

กระบวนการแก้ปัญหาโดยใช้สถิติ ในเรื่องความสัมพันธ์ของสองตัวแปร

ผู้ช่วยศาสตราจารย์อินทิรา เนขุนทด

ภาควิชาสถิติ คณะวิทยาศาสตร์

ห้องพัก STB 308

จุดประสงค์

- นักศึกษาสามารถอธิบายกรณีที่ตัวแปรสองตัวแปรมีความสัมพันธ์กัน และสามารถอธิบายความสัมพันธ์นั้นได้ในรูปแบบของสมการ
- สามารถพยากรณ์ค่าตัวแปรที่สนใจได้
- สามารถบอกขนาดความสัมพันธ์ได้

ลักษณะของข้อมูล

- ข้อมูลเชิงปริมาณ (Quantitative data)
ข้อมูล มีลักษณะที่เป็นตัวเลขหรือค่าวัด เช่น

- อายุ
- น้ำหนัก
- เกรดเฉลี่ย
- รายได้
- คะแนนสอบ
- ระยะทาง

ฯลฯ

ลักษณะของข้อมูล

- ข้อมูลเชิงคุณภาพ (Qualitative data)
ข้อมูลที่แสดงตามคุณลักษณะ สามารถจัดแบ่งเป็นกลุ่ม
เช่น

เพศ (ชาย หญิง)

คณะที่เรียน (มนุษยศาสตร์ วิทยาศาสตร์
เกษตรศาสตร์.....)

ศาสนา (พุทธ คริสต์ อิสลาม)

ความพอใจ (พอใจ ไม่พอใจ)

คุณภาพ (ดี ปานกลาง เลว)

ความสัมพันธ์ของสองตัวแปร

1. กรณีที่ข้อมูลเป็นข้อมูลเชิงปริมาณ

- อธิบายในรูปแบบสมการความสัมพันธ์ของสองตัวแปร
- หาระดับความสัมพันธ์ของสองตัวแปร

2. กรณีที่ข้อมูลเป็นข้อมูลเชิงคุณภาพ

- หาระดับความสัมพันธ์ของสองตัวแปร

ข้อมูลเป็นข้อมูลเชิงปริมาณ

วิธีทางสถิตินี้เรียกว่าการวิเคราะห์การถดถอย (regression analysis)
วิธีการนี้ได้เริ่มมีการศึกษาตั้งแต่ศตวรรษที่ 19 จากนักคณิตศาสตร์ชื่อ
Sir Francis Galton ศึกษาความสัมพันธ์เกี่ยวกับความสูงของพ่อแม่
กับความสูงของบุตร

พบว่าความสูงของคนถดถอยเข้าสู่ความสูงเฉลี่ยของประชากร
การวิเคราะห์การถดถอยเป็นวิธีการที่อธิบายในรูปแบบสมการความสัมพันธ์
ของสองตัวแปร

การศึกษาข้อมูลแบบคู่(คาดว่าสัมพันธ์กัน)

- น้ำหนักของคู่แฝดพี่กับน้อง
- คะแนนสอบกลางเทอม(X)กับคะแนนสอบปลายเทอม(Y)
- รายรับ(X) กับรายจ่าย(Y)ของครัวเรือน
- ยอดขายสินค้า(X) กับกำไรสุทธิ(Y)
- อัตราดอกเบี้ยเงินฝากที่ธนาคารคิดให้ (X) กับจำนวนเงินฝากในธนาคาร(Y)
- อายุของรถยนต์มือสอง(X) กับราคาขาย(Y)
- ระยะทางในการเดินทาง(X) กับค่าน้ำมันเชื้อเพลิง(Y)

หัวข้อเรื่อง การวิเคราะห์การถดถอย

จุดประสงค์

- สร้างสมการความสัมพันธ์ ของสองตัวแปร
- พยากรณ์ค่าตัวแปรหนึ่งจากอีกตัวแปรหนึ่ง

การวิเคราะห์การถดถอย

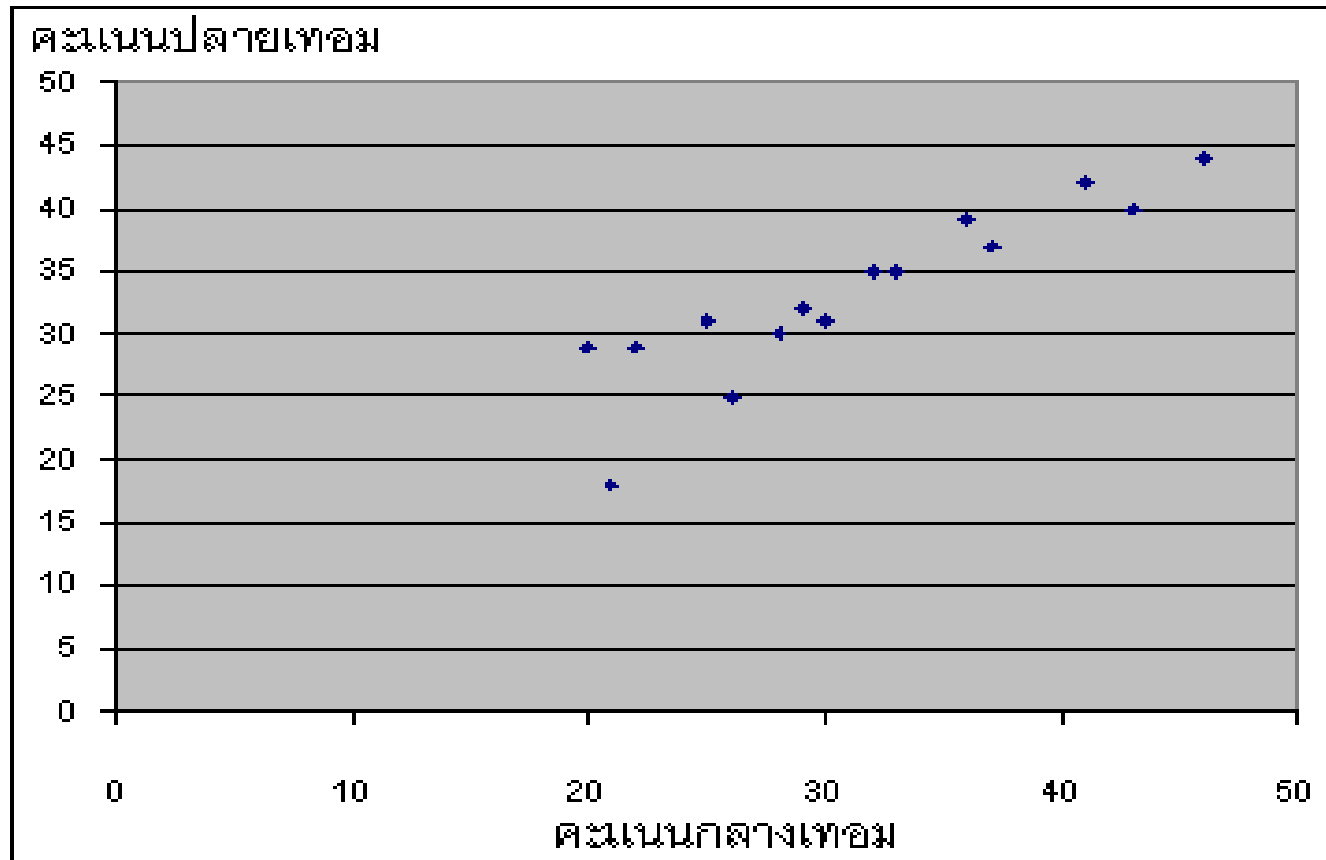
- ลักษณะข้อมูลเป็นข้อมูลเชิงปริมาณ (ข้อมูลมาจากค่าวัดที่เป็นตัวเลขเชิงปริมาณ)
- มีตัวแปรอิสระ(X)
- มีตัวแปรตาม(Y) 1 ตัว (เมื่อตัวแปรอิสระมีค่าเปลี่ยนไปตัวแปรตามจะเปลี่ยนด้วย)
- ถ้ามีตัวแปรอิสระ(X) 1 ตัว จะเรียกว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis)
- ถ้าหากมีตัวแปรอิสระ(X)มากกว่า 1 ตัวจะเรียกว่าการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis)
- ใช้สมการถดถอยพยากรณ์ตัวแปรตาม(Y)จากตัวแปรอิสระ(X)ได้

เมื่อมีข้อมูลที่มีตัวแปรคู่(X,Y)ที่คาดว่าจะมีความสัมพันธ์กันแล้ว ก่อนอื่นต้องทำการพล็อตค่าตัวแปรสองตัวบนแกน X และ y เรียกว่าแผนภาพกระจาย (scatter plot หรือ scatter diagram) เพื่อดูความสัมพันธ์เบื้องต้นเสียก่อน ทำให้เราเห็นแนวโน้มความสัมพันธ์ของข้อมูลว่ามีลักษณะเช่นใด(เป็นเส้นตรง, พาราโบลา, เอ็กซโพเนนเชียล ฯลฯ หรือไม่)

สมมุติข้อมูลคะแนนสอบกลางเทอม และปลายเทอม วิชา คณิตศาสตร์บูรณาการของนักศึกษาจำนวน 15 คน

คนที่	คะแนนกลางเทอม(X)	คะแนนปลายเทอม(Y)
1	22	29
2	41	42
3	33	35
4	21	18
5	37	37
6	46	44
7	36	39
8	25	31
.	.	.
.	.	.

นำข้อมูลตัวอย่าง15 คู่ มาพล็อตแสดงแผนภาพกระจาย



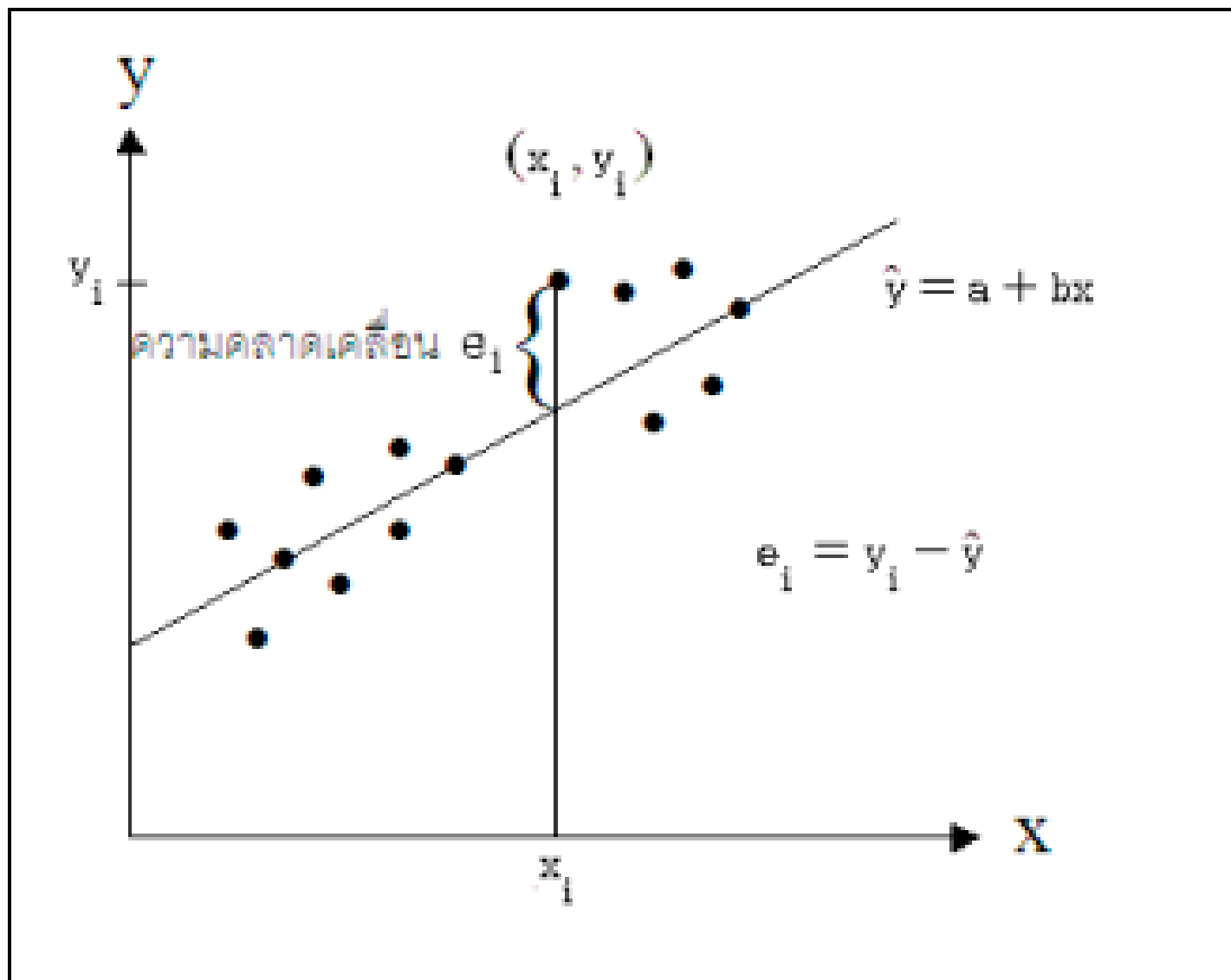
การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

(Simple Linear Regression)

- อาจมีการสร้างสมการได้หลายเส้น เพราะมีหลายวิธีการที่นำมาสร้างสมการเช่น สามารถเลือกจุดสองจุดด้านหัวและท้ายมาเป็นตัวแทนแล้วลากเส้นผ่าน ทำให้ทราบความสัมพันธ์ของสองตัวแปรได้อย่างคร่าว ๆ
- บางคนอาจลากเส้นผ่านเข้าไปในหมู่จุดที่ลงไว้ นั่นก็อาจมีได้หลายเส้น
- จึงหาวิธีที่จะได้เส้นถดถอยที่เหมาะสมที่สุดซึ่งควรจะมีเพียงเส้นเดียวเท่านั้น วิธีที่นิยมใช้มากที่สุดคือ วิธีกำลังสองน้อยที่สุด (method of least squares)

วิธีกำลังสองน้อยที่สุด

- มีหลักการว่า ต้องให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด
- ความคลาดเคลื่อน คือ ความแตกต่างหรือค่าเบี่ยงเบนในแนวตั้งของจุดในแผนภาพกระจายที่ห่างจากเส้นนั้นเอง แสดงได้ดังรูป



จากข้อมูลตัวอย่างสามารถสร้างสมการถดถอยเพื่อประมาณค่า $\underline{y}$ จากสมการถดถอยเชิงเส้น

$$\hat{Y} = a + bX$$

$$\hat{Y} = \text{ค่าประมาณของ } Y$$

$\underline{a}$ = ค่าจุดตัดบนแกน Y หรือเป็นค่า $\hat{Y}$ เมื่อ X มีค่าเป็นศูนย์

$\underline{b}$ = ความชันของสมการ หรือ อัตราการการเปลี่ยนแปลงของ $\hat{Y}$ เมื่อ X เปลี่ยนแปลงไป 1 หน่วย

$\underline{X}$ = ค่าตัวแปรอิสระที่กำหนดให้

การได้สมการที่ดีที่สุดเพื่อประมาณค่า Y

ให้ $e_i = Y_i - \hat{Y}$ เป็นความคลาดเคลื่อนในแต่ละจุดของ (X_i, Y_i)

ในวิธีกำลังสองน้อยที่สุดเป็นวิธีที่ดีที่สุดทำให้ความคลาดเคลื่อนน้อยที่สุดและมีคุณสมบัติดังนี้

$$\sum e_i = \sum (Y_i - \hat{Y}) = 0$$

และ $\sum e_i^2 = \sum (Y_i - \hat{Y})^2$ มีค่าน้อยที่สุด

วิธีกำลังสองน้อยที่สุด

ด้วยวิธีกำลังสองน้อยที่สุดและใช้หลักการทางคณิตศาสตร์ทำให้ได้ค่า a และ b

$$a = \bar{Y} - b\bar{X}$$

เมื่อ $\bar{Y} =$ ค่าเฉลี่ยของตัวแปร Y

$\bar{X} =$ ค่าเฉลี่ยของตัวแปร X

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

หรือ

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

เมื่อ $n =$ จำนวนคู่ของค่าสังเกต

ตัวอย่างที่ 5.3.1 จะสร้างสมการประมาณค่า Y

โจทย์ จากข้อมูลคะแนนสอบวิชา คณิต

บูรณาการกลางเทอมและปลายเทอมของ

นักศึกษา 15 คน จงสร้างสมการเพื่อ

ประมาณคะแนนสอบปลายเทอม จาก

คะแนนสอบกลางเทอม ถ้าหาก

นักศึกษาคคนหนึ่งสอบกลางเทอมได้

คะแนน 30 จงประมาณคะแนนสอบ

ปลายเทอมของเขา

คนที่	X	Y
1	22	29
2	41	42
3	33	35
4	21	18
5	37	37
6	46	44
7	36	39
8	25	31
9	28	30
10	26	25
11	43	40
12	29	32
13	30	31
14	20	29
15	32	35
รวม	469	497

คนที่	X	Y	X^2	Y^2	XY
1	22	29	484	841	638
2	41	42	1,681	1,764	1,722
3	33	35	1,089	1,225	1,155
4	21	18	441	324	378
5	37	37	1,369	1,369	1,369
6	46	44	2,116	1,936	2,024
7	36	39	1,296	1,521	1,404
8	25	31	625	961	775
9	28	30	784	900	840
10	26	25	676	625	650
11	43	40	1,849	1,600	1,720
12	29	32	841	1,024	928
13	30	31	900	961	930
14	20	29	400	841	580
15	32	35	1,024	1,225	1,120
รวม	469	497	15,575	17,117	16,233

ให้ X เป็นคะแนนสอบกลางเทอม

Y เป็นคะแนนสอบปลายเทอม

ในการสร้างสมการประมาณค่า Y โดยทำการหาค่า a และ b จากสูตรดังนี้

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

$$b = \frac{15 (16,233) - (469)(497)}{15 (15,575) - (469)^2}$$

$$\underline{b} = 0.76$$

ให้ X เป็นคะแนนสอบกลางเทอม

Y เป็นคะแนนสอบปลายเทอม

$$a = \bar{Y} - b\bar{X}$$

$$a = \frac{497}{15} - 0.76\left(\frac{469}{15}\right)$$

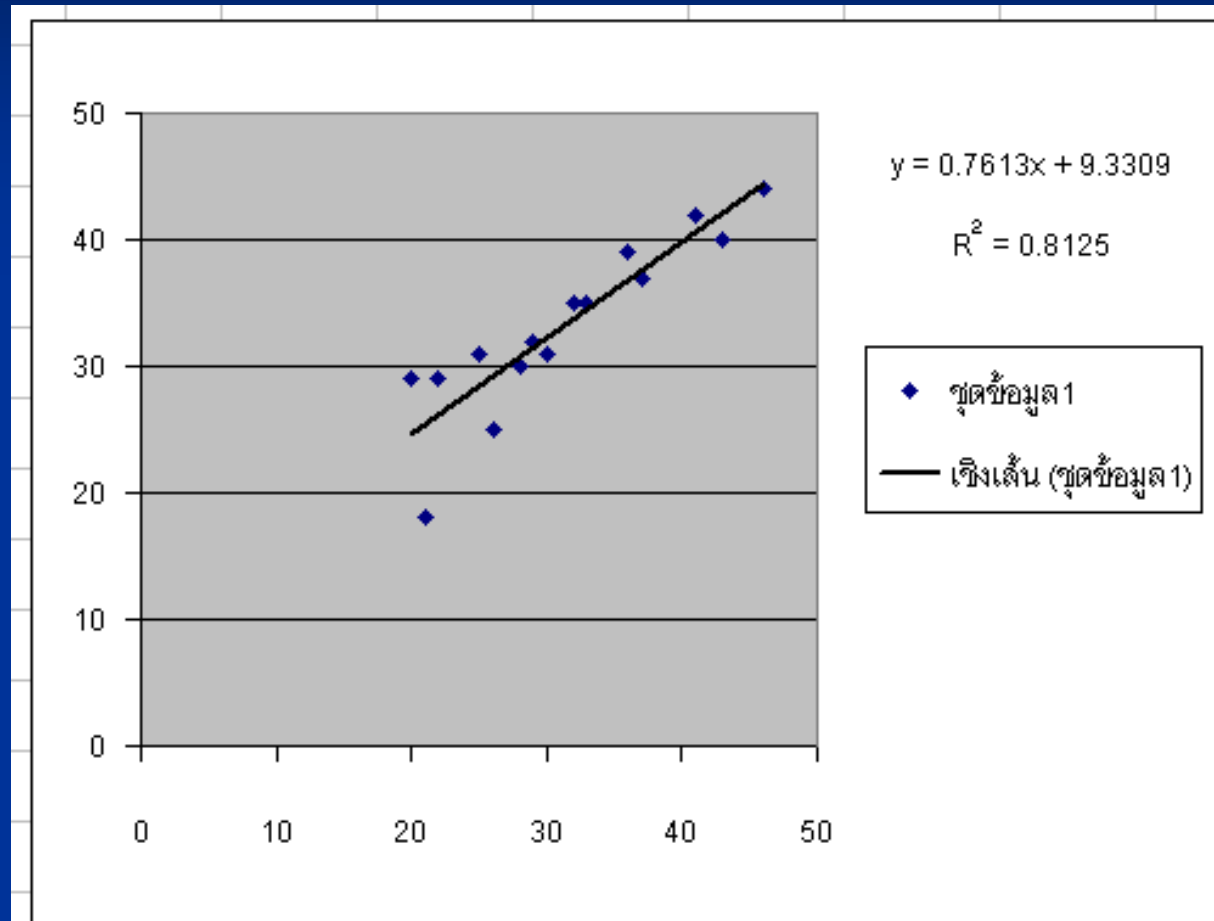
$$a = 9.33$$

ดังนั้นจะได้สมการ $\hat{Y} = 9.33 + 0.76X$

ถ้าหากนักศึกษาสอบกลางเทอมได้คะแนน 30 จะประมาณคะแนนสอบปลายเทอม

$$\hat{Y} = 9.33 + 0.76(30) = 32.13$$

ตัวอย่างการสร้างเส้นความสัมพันธ์จากExcel



ตัวอย่าง 5.3.2 ข้อมูลกำไรสุทธิ (หน่วย : ล้านบาท) ของบริษัทสยามก่อสร้างในช่วงปีพ.ศ. 2540 ถึง 2547
เป็นดังนี้

ปีพ.ศ.	2540	2541	2542	2543	2544	2545	2546	2547
กำไรสุทธิ	5.3	5.4	7.9	11.9	13.2	11.6	11.9	12.8

จงสร้างสมการถดถอยเพื่อประมาณกำไรสุทธิในปีพ.ศ. 2548

วิธีทำ จากข้อมูลข้างต้น ให้ X เป็นปีที่ และ Y เป็นกำไรสุทธิ

ปีพ.ศ.	ปีที่ (X)	กำไรสุทธิ (Y)	XY	X ²
2540	1	5.3	5.3	1
2541	2	5.4	10.8	4
2542	3	7.9	23.7	9
2543	4	11.9	47.6	16
2544	5	13.2	66.0	25
2545	6	11.6	69.6	36
2546	7	11.9	83.3	49
2547	8	12.8	102.4	64
รวม	36	80.0	408.7	204

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

$$b = \frac{8(408.7) - (36)(80)}{8(204) - (36)^2} = 1.160$$

$$a = \bar{Y} - b\bar{X} \quad a = 10 - 1.160(4.5) = 4.782$$

ดังนั้นจะได้สมการถดถอยคือ $\hat{Y} = 4.782 + 1.160X$

ประมาณกำไรสุทธิในปีที่ 9 (พ.ศ. 2548)

$$\hat{Y} = 4.782 + 1.160(9)$$

$$\hat{Y} = 4.782 + 10.44 = 15.222$$

หัวข้อ : สัมประสิทธิ์สหสัมพันธ์

จุดประสงค์ของเนื้อหา

- ต้องการหาระดับความสัมพันธ์ของสองตัวแปรว่ามีความสัมพันธ์กันมาก น้อยเพียงใด

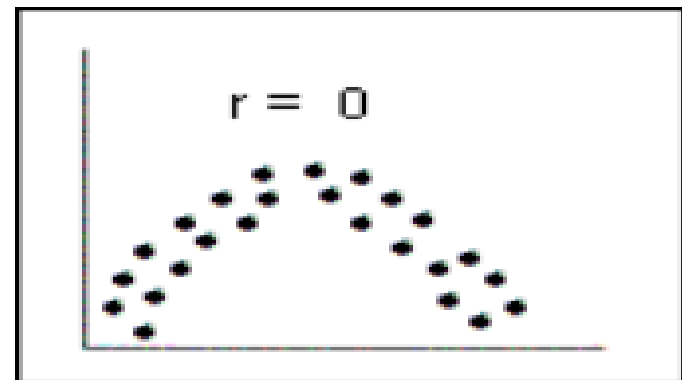
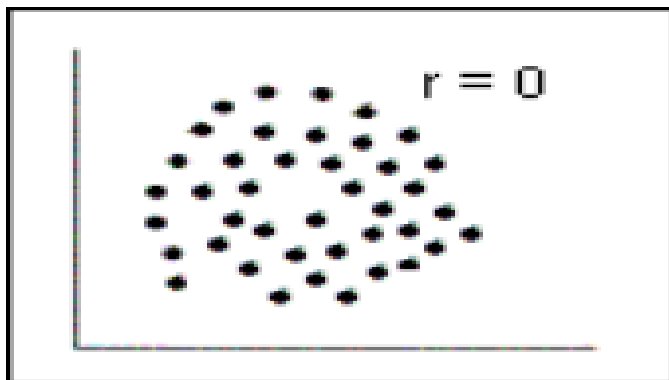
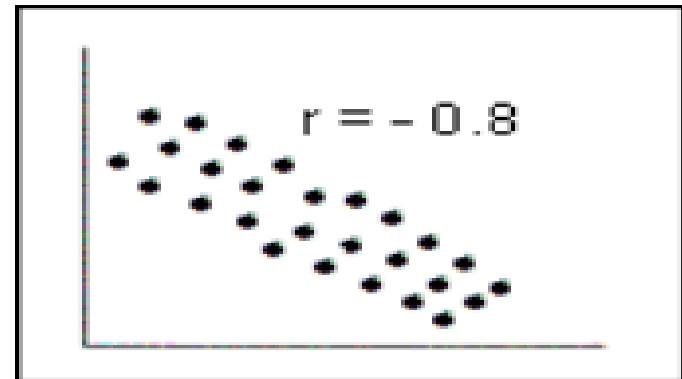
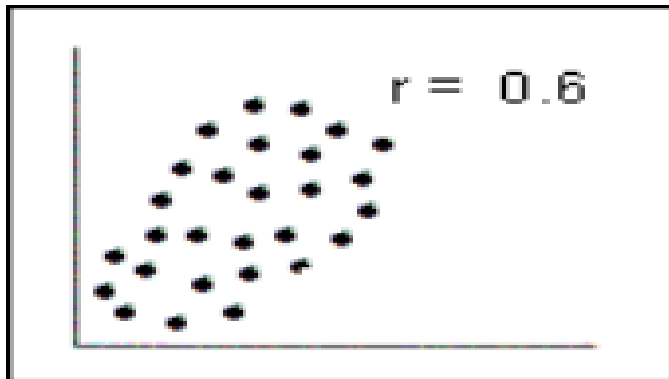
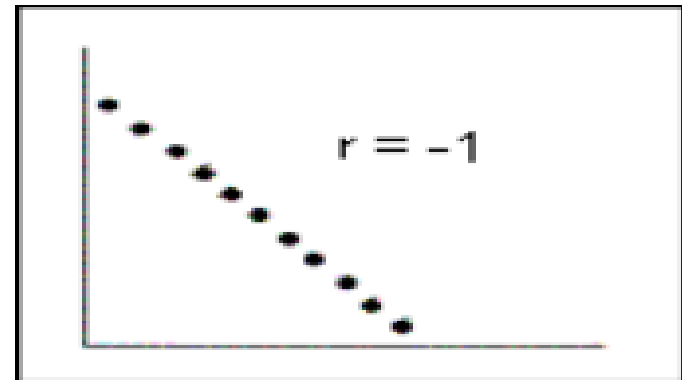
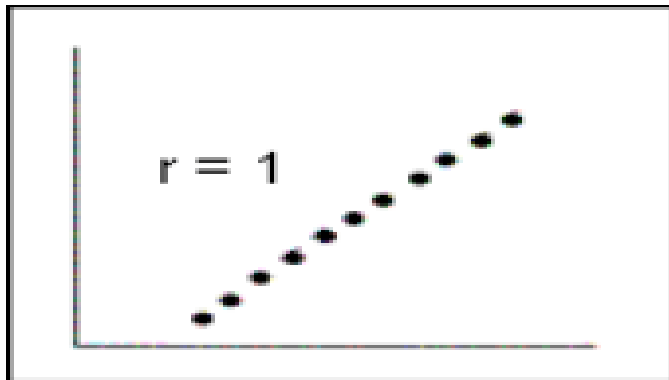
การหาระดับความสัมพันธ์ของสองตัวแปร

- ถ้าตัวแปรคู่ (X,Y) มีความสัมพันธ์กันมาก การคาดคะเนค่าตัวแปรตาม (Y) ก็จะใกล้เคียงกับความจริง
- ในหัวข้อนี้จึงเป็นการศึกษาหาขนาดความสัมพันธ์เชิงเส้นของตัวแปรคู่ ค่าวัดขนาดความสัมพันธ์เชิงเส้นของตัวแปรคู่นี้จะเรียกว่า *สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)*

ความหมายของสัมประสิทธิ์สหสัมพันธ์

ให้ r แทนสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่เชิงปริมาณ (X, Y) จำนวน n คู่
ค่า r มีค่าอยู่ระหว่าง -1 กับ 1 และสรุปความหมายของค่า r ได้ดังนี้

- 1. ถ้า $r = 0$ ไม่มีความสัมพันธ์เชิงเส้นระหว่าง X กับ Y
- 2. ถ้า $r = 1$ มีความสัมพันธ์เชิงเส้นในทางบวกโดยสมบูรณ์
นั่นคือถ้า X มีค่าเพิ่มขึ้น Y จะเพิ่มขึ้นตาม และถ้าหาก X มีค่าลดลง Y จะลดลงตาม
- 3. ถ้า $r = -1$ แสดงว่ามีความสัมพันธ์เชิงเส้นในทางลบโดยสมบูรณ์
นั่นคือถ้า X มีค่าเพิ่มขึ้น Y จะมีค่าลดลง และถ้าหาก X มีค่าลดลง Y จะมีค่าเพิ่มขึ้น
- 4. ถ้า $0 < r < 1$ แสดงว่า X กับ Y มีความสัมพันธ์เชิงบวก
แต่ถ้า $-1 < r < 0$ แสดงว่า X กับ Y มีความสัมพันธ์เชิงลบ
ถ้าค่า r เข้าใกล้ศูนย์ เช่น $r = \pm 0.2$ แสดงว่า X กับ Y มีความสัมพันธ์กันน้อยมาก



การหาค่าสัมประสิทธิ์สหสัมพันธ์

การหาค่า r โดยวิธีการ Pearson product - moment correlation coefficient
หาได้จากสูตรต่อไปนี้

$$\text{สูตร (1)} \quad r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

$$\text{สูตร (2)} \quad r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

ตัวอย่าง จากข้อมูลคะแนนสอบกลางเทอมและปลายเทอมวิชาคณิตศาสตร์
บูรณาการของนักศึกษา 15 คน ตามตัวอย่าง 5.3.1 จงหาค่าสัมประสิทธิ์
สหสัมพันธ์พร้อมทั้งอธิบายค่าที่หาได้

■ วิธีทำ

$$\sum X = 469$$

$$\sum Y = 497$$

$$\sum X^2 = 15,575$$

$$\sum Y^2 = 17,117$$

$$\sum XY = 16,233$$

$$r = \frac{15(16,233) - 469(497)}{\sqrt{[15(15,575) - (469)^2][15(17,117) - (497)^2]}}$$

$$r = \frac{10,402}{\sqrt{133,169,344}} = 0.90$$

ตัวอย่าง นักวิจัยท่านหนึ่งศึกษาความสัมพันธ์ระหว่างอายุ กับอัตราการเต้นของหัวใจสูงสุดขณะออกกำลังกายของผู้หญิง เขาจึงสุ่มผู้หญิงที่ออกกำลังกายจำนวน 5 คน ได้ข้อมูลดังนี้

คนที่	1	2	3	4	5
อายุ (ปี)	23	39	19	44	51
อัตราการเต้นของหัวใจสูงสุด (ครั้งต่อนาที)	210	185	220	164	123

จงหาระดับความสัมพันธ์ระหว่างอายุ กับอัตราการเต้นของหัวใจสูงสุดขณะออกกำลังกายของผู้หญิงกลุ่มนี้

คนที่	X (อายุ)	Y(อัตราการเต้น หัวใจ)	XY	X^2	Y^2
1	23	210	4,830	529	44,100
2	39	185	7,215	1,521	34,225
3	19	220	4,180	361	48,400
4	44	164	7,216	1,936	26,896
5	51	123	6,273	2,601	15,129
รวม	176	902	29,714	6,948	168,750

แทนค่า $\sum X = 176$ $\sum Y = 902$

$$\sum X^2 = 6,948 \quad \sum Y^2 = 168,750 \quad \sum XY = 29,714$$

แทนค่าต่าง ๆ ในสูตรของ r

$$r = \frac{5(29,714) - 176(902)}{\sqrt{[5(6,948) - (176)^2][5(168,750) - (902)^2]}}$$

$$r = \frac{-10,182}{\sqrt{113,469,544}}$$

$$r = -0.96$$

กรณีข้อมูลมีลักษณะเป็นตัวแปรคุณภาพหรือตัวแปรกลุ่ม

- จำแนกข้อมูลลงตารางไขว้หรือตารางสองทางได้
- เช่นต้องการหาระดับความสัมพันธ์ของระดับ I.Q. กับ ความสำเร็จในการทำงาน
- เพศของผู้ซื้อรถยนต์ กับ สีของรถยนต์ที่เลือกซื้อ
- จำนวนสินค้าที่ชำรุด กับ ยี่ห้อของเครื่องจักรที่ใช้ผลิต

การหาระดับความสัมพันธ์จะหาสัมประสิทธิ์สหสัมพันธ์โดยวิธีของ Cramer H. มีชื่อเรียกว่า Cramer' s Contingency Coefficient (V)

ตัวอย่างข้อมูลกลุ่ม

ตารางข้อมูลการสำรวจ ระดับ I.Q. กับความสำเร็จในการทำงานของพนักงานในบริษัทหนึ่ง 432 คน

ระดับ I.Q.	ความสำเร็จในการทำงาน			รวม
	ดีมาก	พอใช้	ไม่ดี	
ระดับสูง	41	12	26	79
ระดับปานกลาง	84	61	48	193
ระดับต่ำ	50	49	61	160
รวม	175	122	135	432

แสดงการคำนวณหาค่าความสัมพันธ์

ให้ v เป็นสัญลักษณ์ แทนสัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และตัวแปร Y

$$v = \sqrt{\frac{\chi^2}{N \times \min(r - 1, c - 1)}}$$

โดยที่ χ^2 เป็นค่าสถิติได้จากสูตร $\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

N เป็นความถี่รวม หรือจำนวนตัวอย่างทั้งหมด

r เป็นจำนวนแถวทั้งหมด

c เป็นจำนวนสดมภ์ทั้งหมด

$\min(r-1, c-1)$ เป็นค่าต่ำสุดของค่า $r-1$ และ $c-1$

การหาค่า Eij

คุณลักษณะ X	คุณลักษณะ Y			รวม
	Y ₁	Y ₂	Y ₃	
X ₁	O ₁₁	O ₁₂	O ₁₃	R ₁
X ₂	O ₂₁	O ₂₂	O ₂₃	R ₂
X ₃	O ₃₁	O ₃₂	O ₃₃	R ₃
รวม	C ₁	C ₂	C ₃	N

ค่า E_{ij} หาได้จากสูตร

เช่น
$$E_{11} = \frac{R_1 \times C_1}{N}$$

$$E_{12} = \frac{R_1 \times C_2}{N}$$

$$E_{21} = \frac{R_2 \times C_1}{N}$$

$$E_{22} = \frac{R_2 \times C_2}{N}$$

คุณลักษณะ X	คุณลักษณะ Y			รวม
	Y_1	Y_2	Y_3	
X_1	E_{11}	E_{12}	E_{13}	R_1
X_2	E_{21}	E_{22}	E_{23}	R_2
X_3	E_{31}	E_{32}	E_{33}	R_3
รวม	C_1	C_2	C_3	N

เมื่อหาค่า E_{ij} ได้ครบทุกช่องแล้วนำไปแทนค่าในสูตร
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

จะได้ค่า χ^2 แล้วแทนค่าในสูตร

$$v = \sqrt{\frac{\chi^2}{N \times \min(r - 1, c - 1)}}$$

คำนวณค่าความถี่คาดหวังต่าง ๆ เช่น

$$E_{11} = \frac{175 \times 79}{432} = 32.00$$

$$E_{12} = \frac{122 \times 79}{432} = 22.31$$

$$E_{21} = \frac{175 \times 193}{432} = 78.18$$

$$E_{22} = \frac{122 \times 193}{432} = 54.54$$

$$E_{13} = 79 - 32 - 22.31 = 24.69$$

$$E_{23} = 193 - 78.19 - 54.50 = 60.31$$

$$E_{31} = 175 - 32 - 78.19 = 64.81$$

$$E_{32} = 122 - 22.31 - 54.50 = 45.19$$

$$E_{33} = 160 - 64.81 - 45.19 = 50.00$$

หาค่าความถี่คาดหวัง(E_{ij})

หาค่า E_{ij} ทุก ๆ ค่าจะได้ดังนี้

E_{11} (32.00)	E_{12} (22.31)	E_{13} (24.69)	R_1 (79)
E_{21} (78.18)	E_{22} (54.50)	E_{23} (60.31)	R_2 (193)
E_{31} (64.81)	E_{32} (45.19)	E_{33} (50.00)	R_3 (160)
C_1 (175)	C_2 (122)	C_3 (135)	N (432)

นำค่า O_{ij} และ E_{ij} แทนในสูตร χ^2 แสดงการคำนวณได้ดังนี้

O_{ij}	E_{ij}	$(O_{ij}-E_{ij})$	$(O_{ij}-E_{ij})^2$	$(O_{ij}-E_{ij})^2/E_{ij}$
41	32	9	81.00	2.53
84	78.19	5.81	33.76	0.43
50	64.81	-14.81	219.34	3.38
12	22.31	-10.31	106.30	4.76
61	54.5	6.5	42.25	0.78
49	45.19	3.81	14.52	0.32
26	24.69	1.31	1.72	0.07
48	60.31	-12.31	151.54	2.51
61	50	11	121.00	2.42
				17.21

จากการคำนวณได้ค่า $\chi^2 = 17.21$ $N=432$ $r-1 = 2$, $c-1 = 2$

หาค่าสัมประสิทธิ์สหสัมพันธ์

$$v = \sqrt{\frac{\chi^2}{N \times \min(r-1, c-1)}}$$

$$v = \sqrt{\frac{17.21}{432 \times 2}} = 0.14$$

สรุป

การวิเคราะห์ข้อมูลที่มีลักษณะเป็นตัวแปรคู่

ก. ข้อมูลปริมาณ

- สร้างสมการแสดงความสัมพันธ์ได้
- บอกระดับความสัมพันธ์ของข้อมูลได้ (หาค่า r)

ข. ข้อมูลคุณภาพ

- บอกระดับความสัมพันธ์ของข้อมูลได้ (หาค่า V)