

การเปรียบเทียบการทดสอบค่าเฉลี่ยหรือค่ากลางของสองประชากรที่มีความ
แปรปรวนเท่ากัน ในกรณีของการแจกแจงปกติและปกติปลอมปน
A Comparison for Testing of Means or Medians of two Population
with Equal Variance in Case of Normal Distribution and
Contaminated Normal Distribution

อัชฌา อระวีพร^{1*} ศุภวิชญ์ สีสลาพิระพันธ์¹ ชุตินา โพชรา¹ สุพัฒน์ เทียนรุ่งโรจน์¹ และ อนุวัตร นามบุญศรี¹
Autcha Araveeporn¹ Supavit Leelapeeraphan¹ Chutima Phochara¹ Supat Thianrungsri¹
and Anuwat Namboonsri¹

¹ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
วันที่ส่ง : 24 มกราคม 2562 วันที่แก้ไข : 30 พฤษภาคม 2562 วันที่ตอบรับ : 20 ธันวาคม 2562

บทคัดย่อ

การวิจัยนี้เป็นการวิจัยเชิงจำลองมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบค่าเฉลี่ยหรือค่ากลางของ 2 ประชากรที่มีความแปรปรวนเท่ากันสำหรับการแจกแจงปกติและปกติปลอมปนโดยใช้ การทดสอบที การทดสอบซี วิธีบูทสเตรป และการทดสอบแวน-เดอร์ วาเดนท์ โดยศึกษาจากข้อมูลที่สุ่มมาจากประชากรที่ การแจกแจงปกติและปกติปลอมปน กำหนดขนาดตัวอย่าง (n_1, n_2) เท่ากันคือ เท่ากับ (5,5) (20,20) (50,50) และไม่เท่ากันคือ (5,10) (20,25) (50,70) ในการคำนวณความน่าจะเป็นของความผิดพลาดแบบที่ 1 กำหนดค่าเฉลี่ยของประชากรทั้งสองกลุ่ม (μ_1, μ_2) เท่ากับ (0,0) และการคำนวณกำลังการทดสอบ กำหนดค่าเฉลี่ยของประชากรทั้งสองกลุ่ม (μ_1, μ_2) เท่ากับ (0,2) ที่ความแปรปรวน (σ^2) เท่ากันคือ 4 และ 16 ตามลำดับ กำหนดระดับนัยสำคัญ 2 ระดับคือ 0.01 และ 0.05 โดยใช้โปรแกรมอาร์ในการจำลอง และวิเคราะห์ข้อมูล ทำการจำลองข้อมูลซ้ำ 1,000 รอบในแต่ละสถานการณ์ จากผลการวิจัยพบว่า เมื่อ ความแปรปรวนเท่ากัน ส่วนใหญ่สถิติอิงพารามิเตอร์ คือ การทดสอบซี และ สถิติไม่อิงพารามิเตอร์ คือ การทดสอบแวน-เดอร์ วาเดนท์ มีความสามารถในการควบคุมค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 นอกจากนี้ยังมีกำลังการทดสอบสูงที่สุดอีกด้วย

คำสำคัญ : การแจกแจงปกติปลอมปน การทดสอบที การทดสอบซี วิธีบูทสเตรป และการทดสอบแวน-เดอร์ วาเดนท์

*ที่อยู่ติดต่อ. E-mail address : autcha.ar@kmitl.ac.th

Abstract

This research is a simulating research that aimed to study and to compare for testing of two equal variance population means or medians in case of normal distribution and contaminated normal distribution by T-test, Z-test, Bootstrap method, and Van der Waerden test. Data is generated from two populations based on a normal distribution and contaminated normal distribution. The sample sizes (n_1, n_2) are equal as (5,5), (20,20) (50,50) and unequal as (5,10), (20,25) and (50,70). For computing probability of type I error, the population mean (μ_1, μ_2) are equal as (0,0), and unequal as (0,2) for computing power of a test. The population variances (σ^2) is defined by 4 and 16. The significant levels are considered on two levels at 0.01 and 0.05. R program is used for simulation and data analysis with 1,000 times for each situation. The results are found that Z-test of parametric statistics and Van der Waerden test of nonparametric statistics can control probability of type I error, and furthermore the highest power of a test.

Keywords : contaminated normal distribution, T-test, Z-test, bootstrap method, Van der Waerden test

1. บทนำ

ในการทดสอบสมมติฐานเพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยหรือค่ากลางของประชากร 2 กลุ่มที่เป็นอิสระกัน การทดสอบที่นิยมใช้ในการวิเคราะห์ข้อมูลเมื่อข้อมูลมีการแจกแจงปกติ ได้แก่ สถิติอิงพารามิเตอร์ เช่น การทดสอบซี (Z-test) การทดสอบที (T-test) โดยพิจารณาจากค่าความแปรปรวนของประชากรทั้ง 2 กลุ่มว่าเท่ากันหรือไม่เท่ากัน และเมื่อข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของสถิติอิงพารามิเตอร์จึงสามารถเลือกใช้สถิติไม่อิงพารามิเตอร์ เช่นวิธีบูทสเตรป (Bootstrap method) และ การทดสอบแวน-เดอร์ วาเดน (Van der Waerden) [1] เป็นต้น

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า มนตรี [2] ศึกษาประสิทธิภาพการทดสอบอิงพารามิเตอร์ และการทดสอบไม่อิงพารามิเตอร์ พบว่า การทดสอบที และ แวน-เดอร์ วาเดน สามารถควบคุมความผิดพลาดแบบที่ 1 ได้และมีกำลังทดสอบสูงสุดเมื่อตัวอย่างทั้ง 2 มีขนาดเล็ก มนตรี [3] ศึกษาความแกร่งและกำลังการทดสอบของการทดสอบอิงพารามิเตอร์และการทดสอบไม่อิงพารามิเตอร์ การศึกษาพบว่าเมื่อประชากรมีการแจกแจงปกติ การทดสอบที สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ และมีกำลังทดสอบสูงสุด กัญญ์พิชญา [4] ศึกษาการเปรียบเทียบประสิทธิภาพของวิธีประมาณการแจกแจงระหว่างวิธีแจ๊คไนฟ์กับวิธีบูทสเตรป โดยประมาณพารามิเตอร์ค่าเฉลี่ย ความแปรปรวน ความเบ้และความโด่ง ทั้งการประมาณค่าแบบจุดและแบบช่วง สำหรับการแจกแจงปกติปลอมปน ผลการวิจัยสรุปได้ว่า กรณีการ

ประมาณค่าแบบจุด พบว่า วิธีบูทสเตรป จะมีประสิทธิภาพดีกว่าวิธีแจ๊คไนฟ์ กรณีการประมาณค่าแบบช่วง พบว่าทุกค่าพารามิเตอร์ในทุกๆการแจกแจง วิธีบูทสเตรป มีประสิทธิภาพดีกว่าวิธีแจ๊คไนฟ์

ในงานวิจัยนี้สนใจกรณีข้อมูลชุดแรกถูกสุ่มมาจากข้อมูลที่มีการแจกแจงปกติ (Normal Distribution) โดยเป็นการแจกแจงของตัวแปรสุ่มชนิดต่อเนื่องที่สำคัญที่สุดและสามารถใช้ได้กับข้อมูลสถานการณ์ และปรากฏการณ์ต่างๆ มากมายเช่น รายได้ของประชากร อายุการใช้งานของสินค้า ปริมาณน้ำตาลในเลือด ลักษณะของเส้นโค้งของการแจกแจงปกติเป็นรูประฆังคว่ำ มีจุดศูนย์กลางที่ค่าเฉลี่ยของประชากรและสมมาตรรอบค่าเฉลี่ยซึ่งทำให้ค่าเฉลี่ย มัชฌิมและฐานนิยมอยู่ที่จุดเดียวกัน

สำหรับข้อมูลชุดที่สองถูกสุ่มมาจากข้อมูลที่มีการแจกแจงปกติปลอมปน (Contaminated Normal Distribution) เป็นการแจกแจงผสมระหว่าง 2 การแจกแจงปกติ โดยการแจกแจงหนึ่งมีข้อมูลเป็นรูปแบบของการแจกแจงปกติที่ค่าเฉลี่ยและความแปรปรวนปกติ และอีกกลุ่มหนึ่งจะถูกสุ่มมาจากข้อมูลที่มีค่าเฉลี่ยปกติแต่มีความแปรปรวนสูงมากโดยถ่วงน้ำหนักด้วยค่าความน่าจะเป็น ซึ่งเรียกว่าค่าสัดส่วนการปลอมปน ซึ่งค่าที่ได้จะแสดงค่านอกเกณฑ์ (outlier)

จากที่ได้กล่าวมาข้างต้นนี้ ผู้วิจัยจึงสนใจที่จะศึกษาการเปรียบเทียบการทดสอบค่าเฉลี่ยหรือค่ากลางของ 2 ประชากรที่เป็นอิสระกัน เมื่อใช้การทดสอบที การทดสอบซี วิธีบูทสเตรป และการทดสอบแวน-เดอร์ วาเดนส์ เพื่อหาว่าการทดสอบใดมีกำลังการทดสอบสูงสุดสำหรับข้อมูลที่มีการแจกแจงการแจกแจงปกติในกลุ่มที่ 1 และปกติปลอมปนในกลุ่มที่ 2 โดยใช้โปรแกรมอาร์ (R) ในการวิเคราะห์ข้อมูล

2. วิธีการวิจัย

ในการวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาค่าความผิดพลาดแบบที่ 1 และกำลังการทดสอบของตัวการทดสอบของประชากร 2 กลุ่มในกรณีค่าเฉลี่ยเท่ากันและความแปรปรวนเท่ากัน กรณีที่ค่าเฉลี่ยไม่เท่ากันและความแปรปรวนเท่ากัน ใช้การทดสอบที (T-test) การทดสอบซี (Z-test) วิธีบูท-สเตรป (Bootstrap) และการทดสอบแวน-เดอร์ วาเดนส์ (Van der Waerden- test)

2.1 การวางแผนการวิจัย

ในการวิจัยครั้งนี้กำหนดสถานการณ์ในการศึกษาเปรียบเทียบดังนี้

2.1.1 กำหนดให้ประชากรทั้ง 2 ประชากรที่เป็นอิสระกัน

2.1.2 ศึกษาในกรณีที่ตัวอย่างเท่ากันและไม่เท่ากัน ดังนี้

ขนาดตัวอย่าง (n_1, n_2)	
เท่ากัน ($n_1 = n_2$)	ไม่เท่ากัน ($n_1 \neq n_2$)
(5,5), (20,20) , (50,50)	(5,10), (20,25) , (50,70)

ขนาดตัวอย่างที่ใช้ในการศึกษาแบ่งเป็น ตัวอย่างขนาดเล็ก ขนาดกลาง และขนาดใหญ่

2.1.3 กำหนดให้ค่าความแปรปรวนทั้งสองกลุ่มเท่ากัน

กรณีข้อมูลที่สุ่มมาจากการแจกแจงปกติ (Normal Distribution) ด้วยพารามิเตอร์ μ และ σ^2 หรือเขียนได้ว่า $N(\mu, \sigma^2)$ โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

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

กรณีข้อมูลที่สุ่มมาจากการแจกแจงปกติปลอมปน (Contaminated Normal Distribution) ด้วยพารามิเตอร์ μ และ σ^2 ที่มีสัดส่วนการปลอมปนของค่านอกกลุ่ม (p) เท่ากับ 0.05 และ 0.1 โดยสร้างมาจากการแจกแจงทวินาม ค่านอกกลุ่มจากการแจกแจงปกติที่มีค่าสเกลแฟคเตอร์ (C) เท่ากับ 5 และ 10 โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(x) = (1-p)N(\mu, \sigma^2) + (p)N(\mu, c^2 \sigma^2)$$

กรณีค่าเฉลี่ยเท่ากันและความแปรปรวนเท่ากัน มีพารามิเตอร์ (μ, σ^2) คือ (0,4) และ (0,16) ทั้ง 2 ประชากร โดยมีสัดส่วนการปลอมปนในกลุ่มที่ 2 คือ 0.05 และ 0.1 สามารถสรุปได้ดังนี้

ประชากรกลุ่มที่ 1	ประชากรกลุ่มที่ 2
$N(0, 4)$	$(0.95)N(0, 4) + (0.05)N(0, (5^2 \times 4))$
	$(0.95)N(0, 4) + (0.05)N(0, (10^2 \times 4))$
	$(0.9)N(0, 4) + (0.1)N(0, (5^2 \times 4))$
	$(0.9)N(0, 4) + (0.1)N(0, (10^2 \times 4))$
$N(0, 16)$	$(0.95)N(0, 16) + (0.05)N(0, (5^2 \times 16))$
	$(0.95)N(0, 16) + (0.05)N(0, (10^2 \times 16))$
	$(0.9)N(0, 16) + (0.1)N(0, (5^2 \times 16))$
	$(0.9)N(0, 16) + (0.1)N(0, (10^2 \times 16))$

กรณีค่าเฉลี่ยไม่เท่ากันและความแปรปรวนเท่ากัน มีพารามิเตอร์ (μ, σ^2) ของประชากรกลุ่มแรก คือ (0,4) และ (0,16) ของประชากรกลุ่มสอง คือ (2,4) และ (2,16) โดยมีสัดส่วนการปลอมปนในกลุ่มที่ 2 คือ 0.05 และ 0.1 สามารถสรุปได้ดังนี้

ประชากรกลุ่มที่ 1	ประชากรกลุ่มที่ 2
$N(0, 4)$	$(0.95)N(2, 4) + (0.05)N(2, (5^2 \times 4))$
	$(0.95)N(2, 4) + (0.05)N(2, (10^2 \times 4))$
	$(0.9)N(2, 4) + (0.1)N(2, (5^2 \times 4))$
	$(0.9)N(2, 4) + (0.1)N(2, (10^2 \times 4))$
$N(0, 16)$	$(0.95)N(2, 16) + (0.05)N(2, (5^2 \times 16))$
	$(0.95)N(2, 16) + (0.05)N(2, (10^2 \times 16))$
	$(0.9)N(2, 16) + (0.1)N(2, (5^2 \times 16))$
	$(0.9)N(2, 16) + (0.1)N(2, (10^2 \times 16))$

2.1.4 ในทุกขนาดตัวอย่างและทุกกรณี กำหนดระดับนัยสำคัญ 2 ระดับ คือ 0.01 และ 0.05

2.1.5 เปรียบเทียบกำลังการทดสอบของตัวการทดสอบและความน่าจะเป็นของความผิดพลาดแบบที่ 1 เพื่อหาตัวการทดสอบที่ดีที่สุดในแต่ละสถานการณ์

2.1.6 ใช้โปรแกรม R เวอร์ชัน 3.4.3 ในการจำลองและวิเคราะห์ข้อมูล

2.2 การทดสอบความเท่ากันของค่าเฉลี่ยหรือค่ากลางของประชากร

ในการทดสอบสมมติฐานเพื่อเปรียบเทียบค่าเฉลี่ยหรือค่ากลางระหว่างประชากร 2 กลุ่มที่เป็นอิสระกันสถิติอิงพารามิเตอร์ที่ใช้ทดสอบสำหรับสมมติฐานดังกล่าวในการศึกษาครั้งนี้มี 2 การทดสอบ คือ การทดสอบซีและการทดสอบที

สมมติฐานในการทดสอบ

$$\text{สมมติฐานหลัก} \quad H_0: \mu_1 = \mu_2$$

$$\text{สมมติฐานรอง} \quad H_1: \mu_1 \neq \mu_2$$

โดยที่ μ_1 และ μ_2 แทน ค่าเฉลี่ยของประชากรกลุ่มที่ 1 และ 2 ตามลำดับ

2.2.1 การทดสอบที (T-test) [5]

เป็นการทดสอบอิงพารามิเตอร์ที่ใช้ทดสอบความแตกต่างของค่าเฉลี่ยระหว่างประชากร 2 กลุ่มที่เป็นอิสระกันและความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2$) จะสามารถใช้การทดสอบที่ได้ก็ต่อเมื่อข้อมูลเป็นไปตามข้อกำหนดเบื้องต้น คือ ประชากร 2 กลุ่มอิสระต่อกัน ประชากร 2 กลุ่มมีการแจกแจงปกติ ตัวอย่างทั้ง 2 กลุ่มมีขนาดเล็ก ($n_1, n_2 < 30$)

$$\text{การทดสอบ } t_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{โดยที่ } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

เมื่อ $\bar{X}_1, \bar{X}_2$ คือ ค่าเฉลี่ยของตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

S_1^2, S_2^2 คือ ค่าความแปรปรวนของตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

S_p^2 คือ ความแปรปรวนร่วม

n_1, n_2 คือ จำนวนตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

อาณาเขตวิกฤต : $t_{cal} > t_{\frac{\alpha}{2}; n_1 + n_2 - 2}$ หรือ $t_{cal} < -t_{\frac{\alpha}{2}; n_1 + n_2 - 2}$

เมื่อ $t_{\frac{\alpha}{2}; n_1 + n_2 - 2}$ คือค่าวิกฤตที่เปิดได้จากตารางการแจกแจงที ที่องศาเสรี $n_1 + n_2 - 2$

2.2.2 การทดสอบซี (Z-test) [5]

เป็นการทดสอบอิงพารามิเตอร์ที่ใช้ทดสอบความแตกต่างของค่าเฉลี่ยระหว่างประชากร 2 กลุ่มที่เป็นอิสระกัน จะสามารถใช้การทดสอบซีได้ก็ต่อเมื่อข้อมูลเป็นไปตามข้อกำหนดเบื้องต้น คือ ประชากร 2 กลุ่มอิสระต่อกัน ประชากร 2 กลุ่มมีการแจกแจงปกติ ตัวอย่างทั้ง 2 กลุ่มมีขนาดใหญ่ ($n_1, n_2 \geq 30$) โดยใช้ทฤษฎีบทขีดจำกัดส่วนกลาง (Central Limit Theorem)

$$\text{การทดสอบ } Z_{cal} \approx \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

เมื่อ $\bar{X}_1, \bar{X}_2$ คือ ค่าเฉลี่ยของตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

S_1^2, S_2^2 คือ ค่าความแปรปรวนของตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

n_1, n_2 คือ จำนวนตัวอย่างของกลุ่มที่ 1 และ 2 ตามลำดับ

อาณาเขตวิกฤต : $Z_{cal} > Z_{\frac{\alpha}{2}}$ หรือ $Z_{cal} < -Z_{\frac{\alpha}{2}}$

เมื่อ $Z_{\frac{\alpha}{2}}$ คือค่าวิกฤตที่เปิดได้จากตารางการแจกแจงปกติมาตรฐาน

สถิติไม่อิงพารามิเตอร์ที่ใช้ทดสอบสำหรับสมมติฐานดังกล่าวในการศึกษาค้างนี้มี 2 การทดสอบคือ วิธีบูทสเตรป และการทดสอบแวน-เดอร์ วาเดนส์
สมมติฐานในการทดสอบ

$$\text{สมมติฐานหลัก} \quad H_0 : M_1 = M_2$$

$$\text{สมมติฐานรอง} \quad H_1 : M_1 \neq M_2$$

โดยที่ M_1 และ M_2 แทน ค่ามัธยฐานของประชากรที่ 1 และ 2 ตามลำดับ

2.2.3 วิธีบูทสแตรป (Bootstrap method) [6]

เป็นการกำหนดตัวอย่างโดยสุ่มแบบคืนที่จำนวน n ตัวจากตัวอย่างสุ่ม $X_1, X_2, \dots, X_n$ ที่มีการแจกแจงปกติและการแจกแจงปกติปลอมปน ซึ่งจะได้ตัวอย่างสุ่มชุดใหม่ คือ $X_1^*, X_2^*, \dots, X_n^*$ ซึ่งเรียกว่าตัวอย่างบูทสแตรป และคำนวณค่า $\bar{X}_j^*$ ของตัวอย่างบูทสแตรป เมื่อ $j = 1, \dots, B$ โดยที่ B เป็นจำนวนการทำซ้ำ

$$\text{การคำนวณค่าวิธีบูทสแตรปคือ} \quad \hat{D}_j = \bar{X}_{j(1)}^* - \bar{X}_{j(2)}^*$$

$$\hat{\eta}_B = \frac{1}{B} \sum_{j=1}^B \hat{D}_j$$

$$SE_{\hat{\eta}_B} = \sqrt{\frac{1}{B-1} \sum_{j=1}^B (\hat{D}_j - \hat{\eta}_B)^2}$$

โดย B คือ จำนวนครั้งการทำซ้ำแบบบูทสแตรป

$\hat{\eta}_B$ คือ ค่าประมาณของผลต่างของค่าประมาณชุดใหม่ของประชากรแต่ละกลุ่ม

$\hat{D}_j$ คือ ผลต่างของค่าประมาณชุดใหม่ของประชากรแต่ละกลุ่ม

$\bar{X}_{j(1)}^*, \bar{X}_{j(2)}^*$ คือ ค่าประมาณชุดใหม่ของประชากรกลุ่มที่ 1, 2 ที่ได้จากการสุ่มแบบแทนที่

$$= \sum_{i=1}^n \frac{X_i^*}{n} ; i \text{ คือ ตัวอย่างบูทสแตรปชุดที่ } i$$

จากทฤษฎีบทขีดจำกัดส่วนกลาง (Central Limit Theorem)

$$\text{สถิติทดสอบ } Z_{cal} \approx \frac{\hat{\eta}_B}{SE_{\hat{\eta}_B}}$$

$$\text{อาณาเขตวิกฤต : } Z_{cal} > Z_{\frac{\alpha}{2}} \text{ หรือ } Z_{cal} < -Z_{\frac{\alpha}{2}}$$

2.2.4 การทดสอบแวน-เดอร์ วาเดนส์ (Van der Waerden test)

การทดสอบแวน-เดอร์ วาเดนส์ เป็นตัวสถิติ ทดสอบที่ถูกต้องค้นพบโดยนักคณิตศาสตร์ชาวดัตช์ ชื่อว่า Bartel Leendert van der Waerden [1] สามารถใช้ทดสอบความแตกต่างระหว่างค่ากลาง 2 กลุ่ม ซึ่งมีรายละเอียดการคำนวณ ดังนี้

$$T_1 = \frac{1}{S^2} \sum_{i=1}^k n_i \bar{A}_i^2$$

เมื่อ $A_{ij} = \phi^{-1} \left[\frac{R(X_{ij})}{N+1} \right]$ คือ คะแนนมาตรฐานตัวที่ j กลุ่มที่ i

เมื่อ $N = n_1 + n_2$

$$A_{ij} = \phi^{-1} \left[\frac{R(X_{ij})}{N+1} \right], i = 1, 2, \dots, k \text{ คือ ค่าเฉลี่ยของคะแนนมาตรฐานกลุ่มที่ } i$$

โดยที่ $R(X_{ij})$ คือลำดับของค่าสังเกต X_{ij} และ ϕ^{-1} คือค่าควอไทล์ฟังก์ชันของการแจกแจงปกติมาตรฐาน

$$S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{n_i} A_{ij}^2 \text{ คือ ค่าความแปรปรวนของคะแนนมาตรฐาน}$$

อาณาเขตวิกฤต : $T_1 > \chi_{\alpha,1}^2$ โดย $\chi_{\alpha,1}^2$ คือค่าวิกฤตที่เปิดได้จากตารางแจกแจงไคกำลังสอง

2.3 ขั้นตอนการดำเนินงาน

2.3.1 จำลองข้อมูลและสุ่มตัวอย่างที่ใช้ในการวิจัยด้วยโปรแกรม R กำหนดค่าเฉลี่ยเท่ากัน ค่าความแปรปรวนเท่ากันระหว่างประชากร 2 กลุ่ม โดยมี การแจกแจง ความแปรปรวน และขนาดตัวอย่าง เป็นไปตามขอบเขตการวิจัย

2.3.2 ทำการทดสอบเปรียบเทียบค่าเฉลี่ยหรือค่ากลางใน 4 การทดสอบ โดยวิธีการบูทสแตรปทำการวนซ้ำ 1,000 รอบ และบันทึกจำนวนครั้งที่ปฏิเสธสมมติฐานหลัก ทำซ้ำจนครบ 1,000 ครั้ง

2.3.3 หาความน่าจะเป็นของความผิดพลาดแบบที่ 1 โดยนำจำนวนครั้งที่ปฏิเสธสมมติฐานหลักหารด้วย 1,000

2.3.4 ทำการเปรียบเทียบความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของแต่ละการทดสอบกับเกณฑ์ของ Bradlay [7] โดยการทดสอบจะสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ก็ต่อเมื่อ α มีค่าอยู่ในช่วง $(0.5\alpha, 1.5\alpha)$ ในการศึกษาครั้งนี้ได้กำหนดระดับนัยสำคัญ (α) 2 ระดับคือ 0.01 และ 0.05 ช่วงของความน่าจะเป็นของความผิดพลาดแบบที่ 1 คือ $[0.005, 0.015]$ และ $[0.025, 0.075]$ ตามลำดับ ถ้าความน่าจะเป็นของความผิดพลาดแบบที่ 1 จากการทดลองอยู่ในช่วงที่กำหนด จะสรุปได้ว่าการทดสอบทั้งหมดนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้

2.3.5 จำลองข้อมูลและสุ่มตัวอย่าง โดยกำหนดค่าเฉลี่ยไม่เท่ากัน ค่าความแปรปรวนเท่ากันระหว่างประชากร 2 กลุ่ม โดยมี การแจกแจง ความแปรปรวน และขนาดตัวอย่าง เป็นไปตามขอบเขตการวิจัย

2.3.6 หากำลังการทดสอบจากการทดสอบ โดยนำจำนวนครั้งที่ปฏิเสธสมมติฐานหลักหารด้วย 1,000

2.3.7 เปรียบเทียบกำลังการทดสอบของการทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ โดยการทดสอบที่มีกำลังการทดสอบที่สูงที่สุด จะถือว่าเป็นการทดสอบที่มีประสิทธิภาพดีที่สุด

3. ผลการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงจำลอง เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของตัวการทดสอบ 4 ตัว ได้แก่ การทดสอบที การทดสอบซี วิธีบูตสเตรป และการทดสอบแวน-เดอร์ วาเดนส์ สำหรับ 2 ประชากรที่เป็นอิสระกัน ในกรณีข้อมูลสุ่มจากประชากรที่มีการแจกแจงปกติในกลุ่มที่ 1 และปกติปโลมปน ในกลุ่มที่ 2 ซึ่งผลที่ได้จากการวิจัยสามารถสรุปได้เป็น 2 ส่วนใหญ่ๆ คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 สรุปผลได้ดังตารางที่ 1 และ 2 และสถิติทดสอบที่ให้กำลังการทดสอบสูงสุดในตารางที่ 3 และ 4 โดยกำหนดให้

T หมายถึง การทดสอบที

Z หมายถึง การทดสอบซี

BT หมายถึง วิธีบูตสเตรป

VW หมายถึง การทดสอบแวน-เดอร์ วาเดนส์

3.1 ความน่าจะเป็นของความผิดพลาดแบบที่ 1

พิจารณาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 กรณีข้อมูลสุ่มจากประชากรที่มีการแจกแจงปกติในตัวอย่างกลุ่มที่ 1 และปกติปโลมปนในตัวอย่างกลุ่มที่ 2 ตามเกณฑ์ของ Bradley สรุปผลได้ดังตารางที่ 1 และ 2

ตารางที่ 1. ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่ระดับนัยสำคัญ 0.01

ความแปรปรวน (σ^2)	p	c	การทดสอบ	ขนาดตัวอย่าง (n_1, n_2)					
				เท่ากัน			ไม่เท่ากัน		
				(5,5)	(20,20)	(50,50)	(5,10)	(20,25)	(50,70)
4	0.05	5	T	0.003	0.002	0.004	0.004	0.002	0.004
			Z	0.032	0.009*	0.009*	0.029	0.011*	0.014*
			BT	0.043	0.011*	0.008*	0.037	0.011*	0.013*
			VW	0.000	0.004	0.011*	0.007*	0.006*	0.011*
		10	T	0.006*	0.003	0.003	0.002	0.001	0.000
			Z	0.039	0.008*	0.006*	0.027	0.004	0.007*
			BT	0.063	0.009*	0.007*	0.036	0.006*	0.007*
			VW	0.000	0.006*	0.012*	0.008*	0.003	0.005*
	0.1	5	T	0.003	0.001	0.005*	0.003	0.003	0.003
			Z	0.027	0.007*	0.010*	0.030	0.008*	0.007*
			BT	0.043	0.010*	0.013*	0.039	0.009*	0.010*
			VW	0.000	0.007*	0.010*	0.005*	0.006*	0.007*
		10	T	0.008*	0.000	0.001	0.004	0.001	0.000
			Z	0.025	0.009*	0.001	0.021	0.005*	0.003
			BT	0.045	0.009*	0.001	0.026	0.008*	0.002
			VW	0.000	0.008*	0.008*	0.004	0.006*	0.002
16	0.05	5	T	0.001	0.005*	0.002	0.004	0.001	0.002
			Z	0.029	0.014*	0.009*	0.028	0.005*	0.008*
			BT	0.042	0.017	0.012*	0.038	0.005*	0.009*
			VW	0.000	0.011*	0.009*	0.005*	0.002	0.007*
		10	T	0.006*	0.006*	0.005*	0.002	0.001	0.002
			Z	0.019	0.009*	0.004	0.022	0.005*	0.006*
			BT	0.034	0.011*	0.004	0.036	0.006*	0.007*
			VW	0.000	0.008*	0.006*	0.004	0.007*	0.008*
	0.1	5	T	0.004	0.002	0.003	0.003	0.000	0.000
			Z	0.032	0.006*	0.008*	0.023	0.01*	0.008*
			BT	0.048	0.011*	0.01*	0.035	0.011*	0.007*
			VW	0.000	0.006*	0.009*	0.004	0.004	0.003
		10	T	0.002	0.000	0.001	0.001	0.001	0.001
			Z	0.022	0.006*	0.008*	0.016	0.002	0.007*
			BT	0.035	0.006*	0.01*	0.024	0.003	0.011*
			VW	0.000	0.009*	0.015*	0.002	0.004	0.009*

* หมายถึงความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ได้ ตามเกณฑ์ของ Bradley (ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตกอยู่ในช่วง [0.005, 0.015])

จากตารางที่ 1 จะได้รับการทดสอบที่สามารถในการควบคุมค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 โดยแบ่งเป็น สถิติอิงพารามิเตอร์ คือ การทดสอบซี สำหรับสถิติไม่อิงพารามิเตอร์ คือ การทดสอบแวน-เดอร์ วาเดนท์ รองลงมาคือวิธีบูทสเตรป

ตารางที่ 2. ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่ระดับนัยสำคัญ 0.05

ความแปรปรวน (σ^2)	p	c	การทดสอบ	ขนาดตัวอย่าง (n_1, n_2)					
				เท่ากัน			ไม่เท่ากัน		
				(5,5)	(20,20)	(50,50)	(5,10)	(20,25)	(50,70)
4	0.05	5	T	0.017	0.022	0.019	0.015	0.016	0.016
			Z	0.080	0.052*	0.044*	0.07*	0.051*	0.052*
			BT	0.100	0.054*	0.048*	0.099	0.059*	0.052*
			VW	0.057*	0.048*	0.044*	0.034*	0.051*	0.046*
		10	T	0.015	0.021	0.018	0.012	0.013	0.008
			Z	0.078	0.047*	0.048*	0.056*	0.043*	0.052*
			BT	0.116*	0.055*	0.050*	0.08	0.051*	0.054*
			VW	0.051*	0.047*	0.055*	0.027*	0.049*	0.047*
	0.1	5	T	0.020	0.023	0.020	0.010	0.012	0.011
			Z	0.071*	0.063*	0.051*	0.073*	0.045*	0.039*
			BT	0.105	0.071*	0.053*	0.097	0.048*	0.044*
			VW	0.049*	0.058*	0.043*	0.029*	0.034*	0.034*
		10	T	0.016	0.016	0.012	0.014	0.006	0.003
			Z	0.066*	0.045*	0.039*	0.050*	0.045*	0.039*
			BT	0.094	0.056*	0.045*	0.067*	0.053*	0.041*
			VW	0.054*	0.059*	0.044*	0.033*	0.046*	0.037*
16	0.05	5	T	0.023	0.012	0.020	0.016	0.015	0.019
			Z	0.067*	0.037*	0.045*	0.076	0.053*	0.050*
			BT	0.099	0.043*	0.048*	0.099	0.059*	0.053*
			VW	0.052*	0.035*	0.044*	0.043*	0.047*	0.051*
		10	T	0.019	0.014	0.021	0.016	0.016	0.012
			Z	0.078	0.037*	0.05*	0.066*	0.034*	0.045*
			BT	0.104	0.042*	0.054*	0.080	0.036*	0.049*
			VW	0.055*	0.047*	0.048*	0.035*	0.043*	0.039*
	0.1	5	T	0.018	0.017	0.027*	0.021	0.008	0.014
			Z	0.075*	0.055*	0.05*	0.075*	0.056*	0.056*
			BT	0.104	0.061*	0.054*	0.101	0.058*	0.059*
			VW	0.064*	0.056*	0.046*	0.042*	0.050*	0.047*
		10	T	0.015	0.009	0.013	0.007	0.009	0.004
			Z	0.061*	0.031*	0.035*	0.049*	0.046*	0.052*
			BT	0.093	0.043*	0.036*	0.066*	0.054*	0.052*
			VW	0.045*	0.044*	0.042*	0.033	0.047*	0.045*

* หมายถึงความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ได้ ตามเกณฑ์ของ Bradley (ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตกอยู่ในช่วง [0.025, 0.075])

จากตารางที่ 2 จะได้การทดสอบที่สามารถในการควบคุมค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 โดยแบ่งเป็น สถิติอิงพารามิเตอร์ คือ การทดสอบซี เกือบทุกกรณี สำหรับสถิติไม่อิงพารามิเตอร์ คือ การทดสอบแวน-เดอร์ วาเดนท์ รองลงมาคือวิธีบูทสเตรป

3.2 กำลังการทดสอบ

การสรุปผลกำลังการทดสอบของการทดสอบจะแสดงในกรณีที่การทดสอบสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เท่านั้น

ตารางที่ 3. การทดสอบที่มีกำลังการทดสอบสูงสุด ที่ระดับนัยสำคัญ 0.01

ความแปรปรวน (σ^2)	p	c	ขนาดตัวอย่าง (n_1, n_2)					
			5,5	20,20	50,50	5,10	20,25	50,70
4	0.05	5	-	BT	VW	VW	BT	VW
		10	T	VW	VW	VW	BT	VW
	0.1	5	-	VW	VW	VW	VW	VW
		10	T	VW	VW	-	VW	-
16	0.05	5	-	Z	VW	VW	BT	VW
		10	T	BT	VW	-	VW	VW
	0.1	5	-	BT	VW	-	BT	BT
		10	-	VW	VW	-	-	VW

จากตารางที่ 3 การทดสอบที่มีกำลังการทดสอบสูงสุดที่สุทธกรณข้อมูลสมจากประชากรที่มีการแจกแจงปกติและปกติปลอมปนที่ระดับนัยสำคัญ 0.01 ส่วนใหญ่ คือ การทดสอบแวน-เดอร์ วาเดนท ร่องลงมาคือวิธีบูทสเตรป

ตารางที่ 4. การทดสอบที่มีกำลังการทดสอบสูง ที่ระดับนัยสำคัญ 0.05

ความแปรปรวน (σ^2)	p	c	ขนาดตัวอย่าง (n_1, n_2)					
			5,5	20,20	50,50	5,10	20,25	50,70
4	0.05	5	VW	VW	VW	Z	VW	VW
		10	VW	VW	VW	Z	VW	VW
	0.1	5	Z	VW	VW	Z	VW	VW
		10	Z	VW	VW	BT	VW	VW
16	0.05	5	Z	BT	VW	VW	BT	VW
		10	VW	VW	VW	Z	VW	BT
	0.1	5	Z	VW	VW	VW	Z	VW
		10	Z	VW	VW	BT	VW	VW

จากตารางที่ 4 การทดสอบที่มีกำลังการทดสอบสูงที่สุทธกรณข้อมูลสมจากประชากรที่มีการแจกแจงปกติและปกติปลอมปนที่ระดับนัยสำคัญ 0.01 ส่วนใหญ่ คือ การทดสอบแวน-เดอร์ วาเดนท ร่องลงมาคือ การทดสอบซี และวิธีบูทสเตรป ตามลำดับ

4. สรุปผลการวิจัย

จากผลการวิเคราะห์ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ ของการทดสอบ 4 ตัว ได้แก่ การทดสอบที การทดสอบซี วิธีบูทสเตรป และการทดสอบ

แวน-เดอร์ วาเดนท์ จากการจำลองข้อมูลตามลักษณะต่างๆ ที่ได้กำหนดไว้ในขอบเขตการวิจัย ข้อมูลสุ่มมาจากประชากรที่มีการแจกแจงปกติและปกติปลอมปน เมื่อความแปรปรวนเท่ากันส่วนใหญ่สถิติอิงพารามิเตอร์ คือ การทดสอบซี และ สถิติไม่อิงพารามิเตอร์ คือ การทดสอบแวน-เดอร์ วาเดนท์ มีกำลังการทดสอบสูงที่สุดใน ซึ่งผลการวิจัยสอดคล้องกับงานวิจัยของ Bakker [8] ซึ่งการทดสอบที่มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับการทดสอบจากสถิติไม่อิงพารามิเตอร์

ดังนั้นในกรณีที่ต้องการทดสอบสมมติฐานของค่าเฉลี่ยและค่ากลางของ 2 ประชากร ถ้าทราบว่าข้อมูลมีค่านอกกลุ่ม ควรใช้สถิติไม่อิงพารามิเตอร์ เช่น การทดสอบแวน-เดอร์ วาเดนท์ ในการทดสอบสมมติฐาน จะทำให้การทดสอบมีความถูกต้องมากขึ้น

เอกสารอ้างอิง (References)

- [1] Van Der Waerden, B.L. 1952. Order Tests for the Two-sample Problem and Their Power. *Indagationes Mathematicae*, 14, 453-458.
- [2] มนตรี สังข์ทอง. 2556. การศึกษาประสิทธิภาพการทดสอบอิงพารามิเตอร์และการทดสอบไม่อิงพารามิเตอร์. *วารสารวิทยาศาสตร์ มข.*, 41(1), 266-238. [Montri Songthong. 2013. A Study of Efficiency of Parametric and Nonparametric Statistics in Testing of Central Difference Between Two Populations. *KKU Science Journal*, 41(1), 266-238. (in Thai)]
- [3] มนตรี สังข์ทอง. 2557. การศึกษาความแกร่งและกำลังการทดสอบของการทดสอบอิงพารามิเตอร์และการทดสอบไม่อิงพารามิเตอร์ มหาวิทยาลัยขอนแก่น. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 22(5), 605-609. [Montri Songthong. 2014. Robustness and Power of the Test of Parametric and Nonparametric Statistics in Testing of Central Difference Between Two Population for Likert-type Data 5 Point. *Thai Science and Technology Journal*, 22(5), 605-609. (in Thai)]
- [4] กัญญ์พิษฐา พุทธะไชยทัศน์. 2552. การเปรียบเทียบการประมาณการแจกแจงของข้อมูลสมมาตรและไม่สมมาตรด้วยวิธีแจ๊คไนฟ์ กับ วิธีบูทสเตรป. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต, สาขาสถิติ, จุฬาลงกรณ์มหาวิทยาลัย. [Kanpichaya Puttachaiyated. 2009. A comparison of estimation methods for symmetric and asymmetric distributed data by jackknifing method and bootstrapping method. M.Sc. Thesis, Statistics Program, Chulalongkorn University. (in Thai)]
- [5] ชูใจ คูหารัตนไชย. 2556. สถิติเบื้องต้น. พิมพ์ครั้งที่ 1, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. [Choojai Kuharatanachai. 2013.1st ed, Introduction to Statistics, Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang. (in Thai)]
- [6] ศรสวรรค์ บุญเพ็ญ, บุญอ้อม โฉมทิ และ อภิญญา หิรัญวงษ์. 2558. การเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ขนาดของการแจกแจงไวบูลล์แบบสองพารามิเตอร์. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 23(4), 579-587. [Sornsawan Boonpen, Boonorm

- Chomtee and Apinya Hirunwong. 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.(in Thai)]
- [7] Bradley, J.V. 1978. Robustness?. *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.
- [8] Bakker, M. 2014. Outlier Removal, Sum Scores, and the Inflation of the Type I Error Rate in Independent Sample t Test : The Power of Alternatives and Recommendations. *Psychol Methods*, 19(3), 409-427.