

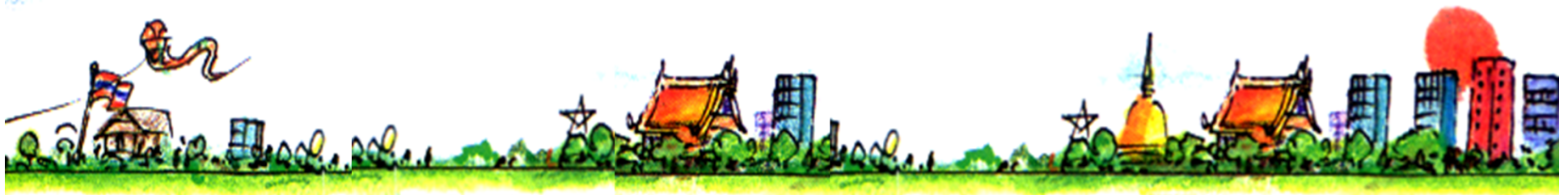
กระบวนการแก้ปัญหาโดยใช้สถิติ



รองศาสตราจารย์พิชญ์ เจียวกุณ
ภาควิชาสถิติ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่

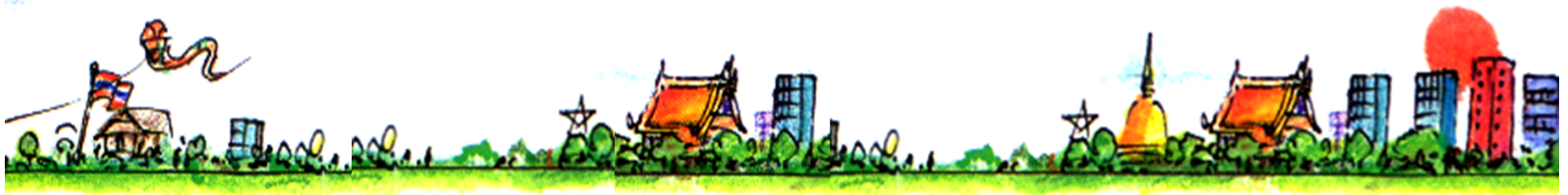
สถิติสำคัญอย่างไร

- ✓ สถิติเป็นศาสตร์ที่ว่าด้วยการจัดการข้อมูล ประกอบด้วย การเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการแปลความหมายของข้อมูล
- ✓ เป็นเครื่องมือที่ใช้ในการตัดสินใจ โดยใช้ความน่าจะเป็น



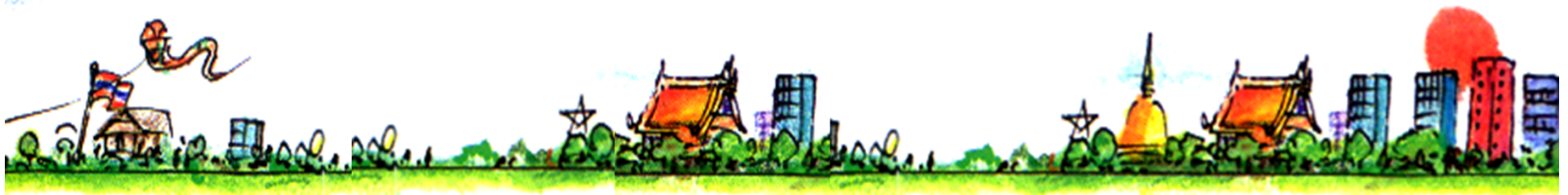
สถิติแก้ปัญหาได้อย่างไร

- ✓ สถิติเป็นศาสตร์ที่ว่าด้วยการจัดการข้อมูล ประกอบด้วย การเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการแปลความหมายของข้อมูล
- ✓ การมีข้อมูลอยู่ในมือมาก ย่อมสามารถหาสารสนเทศจากข้อมูลนั้นได้มาก



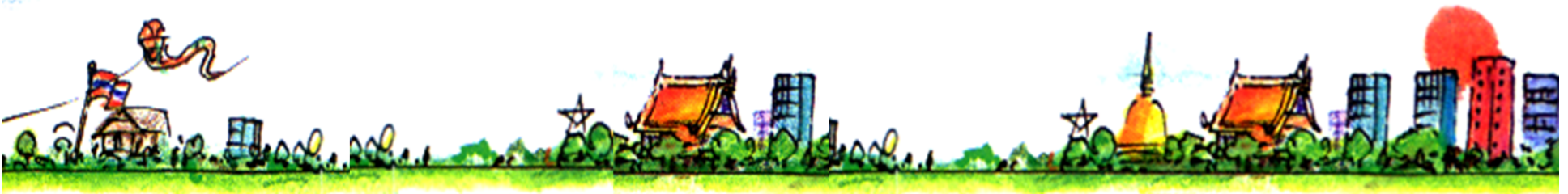
วิธีการทางสถิติ (Statistical Method)

คือการใช้สถิติ เป็นเครื่องมือในการแก้ปัญหา
หรือวิเคราะห์ข้อมูลเพื่อหาคำตอบในสิ่งที่สนใจ
ศึกษา วิธีการทางสถิติที่น่าสนใจเช่น การ
วิเคราะห์การถดถอย (Regression Analysis)
โดยได้**สมการแสดงความสัมพันธ์ระหว่างตัวแปร**



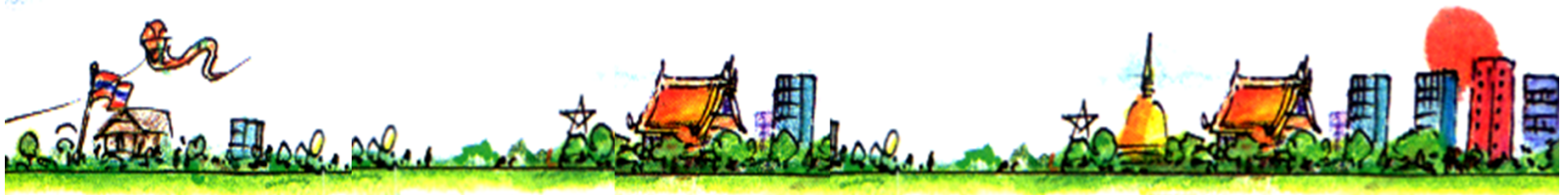
สมการความสัมพันธ์ของสองตัวแปร

- เป็นการศึกษาความสัมพันธ์ของข้อมูลเชิงปริมาณ สองชุด (หรือสองตัวแปร) ในลักษณะตัวแปรคู่ (Bivariate data)
- ตัวแปรหนึ่งเป็น**ตัวแปรอิสระ** และอีกตัวหนึ่งเป็น**ตัวแปรตาม**



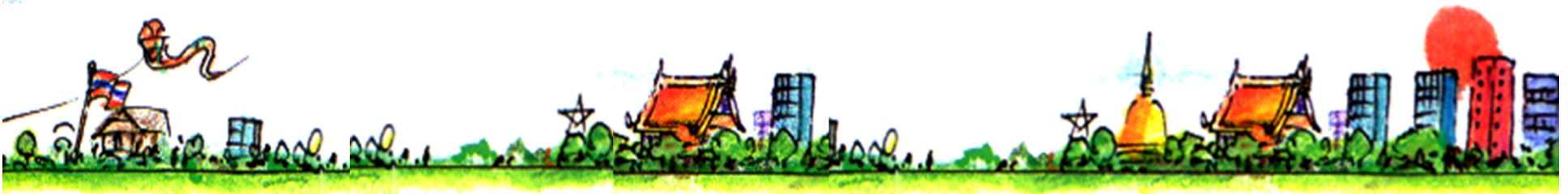
ความหมายของตัวแปรอิสระและตัวแปรตาม

- ตัวแปรอิสระ (Independent Variable)
หมายถึงตัวแปรที่มีอิทธิพลต่อตัวแปรอีกตัว
หนึ่ง
- ตัวแปรตาม (Dependent Variable)
หมายถึงตัวแปรที่ถูกกำหนดโดยตัวแปรอิสระ



ตัวอย่างความสัมพันธ์ระหว่างตัวแปร

- ผลผลิตข้าวต่อไร่ กับ ปริมาณปุ๋ยที่ใส่ต่อไร่
- ยอดขายสินค้ากับรายจ่ายค่าโฆษณาทางโทรทัศน์
- รายได้กับเงินออม
- คะแนนสอบกลางเทอมกับคะแนนสอบไล่



ประเภทของการวิเคราะห์การถดถอย

การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

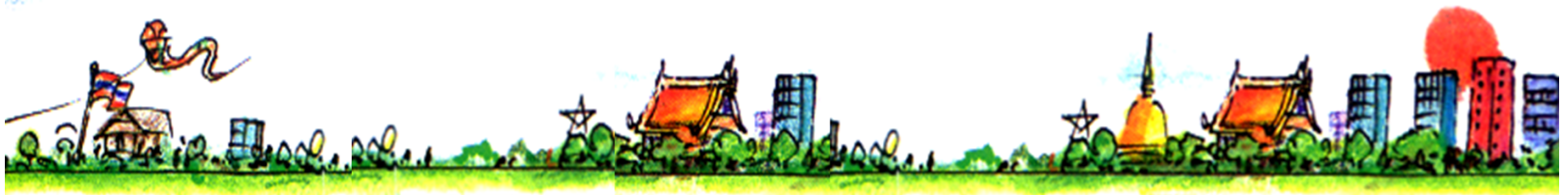
หมายถึงการวิเคราะห์ที่ประกอบด้วยตัวแปรอิสระหนึ่งตัว ตัวแปรตามหนึ่งตัว



ประเภทของการวิเคราะห์การถดถอย

การถดถอยเชิงเส้นพหุคูณ (Multiple
Linear Regression

หมายถึงการวิเคราะห์ที่ประกอบด้วยตัวแปร
อิสระมากกว่าหนึ่งตัว ตัวแปรตามหนึ่งตัว



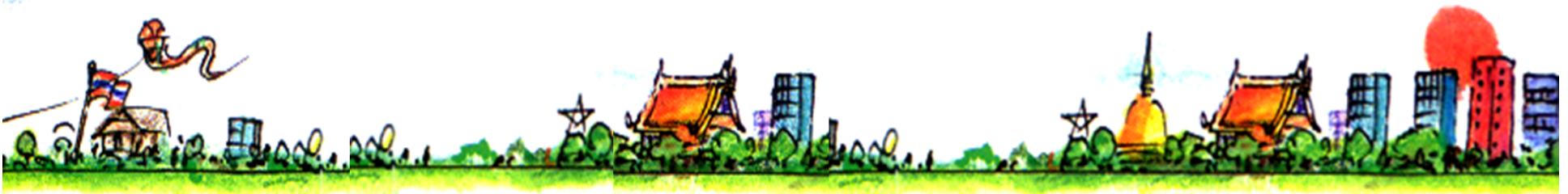
สมการถดถอย

สมการถดถอยของประชากร

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

สมการถดถอยของตัวอย่าง

$$\hat{Y} = a + b x$$

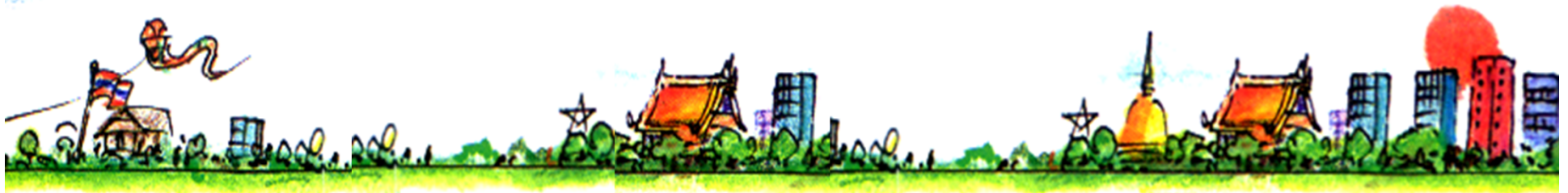


ความหมายของค่าในสมการ

- a หมายถึง ค่าระยะตัดแกน Y
- b หมายถึง ค่าความชัน หรือสัมประสิทธิ์

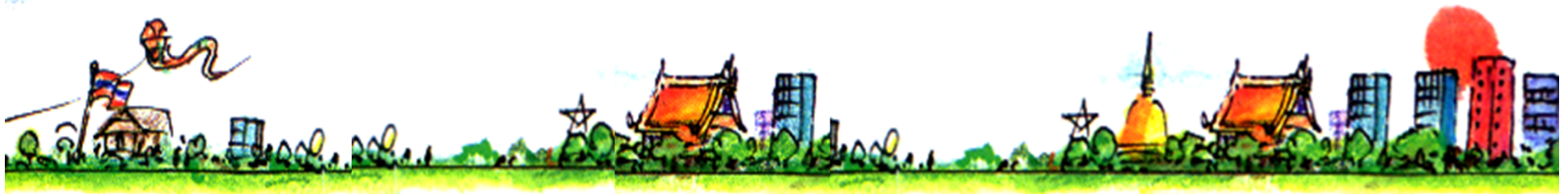
การถดถอย (Regression Coefficient)

ซึ่งอาจจะมีค่าเป็นบวก ลบ หรือศูนย์และหมายถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป 1 หน่วย



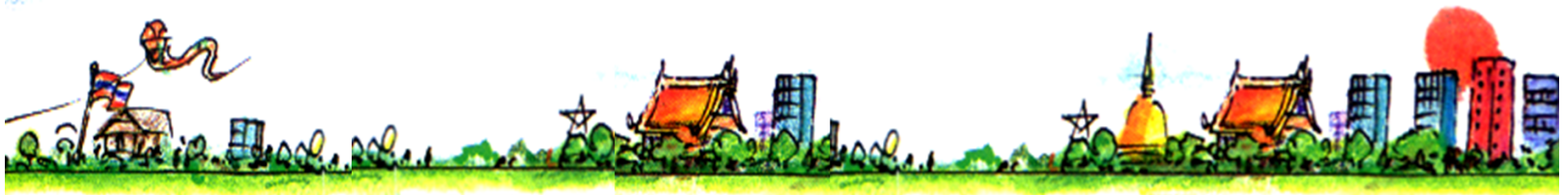
ความหมายของค่าความชัน

ค่า b มีค่าเป็นบวก หมายถึง
ความสัมพันธ์ระหว่าง ตัวแปรอิสระ(X)
และตัวแปรตาม (Y) มีความสัมพันธ์ใน
เชิงบวก มีทิศทางเดียวกัน



ความหมายของค่าความชัน

ค่า b มีค่าเป็นลบ หมายถึงความสัมพันธ์
ระหว่าง ตัวแปรอิสระ(X) และตัวแปรตาม
(Y) มีความสัมพันธ์ในเชิงลบ มีทิศทาง
ตรงกันข้าม



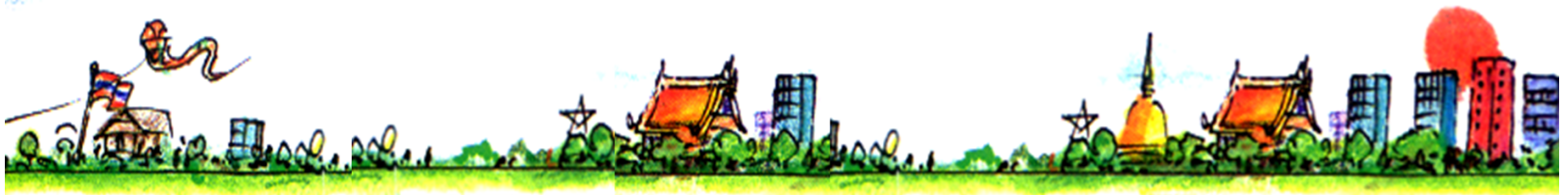
ความหมายของค่าความชัน

ค่า b มีค่าเป็นศูนย์ หมายถึง
ความสัมพันธ์ระหว่าง ตัวแปรอิสระ(X)
และตัวแปรตาม (Y) ไม่มีความสัมพันธ์
กันในลักษณะเชิงเส้น

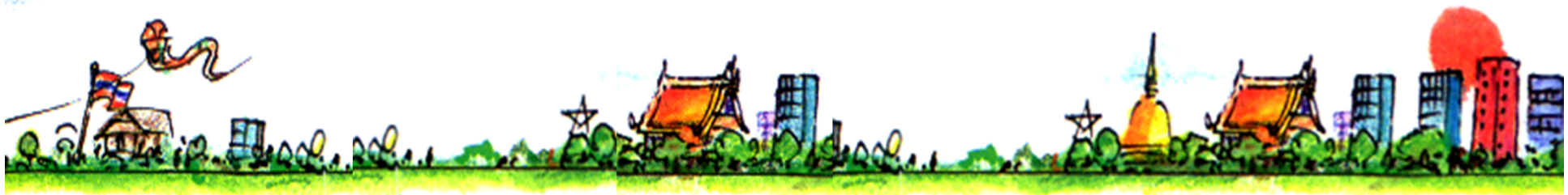
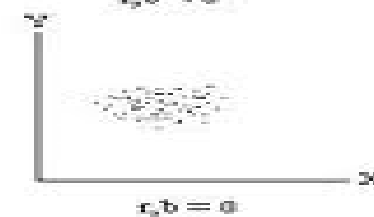
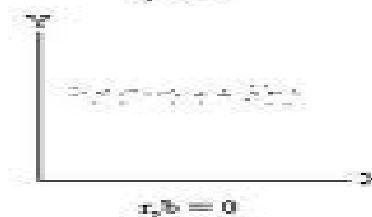
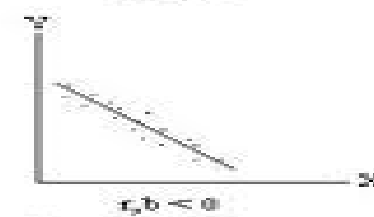
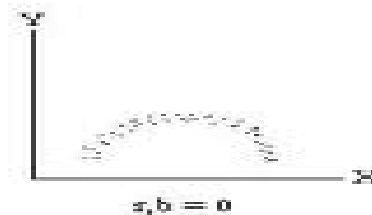
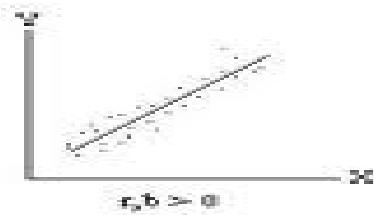


ความหมายของค่า a

ค่า a คือค่าระยะตัดแกน Y หรือระยะ Y intercept แสดงค่าของ Y เมื่อกำหนดให้ค่า $X=0$



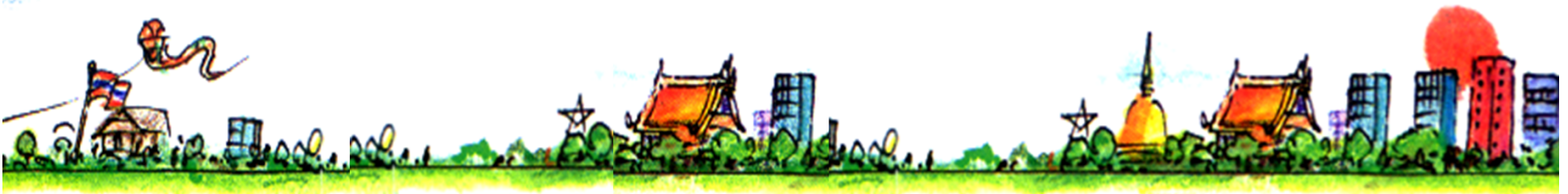
ความสัมพันธ์ระหว่างตัวแปร



การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

จะใช้วิธีกำลังสองน้อยที่สุด (Least squares method) ซึ่งมีหลักการคือให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด นั่นคือ

$$\sum e_i^2 = \sum (y_i - \hat{y})^2 \quad \text{มีค่าน้อยที่สุด}$$



การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

ใช้วิธีการหาอนุพันธ์อันดับที่ 1 เทียบกับค่า a
และ b แล้วให้เท่ากับ 0 จะได้ว่า

$$\sum e_i^2 = \sum (y_i - \hat{y})^2 = Q$$

$$\frac{\partial Q}{\partial b_0} = 0 \quad \frac{\partial Q}{\partial b_1} = 0$$

แก้สมการจะได้ว่า



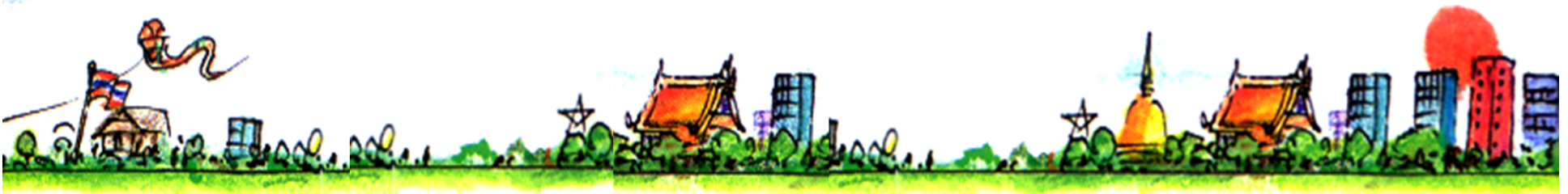
การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

$$a = \bar{y} - b\bar{x}$$

หรือ

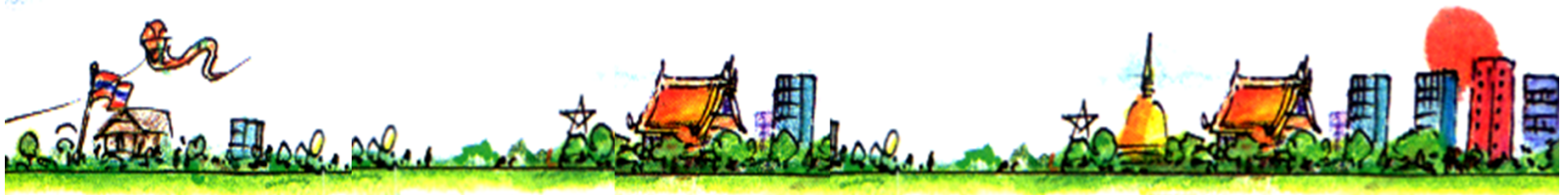
$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$



ตัวอย่างที่ 1

นักวิจัยต้องการหาความสัมพันธ์ระหว่างอายุการใช้
งานของเครื่องถ่ายเอกสาร(ปี) กับค่าใช้จ่ายใน
การบำรุงรักษาต่อเดือน จึงรวบรวมข้อมูลเครื่อง
ถ่ายเอกสารและค่าใช้จ่ายในการบำรุงรักษา ได้
ข้อมูลดังนี้



ตัวอย่างที่ 1

เครื่องถ่ายเอกสาร	อายุการใช้งาน (ปี)	ค่าบำรุงรักษา (พันบาท)
A	1	62
B	2	78
C	3	70
D	4	90
E	4	93
F	6	103



ตัวอย่างที่ 1

เครื่องถ่ายภาพเอกสาร	อายุการใช้งาน (ปี) X	ค่าบำรุงรักษา (พันบาท) Y
A	1	62
B	2	78
C	3	70
D	4	90
E	4	93
F	6	103



ตัวอย่าง

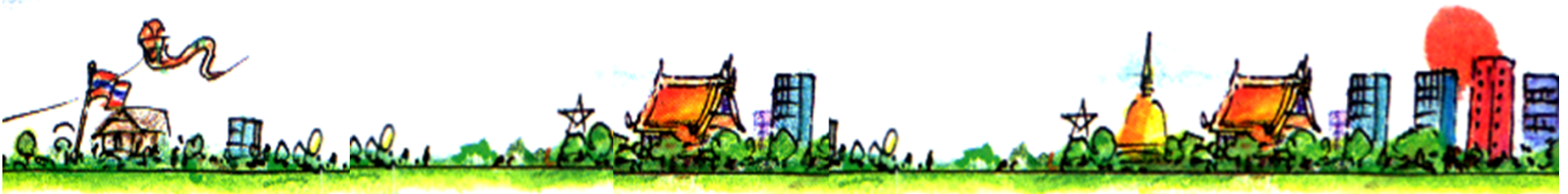
กำหนดสมการแสดงความสัมพันธ์

$$\hat{y} = a + bx$$

หาค่า a และ b จากความสัมพันธ์

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$



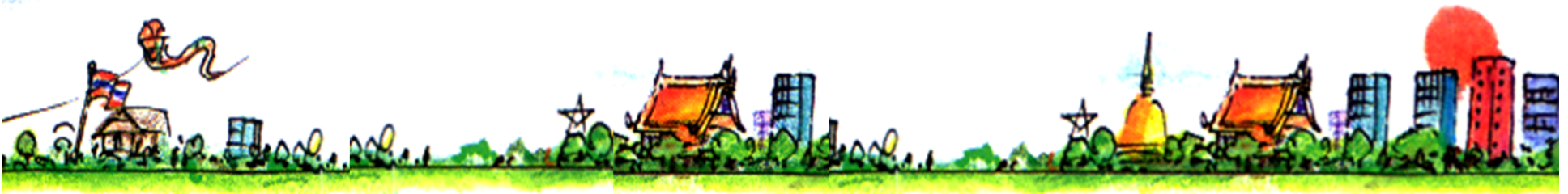
ตัวอย่าง

คำนวณค่า

$$\sum x = 20, \sum y = 496$$

$$\sum x^2 = 82, \sum y^2 = 42,186$$

$$\sum xy = 1,778$$



การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

แทนค่าลงในสมการ

$$a = \bar{y} - b\bar{x}$$

และ

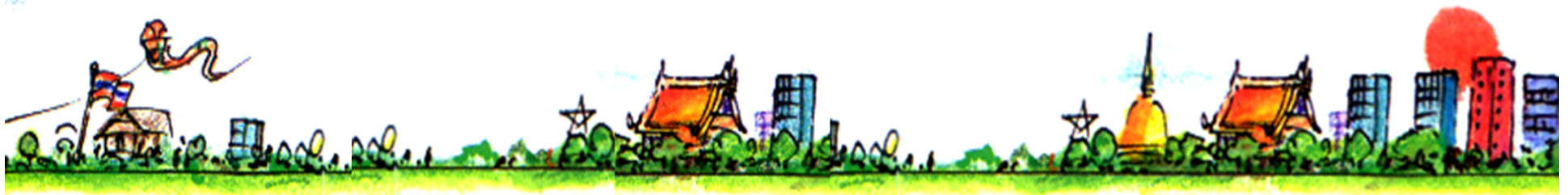
$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$



การสร้างสมการถดถอยเชิงเส้นอย่างง่าย
แทนค่าลงในสมการ

$$b = \frac{6(1,778) - (20)(496)}{6(82) - 400} = 8.13$$

$$a = 82.66 - 8.13(3.33) = 55.58$$

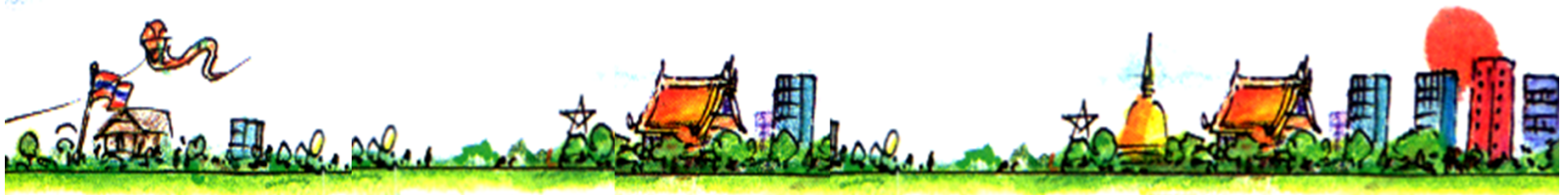


การสร้างสมการถดถอยเชิงเส้นอย่างง่าย
จะได้สมการถดถอยเชิงเส้นอย่างง่ายคือ

$$\hat{y} = 55.58 + 8.13x$$

ถ้าต้องการพยากรณ์ค่าใช้จ่ายเมื่อเครื่องถ่าย
เอกสารมีอายุ 8 ปี 6 เดือน

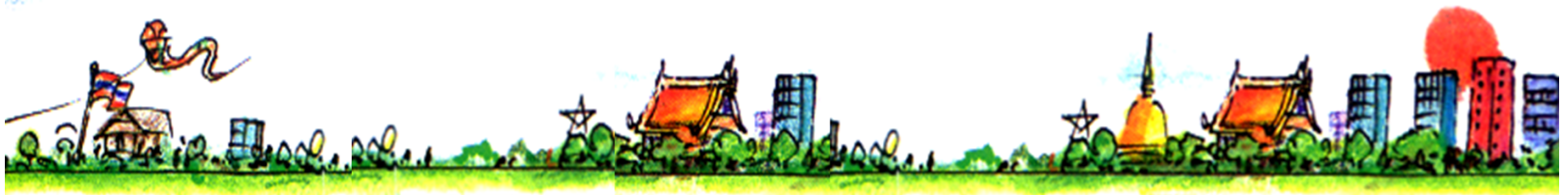
$$\hat{y} = 55.58 + 8.13(8.5) = 124.68$$



สัมประสิทธิ์สหสัมพันธ์ของสองตัวแปร

Correlation Coefficient : (r)

ข้อมูลเชิงปริมาณที่เก็บมาในลักษณะที่เป็นตัวแปรคู่ (Bivariate data) สามารถนำมาค่า r โดยใช้ความสัมพันธ์ของ Pearson product – moment correlation coefficient ดังนี้



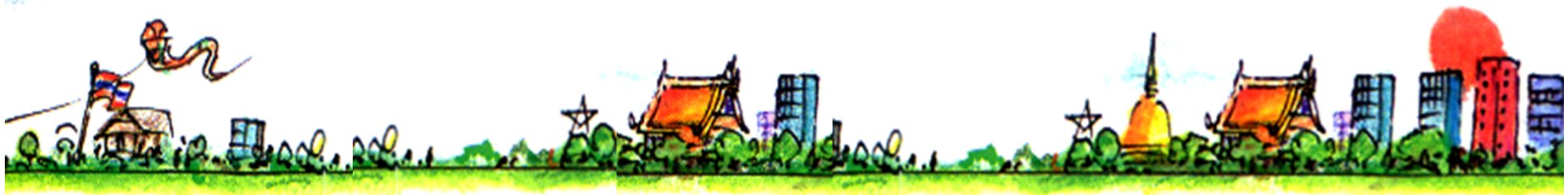
สัมประสิทธิ์สหสัมพันธ์ของสองตัวแปร

Correlation Coefficient : (r)

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

หรือ

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$



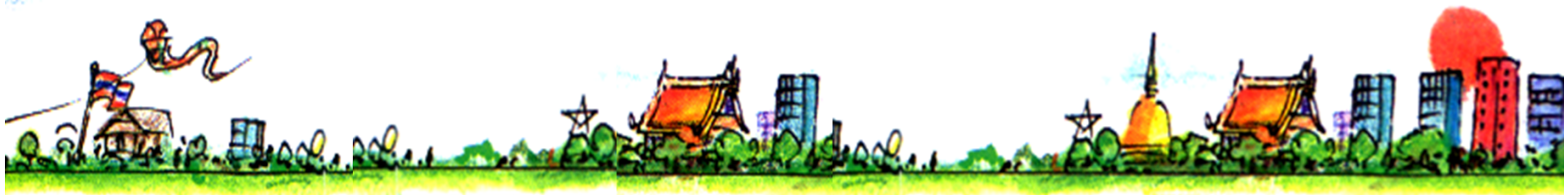
ความหมายสัมประสิทธิ์สหสัมพันธ์ (r)

ถ้า $r = 0$ แสดงว่าไม่มีความสัมพันธ์กันระหว่างตัวแปร X และ Y

ถ้า $r = 1$ แสดงว่ามีความสัมพันธ์กันระหว่างตัวแปร X และ Y ในทิศทางเดียวกัน

ถ้า $r = -1$ แสดงว่ามีความสัมพันธ์กันระหว่างตัวแปร X และ Y ในทิศทางตรงกันข้าม

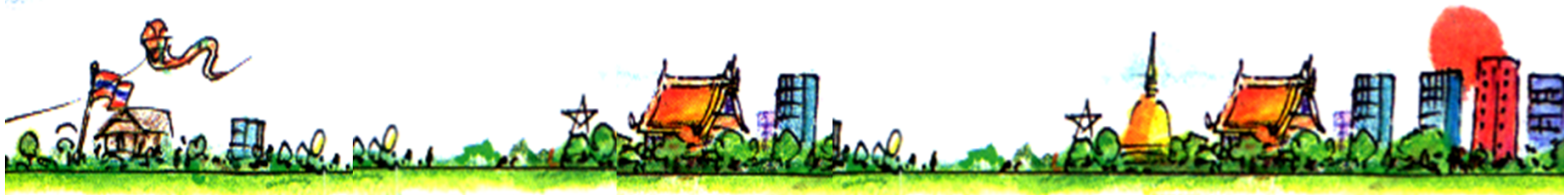
ค่า r มีค่าอยู่ระหว่าง -1 ถึง 1 เท่านั้น



จากข้อมูลตัวอย่างที่ 1 หาค่า r ได้ดังนี้

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

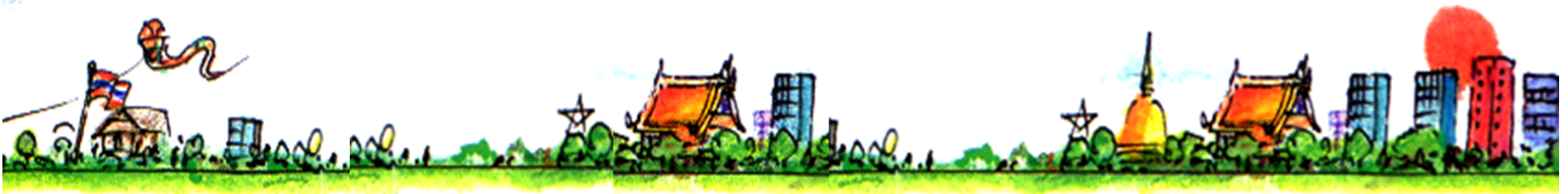
$$r = \frac{6(1,778) - 20(496)}{\sqrt{[6(82) - 400][6(42,186) - 496^2]}} = 0.925$$



ความหมายสัมประสิทธิ์สหสัมพันธ์ (r)

ค่า r มีค่าเท่ากับ 0.925 มีค่าเป็นบวกและมีค่า
เข้าใกล้ 1 แสดงว่าตัวแปรอิสระ X และตัวแปร
ตาม Y มีความสัมพันธ์กันในระดับสูงมาก และมี
ทิศทางเดียวกัน นั่นคือ

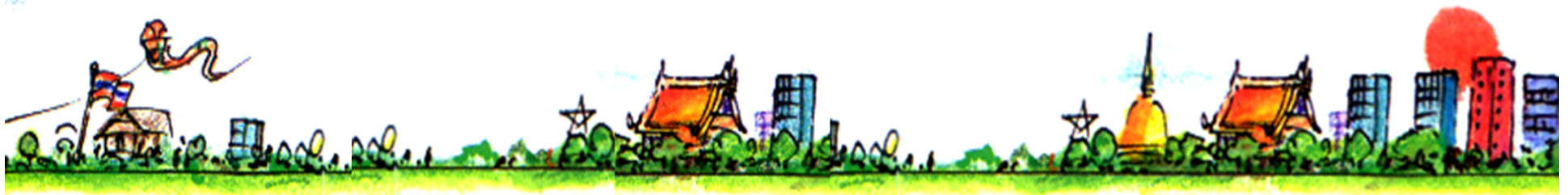
(X เพิ่ม Y เพิ่ม หรือ X ลด Y ลด)



สัมประสิทธิ์การตัดสินใจ (r^2)

คือการนำค่า r ยกกำลังสอง เรียกอีกอย่างหนึ่งว่า
สัมประสิทธิ์การตัดสินใจ (Coefficient of
Determination)

ค่า (r^2) มีค่าอยู่ระหว่าง 0 ถึง 1 มักจะทำให้
อยู่ในรูปของร้อยละ



สัมประสิทธิ์การตัดสินใจ (r^2)

ตัวอย่างเช่น

ค่า $r = 0.80$ ดังนั้นค่า (r^2)

จะมีค่าเท่ากับ 0.64 หรือ 64% หมายความว่า ตัวแปรอิสระ X สามารถอธิบายตัวแปรตาม Y ได้ร้อยละ 64 ที่เหลืออีกร้อยละ 36 เกิดจากอิทธิพลอื่น ซึ่งไม่ทราบสาเหตุ

