บทที่ 7
ทฤษฎีการประมาณค่า
(ESTIMATION THEORY)

วัตถุประสงค์

เมื่อท่านศึกษาเนื้อหาบทที่ 7 โดยละเอียดแล้ว ควรมีความสามารถ ดังนี้

1. บอกวิธีการประมาณค่าของพารามิเตอร์ได้
2. บอกเกณฑ์สำหรับพิจารณาตัวประมาณค่าที่ดีได้
3. ประมาณค่าเฉลี่ย แบบเป็นจุด และแบบเป็นช่วงได้
4. ประมาณผลต่างของค่าเฉลี่ย แบบเป็นช่วงได้
5. ประมาณค่าความแปรปรวน แบบเป็นจุด และแบบเป็นช่วงได้

เนื้อหา

7.1 ความหมายของการประมาณค่า

7.2 เกณฑ์สำหรับพิจารณาตัวประมาณค่าที่ดี

7.3 การประมาณค่าเฉลี่ยของประชากร

7.4 การประมาณผลต่างของค่าเฉลี่ยของประชากร

7.5 การประมาณค่าความแปรปรวนของประชากร

เนื้อหาที่ 7.1 ความหมายของการประมาณค่า

การประมาณค่าหมายถึงการใช้ข้อมูลจากกลุ่มตัวอย่างในรูปของตัวสถิติ ไปประมาณค่าของพารามิเตอร์ซึ่งเป็นค่าของประชากรว่าควรมีค่าเท่าใด หรืออยู่ในช่วงใด ในการประมาณค่าของพารามิเตอร์ อาจจะทำการประมาณค่าแบบเป็นจุด (point estimation) หรือประมาณค่าแบบเป็นช่วง (interval estimation)

7.1.1 การประมาณค่าเป็นจุด

ในการประมาณค่าพารามิเตอร์ เช่น ประมาณค่าของ μ หรือ σ^2 โดยใช้วิธีประมาณค่าเป็นจุดนั้นเป็นการประมาณค่าโดยกำหนดค่าออกมาเป็นตัวเลขโดด ๆ ซึ่งค่าเหล่านี้จะ

เป็นค่าที่คำนวณได้จากข้อมูลที่ได้จากกลุ่มตัวอย่างซึ่งสุ่มมาจากประชากร เช่นใช้ค่าของ $\bar{X}$ ประมาณค่าของ μ และใช้ S^2 ประมาณค่าของ σ^2

โดยปกติแล้วถ้า θ (อ่านว่า เซต้า - Theta) เป็นพารามิเตอร์ของประชากร ตัวสถิติที่ใช้ในการประมาณค่าจะใช้สัญลักษณ์ $\hat{\theta}$ (อ่านว่า เซต้า แฮ็ท) เมื่อเราประมาณค่าโดยใช้ตัวประมาณค่าเราไม่ได้หวังว่าค่าของมันจะเท่ากันจริง ๆ เพราะความจริงแล้วจะมีความคลาดเคลื่อนเกิดขึ้น แต่เราก็คาดหวังว่าตัวสถิติที่เราเลือกมาเป็นตัวประมาณค่าพารามิเตอร์นั้นจะให้ค่าที่ใกล้เคียงกับค่าพารามิเตอร์

7.1.2 การประมาณค่าเป็นช่วง

การทดสอบสมมุติฐานเป็นเพียงวิธีการหนึ่งที่ใช้ในการวิจัย แต่ในสถานะการณ์บางอย่างจุดประสงค์เบื้องต้นของการรวบรวมข้อมูลไม่ใช่เพื่อทดสอบสมมุติฐาน แต่เพื่อประมาณค่าของพารามิเตอร์บางตัว เช่นเราคำนวณหาค่าเฉลี่ยของกลุ่มตัวอย่าง เพื่อใช้เป็นตัวประมาณค่าของค่าเฉลี่ยของประชากร เพราะค่าเฉลี่ยของกลุ่มตัวอย่างเป็นค่าสถิติที่ไม่ลำเอียง จึงถือว่าเป็นตัวประมาณค่าพารามิเตอร์ที่ดีที่สุด

อย่างไรก็ดีเป็นที่ทราบกันแล้วว่าค่าเฉลี่ยของกลุ่มตัวอย่างเป็นค่าโดยประมาณซึ่งไม่เท่ากับค่าเฉลี่ยจริง ๆ ของประชากรทั้งนี้เนื่องจากความคลาดเคลื่อนในการสุ่มตัวอย่าง ดังนั้นจึงมีความจำเป็นที่จะต้องกำหนดดัชนีของความคลาดเคลื่อนนี้เมื่อต้องการจะประมาณค่าพารามิเตอร์ วิธีการนี้นิยมใช้กันมากที่สุดในการกำหนดดัชนีของความคลาดเคลื่อนทำได้โดยการกำหนดช่วงของความเชื่อมั่น (Confidence interval) ซึ่งเป็นวิธีการกำหนดค่าความน่าจะเป็นสูงสุดและต่ำสุดที่จะครอบคลุมค่าที่แท้จริงของประชากร ขอบเขตของความเชื่อมั่นจะกว้างหรือแคบเพียงใดขึ้นอยู่กับความคลาดเคลื่อนในการสุ่มกลุ่มตัวอย่าง ถ้าการสุ่มตัวอย่างมีความคลาดเคลื่อนมากช่วงของความเชื่อมั่นจะกว้าง นั่นคือผลต่างระหว่างค่าสูงสุดและค่าต่ำสุด (range) จะมีค่ามาก (เช่นค่าเฉลี่ยของประชากรมีค่าอยู่ระหว่าง 60-100) แต่ถ้าการสุ่มตัวอย่างมีความคลาดเคลื่อนน้อย ช่วงของความเชื่อมั่นจะแคบลงมา (เช่นค่าเฉลี่ยของประชากรมีค่าอยู่ระหว่าง 80-90) ซึ่งจะทำให้เห็นค่าที่แท้จริงของประชากรได้ชัดเจนยิ่งขึ้น

เนื้อหาที่ 7.2 เกณฑ์สำหรับพิจารณาตัวประมาณค่าที่ดี

เนื่องจากตัวสถิติ $\hat{\theta}$ ซึ่งใช้ในการประมาณค่าพารามิเตอร์ θ หนึ่งตัวนั้นมีหลายแบบ จึงมีปัญหาว่าเราจะเลือกสถิติตัวใดจึงจะให้ค่าใกล้เคียงกับค่าพารามิเตอร์มากที่สุด เช่น ในการประมาณค่า μ นั้นมีค่าสถิติถึง 3 ตัวที่เราสามารถเลือกมาเป็นตัวประมาณค่าได้ เช่น

ค่าเฉลี่ย (mean) มัธยฐาน (median) หรือฐานนิยม (mode) การจะตัดสินใจเลือกค่าสถิติตัวใดตัวหนึ่งนั้นต้องมีเกณฑ์เป็นเครื่องตัดสินว่า สถิติตัวใดเป็นตัวประมาณค่าที่ดี และสถิติตัวใดเป็นตัวประมาณค่าที่ดีกว่าสถิติอีกตัวหนึ่ง ตัวประมาณค่าที่ดีต้องมีลักษณะสำคัญดังนี้

1) **ไม่ลำเอียง** (Unbised)

$\hat{\theta}$ จะเป็นตัวประมาณค่าที่ไม่ลำเอียงของ θ ก็ต่อเมื่อ $E(\hat{\theta}) = \theta$ นั่นคือค่าเฉลี่ยหรือค่าคาดหวังของการแจกแจงของกลุ่มตัวอย่างของตัวประมาณค่ามีค่าเท่ากับพารามิเตอร์ที่เรากำลังประมาณค่า

จากนิยามข้างบนจะเห็นว่า

$\bar{X}$ เป็นตัวประมาณค่าที่ไม่ลำเอียงของ μ เพราะ $E(\bar{X}) = \mu$ และ s^2 เป็นตัวประมาณค่าที่ไม่ลำเอียงของ σ^2 เพราะ $E(s^2) = \sigma^2$

2) **มีความคงเส้นคงวา** (Consistency)

ตัวประมาณค่าที่มีความคงเส้นคงวาจะมีแนวโน้มที่จะมีค่าเข้าใกล้ค่าของพารามิเตอร์ ถ้าขนาดของกลุ่มตัวอย่างใหญ่ขึ้น นั่นคือ $|\theta - \theta|$ จะต้องมีค่าน้อยลง ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ขึ้น

3) **มีประสิทธิภาพ** (Relative Efficiency)

ถ้ามีสถิติ 2 ตัว ซึ่งสามารถใช้แทนค่าพารามิเตอร์ตัวเดียวนี้ได้แล้ว (เช่น $\bar{X}$, มัธยฐาน) สถิติซึ่งมีค่าความคลาดเคลื่อนมาตรฐานของการแจกแจงของกลุ่มตัวอย่างน้อยกว่าจะเป็นตัวประมาณค่าที่มีประสิทธิภาพมากกว่า สถิติซึ่งมีความคลาดเคลื่อนมาตรฐานของการแจกแจงของกลุ่มตัวอย่างมากกว่า ตัวอย่างเช่น

ค่าความคลาดเคลื่อนมาตรฐานของมัธยฐาน (σ^2_{mdn}) $= \dfrac{\pi\sigma^2}{2n}$

ค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย ($\sigma^2_{\bar{X}}$) $= \dfrac{\sigma^2}{2n}$

เนื่องจาก $\frac{\pi\sigma^2}{2n}$ มีค่ามากกว่า $\frac{\sigma^2}{2n}$ ดังนั้นค่าเฉลี่ย ($\bar{X}$) จึงเป็นตัวประมาณค่าสำหรับ μ ที่มีประสิทธิภาพดีกว่ามัธยฐาน

อนึ่งถ้าการแจกแจงของสถิติ 2 ตัว มีค่าเฉลี่ยเท่ากัน แต่มีความเบี่ยงเบนมาตรฐานไม่เท่ากัน (หรือมีความคลาดเคลื่อนมาตรฐานไม่เท่ากัน) สถิติตัวที่มีความเบี่ยงเบนมาตรฐานน้อยกว่าจะเป็นตัวประมาณค่าที่มีประสิทธิภาพดีกว่า

เนื้อหาที่ 7.3 การประมาณค่าเฉลี่ยของประชากร

7.3.1 การประมาณค่าเฉลี่ยของประชากรแบบเป็นจุด

ในการประมาณค่าเฉลี่ยของประชากรแบบเป็นจุดให้ใช้ค่าเฉลี่ยที่คำนวณได้จากกลุ่มตัวอย่าง ($\bar{X}$) เป็นตัวประมาณค่า

ตัวอย่าง 7.1 สุ่มตัวอย่างมา 50 คนจากประชากรที่มีการแจกแจงปกติ ทำการทดสอบความถนัดทางภาษาได้ ค่าเฉลี่ย ($\bar{X}$) เท่ากับ 36.0 และส่วนเบี่ยงเบนมาตรฐาน (S) เท่ากับ 4.0 ให้ประมาณค่าเฉลี่ยของประชากร (μ) แบบเป็นจุด

วิธีทำ เนื่องจากค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{X}$) = 36.0

ฉะนั้นประมาณค่าเฉลี่ยของประชากร (μ) ได้ = 36.0

7.3.2 การประมาณค่าเฉลี่ยของประชากรแบบเป็นช่วง

7.3.2.1 กรณีที่รู้ค่าของ σ หรือไม่รู้ค่าของ σ แต่กลุ่มตัวอย่างมีจำนวนมากกว่า 30

7.3.2.2 กรณีที่ไม่รู้ค่า σ และกลุ่มตัวอย่างมีจำนวนน้อยกว่าหรือเท่ากับ 30

การประมาณค่าทั้ง 2 กรณีมีวิธีการ ดังนี้

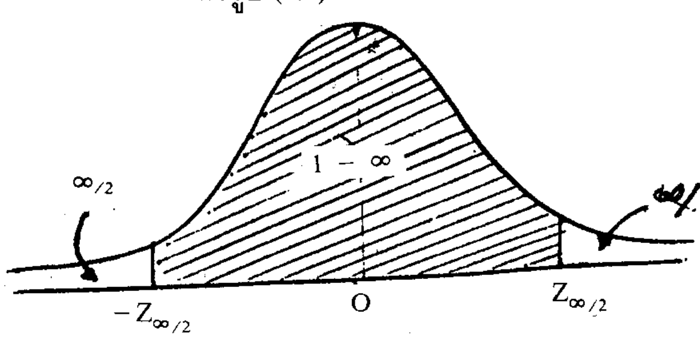
7.3.2.1 กรณีที่รู้ค่าของ σ หรือ ถ้าไม่รู้ค่าของ σ แต่กลุ่มตัวอย่างมีจำนวนมากกว่า 30 มีวิธีการประมาณช่วงของค่าเฉลี่ยดังนี้

ถ้าประชากรมีค่าเฉลี่ยเป็น μ และมีความเบี่ยงเบนมาตรฐานเป็น σ เราใช้ค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{X}$) เป็นค่าสถิติในการประมาณค่า μ

จากการแจกแจงปกติเราสามารถหาค่า Z ซึ่งมีพื้นที่ใต้โค้งเท่ากับ $1-\infty$ ได้ดังนี้

$$P(-Z_{\infty/2} < Z < Z_{\infty/2}) = 1 - \infty$$

แสดงให้เห็นได้ชัดเจนดังรูป (7.1)



รูป 7.1 *แสดงว่า* $P(-Z_{\infty/2} < Z < Z_{\infty/2}) = 1 - \infty$

เนื่องจากเราทราบว่า $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ ดังนั้น

$$P\left(-Z_{\infty/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < Z_{\infty/2}\right) = 1 - \infty$$

$$P\left(-Z_{\infty/2} \cdot \sigma/\sqrt{n} < \mu < Z_{\infty/2} \cdot \sigma/\sqrt{n}\right) = 1 - \infty$$

เอา $\overline{X}$ ลบสมการภายในวงเล็บแล้วคูณตลอดด้วย -1 จะได้

$$P\left(\overline{X} - Z_{\infty/2} \cdot \sigma/\sqrt{n} < \mu < \overline{X} + Z_{\infty/2} \cdot \sigma/\sqrt{n}\right) = 1 - \infty$$

ฉะนั้น ช่วงของความเชื่อมั่นที่ $(1 - \infty)$ 100% สำหรับ μ คือ

$$\overline{X} - Z_{\infty/2} \cdot \sigma/\sqrt{n} < \mu < \overline{X} + Z_{\infty/2} \cdot \sigma/\sqrt{n} \qquad \text{........................(7.1)}$$

เมื่อ n คือจำนวนกลุ่มตัวอย่าง

ในกรณีที่ไม่รู้ค่าของ σ แต่ถ้ากลุ่มตัวอย่างมีจำนวนมากกว่า 30 ก็อาจใช้ค่าของความเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (s) แทนค่า σ ได้ ในการประมาณช่วงดังกล่าว

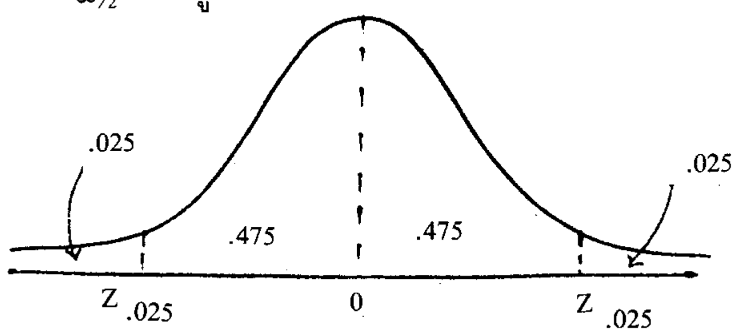
ตัวอย่าง 7.2 ค่าเกรดเฉลี่ยและความเบี่ยงเบนมาตรฐานของเกรดของนักศึกษาปีที่ 1 จำนวน 49 คน คือ 2.4 และ 0.5 ตามลำดับ จงหาช่วงของความเชื่อมั่น 95% และ 99% สำหรับเกรดเฉลี่ยของนักศึกษาปีที่ 1 ทั้งหมด

วิธีทำ

จากโจทย์ $n = 49$ $\overline{X} = 2.4$ $s = 0.5$

1) เมื่อโจทย์กำหนดให้ $1 - \infty = .95$ ฉะนั้น $\infty = .05$

หาค่า $Z_{\infty/2}$ ได้ดังรูป



เนื่องจากพื้นที่ใต้โค้งจาก 0 ถึง $Z_{.025}$ เท่ากับ .475 (แบ่ง .95 ออกเป็น 2 ส่วน) หาค่า Z จากตารางในภาคผนวกว่า Z จะมีค่าเท่าไร โดยดูตารางย้อนกลับ จะพบว่า $z_{.025}$ มีค่าเท่ากับ 1.96 ดังนั้น

ช่วงของความเชื่อมั่นที่ 95% จะเป็น

$$2.4 - (1.96)\left(\frac{0.5}{\sqrt{49}}\right) < \mu < 2.4 + (1.96)\left(\frac{0.5}{\sqrt{49}}\right)$$

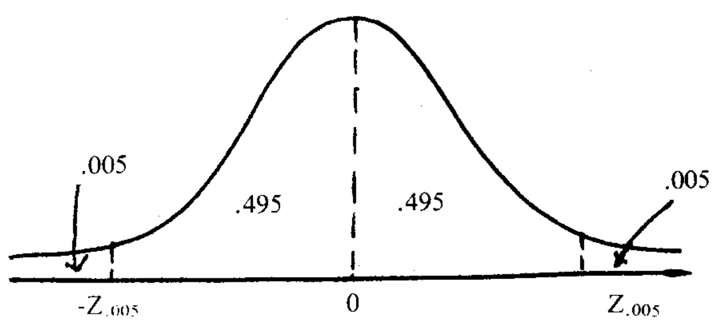
ได้ $2.4 - .14 < \mu < 2.4 + .14$

ได้ $2.26 < \mu < 2.54$

นั่นคือ ช่วงของความเชื่อมั่นที่ 95% สำหรับเกรดเฉลี่ยของนักศึกษาปี 1 ทั้งหมดจะอยู่ระหว่าง 2.26 ถึง 2.54 แสดงว่าเรามีความมั่นใจถึง 95% ว่า μ จะมีค่าอยู่ระหว่าง 2.26 ถึง 2.54

2) ถ้าโจทย์กำหนดให้ $1 - \infty = .99$ ฉะนั้น $\infty = .01$

หาค่า $Z_{\infty/2} = Z_{.005}$ หาค่าได้ดังรูป



เปิดค่า $Z_{.005}$ จากตารางได้ค่าเท่ากับ 2.57

ดังนั้น ช่วงของความเชื่อมั่น 99% จึงเป็น

$$2.4 - (2.57)\left(\frac{.05}{\sqrt{49}}\right) < \mu < 2.4 + (2.57)\left(\frac{.05}{\sqrt{49}}\right)$$

ได้ $2.4 - .18 < \mu < 2.4 + .18$

ได้ $2.22 < \mu < 2.58$

ดังนั้น ช่วงของความเชื่อมั่น 99% สำหรับเกรดเฉลี่ยของนักศึกษาปีที่ 1 ทั้งหมดจะอยู่ระหว่าง 2.22 ถึง 2.58 แสดงว่าเรามีความมั่นใจถึง 99% ว่า μ จะมีค่าอยู่ระหว่าง 2.22 ถึง 2.58

7.3.2.2 กรณีที่ไม่รู้ค่า σ และ $n \leqslant 30$ มีวิธีการประมาณช่วงของค่าเฉลี่ยดังนี้

ในกรณีที่ไม่รู้ค่า σ และกลุ่มตัวอย่างมีจำนวนน้อยกว่าหรือเท่ากับ 30 เราใช้การแจกแจงของที (t -distribution) ในการหาช่วงความเชื่อมั่นเมื่อ

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \quad , \quad df = n-1$$

เราจะได้

$$P(-t_{\infty/2} < \frac{\overline{X} - \mu}{s/\sqrt{n}} < t_{\infty/2}) = 1-\infty$$

โดยวิธีเดียวกันกับการแจกแจงของ Z เราสามารถหาช่วงของความเชื่อมั่นที่ $(1-\infty)$ 100% ได้เป็น

$$\overline{X} - t_{\infty/2} \cdot \frac{s}{\sqrt{n}} < \mu < + t_{\infty/2} \cdot \frac{s}{\sqrt{n}} \qquad \text{.......(7.2)}$$

ตัวอย่าง 7.3 สมมุติว่าความสามารถเฉลี่ยในวิชาความถนัดทางภาษาของนักเรียนชั้นม. 1 จำนวน 25 คน เท่ากับ 58 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 อยากทราบว่าช่วงความสามารถเฉลี่ยของประชากรเป็นเท่าใด ถ้ากำหนดให้ช่วงของความเชื่อมั่นเป็น 95%

วิธีทำ

กำหนดให้ $n = 25 \quad \overline{X} = 58 \quad s = 10$

$$df = n-1 = 25 - 1 = 24$$

เมื่อ $1-\infty = .95$ ฉะนั้น $\infty/2 = .025$

เปิดตารางหาค่า $t_{.025}$ ที่ df = 24 ได้ค่า t = 2.064

ดังนั้นช่วงของความเชื่อมั่น 95% ของ μ คือ

$$58-(2.064)\left(\frac{10}{\sqrt{25}}\right) < \mu < 58 + (2.064)\left(\frac{10}{\sqrt{25}}\right)$$

ได้ $58-4.13 < \mu < 58 + 4.13$

ได้ $53.87 < \mu \; 62.13$

นั่นคือเรามีความมั่นใจถึง 95% ว่า ความสามารถเฉลี่ยของประชากรมีค่าอยู่ระหว่าง *53.87 ถึง 62.13 คะแนน*

เนื้อหาที่ 7.4 การประมาณผลต่างของค่าเฉลี่ยของประชากร

7.4.1 กรณีที่รู้ค่าของ σ หรือไม่รู้ค่าของ σ แต่กลุ่มตัวอย่างมีจำนวนมากกว่า 30

7.4.2 กรณีที่ไม่รู้ค่าของ σ และกลุ่มตัวอย่างมีจำนวนน้อยกว่าหรือเท่ากับ 30

การประมาณค่าทั้ง 2 กรณี มีวิธีการ ดังนี้

7.4.1 กรณีที่รู้ค่าของ σ หรือไม่รู้ค่าของ σ แต่กลุ่มตัวอย่างมากกว่า 30

ถ้าประชากร 2 กลุ่ม มีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานเป็น μ_1, μ_2 และ σ_1, σ_2 ตามลำดับ และถ้าเราสุ่มตัวอย่างมาจำนวน n_1 และ n_2 คน จากประชากร 2 กลุ่มนี้ ค่าสถิติ $X_1 - X_2$ จะเป็นตัวประมาณค่าที่ดีและมีประสิทธิภาพมากที่สุดของ $\mu_1 - \mu_2$ ฉะนั้นเราจึงใช้การแจกแจงของ $\bar{X}_1 - \bar{X}_2$ ในการหาช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$

เนื่องจาก $Z = \dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ มีการแจกแจงปกติ

ดังนั้น $P(-Z_{\infty/2} < Z < Z_{\infty/2}) = 1-\infty$

แทนค่า Z ในสมการจะได้

$$P(-z_{\infty/2} < \frac{(\bar{X}_1-\bar{X}_2)(\mu_1-\mu_2)}{a:/_{,,} + \sigma_{2/n2}^2} Z_{\infty/2}) = 1-\infty$$

ฉะนั้นช่วงของความเชื่อมั่นที่ $(1-\infty)$ 100% ของ $\mu_1-\mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) - Z_{\propto/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1-\bar{X}_2) + Z_{\propto/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{....(7.3)}$$

ถ้าเราไม่รู้ค่า σ_1 และ σ_2 แต่กลุ่มตัวอย่างมีขนาดใหญ่ นั่นคือมีจำนวนมากกว่า 30 แล้ว เราก็อาจใช้ค่า S_1, S_2 แทน σ_1 และ σ_2 ได้ ในการหาช่วงของความเชื่อมั่นของ $\mu_1 - \mu_2$ ดังนั้นช่วงของความเชื่อมันที่ $(1-\infty)$ 100% ของ $\mu_1 - \mu_2$ จึงเปลี่ยนเป็น

$$(\bar{X}_1-\bar{X}_2)-Z_{\infty/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2) + Z_{\infty\,2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

ตัวอย่าง 7.4 นักเรียนชั้นประถมกลุ่มละ 50 คน 2 กลุ่ม ได้รับการสอนให้อ่านด้วยวิธีสอนที่แตกต่างกัน 2 วิธี หลังจากทำการสอนจบหลักสูตรแล้ว ได้ทำการทดสอบความสามารถในการอ่านของนักเรียน 2 กลุ่มนี้ ผลการสอบ เป็นดังนี้

$$\bar{X}_1 = 74,\ \bar{X}_2 = 71,\ s_1 = 9,\ s_2 = 10$$

จงหาช่วงของความเชื่อมั่นที่ 95% ประมาณค่าความแตกต่างนี้

วิธีทำ

$$Z_{\infty/2} = Z_{.025} = 1.96$$

ดังนั้น ช่วงของความเชื่อมั่นที่ 95% ของความแตกต่างของค่าเฉลี่ยของประชากร 2 กลุ่มคือ

$$(74-71)-(1.96)\left(\sqrt{\frac{9^2}{50}+\frac{10^2}{50}}\right) < \mu_1-\mu_2 < (74-71)-(1.96)\left(\sqrt{\frac{9^2}{50}+\frac{10^2}{50}}\right)$$

ได้ $3-(1.96)(\sqrt{3.62}) < \mu_1-\mu_2 < 3 + (1.96)(\sqrt{3.62})$

ได้ $3-(1.96)(1.90) < \mu_1-\mu_2 < 3 + (1.96)(1.90)$

ได้ $3-3.72 < \mu_1-\mu_2 < 3 \text{ t } 3.12$

ได้ $-.72 < \mu_1-\mu_2 < 6.12$

นั่นคือเรามีความมั่นใจ 95% ว่า ความแตกต่างของคะแนนเฉลี่ยของประชากร 2 กลุ่มนี้มีค่าอยู่ระหว่าง $-.72$ ถึง 6.72

7.4.2 กรณีที่ไม่รู้ค่าของ σ และกลุ่มตัวอย่างมีจำนวนน้อยกว่าหรือเท่ากับ 30

ในกรณีที่กลุ่มตัวอย่างมีจำนวนน้อยกว่าหรือเท่ากับ 30 ให้ใช้การแจกแจงของที (t - distribution) ในการหาช่วงของความเชื่อมั่นของความแตกต่างของค่าเฉลี่ยของประชากร 2 กลุ่มเมื่อ

$$t = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad , df = n_1 + n_2 - 2$$

เมื่อ $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

ดังนั้นช่วงของความเชื่อมั่นที่ $(1-\infty)\,100\%$ ของ $\mu_1-\mu_2$ คือ

$$(\bar{X}_1-\bar{X}_2) \pm t_{\infty/2} \cdot \sqrt{s_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \text{.......(7.4)}$$

ตัวอย่าง 7.5 ในการทดลองใช้ตำราเรียน 2 เล่ม กับนักเรียน 2 กลุ่ม ๆ ละ 20 คน นั้น หลังจากเรียนจบหลักสูตรแล้วได้ใช้ข้อสอบมาตรฐานทำการทดสอบความสามารถในการเรียนของนักเรียน 2 กลุ่มนี้ ข้อมูลข้างล่างคือผลจากการสอบ ให้หาช่วงของความเชื่อมั่นที่ 95% ของความแตกต่างของค่าเฉลี่ยของประชากรทั้ง 2 กลุ่ม

ผลจากการทดสอบมีดังนี้

$\bar{X}_1 = 28,\ \bar{X}_2 = 23,\ s_1^2 = 120,\ s_2^2 = 80$

วิธีทำ

$$df = n_1 + n_2 - 2 = 20 + 20 - 2 = 38$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(19)(120) + (19)(80)}{38}$$

$$= 100$$

$t_{\infty/2} = t_{.025}$ เปิดตารางหาค่า $t_{.025}$ ที่ degree of freedom 38 ได้ค่า $t = 2.02$

ดังนั้นช่วงของความเชื่อมั่นที่ 95% ของ $\mu_1-\mu_2$ คือ

$$(28-23) \pm (2.02) \cdot \sqrt{100\left(\frac{1}{20} + \frac{1}{20}\right)}$$

ได้ $5 \pm (2.02)(\sqrt{10})$

ได้ 5 ± 6.39

นั่นคือ $-1.39 < \mu_1 - \mu_2 < 11.39$

แสดงว่าเรามีความมั่นใจ 95% ว่า ความแตกต่างของคะแนนเฉลี่ยของประชากร 2 กลุ่มนี้มีค่าอยู่ระหว่าง -1.39 ถึง 11.39

เนื้อหาที่ 7.5 การประมาณค่าความแปรปรวนของประชากร

7.5.1 การประมาณค่าความแปรปรวนของประชากรแบบเป็นจุด

ในการประมาณค่าความแปรปรวนของประชากรแบบเป็นจุด ให้ใช้ค่าความแปรปรวนที่คำนวณได้จากกลุ่มตัวอย่าง (s^2) เป็นตัวประมาณค่า

ตัวอย่าง 7.6 สุ่มตัวอย่างมา 50 คน จากประชากรที่มีการแจกแจงปกติ ทำการทดสอบความถนัดของภาษาได้ค่าเฉลี่ย ($\bar{X}$) เท่ากับ 36.0 และส่วนเบี่ยงเบนมาตรฐาน (s) เท่ากับ 4.0 ให้ประมาณค่าความแปรปรวนของประชากร (σ^2) แบบเป็นจุด

วิธีทำ เนื่องจาก $s = 4.0$

ดังนั้น $s^2 = 16.0$

ฉะนั้นประมาณค่าความแปรปรวนของประชากร (σ^2) = 16.0

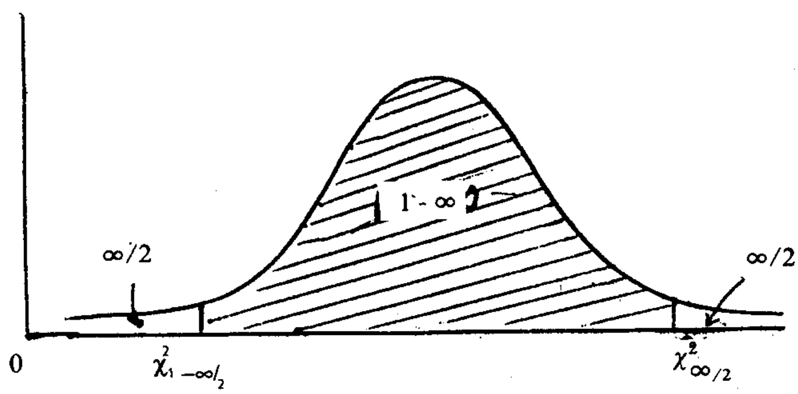
7.5.2 การประมาณค่าความแปรปรวนของประชากรแบบเป็นช่วง

ถ้าประชากรมีการแจกแจงปกติ และมีความแปรปรวนเป็น σ^2 แล้ว s^2 จะเป็นตัวประมาณค่าที่ดีที่สุดสำหรับ σ^2 ในการประมาณช่วงของพารามิเตอร์ σ^2 เราทำได้โดยใช้สถิติ ไค-สแควร์ (χ^2) เมื่อ

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2}, \text{ ชั้นแห่งความอิสระ (df) } = n-1$$

ถ้ากำหนดช่วงของความเชื่อมั่นเป็น $1 - \infty$ เราจะได้

$P(\chi^2_{1-\infty/2} < \chi^2 < \chi^2_{\infty/2}) = 1 - \infty$ ซึ่งอาจแสดงให้เห็นได้ชัดเจน ดังรูปข้างล่าง



(หมายเหตุ : ค่า χ^2_a หมายความถึงค่า χ^2 ที่มีพื้นที่ใต้โค้งตั้งแต่จุดนั้นไปทางขวาเท่ากับ a)

จาก $P(\chi^2_{1-\infty/2} < \chi^2 < \chi^2_{\infty/2}) = 1-\infty$

แทนค่า $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ จะได้

$$P\left(\chi^2_{1-\infty/2} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\infty/2}\right) = 1-\infty.$$

กลับเศษเป็นส่วนของทุกเทอมในวงเล็บแล้วคูณตลอดด้วย $(n-1)s^2$ จะได้

$$P\left(\frac{(n-1)s^2}{\chi^2_{\infty/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\infty/2}}\right) = 1-\infty$$

ฉะนั้นช่วงของความเชื่อมั่นที่ $(1-\infty)\,100\%$ ของ σ^2 คือ

$$\frac{(n-1)s^2}{\chi^2_{\infty/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\infty/2}} \qquad \text{.......(7.5)}$$

ตัวอย่าง 7.7 สมมุติว่าเราต้องการจะประมาณค่าความแปรปรวนของคะแนนของประชากรในวิชาสถิติเบื้องต้น ซึ่งมีผู้เข้าสอบวิชานี้ 31 คน มี $s^2 = 60$ ที่ช่วงของความเชื่อมั่น 90%

วิธีทำ

$$df = n-1 = 31-1 = 30$$

เนื่องจาก σ^2 มีค่าอยู่ระหว่าง

$$\frac{(n-1)s^2}{\chi^2_{n-1;\,\infty/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1;\,1-\infty/2}}$$

$$\frac{(30)(60)}{\chi^2_{30;(.05)}} < \sigma^2 < \frac{(30)(60)}{\chi^2_{30;(.95)}}$$

เนื่องจาก $\chi^2_{.05}$ ที่ degree of freedom 30 มีค่า 43.773
และ $\chi^2_{.95}$ ที่ degree of freedom 30 มีค่า 18.493 ดังนั้นจะได้

$$\frac{(30)(60)}{43.773} < \sigma^2 < \frac{(30)(60)}{18.493}$$

$$41.12 < \sigma^2 < 97.33$$

ฉะนั้นเราสามารถจะประมาณค่าความแปรปรวนของประชากรว่าตกอยู่ระหว่างค่า 41.12 ถึง 97.33 ที่ช่วงของความเชื่อมั่น 90%

สรุปเนื้อหาบทที่ 7

1. การประมาณค่า หมายถึงการใช้ข้อมูลจากกลุ่มตัวอย่างในรูปของตัวสถิติไปประมาณค่าของพารามิเตอร์ว่าควรมีค่าเท่าใด หรืออยู่ในช่วงใด ในการประมาณค่าของพารามิเตอร์อาจประมาณค่าแบบเป็นจุด หรือ แบบเป็นช่วง
2. การประมาณค่าของพารามิเตอร์แบบเป็นจุด ให้ใช้ค่าสถิติที่ได้จากกลุ่มตัวอย่างเป็นตัวประมาณค่า
3. ตัวประมาณค่าที่ดีต้อง **ไม่ลำเอียง มีความคงเส้นคงวา** และ **มีประสิทธิภาพ**
4. ในการประมาณค่าเฉลี่ยของประชากร (μ) เมื่อกลุ่มตัวอย่างมีจำนวนมากกว่า 30 นั้น ค่าเฉลี่ยของประชากรจะมีค่าอยู่ระหว่าง $\bar{X} \pm Z_{\infty/2} \cdot \frac{\sigma}{\sqrt{n}}$ หรือ เขียนได้เป็น

   $$\bar{X} - Z_{\infty/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\infty/2} \cdot \frac{\sigma}{\sqrt{n}}$$

5. ในการประมาณค่าเฉลี่ยของประชากร (μ) เมื่อกลุ่มตัวอย่างมีจำนวนน้อยกว่า 30 นั้น ค่าเฉลี่ยของประชากรจะมีค่าอยู่ระหว่าง $\bar{X} \pm t_{\infty/2} \cdot s/\sqrt{n}$ หรือเขียนได้เป็น

   $$X - t_{\infty/2} \cdot s/\sqrt{n} < \mu < \bar{X} + t_{\infty/2} \cdot s/\sqrt{n}$$

6. ในการประมาณผลต่างของค่าเฉลี่ยของประชากรนั้น ในกรณีที่กลุ่มตัวอย่างมีจำนวนมากกว่า 30 ผลต่างของค่าเฉลี่ยมีค่าอยู่ระหว่าง

   $$(X_1 - X_2) \pm Z_{\infty/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

   แต่ถ้าจำนวนกลุ่มตัวอย่างน้อยกว่า 30 ผลต่างของค่าเฉลี่ยจะมีค่าอยู่ระหว่าง

   $$(X_1 - X_2) \pm t_{\infty/2} \cdot \sqrt{s_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

7. การประมาณค่าความแปรปรวนของประชากรแบบเป็นช่วง ความแปรปรวนจะมีค่าอยู่ระหว่าง

   $\frac{(n-1)s^2}{\chi^2_{\infty/2}}$ กับ $\frac{(n-1)s^2}{\chi^2_{1-\infty/2}}$ หรือเขียนใหม่ได้ดังนี้

   $$\frac{(n-1)s^2}{\chi^2_{\infty/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\infty/2}}$$

คำถามท้ายบทที่ 7

1. นักศึกษา 100 คน มีน้ำหนักเฉลี่ย 45.5 กิโลกรัม ความเบี่ยงเบนมาตรฐาน 3.2 กิโลกรัม จงหาว่าน้ำหนักเฉลี่ยของนักศึกษาทั่ว ๆ ไปจะตกอยู่ในช่วงของน้ำหนัก เท่าใดที่ระดับความเชื่อมั่น 99% และถ้าประมาณค่าน้ำหนักเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของน้ำหนักของนักศึกษาทั่ว ๆ ไปเป็นจุดแล้ว นักศึกษาทั่ว ๆ ไปมีน้ำหนักเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่าใด
2. สุ่มตัวอย่างมา 50 คน จากประชากรที่มีการแจกแจงปกติ หาค่า X ได้เท่ากับ 35.4, S = 3.5 จงหา 95% confidence interval ของ μ
3. จากการสำรวจการใช้จ่ายประจำวันของนักเรียนชั้น ม. 1 จำนวน 35 คน พบว่ามีค่าใช้จ่ายเฉลี่ยเท่ากับ 18 บาท มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.25 บาท ให้หาช่วงของความเชื่อมั่น (confidence interval) ที่ 95% ของค่าใช้จ่ายเฉลี่ยของนักเรียนชั้น ม. 1 ทั้งหมด
4. ถ้าอายุใช้งานของหลอดไฟนีออนมีการแจกแจงปกติ สุ่มตัวอย่างหลอดไฟมา 45 หลอด พบว่าหลอดไฟมีอายุใช้งานโดยเฉลี่ย 800ชั่วโมง มีส่วนเบี่ยงเบนมาตรฐาน 35 ชั่วโมง จงหาช่วงของความเชื่อมั่นที่ 95% ของอายุใช้งานเฉลี่ยของหลอดไฟ
5. ผลการทดสอบวิชาภาษาไทยของนักเรียน ม. 3 จำนวน 25 คนมีดังนี้

   50 28 12 30 32 35 46 37 50 28 32

   13 15 50 29 24 30 36 45 40 48 18

   25 25 20

   ก. ให้กะประมาณค่าเฉลี่ยของผลการสอบวิชาภาษาไทยของนักเรียนชั้น ม. 3 ทั้งหมดแบบเป็นจุด

   ข. ให้กะประมาณค่าเฉลี่ยแบบเป็นช่วงที่ระดับความมั่นใจ 99%

   ค. ให้หา 95% confidence interval ของ σ^2
6. ประชากร 2 กลุ่มซึ่งมีการแจกแจงปกติ มีความแปรปรวนเป็น 9 และ 25 ตามลำดับ ให้หา 95% confidence interval ของ $\mu_1 - \mu_2$ ถ้าสุ่มตัวอย่างมากลุ่มละ 30 คน และแต่ละกลุ่มมีความสามารถทางคณิตศาสตร์เฉลี่ยเท่ากับ 75 และ 80 ตามลำดับ
7. สุ่มตัวอย่างจากประชากร 2 กลุ่มซึ่งมีการแจกแจงปกติมากลุ่มละ 15 คน หาค่าเฉลี่ยที่

ที่ได้จากการทดสอบวิชาภาษาไทยได้เท่ากับ 60 และ 64 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7 และ 10 ตามลำดับ จงหา 99% confidence interval ของ $\mu_1 - \mu_2$

8. ปลูกข้าว 2 พันธุ์บนเนื้อที่ 25 ไร่ ข้าวพันธุ์แรกได้ผลผลิตเฉลี่ยไร่ละ 75 ถัง ส่วนเบี่ยงเบนมาตรฐาน 6 ถัง ส่วนข้าวพันธุ์ที่ 2 ได้ผลผลิตเฉลี่ยไร่ละ 82 ถัง ส่วนเบี่ยงเบนมาตรฐาน 6.5 ถัง จงหา 95% confidence interval ของ $\mu_1 - \mu_2$

9. ปลากระป๋องชนิดหนึ่งจำนวน 10 กระป๋อง มีน้ำหนักดังนี้ (หน่วยเป็นออนซ์) 8.5, 9.4, 8.0, 8.6, 7.9, 9.3, 9.7, 8.7, 8.5, 8.6 จงหา

   ก. 99% confidence interval ของน้ำหนักเฉลี่ยของปลากระป๋องทั่ว ๆ ไป

   ข. 95% confidence interval ของน้ำหนักเฉลี่ยของปลากระป๋องทั่ว ๆ ไป

   ค. 99% confidence interval ของ σ^2

   ง. 95% confidence interval ของ σ^2