การเปรียบเทียบการทดสอบค่าเฉลี่ยภายใต้ความแปรปรวนวิวิธพันธ์ในการทดลองสุ่มสมบูรณ์

A Comparison of Tests for Equality of Means under Heterogeneity of Variances in Completely Randomized Design

ปรีชญา หะสะเล็ม* และ จิราวัลย์ จิตรถเวช

Prechaya Hasalem* and Jirawan Jitthavech

สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Department of Statistics, Graduate School of Applied Statistics, National Institute of Development Administration

Received : 27 September 2017

Accepted : 21 December 2017

Published online : 9 January 2018

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 คือค่าความน่าจะเป็นของจำนวนครั้งที่ปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นจริง และร้อยละของการตัดสินใจถูกต้อง คือร้อยละของจำนวนครั้งที่ปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานทางเลือกเป็นจริงของการทดสอบความเท่ากันของค่าเฉลี่ยประชากร 3 ชุด โดยใช้การทดสอบเอฟ การทดสอบเวลช์ และการทดสอบเอฟวางนัยทั่วไป โดยจำลองข้อมูลในแผนแบบการทดลองแบบสุ่มสมบูรณ์ กำหนดปัจจัยการทดลองมีผลกระทบแบบตรึง กรณีที่ประชากรมีการแจกแจงปกติ กำหนดความแปรปรวนของประชากรเป็น 6 กรณี ขนาดตัวอย่างที่สุ่มจากประชากรแต่ละชุดมีขนาดเท่ากัน $n = 10, 15, 30, 50, 100, 200$ และกระทำการทดสอบซ้ำจำนวน 1000 รอบในแต่ละกรณี ผลการวิจัยพบว่ากรณีความแปรปรวนของประชากรเท่ากัน และมีค่าความแปรปรวนน้อยเท่ากับ 1 ความแปรปรวนปานกลางเท่ากับ 25 และความแปรปรวนมากเท่ากับ 100 พบว่าการทดสอบเอฟสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ และให้ร้อยละของการตัดสินใจถูกต้องมากที่สุด และกรณีความแปรปรวนของประชากรแตกต่างกันน้อย $(\sigma_1^2 = 1,\ \sigma_2^2 = 9,\ \sigma_3^2 = 25)$ ความแปรปรวนของประชากรแตกต่างกันปานกลาง $(\sigma_1^2 = 4,\ \sigma_2^2 = 36,\ \sigma_3^2 = 100)$ และความแปรปรวนของประชากรแตกต่างกันแตกต่างกันมาก $(\sigma_1^2 = 9,\ \sigma_2^2 = 81,\ \sigma_3^2 = 225)$ พบว่าการทดสอบเอฟวางนัยทั่วไปสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ และให้ร้อยละของการตัดสินใจถูกต้องมากที่สุด นอกจากนั้นพบว่าค่าร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟ การทดสอบเวลช์ และการทดสอบเอฟวางนัยทั่วไปมีค่าใกล้เคียงกันและเพิ่มขึ้นเมื่อขนาดตัวอย่างใหญ่ขึ้นในทุกกรณี

คำสำคัญ : แผนแบบการทดลองแบบสุ่มสมบูรณ์ ผลกระทบแบบตรึง การทดสอบเอฟ การทดสอบเวลช์ การทดสอบเอฟวางนัยทั่วไป

* **Correspondent author. E-mail** : prechyay.h@gmail.com

Abstract

The objective of the research is to compare the ability to control probability of type I error of testing the equality of means from a completely randomized design (CRD) by F test, Welch's test and Generalized F test when the population variance is heterogeneous. The ability to control probability of type I error is measured by the probability of the number of "reject" when the null hypothesis is true and the percentage of accuracy is measured by the number of "reject" when the alternative hypothesis is true. The data are generated from the fixed effect CRD simulation of normal distributed error population with zero mean and six cases of variance. The sample sizes are $n = 10, 15, 30, 50, 100, 200$ and the replication in each case is 1,000. The research results can be summarized as follows. In the cases of homogeneous variance with small variance equal to 1, medium variance equal to 25 and large variance equal to 100, it is found that the F test yields the highest result. In the cases of heterogeneous variance with small variation in variance $(\sigma_1^2 = 1,\ \sigma_2^2 = 9,\ \sigma_3^2 = 25)$, medium variation in variance $(\sigma_1^2 = 4,\ \sigma_2^2 = 36,\ \sigma_3^2 = 100)$ and large variation in variance $(\sigma_1^2 = 9,\ \sigma_2^2 = 81,\ \sigma_3^2 = 225)$, it is found that the generalized F test yields the highest result. However, the percentage of accuracy by the F test, Welch's test and generalized F test are nearly the same and increases as the sample size increases in all cases.

Keyword: Completely Randomized Design, Fixed Effect, F test, Welch's test, Generalized F test

บทนำ

ในการศึกษางานวิจัยต่าง ๆ ยกตัวอย่างเช่นในการวิจัยเชิงทดลอง นอกจากผู้วิจัยต้องมีความรู้ความเข้าใจในเรื่องที่ทำการศึกษาแล้ว ผู้วิจัยยังต้องมีความสามารถในการออกแบบการทดลอง และต้องอาศัยความรู้ทางด้านสถิติเพื่อช่วยในการวิเคราะห์ข้อมูลเพื่อหาข้อสรุปในงานวิจัยตามวัตถุประสงค์ที่วางไว้ ความถูกต้องในการสรุปผลจะมากน้อยเพียงใดนั้น ขึ้นอยู่กับวิธีทดสอบที่เลือกว่าเหมาะสมกับลักษณะของข้อมูลที่ทำการวิจัยหรือไม่ วิธีการทดสอบต่าง ๆ นั้นจะมีข้อสมมุติเบื้องต้น (Assumptions) ซึ่งเป็นข้อกำหนดเกี่ยวกับลักษณะต่าง ๆ ของประชากรที่จะนำมาใช้ในการวิเคราะห์ ดังนั้นการเลือกใช้วิธีการทดสอบที่เหมาะสมกับลักษณะของประชากรจะทำให้ผลสรุปที่ได้จากการวิจัยเชิงทดลองมีความถูกต้องและน่าเชื่อถือ

สำหรับงานวิจัยเชิงทดลองมักจะสนใจการทดสอบสมมุติฐานเกี่ยวกับความเท่ากันของค่าเฉลี่ยหลายประชากร การทดสอบที่เหมาะสมคือ การทดสอบเอฟ (F-test) โดยใช้การวิเคราะห์ความแปรปรวน (Analysis of variance : ANOVA) ภายใต้ข้อสมมุติเบื้องต้นที่การสุ่มตัวอย่างต้องมาจากแต่ละประชากรที่มีการแจกแจงปกติและเป็นอิสระต่อกัน แต่ละประชากรต้องมีความแปรปรวน (Variance) เท่ากัน ในทางปฏิบัติข้อมูลที่ได้จากการทดลองอาจมาจากประชากรที่อาจไม่เป็นไปตามข้อสมมุติเบื้องต้นดังกล่าวบางประการที่กำหนดไว้ โดยเฉพาะเมื่อความแปรปรวนของแต่ละประชากรไม่เท่ากัน (ความแปรปรวนวิวิธพันธ์) การใช้การทดสอบเอฟ อาจให้ข้อสรุปที่ไม่ถูกต้อง ในกรณีดังกล่าวอาจทำการแปลงข้อมูลของประชากร (Transformation of data) เพื่อแก้ปัญหาความแปรปรวนวิวิธพันธ์ ก่อนที่จะนำข้อมูลไปวิเคราะห์ความแปรปรวน หรือ อาจใช้การทดสอบค่าเฉลี่ยของหลายประชากรในการวิเคราะห์ความแปรปรวน ที่ได้พัฒนาขึ้นเพื่อใช้กับข้อมูลที่ไม่เป็นไปตามข้อสมมุติเบื้องต้น คือ ประชากรมีการแจกแจงปกติ แต่มีค่าความแปรปรวนของประชากรวิวิธพันธ์

จากปัญหาข้างต้น ทำให้นักสถิติหลายท่านได้พัฒนาและคิดค้น การทดสอบค่าเฉลี่ยของหลายประชากรที่เหมาะสมและมีสมบัติที่ดีเพื่อใช้ในการวิเคราะห์ข้อมูล ในกรณีข้อมูลที่นำมาทดสอบมีความแปรปรวนวิวิธพันธ์ เช่น การทดสอบเวลช์ (Welch, 1951) การทดสอบของสก๊อทและสมิท (Scott and Smith, 1971) การทดสอบของบราวน์ฟอร์ซิตี (Brown and Forsythe, 1974) การทดสอบของเชนและเชน (Chen and Chen, 1998, 2001) คือ การวิเคราะห์ความแปรปรวนขั้นเดียว (One-stage ANOVA test) และการทดสอบพิสัยขั้นเดียว (One-stage range test) การทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิ (Weerahandi, 1995, 2004, 2013) การทดสอบเอฟวางนัยทั่วไปของซูและวัง (Xu and Wang, 2008a, 2008b) และการทดสอบบูตสแตรปแบบใช้พารามิเตอร์ของคริสช์นามอร์ที (Krishnamoorthy *et al.,* 2007) เป็นต้น

งานวิจัยนี้ศึกษาเปรียบเทียบการทดสอบความเท่ากันของค่าเฉลี่ยประชากรที่มีมากกว่า 2 กลุ่ม เมื่อประชากรมีการแจกแจงปกติ และความแปรปรวนวิวิธพันธ์ สำหรับการทดลองสุ่มสมบูรณ์ (Completely Randomized Design: CRD) โดยเปรียบเทียบการทดสอบ 3 วิธี คือ

1. การทดสอบเอฟ ใช้สำหรับทดสอบความเท่ากันของค่าเฉลี่ยของประชากร 2 กลุ่ม หรือมากกว่า 2 กลุ่มขึ้นไปที่มีการแจกแจงปกติ ในกรณีที่มีความแปรปรวนเท่ากัน สถิติทดสอบเอฟภายใต้สมมุติฐานหลักเป็นจริง คำนวณได้จาก

$$F = \frac{\frac{SSTt}{\sigma^2(k-1)}}{\frac{SSE}{\sigma^2(N-k)}} = \frac{MSTt}{MSE}$$ มีการแจกแจงเอฟด้วยองศาเสรี $k-1$ และ $N-k$

เมื่อ SST เป็นผลบวกกำลังสองของยอดรวม, $SST = \sum_i^k \sum_j^{n_i} \left(y_{ij} - \bar{y}_{..}\right)^2$ ด้วยองศาเสรี $N-1$

$SSTt$ เป็นผลบวกกำลังสองระหว่างทรีตเมนต์, $SSTt = \sum_i^k n_i \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2$ ด้วยองศาเสรี $k-1$

และ SSE เป็นผลบวกกำลังสองของความคลาดเคลื่อน, $SSE = \sum_i^k \sum_j^{n_i} \left(y_{ij} - \bar{y}_{i.}\right)^2$ ด้วยองศาเสรี $N-k$

และ $MSE = \frac{SSE}{N-k}$ และ $MSTt = \frac{SSTt}{k-1}$

2. การทดสอบเวลช์ เสนอโดย Welch (1947, 1951) ซึ่งได้ดัดแปลงมาจากการทดสอบของ Cochran เพื่อนำมาใช้ทดสอบความเท่ากันของค่าเฉลี่ย k ประชากรที่มีการแจกแจงปกติ เมื่อความแปรปรวนของประชากรวิวิธพันธ์ สถิติทดสอบเวลช์ภายใต้สมมุติฐานหลักเป็นจริง คำนวณได้จาก

$$W = \frac{\sum_{i=1}^k w_i[(\bar{x}_i - \bar{x})^2/(k-1)]}{1+\frac{2(k-2)}{k^2-1}\sum_{i=1}^k \frac{1}{n_i-1}\left(1-\frac{w_i}{w}\right)^2}$$ มีการแจกแจงเอฟด้วยองศาเสรี $k-1$ และ f

เมื่อ $w_i = \frac{n_i}{S_i^2}, w = \sum_{j=1}^k w_i$ และ $\bar{x} = \frac{\sum_{j=1}^k w_i \bar{x}_i}{w}$

n_i คือ จำนวนซ้ำของหน่วยทดลองในทรีตเมนต์ที่ i

S_i^2 คือ ความแปรปรวนของค่าสังเกตในทรีตเมนต์ที่ i

และ $f = \frac{1}{\frac{3}{k^2-1}\sum_{i=1}^k \frac{1}{n_i-1\left(1-\frac{w_i}{w}\right)^2}}$

3. การทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิ เสนอโดย (Weerahandi, 1995, 2004, 2013) ใช้ในการทดสอบความเท่ากันของค่าเฉลี่ย k ประชากรที่มีการแจกแจงปกติ เมื่อความแปรปรวนของประชากรวิวิธพันธ์ สถิติทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิภายใต้สมมุติฐานหลักเป็นจริง คำนวณได้จาก

$G = \tilde{S}_B\left(\frac{n_1 S_1^2}{Y_1}, \frac{n_2 S_2^2}{Y_2}, \ldots, \frac{n_k S_k^2}{Y_k}\right)$ มีการแจกแจงไคกำลังสองด้วยองศาเสรี $k-1$

เมื่อ $\bar{y}_\sigma = \frac{\sum_{i=1}^{k} \frac{n_i \bar{y}_i}{S_i^2}}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}}, \tilde{S}_i = \frac{n_i}{S_i^2}(\bar{y}_i - \bar{y}_\sigma)^2, Y_i = \frac{n_i S_i^2}{\tilde{S}_i}$ และ $\tilde{S}_B = \sum_{i=1}^{k} \frac{n_i}{S_i^2}(\bar{y}_i - \bar{y}_\sigma)^2$

n_i คือ จำนวนซ้ำของหน่วยทดลองในทรีตเมนต์ที่ i

$\bar{y}_i$ คือ ค่าเฉลี่ยของค่าสังเกตในทรีตเมนต์ที่ i

S_i^2 คือ ความแปรปรวนของค่าสังเกตในทรีตเมนต์ที่ i

เนื่องจากการทดสอบเอฟเป็นการทดสอบที่ใช้กันโดยทั่วไป ในการทดสอบความเท่ากันของค่าเฉลี่ยประชากรที่มีมากกว่า 2 กลุ่ม กรณีที่ความแปรปรวนของประชากรเท่ากัน การทดสอบเวลช์เป็นการทดสอบที่นิยมใช้ทดสอบความเท่ากันของค่าเฉลี่ยประชากรที่มีมากกว่า 2 กลุ่ม กรณีที่ความแปรปรวนของประชากรวิวิธพันธ์ และการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิเป็นการทดสอบที่ได้มีการคิดค้นขึ้นมาใหม่สำหรับทดสอบความเท่ากันของค่าเฉลี่ยประชากรที่มีมากกว่า 2 กลุ่ม กรณีที่ความแปรปรวนของประชากรวิวิธพันธ์ ผู้วิจัยจึงเลือกเปรียบเทียบการทดสอบทั้ง 3 วิธี โดยใช้การจำลองข้อมูลด้วยเทคนิคการจำลองมอนติคาร์โล (Monte Carlo Simulation Technique) ภายใต้สถานการณ์ที่กำหนด

วิธีดำเนินการวิจัย

การศึกษาใช้วิธีการจำลองข้อมูลเพื่อพิจารณาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อสมมุติฐานหลักเป็นจริงเปรียบเทียบกับเกณฑ์ที่กำหนด และร้อยละของการตัดสินใจถูกต้องเมื่อสมมุติฐานทางเลือกเป็นจริง ในกรณีความแปรปรวนเท่ากันหรือความแปรปรวนวิวิธพันธ์ โดยใช้โปรแกรมสำเร็จรูป SAS เวอร์ชัน 9.4 ในการจำลองข้อมูลและทดสอบสมมุติฐาน มีขั้นตอนการดำเนินงาน ดังนี้

1) กำหนดแผนแบบการทดลองแบบสุ่มสมบูรณ์ ให้ปัจจัยการทดลองมีผลกระทบแบบตรึง จำนวนซ้ำของหน่วยทดลองในแต่ละทรีตเมนต์เท่ากัน มีตัวแบบดังนี้

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ; \quad i = 1,2,\ldots,k, j = 1,2,\ldots,n$$

y_{ij} คือ ตัวแปรตอบสนองของหน่วยทดลอง j ที่ได้รับทรีตเมนต์ที่ i

μ คือ พารามิเตอร์ แทนค่าเฉลี่ยรวม

τ_i คือ พารามิเตอร์ แทนอิทธิพลจากทรีตเมนต์ i โดยที่ $\sum_{i=1}^{k} \tau_i = 0$

ε_{ij} คือ ค่าความคลาดเคลื่อนของหน่วยทดลอง j ที่ได้รับทรีตเมนต์ i

k คือ จำนวนทรีตเมนต์ในการทดลอง

n คือ จำนวนซ้ำของหน่วยทดลองในแต่ละทรีตเมนต์

ข้อสมมุติเบื้องต้น ความคลาดเคลื่อนเชิงสุ่ม (ε_{ij}) มีการแจกแจงปกติและเป็นอิสระกัน มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ σ_i^2 นั่นคือ $\varepsilon_{ij} \sim NID(0, \sigma_i^2)$ เมื่อ $i = 1,2,\dots,k, j = 1,2,\dots,n$ ทำให้ $y_{ij} \sim NID(\mu + \tau_i, \sigma_i^2)$

2) กำหนดจำนวนทรีตเมนต์ $k = 3$ และจำนวนซ้ำของหน่วยทดลองในแต่ละทรีตเมนต์เท่ากันทุกทรีตเมนต์ $n = 10, 15, 30, 50, 100, 200$

3) กำหนดค่าเฉลี่ยรวมคงที่ $\mu = 50$ อิทธิพลของทรีตเมนต์ (τ_i) 2 กรณี

3.1) กรณีค่าเฉลี่ยเท่ากัน เมื่อ $\tau_i = 0$ ทุกค่าของ i

3.2) กรณีค่าเฉลี่ยไม่เท่ากัน เมื่อ $\tau_1 = 4, \tau_2 = 2, \tau_3 = -6$

4) กำหนดการแจกแจงของความคลาดเคลื่อนเชิงสุ่ม ε_{ij} ให้มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนตามที่กำหนด

4.1) เมื่อความแปรปรวนเท่ากัน ทำการศึกษา 3 กรณีคือ กรณีความแปรปรวนน้อย ทุกประชากรมีความแปรปรวนเท่ากับ 1 กรณีความแปรปรวนปานกลาง ทุกประชากรมีความแปรปรวนเท่ากับ 25 และกรณีความแปรปรวนมาก ทุกประชากรมีความแปรปรวนเท่ากับ 100

4.2) เมื่อความแปรปรวนแตกต่างกัน ทำการศึกษา 3 กรณีคือ กรณีความแปรปรวนแตกต่างกันน้อย $\sigma_1^2 = 1,\ \sigma_2^2 = 9,\ \sigma_3^2 = 25$ กรณีความแปรปรวนแตกต่างกันปานกลาง $\sigma_1^2 = 4,\ \sigma_2^2 = 36,\ \sigma_3^2 = 100$ และ กรณีความแปรปรวนแตกต่างกันมาก $\sigma_1^2 = 9,\ \sigma_2^2 = 81,\ \sigma_3^2 = 225$

5) สุ่มตัวอย่างจากข้อมูลที่จำลองตามจำนวนซ้ำของหน่วยทดลองในแต่ละทรีตเมนต์เท่ากัน

6) ทดสอบความเท่ากันของความแปรปรวนของตัวแปรตอบสนอง โดยกำหนดระดับนัยสำคัญเท่ากับ 0.05

7) ทดสอบความเท่ากันของค่าเฉลี่ยโดยมีสมมุติฐานดังนี้

$H_o: \mu_1 = \mu_2 = \mu_3$ เทียบกับ $H_a: \mu_i \neq \mu_j$ อย่างน้อย 1 ค่าของ i ไม่เท่ากับ j

โดยใช้การทดสอบ 3 วิธี คือ การทดสอบเอฟ การทดสอบเวลช์ และการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิ

8) กระทำซ้ำ 1,000 รอบในแต่ละสถานการณ์

9) เปรียบเทียบค่าของตัวการทดสอบที่คำนวณได้กับค่าวิกฤตโดยใช้ระดับนัยสำคัญ $\alpha = 0.05$ ทำซ้ำ ๆ กันจนครบจำนวน 1,000 ครั้ง ในแต่ละสถานการณ์ คำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และร้อยละของการตัดสินใจถูกต้อง

เกณฑ์ในการตัดสินใจเมื่อสมมุติฐานหลักเป็นจริงคือ

$$\text{ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1} = \frac{\text{จำนวนครั้งที่ปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นจริง}}{\text{จำนวนครั้งที่ทำการทดสอบ}}$$

เกณฑ์ในการตัดสินใจเมื่อสมมุติฐานทางเลือกเป็นจริงคือ

$$\text{ร้อยละของการตัดสินใจถูกต้อง}(\text{อำนาจการทดสอบ}) = \frac{\text{จำนวนครั้งที่ปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานทางเลือกเป็นจริง}}{\text{จำนวนครั้งที่ทำการทดสอบ}} \times 100$$

ทำการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กับเกณฑ์ของ Cochran (Cochran, 1954) หากอยู่ในเกณฑ์ คือ เมื่อกำหนดระดับนัยสำคัญ $\alpha = 0.05$ และผลการคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อน

ประเภทที่ 1 อยู่ระหว่าง 0.04-0.06 จะสรุปได้ว่าการทดสอบนั้นสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ หรือเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กับเกณฑ์ของ Bradley (Bradley, 1978) หากอยู่ในเกณฑ์ คือ เมื่อกำหนดระดับนัยสำคัญ $\alpha = 0.05$ และผลการคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 อยู่ระหว่าง 0.025-0.075 จะสรุปได้ว่าการทดสอบนั้นสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ และคำนวณร้อยละของการตัดสินใจถูกต้องของการทดสอบทั้ง 3 วิธี ในกรณีที่การทดสอบนั้นสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ เพื่อพิจารณาว่าการทดสอบใดมีร้อยละของการตัดสินใจถูกต้องมากที่สุด กล่าวคือ พิจารณาร้อยละของจำนวนครั้งที่ปฏิเสธสมมุติฐานหลักมากที่สุดเมื่อสมมุติฐานทางเลือกเป็นจริง

ผลการวิจัยและวิจารณ์ผล

ตารางที่ 1 กรณีความแปรปรวนประชากรเท่ากันและความแปรปรวนมีค่าน้อย $\sigma_i^2 = 1; i = 1, 2, 3$ เมื่อสมมุติฐานหลักเป็นจริงและตัวอย่างขนาดเล็กพบว่าการทดสอบเอฟและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran และเมื่อตัวอย่างขนาดใหญ่พบว่าการทดสอบเอฟและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟและการทดสอบเวลช์ พบว่ามีค่าเท่ากันคือมีค่าเท่ากับร้อยละ 100.00

กรณีความแปรปรวนประชากรเท่ากันและความแปรปรวนมีค่าปานกลาง $\sigma_i^2 = 25; i = 1, 2, 3$ เมื่อสมมุติฐานหลักเป็นจริงและมีขนาดตัวอย่าง 10 15 30 และ 50 พบว่าการทดสอบเอฟและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 100 และ 200 พบว่าการทดสอบเอฟและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟและการทดสอบเวลช์ พบว่ามีค่าใกล้เคียงกัน และเมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าร้อยละของการตัดสินใจถูกต้องของทั้งการทดสอบเอฟ และการทดสอบเวลช์มีค่าเท่ากันคือมีค่าเท่ากับร้อยละ 100.00

กรณีความแปรปรวนประชากรเท่ากันและความแปรปรวนมีค่ามาก $\sigma_i^2 = 100; i = 1, 2, 3$ เมื่อสมมุติฐานหลักเป็นจริงและมีขนาดตัวอย่าง 10 15 30 และ 50 พบว่าการทดสอบเอฟและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 100 และ 200 พบว่าการทดสอบเอฟ การทดสอบเวลช์และการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟ การทดสอบเวลช์และการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิ พบว่าการทดสอบเอฟให้ร้อยละของการตัดสินใจถูกต้อง

มากที่สุดเมื่อมีขนาดตัวอย่าง 10 15 และ 30 การทดสอบเอฟและการทดสอบเวลช์ให้ร้อยละของการตัดสินใจถูกต้องไม่แตกต่างกันเมื่อมีขนาดตัวอย่าง 50 และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 100 และ 200 ค่าร้อยละของการตัดสินใจถูกต้องของทั้ง 3 การทดสอบมีค่าเท่ากันคือมีค่าเท่ากับร้อยละ 100.00

ตารางที่ 1 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และร้อยละของการตัดสินใจถูกต้อง ของการทดสอบทั้ง 3 วิธี เมื่อความแปรปรวนเท่ากันทั้ง 3 ประชากร กรณี $\sigma^2 = 1, 25$ และ 100 จำแนกตามขนาดตัวอย่าง

σ_i^2	n_i	ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กรณีค่าเฉลี่ยเท่ากัน			ร้อยละของการตัดสินใจถูกต้อง กรณีค่าเฉลี่ยไม่เท่ากัน		
		เอฟ	เวลช์	วีราแฮนดิ	เอฟ	เวลช์	วีราแฮนดิ [c]
1	10	0.049	0.059	0.077 [a]	100.00	100.00	-
	15	0.043	0.044	0.071 [a]	100.00	100.00	-
	30	0.044	0.051	0.083 [a]	100.00	100.00	-
	50	0.038	0.041	0.085 [b]	100.00	100.00	-
	100	0.035	0.036	0.077 [b]	100.00	100.00	-
	200	0.031	0.037	0.079 [b]	100.00	100.00	-
25	10	0.053	0.054	0.067 [a]	99.50	99.10	-
	15	0.049	0.049	0.064 [a]	99.90	99.90	-
	30	0.048	0.047	0.067 [a]	100.00	100.00	-
	50	0.041	0.041	0.071 [a]	100.00	100.00	-
	100	0.035	0.036	0.081 [b]	100.00	100.00	-
	200	0.035	0.035	0.079 [b]	100.00	100.00	-
100	10	0.054	0.054	0.101 [a]	58.60	53.10	-
	15	0.044	0.049	0.079 [a]	77.60	73.90	-
	30	0.043	0.049	0.071 [a]	98.70	97.60	-
	50	0.040	0.044	0.074 [a]	99.90	99.90	-
	100	0.037	0.040	0.056	100.00	100.00	100.00
	200	0.036	0.036	0.059	100.00	100.00	100.00

a: ไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran

b: ไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley

c: คำนวณค่าร้อยละของการตัดสินใจถูกต้องเมื่อการทดสอบผ่านเกณฑ์ของ Cochran หรือ Bradley

ตารางที่ 2 กรณีความแปรปรวนประชากรวิวิธพันธ์แต่แตกต่างกันน้อย $\sigma_1^2 = 1,\ \sigma_2^2 = 9,\ \sigma_3^2 = 25$ เมื่อสมมุติฐานหลักเป็นจริงและมีขนาดตัวอย่าง 10 15 30 และ 50 พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 100 และ 200 พบว่าทั้ง 3 การทดสอบสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิ การทดสอบเวลช์ และการทดสอบเอฟ พบว่ามีร้อยละของการตัดสินใจถูกต้องใกล้เคียงกัน

กรณีความแปรปรวนประชากรวิวิธพันธ์แต่แตกต่างกันปานกลาง $\sigma_1^2 = 4,\ \sigma_2^2 = 36,\ \sigma_3^2 = 100$ เมื่อสมมุติฐานหลักเป็นจริงและมีขนาดตัวอย่าง 10 และ 15 พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 30 50 100 และ 200 พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้อง พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิมีร้อยละของการตัดสินใจถูกต้องมากที่สุดเมื่อขนาดตัวอย่างเท่ากับ 10 และ 15 และเมื่อขนาดตัวอย่างเพิ่มขึ้น ร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์มีค่าเท่ากัน

กรณีความแปรปรวนประชากรวิวิธพันธ์แต่แตกต่างกันมาก $\sigma_1^2 = 9,\ \sigma_2^2 = 81,\ \sigma_3^2 = 225$ เมื่อสมมุติฐานหลักเป็นจริงและมีขนาดตัวอย่าง 10 15 30 และ 50 พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Cochran และเมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 100 และ 200 พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ ส่วนการทดสอบเอฟไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley และเมื่อเปรียบเทียบค่าร้อยละของการตัดสินใจถูกต้อง พบว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิมีร้อยละของการตัดสินใจถูกต้องมากที่สุดเมื่อขนาดตัวอย่างเท่ากับ 10 15 และ 30 และเมื่อขนาดตัวอย่างเพิ่มขึ้น ร้อยละของการตัดสินใจถูกต้องของการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิและการทดสอบเวลช์มีค่าใกล้เคียงกัน

ตารางที่ 2 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และร้อยละของการตัดสินใจถูกต้อง ของการทดสอบทั้ง 3 วิธี เมื่อความแปรปรวนประชากรแตกต่างกัน กรณี $\sigma_1^2 = 1,\ \sigma_2^2 = 9,\ \sigma_3^2 = 25$, $\sigma_1^2 = 4,\ \sigma_2^2 = 36,\ \sigma_3^2 = 100$ และ $\sigma_1^2 = 9,\ \sigma_2^2 = 81,\ \sigma_3^2 = 225$ จำแนกตามขนาดตัวอย่าง

σ_i^2	n_i	ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กรณีค่าเฉลี่ยเท่ากัน			ร้อยละของการตัดสินใจถูกต้อง กรณีค่าเฉลี่ยไม่เท่ากัน		
		เอฟ	เวลช์	วีราแฮนดิ	เอฟ [c]	เวลช์ [c]	วีราแฮนดิ [c]
$\sigma_1^2 = 1, \sigma_2^2 = 9$ และ $\sigma_3^2 = 25$	10	0.084 [a]	0.043	0.039	-	99.90	100.00
	15	0.087 [a]	0.049	0.034	-	100.00	100.00
	30	0.081 [b]	0.034	0.028	-	100.00	100.00
	50	0.078 [b]	0.037	0.030	-	100.00	100.00
	100	0.075 [b]	0.035	0.026	98.70	100.00	100.00
	200	0.071 [b]	0.038	0.025	99.20	100.00	100.00
$\sigma_1^2 = 4, \sigma_2^2 = 36$ และ $\sigma_3^2 = 100$	10	0.106 [a]	0.054	0.046	-	74.30	85.60
	15	0.101 [a]	0.055	0.040	-	93.10	96.50
	30	0.095 [b]	0.050	0.037	-	99.90	100.00
	50	0.082 [b]	0.047	0.033	-	100.00	100.00
	100	0.078 [b]	0.043	0.030	-	100.00	100.00
	200	0.078 [b]	0.036	0.028	-	100.00	100.00
$\sigma_1^2 = 9, \sigma_2^2 = 81$ และ $\sigma_3^2 = 225$	10	0.142 [a]	0.056	0.054	-	39.50	52.50
	15	0.121 [a]	0.056	0.051	-	58.60	68.60
	30	0.096 [a]	0.053	0.049	-	91.90	94.50
	50	0.082 [a]	0.048	0.042	-	99.60	99.70
	100	0.077 [b]	0.045	0.038	-	100.00	100.00
	200	0.076 [b]	0.038	0.035	-	100.00	100.00

a: **ไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ** Cochran

b: **ไม่สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ** Bradley

c: **คำนวณค่าร้อยละของการตัดสินใจถูกต้องเมื่อการทดสอบผ่านเกณฑ์ของ** Cochran **หรือ** Bradley

สรุปผลการวิจัย

กรณีความแปรปรวนประชากรเท่ากันเมื่อมีความแปรปรวนน้อย, ความแปรปรวนปานกลาง และความแปรปรวนมาก การทดสอบเอฟเป็นการทดสอบที่เหมาะสมที่สุดในการทดสอบค่าเฉลี่ยของประชากร สามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อสมมุติฐานหลักเป็นจริง และให้ร้อยละของการตัดสินใจถูกต้องมากที่สุด เมื่อสมมุติฐานทางเลือกเป็นจริง

ส่วนในกรณีความแปรปรวนประชากรแตกต่างกันน้อย, ความแปรปรวนประชากรแตกต่างกันปานกลาง และความแปรปรวนประชากรแตกต่างกันมาก ในการทดสอบค่าเฉลี่ยของประชากรเมื่อสมมุติฐานหลักเป็นจริง การทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ และเมื่อสมมุติฐานทางเลือกเป็นจริง การทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิให้ร้อยละของการตัดสินใจถูกต้องมากที่สุด จึงเป็นการทดสอบที่เหมาะสม

ส่วนการทดสอบเวลช์นั้นสามารถควบคุมความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อสมมติฐานหลักเป็นจริง ทั้งในกรณีที่ความแปรปรวนประชากรเท่ากันและความแปรปรวนประชากรวิวิธพันธ์ แต่ให้ร้อยละของการตัดสินใจถูกต้องน้อยกว่าการทดสอบเอฟในกรณีความแปรปรวนประชากรเท่ากัน และให้ร้อยละของการตัดสินใจถูกต้องน้อยกว่าการทดสอบเอฟวางนัยทั่วไปของวีราแฮนดิในกรณีความแปรปรวนประชากรไม่เท่ากัน เมื่อสมมุติฐานทางเลือกเป็นจริง

กิตติกรรมประกาศ

ขอขอบพระคุณ รองศาสตราจารย์ ดร.วิชิต หล่อจีระชุณห์กุล ที่กรุณาให้คำปรึกษา แนะนำ แก้ไขข้อบกพร่องต่าง ๆ และขอบคุณเจ้าหน้าที่คณะสถิติประยุกต์ทุกท่านที่ให้ความช่วยเหลือในการติดต่อประสานงานเป็นอย่างดี

เอกสารอ้างอิง

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152.

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics, 16*(1), 129-132.

Chen, S. Y. (2001). One-stage and two-stage statistical inference under heteroscedasticity. *Communications in Statistics-Simulation and Computation, 30*(4), 991-1009.

Chen, S. Y., & Chen, H. J. (1998). Single-stage analysis of variance under heteroscedasticity. *Communications in Statistics-Simulation and Computation, 27*(3), 641-666.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101-129.

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis, 51*(12), 5731-5742.

Scott, A. J., & Smith, T. M. F. (1971). Interval estimates for linear combinations of means. *Applied Statistics,* 276-285.

Weerahandi, S. (1995). ANOVA under unequal error variances. *Biometrics,* 589-599.

Weerahandi, S. (2013). *Exact statistical methods for data analysis.* Springer Science & Business Media.

Weerahandi, S. (2004). *Generalized inference in repeated measures: exact methods in MANOVA and mixed models (Vol. 500)*. John Wiley & Sons.

Welch, B. L. (1947). The generalization ofstudent's' problem when several different population variances are involved. Biometrika, *34*(1/2), 28-35.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika, 38*(3/4), 330-336.

Xu, L. W., & Wang, S. G. (2008a). A new generalized p-value for ANOVA under heteroscedasticity. *Statistics & Probability Letters, 78*(8), 963-969.

Xu, L., & Wang, S. (2008b). A new generalized p-value and its upper bound for ANOVA under unequal error variances. *Communications in Statistics—Theory and Methods, 37*(7), 1002-1010.