

สถิติเบื้องต้น สำหรับสังคมศาสตร์ 2

โดย อาจารย์วรุณา มินแสน

บทที่ 5 การทดสอบแบบไคสแควร์

ในบางครั้งอาจจัดออกเป็นกลุ่มข้อมูล (categorized data) โดยแบ่งค่าสังเกตเป็นกลุ่ม ๆ และจัดข้อมูลอยู่ในรูปของความถี่ และมักจะเป็นการศึกษาทางด้านคุณภาพ ค่าสังเกตจะได้จากการนับ เช่น ในการศึกษาเกี่ยวกับศาสนา อาจจำแนกศาสนาเป็น พุทธ คริสต์ อิสลามและอื่น ๆ ในการตรวจสอบคุณภาพของผลผลิตจากโรงงานอาจจำแนกเป็น ดีมาก มีตำหนิเล็กน้อยและใช้ไม่ได้ หรือศึกษาสภาพของฝนที่ตกในแต่ละปี อาจจำแนกเป็น ตกมาก ตกปานกลาง และตกเล็กน้อย เป็นต้น

ข้อมูลเชิงปริมาณ

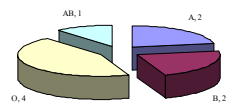
ส่วนสูง.....150, 151.1, 180.55 ซม. เป็นต้น
จำนวนบุตร..... 5, 3, 0 คน เป็นต้น

ข้อมูลเชิงคุณภาพ

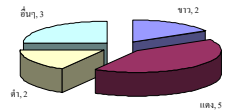
เพศ..... ชาย , หญิง เป็นต้น
สถานะภาพ.....โสด, สมรส, หม้าย เป็นต้น

1. การทดสอบหลายสัดส่วน (การทดสอบอัตราส่วน)

เช่น สัดส่วนคนชอบธงจักรยานยนต์สี
ขาว : แดง : ดำ : อื่นๆ
เป็น 2 : 5 : 2 : 3 หรือไม่



สัดส่วนกรุปเลือด
A : B : O : AB
เป็น 2 : 2 : 4 : 1 หรือไม่



การทดสอบหลายสัดส่วน

ขั้นตอนที่ 1 การตั้งสมมติฐาน

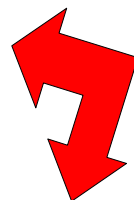
H_0 : สัดส่วน.....เป็น.....
 H_1 : สัดส่วน.....ไม่เป็น.....

เช่น

H_0 : สัดส่วนกรุปเลือด A : B : O : AB เป็น 2:2:4:1
 H_1 : สัดส่วนกรุปเลือด A : B : O : AB ไม่เป็น 2:2:4:1

เช่น

H_0 : ลูกเต๋ายิ่งตรง
 H_1 : ลูกเต๋ามีเอียงตรง



H_0 : ลูกเต๋ามีสัดส่วน 1 : 2 : 3 : 4 : 5 : 6 เป็น 1 : 1 : 1 : 1 : 1 : 1
 H_1 : ลูกเต๋ามีสัดส่วน 1 : 2 : 3 : 4 : 5 : 6 ไม่เป็น 1 : 1 : 1 : 1 : 1 : 1

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \text{ or อื่นๆ}$$



α คือ ระดับนัยสำคัญ
หรือความคาดเคลื่อนชนิดที่ 1
หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$



$$\chi^2 = \sum_{i=1}^k \frac{O_i^2}{E_i} - N$$

O_i = ค่าความถี่สังเกต

E_i = ค่าความถี่คาดหวัง

N = จำนวนครั้งที่ทำการทดลอง
หรือจำนวนที่สังเกตทั้งหมด

ทำผลรวม O และ E ให้เป็นจำนวนรวมเท่ากัน

- โดยการให้ E มีจำนวนเท่ากับ O

เช่น สัตว์ส่วนคนชอบรถจักรยานยนต์สี ขาว : แดง : ดำ : อื่นๆ
เป็น 2 : 5 : 2 : 3 หรือไม่

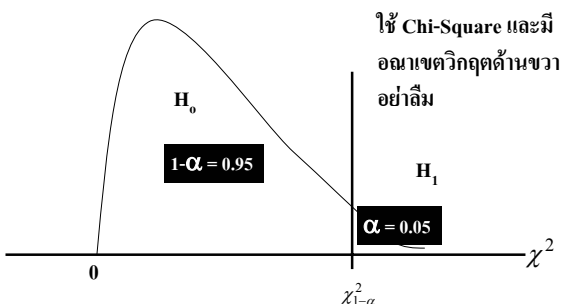
หลังจากได้ทำการสำรวจคน 100 คน ได้ผลดังนี้คนชอบ
รถจักรยานยนต์สี ขาว : แดง : ดำ : อื่นๆ เป็น 20 : 25 : 45 : 10

ทำ E ให้มีจำนวนเท่ากับ O E รวมเป็น 12

สีขาว $100 \times 2 / 12 = 16.67$ สีแดง $100 \times 5 / 12 = 41.67$

สีดำ $100 \times 2 / 12 = 16.67$ อื่นๆ $100 \times 3 / 12 = 25$

4. ทาอณาเขตวิกฤต



***พื้นที่ทั้งหมดมีค่าเท่ากับ 1 ***

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปใน
ขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของ
ใคร

- ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล (สัดส่วนเป็นไปตาม
กำหนด)
- ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล (สัดส่วนไม่เป็นไป
ตามกำหนด)

ลองทำแบบฝึกหัดท้ายบท

2. การทดสอบความเป็นอิสระ

อายุคนขับ	จำนวนครั้งของการเกิดอุบัติเหตุ		
(ปี)	0	1	2
18 - 25	51	183	146
26 - 35	173	104	57
36 - 60	203	154	79

จากข้อมูลนี้พวกเราน่าจะอะไรได้บ้างครับ.....??????

2. การทดสอบความเป็นอิสระ

อายุคนขับ	จำนวนครั้งของการเกิดอุบัติเหตุ		
(ปี)	0	1	2
18 - 25	51	183	146
26 - 35	173	104	57
36 - 60	203	154	79

จากข้อมูลนี้พวกเราน่าจะอะไรได้บ้างครับ.....??????

2. การทดสอบความเป็นอิสระ

อายุคนขับ	จำนวนครั้งของการเกิดอุบัติเหตุ		
(ปี)	0	1	2
18 - 25	51	183	146
26 - 35	173	104	57
36 - 60	203	154	79

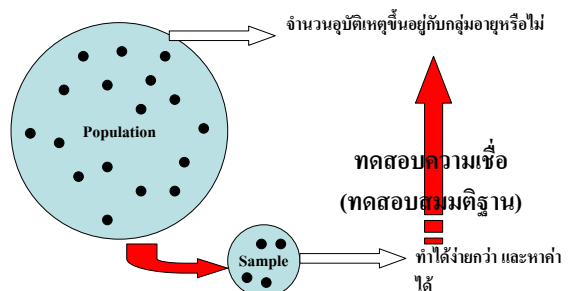
จากข้อมูลนี้พวกเราน่าจะอะไรได้บ้างครับ.....??????

2. การทดสอบความเป็นอิสระ

อายุคนขับ	จำนวนครั้งของการเกิดอุบัติเหตุ		
(ปี)	0	1	2
18 - 25	51	183	146
26 - 35	173	104	57
36 - 60	203	154	79

จะเห็นจำนวนครั้งของการเกิดอุบัติเหตุขึ้นอยู่กับอายุ โดยเฉพาะอายุ 18 - 25 ปี

แต่
เราจึงไม่สามารถสรุปแบบนั้นได้....เพราะข้อมูลที่เห็นในตารางนั้นเป็นเพียง
การสุ่มตัวอย่างเท่านั้น.....



เราเรียกการทดสอบนี้ว่า การทดสอบความเป็นอิสระ

ขั้นตอนที่ 1 การตั้งสมมติฐาน

H_0 : ตัวแปร X และตัวแปร Y ไม่ขึ้นอยู่กับกัน

H_1 : ตัวแปร X และตัวแปร Y ขึ้นอยู่กับกัน

เช่น

H_0 : จำนวนครั้งการเกิดอุบัติเหตุไม่ขึ้นอยู่กับกลุ่มอายุ

H_1 : จำนวนครั้งการเกิดอุบัติเหตุขึ้นอยู่กับกลุ่มอายุ

หรือ

H_0 : จำนวนครั้งการเกิดอุบัติเหตุไม่สัมพันธ์กับกลุ่มอายุ

H_1 : จำนวนครั้งการเกิดอุบัติเหตุสัมพันธ์กับกลุ่มอายุ

หรือ

H_0 : จำนวนครั้งการเกิดอุบัติเหตุเป็นอิสระกับกลุ่มอายุ

H_1 : จำนวนครั้งการเกิดอุบัติเหตุไม่เป็นอิสระกับกลุ่มอายุ

เป็นต้น

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.01$

$\alpha = 0.05$

$\alpha = 0.10$ or อื่นๆ



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ทดสอบ

$$\chi^2 = \sum_{i=1}^{\text{all}} \sum_{j=1}^{\text{all}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \sum_{i=1}^{\text{all}} \sum_{j=1}^{\text{all}} \left(\frac{O_{ij}^2}{E_{ij}} \right) - N$$

O_{ij} = ค่าความถี่สังเกต

E_{ij} = ค่าความถี่คาดหวัง

N = จำนวนครั้งที่ทำการทดลอง

หรือจำนวนที่สังเกตทั้งหมด

ข้อมูลที่ใช้สำรวจมาในตารางเป็นค่า O_{ij}

อายุคนขับ	จำนวนครั้งของการเกิดอุบัติเหตุ		
	0	1	2
(ปี)			
18 - 25	51	183	146
26 - 35	173	104	57
36 - 60	203	154	79

แล้ว E_{ij} หาอย่างไร.....

อายุคนขับ (ปี)	จำนวนครั้งของการเกิดอุบัติเหตุ			รวม
	0	1	2	
18 - 25	427x380/1150	441x380/1150	282x380/1150	380
26 - 35	427x334/1150	441x334/1150	282x334/1150	334
36 - 60	427x436/1150	441x436/1150	282x436/1150	436
รวม	427	441	282	1150

แล้ว E_{ij} หาอย่างไร.....

อายุคนขับ (ปี)	จำนวนครั้งของการเกิดอุบัติเหตุ			รวม
	0	1	2	
18 - 25	427x380/1150	441x380/1150	282x380/1150	380
26 - 35	427x334/1150	441x334/1150	282x334/1150	334
36 - 60	427x436/1150	441x436/1150	282x436/1150	436
รวม	427	441	282	1150

ดังนั้นจะได้ค่า E_{ij} ดังตาราง

อายุคนขับ (ปี)	จำนวนครั้งของการเกิดอุบัติเหตุ		
	0	1	2
18 - 25	141.10	145.72	93.18
26 - 35	124.02	128.08	81.90
36 - 60	161.89	167.20	106.91

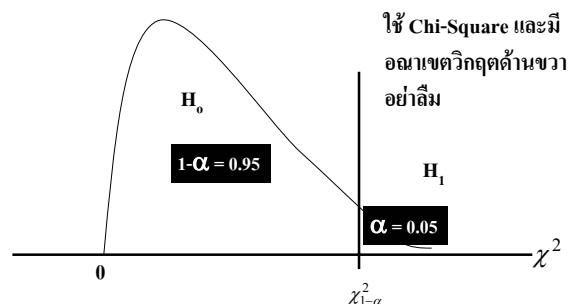
แทนค่าในสูตร

$$\chi^2 = \sum_{i=1}^{\text{all}} \sum_{j=1}^{\text{all}} \left(\frac{O_{ij}^2}{E_{ij}} \right) - N$$

$$\chi^2 = \left(\frac{51^2}{141.10} + \frac{183^2}{145.72} + \dots + \frac{79^2}{106.91} \right) - 1150$$

$$\chi^2 = 147.22$$

4. หาขอบเขตวิกฤต



***พื้นที่ทั้งหมดมีค่าเท่ากับ 1 *** $d.f. = (r-1)(c-1)$

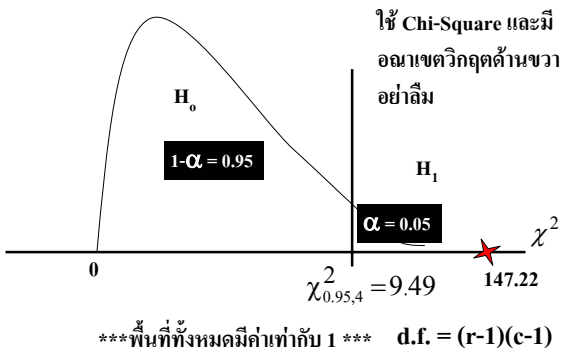
5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล (ตัวแปร X และตัวแปร Y ไม่ขึ้นอยู่กับกัน)
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล (ตัวแปร X และตัวแปร Y ขึ้นอยู่กับกัน)

ลองทำตัวอย่างให้สมบูรณ์เมื่อกำหนดระดับนัยสำคัญ 0.05 และทำแบบฝึกหัดท้ายบท

4. หาอาณาเขตวิกฤต



- ปฏิเสธสมมติฐาน H_0 ที่ระดับนัยสำคัญ 0.05 จำนวนครั้งการเกิดอุบัติเหตุขึ้นอยู่กับกลุ่มอายุ

3. การทดสอบภาวะรูปดี หรือการทดสอบการแจกแจงของข้อมูล

- พิจารณาจากตัวอย่าง

ในการศึกษาความชุกชุมของยุงลาย โดยทำการทดลองนำกระป๋องขนาดใหญ่บรรจุน้ำวางกระจายในท้องที่หนึ่ง เป็นเวลา 1 สัปดาห์ แล้วสำรวจจำนวนลูกน้ำของยุงลายในแต่ละกระป๋อง บันทึกข้อมูลได้ดังนี้

จำนวนลูกน้ำของยุงลาย	0	1	2	3	4	5	6	7
จำนวนกระป๋อง	14	18	29	18	10	7	3	1

ในการปลูกเมล็ดพันธุ์ผัก 450 เมล็ด โดยปลูกเป็นแถว ๆ ละ 5 เมล็ด จำนวน 90 แถว เมล็ดที่งอกในแต่ละแถวเป็นดังนี้

จำนวนเมล็ดพันธุ์ที่งอก/แถว	1	2	3	4	5
จำนวนแถว	1	11	30	38	10

ข้อมูลที่กำหนดให้ต่อไปนี้เป็นความยาวของทารกแรกเกิด จำนวน 125 คน

ความยาว (ซม.)	45-46.9	47-48.9	49-50.9	51-52.9	53-54.9
จำนวน	28	32	35	20	10

สรุปการแจกแจงที่สามารถตั้งข้อสังเกต(ใน 208272)

- ข้อมูลเป็นการแจกแจงแบบทวินาม นั่นคือ มีจำนวนครั้งทำการทดลองที่แน่นอน และข้อมูลเป็นแบบ ไม่ต่อเนื่อง
- ข้อมูลเป็นการแจกแจงแบบปัวซอง นั่นคือ มีเวลาและพื้นที่เป็นข้อจำกัด สิ่งที่เกิดจากการทดลอง มีจำนวนได้ไม่แน่นอน(ไม่จำกัด) และข้อมูลเป็นแบบ ไม่ต่อเนื่อง
- ข้อมูลเป็นการแจกแจงแบบปกติ นั่นคือ ข้อมูลเป็นแบบ ต่อเนื่อง ทุกชนิด

จำนวนลูกน้ำของยุงลาย	0	1	2	3	4	5	6	7
จำนวนกระป๋อง	14	18	29	18	10	7	3	1

จำนวนเมล็ดพันธุ์ที่งอก/แถว	1	2	3	4	5
จำนวนแถว	1	11	30	38	10

ความยาว (ซม.)	45-46.9	47-48.9	49-50.9	51-52.9	53-54.9
จำนวน	28	32	35	20	10

ขั้นตอนที่ 1 การตั้งสมมติฐาน

H_0 : ข้อมูลมีการแจกแจงตามแบบที่กำหนด

H_1 : ข้อมูลไม่มีการแจกแจงตามแบบที่กำหนด

เช่น

H_0 : จำนวนลูกน้ำในกระป๋องมีการแจกแจงแบบปัวซอง

H_1 : จำนวนลูกน้ำในกระป๋องมีการแจกแจงแบบอื่น

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.01$

$\alpha = 0.05$

$\alpha = 0.10$ or อื่นๆ



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ทดสอบ

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$



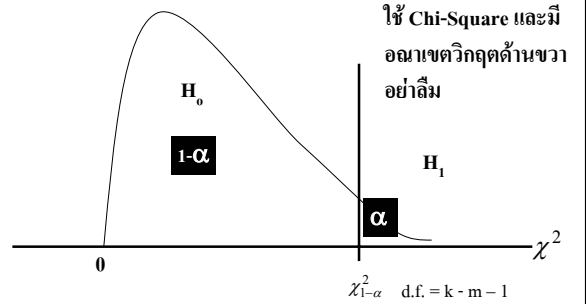
$$\chi^2 = \sum_{i=1}^k \frac{O_i^2}{E_i} - N$$

O_i = ค่าความถี่สังเกต

E_i = ค่าความถี่คาดหวัง (ต้องคำนวณ)

N = จำนวนครั้งที่ทำการทดลอง
หรือจำนวนที่สังเกตทั้งหมด

4. ทาอณาเขตวิกฤต



พื้นที่ทั้งหมดมีค่าเท่ากับ 1 m คือ จำนวนค่าประมาณ

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปใน
ขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของ
ใคร

ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล (การแจกแจงเป็นไป
ตามข้อสงสัย)”

ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล (การแจกแจงเป็นไป
ตามข้อสงสัย)”

ในการปลูกเมล็ดพันธุ์ฝัก 450 เมล็ด โดยปลูกเป็นแถว ๆ ละ 5
เมล็ด จำนวน 90 แถว เมล็ดที่งอกในแต่ละแถวเป็นดังนี้ จากผลการทดลอง
นี้จะพิจารณาว่าข้อมูลดังกล่าวควรมีการแจกแจงความน่าจะเป็นแบบใดและ
ทดสอบสมมุติฐาน โดยใช้ระดับความมีนัยสำคัญ 5 %

จำนวนเมล็ดพันธุ์ที่ งอก/แถว	1	2	3	4	5
จำนวนแถว	1	11	30	38	10

พิจารณาข้อมูลตามค่าสังเกต ที่จำนวนเป็นการงอกของเมล็ดในแต่ละแถว โดยปลูกแถว
ละ 5 เมล็ด ถ้า นิยามตัวแปรสุ่ม X ดังนี้

X : จำนวนเมล็ดที่งอกในแต่ละแถว $x = 0, 1, 2, 3, 4, 5$

จะเห็นว่าตัวแปร X มีคุณสมบัติเบื้องต้นเป็นการแจกแจงแบบทวินาม

ขั้นตอนที่ 1 การตั้งสมมุติฐาน

1

H_0 : ข้อมูลสอดคล้องกับการแจกแจงแบบทวินาม
 H_1 : ข้อมูลไม่มีการแจกแจงตามแบบที่กำหนด

2

$\alpha = 0.05$

3

$$\chi^2_{\text{cal}} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

การหาค่า E_i

- การคำนวณค่าความถี่คาดหวังต้องใช้ความน่าจะเป็นของการแจกแจง
แบบทวินามตาม H_0 ซึ่งจะเห็นว่าไม่ทราบค่า p (ความน่าจะเป็นที่
เมล็ดแต่ละเมล็ดจะงอก) จึงต้องประมาณจากค่าสังเกตดังนี้

$$\begin{aligned}\bar{X} &= \frac{\sum fx}{N} = \frac{(0 \times 0) + (1 \times 1) + (2 \times 11) + (3 \times 30) + (4 \times 38) + (5 \times 10)}{90} \\ &= \frac{315}{90} = 3.5\end{aligned}$$

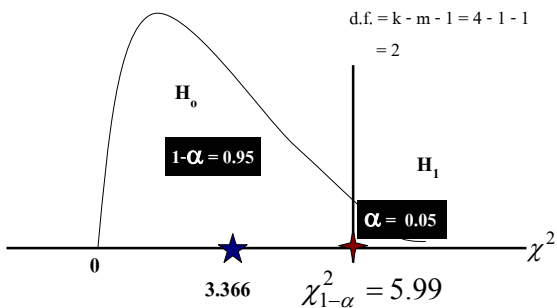
เนื่องจากค่าเฉลี่ยของการแจกแจงแบบทวินาม = np
 ดังนั้น np = 3.5 กรณีนี้ n = 5
 นั่นคือ

$$p = 3.5/5 = 0.7$$

จะหา P(X=x) จากตารางทวินาม b(x; 5, 0.7)
 แล้วนำไปเปิดตารางเพื่อหาค่า p แต่ละตัว

X	O	P(X=x)	= NP(X=x)	$\frac{(O_i - E_i)^2}{E_i}$
0	0	0.0024	0.216	0.489
1	1	0.0284	2.556 = 14.679	
2	11	0.1323	11.907	
3	30	0.3087	27.783	0.177
4	38	0.3602	32.418	0.961
5	10	0.1681	15.129	1.739
รวม	90	1.0000	90	3.366

4 หาดัชนีวิกฤต



4 สรุป

ยอมรับ H_0 ที่ระดับนัยสำคัญ 0.05
 ข้อมูลสอดคล้องกับการแจกแจงแบบทวินาม

บทที่ 6 การถดถอยและสหสัมพันธ์

ลูกบิ๊นที่อิง	เป้าหมายที่ลูก ทำลาย
5 ลูก	1 คน
10 ลูก	2 คน
15 ลูก	3 คน
20 ลูก	4 คน
25 ลูก	???

สวัสดิการ
 จากตารางที่เห็นลองเดาซิว่า
 จะได้เลขอะไรตรง ???

สวัสดิ์ครับ

ลูกปืนที่ยิง	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
10 ลูก	2 คน
15 ลูก	3 คน
20 ลูก	4 คน
25 ลูก	???

เราได้แม่นี่

5 คน
ครบหม

ลูกปืนที่ยิง	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
8 ลูก	1 คน
10 ลูก	4 คน
12 ลูก	5 คน
14 ลูก	???

ทำไมถึงคิด
นานจัง

เออ.....
5 คนมั้ง

6 ครั้ง

6-7 ครั้ง
ไม่แน่ใจ

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
10 ลูก	2 คน
15 ลูก	3 คน
20 ลูก	4 คน
25 ลูก	???

โจทย์ที่ 1

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
8 ลูก	1 คน
10 ลูก	4 คน
12 ลูก	5 คน
14 ลูก	???

โจทย์ที่ 2

สังเกตหรือเปล่าพลทหาร
สมองคนเรามีเครื่อง
คำนวณอยู่สามารถนำ
เครื่องคำนวณนั้นมาใช้
คำนวณเพื่อคาดเดาง่ายๆ
ได้อย่างสบาย โดยไม่
ต้องแสดงวิธีทำ

ใช่ครับ

ครับ

จริงด้วย

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
10 ลูก	2 คน
15 ลูก	3 คน
20 ลูก	4 คน
25 ลูก	???

โจทย์ที่ 1

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
8 ลูก	1 คน
10 ลูก	4 คน
12 ลูก	5 คน
14 ลูก	???

โจทย์ที่ 2

จะเห็นว่าสมองเรา
พยายามนำข้อมูลที่มีอยู่
มาเป็นตัวช่วยในการคาด
เดาของพวกเขา
และเมื่อเพิ่มข้อมูลตัว
หนึ่งเข้าไปก็จะสามารถ
คาดเดาอีกตัวว่าเป็น
เท่าไร

ใช่ครับ

ครับ

จริงด้วย

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
10 ลูก	2 คน
15 ลูก	3 คน
20 ลูก	4 คน
25 ลูก	???

โจทย์ที่ 1

ลูกปืน	เป้าหมายที่ถูกทำลาย
5 ลูก	1 คน
8 ลูก	1 คน
10 ลูก	4 คน
12 ลูก	5 คน
14 ลูก	???

โจทย์ที่ 2

ยากขึ้น

แต่โจทย์ที่สองข้อมูลมีความ
ซับซ้อนกว่าโจทย์แรก จะ
เห็นว่าการคาดเดาของเรา
เราเริ่มแตกต่างกัน.....
เพราะสมองเรามีเกณฑ์การ
คำนวณที่แตกต่างกัน ไม่
สามารถบ่งบอกวิธีการได้มา
ของคำตอบที่ชัดเจนได้ และ
คำตอบถูกต้องหรือไม่ก็
ตรวจสอบได้ยาก

อืมจริงหะ

จริงนะ B3

จริงด้วย B2

ผมจะไม่บอกอะไรถูก
ใครคิด เพราะการคาดเดา
บอกคำตอบที่ชัดเจนไป
ไม่ได้ ถูกคิดไม่รู้อ

แต่.....เราจะลองหาวิธีการ
คาดเดาที่เป็นหลักเป็นเกณฑ์
เพื่อให้เกิดการคำนวณได้
อย่างมีเหตุผล ที่สูงน่ได้
และมีคามน่าเชื่อถือ

ใช้วิธีอะไรครับ

.....

????????

วิธีที่ง่ายคือ
สหสัมพันธ์และการถดถอย

ชื่อข้าง

ต้องตั้งใจ
หน่อยแล้ว

????????
จะพยายามครับ

เดี๋ยวผมจะเปิดวิดีโอสอนวิธีการคาดเดาให้พวก
เราเรียนรู้ ตั้งใจด้วย



ครับผม

บทที่ 9 สหสัมพันธ์และการถดถอย

1. สหสัมพันธ์ (Correlation)
2. สมการถดถอยอย่างง่าย
3. การวิเคราะห์การถดถอย

จะได้กล่าวถึงการวิเคราะห์ข้อมูลที่ประกอบด้วยตัวแปรตั้งแต่ 2 ตัวขึ้นไป เพื่อศึกษาความสัมพันธ์ของตัวแปรเชิงปริมาณ เช่น การศึกษาความสัมพันธ์ของรายได้ต่อเดือนของบริษัท และรายจ่ายต่อเดือนของบริษัทว่ามีความสัมพันธ์กันมากน้อยเพียงใด โดยข้อมูลที่ต้องการวัดความสัมพันธ์ของตัวแปร 2 ตัวนั้น มีมาตราการวัดที่เป็นแบบอันตรภาค (interval scale) หรือแบบอัตราส่วน (ratio scale) ซึ่งเรียกว่า เป็นข้อมูลเชิงปริมาณ (quantitative data) เช่น ต้องการดูความสัมพันธ์ของส่วนสูงกับน้ำหนัก



ส่วนสูง

น้ำหนัก

สังเกต ส่วนสูงและน้ำหนักเป็นข้อมูล
ที่มาจากแหล่งเดียวกัน และต้องมาพร้อมๆ กัน
ข้อมูลแบบนี้เรียกว่า 2 ตัวแปร

การวัดความสัมพันธ์แบบนี้เป็นการวัดที่เรียกว่า
สหสัมพันธ์ (correlation) และสามารถคำนวณออกมา
เป็นตัวเลข เรียกว่า **สัมประสิทธิ์สหสัมพันธ์** (correlation
coefficient) ซึ่งมีวิธีการคำนวณอยู่หลายวิธีด้วยกัน

สมการถดถอยนี้แหละพลทหาร ที่เรา
จะนำไปใช้คาดเดา

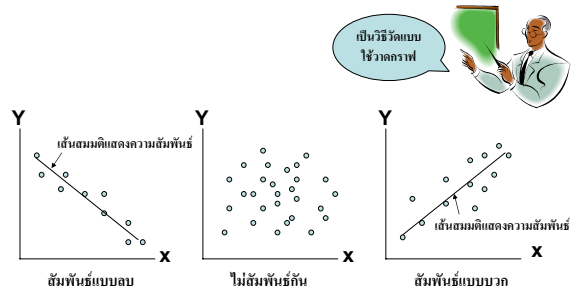
และนอกจากนี้ยังสามารถพิจารณาความสัมพันธ์ในรูปแบบ
เชิงคณิตศาสตร์ที่จะอธิบายตัวแปรหนึ่ง เมื่อทราบตัวแปรอีกตัว
หนึ่ง ซึ่งเรียกว่า **การถดถอย** (regression) โดยสร้างออกมาในรูป
ของสมการแสดงความสัมพันธ์ และเรียกสมการนั้นว่า **สมการ
ถดถอย** (regression equation)

9.1 สหสัมพันธ์ (Correlation)

การพิจารณาสหสัมพันธ์ที่จะกล่าวในที่นี้ จะเป็นการวัดความสัมพันธ์เชิงเส้นตรงของตัวแปรเชิงปริมาณ 2 ตัว เช่น

- ความสัมพันธ์ของคะแนนการสอบวิชาสถิติ (X) กับวิชาคอมพิวเตอร์ (Y) ของนักศึกษา
 - ความสัมพันธ์ของความสูง (X) กับน้ำหนัก (Y) ของนักเรียน
 - ความสัมพันธ์ของราคาส่งออก (X) กับปริมาณส่งออก (Y) ของลำไย
- การศึกษาความสัมพันธ์มักนิยมใช้วิธีการวัดดังนี้

1. Scatter diagram (การวาดกราฟและพิจารณาลักษณะของมัน)



2. สัมประสิทธิ์สหสัมพันธ์ (Pearson Product-Moment Correlation Coefficient)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2\right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i\right)^2\right]}}$$

$$r = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\left[\sum_{i=1}^n X_i^2 - n \bar{X}^2\right] \left[\sum_{i=1}^n Y_i^2 - n \bar{Y}^2\right]}}$$



โดยสัมประสิทธิ์สหสัมพันธ์ (r) จะมีค่าตั้งแต่ -1 ถึง 1
ถ้า r = 1 หมายความว่า ตัวแปร X และ Y มีความสัมพันธ์เชิงบวกแบบตามกันโดยสมบูรณ์

ถ้า r = 0 หมายความว่า ตัวแปร X และ Y ไม่มีความสัมพันธ์เชิงเส้นกัน

ถ้า r = -1 หมายความว่า ตัวแปร X และ Y มีความสัมพันธ์เชิงลบกันโดยสมบูรณ์

ตัวอย่าง

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)
1	36	54
2	26	30
3	12	28
4	40	48
5	24	36
6	18	30
7	30	38
8	30	46
9	14	16
10	34	42

ตัวอย่าง

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)	X ²	Y ²	X * Y
1	36	54	1296	2916	1944
2	26	30	676	900	780
3	12	28	144	784	336
4	40	48	1600	2304	1920
5	24	36	576	1296	864
6	18	30	324	900	540
7	30	38	900	1444	1140
8	30	46	900	2116	1380
9	14	16	196	256	224
10	34	42	1156	1764	1428

$$r = \frac{10 \times 10556 - 264 \times 368}{\sqrt{[10 \times 7768 - 264^2][10 \times 14680 - 368^2]}} = 0.8822$$

ดูตัวอย่างจากข้อมูลนักศึกษา

9.2 สมการถดถอยอย่างง่าย

จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่ ทำให้ทราบความสัมพันธ์เชิงเส้นของตัวแปรคู่นั้น ว่ามีความสัมพันธ์มากน้อยเพียงใด หลังจากที่เราทราบความสัมพันธ์แล้ว ก็จะสามารถนำข้อมูลนั้นมาสร้างความสัมพันธ์ในลักษณะของสมการซึ่งเรียกว่า สมการถดถอย (regression equation) ซึ่งสมการที่แสดงความสัมพันธ์ของข้อมูลตัวแปรคู่ (X และ Y) อาจจะมีรูปแบบสมการความสัมพันธ์ได้หลายลักษณะ เช่น



สมการเส้นตรง $Y = a + bX$
 สมการพาราโบลา $Y = a + bX + cX^2$
 สมการเอกซโพเนนเชียล $Y = ab^x$

ตัวแปร Y ว่าตัวแปรตาม (dependent variable)
 ตัวแปร X ว่าตัวแปรอิสระ (independent variable)
 a เรียกว่า จุดตัดแกน Y (Y - intercept)
 b เรียกว่า ความชันของเส้นตรง (slope)

สมการเส้นตรง $\hat{Y} = a + bX$

ส่วนสมการที่แสดงความสัมพันธ์ที่แท้จริงของข้อมูลประชากร เขียนแทนด้วย $Y = \alpha + \beta X + \varepsilon$ โดย a และ b เป็นตัวสถิติที่สมนัยกับพารามิเตอร์ α และ β ตามลำดับ

$$\begin{cases} b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} \\ b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} , & a = \bar{Y} - b\bar{X} \\ b = \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} \end{cases}$$

ตัวอย่าง

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)	X ²	Y ²	X * Y
1	36	54	1296	2916	1944
2	26	30	676	900	780
3	12	28	144	784	336
4	40	48	1600	2304	1920
5	24	36	576	1296	864
6	18	30	324	900	540
7	30	38	900	1444	1140
8	30	46	900	2116	1380
9	14	16	196	256	224
10	34	42	1156	1764	1428

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$$y = 8.998 + 1.0531x$$

จากสมการเส้นตรงที่ได้สามารถนำมาใช้พยากรณ์ได้ดังนี้

$$y = 8.998 + 1.0531x$$

ถ้าจะลงทุนเพิ่ม โดยสร้างเขตที่ 11 ซึ่งบริษัทคิดจะลงทุน 20 ล้านบาท สามารถประมาณผลตอบแทนได้ คือ

$$x = 20$$

$$y = 8.998 + 1.0531(20)$$

$$y = 30.06$$

ถ้าจะลงทุนเพิ่ม โดยสร้างเขตที่ 11 ซึ่งบริษัทคิดจะลงทุน 20 ล้านบาท สามารถประมาณผลตอบแทนได้เงินประมาณ 30 ล้านบาท เป็นต้น

ดูตัวอย่างจากข้อมูลนักศึกษา

การหาความคลาดเคลื่อนจากการประมาณ

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

สัมประสิทธิ์การตัดสินใจ

$$r^2 (100\%)$$

หมายถึง x พยากรณ์ y ได้ $r^2(100\%)$ เปอร์เซนต์

จากตัวอย่าง $y = 8.998 + 1.0531x$

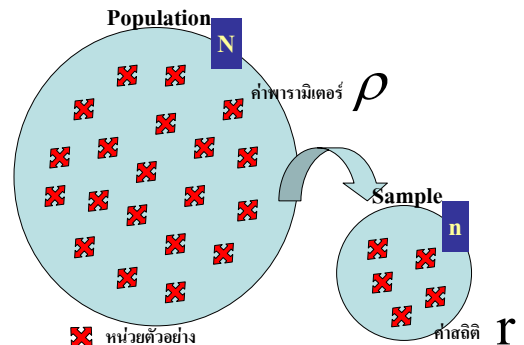
เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)	ค่าพยากรณ์	e^2
1	36	54	46.91	50.27
2	26	30	36.38	40.69
3	12	28	21.64	40.51
4	40	48	51.12	9.75
5	24	36	34.27	2.98
6	18	30	27.95	4.19
7	30	38	40.59	6.71
8	30	46	40.59	29.26
9	14	16	23.74	59.93
10	34	42	44.80	7.86
รวม				252.15

9.3 การวิเคราะห์การถดถอย

1. นัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ (เช่น การทดสอบว่า x และ y มีความสัมพันธ์กันหรือไม่)
2. การประมาณช่วงความเชื่อมั่นของพารามิเตอร์ α และ β
3. การทดสอบภาวะรูปโค้งแบบจำลอง(เป็นการทดสอบว่า Model ที่เราสร้างมาใช้ได้หรือไม่)

1. นัยสำคัญของสัมประสิทธิ์สหสัมพันธ์

(เช่น การทดสอบว่า x และ y มีความสัมพันธ์กันหรือไม่)



การทดสอบแบ่งออกเป็น 2 กรณี

1. ในกรณีที่ต้องการทดสอบ $\rho = 0, \rho \geq 0, \rho \leq 0$
2. ในกรณีที่ต้องการทดสอบ $\rho = c, \rho \geq c, \rho \leq c$
 c คือค่าคงที่ที่ไม่ใช่ศูนย์

1. ตั้งสมมติฐาน

	แบบที่ 1	แบบที่ 2
ก	$H_0 : \rho = 0$ $H_1 : \rho \neq 0$	ก $H_0 : \rho = c$ $H_1 : \rho \neq c$
ข	$H_0 : \rho \geq 0$ $H_1 : \rho < 0$	ข $H_0 : \rho \geq c$ $H_1 : \rho < c$
ค	$H_0 : \rho \leq 0$ $H_1 : \rho > 0$	ค $H_0 : \rho \leq c$ $H_1 : \rho > c$

c คือค่าคงที่ที่ไม่ใช่ศูนย์

2. กำหนดระดับนัยสำคัญ

- $\alpha = 0.01$
 $\alpha = 0.05$
 $\alpha = 0.10$ or อื่นๆ



α คือ ระดับนัยสำคัญ
 หรือความคาดเคลื่อนชนิดที่ 1
 หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. เลือกสถิติทดสอบ

แบบที่ 1
$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

 $d.f. = n - 2$

3. เลือกสถิติทดสอบ

แบบที่ 2
$$Z = \frac{R - \mu_R}{\sigma_R}$$

$$R = \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right] \quad \mu_R = \frac{1}{2} \ln \left[\frac{1+\rho}{1-\rho} \right]$$

$$\sigma_R = \frac{1}{\sqrt{n-3}}$$

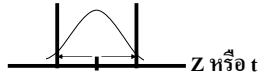
$$\mu_R < \overset{\text{เปิดตาราง}}{\text{-----}} > \rho$$

$$R < \overset{\text{เปิดตาราง}}{\text{-----}} > r$$

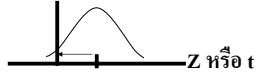
4. หอณาเขตวิกฤต (วาดรูป)

แบบที่ 1 และ 2

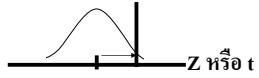
ก $H_0 : \rho = c$
 $H_1 : \rho \neq c$



ข $H_0 : \rho \geq c$
 $H_1 : \rho < c$



ค $H_0 : \rho \leq c$
 $H_1 : \rho > c$



C คือค่าคงที่ที่ใดๆค่า

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

ตัวอย่าง

1. ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากตัวอย่าง 42 คู่เป็น 0.22 ด้วยระดับนัยสำคัญ 0.05 จงทดสอบว่าตัวแปร x และ y มีความสัมพันธ์กันหรือไม่

2. จากตัวอย่างข้างล่าง จงทดสอบว่า เงินลงทุน มีความสัมพันธ์เชิงบวกกับ ผลตอบแทนที่ได้รับ

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)
1	36	54
2	26	30
3	12	28
4	40	48
5	24	36
6	18	30
7	30	38
8	30	46
9	14	16
10	34	42

$$r = \frac{10 \times 10556 - 264 \times 368}{\sqrt{(10 \times 7768 - 264^2)(10 \times 14680 - 368^2)}} = 0.8822$$

ตัวอย่าง

3. ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากตัวอย่าง 39 คู่เป็น 0.85 ด้วยระดับนัยสำคัญ 0.10 จงทดสอบว่าตัวแปร x และ y มีความสัมพันธ์กันน้อยกว่า 0.90 หรือไม่

***การประยุกต์ใช้หาช่วงความเชื่อมั่น ของ ρ

จาก $Z = \frac{R - \mu_R}{\sigma_R}$ หลักการเดิมๆ ในบทที่ 2

$$\mu_R = R \pm Z \cdot \sigma_R$$

หรือ $R - Z_{1-\frac{\alpha}{2}} \cdot \sigma_R < \mu_R < R + Z_{1-\frac{\alpha}{2}} \cdot \sigma_R$

ตัวอย่าง

4. ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากตัวอย่าง 50 คู่เป็น -0.48 จงหา 98% ช่วงความเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร

เปิดตาราง

$$\mu_R \text{ ----- } > \rho$$

การประมาณช่วงความเชื่อมั่นของพารามิเตอร์

1. การประมาณ α

$$t = \frac{a - E(a)}{\sqrt{V(a)}} \quad \text{d.f.} = n - 2$$

$$t = \frac{a - \alpha}{S_E \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}} \rightarrow t = \frac{a - \alpha}{S_E \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}}$$

$$t = \frac{a - \alpha}{S_E \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}} \quad \text{ทำเหมือนบทที่ 2}$$

$$\alpha = a \pm t \cdot S_E \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$$

หรือ

$$a - t_{1-\frac{\alpha}{2}} S_E \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} < \alpha < a + t_{1-\frac{\alpha}{2}} S_E \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$$

2. การประมาณ β

$$t = \frac{b - E(b)}{\sqrt{V(b)}} \quad \text{d.f.} = n - 2$$

$$t = \frac{b - \beta}{\frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}}}$$

$$t = \frac{b - \beta}{\frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}}} \quad \text{ทำเหมือนบทที่ 2}$$

$$\beta = b \pm t \cdot \frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$b - t_{1-\frac{\alpha}{2}, \text{d.f.}} \cdot \frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}} < \beta < b + t_{1-\frac{\alpha}{2}, \text{d.f.}} \cdot \frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}}$$

S_E คือ อะไร....หาได้อย่างไร

S_E คือ ค่าเบี่ยงเบนมาตรฐานของสมการถดถอยที่ Y ขึ้นอยู่กับ X

$$S_E = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}$$

$$S_E = \sqrt{\frac{1}{n-2} [(n-1)S_y^2 - b^2(n-1)S_x^2]}$$

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$



$$r = b \cdot \frac{S_x}{S_y}$$

แต่ถ้าไม่อยากคำนวณ S_y และ S_x

$$S_E = \sqrt{\frac{1}{n-2} [(n-1)S_y^2 - b^2(n-1)S_x^2]}$$



$$S_E = \sqrt{\frac{1}{n-2} [\sum (Y_i - \bar{Y})^2 - b^2 \sum (X_i - \bar{X})^2]}$$

จากตัวอย่าง

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)	X^2	Y^2	$X \cdot Y$
1	36	54	1296	2916	1944
2	26	30	676	900	780
3	12	28	144	784	336
4	40	48	1600	2304	1920
5	24	36	576	1296	864
6	18	30	324	900	540
7	30	38	900	1444	1140
8	30	46	900	2116	1380
9	14	16	196	256	224
10	34	42	1156	1764	1428

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$y = 8.998 + 1.0531x$$

จงหา 95% ช่วงความเชื่อมั่นของ α และ β

จากโจทย์...สรุปสิ่งที่จะต้องหาเพื่อใช้ในการคำนวณ

n

a

b

$$\sum (X_i - \bar{X}_i)^2$$

$$\sum X_i^2$$

$$\sum (Y_i - \bar{Y}_i)^2$$

t

3. การทดสอบภาวะรูปดีของแบบจำลอง(เป็นการทดสอบว่า Model ที่เราสร้างมาใช้ได้หรือไม่)

$$\hat{Y} = a + bX$$

สมการนี้มันใช้ได้จริง
หรือเปล่า.....

1. ตั้งสมมติฐาน

ก $H_0 : \beta = 0$
 $H_1 : \beta \neq 0$

ข $H_0 : \beta \geq 0$
 $H_1 : \beta < 0$

ค $H_0 : \beta \leq 0$
 $H_1 : \beta > 0$

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.01$

$\alpha = 0.05$

$\alpha = 0.10$ or อื่นๆ



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

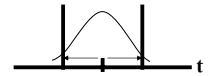
หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

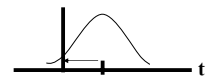
$$t = \frac{b - \beta}{\frac{S_E}{\sqrt{\sum (x_i - \bar{x})^2}}}$$

4. หาอณาเขตวิกฤต (วาดรูป)

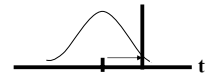
ก $H_0 : \beta = 0$
 $H_1 : \beta \neq 0$



ข $H_0 : \beta \geq 0$
 $H_1 : \beta < 0$



ค $H_0 : \beta \leq 0$
 $H_1 : \beta > 0$



d.f. = n - 2

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล สมการเส้นตรงใช้ไม่ได้ หรือ x กับ y ไม่มีความสัมพันธ์กัน

ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล สมการเส้นตรงใช้ได้ หรือ X กับ Y มีความสัมพันธ์กัน

จากตัวอย่าง

จากการศึกษาข้อมูลจำนวนเงินลงทุน และผลตอบแทนที่ได้ของบริษัทแห่งหนึ่งในเขตพื้นที่การลงทุนทั้งหมด 10 เขต ปรากฏข้อมูลดังนี้

เขตที่	เงินลงทุน (ล้านบาท)	ผลตอบแทน (ล้านบาท)	X ²	Y ²	X * Y
1	36	54	1296	2916	1944
2	26	30	676	900	780
3	12	28	144	784	336
4	40	48	1600	2304	1920
5	24	36	576	1296	864
6	18	30	324	900	540
7	30	38	900	1444	1140
8	30	46	900	2116	1380
9	14	16	196	256	224
10	34	42	1156	1764	1428

$a = \bar{Y} - b\bar{X}$

$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$

$y = 8.998 + 1.0531x$

จงทดสอบว่าสมการเส้นตรงใช้ได้จริงหรือไม่ที่ระดับความเชื่อมั่น 95%

บทที่ 7 การทดสอบนอนพารามตริกสถิติ

ตรงกันข้ามกับ พารามตริก

ถ้าจะพิจารณาที่ข้อมูลโดยตรงเราสามารถดูว่ามันเป็น
Normal Distribution ได้โดย

- 1. Nominal Scale
- 2. Ordinal Scale
- 3. Interval Scale
- 4. Ratio Scale

โดยธรรมชาติแล้ว
จะไม่เป็น Normal
(นอกเสียจากใช้เทคนิค
ขั้นสูงในการปรับเปลี่ยน)

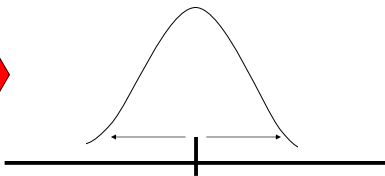
โดยธรรมชาติแล้ว
จะเป็น Normal
(นอกเสียจากรูปร่างของ
ข้อมูลผิดปกติไป ซึ่งเกิด
ขึ้นได้ด้วยหลายๆ สาเหตุ)

ถ้าจะพิจารณาที่การวาดภาพ Histogram และ โค้งความถี่จะ
เป็น Normal Distribution ได้โดย

ข้อมูล

1 2 4 2
4 2 1 2
2 2 2 2
3 5 6 7
7 7 7 7
5 9 9 2
.....

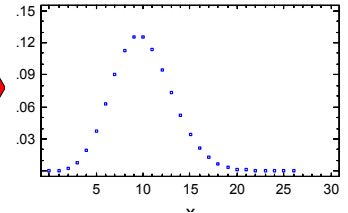
Normal Curve



ไม่ใช่ Normal Curve

ข้อมูล

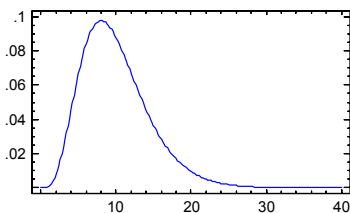
1 2 4 2
4 2 1 2
2 2 2 2
3 5 6 7
7 7 7 7
5 9 9 2
.....



ข้อมูล

1 2 4 2
4 2 1 2
2 2 2 2
3 5 6 7
7 7 7 7
5 9 9 2
.....

ไม่ใช่ Normal Curve



แล้วสถิติ t , F , Chi-Square , Binomial
,Poisson ก็ไม่ใช่ Normal นี้.....



ครับมันไม่ใช่ Normal แต่.....สถิติบางตัวถ้า N มีขนาด
ใหญ่มันก็จะเป็น Normal และบางตัว...ก็พัฒนามาจาก
Normal Distribution อยู่ดีครับ

สถิตินอนพารามตริก

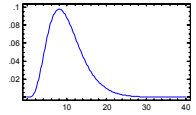
1. การทดสอบโดยใช้เครื่องหมาย (Sign Test)
2. การทดสอบอันดับที่มีเครื่องหมายกำกับของวิลคอกซอน (Wilcoxon)
3. ส.ป.ส.สหสัมพันธ์แบบอันดับของสเปียร์แมน (Spearman)
4. การทดสอบของแมนวิทนีสำหรับตัวอย่าง 2 ชุด ที่เป็นอิสระกัน (Mann-Whitney)
5. การทดสอบของฟริดแมนสำหรับการวิเคราะห์ความแปรปรวนแบบสองข้าง (Friedman)
6. การวิเคราะห์ความแปรปรวนแบบข้างเดียวของค่าอันดับ (Kruskal-Wallis)

1. การทดสอบโดยใช้เครื่องหมาย (Sign Test)

1. Ordinal Scale

ข้อมูลเป็นมาตรการวัดเรียงลำดับ หรือสูงกว่า

หรือรูปโค้งไม่เป็น Normal



1. ตั้งสมมติฐาน

$$\begin{aligned} \text{ก} \quad H_0 &: p(+) = p(-) \\ H_1 &: p(+) \neq p(-) \end{aligned}$$

$$\begin{aligned} \text{ข} \quad H_0 &: p(+) \geq p(-) \\ H_1 &: p(+) < p(-) \end{aligned}$$

$$\begin{aligned} \text{ค} \quad H_0 &: p(+) \leq p(-) \\ H_1 &: p(+) > p(-) \end{aligned}$$

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \quad \text{or} \quad \text{อื่นๆ}$$



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานที่ถูกต้อง

3. สถิติที่ใช้ในการทดสอบ

$$T = \text{จำนวนคู่ที่เป็น} +$$

4. อาณาเขตวิกฤต

พิจารณากรณีที่ไม่มี tie หรือ 0 นั่นคือ ให้ n = จำนวนคู่ทั้งหมดที่ไม่มี tie

$$\text{ก} \quad \begin{aligned} H_0 &: p(+) = p(-) \\ H_1 &: p(+) \neq p(-) \end{aligned} \quad \text{จะปฏิเสธ } H_0 \text{ เมื่อ } T \leq W_{\frac{\alpha}{2}} \text{ หรือ } T \geq W_{1-\frac{\alpha}{2}}$$

$$\text{ข} \quad \begin{aligned} H_0 &: p(+) \geq p(-) \\ H_1 &: p(+) < p(-) \end{aligned} \quad \text{จะปฏิเสธ } H_0 \text{ เมื่อ } T \leq W_{\alpha}$$

$$\text{ค} \quad \begin{aligned} H_0 &: p(+) \leq p(-) \\ H_1 &: p(+) > p(-) \end{aligned} \quad \text{จะปฏิเสธ } H_0 \text{ เมื่อ } T \geq W_{1-\alpha}$$

หาค่า W ได้โดยการเปิดตาราง

สำคัญ

ตารางที่แสดงเปิดแบบสองข้างจะครบ ถ้าจะเปิดแบบข้างเดียวต้องนำค่านัยสำคัญไปคูณ 2 ก่อน

และตารางเป็นการเปิดด้านซ้ายเท่านั้น ถ้าจะเปิดด้านขวาต้องนำ $n - W$

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

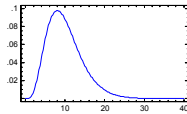
- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

2. การทดสอบอันดับที่มีเครื่องหมายกำกับของวิลคอกซอน (Wilcoxon)

1. Interval Scale

ข้อมูลเป็นมาตรการวัดอัตราภาคหรือสูงกว่า

หรือรูปโค้งไม่เป็น Normal



1. ตั้งสมมติฐาน

$$\begin{aligned} \text{ก. } H_0 &: E(X) = E(Y) \\ H_1 &: E(X) \neq E(Y) \end{aligned}$$

$$\begin{aligned} \text{ข. } H_0 &: E(X) \geq E(Y) \\ H_1 &: E(X) < E(Y) \end{aligned}$$

$$\begin{aligned} \text{ค. } H_0 &: E(X) \leq E(Y) \\ H_1 &: E(X) > E(Y) \end{aligned}$$

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \quad \text{or} \quad \text{อื่นๆ}$$



α คือ ระดับนัยสำคัญ หรือความคาดเคลื่อนชนิดที่ 1 หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

$$T = \sum_{i=1}^n R_i$$

D_i ผลต่างของ X_i กับ Y_i

$R_i = 0$ ถ้า $X_i < Y_i$ นั่นคือ D_i เป็นค่าลบ

$R_i =$ อันดับที่ให้แก่คู่ของ $X_i > Y_i$ นั่นคือ D_i เป็นค่าบวก

4. อาณาเขตวิกฤต

1. ในกรณีตัวอย่างขนาดเล็ก $n \leq 20$

ก $H_0 : E(X) = E(Y)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$
 $H_1 : E(X) \neq E(Y)$

ข $H_0 : E(X) \geq E(Y)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\alpha}$
 $H_1 : E(X) < E(Y)$

ค $H_0 : E(X) \leq E(Y)$ จะปฏิเสธ H_0 เมื่อ $T > W_{1-\alpha}$
 $H_1 : E(X) > E(Y)$

หาค่า W ได้โดยการเปิดตาราง

ตารางเปิดได้เฉพาะด้านซ้าย ถ้าต้องการเปิดด้านขวา
ทำได้โดย

$$W_p = \frac{n(n+1)}{2} - W_{1-p}$$

p คือ พื้นที่ α หรือ $\frac{\alpha}{2}$

4. อาณาเขตวิกฤต

2. ในกรณีตัวอย่างขนาดใหญ่ $n > 20$

$$W_p = \frac{n(n+1)}{4} \pm Z_p \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

★
พิจารณาว่าค่าที่ไม่ใช่ tie หรือ 0 นั่นคือ ให้ $n =$ จำนวนคู่ทั้งหมดที่ไม่ใช่ tie

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปใน
ขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของ
ใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับ
นัยสำคัญ _____ และอธิบายผล

ตัวอย่าง

- บริษัทผลิตยาแห่งหนึ่งต้องการศึกษาถึงประสิทธิภาพของยาลดน้ำหนักรุ่นหนึ่ง จึงได้ทำการทดสอบกับคน 8 คน โดยให้ทานยาชนิดนี้ไปชั่วระยะเวลาหนึ่ง แล้วชั่งน้ำหนักเพื่อเปรียบเทียบกับน้ำหนักก่อนทานยานี้ ได้ผลดังนี้

คน	1	2	3	4	5	6	7	8
น้ำหนักก่อน X	68	49	53	75	49	41	58	75
น้ำหนักหลัง Y	63	41	54	71	39	44	58	67

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่ายาลดน้ำหนักดังกล่าวใช้ได้ผลดี

3. ส.ป.ส.สหสัมพันธ์แบบอันดับ ของสเปียร์แมน (Spearman)

ในการศึกษาความสัมพันธ์ในกรณีที่ข้อมูลไม่มีกรแจกแจงแบบปกติ
หรือข้อมูลเป็นมาตราการวัดแบบ เรียงลำดับ (Ordinal Scale) เราไม่สามารถ
ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (r) ในการหาค่าความสัมพันธ์ได้

ในกรณีแบบนี้แสดงว่าต้องใช้ Nonparametric ช่วยในการศึกษา
ความสัมพันธ์ ซึ่งเราจะใช้ สัมประสิทธิ์สหสัมพันธ์แบบอันดับของสเปียร์
แมนแทน (r_s)

หลักการพิจารณา

- หลักการพิจารณาค่า r_s เหมือนกับกับ r ทุกประการคือ

โดยสัมประสิทธิ์สหสัมพันธ์ (r_s) จะมีค่าตั้งแต่ -1 ถึง 1

ถ้า $r_s = 1$ หมายความว่า ตัวแปร X และ Y มีความสัมพันธ์เชิงบวกแบบตามกันโดยสมบูรณ์

ถ้า $r_s = 0$ หมายความว่า ตัวแปร X และ Y ไม่มีความสัมพันธ์เชิงเส้นกัน

ถ้า $r_s = -1$ หมายความว่า ตัวแปร X และ Y มีความสัมพันธ์เชิงลบกันโดยสมบูรณ์

การคำนวณ

$$r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

$$D_i = R(X_i) - R(Y_i)$$

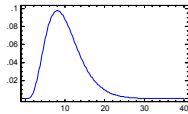
R คือการเรียงลำดับ

3. การทดสอบ ส.ป.ส.สหสัมพันธ์แบบอันดับ ของสเปียร์แมน (Spearman)

1. Ordinal Scale

ข้อมูลเป็นมาตรการวัดเรียงลำดับ
หรือสูงกว่า

หรือรูปโค้งไม่เป็น Normal



1. ตั้งสมมติฐาน

ก $H_0 : \rho = 0$
 $H_1 : \rho \neq 0$

ข $H_0 : \rho \geq 0$ $H_0 : \rho = 0$
 $H_1 : \rho < 0$ หรือ $H_1 : \rho < 0$

ค $H_0 : \rho \leq 0$ $H_0 : \rho = 0$
 $H_1 : \rho > 0$ หรือ $H_1 : \rho > 0$

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \quad \text{or} \quad \text{อื่นๆ}$$



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

$$T = r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

4. อาณาเขตวิกฤต

1. ในกรณีตัวอย่างขนาดเล็ก $n \leq 30$

ก $H_0 : \rho = 0$ จะปฏิเสธ H_0 เมื่อ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$
 $H_1 : \rho \neq 0$

ข $H_0 : \rho \geq 0$ จะปฏิเสธ H_0 เมื่อ $T < W_{\alpha}$
 $H_1 : \rho < 0$

ค $H_0 : \rho \leq 0$ จะปฏิเสธ H_0 เมื่อ $T > W_{1-\alpha}$
 $H_1 : \rho > 0$

หาค่า W ได้โดยการเปิดตาราง

4. อาณาเขตวิกฤต

1. ในกรณีตัวอย่างขนาดใหญ่ $n > 30$

ก $H_0 : \rho = 0$ จะปฏิเสธ H_0 เมื่อ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$
 $H_1 : \rho \neq 0$

ข $H_0 : \rho \geq 0$ จะปฏิเสธ H_0 เมื่อ $T < W_{\alpha}$
 $H_1 : \rho < 0$

ค $H_0 : \rho \leq 0$ จะปฏิเสธ H_0 เมื่อ $T > W_{1-\alpha}$
 $H_1 : \rho > 0$

หาค่า W ได้โดยการคำนวณ

$$W_p = \frac{Z_p}{\sqrt{n-1}} \quad p \text{ คือ พื้นที่ } \alpha \text{ หรือ } \frac{\alpha}{2}$$

ค่าด้านซ้ายหาได้โดย $W_p = -W_{1-p}$

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

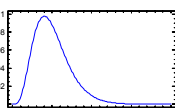
4. การทดสอบสำหรับตัวอย่างสองชุด ที่อิสระกัน

Mann-Whitney Test

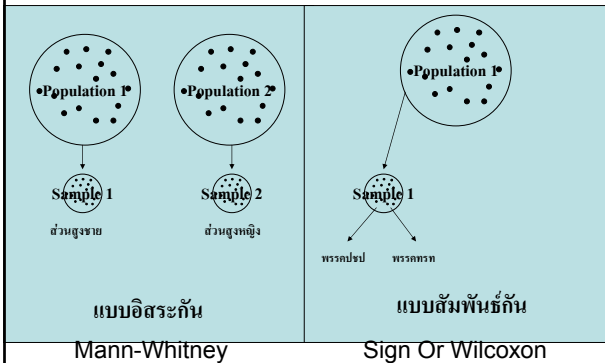
1. Ordinal Scale

ข้อมูลเป็นมาตราการวัดเรียงลำดับหรือสูงกว่า

หรือรูปร่างไม่เป็น Normal



ข้อแตกต่างระหว่างกลุ่มตัวอย่าง 2 กลุ่ม ที่อิสระกันและสัมพันธ์กัน



1. ตั้งสมมติฐาน

ก $H_0 : E(X) = E(Y)$
 $H_1 : E(X) \neq E(Y)$

ข $H_0 : E(X) \geq E(Y)$
 $H_1 : E(X) < E(Y)$

ค $H_0 : E(X) \leq E(Y)$
 $H_1 : E(X) > E(Y)$

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \text{ or อื่นๆ}$$



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

$$T = S - \frac{n(n+1)}{2}$$

S = ผลรวมของค่าอันดับ (Rank) ที่ให้แก่ข้อมูลที่ได้อาจมาจากการสุดที่ 1

$$S = \sum_{i=1}^n R(x_i)$$

ดังนั้นต้องมีการเรียงลำดับข้อมูลทั้งสองชุดก่อน
โดยนำมารวมกันเรียงพร้อมกันทั้งหมด

4. อาณาเขตวิกฤต

1. ในกรณีตัวอย่างขนาดเล็ก

ก $H_0 : E(X) = E(Y)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$
 $H_1 : E(X) \neq E(Y)$

ข $H_0 : E(X) \geq E(Y)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\alpha}$
 $H_1 : E(X) < E(Y)$

ค $H_0 : E(X) \leq E(Y)$ จะปฏิเสธ H_0 เมื่อ $T > W_{1-\alpha}$
 $H_1 : E(X) > E(Y)$

หาค่า W ได้โดยการเปิดตาราง (ข้างซ้าย)

$$W_{\text{ขวา}} = nm - W_{\text{ซ้าย}}$$

$$W_{1-p} = nm - W_p$$

4. อาณาเขตวิกฤต

1. ในกรณีตัวอย่างขนาดใหญ่ n หรือ $m > 20$

ก $H_0 : E(X) = E(Y)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\frac{\alpha}{2}}$ หรือ $T > W_{1-\frac{\alpha}{2}}$
 $H_1 : E(X) \neq E(Y)$

ข $H_0 : E(X) \geq E(X)$ จะปฏิเสธ H_0 เมื่อ $T < W_{\alpha}$
 $H_1 : E(X) < E(X)$

ค $H_0 : E(X) \leq E(Y)$ จะปฏิเสธ H_0 เมื่อ $T > W_{1-\alpha}$
 $H_1 : E(X) > E(Y)$

หาค่า W_p ได้โดยการคำนวณ

$$W_p = \frac{nm}{2} + Z_p \sqrt{\frac{nm(n+m+1)}{12}}$$

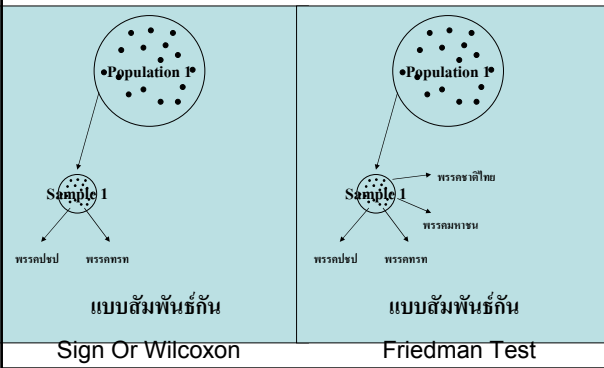
p คือ พื้นที่ α หรือ $\frac{\alpha}{2}$ หรือ $1 - \alpha$ หรือ $1 - \frac{\alpha}{2}$

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

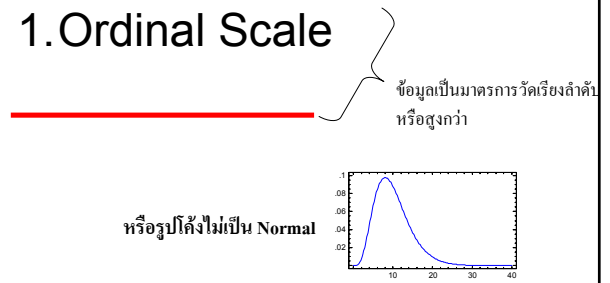
- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

5. การทดสอบของฟริดแมนสำหรับการวิเคราะห์ความแปรปรวนแบบสองข้าง (Friedman) ->RBD



การทดสอบของฟริดแมนสำหรับการวิเคราะห์ความแปรปรวนแบบสองข้าง (Friedman)

1. Ordinal Scale



1. ตั้งสมมติฐาน

H_0 : แต่ละทริทเมนที่มีอิทธิพลเหมือนกัน

H_1 : อย่างน้อย 2 ทริทเมนที่มีอิทธิพลต่างกัน

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.01$

$\alpha = 0.05$

$\alpha = 0.10$ or อื่นๆ



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

$$T = \frac{12}{bk(k+1)} \sum_{j=1}^k R_j^2 - 3b(k+1)$$

$$R_j = \sum_{i=1}^b R(X_{ij}) \text{ สำหรับ } j=1, 2, \dots, k$$

k = จำนวน Treatment , b = จำนวน Block

4. อาณาเขตวิกฤต

จะปฏิเสธ H_0 เมื่อ $T > \chi_{1-\alpha}^2$

$$df = k - 1$$

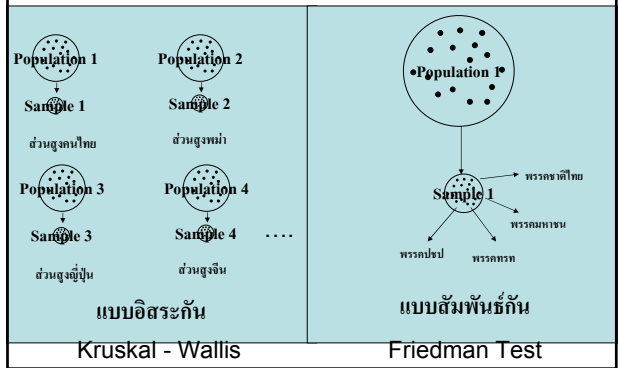
5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

6. การวิเคราะห์ความแปรปรวนแบบข้างเดียวของค่าอันดับ

(Kruskal-Wallis) -> CRD



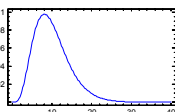
การวิเคราะห์ความแปรปรวนแบบข้างเดียวของค่าอันดับ

(Kruskal-Wallis)

1. Interval Scale

ข้อมูลเป็นมาตรการวัดอัตรา
หรือสูงกว่า

หรือรูปโค้งไม่เป็น Normal



1. ตั้งสมมติฐาน

H_0 : แต่ละทริทเมนต์ที่มีอิทธิพลเหมือนกัน

H_1 : อย่างน้อย 2 ทริทเมนต์ที่มีอิทธิพลต่างกัน

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10 \text{ or อื่นๆ}$$



α คือ ระดับนัยสำคัญ

หรือความคาดเคลื่อนชนิดที่ 1

หรือการปฏิเสธสมมติฐานต่างๆที่เป็นจริง

3. สถิติที่ใช้ในการทดสอบ

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

$$R_i = \sum_{j=1}^n R(X_{ij}) \text{ สำหรับ } j=1, 2, \dots, k$$

N = จำนวนข้อมูลที่สุ่มมาทั้งหมด, n_i = จำนวนข้อมูลที่สุ่มมาในแต่ละ Treatment

4. อาณาเขตวิกฤต

จะปฏิเสธ H_0 เมื่อ $T > \chi_{1-\alpha}^2$

$$df = k - 1$$

5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล

สุ่มข้อมูลเกรดเฉลี่ยของนักศึกษาเรียนสถิติ 2 มาจำนวนคณะละ 5 คน
ได้ข้อมูลเกรดเฉลี่ยดังนี้

คณะสังคม	คณะเศรษฐศาสตร์	คณะบริหาร
2.06	2.75	3.2
2.98	2.89	3.3
2.28	3	2.5
3.5	2.95	2.8
3.24	3.59	2.5

ที่ระดับนัยสำคัญ 0.05 บอกได้หรือไม่ว่าทุกคณะมีเกรดเฉลี่ยเท่ากัน

1

H_0 : แต่ละทริทเมนที่มีอิทธิพลเหมือนกัน
 H_1 : อย่างน้อย 2 ทริทเมนที่มีอิทธิพลต่างกัน

2

$$\alpha = 0.05$$

3

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

เรียกลำดับข้อมูลทั้งหมดด้วยวิธี Watha's Rank

1. เรียงลำดับแต่ละ Treatment

คณะสังคม	คณะเศรษฐศาสตร์	คณะบริหาร
2.06	2.75	2.5
2.28	2.89	2.5
2.98	2.95	2.8
3.24	3	3.2
3.5	3.59	3.3

2. ให้อันดับข้อมูลทั้งหมด

คณะสังคม	คณะเศรษฐศาสตร์	คณะบริหาร
1	5	3.5
2	7	3.5
9	8	6
12	10	11
14	15	13

3. หาผลรวมของแต่ละ Treatment


	คณะสังคม	คณะเศรษฐศาสตร์	คณะบริหาร
1	5		3.5
2	7		3.5
9	8		6
12	10		11
14	15		13
ผลรวม	38	45	37

$$T = \frac{12}{15(15+1)} \left(\frac{38^2}{5} + \frac{45^2}{5} + \frac{37^2}{5} \right) - 3(15+1)$$

3 T = 0.38

4 $\chi^2_{95,2} = 5.99$

5 ยอมรับ H_0 ที่ระดับนัยสำคัญ 0.05 แสดงว่าเกรดเฉลี่ยของนักศึกษาที่เรียนสถิติ 2 ของแต่ละคณะไม่แตกต่างกัน



เรื่องสุดท้าย การทดสอบข้อมูลว่าเป็นการแจกแจงปกติหรือไม่ (Goodness of fit Test)

สรุปการแจกแจงที่สามารถตั้งข้อสังเกต(ใน 208272)

- ข้อมูลเป็นการแจกแจงแบบทวินาม นั่นคือ มีจำนวนครั้งที่การทดลองที่แน่นอน และข้อมูลเป็นแบบ ไม่ต่อเนื่อง
- ข้อมูลเป็นการแจกแจงแบบปัวซอง นั่นคือ มีเวลาและพื้นที่เป็นข้อจำกัด สิ่งที่เกิดจากการทดลอง มีจำนวนได้ไม่แน่นอน(ไม่จำกัด) และข้อมูลเป็นแบบ ไม่ต่อเนื่อง
- ข้อมูลเป็นการแจกแจงแบบปกติ นั่นคือ ข้อมูลเป็นแบบ ต่อเนื่อง ทุกชนิด (ข้อมูลช่วง ตวง วัด)

1. การตั้งสมมติฐาน

H_0 : ข้อมูลมีการแจกแจงเป็นปกติ
 H_1 : ข้อมูลไม่มีการแจกแจงเป็นปกติ

2. กำหนดระดับนัยสำคัญ

$\alpha = 0.01$
 $\alpha = 0.05$
 $\alpha = 0.10$ or อื่นๆ

3. สถิติที่ใช้ทดสอบ

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$



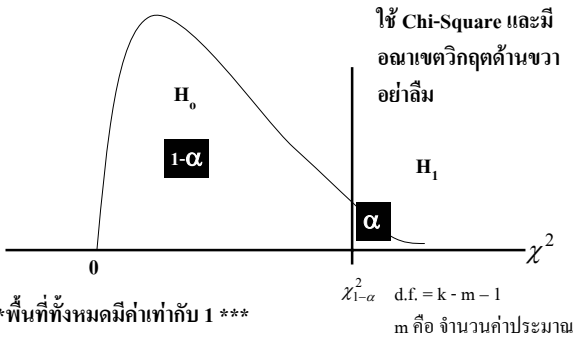
$$\chi^2 = \sum_{i=1}^k \frac{O_i^2}{E_i} - N$$

O_i = ค่าความถี่สังเกต

E_i = ค่าความถี่คาดหวัง (ต้องคำนวณ)

N = จำนวนครั้งที่ทำการทดลอง
หรือจำนวนที่สังเกตทั้งหมด

4. หาอณาเขตวิกฤต



5. สรุปผล และ ตีความหมาย

เอาค่าที่คำนวณได้ในขั้นตอนที่ 3 ไป Plot จุดลงไปในขั้นตอนที่ 4 แล้วพิจารณาว่าค่าที่คำนวณได้ตกอยู่ในพื้นที่ของใคร

- ก. ถ้าตกอยู่ในพื้นที่ของ H_0 ให้ตอบว่า “ยอมรับ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล (การแจกแจงปกติ)
- ข. ถ้าตกอยู่ในพื้นที่ของ H_1 ให้ตอบว่า “ปฏิเสธ H_0 ที่ระดับนัยสำคัญ _____ และอธิบายผล (การแจกแจงเป็นไม่เป็นปกติ)

ความสูง (นิ้ว)	จำนวนนักเรียน
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
รวม	100

จากข้อมูลความสูงของนักเรียน 100 คน จงทดสอบว่าความสูงของนักเรียนกลุ่มนี้มีการแจกแจงแบบปกติหรือไม่ที่ระดับนัยสำคัญ 0.01

1. การตั้งสมมติฐาน

H_0 : ความสูงของนักเรียนมีการแจกแจงเป็นปกติ

H_1 : ความสูงของนักเรียน ไม่มีการแจกแจงเป็นปกติ

2. กำหนดระดับนัยสำคัญ

$$\alpha = 0.01$$

3. เราต้องหา E (ค่าคาดหวัง) ซึ่งต้องทราบ p (ความน่าจะเป็นของแต่ละกลุ่ม) ซึ่งจะทราบ p ได้ต้องทราบ \bar{X} และ SD

ความสูง (นิ้ว)	จำนวนนักเรียน (f)	หาจุดกึ่งกลางชั้น (x)	fx	fx ²
60-62	5	61	305	18605
63-65	18	64	1152	73728
66-68	42	67	2814	188538
69-71	27	70	1890	132300
72-74	8	73	584	42632
รวม	100		6745	455803

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = 67.45$$

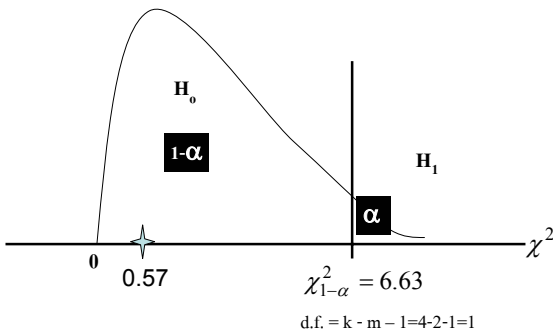
$$S^2 = \frac{\sum f_i x_i^2 - n(\bar{x})^2}{n-1} = 8.6138, S = 2.94$$

} m = 2

ขอบเขตชั้น	h1 Z	h1 p _i	E _i = Np _i	O _i	(ถุน) E _i	(ถุน) O _i	O _i ² /E _i	
<-59.5	(-Infinity) - (-2.7)	0.0035	0.35	0				
59.5 - 62.5	(-2.7) - (-1.68)	0.043	4.3	5				
62.5 - 65.5	(-1.68) - (-0.66)	0.2081	20.81	18	25.46	23	20.78	
65.5 - 68.5	(-0.66) - (0.36)	0.386	38.6	42	38.6	42	45.70	
68.5 - 71.5	(0.36) - (1.38)	0.2756	27.56	27	27.56	27	26.45	
71.5 - 74.5	(1.38) - (2.4)	0.0756	7.56	8	8.38	8	7.64	
74.5 -->	(2.4) - (+Infinity)	0.0082	0.82	0				
รวม							100.57	

$$\chi^2 = \sum_{i=1}^k \frac{O_i^2}{E_i} - N = 100.57 - 100 = .57$$

4. หาค่าขอบเขตวิกฤต



5. สรุปผล และ ตีความหมาย

ตกอยู่ในพื้นที่ของ H₀ ให้ตอบว่า “ยอมรับ H₀ ที่ระดับนัยสำคัญ 0.01 กล่าวได้ว่า การแจกแจงของความสูงของนักเรียนมีการแจกแจงเป็นปกติ

Midterm = 217 หน้า

Final = 193 หน้า

รวม = 410 หน้า

อ่าน 100 หน้า ได้ 24 คะแนน

อ่าน 410 หน้า ได้ 98.4 คะแนน

อีก 1.6 อยู่ที่บุญ/กรรมทำกันมาครับ



Do you have any Question?



Thank you

http://beam.to/statistics

STAT
STATISTICS DEPARTMENT

HOME | CONTACT | BUREAU | LINKS

NEWS

DOWNLOAD

ITIMETABLE

PROFILE

WEBBOARD

มหาวิทยาลัยเชียงใหม่

มหาวิทยาลัยเชียงใหม่

- มหาวิทยาลัยเชียงใหม่
- มหาวิทยาลัยเชียงใหม่
- มหาวิทยาลัยเชียงใหม่

Statistics Department
Faculty of Science, Chiangmai University
© Copyright 2009. All rights reserved.

Created by: Chaiyaporn