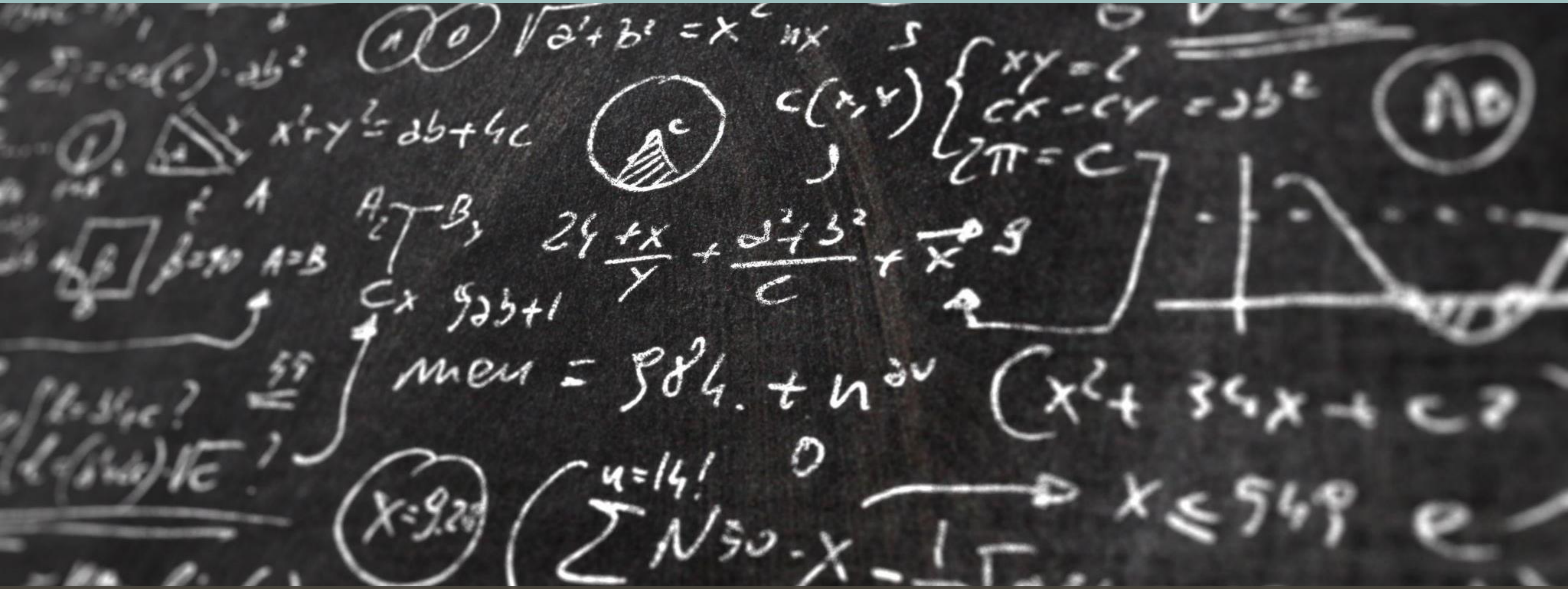


BASIC MATH & STATISTICS (PART III)

พัณณธัญญ์ วิจิตรวงศ์เจริญ

สารบัญ

- การทดสอบสมมติฐาน
- การวิเคราะห์ความสัมพันธ์
  - Scatter plot
  - Correlation
  - Regression
- Optimization Linear Programming

HYPOTHESIS TESTING (การทดสอบสมมติฐาน)

เมื่อต้องการศึกษาค่าพารามิเตอร์ของประชากร โดยการสุ่มตัวอย่างเพื่อตรวจสอบว่าเป็นไปตามที่กำหนดไว้เดิมหรือไม่

สถิติอย่างเดียว ไม่สามารถใช้ พิสูจน์สิ่งใดๆ ได้

เราใช้อนุมานทางสถิติเพื่อยอมรับ หรือ ปฏิเสธ คำอธิบายต่างๆ โดยอิงกับความเป็นไปได้

การอนุมาน

วัคซีนโควิด
ใช้ได้ดีหรือไม่

ลูกค้าสองกลุ่ม
มีพฤติกรรมเหมือนกันหรือไม่

แคมเปญโฆษณา
เพิ่มยอดขายได้จริงหรือไม่

HYPOTHESIS TESTING (การทดสอบสมมติฐาน)

1. แคมเปญโฆษณา A สามารถทำให้ลูกค้าซื้อมากกว่า 3 ชิ้น/เดือน

=> สมมติฐานเกี่ยวกับประชากร 1 กลุ่ม

2. ลูกค้าวัยรุ่นตอนต้น (13-18 ปี) มียอดใช้จ่ายในการซื้อบัตรคอนเสิร์ต มากกว่า วัยรุ่นตอนปลาย (19-25 ปี)

=> สมมติฐานเกี่ยวกับประชากร 2 กลุ่ม

3. เมื่อใช้วัคซีนโควิดชนิด A สามารถป้องกันโควิดได้ 80%

=> สมมติฐานเกี่ยวกับสัดส่วนประชากร 1 กลุ่ม

HYPOTHESIS TESTING (การทดสอบสมมติฐาน)

แคมเปญโฆษณา A สามารถทำให้ลูกค้าซื้อมากกว่า 3 ชิ้น/เดือน

$$\mu \geq 3$$

ครั้งที่ 1 สุ่มตัวอย่าง n = 40 คน , $\bar{X}$ = 5.5

ครั้งที่ 2 สุ่มตัวอย่าง n = 40 คน , $\bar{X}$ = 2.2

ครั้งที่ 3 สุ่มตัวอย่าง n = 40 คน , $\bar{X}$ = 3.1

ครั้งที่ 4 สุ่มตัวอย่าง n = 40 คน , $\bar{X}$ = 2.98

HYPOTHESIS TESTING (การทดสอบสมมติฐาน)

ตั้งข้อสมมติฐานไว้ 2 ข้อสมมติฐานคือ

- H_0 หรือข้อสมมติฐานหลัก (Null Hypothesis)
  - สิ่งที่คาดไว้ มีเครื่องหมาย $=$, $\geq$, $\leq$
- H_1 หรือ H_a หรือข้อสมมติฐานรอง หรือ ข้อสมมติฐานทางเลือก (Alternative Hypothesis)
  - มีเครื่องหมายตรงข้ามกับสมมติฐานหลัก $\neq$, $<$, $>$

ประเภทของการทดสอบสมมติฐาน

1. การทดสอบสองทาง (Two-sided test)
- สมมติฐาน H_1 มีเครื่องหมาย $\neq$
- เขตปฏิเสธการทดสอบจะมีสองด้าน

$$H_0 : \; \ldots = \ldots$$
$$H_1 : \; \ldots \neq \ldots$$

2. การทดสอบทางเดียว (One-sided test)

การทดสอบทางขวา (Right-tailed test)
- สมมติฐาน H_1 มีเครื่องหมาย $>$
- เขตปฏิเสธการทดสอบจะอยู่ทางด้านขวา

$$H_0 : \; \ldots \leq \ldots$$
$$H_1 : \; \ldots > \ldots$$

การทดสอบทางซ้าย (Left-tailed test)
- สมมติฐาน H_1 มีเครื่องหมาย $<$
- เขตปฏิเสธการทดสอบจะอยู่ทางด้านซ้าย

$$H_0 : \; \ldots \geq \ldots$$
$$H_1 : \; \ldots < \ldots$$

การตั้งสมมติฐาน

1.นักวิชาการเชื่อว่าเด็กไทยเฉลี่ยเล่นอินเตอร์เนต 4.5 ชั่วโมงต่อวัน

$$H_0: \mu = 4.5$$
$$H_1: \mu \neq 4.5$$

2.สัตวแพทย์คาดว่าสุนัขจะใช้เวลาฟื้นตัวหลังผ่าตัดเฉลี่ยอย่างน้อย 5 ชั่วโมง

$$H_0: \mu \geq 5$$
$$H_1: \mu < 5$$

การตั้งสมมติฐาน

3.โดยปกติ ยาแก้ปวดจะมีฤทธิ์เฉลี่ยน้อยกว่า 4 ชั่วโมง

$$H_0 : \mu \leq 4$$
$$H_1 : \mu > 4$$

4.นักเรียนชายและนักเรียนหญิงใช้เวลาทบทวนหนังสือก่อนสอบต่างกัน 50 นาที

$$H_0 : \mu_1 - \mu_2 = 50$$
$$H_1 : \mu_1 - \mu_2 \neq 50$$

การตั้งสมมติฐาน

5.ผู้จบปริญญาโทมีเงินเดือนสูงกว่าผู้จบปริญญาตรีในสาขาเดียวกันไม่น้อยกว่า 4,000 บาท

$$H_0 : \mu_1 - \mu_2 \geq 4,000$$
$$H_1 : \mu_1 - \mu_2 < 4,000$$

6.ชาวสวนทุเรียนในจังหวัดระยองมีรายได้เฉลี่ยแตกต่างจากชาวสวนในจังหวัดจันทบุรี

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

การตั้งสมมติฐาน

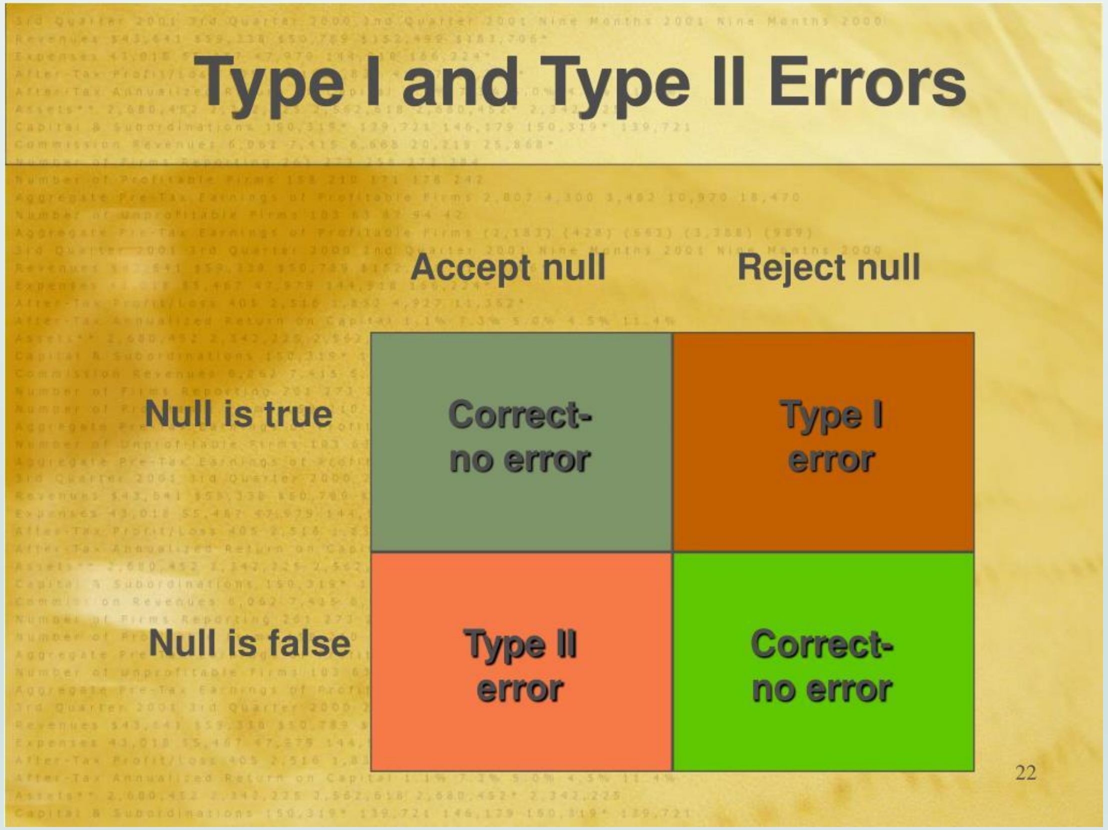
7.ประชาชนในเขตลาดกระบัง 45% เป็นคนกรุงเทพโดยกำเนิด

$$H_0 : p = 0.45$$
$$H_1 : p \neq 0.45$$

8.เด็กไทยมากกว่า 72% มีผลการทดสอบ IQ ต่ำกว่า 115

$$H_0 : p \leq 0.72$$
$$H_1 : p > 0.72$$

ความน่าจะเป็นของการเกิดข้อผิดพลาดจากการสรุปผล (TYPES OF ERROR)



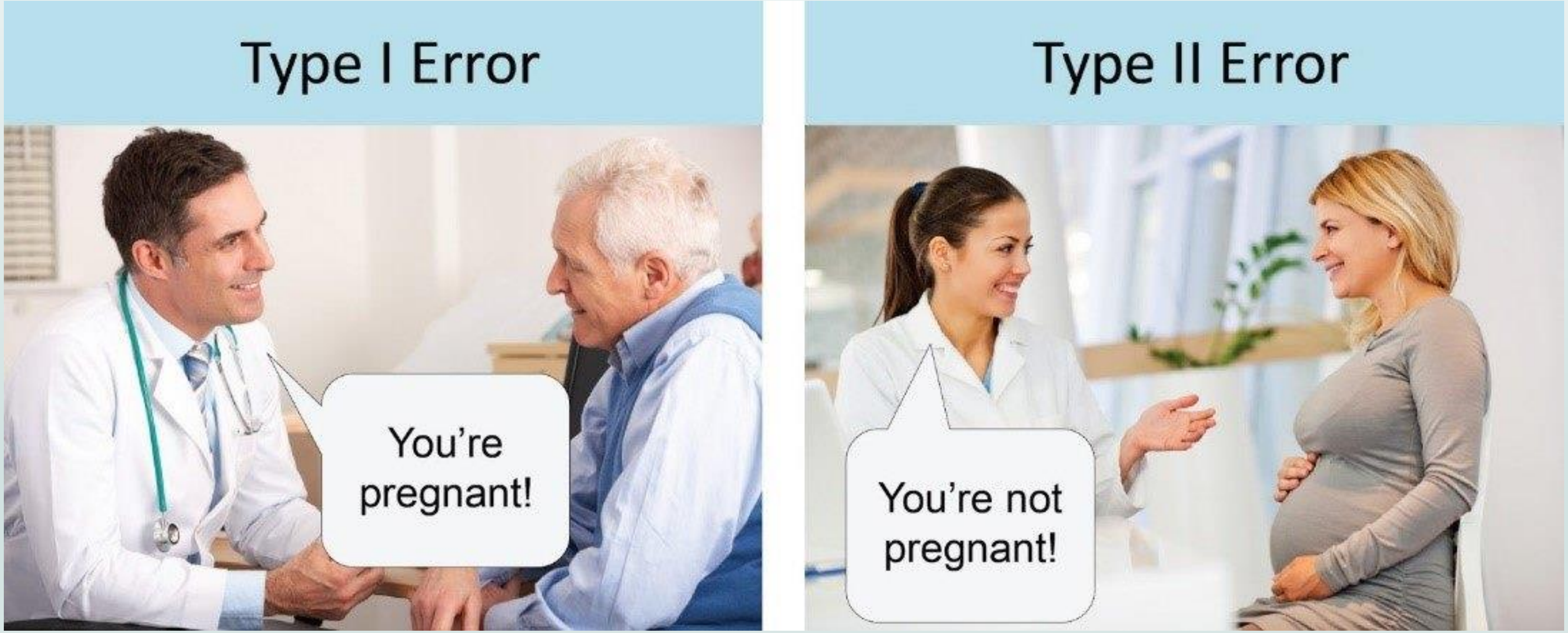
Type I : ผลบวกลวง (Positive False)

Type II : ผลลบลวง (Negative false)

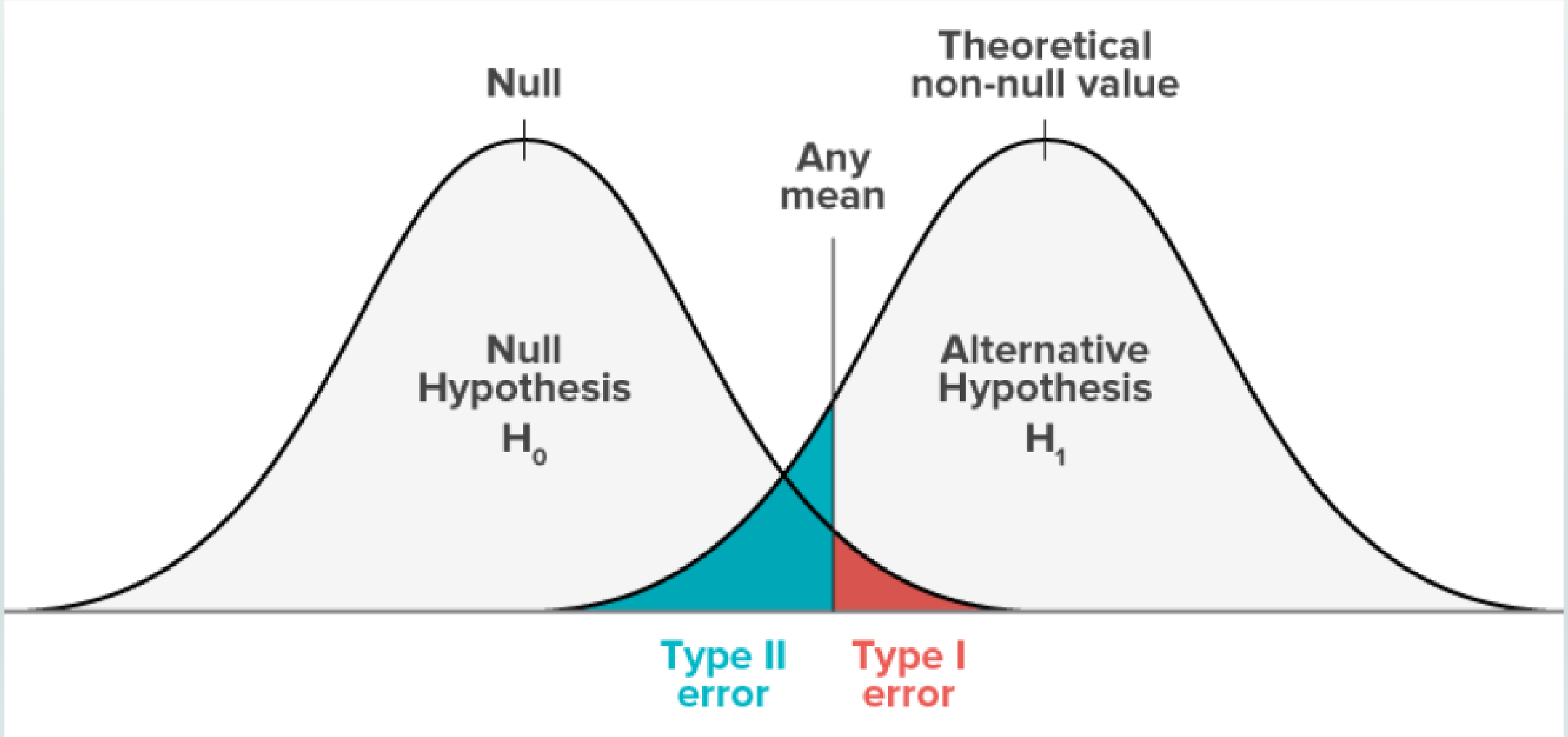
ความน่าจะเป็นของการเกิดข้อผิดพลาดจากการสรุปผล (TYPES OF ERROR)

สถานการณ์	Type I error ผลบวกลวง	Type II error ผลลบลวง
ตัวกรองสแปมอีเมลล์	กรองอีเมลล์ที่ไม่ใช้สแปมทิ้งไปด้วย	ปล่อยอีเมลล์สแปมเข้ามาในกล่องข้อความ
การตรวจคัดกรองโควิด	ผลเป็นบวก ทั้งๆที่ไม่ได้เป็น	ผลเป็นลบ ทั้งๆที่เป็น
การจับกุมผู้ต้องหา	ผู้บริสุทธิ์ติดคุก	ผู้ต้องหาลอยนวล

ความน่าจะเป็นของการเกิดข้อผิดพลาดจากการสรุปผล (TYPES OF ERROR)



ความน่าจะเป็นของการเกิดข้อผิดพลาดจากการสรุปผล (TYPES OF ERROR)



ความน่าจะเป็นของการเกิดข้อผิดพลาดจากการสรุปผล (TYPES OF ERROR)

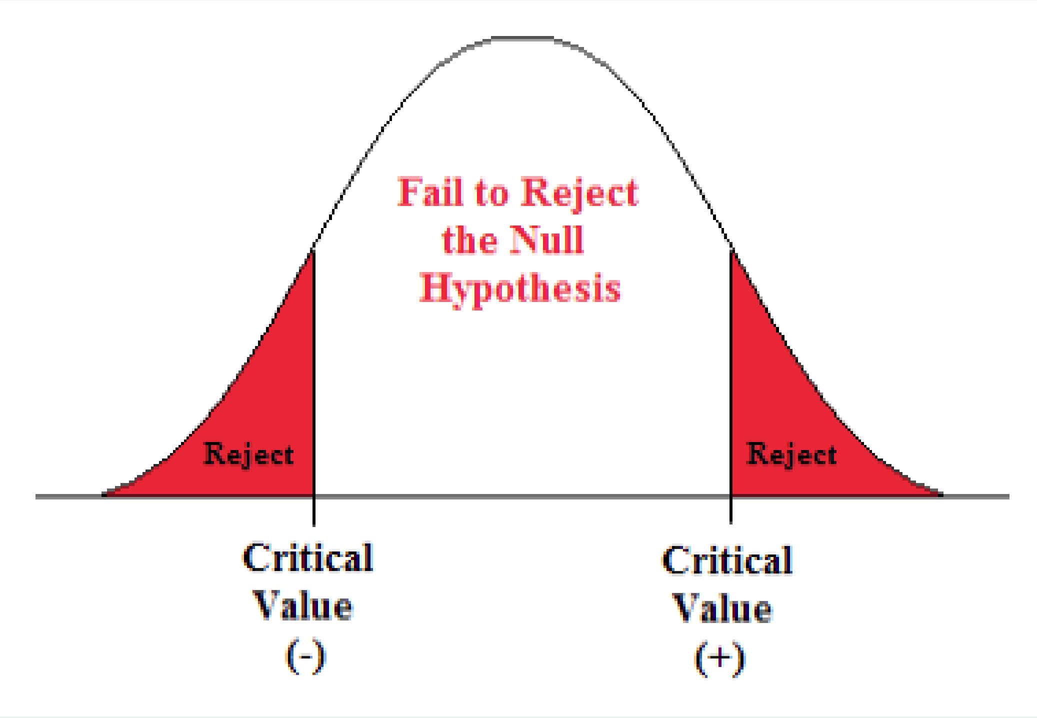
วิธีการลดค่าความผิดพลาด

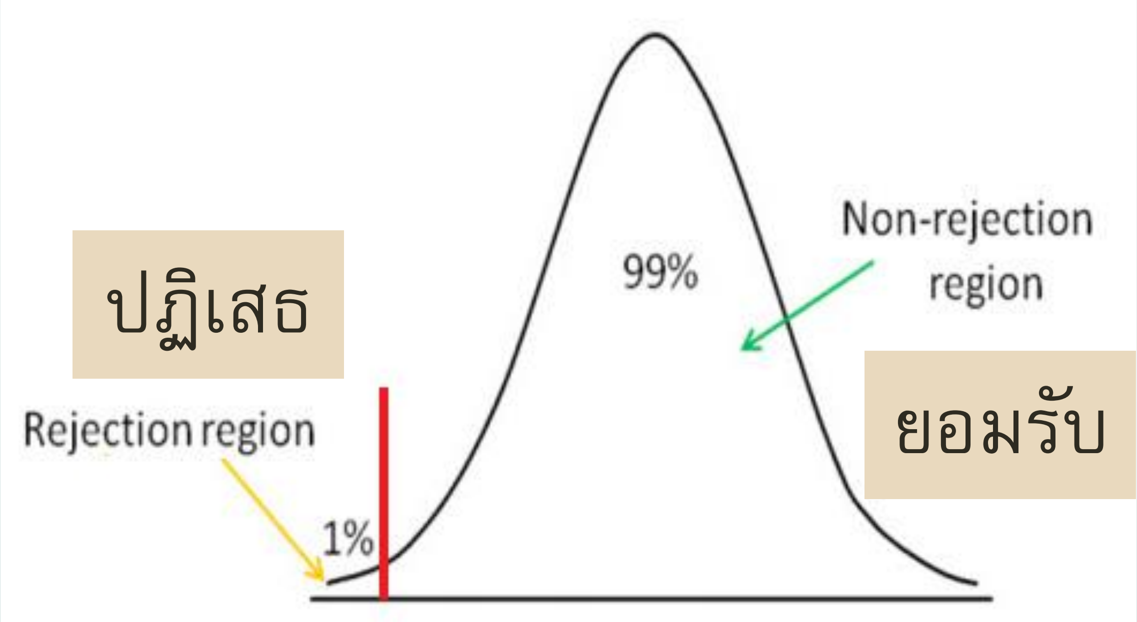
- เพิ่มช่วงการยอมรับ
- เพิ่มขนาดของตัวอย่าง **n**

ขั้นตอนการทดสอบสมมติฐาน

1) ตั้งสมมติฐานหลักและสมมติฐานรอง
2) กำหนดตัวสถิติทดสอบเป็น Z หรือ t
3) หาค่าวิกฤต z_α , $z_{\alpha/2}$, t_α, $t_{\alpha/2}$
4) กำหนดกฎของการตัดสินใจ
5) คำนวณค่าสถิติ
6) สรุปผล

กฎการตัดสินใจ





ปฏิเสธ H_0 เมื่อ $z < -z_{\alpha}$

ปฏิเสธ H_0 เมื่อ $z < -z_{\alpha/2}$ หรือ $z > z_{\alpha/2}$

การทดสอบค่าพารามิเตอร์

การทดสอบค่าเฉลี่ย

การทดสอบผลต่างของค่าเฉลี่ย

การทดสอบความเท่ากันของค่าเฉลี่ยเมื่อค่าสังเกตเป็นคู่

การทดสอบสัดส่วน

การทดสอบค่าความแปรปรวน

การทดสอบอัตราส่วนของค่าความแปรปรวน

I.การทดสอบสมมติฐานค่าเฉลี่ยประชากรหนึ่งกลุ่ม

สมมติฐานหลักและสมมติฐานรอง

$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$

สถิติทดสอบ

กรณีทราบความแปรปรวนประชากร (σ^2)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

กรณีไม่ทราบความแปรปรวนประชากร (σ^2)

n > 30

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

n < 30

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}},$$

$$d.f. = n - 1$$

ตัวอย่าง

ผู้จัดการรีสอร์ตกำหนดระดับความพึงพอใจของผู้ใช้บริการไว้ว่าจะต้องไม่ต่ำกว่า 4.2 คะแนน จึงต้องสำรวจความพึงพอใจอย่างสม่ำเสมอ เมื่อสำรวจผู้ใช้บริการจำนวน 49 คนพบว่ากลุ่มตัวอย่างมีความพึงพอใจ 4.02 คะแนน ค่าเบี่ยงเบนมาตรฐาน 1.24 จงทดสอบว่าเป็นไปอย่างที่กำหนดหรือไม่ ที่ระดับนัยสำคัญทางสถิติ 0.1

ตัวอย่าง

1) ตั้งสมมติฐานหลักและสมมติฐานรอง

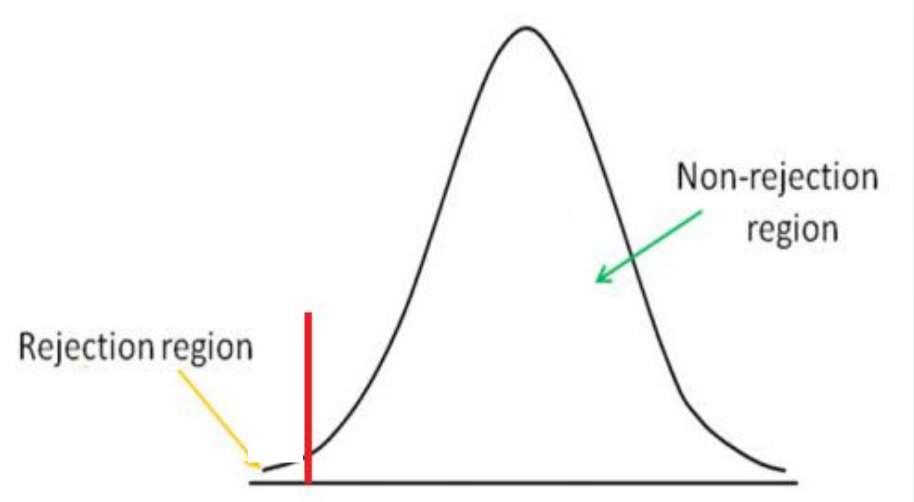
$$H_0 : \mu \geq 4.2$$
$$H_1 : \mu < 4.2$$

2) กำหนดสถิติทดสอบเป็น z

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

3) หาค่าวิกฤต

- เป็นการทดสอบทางเดียวแบบทางซ้าย ใช้ค่า z_α
- ที่ $\alpha = 0.1$ เปิดตาราง ค่า $z_\alpha = 1.28$



$-z_\alpha$ = -1.28

ตัวอย่าง

4) กำหนดกฎของการตัดสินใจ

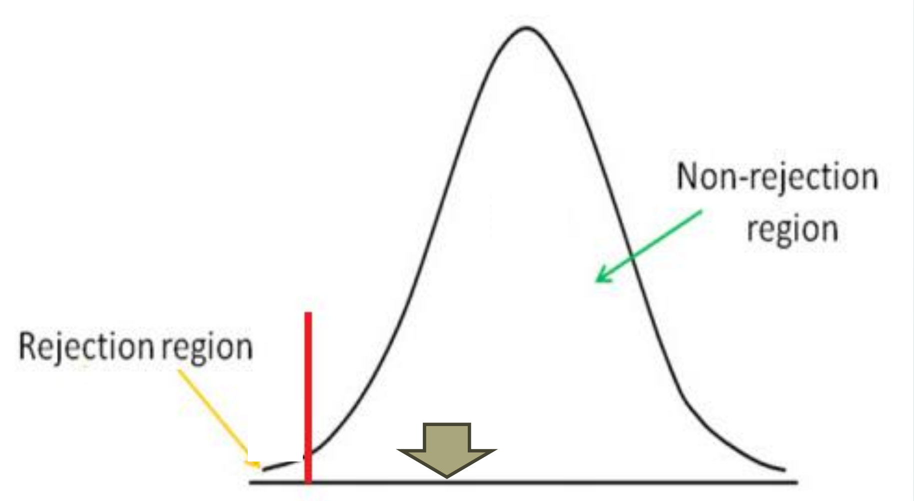
จะปฏิเสธ H_0 เมื่อ $z < -1.28$

5) คำนวณค่าสถิติทดสอบ

$$z = \frac{4.02 - 4.2}{1.24/\sqrt{49}}$$
$$= -1.02$$

6) สรุปผล

ยอมรับ H0 ความพึงพอใจของผู้ใช้บริการไม่ต่ำกว่า 4.2 ที่ระดับนัยสำคัญ 0.1



-1.28 -1.02

II.การทดสอบสมมติฐานของผลต่างค่าเฉลี่ยประชากรสองกลุ่ม

สมมติฐานหลักและสมมติฐานรอง

$H_0: \mu_1 - \mu_2 = \mu_0$ $H_1: \mu_1 - \mu_2 \neq \mu_0$	$H_0: \mu_1 - \mu_2 \geq \mu_0$ $H_1: \mu_1 - \mu_2 < \mu_0$	$H_0: \mu_1 - \mu_2 \geq \mu_0$ $H_1: \mu_1 - \mu_2 < \mu_0$

<u>สถิติทดสอบ</u>

กรณีทราบความแปรปรวนประชากร (σ_1^2, σ_2^2)

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

กรณีไม่ทราบความแปรปรวนประชากร (σ_1^2, σ_2^2)

$n > 30$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$$

ตัวอย่าง

สุ่มตัวอย่างหลอดไฟจากบริษัท A 60 หลอด ปรากฏว่ามีอายุการใช้งานเฉลี่ย 1,300 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 90 ชั่วโมง และสุ่มตัวอย่างบริษัท B มา 50 หลอด ปรากฏว่ามีอายุการใช้งานเฉลี่ย 1,060 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 70 ชั่วโมง จงทดสอบสมมติฐานว่า หลอดไฟจากบริษัท A มีอายุการใช้งานนานกว่าหลอดไฟจากบริษัท B มากกว่า 200 ชั่วโมง ที่ระดับนัยสำคัญทางสถิติ 0.01 หรือไม่

ตัวอย่าง

1) ตั้งสมมติฐานหลักและสมมติฐานรอง

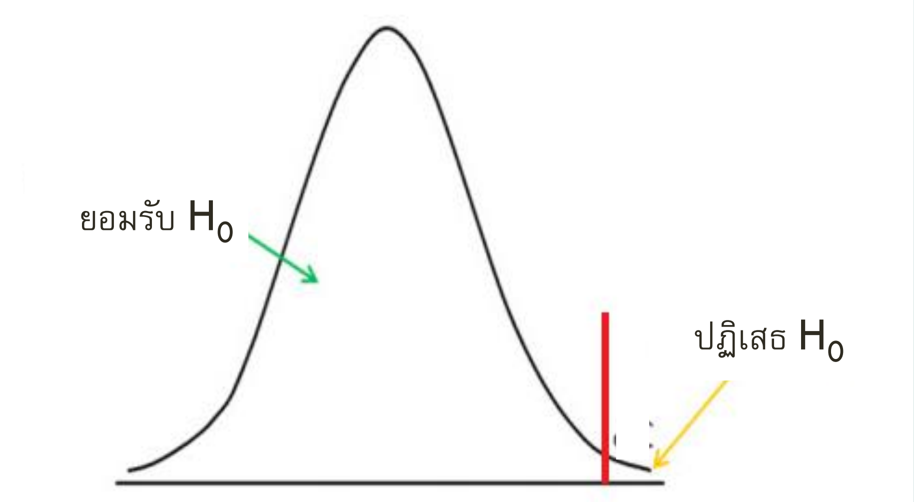
$$H_0 : \mu_1 - \mu_2 \geq \mu_0$$
$$H_1 : \mu_1 - \mu_2 < \mu_0$$

2) กำหนดสถิติทดสอบเป็น z

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$$

3) หาค่าวิกฤต

- เป็นการทดสอบทางเดียวแบบทางขวา ใช้ค่า z_α
- ที่ $\alpha = 0.01$ เปิดตาราง ค่า $z_\alpha = 2.33$



$z_\alpha = 2.33$

ตัวอย่าง

4) กำหนดกฎของการตัดสินใจ

จะปฏิเสธ H_0 เมื่อ $z > 2.33$

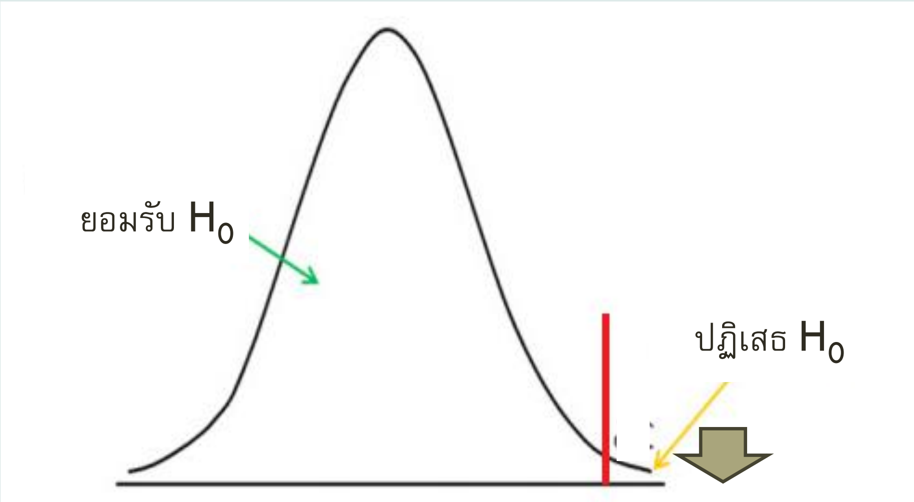
5) คำนวณค่าสถิติทดสอบ

$$z = \frac{(1300 - 1060) - (200)}{\sqrt{\frac{90^2}{60} + \frac{70^2}{50}}}$$

$$= 2.62$$

6) สรุปผล

ปฏิเสธ H0 หลอดไฟจากบริษัท A มีอายุการใช้งานนานกว่าหลอดไฟจากบริษัท B มากกว่า 200 ชม. ที่นัยสำคัญ 0.01



2.33 2.62

III.การทดสอบสมมติฐานของค่าสัดส่วนประชากร

สมมติฐานหลักและสมมติฐานรอง

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

$$H_0 : p \geq p_0$$
$$H_1 : p < p$$

$$H_0 : p \geq p_0$$
$$H_1 : p < p_0$$

<u>สถิติทดสอบ</u>

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

เมื่อ q = 1-p

ตัวอย่าง

จากการสำรวจโพลของ ABC เชื่อว่า 72% ของคนที่เล่นอินเตอร์เน็ต จะถูกขโมยข้อมูลทางอินเตอร์เน็ต จึงทำการสุ่มตัวอย่างนักศึกษาที่เล่นอินเตอร์เน็ตจำนวน 300 คน จากมหาวิทยาลัยแห่งหนึ่ง พบว่า 228 คน ถูกขโมยข้อมูล จงทดสอบความเชื่อนี้ที่ระดับนัยสำคัญ 0.10

ตัวอย่าง

1) ตั้งสมมติฐานหลักและสมมติฐานรอง

$$H_0 : p = 0.72$$
$$H_1 : p \neq 0.72$$

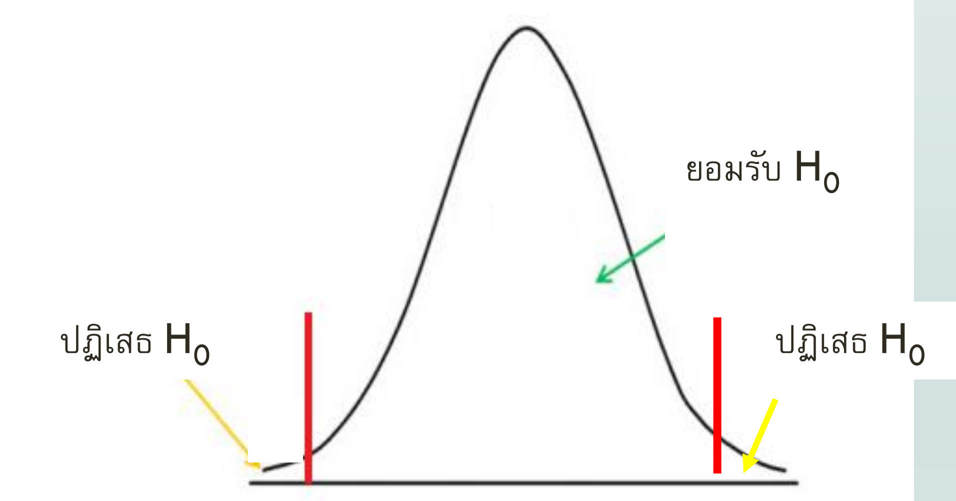
2) กำหนดสถิติทดสอบเป็น z

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

3) หาค่าวิกฤต

- เป็นการทดสอบสองทางใช้ค่า $z_{\alpha/2}$
- ที่ $\alpha = 0.1$, $\frac{\alpha}{2} = 0.05$ เปิดตาราง

ค่า $z_{\alpha/2} = 1.645$



$-z_\alpha$ = -1.645 $z_\alpha = 1.645$

ตัวอย่าง

4) กำหนดกฎของการตัดสินใจ

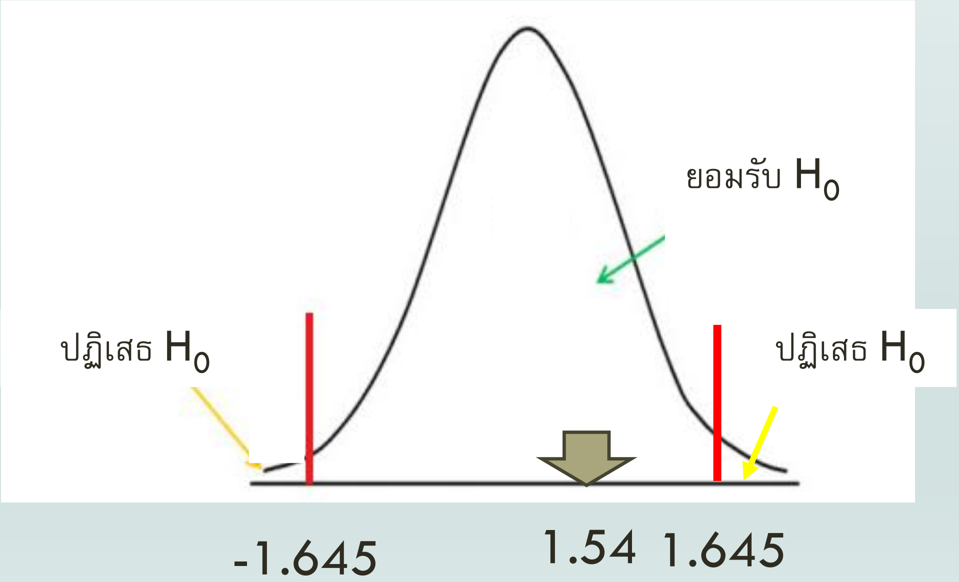
จะปฏิเสธ H_0 เมื่อ $z < -1.645$ หรือ $z > 1.645$

5) คำนวณค่าสถิติทดสอบ

$$z = \frac{0.76 - 0.72}{300} = 1.54$$

6) สรุปผล

ยอมรับ H0 มีคนที่เล่นอินเทอร์เน็ตและถูกขโมยข้อมูลทางอินเตอร์เน็ต 72% จริงที่ระดับนัยสำคัญ 0.10



P-VALUE

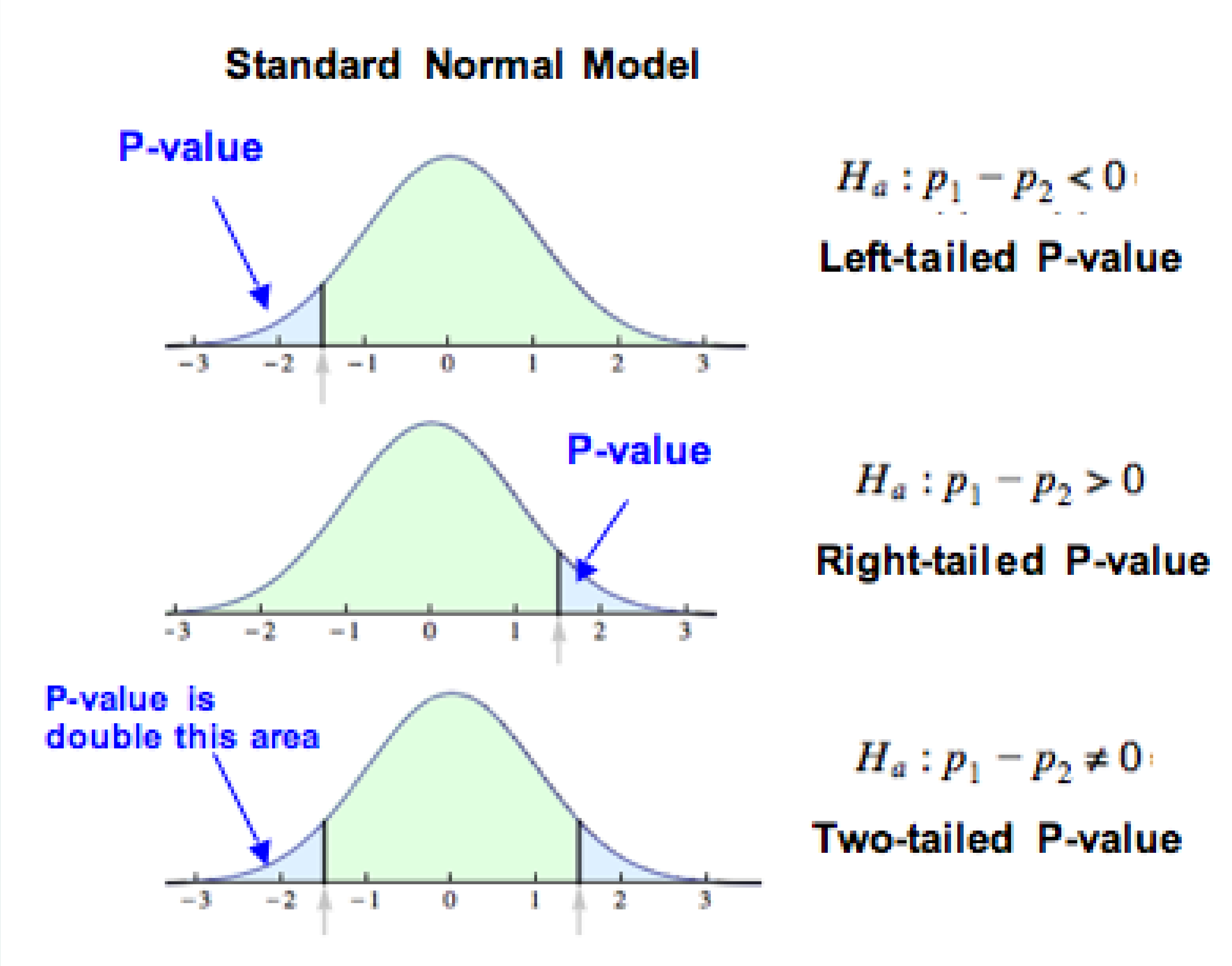
ความน่าจะเป็นที่น้อยที่สุด ที่สามารถจะปฏิเสธสมมติฐานหลักได้

ระดับความเชื่อมั่นสูงสุดที่จะสามารถปฏิเสธสมมติฐานหลักได้ (1-P-value)

$$P - Value = \Pr(Z > |z_c|)$$

ค่าความน่าจะเป็นที่ค่าสถิติที่เหมาะสมนั้นมีค่ามากกว่าค่าสัมบูรณ์ของค่าสถิติที่คำนวณได้

P-VALUE


Standard Normal Model
P-value
$H_a: p_1 - p_2 < 0$
Left-tailed P-value
P-value
$H_a: p_1 - p_2 > 0$
Right-tailed P-value
P-value is double this area
$H_a: p_1 - p_2 \neq 0$
Two-tailed P-value
-3 -2 -1 0 1 2 3

P-VALUE

กรณีที่	สรุปผล
P-Value < α	Reject H_0 (ปฏิเสธสมมติฐานหลัก)
P-Value > α	Accept H_0 (ยอมรับสมมติฐานหลัก)

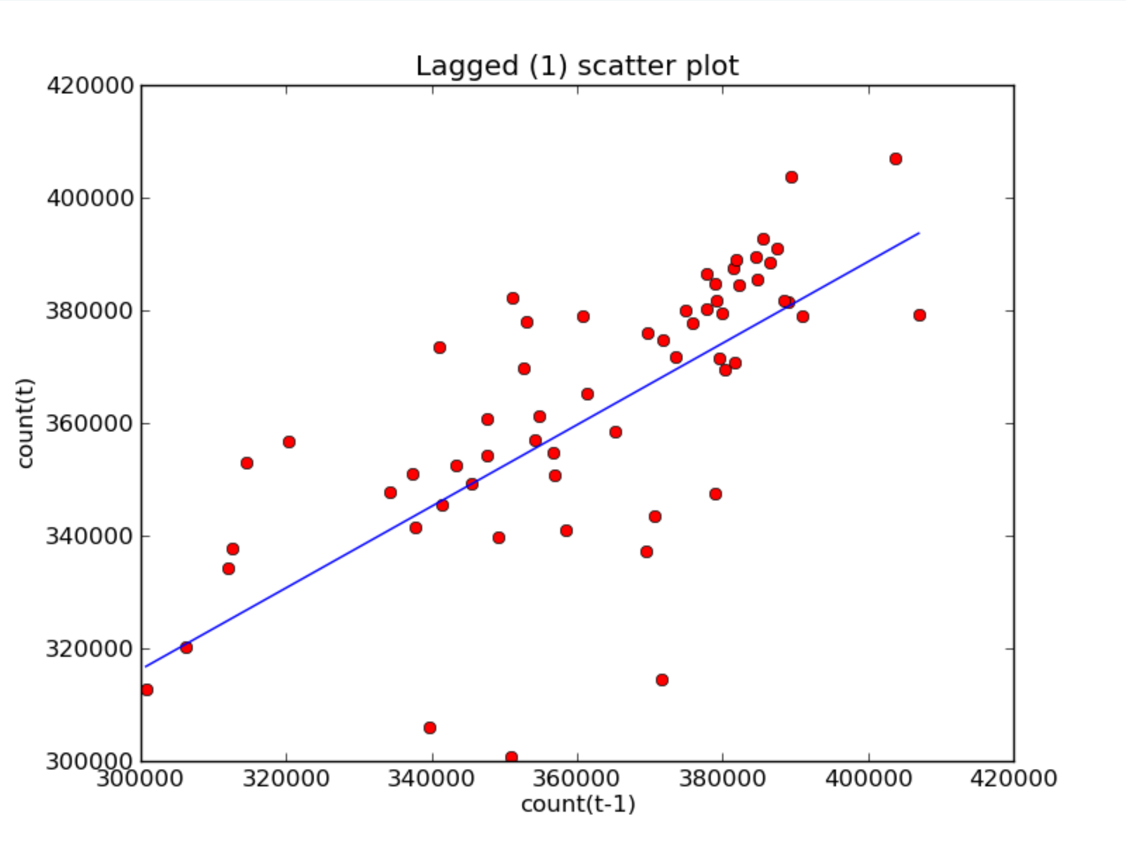
การวิเคราะห์ความสัมพันธ์

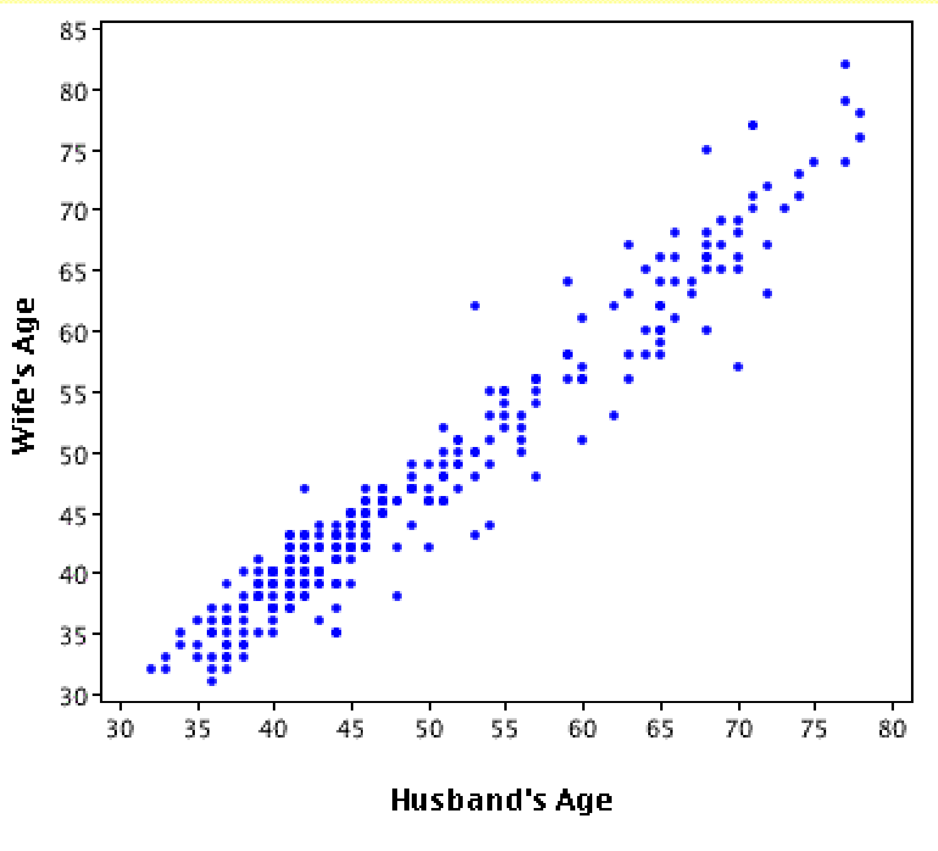
การวิเคราะห์ความสัมพันธ์

ตัวแปร X กับ Y มีความสัมพันธ์กันอย่างไร

- **Netflix**
- อุณหภูมิ กับ ผลผลิต
- ฤดูกาล กับ ยอดขาย

SCATTER PLOT (การกระจาย)





CORRELATION สหสัมพันธ์

ทำหน้าที่ชี้วัดว่า ปรากฏการณ์สองอย่างสัมพันธ์กันเพียงใด

อุณหภูมิในฤดูร้อน	ยอดขายไอศครีม	+
ส่วนสูง	น้ำหนัก	-
การออกกำลังกาย	น้ำหนัก	-

CORRELATION สหสัมพันธ์

Correlation VS Causation

vs

ยอดขายชุดว่ายน้ำ

ยอดขายไอศครีม

CORRELATION สหสัมพันธ์

- **Linear Correlation** สหสัมพันธ์เชิงเส้น เป็นการศึกษาว่ามีความสัมพันธ์เป็นเส้นตรงมากน้อยเพียงใด

- สัมประสิทธิ์สหสัมพันธ์ **Coefficient of Correlation**

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

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples y_i = y variable sample

$\bar{x}$ = mean of values in x variable $\bar{y}$ =mean of values in y variable

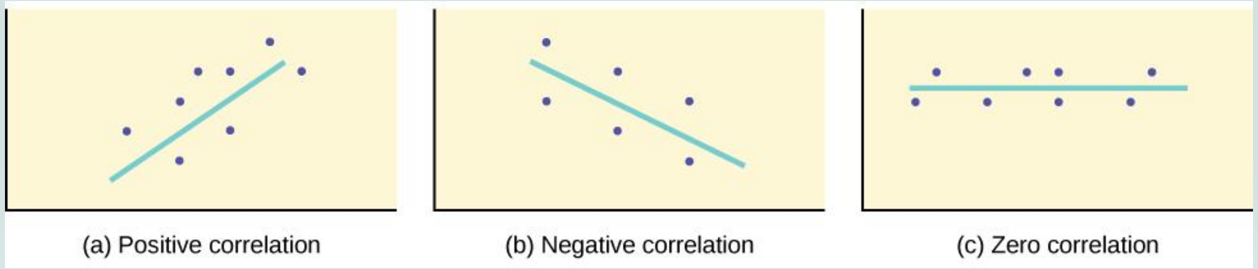
CORRELATION สหสัมพันธ์

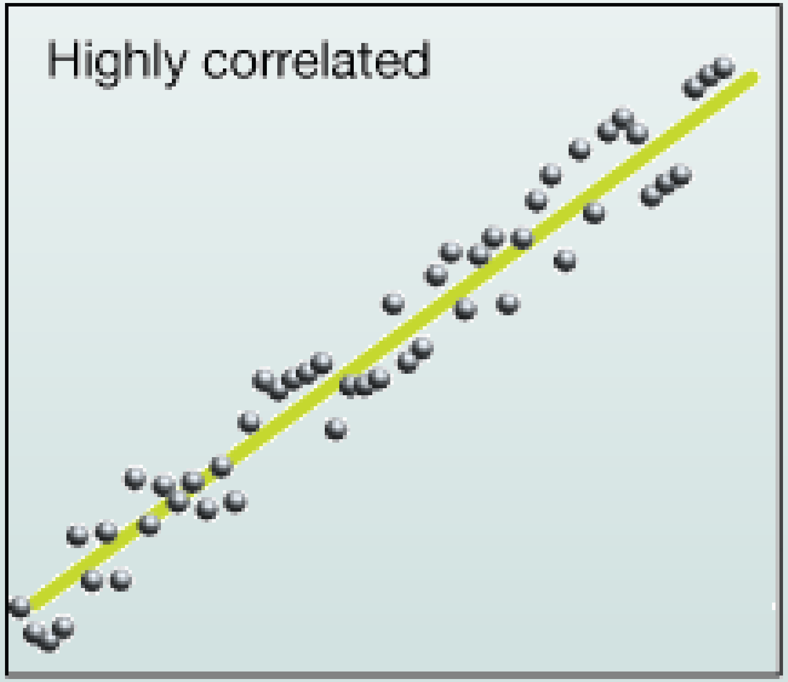
- มีค่าระหว่าง -1 กับ 1

ค่าเข้าใกล้ 1 => สัมพันธ์เชิงบวกอย่างสมบูรณ์

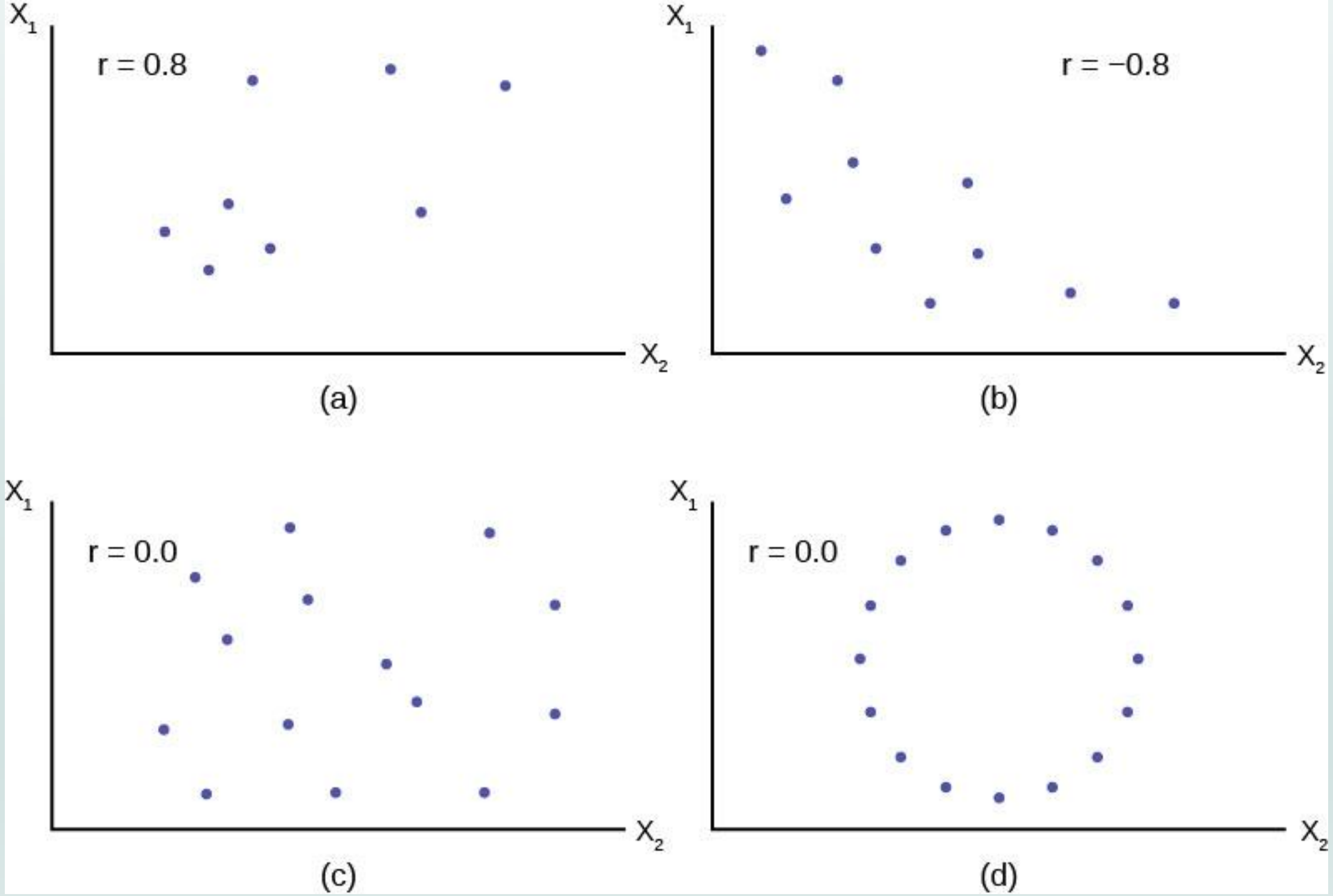
ค่าเข้าใกล้ -1 => สัมพันธ์เชิงลบอย่างสมบูรณ์

- ไม่ยึดติดกับหน่วยวัดใดๆ

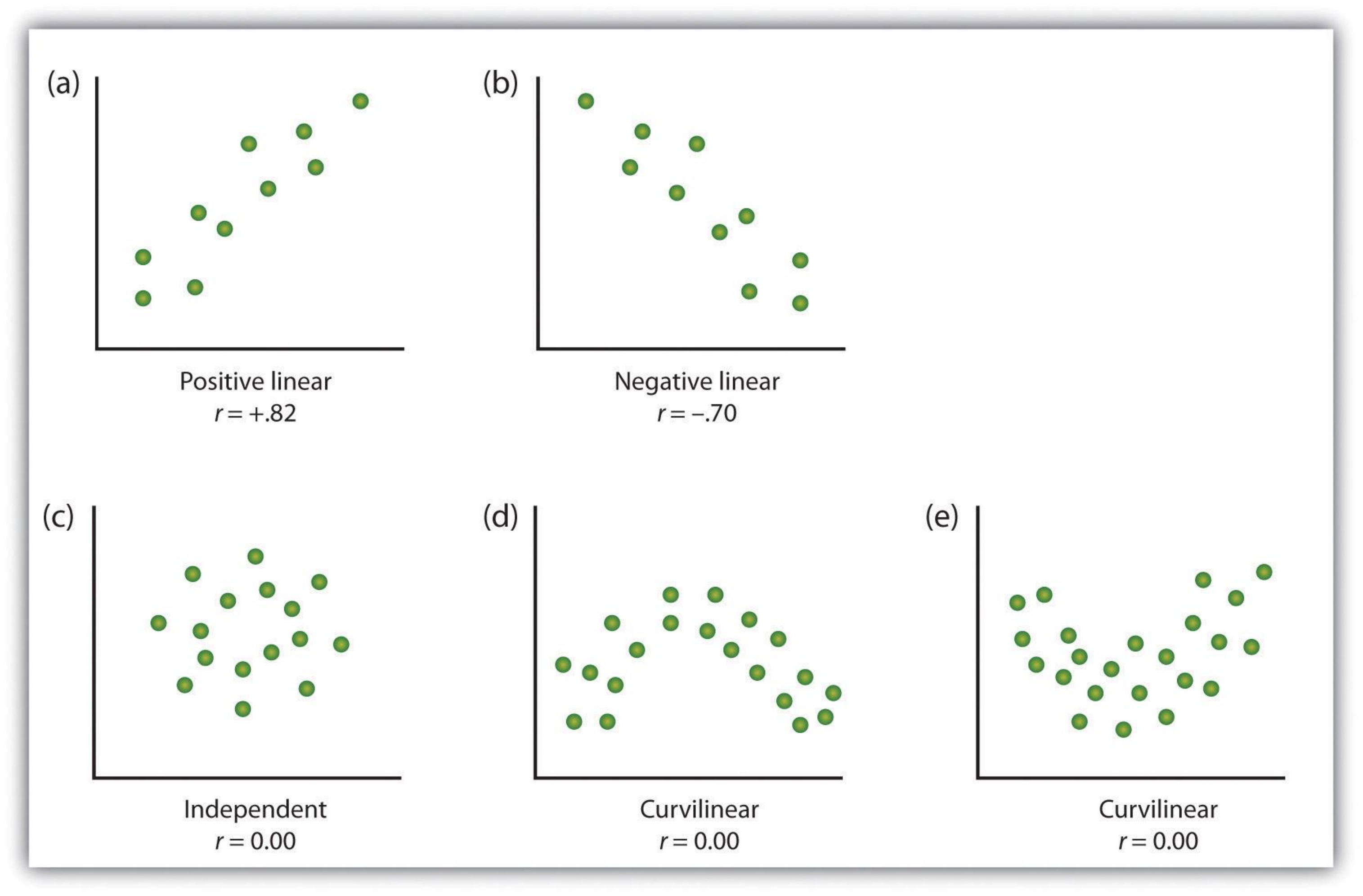




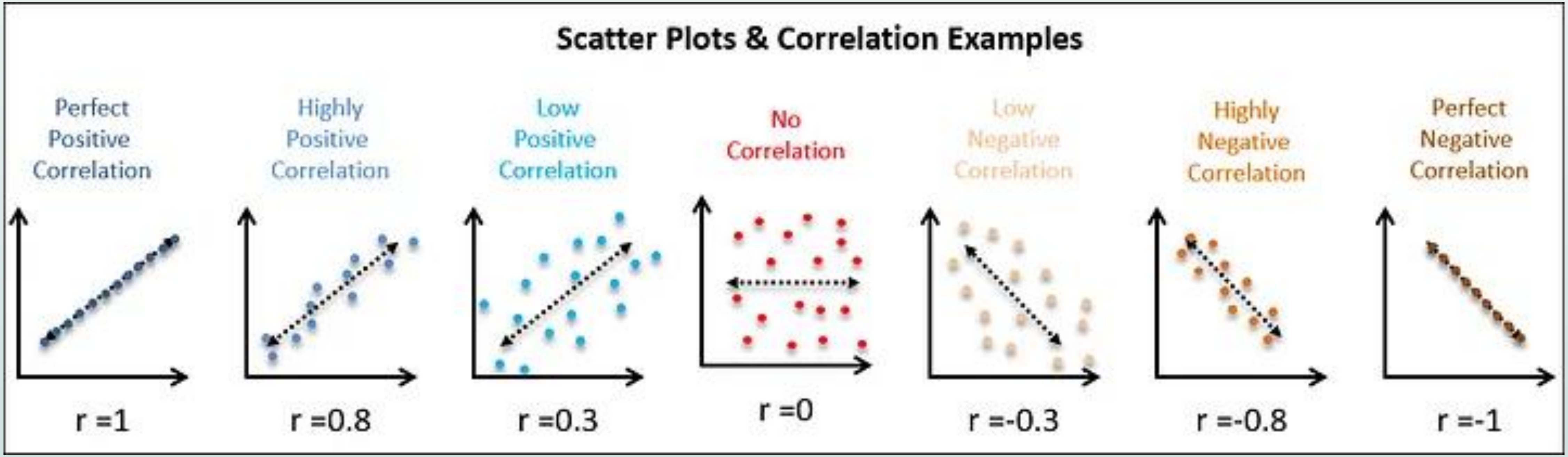
CORRELATION สหสัมพันธ์



CORRELATION สหสัมพันธ์



CORRELATION สหสัมพันธ์



ตัวอย่าง

เจ้าของร้านน้ำแข็งใส อยากรู้ว่าอุณหภูมิมีความสัมพันธ์กับปริมาณน้ำแข็งใสที่ถูกซื้อไปหรือไม่

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

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples

y_i = y variable sample

$\bar{x}$ = mean of values in x variable

$\bar{y}$ =mean of values in y variable

เวลาที่	อุณหภูมิ °C (x_i)	#น้ำแข็งใส (y_i)
1	23	66
2	26	75
3	25	70
4	30	81
5	33	86
6	26	72
7	26	90
8	30	80
9	25	73
10	29	78
11	41	99
12	37	95
13	23	60
14	23	63
15	33	86

เวลาที่	อุณหภูมิ °C (xi)	#น้ำแข็งใส (yi)	xi_bar	yi_bar	xi - xi_bar (Xi)	yi - yi_bar (Yi)	Xi*Yi	∑Xi*Yi	Xi^2	Yi^2	$\sum Xi^2$	$\sum Yi^2$	sqrt_$\sum Xi^2 * \sum Yi^2$	Correlation Coefficient (r)
1	23	66	28.67	78.27	-5.67	-12.27	69.51	771.33	32.11	150.47	407.33	1840.93	865.95	0.89
2	26	75	28.67	78.27	-2.67	-3.27	8.71	771.33	7.11	10.67	407.33	1840.93	865.95	0.89
3	25	70	28.67	78.27	-3.67	-8.27	30.31	771.33	13.44	68.34	407.33	1840.93	865.95	0.89
4	30	81	28.67	78.27	1.33	2.73	3.64	771.33	1.78	7.47	407.33	1840.93	865.95	0.89
5	33	86	28.67	78.27	4.33	7.73	33.51	771.33	18.78	59.80	407.33	1840.93	865.95	0.89
6	26	72	28.67	78.27	-2.67	-6.27	16.71	771.33	7.11	39.27	407.33	1840.93	865.95	0.89
7	26	90	28.67	78.27	-2.67	11.73	-31.29	771.33	7.11	137.67	407.33	1840.93	865.95	0.89
8	30	80	28.67	78.27	1.33	1.73	2.31	771.33	1.78	3.00	407.33	1840.93	865.95	0.89
9	25	73	28.67	78.27	-3.67	-5.27	19.31	771.33	13.44	27.74	407.33	1840.93	865.95	0.89
10	29	78	28.67	78.27	0.33	-0.27	-0.09	771.33	0.11	0.07	407.33	1840.93	865.95	0.89
11	41	99	28.67	78.27	12.33	20.73	255.71	771.33	152.11	429.87	407.33	1840.93	865.95	0.89
12	37	95	28.67	78.27	8.33	16.73	139.44	771.33	69.44	280.00	407.33	1840.93	865.95	0.89
13	23	60	28.67	78.27	-5.67	-18.27	103.51	771.33	32.11	333.67	407.33	1840.93	865.95	0.89
14	23	63	28.67	78.27	-5.67	-15.27	86.51	771.33	32.11	233.07	407.33	1840.93	865.95	0.89
15	33	86	28.67	78.27	4.33	7.73	33.51	771.33	18.78	59.80	407.33	1840.93	865.95	0.89

ค่า **Correlation Coefficient** = 0.89

อุณหภูมิมีความสัมพันธ์กับปริมาณการซื้อน้ำแข็งใส

CORRELATION สหสัมพันธ์

Netflix Example

Suppose we have 4 people who have rated movies on a scale of 1 to 5 stars with 1 being disliking the movie and 5 being loving the movie.

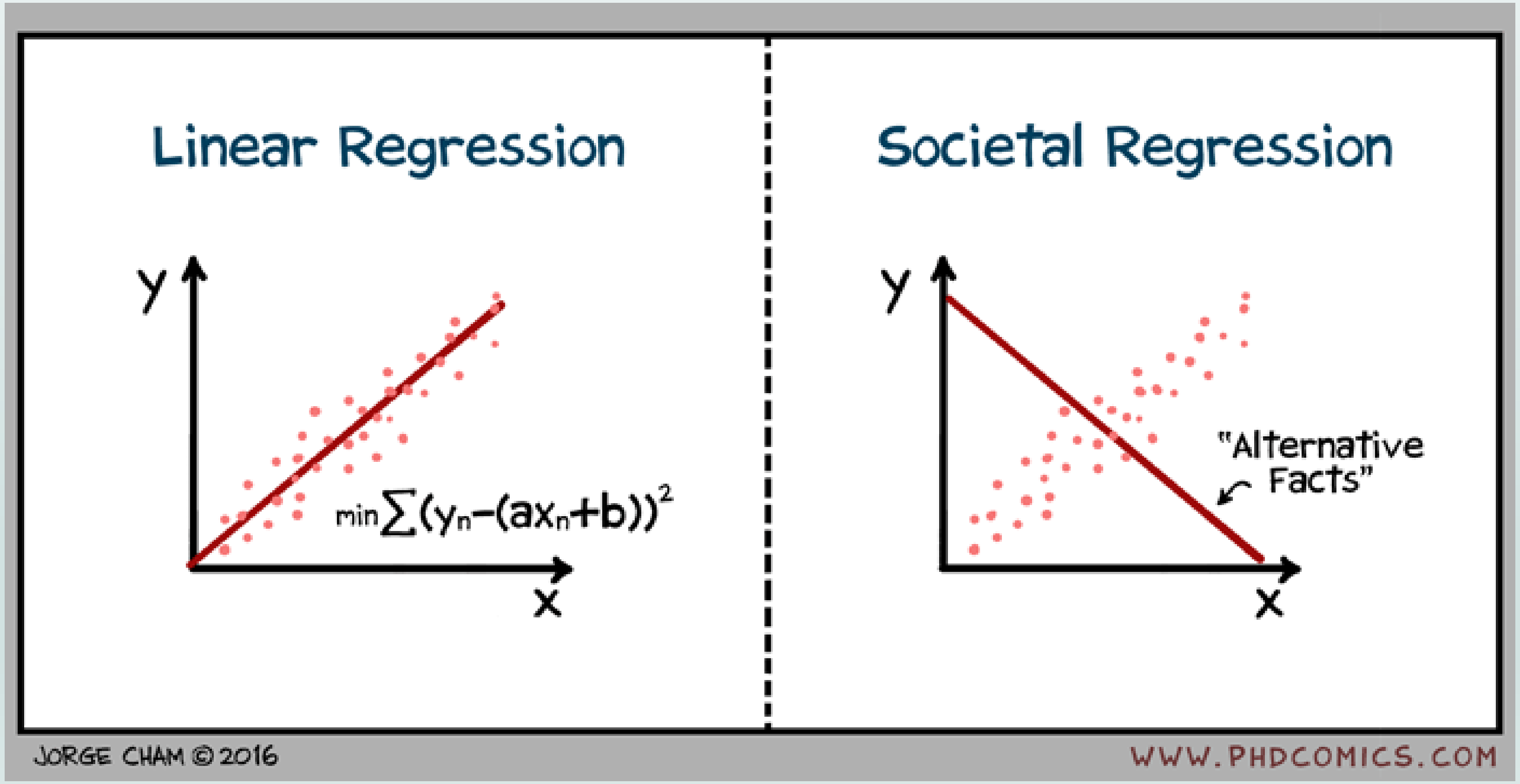
Movie	Adam	Lindsay	Austin	Sarah
Top Gun	4	1	5	2
Jurassic Park	5	2	5	3
Office Space	5	3	5	1
Message in a Bottle	1	4	1	5
Sleepless in Seattle	1	5	1	1
Titanic	4	1	5	3
Predator	5	2	5	2
Terminator	5	3	5	2
Anchorman	5	4	5	2

CORRELATION สหสัมพันธ์

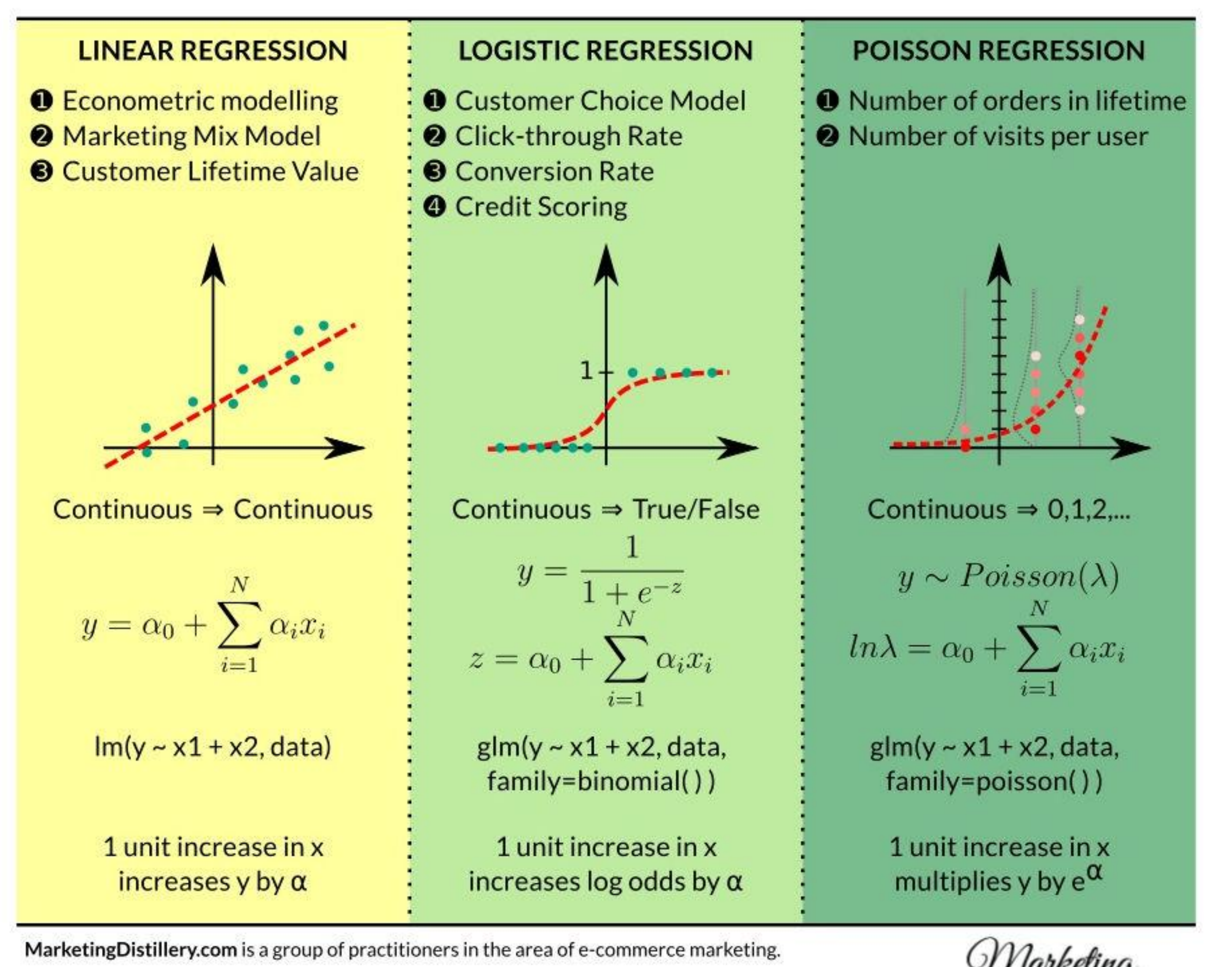
ค่า r

	Adam	Lindsay	Austin	Sarah
Adam	1	-0.54	0.97	-0.34
Lindsay	-0.54	1	-0.7	-0.1
Austin	0.97	-0.7	1	-0.31
Sarah	-0.34	-0.1	-0.31	1

REGRESSION สมการถดถอย



REGRESSION สมการถดถอย



LINEAR REGRESSION การถดถอยเชิงเส้น

เป็นการอธิบายลักษณะความสัมพันธ์ระหว่างตัวแปร ว่าตัวแปรตัวหนึ่งหรือหลายตัว มีอิทธิพลต่อตัวแปรอีกตัวหนึ่งอย่างไร

X ตัวแปรต้น (**Independent variable)** : ควบคุมและวัดค่าได้

Y ตัวแปรตาม **(Dependent variable)** : เปลี่ยนแปลงไปตามการเปลี่ยนแปลงของตัวแปรต้น

ใช้ในการสร้างแบบจำลองเพื่อพยากรณ์ค่าต่าง ๆ

Simple linear regression :

$$Y = aX + b$$

a คือ ค่าความชันของเส้นตรง

b คือ ค่าที่ตัดกับเส้นแกน Y

LINEAR REGRESSION การถดถอยเชิงเส้น

การประมาณค่า a กับ b โดยใช้วิธีกำลังสองน้อยที่สุด (Least square estimation)

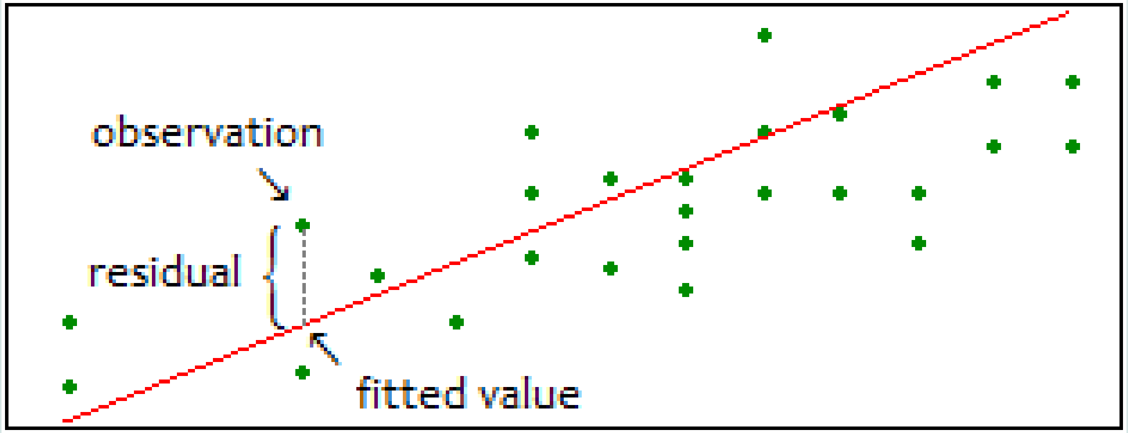
$$Y = aX + b$$

a คือ ค่าความชันของเส้นตรง

b คือ ค่าที่ตัดกับเส้นแกน Y

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

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



ตัวอย่าง

การศึกษาราคาบ้าน (Housing Price)

บ้านหลังที่	ขนาดบ้าน (X)	ราคาบ้าน (Y)
1	1705	30859
2	1785	39759
3	1430	33232
4	1429	24158
5	2120	56310
6	2002	46101
7	1460	25727
8	1787	34780
9	2021	45137
10	2049	47411

ตัวอย่าง

บ้านหลังที่	ขนาดบ้าน (xi)	ราคาบ้าน (yi)	x_bar	y_bar	Xi = xi-x_bar	Yi = yi-y_bar	Xi*Yi	ΣXi*Yi	Xi^2	ΣXi^2	a_hat
1	1705	30859	1778.8	38347.4	-73.8	-7488.4	552643.92	23256098.8	5446.44	649111.6	35.828
2	1785	39759	1778.8	38347.4	6.2	1411.6	8751.92	23256098.8	38.44	649111.6	
3	1430	33232	1778.8	38347.4	-348.8	-5115.4	1784251.52	23256098.8	121661.44	649111.6	
4	1429	24158	1778.8	38347.4	-349.8	-14189.4	4963452.12	23256098.8	122360.04	649111.6	
5	2120	56310	1778.8	38347.4	341.2	17962.6	6128839.12	23256098.8	116417.44	649111.6	
6	2002	46101	1778.8	38347.4	223.2	7753.6	1730603.52	23256098.8	49818.24	649111.6	
7	1460	25727	1778.8	38347.4	-318.8	-12620.4	4023383.52	23256098.8	101633.44	649111.6	
8	1787	34780	1778.8	38347.4	8.2	-3567.4	-29252.68	23256098.8	67.24	649111.6	
9	2021	45137	1778.8	38347.4	242.2	6789.6	1644441.12	23256098.8	58660.84	649111.6	
10	2049	47411	1778.8	38347.4	270.2	9063.6	2448984.72	23256098.8	73008.04	649111.6	

ตัวอย่าง

Housing Price
y = 35.828x - 25383
60000
55000
50000
45000
40000
35000
30000
25000
20000
15000
10000
1000
1200
1400
1600
1800
2000
2200

ตัวอย่าง

บ้านหลังที่	ขนาดบ้าน (X)	ราคาบ้าน (Y)	ราคาบ้านคาดการณ์ (Y_hat)	Error Term (e = Y-Y_hat)	e^2	LSE (Σe^2)
1	1705	30859	35703.74	-4844.74	23471505.67	135,387,027.64
2	1785	39759	38569.98	1189.02	1413768.56	
3	1430	33232	25851.04	7380.96	54478570.52	
4	1429	24158	25815.212	-1657.212	2746351.613	
5	2120	56310	50572.36	5737.64	32920512.77	
6	2002	46101	46344.656	-243.656	59368.24634	
7	1460	25727	26925.88	-1198.88	1437313.254	
8	1787	34780	38641.636	-3861.636	14912232.6	
9	2021	45137	47025.388	-1888.388	3566009.239	
10	2049	47411	48028.572	-617.572	381395.1752	

MULTIPLE REGRESSION

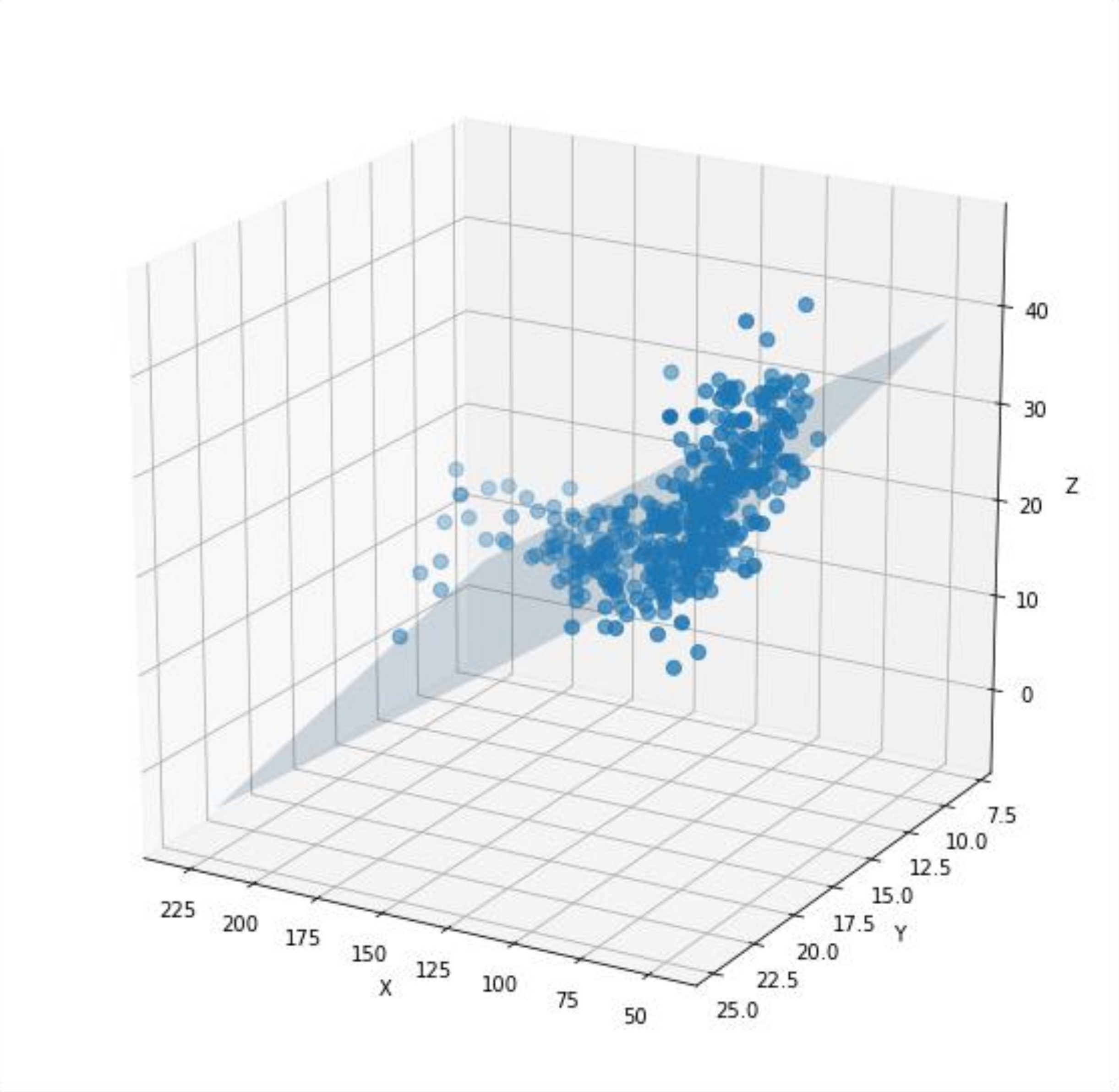
ในความเป็นจริงแล้วเราไม่ได้ตัดสินใจว่าราคาบ้านควรจะเป็นเท่าไหร่

“จากปัจจัยขนาดของบ้านเท่านั้น”

แต่เราอาจจะมีปัจจัยอื่นๆมาใช้ในการตัดสินใจด้วย เช่น จำนวนห้อง, ปริมาณที่จอดรถ, เฟอร์นิเจอร์ภายในบ้าน

ตัวแปรต้น (X) ที่ใช้จะมีมากกว่า 1 ตัว

$$Y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$



40
30
Z
20
10
0
7.5
10.0
12.5
15.0
17.5
Y
20.0
22.5
25.0
225
200
175
150
125
100
75
50
X

OPTIMIZATION LINEAR PROGRAMMING

เป็นวิธีการหาค่าที่ดีที่สุด ภายใต้เงื่อนไข/ข้อจำกัด/ทรัพยากร

$$
\begin{array}{ll}
\text{maximize} & c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
\text{subject to} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 \\
 & \vdots \\
 & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m \\
 & x_1,\ x_2,\ \ldots\ x_n \geq 0.
\end{array}
$$

We shall refer to linear programs formulated this way as linear programs in *standard form*. We shall always use m to denote the number of constraints, and n to denote the number of decision variables.

OPTIMIZATION LINEAR PROGRAMMING

การประยุกต์ใช้ => กำไรสูงสุด / ต้นทุนต่ำสุด

- อุตสาหกรรมปิโตรเลียม : สูตรผสมน้ำมันดิบ
- ภาคการผลิต : inventory, overtime
- อุตสาหกรรมอาหาร : shipping plan
- การแพทย์ : nutrition
- การสื่อสาร : network

Maximization problem

Objective function $c^T x = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

ที่สอดคล้องกับ constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\le b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_n \end{aligned}\right\} \qquad (Ax \le b)$$

$$x_i \ge 0, \quad i = 1,2,\ldots,n$$

จะได้แบบฟอร์มทั่วไปของปัญหาในการหาค่าสูงสุด คือ

Maximize $c^T x$

Subject to: $Ax \le b$

$x \ge 0$

Minimization problem

Objective function $y^T b = y_1b_1 + y_2b_2 + \ldots + y_mb_m$

ที่สอดคล้องกับ constraints

$$\left.\begin{aligned} y_1a_{11} + y_2a_{21} + \cdots + y_ma_{m1} &\ge c_1 \\ y_1a_{12} + y_2a_{22} + \cdots + y_ma_{m2} &\ge c_2 \\ &\vdots \\ y_1a_{1n} + y_2a_{2n} + \cdots + y_ma_{mn} &\ge c_n \end{aligned}\right\} \qquad (y^T A \ge c^T)$$

$$y_i \ge 0, \qquad i = 1,2,\ldots,n$$

จะได้แบบฟอร์มทั่วไปของปัญหาในการหาค่าต่ำสุด คือ

Minimize $y^T b$

Subject to: $y^T A \ge c^T$

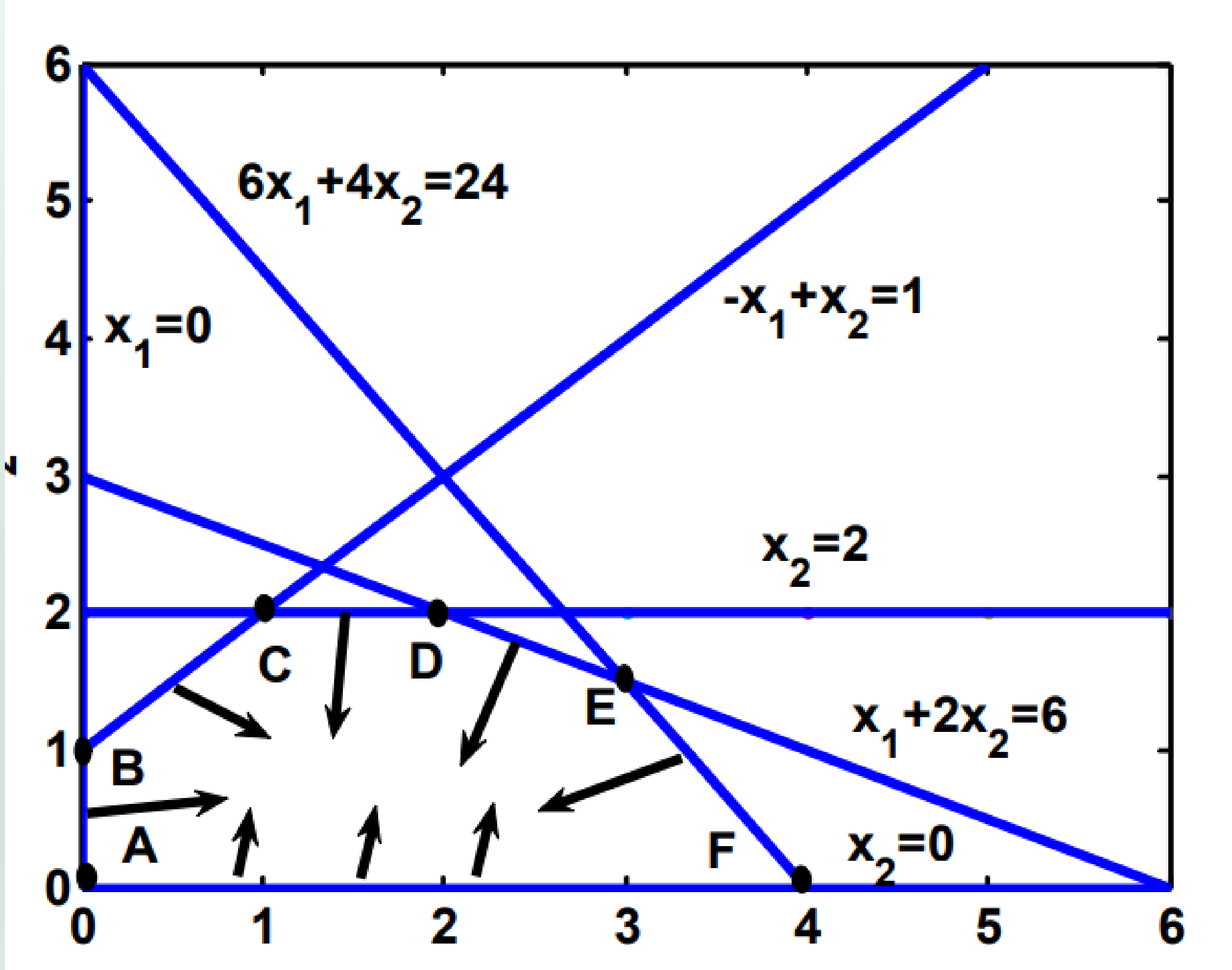
$y \ge 0$

GRAPHICAL METHOD

Maximize $z = 5x_1 + 4x_2$

ที่สอดคล้องกับ constraints:

$$6x_1 + 4x_2 \le 24 \quad (1)$$
$$x_1 + 2x_2 \le 6 \quad (2)$$
$$-x_1 + x_2 \le 1 \quad (3)$$
$$x_2 \le 2 \quad (4)$$
$$x_1 \ge 0 \quad (5)$$
$$x_2 \ge 0 \quad (6)$$



PIVOT METHOD

	s_1	$\cdots$	s_j	$\cdots$	s_n
y_1	a_{11}	$\cdots$	a_{1j}	$\cdots$	a_{1n}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
y_i	a_{i1}	$\cdots$	a_{ij}	$\cdots$	a_{in}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
y_m	a_{m1}	$\cdots$	a_{mj}	$\cdots$	a_{mn}

$\longrightarrow$

	s_1	$\cdots$	y_i	$\cdots$	s_n
y_1	$\hat{a}_{11}$	$\cdots$	$\hat{a}_{1j}$	$\cdots$	$\hat{a}_{1n}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
s_j	$\hat{a}_{i1}$	$\cdots$	$\hat{a}_{ij}$	$\cdots$	$\hat{a}_{in}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
y_m	$\hat{a}_{m1}$	$\cdots$	$\hat{a}_{mj}$	$\cdots$	$\hat{a}_{mn}$

SIMPLEX METHOD

$$-y^T A + s^T = -c^T$$

y_1	y_2	$\ldots$	y_n	s_1	s_2	$\ldots$	s_m	Solution constant
$-a_{11}$	$-a_{12}$	$\ldots$	$-a_{1n}$	1	0	$\ldots$	0	$-c_1$
$-a_{21}$	$-a_{22}$	$\ldots$	$-a_{2n}$	0	1	$\ldots$	0	$-c_2$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$-a_{m1}$	$-a_{m2}$	$\ldots$	$-a_{mn}$	0	0	$\ldots$	1	$-c_m$
$-b_1$	$-b_2$	$\ldots$	$-b_n$	0	0	$\ldots$	0	0

Simplex Tableau ของ minimization problem

WRAP UP

- การทดสอบสมมติฐาน
- การวิเคราะห์ความสัมพันธ์
  - Scatter plot
  - Correlation
  - Regression
- Optimization Linear Programming

Q & A

Q & A