

“...การวิเคราะห์การถดถอยเป็นเครื่องมือทางสถิติที่นำมาใช้เพื่อศึกษาและสร้างรูปแบบความสัมพันธ์ระหว่างตัวแปร 2 ตัว หรือมากกว่า 2 ตัว...”

การนำโปรแกรมเอ็กเซล (Excel) และโปรแกรม ในอินเทอร์เน็ต (Internet) มาใช้ในการวิเคราะห์ การถดถอย (Regression Analysis)

นันทนา ณ ระนอง *

การวิเคราะห์การถดถอยเป็นเครื่องมือทางสถิติที่นำมาใช้เพื่อศึกษาและสร้างรูปแบบความสัมพันธ์ระหว่างตัวแปร 2 ตัว หรือมากกว่า 2 ตัว เมื่อทราบลักษณะความสัมพันธ์ระหว่างตัวแปรดังกล่าวแล้ว ก็จะสามารถพยากรณ์ค่าของตัวแปรหนึ่งจากตัวแปรอีกกลุ่มหนึ่งได้ กรณีศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัว (X) กับตัวแปรตามหรือตัวแปรที่เป็นผลตอบสนอง (Y) ที่มีแนวโน้มเป็นเส้นตรง เรียกว่า simple linear regression

ในการวิเคราะห์การถดถอยจะใช้โปรแกรม Excel มาช่วยในการวิเคราะห์ ดังนั้น รูปแบบการถดถอยคือ

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

ซึ่ง β_0 , β_1 เป็นค่าพารามิเตอร์ และ ε เป็นความคลาดเคลื่อนที่เกิดขึ้นโดยสุ่ม

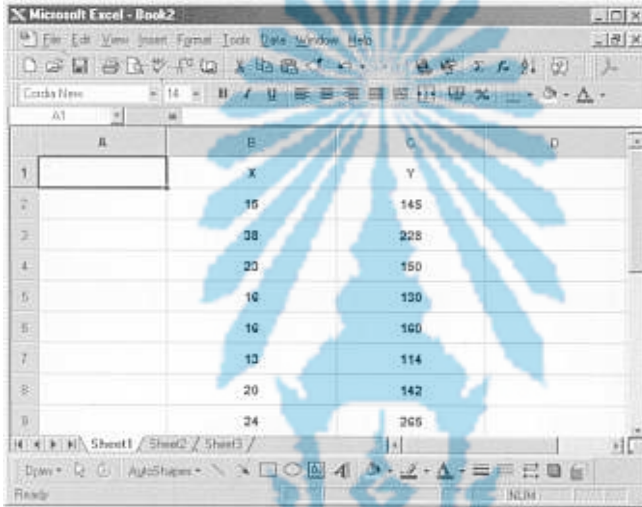
ตัวอย่าง ต้องการวิเคราะห์ ความสัมพันธ์เชิงเส้นระหว่าง ราคาขายกับพื้นที่ใช้สอย

X : พื้นที่ใช้สอย (ร้อยตารางฟุต)	15	38	23	16	16	13	20	24
Y : ราคาขาย (พันดอลลาร์)	145	228	150	130	160	114	142	265

* อาจารย์ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ

ขั้นตอนในการวิเคราะห์การถดถอย

กรอกข้อมูลลงในเซลล์ต่าง ๆ ตามตัวอย่างข้างล่าง



	A	B	C	D
1		X	Y	
2		16	145	
3		28	228	
4		20	150	
5		10	120	
6		10	100	
7		13	114	
8		20	142	
9		24	266	

หาค่าเฉลี่ย \bar{x}

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

จะได้ AVERAGE (B2:B9) เท่ากับ 20.625

หาค่าสัมประสิทธิ์สหสัมพันธ์ (r) :

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

จะได้ CORREL (B2:B9 ; C2:C9) เท่ากับ 0.716708

$$SSR = b_1^2 \sum(x_i - \bar{x})^2$$

จากค่า b_1 , $\sum(x_i - \bar{x})^2$ ที่หาค่าแล้วข้างบน จะได้ SSR = 9733.792

หาค่าสัมประสิทธิ์การถดถอย (b_0)

$$b_0 = \bar{y} - b\bar{x}$$

INTERCEPT (C2:C9; B2:B9) เท่ากับ 71.0249

หาค่าสัมประสิทธิ์การถดถอย (b_1)

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

SLOPE (C2:C9; B2:B9) เท่ากับ 4.641217

หาค่าสัมประสิทธิ์กำหนด (r^2)

จะได้ RSQ (C2:C9; B2:B9) เท่ากับ 0.51367

หาค่า standard error of regression (s_{y-x})

$$s_{y-x} = \sqrt{\frac{\sum(y - \hat{y})^2}{n-2}}$$

STEYX (C2:C9; B2:B9) เท่ากับ 39.19121

หาค่า $\sum(x_i - \bar{x})^2$

DEVSQ (B2:B9) เท่ากับ 451.875

หาค่า b_1^2

POWER (4.641217; 2) เท่ากับ 21.5409

หาค่า SS Regression (SSR)

$$SSR = b_1^2 \sum(x_i - \bar{x})^2$$

จากค่า b_1^2 , $\sum(x_i - \bar{x})^2$ ที่หาค่าแล้วข้างบน จะได้ SSR = 9733.792

หาค่า MS Error (MSE)

$$MSE = s_{y-x}^2$$

POWER (39.19121; 2) เท่ากับ 1535.951

หาค่า สถิติ F

$$F = \frac{SSR/1}{MSE}$$

จากค่า SSR, MSE ข้างบน จะได้ F 6.34

หาค่า P-value

จะได้ FDIST (6.34 ;1;6) เท่ากับ 0.045

การทดสอบ $H_0 : \beta_1 = 0$ กับ $H_1 : \beta_1 \neq 0$

ถ้า P-value > α แสดงว่า $\beta_1 = 0$

หาค่าประมาณความแปรปรวนของ b_1 ($s^2_{b_1}$)

$$s^2(b_1) = \frac{MSE}{\sum(x_i - \bar{x})^2}$$

จากค่า MSE , $\sum(x_i - \bar{x})^2$ ข้างบน จะได้ $s^2(b_1) = 3.399062$

หาค่า $s(b_1)$

จะได้ SQRT (3.399062) เท่ากับ 1.843655

หาค่าสถิติ t :

$$t = \frac{b_1}{s(b_1)}$$

จากค่า b_1 , s_{b_1} ข้างบนจะได้ t 2.517



หาค่า P-value

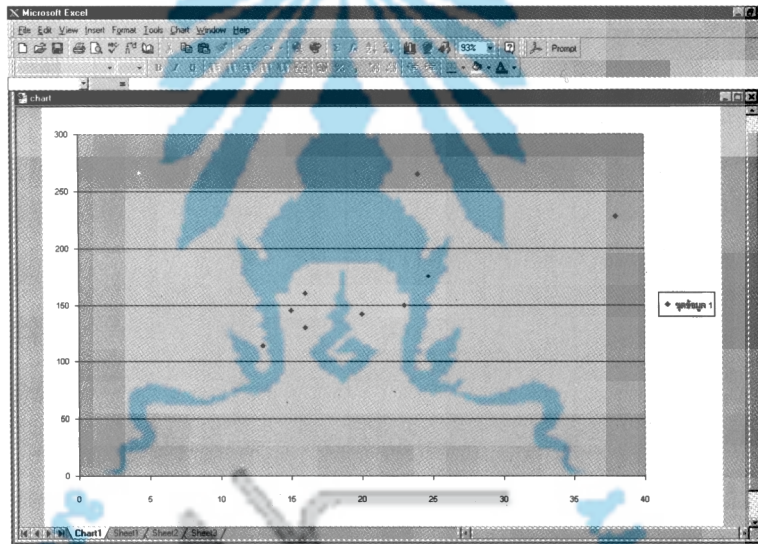
จะได้ TDIST (2.517 ;6;2) เท่ากับ 0.045

การทดสอบ $H_0 : \beta_1 = 0$ กับ $H_1 : \beta_1 \neq 0$

ถ้า P-value > α แสดงว่า $\beta_1 = 0$

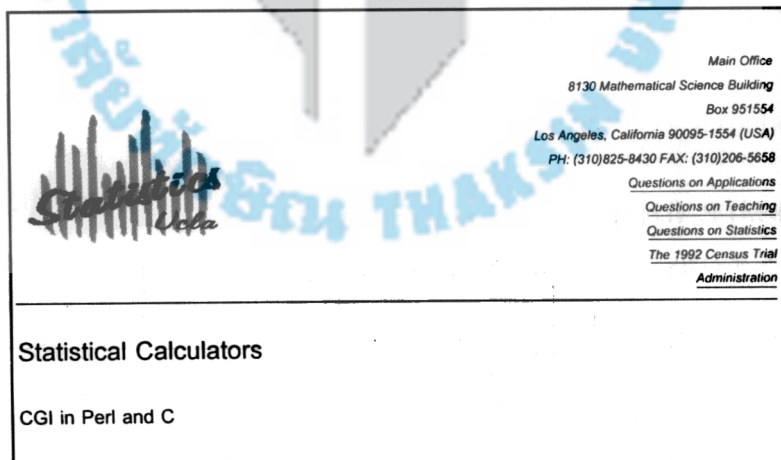
เรานำข้อมูลมาเขียนแผนภาพการกระจาย (scatter diagram) โดย

- 1) นำแถบที่ขอบเขตข้อมูลที่ B2 : C9
- 2) เลือกเครื่องมือ 
- 3) เลือกชนิดกราฟ xy (Scatter) 
- 4) เลือกคำสั่ง ok
- 5) จะได้ผลลัพธ์ ดังรูปข้างล่าง



โปรแกรมในอินเทอร์เน็ตใช้ในการวิเคราะห์การถดถอย เรามีวิธีการดังนี้

1. พิมพ์เว็บเพจ <http://www.stat.ucla.edu/calculators/>

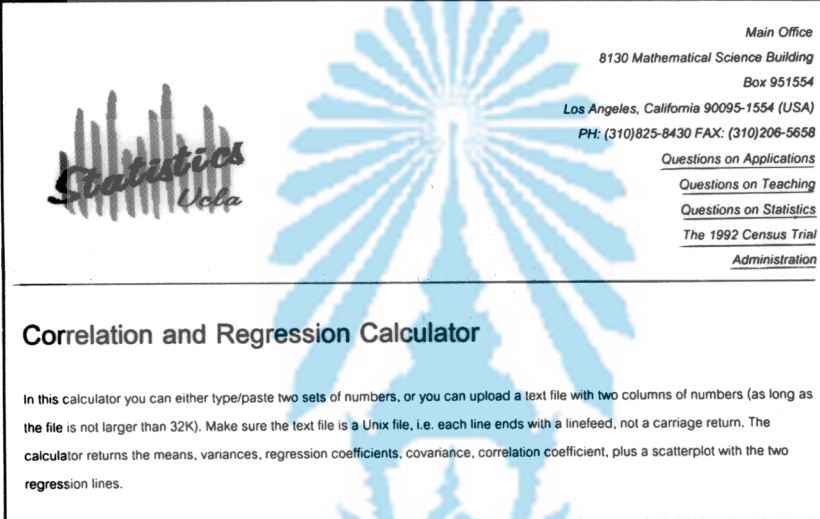


Statistical Calculators
CGI in Perl and C

Main Office
8130 Mathematical Science Building
Box 951554
Los Angeles, California 90095-1554 (USA)
PH: (310)825-8430 FAX: (310)206-5658

[Questions on Applications](#)
[Questions on Teaching](#)
[Questions on Statistics](#)
[The 1992 Census Trial](#)
[Administration](#)

2. คลิกที่ ลิงค์ Correlation and Regression Calculator



Statistics Ucla

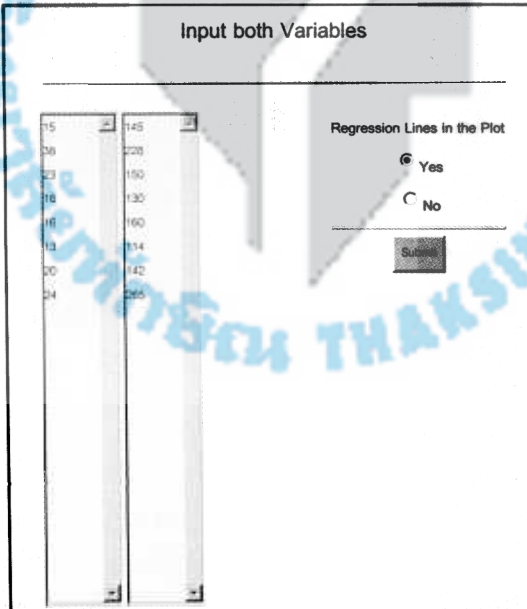
Main Office
8130 Mathematical Science Building
Box 951554
Los Angeles, California 90095-1554 (USA)
PH: (310)825-8430 FAX: (310)206-5658

[Questions on Applications](#)
[Questions on Teaching](#)
[Questions on Statistics](#)
[The 1992 Census Trial Administration](#)

Correlation and Regression Calculator

In this calculator you can either type/paste two sets of numbers, or you can upload a text file with two columns of numbers (as long as the file is not larger than 32K). Make sure the text file is a Unix file, i.e. each line ends with a linefeed, not a carriage return. The calculator returns the means, variances, regression coefficients, covariance, correlation coefficient, plus a scatterplot with the two regression lines.

3. พิมพ์ข้อมูล ค่า X ในคอลัมน์ที่ 1 และพิมพ์ ค่า Y ในคอลัมน์ที่ 2 ที่สอดคล้องกัน ตามโจทย์
ในตัวอย่าง
เลือกปุ่ม yes เพื่อกำหนดให้พล็อตเส้นถดถอย
เลือกปุ่ม Submit



Input both Variables

15	145
38	220
23	150
18	130
11	160
13	114
20	142
24	285

Regression Lines in the Plot

Yes
 No

Submit

4. จะได้ผลลัพธ์ดังข้างล่าง

Statistics

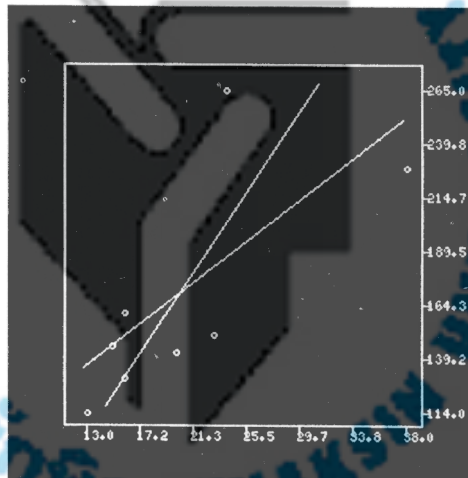
Mean X	20.625000
Mean Y	166.750000
Variance X	56.484375
Variance Y	2368.687500
Covariance X and Y	262.156250
Correlation X and Y	0.716708
Regression Y on X	4.641217
Regression X on Y	0.110676

Lines

$$y = 166.750000 + 4.641217(x - 20.625000)$$

$$x = 20.625000 + 0.110676(y - 166.750000)$$

Scatterplot



หนังสืออ้างอิง

นันทนา ณ ระนอง แผ่นปลิวคำสอนประกอบการเรียนวิชา ST242 เรื่องการวิเคราะห์การถดถอยเชิงเส้นสหสัมพันธ์,
ลำดับที่ 915 หมายเลขเอกสาร 26007 การผลิตเอกสารและตำรามหาวิทยาลัยทักษิณ.
ภุริทัต ทองปรีชา คู่มืออ้างอิงเวิร์คชีตฟังก์ชัน Excel 5, บริษัทซีเอ็ดดูเคชั่น จำกัด (มหาชน), 2538.