การเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวนสำหรับข้อมูลที่มีการแจกแจงแบบไม่ปรกติ

An efficiency comparison of test statistics for testing homogeneity of variances for non-normally distributed data

สุกัญญา ยอดนวล[1], จุฬารัตน์ ชุมนวล[2] และ กรกช วิจิตรสงวน เจ็ดวรรณะ[2*]
Sukanya Yodnual[1], Julalat Chumnaul[2] and Korakot Wichitsa-nguan Jetwanna[2*]

Received: 24 February 2023; **Revised**: 11 April 2023; **Accepted**: 28 April 2023

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวนของประชากร 3 วิธี ได้แก่ สถิติทดสอบของ Levene สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen เมื่อข้อมูลไม่มีการแจกแจงปรกติ โดยจำนวนกลุ่มประชากรที่ศึกษา คือ 3 กลุ่ม การแจกแจงของประชากรที่ศึกษา คือ การแจกแจงแกมมา ((α, β) = (2, 2), (3, 2), (4, 2), (6, 2), (10, 2)) การแจกแจงไวบูล ((α, β) = (2, 6.105), (2, 7.478), (2, 8.635), (2, 10.575), (2, 13.652)) การแจกแจงลอจิสติก ((μ, s) = (2, 1.559), (2, 1.910), (2, 2.205), (2, 2.701), (2, 3.487)) และการแจกแจงเอกรูป ((a, b) = (0, 9.798), (0, 12), (0, 13.856), (0, 16.971), (0, 21.909)) และพิจารณาเฉพาะกรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบทั้ง 3 วิธี คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ความแกร่ง และกำลังการทดสอบ โดยสถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้และให้ค่าประมาณกำลังการทดสอบสูงที่สุดจะสรุปว่าเป็นสถิติทดสอบที่มีประสิทธิภาพดีที่สุด ผลการศึกษาพบว่า กรณีข้อมูลมีการแจกแจงลอจิสติกและการแจกแจงเอกรูป สถิติทดสอบของ Levene มีประสิทธิภาพดีกว่าสถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen เกือบทุกกรณีเนื่องจากสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้และให้ค่าประมาณกำลังการทดสอบสูงที่สุด ส่วนกรณีข้อมูลมีการแจกแจงไวบูล สถิติทดสอบของ Levene ยังคงมีประสิทธิภาพดีกว่าสถิติทดสอบทั้งสองตัวเมื่อตัวอย่างแต่ละกลุ่มมีขนาดตั้งแต่ 30 ขึ้นไป และกรณีข้อมูลมีการแจกแจงแกมมา สถิติทดสอบของ Fligner-Killeen มีประสิทธิภาพดีที่สุดในทุกกรณี

คำสำคัญ: ความเท่ากันของความแปรปรวน, ความน่าจะเป็นที่จะเกิดความผิดพลาดแบบที่ 1, ความแกร่ง, กำลังการทดสอบ

Abstract

This research aimed to study and compare the efficiency of three test statistics for testing homogeneity of variances, (Levene's test, Brown-Forsythe's test, and Figner-Killeen's test) when data are not normally distributed. The number of populations considered in this study was three groups and considered only when sample sizes of all groups were equal. The distributions considered in this study were Gamma distributions ((α, β) = (2, 2), (3, 2), (4, 2), (6, 2), (10, 2))), Weibull distributions ((α, β) = (2, 6.105), (2, 7.478), (2, 8.635), (2, 10.575), (2, 13.652)), Logistic distributions ((μ, s) = (2, 1.559), (2, 1.910), (2, 2.205), (2, 2.701), (2, 3.487)), and Uniform distributions ((a, b) = (0, 9.798), (0, 12), (0, 13.856), (0, 16.971), (0, 21.909)). The criteria used to compare the efficiency of proposed test statistics

[1] นักศึกษา, หลักสูตรวิทยาศาสตรบัณฑิต (สาขาวิชาสถิติ) สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์
[2] ผู้ช่วยศาสตราจารย์, สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์
[1] Bachelor of Science (Statistics), Division of Computational Science, Faculty of Science, Prince of Songkla University
[2] Assistant Professor, Division of Computational Science, Faculty of Science, Prince of Songkla University
* Corresponding author: Assistant Professor, Division of Computational Science, Faculty of Science, Prince of Songkla University, Email: korakot.w@psu.ac.th

were the ability to control the probability of type 1 error, robustness, and power of a test. In this study, the test statistic that could control the probability of type 1 error and had the highest empirical power was concluded to be the best test statistic. The results showed that Levene's test performed better than Brown-Forsythe's test and Figner-Killeen's test in almost all cases when data followed Logistic and Uniform distributions because it could control the probability of type 1 error and had the higher empirical power. In the case of Weibull distribution, Levene's test still performed better than other tests when the sample size of each group was 30 or more. Finally, Figner-Killeen's was the best in all cases when data followed the Gamma distribution.

Keywords: Homogeneity of variances, probability of type I error, robustness, power of a test

บทนำ

ในการวิเคราะห์ข้อมูลงานวิจัยโดยใช้สถิติเชิงอนุมาน (statistical inference) ผู้วิจัยจำเป็นต้องเลือกสถิติวิเคราะห์ให้เหมาะสมกับข้อมูล เพื่อให้ได้ข้อสรุปไปยังประชากรที่ถูกต้องและน่าเชื่อถือ ยกตัวอย่างเช่น การเปรียบเทียบความแตกต่างของค่าเฉลี่ยประชากรมากกว่า 2 กลุ่มโดยใช้สถิติทดสอบเอฟ (F-test) ของการวิเคราะห์ความแปรปรวน (analysis of variance) มีเงื่อนไขหรือข้อตกลงเบื้องต้น (assumptions) ที่สำคัญคือ ตัวอย่างต้องถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติ (normal distribution) และมีความแปรปรวนเท่ากัน (homogeneity of variances) แต่ในทางปฏิบัติเรามักพบว่าลักษณะของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น และหากนักวิจัยยังคงใช้การวิเคราะห์ความแปรปรวนด้วยสถิติทดสอบเอฟในการวิเคราะห์ข้อมูล ผลที่ได้อาจมีความคลาดเคลื่อน ดังนั้นก่อนการศึกษาเปรียบเทียบความแตกต่างของค่าเฉลี่ยประชากรมากกว่า 2 กลุ่มโดยใช้สถิติทดสอบเอฟของการวิเคราะห์ความแปรปรวน ผู้วิจัยจึงจำเป็นต้องตรวจสอบข้อตกลงเบื้องต้นของข้อมูลก่อนเสมอเพื่อให้ได้ผลการวิเคราะห์ข้อมูลที่ถูกต้องและน่าเชื่อถือ

สำหรับการทดสอบภาวะความเท่ากันของความแปรปรวนซึ่งเป็นหนึ่งในข้อตกลงเบื้องต้นของการเปรียบเทียบความแตกต่างของค่าเฉลี่ยประชากรที่เป็นอิสระกัน เป็นที่ทราบกันดีว่าในกรณีประชากร 2 กลุ่มมีการแจกแจงปรกติ สถิติทดสอบเอฟ เป็นสถิติทดสอบที่มีประสิทธิภาพดีที่สุด ส่วนกรณีประชากรมากกว่า 2 กลุ่ม และประชากรแต่ละกลุ่มมีการแจกแจงปรกติ สถิติทดสอบของ Bartlett (Bartlett's test) เป็นสถิติทดสอบที่มีประสิทธิภาพดีที่สุด อย่างไรก็ตาม หากข้อมูลที่นำมาวิเคราะห์มีการแจกแจงแบบอื่นหรือไม่ได้มาจากประชากรที่มีการแจกแจงปรกติ สถิติทดสอบดังกล่าวข้างต้นอาจมีประสิทธิภาพลดลง ยกตัวอย่างเช่น สถิติทดสอบของ Bartlett มีแนวโน้มไม่สามารถความคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ เมื่อข้อมูลเบนออกการแจกแจงปรกติ (Conover *et al*., 1981; Lim & Loh, 1996; Wang *et al*., 2017) เป็นต้น

ในหลายปีที่ผ่านมา นักสถิติหลายท่านได้คิดค้นและพัฒนาสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวนสำหรับกรณีประชากรไม่มีการแจกแจงปรกติ ยกตัวอย่างเช่น Levene (1960) ได้เสนอสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน โดยใช้การแปลงข้อมูลให้อยู่ในรูปของค่าสัมบูรณ์ของส่วนเบี่ยงเบนระหว่างค่าสังเกตกับค่าเฉลี่ยตัวอย่างที่ถูกสุ่มมาจากประชากรแต่ละชุด โดยจากการศึกษาพบว่า สถิติทดสอบของ Levene มีความแกร่งในกรณีที่ประชากรไม่มีการแจกแจงปรกติ ต่อมา Brown และ Forsythe (Brown & Forsythe, 1974) ได้พัฒนาสถิติทดสอบจากวิธีการของ Levene โดยใช้ค่ามัธยฐานแทนค่าเฉลี่ยตัวอย่าง โดยแทนค่าสังเกตแต่ละค่าด้วยค่าสัมบูรณ์ของส่วนเบี่ยงเบนระหว่างค่าสังเกตกับค่ามัธยฐานของตัวอย่าง ทำให้สถิติทดสอบนี้มีความแกร่งเมื่อประชากรไม่มีการแจกแจงปรกติเช่นเดียวกับสถิติทดสอบของ Levene (Conover *et al*., 1981) สำหรับในประเทศไทยมีนักสถิติหลายท่านได้ทำการศึกษาประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน ทั้งนี้ เพื่อเป็นแนวทางในการเลือกใช้สถิติทดสอบให้เหมาะสมกับลักษณะข้อมูล ยกตัวอย่างเช่น วรางคณา เรียนสุทธิ์ (2561) ได้ทำการเปรียบเทียบประสิทธิภาพของสถิติทดสอบอิงพารามิเตอร์สำหรับการทดสอบภาวะความเท่ากันของความแปรปรวน ผลการศึกษาพบว่า สถิติทดสอบของ Bartlett สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์และให้กำลังการทดสอบสูงเมื่อประชากรมีการแจกแจงปรกติ แต่เมื่อประชากรมีการแจกแจงเบ้ซ้ายและเบ้ขวาที่มีความโด่งสูงหรือต่ำมากกว่าปรกติ สถิติทดสอบของ Bartlett ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และเมื่อประชากรมีการแจกแจงเบ้ซ้ายหรือเบ้ขวาที่มีความโด่งสูงกว่าปรกติ สถิติทดสอบของ Brown-Forsythe สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้มากกว่าสถิติทดสอบของ Bartlett และสถิติทดสอบของ Levene (วรางคณา เรียนสุทธิ, 2561) ในขณะที่สายชล สินสมบูรณ์ทอง (2561) ได้ทำการเปรียบเทียบประสิทธิภาพ

ของสถิติทดสอบสำหรับทดสอบความเท่ากันของความแปรปรวนของประชากร 3 กลุ่ม ภายใต้การแจกแจงที่มีความโด่งมากและความเบ้มาก ผลการศึกษาพบว่า ในกรณีที่ข้อมูลมีการแจกแจงปรกติที่มีลักษณะโด่งมาก สถิติทดสอบของ Lehman (Lehman's test) มีกำลังการทดสอบสูงสุดในเกือบทุกสถานการณ์ ยกเว้นในกรณีตัวอย่างขนาดเล็ก สถิติทดสอบของ Bartlett มีกำลังการทดสอบสูงสุด ในกรณีที่ข้อมูลมีการแจกแจงแกมมาที่มีความโด่งมาก สถิติทดสอบของ Levene และสถิติทดสอบของ Bartlett มีกำลังการทดสอบสูงสุด (สายชล สินสมบูรณ์ทอง, 2561) และวรางคณา เรียนสุทธิ์ (2562) ได้ทำการเปรียบเทียบประสิทธิภาพของสถิติทดสอบไม่อิงพารามิเตอร์สำหรับการทดสอบภาวะความเท่ากันของความแปรปรวน ผลการศึกษาพบว่า สถิติทดสอบของ Mood (Mood's test) มีความเหมาะสมมากกว่าสถิติทดสอบของ Klotz (Klotz's test) เมื่อตัวอย่างมีขนาดเล็กและขนาดกลาง และข้อมูลมีการแจกแจงเบ้ซ้ายหรือเบ้ขวาที่มีความโด่งสูงกว่าปรกติ ส่วนในกรณีที่ความแตกต่างของความแปรปรวนประชากรน้อย หรือตัวอย่างมีขนาดกลางที่มีขนาดเท่ากันและข้อมูลมีการแจกแจงเบ้ซ้ายหรือเบ้ขวาที่มีความโด่งต่ำกว่าปรกติ หรือตัวอย่างมีขนาดใหญ่ และข้อมูลมีการแจกแจงปรกติหรือมีการแจกแจงเบ้ซ้ายหรือเบ้ขวาที่มีความโด่งต่ำกว่าปรกติ สถิติทดสอบของ Klotz มีความเหมาะสมมากกว่าสถิติทดสอบของ Mood (วรางคณา เรียนสุทธิ, 2562)

ในการวิจัยครั้งนี้ ผู้วิจัยสนใจศึกษาและเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน 3 วิธี ได้แก่ สถิติทดสอบของ Levene สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบ คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 (probability of type I error) ความแกร่ง (robustness) และกำลังการทดสอบ (power of a test) ทั้งนี้ เพื่อหาสถิติทดสอบที่มีประสิทธิภาพดีที่สุดในการทดสอบภาวะความเท่ากันของความแปรปรวนกรณีที่ข้อมูลไม่มีการแจกแจงปรกติและเป็นแนวทางในการเลือกใช้สถิติวิเคราะห์ให้เหมาะสมกับข้อมูล

สถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน

ในการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน 3 วิธี ได้แก่ สถิติทดสอบของ Levene สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen โดยมีสมมุติฐานการทดสอบคือ

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$$H_1: \sigma_i^2 \neq \sigma_j^2$$

อย่างน้อย 1 คู่ โดยที่ $i = j$ และ $i,j = 1,2,3$ (1)

สำหรับสถิติทดสอบทั้ง 3 วิธีมีรายละเอียดดังต่อไปนี้

1. สถิติทดสอบของ Levene (Levene's test)

Levene (1960) ได้เสนอสถิติทดสอบแบบอิงพารามิเตอร์ที่ใช้สำหรับทดสอบภาวะความเท่ากันของความแปรปรวนสำหรับประชากร k กลุ่ม ซึ่งสถิติทดสอบนี้จะมีความแกร่งในกรณีที่ประชากรไม่มีการแจกแจงปรกติ โดยค่าสถิติทดสอบของ Levene สามารถคำนวณได้ดังสมการ (2)

$$L = \frac{(n-k)}{(k-1)} \left(\frac{\sum_{i=1}^{k} n_i (\overline{z}_{i.} - \overline{z})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} n_i (z_{ij} - \overline{z}_{i.})^2} \right) \quad (2)$$

โดยที่ $z_{ij} = (x_{ij} - \overline{x}_i)^2$, $\overline{z}_{i.} = \frac{\sum_{j=1}^{n_i} z_{ij}}{n_i}$, $\overline{z} = \frac{\sum_{i=1}^{k} n_i \overline{z}_{i.}}{n}$

เมื่อ

n คือ จำนวนข้อมูลทั้งหมด

n_i คือ ขนาดตัวอย่างกลุ่มที่ i

$\overline{X}_i$ คือ ค่าเฉลี่ยตัวอย่างกลุ่มที่ i

k คือ จำนวนกลุ่มของประชากร

สำหรับสถิติทดสอบของ Levene มีการแจกแจงใกล้เคียงกับการแจกแจงเอฟ (F distribution) โดยมีองศาเสรีเท่ากับ $k - 1$ และ $n - k$ ดังนั้น ในการทดสอบสมมุติฐานภาวะความเท่ากันของความแปรปรวน เราจะปฏิเสธสมมุติฐาน H_0 ใน (1) เมื่อ $L \geq f_{\alpha;k-1,n-k}$ โดย α คือ ระดับนัยสำคัญของการทดสอบ

2. สถิติทดสอบของ Brown-Forsythe (Brown-Forsythe's test)

Brown and Forsythe (1974) สายชล สินสมบูรณ์ทอง, 2561) ได้เสนอสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวนที่ได้พัฒนามาจากสถิติทดสอบของ Levene โดยใช้ค่ามัธย-ฐานแทนค่าเฉลี่ยตัวอย่าง ด้วยเหตุนี้จึงทำให้สถิติทดสอบของ Brown-Forsythe มีความแกร่งในกรณีที่ประชากรไม่มีการแจกแจงปรกติ โดยค่าสถิติทดสอบของ Brown-Forsythe สามารถคำนวณได้ดังสมการ (3)

$$BF = \frac{\sum_{i=1}^{k} n_i(\bar{z}_i - \bar{z})^2 \Big/ (k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_j} n_i(\bar{z}_{ij} - \bar{z})^2 \Big/ \sum_{i=1}^{k}(n_i - 1)} \quad (3)$$

โดยที่ $\bar{z}_i = | x_{ij} - \tilde{x}_{ij} |$ และ $\tilde{x}_{ij}$ คือ มัธยฐานของตัวอย่างกลุ่มที่ i

สำหรับการทดสอบสมมุติฐานภาวะความเท่ากันของความแปรปรวน เราจะปฏิเสธสมมุติฐาน H_0 ใน (1) เมื่อ $BF \geq f_{\alpha;\, k-1,\, n-k}$

3. สถิติทดสอบของ Fligner-Killeen (Fligner-Killeen's test)

สถิติทดสอบของ Fligner และ Killeen เป็นสถิติทดสอบแบบไม่อิงพารามิเตอร์สำหรับทดสอบภาวะความเท่ากันของความแปรปรวน ซึ่งเหมาะสำหรับข้อมูลที่ไม่มีการแจกแจงปรกติ เป็นสถิติทดสอบที่แก้ไขโดย Conover *et al.* (1981) โดยการใช้อันดับของ $| x_{ij} - \tilde{x}_j |$ เมื่อ $\tilde{x}_j$ คือ มัธยฐานของประชากรที่ j ดังนั้น สถิติทดสอบนี้จึงถูกเรียกว่า สถิติทดสอบของ Fligner-Killeen แบบมัธยฐาน (Niu, 2004) โดยค่าสถิติทดสอบของ Levene สามารถคำนวณได้ดังสมการ (4)

$$FK = \frac{\sum_{j=1}^{k} n_j - (\bar{A}_j - \bar{a})^2}{V^2} \quad (4)$$

โดยที่ $\bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_{N,i}$

และ $V^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_{N,i} - \bar{a})^2$

เมื่อ

$\bar{A}_j$ คือ คะแนนเฉลี่ยสำหรับตัวอย่างอันดับที่ j

$\bar{a}$ คือ คะแนนเฉลี่ยรวม

สำหรับตัวอย่างขนาดใหญ่ สถิติทดสอบของ Fligner-Killeen จะมีการแจกแจงแบบไคกำลังสอง (Chi-square distribution) โดยมีองศาเสรีเท่ากับ $k-1$

เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบ

สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบทั้ง 3 วิธีจะพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ความแกร่งและกำลังการทดสอบ โดยมีรายละเอียดดังนี้

1. ความน่าจะเป็นของความผิดพลาดแบบที่ 1 (Probability of type I error)

ความผิดพลาดแบบที่ 1 (type I error) คือ ความผิดพลาดที่เกิดจากการปฏิเสธสมมุติฐานหลัก (H_0) เมื่อสมมุติฐานหลักเป็นจริง โดยความน่าจะเป็นของความผิดพลาดแบบที่ 1 เขียนแทนด้วย α

สำหรับการวิจัยครั้งนี้ จะใช้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 (empirical probability of type I error, $\hat{\alpha}$) ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบ โดยค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 จะคำนวณจากจำนวนครั้งในการปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นจริงหารด้วยจำนวนครั้งในการทำซ้ำ (10,000 รอบ) ถ้าสถิติทดสอบใดให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley (Bradley, 1978) กล่าวคือ ให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วง [0.025, 0.075] ที่ระดับนัยสำคัญ 0.05 จะสรุปว่า สถิติทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้

2. ความแกร่ง (robustness)

ความแกร่ง เป็นคุณสมบัติที่ตัวสถิติทดสอบใดๆ ไม่ไวเมื่อข้อมูลมีการเบี่ยงเบนไปจากข้อสมมุติที่กำหนดหรือยังคงสมบัติเดิม

สำหรับการวิจัยครั้งนี้ จะใช้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ($\hat{\alpha}$) ในการพิจารณาความแกร่งของสถิติทดสอบ ถ้าสถิติทดสอบใดให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Cochran (Cochran, 1952) กล่าวคือ ให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วง [0.04, 0.06] ที่ระดับนัยสำคัญ 0.05 จะสรุปว่า สถิติทดสอบนั้นมีความแกร่ง

3. กำลังการทดสอบ (Power of a test)

กำลังการทดสอบ คือ ความน่าจะเป็นของการปฏิเสธสมมุติฐานหลัก (H_0) เมื่อสมมุติฐานหลักไม่เป็นจริง เขียนแทนด้วย $1 - \beta$

สำหรับการวิจัยครั้งนี้ จะใช้ค่าประมาณกำลังการทดสอบ ($1 - \hat{\beta}$) ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบ โดยค่าประมาณกำลังการทดสอบจะคำนวณจากจำนวนครั้งในการปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นเท็จหารด้วยจำนวนครั้งในการทำซ้ำ (10,000 รอบ) ถ้าสถิติทดสอบใดสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้และให้ค่าประมาณกำลังการทดสอบสูงสุดจะสรุปว่าสถิติทดสอบนั้นมีประสิทธิภาพดีที่สุด

วิธีการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงจำลองข้อมูล (simulation study) ด้วยโปรแกรม RStudio โดยผู้วิจัยได้กำหนดขอบเขตและขั้นตอนการดำเนินการวิจัยดังนี้

1. ขอบเขตการวิจัย

ผู้วิจัยได้กำหนดขอบเขตการดำเนินการวิจัยดังนี้

1. กำหนดจำนวนกลุ่มประชากรที่ศึกษาเท่ากับ 3 กลุ่มและขนาดตัวอย่างแต่ละกลุ่มเท่ากันโดยกำหนดขนาดตัวอย่างครอบคลุมทั้งกรณีตัวอย่างขนาดเล็ก ขนาดปานกลาง และขนาดใหญ่ รายละเอียดดัง Table 1

Table 1 Sample sizes considered in this study.

Cases	Sample sizes (n_1, n_2, n_3)
Small	(5, 5, 5), (10, 10, 10)
Medium	(20, 20, 20), (30, 30, 30)
Large	(70, 70, 70), (100, 100, 100)

2. กำหนดความแตกต่างของความแปรปรวนโดยใช้ค่าพารามิเตอร์ไม่มีศูนย์กลาง (non-centrality parameter, ϕ) เป็นเกณฑ์วัดความแตกต่างของความแปรปรวนของประชากร รายละเอียดดัง Table 2

Table 2 Population variance ratios based on ϕ.

Levels	Ratios	ϕ
Slightly ($0 < \phi < 1.5$)	8:12:16	1.155
Moderately ($1.5 \le \phi < 3$)	8:16:24	2.309
Highly ($\phi \ge 3$)	8:24:40	4.619

3. กำหนดการแจกแจงของประชากรที่ศึกษา 4 การแจกแจง คือ การแจกแจงแกมมา (Gamma distribution) การแจกแจงไวบูล (Weibull distribution) การแจกแจงลอจิสติก (logistic distribution) และการแจกแจงเอกรูป (uniform distribution) รายละเอียดการกำหนดค่าพารามิเตอร์ของการแจกแจงต่างๆ เพื่อให้ข้อมูลมีความเบ้ความโด่งที่แตกต่างกัน แสดงดัง Table 3-6 และลักษณะการแจกแจงต่างๆ ภายใต้พารามิเตอร์ที่ศึกษาแสดงดัง Figure 1-4

4. กำหนดระดับนัยสำคัญ (α) เท่ากับ 0.05

5. กำหนดเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสถิติทดสอบ คือ ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ความแกร่งและกำลังการทดสอบ

6. โปรแกรมที่ใช้ในการศึกษาคือ โปรแกรม R-Studio เวอร์ชั่น 4.2.2

Table 3 Parameters of Gamma distribution.

(α, β)	$E(X)$*	$Var(X)$**
(2, 2)	4	8
(3, 2)	6	12
(4 2)	8	16
(6, 2)	12	24
(10, 2)	20	40

* $E(X) = \alpha\beta$, ** $Var(X) = \alpha\beta^2$

Table 4 Parameters of Weibull distribution.

(α, β)	$E(X)$*	$Var(X)$**
(2, 6.105)	5.41	8
(2, 7.478)	6.63	12
(2, 8.635)	7.65	16
(2, 10.575)	9.37	24
(2, 13.652)	12.10	40

* $E(X) = \alpha\Gamma(1+1/\beta)$, ** $Var(X) = \alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2\right]$

Table 5 Parameters of Logistic distribution.

(μ, s)	$E(X)$*	$Var(X)$**
(2, 1.559)	2	8
(2, 1.910)	2	12
(2, 2.205)	2	16
(2, 2.701)	2	24
(2, 3.487)	2	40

* $E(X) = \mu$, ** $Var(X) = \frac{s^2\pi^2}{3}$

Table 6 Parameters of Uniform distribution.

(a, b)	$E(X)^*$	$Var(X)^{**}$
(0, 9.798)	2	8
(0, 12)	2	12
(0, 13.856)	2	16
(0, 16.971)	2	24
(0, 21.909)	2	40

* $E(X) = \frac{a+b}{2}$, ** $Var(X) = \frac{1}{12}(b-a)^2$

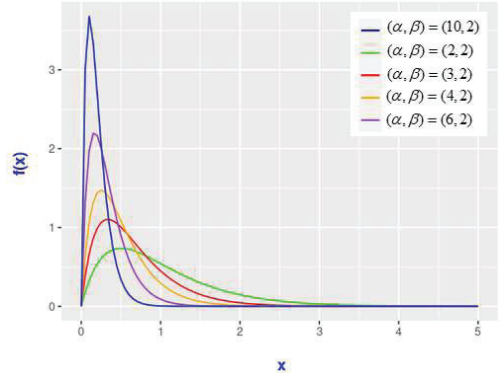

Figure 1 Gamma distributions under various parameters.

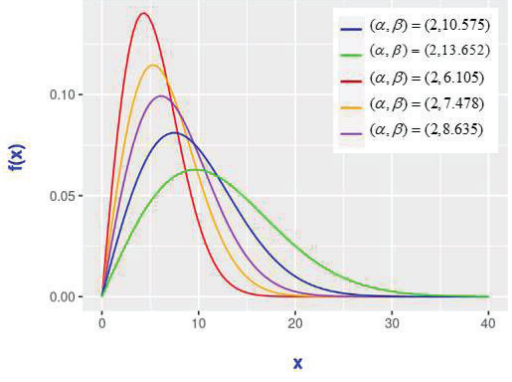

Figure 2 Weibull distributions under various parameters.

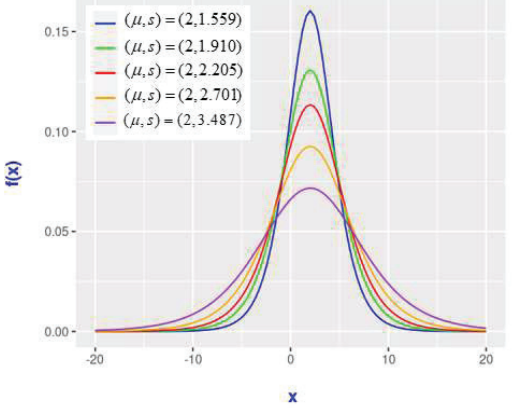

Figure 3 Logistic distributions under various parameters.

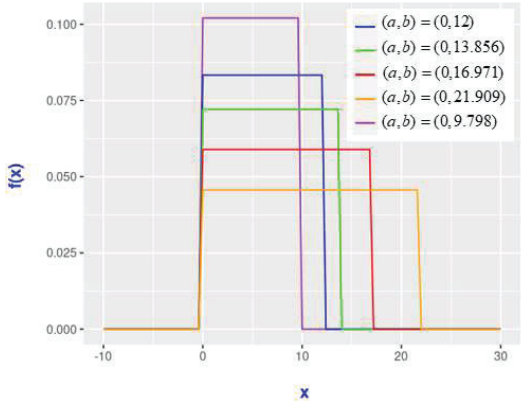

Figure 4 Uniform distributions under various parameters.

2. ขั้นตอนการดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัยแยกเป็น 2 กรณี ดังนี้

2.1 การศึกษาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1

ขั้นตอนการจำลองข้อมูลเพื่อศึกษาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบทั้ง 3 วิธีมีดังต่อไปนี้

1) กำหนดสมมุติฐานสำหรับการทดสอบภาวะความเท่ากันของความแปรปรวนดังนี้

$$H_0 : \sigma^2_1 = 8, \sigma^2_2 = 8, \sigma^2_3 = 8$$
$$H_1 : \sigma^2_i \neq 8, \sigma^2_j \text{ อย่างน้อย 1 คู่}$$

โดยที่ $i \neq j$ และ $i, j = 1, 2, 3$

โดยที่ σ^2_1, σ^2_2, σ^2_3 คือ ความแปรปรวนของประชากรกลุ่มที่ 1, 2, และ 3 ตามลำดับ

2) จำลองข้อมูลภายใต้สมมุติฐานหลัก (H_0) ในข้อ 1. ที่เป็นจริง โดยกำหนดค่าของเลขสุ่มเริ่มต้นคือ 23456

3) คำนวณค่าสถิติทดสอบและ *p*-value ทั้ง 3 วิธี

4) เปรียบเทียบ *p*-value กับระดับนัยสำคัญ 0.05 โดยถ้า *p*-value น้อยกว่าหรือเท่ากับระดับนัยสำคัญ จะตัดสินใจปฏิเสธสมมุติฐานหลัก และถ้า *p*-value มากกว่าระดับนัยสำคัญจะตัดสินใจยอมรับสมมุติฐานหลัก

5) บันทึกจำนวนครั้งที่ปฏิเสธสมมุติฐานหลัก

6) ทำซ้ำข้อ 1-4 จำนวน 10,000 ครั้ง

7) คำนวณค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบแต่ละตัว

8) เปรียบเทียบค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบทั้ง 3 วิธีกับเกณฑ์ของ Bradley ถ้าค่าประมาณความน่าจะเป็นของความผิดพลาด

แบบที่ 1 อยู่ในช่วง [0.5 α, 1.5 α] หรือ [0.025, 0.075] จะสรุปว่า สถิติทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และสถิติทดสอบที่ให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วง [0.04, 0.06] ตามเกณฑ์ของ Cochran จะสรุปว่า สถิติทดสอบนั้นมีความแกร่ง

2.2 การศึกษากำลังการทดสอบ

ขั้นตอนการจำลองข้อมูลเพื่อศึกษากำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธีมีดังต่อไปนี้

1) กำหนดสมมุติฐานสำหรับการทดสอบภาวะความเท่ากันของความแปรปรวน 3 กรณี ดัง Table 7

Table 7 Hypotheses for power study.

Cases	Hypotheses	ϕ
1	$H_0 : \sigma^2_1, \sigma^2_2, \sigma^2_3$ $H_1 : \sigma^2_1 = 8, \sigma^2_2 = 12, \sigma^2_3 = 16$	1.155
2	$H_0 : \sigma^2_1, \sigma^2_2, \sigma^2_3$ $H_1 : \sigma^2_1 = 8, \sigma^2_2 = 16, \sigma^2_3 = 24$	2.309
3	$H_0 : \sigma^2_1, \sigma^2_2, \sigma^2_3$ $H_1 : \sigma^2_1 = 8, \sigma^2_2 = 24, \sigma^2_3 = 40$	4.619

2) จำลองข้อมูลภายใต้สมมุติฐานทางเลือก (H_1) ในข้อ 1. ที่เป็นจริง โดยกำหนดค่าของเลขสุ่มเริ่มต้นคือ 23456

3) คำนวณค่าสถิติทดสอบและ *p*-value ทั้ง 3 วิธี

4) เปรียบเทียบ *p*-value กับระดับนัยสำคัญ 0.05 โดยถ้า *p*-value น้อยกว่าหรือเท่ากับระดับนัยสำคัญ จะตัดสินใจปฏิเสธสมมุติฐานหลัก และถ้า *p*-value มากกว่าระดับนัยสำคัญจะตัดสินใจยอมรับสมมุติฐานหลัก

5) บันทึกจำนวนครั้งที่ปฏิเสธสมมุติฐานหลัก

6) ทำซ้ำข้อ 1-4 จำนวน 10,000 ครั้ง

7) คำนวณค่าประมาณกำลังการทดสอบ (empirical power) ของสถิติทดสอบแต่ละตัว

8) เปรียบเทียบค่าประมาณกำลังการทดสอบ เฉพาะสถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เท่านั้น โดยสถิติทดสอบที่มีค่าประมาณกำลังการทดสอบสูงสุด จะสรุปว่าสถิติทดสอบนั้นเป็นสถิติทดสอบที่มีประสิทธิภาพดีที่สุด

ผลการวิจัย

ผลการวิจัยจำแนกออกเป็น 2 ส่วนดังนี้

1. ผลการศึกษาความสามารถในการควบคุมความน่าเป็นของความผิดพลาดแบบที่ 1

ผลการศึกษาความสามารถในการควบคุมความน่าเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบทั้ง 3 วิธี แสดงดัง Table 8-11 และ Figure 5

Table 8 Empirical probability of type I error when data follow Gamma distribution.

Sample sizes	Test statistics		
	L	BF	FK
(5, 5, 5)	0.1407	0.0087	0.0003
(10, 10, 10)	0.1363	0.0448*	0.0531*
(20, 20, 20)	0.1289	0.0419*	0.0556*
(30, 30, 30)	0.1244	0.0455*	0.0616*
(70, 70, 70)	0.1266	0.0489*	0.0724*
(100, 100, 100)	0.1247	0.0480*	0.0754

Note: L = Levene's test, BF = Brown-Forsythe's test,
FK = Fligner-Killeen's test, * Probability of type I error in control

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงแกมมา ผลการศึกษาพบว่า สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้นกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5)) ส่วนสถิติทดสอบของ Levene ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สูงกว่าระดับนัยสำคัญของการทดสอบทุกกรณี อย่างไรก็ตาม สถิติทดสอบของ Levene มีแนวโน้มที่จะให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น (Table 8)

Table 9 Empirical probability of type I error when data follow Weibull distribution.

Sample sizes (n_1, n_2, n_3)	Test statistics		
	L	BF	FK
(5, 5, 5)	0.0966	0.0038	0.0002
(10, 10, 10)	0.0816	0.0363*	0.0351*
(20, 20, 20)	0.0754	0.0376*	0.0369*
(30, 30, 30)	0.0725*	0.0398*	0.0405*
(70, 70, 70)	0.0686*	0.0460*	0.0483*
(100, 100, 100)	0.0685*	0.0483*	0.0510*

Note: L = Levene's test, BF = Brown-Forsythe's test, FK = Fligner-Killeen's test, * Probability of type I error in control

Table 10 Empirical probability of type I error when data follow Logistic distribution.

Sample sizes (n_1, n_2, n_3)	Test statistics		
	L	BF	FK
(5, 5, 5)	0.0859	0.0048	0.0002
(10, 10, 10)	0.0690*	0.0342*	0.0339*
(20, 20, 20)	0.0595*	0.0392*	0.0386*
(30, 30, 30)	0.0531*	0.0404*	0.0398*
(70, 70, 70)	0.0532*	0.0472*	0.0450*
(100, 100, 100)	0.0530*	0.0494*	0.0471*

Note: L = Levene's test, BF = Brown-Forsythe's test, FK = Fligner-Killeen's test, * Probability of type I error in control

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงไวบูล ผลการศึกษาพบว่า สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้นกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5)) ส่วนสถิติทดสอบของ Levene ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5) และ (10, 10, 10)) และขนาดกลาง ((n_1, n_2, n_3) = (20, 20, 20)) และให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สูงกว่าระดับนัยสำคัญของการทดสอบทุกกรณี อย่างไรก็ตาม สถิติทดสอบของ Levene มีแนวโน้มที่จะให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นเช่นเดียวกับกรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงแกมมา (Table 9)

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจง ลอจิสติก ผลการศึกษาพบว่า สถิติทดสอบของ Levene สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้นกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5)) (Table 10)

Table 11 Empirical probability of type I error when data follow Uniform distribution.

Sample sizes (n_1, n_2, n_3)	Tests		
	L	BF	FK
(5, 5, 5)	0.0906	0.0019	0.0002
(10, 10, 10)	0.0704*	0.0303*	0.0245
(20, 20, 20)	0.0573*	0.0288*	0.0265*
(30, 30, 30)	0.0548*	0.0316*	0.0294*
(70, 70, 70)	0.0496*	0.0368*	0.0398*
(100, 100, 100)	0.0521*	0.0436*	0.0440*

Note: L = Levene's test, BF = Brown-Forsythe's test, FK = Fligner-Killeen's test, * Probability of type I error in control

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงเอกรูป ผลการศึกษาพบว่า สถิติทดสอบของ Levene และสถิติทดสอบของ Brown-Forsythe สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้นกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5)) และสถิติทดสอบของ Fligner-Killeen ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีตัวอย่าง (n_1, n_2, n_3) มีขนาด (5, 5, 5) และ (10, 10, 10) (Table 11)

2. ผลการศึกษาความแกร่ง

จาก Table 8-11 พบว่า กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงแกมมา สถิติทดสอบของ Brown-Forsythe มีความแกร่งเกือบทุกสถานการณ์ ยกเว้นกรณีตัวอย่างขนาดเล็ก ((n_1, n_2, n_3) = (5, 5, 5)) สำหรับกรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงไวบูลและการแจกแจงลอจิสติก สถิติ

ทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen มีความแกร่งเมื่อตัวอย่างมีขนาดใหญ่ ((n_1, n_2, n_3) = (70, 70, 70) และ (100, 100, 100)) ส่วนสถิติทดสอบของ Levene มีความแกร่งในกรณีที่ประชากรทั้ง 3 กลุ่มมีการแจกแจงลอจิสติกและการแจกแจงเอกรูป และตัวอย่างแต่ละกลุ่มมีขนาด 20 ขึ้นไป

3. ผลการศึกษากำลังการทดสอบ

ผลการศึกษากำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธี แสดงดัง Table 12 และ Figure 6

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงแกมมา ผลการศึกษาพบว่า สถิติทดสอบของ Fligner-Killeen มีประสิทธิภาพดีที่สุดเนื่องจากให้กำลังการทดสอบสูงสุดทุกกรณี

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงไวบูล สถิติทดสอบของ Brown-Forsythe ให้กำลังการทดสอบสูงสุดสำหรับทุกค่า ϕ เมื่อตัวอย่างมีขนาดเล็ก สำหรับตัวอย่างขนาดกลางและขนาดใหญ่ สถิติทดสอบของ Levene ให้กำลังการทดสอบสูงสุดในกรณีค่า ϕ เท่ากับ 1.155 และ 2.309 ส่วนกรณีค่า ϕ เท่ากับ 4.619 สถิติทดสอบของ Levene ให้กำลังการทดสอบสูงสุดสำหรับตัวอย่างขนาดกลาง และเมื่อตัวอย่างมีขนาดใหญ่ สถิติทดสอบทั้ง 3 วิธี ให้กำลังการทดสอบสูงไม่แตกต่างกัน

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงลอจิสติก สถิติทดสอบของ Levene มีประสิทธิภาพดีที่สุดเนื่องจากให้กำลังการทดสอบสูงสุดทุกกรณี รองลงมาคือ สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen ตามลำดับ

กรณีประชากรทั้ง 3 กลุ่มมีการแจกแจงเอกรูป สถิติทดสอบของ Levene ให้กำลังการทดสอบสูงสุดเกือบทุกกรณี ยกเว้นกรณีตัวอย่างขนาดใหญ่และค่า ϕ เท่ากับ 1.155 และ 2.309 สถิติทดสอบของ Fligner-Killeen ให้กำลังการทดสอบสูงสุด และสถิติทดสอบทั้ง 3 วิธี ให้กำลังการทดสอบสูงไม่แตกต่างกันเมื่อตัวอย่างมีขนาดใหญ่และค่า ϕ เท่ากับ 4.619

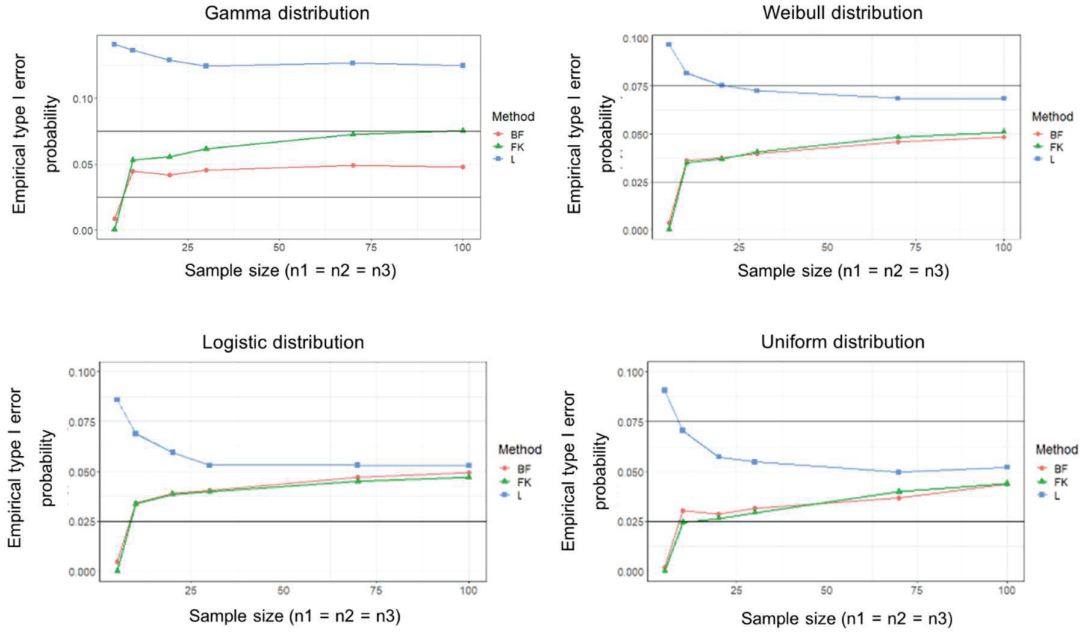

Figure 5 Empirical probability of type I error of three test statistics when data follow Gamma, Weibull, Logistic, and Uniform distribution.

Table 12 Empirical power of Levene's test, Brown-Forsythe's test and Fligner-Killeen's test.

Sample Sizes	ϕ	Gamma distribution			Weibull distribution			Logistic distribution			Uniform distribution		
		L	BF	FK	L	BF	FK	L	BF	FK	L	BF	FK
(5, 5, 5)	1.155	-	-	-	-	-	-	-	-	-	-	-	-
	2.309	-	-	-	-	-	-	-	-	-	-	-	-
	4.619	-	-	-	-	-	-	-	-	-	-	-	-
(10, 10, 10)	1.155	-	0.0852	0.0917*	-	0.0800*	0.0756	0.1269*	0.0725	0.0697	0.1782*	0.0893	0.0785
	2.309	-	0.1568	0.1626*	-	0.1496*	0.1390	0.2106*	0.1311	0.1242	0.3380*	0.1866	0.1699
	4.619	-	0.2918	0.3035*	-	0.285*	0.2723	0.3716*	0.2452	0.2362	0.5950*	0.3797	0.3461
(20, 20, 20)	1.155	-	0.1681	0.1894*	-	0.1608*	0.1554	0.1834*	0.1402	0.1354	0.3244*	0.2171	0.2344
	2.309	-	0.3767	0.3960*	-	0.3731*	0.3586	0.3853*	0.3148	0.2964	0.6677*	0.5343	0.5411
	4.619	-	0.7059	0.7149*	-	0.6945*	0.6821	0.6990*	0.6248	0.6041	0.9378*	0.8833	0.8688
(30, 30, 30)	1.155	-	0.2628	0.2964*	0.3384*	0.2553	0.2519	0.2592*	0.2209	0.2058	0.4746*	0.3812	0.4420
	2.309	-	0.5800	0.6093*	0.6686*	0.5829	0.5672	0.5597*	0.5099	0.4859	0.8664*	0.8040	0.8338
	4.619	-	0.9064	0.9150*	0.9435*	0.9138	0.9007	0.8820*	0.8530	0.8345	0.9955*	0.9909	0.9884
(70, 70, 70)	1.155	-	0.5799	0.6482*	0.6724*	0.6151	0.6092	0.5576*	0.5374	0.5083	0.8750	0.8501	0.9344*
	2.309	-	0.9458	0.9610*	0.9711*	0.9610	0.9544	0.9295*	0.9242	0.9105	0.9990	0.9987	0.9997*
	4.619	-	0.9996	0.9997*	0.9999	0.9999	0.9998	0.9995*	0.9994	0.9987	1.0000	1.0000	1.0000
(100, 100, 100)	1.155	-	0.7571	0.8190*	0.8289*	0.7973	0.7903	0.7279*	0.7168	0.6959	0.9676	0.9622	0.9914*
	2.309	-	0.9906	0.9948*	0.9971*	0.9956	0.9949	0.9872*	0.9861	0.9811	1.0000	1.0000	1.0000
	4.619	-	0.9999	1.0000*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: L = Levene's test, BF = Brown-Forsythe's test, FK = Fligner-Killeen's test, - Probability of type I error not in control, * Highest power

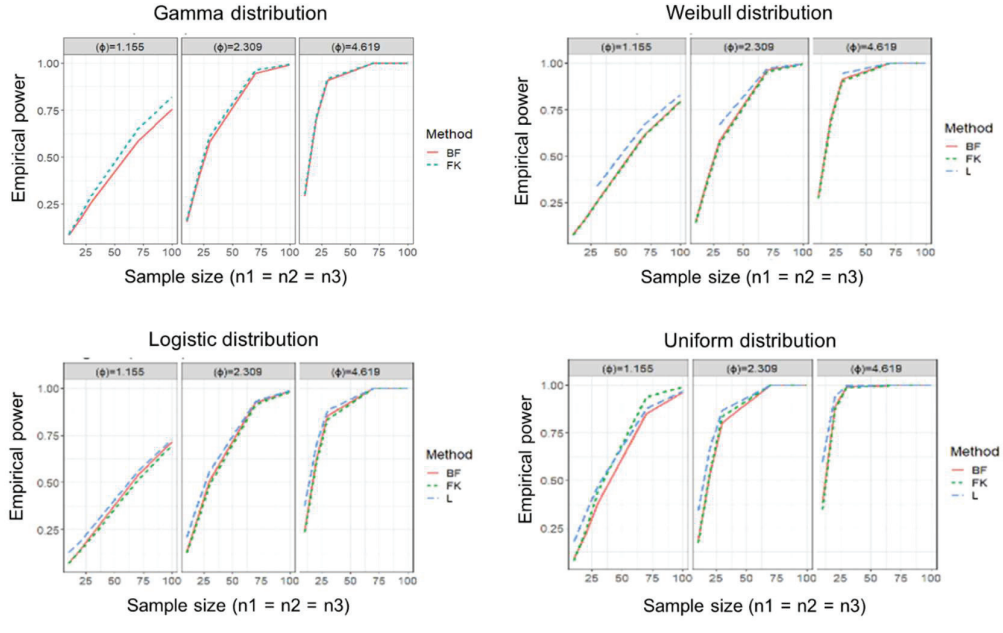

Figure 6 Empirical power of three test statistics when data follow Gamma, Weibull, Logistic, and Uniform distribution.

นอกจากนี้ จาก Figure 6 แสดงให้เห็นว่า กำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธีมีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างทั้ง 3 กลุ่มเพิ่มขึ้นและเมื่ออัตราส่วนความแปรปรวนของประชากรทั้ง 3 กลุ่มเพิ่มขึ้น (ค่า ϕ เพิ่มขึ้น)

สรุปและวิจารณ์ผลการวิจัย

จากผลการเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวน 3 วิธี คือ สถิติทดสอบของ Levene สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-killeen โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ความแกร่ง และกำลังการทดสอบ สามารถสรุปได้ดังนี้

1. ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และความแกร่ง

จากการศึกษาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley (ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วง [0.025, 0.075] ที่ระดับนัยสำคัญ 0.05) และความแกร่งตามเกณฑ์ของ Cochran (ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วง [0.04, 0.06] ที่ระดับนัยสำคัญ 0.05) ผลการศึกษาพบว่า สถิติทดสอบของ Levene มีประสิทธิภาพต่ำที่สุดในกรณีที่ประชากรมีการแจกแจงแกมมาที่มีลักษณะเบ้ขวาและโด่งต่ำ (α, β = (2, 2)) เนื่องจากไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์ของ Bradley และให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สูงกว่าระดับนัยสำคัญของการทดสอบทุกกรณี อย่างไรก็ตาม สถิติทดสอบของ Levene สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ดีขึ้นเมื่อประชากรมีการแจกแจงไวบูล (มีความเบ้น้อยกว่าการแจกแจงแกมมา) และตัวอย่างมีขนาดใหญ่ และสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ดีและมีความแกร่งเมื่อประชากรมีลักษณะการแจกแจงที่สมมาตรมากขึ้น (การแจกแจงลอจิสติกและการแจกแจงเอกรูป) ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของดวงพร หัชชะวณิช (2557) ที่พบว่า สถิติทดสอบของ Levene ไม่ใช่สถิติทดสอบที่ดีที่สุด ยังมีสถิติทดสอบอื่นๆ ที่ให้ผลการทดสอบดีกว่า ทั้งนี้ขึ้นอยู่กับการแจกแจงของข้อมูล

สำหรับสถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen จากผลการศึกษาพบว่า สถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen สามารถควบคุมความน่าเป็นของความผิดพลาดแบบที่ 1 ได้สำหรับทุกการแจกแจงที่ศึกษาและในเกือบทุกสถานการณ์ ยกเว้นกรณีตัวอย่างขนาดเล็ก

2. กำลังการทดสอบ

จากการศึกษากำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธี โดยพิจารณาเฉพาสถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้ ผลการศึกษาพบว่า กรณีประชากรมีการแจกแจงลอจิสติกและการแจกแจงเอกรูปสถิติทดสอบของ Levene มีประสิทธิภาพดีกว่าสถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen ในเกือบทุกสถานการณ์ เนื่องจากให้ค่าประมาณกำลังการทดสอบที่สูงกว่า ส่วนกรณีประชากรมีการแจกแจงไวบูล สถิติทดสอบของ Levene จะมีประสิทธิภาพดีกว่าสถิติทดสอบของ Brown-Forsythe และสถิติทดสอบของ Fligner-Killeen เมื่อตัวอย่างแต่ละกลุ่มมีขนาดตั้งแต่ 30 ขึ้นไป ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของ Conover *et al*. (1981) Lim & Loh (1996) และ Wang *et al*. (2017) ที่พบว่า ถึงแม้ว่าสถิติทดสอบของ Brown-Forsythe จะสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกลักษณะการแจกแจงของข้อมูลแต่สถิติทดสอบวิธีนี้มีแนวโน้มให้ค่ากำลังการทดสอบต่ำ สำหรับกรณีประชากรมีการแจกแจงแกมมา สถิติทดสอบของ Fligner-Killeen มีประสิทธิภาพดีกว่าสถิติทดสอบตัวอื่นๆ ในทุกสถานการณ์ ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของ Conover *et al*. (1981) ที่พบว่า สถิติทดสอบของ Fligner-Killeen มีความแกร่งและกำลังการทดสอบสูงเมื่อข้อมูลมีการแจกแจงแกมมาที่มีพารามิเตอร์ β = 2

3. ความสัมพันธ์ระหว่างขนาดตัวอย่าง ค่าพารามิเตอร์ไม่มีศูนย์กลาง และประสิทธิภาพของสถิติทดสอบ

สำหรับสถิติทดสอบของ Levene ผลการศึกษาพบว่า สถิติทดสอบของ Levene มักจะให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สูงกว่าระดับนัยสำคัญของการทดสอบเสมอ แต่เมื่อขนาดตัวอย่างเพิ่มขึ้น สถิติทดสอบของ Levene มีแนวโน้มที่จะให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ลดลงและเพิ่มความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 นอกจากนี้ยังพบว่า การเพิ่มขนาดตัวอย่าง ส่งผลให้กำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธีสูงขึ้น และเมื่อค่าพารามิเตอร์ไม่มีศูนย์กลาง (ϕ) เพิ่มขึ้นหรืออัตราส่วนความแปรปรวนของประชากรทั้ง 3 กลุ่มเพิ่มขึ้น จะทำให้กำลังการทดสอบเพิ่มขึ้นตามไปด้วย ดังนั้น การเพิ่มขนาดตัวอย่างและค่าพารามิเตอร์ไม่มีศูนย์กลางจะส่งผลให้ประสิทธิภาพของสถิติทดสอบทั้ง 3 วิธีสูงขึ้น

ข้อเสนอแนะ

ในงานวิจัยครั้งต่อไป สามารถขยายขอบเขตการศึกษาประสิทธิภาพของสถิติทดสอบสำหรับทดสอบภาวะความเท่ากันของความแปรปรวนของประชากรให้กว้างขึ้น โดยอาจศึกษาการแจกแจงที่นำเสนอในบทความนี้ที่มีค่าพารามิเตอร์อื่นๆ แตกต่างจากที่ผู้วิจัยได้ศึกษาไว้หรือใช้การแจกแจงอื่นๆ ที่มีลักษณะความเบ้ และความโด่งแตกต่างกัน เช่น การแจกแจงพาเรโต (Pareto distribution) การแจกแจงเลขชี้กำลังคู่ (double exponential distribution) การแจกแจงบีตา (beta distribution) เป็นต้น และควรเพิ่มสถิติทดสอบวิธีอื่นๆ ที่นอกเหนือจากที่ได้ศึกษาไปแล้ว

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนทุนวิจัยจากคณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ปีการศึกษา 2565

เอกสารอ้างอิง

ดวงพร หัชชะวณิช. (2557). การเปรียบเทียบความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบของสถิติทดสอบความเป็นเอกพันธ์ของความแปรปรวน. *Journal of Science Ladkrabang*, *23*(1), 17-28.

วรางคณา เรียนสุทธิ์. (2561). การเปรียบเทียบประสิทธิภาพของสถิติทดสอบอิงพารามิเตอร์ สำหรับการทดสอบภาวะความเท่ากันของความแปรปรวน. *Naresuan University Journal: Science and Technology (NUJST), 26*(3), 170-180.

วรางคณา เรียนสุทธิ์. (2562). การเปรียบเทียบประสิทธิภาพของสถิติทดสอบไม่อิงพารามิเตอร์สำหรับการทดสอบภาวะความเท่ากันของความแปรปรวน. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี*, *21*(2), 163-170.

สายชล สินสมบูรณ์ทอง. (2561). การเปรียบเทียบประสิทธิภาพความเท่ากันของความแปรปรวนประชากร 3 กลุ่ม ภายใต้การแจกแจงที่มีความโด่งมากและความเบ้มาก. *Thai Science and Technology Journal*, 721-738.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association, 69*(346), 364-367. https://doi.org/10.2307/2285659

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Cochran, W. G. (1952). The C^2 test of goodness of fit. *The Annals of Mathematical Statistics, 23*(3), 315-345. https://doi.org/10.1214/aoms/1177729380

Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, *23*(4), 351-361. https://doi.org/10.1080/00401706.1981.10487680

Levene, H. (1960). Robust tests for equality of variances. *Contributions to Probability and Statistics*, 278-292.

Lim, T. S., & Loh, W. Y. (1996). A comparison of tests of equality of variances. *Computational Statistics and Data Analysis, 22*, 287-301. doi:10.1016/0167-9473(95)000 54-2

Niu, X. (2004). *Statistical procedures for testing homogeneity of water quality parameters.* Department of Statistics Florida State University Tallahassee.

Wang, Y., de Gil, P.R., Chen, Y. Kromrey, J. D., Kim, E. S., Pham, T., Nguyen, D., & Romano, J. L. (2017). Comparing the performance of approaches for testing the homogeneity of variance assumption in one-factor ANOVA models. *Educational and Psychological Measurement*, *77*(2), 305-329. https://doi.org/10.1177/0013164416645162