บทที่ 11
สหสัมพันธ์ และการถดถอย (Correlation and Regression)

11.1 บทนำ

ในบทที่แล้วเราได้ศึกษาถึงตัวแปรแต่ละตัว สำหรับในบทนี้เราจะศึกษาตัวแปร 2 ตัว ให้เป็นตัวแปร X กับตัวแปร Y โดยเราจะดูว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันหรือไม่ และถ้าตัวแปรทั้ง 2 มีความสัมพันธ์กัน จะมีความสัมพันธ์กันอย่างไร และสามารถเขียนสมการแสดงความสัมพันธ์ได้หรือไม่ ซึ่งเป็นเรื่องของการศึกษาสหสัมพันธ์และการถดถอย เช่นจะศึกษาดูว่าน้ำหนักของคนจะมีความสัมพันธ์กับส่วนสูงหรือไม่ หรือจะศึกษาดูว่าคนที่เรียนวิชาคณิตศาสตร์เก่งจะเรียนวิชาสถิติเก่งหรือไม่ เป็นต้น

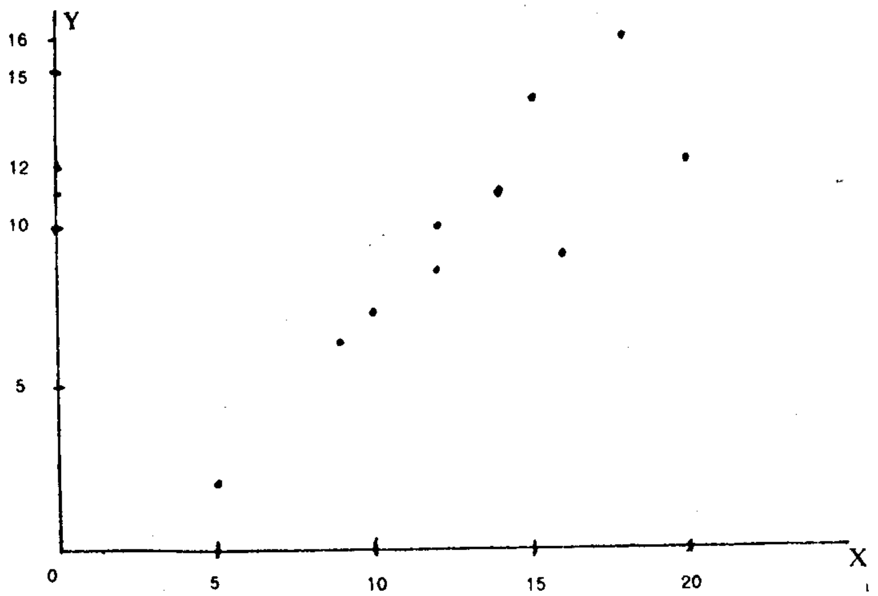
ในการที่จะตรวจดูว่าตัวแปรทั้ง 2 นั้นมีความสัมพันธ์กันหรือไม่นั้น เราอาจทำอย่างคร่าว ๆ ได้ โดยการ plot ค่าสังเกตของตัวแปรทั้ง 2 ทีละคู่ลงบนกราฟที่มีแกน X แทนตัวแปรตัวหนึ่ง และแกน Y ซึ่งแทนตัวแปรอีกตัวหนึ่ง กราฟที่ได้นี้เรียกว่า Scattergram หรือ Scatter diagram ตัวอย่างเช่น จะศึกษาคะแนนสอบวิชาคณิตศาสตร์และวิชาสถิติของนักศึกษาภาควิชาสถิติ 10 คน ดังนี้

ให้ X เป็นคะแนนสอบวิชาคณิตศาสตร์

Y เป็นคะแนนสอบวิชาสถิติ

X	20	18	16	15	14	12	12	10	8	5
Y	12	16	10	14	12	10	9	8	7	2

เราจะดูความสัมพันธ์ระหว่างตัวแปรทั้ง 2 ได้อย่างคร่าว ๆ โดย plot จุด (X, Y) ใด ๆ ลงบน X – Y plane ดังนี้



Scatter diagram แสดงความสัมพันธ์ระหว่างตัวแปร X และ Y

11.2 สหสัมพันธ์ (Correlation)

ในการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร เราเรียกว่า การวิเคราะห์ **สหสัมพันธ์** (Correlation Analysis) จุดมุ่งหมายในการศึกษาสหสัมพันธ์คือ เราจะพิจารณาดูว่า ตัวแปรทั้ง 2 มีความสัมพันธ์กันหรือไม่ และถ้ามีความสัมพันธ์กัน ขนาดความสัมพันธ์จะมีมากน้อยแค่ไหน ซึ่งการวัดขนาดของความสัมพันธ์ระหว่างตัวแปรทั้ง 2 เราใช้สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) เป็นตัววัด สหสัมพันธ์ระหว่างตัวแปรอาจจะเป็นเชิงเส้นตรง (linear) ถ้าทุกจุด (X, Y) บน Scatter diagram เรียงกันเป็นเส้นตรงหรือเกือบเป็นเส้นตรง หรืออาจจะมีสหสัมพันธ์ไม่เป็นเชิงเส้นตรง (Nonlinear) ก็ได้ ในที่นี้จะศึกษาเฉพาะกรณีเชิงเส้นตรง (linear) เท่านั้น

สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

ใช้สัญลักษณ์แทนด้วย r

การพิจารณาสัมประสิทธิ์สหสัมพันธ์เราจะดู 2 อย่างด้วยกันคือ

1. **ดูเครื่องหมายของ** r ว่าเป็นเครื่องหมายบวก (+) หรือเครื่องหมายลบ (−) ถ้ามีเครื่องหมาย + แสดงว่ามีความสัมพันธ์ไปในทิศทางเดียวกัน คือถ้าค่าของตัวแปรตัวหนึ่งเพิ่มค่าของอีกตัวแปรก็จะเพิ่มตามด้วย และเครื่องหมายลบ แสดงถึงความสัมพันธ์ในทิศทางตรงกันข้าม คือถ้าตัวแปรตัวหนึ่งเพิ่ม ค่าของอีกตัวแปรหนึ่งจะลดลง

2. **ดูขนาดของตัวเลข** ถ้า $r \geqslant 0.8$ ถือว่าขนาดความสัมพันธ์มีมาก ถ้า r มีค่าประมาณ 0.5 ถือว่ามีความสัมพันธ์ปานกลาง ถ้า $r \leqslant 0.3$ ถือว่ามีความสัมพันธ์ระดับต่ำหรือเกือบไม่มีเลย

คุณสมบัติของ r

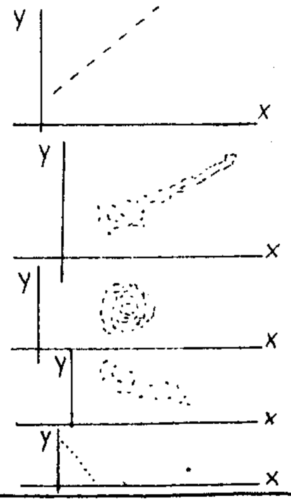
1. ค่าของ r จะอยู่ในพิสัย −1.00 ถึง +1.00 คือ

$$-1.00 \leqslant r \leqslant +1.00$$

3. ความสัมพันธ์ทางบวก (r มีค่า +) ระหว่างตัวแปรทั้ง 2 หมายถึง ถ้าค่าของตัวแปรหนึ่งสูง ค่าของอีกตัวแปรหนึ่งก็สูงด้วย หรือถ้าค่าของตัวแปรตัวหนึ่งต่ำ ค่าของอีกตัวแปรหนึ่งก็จะต่ำด้วย

3. ความสัมพันธ์ทางลบ (r มีค่า −) ระหว่างตัวแปรทั้ง 2 หมายถึง ถ้าค่าของตัวแปรตัวหนึ่งสูง ค่าของตัวแปรอีกตัวหนึ่งจะต่ำ

4. ถ้า r = 0 หมายถึง ไม่มีความสัมพันธ์ระหว่างตัวแปรทั้ง 2 เลย ดังรูป

ค่าของ r	Description of linear relationship	Scatter diagram
+1.00	Perfect positive relationship	
ประมาณ +.70	Moderate, positive relationship	
.00	No relationship	
ประมาณ −.70	Moderate, negative relationship.	
−1.00	Perfect, negative relationship	

การคำนวณหาค่า r

เราคำนวณค่าของ r ได้จากสูตรดังต่อไปนี้

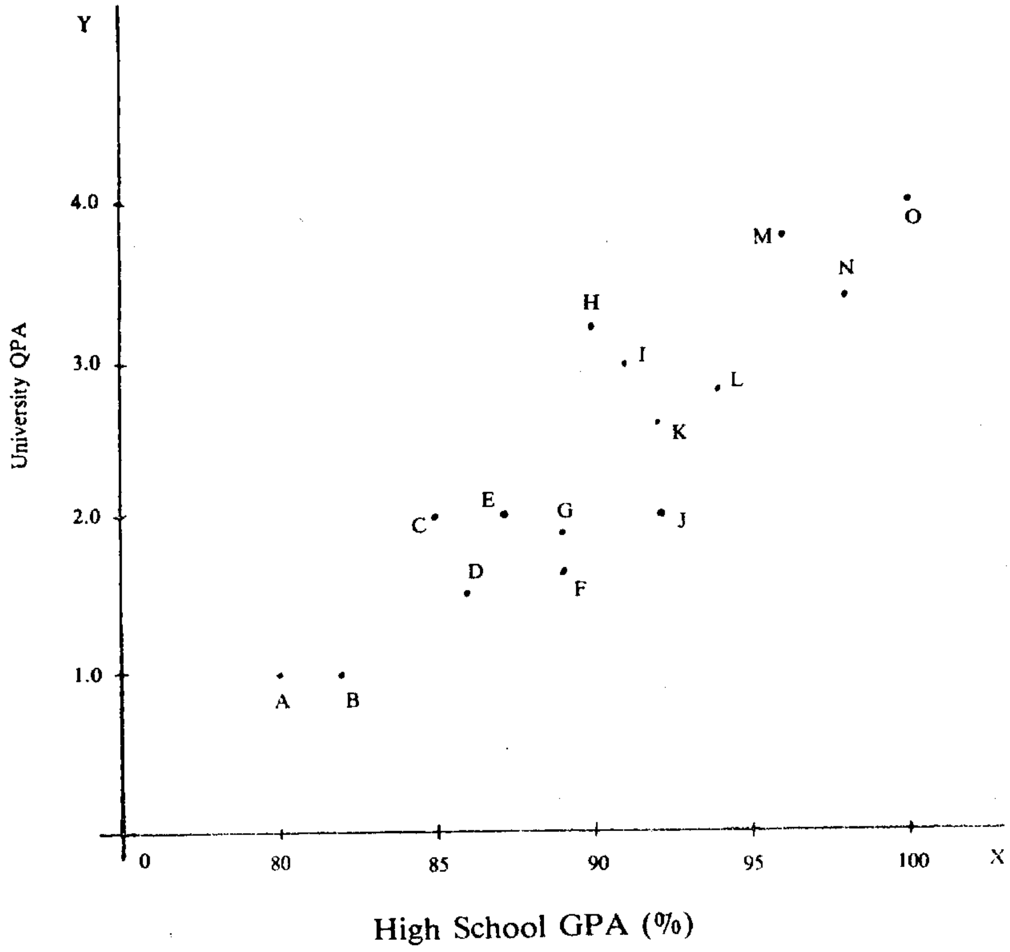
$$r = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[n(\Sigma X^2) - (\Sigma X)^2][n(\Sigma Y^2) - (\Sigma Y)^2]}}$$

ตัวอย่าง 11.2.1

จากการศึกษาผลการเรียนของนักศึกษา 15 คน เพื่อดูว่าผลการเรียนในระดับมัธยมกับผลการเรียนในระดับมหาวิทยาลัยจะมีความสัมพันธ์กันหรือไม่ ได้ข้อมูลมาดังนี้

จำนวน	ชื่อนักศึกษา	High School grade point average (GPA) (%)	University quality point average (QPA)
1	A	80	1.0
2	B	82	1.0
3	C	84	2.1
4	D	85	1.4
5	E	87	2.1
6	F	88	1.7
7	G	88	2.0
8	H	89	3.5
9	I	90	3.1
10	J	91	2.4
11	K	91	2.7
12	L	92	3.0
13	M	94	3.9
14	N	96	3.6
15	O	98	4.0

ถ้าเรานำมา plot ลงใน X – Y plane เพื่อดูความสัมพันธ์อย่างคร่าว ๆ จะได้ Scatter diagram ข้างล่างนี้



จาก Scatter diagram จะเห็นได้ว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันในทางบวก (Positive correlation) และเราสามารถหาขนาดของความสัมพันธ์ r ได้ดังนี้

จำนวน	ชื่อ	X_i GPA	Y_i QPA	X_iY_i	X_i^2	Y_i^2
1	A	80	1.0	80.0	6400	1.00
2	B	82	1.0	82.00	6724	1.00
3	C	84	2.1	176.4	7056	4.41
4	D	85	1.4	119.0	7225	1.96
5	E	87	2.1	182.7	7569	4.41
6	F	88	1.7	149.6	7744	2.89
7	G	88	2.0	176.0	7744	4.00
8	H	89	3.5	311.5	7921	12.25
9	I	90	3.1	279.0	8100	9.61
10	J	91	2.4	218.4	8281	5.76
11	K	91	2.7	245.7	8281	7.29
12	L	92	3.0	276.0	8464	9.00
13	M	94	3.9	366.6	8836	15.21
14	N	96	3.6	345.6	9216	12.96
15	O	98	4.0	392.0	9604	16.00
Total		$\Sigma X_i = 1,335$	$\Sigma Y_i = 3.75$	$\Sigma X_iY_i = 3400.5$	$\Sigma X_i^2 = 119,155$	$\Sigma Y_i^2 = 107.75$

จากสูตร $$r = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[n(\Sigma X^2) - (\Sigma X)^2][n(\Sigma Y^2)(\Sigma X)^2]}}$$

แทนค่า

$$\therefore\ r = \frac{15(3400.5) - 1335(37.5)}{\sqrt{[15(119155) - (1335)^2][15(107.75) - (37.5)^2]}}$$

$$= +0.90$$

แสดงว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันในระดับสูง

การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์

ให้ r = สัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง

ρ = สัมประสิทธิ์สหสัมพันธ์ของประชากร

เราต้องการจะทดสอบดูว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันจริงหรือไม่ โดยตั้งสมมติฐานว่า สัมประสิทธิ์สหสัมพันธ์ที่แท้จริง(ของประชากร) เป็น 0 หรือเขียนว่า $H_0 : \rho = 0$ สำหรับขั้นตอนในการทดสอบมีดังนี้

1. ตั้งสมมติฐาน

$H_0 : \rho = 0$

$H_a : \rho \neq 0$

2. กำหนด α และ n

3. ตัวสถิติที่ใช้ทดสอบคือ

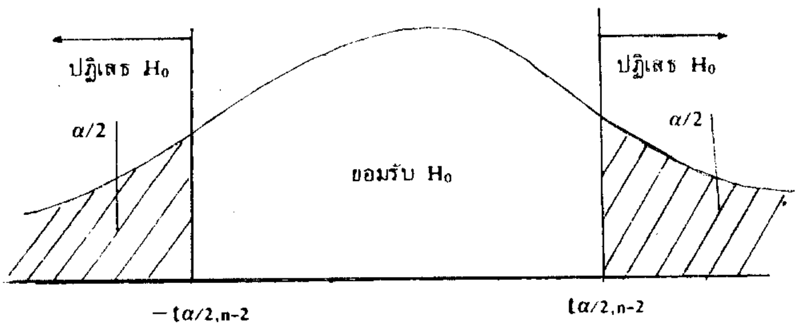
$$t = \frac{r-0}{\sqrt{\frac{1-r^2}{(n-2)}}}$$ มีการแจกแจงแบบ t ที่มี $df = n-2$

หรือ $$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

4. CR: ปฏิเสธ H_0 ถ้า

$t_c < -t_{\alpha/2,n-2}$ หรือ $t_c > t_{\alpha/2,n-2}$

หรือ $|t_c| > t_{\alpha/2,n-2}$



5. คำนวณหา t_c จาก

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

โดยที่ $$r = \frac{n(\Sigma XY)-(\Sigma X)(\Sigma Y)}{\sqrt{[n(\Sigma X^2)-(\Sigma X)^2][n(\Sigma Y^2)-(\Sigma Y)^2]}}$$

6. สรุปผล

   1. ถ้า t_c ตกอยู่ใน CR จะปฏิเสธ H_0 และยอมรับ H_a
   2. ถ้า t_c ตกอยู่นอก CR จะยอมรับ H_0

ตัวอย่างที่ 11.2.2

สุ่มตัวอย่าง $n = 24$ ได้ค่า $r = 0.50$ จงทดสอบสมมติฐาน

$H_0 : \rho = 0$ ที่ $\alpha = .01$

1. $H_0 : \rho = 0$

   $H_a : \rho \neq 0$

2. $\alpha = .01$, $n = 24$

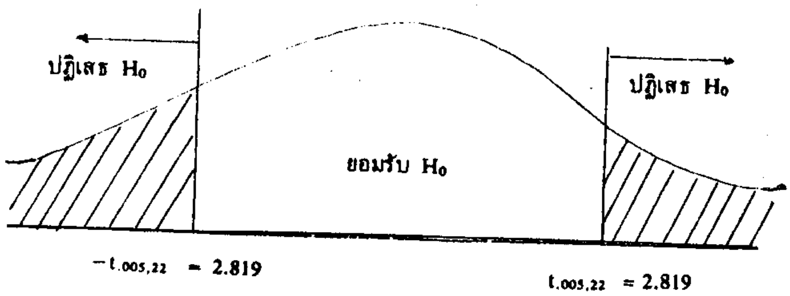
3. ตัวสถิติที่ใช้ทดสอบ คือ

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$df = 24-2 = 22$

4. CR: ปฏิเสธ H_0 ถ้า $|t_c| > t_{.005,22}$

   คือ $|t_c| > 2.819$



5. คำนวณหาค่า t_c จากสูตร

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$= \frac{.50\sqrt{24-2}}{\sqrt{1-(.50)^2}} = \frac{.50\sqrt{22}}{\sqrt{1-.25}}$$

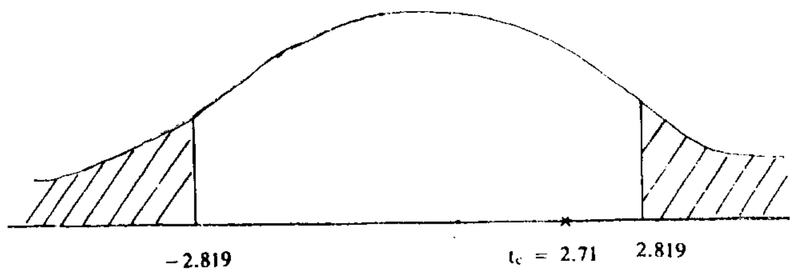
$$= \frac{(.50)(4.69)}{\sqrt{.75}} = \frac{2.3450}{0.8660}$$

$$= 2.71$$

6. สรุปผล $\because$ t_c ตกอยู่นอก CR

$\therefore$ ยอมรับ $H_0 : \rho = 0$

สรุปได้ว่าไม่มีสหสัมพันธ์ในประชากรทั้ง 2



ตัวอย่างที่ 11.2.3

จากข้อมูลในตัวอย่าง 11.2.1 จงทดสอบ $H_0 : \rho = 0$ ที่ $\alpha = .05$

1. $H_0 : \rho = 0$

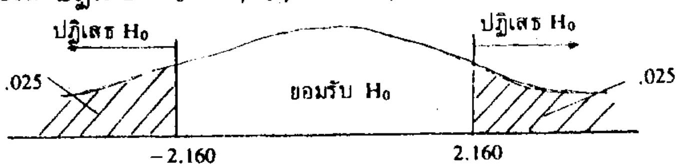
   $H_a : \rho \neq 0$

2. $\alpha = .05$, $n = 15$

3. ตัวสถิติที่ใช้ทดสอบคือ $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

   $df = 15 - 2 = 13$

4. CR: ปฏิเสธ H_0 ถ้า $|t_c| > t_{.025,13}$ คือ $|t_c| > 2.160$



5. คำนวณค่า t_c จาก $t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

เมื่อ $r = 0.90$

$$\therefore t_c = \frac{.90\sqrt{15-2}}{\sqrt{1-(.90)^2}}$$

$$= \frac{(.90)(3.61)}{\sqrt{1-.81}} = \frac{(.90)(3.61)}{\sqrt{0.91}}$$

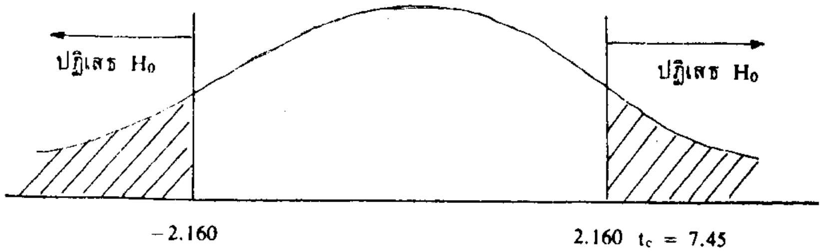
$$= \frac{3.249}{0.436} = 7.45$$

6. สรุปผล $\because$ t_c ตกอยู่ใน CR

$\therefore$ ปฏิเสธ H_0 ที่ว่า $\rho = 0$

และยอมรับ $H_a : \rho \neq 0$

แสดงว่ามีสหสัมพันธ์ในประชากรทั้ง 2



11.3 การถดถอย (Regression)

ในการศึกษาความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งซึ่งเรียกว่าตัวแปรตาม (Dependent Variable ; Y) และตัวแปรอื่น ๆ (อาจจะมี 1 ตัวแปรหรือมากกว่า) ซึ่งเราเรียกว่าตัวแปรอิสระ (Independent variable ; X) ว่า ตัวแปรทั้ง 2 มีความสัมพันธ์กันอย่างไร และความสัมพันธ์นั้นสามารถเขียนได้ในรูปของตัวแบบทางคณิตศาสตร์ (Mathematical model) เพื่อประโยชน์ในการทำนาย (Prediction) ค่าของตัวแปรหนึ่งจากตัวแปรอื่น ๆ การศึกษาปัญหานี้อาศัยหลักการวิเคราะห์เกี่ยวกับการถดถอย (Regression analysis)

รูปของตัวแบบทางคณิตศาสตร์ที่ง่ายที่สุดของความสัมพันธ์ระหว่างตัวแปรทั้ง 2 คือตัวแปร X กับตัวแปร Y เรียกว่า **ตัวแบบถดถอยเชิงเส้นแบบเชิงเดี่ยว (Simple linear Regression Model)** หรือเขียนย่อ ๆ ว่า **ตัวแบบถดถอยเชิงเส้น (Linear Regression Model)** ตัวแบบนี้มีรูปฟอร์มเป็น

$$Y_i = \beta_1 + \beta_1 X_i + \in_i \qquad \ldots\ldots\ldots\ldots 1$$

เมื่อ Y_i เป็นตัวแปรตาม (dependent variable)

X_i เป็นตัวแปรอิสระ (Independent variable)

β_1 , β_2 เป็นพารามิเตอร์ของการถดถอย (Regression parameter) ที่ไม่ทราบค่า

$\in_i$ เป็นตัวคลาดเคลื่อนซึ่งไม่สามารถวัดและสังเกตได้

ข้อสมมติเบื้องต้นของตัวแบบถดถอยเชิงเส้นชนิดเชิงเดี่ยว มีดังนี้

1. $E(\in_i) = 0$ และ $V(\in_i) = \sigma^2$
2. $\in_i$ และ $\in_j$ ของข้อมูลตัวที่ i และตัวที่ j $(i \neq j)$ ไม่มีความสัมพันธ์ (Uncorrelated) กัน
3. β_1 และ β_2 ในตัวแบบเป็นพารามิเตอร์ที่ไม่ทราบค่า ซึ่งจะต้องทำการประมาณค่าจากตัวอย่าง และให้ X_i เป็นตัวแปรแบบกำหนดค่าได้หรือแบบคงที่

ข้อสมมติฐานต่าง มีความหมายดังนี้

1. ถ้าสังเกตค่าของ Y ที่ i จะเห็นว่า Y_i ประกอบด้วย 2 ส่วนคือ ค่าคงที่ (Constant) และตัวแปรเชิงสุ่ม (random variable) ดังนี้ จาก

$$Y_i = \underbrace{\beta_1 + \beta_2 X_i}_{\text{constant}} + \underbrace{\in_i}_{\text{random variable}}$$

2. จาก $Y_i = \beta_1 + \beta_2 X_i + \in_i$

ถ้าใส่ค่า Expected จะได้

$$E(Y_i) = E[\beta_1 + \beta_2 X_i + \in_i]$$
$$= \beta_1 + \beta_2 X_i + E(\in_i)$$

แต่ $E(\in_i) = 0$

$$\therefore E(Y_i) = \beta_1 + \beta_2 X_i$$

นั่นคือค่าเฉลี่ยแบบมีเงื่อนไขของ Y เมื่อกำหนด X ให้ คือ $\mu_{y/x}$ หรือ $E(Y/X)$ จะมีค่าเท่ากับ $\beta_1 + \beta_2 X_i$

ดังนั้น ฟังก์ชันการถดถอย (regression function) ซึ่งสอดคล้องกับตัวแบบการถดถอย (regression model) คือ

$$E(Y) = \beta_1 + \beta_2 X$$

3. จาก

$$Y_i = \beta_1 + \beta_2 X_i$$
$$\therefore V(Y_i) = V(\beta_1 + \beta_2 X_i + \in_i)$$
$$= V(\beta_1 + \beta_2 X_i) + V(\in_i)$$
$$= 0 + V(\in_i)$$

แต่ $V(\in_i) = \sigma^2$

$$\therefore V(Y_i) = \sigma^2$$ ด้วย

นั่นคือ $\sigma^2_{y/x} = \sigma^2$

จาก $Y_i = \beta_1 + \beta_2 X_i + \in_i$

และ $E(Y_i) = \beta_1 - \beta_2 X_i$ ซึ่งให้ค่าเฉลี่ยของ Y สำหรับแต่ละค่าของ X สมการนี้เรียกว่า เส้นถดถอยประชากร (Population Regression line) ซึ่งเขียนใหม่ได้เป็นสมการเชิงเส้นถดถอย

$$Y_i = \beta_1 + \beta_2 X_i$$

เมื่อ β_1 เป็นจุดตัดแกนของเส้นซึ่งจะวัดค่าเฉลี่ยของ Y เมื่อ X = 0

β_2 เป็นความชันของเส้นถดถอย ซึ่งแสดงถึงอัตราการเปลี่ยนแปลงของค่า Y ต่อ 1 หน่วยค่า X ที่เปลี่ยนไป และเรียก β_2 ว่าเป็นสัมประสิทธิ์การถดถอย (Regression coefficient)

เนื่องจากค่าพารามิเตอร์เหล่านี้เราไม่ทราบ ดังนั้นเราจึงไม่ทราบเส้นถดถอยประชากร เราจึงใช้วิธีการสุ่มตัวอย่างแล้วทำการประมาณค่าพารามิเตอร์ β_1 และ β_2 เราก็จะได้เส้นถดถอยของตัวอย่าง (Sample Regression line) ซึ่งเป็นค่าประมาณของเส้นถดถอยประชากร

ถ้าให้ β_1 และ β_2 เป็นตัวประมาณค่าของ β_1 และ β_2 เราจะได้เส้นถดถอยตัวอย่างดังนี้

$$Y_i = \beta_1 + \beta_2 X_i$$

หรือ
$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

เมื่อ e_i เป็นความคลาดเคลื่อน ซึ่งเป็นผลต่างของ Y_i กับ $\hat{Y}_i$ โดยทั่วไป e_i จะต่างจาก ϵ_i และ e_i เป็นค่าประมาณของ ϵ_i

การประมาณค่าของพารามิเตอร์ β_1 และ β_2

มีหลายวิธีด้วยกัน แต่จะขอกล่าวถึงวิธีการประมาณค่า β_1 และ β_2 โดยวิธีกำลังสองต่ำสุด (Least Square Method) ซึ่งคือการทำให้ผลรวมของความเบี่ยงเบนกำลังสอง (Squared Deviation) ของค่าสังเกตที่ห่างจากค่าเฉลี่ยของมันให้มีค่าน้อยที่สุด

โดยวิธีกำลังสองต่ำสุด เราจะหาค่า β_1 และ β_2 ได้ดังนี้

$$\beta_2 = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$$

$$\beta_1 = \overline{Y} - \beta_2 \overline{X}$$

ซึ่งจะได้สมการถดถอยเป็น $Y_i = \beta_1 + \beta_2 X_i$

ตัวอย่าง 11.3.1 จากข้อมูลเกี่ยวกับจำนวนไมล์ที่ใช้กับราคาขายของรถรุ่น1975 ยี่ห้อหนึ่งเป็นดังนี้

ค่าสังเกต	จำนวนไมล์ที่ใช้ (หน่วยเป็นพัน) X	ราคาขาย (ดอลล่า) Y	XY	X^2	Y^2
1	40	$ 1,000	40,000	1,600	1,000,000
2	30	1,500	45,000	900	2,250,000
3	30	1,200	36,000	900	1,440,000
4	25	1,800	45,000	625	3,240,000
5	50	800	40,000	2,500	640,000
6	60	1,000	60,000	3,600	1,000,000
7	65	500	32,500	4,225	250,000
8	10	3,000	30,000	100	9,000,000
9	15	2,500	37,500	225	6,250,000
10	20	2,000	40,000	400	4,000,000
11	55	800	44,000	3,025	640,000
12	40	1,500	60,000	1,600	2,250,000
13	35	2,000	70,000	1,225	4,000,000
14	30	2,000	60,000	900	4,000,000
รวม	$\Sigma X = 505$	$\Sigma Y = 21{,}6000$	$\Sigma XY = 640{,}000$	$\Sigma X^2 = 21{,}825$	$\Sigma Y^2 = 39{,}960{,}000$

จากตารางจะได้

$$\beta_2 = \frac{14(640{,}000)-(505)(21{,}600)}{14(21{,}825)-(505)^2}$$

$$= \frac{8{,}960{,}000 - 10{,}908{,}000}{305{,}550 - 255{,}025}$$

$$= \frac{-1{,}948{,}000}{50{,}525} = -38.56$$

$$\therefore \beta_1 = \bar{Y} - \beta_2\bar{X}$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{21,600}{14}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{505}{14}$$

$$\therefore \beta_1 = \frac{21,600}{14} - (-38.56)(\frac{505}{14})$$

$$= \frac{21,600-(-38.56)(505)}{14}$$

$$= \frac{40,979.14}{14} = 2,934$$

∴ สมการถดถอยคือ

$$\hat{Y} = 2930 - 38.56X$$

จากสมการถดถอยนั้น ถ้าเราทราบจำนวนไมล์ที่ใช้เราก็สามารถพยากรณ์ค่าของราคาขายได้ เช่น

ถ้าจำนวนไมล์ที่ใช้ = 20,000 ไมล์ ($X = 20$)

เราจะพยากรณ์ค่าของราคาขายได้ $= 2934-(38.56)(20)$

$= \$\ 2,163$

ความคลาดเคลื่อนมาตรฐานของค่าประมาณ (Standard Error of Estimate)

ความคลาดเคลื่อนมาตรฐานของค่าประมาณ หาได้จาก

$$S_{y/x} = \sqrt{\frac{\Sigma(Y_i-\hat{Y}_i)^2}{n-2}}$$

แต่สูตรนี้ไม่ค่อยนิยมใช้กันในทางปฏิบัติเพราะว่ามีสูตรที่คำนวณค่า $S_{y/x}$ ได้ง่ายและรวดเร็วกว่า ดังนี้

$$S_{y/x} = \sqrt{\frac{\Sigma Y_i^2 - \beta_1\Sigma Y_i - \beta_2\Sigma X_iY_i}{n-2}}$$

ตัวอย่าง 11.3.2 จากข้อมูลในตัวอย่าง 11.3.1

จงหาความคลาดเคลื่อนมาตรฐานของค่าประมาณสำหรับจำนวนไมล์ที่ใช้กับราคาขาย

$$\text{จาก} \quad S_{y/x} = \sqrt{\frac{\Sigma Y_i^2 - \beta_1 \Sigma Y_i - \beta_2 \Sigma X_i Y_i}{n-2}}$$

$$= \sqrt{\frac{39{,}960{,}000 - 2{,}934(21{,}600) - (-38.56)(640{,}000)}{14-2}}$$

$$= \sqrt{\frac{39{,}960{,}000 - 63{,}374{,}400 + 24{,}678{,}400}{12}}$$

$$= \sqrt{105{,}333} = 324.55$$

การทดสอบสมมติฐานเกี่ยวกับ β_2

เป็นการทดสอบดูว่าไม่มีความสัมพันธ์ระหว่างตัวแปรที่อธิบายได้ (X) กับตัวแปรตาม (Y) สำหรับขั้นตอนในการทดสอบมีดังนี้

1. ตั้งสมมติฐาน

$$H_0 : \beta_2 = 0$$

$$H_a : \beta_2 \neq 0$$

3. กำหนด α , n

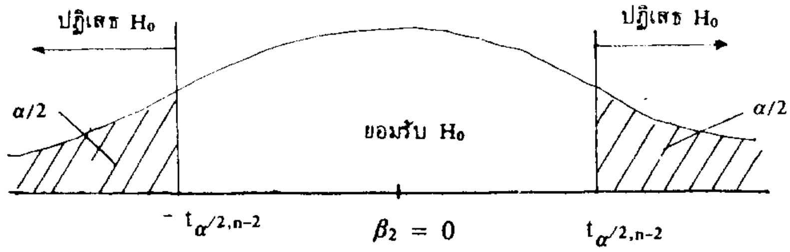
3. ตัวสถิติที่ใช้ทดสอบ คือ

$$t = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}}$$

$$df = n-2$$

$$\hat{\beta}_2 = S_{y/x}\sqrt{\frac{1}{\Sigma X^2 - [\frac{(\Sigma X)^2}{n}]}}$$

4. CR: ปฏิเสธถ้า $t_c < -t_{\alpha/2, n-2}$ หรือ $t_c > t_{\alpha/2, n-2}$



5. คำนวณค่า t_c จาก

$$t_c = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}_2}}$$

6. สรุปผล

1. ถ้า t_c ตกอยู่ใน CR จะปฏิเสธ H_0 และยอมรับ H_a
2. ถ้า t_c ตกอยู่นอก CR จะยอมรับ H_0

ตัวอย่าง 11.3.3

จากตัวอย่าง 11.3.2 จงทดสอบสมมติฐาน เกี่ยวกับ $\beta_2 = 0$ ที่ $\alpha = .01$

ได้

$$S_{y/x} = 324.55$$
$$\Sigma X^2 = 21,825$$
$$(\Sigma X)^2 = 255,025$$
$$\hat{\beta}_2 = -38.56$$

1. $H_0 : \beta_2 = 0$

   $H_a: \beta_2 \neq 0$

2. $\alpha = .01$, $n = 14$

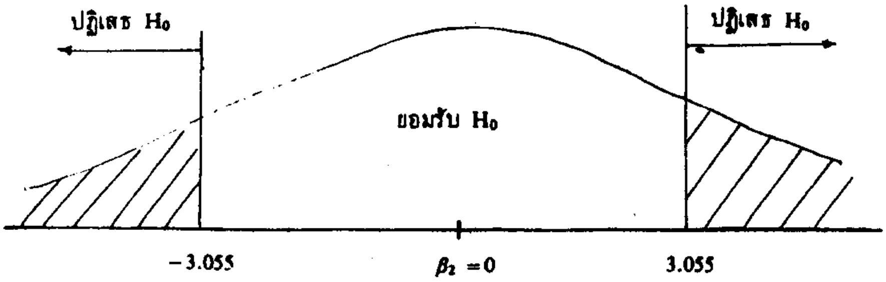
3. ตัวสถิติที่ใช้ทดสอบ คือ

$$t = \frac{\hat{\beta}_2 - 0}{\hat{\beta}_2}$$

$$dt = 14 - 2 = 12$$

4. CR: ปฏิเสธ H_0 ถ้า $t_c < -t_{.005,12}$ หรือ $t_c > t_{.005,12}$

คือ $t_c < -3.055$ หรือ $t_c > 3.055$



5. คำนวณค่า t_c จาก

$$t_c = \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}}$$

$$\hat{\beta}_2 = -38.56$$

$$\hat{\beta}_2 = S_{y/x}\sqrt{\frac{1}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}}$$

$$= 324.55\sqrt{\frac{1}{21{,}825 - \frac{(255{,}025)}{14}}}$$

$$= 324.55\sqrt{\frac{1}{3608.9}}$$

$$= 5.40$$

$$\therefore t_c = \frac{-38.56}{5.40}$$

$$= -7.14$$

สรุปผล $\because t_c$ ตกอยู่ใน CR

$\therefore$ ปฏิเสธ $H_0 : \beta_2 = 0$ และ

ยอมรับ $H_a : \beta_2 \neq 0$ แสดงว่า

มีความสัมพันธ์ระหว่างจำนวนไมล์ที่ใช้กับราคาขาย

นอกจากการทดสอบสมมติฐานเกี่ยวกับ β_2 แล้ว เรายังสามารถหา $100(1-\alpha)\%$ ช่วงเชื่อมั่นของ β_2 ได้จากสูตร

$$\hat{\beta}_2 - t_{\alpha/2,n-2}\, S_{\hat{\beta}_2} \leqslant \beta_2 \leqslant \hat{\beta}_2 + t_{\alpha/2,n-2}\, S_{\hat{\beta}_2}$$

เช่นจะหา 95% ช่วงเชื่อมั่นของ β_2 จากตัวอย่าง 11.3.3 จะได้

$$-38.56 - t_{.025,12}\, S_{\hat{\beta}_2} \leqslant \beta_2 \leqslant -38.56 + t_{.025,12}\, S_{\hat{\beta}_2}$$

$$-38.56 - 2.179(5.4) \leqslant \beta_2 \leqslant -38.56 + 2.179(5.4)$$

$$-38.56 - 11.77 \leqslant \beta_2 \leqslant -38.56 + 11.77$$

$$-50.33 \leqslant \beta_2 \leqslant -26.79$$

$\therefore$ 95% ช่วงเชื่อมั่นของ β_2 จะอยู่ระหว่าง -50.33 ถึง -26.79

แบบฝึกหัด

1. สหสัมพันธ์และการถดถอยคืออะไร
2. สัมประสิทธิ์สหสัมพันธ์คืออะไร มีคุณสมบัติอย่างไร และคำนวณหาสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างได้จากสูตรไหน
3. ตัวแบบถดถอยเชิงเส้นแบบเชิงเดี่ยวมีรูปฟอร์มเป็นอย่างไร และมีข้อสมมติเบื้องต้นอะไรบ้างเกี่ยวกับตัวแบบดังกล่าว
4. จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์จากข้อมูลแต่ละชุดดังนี้

ก.

x	y
34	21
30	22
40	25
34	28
39	15
35	24
42	24
45	22
43	17

$\bar{x} = 38$, $\bar{y} = 22$

$s_x = 5$, $s_y = 4$

ข.

x	y
3.9	46
4.6	46
6.0	52
2.8	50
3.1	48
3.4	40
4.2	42
4.0	44

$\bar{x} = 4$, $\bar{y} = 46$

$s_x \doteq 1$, $s_y = 4$

5. จงหาสัมประสิทธิ์สหสัมพันธ์ของคะแนนทดสอบที่ได้ 2 ครั้ง ดังนี้

นักศึกษาคนที่	ทดสอบครั้งที่ 1	ทดสอบครั้งที่ 2
1	82	92
2	84	91
3	86	90
4	83	92
5.	88	87
6	87	86
7	85	89
8	83	90
9	86	92
10	85	90
11	87	91
	$\bar{x} = 85$	$\bar{y} = 90$

6. กำหนดค่าต่าง ๆ ต่อไปนี้ให้ในแต่ละหัวข้อจงหาค่า

	n	Σx	Σy	Σxy	Σx^2	Σy^2
ก.	25	60	52	200	400	592
ข.	50	15	20	146	204.5	400
ค.	100	-20	25	-3.5	5	12.5

7. ตารางข้างล่างแสดงถึงส่วนสูง (นิ้ว) และน้ำหนักเป็นปอนด์ของนักเรียนชาย 12 คน

ก. จงเขียน Scatter diaqram

ข. จงหาสมการของเส้นถดถอย

ค. จงประมาณน้ำหนักของนักเรียนคนหนึ่งถ้าความสูงของเขาเท่ากับ 63 นิ้ว

ง. จงประมาณความสูงของนักเรียนคนหนึ่งถ้าน้ำหนักของเขาเท่ากับ 168 ปอนด์

ส่วนสูง (นิ้ว)	70	63	72	60	66	72	74	63	63	67	65	68
น้ำหนัก (ปอนด์)	155	150	180	135	156	168	178	160	132	145	149	152

8. จากโจทย์ข้อ (7) จงคำนวณหาค่า r

9. จากโจทย์ข้อ (7) จงทดสอบสมมติฐาน $H_o : \rho = 0$

$H_1 : \rho \neq 0$

เมื่อ $\alpha = 0.05$

10. จากตัวอย่าง 42 คู่ ของค่าสังเกต x และ y คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ได้เท่ากับ 0.22 จงทดสอบสมมติฐานที่ว่า $H_o : \rho = 0$, $H_1 : \rho \neq 0$ เมื่อ $\alpha = 0.05$

11. กำหนดให้

x	0.5	1.5	3.2	4.2	5.1	6.5
y	1.3	3.4	6.7	8.0	10.0	13.2

ก. จงประมาณเส้นถดถอย

ข) จงทดสอบที่ระดับนัยสำคัญ 0.01 ว่าเส้นถดถอยจะผ่านจุดกำเนิด

12. กำหนดให้ $\Sigma x = 766$, $\Sigma y = 1,700$, $\Sigma xy = 109,380$

$\Sigma x^2 = 49,068$, $\Sigma y^2 = 246,100$, $n = 12$

ก. จะประมาณเส้นถดถอย

ข. จงทำนายค่า y เมื่อ $x = 80$

ค. จงทดสอบที่ระดับนัยสำคัญ .05 ว่า y และ x ไม่มีความสัมพันธ์กัน

13. จากข้อมูลในข้อ (4) จงหา

ก. β_1 และ β_2

ข. ความคลาดเคลื่อนมาตรฐานของค่าประมาณ

ค. จงทดสอบ $H_o : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$ ที่ $\alpha = .01$

ง. จงสร้าง 95% ช่วงเชื่อมั่นของ β_2

14. จากข้อมูลในข้อ (5) จงหา

ก. β_1 และ β_2

ข. ความคลาดเคลื่อนมาตรฐานของค่าประมาณ

ค. จงทดสอบ $H_o : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$ ที่ $\alpha = .01$

15 จากข้อมูลในข้อ (6)

ก. จงสร้างสมการถดถอย

ข. จงหาความคลาดเคลื่อนมาตรฐานของค่าประมาณ

ค. จงทดสอบ $H_o : \rho = 0$, $H_1 : \rho \neq 0$ ที่ $\alpha = .05$