บทที่ 6

สมมติฐานและความสำคัญของรูปแบบสมการถดถอยอย่างง่าย

(Assumptions and Significance of the Simple Regression Model)

วัตถุประสงค์ : เพื่อศึกษาถึง สมมติฐานของสมการถดถอยเส้นตรง ในหัวข้อการได้มาซึ่งข้อมูลการกระจายของค่า Y (ตัวแปรตาม) การกระจายของสัมประสิทธิ์ความถดถอย การกระจายของค่าผิดพลาด หรือค่าคลาดเคลื่อน นอกจากนี้แล้วนักศึกษาจะได้ศึกษาถึง ค่าสัมประสิทธิ์สหสัมพันธ์ ในด้านความหมาย การตรวจสอบนัยสำคัญของค่า r การตรวจสอบนัยสำคัญของค่า $\hat{\alpha}$ และ $\hat{\beta}$

บทที่ 6

สมมติฐานและความสำคัญของรูปแบบสมการถดถอยอย่างง่าย

(Assumptions and Significance of the Simple Regression Model)

6.1 สมมติฐานของสมการถดถอยเส้นตรง

6.1.1 การได้มาซึ่งข้อมูล

ตัวอย่างที่ได้ศึกษาในบทก่อนหน้าเกี่ยวกับผู้ผูกขาดที่ตั้งราคาแตกต่างกันสำหรับสินค้าของเขาในแต่ละช่วงของเดือน และสังเกตผลของ ความต้องการ ในช่วง 6 เดือนที่เขาเลือก ข้อมูลจากการสังเกตถูกแสดงที่ตาราง 5.1 สมมติว่าเขาต้องการจะทดลองซ้ำอีกครั้งหนึ่งในอีก 6 เดือนถัดมา โดยใช้รูปแบบการตั้งราคาแบบเดิม เราจะแปลกใจถ้าปริมาณของความต้องการเหมือนกันกับ 6 เดือนแรก ปัจจัยอื่นนอกจากราคาได้ส่งผลกระทบต่อปริมาณความต้องการ และไม่มีเหตุผลที่จะคาดหวังว่าปัจจัยเหล่านี้จะเหมือนกันในช่วง 6 เดือนถัดมา

ตารางที่ 6.1 แสดงข้อมูลของการทดลองซ้ำ

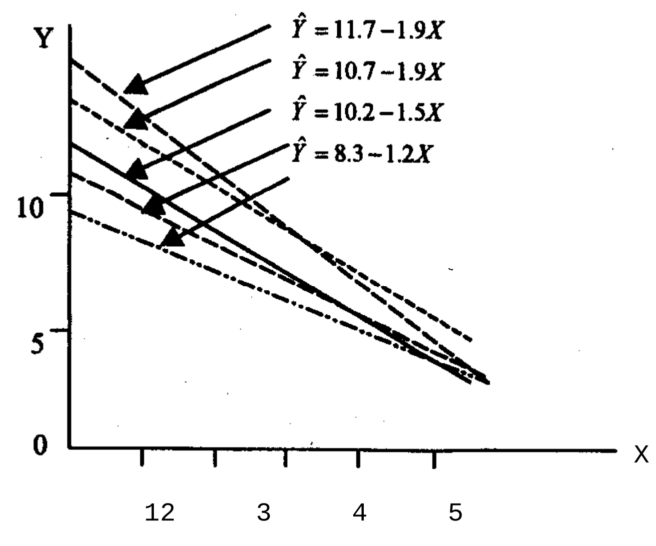
Month	Price(X)	Experiment				
		1	2	3	4	5
1	2	8	10	6	9	8
2	4	3	2	6	6	3
3	3	4	5	4	6	4
4	1	7	8	8	7	10
5	3	8	8	2	5	8
6	5	0	0	2	1	3

ผู้ผูกขาดทดลองซ้ำหลายครั้ง สมมติว่า 5 ครั้งแบบ 6 เดือน สมมติให้รูปแบบราคาถูกตั้งไว้ในเดือนที่ 1,2,...,6 ของแต่ละช่วง เหมือนกับการทดลองแบบแรก ผลของปริมาณความต้องการจะแสดงในตาราง 6.1 ที่การทดลองครั้งที่ 1 กล่าวถึงการสังเกตขั้นต้นที่เราวิเคราะห์ในบทก่อน และการทดลองครั้งที่ 2-5 จะกล่าวถึงระยะ 6 เดือนถัดมา

ตารางที่ 6.2 แสดงค่าสัมประสิทธิ์สมการถดถอยจากข้อมูลการทดลอง

Experiment	$\hat{\alpha}$	$\hat{\beta}$
1	10.7	-1.9
2	12.7	-2.4
3	8.3	-1.2
4	10.2	-1.5
5	11.7	-1.9

สมมติให้เราใช้เทคนิคสมการถดถอยอย่างง่ายสำหรับแต่ละชุดของข้อมูลในตาราง 6.1 ยกตัวอย่างเช่น การให้ผลของการทดลองแยกต่างหาก สัมประสิทธิ์สมการถดถอยที่เกี่ยวกับชุดของมูลค่า Y ในการทดลองครั้งที่ 1 กับชุดของมูลค่า x ที่ได้ในบทก่อนหน้า $\hat{\alpha}$ = 10.7, $\hat{\beta}$ =-1.9 และทำกระบวนการประมาณค่าการถดถอยครั้งที่ 2,3,...... ชุดของมูลค่า y ต่อชุดของมูลค่า x ให้ค่าของสัมประสิทธิ์สมการถดถอยตามตารางที่ 6.2 แต่ละชุดของข้อมูลนำไปสู่สัมประสิทธิ์ของสมการถดถอยที่แตกต่างกัน เส้นถดถอยที่ประมาณค่าทั้ง 5 จากการทดลองแยกกันทั้ง 5 ครั้ง ใน รูปที่ 6.1



รูปที่ 6.1 แสดงเส้นถดถอยที่ประมาณจากการทดลองทั้ง 5

เราสามารถอธิบายความแตกต่างระหว่างเส้นถดถอยในเทอมของปัจจัยภายนอกที่ได้กล่าวไปแล้ว การทดลองดังกล่าวไม่ได้ถูกควบคุมมากกว่าปัจจัยสุ่มที่มีผลต่อปริมาณความต้องการในแต่ละเดือน ผลกระทบจะแสดงให้เห็นเป็นการหลากหลายในตัวแปร Y นอกจากนี้ ข้อผิดพลาดในการวัดอาจทำให้ข้อมูลผิดไปในบางเดือน

ดังนั้นจะเห็นชัดว่าไม่ว่าจะเป็นชุดของข้อมูลแบบเดี่ยว ผลการทดลองแบบเดี่ยวจะถือเป็นตัวอย่างของการสังเกต ราคากับปริมาณสำหรับจำนวนประชากรที่มาก ในความเป็นจริงประชากรของการสังเกตราคากับปริมาณในตัวอย่างจะไม่ถูกกำหนดเพราะอาจจะเป็นไปได้ว่ามีการทดลองซ้ำแล้วซ้ำอีก และแต่ละครั้งก็จะมีชุดข้อมูลใหม่ที่ได้จากการสังเกต

6.1.2 การกระจายของค่า Y

ถ้าเราใช้ตัวอย่างซ้ำจากประชากรดังกล่าว คือทำการทดลองครั้งที่ 6,7,8,.... เราจะค้นพบหลักเกณฑ์บางอย่างจากผลการทดลอง ยกตัวอย่างเช่นให้X =2 ซึ่งแสดงในเดือนที่ 1 ในตารางที่ 6.1 ดูการโต้ตอบของ Y ซึ่งอยู่แถวบนของตาราง ค่าเฉลี่ยของค่าทั้ง 5 คือ 8.2 เราสามารถสังเกตได้ว่าถึงแม้มูลค่าแค่ 5 ตัว แนวโน้มที่ค่าที่ได้จากการสังเกตจะผิดไปจากค่าเฉลี่ย ถ้าเราสามารถทำการทดลองสัก 100 ครั้ง เราก็จะเห็นแนวโน้มดังกล่าวชัดเจนยิ่งขึ้น ในทางปฏิบัติ การกระทำดังกล่าวต้องใช้เวลาถึง 50 ปี(การทดลอง 100 ครั้ง ๆ ละ 6 เดือน) แทนที่จะรอเวลายาวขนาดนี้ เราสามารถเลียนแบบพฤติกรรมของระบบนี้ด้วยคอมพิวเตอร์ ซึ่งจะจัดการทดลองนี้ให้สำหรับเราและช่วยแจกแจงผลของการสังเกตให้

ตารางที่ 6.3 แสดงผลการทดลอง 100 ครั้งของค่า Y เมื่อ X = 2

8	10	6	9	8	8	7	7	6	10
6	9	3	5	5	4	9	7	11	6
8	8	7	8	7	4	7	10	6	6
6	6	4	7	5	8	10	4	4	11
8	7	10	6	8	8	9	7	7	5
7	4	6	7	7	7	6	10	7	10
5	6	7	6	5	9	6	7	4	6
6	5	2	8	5	8	11	5	5	5
10	7	5	7	10	5	5	5	10	7
5	8	10	5	5	5	11	7	6	8

ตารางที่ 6.3 แสดงถึงค่าที่ได้จากการสังเกตของ Y เมื่อ X = 2 หรือในแง่หนึ่งตัวเลขในตารางที่ 6.3 จะพบโดยการขยายแถวแรกของตาราง 6.1 ผ่านสดมภ์ 100 อัน ตามการทดลองที่ 1 ตัวเลขในตารางที่ 6.3 จะถูกปัดตัวเลขให้เข้าใกล้จำนวนเต็ม

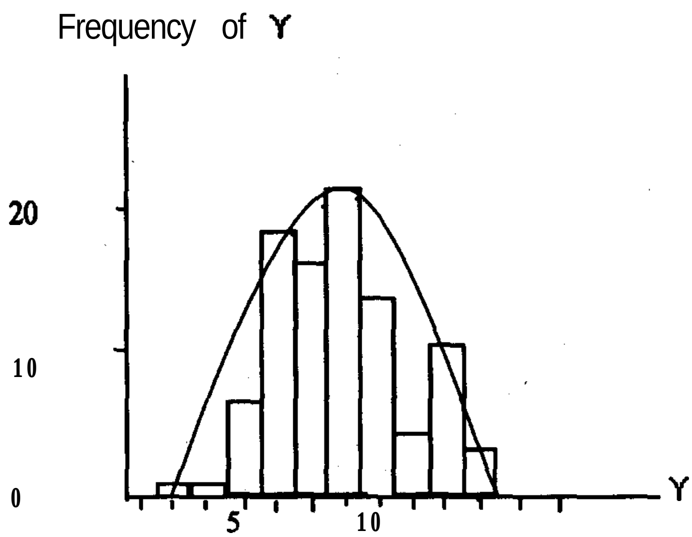
ค่าเฉลี่ยของค่า Y ที่ได้จากวิธีนี้ประมาณ 7 ตารางที่ 6.4 แสดงให้เห็นถึงความถี่ที่จะเกิด Y จาก 1-12 ในการทดลองทั้ง 100 ครั้ง เราสามารถเห็นได้ว่ามูลค่า Y ที่ได้จากการสังเกตมีแนวโน้มจะคลาดเคลื่อนออกจากค่าเฉลี่ย

ตารางที่ 6.4 แสดงความถี่ของค่า Y ในการทดลอง 100 ครั้ง

ค่าของ Y	ความถี่ของค่า Y
1	0
3	1
4	7
5	19
6	17
7	21
8	14
9	5
10	11
11	4
12	0

เราสามารถทำกระบวนการดังกล่าวซ้ำสำหรับแต่ละค่าของ X ในตารางที่ 6.1 ค่าต่าง ๆของ Y ต่อค่าของ X แต่ละตัวก็เบี่ยงเบนไปจากค่าเฉลี่ยอีกเช่นกัน

เมื่อเรามาดูตารางที่ 6.4 ในรูปกราฟ ให้แนวตั้งเป็นความถี่ของมูลค่า Y และแนวนอนเป็นมูลค่าตัวมันเอง เราจะได้รูปการกระจายความถี่ของค่าเหล่านี้ ซึ่งจะแสดงออกมาให้เห็นในรูปที่ 6.2



รูปที่ 6.2 แสดงความถี่ของมูลค่า Y

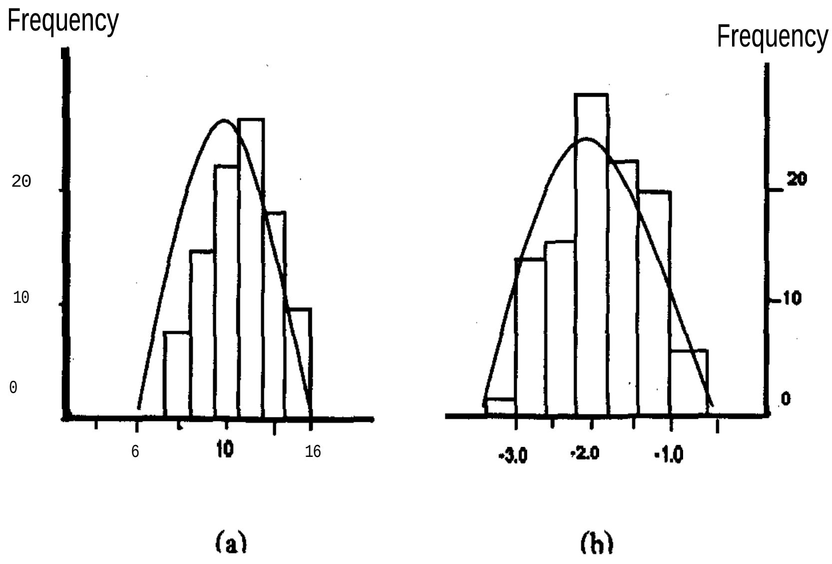
จะพบว่ามูลค่าของ Y จะเป็นไปตามการกระจายแบบปกติ เส้นโค้งที่บรรยายการกระจายชุดของข้อมูลเหล่านี้จะแสดงดังรูปที่ 6.2 ถ้าเราทำการทดลองเป็นจำนวนพันครั้งแทนที่ร้อยครั้ง การกระจายของความถี่จะเข้าใกล้เส้นโค้งการกระจายปกติ

ลักษณะของการกระจายแบบปกติจะถูกรวมอยู่ในการกระจายของค่าเฉลี่ยและค่าความแปรปรวน ในการสังเกตค่า Y ที่ X = 2 ในตารางที่ 6.3 ค่าเฉลี่ยของ Y ประมาณ 7 และค่าเบี่ยงเบนประมาณ 4 การกระจายของ Y ณ ระดับ X ต่างๆ กัน ถูกสมมติว่าเป็นแบบปกติ นอกจากนั้นยังสมมติว่ามีค่าเบี่ยงเบนที่เท่ากันในค่าเฉลี่ยต่างกันสำหรับค่า Y ที่ X =2

6.1.3 การกระจายของสัมประสิทธิ์ความถดถอย

สมมติให้ชุดข้อมูลทั้ง 100 ชุด ถูกจัดให้เป็นปัญหาสำหรับผู้ผูกขาด แสดงว่ามีการตอบสนองด้วยสัมประสิทธิ์ความถดถอยที่ถูกประมาณมาทั้ง 100 ชุด ยกตัวอย่างเช่น มูลค่า $\hat{\alpha}$ และ $\hat{\beta}$ ทั้ง 100 คู่ เราสามารถใส่ค่าในตารางได้เหมือนกับตารางที่ 6.3 และทำการกระจายความถี่ของทั้ง 2 ตัวอย่างที่ทำกับ Y ในตารางที่ 6.4 ผลก็คือ เราจะได้รูปของการกระจายของสัมประสิทธิ์เหล่านี้

การทดลองดังกล่าวมักจะใช้คอมพิวเตอร์ในการแก้ปัญหา การกระจายของผลของ $\hat{\alpha}$ แสดงในรูป 6.3a และผลของ $\hat{\beta}$ ในรูป 6.3b การกระจายเป็นแบบปกติ ค่าเฉลี่ยของ $\hat{\alpha}$ คือ 10.8 และค่าเบี่ยงเบนคือ 4.2 ค่าเฉลี่ยของ $\hat{\beta}$ คือ –1.9 โดยมีค่าเบี่ยงเบนประมาณ 0.4



รูปที่ 6.3 แสดงการกระจายของ $\hat{\alpha}$ และ $\hat{\beta}$

ดังนั้นเราสามารถพบว่าค่าประมาณของ $\hat{\alpha}$ และ $\hat{\beta}$ มีแนวโน้มที่จะเบี่ยงเบนไปจากค่าค่าหนึ่งซึ่งเรียกว่าค่าที่แท้จริงของ α และ β นี่คือค่าที่จะได้ ถ้าค่าประชากรทั้งหมดของค่าสังเกต ราคา - ปริมาณ (Price, Quantity) ใช้ในสมการถดถอยเชิงเดี่ยวอย่างใหญ่ (single gigantic regression) ขณะนี้จึงเข้าใจได้ว่าทำไมค่า $\hat{\alpha}$ และ $\hat{\beta}$ ที่ได้จากสมการจึงถือได้ว่าเป็นได้แค่ค่าประมาณ เนื่องจากค่าเหล่านั้นคิดจาก กลุ่มตัวอย่างที่เป็นจริงแต่ไม่ทราบค่าของประชากร α และ β ในความเป็นจริง เราใช้คำว่า ค่าประมาณกำลังสองน้อยที่สุด (least square estimators) ในการพูดถึงค่า $\hat{\alpha}$ และ $\hat{\beta}$ งานทางเศรษฐมิติส่วนมากเพื่อพยายามอธิบายค่าของค่าประมาณเหล่านี้

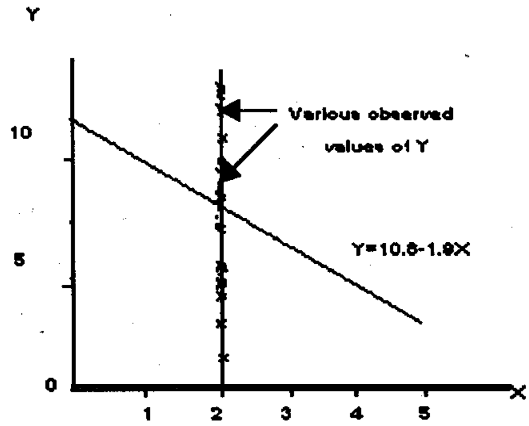
6.1.4 การกระจายของค่าผิดพลาด(The Distribution of the Error Term)

เราสามารถแปลคุณสมบัติของการกระจายของ Y ที่ได้กล่าวไปแล้วว่าเป็นคุณสมบัติของค่าผิดพลาดสำหรับค่าประชากรทั้งหมดของค่าสังเกต ราคา และ ปริมาณ ลองสมมติเพื่อความง่ายว่ากลุ่มของค่าทดลอง 100 ตัวเป็นค่าประชากรทั้งหมดของค่าสังเกต ราคา และ ปริมาณ ด้วยเหตุนี้เราสามารถระบุความสัมพันธ์ที่แท้จริงระหว่างตัวแปร ตัวอย่างเช่น สมการ

$$Y = 10.8 - 1.9X \qquad (6.1)$$

อย่างที่ได้คำนวณข้างต้น รูป 6.4 แสดงความสัมพันธ์นี้ รูป 6.4 ได้แสดงค่าสังเกตหลายค่าของ Y สำหรับ X = 2 ที่นำมาจากตาราง 6.3 เราสามารถมองการเบี่ยงเบนเกี่ยวกับเส้นถดถอย "ที่แท้จริง" ถ้าเราสามารถวัดระยะทางระหว่างค่าสังเกตแต่ละตัวของ Y จากเส้นถดถอยได้ เราสามารถ

หาค่าผิดพลาด หรือค่าที่เหลือสำหรับประชากรทั้งหมดของค่า Y เหล่านี้ ซึ่งตอบสนองกับค่าผิดพลาด e สำหรับกลุ่มตัวอย่าง การวัดค่าที่เหลือไม่ใช่สำหรับ Y ที่ x= 2 แต่สำหรับค่า Y ทุกค่าในกลุ่มประชากร เราเรียกค่าที่เหลือนี้ว่า u ในเศรษฐมิติเราเรียกว่า disturbance term



รูปที่ 6.4 แสดงค่าสังเกตของ Y เมื่อค่า X = 2

เป็นที่เห็นได้ชัดเจนจากการกระจายของค่า Y และจากการพิจารณา รูป 6.4 ว่าระยะทางระหว่างจุดใดใดจากเส้นถดถอยที่แท้จริง จะเป็นบวกมากพอๆ กับค่าลบและในภาพรวม ค่าบวกและค่าลบของ u จะหักล้างกันเองในความจริงถ้าให้ค่า Y กระจายอย่างปกติด้วยค่าเฉลี่ยที่แน่นอนและมีค่าความแปรปรวน= δ^2 ค่า u จะมีการกระจายอย่างปกติด้วยค่าเฉลี่ยที่ 0 และค่าแปรปรวนของ u คือ δ^2

6.1.5 สรุป

ผลของบทนี้คือการแสดงกลุ่มของสมมติฐานที่รูปแบบสมการเชิงเส้นใช้เป็นเกณฑ์ ให้ละเอียดยิ่งขึ้นก็คือแสดงกลุ่มของสมมติฐานให้เห็นว่าข้อมูลจากการปฏิบัติเกิดขึ้นได้อย่างไร

ค่าของตัวแปรตามในกลุ่มประชากรเกี่ยวพันกับตัวแปรอิสระดังสมการ

$$Y = \alpha + \beta X + u \tag{6.2}$$

เราสมมติว่า disturbance term u มีการกระจายอย่างปกติด้วยค่าเฉลี่ยที่ 0 และค่าแปรปรวนที่ δ^2 เรามีกลุ่มตัวอย่าง n ค่าของตัวแปรอิสระและตัวแปรตาม X_i และ Y_i ตามลำดับ (i = 1,2,…,n) ค่าสังเกตตัวอย่างเกี่ยวข้องกันดังสมการ

$$Y_I = \hat{\alpha} + \hat{\beta} X_i + e_I \tag{6.3}$$

ในขณะที่สัมประสิทธิ์ความถดถอย $\hat{\alpha}$ และ $\hat{\beta}$ เป็นการประมาณอย่างจริง แต่ค่าสัมประสิทธิ์ของประชากร α และ β ยังไม่ทราบค่าอย่างที่เราจะเห็นในส่วนที่ 6.2.2 เราอาจจะใช้ข้อมูลตัวอย่างในการคำนวณและประมาณค่าความแปรปรวนของตัวรบกวน δ^2 ภายใต้ข้อสมมติฐานข้างต้น ร่วมกับข้อสมมติฐานอื่นๆ ที่เราไม่ได้กล่าวถึงในที่นี้ ค่าประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ ของเราได้จากกลุ่มของข้อมูลจะเป็นการประมาณที่น่าพอใจอย่างที่เราต้องการ ขั้นตอนทางสถิติบางส่วนที่จะมีการแสดงให้เห็นในส่วนถัดไปเหมาะที่จะใช้สำหรับข้อมูลตัวอย่างที่จะใช้หาข้อสรุปเกี่ยวกับค่าประชากรจากกลุ่มตัวอย่างที่ถูกเลือกมา

6.2 สัมประสิทธิ์สหสัมพันธ์

6.2.1 ความหมาย

ในการวิเคราะห์สมการถดถอยแบบเส้นตรง เราจะสนใจในการอธิบายความแปรปรวนในตัวแปรตาม Y โดยการเชื่อมโยงกับเส้นตรงที่มีตัวแปรอิสระ X โดยต้องการที่จะทราบว่าสามารถอธิบายความแปรปรวนของ Y ได้มากเท่าไร และอธิบายไม่ได้เท่าไร ในความเป็นจริงเราสามารถแยกความแปรปรวนของ Y ได้เป็น 2 ส่วน คือ ส่วนที่อธิบายได้ด้วยการถดถอย (regression) และส่วนที่ไม่สามารถอธิบายได้ด้วยการถดถอย เราสามารถเขียนให้อยู่ในรูป

ความแปรปรวนของ Y = ความแปรปรวนที่อธิบายได้ + ความแปรปรวนที่อธิบายไม่ได้ (6.4)

ถ้าเราให้ความแปรปรวนของ Y ซึ่งเป็นการวัดค่าสังเกตของกลุ่มตัวอย่างของ Y แทนด้วย S_y^2 และ r^2 แสดงสัดส่วนของความแปรปรวนใน Y ที่อธิบายได้ด้วยการถดถอย เราอาจเขียนสมการ (6.4) ได้เป็น

$$S_y^2 = r^2 S_y^2 + (1-r^2)S_y^2 \qquad (6.5)$$

สูตรการคำนวณ r^2 จากกลุ่มตัวอย่าง คือ

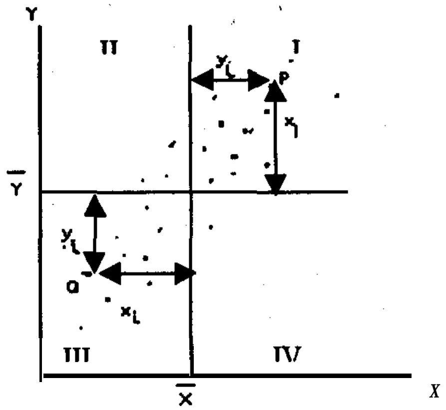
$$r^2 = \frac{(\Sigma yx)^2}{\Sigma x^2 \Sigma y^2} \qquad (6.6)$$

เรียกว่า สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) ซึ่งเป็นตัวที่แสดงความเหมาะสมในการใช้เส้นถดถอยนั้นกับค่าสังเกต โดยจะมีค่าระหว่าง 0 และ 1 โดย 1 หมายถึง เหมาะสมที่สุด ในขณะที่ 0 แสดงว่าไม่มีความสัมพันธ์แบบเส้นตรงเลยระหว่าง X และ Y

รากที่สองของ r^2 เรียกว่า สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

$$r = \frac{\Sigma yx}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

เครื่องหมายของค่า $\sum yx$ จะแสดงความสัมพันธ์ระหว่าง Y และ X สมมติว่าเราแบ่งค่า X และ Y ออกเป็นควอเดนท์ (quadrant) 4 ส่วนดังรูป 6.5 ค่า $x_i = X_i - \bar{X}$ และ $y_i = Y_i - \bar{Y}$ จะแสดงที่จุด P ซึ่งจะเห็นได้ชัดว่าที่จุด P ทั้ง x_i และ y_i เป็นบวก อย่างไรก็ตาม ที่จุด Q ซึ่งอยู่ในควอเดนท์ที่ 3 ทั้ง x_i และ y_i เป็นลบ



รูปที่ 6.5 แสดงทิศทางของสัมประสิทธิ์สหสัมพันธ์

ด้วยเหตุนี้เราจะสังเกตได้ว่า

- จุดที่อยู่ในควอเดนท์ I ค่า yx จะเป็นบวก
- จุดที่อยู่ในควอเดนท์ II ค่า yx จะเป็นลบ
- จุดที่อยู่ในควอเดนท์ III ค่า yx จะเป็นบวก
- จุดที่อยู่ในควอเดนท์ IV ค่า yx จะเป็นลบ

ด้วยเหตุนี้ ถ้ามีแนวโน้มว่าจะมีความสัมพันธ์แบบเป็นบวก (upward sloping) ระหว่าง y และ x จุดต่างๆ ส่วนมากจะอยู่ในควอเดนท์ I และ III จึงทำให้ค่า $\sum yx$ มีแนวโน้มเป็นบวก ในทางตรงกันข้าม ถ้าความสัมพันธ์เป็นแบบ downward จุดต่างๆ ส่วนมากจะอยู่ในควอเดนท์ที่ II และ IV และค่า $\sum yx$ มีแนวโน้มที่จะเป็นลบ ด้วยเหตุนี้สัมประสิทธิ์สหสัมพันธ์ r สามารถมีค่าตั้งแต่ –1 ถึง +1 โดยเครื่องหมายจะแสดงถึงความสัมพันธ์ว่าเป็นบวกหรือลบ

6.2.2 การตรวจสอบนัยสำคัญของ r

สมมติว่าเราได้ค่า r^2 เท่ากับ 0.64 สำหรับกลุ่มข้อมูลกลุ่มหนึ่ง ค่านี้จะบอกเราว่า 64 เปอร์เซ็นต์ของความแปรปรวนของตัวแปรสามารถอธิบายได้ด้วยความสัมพันธ์แบบเส้นตรงจากตัวแปรอิสระ ส่วน 36 เปอร์เซ็นต์ที่เหลือไม่สามารถอธิบายได้ แล้วเราจะพิจารณาว่าผลที่ได้นี้น่าพอใจ

หรือไม่? มีโอกาสที่ทำการสุ่มทางสถิติแล้วได้ค่านี้ จะหมายถึงการที่ Y และ X ไม่ได้มีความสัมพันธ์กันหรือไม่ การที่จะตอบคำถามนี้ได้ก็ต้องทำการตรวจสอบนัยสำคัญของค่า r

การที่จะตรวจสอบค่า r ได้นั้น เราจะต้องทำการตั้งสมมติฐานว่า X และ Y นั้นไม่ได้มีความสัมพันธ์กันในแบบเส้นตรง ซึ่งเราเรียกการตั้งสมมติฐานนี้ว่า null hypothesis ใช้แทนด้วยสัญลักษณ์ H_0 ด้วยสมมติฐานนี้ ค่า r สามารถแสดงได้ด้วยการแจกแจงแบบ t ดังต่อไปนี้

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (6.8)$$

เมื่อเราใช้ตารางการแจกแจงแบบ t ที่องศาอิสระ n - 2 และที่ค่านัยสำคัญตามที่ต้องการในการหาค่า t ค่า t ที่ได้จะเรียกว่าค่า critical value ถ้าค่า t ที่เราคำนวณหาได้จากสมการที่ (6.8) มีค่าสัมบูรณ์มากกว่า critical value เราจะปฏิเสธสมมติฐานที่ตั้งไว้ (null hypothesis)

เราสามารถแสดงให้เห็นได้ว่าการทดสอบการแจกแจงแบบ t มีค่าเท่ากับการทดสอบแบบ F ในการหาอัตราส่วนระหว่างความแปรปรวนใน Y ที่สามารถอธิบายได้ด้วยการถดถอยและส่วนที่ไม่สามารถอธิบายได้ การเขียนสูตรของ F สามารถทำได้หลายวิธี ตัวอย่างหนึ่งเช่น

$$F = \frac{\hat{\beta}^2\sum x^2}{\sum e^2 \big/ (n-2)} \qquad (6.9)$$

ตัวหารของสมการข้างต้น คือ ค่าความแปรปรวนของค่าผิดพลาด e ปกติแสดงด้วย s^2 ซึ่งเป็นค่าประมาณของความแปรปรวนของประชากรที่ไม่ทราบค่า δ^2 ค่า F ที่ได้จากสมการ (6.9) จะต้องนำมาเปรียบเทียบกับค่า critical value ของ F ที่อ่านได้จากตารางของการแจกแจงแบบ F ที่ระดับความเชื่อมั่นที่กำหนด สำหรับองศาอิสระที่ (1,n-2) ถ้าค่า F ที่คำนวณได้มีค่ามากกว่าค่า critical value ของ F เราจะปฏิเสธ null hypothesis

6.3 การตรวจสอบนัยสำคัญของ $\hat{\alpha}$ และ $\hat{\beta}$

เราได้เห็นตัวอย่างที่แสดงว่าค่าประมาณแบบกำลังสองน้อยที่สุด $\hat{\alpha}$ และ $\hat{\beta}$ มีการแจกแจงแบบปกติ ในความเป็นจริง เราอาจจะเพิ่มเติมอีกได้ว่า ถ้าสมมติฐานที่ตั้งไว้ผ่านการทดสอบ โดยเฉพาะถ้าค่าผิดพลาดมีการแจกแจงด้วยค่าเฉลี่ยที่ 0 และความแปรปรวนคงที่ δ^2 ค่าประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ ก็จะมีการแจกแจงแบบปกติด้วย ด้วยค่าเฉลี่ย α และ β ตามลำดับ ค่าความแปรปรวน $\delta^2(\sum X^2/n\sum x^2)$ และ $\delta^2/\sum x^2$ ตามลำดับ แน่นอนว่าในการคำนวณค่าความแปรปรวนเหล่านี้ เราไม่ทราบค่า δ^2 จะต้องแทนค่าด้วยค่าประมาณของ δ^2 คือด้วย S^2 ตามที่กล่าวไปแล้ว เรายังสามารถใส่

รากที่สองเพื่อให้ค่าประมาณของ standard errors ของค่าสัมประสิทธิ์ของการถดถอย 2 ค่า ทำให้เราได้

$$\mathrm{SE}(\hat{\alpha}) = s\sqrt{\frac{\sum X^2}{n\sum x^2}} \qquad (6.10)$$

และ

$$\mathrm{SE}(\hat{\beta}) = \frac{S}{\sqrt{\sum x^2}} \qquad (6.11)$$

สมมติฐานหลักที่เราตั้งไว้ว่า Y และ X ไม่มีความสัมพันธ์กันในแบบเส้นตรงอาจแสดงได้ในรูป $\beta = 0$ เราสามารถทำการตรวจสอบแบบ t ในค่าประมาณ $\hat{\beta}$ ด้วยการหาอัตราส่วนของค่าแตกต่างระหว่างค่าประมาณและค่า standard error ของข้อมูล ดังนั้นเราอาจจะเขียนได้ว่า

$$t = \frac{\hat{\beta} - \beta}{S/\sqrt{\sum x^2}} \qquad (6.12)$$

และเนื่องจากสมมติฐานหลักมีว่า $\beta = 0$ สมการ (6.12) อาจจะเขียนได้ว่า

$$t = \frac{\hat{\beta}}{S/\sqrt{\sum x^2}} \qquad (6.13)$$

เมื่อ t มีค่าองศาอิสระที่ n-2 และเช่นกันว่าเราจะปฏิเสธสมมติฐานหลักถ้าค่า t ที่คำนวณได้เกินกว่าค่า critical value ของ t ที่ได้จากระดับความมั่นใจที่กำหนด ในความเป็นจริงการทดสอบแบบ t นี้มีความเหมือนกันทุกประการกับการหาสัมประสิทธิ์สหสัมพันธ์ r ที่ได้แสดงไปข้างต้น

6.4 การทดสอบนัยสำคัญของกลุ่มตัวอย่างที่เป็นตัวเลข (Numerical Example)

6.4.1 การทดสอบแบบ t ของ r

ลองย้อนกลับไปยังข้อมูลที่เป็นตัวเลขของการทดลองที่ 1 ของปัญหาแบบผูกขาด (monopolist 's problem) ที่ได้ศึกษามาแล้วในบทก่อน โดยการใช้ผลจากตาราง 5.2 และ 5.3 เราคำนวณค่าแรกของสัมประสิทธิ์สหสัมพันธ์และสัมประสิทธิ์การตัดสินใจ โดยการใช้สมการ(6.7) เราจะได้

$$r = \frac{-1.9}{\sqrt{10 \times 52}} = -0.833 \qquad (6.14)$$

ด้วยเหตุนี้ $r^2 = 0.694$ (6.15)

หรือเราอาจกล่าวได้ว่า สมการ (6.15) แสดงให้เห็นว่า 69.4 เปอร์เซ็นต์ของความแปรปรวนในปริมาณที่ต้องการในกลุ่มตัวอย่างนี้มีความสัมพันธ์แบบเส้นตรงกับราคา จากสมการ (6.8) เราจะได้

$$t = \frac{-0.833\sqrt{6-2}}{\sqrt{1-0.694}} = -3.01 \quad (6.16)$$

จากตารางการแจกแจงแบบ t เราจะได้ค่า critical value ที่ 2.776 สำหรับองศาอิสระที่ 4 และความเชื่อมั่น 95 เปอร์เซ็นต์ เนื่องจากค่า t ที่คำนวณได้จากสมการ (6.16) เกินกว่าค่าสัมบูรณ์ของ critical value เราจึงปฏิเสธสมมติฐานหลักที่ว่าไม่มีความสัมพันธ์แบบเส้นตรงระหว่างราคาและปริมาณ หรือเราอาจกล่าวได้ว่า โอกาสที่ค่า t ที่คำนวณได้จะเพิ่มขึ้นเนื่องจากค่าผันผวนจากการสุ่มจะน้อยกว่า 5 เปอร์เซ็นต์

6.4.2 การทดสอบแบบ F

เพื่อที่จะใช้สมการที่ (6.9) คำนวณหา F เราต้องคำนวณค่าความแปรปรวนที่เหลือ (residual variance) ก่อน ค่าผลรวมของกำลังสองของความผิดพลาด $\sum e^2$ เราสามารถคำนวณได้จาก

$$\sum e^2 = \sum y^2 - \sum \hat{y}^2 \quad (6.17)$$

ด้วยการยกกำลัง การรวมและจัดรูปสมการที่ (5.5) ใหม่ ปริมาณ $\sum \hat{y}^2$ อาจคำนวณได้โดยสูตร

$$\sum \hat{y}^2 = \hat{\beta}^2 \sum x^2 \quad (6.18)$$

ถ้าเราใช้สมการที่(6.18) กับข้อมูลเราจะได้

$$\sum \hat{y}^2 = (-1.9)^2 \times 10 = 36.09 \quad (6.19)$$

และแทนค่าในสมการที่ (6.17) จะได้

$$\sum e^2 = 52-36.09 = 15.91 \quad (6.20)$$

ในขณะที่ $S^2 = 15.91/4 = 3.98$ (6.21)

และถ้าใช้สมการที่ (6.9) จะได้

$$F = \frac{36.09}{3.98} = 9.07 \quad (6.22)$$

จากตารางการแจกแจงแบบ F องศาอิสระที่ (1, 4) มีค่า criticalvalue ที่ 7.71 สำหรับความเชื่อมั่นที่ 95 เปอร์เซ็นต์ และ 21.2 สำหรับความเชื่อมั่นที่ 99 เปอร์เซ็นต์ ด้วยเหตุนี้เราจะปฏิเสธสมมติฐานหลักที่ระดับความเชื่อมั่นที่ได้กล่าวไปแล้ว แต่เราจะไม่ปฏิเสธถ้าระดับความเชื่อมั่นสูงกว่านั้น

6.4.3 การทดสอบ t ของ $\hat{\beta}$

ค่า standard errors ของค่าสัมประสิทธิ์ถดถอยโดยประมาณอาจจะคำนวณได้โดยใช้สมการ (6.10) และ (6.11)

แทนค่าจะได้ $\quad SE(\hat{\alpha}) \quad = \quad 1.99\sqrt{64/(6\times 10)} \quad = 2.509 \qquad (6.23)$

และ $\quad SE(\hat{\beta}) \quad = \quad 1.99/G- \quad = 0.631 \qquad (6.24)$

ด้วยเหตุนี้ เพื่อที่จะทดสอบสมมติฐานหลักที่ $\beta = 0$ เราจะใช้สมการ (6.13) จะได้

$$t = -1.9/0.631 = -3.01 \qquad (6.25)$$

ซึ่งตรงกับสมการที่ (6.16)

คำถามท้ายบทที่ 6

1. จงใช้ตัวเลขจากแบบฝึกหัดข้อที่ 4 บทที่ 5 คำนวณสัมประสิทธิ์การตัดสินใจและสัมประสิทธิ์สหสัมพันธ์
2. ตรวจสอบนัยสำคัญของ r จากข้อ 1
3. คำนวณค่า standard errors ของค่าสัมประสิทธิ์การถดถอยโดยประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ ในตัวอย่างข้างต้น
4. ใช้ตัวอย่างข้างต้นแสดงว่าการทดสอบแบบ t ที่สมมติฐานหลัก คือ $\beta = 0$ จะมีค่าเท่ากับการทดสอบแบบ t ของ r ภายใต้สมมติฐานเดียวกัน
5. คำนวณค่าความแปรปรวนของความผิดพลาดในตัวอย่างข้างต้น โดยใช้สูตรดังสมการที่ (6.9)
6. หน่วยงานรัฐบาลได้ทำโครงการอิสระเกี่ยวกับราคาการส่งออกเนื้อวัวในอนาคต มีการประมาณว่าในปี 2540 ราคาจะเท่ากับ 8 บาท ต่อหน่วยและในปี 2542 ราคาจะเท่ากับ 12 บาท จงวิจารณ์การใช้และข้อจำกัดของแบบจำลองการถดถอยในแบบฝึกหัดข้อ 4 บทที่ 5 ในการพิจารณาราคาการส่งออกในอนาคต