การเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับค่าเฉลี่ยประชากรมากกว่า 2 กลุ่ม ภายใต้การแจกแจงแบบสมมาตร

Comparing the Performance of Test Statistics for More Than Two Population Means under Symmertric Distribution

บำรุงศักดิ์ เผื่อนอารีย์* และ โชคศรัญญ์ สิริลีพันธุ์

Bumrungsak Phuenaree and Choksarun Sirileepun

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

Received : 26 December 2017

Accepted : 15 March 2018

Published online : 16 March 2018

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับค่ากลางของประชากรมากกว่า 2 กลุ่ม 3 วิธีคือ สถิติทดสอบเอฟ สถิติทดสอบครัสคอล-วอลลิส และสถิติทดสอบการเรียงสับเปลี่ยน ภายใต้การแจกแจงสมมาตร ได้แก่ การแจกแจงลอจิสติก การแจกแจงลาปลาซ และการแจกแจงเอกรูป กำหนดขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 5, 10, 15 และ 20 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และ 2 ค่าเฉลี่ยมีค่าเป็น 0 และค่าความแตกต่างของค่าเฉลี่ยเท่ากับ 0.0, 0.5 และ 1.5 กำหนดระดับนัยสำคัญเท่ากับ 0.05 โดยมีเกณฑ์พิจารณา คือ ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลทำซ้ำจำนวน 10,000 ครั้ง และทดลองซ้ำโดยวิธีการเรียงสับเปลี่ยนจำนวน 2,000 ครั้ง พบว่าสถิติทดสอบทั้งสามวิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี เมื่อพิจารณากำลังการทดสอบ พบว่าสถิติทดสอบการเรียงสับเปลี่ยนจะมีกำลังการทดสอบมากที่สุด ภายใต้การแจกแจงลอจิสติก สำหรับข้อมูลที่มีการแจกแจงลาปลาซ สถิติทดสอบครัสคอล-วอลลิสจะมีกำลังการทดสอบมากที่สุด ยกเว้นกรณีตัวอย่างขนาดเท่ากับ 5 สถิติทดสอบการเรียงสับเปลี่ยนจะมีกำลังการทดสอบมากที่สุด สำหรับภายใต้การแจกแจงเอกรูป พบว่าสถิติทดสอบเอฟจะมีกำลังการทดสอบมากที่สุด

คำสำคัญ : สถิติทดสอบการเรียงสับเปลี่ยน สถิติทดสอบเอฟ สถิติทดสอบครัสคอล-วอลลิส กำลังการทดสอบ ความผิดพลาดแบบที่ 1

*Corresponding author. E-mail : bumrungsak@buu.ac.th

Abstract

The purpose of this study is to compare the performance of three test statistics which are F– test, Kruskal – Wallis test and permutation test. Three symmetric distributions of data, logistic, Laplace and uniform distribution are considered. These data sets have standard deviations 1 and 2. The sample sizes are 5, 10, 15 and 20. The mean is 0 and the mean differences are 0.0, 0.5 and 1.5. Significance level will be set at 0.05. The considered criterions are type I error rates and power of the tests. A Monte Carlo simulation is performed with repeated 10,000 times. The number of permutation is 2,000. The results show that three test statistics control the nominal level well. Considering the power of a test, the permutation test has the highest power when the data are chosen from logistic distribution. For Laplace distribution, Kruskal – Wallis test test has the highest power except for the case sample size 5 which the permutation test performs well. For uniform distribution, F-test statistic test has the highest power.

Keywords : permutation test, F-test, Kruskal – Wallis test, power of a test, type I error

บทนำ

การวิเคราะห์ข้อมูลด้วยวิธีการทางสถิตินั้น สิ่งหนึ่งที่ผู้วิจัยต้องคำนึงถึงอยู่เสมอคือข้อสมมุติเบื้องต้นของวิธีการทางสถิติที่เลือกใช้ ตัวอย่างเช่น กรณีที่ต้องการทดสอบสมมุติฐานเพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยสำหรับประชากรมากกว่า 2 กลุ่มขึ้นไป เทคนิคการทดสอบสมมุติฐานทางสถิติที่นิยมใช้กัน คือ การวิเคราะห์ความแปรปรวน (Analysis of Variance) โดยใช้สถิติทดสอบเอฟ (F - Test) ซึ่งมีข้อสมมุติเบื้องต้น คือ 1) ข้อมูลแต่ละกลุ่มถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติ (Normal Distribution) 2) ข้อมูลแต่ละกลุ่มสุ่มมาจากประชากรที่มีความแปรปรวนเท่ากันหรือคงที่ (Homogeneity of Variance) และ 3) ข้อมูลแต่ละกลุ่มถูกสุ่มมาจากประชากรที่เป็นอิสระกัน แต่อย่างไรก็ตามในความเป็นจริงแล้วข้อมูลที่มีอยู่อาจไม่เป็นไปตามข้อสมมุติเบื้องต้นทุกประการ โดยเฉพาะเรื่องการแจกแจงของข้อมูล ซึ่งในทางปฏิบัติข้อมูลที่นำมาวิเคราะห์อาจไม่มีการแจกแจงปรกติ ดังนั้นการใช้สถิติทดสอบเอฟในการทดสอบสมมุติฐานอาจนำไปสู่ผลสรุปที่ผิดพลาดได้

นักวิจัยหลายท่านได้ศึกษาหาสถิติทดสอบที่สามารถนำมาใช้ในกรณีดังกล่าวนี้โดยไม่ต้องคำนึงถึงข้อสมมุติเบื้องต้นบางประการ Kruskal and Wallis (1952) ได้เสนอตัวสถิติทดสอบที่ใช้ได้ดีกับตัวอย่างขนาดเล็ก คือ สถิติทดสอบครัสคอล-วอลลิส (Kruskal - Wallis Test) โดยสามารถวิเคราะห์ข้อมูลที่ได้จากการวางแผนการทดลองแบบสุ่มสมบูรณ์ แต่สถิติทดสอบนี้มีข้อจำกัดที่ว่าหากข้อมูลมีลักษณะตรงตามข้อสมมุติเบื้องต้นของสถิติทดสอบเอฟ จะทำให้ประสิทธิภาพของสถิติทดสอบครัสคอล-วอลลิส ด้อยกว่าสถิติทดสอบเอฟ นอกจากนั้นยังพบว่าเมื่อข้อมูลมีการแจกแจงเอกรูป สถิติทดสอบครัสคอล-วอลลิสจะมีค่ากำลังการทดสอบน้อยกว่าสถิติทดสอบเอฟด้วยเช่นกัน (Gleason, 2013)

อีกวิธีหนึ่งที่สามารถนำมาประยุกต์ใช้ในกรณีที่ไม่ทราบการแจกแจงของข้อมูล คือวิธีการเรียงสับเปลี่ยน (Permutation) โดย Fisher (1971) เป็นคนแรกที่นำทฤษฎีดังกล่าวมาใช้ในการอนุมานทางสถิติ Ernst (2004) กล่าวถึงการใช้

สถิติทดสอบการเรียงสับเปลี่ยน (Permutation Test) เพื่อทดสอบสมมุติฐานกรณีประชากรมากกว่า 2 กลุ่มโดยนำข้อมูลชุดตั้งต้นไปวิเคราะห์ทางสถิติ จากนั้นทำการสุ่มเรียงสับเปลี่ยนชุดข้อมูลใหม่ โดยที่ทำการเปรียบเทียบค่าสถิติทดสอบจากวิธีเรียงสับเปลี่ยนกับสถิติทดสอบจากข้อมูลชุดตั้งต้นตามจำนวนวิธีเรียงสับเปลี่ยนทั้งหมดที่เป็นไปได้ นอกจากนั้นทฤษฎีดังกล่าวยังสามารถปรับเลือกใช้สถิติทดสอบอื่น ๆ ได้อีกด้วย

จากที่กล่าวมาข้างต้นผู้วิจัยจึงสนใจที่จะศึกษาการเปรียบเทียบประสิทธิภาพของสถิติทดสอบเอฟ สถิติทดสอบครัสคอล-วอลลิส และสถิติทดสอบการเรียงสับเปลี่ยน ภายใต้การแจกแจงสมมาตร ได้แก่ การแจกแจงลอจิสติก (Logistic Distribution) การแจกแจงลาปลาซ (Laplace Distribution) และ การแจกแจงเอกรูป (Uniform Distribution) ในกรณีที่ข้อมูลทุกกลุ่มมีความแปรปรวนเท่ากันและตัวอย่างมีขนาดเล็ก

วิธีดำเนินการวิจัย

งานวิจัยนี้ต้องการเปรียบเทียบประสิทธิภาพของสถิติทดสอบค่ากลางของประชากรมากกว่า 2 กลุ่ม ทำการศึกษาโดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ทดลองจำนวน 10,000 ครั้ง และทดลองซ้ำโดยวิธีการเรียงสับเปลี่ยนจำนวน 2,000 ครั้ง กำหนดขนาดตัวอย่างเท่ากับ 5, 10, 15 และ 20 ค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 และ 2 กำหนดจำนวนประชากรเท่ากับ 3 และ 4 กลุ่ม *(k=3, 4)* โดยมีเกณฑ์การพิจารณาคือค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ซึ่งจะกำหนดค่าเฉลี่ยของประชากรทุกกลุ่มเท่ากับ 0 ในทุกกรณีที่ศึกษา และพิจารณาค่ากำลังการทดสอบ ซึ่งจะกำหนดค่าเฉลี่ยในแต่ละกลุ่มตากรณีที่ศึกษา คือ กรณีประชากร 3 กลุ่ม กำหนดค่าเฉลี่ย (μ, μ, μ) เป็น (0, 0, 0.5), (0, 0, 1.5) และกรณีประชากร 4 กลุ่ม กำหนดค่าเฉลี่ย (μ, μ, μ, μ) เป็น (0, 0, 0, 0.5), (0, 0, 0, 1.5), (0, 0, 0.5, 0.5), (0, 0, 1.5, 1.5) กำหนดระดับนัยสำคัญ 0.05 ศึกษาข้อมูลภายใต้การแจกแจงลอจิสติก การแจกแจงลาปลาซ และการแจกแจงเอกรูป โดยเปรียบเทียบประสิทธิภาพของสถิติทดสอบทั้งหมด 3 วิธีได้แก่ สถิติทดสอบเอฟ สถิติทดสอบครัสคอล-วอลลิส และสถิติทดสอบการเรียงสับเปลี่ยน ซึ่งมีรายละเอียดดังนี้

1. สถิติทดสอบเอฟ เป็นสถิติทดสอบที่ใช้ทดสอบค่าเฉลี่ยตั้งแต่สองกลุ่มขึ้นไป โดยมีข้อสมมติเบื้องต้นคือ ข้อมูลต้องถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติ และมีความแปรปรวนคงที่ สามารถคำนวณได้จากสูตร (Casella & Berger, 2001)

$$F = \frac{\sum_{i=1}^{k} n_i \left(\bar{y}_{i.} - \bar{y}\right)^2 \Big/ k-1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_{i.}\right)^2 \Big/ n-k} \qquad (1)$$

โดยที่ n คือ จำนวนค่าสังเกตทั้งหมด, k คือ จำนวนกลุ่ม, n_i คือ ขนาดตัวอย่างที่ i, y_{ij} คือ ค่าสังเกตที่ j ของตัวอย่างที่ i, $\bar{y}_{i.}$ คือ ค่าเฉลี่ยของตัวอย่างที่ i, $\bar{y}$ คือ ค่าเฉลี่ยรวม

2. สถิติทดสอบครัสคอล-วอลลิส ใช้ทดสอบความแตกต่างของค่ามัธยฐานประชากร k กลุ่ม โดยมีข้อสมมติเบื้องต้น คือ ข้อมูลที่จะใช้วิเคราะห์ต้องมีมาตราวัดอันดับ (Ordinal Scale) เป็นอย่างน้อย และมีการแจกแจงต่อเนื่อง โดยสามารถคำนวณได้จากสูตร

$$H = \left[\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right] - 3(n+1) \qquad (2)$$

โดยที่ n คือ จำนวนค่าสังเกตทั้งหมด, k คือ จำนวนกลุ่ม, n_i คือ ขนาดตัวอย่างที่ i, R_i คือ ผลรวมของลำดับที่ในตัวอย่างที่ i

3. สถิติทดสอบการเรียงสับเปลี่ยน โดย Fisher (1971) เป็นคนแรกที่นำทฤษฎีของวิธีการเรียงสับเปลี่ยนมาใช้ในการอนุมานทางสถิติ ซึ่งมีหลักการคือ สมมติตัวอย่างที่สุ่มมาจากประชากร k กลุ่ม แต่ละกลุ่มมีขนาด n_i เมื่อ $i = 1, 2, ..., k$ โดยที่ $n_1 + n_2 + \ldots + n_k = n$ ทำการคำนวณค่าสถิติทดสอบของข้อมูลตั้งต้นดังสมการ (3) (Ernst, 2004; Butar, 2007)

$$T_{obs} = \sum_{i=1}^{k} n_i \bar{y}_{i.}^2 \qquad (3)$$

เมื่อ k คือ จำนวนกลุ่ม, n_i คือ ขนาดตัวอย่างที่ i, $\bar{y}_{i.}$ คือ ค่าเฉลี่ยของตัวอย่างที่ i จากนั้นเริ่มดำเนินการโดยการนำตัวอย่างทั้งหมดมาทำการสุ่มแบบไม่ใส่คืนจากตัวอย่างทั้งหมดขนาด n โดยสุ่มครั้งแรกขนาด n_1 ให้กับกลุ่มที่ 1 เหลือตัวอย่างขนาด $n - n_1$ จากนั้นสุ่มครั้งที่ 2 ขนาด n_2 ให้กับกลุ่มที่ 2 เหลือตัวอย่างขนาด $n - n_1 - n_2$ สุ่มไปเรื่อย ๆ จนถึงการสุ่มครั้งที่ *k-1* ขนาด n_{k-1} ให้กับกลุ่มที่ *k-1* เหลือตัวอย่างขนาด $n - n_1 - n_2 - \ldots - n_{k-1}$ กำหนดให้กลุ่มที่ k จะได้เป็นจำนวนวิธีการเรียงสับเปลี่ยนทั้งหมดเท่ากับ $\binom{n}{n_1 n_2 \ldots n_k}$ ซึ่งในแต่ละรอบของชุดตัวอย่างสุ่มที่ได้มานั้น นำมาหาค่าสถิติทดสอบดังสมการ (3) และคำนวณค่าพี (p- value) โดยใช้สัดส่วนของค่าสถิติทดสอบในแต่ละครั้งที่มากกว่าค่าสถิติทดสอบของข้อมูลตั้งต้น ซึ่งหากข้อมูลมีจำนวนเพิ่มมากขึ้นชุดข้อมูลที่ได้จากวิธีการเรียงสับเปลี่ยนอาจมีจำนวนมากเกินไป ดังนั้นในงานวิจัยนี้จึงกำหนดให้มีจำนวนการทำซ้ำของการเรียงสับเปลี่ยนทั้งหมด 2,000 ครั้ง

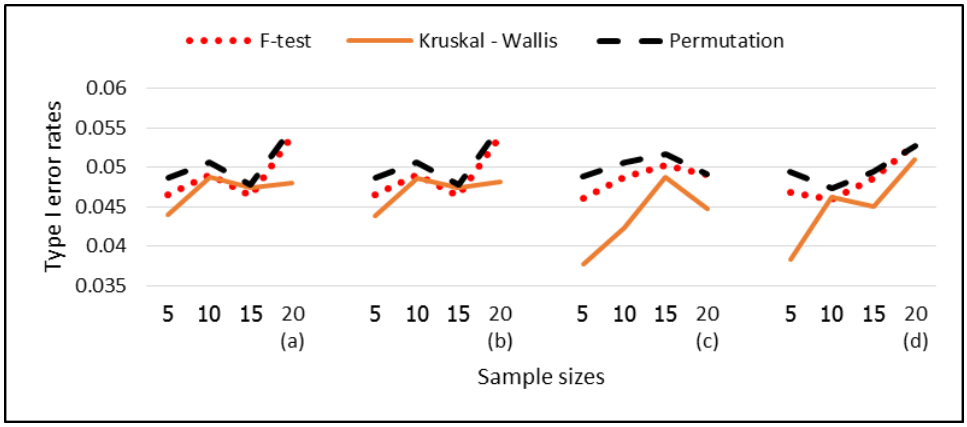
4. เกณฑ์ที่ใช้ในงานวิจัย แบ่งออกเป็น 2 เกณฑ์ ได้แก่

4.1 พิจารณาค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ($\hat{\alpha}$) คำนวณโดยการจำลองข้อมูลตามสมมุติฐานหลัก (ค่าเฉลี่ยประชากรทุกกลุ่มเท่ากับ 0) จากนั้นนับสัดส่วนของการปฏิเสธสมมุติฐานหลัก และนำมาเปรียบเทียบกับเกณฑ์ของ Bradley โดย $\hat{\alpha} \in$ (0.025,0.075) และเกณฑ์ของ Cochran โดย $\hat{\alpha} \in$ (0.04,0.06) หากสถิติทดสอบใดมีค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 อยู่ในช่วงดังกล่าวจะถือว่ามีความสามารถในการควบคุมการเกิดความผิดพลาดแบบที่ 1

4.2 พิจารณาเปรียบเทียบค่ากำลังการทดสอบของสถิติทดสอบที่ผ่านเกณฑ์ในข้อ 4.1 คำนวณโดยการจำลองข้อมูลที่ไม่เป็นไปตามสมมุติฐานหลัก (ค่าเฉลี่ยอย่างน้อยหนึ่งกลุ่มแตกต่างจากกลุ่มอื่น) จากนั้นนับสัดส่วนของการปฏิเสธสมมุติฐานหลัก และพิจารณาจากค่ากำลังการทดสอบที่มากที่สุด

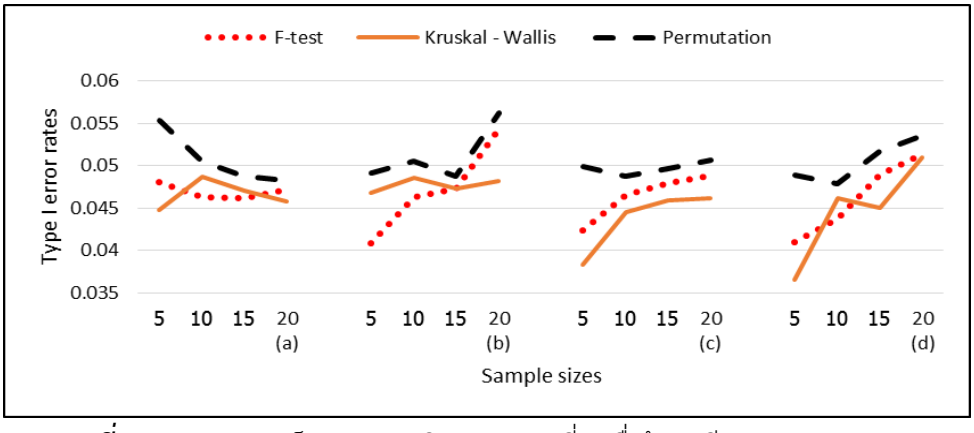
ผลการวิจัยและวิจารณ์ผล

การนำเสนอผลการวิจัยจะแบ่งออกเป็น 2 ส่วนได้แก่ ส่วนที่ 1 แสดงกราฟของความน่าจะเป็นของความผิดพลาดแบบที่ 1 และ ส่วนที่ 2 แสดงตารางของค่ากำลังการทดสอบ โดยกำหนดสัญลักษณ์ คือ F แทน สถิติทดสอบเอฟ K แทน สถิติทดสอบครัสคอล-วอลลิส และ P แทน สถิติทดสอบการเรียงสับเปลี่ยน



ภาพที่ 1 ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงลอจิสติก

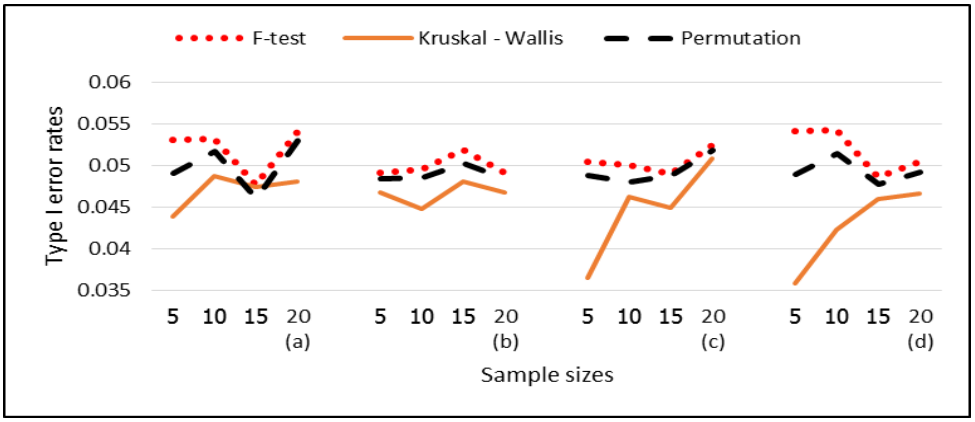
(a) $k = 3, \sigma = 1$,(b) $k = 3, \sigma = 2$, (c) $k = 4, \sigma = 1$ และ (d) $k = 4, \sigma = 2$



ภาพที่ 2 ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงลาปลาซ

(a) $k = 3, \sigma = 1$,(b) $k = 3, \sigma = 2$, (c) $k = 4, \sigma = 1$ และ (d) $k = 4, \sigma = 2$

พิจารณาภาพที่ 1-3 ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงลอจิสติก การแจกแจงลาปลาซ และการแจกแจงเอกรูป พบว่าสถิติทดสอบทั้ง 3 วิธี สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ยกเว้นกรณีประชากร 4 กลุ่ม เฉพาะที่ขนาดตัวอย่างเท่ากับ 5 สถิติทดสอบครัสคอล-วอลลิส ไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Cochran ได้ แต่ยังสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์ของ Bradley



ภาพที่ 3 ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงเอกรูป

(a) $k = 3, \sigma = 1$,(b) $k = 3, \sigma = 2$, (c) $k = 4, \sigma = 1$ และ (d) $k = 4, \sigma = 2$

พิจารณาค่ากำลังการทดสอบจากตารางที่ 1-3 เมื่อข้อมูลมีการแจกแจงลอจิสติก พบว่าสถิติทดสอบการเรียงสับเปลี่ยนมีค่ากำลังการทดสอบสูงที่สุดในทุกกรณีที่ศึกษา เมื่อข้อมูลมีการแจกแจงลาปลาซ พบว่าสถิติทดสอบครัสคอล-วอลลิสมีกำลังการทดสอบสูงที่สุด ยกเว้นในกรณีที่ขนาดตัวอย่างเท่ากับ 5 สถิติทดสอบการเรียงสับเปลี่ยนมีกำลังการทดสอบสูงที่สุด สำหรับข้อมูลที่มีการแจกแจงเอกรูปพบว่าสถิติทดสอบเอฟมีกำลังการทดสอบมากที่สุดในทุกกรณี

ตารางที่ 1 ค่ากำลังการทดสอบ สำหรับการแจกแจงลอจิสติก

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 0.5)	5	0.1072	0.0905	**0.1111***	0.0596	0.0551	**0.0636***
	10	0.1858	0.1537	**0.1906***	0.0803	0.0678	**0.0829***
	15	0.2685	0.2201	**0.2703***	0.0934	0.0793	**0.0954***
	20	0.3431	0.2839	**0.3443***	0.1157	0.0871	**0.1166***
(0, 0, 1.5)	5	0.5848	0.5307	**0.5905***	0.1863	0.1560	**0.1910***
	10	0.9115	0.8836	**0.9119***	0.3666	0.3082	**0.3709***
	15	0.9842	0.9800	**0.9848***	0.5312	0.4700	**0.5341***
	20	0.9979	0.9976	**0.9981***	0.6693	0.5967	**0.6712***
(0, 0, 0, 0.5)	5	0.0967	0.0753	**0.0984***	0.0562	0.0458	**0.0610***
	10	0.1732	0.1398	**0.1754***	0.0724	0.0636	**0.0751***
	15	0.2504	0.2004	**0.2528***	0.0922	0.0775	**0.0938***
	20	0.3255	0.2644	**0.3277***	0.1060	0.0849	**0.1075***

ตารางที่ 1(ต่อ) ค่ากำลังการทดสอบ สำหรับการแจกแจงลอจิสติก

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 0, 1.5)	5	0.5731	0.4869	**0.5790***	0.1702	0.1360	**0.1760***
	10	0.9142	0.8788	**0.9147***	0.3602	0.2998	**0.3624***
	15	0.9891	0.9819	**0.9894***	0.5214	0.4549	**0.5236***
	20	**0.9989***	0.9974	**0.9989***	0.6582	0.5878	**0.6585***
(0, 0, 0.5, 0.5)	5	0.1167	0.0867	**0.1206***	0.0640	0.0473	**0.0663***
	10	0.2231	0.1752	**0.2261***	0.0847	0.0699	**0.0863***
	15	0.3232	0.2624	**0.3266***	0.1089	0.0869	**0.1105***
	20	0.4254	0.3494	**0.4267***	0.1309	0.1049	**0.1315***
(0, 0, 1.5, 1.5)	5	0.7138	0.6405	**0.7143***	0.2114	0.1611	**0.2174***
	10	0.9721	0.9664	**0.9728***	0.4507	0.3877	**0.4556***
	15	0.9984*	0.9982	**0.9984***	0.6516	0.5860	**0.6529***
	20	**1.0000***	0.9999	**1.0000***	0.7960	0.7410	**0.7973***

* สถิติทดสอบที่มีกำลังการทดสอบสูงที่สุด

ตารางที่ 2 ค่ากำลังการทดสอบ สำหรับการแจกแจงลาปลาซ

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 0.5)	5	0.1110	0.1206	**0.2002***	0.0572	0.0586	**0.0685***
	10	0.1963	**0.2284***	0.204	0.0805	**0.0883***	0.0868
	15	0.2651	**0.3438***	0.2702	0.0986	**0.1152***	0.1028
	20	0.3570	**0.4505***	0.3595	0.1180	**0.1440***	0.1215
(0, 0, 1.5)	5	0.6148	0.5970	**0.6248***	0.1974	0.2084	**0.2115***
	10	0.9052	**0.9271***	0.9069	0.3806	**0.4396***	0.3923
	15	0.9826	**0.9903***	0.9831	0.5431	**0.6413***	0.5491
	20	0.9973	**0.9997***	0.9978	0.6677	**0.7821***	0.6714

ตารางที่ 2 (ต่อ) ค่ากำลังการทดสอบ สำหรับการแจกแจงลาปลาซ

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 0, 0.5)	5	0.0967	0.0975	**0.1086***	0.0549	0.0532	**0.0625***
	10	0.1766	**0.2094***	0.1858	0.0730	**0.0828***	0.0786
	15	0.2587	**0.3227***	0.2661	0.0937	**0.1068***	0.0971
	20	0.3321	**0.4331***	0.3388	0.1078	**0.1317***	0.1117
(0, 0, 0, 1.5)	5	0.6001	0.5505	**0.6104***	0.1793	0.1829	**0.1939***
	10	0.9103	**0.9239***	0.9121	0.3645	**0.4202***	0.3753
	15	0.9847	**0.9924***	0.9857	0.5290	**0.6250***	0.5340
	20	0.9979	**0.9997***	0.9978	0.6700	**0.7792***	0.6743
(0, 0, 0.5, 0.5)	5	0.1224	0.1265	**0.1334***	0.0634	0.0583	**0.0691***
	10	0.2292	**0.2809***	0.2375	0.0838	**0.0969***	0.0896
	15	0.3330	**0.4254***	0.3390	0.1112	**0.1321***	0.1158
	20	0.4331	**0.5597***	0.4383	0.1337	**0.1669***	0.1368
(0, 0, 1.5, 1.5)	5	0.7146	0.7086	**0.7213***	0.2241	0.2310	**0.2426***
	10	0.9627	**0.9822***	0.9629	0.4638	**0.5438***	0.4708
	15	0.9971	**0.9998***	0.9972	0.6547	**0.7732***	0.6601
	20	0.9999	**1.0000***	0.9998	0.7960	**0.8958***	0.8005

* สถิติทดสอบที่มีกำลังการทดสอบสูงที่สุด

ตารางที่ 3 ค่ากำลังการทดสอบ สำหรับการแจกแจงเอกรูป

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 0.5)	5	**0.0977***	0.0656	0.0918	**0.0650***	0.0494	0.0599
	10	**0.1747***	0.0885	0.1710	**0.0828***	0.0560	0.0807
	15	**0.2446***	0.1155	0.2406	**0.0898***	0.0590	0.0876
	20	**0.3299***	0.1359	0.3284	**0.1143***	0.0663	0.1130

ตารางที่ 3 (ต่อ) ค่ากำลังการทดสอบ สำหรับการแจกแจงเอกรูป

ค่าเฉลี่ย	n	$\sigma = 1$			$\sigma = 2$		
		F	K	P	F	K	P
(0, 0, 1.5)	5	**0.5456***	0.3882	0.5323	**0.1605***	0.0926	0.1521
	10	**0.9226***	0.7229	0.9207	**0.3447***	0.1557	0.3383
	15	**0.9930***	0.8929	0.9927	**0.5156***	0.2341	0.5123
	20	**0.9991***	0.9662	**0.9991***	**0.6534***	0.2928	0.6499
(0, 0, 0, 0.5)	5	**0.0973***	0.0534	0.0909	**0.0680***	0.0454	0.0637
	10	**0.1678***	0.0837	0.1639	**0.0765***	0.0520	0.0748
	15	**0.2387***	0.1064	0.2372	**0.0876***	0.0575	0.0863
	20	**0.3189***	0.1263	0.3152	**0.1058***	0.0607	0.1050
(0, 0, 0, 1.5)	5	**0.5355***	0.3471	0.5255	**0.1572***	0.0864	0.1495
	10	**0.9273***	0.7121	0.9251	**0.3246***	0.1498	0.3196
	15	**0.9931***	0.8860	0.9929	**0.5054***	0.2078	0.5015
	20	**0.9995***	0.9650	**0.9995***	**0.6484***	0.2808	0.6446
(0, 0, 0.5, 0.5)	5	**0.1134***	0.0567	0.1083	**0.0695***	0.0393	0.0639
	10	**0.2058***	0.0953	0.2014	**0.0852***	0.0538	0.0829
	15	**0.3109***	0.1229	0.3081	**0.1071***	0.0622	0.1041
	20	**0.4179***	0.1561	0.4129	**0.1267***	0.0680	0.1244
(0, 0, 1.5, 1.5)	5	**0.6844***	0.4519	0.6739	**0.1959***	0.0913	0.1867
	10	**0.9820***	0.8528	0.9815	**0.4388***	0.1875	0.4318
	15	**0.9993***	0.9761	**0.9993***	**0.6446***	0.2804	0.6399
	20	**0.9999***	0.9965	**0.9999***	**0.7903***	0.3641	0.7887

* สถิติทดสอบที่มีกำลังการทดสอบสูงที่สุด

สรุปผลการวิจัย

จากผลการศึกษาสรุปได้ว่าสถิติทดสอบทั้ง 3 วิธี สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณีที่ศึกษา เมื่อพิจารณาถึงกำลังการทดสอบของข้อมูลที่มีการแจกแจงสมมาตรทั้ง 3 แบบ พบว่าเมื่อขนาดตัวอย่างและค่าส่วนเบี่ยงเบนมาตรฐานเพิ่มมากขึ้น ค่ากำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธีมีค่าเพิ่มมากขึ้น และเมื่อขนาดความแตกต่างของค่าเฉลี่ยในแต่ละกลุ่มเพิ่มขึ้นค่ากำลังการทดสอบของสถิติทดสอบทั้ง 3 วิธีมีค่าเพิ่มมากขึ้นเช่นเดียวกัน

สถิติทดสอบการเรียงสับเปลี่ยนจะมีประสิทธิภาพดีที่สุดเมื่อข้อมูลมีการแจกแจงลอจิสติก ส่วนสถิติทดสอบเอฟจะมีประสิทธิภาพดีที่สุดเมื่อข้อมูลมีการแจกแจงเอกรูป และสำหรับสถิติทดสอบครัสคอล-วอลลิส ส่วนใหญ่จะมีประสิทธิภาพดีที่สุดเมื่อข้อมูลมีการแจกแจงลาปลาซ ยกเว้นขนาดตัวอย่างเท่ากับ 5 สถิติทดสอบการเรียงสับเปลี่ยนจะเป็นวิธีดีที่สุด นอกจากนี้ยังพบว่าถึงแม้การแจกแจงข้อมูลจะมีลักษณะสมมาตรเหมือนกัน รวมไปถึงกำหนดค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานเท่ากัน แต่ผลลัพธ์ที่ได้แตกต่างกัน ซึ่งเป็นเพราะลักษณะของการแจกแจงนั้นต่างกันคือ การกระจายของข้อมูลรอบค่าเฉลี่ย โดยสำหรับการแจกแจงลอจิสติกและการแจกแจงลาปลาซ ข้อมูลส่วนใหญ่จะกระจายอยู่รอบค่าเฉลี่ยมากกว่าการแจกแจงแบบเอกรูป เป็นต้น ดังนั้นจึงส่งผลให้ประสิทธิภาพของสถิติทดสอบแต่ละวิธีแตกต่างกัน

อย่างไรก็ตามเมื่อพิจารณาค่ากำลังการทดสอบของสถิติทดสอบการเรียงสับเปลี่ยนในทุกกรณีที่ศึกษา พบว่ามีค่าสูงที่สุดหรือบางกรณีก็มีค่าใกล้เคียงกันกับวิธีที่ดีที่สุดเสมอ ยกตัวอย่างเช่น เมื่อข้อมูลมีการแจกแจงเอกรูป สถิติทดสอบการเรียงสับเปลี่ยนและสถิติทดสอบเอฟจะมีกำลังการทดสอบใกล้เคียงกัน สอดคล้องกับงานวิจัยของ Gleason (2013) ดังนั้นในทางปฏิบัติ หากข้อมูลไม่มีการแจกแจงปรกติแต่มีลักษณะสมมาตรแล้ว ควรเลือกใช้สถิติทดสอบการเรียงสับเปลี่ยนเนื่องจากมีประสิทธิภาพดีสุดหรือใกล้เคียงกับวิธีที่มีประสิทธิภาพดีที่สุดเสมอ

เอกสารอ้างอิง

Butar, B.F. (2007). Permutation tests for more than two samples. *Journal of Mathematical Sciences & Mathematics Education, 2*(2), 20-29.

Casella, G. & Berger, R.L. (2001). *Statistical Inference*. United States of America: Duxbury.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the Bootstrap*. New York: Chapman & Hall.

Ernst, M.D. (2004). Permutation methods: A basis for exact inference. *Statistical Science, 19*(4), 676-685.

Fisher, R.A. (1971). *The Design of Experiments*. New York: Hafner Publishing Company.

Gleason, J.H. (2013). *Comparative power of The Anova, Randomization Anova, and Kruskal-Wallis test*. Doctoral Dissertation, Evaluation and Research, Graduate School, Wayne State University.

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