

ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบช่วง สำหรับข้อมูลแบบต่อเนื่อง ด้วยวิธี แจ็กไนฟ์และวิธีบูตสเตรป

An Efficiency of Interval Estimation for Continuous Variable by Jackknifing and Bootstrapping Method

ภัททิตา เลิศจริยพร¹ และณภัทน์จันทร์ ด้านสวัสดิ์² *

¹ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

² คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์

*ผู้เขียนหลัก (Corresponding Author) E-mail: napattchan@vru.ac.th

Received: October 5,2021

Revised: November 22,2021

Accepted: November 30,2021

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์แบบช่วง ได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง สำหรับข้อมูลแบบต่อเนื่อง ระหว่างวิธีแจ็กไนฟ์กับวิธีบูตสเตรป โดยใช้ข้อมูลจากการจำลองด้วยเทคนิคมอนติคาร์โล ด้วยโปรแกรม R เมื่อกำหนดให้ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบที่ การแจกแจงแบบเอ็กซ์โพเนนเชียล และการแจกแจงแบบแกมมา โดยกำหนดขนาดตัวอย่าง 100, 200, 300, 400 และ 500 ที่ระดับนัยสำคัญ 0.01 และ 0.05 ในแต่ละสถานการณ์จะมีการจำลองข้อมูลกระทำซ้ำจำนวน 1,000 ครั้ง เกณฑ์ที่ใช้ในการเปรียบเทียบการประมาณค่าแบบช่วง คือ สัมประสิทธิ์ความเชื่อมั่น ผลการศึกษา พบว่า ทุกค่าพารามิเตอร์ในทุก ๆ ลักษณะการแจกแจงวิธีบูตสเตรปจะมีค่าสัมประสิทธิ์ความเชื่อมั่นสูง และมีประสิทธิภาพดีกว่าวิธีแจ็กไนฟ์

คำสำคัญ: แจ็กไนฟ์, บูตสเตรป, การประมาณค่าพารามิเตอร์, ช่วงความเชื่อมั่น

Abstract

The purpose of this study was to compare the efficiency of interval estimation mean, variance, skewness and kurtosis of continuous variable between Jackknifing and

Bootstrapping method. Data are generated using Monte Carlo Simulation Technique by R-Program. Standard normal distribution, t-distribution, exponential distribution, and gamma distribution were generated at the sample of size 100, 200, 300, 400 and 500 at significance levels of 0.01 and 0.05. In each situation data was simulated and repeated 1,000 times. The efficiency was considered by confidence coefficient of the interval estimation. The results showed that Bootstrapping method gave the higher confidence coefficient than Jackknifing method in four parameters of each distribution. Therefore, Bootstrapping method was more efficient than Jackknifing method.

Keywords: Jackknifing, Bootstrapping, Parameter estimation, Confidence interval

บทนำ

โดยทั่วไปในการศึกษาข้อมูลจากทุกหน่วยในประชากร (Population) ที่สนใจศึกษานับเป็นเรื่องที่ทำได้ค่อนข้างยาก โดยเฉพาะเมื่อประชากรที่สนใจศึกษามีขนาดใหญ่ อาจเนื่องจากปัญหาหรือข้อจำกัดในด้านต่าง ๆ ดังนั้นผู้วิจัยหลายคนจึงนิยมทำการศึกษาข้อมูลของประชากรที่สนใจผ่านการศึกษาจากข้อมูลของตัวแทนที่ได้จากสุ่มหรือเลือกมาจากประชากรแทน ซึ่งในทางทฤษฎีจะเรียกว่าตัวอย่าง (Sample) โดยข้อมูลที่ได้จากตัวอย่างนั้นจะต้องมีคุณสมบัติที่ครอบคลุมถึงลักษณะของประชากรที่ต้องการศึกษาให้ได้มากที่สุด แล้วจะนำค่าที่คำนวณจากตัวอย่าง ซึ่งเรียกว่า ค่าสถิติ (Statistic) ไปสรุปหรืออ้างอิงถึงคุณลักษณะบางประการของประชากรที่ต้องการศึกษา ซึ่งเรียกว่า ค่าพารามิเตอร์ (Parameter) และจะเรียกวิธีการดังกล่าวว่า การอนุมานทางสถิติ (Inferential Statistics) การประมาณค่า (Estimation) เป็นวิธีการหนึ่งในการอนุมานเชิงสถิติ เป็นการนำค่าสถิติที่คำนวณได้จากข้อมูลตัวอย่างไปประมาณค่าพารามิเตอร์ที่เป็นคุณลักษณะที่แท้จริงของประชากร เช่น การนำค่าเฉลี่ยของตัวอย่าง ค่าสัดส่วนของตัวอย่าง และค่าความแปรปรวนของตัวอย่าง ไปสรุปหรืออนุมานถึงค่าเฉลี่ยของประชากร ค่าสัดส่วนของประชากร และค่าความแปรปรวนของประชากร ตามลำดับ ซึ่งการประมาณค่าพารามิเตอร์นี้สามารถแบ่งออกได้เป็น 2 ประเภท คือ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval estimation) การประมาณค่าแบบจุด เป็นการประมาณค่าพารามิเตอร์ที่ใช้ตัวประมาณเป็นฟังก์ชันของตัวแปรสุ่ม ค่าประมาณที่ได้จึงมีเพียงค่าเดียว จึงมีโอกาสน้อยมากที่ค่าประมาณที่ได้จะมีค่าที่เท่ากับค่าพารามิเตอร์ของประชากร ดังนั้นผู้วิจัยจะการใช้การประมาณค่าแบบช่วงมากกว่า และคาดหวังว่าช่วงความเชื่อมั่นที่กำหนดจะสามารถครอบคลุมค่าพารามิเตอร์ได้ ซึ่งการประมาณค่าแบบช่วงจะครอบคลุมค่าพารามิเตอร์มากกว่าการประมาณค่าแบบจุด เนื่องจากจะใช้ค่าประมาณแบบจุดและการแจก

แรงความน่าจะเป็นของตัวสถิติในการสร้างช่วงของตัวเลขช่วงหนึ่ง ซึ่งช่วงของตัวเลขนั้นจะแคบหรือกว้าง จะขึ้นอยู่กับระดับความเชื่อมั่นและลักษณะของประชากรที่สนใจศึกษา

ในกรณีที่ผู้วิจัยอยากทราบว่าตัวอย่างที่ถูกสุ่มมา มาจากประชากรที่มีการแจกแจงแบบใด ผู้วิจัยสามารถนำข้อมูลจากตัวอย่างดังกล่าวไปประมาณการแจกแจงของประชากรผ่านการศึกษาจากการประมาณค่าพารามิเตอร์ต่าง ๆ เช่น ค่าเฉลี่ย ความแปรปรวน ค่าความเบ้ และค่าความโด่ง ได้ แต่ในกรณีที่ ไม่ทราบการแจกแจงของประชากรมาก่อน หรือข้อมูลตัวอย่างมีจำนวนน้อย หรือตัวสถิติที่ใช้ประมาณค่าพารามิเตอร์อยู่ในรูปที่ยุ่งยาก การประมาณค่าดังกล่าว ไม่อาจจะหาได้ด้วยวิธีง่าย ๆ จึงมีผู้เสนอให้ใช้วิธีการประมาณค่าพารามิเตอร์ ด้วยวิธีนอนพารามेटริก (Nonparametric) โดยหนึ่งในวิธีนอนพารามेटริกที่นิยมได้แก่ วิธีแจ๊คไนฟ์ ซึ่งนำเสนอโดย Quenouille (1956) และวิธีบูตสเตรป ซึ่งนำเสนอโดย Efron (1979) ซึ่งวิธีการดังกล่าว เป็นวิธีที่ไม่มีข้อจำกัดเกี่ยวกับลักษณะการแจกแจงของข้อมูล โดยจะอาศัยหลักการสุ่มซ้ำ (Resampling) จากตัวอย่างเพียงชุดเดียว เพื่อสร้างความผันแปรให้กับค่าของตัวสถิติ โดยจะถือเสมือนว่า ตัวอย่างที่ได้จากการสุ่มซ้ำจากตัวอย่างชุดเดียวนี้ เป็นตัวอย่างสุ่มซ้ำ ๆ จากประชากรจริง ๆ แต่อย่างไรก็ตามทั้งวิธีแจ๊คไนฟ์และวิธีบูตสเตรปมีข้อจำกัดคือ อาจทำให้ค่าบางค่าไม่อาจมีโอกาสที่จะถูกสุ่มตัวอย่างซ้ำได้เลย เนื่องจากการสุ่มตัวอย่างซ้ำที่ขึ้นอยู่กับตัวอย่างเพียงชุดเดียวที่สุ่มได้จากประชากร

ในการวิจัยครั้งนี้ผู้วิจัยจึงมีความสนใจเปรียบเทียบประสิทธิภาพระหว่างวิธีแจ๊คไนฟ์และวิธีบูตสเตรป ในการประมาณค่าพารามิเตอร์แบบช่วง สำหรับข้อมูลที่ต่อเนื่อง โดยจะประมาณพารามิเตอร์จำนวน 4 ค่า ได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์แบบช่วง ได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง สำหรับข้อมูลแบบต่อเนื่อง ระหว่างวิธีแจ๊คไนฟ์กับวิธีบูตสเตรป เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบที การแจกแจงแบบเอ็กซ์โพเนนเชียล และการแจกแจงแบบแกมมา

วิธีดำเนินการวิจัย

ผู้วิจัยได้กำหนดขอบเขตของการวิจัยได้ดังนี้

1. สร้างข้อมูลที่มีการแจกแจงตามลักษณะที่ต้องการศึกษา สำหรับการประมาณค่าพารามิเตอร์แบบช่วง ภายใต้ขอบเขตที่ต้องการศึกษาดังนี้

1.1 การแจกแจงแบบปกติมาตรฐาน ที่มีค่าเฉลี่ย และความแปรปรวน (μ, σ^2) 0 และ 1 ตามลำดับ

1.2 การแจกแจงแบบท ที่ มีจำนวนองศาอิสระ (Degree of freedom: df) เท่ากับ 5

1.3 การแจกแจงแบบเอ็กซ์โพเนนเชียล ที่มีค่าพารามิเตอร์ (θ) เท่ากับ 1

1.4 การแจกแจงแบบแกมมา ที่มีค่าพารามิเตอร์ (β) เท่ากับ 0.1

2. กำหนดระดับนัยสำคัญ 0.01 และ 0.05

3. กำหนดขนาดตัวอย่างที่ต้องการศึกษา (n) 100, 200, 300, 400 และ 500

4. กำหนดการทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ (M)

5. เปรียบเทียบประสิทธิภาพของการประมาณค่า พิจารณาจากค่าสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient) เพื่อหาวิธีการประมาณที่เหมาะสมที่สุดในแต่ละสถานการณ์ กล่าวคือ วิธีการประมาณค่าแบบใดที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงจะถือว่าการประมาณค่าจากวิธีนั้นเป็นวิธีที่เหมาะสมกว่าอีกวิธีหนึ่ง (Puttachaiyatad, 2009)

ขั้นตอนการจำลองข้อมูลและการประมาณค่า

1. สร้างข้อมูลที่มีลักษณะการแจกแจง 4 ชนิดที่ต้องการศึกษา

2. ทำการหาค่าประมาณของพารามิเตอร์ที่ต้องการศึกษา โดยใช้วิธีแจ๊คไนฟ์ และวิธีบูตสเตรป

3. จากนั้นนำค่าประมาณของพารามิเตอร์ที่ได้มาจัดเรียงจากค่าน้อยไปหาค่ามาก แล้วคำนวณหาค่าขอบล่างของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ที่ตรงกับตำแหน่งเปอร์เซ็นต์ไทล์ที่ $100(\alpha/2)$ และหาค่าขอบบนของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ที่ตรงกับตำแหน่งเปอร์เซ็นต์ไทล์ที่ $100(1-\alpha/2)$ โดยการทำการทดลองซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ ทั้งนี้ α คือค่าระดับนัยสำคัญที่กำหนด ซึ่งในการวิจัยครั้งนี้ได้กำหนดให้มีค่าเท่ากับ 0.01 และ 0.05

4. นับจำนวนที่ค่าประมาณของพารามิเตอร์ตกอยู่ในช่วงความเชื่อมั่นที่คำนวณได้จากข้างต้น เพื่อนำมาคำนวณค่าสัมประสิทธิ์ความเชื่อมั่น

5. นำค่าสัมประสิทธิ์ความเชื่อมั่น ที่ได้จากวิธีแจ๊คไนฟ์ และวิธีบูตสเตรป มาเปรียบเทียบภายใต้สถานการณ์เดียวกัน โดยที่ค่าสัมประสิทธิ์ความเชื่อมั่นของวิธีการประมาณ เป็นค่าที่ใช้วัดประสิทธิภาพของวิธีการประมาณแบบช่วง ซึ่งค่านี้หาได้จากการตรวจสอบว่าช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธีการประมาณค่า ครอบคลุมค่าพารามิเตอร์ที่ต้องการประมาณหรือไม่ หากช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณใดครอบคลุมค่าพารามิเตอร์ที่ต้องการประมาณ จะทำการนับจำนวนครั้งและบวกสะสมไว้ โดยแต่ละสถานการณ์จะคำนวณช่วงความเชื่อมั่นซ้ำกัน M ครั้ง ผลบวกสะสมที่ได้คือจำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์ที่ต้องการประมาณ ซึ่งคำนวณได้ดังนี้

$$\text{ค่าสัมประสิทธิ์ความเชื่อมั่น} = \frac{\text{จำนวนครั้งที่ทั้งหมดในช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์ที่ต้องการประมาณ}}{\text{จำนวนรอบที่ทำซ้ำในแต่ละสถานการณ์ (M)}}$$

(1)

โดย M คือ จำนวนรอบที่กระทำซ้ำในแต่ละสถานการณ์ในงานวิจัยครั้งนี้ กำหนดให้ M มีค่าเท่ากับ 1,000

ถ้าวิธีการประมาณค่าวิธีใดที่ทำให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าอีกวิธี จะถือว่าการประมาณค่าจากวิธีนั้นเป็นวิธีที่เหมาะสมและมีประสิทธิภาพมากกว่าอีกวิธีในการประมาณค่าแบบช่วง

6. สรุปผลการวิจัย

การประมาณค่าพารามิเตอร์

1. การประมาณค่าพารามิเตอร์ด้วยวิธีแจ๊คไนฟ์ (Jackknifing Method)

วิธีแจ๊คไนฟ์ ถูกนำเสนอครั้งแรกโดย Quenouille (1956) ต่อมาได้ถูกพัฒนาโดย Miller (1974) เพื่อใช้ลดความเอนเอียงของตัวประมาณ และใช้ในการหาค่าประมาณของตัวพารามิเตอร์ รวมถึงค่าความแปรปรวนของตัวประมาณต่าง ๆ จัดเป็นวิธีที่ใช้ในการเลือกตัวอย่างซ้ำวิธีหนึ่ง ซึ่งเป็นการเลือกตัวอย่างใหม่จากตัวอย่างสุ่มเพียงชุดเดียว มีวิธีดำเนินการดังนี้

สุ่มตัวอย่างมา n จำนวน คือ $x_1, x_2, \dots, x_n$ จากประชากรที่มีการแจกแจงแบบต่าง ๆ ให้ $\theta = \{\mu, \sigma^2, v_1, v_2\}$ เป็นพารามิเตอร์ที่ต้องการประมาณ และให้ $\theta_j = \{\mu_j, \sigma_j^2, v_{1j}, v_{2j}\}$ เป็นค่าประมาณของพารามิเตอร์ด้วยวิธีแจ๊คไนฟ์ที่คำนวณจากข้อมูลตัวอย่างขนาด n ในการหาค่าประมาณด้วยวิธีแจ๊คไนฟ์จะเริ่มจาก

การสุ่มตัวอย่างครั้งที่ 1 จะตัดค่า x_1 ออกจากตัวอย่าง แล้วคำนวณค่าประมาณจาก

$$x_2, x_3, \dots, x_n \text{ จะได้ค่าประมาณคือ } \mu_{j(1)}, \sigma_{j(1)}^2, v_{1j(1)}, v_{2j(1)}$$

การสุ่มตัวอย่างครั้งที่ 2 จะตัดค่า x_2 ออกจากตัวอย่าง แล้วคำนวณค่าประมาณจาก

$$x_1, x_3, \dots, x_n \text{ จะได้ค่าประมาณคือ } \mu_{j(2)}, \sigma_{j(2)}^2, v_{1j(2)}, v_{2j(2)}$$

⋮

การสุ่มตัวอย่างครั้งที่ n จะตัดค่า x_n ออกจากตัวอย่าง แล้วคำนวณค่าประมาณจาก

$$x_1, x_2, \dots, x_{n-1} \text{ จะได้ค่าประมาณคือ } \mu_{j(n)}, \sigma_{j(n)}^2, v_{1j(n)}, v_{2j(n)}$$

และจากการทำซ้ำ จำนวน n ครั้ง ดังที่กล่าวมาข้างต้น จะได้

ค่าประมาณของพารามิเตอร์ค่าเฉลี่ย จำนวน n ตัว คือ $\mu_{j(1)}, \mu_{j(2)}, \dots, \mu_{j(n)}$ และช่วงความเชื่อมั่นของ μ คือ $[\mu_{jL}, \mu_{jU}]$

ค่าประมาณของพารามิเตอร์ค่าความแปรปรวน จำนวน n ตัว คือ $\sigma_{j(1)}^2, \sigma_{j(2)}^2, \dots, \sigma_{j(n)}^2$ และช่วงความเชื่อมั่นของ σ^2 คือ $[\sigma_{jL}^2, \sigma_{jU}^2]$

ค่าประมาณของพารามิเตอร์ค่าความเบ้ จำนวน n ตัว คือ $v_{1j(1)}, v_{1j(2)}, \dots, v_{1j(n)}$ และช่วงความเชื่อมั่นของ v_1 คือ $[v_{1jL}, v_{1jU}]$

ค่าประมาณของพารามิเตอร์ค่าความโด่ง จำนวน n ตัว คือ $v_{2j(1)}, v_{2j(2)}, \dots, v_{2j(n)}$ และช่วงความเชื่อมั่นของ v_2 คือ $[v_{2jL}, v_{2jU}]$

2. การประมาณค่าพารามิเตอร์ด้วยวิธีบูตสเตรป (Bootstrapping Method)

วิธีบูตสเตรป ถูกนำเสนอครั้งแรกเสนอโดย Efron (1979) และได้พัฒนาต่อมาโดย Efron (1982) เป็นวิธีการประมาณค่าพารามิเตอร์ โดยใช้การสุ่มตัวอย่างซ้ำเช่นเดียวกับวิธีแจ๊คไนฟ์ แต่วิธีการสุ่มตัวอย่างซ้ำจะใช้การสร้างตัวอย่างชุดใหม่จากตัวอย่างสุ่มที่มีเพียงชุดเดียว โดยการสุ่มตัวอย่างแบบคืนที่ (Resampling with Replacement) มีวิธีดำเนินการดังนี้

สุ่มตัวอย่างมา n จำนวน คือ $x_1, x_2, \dots, x_n$ ที่เป็นอิสระกันมาจากการประชากรที่มีการแจกแจงแบบต่าง ๆ ให้ $\theta = \{\mu, \sigma^2, v_1, v_2\}$ เป็นพารามิเตอร์ที่ต้องการประมาณ และให้ $\theta_B = \{\mu_B, \sigma_B^2, v_{1B}, v_{2B}\}$ เป็นค่าประมาณของพารามิเตอร์ด้วยวิธีบูตสเตรปที่คำนวณจากข้อมูลตัวอย่างขนาด n ในการหาค่าประมาณด้วยวิธีบูตสเตรป จะเริ่มจาก

การสุ่มตัวอย่างครั้งที่ 1 จะทำการสุ่มตัวอย่างทีละ 1 ค่า แบบคืนที่จำนวน n ครั้ง จากชุดตัวอย่างจะได้ $x_1^*, x_2^*, \dots, x_n^*$ แล้วคำนวณหาค่าประมาณ จะได้ค่าประมาณของ θ คือ $\mu_{B1}^*, \sigma_{B1}^{2*}, v_{1B1}^*, v_{2B1}^*$

การสุ่มตัวอย่างครั้งที่ 2 ทำการสุ่มตัวอย่างทีละ 1 ค่า แบบคืนที่จำนวน n ครั้ง จากชุดตัวอย่างจะได้ $x_1^*, x_2^*, \dots, x_n^*$ แล้วคำนวณหาค่าประมาณ จะได้ค่าประมาณของ θ คือ $\mu_{B2}^*, \sigma_{B2}^{2*}, v_{1B2}^*, v_{2B2}^*$

⋮

การสุ่มตัวอย่างครั้งที่ B ทำการสุ่มตัวอย่างทีละ 1 ค่า แบบคืนที่จำนวน $1n$ ครั้ง จากชุดตัวอย่างจะได้ $x_1^*, x_2^*, \dots, x_n^*$ แล้วคำนวณหาค่าประมาณ จะได้ค่าประมาณของ θ คือ $\mu_{BB}^*, \sigma_{BB}^{2*}, v_{1BB}^*, v_{2BB}^*$

และจากการทำซ้ำ จำนวน B ครั้ง ดังที่กล่าวมาข้างต้น จะได้

ค่าประมาณของพารามิเตอร์ค่าเฉลี่ย จำนวน B ตัว คือ $\mu_{B1}^*, \mu_{B2}^*, \dots, \mu_{BB}^*$ และช่วงความเชื่อมั่นของ μ คือ $[\mu_{BL}, \mu_{BU}]$

ค่าประมาณของพารามิเตอร์ค่าความแปรปรวน จำนวน B ตัว คือ $\sigma_{B1}^{2*}, \sigma_{B2}^{2*}, \dots, \sigma_{BB}^{2*}$ และช่วงความเชื่อมั่นของ σ^2 คือ $[\sigma_{BL}^2, \sigma_{BU}^2]$

ค่าประมาณของพารามิเตอร์ค่าความเบ้ จำนวน B ตัว คือ $v_{1B1}^*, v_{1B2}^*, \dots, v_{1BB}^*$ และช่วงความ

เชื่อมั่นของ v_1 คือ $[v_{1BL}, v_{1BU}]$

ค่าประมาณของพารามิเตอร์ค่าความโด่ง จำนวน B ตัว คือ $v_{2B1}^*, v_{2B2}^*, \dots, v_{2BB}^*$ และช่วงความ

เชื่อมั่นของ v_2 คือ $[v_{2BL}, v_{2BU}]$

ผลและอภิปราย

จากการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบช่วงของข้อมูลที่มีการแจกแจงแบบต่อเนื่อง ด้วยวิธีแจ๊คไนฟ์และวิธีบูตสเตรป โดยประมาณพารามิเตอร์ค่าเฉลี่ย ค่าความแปรปรวน ความเบ้ และความโด่ง เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบท การแจกแจงแบบเอ็กซีโพเนนเชียล และการแจกแจงแบบแกมมา ได้ผลการวิจัยแสดงดังตารางที่ 1 และ 2 ดังนี้

ตารางที่ 1 การเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นระหว่างวิธีแจ๊คไนฟ์ (J) และวิธีบูตสเตรป (B) ที่ระดับนัยสำคัญ 0.01

พารามิเตอร์	n	การแจกแจงของข้อมูล							
		แบบปกติมาตรฐาน		แบบท		แบบเอ็กซีโพเนนเชียล		แบบแกมมา	
		J	B	J	B	J	B	J	B
ค่าเฉลี่ย	100	0.1747	0.9002	0.1576	0.8835	0.0961	0.8472	0.1055	0.8645
	200	0.1511	0.9104	0.1370	0.8944	0.0754	0.8581	0.0951	0.8723
	300	0.1018	0.9168	0.0974	0.9033	0.0652	0.8746	0.0744	0.8968
	400	0.0991	0.9250	0.0809	0.9128	0.0438	0.8872	0.0671	0.8979
	500	0.0954	0.9319	0.0723	0.9267	0.0291	0.8947	0.0598	0.9131
ความแปรปรวน	100	0.1291	0.8422	0.1049	0.8163	0.0823	0.7559	0.0702	0.7847
	200	0.0934	0.8547	0.7897	0.8317	0.0613	0.7763	0.0569	0.8198
	300	0.0906	0.8572	0.0612	0.8491	0.0487	0.7852	0.0421	0.8297
	400	0.0872	0.8619	0.0750	0.8563	0.0242	0.8144	0.0142	0.8355
	500	0.0795	0.8690	0.0648	0.8605	0.0149	0.8254	0.0094	0.8432

ตารางที่ 1 (ต่อ)

พารามิเตอร์	n	การแจกแจงของข้อมูล							
		แบบปกติมาตรฐาน		แบบที		แบบเอ็กซ์โพเนนเชียล		แบบแกมมา	
		J	B	J	B	J	B	J	B
ความเบ้	100	0.0876	0.7851	0.0724	0.7533	0.0887	0.7129	0.0791	0.7334
	200	0.0595	0.8020	0.0484	0.7773	0.0517	0.7315	0.0644	0.7532
	300	0.0325	0.8146	0.0161	0.7843	0.0270	0.7522	0.0421	0.7686
	400	0.0164	0.8242	0.0094	0.8115	0.0147	0.7593	0.0152	0.8015
	500	0.0149	0.8385	0.0087	0.8247	0.0108	0.7762	0.0132	0.8142
ความโด่ง	100	0.0943	0.7925	0.0842	0.6805	0.0875	0.6203	0.0961	0.6672
	200	0.0730	0.7830	0.0611	0.6304	0.0711	0.6086	0.0803	0.6235
	300	0.0671	0.6811	0.0529	0.5920	0.0657	0.5407	0.0833	0.5701
	400	0.0576	0.6632	0.0357	0.5734	0.0468	0.5294	0.0637	0.5322
	500	0.0420	0.6559	0.0204	0.5679	0.0355	0.5004	0.0508	0.5189

ตารางที่ 2 การเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นระหว่างวิธีแจ๊คไนฟ์ (J) และวิธีบูตสเตรป (B) ที่ระดับนัยสำคัญ 0.05

พารามิเตอร์	n	การแจกแจงของข้อมูล							
		แบบปกติมาตรฐาน		แบบที		แบบเอ็กซ์โพเนนเชียล		แบบแกมมา	
		J	B	J	B	J	B	J	B
ค่าเฉลี่ย	100	0.1922	0.9535	0.1840	0.9250	0.0961	0.8824	0.1671	0.9130
	200	0.1795	0.9690	0.1633	0.9430	0.0754	0.8961	0.1499	0.9326
	300	0.1674	0.9787	0.1523	0.9584	0.0652	0.9080	0.1374	0.9283
	400	0.1449	0.9821	0.1346	0.9741	0.0438	0.9256	0.1107	0.9680
	500	0.1297	0.9923	0.1023	0.9807	0.0291	0.9343	0.0969	0.9538
ความแปรปรวน	100	0.1560	0.8543	0.1385	0.8419	0.0823	0.8202	0.0942	0.8355
	200	0.1398	0.8634	0.1256	0.8476	0.0613	0.8299	0.0734	0.8369
	300	0.1179	0.8658	0.1051	0.8571	0.0487	0.8400	0.0646	0.8428
	400	0.0976	0.8783	0.0935	0.8613	0.0242	0.8472	0.0596	0.8556
	500	0.0881	0.8908	0.0711	0.8824	0.0149	0.8349	0.0460	0.8676

ตารางที่ 2 (ต่อ)

พารามิเตอร์	n	การแจกแจงของข้อมูล							
		แบบปกติมาตรฐาน		แบบที		แบบเอ็กซ์โพเนนเชียล		แบบแกมมา	
		J	B	J	B	J	B	J	B
ความเบ้	100	0.1029	0.8009	0.0694	0.7865	0.0887	0.7535	0.0891	0.7737
	200	0.0973	0.8359	0.0496	0.8093	0.0517	0.7690	0.0744	0.7854
	300	0.0621	0.8619	0.0279	0.8223	0.0270	0.7890	0.0521	0.8045
	400	0.0565	0.8628	0.0109	0.8495	0.0147	0.7932	0.0420	0.8178
	500	0.0481	0.8734	0.0093	0.8612	0.0108	0.8082	0.0305	0.8395
ความโด่ง	100	0.1280	0.8151	0.0945	0.7402	0.0875	0.6372	0.1459	0.6896
	200	0.1181	0.8072	0.0819	0.7385	0.0711	0.6189	0.1265	0.6696
	300	0.0971	0.7727	0.0614	0.7207	0.0657	0.6018	0.1020	0.6504
	400	0.0824	0.7512	0.0540	0.7174	0.0468	0.5877	0.0947	0.6350
	500	0.0606	0.7261	0.0500	0.7031	0.0355	0.5730	0.0737	0.6279

จากตารางที่ 1 และตารางที่ 2 ที่ระดับนัยสำคัญ 0.01 และ 0.05 เมื่อข้อมูลมีการแจกแจงแบบแบบปกติมาตรฐาน การแจกแจงแบบที การแจกแจงแบบเอ็กซ์โพเนนเชียล และการแจกแจงแบบแกมมา พบว่า ค่าสัมประสิทธิ์ความเชื่อมั่นของวิธีบูตสเตรปสูงกว่าวิธีแจ๊คไนฟ์ ในทุกขนาดตัวอย่างและทุก ๆ ค่าพารามิเตอร์ ซึ่งได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง เนื่องจากค่าเฉลี่ยความแปรปรวนของการประมาณค่าพารามิเตอร์แบบช่วงด้วยวิธีบูตสเตรปจะมีค่ามากกว่าวิธีแจ๊คไนฟ์ในทุกขนาดตัวอย่าง จึงทำให้ช่วงความเชื่อมั่นที่ประมาณด้วยวิธีบูตสเตรปจะกว้างกว่าช่วงความเชื่อมั่นที่ประมาณด้วยวิธีแจ๊คไนฟ์ ดังนั้นการประมาณค่าพารามิเตอร์แบบช่วงด้วยวิธีบูตสเตรปจึงครอบคลุมค่าพารามิเตอร์ได้มากกว่าการประมาณค่าพารามิเตอร์แบบช่วงด้วยวิธีแจ๊คไนฟ์ ซึ่งผลที่ได้จากการวิจัยฉบับนี้สอดคล้องกับผลที่ได้จากงานวิจัยของ Aiemsuwan & Maiklad (2013) และ Boonpen et al. (2015)

บทสรุป

จากการเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นของวิธีการประมาณค่าพารามิเตอร์แบบช่วงระหว่างวิธีแจ๊คไนฟ์กับวิธีบูตสเตรป ในการประมาณพารามิเตอร์จำนวน 4 ค่า ได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง ของข้อมูลที่มีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบที การแจกแจงแบบเอ็กซ์โพเนนเชียล และการแจกแจงแบบแกมมา เมื่อกำหนดขนาดตัวอย่างเท่ากับ 100,

200, 300, 400 และ 500 ที่ระดับนัยสำคัญเท่ากับ 0.01 และ 0.05 สามารถสรุปผลการวิจัยได้ดังนี้

1. เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน วิธีบูตสเตรปจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของวิธีแจ๊คไนฟ์ ในทุกขนาดตัวอย่าง และทุกค่าพารามิเตอร์
 2. เมื่อข้อมูลมีการแจกแจงแบบที วิธีบูตสเตรปจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของวิธีแจ๊คไนฟ์ ในทุกขนาดตัวอย่าง และทุกค่าพารามิเตอร์
 3. เมื่อข้อมูลมีการแจกแจงแบบเอ็กซ์โพเนนเชียล วิธีบูตสเตรปจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของวิธีแจ๊คไนฟ์ ในทุกขนาดตัวอย่าง และทุกค่าพารามิเตอร์
 4. เมื่อข้อมูลมีการแจกแจงแบบแกมมา วิธีบูตสเตรปจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของวิธีแจ๊คไนฟ์ ในทุกขนาดตัวอย่าง และทุกค่าพารามิเตอร์
- ดังนั้นจึงสรุปได้ว่าการประมาณพารามิเตอร์แบบช่วงด้วยวิธีบูตสเตรปจะมีประสิทธิภาพดีกว่าวิธีแจ๊คไนฟ์ ทุกขนาดตัวอย่าง ทุกค่าพารามิเตอร์ ในทั้ง 4 ลักษณะการแจกแจง

เอกสารอ้างอิง

- Aiemsuwan, P. & Maiklad, T. (2013). A comparisom of confidence intervals for the quartile range. *The Journal of KMUTNB*. 23(2), 410-418.
- Boonpen, S., Chomtee, B., & Hirunwong, A. (2015). A comparison of Intervals estimation methods for scale parameter of the two-parameter Weibull distribution. *Thai Science and Technology Journal*. 23(4), 579-587.
- Efron, B. (1979). Bootstrapping methods: Another look at the Jackknife. *The Annals of Statistics*. 7, 1-26.
- Efron, B. (1982). *The Jackknife the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.
- Puttachaiyatad, K. (2009). *A comparison of Estimation Methods for symmetric and asymmetric distributed data by Jackknifing methos and Bootstrapping method*. In the 19 th Thaksin University annual conference (pp. 157-167). Songkhla: Thailand.
- Miller, R. G. (1974). The Jackknife – a Review. *Biometrika*. 61, 1-15.
- Quenouille, M. H. (1956). Notes on Bias in estimation. *Biometrika*. 43, 353-360.