การเปรียบเทียบประสิทธิภาพในการทดสอบของวิธีการเปรียบเทียบพหุคูณด้วยวิธีของตูกีย์และวิธีสเต็ปดาวน์บูทสแตรปมินพี

A Comparison of the Performance of Tukey's Multiple Comparison Test and Step-Down Bootstrap Min P Procedure

บำรุงศักดิ์ เผื่อนอารีย์[1] และมงคล ลีลาไพบูลย์[2*]

Bumrungsak Phuenaree[1] and Mongkol Leelaphaiboon[2*]

[1] *ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา*

[2] *สาขาคณิตศาสตร์และสถิติประยุกต์ คณะวิทยาศาสตร์และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน*

[1] *Department of Mathematics, Faculty of Science, Burapha University*

[2] *Department of Applied Mathematics and Statistics, Faculty of Sciences and Liberal Arts,*

Rajamangala University of Technology Isan

Received : 30 October 2018

Revised : 2 January 2019

Accepted : 16 January 2019

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพในการทดสอบของวิธีการเปรียบเทียบพหุคูณ ด้วยสถิติทดสอบ 3 วิธีคือ สถิติทดสอบตูกีย์ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี ภายใต้การแจกแจงปรกติ การแจกแจงลาปลาซ การแจกแจงลอจิสติก การแจกแจงล็อกนอร์มอล และการแจกแจงแกมมา ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จำนวน 5,000 ครั้ง และทดลองซ้ำด้วยวิธีบูทสแตรป 1,000 ครั้ง กำหนดจำนวนทรีทเมนต์ที่ทำการศึกษาเท่ากับ 3 และ 4 ทรีทเมนต์ กำหนดขนาดตัวอย่างและความแปรปรวนเท่ากันทุกกลุ่ม โดยที่ขนาดตัวอย่างมีค่า 5, 7, 10 และ 15 และความแปรปรวนมีค่าเท่ากับ 0.5 และ 1.0 กำหนดระดับนัยสำคัญ 0.05 ผลการศึกษาพบว่าสถิติทดสอบทั้งหมด สามารถควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบ ได้ทุกกรณีที่ศึกษา เมื่อพิจารณาค่ากำลังการทดสอบ พบว่าสถิติทดสอบบูทสแตรปทั้งสองวิธี ส่วนใหญ่มีกำลังการทดสอบสูงกว่าสถิติทดสอบตูกีย์ ในทุกการแจกแจงที่ศึกษา

คำสำคัญ : การเปรียบเทียบพหุคูณ, สถิติทดสอบตูกีย์, สถิติทดสอบสเต็ปดาวน์บูทสแตรปมินพี

***Corresponding author. E-mail** : mongkol.le@rmuti.ac.th

Abstract

The purpose of this research is to compare the performance of three test statistics for multiple comparison test which are Tukey's test, Step-down independent bootstrap min P and Step-down dependent bootstrap min P. Five distributions of data sets, Normal, Laplace, Logistic, log-normal and Gamma distributions are considered. A Monte Carlo simulation is performed with repeated 5,000 times. The number of bootstrap resampling is 1,000 times. The number of treatments are 3 and 4 with equal sample sizes and equal variance. The sample sizes are 5, 7, 10 and 15. The variance of all data sets are 0.5 and 1.0. The significance level will be set at 0.05. The results show that the familywise error rate of three test statistics can be controlled for all cases. Most power of the test of both step-down bootstrap min P procedures are higher than that of Tukey's test for all distributions.

Keywords : multiple comparison test, Tukey's test, Step-down bootstrap min P procedure

บทนำ

การทดสอบสมมุติฐานสำหรับค่าเฉลี่ยมากกว่า 2 ประชากรขึ้นไปนั้น จะใช้วิธีการทางสถิติที่เรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance) ซึ่งใช้สถิติทดสอบเอฟ โดยการวิเคราะห์ความแปรปรวนเป็นการทดสอบในภาพรวม คือทดสอบความแตกต่างของค่าเฉลี่ยพร้อมกันทุกค่า ซึ่งถ้าผลการทดสอบไม่มีนัยสำคัญแสดงว่าค่าเฉลี่ยของทุกประชากรไม่แตกต่างกัน แต่ถ้าผลการทดสอบมีนัยสำคัญทางสถิติ แสดงว่ามีค่าเฉลี่ยอย่างน้อย 2 ค่าที่แตกต่างกัน ซึ่งผลสรุปดังกล่าวยังไม่สามารถบอกได้ว่าค่าเฉลี่ยของประชากรกลุ่มใดบ้างที่แตกต่างกัน ขั้นตอนต่อไปจึงต้องมีการทดสอบเป็นรายคู่เรียกว่า การเปรียบเทียบพหุคูณ (Multiple comparison) เพื่อตรวจสอบให้แน่ชัดว่าค่าเฉลี่ยประชากรกลุ่มใดบ้างที่แตกต่างกัน

ในการทดสอบสมมุติฐานแต่ละคู่จะมีความผิดพลาดแบบที่ 1 (Type I error) เกิดขึ้น โดยการทดสอบด้วยวิธีการทางสถิติที่เลือกใช้จะต้องสามารถควบคุมอัตราการเกิดความผิดพลาดดังกล่าวให้มีค่าคงที่เท่ากับระดับนัยสำคัญที่กำหนด แต่การเปรียบเทียบพหุคูณเป็นการทดสอบสมมุติฐานหลายสมมุติฐานพร้อมกัน (Simultaneous tests) ซึ่งวิธีการทางสถิติที่เลือกใช้จะต้องสามารถควบคุมความผิดพลาดของทุกคู่ได้ นั่นคือต้องสามารถควบคุมอัตราการเกิดความผิดพลาดแบบที่ 1 อย่างน้อยหนึ่งครั้งในทุกคู่การทดสอบให้มีค่าคงที่เท่ากับระดับนัยสำคัญที่กำหนด โดยอัตราความผิดพลาดนี้เรียกว่า "อัตราความผิดพลาดต่อวงศ์การทดสอบ" (Familywise error rate: FWER) (Richter and McCann, 2013) ซึ่งวิธีการที่มีประสิทธิภาพและนิยมใช้ในการเปรียบเทียบพหุคูณคือ การทดสอบของตูกีย์ (Tukey's test)

อย่างไรก็ตามในบางสถานการณ์ที่ไม่ทราบการแจกแจงร่วมของสถิติเนื่องจากไม่ทราบการแจกแจงของประชากรที่นำมาศึกษา จึงต้องอาศัยการสุ่มข้อมูลซ้ำด้วยวิธีบูตสแตรป (Bootstrap resampling) ซึ่งวิธีการดังกล่าวสามารถใช้ประมาณค่าพี (*p*-value) โดยไม่ต้องคำนึงถึงข้อตกลงเกี่ยวกับค่าพารามิเตอร์ของการแจกแจงร่วมของสถิติทดสอบ (Efron and Tibshirani, 1993)

Komonniramit (2008) ได้ทำการศึกษาวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี (Step-down Dependent Bootstrap min P) เปรียบเทียบกับสถิติทดสอบของดันเนตต์ ในกรณีที่ข้อมูลมีการแจกแจงปรกติ กำหนดให้มี 3 ทรีทเมนต์และมีทรีทเมนต์ควบคุมอีก 1 ทรีทเมนต์ ในแต่ละทรีทเมนต์มีขนาดตัวอย่างเท่ากัน ผลการศึกษาพบว่าวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพีมีประสิทธิภาพใกล้เคียงกับสถิติทดสอบของดันเนตต์

Budsaba and Phetcharat (2011) ได้ทำการศึกษาความแกร่งของวิธีสเต็ปดาวน์บูทสแตรปมินพี (Step-down Bootstrap min P) สำหรับเปรียบเทียบค่าเฉลี่ยกรณีที่มีทรีทเมนต์ควบคุม เปรียบเทียบกับสถิติทดสอบของดันเนตต์ ศึกษาภายใต้ข้อมูลที่มีการแจกแจงล็อกนอร์มอลที่มีความแปรปรวนเท่ากัน กำหนดให้มี 3 ทรีทเมนต์และมีทรีทเมนต์ควบคุมอีก 1 ทรีทเมนต์ ในแต่ละทรีทเมนต์มีขนาดตัวอย่างเท่ากัน ผลการศึกษาพบว่า วิธีสเต็ปดาวน์บูทสแตรปมินพี มีประสิทธิภาพสูงกว่าสถิติทดสอบของดันเนตต์ ยกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 3

จากผลของงานวิจัยที่กล่าวมานั้น ผู้วิจัยจึงมีความสนใจที่จะนำวิธีสเต็ปดาวน์บูทสแตรปมินพี มาปรับใช้สำหรับการเปรียบเทียบพหุคูณทั่วไป โดยนำมาประยุกต์ใช้กับวิธีการทดสอบของตูกีย์ โดยศึกษาและเปรียบเทียบประสิทธิภาพของสถิติทดสอบจากวิธีการดังที่กล่าวมาแล้ว

วิธีดำเนินการวิจัย

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการทดสอบของวิธีการเปรียบเทียบพหุคูณ ด้วยสถิติทดสอบ 3 วิธี คือ สถิติทดสอบตูกีย์ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จำนวน 5,000 ครั้ง และทดลองซ้ำด้วยวิธีบูทสแตรป 1,000 ครั้ง กำหนดจำนวนทรีทเมนต์ที่ทำการศึกษาเท่ากับ 3 และ 4 ทรีทเมนต์ กำหนดขนาดตัวอย่างและความแปรปรวนเท่ากันทุกกลุ่ม โดยที่ขนาดตัวอย่างมีค่า 5, 7, 10 และ 15 และความแปรปรวนมีค่าเท่ากับ 0.5 และ 1.0 กำหนดค่าเฉลี่ยในแต่ละสถานการณ์ที่ศึกษา คือ กรณีจำนวนทรีทเมนต์เท่ากับ 3 ทรีทเมนต์ ค่าเฉลี่ยคือ $(\mu, \mu, \mu+\Delta)$ และ กรณีจำนวนทรีทเมนต์เท่ากับ 4 ทรีทเมนต์ ค่าเฉลี่ยคือ $(\mu, \mu, \mu, \mu+\Delta)$ เมื่อ $\mu=0$ และ Δ = 0.0, 0.5, 1.0 โดยมีเกณฑ์การพิจารณาคือ ค่าอัตราความผิดพลาดต่อวงศ์การทดสอบ และค่ากำลังการทดสอบ กำหนดระดับนัยสำคัญ 0.05 ศึกษาข้อมูลภายใต้การแจกแจงสมมาตร ได้แก่ การแจกแจงปรกติ การแจกแจงลาปลาซ การแจกแจงลอจิสติก และการแจกแจงเบ้ขวา ได้แก่ การแจกแจงล็อกนอร์มอล และการแจกแจงแกมมา ซึ่งมีรายละเอียดของสถิติทดสอบที่ศึกษาดังนี้

1. สถิติทดสอบตูกีย์

เป็นการเปรียบเทียบค่าเฉลี่ยของทรีทเมนต์ทุกคู่ที่เป็นไปได้จำนวน $\frac{k(k-1)}{2}$ คู่ โดยสามารถคำนวณค่าสถิติทดสอบได้ดังนี้ (Montgomery, 2009)

$$T=\frac{\left(\bar{x}_i-\bar{x}_j\right)}{\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}} \quad \text{เมื่อ } i \neq j \ ; \ i, j=1,2,\ldots,k \qquad (1)$$

โดยที่ $MSE = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i.}\right)^2}{n-k}$, n คือ จำนวนค่าสังเกตทั้งหมด, k คือ จำนวนทรีทเมนต์, n_i คือ ขนาดตัวอย่างของทรีทเมนต์ที่ i , x_{ij} คือ ค่าสังเกตที่ j ของทรีทเมนต์ที่ i , $\bar{x}_{i.}$ คือ ค่าเฉลี่ยของทรีทเมนต์ที่ i

2. สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนบูทสแตรปมินพี

สูตรที่ใช้

$$\tilde{p}_{(h)} = \max_{m=1,\dots,h}\left\{\Pr\left(\min_{r\in\left\{m,\dots,\frac{k!}{(k-2)!2!}\right\}} \mathrm{P}_r \le p_{(m)} \mid H_0\right)\right\} \tag{2}$$

งานวิจัยนี้ได้กำหนดให้มีจำนวนทรีทเมนต์ที่ทำการศึกษาเท่ากับ 3 และ 4 ทรีทเมนต์ ดังนั้นสมมุติฐานของการทดสอบเปรียบเทียบเชิงพหุคูณมีดังนี้

กรณี 3 ทรีทเมนต์

$$H_{01}:\mu_1=\mu_2 \text{ vs } H_{11}:\mu_1\ne\mu_2 \;;\; H_{02}:\mu_1=\mu_3 \text{ vs } H_{12}:\mu_1\ne\mu_3 \;;\; H_{03}:\mu_2=\mu_3 \text{ vs } H_{13}:\mu_2\ne\mu_3$$

กรณี 4 ทรีทเมนต์

$$H_{01}:\mu_1=\mu_2 \text{ vs } H_{11}:\mu_1\ne\mu_2 \;;\; H_{02}:\mu_1=\mu_3 \text{ vs } H_{12}:\mu_1\ne\mu_3 \;;\; H_{03}:\mu_1=\mu_4 \text{ vs } H_{13}:\mu_1\ne\mu_4$$
$$H_{04}:\mu_2=\mu_3 \text{ vs } H_{14}:\mu_2\ne\mu_3 \;;\; H_{05}:\mu_2=\mu_4 \text{ vs } H_{15}:\mu_2\ne\mu_4 \;;\; H_{06}:\mu_3=\mu_4 \text{ vs } H_{16}:\mu_3\ne\mu_4$$

สถิติทดสอบ

$$T_h = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\dfrac{MSE}{2}\left(\dfrac{1}{n_i}+\dfrac{1}{n_j}\right)}} \sim t_{n-k} \text{ , โดยที่ } h=1,2,\dots,\frac{k(k-1)}{2} \text{ และ } i\ne j\;;\; i,j=1,2,\dots,k \tag{3}$$

ทำการปรับค่าพีจากสูตร โดยแบ่งการพิจารณาตามจำนวนทรีทเมนต์ที่ศึกษาดังนี้ (Budsaba and Phetcharat, 2011)

กรณี 3 ทรีทเมนต์

$$\tilde{p}_{(1)} = \max\left\{\Pr\left(\min_{r\in\{1,2,3\}} \mathrm{P}_r \le p_{(1)}\right)\right\} = \Pr\left(\min(P_1,P_2,P_3) < p_{(1)}\right) \tag{4}$$

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{r\in\{1,2,3\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3\}} \mathrm{P}_r \le p_{(2)}\right)\right\} \tag{5}$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{r\in\{1,2,3\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3\}} \mathrm{P}_r \le p_{(2)}\right), \Pr\left(\min_{r\in\{3\}} \mathrm{P}_r \le p_{(3)}\right)\right\} \tag{6}$$

กรณี 4 ทรีทเมนต์

$$\tilde{p}_{(1)} = \max\left\{\Pr\left(\min_{r\in\{1,2,\dots,6\}} \mathrm{P}_r \le p_{(1)}\right)\right\} = \Pr\left(\min(P_1,P_2,\dots,P_6) < p_{(1)}\right) \tag{7}$$

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{r\in\{1,2,...,6\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3,...,6\}} \mathrm{P}_r \le p_{(2)}\right)\right\} \quad (8)$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{r\in\{1,2,...,6\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3,...,6\}} \mathrm{P}_r \le p_{(2)}\right), \Pr\left(\min_{r\in\{3,4,5,6\}} \mathrm{P}_r \le p_{(3)}\right)\right\} \quad (9)$$

$$\tilde{p}_{(4)} = \max\left\{\Pr\left(\min_{r\in\{1,2,...,6\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3,...,6\}} \mathrm{P}_r \le p_{(2)}\right), \Pr\left(\min_{r\in\{3,4,5,6\}} \mathrm{P}_r \le p_{(3)}\right),\right. \quad (10)$$
$$\left.\Pr\left(\min_{r\in\{4,5,6\}} \mathrm{P}_r \le p_{(4)}\right)\right\}$$

$$\tilde{p}_{(5)} = \max\left\{\Pr\left(\min_{r\in\{1,2,...,6\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3,...,6\}} \mathrm{P}_r \le p_{(2)}\right), \Pr\left(\min_{r\in\{3,4,5,6\}} \mathrm{P}_r \le p_{(3)}\right),\right. \quad (11)$$
$$\left.\Pr\left(\min_{r\in\{4,5,6\}} \mathrm{P}_r \le p_{(4)}\right), \Pr\left(\min_{r\in\{5,6\}} \mathrm{P}_r \le p_{(5)}\right)\right\}$$

$$\tilde{p}_{(6)} = \max\left\{\Pr\left(\min_{r\in\{1,2,...,6\}} \mathrm{P}_r \le p_{(1)}\right), \Pr\left(\min_{r\in\{2,3,...,6\}} \mathrm{P}_r \le p_{(2)}\right), \Pr\left(\min_{r\in\{3,4,5,6\}} \mathrm{P}_r \le p_{(3)}\right),\right. \quad (12)$$
$$\left.\Pr\left(\min_{r\in\{4,5,6\}} \mathrm{P}_r \le p_{(4)}\right), \Pr\left(\min_{r\in\{5,6\}} \mathrm{P}_r \le p_{(5)}\right), \Pr\left(\min_{r\in\{6\}} \mathrm{P}_r \le p_{(6)}\right)\right\}$$

เมื่อ $p_{(m)}$ คือ ค่าพีที่คำนวณได้จากการทดสอบสมมุติฐานดังที่กล่าวไว้ข้างต้นจำนวน 3 ค่า (กรณี 3 ทรีทเมนต์; สมการ (4)-(6)) และ 6 ค่า (กรณี 4 ทรีทเมนต์; สมการ (7)-(12)) ซึ่งคำนวณจากข้อมูลดั้งเดิม (Original data)

P_r คือ ค่าพีที่คำนวณได้จากการทดสอบสมมุติฐานดังที่กล่าวไว้ข้างต้นจำนวน 3 ค่า (กรณี 3 ทรีทเมนต์) และ 6 ค่า (กรณี 4 ทรีทเมนต์) ซึ่งคำนวณหลังจากการสุ่มซ้ำจากข้อมูลทั้งหมดโดยวิธีบูทสแตรปภายใต้สมมุติฐานว่าง

3. สถิติทดสอบสเต็ปดาวน์ดิเพนเดนบูทสแตรปมินพี

การคำนวณค่าสถิติทดสอบสามารถคำนวณได้โดยใช้วิธีเช่นเดียวกับ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนบูท-สแตรปมินพีทุกขั้นตอน แต่ส่วนที่แตกต่างกันคือค่า P_r จะคำนวณได้จากการทดสอบสมมุติฐานดังที่กล่าวไว้ข้างต้นจำนวน 3 ค่า (กรณี 3 ทรีทเมนต์) และ 6 ค่า (กรณี 4 ทรีทเมนต์) โดยคำนวณหลังจากการสุ่มซ้ำด้วยวิธีบูทสแตรป ที่ได้จากการคัดลอกข้อมูลดั้งเดิมทั้งชุด ตัวอย่างเช่น พิจารณากรณี 3 ทรีทเมนต์ ทำการจำลองข้อมูลได้ทั้งหมด 3 กลุ่ม $\underset{\sim}{x}_{1n}, \underset{\sim}{x}_{2n}, \underset{\sim}{x}_{3n}$ เรียกว่าข้อมูลดั้งเดิม จากนั้นทำการคัดลอกข้อมูลจากทั้ง 3 กลุ่มเดิม $\underset{\sim}{x}_{1n}, \underset{\sim}{x}_{2n}, \underset{\sim}{x}_{3n}, \underset{\sim}{x}_{1n}, \underset{\sim}{x}_{2n}, \underset{\sim}{x}_{3n}$ รวมมีข้อมูลทั้งหมด 6 กลุ่มขนาด 6n จากนั้นสุ่มข้อมูล มาทั้งหมด 3n ค่า แล้วนำข้อมูลที่ได้มาวิเคราะห์ในขั้นต่อไป (Budsaba and Phetcharat, 2011)

4. อัตราความผิดพลาดต่อวงศ์การทดสอบ (Familywise error rate : FWER)

คือความน่าจะเป็นที่ปฏิเสธสมมุติฐานอย่างน้อยหนึ่งคู่ เมื่อสมมุติฐานว่างทุกคู่ในชุดการทดสอบเป็นจริง โดยจะต้องมีค่าใกล้เคียงกับระดับนัยสำคัญที่กำหนด (Horn and Dunnett, 2004) ซึ่งค่าประมาณดังกล่าวคำนวณได้จาก $\hat{\alpha} = \dfrac{\theta}{5000}$, เมื่อ θ คือ จำนวนครั้งที่ปฏิเสธสมมุติฐานว่างอย่างน้อยหนึ่งครั้งในชุดการทดสอบเมื่อสมมุติฐานว่างทุกคู่ในชุดการทดสอบเป็นจริง สำหรับเกณฑ์ที่ใช้พิจารณาอัตราความผิดพลาดต่อวงศ์การทดสอบ คือ เกณฑ์ของ Bradley (อ้างถึงใน Haidous and Sawilowsky, 2013) เมื่อกำหนดระดับนัยสำคัญเท่ากับ 0.05 จะใช้เกณฑ์ $(0.025, 0.075)$ กล่าวคือ หาก

ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ $(\hat{\alpha})$ ตกอยู่ในเกณฑ์ดังกล่าวจะถือว่าสถิติทดสอบมีความสามารถในการควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบ

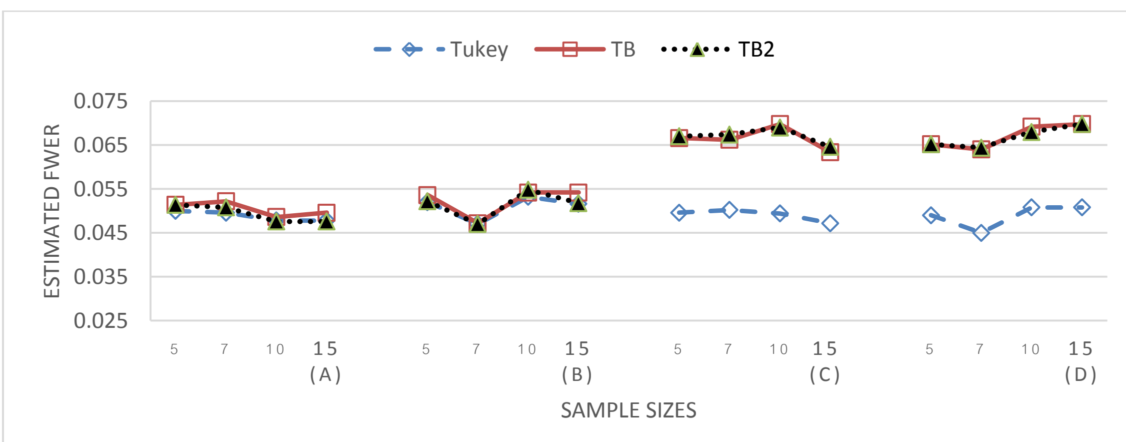
5. กำลังการทดสอบ (Power of the test)

คือความน่าจะเป็นที่ปฏิเสธสมมุติฐานว่างได้ถูกต้องทุกคู่ เมื่อสมมุติฐานว่างเป็นเท็จ ซึ่งสามารถประมาณค่ากำลังการทดสอบได้จาก $\frac{I}{5000}$ เมื่อ I คือจำนวนครั้งที่ปฏิเสธสมมุติฐานว่างได้ถูกต้องทุกคู่ เมื่อสมมุติฐานว่างเป็นเท็จ

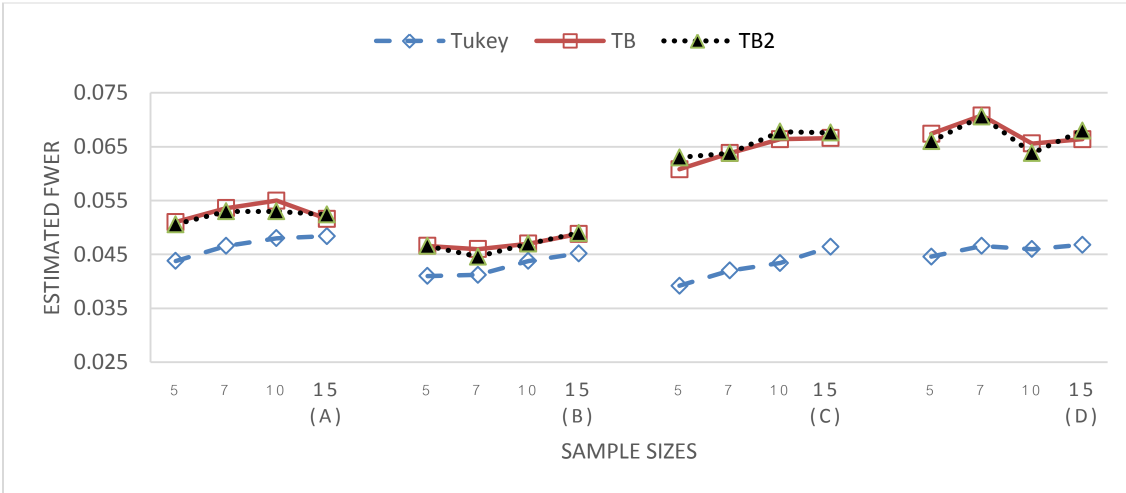
ผลการวิจัย

สำหรับการนำเสนอผลการวิจัยแบ่งออกเป็น 2 ส่วนได้แก่ ส่วนที่ 1 แสดงกราฟของค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ และส่วนที่ 2 แสดงตารางของค่าประมาณกำลังการทดสอบ โดยกำหนดสัญลักษณ์ คือ Tukey แทน สถิติทดสอบตูกีย์ TB แทน สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และ TB2 แทน สถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี และได้ผลการวิจัยดังนี้

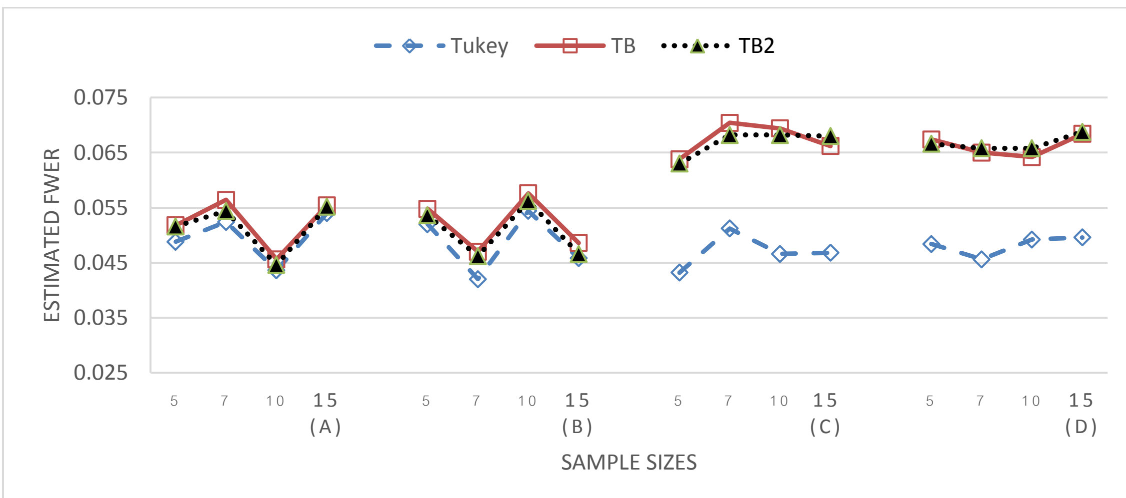
พิจารณาภาพที่ 1-5 สำหรับข้อมูลทุกการแจกแจงพบว่ากรณีจำนวนทรีทเมนต์เท่ากับ 3 กลุ่ม สถิติทดสอบตูกีย์ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี มีค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบใกล้เคียงกัน ส่วนกรณีที่จำนวนกลุ่มทรีทเมนต์เท่ากับ 4 กลุ่ม ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ ของสถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพีมีค่าใกล้เคียงกันแต่มีค่ามากกว่าสถิติทดสอบตูกีย์ค่อนข้างมาก นอกจากนี้ยังพบอีกว่าค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบของสถิติทดสอบตูกีย์ใกล้เคียงกับค่าระดับนัยสำคัญที่กำหนดทุกกรณีที่ศึกษา โดยค่าไม่ได้เพิ่มตามจำนวนทรีทเมนต์ที่เพิ่มขึ้น อย่างไรก็ตามเมื่อพิจารณารวมทุกกรณีที่ศึกษาพบว่าสถิติทดสอบทั้งหมดสามารถควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบ ได้ตามเกณฑ์ของ Bradley



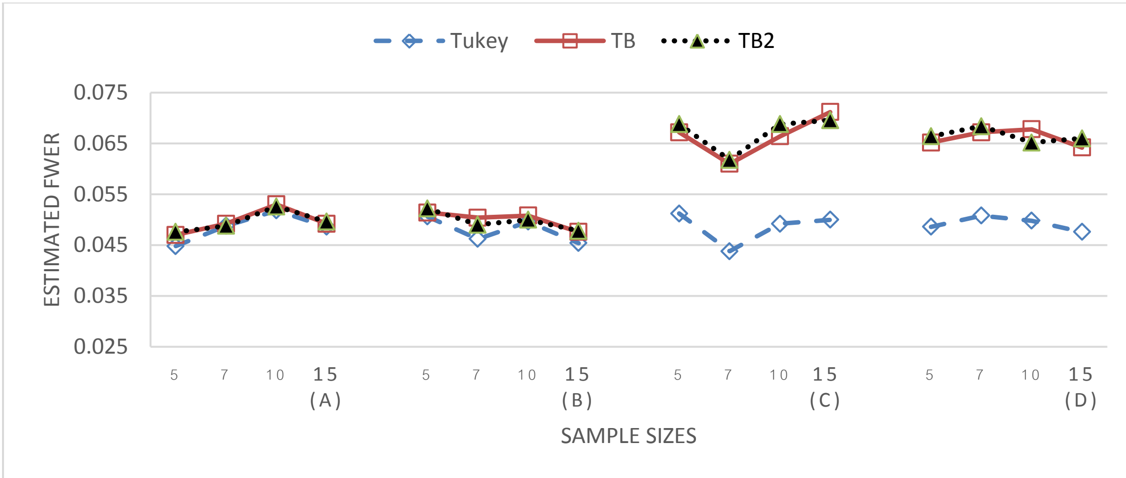
ภาพที่ 1 ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ เมื่อข้อมูลมีการแจกแจงปรกติ แบ่งตามกรณีดังนี้ (A) $k=3, \sigma=0.5$, (B) $k=3, \sigma=1$, (C) $k=4, \sigma=0.5$ และ (D) $k=4, \sigma=1$



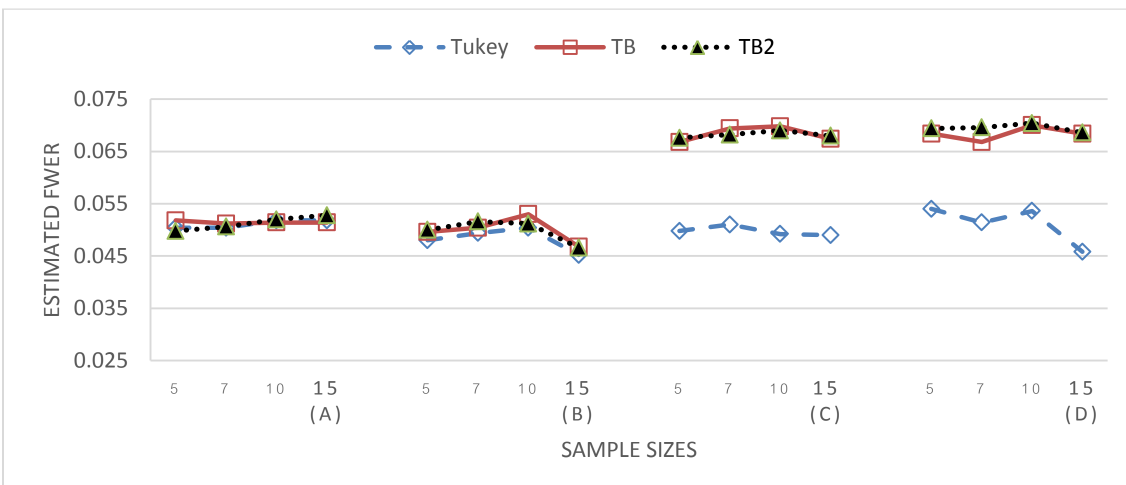
ภาพที่ 2 ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ เมื่อข้อมูลมีการแจกแจงลาปลาซ แบ่งตามกรณีดังนี้ (A) $k=3, \sigma=0.5$, (B) $k=3, \sigma=1$, (C) $k=4, \sigma=0.5$ และ (D) $k=4, \sigma=1$



ภาพที่ 3 ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ เมื่อข้อมูลมีการแจกแจงลอจิสติก แบ่งตามกรณีดังนี้ (A) $k=3, \sigma=0.5$, (B) $k=3, \sigma=1$, (C) $k=4, \sigma=0.5$ และ (D) $k=4, \sigma=1$



ภาพที่ 4 ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ เมื่อข้อมูลมีการแจกแจงล็อกนอร์มอลแบ่งตามกรณีดังนี้ (A) $k=3, \sigma=0.5$, (B) $k=3, \sigma=1$, (C) $k=4, \sigma=0.5$ และ (D) $k=4, \sigma=1$



ภาพที่ 5 ค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบ เมื่อข้อมูลมีการแจกแจงแกมมา แบ่งตามกรณีดังนี้ (A) $k=3, \sigma=0.5$, (B) $k=3, \sigma=1$, (C) $k=4, \sigma=0.5$ และ (D) $k=4, \sigma=1$

พิจารณาตารางที่ 1-4 ค่าประมาณกำลังการทดสอบของสถิติทดสอบตูกีย์ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี จะเพิ่มขึ้นเมื่อขนาดตัวอย่างและความแตกต่างของทรีทเมนต์เพิ่มขึ้น ในขณะที่เมื่อความแปรปรวนของข้อมูลทุกกลุ่มเพิ่มมากขึ้น ค่าประมาณกำลังการทดสอบของทุกสถิติทดสอบจะมีค่าลดลง เนื่องจากข้อมูลมีการกระจายมากขึ้น สำหรับทุกการแจกแจงที่ศึกษาคือ การแจกแจงปรกติ การแจกแจงลาปลาซ การแจกแจงลอจิสติก การแจกแจงล็อกนอร์มอล และการแจกแจงแกมมา

ตารางที่ 1 ค่าประมาณกำลังการทดสอบกรณีจำนวนทรีทเมนต์เท่ากับ 3 กลุ่ม ($\mu_1 = \mu_2 = 0, \mu_3 = 0.5$)

การแจกแจง	n	$\sigma^2 = 0.5$			$\sigma^2 = 1$		
		Tukey	TB	TB2	Tukey	TB	TB2
Normal	5	0.1558	0.1596*	0.1576	0.0992	0.1004	0.1010*
	7	0.2200	0.2224*	0.2216	0.1372	0.1374	0.1394*
	10	0.3146	0.3194	0.3198*	0.1712	0.1726	0.1738*
	15	0.4616	0.4622	0.4636*	0.2548	0.2552	0.2566*
Laplace	5	0.1880	0.1964	0.2028*	0.1116	0.1246*	0.1230
	7	0.2490	0.2612	0.2634*	0.1384	0.1468	0.1482*
	10	0.3458	0.3558	0.3564*	0.1930	0.1980	0.1982*
	15	0.4880	0.4960	0.4986*	0.2684	0.2792*	0.2774
Logistic	5	0.1694	0.1702	0.1708*	0.1052	0.1112	0.1114*
	7	0.2316	0.2364*	0.2360	0.1374	0.1422*	0.1418
	10	0.3306	0.3368*	0.3366	0.1866	0.1914*	0.1910
	15	0.4794	0.4852*	0.4842	0.2640	0.2706*	0.2672
Lognormal	5	0.1604	0.1630*	0.1620	0.1084	0.1100	0.1118*
	7	0.2170	0.2180*	0.2174	0.1284	0.1336*	0.1336*
	10	0.3222	0.3258*	0.3238	0.1706	0.1744*	0.1712
	15	0.4740	0.4746	0.4766*	0.2552	0.2596	0.2598*
Gamma	5	0.1542	0.1572	0.1578*	0.0994	0.1040*	0.1032
	7	0.2088	0.2132	0.2146*	0.1370	0.1378*	0.1372
	10	0.3184	0.3204	0.3208*	0.1740	0.1758	0.1760*
	15	0.4714	0.4724	0.4726*	0.2600	0.2628*	0.2618

* สถิติทดสอบที่มีค่าประมาณกำลังการทดสอบสูงที่สุด

ตารางที่ 2 ค่าประมาณกำลังการทดสอบกรณีจำนวนทรีทเมนต์เท่ากับ 3 กลุ่ม ($\mu_1 = \mu_2 = 0, \mu_3 = 1.0$)

การแจกแจง	n	$\sigma^2 = 0.5$			$\sigma^2 = 1$		
		Tukey	TB	TB2	Tukey	TB	TB2
Normal	5	0.5156	**0.5196***	0.5166	0.2742	**0.2776***	0.2750
	7	0.6968	0.6996	**0.7000***	0.4092	0.4094	**0.4124***
	10	0.8676	**0.8704***	0.8680	0.5698	**0.5714***	0.5704
	15	0.9782	0.9782	**0.9788***	0.7754	**0.7770***	0.7760
Laplace	5	0.5516	0.5624	**0.5636***	0.3142	0.3300	**0.3306***
	7	0.7174	**0.7250***	0.7248	0.4460	0.4592	**0.4610***
	10	0.8698	0.8742	**0.8748***	0.5920	0.6020	**0.6066***
	15	0.9658	0.9658	**0.9668***	0.7774	0.7846	**0.7850***
Logistic	5	0.5246	**0.5292***	0.5252	0.2936	**0.3000***	0.2980
	7	0.7126	**0.7166***	**0.7166***	0.4228	0.4274	**0.4320***
	10	0.8712	0.8700	**0.8728***	0.5800	0.5836	**0.5848***
	15	0.9694	0.9696	**0.9708***	0.7762	0.7790	**0.7808***
Lognormal	5	0.4972	**0.4992***	0.4990	0.2812	0.2862	**0.2864***
	7	0.6969	**0.6984***	0.6978	0.4164	0.4176	**0.4182***
	10	0.8822	**0.8836***	0.8828	0.5714	0.5730	**0.5732***
	15	0.9748	0.9756	**0.9762***	0.7702	**0.7718***	0.7712
Gamma	5	0.5082	0.5092	**0.5112***	0.2840	**0.2878***	0.2856
	7	0.7086	0.7112*	**0.7098**	0.3976	**0.3992**	0.4012*
	10	0.8788	0.8784	**0.8802***	0.5774	**0.5820***	0.5788
	15	0.9760	0.9766*	**0.9762**	0.7722	**0.7724**	0.7732*

* สถิติทดสอบที่มีค่าประมาณกำลังการทดสอบสูงที่สุด

ตารางที่ **3** ค่าประมาณกำลังการทดสอบกรณีจำนวนทรีทเมนต์เท่ากับ 4 กลุ่ม ($\mu_1 = \mu_2 = \mu_3 = 0, \mu_4 = 0.5$)

การแจกแจง	n	$\sigma^2 = 0.5$			$\sigma^2 = 1$		
		Tukey	TB	TB2	Tukey	TB	TB2
Normal	5	0.1480	**0.1848***	0.1824	0.0882	0.1166	**0.1182***
	7	0.2130	**0.2574***	0.2558	0.1214	0.1528	**0.1538***
	10	0.3158	0.3676	**0.3688***	0.1712	0.2072	**0.2098***
	15	0.4830	0.5360	**0.5412***	0.2454	0.2922	**0.2972***
Laplace	5	0.1724	**0.2224***	0.2218	0.0978	0.1334	**0.1380***
	7	0.2328	0.2870	**0.2874***	0.1238	0.1648	**0.1686***
	10	0.3262	0.3900	**0.3906***	0.1178	0.1504	**0.1518***
	15	0.4740	**0.5352***	0.5340	0.2564	**0.3148***	0.3112
Logistic	5	0.1596	**0.1966***	0.1956	0.1080	**0.1388***	0.1386
	7	0.2184	0.2664	**0.2666***	0.1260	**0.1662***	0.1650
	10	0.3108	**0.3670***	0.3668	0.1708	0.2094	**0.2108***
	15	0.4726	**0.5344***	0.5342	0.2346	0.2832	**0.2876***
Lognormal	5	0.1446	**0.1860***	0.1848	0.1066	**0.1352***	0.1328
	7	0.2048	0.2530	**0.2534***	0.1320	0.1608	**0.1638***
	10	0.3138	**0.3654***	0.3646	0.1676	**0.2112***	0.2082
	15	0.4534	**0.5120***	0.5094	0.2376	**0.2834***	0.2832
Gamma	5	0.1552	0.1892	**0.1908***	0.1036	**0.1302***	0.1296
	7	0.2178	**0.2696***	0.2654	0.1204	**0.1524***	**0.1524***
	10	0.2968	**0.3546***	0.3524	0.1706	**0.2154***	0.2136
	15	0.4626	0.5224	**0.5256***	0.2378	0.2882	**0.2890***

* สถิติทดสอบที่มีค่าประมาณกำลังการทดสอบสูงที่สุด

ตารางที่ 4 ค่าประมาณกำลังการทดสอบกรณีจำนวนทรีทเมนต์เท่ากับ 4 กลุ่ม ($\mu_1 = \mu_2 = \mu_3 = 0, \mu_4 = 1.0$)

การแจกแจง	n	$\sigma^2 = 0.5$			$\sigma^2 = 1$		
		Tukey	TB	TB2	Tukey	TB	TB2
Normal	5	0.4912	0.5502	**0.5524***	0.2736	0.3180	**0.3188***
	7	0.6966	0.7486	**0.7490***	0.3924	**0.4508***	0.4502
	10	0.8722	0.9018	**0.9022***	0.5662	**0.6222***	0.6212
	15	0.9810	**0.9878***	0.9870	0.7726	**0.8150***	0.8128
Laplace	5	0.5380	0.5990	**0.5994***	0.3072	0.3596	**0.3636***
	7	0.7146	**0.7702***	0.7692	0.4240	**0.4924***	0.4920
	10	0.8754	**0.9008***	0.8998	0.5886	**0.6484***	**0.6484***
	15	0.9694	**0.9772***	0.9768	0.7800	0.8178	**0.8216***
Logistic	5	0.5298	**0.5970***	0.5954	0.2768	**0.3300***	0.3290
	7	0.7138	0.7582	**0.7606***	0.3984	0.4546	**0.4584***
	10	0.8802	0.9086	**0.9090***	0.5890	**0.6462***	0.6456
	15	0.9744	0.9820	**0.9822***	0.7844	0.8238	**0.8248***
Lognormal	5	0.4964	0.5590	**0.5602***	0.2632	0.3114	**0.3150***
	7	0.6904	0.7412	**0.7428***	0.3986	0.4580	**0.4588***
	10	0.8780	0.9066	**0.9076***	0.5576	0.6100	**0.6112***
	15	0.9768	**0.9850***	**0.9850***	0.7666	0.8112	**0.8132***
Gamma	5	0.5084	0.5720	**0.5724***	0.2748	0.3226	**0.3252***
	7	0.7170	0.7666	**0.7672***	0.3930	**0.4504***	0.4498
	10	0.8780	0.9062	**0.9068***	0.5672	0.6210	**0.6222***
	15	0.9796	0.9850	**0.9860***	0.7760	0.8250	**0.8262***

* สถิติทดสอบที่มีค่าประมาณกำลังการทดสอบสูงที่สุด

เมื่อพิจารณาทั้งจำนวนทรีทเมนต์เท่ากับ 3 และ 4 กลุ่ม พบว่า โดยภาพรวมแล้วค่าประมาณกำลังการทดสอบของสถิติทดสอบทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพีสูงกว่าสถิติทดสอบของตูกีย์อย่างชัดเจน ยกเว้นกรณีดังนี้

กรณีจำนวนทรีทเมนต์เท่ากับ 3 กลุ่ม ($\mu_1 = \mu_2 = 0, \mu_3 = 1.0$) ที่ความแปรปรวนเท่ากับ 0.5 ขนาดตัวอย่างเท่ากับ 15 พบว่าสถิติทดสอบตูกีย์และสถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพีมีค่ากำลังการทดสอบเท่ากันในการแจกแจงปรกติ และการแจกแจงลาปลาซ รวมถึงกรณีที่ขนาดตัวอย่างเท่ากับ 10 พบว่าสถิติทดสอบตูกีย์มีค่ากำลังการทดสอบสูงกว่าสถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพีเล็กน้อย ในการแจกแจงลอจิสติก และการแจกแจงแกมม่า
สังเกตได้จากตารางที่ 2

วิจารณ์ผลการวิจัย

จากผลการวิจัย พบว่าสถิติทดสอบบูทสแตรปมินพีทั้งสองวิธีจะมีประสิทธิภาพดีกว่าวิธีดั้งเดิม ในกรณีนี้ คือ สถิติทดสอบตูกีย์ ซึ่งสอดคล้องกับงานวิจัยของ Budsaba and Phetcharat (2011) ที่ศึกษากรณีที่มีกลุ่มควบคุม พบว่าสถิติทดสอบบูทสแตรปมินพีมีประสิทธิภาพดีกว่าวิธีดั้งเดิม คือ สถิติทดสอบดันเนทท์ นอกจากนั้นเมื่อพิจารณาค่ากำลังการทดสอบของสถิติทดสอบทั้งหมด ยังพบว่ากรณีที่มีจำนวนทรีทเมนต์ 4 กลุ่ม ค่าประมาณกำลังการทดสอบของสถิติทดสอบบูทสแตรปมินพีทั้งสองแบบมีค่ามากกว่าสถิติทดสอบตูกีย์อย่างเห็นได้ชัดกว่ากรณีที่มีจำนวนทรีทเมนต์ 3 กลุ่ม

สรุปผลการวิจัย

พิจารณาค่าประมาณอัตราความผิดพลาดต่อวงศ์การทดสอบภายใต้ข้อมูลที่มีการแจกแจงสมมาตร ได้แก่ การแจกแจงปรกติ การแจกแจงลาปลาซ และการแจกแจงลอจิสติก รวมไปถึงการแจกแจงเบ้ขวา ได้แก่ การแจกแจงล็อกนอร์มอล และการแจกแจงแกมมา สรุปได้ว่าสถิติทดสอบทั้ง 3 วิธี คือ สถิติทดสอบตูกีย์ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี สามารถควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบ ได้ทุกกรณีที่ศึกษา เมื่อพิจารณาค่ากำลังการทดสอบ พบว่า สถิติทดสอบทั้ง 3 วิธีมีค่าประมาณกำลังการทดสอบเพิ่มขึ้น เมื่อขนาดตัวอย่างและขนาดความแตกต่างของทรีทเมนต์เพิ่มขึ้น ในขณะที่เมื่อความแปรปรวนของทุกกลุ่มเพิ่มขึ้น ค่าประมาณกำลังการทดสอบของสถิติทดสอบทุกวิธีลดลง เนื่องจากข้อมูลมีการกระจายมากขึ้น

สำหรับสถิติทดสอบบูทสแตรปทั้งสองวิธี คือ สถิติทดสอบสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรปมินพี และสถิติทดสอบสเต็ปดาวน์ดิเพนเดนท์บูทสแตรปมินพี นั้น ส่วนใหญ่มีประสิทธิภาพดีกว่าสถิติทดสอบตูกีย์ ทุกการแจกแจงที่ศึกษา ซึ่งแสดงให้เห็นว่าสถิติทดสอบดังกล่าวมีความแกร่ง กล่าวคือ สามารถใช้ได้ดีกับข้อมูลที่ไม่มีการแจกแจงปรกติ ทั้งที่มีลักษณะสมมาตรและเบ้ขวา

เอกสารอ้างอิง

Budsaba, K. & Phetcharat, A. (2011). Robustness study of step-down bootstrap min P for comparing several means with a control. *Journal of Science and Technology Thammasat University*, *19*(3), 29-39. (in Thai)

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

Haidous, N N., and Sawilowsky. S. S. (2013) "Robustness and Power of the Kornbrot Rank Difference, Signed Ranks, and Dependent Samples T-test." *American Journal of Applied Mathematics and Statistics*, *1*(5), 99-102.

Horn, M. & Dunnett, C.W. (2004). Power and sample size comparisons of stepwise FEW and FDR controlling test procedures in the normal many-one case. *Lecture Notes Monograph Series*, *47*, 48-64.

Komonniramit, Y. (2008). *Step-down dependent bootstrap min p for comparison several means with a control*. Master's thesis, Faculty of Science and Technology, Thammasat University. (In Thai)

Montgomery D.C. (2009). *Design and analysis of experiments.* (7th edition). Asia: John Wiley & Sons.

Richter, S. J., & McCann, M. H. (2013). Simultaneous multiple comparisons with a control using median differences and permutation tests. *Statistics & Probability Letters*, *83*(4), 1167-1173