JASET KMUTNB Research Article

Journal of Applied Science and Emerging Technology (JASET)
Published by Faculty of Applied Science, KMUTNB, Thailand
doi: 10.14416/JASET.KMUTNB.2023.02.007
Vol. 22, No. 2 [2023]: e249712

การเปรียบเทียบประสิทธิภาพของการทดสอบอิงพารามิเตอร์และไม่อิงพารามิเตอร์สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ

Efficiency Comparison of Parametric and Non-Parametric Tests for Testing the Difference Between Central Values of More Than Two Independent Populations Using Counting Data

สุภาวดี สุวิธรรมา[1] จุฬารัตน์ ชุมนวล[2,3*] ศิริพร สมุทรวชิรวงษ์[4] และ สมพร เทพฉิม[5]

Suphawadee Suwithamma[1], Julalat Chumnaul[2,3*], Siriporn Samutwachirawong[4] and Somporn Thepchim[5]

[1]สาขาวิชาสถิติประยุกต์และวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครราชสีมา ประเทศไทย

[2]สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ประเทศไทย

[3]หน่วยวิจัยสถิติและการประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ประเทศไทย

[4]สาขาวิชาสถิติและการจัดการสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้ ประเทศไทย

[5]สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี ประเทศไทย

[1] Program of Applied Statistics and Data Science, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University, Thailand

[2]Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand

[3]Statistics and Applications Research Unit, Faculty of Science, Prince of Songkla University, Thailand

[4]Section of Statistics and Information Management, Faculty of Science, Maejo University, Thailand

[5]Program of Applied Statistics, Faculty of Science, Ubon Ratchathani Rajabhat University, Thailand

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการทดสอบอิงพารามิเตอร์และไม่อิงพารามิเตอร์ 5 วิธี ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) การทดสอบของเวลช์ (Welch's test) การทดสอบครัสคัล-วัลลิส (Kruskal-Wallis test) การทดสอบแวน เดอ แวร์เดน (Van der Waerden test) และการทดสอบมัธยฐาน (median test) สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับที่มีการแจกแจงทวินามและการแจกแจงปัวซง โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของการทดสอบ คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ ผลการศึกษาพบว่า กรณีที่ข้อมูลมีการแจกแจงทวินาม การทดสอบทั้ง 5 วิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และความแปรปรวนของข้อมูลแต่ละกลุ่มน้อยกว่า 4 และกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 15 และ (n, p) เท่ากับ (10, 0.1) และ (10, 0.7) สำหรับกรณีที่ข้อมูลมีการแจกแจงปัวซง การทดสอบทั้ง 5 วิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นการทดสอบของเวลช์ การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน ที่ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และ λ = 30 และเมื่อพิจารณากำลังการทดสอบพบว่า การวิเคราะห์ความแปรปรวนให้กำลังการทดสอบสูงกว่าวิธีอื่น ๆ เกือบทุกสถานการณ์ ทั้งนี้ ผลการศึกษาที่ได้จะเป็นแนวทางให้กับผู้วิเคราะห์ข้อมูลในการเลือกใช้สถิติทดสอบสำหรับการทดสอบสมมุติฐานความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันในกรณีที่ข้อมูลเป็นจำนวนนับได้อย่างเหมาะสม

คำสำคัญ: การทดสอบครัสคัล-วัลลิส, การทดสอบแวน เดอ แวร์เดน, การทดสอบของเวลช์, กำลังการทดสอบ, ความผิดพลาดแบบที่ 1

ABSTRACT

This research aimed to compare the performance of the one-way analysis of variance (F test), Welch's test, Kruskal-Wallis test, Van der Waerden test, and Median test for testing the difference between central values of more than two independent populations using counting data with binomial and Poisson distributions. The criteria used to evaluate and compare the efficiency of the proposed

tests are the ability to control the probability of type I error and power. The results showed that for the binomial distribution data, all tests could control the probability of type I error for most situations, except when the sample size of each group was 10, and the variance of each group was less than 4, and when the sample size of each group was 15 and (n, p) were (10, 0.1) and (10, 0.7). For the Poisson distribution data, all tests could also control the probability of type I error in almost situations. However, except for the Welch's Test, the Van der Waerden Test, and the Median Test, the probability of type 1 error could not be controlled in cases when the sample size of each group was ten and λ = 30. When considering the power, it was found that the analysis of variance provided the highest power than other tests in almost situations. The results of this study will guide data analysts in choosing an appropriate statistic for testing the difference between central values of more than two independent populations when data is count data.

KEYWORDS: Kruskal-Wallis test, Van der Waerden test, Welch's test, power of a test, type I error

***Corresponding Author: pongsathorn.cho@nmu.ac.th**

Received: 28/08/2022; Revised: 29/06/2023; Accepted: 14/07/2023

1. บทนำ

ในการทดสอบสมมุติฐานทางสถิติเพื่อเปรียบเทียบค่ากลางของประชากรมากกว่า 2 กลุ่ม สถิติที่นิยมใช้กันอย่างแพร่หลาย คือ การวิเคราะห์ความแปรปรวน (analysis of variance หรือ ANOVA) (Moder, 2007) ซึ่งเป็นการทดสอบอิงพารามิเตอร์ (parametric test) โดยตัวสถิติทดสอบที่ใช้ในการวิเคราะห์ความแปรปรวนคือ ตัวสถิติทดสอบเอฟ (F test statistic) ซึ่งการทดสอบสมมุติฐานด้วยสถิติดังกล่าวมีข้อตกลงเบื้องต้น คือ ประชากรแต่ละกลุ่มมีการแจกแจงปรกติ (normal distribution) ความแปรปรวนของประชากรแต่ละกลุ่มไม่แตกต่างกัน (homoscedastic) และตัวอย่างแต่ละหน่วยเป็นอิสระกัน แต่ในทางปฏิบัติ ข้อมูลที่เก็บรวบรวมมาอาจไม่เป็นไปตามข้อตกลงเบื้องต้น กล่าวคือ ข้อมูลอาจไม่ได้มีการแจกแจงปรกติเสมอไปโดยเฉพาะข้อมูลในงานวิจัยเชิงประยุกต์ (Urawong et al., 2022) เช่น จำนวนผู้ป่วยโควิดที่เข้ารับการรักษาในโรงพยาบาล จำนวนผู้ใช้บริการของหน่วยงานราชการ จำนวนอุบัติเหตุรายวัน เป็นต้น การวิเคราะห์ข้อมูลด้วยการวิเคราะห์ความแปรปรวนอาจส่งผลให้ข้อสรุปไม่ถูกต้อง หรือผลการวิเคราะห์ข้อมูลที่ได้ไม่น่าเชื่อถือและไม่สามารถนำไปอ้างอิงหรือไปอธิบายสิ่งที่ต้องการศึกษาได้ โดย Scheffe (1959) ได้อธิบายไว้ว่า ถ้าลักษณะการแจกแจงของประชากรไม่เป็นไปตามข้อตกลงเบื้องต้นจะส่งผลให้การประมาณค่าหรือการทดสอบสมมุติฐานเกี่ยวกับความแปรปรวนของประชากรเกิดความผิดพลาดได้ และจะส่งผลอย่างมากต่อการสรุปผลที่ได้จากการใช้ตัวสถิติทดสอบเอฟของการวิเคราะห์ความแปรปรวน โดยเฉพาะเมื่อขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน ซึ่งสอดคล้องกับงานวิจัยของ Glass, Peckham & Sander (1972) ที่พบว่า การวิเคราะห์ความแปรปรวนจะมีความแกร่ง (robustness) เมื่ออยู่ภายใต้เงื่อนไขที่ประชากรมีการแจกแจงปรกติ ดังนั้น เมื่อข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น การทดสอบไม่อิงพารามิเตอร์ (non-parametric test) จึงเป็นอีกทางเลือกหนึ่งที่ถูกนำมาใช้แทนการทดสอบอิงพารามิเตอร์

ในอดีตที่ผ่านมา มีงานวิจัยหลายงานที่ศึกษาเกี่ยวกับประสิทธิภาพของการทดสอบแบบอิงพารามิเตอร์และไม่อิงพารามิเตอร์สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันในกรณีที่ข้อมูลไม่มีการแจกแจงปรกติ ยกตัวอย่างเช่น Hecke (2012) ได้ศึกษากำลังการทดสอบของการวิเคราะห์ความแปรปรวนและการทดสอบครัสคัล-วัลลิส ผลการศึกษาพบว่า เมื่อประชากรมีการแจกแจงที่ไม่สมมาตร การทดสอบครัสคัล-วัลลิสจะให้กำลังการทดสอบที่สูงกว่าการวิเคราะห์ความแปรปรวน ต่อมา Bowarnkitiwong & Areekul (2017) ได้ศึกษาความแกร่งของสถิติทดสอบเอฟเมื่อความแปรปรวนของประชากรไม่เท่ากันและพบว่าสถิติทดสอบเอฟมีความแกร่งทั้งในกรณีที่ตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน ในขณะที่ Khomduean & Araveeporn (2017) ได้ศึกษาประสิทธิภาพของสถิติทดสอบที่ใช้ทดสอบค่าเฉลี่ยของประชากร 3 กลุ่มกรณีที่ความแปรปรวนของประชากรทั้ง 3 กลุ่มเท่ากันและไม่เท่ากัน ผลการศึกษาพบว่า ในกรณีที่ประชากรทั้ง 3 กลุ่มมีการแจกแจงปรกติและมีความแปรปรวนเท่ากัน สถิติทดสอบเอฟของการวิเคราะห์ความแปรปรวน สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Brown-Forsythe ที่ปรับแล้ว สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้และให้กำลังการทดสอบสูงในเกือบทุกสถานการณ์ และในกรณีที่ประชากรทั้ง 3 กลุ่มมีการแจกแจงปรกติแต่มีความแปรปรวนแตกต่างกัน ผลการศึกษาพบว่า สถิติทดสอบของเวลช์และสถิติทดสอบของ Marascuilo สามารถควบคุมความน่าจะ

เป็นของความผิดพลาดแบบที่ 1 ได้ และมีกำลังการทดสอบสูงที่สุดในทุกกรณี ส่วน Lüpsen (2017) ได้ศึกษาประสิทธิภาพของสถิติไม่อิงพารามิเตอร์สำหรับการวิเคราะห์ความแปรปรวนและพบว่า การทดสอบแวน เดอ แวร์เดน มีประสิทธิภาพดีที่สุดเกือบทุกกรณี ส่วนการทดสอบของ Puri & Sen และการทดสอบของ Akritas & Brunners ATS ให้กำลังการทดสอบที่ค่อนข้างต่ำและไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในหลาย ๆ สถานการณ์เช่นเดียวกับการทดสอบอื่น ๆ โดยเฉพาะในกรณีที่ประชากรมีการแจกแจงล็อกนอร์มัล (log-normal distribution) อย่างไรก็ตาม ถึงแม้ว่าจะมีงานวิจัยหลาย ๆ งานที่ได้ทำการศึกษาเกี่ยวกับประสิทธิภาพของการทดสอบแบบอิงพารามิเตอร์และไม่อิงพารามิเตอร์สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันในกรณีที่ข้อมูลไม่มีการแจกแจงปรกติ แต่งานวิจัยส่วนใหญ่ทำการศึกษาเฉพาะกรณีข้อมูลเชิงปริมาณแบบต่อเนื่องเท่านั้น สำหรับกรณีข้อมูลเชิงปริมาณไม่ต่อเนื่องยังมีการศึกษาค่อนข้างน้อย ยกตัวอย่างเช่น งานวิจัยของ Sangthong (2018) ได้ทำการศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบค่ากลางระหว่างประชากรหลายกลุ่มเมื่อข้อมูลเป็นแบบลิเคิร์ท 5 ระดับ ผลการศึกษาพบว่า การวิเคราะห์ความแปรปรวนเป็นการทดสอบที่มีประสิทธิภาพดีที่สุดในเกือบทุกสถานการณ์ ยกเว้นกรณีที่ประชากรมีการแจกแจงเบ้เชิงบวกและความโด่งสูงกว่าปกติ การทดสอบครัสคัล-วัลลิส เป็นการทดสอบที่มีประสิทธิภาพสูงสุดกรณีที่ประชากรมีความแปรปรวนเท่ากันและตัวอย่างมีขนาดกลาง และการทดสอบแวน เดอ แวร์เดน เป็นการทดสอบที่มีประสิทธิภาพสูงสุดกรณีที่ประชากรมีความแปรปรวนไม่เท่ากันและตัวอย่างมีขนาดกลางและขนาดใหญ่ ส่วนกรณีที่ประชากรมีการแจกแจงเบ้เชิงลบและมีความโด่งสูงกว่าปกติ ผลการศึกษาพบว่า การวิเคราะห์ความแปรปรวน การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานไม่มีประสิทธิภาพ ดังนั้น ในการศึกษาครั้งนี้จึงต้องการศึกษาประสิทธิภาพของการทดสอบอิงพารามิเตอร์และไม่อิงพารามิเตอร์ 5 วิธี ได้แก่ การวิเคราะห์ความแปรปรวน การทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกัน ในกรณีที่ข้อมูลเป็นข้อมูลเชิงปริมาณไม่ต่อเนื่อง โดยการแจกแจงที่ศึกษาคือ การแจกแจงทวินาม (Binomial distribution) และการแจกแจงปัวซง (Poisson distribution) ซึ่งการแจกแจงทั้งสองการแจกแจงเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่องที่สำคัญที่มีค่าของตัวแปรสุ่มเป็นจำนวนนับและเป็นการแจกแจงที่ใช้กันอย่างแพร่หลายในงานวิจัยเชิงประยุกต์ ทั้งนี้ ผลการศึกษาที่ได้จะเป็นแนวทางให้กับผู้วิเคราะห์ข้อมูลในการเลือกใช้สถิติทดสอบสำหรับการทดสอบสมมุติฐานความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันในกรณีที่ข้อมูลเป็นจำนวนนับได้อย่างเหมาะสม สำหรับเนื้อหาในบทความมีการจัดเรียงตามลำดับดังนี้ ส่วนที่ 2 นำเสนอแนวคิดและรายละเอียดวิธีการทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกัน 5 วิธี ส่วนที่ 3 นำเสนอวิธีการวิจัย ซึ่งประกอบด้วยขอบเขตการวิจัย (จำนวนประชากร ขนาดตัวอย่าง การแจกแจงที่ศึกษา การกำหนดค่าพารามิเตอร์ และเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของการทดสอบ) และขั้นตอนการดำเนินการวิจัย และส่วนที่ 4 และ 5 นำเสนอผลการวิจัยและสรุปและวิจารณ์ผลการวิจัยตามลำดับ

2. การทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกัน

2.1 การวิเคราะห์ความแปรปรวน (ANOVA)

การวิเคราะห์ความแปรปรวน (analysis of variance) เป็นสถิติที่ใช้ในการเปรียบเทียบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกัน โดยมีข้อสมมติเบื้องต้นที่สำคัญ คือ ข้อมูลแต่ละกลุ่มที่นำมาวิเคราะห์ต้องมาจากประชากรที่มีการแจกแจงปรกติและมีความแปรปรวนไม่แตกต่างกัน หลักการสำคัญในการวิเคราะห์ความแปรปรวนคือ การแบ่งความแปรปรวนของข้อมูลทั้งหมดออกตามสาเหตุที่ทำให้ข้อมูลแตกต่างกัน โดยที่ ความแปรปรวนรวม = ความแปรปรวนระหว่างกลุ่ม + ความแปรปรวนภายในกลุ่ม และการคำนวณค่าสถิติทดสอบสำหรับการวิเคราะห์ความแปรปรวนสามารถทำได้ดังนี้

2.1.1 คำนวณผลบวกกำลังสอง (sum of squares) คำนวณผลบวกกำลังสอง หรือเรียกย่อ ๆ ว่า SS แยกตามแหล่งความแปรปรวน (source of variation) ดังนี้

คำนวณผลบวกกำลังสองของค่าสังเกตทั้งหมด (total sum of squares) หรือเรียกย่อ ๆ ว่า SST โดยที่

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}..)^2$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij}^2 - \frac{\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij}\right)^2}{n} \quad (1)$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij}^2 - \frac{y_{..}^2}{n}$$

และองศาเสรี (degree of freedom) ของค่าสังเกตทั้งหมดเท่ากับ

คำนวณผลบวกกำลังสองระหว่างทรีตเมนต์ (between treatment sum of squares) หรือเรียกย่อ ๆ ว่า SSTr โดยที่

$$SSTr = \sum_{j=1}^{k} n_j(\bar{y}_{.j} - \bar{y}_{..})^2 = \sum_{j=1}^{k}\left(\frac{y_{.j}^2}{n_j}\right) - \frac{y_{..}^2}{n} \quad (2)$$

และองศาเสรีของทรีตเมนต์เท่ากับ $k-1$

คำนวณผลบวกกำลังสองของความคลาดเคลื่อน (error sum of squares) หรือเรียกย่อ ๆ ว่า SSE โดยที่

$$SSE = \sum_{j=1}^{k}\left[\sum_{i=1}^{n_j} y_{ij}^2 - \frac{y_{.j}^2}{n_j}\right] \quad (3)$$

หรือ

$$SSE = SST - SSTr \quad (4)$$

และองศาเสรีของความคลาดเคลื่อนเท่ากับ

2.1.2 คำนวณกำลังสองเฉลี่ย (mean square) คำนวณกำลังสองเฉลี่ยหรือเรียกย่อ ๆ ว่า MS ดังนี้

คำนวณทรีตเมนต์กำลังสองเฉลี่ย (mean square treatment) หรือเรียกย่อ ๆ ว่า MSTr โดยที่

$$MSTr = \frac{SSTr}{k-1} \quad (5)$$

คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error) หรือเรียกย่อ ๆ ว่า MSE โดยที่

$$MSE = \frac{SSE}{n-k} \quad (6)$$

2.1.3 คำนวณค่าสถิติทดสอบ

สำหรับสถิติทดสอบที่ใช้ในการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว คือ สถิติทดสอบที่มีการแจกแจงเอฟ (F distribution) องศาเสรี และ โดย ค่าสถิติทดสอบคำนวณได้จาก

$$F = \frac{MSTr}{MSE} \quad (7)$$

การสรุปผล เราจะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อ $F \geq f_{\alpha,[(k-1),(n-k)]}$ โดย $f_{\alpha,[(k-1),(n-k)]}$ คือ ค่าวิกฤติที่ได้จากตารางการแจกแจงเอฟ (Sinsomboonthong, 2017)

2.2 การทดสอบของเวลช์ (Welch's test)

Welch (1951) ได้พัฒนาการทดสอบวิธีนี้ขึ้นมาจากการปรับปรุงการทดสอบของ Cochran เดิมโดยถ่วงน้ำหนักค่าสถิติทดสอบสำหรับทดสอบความแตกต่างค่าเฉลี่ยของประชากร k กลุ่ม โดยแนวคิดในการถ่วงน้ำหนักดังกล่าวก็เพื่อลดอิทธิพลของความต่างทั้งในรูปแบบการแจกแจงความแปรปรวน หรือพารามิเตอร์ต่าง ๆ (heterogeneity) สำหรับการวิเคราะห์ความแปรปรวนนั้น ภาวะความต่างจะเกิดขึ้นเมื่อความแปรปรวนของประชากรแต่ละกลุ่มแตกต่างกัน โดยค่าสถิติทดสอบของเวลช์สามารถคำนวณได้ดังนี้

$$F_W = \frac{\sum_{j=1}^{k}\left[\frac{W_j(\bar{y}_{j.} - \bar{y}_{w..})^2}{(k-1)}\right]}{1 + \frac{2(k-2)}{(k^2-1)}\sum_{j=1}^{k}\left[\frac{\left(1-\frac{W_j}{u}\right)}{(n_j-1)}\right]} \quad (8)$$

เมื่อ $W_i = \frac{n_i}{S_i^2}$, $\bar{y}_{w..} = \sum_{j=1}^{k}\left(\frac{W_j\bar{y}_j}{u}\right)$, และ $u = \sum_{j=1}^{k} W_j$

โดยที่

F_W คือ สถิติทดสอบของเวลช์ที่มีการแจกแจงเอฟ (F distribution) มีองศาเสรี $k-1$ และ ν โดยที่

$$\nu = \left[\frac{3}{k^2-1}\sum_{j=1}^{k}\left(\frac{(1-(W_i/u))^2}{(n_j-1)}\right)\right]^{-1} \quad (9)$$

$\bar{y}_{j.}$ คือ ค่าเฉลี่ยตัวอย่างกลุ่มที่ j

n_j คือ ขนาดตัวอย่างกลุ่มที่ j

S_i^2 คือ ความแปรปรวนตัวอย่างกลุ่มที่ j

การสรุปผล เราจะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อ $F_W \geq f_{\alpha,[(k-1),\nu]}$ โดย $f_{\alpha,[(k-1),\nu]}$ คือ ค่าวิกฤติที่ได้จากตารางการแจกแจงเอฟ

2.3 การทดสอบครัสคัล-วัลลิส (Kruskal-Wallis test)

การทดสอบครัสคัล-วัลลิสเป็นการทดสอบไม่อิงพารามิเตอร์ ถูกพัฒนาขึ้นโดย Kruskal และ Wallis ในปี ค.ศ. 1952 (Kruskal & Wallis, 1952) โดยมีแนวคิดใน

การใช้ลำดับที่ของข้อมูลมาคำนวณค่าสถิติทดสอบแทนข้อมูลเดิมเพื่อลดอิทธิพลของการแจกแจงไม่ปรกติของข้อมูล (McDonald, 2014)

สำหรับสถิติทดสอบครัสคัล-วัลลิสมีการแจกแจงไคกำลังสอง (Chi-squared distribution) โดยประมาณและสามารถคำนวณได้ดังนี้

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \left(\frac{(\sum_{j=1}^{k} R_j)^2}{n_j} \right) - 3(n+1) \quad (10)$$

เมื่อ

H คือ สถิติทดสอบครัสคัล-วัลลิส

R_j คือ อันดับของค่าสังเกตของกลุ่มที่ j

k คือ จำนวนกลุ่ม

n คือ ขนาดตัวอย่างทั้งหมด

การสรุปผล เราจะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อ $H \geq \chi^2_{\alpha,k-1}$ โดย $\chi^2_{\alpha,k-1}$ ค่าวิกฤติที่ได้จากตารางการแจกแจงไคกำลังสอง (Sinsomboonthong, 2020; Sheskin, 2000)

2.4 การทดสอบแวน เดอ แวร์เดน (Van der Waerden test)

การทดสอบแวน เดอ แวร์เดน หรือการทดสอบแวน เดอ แวร์เดน นอร์มอล สกอร์ (Van der Waerden normal-scores test) เป็นการทดสอบไม่อิงพารามิเตอร์เช่นเดียวกับการทดสอบครัสคัล-วัลลิส ถูกพัฒนาขึ้นโดยแวน เดอ แวร์เดน ในปี ค.ศ. 1953 (Van der Waerden, 1952; Van der Waerden, 1953) โดยมีแนวคิดในการใช้ค่าลำดับที่จากวิธีของครัสคัล-วัลลิสมาแปลงเป็นควอนไทล์ของการแจกแจงปรกติมาตรฐาน (standard normal distribution) เรียกว่า inverse-normal scores (Z(i)) จากนั้นนำมาใช้ในการคำนวณค่าสถิติทดสอบแทนค่าสังเกตเดิม โดยข้อดีของการทดสอบแวน เดอ แวร์เดน คือให้ประสิทธิภาพที่ดีกว่าการทดสอบเอฟของการวิเคราะห์ความแปรปรวนเมื่อข้อมูลมีการแจกแจงปรกติและให้ประสิทธิภาพที่ดีกว่าการทดสอบครัสคัล-วัลลิสเมื่อข้อมูลไม่มีการแจกแจงปรกติ สำหรับสถิติทดสอบของแวนเดอแวร์เดนสามารถคำนวณได้ดังนี้

$$V - W = \frac{\sum_{j=1}^{k} n_j (\bar{Z}_j)^2}{\tilde{S}^2} \quad (11)$$

เมื่อ

$V-W$ คือ สถิติทดสอบแวนเดอแวร์เดนที่มีการแจกแจงไคกำลังสองโดยประมาณ

$\bar{Z}_j$ คือ คะแนนมาตรฐานเฉลี่ยของกลุ่มที่ j

$\tilde{S}^2$ คือ ความแปรปรวนตัวอย่าง

การสรุปผล เราจะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อ $V-W \geq \chi^2_{\alpha,k-1}$ โดย $\chi^2_{\alpha,k-1}$ ค่าวิกฤติที่ได้จากตารางการแจกแจงไคกำลังสอง (Sangthong, 2018; Sheskin, 2000)

2.5 การทดสอบมัธยฐาน (Median test)

การทดสอบมัธยฐานเป็นการทดสอบไม่อิงพารามิเตอร์ที่ใช้ทดสอบค่ากลางของประชากรตั้งแต่ 3 กลุ่มขึ้นไป โดยใช้จำนวนค่าสังเกตที่มีค่ามากกว่าหรือน้อยกว่าค่ามัธยฐานของตัวอย่างทั้งหมดเป็นตัวสถิติทดสอบและสามารถคำนวณได้ดังนี้

$$M = \frac{n^2}{ab} \sum_{j=1}^{k} \frac{\left(O_{1j} - \frac{n_j a}{n} \right)^2}{n_j} \quad (12)$$

เมื่อ

M คือ สถิติทดสอบมัธยฐาน มีการแจกแจงไคกำลังสองโดยประมาณ

a คือ ผลรวมของจำนวนค่าสังเกตที่มีค่ามากกว่ามัธยฐานร่วม

b คือ ผลรวมของจำนวนค่าสังเกตที่มีค่าน้อยกว่าหรือเท่ากับมัธยฐานร่วม

O_{1j} คือ จำนวนค่าสังเกตที่มีค่ามากกว่ามัธยฐานร่วมของกลุ่มที่ j

การสรุปผล เราจะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อ $M \geq \chi^2_{\alpha,k-1}$ โดย $\chi^2_{\alpha,k-1}$ ค่าวิกฤติที่ได้จากตารางการแจกแจงไคกำลังสอง (Sangthong, 2018)

3. วิธีการวิจัย

การวิจัยครั้งนี้ใช้การจำลองมอนติคาร์โล (Monte Carlo simulation) (Harrison, 2010) ในการจำลองข้อมูลเพื่อตรวจสอบประสิทธิภาพของการทดสอบอิงพารามิเตอร์และไม่อิงพารามิเตอร์สำหรับทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ โดยผู้วิจัยได้กำหนดขอบเขตและขั้นตอนการดำเนินการวิจัยดังนี้

3.1 ขอบเขตการวิจัย

1. กำหนดจำนวนประชากรที่ทำการศึกษา คือ 3 กลุ่ม

2. ขนาดตัวอย่างที่ศึกษา ในที่นี้จะพิจารณาเฉพาะกรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน คือ

(n_1, n_2, n_3) = (10, 10, 10), (15, 15, 15), (20, 20, 20), (25, 25, 25), (30, 30, 30)

3. การแจกแจงที่ศึกษา คือ การแจกแจงทวินาม (Binomial distribution) และการแจกแจงปัวซง (Poisson distribution) กำหนดพารามิเตอร์ของการแจกแจงทั้ง 2 การแจกแจงให้ครอบคลุมทั้งกรณีการแจกแจงเบ้และการแจกแจงสมมาตร (ดังรูปที่ 1-2) โดยมีรายละเอียดดังนี้

ความน่าจะเป็นของความผิดพลาดแบบที่ 1 (Probability of type I error)

สำหรับการแจกแจงทวินาม กำหนดพารามิเตอร์ของข้อมูลทั้ง 3 กลุ่ม ดังนี้

$[(n, p_1), (n, p_2), (n, p_3)]$ = [(10, 0.1), (10, 0.1), (10, 0.1)], [(10, 0.5), (10, 0.5), (10, 0.5)], [(10, 0.7), (10, 0.7), (10, 0.7)], [(30, 0.1), (30, 0.1), (30, 0.1)], [(30, 0.5), (30, 0.5), (30, 0.5)], [(30, 0.7), (30, 0.7), (30, 0.7)], [(50, 0.1), (50, 0.1), (50, 0.1)], [(50, 0.5), (50, 0.5), (50, 0.5)], [(50, 0.7), (50, 0.7), (50, 0.7)]

สำหรับการแจกแจงปัวซง กำหนดพารามิเตอร์ของข้อมูลทั้ง 3 กลุ่ม ดังนี้

$[\lambda_1, \lambda_2, \lambda_3]$ = [5, 5, 5], [10, 10, 10], [15, 15, 15], [20, 20, 20], [25, 25, 25], [30, 30, 30]

กำลังการทดสอบ (Power of a test)

สำหรับการแจกแจงทวินาม กำหนดพารามิเตอร์ของข้อมูลทั้ง 3 กลุ่ม ดังนี้

$[(n, p_1), (n, p_2), (n, p_3)]$ = [(10, 0.1), (10, 0.125), (10, 0.1)], [(10, 0.1), (10, 0.125), (10, 0.15)], [(10, 0.5), (10, 0.525), (10, 0.5)], [(10, 0.7), (10, 0.8), (10, 0.7)], [(30, 0.1), (30, 0.125), (30, 0.1)], [(30, 0.1), (30, 0.125), (30, 0.15)], [(30, 0.5), (30, 0.525), (30, 0.5)], [(30, 0.7), (30, 0.8), (30, 0.7)], [(50, 0.1), (50, 0.125), (50, 0.1)], [(50, 0.1), (50, 0.125), (50, 0.15)], [(50, 0.5), (50, 0.525), (50, 0.5)], [(50, 0.7), (50, 0.8), (50, 0.7)]

สำหรับการแจกแจงปัวซง กำหนดพารามิเตอร์ของข้อมูลทั้ง 3 กลุ่ม ดังนี้

$[\lambda_1, \lambda_2, \lambda_3]$ = [5, 6, 5], [5, 6, 7], [5, 6, 8], [20, 22, 20], [20, 22, 24], [20, 22, 26], [30, 35, 30], [30, 35, 40], [30, 35, 45]

4. กำหนดระดับนัยสำคัญในการศึกษา α = 0.05

5. ประสิทธิภาพของการทดสอบพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 (type I error) และกำลังการทดสอบ (power of a test)

6. ทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์ที่กำหนด

3.2 ขั้นตอนการดำเนินการวิจัย

3.2.1 การศึกษาความน่าจะเป็นของความผิดพลาดแบบที่ 1

การจำลองข้อมูลเพื่อศึกษาความน่าจะเป็นของความผิดพลาดแบบที่ 1 แสดงดังรูปภาพ 3 โดยมีขั้นตอนดังต่อไปนี้

1. กำหนดสมมุติฐานการทดสอบสำหรับทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่ม

H_0: ประชากรทั้ง 3 กลุ่มมีค่ากลาง (ค่าเฉลี่ย/มัธยฐาน) ไม่แตกต่างกัน

H_1: มีประชากรอย่างน้อย 1 กลุ่มที่มีค่ากลาง (ค่าเฉลี่ย/มัธยฐาน) แตกต่างจากกลุ่มอื่น

2. สร้างข้อมูลตามสถานการณ์ต่าง ๆ ที่กำหนดไว้ในขอบเขตการวิจัยภายใต้สมมุติฐานว่าง (null hypothesis, H_0)

3. คำนวณค่าสถิติทดสอบสำหรับการวิเคราะห์ความแปรปรวน การทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน

4. เปรียบเทียบค่าสถิติทดสอบที่ได้ในข้อ 2 กับค่าวิกฤต (critical values) ที่ได้จากการคำนวณโดยใช้โปรแกรมทางสถิติ จากนั้นทำการสรุปผลการทดสอบตามสมมุติฐานที่กำหนดไว้ในข้อ 1 ว่าปฏิเสธหรือยอมรับสมมุติฐานว่าง (H_0)

5. ดำเนินการตามข้อ 1-3 ซ้ำ จำนวน 10,000 ครั้งในแต่ละสถานการณ์

6. นับจำนวนครั้งในการปฏิเสธสมมุติฐานว่าง

7. คำนวณค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 (empirical type 1 error, $\hat{\alpha}$) โดยที่

$$\hat{\alpha} = \frac{\text{จำนวนครั้งในการปฏิเสธ } H_0 \text{ เมื่อ } H_0 \text{ เป็นจริง}}{\text{จำนวนครั้งในการทำซ้ำ}}$$

8. เปรียบเทียบค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของการทดสอบทั้ง 5 วิธีกับเกณฑ์ของ Cochran (Cochran, 1947) ที่ระดับนัยสำคัญ 0.05 ถ้า

ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตกอยู่ในช่วง [0.04, 0.06] จะสรุปว่า การทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้และเป็นสถิติทดสอบที่มีสมบัติความแกร่ง (robustness)

3.2.2 การศึกษากำลังการทดสอบ

การจำลองข้อมูลเพื่อศึกษากำลังการทดสอบแสดงดังรูปภาพ 4 โดยมีขั้นตอนดังต่อไปนี้

1. กำหนดสมมุติฐานการทดสอบสำหรับทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่ม (เช่นเดียวกับการศึกษาความน่าจะเป็นของความผิดพลาดแบบที่ 1)

2. สร้างข้อมูลตามสถานการณ์ต่าง ๆ ที่กำหนดไว้ในขอบเขตการวิจัยภายใต้สมมุติฐานทางเลือก (alternative hypothesis, H_1)

3. คำนวณค่าสถิติทดสอบสำหรับการวิเคราะห์ความแปรปรวน การทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน

4. เปรียบเทียบค่าสถิติทดสอบที่ได้ในข้อ 2 กับค่าวิกฤต (critical value) ที่ได้จากการคำนวณโดยใช้โปรแกรมทางสถิติ จากนั้นทำการสรุปผลการทดสอบตามสมมุติฐานที่กำหนดไว้ในข้อ 1 ว่าปฏิเสธหรือยอมรับสมมุติฐานว่าง (H_0)

5. ดำเนินการตามข้อ 1-3 ซ้ำ จำนวน 10,000 ครั้ง ในแต่ละสถานการณ์

6. นับจำนวนครั้งในการปฏิเสธสมมุติฐานว่าง

7. คำนวณค่าประมาณกำลังการทดสอบ (empirical power, $1-\hat{\beta}$) ของการทดสอบทั้ง 5 วิธี โดยที่

$$1-\hat{\beta} = \frac{\text{จำนวนครั้งในการปฏิเสธ } H_0 \text{ เมื่อ } H_1 \text{ เป็นจริง}}{\text{จำนวนครั้งในการทำซ้ำ}}$$

8. เปรียบเทียบกำลังการทดสอบของการทดสอบทั้ง 5 วิธี ถ้าวิธีใดสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และให้กำลังการทดสอบที่มากกว่าจะสรุปว่า การทดสอบนั้นมีประสิทธิภาพมากกว่า

4. ผลการวิจัย

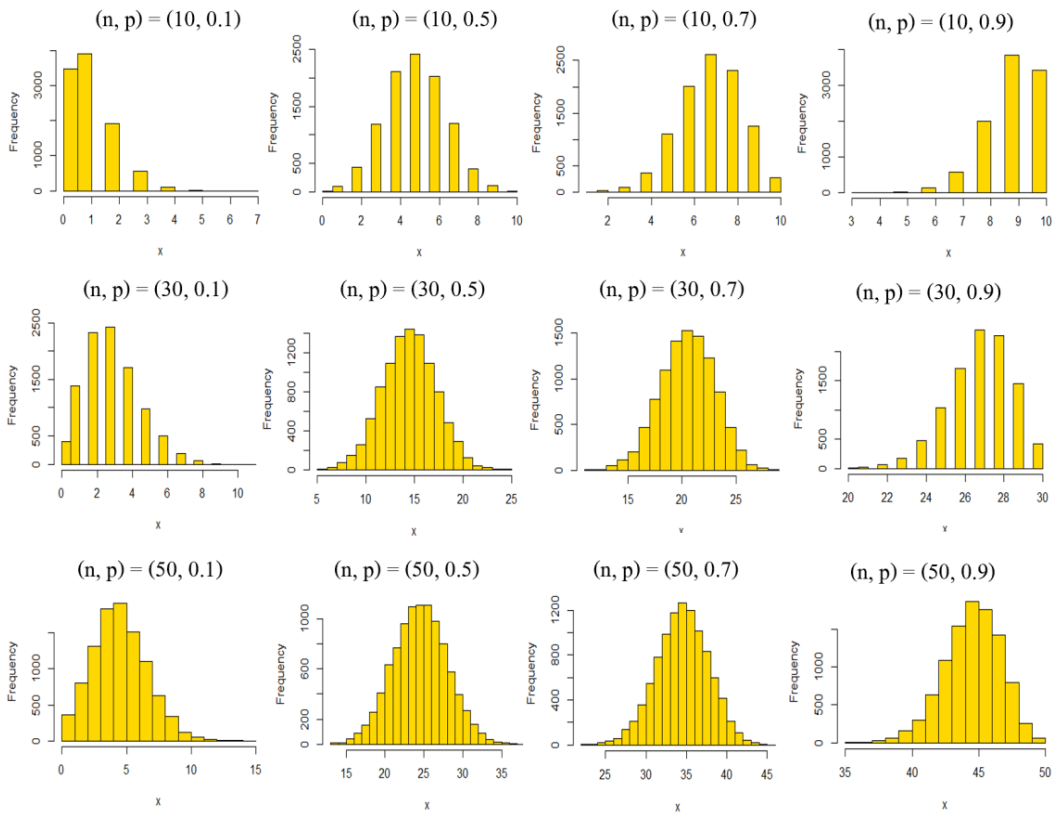
ผลการเปรียบเทียบประสิทธิภาพของการวิเคราะห์ความแปรปรวน การทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานสำหรับทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ แสดงดังตารางที่ 1-4 โดยมีรายละเอียดดังนี้

4.1 ผลการศึกษาความน่าจะเป็นของความผิดพลาดแบบที่ 1

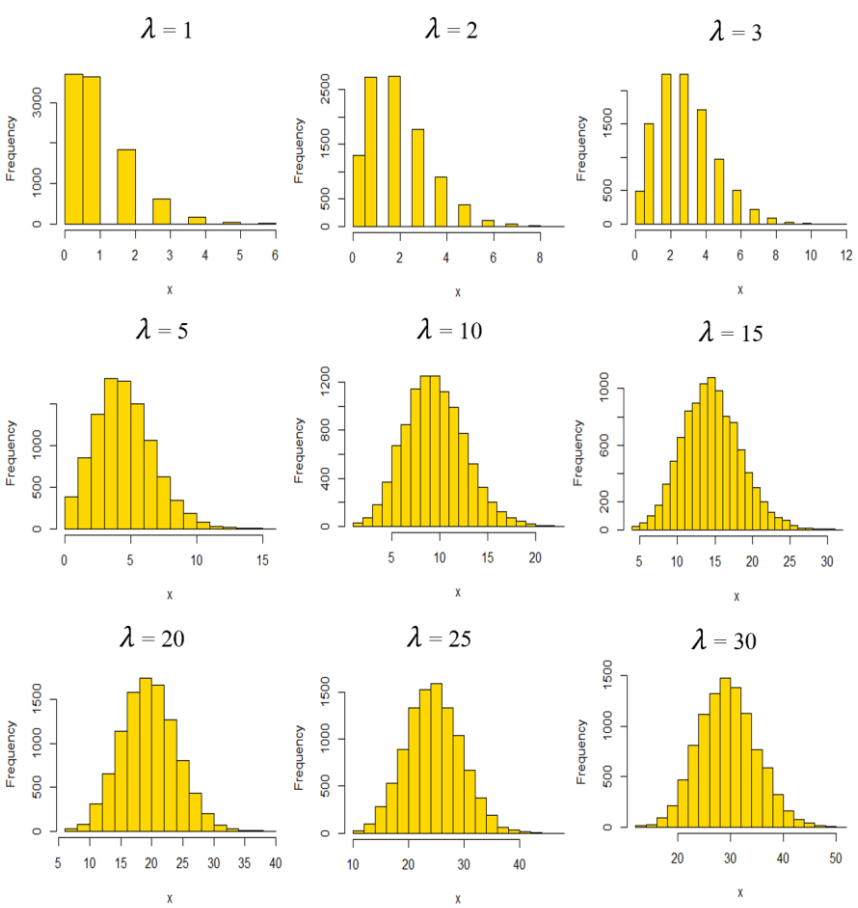
กรณีที่ข้อมูลมีการแจกแจงทวินาม การทดสอบทั้ง 5 วิธี สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และความแปรปรวน (np(1 - p))ของข้อมูลแต่ละกลุ่มน้อยกว่า 4 และกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 15 และ (n, p) เท่ากับ (10, 0.1) และ (10, 0.7) นอกจากนี้ ในกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และความแปรปรวนของข้อมูลแต่ละกลุ่มมากกว่า 10 การทดสอบมัธยฐานเป็นเพียงการทดสอบเดียวที่ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ อย่างไรก็ตาม เมื่อขนาดตัวอย่างแต่ละกลุ่มเพิ่มขึ้น การทดสอบทั้ง 5 วิธีจะสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ดีขึ้นโดยจะเห็นได้จากค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่ใกล้เคียงกับระดับนัยสำคัญที่กำหนด 0.05 (ตารางที่ 1) สำหรับกรณีที่ข้อมูลมีการแจกแจงปัวซง การทดสอบทั้ง 5 วิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นการทดสอบของเวลช์ การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานที่ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และ λ = 30 (ตารางที่ 2) ทั้งนี้ สาเหตุที่สถิติทดสอบไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้อาจเนื่องมาจากตัวอย่างมีขนาดเล็กทำให้การแจกแจงของตัวสถิติทดสอบไม่เป็นไปตามทฤษฎี

4.2 ผลการศึกษากำลังการทดสอบ

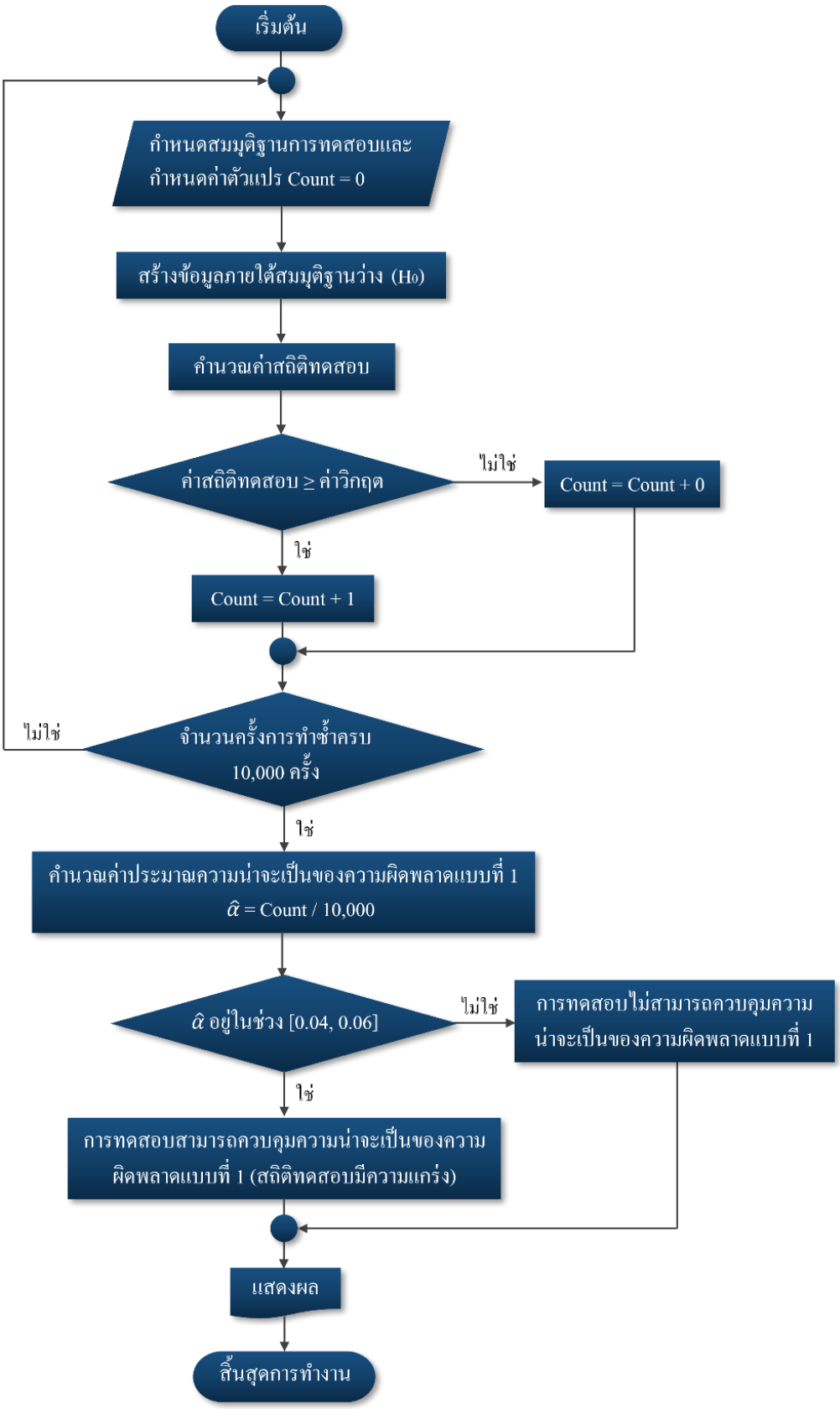
จากตารางที่ 3 และตารางที่ 4 พบว่า การวิเคราะห์ความแปรปรวนให้กำลังการทดสอบสูงกว่าการทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานเกือบทุกกรณีที่ศึกษา และมีบางกรณีที่การทดสอบของเวลช์ให้กำลังการทดสอบสูงกว่าการวิเคราะห์ความแปรปรวนเล็กน้อย ทั้งนี้ สาเหตุอาจเนื่องมาจากข้อมูลที่ทำการศึกษาในหลาย ๆ สถานการณ์ยังมีการแจกแจงที่ค่อนข้างสมมาตร (ดังรูปที่ 1-2) หรือมีความเบ้ไม่มากนัก ส่งผลให้การแจกแจงของตัวสถิติทดสอบของการวิเคราะห์ความแปรปรวนยังสามารถประมาณได้ด้วยการแจกแจงเอฟ เมื่อพิจารณาเฉพาะการทดสอบไม่อิงพารามิเตอร์สำหรับทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ พบว่า การทดสอบครัสคัล-วัลลิส และการทดสอบแวน เดอ แวร์เดน มีประสิทธิภาพที่ดีกว่า



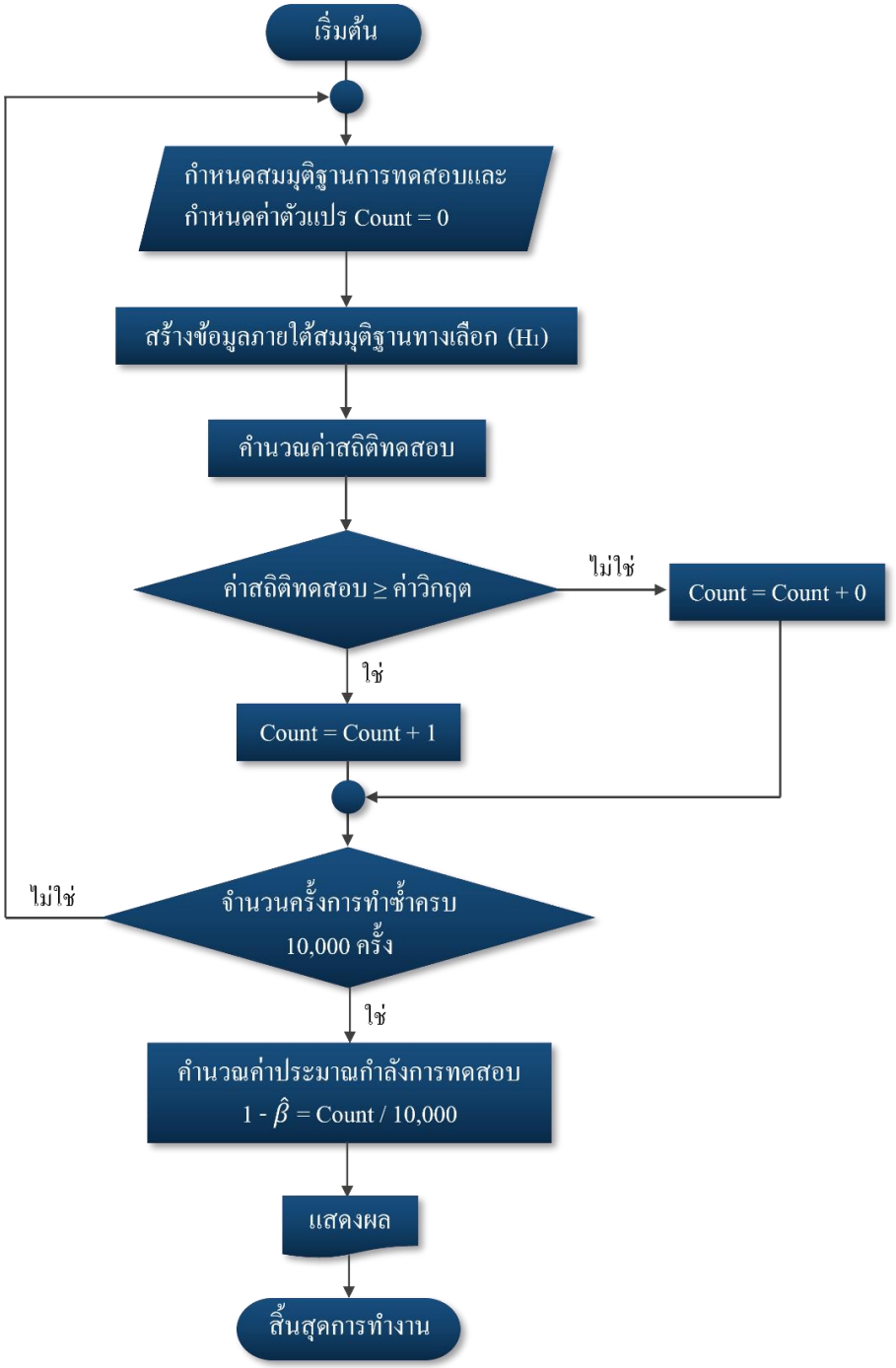
รูปที่ 1 ฮิสโทแกรมแสดงลักษณะข้อมูลที่มีการแจกแจงทวินามภายใต้พารามิเตอร์ค่าต่าง ๆ



รูปที่ 2 ฮิสโทแกรมแสดงลักษณะข้อมูลที่มีการแจกแจงปัวซงภายใต้พารามิเตอร์ค่าต่าง ๆ



รูปที่ 3 แผนภาพแสดงขั้นตอนการหาค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1



รูปที่ 4 แผนภาพแสดงขั้นตอนการหาค่าประมาณกำลังการทดสอบ

ตารางที่ 2 ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่มีการแจกแจงปัวซง

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
λ_1	λ_2	λ_3	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
5	5	5	F (ANOVA)	0.0533	0.0483	0.0548	0.0470	0.0495
			Welch	0.0533	0.0498	0.0560	0.0466	0.0479
			Kruskal–Wallis	0.0498	0.0476	0.0525	0.0474	0.0496
			Van der Waerden	0.0478	0.0468	0.0523	0.0454	0.0469
			Median	0.0430	0.0491	0.051	0.0472	0.0483
10	10	10	F (ANOVA)	0.0490	0.0475	0.0470	0.0508	0.0484
			Welch	0.0467	0.0468	0.0466	0.0516	0.0493
			Kruskal–Wallis	0.0435	0.0434	0.0485	0.0492	0.0494
			Van der Waerden	0.0424	0.0458	0.0457	0.0477	0.0492
			Median	0.0405	0.0475	0.0511	0.0541	0.0496
15	15	15	F (ANOVA)	0.0538	0.0514	0.0524	0.0528	0.0487
			Welch	0.0542	0.0502	0.0542	0.0516	0.0480
			Kruskal–Wallis	0.0498	0.0484	0.0509	0.0501	0.0457
			Van der Waerden	0.0495	0.0471	0.0490	0.0488	0.0451
			Median	0.0416	0.0462	0.0494	0.0515	0.0515
20	20	20	F (ANOVA)	0.0501	0.0512	0.0475	0.0485	0.0495
			Welch	0.0485	0.0519	0.0486	0.0474	0.0493
			Kruskal–Wallis	0.0454	0.0493	0.0458	0.0473	0.0464
			Van der Waerden	0.0447	0.0461	0.0446	0.0469	0.0477
			Median	0.0404	0.0515	0.0456	0.0503	0.0450
25	25	25	F (ANOVA)	0.0494	0.0479	0.0488	0.0473	0.0501
			Welch	0.0507	0.0467	0.0481	0.0481	0.0506
			Kruskal–Wallis	0.0451	0.0460	0.0460	0.0467	0.0492
			Van der Waerden	0.0455	0.0450	0.0459	0.0460	0.0474
			Median	0.0427	0.0478	0.0482	0.0513	0.0476
30	30	30	F (ANOVA)	0.0426	0.0457	0.0535	0.0468	0.0503
			Welch	0.0394*	0.0465	0.0522	0.0489	0.0510
			Kruskal–Wallis	0.0408	0.0443	0.0480	0.0450	0.0481
			Van der Waerden	0.0398*	0.0428	0.0492	0.0440	0.0477
			Median	0.0358*	0.0518	0.0481	0.0539	0.0471

หมายเหตุ * หมายถึง การทดสอบไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้

ตารางที่ 3 ค่าประมาณกำลังการทดสอบ สำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่มีการแจกแจงทวินาม

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
(n, p_1)	(n, p_2)	(n, p_3)	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
(10, 0.1)	(10, 0.125)	(10, 0.1)	F (ANOVA)	**0.0018**	**0.0827**	**0.1121**	**0.1368**	**0.1605**
			Welch	**0.0018**	0.0785	0.1098	0.1295	0.1566
			Kruskal–Wallis	0.0017	0.0760	0.1083	0.1273	0.1479
			Van der Waerden	0.0017	0.0786	0.1109	0.1325	0.1581
			Median	0.0016	0.0741	0.0958	0.1114	0.1310
(10, 0.1)	(10, 0.125)	(10, 0.15)	F (ANOVA)	0.0099	**0.1878**	**0.2385**	**0.3035**	**0.3496**
			Welch	**0.0107**	0.1877	0.2377	0.3035	0.3484
			Kruskal–Wallis	0.0097	0.1783	0.2282	0.2923	0.3323
			Van der Waerden	0.0096	0.1807	0.2350	0.3027	0.3467
			Median	0.0083	0.1491	0.1821	0.2281	0.2691

ตารางที่ 3 (ต่อ)

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
(n, p_1)	(n, p_2)	(n, p_3)	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
(10, 0.5)	(10, 0.525)	(10, 0.5)	F (ANOVA)	**0.0253**	0.0672	0.0720	**0.0796**	**0.0870**
			Welch	0.0250	**0.0690**	**0.0741**	0.0793	0.0864
			Kruskal–Wallis	0.0236	0.0652	0.0699	0.0741	0.0862
			Van der Waerden	0.0225	0.0629	0.0693	0.0760	0.0852
			Median	0.0224	0.0651	0.0671	0.0687	0.0763
(10, 0.7)	(10, 0.8)	(10, 0.7)	F (ANOVA)	0.0879	0.4920	0.6293	0.7445	0.8315
			Welch	**0.0887**	**0.5003**	**0.6408**	**0.7545**	**0.8397**
			Kruskal–Wallis	0.0842	0.4782	0.6225	0.7336	0.8195
			Van der Waerden	0.0833	0.4814	0.6235	0.7399	0.8274
			Median	0.0603	0.3527	0.4673	0.5524	0.6357
(30, 0.1)	(30, 0.125)	(30, 0.1)	F (ANOVA)	**0.0645**	**0.2095**	**0.2717**	**0.3331**	**0.3893**
			Welch	0.0604	0.1895	0.2539	0.3135	0.3696
			Kruskal–Wallis	0.0594	0.1874	0.2482	0.3098	0.3624
			Van der Waerden	0.0589	0.1915	0.2565	0.3189	0.3758
			Median	0.0441	0.1541	0.1908	0.2308	0.2752
(30, 0.1)	(30, 0.125)	(30, 0.15)	F (ANOVA)	**0.3211**	**0.4865**	**0.6318**	**0.7271**	**0.8070**
			Welch	0.3076	0.4711	0.6202	0.7190	0.8006
			Kruskal–Wallis	0.3002	0.4617	0.6012	0.7035	0.7840
			Van der Waerden	0.3034	0.4690	0.6172	0.7176	0.7978
			Median	0.2039	0.3417	0.4388	0.5283	0.6074
(30, 0.5)	(30, 0.525)	(30, 0.5)	F (ANOVA)	**0.0793**	**0.1113**	**0.1294**	**0.1571**	**0.1736**
			Welch	0.0789	0.1088	0.1264	0.1524	0.1722
			Kruskal–Wallis	0.0736	0.1030	0.1210	0.1465	0.1649
			Van der Waerden	0.0726	0.1029	0.1204	0.1480	0.1678
			Median	0.0618	0.0851	0.1015	0.1158	0.1240
(30, 0.7)	(30, 0.8)	(30, 0.7)	F (ANOVA)	**0.7894**	**0.9470**	**0.9856**	**0.9975**	0.9991
			Welch	0.7814	0.9433	0.9854	0.9971	**0.9992**
			Kruskal–Wallis	0.7674	0.9360	0.9811	0.9957	0.9990
			Van der Waerden	0.7709	0.9404	0.9835	0.9970	**0.9992**
			Median	0.5887	0.8207	0.916	0.9665	0.9856
(50, 0.1)	(50, 0.125)	(50, 0.1)	F (ANOVA)	**0.2204**	**0.3205**	**0.4163**	**0.5163**	**0.5856**
			Welch	0.2011	0.2968	0.3881	0.4920	0.5598
			Kruskal–Wallis	0.1991	0.2925	0.3807	0.4822	0.5522
			Van der Waerden	0.2031	0.2998	0.3954	0.4964	0.5681
			Median	0.1395	0.2266	0.2888	0.3620	0.4054
(50, 0.1)	(50, 0.125)	(50, 0.15)	F (ANOVA)	**0.5125**	**0.7180**	**0.8513**	**0.9279**	**0.9619**
			Welch	0.4862	0.7001	0.8393	0.9235	0.9601
			Kruskal–Wallis	0.4794	0.6908	0.8293	0.9130	0.9533
			Van der Waerden	0.4882	0.7007	0.8426	0.9194	0.9584
			Median	0.3194	0.5156	0.6502	0.7681	0.8373
(50, 0.5)	(50, 0.525)	(50, 0.5)	F (ANOVA)	**0.1108**	**0.1471**	**0.1899**	**0.2263**	**0.2676**
			Welch	0.1083	0.1437	0.1840	0.2199	0.2630
			Kruskal–Wallis	0.1026	0.1370	0.1806	0.2143	0.2497
			Van der Waerden	0.1036	0.1377	0.1820	0.2185	0.2608
			Median	0.0805	0.1124	0.1389	0.1635	0.1807

ตารางที่ 3 (ต่อ)

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
(n, p_1)	(n, p_2)	(n, p_3)	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
(50, 0.7)	(50, 0.8)	(50, 0.7)	F (ANOVA)	**0.9560**	**0.9972**	**0.9998**	**1.0000**	**1.0000**
			Welch	0.9494	0.9962	**0.9998**	**1.0000**	**1.0000**
			Kruskal–Wallis	0.9446	0.9959	0.9997	**1.0000**	**1.0000**
			Van der Waerden	0.9493	0.9961	0.9997	**1.0000**	**1.0000**
			Median	0.8084	0.9569	0.9916	0.9980	**1.0000**

หมายเหตุ ตัวเลขตัวหนา หมายถึง การทดสอบที่ให้กำลังการทดสอบสูงสุด

ตารางที่ 4 ค่าประมาณกำลังการทดสอบสำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่มีการแจกแจงปัวซง

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
λ_1	λ_2	λ_3	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
5	6	5	F (ANOVA)	**0.1468**	**0.2029**	**0.2671**	**0.3277**	**0.3732**
			Welch	0.1296	0.1871	0.2507	0.3107	0.3553
			Kruskal–Wallis	0.1325	0.1822	0.2465	0.3024	0.3476
			Van der Waerden	0.1307	0.1864	0.2521	0.3134	0.3600
			Median	0.0967	0.1451	0.1839	0.2282	0.2565
5	6	7	F (ANOVA)	**0.3266**	**0.4751**	**0.5996**	**0.7221**	**0.8083**
			Welch	0.3040	0.4636	0.5878	0.7126	0.8003
			Kruskal–Wallis	0.3019	0.4482	0.5734	0.6970	0.7857
			Van der Waerden	0.3018	0.4567	0.5832	0.7086	0.7982
			Median	0.1970	0.3221	0.4270	0.5250	0.6023
5	6	8	F (ANOVA)	**0.6201**	**0.8194**	**0.9277**	**0.9691**	**0.9876**
			Welch	0.5745	0.7924	0.9132	0.9622	0.9865
			Kruskal–Wallis	0.5794	0.7875	0.9076	0.9580	0.9835
			Van der Waerden	0.5848	0.7979	0.9151	0.9631	0.9863
			Median	0.4109	0.6246	0.7728	0.8585	0.9145
20	22	20	F (ANOVA)	**0.1450**	**0.2022**	**0.2685**	**0.3260**	**0.3820**
			Welch	0.1316	0.1938	0.2593	0.3115	0.3722
			Kruskal–Wallis	0.1287	0.1881	0.2510	0.3057	0.3642
			Van der Waerden	0.1289	0.1896	0.2557	0.3167	0.3696
			Median	0.0924	0.1512	0.1888	0.2288	0.2571
20	22	24	F (ANOVA)	**0.3520**	**0.5085**	**0.6457**	**0.7593**	**0.8354**
			Welch	0.3330	0.4937	0.6337	0.7491	0.8305
			Kruskal–Wallis	0.3270	0.4804	0.6145	0.7323	0.8185
			Van der Waerden	0.3287	0.4888	0.6307	0.7474	0.8269
			Median	0.2224	0.3539	0.4545	0.5600	0.6323
20	22	26	F (ANOVA)	**0.6799**	**0.8643**	**0.9529**	**0.9830**	**0.9932**
			Welch	0.6401	0.8427	0.9440	0.9816	0.9924
			Kruskal–Wallis	0.6367	0.8344	0.9384	0.9781	0.9909
			Van der Waerden	0.6432	0.8447	0.9467	0.9819	0.9928
			Median	0.4416	0.6732	0.8054	0.8920	0.9364
30	35	30	F (ANOVA)	**0.4837**	**0.6656**	**0.8074**	**0.8857**	**0.9431**
			Welch	0.4460	0.6326	0.7880	0.8720	0.9328
			Kruskal–Wallis	0.4427	0.6289	0.7789	0.8647	0.9258
			Van der Waerden	0.4464	0.6394	0.7931	0.8758	0.9351
			Median	0.3073	0.4675	0.6007	0.7140	0.7877

ตารางที่ 4 (ต่อ)

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2, n_3)				
λ_1	λ_2	λ_3	(Test)	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)	(25, 25, 25)	(30, 30, 30)
30	35	40	F (ANOVA)	**0.9001**	**0.9852**	**0.9986**	**0.9996**	**1.0000**
			Welch	0.8808	0.9827	0.9983	0.9995	**1.0000**
			Kruskal-Wallis	0.8774	0.9789	0.9974	0.9993	**1.0000**
			Van der Waerden	0.8820	0.9823	0.9983	0.9995	**1.0000**
			Median	0.6822	0.9033	0.9666	0.9898	0.9974
30	35	45	F (ANOVA)	**0.9984**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
			Welch	0.9967	**1.0000**	**1.0000**	**1.0000**	**1.0000**
			Kruskal-Wallis	0.9968	**1.0000**	**1.0000**	**1.0000**	**1.0000**
			Van der Waerden	0.9975	**1.0000**	**1.0000**	**1.0000**	**1.0000**
			Median	0.9557	0.9981	0.9997	**1.0000**	**1.0000**

หมายเหตุ ตัวเลขตัวหนา หมายถึง การทดสอบที่ให้กำลังการทดสอบสูงสุด

กำลังการทดสอบของการทดสอบทั้ง 5 วิธีมีค่าเพิ่มขึ้นตามไปด้วย

5. สรุปและวิจารณ์ผลการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการวิเคราะห์ความแปรปรวน การทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานสำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับที่มีการแจกแจงทวินามและการแจกแจงปัวซง โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของการทดสอบ คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ และกำหนดระดับนัยสำคัญของการศึกษาที่ 0.05

สำหรับผลการศึกษาความสามารถในการควบคุมความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ของการทดสอบทั้ง 5 วิธีพบว่า กรณีที่ข้อมูลมีการแจกแจงทวินาม การทดสอบทั้ง 5 วิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และความแปรปรวน (np(1 - p)) ของข้อมูลแต่ละกลุ่มน้อยกว่า 4 และกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 15 และ (n, p) เท่ากับ (10, 0.1) และ (10, 0.7) และกรณีที่ข้อมูลมีการแจกแจงปัวซง การทดสอบทั้ง 5 วิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกสถานการณ์ ยกเว้นการทดสอบของเวลช์ การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานที่ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากับ 10 และ λ = 30 เมื่อพิจารณากำลังการทดสอบพบว่า การวิเคราะห์ความแปรปรวนให้กำลังการทดสอบสูงกว่าการทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน เกือบทุกกรณีที่ศึกษา และมีบางกรณีที่การทดสอบของเวลช์ ให้กำลังการทดสอบสูงกว่าการวิเคราะห์ความแปรปรวนเล็กน้อย

จากผลการวิจัยในครั้งนี้จะเห็นได้ว่า สำหรับข้อมูลจำนวนนับที่มีการแจกแจงทวินามและการแจกแจงปัวซง การทดสอบเอฟของการวิเคราะห์ความแปรปรวนเป็นการทดสอบความแตกต่างระหว่างค่ากลางของประชากรมากกว่า 2 กลุ่มที่เป็นอิสระกันที่มีประสิทธิภาพดีกว่าการทดสอบของเวลช์ การทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐานในหลาย ๆ สถานการณ์ เนื่องจากสามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ได้ และมีกำลังการทดสอบที่สูงกว่าถึงแม้ว่าข้อมูลจะไม่เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวน กล่าวคือ ข้อมูลไม่ได้มาจากประชากรที่มีการแจกแจงปรกติ ความแปรปรวนของประชากรไม่เท่ากัน และข้อมูลไม่ได้มีมาตราการวัดอยู่ในมาตราช่วงเป็นอย่างน้อย ทั้งนี้ สาเหตุอาจเนื่องมาจากสถิติทดสอบเอฟเป็นสถิติทดสอบที่มีความแกร่งต่อการละเมิดข้อตกลงเบื้องต้น ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยบางส่วนของ Blanca et al. (2017) และ Klubnual (2018) ที่ได้ทำการศึกษาประสิทธิภาพของการวิเคราะห์ความแปรปรวนโดยใช้สถิติทดสอบเอฟสำหรับทดสอบค่าเฉลี่ยประชากรมากกว่า 2 กลุ่มเมื่อข้อมูลไม่มีการแจกแจงปรกติ โดยผลการวิจัยพบว่า สถิติทดสอบเอฟเป็นสถิติทดสอบที่มีความแกร่งและให้กำลังทดสอบสูงถึงแม้ว่าข้อมูลจะไม่ผ่านข้อตกลงเบื้องต้นของการแจกแจงปรกติ นอกจากนี้ยังพบว่า ความไม่เท่ากันของความแปรปรวนส่งผลต่อความแกร่งของสถิติทดสอบเอฟมากกว่าการแจกแจงไม่ปรกติของข้อมูล

ซึ่งสอดคล้องกับงานวิจัยหลาย ๆ งาน (Alexander & Govern, 1994; Büning, 1997; Gamage & Weerahandi, 1998; Harwell et al., 1992; Lee & Ahn, 2003; Lix et al., 1996; Moder, 2010; Patrick, 2007; Yiğit & Gökpinar, 2010; Zijlstra, 2004) อย่างไรก็ตาม ผลการวิจัยครั้งนี้บางส่วนไม่สอดคล้องกับงานวิจัยของ Hecke (2012) ที่ได้ทำการศึกษากำลังการทดสอบของการวิเคราะห์ความแปรปรวนและการทดสอบครัสคัล-วัลลิสและพบว่าการทดสอบครัสคัล-วัลลิส มีประสิทธิภาพที่ดีกว่าการวิเคราะห์ความแปรปรวนเมื่อข้อมูลมีการแจกแจงที่ไม่สมมาตร (เช่น การแจกแจงล็อกนอร์มัล) ทั้งนี้สาเหตุอาจเนื่องมาจากในงานวิจัยของ Hecke (2012) พิจารณาเฉพาะกรณีข้อมูลมีการแจกแจงแบบต่อเนื่อง ซึ่งอาจเป็นปัจจัยหนึ่งที่ส่งผลต่อกำลังการทดสอบ

นอกจากนี้ จากผลการวิจัยจะเห็นได้ว่ากำลังการทดสอบจะแปรผันตรงกับขนาดตัวอย่างและอัตราส่วนค่าเฉลี่ยของประชากรทั้ง 3 กลุ่ม กล่าวคือ เมื่อขนาดตัวอย่างหรืออัตราส่วนค่าเฉลี่ยของประชากรทั้ง 3 กลุ่มเพิ่มขึ้น กำลังการทดสอบของการทดสอบทั้ง 5 วิธี จะเพิ่มขึ้นตามไปด้วย (Everitt, 2002) ซึ่งเป็นไปตามคุณสมบัติของกำลังการทดสอบ สำหรับข้อมูลที่มีการแจกแจงปัวซง เมื่อค่าพารามิเตอร์ λ เพิ่มขึ้น การทดสอบทั้ง 5 วิธีมีแนวโน้มให้กำลังการทดสอบสูงขึ้น ทั้งนี้ สาเหตุอาจเนื่องมาจากเมื่อค่าพารามิเตอร์ λ เพิ่มขึ้นข้อมูลจะมีลักษณะที่สมมาตรมากขึ้นจนใกล้เคียงการแจกแจงปรกติหรือสามารถประมาณได้ด้วยการแจกแจงปรกติ ส่งผลให้การแจกแจงของตัวสถิติทดสอบทั้ง 5 วิธีเข้าใกล้การแจกแจงตามทฤษฎีมากขึ้น ในทำนองเดียวกัน สำหรับข้อมูลที่มีการแจกแจงทวินาม เมื่อค่า $p = 0.5$ หรือ np และ $n(1-p)$ มากกว่า 5 การทดสอบทั้ง 5 วิธีมีแนวโน้มที่จะให้ประสิทธิภาพที่ดีเนื่องจากข้อมูลมีลักษณะที่สมมาตรมากขึ้น (Lesch & Jeske, 2009)

ในทางปฏิบัติ สำหรับการทดสอบความแตกต่างระหว่างค่ากลางของประชากร 3 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับที่มีการแจกแจงทวินามและขนาดตัวอย่างแต่ละกลุ่มเท่ากันและมากกว่าหรือเท่ากับ 20 ผู้วิจัยหรือผู้วิเคราะห์ข้อมูลสามารถเลือกใช้สถิติทดสอบเอฟของการวิเคราะห์ความแปรปรวนแทนการใช้สถิติทดสอบไม่อิงพารามิเตอร์ เนื่องจากผลการวิจัยในครั้งนี้แสดงให้เห็นว่าสถิติทดสอบเอฟสามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ได้ในเกือบทุกสถานการณ์และให้กำลังการทดสอบสูงกว่าการทดสอบครัสคัล-วัลลิส การทดสอบแวน เดอ แวร์เดน และการทดสอบมัธยฐาน และสำหรับการทดสอบความแตกต่างระหว่างค่ากลางของประชากร 3 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับที่มีการแจกแจงปัวซง ผู้วิจัยหรือผู้วิเคราะห์ข้อมูลสามารถใช้สถิติทดสอบเอฟของการวิเคราะห์ความแปรปรวนแทนการใช้สถิติทดสอบไม่อิงพารามิเตอร์ได้เช่นเดียวกัน เนื่องจากผลการวิจัยในครั้งนี้แสดงให้เห็นว่าสถิติทดสอบเอฟเป็นสถิติทดสอบที่ให้ประสิทธิภาพที่ดีกว่าสถิติทดสอบอื่น ๆ ในทุกสถานการณ์

กิตติกรรมประกาศ

ผู้เขียนขอขอบพระคุณผู้ทรงคุณวุฒิทุกท่านที่ได้กรุณาสละเวลาในการอ่านและพิจารณาบทความ ให้ข้อคิดเห็นและคำชี้แนะที่เป็นประโยชน์และทำให้บทความนี้มีความถูกต้องและสมบูรณ์ยิ่งขึ้น

เอกสารอ้างอิง

Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational and Behavioral Statistics*, *19*, 91-101.

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. *Psicothema*, *29*(4), 552-557.

Bowarnkitiwong, S., & Areekul, K. (2017). Robustness of f-test for heterogeneity of population variances. *SDU Research Journal*, *13*(1), 1-16.

Büning, H. (1997). Robust analysis of variance. *Journal of Applied Statistics*, *24*, 319-332.

Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, *3*, 22-38.

Everitt, B. S. (2002). *The Cambridge Dictionary of Statistics*. Cambridge University Press.

Gamage, J., & Weerahandi, S. (1998). Size performance of some tests in one-way ANOVA. *Communications in Statistics - Simulation and Computation*, *27*, 625-640.

Glass, G., Peckham, P., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, *42*, 237-288.

Harrison, R. L. (2010). *Introduction to monte carlo simulation. AIP conference proceedings*, *1204*(1), 17-21.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fi xed effects ANOVA cases. *Journal of Educational and Behavioral Statistics*, *17*, 315-339.

Hecke, T. V. (2012). Power Study of ANOVA Versus Kruskal-Wallis Test. *Journal of Statistics and Management Systems*, *15*(2-3), 241-247.

Khomduean, J., & Araveeporn, A. (2017). Efficiency Comparison of Statistic for Testing Three Population Means in Case of Homogeneity and Heterogeneity of Variance. *Journal of Applied Statistics and Information Technology*, *2*(2), 61-77.

Khomduean, J., & Araveeporn, A. (2017). Efficiency Comparison of Statistic for Testing Three Population Means in Case of Heterogeneity of Variance. *Thai Science and Technology Journal*, *25*(6), 918-928.

Klubnual, P. (2018). The efficiency of parametric and non-parametric statistics on location testing with multiple population groups. *RMUTSB Academic Journal, 6*(1), 84-100.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*(260), 583-621.

Lee, S., & Ahn, C. H. (2003). Modified ANOVA for unequal variances. *Communications in Statistics - Simulation and Computation, 32*, 987-1004.

Lesch, S. M., & Jeske, D. R. (2009). Some Suggestions for Teaching About Normal Approximations to Poisson and Binomial Distribution Functions. *The American Statistician, 63*(3), 274-277.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research, 66*, 579-619.

Lüpsen, H. (2017). Comparison of nonparametric analysis of variance methods - A Vote for van der Waerden. *Communications in Statistics - Simulation and Computation, 30*, 1-30.

McDonald, J. H. (2014). *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

Moder, K. (2007). How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic. *Austrian Journal of Statistics, 36*(3), 179-188.

Moder, K. (2010). Alternatives to F-test in one way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling, 52*, 343-353.

Patrick, J. D. (2007). Simulations to analyze Type I error and power in the ANOVA F test and nonparametric alternatives [Unpublished master's thesis]. University of West Florida. http://etd.fcla.edu/WF/WFE0000158/Patrick_Joshua_Daniel_200905_MS.pdf

Sangthong, M. (2018). Efficiency Comparison of Parametric and Nonparametric Statistics on Central Testing Between Population Groups When Using Five-level Likert-type Scales. *Thai Science and Technology Journal, 28*(3), 403-418.

Scheffe, H. (1959). *The analysis of variance* (6th ed.). New York: Jonh Wiley & Sons.

Sheskin, D. J. (2000). *Hand book of Parametric and Nonparametric Statistical Procedures* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Sinsomboonthong, S. (2017). *Experimental Designs in Agriculture*. Bangkok, Thailand: Chamchuree Products Co., LTD.

Sinsomboonthong, S. (2020). *Nonparametric Statistics*. Bangkok, Thailand: Chamchuree Products Co., LTD.

Urawong, K., Chumnaul, J., Phaochamrun, C., & Singhabumrung, A. (2022). A comparing on performance of t-test and Wilcoxon rank sum test for two independent populations using counting data. *The Journal of Applied Science, 21*(1), 1-17.

Van der Waerden, B. L. (1952). Order tests for the two-sample problem and their power. *Indagationes Mathematicae, 14*, 453-458.

Van der Waerden, B. L. (1953). *Order tests for the two-sample problem. II, III, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Serie A, 564*, 303-310.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika, 38*(1), 330-336.

Yiğit, E., & Gökpınar, F. (2010). A Simulation study on tests for one-way ANOVA under the unequal variance assumption. *Communications Faculty of Sciences University of Ankara, Series A1, 59*, 15-34.

Zijlstra, W. (2004). Comparing the Student´s t and the ANOVA contrast procedure with five alternative procedures [Unpublished master's thesis]. Rijksuniversiteit Groningen.