

การเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้ทดสอบค่าเฉลี่ย ของ 3 ประชากร ในกรณีที่ความแปรปรวนไม่เท่ากัน

Efficiency Comparison of Statistic for Testing 3 Population Means in Case of Heterogeneity of Variance

จิราภา โคมเดือน* และอัชฌา อระวีพร

ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ถนนฉลองกรุง เขตลาดกระบัง กรุงเทพมหานคร 10520

Jirapa Khomduean* and Autcha Araveeporn

Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang,

Chalongkrung Road, Ladkrabang, Bangkok 10520

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้ทดสอบค่าเฉลี่ยของ 3 ประชากร ในกรณีที่ความแปรปรวนไม่เท่ากัน ในการวิเคราะห์ความแปรปรวนด้วยสถิติทดสอบเอฟ (F) สถิติทดสอบเวลช์ (W) สถิติทดสอบมาราสคิโล (M) และสถิติทดสอบบูตสเตรปแบบใช้พารามิเตอร์ (PB) เป็นการทดสอบสถิติเพื่อเปรียบเทียบความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ โดยใช้เกณฑ์ของ Bradley เป็นเกณฑ์ ในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่ระดับนัยสำคัญ 0.01 และ 0.05 ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลและแต่ละกรณีถูกทำซ้ำ 1,000 ครั้ง จากการแจกแจงปกติ และแกมมา เมื่อขนาดของกลุ่มตัวอย่างเท่ากันและไม่เท่ากันโดยขึ้นอยู่กับอัตราส่วนความแตกต่างของความแปรปรวน ผลการวิจัยพบว่า สถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และมีกำลังการทดสอบสูงที่สุดทุกกรณีที่มีการแจกแจงปกติ สำหรับการแจกแจงแกมมา สถิติทดสอบ F, W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และมีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี

คำสำคัญ : ความแปรปรวนไม่เท่ากัน; ความน่าจะเป็นของความผิดพลาดแบบที่ 1; กำลังการทดสอบ

Abstract

The objective of this research is to compare the efficiency of statistic for testing three population means in case of heterogeneity of variance. In analysis of variance using F-test (F), Welch's test (W), Marascuilo's test (M), and Parametric Bootstrap approach test (PB) are testing statistics for computing probability of type I error and power of a test. The Bradley's is a criterion

to control the probability of type I error at 0.01 and 0.05 significance level. The data of this research is simulated by using the Monte Carlo technique and each case is replicated 1,000 times. The normal and gamma distributions are generated from equal and unequal sample sizes based on difference ratio of variances. The results of this research are found that W and M can control the probability of type I error and shows the highest power of a test all cases with normal distribution. For gamma distribution, F, W, and M can control the probability of type I error and shows the highest power of a test in almost case.

Keywords: heterogeneity of variance; probability of type I error; power of a test

1. บทนำ

งานวิจัยทั่วไปการศึกษาลักษณะของประชากรเป็นไปได้อย่างยาก เนื่องจากขนาดของประชากรมีจำนวนมากทำให้ต้องเสียค่าใช้จ่ายและใช้เวลานานในการจัดเก็บข้อมูล จึงต้องสุ่มตัวอย่างมาเพียงบางส่วนแล้วนำข้อมูลไปคำนวณหาค่าสถิติ (statistic) เพื่อประมาณค่าพารามิเตอร์ (parameter) และใช้อ้างอิงไปยังลักษณะของประชากร เรียกว่า การอนุมานเชิงสถิติ (statistical inference) ซึ่งการอนุมานเชิงสถิตินี้จะมีวัตถุประสงค์หลักที่สำคัญ คือ การประมาณค่า (estimation) และการทดสอบสมมติฐาน (test of hypothesis) การที่งานวิจัยจะให้ผลสรุปที่น่าเชื่อถือ มีความเที่ยงตรงและแม่นยำใกล้เคียงกับความเป็นจริงนั้น ผู้วิจัยต้องมีความรู้ในการทดสอบสมมติฐาน การสุ่มตัวอย่าง การเลือกใช้สถิติให้เหมาะสมกับลักษณะของข้อมูล และที่สำคัญอีกประการหนึ่ง คือ ผู้วิจัยต้องมีความรู้เกี่ยวกับข้อตกลงเบื้องต้น (assumption) ของสถิติแต่ละตัวซึ่งก็มีลักษณะของข้อมูลที่จะนำมาวิเคราะห์แตกต่างกันไป

กรณีที่ต้องการทดสอบสมมติฐานเพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่ม ขึ้นไป เทคนิคการทดสอบสมมติฐานทางสถิติที่นิยมใช้กัน คือ การวิเคราะห์ความแปรปรวน (analysis of variance, ANOVA) โดยใช้สถิติทดสอบเอฟ (F-

test) ซึ่งมีข้อตกลงเบื้องต้น คือ ข้อมูลแต่ละประชากรถูกสุ่มจากประชากรที่มีการแจกแจงปกติ (normal distribution) ข้อมูลแต่ละประชากรถูกสุ่มจากประชากรที่มีความแปรปรวนเท่ากันหรือคงที่ (homogeneity of variance) และข้อมูลแต่ละประชากรถูกสุ่มจากประชากรเป็นอิสระกัน อย่างไรก็ตาม ในความเป็นจริงแล้วข้อมูลที่มีอยู่อาจไม่เป็นไปตามข้อตกลงเบื้องต้นทุกประการ ดังนั้นจึงต้องใช้สถิติทดสอบอื่นแทน

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่าในปี พ.ศ. 2549 พลชาติ [1] ได้เปรียบเทียบตัวสถิติทดสอบเวลช์ (W) บราวน์-ฟอร์ลิตส์ (BF) และมาร์สควิลโล (M) สำหรับการทดสอบความเท่ากันของค่าเฉลี่ย เมื่อความแปรปรวนของประชากรไม่เท่ากัน พบว่าสถิติทดสอบ BF สามารถควบคุมความผิดพลาดแบบที่ 1 และมีอำนาจการทดสอบสูงสุด เมื่อประชากรมีการแจกแจงเอกซ์โปเนนเชียล ส่วนสถิติทดสอบ M สามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อประชากรมีการแจกแจงที่ และมีอำนาจการทดสอบสูงสุด เมื่อประชากรมีการแจกแจงปกติ และที่ ถัดมาในปี ค.ศ. 2007 Krishnamoorthy และคณะ [2] ได้ศึกษาวิธีบูตสเตรปแบบใช้พารามิเตอร์ สำหรับการวิเคราะห์ความแปรปรวนในขณะที่มีความแปรปรวนไม่เท่ากัน : ตัวแบบคงที่ และตัวแบบสุ่ม เปรียบเทียบกับการทดสอบของ 3 วิธี ได้แก่ การทดสอบเวลช์ (W) การทดสอบเจมส์ และ

การทดสอบ generalized F (GF) พบว่าการทดสอบ PB เป็นวิธีที่ดีที่สุด และหลังจากนั้น ปี ค.ศ. 2010 Yigit และ Gokpinar [3] ได้ศึกษาการจำลองข้อมูลการทดสอบการวิเคราะห์ความแปรปรวนทางเดียวภายใต้ความแปรปรวนไม่เท่ากัน ด้วยใช้สถิติทดสอบ ดังนี้ การทดสอบ F, บราวน์-ฟอร์ลิต (BF), Weerahandi's generalized F (GF), วิธีบูตสเตรปแบบใช้พารามิเตอร์ (PB), Scott-Smith (SS), one-stage (OS), one-stage range (OSR), เวลช์ (W) และ Xu-Wang's generalized F (XW) พบว่าการทดสอบ W, GF และ PB มีผลดีที่สุด

จากการศึกษางานวิจัยต่าง ๆ จึงสรุปได้ว่า ผู้วิจัยจะศึกษาเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบทั้ง 4 ตัว ได้แก่ การวิเคราะห์ความแปรปรวนด้วยสถิติทดสอบเอฟ (F) สถิติทดสอบเวลช์ (W) สถิติทดสอบมาราสควิลโล (M) และสถิติทดสอบบูตสเตรปแบบใช้พารามิเตอร์ (PB) ในกรณีที่ข้อมูลมีความแปรปรวนไม่เท่ากัน และประชากรมีการแจกแจงปกติ และการแจกแจงแกมมา ภายใต้การสร้างข้อมูลให้มีลักษณะตามที่กำหนดในประชากร 3 กลุ่ม โดยใช้โปรแกรมอาร์ (R) ในการจำลองและวิเคราะห์ข้อมูล ส่วนสถิติทดสอบบราวน์-ฟอร์ลิต (BF) ที่ไม่นำมาศึกษาเพราะค่าความผิดพลาดแบบที่ 1 เกินเกณฑ์ที่กำหนดในงานวิจัยของ Yigit และ Gokpinar [3] ส่วนสถิติทดสอบ generalized F ที่ไม่นำมาศึกษาเพราะค่าความผิดพลาดแบบที่ 1 ค่อนข้างต่ำในงานวิจัยของ Krishnamoorthy และคณะ [2] แต่ส่วนสถิติทดสอบเวลช์ที่นำมาศึกษาเนื่องจากในโปรแกรมสำเร็จรูป เช่น โปรแกรม SPSS มีการใช้สถิติทดสอบเวลช์ในการทดสอบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่ม เมื่อความแปรปรวนไม่เท่ากัน

ในการวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาวิธีการทดสอบโดยใช้สถิติทดสอบ 4 ตัว ดังนี้ (1) การ

วิเคราะห์ความแปรปรวนด้วยสถิติทดสอบเอฟ (F) (2) สถิติทดสอบเวลช์ (W) (3) สถิติทดสอบมาราสควิลโล (M) และ (4) สถิติทดสอบบูตสเตรปแบบใช้พารามิเตอร์ (PB) เพื่อเปรียบเทียบค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบทั้ง 4 ตัว ตามเกณฑ์ของ Bradley และเพื่อเปรียบเทียบกำลังการทดสอบของสถิติทดสอบทั้ง 4 ตัว

2. ตัวสถิติที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้ศึกษาตัวสถิติที่ใช้ทดสอบการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่ม โดยมีสมมติฐานว่าง (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่ โดยที่ } i \neq j$$

ตัวสถิติที่ใช้ในการวิจัยครั้งนี้ ประกอบด้วย

2.1 การวิเคราะห์ความแปรปรวน (analysis of variance, ANOVA) [4]

ในการทดสอบความเท่ากันของค่าเฉลี่ยในกรณีตั้งแต่ 3 ประชากร ขึ้นไป และข้อมูลที่นำมาวิเคราะห์เป็นไปตามข้อกำหนดเบื้องต้นทั้ง 3 ข้อ ดังนี้ (1) ข้อมูลแต่ละประชากรสุ่มจากประชากรที่มีการแจกแจงปกติ (2) ข้อมูลแต่ละประชากรสุ่มจากประชากรที่มีความแปรปรวนเท่ากันหรือคงที่ และ (3) ข้อมูลแต่ละประชากรสุ่มจากประชากรที่เป็นอิสระกัน

สถิติที่ใช้ในการทดสอบ คือ

$$F_{cal} = \frac{\sum_{i=1}^k \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}}{\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^k \frac{x_{i.}^2}{n_i}} \bigg/ (N - k)$$

โดยที่ x_{ij} คือ ค่าสังเกตที่ j ที่ได้รับจากประชากรที่ i ; k คือ จำนวนประชากร; $N = \sum_{i=1}^k n_i$; $x_{i.} =$

$$\sum_{j=1}^{n_i} x_{ij}; x_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}$$

โดยจะปฏิเสธสมมุติฐานว่างเมื่อค่าสถิติทดสอบ F_{cal} มากกว่าค่า $F_{\alpha; k-1, N-k}$ เมื่อ $F_{\alpha; k-1, N-k}$ คือ ค่าวิกฤตที่เปิดจากตารางการแจกแจงสถิติ F ; α คือ ระดับนัยสำคัญ

2.2 สถิติทดสอบเวลช์ (Welch's test, W) [5]

คิดค้นโดย Welch (1947, 1951) ซึ่งได้ดัดแปลงมาจากการทดสอบของ Cochran เพื่อนำมาใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ย k ประชากร เมื่อความแปรปรวนของประชากรไม่เท่ากัน โดยพบว่าสถิติทดสอบเวลช์มีความแม่นยำมากกว่าสถิติทดสอบเอฟ เมื่อความแปรปรวนของประชากรแตกต่างกัน โดยเฉพาะเมื่อขนาดตัวอย่างไม่เท่ากันสำหรับกรณีที่ประชากรมีการแจกแจงปกติ

สถิติที่ใช้ในการทดสอบ คือ

$$W_0 = \frac{\sum_{i=1}^k w_i [(\bar{x}_i - \bar{x})^2 / (k-1)]}{1 + \frac{2(k-2)}{(k^2-1)} \left[\sum_{i=1}^k \left(1 - \frac{w_i}{u}\right)^2 / n_i - 1 \right]}$$

โดยที่ $w_i = n_i / S_i^2$; $u = \sum_{i=1}^k w_i$; $\bar{x} = \sum_{i=1}^k w_i \bar{x}_i / u$

$$; \bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}; S_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{(n_i - 1)}$$

โดยจะปฏิเสธสมมุติฐานว่างเมื่อสถิติทดสอบเวลช์ W_0 มีค่ามากกว่าค่าวิกฤตจากตารางการแจกแจงสถิติ F ที่องศาเสรี df_1 เป็น $k-1$ และ df_2 เป็น f เมื่อ

$$f = \left[\left(3 / (k^2 - 1) \right) \sum_{i=1}^k \left(1 - \frac{w_i}{u} \right)^2 \right] \times \frac{1}{n_i - 1} \Bigg]^{-1}$$

2.3 สถิติทดสอบมาราสคิลโล (Marascuilo's test, M) [6]

เสนอโดย Marascuilo (1971) ได้ดัดแปลงจากการทดสอบของ Welch ทำให้รูปแบบการคำนวณที่ง่ายขึ้น ซึ่งเป็นวิธีการทดสอบความเท่ากันของค่าเฉลี่ยประชากรที่มากกว่า 2 กลุ่ม ในกรณีที่ความ

แปรปรวนของประชากรไม่เท่ากัน โดยเหมาะสมกับกลุ่มตัวอย่างที่มีขนาดใหญ่

สถิติที่ใช้ในการทดสอบ คือ

$$M = \sum_{i=1}^k w_i [(\bar{x}_i - \bar{x})^2 / (k-1)]$$

โดยที่ $w_i = n_i / S_i^2$; $u = \sum_{i=1}^k w_i$; $\bar{x} = \sum_{i=1}^k w_i \bar{x}_i / u$

โดยจะปฏิเสธสมมุติฐานว่างเมื่อสถิติทดสอบมาราสคิลโล M มีค่ามากกว่าค่าวิกฤตจากตารางการแจกแจงสถิติ F ที่องศาเสรี df_1 เป็น $k-1$ และ df_2 เป็น f เมื่อ f แทนค่าเหมือนของสถิติทดสอบเวลช์

2.4 วิธีบูตสเตรปแบบใช้พารามิเตอร์ (parametric bootstrap approach, PB) [2]

วิธีบูตสเตรปแบบใช้พารามิเตอร์เป็นวิธีการประมาณค่าโดยใช้การสุ่มตัวอย่างจากประชากรแบบใส่คืน (replacement) ภายใต้เงื่อนไขของการใช้สถิติมีพารามิเตอร์ที่ว่าตัวแปรที่ศึกษาอยู่ในมาตรวัดระดับช่วงขึ้นไป ข้อมูลที่ได้จากกลุ่มตัวอย่างมีการแจกแจงปกติและประชากรแต่ละกลุ่มมีความแปรปรวนเท่ากัน

สถิติที่ใช้ในการทดสอบ คือ

$$T_{NB}(Z_i, \chi_{n_i-1}^2; S_i^2) = \sum_{i=1}^k \frac{Z_i^2(n_i-1)}{\chi_{n_i-1}^2} - \frac{\left[\sum_{i=1}^k \left(\sqrt{n_i} Z_i(n_i-1) / \{S_i \chi_{n_i-1}^2\} \right) \right]^2}{\sum_{i=1}^k (n_i(n_i-1) / \{S_i^2 \chi_{n_i-1}^2\})}$$

โดยที่ $Z_i \sim N(0,1)$; $T_{N0} = \sum_{i=1}^k \frac{n_i}{S_i^2} \bar{x}_i^2 - \left[\left(\sum_{i=1}^k n_i \bar{x}_i / S_i^2 \right)^2 / \sum_{i=1}^k n_i / S_i^2 \right]$

; $\chi_{n_i-1}^2$ คือ ค่าวิกฤตที่เปิดจากตารางการแจกแจงไคกำลังสอง ที่องศาเสรี df เป็น $n_i - 1$

โดยจะปฏิเสธสมมุติฐานว่าง เมื่อ

$$P(T_{NB}(Z_i, \chi_{n_i-1}^2; S_i^2) > T_{N0}) < \alpha$$

3. วิธีการดำเนินการวิจัย

ในการวิจัยครั้งนี้มีขั้นตอนดังนี้ คือ

3.1 กำหนดจำนวนประชากรเป็น 3 ประชากร

3.2 กำหนดกลุ่มตัวอย่างของประชากรเป็น 2 กรณี คือ กรณีกลุ่มตัวอย่างมีขนาดเท่ากัน และกรณีกลุ่มตัวอย่างมีขนาดไม่เท่ากัน โดยมีรายละเอียดขนาดตัวอย่างที่ใช้ในการศึกษา ดังนี้

ลักษณะ	รายละเอียด
ขนาดเท่ากัน	(10,10,10),(30,30,30),(50,50,50)
ขนาดไม่เท่ากัน	(5,10,15),(25,30,35),(45,50,55)

3.3 กำหนดการแจกแจงของประชากร 2 แบบ ได้แก่

3.3.1 การแจกแจงปกติ (normal distribution) ประกอบด้วยค่าพารามิเตอร์เป็น μ และ σ^2 โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็น ดังนี้

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

โดยมีค่าเฉลี่ย คือ $E(X) = \mu$ และค่าความแปรปรวน คือ $Var(X) = \sigma^2$

3.3.2 การแจกแจงแกมมา (gamma distribution) ประกอบด้วยค่าพารามิเตอร์เป็น α และ β โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็น ดังนี้

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & , x > 0, \alpha > 0, \beta > 0 \\ 0 & , x \text{ มีค่าอื่น ๆ} \end{cases}$$

โดยมีค่าเฉลี่ย คือ $E(X) = \alpha\beta$ และค่าความแปรปรวน คือ $Var(X) = \alpha\beta^2$

3.4 กำหนดความแตกต่างของความแปรปรวน โดยใช้ค่าอนเซนทรลิตีพารามิเตอร์ ϕ (noncentrality parameter) [7] เป็นเกณฑ์วัดความแตกต่างของความแปรปรวน โดยที่

$$\phi = \left[\left(\sum_{i=1}^k (\sigma_i^2 - \overline{\sigma^2})^2 / k \right) / \sigma_1^2 \right]^{1/2}$$

เมื่อ σ_1^2 คือ ค่าความแปรปรวนของประชากรที่มีค่า

ต่ำสุด; σ_i^2 คือ ค่าความแปรปรวนของประชากรที่ i โดย $i = 1, 2, \dots, k$; $\overline{\sigma^2}$ คือ ค่าเฉลี่ยความแปรปรวนของประชากรทั้ง k กลุ่ม

รายละเอียดระดับความแตกต่างของความแปรปรวนที่ใช้ในวิจัยมีดังนี้

ระดับความแตกต่างของความแปรปรวน	ความแปรปรวนแต่ละประชากร	ϕ
มีความแตกต่างกันน้อย ($0 < \phi < 1.5$)	1.5 : 2.4 : 3	0.5
มีความแตกต่างกันปานกลาง ($1.5 \leq \phi < 3.0$)	1.5 : 6 : 9	2.5
มีความแตกต่างกันมาก ($\phi \geq 3.0$)	1.5 : 18 : 36	11.5

3.5 กำหนดระดับนัยสำคัญ (α) 2 ระดับ คือ 0.01 และ 0.05

3.6 ใช้โปรแกรมอาร์ (R) เวอร์ชัน 3.2.2 ในการจำลองและวิเคราะห์ข้อมูล ทดลองซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์

3.7 คำนวณหาความน่าจะเป็นของความผิดพลาดแบบที่ 1 และเปรียบเทียบค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley [8] สามารถจำแนกตามระดับนัยสำคัญได้ดังนี้ $\alpha = 0.01$ ช่วงที่กำหนด คือ [0.005, 0.015] และ $\alpha = 0.05$ ช่วงที่กำหนด คือ [0.025, 0.075] โดยกำหนดให้แต่ละการแจกแจงมีค่าเฉลี่ยเท่ากัน

3.8 การคำนวณหากล้าการทดสอบ จะกำหนดอัตราส่วนค่าเฉลี่ยของประชากรเป็น 1:2:3 โดยกำหนดให้แต่ละการแจกแจงให้มีค่าเฉลี่ยไม่เท่ากัน

3.9 เปรียบเทียบกำลังการทดสอบของสถิติทดสอบ โดยสถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์ของ Bradley และมีกำลังการทดสอบสูงที่สุดจะเป็นสถิติ

ทดสอบที่ดีที่สุด

4. ผลการวิจัย

ผลการวิจัยครั้งนี้แบ่งเป็น 2 ส่วน คือ ความน่าจะเป็นของความผิดพลาดแบบที่ 1 และ กำลังการทดสอบ

4.1 ความน่าจะเป็นของความผิดพลาดแบบที่ 1

พิจารณาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่ระดับนัยสำคัญ 0.01 และ 0.05 ของการแจกแจงปกติและการแจกแจงแกมมา

4.1.1 ระดับนัยสำคัญ 0.01

พิจารณาตามเกณฑ์ของ Bradley โดยสถิติทดสอบจะสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่ออยู่ในช่วงที่กำหนด คือ $[0.005, 0.015]$ แสดงดังตารางที่ 1 และ 2

(1) เมื่อประชากรมีการแจกแจงปกติ และมีค่าอนเซนทรีลิตี้พารามิเตอร์ (ϕ) ต่างกัน จากตารางที่ 1 สรุปผลดังนี้

$\phi = 0.5$ สถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น ขนาดตัวอย่าง (n) เท่ากับ (5,10,15) และ (25,30,25) ส่วนสถิติทดสอบ PB สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เมื่อ n เท่ากับ (30,30,30) และ (25,30,35)

$\phi = 2.5$ สถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (50,50,50)

ตารางที่ 1 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบ F, W, M และ PB เมื่อประชากรมีการแจกแจงปกติ ที่ระดับนัยสำคัญ 0.01

ขนาดตัวอย่าง (n_1, n_2, n_3)	สถิติทดสอบ	ค่าอนเซนทรีลิตี้พารามิเตอร์ ϕ		
		0.5	2.5	11.5
(10,10,10)	F	0.010*	0.009*	0.013*
	W	0.007*	0.009*	0.017
	M	0.008*	0.011*	0.017
	PB	0.192	0.186	0.184
(30,30,30)	F	0.009*	0.015*	0.013*
	W	0.009*	0.007*	0.009*
	M	0.009*	0.007*	0.009*
	PB	0.005*	0.004	0.004
(50,50,50)	F	0.009*	0.020	0.022
	W	0.011*	0.014*	0.014*
	M	0.011*	0.015*	0.014*
	PB	0	0	0
(5,10,15)	F	0.010*	0.006*	0.004
	W	0.016	0.008*	0.011*
	M	0.016	0.011*	0.012*
	PB	0.225	0.182	0.188
(25,30,35)	F	0.013*	0.008*	0.005*
	W	0.017	0.009*	0.005*
	M	0.018	0.010*	0.005*
	PB	0.007*	0.001	0.003
(45,50,55)	F	0.009*	0.008*	0.010*
	W	0.007*	0.007*	0.008*
	M	0.008*	0.007*	0.008*
	PB	0	0.001	0.001

*ควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้

ตารางที่ 2 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบ F, W, M และ PB เมื่อประชากรมีการแจกแจงแกมมา ที่ระดับนัยสำคัญ 0.01

ขนาดตัวอย่าง (n_1, n_2, n_3)	สถิติทดสอบ	ค่าอนเซนทริลิตี้พารามิเตอร์ ϕ		
		0.5	2.5	11.5
(10,10,10)	F	0.011*	0.013*	0.025
	W	0.010*	0.017	0.038
	M	0.012*	0.018	0.039
	PB	0.179	0.212	0.232
(30,30,30)	F	0.008*	0.019	0.017
	W	0.008*	0.014*	0.018
	M	0.008*	0.016	0.020
	PB	0.007*	0.008*	0.016
(50,50,50)	F	0.015*	0.014*	0.027
	W	0.008*	0.013*	0.030
	M	0.009*	0.013*	0.032
	PB	0.001	0	0.003
(5,10,15)	F	0.005*	0.007*	0.012*
	W	0.011*	0.014*	0.027
	M	0.011*	0.016	0.029
	PB	0.228	0.185	0.215
(25,30,35)	F	0.008*	0.006*	0.006*
	W	0.010*	0.012*	0.016
	M	0.010*	0.013*	0.017
	PB	0.005*	0.003	0.007*
(45,50,55)	F	0.014*	0.013*	0.011*
	W	0.016	0.009*	0.019
	M	0.017	0.009*	0.020
	PB	0	0	0.002

*ควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้

$\phi = 11.5$ สถิติทดสอบ W และ M

สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (10,10,10) ส่วนสถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (50,50,50) และ (5,10,15)

(2) เมื่อประชากรมีการแจกแจงแกมมา และมีย่านอนเซนทริลิตี้พารามิเตอร์ (ϕ) ต่างกัน จากตารางที่ 2 สรุปผลดังนี้

$\phi = 0.5$ สถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (45,50,55) ส่วนสถิติทดสอบ PB สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เมื่อ n เท่ากับ (30,30,30) และ (25,30,35)

$\phi = 2.5$ สถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (30,30,30) ส่วนสถิติทดสอบ W สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (10,10,10) ส่วนสถิติทดสอบ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (10,10,10), (30,30,30) และ (5,10,15) ส่วนสถิติทดสอบ PB สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เมื่อ n เท่ากับ (30,30,30)

$\phi = 11.5$ สถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้พิจารณาตามเกณฑ์ของ Bradley โดยสถิติทดสอบจะสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่ออยู่ในช่วงที่กำหนด คือ [0.025,0.075] แสดงดังตารางที่ 3 และ 4

ตารางที่ 3 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบ F, W, M และ PB เมื่อประชากรมีการแจกแจงแกมมา ที่ระดับนัยสำคัญ 0.05

ขนาดตัวอย่าง (n_1, n_2, n_3)	สถิติทดสอบ	ค่านอนเซนทรัลลิตี้พารามิเตอร์ ϕ		
		0.5	2.5	11.5
(10,10,10)	F	0.052*	0.051*	0.059*
	W	0.049*	0.043*	0.049*
	M	0.057*	0.052*	0.052*
	PB	0.247	0.255	0.241
(30,30,30)	F	0.047*	0.053*	0.051*
	W	0.045*	0.047*	0.044*
	M	0.048*	0.048*	0.046*
	PB	0.012	0.010	0.012
(50,50,50)	F	0.044*	0.070*	0.069*
	W	0.045*	0.066*	0.064*
	M	0.047*	0.066*	0.064*
	PB	0	0	0
(5,10,15)	F	0.040*	0.025*	0.024
	W	0.056*	0.047*	0.057*
	M	0.065*	0.056*	0.059*
	PB	0.286	0.245	0.244
(25,30,35)	F	0.046*	0.041*	0.029*
	W	0.050*	0.048*	0.037*
	M	0.051*	0.049*	0.038*
	PB	0.018	0.009	0.006
(45,50,55)	F	0.043*	0.026*	0.053*
	W	0.045*	0.035*	0.050*
	M	0.045*	0.037*	0.051*
	PB	0	0.001	0.001

*ควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้

เกือบทุกกรณี ยกเว้นขนาดตัวอย่างเท่ากัน ส่วนสถิติทดสอบ PB สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เมื่อ n เท่ากับ (25,30,35)

4.1.2 ระดับนัยสำคัญ 0.05

(1) เมื่อประชากรมีการแจกแจงปกติ และมีค่านอนเซนทรัลลิตี้พารามิเตอร์ (ϕ) ต่างกัน จากตารางที่ 3 สรุปผลดังนี้

$\phi = 0.5$ และ 2.5 สถิติทดสอบ F, W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี

$\phi = 11.5$ สถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (5,10,15)

(2) เมื่อประชากรมีการแจกแจงแกมมา และมีค่านอนเซนทรัลลิตี้พารามิเตอร์ (ϕ) ต่างกัน จากตารางที่ 4 สรุปผลดังนี้

$\phi = 0.5$ สถิติทดสอบ F, W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี

$\phi = 2.5$ สถิติทดสอบ W และ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณี ส่วนสถิติทดสอบ F สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (5,10,15)

$\phi = 11.5$ สถิติทดสอบ F และ W สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (10,10,10) ส่วนสถิติทดสอบ M สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เกือบทุกกรณี ยกเว้น n เท่ากับ (10,10,10) และ (5,10,15)

ตารางที่ 4 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของสถิติทดสอบ F, W, M และ PB เมื่อประชากรมีการแจกแจงแกมมา ที่ระดับนัยสำคัญ 0.05

ขนาดตัวอย่าง (n_1, n_2, n_3)	สถิติทดสอบ	ค่าอนเซนทรลิตี้พารามิเตอร์ ϕ		
		0.5	2.5	11.5
(10,10,10)	F	0.053*	0.068*	0.090
	W	0.049*	0.066*	0.105
	M	0.051*	0.072*	0.117
	PB	0.234	0.275	0.281
(30,30,30)	F	0.053*	0.075*	0.068*
	W	0.050*	0.067*	0.072*
	M	0.053*	0.069*	0.074*
	PB	0.010	0.017	0.025*
(50,50,50)	F	0.050*	0.053*	0.073*
	W	0.050*	0.043*	0.074*
	M	0.050*	0.047*	0.074*
	PB	0.001	0.001	0.005
(5,10,15)	F	0.034*	0.022	0.031*
	W	0.044*	0.048*	0.073*
	M	0.053*	0.050*	0.076
	PB	0.280	0.227	0.248
(25,30,35)	F	0.043*	0.046*	0.026*
	W	0.047*	0.058*	0.062*
	M	0.051*	0.060*	0.066*
	PB	0.015	0.015	0.018
(45,50,55)	F	0.047*	0.053*	0.050*
	W	0.055*	0.053*	0.065*
	M	0.055*	0.054*	0.066*
	PB	0.001	0	0.003

*ควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้

4.2 กำลังการทดสอบ

การเปรียบเทียบกำลังการทดสอบของสถิติทดสอบจะเปรียบเทียบเฉพาะกรณีที่สถิติทดสอบสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เท่านั้น โดยพิจารณาจากระดับนัยสำคัญ 0.01 และ 0.05

4.2.1 ระดับนัยสำคัญ 0.01

ตัวสถิติที่ให้กำลังการทดสอบสูงสุดแสดงดังตารางที่ 5 และสรุปผลได้ดังนี้

(1) เมื่อประชากรมีการแจกแจงปกติ สถิติทดสอบ W และ M มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 0.5$ $n = (5,10,15)$ และ $(25,30,35)$ และ $\phi = 11.5$ $n = (10,10,10)$ ส่วนสถิติทดสอบ F มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 2.5$ $n = (50,50,50)$ และ $\phi = 11.5$ $n = (50,50,50)$ และ $(5,10,15)$ และส่วนสถิติทดสอบ PB มีกำลังการทดสอบสูงที่สุด เมื่อ $\phi = 0.5$ $n = (30,30,30)$ และ $(25,30,35)$

(2) เมื่อประชากรมีการแจกแจงแกมมา สถิติทดสอบ F มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 2.5$ $n = (30,30,30)$ และ $\phi = 11.5$ n เท่ากัน ส่วนสถิติทดสอบ W มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 0.5$ $n = (45,50,55)$ $\phi = 2.5$ $n = (10,10,10)$ และ $\phi = 11.5$ ทุกกรณี ส่วนสถิติทดสอบ M มีกำลังทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 0.5$ $n = (45,50,55)$ $\phi = 2.5$ $n = (10,10,10)$ $(30,30,30)$ และ $(5,10,15)$ และ $\phi = 11.5$ ทุกกรณี และส่วนสถิติทดสอบ PB มีกำลังทดสอบสูงที่สุด เมื่อ $\phi = 0.5$ $n = (30,30,30)$ และ $(25,30,35)$ $\phi = 2.5$ $n = (30,30,30)$ และ $\phi = 11.5$ $n = (25,30,35)$

4.2.2 ระดับนัยสำคัญ 0.05

ตัวสถิติที่ให้กำลังการทดสอบสูงสุดแสดงดังตารางที่ 6 และสรุปผลได้ดังนี้

(1) เมื่อประชากรมีการแจกแจงปกติ สถิติทดสอบ W และ M มีกำลังการทดสอบสูงที่สุดทุกกรณี ส่วนสถิติทดสอบ F มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 11.5$ $n = (5,10,15)$

(2) เมื่อประชากรมีการแจกแจงแกมมา สถิติทดสอบ W มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 11.5$ $n = (10,10,10)$ ส่วน

สถิติทดสอบ F มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 2.5$ $n = (5,10,15)$ และ $\phi = 11.5$ $n = (10,10,10)$ ส่วนสถิติทดสอบ M มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี ยกเว้นกรณี $\phi = 11.5$ $n = (10,10,10)$ และ $(25,30,35)$ และส่วนสถิติทดสอบ PB มีกำลังการทดสอบสูงที่สุด เมื่อ $\phi = 11.5$ $n = (30,30,30)$

ตารางที่ 5 สถิติทดสอบที่ให้กำลังการทดสอบสูงที่สุดของการแจกแจงปกติและแกมมา ที่ระดับนัยสำคัญ 0.01

การแจกแจง	ϕ	ขนาดตัวอย่าง (n_1, n_2, n_3)					
		10,10,10	30,30,30	50,50,50	5,10,15	25,30,35	45,50,55
ปกติ	0.5	F,W,M	F,W,M,PB	F,W,M	F	F,PB	F,W,M
	2.5	F,W,M	F,W,M	W,M	F,W,M	F,W,M	F,W,M
	11.5	F	F,W,M	W,M	W,M	F,W,M	F,W,M
แกมมา	0.5	F,W,M	F,W,M,PB	F,W,M	F,W,M	F,W,M,PB	F
	2.5	F	W,PB	F,W,M	F,W	F,W,M	F,W,M
	11.5	-	-	-	F	F,PB	F

ตารางที่ 6 สถิติทดสอบที่ให้กำลังการทดสอบสูงที่สุดของการแจกแจงปกติและแกมมา ที่ระดับนัยสำคัญ 0.05

การแจกแจง	ϕ	ขนาดตัวอย่าง (n_1, n_2, n_3)					
		10,10,10	30,30,30	50,50,50	5,10,15	25,30,35	45,50,55
ปกติ	0.5	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M
	2.5	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M
	11.5	F,W,M	F,W,M	F,W,M	W,M	F,W,M	F,W,M
แกมมา	0.5	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M	F,W,M
	2.5	F,W,M	F,W,M	F,W,M	W,M	F,W,M	F,W,M
	11.5	-	F,W,M,PB	F,W,M	F,W,M	F,W	F,W,M

5. สรุป

ในงานวิจัยเมื่อผู้วิจัยต้องการทดสอบค่าเฉลี่ยของประชากรสามกลุ่ม แต่ค่าความแปรปรวนของประชากรไม่เท่ากัน กรณีที่ประชากรมีการแจกแจงปกติควรใช้สถิติทดสอบ Welch (W) และสถิติทดสอบมาราสควิลโล (M) มากกว่าการวิเคราะห์ความแปรปรวน

ด้วยสถิติทดสอบเอฟ (F) โดยเฉพาะเมื่อขนาดตัวอย่างมีจำนวนไม่มากและไม่เท่ากัน เมื่ออัตราส่วนความแตกต่างของความแปรปรวนอยู่ในระดับปานกลางและมาก เนื่องจากสถิติทดสอบทดสอบ Welch (W) และสถิติทดสอบมาราสควิลโล (M) ได้มีการถ่วงน้ำหนักให้มีความแข็งแกร่งมากกว่าการวิเคราะห์ความแปรปรวนด้วย

สถิติทดสอบเอฟ (F) กรณีที่ประชากรไม่ใช้การแจกแจงปกติผู้วิจัยควรเลือกใช้การวิเคราะห์ความแปรปรวนด้วยสถิติทดสอบเอฟ (F) สถิติทดสอบเวลช์ (W) และสถิติทดสอบมาราสคิลโล (M) เมื่ออัตราส่วนความแตกต่างของความแปรปรวนอยู่ในระดับน้อยและปานกลาง เนื่องจากระดับนัยสำคัญเพิ่มขึ้น ทำให้สถิติทดสอบดังกล่าวนี้สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้มากขึ้น และกำลังการทดสอบเข้าใกล้ 1 ส่วนสถิติทดสอบวิธีบูตสเตรปแบบใช้พารามิเตอร์ (PB) ควรเลือกใช้ในกรณีที่ขนาดตัวอย่างใกล้เคียง 30

6. รายการอ้างอิง

- [1] พลชาติ หาญอนุรักษ์, 2549, การเปรียบเทียบตัวสถิติทดสอบเวลช์ บราวน์-ฟอร์ลิตี และมาราสคิลโล สำหรับความเท่ากันของค่าเฉลี่ย เมื่อความแปรปรวนของประชากรไม่เท่ากัน, วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ, 71 น.
- [2] Krishnamoorthy, K., Lu, F. and Mathew, T., 2007, A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models, *Comput. Stat. Data Anal.* 51: 5731–5742.
- [3] Yigit, E. and Gokpinar, F., 2010, A simulation study on tests for one – way ANOVA under the unequal variance assumption, *Commun. Fac. Sci. Univ. Ank. Ser. A1* 59(2): 15-34.
- [4] Montgomery, D.C., 2013, *Design and Analysis of Experiments*, John Wiley and Sons, Inc., New York.
- [5] Welch, B.L., 1951, On the comparison of several mean values: An alternative approach, *Biometrika* 38: 330-336.
- [6] Marascuilo, L.A., 1971, *Statistical Methods for Behavioural Science Research*, McGraw-Hill Book Co., New York, 578 p.
- [7] Games, P.A., Winkler, H.B., and Probert, D.A., 1972, Robust tests for homogeneity of variance, *Educ. Psychol. Meas.* 32: 887-909.
- [8] Bradley, J.V., 1978, Robustness?, *Br. J. Math. Stat. Psychol.* 31: 144-152.