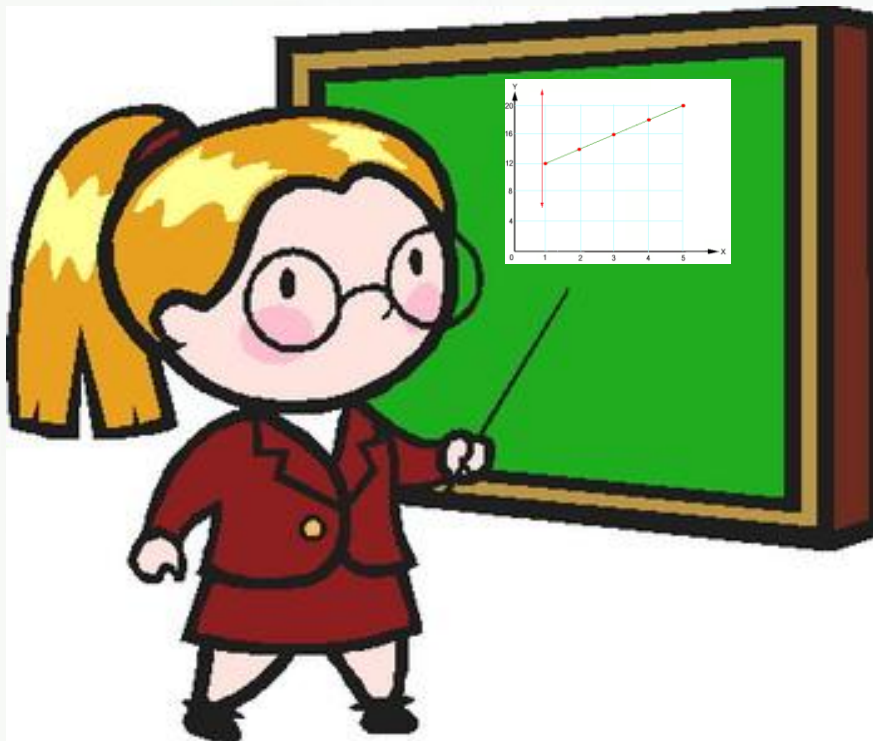




บทที่ 6

การถดถอยและสหสัมพันธ์เชิงเส้นอย่างง่าย





การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติ ที่ศึกษาถึงความสัมพันธ์
ระหว่าง ตัวแปรตัวหนึ่ง ซึ่งเรียกว่า “ตัวแปรตาม (Dependent variable)
แทนด้วย Y ” และตัวแปรอื่น ๆ ซึ่งเรียกว่า “ตัวแปรอิสระ
(Independent variable) แทนด้วย X ” ว่า
มีความสัมพันธ์กันอย่างไร ซึ่งความสัมพันธ์
ดังกล่าวนี้จะอยู่ในรูปแบบคณิตศาสตร์โดยอาจ
จะเป็นเชิงเส้นตรงก็ได้หรือไม่เป็นก็ได้ มีประโยชน์
เพื่อใช้ในการทำนาย (Prediction) นั่นคือ ใช้ตัวแปรอิสระ X เป็นเกณฑ์ใน
การประมาณตัวแปรตาม Y โดยสำหรับในบทนี้จะกล่าวถึง การวิเคราะห์
ถดถอยที่เป็นเชิงเส้นอย่างง่ายเท่านั้น





6.1 แผนภาพการกระจาย

ในการวิเคราะห์การถดถอยเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว มักเริ่มด้วยการนำข้อมูลจากตัวอย่างมาเขียนเป็นแผนภาพการกระจาย (Scatter diagram) ให้ตัวแปรตามแทนด้วย Y และตัวแปรอิสระแทนด้วย X ซึ่งอาจได้แผนภาพแสดงความสัมพันธ์เป็นแบบเชิงเส้นตรง เส้นพาราโบลา เส้นเอ็กซ์โพเนนเชียล ฯลฯ หรืออาจไม่มีความสัมพันธ์กันก็ได้ เช่น





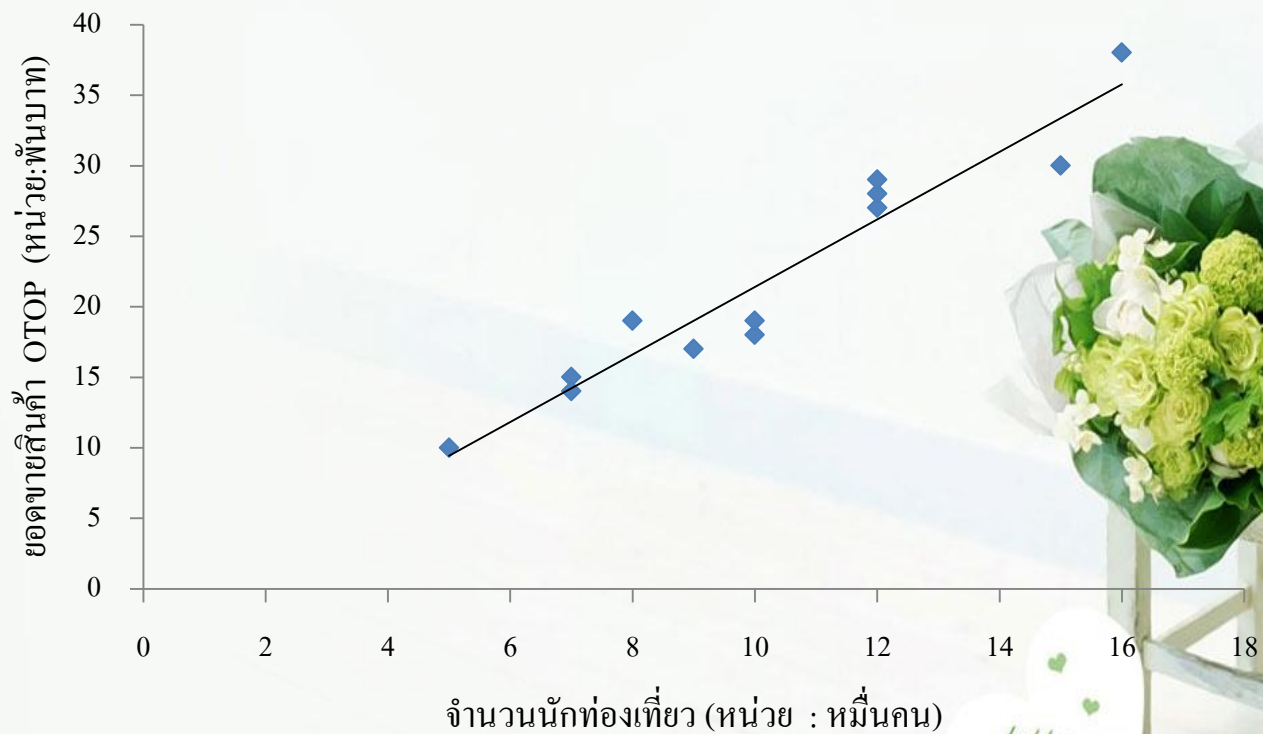
ตัวอย่าง จากข้อมูลจำนวนนักท่องเที่ยวที่เดินทางเข้ามาเที่ยวในจังหวัด
บุรีรัมย์ปี 2556 กับยอดขายสินค้า OTOP เป็นดังนี้

เดือน	จำนวนนักท่องเที่ยว (หน่วย : หมื่นคน)	ยอดขายสินค้า OTOP (หน่วย : พันบาท)
มกราคม	8	19
กุมภาพันธ์	7	15
มีนาคม	12	27
เมษายน	10	18
พฤษภาคม	5	10
มิถุนายน	7	14
กรกฎาคม	9	17
สิงหาคม	12	28
กันยายน	10	19
ตุลาคม	15	30
พฤศจิกายน	16	38
ธันวาคม	12	29





นำข้อมูลไปเขียนกราฟแสดงความสัมพันธ์ได้ดังนี้





6.2 ตัวแบบการถดถอยเชิงเส้นอย่างง่าย

กำหนดให้ตัวแบบการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Model) คือ $Y = \beta_0 + \beta_1 X + \varepsilon$

Y คือ ตัวแปรตามเป็นตัวแปรที่ต้องการสร้างตัวแบบในการพยากรณ์

X คือ ตัวแปรอิสระหรือตัวแปรต้นหรือตัวแปรที่ใช้ในการพยากรณ์ค่าของตัวแปรตาม Y

$$E(Y) = \mu_{Y \cdot X} = \beta_0 + \beta_1 X$$

β_0 และ β_1 คือ พารามิเตอร์ที่แทนค่าสัมประสิทธิ์

การถดถอยของประชากร

โดยที่ β_0 คือ ค่าที่ตัดแกน Y (Y - intercept) เป็นค่าคงที่เมื่อ $X = 0$

β_1 คือ ความชัน (Slope) ของเส้นถดถอย เป็นอัตราการเปลี่ยนแปลงของค่าเฉลี่ย Y เมื่อ X เปลี่ยนไป 1 หน่วย





ε คือ ความคลาดเคลื่อนที่เกิดจากการที่ Y แตกต่างจากค่าเฉลี่ยของ
ประชากร $\mu_{Y \cdot X}$

สำหรับความคลาดเคลื่อนมีคุณสมบัติหรือข้อสมมติ ดังนี้

1. มีการแจกแจงปกติ มีค่าเฉลี่ยเป็นศูนย์กล่าวคือ

$$E(\varepsilon_i) = 0$$

2. มีค่าความแปรปรวนคงที่คือ $\sigma_\varepsilon^2 = \sigma^2$

3. ε_i และ ε_j เป็นอิสระกัน มีค่าความแปรปรวนร่วม
เท่ากับศูนย์ $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ เมื่อ $i \neq j$





6.3 สมการถดถอยเชิงเส้นอย่างง่าย

จากตัวแบบถดถอยเชิงเส้นอย่างง่าย β_0 และ β_1 เป็นพารามิเตอร์

ที่ไม่ทราบค่าและในทางปฏิบัติไม่สามารถเก็บข้อมูลได้ครบถ้วนทุกค่าในประชากร จึงต้องประมาณโดยใช้ข้อมูลตัวอย่าง สำหรับค่าประมาณของ

β_0 และ β_1 คือ b_0 และ b_1 โดยวิธีการประมาณจะใช้วิธีกำลัง

สองน้อยที่สุด (method of least square) วิธีการนี้

เป็นวิธีการที่หาค่าของ b_0 และ b_1 ในเทอมของตัวแปรอิสระ (X)

และตัวแปรตาม (Y) โดยให้ผลรวมของความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด

ดังนั้นสมการถดถอยที่คำนวณได้จะมีคุณสมบัติดังนี้

1. ผลรวมของความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด ($\sum e^2$)
2. ผลรวมของความคลาดเคลื่อนมีค่าเท่ากับศูนย์ ($\sum e = 0$)
3. เส้นถดถอยที่ได้จะผ่านจุดตัดของ \bar{X} และ \bar{Y}





เมื่อทำการหาค่าประมาณ b_0 และ b_1 โดยวิธีกำลังสองน้อยที่สุดแล้วจะได้ว่า

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

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

จากค่า b_0 และ b_1 จะทำให้สามารถคาดคะเนสมการถดถอย

$$\hat{Y} = b_0 + b_1X \quad \text{ได้ตามต้องการ}$$





ตัวอย่าง อาจารย์ผู้สอนในรายวิชาสถิติท่านหนึ่งรวบรวมข้อมูลจำนวนชั่วโมงที่
นักศึกษาทบทวนบทเรียนก่อนสอบ (X) และคะแนนที่สอบได้ (Y) เป็นดังนี้

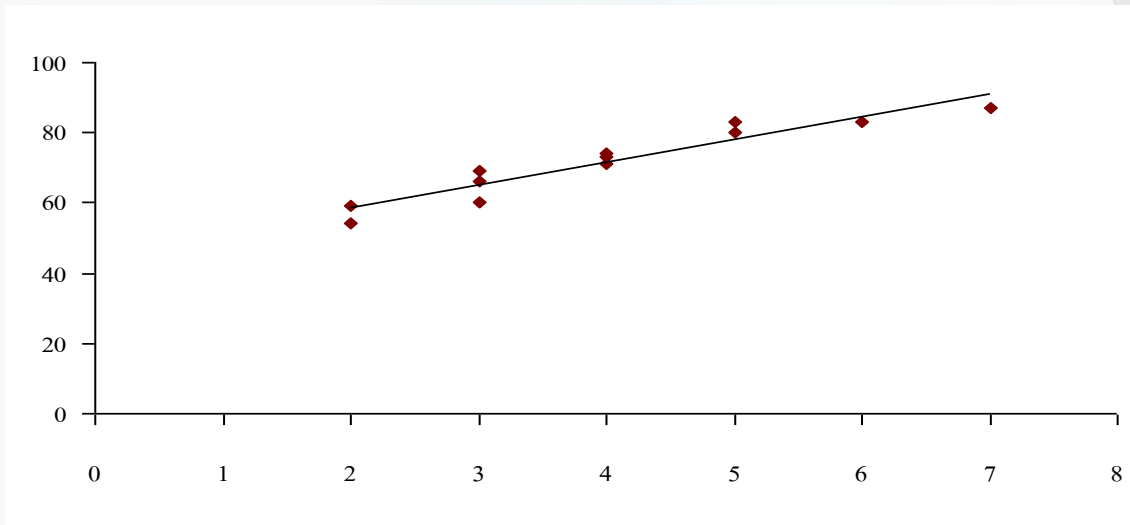
ชั่วโมงทบทวน (X)	คะแนนที่สอบได้ (Y)
2	59
3	69
4	71
4	73
5	80
2	54
3	60
7	87
6	83
4	74
5	83
3	66





- ก. จงเขียนแผนภาพการกระจาย
- ข. จงคำนวณหาสมการถดถอยเชิงเส้นอย่างง่าย
- ค. ถ้านายมาตีใช้เวลาในการทบทวนบทเรียน 8 ชั่วโมง จะได้คะแนนสอบ
ประมาณเท่าใด

วิธีทำ ก. เขียนแผนภาพการกระจาย ได้ดังนี้





ข. คำนวณหาสมการถดถอยเชิงเส้นอย่างง่าย ดังนี้

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

ชั่วโมงทบทวน (X)	คะแนนที่สอบได้ (Y)	X ²	XY
2	59	4	118
3	69	9	207
4	71	16	284
4	73	16	292
5	80	25	400
2	54	4	108
3	60	9	180
7	87	49	609
6	83	36	498
4	74	16	296
5	83	25	415
3	66	9	198
$\sum X = 48$	$\sum Y = 859$	$\sum X^2 = 218$	$\sum XY = 3,605$





แทนค่าต่างๆลงในสูตร ดังนี้

$$\begin{aligned}b_1 &= \frac{3,605 - \frac{(48)(859)}{12}}{218 - \frac{(48)^2}{12}} \\&= \frac{3,605 - 3,436}{218 - 192} \\&= \frac{169}{26} = 6.5 \\b_0 &= \bar{Y} - b_1 \bar{X} \\&= \left(\frac{859}{12} \right) - 6.5 \left(\frac{48}{12} \right) \\&= 71.583 - 26 = 45.583\end{aligned}$$

ดังนั้นจะได้สมการถดถอยเชิงเส้นอย่างง่ายจาก รูปแบบสมการถดถอย $Y = b_0 + b_1X$ คือ
 $Y = 45.583 + 6.5X$





ค. ถ้านายมาดีใช้เวลาในการทบทวนบทเรียน 8 ชั่วโมง จะได้คะแนนสอบ
ประมาณเท่าใด

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

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

6.4 สหสัมพันธ์เชิงเส้นอย่างง่าย

ในการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว

ถ้าไม่ได้กำหนดว่าตัวแปรใดเป็นตัวแปรตาม ตัวแปรใดเป็นตัวแปรต้น
หรือตัวแปรอิสระ จะเป็นเรื่องของการวิเคราะห์สหสัมพันธ์ (correlation analysis)
ซึ่งเป็นการศึกษาว่าตัวแปรคู่หนึ่งมีขนาดของความสัมพันธ์ระดับใด โดยวัดด้วยค่าสัมประสิทธิ์
สหสัมพันธ์ (correlation coefficient) และถ้าหากว่ามีข้อสงสัยว่าตัวแปรแต่ละคู่มี
สหสัมพันธ์กันจริงหรือไม่ ก็สามารถทำการทดสอบสมมติฐานได้ ค่าสัมประสิทธิ์สหสัมพันธ์จะ
มีค่าอยู่ระหว่าง -1 ถึง 1 เมื่อค่าเข้าใกล้ศูนย์ หมายความว่า ตัวแปรทั้งสองไม่มี
ความสัมพันธ์กัน แต่ถ้าค่าเข้าใกล้ 1 หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่าง
มาก ส่วนเครื่องหมาย + และเครื่องหมาย - จะแสดงถึงทิศทางความสัมพันธ์





ถ้าเป็นเครื่องหมาย + แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางเดียวกัน แต่ถ้าเป็นเครื่องหมาย - แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางตรงกันข้าม โดยในการวัดความสัมพันธ์ระหว่างตัวแปรจะต้องพิจารณาด้วยว่าตัวแปรที่นำมาวัดความสัมพันธ์กันนั้นอยู่ในมาตราการวัดใด

6.4.1 สัมประสิทธิ์สหสัมพันธ์

6.4.1.1 สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation)

ถ้าข้อมูลที่ต้องการวัดความสัมพันธ์อยู่ในมาตราอันดับภาค หรือมาตราอัตราส่วนค่าสัมประสิทธิ์สหสัมพันธ์จะหาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเขียนแทนด้วยสัญลักษณ์ ρ แต่ในทางปฏิบัติจะหาค่า ρ ไม่ได้เนื่องจากเป็นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์จากตัวอย่างจึงต้องทำการประมาณค่าของ ρ ด้วย r โดยมีสูตรการคำนวณดังนี้





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

ตัวอย่าง จากข้อมูลการผลิตสินค้าของบริษัทแห่งหนึ่ง โดยทำการบันทึกข้อมูล
ปริมาณการผลิต (X) กับเวลาที่ใช้ในการผลิต (Y) ในช่วงเวลา 15 เดือน ได้ข้อมูล
ดังนี้

เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)	X ²	Y ²	XY
1	30	75	900	5,625	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





เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)	X^2	Y^2	XY
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	1,667	43,500	209,139	95,290

จงหาค่าสัมประสิทธิ์สหสัมพันธ์





วิธีทำ จากสูตร

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

แทนค่าต่างๆลงในสูตร ดังนี้

$$\begin{aligned} r_{XY} &= \frac{95,290 - 15\left(\frac{750}{15}\right)\left(\frac{1,667}{15}\right)}{\sqrt{\left[(43,500) - 15\left(\frac{750}{15}\right)^2\right]\left[(209,139) - 15\left(\frac{1,667}{15}\right)^2\right]}} \\ &= \frac{95,290 - 83349.75}{\sqrt{[43,500 - 37,500][209,139 - 185,258.15]}} \\ &= \frac{11,940.25}{\sqrt{(6,000)(23,880.845)}} = \frac{11,940.25}{\sqrt{143,285,070}} \\ &= \frac{11,940.25}{11,970.174} = 0.9975 \end{aligned}$$



ดังนั้น แสดงว่า ปริมาณการผลิตและเวลาที่ใช้ในการผลิตมีความสัมพันธ์กัน 0.9975



6.4.1.2 สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนแรงค์

(Spearman Rank Correlation)

ถ้าข้อมูลที่ต้องการวัดความสัมพันธ์อยู่ในมาตราเรียงลำดับค่าสัมประสิทธิ์สหสัมพันธ์จะหาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนแรงค์ โดยสูตรที่ใช้จะเป็นดังนี้

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

เมื่อ ρ แทน ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์
 d แทน ผลต่างของอันดับที่
 n แทน จำนวนคู่ของข้อมูล





ตัวอย่าง จากการประเมินประสิทธิภาพการใช้งานของเครื่องคอมพิวเตอร์ 12
ยี่ห้อ โดยผู้เชี่ยวชาญ 2 คน ทำการจัดอันดับประสิทธิภาพการใช้งานได้ผลดัง
ตาราง

ยี่ห้อคอมพิวเตอร์	ผลการประเมินประสิทธิภาพการใช้งาน (เป็นอันดับที่)	
	ผู้เชี่ยวชาญคนที่ 1	ผู้เชี่ยวชาญคนที่ 2
1	2	3
2	3	4
3	5	5
4	8	6
5	6	8
6	4	2
7	1	1
8	7	7
9	10	9
10	9	11
11	12	10
12	11	12

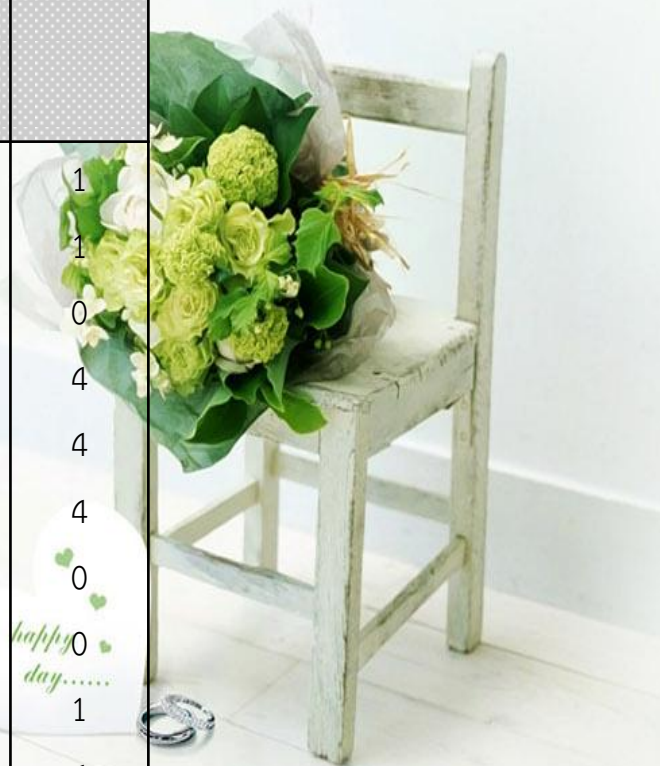


จึงหาค่าสัมประสิทธิ์
สหสัมพันธ์ระหว่างการจัดอันดับที่ของกรรมการ
สองคนที่ได้จัดไว้



วิธีทำ หาค่า d คือ หาผลต่างของอันดับที่ ที่ผู้เชี่ยวชาญทั้งสองจัดอันดับให้ คอมพิวเตอร์แต่ละยี่ห้อพร้อมทั้งหาค่ากำลังสองของ d แต่ละค่า แล้วหาผลรวม ทั้งหมดของ d^2 จะได้ดังนี้

ยี่ห้อคอมพิวเตอร์	ผลการประเมินประสิทธิภาพการใช้งาน (เป็นอันดับที่)		d	d^2
	ผู้เชี่ยวชาญคนที่ 1	ผู้เชี่ยวชาญคนที่ 2		
1	2	3	-1	1
2	3	4	-1	1
3	5	5	0	0
4	8	6	2	4
5	6	8	-2	4
6	4	2	2	4
7	1	1	0	0
8	7	7	0	0
9	10	9	1	1
10	9	11	-2	4
11	12	10	2	4
12	11	12	-1	1
รวม				24





แทนค่าต่างลงในสูตร ดังนี้

$$\begin{aligned}\rho &= 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \\ &= 1 - \frac{6(24)}{12(12^2 - 1)} \\ &= 1 - \frac{144}{12(143)} \\ &= 1 - \frac{144}{1716} \\ &= 1 - 0.084 = 0.916\end{aligned}$$

นั่นคือ การตัดสินใจของกรรมการทั้งสองมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.916





6.4.2 การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์

ในการทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์จะแบ่งออกเป็น 2 กรณีดังนี้คือ

6.4.2.1 การทดสอบว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันหรือไม่ จะกำหนดสมมติฐานการทดสอบ ดังนี้

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

หรือ $H_1 : \rho > 0$, $H_1 : \rho < 0$ เมื่อต้องการทดสอบว่าตัวแปรทั้งสองมีสหสัมพันธ์ในทิศทางตรงข้ามกันหรือไม่ ตัวสถิติที่ใช้ในการทดสอบคือ

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \text{ที่ } d.f = n - 2$$





6.4.2.2 การทดสอบสมมติฐานว่าตัวแปรทั้งสองมีความสัมพันธ์กันในระดับใด ๆ หรือไม่ จะกำหนดสมมติฐานในการทดสอบดังนี้ คือ

$$H_0 : \rho = \rho_0$$

$$H_1 : \rho \neq \rho_0 \text{ หรือ } H_1 : \rho > \rho_0 , H_1 : \rho < \rho_0$$

สำหรับตัวสถิติที่ใช้ในการทดสอบคือ

$$Z = \frac{Z_r - Z_p}{\sigma_{Z_r}}$$

เมื่อ

$$Z_r = \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right] , Z_p = \frac{1}{2} \ln \left[\frac{1+\rho_0}{1-\rho_0} \right] \text{ และ}$$

$$\sigma_{Z_r} = \sqrt{\frac{1}{n-3}}$$





ตัวอย่าง ในการทดลองใช้ยาเบื่อหนูปริมาณต่างๆ 15 ขนาด กับหนูนา แล้วนับจำนวนหนูนาที่ตายในแต่ละปริมาณของยาเบื่อหนู คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปริมาณยาเบื่อหนูและจำนวนหนูนาที่ตายได้เป็น 0.75 ที่ระดับนัยสำคัญ 0.01 จงทดสอบว่า

- ก. ปริมาณยาเบื่อหนูและจำนวนหนูนาที่ตายมีความสัมพันธ์กันหรือไม่
- ข. ปริมาณยาเบื่อหนูและจำนวนหนูนาที่ตายมีความสัมพันธ์กันน้อยกว่า 0.85 หรือไม่

วิธีทำ ก. 1. กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$





2. กำหนดตัวสถิติที่ใช้ในการทดสอบ

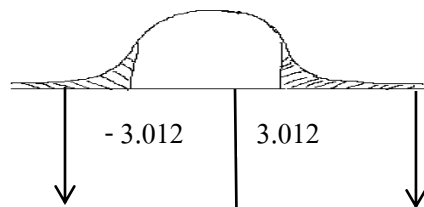
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \text{ที่ } d.f = n - 2$$

3. กำหนดบริเวณอาณาเขตวิกฤต

$$\alpha = 0.01, d.f. = n - 2 = 15 - 2 = 13$$

$$\text{เปิดตารางที่ } t_{\frac{\alpha}{2}, n-2} = t_{0.005, 13} = 3.012$$

การทดสอบเป็นแบบสองทาง ดังนั้นบริเวณ
อาณาเขตวิกฤต คือ



บริเวณอาณาเขตวิกฤต (ปฏิเสธ H_0)

บริเวณอาณาเขตวิกฤต (ปฏิเสธ H_0)

บริเวณยอมรับสมมติฐาน H_0





4. คำนวณค่าสถิติที่ใช้ในการทดสอบ

$$\begin{aligned} t &= \frac{0.75\sqrt{15-2}}{\sqrt{1-0.75^2}} \\ &= \frac{(0.75)(3.606)}{\sqrt{1-0.5625}} \\ &= \frac{2.7045}{\sqrt{0.4375}} \\ &= \frac{2.7045}{0.661} = 4.092 \end{aligned}$$

5. สรุปผล

ค่า t ที่คำนวณได้มีค่าเท่ากับ 4.092 จึงตกอยู่ในบริเวณอาณาเขตวิกฤตดังนั้นจะปฏิเสธ H_0 นั่นคือ ปริมาณยาเบื่อหนูและจำนวนหนูนาที่ตายมีความสัมพันธ์กัน





ข. เงินเดือนและค่าใช้จ่ายส่วนตัวของพนักงานมีสหสัมพันธ์กันน้อยกว่า 0.85 หรือไม่

1. กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \rho = 0.85$$

$$H_1 : \rho < 0.85$$

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ

$$Z = \frac{Z_r - Z_p}{\sigma_{Z_r}}$$

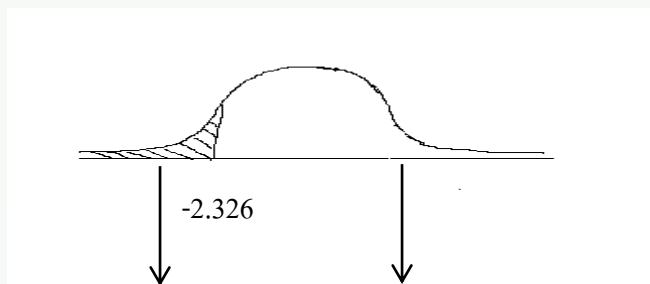
3. กำหนดบริเวณอาณาเขตวิกฤต

$$\alpha = 0.01 \text{ เปิดตารางที่ } Z_\alpha = Z_{0.01} = 2.326$$

การทดสอบเป็นแบบทางเดียวด้านซ้าย ดังนั้น

บริเวณอาณาเขตวิกฤต คือ บริเวณสถิติทดสอบ $Z < -2.326$





บริเวณอาณาเขตวิกฤต (ปฏิเสธ H_0) บริเวณยอมรับสมมติฐาน H_0

4. คำนวณค่าสถิติทดสอบ ดังนี้

$$\begin{aligned}
 Z_r &= \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right] \\
 &= \frac{1}{2} \ln \left[\frac{1+0.75}{1-0.75} \right] \\
 &= \frac{1}{2} \ln \left[\frac{1.75}{0.25} \right] = \frac{1}{2} \ln 7 \\
 &= \frac{1}{2} (1.946) = 0.973
 \end{aligned}$$





$$Z_p = \frac{1}{2} \ln \left[\frac{1 + \rho_0}{1 - \rho_0} \right]$$

$$= \frac{1}{2} \ln \left[\frac{1 + 0.85}{1 - 0.85} \right]$$

$$= \frac{1}{2} \ln \left[\frac{1.85}{0.15} \right] = \frac{1}{2} \ln 12.333$$

$$= \frac{1}{2} (2.512) = 1.256$$

$$\sigma_{Z_r} = \sqrt{\frac{1}{n-3}}$$

$$= \sqrt{\frac{1}{15-3}}$$

$$= \sqrt{\frac{1}{12}} = \sqrt{0.083} = 0.288$$





$$\begin{aligned}\therefore Z &= \frac{Z_r - Z_p}{\sigma_{Z_r}} \\ &= \frac{0.973 - 1.256}{0.288} \\ &= \frac{-0.283}{0.288} \\ &= -0.983\end{aligned}$$

5. สรุปผล

ค่า Z ที่คำนวณได้มีค่าเท่ากับ -0.983 จึงไม่ตกอยู่ในบริเวณอาณาเขตวิกฤตดังนั้นจะยอมรับ H_0 นั่นคือ ปริมาณยาเบื่อหนูและจำนวนหนูนานาที่ตายมีความสัมพันธ์กันไม่น้อยกว่า 0.85





6.5 บทสรุป

การวิเคราะห์ถดถอยอย่างง่าย เป็นวิธีการทางสถิติ ที่ศึกษาถึงความสัมพันธ์ระหว่างตัวแปรสองตัว โดยตัวแปรตัวหนึ่งเรียกว่า “ตัวแปรตาม (Dependent variable) แทนด้วย Y ” และตัวแปรอีกตัวหนึ่งเรียกว่า “ตัวแปรต้นหรือตัวแปรอิสระ (Independent variable) แทนด้วย X ” ในรูปแบบทางคณิตศาสตร์ โดยมีประโยชน์เพื่อใช้ในการทำนาย (Prediction) นั่นคือจะใช้ตัวแปรอิสระ X เป็นเกณฑ์ในการประมาณตัวแปรตาม Y

แต่ถ้าหากไม่ได้กำหนดว่าตัวแปรใดเป็นตัวแปรตาม ตัวแปรใดเป็นตัวแปรต้นหรือตัวแปรอิสระ จะเป็นเรื่องของการวิเคราะห์สหสัมพันธ์ (correlation analysis) ซึ่งเป็นการศึกษาว่าตัวแปรคู่่นั้นมีขนาดของความสัมพันธ์ระดับใด โดยวัดด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) และถ้ามีข้อสงสัยว่าตัวแปรแต่ละคู่มีสหสัมพันธ์กันจริงหรือไม่ ก็สามารถทำการทดสอบสมมติฐานได้ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าอยู่ระหว่าง -1 ถึง 1

