การเปรียบเทียบประสิทธิภาพการประมาณค่าช่วงความเชื่อมั่นของความแปรปรวนของการแจกแจงปกติและปกติปลอมปนด้วยเทคนิคการสุ่มตัวอย่างซ้ำ

A COMPARISON OF EFFICIENCY ON INTERVAL ESTIMATION OF VARIANCE ON NORMAL AND CONTAMINTED NORMAL DISTRIBUTIONS BY RESAMPLING TECHNIQUES

กิตติคุณ สุภาวณิชย์, *อัชฌา อระวีพร**

Kittikhun Supawnith, *Autcha Araveeporn**

ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang.

**Corresponding author, e-mail: autcha.ar@kmitl.ac.th*

Received: *12 January 2021;* ***Revised:*** *1 March 2022.;* ***Accepted:*** *23 June 2022*

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการประมาณช่วงความเชื่อมั่นของความแปรปรวนในการแจกแจงปกติและปกติปลอมปน ด้วยวิธีความแปรปรวนของตัวอย่าง และเทคนิคการสุ่มตัวอย่างซ้ำซึ่งประกอบไปด้วยวิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน โดยสุ่มข้อมูลปกติ มีค่าเฉลี่ย (μ) เท่ากับ 2 และความแปรปรวน (σ^2) เท่ากับ 2 และ 6 และจากประชากรที่มีการแจกแจงปกติปลอมปน ที่มีค่าเฉลี่ยค่าเฉลี่ยและค่าความแปรปรวนเท่ากับการแจกแจงปกติ โดยกำหนดสัดส่วนการปลอมปน (P) เท่ากับ 0.1 และสเกลแฟคเตอร์ (C) เท่ากับ 2 และ 5 ขนาดตัวอย่าง (n) ที่ใช้ในการศึกษา คือ 10, 20, 30 และ 50 ที่ระดับความเชื่อมั่น 95% และ 99% ในการจำลองข้อมูลใช้โปรแกรมอาร์ในการวิจัย โดยทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ เกณฑ์ที่ใช้พิจารณาหาวิธีการประมาณช่วงความเชื่อมั่นที่ดีที่สุดจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด จากผลการวิจัยพบว่า เมื่อข้อมูลเป็นการแจกแจงแบบปกติ วิธีความแปรปรวนของตัวอย่างให้ผลการประมาณช่วงความเชื่อมั่นที่ดีที่สุด ส่วนการแจกแจงแบบปกติปลอมปนวิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐานจะให้ผลที่ดีกว่าวิธีความแปรปรวนตัวอย่าง และวิธีบูตสแตรปมาตรฐานมีค่าความกว้างเฉลี่ยความเชื่อมั่นแคบกว่าวิธีแจ็คไนฟ์

คำสำคัญ: การแจกแจงปกติ การแจกแจงปกติปลอมปน ความแปรปรวนของตัวอย่าง แจ็คไนฟ์ บูตสแตรปมาตรฐาน

Abstract

The objective of this research is to compare the efficient confidence interval of variance based on normal and contaminated normal distribution by sample variance method and resampling techniques consisted of jackknife and standard bootstrap methods. The data sample are generated from the normal distribution with the parameter of mean (μ) 2 and variance (σ^2) 2 and 6. The mean and variance of contaminated normal distribution are the same as the normal distribution, but proportion of contamination (P) is 0.1 and scale factor (c) is 2 and 5. The sample sizes (n) for this study are set as 10, 20, 30 and 50, and the 95% and 99% confidence intervals. The simulated data is generated by R program and repeated 1,000 times in each situation. The criterion to consider the best confidence interval is a minimum value of average width. On normal distribution, the results are shown that sample variance method is the best confidence interval estimation. For contaminated normal distribution, jackknife and standard bootstrap methods outperform the sample mean method, and the standard bootstrap method has the average width narrower than jackknife method.

Keywords: Normal Distribution, Contaminated Normal Distribution, Sample Variance, Jackknife, Standard Bootstrap

บทนำ

การประมาณค่า วิธีการหนึ่งที่สำคัญสำหรับการอนุมานเชิงสถิติ (Statistical Inference) โดยอาศัยข้อมูลจากตัวอย่างในการอธิบายลักษณะที่สำคัญของประชากรที่เรียกว่าพารามิเตอร์ (Parameter) โดยสรุปลักษณะของประชากรในรูปของเชิงตัวเลขในการสร้างตัวประมาณ (Estimator) สำหรับการประมาณค่าพารามิเตอร์ การประมาณค่าพารามิเตอร์ในทางสถิติสามารถกระทำได้ 2 รูปแบบ คือ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation)

การประมาณค่าแบบจุด (Point Estimation) เป็นวิธีการประมาณค่าที่ใช้ค่าประมาณเพียงค่าเดียวในการประมาณค่าพารามิเตอร์ แต่ในการประมาณค่าด้วยรูปแบบนี้มีโอกาสที่ค่าประมาณมีความคลาดเคลื่อนไปจากพารามิเตอร์ได้มากเนื่องจากใช้ค่าประมาณเพียงค่าเดียว ดังนั้น การประมาณค่าแบบช่วง (Interval Estimation) จึงเป็นเป็นวิธีการประมานค่าพารามิเตอร์อีกรูปแบบหนึ่งที่ช่วยลดความคลาดเคลื่อนจากการประมาณค่าพารามิเตอร์มากกว่าการประมาณค่าพารามิเตอร์รูปแบบจุด โดยอาศัยการประมาณค่าที่เป็นช่วงของตัวเลขและเชื่อมั่นว่าช่วงดังกล่าวจะครอบคลุมค่าที่แท้จริงของพารามิเตอร์ สิ่งแรกในการหาค่าประมาณพารามิเตอร์แบบช่วงจะต้องกำหนดสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ ในการสร้างขอบเขตของช่วง 2 ค่า คือ ขีดจำกัดล่าง (Lower Limit : L) และขีดจำกัดบน (Upper Limit : U) เพื่ออยู่ในรูปแบบ $P(\mathrm{L} \le \theta \le \mathrm{U}) = 1-\alpha$ แล้วจะเรียก $(\mathrm{L} \le \theta \le \mathrm{U})$ ว่าช่วงความเชื่อมั่น (Confidence Interval)

ในการประมาณค่าพารามิเตอร์นั้นพิจารณาจากรูปแบบของการแจกแจงว่ามีพารามิเตอร์ใดบ้าง ในงานวิจัยนี้สนใจประมาณค่าความแปรปรวนของรูปของการแจกแจงปกติ (Normal Distribution) ซึ่งข้อมูลในชีวิตประจำวันส่วนใหญ่อยู่ในรูปแบบของการแจกแจงนี้ กล่าวคือ ข้อมูลมีลักษณะเส้นโค้งรูประฆังคว่ำโดยสมมาตรรอบค่าเฉลี่ย (μ) และลักษณะการกระจายของการแจกแจงด้วยค่าความแปรปรวน (σ^2) โดยปกติแล้วจะไม่สามารถหาค่าเฉลี่ยและความแปรปรวน ของประชากรได้จึงทำการสุ่มตัวอย่างเพื่อประมาณค่าเฉลี่ย $(\bar{X})$ สำหรับหาตัวแทนของข้อมูลและค่าความแปรปรวน (s^2) ในการวัดค่าการกระจายของข้อมูล โดยใช้ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ในการประมาณค่าความแปรปรวนและใช้ตัวสถิติไคกำลังสอง (χ^2) ในการสร้างช่วง

ความเชื่อมั่นของ s^2 โดยตัวประมาณดังกล่าวมีคุณสมบัติ ตัวประมาณที่ไม่เอนเอียงซึ่งเป็นเกณฑ์หนึ่งที่สำคัญสำหรับการประมาณค่า [1] แต่เนื่องจากสูตร s^2 ดังกล่าวถูกสร้างขึ้นมาจากฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่มที่มีการแจกแจงปกติ ดังนั้นหากข้อมูลไม่ได้เป็นไปตามตัวแปรสุ่มดังกล่าว อาทิ ข้อมูลมีค่านอกเกณฑ์ การประมาณค่าความแปรปรวนด้วยวิธีดังกล่าวอาจจะไม่เหมาะสม

การแจกแจงปกติปลอมปน (Contaminated Normal Distribution) เป็นการแจกแจงผสมระหว่าง 2 การแจกแจงปกติ โดยการแจกแจงหนึ่งมีข้อมูลเป็นรูปแบบของการแจกแจงปกติที่ค่าเฉลี่ยและความแปรปรวนปกติ และอีกกลุ่มหนึ่งจะถูกสุ่มจากข้อมูลทีค่าเฉลี่ยสูงมากโดยถ่วงน้ำหนักด้วยค่าความน่าจะเป็น ซึ่งเรียกว่าสัดส่วนการปลอมปน ซึ่งค่าที่ได้จะแสดงข้อมูลที่มีลักษณะเป็นค่านอกเกณฑ์ (Outlier) คือข้อมูลบางค่าที่มีความแตกต่างจากข้อมูลอื่น ๆ มาก บางค่าอาจสูงเกินไปหรือต่ำเกินไป ซึ่งข้อมูลเหล่านี้มีสาเหตุเกิดความผิดพลาดจากขั้นตอนการเก็บข้อมูลซึ่งแก้ไขได้โดยตัดข้อมูลที่ผิดปกติออกจากการวิเคราะห์ หรือ อาจเกิดความผิดพลาดจากธรรมชาติในตัวข้อมูลที่เรียกว่าความผิดพลาดเชิงสุ่ม [2] ซึ่งค่านอกเกณฑ์ในกรณีดังกล่าวอาจบ่งบอกถึงลักษณะที่สำคัญได้ ดังนั้นจึงต้องนำข้อมูลที่ค่าผิดปกติมาวิเคราะห์ด้วย โดยนักวิจัยใช้การประมาณค่าความแปรปรวนด้วยวิธีการอื่น ๆ ที่เหมาะสมในการสร้างตัวประมาณและช่วงความเชื่อมั่นเมื่อข้อมูลไม่ได้อยู่ในรูปแบบการแจกแจงปกติ

การประมาณค่าพารามิเตอร์โดยวิธีแจ็คไนฟ์ (Jackknife Method) [3] เป็นวิธีการประมาณจากข้อมูลเดิมจำนวน n ค่า แล้วทำการสร้างตัวอย่างชุดใหม่โดยตัดค่าสังเกตหนึ่งครั้งออกจากตัวอย่างตามลำดับ จะเรียกตัวอย่างสุ่มชุดใหม่ว่า ตัวอย่างแจ็คไนฟ์ (Jackknife Samples) แล้วคำนวณค่าสถิติที่ต้องการประมาณด้วยข้อมูลตัวที่เหลือโดยค่าที่ถูกตัดออกจะใส่คืนกลับไปในตัวอย่างก่อนที่จะสร้างชุดข้อมูลชุดใหม่ในครั้งถัดไป ทำไปเรื่อย ๆ จนครบ n ครั้ง หลังจากนั้นทำการประมาณค่าพารามิเตอร์และสร้างช่วงความเชื่อมั่นด้วยค่าสถิติที (t) ในการประมาณค่าพารามิเตอร์ด้วยวิธีแจ็คไนฟ์เป็นวิธีการหนึ่งที่ช่วยลดความเอนเอียงของตัวประมาณ และสังเกตตัวอย่างชุดใหม่ แต่ละชุดได้ว่าจะเกิดอะไรขึ้นเมื่อตัดค่าสังเกตหนึ่งค่าออกจากชุดข้อมูล แต่เนื่องจากวิธีแจ็คไนฟ์มีข้อจำกัดเรื่องการสร้างข้อมูลชุดใหม่สำหรับการทำซ้ำมักขึ้นอยู่กับ ขนาดตัวอย่างของข้อมูลเดิม n ค่าเท่านั้น

การประมาณค่าพารามิเตอร์โดยวิธีบูตสแตรป (Bootstrap Method) [4] เป็นวิธีการประมาณจากข้อมูลเดิมจำนวน n ค่า แล้วทำการสร้างข้อมูลชุดใหม่โดยทำการสุ่มตัวอย่างซ้ำจากข้อมูลเดิมแบบคืนที่ (With Replacement) จำนวน n ค่าและจะเรียกตัวอย่างสุ่มชุดใหม่ว่า ตัวอย่างบูตสแตรป (Bootstrap Samples) จากนั้นคำนวณค่าสถิติที่ต้องการประมาณ โดยวิธีบูตสแตรปไม่ได้มีข้อจำกัดในการกำหนดจำนวนชุดข้อมูล กล่าวคือกำหนดจำนวนครั้งของการทำซ้ำแบบบูตสแตรปตามผู้ใช้งาน ซึ่งปกติมักจะกำหนดจำนวนครั้งของการทำซ้ำแบบบูตสแตรปเป็นจำนวนมาก ๆ พอจนสามารถสร้างการแจกแจงของตัวสถิติในการประมาณค่าพารามิเตอร์แบบช่วงได้ ส่วนใหญ่มักจะให้การแจกแจงของตัวสถิติอยู่ในรูปแบบการแจกแจงปกติเพื่อสร้างช่วงความเชื่อมั่นของพารามิเตอร์โดยวิธีบูตสแตรปมาตรฐาน (Standard Bootstrap)

วิธีแจ็คไนฟ์และวิธีบูตสแตรปมาตรฐานเป็นวิธีของเทคนิคการสุ่มตัวอย่างซ้ำ (Resampling Techniques) ซึ่งเป็นวิธีการทางสถิติที่ไม่ใช้พารามิเตอร์ (Nonparametric Statistics) โดยไม่จำเป็นต้องมีข้อสมมุติลักษณะการแจกแจงของประชากรเข้ามาเกี่ยวข้อง ทำให้นักวิจัยหลายท่านได้ทำการศึกษาประสิทธิภาพการประมาณช่วงความเชื่อมั่นด้วยเทคนิคการสุ่มตัวอย่างซ้ำ การเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ขนาดของการแจกแจงไวบูลล์แบบสองพารามิเตอร์ [5] โดยเปรียบเทียบระหว่างวิธีบูตสแตรปทีและวิธีแจ็คไนฟ์ พบว่าวิธีบูตสแตรปทีเป็นวิธีที่ดีที่สุดในการประมาณค่าทุก ๆ สถานการณ์ที่ศึกษา การประมาณช่วงความเชื่อมั่นของค่ามัธยฐานสำหรับการแจกแจงเลขชี้กำลัง 3 วิธี [6] ประกอบด้วย วิธีใช้ค่ามัธยฐานของตัวอย่างเป็นตัวประมาณค่า วิธีบูตสแตรปมาตรฐาน และวิธีเปอร์เซ็นไทล์บูตสแตรป ผลการศึกษาพบว่าวิธีบูตสแตรปมาตรฐานให้ค่าความ

น่าจะเป็นครอบคลุมไม่แตกต่างจากค่าสัมประสิทธิ์ความเชื่อมั่นทุกสถานการณ์ที่ศึกษา และเมื่อพิจารณาความกว้างเฉลี่ยพบว่าวิธีบูตสแตรปมาตรฐานให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเมื่อขนาดตัวอย่างน้อยกว่าห้าสิบ ศึกษาการประมาณช่วงความเชื่อมั่นของอัตราส่วนความแปรปรวนสองประชากร [7] เมื่อข้อมูลมีการแจกแจงปกติ การแจกแจงแกมมา และการแจกแจงแบบเลขชี้กำลัง ด้วยวิธีเอฟ วิธีขีดจำกัดส่วนกลาง วิธีบูตสแตรป และการประยุกต์ระหว่างวิธีบูตสแตรป พบว่าวิธีเอฟเหมาะสมในการประมาณช่วงความเชื่อมั่นอัตราส่วนความแปรปรวนสองประชากร เมื่อข้อมูลมีการแจกแจงปกติ ส่วนกรณีข้อมูลการแจกแจงแกมมาและการแจกแจงแบบเลขชี้กำลัง การประยุกต์ระหว่างวิธีบูตสแตรป มีประสิทธิภาพในการประมาณค่าช่วงความเชื่อมั่นดีกว่าวิธีอื่น ๆ นอกจากนี้ยังมีงานวิจัยที่ศึกษาการประมาณช่วงความเชื่อมั่นของอัตราส่วนความแปรปรวน [8] โดยใช้วิธีความโด่งของเพียร์สัน และมีการศึกษาความประมาณช่วงความเชื่อมั่นของสัมประสิทธิ์ความแปรผัน [9] ของการแจกแจงแกมมาโดยใช้วิธีการประมาณค่าความแปรปรวนแบบย้อนกลับ

แนวคิดวิธีการประมาณค่าความแปรปรวนที่กล่าวไว้ข้างต้น จึงเป็นแนวทางในการศึกษาการประมาณช่วงความเชื่อมั่นของความแปรปรวน 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่างซึ่งเป็นวิธีที่ผู้คนส่วนใหญ่นิยมใช้ในการศึกษาและประมาณค่าความแปรปรวน วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน เป็นเทคนิคการสุ่มตัวอย่างซ้ำสำหรับการประมาณค่าพารามิเตอร์ซึ่งโดยส่วนใหญ่วิธีนี้จะใช้ในการประมาณค่าเฉลี่ยของข้อมูลที่มีการแจกแจงแบบปกติปลอมปนเพื่อขจัดอิทธิพลของข้อมูลที่มีค่านอกเกณฑ์ จากแนวความคิดดังกล่าวผู้วิจัยจึงนำมาประยุกต์เพื่อใช้ในการประมาณค่าความแปรปรวนกรณีข้อมูลมีลักษณะการแจกแจงปกติ และการแจกแจงปกติปลอมปน

วัตถุประสงค์ของการวิจัย

1. เปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นของความแปรปรวน 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่าง วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน กรณีข้อมูลมีลักษณะการแจกแจงปกติ

2. เปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นของความแปรปรวน 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่าง วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน กรณีข้อมูลมีค่านอกเกณฑ์โดยจำลองข้อมูลให้อยู่ในรูปการแจกแจงปกติปลอมปน

วิธีดำเนินการวิจัย

1. จำลองข้อมูลและสุ่มตัวอย่างโดยใช้โปรแกรม R กำหนดขอบเขตงานวิจัยดังต่อไปนี้

1.1 ตัวแปรสุ่ม X มีการแจกแจงปกติ (Normal Distribution) $X \sim N(\mu,\sigma^2)$ เมื่อ μ คือค่าเฉลี่ย และ σ^2 คือความแปรปรวน โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

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

โดยที่ ค่าเฉลี่ยของตัวแปรสุ่ม X คือ $E(X)=\mu$ และความแปรปรวนของตัวแปรสุ่ม คือ $Var(X)=\sigma^2$ ซึ่งจำลองข้อมูลให้มีลักษณะการแจกแจงปกติ 2 รูปแบบ คือ $N(2,2)$ และ $N(2,6)$

1.2 ตัวแปรสุ่ม X มีการแจกแจงปกติปลอมปน (Contaminated Normal Distribution) ซึ่งฟังก์ชันความหนาแน่นความน่าจะเป็น X อยู่ในรูปแบบ

$$f(x)=(1-P)N(\mu,\sigma^2)+PN(\mu,\mathrm{c}^2\sigma^2)$$

เมื่อ P คือ สัดส่วนการปลอมปน โดยที่ $0\le P\le 1$ และ c คือ สเกลแฟกเตอร์ โดยที่ $\mathrm{c}>0$

โดยจำลองข้อมูลให้มีลักษณะการแจกแจงปกติปลอมปนทั้งหมด 4 รูปแบบโดยกำหนดให้

$(1-P)N(2,2)+PN(2,2c^2)$ เมื่อ P เท่ากับ 0.1 และค่า c เท่ากับ 2 และ 5 ตามลำดับ

$(1-P)N(2,6)+PN(2,6c^2)$ เมื่อ P เท่ากับ 0.1 และค่า c เท่ากับ 2 และ 5 ตามลำดับ

1.3 กำหนดขนาดตัวอย่างในการจำลองข้อมูลแต่ละรูปแบบ เท่ากับ 10, 20, 30 และ 50

2. นำข้อมูลที่ได้จากการจำลองข้อมูลในขั้นตอนที่ 1 หาค่าขีดจำกัดล่าง (Lower Limit : L) และขีดจำกัดบน (Upper Limit : U) เพื่อสร้างช่วงความเชื่อมั่นของความแปรปรวน 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่าง วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน โดยกำหนดช่วงความเชื่อมั่น เท่ากับ 95% และ 99% วิธีการประมาณช่วงความเชื่อมั่นของความแปรปรวนทั้ง 3 วิธี มีรายละเอียดดังนี้

2.1 วิธีที่ใช้ความแปรปรวนของตัวอย่าง (Sample Variance)

กำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีขนาดตัวอย่าง n และ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ เป็นค่าความแปรปรวนของตัวอย่างที่ใช้เป็นตัวประมาณในการหาช่วงความเชื่อมั่นของความแปรปรวนเนื่องจาก $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ โดยที่ $E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1$ และ $Var\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1)$ [1]

จะได้ว่า

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

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

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

นั่นคือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ^2 คือ

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

(1)

เมื่อ ขีดจำกัดล่าง (Lower Limit : L) เท่ากับ $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}$

และ ขีดจำกัดบน (Upper Limit : U) เท่ากับ $\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$

2.2 วิธีแจ็คไนฟ์ (Jackknife)

การประมาณค่าช่วงความเชื่อมั่นด้วยวีธีแจ็คไนฟ์ มีวิธีการดังนี้ [8]

2.2.1 กำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีขนาดตัวอย่าง n โดยจะเรียกข้อมูลชุดนี้ว่า ข้อมูลดั้งเดิม (Original data)

2.2.2 นำข้อมูลที่ได้มาจากขั้นตอนที่ 2.2.1 มาสร้างข้อมูลชุดใหม่ n ครั้ง ดังนี้

ครั้งที่ 1 สร้างตัวอย่างชุดใหม่ โดยตัดข้อมูล X_1 ออกจากข้อมูลดั้งเดิม แล้วคำนวณด้วยข้อมูลตัวที่เหลือในการประมาณค่าของ σ^2 งานวิจัยนี้กำหนดให้เป็น $S_{(1)}^2$

ครั้งที่ 2 สร้างตัวอย่างชุดใหม่ โดยตัดข้อมูล X_2 ออกจากข้อมูลดั้งเดิม แล้วคำนวณด้วยข้อมูลตัวที่เหลือในการประมาณค่าของ σ^2 งานวิจัยนี้กำหนดให้เป็น $S_{(2)}^2$ ทำไปเรื่อย ๆ จนถึงครั้งที่ n

ครั้งที่ n สร้างตัวอย่างชุดใหม่ โดยตัดข้อมูล X_n ออกจากข้อมูลดั้งเดิม แล้วคำนวณด้วยข้อมูลตัวที่เหลือในการประมาณค่าของ σ^2 งานวิจัยนี้กำหนดให้เป็น $S_{(n)}^2$ จะได้ $S_{(1)}^2, S_{(2)}^2, ..., S_{(n)}^2$ จำนวน n ค่า โดยนำค่าเหล่านี้มาประมาณช่วงความเชื่อมั่นของความแปรปรวน

2.2.3 ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ^2 ด้วยวีธีด้วยวิธีแจ็คไนฟ์ คือ

$$P\left(\hat{\sigma}_{Jack}^2 - t_{(\frac{\alpha}{2},n-1)} se(\hat{\sigma}_{Jack}^2) < \sigma^2 < \hat{\sigma}_{Jack}^2 + t_{(\frac{\alpha}{2},n-1)} se(\hat{\sigma}_{Jack}^2)\right) = 1-\alpha$$

(2)

เมื่อ ขีดจำกัดล่าง (Lower Limit : L) เท่ากับ $\hat{\sigma}_{Jack}^2 - t_{(\frac{\alpha}{2},n-1)} se(\hat{\sigma}_{Jack}^2)$

และ ขีดจำกัดบน (Upper Limit : U) เท่ากับ $\hat{\sigma}_{Jack}^2 + t_{(\frac{\alpha}{2},n-1)} se(\hat{\sigma}_{Jack}^2)$

โดยที่ $\hat{\sigma}_{Jack}^2$ คือ ตัวประมาณความแปรปรวนรูปแบบจุดของวิธีแจ็คไนฟ์ ซึ่ง $\hat{\sigma}_{Jack}^2 = \frac{\sum_{i=1}^{n} S_{(i)}^2}{n}$

และ $$se(\hat{\sigma}_{Jack}^2) = \sqrt{\frac{\left[\sum_{i=1}^{n}\left(S_{(i)}^2 - \hat{\sigma}_{Jack}^2\right)^2\right]}{n(n-1)}}$$

2.3 วิธีบูตสแตรปมาตรฐาน (Standard Bootstrap) [10]

การประมาณค่าช่วงความเชื่อมั่นด้วยวีธีบูตสแตรปมาตรฐานมีวิธีการดังนี้

2.3.1 กำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีขนาดตัวอย่าง n โดยจะเรียกข้อมูลชุดนี้ว่า ข้อมูลดั้งเดิม (Original data)

2.3.2 นำข้อมูลที่ได้มาจากขั้นตอนที่ 2.3.1 มาสร้างข้อมูลชุดใหม่ B ครั้งโดย

ครั้งที่ 1 สุ่มตัวอย่างซ้ำจากข้อมูลดั้งเดิมโดยสุ่มตัวอย่างแบบคืนที่จำนวน n ค่า จะได้ตัวอย่างบูตสแตรป $X_1^*, X_2^*, ..., X_n^*$ จากนั้นนำตัวอย่างบูตสแตรปมาคำนวณค่าประมาณของ σ^2 งานวิจัยนี้กำหนดให้เป็น S_1^{2*}

ครั้งที่ 2 สุ่มตัวอย่างซ้ำจากข้อมูลดั้งเดิมโดยสุ่มตัวอย่างแบบคืนที่จำนวน n ค่า จะได้ตัวอย่างบูตสแตรป $X_1^*, X_2^*, ..., X_n^*$ จากนั้นนำตัวอย่างบูตสแตรปมาคำนวณค่าประมาณของ σ^2 งานวิจัยนี้กำหนดให้เป็น S_2^{2*} ทำไปเรื่อย ๆ จนถึงครั้งที่ B

ครั้งที่ B สุ่มตัวอย่างซ้ำจากข้อมูลดั้งเดิมโดยสุ่มตัวอย่างแบบคืนที่จำนวน n ค่าจะได้ จะได้ตัวอย่างบูตสแตรป $X_1^*, X_2^*, \ldots, X_n^*$ จากนั้นนำตัวอย่างบูตสแตรปมาคำนวณค่าประมาณของ σ^2 งานวิจัยนี้กำหนดให้เป็น S_B^{2*} จะได้ $S_1^{2*}, S_2^{2*}, \ldots, S_B^{2*}$ จำนวน B ค่า โดยปกติแล้วมักกำหนดขนาดตัวอย่างของบูตสแตรปตามผู้ใช้งาน ดังนั้นในงานวิจัยนี้จึงกำหนดขนาดตัวอย่างของบูตสแตรป B เท่ากับ 10,000 ค่า

2.3.3 เนื่องจากหากกำหนดขนาดตัวอย่างของบูตสแตรป B มากพอทำให้ $\hat{\sigma}^2_{Boot}$ มีลักษณะไกล้เคียงการแจกแจงปกติ จะได้ว่า ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ^2 ด้วยวิธีบูตสแตรปมาตรฐาน คือ

$$P\left(\hat{\sigma}^2_{Boot} - z_{\frac{\alpha}{2}} se(\hat{\sigma}^2_{Boot}) < \sigma^2 < \hat{\sigma}^2_{Boot} + z_{\frac{\alpha}{2}} se(\hat{\sigma}^2_{Boot})\right) = 1-\alpha$$

(3)

เมื่อ ขีดจำกัดล่าง (Lower Limit : L) เท่ากับ $\hat{\sigma}^2_{Boot} - z_{\frac{\alpha}{2}} se(\hat{\sigma}^2_{Boot})$

และ ขีดจำกัดบน (Upper Limit : U) เท่ากับ $\hat{\sigma}^2_{Boot} + z_{\frac{\alpha}{2}} se(\hat{\sigma}^2_{Boot})$

โดยที่ $\hat{\sigma}^2_{Boot}$ คือ ตัวประมาณความแปรปรวนรูปแบบจุดของวิธีบูตสแตรป ซึ่ง $\hat{\sigma}^2_{Boot} = \dfrac{\sum_{i=1}^{B} s_i^{2*}}{B}$

และ $se(\hat{\sigma}^2_{Boot}) = \sqrt{\dfrac{\left[\sum_{i=1}^{B}\left(s_i^{2*} - \hat{\sigma}^2_{Boot}\right)^2\right]}{B-1}}$

3. ทำขั้นตอนที่ 1 และ ขั้นตอนที่ 2 ซ้ำจำนวน 1,000 รอบ/สถานการณ์

4. เปรียบเทียบประสิทธิภาพในการประมาณช่วงความเชื่อมั่นของค่าความแปรปรวน 3 วิธีข้างต้น ดังนี้

4.1 คำนวณค่าประมาณสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient: CC)

โดยนำช่วงความเชื่อมั่นที่ได้จากการประมาณค่าความแปรปรวนของทั้ง 3 วิธีมาพิจารณาว่าครอบคลุมพารามิเตอร์ σ^2 หรือไม่ ถ้าช่วงความเชื่อมั่นใดครอบคลุมพารามิเตอร์ σ^2 จะทำการนับจำนวนครั้ง หลังจากนั้นนำจำนวนครั้งทั้งหมดหารด้วยจำนวนรอบ (ในงานวิจัยนี้จำนวนรอบมีค่าเท่ากับ 1,000 รอบ)

4.2 เปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

พิจารณาค่าประมาณสัมประสิทธิ์ความเชื่อมั่น (CC) ที่ได้ในข้อ 4.1 ว่าอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่ โดยช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดคำนวณจาก

$$P_0 - Z_{\frac{\alpha}{2}}\sqrt{\frac{P_0(1-P_0)}{1,000}} \le \hat{P} \le P_0 + Z_{\frac{\alpha}{2}}\sqrt{\frac{P_0(1-P_0)}{1,000}}$$

(4)

เมื่อ P_0 คือ ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

ซึ่งงานวิจัยนี้กำหนดระดับนัยสำคัญสำหรับการทดสอบที่ $\alpha = 0.05$ จะได้ $Z_{\frac{\alpha}{2}} = 1.96$

และในงานวิจัยนี้กำหนด P_0 เท่ากับ 0.95 และ 0.99 หรือที่ระดับช่วงความเชื่อมั่น 95% และ 99% ตามลำดับ ในการคำนวณค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดแสดงได้ในตารางที่ 1

ตารางที่ 1 เกณฑ์การตรวจสอบค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดที่ระดับช่วงความเชื่อมั่น 95% และ 99%

ช่วงความเชื่อมั่น	P_0	ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด
95%	0.95	$0.93650 \le \hat{P} \le 0.96350$
99%	0.99	$0.98380 \le \hat{P} \le 0.99620$

4.3 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (Average Width)

พิจารณาเฉพาะวิธีที่ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น (จากหัวข้อ 4.1) ที่อยู่ในช่วงค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (จากหัวข้อ 4.2) แล้วจึงนำช่วงความเชื่อมั่นจากวิธีดังกล่าวมาหาค่าความกว้างเฉลี่ยได้จากสูตร

$$AW_{\hat{\sigma}^2} = \frac{\sum_{i=1}^{1,000} (\mathrm{U}_{\hat{\sigma}_i^2} - L_{\hat{\sigma}_i^2})}{1,000}$$

(5)

เมื่อ $L_{\hat{\sigma}_i^2}$ คือ ขอบเขตล่างของช่วงความเชื่อมั่นที่ได้จากการประมาณค่าพารามิเตอร์ σ^2

ครั้งที่ i โดยที่ $i = 1, 2, ..., 1,000$

$U_{\hat{\sigma}_i^2}$ คือ ขอบเขตบนของช่วงความเชื่อมั่นที่ได้จากการประมาณค่าพารามิเตอร์ σ^2

ครั้งที่ i โดยที่ $i = 1, 2, ..., 1,000$

หากวิธีใดให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด จะถือว่าเป็นวิธีที่ใช้ในการประมาณช่วงความเชื่อมั่นของความแปรปรวนได้เหมาะสมที่สุด

ผลการวิจัย

ผลการวิจัยที่ได้จากการเปรียบเทียบประสิทธิภาพการประมาณช่วงความเชื่อมั่น 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่าง วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน โดยวัดประสิทธิภาพการประมาณช่วงความเชื่อมั่นทั้ง 3 วิธีได้จาก ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient) และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (Average Width) โดยผลจากการวิจัยสำหรับการแจกแจงปกตินำเสนอในตารางที่ 2-3 และผลจากการวิจัยสำหรับการแจกแจงปกติปลอมปนนำเสนอในตารางที่ 4-5 โดยใช้สัญลักษณ์แทนความหมายดังนี้

กำหนดให้

σ^2 แทน ความแปรปรวนของประชากร

n แทน ขนาดตัวอย่าง

C แทน สเกลแฟกเตอร์

Sample Variance แทน วิธีความแปรปรวนของตัวอย่าง

Jackknife แทน วิธีแจ็คไนฟ์

Standard Bootstrap แทน วิธีบูตสแตรปมาตรฐาน

ตารางที่ 2 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนที่ได้จากการประมาณค่าแต่ละวิธี เมื่อกำหนดลักษณะการแจกแจงปกติ $X \sim N(2,\sigma^2)$ ที่ช่วงความเชื่อมั่น 95%

σ^2	n	ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น			ค่าความกว้างเฉลี่ย		
		Sample Variance	Jack knife	Standard Bootstrap	Sample Variance	Jack knife	Standard Bootstrap
2	10	**0.9470**	0.8640	0.7700	5.7230*	-	-
	20	**0.9380**	0.8820	0.8440	3.1099*	-	-
	30	**0.9390**	0.9100	0.8780	2.3629*	-	-
	50	**0.9500**	0.9290	0.9060	1.7176*	-	-
6	10	**0.9470**	0.8640	0.7700	17.1691*	-	-
	20	**0.9380**	0.8820	0.8440	9.3297*	-	-
	30	**0.9390**	0.9100	0.8780	7.0886*	-	-
	50	**0.9500**	0.9290	0.9060	5.1527*	-	-

หมายเหตุ: ตัวเลขตัวหนา หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นมีค่าอยู่ระหว่าง 0.9365 และ 0.9635)

* หมายถึง วิธีที่ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดโดยพิจารณาเฉพาะวิธีที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

ตารางที่ 3 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนที่ได้จากการประมาณค่าแต่ละวิธี เมื่อกำหนด ลักษณะการแจกแจงปกติ $X \sim N(2,\sigma^2)$ ที่ช่วงความเชื่อมั่น 99%

σ^2	n	ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น			ค่าความกว้างเฉลี่ย		
		Sample Variance	Jack knife	Standard Bootstrap	Sample Variance	Jack knife	Standard Bootstrap
2	10	**0.9880**	0.9210	0.8320	9.6180*	-	-
	20	**0.9860**	0.9340	0.8880	4.5675*	-	-
	30	**0.9890**	0.9560	0.9290	3.3361*	-	-
	50	**0.9940**	0.9670	0.9490	2.3539*	-	-
6	10	**0.9880**	0.9210	0.8320	28.8540*	-	-
	20	**0.9860**	0.9340	0.8880	13.7024*	-	-
	30	**0.9890**	0.9560	0.9290	10.0084*	-	-
	50	**0.9940**	0.9670	0.9490	7.0617*	-	-

หมายเหตุ: ตัวเลขตัวหนา หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นมีค่าอยู่ระหว่าง 0.9838 และ 0.9962)

* หมายถึง วิธีที่ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดโดยพิจารณาเฉพาะวิธีที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

พิจารณาผลการวิจัยจากตารางที่ 2 และ 3 พบว่าเมื่อข้อมูลมีการแจกแจงปกติวิธีที่ใช้ความแปรปรวนของตัวอย่างให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดทุกกรณี ส่วนวิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐานให้ค่าให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ได้อยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดทุกกรณี ดังนั้นหากพิจารณาค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนเฉพาะวิธีที่ใช้ความแปรปรวนของตัวอย่าง พบว่าค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบลงเมื่อขนาดตัวอย่างเพิ่มขึ้น (n มาก) หากพิจารณาลักษณะการกระจายของการแจกแจง (σ^2) เมื่อขนาดตัวอย่างคงที่ (n คงที่) พบว่า หากการกระจายของข้อมูลน้อยลง (σ^2 น้อย) ทำให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบลง และเมื่อช่วงความเชื่อมั่นเพิ่มขึ้นค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นจะกว้างขึ้น

ตารางที่ 4 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนที่ได้จากการประมาณค่าแต่ละวิธี เมื่อกำหนดลักษณะการแจกแจงปกติปลอมปนรูปแบบ $f(X)=(0.9)N(2,\sigma^2)+(0.1)N(2,\mathrm{c}^2\sigma^2)$ ที่ช่วงความเชื่อมั่น 95%

σ^2	n	c	ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น			ค่าความกว้างเฉลี่ย		
			Sample Variance	Jack knife	Standard Bootstrap	Sample Variance	Jack knife	Standard Bootstrap
2	10	2	0.8720	0.9210	0.8380	-	-	-
		5	0.5860	**0.9480**	0.8910	-	21.0599*	-
	20	2	0.8320	**0.9520**	0.9170	-	3.8380*	-
		5	0.3710	0.9690	**0.9460**	-	-	13.7010*
	30	2	0.7780	**0.9580**	**0.9370**	-	3.1508	2.8120*
		5	0.2460	0.9670	**0.9400**	-	-	12.9626*
	50	2	0.6840	**0.9370**	0.9230	-	2.5368*	-
		5	0.1090	0.8620	0.8150	-	-	-
6	10	2	0.8720	0.9210	0.8380	-	-	-
		5	0.5860	**0.9480**	0.8910	-	63.1798*	-
	20	2	0.8320	**0.9520**	0.9170	-	11.5140*	-
		5	0.3710	0.9690	**0.9460**	-	-	41.1030*
	30	2	0.7780	**0.9580**	**0.9370**	-	9.4523	8.4360*
		5	0.2460	0.9670	**0.9400**	-	-	38.8877*
	50	2	0.6840	**0.9370**	0.9230	-	7.6103*	-
		5	0.1090	0.8620	0.8150	-	-	-

หมายเหตุ: ตัวเลขตัวหนา หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นมีค่าอยู่ระหว่าง 0.9365 และ 0.9635)

* หมายถึง วิธีที่ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดโดยพิจารณาเฉพาะวิธีที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

พิจารณาผลการวิจัยจากตารางที่ 4 พบว่า เมื่อข้อมูลมีการแจกแจงปกติปลอมปน ที่ช่วงความเชื่อมั่น 95% วิธีที่ใช้ความแปรปรวนของตัวอย่างให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่อยู่ในค่าสัมประสิทธ์ควมเชื่อมั่นที่กำหนดทุกกรณี ส่วนวิธีแจ็คไนฟ์ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 สเกลแฟกเตอร์เท่ากับ 2 เมื่อ σ^2 เท่ากับ 2 และ 6 และกรณีที่ขนาดตัวอย่างเท่ากับ 10 สเกลแฟกเตอร์เท่ากับ 5 เมื่อ σ^2 เท่ากับ 2 และ 6 ส่วนวิธีบูตสแตรปมาตรฐานให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 สเกลแฟกเตอร์ เท่ากับ 5 ทุกความแปรปรวนของประชากรที่ทำการศึกษา กรณีที่ขนาดตัวอย่างเท่ากับ 30 สเกลแฟกเตอร์เท่ากับ 2 เมื่อ σ^2 เท่ากับ 2 และ 6 ทั้งวิธีแจ็คไนฟ์และวิธีบูตสแตรปมาตรฐานให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จึงพิจารณาโดยเปรียบเทียบค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนพบว่า วิธีแจ็คไนฟ์และวิธีบูตสแตรปมาตรฐานมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเท่า ๆ กัน

ตารางที่ 5 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนที่ได้จากการประมาณค่าแต่ละวิธี เมื่อกำหนดลักษณะการแจกแจงปกติปลอมปนรูปแบบ $f(X)=(0.9)N(2,\sigma^2)+(0.1)N(2,c^2\sigma^2)$ ที่ช่วงความเชื่อมั่น 99%

σ^2	n	c	ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น			ค่าความกว้างเฉลี่ย		
			Sample Variance	Jack knife	Standard Bootstrap	Sample Variance	Jack knife	Standard Bootstrap
2	10	2	0.9370	0.9570	0.8940	-	-	-
		5	0.6530	0.9730	0.9310	-	-	-
	20	2	0.9100	0.9700	0.9570	-	-	-
		5	0.4300	**0.9850**	0.9750	-	22.5927*	-
	30	2	0.8820	**0.9870**	0.9700	-	4.2464*	-
		5	0.3130	0.9980	**0.9930**	-	-	17.0357*
	50	2	0.8040	**0.9920**	**0.9880**	-	3.3830	3.1143*
		5	0.1410	0.9980	**0.9900**	-		15.1547*
6	10	2	0.9370	0.9570	0.8940	-	-	-
		5	0.6530	0.9730	0.9310	-	-	-
	20	2	0.9100	0.9700	0.9570	-	-	-
		5	0.4300	**0.9850**	0.9750	-	67.7780*	-
	30	2	0.8820	**0.9870**	0.9700	-	12.7391*	-
		5	0.3130	0.9980	**0.9930**	-	-	51.1071*
	50	2	0.8040	**0.9920**	**0.9880**	-	10.1490	9.3430*
		5	0.141	0.9980	**0.9900**	-	-	45.4641*

หมายเหตุ: ตัวเลขตัวหนา หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นมีค่าอยู่ระหว่าง 0.9838 และ 0.9962)

* หมายถึง วิธีที่ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดโดยพิจารณาเฉพาะวิธีที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

พิจารณาผลการวิจัยจากตารางที่ 5 พบว่า เมื่อข้อมูลมีการแจกแจงปกติปลอมปนที่ช่วงความเชื่อมั่น 99% วิธีที่ใช้ความแปรปรวนของตัวอย่าง ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่อยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดทุกกรณี ส่วนวิธีแจ็คไนฟ์ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเมื่อขนาดตัวอย่างเท่ากับ 20 สเกลแฟกเตอร์ เท่ากับ 5 เมื่อ σ^2 เท่ากับ 2 และ 6 และกรณีที่ขนาดตัวอย่างเท่ากับ 30 และ 50 สเกลแฟกเตอร์ เท่ากับ 2 เมื่อ σ^2 เท่ากับ 2 และ 6 ส่วนวิธีบูตสแตรปมาตรฐานให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธ์ความเชื่อมั่นที่กำหนดเมื่อ ขนาดตัวอย่างเท่ากับ 30 และ 50 สเกลแฟกเตอร์เท่ากับ 5 ทุกความแปรปรวนของประชากรที่ทำการศึกษา ในกรณีที่ขนาดตัวอย่างเท่ากับ 50 สเกลแฟกเตอร์เท่ากับ 2 เมื่อ σ^2 เท่ากับ 2 และ 6 ทั้งวิธีแจ็คไนฟ์และวิธีบูตสแตรปมาตรฐานให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในช่วงของค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จึงพิจารณาโดยเปรียบเทียบค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของความแปรปรวนพบว่า วิธีบูตสแตรปมาตรฐานมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดมากกว่าวิธีแจ็คไนฟ์

สรุปและอภิปรายผล

การศึกษาการประมาณค่าช่วงความเชื่อมั่นของความแปรปรวนจากการเปรียบเทียบประสิทธิภาพการประมาณความแปรปรวนทั้ง 3 วิธี คือ วิธีที่ใช้ความแปรปรวนของตัวอย่าง วิธีแจ็คไนฟ์ และวิธีบูตสแตรปมาตรฐาน พบว่าวิธีที่ใช้ความแปรปรวนของตัวอย่างให้ผลเป็นไปตามทฤษฎีสำหรับการประมาณค่าแบบช่วงของความแปรปรวนของประชากรที่มีการแจกแจงปกติ แต่อย่างไรก็ตาม วิธีที่กล่าวมาข้างต้นไม่เหมาะสมในการประมาณค่าความแปรปรวนในกรณีที่ข้อมูลมีค่านอกเกณฑ์เกิดขึ้น ดังนั้น วิธีแจ็คไนฟ์และวิธีบูตสแตรปมาตรฐานซึ่งทั้ง 2 วิธี ใช้เทคนิคการสุ่มตัวอย่างซ้ำซึ่งเป็นวิธีการทางสถิติที่ไม่ใช้พารามิเตอร์มีความเหมาะสมในการประมาณค่าความแปรปรวนของประชากรมากกว่าวิธีที่ความแปรปรวนของตัวอย่างซึ่งคนส่วนใหญ่มักจะนิยมใช้ นอกจากนี้ ขนาดตัวอย่าง เป็นปัจจัยที่สำคัญสำหรับการประมาณค่าความแปรปรวน ซึ่งหากมีขนาดตัวอย่างสำหรับการวิเคราะห์มากขึ้นทำให้การประมาณค่าความแปรปรวนมีประสิทธิภาพยิ่งขึ้น แต่เมื่อพิจารณาในการประมวลผลของโปรแกรมวิธีบูตสแตรปจะใช้เวลามากกว่าวิธีอื่นเพราะมีจำนวนในการสุ่มซ้ำมากกว่าวิธีอื่น ๆ

เอกสารอ้างอิง

[1] Prachoom Suwattee. (2010). *Theory of Statistical Inference.* 3th ed. Bangkok: WVO officer of Printing Mill.

[2] John R. Taylor. (2013). *An Introduction to Error Analysis the Study of Uncertainties in Physical Measurements.* (Jirapong Kasivitamnuay, Translator). Bangkok: Chulalongkorn University Printing House.

[3] Quenouille, M.H. (1956, December). Notes on Bias in Estimation. *Biometrika, 43*, 353-360.

[4] Efron, B. (1979, January). Bootstrap Methods: Another Loot at the Jackknife. *The Annals of statistics, 7*, 1-26.

[5] Sornsawan Boonpen, Boonorm Chomtee; and Apinya Hirunwong. (2015, October-December). A Comparison of Intervals Estimation Methods for Scale Parameter of the Two-Parameter Weibull Distribution. *Thai Science and Technology Journal Thammasat University, 23*(4), 579-587.

[6] Bumrungsak Phuenaree, and Kornrawee Rungsawang. (2017). An Estimation of Exponential Distribution Median by Bootstrap Methods. In *9th Science Research Conference*. MA 187-194. Chonburi: Burapha University.

[7] Anuwat Khamma, Manachai Rodchuen; and Putipong Bookkamana. (2017, October-December). Confidence Interval Estimation for the Ratio of Two Population Variances with Non-normal Distributions. *KKU Research Journal (Graduate Studies)*, *17*(4), 12-23.

[8] Bonett, D.G. (2006) Confidence Interval for a Ratio of Variance in Bivariate Nonnormal Distribution, *Journal of Statistical Computation and Simulation*, *76*(7), 637-644.

[9] Sangnawakij, P., Niwitpong, S. (2010) Interval Estimation for the Common Coefficient of Variation of Gamma Distributions. *Thailand Statistician, 18*(3), 340-353.

[10] Efron, B; and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* New York: Chapman&Hall.