การเปรียบเทียบตัวสถิติทดสอบไม่อิงพารามิเตอร์ ระหว่างประชากร 2 กลุ่ม เมื่อความแปรปรวนไม่เท่ากัน

A Comparison Study of Nonparametric Test Statistics for Two Populations in Case of Heterogeneity of Variances

นพดล วันชนะชัย และ ชินนะพงษ์ บำรุงทรัพย์
ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

บทคัดย่อ

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบตัวสถิติทดสอบไม่อิงพารามิเตอร์ 3 วิธี คือ สถิติทดสอบโดยวิธี Brunner and Munzel วิธีBootstrap Brunner and Munzel และ วิธีBootstrap Rank Welch ระหว่างประชากร 2 กลุ่ม เมื่อความแปรปรวนไม่เท่ากัน เพื่อเปรียบเทียบประสิทธิภาพของวิธีการทดสอบด้วยค่าอำนาจการทดสอบ โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ศึกษาในกรณีที่ประชากรมีการแจกแจงลอกนอร์มอล ที่ระดับนัยสำคัญ 0.05 ขนาดตัวอย่าง $(n_1, n_2) = (10,10)$, $(10,15)$, $(15,15)$ ผลการศึกษา พบว่า สถิติทดสอบโดยวิธี Bootstrap Rank Welch ให้ค่าอำนาจการทดสอบสูงที่สุด แต่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ ในขณะที่สถิติทดสอบโดยวิธี Brunner and Munzel ให้ค่าอำนาจการทดสอบสูงใกล้เคียงกับสถิติทดสอบโดยวิธี Bootstrap Rank Welch และสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel ไม่มีกรณีใดสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ และยังให้ค่าอำนาจการทดสอบต่ำสุดด้วย

คำสำคัญ: สถิติไม่อิงพารามิเตอร์ ความแปรปรวนไม่เท่ากัน

Abstract

This research is to compare 3 methods of nonparametric test statistics for two populations in case of heterogeneity of variances; the methods are Brunner and Munzel, Bootstrap Brunner and Munzel, and Bootstrap Rank Welch. The methods are to be compared by considering the ability to control the probability of type I error and the power of the test. The test is based on 0.05 level of significance whereas the populations are assumed to have a log-normal distribution with the sample sizes $(n_1, n_2) = (10,10),\ (10,15),\ (15,15)$. The results of the study are as follows: Bootstrap Rank Welch test statistic has the highest power of the test but lacks of ability to control probability of type I error whereas Brunner and Munzel has ability to control probability of type I error in all cases and has the power of the test close to Bootstrap Rank Welch. Bootstrap Brunner and Munzel test statistic has the lowest power of the test and lacks of ability to control probability of type I error.

Keywords: nonparametric statistics, heterogeneity of variance

1. บทนำ

ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างประชากร 2 กลุ่ม ที่เป็นอิสระกันเป็นการตรวจสอบว่าค่ากลางของลักษณะที่สนใจศึกษาในประชากรนั้นๆ แตกต่างกันหรือไม่ ซึ่งตัวสถิติที่เหมาะสมสำหรับการทดสอบความแตกต่างระหว่างประชากร 2 กลุ่ม คือ สถิติทดสอบ Z เมื่อทราบส่วนเบี่ยงเบนมาตรฐานของประชากร และสถิติทดสอบ t เมื่อไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร ภายใต้ข้อตกลงเบื้องต้นสำหรับการใช้ตัวสถิติทดสอบ คือ ข้อมูลมาจากประชากรที่นำมาใช้ในการทดสอบมีการแจกแจงปกติ แต่ในทางปฏิบัติข้อสมมตินี้อาจไม่เป็นจริง เพราะลักษณะของข้อมูลที่นำมาวิเคราะห์อาจมีการแจกแจงที่ไม่ใช่การแจกแจงปกติ ถ้าผู้วิจัยยังคงใช้สถิติทดสอบ t ผลสรุปที่ได้อาจขัดแย้งกับความเป็นจริง

สำหรับการศึกษาเปรียบเทียบประชากร 2 กลุ่ม กรณีที่ข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นมีผู้ศึกษาไว้หลายท่าน เช่น Brunner and Munzel ได้เสนอสถิติทดสอบที่ใช้สำหรับการทดสอบประชากร 2 กลุ่ม เมื่อความแปรปรวนไม่เท่ากัน ข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น ใช้สำหรับขนาดตัวอย่างมีขนาดเล็ก$(n_1, n_2 \le 10)$และ Brunner and Munzel พบว่าสถิติทดสอบ Wilcoxon Mann Whitney มีความไวต่อการปฏิเสธสมมติฐานหลัก(H_0) โดยที่การปฏิเสธสมมติฐานหลัก(H_0)ของสถิติทดสอบWilcoxon Mann Whitney ขึ้นอยู่กับอัตราส่วนของขนาดตัวอย่างและความแปรปรวนของประชากร โดยแสดงด้วยการจำลองข้อมูล[2] ในปี 2005 Reiczigel, Zakarias and Rozsa ได้เสนอสถิติทดสอบไม่อิงพารามิเตอร์ Bootstrap Rank Welch โดยใช้วิธีการสุ่มตัวอย่างโดยวิธี bootstrap เพื่อใช้ในการทดสอบความแตกต่างระหว่างประชากร 2 กลุ่ม ภายใต้สมมติฐาน H_0: $(X<Y) = P(X>Y)$ H_1: $P(X<Y) \neq P(X>Y)$ ซึ่งเรียกว่าความเท่ากันและไม่เท่ากันของสโทแคสติก (stochastic equality and inequality) และได้นำสถิติทดสอบนี้ไปเปรียบเทียบกับสถิติทดสอบอีก 3 วิธี คือ สถิติทดสอบWilcoxon Mann Whitney สถิติทดสอบ

Rank Welch และ สถิติทดสอบBrunner and Munzel พบว่า สถิติทดสอบ Bootstrap Rank Welch ให้ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าสถิติทดสอบRank Welch และ สถิติทดสอบ Wilcoxon Mann Whitney และสถิติทดสอบ Bootstrap Rank Welch ให้อำนาจการทดสอบสูงกว่าสถิติทดสอบ Wilcoxon Mann Whitney[5] ในปี 2006 Neuhauser, Losch and Jockel ศึกษาเปรียบเทียบสถิติทดสอบแบบไม่อิงพารามิเตอร์ 3 วิธี คือ Wilcoxon Mann Whitney, Brunner and Munzel และ Cliff's กับตัวสถิติทดสอบ Chen Luo พบว่า สถิติทดสอบ Wilcoxon Mann Whitney และ Chen Luo มีความไวต่อการปฏิเสธสมมติฐานหลัก โดยที่การปฏิเสธสมมติฐานหลักของสถิติทดสอบ Wilcoxon Mann Whitney และ Chen Luo ขึ้นอยู่กับอัตราส่วนของขนาดตัวอย่าง และความแปรปรวนของประชากร Neuhauser, Losch and Jockel จึงแนะนำให้ใช้สถิติทดสอบ Brunner and Munzel และ Cliff's [4]

จากการศึกษาผลงานวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงสนใจนำวิธี Bootstrap มาปรับใช้กับวิธี Brunner and Munzel และได้ทำการศึกษาเปรียบเทียบตัวสถิติทดสอบ Brunner and Munzel, Bootstrap Brunner and Munzel และ Bootstrap Rank Welch ระหว่างประชากร 2 กลุ่ม เมื่อความแปรปรวนไม่เท่ากัน โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ ด้วยค่าอำนาจการทดสอบ โดยศึกษาในกรณีที่ประชากร 2 กลุ่มมีความแปรปรวนไม่เท่ากัน

2. ขอบเขตของการวิจัย

1. กำหนดลักษณะของประชากรมีการแจกแจงลอกนอร์มอล(Log-normal distribution) มีพารามิเตอร์ ดังนี้

ตารางที่ 1 แสดงลักษณะของประชากรมีการแจกแจงสำหรับการคำนวณหาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และกำลังการทดสอบ

Type I error	Power of the test
$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,2)$	$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,2)$ $(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,2)$
$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,3)$	$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,3)$ $(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,3)$
$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,4)$	$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,4)$ $(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,4)$

ml_i คือ ค่าเฉลี่ยของการแจกแจงลอกนอร์มอลกลุ่มที่ i (i = 1, 2)

sl_i คือ ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงลอกนอร์มอล กลุ่มที่ i (i = 1, 2)

2. ตัวสถิติทดสอบที่ใช้ในการเปรียบเทียบมี 3 วิธี ได้แก่ สถิติทดสอบโดยวิธี Brunner and Munzel (BM) สถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel (BBM) และ สถิติทดสอบโดยวิธี Bootstrap Rank Welch (BRW)

3. กำหนดขนาดตัวอย่าง(n) ที่ใช้ในการศึกษาคือ $(n_1, n_2) = (10,10), (10,15), (15,15)$

4. กำหนดระดับนัยสำคัญ(α)เท่ากับ 0.05

5. จำนวนรอบของการสุ่มตัวอย่างโดยวิธี Bootstrap เท่ากับ 1,000 รอบ

6. จำลองข้อมูลให้มีสถานการณ์ตามที่กำหนดข้างต้นโดยวิธีการจำลองมอนติคาร์โล โดยใช้โปรแกรม R และทำการจำลองข้อมูลซ้ำ ๆ กัน 1,000 รอบ ในแต่ละสถานการณ์

7. คำนวณหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ในแต่ละสถานการณ์

8. คำนวณหาค่าอำนาจการทดสอบในแต่ละสถานการณ์

เกณฑ์ที่ใช้ในการศึกษาประสิทธิภาพของตัวสถิติทดสอบ

พิจารณาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1ในแต่ละสถานการณ์ ซึ่งในการวิจัยครั้งนี้ใช้เกณฑ์ของ Bradley [1] ดังนี้ กำหนดให้ τ คือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ที่เกิดจากการทดลอง สถิติทดสอบจะควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ก็ต่อเมื่อค่า τ อยู่ในช่วง $[0.5\alpha, 1.5\alpha]$ ที่ระดับนัยสำคัญ α นั่นคือ ค่า τ อยู่ในช่วง $[0.025, 0.075]$ ที่ระดับนัยสำคัญ 0.05

3. วิธีดำเนินการวิจัย

1. สร้างแบบจำลองข้อมูลของตัวอย่างทั้ง 2 กลุ่ม โดยกำหนดค่าเฉลี่ย ความแปรปรวน ขนาดตัวอย่าง และการแจกแจง ตามสถานการณ์ข้างต้น

2. คำนวณหาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 สามารถหาได้โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานหลัก(H_0) แล้วหารด้วยจำนวนรอบทั้งหมด ซึ่งมีขั้นตอนดังนี้

สมมติฐานของการทดลอง

H_0 : ประชากรทั้ง 2 กลุ่ม มีลักษณะไม่แตกต่างกัน

H_1 : ประชากรทั้ง 2 กลุ่ม มีลักษณะแตกต่างกัน

กำหนดลักษณะการแจกแจงของตัวอย่างทั้ง 2 กลุ่ม ให้เป็นไปตามสมมติฐานหลัก(H_0)

3. คำนวณค่าสถิติทดสอบทั้ง 3 วิธี โดยมีสูตรและวิธีการคำนวณดังต่อไปนี้

3.1 สถิติทดสอบโดยวิธี Brunner and Munzel มีวิธีการคำนวณค่าสถิติดังนี้

$$t_{BM} = \frac{n_1 n_2 \left(\overline{r}_2 - \overline{r}_1\right)}{\left(n_1 + n_2\right)\sqrt{n_1 s_1^2 + n_2 s_2^2}}$$

โดยที่ $$s_i^2 = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(r_{ik} - w_{ik} - \overline{r}_i + \frac{n_i + 1}{2}\right)^2$$

$$df_{BM} = \frac{\left(n_1 s_1^2 + n_2 s_2^2\right)^2}{\frac{\left(n_1 s_1^2\right)^2}{n_1 - 1} + \frac{\left(n_2 s_2^2\right)^2}{n_2 - 1}}$$

เมื่อ $\overline{r}_1$ คือ ค่าเฉลี่ยลำดับของตัวอย่างกลุ่มที่ 1

$\overline{r}_2$ คือ ค่าเฉลี่ยลำดับของตัวอย่างกลุ่มที่ 2

n_1 คือ ขนาดตัวอย่างกลุ่มที่ 1

n_2 คือ ขนาดตัวอย่างกลุ่มที่ 2

r_{ik} คือ ลำดับของตัวอย่างที่ k ของกลุ่มที่ i เมื่อนำข้อมูลทั้ง 2 กลุ่ม มาเรียงลำดับ

w_{ik} คือ ลำดับของตัวอย่างที่ k ของกลุ่มที่ i เมื่อนำข้อมูลในแต่ละกลุ่มมาเรียงลำดับ

s_i^2 คือ ความแปรปรวนของลำดับตัวอย่างกลุ่มที่ i

เงื่อนไขในการปฏิเสธสมมติฐานหลัก(H_0) จะปฏิเสธสมมติฐานหลัก(H_0) เมื่อค่าสถิติที่คำนวณได้(t_{BM}) มีค่าน้อยกว่า $-t_{\left(\alpha/2,\, df_{BM}\right)}$ หรือ มากกว่า $t_{\left(\alpha/2,\, df_{BM}\right)}$ โดยที่ $t_{\left(\alpha/2,\, df_{BM}\right)}$ แทนบริเวณวิกฤติทางด้านขวาที่ได้จากตาราง t และ $-t_{\left(\alpha/2,\, df_{BM}\right)}$ แทนบริเวณวิกฤติทางด้านซ้ายที่ได้จากตาราง t

3.2 สถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel มีขั้นตอนการสุ่มตัวอย่างโดยวิธี Bootstrap และ วิธีการคำนวณค่าสถิติดังนี้

- จากตัวอย่างทั้ง 2 กลุ่ม คือ X และ Y ทำการสุ่มตัวอย่างแบบแทนที่ (with replacement) จาก X และ Y ขนาด n_1 และ n_2 เขียนแทนด้วย X' และ Y' ตามลำดับ

- นำ X' และ Y' ที่ได้มาคำนวณสถิติทดสอบ Brunner and Munzel จำนวน B ครั้ง เมื่อ B คือ จำนวนรอบของการสุ่มตัวอย่างโดยวิธี Bootstrap ในที่นี้กำหนด B เท่ากับ 1,000 ครั้ง

- นำค่าสถิติทดสอบที่ได้จากการสุ่มตัวอย่างโดยวิธี Bootstrap ทั้ง 1,000 ครั้ง มาหาค่าเฉลี่ย และ เปรียบเทียบค่าเฉลี่ยของสถิติทดสอบที่ได้กับบริเวณวิกฤต โดยมีวิธีการคำนวณค่าสถิติดังที่กล่าวมาแล้วในข้อ 1.

3.3 สถิติทดสอบโดยวิธี Bootstrap Rank Welch มีขั้นตอนการสุ่มตัวอย่างโดยวิธี bootstrap ข้อมูล และ วิธีการคำนวณค่าสถิติดังนี้

- จากตัวอย่างทั้ง 2 กลุ่ม คือ X และ Y ทำการสุ่มตัวอย่างแบบแทนที่ (with replacement) จาก X และ Y ขนาด n_1 และ n_2 เขียนแทนด้วย X' และ Y' ตามลำดับ

- นำ X' และ Y' ที่ได้มาคำนวณสถิติทดสอบ Rank Welch จำนวน B ครั้ง เมื่อ B คือ จำนวนรอบของการ สุ่มตัวอย่างโดยวิธี Bootstrap ในที่นี้กำหนด B เท่ากับ 1,000 ครั้ง

- นำค่าสถิติทดสอบที่ได้จากการสุ่มตัวอย่างโดยวิธี Bootstrap ทั้ง 1,000 ครั้ง มาหาค่าเฉลี่ย และ เปรียบเทียบค่าเฉลี่ยของสถิติทดสอบที่ได้กับบริเวณวิกฤต โดยมีวิธีการคำนวณค่าสถิติดังนี้

$$t_{RW} = (\overline{r}_2 - \overline{r}_1) \Big/ \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

โดยที่ $$df_{RW} = \frac{\left(n_2 s_1^2 + n_1 s_2^2\right)^2}{\frac{\left(n_2 s_1^2\right)^2}{n_1 - 1} + \frac{\left(n_1 s_2^2\right)^2}{n_2 - 1}}$$

เมื่อ $\overline{r}_1$ คือ ค่าเฉลี่ยลำดับของตัวอย่างกลุ่มที่ 1

$\overline{r}_2$ คือ ค่าเฉลี่ยลำดับของตัวอย่างกลุ่มที่ 2

n_1 คือ ขนาดตัวอย่างกลุ่มที่ 1

n_2 คือ ขนาดตัวอย่างกลุ่มที่ 2

s_1^2 คือ ความแปรปรวนของลำดับตัวอย่างกลุ่มที่ 1

s_2^2 คือ ความแปรปรวนของลำดับตัวอย่างกลุ่มที่ 2

เงื่อนไขในการปฏิเสธสมมติฐานหลัก (H_0) จะปฏิเสธสมมติฐานหลัก (H_0) เมื่อค่าสถิติที่คำนวณได้ (t_{RW}) มีค่าน้อยกว่า $-t_{\left(\alpha/2,\, df_{RW}\right)}$ หรือ มากกว่า $t_{\left(\alpha/2,\, df_{RW}\right)}$

3.4 นำค่าสถิติทดสอบที่คำนวณได้ มาเปรียบเทียบกับค่าวิกฤตในแต่ละตัวสถิติทดสอบ กระทำการทดลองซ้ำ ๆ กัน 1,000 รอบ ในแต่ละสถานการณ์แล้วนับจำนวนครั้งของการปฏิเสธสมมติฐานหลัก (H_0) ทั้งหมด แล้วนำมาหารด้วยจำนวนรอบทั้งหมด 1,000 รอบ

3.5 คำนวณหาอำนาจการทดสอบ (Power of the test) โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานหลัก (H_0) แล้วหารด้วยจำนวนรอบทั้งหมด โดยกำหนดให้ลักษณะการแจกแจงของตัวอย่างให้เป็นไปตามสมมติฐานรอง (H_1) นั่นคือ มีการกำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของทั้ง 2 กลุ่มแตกต่างกันดังแสดงในตารางที่ 1 ส่วนขั้นตอนต่าง ๆ เหมือนกับการหาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ข้างต้น

4. ผลการวิจัย

จากตารางผนวกที่ 1 แสดงค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อ $\alpha = 0.05$ การแจกแจงลอกนอร์มอล และ ความแปรปรวนของประชากรแตกต่างกันพบว่า สถิติทดสอบโดยวิธี BM สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ดีทีสุดเมื่อ $n_1 = 10, n_2 = 10$ โดยมีค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เท่ากับ 0.053, 0.050 และ 0.060 เมื่อ $n_1 = 10, n_2 = 15$ มีค่าเท่ากับ 0.058, 0.049 และ 0.052 และ เมื่อ $n_1 = 15, n_2 = 15$ มีค่าเท่ากับ 0.051, 0.052 และ 0.055 สถิติทดสอบโดยวิธี BBM และ BRW ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

จากตารางผนวกที่ 2 แสดงค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ $n_1 = 10, n_2 = 10$ พบว่า สถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงที่สุด โดยมีค่าอำนาจการทดสอบ เท่ากับ 0.757 และ 1, 0.545 และ 0.956, 0.402 และ 0.849 ทั้ง 2 แบบของพารามิเตอร์ รองลงมา คือ สถิติทดสอบโดยวิธี BM และ BBM ตามลำดับ เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันน้อย สถิติทดสอบโดยวิธี BM และ BRW มีค่าอำนาจการทดสอบใกล้เคียงกัน เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันมาก ค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าลดลง และสถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงกว่า สถิติทดสอบโดยวิธี BM และ BBM เมื่อค่าเฉลี่ยแตกต่างกันมาก ค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าเพิ่มขึ้น

จากตารางผนวกที่ 3 แสดงค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ $n_1 = 10, n_2 = 15$ พบว่า สถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงที่สุด มีค่าอำนาจการทดสอบ เท่ากับ 0.902 และ 1, 0.677 และ 0.988, 0.492 และ 0.940 ทั้ง 2 แบบของพารามิเตอร์ รองลงมา คือ สถิติทดสอบโดยวิธี BM และ BBM ตามลำดับ เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันน้อย สถิติทดสอบโดยวิธี BM และ BRW มีค่าอำนาจการทดสอบใกล้เคียงกัน เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันมาก ค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าลดลง และสถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงกว่าสถิติทดสอบโดยวิธี BM และ BBM

เมื่อค่าเฉลี่ยแตกต่างกันมาก ค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าเพิ่มขึ้นจากตารางผนวกที่ 4 แสดงค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ $n_1 = 15, n_2 = 15$ พบว่า สถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงที่สุด มีค่าอำนาจการทดสอบ เท่ากับ 0.924 และ 1, 0.692 และ 0.993, 0.505 และ 0.949 ทั้ง 2 แบบของพารามิเตอร์ รองลงมา คือ สถิติทดสอบโดยวิธี BM และ BBM ตามลำดับ เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันน้อย สถิติทดสอบโดยวิธี BM และ BRW มีค่าอำนาจการทดสอบใกล้เคียงกัน เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกันมากค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าลดลง และสถิติทดสอบโดยวิธี BRW มีค่าอำนาจการทดสอบสูงกว่า สถิติทดสอบโดยวิธี BM และ BBM เมื่อค่าเฉลี่ยแตกต่างกันมาก ค่าอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธี มีค่าเพิ่มขึ้น

จากตารางผนวกที่ 2, 3 และ 4 พบว่า เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นค่าอำนาจการทดสอบของสถิติทดสอบ BM และ BRW จะเพิ่มขึ้น ยกเว้นสถิติทดสอบ BBM

5. สรุปผลการวิจัยและข้อเสนอแนะ

- ผลการเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

ที่ระดับนัยสำคัญ เท่ากับ 0.05 การแจกแจงลอกนอร์มอล สถิติทดสอบโดยวิธี Brunner and Munzel สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ทุกกรณี สถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel และ Bootstrap Rank Welch ไม่มีกรณีใดสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

- ผลการเปรียบเทียบอำนาจการทดสอบ

เมื่อความคลาดเคลื่อนของตัวอย่างทั้ง 2 กลุ่มแตกต่างกันน้อย สถิติทดสอบโดยวิธี Brunner and Munzel และ Bootstrap Rank Welch ให้ค่าอำนาจการทดสอบสูงใกล้เคียงกัน สถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel ให้ค่าอำนาจการทดสอบต่ำที่สุดใน 3 วิธี

เมื่อส่วนเบี่ยงเบนมาตรฐานของตัวอย่างทั้ง 2 กลุ่มแตกต่างกันมาก สถิติทดสอบโดยวิธี Bootstrap Rank Welch ให้ค่าอำนาจการทดสอบสูงกว่าวิธี Brunner and Munzel และ วิธี Bootstrap Brunner and Munzel

เมื่อกำหนดให้ค่าเฉลี่ยของตัวอย่างทั้ง 2 กลุ่ม แตกต่างกันมาก สถิติทดสอบโดยวิธี Brunner and Munzel วิธี Bootstrap Rank Welch และวิธี Bootstrap Brunner and Munzel ให้ค่าอำนาจการทดสอบสูงกว่าเมื่อกำหนดให้ค่าเฉลี่ยของตัวอย่างทั้ง 2 กลุ่มแตกต่างกันน้อยในทุกกรณี

อภิปรายผลการวิจัย

แม้ว่าสถิติทดสอบโดยวิธี Bootstrap Rank Welch ให้ค่าอำนาจการทดสอบสูงที่สุด เมื่อส่วนเบี่ยงเบนมาตรฐานของตัวอย่างแตกต่างกันน้อย และแตกต่างกันมาก แต่สถิติทดสอบโดยวิธี Bootstrap Rank Welch ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ในขณะที่สถิติทดสอบโดยวิธี Brunner and Munzel ให้ค่าอำนาจการทดสอบสูงใกล้เคียงกับ สถิติทดสอบโดยวิธี Bootstrap Rank Welch เมื่อส่วนเบี่ยงเบนมาตรฐานของตัวอย่างแตกต่างกันน้อย และสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบโดยวิธี Bootstrap Brunner and Munzel ไม่มีกรณีใดสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ และยังให้ค่าอำนาจการทดสอบต่ำสุดด้วย

ข้อเสนอแนะ

ในการทำวิจัยครั้งต่อไปสำหรับผู้วิจัยท่านอื่น ๆ มีดังนี้ ผู้สนใจอาจศึกษากรณีที่มีการแจกแจงแบบอื่นๆ ที่หลากหลายมากขึ้น โดยอาจเพิ่มกรณีศึกษาที่แตกต่างกันไป ศึกษากรณีที่ขนาดตัวอย่างมีขนาดใหญ่ และ ขนาดตัวอย่างแตกต่างกันมาก ๆ

6. เอกสารอ้างอิง

[1] Bradley, J. V., Robustness, British Journal of Mathematical and Statistical Psychology, Vol. 31, 144-152, 1978.

[2] Brunner, E., Munzel, U., The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small Sample Approximation Biometrical J.,Vol. 42, 17-25, 2000.

[3] Hollander, M., Wolfe, DA., Nonparametric Statistical Methods, 1973.

[4] Neuhauser, M., Losch, C., Jockel, K.H., The Chen-Luo Test in Case of Heteroscedasticity, Comput. Statist. Data Anal.,Vol. 51, pp. 5055-5060, 2006.

[5] Reiczigel, J., Zakarias, I., Rozsa, L., A Bootstrap Test of Stochastic Equality of Two Populations, Amer. Statist.,Vol. 59, 156-161, 2005.

[6]. Welch, B. L., The Significance of the Difference Between Two Means When the Population Variances are Unequal, Biometrika.,Vol. 29, 350-362, 1938.

ตารางผนวกที่ 1 ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อ $\alpha = 0.05$ การแจกแจงลอกนอร์มอล

$\alpha = 0.05$		สถิติทดสอบ	ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1		
n_1	n_2		$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,2)$	$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,3)$	$(ml_1, ml_2) = (1,1)$ $(sl_1, sl_2) = (1,4)$
10	10	BM	**0.053**	**0.050**	**0.060**
		BBM	0.015	0.012	0.010
		BRW	0.078	0.101	0.101
10	15	BM	**0.058**	**0.049**	**0.052**
		BBM	0.014	0.011	0.006
		BRW	0.082	0.086	0.103
15	15	BM	**0.051**	**0.052**	**0.055**
		BBM	0.005	0.006	0.001
		BRW	0.076	0.081	0.108

ตารางผนวกที่ 2 ค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ การแจกแจงลอกนอร์มอล $n = 10,10$

$\alpha = 0.05$		สถิติทดสอบ	อำนาจการทดสอบ	
n_1	n_2		$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,2)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,2)$
10	10	BM	0.735	0.997
		BBM	0.558	0.974
		BRW	**0.757**	**1**
			$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,3)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,3)$
		BM	0.434	0.916
		BBM	0.215	0.745
		BRW	**0.545**	**0.956**
			$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,4)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,4)$
		BM	0.283	0.733
		BBM	0.180	0.483
		BRW	**0.402**	**0.849**

ตารางผนวกที่ 3 ค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ การแจกแจงลอกนอร์มอล $n = 10,15$

$\alpha = 0.05$		สถิติทดสอบ	อำนาจการทดสอบ	
n_1	n_2		$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,2)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,2)$
10	15	BM	0.866	1
		BBM	0.632	0.999
		BRW	**0.902**	**1**
			$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,3)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,3)$
		BM	0.587	0.975
		BBM	0.295	0.903
		BRW	**0.677**	**0.988**
			$(ml_1, ml_2) = (1,3),\ (sl_1, sl_2) = (1,4)$	$(ml_1, ml_2) = (1,5),\ (sl_1, sl_2) = (1,4)$
		BM	0.372	0.877
		BBM	0.149	0.636
		BRW	**0.492**	**0.940**

ตารางผนวกที่ 4 ค่าอำนาจการทดสอบ เมื่อ $\alpha = 0.05$ การแจกแจงลอกนอร์มอล $n = 15,15$

$\alpha = 0.05$		สถิติทดสอบ	อำนาจการทดสอบ	
n_1	n_2		$(ml_1, ml_2) = (1,3), (sl_1, sl_2) = (1,2)$	$(ml_1, ml_2) = (1,5), (sl_1, sl_2) = (1,2)$
15	15	BM	0.891	1
		BBM	0.680	0.998
		BRW	**0.924**	**1**
			$(ml_1, ml_2) = (1,3), (sl_1, sl_2) = (1,3)$	$(ml_1, ml_2) = (1,5), (sl_1, sl_2) = (1,3)$
		BM	0.599	0.985
		BBM	0.327	0.913
		BRW	**0.692**	**0.993**
			$(ml_1, ml_2) = (1,3), (sl_1, sl_2) = (1,4)$	$(ml_1, ml_2) = (1,5), (sl_1, sl_2) = (1,4)$
		BM	0.401	0.900
		BBM	0.180	0.705
		BRW	**0.505**	**0.949**