

ประสิทธิภาพของช่วงความเชื่อมั่นบูตสเตรป สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์

An Efficiency of Bootstrap Confidence Intervals for Coefficient of Quartile Variation

พส สมบูรณ์ และธิดาเดียว มยุรีสุวรรณ*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ถนนมิตรภาพ อำเภอเมือง จังหวัดขอนแก่น 40002

Pot Somboon and Tidadeaw Mayureesawan*

Department of Statistics, Faculty of Science, Khon Kaen University,

Mittraphap Road, Muang, Khon Kaen 40002

บทคัดย่อ

งานวิจัยนี้ได้ศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ 3 วิธี ได้แก่ วิธี Bonett Bootstrap วิธี Nonparametric Bootstrap และวิธี Parametric Bootstrap เมื่อประมาณค่าควอไทล์ของประชากรจาก 5 วิธี ได้แก่วิธี Tukey วิธี Moore and McCabe (M&M) วิธี Mendenhall and Sincich (M&S) วิธี Freund and Perles (F&P) และวิธี Minitab กรณีข้อมูลมีการแจกแจงสมมาตร ($N(4, 1)$) การแจกแจงเบ้เชิงบวกน้อย ($LN(0, 0.25)$, $G(9, 2)$) การแจกแจงเบ้เชิงบวกปานกลาง ($LN(0, 0.5)$, $G(6, 1)$) และการแจกแจงเบ้เชิงบวกมาก ($LN(0, 0.75)$, $LN(0, 1)$, $LN(0, 1.5)$, $G(2, 0.5)$, $G(0.5, 0.5)$) ในหลายระดับขนาดตัวอย่าง โดยพิจารณาประสิทธิภาพของช่วงความเชื่อมั่นจากค่าความน่าจะเป็นคุ้มครองและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ที่ระดับความเชื่อมั่น 0.95 ด้วยวิธีการจำลองแบบมอนติคาร์โล ผลการวิจัยสำหรับช่วงความเชื่อมั่นวิธี Bonett Bootstrap พบว่ามีประสิทธิภาพหากประมาณค่าควอไทล์ด้วยวิธี Minitab เมื่อข้อมูลมีการแจกแจง $N(4, 1)$ และ $LN(0, 0.25)$ ที่ตัวอย่างขนาดเล็ก และเมื่อข้อมูลมีการแจกแจง $G(0.5, 0.5)$ ในหลายระดับขนาดตัวอย่าง สำหรับช่วงความเชื่อมั่นวิธี Nonparametric Bootstrap พบว่ามีประสิทธิภาพในหลายการแจกแจงหากประมาณค่าควอไทล์ด้วยวิธี F&P ได้แก่ การแจกแจง $N(4, 1)$, $G(9, 2)$ และ $G(6, 1)$ ที่ตัวอย่างขนาดเล็ก และเมื่อข้อมูลมีการแจกแจง $LN(0, 0.75)$, $LN(0, 1)$ และ $G(2, 0.5)$ ในหลายระดับขนาดตัวอย่าง สำหรับช่วงความเชื่อมั่นวิธี Parametric Bootstrap พบว่ามีประสิทธิภาพหากประมาณค่าควอไทล์ด้วยวิธี F&P เมื่อข้อมูลมีการแจกแจง $N(4, 1)$, $LN(0, 0.5)$, $G(6, 1)$ และ $G(9, 2)$ ที่ตัวอย่างขนาดเล็ก นอกจากนี้ยังพบว่าช่วงความเชื่อมั่นวิธี Parametric Bootstrap มีประสิทธิภาพดีในหลายระดับขนาดตัวอย่างหากประมาณค่าควอไทล์ด้วยวิธี Minitab เมื่อข้อมูลมีการแจกแจง $LN(0, 0.75)$, $LN(0, 1)$, $LN(0, 1.5)$ และ $G(2, 0.5)$

คำสำคัญ : ช่วงความเชื่อมั่นบูตสตรป; สัมประสิทธิ์การแปรผันควอไทล์; โบนีตบูตสตรป

Abstract

The purpose of this research is to study the efficiency of bootstrap confidence intervals for coefficient of quartile variation in three methods, namely, Bonett Bootstrap (BB), Nonparametric Bootstrap (NB), and Parametric Bootstrap (PB) method when estimating the population of quartile from five quartile methods, i.e. Tukey, Moore and McCabe (M&M), Mendenhall and Sincich (M&S), Freund and Perles (F&P) and Minitab. The data cases are symmetrical ($N(4, 1)$), fairly symmetrical ($LN(0, 0.25)$, $G(9, 2)$), moderately skewness ($LN(0, 0.5)$, $G(6, 1)$), highly skewness ($LN(0, 0.75)$, $LN(0, 1)$, $LN(0, 1.5)$, $G(2, 0.5)$, $G(0.5, 0.5)$) with the various sizes of samples. The criteria of confidence interval are considered from coverage probability and the average width of confidence interval at confidence level of 0.95 by using Monte Carlo simulation. The result for confidence interval of BB method found that the population of quartile estimated by Minitab method performs better when distributions are $N(4, 1)$ and $LN(0, 0.25)$ at small level of sample and $G(0.5, 0.5)$ in various sizes of samples. For confidence interval of NB method, quartile estimated by F&P method has an efficiency in several distributions, namely, $N(4, 1)$, $G(9, 2)$ and $G(6, 1)$ at small levels of samples and $LN(0, 0.75)$, $LN(0, 1)$ and $G(2, 0.5)$ in many levels of samples. For confidence interval of PB method performs better than quartile estimated by F&P method when distributions are $N(4, 1)$, $LN(0, 0.5)$, $G(6, 1)$ and $G(9, 2)$ at small levels of samples. It is also found that the PB confidence interval based on Minitab quartile method has an efficiency in various sizes of samples when distributions are $LN(0, 1.5)$, $LN(0, 1)$, $LN(0, 0.75)$ and $G(2, 0.5)$ distribution.

Keywords: bootstrap confidence interval; coefficient of quartile variation; bonett bootstrap

1. บทนำ

สัมประสิทธิ์การแปรผันควอไทล์ (coefficient of quartile variation, CQV) เป็นค่าวัดการกระจายสัมพัทธ์ที่เหมาะสมสำหรับใช้วัดการกระจายของข้อมูลที่ไม่ได้มีการแจกแจงปกติ เนื่องจากการคำนวณที่หลีกเลี่ยงผลกระทบจากข้อมูลปลายสุดด้านใดด้านหนึ่งหรือสองด้านมีค่าผิดปกติ โดยค่า CQV คำนวณจาก $CQV = \frac{Q_3 - Q_1}{(Q_3 + Q_1)}$ เมื่อ Q_1 และ Q_3 คือค่าควอไทล์ที่ 1 และควอไทล์ที่ 3 ตามลำดับ และค่า $Q_3 - Q_1$ คือ

พิสัยระหว่างควอไทล์ (Interquartile Range: IQR) [1]

Langford [2] ศึกษาว่า IQR เมื่อประมาณค่าควอไทล์ที่ 1 และควอไทล์ที่ 3 ด้วยวิธีการที่ต่างกัน 5 วิธี ได้แก่ วิธี Tukey วิธี Moore and McCabe วิธี Mendenhall and Sincich วิธี Freund and Perles และวิธีที่อยู่ในโปรแกรม Minitab ผลการวิจัยพบว่าวิธีประมาณค่าควอไทล์ที่ต่างกันมีผลต่อค่า IQR แต่ความแตกต่างของค่า IQR ที่ได้ขึ้นอยู่กับขนาดตัวอย่างและลักษณะของข้อมูลที่เป็นจำนวนคู่หรือจำนวนคี่

Cangur และคณะ ศึกษาค่าควอไทล์ที่คำนวณในโปรแกรม Microsoft Excel และโปรแกรม Minitab โดยวิธีคำนวณค่าควอไทล์ในโปรแกรม Microsoft Excel คือ วิธีของ Freund and Perles โดยศึกษาทั้งกรณีข้อมูลมีลักษณะสมมาตร ได้แก่ การแจกแจงปกติมาตรฐาน และข้อมูลมีลักษณะไม่สมมาตร ได้แก่ การแจกแจงโคเคาลังสอง และการแจกแจงค่าต่ำสุดขีด (smallest extreme value) จากการพิจารณาค่า standard error ของ IQR พบว่าวิธีในโปรแกรม Microsoft Excel เหมาะสมกับข้อมูลที่มีลักษณะไม่สมมาตร ส่วนวิธีในโปรแกรม Minitab เหมาะสมกับข้อมูลที่มีลักษณะสมมาตร

การศึกษาผลของวิธีประมาณค่าควอไทล์ที่พบว่าให้ค่าควอไทล์ต่างกัน แต่ขึ้นอยู่กับวิธีการแจกแจงของข้อมูลและขนาดตัวอย่างที่ใช้คำนวณ ซึ่งค่าประมาณค่าควอไทล์ที่ 1 และควอไทล์ที่ 3 ส่งผลต่อค่าประมาณแบบจุดของสัมประสิทธิ์การแปรผันควอไทล์ และมีนักวิจัยหลายคนศึกษาเกี่ยวกับช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ โดย Bonett [4] นำเสนอช่วงความเชื่อมั่น 95 % สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ โดยใช้สมบัติการลดรูปของลอการิทึมและความแปรปรวนผลต่าง เพื่อหาความแปรปรวนของค่าสัมประสิทธิ์การแปรผันควอไทล์ และศึกษาประสิทธิภาพช่วงความเชื่อมั่นดังกล่าวกรณีข้อมูลมีการแจกแจงปกติ และไม่ได้มีการแจกแจงปกติ ในหลายระดับขนาดตัวอย่าง โดยใช้ความน่าจะเป็นคัมรวม (coverage probability, CP) เป็นเกณฑ์การพิจารณาประสิทธิภาพของช่วงความเชื่อมั่น ผลการวิจัยพบว่าช่วงความเชื่อมั่นของ Bonett มีประสิทธิภาพดีกรณีข้อมูลไม่ได้มีการแจกแจงปกติ และตัวอย่างมีขนาดเล็ก

Tongkaw และ Pongsakchat [5] นำเสนอช่วงความเชื่อมั่น 95 % สำหรับค่าสัมประสิทธิ์การแปร

ผันควอไทล์ 2 วิธี คือ วิธี Bootstrap และวิธี Bonett Bootstrap โดยวิธี Bootstrap ใช้เทคนิคการสุ่มซ้ำ (resampling technique) ส่วนวิธี Bonett Bootstrap เป็นวิธีที่ปรับจากวิธี Bonett และใช้เทคนิคการสุ่มซ้ำ Tongkaw และ Pongsakchat เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่น 2 วิธีนี้กับช่วงความเชื่อมั่นวิธี Bonett กรณีที่ข้อมูลมีการแจกแจงปกติ การแจกแจงลอกลอนอร์มอล และการแจกแจงแกมมา โดยใช้ความน่าจะเป็นคัมรวมและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (average width of confidence interval, AW) เป็นเกณฑ์การพิจารณาประสิทธิภาพของช่วงความเชื่อมั่น ผลการวิจัยพบว่าช่วงความเชื่อมั่นวิธี Bootstrap และวิธี Bonett Bootstrap มีประสิทธิภาพดีกว่าวิธี Bonett ในกรณีตัวอย่างมีขนาดเล็กสำหรับการแจกแจงที่ศึกษา

Altunkaynak และ Gamgam [6] นำเสนอช่วงความเชื่อมั่น 95 % สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ 2 วิธี คือ วิธี Nonparametric Bootstrap และวิธี Parametric Bootstrap โดยวิธี Nonparametric Bootstrap ใช้เทคนิคที่เรียกว่าเปอร์เซ็นต์ไทล์บูตสเตรป (percentile bootstrap) ส่วนวิธี Parametric Bootstrap ใช้เทคนิคตัวประมาณค่าแบบจุดโดยวิธี Bootstrap และเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นทั้ง 2 วิธีนี้กับช่วงความเชื่อมั่นวิธี Bonett กรณีข้อมูลมีการแจกแจงปกติ และไม่ได้มีการแจกแจงปกติ ในหลายระดับขนาดตัวอย่าง โดยใช้ความน่าจะเป็นคัมรวมและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น เป็นเกณฑ์การพิจารณาประสิทธิภาพของช่วงความเชื่อมั่น ผลการวิจัยพบว่าช่วงความเชื่อมั่นทั้ง 2 วิธีนี้มีประสิทธิภาพดีกว่าวิธี Bonett กรณีข้อมูลไม่ได้มีการแจกแจงปกติ และตัวอย่างมีขนาดเล็ก

ช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ที่นำเสนอโดย Bonett [4] Tongkaw

และ Pongsakchat [5] และ Altunkaynak และ Gamgam [6] ไม่พบว่าได้กล่าวถึงวิธีประมาณค่าควอไทล์ของประชากรด้วยวิธีใด การศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผัน ควอไทล์เมื่อประมาณค่าควอไทล์ด้วยวิธีต่าง ๆ จึงเป็นสิ่งที่น่าสนใจศึกษาว่าให้ประสิทธิภาพต่างกันหรือไม่ ดังนั้นงานวิจัยนี้จึงศึกษาช่วงความเชื่อมั่น Bootstrap สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ 3 วิธี ได้แก่ วิธี Bonett Bootstrap (BB) วิธี Nonparametric Bootstrap (NB) และวิธี Parametric Bootstrap (PB) เมื่อประมาณค่าควอไทล์ของประชากรจาก 5 วิธี ได้แก่ วิธี Tukey วิธี Moore and McCabe (M&M) วิธี Mendenhall and Sincich (M&S) วิธี Freund and Perles (F&P) ในโปรแกรม Microsoft Excel และวิธีในโปรแกรม Minitab

2. ทฤษฎีที่เกี่ยวข้อง

2.1 วิธีการหาค่าตำแหน่งควอไทล์

ข้อมูลตัวอย่างขนาด n คือ $X_1, X_2, \dots, X_n$ เรียงข้อมูลจากน้อยไปหามาก แทนด้วย $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ วิธีการหาค่าตำแหน่งควอไทล์สำหรับข้อมูลที่ไม่ได้จัดกลุ่ม ได้แก่ ควอไทล์ที่ 1 (Q_1) ควอไทล์ที่ 2 (Q_2) และควอไทล์ที่ 3 (Q_3) ด้วยวิธีต่างๆ มีรายละเอียดดังนี้ [3]

2.1.1 วิธี Tukey (Tukey method)

วิธี Tukey มีขั้นตอนการหาค่าตำแหน่ง Q_1 และ Q_3 โดยหาค่าตำแหน่ง Q_2 ก่อนจากสูตร $(n+1)/2$ ส่วนตำแหน่ง Q_1 คือ ตำแหน่งมัธยฐานของข้อมูลส่วนล่าง เมื่อข้อมูลส่วนล่างคือข้อมูลลำดับที่ 1 ($X_{(1)}$) ถึงข้อมูลตำแหน่ง Q_2 ส่วนการหาค่าตำแหน่ง Q_3 หาจากตำแหน่งมัธยฐานของข้อมูลส่วนบน เมื่อข้อมูลส่วนบน คือ ข้อมูลตั้งแต่ตำแหน่ง Q_2 ถึงข้อมูล

ลำดับที่ n ($X_{(n)}$)

2.1.2 วิธี Moore and McCabe (M&M method)

วิธี M&M จะแยกพิจารณาการหาค่าตำแหน่ง Q_1 และ Q_3 เป็น 2 กรณี คือ กรณีที่จำนวนข้อมูลเป็นเลขคี่และกรณีจำนวนข้อมูลเป็นเลขคู่ เมื่อจำนวนข้อมูลเป็นเลขคี่สูตรการหาค่าตำแหน่ง Q_1 คือ $(n+1)/4$ และตำแหน่ง Q_3 คือ $3(n+1)/4$ ส่วนกรณีจำนวนข้อมูลเป็นเลขคู่สูตรการหาค่าตำแหน่ง Q_1 คือ $(n+2)/4$ และตำแหน่ง Q_3 คือ $(3n+2)/4$

2.1.3 วิธี Mendenhall and Sincich (M&S method)

วิธี M&S มีสูตรการหาค่าตำแหน่ง Q_1 และ Q_3 สัมพันธ์กับวิธี M&M แต่วิธีนี้จะไม่พิจารณาการคำนวณว่าจำนวนข้อมูลเป็นเลขคี่หรือเลขคู่ โดยตำแหน่ง Q_1 จะเป็นสูตรเดียวกันกับการหาค่าตำแหน่ง Q_1 วิธี M&M เมื่อจำนวนข้อมูลเป็นเลขคี่ ส่วนตำแหน่ง Q_3 จะเป็นสูตรเดียวกันกับการหาค่าตำแหน่ง Q_3 วิธี M&M เมื่อจำนวนข้อมูลเป็นเลขคู่ ดังนั้น สูตรการหาค่าตำแหน่ง Q_1 คือ $(n+1)/4$ และตำแหน่ง Q_3 คือ $(3n+2)/4$

2.1.4 วิธี Freund and Perles (F&P method)

วิธี F&P เป็นวิธีการหาค่าควอไทล์ที่ใช้ในโปรแกรม Microsoft Excel มีสูตรการหาค่าตำแหน่ง Q_1 และ Q_3 ต่างจาก 3 วิธีข้างต้น และวิธีนี้จะไม่พิจารณาว่าจำนวนข้อมูลเป็นเลขคี่หรือเลขคู่ โดยมีสูตรการหาค่าตำแหน่ง Q_1 คือ $(n+3)/4$ และตำแหน่ง Q_3 คือ $(3n+1)/4$

2.1.5 วิธี Minitab (Minitab method)

วิธี Minitab เป็นวิธีการหาค่าควอไทล์

ที่ใช้ในโปรแกรม Minitab วิธีนี้จะไม่พิจารณาว่าจำนวนข้อมูลเป็นเลขคี่หรือเลขคู่ โดยมีสูตรการหาตำแหน่ง Q_1 และ Q_3 เหมือนกับวิธี M&M เมื่อจำนวนข้อมูลเป็นเลขคี่ ดังนั้นสูตรการหาตำแหน่ง Q_1 คือ $(n+1)/4$ และตำแหน่ง Q_3 คือ $3(n+1)/4$

2.2 วิธี Bootstrap

วิธี Bootstrap เป็นวิธีการสุ่มตัวอย่างซ้ำแบบใส่คืน (resampling with replacement) วิธีการนี้ถูกเสนอโดย Efron และ Tibshirani [7] เมื่อกำหนดให้ θ เป็นพารามิเตอร์ของประชากรที่มีการแจกแจงแบบใด ๆ $\hat{\theta}_B$ เป็นค่าประมาณของพารามิเตอร์ด้วยวิธี Bootstrap และ $x_1^*, x_2^*, \dots, x_n^*$ เป็นตัวอย่าง Bootstrap หลักการสุ่มวิธี Bootstrap มีขั้นตอนและรายละเอียดดังนี้

2.2.1 สุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบใด ๆ แทนด้วย $x_1, x_2, \dots, x_n$

2.2.2 สุ่มตัวอย่าง 1 ค่าจากชุดของตัวอย่าง $x_1, x_2, \dots, x_n$ แล้วสุ่มแบบใส่คืนจำนวน n ครั้ง จะได้ตัวอย่างที่เรียกว่าตัวอย่าง Bootstrap คือ $x_1^*, x_2^*, \dots, x_n^*$

2.2.3 ดำเนินการในขั้นตอนที่ 2 จำนวน B รอบ จะได้ $x_1^*, x_2^*, \dots, x_n^*$ จำนวน B ชุด แต่ละชุดคำนวณค่า $\hat{\theta}_i^*$ ดังนั้นจะได้ $\hat{\theta}_i^*$ จำนวนค่า B ค่า คือ $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$

2.2.4 นำค่า $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ มาหาค่าเฉลี่ย ($\hat{\theta}_B$) และความแปรปรวน (s_B^2) จาก $\hat{\theta}_B = \sum_{i=1}^B \hat{\theta}_i^* / B$ และ $s_B^2 = \sum_{i=1}^B (\hat{\theta}_i^* - \hat{\theta}_B)^2 / B$ ตามลำดับ โดยค่า $\hat{\theta}_B$ คือค่าประมาณแบบจุดของพารามิเตอร์ด้วยวิธี Bootstrap และ s_B^2 คือค่าความแปรปรวนของค่าประมาณแบบจุดของพารามิเตอร์ด้วยวิธี Bootstrap

2.3 ช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์

การแปรผันควอไทล์วิธี Bonett

กำหนดให้ $D = \hat{Q}_3 - \hat{Q}_1$ และ $S = \hat{Q}_3 + \hat{Q}_1$ เมื่อ $\hat{Q}_1$ และ $\hat{Q}_3$ เป็นตัวประมาณค่าของข้อมูลตำแหน่งควอไทล์ที่ 1 และควอไทล์ที่ 3 ที่คำนวณจากตัวอย่าง

ช่วงความเชื่อมั่น 95 % สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Bonett คำนวณได้ดังนี้

$$\exp\left\{\ln(\text{COV})c \pm z_{1-\alpha^*/2} v^{1/2}\right\} \quad (1)$$

เมื่อ $c = n/(n-1)$ คือ ค่าปรับแก้ศูนย์กลาง; V คือ ค่าความแปรปรวนของสัมประสิทธิ์การแปรผันควอไทล์ คำนวณได้ดังนี้

$$V = \left(\frac{1}{16n} \right) \left[\frac{\left(\frac{3}{f_1^2} + \frac{3}{f_3^2} - \frac{2}{f_1 f_3} \right)}{D^2} + \frac{\left(\frac{3}{f_1^2} + \frac{3}{f_3^2} + \frac{2}{f_1 f_3} \right)}{S^2} - 2 \frac{\left(\frac{3}{f_1^2} - \frac{3}{f_3^2} \right)}{DS} \right] \quad (2)$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวประมาณค่าควอไทล์ที่ 1 ($\hat{f}_1$) และควอไทล์ที่ 3 ($\hat{f}_3$) คำนวณได้ดังนี้

$$\hat{f}_1 = \sqrt{3(z_{1-\alpha^*/2})^2 / \{4n(Y_{(b)} - Y_{(a)})^2\}} \quad (3)$$

$$\hat{f}_3 = \sqrt{3(z_{1-\alpha^*/2})^2 / \{4n(Y_{(d)} - Y_{(c)})^2\}} \quad (4)$$

เมื่อ $a = n/4 - 1.96\sqrt{3n/16}$, $b = n/4 + 1.96\sqrt{3n/16}$, $c = n+1-b$, $d = n+1-a$; $Y_{(j)}$ เป็นค่าสถิติลำดับที่ j ของตัวอย่าง โดยที่ $j = a, b, c, d$; ค่า $z_{1-\alpha^*/2}$ คือ ค่าควอไทล์ที่ $(1 - \alpha^*/2)$ ของการแจกแจงปกติมาตรฐาน โดยที่ α^* ขึ้นอยู่กับขนาดตัวอย่างดังนี้

$$\alpha^* = \begin{cases} 1 - \sum_{i=a}^{b-1} \binom{n}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{n-i}, & n < 30 \\ 0.05, & n \geq 30 \end{cases} \quad (5)$$

2.4 ช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Bonett Bootstrap (BB method)

กำหนดให้ $\hat{Q}_{1i}^*$ และ $\hat{Q}_{3i}^*$ เป็นตัวประมาณค่าควอไทล์ที่ 1 และควอไทล์ที่ 3 รอบที่ i ด้วยวิธี Bootstrap ตามลำดับ ค่าสัมประสิทธิ์การแปรผันควอไทล์รอบที่ i วิธี Bootstrap (CQV_i^*) คำนวณได้ดังนี้

$$CQV_i^* = \frac{\hat{Q}_{3i}^* - \hat{Q}_{1i}^*}{\hat{Q}_{3i}^* + \hat{Q}_{1i}^*} \quad (6)$$

จากสมการ (6) สามารถคำนวณค่าประมาณแบบจุดของสัมประสิทธิ์การแปรผันควอไทล์วิธี Bootstrap (CQV_B) ดังนี้

$$CQV_B = \frac{\sum_{i=1}^B CQV_i^*}{B} \quad (7)$$

กำหนดให้ $S_{CQV_B}^2$ คือ ค่าความแปรปรวนของตัวประมาณแบบจุดของสัมประสิทธิ์การแปรผันควอไทล์วิธี Bootstrap ซึ่งคำนวณได้ดังนี้

$$S_{CQV_B}^2 = \frac{\sum_{i=1}^B (CQV_i^* - CQV_B)^2}{B} \quad (8)$$

เมื่อ B คือ จำนวนครั้งการสุ่มซ้ำโดยวิธี Bootstrap

ฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวประมาณค่าควอไทล์ที่ 1 ($\hat{f}_{1B}$) และควอไทล์ที่ 3 ($\hat{f}_{3B}$) วิธี Bootstrap คำนวณได้ดังนี้

$$\hat{f}_{1B} = \sqrt{3(z_1 - \alpha/2)^2 / \{4n(\hat{Y}_{B(a)} - \hat{Y}_{B(a)})^2\}} \quad (9)$$

$$\hat{f}_{3B} = \sqrt{3(z_1 - \alpha/2)^2 / \{4n(\hat{Y}_{B(d)} - \hat{Y}_{B(d)})^2\}} \quad (10)$$

เมื่อ $\hat{Y}_{B(j)} = \sum_{i=1}^B \hat{Y}_{(j)i}^* / B$; $j = a, b, c, d$ โดยที่ $\hat{Y}_{(j)i}^*$ เป็นค่าสถิติลำดับที่ j รอบที่ i ของตัวอย่าง Bootstrap และ

$$(\hat{Y}_{(a)i}^*, \hat{Y}_{(b)i}^*) = \left(\hat{Y}_{(n/4 - 1.96\sqrt{3n/16})i}^*, \hat{Y}_{(n/4 + 1.96\sqrt{3n/16})i}^* \right)$$

$$(\hat{Y}_{(d)i}^*, \hat{Y}_{(c)i}^*) = (\hat{Y}_{(n+1-b)i}^*, \hat{Y}_{(n+1-a)i}^*)$$

$$\text{กำหนดให้ } D_B = \sum_{i=1}^B (\hat{Q}_{3i}^* - \hat{Q}_{1i}^*) / B$$

$$\text{และ } S_B = \sum_{i=1}^B (\hat{Q}_{3i}^* + \hat{Q}_{1i}^*) / B \text{ เมื่อแทน } \hat{f}_{1B}, \hat{f}_{3B}, D$$

และ S ด้วย $\hat{f}_{1B}, \hat{f}_{3B}, D_B$ และ S_B ตามลำดับ ในสมการที่ (2) จะได้ค่าความแปรปรวนของสัมประสิทธิ์การแปรผันควอไทล์วิธี Bonett Bootstrap (V_B) ดังนี้

$$V_B = \left(\frac{1}{16n} \right) \left\{ \frac{\left(\frac{3}{\hat{f}_{1B}^2} + \frac{3}{\hat{f}_{3B}^2} - \frac{2}{\hat{f}_{1B}\hat{f}_{3B}} \right)}{D_B^2} + \frac{\left(\frac{3}{\hat{f}_{1B}^2} + \frac{3}{\hat{f}_{3B}^2} - \frac{2}{\hat{f}_{1B}\hat{f}_{3B}} \right)}{S_B^2} - 2 \frac{\left(\frac{3}{\hat{f}_{1B}^2} - \frac{3}{\hat{f}_{3B}^2} \right)}{D_B S_B} \right\} \quad (11)$$

กำหนดให้ T_i^* คือ ค่า Bootstrap - t สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์คำนวณได้ดังนี้

$$T_i^* = \frac{(CQV_i^* - CQV_B)}{S_{CQV_B}}; \quad i = 1, 2, \dots, B \quad (12)$$

จากสมการ (1) เมื่อแทน CQV, V และ $Z_1 - \alpha^*/2$ ด้วย CQV_B, V_B และ $T_{(1-\alpha/2)B}^*$ ตามลำดับ โดยเมื่อเรียงค่า T_i^* จากค่าน้อยไปหาค่ามาก ซึ่งแทนด้วย $T_{(1)}^*, T_{(2)}^*, \dots, T_{(B)}^*$ ค่า $T_{(1-\alpha/2)B}^*$ คือ T^* ลำดับที่ $(1-\alpha/2)B$ จะได้ช่วงความเชื่อมั่น $(1-\alpha) 100 \%$ สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Bonett Bootstrap ดังนี้

$$\exp \left[\ln(CQV_B) \pm T_{(1-\alpha/2)B}^* V_B^{1/2} \right] \quad (13)$$

2.5 ช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Nonparametric Bootstrap (NB method)

ช่วงความเชื่อมั่นวิธี Nonparametric Bootstrap จะใช้ค่าที่เรียกว่า Percentile Bootstrap โดยจะนำค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Bootstrap แต่ละรอบมาเรียงจากค่าน้อยไปหาค่ามาก ซึ่งแทนด้วย $CQV_{(1)}^*$, $CQV_{(2)}^*$, ..., $CQV_{(B)}^*$ กำหนดให้ $CQV_{(BL)}^*$ คือค่า CQV^* ลำดับที่ $(\alpha/2)B$ และ $CQV_{(BU)}^*$ คือค่า CQV^* ลำดับที่ $(1 - \alpha/2)B$ จะได้ช่วงความเชื่อมั่น $(1 - \alpha)$ 100% สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Nonparametric Bootstrap ดังนี้

$$[CQV_{(BL)}^*, CQV_{(BU)}^*] \quad (14)$$

ดังนั้น $CQV_{(BL)}^*$ และ $CQV_{(BU)}^*$ คือ ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่น $(1 - \alpha)$ 100 % สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Nonparametric Bootstrap

2.6 ช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Parametric Bootstrap (PB method)

ช่วงความเชื่อมั่น Parametric Bootstrap จะใช้ตัวประมาณแบบจุดและความแปรปรวนของตัวประมาณแบบจุดของสัมประสิทธิ์การแปรผันควอไทล์วิธี Bootstrap จากสมการ (7) และ (8) มาคำนวณและใช้ค่า Bootstrap - t จากวิธี Bonett Bootstrap ตามสมการ (13) มาสร้างช่วงความเชื่อมั่น จะได้ช่วงความเชื่อมั่น $(1 - \alpha)$ 100% สำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์วิธี Parametric Bootstrap ดังนี้

$$[CQV_B - T_{(1 - \alpha/2)B}^* S_{CQV_B}, CQV_B + T_{(1 - \alpha/2)B}^* S_{CQV_B}] \quad (15)$$

3. วิธีวิจัย

3.1 ขอบเขตการวิจัย

3.1.1 งานวิจัยนี้ได้ศึกษากรณีข้อมูลมีการแจกแจงสมมาตร คือ $N(4, 1)$ และกรณีข้อมูลมีการแจกแจงเบ้เชิงบวกระดับต่าง ๆ คือการแจกแจง $LN(\mu, \sigma^2)$ และ $G(\alpha, \beta)$ ดังนี้ การแจกแจงเบ้เชิงบวกน้อย โดยมีค่าสัมประสิทธิ์ความเบ้มากกว่า 0 ถึง 0.5 ได้แก่การแจกแจง $LN(0, 0.25)$ และ $G(9, 2)$ การแจกแจงเบ้เชิงบวกปานกลาง โดยมีค่าสัมประสิทธิ์ความเบ้มากกว่า 0.5 ถึง 1 ได้แก่การแจกแจง $LN(0, 0.5)$ และ $G(6, 1)$ และการแจกแจงเบ้เชิงบวกมาก โดยมีค่าสัมประสิทธิ์ความเบ้มากกว่า 1 ได้แก่การแจกแจง $LN(0, 0.75)$, $LN(0, 1)$, $LN(0, 1.5)$, $G(2, 0.5)$ และ $G(0.5, 0.5)$ [8]

3.1.2 กำหนดขนาดตัวอย่าง่างเพื่อให้ครอบคลุมทั้งจำนวนคู่และจำนวนคี่ กรณีตัวอย่างขนาดเล็ก ($n = 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 25, 26, 29, 30$) ตัวอย่างขนาดกลาง ($n = 35, 36, 39, 40, 45, 46, 49, 50$) และตัวอย่างขนาดใหญ่ ($n = 99, 100$)

3.1.3 กำหนดระดับความเชื่อมั่น $(1 - \alpha)$ เท่ากับ 0.95

3.1.4 กำหนดจำนวนครั้งการสุ่มซ้ำโดยวิธี Bootstrap เท่ากับ 1,000 ครั้ง

3.1.5 จำลองข้อมูลในสถานการณ์ต่าง ๆ ซ้ำจำนวน 5,000 รอบ ด้วยโปรแกรม R

3.1.6 วิเคราะห์ประสิทธิภาพของช่วงความเชื่อมั่นจากค่าความน่าจะเป็นคัมรวม (coverage probability, CP) โดยพิจารณาเฉพาะช่วงความเชื่อมั่นที่มีค่า CP อยู่ในช่วง $[0.9444, 0.9561]$ และมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (average width of confidence interval, AW) แคบที่สุดจะถือว่าเป็นช่วงความเชื่อมั่นที่มีประสิทธิภาพตรงตามเกณฑ์ที่ระดับความเชื่อมั่น 0.95 โดยสำหรับการหาค่า AW จะ

ทำการหาเฉพาะช่วงความเชื่อมั่นที่มีค่า CP อยู่ในช่วง [0.9444, 0.9561] เท่านั้น

3.2 ขั้นตอนการวิจัย

3.2.1 สร้างข้อมูลประชากรจำนวน 5,000 ค่าสังเกตตามการแจกแจงและค่าพารามิเตอร์ที่กำหนด

3.2.2 คำนวณค่าสัมประสิทธิ์การแปรผัน ควอไทล์ของประชากร

3.2.3 สุ่มตัวอย่างขนาด n จากประชากร ในขั้นตอนที่ 1 และตรวจสอบว่าตัวอย่างมีสมบัติตามที่กำหนด

3.2.4 คำนวณค่าควอไทล์ 5 วิธี ได้แก่ วิธี Tukey วิธี M&M วิธี M&S วิธี F&P และวิธี Minitab

3.2.5 นำค่าควอไทล์ 5 วิธี ที่ได้ในขั้นตอนที่ 4 มาคำนวณช่วงความเชื่อมั่น 95 % สำหรับค่าสัมประสิทธิ์การแปรผัน ควอไทล์แต่ละวิธี ได้แก่วิธี BB วิธี NB และวิธี PB ตามสมการ (13), (14) และ (15) ตามลำดับ

3.2.6 ตรวจสอบช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์รอบที่ i ว่าครอบคลุมค่าพารามิเตอร์หรือไม่ เมื่อกำหนดให้ C_i คือดัชนีที่แสดงการครอบคลุมค่าพารามิเตอร์ของช่วงความเชื่อมั่นรอบที่ i โดย C_i เท่ากับ 1 เมื่อช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์ และ C_i เท่ากับ 0 เมื่อช่วงความเชื่อมั่นไม่ครอบคลุมค่าพารามิเตอร์ และคำนวณความกว้างของช่วงความเชื่อมั่นจาก $U_i - L_i$ เมื่อ U_i คือ ขีดจำกัดบนของช่วงความเชื่อมั่นรอบที่ i และ L_i คือ ขีดจำกัดล่างของช่วงความเชื่อมั่นรอบที่ i

3.2.7 ดำเนินขั้นตอนที่ 3 ถึง 6 ซ้ำจำนวน 5,000 รอบ

3.2.8 คำนวณค่า CP และค่า AW จากสูตร $CP = \sum_{i=1}^{5,000} C_i / 5,000$ และ $AW = \sum_{i=1}^{5,000} (U_i - L_i) / 5,000$

3.2.9 ทำซ้ำขั้นตอนที่ 1 ถึง 8 จนครบทุก

สถานการณ์ที่กำหนดในขอบเขตการวิจัย

4. ผลการวิจัย

เมื่อพิจารณาแต่ละวิธีของช่วงความเชื่อมั่น ได้แก่ ช่วงความเชื่อมั่นวิธี Bonett Bootstrap วิธี Nonparametric Bootstrap และวิธี Parametric Bootstrap ว่าวิธีประมาณค่าควอไทล์ของประชากรวิธีใดที่ให้ช่วงความเชื่อมั่นมีประสิทธิภาพ ที่ระดับความเชื่อมั่น 0.95 ในแต่ละระดับความเบ้ แต่ละระดับการแจกแจง และแต่ละระดับขนาดตัวอย่าง สรุปผลดังแสดงในตารางที่ 1 ซึ่งพบว่าแต่ละวิธีประมาณค่าควอไทล์ให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพต่างกัน โดยขึ้นอยู่กับทั้งระดับความเบ้ การแจกแจง และขนาดตัวอย่าง แต่ก็มีขนาดตัวอย่างที่เป็นจำนวนคู่และวิธีประมาณค่าควอไทล์บางวิธีที่ให้ช่วงความเชื่อมั่นไม่มีประสิทธิภาพเลย ได้แก่ ช่วงความเชื่อมั่นวิธี Bonett Bootstrap เมื่อประมาณค่าควอไทล์ด้วยวิธี Mendenhall and Sincich และด้วยวิธี Freund and Perles และช่วงความเชื่อมั่นวิธี Nonparametric Bootstrap เมื่อประมาณค่าควอไทล์ด้วยวิธี Mendenhall and Sincich

5. สรุปผลการศึกษา

งานวิจัยนี้ได้ศึกษาประสิทธิภาพของช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การแปรผันควอไทล์ทั้ง 3 วิธี คือ วิธี Bonett Bootstrap วิธี Nonparametric Bootstrap และวิธี Parametric Bootstrap ซึ่งแต่ละวิธีช่วงความเชื่อมั่นประมาณควอไทล์ด้วยวิธีต่าง ๆ 5 วิธี ได้แก่ วิธี Tukey วิธี Moore and McCabe วิธี Mendenhall and Sincich วิธี Freund and Perles และวิธี Minitab ผลการวิจัยพบว่าช่วงความเชื่อมั่นวิธี Bonett Bootstrap หากประมาณค่าควอไทล์ด้วยวิธี Minitab จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเมื่อ

Table 1 The quartile estimation methods has efficiency of confidence intervals skewnesses distribution and sizes of sample at confidence level 0.95.

Confidence interval methods	Quartile estimation methods	Levels of skewnesses	Distribution	Sizes of sample (n)
BB	Tukey	-	-	-
	M&M	highly	G(0.5, 0.5)	19 ≤ n ≤ 99 ; only n is odd
	M&S	-	-	-
	F&P	-	-	-
	Minitab	symmetric	N(4, 1)	5 ≤ n ≤ 8
		fairly	LN(0, 0.25)	
		highly	G(0.5, 0.5)	19 ≤ n ≤ 99
NB	Tukey	fairly	G(9, 2)	9 ≤ n ≤ 15 ; only n is odd
		moderately	G(6, 1)	
		highly	LN(0, 0.75)	
		symmetric	N(4, 1)	9 ≤ n ≤ 19 ; only n is odd
		highly	LN(0, 1) and G(2, 0.5)	
	M&M	highly	G(0.5, 0.5)	7 ≤ n ≤ 13 ; only n is odd
	M&S	-	-	-
	F&P	fairly	G(9, 2)	9 ≤ n ≤ 15
		moderately	G(6, 1)	
		highly	LN(0, 0.75)	
		symmetric	N(4, 1)	9 ≤ n ≤ 20
		highly	G(2, 0.5)	
		highly	LN(0, 1)	
		highly	G(0.5, 0.5)	35 ≤ n ≤ 46
	Minitab	highly	G(0.5, 0.5)	7 ≤ n ≤ 13
PB	Tukey	fairly	G(9, 2)	9 ≤ n ≤ 15 ; only n is odd
		moderately	G(6, 1)	
		symmetric	N(4, 1)	11 ≤ n ≤ 15 ; only n is odd
		moderately	LN(0, 0.5)	
	M&M	highly	LN(0, 1.5)	13 ≤ n ≤ 99 ; only n is odd
	M&S	highly	LN(0, 1.5)	10 ≤ n ≤ 20 ; only n is odd
	F&P	moderately	LN(0, 0.5) and G(6, 1)	10 ≤ n ≤ 15
		symmetric	N(4, 1)	10 ≤ n ≤ 20
		fairly	G(9, 2)	
	Minitab	highly	LN(0, 0.75)	7 ≤ n ≤ 13
		highly	LN(0, 1)	11 ≤ n ≤ 49
		highly	LN(0, 1.5)	13 ≤ n ≤ 25 ; only n is odd
		highly	G(2, 0.5)	25 ≤ n ≤ 36
		highly	LN(0, 1.5)	29 ≤ n ≤ 99

ข้อมูลมีการแจกแจงสมมาตรหรือเบ้เชิงบวกเพียงเล็กน้อยเฉพาะกรณีที่ตัวอย่างมีขนาดเล็ก แต่วิธี Minitab จะไม่ให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเลย เมื่อข้อมูลมีการแจกแจงเบ้เชิงบวกระดับปานกลาง แต่กรณีข้อมูลมีการแจกแจงเบ้เชิงบวกมากวิธี Minitab จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพในหลาย ๆ ระดับของขนาดตัวอย่าง สำหรับช่วงความเชื่อมั่นวิธี Nonparametric Bootstrap เมื่อประมาณค่าควอไทล์ด้วยวิธี minitab จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเมื่อข้อมูลมีการแจกแจงเบ้เชิงบวกมาก เฉพาะกรณีที่ตัวอย่างมีขนาดเล็ก แต่พบว่าวิธี Freund and Perles จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเสมอไม่ว่าข้อมูลจะมีการแจกแจงสมมาตรหรือเบ้เชิงบวกระดับใดทั้งกรณีที่ตัวอย่างมีขนาดเล็กและขนาดกลาง ส่วนวิธี Tukey จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเสมอไม่ว่าข้อมูลจะมีการแจกแจงสมมาตรหรือเบ้เชิงบวกระดับใดเฉพาะกรณีที่ตัวอย่างมีขนาดเล็กและเป็นจำนวนคี่ สำหรับช่วงความเชื่อมั่นวิธี Parametric Bootstrap พบว่าหากประมาณค่าควอไทล์ด้วยวิธี Tukey จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเมื่อข้อมูลมีการแจกแจงสมมาตรหรือเบ้เชิงบวกน้อยถึงปานกลางเฉพาะกรณีที่ตัวอย่างมีขนาดเล็กและเป็นจำนวนคี่ ส่วนวิธี Freund and Perles จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเมื่อข้อมูลมีการแจกแจงสมมาตรหรือเบ้เชิงบวกน้อยถึงปานกลาง เฉพาะกรณีที่ตัวอย่างมีขนาดเล็ก ส่วนเมื่อข้อมูลมีการแจกแจงเบ้เชิงบวกมากจะไม่ให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเลย ส่วนการประมาณค่าควอไทล์ด้วยวิธี Minitab จะให้ช่วงความเชื่อมั่นที่มีประสิทธิภาพเมื่อข้อมูลมีการแจกแจงเบ้เชิงบวกสูงในหลายระดับขนาดตัวอย่าง

6. References

- [1] Ruengpraprapan, C., 2000, Basic Statistics, Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen, 509 p. (in Thai)
- [2] Langford, E., 2006, Quartiles in elementary statistics, J. Stat. Edu. 14: 1-16.
- [3] Cangur, S., Pasin, O. and Ankarali, H., 2015, Comparison of sampling distributions and performances of Minitab and Freund & Perles quartile, Pak. J. Stat. 31: 1-20.
- [4] Bonett, D.G., 2006, Confidence interval for a coefficient of quartile variation, Comput. Stat. Data Anal. 50: 2953-2957.
- [5] Tongkaw, A. and Pongsakchat, V., 2014, Confidence intervals for a coefficient of quartile variation with bootstrap method, pp. 19-21, ICAS 2014, Khon Kaen.
- [6] Altunkaynak, B. and Gamgam, H., Bootstrap Confidence Intervals for the Coefficient of Quartile Variation, Communications in Statistics – Simulation and Computation, Available Source: <https://doi.org/10.1080/03610918.2018.1435800>, February 18, 2018.
- [7] Efron, B. and Tibshirani, R.J., 1993, An Introduction to the Bootstrap, Chapman & Hall, London, 436 p.
- [8] Bulmer, M.G., 1967, Principle of Statistics, Dover, New York, 252 p.