

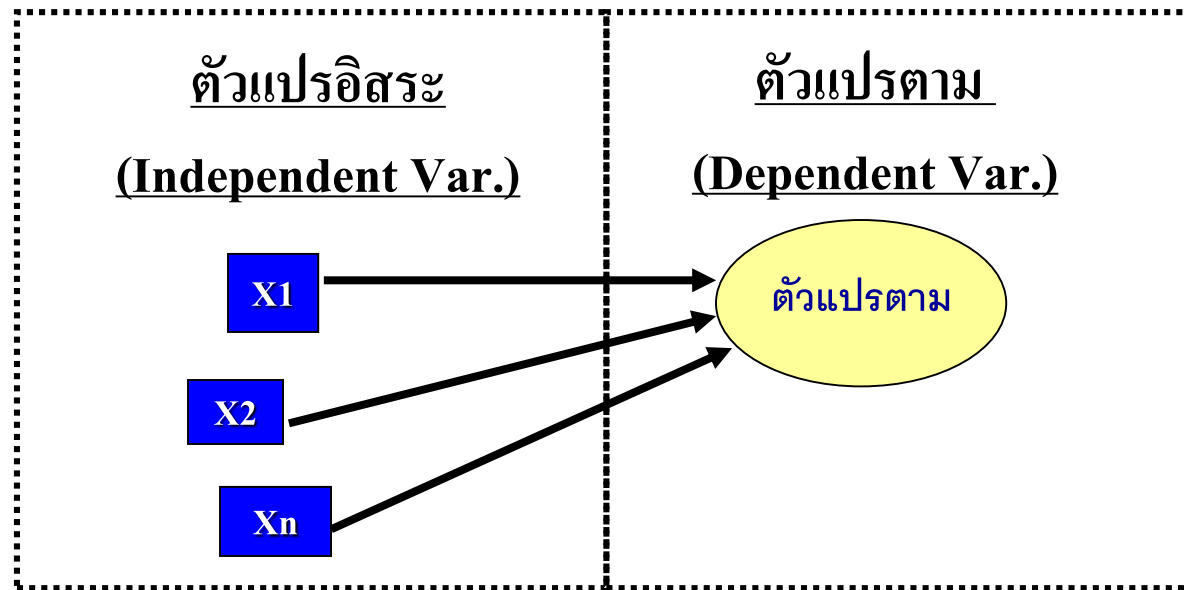
บทที่ 7

การวิเคราะห์ถดถอยและสหสัมพันธ์อย่างง่าย

(Simple linear regression and Correlation)

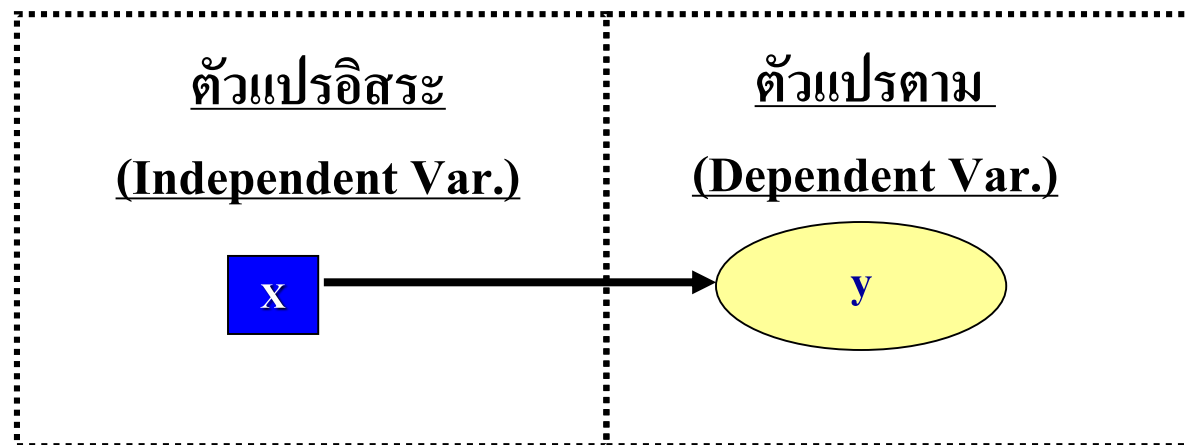
การวิเคราะห์ความถดถอย (Regression Analysis)

- การวิเคราะห์ความถดถอย คือ **การศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป** เพื่อทำให้สามารถประมาณค่าของตัวแปรหนึ่งจากตัวแปรอื่นๆ โดยจะต้องทราบค่าตัวแปรอื่นๆ ล่วงหน้า เช่น ความสัมพันธ์ระหว่างเกรดเฉลี่ยกับคะแนนของรายวิชาต่างๆ ที่ลงทะเบียนเรียน



การวิเคราะห์ความถดถอยอย่างง่าย (Simple regression)

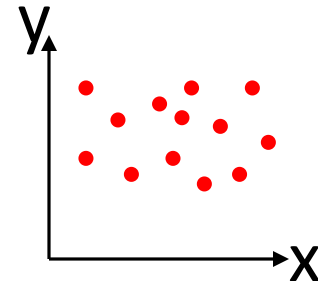
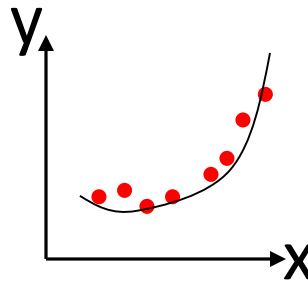
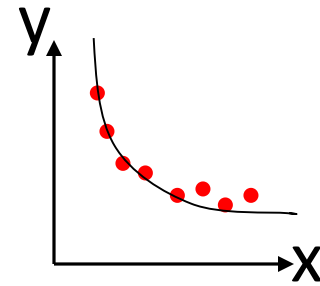
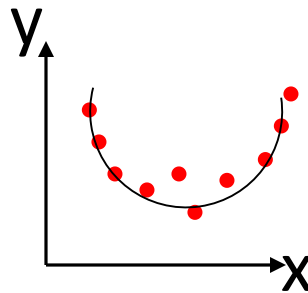
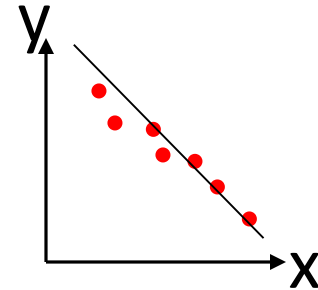
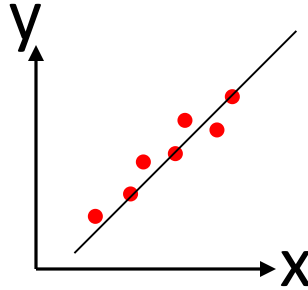
- การวิเคราะห์ความถดถอยอย่างง่าย คือ **การศึกษาความสัมพันธ์ระหว่าง 2 ตัวแปร** เพื่อทำให้สามารถประมาณค่าของตัวแปรตาม (y) ได้จากตัวแปรอิสระ (x) เช่น
 - ความสัมพันธ์ระหว่างรายจ่ายกับรายรับ
 - ความสัมพันธ์ระหว่างแรงดัน **output** กับแรงดัน **input**
 - ความสัมพันธ์ระหว่างเกรดกับคะแนนรายวิชาสถิติ



แผนภาพการกระจาย (Scatter Diagram)

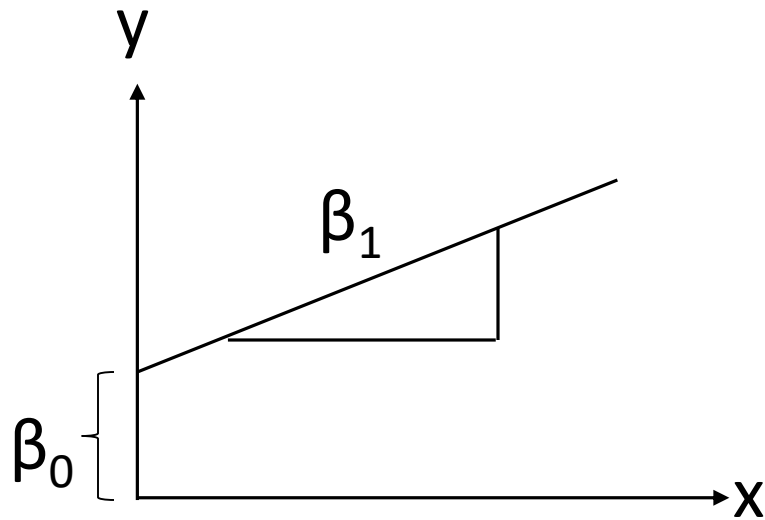
- แผนภาพการกระจาย คือ แผนภาพที่แสดงความสัมพันธ์ระหว่าง 2 ตัวแปร ซึ่งสามารถอยู่ในรูปแบบของ

- เส้นตรง
- พาราโบลา
- เอ็กซ์โพเนนเชียล



การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis)

- การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย คือ การศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ที่มี **ความสัมพันธ์ในรูปเชิงเส้น** ซึ่งสามารถแสดงให้อยู่ในรูปสมการเชิงเส้นได้ดังนี้



$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

- Y ตัวแปรตาม
- X ตัวแปรอิสระ
- β_0 ส่วนตัดแกน Y
- β_1 ความชัน (slope)
- e ความคลาดเคลื่อนอย่างสุ่ม

การประมาณค่าพารามิเตอร์ของสมการความถดถอย

- การจะหาค่า β_0 และ β_1 เป็นไปได้ยากในทางปฏิบัติ จึงต้องใช้ข้อมูลตัวอย่างขนาด n ในการประมาณค่าโดยวิธี **กำลังสองน้อยสุด (Least Square Method)** ดังนี้

จากเดิม
$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

ประมาณได้

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

โดย

$$\bullet \beta_1 = b = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{S^2_{XY}}{S^2_X}$$

$$\bullet \beta_0 = a = \bar{y} - b\bar{x}$$

การประมาณค่าพารามิเตอร์ของสมการความถดถอย

- **Ex** ในการหาความสัมพันธ์ระหว่างแรงดัน **output** และแรงดัน **input** ของวงจรจ่ายแรงดัน (**mV**) เพื่อนำมาประมาณค่าของการจ่ายแรงดัน **output** โดยเก็บข้อมูลตัวอย่างมาดังนี้ จงหาสมการความถดถอย

ครั้งที่	1	2	3	4	5
input	1	2	3	4	5
output	1	1	2	2	4

การประมาณค่าพารามิเตอร์ของสมการความถดถอย

Output(Y)	Input(X)	$X_i Y_i$	x_i^2
1	1		
1	2		
2	3		
2	4		
4	5		
$\Sigma Y =$ $\bar{Y} =$	$\Sigma X =$ $\bar{X} =$	$\Sigma XY =$	$\Sigma X^2 =$

การประมาณค่าพารามิเตอร์ของสมการความถดถอย

- **Ex1** ในการหาความสัมพันธ์ระหว่างแรงดัน **output** และแรงดัน **input** ของวงจรจ่ายแรงดัน (mV) เพื่อนำมาประมาณค่าของการจ่ายแรงดัน **output** โดยเก็บข้อมูลตัวอย่างมาดังนี้ จงหาสมการความถดถอย

ครั้งที่	1	2	3	4	5	6	7	8	9	10	11	12
in	0.8	1.0	1.6	2.0	2.2	2.6	3.0	3.0	4.0	4.0	4.0	4.6
out	22	28	22	26	34	18	30	38	30	40	50	46

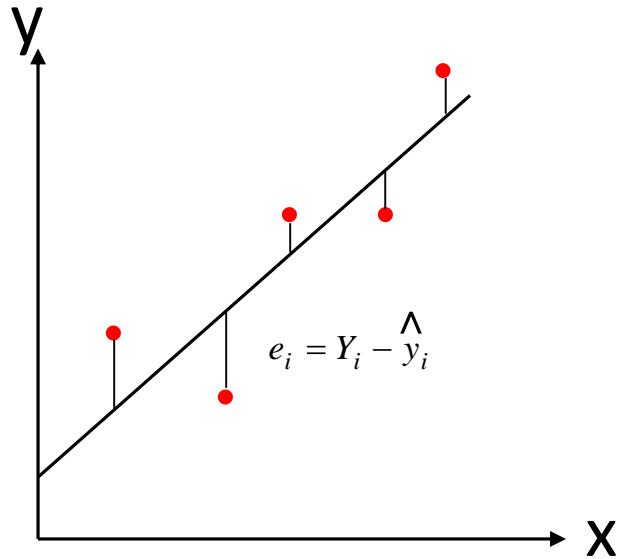
การประมาณค่าพารามิเตอร์ของสมการความถดถอย

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
0.8	22			
1.0	28			
1.6	22			
2.0	26			
2.2	34			
2.6	18			
3.0	30			
3.0	38			
4.0	30			
4.0	40			
4.0	50			
4.6	46			
$\Sigma X =$ $\overline{X} =$	$\Sigma Y =$ $\overline{Y} =$	$\Sigma XY =$	$\Sigma X^2 =$	$\Sigma Y^2 =$

การประมาณค่าแปรปรวนของความคาดเคลื่อน(S^2)

- ค่าคาดเคลื่อนในการประมาณ Y_i ด้วย \hat{y}_i สามารถคำนวณได้ดังนี้

การประมาณค่าแปรปรวนของความคาดเคลื่อน



$$\bullet S^2 = \frac{SSE}{n - 2}$$

$$SSE = S_Y^2 - bS_{XY}^2$$

$$S_Y^2 = \frac{\sum Y^2}{n} - \frac{(\sum Y)^2}{n^2}$$

- จาก **EX1** จงหาความแปรปรวนความคาดเคลื่อน

การประมาณค่าแบบช่วงค่า β_1

- การประมาณค่า β_1 แบบช่วง ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n \geq 30$** สามารถทำได้ดังนี้

$$\beta_1 \pm Z_{\frac{\alpha}{2}} \frac{S}{S_X}$$

- การประมาณค่า β_1 แบบช่วง ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n < 30$** สามารถทำได้ดังนี้

$$\beta_1 \pm t_{\frac{\alpha}{2}; n-2} \frac{S}{S_X}$$

การประมาณค่าแบบช่วงค่า β_1

- จากตัวอย่าง **EX1** จงประมาณค่า β_1 แบบช่วงที่ระดับความเชื่อมั่น 95%

การประมาณค่าแบบช่วงค่า β_0

- การประมาณค่า β_0 แบบช่วง ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n \geq 30$** สามารถทำได้ดังนี้

$$\beta_0 \pm Z_{\frac{\alpha}{2}} S \sqrt{(1/n) + (\bar{X}^2 / S_X^2)}$$

- การประมาณค่า β_0 แบบช่วง ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n < 30$** สามารถทำได้ดังนี้

$$\beta_0 \pm t_{\frac{\alpha}{2}, n-2} S \sqrt{(1/n) + (\bar{X}^2 / S_X^2)}$$

การประมาณค่าแบบช่วงค่า β_0

- จากตัวอย่าง **EX1** จงประมาณค่า β_0 แบบช่วงที่ระดับความเชื่อมั่น 95%

การประมาณค่า **Y** แบบช่วงเมื่อกำหนดค่า **X**

- การประมาณค่า **Y** แบบช่วงที่ระดับความเชื่อ $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n \geq 30$** สามารถทำได้ดังนี้

$$\hat{y}_p \pm Z_{\frac{\alpha}{2}} S \sqrt{1 + (1/n) + (X_p - \bar{X})^2 / S_X^2}$$

- การประมาณค่า **Y** แบบช่วงที่ระดับความเชื่อ $(1-\alpha)100\%$ เมื่อ **ขนาดตัวอย่าง $n < 30$** สามารถทำได้ดังนี้

$$\hat{y}_p \pm t_{\frac{\alpha}{2}; n-2} S \sqrt{1 + (1/n) + (X_p - \bar{X})^2 / S_X^2}$$

การประมาณค่า **Y** แบบช่วงเมื่อกำหนดค่า **X**

- จากตัวอย่าง **EX1** ถ้าต้องการประมาณค่าแรงดัน **output** แบบช่วง เมื่อ กำหนดแรงดัน **input** เป็น **2.5 mV** ที่ระดับความเชื่อมั่น **95%**

การประมาณค่าเฉลี่ย Y แบบช่วงเมื่อกำหนดค่า X

- การประมาณค่าเฉลี่ย Y ($\mu_{Y.X}$) แบบช่วงที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อขนาดตัวอย่าง $n \geq 30$

$$\hat{y}_p \pm Z_{\frac{\alpha}{2}} S \sqrt{(1/n) + (X_p - \bar{X})^2 / S_X^2}$$

- การประมาณค่าเฉลี่ย Y ($\mu_{Y.X}$) แบบช่วงที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อขนาดตัวอย่าง $n < 30$

$$\hat{y}_p \pm t_{\frac{\alpha}{2}; n-2} S \sqrt{(1/n) + (X_p - \bar{X})^2 / S_X^2}$$

การประมาณค่าเฉลี่ย **Y** แบบช่วงเมื่อกำหนดค่า **X**

- จากตัวอย่าง **EX1** ถ้าต้องการประมาณค่าเฉลี่ยแรงดัน **output** แบบช่วงเมื่อกำหนดแรงดัน **input** เป็น **2.5 mV** ที่ระดับความเชื่อมั่น **95%**

สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination : R^2)

- สัมประสิทธิ์การตัดสินใจ (R^2) คือ ค่าที่ใช้อธิบายเปอร์เซ็นต์การเปลี่ยนแปลงของตัวแปรตาม (Y) ว่ามีความสัมพันธ์กับตัวแปรอิสระ (X) มากหรือน้อย ซึ่งสามารถคำนวณได้ดังนี้

$$\begin{aligned} \bullet R^2 &= 1 - \frac{SSE}{SST} \\ SST &= SSR + SSE \\ SSR &= \frac{(S_{XY}^2)^2}{S_X^2} \end{aligned}$$

- $0 \leq R^2 \leq 1$

สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination : R^2)

- คุณสมบัติของ R^2
 - R^2 มีค่าเข้าใกล้ 1 คือ **X และ Y มีความสัมพันธ์กันมาก** หรือ
แสดงว่าเปอร์เซ็นต์ที่ **X** สามารถอธิบายการเปลี่ยนแปลงของ **Y** มีค่ามาก
 - R^2 มีค่าเข้าใกล้ 0 คือ **X และ Y มีความสัมพันธ์กันน้อย** หรือ
แสดงว่าเปอร์เซ็นต์ที่ **X** สามารถอธิบายการเปลี่ยนแปลงของค่า **Y** มีค่าน้อย
 - R^2 มีค่าอยู่ระหว่าง 0.3 ถึง 0.7 คือ **X และ Y มีความสัมพันธ์ปานกลาง**
หรือ แสดงว่าเปอร์เซ็นต์ที่ **X** สามารถอธิบายการเปลี่ยนแปลงของ **Y** ได้
ปานกลาง

สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination : R^2)

- จากตัวอย่าง **EX1** จงหาสัมประสิทธิ์การตัดสินใจ พร้อมทั้งอธิบายความหมาย

สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient: r)

- สัมประสิทธิ์สหสัมพันธ์ (r) คือ ค่าที่แสดงถึง **ประเภทความสัมพันธ์เชิงเส้นตรง** ระหว่าง 2 ตัวแปร

$$r = \frac{S_{XY}^2}{\sqrt{S_X^2 S_Y^2}}$$

- คุณสมบัติของ r
 - ค่า r เป็นบวก แสดงว่า **X** และ **Y** มีความ **สัมพันธ์เชิงบวก** (แปรผันตรง)
ถ้า **X** **เพิ่ม** **Y** ก็ **เพิ่ม** หรือ **X** **ลด** **Y** ก็ **ลด**
 - ค่า r เป็นลบ แสดงว่า **X** และ **Y** มีความ **สัมพันธ์เชิงลบ** (แปรผกผัน)
ถ้า **X** **เพิ่ม** **Y** จะ **ลด** หรือ **X** **ลด** **Y** จะ **เพิ่ม**
 - $r = 0$ หมายถึง **X** และ **Y** **ไม่มีความสัมพันธ์กัน**

สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient: r)

- จาก **Ex1** จงหาค่าสัมประสิทธิ์สหสัมพันธ์

Excel

- การวิเคราะห์ความถดถอยอย่างง่าย

Data -> Data Analysis -> Regression

The image shows an Excel spreadsheet with two columns: 'input (x)' and 'output (y)'. The data is as follows:

	A	B
1	input (x)	output (y)
2	0.8	22
3	1	28
4	1.6	22
5	2	26
6	2.2	34
7	2.6	18
8	3	30
9	3	38
10	4	30
11	4	40
12	4	50
13	4.6	46

Overlaid on the spreadsheet is the 'Data Analysis' dialog box. The 'Analysis Tools' list includes: F-Test Two-Sample for Variances, Fourier Analysis, Histogram, Moving Average, Random Number Generation, Rank and Percentile, **Regression** (highlighted), Sampling, t-Test: Paired Two Sample for Means, and t-Test: Two-Sample Assuming Equal Variances. The 'OK' button is highlighted.

Excel

- การวิเคราะห์ความถดถอยอย่างง่าย

ตัวแปรอิสระ, ตัวแปรตาม

Access Web Text Sources Connections All Edit Links Advanced Columns Duplicate

Get External Data

C13 fx

	A	B	C
1	input (x)	output (y)	
2	0.8	22	
3	1	28	
4	1.6	22	
5	2	26	
6	2.2	34	
7	2.6	18	
8	3	30	
9	3	38	
10	4	30	
11	4	40	
12	4	50	
13	4.6	46	
14			
15			
16			
17			

Regression

Input

Input Y Range: \$B\$2:\$B\$13

Input X Range: \$A\$2:\$A\$13

☒ Labels ☐ Constant is Zero

☒ Confidence Level: 95 %

Output options

☐ Output Range:

☒ New Worksheet Ply: regression

☐ New Workbook

Residuals

☒ Residuals ☒ Residual Plots

☒ Standardized Residuals ☒ Line Fit Plots

Normal Probability

☒ Normal Probability Plots

OK Cancel Help

ข้อตัวแปร

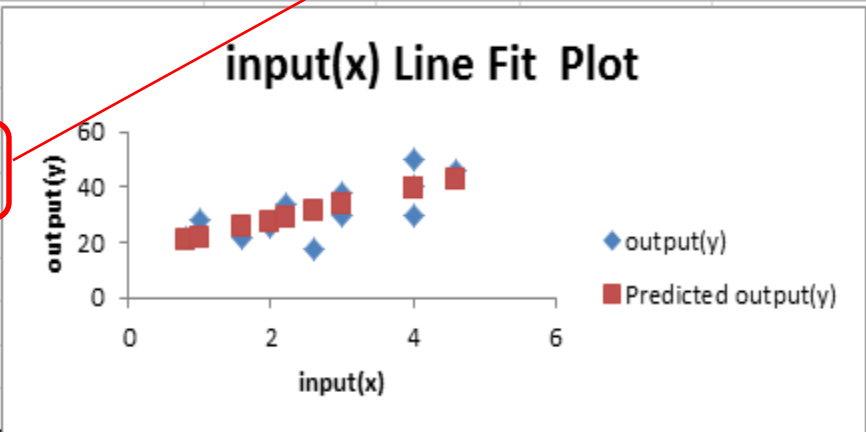
ระดับความเชื่อมั่น

ค่าความคาดเคลื่อนและข้อมูลอื่นๆ

Excel

- การวิเคราะห์ความถดถอยอย่างง่าย

R^2, r

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT		<div>input(x) Line Fit Plot</div> 						
2									
3	Regression Statistics								
4	Multiple R	0.734370327							
5	R Square	0.539299777							
6	Adjusted R Square	0.493229755							
7	Standard Error	7.053766656							
8	Observations	12							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	582.4437596	582.44376	11.70609	0.006531887			
13	Residual	10	497.5562404	49.755624					
14	Total	11	1080						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	16.14329738	5.062143455	3.18902408	0.00967	4.864138875	27.422456	4.86413888	27.4224559
18	input(x)	5.801232666	1.695564731	3.42141622	0.006532	2.023279013	9.5791863	2.02327901	9.57918632

SSE
SST

β_0

β_1

การบ้าน

- ในการหาความสัมพันธ์ระหว่างแรงดัน **output** และแรงดัน **input** ของวงจรจ่ายแรงดัน (**mV**) เพื่อนำมาประมาณค่าของการจ่ายแรงดัน **output** โดยเก็บข้อมูลตัวอย่างมาดังนี้ จงหาสมการความถดถอย

	1	2	3	4	5	6	7	8	9	10
in	1.2	0.8	1.0	1.3	0.7	0.8	1.0	0.6	0.9	1.1
out	101	92	110	120	90	82	93	75	91	105

- จงหาสมการถดถอย
- จงประมาณค่าแปรปรวนของความคาดเคลื่อน
- จงประมาณค่าแบบช่วงค่า β_1 และ β_0 ที่ความเชื่อมั่น 95%
- จงประมาณค่า Y แบบช่วง เมื่อกำหนด $X = 2.5 \text{ mV}$
- จงประมาณค่าเฉลี่ย Y แบบช่วง เมื่อกำหนด $X = 2.5 \text{ mV}$
- หาค่า R^2 และ r พร้อมอธิบายความหมาย