บทที่ 7
การประมาณค่าของพารามิเตอร์ด้วยช่วงเชื่อมั่น
(Interval Estimate)

7.1 บทนำ

ดังที่ได้กล่าวมาแล้วในตอนแรก ๆ ว่าการประมาณค่ามีอยู่ 2 วิธีด้วยกันคือ

แบบที่ 1 ประมาณค่าของพารามิเตอร์ด้วยค่าคงที่ (Point estimate)

แบบที่ 2 ประมาณค่าของพารามิเตอร์ด้วยช่วง (Interval estimate) แต่โดยปกติเรามักจะใช้คำว่าช่วงเชื่อมั่น (Confidence interval) แทน ทั้งนี้ก็เพราะว่าเรามักจะกำหนดค่าของความเชื่อมั่นลงไปด้วยในแต่ละปัญหาที่จะสร้างค่าประมาณด้วยช่วงด้วยการที่จะกำหนดว่าจะใช้ความเชื่อมั่นเท่าไร ก็ขึ้นอยู่กับลักษณะความสำคัญของแต่ละปัญหา เช่น ในการทดลองทางการเกษตรในเรื่องของ การใช้ปุ๋ย วิธีการเพาะปลูกให้ได้ผลดี หรือในการวิจัยอื่น ๆ ที่ไม่เป็นปัญหาวิกฤติ ผู้วิจัยก็มักจะใช้ระดับความเชื่อมั่นสูงแค่ 90%, 95%, หรือ 99% เท่านั้น แต่ถ้าเป็นกรณีที่เป็นเรื่องวิกฤติเช่น ศึกษาเรื่องยาที่จะใช้กับคนไข้, อานุภาพของสารปรมาณู ฯลฯ เหล่านี้ ผู้วิจัยก็จำเป็นที่จะต้องใช้ระดับความเชื่อมั่นสูงมาก เช่นใช้ 99.99% หรือ 99.9999% เป็นต้น

การประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่นคือการหาช่วงเชื่อมั่น (a ถึง b) ที่ให้ความเชื่อมั่นสูงถึง $(1-\alpha)100\%$ ช่วงดังกล่าวจะรวมค่าพารามิเตอร์ที่เราสนใจอยู่ด้วย

สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้คือ

$$Pr(a > \theta < b) = 1 - \alpha$$

โดยที่ a และ b คือค่าของตัวแปรเชิงสุ่มที่เกี่ยวข้องอยู่กับตัวสถิติ $(\hat{\theta})$ ที่ใช้ประมาณค่าของพารามิเตอร์ (θ)

$b - a$ คือค่าที่ใช้วัดความแม่นยำ (Precision) ของการประมาณค่า

$1 - \alpha$ คือค่าที่ใช้วัดความน่าเชื่อถือ (Reliability)

ข้อสังเกตุ ถึงแม้ว่าจะใช้วิธีการประมาณค่าด้วยช่วงเชื่อมั่นก็ตามแต่พื้นฐานที่นำมาใช้ก็ยังต้องพึ่งพิงวิธีการประมาณค่าพารามิเตอร์ด้วยค่าคงที่อยู่ (Point estimate)

ตัวอย่าง ในการสร้างช่วงเชื่อมั่นสำหรับค่าเฉลี่ยของประชากร μ โดยที่ใช้วิธีการคือสุ่มตัวอย่างมาขนาด n ประกอบด้วย $X_1,\ldots,X_n$ โดยที่ $\overline{X}$ คือค่าเฉลี่ยที่ได้จากตัวอย่าง

จะได้ว่า $\Pr(\overline{X} - Z_{\alpha/2}\,\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\,\sigma_{\overline{X}}) = 1 - \alpha$

โดยที่ $\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}}$ คือขีดจำกัดล่าง (a)

$\overline{X} + Z_{\alpha/2}\,\sigma_{\overline{X}}$ คือขีดจำกัดบน (b)

จะเห็นได้ว่า ค่าของ μ ที่เราสร้างขึ้นนั้นถูกจำกัดด้วยปัจจัย 3 ประการคือ

ปัจจัยที่ 1 ค่าเฉลี่ยจากตัวอย่าง $\overline{X}$ ($\overline{X}$ คือ Point estimate ของ μ)

ปัจจัยที่ 2 $\sigma_{\overline{X}}$ ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย

และ ปัจจัยที่ 3 ค่าคะแนนมาตรฐานของ Z ซึ่งขึ้นอยู่กับการกำหนดว่าจะให้เป็นเท่าไร โดยดูจากปัญหาที่ศึกษาประกอบด้วยดังที่กล่าวไว้แล้วในตอนต้น การที่เราใช้ Z ในปัญหานี้ก็เพราะผลจากทฤษฎีการเข้าสู่เกณฑ์กลาง (Central Limit Theorem) ได้กล่าวไว้ว่าไม่ว่าข้อมูลที่ได้จากการสังเกตุจะมาจากประชากรที่มีการแจกแจงแบบใดก็ตาม ตัวสถิติที่อยู่ในรูปของ $\overline{X}$ (n มีขนาดใหญ่พอ โดยปกติจะใช้ค่า 30 เป็นเครื่องวัด) แล้ว $\overline{X}$ จะมีการแจกแจงแบบปกติ

ในเรื่องของความแม่นยำ (Precision) จะพบว่าค่าของมันจะมีค่าเปลี่ยนแปลงไปในทิศทางเดียวกับขนาดของตัวอย่าง n แต่จะมีทิศทางตรงกันข้ามกับขนาดของความแปรปรวนของประชากร

ถ้ากำหนดความเชื่อมั่น คือ $(1 - \alpha)100\%$ โดยที่ใช้ขนาดตัวอย่าง n ความแม่นยำของค่าที่จะประมาณด้วยความเชื่อมั่น $(1 - \alpha)100\%$ จะขึ้นอยู่กับขนาดของความแปรปรวน นั่นคือ ถ้าความแปรปรวนสูงจะส่งผลให้ความแม่นยำมีขนาดต่ำ และช่วงเชื่อมั่นก็จะมีขอบเขตกว้างมากขึ้นตาม

เรื่องของช่วงเชื่อมั่นสำหรับพารามิเตอร์จะแบ่งออกเป็นตอน ๆ เพื่อง่ายแก่การศึกษา

ตอนที่ 1 การประมาณค่าเฉลี่ยของประชากรด้วยช่วงเชื่อมั่น

กรณีที่ 1 ศึกษาเฉพาะประชากรเดียว

ตัวอย่าง กำหนดให้ตัวแปรสุ่ม X มีการแจกแจงแบบปกติ $N(\mu,\sigma^2)$ ถ้าสุ่มตัวอย่างชุดหนึ่งได้ $x_1,\ldots,x_n$ จะได้ว่าค่าเฉลี่ยจากกลุ่มตัวอย่างนี้คือ $\bar{X}$ จะมีลักษณะดังนี้

$$\Pr(\mu - 1.96\,\sigma/\sqrt{n} \leq \bar{x} \leq \mu + 1.96\,\sigma/\sqrt{n}) = 0.95 \qquad \ldots\ldots\ldots(1)$$

โดยที่เรามีขนาดความเชื่อมั่น $(1\alpha)100\% = 95\%$ นั่นเอง

ในที่นี้ ค่า 1.96 คือค่าของ $Z_{\alpha/2}$ จากตารางการแจกแจงปกติเราสามารถจะเขียนความน่าจะเป็นของ μ ในช่วงเชื่อมั่นดังกล่าวได้อีกแบบคือ

$$\Pr(\bar{x} - 1.96\,\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.96\,\sigma/\sqrt{n}) = 0.95 \qquad \ldots\ldots\ldots(2)$$

สมการที่ (1) และ (2) เป็นค่าที่สมนัยกัน

ดังนั้น $\Pr(\bar{x} - 1.96\,\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.96\,\sigma/\sqrt{n}) = 0.95$

ถ้าหากว่าเราทราบค่า σ ก็จะสามารถนำค่าสังเกตุที่ได้มาสร้างค่าของ

$$\bar{x} - 1.96\,\sigma/\sqrt{n} \qquad \text{และ } \bar{x} + 1.96\,\sigma/\sqrt{n}$$

นั่นหมายความว่า ช่วงเชื่อมั่น $(\bar{x} - 1.96\,\sigma/\sqrt{n}\ ,\ \bar{x} + 1.96\,\sigma/n)$ จะรวมค่าของพารามิเตอร์ μ (ที่ไม่ทราบค่า) ด้วยความน่าจะเป็นถึง 0.95

เราสามารถเขียนเป็นรูปทั่ว ๆ ไปได้ดังนี้ $\Pr(\theta_0 < \theta < \theta_1) = 1 - \alpha$

บางที่เราเรียก (θ_0, θ_1) นี้ว่า ช่วงเชื่อมั่นเชิงสุ่มก็ได้ เปรียบเทียบกับในตัวอย่างของ

$$\theta_0 = \bar{x} - 1.96\,\sigma/\sqrt{n} \qquad \theta_1 = \bar{x} + 1.96\,\sigma\sqrt{n}$$

การที่จะสร้างช่วงเชื่อมั่นของตัวพารามิเตอร์ใด ๆ ขึ้นมามีหลักเกณฑ์ว่าควรจะสร้างตัวสถิติที่เกี่ยวข้องกับค่าสังเกตุของตัวอย่างและตัวพารามิเตอร์นั้น โดยมีข้อแม้ว่าการแจกแจงของตัวสถิติที่สร้างนี้จะต้องไม่เกี่ยวข้องกับพารามิเตอร์ใด ๆ ตัวสถิติดังกล่าวเรียกว่า Pivotal quantity

ตัวอย่างของ Pivotal quantity

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad , \quad t_{(n-1)} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

$$X^2(n-1) = \frac{ns^2}{\sigma^2} \quad , \quad \text{เป็นต้น}$$

ตัวอย่าง นักสัตววิทยาได้ทดลองสุ่มหนูมา 16 ตัวเพื่อทดสอบปฏิกริยาของยากับน้ำหนักของหนู ปรากฏว่า หลังจากให้ยาแล้วน้ำหนักของหนูโดยเฉลี่ยเป็น 20 กรัม สมมุติว่าการเพิ่มน้ำหนักของหนูมีการแจกแจงแบบปกติ $N(\mu,\sigma^2)$ โดยที่ $\sigma^2 = 4$.

ให้สร้างช่วงเชื่อมั่น ณ ระดับ 95% สำหรับค่าคาดหมายของน้ำหนักของหนูที่เพิ่มขึ้นหลังจากใช้ยานี้

Pivotal quantity ที่จะใช้คือ

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - \mu}{2/\sqrt{n}}$$

เนื่องจาก $Z \sim N(0,1)$ ซึ่งมีการแจกแจงแบบอิสระจากพารามิเตอร์

โดยที่ $\Pr(-1.96 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$

$$\Pr(-1.96\,\sigma/\sqrt{n} \leq \bar{x} - \mu \leq 1.96\,\sigma/\sqrt{n}) = 0.95$$

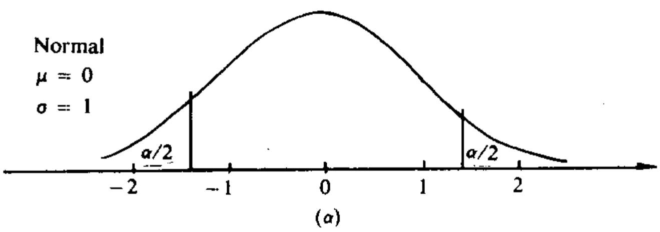
$$\Pr(\bar{x} - 1.96/\sqrt{n} \leq \mu \leq \bar{x} + 1.96/\sqrt{n}) = 0.95$$

เนื่องจาก $\bar{x} = 20,\ n = 16,\ = 2$

ดังนั้น $\Pr(19.02 \leq \mu \leq 20.98) = 0.95$

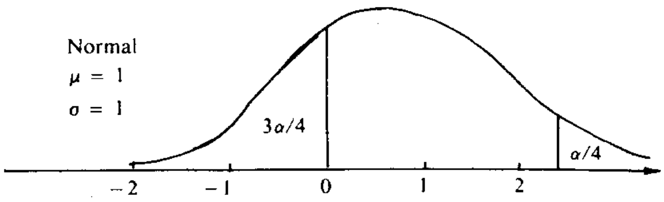
นั่นหมายความว่า ความน่าจะเป็นที่ช่วงเชิงสุ่ม (19.02, 20.98) จะรวม μ ไว้มีค่าถึง 95% ถึง 95%

ข้อสังเกตุ ในกรณีที่สร้างช่วงเชิงสุ่ม โดยการกำหนดให้พื้นที่แต่ละปลายเท่ากัน (Equal tail area) เราอาจจะเรียกจำกัดลงไปว่าเป็นช่วงเชิงสุ่มแบบ "Central interval" ดังรูปที่ 1



100(1 − α)% Confidence interval

อาจสร้างช่วงเชิงสุ่มแบบไม่ใช่ Central ได้เช่นกัน ดังรูปที่ 2



100(1 − α)% Confidence interval

ตัวอย่างที่ 2 ช่างตัดเสื้อสตรีได้สังเกตว่าสตรี 9 คนที่มาตัดเสื้อในวันหนึ่ง มีขนาดหน้าอกเฉลี่ย 36 นิ้ว โดยที่มีส่วนเบี่ยงเบนมาตรฐาน 2 นิ้ว ถ้าหากว่าการวัดมีการแจกแจงแบบปกติ ให้สร้างช่วงเชื่อมั่นเชิงสุ่ม ณ ระดับ 99% เพื่อหาค่าคาดหมายของขนาดที่วัด

$$\text{Pivotal quantity} \quad t_{(n-1)} = \frac{\bar{X} - \mu}{s'\sqrt{n}}$$

$$s'^2 = \Sigma(x_i - \bar{x})^2/n = 4, \bar{x} = 39, n = 9$$

$$s'^2 = \Sigma(x_i - \bar{x})^2/(n-1) = 9(4)/8 = 4.5$$

ใช้ Central interval **พื้นที่ปลายทั้งสองข้างเท่ากัน**

$\alpha/2 = (1-0.99)/2 = .005$ **ตรงกับ** $t_{8;.005} = 3.355$

ดังนั้น

$$Pr(-3.355 \leqslant \frac{\bar{x}-\mu}{s'/\sqrt{n}} \leqslant 3.355) = 0.99$$

$$pr(\bar{x} - 3.3558s'/\sqrt{n} \leqslant \mu \leqslant \bar{x} + 3.355s'/\sqrt{n} = 0.99$$

$$Pr(33.655 \leqslant \mu \leqslant 38.365) = 0.99$$

ช่วงเชื่อมั่นระดับ 99% สำหรับ μ คือ (33.655, 38.365)

แบบฝึกหัด

1. จงสร้างช่วงเชิงสุ่ม ณ ระดับ 95.45% และ 99.73% ของค่าเฉลี่ยของตัวแปร X ที่มีการแจกแจงแบบปกติ $N(\mu,\sigma^2)$ โดยที่มีเงื่อนไขของค่าเฉลี่ย $\overline{X}$ จากกลุ่มตัวอย่าง n ดังต่อไปนี้

   a. $\overline{X} = 60, n = 100, \sigma^2 = 2500$ c. $\overline{X} = 20, n = 16, \sigma^2 = 100$

   b. $\overline{X} = 30, n = 25, \sigma^2 = 144$ d. $\overline{X} = 10, n = 25, \sigma^2 = 64$

2. โรงงานผลิตตลับลูกปืนแห่งหนึ่งได้ทำการทดสอบขนาดของตลับลูกปืนที่ผลิตได้ โดยการสุ่มตัวอย่างมา 100 หน่วย แล้ววัดค่าเฉลี่ยเส้นผ่าศูนย์กลางของตลับลูกปืนที่สุ่มได้ ปรากฏผลว่าค่าเฉลี่ยเป็น 0.75 นิ้ว ถ้ากำหนดให้ความแปรปรวนของความยาวของเส้นผ่าศูนย์กลางเป็น 0.0025 นิ้ว จงสร้างช่วงเชื่อมั่นเชิงสุ่ม ณ ระดับ 95% และ 99% สำหรับค่าของ μ ของประชากร

3. กำหนดให้ X เป็นอายุของยางยี่ห้อ Fairyear deluxe trie โดยที่ X มีการแจกแจงแบบปกติไม่ทราบค่า μ, $\sigma^2 = 4{,}000$ ไมล์ เพื่อที่จะหาช่วงเชื่อมั่นของอายุของยางชนิดนี้ (μ) ผู้วิจัยได้สุ่มยางมา 100 เส้น แล้วหาอายุเฉลี่ยของยางตัวอย่างนี้พบว่า อายุเฉลี่ยเท่ากับ 30,000 ไมล์ จงหาช่วงเชื่อมั่นของ μ ณ ระดับ 95% ถ้าท่านเป็นผู้จัดการบริษัทแห่งนี้ ท่านกล้ารับประกันอายุการใช้งานของยางว่าจะใช้ทนทานนานกว่า 25,000 ไมล์หรือไม่

4. เกษตรกรผู้หนึ่งมีความประสงค์จะขายไก่ของเขาในราคากิโลกรัมละ 15 บาท เขาสุ่มตัวอย่างไก่มา 40 ตัว แล้วนำมาชั่งน้ำหนักปรากฏว่าได้ค่าเฉลี่ยเป็นน้ำหนัก 3 กิโลกรัม จงสร้างช่วงเชื่อมั่น ณ ระดับ 99% สำหรับรายได้ที่เขาจะได้รับจากการขายไก่ทั้งหมดที่มีอยู่ 10,000 ตัว (ส่วนเบี่ยงเบนมาตรฐานของน้ำหนักไก่คือ 0.25 กิโลกรัม)

กรณีที่ 2 การประมาณช่วงเชื่อมั่นเชิงสุ่มของความแตกต่างระหว่างค่าเฉลี่ยของสองประชากร กรณีที่ 2 นี้คล้ายคลึงกับกรณีที่ 1 เพียงแต่ว่าเราจะต้องศึกษาประชากรเพิ่มขึ้นดังนี้

ให้ X_1 เป็นตัวแปรเชิงสุ่มจากประชากรที่ 1 ซึ่งมีค่าเฉลี่ยและความแปรปรวนของประชากรเป็น μ_1, σ_1^2 ตามลำดับ

X_2 เป็นตัวแปรเชิงสุ่มจากประชากรที่ 2 ซึ่งมีค่าเฉลี่ยและความแปรปรวนของประชากรเป็น μ_2, σ_2^2 ตามลำดับ

โดยที่ X_1 และ X_2 เป็นตัวแปรเชิงสุ่มที่เป็นอิสระต่อกัน ถ้าสุ่มตัวอย่าง n_1 จากประชากรกลุ่มที่ 1 และสุ่มตัวอย่าง n_2 จากประชากรกลุ่มที่ 2 จะได้ค่าเฉลี่ยจากตัวอย่างเป็น $\overline{X}_1, \overline{X}_2$

กำหนดให้ความแตกต่างระหว่างค่าเฉลี่ยจากตัวอย่างทั้งสองคือ $\overline{D}$ นั่นคือ $\overline{D} = \overline{X}_1 - \overline{X}_2$

$\overline{D}$ เป็น Linear combination ของ $\overline{X}_1$ และ $\overline{X}_2$ ดังนั้นการแจกแจงของ $\overline{D}$ จึงขึ้นอยู่กับ $\overline{X}_1$ และ $\overline{X}_2$

ถ้า X_1 และ X_2 มีการแจกแจงแบบปกติจะยังผลให้ $\overline{X}_1$ และ $\overline{X}_2$ แจกแจงแบบปกติด้วย นั่นหมายความว่า $\overline{D}$ จะมีการแจกแจงแบบปกติด้วย

ถ้า X_1 และ X_2 ไม่ได้แจกแจงแบบปกติ แต่ขนาดตัวอย่าง n_1 และ n_2 มีขนาดใหญ่พอก็จะยังผลให้ $\overline{X}_1$ และ $\overline{X}_2$ มีการแจกแจงเข้าสู่การแจกแจงแบบปกติ นั่นคือ $\overline{D}$ จะเข้าสู่การแจกแจงแบบปกติด้วย

$\overline{D}$ เป็นตัวสถิติที่มีค่าเฉลี่ย $\mu_1 - \mu_2 = \delta$

δ คือพารามิเตอร์ที่ไม่ทราบค่า

จุดมุ่งหมายก็คือสร้างค่าประมาณของ δ ขึ้นมา การที่จะสร้างช่วงเชื่อมั่นดังกล่าวของ δ ขึ้นมานั้นต้องอาศัยส่วนเบี่ยงเบนมาตรฐานของ $\overline{D}$ เข้ามาช่วย

$$\sigma^2_{\overline{D}} = \sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \sigma_1^2/n_1 + \sigma_2^2/n_2$$

นั่นก็คือ ถ้า $\overline{X}_1 \sim N(\mu_1, \sigma_1^2)$ และ $\overline{X}_2 \sim N(\mu_2, \sigma_2^2)$

โดยที่ $\overline{X}_1$ และ $\overline{X}_2$ เป็นอิสระต่อกัน จะได้ผลตามมาว่า

$$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

ส่วนเบี่ยงเบนมาตรฐานของ $\overline{D}$

$$\sigma^2_{\overline{D}} = \sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2/n_1 + \sigma^2/n_2$$

ดังนั้นช่วงเชื่อมั่นของ δ ณ ระดับ $(1 - \alpha)$ 100%

$$\Pr(\dot{\overline{D}} - Z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \geqslant \delta < \overline{D} + Z_{\alpha/2}\sqrt{\sigma_1^2 n_1 + \sigma_2^2/n_2}) = 1 - \alpha$$

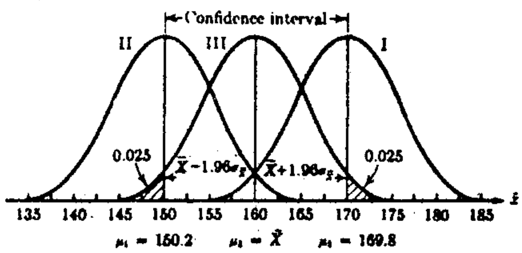
ข้อสังเกตุ 1. n_1 และ n_2 ไม่จำเป็นต้องเท่ากัน ในกรณีที่ X_1 และ X_2 มีการแจกแจงไม่เป็นปกติ ให้ใช้ขนาด n_1 และ n_2 ให้มากพอ คือ อย่างน้อยที่สุดไม่ควรต่ำกว่า 30

2. ในกรณีที่ไม่ทราบค่าความแปรปรวนของประชากร ก็ให้ใช้ความแปรปรวน จากตัวอย่างเป็นตัวทดแทนโดยยึดหลักเช่นเดียวกันกับกรณีประชากรเดียว คือใช้ตัวสถิติสถิติ t แทน

หมายเหตุ จะสังเกตุได้ว่าช่วงเชื่อมั่นที่หาเป็นช่วงเชิงสุ่มก็เพราะว่าทั้งค่าขีดจำกัดล่างและขีดจำกัดบนล้วนแต่ขึ้นอยู่กับค่าสังเกตุทั้งสิ้น การสุ่มตัวอย่างแต่ละครั้งก็จะได้ค่าสังเกตุเป็นชุด ๆ หนึ่ง อาจจะเหมือนหรือแตกต่างกันก็เป็นไปได้ ในกรณีที่ค่าสังเกตุในแต่ละชุดแตกต่างกัน เช่นสุ่มครั้งที่ 1 ได้ X_1 (ขนาด n_1) สุ่มครั้งที่ 2 ได้ X_2 (ขนาด n_2) และสุ่มครั้งที่ 3 ได้ X_3 (ขนาด n_3) ในเมื่อ

$\overline{X}_1 \neq \overline{X}_2 \neq \overline{X}_3$ ก็จะส่งผลให้ช่วงเชิงสุ่มในแต่ละกรณีแตกต่างกันไปได้ (ดูจากรูปประกอบ)

Statistics: methods and analyses



ปัญหา

1. โรงงานผลิตดุมล้อรถแห่งหนึ่ง ได้แยกการผลิตออกเป็น 2 ส่วน โดยที่ทั้งสองส่วนผลิตดุมล้อรถอย่างเดียวกัน แต่เป็นที่น่าสังเกตุว่าจำนวนดุมล้อมที่ผลิตออกมาต่อชั่วโมงจะได้ไม่เท่ากัน จึงมีการศึกษาผลผลิตที่ได้โดยการสุ่มตัวอย่าง โดยสุ่มจำนวนชั่วโมงจากส่วนส่วนที่ 1 มา 64 ชั่วโมงได้ดุมล้อที่ผลิตได้เฉลี่ยชั่วโมงละ 100 อัน จากส่วนที่ 2 สุ่มมา 49 ชั่วโมงได้ดุมล้อเฉลี่ยแล้วชั่วโมงละ 90 อัน โดยที่ความแปรปรวนของดุมล้อที่ผลิตได้จากส่วนที่ 1 และส่วนที่ 2 คือ $\sigma_1^2 = 256, \sigma_2^2 = 196$ ให้สร้างช่วงเชื่อมั่น ณ ระดับ 95% สำหรับค่าของ δ

2. ยานอนหลับชนิดใหม่ได้ผลิตออกมาสู่ท้องตลาด เภสัชกรผู้วิเคราะห์ยาได้ตั้งข้อสังเกตุไว้ว่า ยาดังกล่าวมีผลต่อเพศชายและเพศหญิงแตกต่างกัน จึงได้มีการทดลองกับคนไข้ผู้ใช้ยานี้ ให้ X_1, X_2 เป็นจำนวนชั่วโมงที่หลับ เมื่อใช้ยาของคนไข้ชายและคนไข้หญิงตามลำดับผลการทดลองได้ออกมาดังนี้

	ชาย	หญิง
ขนาดตัวอย่าง	$n_1 = 36$	$n_2 = 64$
ค่าเฉลี่ยจำนวนชั่วโมงที่หลับ	$\overline{X}_1 = 8.75$	$\overline{X}_2 = 7.25$
ความแปรปรวน	$\sigma_1^2 = 9.0$	$\sigma_2^2 = 4.0$

ให้สร้างช่วงเชื่อมั่น ณ ระดับ 95% และ 99% ของ $\mu_1 - \mu_2$

3. บริษัทซ่อมรถแห่งหนึ่งมีแผนกซ่อมเครื่องยนต์และแผนกทาสี ต้องการที่จะศึกษาถึงรายได้ของคนงานในแต่ละแผนก ว่าแตกต่างกันอย่างไร

โดยการสุ่มคนงานในแต่ละแผนก ๆ ละ 100 คน เก็บข้อมูลเกี่ยวกับรายได้ในแต่ละเดือนได้ข้อมูลดังนี้

	แผนกเครื่องยนต์	แผนกสี
ขนาดตัวอย่าง	$n_1 = 100$	$n_2 = 100$
ค่าเฉลี่ย	$\overline{X}_1 = 345$	$\overline{X}_2 = 340$
ความแปรปรวน	$\sigma_1^2 = 196$	$\sigma_2^2 = 204$

จงสร้างช่วงเชื่อมั่น ณ ระดับ 95%, 99% ของ δ (ความแตกต่างในรายได้ของคนงานทั้งสองแผนก)

ตอนที่ 2 ช่วงเชื่อมั่นของสัดส่วน

กรณีที่ 1 ช่วงเชื่อมั่นของสัดส่วนจากประชากรเดียว

ในกรณีที่พารามิเตอร์ของประชากรที่เราสนใจจะศึกษาอยู่ในรูปของสัดส่วนนั้นหมายความว่าเราสนใจจะวิเคราะห์ข้อมูลในเชิงคุณภาพ (quality) มากกว่าข้อมูลเชิงปริมาณ (quantity) ดังในตอนแรก ๆ ที่กล่าวมาแล้ว ข้อมูลพวกนี้เช่น สินค้าที่ผลิตจากโรงงานจะมีคุณภาพได้มาตรฐานเท่าไร จำนวนคนที่ตายด้วยอุบัติเหตุในกรุงเทพฯ ฯลฯ ซึ่งข้อมูลพวกนี้จะสนใจในรูปของสัดส่วน p

กำหนดให้ X เป็นตัวแปรในเรื่องของสัดส่วน X จะมีค่าเป็น 0 หรือ จะเป็น 1 หรือ แล้วแต่ว่าเราจะสนใจจะศึกษาอะไร เช่น สุ่มสินค้ามาจากโรงงานมา 10 ชิ้นจากการตรวจสอบคุณภาพ ถ้าไม่ได้มาตรฐานให้เป็น 0 ได้มาตรฐานให้เป็น 1 เราจะได้ค่าของกลุ่มตัวอย่างนี้

$$X_1, X_2, \ldots, X_{10} = 0, 1, 0, \ldots, 0$$

$$\hat{\pi} = \overline{X} = \Sigma X_i / n = (0 + 1 + \ldots + 1)/10$$

เนื่องจาก X_i จะมีการแจกแจงแบบทวินามมีค่าเฉลี่ยเป็น np (p คือสัดส่วนของประชากรในเรื่องที่ศึกษา)

ดังนั้น $\hat{\pi}$ เป็นตัวแปรสุ่มที่มีค่าเฉลี่ยเป็น p และความแปรปรวนเป็น $p(1-p)/n$

โดยทฤษฎีการเข้าสู่เกณฑ์กลาง (Central Limit Theorem) จะได้ว่า (ถ้าค่า n มีขนาดใหญ่พอ)

$$Z = \frac{\hat{\pi} - p}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} \sim N(0,1)$$

นั่นก็คือเราสามารถสร้างได้ว่า

$$\Pr\left(\hat{\pi} - Z_{\alpha/2}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < p < \hat{\pi} + Z_{\alpha/2}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1 - \alpha$$

ดังนั้นช่วงเชื่อมั่น ณ ระดับ $(1-\alpha)100\%$ ของ p ก็คือ

$$\hat{\pi} \pm Z_{\alpha/2}\sqrt{\hat{\pi}(1-\hat{\pi})/n}$$

ตัวอย่างที่ 1 บริษัทผู้ผลิตเครื่องดูดฝุ่นแห่งหนึ่งต้องการวิจัยว่าแม่บ้านในกรุงเทพฯ นิยมใช้เครื่องดูดฝุ่นมากน้อยเพียงใด โดยการส่งพนักงานไปสัมภาษณ์แม่บ้านมา 100 ราย ผลปรากฏว่ามีแม่บ้านนิยมใช้เครื่องดูดฝุ่น 20 ราย อีก 80 รายไม่นิยมใช้ ถ้าจะสร้างช่วงเชื่อมั่น ณ ระดับ 95% จะได้ค่าอย่างไร

$$Pr(200/100 - 1.96\frac{0.20(0.80)}{100} < p < 20/100 + 1.96\frac{0.20(0.80)}{100}) = 1 - \alpha$$

ช่วงเชื่อมั่น π ณ ระดับ 95% คือ (0.1216, 0.2784)

กรณีที่ใช้หาความแตกต่างของสัดส่วนของสองประชากร จะได้ว่า ช่วงเชื่อมั่น ณ ระดับ $(1 - \alpha)100\%$ ของความแตกต่างของสองประชากรคือ

$$\hat{\pi}_1 - \hat{\pi}_2 \pm Z_{\alpha/2}\left[\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}\right]^{\frac{1}{2}}$$

แบบฝึกหัด

1. การศึกษาถึงชีวิตของสัตว์ในขั้วโลกเหนือ นักวิทยาศาสตร์ได้สุ่มตัวอย่างไข่ของนกกีวีมา 100 ฟอง แล้วนำมาฟักในห้องทดลอง พบว่าไข่ของนกกีวีฟักออกมาเป็นตัวเพียง 90 ฟอง สมมุติว่าเราประมาณได้ว่าในขั้วโลกเหนือมีไข่ของนกกีวีในช่วงระยะเวลาที่ศึกษามีอยู่ 10,000 ฟอง ถามว่าในช่วงระยะเวลาดังกล่าวจะมีลูกนกกีวีกี่ตัว (ใช้ระดับความเชื่อมั่น $(1-\alpha)100\%$)

2. ในการส่งกล้วยหอมเป็นสินค้าขาออกจากประเทศไทยไปยังญี่ปุ่นโดยทางเรือ เราพบว่าหลังจากที่เรือถึงจุดหมายปลายทางแล้ว ผู้รับสินค้าจะตรวจสอบโดยสุ่มกล้วยมา 30 หวี ปรากฏว่าเน่าเสีย 2 หวี จงสร้างช่วงเชื่อมั่น ณ ระดับ 95% ของสัดส่วนของกล้วยเสียทั้งหมดในเรือบรรทุกลำนั้น

7.4

ตอนที่ 3 การหาช่วงเชื่อมั่น ณ ระดับ $(1 - \alpha)100\%$ ของความแปรปรวนของประชากร

กรณีของประชากรเดียว

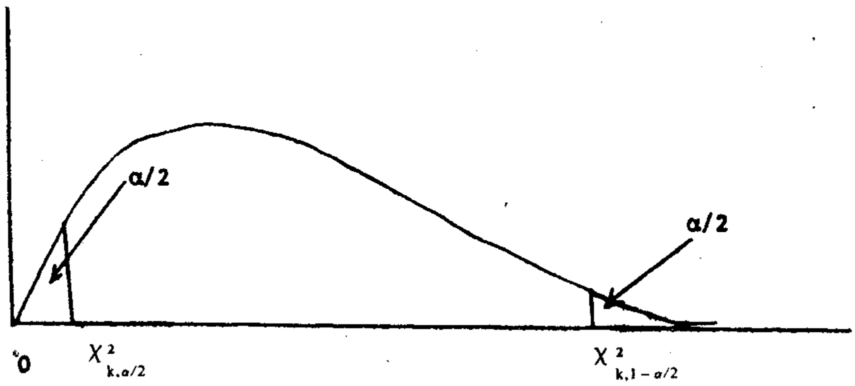
ในส่วนที่เราสนใจจะศึกษาถึงค่าของความแปรปรวนของประชากรคือ σ^2 เราสามารถใช้ตัวสถิติที่เกี่ยวข้องในการสร้างความสัมพันธ์ดังกล่าวได้คือ

$$\chi^2_{(n-1)} = \Sigma(X_i - \overline{X})^2/\sigma^2 = \frac{nS^2}{\sigma^2}$$

ดังนั้น

$$\Pr(\chi^2_{(n-1),\alpha/2} < \frac{ns^2}{\sigma^2} < \chi^2_{(n-1),1-\alpha/2}) = 1 - \alpha$$

$$\Pr\left(\frac{nS^2}{\chi^2_{(n-1),\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{(n-1),1-\alpha/2}}\right) = 1 - \alpha$$



รูป แสดงพื้นที่ของไคกำลังสองโดยการใช้ Central Interval

ในเรื่องการหาช่วงเชื่อมั่น ณ ระดับ $(1 - \alpha)100\%$ สำหรับความแตกต่างของความแปรปรวนระหว่างสองประชากร ให้นักศึกษาลองทำเป็นแบบฝึกหัด โดยการใช้แนวทางของตัวสถิติ F เป็นหลักคือ

$$F_{(n_1,n_2)} = \frac{\chi^2_1/(n_1-1)}{\chi^2_2(n_2-1)}$$

$$= \frac{n_1s_1^2/(n_1-1)}{n_2s_2^2/(n_2-1)}$$