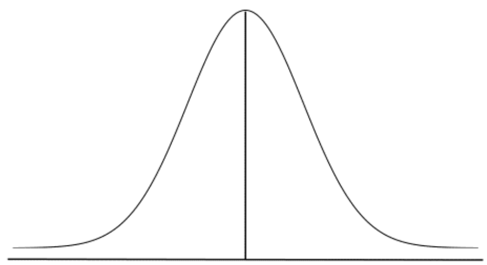
การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่องที่สำคัญ เนื่องจากการแจกแจงแบบปกติใช้ในการประมาณปรากฏการณ์ทางธรรมชาติหลายอย่าง โดยการแจกแจงแบบปกติปกติจะมีลักษณะของการแจกแจงคล้ายกับโค้งรูประฆังคว่ำ (Bell Shape) เส้นโค้งดังกล่าวของการแจกแจงแบบปกติ จะเรียกว่า โค้งปกติ (Normal Curve)



รูป แสดงการแจกแจงแบบปกติ

กำหนดให้ X มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 นั่นคือฟังก์ชันการแจกแจงแบบปกติคือ

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

คุณสมบัติของเส้นโค้งปกติ

1. เส้นโค้งปกติเป็นรูประฆังคว่ำ ข้อมูลส่วนใหญ่มีค่ากลาง ๆ มีลักษณะสมมาตรที่จุดกึ่งกลาง คือค่าเฉลี่ย ปลายของเส้นโค้งค่อย ๆ ลาดเข้าสู่แกนนอน แต่จะไม่ตัดแกนนอน มีระยะตั้งแต่ $-\infty$ ถึง ∞

2. ค่าของตัวแปรสุ่มมีค่าอยู่ระหว่าง $-\infty$ ถึง ∞

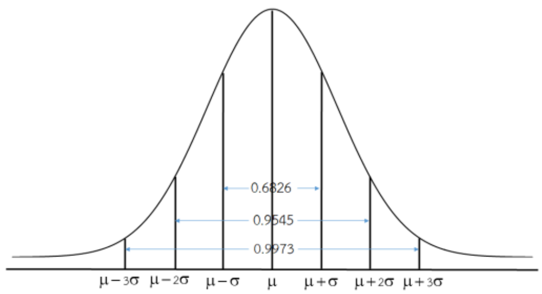
3. พื้นที่ใต้เส้นโค้งปกติ หมายถึง ความน่าจะเป็นทั้งหมดมีค่าเท่ากับ 1 เนื่องจากเส้นโค้งปกติมีลักษณะสมมาตรที่ค่าเฉลี่ย เพราะฉะนั้นพื้นที่ใต้เส้นโค้งตั้งแต่ค่าเฉลี่ยถึงค่าทางขวามือ เท่ากับพื้นที่ตั้งแต่ทางว้ายมือถึงค่าเฉลี่ยมีค่าเท่ากับ 0.5

4. เส้นโค้งปกติมีค่าสูงที่สุดที่ค่าเฉลี่ยและมีจุดเปลี่ยนเว้า (Inflectional Point) ที่ $\mu \pm \sigma$ โดยมีพื้นที่ใต้เส้นโค้งดังนี้

พื้นที่ระหว่าง $\mu - \sigma$ ถึง $\mu + \sigma$ เท่ากับ 0.6826 หรือ 68.26%

พื้นที่ระหว่าง $\mu - 2\sigma$ ถึง $\mu + 2\sigma$ เท่ากับ 0.9545 หรือ 95.45%

พื้นที่ระหว่าง $\mu - 3\sigma$ ถึง $\mu + 3\sigma$ เท่ากับ 0.9973 หรือ 99.73%

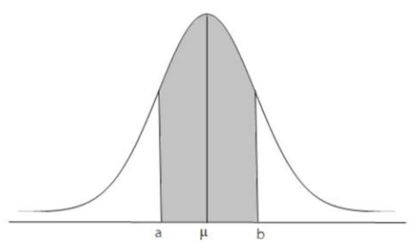


3.7.1 การหาพื้นที่ใต้โค้ง

การหาความน่าจะเป็นของการแจกแจงแบบปกติ หมายถึงการหาพื้นที่ใต้โค้งปกติ โดยที่ พื้นที่ใต้เส้นโค้งปกติทั้งหมดมีค่าเท่ากับ 1 นั่นคือ

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

ถ้ากำหนดให้ x มีการแจกแจงแบบปกติ การหาความน่าจะเป็นของ x ที่มีค่าระหว่างค่า a ถึงค่า b แทนด้วย $P(a < x < b)$ ดังรูป



โดยที่ $P(a < x < b) = \int_{a}^{b} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

การหาความน่าจะเป็นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติจะต้องใช้การอินทิกรัลเข้ามาช่วยในการหาพื้นที่ซึ่งค้อนข้างยุ่งยาก จึงมีการสร้างการแจกแจงเพื่อช่วยในการหาความน่าจะเป็นให้ง่ายขึ้น คือ การแปลงการแจกแจงแบบปกติให้อยู่ในรูปของการแจกแจงแบบปกติมาตรฐาน

3.7.2. การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution)

การแจกแจงแบบปกติมาตรฐาน อาจเรียกแทนด้วยการแจกแจงแบบ Z คือการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ 0 และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 แทนด้วย $Z \sim N(0,1)$ การแปลงตัวแปรสุ่ม X ที่มีการแจกแจงแบบปกติ เป็นการแจกแจงแบบปกติมาตรฐาน สามารถทำได้โดย

$$Z = \frac{x-\mu}{\sigma}$$

เขียนฟังก์ชันความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐานได้ คือ

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad \text{เมื่อ } -\infty < Z < \infty$$

ดังนั้นการหาความน่าจะเป็นของตัวแปรสุ่ม x ที่มีค่าระหว่างค่า a ถึงค่า b สามารถแปลงให้เป็นการแจกแจงแบบปกติมาตรฐาน คือ

$$\begin{aligned} P(a < x < b) &= P\left(\frac{a-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\ &= P(Z_1 < Z < Z_2) \end{aligned}$$

โดยที่ $Z_1 = \frac{a-\mu}{\sigma}$ และ $Z_2 = \frac{b-\mu}{\sigma}$

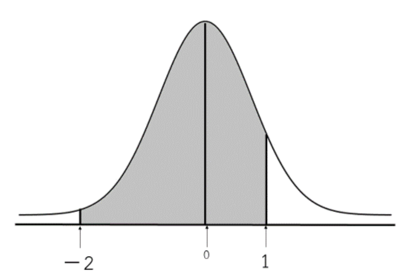
ตัวอย่าง ให้ x มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ 20 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 จงหาความน่าเป็นที่

1. ตัวแปรสุ่ม x มีค่าอยู่ระหว่าง 10 ถึง 25
2. ตัวแปรสุ่ม x มีค่ามากกว่า 30

__วิธีทำ__ กำหนดให้ x มีการแจกแจงแบบปกติ โดยมี $\mu = 20$ และ $\sigma = 5$

1. ความน่าจะเป็นที่ตัวแปรสุ่ม x มีค่าอยู่ระหว่าง 10 ถึง 25 คือ

$$\begin{aligned} P(10 < x < 25) &= P\left(\frac{10-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{25-\mu}{\sigma}\right) \\ &= P\left(\frac{10-20}{5} < z < \frac{25-20}{5}\right) \\ &= P(-2 < z < 1) \end{aligned}$$



$$= P(-2 < z < 0) + P(0 < z < 1)$$

$$= 0.4773 + 0.3413$$

$$= 0.8186$$

2. ความน่าจะเป็นที่ตัวแปรสุ่ม x มีค่ามากกว่า 30 คือ

The Central Limit Theorem

นอกจากนักสถิติจะสนใจว่า “ค่าข้อมูลแต่ละตัวในประชากร” มีการกระจายรอบค่าเฉลี่ยประชากรอย่างไรแล้ว พวกเขายังสนใจว่า “ค่าเฉลี่ยของกลุ่มตัวอย่าง (samples) ที่มีขนาดเท่ากันและมาจากประชากรเดียวกัน” จะกระจายรอบค่าเฉลี่ยของประชากรอย่างไรด้วย

การแจกแจงของค่าเฉลี่ยตัวอย่าง (Distribution of Sample Means)

สมมติว่านักวิจัยเลือกกลุ่มตัวอย่างชายวัยผู้ใหญ่ 30 คน และคำนวณค่าเฉลี่ยของระดับไตรกลีเซอไรด์ในเลือด ได้ **187 mg/dl** จากนั้นเลือกตัวอย่างอีก 30 คน ได้ค่าเฉลี่ย **192 mg/dl** ทำต่อไปเรื่อย ๆ รวม 100 กลุ่มตัวอย่าง ⟶ เราจะได้ค่าเฉลี่ยหลาย ๆ ค่า เช่น 187, 192, 184, ..., 196

สิ่งที่เกิดขึ้นคือ **ค่าเฉลี่ยเองกลายเป็นตัวแปรสุ่ม** และค่าเฉลี่ยของตัวอย่างเหล่านี้จะรวมกันเป็นสิ่งที่เรียกว่า **การแจกแจงของค่าเฉลี่ยตัวอย่าง (sampling distribution of sample means)**

The Central Limit Theorem

เมื่อขนาดตัวอย่าง n มีขนาดใหญ่ การแจกแจงของค่าเฉลี่ยตัวอย่าง ที่สุ่มมาจากประชากรที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ การแจกแจงของค่าเฉลี่ยตัวอย่างจะมีการแจกแจงใกล้เคียงการแจกแจงปกติ โดยที่ $\mu_{\bar{x}} = \mu$ และ $\sigma_{\bar{x}} = \frac{\sigma}{n}$

ตัวอย่าง เด็กอายุระหว่าง 2 ถึง 5 ปี ดูโทรทัศน์โดยเฉลี่ย 25 ชั่วโมงต่อสัปดาห์ สมมติว่าตัวแปรมีการแจกแจงแบบปกติและค่าเบี่ยงเบนมาตรฐานคือ 3 ชั่วโมง หากสุ่มเลือกเด็กอายุระหว่าง 2 ถึง 5 ปี จำนวน 20 คน จงหาความน่าจะเป็นที่ค่าเฉลี่ยของจำนวนชั่วโมงที่พวกเขาดูโทรทัศน์จะมากกว่า 26.3 ชั่วโมง

วิธีทำ ความน่าจะเป็นที่ค่าเฉลี่ยของจำนวนชั่วโมงที่พวกเขาดูโทรทัศน์จะมากกว่า 26.3 ชั่วโมง

$$P(\bar{x} > 30) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > \frac{26.3-\mu}{\sigma/\sqrt{n}}\right)$$

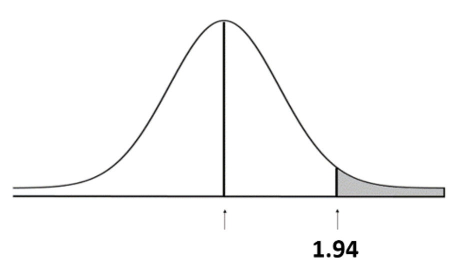
$$= P\left(z > \frac{26.3-25}{3/\sqrt{20}}\right)$$

$$= P\left(z > \frac{1.3}{0.671}\right)$$

$$= P(z > 1.94)$$

$$= 1 - 0.9738$$

$$= 0.0262$$



ตัวอย่าง อายุเฉลี่ยของรถยนต์ที่จดทะเบียนในสหรัฐอเมริกาคือ 8 ปี หรือ 96 เดือน สมมติว่าค่าเบี่ยงเบนมาตรฐานคือ 16 เดือน หากเลือกกลุ่มตัวอย่างแบบสุ่มจำนวน 36 คัน จงหาความน่าจะเป็นที่อายุเฉลี่ยของรถยนต์เหล่านั้นจะอยู่ระหว่าง 90 ถึง 100 เดือน