ช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่ม เมื่อไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่ม

The Confidence Interval for Ratio of Means of Two Lognormal Distributions when One Population Mean and Variance are Unknown

พรพชร บุนนาค (Ponpachara Boonnak)* ดร.วรารฤทธิ์ พานิชกิจโกศลกุล (Dr.Wararit Panichkitkosolkul)**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อปรับปรุงช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติของสองประชากร โดยกำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนหนึ่งกลุ่ม สำหรับวิธีพาราเมตริกบูตสแตรป (Parametric bootstrap method) ซึ่งเป็นวิธีใหม่ที่แตกต่างและมีประสิทธิภาพกว่าวิธีพาราเมตริกบูตสแตรปในอดีต ผลลัพธ์จากการจำลองโดยวิธีมอนติคาร์โลแสดงให้เห็นถึงช่วงความเชื่อมั่นโดยวิธีใหม่ที่กำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนหนึ่งกลุ่มให้ค่าความน่าจะเป็นคุ้มรวมเข้าใกล้ระดับความเชื่อมั่นที่กำหนดในบางกรณีและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่สั้นกว่าวิธีการประมาณแบบพาราเมตริกบูตสแตรปเดิมในทุกๆกรณี

ABSTRACT

The objective of this research is to improve confidence interval for the ratio of means of two lognormal distributions of the old method, which is a new way to different and more effective methods than in parametric bootstrap method, we investigate by one population mean and variance are unknown. This research finding the confidence interval performed using the parametric bootstrap method. Monte Carlo simulations results indicate that the new confidence interval for the ratio of means of two lognormal distributions when one population mean and variance are unknown give coverage probability close to confidence level in some case and give the shorter average length than that of the parametric bootstrap method in all case.

คำสำคัญ: ช่วงความเชื่อมั่น การแจกแจงแบบล็อกปกติ ความน่าจะเป็นคุ้มรวม

Keywords: Confidence interval, Log-normal distribution, Coverage probability

** นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*

*** รองศาสตราจารย์ สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*

บทนำ

โดยปกติการอนุมานทางสถิติแบบช่วงของค่าเฉลี่ยของประชากร 2 กลุ่มขึ้นไปนั้น สถิติทดสอบที่ใช้มักจะเป็นแบบใช้พารามิเตอร์ ซึ่งข้อมูลที่นำมาประมาณค่าช่วงความเชื่อมั่นจะต้องอยู่ภายใต้ข้อตกลงเบื้องต้น คือ ประชากรมีการแจกแจงแบบปกติ สถิติทดสอบที่นิยมใช้กันอย่างมากเมื่อเราทราบค่าแปรปรวนของประชากร คือ สถิติทดสอบซี (Z-test) และสำหรับการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อทราบความแปรปรวนของตัวอย่างจะใช้สถิติทดสอบที (T-test) และในบางประชากรที่เราสนใจอาจเป็นการประมาณค่าแบบช่วงของค่าอัตราส่วนของค่าเฉลี่ยของประชากร 2 กลุ่ม ภายใต้ข้อมูลที่มีการแจกแจงแบบปกติ แต่บางครั้งในทางปฏิบัติประชากรที่เราสนใจศึกษาอาจไม่ได้มีการแจกแจงแบบปกติเช่นข้อมูลการสำรวจและคำนวณค่าดรรชนี SPI ในพื้นที่ลุ่มแม่น้ำเพิร์ล (Zhang , 2009) ในสาธารณรัฐประชาชนจีนโดยข้อมูลนี้จะมีการแจกแจงแบบล็อกปกติ (Log-normal Distribution) ดังนั้นในการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ยนั้นจึงไม่สามารถประมาณได้ด้วยวิธีการแจกแจงแบบปกติ แต่ก็ได้มีผู้วิจัยและคิดค้นวิธีในการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ยสำหรับประชากรที่มีการแจกแจงแบบล็อกปกติอยู่เรื่อยมา

การแจกแจงแบบล็อกปกติเป็นการแจกแจงหนึ่งที่ข้อมูลประชากรที่สนใจไม่ได้มีการแจกแจงแบบปกติในทางทฤษฎีความน่าจะเป็น การแจกแจงแบบล็อกปกติ เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่อง ของตัวแปรสุ่มซึ่งเป็นการใส่ล็อกการิทึมในการแจกแจงแบบปกติ ซึ่งตัวอย่างประชากรที่มีการแจกแจงแบบล็อกปกตินั้นสามารถประยุกต์ใช้อย่างแพร่หลายในหลายๆ องค์กร ไม่ว่าจะเป็นงานในทางด้านการแพทย์ เช่น การเปรียบเทียบการใช้เซรุ่มต่างๆ หรือ เวลาที่เชื้อโรคจะแพร่กระจายสู่ผู้ป่วย งานทางด้านวิศวกรรมศาสตร์ เช่น ระยะเวลาซ่อมของระบบไฟฟ้าจากเหตุไฟฟ้าขัดข้อง งานทางด้านประกันภัย เช่น ค่าชดใช้ความเสียหายของผู้เอาประกัน งานด้านเศรษฐศาสตร์ เช่น ระยะเวลาการจ้างพนักงาน งานด้านอุตุนิยมวิทยาและธรณีวิทยา เช่น ระดับความสูงของน้ำทะเล ณ ช่วงเวลาหนึ่ง หรือ ปริมาณการปล่อยน้ำในแม่น้ำ เป็นต้น โดยจะพบว่าข้อมูลส่วนใหญ่เป็นข้อมูลที่มีระยะสั้น ดังนั้น ถ้าตัวแปรสุ่ม X ที่มีการแจกแจงแบบล็อกปกติ โดย $X \sim LN(\mu,\sigma^2)$ ที่มีพารามิเตอร์คือ (μ,σ^2) และให้ $Y = \ln X$ จะมีการแจกแจงแบบปกติ ซึ่งมีพารามิเตอร์ คือ (μ,σ^2) เช่นเดียวกัน โดยที่ μ คือค่าเฉลี่ยของประชากร และ σ^2 คือความแปรปรวนของประชากร ซึ่งเราพบว่าค่าของ X นั้นมีค่าเป็นจำนวนบวกเสมอ โดยฟังก์ชันความหนาแน่นของความน่าจะเป็นของการแจกแจงแบบล็อกปกติเขียนได้ ดังนี้

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{\frac{-(\log x-\mu)^2}{2\sigma^2}}, x > 0, -\infty < \mu < \infty, \sigma > 0$$

ดังนั้นในงานวิจัยนี้จึงมุ่งเน้นศึกษาและปรับปรุงวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มให้มีค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่สั้นลงและให้ค่าความน่าจะเป็นคุ้มรวมเข้าใกล้ระดับความเชื่อมั่น โดยจะกำหนดให้เราทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่มซึ่งเป็นข้อมูลจากกลุ่มควบคุมที่ผู้วิจัยได้ทำการศึกษามาอย่างดี และอีกกลุ่มเป็นกลุ่มประชากรใหม่ที่ผู้วิจัยไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากร โดยทำการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นและค่าความน่าจะเป็นคุ้มรวมกับวิธีพาราเมตริกบูตสแตรปที่มีการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มโดยสถานการณ์เช่นนี้อาจเกิดขึ้นได้ ตัวอย่างเช่น แพทย์ต้องการเปรียบเทียบวิธีการรักษาคนไข้โรคเบาหวานว่าต้องใช้ระยะเวลานานเท่าไรกว่าคนไข้จะหายจากโรคเบาหวานด้วยวิธีการรักษาแบบปกติ (ทราบค่าพารามิเตอร์) กับวิธีการรักษาด้วยวิธีใหม่ (ไม่ทราบค่าพารามิเตอร์)

วัตถุประสงค์การวิจัย

เพื่อปรับปรุงวิธีการประมาณช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติของสองประชากร โดยใช้วิธีการประมาณแบบพาราเมตริกบูตสแตรป (Parametric bootstrap method) เมื่อไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่ม และหาวิธีที่เหมาะสมของแต่ละสถานการณ์

วิธีการวิจัย

งานวิจัยนี้สนใจศึกษาการหาค่าประมาณแบบช่วงสำหรับอัตราส่วนของค่าเฉลี่ยของประชากร 2 กลุ่ม ที่มีการแจกแจงแบบล็อกปกติเมื่อไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่ม และนำผลลัพธ์จากการจำลองมาเปรียบเทียบกับวิธีการหาช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากร 2 กลุ่ม กับวิธีพาราเมตริกบูตสแตรปเดิม โดยพิจารณาที่ค่าความน่าจะเป็นคุ้มรวม และ ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น โดยมีขั้นตอนการศึกษาดังนี้

1) สร้างข้อมูลที่มีการแจกแจงแบบล็อกปกติ โดยจำลองข้อมูลให้มีขนาดตัวอย่างต่างๆในแต่ละสถานการณ์ทั้งหมด 3 กรณีดังนี้

1.1 $(\mu_1,\mu_2)=(0,0)$, $(\sigma_1^2,\sigma_2^2)=(3,3)$

1.2 $(\mu_1,\mu_2)=(0.75,0)$, $(\sigma_1^2,\sigma_2^2)=(0.5,2)$

1.3 $(\mu_1,\mu_2)=(0,0.75)$, $(\sigma_1^2,\sigma_2^2)=(2,0.5)$

โดยกำหนดขนาดตัวอย่างทั้งสองกลุ่ม 7 กรณี คือ $(n_1,n_2)=(5,5),(5,25),(25,5),(25,25),(25,50),$ $(50,25),(50,50)$

2) แปลงข้อมูล $X_{i1},X_{i2},...,X_{in_i}$ ที่มีการแจกแจงแบบล็อกปกติให้เป็นข้อมูล $Y_{i1},Y_{i2},...,Y_{in_i}$ ที่มีการแจกแจงใกล้เคียงการแจกแจงปกติ (โดย $i=1,2$) ทำได้โดยการกำหนดดังนี้

$$Y_{in}=\ln(X_{in})$$

จะได้ $$Y_i=\ln(X_i)\sim N(\mu_i,\sigma_i^2)$$

3) คำนวณค่าเฉลี่ยและความแปรปรวน จากตัวแปรสุ่ม Y ที่มีการแจกแจงแบบปกติ โดยคำนวณหาค่าเฉลี่ยและความแปรปรวน ดังนี้

$$\overline{Y}_i=\frac{\sum_{j=1}^{n_i}Y_{ij}}{n_i} \quad \text{และ} \quad S_i^2=\frac{\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}\right)^2}{n_i-1}$$

โดย $\overline{Y}$ แทนค่าเฉลี่ยของตัวอย่าง

S_i^2 แทนความแปรปรวนของตัวอย่าง

จะได้ $$\overline{Y}\sim N\left(\mu_i,\frac{\sigma_i^2}{n_i}\right) \quad \text{และ} \quad \frac{(n_i-1)S_i^2}{\sigma_i^2}\sim\chi^2_{(n_i-1)} \quad \text{โดย } i=1,2$$

4) คำนวณช่วงความเชื่อมั่นของอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับ 2 ประชากร

โดยอัตราส่วนของค่าเฉลี่ยของประชากรสองกลุ่มสามารถคำนวณได้ดังนี้

$$\frac{m_1}{m_2} = \frac{\exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right)}{\exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right)}$$

จะได้ $$\frac{m_1}{m_2} = \exp\left(\mu_1 + \frac{\sigma_1^2}{2} - \mu_2 - \frac{\sigma_2^2}{2}\right) \quad (1)$$

เมื่อพิจารณา $$\psi = \log\left(\frac{m_1}{m_2}\right)$$
$$= \log(m_1) - \log(m_2)$$
$$= \mu_1 + \frac{\sigma_1^2}{2} - \mu_2 - \frac{\sigma_2^2}{2}$$

และกำหนดให้ $\eta_i = \mu_i + \frac{\sigma_i^2}{2}$ โดย $i = 1, 2$

จะได้ว่า $$\psi = \eta_1 - \eta_2 = \mu_1 + \frac{\sigma_1^2}{2} - \mu_2 - \frac{\sigma_2^2}{2} \quad (2)$$

เพื่อให้ง่ายต่อการคำนวณเราจะพิจารณาการหาช่วงความเชื่อมั่นของ $\eta_1 - \eta_2$ ในสมการ (2) แทนการหาช่วงความเชื่อมั่นอัตราส่วนของค่าเฉลี่ยของ $\frac{m_1}{m_2}$ จากสมการ (1)

4.1) วิธีพาราเมตริกบูตสแตรป

ช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ $\eta_1 - \eta_2$ คือ

$$(\eta_1 - \eta_2) \in (\eta_1 - \eta_2) \pm q_\alpha^B \sqrt{V_{12}}$$

โดย $$Var(\eta_1 - \eta_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^4}{2(n_1 - 1)} + \frac{\sigma_2^2}{n_2} + \frac{\sigma_2^4}{2(n_2 - 1)}$$

และเนื่องจาก

$$E\left(\frac{S_i^2}{n_i}+\frac{S_i^4}{2(n_i+1)}\right)=E\left(\frac{S_i^2}{n_i}\right)+E\left(\frac{S_i^4}{2(n_i+1)}\right)$$
$$=\frac{1}{n_i}E\left(S_i^2\right)+\frac{1}{2(n_i+1)}E\left(S_i^4\right) \qquad \because E\left(S_i^4\right)=\left(\frac{n_i+1}{n_i-1}\right)\sigma_i^4, E\left(S_i^2\right)=\sigma_i^2$$
$$=\frac{\sigma_i^2}{n_i}+\frac{1}{2(n_i+1)}\left(\frac{n_i+1}{n_i-1}\right)\sigma_i^4$$
$$=\frac{\sigma_i^2}{n_i}+\frac{\sigma_i^4}{2(n_i-1)}$$

จะได้ตัวประมาณที่ไม่เอนเอียงของ $Var\left(\eta_1-\eta_2\right)$ คือ

$$V_{12}=\frac{S_1^2}{n_1}+\frac{S_1^4}{2(n_1+1)}+\frac{S_2^2}{n_2}+\frac{S_2^4}{2(n_2+1)}$$

และ $\eta_i=\mu_i+\frac{\sigma_i^2}{2}, i=1,2$

เนื่องจาก $E\left(\overline{Y}_i\right)=\mu_i$ และ $E\left(S_i^2\right)=\sigma_i^2$

จะได้ตัวประมาณค่าที่ไม่เอนเอียงของ η_i คือ $\eta_i=\overline{Y}_i+\frac{S_i^2}{2}, i=1,2$

ซึ่ง q_α^B แทนควอนไทล์ที่ $1-\alpha$ ของการแจกแจง T^B โดย

$$T^B=\left|\frac{\overline{Y}_1^B-\overline{Y}_2^B+\frac{1}{2}\left(S_1^{2B}-S_2^{2B}\right)-\frac{1}{2}\left(S_1^2-S_2^2\right)}{\sqrt{V_{12}^B}}\right| \tag{3}$$

ซึ่งเราจะใช้กระบวนการทางบูตสแตรปในการประมาณช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ $\eta_1-\eta_2$ ดังนี้

1. สร้างตัวอย่างสุ่ม $X_{11}, X_{12},..., X_{1n_1}$ และ $X_{21}, X_{22},..., X_{2n_2}$ ที่มีการแจกแจงแบบล็อกปกติ โดยมีพารามิเตอร์คือ $\left(\mu_1,\sigma_1^2\right)$ และ $\left(\mu_2,\sigma_2^2\right)$ ตามลำดับ
2. ทำการแปลงค่า $Y_{in}=\ln\left(X_{in}\right)$ ซึ่งจะได้ตัวอย่างสุ่ม $Y_{i1}, Y_{i2},..., Y_{in_i}$ โดย $i=1,2$
3. คำนวณค่า $S_i^2, i=1,2$ จากข้อมูลที่จำลองขึ้นมา
4. สร้างตัวแปรสุ่ม $\overline{Y}_i^B \sim \frac{S_i}{\sqrt{n_i}}N(0,1)$ และ $S_i^{2B} \sim \frac{S_i^2}{n_i-1}\chi^2_{(n_i-1)}, i=1,2$
5. ทำซ้ำในขั้นที่ 4 หลายๆครั้ง (จำนวน $M=10,000$ รอบ) จนได้การแจกแจงที่ชัดเจนของ T^B จากสมการ 3

6. เรียงลำดับค่า T^{B} จากน้อยไปมาก จากนั้นให้ q_{α}^{B} เป็นเปอร์เซ็นต์ไทล์ที่ $100(1-\alpha)$

โดย $V_{12}^{B}=\dfrac{s_1^{2B}}{n_1}+\dfrac{s_1^{4B}}{2(n_1+1)}+\dfrac{s_2^{2B}}{n_2}+\dfrac{s_2^{4B}}{2(n_2+1)}$

4.2) วิธีพาราเมตริกบูตสแตรปเมื่อไม่ทราบค่าค่าเฉลี่ยและความแปรปรวนของประชากร 1 กลุ่ม ช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ $\eta_1-\eta_2$ คือ

$$(\eta_1-\eta_2)\in\left(\eta_1-\eta_1\right)\pm q_{\alpha}^{B}\sqrt{V_{12}}$$

โดยกำหนดให้ทราบค่าพารามิเตอร์ของประชากรกลุ่มที่ 2 คือ $\eta_2=\eta_2,\ \bar{Y}_2=\mu_2$ และ $S_2^2=\sigma_2^2$

โดย $Var\left(\eta_1-\eta_2\right)$ หาได้ดังนี้

$$\begin{aligned}Var\left(\eta_1-\eta_2\right)&=Var\left(\eta_1-\eta_2\right)\\&=Var\left(\eta_1\right)+Var\left(\eta_2\right)\\&=Var\left(\eta_1\right)\\&=\frac{\sigma_1^2}{n_1}+\frac{\sigma_1^4}{2(n_1-1)}\end{aligned}$$

และจะได้ตัวประมาณที่ไม่เอนเอียงของ $Var\left(\eta_1-\eta_2\right)$ คือ

$$V_{12}=\frac{S_1^2}{n_1}+\frac{S_1^4}{2(n_1+1)}$$

ซึ่งการคำนวณค่า q_{α}^{B} จะคำนวณเช่นเดียวกับวิธีในข้อ 4.1)

5) ตรวจสอบวิธีการประมาณแบบช่วงสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับสองประชากรเมื่อไม่ทราบค่าค่าเฉลี่ยและความแปรปรวนหนึ่งกลุ่มกับวิธีแบบพาราเมตริกบูตสแตรปโดยเปรียบเทียบค่าความน่าจะเป็นคุ้มรวมที่ได้จากการทดลองว่ามีค่าไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่ โดยถ้าสถานการณ์ใดให้ค่าความน่าจะเป็นคุ้มรวมของช่วงความเชื่อมั่นที่สัมประสิทธิ์ความเชื่อมั่น 0.95 มีค่าอยู่ในช่วง 0.9464 ถึง 0.9536 จะสรุปได้ว่าวิธีการประมาณนั้นให้ค่าความน่าจะเป็นคุ้มรวมไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดไว้ และวิธีการนั้นจะถือว่ามีประสิทธิภาพโดยเมื่อทำการคำนวณค่าความน่าจะเป็นคุ้มรวมแล้วเราจะทำการคำนวณค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีการประมาณแบบพาราเมตริกบูตสแตรปกับวิธีพาราเมตริกบูตสแตรปโดยกำหนดให้ไม่ทราบค่าค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่ม ว่าวิธีการประมาณแบบใดให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดของในแต่ละสถานการณ์ ซึ่งวิธีการคำนวณจะแสดงในขั้นตอนถัดไป

6) ความน่าจะเป็นคุ้มรวม (Coverage Probability : CP) คือ ค่าความน่าจะเป็นที่ช่วงความเชื่อมั่นคลุมค่าพารามิเตอร์ โดยคำนวณจากข้อมูลที่ทำซ้ำหลายๆครั้ง ซึ่งคำนวณได้จาก

$$CP = \frac{\sum_{i=1}^{N} I_{[L,U]}(\eta_1 - \eta_2)}{N}$$

เมื่อ $I_{[L,U]}(\eta_1 - \eta_2)$ จะมีค่าเท่ากับ 1 เมื่อ $\eta_1 - \eta_2$ อยู่ในช่วง $[L_i, U_i]$ และ เท่ากับ 0 เมื่ออยู่นอกช่วง $[L_i, U_i]$

L_i และ U_i เป็นขอบเขตล่างและบนของช่วงความเชื่อมั่นรอบที่ i

N แทน จำนวนครั้งที่ทำการจำลองของข้อมูล

ความยาวเฉลี่ยของช่วงความเชื่อมั่น (Average Length : AL) ซึ่งสามารถคำนวณได้ดังนี้

$$AL = \frac{\sum_{i=1}^{M} (U_i - L_i)}{N}$$

โดยที่ U_i คือ ขอบเขตบนของช่วงความเชื่อมั่น

L_i คือ ขอบเขตล่วงของช่วงความเชื่อมั่น

N คือ จำนวนครั้งที่ทำการจำลองข้อมูล

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อปรับปรุงช่วงความเชื่อมั่นของอัตราส่วนของค่าเฉลี่ยสำหรับการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มเมื่อไม่ทราบทราบค่าเฉลี่ยและความแปรปรวนเพียงหนึ่งกลุ่ม โดยมีเกณฑ์ที่ใช้ในการเปรียบเทียบระหว่างวิธีการหาโดยใช้บูตสแตรปและวิธีของผู้วิจัย

โดยในการนำเสนอผลการวิจัย ผู้วิจัยได้กำหนดสัญลักษณ์แทนความหมายต่างๆดังนี้

n_1, n_2 แทน ขนาดตัวอย่างกลุ่มที่ 1 และ 2 ตามลำดับ

μ_1, μ_2 แทน ค่าเฉลี่ยประชากรกลุ่มที่ 1 และ 2 ตามลำดับ

σ_1^2, σ_2^2 แทน ความแปรปรวนประชากรกลุ่มที่ 1 และ 2 ตามลำดับ

CP แทน ความน่าจะเป็นคุ้มรวม

AL แทน ความยาวเฉลี่ยของช่วงความเชื่อมั่น

$PB1$ แทน วิธีพาราเมตริกบูตสแตรป

$PB2$ แทน วิธีพาราเมตริกบูตสแตรปโดยกำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่ม

* แทน วิธีที่ให้ค่าความน่าจะเป็นคุ้มรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

** แทน กรณีที่วิธีการประมาณให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสั้นที่สุด

ผลลัพธ์จากการจำลองโดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo simulation method) ทำซ้ำ 10,000 ครั้ง

กรณีสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่ากับ 0.95 นำเสนอดังตารางที่ 1 โดยถ้าวิธีใดให้ค่าความน่าจะเป็นคุ้มรวมจากผลการจำลองอยู่ในช่วง 0.9464 ถึง 0.9536 ถือว่าวิธีนั้นมีประสิทธิภาพที่ดี

ตารางที่ 1 แสดงค่าความน่าจะเป็นคุ้มรวมและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มด้วยวิธีพาราเมติกบูตสแตรปและวิธีพาราเมติกบูตสแตรปเมื่อทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่มที่ระดับความเชื่อมั่น 95%

พารามิเตอร์	(n_1,n_2)	PB1		PB2	
		CP	AL	CP	AL
$(\mu_1,\mu_2)=(0,0)$ $(\sigma_1^2,\sigma_2^2)=(3,3)$	(5,5)	0.9827*	9.3058	0.8909	5.3928**
	(5,25)	0.9111	7.6104	0.8512	6.7904**
	(25,25)	0.9594*	3.0612	0.9317	2.1262**
	(50,50)	0.9487*	2.1413	0.9394	1.5011**
	(25,50)	0.9513*	2.6795	0.9284	2.1638**
	(25,5)	0.9121	7.5565	0.9811*	2.7967**
	(50,25)	0.9481	2.6728	0.9457	1.5291**
$(\mu_1,\mu_2)=(0.75,0)$ $(\sigma_1^2,\sigma_2^2)=(0.5,2)$	(5,25)	0.9548*	2.2431	0.8865	1.4046**
	(5,5)	0.9313	5.7181	0.9590*	1.7949**
	(25,25)	0.9477*	1.7860	0.9483*	0.6524**
	(50,50)	0.9462	1.2215	0.9518*	0.4479**
	(25,5)	0.9142	6.0565	0.9956*	1.0446**
	(25,50)	0.9493*	1.2923	0.9425	0.6249**
	(50,25)	0.943	1.7477	0.9626*	0.4724**
$(\mu_1,\mu_2)=(0,0.75)$ $(\sigma_1^2,\sigma_2^2)=(2,0.5)$	(5,5)	0.9319	5.6437	0.8838	5.0556**
	(5,25)	0.9110	6.0859	0.9016	5.9661**
	(25,25)	0.9499*	1.7903	0.9426	1.6583**
	(50,50)	0.9492*	1.2210	0.9476*	1.1327**
	(25,50)	0.9484*	1.7545	0.9438	1.6860**
	(25,5)	0.9586*	2.2383	0.9422	1.6303**
	(50,25)	0.9505*	1.2927	0.9480*	1.1245**

จากตารางที่ 1 ได้แสดงความน่าจะเป็นคุ้มรวมและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสำหรับอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มด้วยวิธีพาราเมติกบูตสแตรปและวิธีพาราเมติกบูตสแตรปเมื่อทราบค่าเฉลี่ยและความแปรปรวนของประชากรหนึ่งกลุ่มที่ระดับความเชื่อมั่น 95% โดยมีขนาดตัวอย่างต่างๆ และค่าพารามิเตอร์ตามที่กำหนดไว้ข้างต้น ผู้วิจัยสามารถสรุปผลการเปรียบเทียบได้ดังนี้

1. วิธีการประมาณแบบพาราเมตริกบูตสแตรปให้ค่าความน่าจะเป็นคุ้มรวมไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดในสถานการณ์ต่อไปนี้

เมื่อขนาดตัวอย่าง (n_1,n_2) มีค่า (25,50), (50,25) และ (50,50) ในกรณี $(\mu_1,\mu_2)=(0,0)$ และ $(\sigma_1^2,\sigma_2^2)=(3,3)$

เมื่อขนาดตัวอย่าง (n_1,n_2) มีค่า (25,25) และ (25,50) ในกรณี $(\mu_1,\mu_2)=(0.75,0)$ และ $(\sigma_1^2,\sigma_2^2)=(0.5,2)$

เมื่อขนาดตัวอย่าง (n_1,n_2) มีค่า (25,25), (50,25), (25,50) และ (50,50) ในกรณี $(\mu_1,\mu_2)=(0,0.75)$ และ $(\sigma_1^2,\sigma_2^2)=(2,0.5)$

2. วิธีการประมาณแบบพาราเมตริกบูตสแตรปโดยกำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนเพียงหนึ่งกลุ่มให้ค่าความน่าจะเป็นคุ้มรวมไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดในสถานการณ์ต่อไปนี้

ไม่มีขนาดตัวอย่าง (n_1,n_2) ในกรณี $(\mu_1,\mu_2)=(0,0)$ และ $(\sigma_1^2,\sigma_2^2)=(3,3)$ ที่ความน่าจะเป็นคุ้มรวมไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนด

เมื่อขนาดตัวอย่าง (n_1,n_2) มีค่า (25,25) และ (50,50) ในกรณี $(\mu_1,\mu_2)=(0.75,0)$ และ $(\sigma_1^2,\sigma_2^2)=(0.5,2)$

เมื่อขนาดตัวอย่าง (n_1,n_2) มีค่า (50,25) และ (50,50) ในกรณี $(\mu_1,\mu_2)=(0,0.75)$ และ $(\sigma_1^2,\sigma_2^2)=(2,0.5)$

3. ความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีการประมาณแบบพาราเมตริกบูตสแตรปโดยกำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนเพียงหนึ่งกลุ่มให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่สั้นกว่าวิธีการประมาณแบบพาราเมตริกบูตสแตรปในทุกๆขนาดตัวอย่าง (n_1,n_2) และทุกค่าของพารามิเตอร์ (μ_1,μ_2) และ (σ_1^2,σ_2^2)

อภิปรายและสรุปผลการวิจัย

ช่วงความเชื่อมั่นของอัตราส่วนของค่าเฉลี่ยของการแจกแจงแบบล็อกปกติสำหรับประชากรสองกลุ่มโดยกำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนหนึ่งกลุ่มโดยใช้วิธีพาราเมตริกบูตแสตรปนั้นจะมีประสิทธิภาพเมื่อขนาดตัวอย่างของทั้งสองกลุ่มประชากรมีค่าที่มากพอหรือไม่เล็กจนเกินไป ซึ่งเราจะพบว่าวิธีที่กำหนดให้ไม่ทราบค่าเฉลี่ยและความแปรปรวนหนึ่งกลุ่มของผู้วิจัยนั้นจะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่สั้นกว่าวิธีพาราเมตริกบูตสแตรปในทุกๆกรณี เมื่อกำหนดระดับความเชื่อมั่นที่ 95%

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณครอบครัวและเพื่อนๆ ที่คอยห่วงใย เป็นกำลัง ให้กับผู้วิจัย และขอบคุณทุนเรียนดีคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ สำหรับการส่งเสริมและสนับสนุนด้านการเรียนของผู้วิจัยตลอดงานวิจัยนี้

เอกสารอ้างอิง

ดนุพล ทองคำสองเมือง เหล่าตระกูล และ วุฒิชัย ศรีโสดาพล.ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกปกติและ การประยุกต์ ภาควิชาสถิติ. คณะวิทยาศาสตร์. มหาวิทยาลัยขอนแก่น. ขอนแก่น; 2556.

วีรวรรณ ศักดาจิวะเจริญ. ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาม. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ. คณะพาณิชยศาสตร์และการบัญชี. จุฬาลงกรณ์มหาวิทยาลัย; 2544.

Angus JE. Bootstrap one-sided intervals for the log-normal mean. The Statistician 1994; 43 (3):395–401.

Chen Y-H, Zhou, X-H. Interval estimates for the ratio and difference of two lognormal means. Statistics in Medicine 2006; 25:4099–4113.

Krishnamoorthy K, Mathew T. Inferences on the means of log-normal distributions using generalized p-values and generalized confidence intervals. Journal of Statistical Planning and Inference. 2003; 115: 103-21.

Sadooghi-Alvandi SM, Malekzadeh A. Simultaneous confidence intervals for ratios of means of several lognormal distributions: A parametric bootstrap approach. Computational Statistics and Data Analysis 2014; 69:133-140

Suparat N. Confidence intervals for the difference of two normal population means with one variance unknown. King Mongkut's University of Technology North Bangkok ,Thailand Statistician 2009; 7(2):161-177

Zhou XH, Gao S. Confidence intervals for the lognormal Mean. Statistics in Medecien 1997; 16:783-790.

Zhou X-H, Tu W. Interval estimation for the ratio in means of log-normally distributed medical costs withzero values. Computational Statistics and Data Analysis 2000; 35:201–210.