่นของสัมประสิทธิ์ถดถอยที่แท้จริง :: $\beta_2 = b_2 \pm t_{\alpha/2} S_{b2}$

ช่วงความเชื่อมั่นของสัมประสิทธิ์ถดถอยที่แท้จริง :: $\beta_k = b_k \pm t_{\alpha/2} S_{bk}$

5 <u>ค่าสถิติ t และการทดสอบ (t - test)</u>

การทดสอบสัมประสิทธิ์ถดถอยทีละตัวจะใช้ทดสอบด้วย t ถ้าเราตั้งสมมติฐานให้เป็น 0 สูตรทดสอบคือ :

$t_a = \frac{a}{S_a}$, $t_b = \frac{b}{S_b}$, $t_c = \frac{c}{S_c}$ เขตวิกฤติ ถ้า t > $t_{\alpha/2,n-k-1}$ และ $-t < -t_{\alpha/2,n-k-1}$

<u>ตัวอย่างที่ 3.3</u>

ให้ Y เป็นต้นทุนการผลิต (หมื่นบาท) X_1 เป็นผลผลิต (ตัน) X_2 เป็นประสบการณ์ของผู้จัดการ(ปี)คอมพิวเตอร์ได้พิมพ์ผลของการวิเคราะห์สมการถดถอยจากตัวอย่าง 9 ตัวอย่างออกมาดังนี้

$$Y = 1.324 + 0.779 X_1 + 0.167 X_2$$

Predictor	Coef.	Stdev.	t - ratio
Constant	1.3236	0.1432	9.2430
X_1	0.7792	0.0335	23.2043
X_2	- 0.16692	0.0322	- 5.1839

ให้ใช้ผลเหล่านี้คำนวณช่วงความเชื่อมั่น 95 เปอร์เซ็นต์ของ β_1 และทดสอบสมมติฐานว่า Ho : $\beta_2 = 0$, Ha : $\beta_2 \neq 0$ ณ ระดับนัยสำคัญ .05 .

<u>วิธีทำ</u>

$$\begin{aligned}\beta_1 &= b_1 \pm t_{\alpha/2,n-k-1} S_{b1}\\ &= 0.7792 \pm t_{.05/2,9-2-1}(.034)\\ &= 0.7792 \pm 2.447(.034)\\ &= 0.7792 \pm 0.0832\\ &= \{0.6960 , 0.8624\}\end{aligned}$$

ตีความ : เมื่อผลผลิต (X_1) เพิ่มขึ้น 1 ตัน ต้นทุนการผลิต (Y)จะเพิ่มขึ้นอยู่ในช่วงระหว่าง 6,960 บาท ถึง 8,634 บาท ด้วยความเชื่อมั่น 95 เปอร์เซ็นต์

<u>การทดสอบ</u> (สรุปผลจากค่า t ratio ได้เลย แต่ข้างล่างแสดงการหาค่า t ให้ดู)

Ho : $\beta_2 = 0$ (ประสบการณ์ของผู้จัดการอธิบายความผันแปรในต้นทุนการผลิตไม่ได้)

Ha : $\beta_2 \neq 0$

สูตร $$t_{b2} = \frac{b_2}{S_{b2}} = \frac{-0.16692}{0.0322} = -5.1839$$

สรุปผล : ค่าสถิติ t (- 5.1839) < $-t_{.025}$ (-2.447) เราปฏิเสธสมมติฐานที่ว่า $\beta_2 = 0$ เรียกว่า β_2 มีนัยสำคัญทางสถิติ ประสบการณ์ของผู้จัดการอธิบายความผันแปรในต้นทุนการผลิตได้ หรืออาจกล่าวว่า ต้นทุนการผลิตซึ่งเป็นผลจากผู้จัดการมีประสบการณ์เพิ่มขึ้น 1 ปีต่างไปจาก 0 อย่างมีนัยสำคัญ นั่นคือ ประสบการณ์ของผู้จัดการมีผลทำให้ต้นทุนการผลิตลดลงจริง

<u>การวิเคราะห์ความแปรปรวน</u> (Analysis of Varience)

นักสถิติได้ใช้การวิเคราะห์ความแปรปรวนเพื่อทดสอบนัยสำคัญทางสถิติทั้งหมดของสมการถดถอย กล่าวคือ ใช้ทดสอบว่า สัมประสิทธิ์ถดถอยที่แท้จริงทุกตัวในสมการเป็น 0 หรือไม่ เช่นเดียวกับที่ใช้ในสมการ ถดถอยอย่างง่ายเพื่อทดสอบค่า β

ตั้งสมมติฐาน Ho : $\beta_1 = \beta_2 = 0$

Ha : $\beta_1 \neq \beta_2 \neq 0$

สูตร

$$F = \frac{\text{ความแปรปรวนที่อธิบายได้ด้วย Regression}}{\text{ความแปรปรวนที่อธิบายไม่ได้ด้วย Regression}}$$

$$= \frac{S_1^2}{S_2^2}$$

เขตวิกฤติ : F > F_α ที่ d.f. n และ n - k - 1

ตารางที่ 3.1 การวิเคราะห์ความแปรปรวน

Source of variation	degree of freedom	Sum of squares	Mean square	F ratio
Explained by Regression	k	$\sum(\hat{Y}_i - \bar{Y})^2$	$\frac{\sum(Y_i - \bar{Y})^2}{k} = S_1^2$	$\frac{S_1^2}{S_2^2}$
Unexplained By regression	n - k - l	$\sum(Y_i - \hat{Y}_i)^2$	$\frac{\sum(Y_i - \hat{Y}_i)^2}{n-k-1} = S_2^2$	
Total	n-l	$\sum(Y_i - \bar{Y})^2$		

ตัวอย่าง ตาราง ANOVA ที่ใช้ความแปรปรวนอีกรูปหนึ่ง

ตารางที่ 3.2 การวิเคราะห์ความแปรปรวน

S V	d.f.	S S	M S	F ratio
EV	k	$b\sum xy + c\sum yz$ หรือ $\sum (bx+cz)^2$	$\frac{b\sum xy + c\sum yz}{k} = S_1^2$	
U V	n - k - l	$\sum e^2$ หรือ $\sum y^2 - b\sum xy - c\sum yz$ หรือ $\sum y^2 - \sum (bx+cz)^2$	$\frac{\sum e^2}{\text{n - k - l}} = S_2^2$	$= \frac{S_1^2}{S_2^2}$
T V	n-1	$\sum y^2$		

การทดสอบว่าสัมประสิทธิ์ถดถอยทุกตัวจะเป็น 0 หรือไม่ ถ้าสัมประสิทธิ์ถดถอยทุกตัวเป็น 0 จริง อัตราส่วน $\frac{S_1^2}{S_2^2}$ จะมีการแจกแจงเป็นแบบ F ซึ่งมี d.f.= k และ n – k – 1 (ในตาราง F, d.f. = k อยู่แถวนอน d.f. = n – k – 1 อยู่แถวตั้ง) แต่ถ้าสัมประสิทธิ์ถดถอยไม่เท่ากับ 0 แล้วอัตราส่วนนี้มีแนวโน้มที่จะให้ตัวเลขสูง (ค่า F สูง)

การสรุปผลว่า จะปฏิเสธหรือยอมรับสมมติฐานดูที่ค่าวิกฤติเช่นเดียวกับในสมการถดถอยอย่างง่าย

ตัวอย่างที่ 3.4 สมการถดถอยของต้นทุนการผลิตขึ้นอยู่กับผลผลิตและประสบการณ์ของผู้จัดการในตัวอย่างที่3.3 สมมติมีค่าคำนวณ F =50 ให้ทดสอบสัมประสิทธิ์ถดถอยที่แท้จริงทุกตัวว่า = 0 ตั้ง α = .05

วิธีทำ สมการถดถอยนี้มี n = 9, และ k = 2 จำนวน degree of freedom คือ 2 กับ 6 (k=2และ n – k –1 = 6) $F_{.05}$ = 4.76 (จากตาราง F ท้ายเล่ม) เนื่องจาก F ที่คำนวณได้มากกว่า 4.76 เราปฏิเสธHo ที่ว่าสัมประสิทธิ์แท้จริงทุกตัวเป็น 0 แสดงว่าผลผลิตและประสบการณ์ของผู้จัดการอธิบายความผันแปรในต้นทุนการผลิตได้ นั่นคือ ตัวแปรทั้ง 3 มีความสัมพันธ์กัน ผลผลิตและประสบการณ์ของผู้จัดการมีผลต่อต้นทุนการผลิต

<u>ตัวอย่างที่ 3.5</u>

จงสร้างสมการเส้นถดถอยของจำนวน demand ขึ้นอยู่กับราคาและรายได้ ดังข้อมูลข้างล่างนี้

Q_d	100	75	80	70	50	65	90	100	110	60
P	5	7	6	6	8	7	5	4	3	9
Y	1,000	600	1,200	500	300	400	1,300	1,000	1,300	300

วิธีทำ

1. ตัวแบบ $Q_d = a + bP + cY + 2$

2. การคำนวณ

เพื่อความสะดวกในการคำนวณ เราใช้ตัวแบบทั่วไปก่อนคือ $Y = a + bX + cZ$

ตารางที่ 3.3

n	Y Quantity demanded	X Price	Z Income	y $(Y-\bar{Y})$	x $(X_1-\bar{X}_1)$	z $(Z-\bar{Z})$	y^2	x^2	z^2	yx	yz	xz
1	100	5	1,000	20	-1	200	400	1	40,000	-20	4,000	-200
2	75	7	600	-5	1	-200	25	1	40,000	-5	000	-200
3	80	6	1,200	0	0	400	0	0	160,000	0	3, 0	0
4	70	6	500	-10	0	-300	100	0	90.000	0	15,000	0
5	50	8	300	-30	2	-500	900	4	250,000	-60	6,000	-1,000
6	65	7	400	-15	1	-400	225	1	160,000	-15	6.000	-400
7	90	5	1,300	10	-1	500	100	1	250,000	-10	5,000	-500
8	100	4	1,100	20	-2	300	400	4	Qb 000	-40	6,000	-600
9	110	3	1,300	30	-3	500	900	9	250,000	-90	15,000	-1 ,500
10	60	9	300	-20	3	500	400	9	250,000	-60	10.000	-1,500
$n = 10$	$\Sigma Y_i = 800$	$\Sigma X = 60$	8,000				3,650	30	1,580,000	-300	65,000	-5,900

$\bar{Y} = 80,\quad \bar{X} = 6,\quad \bar{Z} = 800.$

$$\hat{b} = \frac{\Sigma z^2 \Sigma xy - \Sigma xz \; cyz}{\Sigma x^2 \Sigma z^2 - (\Sigma xz)^2}$$

$$= \frac{(1,580,000)(-300)-(-5,900)(65,000)}{(30)(1,580,000)-(-5,900)^2}$$

$$= \frac{-90,500}{12,590}$$

$$= -7.1882$$

$$\hat{c} = \frac{\Sigma x^2 \Sigma yz - \Sigma xz \; cxy}{\Sigma x^2 \; \Sigma z^2 - (\Sigma xz)^2}$$

$$= \frac{(30)(65,000)-(-5,900)(-300)}{(30)(1,580,000)-(-5,900)^2}$$

$$= \frac{180}{12,590}$$

$$= 0.0143$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} - \hat{c}\bar{Z}$$

$$= 80 - (-7.19)(6) - (0.0143)(800)$$

$$= 111.69$$

3. สมการ regression

$$Y = 111.69 - 7.19\,x + 0.0143\,z$$

4. สมการ demand

$$Q_d = 11.69 - 7.19\ P + 0.0143\ Y$$

<u>ค่าสถิติที่จำเป็น</u> :-

1. $$R^2 = \frac{b\ cxy + c\ \Sigma yz}{\Sigma y^2}$$

$$= \frac{(-7.19)(-300) + (0.0143)(65,000)}{3,450}$$

$$= 0.894$$

หรือจะใช้อีกสูตรหนึ่งคือ

$$R^2 = 1 - \frac{\Sigma e^2}{y^2}$$

$$\Sigma e^2 = y^2 - b\ Cxy - c\ \Sigma yz$$

$$= 3,450 - (-7.19)(-300) - (0.0143)(65,000)$$

$$= 363.5$$

$$R^2 = 1 - \frac{363.5}{3,450}$$

$$= 0.894$$

2. $\bar{R}^2$

$$\bar{R}^2 = 1 - \frac{(n-1).\ e^2}{(n-k-1)\ y^2}$$

$$= 1 - \frac{(10-1)}{(10-2-1)} \cdot \frac{363.5}{3,450} = 1 - \frac{3,271.5}{24,150}$$

$$= 0.86$$

3. $$SEE = \sqrt{S^2} = \sqrt{\frac{\Sigma e^2}{n-3}}$$

$$= \sqrt{\frac{363.5}{10-3}} = \sqrt{51.93}$$

$$= 7.21$$

4. $$\hat{s}_a = \sqrt{\frac{S^2\ \Sigma X^2\ \Sigma Z^2 - (\Sigma XZ)^2}{n\left[\Sigma x^2\ \Sigma z^2 - (\ \Sigma xz)^2\right]}}$$

$$= \sqrt{553.69}$$

$$= 23.5$$

สูตรนี้จะต้องตีช่องเพิ่มสำหรับ X^2, Z^2 และ XZ แล้วหา Σ ของมัน สำหรับ XZ จะต้องนำ Σ XZ ไปยกกำลังสอง

$$\hat{S}_b = \sqrt{\frac{S^2 \, \Sigma z^2}{\Sigma x^2 \, \Sigma z^2 - (\Sigma xz)^2}}$$

$$= \sqrt{\frac{(51.93)(1,580,000)}{(30)(1,580,000)-(-5,900)^2}}$$

$$= \sqrt{\frac{820,494}{34,814,740}} = 2.25$$

$$\hat{S}_c = \sqrt{S_c^2}$$

$$= \sqrt{\frac{S^2 \Sigma x^2}{\Sigma x^2 \Sigma z^2 - (\Sigma xz)^2}}$$

$$= \sqrt{\frac{(51.93)(30)}{(30)(1,580,000)-(-5,900)^2}}$$

$$= \sqrt{\frac{1,557.9}{34,814,740}} = 0.01$$

5. ค่าสถิติ t

$$\hat{t}_a = \frac{\hat{a}}{\hat{S}_a}$$

$$= \frac{111.69}{23.5} = 4.75$$

$$t_b = \frac{\hat{b}}{\hat{S}_b}$$

$$= \frac{-7.19}{2.25} = -2.8$$

$$\hat{t}_c = \frac{\hat{c}}{\hat{S}_c}$$

$$= \frac{0.014}{0.01} = 1.28$$

หมายเหตุ :

ค่า t ใช้ทดสอบสัมประสิทธิ์ถดถอยทีละตัว ในสมการเชิงซ้อน โดยมีสมมติฐานว่า $\alpha_1=0$ $\beta_1 = 0$ และ $\alpha_2=0$ $\beta_2 = 0$

6. ค่าสถิติ F

ตารางที่ 3 4 ANOVA Table

Source of Variation	d.f	S.S	M.S	F
Explained (by Regression)	k = 2	$\Sigma y^2 - \Sigma e^2$ =3450-363.5 =3086.5	$S_1^2 = \Sigma y^2 - \Sigma e^2/2$ = 3,086/2 = 1,543	$\frac{S_1^2}{S_2^2} = \frac{1,543}{51.93}$
Unexplained (error)	n-k-l = 10-2-l = 7	Σe^2 = 363.5	$S_2^2 = \Sigma e^2/n-k-1$ = 363.517 = 51.93	= 29.71
Total Variation	n-l 10-1=9	$\Sigma y^2 = 3,450$		

(มีนัยสำคัญ ระดับความเชื่อมั่น 99%)

หมายเหตุ:

ค่าสถิติ F ใช้ทดสอบ

1. สัมประสิทธิเส้นถดถอย ซึ่งมีสมมุติฐาน $\hat{b} = \hat{c} = 0$ ใช้สูตรดังในตาราง 3.3

2. ใช้ทดสอบ $\bar{R}^2$ ซึ่งมีสมมุติฐาน $R^2 = 0$ สูตร F คือ

$$F = \frac{R^2/2}{(1-R^2)/(n-3)}$$

แล้วนำค่าที่คำนวณได้ (ปกติค่า F ในสูตรข้อ 2 จะเท่ากับค่า F ในข้อ 1) ไปเปรียบเทียบกับตาราง F ถ้ามากกว่า ก็สรุปว่า มีนัยสำคัญ (ปฏิเสธ $R^2 = 0$)

F-ratio ที่คำนวณได้ ถ้ามีค่ายิ่งสูงมาก (ปฏิเสธ $R^2 = 0$) ทำให้เราแน่ใจยิ่งขึ้นว่า ตัวแปร X และ Z มีความสัมพันธ์กับ Y แน่นอน

จะได้สมการเส้น demand คือ

$$Q^d = a + bP + cY$$

$$= 111.69 - 7.19\ P + 0.0143\ Y \quad \underline{1/}$$

$$(23.5) \quad (2.25) \quad (0.01)$$

$$t \quad 4.75 \quad -2.8 \quad 1.28$$

$R^2 = 0.874$, $\bar{R}^2 = 0.864$

SEE = 7.21

F = 29.71

1/ จากสมการนี้ ความยืดหยุ่นของราคาและความยืดหยุ่นของรายได้จะหาได้โดย :

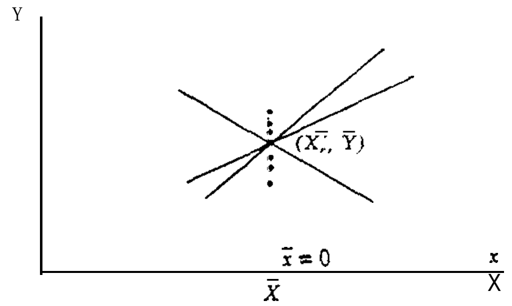
$$e_p = \frac{\partial q}{\partial p} \cdot \frac{p}{q} = b. \frac{\bar{p}}{4} = (-7.19)\frac{6}{80} = -0.60$$

และ $$e_y = \frac{\partial q}{\partial y} \cdot \frac{y}{q} = c. \frac{\bar{y}}{\bar{q}} = (0.0143) \frac{800}{80} = 0.14$$

ปัญหาในการวิเคราะห์เส้นถดถอย

1. ปัญหา Multicollinearity

1.1 ใน Simple Regression เรียกว่า ปัญหา Collinearity เกิดขึ้นเมื่อค่าสังเกตุของตัวแปรอิสระ (X) อยู่ใกล้กันมากเกินไปหรืออยู่เป็นกระจุก มีการกระจายน้อยมาก ค่า X ทุกค่าจึงอยู่ใกล้กับ $\bar{X}$ มาก (ดังรูปที่ 3.2)ทำให้ $(X - \bar{X}) = 0$



รูปที่ 3.2 เส้น regression ที่เกิดจากไม่มีการกระจายใน X

เมื่อเป็นเช่นนี้ เราไม่สามารถคำนวณค่า b ได้เพราะ

$$\hat{b} = \frac{\Sigma xy}{\Sigma x^2}$$

แต่ $x = (X - \bar{X}) = 0$

$$\Sigma xy = 0, \quad \Sigma x^2 = 0$$

$$\hat{b} = 0$$

ดังนั้น เราไม่สามารถจะอธิบาย Y ได้ การที่ X ไม่มีกระจายหรือมีน้อย ทำให้เกิดเส้นถดถอยได้หลายเส้น แต่ละเส้นมีความลาดชันต่างกันแต่มีความสำคัญพอๆกัน เพราะการพยากรณ์ Y ของแต่ละเส้นทำได้จำกัดด้วยค่า X ค่าเดียวกันเส้น regression ทุกเส้นจึงพยากรณ์ค่า Y ได้ดีเท่าๆกันและให้ผลเหมือนกัน คือ $\hat{Y} = a + 0 = \frac{\Sigma Y}{n} = \bar{Y}$ จึงถือว่า X ไม่มีอิทธิพลต่อ Y

1.2 ใน Multiple Regression จะเกิดปัญหา Multicollinearity ซึ่งคือการที่ตัวแปรอิสระในสมการ Multiple Regression มีความสัมพันธ์ซึ่งกันและกันเป็นเส้นตรง เช่น สมการ $Y = a + bX + cZ$ Multicollinearity เกิดขึ้นเมื่อ X และ Z มีความสัมพันธ์กัน (เช่น $X = 1.5Z$ ซึ่งปกติเราจะไม่ทราบความสัมพันธ์ที่แน่นอนเช่นนี้) R. Frish เป็นคนแรกที่เรียกความสัมพันธ์เส้นตรงระหว่างตัวแปรอิสระ 2 ตัวหรือมากกว่า 2 ตัวนี้ว่า Multicollinearity

เมื่อมี Multicollinearity เกิดขึ้น สัมประสิทธิ์ที่อยู่หน้าตัวแปรอิสระ (b, c,...) นั้น ๆ จะอธิบายอะไรไม่ได้เลย สัมประสิทธิ์อาจจะใหญ่มากเกินไป จนไม่มีความสำคัญหรืออาจจะเท่ากับ 0

การที่เราจะทราบว่ามี Multicollinearity หรือไม่ ดูจาก Standard Error ของสัมประสิทธิ ($\hat{S}_a$, $\hat{S}_b$,...) ถ้าค่าของ Standard Error สูง ตัวแปรอิสระนั้นก็อาจจะมี Multicollinearity แต่วิธีนี้ก็ไม่แน่เสมอไป วิธีดูอีกวิธีหนึ่งคือ ใช้การวิเคราะห์อิทธิพลร่วม (Confluence - analysis) โดยการคำนวณ Simple Regression ของ Y ต่อตัวแปรอิสระที่สงสัยว่าจะมี Multicollinearity แล้วคำนวณ Multiple Regression ของ Y โดยนำเอาตัวแปรอิสระเข้ามาในสมการทีละตัว เช่น กรณี $Y = a + bX + cZ$ เราสงสัยว่า X และ Z จะมีความสัมพันธ์กัน เราก็คำนวณหา Simple Regression ของ Y on X, ($\hat{Y} = \hat{a} + \hat{b}X$), Y on Z ($\hat{Y} = \hat{a} + \hat{c}Z$) แล้ว Y on X on Z ($\hat{Y} = \hat{a} + \hat{b}X + \hat{c}Z$)

และแต่ละสมการจะให้ค่า $\hat{b}$, $\hat{c}$, R^2 ทั้งหมดจะได้ค่า $\hat{b}$ และ $\hat{c}$ อย่างละ 2 ค่า R^2 ได้ 3 ค่า แล้วมาดูว่าค่าที่ได้นี้เป็นอย่างไร ถ้า R^2 ได้เท่ากันหรือใกล้เคียงกัน โดยมิได้สูงขึ้น (ปกติถ้าเพิ่มตัวแปรอิสระเข้าไป R^2 จะค่อย ๆ สูงขึ้น) และค่า b, c แต่ละสมการต่างกัน โดยเปลี่ยนแปลงไปแทนที่จะค่อนข้างคงที่ แสดงว่ามี Multicollinearity

ทางแก้ปัญหา Multicollinearity ก็คือ ควรเก็บตัวอย่างข้อมูลให้ขนาดใหญ่ขึ้น พยายามเลือกตัวแปรอิสระที่มีความหมายพอและไม่แสดงความสัมพันธ์ระหว่างกันอย่างเห็นชัด โดยใช้เหตุและผลทางทฤษฎีเศรษฐศาสตร์ช่วย

2. <u>การใช้ Dummy Variables</u>

ตัวแปรที่เราเกี่ยวข้องในการวิเคราะห์เส้นถดถอยมาตั้งแต่ต้นล้วนเป็นตัวแปรที่วัดได้ เช่น ปริมาณซื้อ (Q_d) ราคา (P) รายได้ (Y) และการบริโภค (C) เป็นต้น ปัญหาในการวิเคราะห์เส้นถดถอยอีกอย่างหนึ่งคือ จะต้องเกี่ยวข้องกับตัวแปรที่ไม่สามารถจะวัดได้ ซึ่งมีลักษณะเป็นตัวแปรคุณภาพ (quanlity variable) และเราเชื่อกันว่า ตัวแปรคุณภาพนี้มีส่วนสำคัญซึ่งจะกระทบกระเทือนถึงตัวแปรอื่น ๆ ตัวอย่างเช่น ค่าใช้จ่ายในการบริโภคในครอบครัวไม่สามารถจะถูกอธิบายทั้งหมดได้โดยรายได้ของครอบครัว หรือกล่าวง่าย ๆ ว่าการบริโภคมิได้ขึ้นอยู่กับรายได้อย่างเดียว ยังมีตัวแปรอื่น ๆ เช่น เพศ ศาสนา เชื้อชาติ การศึกษาของหัวหน้าครอบครัว อาจจะกระทบกระเทือนค่าใช้จ่ายในการบริโภค ซึ่งตัวแปรเหล่านี้วัดออกมาเป็นตัวเลขไม่ได้ เราจะเอาตัวแปรเหล่านี้เข้ามารวมในโมเดลเส้นถดถอยได้อย่างไร และถ้าหากไม่นำเข้ามาพิจารณาด้วยก็จะเป็นการไม่ถูกต้องจะทำให้ค่าพยากรณ์ผิดพลาดไป ด้วยเหตุนี้จึงมีผู้คิด "ตัวแปรหุ่น" (Dummy Variable) ขึ้นมา Dummy Variable คือ ตัวแปรที่กำหนดขึ้นมาโดยพละการแทนตัวแปรที่วัดไม่ได้ มีลักษณะเป็นตัวแปรอิสระหรือตัวแปรภายนอกทำหน้าที่เหมือนตัวแปรอิสระอื่น ๆ และอยู่ทางขวามือของสมการ คือ

$$Y_i = a + bX + c_1D_1 + c_2D_2 + \ldots\ldots + c_jD_3$$

$D_l \ldots\ldots\ldots D_j$ คือ ตัวแปรหุ่นแทนตัวแปรคุณภาพที่วัดเป็นตัวเลขไม่ได้ ตัวแปรคุณภาพนี้ยังรวมไปถึงผลกระทบจากกฎเกณฑ์ที่เราตั้งขึ้นหรือจากเหตุการณ์ต่าง ๆ ซึ่งไม่สามารถจะวัดออกมาได้เช่นกัน เช่น ผลกระทบจากสงคราม จากการเปลี่ยนแปลงทางการเมือง จากการวางแผนพัฒนาเศรษฐกิจ, จากการค้าต่างประเทศ จากการขึ้นราคาสินค้า ภาษี การควบคุม, ข้อห้ามต่าง ๆ ฯลฯ เป็นต้น

ค่าของ Dummy Variable จะมี 2 ค่า คือ 1 และ 0 ค่า 1 จะแทนพฤติกรรมที่เหมาะกับกรณีหรือแทนสถานะการณ์ที่มีผลแน่ในช่วงเวลาที่ศึกษา ส่วนพฤติกรรมที่ตรงข้ามหรือไม่มีผลก็จะกำหนดค่าให้ = 0 เช่น ให้ D เป็น Dummy Variable แทนเพศของหัวหน้าครอบครัว เราอาจกำหนดค่า 1 ถ้าหัวหน้าครอบครัวเป็นหญิง และ 0 ถ้าหัวหน้าครอบครัวเป็นชาย หัวหน้าครอบครัวเป็นหญิงมีแนวโน้มที่มีรายได้ต่ำกว่าหัวหน้าครอบครัวเป็นชาย ดังนั้นในระดับรายได้ที่กำหนดให้ ครอบครัวที่มีหัวหน้าเป็นหญิงจะบริโภคมากกว่า (ความโน้มเอียงในการบริโภคเฉลี่ย) ในการสำรวจเก็บข้อมูลตัวอย่างจะปรากฏข้อมูลเป็น 2 กลุ่ม ดังรูป 3.3 ถ้าหากเราไม่พิจารณาเพศของหัวหน้าครอบครัว จะได้เส้น regression ซึ่งมี slope ค่อนข้างราบ (เส้นประในรูป) แต่ถ้านำเพศมาพิจารณาด้วย เราจะได้ slope ที่แท้จริง แม้เส้น regression จะแยกกัน แต่ slope ก็เท่ากัน ซึ่งแสดงว่า แม้ว่าความโน้มเอียงในการบริโภคเฉลี่ยในครอบครัวที่มีหัวหน้าเป็นหญิงจะสูงกว่าชาย แต่ความโน้มเอียงในการบริโภคเพิ่ม (Marginal Propensity to Consume) จะเท่ากันหรือเหมือนกัน