

ประสิทธิภาพของสถิติทดสอบสำหรับการเปรียบเทียบค่าเฉลี่ยภายใต้สถานการณ์ที่
ความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์
EFFICIENCY OF STATISTICAL TESTS FOR MEAN COMPARISON IN HETEROGENEITY
OF VARIANCE CONDITIONS

พิทยา ระยัศศรี¹ และ สีวะโชติ ศรีสุทธียากร²

Pittaya Rayubsri¹ and Siwachoat Srisuttiyakorn²

¹นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹Student, Faculty of Education, Chulalongkorn University

²คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

²Faculty of Education, Chulalongkorn University

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับเปรียบเทียบค่าเฉลี่ยภายใต้สถานการณ์ที่ความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์ ได้แก่ สถิติทดสอบที่แบบดั้งเดิม สถิติทดสอบ Welch สถิติทดสอบ Browne-Forsythe สถิติทดสอบ James สถิติทดสอบ Yuen สถิติทดสอบ Alexander-Govern และสถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ย (SMM) ประกอบด้วย สถิติทดสอบไคสแควร์ที่ประมาณค่าด้วยวิธีภาวะความควรจะเป็นสูงสุด และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก และสถิติทดสอบไคสแควร์ของ Yuen และ Bentler การศึกษาใช้ข้อมูลจำลองด้วยวิธีการมอนติคาร์โล โดยกำหนดเงื่อนไขในการจำลองดังนี้ (1) ความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มกำหนดจากค่าอัตราส่วนความแปรปรวนระหว่างกลุ่มประชากรที่มีค่าเท่ากับ 1, 4, 16 และ 64 เท่า (2) ความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มประชากรทั้งสองกำหนดจากค่าขนาดอิทธิพล d ที่มีค่าเท่ากับ 0, 0.2 และ 0.8 (3) ขนาดตัวอย่างในแต่ละกลุ่มประชากรกำหนดให้มีค่าเท่ากับ (10,10) (30,30) (50,50) (100,100) (10,30) (10,50) และ (10,100) หน่วย และ (4) ระดับนัยสำคัญของการทดสอบกำหนดให้เท่ากับ .05 โดยคิดเป็นสถานการณ์จำลองทั้งสิ้น 96 สถานการณ์ กระทำซ้ำในแต่ละสถานการณ์จำนวน 500 รอบ ประสิทธิภาพของสถิติทดสอบพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่หนึ่ง และอำนาจการทดสอบ

ผลการวิจัย พบว่า (1) กรณีที่ตัวอย่างสองกลุ่มมีขนาดเท่ากัน ประสิทธิภาพของสถิติทดสอบทุกตัวมีแนวโน้มใกล้เคียงกันทั้งภายใต้สถานการณ์ที่ความแปรปรวนเป็นเอกพันธ์กันและไม่เป็นเอกพันธ์กัน อย่างไรก็ตามอำนาจการทดสอบของสถิติทดสอบ Yuen มีแนวโน้มที่จะต่ำกว่าสถิติทดสอบอื่นๆ และ (2) กรณีที่ตัวอย่างสองกลุ่มมีขนาดไม่เท่ากัน และความแปรปรวนไม่เป็นเอกพันธ์กันระหว่างกลุ่ม สถิติทดสอบที่มีแนวโน้มที่จะมีประสิทธิภาพต่ำเมื่อเปรียบเทียบกับสถิติทดสอบตัวอื่นๆ โดยสถิติทดสอบของ James, Welch, BF และ AG มีแนวโน้มที่จะมีประสิทธิภาพสูงที่สุด

คำสำคัญ: สถิติทดสอบสำหรับเปรียบเทียบค่าเฉลี่ย ความไม่เป็นเอกพันธ์กันของความแปรปรวน
อัตราความคลาดเคลื่อนประเภทที่หนึ่ง อำนาจการทดสอบ

¹Corresponding author: rayubsri.p@gmail.com

ABSTRACT

The purpose of this research is to compare the efficiency of statistical tests for comparing mean under heterogeneity of variance conditions among the traditional T-test, Welch test, Browne-Forsythe test, James test, Yuen test, Alexander-Govern test, and Structural Mean Model (SMM) based statistical tests. These also included chi-square tests, which are estimated by the Maximum Likelihood method (ML), Weighted Least Squares methods (WLS), and Yuan-Bentler chi-square. The study used simulated data generated by Monte Carlo methods under the following conditions: 1) ratio of variance between two groups were equal to 1, 4, 16 and 64; 2) the mean parameters of each group, set by the magnitude of effect size d , which were equal to 0, 0.2 and 0.8; 3) the sample sizes of each group were (10,10), (30,30), (50,50), (100,100), (10,30), (10,50) and (10,100) units; and 4) the significance level was chosen to be .05. The total number of simulations was 96, with 500 repetitions for each situation. The efficiency of the statistical test was based on the rate of type I errors and the statistical power of the test.

The results revealed that 1) in cases of equal sample size and equal variance, all statistical tests tended to have the same efficiency under both homogeneity and heterogeneity of variances. However, Yuen tended to have lower statistical power than other tests and 2) in cases of unequal sample sizes and unequal variances, T statistics are likely to display lower efficiency compared to James's test, Welch's test, BF-test, and AG-test, which yield the highest efficiency values.

KEYWORDS: Mean Comparison Test Statistics, Heterogeneity of Variances, Type I Error Rate, Power of the Test

บทนำ

สถิติทดสอบที (t-test) เป็นสถิติทดสอบแบบดั้งเดิมที่นักวิจัยเชิงปริมาณทางสังคมศาสตร์ นิยมใช้เป็นสถิติทดสอบมาตรฐานสำหรับการทดสอบเพื่อหาข้อสรุปเกี่ยวกับความแตกต่างของค่าเฉลี่ยระหว่างตัวอย่างสองกลุ่ม โดยใช้สถิติทดสอบที่ดังกล่าวมีข้อกำหนดเบื้องต้น (Assumptions) ที่สำคัญสามประการ ได้แก่ (1) ตัวอย่างทั้งสองกลุ่มต้องมีการสุ่มมาจากรประชากรที่มีการแจกแจงแบบปกติ (normal distribution) (2) ความแปรปรวนของประชากรทั้งสองกลุ่มต้องมีความเป็นเอกพันธ์กัน (Homogeneity of variances) และ (3) ข้อมูลตัวอย่างต้องมีความเป็นอิสระซึ่งกันและกัน (Independence) อย่างไรก็ตามภายใต้สถานการณ์จริงของการวิเคราะห์ข้อมูลในการวิจัย พบว่ามีความเป็นไปได้ไม่น้อยที่ข้อมูลตัวอย่างที่เก็บรวบรวมมานั้นจะเป็นไปตามข้อกำหนดเบื้องต้นดังกล่าว การละเมิดข้อตกลงเบื้องต้นย่อมส่งผลกระทบต่อความถูกต้องของผลสรุปที่ได้จากการทดสอบซึ่งอาจเป็นผลให้นักวิจัยได้ข้อสรุปของการวิจัยที่คลาดเคลื่อนไปจากสภาพจริง (Field, et al., 2012)

จากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องพบว่า ในเชิงปฏิบัติการละเมิดข้อกำหนดเบื้องต้นด้านความเป็นอิสระซึ่งกันและกันของข้อมูลสามารถป้องกันได้ตั้งแต่ขั้นตอนการออกแบบการเลือกตัวอย่าง โดยใช้กระบวนการสุ่ม (Randomization) แต่การป้องกันหรือหลีกเลี่ยงการละเมิดข้อกำหนดเบื้องต้นด้านการแจกแจงปกติ หรือความไม่เป็นเอกพันธ์กันของความแปรปรวนนั้นเป็นไปได้ยาก (Vexler, et al., 2016) จากการศึกษาที่ผ่านมาด้วยวิธีการจำลองแบบมอนติคาร์โลพบว่า (1) การละเมิดข้อกำหนดเบื้องต้นด้านการแจกแจงแบบปกติของข้อมูลจะส่งผลให้การแจกแจงความน่าจะเป็นของสถิติทดสอบที่ไม่สอดคล้องกับการแจกแจงแบบที (t distribution) ซึ่งทำให้เกิดความคลาดเคลื่อนในการสรุปผลการทดสอบ อย่างไรก็ตามการแจกแจงความน่าจะเป็นของสถิติทดสอบที่มีแนวโน้มจะเข้าสู่การแจกแจงแบบทีเมื่อขนาดตัวอย่างเพิ่มขึ้น จึงทำให้สถิติทดสอบที่เป็นสถิติทดสอบที่มีความแข็งแกร่งต่อสถานการณ์ที่มีการละเมิดข้อกำหนดเบื้องต้นด้านการแจกแจงแบบปกติของข้อมูล กล่าวคือยังเป็นสถิติทดสอบที่สามารถควบคุมระดับนัยสำคัญได้ตามที่กำหนด และมีอำนาจการทดสอบสูง อย่างไรก็ตามประสิทธิภาพดังกล่าวมีแนวโน้มลดลงหากข้อมูลมีความเบี่ยงเบนไปมากจากการแจกแจงแบบปกติ โดยเฉพาะในกรณีที่มีการแจกแจงของข้อมูลมีลักษณะไม่สมมาตรอย่างรุนแรง หรือมีค่าผิดปกติ (Outlier) (Field, et al., 2012; Algina, et al., 2005) และ (2) ในกรณีที่ความแปรปรวนไม่เป็นเอกพันธ์กันระหว่างกลุ่มตัวอย่างพบว่า ประสิทธิภาพของสถิติทดสอบที่มีความแตกต่างกันออกไป โดยขึ้นกับขนาดความแตกต่างของความแปรปรวน และขนาดตัวอย่างระหว่างกลุ่มตัวอย่าง โดยสถิติทดสอบที่ยังสามารถควบคุมระดับนัยสำคัญ และยังมีอำนาจการทดสอบอยู่ในระดับที่สูงอยู่ในกรณีที่ขนาดของความแปรปรวนในแต่ละกลุ่มตัวอย่างแปรผันตรงกับขนาดตัวอย่างในแต่ละกลุ่ม แต่ในกรณีที่ขนาดของความแปรปรวนในแต่ละกลุ่มตัวอย่างแปรผกผันกับขนาดตัวอย่างในแต่ละกลุ่ม พบว่าประสิทธิภาพของสถิติทดสอบที่มีแนวโน้มลดลงซึ่งเป็นการเพิ่มความเสี่ยงในการได้ข้อสรุปของการวิจัยที่คลาดเคลื่อน (Overall, et al., 1995)

การแก้ปัญหาเมื่อมีการละเมิดข้อกำหนดเบื้องต้นดังกล่าวสามารถกระทำได้หลายวิธีการ เช่น วิธีการแรกคือ วิธีการแปลงค่า (Transformation methods) เพื่อแปลงค่าของข้อมูลผ่านฟังก์ชันการแปลงที่เหมาะสมซึ่งจะช่วยให้การแจกแจงของข้อมูลมีแนวโน้มที่ใกล้เคียงกับการแจกแจงแบบปกติ รวมถึงในบางกรณียังช่วยให้มีความแปรปรวนระหว่างกลุ่มมีความใกล้เคียงกันมากขึ้น (Box & Cox, 1964; Hair et al., 2010) อย่างไรก็ตาม วิธีการนี้มีข้อจำกัดที่สำคัญคือทำให้ความหมายของตัวแปรที่ศึกษาเปลี่ยนแปลงไปและยากที่จะแปลผลให้มีความหมาย วิธีการที่สอง คือใช้สถิติทดสอบแบบไม่ใช้พารามิเตอร์ (Nonparametric test) เช่น Wilcoxon rank-sum test หรือ Mann-Whitney U-tests ที่ข้อกำหนดเบื้องต้นของสถิติทดสอบไม่ขึ้นอยู่กับการแจกแจงของข้อมูล อย่างไรก็ตาม การใช้สถิติแบบไม่ใช้พารามิเตอร์ดังกล่าวมีข้อจำกัดหลายประการ ประการแรก คือข้อจำกัดด้านสารสนเทศที่ได้รับและการสรุปผลจากสถิติไม่ใช้พารามิเตอร์ ที่ทำให้การอธิบายคุณลักษณะหรือการเปรียบเทียบความแตกต่างของประชากรที่สนใจกระทำได้ในขอบเขตจำกัด ประการที่สอง อำนาจของสถิติทดสอบแบบไม่ใช้พารามิเตอร์มักมีค่าต่ำเมื่อเปรียบเทียบกับสถิติที่ใช้พารามิเตอร์โดยเฉพาะในกรณีที่ตัวอย่างมีขนาดเล็ก ดังนั้นสถิติแบบไม่ใช้พารามิเตอร์จึงไม่ใช่วิธีการที่มีประสิทธิภาพสูงสุด (Whitley & Ball, 2002)

จากข้อจำกัดของวิธีการทั้งสองจึงมีการพัฒนาวิธีการที่สามคือ สถิติทดสอบที่มีความแข็งแกร่ง (Robust test statistics) วิธีการในกลุ่มนี้จัดอยู่ในกลุ่มของสถิติแบบไม่ใช้พารามิเตอร์ที่ถูกพัฒนาขึ้นเพื่อใช้เป็นสถิติทางเลือกในกรณีที่ข้อมูลไม่เป็นไปตามข้อกำหนดเบื้องต้นของการใช้สถิติทดสอบที่แบบดั้งเดิม สถิติทดสอบที่มีความแข็งแกร่งดังกล่าวถูกพัฒนาให้ได้รับผลกระทบจากการที่ข้อมูลไม่เป็นไปตามข้อกำหนด

เบื้องต้นน้อยลง ทำให้มีประสิทธิภาพทั้งด้านการควบคุมระดับนัยสำคัญ และอำนาจการทดสอบที่สูงกว่า สถิติทดสอบที่แบบดั้งเดิม รวมทั้งสถิติทดสอบแบบไม่ใช้พารามิเตอร์ (Wilcox & Schonbrodt, 2016; Wong, et al., 2014) ในปัจจุบันมีการพัฒนาสถิติที่มีความแข็งแกร่งสำหรับทดสอบความแตกต่างระหว่างค่าเฉลี่ยของสองกลุ่มตัวอย่างไวก์หลายตัว โดยอาจแบ่งได้เป็นสองกลุ่มคือ (1) สถิติทดสอบที่พัฒนาให้มีความแข็งแกร่งต่อความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม ได้แก่ สถิติทดสอบของ Welch สถิติทดสอบของ Brown-Forsyth สถิติทดสอบของ Yuen สถิติทดสอบของ James และ Alexander-Govern สถิติทดสอบในกลุ่มนี้เป็นสถิติทดสอบที่มีประสิทธิภาพสูงกว่าสถิติทดสอบที่ในกรณีที่ความแปรปรวนระหว่างกลุ่มไม่มีความเป็นเอกพันธ์กัน แต่เมื่อข้อมูลมีการแจกแจงแบบไม่ปกติ สถิติทดสอบในกลุ่มนี้มีแนวโน้มที่จะควบคุมระดับนัยสำคัญได้ยากขึ้น และ (2) สถิติทดสอบที่พัฒนาให้มีความแข็งแกร่งต่อการแจกแจงแบบไม่ปกติและความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม ได้แก่ สถิติทดสอบของ Yuen-Welch เป็นต้น (Alexander & Govern, 1994; Brown & Forsythe, 1974; James, 1951; Welch, 1938, 1951; Yuen, 1974)

วิธีการที่สี่คือ การใช้โมเดลโครงสร้างค่าเฉลี่ย (Structural mean model: SMM) วิธีการนี้พัฒนาจากโมเดลสมการโครงสร้าง (Structural equation model: SEM) โดยลดส่วนประกอบเกี่ยวกับตัวแปรแฝงออกจากโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน ดังนี้ $x = v_k + \delta$ เมื่อ $\delta \sim \varepsilon(0, \sigma_k^2)$ คือค่าคลาดเคลื่อนสุ่มของโมเดล σ_k^2 คือความแปรปรวนของกลุ่มประชากรที่ k และ $k = 1, 2$ คือจำนวนกลุ่มประชากรที่ต้องเปรียบเทียบ จากโมเดลข้างต้นจะได้ว่า $H_0: v_1 = v_2$ คือสมมติฐานการทดสอบเพื่อเปรียบเทียบค่าเฉลี่ยสองกลุ่ม การประมาณค่าสถิติทดสอบสามารถทำได้หลายวิธีการ เช่น หากข้อมูลมีการแจกแจงแบบปกติหรือใกล้เคียงสามารถใช้วิธีประมาณแบบภาวะความควรจะเป็นสูงสุด (maximum likelihood: ML) ที่มีฟังก์ชันความแตกต่างพหุกลุ่ม (Multiple group discrepancy function) ดังนี้

$$F_{ML} = \sum_{k=1}^2 \left(\frac{n_k}{N} \right) F_k$$

โดยที่ $F_k = [\ln|\Sigma_k| + \text{tr}(S_k \Sigma_k^{-1}) - \ln|S_k| - p] + (m_k - \mu_k)^T \Sigma_k^{-1} (m_k - \mu_k)$ คือฟังก์ชันความแตกต่างของกลุ่มประชากรที่ k ฟังก์ชันความแตกต่างในข้างต้นนำมาคำนวณเป็นสถิติทดสอบ T_{ML} สำหรับสมมติฐานในข้างต้นได้ดังนี้

$$T_{ML} = (N - 1) F_{ML} \sim \chi_{df=K+1}^2$$

โดยที่ K คือ จำนวนพารามิเตอร์ (ความแปรปรวนและค่าเฉลี่ยประชากร) ในโมเดล แต่หากข้อมูลมีการแจกแจงแบบไม่ปกติ ผู้วิจัยสามารถเปลี่ยนวิธีประมาณค่าพารามิเตอร์ที่ไม่อิงการแจกแจงของข้อมูล เช่น วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (Weighted least squares: WLS) ที่มีฟังก์ชันความแตกต่างเป็น $F_{WLS} = \sum_{k=1}^2 (s_k - \sigma_k)^T W_k^{-1} (s_k - \sigma_k)$ โดยสถิติทดสอบสามารถคำนวณได้ในทำนองเดียวกับกรณีที่ประมาณค่าภาวะความควรจะเป็นสูงสุดดังนี้ $T_{WLS} = (N - 1) F_{WLS}$ อย่างไรก็ตามการประมาณค่าด้วยวิธี WLS มีข้อจำกัดในกรณีที่ตัวอย่างมีขนาดเล็ก Yuan and Bentler (1997, 1999) จึงพัฒนาสถิติทดสอบ $T_{YB} = T_{WLS} / (1 + T_{WLS} / N)$ เพื่อให้มีความแข็งแกร่งต่อกรณีขนาดตัวอย่างเล็ก

จากประเด็นสำคัญในข้างต้น ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพของสถิติทดสอบระหว่างสถิติทดสอบที่แบบดั้งเดิม สถิติทดสอบที่มีความแกร่ง (RSM) และการใช้โมเดลโครงสร้างค่าเฉลี่ย (SMM) ภายใต้สถานการณ์ที่ความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์กัน

วัตถุประสงค์การวิจัย

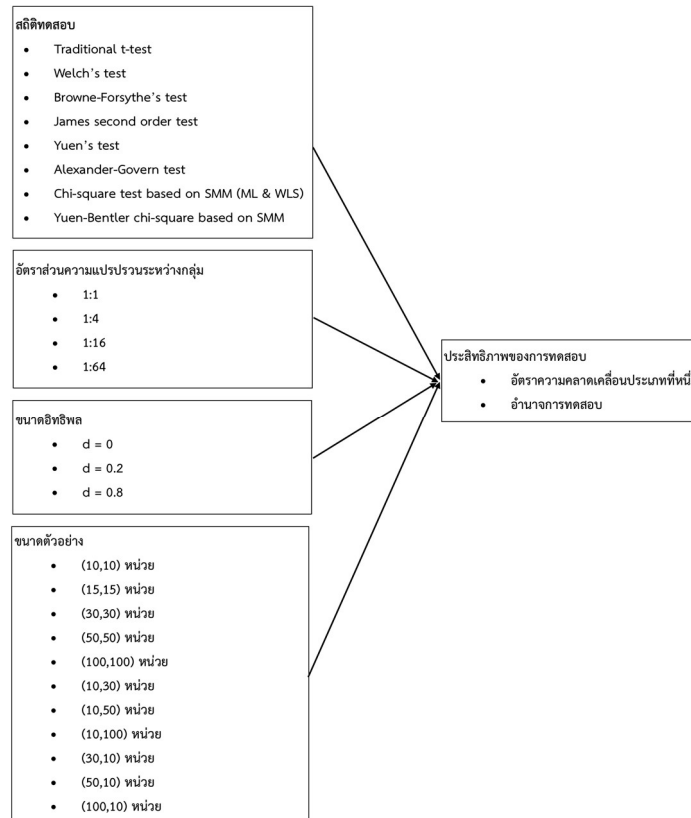
เพื่อเปรียบเทียบประสิทธิภาพของสถิติทดสอบ ได้แก่ สถิติทดสอบที่แบบดั้งเดิม สถิติทดสอบ Welch สถิติทดสอบ Browne-Forsythe สถิติทดสอบ James สถิติทดสอบ Yuen สถิติทดสอบ Alexander-Govern และสถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ย ประกอบด้วย สถิติทดสอบไคสแควร์ที่ประมาณค่าด้วยวิธีภาวะความควรจะเป็นสูงสุด และวิธีกำลังสองน้อยสุด และสถิติทดสอบไคสแควร์ของ Yuen and Bentler (1997, 1999) ภายใต้สถานการณ์ที่ความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์กัน ขนาดอิทธิพล และขนาดตัวอย่างมีความแตกต่างกัน

นิยามศัพท์เฉพาะ

ประสิทธิภาพของการทดสอบเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง คือ อัตราความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error rate) และอำนาจการทดสอบ (Power of the test) ของสถิติทดสอบสำหรับเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่างสองกลุ่ม

กรอบแนวคิดการวิจัย

จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง (1) ตัวแปรตามของการวิจัย ได้แก่ ประสิทธิภาพของการทดสอบเพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยสองกลุ่มตัวอย่าง ที่วัดจากตัวบ่งชี้สองตัว ได้แก่ อัตราความคลาดเคลื่อนประเภทที่หนึ่ง ซึ่งใช้บ่งชี้ประสิทธิภาพในด้านการควบคุมระดับนัยสำคัญของการทดสอบ และอำนาจการทดสอบ ซึ่งใช้บ่งชี้ประสิทธิภาพในด้านความไวของการตรวจจับความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม (2) ส่วนตัวแปรอิสระในการศึกษาประกอบด้วย สถิติทดสอบ จำนวน 8 ตัว อัตราส่วนความแปรปรวนระหว่างกลุ่ม ที่กำหนดให้มีจำนวน 4 ระดับ ขนาดอิทธิพล ที่มี 3 ระดับ และขนาดตัวอย่างที่มี 11 ระดับ ดังรายละเอียดในกรอบแนวคิดการวิจัย รูปที่ 1



รูปที่ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

งานวิจัยนี้ใช้ระเบียบวิธีวิจัยเชิงทดลองด้วยการจำลองสถานการณ์แบบมอนติคาร์โล (Monte Carlo simulation) แผนการวิจัยประกอบด้วยสี่ส่วน ได้แก่ สถานการณ์การจำลองที่ใช้ในการวิจัย การจำลองข้อมูลในการวิจัย เกณฑ์การพิจารณาประสิทธิภาพของสถิติทดสอบ และการวิเคราะห์ข้อมูล รายละเอียดเป็นดังนี้

1. สถานการณ์จำลองที่ใช้ในการวิจัย การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่างสองกลุ่ม ภายใต้สถานการณ์ที่ความแปรปรวนประชากรทั้งสองกลุ่มไม่เป็นเอกพันธ์กัน

1) ข้อมูลค่าสังเกต x_{ij} ของหน่วยตัวอย่างที่ i ในกลุ่มตัวอย่างที่ j จำลองจากประชากรมีการแจกแจงแบบปกติ (Normal distribution) $x_{ij} \sim N(\mu_j, \sigma_j^2)$ เมื่อ $i = 1, 2, \dots, n$ และ $j = 1, 2$

2) พารามิเตอร์ความแปรปรวนของการแจกแจงใน (1) กำหนดให้ $\sigma_1^2 = 1$ และ σ_2^2 กำหนดจากอัตราส่วนความแปรปรวนระหว่างกลุ่มตามกรอบแนวคิดของการวิจัยที่มีค่าเท่ากับ 1:1, 1:4, 1:16 และ 1:64

3) พารามิเตอร์ค่าเฉลี่ยของการแจกแจงใน (1) กำหนดให้ $\mu_1 = 0$ และ μ_2 กำหนดจากขนาดอิทธิพล $d = |\mu_1 - \mu_2|/\sigma_p$ โดยที่ σ_p คือส่วนเบี่ยงเบนมาตรฐานรวม (Pooled SD) และในการศึกษากำหนดให้มีค่าเท่ากับ 0.0, 0.2 และ 0.8

4) ขนาดตัวอย่างในการศึกษามีสองกรณี ได้แก่ กรณีที่ทั้งสองกลุ่มมีขนาดตัวอย่างเท่ากันคือ (10,10) (15,15) (30,30) (50,50) และ (100,100) หน่วย และกรณีที่ทั้งสองกลุ่มมีขนาดตัวอย่างไม่เท่ากันคือ (10,30) (10,50) และ (10,100) หน่วย

2. การจำลองข้อมูลในการวิจัย ข้อมูลจำลองตามสถานการณ์จำลองข้างต้นจำลองโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo simulation) ดำเนินการภายใต้โปรแกรม R โดยใช้อัลกอริทึม Mersenne-Twister (Matsumoto & Nishimura, 1998) ในการสร้างเลขสุ่มเทียม และใช้วิธีการแปลงผกผัน (Inverse method) ในการสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (Johnson, et al., 1996) โดยในแต่ละสถานการณ์จำลองผู้วิจัยกำหนดให้กระทำซ้ำเป็นจำนวน 500 รอบ

3. เกณฑ์การพิจารณาประสิทธิภาพของสถิติทดสอบ ได้แก่ (1) อัตราความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error rate) คำนวณจากจำนวนครั้งที่สถิติทดสอบแต่ละตัวปฏิเสธสมมติฐาน $H_0: \mu_1 = \mu_2$ ภายใต้สถานการณ์ที่ H_0 เป็นจริง และ (2) อำนาจการทดสอบสมมติฐาน (Power of the test) คำนวณจากจำนวนครั้งที่สถิติทดสอบแต่ละตัวปฏิเสธสมมติฐาน $H_0: \mu_1 = \mu_2$ ภายใต้สถานการณ์ที่ H_0 เป็นเท็จ

4. การวิเคราะห์ข้อมูล ประกอบด้วย (1) การวิเคราะห์ข้อมูลด้วยสถิติเชิงบรรยาย โดยนำเสนอการแจกแจงของค่าอัตราความคลาดเคลื่อนประเภทที่หนึ่ง และอำนาจการทดสอบ ของแต่ละสถิติทดสอบแต่ละตัว ภายใต้สถานการณ์จำลองในรูปของแผนภาพกล่อง และ (2) การวิเคราะห์ความไว (Sensitivity) ของสถิติทดสอบต่อปัจจัยระดับความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม (Var) ขนาดอิทธิพล (d) ขนาดตัวอย่าง (n) และปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว ที่มีต่ออัตราความคลาดเคลื่อนประเภทที่หนึ่ง และอำนาจการทดสอบของสถิติทดสอบทั้ง 9 ตัว ค่าความไวของสถิติทดสอบดังกล่าวคำนวณจากค่าขนาดอิทธิพล partial eta squares (η^2) ที่ได้จากการวิเคราะห์ความแปรปรวนตามโมเดลดังสมการ $criterion = \mu + Var + d + n + Var \times d + Var \times n + d \times n + Var \times d \times n + errors$ เพื่อคำนวณค่าผลรวมกำลังสองบางส่วน (Partial sum squares: SS) ตามแหล่งที่มาของความผันแปรในค่าอัตราความคลาดเคลื่อนประเภทที่หนึ่ง และอำนาจการทดสอบ ดังนี้

$$\eta^2_{source} = \frac{SS_{source}}{SS_{source} + SS_{errors}}$$

โดยที่ SS_{source} คือ ผลรวมกำลังสองของปัจจัยภายใต้สถานการณ์จำลองที่ประกอบด้วย ระดับความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม ขนาดอิทธิพล ขนาดตัวอย่าง และปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว และ SS_{errors} คือผลรวมกำลังสองของความคลาดเคลื่อน ที่คำนวณโดยใช้เทคนิคการวิเคราะห์ความแปรปรวนสามทางแบบมีปฏิสัมพันธ์

สรุปและอภิปรายผลการวิจัย

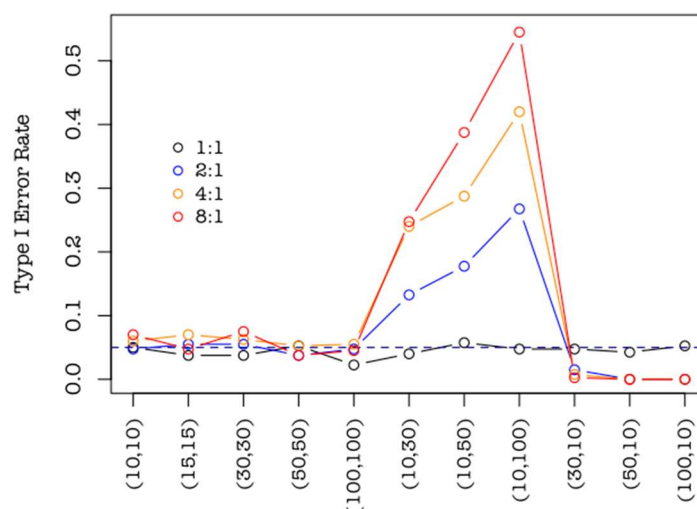
สรุปผลการวิจัย

ผลการวิจัยจำแนกออกได้เป็นสองส่วน ได้แก่ (1) ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่หนึ่งของสถิติทดสอบแต่ละตัว ประกอบด้วย การนำเสนอค่าอัตราความคลาดเคลื่อนประเภทที่หนึ่งจำแนกตามสถานการณ์จำลอง และการวิเคราะห์ความไวของสถิติทดสอบด้านอัตราความคลาดเคลื่อนประเภทที่หนึ่งต่อปัจจัยความไม่เป็นเอกพันธ์กันของความแปรปรวน ขนาดตัวอย่าง และปฏิสัมพันธ์ระหว่างปัจจัยทั้งสอง และ (2) ผลการวิเคราะห์อำนาจการทดสอบของสถิติทดสอบแต่ละตัว ประกอบด้วย การนำเสนอค่าอำนาจการทดสอบจำแนกตามสถานการณ์จำลอง และการวิเคราะห์ความไวของสถิติทดสอบด้านอำนาจการทดสอบต่อปัจจัยความไม่เป็นเอกพันธ์ ขนาดอิทธิพล ขนาดตัวอย่าง และปฏิสัมพันธ์ระหว่างปัจจัย รายละเอียดเป็นดังนี้

1. อัตราความคลาดเคลื่อนประเภทที่หนึ่ง

จากการวิเคราะห์ความแปรปรวนสองทางพบว่า สถิติทดสอบที (t-test) เป็นสถิติทดสอบเพียงตัวเดียวที่ได้รับอิทธิพลจากปฏิสัมพันธ์ระหว่างปัจจัยความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มกับขนาดตัวอย่างอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = .000$) รูปที่ 2 แสดงค่าอัตราความคลาดเคลื่อนประเภทที่หนึ่งของสถิติทดสอบที จำแนกตามระดับความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มและขนาดตัวอย่าง จากรูปจะเห็นว่าสถิติทดสอบทียังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่กำหนด ($=.05$) ได้สองกรณี ได้แก่ (1) กรณีที่ตัวอย่างสองกลุ่มมีขนาดเท่ากัน ($n_1 = n_2$) พบว่ามีอัตราความคลาดเคลื่อนประเภทที่หนึ่งเฉลี่ยไม่แตกต่างจาก .05 ที่ระดับนัยสำคัญเท่ากับ .05 ($p\text{-value} = .747$) และ (2) กรณีที่ตัวอย่างสองกลุ่มมีขนาดตัวอย่างไม่เท่ากันและแปรผันตรงกับระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน ($n_1 > n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่าอัตราความคลาดเคลื่อนประเภทที่หนึ่งเฉลี่ยต่ำกว่า .05 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($p\text{-value} = .000$) แต่พบว่าไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่กำหนดได้ในกรณีที่ตัวอย่างสองกลุ่มมีขนาดตัวอย่างไม่เท่ากันและแปรผกผันกับระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน ($n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$) และพบว่าอัตราความคลาดเคลื่อนประเภทที่หนึ่งของสถิติทดสอบทีมีแนวโน้มสูงขึ้นเมื่อขนาดตัวอย่างสองกลุ่มมีความแตกต่างกันมากขึ้นและระดับความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มมีค่ามากขึ้น

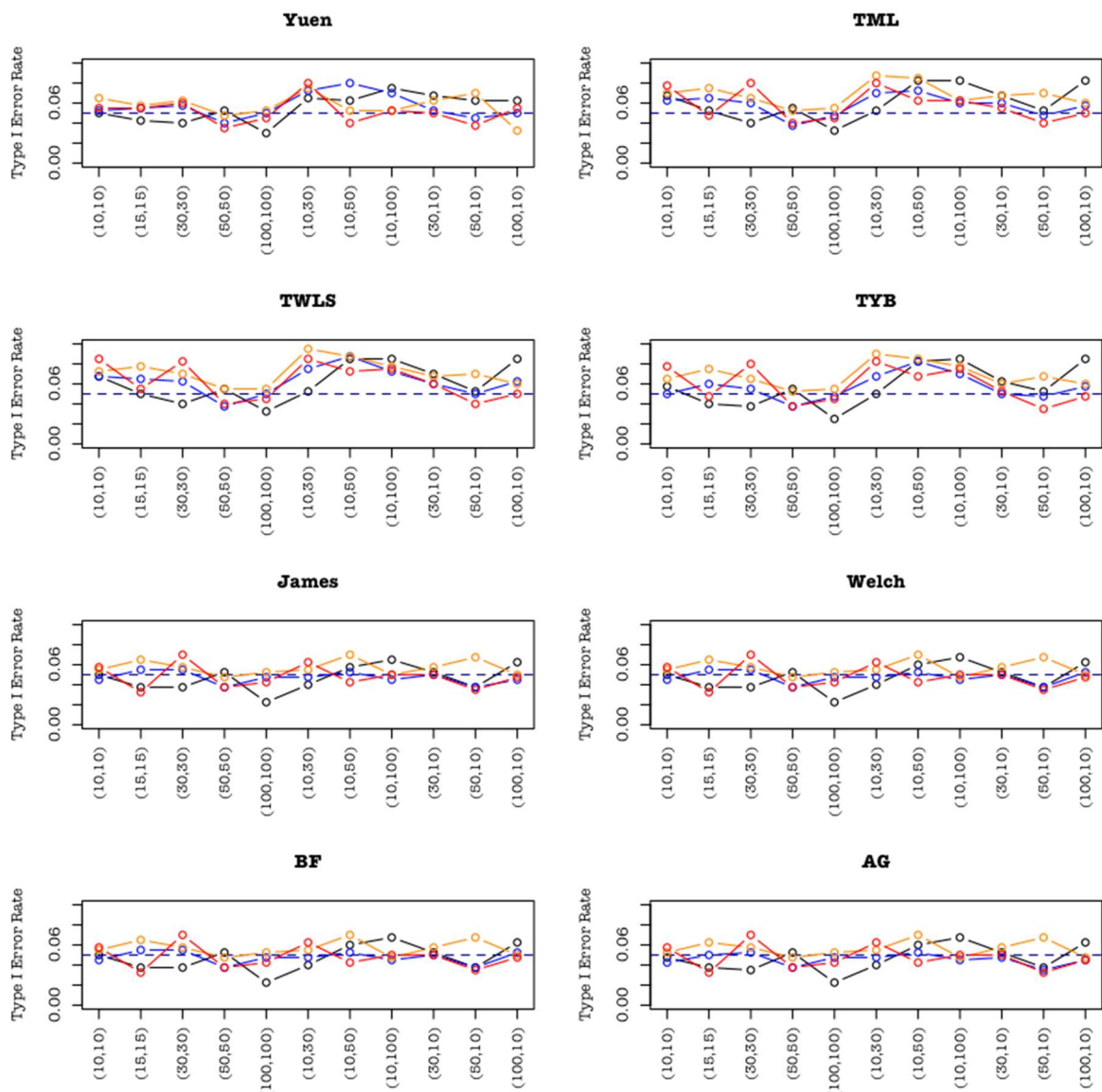
ผลการวิเคราะห์ข้างต้นแสดงให้เห็นว่าในกรณีที่ตัวอย่างสองกลุ่มมีขนาดเท่ากันหรือไม่เท่ากัน แต่ขนาดตัวอย่างในแต่ละกลุ่มแปรผันตรงกับระดับของความไม่เป็นเอกพันธ์กันของความแปรปรวน สถิติทดสอบทียังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งได้อย่างมีประสิทธิภาพ แต่ในกรณีที่ขนาดตัวอย่างในแต่ละกลุ่มแปรผกผันกับระดับของความไม่เป็นเอกพันธ์กันของความแปรปรวน สถิติทดสอบทีมีแนวโน้มไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่กำหนดได้ และปัญหามีแนวโน้มรุนแรงมากยิ่งขึ้นเมื่อระดับความไม่เป็นเอกพันธ์กันของความแปรปรวนมีเพิ่มมากขึ้น



รูปที่ 2 อัตราความคลาดเคลื่อนประเภทที่หนึ่งของสถิติทดสอบทีภายใต้สถานการณ์ที่ระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน และขนาดตัวอย่าง แตกต่างกัน

ผลการวิเคราะห์ความแปรปรวนพบว่าสถิติทดสอบของ Welch, BF, Yuen, James, AG, T_{ML} , T_{WLS} และ T_{YB} เป็นสถิติทดสอบที่ไม่ได้รับอิทธิพลจากปฏิสัมพันธ์ระหว่างความไม่เป็นเอกพันธ์กันของความแปรปรวนกับขนาดตัวอย่าง ที่ระดับนัยสำคัญทางสถิติ .05 ($p\text{-value} > .141$) และเมื่อพิจารณาอิทธิพลหลักของปัจจัยความไม่เป็นเอกพันธ์กันของความแปรปรวนกับขนาดตัวอย่างที่มีต่อสถิติทดสอบดังกล่าวได้ข้อค้นพบโดยสรุปดังนี้ (1) สถิติที่ได้รับอิทธิพลจากความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติมีจำนวน 4 ตัว เรียงลำดับตามขนาดอิทธิพล ได้แก่ สถิติทดสอบของ James ($p\text{-value}=.021$, $\eta^2 = .196$), AG ($p\text{-value} = .030$, $\eta^2 = .182$), BF ($p\text{-value}=.036$, $\eta^2 = .174$) และ Welch ($p\text{-value} = .036$, $\eta^2 = .174$) (2) สถิติทดสอบที่ไม่ได้รับอิทธิพลจากความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่มที่ระดับนัยสำคัญเท่ากับ .05 มีจำนวน 4 ตัว เรียงลำดับตามขนาดอิทธิพล ได้แก่ สถิติทดสอบ T_{YB} ($p\text{-value} = .061$, $\eta^2 = .153$), T_{WLS} ($p\text{-value} = .121$, $\eta^2 = .122$), T_{ML} ($p\text{-value} = .129$, $\eta^2 = .120$) และ Yuen ($p\text{-value} = .660$, $\eta^2 = .035$) (3) สถิติทดสอบที่ได้รับอิทธิพลจากขนาดตัวอย่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีจำนวน 3 ตัว เรียงลำดับตามขนาดอิทธิพล ได้แก่ T_{WLS} ($p\text{-value} = .000$, $\eta^2 = .542$), T_{YB} ($p\text{-value} = .000$, $\eta^2 = .518$) และ T_{ML} ($p\text{-value} = .003$, $\eta^2 = .424$) และ Yuen ($p\text{-value} = .042$, $\eta^2 = .294$) และ (4) สถิติทดสอบที่ไม่ได้รับอิทธิพลจากขนาดตัวอย่างที่ระดับนัยสำคัญทางสถิติ .05 มีจำนวน 5 ตัว เรียงลำดับตามขนาดอิทธิพล ได้แก่ สถิติทดสอบของ BF ($p\text{-value}=.256$, $\eta^2 = .229$), Welch ($p\text{-value} = .256$, $\eta^2 = .229$), AG ($p\text{-value} = .308$, $\eta^2 = .217$) และ James ($p\text{-value}=.297$, $\eta^2 = .219$)

จากการใช้ผลการวิเคราะห์ข้างต้นประกอบกับผลการวิเคราะห์ในรูปที่ 3 สรุปข้อค้นพบได้ดังนี้ (1) สถิติทดสอบ Yuen, T_{ML} , T_{WLS} และ T_{YB} เป็นสถิติที่ได้รับอิทธิพลจากขนาดตัวอย่างอยู่ในระดับที่ค่อนข้างสูงกล่าวคือเป็นสถิติที่ต้องการขนาดตัวอย่างมาก และมีความไวต่อสถานการณ์ที่ขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากันและมีขนาดแปรผันตามกับความแปรปรวนในแต่ละกลุ่ม (2) สถิติทดสอบ James, Welch, BF และ AG เป็นสถิติที่ได้รับอิทธิพลจากความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม และขนาดตัวอย่างอยู่ในระดับต่ำ และมีแนวโน้มควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ในระดับที่กำหนด ($=.05$) ได้ในทุกสถานการณ์ที่ศึกษา

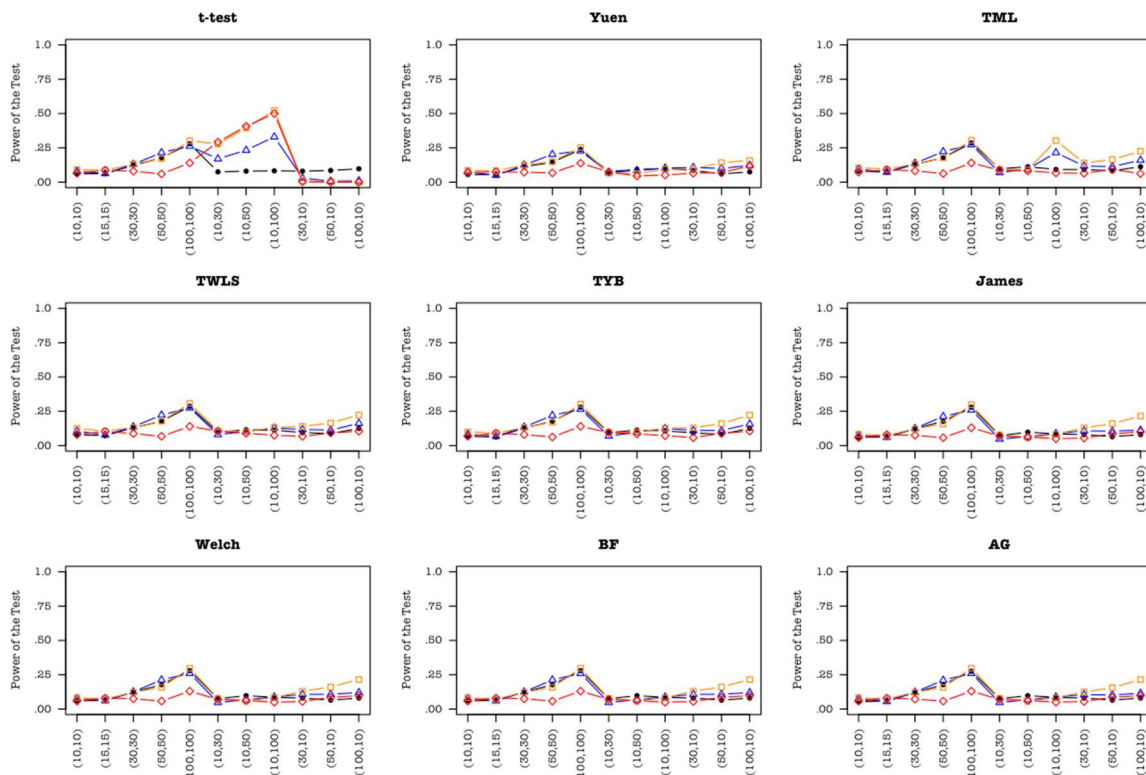


รูปที่ 3 อัตราความคลาดเคลื่อนประเภทที่หนึ่งของสถิติทดสอบ Yuen, TML, TWLS, TYB, James, Welch, BF และ AG ภายใต้สถานการณ์ที่ระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน และขนาดตัวอย่าง แตกต่างกัน

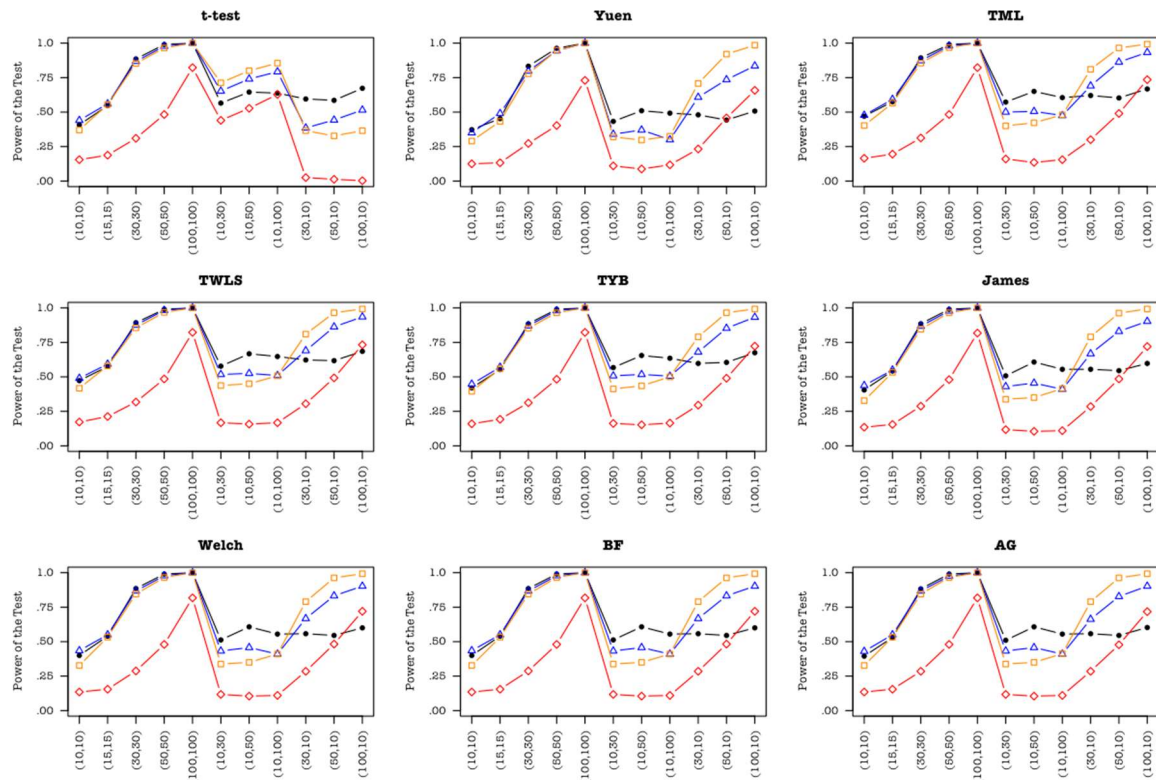
2. อำนาจการทดสอบ

จากผลการวิเคราะห์ความแปรปรวนสองทางพบว่าอำนาจการทดสอบของสถิติทดสอบทั้ง 9 ตัวได้รับอิทธิพลจากปฏิสัมพันธ์ระหว่างปัจจัยความไม่เป็นเอกพันธ์กันของความแปรปรวนระหว่างกลุ่ม ขนาดอิทธิพล และขนาดตัวอย่าง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($p\text{-value} = .000$) และเมื่อพิจารณาผลการวิเคราะห์อำนาจการทดสอบของสถิติแต่ละตัวในแต่ละสถานการณ์จากรูปที่ 4 และ 5 ได้ข้อค้นพบที่สามารถสรุปได้ดังนี้ (1) อำนาจการทดสอบของสถิติทดสอบแต่ละตัวมีแนวโน้มต่ำที่สุดในกรณีที่ความแปรปรวนระหว่างกลุ่มมีความไม่เป็นเอกพันธ์กันในระดับสูง ($\sigma_1^2 : \sigma_2^2 = 8:1$) (2) กรณีที่ตัวอย่างมีขนาดเท่ากัน ($n_1 = n_2$) อำนาจการทดสอบของสถิติทดสอบแต่ละตัวมีแนวโน้มแปรผันโดยตรงกับขนาดตัวอย่าง (3) กรณีที่ตัวอย่างมีขนาดไม่เท่ากันและมีขนาดตัวอย่างในแต่ละกลุ่มที่แปรผันโดยตรงกับระดับ

ความไม่เป็นเอกพันธ์กันของความแปรปรวน ($n_1 > n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่า อำนาจการทดสอบของสถิติทดสอบที่มีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างใหญ่ขึ้น แต่ไม่พบแนวโน้มการเปลี่ยนแปลงของอำนาจการทดสอบในสถิติทดสอบตัวอื่นๆ (4) กรณีที่ตัวอย่างมีขนาดไม่เท่ากันและมีขนาดตัวอย่างในแต่ละกลุ่มที่แปรผกผันกับระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน ($n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่า อำนาจการทดสอบของสถิติทดสอบที่ไม่มีแนวโน้มเปลี่ยนแปลงไปตามขนาดตัวอย่างที่เพิ่มขึ้น แต่พบว่าเมื่อระดับความไม่เป็นเอกพันธ์ของความแปรปรวนเพิ่มขึ้นอำนาจการทดสอบของสถิติทดสอบที่มีแนวโน้มลดลง ส่วนอำนาจการทดสอบของสถิติทดสอบตัวอื่นๆ พบว่ามีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้นสำหรับกรณีที่ความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์กัน และ (5) อำนาจการทดสอบของสถิติทดสอบแต่ละตัวมีแนวโน้มเพิ่มขึ้นเมื่อขนาดอิทธิพลเพิ่มขึ้น

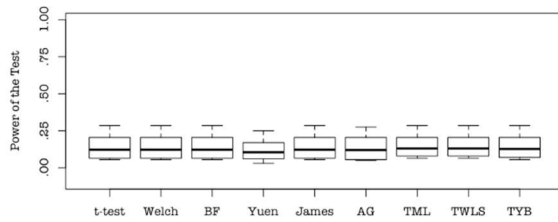
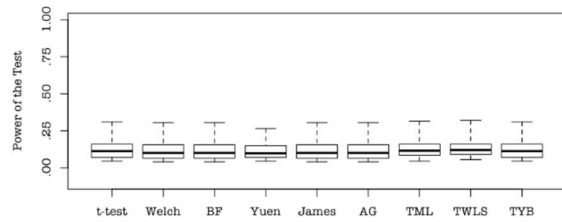
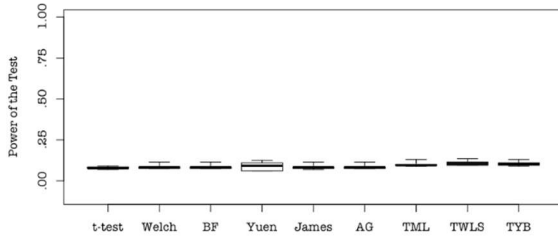
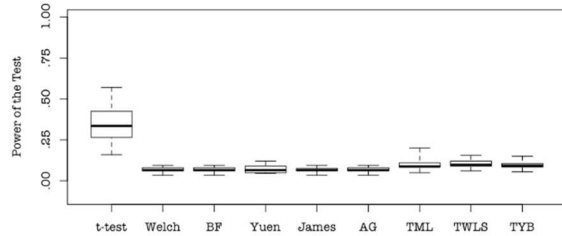
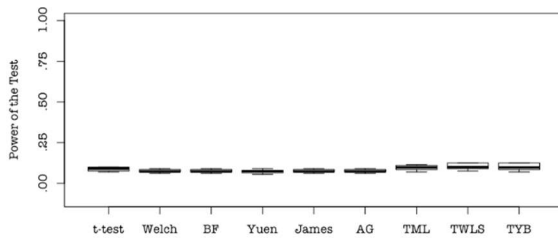
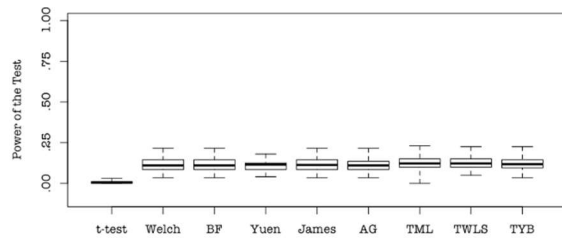


รูปที่ 4 อำนาจการทดสอบของสถิติทดสอบ Yuen, TML, TWLS, TYB, James, Welch, BF และ AG ภายใต้สถานการณ์ที่ขนาดอิทธิพล $d=0.2$ ระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน และขนาดตัวอย่างมีค่าแตกต่างกัน

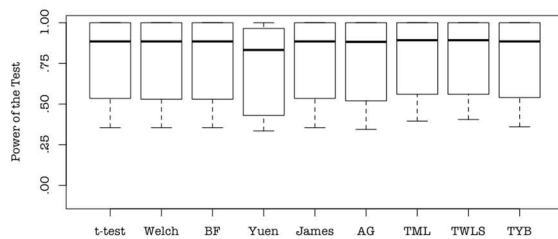
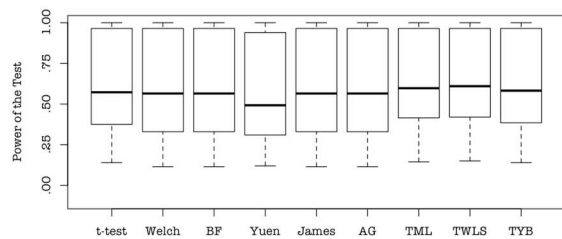
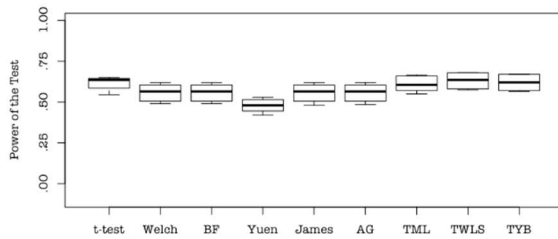
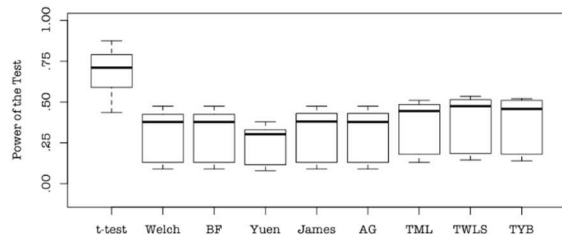
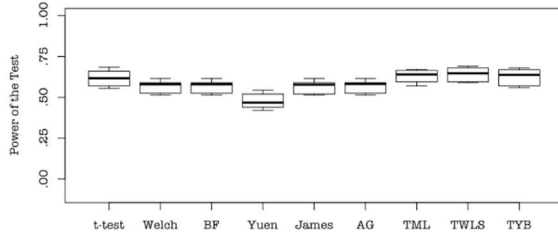
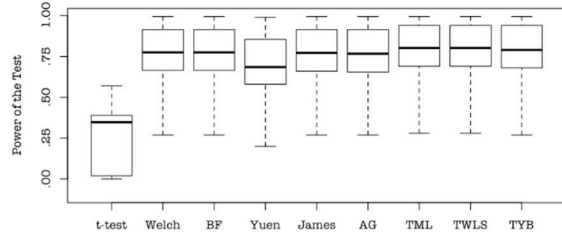


รูปที่ 5 อำนาจการทดสอบของสถิติทดสอบ Yuen, TML, TWLS, TYB, James, Welch, BF และ AG ภายใต้สถานการณ์ที่ขนาดอิทธิพล $d=0.8$ ระดับความไม่เป็นเอกพันธ์กันของความแปรปรวน และขนาดตัวอย่างมีค่าแตกต่างกัน

รูปที่ 6 และ 7 แสดงการเปรียบเทียบอำนาจการทดสอบของสถิติทดสอบแต่ละตัวภายใต้สถานการณ์จำลองต่างๆ พบว่า (1) กรณีที่ตัวอย่างมีขนาดเท่ากันระหว่างกลุ่ม สถิติทดสอบแต่ละตัวมีแนวโน้มให้อำนาจการทดสอบสูงที่สุดเมื่อเปรียบเทียบกับกรณีที่ตัวอย่างมีขนาดไม่เท่ากัน และในแต่ละสถานการณ์จำลองยังมีค่าอำนาจการทดสอบที่ใกล้เคียง (2) กรณีที่ตัวอย่างมีขนาดไม่เท่ากัน แต่ความแปรปรวนระหว่างกลุ่มเป็นเอกพันธ์กันพบว่าสถิติทดสอบแต่ละตัวมีแนวโน้มให้ค่าอำนาจการทดสอบที่ใกล้เคียงกัน แต่มีแนวโน้มให้ค่าอำนาจการทดสอบที่ลดลงเมื่อเปรียบเทียบกับกรณีที่ขนาดตัวอย่างเท่ากัน (3) กรณีที่ขนาดตัวอย่างไม่เท่ากัน และความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์กันพบว่า หากขนาดตัวอย่างในแต่ละกลุ่มแปรผันโดยตรงกับขนาดของความแปรปรวน ($n_1 > n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่า สถิติทดสอบที่มีแนวโน้มให้อำนาจการทดสอบสูงกว่าสถิติทดสอบตัวอื่นๆ และสถิติทดสอบแต่ละตัวยกเว้นสถิติทดสอบที่มีแนวโน้มให้ค่าอำนาจการทดสอบต่ำกว่ากรณีที่ความแปรปรวนระหว่างกลุ่มมีความเป็นเอกพันธ์กัน แต่หากขนาดตัวอย่างในแต่ละกลุ่มแปรผกผันกับขนาดของความแปรปรวน ($n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่าสถิติทดสอบที่มีแนวโน้มให้อำนาจการทดสอบต่ำกว่าสถิติทดสอบตัวอื่นๆ และสถิติทดสอบแต่ละตัวยกเว้นสถิติทดสอบที่มีแนวโน้มให้อำนาจการทดสอบที่สูงใกล้เคียงกับกรณีที่ตัวอย่างมีขนาดเท่ากัน

ก. $n_1 = n_2$ และ $\sigma_1^2 = \sigma_2^2$ ข. $n_1 = n_2$ และ $\sigma_1^2 > \sigma_2^2$ ค. $n_1 > n_2$ และ $\sigma_1^2 = \sigma_2^2$ ง. $n_1 > n_2$ และ $\sigma_1^2 > \sigma_2^2$ จ. $n_1 < n_2$ และ $\sigma_1^2 = \sigma_2^2$ ฉ. $n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$

รูปที่ 6 อำนาจการทดสอบของสถิติทดสอบแต่ละตัวภายใต้สถานการณ์ที่ขนาดอิทธิพล d=0.2

ก. $n_1 = n_2$ และ $\sigma_1^2 = \sigma_2^2$ ข. $n_1 = n_2$ และ $\sigma_1^2 > \sigma_2^2$ ค. $n_1 > n_2$ และ $\sigma_1^2 = \sigma_2^2$ ง. $n_1 > n_2$ และ $\sigma_1^2 > \sigma_2^2$ จ. $n_1 < n_2$ และ $\sigma_1^2 = \sigma_2^2$ ฉ. $n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$

รูปที่ 7 อำนาจการทดสอบของสถิติทดสอบแต่ละตัวภายใต้สถานการณ์ที่ขนาดอิทธิพล d=0.8

อภิปรายผลการวิจัย

1. ในกรณีที่ตัวอย่างสองกลุ่มมีขนาดไม่เท่ากันและขนาดตัวอย่างในแต่ละกลุ่มแปรผกผันกับความแปรปรวนในแต่ละกลุ่ม ($n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$) พบว่าสถิติทดสอบที่มีประสิทธิภาพในการทดสอบต่ำที่สุด โดยไม่สามารถควบคุมระดับนัยสำคัญของการทดสอบได้ และมีอำนาจการทดสอบที่ต่ำเมื่อเปรียบเทียบกับสถิติทดสอบอื่นๆ สาเหตุเนื่องมาจากสูตรการคำนวณค่าสถิติทดสอบที่เป็น $t = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ โดยที่ S_p คือ

ส่วนเบี่ยงเบนมาตรฐานรวม (pooled SD) ที่มีค่าเท่ากับ $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$ เมื่อสมมติฐานหลักเป็น

จริง และ $n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$ จะได้ว่า มีแนวโน้มจะลดต่ำลงเรื่อยๆ เมื่อขนาดตัวอย่างมีความแตกต่างกันมากขึ้น เพราะมีการให้น้ำหนักกับความแปรปรวน σ_1^2 ที่เป็นของกลุ่มตัวอย่างขนาดเล็กมากขึ้น ทำให้ค่าสถิติที่ภายใต้สถานการณ์ดังกล่าวมีแนวโน้มลดลงเมื่อความไม่เป็นเอกพันธ์ของความแปรปรวนระหว่างประชากรและขนาดตัวอย่างทั้งสองกลุ่มมีมากขึ้น เป็นการเพิ่มโอกาสของการปฏิเสธสมมติฐานหลักที่ถูกต้องให้สูงขึ้น ทำให้สถิติทดสอบที่ควบคุมระดับนัยสำคัญของการทดสอบได้ยากขึ้น ในขณะที่เมื่อสมมติฐานทางเลือกเป็นจริง และ $n_1 < n_2$ และ $\sigma_1^2 > \sigma_2^2$ จะได้ว่าสถิติทดสอบที่มีแนวโน้มจะมีค่าต่ำกว่าในกรณีอื่นๆ ทั้งนี้เพราะเทอม S_p จะมีค่าสูงเมื่อเปรียบเทียบกับกรณีที่ความแปรปรวนระหว่างกลุ่มประชากรเป็นเอกพันธ์ ทำให้สถิติทดสอบที่มีค่าต่ำ ส่งผลให้มีโอกาสปฏิเสธสมมติฐานน้อยลงเมื่อเปรียบเทียบกับสถานการณ์อื่นๆ ทำให้อำนาจการทดสอบของสถิติที่ต่ำในสถานการณ์ดังกล่าว ผลการวิจัยในส่วนนี้มีแนวโน้มที่สอดคล้องกับข้อค้นพบของ Overall, et al. (1995) และ Yuen (1974)

2. ถึงแม้ว่าในทางทฤษฎีสถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ย (SMM) ทั้งสามตัว ได้แก่ T_{ML} , T_{WLS} และ T_{VB} มีแนวโน้มจะให้ผลการทดสอบที่มีประสิทธิภาพเนื่องจากเป็นสถิติทดสอบสร้างขึ้นโดยอิงโมเดลที่มีข้อสมมติว่าความแปรปรวนระหว่างกลุ่มไม่เป็นเอกพันธ์กันแตกต่างจากการแก้ปัญหาในสถิติทดสอบตัวอื่นๆ ที่ใช้วิธีการปรับแก้สูตรการคำนวณค่าสถิติทดสอบ เช่น การใช้ Trimmed mean การใช้ค่ามัธยฐานของตัวอย่างแทนค่าเฉลี่ยตัวอย่าง หรือการใช้วิธีการถ่วงน้ำหนัก เป็นต้น อย่างไรก็ตาม เนื่องจากสถิติทดสอบดังกล่าวอิงอยู่กับสถิติทดสอบไคสแควร์ที่มีความไวต่อขนาดตัวอย่าง จึงทำให้สถิติทดสอบดังกล่าวมีประสิทธิภาพไม่ดีเท่าที่ควรเมื่อเปรียบเทียบกับสถิติทดสอบตัวอื่นๆ และเนื่องจากสถิติทดสอบ T_{ML} , T_{WLS} และ T_{VB} ดังกล่าวเป็นกรณีเฉพาะของการทดสอบความสอดคล้องเชิงประจักษ์ในการวิเคราะห์โมเดลสมการโครงสร้าง เมื่อเปรียบเทียบผลการวิจัยนี้กับผลการวิจัยที่ศึกษาประสิทธิภาพของการทดสอบไคสแควร์สำหรับการตรวจสอบความสอดคล้องเชิงประจักษ์ของโมเดลสมการโครงสร้างพบว่า มีพฤติกรรมที่สอดคล้องกัน (Hu & Bentler, 1998; Fan & Sivo, 2005, 2007; McNeish & Hancock, 2017) นอกจากนี้ สถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ยนี้มีข้อเด่นเหนือกว่าสถิติทดสอบตัวอื่นๆ คือสามารถประยุกต์ใช้ได้ในทุกกรณีทั้งการเปรียบเทียบค่าเฉลี่ยที่มากกว่าสองกลุ่ม และการเปรียบเทียบเวกเตอร์ค่าเฉลี่ยหลายกลุ่ม ภายใต้สถานการณ์ที่เมทริกซ์ความแปรปรวนร่วมไม่เป็นเอกพันธ์กัน ผลการวิจัยพบว่าสถิติทดสอบ Yuen มีแนวโน้มให้อำนาจการทดสอบต่ำที่สุดในทุกกรณีเมื่อเปรียบเทียบกับสถิติทดสอบที่มีความแกร่งตัวอื่นๆ สาเหตุเป็นเพราะสถิติทดสอบ Yuen แก้ปัญหาความไม่เป็นเอกพันธ์กันของความแปรปรวนด้วยการใช้ Trimmed mean แทนการใช้ค่าเฉลี่ยปกติ วิธีการดังกล่าวสามารถแก้ไขได้ทั้งปัญหาด้านการแจกแจงไม่ปกติ ความไม่เป็นเอกพันธ์กันของความแปรปรวน และค่าผิดปกติ อย่างไรก็ตาม ผู้วิจัยมองว่าเป็นวิธีการแก้ปัญหาที่ไม่มีประสิทธิภาพเนื่องจากการใช้ Trimmed mean ทำให้ต้องเสียข้อมูลบางส่วนในการวิเคราะห์ซึ่งส่งผลโดยตรงต่อประสิทธิภาพของการทดสอบ อีกทั้งการตัดข้อมูลส่วนหัวและส่วนท้ายออกไปจากชุดข้อมูลยังอาจไม่

สามารถแก้ปัญหาได้หมดโดยเฉพาะในกรณีที่การแจกแจงของข้อมูลมีความเบ้อยู่ในระดับสูง หรือมีค่าผิดปกติอยู่จำนวนมาก หรือมีความไม่เป็นเอกพันธ์กันของความแปรปรวนสูง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. กรณีที่ตัวอย่างสองกลุ่มมีขนาดเท่ากัน หากตัวอย่างมีขนาดใหญ่เพียงพอ (15 หน่วยขึ้นไป) ผู้วิจัยสามารถใช้สถิติทดสอบใดก็ได้ในการวิเคราะห์ เนื่องจากสถิติทดสอบแต่ละตัวมีประสิทธิภาพที่ใกล้เคียงกัน แต่หากตัวอย่างมีขนาดเล็กไม่ควรเลือกใช้สถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ย (SMM) ทั้งสามตัว ได้แก่ T_{ML} , T_{WLS} และ T_{YB} เนื่องจากมีแนวโน้มที่จะควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ในระดับที่กำหนดได้ยากกว่าสถิติทดสอบตัวอื่นๆ
2. กรณีที่ตัวอย่างสองกลุ่มมีขนาดไม่เท่ากัน และความแปรปรวนระหว่างกลุ่มมีความเป็นเอกพันธ์กันควรเลือกใช้สถิติทดสอบที่เนื่องจากเป็นสถิติทดสอบที่มีประสิทธิภาพในด้านการควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งดีที่สุด และมีอำนาจการทดสอบที่สูงเมื่อเปรียบเทียบกับสถิติทดสอบตัวอื่นๆ
3. กรณีที่ตัวอย่างสองกลุ่มมีขนาดไม่เท่ากัน และความแปรปรวนระหว่างกลุ่มไม่มีความเป็นเอกพันธ์กัน ให้ผู้วิจัยตรวจสอบว่าขนาดตัวอย่างในแต่ละกลุ่มแปรผันตรงหรือแปรผกผันกับขนาดของความแปรปรวน หากขนาดตัวอย่างในแต่ละกลุ่มแปรผันตรงกับขนาดของความแปรปรวนผู้วิจัยอาจเลือกใช้สถิติทดสอบของ James, Welch, BF หรือ AG เนื่องจากเป็นสถิติทดสอบที่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งได้ดี และมีอำนาจการทดสอบที่สูง แต่หากขนาดตัวอย่างในแต่ละกลุ่มแปรผกผันกับขนาดของความแปรปรวนผู้วิจัยสามารถเลือกใช้สถิติทดสอบตัวใดก็ได้ยกเว้นสถิติทดสอบที่ และสถิติทดสอบ Yuen

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

สถิติทดสอบที่ใช้ในการวิจัยนี้หลายตัว ถูกออกแบบมาให้สามารถใช้กับโมเดลการเปรียบเทียบค่าเฉลี่ยที่มีมากกว่าสองกลุ่มตัวอย่างได้ เช่น สถิติทดสอบ Welch สถิติทดสอบ Browne-Forsythe สถิติทดสอบ James และ สถิติทดสอบ Alexander-Govern สามารถใช้เพื่อเปรียบเทียบค่าเฉลี่ยหลายกลุ่มตัวอย่าง (Alexander, & Govern, 1994; Brown, & Forsythe, 1974; James, 1951; Welch, 1951) และสถิติทดสอบที่อิงโมเดลโครงสร้างค่าเฉลี่ย สามารถใช้เพื่อเปรียบเทียบค่าเฉลี่ยหลายกลุ่มตัวอย่าง ไปจนถึงการเปรียบเทียบเซนทรอยด์ ภายใต้สถานการณ์ที่ความแปรปรวนหรือเมทริกซ์ความแปรปรวนร่วมไม่เป็นเอกพันธ์กันระหว่างกลุ่ม (Yuan, & Bentler, 1997, 1999) ซึ่งปัจจุบันยังขาดการวิจัยที่ศึกษาเปรียบเทียบประสิทธิภาพระหว่างสถิติทดสอบภายใต้สถานการณ์ดังกล่าวอยู่ จึงควรมีการศึกษาประสิทธิภาพของสถิติทดสอบดังกล่าวเพิ่มเติมเพื่อขยายขอบเขตของผลการวิจัยนี้

เอกสารอ้างอิง

- Alexander, R. A., & Govern, D. M. (1994). A new and Simpler Approximation for ANOVA under Variance Heterogeneity. *Journal of Educational Statistics*, 19(2), 91-101.
- Algina, J., Keselman H. J., & Penfield, R. D. (2005). An Alternative to Coden's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case. *Psychological Methods*, 10, 317-328.

- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformation. **Journal of the Royal Statistical Society B**, 26, 211-252.
- Brown, M. B., & Forsythe, A. B. (1974). The Small Sample Behavior of Some Statistics Which Test the Equality of Means. **Technometrics**, 16(1), 129-132.
- Fan, X., & Sivo, S. A. (2005). Sensitivity of Fit Indexes to Misspecified Structural or Measurement Model Components: Rationale of Two-Index Strategy Revisited. **Structural Equation Modeling**, 12(3), 343-367.
- _____. (2007). Sensitivity of Fit Indices to Model Misspecification and model Types. **Multivariate Behavioral Research**, 42(3), 509-529.
- Field, A. P., Miles, J., & Field, Z. (2012). **Discovering Statistics Using R**. London: Sage Publication.
- Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). **Multivariate data analysis**. 7th ed. Upper saddle River, NJ: Pearson Education International.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. **Psychological Methods**, 3(4), 424-453.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. **Eiomerrika**, 38, 324-329.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1996). **Continuous Univariate Distributions**. 2nd Ed. New York: John Wiley.
- Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. **ACM Transactions on Modeling and Computer Simulation**, 8(1), 3-30.
- McNeish, D., An, J., & Hancock, G. R. (2017). The Thorny Relation Between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. **Journal of Personality Assessment**, 1, 1-10.
- Overall, J. E., Atlas, R. S., & Gibson, J. M. (1995). Tests that are robust against variance heterogeneity in k x 2 designs with unequal cell frequencies. **Psychological Report**, 76, 1011-10.
- Vexler, A., Hutson, A. D., & Chen, X. (2016). **Statistical testing strategies in the health sciences**. New York: Chapman and Hall.
- Welch, B. L. (1938). The significance of the difference between two means when population variances are unequal. **Biometrika**, 29, 350-362.
- _____. (1951). On the comparison of several means: An alternative approach. **Biometrika**, 38(3-4), 330-336.
- Whitley, E., & Ball, J. (2002). Statistics Review 4: Sample Size Calculations. **Critical Care**, 6, 335-341.

- Wilcox, R. R., & Schonbrodt, F. D. (2016). **The WRS package for robust statistics in R (version 0.30)**. Retrieved December 15, 2017, from <http://github.com/nicebread/WRS>
- Wong, R. K. W., Yao, F., & Lee, T. C. M. (2014). Robust Estimation for Generalized Additive Models. **Journal of Computational and Graphical Statistics**, **23**(1), 270–289.
- Yuen, K. (1974). The two-sample trimmed t for unequal population variances. **Biometrika**, **61**(1), 165–170.
- Yuan, K. H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. **Journal of the American Statistical Association**, **92**, 767-774.
- _____. (1999). F tests for mean and covariance structure analysis. **Journal of Educational and Behavioral Statistics**, **24**, 225-243.