บทที่ 12
การถดถอยและสหสัมพันธ์

การถดถอย เป็นวิธีการอันหนึ่งที่จะใช้เป็นเครื่องมือตรวจหาลักษณะธรรมชาติของความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรขึ้นไป ทั้งนี้เพื่อประโยชน์ในการประมาณค่าสิ่งที่เราไม่ทราบโดยอาศัยค่าสังเกตหรือข้อมูลจากปรากฏการณ์ที่ผ่านมา อย่างเช่น การศึกษาอิทธิพลของปุ๋ยที่ใช้ใน 1 ไร่ของข้าวสาลี (X) ต่อผลผลิตที่ได้ใน 1 ไร่ (Y) สิ่งที่เราต้องการทราบคืออิทธิพลของปุ๋ย (X) ที่มีต่อผลผลิต (Y) ในที่นี้ X เป็นตัวแปรอิสระ Y เป็นตัวแปรตาม สังเกตว่าตามความเป็นจริงแล้วผลผลิตข้าวสาลี (Y) ไม่ได้ขึ้นอยู่กับจำนวนปุ๋ย (X_1) เพียงอย่างเดียว ยังขึ้นอยู่กับตัวแปรอิสระอื่น ๆ เช่น ดินฟ้าอากาศ (X_2) พันธุ์ (X_3) น้ำ (X_4) การดูแลรักษา (X_5) ฯลฯ. แต่ถ้าเราพิจารณาอิทธิพลเพียงปัจจัยเดียวคือตัวแปรอิสระเพียงตัวเดียว เรียก การถดถอยนั้นว่า การถดถอยอย่างง่าย เมื่อใดที่เราพิจารณาอิทธิพลของปัจจัยมากกว่าหนึ่งตัว การถดถอยนั้นเรียกว่า การถดถอยเชิงซ้อน

ความสัมพันธ์ของ X กับ Y อาจเป็นรูปเส้นตรงหรือเส้นโค้ง ฯ. แต่ในที่นี้เราจะอธิบายเฉพาะความสัมพันธ์ในรูปของเส้นตรงที่เรียกว่า “การถดถอยอย่างง่ายที่เป็นเส้นตรง” การศึกษาเรื่อง การถดถอยนั้น เราจะต้องอาศัยความรู้ในอดีตหรือรายละเอียดเกี่ยวกับข้อมูลที่จะนำมาวิเคราะห์ เพื่อที่จะมาพิจารณาชี้ให้เห็นว่า ข้อมูลชุดใดเป็นตัวแปรอิสระ (X) หรือเป็นสาเหตุ ข้อมูลชุดใดเป็นตัวแปรตาม (Y) หรือเป็นผล

การถดถอยอย่างง่ายที่เป็นเส้นตรง

เป็นเรื่องของการวิเคราะห์ความสัมพันธ์ของตัวแปร 2 ชุด โดยที่ตัวแปรชุดหนึ่งเป็นตัวแปรอิสระหรือสาเหตุเราใช้อักษร X_i ซึ่งเป็นตัวที่มีอิทธิพลต่อตัวแปรอีกชุดหนึ่ง เป็นตัวแปรตามหรือผลใช้อักษร Y_i ค่าของ Y_i จะมากหรือน้อยหรือเปลี่ยนแปลงไปในรูปใดต้องขึ้นอยู่กับค่าของ X_i โดยเราจะศึกษาว่า ถ้า X_i เปลี่ยนไป 1 หน่วย จะทำให้ค่าของ Y_i เปลี่ยนไปกี่หน่วย อย่างเช่น จำนวนเงินรายได้กับจำนวนเงินออมทรัพย์ ผู้มีรายได้มากย่อมมีโอกาสออมทรัพย์ได้มาก ผู้มีรายได้น้อยย่อมมีโอกาสออมทรัพย์ได้น้อย จำนวนเงินรายได้เป็น X_i จำนวนเงิน

ออมทรัพย์เป็น Y_i หรือเรามักจะได้ยินอยู่เสมอว่าผู้มีรายได้มาก (X_i) ย่อมมีรายจ่ายมาก (Y_i) เป็นเงาตามตัว

การวิเคราะห์ การถดถอยและสหสัมพันธ์นี้ถ้าเราวิเคราะห์โดยใช้ข้อมูล 2 ชุด ชุดหนึ่งเป็นสาเหตุและอีกชุดหนึ่งเป็นผล เราถือว่าเป็นการวิเคราะห์แบบ การถดถอยอย่างง่าย และสัมพันธ์อย่างง่าย เช่น ปุ๋ยเป็นสาเหตุหรือตัวแปรอิสระกับผลิตผลเป็นผลหรือตัวแปรตามความตั้งใจเรียนเป็นสาเหตุหรือตัวแปรอิสระกับผลสอบเป็นผลหรือตัวแปรตาม แต่ถ้าเราวิเคราะห์การถดถอยหรือสหสัมพันธ์ โดยใช้ข้อมูลมากกว่า 2 ชุด โดยที่ให้ชุดหนึ่งเป็นผลและชุดอื่น ๆ เป็นสาเหตุซึ่งมีมากกว่า 1 สาเหตุเราถือว่าเป็นการวิเคราะห์ การถดถอยเชิงซ้อนหรือสหสัมพันธ์เชิงซ้อน ดังตัวอย่าง เราต้องการวิเคราะห์ความสัมพันธ์ระหว่างผลิตผลของดอกกล้วยไม้เพื่อดูว่า จะมีปริมาณมากน้อยเท่าใด ปริมาณเพิ่มขึ้นหรือลดลงเท่าใดขึ้นอยู่กับสาเหตุใดบ้างซึ่งสาเหตุดังกล่าวอาจจะเป็นปุ๋ย พันธ์ น้ำ การรักษาดูแล ดินฟ้าอากาศ ฯลฯ

การวิเคราะห์ การถดถอยและสหสัมพันธ์จะต้องคำนึงถึงรูปลักษณะของเส้นที่แสดงความสัมพันธ์ที่คาดว่า ความสัมพันธ์จะเป็นไปในรูป ซึ่งอาจจะเป็นเส้นตรงหรือเส้นโค้งเพื่อที่จะเลือกแบบของสมการที่จะนำมาเป็นหลักในการวิเคราะห์ได้ถูกต้อง จึงพอสรุปได้ว่าในเนื้อเรื่องของคำบรรยายเกี่ยวกับการวิเคราะห์ การถดถอยและสหสัมพันธ์นี้อาจจะแยกหัวข้อได้ดังนี้

การวิเคราะห์ การถดถอยแบ่งออกเป็น

(1) การวิเคราะห์ การถดถอยอย่างง่าย
- (ก) การถดถอยอย่างง่ายที่เป็นเส้นตรง
- (ข) การถดถอยอย่างง่ายที่ไม่เป็นเส้นตรง

(2) การวิเคราะห์ การถดถอยเชิงซ้อน
- (ก) การถดถอยเชิงเส้นที่มีหลาย ๆ ตัวแปรอิสระ
- (ข) การถดถอยไม่เชิงเส้นที่มีหลาย ๆ ตัวแปรอิสระ

การวิเคราะห์สหสัมพันธ์แบ่งออกเป็น

(1) การวิเคราะห์สหสัมพันธ์อย่างง่ายที่เป็นเส้นตรง
- (ก) สหสัมพันธ์อย่างง่ายที่เป็นเส้นตรง
- (ข) สหสัมพันธ์อย่างง่ายที่ไม่เป็นเส้นตรง

(2) การวิเคราะห์สหสัมพันธ์เชิงซ้อน

(ก) สหสัมพันธ์เชิงเส้นที่มีหลาย ๆ ตัวแปรอิสระ

(ข) สหสัมพันธ์ไม่เชิงเส้นที่มีหลาย ๆ ตัวแปรอิสระ

แต่สำหรับในบทนี้จะอธิบายแต่วิธีวิเคราะห์ การถดถอยอย่างง่ายที่เป็นเส้นตรง และวิธีวิเคราะห์สหสัมพันธ์อย่างง่ายที่เป็นเส้นตรงเท่านั้น

ในการวิเคราะห์ การถดถอยอย่างง่ายที่เป็นเส้นตรง เรามีตัวแบบทางคณิตศาสตร์ หรือสมการถดถอยของประชากร (Population Regression) ซึ่งเป็นสมการที่แสดงถึงความสัมพันธ์ที่แท้จริงของตัวแปร 2 ตัว สมการที่แสดงถึงความสัมพันธ์ที่แท้จริงของการถดถอยเชิงเส้นของประชากรเขียนได้เป็น

$$Y = \beta_0 + \beta_1 X + \epsilon$$

ในที่นี้ ค่าของ Y แต่ละค่าจะประกอบด้วยส่วนใหญ่ ๆ 2 ส่วน คือ $\beta_0 + \beta_1 X$ กับ ϵ

$\beta_0 + \beta_1 X$ หมายถึงค่าที่คาดหวังว่า Y ควรจะเป็นโดยผลของ X โดยที่

β_0 เป็นจุดตัดบนแกน Y เมื่อ X มีค่าเป็นศูนย์

β_1 เป็นความชันของเส้นถดถอย

ϵ เป็นค่าคาดหวังที่ทำให้ Y คลาดเคลื่อนไปจากสาเหตุอื่นนอกเหนือไปจาก X

แต่ในทางปฏิบัติ เราไม่สามารถสังเกตค่าของประชากรได้หมด นั่นคือเราไม่สามารถสังเกตความสัมพันธ์จริงนี้ได้ สิ่งที่เราจะทำได้ก็เพียงหาค่าประมาณเพื่อที่จะเป็นตัวแทนของประชากร ตัวประมาณค่าที่ดีนั้นจะต้องมีคุณสมบัติคือไม่มีความเอนเอียง มีความแปรปรวนน้อยที่สุด วิธีการกำลังสองน้อยที่สุดเป็นวิธีการที่เข้ากับคุณสมบัติข้างต้นในการกำหนดเส้นตรงให้กับข้อมูล สมการของความถดถอยเชิงเส้นเขียนได้ดังนี้

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

ค่าของ Y แต่ละตัวจะประกอบด้วย $\hat{\beta}_0 + \hat{\beta}_1 X$ กับ e ซึ่งมีความหมายเช่นเดียวกับสมการถดถอยของประชากรคือ $\hat{\beta}_0 + \hat{\beta}_1 X$ แสดงถึงส่วนหนึ่งของค่า Y ที่คาดหวังจากผลของ X

$\hat{\beta}_0$ เป็นจุดตัดบนแกน Y ซึ่งเป็นตัวค่าประมาณของ β_0

$\hat{\beta}_1$ เป็นความชันของเส้นถดถอยซึ่งเป็นตัวค่าประมาณของ β_1

และ e เป็นค่าความแตกต่างหรือค่าคลาดเคลื่อนระหว่างค่า Y กับเส้นถดถอย ซึ่ง

เกิดขึ้นเนื่องมาจากสาเหตุอื่นที่นอกเหนือไปจาก X และเป็นตัวค่าประมาณของ ϵ

$$\text{จาก } Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

$$Y - e = \hat{\beta}_0 + \hat{\beta}_1 X$$

ให้ $Y - e$ เป็น $\hat{Y}$ ซึ่งหมายถึงตัวค่าประมาณของ Y

$$\therefore \quad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

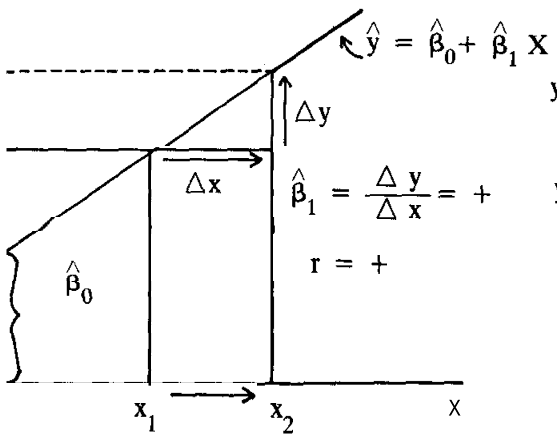
เพราะฉะนั้น สมการถดถอยอย่างง่ายที่เป็นเส้นตรงของตัวอย่างที่เราต้องการวิเคราะห์คือสมการข้างต้น

สมการ $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ อาจเรียกว่าสมการเส้นถดถอย Y on X, $\hat{Y}$ อาจเขียน Y_c, Y' ได้เช่นเดียวกัน

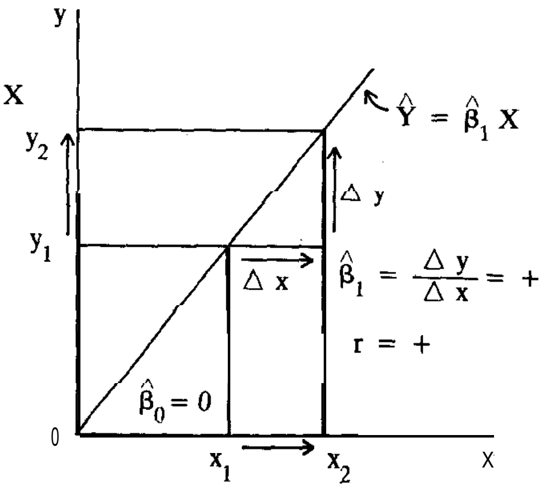
$\hat{\beta}_0$ เป็นจุดตัดบนแกน Y มีค่าอาจ เป็น +, – หรือ 0

$\hat{\beta}_1$ เป็นตัววัดค่าความชันของเส้นถดถอย เรียกว่าสัมประสิทธิ์ การถดถอย ค่า $\hat{\beta}_1$ นี้อาจจะเป็น +, – หรือ 0

ถ้า $\hat{\beta}_1$ มีค่าเป็นบวกแสดงว่าค่าของ X มีผลที่ทำให้ Y มีค่าเปลี่ยนแปลงไปในทางเดียวกัน คือ X มีค่าเพิ่ม Y จะมีค่าเพิ่มตามและถ้า X มีค่าลดลง Y ก็จะมีค่าลดตาม ถ้าเขียนกราฟเส้นถดถอย จะได้ดังภาพ

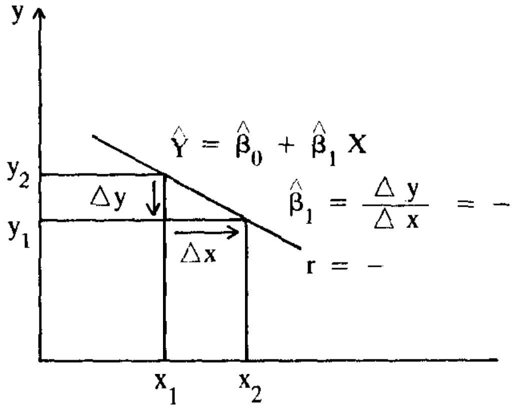


รูปที่ 1

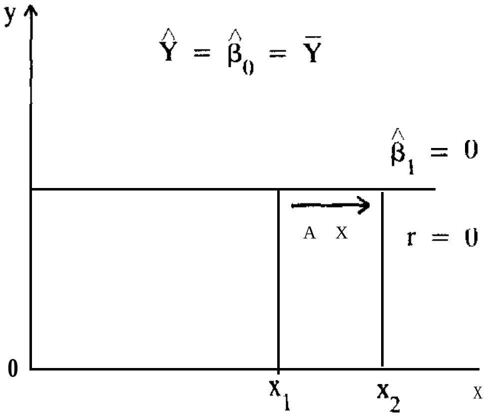


รูปที่ 2

ถ้า $\hat{\beta}_1$ มีค่าเป็นลบแสดงว่าค่าของ X มีผลที่ทำให้ Y มีค่าเปลี่ยนแปลงไปในทางตรงข้ามกันคือ X มีค่าเพิ่มขึ้น Y จะมีค่าลดลงหรือ X มีค่าลดลง Y จะมีค่าเพิ่มขึ้น ถ้าเขียนกราฟเส้นถดถอยจะได้ดังภาพ



รูปที่ 3



รูปที่ 4

ถ้า $\hat{\beta}_1 = 0$ แสดงว่า X ไม่มีผลต่อ Y และ กล่าวคือ X จะเปลี่ยนอย่างไรก็ไม่ทำให้ Y มีการเปลี่ยนแปลงเลย ถ้าเราเขียนกราฟเส้นถดถอยจะได้ดังรูปที่ 4

สมการเส้นถดถอยมีสองชนิดด้วยกัน หนึ่งสำหรับทำนาย Y จาก X สมการเขียนได้เป็น $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ สองสำหรับทำนาย X จาก Y สมการเขียนได้เป็น $\hat{X} = \hat{\beta}_0^1 + \hat{\beta}_1^1 Y$ โดยทั่ว ๆ ไปเรามักจะพบสมการที่หนึ่ง ซึ่งใช้กันมากดังรายละเอียดจะกล่าวต่อไป

คุณสมบัติของ การถดถอยเชิงเส้นที่คำนวณได้โดยวิธีการกำลังสองน้อยที่สุด

(1) $\bar{X}, \bar{Y}$ จะอยู่บนเส้นถดถอยหรือเส้นถดถอยจะผ่านจุด $(\bar{X}, \bar{Y})$

(2) ผลบวกของส่วนเบี่ยงเบนจากเส้นถดถอยจะมีค่าเป็นศูนย์ $\Sigma(Y - \hat{Y}) = 0$

(3) ผลบวกของส่วนเบี่ยงเบนจากเส้นถดถอยกำลังสอง $\Sigma(Y - \hat{Y})^2$ มีค่าน้อยที่สุด

จากคุณสมบัติข้อที่ 3 ถ้าเรา differentiate เทียบกับ $\hat{\beta}_0$ และ $\hat{\beta}_1$ แล้วทำให้เท่ากับศูนย์ จะได้สมการปกติ สำหรับหาค่า $\hat{\beta}_0, \hat{\beta}_1$ ที่จะทำให้ $\Sigma(Y - \hat{Y})^2$ มีค่าน้อยที่สุด สมการปกตินั้นเขียนได้เป็น

$$n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X = \Sigma Y \qquad (1)$$

$$\hat{\beta}_0 \Sigma X + \hat{\beta}_1 \Sigma X^2 = \Sigma XY \qquad (2)$$

ΣX, ΣX^2, ΣXY และ ΣY คำนวณหาได้จากข้อมูลที่เราได้มาแล้ว นำไปแทนค่าลงในสมการ (1) และ (2) จากนั้นก็ทำการหาคำตอบของค่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ ได้โดยวิธีการทางพีชคณิตได้

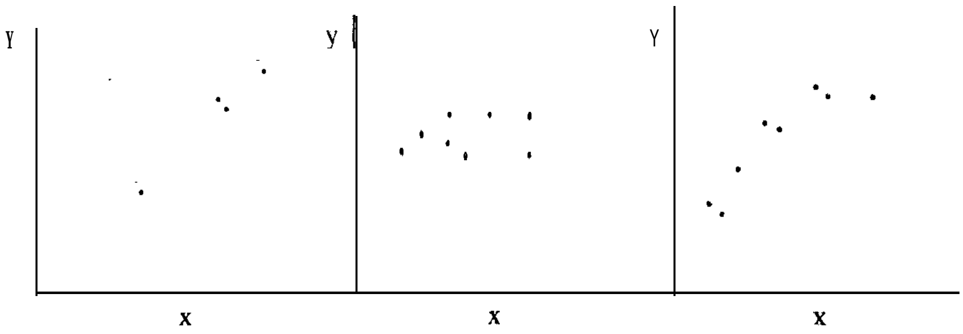
$$\hat{\beta}_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2}$$

และ

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

เมื่อคำนวณหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$ ได้แล้วก็นำไปแทนค่าในสมการเส้นถดถอย $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ เมื่อต้อง การประมาณค่า Y ก็หาค่า $\hat{Y}$ โดยแทนค่า X ลงในสมการนี้ ค่า $\hat{\beta}_0$ เป็นจุดตัดแกน Y เมื่อ X มีค่าเป็นศูนย์ ส่วนค่า $\hat{\beta}_1$ เป็นความชันหรืออัตราการเปลี่ยน-แปลงระหว่าง X กับ Y คือ เมื่อ X เปลี่ยนแปลงไป 1 หน่วย Y จะเปลี่ยนแปลงไป $\hat{\beta}_1$ หน่วย ดูจากรูปที่ $\hat{\beta}_1 \frac{\Delta y}{\Delta x}$ ถ้า $\Delta x = 1$ ค่า $\hat{\beta}_1$ จะเท่ากับ Δy นั่นคือการเปลี่ยนแปลงของ Y จะมีค่าเท่ากับ $\hat{\beta}_1$ หน่วย ในทางเพิ่มขึ้น เมื่อการเปลี่ยนแปลงของ X = 1 หน่วย แต่ถ้ามาดูรูปที่ 3 การเปลี่ยนแปลงของ Y จะมีค่าเท่ากับ $\hat{\beta}_1$ หน่วยในทางลดลง เมื่อการเปลี่ยนแปลงของ X = 1 หน่วย

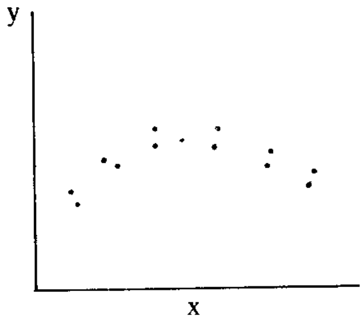
ลักษณะตัวแบบที่ใช้กำหนดกับข้อมูล



กรณีที่ 1 **ใช้ตัวแบบ** $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

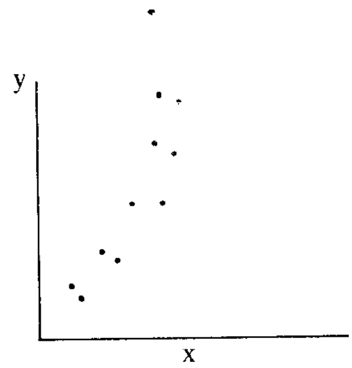
กรณีที่ 2 **ใช้ตัวแบบ** $\hat{Y} = \bar{Y}$

กรณีที่ 3 **พยายามใช้ตัวแบบ** $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_{11} X^2$



กรณีที่ 4

พยายามใช้ตัวแบบ $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_{11} X^2$



กรณีที่ 5

พยายามใช้ตัวแบบ $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_{11} X^2$

ตัวอย่างการวิเคราะห์ การถดถอยอย่างง่ายที่เป็นเส้นตรง

ในการสำรวจนักศึกษา 12 คน เกี่ยวกับความสัมพันธ์ว่าคะแนนทดสอบเชาว์ปัญญามีอิทธิพลต่อเกรดเฉลี่ยมากน้อยแค่ไหน ดังตาราง

เกรดเฉลี่ย	2.1	2.2	3.1	2.3	3.4	2.9	2.9	2.7	2.1	1.7	3.3	3.5
เชาว์ปัญญา	116	129	123	121	131	134	126	122	114	118	132	129

วิธีทำ สร้างตารางเพื่อคำนวณหา ΣX, ΣY, ΣX^2, ΣXY แทนลงในสูตรเพื่อหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$ จากสูตร

$$\hat{\beta}_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

ตัวแบบ

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

เกรดเฉลี่ย (Y)	เชาว์ปัญญา (X)	X^2	Y^2	XY	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
2.1	.16	13456	4.41	243.6	2.1192	−0.0192	.0004
2.2	129	16641	4.48	283.8	2.9733	−0.7733	.5929
3.1	123	15129	6.61	381.3	2.5791	0.5209	.2704
2.3	121	14641	5.29	278.3	2.4477	−0.1477	.0225
3.4	131	17161	11.56	445.4	3.1047	0.2953	.0900
2.9	134	17956	8.41	388.6	3.3018	−0.4018	.1600
2.9	126	15876	8.41	365.4	2.7762	0.1238	.0144
2.7	122	14884	7.29	329.4	2.5134	0.1866	.0361
2.1	114	12996	4.41	239.4	1.9878	0.1122	.0121
1.7	118	13924	2.89	200.6	2.2506	−0.5506	.3025
3.3	132	17424	10.89	435.6	3.1704	0.1296	.0169
3.5	129	16641	12.25	451.5	2.9733	0.5267	.2769
32.2	1495	186729	90.26	4042.9			2.0651

ในที่นี้เราได้ $\Sigma X = 1495$, $\Sigma Y = 32.2$, $\Sigma XY = 4042.9$ แทนลงในสูตร

$$\hat{\beta}_1 = \frac{12\,(4042.9) - (1495)\,(32.2)}{12\,(186729) - (1495)^2}$$

$$= \frac{48514.8 - 48139}{2240748 - 2235025} = \frac{375.8}{5723}$$

$$= 0.0657$$

$$\hat{\beta}_0 = \frac{32.2}{12} - (0.0657)\,\frac{(1495)}{12}$$

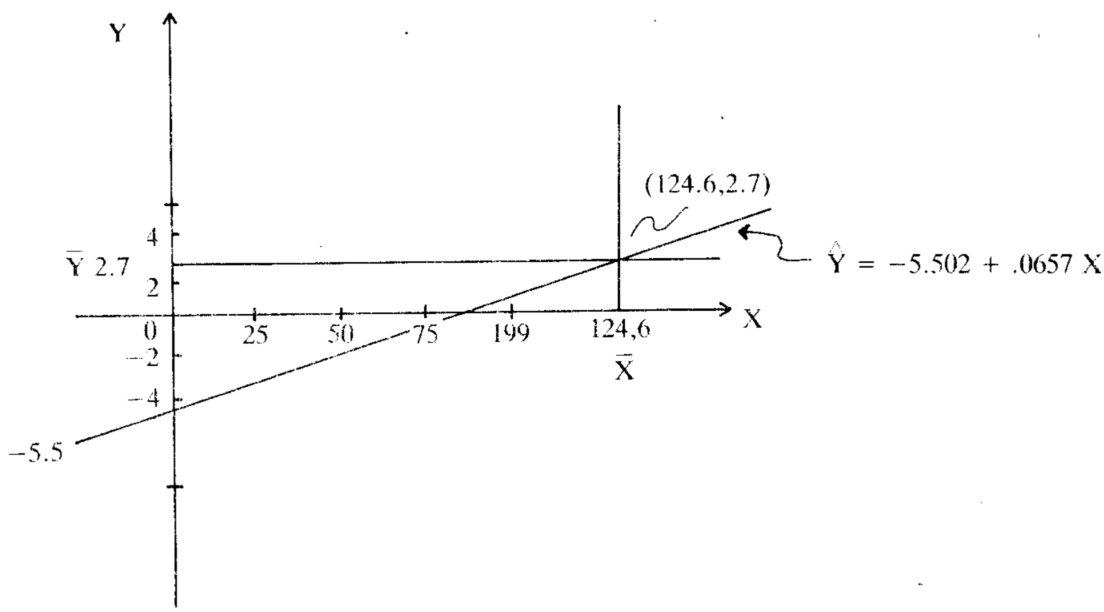
$= 2.683 - 8.185$

$= - 5.502$

แทนค่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ ลงในตัวแบบ

$\therefore$ สมการถดถอย $\hat{Y} = -5.502 + 0.0657\,X$

จะเห็นว่าค่า $\hat{\beta}_1$ แสดงถึงความสัมพันธ์ของ X กับ Y $\hat{\beta}_1$ มีค่าเป็นบวก แสดงว่าความสัมพันธ์ของ X กับ Y ไปทางเดียวกัน $\hat{\beta}_1 = 0.0657$ หมายความว่าเกรดเฉลี่ย (Y) จะเพิ่มขึ้นโดยเฉลี่ย 0.0657 หน่วยต่อเชาว์ปัญญา (X) ที่เพิ่มขึ้น 1 หน่วย ถ้าหากว่านักศึกษามีเชาว์ปัญญา (X) 140 และเราใช้สมการ $\hat{Y} = -5.502 + .065X$ ประมาณค่า Y ค่า $\hat{Y}$ จะมีค่าเท่ากับ $-5.502 + 0.0657\ (140) = 3.696$ หน่วย อาจจะมีความคลาดเคลื่อนไปจากความจริงที่ควรเป็น เพราะเกรดเฉลี่ยของนักศึกษาไม่ได้มีผลมาจากเชาว์ปัญญาอย่างเดียว ยังมีสาเหตุที่มีผลต่อเกรดเฉลี่ยอีกหลายสาเหตุ ทั้งนี้ต้องเข้าใจว่าตราบใดที่เราใช้สมการถดถอย $\hat{Y} = 5.502 + 0.0657\ X$ เป็นหลักในการคาดคะเนเกรดเฉลี่ย เราคำนึงถึงอิทธิพลของเชาว์ปัญญาเพียงอย่างเดียว ส่วนค่าคลาดเคลื่อนจากสาเหตุอื่นนั้น ให้นักศึกษาดูตารางคอลัมน์ที่ 7



ถ้าลองพลอตกราฟเพื่อพิจารณาเส้นถดถอยของตัวอย่างนี้ จะได้รูปข้างต้น จากรูป $\hat{\beta}_0 = -5.502$ แสดงถึงเส้นถดถอยตัดแกน Y ที่จุด -5.502

ความแปรปรวนของ Y

จากสมการที่แสดงถึงความสัมพันธ์ที่แท้จริงของ การถดถอยเชิงเส้นของประชากร $Y = \beta_0 + \beta_1 X + \epsilon$ โดยที่ ϵ เป็นค่าคลาดเคลื่อน มีมัชฌิมเลขคณิตศูนย์และความแปรปรวน $\sigma^2_{y/x}$ (ไม่ทราบค่า) เราจึงใช้ตัวประมาณค่า $S^2_{y/x}$ แทนค่า $\sigma^2_{y/x}$ และคำนวณหาได้จากสูตร

$$S^2_{y/x} = \frac{\Sigma (Y - \hat{Y})^2}{n - 2}$$

หรือ

$$S^2_{y/x} = \frac{1}{n-2}\left[\Sigma Y^2 - n\bar{Y}^2 - \frac{(\Sigma XY - n\bar{X}\bar{Y})^2}{\Sigma X^2 - n\bar{X}^2}\right]$$

มัชฌิมเลขคณิตและความแปรปรวนของ $\hat{\beta}_0$ และ $\hat{\beta}_1$

เราทราบแล้วว่า $\hat{\beta}_0$ เป็นตัวค่าประมาณของ β_0 และ $\hat{\beta}_1$ เป็นตัวค่าประมาณของ β_1 ทั้ง $\hat{\beta}_0$ และ $\hat{\beta}_1$ ต่างก็เป็นตัวค่าประมาณที่ดีและปราศจากความเอนเอียง นั่นคือ

$$E(\hat{\beta}_0) = \beta_0$$

$$V(\hat{\beta}_0) = \frac{\Sigma X^2 \sigma^2_{y/x}}{n \Sigma (X - \bar{X})^2}$$

$$E(\hat{\beta}_1) = \beta_1$$

$$V(\hat{\beta}_1) = \frac{\sigma^2_{y/x}}{\Sigma (X - \bar{X})^2}$$

เรามาพิสูจน์ค่าคาดหวังและค่าความแปรปรวนของ $\hat{\beta}_1$ เท่านั้น จากสูตร

$$\hat{\beta}_1 = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2}$$

$$= \frac{\Sigma (X - \bar{X}) Y_i}{\Sigma (X_i - \bar{X})^2}$$

เนื่องจากเทอม $\Sigma (X_i - \bar{X})\bar{Y} = \bar{Y}\Sigma (X - \bar{X}) = 0$

ดังนั้น

$$\hat{\beta}_1 = \{ (X_1 - \bar{X}) Y_1 + (X_2 - \bar{X}) Y_2 + \ldots + (X_n - \bar{X}) Y_n \} / \Sigma (X - \bar{X})^2$$

ให้ $a_i = (X_i - \bar{X}) / \Sigma (X_i - \bar{X})^2$ และเป็นตัวคงที่

$$\hat{\beta}_1 = a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n$$

$$= \Sigma a_i Y_i$$

ค่าคาดหวังของ $\hat{\beta}_1 = \Sigma a_i Y_i$

$$E(\hat{\beta}_1) = \Sigma a_i E(Y_i)$$

$$= \Sigma a_i (\beta_o + \beta_1 X_i)$$

$$= \beta_o \Sigma a_i + \beta_1 \Sigma a_i X_i$$

$$= \beta_o \Sigma \frac{(X_i - \bar{X})}{\Sigma (X_i - \bar{X})^2} + \beta_1 \Sigma \frac{(X_i - \bar{X}) X_i}{\Sigma (X_i - \bar{X})^2}$$

เนื่องจากว่า

$$\Sigma (X_i - \bar{X}) = 0$$

$$E(\hat{\beta}_1) = 0 + \beta_1 \Sigma \frac{(X_1 - \bar{X})(X_i - \bar{X})}{\Sigma (X_i - \bar{X})^2}$$

$$= 0 + \beta_1 \frac{\Sigma (X_i - \bar{X})^2}{\Sigma (X_i - \bar{X})^2}$$

$$= \beta_1$$

สำหรับความแปรปรวนของ $\hat{\beta}_1$ เราได้

$$V(\hat{\beta}_1) = \Sigma a_i^2 V(Y_i)$$

$$= \Sigma a_i^2 \sigma^2_{y/x}$$

$$= \Sigma \frac{(X_i - \bar{X})^2}{[\Sigma (X_i - \bar{X})^2]^2} \sigma^2_{y/x}$$

$$= \frac{\Sigma (X_i - \bar{X})^2 \sigma^2_{y/x}}{[\Sigma (X_i - \bar{X})^2]^2}$$

$$= \frac{\sigma^2_{y/x}}{\Sigma (X_i - \bar{X})^2}$$

เมื่อเราทราบค่าคาดหวังและความแปรปรวนของ $\hat{\beta}_0$, $\hat{\beta}_1$ ก็เท่ากับว่าเราทราบการแจกแจงของ $\hat{\beta}_0$, $\hat{\beta}_1$

การทดสอบสมมติฐานและช่วงความเชื่อมั่นของ β_1

เมื่อไรเราทราบการแจกแจงของ $\hat{\beta}_1$ ว่ามีการแจกแจงปกติ มีมัชฌิมเลขคณิต β_1 ความแปรปรวน $\sigma^2_{y/x} / \Sigma (X_i - \bar{X})^2$ การทดสอบเกี่ยวกับ β_1 ก็สามารถใช้การทดสอบแบบ Z ได้

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\Sigma (X_i - \bar{X})^2}}}$$

แต่โดยทั่ว ๆ ไปแล้วเราไม่ทราบความแปรปรวน $\sigma^2_{y/x}$ จึงจำเป็นต้องประมาณด้วย $S^2_{y/x}$ ซึ่งเป็นตัวประมาณค่าที่ปราศจากความเอนเอียงของ $\sigma^2_{y/x}$ เมื่อไรที่ประมาณค่า $\sigma^2_{y/x}$ ด้วย $S^2_{y/x}$ แล้ว ค่าประมาณ $\hat{\beta}_1$ จะมีการแจกแจงแบบ t นั่นคือ

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{S^2_{y/x}}{\Sigma (X_i - \bar{X})^2}}}$$

มีองศาแห่งความอิสระเป็น n − 2

ช่วงความเชื่อมั่นของ β_1

จากตารางเราทราบว่าพื้นที่ใต้โค้งการแจกแจงของ t เท่ากับ 0.95 ระหว่างค่า $-t_{.025}$ กับ $+t_{.025}$ ที่องศาแห่งความอิสระ n − 2 นั่นคือ

$$P(-t_{.025} < t < +t_{.025}) = 0.95$$

ถ้าหากว่าเราเขียน t ในรูป $\dfrac{\hat{\beta}_1 - \hat{\beta}_1}{S_{y/x}/\sqrt{\Sigma (X - X)^2}}$ เพื่อหาค่าช่วงของค่าอะไรที่ครอบคลุมค่า β_1 ด้วยความน่าจะเป็น 0.95 เราจะได้

$$P\left(-t_{.025} < \frac{\hat{\beta}_1 - \hat{\beta}_1}{S_{y/x}/\sqrt{\Sigma (X - \bar{X})^2}} < +t_{.025}\right) = 0.95$$

สมการไม่เท่ากันในวงเล็บแสดงได้เป็น

$$P\left(\hat{\beta}_1 - t_{.025}\frac{S_{y/x}}{\sqrt{\Sigma (X - \bar{X})^2}} < \beta_1 < \hat{\beta}_1 + t_{.025}\frac{S_{y/x}}{\sqrt{\Sigma (X - \bar{X})^2}}\right) = 0.95$$

ซึ่งให้ 95% ช่วงความเชื่อมั่นสำหรับ β_1 กล่าวคือ

$$\hat{\beta}_1 \pm t_{.025}\frac{S_{y/x}}{\sqrt{\Sigma (X - \bar{X})^2}}$$

ในเมื่อ $\hat{\beta}_1 - t_{.025}\dfrac{S_{y/x}}{\sqrt{\Sigma (X - \bar{X})^2}}$ เป็นขีดจำกัดล่าง

$\hat{\beta}_1 + t_{.025}\dfrac{S_{y/x}}{\sqrt{\Sigma (X - \bar{X})^2}}$ เป็นขีดจำกัดบน

จากตัวอย่างก่อนระหว่างเกรดเฉลี่ยกับเชาว์ปัญญา หากเราต้องการทดสอบว่าเชาว์ปัญญาไม่มีอิทธิพลต่อเกรดเฉลี่ยโดยเปรียบเทียบกับ alternative สองข้าง เราจะได้ว่า

$$\text{Ho}: \beta_1 = 0$$
$$\text{Ha}: \beta_1 \neq 0$$

ไม่ยอมรับ Ho ที่ระดับนัยสำคัญ 5%

$$t = \frac{\hat{\beta}_1 - \beta_1}{S_{y/x} / \sqrt{\Sigma (X - \bar{X})^2}}$$

ในเมื่อ $$S^2_{y/x} = \frac{\Sigma (Y - \hat{Y})^2}{n - 2} = \frac{2.0651}{10}$$

$$= 0.20651$$

$$\Sigma (X - \bar{X})^2 = 476.917$$

$$\therefore t = \frac{0.0657}{\sqrt{\frac{0.20651}{476.917}}}$$

$$= \frac{0.0657}{0.0208} = 3.159$$

เราจะไม่ยอมรับ Ho เมื่อ $t > t_{.025\,;\,10} = 2{,}228$ หรือ $t < -t_{.025\,:\,10} = -2.228$

$\therefore$ เราจึงไม่สามารถยอมรับ Ho ที่ระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าเชาว์ปัญญามีอิทธิพลต่อเกรดเฉลี่ยหรือเชาว์ปัญญามีความสัมพันธ์เชิงเส้นกับเกรดเฉลี่ย (Ho : $\beta_1 = 0$ หมายความว่าเชาว์ปัญญากับเกรดเฉลี่ยไม่มีความสัมพันธ์เชิงเส้น)

ความคลาดเคลื่อนมาตรฐานของ $\hat{Y}$

เราทราบแล้วว่าสมการถดถอย คือ

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

ในเมื่อ

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\therefore \hat{Y} = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X$$

$$= \bar{Y} + \hat{\beta}_1 (X - \bar{X})$$

ในเมื่อทั้ง $\bar{Y}$ และ $\hat{\beta}_1$ มีส่วนทำให้ $\hat{Y}$ มีความคลาดเคลื่อนและมีความอิสระต่อกันด้วย ดังนั้นความแปรปรวนของค่าคาดหวังของ Y คือ $\hat{Y}_k$ เมื่อกำหนดค่า X_k ของ X คือ

$$V(\hat{Y}) = V(\bar{Y}) + V[\hat{\beta}_1 (X_k \quad \bar{X})] + \text{cov}(\bar{Y}, \hat{\beta}_1 (X_k - \bar{X}))$$

$$\text{Cov}(\bar{Y}, \hat{\beta}_1 (X_k - \bar{X})) = 0$$

$$V(\hat{Y}_k) = V(\bar{Y}) + V[\hat{\beta}_1 (X_k - \bar{X})]$$

$$= V(\bar{Y}) + (X_k - \bar{X})^2 V(\hat{\beta}_1)$$

$$= \frac{\sigma^2_{y/x}}{n} + \frac{(X_k - \bar{X})^2}{\Sigma (X - \bar{X})^2} \sigma^2_{y/x}$$

ตัวค่าประมาณของ $V(\hat{Y}_k)$ **da**

$$V(\hat{Y}_k) = S^2_{y/x} \left[\frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma (X - \bar{X})^2} \right]$$

ความคลาดเคลื่อนมาตรฐานของ $\hat{Y}_k$ คือ

$$\sqrt{V(\hat{Y}_k)} = S_{y/x} \left[\frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma (X - \bar{X})^2} \right]^{\frac{1}{2}}$$

ค่านี้จะมีค่าน้อยที่สุดเมื่อ $X_k = \bar{X}$ และจะเพิ่มขึ้นเมื่อไรที่ค่า X_k เคลื่อนออกจาก $\bar{X}$ ไม่ว่าในทิศทางใดทางหนึ่ง ดังนั้นความคาดหวังในการทำนายของเราที่ดีที่สุดก็คือพยายามทำนายค่ากลาง ๆ ของช่วงค่าสังเกตของ X

ตัวอย่าง เราใช้โจทย์ของตัวอย่างก่อนในการหา $V(\hat{Y}_k)$

$$n = 12, \Sigma (X - \bar{X})^2 = 476,917, \; S^2_{y/x} = 0.2065 : \bar{X} = 124.583$$

$$V(\hat{Y}_k) = S^2_{y/x} \left[\frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma (X - \bar{X})^2} \right]$$

$$= 0.2065 \left[\frac{1}{12} + \frac{(X_k - 124.583)^2}{476.917} \right]$$

หาก $X_k = \bar{X}$

$$V(\hat{Y}_k) = 0.2065 \left(\frac{1}{12}\right) = 0.0172$$

$$\sqrt{V(\hat{Y}_k)} = \sqrt{0.0172} = 0.131$$

ถ้าหาก $X_k = 120$

$$V(\hat{Y}_k) = 0.2065 \left[\frac{1}{12} + \frac{(120 - 124.583)^2}{476.917}\right]$$

$$= 0.1839$$

$$\sqrt{V(\hat{Y}_k)} = 0.429$$

ต้องการคำนวณหาช่วงความเชื่อมั่นที่ 90% สำหรับเกรดเฉลี่ยที่คาดหวังโดยกำหนดคะแนนเชาว์ปัญญาเท่ากับ 120 คือ ($t_{.05\,;\,10} = 1.812$)

$$\begin{aligned}\hat{Y}_k &= 5.5021 + 0.0657\,(120) \\ &= -5.5021 + 7.884 \\ &= 2.3819\end{aligned}$$

$$\begin{aligned}E\,(Y/x = 120) &= \hat{Y}_k \pm t_{.05\,:\,10} S_{y/x} \sqrt{\frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma\,(X - \bar{X})^2}} \\ &= 2.38 \pm (1.812)\,(.454) \sqrt{\frac{1}{12} + \frac{(120 - 124.583)^2}{476.9167}} \\ &= 2.38 \pm (1.812)\,(.1384) \\ &= 2.38\ 4\ .251\end{aligned}$$

นั่นคือ $2.129 < E\,(y/x = 120) < 2.831$

ความแปรปรวนและความคลาดเคลื่อนมาตรฐานที่ได้กล่าวข้างต้นใช้สำหรับหาค่าคาดหวังของ Y กำหนด X_k แต่เนื่องจากว่าค่าสังเกตจริงของ Y แปรรวมค่าคาดหวังจริงด้วย

ความแปรปรวน $\sigma^2_{y/x}$ ซึ่งมีความอิสระกับ $V(\hat{Y})$ ค่าที่ถูกทำนายของค่าสังเกต แต่ละค่าก็ยังกำหนดได้โดย $\hat{Y}$ แต่จะมีความแปรปรวน

$$\sigma^2_{y/x} + V(\hat{Y}_k) \quad = \quad \sigma^2_{y/x} + \frac{\sigma^2_{y/x}}{n} + \frac{(X_k - \bar{X})^2}{\Sigma(X - \bar{X})^2}\sigma^2_{y/x}$$

$$V(Y_p) \quad = \quad \sigma^2_{y/x}\left[1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma(X - \bar{X})^2}\right]$$

ใช้ $S^2_{y/x}$ แทนค่า $\sigma^2_{y/x}$ เราได้

$$V(Y_p) \quad = \quad S^2_{y/x}\left[1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{X(X \quad \bar{X})^2}\right]$$

ค่าความเชื่อมั่นก็สามารถหาได้ด้วยวิธีเดียวกันกับที่ได้กล่าวมาก่อน นั่นคือ เราสามารถคำนวณ 90% ช่วงความเชื่อมั่น สำหรับค่าสังเกตใหม่ซึ่งจะขึ้นอยู่กับตัวค่าประมาณของค่าความแปรปรวนใหม่นี้

$$Y_p = \hat{Y}_k \pm t_{.05\,:\,n-2}\left(1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma(X - \bar{X})^2}\right)^{\frac{1}{2}} S_{y/x}$$

เพราะฉะนั้น 90% ช่วงความเชื่อมั่นของการทำนายเกรดเฉลี่ยของนักศึกษาโดยกำหนดคะแนนเชาว์ปัญญาเท่ากับ 120 เป็น

$$Y_p = \hat{Y}_k \pm t_{.05\,;\,n-2}\, S_{y/x}\sqrt{1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$$

$$= 2.38 \pm (1.812)\,(.454)\sqrt{1 + \frac{1}{12} + \frac{(120 \quad 124.583)^2}{476.9167}}$$

$$=: 2.38 \pm (1.812)\,(.454)\,(1.062)$$

$$= 2.38 \pm .8735$$

$$1.5065 < Y_p < 3.2535$$

สหสัมพันธ์

สหสัมพันธ์เป็นการแสดงถึงองศาความสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือมากกว่าว่ามีความสัมพันธ์มากน้อยแค่ไหน หากความสัมพันธ์นั้นมีเพียง 2 ตัวแปรก็เป็นสหสัมพันธ์อย่างง่าย ถ้าหากว่าความสัมพันธ์นั้นมีตัวแปรตั้งแต่สามตัวขึ้นไปก็เป็นสหสัมพันธ์เชิงซ้อน ในหัวข้อนี้จะกล่าวแต่สหสัมพันธ์อย่างง่ายและเป็นเชิงเส้นเท่านั้น นักสถิติใช้สัญลักษณ์ r แสดงค่าสัมประสิทธิสหสัมพันธ์อย่างง่ายของตัวอย่างและ ρ (rho) เป็นค่าสัมประสิทธิสหสัมพันธ์อย่างง่ายของประชากร

ในการศึกษาสหสัมพันธ์อย่างง่าย เราใช้ค่า r อนุมานค่า ρ เพราะในทางปฏิบัติเราไม่สามารถเอาค่าสังเกตทั้งหมดจากประชากรได้ r จึงเป็นค่าแสดงองศาแห่งความสัมพันธ์ระหว่างข้อมูล 2 ชุด r จะมีค่าอยู่ระหว่าง 1 กับ +1 $(-1 \leqslant r \leqslant +1)$ หาก r มีค่า $= +1$ แสดงว่าข้อมูลทั้งสองมีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางเดียวกัน นั่นคือ ข้อมูล X เปลี่ยนไป ข้อมูล Y จะเปลี่ยนตามไปในทางเดียวกันกล่าวคือ X เพิ่ม Y เพิ่ม X ลด Y ก็ลดตาม แต่ถ้า r มีค่าเท่ากับ -1 ก็แสดงว่าข้อมูล X กับ Y มีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางตรงข้าม นั่นคือ ข้อมูล X เพิ่มขึ้นจะมีผลให้อีกข้อมูล Y ลดลง หรือข้อมูล X ลดลงจะมีผลให้ข้อมูล Y เพิ่มขึ้น นอกจากนี้ค่า r ยังใช้วัด goodness of fit ของเส้นกำลังสองน้อยที่สุดหรือเส้นถดถอยอีกด้วย กล่าวคือเราสามารถใช้ข้อมูล X ไปทำนายค่า Y ในเส้นถดถอยที่คำนวณได้ ได้ถูกต้องร้อยเปอร์เซ็นต์ หากเราหาค่า $r = \pm 1$ หาก r มีค่าเข้าใกล้ศูนย์ก็แสดงว่าข้อมูล X กับข้อมูล Y ไม่มีความสัมพันธ์กันเลย (หรือน้อยมาก) หรือ fit is poor

วิธีการวิเคราะห์ค่าสัมประสิทธิสหสัมพันธ์อย่างง่ายที่นิยมทำมีดังนี้

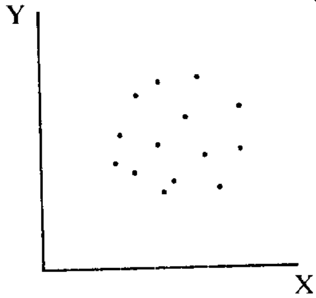
วิธีที่ 1 พิจารณาลักษณะของการกระจายของข้อมูลจากแผนภาพกระจาย

วิธีที่ 2 พิจารณาลักษณะของเส้นถดถอยโดยพิจารณาความชันของเส้นและทิศทางของเส้น

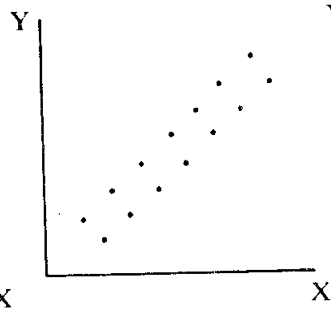
วิธีที่ 3 โดยใช้สูตร

วิธีที่ 1 พิจารณาลักษณะของการกระจายของข้อมูล

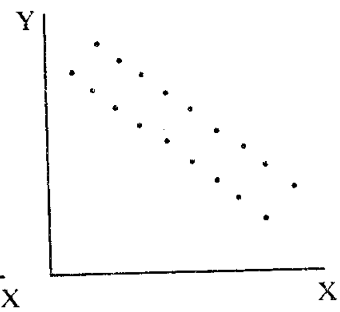
วิธีการนี้เป็นวิธีการอ่านความสัมพันธ์ของข้อมูล 2 ชุด อย่างง่าย ๆ โดยการนำค่าข้อมูลทั้งสองมาพลอตกราฟ หากจุดของข้อมูลในกราฟมีการกระจายกันมากหรือจับกันมีลักษณะเป็นกลุ่ม ก็แสดงว่า X และ Y มีองศาแห่ง ความสัมพันธ์น้อยหรือไม่มีเลย ดังรูป (ก)



สหสัมพันธ์น้อยหรือศูนย์
(ก)



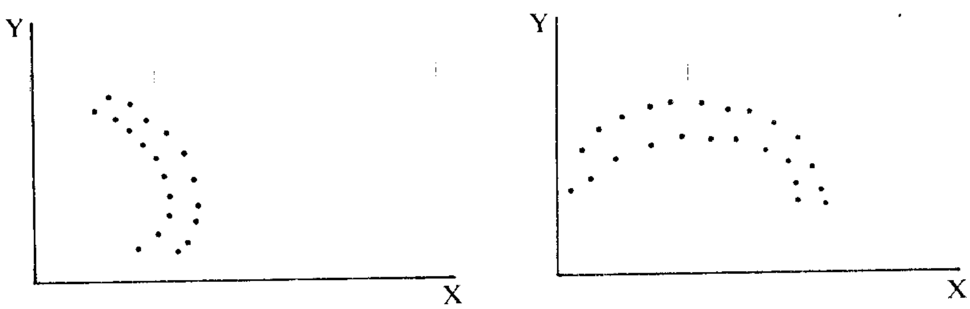
สหสัมพันธ์มากทางบวก
(ข)



สหสัมพันธ์มากทางลบ
(ค)

แต่ถ้าจุดของข้อมูลกระจายเป็นแนวเดียวกันก็แสดงว่าข้อมูล 2 ชุดนั้นมีสหสัมพันธ์สูง ดังรูป (ข) (ค)

ในบางครั้งข้อมูลมีความสัมพันธ์กันแต่เป็นไปในรูปของเส้นโค้ง เราเรียกข้อมูลทั้งสองมีสหสัมพันธ์แบบไม่เชิงเส้น

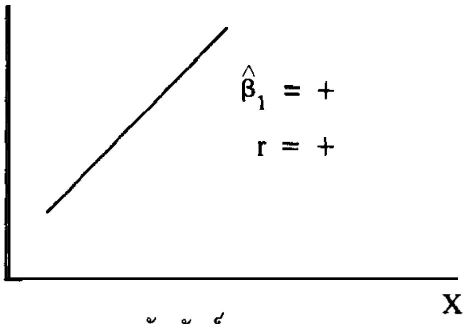


สหสัมพันธ์แบบไม่เชิงเส้น

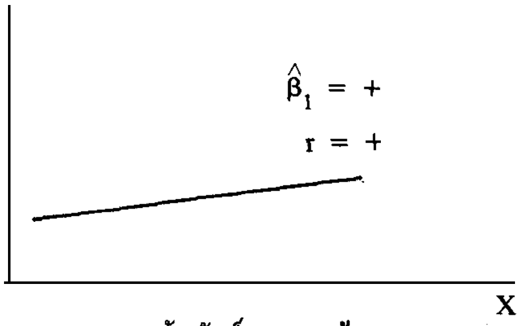
วิธีที่ 2 พิจารณาลักษณะของเส้นถดถอย

เราอ่านค่าสัมประสิทธิสหสัมพันธ์ได้จากการพิจารณาลักษณะความชันและทิศทางของเส้นถดถอยโดยประมาณอย่างหยาบ ๆ

ก. เส้นถดถอยที่มีความชันเป็นบวก ค่าสัมประสิทธิสหสัมพันธ์ก็จะมีค่าเป็นบวก ดังรูป

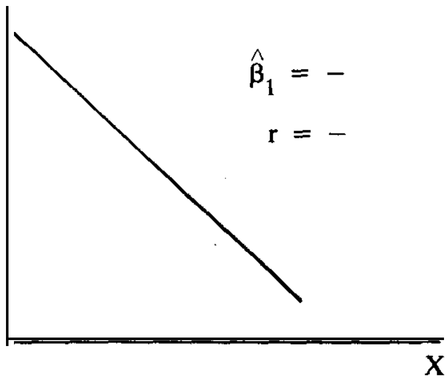


สหสัมพันธ์ทางบวกมาก

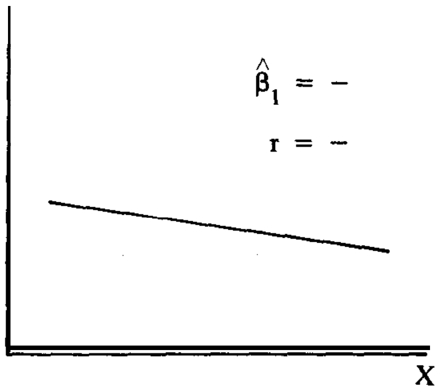


สหสัมพันธ์ทางบวกน้อย

ข. เส้นถดถอยที่มีความชันเป็นลบ ค่าสัมประสิทธิสหสัมพันธ์ก็จะมีค่าเป็นลบ ดังรูป

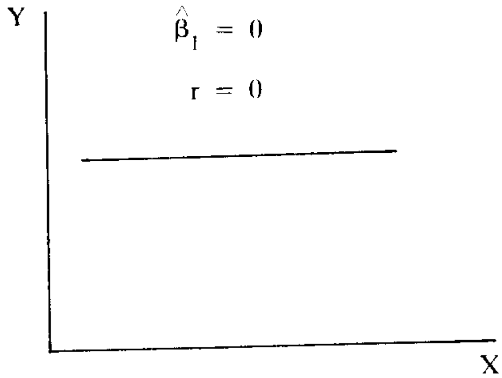


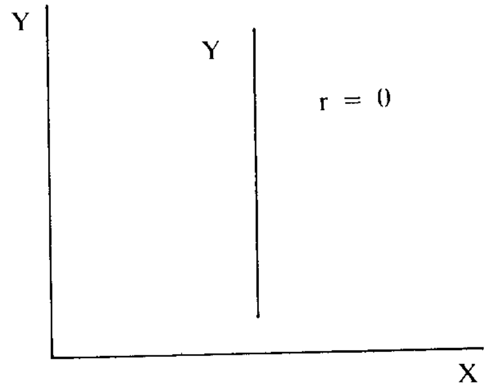
สหสัมพันธ์ทางลบมาก



สหสัมพันธ์ทางลบน้อย

ค. เส้นถดถอยที่มีความชันเป็นศูนย์ ค่าสัมประสิทธิสหสัมพันธ์จะมีค่าเป็นศูนย์หรือไม่มีความสัมพันธ์กันเลย





สหสัมพันธ์เป็นศูนย์

วิธีที่ 3 โดยสูตร

สูตรที่ใช้คำนวณหาค่าสัมประสิทธิสหสัมพันธ์ที่นิยมใช้กันมี

$$1. \quad r = \sqrt{\frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}}$$

สูตรนี้ใช้คำนวณสัมประสิทธิสหสัมพันธ์ของตัวแปรตั้งแต่สองตัวขึ้นไปทั้งที่มีสหสัมพันธ์เชิงเส้นและไม่เชิงเส้น ส่วนทิศทางของความสัมพันธ์นั้น จะต้องดูความชันของเส้นถดถอย

$$2. \quad r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{(\Sigma(X - \bar{X})^2)(\Sigma(Y - \bar{Y})^2)}}$$

หรือ

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

สูตรนี้ใช้คำนวณสัมประสิทธิสหสัมพันธ์ของตัวแปรเพียงสองตัวเท่านั้น ในกรณีตัวแปรเพียงสองตัว สูตรทั้งสองจะให้ค่า r เท่ากันหรือใกล้เคียงกัน

ตัวอย่าง คำนวณหาสัมประสิทธิสหสัมพันธ์ r ระหว่างคะแนนกับจำนวนเงินที่ขายได้ในตาราง ต่อไปนี้

X	Y	Y^2	X^2	XY	$\hat{Y}$	$(\hat{Y} - \bar{Y})^2$
48	312	97,344	2,304	14,976	307.04	5,218.62
32	164	26,896	1,024	5,248	196.96	1,431.87
40	280	78,400	1,600	11,200	252.00	295.84
34	196	38,416	1,156	6,664	210.72	579.85
30	200	40,000	900	6,000	183.20	2,662.56
50	288	82,944	2,500	14,400	320.80	7,396.–
26	146	21,316	676	3,796	155.68	6,259.97
50	361	130,321	2,500	18,050	320.80	7,396.–
22	149	22,201	484	3,278	128.16	11,372.09
43	252	63,504	1,849	10,836	272.64	1,431.87
375	2,348	601,342	14,993	94,448	2.348	39,405.89

ใช้สูตรที่ 1

$$r = \sqrt{\frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}}$$

ในที่นี้ $\Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - n\bar{Y}^2$

$$= 601{,}342 - 10\left(\frac{2348}{10}\right)^2$$

$$= 50{,}032$$

$$r = \sqrt{\frac{39{,}405.89}{50{,}032}}$$

$$= \sqrt{0.7876}$$

$$= 0.89$$

ใช้สูตรที่ 2

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

ในเมื่อ $n = 10; \Sigma X = 375, \Sigma Y = 2{,}348; \Sigma X^2 = 14{,}993$

$\Sigma Y^2 = 601{,}342, \Sigma XY = 94{,}448$

$$r = \frac{10\,(94{,}448) - (375)\,(2{,}348)}{\sqrt{|10\,(14{,}993) - (375)^2|\,|10\,(601{,}342) - (2{,}348)^2|}}$$

$$= \frac{944480 - 880500}{\sqrt{(9305)\,(500316)}} = \frac{63980}{68230.773} = .94$$

สัมประสิทธิของการตัดสินใจ (Coefficient of Determination)

ค่าสัมประสิทธิสหสัมพันธ์ r เมื่อยกกำลังสองก็จะได้ค่าสัมประสิทธิของการตัดสินใจ r^2 ซึ่งเป็นตัวใช้อธิบายการกระจายที่เกิดขึ้นทั้งหมดในค่าสังเกต Y ได้มากน้อยเพียงใดหรือกล่าวอีกนัยหนึ่ง r^2 เป็นตัวชี้ให้เห็นว่าเส้นถดถอยที่เราคำนวณหาได้นั้นเหมาะสมกับข้อมูลเพียงใด โดยทั่ว ๆ ไปเราอ่านค่า r^2 ในรูปเปอร์เซ็นต์ อย่างเช่น หากเราคำนวณค่า $100\,r^2 = 0$ หมายความว่าการกระจายในค่าสังเกต Y ที่เกิดขึ้นไม่ได้ถูกอธิบายได้โดยเส้นถดถอยเลย ซึ่งเป็นผลให้การกระจายที่ อธิบายไม่ได้ $\Sigma\,(Y - \hat{Y})^2$ เท่ากับการกระจายทั้งหมด $\Sigma\,(Y - \bar{Y})^2$ ถ้าหากว่า $100\ r^2 = 100\%$ ก็หมายความว่าเส้นถดถอยที่เราคำนวณได้นั้นอธิบายการกระจายทั้งหมดของค่าสังเกต Y ได้ ซึ่งมีผลให้การกระจายที่อธิบายไม่ได้ ไม่มีเลย $(\Sigma\,(Y - \hat{Y})^2 = 0)$

ความสัมพันธ์ระหว่าง r กับ β_1

กำหนดให้ข้อมูล 2 ชุด โดยชุดที่หนึ่ง X เป็นตัวแปรอิสระ อีกชุดหนึ่ง Y เป็นตัวแปรตาม สมการถดถอยก็เขียนได้เป็น

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

ในทำนองเดียวกัน หาก X เป็นตัวแปรตาม Y เป็นตัวแปรอิสระ สมการถดถอยก็เขียนได้เป็น (ข้อชุดเดียวกัน)

$$\hat{X} = \hat{\beta}'_0 + \hat{\beta}'_1\,Y$$

จากสมการถดถอยแรกเราคำนวณค่า $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2}$$

สมการถดถอยที่สอง

$$\hat{\beta}_1^1 = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (Y - \bar{Y})^2}$$

นิยาม สัมประสิทธิสหสัมพันธ์ r คือรูทสองของผลคูณ $\hat{\beta}_1$ กับ $\hat{\beta}_1^1$ ซึ่งเป็นความชันของสมการถดถอยสองเส้นนั่นคือ

$$r = \sqrt{\hat{\beta}_1 \hat{\beta}_1^1}$$

$$r = \sqrt{\left[\frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2}\right]\left[\frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (Y - \bar{Y})^2}\right]}$$

$$= \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2 \Sigma (Y - \bar{Y})^2}}$$

ตัวอย่าง

X	2	6	4	7
Y	−15	41	13	55

(ก) จงคำนวณหาสมการถดถอย Y on X

(ข) จงคำนวณหาสมการถดถอย X on Y

(ค) จงคำนวณหาสัมประสิทธิสหสัมพันธ์

X	Y	X^2	Y^2	XY
2	−15	4	225	−30
6	41	36	1681	246
4	13	16	169	52
7	55	49	3025	385
19	194	105	5100	653

(ก) สมการถดถอย $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

$$\hat{\beta}_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{4(653) - (19)(94)}{4(105) - (19)^2}$$

$$= \frac{826}{59} = 14$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{94}{4} - \frac{(14)(19)}{4}$$

$$= 23.5 - 66.5 = -43$$

$$\hat{Y} = -43 + 14X$$

(ข) สมการถดถอย $\hat{X} = \hat{\beta}_0^1 + \hat{\beta}_1^1 Y$

$$\hat{\beta}_1^1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{4(653) - (19)(94)}{4(5100) - (94)^2}$$

$$= \frac{826}{11564} = 0.07143$$

$$\hat{\beta}_0^1 = \bar{X} - \hat{\beta}_1^1 \bar{Y}$$

$$= \frac{(19)}{4} - \frac{(0.07143)\ (94)}{4}$$

$$= 4.75 - 1.6786$$

$$= 3.0714$$

$$\hat{X} = 3.0714 - 0.07143\ Y$$

$$r = \sqrt{\hat{\beta}_1 \hat{\beta}_1^1}$$

$$= \sqrt{(14)\ (0.07143)}$$

$$= 1$$

ในกรณีที่เราทราบค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรค่าทั้งสองและความชันของเส้นถดถอย เราสามารถคำนวณหาสัมประสิทธิของสหสัมพันธ์ได้นั่นคือ

$$r = \hat{\beta}_1 \frac{S_x}{S_y}$$

หรือ

$$\hat{\beta}_1 = r \frac{S_y}{S_x}$$

ในเมื่อสมการถดถอยเป็น

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

หากสมการถดถอยเป็น

$$\hat{X} = \hat{\beta}'_0 + \hat{\beta}'_1 Y$$

เราได้

$$r = \hat{\beta}_1^1 \frac{S_y}{S_x}$$

หรือ

$$\hat{\beta}_1^1 = r \frac{S_x}{S_y}$$

ตัวอย่าง มีนักศึกษา 100 คน สอบวิชาคณิตศาสตร์กับวิชาสถิติ ซึ่งมีความชัน $\hat{\beta}_1$ ของสมการเส้นถดถอยของวิชาสถิติ เมื่อกำหนดวิชาคณิตศาสตร์เท่ากับ 0.8 และมัชฌิมเลขคณิต และส่วนเบี่ยงเบนมาตรฐานมี ดังนี้

	คณิตศาสตร์	สถิติ
มัชฌิมเลขคณิต	50	55
ส่วนเบี่ยงเบนมาตรฐาน	5	6

$$S_x = 5 \; ; \; S_y = 6 \; ; \; \hat{\beta}_1 = 0.8$$

เนื่องจากว่า $r = \hat{\beta}_1 \frac{S_x}{S_Y}$ (เป็นสมการถดถอย Y on X)

$$= (0.8)\frac{5}{6}$$

$$= \frac{4}{6} = \frac{2}{3}$$

เพราะฉะนั้นสัมประสิทธิสหสัมพันธ์ $r = \frac{2}{3}$

หมายเหตุ ค่าสัมประสิทธิสหสัมพันธ์จะมีค่าคงที่เสมอจากข้อมูล 2 ชุด ไม่ว่าข้อมูล 2 ชุดหรือชุดหนึ่งบวกหรือลบด้วยตัวคงที่ทุก ๆ ข้อมูล หรือข้อมูล 2 ชุดหรือชุดหนึ่งคูณหรือหารด้วยตัวคงที่ทุก ๆ ข้อมูล และค่าสัมประสิทธิสหสัมพันธ์ของข้อมูลคู่หนึ่งจะมีค่าเท่ากับศูนย์ 2 คู่จะมีค่าเท่ากับ +1 หรือ −1 ขึ้นอยู่กับข้อมูลคู่นั้น มีทิศทางไปทางเดียวกันหรือทิศทางตรงกันข้ามกัน

ตัวอย่าง มีข้อมูล 2 ชุด X และ Y มีมัชฌิมเลขคณิต $\bar{X}, \bar{Y}$ ค่าสัมประสิทธิสหสัมพันธ์ r_{xy} คำนวณได้จากสูตร

$$r_{xy} = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2 \, \Sigma (Y - \bar{Y})^2}}$$

หากข้อมูล 2 ชุด X และ Y บวกหรือลบด้วยตัวคงที่ C นั่นคือ $Z = X \pm C$; $U = Y \pm C$ เพราะฉะนั้น มัชฌิมเลขคณิตของ Z และ U คือ $\bar{Z} = \bar{X} \pm C$; $\bar{U} = \bar{Y} \pm C$

ค่าสัมประสิทธิสหสัมพันธ์ของ Z กับ U คือ

$$r_{zu} = \frac{\Sigma (Z - \bar{Z})(U - \bar{U})}{\sqrt{\Sigma (Z - \bar{Z})^2 \, \Sigma (U - \bar{U})^2}}$$

$$= \frac{\Sigma(X \pm C - (\bar{X} \pm C))(Y \pm C - (\bar{Y} \pm C))}{\sqrt{\Sigma(X \pm C - (\bar{X} \pm C))^2 \Sigma(Y \pm C - (\bar{Y} \pm C))^2}}$$

$$= \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

$$= r_{xy}$$

ในกรณี X และ Y คูณด้วยตัวคงที่ C นั่นคือ $Z = XC$; Y มัชฌิมเลขคณิต Z และ Y คือ $\bar{X}C$ และ $\bar{Y}$

$$r_{zy} = \frac{\Sigma(Z - \bar{Z})(Y - \bar{Y})}{\sqrt{\Sigma(Z - \bar{Z})^2 \Sigma(Y - \bar{Y})^2}}$$

$$r_{zy} = \frac{\Sigma(XC - \bar{X}C)(Y - \bar{Y})}{\sqrt{\Sigma(XC - \bar{X}C)^2 \Sigma(Y - \bar{Y})^2}}$$

$$= \frac{C\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{C^2\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

$$= \frac{C\,\Sigma(X - \bar{X})(Y - \bar{Y})}{C\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

$$= r_{xy}$$

ในทำนองเดียวกันหากข้อมูล 2 ชุดจะหารด้วยตัวคงที่ก็จะได้ค่า r_{xy} เหมือนเดิมเช่นเดียวกัน

การใช้สมการถดถอยไปประมาณค่าหรือทำนายค่านั้น กระทำได้เพียงบางค่าหรือบางช่วงเท่านั้น นอกเหนือจากนั้นจะให้ค่าผิดความเป็นจริงหรือเป็นไปไม่ได้ดังตัวอย่าง ระหว่างราคารถที่ใช้แล้วกับอายุของรถยนต์ยี่ห้อหนึ่งจากนายหน้าขายรถ

X อายุ (ปี)	1	2	2	3	4	5	6	8	9	10
Y ราคา ($ 1000)	3.8	3.4	3.3	3.0	2.5	2.0	2.2	1.0	1.0	0.8

เนื่องจากว่าอายุของรถเป็นตัวกำหนดราคารถยนต์ สมการถดถอยก็เขียนได้

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

อายุ X	ราคารถยนต์ Y	XY	X^2
1	3.8	3.8	1
2	3.4	6.8	4
3	3.3	6.6	4
4	3.0	9.0	9
5	2.5	10.0	16
6	2.0	10.0	10
7	2.2	13.2	12
8	1.0	8.0	16
9	1.0	9.0	18
10	0.8	8.0	20
รวม 50	23.0	84.4	340
$\bar{X}$ 5	$\bar{Y} = 2.3$		

แทนค่า $\Sigma X = 50; \Sigma Y = 23.0$; $\Sigma XY = 84.4$; $\Sigma X^2 = 340$
ลงในสูตร

$$\hat{\beta}_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{10\,(84.4) - (50)\,(23)}{10\,(340) - (50)^2} = \frac{-306}{900}$$

$$= -0.34$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$= 2.3 - (-.34)\,5 = 40$$

สมการถดถอย คือ

$$\hat{Y} = 4.0 - .34\,X$$

เมื่อไร $X = 0$ จุดตัดที่แกน Y คือ 4.0 (4000) แต่เนื่องจากราคารถยนต์ลดลงอย่างรวดเร็วสำหรับปีแรกและเนื่องจากว่าไม่มีรถใหม่ในตัวอย่าง 4.0 (4000) ไม่อาจเป็นตัวประมาณค่าที่ดีสำหรับราคาเฉลี่ยจริงเมื่อไร $X = 0$ เส้นถดถอยจะให้ค่าประมาณอย่างสมเหตุสมผลสำหรับค่า X จากหนึ่งถึงสิบปี แต่จะให้ค่าไม่เป็นจริงสำหรับ $X < 1$ หรือ $X > 10$ ปี

คำและประโยคที่ควรจำ

การถดถอย	สหสัมพันธ์
สมการถดถอย	เกณฑ์กำลังสองน้อยที่สุด
ตัวแปรอิสระ	ตัวแปรตาม
การถดถอย Y on X	การถดถอย X on Y
จุดตัดแกน Y	ความชันของเส้น
ค่าคลาดเคลื่อนของตัวประมาณค่า	แผนภาพกระจาย
สัมประสิทธิ์สหสัมพันธ์	ค่าลบของ r
การทำนาย	$r = \pm 1$
ค่าบวกของ r	การถดถอยเชิงซ้อน
การถดถอยเชิงเส้นอย่างง่าย	ช่วงความเชื่อมั่นของ-
ส่วนเบี่ยงเบนมาตรฐานของ $\hat{\beta}_1$	ค่าคาดหวังเฉลี่ย
สัมประสิทธิ์ของการตัดสินใจ	
ช่วงความเชื่อมั่นของค่าทำนาย	

เติมคำลงในช่องว่าง บทที่ 12

จงเติมคำลงในช่องว่างให้สมบูรณ์ โดยคำตอบที่ถูกต้องอยู่ด้านขวามือ ใช้ไม้บรรทัดปิดคำตอบสำหรับคำถามซึ่งท่านยังไม่ตอบ

แนวทางศึกษา การถดถอย และกระบวนการกำลังสองน้อยที่สุดค่อนข้างยากในการตีความหมาย r (สัมประสิทธิ์ของเปียสันของสหสัมพันธ์เชิงเส้น) และการทดสอบการถดถอยเชิงเส้นอย่างมีความหมาย $Ha : \beta_1 \neq 0$ หัวข้อต่อไปนี้จะเกี่ยวกับเรื่องเหล่านี้โดยตรง

การดำเนินการวิเคราะห์กำลังสองน้อยที่สุดเกี่ยวกับกลุ่มของข้อมูล ตัวอย่าง เมื่อทราบจุดตัดแกน Y เป็นศูนย์ นั่นคือ β_0 = (1).........., เพื่อว่า $Y = 0 + \beta_1 X + \epsilon$ ให้กำลังสองน้อยที่สุดในรูป $\Sigma(\epsilon_i^2) = \Sigma[Y - (0 + (2)..........)]^2$ นี้ให้สมการปกติหนึ่ง, $\Sigma(XY) - \hat{\beta}_1 \Sigma(X^2) = 0$ หรือ $\hat{\beta}_1$ = (3)........

(1) 0

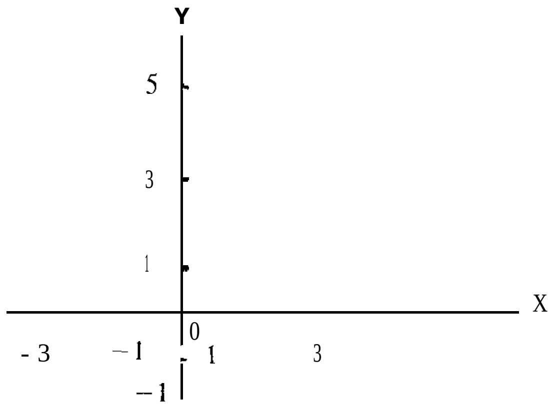
(2) $\hat{\beta}_1 X$

(3) $\Sigma(xy) / \Sigma x^2$

สำหรับค่าของตัวอย่าง (X, Y) = (−1, −.4), (1, .5), (2, .9), (2.5, 1.3), (3, 1.6), (4.1.8), (5, 2.2) นี้ให้ $\Sigma(xy) = 29.15$ และ $\Sigma(x^2) = 62.25$ ดังนั้น $\hat{\beta}_1$ = (4)..........พลอตจุดของข้อมูลและเส้นถดถอยกำลังสองน้อยที่สุด (5)..........

(4) .47

(5) ดูคำตอบหัวข้อสำหรับพลอตเสร็จเรียบร้อยแล้ว



โดยการสังเกต การปรับด้วยการมองดู (6)..........(ดี, เลว, ไม่สามารถบอกได้)	(6) ดี
การตีความสัมประสิทธิ์สหสัมพันธ์ที่ดีที่สุดโดยการเปรียบเทียบสองสัมประสิทธิ์หรือมากกว่า จากการศึกษาเหมือนกัน การขจัดความแตกต่างใน (7)..........ของค่าสำหรับ r^2 มากกว่าเปรียบเทียบ r เทอม r^2 มีความหมายทางปฏิบัติด้วย การบรรยาย (8)........ **ของความผันแปรทั้งหมดที่อธิบายด้วยเส้นถดถอย ดังนั้น ถ้าหากว่าการศึกษาความสัมพันธ์สองชนิดให้** $r_{1y} = -.8$ **กับ** $r_{2y} = +.75$ **ตัวแปร** (9)..........(X_1,X_2) **มีความสัมพันธ์มากกว่า** (10) ความสัมพันธ์.......... กับ Y เพราะว่าการมีสัมประสิทธิ์ของ (11).......... มากกว่า	(7) เครื่องหมายบวกหรือลบ (8) สัดส่วน (9) X_1 **(10) เชิงเส้น** **(11)** การตัดสินใจ
ต่อไปนี้เป็นตัวอย่างสำหรับการทดสอบว่าการถดถอยเชิงเส้นมีความหมายมากหรือไม่ บริษัทขายกระดาษบริษัทหนึ่ง พบว่า บริษัทสามารถประมาณค่าการขายของของบริษัทได้อย่างสมเหตุสมผล Y (ล้านบาท) จากการถดถอย $\hat{Y} = 5 + .03X$ **ในเมื่อ X = รายได้**จำพวกเครื่องอุปโภคบริโภคหน่วยล้านบาท บริษัทคำนวณหา $\sqrt{V(\hat{\beta}_1)} = 0.15$ ตัวเลขเหล่านี้แสดงว่า ขาย (12)..........(เพิ่มขึ้น, ไม่เปลี่ยน, ลดลง) ประมาณ (13)..........บาท สำหรับแต่ละ (14) ....ล้านบาท เพิ่มขึ้นของรายได้จำพวกเครื่องอุปโภคบริโภค X ถ้าหากว่ารายได้จำพวกเครื่องอุปโภคบริโภคเป็น 200 ล้านบาท การขายของบริษัทที่ควรสัมพันธ์กันประมาณ (15)..........ล้านบาท เราทราบว่า ข้อมูลที่ขายได้ใช้สร้างสมการข้างต้นมาจากข้อมูล 36 จุด เราสามารถทดสอบความสัมพันธ์เชิงเส้นอย่างมีความหมาย นั่นคือ No : (16)..........เทียบกับ Ha : (17)..........สำหรับ α = .10 ตัวสถิติที่ใช้ทดสอบt (หรือ Z ก็เหมือนกันเพราะองศาแห่งอิสระ df = n − (18).......... = (19)..........มากกว่า 30) มีค่า (20).......... แล้วค่าที่คำนวณได้คือ $(\hat{\beta}_1 - 0) / \sqrt{V(\hat{\beta}_1)}$ = (21).......... ดังนั้น การสรุปของเราคือ (22)..........(ปฏิเสธ, ไม่ปฏิเสธ) ความชันเป็นศูนย์ นี้แสดงว่าการทดสอบนี้และข้อมูลนี้ ตัวแบบเชิงเส้นที่ได้สมมตินี้ (23)..........(มี, ไม่มี) ความหมายมาก อย่างไรก็ตาม	**(12)** เพิ่มขึ้น **(13) 30,000** (14) 1 **(15) 11** (16) $\beta_1 = 0$ **(17)** $\beta_1 \neq 0$ (18) 2 **(19) 34** **(20) 1.64** **(21) 2.0** **(22)** ปฏิเสธ **(23)** มี

เราไม่ควรละเลยความน่าจะเป็นไปได้สำหรับการพรรณนาที่ปรับปรุงให้ดีขึ้นในลำดับสูงขึ้น (higher--order) ตัวแบบ (24)..........หรือ (24) Polynomial exponential

ปัญหาถูกหรือผิด บทที่ 12

ข้อความต่อไปนี้ถูกหรือผิด ถ้าถูกให้ขีดเส้นใต้ที่ตัวอักษร T หน้าข้อ ถ้าผิดให้ขีดเส้นใต้ที่ตัวอักษร F หน้าข้อนั้น ๆ

T - F 1. ขั้นแรกของปัญหา การถดถอยเชิงเส้นต้องเขียนกราฟจุดของข้อมูล

T - F 2. วัตถุประสงค์ของการอนุมานทางสถิติ เส้นที่ได้จากการกะเอาดีกว่าเส้นกำลังสองน้อยที่สุด

T - F 3. ค่าสหสัมพันธ์อยู่ระหว่าง 0 กับ + 1

T - F 4. สัมประสิทธิ์สหสัมพันธ์ใกล้ 0 แสดงว่า การถดถอยเชิงเส้นมีค่าน้อยแต่บอกความสัมพันธ์ของ higher-order regression เล็กน้อย อย่างเช่น quadratic

T - F 5. $r_1 = -.96$ แสดงความสัมพันธ์เชิงเส้นมากกว่า $r_2^2 = .81$

T - F 6. สหสัมพันธ์เชิงเส้นสัมบูรณ์ + 1 หรือ − 1 ไม่ได้หมายความถึงต้นเหตุ

T - F 7. แนวโน้มแบบพาราโบล่าจะเหมาะกับเซทของข้อมูล อย่างน้อยเช่นเดียวกับแนวโน้มเชิงเส้น

T - F 8. การคำนวณอนุกรมเวลาต้องการตัวแปร X ต้องเป็นค่าเวลา

T - F 9. เมื่อไรการปรับข้อมูลของเวลาต่าง ๆ ให้เข้ากับแนวโน้มระยะยาว ความผันแปรที่เหลือจะสะท้อนถึงการเปลี่ยนแปลงวัฏจักร์ ฤดูกาล และไม่สม่ำเสมอ

T - F 10. ค่าของ $(y - \hat{y})$ จะเป็น +, −, +, −, +, −, +, − ฯลฯ ตามลำดับการเพิ่มขึ้นของค่า X แสดงว่าเป็นการกำหนดที่เลว

T - F 11. เมื่อไรมีใครคนหนึ่งถามขึ้นว่า "เงินปันผลเกี่ยวกับใบหุ้นมีความสัมพันธ์กับธนบัตรของเขาในการแลกเปลี่ยน" หรือ "เขากำลังเกี่ยวข้องอยู่กับสหสัมพันธ์"

T - F 12. ทั้งนาย ก. และ น.ส.มา ได้คะแนนเท่ากัน ในการสอบวิชาหนึ่ง มันเป็นการถูกต้องที่จะกล่าวว่า เกรดของเขาทั้งสองมีสหสัมพันธ์สูง

T - F 13. จุดในแผนภาพกระจายใช้แทนคะแนนที่เป็นคู่กัน (paired scores)

T - F 14. ถ้าหากว่าจุดในแผนภาพกระจายฟอร์มเป็นรูป ellipse ขยายจากมุมซ้ายมือข้างบนลงมาข้างล่างขวามือ จะมีสหสัมพันธ์เป็นลบ

T - F 15. สหสัมพันธ์ลบสูง หมายความว่า แต่ละค่าที่ได้คะแนน Z เหมือนกันโดยประมาณทั้งสองตัวแปรถึงแม้คะแนน Z จะมีเครื่องหมายตรงข้ามกัน

T - F 16. ขณะ Pearson r เข้าใกล้ศูนย์ ผลบวกของผลคูณของคะแนน Z ที่เข้าคู่กันเข้าใกล้ค่าสูงสุด

T – F 17. ถ้าหากว่า $\Sigma(X-\bar{X})(Y-\bar{Y})$ เท่ากับ $\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}$ สหสัมพันธ์เป็น 1.00

T – F 18. ถ้าหาก $\Sigma(X-\bar{X})(Y-\bar{Y})$ เป็นบวกมากที่สุด สหสัมพันธ์เป็น 1.00

T – F 19. Pearson r เป็นมาตรวัดที่เหมาะสมของสหสัมพันธ์ เมื่อตัวแปรมีความสัมพันธ์ไม่เป็นเชิงเส้น

T – F 20. Pearson r วัดความสัมพันธ์เชิงเส้นระหว่างตัวแปร

T – F 21. ความสำคัญของความสัมพันธ์แสดงได้ด้วยขนาดของสัมประสิทธิ์สหสัมพันธ์

T – F 22. ทิศทางของความสัมพันธ์แสดงได้ด้วยเครื่องหมายของสัมประสิทธิ์สหสัมพันธ์

T – F 23. ขนาดของ Pearson r ขึ้นอยู่กับช่วงของสองตัวแปร มากกว่าขึ้นอยู่กับจำนวนของข้อมูล

T – F 24. เพื่อจะทำนายค่าหนึ่งของตัวแปรหนึ่ง จากความรู้ของค่าหนึ่งของตัวแปรที่สองข่าวสารที่เกี่ยวกับสหสัมพันธ์ระหว่างสองตัวแปรเป็นสิ่งจำเป็น

T – F 25. ถ้าหากว่า $r = 0.00$ และนาย ก. ได้คะแนนวิชา X ต่ำกว่ามัชฌิมเลขคณิตสองเท่าของส่วนเบี่ยงเบนมาตรฐาน นาย ก. น่าจะได้คะแนนวิชา Y ต่ำกว่ามัชฌิมเลขคณิต

T – F 26. ถ้าหากว่า Y ใช้แทนรายได้ต่อปีของบุคคลหนึ่ง และถ้าหากว่าเขาจ่ายเงินออกสองสัปดาห์ต่อครั้ง สูตรความสัมพันธ์ระหว่างรายได้ต่อปีกับรายจ่ายสองสัปดาห์ต่อครั้ง เป็น $y = 26X$

T – F 27. สูตร $Y = a + bX$, b_y ใช้แทนความชันของเส้นที่สัมพันธ์ของค่า Y ต่อค่า X

T – F 28. การวิเคราะห์สหสัมพันธ์ เส้นตรงที่ปรับเข้ากับข้อมูลดีที่สุดคือ เส้นถดถอย

T – F 29. สูตร $b_y = r\frac{s_y}{s_x}$ เป็นที่ทราบว่า เป็นความชันของเส้นถดถอย Y on X

T – F 30. ถ้า $r = 1.00$, $\bar{Y} = 10$, $s_x = 20$, $s_y = 5$, $X = 52$, $\bar{X} = 80$ แล้ว $\hat{Y} = 3$

T – F 31. ในเหตุการณ์ของสหสัมพันธ์สมบูรณ์ ส่วนเบี่ยงเบนที่ได้ทำนายไว้จากมัชฌิมเลขคณิตตัวอย่างเป็นศูนย์

T – F 32. เมื่อไรสหสัมพันธ์สมบูรณ์ ส่วนเบี่ยงเบนที่ได้ทำนายไว้จากมัชฌิมเลขคณิตมีค่ามากที่สุด

T – F 33. เส้นถดถอยจะตั้งฉากกันเมื่อ $r = -1.00$

T – F 34. ถ้าหากว่า นายโอโมได้คะแนน Z ของวิชา ก. 0.75 และสหสัมพันธ์เป็นบวกพอประมาณ คะแนน Z ที่ทำนายของวิชา ข. น้อยกว่า 0.75 แต่มากกว่า 0.00

T – F 35. ถ้าหากว่าเส้นถดถอยตัดที่มัชฌิมเลขคณิตของ X และ Y , r = 0.00

T – F 36. ผลบวกทางพีชคณิตของ $(y - \hat{y})$ เท่ากับศูนย์

T – F 37. residual variances อาจนิยามได้เป็นผลบวกของส่วนเบี่ยงเบนกำลังสองของค่าสังเกต จากเส้นถดถอยหารด้วยองศาแห่งความอิสระ

T – F 38. เมื่อไร r = 0.00 ความคลาดเคลื่อนมาตรฐานของค่าประมาณของ Y เท่ากับ

T – F 39. เมื่อไร r = – 1.00 ความคลาดเคลื่อนมาตรฐานของค่าประมาณของ Y เท่ากับ $-s_y$

T – F 40. ถ้าหากว่า r = 0.06, 36% ของความผันแปรอธิบายได้ในรูปของจำนวนขนาดของสัมประสิทธิ์สหสัมพันธ์ของมัน

T – F 41. ถ้าหาก $\Sigma (y - \hat{y})^2$ มีค่าน้อยที่สุด, r = 0.00

T – F 42. เมื่อไรส่วนเบี่ยงเบนที่ทำนายมีค่ามากที่สุด ความผันแปรที่อธิบายได้เป็น 100%

T – F 43. เมื่อไร r = 1.00, $\Sigma (\hat{y} - \bar{y})^2 = 0.00$

T – F 44. ในเหตุการณ์ที่สหสัมพันธ์ระหว่างสองเหตุการณ์เป็น 1.00 เราอาจสรุปได้ว่าตัวแปรก่อนมีความสัมพันธ์กับตัวแปรที่สองอย่างมีเหตุผล

เฉลยปัญหาถูกหรือผิด บทที่ 12

1. T	2. F	3. F	4. T
5. T	6. T	7. T	8. F
9. T	10. F	11. T	12. F
13. T	14. T	15. T	16. F
17. T	18. T	19. F	20. T
21. T	22. T	23. T	24. T
25. F	26. T	27. T	28. T
29. T	30. T	31. F	32. T
33. F	34. T	35. F	36. T
37. T	38. T	39. F	40. T
41. F	42. T	43. F	44. F

คำถามบทที่ 12

จงเลือกคำตอบที่ถูกต้อง

1. กำหนดให้ $\Sigma x = 43$, $\bar{X} = 6.1$, $\Sigma Y = 41$, $\bar{Y} = 5.9$, $\Sigma x^2 = 28.8$ $\Sigma y^2 = 22.8$, $cxy = 15.2$

   จงหาสัมประสิทธิ์สหสัมพันธ์ (r)

   (1) $-.95$ (2) $+.95$ (3) $\pm .53$ (4) $-.59$ (5) $+.59$

2. จากข้อกำหนด ข้อ 1. จงหา $\hat{\beta}_1$

   (1) $-.59$ (2) $+.59$ (3) $+.53$ (4) $-.53$ (5) $\pm .95$

3. ค่าไหนของ r แสดงองศามากที่สุดของความสัมพันธ์

   (1) .87 (2) .62 (3) $-.79$ (4) $-.91$ (5) 0

4. ค่าไหนของ r แสดงความสัมพันธ์น้อยที่สุด

   (1) .08 (2) .12 (3) $-.85$ (4) $-.98$ (5) .50

5. ถ้าหากว่าฐานะของ X ไม่ได้ช่วยในการทำนายฐานะของ Y ดังนั้น r เป็น

   (1) 0 (2) -1.0 (3) น้อยกว่า -1.00

   (4) 1 (5) ไม่สามารถคำนวณหาได้

6. ในแผนภาพกระจาย ถ้าจุดหนึ่งของหลาย ๆ จุดไม่ได้ตกบนเส้นตรงของ best fit ต่อจุดทั้งหมด r ไม่สามารถเป็น.–

   (1) 0 (2) ± 1 (3) เป็นบวก

   (4) เป็นลบ (5) ไม่สามารถคำนวณได้

7. อาจคำนวณสัมประสิทธิ์สหสัมพันธ์ ถ้าเรามี .–

   (1) คู่หนึ่งของข้อมูลโดยเฉพาะ

   (2) เซทหนึ่งของข้อมูลโดยเฉพาะ

   (3) สองเซทของข้อมูลสำหรับกลุ่มที่เหมือนกันโดยเฉพาะ

   (4) เซทหนึ่งของข้อมูลสำหรับกลุ่มหนึ่งโดยเฉพาะกับอีกเซทหนึ่ง สำหรับกลุ่มอื่น ๆ โดยเฉพาะ

   (5) ถูกทั้งหมด

8. เราสามารถทำนาย Y จากความรู้ของ X ได้ดีอย่างไร เราคำนวณ r ได้ -1.16 จากความรู้เหล่านี้ เราทราบว่า

   (1) ค่ามากของ X ทำนายค่าน้อยของ Y

(2) ข้อมูลใน Y โดยทั่ว ๆ ไปต่ำ

(3) ส่วนเฉลี่ยของ X สูงกว่าส่วนเฉลี่ยของ Y

(4) มีความสัมพันธ์กันมาก

(5) เราได้ทำการคำนวณผิดพลาด

9. นักศึกษาห้าสิบคนได้ทำการสอบข้อสอบแบบถูกผิด 100 ข้อ ให้ X เป็นจำนวนปัญหาที่ได้ตอบถูกต้อง กับ Y เป็นจำนวนที่ได้ตอบผิด เราควรหวังว่า r เป็น.–

(1) + 1.00 (2) ศูนย์ (3) – 1.00

(4) ระหว่างศูนย์กับ – 1.00 (5) ระหว่างศูนย์กับ + 1.00

10. ในการศึกษาครั้งหนึ่งของผู้ขับรถ พบว่า เวลาของผลอันเกิดขึ้นมีสหสัมพันธ์ในทางบวกกับจำนวนของอุบัติเหตุนี้หมายความว่า.–

(1) เวลาของผลอันเกิดขึ้นช้าลง อุบัติเหตุมากขึ้น

(2) เวลาของผลอันเกิดขึ้นช้าลง อุบัติเหตุน้อยลง

(3) เวลาของผลอันเกิดขึ้นเร็วขึ้น อุบัติเหตุมากขึ้น

(4) เวลาของผลอันเกิดขึ้นช้าลง ความปลอดภัยของผู้ขับมากขึ้น

(5) ไม่มีข้อใดถูกต้อง

11. ถ้า r = – 1.00

(1) ใช้ X ทำนาย Y ไม่ได้

(2) ค่าของ X สามารถทำนายค่าของ Y โดยไม่มีความคลาดเคลื่อน

(3) ส่วนเฉลี่ยเลขคณิตของ X น้อยกว่าส่วนเฉลี่ยเลขคณิตของ Y

(4) ส่วนเบี่ยงเบนมาตรฐานของ X น้อยกว่าส่วนเบี่ยงเบนมาตรฐานของ Y

(5) ข้อ (3) และ ข้อ (4) ถูก

12. สูตรของข้อมูลเบี่ยงเบนสำหรับ r มีประโยชน์สำหรับ.–

(1) คำนวณค่า r เมื่อข้อมูลไม่ได้จัดเป็นกลุ่ม

(2) คำนวณค่า r เมื่อ n โต

(3) เข้าใจสภาพของ r

(4) ถูกทั้งหมด (5) ผิดทั้งหมด

13. เริ่มต้นการคำนวณ r จากข้อมูลดิบ

(1) ข้อมูล X ควรจัดตามลำดับขนาด

(2) ทั้งข้อมูล X และ Y ควรจัดตามลำดับขนาด

(3) คู่ของข้อมูล X และ Y อาจอยู่ในลำดับใด ๆ ยาวเท่าที่มีรายละเอียด

(4) กระบวนการใด ๆ ข้างต้นเป็นที่พอใจ

(5) กระบวนการข้อ (2) และข้อ (3) เป็นที่พอใจ

14. พิจารณาข้อมูลสองคู่เหล่านี้

X	Y
2	3
4	5

สำหรับข้อมูลเหล่านี้ ค่าของ ΣXY คือ

(1) 14 (2) 22 (3) 23 (4) 26 (5) 25

15. ระหว่างเด็กกลุ่มหนึ่ง สหสัมพันธ์คะแนนทดสอบในวิชาวิทยาศาสตร์ กับคะแนนทดสอบในวิชาภาษาอังกฤษเป็น + .45 จากการศึกษาพบว่า คะแนนทดสอบวิชาวิทยาศาสตร์แต่ละครั้งเพิ่มขึ้น 5 คะแนน คะแนนจึงจะถูกต้อง คำนวณค่า r ใหม่ เราหวังว่าค่า r จะ.-

(1) มากกว่า + .45 (2) น้อยกว่า + .45 (3) เปลี่ยนไป

(4) ไม่เปลี่ยน (5) ไม่มีข้อใดถูกต้อง

16. สหสัมพันธ์ระหว่างน้ำหนักวัตถุเป็นปอนด์ กับความสูงวัดเป็นฟุต เป็น + .68 ข้อใดควรจะเปลี่ยนค่า r

(1) ความสูงมีหน่วยเป็นเซ็นติเมตร

(2) น้ำหนักมีหน่วยเป็นกิโลกรัม

(3) ทั้งข้อ 1 และข้อ 2 เปลี่ยนจะมีผลต่อค่า r

(4) ทั้งข้อ (1) และข้อ (2) เปลี่ยนจะไม่มีผลต่อค่า r

(5) ถูกทั้งหมด

17. ก่อนการคำนวณหาสหสัมพันธ์ แต่ละคะแนน X บวกด้วย 10 และแต่ละคะแนน Y หารด้วย 2 นี้จะ.-

(1) ทำให้ r ไม่เปลี่ยน (2) เพิ่ม r (3) ลด r

(4) r เปลี่ยนไป (5) r เพิ่มขึ้น .5

18. คำนวณค่าของ r จากข้อมูลดิบข้อไหนที่จะเปลี่ยนค่าของ r

(1) แต่ละข้อมูล X คูณด้วย − 1

(2) แต่ละข้อมูล X ถูกสับเปลี่ยนกับตำแหน่งที่สมนัยกัน

(3) แต่ละข้อมูล X เป็นจำนวนที่ผิดมากกว่าจำนวนที่ถูก

(4) ทั้งสามข้อข้างต้นไม่เปลี่ยนค่าของ r

(5) ทั้งสามข้อข้างต้นเปลี่ยนค่าของ r

19. กล่าวได้ว่า เซทสองเซทของข้อมูลมีความสัมพันธ์เชิงเส้น หมายความว่า เส้นของ best fit คือ

(1) เส้นซึ่งอาจเป็นเส้นโค้ง ในทิศทางหนึ่งเท่านั้น

(2) เส้นซึ่งอาจเป็นเส้นโค้งในทิศทางใดทางหนึ่ง

(3) เส้นตรง

(4) ข้อใดข้อหนึ่งทั้งสามข้อ

(5) ไม่ใช่ข้อใดข้อหนึ่งทั้งสามข้อ

20. จุดหลายจุดในการแจกแจงความถี่ ตัวแปรทั้งสองยึดมั่นอยู่ในเส้นตรงของ best fit อย่างใกล้ชิดเมื่อ.–

(1) r เข้าใกล้ + 1.00

(2) r เข้าใกล้ – 1.00

(3) ข้อ (1) หรือข้อ (2) ข้อใดข้อหนึ่ง

(4) ไม่ใช่ทั้งข้อ (1) หรือข้อ (2)

(5) r เข้าใกล้ ± .90

21. การกระจายของการเลือกตัวอย่างสุ่มระหว่างค่าของ r มากที่สุดเมื่อ .–

(1) หน่วยของการวัดใหญ่ (2) s_x และ s_y ใหญ่ (3) n มีค่าใหญ่

(4) n มีค่าเล็ก (5) ความสัมพันธ์ไม่เป็นเชิงเส้น

22. ถ้า s_{yx} เป็นศูนย์ ดังนั้น

(1) s_y ต้องเป็นศูนย์ (2) r ต้องเป็นศูนย์ ถ้า $s_y > 0$

(3) r ต้องเป็นหนึ่ง ถ้า $s_y > 0$ (4) r ต้องเป็นลบ

(5) r ต้องเป็นบวก

23. ค่าไหนของ r ที่จะให้ค่าที่ถูกต้องที่สุดในการทำนาย

(1) + .78 (2) + .27 (3) – .37 (4) – .81 (5) + .50

24. อะไรที่จะทำให้สมการถดถอยใช้ทำนายได้ถูกต้องแน่นอนยิ่งขึ้น

(1) ตัวอย่างที่ใหญ่ (2) ความสัมพันธ์เชิงเส้น

(3) ตัวอย่างที่เล็ก (4) ข้อ (1) และข้อ (2) ถูก

(5) ข้อ (2) และข้อ (3)ถูก

25. ในการทำนาย Y จาก X กำหนดเส้นถดถอยเพื่อให้ผลต่างกำลังสองระหว่างจุดกับเส้นมีค่าน้อยที่สุด

(1) ในมิติของ X (2) ในมิติของ Y

(3) ในทิศทางตั้งฉากกับเส้น (4) ในมิติทั้งหมดที่ตอบข้างต้น

(5) ไม่มีข้อใดถูก

26. ใน Concept เส้นถดถอยมีความสัมพันธ์ใกล้ชิดที่สุดกับ

(1) $\bar{X}$ (2) M_e (3) M_o (4) s_x (5) Z

27. เทอมไหนไม่ได้ปรากฏในข้อมูลดิบของสมการถดถอย

(1) $\bar{X}$ (2) s_y (3) r

(4) ไม่มีข้อใดปรากฏในสมการข้อมูลดิบ

(5) ทุกข้อปรากฏในสมการข้อมูล

28. ถ้าสมการถดถอย $\hat{Y} = 31 - .75X$ เราทราบว่า.-

(1) $\bar{X}$ มากกว่า $\bar{Y}$ (2) s_y มากกว่า s_x (3) สหสัมพันธ์เป็นลบ

(4) คำนวณค่า r ไม่ถูกต้อง (5) ข้อ (1) และข้อ (2) ถูก

29. ถ้าสมการถดถอย $\hat{Y} = 23 + 1.25X$ เราทราบว่า.-

(1) $\bar{X}$ มากกว่า $\bar{Y}$ (2) s_y มากกว่า s_x (3) สหสัมพันธ์เป็นลบ

(4) คำนวณค่า r ไม่ถูกต้อง (5) Y เพิ่ม 1 หน่วย X เพิ่ม 2 หน่วย

30. ถ้าสมการถดถอย $\hat{Y} = -3 + .50X$ เราทราบว่า.-

(1) Y เพิ่ม 1 หน่วย สำหรับ X เพิ่มทุก ๆ 2 หน่วย

(2) Y เพิ่ม 2 หน่วย สำหรับ X เพิ่มทุก ๆ 1 หน่วย

(3) $\bar{Y}$ น้อยกว่า $\bar{X}$ 3 หน่วย

(4) สหสัมพันธ์เป็นลบ

(5) คำนวณค่า r ไม่ถูก

31. ถ้าสมการถดถอย $\hat{Y} = 5 - .50X$ เราพบว่า.-

(1) Y เพิ่มขึ้น 5 หน่วย สำหรับ X เพิ่มขึ้นทุก ๆ 1 หน่วย

(2) Y ลดลง 2 หน่วย สำหรับ X เพิ่มขึ้นทุก ๆ 1 หน่วย

(3) Y ลดลง ขณะ X ลดลง

(4) Y เพิ่มขึ้นขณะ X เพิ่มขึ้น

(5) ไม่มีข้อใดถูก

32. กำหนดให้

$\bar{Y} = 100$ $\bar{X} = 50$

$s_y = 20$ $s_x = 10$

ถ้า r = + 1.00 และ X เท่ากับ 40 ค่าทำนายสำหรับ Y คือ

(1) 120 (2) 100 (3) 90 (4) 0 (5) 80

33. จากโจทย์ข้อ 32. ถ้า r = -1.00 และ X เท่ากับ 40 ค่าทำนายสำหรับ Y คือ

(1) 120 (2) 100 (3) 90 (4) 80 (5) 0

34. จากโจทย์ข้อ 32. ถ้า r = 0 และ X เท่ากับ 40 ค่าทำนายสำหรับ Y คือ

(1) 120 (2) 100 (3) 80 (4) 0 (5) 90

35. จากโจทย์ข้อ 32.61 r = + .50 และ X เท่ากับ 40 ค่าทำนายสำหรับ Y คือ

(1) 120 (2) 110 (3) 90 (4) 80 (5) 0

36. จากโจทย์ข้อ 32. ถ้า r = – SO และ X เท่ากับ 40 ค่าทำนายสำหรับ Y คือ

(1) 120 (2) 110 (3) 90 (4) 80 (5) 0

37. เมื่อไรที่สหสัมพันธ์สมบูรณ์ (± 1) ทุก ๆ ค่าของข้อใดจะเป็นศูนย์

(1) $(y - \hat{y})$ (2) $(y - \bar{y})$ (3) $(\hat{y} - \bar{y})$

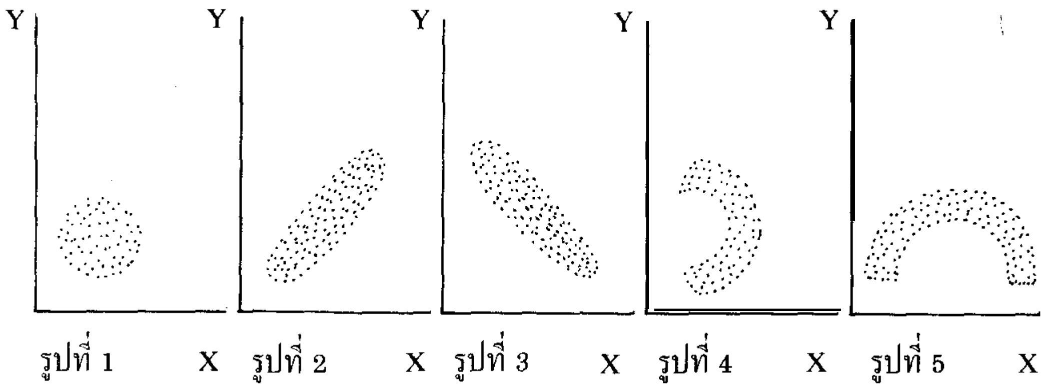
(4) $(\bar{y} - \bar{x})$ (5) ถูกทั้งหมด

38. $(y - \hat{y}) = (y - \bar{y})$ 6&l-

(1) r = 1.00 (2) r = 1.00 (3) r ± 1.00

(4) r ± SO (5) r = 0

39.



รูปที่ 1 รูปที่ 2 รูปที่ 3 รูปที่ 4 รูปที่ 5

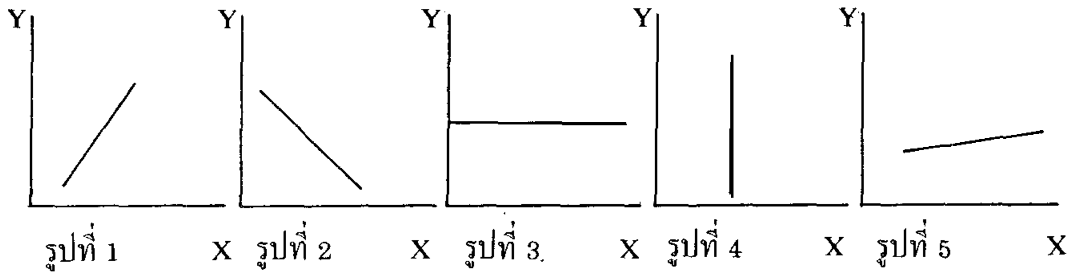
จงตอบคำถามว่ารูปใดที่ตัวแปรทั้งสองมีสหสัมพันธ์น้อยหรือสหสัมพันธ์เป็นศูนย์

(1) รูปที่ 1 (2) รูปที่ 2 (3) รูปที่ 3 (4) รูปที่ 4 (5) รูปที่ 5

40. จากโจทย์ข้อ 39. รูปใดที่ไม่มีสหสัมพันธ์เชิงเส้น

(1) รูปที่ 1,2 (2) รูปที่ 1,3 (3) รูปที่ 2,3 (4) รูปที่ 3,4 (5) รูปที่ 4,5

41. จงพิจารณาดูจากรูปต่อไปนี้โดยสมการเส้นถดถอยเขียนได้เป็น $Y = b_o + b_1X$



รูปใดที่ให้ค่า b_1 และ r เป็นค่าบวกที่มีค่าน้อย

(1) รูปที่ 1 (2) รูปที่ 2 (3) รูปที่ 3 (4) รูปที่ 4 (5) รูปที่ 5

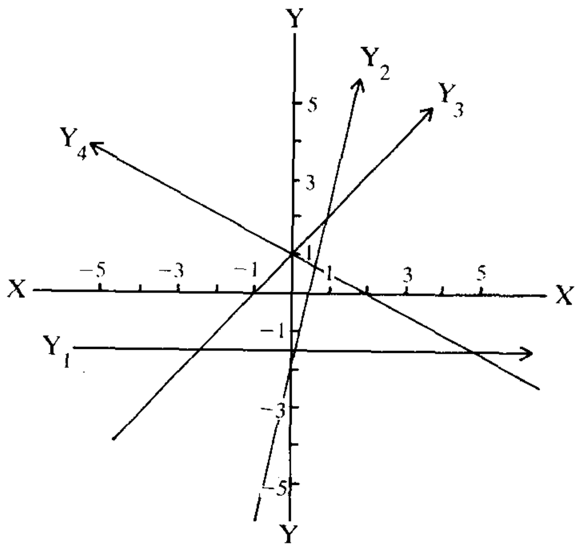
42. จากโจทย์ข้อ 41. รูปใดที่ค่า b_1 และ r เป็นศูนย์

(1) รูปที่ 1 (2) รูปที่ 2 (3) รูปที่ 3 (4) รูปที่ 4 (5) รูปที่ 5

43. จากโจทย์ข้อ 39. รูปใดที่มีค่า b_1 และ r เป็นลบ

(1) รูปที่ 1 (2) รูปที่ 2 (3) รูปที่ 3 (4) รูปที่ 4 (5) รูปที่ 5

44. กำหนดให้สมการเส้นถดถอยเขียนได้เป็น $Y_i = b_{ei} + b_{1i} X_i$; i = 1, 2, 3, 4 ใช้แทนเส้นทั้งสี่ดังรูปข้างล่าง



จงหาเส้นถดถอยของ Y, = b_{o1} + b,, X,

(1) Y_1 = -2 + 4x, (2) Y, = 1 + x, (3) Y, = − 1.5

(4) Y, = 1 + $.5X_1$ (5) Y, = 1.5 + x,

45. จากโจทย์ข้อ 44. จงหาเส้นถดถอยของ $Y_4 = b_{o4} + b_{14} X_4$

(1) $Y_4 = -1 + .5X_4$ (2) $Y_4 = -2 + 4X_4$ (3) $Y_4 = 1 + X_4$

(4) $Y_4 = 1 + .5X_4$ (5) $Y_4 = 1 - .5X_4$

เฉลยคำถาม

1. (5)	2. (3)
3. (4)	4. (1)
5. (1)	6. (2)
7. (3)	8. (5)
9. (3)	10. (1)
11. (2)	12. (3)
13. (3)	14. (4)
15. (4)	16. (4)
17. (1)	18. (5)
19. (3)	20. (3)
21. (5)	22. (3)
23. (4)	24. (4)
25. (2)	26. (1)
27. (5)	28. (3)
29. (2)	30. (1)
31. (5)	32. (5)
33. (1)	34. (2)
35. (3)	36. (2)
37. (1)	38. (5)
39. (1)	40. (5)
41. (5)	42. (3)
43. (2)	44. (3)
45. (5)	

นี้เป็น **WATFIN routine** สำหรับการประมาณค่ากำลังสองน้อยที่สุดของสัมประสิทธิความถดถอยเชิงเส้น input ที่ต้องการคือ N = ขนาดตัวอย่างและชุดข้อมูล N ยังเป็นค่าดัชนีสำหรับ X ( ) และ Y ( ) ด้วย output จะมี N, $\bar{X}$, $\bar{Y}$, s_x, s_y, r, b_o, b_1 และ s_{b1} รวมอยู่ด้วย ท่านจะต้องจัดหา JCL เอง

บรรทัด ข้อความ

```
REAL X (1Ø), Y (1Ø), SUMX, SUMY, CROSS, XSQ, YSQ,
XMEAN, YMEAN, 1SSQX, SSQY, XDEVN, YDEVN, CORR, RSQ,
BZERO, BONE, XDIFF, YDIFF, 2XYDIF, SYONX, VARB, SDEVB
INTEGER N,I
READ, N
SUMX = 0.0
SUMY = 0.0
CROSS = 0.0
XSQ = 0.0
YSQ = 0.0
DO1I = 1, N
READ, X (I), Y (I)
SUMX = SUMX + X (I)
SUMY = SUMY + Y (I)
CROSS = CROSS + (X (I) *Y (I))
XSQ = XSQ + (X (I) *X(I))
YSQ = YSQ + (Y (I) *Y (I))
1 CONTINUE'
XMEAN = SUMX/N
YMEAN = SUMY/N
XDIFF = (N*XSQ) - (SUMX * SUMX)
YDIFF = (N* YSQ) - (SUMY * SUMY)
XYDIF = (N* CROSS) - (SUMX*SUMY)
SSQX = XDIFF / (N*N - N)
```

```
XDEVN = SQRT (SSQX)
SSQY = YDIFF/(N*N - N)
YDEVN = SQRT (SSQY)
RSQ = (XYDIF*XYDIF) (XDIFF*YDIFF)
CORR = SQRT (RSQ)
BONE = XYDIF/XDIFF
BZERO = YMEAN - (BONE*XMEAN)
SYONX = (YDIFF - (BONE*XYDIF)) / (N*N - 2*N)
VARB = SYONX/XDIFF
SDEVB = SQRT (VARB)
PRINT, N, XMEAN, YMEAN, XDEVN, YDEVN, CORR, BZERO,
BONE, SDEVB
RETURN
END
```