บทที่ 2
ทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง
(CENTRAL LIMIT THOREM)

2.1 ทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง

ก่อนที่จะศึกษาถึงตัวทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง ซึ่งเป็นทฤษฎีที่มีประโยชน์อย่างสูงในทางปฏิบัติ ขออธิบายความหมายของชื่อทฤษฎีเพื่อเป็นการปูพื้นฐานของความรู้ความเข้าใจในเรื่องนี้เสียก่อน

คำว่าเกณฑ์กลาง หมายถึงตัวแบบหรือค่ากลาง ๆ ที่ปรากฏการณ์ต่าง ๆ จะโน้มเข้าไปใกล้ตัวแทนนั้นหรือค่านั้นเมื่อการสุ่มตัวอย่างดำเนินด้วยการสุ่มตัวอย่างขนาดใหญ่ หรืออาจกล่าวอย่างไม่เป็นทางการได้ 2 นัยคือ

(1) เกณฑ์กลางหมายถึงค่าเฉลี่ยซึ่งเป็นค่ากลางหรือค่าร่วมกัน (Common Point) ที่ค่าของปรากฏการณ์ต่าง ๆ โดยส่วนใหญ่จะมีค่าใกล้เคียงกับค่านี้เมื่อใช้กลุ่มตัวอย่างขนาดใหญ่

(2) เกณฑ์กลางหมายถึงตัวแปรสุ่มตัวใหม่ที่มีการแจกแจงแบบปกติ กล่าวคือ ไม่ว่าตัวแปรสุ่มเดิมจะมีการแจกแจงเป็นรูปใดก็ตาม ยอดรวมหรือค่าเฉลี่ยของตัวแปรนั้น (เหล่านั้น) ที่ผ่านการแปลงรูปโดยการหักออกด้วยค่าคาดหมายและหารตลอดด้วยค่าความเบี่ยงเบนมาตรฐานแล้วจะมีการแจกแจงแบบปกติ $N(0, 1)$ เมื่อกลุ่มตัวอย่างที่ใช้มีขนาดใหญ่

ในที่นี้มุ่งหวังให้เข้าใจตามความหมายที่สองซึ่งเกี่ยวข้องกับตัวทฤษฎีโดยตรง ส่วนความหมายที่หนึ่งเป็นเพียงความหมายที่เนื่องกันอยู่ และเป็นความรู้ความเข้าใจเดิมที่จำเป็นต้องทราบเมื่อเริ่มการศึกษาวิชาสถิติแล้ว

ทฤษฎีการโน้มเข้าสู่เกณฑ์กลางถูกนำไปใช้ในทางปฏิบัติอย่างกว้างขวาง โดยถูกนำไปใช้ในการคำนวณค่าความน่าจะเป็นต่าง ๆ การกำหนดความแม่นยำ (Precision) ของการประมาณหรือคาดทำนาย การกำหนดขนาดของตัวอย่างเพื่อการสำรวจ การสร้างตัวสถิติ (Test Statistics)

และประการที่สำคัญยิ่งก็คือการประยุกต์ทฤษฎีต่าง ๆ ของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ ซึ่งเรื่องการใช้ทฤษฎีดังกล่าวเป็นปัญหาของนักวิจัยเป็นอย่างมากและมักมีคำถามเสมอว่า "ทราบได้อย่างไรว่าสถานการณ์เช่นนั้นคือสถานการณ์ของตัวแปรแบบปกติ" หรือ "ทำไมจึงกำหนดให้ตัวแปรเป็นปกติ" หรือ แบบอื่น ๆ ซึ่งนักวิจัยที่มีความรู้ความเข้าใจในทฤษฎีการโน้มเข้าสู่เกณฑ์กลางแล้ว จะอำนวยความสะดวกในการประยุกต์ทฤษฎีต่าง ๆ ของตัวแปรสุ่มดังกล่าวได้เป็นอย่างสูง

ทฤษฎี 2.1 (Central Limit Theoem)

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน ซึ่งต่างก็มีค่าคาดหมายและค่าความแปรปรวนเป็น $\mu_1, \mu_2, \ldots, \mu_n$ และ $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ ตามลำดับ ให้ $S_n = X_1 + X_2 + \ldots + X_n$ เป็นยอดรวมของตัวแปรเหล่านั้น

ดังนั้น ถ้า a และ b เป็นจำนวนจริงใด ๆ โดยที่ $a < b$ และภายใต้เงื่อนไขทั่ว ๆ ไปประการหนึ่งแล้ว

$$\lim_{n\to\infty} \Pr\{a \leqslant \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leqslant b\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

โดยที่ $E(S_n) = E(X_1 + X_2 + \ldots + X_n) = \mu_1 + \mu_2 + \ldots + \mu_n$

และ $V(S_n) = V(X_1 + X_2 + \ldots + X_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$

ข้อสังเกต

จากทฤษฎี 2.1 เราสามารถให้ข้อสังเกตและคำอธิบายเพิ่มเติมได้ดังนี้

1. $Z_n = \dfrac{S_n - E(S_n)}{\sqrt{V(S_n)}} \to N(0, 1)$ เมื่อ $n\to\infty$

2. $S_n = X_1 + X_2 + \ldots + X_n = \sum_i^n X_i$ หมายความว่า X_i จะต้องมาจากกลุ่มประชากรเดียวกัน ทั้งนี้อาจจะเป็นกลุ่มประชากรใด ๆ ก็ได้ แต่ค่าคาดหมายและค่าความแปรปรวนของแต่ละ X_i จะต้องเป็นค่าที่นับได้ (Finite Expectation และ Finite Variance)

3. "เงื่อนไขทั่วไปประการหนึ่ง" ในตัวทฤษฎีบทคืออะไร เงื่อนไขทั่วไปคือเงื่อนไขที่ว่าในยอดรวม S_n ซึ่งประกอบไปด้วยผลรวมของตัวแปรสุ่มนั้น ตัวแปรสุ่มแต่ละตัวจะต้องไม่ก่อให้เกิดความผันแปรแก่ยอดรวมมากนัก หรือนัยหนึ่งค่าความผิดพลาด (Error of Measurement) ของยอดรวม S_n จะเกิดจากผลรวมของความผิดพลาดจากตัวแปรสุ่ม X_i แต่ละตัว และตัวแปร

ตัวนั้นจะต้องไม่ส่งผลแห่งความผิดพลาดอย่างมากมายผิดปกติแก่ยอดรวม เพราะมิเช่นนั้นแล้ว $V(S_n)$ จะมีค่าสูงมากจนทำให้ไม่อาจยอมรับหรือเชื่อถือในผลลัพธ์ได้ เพราะ $V(S_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$ หรือ (อาจกล่าวอย่างไม่เป็นทางการได้ว่า $V(X_i)$; $i = 1, 2, \ldots, n$ ควรมีค่าน้อย ๆ)

4. $Z_n = \dfrac{S_n - E(S_n)}{\sqrt{V(S_n)}} \rightarrow N(0, 1)$ เมื่อ $n \rightarrow \infty$ มิได้หมายความว่า S_n จะมีการแจกแจงที่โน้มเข้าสู่ $N(0, 1)$ ข้อนี้เข้าใจผิดกันมาก[1] แต่หมายความว่าเมื่อเราแปลงรูป S_n โดยการหักออกด้วย $E(S_n)$ และหารตลอดด้วย $\sqrt{V(S_n)}$ แล้วจะได้ตัวแปรสุ่มตัวใหม่คือ Z_n ซึ่งตัวแปรใหม่นี้จะมีการแจกแจงที่โน้มเข้าสู่ $N(0, 1)$ เมื่อ $n \rightarrow \infty$ มิใช่ S_n โน้มเข้าสู่ $N(0, 1)$

ข้อสังเกตประการนี้มีความสำคัญมาก เพราะถ้าไม่ทำความเข้าใจให้ชัดเจนแล้วจะเกิดความสับสนและจะนำไปปะปนกับกรณีเฉพาะเมื่อ $X_i \sim N(\mu_i, \sigma_i^2)$ [2] แล้วจะพากันเข้าใจผิดตลอดไป

ในกรณีที่ $X_i \sim N(\mu_i, \sigma_i^2)$ นั้นจะเป็นจริงได้ทั้งสองกรณีดังต่อไปนี้คือ

ก. $S_n = X_1 + X_2 + \ldots + X_n$ จะมีการแจกแจงแบบปกติเป็น $N(\sum_i^n \mu_i, \sum_i^n \sigma_i^2)$ ซึ่งหมายความว่ามิจำเป็นต้องแปลงรูปของ S_n แต่ประการใด ก็จะยังคงมีการแจกแจงแบบปกติ เพียงแต่ไม่เป็น $N(0, 1)$ เท่านั้น ผิดกับ $S_n = X_1 + X_2 + \ldots + X_n$ ที่ X_i มีการแจกแจงแบบใด ๆ คือมีการแจกแจงที่เราไม่ทราบแน่ชัดว่าเป็นแบบใด ที่ต้องแปลงรูปเป็น Z_n เสียก่อนจึงจะมีการแจกแจงแบบปกติและเป็นได้เฉพาะ $N(0, 1)$ เท่านั้น

ข. $S_n = X_1 + X_2 + \ldots + X_n$ ที่แปลงรูปเป็น Z_n จะมีการแจกแจงแบบปกติ $N(0, 1)$ เช่นเดียวกับทฤษฎี 2.1 แต่ไม่จำเป็นที่ขนาดตัวอย่างจะต้องมีขนาดใหญ่ คือ n จะมีขนาดใด ๆ ก็ได้ ถ้า $X_i \sim N(\mu_i, \sigma_i^2)$ แล้ว $Z_n = \dfrac{S_n - E(S_n)}{\sqrt{V(S_n)}}$ จะต้องมีการแจกแจงแบบ $N(0, 1)$ เสมอไป

ความจริงทั้งสองประการ (ก, ข) นี้จะพิสูจน์ให้เห็นจริงในบทที่ 4 ในตอนที่ว่าด้วยการแจกแจงแบบปกติ

5. คำว่า "กลุ่มตัวอย่างขนาดใหญ่" หรือ "$n \rightarrow \infty$" ในทฤษฎี 2.1 เป็นคำคลุมเคลือและทำให้เข้าใจไปในลักษณะที่น่าวิตกว่า "จะต้องสุ่มตัวอย่างเป็นพันเป็นหมื่นชุด" จึงจะใช้ทฤษฎี

---

[1] นักศึกษาสามารถตรวจสอบความเป็นจริงข้อนี้ได้เมื่อศึกษาถึงบทที่ 3 และบทที่ 4

[2] $X_i \sim N(\mu_i, \sigma_i^2)$ อ่านว่าตัวแปรสุ่ม X_i มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น μ_i และค่าความแปรปรวนเป็น σ_i^2

นี้ได้ ความจริงแล้วถ้ากลุ่มประชากรมีลักษณะที่สมมาตรกัน[1]แล้ว กลุ่มตัวอย่าง n จะต้องมีขนาดไม่ต่ำกว่า 8 หน่วย[2] ก็สามารถประยุกต์ทฤษฎีที่ 2.1 ไปใช้ได้ แต่ถ้ากลุ่มประชากรไม่สมมาตรกัน กลุ่มตัวอย่างที่ใช้ก็ควรเพิ่มให้สูงขึ้นและในกรณีที่ไม่อาจทราบว่ากลุ่มประชากรที่มุ่งศึกษานั้นสมมาตรหรือไม่ หรือไม่มีความรู้เกี่ยวกับประชากรมากพอที่จะคาดหมายความสมมาตรได้ "กลุ่มตัวอย่างขนาดใหญ่" ที่เหมาะสมแก่งานจะต้องมีขนาดไม่ต่ำกว่า 30 หน่วย[3] และถ้ายิ่งต้องการให้มีความถูกต้องมากขึ้น กลุ่มตัวอย่างที่ใช้ก็ควรเพิ่มให้มากขึ้น ๆ ตามลำดับซึ่งจะมีขนาดเท่าใดจึงจะเหมาะสมกับข้อกำหนดของผู้วิจัยแต่ละรายนั้น เราจะกล่าวถึงในลำดับต่อไปในตัวอย่างท้ายตอน 2.1 นี้

6. จะพิสูจน์ความถูกต้องของตัวทฤษฎี 2.1 และบทแทรกต่าง ๆ ในบทที่ 3 ภายหลังจากที่นักศึกษาได้เรียนรู้เรื่อง Moment Generating Function แล้ว

7. ในกรณีที่มี Subscript n กำกับไว้ในตัวแปรเชิงสุ่ม เช่น Z_n หรือ $\bar{X}_n$ ก็เพียงเพื่อแสดงให้เห็นว่า ขณะนี้เรากำลังเกี่ยวข้องอยู่กับขนาดของตัวอย่างเท่านั้น มิได้มุ่งหมายเป็นอย่างอื่น และโดยปกติ (โดยเฉพาะในทางปฏิบัติ) ก็จะไม่เขียนกำกับไว้ให้รุงรัง หรือน่า "เวียนหัว" แต่อย่างไร อีกประการหนึ่ง $\bar{X}_n$ นั้นคือตัวแปรสุ่มของ Sample Mean (กล่าวถึงแล้วในตอน 1.1) อย่าเข้าใจว่าเป็น Population Mean เพราะเห็นว่าใช้ X ตัวพิมพ์ใหญ่ โดยปกติจะเขียนเป็น $\bar{x}$

8. การที่เรามีความรู้เกี่ยวกับการแจกแจงของประชากร มิได้หมายความว่าจะใช้ทฤษฎี 2.1 และบทแทรกต่าง ๆ ของทฤษฎีนี้ไม่ได้ ตรงกันข้าม การมีความรู้ดังกล่าวกับยิ่งส่งเสริมให้การประมาณค่าตามทฤษฎีนี้ถูกต้องมากยิ่งขึ้น

บทแทรก 2.1 ถ้า $\bar{X}_n$ เป็น Sample Mean จากกลุ่มตัวอย่างขนาด n ที่สุ่มมาจากประชากรใด ๆ ที่มีค่าเฉลี่ย (ค่าคาดหมาย) และความแปรปรวนเป็น μ และ σ^2 ตามลำดับ ดังนั้นภายใต้เงื่อนไขทั่วไปประการหนึ่ง และ a, b เป็นจำนวนจริงใด ๆ ที่ $a < b$ แล้ว

$$\lim_{n\to\infty} \Pr\{a \leqslant \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant b\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

---

[1] จะกล่าวถึงและแสดงให้เห็นด้วยภาพของความสมมาตรของกลุ่มประชากรซึ่งมีการแจกแจงแบบต่าง ๆ ในบทที่ 4 เช่นกรณี Binomial จะสมมาตรเมื่อ $p = \frac{1}{2}$

[2] Isac N. Gibra, **Probability and Statistical Inference for Scientists and Engineers. p. 125**

[3] H.D. Brunk, **An Introduction to Mathematical Statistics, p. 167.**

บทแทรก 2.2 ถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและสุ่มมาจากประชากรเดียวกัน (หรือจากต่างประชากร แต่กลุ่มประชากรเหล่านั้นมีการแจกแจงแบบเดียวกัน มีค่าเฉลี่ยและค่าความแปรปรวนเดียวกัน) นั่นคือ

$$E(X_1) = E(X_2) = \ldots = E(X_n) = \mu$$

และ

$$V(X_1) = V(X_2) = \ldots = V(X_n) = \sigma^2$$

แล้ว

$$\lim_{n\to\infty} \{a \leqslant \frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant b\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

โดยที่ $S_n = X_1 + X_2 + \ldots + X_n$, $E(S_n) = \mu + \mu + \ldots + \mu = n\mu$

และ $V(S_n) = \sigma^2 + \sigma^2 + \ldots + \sigma^2 = n\sigma^2$

บทแทรก 2.1 และ 2.2 เป็นกรณีเฉพาะของทฤษฎี 2.1 เป็นกรณีเฉพาะเมื่อกลุ่มตัวอย่างที่ได้นั้นถูกสุ่มมาจากกลุ่มประชากรเดียวกัน ขอให้สังเกตว่าค่าคาดหมายและค่าความแปรปรวนของตัวแปรสุ่ม $X_1, X_2, \ldots, X_n$ มีค่าร่วมกันเป็น μ และ σ^2 ตามลำดับ

บทแทรกที่ 2.1 และ 2.2 นั้น ความจริงเป็นเรื่องเดียวกันแต่เสนอกันคนละรูป บทแทรกที่ 2.1 เสนอในรูปของค่าเฉลี่ย (Sample Mean) ส่วนบทที่ 2.2 เสนอในรูปของยอดรวม (Sample Total)[1] การนำไปใช้ประโยชน์จึงเป็นไปในลักษณะเดียวกัน แต่อาจทำให้มองเห็นภาพของสถานการณ์ของเรื่องนั้นได้ต่างกัน

อนึ่ง ถ้าตัวแปรสุ่ม X_i มีค่าได้เพียง 2 ค่าคือ 0 และ 1 บทแทรกทั้งสองจะกลายเป็นอีกรูปหนึ่งคือ De Moivre's Theorem ซึ่งก็มีประโยชน์เป็นอย่างสูงในทางปฏิบัติเช่นกัน

บทแทรก 2.3 (De Moivre Theorem)

ถ้า S_n คือยอดรวมของจำนวน Success ทั้งหมดจากการทดลอง (หรือการสังเกต การวัดค่า) ที่เป็นอิสระต่อกัน n ครั้ง โดยที่การทดลองแต่ละครั้งมีโอกาสที่จะประสบความสำเร็จ (Success) เท่ากับ p

ถ้า $a < b$

แล้ว

$$\lim_{n\to\infty} \{a < \frac{S_n - np}{\sqrt{npq}} < b\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

---

[1] ขอให้สังเกตว่า $\bar{X} = \frac{S_n}{n}$ และ $V(\bar{X}) = V\frac{(S_n)}{n} = \frac{1}{n^2} V(S_n) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$

ตัวอย่าง 2.1 สมมุติว่าความยาวขาหน้าของสุนัขไทยพื้นเมือง มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 นิ้ว[1] ถ้าจะให้เชื่อถือได้ถึง 99% เป็นอย่างน้อยว่า ความยาวเฉลี่ยของขาหน้าคลาดไปจากค่าจริงเพียงไม่ถึง 0.1 นิ้ว อยากทราบว่าจะต้องนำสุนัข (โดยสุ่ม) มากี่ตัวเพื่อทดสอบความยาวของขาหน้าเพื่อให้ตรงตามเกณฑ์ที่กำหนดไว้นั้น

วิธีทำ

สิ่งที่ต้องการคือ จาก $\Pr\{|\bar{x} - \mu| < 0.1\} = .99$, $n = ?$

อาศัยทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง (C.L.T)

$$=> \quad \Pr\{|\frac{\bar{x} - \mu}{\sqrt{V(\bar{X})}}| < \frac{0.1}{\sqrt{V(\bar{X})}}\} = .99$$

$$=> \quad \Pr\{|Z| < \frac{0.1}{\sigma/\sqrt{n}}\} = .99$$

$$= \Pr\{|Z| < 2.575\}$$

$$=> \quad \frac{0.1}{\sigma/\sqrt{n}} = 2.575$$

$$\because \quad \sigma = 1$$

$$=> \quad \frac{0.1}{1/\sqrt{n}} = 2.575$$

$$=> \quad \sqrt{n} = \frac{2.575}{0.1}$$

$$\sqrt{n} = 25.75$$

$$n = 663$$

---

[1] ส่วนเบี่ยงเบนมาตรฐาน เป็นค่าที่บ่งบอกไว้ให้ทราบว่า โดยถัวเฉลี่ยแล้ว ค่าสังเกตต่าง ๆ (อาจหมายถึงจากกลุ่มตัวอย่างหรือจากกลุ่มประชากรก็ได้) คลาดเคลื่อนไปจากค่าคาดหมายมากน้อยเพียงใด ถ้าค่าส่วนเบี่ยงมาตรฐานต่ำ แสดงว่าในกลุ่มประชากรนั้นมีความละม้ายคล้ายคลึงกันมาก (แตกต่างกันน้อย) ถ้าค่าส่วนเบี่ยงเบนมาตรฐานสูง แสดงว่ากลุ่มประชากรนั้นมีความแตกต่างกันมาก ยกตัวอย่างเช่นการสำรวจทัศนคติของประชากร ถ้าส่วนเบี่ยงเบนมาตรฐาน (ของทัศนคติ) มีค่าต่ำแสดงว่าประชากรกลุ่มนั้นมีทัศนคติหรือความคิดเห็นคล้าย ๆ กัน ซึ่งถ้าจะสุ่มตัวอย่างจากประชากรกลุ่มนั้นมาเพื่อสำรวจก็ไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ขณะเดียวกัน ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าสูงแสดงว่ากลุ่มประชากรมีความเห็นไม่ค่อยลงกัน

นั่นคือ ต้องสุ่มสุนัขไทยเป็นจำนวนทั้งสิ้น 663 ตัว จึงจะทำให้สามารถเชื่อถือได้ถึง 99% เป็นอย่างน้อยว่า ความยาวขาหน้าที่วัดได้จากกลุ่มตัวอย่างจะคลาดเคลื่อนไปจากค่าจริงไม่ถึง 0.1 นิ้ว

หมายเหตุ การที่ใช้ Absolute Deviation $|\bar{x} - \mu|$ มิได้ใช้ Deviation ตามธรรมดาคือ $(\bar{x} - \mu)$ เพราะเราไม่อาจได้ทราบว่า $\bar{x} < \mu$ หรือ $\bar{x} > \mu$ เมื่อใช้ $|\bar{x} - \mu|$ จึงครอบคลุมความคลาดเคลื่อนทั้งสองกรณีดังกล่าวได้ครบ

ตัวอย่าง 2.2 ท่านกำลังวางแผนที่จะวิจัยพันธุ์ข้าวโพดพันธุ์ใหม่ โดยต้องการทราบผลผลิตเฉลี่ยต่อไร่ และวางแผนว่าจะปลูกทั้งสิ้นไร่ละ n หลุม จากการทดลองและวิจัยครั้งก่อนซึ่งเคยมีผู้วิจัยผลผลิตของข้าวโพดพันธุ์นี้ไว้ ปรากฏว่ามีส่วนเบี่ยงเบนมาตรฐานของผลผลิตต่อไร่เท่ากับ 2 ถัง และท่านมีเหตุผลที่จะถือว่าค่านี้เป็นตัวแทนของ Population Standard Deviation

ก. จงใช้อสมการของเชพบิเชฟ กำหนดขนาดตัวอย่าง n (จำนวนหลุมที่ปลูก) ที่จะทำให้สามารถเชื่อถือได้ถึง 80% เป็นอย่างน้อยว่าผลผลิตเฉลี่ยต่อไร่จากการทดลองปลูก คลาดไปจากผลผลิตจริงของสายพันธุ์ใหม่นี้ไม่เกินครึ่งถัง

ข. จงอาศัยทฤษฎีการโน้มเข้าสู่เกณฑ์กลางคำนวณหาขนาดของตัวอย่างตามข้อ ข.

วิธีทำ

ก. สิ่งที่ต้องการคือจาก $\Pr\{|\bar{x} - \mu| \leqslant 0.5\} \geqslant .80$ อยากทราบขนาดของตัวอย่าง n ควรมีค่าเท่าใดจึงจะสอดคล้องกับข้อกำหนด

จากอสมการของเชพบิเชฟ พบว่า

$$\Pr\{|\bar{x} - \mu| \leqslant .05\} \geqslant 1 - \frac{V(\bar{X})}{(0.5)^2}$$

$$\because \quad V(\bar{X}) = \frac{\sigma^2}{n} \; ; \quad s^2 = 2^2 = 4$$

$$=> \quad \hat{V}(\bar{X}) = \frac{s^2}{n} = \frac{4}{n}$$

$$\Pr\{|\bar{x} - \mu| \leqslant 0.5\} \geqslant 1 - \frac{4}{n(0.5)^2}$$

นั่นคือ $1 - \dfrac{4}{n(0.5)^2} \geqslant .80$

$$=> \quad n \geqslant \frac{4}{(.20)(.25)}$$

$$n \geqslant 80$$

แสดงว่า ต้องปลูกข้าวโพดอย่างน้อย 80 หลุม จึงจะสามารถเชื่อถือได้ถึง 80% เป็นอย่างน้อยว่าผลผลิตเฉลี่ยต่อไร่จากการทดลอง จะคลาดไปจากผลผลิตจริงของสายพันธุ์ใหม่นี้ถึงครึ่งถัง

ข. จาก $Pr\{|\bar{x} - \mu| \leqslant 0.5\} \geqslant .80$; $n = ?$

โดยอาศัยทฤษฎีการโน้มสู่เกณฑ์กลาง (C.L.T) จะพบว่า

$$Pr\{|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leqslant \frac{0.5}{\sigma/\sqrt{n}}|\} \geqslant .80$$

$$=> \quad Pr\{|Z| \leqslant \frac{0.5}{\sigma/\sqrt{n}}\} \geqslant .80 = Pr\{|z| \leqslant 1.285\}$$

$\because$ $s = 2$ เป็นตัวแทนของ σ

$$=> \quad \frac{0.5}{2/\sqrt{n}} \geqslant 1.285$$

$$\sqrt{n} \geqslant 5.1440$$

$$=> \quad n \geqslant 26$$

แสดงว่า ถ้าจะให้สามารถคาดหมายได้ถูกต้องเป็นอย่างน้อยถึง 80% ว่าผลผลิตต่อไร่จากการทดลองปลูกจะคลาดเคลื่อนไปจากผลผลิตจริงของสายพันธุ์ประมาณครึ่งถัง ผู้วิจัยจะต้องทำการทดลองปลูกประมาณ 26 หลุม (เป็นอย่างน้อย)

ข้อสังเกต

ขอให้สังเกตว่า เมื่อใช้อสมการของเชพบิเชฟ เป็นเครื่องมือในการกำหนดขนาดตัวอย่าง ๆ ที่จะได้จะใหญ่กว่าที่ได้รับจากการคำนวณโดยอาศัยทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง ทั้งนี้เพราะอสมการของเชพบิเชฟอาศัยข้อสนเทศ (Information) จากกลุ่มประชากรน้อยกว่า เมื่อเป็นเช่นนี้ ขนาดตัวอย่างจึงใหญ่กว่าขาดความกระชับรัดกุมกว่าทั้งนี้เพื่อชดเชยข้อสนเทศที่ขาดไป

ตัวอย่าง 2.3

หลอดอีเลคโทรนิค 30 ตัวคือ $D_1, D_2, ..., D_n$ ถูกนำมาใช้ในลักษณะดังต่อไปนี้คือ

ก. ต่อเข้าวงจรคราวละ 1 หลอด เมื่อหลอดที่ 1 เสื่อมสภาพก็นำหลอดที่ 2 ต่อเข้าในวงจรแทน ทำดังนี้เรื่อยไปจนครบ 30 หลอด แต่ละครั้งจะบันทึกอายุการใช้งานไว้ทุกหลอด และทราบว่าอายุการใช้งานของหลอดอีเลคโทรนิคมีการแจกแจงแบบ Exponential อายุใช้งานเฉลี่ย

เท่ากับ 10 เดือน จงหาความน่าจะเป็นที่อายุการทำงานรวมของหลอดอิเลคโทรนิคทั้ง 30 หลอดจะนานเกินกว่า 350 เดือน

ข. ต่อเข้าวงจรพร้อมกันทั้ง 30 หลอด โดย หลอดถูกนำต่อเข้าสู่วงจรย่อยวงจรละหลอด และวงจรย่อยทั้ง 30 วงจรนั้นทำงานอย่างอิสระต่อกัน ซึ่งเมื่อกระทำดังนี้จะทำให้หลอดแต่ละหลอดมีโอกาสที่จะเสื่อมสภาพในระหว่างทำงานภายหลังเมื่อใช้ไปแล้ว 10 เดือน เท่ากับ 10% และวงจรนี้ยังคงที่จะทำงานต่อไปได้ก็ต่อเมื่อวงจรย่อยยังคงทำงานได้อย่างน้อย 25 วงจร จงหาความน่าจะเป็นที่วงจรไฟฟ้าสามารถทำงานต่อไปได้

วิธีทำ

ก. ให้ $T_i, i = 1, 2, \ldots, 30$ แทนอายุใช้งานของหลอดอิเลคโทรนิคที่ i

และ $f_{T_i}(t) = \lambda e^{\lambda t} \quad ; \quad \lambda = 0.1$[1]

นั่นคือ $f_{T_i}(t) = 0.1e^{-(0.1)t}, \; 0 < t < \infty$

ให้ T แทนอายุของหลอดอิเลคโทรนิค $D_1, D_2, \ldots, D_n$

นั่นคือ $T = \sum^{30} T_i$

ดังนั้น

สิ่งที่ต้องการคือ

$$\Pr\{ \sum^{30} T_i > 350\} = ?$$

อาศัยทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง และเพราะเหตุว่าการทำงานของหลอดอิเลคโทรนิค $D_1, D_2, \ldots, D_i$ เป็นอิสระต่อกัน

---

[1] ตัวแปรสุ่ม T จะมีการแจกแจงแบบ Exponential ถ้าตัวแปรสุ่ม T มี Probability Density เป็น $f(x) = \lambda e^{-\lambda t}; 0 < t < \infty, \lambda > 0$ ซึ่งจะพบว่า $E(T) = 1/\lambda$, $V(T) = 1/\lambda^2$ ในกรณีนี้ $E(T) = 1/\lambda = 10$ ดังนั้น $\lambda = 1/10 = 0.1$

$$\Pr\{\sum_{i}^{30} T_i > 350\} = \Pr\left\{ \frac{\sum_{i}^{30} T_i - E(T_i)}{\sqrt{\sum_{i}^{30} V(T_i)}} > \frac{350 - \sum^{30} E(T_i)}{\sqrt{\sum_{i}^{30} V(T_i)}} \right\}$$

$$\sum_{i}^{30} E(T_i) = 10 + 10 + 10 + + 10 = 300$$

$$\sqrt{\sum_{i}^{30} V(T_i)} = \sqrt{\frac{1}{.1} + \frac{1}{.1} + \ldots + \frac{1}{.1}} = \sqrt{3000} = 54.75$$

$$\Pr\{\sum^{30} T_i > 350\} = \Pr\{Z > \frac{350 - 300}{54.75}\}$$

$$= \Pr\{Z > .913\}$$

$$= .1814$$

นั่นคือหลอดอิเลคโทรนิคทั้ง 30 หลอดจะมีอายุใช้งานรวมกันเกินกว่า 350 เดือน ด้วยโอกาสที่จะเป็นไปได้เพียงประมาณ 18%

ข. ให้ p แทนโอกาสที่หลอดอิเลคโทรนิค (วงจรย่อย) ยังคงทำงานได้ต่อไปเมื่อใช้งานมาแล้ว 10 เดือน

q แทนโอกาสที่หลอดอิเลคโทรนิค (วงจรย่อย) จะเสื่อมสภาพขณะทำงานหลังจากใช้งานมาแล้วนาน 10 เดือน

ดังนั้น $p = .90$ และ $q = .10$

ให้ S_i, $i = 1, 2, \ldots, 30$ แทนสมรรถนะของหลอดอิเลคโทรนิคที่ i (วงจรย่อยที่ i) ดังนั้นสิ่งที่ต้องการทราบก็คือ

$$\Pr\{\sum_{i}^{30} S_i < 25\} = ?$$

เพราะว่า S_i มีค่าได้ 2 ค่าคือ

$$S_i = \begin{cases} 1 & \text{ถ้าหลอดยังคงทำงานได้} \\ 0 & \text{ถ้าหลอดเสื่อมสภาพ} \end{cases}$$

แสดงว่า S_i มีการแจกแจงแบบ Bernoulli

ดังนั้น $\sum_{i}^{30} S_i$ มีการแจกแจงแบบ Binomial มีพารามิเตอร์ $n = 30$ และ $p = .90$

$$E(\sum_{i}^{30} S_i) = np = (30) \times (.90) = 27^{1}$$

$$V(\sum_{i}^{30} S_i) = npq = (30) \times (.90) \times (.10) = 2.7$$

ดังนั้น

$$\Pr\{\sum_{i}^{30} S_i < .25\} = \Pr\left\{\frac{\sum_{i}^{30} S_i - \sum_{i}^{30} E(S_i)}{\sqrt{\sum_{i}^{30} V(S_i)}} < \frac{25 - \sum_{i}^{30} E(S_i)}{\sqrt{\sum_{i}^{30} V(S_i)}}\right\}$$

$$= \Pr\{Z < \frac{25 - 27}{\sqrt{2.7}}\}$$

$$= \Pr\{Z < -1.219\} = .1112$$

นั่นคือ เมื่อต่อหลอดอิเลคโทรนิคเข้าวงจรย่อยละ 1 หลอด ภายหลังเมื่อใช้งานไปแล้ว 10 เดือน วงจรนี้จะมีโอกาสที่จะสามารถทำงานต่อไปได้ (หลอดเสื่อมสภาพไม่เกิน 5 หลอด) เพียงประมาณ 11%

---

[1] หมายความว่าเมื่อวงจรนี้ถูกใช้งานไปแล้ว 10 เดือน คาดว่าหลังจากนั้นจะยังคงมีวงจรย่อยที่ทรงสภาพเดิมที่สามารถจะทำงานต่อได้ประมาณ 27 วงจร (หลอด)