



การวิเคราะห์การถดถอยเชิงเส้นและสหสัมพันธ์อย่างง่าย (Simple Linear Regression and Correlation Analysis)





เนื้อหา

- การวิเคราะห์การถดถอยอย่างง่าย
 - ตัวแบบการถดถอยเชิงเส้นอย่างง่าย
 - กราฟแสดงความสัมพันธ์ระหว่าง X และ Y และสมการประมาณค่า
 - สมการถดถอยเชิงเส้น
 - ขั้นตอนการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย
- การวิเคราะห์สหสัมพันธ์อย่างง่าย
 - แผนภาพการกระจาย
 - สัมประสิทธิ์สหสัมพันธ์



การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

เป็นการหาสมการที่แสดงความสัมพันธ์ระหว่างตัวแปรต้น (X) และตัวแปรตาม (Y) ซึ่งจะทำให้พยากรณ์หรือคาดคะเนสิ่งที่จะเกิดขึ้น สำหรับการวิเคราะห์สัมพัทธ์ก็เป็นการวัดว่าตัวแปรต้นมีความสัมพันธ์กับตัวแปรตามอย่างไร



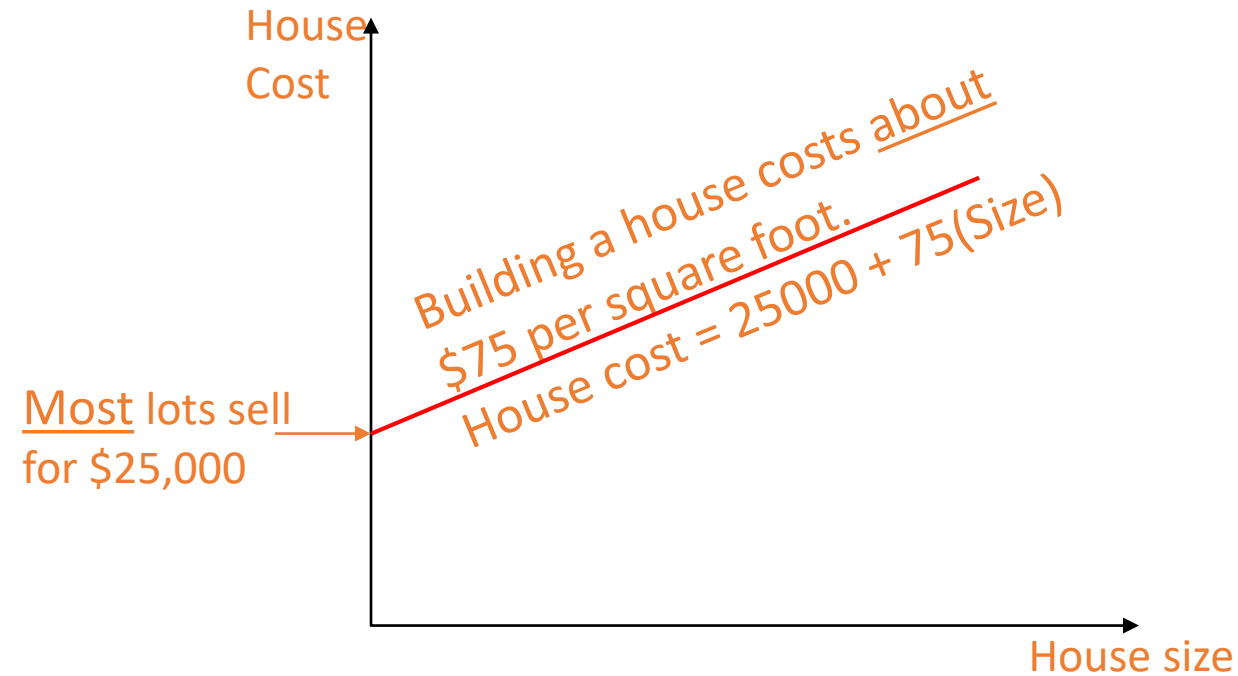
การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

- ตัวอย่างคำถาม เช่น
 - เปิดเครื่องปรับอากาศวันละ 5 ชั่วโมง เดือนนี้จ่ายค่าไฟ 560 บาท ถ้าอยากจ่ายเดือนละ 450 บาท หรือน้อยกว่าควรจะเปิดวันละกี่ชั่วโมง
 - ถ้าขับรถด้วยความเร็วเฉลี่ย 90 กิโลเมตรต่อชั่วโมง เติมน้ำมันเต็มถังจะวิ่งรถได้รวม 560 กิโลเมตร ถ้าขับที่ความเร็วเฉลี่ย 110 กิโลเมตรต่อชั่วโมง จะวิ่งได้กี่กิโลเมตร แล้วถ้าเป็น 130 กิโลเมตรต่อชั่วโมง จะได้กี่กิโลเมตร ถ้าเติมน้ำมันเต็มถังเหมือนกัน
 - ถ้าตั้งอุณหภูมิหม้อต้มน้ำที่ 100 องศาเซลเซียส จะสามารถฆ่าเชื้อแบคทีเรียได้ร้อยละ 85 ภายใน 1 นาที อยากทราบว่าถ้าต้องการให้ฆ่าได้ร้อยละ 90 ในเวลาเท่าเดิม จะต้องใช้อุณหภูมิเท่าใด และถ้าต้องการให้ฆ่าได้หมดร้อยละ 100 จะต้องใช้อุณหภูมิเท่าใด



The Model

The model has a deterministic and a probabilistic components



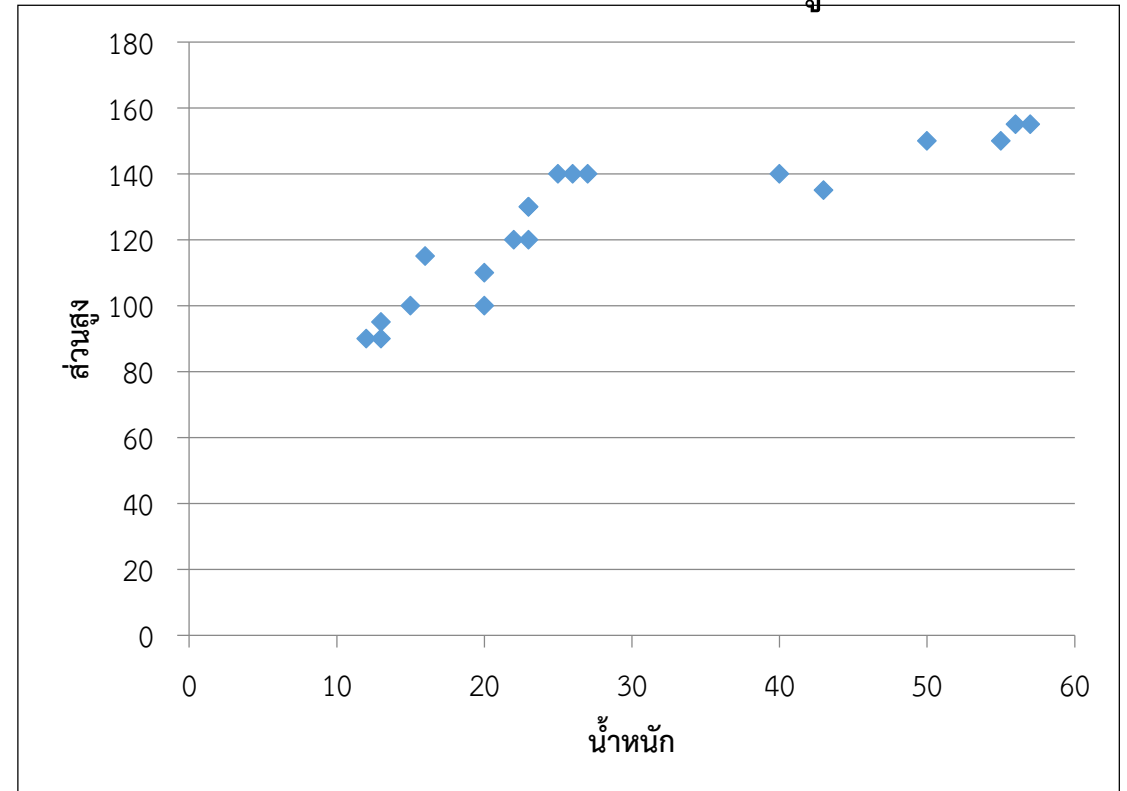


ตัวอย่าง

- จากข้อมูลแสดงส่วนสูงและน้ำหนักของนักเรียนโรงเรียนแห่งหนึ่งจำนวน 20 คน

ส่วนสูง	น้ำหนัก	ส่วนสูง	น้ำหนัก
90	12	140	27
100	15	150	50
100	20	150	55
110	20	155	57
130	23	135	43
115	16	140	40
95	13	130	23
120	22	120	23
140	25	90	13
140	26	155	56

- กราฟแสดงความสัมพันธ์ระหว่างส่วนสูงและน้ำหนัก





ตัวแบบการถดถอยเชิงเส้นอย่างง่าย

รูปแบบการถดถอยเชิงเส้นอย่างง่ายตรงที่แทนความสัมพันธ์ของ X และ Y คือ

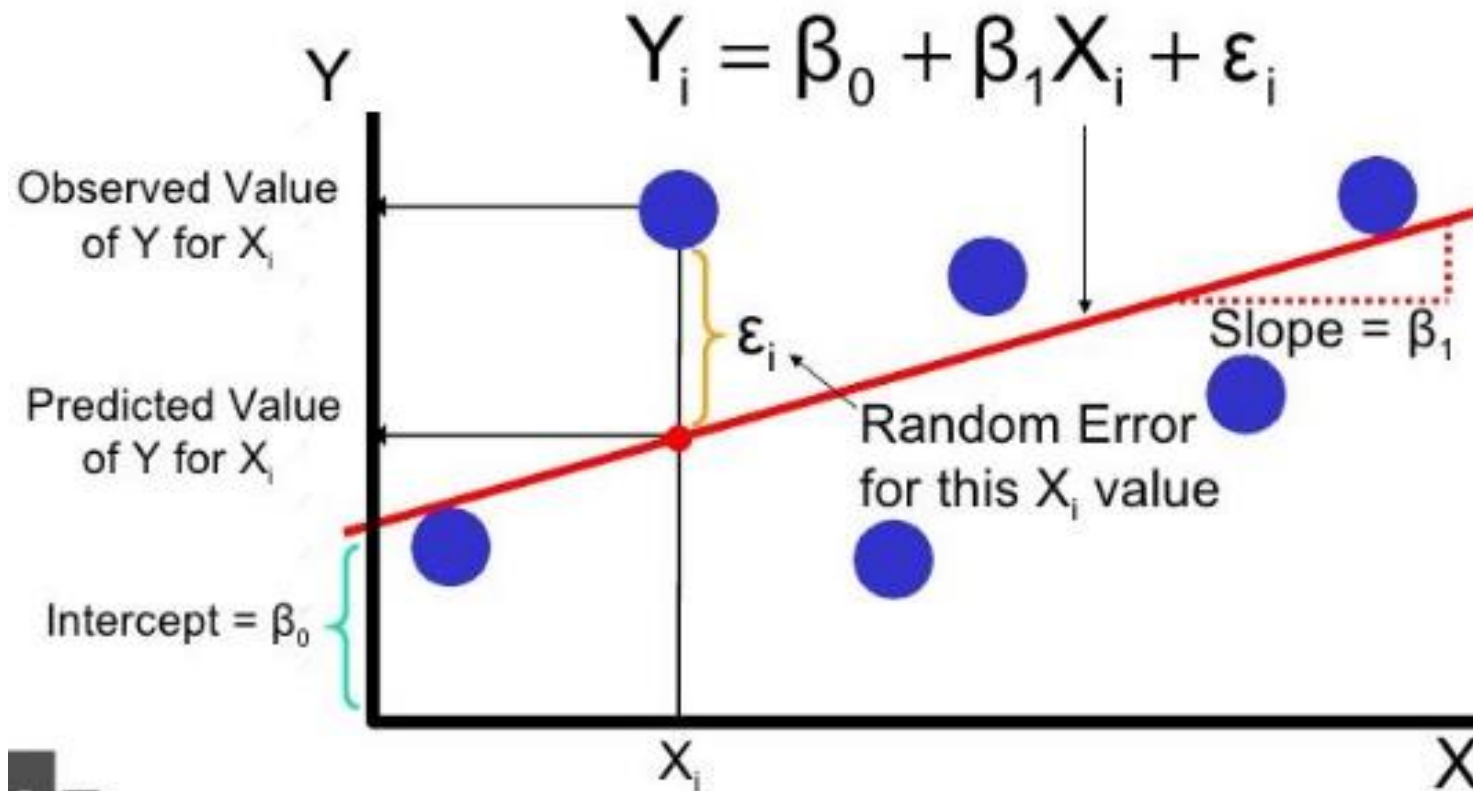
$$Y = \beta_0 + \beta_1 X_i + \varepsilon_i$$

เมื่อ	Y	คือ ตัวแปรตามเป็นตัวแปรที่ต้องการสร้างตัวแบบในการพยากรณ์
	X_i	คือ ตัวแปรอิสระหรือตัวแปรต้นหรือตัวแปรที่ใช้ในการพยากรณ์ค่าของตัวแปรตาม
Y		
	β_0	เป็นจุดที่เส้นถดถอยตัดแกน Y หรือค่าของ Y เมื่อ $X = 0$
	β_1	เป็นค่าความชันของเส้นตรง ซึ่งเป็นค่าที่แสดงถึงอัตราการเปลี่ยนแปลง
ของ Y		เมื่อ X เปลี่ยนไป 1 หน่วย
	ε_i	เป็นความคลาดเคลื่อนที่เกิดจากการที่ Y แตกต่าง
จากค่าเฉลี่ยของประชากร		



การวิเคราะห์การถดถอย

- กราฟแสดงความสัมพันธ์ระหว่าง X และ Y และสมการประมาณค่า





สมการถดถอยเชิงเส้น

- จากตัวแบบถดถอยเชิงเส้นอย่างง่าย β_0 และ β_1 เป็นพารามิเตอร์ที่ไม่ทราบค่าและในทางปฏิบัติไม่สามารถเก็บข้อมูลได้ครบถ้วนทุกค่าในประชากร จึงต้องประมาณโดยใช้ข้อมูลตัวอย่าง
- สำหรับ β_0 และ β_1 คือ b_0 และ b_1 โดยวิธีการประมาณจะใช้วิธีกำลังสองน้อยที่สุด (method of least square)
- ค่าสัมประสิทธิ์การถดถอยหาได้จากสมการ

$$b_1 = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$



การวิเคราะห์การถดถอย

ความหมายของค่าสัมประสิทธิ์การถดถอย (b_0 และ b_1)

- b_0 เป็นค่าที่บอกให้ทราบว่า ถ้าค่า X เป็น 0 แล้วค่า Y จะมีค่าเป็น b_0 โดยเฉลี่ย
เครื่องหมาย b_0 เป็นได้ทั้งบวกและลบ ถ้าเป็นลบแสดงว่าเส้นถดถอยตัดแกน Y ต่ำกว่าจุดกำเนิด (Origin) แต่ถ้าเครื่องหมายเป็นบวก ก็แสดงว่า เส้นถดถอยตัดแกน Y เหนือจุดกำเนิด
- b_1 เป็นอัตราการเพิ่มหรือลดลงของ Y เมื่อ X มีค่าเพิ่มขึ้น 1 หน่วย เมื่อ b_1 มีค่าเป็นลบจะบอกให้ทราบว่าตัวแปร X กับ Y มีความสัมพันธ์กันในทางตรงกันข้าม นั่นคือ เมื่อ X มีค่าเพิ่มขึ้น Y จะมีค่าลดลง แต่ถ้า b_1 มีค่าเป็นบวก X กับ Y จะมีความสัมพันธ์ไปในทิศทางเดียวกัน คือ เมื่อ X มีค่าเพิ่มขึ้น Y ก็จะมีเพิ่มด้วย



ขั้นตอนการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

1. ตรวจสอบว่าตัวแปร X และ Y มีความสัมพันธ์กันเชิงเส้นหรือไม่
 2. สร้างสมการการพยากรณ์ เพื่อใช้ประมาณค่า Y
 3. ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่ายของตัวแปรทั้งสอง
-

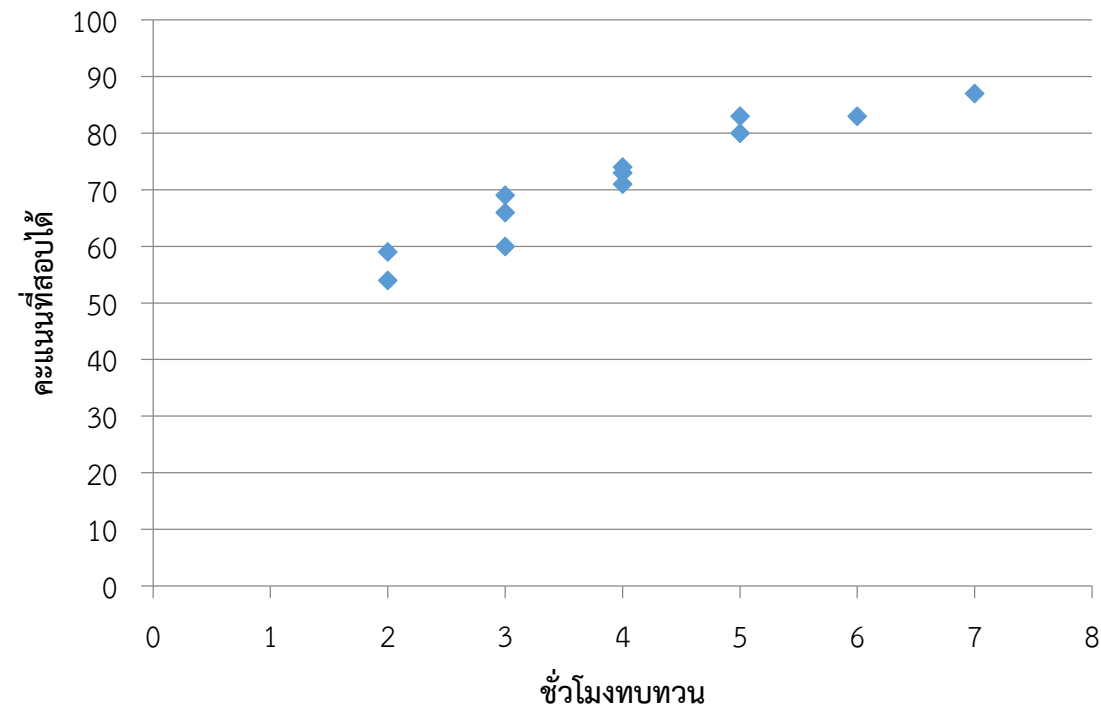


ตัวอย่าง

- อาจารย์สอนในรายวิชาสถิติผู้หนึ่งรวบรวมข้อมูลจำนวนชั่วโมงที่นักศึกษาทบทวนบทเรียนก่อนสอบ (X) และคะแนนที่สอบได้ (Y) เป็นดังนี้

ชั่วโมงทบทวน(X)	คะแนนที่สอบได้ (Y)
2	59
3	69
4	70
4	73
5	80
2	54
3	60
7	87
6	83
4	74
5	83
3	66

แผนภาพการกระจาย (Scatter plot)





ชั่วโมงทบทวน (X)	คะแนนที่สอบได้ (Y)	X ²	XY
2	59		
3	69		
4	71		
4	73		
5	80		
2	54		
3	60		
7	87		
6	83		
4	74		
5	83		
3	66		

• จากสูตร

$$b_1 = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$\begin{aligned} b_1 &= \frac{3,605 - \frac{(48)(859)}{12}}{218 - \frac{(48)^2}{12}} \\ &= \frac{3,605 - 3,436}{218 - 192} \\ &= \frac{169}{26} = 6.5 \end{aligned}$$



ตัวอย่าง

- จากสูตร

$$b_0 = \bar{Y} - b_1 \bar{X}$$

$$= \left(\frac{859}{12} \right) - 6.5 \left(\frac{48}{12} \right) \\ = 71.583 - 26 = 45.583$$

- ดังนั้นจะได้สมการถดถอยเชิงเส้นอย่างง่ายคือ $Y = b_0 + b_1 X$

$$Y = 45.583 + 6.5X$$

- ถ้านายมาดีใช้เวลาในการทบทวนบทเรียน 8 ชั่วโมง จะได้คะแนนสอบประมาณเท่าใด

$$Y = 45.583 + 6.5(8)$$

$$= 45.583 + 52 = 97.583 \text{ คะแนน}$$



การวิเคราะห์สหสัมพันธ์อย่างง่าย



สัมประสิทธิ์สหสัมพันธ์

- ค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple Correlation Coefficient) เป็นค่าที่วัดความสัมพันธ์ของ X กับ Y ว่ามีขนาดและทิศทางของความสัมพันธ์อย่างไร

กำหนดให้ ρ คือ สัมประสิทธิ์สหสัมพันธ์อย่างง่ายของประชากร

$$\text{เมื่อ } -1 \leq \rho \leq 1$$

- เนื่องจากในการวิเคราะห์สหสัมพันธ์ ไม่ได้เก็บข้อมูลจากประชากร แต่เป็นการเก็บข้อมูลจากตัวอย่าง
- ดังนั้น ค่าสัมประสิทธิ์ที่ได้ จึงเป็นสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของตัวอย่าง โดยให้สัญลักษณ์ว่า “ r ”



สัมประสิทธิ์สหสัมพันธ์

- ค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของตัวอย่าง (r) คำนวณได้ดังนี้

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}}$$

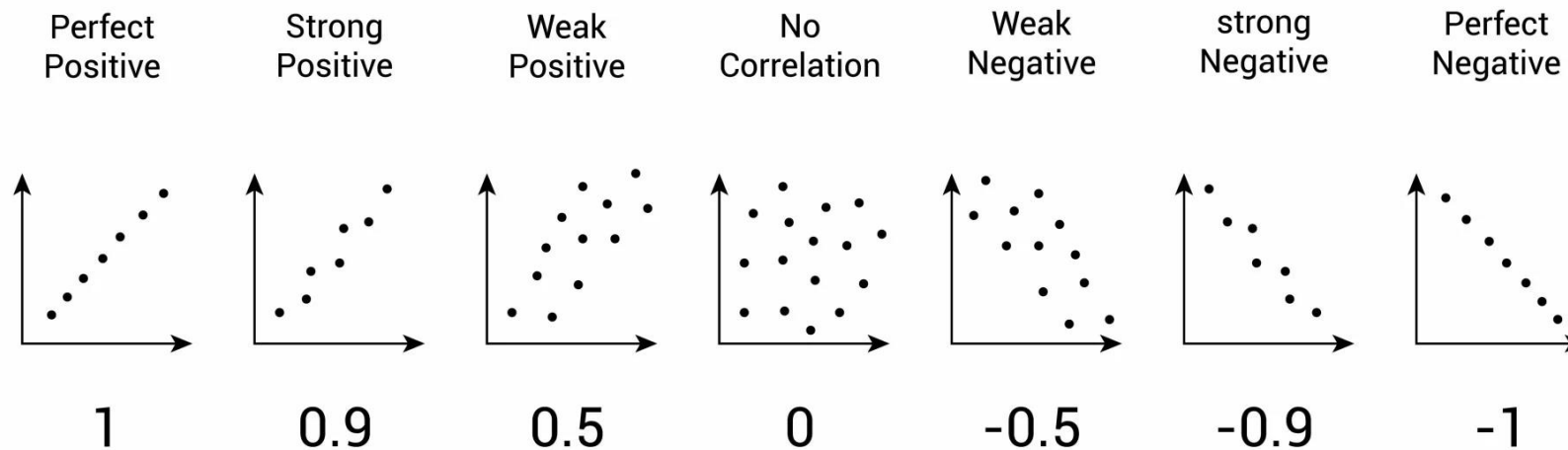
ความหมายของค่า r

- ค่า r เป็น + แสดงว่า X กับ Y มีความสัมพันธ์ในทิศทางเดียวกัน
- ค่า r เป็น - แสดงว่า X กับ Y มีความสัมพันธ์ในทิศทางตรงกันข้าม
- ค่า r เป็น 0 แสดงว่า X กับ Y ไม่มีความสัมพันธ์กัน
- ค่า $|r|$ มีค่าเข้าใกล้ 1 แสดงว่า X กับ Y มีความสัมพันธ์กันมาก
- ค่า $|r|$ มีค่าเข้าใกล้ 0 แสดงว่า X กับ Y มีความสัมพันธ์กันน้อย



แผนภาพการกระจาย

- ลักษณะของ Scatter Plot ดังนี้





ตัวอย่าง

- จากข้อมูลการผลิตสินค้าของบริษัทแห่งหนึ่ง โดยทำการบันทึกข้อมูลปริมาณการผลิต (X) กับเวลาที่ใช้ในการผลิต (Y) ในช่วงเวลา 15 เดือน ได้ข้อมูลดังนี้

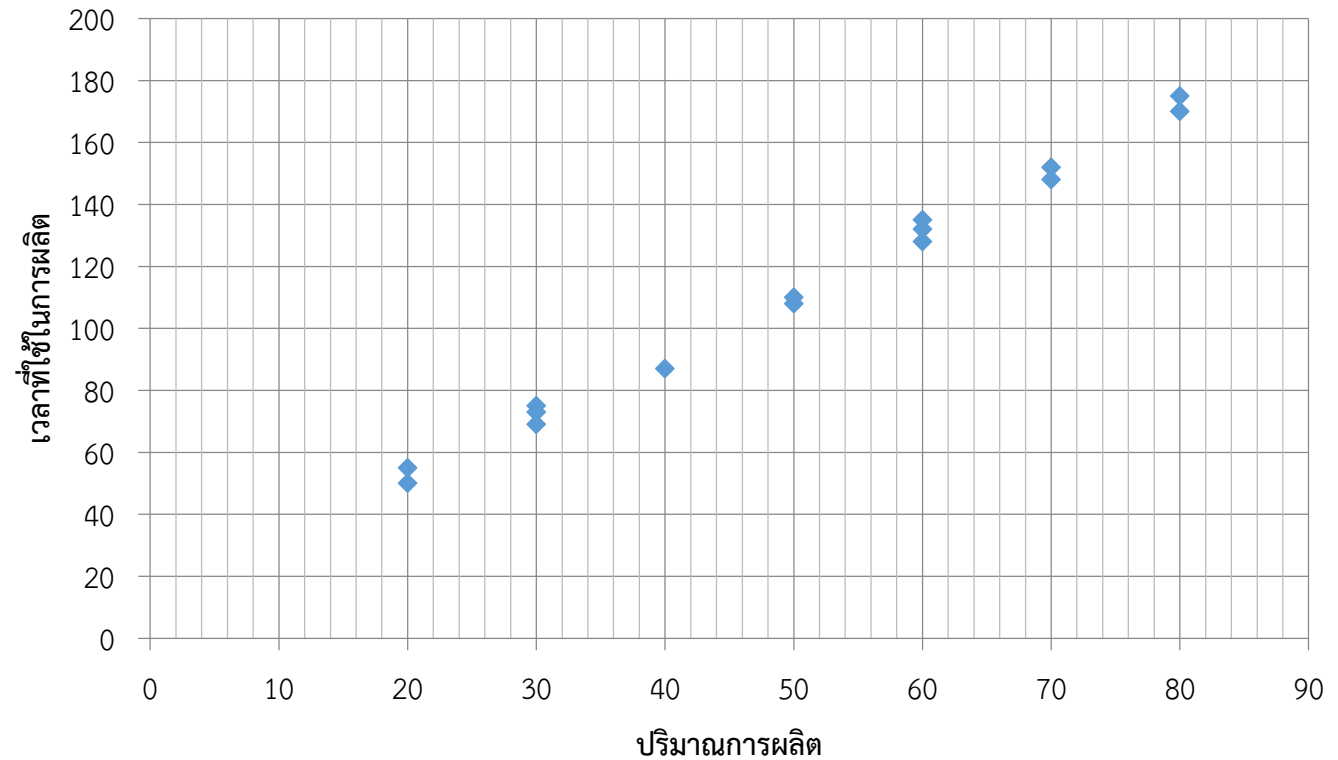
เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)
1	30	75
2	70	152
3	20	55
4	50	110
5	80	175
6	30	73
7	20	50
8	60	128
9	80	170
10	40	87
11	50	108
12	60	135
13	30	69
14	70	148
15	60	132

ตัวอย่าง



- จงเขียนแผนภาพกระจาย

เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)
1	30	75
2	70	152
3	20	55
4	50	110
5	80	175
6	30	73
7	20	50
8	60	128
9	80	170
10	40	87
11	50	108
12	60	135
13	30	69
14	70	148
15	60	132





เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)	X^2	Y^2	XY
1	30	75	$30 \times 30 = 900$	$75 \times 75 = 5,625$	$30 \times 75 = 2,250$
2	70	152	4,900	23,104	10,640
3	20	55	400	3,025	1,100
4	50	110	2,500	12,100	5,500
5	80	175	6,400	30,625	14,000
6	30	73	900	5,329	2,190
7	20	50	400	2,500	1,000
8	60	128	3,600	16,384	7,680
9	80	170	6,400	28,900	13,600
10	40	87	1,600	7,569	3,480
11	50	108	2,500	11,664	5,400
12	60	135	3,600	18,225	8,100
13	30	69	900	4,761	2,070
14	70	148	4,900	21,904	10,360
15	60	132	3,600	17,424	7,920
รวม	750.00	1,667.00	43,500.00	209,139.00	95,290.00

การคำนวณ



จากสูตร

$$\sum_{i=1}^n x = 750$$

$$\sum_{i=1}^n y = 1,667$$

$$n = 15$$

$$\sum_{i=1}^n x_i y_i = 95,290$$

$$\sum_{i=1}^n x_i^2 = 43,500$$

$$\sum_{i=1}^n y_i^2 = 209,139$$

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}}$$

$$r = \frac{95,290 - 15 \left(\frac{750}{15} \right) \left(\frac{1,667}{15} \right)}{\sqrt{(43,500 - 15 \left(\frac{750}{15} \right)^2) (209,139 - 15 \left(\frac{1,667}{15} \right)^2)}}$$

$$r = \frac{95,290 - 83,349.75}{\sqrt{(43,500 - 37,500)(209,139 - 185,258.155)}}$$

$$r = \frac{11,940.25}{\sqrt{(6,000)(23,880.845)}} = \frac{11,940.25}{\sqrt{143,285,070}}$$

$$r = \frac{11,940.25}{11,970.174} = 0.9975$$

ดังนั้น ปริมาณการผลิตและเวลาที่ใช้ในการผลิตมีความสัมพันธ์เท่ากับ **0.9975**



เอกสารอ้างอิง

- ชลิดา ตระกูลสุนทร. (2567). เอกสารประกอบการสอน 4113105 สถิติเพื่อการวิจัย.
ค้นเมื่อ มกราคม 10, 2567, จาก <https://pws.npru.ac.th/chalida/data/files/บทที่ 8 สหสัมพันธ์และการถดถอย.pdf>
- มานะชัย รอดชื่น. (2567). เอกสารประกอบการสอนวิชา 208272 สถิติเบื้องต้นสำหรับ
สังคมศาสตร์ 2. ค้นเมื่อ มกราคม 10, 2567, จาก
<https://www.manachai.cmustat.com/208272/chapter6.pdf>



เอกสารอ้างอิง

สันหัตถ์ พรประเสริฐมานิต. (2568). การวิเคราะห์ความสัมพันธ์และวิเคราะห์ถดถอย. ค้นเมื่อ มกราคม 10, 2568, จาก

<https://www.sunthud.com/media/Teaching/ComPsy/ComPsy%20Correlation%20and%20Regression.pdf>

สำนักเทคโนโลยีการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช. (2568). คณิตศาสตร์สำหรับวิทยาการคอมพิวเตอร์. ค้นเมื่อ มกราคม 10, 2568, จาก

<https://media.stou.ac.th/>

