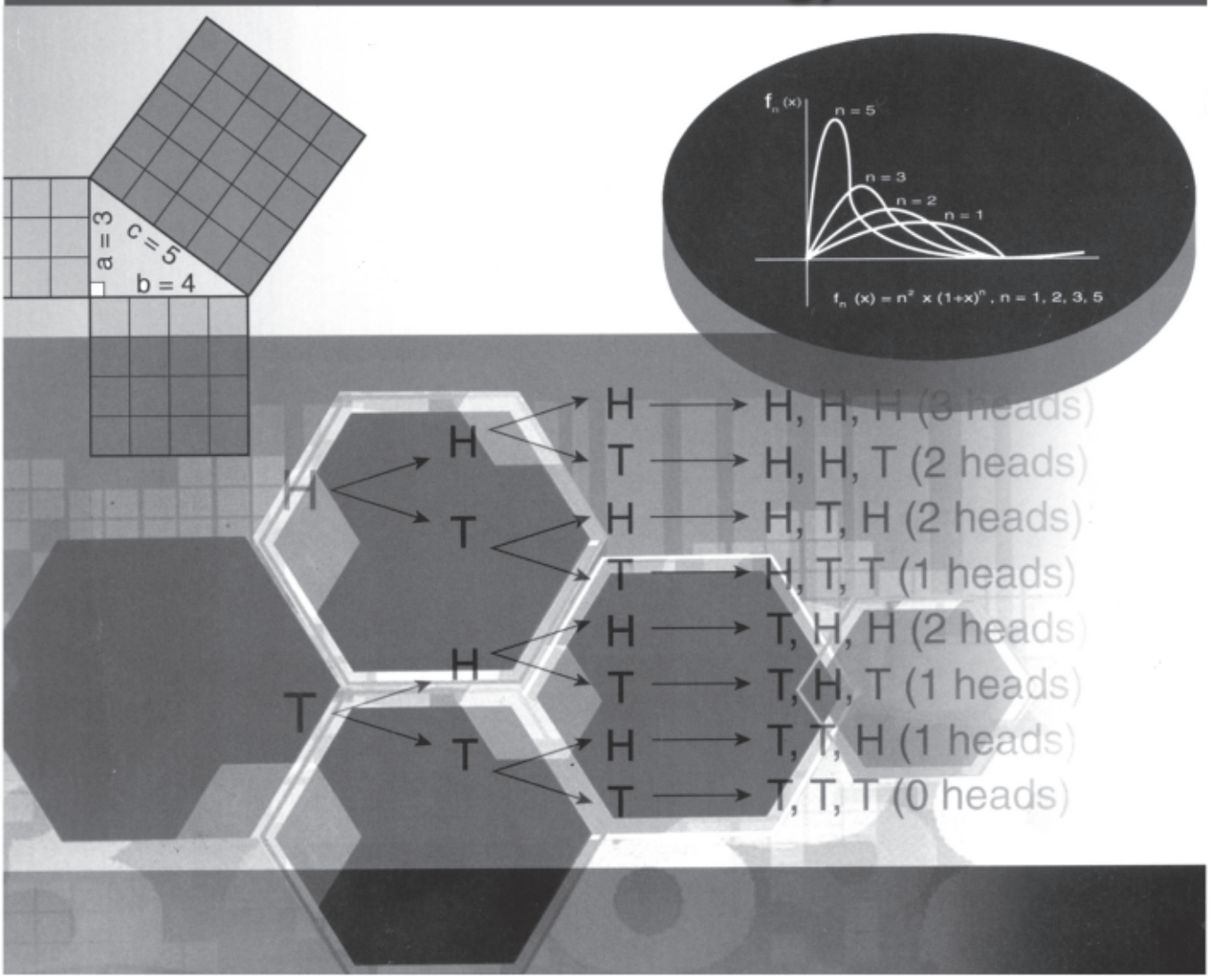
สาขาวิชาวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยสุโขทัยธรรมาธิราช

96102
การสอนเสริมครั้งที่ 3

เอกสารโสตทัศน์ชุดวิชา

คณิตศาสตร์และสถิติสำหรับวิทยาศาสตร์และเทคโนโลยี

Mathematics and Statistics for Science and Technology



เอกสารโสตทัศน์ชุดวิชา คณิตศาสตร์และสถิติสำหรับวิทยาศาสตร์และเทคโนโลยี การสอนเสริมครั้งที่ 3
จัดทำขึ้นเพื่อเป็นบริการแก่นักศึกษาในการสอนเสริม

จัดทำต้นฉบับ : คณะกรรมการกลุ่มผลิตชุดวิชา

บรรณาธิการ/ออกแบบ : หน่วยผลิตสื่อสอนเสริม ศูนย์โสตทัศนศึกษา
สำนักเทคโนโลยีการศึกษา

จัดพิมพ์ : สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช

พิมพ์ที่ : โรงพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช

พิมพ์ครั้งที่ 9 ภาค 1/2553 ปรับปรุง

เอกสารโสตทัศน์
ประกอบการสอนเสริม
ครั้งที่ 3

ชุดวิชา คณิตศาสตร์และสถิติสำหรับวิทยาศาสตร์และเทคโนโลยี

หน่วยที่ 11–15

หน่วยที่ 11 การแจกแจงตัวแปรสุ่ม

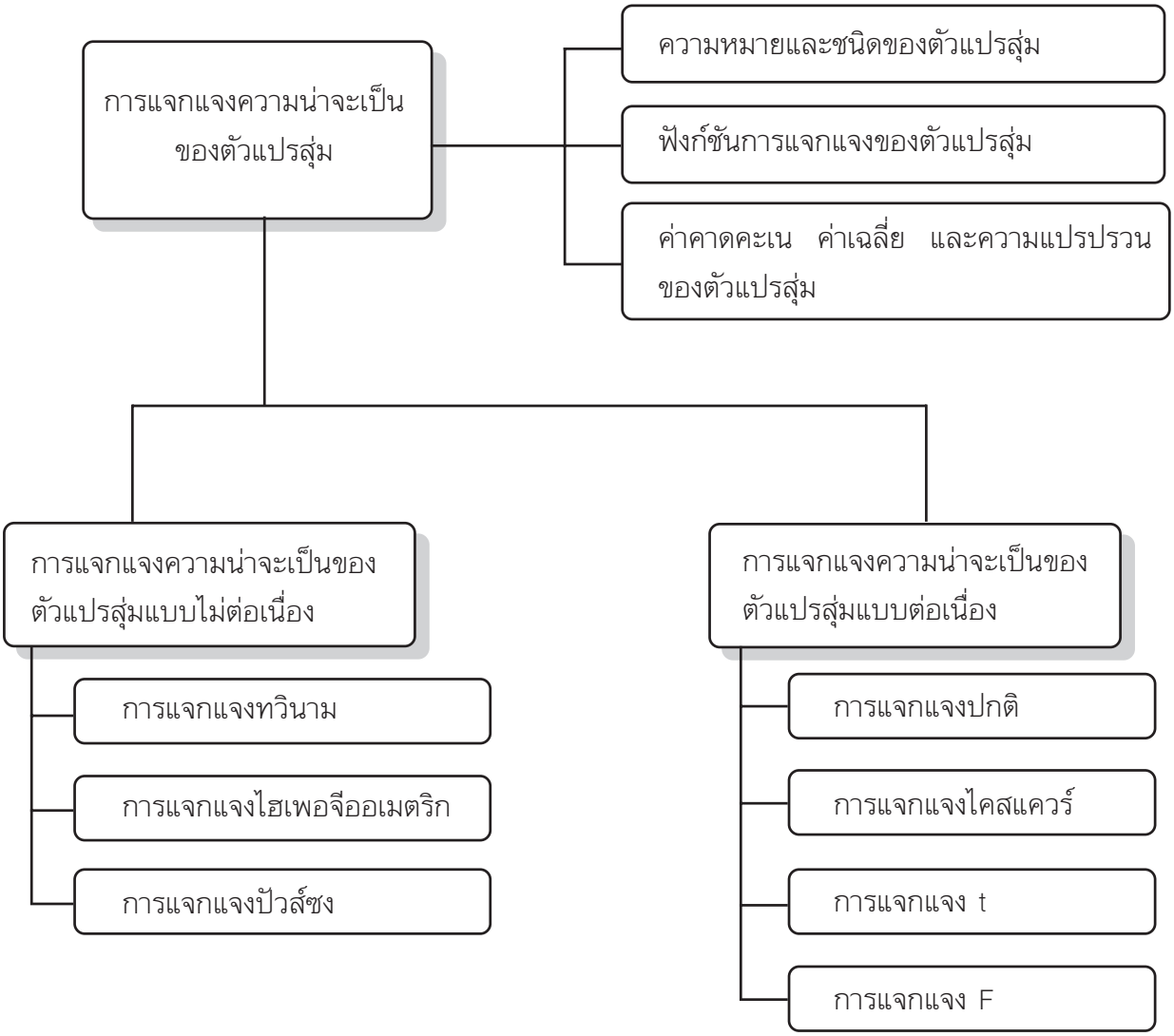
หน่วยที่ 12 สถิติศาสตร์อิงพารามิเตอร์เบื้องต้น

หน่วยที่ 13 สถิติศาสตร์ไม่อิงพารามิเตอร์เบื้องต้น

หน่วยที่ 14 การวิเคราะห์สหสัมพันธ์และการถดถอยเชิงเส้นอย่างง่าย

หน่วยที่ 15 การประยุกต์สถิติทางวิทยาศาสตร์และเทคโนโลยี

หน่วยที่ 11
การแจกแจงตัวแปรสุ่ม



โสตทัศน์ # 11.1 ตัวแปรสุ่ม

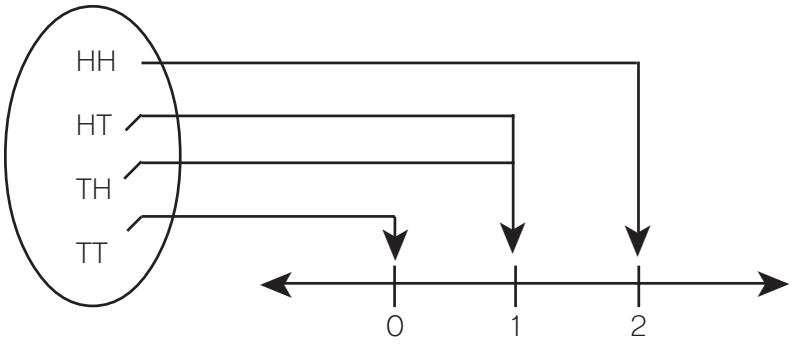
ตัวแปรสุ่ม คือฟังก์ชันที่มีโดเมนเป็นปริภูมิตัวอย่างของการทดลองเชิงสุ่มหนึ่ง และมีเรนจ์เป็นสับเซตของเซตของจำนวนจริง

ตัวอย่าง เช่น ในการโยนเหรียญสองอันหนึ่งครั้ง เราสนใจว่าเหรียญจะหงายด้านหัวกี่เหรียญ ถ้าแทนหัวด้วย H และแทนก้อยด้วย T เหตุการณ์ที่เป็นไปได้มีดังนี้

1. ไม่มีเหรียญหงายด้านหัว แทนด้วย TT หรือมีจำนวนหัวเท่ากับ 0
2. มีหนึ่งเหรียญที่หงายด้านหัว แทนด้วย HT หรือ TH หรือมีจำนวนหัวเท่ากับ 1

และ 3. เหรียญทั้งสองหงายด้านหัว แทนด้วย HH หรือมีจำนวนหัวเท่ากับ 2

ถ้าให้ตัวแปรสุ่ม X แทนจำนวนหัว X คือฟังก์ชันที่จับคู่เหตุการณ์กับจำนวนจริง แสดงได้ดังแผนภาพ



ตัวอย่างของตัวแปรสุ่ม เช่น

1. ครอบครัวหนึ่งมีบุตร 3 คน ถ้าให้ X แทนจำนวนบุตรชาย ตัวแปรสุ่ม X มีค่าเป็น 0, 1, 2, 3
2. ถ้าครอบครัวหนึ่งต้องการมีบุตรชาย 1 คน ให้ Y แทนจำนวนบุตรของครอบครัวนี้ก่อนที่จะมีบุตรชาย ตัวแปรสุ่ม Y มีค่าเป็น 1, 2, 3, 4, . . .
3. ให้ Y แทนรายได้ต่อปีของครอบครัวหนึ่ง ตัวแปรสุ่ม Y มีค่าเป็นจำนวนจริง
4. ให้ Z แทนอายุการใช้งานของเครื่องใช้ไฟฟ้าชนิดหนึ่ง ตัวแปรสุ่ม Z มีค่าเป็นจำนวนจริงที่ไม่น้อยกว่าศูนย์ เป็นต้น

ตัวอย่าง ในการทอดลูกเต๋าเที่ยงตรงสองลูกหนึ่งครั้ง ถ้าตัวแปรสุ่มที่สนใจได้แก่

X แทนผลรวมของหน้าลูกเต๋าทั้งสองลูก

Y แทนผลต่างของหน้าลูกเต๋าทั้งสองลูก

Z แทนค่าสัมบูรณ์ของผลต่างของหน้าลูกเต๋าทั้งสองลูก

ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือจำนวนที่อยู่ในเซต {2, 3, 4, . . . , 12}

ค่าที่เป็นไปได้ของตัวแปรสุ่ม Y คือจำนวนที่อยู่ในเซต { –5, –4, . . . , –1, 0, 1, . . . , 4, 5}

ค่าที่เป็นไปได้ของตัวแปรสุ่ม Z คือจำนวนที่อยู่ในเซต {0, 1, 2, 3, 4, 5}

โสตทัศน์ # 11.2 ชนิดของตัวแปรสุ่ม

ตัวแปรสุ่มมีลักษณะที่แตกต่างกันแบ่งได้เป็น 2 แบบได้แก่

1. ตัวแปรสุ่มแบบไม่ต่อเนื่อง (discrete random variable) คือฟังก์ชันจากปริภูมิตัวอย่างของการทดลองเชิงสุ่มหนึ่งที่มีเรนจ์เป็นเซตที่นับได้ (countable set) ซึ่งอาจเป็นได้ทั้งเซตจำกัดหรือเซตอนันต์ เช่น จำนวนแต้มบนหน้าที่หงายในการทอดลูกเต๋าหนึ่งลูกหนึ่งครั้ง จำนวนเด็กชายเกิดใหม่ต่อวันในโรงพยาบาลแห่งหนึ่งจำนวนครั้งของการโยนเหรียญเที่ยงตรงอันหนึ่งจนกว่าจะเหรียญจะหงายหัว จำนวนครั้งของโทรศัพท์ที่เรียกเข้ามาในสำนักงานหนึ่งใน 1 ชั่วโมง เป็นต้น

2. ตัวแปรสุ่มแบบต่อเนื่อง (continuous random variable) คือฟังก์ชันจากปริภูมิตัวอย่างของการทดลองเชิงสุ่มหนึ่งที่มีเรนจ์เป็นช่วงซึ่งเป็นสับเซตของเซตจำนวนจริง เช่น ปริมาณน้ำฝนในพื้นที่แห่งหนึ่งในเดือนพฤษภาคม ระยะเวลาที่รอรถประจำทางสายหนึ่ง อายุการใช้งานของหลอดไฟฟ้า ความเร็วเฉลี่ยของรถยนต์บนทางด่วน เป็นต้น

โสตทัศน์ # 11.3 ฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่อง

ถ้า X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่องซึ่งมีค่าอยู่ใน $\{x_1, x_2, x_3, \ldots\}$ ความน่าจะเป็นที่ตัวแปรสุ่มมีค่าค่าหนึ่งกำหนดโดย $P(X = x_i) = P(\{s ; X(s) = x_i\})$; $i = 1, 2, 3, \ldots$ และเราเรียก $P(X = x_i)$ ว่าฟังก์ชันความน่าจะเป็น (the probability function) ของตัวแปรสุ่ม X ซึ่งต่อไปจะเขียนแทนด้วย f(x)

โดยที่ $P(A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} f(x)$

หมายเหตุ $\sum_{x \in A} f(x)$ หมายถึงผลรวมของค่า f(x) สำหรับทุกๆ ค่า x ที่อยู่ในเซต A

บทนิยาม ฟังก์ชันความน่าจะเป็น (Probability function)

ถ้า X คือตัวแปรสุ่มแบบไม่ต่อเนื่องที่มีค่าที่เป็นไปได้อยู่ในเซตที่นับได้ $\{x_1, x_2, x_3, \ldots\}$ แล้วฟังก์ชัน f เป็นฟังก์ชันที่มีโดเมนคือเซตจำนวนจริง และเรนจ์คือช่วง [0,1] เรียก f ว่าเป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X เมื่อ

1. $f(x_i) > 0$ สำหรับ $i = 1, 2, 3, \ldots$
2. $f(x) = 0$ สำหรับ $x \neq x_i$; $i = 1, 2, 3, \ldots$
3. $\sum f(x_i) = 1$ สำหรับ $i = 1, 2, 3, \ldots$
4. $P(A) = \sum_{x \in A} f(x)$ เมื่อ A คือเหตุการณ์บนปริภูมิตัวอย่างหนึ่ง

หมายเหตุ $\sum f(x_i)$ คือผลรวมของค่า $f(x_i)$ สำหรับทุกค่า i

โสตทัศน์ # 11.3 ฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่อง

เช่น ในการทอดลูกเต๋าเที่ยงตรง 2 ลูก 1 ครั้ง

ถ้าให้ X แทนผลรวมของแต้มที่ปรากฏ ซึ่ง X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง

ปริภูมิตัวอย่าง $S = \{(1,1), (1,2), \ldots, (6,6)\}$ ซึ่ง $n(S) = 36$

จะได้ว่า X อาจมีค่าเป็น 2, 3, 4, . . . , 12

เราสามารถหาความน่าจะเป็นที่ X มีค่าต่างๆ ดังกล่าวได้ เช่น

$\{ X = 5 \}$ คือเหตุการณ์ $\{(1,4), (4,1), (2,3), (3,2)\}$ จะได้ $P(X = 5) = \frac{4}{36}$

$\{ X = 7 \}$ คือเหตุการณ์ $\{(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)\}$ จะได้ $P(X = 7) = \frac{6}{36}$ เป็นต้น

ดังนั้นเราอาจแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ได้ดังตาราง

x	2	3	4	5	6	7	8	9	10	11	12
P(X=x) หรือ f(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

โสตทัศน์ # 11.4 ฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่อง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่อง X เขียนแทนได้ด้วยเส้นโค้งของฟังก์ชันที่เรียกว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) คือ f(x) ซึ่งใช้ในการคำนวณความน่าจะเป็นของเหตุการณ์ A

บทนิยาม ฟังก์ชันความหนาแน่นของความน่าจะเป็น

ฟังก์ชัน f ใดๆที่มีโดเมนเป็นเซตจำนวนจริงและเรนจ์คือ $[0,\infty)$ เรา**เรียก f ว่าฟังก์ชันความหนาแน่นของความน่าจะเป็น** (probability density function หรือเรียกสั้นๆ ว่า p.d.f.) ก็ต่อเมื่อ

1. $f(x) > 0$ สำหรับทุกๆค่า x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

โสตทัศน์ # 11.4 (ต่อ)

ถ้า X เป็นตัวแปรสุ่มแบบต่อเนื่อง ความน่าจะเป็นของเหตุการณ์ A ที่ X มีค่าบนช่วง (a,b) คือ

$$P(A) = P(\{s\in S \mid a < X(s) < b\}) = P(a < X < b) = \int_a^b f(x)dx \text{ เมื่อ } a < b$$

ถ้า X เป็นตัวแปรสุ่มแบบต่อเนื่อง จะได้ว่า **ความน่าจะเป็นของเหตุการณ์ A** ที่ตัวแปรสุ่ม X มีค่าบนช่วง (a,b) คือ**พื้นที่ใต้เส้นกราฟของ f(x) ที่อยู่ระหว่าง x = a และ x = b** และความน่าจะเป็นที่ X = a หรือ P(X = a) = 0 เพราะว่า

$$P(X = a) = \int_a^a f(x)dx = 0$$

ดังนั้น $P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$

โสตทัศน์ # 11.5 ฟังก์ชันการแจกแจง (Distribution functions)

ฟังก์ชันการแจกแจงของตัวแปรสุ่ม X บนปริภูมิตัวอย่าง S กำหนดโดยฟังก์ชัน F(x) ซึ่ง

$$F(x) = P(\{s\in S \mid -\infty < X(s) \le x\}) = P(X \le x);\ x \in \mathbf{R}$$

1. สำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง

$$F(x) = P(X < x) = \sum_{x_i \le x} f(x_i)$$

ดังนั้น $P(X = x_i) = F(x_i) - F(x_{i-1}) = P(X \le x_i) - P(X \le x_i - 1)$

หมายเหตุ $\sum_{x_i \le x} f(x_i)$ หมายถึงผลรวมของ $f(x_i)$ สำหรับ x_i ทุกๆค่าที่น้อยกว่า x

ตัวอย่าง ถ้า X คือตัวแปรสุ่มแบบไม่ต่อเนื่อง และมีฟังก์ชันความน่าจะเป็นคือ $f(x) = \frac{x}{6}$; x = 1, 2, 3

จงหา 1. $P(X \le 1)$

2. $P(X \le 3)$

โสตทัศน์ # 11.5 (ต่อ)

3. $P(X \leq \frac{3}{2})$

4. $P(X \leq \frac{7}{3})$

5. จงเขียนฟังก์ชันการแจกแจงของ X

2. สำหรับตัวแปรสุ่มแบบต่อเนื่อง

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

โดยทฤษฎีบทของแคลคูลัสและโดยที่ $P(X = a) = \int_{a}^{a} f(x)dx = 0$ จะได้ว่า

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$$

ตัวอย่าง กำหนดตัวแปรสุ่ม X เป็นตัวแปรสุ่มแบบต่อเนื่องที่มี p.d.f. คือ

$$f(x) = \begin{cases} x \ ; \ 0 \leq x < 1 \\ 2-x \ ; \ 1 \leq x < 2 \\ 0 \ ; \ x \text{ มีค่าอื่น ๆ} \end{cases}$$

1. จงหาฟังก์ชันการแจกแจงของ X
2. จงหาความน่าจะเป็นต่อไปนี้

2.1 $P(-1 < X \leq \frac{1}{2})$

2.2 $P(X \leq \frac{3}{2})$

2.3 $P(X \leq 3)$

2.3 $P(X \geq 2.5)$

โสตทัศน์ # 11.6 ค่าคาดคะเน ค่าเฉลี่ย และความแปรปรวนของตัวแปรสุ่ม

ให้ X เป็นตัวแปรสุ่ม **ค่าเฉลี่ยหรือค่าคาดคะเนของ X** เขียนแทนด้วย E(X) หรือ μ_x หรือ μ

1. ถ้า X เป็นตัวแปรสุ่มแบบ**ไม่ต่อเนื่อง**ที่มีค่า $x_1, x_2, \ldots, x_n$

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

2. ถ้า X เป็นตัวแปรสุ่มแบบ**ต่อเนื่อง**

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

ถ้า X คือตัวแปรสุ่ม เราจะได้สูตรการหา**ค่าคาดคะเนหรือค่าเฉลี่ยของ g(X)** ดังนี้

$E[g(X)] = \sum g(x)f(x)$ เมื่อ X คือตัวแปรสุ่มแบบ**ไม่ต่อเนื่อง**

และ $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$ เมื่อ X คือตัวแปรสุ่มแบบ**ต่อเนื่อง**

ตัวอย่าง ในการตรวจสอบคุณภาพของสินค้าตัวอย่างจำนวน 7 ชิ้นพบว่ามีสินค้าที่ไม่มีตำหนิ 4 ชิ้นและมีสินค้าที่มีตำหนิ 3 ชิ้น ถ้าผู้ตรวจสอบนำสินค้ามาตรวจสอบ 3 ชิ้น จงหาค่าเฉลี่ยของจำนวนสินค้าที่ไม่มีตำหนิในสินค้าตัวอย่างชุดนี้

ตัวอย่าง ให้ตัวแปรสุ่ม X แทนปริมาณ (เปอร์เซนต์) ของสารประกอบหนึ่งที่ผสมในน้ำมันเครื่องยนต์ชนิดหนึ่ง ซึ่ง p.d.f. กำหนดโดย

$$f(x) = \begin{cases} 20x^3(1-x) & ; 0 < x < 1 \\ 0 & ; x \text{ มีค่าอื่นๆ} \end{cases}$$

จงหาว่าน้ำมันเครื่องยนต์ชนิดนี้มีปริมาณของสารประกอบหนึ่งโดยเฉลี่ยเท่าใด

โสตทัศน์ # 11.7 สมบัติของค่าคาดคะเน

สมบัติของค่าคาดคะเน

ให้ X คือตัวแปรสุ่ม

1. E(c) = c เมื่อ c คือค่าคงตัว
2. E[cg(X)] = cE[g(X)] เมื่อ c คือค่าคงตัว
3. $E[c_1g_1(X) + c_2g_2(X)] = c_1E[g_1(X)] + c_2E[g_2(X)]$ เมื่อ c_1 และ c_2 คือค่าคงตัว

บทนิยาม

ให้ X คือตัวแปรสุ่มที่มีค่าเฉลี่ย μ ความแปรปรวนของ X คือ

$$Var(X) = \sigma^2 = E[(X - \mu)^2]$$

และเรียก $\sqrt{\sigma^2} = \sigma$ ว่าส่วนเบี่ยงเบนมาตรฐานของ X

ถ้า X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง $Var(X) = \sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2 f(x_i)$

และถ้า X เป็นตัวแปรสุ่มแบบต่อเนื่อง $Var(X) = \sigma^2 = \int_{-\infty}^{\infty}(x - \mu)^2 f(x)dx$

ตัวอย่าง ให้ X คือตัวแปรสุ่มแทนจำนวนชิ้นส่วนที่มีตำหนิของเครื่องยนต์เครื่องหนึ่งถ้าสุ่มชิ้นส่วนของเครื่องยนต์มาตรวจสอบ 3 ชิ้น ความน่าจะเป็นของแต่ละค่า X ดังปรากฏในตาราง

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

จงหาค่าเฉลี่ย และความแปรปรวนของ X

ตัวอย่าง ให้ X แทนปริมาณความต้องการน้ำดื่ม (ลิตร) ในชุมชนแห่งหนึ่งเป็นตัวแปรสุ่มแบบต่อเนื่องที่มี p.d.f. คือ

$$f(x) = \begin{cases} 2(x - 1) & ; \ 1 < x < 2 \\ 0 & ; \ x \text{ มีค่าอื่นๆ} \end{cases}$$

จงหาค่าเฉลี่ย และความแปรปรวนของ X

โสตทัศน์ # 11.8 สมบัติของความแปรปรวน

1. ถ้า $g(X) = X + c$ โดยที่ c คือค่าคงตัว
   $Var(X + c) = Var(X)$

เช่น ถ้า $Var(X) = 6$ จะได้ว่า $Var(X + 9) = Var(X) = 6$

2. ถ้า $g(X) = aX$ โดยที่ a คือค่าคงตัว
   $Var(aX) = a^2 Var(X)$

เช่น ถ้า $Var(X) = 6$ จะได้ว่า $Var(5X) = 25Var(X) = 25(6) = 150$

โสตทัศน์ # 11.9 การแจกแจงทวินาม

ถ้าในการทดลองเชิงสุ่มหนึ่ง ผลลัพธ์ที่เป็นไปได้มีเพียงสองทางคือ “สำเร็จ” และ “ไม่สำเร็จ”

ตัวแปรสุ่ม X ดังกล่าวนี้เราเรียกว่า**ตัวแปรสุ่มเบอร์นูลลิ** (Bernulli random variable) และเรียกการทดลองเชิงสุ่มนี้ว่า**การทดลองสุ่มเบอร์นูลลิ** (Bernulli trial)

ถ้าให้ฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มเบอร์นูลลิ X กำหนดโดย

$$P(X = 1) = p,\ P(X = 0) = 1 - p$$

หรือ
$$f(x) = \begin{cases} p^x (1 - p)^{1-x} & ; x = 0, 1 \\ 0 & ; x \text{ มีค่าอื่นๆ} \end{cases}$$

เราเรียกว่าเป็น**การแจกแจงเบอร์นูลลิ** (Bernulli distribution) ซึ่งฟังก์ชันการแจกแจงคือ

$$F(x) = \begin{cases} 0 & ; x < 0 \\ p & ; 0 \le x < 1 \\ 1 & ; x \ge 1 \end{cases}$$

และมีค่าเฉลี่ย $E(X) = p$

ค่าความแปรปรวน $Var(X) = p(1 - p)$

โสตทัศน์ # 11.9 (ต่อ)

ถ้าในการทดลองเชิงสุ่มหนึ่งที่เป็นการลองสุ่มเบอร์นูลลิซ้ำๆ กัน n ครั้งและแต่ละครั้งเป็นอิสระต่อกันซึ่งมีความน่าจะเป็นของผลลัพธ์ที่สำเร็จ p ถ้าให้ X แทนจำนวนครั้งที่สำเร็จในการทดลอง n ครั้ง เราเรียก X ว่า**ตัวแปรสุ่มทวินาม** (binomial random variable) และการแจกแจงความน่าจะเป็นของ X เราเรียกว่าการแจกแจงทวินาม (binomial distribution) ซึ่งเขียนแทนด้วย b(x; n, p) โดยที่

$$b(x;\ n,\ p) \ = \ P(X = x) = f(x) = \binom{n}{x} p^x (1 - p)^{n - x}\ ,\ x = 0,\ 1,\ 2,\ \ldots,\ n$$

ในการคำนวณ b(x; n, p) เราต้องทราบค่า n และ p ซึ่งจะเรียกว่าพารามิเตอร์ของการแจกแจงทวินาม ค่าของ b(x; n, p) หาได้จากตาราง ตัวเลขในตารางแสดงค่าของ $P(X \leq x)$ สำหรับแต่ละค่า n, p และ x ที่กำหนด ซึ่ง

$$P(X \leq x) = \sum_{k=0}^{x} b(x;\ n,\ p) = \sum_{k=0}^{x} \binom{n}{x} p^k (1 - p)^{n - k}$$

ค่าเฉลี่ยและค่าแปรปรวนของการแจกแจงทวินาม

ถ้าตัวแปรสุ่ม X มีการแจกแจงทวินาม

$\mu = E(X) = np$ และ $\sigma^2 = Var(X) = np(1 - p)$

ตัวอย่าง ความน่าจะเป็นที่ผู้ป่วยโรคโลหิตจางที่ได้รับยาชนิดใหม่จะหายป่วยเท่ากับ 0.4 ถ้ามีผู้ป่วยโรคโลหิตจางได้รับยาชนิดนี้ 15 คน จงหาความน่าจะเป็นที่จะมีผู้หายป่วย

1. จำนวน 5 คน
2. อย่างน้อย 10 คน
3. อย่างน้อย 3 คนแต่ไม่เกิน 8 คน

โสตทัศน์ # 11.10 การแจกแจงไฮเพอร์จีออเมตริก

ในการทดลองสุ่ม เลือกตัวอย่างจำนวน n จากของทั้งหมดจำนวน N เราต้องการหาความน่าจะเป็นที่จะได้ผลลัพธ์ที่สนใจ (สำเร็จ) จำนวน x จากจำนวนสิ่งที่สนใจที่มีอยู่ทั้งหมด k และเลือกได้จำนวนผลลัพธ์ที่ไม่สนใจ (ไม่สำเร็จ) จำนวน n – x จากจำนวนสิ่งที่ไม่สนใจทั้งหมด N – k การทดลองสุ่มนี้เรียกว่า**การทดลองไฮเพอร์จีออเมตริก** (hypergeometric experiment) หรือกล่าวได้ว่าการทดลองไฮเพอร์จีออเมตริก มีสมบัติดังนี้

1. เลือกตัวอย่างสุ่มขนาด n โดยเลือกครั้งละตัวอย่างแต่ละครั้งไม่ใส่กลับคืนหรือสุ่มเลือกทีเดียว n ครั้งจากจำนวนของทั้งหมด N

2. ของทั้งหมดจำนวน N แบ่งเป็น 2 พวกคือสิ่งที่สนใจหรือที่เรียกว่าผลลัพธ์ที่สำเร็จจำนวน k และสิ่งที่ไม่สนใจหรือที่เรียกว่าผลลัพธ์ที่ไม่สำเร็จจำนวน N – k

ถ้าให้ X แทนจำนวนผลลัพธ์ที่สำเร็จในการทดลองไฮเพอร์จีออเมตริกเราเรียกว่า X เป็นตัวแปรสุ่มไฮเพอร์จีออเมตริก (hypergeometric random variable) และการแจกแจงความน่าจะเป็นของ X เรียกว่าการแจกแจงไฮเพอร์จีออเมตริก (hypergeometric distribution) และจะเขียนแทน P(X = x) ด้วย h(x; N, n, k)

สูตรโดยทั่วไปของ h(x; N, n, k) ซึ่งเป็นจำนวนวิธีที่หยิบของจำนวน n จากจำนวนทั้งหมด N เท่ากับ $\binom{N}{n}$

จำนวนวิธีที่หยิบได้ผลลัพธ์ที่สนใจ (สำเร็จ) จำนวน x จากจำนวนที่สนใจทั้งหมด k เท่ากับ $\binom{k}{x}$

ซึ่งแต่ละวิธีนี้จะหยิบได้ผลลัพธ์ที่ไม่สนใจ (ไม่สำเร็จ) จำนวน n – x จากจำนวนที่ไม่สนใจทั้งหมด N – k

ได้ $\binom{N-k}{n-x}$ วิธี

ดังนั้นจำนวนวิธีที่หยิบได้ของที่สนใจจำนวน x ในการหยิบจำนวน n เท่ากับ $\binom{k}{x}\binom{N-k}{n-x}$ วิธี

และเราเขียนสูตร h(x; N, n, k) ได้ดังนี้

ถ้าตัวแปรสุ่ม X มีการแจกแจงไฮเพอร์จีออเมตริกที่แทนจำนวนผลลัพธ์ที่สนใจ (สำเร็จ) ในการสุ่มตัวอย่างขนาด n จากสิ่งของทั้งหมดจำนวน N ที่แบ่งออกเป็นสิ่งที่สนใจ (สำเร็จ) จำนวน k และไม่สนใจ (ไม่สำเร็จ) จำนวน N – k การแจกแจงของ X กำหนดโดย

$$P(X = x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}; \quad x = 0, 1, 2, \ldots, n$$

โสตทัศน์ # 11.10 (ต่อ)

ทฤษฎีบท

การแจกแจงไฮเพอร์จีออเมตริกมี**ค่าเฉลี่ย**เท่ากับ $\frac{nk}{N}$

และ**ค่าแปรปรวน**เท่ากับ $\frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1-\frac{k}{N}\right)$

ตัวอย่าง หลอดไฟกล่องหนึ่งมี 10 หลอด ถ้าทราบว่าหลอดไฟในกล่องนี้มีหลอดสีแดง 4 หลอด นอกนั้นเป็นสีน้ำเงิน สุ่มหยิบหลอดไฟมา 3 หลอด จงหาความน่าจะเป็นที่

1. หลอดไฟทั้ง 3 เป็นสีเดียวกัน
2. มีหลอดไฟสีแดงอย่างมาก 2 หลอด

โสตทัศน์ # 11.11 การแจกแจงปัวส์ซง

ในการทดลองเชิงสุ่มหนึ่ง ที่มีผลลัพธ์ที่อาจเกิดขึ้นได้สองแบบคือ ผลลัพธ์ที่เราสนใจ และผลลัพธ์ที่เราไม่สนใจและเราต้องการนับจำนวนครั้งของการเกิดผลลัพธ์ที่สนใจในช่วงเวลา (อาจเป็นนาที ชั่วโมง วัน สัปดาห์ เดือนหรือปี ก็ได้) หรือในบริเวณที่กำหนด (ซึ่งอาจเป็นเส้นตรง พื้นที่ หรือปริมาตร) ตัวอย่าง เช่น จำนวนครั้งของเสียงเรียกโทรศัพท์ที่เข้ามายังสำนักงานแห่งหนึ่งตั้งแต่เวลา 9.00 – 11.00 น. จำนวนลูกค้าที่เข้ามาซื้อตั๋วดูภาพยนตร์ในระหว่างเวลา 12.00 – 14.00 น. จำนวนตำหนิที่เกิดขึ้นในการทอพรมผืนหนึ่งที่มีขนาดกว้าง 2 หลา และยาว 5 หลา จำนวนคำที่พิมพ์ผิดในหนังสือหนึ่งหน้า เป็นต้น **จำนวนผลลัพธ์ที่สนใจที่เกิดขึ้นดังกล่าวเรียกได้ว่าเป็นตัวแปรสุ่มในกระบวนการปัวส์ซง** (Poisson process) ซึ่งมีเงื่อนไขดังนี้

1. จำนวนของผลลัพธ์ที่สนใจที่เกิดขึ้นในช่วงเวลาหนึ่งที่กำหนดหรือในบริเวณหนึ่งที่เจาะจงไม่ขึ้นอยู่กับจำนวนผลลัพธ์ที่เกิดขึ้นในช่วงเวลาหรือบริเวณอื่นๆ
2. ความน่าจะเป็นที่ผลลัพธ์ที่สนใจหนึ่งเกิดขึ้นในช่วงเวลาสั้นๆ หรือบริเวณที่เล็กๆ เป็นสัดส่วนกับความยาวของช่วงเวลาหรือขนาดของบริเวณที่กำหนด และไม่ขึ้นอยู่กับเวลาหรือบริเวณที่ผลลัพธ์ที่สนใจนั้นจะเกิดขึ้น
3. ความน่าจะเป็นที่จะมีผลลัพธ์ที่สนใจเกิดขึ้นมากกว่า 1 ครั้งในช่วงเวลาสั้นมากๆหรือในบริเวณที่เล็กมากๆเป็นศูนย์

ถ้า X เป็นตัวแปรสุ่มในกระบวนการปัวส์ซงเราเรียก X ว่า**ตัวแปรสุ่มปัวส์ซง** และการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X นี้เรียกว่า**การแจกแจงปัวส์ซง** ถ้าให้ λ แทนจำนวนเฉลี่ยของผลลัพธ์ที่สนใจที่เกิดขึ้นในหนึ่งหน่วยเวลาหรือในหนึ่งหน่วยของบริเวณที่กำหนด และถ้า t แทนช่วงเวลาหรือบริเวณที่กำหนด ค่าเฉลี่ย μ ของจำนวนผลลัพธ์ที่สนใจที่เกิดขึ้นในช่วงเวลาหรือบริเวณของจำนวนผลลัพธ์ที่สนใจที่เกิดขึ้นในช่วงเวลาหรือบริเวณที่กำหนดคือ $\mu = \lambda t$

บทนิยาม

ถ้า X แทนจำนวนของผลลัพธ์ที่สนใจที่เกิดขึ้นในช่วงเวลาหนึ่งที่กำหนดหรือในบริเวณหนึ่งที่เจาะจง ซึ่งเป็นไปตามกระบวนการปัวส์ซง X คือ ตัวแปรสุ่มที่มีการแจกแจงปัวส์ซง และมีฟังก์ชันความน่าจะเป็นคือ

$$f(x;\mu) = P(X = x) = \frac{\mu^x e^{-\mu}}{x!} \quad ; \quad x = 0, 1, 2, 3, \ldots$$

โดยที่ e = 2.71828 . . . และ μ **คือจำนวนผลลัพธ์ที่สนใจที่เกิดขึ้นโดยเฉลี่ยในช่วงเวลาหรือบริเวณที่กำหนด**

ทฤษฎีบท

ค่าเฉลี่ยและค่าแปรปรวนของการแจกแจงปัวส์ซง $f(x;\mu)$ เท่ากันคือ μ

ตัวอย่าง ถ้ามีสัญญาณเรียกโทรศัพท์เข้ามายังสำนักงานแห่งหนึ่งโดยเฉลี่ย 2 ครั้งในทุก ๆ 3 วินาที ถ้าจำนวนสัญญาณเรียกโทรศัพท์มีการแจกแจงปัวส์ซง จงหาความน่าจะเป็นที่จะมีสัญญาณเรียกโทรศัพท์เข้ามายังสำนักงานแห่งนี้ใน 9 วินาทีเป็นจำนวน

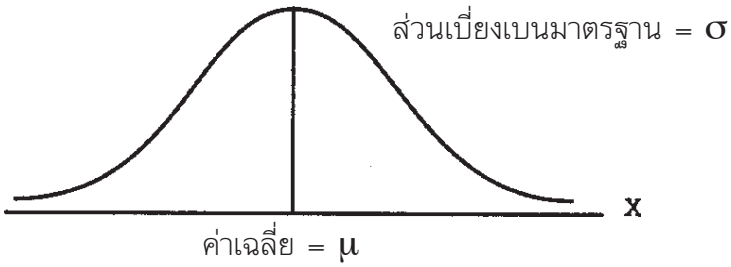
1. 3 ครั้ง
2. มากกว่า 5 ครั้ง

โสตทัศน์ # 11.12 การแจกแจงปกติ

ตัวแปรสุ่มแบบต่อเนื่อง X ที่มีการแจกแจงความน่าจะเป็นแบบปกติเรียกว่า**ตัวแปรสุ่มปกติ** และ**ฟังก์ชันการแจกแจงความน่าจะเป็นปกติ** (Normal probability distribution function) คือ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

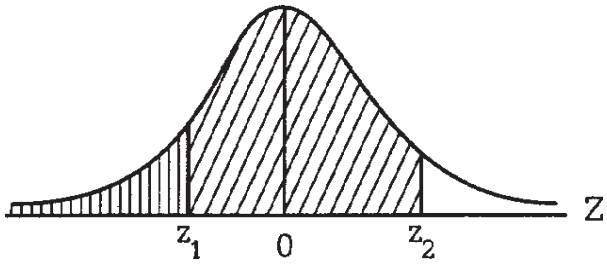
โดยที่ μ **คือค่าเฉลี่ย** และ σ **คือส่วนเบี่ยงเบนมาตรฐานของการแจกแจงปกติ** มีกราฟที่เรียกว่าโค้งปกติ คือมีลักษณะเส้นโค้งเป็นรูประฆังคว่ำ (a bell–shaped curve) ดังภาพ



บทนิยาม

การแจกแจงปกติที่มี $\mu = 0$ และ $\sigma = 1$ เรียกว่าการแจกแจงปกติมาตรฐานซึ่งมีฟังก์ชันการแจกแจงความน่าจะเป็นคือ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

ในการหา $P(z_1 < Z < z_2)$ เราสามารถใช้ตารางพื้นที่ใต้โค้งปกติมาตรฐาน ดังแสดงตัวอย่างโดยพื้นที่แรเงาดังรูป



จากรูป จะสังเกตได้ว่า $P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$

ทฤษฎีบท

ถ้า X เป็นตัวแปรสุ่มทวินามที่มีค่าเฉลี่ย $\mu = np$ และค่าแปรปรวน $\sigma^2 = np(1-p)$ แล้ว ตัวแปรสุ่ม $Z = \frac{X - np}{\sqrt{np(1-p)}}$ จะมีการแจกแจงเข้าใกล้การแจกแจงปกติมาตรฐานเมื่อ n มีค่ามากๆ $(n \to \infty)$

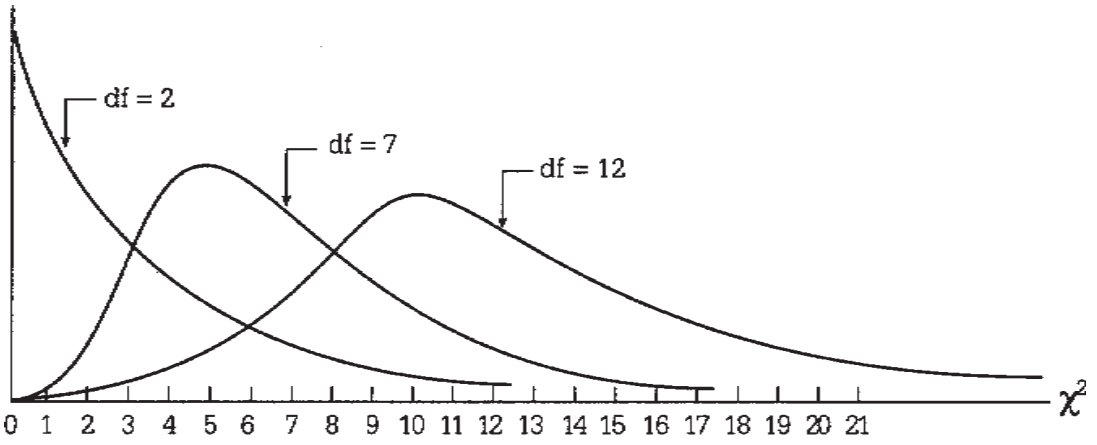
โสตทัศน์ # 11.12 (ต่อ)

ตัวอย่าง สมมติว่าปริมาณการใช้ไฟฟ้าของครัวเรือนในท้องที่แห่งหนึ่งในแต่ละเดือนมีการแจกแจงปกติที่มีค่าเฉลี่ย 1,650 กิโลวัตต์ชั่วโมงและส่วนเบี่ยงเบนมาตรฐาน 320 กิโลวัตต์ชั่วโมง

1. จงหาความน่าจะเป็นที่ ถ้าในเดือนหนึ่งเลือกบ้านหนึ่งในท้องที่แห่งนี้มาอย่างสุ่มแล้ว บ้านหลังนี้ใช้ไฟฟ้าปริมาณน้อยกว่า 2,050 กิโลวัตต์ชั่วโมง
2. จงหาว่ามีจำนวนครัวเรือนกี่เปอร์เซนต์ที่ใช้ไฟฟ้าปริมาณอยู่ระหว่าง 930 ถึง 1,346 กิโลวัตต์ชั่วโมงต่อเดือน

โสตทัศน์ 11.13 การแจกแจงไคสแควร์

การแจกแจงไคสแควร์ เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่องที่มีค่ามากกว่า 0 กราฟของฟังก์ชันการแจกแจงจะมีลักษณะแตกต่างกันขึ้นอยู่กับค่าพารามิเตอร์ตัวหนึ่งที่เราเรียกว่า จำนวนองศาแห่งความเป็นอิสระ (the number of degrees of freedom) โค้งของการแจกแจงไคสแควร์จะมีลักษณะเบ้ทางขวา ซึ่งจะเบ้มากหรือน้อยขึ้นอยู่กับจำนวนองศาแห่งความเป็นอิสระ ตัวอย่างดังภาพ

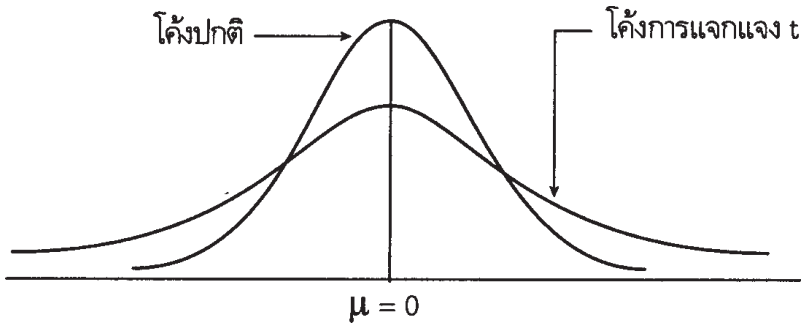


ตารางแสดงค่าของตัวแปรสุ่มที่มีการแจกแจงไคสแควร์หรือเรียกว่าค่าไคสแควร์สำหรับบางค่าของจำนวนองศาแห่งความเป็นอิสระและพื้นที่ใต้เส้นโค้ง สดมภ์แรกของตารางแสดงจำนวนองศาแห่งความเป็นอิสระและตัวเลขแถวแรกแสดงพื้นที่ใต้เส้นโค้งทางด้านซ้ายและตัวเลขแถวที่สองแสดงพื้นที่ใต้เส้นโค้งทางด้านขวาซึ่งสมนัยกันสำหรับตัวเลขที่อยู่ในสดมภ์เดียวกันเพราะพื้นที่ใต้เส้นโค้งทั้งหมดเท่ากับ 1 ส่วนตัวเลขในแต่ละแถวถัดจากแถวที่สองลงไปแสดงค่าของตัวแปรสุ่มที่มีการแจกแจงไคสแควร์ เช่น

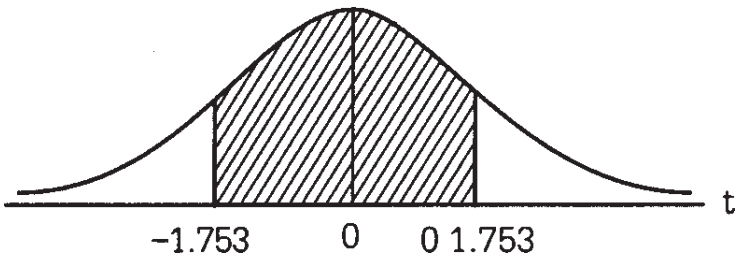
ตัวอย่าง กำหนดพื้นที่ใต้โค้งไคสแควร์ทางด้านขวาเท่ากับ 0.025 จงหาค่าไคสแควร์ที่สอดคล้องกันที่องศาแห่งความเป็นอิสระ 20 จากตาราง

โสตทัศน์ # 11.14 การแจกแจง t

การแจกแจง t หรือเรียกว่าการแจกแจง student's t เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่องที่มีโค้งของการแจกแจงคล้ายกับการแจกแจงปกติมากคือโค้งเป็นรูประฆังคว่ำซึ่งสมมาตร และมีค่าเฉลี่ยเป็น 0 แต่โค้งมีความโด่งและความชันน้อยกว่าโค้งปกติ นั่นคือ มีส่วนเบี่ยงเบนมาตรฐานมากกว่าการแจกแจงปกติ ซึ่งลักษณะของโค้งจะแตกต่างกันถ้ากำหนดพารามิเตอร์ของฟังก์ชันการแจกแจงแตกต่างกัน เราเรียกค่าพารามิเตอร์นี้ว่า **จำนวนองศาแห่งความเป็นอิสระ** สำหรับ**การแจกแจง t มีจำนวนองศาแห่งความเป็นอิสระเท่ากับจำนวนของขนาดตัวอย่างลบด้วย 1** เพราะเสียจำนวนแห่งองศาอิสระไป 1 ในการคำนวณค่าเฉลี่ย เมื่อจำนวนองศาแห่งความเป็นอิสระมีค่ามากๆ จนใกล้อนันต์ โค้งการแจกแจง t จะใกล้เคียงกับโค้งปกติมาก ดังภาพ



พื้นที่ใต้โค้งการแจกแจง t ทั้งหมดคิดเป็น 1 หรือ 100% ขึ้นอยู่กับจำนวนองศาแห่งความเป็นอิสระ ซึ่งตารางการแจกแจง t จะแสดงค่า t ที่สมนัยกับพื้นที่บางค่า เช่น ค่า t = 1.753 คือค่า t ที่ทำให้ P(– t < T < t) = 0.95 ก็คือค่า t ที่ทำให้ P(T < t) = 0.95 ดังภาพ

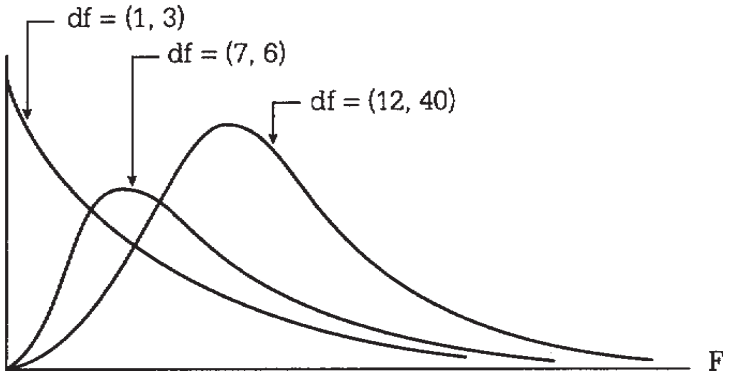


โสตทัศน์ # 11.14 การแจกแจง F

การแจกแจง F เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่องที่สำคัญแบบหนึ่ง โดยที่**ตัวแปรสุ่ม F เป็นฟังก์ชัน**

ที่อยู่ในรูปอัตราส่วนของตัวแปรสุ่มไคสแควร์คือ $F = \frac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}}$ เมื่อ U และ V คือตัวแปรสุ่มไคสแควร์ที่มีองศา

แห่งความเป็นอิสระ ν_1 และ ν_2 ตามลำดับ การแจกแจงความน่าจะเป็น F จึงขึ้นอยู่กับจำนวนองศาแห่งความเป็นอิสระ 2 จำนวนซึ่งเรียกว่าเป็นพารามิเตอร์ของฟังก์ชันการแจกแจง เช่นเดียวกับการแจกแจงไคสแควร์และการแจกแจง t เส้นโค้งของการแจกแจง F แตกต่างกันที่องศาแห่งความเป็นอิสระ ตัวอย่างดังภาพ



กิจกรรม

1. ในการศึกษาเกี่ยวกับความนิยมของผู้ชมรายการทีวีในท้องที่แห่งหนึ่งว่าชอบชมรายการละครหลังข่าวหรือไม่ ถ้าสุ่มผู้ชมรายการทีวีในท้องที่แห่งนี้มา 10 คน เพื่อสัมภาษณ์
   1.1 จงกำหนดตัวแปรสุ่ม
   1.2 ค่าที่เป็นไปได้ของตัวแปรสุ่มใน 1.1 คือค่าใดบ้าง

2. ตัวแปรสุ่มที่กำหนดให้ต่อไปนี้เป็นตัวแปรสุ่มแบบไม่ต่อเนื่องหรือเป็นตัวแปรสุ่มแบบต่อเนื่อง
   2.1 เวลาที่นักวิ่งมาราธอนใช้ในการแข่งขันครั้งหนึ่ง
   2.2 ราคาตั๋วเข้าชมภาพยนตร์
   2.3 จำนวนหน้าในหนังสือเล่มหนึ่งที่มีการพิมพ์ผิด
   2.4 อายุของบ้าน
   2.5 จำนวนไข่ที่เสียในตะกร้าใบหนึ่ง

โสตทัศน์ # 11.14 (ต่อ)

3. ร้านประกอบเครื่องคอมพิวเตอร์แห่งหนึ่งประกอบเครื่องคอมพิวเตอร์ได้ 20 เครื่อง ถ้าทราบว่ามีเครื่องที่ประกอบเสร็จแล้วแต่ยังใช้งานไม่ได้ 6 เครื่อง ถ้าผู้ตรวจสอบคุณภาพสุ่มเลือกเครื่องคอมพิวเตอร์ที่ประกอบได้มา 5 เครื่อง ให้ X แทนจำนวนเครื่องคอมพิวเตอร์ที่ประกอบเสร็จแล้วแต่ยังใช้งานไม่ได้ในจำนวน 5 เครื่องที่สุ่มมา จงเขียนฟังก์ชันความน่าจะเป็นของ X
4. จำนวนคำที่พิมพ์ผิดในหน้าหนึ่งที่สุ่มมาของหนังสือเล่มหนึ่งเป็น 0, 1, 2, 3, และ 4 คำด้วย ความน่าจะเป็น 0.86, 0.05, 0.04, 0.03, และ 0.02 ตามลำดับ จงหาจำนวนคำที่พิมพ์ผิดโดยเฉลี่ยในหน้าหนึ่งของหนังสือเล่มนี้

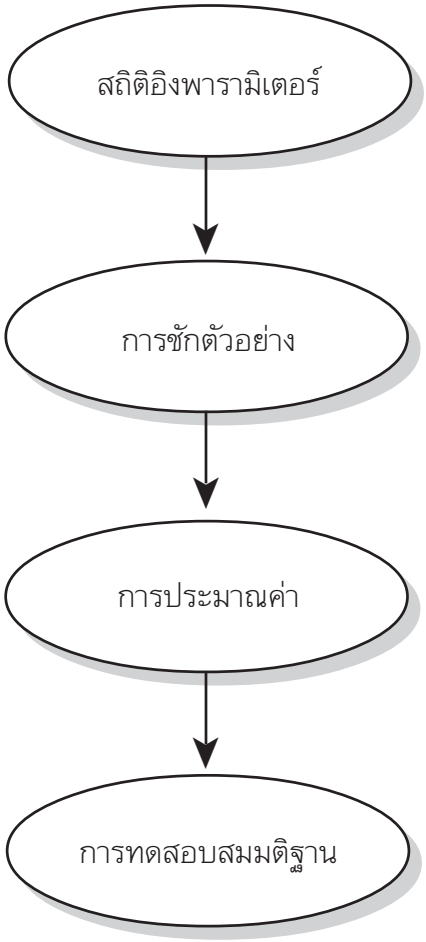
5. ถ้า 20% ของต้นไม้ที่ปลูกใหม่ในพื้นที่แห่งหนึ่งตาย จงหาความน่าจะเป็นที่ต้นไม้ที่ปลูกใหม่ 10 ต้นตายไปไม่เกิน 2 ต้น

6. มีสีบรรจุกระป๋อง 10 กระป๋องโดยที่ไม่มีฉลากบอกสี แต่ทราบว่ามีสีแดงอยู่ 4 กระป๋องและมีสีดำอยู่ 6 กระป๋อง จงหาความน่าจะเป็นที่ในการสุ่มหยิบสีมา 4 กระป๋อง
   6.1 ทุกกระป๋องเป็นสีแดง
   6.2 มีสีดำมากกว่า 1 กระป๋อง

7. ในการทอผ้าของโรงงานทอผ้าแห่งหนึ่ง จะเกิดตำหนิ 1 แห่งในพื้นที่ 150 ตารางเมตรถ้าจำนวนของการเกิดตำหนิในการทอผ้ามีการแจกแจงปัวส์ซง จงหาความน่าจะเป็นที่จะเกิดตำหนิในการทอผ้าอย่างมากแห่งเดียวบนผ้า 225 ตารางเมตร

8. ถ้าความเร็วของรถยนตร์ที่วิ่งบนถนนใหญ่ถนนหนึ่งมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 120 กิโลเมตรต่อชั่วโมงและส่วนเบี่ยงเบนมาตรฐาน 10 กิโลเมตรต่อชั่วโมง จงหาว่ามีจำนวนรถยนตร์กี่เปอร์เซนต์ที่วิ่งบนถนนนี้ด้วยความเร็ว
   8.1 95 ถึง 110 กิโลเมตรต่อชั่วโมง 8.2 มากกว่า 140 กิโลเมตรต่อชั่วโมง

หน่วยที่ 12
สถิติศาสตร์อิงพารามิเตอร์เบื้องต้น



โสตทัศน์ # 12.1 สถิติศาสตร์อิงพารามิเตอร์

ข้อตกลงเบื้องต้นที่สำคัญ

1. ข้อมูลของตัวอย่างที่ใช้ในการศึกษาชักจากประชากรที่มีรูปแบบการแจกแจงที่ชัดเจนอาจเป็นการแจกแจงแบบปกติ (normal distribution) หรือการแจกแจงแบบอื่นๆ
2. ตัวอย่างที่ใช้ในการศึกษามีขนาดเหมาะสมไม่เล็กมากจนเกินไป ซึ่งการกำหนดขนาดตัวอย่างที่เหมาะสมในการศึกษามีหลายวิธี
3. ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลเชิงปริมาณได้จากการวัดในมาตราอันตรภาคหรือมาตราอัตราส่วน

ขอบข่ายเนื้อหา

1. การชักตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น
2. การประมาณค่า (estimation)
3. การทดสอบสมมติฐาน (hypothesis testing)

โสตทัศน์ # 12.2 การชักตัวอย่าง

การชักตัวอย่าง (sampling) หมายถึง กระบวนการจัดกระทำให้ได้มาซึ่งตัวอย่าง (sample) ที่เป็นตัวแทนของประชากรเพื่อศึกษาข้อมูลจากตัวอย่างแล้วสรุปอ้างอิงไปยังลักษณะของประชากรได้อย่างน่าเชื่อถือ

การชักตัวอย่างแบ่งเป็น 2 ประเภท

1. การชักตัวอย่างโดยไม่อาศัยทฤษฎีความน่าจะเป็น (non– probability sampling)
2. การชักตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น (probability sampling)
    1. การสุ่มตัวอย่างแบบง่าย (simple random sampling)
    2. การสุ่มตัวอย่างแบบเป็นระบบ (systematic sampling)
    3. การสุ่มแบบแบ่งชั้น (stratified sampling)
    4. การสุ่มแบบกลุ่ม (cluster sampling)

โสตทัศน์ # 12.3 การแจกแจงของค่าเฉลี่ยตัวอย่าง

การแจกแจงของค่าเฉลี่ยตัวอย่าง (sampling distribution of mean)

1. กรณีการสุ่มแบบใส่คืน

ค่าเฉลี่ยของค่าเฉลี่ยตัวอย่าง $\mu_{\bar{x}} = \mu$

ความแปรปรวนของค่าเฉลี่ยตัวอย่าง $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

2. กรณีการสุ่มแบบไม่ใส่คืน

ค่าเฉลี่ยของค่าเฉลี่ยตัวอย่าง $\mu_{\bar{x}} = \mu$

ความแปรปรวนของค่าเฉลี่ยตัวอย่าง $\sigma^2 = \frac{\sigma^2}{n} \cdot \left(\frac{N-n}{N-1}\right)$

กรณีที่ N มีค่ามากและ n มีค่าน้อย $\frac{N-n}{N-1}$ จะมีค่าใกล้ 1

ในที่นี้พิจารณาการแจกแจงของค่าเฉลี่ยตัวอย่าง 2 กรณีคือ

กรณีที่ 1 การแจกแจงของค่าเฉลี่ยตัวอย่างเมื่อ ประชากรมีการแจกแจงแบบปกติ **ไม่**ทราบค่าความแปรปรวนของประชากร และกลุ่มตัวอย่างมีขนาดใหญ่

กรณีที่ 2 การแจกแจงของค่าเฉลี่ยตัวอย่างเมื่อ ประชากร**ไม่ได้**มีการแจกแจงแบบปกติ และตัวอย่างมีขนาดใหญ่

ความรู้เกี่ยวกับการแจกแจงของค่าเฉลี่ยตัวอย่างนำไปประยุกต์หาค่าความน่าจะเป็นในกรณีต่างๆ ได้ดังตัวอย่าง 12.1.3(2) และ 12.1.3(3) ในเอกสารการสอน

โสตทัศน์ # 12.4 การแจกแจงค่าสัดส่วนของตัวอย่าง

การแจกแจงค่าสัดส่วนของตัวอย่าง (sampling distribution of sample proportions)

เมื่อ $\mu_{\hat{p}}$ แทน ค่าเฉลี่ยของค่าสัดส่วนตัวอย่าง

$\sigma^2_{\hat{p}}$ แทน ค่าความแปรปรวนของค่าสัดส่วนตัวอย่าง

อาศัยทฤษฎีขีดจำกัดกลางจะได้ว่าถ้าตัวอย่างมีขนาดใหญ่พอ (เมื่อ $np \geq 5$ และ $nq \geq 5$) การแจกแจงของค่าสัดส่วนตัวอย่างที่มีลักษณะตามที่ต้องการจะมีลักษณะ**ใกล้เคียง**กับการแจกแจงแบบปกติ โดยที่

1. $\mu_{\hat{P}} = p$

2. $\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$ เมื่อ $\frac{n}{N} < 0.5$

ดังนั้น

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

มีการแจกแจงใกล้เคียงการแจกแจงแบบปกติมาตรฐาน

ความรู้เกี่ยวกับการแจกแจงของค่าสัดส่วนตัวอย่างนำไปประยุกต์ใช้ในการหาความน่าจะเป็นได้ดังตัวอย่างที่ 12.1.3(4) ในเอกสารการสอน

โสตทัศน์ # 12.5 การประมาณค่า

การประมาณค่า (estimation) หมายถึงวิธีการของสถิติศาสตร์อิงพารามิเตอร์ที่อาศัยค่าสถิติในการประมาณค่าพารามิเตอร์

ตัวประมาณค่า (estimator) หมายถึง ตัวสถิติที่ใช้ประมาณพารามิเตอร์

ค่าประมาณ (estimate) หมายถึง ค่าสถิติที่ใช้ในการประมาณค่าพารามิเตอร์

วิธีการประมาณค่าพารามิเตอร์มี 2 วิธีคือ การประมาณค่าแบบจุด (point estimation) และ การประมาณค่าแบบช่วง (interval estimation)

โสตทัศน์ # 12.5 (ต่อ)

สมบัติของตัวประมาณค่าที่ดี

1. ไม่ลำเอียง (unbiased)
2. มีประสิทธิภาพ (efficient)
3. มีความคงเส้นคงวา(consistent)

ระดับความเชื่อมั่น $(1-\alpha)$ หมายถึง ความน่าจะเป็นที่ค่าพารามิเตอร์ (θ) ที่ต้องการประมาณค่าจะตกอยู่ในช่วงจำนวนจริงที่คำนวณได้จากค่าสถิติ

โดยทั่วไปกำหนดระดับความเชื่อมั่น 3 ระดับคือ .90, .95 และ .99

เมื่อสุ่มตัวอย่างขนาดเท่ากันจากประชากร และสร้างช่วงของจำนวนจริงเพื่อประมาณค่าพารามิเตอร์ ดังนั้น ระดับความเชื่อมั่น $(1-\alpha) = .90$ หมายความว่ามีความน่าจะเป็น 90% ที่ช่วงของจำนวนจริงที่คำนวณได้จากค่าสถิติจะคลุมค่าพารามิเตอร์ที่ต้องการไว้ นั่นคือจำนวนช่วงที่สร้างขึ้น 90% จะคลุมค่าพารามิเตอร์ และ 10% ของช่วงที่สร้างขึ้นไม่คลุมค่าพารามิเตอร์

โสตทัศน์ # 12.6 การประมาณค่าเฉลี่ยของประชากรชุดเดียวเมื่อไม่ทราบค่าความแปรปรวนของประชากรและตัวอย่างมีขนาดเล็ก

กรณีที่สุ่มตัวอย่างขนาดเล็ก $\left(n<30\right)$ จากประชากรที่มีการแจกแจงแบบปกติหรือใกล้เคียงการแจกแจงแบบปกติ จะประมาณค่า σ^2 ด้วย s^2

ดังนั้นเมื่อกำหนดระดับความเชื่อมั่น $(1-\alpha) \times 100\%$ ช่วงการประมาณค่าของ μ คือ

$$\bar{x} - \left(t_{\frac{\alpha}{2}}\right)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + \left(t_{\frac{\alpha}{2}}\right)\frac{s}{\sqrt{n}}$$

เมื่อ $\bar{x}$ = ค่าเฉลี่ยของตัวอย่าง

S = ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

n = ขนาดของตัวอย่าง

การประมาณค่าเฉลี่ยกรณีนี้ศึกษาได้จากตัวอย่าง 12.2.2 (1) ในเอกสารการสอน

โสตทัศน์ # 12.7 การประมาณค่าเฉลี่ยเมื่อไม่ทราบค่าความแปรปรวนของประชากรและตัวอย่างมีขนาดใหญ่

กรณีที่สุ่มตัวอย่างขนาดใหญ่ $\left(n \geq 30\right)$ จากประชากรที่มีการแจกแจงแบบใด ๆ ก็ตาม การแจกแจงของค่าเฉลี่ยตัวอย่างจะมีการแจกแจงเป็นแบบปกติด้วยค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ $\frac{\sigma^2}{n}$ ในกรณีที่ไม่ทราบค่าความแปรปรวนของประชากรจะประมาณค่า σ^2 ด้วย s^2

ดังนั้นเมื่อกำหนดระดับความเชื่อมั่น $(1 - \alpha) \times 100\%$ ช่วงการประมาณค่าของ μ คือ

$$\bar{x} - \left(z_{\frac{\alpha}{2}}\right)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + \left(z_{\frac{\alpha}{2}}\right)\frac{s}{\sqrt{n}}$$

เมื่อ $\bar{x}$ = ค่าเฉลี่ยของตัวอย่าง
S = ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
n = ขนาดของตัวอย่าง

การประมาณค่าเฉลี่ยกรณีนี้ศึกษาได้จากตัวอย่าง 12.2.2 (2) ในเอกสารการสอน

โสตทัศน์ # 12.8 การประมาณค่าสัดส่วน

ในการประมาณค่าสัดส่วนของประชากรมักใช้ข้อมูลจากตัวอย่างที่มีขนาดใหญ่ กรณีที่ตัวอย่างมีขนาดใหญ่ การแจกแจงของค่าสัดส่วนตัวอย่างจะมีลักษณะใกล้เคียงกับการแจกแจงแบบปกติ

ดังนั้น เมื่อกำหนดระดับความเชื่อมั่น $(1 - \alpha) \times 100\%$ ช่วงการประมาณค่าของ P คือ

$$\hat{P} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq p + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

เมื่อ $\hat{P}$ = ค่าประมาณแบบจุดของ P
= $\frac{x}{n}$
q = $1 - \hat{p}$
n = ขนาดของตัวอย่าง

การประมาณค่าสัดส่วนศึกษาได้จากตัวอย่าง 12.2.3 (1) และ12.2.3(2) ในเอกสารการสอน

โสตทัศน์ # 12.9 ความรู้พื้นฐานเกี่ยวกับการทดสอบสมมติฐาน

สมมติฐาน (hypothesis) หมายถึง ข้อสมมติที่เสนอขึ้นสำหรับอธิบายเรื่องที่สนใจ อาจเป็นจริงหรือไม่เป็นจริงก็ได้ โดยข้อสมมติที่เสนอต้องสมเหตุสมผล โดยอาศัยประสบการณ์ ทฤษฎี และผลจากการศึกษาค้นคว้าในเรื่องที่เกี่ยวข้องที่มีผู้ศึกษาไว้แล้ว ข้อสมมติดังกล่าวใช้เป็นแนวทางในการค้นหาข้อมูลเพื่อทำการทดสอบต่อไป

สมมติฐานทางสถิติ (statistical hypothesis) หมายถึงข้อสมมติเกี่ยวกับค่าพารามิเตอร์ซึ่งอาจเป็นจริงหรือไม่เป็นจริงก็ได้

การทดสอบสมมติฐานทางสถิติ (Hypothesis Testing หรือ Significant Testing) หรือเรียกสั้นๆ ว่าการทดสอบสมมติฐาน หมายถึง วิธีการสังเคราะห์ข้อมูลที่เก็บรวบรวมจากตัวอย่างเพื่อทดสอบและหาข้อสรุปเกี่ยวกับพารามิเตอร์โดยอาศัยการตัดสินใจเชิงสถิติ

สมมติฐานทางสถิติแบ่งเป็น 2 ประเภทคือ

1. สมมติฐานว่าง (Null Hypothesis) ใช้สัญลักษณ์ H_0
2. สมมติฐานทางเลือก (Alternative Hypothesis) ใช้สัญลักษณ์ H_1

ความคลาดเคลื่อนในการทดสอบสมมติฐานทางสถิติแบ่งได้ 2 ประเภท คือ

1. ความคลาดเคลื่อนประเภทที่ 1 (Type I error) คือ ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติว่าง (H_0) ทั้ง ๆ ที่สมมติฐานว่างเป็นจริง
2. ความคลาดเคลื่อนประเภทที่ 2 (Type II error) คือ ความคลาดเคลื่อนที่เกิดจากการยอมรับสมมติฐานว่าง (H_0) ทั้งที่สมมติฐานว่างเป็นเท็จ

การทดสอบสมมติฐาน แบ่งเป็น 2 ประเภทคือ

1. การทดสอบสมมติฐานทางเดียว (one – tailed test)
2. การทดสอบสมมติฐานสองทาง (two – tailed test)

โสตทัศน์ # 12 .10 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย

1. กรณีเป็นการทดสอบทางเดียว

$H_0 : \mu = \mu_0$ หรือ $H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$ $H_1 : \mu > \mu_0$

2. กรณีเป็นการทดสอบสองทาง

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

โดยที่ μ_0 เป็นค่าคงที่ที่สมมติหรือคาดคะเนไว้ล่วงหน้า

โสตทัศน์ # 12 .10 (ต่อ)

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรชุดเดียวเมื่อ**ไม่ทราบค่าความแปรปรวนของประชากร** และ**ตัวอย่างมีขนาดเล็ก**

สถิติทดสอบที่ใช้ในการทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ยของประชากรกรณีนี้คือ

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}},\ df = n-1$$

เมื่อ $\bar{x}$ = ค่าเฉลี่ยของตัวอย่างขนาด n
μ_0 = ค่าเฉลี่ยของประชากร
s = ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
n = ขนาดของตัวอย่าง

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากรกรณีนี้ศึกษาได้จากตัวอย่าง 12.3.2(1) ในเอกสารการสอน

โสตทัศน์ # 12.11 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรชุดเดียวเมื่อไม่ทราบค่าความแปรปรวนของประชากร และตัวอย่างมีขนาดใหญ่

สถิติทดสอบที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรกรณีนี้คือ

$$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

เมื่อ $\bar{x}$ = ค่าเฉลี่ยของตัวอย่างขนาด n
μ_0 = ค่าเฉลี่ยประชากร
s = ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
n = ขนาดของตัวอย่าง

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากรกรณีนี้ศึกษาได้จากตัวอย่าง12.3.2(2) ในเอกสารการสอน

โสตทัศน์ # 12.12 การทดสอบค่าสัดส่วนของประชากรชุดเดียว

กำหนดสมมติฐานทางสถิติดังนี้

1. กรณีการทดสอบสมมติฐานทางเดียว

$H_0 : p = p_0$ หรือ $H_0 : p = p_0$

$H_1 : p < p_0$ $H_1 : p > p_0$

2. กรณีการทดสอบสมมติฐานสองทาง

$H_0 : p = p_0$

$H_1 : p \neq p_0$

เมื่อ p_0 เป็นค่าคงที่ที่สมมติหรือคาดคะเนไว้ล่วงหน้า

ดังนั้น สถิติที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าสัดส่วนของประชากรชุดเดียว คือ

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

เมื่อ p_0 คือ ค่าสัดส่วนของประชากรที่สมมติหรือคาดคะเนไว้ล่วงหน้า

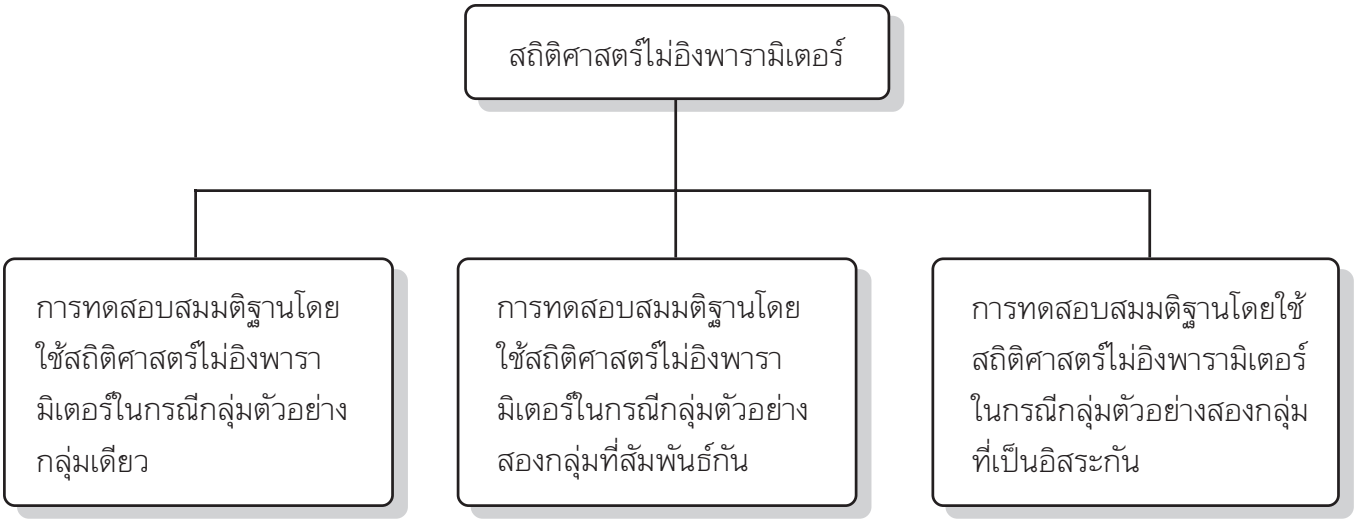
$\hat{p}$ คือ ค่าสัดส่วนของตัวอย่าง

$q_0 = 1 - p_0$

n คือ ขนาดของตัวอย่าง

การประมาณค่าสัดส่วนขอประชากรกรณีนี้ศึกษาได้จากตัวอย่าง12.3.3(1)และ12.3.3(2) ในเอกสารการสอน

หน่วยที่ 13
สถิติศาสตร์ไม่อิงพารามิเตอร์เบื้องต้น



โสตทัศน์ # 13.1 ความแตกต่างระหว่างสถิติศาสตร์ไม่อิงพารามิเตอร์ กับ สถิติศาสตร์อิงพารามิเตอร์

สถิติศาสตร์ไม่อิงพารามิเตอร์ เป็นสถิติอนุมานเช่นเดียวกับสถิติศาสตร์อิงพารามิเตอร์ แต่แตกต่างกัน ดังนี้

1. มีข้อตกลงเบื้องต้นน้อยกว่าสถิติศาสตร์อิงพารามิเตอร์ เช่นไม่มีข้อกำหนดเกี่ยวกับการแจกแจงของประชากร
2. ใช้ทดสอบข้อมูลที่วัดในมาตรานามบัญญัติ หรือเรียงลำดับได้
3. ใช้ได้กับกลุ่มตัวอย่างขนาดเล็ก
4. ใช้วิธีการคำนวณง่ายๆ เช่น การนับความถี่ การนับเครื่องหมาย การพิจารณาลำดับที่

โสตทัศน์ # 13.2 ขั้นตอนการทดสอบสมมติฐานด้วยสถิติศาสตร์ไม่อิงพารามิเตอร์

ขั้นที่ 1 กำหนดสมมติฐานทางสถิติ ซึ่งประกอบด้วยสมมติฐานว่าง (H_0) และสมมติฐานทางเลือก (H_1)

ขั้นที่ 2 เลือกสถิติทดสอบให้เหมาะสม โดยพิจารณาจากจุดมุ่งหมายของการทดสอบ มาตราการวัดของข้อมูล จำนวนลักษณะและขนาดของกลุ่มตัวอย่าง

ขั้นที่ 3 กำหนดระดับนัยสำคัญของการทดสอบ ซึ่งนิยมกำหนดให้เท่ากับ 0.05 และ 0.01 หาค่าวิกฤตและกำหนดบริเวณวิกฤต ซึ่งได้จากการเปิดตารางค่าวิกฤตของสถิติทดสอบแต่ละตัว

ขั้นที่ 4 คำนวณค่าสถิติทดสอบจากข้อมูลของกลุ่มตัวอย่าง

ขั้นที่ 5 สรุปผลการทดสอบทางสถิติ โดยการนำค่าสถิติทดสอบจากขั้นที่ 4 ไปเปรียบเทียบกับค่าวิกฤตในขั้นที่ 3 ถ้าค่าสถิติทดสอบอยู่ในบริเวณวิกฤตก็จะปฏิเสธสมมติฐานว่าง (H_0) แต่ถ้าค่าสถิติทดสอบไม่อยู่ในบริเวณวิกฤตก็จะยอมรับสมมติฐานว่าง (H_0)

โสตทัศน์ # 13.3 การทดสอบสมมติฐานโดยใช้สถิติศาสตร์ไม่อิงพารามิเตอร์ในกรณีกลุ่มตัวอย่างกลุ่มเดียว

1. การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างกลุ่มเดียว
2. การทดสอบการสุ่ม

โสตทัศน์ # 13.4 การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างกลุ่มเดียว

การทดสอบไคสแควร์ เป็นสถิติทดสอบที่ใช้ทดสอบเกี่ยวกับการแจกแจงความถี่ของตัวแปรที่สนใจว่ามีการแจกแจงตามความถี่ที่คาดหวังหรือไม่

ข้อมูลที่จะทำการทดสอบสมมติฐาน ต้องมีลักษณะ ดังนี้

1. ข้อมูลวัดในมาตรานามบัญญัติ หรือสูงกว่า
2. ตัวแปรที่ศึกษาแบ่งได้เป็น k ประเภท
3. สัดส่วนของสมาชิกในแต่ละประเภทเท่ากับ $p_1, p_2, p_3, ..., p_k$ โดยที่ $p_1 + p_2 + p_3 + ... + p_k = 1$
4. สมาชิกตัวหนึ่งจะอยู่ในหนึ่งประเภทเท่านั้น

การตั้งสมมุติฐาน

H_0 : ความถี่ที่สังเกตได้กับความถี่ที่คาดหวังไม่แตกต่างกัน

H_1 : มีอย่างน้อยหนึ่งกลุ่มที่มีความถี่ที่สังเกตได้กับความถี่ที่คาดหวังแตกต่างกัน

สูตรที่ใช้คำนวณ

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E} \quad df = k-1$$

เมื่อ χ^2 คือสถิติทดสอบไคสแควร์

O_i คือความถี่ที่สังเกตได้ในประเภทที่ i

E_i คือความถี่ที่คาดหวังในประเภทที่ i

k คือจำนวนประเภทของตัวแปร

การปฏิเสธสมมุติฐานว่าง

การปฏิเสธ สมมติฐานว่าง (H_o) กระทำได้เมื่อค่า χ^2 ที่คำนวณได้อยู่ในบริเวณวิกฤต

ตัวอย่าง จากการสุ่มสำรวจนักศึกษาสาขาวิชาวิทยาศาสตร์และเทคโนโลยีจำนวน 120 คน พบว่ามีนักศึกษาที่สมัครโดยใช้วุฒิการศึกษาอนุปริญญา 60 คน ใช้วุฒิปริญญาตรี 35 คน และใช้วุฒิสูงกว่าปริญญาตรีจำนวน 25 คน อยากทราบว่านักศึกษาสาขาวิชาวิทยาศาสตร์และเทคโนโลยีใช้วุฒิการศึกษาที่สมัครแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ 0.05 (การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.2.1(1) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ความถี่ที่คาดหวังคือนักศึกษาใช้วุฒิการศึกษาอนุปริญญา ปริญญาตรี และวุฒิสูงกว่าปริญญาตรี เท่ากัน

   จำนวนวุฒิการศึกษาละ $\frac{120}{3} = 40$ คน

2. ความถี่ที่สังเกตได้คือ มีนักศึกษาที่สมัครโดยใช้วุฒิการศึกษาอนุปริญญา 60 คน ใช้วุฒิปริญญาตรี 35 คน และใช้วุฒิสูงกว่าปริญญาตรีจำนวน 25 คน

โสตทัศน์ # 13.4 (ต่อ)

3. เปิดตารางไคสแคว์ ที่ชั้นแห่งความเป็นอิสระ 2 ระดับนัยสำคัญ 0.05 ได้ค่าวิกฤต 5.991 หรือกล่าวได้ว่าบริเวณวิกฤติเริ่มตั้งแต่ค่า 5.991 ถึงค่าอนันต์
4. คำนวณค่าไคสแคว์จากข้อมูลของกลุ่มตัวอย่างได้เท่ากับ 16.250 ดังนั้น ค่าที่คำนวณได้อยู่ในบริเวณวิกฤตจึงปฏิเสธสมมติฐานว่าง สรุปผลได้ว่านักศึกษาสาขาวิชาวิทยาศาสตร์และเทคโนโลยีสมัครโดยใช้วุฒิการศึกษาที่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

โสตทัศน์ # 13.5 การทดสอบการสุ่ม

การทดสอบการสุ่ม (The One–Sample Run Test of Randomness) เป็นสถิติทดสอบที่ใช้สำหรับทดสอบว่าการเกิดเหตุการณ์อย่างหนึ่งที่เป็นได้สองแบบว่าเป็นไปอย่างสุ่มหรือไม่ โดยพิจารณาจากจำนวนครั้งของการเกิดชุดของเหตุการณ์แต่ละแบบที่ต่อเนื่องกัน หรือเรียกว่า รัน (run)

ตัวแปรที่จะนำมาทดสอบการสุ่มจะต้องเป็นตัวแปรที่วัดในมาตรานามบัญญัติที่แบ่งได้ 2 ประเภท หรือตัวแปรที่วัดในมาตราเรียงลำดับ หรือสูงกว่า

การตั้งสมมุติฐาน

H_0 : ข้อมูลเกิดขึ้นตามลำดับอย่างสุ่ม

H_1 : ข้อมูลไม่ได้เกิดขึ้นตามลำดับอย่างสุ่ม

วิธีการทดสอบ

(1) กรณีกลุ่มตัวอย่างขนาดเล็ก (พิจารณาจากจำนวนการเกิดเหตุการณ์ทั้งสองประเภท ประเภทละไม่เกิน 20 จัดเป็นกลุ่มตัวอย่างขนาดเล็ก โดยการนำจำนวนรันของการทดลองหรือการสังเกตไปเปรียบเทียบกับค่าในตารางที่ 7 ในภาคผนวก ค่าวิกฤตที่ได้จากตารางจะประกอบด้วยค่า 2 ค่า ซึ่งเป็นช่วงของค่าวิกฤตที่ระดับนัยสำคัญ 0.05 ถ้าจำนวนรันตกอยู่ระหว่างค่าวิกฤตก็ไม่สามารถปฏิเสธสมมติฐานว่างได้ ถ้าจำนวนรันน้อยกว่าหรือเท่ากับค่าน้อยของค่าวิกฤต หรือมากกว่าหรือเท่ากับค่ามากของค่าวิกฤตก็จะปฏิเสธสมมติฐานว่าง

ตัวอย่าง ในการวิจัยเชิงสำรวจ ผู้วิจัยเก็บข้อมูลจากตัวอย่างจำนวน 30 คน โดยการเลือกแบบบังเอิญพบว่าได้กลุ่มตัวอย่างเป็นชาย 16 คน เป็นหญิง 14 คน โดยมีการเรียงลำดับของตัวอย่างชายและหญิงดังนี้

ช ช ญ ญ ญ ช ญ ญ ญ ช ช ช ช ญ ช ช ช ญ ญ ญ ช ญ ญ ช ช ช ญ ญ ช ช

อยากทราบว่าลำดับที่ของตัวอย่างชายและหญิงเป็นไปอย่างสุ่มหรือไม่ ที่ระดับนัยสำคัญ 0.05 (การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.2.2(1) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบและการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ถ้าให้ m แทนจำนวนตัวอย่างเพศชาย จะได้ m เท่ากับ 16
   n แทนจำนวนตัวอย่างเพศหญิง จะได้ n เท่ากับ 14
2. จากการเปิดตารางที่ 7 ค่าวิกฤตของรัน ในภาคผนวก เมื่อ m = 16 n = 14 ได้ค่าวิกฤต 10 และ 22
3. จากการทดลอง จำนวนรันเท่ากับ 13
4. ค่ารัน**ไม่อยู่**ในช่วงวิกฤต ดังนั้นจึงไม่สามารถปฏิเสธสมมติฐานว่าง สรุปได้ว่าลำดับที่ของตัวอย่างชายและหญิงเป็นไปอย่างสุ่ม

โสตทัศน์ # 13.6 กรณีกลุ่มตัวอย่างขนาดใหญ่

เมื่อกลุ่มตัวอย่างมีขนาดใหญ่การแจกแจงของรันจะเข้าสู่การแจกแจงปกติ (normal distribution) โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ดังนี้

$$\text{ค่าเฉลี่ย} = \mu_r = \frac{2mn}{N} + 1$$

$$\text{ส่วนเบี่ยงเบนมาตรฐาน} = \sigma_r = \sqrt{\frac{2mn\left(2mn - N\right)}{N^2\left(N-1\right)}}$$

สำหรับสถิติที่ใช้ทดสอบคือสถิติทดสอบ Z (Z test) ซึ่งมีสูตรดังนี้

$$Z = \frac{r - \mu_r}{\sigma_r} = \frac{r - \left(\frac{2mn}{N} + 1\right)}{\sqrt{\frac{2mn\left(2mn - N\right)}{N^2\left(N-1\right)}}}$$

เมื่อ r คือจำนวนรัน

μ_r คือค่าเฉลี่ยของรัน

σ_r คือค่าส่วนเบี่ยงเบนมาตรฐานของรัน

m คือจำนวนครั้งของการเกิดเหตุการณ์อย่างหนึ่ง

n คือจำนวนครั้งของการเกิดเหตุการณ์อีกอย่างหนึ่ง

N เท่ากับ m+n

ตัวอย่าง แบบทดสอบถูกผิดฉบับหนึ่งมี 50 ข้อ เฉลย ถ 27 ข้อ และเฉลย ผ 23 ข้อ คำตอบที่ถูกต้องมีการเรียงลำดับตั้งแต่ข้อ 1 ถึงข้อ 50 ดังนี้

ถ ถ ผ ถ ถ ถ ผ ผ ถ ผ ถ ถ ถ ผ ถ ผ ผ ถ ถ ถ ผ ถ ผ ผ

ถ ถ ผ ผ ถ ผ ถ ถ ผ ผ ผ ถ ผ ถ ถ ผ ผ ถ ถ ผ ถ ผ ถ ถ ผ ผ

อยากทราบว่าคำเฉลยของแบบทดสอบฉบับนี้เป็นไปอย่างสุ่มหรือไม่ ที่ระดับนัยสำคัญ 0.05 (การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.2.2(3) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ที่ระดับนัยสำคัญ .05 ค่าวิกฤตของ Z เท่ากับ ± 1.96
2. คำนวณค่าสถิติทดสอบจากข้อมูลของกลุ่มตัวอย่าง ได้ค่า Z = –1.197
3. ค่าสถิติทดสอบที่คำนวณได้ไม่อยู่ในบริเวณวิกฤต จึงไม่สามารถปฏิเสธสมมติฐานว่าง สรุปว่า คำเฉลยของแบบทดสอบฉบับนี้เป็นไปอย่างสุ่ม

โสตทัศน์ # 13.7 การทดสอบสมมติฐานโดยใช้สถิติศาสตร์ไม่อิงพารามิเตอร์ในกรณีกลุ่มตัวอย่างสองกลุ่มที่สัมพันธ์กัน

1. การทดสอบเครื่องหมาย
2. การทดสอบเครื่องหมายและลำดับที่ของวิลคอกซัน

โสตทัศน์ # 13.8 การทดสอบเครื่องหมาย

การทดสอบเครื่องหมาย (The Sign Test) เป็นสถิติศาสตร์ไม่อิงพารามิเตอร์ที่ใช้เครื่องหมายบวก (+) และเครื่องหมายลบ (–) แทนความแตกต่างของข้อมูลทีละคู่ แล้วนับจำนวนความถี่ของเครื่องหมายบวกและลบนั้นว่ามีจำนวนแตกต่างกันอย่างมีนัยสำคัญหรือไม่ และใช้ทดสอบว่ามัธยฐานของประชากรเท่ากับค่าที่กำหนดให้หรือไม่

การทดสอบเครื่องหมายใช้ทดสอบกับข้อมูลที่วัดในมาตราเรียงลำดับหรือสูงกว่า

การตั้งสมมติฐาน

ถ้าตัวแปรที่ต้องการเปรียบเทียบสองตัวคือตัวแปร X และ Y การทดสอบเครื่องหมายมีการตั้งสมมติฐานทั้งการทดสอบทางเดียวและการทดสอบสองทาง แต่ในที่นี้จะกล่าวเฉพาะการทดสอบสองทาง ดังนี้

H_o : ความน่าจะเป็นที่ X มากกว่า Y เท่ากับ ความน่าจะเป็นที่ X น้อยกว่า Y และเท่ากับ 0.5

หรือ $H_o : P[X_1 > Y_1] = P[X_1 < Y_1] = \frac{1}{2}$

H_1 : ความน่าจะเป็นที่ X มากกว่า Y ไม่เท่ากับ ความน่าจะเป็นที่ X น้อยกว่า Y

หรือ $H_1 : P[X_1 > Y_1] \neq P[X_1 < Y_1]$

วิธีการทดสอบ

(1) กรณีกลุ่มตัวอย่างขนาดเล็ก (จำนวนคู่ที่มีความแตกต่างน้อยกว่าหรือเท่ากับ 35 คู่

การทดสอบใช้การนับจำนวนเครื่องหมายบวก และเครื่องหมายลบ จำนวนเครื่องหมายที่น้อยกว่าให้เป็นค่า x ผลรวมของเครื่องหมายบวกและลบเป็นค่า N แล้วนำค่า N และ x ไปเปิด ตารางที่ 8 ในภาคผนวกค่าที่ปรากฏในภาคผนวกคือค่าความน่าจะเป็นที่ $p = \frac{1}{2}$ การปฏิเสธสมมุติฐานจะกระทำได้เมื่อค่าความน่าจะเป็นที่ได้จากการเปิดตารางน้อยกว่าระดับนัยสำคัญที่ตั้งไว้

โสตทัศน์ # 13.8 (ต่อ)

ตัวอย่าง ในการประเมินว่าใครเป็นผู้มีบทบาทต่อการตัดสินใจซื้อบ้านมากกว่ากันของสามีภรรยา 17 คู่ โดยให้ตอบลงในแบบสอบถามซึ่งเป็นแบบมาตรประมาณค่า 7 ค่า คะแนน 7 หมายถึง มีบทบาทต่อการตัดสินใจซื้อบ้านมากที่สุด คะแนนจะลดหลั่นลงมาเรื่อยๆ จนถึง 1 มีบทบาทต่อการตัดสินใจซื้อบ้านน้อยที่สุด ผลการประเมินปรากฏดังตาราง อยากทราบว่าสามีมีบทบาทต่อการตัดสินใจซื้อบ้านมากกว่าภรรยาหรือไม่ ที่ระดับนัยสำคัญ 0.05

คู่ที่	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
สามี	5	4	6	6	3	2	5	3	1	4	5	4	4	7	5	5	5
ภรรยา	3	3	4	5	3	3	2	3	2	3	2	2	5	2	5	3	1

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.3.1(1) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. มีเครื่องหมายบวก 11 เครื่องหมาย เครื่องหมายลบ 3 เครื่องหมาย นอกนั้นไม่มีเครื่องหมาย ดังนั้น N = 14 และ x = 3
2. เปิดตารางที่ 8 ที่ N = 14 และ x = 3 จะได้ค่าความน่าจะเป็นเท่ากับ 0.029
3. นำค่าความน่าจะเป็นที่เปิดตารางได้ไปเปรียบเทียบกับระดับนัยสำคัญที่ตั้งไว้ พบว่าน้อยกว่า จึงปฏิเสธสมมติฐานว่าง สรุปผลว่าสามีมีบทบาทต่อการตัดสินใจซื้อบ้านมากกว่าภรรยาที่ระดับนัยสำคัญ .05

โสตทัศน์ # 13.9 กรณีกลุ่มตัวอย่างขนาดใหญ่

ถ้า N มากกว่า 35 การแจกแจงทวินามจะเข้าใกล้การแจกแจงปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดังนี้

$$\text{ค่าเฉลี่ย} = \mu_x = N_p = \frac{N}{2}$$

$$\text{และ ความแปรปรวน} = \sigma_x^2 = N_{pq} = \frac{N}{4}$$

เนื่องจากการแจกแจงปกติเป็นการแจกแจงแบบต่อเนื่อง แต่ถูกนำมาประยุกต์ใช้กับการแจกแจงทวินาม ซึ่งเป็นการแจกแจงของตัวแปรไม่ต่อเนื่อง จึงใช้สูตรการปรับแก้ ดังนี้

$$Z = \frac{2x + 1 - N}{\sqrt{N}}$$

โสตทัศน์ # 13.9 (ต่อ)

ตัวอย่าง นักวิจัยมีความเชื่อว่าการจัดกิจกรรม 5 ส จะช่วยให้พนักงานของโรงพิมพ์มีความตระหนักในการประกันคุณภาพเพิ่มขึ้น จึงสุ่มพนักงานมา 100 คน ให้แต่ละคนประเมินความตระหนักในการประกันคุณภาพของตนเอง จากนั้นจัดกิจกรรม 5 ส เมื่อการจัดกิจกรรม 5 ส สิ้นสุดลง ก็ให้แต่ละคนประเมินความตระหนักในการประกันคุณภาพของตนเองอีกครั้ง ผลปรากฏดังตาราง อยากทราบว่าผลการวิจัยเป็นไปตามความเชื่อของนักวิจัยหรือไม่ ที่ระดับนัยสำคัญ 0.05

ผลการประเมินความตระหนัก	จำนวน
ความตระหนักเพิ่มขึ้น (+)	59
ความตระหนักลดลง (–)	26
ความตระหนักไม่เปลี่ยนแปลง	15

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.3.1(2) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ที่ระดับนัยสำคัญ .05 ค่าวิกฤตของ Z เท่ากับ – 1.645
2. คำนวณค่าสถิติทดสอบจากข้อมูลของกลุ่มตัวอย่าง ได้ค่า Z = –3.471
3. ค่าสถิติทดสอบที่คำนวณได้อยู่ในบริเวณวิกฤต
4. ค่า Z ที่คำนวณได้ตกอยู่บริเวณปฏิเสธสมมติฐานว่าง จึงสรุปผลได้ว่าการจัดกิจกรรม 5 ส จะช่วยให้พนักงานของโรงพิมพ์มีความตระหนักในการประกันคุณภาพมากขึ้น

โสตทัศน์ # 13.10 การทดสอบเครื่องหมายและลำดับที่ของวิลคอกซัน

การทดสอบเครื่องหมายและลำดับที่ของวิลคอกซัน เป็นสถิติศาสตร์ไม่อิงพารามิเตอร์ที่ใช้สำหรับทดสอบความแตกต่าง คล้ายกับการทดสอบ t ของกลุ่มตัวอย่างสองกลุ่มที่ไม่เป็นอิสระกัน แต่การทดสอบเครื่องหมายและลำดับที่ของวิลคอกใช้ลำดับที่ของข้อมูลมาประกอบการทดสอบ

ข้อมูลที่นำมาทดสอลต้องวัดในมาตราอันตรภาค หรือสูงกว่า และ ข้อมูลทั้งสองชุดมาจากกลุ่มตัวอย่างเดียวกันที่มีการวัดซ้ำ หรือเป็นกลุ่มตัวอย่างสองกลุ่มที่สัมพันธ์กัน เช่น การใช้คู่แฝด การจับคู่ เป็นต้น

. การตั้งสมมติฐาน

การตั้งสมมติฐานเป็นได้ทั้งการทดสอบทางเดียว และการทดสอบสองทาง แต่ในที่นี้จะกล่าวเฉพาะการทดสอบสองทาง

H_o : ผลบวกของลำดับที่ที่มีเครื่องหมายบวกเท่ากับผลบวกของลำดับที่ที่มีเครื่องหมายลบ

H_1 : ผลบวกของลำดับที่ที่มีเครื่องหมายบวกไม่เท่ากับผลบวกของลำดับที่ที่มีเครื่องหมายลบ

โสตทัศน์ # 13.10 (ต่อ)

วิธีการทดสอบ

(1) กรณีกลุ่มตัวอย่างขนาดเล็ก (จำนวนคู่ที่มีความแตกต่างน้อยกว่าหรือเท่ากับ 30 คู่)

ขั้นตอนการคำนวณดังนี้

(1) คำนวณผลต่างของข้อมูลแต่ละคู่โดยคิดเครื่องหมายตามหลักคณิตศาสตร์

(2) จัดลำดับที่ของผลต่างจากน้อยไปหามาก (โดยไม่คิดเครื่องหมาย) ในกรณีที่ผลต่างของข้อมูลเท่ากันให้คำนวณลำดับที่เฉลี่ย แล้วให้ข้อมูลทุกตัวใช้ลำดับเฉลี่ยนั้น ถ้าผลต่างของข้อมูลคู่ใดเป็น 0 จะไม่นำมาจัดลำดับ

(3) บันทึกเครื่องหมายของลำดับที่ตามเครื่องหมายที่คำนวณได้ในขั้นที่ (1)

(4) คำนวณผลรวมของลำดับที่ โดยแยกเป็นสองกลุ่มคือกลุ่มที่มีเครื่องหมายบวก กับ กลุ่มที่มีเครื่องหมายลบ

(5) ให้ค่าผลรวมของลำดับที่ที่มีค่าน้อยกว่าในขั้นที่ (4) เป็นค่า T

(6) นับจำนวนคู่ที่ผลต่างไม่เป็น 0 ให้เป็นค่า n

(7) นำค่า n และค่าระดับนัยสำคัญ ไปเปิดตารางที่ 9 ในภาคผนวก เพื่อหาค่าวิกฤตของ T

(8) เปรียบเทียบค่า T ในขั้นที่ (5) กับค่าวิกฤตที่เปิดตารางได้ในขั้นที่ (7) ถ้าค่า T อยู่ในบริเวณวิกฤต ก็จะสรุปว่าปฏิเสธสมมติฐานว่าง

ตัวอย่าง บริษัทผลิตยาแห่งหนึ่งต้องการทดสอบว่ายาสองชนิด จะมีประสิทธิภาพในการลดปริมาณโคเรสเตอรอลในกระแสเลือดได้มากน้อยต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.01 บริษัทจึงรับสมัครฝาแฝดเหมือน 10 คู่ ในฝาแฝดแต่ละคู่ให้ได้รับยาชนิดที่ 1 และชนิดที่ 2 โดยการสุ่ม เมื่อครบ 1 เดือน จึงวัดปริมาณโคเรสเตอรอลในกระแสเลือดที่ลดลง ผลปรากฏดังตาราง

คู่ที่	ยาชนิดที่ 1	ยาชนิดที่ 2
1	74	63
2	55	58
3	61	49
4	41	47
5	53	50
6	74	69
7	52	67
8	31	40
9	50	44
10	58	38

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.3.2(12) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

โสตทัศน์ # 13.10 (ต่อ)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ผลต่างที่เป็นบวกและลบรวมกันได้ 10 จำนวน ไม่มีผลต่างคู่ใดเป็น 0 ดังนั้น n = 10
2. ผลรวมของลำดับที่น้อยกว่าคือ 21 ดังนั้น T = 21
3. เปิดตารางที่ 9 ที่ n = 10, α = .01 และเป็นการทดสอบแบบสองทาง ได้ค่าวิกฤตเท่ากับ 3
4. ค่า T ที่คำนวณได้ไม่อยู่ในบริเวณวิกฤต จึงไม่สามารถปฏิเสธสมมติฐานว่าง สรุปผลได้ว่ายาชนิดที่ 1 และ 2 มีประสิทธิภาพในการลดปริมาณโคเรสเตอรอลในกระแสเลือดได้ไม่แตกต่างกันอย่างมีนัยสำคัญ

โสตทัศน์ # 13.11 กรณีกลุ่มตัวอย่างขนาดใหญ่

N มีค่ามากกว่า 30 การแจกแจงของ T จะมีการแจกแจงใกล้เคียงการแจกแจงปกติ ที่มีค่าเฉลี่ย และความแปรปรวนดังนี้

ค่าเฉลี่ย $= \mu_T = \dfrac{n(n+1)}{4}$

ความแปรปรวน $= \sigma_T^2 = \dfrac{n(n+1)(2n+1)}{24}$

ดังนั้น $Z = \dfrac{T - \mu_T}{\sigma_T} = \dfrac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$

เมื่อ Z มีการแจกแจงปกติมาตรฐาน

โสตทัศน์ # 13.11 กรณีกลุ่มตัวอย่างขนาดใหญ่

ตัวอย่าง โรงเรียนแห่งหนึ่งนำกระบวนการวิจัยมาใช้ในการเรียนการสอนทุกวิชา โดยมีความคาดหวังว่ากระบวนการจัดการเรียนการสอนดังกล่าวจะช่วยพัฒนากระบวนการคิดอย่างมีวิจารณญาณของนักเรียนได้ การเก็บข้อมูลใช้การวัดก่อน–หลังการทดลอง ผลปรากฏดังตาราง จากการทดสอบเบื้องต้นพบว่าข้อมูลมีความเบ้ อยากทราบว่าผลการทดลองเป็นไปตามสมมติฐานที่ตั้งไว้หรือไม่ ที่ระดับนัยสำคัญ 0.05

คนที่	คะแนนก่อนทดลอง	คะแนนหลังทดลอง	คนที่	คะแนนก่อนทดลอง	คะแนนหลังทดลอง	คนที่	คะแนนก่อนทดลอง	คะแนนหลังทดลอง
1	35	37	12	35	40	23	35	34
2	39	45	13	47	52	24	46	47
3	27	32	14	28	30	25	39	42
4	36	42	15	32	40	26	40	44
5	29	26	16	29	48	27	28	29
6	43	49	17	30	37	28	34	29
7	47	47	18	25	20	29	29	38
8	36	37	19	24	30	30	35	37
9	42	49	20	38	40	31	35	35
10	38	36	21	42	48	32	42	44
11	28	32	22	23	29	33	36	39

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.3.2(2) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. ที่ระดับนัยสำคัญ .05 ค่าวิกฤตของ Z เท่ากับ – 1.645
2. คำนวณค่าสถิติทดสอบจากข้อมูลของกลุ่มตัวอย่าง ได้ค่า Z = –3.72
3. ค่าสถิติทดสอบที่คำนวณได้อยู่ในบริเวณวิกฤต
4. คะแนนก่อนทดลองน้อยกว่าคะแนนหลังทดลอง หรือกล่าวอีกนัยหนึ่งว่า กระบวนการจัดการเรียนการสอนโดยใช้การวิจัยจะช่วยพัฒนากระบวนการคิดอย่างมีวิจารณญาณของนักเรียนได้อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

โสตทัศน์ # 13.12 การทดสอบสมมติฐานโดยใช้สถิติศาสตร์ไม่อิงพารามิเตอร์ในกรณีกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระกัน

1. การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระกัน
2. การทดสอบวิลคอกซัน–แมนน์–วิทนีย์

1. การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระกัน

การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระกัน ใช้สำหรับทดสอบว่า เมื่อจำแนกกลุ่มตัวอย่างออกเป็นประเภทๆ ตามคุณลักษณะสองอย่างแล้ว คุณลักษณะสองอย่างนั้นมีความสัมพันธ์กันหรือเป็นอิสระกัน

ข้อตกลงเบื้องต้นสำหรับการทดสอบ มีดังนี้

1. กลุ่มตัวอย่างมาจากการสุ่ม และควรมีขนาดค่อนข้างใหญ่
2. ตัวแปรที่ใช้ในการจำแนกคุณลักษณะวัดในมาตรานามบัญญัติ ถ้าตัวแปรนั้นวัดในมาตราที่สูงกว่านามบัญญัติต้องแปลงให้อยู่ในมาตรานามบัญญัติ
3. การจำแนกข้อมูลออกเป็นสองลักษณะต้องเป็นอิสระกัน นั่นคือไม่มีข้อมูลตัวใดตกอยู่เกินหนึ่งเซลล์

การตั้งสมมติฐาน

การทดสอบไคสแควร์กรณีกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระกัน มีการตั้งสมมติฐาน ดังนี้

H_0 : คุณลักษณะสองลักษณะในประชากรเป็นอิสระกัน

H_1 : คุณลักษณะสองลักษณะในประชากรไม่อิสระกัน

วิธีการทดสอบ

ใช้หลักการเดียวกับการทดสอบไคสแควร์สำหรับกลุ่มตัวอย่างเดียว คือเปรียบเทียบความถี่ที่สังเกตได้กับความถี่ที่คาดหวัง หรือความถี่ตามทฤษฎี โดยมีสูตรการคำนวณ ดังนี้

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad ; \quad df = (r-1)(c-1)$$

เมื่อ O_{ij} คือความถี่ที่สังเกตได้ในแถวที่ i สดมภ์ที่ j

r คือจำนวนแถวทั้งหมด

c คือจำนวนสดมภ์ทั้งหมด

E_{ij} คือความถี่ที่คาดหวังในแถวที่ i สดมภ์ที่ j ซึ่งคำนวณได้จากสูตร

$$E_{ij} = \frac{R_i C_j}{n}$$

R_i คือผลรวมของความถี่ที่สังเกตได้ในแถวที่ i

C_i คือผลรวมของความถี่ที่สังเกตได้ในสดมภ์ที่ j

n คือขนาดของกลุ่มตัวอย่าง

โสตทัศน์ # 13.12 (ต่อ)

ตัวอย่าง บริษัทผลิตยาสีฟันแห่งหนึ่งผลิตยาสีฟันออกมา 3 รสคือ รสมินท์ รสสตรอเบอรี และรสส้ม คณะผู้วิจัยได้นำผลิตภัณฑ์ทั้ง 3 รส ไปให้กลุ่มตัวอย่างที่มีวัยต่างกัน 3 กลุ่ม คือ เด็ก ผู้ใหญ่ที่อายุต่ำกว่า 40 ปี และผู้ที่อายุ 40 ปีขึ้นไป ที่สุ่มมากลุ่มละ 50 คน รวม 150 คน ทดลองใช้ แล้วบอกว่าตนเองชอบยาสีฟันรสใดมากที่สุด ผลการเก็บรวบรวมข้อมูล ดังตาราง จงทดสอบว่าการชอบยาสีฟันรสใดขึ้นอยู่กับวัยหรือไม่ที่ระดับนัยสำคัญ 0.05

กลุ่มตัวอย่าง	จำนวนผู้ที่ตอบว่าชอบมากที่สุด			รวม
	รสมินท์	รสสตรอเบอรี	รสส้ม	
เด็ก	30	35	31	96
ผู้ใหญ่ที่อายุต่ำกว่า 40 ปี	8	11	11	30
ผู้ที่อายุ 40 ปีขึ้นไป	12	4	8	24
รวม	50	50	50	150

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.4.1(1) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. เปิดตาราง χ^2 ในภาคผนวกที่ระดับนัยสำคัญ 0.05 df = (3–1)(3–1) =4 ได้ค่า χ^2 = 9.488
2. คำนวณค่า χ^2 จากข้อมูลของกลุ่มตัวอย่างได้ค่า χ^2 = 5.0375
3. ค่า χ^2 ที่คำนวณได้ตกไม่อยู่ในบริเวณวิกฤต จึงไม่สามารถปฏิเสธสมมติฐานว่าง จึงสรุปผลได้ว่าการชอบยาสีฟันรสใดกับวัยเป็นอิสระกัน

ในกรณีที่ตารางการจรมี 2 แถว และ 2 สดมภ์ นั่นคือตัวแปรแบ่งออกเป็น 2 ประเภททั้งสองตัวแปรเรียกตารางการจรแบบนี้ว่า แบบ 2x2 การคำนวณค่าไคสแควร์จะใช้สูตรแก้ไขความต่อเนื่อง ดังนี้

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\left|O_{ij} - E_{ij}\right| - \frac{1}{2}\right)^2}{E_{ij}}$$

โสตทัศน์ # 13.13 การทดสอบ วิลคอกซัน–แมน–วิทนีย์

การทดสอบ วิลคอกซัน–แมน–วิทนีย์ เป็นสถิติศาสตร์ไม่อิงพารามิเตอร์ที่ใช้กับข้อมูลที่วัดในระดับเรียงลำดับหรือสูงกว่า เพื่อทดสอบว่ากลุ่มตัวอย่างที่เป็นอิสระกันสองกลุ่มมาจากประชากรที่มีการแจกแจงเหมือนกันหรือไม่ โดยใช้ลำดับที่ของข้อมูลเป็นตัวทดสอบ

การทดสอบ วิลคอกซัน–แมน–วิทนีย์ มีข้อตกลงเบื้องต้นที่สำคัญ ดังนี้

1. ตัวแปรที่ต้องการทดสอบมีการวัดในมาตราเรียงลำดับหรือสูงกว่า และเป็นตัวแปรแบบต่อเนื่อง (ตัวแปรที่มีการวัดในระดับนามบัญญัติ หรือตัวแปรไม่ต่อเนื่องใช้ไม่ได้)
2. ขนาดของกลุ่มตัวอย่างทั้งสองกลุ่มรวมกันไม่น้อยกว่า 8 และอีกกลุ่มหนึ่งมีขนาดของกลุ่มตัวอย่างไม่น้อยกว่า 3

การตั้งสมมติฐาน

การตั้งสมมุติฐานอาจตั้งเกี่ยวกับการแจกแจง หรือตั้งเกี่ยวกับค่ากลางของประชากรสองกลุ่มก็ได้ และสามารถตั้งสมมติฐนการทดสอบแบบทางเดียว หรือสองทางก็ได้ แต่ในที่นี้จะยกตัวอย่างเฉพาะการทดสอบแบบสองทาง ดังนี้

H_0 : การแจกแจงของประชากรกลุ่มที่ 1 และกลุ่มที่ 2 เหมือนกัน

H_1 : การแจกแจงของประชากรกลุ่มที่ 1 และกลุ่มที่ 2 ต่างกัน

หรือ

H_0 : ค่ากลางของประชากรกลุ่มที่ 1 และกลุ่มที่ 2 เท่ากัน

H_1 : ค่ากลางของประชากรกลุ่มที่ 1 และกลุ่มที่ 2 ต่างกัน

วิธีการทดสอบ

หลักการทดสอบใช้การเรียงลำดับข้อมูลของกลุ่มตัวอย่างทั้งสองกลุ่มด้วยกัน แล้วหาผลรวมของตำแหน่งของกลุ่มตัวอย่างแยกทีละกลุ่ม ถ้าประชากรทั้งสองกลุ่มมีการแจกแจงเหมือนกัน ผลรวมของตำแหน่งควรเท่ากัน หรือต่างกันเพียงเล็กน้อย ถ้าผลรวมของตำแหน่งต่างกันมากกว่าค่าวิกฤต แสดงว่าประชากรทั้งสองกลุ่มมีการแจกแจงต่างกัน

ขั้นตอนการทดสอบมีดังนี้

1. นำคะแนนของกลุ่มตัวอย่างที่ 1 และกลุ่มที่ 2 มารวมเป็นกลุ่มเดียวกัน
2. เรียงลำดับคะแนนในข้อ 1 จากน้อยไปมากแล้วให้ข้อมูลตัวที่น้อยที่สุดเป็นตำแหน่งที่ 1 ถ้าคะแนนซ้ำกันให้ใช้ตำแหน่งเฉลี่ย
3. คำนวณผลรวมของตำแหน่งของกลุ่มตัวอย่างแต่ละกลุ่ม ให้ U เป็นผลรวมของตำแหน่งของกลุ่มที่น้อยกว่า
4. คำนวณค่า สถิติทดสอบ จากสูตร

$$U_{obt} = n_1 n_2 + \frac{n_1\left(n_1 + 1\right)}{2} - U$$

เมื่อ n_1 เป็นขนาดของกลุ่มตัวอย่างที่เล็กกว่า

n_2 เป็นขนาดของกลุ่มตัวอย่างอีกกลุ่มหนึ่ง

U เป็นผลรวมของตำแหน่งที่น้อยกว่า

นำค่า U_{obt} ไปเปรียบเทียบกับค่าวิกฤตที่ได้จากการเปิดตารางที่ 10 ในภาคผนวก

โสตทัศน์ # 13.13 (ต่อ)

ตัวอย่าง ผู้วิจัยต้องการเปรียบเทียบระยะเวลาในการสำรองไฟ (หน่วย: ชั่วโมง) ของแบตเตอรีก่อนที่จะต้องชาร์จครั้งต่อไปของเครื่องคอมพิวเตอร์โน้ตบุกสองตราสินค้า ว่าแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ .05 โดยสุ่มคอมพิวเตอร์โน๊ตบุกมาตราสินค้าละ 4 เครื่อง ผลการทดลองเป็นดังนี้

ตราสินค้าที่ 1	3.6	3.9	4.0	4.3
ตราสินค้าที่ 2	3.8	4.1	4.5	4.8

(การทดสอบโดยละเอียดแสดงในตัวอย่างที่ 13.4.2(1) ในที่นี้จะเน้นเฉพาะการสรุปผลการทดสอบ และการแปลความหมายเท่านั้น)

จากการทดสอบมีประเด็นสำคัญดังนี้

1. เปิดตารางที่ 10 ที่ระดับนัยสำคัญ .05 $n_1 = 4$ และ $n_2 = 4$ ได้ค่าวิกฤตเท่ากับ 0 หมายความว่า ค่า U_{obt} จะต้องน้อยกว่าหรือเท่ากับ 0 การทดสอบจึงจะมีนัยสำคัญทางสถิติที่ระดับ .05
2. คำนวณผลรวมของลำดับที่ของตราสินค้าที่ 1 ได้ 1+3+4+6 = 14
3. คำนวณผลรวมของลำดับที่ของตราสินค้าที่ 2 ได้ 2+5+7+8 = 22
4. คำนวณค่า U_{obt} ได้เท่ากับ 12
5. นำค่า U_{obt} ที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตพบว่าค่า U_{obt} ไม่อยู่บริเวณวิกฤต จึงไม่สามารถปฏิเสธสมมติฐานว่าง สรุปผลการทดลองว่า ระยะเวลาในการสำรองไฟของแบตเตอรีของเครื่องคอมพิวเตอร์โน้ตบุก ตราสินค้าที่ 1 และตราสินค้าที่ 2 ไม่ต่างกัน

ในกรณีที่กลุ่มตัวอย่างทั้งสองกลุ่มมีขนาดมากกว่าหรือเท่ากับ 10 ตัวสถิติ U จะมีการแจกแจงใกล้เคียงกับการแจกแจงปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ดังนี้

$$\mu_u = \frac{n_1\left(n_1 + n_2 + 1\right)}{2}$$

$$\sigma_u = \sqrt{\frac{n_1 n_2\left(n_1 + n_2 + 1\right)}{12}}$$

n_1 คือ ขนาดของกลุ่มตัวอย่างกลุ่มที่ 1

n_2 คือ ขนาดของกลุ่มตัวอย่างกลุ่มที่ 2

ดังนั้นในการทดสอบวิลคอกซัน–แมน–วิทนีย์ ที่กลุ่มตัวอย่างมีขนาดใหญ่จึงใช้สถิติทดสอบ Z ที่มีสูตรดังนี้

สูตร $$Z = \frac{u - \mu_u}{\sigma_u}$$

โสตทัศน์ # 13.13 (ต่อ)

กิจกรรม

จงเลือกสถิติทดสอบให้เหมาะสมกับสถานการณ์ต่อไปนี้ พร้อมทั้งแสดงการทดสอบด้วย

1. จากการรวบรวมข้อมูลคะแนนสอบวิชาสถิติของนักศึกษาหญิงและนักศึกษาชาย กลุ่มละ 20 คน ได้ผลดังตาราง ซึ่งคะแนนมีความเบ้ จงเปรียบเทียบว่าคะแนนของนักศึกษาหญิงกับนักศึกษาชายต่างกันหรือไม่ที่ระดับนัยสำคัญ .05

ชาย	39	32	36	52	32	33	15	43	78	42	17	19	50	32	40	51	47	55	49	37
หญิง	26	47	60	12	58	50	29	63	32	34	55	60	26	60	39	25	62	51	34	62

2. ในการจับฉลากเด็กนักเรียนเพื่อเข้าเรียนชั้น มัธยมศึกษาปีที่ 1 ของโรงเรียนแห่งหนึ่งได้ผลดังนี้

   ช ช ช ช ญ ญ ช ญ ญ ช ญ ญ

   ช ญ ช ญ ช ญ ช ญ

   การจับฉลากได้นักเรียน ชาย หญิงเป็นไปอย่างสุ่มหรือไม่ ที่ระดับนัยสำคัญ .05

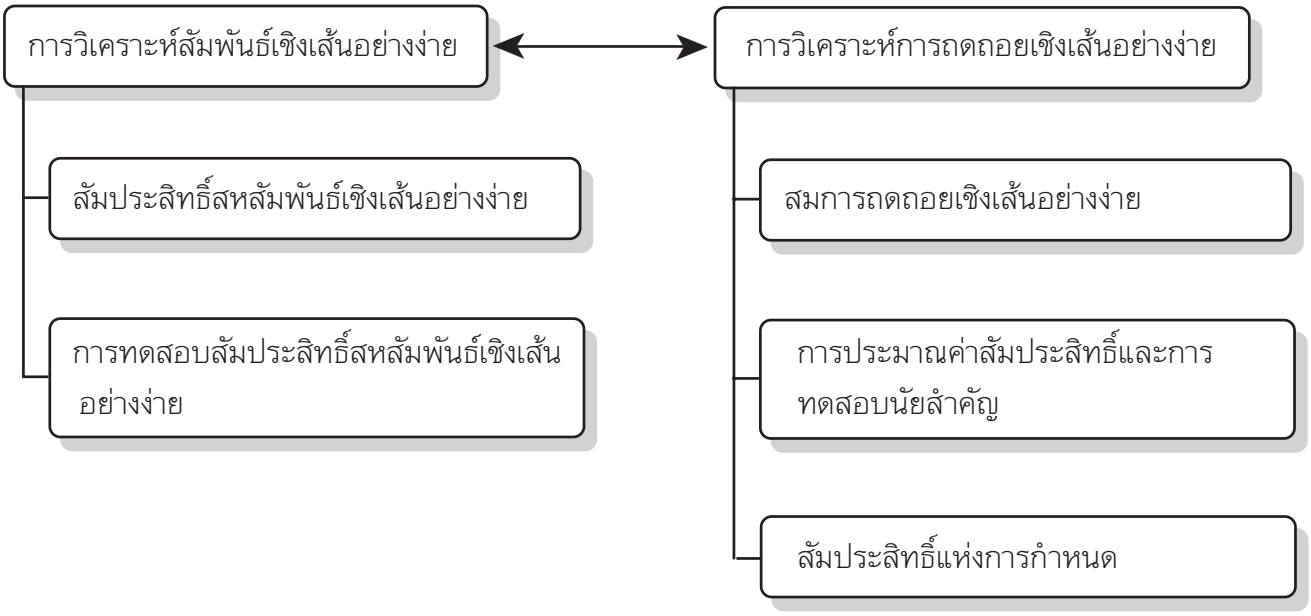
3. ในการทดสอบวัดความจุของปอดของกลุ่มตัวอย่าง 10 คน ในท่านั่งและท่านอนหงายได้ความจุของปอด (หน่วย: ลิตร) ดังตาราง จงทดสอบว่าไม่มีความแตกต่างของความจุปอดเมื่อวัดในท่านั่งและท่านอนหงายที่ระดับนัยสำคัญ .05

ท่านั่ง	2.96	4.65	3.27	2.50	2.59	5.97	1.74	3.51	4.37	4.02
ท่านอนหงาย	1.97	3.05	2.29	1.68	1.58	4.43	1.53	2.81	2.70	2.70

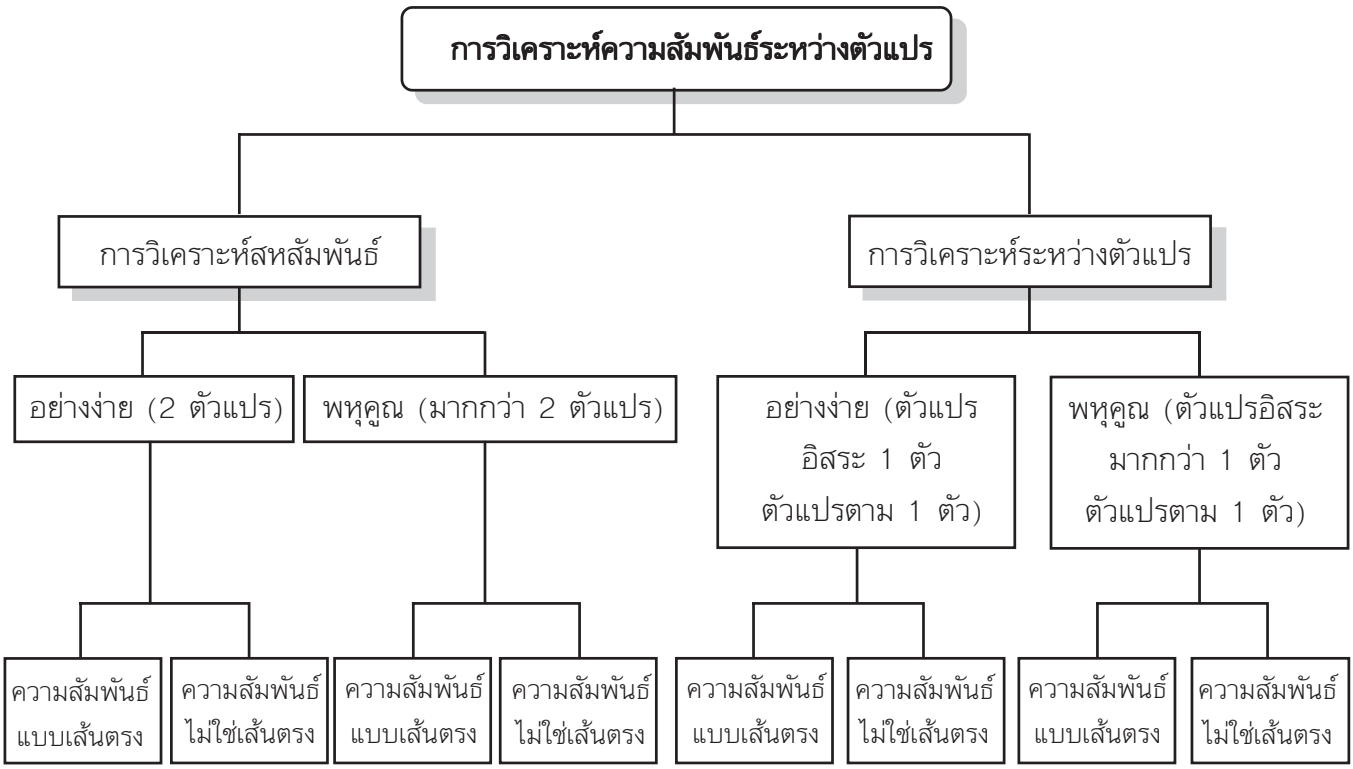
4. สภากาชาดไทยรายงานว่ากลุ่มเลือดของคนไทย A B AB และ O คิดเป็นร้อยละ 40 10 5 และ 45 ตามลำดับ จากการสำรวจผู้มาบริจาคโลหิตจำนวน 80 คน พบว่ามีกลุ่มเลือด A B AB และ O จำนวน 32 10 6 และ 32 ตามลำดับ อยากทราบว่าการสำรวจสอดคล้องกับรายงานของสภากาชาดไทยหรือไม่ที่ระดับนัยสำคัญ 0.05
5. โรงพยาบาลแห่งหนึ่งวัดไข้คนไข้ได้สูงกว่า 37 องศาเซลเซียส อยู่ 68 คน ต่ำกว่า 37 องศาเซลเซียส อยู่ 23 คน และเท่ากับ 37 องศาเซลเซียส อยู่ 15 คน จงทดสอบว่ามัธยฐานของอุณภูมิของคนไข้เท่ากับ 37 องศาเซลเซียส หรือไม่ที่ระดับนัยสำคัญ 0.01
6. นักวิจัยต้องการทดสอบว่าการสูบบุหรี่จะมีผลต่อการเกิดรอยเหี่ยวย่นรอบดวงตาหรือไม่ จึงสุ่มผู้ที่สูบบุหรี่และไม่สูบบุหรี่มา 500 คน แล้วสังเกตรอยเหี่ยวย่นรอบดวงตา พบข้อมูลดังตาราง จงทดสอบว่าการสูบบุหรี่กับการเกิดรอยเหี่ยวย่นรอบดวงตาสัมพันธ์กันหรือไม่ ที่ระดับนัยสำคัญ .05

การสูบบุหรี่	มี	ไม่มี
สูบ	103	52
ไม่สูบ	112	233

หน่วยที่ 14
การวิเคราะห์สหสัมพันธ์และการถดถอยเชิงเส้นอย่างง่าย



โสตทัศน์ # 14.1 การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร



การวิเคราะห์สหสัมพันธ์ เป็นวิธีการตรวจหาระดับความสัมพันธ์ระหว่างตัวแปรต่างๆ ว่ามีมากน้อยเพียงใด จึงไม่สนใจว่าตัวแปรใดเป็นเหตุ และตัวแปรใดเป็นผล ค่าประมาณของความสัมพันธ์เรียกว่าสัมประสิทธิ์สหสัมพันธ์

การวิเคราะห์การถดถอย เป็นวิธีการที่เกี่ยวข้องกับการหากฎเกณฑ์ หรือสมการพีชคณิตหรือตัวแบบที่ใช้กำหนดความสัมพันธ์เชิงคณิตศาสตร์ระหว่างตัวแปรที่ต้องการศึกษา เพื่อนำมาใช้ในการพยากรณ์หรือคาดคะเนเกี่ยวกับตัวแปรที่ต้องการศึกษาโดยอาศัยค่าของตัวแปรที่เกี่ยวข้อง

ดังนั้นในการวิเคราะห์การถดถอยจะต้องสร้างสมการที่เป็นตัวแทนของความสัมพันธ์ระหว่างตัวแปรที่ต้องการศึกษาเรียกว่า**ตัวแปรตาม**กับตัวแปรที่เกี่ยวข้องที่เรียกว่า**ตัวแปรอิสระ** ค่าประมาณของความสัมพันธ์เรียกว่า **สัมประสิทธิ์การถดถอย**

การวิเคราะห์การถดถอยจึงมุ่งศึกษาอิทธิพลที่ตัวแปรอิสระมีต่อตัวแปรตามและนำไปใช้ประโยชน์ในการพยากรณ์ค่าตัวแปรตาม

ขอบข่ายของการศึกษาความสัมพันธ์ระหว่างตัวแปรในหน่วยนี้เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวที่มีความสัมพันธ์แบบเส้นตรง จึงเป็นการศึกษา**การวิเคราะห์สหสัมพันธ์และการถดถอยเชิงเส้นอย่างง่าย**

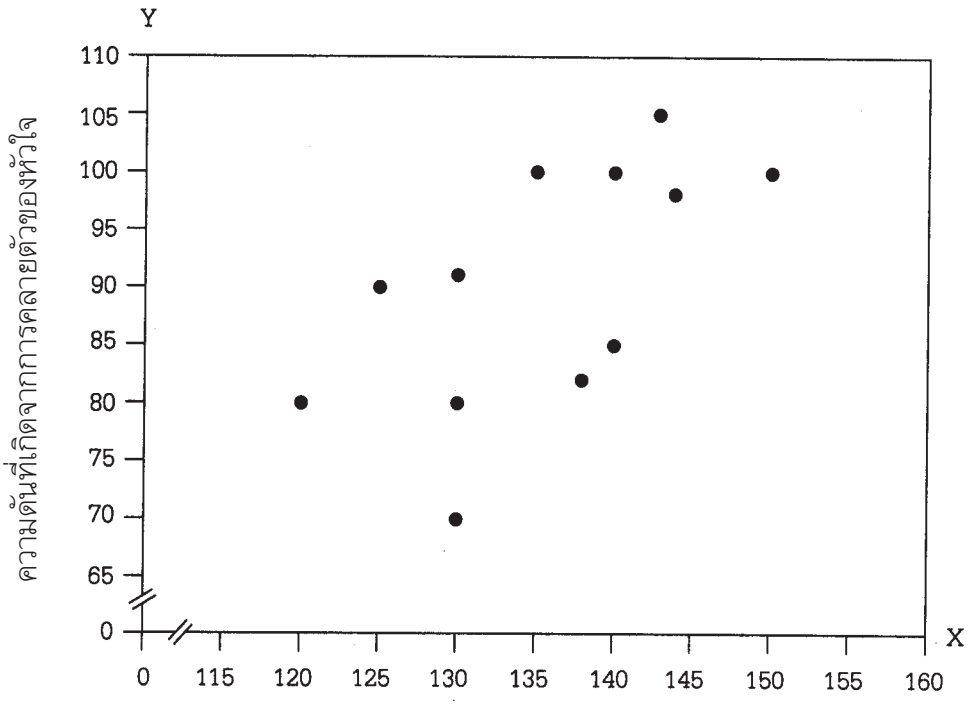
โสตทัศน์ # 14.1 แนวคิดเกี่ยวกับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

1. การสังเกตความสัมพันธ์ระหว่างตัวแปร

1.1 การศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวใด สามารถดูได้จากแผนภาพการกระจายซึ่งเป็นการนำค่าของตัวแปร x และ y ทีละคู่ มาลงบนจุดระนาบ xy เพื่อศึกษาทิศทางของความสัมพันธ์ระหว่างตัวแปร x กับ y

1.2 **ตัวแปร** คือ สิ่งที่ต้องการศึกษาหรือลักษณะที่ต้องการศึกษา ซึ่งเปลี่ยนแปลงค่าได้

ตัวอย่าง การศึกษาความสัมพันธ์ระหว่างความดันที่เกิดจากการบีบตัวของหัวใจและความดันที่เกิดจากการคลายตัวของหัวใจในการวัดความดันโลหิต



ภาพที่ 1 แผนภาพการกระจายของความดันที่เกิดจากการบีบตัวของหัวใจกับความดันที่เกิดจากการคลายตัวของหัวใจ

2. ลักษณะความสัมพันธ์ระหว่างตัวแปร

เป็นเส้นตรง เส้นโค้ง พาราโบลา วิธีการตรวจสอบลักษณะความสัมพันธ์ที่ง่ายที่สุดคือการเขียนกราฟแสดงความสัมพันธ์ของตัวแปรทั้งสองในแผนภาพการกระจาย ลักษณะของความสัมพันธ์ที่เป็นเส้นตรง มีทั้งความสัมพันธ์ที่เป็นทางบวก คือความสัมพันธ์ระหว่างตัวแปรมีทิศทางไปในทิศทางเดียวกันหรือมีความสัมพันธ์ทางลบคือความสัมพันธ์ระหว่างตัวแปรสองตัวมีทิศทางตรงข้ามกัน

โสตทัศน์ # 14.3 สัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่าย

1. การหาค่าสัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่าย

แผนภาพการกระจายแสดงให้เป็นทิศทางความสัมพันธ์ระหว่างตัวแปรสองตัวและลักษณะของความสัมพันธ์ว่าเป็นเส้นตรงหรือไม่เท่านั้น แต่ไม่สามารถบอกค่าระดับความสัมพันธ์ว่ามีขนาดเท่าใดต้องใช้**ค่าสัมประสิทธิ์สหสัมพันธ์**เป็นตัวบอก

1.1 **สัมประสิทธิ์สหสัมพันธ์ "r"** เป็นค่าที่ใช้วัดระดับความสัมพันธ์เชิงเส้นระหว่างตัวแปร x และตัวแปร y สำหรับข้อมูลตัวอย่างที่รวบรวมไว้

1.2 **สัมประสิทธิ์สหสัมพันธ์ "ρ"** เป็นค่าที่ใช้วัดระดับความสัมพันธ์ระหว่างตัวแปร x และ ตัวแปร y สำหรับข้อมูลในประชากร การคำนวณค่า ρ จะต้องทราบข้อมูลทั้งหมดในประชากร ซึ่งในทางปฏิบัติแล้วจะไม่ทราบค่าเหล่านี้ ดังนั้นจึงประมาณค่า ρ ด้วยค่า r

สัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่าย สามารถคำนวณได้จากสูตร

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\left(\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\right)\left(\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}\right)}$$

1.3 ค่าของตัวแปร x และตัวแปร y ที่ใช้ในสูตรต้องเป็นข้อมูลเชิงปริมาณ ปัจจุบันสามารถใช้โปรแกรมสำเร็จรูปในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ได้หลายโปรแกรม

1.4 r มีค่าตั้งแต่ –1 ถึง 1 นั่นคือ $-1 \leq r \leq 1$

1.5 กรณีที่ความสัมพันธ์เชิงเส้นระหว่างตัวแปร x และตัวแปร y เป็นไปในทิศทางเดียวกัน ค่า r ที่คำนวณได้จะมีค่าเป็นบวก

1.6 แต่ถ้าความสัมพันธ์เชิงเส้นเป็นไปในทิศทางตรงข้ามกัน ค่า r ที่คำนวณได้จะมีค่าเป็นลบ

1.7 หากไม่มีความสัมพันธ์ระหว่างตัวแปร x และตัวแปร y ค่า r ที่คำนวณได้จะมีค่าเป็น 0

2. ข้อสังเกตในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์

1. การอธิบายความหมายค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ จะอธิบายในรูปของระดับความสัมพันธ์ไม่อธิบายในความหมายของการเป็นเหตุเป็นผลกัน

2. ก่อนคำนวณค่าตามสูตรควรตรวจสอบลักษณะของความสัมพันธ์ว่าเป็นเส้นตรง

โสตทัศน์ # 14.3 (ต่อ)

ตัวอย่าง ผู้ผลิตชิ้นส่วนโลหะใช้วิธีการทดสอบความแข็งของโลหะ 2 วิธี คือ วิธีการใช้เครื่องจักรได้ค่าข้อมูลออกมาชุดหนึ่ง (x) และขณะเดียวกันก็ทดสอบด้วยวิธีการไม่ใช้เครื่องจักรได้ค่าข้อมูลออกมาอีกชุดหนึ่ง (y) โดยค่าความแข็งของโลหะที่วัดได้มีดังนี้

x	4.1	5.1	9.8	7.5	8.1	6.5	7.1	9.1	9.6	8.4
y	4.3	5.0	9.6	7.6	8.0	6.4	7.1	9.4	9.5	8.5

จงหาความสัมพันธ์ระหว่างความแข็งของโลหะที่วัดโดยใช้เครื่องจักรกับที่ไม่ใช้เครื่องจักร และอธิบายความหมายของค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้

วิธีทำ 1) คำนวณค่าต่าง ๆ ที่ต้องใช้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์

$\Sigma x = 75.3 \quad \Sigma y = 75.4 \quad \Sigma xy = 599.16 \quad \Sigma x^2 = 598.91 \quad \Sigma y^2 = 598.64$

2) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\left(\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\right)\left(\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}\right)}$$

$$= \frac{10\left(599.16\right) - \left(75.3\right)\left(75.4\right)}{\left(\sqrt{10\left(598.91\right) - \left(75.3\right)^2}\right)\left(\sqrt{10\left(599.64\right) - \left(75.4\right)^2}\right)}$$

$$= \frac{313.98}{315.101} = 0.996$$

3) ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าเท่ากับ 0.996 และมีค่าเป็นบวก แสดงถึงความสัมพันธ์ระหว่างความแข็งของโลหะที่วัดโดยใช้เครื่องจักรกับที่ไม่ใช้เครื่องจักรเป็นไปในทิศทางเดียวกันและมีค่าค่อนข้างสูง

โสตทัศน์ # 14.4 การทดสอบสมมติฐานของสัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่าย

1. การทดสอบสมมุติฐาน

ค่าสัมประสิทธิ์สหสัมพันธ์ r ที่คำนวณได้จากข้อมูลกลุ่มตัวอย่าง หากต้องการอ้างอิงไปยังความสัมพันธ์ในประชากรจะต้องทำการทดสอบสมมติฐานก่อนโดยใช้สถิติทดสอบ t ที่ df เท่ากับ n–2

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

ในการทดสอบสมมติฐานนั้นสามารถทดสอบสมมติฐานแบบสองทางและทางเดียวได้ ทั้งนี้ขึ้นอยู่กับสมมติฐานที่กำหนดขึ้น

1.1 การทดสอบสมมติฐานสองทาง

การทดสอบสมมติฐานสองทางสำหรับสัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่ายใช้ในกรณีที่ไม่ทราบว่าตัวแปรที่ต้องการศึกษาทั้งสองตัวมีความสัมพันธ์แบบทางตรงหรือแบบตรงข้ามกัน

1.2 การทดสอบสมมติฐานทางเดียว

1.2.1 การทดสอบสมมติฐานสำหรับความสัมพันธ์ทางบวก

$H_0 : \rho = 0$ หรือ $H_0 : \rho \leq 0$

$H_1 : \rho > 0$ $H_1 : \rho > 0$

ยอมรับสมมติฐาน เมื่อค่าสถิติทดสอบน้อยกว่าค่าวิกฤต

1.2.2 การทดสอบสมมติฐานสำหรับความสัมพันธ์ทางลบ

$H_0 : \rho = 0$ หรือ $H_0 : \rho \geq 0$

$H_1 : \rho < 0$ $H_1 : \rho < 0$

ยอมรับสมมติฐาน H_0 เมื่อค่าสถิติทดสอบมากกว่าค่าวิกฤต

โดยมีขั้นตอนของการทดสอบสมมติฐานและสถิติทดสอบเช่นเดียวกับการทดสอบสมมติฐานสองทาง

โสตทัศน์ # 14.4 (ต่อ)

ขั้นตอนการทดสอบสมมติฐาน

1. การกำหนดสมมติฐาน

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

↓

2. เลือกระดับนัยสำคัญทางสถิติ (α)

↓

3. คำนวณค่าสถิติ r

↓

4. คำนวณค่าสถิติทดสอบ t คือ

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

ค่าวิกฤต t เปิดจากตาราง

ด้วยองศาแห่งความเป็นอิสระ n–2 และระดับนัยสำคัญที่กำหนด

↓

5. การตัดสิน:

5.1 ถ้าค่าสถิติทดสอบตกอยู่ในบริเวณวิกฤตคือปฏิเสธสมมติฐานว่าง ($H_0 : \rho = 0$)

5.2 ถ้าค่าสถิติทดสอบอยู่นอกบริเวณวิกฤตคือยอมรับสมมติฐานว่าง

↓

6. การแปลความหมาย:

6.1 กรณีปฏิเสธสมมติฐานว่างอธิบายได้ว่ามีหลักฐานเพียงพอที่จะสรุปว่าตัวแปร x และตัวแปร y มีความสัมพันธ์กัน

6.2 กรณียอมรับสมมติฐานว่างอธิบายได้ว่ายังมีหลักฐานไม่เพียงพอที่จะสรุปว่าตัวแปร x ตัวแปร y มีความสัมพันธ์กัน

2. ข้อสังเกตในการทดสอบสมมุติฐานของสัมประสิทธิ์สหสัมพันธ์

1. สถิติทดสอบ t จะใช้ทดสอบสมมติฐาน $H_0 : \rho = 0$ หรือ $\rho > 0$ หรือ $\rho < 0$ เท่านั้น ถ้าต้องการทดสอบว่า $\rho =$ 0.80 หรือค่าอื่น ๆ ที่นอกเหนือไปจาก 0 จะต้องใช้ Fisher's transformation

โสตทัศน์ # 14.4 (ต่อ)

2. ผลจากการทดสอบค่า r ว่า มีความแตกต่างจากศูนย์อย่างมีนัยสำคัญ มิใช่เหตุผลที่จะใช้ในการพิจารณาว่าตัวแปรดังกล่าวมีความสัมพันธ์กันมากพอ เช่น ในกรณีที่ตัวอย่างขนาดใหญ่คำนวณได้ค่า r น้อย แต่นำไปทดสอบนัยสำคัญ ปรากฏว่าค่า r แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ

3. ในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ตัวแปร x และตัวแปร y เป็นตัวแปรสุ่มทั้งคู่ แต่ถ้าค่า x ถูกกำหนดไว้ก่อน เช่น การศึกษาความสัมพันธ์ระหว่างน้ำหนักกับส่วนสูง โดยกำหนดค่าของน้ำหนักที่จะศึกษาไว้ที่ 10 15 20 25 30 ... กิโลกรัม เป็นต้น ในกรณีเช่นนี้ ไม่ควรวิเคราะห์โดยใช้สัมประสิทธิ์สหสัมพันธ์

ตัวอย่าง จากข้อมูลในการวัดความดันโลหิตของนักศึกษา 14 คน คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของความดันที่เกิดจากการบีบตัวของหัวใจและความดันที่เกิดจากการคลายตัวของหัวใจได้ค่า r = 0.658 จงทดสอบนัยสำคัญทางสถิติของ r ที่ระดับนัยสำคัญ .05 และอธิบายความหมายที่ได้

วิธีทำ 1.1) กำหนดสมมุติฐาน

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

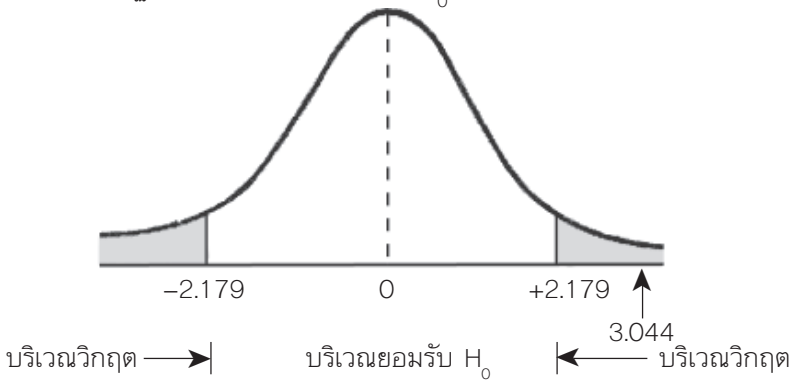
1.2) กำหนดระดับนัยสำคัญในการทดสอบเท่ากับ 0.05

1.3) คำนวณค่าสถิติทดสอบ t

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.658\sqrt{14-2}}{\sqrt{1-\left(0.658\right)^2}} = 3.044$$

1.4) เปิดค่าวิกฤตจากตาราง t ณ ระดับนัยสำคัญ $\frac{0.05}{2}$ = 0.025 เนื่องจากเป็นการทดสอบสมมติฐานสองทางและองศาแห่งความเป็นอิสระเท่ากับ n – 2 = 14 – 2 = 12 จะได้ค่า t = ± 2.179

1.5) เปรียบเทียบค่าที่คำนวณได้กับค่าวิกฤต จากภาพบริเวณวิกฤตคือส่วนที่น้อยกว่า –2.179 หรือมากกว่า 2.179 ซึ่งค่าสถิติทดสอบ ที่คำนวณได้เท่ากับ 3.044 มากกว่าค่าวิกฤต (ซึ่งเท่ากับ 2.179) ดังนั้นจึงตกอยู่ในบริเวณที่ปฏิเสธสมมติฐานว่าง นั่นคือปฏิเสธ $H_0 : \rho = 0$



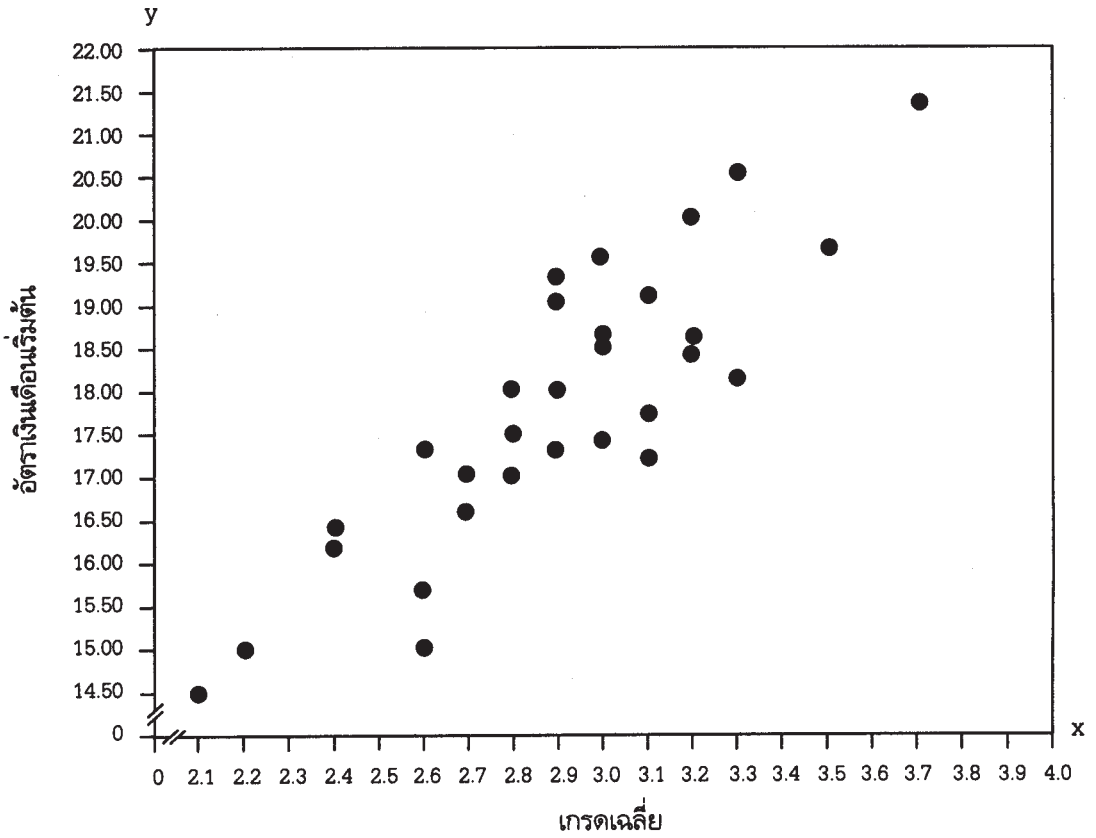
2) จากการเปรียบเทียบจะพบว่าความแตกต่างที่เกิดขึ้นมีนัยสำคัญนั่นคือ ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าแตกต่างไปจากศูนย์ ดังนั้น สามารถอธิบายได้ว่ามีหลักฐานเพียงพอที่จะสรุปว่า ความดันที่เกิดจากการบีบตัวของหัวใจและความดันที่เกิดจากการคลายตัวของหัวใจ มีความสัมพันธ์เชิงเส้นแบบทางตรงที่ระดับนัยสำคัญ 0.05

โสตทัศน์ # 14.5 การสร้างสมการถดถอยเชิงเส้นอย่างง่าย

1. การตรวจสอบความสัมพันธ์เชิงเส้น

1.1 ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเป็นการสร้างตัวแบบเชิงเส้นที่แสดงความสัมพันธ์เชิงเส้นระหว่างตัวแปรตามหนึ่งตัวและตัวแปรอิสระหนึ่งตัว จึงควรตรวจสอบความสัมพันธ์ของตัวแปรว่าเป็นเส้นตรงด้วยแผนภาพการกระจาย ก่อนนำมาวิเคราะห์

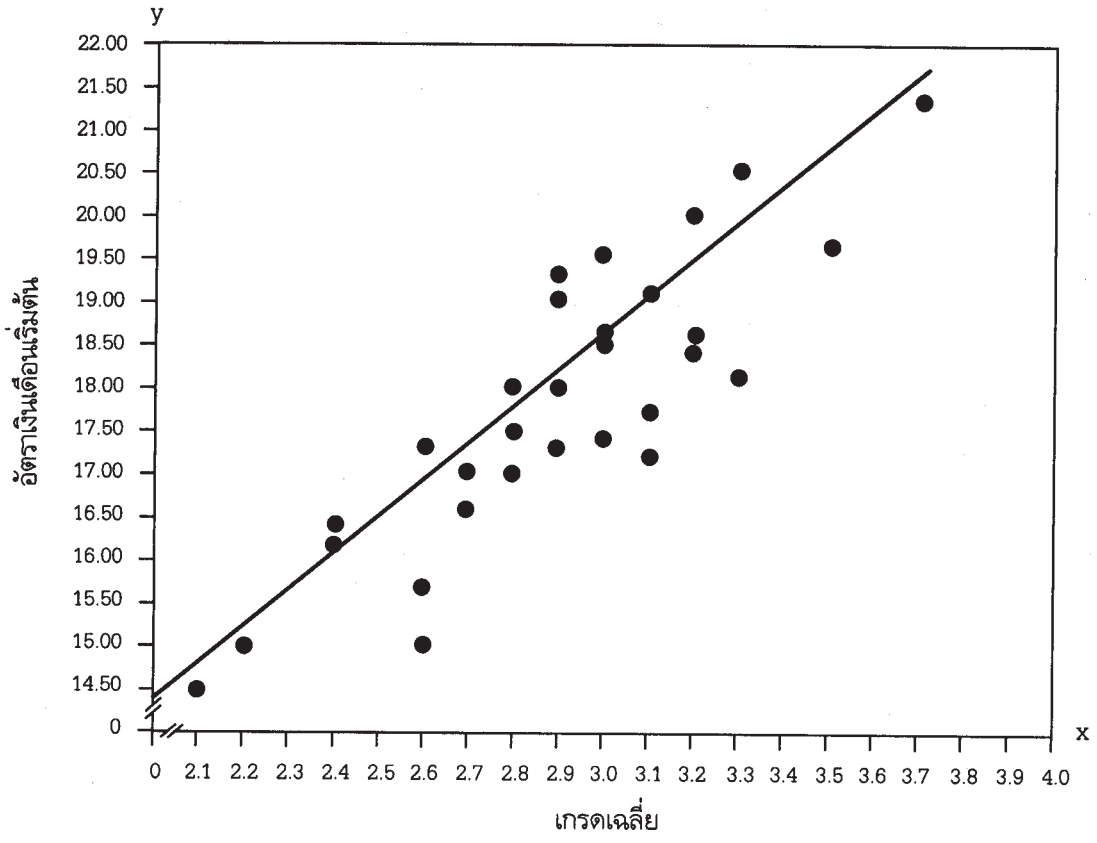
ตัวอย่าง การบรรจุพนักงานของบริษัทที่ต้องการนำข้อมูลเกรดเฉลี่ยของผู้ที่จบการศึกษาทางด้านคอมพิวเตอร์ (x) กับเงินเดือนเริ่มต้น (y) (หน่วยเป็น 1,000 บาท) จากพนักงานจำนวน 30 คน มาสร้างตัวแบบการพยากรณ์เงินเดือนเริ่มต้นจากเกรดเฉลี่ย



ภาพที่ 2 แผนภาพการกระจายของเกรดเฉลี่ยและอัตราเงินเดือนเริ่มต้นของพนักงาน

1.2 แผนภาพการกระจายข้างต้น จะทำให้เห็นลักษณะความสัมพันธ์ได้ชัดเจนว่ามีความสัมพันธ์เชิงเส้นทางบวกหรือความสัมพันธ์ทางตรง จากนั้นเลือกตัวแบบเชิงคณิตศาสตร์ระหว่างตัวแปร x และตัวแปร y ที่ง่ายที่สุดคือสมการเส้นตรงหรือความสัมพันธ์เชิงเส้นนั่นเอง ดังภาพ

โสตทัศน์ # 14.5 (ต่อ)



ภาพที่ 3 เส้นการถดถอยของเกรดเฉลี่ยและอัตราเงินเดือนเริ่มต้น

2. ตัวแบบการถดถอยเชิงเส้นอย่างง่าย

2.1 สำหรับประชากรตัวแบบเส้นตรงหรือเชิงเส้นสามารถแสดงได้ดังนี้

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

2.2 β_1 เป็นความชันของเส้นตรงที่เรียกว่า**สัมประสิทธิ์การถดถอย** แทนค่าที่เปลี่ยนแปลงใน y เมื่อค่า x มีการเปลี่ยนแปลง 1 หน่วย

2.3 β_0 เป็นส่วนตัดแกน y หรือเป็นค่าของ y เมื่อ x มีค่าเป็นศูนย์

2.4 ε เป็นความคลาดเคลื่อนเชิงสุ่มใน y ที่ประมาณด้วย e ซึ่งเท่ากับ $y - \hat{y}$ และ $\hat{y}$ คือ ค่าพยากรณ์ของ y ที่ x ค่าหนึ่ง

2.5 ในการศึกษาข้อมูลจากกลุ่มตัวอย่างเพื่ออ้างอิงความสัมพันธ์สำหรับตัวแปร x และ ตัวแปร y ในประชากรจึงประมาณค่า β_0 ด้วย b_0 และ β_1 ด้วย b_1

2.6 สมการถดถอยในตัวอย่างที่เป็นตัวแบบเชิงเส้น คือ

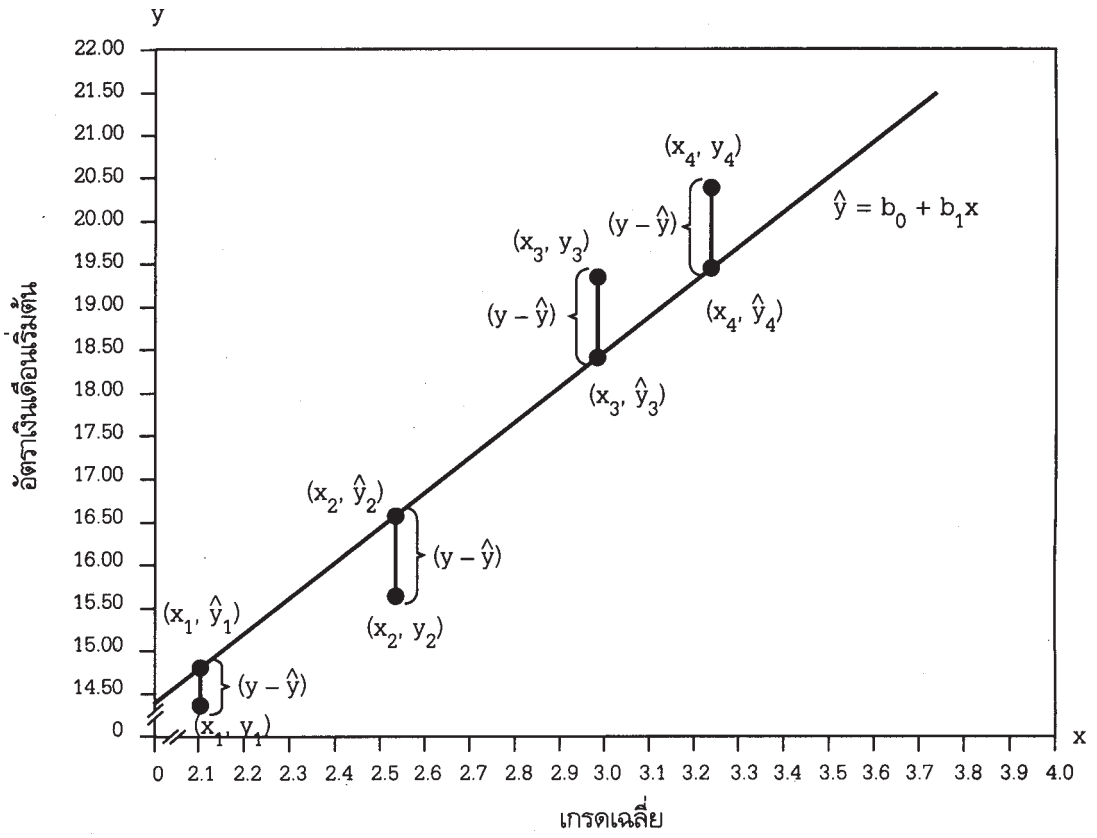
$$\hat{y} = b_0 + b_1 x$$

โสตทัศน์ # 14.5 (ต่อ)

2.7 ประมาณค่า β_0 และ β_1 โดยใช้วิธีกำลังสองน้อยที่สุด

1) **วิธีกำลังสองน้อยที่สุด** เป็นวิธีการหาตัวแบบการถดถอยที่ให้ค่าความแตกต่างระหว่างค่าจริง y กับค่าประมาณ y หรือ $\hat{y}$ ซึ่งมีค่าเท่ากับ $(y - \hat{y})$ แต่ละค่าน้อยที่สุด

2) **หลักการของวิธีกำลังสองน้อยที่สุด** คือ นำความคลาดเคลื่อนทั้งหมดมายกกำลังสองและหาผลรวมของความคลาดเคลื่อนกำลังสอง เมื่อทำให้มีค่าต่ำสุดด้วยวิธีการแคลคูลัสจะทำให้ได้สูตรคำนวณค่าประมาณ b_0 (ส่วนตัดแกน y) และค่าประมาณ b_1 (ความชัน) ของสมการเส้นตรงนั้นได้



ภาพที่ 4 ความคลาดเคลื่อนระหว่างค่าสังเกตกับค่าประมาณสำหรับแต่ละค่าของ x

โสตทัศน์ # 14.6 การประมาณค่าและทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอยเชิงเส้นอย่างง่ายและการพยากรณ์ค่าตัวแปร

1. การประมาณค่าสัมประสิทธิ์การถดถอย และส่วนตัดแกน y

1.1 $$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

หรือ $$b_0 = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

1.2 เมื่อนำค่า b_0 และ b_1 ที่คำนวณค่าได้แทนในสมการถดถอยเชิงเส้นอย่างง่าย สามารถเขียนได้ในรูป

$$\hat{y} = b_0 + b_1 x$$

1.3 b_1 จะมีเครื่องหมายเดียวกันกับ r คือ ถ้า r มีค่าเป็นบวก b_1 มีค่าเป็นบวกด้วย หากค่าของ r มีค่าเป็นลบ b_1 มีค่าเป็นลบด้วย

1.4 ทั้ง b_0 และ b_1 เป็นค่าที่แสดงถึงความสัมพันธ์ระหว่างตัวแปร x และ y

2. ความคลาดเคลื่อนมาตรฐานของการประมาณค่า

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า คือความแตกต่างระหว่างค่าสังเกตของ $\hat{y}$ กับค่าที่ได้จากการพยากรณ์ ด้วยสมการถดถอย (y) ถ้าค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีค่าน้อย แสดงว่าจุดตัวอย่างอยู่ใกล้เส้นถดถอย แต่ถ้ามีค่ามากแสดงว่าจุดตัวอย่างอยู่ห่างจากเส้นถดถอย คำนวณ โดยใช้สูตร

$$s_e = \sqrt{\frac{\sum\left(y - \hat{y}\right)^2}{n-2}}$$

หรือ $$s_e = \sqrt{\frac{\sum y^2 - b_0\sum y - b_1\sum xy}{n-2}}$$

โสตทัศน์ # 14.6 (ต่อ)

3. การทดสอบสมมติฐานของสัมประสิทธิ์การถดถอย

3.1 สมมติฐานสองทาง

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

3.2 สมมติฐานทางเดียว

$H_0 : \beta_1 = 0$ หรือ $H_0 : \beta_1 \leq 0$

$H_1 : \beta_1 > 0$ $H_1 : \beta_1 > 0$

$H_0 : \beta_1 = 0$ หรือ $H_0 : \beta_1 \geq 0$

$H_1 : \beta_1 < 0$ $H_1 : \beta_1 < 0$

3.3 ใช้สถิติทดสอบ t

$$t = \frac{b_1}{s_{b_1}}$$

ที่องศาแห่งความเป็นอิสระ df เท่ากับ n – 2

3.4 $$s_{b_1} = \frac{s_e}{\sqrt{ss_x}}$$

3.5 $$ss_x = \sum \left(x - \bar{x} \right)^2$$

$$= \sum x^2 - \frac{\left(\sum x \right)^2}{n}$$

3.6 กำหนดระดับนัยสำคัญ ของการทดสอบ

3.7 ค่าวิกฤตของการแจกแจง t สามารถหาค่า ณ องศาแห่งความเป็นอิสระ n – 2 และระดับนัยสำคัญ

ตามที่กำหนดหารด้วย 2 หรือ $\frac{\alpha}{2}$ ในกรณีการทดสอบสมมติฐานสองทาง

3.8 เปรียบเทียบค่าสถิติทดสอบกับค่าวิกฤต จะปฏิเสธสมมติฐาน ถ้าค่าสถิติทดสอบ t ที่คำนวณได้มากกว่าค่าวิกฤต (ค่าวิกฤตทางขวา) หรือถ้าค่าสถิติทดสอบ t ที่คำนวณได้น้อยกว่าค่าวิกฤต (ค่าวิกฤตทางซ้าย)

4. การใช้สมการถดถอยสำหรับการพยากรณ์

4.1 การใช้สมการถดถอยสำหรับพยากรณ์ค่าตัวแปรตาม เมื่อทราบค่าตัวแปรอิสระ

4.2 การใช้สมการถดถอยสำหรับพยากรณ์ควรใช้เฉพาะกับชุดข้อมูลที่ค่าสัมประสิทธิ์สหสัมพันธ์เชิงเส้นอย่างง่าย (r) มีนัยสำคัญเท่ากัน

โสตทัศน์ # 14.6 (ต่อ)

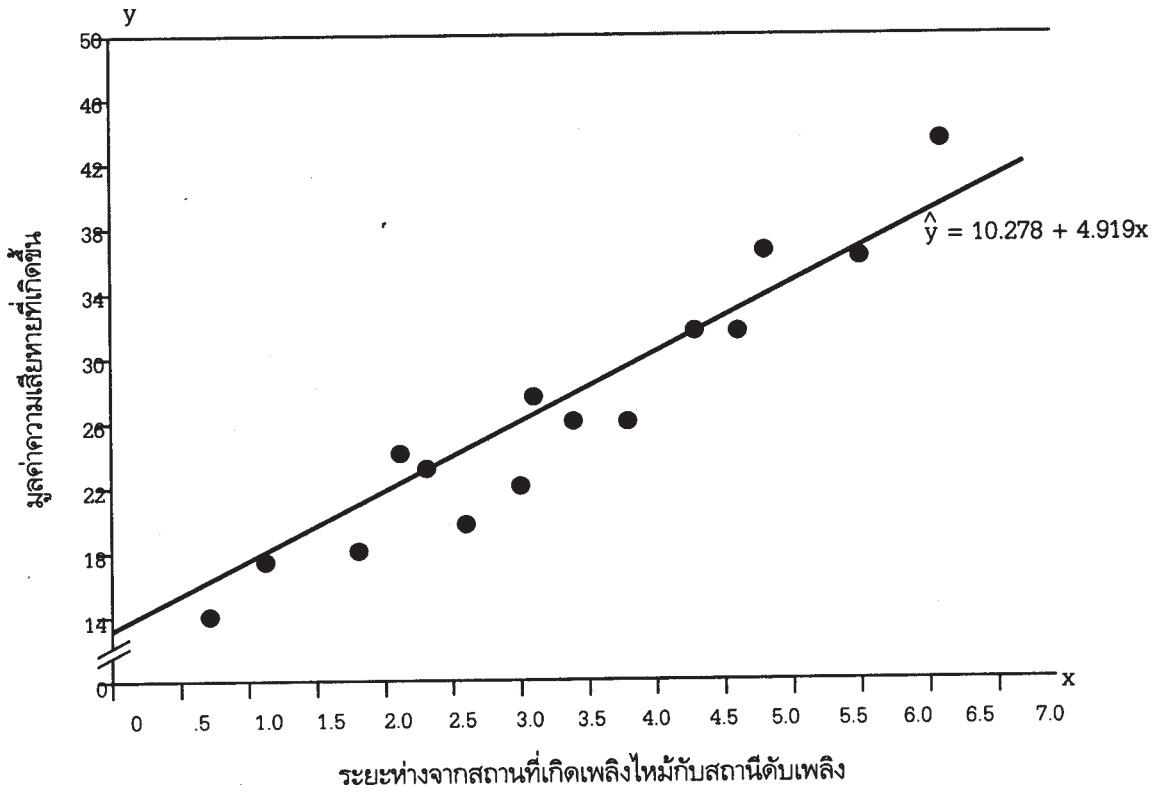
5. ข้อควรระวังสำหรับการใช้สมการถดถอย

1. ถ้าความสัมพันธ์ไม่ใช่เชิงเส้น จะไม่ใช้สมการถดถอยเชิงเส้นในการพยากรณ์
2. เมื่อใช้สมการถดถอยสำหรับการพยากรณ์จะต้องอยู่ภายในขอบเขตของข้อมูลกลุ่มตัวอย่างที่หาได้
3. สมการถดถอยที่สร้างจากข้อมูลในอดีตอาจจะให้ผลที่ไม่ถูกต้องนัก หากนำมาใช้ในปัจจุบัน
4. ต้องไม่นำสมการถดถอยไปพยากรณ์ในประชากรที่แตกต่างจากข้อมูลของกลุ่มตัวอย่างที่สุ่มมา

ตัวอย่าง จากข้อมูลในการวิเคราะห์ความสัมพันธ์ระหว่างระยะห่างของสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิง และมูลค่าความเสียหายที่เกิดขึ้น 15 แห่ง สามารถคำนวณค่าต่าง ๆ ได้ดังนี้ $b_0 = 10.278$, $b_1 = 4.919$, $\Sigma x = 49.20$, $\Sigma y = 396.20$, $\Sigma xy = 1{,}470.65$, $\Sigma x^2 = 196.16$, $\Sigma y^2 = 11.376.48$, $\bar{x} = 6.33$, $\bar{y} = 51.08$ และ $S_e = 2.324$

1. จงหาสมการถดถอยที่ใช้ในการพยากรณ์มูลค่าความเสียหาย และอธิบายความหมายของค่าคงที่และค่าสัมประสิทธิ์การถดถอย
2. จงทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอยเชิงเส้นที่ระดับนัยสำคัญ 0.01 และอธิบายความหมายของผลการทดสอบที่ได้
3. ให้พยากรณ์มูลค่าความเสียหายที่เกิดขึ้นเมื่อสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิงอยู่ห่างกัน 4 ไมล์

วิธีทำ 1.1 สมการถดถอยเชิงเส้นอย่างง่าย คือ $\hat{y} = 10.278 + 4.919x$



โสตทัศน์ # 14.5 (ต่อ)

1.2 ค่าคงที่ b_0 ในสมการมีค่าเท่ากับ 10.278 (หน่วย: พันดอลลาร์) หมายความว่าถ้าระยะห่างระหว่างสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิง (x) มีค่าเป็น 0 หรือกล่าวได้ว่าสถานที่ที่เกิดเพลิงไหม้อยู่ติดกับสถานีดับเพลิง มูลค่าความเสียหายที่เกิดขึ้นจะมีค่าเท่ากับ 10.278 พันดอลลาร์ หรือเท่ากับ 10,278 ดอลลาร์

2.1 ค่าสัมประสิทธิ์การถดถอย b_1 มีค่าเท่ากับ 4.919 หมายความว่าถ้าระยะห่างระหว่างสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิงเปลี่ยนแปลงไป 1 ไมล์ มูลค่าความเสียหายที่เกิดขึ้นจะเปลี่ยนแปลงไปโดยเฉลี่ยเท่ากับ 4.919 พันดอลลาร์ หรือ 4,919 ดอลลาร์ และเป็นการเปลี่ยนแปลงไปในทิศทางเดียวกัน หรือกล่าวได้ว่าถ้าระยะห่างระหว่างสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิงเพิ่มขึ้น 1 ไมล์ มูลค่าความเสียหายจะเพิ่มขึ้น 4.919 พันดอลลาร์

2.1.1 กำหนดสมมติฐาน

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

2.1.2 กำหนดระดับนัยสำคัญในการทดสอบเท่ากับ 0.01

2.1.3 คำนวณค่า S_{b_1}

$$S_{b_1} = \frac{S_e}{\sqrt{SS_x}}$$

โดยที่ $$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 196.16 - \frac{\left(49.20\right)^2}{15} = 34.784$$

ดังนั้น $$S_{b_1} = \frac{2.324}{\sqrt{34.784}} = 0.393$$

2.1.4 คำนวณค่าสถิติทดสอบ

$$t = \frac{b_1}{S_{b_1}} = \frac{4.919}{0.393} = 12.525$$

2.1.5 เปิดค่าวิกฤตจากตาราง t ณ ระดับนัยสำคัญที่ $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$ องศาแห่งความเป็นอิสระเท่ากับ $n - 2 = 15 - 2 = 13$ จะได้ $t = \pm 3.012$

2.1.6 เปรียบเทียบค่าที่คำนวณได้กับค่าวิกฤต บริเวณวิกฤตคือ ค่าที่น้อยกว่า –3.012 หรือ ค่ามากกว่า 3.12 ซึ่งค่าที่คำนวณได้มากกว่า 3.012 จึงอยู่ในบริเวณที่ปฏิเสธสมมติฐานว่าง นั่นคือ ปฏิเสธ H_0

2.2 ดังนั้น สามารถอธิบายได้ว่ามีหลักฐานเพียงพอที่จะสรุปว่าระยะห่างจากสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิงมีผลต่อมูลค่าความเสียหายที่เกิดขึ้นที่ระดับนัยสำคัญ 0.01

3. มูลค่าความเสียหายที่เกิดขึ้นเมื่อสถานที่ที่เกิดเพลิงไหม้กับสถานีดับเพลิงห่างกัน 4 ไมล์

$\hat{y} = 10.278 + 4.919\ (4)$

$= 29.954$ พันดอลลาร์

โสตทัศน์ # 14.6 สัมประสิทธิ์แห่งการกำหนด

สัมประสิทธิ์แห่งการกำหนด (r^2) คือ สัดส่วนความแปรปรวนที่อธิบายได้ด้วยเส้นถดถอยต่อความแปรปรวนทั้งหมด

$$r^2 = \frac{\text{ความแปรปรวนที่อธิบายได้ด้วยเส้นถดถอย}}{\text{ความแปรปรวนทั้งหมด}} = \frac{\sum\left(\hat{y} - \bar{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}$$

หรือ $$r^2 = 1 - \frac{\sum\left(y - \hat{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}$$

ผลรวมของส่วนเบี่ยงเบนที่ไม่สามารถอธิบายได้ด้วยเส้นถดถอย ก็คือผลรวมกำลังสองของความคลาดเคลื่อนในการประมาณค่า (residual) ถ้ามีค่าต่ำ r^2 ก็จะมีค่าสูง

ค่าเบี่ยงเบนที่อธิบายได้ (explained deviation) ของจุด (x, y) คือค่าความแตกต่างของค่าประมาณ y หรือค่า $\hat{y}$ ค่าหนึ่งกับค่าเฉลี่ยตัวอย่าง y (หรือค่า $\bar{y}$) ณ x ค่าหนึ่ง ดังนั้นค่าเบี่ยงเบนที่อธิบายได้คือ $\hat{y} - y$

ค่าเบี่ยงเบนที่อธิบายไม่ได้ (unexplained deviation) ของจุด (x, y) คือค่าความแตกต่างของค่าสังเกต หรือค่า y ค่าหนึ่งกับ ค่าประมาณของ y (หรือค่า $\hat{y}$) ณ x ค่าหนึ่ง ดังนั้นค่าเบี่ยงเบนที่อธิบายไม่ได้คือ $y - \hat{y}$ หรือเรียกว่า residual หรือความคลาดเคลื่อนมาตรฐานของการประมาณค่าตัวแปรตาม

ค่าเบี่ยงเบนรวม (total deviation) ของจุด (x, y) คือค่าความแตกต่างของค่าสังเกต หรือค่า y ค่าหนึ่งกับค่าเฉลี่ยของ y (หรือค่า $\bar{y}$) ณ x ค่าหนึ่ง ดังนั้น ค่าเบี่ยงเบนรวมคือ $y - \bar{y}$

ตัวอย่าง จากข้อมูลการศึกษาความสัมพันธ์ระหว่างปริมาณแคลอรี (หน่วย : กรัมของผลิตภัณฑ์ธัญพืช) กับจำนวนกรัมของไขมัน (หน่วย : กรัมของผลิตภัณฑ์ธัญพืช) จากธัญพืช 16 ชนิด คำนวณค่า r ได้เท่ากับ 0.359 จงหาค่าสัมประสิทธิ์แห่งการกำหนดและอธิบายความหมายของค่าสัมประสิทธิ์

วิธีทำ $r = 0.359$ ดังนั้น $r^2 = 0.129$ หรือ 12.9%

หมายความว่าร้อยละ 12.9 ของความแปรปรวนทั้งหมดใน y สามารถอธิบายได้ด้วยเส้นถดถอย นั่นคือ ร้อยละ 87.1 ของความแปรปรวนใน y ไม่สามารถอธิบายได้ด้วยเส้นถดถอย

โสตทัศน์ # 14.6 (ต่อ)

กิจกรรม

1. จากข้อมูลในตารางให้สร้างแผนภาพการกระจายและอธิบายลักษณะของความสัมพันธ์

อุณหภูมิร่างกาย (x) (องศาฟาเรนไฮต์)	ระยะเวลาที่ใช้ในการวิ่งออกกำลังกาย (y) (นาที)
55	145.3
61	148.7
49	148.3
62	148.1
70	147.6
73	146.4
51	144.7
57	147.5

2. จากข้อมูลจำนวนพนักงานของบริษัท (หน่วย : คน) กับยอดขายของบริษัท (หน่วย : 10 ล้านบาท) ซึ่งเก็บข้อมูลรายละเอียดย้อนหลัง 10 ปี เมื่อนำมาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (r) ได้เท่ากับ 0.99 จงอธิบายความหมายของสัมประสิทธิ์สหสัมพันธ์ และทดสอบว่าจำนวนพนักงานขายของบริษัทผู้ผลิตฮาร์ดแวร์มีความสัมพันธ์กับยอดขายต่อปีที่ระดับนัยสำคัญ 0.05 พร้อมทั้งอธิบายความหมายของผลการทดสอบที่ได้

3. วิธีกำลังสองน้อยที่สุดที่นำมาใช้ในการประมาณค่าส่วนตัดแกน y และความชันในตัวแบบถดถอยเชิงเส้นอย่างง่ายมีหลักการอย่างไร

4. จากสมการถดถอยเชิงเส้นอย่างง่ายต่อไปนี้

$$\hat{y} = 1.40 + 0.1010x$$

เมื่อ x คือ จำนวนเรคอร์ดของข้อมูลที่จัดเก็บ (หน่วย : พันเรคอร์ด)

y คือ จำนวนของเนื้อที่บนเทปแม่เหล็กที่ใช้ (หน่วยเมกะไบต์)

จงอธิบายความหมายของค่าคงที่และค่าสัมประสิทธิ์การถดถอยที่ได้

โสตทัศน์ # 14.6 (ต่อ)

5. บริษัทผลิตฉนวนป้องกันไฟฟ้ารั่วแห่งหนึ่ง กำลังทดลองเพื่อพัฒนาฉนวนกันไฟฟ้ารั่วชนิดใหม่โดยศึกษาความสัมพันธ์ระหว่างแรงดันที่ต้องใช้ในการอัดวัสดุที่นำมาทำฉนวน (หน่วย : 10 ปอนด์ต่อ ตารางนิ้ว) กับขนาดความหนาของฉนวน (หน่วยเป็น 0.1 นิ้ว) ซึ่งจากการทดลองได้เก็บบันทึกข้อมูลดังนี้

การทดลองครั้งที่	แรงดันที่ใช้อัดวัสดุ (x)	ความหนาของฉนวน (y)
1	1	1
2	2	1
3	3	2
4	4	2
5	5	4

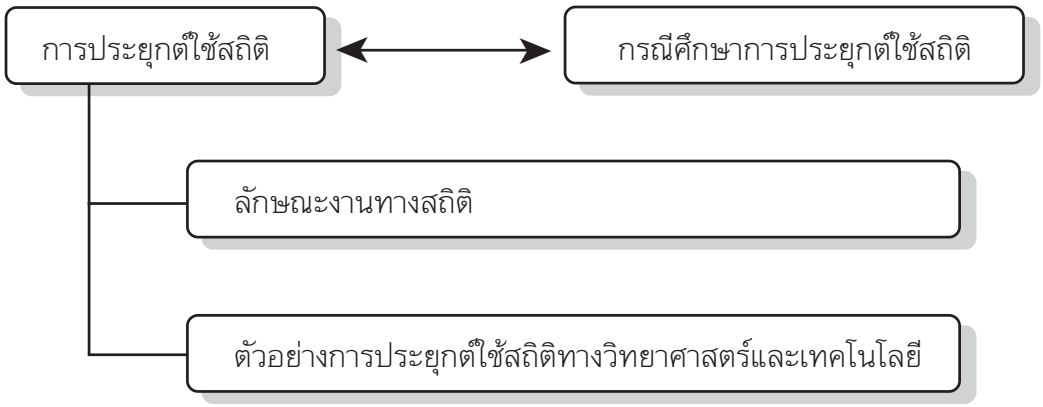
5.1 จงสร้างสมการถดถอยเชิงเส้นอย่างง่ายและอธิบายความหมายของค่า b_0 และ b_1

5.2 จงทดสอบนัยสำคัญ ของค่าสัมประสิทธิ์การถดถอยที่ $\alpha = 0.01$ และสรุปผลการทดสอบที่ได้

6. สัมประสิทธิ์แห่งการกำหนดหมายถึงอะไร

7. จากคำถามท้ายประเด็นข้อ 2 คำนวณค่า r^2 ได้ = 0.98 หมายความว่าอย่างไร

หน่วยที่ 15
การประยุกต์สถิติทางวิทยาศาสตร์และเทคโนโลยี



โสตทัศน์ # 15.1 ลักษณะงานทางสถิติ

การชักตัวอย่าง (sampling) เป็นสิ่งที่สำคัญสำหรับการใช้สถิติเพื่ออ้างอิง

เทคนิคทางสถิติที่เรียกว่า **"การชักตัวอย่าง"** นั้นมีประโยชน์อย่างยิ่งในการหาข้อมูลเกี่ยวกับประชากรของกลุ่มประชากรจริงๆ หรือกลุ่มของสิ่งต่างๆ ซึ่งทำโดยใช้การสำรวจเพียงส่วนหนึ่งของประชากรนั้นๆ

ลักษณะงานทางสถิติ

ลักษณะของงานทางสถิติเพื่อการอ้างอิงเกี่ยวข้องกับการกำหนดวิธีการที่ใช้ชักตัวอย่างและการเก็บรวบรวมข้อมูล การประมวลผลข้อมูล การระบุถึงความเชื่อถือได้ของผลลัพธ์ที่ได้ รวมทั้งข้อจำกัดต่างๆ ในการใช้ผลลัพธ์นั้นๆ ผู้ใช้สถิติจะต้องเข้าใจในเรื่องของความไม่แน่นอนและสามารถเชื่อมโยงผลสรุปที่ได้ให้เกี่ยวข้องกับเรื่องเฉพาะหนึ่งๆ ที่กำลังใช้สถิติอยู่

ตัวอย่างของลักษณะงานทางสถิติ

1. **งานสำรวจทั่วๆ ไป :** เป็นการรวบรวมข้อสนเทศโดยอาศัยข้อมูลจากตัวอย่างโดยจะนำผลลัพธ์ที่ได้จากตัวอย่างขยายผลไปยังประชากร อาจเป็นส่วนหนึ่งของการวิจัยหรือส่วนหนึ่งของการบริหารจัดการก็ได้
2. **งานเพื่อการดำเนินการของรัฐบาล :** เป็นการศึกษาในเรื่องต่างๆ เพื่อใช้ในการพัฒนาประเทศโดยอาศัยสถิติเป็นเครื่องมือช่วยในการกำหนดนโยบายต่างๆ ของรัฐ รวมทั้งโครงการบริการสังคมต่างๆ
3. **งานวิจัย :** งานวิจัยต้องอาศัยสถิติเพื่อเป็นเครื่องมือในการประเมินความสมเหตุสมผลของการสรุปผลและ การอ้างอิง
4. **งานด้านธุรกิจและอุตสาหกรรม :** เป็นการศึกษาเพื่อกำหนดใช้ทรัพยากรที่มีอยู่ให้เกิดประโยชน์สูงสุด การพยากรณ์ การตรวจสอบคุณภาพของสินค้าหรือบริการ และการวิจัยตลาด

โสตทัศน์ # 15.2 ตัวอย่างการประยุกต์ใช้สถิติทางวิทยาศาสตร์และเทคโนโลยี

การใช้สถิติในงานด้านวิทยาศาสตร์และเทคโนโลยีมักเกี่ยวข้องกับการวิจัยต่างๆ ของนักวิทยาศาสตร์ เช่น

1. ชีวสถิติศาสตร์ (Biostatistics) หรือชีวมิติ (Biometric)

เป็นการประยุกต์ใช้เทคนิคทางสถิติในงานวิจัยหรือศึกษาปรากฏการณ์ทางวิทยาศาสตร์ต่างๆ ที่เกี่ยวข้องกับสุขภาพ เช่น การแพทย์ ระบาดวิทยาและสาธารณสุข รวมทั้งชีววิทยา โดยทำการออกแบบการทดลอง กำหนดวิธีการชักตัวอย่าง การรวบรวมข้อมูล และการวิเคราะห์ทางสถิติเพื่อตอบคำถามทางชีววิทยาในลักษณะต่างๆ

2. เคมี (Chemistry)

การประยุกต์ใช้สถิติในทางเคมีอาจแบ่งเป็นสองส่วน ได้แก่การใช้สถิติในเรื่องของการวัดค่า ความแม่นและความเที่ยงในการวัดต่างๆ ในห้องปฏิบัติการ และการใช้สถิติเพื่อวิเคราะห์ข้อมูลที่เกี่ยวข้องกับสารเคมีหรือสร้างตัวแบบทางคณิตศาสตร์เพื่ออธิบายความสัมพันธ์ต่างๆ ทางเคมีหรือทำนายสมบัติของสาร

โสตทัศน์ # 15.2 (ต่อ)

3. ศาสตร์คอมพิวเตอร์ และสารสนเทศศาสตร์

สถิติมีบทบาทในส่วนของการออกแบบการทดลองและวิเคราะห์ข้อมูลเพื่อประเมินผลการทดลองทั้งด้านฮาร์ดแวร์และซอฟต์แวร์ นักวิเคราะห์ระบบ (system analysis) อาศัยสถิติเพื่อเปรียบเทียบระบบงานคอมพิวเตอร์ นอกจากนี้ยังใช้สถิติในงานอื่นๆ ที่เกี่ยวข้องกับคอมพิวเตอร์ เช่น การทำเหมืองข้อมูลหรือดาตาไมนิง (data mining) การวิเคราะห์เชิงภาพ (vision and image analyses) และ ปัญญาประดิษฐ์ (artificial intelligence) เป็นต้น

4. วิศวกรรม (Engineering)

เนื่องจากงานหลายประเภทต้องใช้วิศวกร และอุตสาหกรรมแต่ละประเภทมีการแข่งขันกันในด้านมาตรฐาน ระยะเวลาในการพัฒนา และการรับประกันในเรื่องความปลอดภัยของผลิตภัณฑ์ ดังนั้นการใช้สถิติจึงได้แก่การเสนอวิธีการที่ลดค่าใช้จ่ายและเพิ่มคุณภาพในกระบวนการผลิตในขั้นตอนตั้งแต่การออกแบบ การผลิต การบำรุงรักษา และการจัดการด้านการพาณิชย์

5. กระบวนการผลิต (Manufacturing) และการปรับปรุงคุณภาพ (Quality Improvement)

บทบาทของสถิติในงานอุตสาหกรรมได้แก่การมีส่วนในการผลิตสินค้าหรือบริการเพื่อให้ลูกค้าพึงพอใจและเพิ่มส่วนแบ่งทางการตลาดรวมทั้งผลกำไร ทั้งนี้ตั้งแต่ขั้นตอนการออกแบบผลิตภัณฑ์ การดำเนินการผลิต การรับประกันคุณภาพและ การบำรุงรักษา

6. ความเชื่อถือได้ (Reliability)

ความเชื่อถือได้ หมายถึง ความสามารถของสินค้า บริการหรือระบบที่จะทำงานได้ตามที่กำหนดในสภาพแวดล้อมและช่วงเวลาหนึ่งๆ ซึ่งสามารถใช้วิธีการทางสถิติในการวัดวัดความเชื่อถือได้นี้โดยการทำการคาดหมาย ทดลอง และประเมินผล

7. การประเมินความเสี่ยง (Risk Assessment)

การประเมินความเสี่ยงใช้ศาสตร์หลายแขนงรวมทั้งสถิติในการระบุอันตราย (hazard identification) การประเมินความสัมพันธ์ระหว่างผลตอบสนองและขนาดสัมผัส (dose–response assessment) การประเมินการสัมผัสหรือการอุบัติขึ้น (exposure assessment) และการกำหนดค่าความเสี่ยงหรือการคำนวณความเสี่ยง (risk characterization)

โสตทัศน์ # 15.3 กรณีศึกษา : การเปรียบเทียบสูตรการผสมสีสำหรับผลิตม่านปรับแสงอลูมิเนียม

การเปรียบเทียบสูตรการผสมสีสำหรับผลิตม่านปรับแสงอลูมิเนียมเป็นการประยุกตสถิติในด้านกระบวนการผลิตเพื่อปรับปรุงคุณภาพผลิตภัณฑ์

สถิติหลักที่ใช้ได้แก่

- การออกแบบการทดลอง
- การเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่สองกลุ่มขึ้นไปด้วยสถิติเอฟ
- การศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวด้วยการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

โสตทัศน์ # 15.3 (ต่อ)

วัตถุประสงค์ของการศึกษา คือการหาสูตรผสมสีที่ผลิตม่านปรับแสงอลูมิเนียมได้ตามมาตรฐานที่กำหนด โดยมีสมมุติฐานของการวิจัย คือเมื่อเคลือบแผ่นอลูมิเนียมด้วยสีที่ผสมตามสูตร 12 สูตรแล้วค่าวัดต่างๆ ที่แสดงสมบัติทางกายภาพและสมบัติทางเคมีของสีจะแตกต่างกัน

ในที่นี้พารามิเตอร์ได้แก่ค่าเฉลี่ยประชากรของตัวแปรแต่ละตัว ตัวอย่างของสมมติฐานทางสถิติคือ

H_0 : ความแข็งของสีโดยเฉลี่ยในแต่ละสูตร (12 สูตร) ไม่แตกต่างกัน

($\mu_1 = \mu_1 = \mu_3 = ...\mu_{12}$)

H_1 : ไม่จริงที่ว่าความแข็งของสีโดยเฉลี่ยของแต่ละสูตรไม่แตกต่างกัน

(ไม่จริงที่ว่า $\mu_1 = \mu_1 = \mu_3 = ...\mu_{12}$)

หรือ

H_0 : ความทนต่อแรงขีดข่วนของสีโดยเฉลี่ยในแต่ละสูตร (12 สูตร)ไม่แตกต่างกัน

($\mu_1 = \mu_1 = \mu_3 = ...\mu_{12}$)

H_1: ไม่จริงที่ว่าความทนต่อแรงขีดข่วนของสีโดยเฉลี่ยของแต่ละสูตรไม่แตกต่างกัน

(ไม่จริงที่ว่า $\mu_1 = \mu_1 = \mu_3 = ...\mu_{12}$)

ให้สังเกตว่าสีแต่ละสูตรจัดว่าเป็นกลุ่มประชากรกลุ่มหนึ่งที่เป็นอิสระต่อกัน ดังนั้นจึงมีทั้งสิ้น 12 ประชากร

การทดลองทำโดยการแบ่งแผ่นอลูมิเนียมเป็นชิ้นๆ จำนวน 120 ชิ้น ทำการสุ่มแผ่นอลูมิเนียมเพื่อเคลือบสีด้วยสูตรทั้ง 12 สูตรโดยใช้สูตรละ 10 ชิ้น (ใช้ขนาดตัวอย่างเท่ากับ 10 หน่วยในแต่ละประชากร)

ค่าที่วัดมาศึกษาได้แก่ค่าที่แสดงถึงสมบัติทางกายภาพและสมบัติทางเคมี คำนวณค่าเฉลี่ยตัวอย่างจากแต่ละกลุ่ม โดยสมมุติว่าข้อมูลแต่ละกลุ่มเป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบปกติ และมีความแปรปรวนของแต่ละประชากรเท่ากัน จากนั้นใช้วิธีการทางสถิติที่ใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยประชากรหลายกลุ่ม ที่เรียกว่า “การวิเคราะห์ความแปรปรวน ” สถิติที่ใช้คือสถิติเอฟ ส่วนการศึกษาความสัมพันธ์ระหว่างอัตราส่วนของกาวชนิดที่ 1 กับสมบัติต่างๆ ของสี ใช้การวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย เนื่องจากเป็นการศึกษาความสัมพันธ์ระหว่างข้อมูลเชิงปริมาณทั้งคู่

โสตทัศน์ # 15.4 กรณีศึกษา : การปนเปื้อนของเชื้อซัลโมเนลลาในอาหารสัตว์และการควบคุม

การปนเปื้อนของเชื้อซัลโมเนลลาในอาหารสัตว์และการควบคุม เป็นการประยุกต์สถิติในด้านวิทยาศาสตร์และเทคโนโลยีการอาหารหรือสาธารณสุขศาสตร์

สถิติหลักที่ใช้ ได้แก่การเปรียบเทียบความแตกต่างของสัดส่วนหลายประชากรด้วยสถิติไคสแควร์

วัตถุประสงค์หนึ่งของการศึกษาได้แก่การสำรวจแหล่งที่มาของวัตถุดิบต่างๆ ที่นำมาผสมเป็นอาหารไก่ โดยมีสมมติฐานหนึ่งของการวิจัยได้แก่สัดส่วนของเชื้อซัลโมเนลลาที่ตรวจพบในอาหารสัตว์ของโรงงานแต่ละโรงนั้นแตกต่างกัน เนื่องจากพารามิเตอร์ที่สนใจศึกษาได้แก่สัดส่วน ของเชื้อซัลโมเนลลาที่ตรวจพบในอาหารสัตว์ของแต่ละโรงงาน และมีโรงงานทั้งสิ้น 12 โรงงานที่เป็นอิสระกัน นั่นคือมีจำนวนประชากรทั้งสิ้น 12 ประชากร ทำให้ได้สมมติฐานทางสถิติดังนี้

H_0 : สัดส่วนของเชื้อซัลโมเนลลาที่ตรวจพบในอาหารสัตว์ของแต่ละโรงงาน (12 โรงงาน) ไม่แตกต่างกัน

($P_1 = P_1 = P_3 = \ldots P_{12}$)

H_1 : ไม่จริงที่ว่าสัดส่วนของเชื้อซัลโมเนลลาที่ตรวจพบในอาหารสัตว์ของแต่ละโรงงาน (12 โรงงาน) ไม่แตกต่างกัน

(ไม่จริงที่ว่า $P_1 = P_1 = P_3 = \ldots P_{12}$)

ในที่นี้ P_i แทนสัดส่วนประชากรของเชื้อซัลโมเนลลาที่ตรวจพบในอาหารสัตว์ของโรงงานที่ i (i=1, 2, 3,..., 12)

การเก็บข้อมูลทำโดยการสุ่มเก็บตัวอย่างวัตถุดิบต่างๆ ที่ใช้ผสมเป็นอาหารไก่ และอาหารไก่สำเร็จรูปจากโรงงานผลิตอาหารสัตว์จำนวน 149 ตัวอย่าง (ขนาดตัวอย่างรวมจาก 12 โรงงาน) จากนั้นบันทึกการพบหรือไม่พบเชื้อซัลโมเนลลาในตัวอย่างของโรงงานทั้ง 12 แห่ง ทำการคำนวณสัดส่วนตัวอย่างแล้วเปรียบเทียบความแตกต่างของสัดส่วนของตัวอย่างที่พบเชื้อซัลโมเนลลาของแต่ละโรงงานด้วยสถิติไคสแควร์ เนื่องจากข้อมูลเป็นจำนวนนับและสิ่งที่สนใจศึกษาได้แก่สัดส่วนประชากรที่พบเชื้อซัลโมเนลลา นอกจากนี้ การพบเชื้อหรือไม่พบเชื้อซัลโมเนลลาของแต่ละโรงงานสมมติได้ว่าเป็นอิสระต่อกัน