

การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าเบี่ยงเบนมาตรฐาน สำหรับการแจกแจงปกติเมื่อข้อมูลมีค่านอกเกณฑ์

Efficiency Comparison of the Standard Deviation Estimation Methods for a Normal Distribution When Data Set Containing Outliers

ภัทรธร บุญยะวาหะ*, ณัฐชัชชา ปรัชชญาลัมเลิศ, เบญจจิรา บุญเพชร
และจuthาภรณ์ สิ้นสมบุญรณทอง

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

Pattaratorn Boonyavaha*, Nutchicha Pratchayalamlert, Benjira Boonpet
and Juthaphorn Sinsomboonthong

Department of Statistics, Faculty of Science, Kasetsart University, Ladyao, Chatuchak, Bangkok 10900

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากร สำหรับการแจกแจงปกติเมื่อข้อมูลมีค่านอกเกณฑ์ 5 วิธี คือ วิธี sample standard deviation (SD) วิธี mean absolute deviation (MAD) วิธี adjusted range (AR) วิธี percentile tab-standard deviation (PSD) และวิธี adjusted standard deviation (ASD) ในการศึกษาได้จำลองข้อมูลที่มีการแจกแจงปกติค่าเฉลี่ยเท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานประชากรเท่ากับ 1, 5, 10, 15 และ 20 กำหนดขนาดตัวอย่างเท่ากับ 10, 20, 30, 50, 70 และ 100 โดยใช้โปรแกรม SAS เมื่อร้อยละค่านอกเกณฑ์เท่ากับ 0, 10 และ 20 ของขนาดตัวอย่าง รวมสถานการณ์ที่ศึกษาทั้งหมด 90 สถานการณ์ ซึ่งเกณฑ์ในการเปรียบเทียบประสิทธิภาพ คือ ค่าความเอนเอียงสมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ผลการวิจัยพบว่าเมื่อข้อมูลไม่มีค่านอกเกณฑ์ปลอมปน ถ้าพิจารณาจากค่า ABS การประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี ASD มีประสิทธิภาพดีที่สุดในทุกสถานการณ์ และถ้าพิจารณาจากค่า MSE พบว่าการประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี SD มีประสิทธิภาพดีที่สุดในทุกสถานการณ์ อย่างไรก็ตาม เมื่อข้อมูลมีค่านอกเกณฑ์ปลอมปนร้อยละ 10 และ 20 การประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี MAD มีแนวโน้มให้ค่า ABS และ MSE ต่ำที่สุดเกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 30 ร้อยละค่านอกเกณฑ์เท่ากับ 20 ของขนาดตัวอย่าง และค่าเบี่ยงเบนมาตรฐานประชากรเท่ากับ 5 พบว่าการประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี PSD มีประสิทธิภาพดีที่สุด

คำสำคัญ : ค่าเบี่ยงเบนมาตรฐาน; การแจกแจงปกติ; ความเอนเอียงสัมบูรณ์; ความคลาดเคลื่อนกำลังสองเฉลี่ย

Abstract

The objective of this research was to compare the efficiency of five population standard deviation estimation methods – sample standard deviation (SD), mean absolute deviation (MAD), adjusted range (AR), percentile tab-standard deviation (PSD) and adjusted standard deviation (ASD) methods – for a normal distribution when data set containing outliers. The simulation data in the form of normal distribution with mean (μ) equals 30 and the population standard deviation (σ) equals 1, 5, 10, 15 and 20 were generated by SAS programming. In addition, the sample sizes (n) in this study were determined at 10, 20, 30, 50, 70 and 100, and the percentages of mild outliers were set at 0, 10 and 20 of the sample sizes. The totals of 90 situations were studied. The criteria for efficiency comparison were absolute bias (ABS) and mean square error (MSE). The conclusions of this research were as follows: in the case of no-outliers, adjusted standard deviation (ASD) method was the most efficient estimator for all situations based on considering the amount of ABS. In addition, sample standard deviation (SD) method was the most efficient estimator for all situations based on considering the amount of MSE. However, when the percentages of outliers were contaminated with the data equal 10 and 20, mean absolute deviation (MAD) method tended to have the lowest ABS and MSE for almost all situations. Unless, the percentage of mild outliers was 20 for a sample size of 30 and population standard deviation (σ) was 5, it was found that percentile tab-standard deviation (PSD) method was the most efficient estimator.

Keywords: standard deviation; normal distribution; absolute bias; mean square error

1. บทนำ

ปัจจุบันพบว่าสถิติเข้ามามีบทบาทในงานวิจัยมากขึ้น กล่าวคือ เป็นเครื่องมือสำคัญที่ช่วยในการวิเคราะห์ข้อมูล ค่าเบี่ยงเบนมาตรฐาน (standard deviation) เป็นค่าวัดการกระจายแบบหนึ่งของข้อมูลว่าข้อมูลมีการกระจายมากน้อยเพียงใด โดยค่าเบี่ยงเบนมาตรฐานควรจะมีค่าน้อย ๆ ถ้าค่าเบี่ยงเบนมาตรฐานมีค่าน้อยแสดงว่าข้อมูลมีการกระจายน้อย ซึ่งค่าวัดแนวโน้มสู่ส่วนกลางที่คำนวณได้จะมีความแม่นยำ แต่ถ้าค่าเบี่ยงเบนมาตรฐานมีค่ามากแสดงว่าข้อมูลมี

การกระจายมากจะทำให้ค่าวัดแนวโน้มสู่ส่วนกลางที่คำนวณออกมามีความเบี่ยงเบนมากและการประมาณค่าเบี่ยงเบนมาตรฐานประชากรจึงมีความจำเป็นต่อการอนุมานทางสถิติซึ่งถ้าค่าประมาณที่ได้ใกล้เคียงกับค่าจริงจะส่งผลทำให้การวิเคราะห์ข้อมูลมีความถูกต้องแม่นยำสูงขึ้น ในการเก็บรวบรวมข้อมูลที่จะนำมาวิเคราะห์นั้น ผู้วิจัยอาจไม่สามารถเก็บรวบรวมข้อมูลมาได้จากทุกหน่วยในประชากร เนื่องจากมีค่าใช้จ่ายสูงและใช้ระยะเวลานานในการเก็บรวบรวมข้อมูล ดังนั้นจึงต้องมีการเก็บข้อมูลจากตัวอย่างเพื่อนำมาวิเคราะห์

ภายหลังจากการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างแล้ว จะใช้ผลวิเคราะห์ข้อมูลที่ได้จากข้อมูลตัวอย่างไปสรุปผลของประชากร ซึ่งเป็นเป้าหมายหลักของการศึกษาวิจัยต่อไปได้ ในการเก็บรวบรวมข้อมูลบ่อยครั้งที่พบว่าจะมีข้อมูลที่มีค่าต่ำเกินไปหรือสูงเกินไปจากข้อมูลส่วนใหญ่ปลอมปนอยู่ ซึ่งเรียกว่าค่านอกเกณฑ์ (Outliers) โดยสาเหตุของการเกิดค่านอกเกณฑ์ [1] อาจเกิดจากความคลาดเคลื่อนของการวัด (measurement error) หรือความคลาดเคลื่อนจากการปฏิบัติการ (execution error) เช่นการบันทึกข้อมูลผิด รวมไปถึงความคลาดเคลื่อนจากการผันแปรที่มีอยู่ในประชากรที่ทำการศึกษา (inherent variability) ซึ่งไม่ได้เกิดจากการวัดหรือการบันทึกข้อมูลที่ผิดพลาด แต่เป็นความผิดพลาดที่หลีกเลี่ยงไม่ได้ เนื่องจากไม่ว่าจะควบคุมการวัดหรือการเก็บข้อมูลได้ดีเพียงใด ความผันแปรในประชากรก็ยังคงมีอยู่ ดังนั้นการนำข้อมูลที่มีค่านอกเกณฑ์มาใช้ อาจทำให้การวิเคราะห์ข้อมูลมีความผิดพลาดเกิดขึ้น ซึ่งอาจแก้ปัญหาได้โดยการตัดข้อมูลที่มีค่านอกเกณฑ์ออกหรือปรับค่าให้เหมาะสม ซึ่งขึ้นอยู่กับดุลยพินิจของผู้ทำวิจัย [2]

การประมาณค่าเบี่ยงเบนมาตรฐานประชากรมีหลายวิธี และในปัจจุบันมักนิยมใช้วิธี sample standard deviation (SD) ในการประมาณค่าเบี่ยงเบนมาตรฐานประชากร แต่พบว่าวิธีดังกล่าวมีความเอนเอียง ซึ่งจะทำให้ค่าประมาณที่คำนวณได้มีความคลาดเคลื่อนสูงและมีประสิทธิภาพน้อยเมื่อขนาดตัวอย่างเล็ก [3] นอกจากนี้พบว่ามีวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธีอื่น ๆ ที่มีประสิทธิภาพอีก เช่น วิธี mean absolute deviation (MAD) วิธี adjusted range (AR) วิธี percentile tab-standard deviation (PSD) วิธี adjusted standard deviation (ASD) โดย Woodall และ Montgomery [4] ได้เสนอไว้ว่าสำหรับข้อมูลที่มีการแจกแจงปกติ

การประมาณค่าเบี่ยงเบนมาตรฐานประชากรด้วยวิธี adjusted range (AR) และวิธี adjusted standard deviation (ASD) จะทำให้ได้ตัวประมาณที่ไม่เอนเอียง และในการเปรียบเทียบประสิทธิภาพวิธีการประมาณทั้ง 2 วิธี พบว่าวิธี adjusted standard deviation (ASD) มีประสิทธิภาพมากกว่า นอกจากนี้ Montgomery [3] ได้เปรียบเทียบประสิทธิภาพตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากร 2 วิธี คือ วิธี adjusted range (AR) และวิธี adjusted standard deviation (ASD) ในการประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยใช้ค่าประสิทธิภาพสัมพัทธ์เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพ ผลการวิจัยพบว่าเมื่อขนาดตัวอย่าง (n) เพิ่มขึ้น ค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี adjusted range (AR) เทียบกับตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี adjusted standard deviation (ASD) จะลดลงซึ่งหมายความว่า การประมาณค่าเบี่ยงเบนมาตรฐานประชากรโดยวิธี adjusted standard deviation (ASD) จะมีประสิทธิภาพมากกว่า

ในการวิจัยครั้งนี้ ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพการประมาณค่าเบี่ยงเบนมาตรฐานสำหรับข้อมูลที่มีการแจกแจงปกติเมื่อข้อมูลมีค่านอกเกณฑ์ทั้งหมด 5 วิธี คือ วิธี sample standard deviation (SD) วิธี mean absolute deviation (MAD) วิธี adjusted range (AR) วิธี percentile tab-standard deviation (PSD) และวิธี adjusted standard deviation (ASD) โดยจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพทั้งหมด 90 สถานการณ์ เพื่อเป็นแนวทางให้แก่ผู้วิจัยในงานด้านต่าง ๆ เลือกใช้วิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรให้เหมาะสมกับข้อมูลที่มีอยู่ และให้ความสำคัญกับค่านอกเกณฑ์ของข้อมูลมากขึ้น เพื่อให้ได้ผลลัพธ์ของงาน

วิจัยที่มีความแม่นยำมากขึ้น

2. วิธีการวิจัย

การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปรกติเมื่อข้อมูลมีค่านอกเกณฑ์ทั้ง 5 วิธี คือ วิธี sample standard deviation (SD) วิธี mean absolute deviation (MAD) วิธี adjusted range (AR) วิธี percentile tab-standard deviation (PSD) และวิธี adjusted standard deviation (ASD) มีขั้นตอนการวิจัย ดังต่อไปนี้

2.1 จำลองข้อมูลที่มีการแจกแจงปรกติ โดยมีค่าเฉลี่ย (μ) เท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานประชากร (σ) เท่ากับ 1, 5, 10, 15 และ 20

2.2 กำหนดขนาดตัวอย่าง (n) เท่ากับ 10, 20, 30, 50, 70 และ 100

2.3 กำหนดร้อยละของค่านอกเกณฑ์ ในระดับปานกลาง (mild outlier) ด้านมาก โดยร้อยละของค่านอกเกณฑ์จากกลุ่มมีทั้งหมด 3 ระดับ คือ 0, 10 และ 20 % ของขนาดตัวอย่าง ซึ่งสร้างค่านอกเกณฑ์จากกลุ่มตามหลักการที่ Barnett และ Lewis [5] กล่าวถึง ซึ่งค่านอกเกณฑ์อยู่ในช่วง ($Q_3 + 1.5 \text{ IQR}, Q_3 + 3 \text{ IQR}$) เมื่อ Q_1 คือ ค่าควอไทล์ที่ 1; Q_3 คือ ค่าควอไทล์ที่ 3; IQR คือ ระยะระหว่างควอไทล์ซึ่งมีค่าเท่ากับ $Q_3 - Q_1$

2.4 การตรวจสอบค่านอกเกณฑ์ สามารถตรวจสอบได้หลายวิธี คือ การตรวจสอบโดยวิธีกราฟซึ่งเป็นวิธีการตรวจสอบที่ง่ายและสะดวก หรือการตรวจสอบด้วยการทดสอบทางสถิติซึ่งตัวสถิติทดสอบมีหลายวิธี เช่น ตัวสถิติทดสอบ T_{NI} [5] ที่ใช้ในกรณีไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากร หรือตัวสถิติทดสอบของ Furguson [6] ที่ใช้สัมประสิทธิ์ความเบ้และสัมประสิทธิ์ความโค้งเป็นพื้นฐานในการ

ทดสอบค่านอกเกณฑ์ สำหรับงานวิจัยนี้ได้ตรวจสอบค่านอกเกณฑ์โดยวิธีกราฟ ซึ่งตรวจสอบด้วยแผนภาพกล่อง (box plot) โดยค่าสังเกตที่มีค่านอกเกณฑ์ในระดับปานกลางทางด้านมากจะมีค่ามากกว่า Q_3 ระหว่าง 1.5 ถึง 3 เท่าของ IQR

2.5 ใช้โปรแกรม SAS เวอร์ชัน 9.4 ในการจำลองข้อมูลที่มีการแจกแจงปรกติ โดยเทคนิคมอนติคาร์โล มีการทำซ้ำ 2,000 ครั้ง ในแต่ละสถานการณ์

2.6 ให้ x_1, x_2, \dots, x_n เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงปรกติด้วยค่าเฉลี่ย μ และความแปรปรวน σ^2 วิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากร (σ) ที่ใช้ในงานวิจัยมีดังนี้

2.6.1 วิธี sample standard deviation (SD) เป็นวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรทั่วไปที่นิยมใช้กัน ซึ่งเป็นตัวประมาณที่มีความเอนเอียง สูตรที่ใช้ในการคำนวณ [3] คือ

$$\hat{\sigma}_{SD} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

โดยที่ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ คือ ค่าเฉลี่ยตัวอย่าง; n คือ ขนาดตัวอย่าง

2.6.2 วิธี mean absolute deviation (MAD) บางครั้งเรียกว่าวิธีค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย (average absolute deviation) สูตรที่ใช้ในการคำนวณ [7] คือ

$$\hat{\sigma}_{MAD} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

โดยที่ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ คือ ค่าเฉลี่ยตัวอย่าง; n คือ ขนาดตัวอย่าง

2.6.3 วิธี adjusted range (AR) เป็นการใช้พิสัยของตัวอย่าง มาประมาณค่าเบี่ยงเบนมาตรฐานประชากร เพื่อให้ได้ตัวประมาณที่ไม่เอนเอียง สูตรที่ใช้

ในการคำนวณ [8] คือ

$$\hat{\sigma}_{AR} = \frac{R}{d_2}$$

โดยที่ $R = x_{max} - x_{min}$ คือ ค่าพิสัยตัวอย่าง; d_2 คือ ค่าคงที่ที่เปิดจากตารางที่ 1 [9] ซึ่งขึ้นกับค่า n เมื่อ $n=1, 2, 3, \dots, 50$; $d_2 = 3.4873 + 0.0250141 \times n - 0.00009823 \times n^2$ เมื่อ $n=51, 52, 53, \dots, 100$

2.6.4 วิธี percentiles tab-standard deviation (PSD) เป็นวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปกติ โดยการใช้เปอร์เซ็นต์ไทล์ 2 ตัว มาประมาณค่าเบี่ยงเบนมาตรฐานประชากร สูตรที่ใช้ในการคำนวณ [10] คือ

$$\hat{\sigma}_{PSD} = \left| \frac{X_2 - X_1}{Z\left(\frac{P_2}{100}\right) - Z\left(\frac{P_1}{100}\right)} \right|$$

โดยที่ x_1 คือ ค่าเปอร์เซ็นต์ไทล์ที่ 25; x_2 คือ ค่าเปอร์เซ็นต์ไทล์ที่ 75; $P_1 = 25$, $P_2 = 75$; $Z\left(\frac{P}{100}\right)$ คือ ค่า z ที่เปิดจากตารางการแจกแจงปกติมาตรฐานซึ่งทำให้พื้นที่ทางซ้ายมีค่าเท่ากับ $\frac{P}{100}$

2.6.5 วิธี adjusted standard deviation (ASD) เป็นการนำค่าเบี่ยงเบนมาตรฐานตัวอย่างมาปรับค่า เพื่อให้ได้ตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากรที่ไม่เอนเอียง สูตรที่ใช้ในการคำนวณ [11] คือ

$$\hat{\sigma}_{ASD} = \frac{\hat{\sigma}_{SD}}{c_4}$$

โดยที่ $\hat{\sigma}_{SD} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$; $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ คือ ค่าเฉลี่ยตัวอย่าง; c_4 คือ ค่าคงที่ที่เปิดจากตารางที่ 1 [3] ซึ่งขึ้นกับค่า n เมื่อ $n=1, 2, 3, \dots, 25$;

$$c_4 = \sqrt{\frac{2}{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \text{ เมื่อ } n=26, 27, 28, \dots, 100;$$

$$\Gamma(n) \text{ เป็นฟังก์ชันแกมมา ซึ่ง } \Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$$

ตารางที่ 1 ค่าคงที่สำหรับการคำนวณ $\hat{\sigma}_{AR}$ และ $\hat{\sigma}_{ASD}$

n	d ₂	c ₄	n	d ₂	c ₄	n	d ₂
2	1.128	0.7979	19	3.689	0.9862	36	4.236
3	1.693	0.8862	20	3.735	0.9869	37	4.259
4	2.059	0.9213	21	3.778	0.9876	38	4.280
5	2.326	0.9400	22	3.819	0.9882	39	4.301
6	2.534	0.9515	23	3.858	0.9887	40	4.322
7	2.704	0.9594	24	3.895	0.9892	41	4.341
8	2.847	0.9650	25	3.931	0.9896	42	4.361
9	2.970	0.9693	26	3.964		43	4.379
10	3.078	0.9727	27	3.997		44	4.398
11	3.173	0.9754	28	4.027		45	4.415
12	3.258	0.9776	29	4.057		46	4.433
13	3.336	0.9794	30	4.086		47	4.450
14	3.407	0.9810	31	4.113		48	4.466
15	3.472	0.9823	32	4.139		49	4.482
16	3.532	0.9835	33	4.165		50	4.498
17	3.588	0.9845	34	4.189			
18	3.640	0.9854	35	4.213			

2.7 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ

เกณฑ์การตัดสินว่าการประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธีใดมีประสิทธิภาพในการประมาณค่า จะพิจารณาจากค่าความเอนเอียงสัมบูรณ์ (absolute bias, ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error, MSE)

ซึ่งวิธีการประมาณค่าวิธีใดที่ให้ค่าความเอนเอียงสัมบูรณ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด จะถือว่าเป็นวิธีประมาณค่าเบี่ยงเบนมาตรฐานประชากรที่มีประสิทธิภาพดีที่สุด โดยค่าความเอนเอียงสัมบูรณ์ (ABS) ของตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากร ($\hat{\sigma}$) สามารถประมาณได้จากสูตร

$$\overline{ABS} = \left| \frac{\sum_{k=1}^m (\hat{\sigma})_k}{m} - \sigma \right|$$

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณค่าเบี่ยงเบนมาตรฐานประชากร ($\hat{\sigma}$) สามารถประมาณได้จากสูตร

$$\overline{MSE} = \frac{1}{m} \sum_{k=1}^m [(\hat{\sigma})_k - \sigma]^2$$

เมื่อ $(\hat{\sigma})_k$ คือ ค่าประมาณค่าเบี่ยงเบนมาตรฐานประชากรของแต่ละวิธีในการทำซ้ำรอบที่ k สำหรับสถานการณ์ต่าง ๆ; σ คือ ค่าเบี่ยงเบนมาตรฐานประชากรที่กำหนด; m คือ จำนวนรอบที่ทำซ้ำ ในที่นี้ $m = 2,000$ รอบ

3. ผลการวิจัย

งานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปกติเมื่อข้อมูลมีค่านอกเกณฑ์ทั้งหมด 5 วิธี คือ วิธี sample standard deviation (SD) วิธี mean absolute deviation (MAD) วิธี adjusted range (AR) วิธี percentile tab-standard deviation (PSD) และวิธี adjusted standard deviation (ASD) โดยเกณฑ์ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ซึ่งวิธีใดที่ให้ค่า ABS และค่า MSE ต่ำกว่า จะมีประสิทธิภาพมากกว่า และถือว่าเป็นวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปกติที่มีประสิทธิภาพ

มากที่สุด การศึกษาครั้งนี้ได้จำลองข้อมูลประชากรที่มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ย (μ) เท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานประชากร (σ) เท่ากับ 1, 5, 10, 15 และ 20 โดยเทคนิคมอนติคาร์โล (Monte Carlo simulation) ด้วยโปรแกรม SAS 9.4 และสุ่มตัวอย่างขนาด (n) เท่ากับ 10, 20, 30, 50, 70 และ 100 จากประชากรดังกล่าว นอกจากนี้ยังกำหนดร้อยละค่านอกเกณฑ์ในระดับที่แตกต่างกันตั้งแต่ 0, 10 และ 20 % ของขนาดตัวอย่าง งานวิจัยนี้ได้ตรวจสอบค่านอกเกณฑ์ของข้อมูลที่จำลองทุกสถานการณ์โดยใช้แผนภาพกล่อง (box plot) โดยค่าสังเกตที่มีค่านอกเกณฑ์ในระดับปานกลางจะมีค่ามากกว่า Q_3 ระหว่าง 1.5 ถึง 3 เท่าของ IQR รวมสถานการณ์ที่จำลองข้อมูลทั้งหมด 90 สถานการณ์ และในแต่ละสถานการณ์ทดลองซ้ำ 2,000 ครั้ง ผลการวิจัยแสดงในตารางที่ 2 ถึง 4

เมื่อพิจารณาจากตารางที่ 2 คือ ข้อมูลไม่มีค่านอกเกณฑ์ปลอมปน ถ้าพิจารณาจากค่าความเอนเอียงสัมบูรณ์ (ABS) การประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธี adjusted standard deviation (ASD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพดีที่สุดในทุกสถานการณ์ แต่ถ้าพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) การประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธี sample standard deviation (SD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพดีที่สุดในทุกสถานการณ์ นอกจากนี้ถ้าขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีแนวโน้มลดลง

เมื่อพิจารณาจากตารางที่ 3 คือข้อมูลมีค่านอกเกณฑ์ปลอมปนร้อยละ 10 ถ้าพิจารณาจากค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) การประมาณค่า

เบี่ยงเบนมาตรฐานประชากรวิธี mean absolute deviation (MAD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพดีที่สุดในทุกสถานการณ์ เนื่องจากให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ต่ำที่สุด นอกจากนี้ถ้าขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีแนวโน้มลดลง

เมื่อพิจารณาจากตารางที่ 4 คือข้อมูลมีค่า นอกเกณฑ์ปลอมปนร้อยละ 20 ถ้าพิจารณาจากค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) การประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธี mean absolute deviation (MAD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพดีที่สุดในเกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่าง (n) เท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานประชากร (σ) เท่ากับ 5 การประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธี percentile tab-standard deviation (PSD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพดีที่สุดในเกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีแนวโน้มลดลง

ตารางที่ 2 ถึงตารางที่ 4 จะเห็นได้ว่าเมื่อร้อยละของค่านอกเกณฑ์มีค่าเพิ่มขึ้น ค่า ABS และค่า MSE ของวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรทั้ง 5 วิธี จะมีแนวโน้มเพิ่มขึ้นอย่างเห็นได้ชัด

4. วิจัยณ์

ในงานวิจัยนี้ใช้เกณฑ์การเปรียบเทียบประสิทธิภาพ 2 เกณฑ์ คือ ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) สำหรับกรณีที่ข้อมูลไม่มีค่านอกเกณฑ์ปลอมปนนั้น

การประมาณค่าเบี่ยงเบนมาตรฐานประชากร (σ) โดยวิธี sample standard deviation (SD) ค่อนข้างนิยมใช้กันเนื่องจากในผลการวิจัยนี้พบว่าให้ค่า MSE ต่ำที่สุดในทุกขนาดตัวอย่าง (n) แต่เมื่อพิจารณาจากเกณฑ์ค่าความเอนเอียงสัมบูรณ์ พบว่ากรณีที่ข้อมูลไม่มีค่านอกเกณฑ์ปลอมปน ในทุกขนาดตัวอย่าง (n) การประมาณค่าเบี่ยงเบนมาตรฐานประชากร (σ) ที่มีค่า ABS ต่ำที่สุดคือ วิธี adjusted standard deviation (ASD) นอกจากนี้วิธี adjusted range (AR) จะให้ประสิทธิภาพที่ใกล้เคียงกับวิธี adjusted standard deviation (ASD) เมื่อขนาดตัวอย่าง (n) น้อยและค่าเบี่ยงเบนมาตรฐานประชากร (σ) ไม่เกิน 5 ซึ่งสอดคล้องกับการศึกษาของ Woodall และ Montgomery [4] และโดยทั่วไปพบว่าการสร้างขีดจำกัดควบคุมในแผนภูมิควบคุมคุณภาพของชีวฮาร์ท [3] มักจะนิยมประมาณค่าเบี่ยงเบนมาตรฐานประชากร (σ) ด้วยวิธี adjusted standard deviation (ASD) มากกว่าวิธี sample standard deviation (SD) เนื่องจากในกระบวนการควบคุมคุณภาพจะสุ่มวัดจุดหรือผลิตภัณฑ์มาตรวจสอบคุณภาพแต่ละครั้งด้วยขนาดตัวอย่าง (n) น้อยซึ่งบางครั้งสุ่มตรวจสอบครั้งละไม่เกิน 10 ชิ้นแต่จะสุ่มตรวจบ่อยครั้ง และตัวประมาณโดยวิธี ASD เป็นตัวประมาณไม่เอนเอียงของค่าเบี่ยงเบนมาตรฐานประชากร [3] นอกจากนี้เมื่อพิจารณาจากตารางที่ 2 พบว่าค่า MSE ของตัวประมาณโดยวิธี ASD ให้ค่าใกล้เคียงกับการประมาณโดยวิธี sample standard deviation (SD) ในทุกระดับของค่าเบี่ยงเบนมาตรฐานประชากร (σ) และขนาดตัวอย่าง (n) เมื่อข้อมูลไม่มีค่านอกเกณฑ์ปลอมปน

ผลการวิจัยครั้งนี้ยังพบว่าเมื่อข้อมูลไม่มีค่านอกเกณฑ์ปลอมปน ค่าความเอนเอียงสัมบูรณ์ของวิธี adjusted standard deviation (ASD) และวิธี

ตารางที่ 2 ค่า ABS และ MSE เมื่อร้อยละของค่านอกเกณฑ์เท่ากับ 0

n	σ	SD		MAD		AR		PSD		ASD	
		ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE
10	1	0.247	0.051*	0.247	0.095	0.005	0.061	0.097	0.130	0.004*	0.053
	5	1.234	1.274*	1.234	2.370	0.025	1.525	0.488	3.242	0.021*	1.321
	10	2.468	5.096*	2.468	9.479	0.049	6.099	0.975	12.968	0.042*	5.284
	15	3.702	11.466*	3.702	21.327	0.074	13.722	1.462	29.175	0.064*	11.888
	20	4.936	20.385*	4.936	37.915	0.098	24.396	1.950	51.867	0.085*	21.135
20	1	0.110	0.025*	0.227	0.069	0.008	0.037	0.037	0.062	0.007*	0.026
	5	0.099	0.628*	1.136	1.717	0.038	0.921	0.187	1.542	0.034*	0.636
	10	0.198	2.514*	2.273	6.868	0.076	3.684	0.373	6.170	0.068*	2.545
	15	0.297	5.656*	3.409	15.457	0.114	8.288	0.560	13.882	0.102*	5.727
	20	0.396	10.055*	4.545	27.473	0.152	14.734	0.746	24.680	0.136*	10.181
30	1	0.011	0.017*	0.216	0.058	0.006	0.029	0.029	0.044	0.001*	0.017*
	5	0.053	0.417*	1.079	1.457	0.028	0.724	0.148	1.098	0.004*	0.423
	10	0.106	1.669*	2.158	5.827	0.057	2.896	0.295	4.393	0.009*	1.691
	15	0.158	3.756*	3.237	13.111	0.085	6.517	0.443	9.884	0.013*	3.804
	20	0.211	6.677*	4.317	23.308	0.113	11.586	0.590	17.571	0.017*	6.763
50	1	0.007	0.010*	0.211	0.052	0.007	0.021	0.020	0.028	0.001*	0.010*
	5	0.035	0.256*	1.055	1.296	0.036	0.515	0.097	0.688	0.007*	0.257
	10	0.070	1.022*	2.110	5.182	0.072	2.058	0.195	2.752	0.014*	1.029
	15	0.105	2.300*	3.164	11.661	0.108	4.630	0.292	6.193	0.021*	2.315
	20	0.140	4.089*	4.219	20.730	0.144	8.232	0.389	11.009	0.028*	4.116
70	1	0.006	0.007*	0.209	0.049	0.005	0.018	0.012	0.019	0.002*	0.007*
	5	0.028	0.179*	1.044	1.215	0.026	0.444	0.062	0.462	0.009*	0.180
	10	0.056	0.717*	2.088	4.859	0.053	1.774	0.123	1.849	0.017*	0.720
	15	0.084	1.613*	3.132	10.932	0.079	3.991	0.185	4.160	0.026*	1.619
	20	0.112	2.868*	4.177	19.434	0.106	7.095	0.246	7.395	0.034*	2.879
100	1	0.005	0.005*	0.208	0.047	0.004	0.015	0.008	0.014	0.002*	0.005*
	5	0.026	0.136*	1.038	1.174	0.021	0.366	0.039	0.342	0.010*	0.136*
	10	0.053	0.543*	2.076	4.694	0.041	1.463	0.079	1.369	0.020*	0.544
	15	0.079	1.221*	3.114	10.561	0.062	3.292	0.118	3.081	0.030*	1.224
	20	0.106	2.171*	4.152	18.776	0.083	5.853	0.157	5.476	0.040*	2.176

* คือ วิธีที่ให้ค่า ABS และ MSE ต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 3 ค่า ABS และ MSE เมื่อร้อยละของค่าอันอกเกณฑ์เท่ากับ 10

n	σ	SD		MAD		AR		PSD		ASD	
		ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE
10	1	0.474	0.428	0.041*	0.093*	0.654	0.710	0.274	0.224	0.515	0.481
	5	2.369	10.712	0.207*	2.317*	3.271	17.752	1.369	5.608	2.575	12.024
	10	4.737	42.850	0.414*	9.270*	6.541	71.012	2.739	22.433	5.151	48.102
	15	7.106	96.409	0.621*	20.856*	9.812	159.773	4.108	50.471	7.726	108.225
	20	9.474	171.393	0.827*	37.077*	13.082	284.039	5.477	89.725	10.302	192.398
20	1	0.467	0.308	0.065*	0.048*	0.548	0.413	0.203	0.121	0.486	0.329
	5	2.334	7.700	0.326*	1.196*	2.741	10.319	1.013	3.026	2.431	8.224
	10	4.668	30.800	0.652*	4.784*	5.483	41.274	2.026	12.102	4.863	32.896
	15	7.002	69.298	0.978*	10.763*	8.224	92.865	3.038	27.231	7.294	74.015
	20	9.336	123.199	1.304*	19.134*	10.965	165.097	4.051	48.410	9.725	131.585
30	1	0.474	0.292	0.082*	0.039*	0.507	0.336	0.197	0.094	0.488	0.307
	5	2.369	7.289	0.408*	0.973*	2.537	8.391	0.984	2.351	2.441	7.669
	10	4.737	29.156	0.816*	3.894*	5.074	33.561	4.016	17.508	4.882	30.677
	15	7.106	65.601	1.225*	8.761*	7.611	75.513	2.953	21.158	7.323	69.023
	20	9.475	116.621	1.633*	15.575*	10.148	134.240	3.937	37.614	9.763	122.704
50	1	0.476	0.266	0.090*	0.028*	0.459	0.256	0.175	0.066	0.484	0.274
	5	2.378	6.651	0.450*	0.694*	2.293	6.393	0.877	1.641	2.420	6.862
	10	4.756	26.604	0.900*	2.778*	4.586	25.571	1.755	6.566	4.839	27.446
	15	7.134	59.861	1.351*	6.250*	6.879	57.537	2.632	14.773	7.259	61.756
	20	9.512	106.419	1.801*	11.112*	9.172	102.287	3.509	26.264	9.678	109.788
70	1	0.474	0.252	0.093*	0.022*	0.423	0.208	0.169	0.053	0.480	0.258
	5	2.371	6.294	0.464*	0.547*	2.114	5.207	0.846	1.312	2.400	6.437
	10	4.743	25.176	0.929*	2.188*	4.227	20.829	1.692	5.249	4.801	25.748
	15	7.114	56.642	1.393*	4.922*	6.340	46.863	2.538	11.808	7.201	57.928
	20	9.485	100.698	1.857*	8.751*	8.454	83.313	2.462	11.431	9.601	102.985
100	1	0.471	0.242	0.094*	0.019*	0.392	0.175	0.163	0.044	0.475	0.246
	5	2.357	6.058	0.468*	0.468*	1.958	4.380	0.816	1.100	2.377	6.155
	10	4.714	24.232	0.935*	1.874*	3.917	17.518	1.631	4.402	4.754	24.620
	15	7.071	54.521	1.403*	4.217*	5.875	39.415	2.447	9.904	7.131	55.394
	20	9.428	96.925	1.871*	7.496*	7.833	70.072	3.262	17.607	9.508	98.478

* คือ วิธีที่ให้ค่า ABS และ MSE ต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 4 ค่า ABS และ MSE เมื่อร้อยละของค่านอกเกณฑ์เท่ากับ 20

n	σ	SD		MAD		AR		PSD		ASD	
		ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS	MSE
10	1	0.769	0.923	0.331*	0.285*	0.746	0.866	0.803	0.958	0.819	1.021
	5	3.845	23.077	1.655*	7.135*	3.731	21.662	4.016	23.940	4.094	25.520
	10	7.691	92.315	3.310*	28.543*	7.462	86.652	8.032	95.767	8.187	102.087
	15	11.536	207.702	4.965*	64.219*	11.193	194.962	12.048	215.468	12.281	229.687
	20	15.382	369.247	6.620*	114.168*	14.923	346.598	16.063	383.052	16.375	408.333
20	1	0.752	0.712	0.358*	0.211*	0.607	0.483	0.579	0.476	0.776	0.752
	5	3.763	17.809	1.793*	5.277*	3.038	12.081	2.895	11.890	3.879	18.795
	10	7.526	71.235	3.587*	21.109*	6.076	48.323	5.789	47.560	7.758	75.181
	15	11.289	160.277	5.380*	47.495*	9.114	108.725	8.684	107.008	11.638	169.155
	20	15.052	284.940	7.173*	84.437*	12.152	193.291	11.578	190.235	15.517	300.723
30	1	0.752	0.673	0.373*	0.201*	0.544	0.373	0.567	0.417	0.769	0.701
	5	3.759	16.824	1.867	5.015	2.719	9.328	1.061*	0.492*	3.845	17.530
	10	7.519	67.294	3.734*	20.060*	5.437	37.308	5.668	41.640	7.690	70.116
	15	11.278	151.414	5.600*	45.135*	8.156	83.943	8.502	93.691	11.535	157.762
	20	15.037	269.172	7.467*	80.238*	10.875	149.227	11.336	166.561	15.380	280.458
50	1	0.759	0.645	0.388*	0.190*	0.482	0.276	0.535	0.349	0.769	0.661
	5	3.795	16.118	1.939*	4.753*	2.408	6.895	3.464	12.064	3.845	16.516
	10	7.591	64.470	3.878*	19.013*	4.816	27.579	5.353	34.897	7.690	66.061
	15	11.387	145.064	5.818*	42.782*	7.224	62.055	8.030	78.516	11.535	148.644
	20	15.182	257.887	7.757*	76.056*	9.632	110.318	10.706	139.581	15.380	264.252
70	1	0.749	0.606	0.386*	0.175*	0.435	0.217	0.523	0.317	0.756	0.617
	5	3.746	15.156	1.929*	4.374*	2.173	5.433	2.614	7.915	3.780	15.423
	10	7.492	60.626	3.859*	17.498*	4.347	21.732	5.228	31.660	7.560	61.692
	15	11.237	136.401	5.788*	39.368*	6.520	48.895	7.842	71.233	11.340	138.800
	20	14.983	242.493	7.717*	69.988*	8.693	86.925	10.455	126.633	15.120	246.757
100	1	0.752	0.600	0.392*	0.173*	0.401	0.183	0.516	0.298	0.757	0.607
	5	3.760	14.988	1.959*	4.330*	2.004	4.563	2.581	7.455	3.784	15.171
	10	7.521	59.951	3.917*	17.319*	4.009	18.250	5.161	29.821	7.568	60.685
	15	11.281	134.887	5.876*	38.968*	6.013	41.062	7.742	67.100	11.352	136.539
	20	15.042	239.801	7.834*	69.276*	8.018	72.999	10.322	119.287	15.137	242.737

* คือ วิธีที่ให้ค่า ABS และ MSE ต่ำที่สุดในแต่ละสถานการณ์

adjusted range (AR) มีค่าเข้าใกล้ศูนย์ในทุกสถานการณ์ ซึ่งหมายความว่าวิธี adjusted standard deviation (ASD) และวิธี adjusted range (AR) มีแนวโน้มเป็นตัวประมาณที่ไม่เอนเอียงของค่าเบี่ยงเบนมาตรฐานประชากร ซึ่งสอดคล้องกับการศึกษาของ Montgomery [3] อย่างไรก็ตาม เมื่อข้อมูลมีค่านอกเกณฑ์ปลอมปนร้อยละ 10 และ 20 พบว่าวิธี sample standard deviation (SD) ไม่เหมาะสมสำหรับการประมาณค่าเบี่ยงเบนมาตรฐานประชากร ในทุกระดับของค่าเบี่ยงเบนมาตรฐานประชากร (σ) และขนาดตัวอย่าง (n) ซึ่งสอดคล้องกับงานวิจัยของ Leys และ

คณะ [12]

5. สรุปผลการวิจัย

งานวิจัยนี้ใช้เกณฑ์การเปรียบเทียบประสิทธิภาพ 2 เกณฑ์ คือ ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ซึ่งวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปกติเมื่อข้อมูลมีค่านอกเกณฑ์ที่ให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ต่ำที่สุดในแต่ละสถานการณ์สามารถสรุปได้ดังตารางที่ 5

ตารางที่ 5 วิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรที่ให้ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ต่ำที่สุดในแต่ละสถานการณ์

ร้อยละของค่า นอกเกณฑ์	n	ค่าเบี่ยงเบนมาตรฐานประชากร (σ)									
		1		5		10		15		20	
		ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด
0	10	ASD	SD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
	20	ASD	SD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
	30	ASD	SD, ASD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
	50	ASD	SD, ASD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
	70	ASD	SD, ASD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
	100	ASD	SD, ASD	ASD	SD	ASD	SD	ASD	SD	ASD	SD
10	10	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	20	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	30	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	50	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	70	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	100	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
20	10	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	20	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	30	MAD	MAD	PSD	PSD	MAD	MAD	MAD	MAD	MAD	MAD
	50	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	70	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD
	100	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD	MAD

กรณีไม่มีค่านอกเกณฑ์ปลอมปน จากตารางที่ 5 เมื่อพิจารณาค่าความเอนเอียงสัมบูรณ์ (ABS) จะได้ว่าวิธี adjusted standard deviation (ASD) เป็นวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรที่มีประสิทธิภาพมากที่สุดในทุกสถานการณ์ แต่เมื่อพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) จะพบว่าวิธี sample standard deviation (SD) มีประสิทธิภาพมากที่สุดในทุกสถานการณ์ และเมื่อค่าเบี่ยงเบนมาตรฐานประชากร (σ) เท่ากับ 1 และขนาดตัวอย่าง (n) เท่ากับ 30, 50, 70, 100 วิธี adjusted standard deviation (ASD) มีประสิทธิภาพเท่ากับวิธี sample standard deviation (SD) เนื่องจากมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากัน ทั้งนี้ในกรณีที่ไม่มีค่านอกเกณฑ์ปลอมปนนั้นการตัดสินใจเลือกใช้ตัวประมาณใดขึ้นอยู่กับสถานการณ์ที่จะนำไปใช้ เช่นกรณีที่ขนาดตัวอย่างน้อยโดยเฉพาะอย่างยิ่งขนาดตัวอย่างไม่เกิน 10 อาจเลือกใช้วิธี adjusted standard deviation (ASD) เนื่องจากให้ค่าความเอนเอียงสัมบูรณ์ต่ำที่สุด และเมื่อพิจารณาข้อมูลจากตารางที่ 2 ประกอบพบว่าค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณโดยวิธี adjusted standard deviation (ASD) ให้ค่าแตกต่างจากการประมาณโดยวิธี sample standard deviation (SD) น้อยมาก (ไม่เกิน 1) ในทุกระดับของค่าเบี่ยงเบนมาตรฐานประชากร (σ) อย่างไรก็ตาม ในกรณีขนาดตัวอย่าง (n) มากกว่า 10 ในทุกระดับค่าเบี่ยงเบนมาตรฐานประชากร อาจเลือกใช้วิธี adjusted standard deviation (ASD) ได้เช่นกันเนื่องจากให้ค่าความเอนเอียงสัมบูรณ์ต่ำที่สุดในสถานการณ์ดังกล่าวข้างต้น และให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (พิจารณาจากตารางที่ 2) ใกล้เคียงกับวิธี sample standard deviation (SD) ในสถานการณ์ดังกล่าวนี้ด้วย หรืออาจเลือกใช้วิธี sample standard deviation (SD) สำหรับกรณีที่

ขนาดตัวอย่างใหญ่ก็ได้ แต่ถ้าขนาดตัวอย่างไม่เกิน 10 วิธี sample standard deviation (SD) อาจไม่เหมาะสมเนื่องจากให้ค่าความเอนเอียงสัมบูรณ์ค่อนข้างมากโดยการพิจารณาจากตารางที่ 2 ประกอบการตัดสินใจเลือกวิธีการประมาณค่าที่เหมาะสมในกรณีไม่มีค่านอกเกณฑ์ปลอมปนดังกล่าวนี้

กรณีมีค่านอกเกณฑ์ปลอมปนร้อยละ 10 และ 20 พบว่าวิธี mean absolute deviation (MAD) เป็นวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรที่มีประสิทธิภาพมากที่สุดในเกือบทุกสถานการณ์เนื่องจากเป็นวิธีที่ให้ทั้งค่าความเอนเอียงสัมบูรณ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด ยกเว้นกรณีที่ค่านอกเกณฑ์ปลอมปนร้อยละ 20 ขนาดตัวอย่าง (n) เท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานประชากร (σ) เท่ากับ 5 การประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธี percentile tab-standard deviation (PSD) จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพมากที่สุด เนื่องจากเป็นวิธีที่ให้ทั้งค่าความเอนเอียงสัมบูรณ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด

6. ข้อเสนอแนะ

ควรมีการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรวิธีอื่น ๆ อีกในกรณีที่ข้อมูลมีค่านอกเกณฑ์ปลอมปน เช่น วิธี median absolute deviation [13] หรือใช้วิธี mean absolute deviation (MAD) ตามงานวิจัยนี้แต่เปลี่ยนเป็นการหาค่าเฉลี่ยของผลต่างระหว่างค่าของข้อมูลกับค่ามัธยฐานแทน [14] เพื่อเป็นแนวทางในการเลือกใช้วิธีการประมาณค่าเบี่ยงเบนมาตรฐานประชากรสำหรับข้อมูลที่มีการแจกแจงปกติในแต่ละสถานการณ์ตามความต้องการของผู้วิจัย และได้ผลลัพธ์ที่มีประสิทธิภาพมากขึ้นในกรณีที่ข้อมูลมีค่านอกเกณฑ์ปลอมปน

7. กิตติกรรมประกาศ

ขอขอบคุณ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ที่สนับสนุนค่าใช้จ่ายในการทำวิจัยครั้งนี้

8. รายการอ้างอิง

- [1] Anscombe, F.J., 1960, Rejection of outliers, J. Amer. Stat. Assoc. Amer. Soc. Qual. 2: 123-147.
- [2] Baragona, R., Battaglia, F. and Poli, I., 2011, Outliers, Available Source: http://link.springer.com/chapter/10.1007/978-3-642-16218-3_6?no-access=true, October 6, 2017.
- [3] Montgomery, D.C., 2012, Introduction to Statistical Quality Control, 7th Ed., John Wiley, New York.
- [4] Woodall, W.H. and Montgomery, D.C., 2000-2001, Using ranges to estimate variability, Qual. Eng. 13(2): 211-217.
- [5] Barnett, V. and Lewis, T., 1995, Outlier in Statistical Data, 3th Ed., John Wiley, New York.
- [6] Ferguson, T.S., 1961, On the rejection of outliers, pp. 253-287, Neyman, J. (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, California.
- [7] Geary, R.C., 1935, The ratio of the mean deviation to the standard deviation as a test of normality, Biometrika 27: 310-332.
- [8] Chen, G., 1997, The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated, Statistica Sinica 7: 789-798.
- [9] Leon, H., 1960, Tables of range and studentized range, Ann. Math. Stat. 31(4): 1122-1147.
- [10] Hintze, J.L., 2008, Standard deviation estimator, Sample Size Software, NCSS, Kaysville.
- [11] Vardeman, S.B., 1999, A brief tutorial on the estimation of the process standard deviation, IIE Transact. 31: 503-507.
- [12] Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L., 2013, Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol. 49: 764-766.
- [13] Huber, P.J., 1981, Robust statistics, John Wiley, New York.
- [14] พจนานุกรมศัพท์สถิติศาสตร์ ฉบับราชบัณฑิตยสถาน, 2558, สำนักพิมพ์คณะรัฐมนตรีและราชกิจจานุเบกษา, กรุงเทพฯ.