

8 การถดถอยและสหสัมพันธ์ (REGRESSION AND CORRELATION)

กระบวนวิชา 208263 - สถิติเบื้องต้น
 โดย ผศ. ดร. สุคนธ์ ประสิทธิ์วัฒน์เสรี
 ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

1

เนื้อหา

- บทนำ
- การถดถอย
 - สมการถดถอยอย่างง่าย
 - สมการถดถอยพหุคูณ
 - การอนุมานการถดถอย
- สหสัมพันธ์
 - สัมประสิทธิ์สหสัมพันธ์
 - การอนุมานสัมประสิทธิ์สหสัมพันธ์

วัตถุประสงค์

- สามารถสร้างสมการถดถอย และทำการวิเคราะห์การถดถอยได้อย่างถูกต้องเหมาะสม
- สามารถคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ และทำการวิเคราะห์สหสัมพันธ์ได้อย่างถูกต้องเหมาะสม

2

บทนำ (Introduction)

- การศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว (Bivariate data)
 - กรณีตัวแปรคุณภาพ
 - ใช้การวิเคราะห์ไคสแควร์
 - เช่น การศึกษาความสัมพันธ์ระหว่างเพศกับการสูบบุหรี่
 - กรณีตัวแปรปริมาณ
 - ใช้การวิเคราะห์การถดถอยและสหสัมพันธ์
 - เช่น การศึกษาความสัมพันธ์ระหว่าง ความสูง + น้ำหนัก ยอดขาย + ราคาสินค้า

3

การถดถอยและสหสัมพันธ์

เป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัว หรือลักษณะที่สนใจศึกษา 2 ลักษณะ

- หากต้องการสร้างสมการคณิตศาสตร์อธิบายความสัมพันธ์ที่ตัวแปรหนึ่ง **ส่งผลต่อ** ค่าของอีกตัวแปรหนึ่ง จะใช้ **การวิเคราะห์การถดถอย (Regression analysis)**
- ศึกษา **ขนาดความสัมพันธ์** ระหว่างตัวแปรทั้ง 2 ตัว จะใช้ **การวิเคราะห์สหสัมพันธ์ (Correlation analysis)**

4

การถดถอย

- **สมการถดถอยอย่างง่าย**
- **การอนุมานการถดถอย**

5

การถดถอยอย่างง่าย - Regression

- เป็นการสร้างแบบจำลองหรือตัวแบบเพื่ออธิบายความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยทำการสร้าง **สมการถดถอย - regression equation**

$$Y = A + BX + \epsilon$$
- มักแทน ตัวแปรอิสระ -Independent Variable -> X
 ตัวแปรตาม-Dependent Variable -> Y
- โดย A = ค่า intercept
 B = ค่า slope หรือค่าสปส.การถดถอย
 ϵ = ความคลาดเคลื่อน

6

ตัวแปรอิสระกับตัวแปรตาม

- คะแนนวิชาสถิติของน.ศ.ขึ้นกับคะแนนวิชาคณิตศาสตร์
 - ตัวแปรตาม
 - ตัวแปรอิสระ
- เงินที่ใช้ในการประชาสัมพันธ์ส่งผลต่อยอดขายสินค้า
 - ตัวแปรอิสระ
 - ตัวแปรตาม

7

ลักษณะข้อมูลในการวิเคราะห์การถดถอย

ตัวแปรอิสระ (X)	ตัวแปรตาม (Y)
X_1	Y_1
X_2	Y_2
X_3	Y_3
...	...
X_n	Y_n

ตัวแปรอิสระ (X)	X_1	X_2	X_3	...	X_n
ตัวแปรตาม (Y)	Y_1	Y_2	Y_3	...	Y_n

8

การหารูปแบบความสัมพันธ์ระหว่าง X กับ Y

นำข้อมูลของตัวแปรทั้ง 2 มาเขียนกราฟแผนภาพการกระจาย (Scatter diagram) อาจพิจารณาเห็นได้คร่าว ๆ ถึงรูปแบบความสัมพันธ์ที่เกิดขึ้น

เช่น

X	Y
5	6
2.5	2
7.5	10

9

ความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรอิสระ- X กับตัวแปรตาม- Y

- X เพิ่ม \rightarrow Y เพิ่ม
- X ลด \rightarrow Y ลด
- X เพิ่ม \rightarrow Y ลด
- X ลด \rightarrow Y เพิ่ม

ความสัมพันธ์ทางเดียวกัน หรือเชิงบวก

ความสัมพันธ์ทางตรงข้าม หรือเชิงลบ

\rightarrow ตัวแปรตามขึ้นอยู่กับตัวแปรอิสระ

10

รูปแบบความสัมพันธ์ระหว่างตัวแปร X กับ Y

แผนภาพการกระจาย (Scatter Diagram)

11

การสร้างสมการถดถอย

สมการเส้นตรงทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่าง X กับ Y

$$Y_i = A + BX_i + \varepsilon_i$$

เมื่อ A = ค่าเริ่มต้นของ Y ที่ $X = 0$,
 B = อัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนไป 1 หน่วย
 $= \Delta Y / \Delta X$
 ε = ความคลาดเคลื่อน

12

สมการถดถอยระหว่าง X กับ Y

สมการถดถอยในประชากร $Y_i = A + BX_i + \varepsilon_i$

↑
อนุมาน

เมื่อ $A = \text{intercept (parameter)}$,
 $B = \text{coefficient regression (parameter)}$,
 $\varepsilon = \text{residual } i (\varepsilon \sim N(0, \sigma^2))$

สมการถดถอยจากตัวอย่าง $\hat{Y}_i = E(Y_i) = a + bX_i$

เมื่อ $a = \text{intercept (statistic)}$,
 $b = \text{coefficient regression (statistic)}$

13

การสร้างสมการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square)

$Y_i = A + BX_i + \varepsilon_i$

X	Y	\hat{Y}	e
X_1	Y_1	\hat{Y}_1	e_1
X_2	Y_2	\hat{Y}_2	e_2
X_3	Y_3	\hat{Y}_3	e_3
...
X_n	Y_n	\hat{Y}_n	e_n

ต้องการสร้างสมการถดถอย $\hat{Y}_i = a + bX_i$

ที่ทำให้ $Y_i - \hat{Y}_i = e_i$

มีค่าน้อยที่สุด ($e_i = \text{ค่าความคลาดเคลื่อนในการประมาณ}$)

$\sum e_i = 0$

ต้องการหาค่า a และ b ที่จะทำให้ค่า e_i มีค่าน้อย ๆ

14

การสร้างสมการถดถอยด้วยวิธีกำลังสองน้อยที่สุด

การหาค่า a และ b เพื่อให้ค่า e_i มีค่าน้อย ๆ

- เนื่องจาก $\sum e_i$ มีค่าเข้าใกล้ 0 เสมอ
- ดังนั้นจึงพิจารณาหา a และ b ที่ทำให้ $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ มีค่าต่ำที่สุด

แทนค่า $\hat{Y}_i = a + bX_i$ → $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2$

15

ค่า a หาได้จากการใช้อนุพันธ์เชิงส่วน (Partial Derivative) ของ $\sum e_i^2$ เทียบกับ a แล้วให้เท่ากับ 0 นั่นคือ

$$\frac{\partial}{\partial a} \sum (Y_i - a - bX_i)^2 = 0$$

→ $-2 \sum (Y_i - a - bX_i) = 0$

→ $a \cdot n + b \sum X_i = \sum Y_i \quad \dots(1)$

ค่า b หาได้จาก

$$\frac{\partial}{\partial b} \sum (Y_i - a - bX_i)^2 = 0$$

→ $a \sum X_i + b \sum X_i^2 = \sum X_i Y_i \quad \dots(2)$

16

จาก (1) และ (2) $S_{XY} = \text{ความแปรปรวนร่วม X และ Y}$

$$a \cdot n + b \sum X_i = \sum Y_i \quad \dots(1) \quad S_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

$$a \sum X_i + b \sum X_i^2 = \sum X_i Y_i \quad \dots(2) \quad S_X^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

จัดรูปใหม่ จะได้

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{S_{XY}}{S_X^2}$$

เมื่อคำนวณได้ b แทนค่าเพื่อหา a → $a = \bar{Y} - b\bar{X}$

จะได้สมการถดถอย $\hat{Y}_i = a + bX_i$

17

การคำนวณ b และ a

$$b = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \quad \text{และ} \quad a = \bar{Y} - b\bar{X}$$

X	Y
X_1	Y_1
X_2	Y_2
...	...
X_n	Y_n

$$\bar{Y} = \frac{\sum Y}{n}$$

$$\bar{X} = \frac{\sum X}{n}$$

$\sum X \quad \sum Y \quad \sum XY \quad \sum X^2$

18

การอธิบาย หรือการทำนายค่า

- จากสมการถดถอยที่ได้ $\hat{Y}_i = a + bX_i$
 - ค่า a คือ ค่าประมาณ Y เมื่อ X = 0
 - ค่า b คือ ค่าการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป 1 หน่วย
 - เมื่อทราบค่า X จะประมาณค่า Y ที่สอดคล้อง สามารถแทนค่า X ลงในสมการถดถอยเพื่อประมาณค่า Y ได้

ข้อควรระวัง!!!

- เมื่อทราบค่า Y และต้องการพยากรณ์ค่า X ที่สอดคล้อง ไม่สามารถประมาณค่าได้ (โดยตรง)

19

ตัวอย่าง 1

ศึกษาความสัมพันธ์ระหว่างน้ำหนักที่ลดได้ด้วยกระบวนการลดน้ำหนักวิธีหนึ่งกับระยะเวลาที่ทำการลดน้ำหนักโดยการรวบรวมข้อมูลของผู้เข้ากระบวนการลดน้ำหนักจำนวน 10 คน ดังนี้

ระยะเวลา (เดือน)	4	16	14	2	10	22	9	12	4	7
น้ำหนักที่ลด (กก.)	8	32	26	1	22	35	19	20	5	12

- ก. จงสร้างสมการแสดงความสัมพันธ์เพื่อประมาณน้ำหนักที่ลดจากรยะเวลาที่ทำการลดน้ำหนัก
- ข. คุณมยุรีจะใช้เวลาลดน้ำหนัก 6 เดือน จงประมาณว่า เธอจะลดน้ำหนักได้กี่กิโลกรัม

20

- ก. จงสร้างสมการแสดงความสัมพันธ์เพื่อประมาณน้ำหนักที่ลดจากรยะเวลาที่ทำการลดน้ำหนัก

ต้องการประมาณน้ำหนักที่ลด เมื่อทราบระยะเวลาที่ใช้ลดน้ำหนัก

ตัวแปรตาม Y

ตัวแปรอิสระ X

สมการถดถอย

$$\hat{Y}_i = a + bX_i$$

$$b = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \quad a = \bar{Y} - b\bar{X}$$

21

$$b = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \quad a = \bar{Y} - b\bar{X}$$

ระยะเวลา X	4	16	14	2	10	22	9	12	4	7
น้ำหนักที่ลด Y	8	32	26	1	22	35	19	20	5	12

$$\sum X = 4 + 16 + \dots + 7 = 100 \quad \sum X^2 = 4^2 + \dots + 7^2 = 1346$$

$$\sum Y = 8 + 32 + \dots + 12 = 180 \quad \sum XY = (4 \times 8) + \dots + (7 \times 12) = 2415$$

$$b = \frac{2415 - (10)(10)(18)}{1346 - (10)10^2} = 1.7775$$

$$a = 18 - (1.7775 \times 10) = 0.225$$

$$\hat{Y}_i = 0.225 + 1.7775 X_i$$

22

- ข. คุณมยุรีจะใช้เวลาลดน้ำหนัก 6 เดือน จงประมาณว่า เธอจะลดน้ำหนักได้กี่กิโลกรัม

จากสมการถดถอยที่ได้ $\hat{Y}_i = 0.225 + 1.7775 X_i$

ต้องการประมาณน้ำหนักที่ลดลง เมื่อใช้เวลาลดน้ำหนัก 6 เดือน

- X = 6 Y = ?

$$\hat{Y}_i = 0.225 + (1.7775 \times 6) = 10.89$$

23

ค่าความคลาดเคลื่อนมาตรฐานของการประมาณ

$$\hat{Y}_i = a + bX_i$$

ค่าพยากรณ์ \hat{Y}_i ที่ประมาณได้ อาจมีค่าเท่ากับหรือคลาดเคลื่อนไปจากค่า Y_i ก็ได้

หากต้องการทราบถึงความคลาดเคลื่อนที่เกิดขึ้นจากการประมาณหรือการพยากรณ์

- ค่าความคลาดเคลื่อนมาตรฐาน (Standard error, s_e) ของการประมาณ เป็นค่าที่แสดงค่าเฉลี่ยการกระจายระหว่างค่าพยากรณ์กับค่าจริง

24

■ สูตรสำหรับการหาค่าความแปรปรวนทั่วไป (S^2)
 ความคลาดเคลื่อนจากการประมาณ X_i ด้วย \bar{X} คือ $X_i - \bar{X}$

$$\Rightarrow s_x^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \Rightarrow \frac{SS}{df}$$
 (Sum of Square, degree of freedom)

■ ความแปรปรวนของความคลาดเคลื่อนจากการประมาณ (S_e^2)
 ความคลาดเคลื่อนในการประมาณคือ $e_i = Y_i - \hat{Y}_i$

$$\Rightarrow s_e^2 = \frac{SS}{df} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$$

$$df = n-2$$
 เนื่องจากในการประมาณ A และ B ทำให้ต้องสูญเสียความเป็นอิสระไป 2 ตัว

$$s_e^2 = \frac{1}{n-2} \left[\sum (Y - \bar{Y})^2 - b^2 \sum (X - \bar{X})^2 \right] = \frac{n-1}{n-2} [s_y^2 - b^2 s_x^2]$$

25

■ ตัวอย่าง 3
 จากตัวอย่างที่ 1 ผลการสร้างสมการถดถอยเพื่อประมาณน้ำหนักที่ลดลง (Y) จากระยะเวลาที่ใช้ในการลดน้ำหนัก (X) ได้เป็น

$$\hat{Y}_i = 0.225 + 1.7775 X_i$$

จงประมาณค่าความคลาดเคลื่อนมาตรฐาน

X	4	16	14	2	10	22	9	12	4	7
Y	8	32	26	1	22	35	19	20	5	12
Y^	7.34	28.67	25.11	3.78	18.00	39.33	16.22	21.56	7.34	12.67
e	0.66	3.33	0.89	-2.78	4.00	-4.33	2.78	-1.56	-2.34	-0.67

$$s_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{.66^2 + 3.33^2 + \dots + (-.67)^2}{8} = \frac{70.85}{8} = 8.86 \Rightarrow S_e = 2.98$$

26

■ ตัวอย่าง 3

X	4	16	14	2	10	22	9	12	4	7	100
Y	8	32	26	1	22	35	19	20	5	12	180
Y^2	64	1024	676	1	484	1225	361	400	25	144	4404
X^2	16	256	196	4	100	484	81	144	16	49	1346

$$s_e^2 = \frac{n-1}{n-2} [s_y^2 - b^2 s_x^2] = \frac{9}{8} [129.33 - (1.7775^2 \times 38.44)] = 8.86$$

$$s_y^2 = \frac{\sum Y_i^2 - n\bar{Y}^2}{n-1} = \frac{4404 - (10 \times 18^2)}{9} = \frac{1164}{9} = 129.33 \Rightarrow S_e = 2.98$$

$$s_x^2 = \frac{\sum X_i^2 - n\bar{X}^2}{n-1} = \frac{1346 - (10 \times 10^2)}{9} = \frac{346}{9} = 38.44$$

27

■ การถดถอย

- สมการถดถอยอย่างง่าย
- การอนุมานการถดถอย

28

■ การอนุมานค่าพารามิเตอร์ของการถดถอย

- ประมาณค่าพารามิเตอร์ A, B โดยการ
 - ประมาณค่าแบบจุด
 - ประมาณค่าแบบช่วง
- ทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ A, B เพื่อพิจารณาว่า สมการถดถอยระหว่าง X และ Y ที่ได้จากข้อมูลตัวอย่างเพียงพอที่จะสรุปได้หรือไม่ว่าสามารถใช้อธิบาย 2 ตัวแปรนี้ในข้อมูลประชากร

29

■ การประมาณค่าแบบช่วงของ A

- ลักษณะการแจกแจงของ a
 เมื่อทำการสุ่มตัวอย่าง k ครั้ง แต่ละครั้งด้วยขนาดตัวอย่าง n จะได้ a_1, a_2, \dots, a_k พบว่า ตัวแปรสุ่ม a จะมีการแจกแจงแบบ t คือ

$$t = \frac{a - E(a)}{\sqrt{Var(a)}} = \frac{a - A}{S_a}, \quad \nu = n-2$$
 โดย $E(a) = A$

$$Var(a) = S_a^2 = s_e^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] = \frac{s_e^2}{(n-1)s_x^2} \cdot \frac{\sum X_i^2}{n}$$

ช่วงความเชื่อมั่นที่ $(1 - \alpha)100\%$ ของ A คือ

$$a - t_{\frac{\alpha}{2}, \nu} \cdot s_a < A < a + t_{\frac{\alpha}{2}, \nu} \cdot s_a \Rightarrow A = a \pm t_{\frac{\alpha}{2}, \nu} \cdot s_a$$

30

การประมาณค่าแบบช่วงของ B

- ลักษณะการแจกแจงของ b

เมื่อทำการสุ่มตัวอย่าง k ครั้ง ๆ ละ n หน่วย จะได้ b_1, b_2, \dots, b_k พบว่า ตัวแปรสุ่ม b จะมีการแจกแจงแบบ t คือ

$$t = \frac{b - E(b)}{\sqrt{\text{Var}(b)}} = \frac{b - B}{S_b}, \quad v = n - 2 \quad \text{โดย } E(b) = B$$

$$\text{Var}(b) = S_b^2 = \frac{s_e^2}{\sum (X_i - \bar{X})^2} = \frac{s_e^2}{(n-1)s_x^2}$$

ช่วงความเชื่อมั่นที่ $(1 - \alpha)100\%$ ของ B คือ

$$b - (t_{1-\frac{\alpha}{2}, v} \cdot s_b) < B < b + (t_{1-\frac{\alpha}{2}, v} \cdot s_b) \quad \Rightarrow \quad B = b \pm (t_{1-\frac{\alpha}{2}, v} \cdot s_b)$$

31

การทดสอบสมมติฐาน

ในการวิเคราะห์การถดถอย มักต้องการศึกษาว่า X มีผลต่อ Y หรือไม่ในกลุ่มประชากรเป้าหมาย ?

- ตัวพารามิเตอร์ที่จะช่วยตอบคำถามนี้ได้คือ B
- ดังนั้น ในการทดสอบสมมติฐานโดยอาศัยผลที่ได้จากกลุ่มตัวอย่างไปอธิบายลักษณะความสัมพันธ์ในประชากร จะทำการวิเคราะห์ใน B เป็นส่วนมาก

32

การทดสอบสมมติฐานเกี่ยวกับ B

- สมมติฐานในการทดสอบ ▪ เขตวิกฤต ที่ α

การทดสอบ 2 ทาง

- $H_0 : B = B_0$
- $H_1 : B \neq B_0$

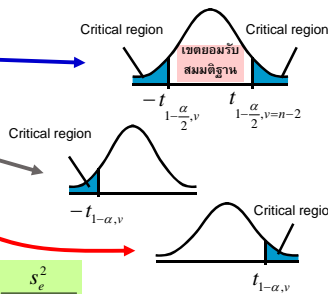
การทดสอบทางเดียว

- $H_0 : B = B_0$
- $H_1 : B < B_0$

- $H_0 : B = B_0$
- $H_1 : B > B_0$

- สถิติที่ใช้ทดสอบ

$$t = \frac{b - B_0}{s_b} \quad \text{โดย } s_b^2 = \frac{s_e^2}{(n-1)s_x^2}$$



33

การทดสอบสมมติฐานเกี่ยวกับ A

- สมมติฐานในการทดสอบ ▪ เขตวิกฤต ที่ α

การทดสอบ 2 ทาง

- $H_0 : A = A_0$
- $H_1 : A \neq A_0$

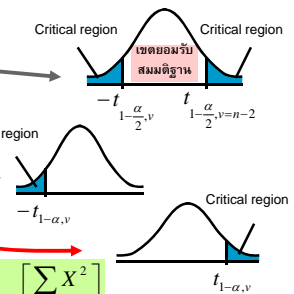
การทดสอบทางเดียว

- $H_0 : A = A_0$
- $H_1 : A < A_0$

- $H_0 : A = A_0$
- $H_1 : A > A_0$

- สถิติที่ใช้ทดสอบ

$$t = \frac{a - A_0}{s_a} \quad \text{โดย } s_a^2 = \frac{s_e^2}{(n-1)s_x^2} \left[\frac{\sum X^2}{n} \right]$$



34

ตัวอย่าง 4

ในการศึกษาเกี่ยวกับความสัมพันธ์ของระยะทางการใช้งานของรถ (X: หน่วยเป็น 1000 กม.) และค่าใช้จ่ายในการตรวจสอบ (Y: หน่วยเป็น 100 บาท) โดยสุ่มตัวอย่างรถยนต์ยี่ห้อหนึ่ง ซึ่งมีขนาดของเครื่องยนต์เท่ากัน จำนวน 20 คัน ปรากฏผลการศึกษาคำนวณเบื้องต้นดังนี้

$$\sum_{i=1}^{20} x_i = 600, \quad \sum_{i=1}^{20} y_i = 300, \quad \sum_{i=1}^{20} (x_i - \bar{x})^2 = 8540,$$

$$\sum_{i=1}^{20} (y_i - \bar{y})^2 = 1542, \quad \sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = 3499$$

- ก. จงทดสอบระยะทางการใช้งานของรถยนต์ และค่าใช้จ่ายในการตรวจสอบ มีความสัมพันธ์เชิงเส้นตรงหรือไม่ กำหนดให้ระดับนัยสำคัญ 5% หาก $B = 0$ แสดงว่า X กับ Y ไม่มีความสัมพันธ์เชิงเส้นตรง
- ข. จงประมาณค่า A โดยใช้ระดับความเชื่อมั่น 95%

35

- สมมติฐานทางสถิติ

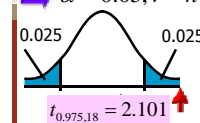
- $H_0 : B = 0$
- $H_1 : B \neq 0$

- สถิติทดสอบ

$$t = \frac{b - B_0}{s_b} = \frac{0.41}{\sqrt{0.0007}} = 13.67$$

- เขตวิกฤต

$$\alpha = 0.05, v = n - 2 = 18$$



- ค่าตัวเลข $b = \frac{3499}{8540} = 0.41$

$$s_b^2 = \frac{s_e^2}{\sum (X - \bar{X})^2} = \frac{5.92}{8540} = 0.0007$$

$$s_e^2 = \frac{1}{n-2} [\sum (Y - \bar{Y})^2 - b^2 \sum (X - \bar{X})^2]$$

$$= \frac{1}{18} [1542 - (0.41^2)(8540)] = 5.92$$

- สรุปผล

ปฏิเสธ H_0

นั่นคือที่ระดับนัยสำคัญ 0.05 สรุปได้ว่าตัวแปรทั้ง 2 มีความสัมพันธ์เป็นเส้นตรง

$$t < -2.101 \text{ หรือ } t > 2.101$$

36

$$\sum_{i=1}^{20} x_i = 600, \quad \sum_{i=1}^{20} y_i = 300,$$

$$\sum_{i=1}^{20} (x_i - \bar{x})^2 = 8540,$$

$$\sum_{i=1}^{20} (y_i - \bar{y})^2 = 1542,$$

$$\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = 3499$$

$$a = 15 - 0.41(30) = 2.7$$

$$s_e^2 = 5.92$$

$$s_a^2 = s_e^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X - \bar{X})^2} \right]$$

$$= 0.9199$$

$$S_a = \sqrt{0.9199} = 0.96$$

$$a \pm t_{1-\frac{\alpha}{2}, v} \cdot s_a$$

$2.7 \pm 2.101(0.96)$
 2.7 ± 2.02
 $0.68 < A < 4.72$

ที่ความเชื่อมั่น 95% ค่าประมาณของ A มีค่าระหว่าง 0.68 ถึง 4.72

$t_{0.975, 18} = 2.101$

37

สหสัมพันธ์

- สัมประสิทธิ์สหสัมพันธ์
- การอนุมานสัมประสิทธิ์สหสัมพันธ์

38

สหสัมพันธ์ Correlation

- ศึกษาขนาดความสัมพันธ์ระหว่างตัวแปรปริมาณ 2 ตัว
 - ตัวแปรอิสระ และตัวแปรตาม
 - ตัวแปรปริมาณ 2 ตัวใด ๆ

วัดขนาดความสัมพันธ์ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

สำหรับประชากร $\Rightarrow \rho$ (rho), $-1 \leq \rho \leq 1$

สำหรับกลุ่มตัวอย่าง $\Rightarrow r$, $-1 \leq r \leq 1$

39

การคำนวณค่าสปส.สหสัมพันธ์ของตัวอย่าง

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{[\sum (X_i - \bar{X})^2][\sum (Y_i - \bar{Y})^2]}}$$

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

$$= \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{(\sum X^2 - n\bar{X}^2)(\sum Y^2 - n\bar{Y}^2)}} = b \cdot \frac{s_X}{s_Y}$$

$s_X = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$
 $s_Y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n-1}}$

40

ขนาด & รูปแบบความสัมพันธ์ระหว่าง X กับ Y

$r \rightarrow +1$

แสดงว่า มีความสัมพันธ์ในทิศทางเดียวกันอย่างสมบูรณ์

$r \rightarrow -1$

แสดงว่า มีความสัมพันธ์ในทิศทางตรงข้ามอย่างสมบูรณ์

$r = 0$

แสดงว่า ไม่มีความสัมพันธ์กันเชิงเส้นตรง

41

ตัวอย่าง 5

ข้อมูลของบริษัทจำหน่ายรถยนต์ที่ใช้แล้ว โดยบันทึกข้อมูลอายุการใช้งานของรถยนต์ (X: หน่วย ปี) และราคาขาย (Y: หน่วย พันบาท) ของรถจำนวน 6 คัน ดังนี้

X : อายุการใช้งาน (ปี)	4	10	2	1	3	5
Y : ราคาขาย (พันบาท)	895	125	1395	1795	1245	695

จงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอายุการใช้งานของรถยนต์และราคาขาย

42

X	Y	XY	X ²	Y ²
4	895	3580	16	801025
10	125	1250	100	15625
2	1395	2790	4	1946025
1	1795	1795	1	3222025
3	1245	3735	9	1550025
5	695	3475	25	483025
รวม	25	6150	155	8017750

$\bar{X} = 25/6 = 4.17$
 $\bar{Y} = 6150/6 = 1025$
 $\sum XY = 16625$
 $\sum X^2 = 155$
 $\sum Y^2 = 8017750$

คำนวณค่า สปส. สหสัมพันธ์

$$r = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{[\sum X_i^2 - n\bar{X}^2][\sum Y_i^2 - n\bar{Y}^2]}} = -0.964$$

สหสัมพันธ์

- สัมประสิทธิ์สหสัมพันธ์
- การอนุมานสัมประสิทธิ์สหสัมพันธ์

การอนุมานค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ)

- การประมาณค่า (Estimation)
 - การประมาณค่าแบบจุด (Point estimate)
 - การประมาณค่าแบบช่วง (Interval estimate)
- การทดสอบสมมติฐาน (Hypothesis testing)

ในการอนุมานค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) จำเป็นต้องพิจารณาการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r)

การแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง

- ลักษณะการแจกแจงของค่า r มี 2 กรณี คือ
 - เมื่อ $\rho = 0$
r มีการแจกแจงแบบที่ (t) $\rightarrow t = r / \sqrt{\frac{1-r^2}{n-2}}$ $v = n - 2$
 - เมื่อ $\rho = \rho_0$ (ρ_0 เป็นค่าคงที่ใด ๆ)
แปลงค่า r เป็น R โดย $R = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$
พบว่า R มีการแจกแจงแบบปกติ โดย $R \sim N(\mu_R, \sigma_R^2)$
 $\rightarrow Z = \frac{R - \mu_R}{\sigma_R}$ เมื่อ $\mu_R = \frac{1}{2} \ln \left(\frac{1+\rho_0}{1-\rho_0} \right)$ $\sigma_R = \frac{1}{\sqrt{n-3}}$

การใช้ตารางสำเร็จรูปในการแปลงค่า $r \leftrightarrow R, \rho \leftrightarrow \mu_R$

- ใช้ตารางที่ 7 หน้า 319 R, μ_R

r	ρ	R	μ _R
.01	.010	.34	.354
.02	.020	.35	.366
...
.33	.343	.66	.793

$r = .35 \quad R = .366 \quad \rho = .68 \quad \mu_R = .829$
 $R = .35 \quad r = .34 \quad \mu_R = .79 \quad \rho = .66$

การประมาณค่าสปส. สหสัมพันธ์ (ρ)

- การประมาณค่าแบบจุด จะประมาณค่า ρ ด้วยค่า r
- การประมาณค่าแบบช่วง
ในการประมาณค่า มักเป็นกรณีที่คาดว่า $\rho \neq 0$ นั่นคือ $\rho = \rho_0$ ดังนั้นจึงควรใช้ตัวสถิติ Z ที่ระดับความเชื่อมั่น $(1 - \alpha)100\%$
 $R - (Z_{\frac{1-\alpha}{2}} \cdot \sigma_R) < \mu_R < R + (Z_{\frac{1-\alpha}{2}} \cdot \sigma_R) \rightarrow \mu_R = R \pm (Z_{\frac{1-\alpha}{2}} \cdot \sigma_R)$
 แปลงค่าประมาณ μ_R กลับไปเป็น ρ

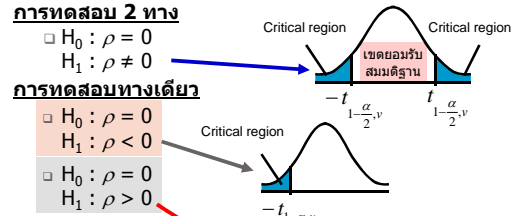
การทดสอบสมมติฐานเกี่ยวกับค่า ρ

- เนื่องจากลักษณะการแจกแจงของ r จำแนกเป็น 2 กรณี ตามลักษณะของ ρ ($\rho = 0$ หรือ $\rho = \rho_0$)
- ดังนั้นในการทดสอบสมมติฐานเกี่ยวกับค่า ρ จึงจำแนกเป็น 2 กรณี ตามลักษณะข้อสงสัยเกี่ยวกับค่า ρ
 - การทดสอบสมมติฐานว่า $\rho = 0$?
 - การทดสอบสมมติฐานว่า $\rho = \rho_0$?

49

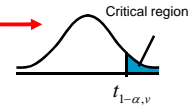
การทดสอบสมมติฐานเกี่ยวกับ $\rho = 0$

- สมมติฐานในการทดสอบ
 - การทดสอบ 2 ทาง
 - $H_0 : \rho = 0$
 - $H_1 : \rho \neq 0$
 - การทดสอบทางเดียว
 - $H_0 : \rho = 0$
 - $H_1 : \rho < 0$
 - $H_0 : \rho = 0$
 - $H_1 : \rho > 0$



- สถิติที่ใช้ทดสอบ

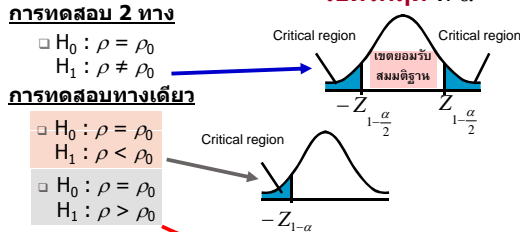
$$t = r / \sqrt{\frac{1-r^2}{n-2}} \quad v = n - 2$$



50

การทดสอบสมมติฐานเกี่ยวกับ $\rho = \rho_0$

- สมมติฐานในการทดสอบ
 - การทดสอบ 2 ทาง
 - $H_0 : \rho = \rho_0$
 - $H_1 : \rho \neq \rho_0$
 - การทดสอบทางเดียว
 - $H_0 : \rho = \rho_0$
 - $H_1 : \rho < \rho_0$
 - $H_0 : \rho = \rho_0$
 - $H_1 : \rho > \rho_0$



- สถิติที่ใช้ทดสอบ

$$Z = \frac{R - \mu_R}{\sigma_R}$$

51

ตัวอย่าง 6

สุ่มตัวอย่างนักศึกษาสายวิทยาศาสตร์และเทคโนโลยี จำนวน 30 คน ศึกษาความสัมพันธ์ระหว่างลำดับชั้นเฉลี่ยเมื่อสำเร็จชั้นมัธยมศึกษาตอนปลาย และลำดับชั้นเฉลี่ยเมื่อเรียนจบ 2 ภาคการศึกษาแรกในมหาวิทยาลัย พบว่าได้ค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.35

ด้วยระดับนัยสำคัญ 5% จงทดสอบว่าลำดับชั้นเฉลี่ยของผลการศึกษาใน 2 ระดับชั้นการศึกษาของนักศึกษาดังกล่าวมีความสัมพันธ์กันหรือไม่

คำถาม ตัวแปรทั้งสองมีความสัมพันธ์กันเชิงเส้นตรงหรือไม่

- ค่า ρ แตกต่างจาก 0 แสดงว่า มีความสัมพันธ์เชิงเส้นตรง
- $\rho \neq 0$?

52

- สมมติฐานทางสถิติ
 - $H_0 : \rho = 0$
 - $H_1 : \rho \neq 0$

- สถิติทดสอบ

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

- เขตวิกฤต

$$\alpha = 0.05, v = n - 2 = 28$$



$$t < -2.048 \text{ หรือ } t > 2.048$$

53

- คำนวณค่า

$$t = \frac{0.35}{\sqrt{\frac{1-0.35^2}{30-2}}} = 1.977$$

- สรุปผล

ไม่ปฏิเสธ H_0
นั่นคือที่ระดับนัยสำคัญ 0.05 สรุปว่าผลการเรียนม.ปลายกับมหาวิทยาลัยไม่มีความสัมพันธ์เชิงเส้นตรง

ตัวอย่าง 7

สุ่มตัวอย่างมารดาและทารกเกิดใหม่ 35 คู่ ศึกษาความสัมพันธ์ระหว่างน้ำหนักของมารดาที่เพิ่มขึ้นระหว่างตั้งครรภ์และน้ำหนักแรกเกิดของบุตรพบว่าค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.8

- ด้วยระดับนัยสำคัญ 10% จงทดสอบสมมติฐานว่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มประชากรมีค่าต่ำกว่า 0.9
- จงหาช่วงความเชื่อมั่น 90% ของสัมประสิทธิ์สหสัมพันธ์ของประชากร

54

ก. จงทดสอบว่า ส.ป.ส. สหสัมพันธ์ของประชากรต่ำกว่า 0.9 ($\rho < 0.9$)

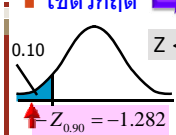
- สมมติฐานทางสถิติ
 - $H_0: \rho = 0.9$
 - $H_1: \rho < 0.9$
- สถิติทดสอบ

คำนวณค่า

จาก $r = 0.8 \Rightarrow R = 1.099$
 $\rho_0 = 0.9 \Rightarrow \mu_R = 1.472$
 $\sigma_R = 1/\sqrt{n-3} = 1/\sqrt{32}$

$$Z = \frac{R - \mu_R}{\sigma_R} = \frac{1.099 - 1.472}{1/\sqrt{32}} = -2.06$$

เขตวิกฤต $\alpha = 0.10$ $Z < -1.282$



สรุปผล

ปฏิเสธ H_0
 นั่นคือที่ระดับนัยสำคัญ 0.10 สรุปว่า ส.ป.ส. สหสัมพันธ์มีค่าต่ำกว่า 0.9

55

ข. ให้หาช่วงความเชื่อมั่นของ ρ โดยใช้ระดับความเชื่อมั่น 90%

$R \pm Z_{1-\frac{\alpha}{2}} \cdot \sigma_R$

$R = 1.099$
 $\sigma_R = 1/5.66$
 $Z_{1-\frac{\alpha}{2}} = Z_{0.95} = 1.645$

$1.099 \pm 1.645(1/5.66)$
 1.099 ± 0.291
 $0.81 < \mu_R < 1.39$
 $0.67 < \rho < 0.88$

56

ทดสอบตั้งสมมติฐานทางสถิติ

- ต้องการทดสอบว่า X กับ Y มีความสัมพันธ์เชิงเส้นตรงหรือไม่

ก) $H_0: \rho = 0$ $H_1: \rho \neq 0$
 ข) $H_0: A = 0$ $H_1: A \neq 0$
 ค) $H_0: B = 0$ $H_1: B \neq 0$
 ง) $H_0: B = 0$ $H_1: B > 0$
 จ) $H_0: \rho = 0$ $H_1: \rho > 0$

57

ทดสอบตั้งสมมติฐานทางสถิติ

- ต้องการทดสอบว่า สมการถดถอยที่ได้เป็นสมการเชิงเส้นตรงหรือไม่

ก) $H_0: \rho = 0$ $H_1: \rho \neq 0$
 ข) $H_0: A = 0$ $H_1: A \neq 0$
 ค) $H_0: B = 0$ $H_1: B \neq 0$

58

ทดสอบตั้งสมมติฐานทางสถิติ

- อยากทราบว่า X กับ Y มีความสัมพันธ์กันในทิศทางเดียวกันหรือไม่

ก) $H_0: \rho = 0$ $H_1: \rho \neq 0$
 ข) $H_0: \rho = 0$ $H_1: \rho > 0$
 ค) $H_0: B = 0$ $H_1: B \neq 0$
 ง) $H_0: B = 0$ $H_1: B > 0$

59

การถดถอยพหุคูณ

MULTIPLE LINEAR REGRESSION

60

การถดถอยพหุคูณ (Multiple Linear Regression)

- ในกรณีที่ต้องการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม Y กับชุดของตัวแปรอิสระ X_1, X_2, \dots, X_p

สามารถทำการวิเคราะห์ได้โดยสร้างสมการถดถอยพหุคูณ

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} + \dots + b_p X_{pi}$$

- ในที่นี้จะขอศึกษากรณีความสัมพันธ์ระหว่างตัวแปรตาม Y กับชุดของตัวแปรอิสระ X_1, X_2

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

61

สมการถดถอยพหุคูณ

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

- Y = ตัวแปรตาม
- X_1 = ตัวแปรอิสระตัวที่ 1
- X_2 = ตัวแปรอิสระตัวที่ 2
- a = intercept
- b_1 = สัมประสิทธิ์การถดถอยของตัวแปร X_1
สามารถอธิบายค่าได้ว่า เมื่อ X_1 เปลี่ยนแปลงไป 1 หน่วย จะทำให้ Y เปลี่ยนค่าไป b_1 หน่วย เมื่อกำหนดให้ X_2 มีค่าคงที่
- b_2 = สัมประสิทธิ์การถดถอยของตัวแปร X_2 (เมื่อ X_2 เปลี่ยนแปลงไป 1 หน่วย จะทำให้ Y เปลี่ยนค่าไป b_2 หน่วย เมื่อกำหนดให้ X_1 มีค่าคงที่)

62

ข้อมูลในการวิเคราะห์การถดถอยพหุคูณ

Case	ตัวแปรอิสระ		ตัวแปรตาม (Y)
	X_1	X_2	
1	X_{11}	X_{12}	Y_1
2	X_{21}	X_{22}	Y_2
3	X_{31}	X_{32}	Y_3
...
n	X_{n1}	X_{n2}	Y_n

63

วิธีกำลังสองน้อยที่สุด (Least Square)

$$Y_i = A + B_1 X_{1i} + B_2 X_{2i} + \varepsilon_i \Rightarrow \hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

$$Y_i - \hat{Y}_i = e_i \Rightarrow \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

หาค่า a, b_1 และ b_2 เพื่อให้ $\sum_{i=1}^n e_i^2$ มีค่าน้อยที่สุด

$$\text{แทนค่า } \hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - (a + b_1 X_{1i} + b_2 X_{2i}))^2$$

64

- ค่า a หาได้จากอนุพันธ์เชิงส่วน (Partial Derivative) ของ $\sum e_i^2$ เทียบกับ a แล้วให้เท่ากับ 0 นั่นคือ

$$\frac{\partial}{\partial a} \sum_i (Y - a - b_1 X_1 - b_2 X_2)^2 = 0$$

$$\Rightarrow a \cdot n + b_1 \sum_i X_{1i} + b_2 \sum_i X_{2i} = \sum_i Y \quad \dots(1)$$

- ค่า b_1 หาได้จาก $\frac{\partial}{\partial b_1} \sum_i (Y - a - b_1 X_1 - b_2 X_2)^2 = 0$

$$\Rightarrow a \sum_i X_{1i} + b_1 \sum_i X_{1i}^2 + b_2 \sum_i X_{1i} X_{2i} = \sum_i X_{1i} Y \quad \dots(2)$$

- ค่า b_2 หาได้จาก $\frac{\partial}{\partial b_2} \sum_i (Y - a - b_1 X_1 - b_2 X_2)^2 = 0$

$$\Rightarrow a \sum_i X_{2i} + b_1 \sum_i X_{1i} X_{2i} + b_2 \sum_i X_{2i}^2 = \sum_i X_{2i} Y \quad \dots(3)$$

65

ตัวอย่าง 2

ในการศึกษาความสัมพันธ์ระหว่างตัวแปร X_1, X_2 และ Y จึงสุ่มตัวอย่างบันทึกข้อมูลได้ดังนี้

หน่วยตัวอย่าง	X_1	X_2	Y
1	-2	-2	2
2	-1	-2	1
3	0	0	4
4	1	0	3
5	2	4	5

ก. จงสร้างสมการถดถอยพหุคูณเชิงเส้นตรงเพื่อประมาณ Y

ข. ถ้า $X_1=3$ และ $X_2=3$ จงประมาณค่า Y

66

ก. จงสร้างสมการถดถอยพหุคูณเชิงเส้นตรง เพื่อประมาณ Y

$$a.n + b_1 \sum X_1 + b_2 \sum X_2 = \sum Y \rightarrow 5.a = 15$$

$$a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y \rightarrow 10.b_1 + 14.b_2 = 8$$

$$a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 = \sum X_2 Y \rightarrow 14.b_1 + 24.b_2 = 14$$

ตัวอย่าง	X_1	X_2	Y	$X_1 X_2$	$X_1 Y$	$X_2 Y$	X_1^2	X_2^2
1	-2	-2	2	4	-4	-4	4	4
2	-1	-2	1	2	-1	-2	1	4
3	0	0	4	0	0	0	0	0
4	1	0	3	0	3	0	1	0
5	2	4	5	8	10	20	4	16
Total	0	0	15	14	8	14	10	24

67

- จาก $5.a = 15$
 $10.b_1 + 14.b_2 = 8$
 $14.b_1 + 24.b_2 = 14$
- จะได้ $a = 3$ $b_1 = -0.09$ $b_2 = 0.64$

$$\hat{Y}_i = 3 - 0.09X_1 + 0.64X_2$$

ข. ถ้า $X_1 = 3$ และ $X_2 = 3$ จงประมาณค่า Y

$$\hat{Y}_i = 3 - 0.09(3) + 0.64(3) = 4.65$$

68