

ความน่าจะเป็นและสถิติสำหรับวิศวกร



Probability and Statistics
for Engineers

e-mail: reudee.ni@ssru.ac.th

Web site: http://www.elfit.ssru.ac.th/reudee_ni/

รศ.ดร.ฤดี นิยมรัตน์

คำอธิบายรายวิชา

ความน่าจะเป็น ตัวแปรสุ่ม
การทดสอบสมมติฐาน การแจกแจง
การแจกแจงแบบมีเงื่อนไข ฟังก์ชันความ
หนาแน่น ความน่าจะเป็น ค่าความคาดหวัง

2 ความแปรปรวน

1 การอนุมานทางสถิติ

3 การวิเคราะห์การถดถอย

การประยุกต์ใช้งาน

Probability, random variables,
hypotheses testing, distribution,
conditional distribution, probability
density function, expectations,

variances,

statistical inferences,

regression analysis,

implementation



กำหนดการสอน 5 ครั้ง

- การประเมินผล
1. คะแนนการร่วมกิจกรรมในชั้นเรียน
 2. คะแนนการทดสอบ รวม 30 คะแนน

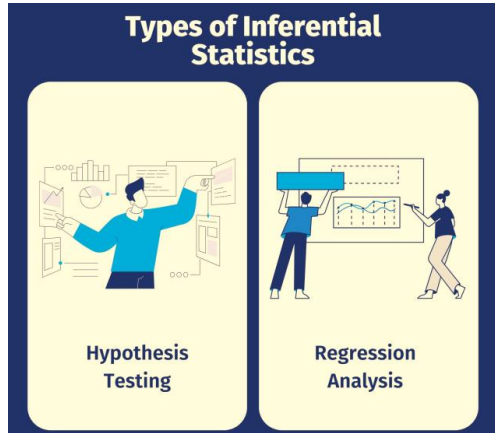
หัวข้อการเรียน	การทดสอบ (คะแนน)
การอนุมานทางสถิติ	20
t-test	20
F-test (one way ANOVA)	20
การวิเคราะห์การถดถอยอย่างง่าย	20
การวิเคราะห์การถดถอยพหุคูณ	20

การทุจริต
ในการทดสอบ
มีคะแนนการ
ประเมินผล
ทั้ง 5 ครั้ง เป็น 0

3

การวิเคราะห์
การถดถอย

เป็นสถิติที่ใช้ประมาณค่า**ความสัมพันธ์**ระหว่างตัวแปรอิสระ (Independent Variable, IV: X) และตัวแปรตาม (Dependent Variable, DV: Y)

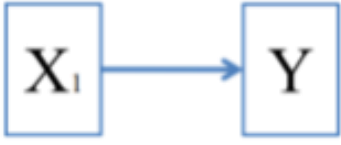


Regression Analysis

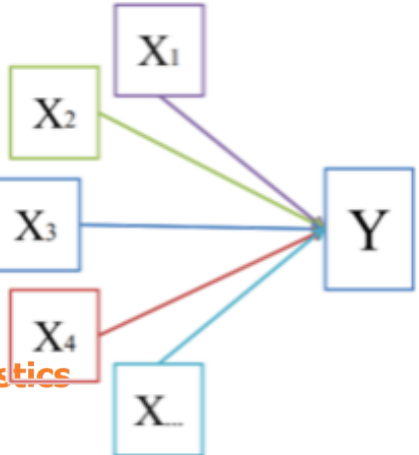
เชิงเส้น
Linear Regression

ไม่เป็นเชิงเส้น
Non-Linear Regression

Simple Linear Regression



Multiple Linear Regression



ข้อตกลงเบื้องต้นสำหรับการวิเคราะห์การถดถอยเชิงเส้น

1.นามบัญญัติ (Nominal)	ชื่อ ประเภท แบ่งกลุ่ม แบ่งพวก ชนิดของสิ่งต่าง ๆ	<ul style="list-style-type: none"> -เพศ (หญิง/ ชาย) -กลุ่มเลือด (A B AB และ O) -สถานภาพสมรส (โสด คู่ หม้าย หย่า และ แยก) -บ้านเลขที่ เบอร์โทรศัพท์ หมายเลขเสื้อนักกีฬา ฯลฯ
2.เรียงอันดับ (Ordinal)	บอกลำดับข้อมูลได้ว่ามาก น้อย สูงหรือต่างกัน แต่บอกไม่ได้ว่าต่างกันเท่าใด	<ul style="list-style-type: none"> -ระดับความเจ็บปวด (มาก ปานกลาง น้อย) -ระดับความคิดเห็น (เห็นด้วย เฉย ๆ ไม่เห็นด้วย) -ระดับความพึงพอใจ (มาก น้อย ไม่พอใจ)
3.อันดับภาค (Interval)	เป็นตัวเลขต่อเนื่อง แต่เป็นศูนย์ไม่แท้ (คือที่จุด 0 ยังมีค่าอยู่)	<ul style="list-style-type: none"> -อุณหภูมิ - I.Q. -คะแนนความรู้ -คะแนนทัศนคติ -คะแนนการปฏิบัติ ฯลฯ
4.อัตราส่วน (Ratio)	เป็นตัวเลขต่อเนื่อง มีจุดศูนย์แท้ (คือค่า 0 มีความหมายว่าไม่มีค่า)	<ul style="list-style-type: none"> -น้ำหนัก -ส่วนสูง -น้ำตาลในเลือด -ความดันโลหิต -จำนวนผู้ป่วย ฯลฯ

1. ตัวแปรอิสระและตัวแปรตาม มีมาตรวัดเป็น Interval ขึ้นไป (อนุญาตให้ตัวแปรอิสระ บางตัวมีมาตรวัดเป็น Nominal หรือ Ordinal ได้บาง โดยจะต้องเปลี่ยนตัวแปรอิสระเป็นตัวแปรหุ่นก่อน)

2. สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ (Normal Curve Distribution)

3. Independent Variable, IV: X ต้องไม่มีความสัมพันธ์กันเอง

4. ข้อมูล ต้องไม่มีความสัมพันธ์ภายในตัวเอง

5. ค่าความคลาดเคลื่อนที่เกิดจากการพยากรณ์/วิเคราะห์จะต้อง

5.1 มีการแจกแจงแบบปกติ (Assumption of Normality)

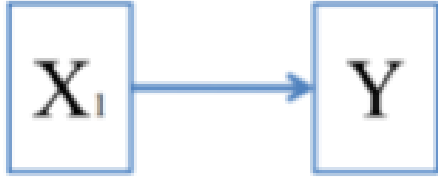
5.2 มีค่าเฉลี่ยเท่ากับ 0

5.3 มีความแปรปรวนคงที่ (Homogeneity of Variance)

5.4 ความคลาดเคลื่อนเป็นอิสระกัน (Assumption of Autocorrelation) (เฉพาะการวิเคราะห์ Multiple Linear Regression)



Simple Linear Regression



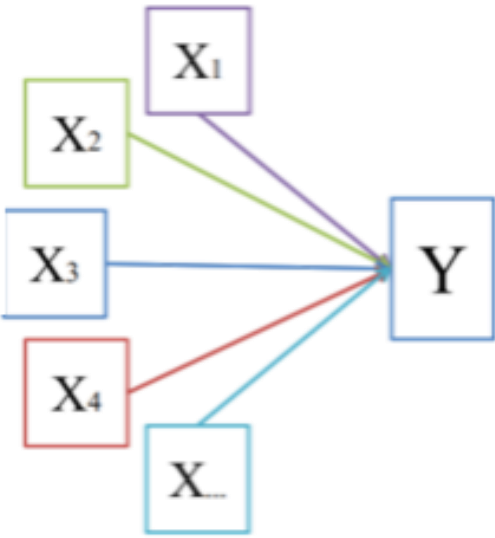
ความดันเลือดของแม่ (X) (ตัวแปรอิสระ) สามารถทำนาย น้ำหนักของเด็กแรกเกิด (Y) (ตัวแปรตาม) ได้หรือไม่

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

Labels:
 - Y_i : Dependent Variable
 - β_0 : Population Y intercept
 - β_1 : Population Slope Coefficient
 - X_i : Independent Variable
 - ϵ_i : Random Error term
 - $\beta_0 + \beta_1 X_i$: Linear component
 - ϵ_i : Random Error component

$$Y = a + bX$$

Multiple Linear Regression



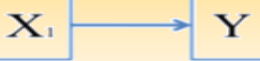
อายุ (X1) จำนวนปีที่สมรส (X2) ระยะเวลาที่รู้จักกันก่อน แต่งงาน (X3) รายได้ของครอบครัว (X4) และจำนวนบุตร (X5) สามารถทำนายความพึงพอใจในชีวิตสมรส (Y) ได้หรือไม่

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \epsilon$$

Labels:
 - Y : Dependent Variable (Response Variable)
 - β_0 : Y intercept
 - β_1, β_2, \dots : Slope Coefficient
 - X_1, X_2, \dots : Independent Variables (Predictors)
 - ϵ : Error Term



Simple Linear Regression



Simple Linear Regression

การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

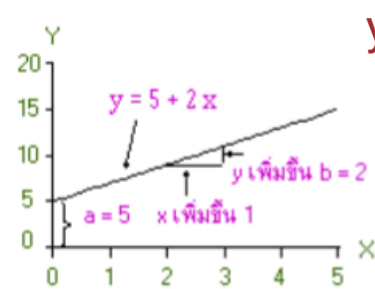
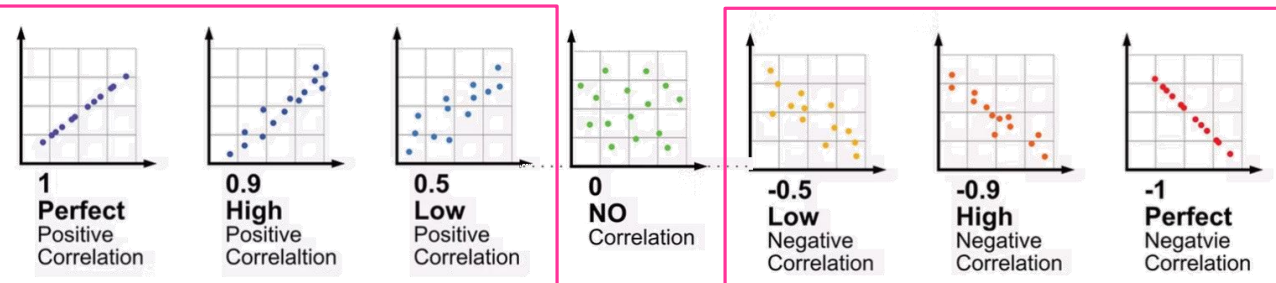
ศึกษาระดับหรือขนาดของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสองตัวแปร (IV: X, DV: Y) ว่ามีมากน้อยเพียงใด

หาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient, r)

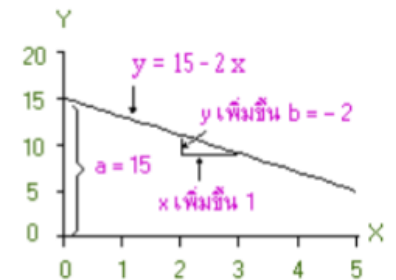
การวิเคราะห์การถดถอย (Regression Analysis)

ประมาณตัวแปรตาม (Y) เรียกตัวตอบสนอง (Response) ด้วยตัวแปรอิสระ (X) เรียกตัวประมาณการ (Predictor)

การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)



$y = 5 + 2x$



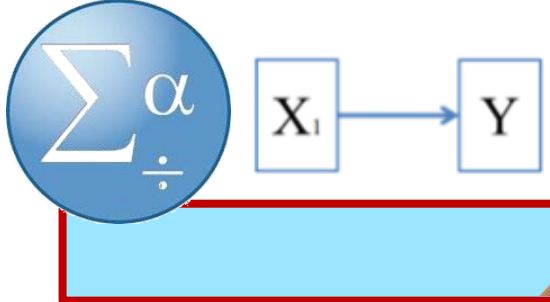
$y = 15 - 2x$

ราคาไข่ มีความสัมพันธ์ทางบวก กับราคาอาหารไก่

ไข่ 10 ฟอง ราคา 38 บาท ไข่ 50 ฟอง ราคา $3.8 \times 50 = 190$ บาท



Simple Linear Regression



ข้อมูล

Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	a	Numeric	8	2	ปริมาณสารA	None	None	8	Right	Scale	Input
2	b	Numeric	8	2	ปริมาณสารB	None	None	8	Right	Scale	Input
3	w	Numeric	8	2	น้ำหนัก	None	None	8	Right	Scale	Input
4	s	Numeric	8	2	สมบัติS	None	None	8	Right	Scale	Input
5	g	Numeric	8	2	สมบัติG	None	None	8	Right	Scale	Input
6	c	Numeric	8	2	สาร C ในร่างกาย	None	None	8	Right	Scale	Input

Data View

	a	b	w	s	g	c
1	12000.00	12000.00	25.00	5.00	3.22	200.00
2	12500.00	12000.00	22.00	5.00	2.98	150.00
3	20000.00	12000.00	28.00	4.00	2.99	250.00
4	25000.00	12500.00	22.00	4.00	3.01	176.00

การวิเคราะห์ โจทย์: สมบัติ S (X) สามารถทำนายผลของสาร C ในร่างกาย (Y) ได้หรือไม่

1

Regression

Linear

2

ตัวแปรตาม Dependent Variable, DV: Y

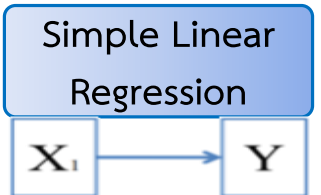
ตัวแปรอิสระ Independent Variable, IV: X

Linear Regression: Statistics

- Estimates
- Descriptives
- Model fit
- R squared change
- Confidence intervals
- Part and partial correlations
- Covariance matrix
- Collinearity diagnostics



การแปลค่า



Output

- Regression
 - Title
 - Notes
 - Descriptive Statist
 - Correlations
 - Variables Entered/Removed**
 - Model Summary**
 - ANOVA
 - Coefficients

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	สมมติ S ^b		Enter

a. Dependent Variable: สาร C ในร่างกาย
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810 ^a	.656	.644	20.40871

a. Predictors: (Constant), สมมติ S

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22279.033	1	22279.033	53.489	.000 ^b
	Residual	11662.433	28	416.515		
	Total	33941.467	29			

a. Dependent Variable: สาร C ในร่างกาย
b. Predictors: (Constant), สมมติ S

Coefficients^a

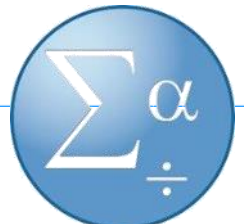
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41.633	19.435		2.142	.041
	สมมติ S	21.244	2.905	.810	7.314	.000

a. Dependent Variable: สาร C ในร่างกาย

1-พิจารณาค่า**ความสัมพันธ์** (R-Square) เพื่อระดับการมีอิทธิพลของตัวแปรอิสระ ต่อตัวแปรตาม บอกเป็นร้อยละ ส่วนที่เหลือเป็นอิทธิพลของตัวแปรหรือปัจจัยอื่นๆ

ค่าความสัมพันธ์ (R-Square) = .656 หมายถึงปริมาณสาร A ที่ใช้ (X) มีอิทธิพลต่อสมบัติ S (Y) ร้อยละ 65.6 (**สูง**) อีกร้อยละ 34.4 เป็นอิทธิพลของปัจจัยอื่น

Coefficient Range	Strength of Relationship
0.00 - 0.20	Very Low
0.21 - 0.40	Low
0.41 - 0.60	Moderate
0.61 - 0.80	High Moderate
0.81 - 1.00	Very High



Simple Linear Regression

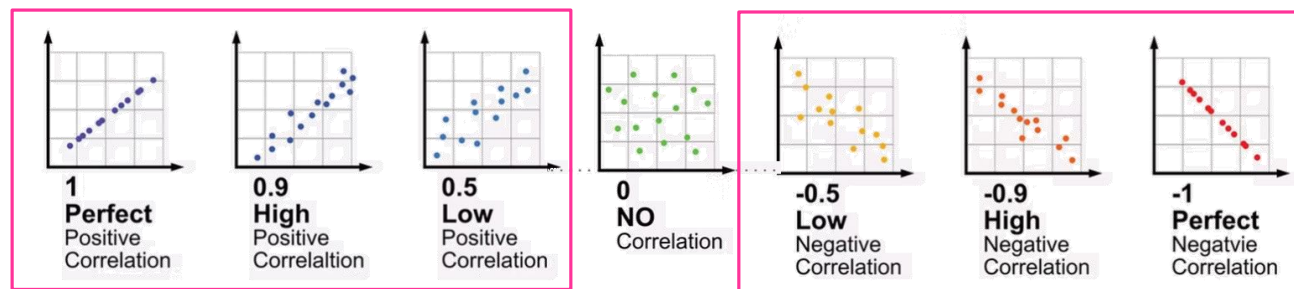


การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

ศึกษาระดับหรือขนาดของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสองตัวแปร (IV: X, DV: Y) ว่ามีมากน้อยเพียงใด

หาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient, r)

	สัญลักษณ์ค่าพารามิเตอร์	สัญลักษณ์ค่าสถิติ
ค่าเฉลี่ย(Mean)	μ (มิว)	\bar{x}
ค่าสัดส่วน(Proportion)	π (พาย) หรือ P	p
ส่วนเบี่ยงเบนมาตรฐาน(Standard Deviation)	σ (ซิกมา)	s หรือ s.d.
ความแปรปรวน(Variance)	σ^2	s^2 หรือ s.d. ²
สัมประสิทธิ์สหสัมพันธ์(Correlation Coefficient)	ρ (โร)	r
สัมประสิทธิ์การถดถอย(Regression Coefficient)	β (เบต้า)	b



$-1 \leq \rho \leq 1$

- $\rho = 0$
- $\rho \neq 0$
- $\rho > 0$
- $\rho < 0$

x และ y ไม่สัมพันธ์กัน
 x และ y สัมพันธ์กัน
 x และ y สัมพันธ์เชิงบวก
 x และ y สัมพันธ์เชิงลบ

Model Summary

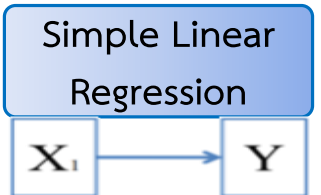
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810 ^a	.656	.644	20.40871

a. Predictors: (Constant), สมมติS

Coefficient Range	Strength of Relationship
0.00 - 0.20	Very Low
0.21 - 0.40	Low
0.41 - 0.60	Moderate
0.61 - 0.80	High Moderate
0.81 - 1.00	Very High



การแปลค่า



Output

- Regression
 - Title
 - Notes
 - Descriptive Statist
 - Correlations
 - Variables Entered/Removed
 - Model Summary
 - ANOVA
 - Coefficients

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	สมมติ S ^b		Enter

a. Dependent Variable: สาร C ในร่างกาย
 b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810 ^a	.656	.644	20.40871

a. Predictors: (Constant), สมมติ S

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	22279.033	1	22279.033	53.489	.000 ^b
	Residual	11662.433	28	416.515		
	Total	33941.467	29			

a. Dependent Variable: สาร C ในร่างกาย
 b. Predictors: (Constant), สมมติ S

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41.633	19.435		2.142	.041
	สมมติ S	21.244	2.905	.810	7.314	.000

a. Dependent Variable: สาร C ในร่างกาย

1-พิจารณาค่าความสัมพันธ์ (R-Square) เพื่อระดับการมีอิทธิพลของตัวแปรอิสระ ต่อตัวแปรตาม บอกเป็นร้อยละ ส่วนที่เหลือเป็นอิทธิพลของตัวแปรหรือปัจจัยอื่นๆ

ค่าความสัมพันธ์ (R-Square) = .656 หมายถึงปริมาณสาร A ที่ใช้ (X) มีอิทธิพลต่อสมบัติ S (Y) ร้อยละ 65.6 (น้อย) อีกร้อยละ 34.4 เป็นอิทธิพลของปัจจัยอื่น

2-พิจารณาค่า Sig. ของ F ว่าสมการถดถอยที่จะเขียน สามารถเขียนได้หรือไม่ โดยค่า Sig. มีค่าน้อยกว่าค่าอัลฟา (.05)

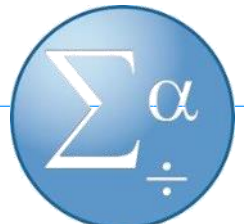
ค่า Sig. หรือค่า P-value = .000 มีค่าน้อยกว่าค่า $\alpha = .05$ หมายถึง สมการถดถอยสามารถเขียนและใช้งานได้

2-พิจารณาว่าปฏิเสธสมมติฐาน H0 หรือไม่ มีค่าน้อยกว่าค่าอัลฟา (.05) จึงปฏิเสธ H0 หมายถึงแตกต่างกัน คือมีอิทธิพล ใช้ทำนายได้

ค่า Sig. หรือค่า P-value = .000 มีค่าน้อยกว่าค่า $\alpha = .05$ หมายถึง ปริมาณสาร A ที่ใช้ (X) สามารถทำนายสมบัติ S (Y) ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

Regression=ส่วนที่ทำนายได้

Residual=ส่วนที่ทำนายไม่ได้



การเขียนสมการทำนาย

Simple Linear Regression



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

Population Y intercept: β_0
 Population Slope Coefficient: β_1
 Independent Variable: X_i
 Random Error term: ϵ_i
 Linear component: $\beta_0 + \beta_1 X_i$
 Random Error component: ϵ_i

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	a	41.633	19.435	2.142	.041
	สมบัติ S	b	21.244	2.905	.810	.000

a. Dependent Variable: สาร C ในร่างกาย

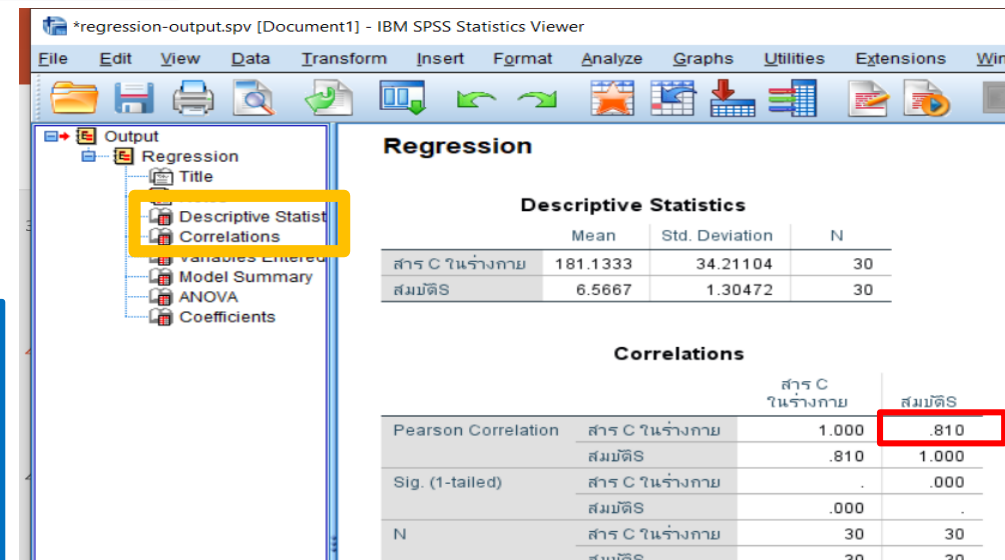
$$Y = a + bX$$

$$\text{สาร C ในร่างกาย (Y)} = 41.633 + 21.244 (\text{สมบัติ S (X)})$$

แทนค่า

หากสินค้านี้มีสมบัติ S = 10 จะมีผลต่อสาร C ในร่างกายเท่าไร จากสมการทำนาย

$$\begin{aligned} \text{สาร C ในร่างกาย} &= 41.633 + 21.244 (\text{สมบัติ S}) \\ &= 41.633 + 21.244 (10) = 254.073 \end{aligned}$$

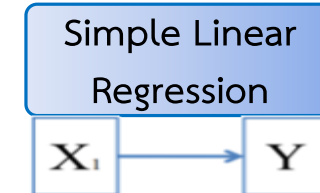
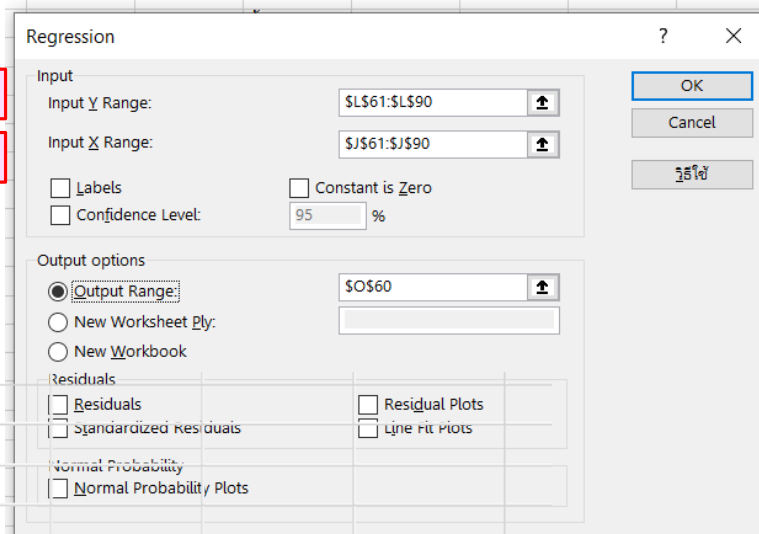


ค่าความสัมพันธ์ (สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson correlation coefficient) = .810 หมายถึงมีความสัมพันธ์กันสูง



ตัวแปรตาม Dependent Variable, DV: Y

ตัวแปรอิสระ Independent Variable, IV: X



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.810
R Square	0.656
Adjusted R	0.644
Standard E	20.409
Observatio	30.000

1-ค่าความสัมพันธ์ (R-Square) = .656 หมายถึงปริมาณสาร A ที่ใช้ (X) มีอิทธิพลต่อสมบัติ S (Y) ร้อยละ 65.6 (น้อย) อีกร้อยละ 34.4 เป็นอิทธิพลของปัจจัยอื่น

ANOVA

	df	SS	MS	F	Significance F
Regression	1	22279.033	22279.033	53.489	0.000
Residual	28	11662.433	416.515		
Total	29	33941.467			

2-ค่า Sig. หรือค่า P-value = .000 มีค่าน้อยกว่าค่า $\alpha = .05$ หมายถึง สมการถดถอยสามารถเขียนและใช้งานได้

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
a Intercept	41.633	19.435	2.142	0.041	1.823	81.443	1.823	81.443
b X Variable	21.244	2.905	7.314	0.000	15.294	27.194	15.294	27.194

2-ค่า Sig. หรือค่า P-value = .000 มีค่าน้อยกว่าค่า $\alpha = .05$ หมายถึง ปริมาณสาร A ที่ใช้ (X) สามารถทำนายสมบัติ S (Y) ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

สาร C ในร่างกาย (Y) = 41.633 + 21.244 (สมบัติ S (X))



? ปริมาณสาร A (X) สามารถทำนายผลของสาร C ในร่างกาย (Y) ได้หรือไม่

? เขียนสมการทำนาย

? พยากรณ์สาร C ในร่างกาย หากใช้สาร A มีปริมาณดังต่อไปนี้

ปริมาณสาร A	ปริมาณสาร C ในร่างกาย
1400	
1300	
1200	
1100	
3500	
3600	
3700	
3800	



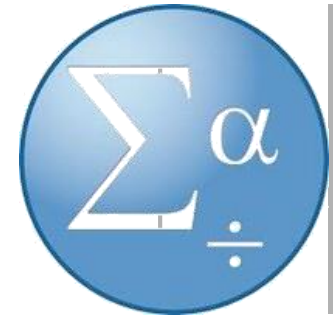
EXAMPLE

Simple Linear
Regression



ค่าสัมประสิทธิ์สหสัมพันธ์	ระดับความสัมพันธ์
(+-) 0.81-1.00	มาก
(+-) 0.51-0.80	ปานกลาง
(+-) 0.21-0.50	น้อย/ต่ำ
(+-) 0.01-0.20	น้อยมาก
0.00	ไม่สัมพันธ์





ทดสอบ 20 คะแนน

เรื่อง การวิเคราะห์การถดถอยอย่างง่าย (Simple Linear Regression)

ปัญหาที่พบ

- * ตัวเลขอ่านไม่ออก
- * เขียนสัญลักษณ์ไม่ชัดเจน =, +, -
- * ไม่ใช้ทศนิยม 3 ตำแหน่ง
- * ไม่แทนค่าตัวแปร X, Y
- * สกปรก ขีดฆ่าในสูตร/สมการ



**Probability and Statistics
for Engineers**

รศ.ดร.ฤดี นิยมรัตน์



**Probability and Statistics
for Engineers**