

ความแกร่งของสถิติทดสอบเอฟ เมื่อความแปรปรวนของประชากรไม่เท่ากัน

Robustness of F-test for Heterogeneity of Population Variances

สุชาดา บวรกิติวงศ์* และกุนเฑรา อารีกุล
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Suchada Bowarnkitiwong* and Kuntera Areekul
Faculty of Education, Chulalongkorn University

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความแกร่งของสถิติทดสอบเอฟ ในการทดสอบค่าเฉลี่ยของประชากรภายใต้ความแปรปรวนแตกต่างกันและเพื่อศึกษาขอบเขตของความแปรปรวนที่ทำให้สถิติทดสอบเอฟยังคงมีความแกร่ง โดยความแกร่งพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่ 1 และจากค่าอำนาจการทดสอบ ภายใต้ข้อกำหนดของขนาดตัวอย่างที่เท่ากัน (10,10,10) (30,30,30) (80,80,80) และขนาดตัวอย่างต่างกัน (5,10,15) (20,30,40) (60,80,100) เมื่อมีความแปรปรวนแตกต่างกัน 3 ระดับ คือ น้อย ปานกลาง มาก ที่ระดับนัยสำคัญ .01 และ .05 เมื่อประชากรมีการแจกแจงปกติ ใช้ข้อมูลจำลองด้วยเทคนิคมอนติคาร์โลจากโปรแกรม MATLAB ทำการทดลองซ้ำ 5,000 ครั้งในแต่ละสถานการณ์ ผลวิจัยสรุปได้ดังนี้ ประเด็นแรก กรณีที่ตัวอย่างมีขนาดเท่ากันและความแปรปรวนของประชากรเท่ากัน สถิติทดสอบเอฟจะมีความแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01 ประเด็นที่สอง กรณีที่ตัวอย่างมีขนาดเท่ากันและความแปรปรวนของประชากรแตกต่างกัน สถิติทดสอบเอฟจะมีความแกร่งในทุกสถานการณ์ที่ศึกษาที่ระดับนัยสำคัญ .05 และในตัวอย่างขนาดใหญ่ที่ระดับนัยสำคัญ .01 ประเด็นที่สาม กรณีที่ตัวอย่างมีขนาดไม่เท่ากันและความแปรปรวนของประชากรเท่ากัน สถิติทดสอบเอฟจะมีความแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01 ประเด็นที่สี่ กรณีที่ตัวอย่างมีขนาดไม่เท่ากันและความแปรปรวนของประชากรไม่เท่ากัน สถิติทดสอบเอฟจะมีความแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01 ประเด็นที่ห้า เมื่อตัวอย่างมีขนาดเล็กทั้งที่มีจำนวนเท่ากันและแตกต่างกัน

ค่าอำนาจการทดสอบจะลดลงเมื่อความแปรปรวนของประชากรต่างกันเพิ่มขึ้น แต่เมื่อตัวอย่างขนาดกลางและขนาดใหญ่ทั้งที่มีจำนวนเท่ากันและแตกต่างกัน ความต่างของความแปรปรวนของประชากรที่เพิ่มขึ้นจะมีผลต่อค่าอำนาจการทดสอบเพียงเล็กน้อย ทั้งที่ระดับนัยสำคัญ .05 และ .01

คำสำคัญ: ความแกร่ง สถิติทดสอบเอฟ ความคลาดเคลื่อนประเภทที่ 1 อำนาจการทดสอบ

Abstract

The purpose of this research was to study robustness of F – test in testing means when heterogeneity of population variances and bounds of variances that make F – test still robust. Robustness is considered from type I error rates based on Bradley criterion and power of the test under requirements of sample sizes were to equal (10,10,10), (30,30,30) and (80,80,80) and unequal (5,10,15), (20,30,40) and (60,80,100) with 3 variance ratios: small, moderate and high at .05 and .01 significant levels under normal distribution. Data are simulated by using Monte Carlo simulation technique from MATLAB with 5,000 repetitions for each case. Results could be summarized as follows: 1) When sample sizes are equal and homogeneity of population variances, F – test is robust in every cases at .05 and .01. 2) When sample sizes are equal and heterogeneity of population variances, F – test is robust in every cases at .05 and in large sample size at .01. 3) When sample sizes are unequal and homogeneity of population variances, F – test is robust in every cases at .05 and .01. 4) When sample sizes are unequal and heterogeneity of population variances, F – test is robust in every cases at .05 and .01. 5) When sample sizes are small both equal and unequal, power of the test decreases for the variance ratio of population increases but when sample sizes are moderate and high both equal and unequal sample size the variance ratio of are increases will effect on power of the test a little at the hypothesis testing are .05 and .01 population

Keywords: Robustness, F-test, Type One Error, Power of the Test

บทนำ

ในการออกแบบการวิจัยนั้น นอกจากผู้วิจัยจะต้องมีความเข้าใจเรื่องที่กำลังศึกษาแล้วยังต้องมีความรู้ความเข้าใจด้านสถิติในการวิเคราะห์ข้อมูล เพื่อสรุปสมมติฐานที่ผู้วิจัยตั้งขึ้น ข้อสรุปจะมีความถูกต้องมากน้อยเพียงใดขึ้นอยู่กับทางเลือกใช้สถิติที่เหมาะสมกับลักษณะของข้อมูล เนื่องจากสถิติทดสอบแต่ละตัวมีข้อตกลงเบื้องต้น (assumption) แตกต่างกัน จากการศึกษาพบว่า การฝ่าฝืนข้อตกลงเบื้องต้นบางข้อไม่ทำให้เกิดผลกระทบต่อผลการวิเคราะห์ข้อมูลมากนัก แต่การฝ่าฝืนบางข้อจะทำให้เกิดความคลาดเคลื่อนต่อผลการวิเคราะห์มาก ถ้าผู้วิจัยเลือกใช้สถิติโดยไม่ตรวจสอบข้อตกลงเบื้องต้นก่อน อาจทำให้ได้ข้อสรุปที่ผิดพลาด

สำหรับงานวิจัยเชิงทดลอง แผนการทดลองที่ง่ายและพบ่อยที่สุดคือ แผนการทดลองแบบสุ่มสมบูรณ์หรือสุ่มตลอด (completely randomized design: CR-p) ใช้การวิเคราะห์ความแปรปรวนทางเดียว (one-way ANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของประชากรตั้งแต่ 2 กลุ่มขึ้นไป เป็นเทคนิคการวิเคราะห์ทางสถิติที่ถูกนำมาใช้มากที่สุด สถิติทดสอบเอฟเป็นสัดส่วนระหว่างความแปรปรวนระหว่างกลุ่ม (mean square between) กับความแปรปรวนภายในกลุ่มหรือความคลาดเคลื่อน (mean square within) โดยมีข้อตกลงเบื้องต้น 1) ตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงปกติ (normality) 2) ตัวอย่างถูกสุ่มมาจากประชากรที่เป็นอิสระต่อกัน (independence) และ 3) ความแปรปรวนของประชากรแต่ละกลุ่มต้องเท่ากัน (homogeneity) นั่นคือ $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$

ในการตรวจสอบข้อตกลงเบื้องต้นของสถิติเอฟ มีวิธีการดังนี้ (Bowamkitiwong, 2005) 1) การตรวจสอบความเป็นปกติ (normality) ใช้การทดสอบภาวะสารูปสนิทธิ (goodness of fit) หรือใช้การทดสอบโคโมโกรอฟ-สเมโนฟ (Kolmogorov-Smirnov test) 2) การตรวจสอบความเป็นอิสระ (Independence) ใช้การทดสอบความสุ่ม (run test) และ 3) การตรวจสอบความเป็นเอกพันธ์ของความแปรปรวน (homogeneity of variance) สามารถตรวจสอบ 3.1) กรณีที่มีตัวอย่าง 2 กลุ่มใช้ F-ratio เป็นสถิติทดสอบ 3.2) กรณีที่มีตัวอย่างมากกว่า 2 กลุ่ม ใช้การทดสอบของฮาร์ทลีย์ (Hartley) ที่มี F-max เป็นสถิติทดสอบ จากทั้ง 2 กรณีให้นำค่า sig ที่คำนวณได้มาเปรียบเทียบกับระดับนัยสำคัญที่กำหนด หากค่า sig มีค่ามากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าข้อมูลแต่ละชุดสุ่มมาจากประชากรที่มีความแปรปรวนไม่แตกต่างกัน อย่างไรก็ตามเพื่อความรวดเร็วในระยะหลังได้มีงานวิจัยออกมาสรุปว่า หากความแปรปรวนมากที่สุดหารด้วยความแปรปรวนที่ต่ำที่สุดมีค่าไม่เกิน 4 สรุปได้ว่าความแปรปรวนในประชากรไม่แตกต่างกัน

Khuantham (2006) สรุปว่าการทดสอบความเป็นเอกพันธ์ของความแปรปรวนของเลวิน (Levene's test) มีประสิทธิภาพสูงกว่าการทดสอบของฮาร์ทลีย์ (Hartley's test) การทดสอบของคอคดราน (Cochran's test) และการทดสอบของบาร์ทเลต (Bartlett's test) ระพินทร์ โพธิ์ศรี (2551) เสนอให้ใช้วิธีการของครัสคาล วอลลิส (Kruskal-Wallis) หากผลการทดสอบของเลวิน (Levene's test) ไม่ผ่านข้อตกลงเบื้องต้นการเท่ากันของความแปรปรวนของประชากร

Bonarnkitiwong (2005) กล่าวว่า การฝ่าฝืนข้อตกลงเกี่ยวกับความเป็นเอกพันธ์ของความแปรปรวน (homogeneity of variance) จะมีผลต่อระดับนัยสำคัญและอำนาจการทดสอบในกรณีที่ขนาดตัวอย่างไม่เท่ากันเท่านั้น ในกรณีที่ขนาดตัวอย่างเท่ากันจะมีผลต่อระดับนัยสำคัญและอำนาจการทดสอบไม่มากนัก ดังนั้นการป้องกันทำได้โดยนักวิจัยเก็บข้อมูลแต่ละกลุ่มตัวอย่างจำนวนเท่าและมากพอ

Cochran and Cox (1976) กล่าวว่า ถ้าสถิติทดสอบเอฟไม่เป็นไปตามข้อตกลงเบื้องต้นจะเกิดผลกระทบถึงระดับความมีนัยสำคัญและความไวของการทดสอบ (sensitivity) ถ้าผู้วิจัยทำการทดสอบที่ระดับนัยสำคัญ $\alpha = .05$ แต่ความเป็นจริงแล้วระดับนัยสำคัญ α อาจจะมีค่ามากกว่านั้น

Glass (1970: as cited in Suboonson, 1988) ได้สรุปเรื่องวิธีพินธ์ของความแปรปรวน (heterogeneous variance) ดังนี้ 1) เมื่อขนาดตัวอย่างเท่ากัน การไม่เท่ากันของความแปรปรวนของประชากรมีต่อระดับนัยสำคัญเพียงเล็กน้อย 2) เมื่อขนาดของตัวอย่างไม่เท่ากันและความแปรปรวนไม่เท่ากัน จะเกิดผลดังนี้ 2.1) ตัวอย่างขนาดเล็กและความแปรปรวนของประชากรมีขนาดใหญ่ ความคลาดเคลื่อนประเภทที่ 1 (Type I error) จะใหญ่กว่า α ผลกระทบนี้ทำให้การแจกแจงเอฟเลื่อนไปทางขวา 2.2) ตัวอย่างขนาดใหญ่ และความต่างของแปรปรวนของประชากรมีขนาดเล็กความคลาดเคลื่อนประเภทที่ 1 จะเล็กกว่า α ผลกระทบของกรณีนี้จะทำให้การแจกแจงเอฟเลื่อนไปทางซ้าย (Hsu, 1938; Scheffe, 1959; Lindquist, 1953, Boneau, 1960; Cochran, 1947; Godard & Lindquist, 1940; Horsnell, 1953; Welch, 1937)

Scheffe (1970) ได้ศึกษาผลต่อการฝ่าฝืนข้อตกลงด้านความแปรปรวนของประชากร เมื่อขนาดตัวอย่างเท่ากันและมีขนาดใหญ่พบว่า ความคลาดเคลื่อนประเภทที่ 1 จะใหญ่กว่า α สำหรับตัวอย่างขนาดเล็กการไม่ผ่านข้อตกลงเบื้องต้นข้อนี้จะมีผลกระทบอย่างมากต่อความคลาดเคลื่อนประเภทที่ 1

Suboonson (1988) สรุปว่าสถิติทดสอบเอฟสตาร์ ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้มากกว่าสถิติทดสอบเอฟและสถิติทดสอบยูเฉพาะเมื่อความแปรปรวนของประชากรเท่ากันทั้งตัวอย่างขนาดเล็กและใหญ่ที่เท่ากันและไม่เท่ากัน กรณีของตัวอย่างขนาดเล็กและไม่เท่ากัน เมื่อความแปรปรวนของประชากรต่างกัน สถิติทดสอบ 3 แบบ ไม่สามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามที่ระบุ

Youngyootmeechai (1996) สรุปตัวสถิติทดสอบทั้ง 5 วิธีคือตัวสถิติทดสอบ F ตัวสถิติทดสอบ F ที่มีการแปลงข้อมูลเป็นลอการิทึม (the logarithms transformation) ตัวสถิติทดสอบแบบ Trimmed F ตัวสถิติทดสอบ Brown and Forsythe และตัวสถิติทดสอบเอฟที่ไขค่าเฉลี่ยของ Graybill and Deal ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ในกรณีที่ประชากรมีการแจกแจงปกติและมีอัตราส่วนของส่วนเบี่ยงเบนมาตรฐานเท่ากัน กรณีที่อัตราส่วนของส่วนเบี่ยงเบนมาตรฐานแตกต่างกัน ตัวสถิติทดสอบเอฟของ Graybill and Deal และตัวสถิติทดสอบ Brown and Forsythe ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

Shamaboon (2003) กล่าวว่าสถิติทดสอบเอฟควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้สูงสุดรองลงมาเป็นสถิติทดสอบนอร์มอล-สกอว์ และสถิติทดสอบพรีดแมน

Bamrungsawas (1992) สรุปว่าเมื่อความแปรปรวนของประชากรเท่ากัน สถิติทดสอบ Brown and Forsythe สถิติทดสอบ Marascuilo และสถิติทดสอบ F-test ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ แต่ถ้าอัตราส่วนความแปรปรวนไม่เท่ากันสถิติทดสอบ Brown and Forsythe และสถิติทดสอบ Marascuilo ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าสถิติทดสอบ F ส่วนอำนาจการทดสอบ เมื่อตัวอย่างขนาดเล็กและความแปรปรวนเท่ากัน สถิติทดสอบ F มีอำนาจการทดสอบสูงสุดสำหรับตัวอย่างขนาดใหญ่ สถิติทดสอบทั้ง 3 ตัวมีอำนาจการทดสอบเท่ากัน เมื่อประชากรมีอัตราส่วนความแปรปรวนไม่เท่ากัน สถิติ Brown and Forsythe และสถิติ Marascuilo ให้ค่าอำนาจการทดสอบสูงกว่าสถิติทดสอบ F

Rogan and Keselman (1977) สำหรับตัวอย่างขนาดเท่ากัน หากความแตกต่างของความแปรปรวนมีขนาดใหญ่ อัตราความคลาดเคลื่อนจะใหญ่กว่าอัตราความคลาดเคลื่อนที่ระบุอย่างน้อย 2-4%

Tomarken and Serlin (1986) พบว่า สถิติทดสอบเอฟไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ยกเว้นเมื่อความแปรปรวนเท่ากัน

Clinch and Keselman (1982) เมื่อการแจกแจงของประชากรเป็นแบบปกติ ความแปรปรวนของประชากรต่างกัน และกลุ่มตัวอย่างมีขนาดเท่ากัน สถิติทดสอบเอฟจะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ในขณะที่สถิติทดสอบเอฟสตาร์สามารถควบคุมได้ดี ส่วนอำนาจการทดสอบมีค่าใกล้เคียงกัน สำหรับกลุ่มตัวอย่างขนาดเท่ากัน และสถิติทดสอบของ Welch มีอำนาจการทดสอบมากที่สุดตามด้วยสถิติทดสอบเอฟ และสถิติทดสอบเอฟสตาร์

Guo and Luh (2008) เมื่อความแปรปรวนเท่ากันสถิติทดสอบ F สถิติทดสอบเอฟจะมีค่าอำนาจการทดสอบมากกว่าสถิติทดสอบ Welch และ Welch จะมีค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าสถิติทดสอบเอฟ

Feir and Toothaker (1974) พบว่า สถิติทดสอบเอฟและสถิติทดสอบ Kruskal-Wallis ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้พอๆ กัน แต่สถิติทดสอบเอฟจะมีค่าอำนาจการทดสอบสูงกว่าสถิติทดสอบ Kruskal-Wallis

Alexander and Goven (1994) พบว่า สถิติทดสอบ New normalized t approximation มีค่าความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับสถิติทดสอบ Jame's second order approximation และสถิติทั้ง 2 ตัว ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าสถิติทดสอบเอฟ เมื่อมีความแปรปรวนแตกต่างกัน ทั้งในกลุ่มตัวอย่างขนาดเท่ากันและไม่เท่ากัน ส่วนค่าอำนาจการทดสอบสถิติทดสอบ Jame's second order approximation จะมีอำนาจการทดสอบสูงกว่าสถิติทดสอบทั้ง 2 ตัว

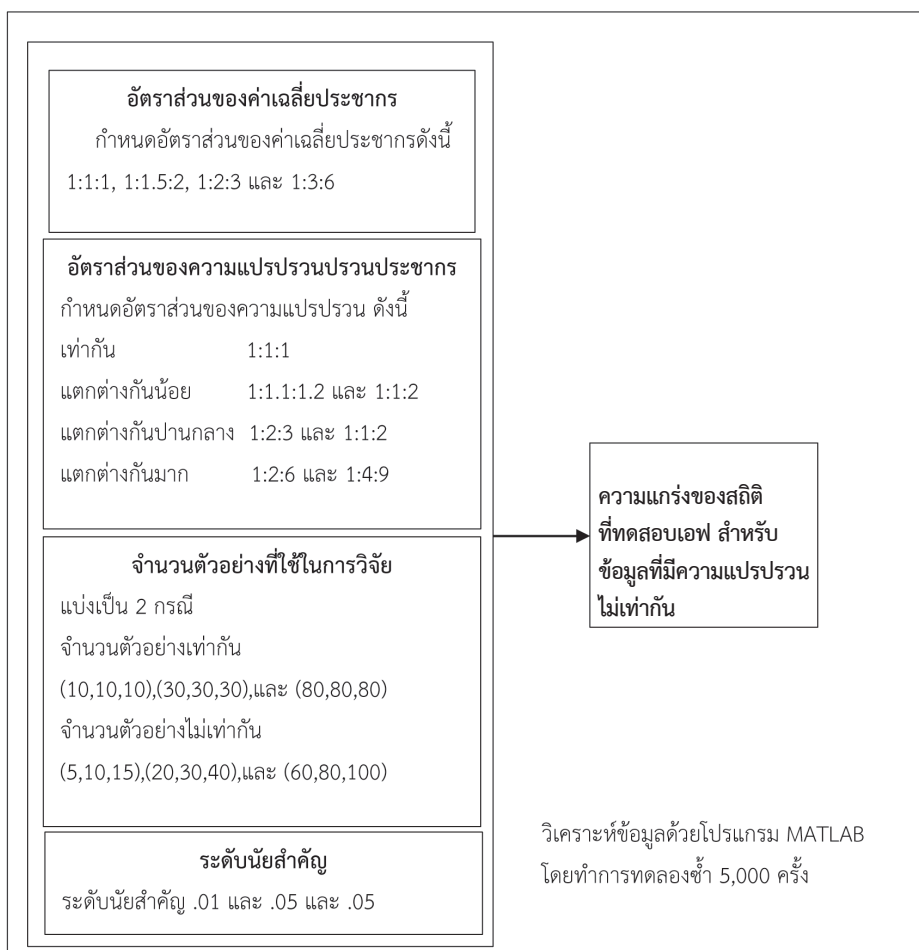
ข้อสรุปงานวิจัยเพื่อแก้ปัญหาการฝ่าฝืนข้อตกลงเบื้องต้น (assumption violation) เกี่ยวกับความเป็นเอกพันธ์ของความแปรปรวน ยังไม่มีความชัดเจน นักวิจัยบางกลุ่มเสนอให้ยังคงใช้สถิติทดสอบเอฟต่อไปได้ เนื่องจากการฝ่าฝืนข้อตกลงเบื้องต้นนี้มีผลเพียงเล็กน้อย บางกลุ่มเสนอให้เปลี่ยนไปใช้สถิติทดสอบอื่น หรือใช้การแปลงข้อมูล (data transformation) ก่อนทำการวิเคราะห์ ผู้วิจัยจึงสนใจศึกษาขอบเขตของ

ความแตกต่างของความแปรปรวนที่นักวิจัยยังสามารถใช้สถิติทดสอบเอฟ โดยค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบยังอยู่ในระดับที่ยอมรับได้

วัตถุประสงค์

1. เพื่อศึกษาความแกร่งของสถิติทดสอบเอฟ ในการทดสอบค่าเฉลี่ยของประชากรภายใต้ความแปรปรวนแตกต่างกัน
2. เพื่อศึกษาขอบเขตของความแปรปรวน ที่ทำให้สถิติทดสอบเอฟยังคงมีความแกร่ง

กรอบแนวคิด



ภาพที่ 1 กรอบแนวคิดในการวิจัย

ระเบียบวิธีการวิจัย

การวิจัยนี้มุ่งศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของสถิติทดสอบเอฟ ในกรณีข้อมูลมีการแจกแจงปกติและความแปรปรวนต่างกัน ผู้วิจัยเริ่มจากการกำหนดจำนวนประชากร 3 กลุ่ม และกำหนดสถานการณ์ที่คาดว่าจะมีผลต่อความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติทดสอบ มีดังนี้

1. กำหนดขนาดตัวอย่าง (Sample size) ทั้ง 3 กลุ่ม แบ่งออกเป็น 2 ประเภท คือ
 - 1.1 ขนาดตัวอย่างเท่ากันคือ (10,10,10) (30,30,30) และ (80,80,80)
 - 1.2 ขนาดตัวอย่างต่างคือ (5,10,15) (20,30,40) และ (60,80,100)
2. กำหนดอัตราส่วนของค่าเฉลี่ยประชากร มีค่าเป็น 1:1:1, 1:1.5:2, 1:2:3 และ 1:3:6
3. กำหนดอัตราส่วนของความแปรปรวนทั้ง 3 ประชากร
 - 3.1 อัตราส่วนของความแปรปรวนเท่ากันคือ 1:1:1
 - 3.2 อัตราส่วนของความแปรปรวนแตกต่างกันน้อย คือ 1:1.1:1.2 และ 1:1:2
 - 3.3 อัตราส่วนของความแปรปรวนแตกต่างกันปานกลาง คือ 1:2:3 และ 1:1:4
 - 3.4 อัตราส่วนของความแปรปรวนแตกต่างกันมาก คือ 1:2:6 และ 1:4:9
4. กำหนดระดับนัยสำคัญของการทดสอบ 2 ระดับ คือ $\alpha = .01$ และ $.05$

การวิเคราะห์ข้อมูล

1. จำลองข้อมูลด้วยโปรแกรม MATLAB คำสั่ง normrand (mu,sigma,m,1) ตามสถานการณ์ที่กำหนด โดยที่ mu แทนค่าเฉลี่ยของประชากร Sigma แทน ค่าส่วนเบี่ยงเบนมาตรฐานของประชากร ซึ่งหาได้จากรากที่สองของความแปรปรวนของประชากร M แทน ขนาดตัวอย่าง
2. สุ่มตรวจสอบค่าสถิติที่คำนวณได้จากการจำลองข้อมูลในข้อ 1 นั่นคือ ตรวจสอบค่าเฉลี่ยและค่าความแปรปรวน โดยคำนวณจากโปรแกรม MATLAB ได้ค่าสถิติดังตารางที่ 1

ตารางที่ 1 ตัวอย่างค่าสถิติที่ได้จากการจำลองข้อมูล

ขนาดตัวอย่าง	ค่าพารามิเตอร์		ค่าสถิติ	
	ค่าเฉลี่ย	ค่าความแปรปรวน	ค่าเฉลี่ย	ค่าความแปรปรวน
10	1	1	0.848	1.088
	1.5	1	1.420	1.161
	2	2	1.556	2.719
80	1	1	0.833	1.041
	3	4	2.661	3.956
	6	9	5.849	8.795

3. คำนวณค่าสถิติทดสอบเอฟ

4. ทดสอบความถูกต้องในการคำนวณค่าสถิติเอฟจากโปรแกรมของผู้วิจัยเปรียบเทียบกับการคำนวณด้วยโปรแกรม SPSS ผลการสุ่มทดสอบพบว่าค่าสถิติทดสอบเอฟที่ได้จากการคำนวณด้วยโปรแกรมที่ผู้วิจัยสร้างขึ้นและโปรแกรม SPSS มีค่าเท่ากันในทุกกรณีที่ทำการสุ่ม ดังแสดงในตารางที่ 2

ตารางที่ 2 การเปรียบเทียบค่าสถิติทดสอบเอฟที่ได้จากโปรแกรมที่ผู้วิจัยสร้างขึ้นกับโปรแกรม SPSS

อัตราส่วน ค่าเฉลี่ย	อัตราส่วนความ แปรปรวน	ขนาดกลุ่ม ตัวอย่าง	ค่า F ที่ได้จากการคำนวณโดย	
			โปรแกรม MATLAB	โปรแกรม SPSS
1:1:1	1:1:1	10,10,10	4.177	4.177
		30,30,30	30.632	6.0632
		80,80,80	71.062	71.062
1:1.5:2	1:1:2	10,10,10	5.882	5.882
		30,30,30	1.469	1.469
		80,80,80	23.004	23.004
1:3:6	1:2:3	5,10,15	17.459	17.459
		20,30,40	77.850	77.850
		60,80,100	311.495	311.495
1:2:3	1:4:9	5,10,15	3.006	3.006
		20,30,40	12.495	12.495
		60,80,100	14.959	14.959

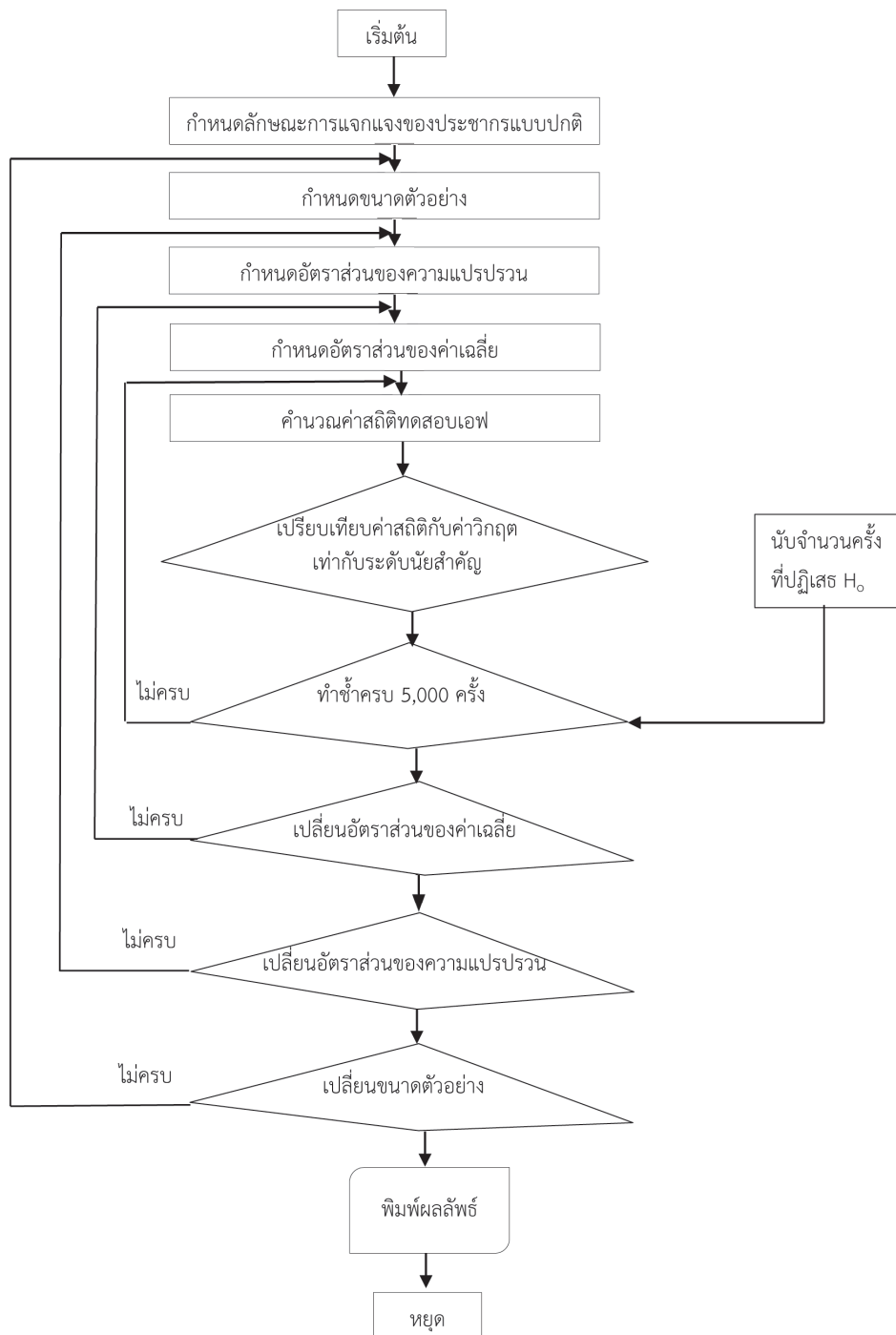
5. การหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบ

ในกรณีที่อัตราส่วนของค่าเฉลี่ยเป็น 1:1:1 หาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งคำนวณโดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นจริงหารด้วย 5,000 ครั้ง นำค่าที่ได้เทียบกับระดับนัยสำคัญที่กำหนดตามเกณฑ์ของ Bradley คือที่ระดับนัยสำคัญ .05 ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วง (.025, .075) และที่ระดับนัยสำคัญ .01 ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วง (.005, .015) จึงถือว่าสถิติทดสอบนั้นสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ในส่วนกรณีที่อัตราส่วนของค่าเฉลี่ยไม่เท่ากัน จะหาค่าอำนาจการทดสอบซึ่งคำนวณโดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นเท็จหารด้วยจำนวนครั้งของการทดสอบ (5,000 ครั้ง)

6. การทดสอบสมมติฐาน

เมื่อหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบครบ 5,000 ครั้งแล้วนำมาทดสอบสมมติฐาน $H_0 = P \geq P_0$ $H_1 = P < P_0$ สถิติที่ใช้ในการทดสอบ

$$Z = \frac{P - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$



ภาพที่ 2 ขั้นตอนการทำงานของโปรแกรม

ผลการวิจัย

ผลการทดสอบทั้งสิ้น 336 สถานการณ์ มีผลสรุปแสดงในตารางที่ 3 และ 4 ดังนี้

ตารางที่ 3 ความแกร่ง จำแนกตามอัตราส่วนความแปรปรวนและขนาดตัวอย่างที่ระดับนัยสำคัญ .05

อัตราส่วน ความแปรปรวน	ขนาดกลุ่ม ตัวอย่าง	ความแกร่ง	ค่าอำนาจการทดสอบเมื่อค่าเฉลี่ย		
			1:1.5:2	1:2:3	1:3:6
1:1:1	10,10,10	✓	0.4584	0.9702	1.0000
	30,30,30	✓	0.9370	1.0000	1.0000
	80,80,80	✓	1.0000	1.0000	1.0000
	5,10,15	✓	0.4750	0.9640	1.0000
	20,30,40	✓	0.9190	1.0000	1.0000
	60,80,100	✓	1.0000	1.0000	1.0000
1:1:1:1.2	10,10,10	✓	0.4240	0.9582	1.0000
	30,30,30	✓	0.9158	1.0000	1.0000
	80,80,80	✓	0.9998	1.0000	1.0000
	5,10,15	✓	0.4142	0.9458	1.0000
	20,30,40	✓	0.9038	1.0000	1.0000
	60,80,100	✓	0.9996	1.0000	1.0000
1:1:2	10,10,10	✓	0.3626	0.9074	1.0000
	30,30,30	✓	0.8228	1.0000	1.0000
	80,80,80	✓	0.9990	1.0000	1.0000
	5,10,15	✓	0.2976	0.8648	1.0000
	20,30,40	✓	0.7836	0.9998	1.0000
	60,80,100	✓	0.9970	1.0000	1.0000
1:2:3	10,10,10	✓	0.2600	0.7640	1.0000
	30,30,30	✓	0.6564	0.9992	1.0000
	80,80,80	✓	0.9872	1.0000	1.0000
	5,10,15	✓	0.1920	0.6878	1.0000
	20,30,40	✓	0.5920	0.9982	1.0000
	60,80,100	✓	0.9808	1.0000	1.0000
1:2:4	10,10,10	✓	0.2174	0.668	1.0000
	30,30,30	✓	0.5872	0.9976	1.0000
	80,80,80	✓	0.9678	1.0000	1.0000
	5,10,15	✓	0.1490	0.5758	1.0000
	20,30,40	✓	0.5092	0.9946	1.0000
	60,80,100	✓	0.9564	1.0000	1.0000

ตารางที่ 3 (ต่อ)

อัตราส่วน ความแปรปรวน	ขนาดกลุ่ม ตัวอย่าง	ความแกร่ง	ค่าอำนาจการทดสอบเมื่อค่าเฉลี่ย		
			1:1.5:2	1:2:3	1:3:6
1:2:6	10,10,10	✓	0.1814	0.5618	0.9998
	30,30,30	✓	0.4746	0.9750	1.0000
	80,80,80	✓	0.9124	1.0000	1.0000
	5,10,15	✓	0.1072	0.4320	0.9992
	20,30,40	✓	0.3658	0.9552	1.0000
	60,80,100	✓	0.8636	1.0000	1.0000
1:4:9	10,10,10	✓	0.1346	0.4048	0.9914
	30,30,30	✓	0.3148	0.8892	1.0000
	80,80,80	✓	0.7314	1.0000	1.0000
	5,10,15	✓	0.0764	0.2650	0.9866
	20,30,40	✓	0.2254	0.8336	1.0000
	60,80,100	✓	0.6752	1.0000	1.0000

ตารางที่ 4 ความแกร่ง จำแนกตามอัตราส่วนความแปรปรวนและขนาดตัวอย่าง ที่ระดับนัยสำคัญ .01

อัตราส่วน ความแปรปรวน	ขนาดกลุ่ม ตัวอย่าง	ความแกร่ง	ค่าอำนาจการทดสอบเมื่อค่าเฉลี่ย		
			1:1.5:2	1:2:3	1:3:6
1:1:1	10,10,10	✓	0.2206	0.8860	1.0000
	30,30,30	✓	0.8074	1.0000	1.0000
	80,80,80	✓	0.9998	1.0000	1.0000
	5,10,15	✓	0.2368	0.8700	1.0000
	20,30,40	✓	0.7862	1.0000	1.0000
	60,80,100	✓	0.9996	1.0000	1.0000
1:1:1:1.2	10,10,10	✓	0.1932	0.8502	1.0000
	30,30,30	✓	0.7520	1.0000	1.0000
	80,80,80	✓	0.9992	1.0000	1.0000
	5,10,15	✓	0.1930	0.8214	1.0000
	20,30,40	✓	0.7388	1.0000	1.0000
	60,80,100	✓	0.9988	1.0000	1.0000
1:1:2	10,10,10	✓	0.1594	0.7408	1.0000
	30,30,30	✓	0.6200	0.9998	1.0000
	80,80,80	✓	0.9912	1.0000	1.0000
	5,10,15	✓	0.1206	0.6604	1.0000
	20,30,40	✓	0.5734	0.9994	1.0000
	60,80,100	✓	0.9876	1.0000	1.0000

ตารางที่ 4 (ต่อ)

อัตราส่วน ความแปรปรวน	ขนาดกลุ่ม ตัวอย่าง	ความแกร่ง	ค่าอำนาจการทดสอบเมื่อค่าเฉลี่ย		
			1:1.5:2	1:2:3	1:3:6
1:2:3	10,10,10	✓	0.0964	0.5078	1.0000
	30,30,30	✓	0.4010	0.9926	1.0000
	80,80,80	✓	0.9430	1.0000	1.0000
	5,10,15	✓	0.0676	0.4116	1.0000
	20,30,40	✓	0.3294	0.9850	1.0000
	60,80,100	✓	0.9222	1.0000	1.0000
1:2:4	10,10,10	✓	0.0790	0.4316	1.0000
	30,30,30	✓	0.3392	0.9764	1.0000
	80,80,80	✓	0.8860	1.0000	1.0000
	5,10,15	✓	0.0478	0.3022	1.0000
	20,30,40	✓	0.2560	0.9616	1.0000
	60,80,100	✓	0.8504	1.0000	1.0000
1:2:6	10,10,10	✗	-	-	-
	30,30,30	✗	-	-	-
	80,80,80	✓	0.7660	1.0000	1.0000
	5,10,15	✓	0.0296	0.2006	0.9892
	20,30,40	✓	0.1670	0.8488	1.0000
	60,80,100	✓	0.678	1.0000	1.0000
1:4:9	10,10,10	✗	-	-	-
	30,30,30	✗	-	-	-
	80,80,80	✓	0.5130	0.9994	1.0000
	5,10,15	✓	0.0206	0.0996	0.9226
	20,30,40	✓	0.0804	0.5942	1.0000
	60,80,100	✓	0.4234	0.9974	1.0000

จากตารางแสดงความแกร่งของสถิติทดสอบเอฟ พบว่า

1. กรณีตัวอย่างมีขนาดเท่ากันและอัตราส่วนความแปรปรวนของประชากรเท่ากัน สถิติทดสอบเอฟ มีความแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01
2. กรณีตัวอย่างมีขนาดเท่ากันแต่อัตราส่วนความแปรปรวนของประชากรต่างกัน สถิติทดสอบเอฟ มีความแกร่ง ในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และในตัวอย่างขนาดใหญ่ที่ระดับนัยสำคัญ .01
3. ตัวอย่างที่มีขนาดต่างกันและอัตราส่วนความแปรปรวนของประชากรเท่ากัน สถิติทดสอบเอฟ มีความแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01

4. ตัวอย่างที่มีขนาดต่างกันแต่อัตราส่วนความแปรปรวนของประชากรไม่เท่ากัน สถิติทดสอบเอฟมีความแข็งแกร่งในทุกสถานการณ์ที่ศึกษาทั้งที่ระดับนัยสำคัญ .05 และ .01

5. เมื่อพิจารณาค่าอำนาจการทดสอบพบว่า ในตัวอย่างขนาดเล็ก และขนาดกลางทั้งที่เท่ากัน และไม่เท่ากัน อัตราส่วนความแปรปรวนแตกต่างกันเพิ่มขึ้น ค่าอำนาจการทดสอบจะต่ำลง แต่จะมีค่าเพิ่มขึ้นเมื่ออัตราส่วนค่าเฉลี่ยเพิ่มขึ้น ส่วนในตัวอย่างขนาดใหญ่ ค่าอำนาจการทดสอบมีค่าสูงในทุกสถานการณ์ที่ศึกษา

6. การเพิ่มขนาดกลุ่มตัวอย่าง มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 โดยควบคุมความคลาดเคลื่อนได้ดีเมื่อมีการเพิ่มขนาดตัวอย่าง

7. อัตราส่วนความแปรปรวนของประชากรมีส่วนในการพิจารณาการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งจากผลการทดสอบพบว่า ภายในสถานการณ์เดียวกัน (ลักษณะการแจกแจงปกติ, ขนาดกลุ่มตัวอย่าง และอัตราส่วนค่าเฉลี่ยประชากร) เมื่ออัตราส่วนความเปลี่ยนแปลงของประชากรต่างกันมาก (1:2:6 และ 1:4:9) จะทำให้การทดสอบด้วยสถิติทดสอบเอฟไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ โดยเฉพาะในกรณีที่มีตัวอย่างขนาดเล็ก (10,10,10) และขนาดกลาง (30,30,30) นั้นอาจเนื่องมาจากอัตราส่วนความแปรปรวนที่แตกต่างกันมากและตัวอย่างมีขนาดเล็ก

8. อำนาจการทดสอบจะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้นหรือระดับนัยสำคัญเพิ่มขึ้น เนื่องจากการเพิ่มระดับนัยสำคัญ (α) จะทำให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 2 (β) มีค่าน้อยลง จึงทำให้ค่าอำนาจการทดสอบซึ่งมีค่าเท่ากับ $1-\beta$ สูงขึ้น

9. การเพิ่มขนาดตัวอย่าง และเพิ่มอัตราส่วนค่าเฉลี่ยของประชากรมากขึ้น จะมีผลทำให้ค่าอำนาจการทดสอบของสถิติเอฟมีค่ามากขึ้น แต่การเพิ่มอัตราส่วนความแปรปรวนของประชากรมากขึ้น และตัวอย่างมีขนาดเล็ก จะมีผลทำให้ค่าอำนาจการทดสอบเอฟมีค่าน้อยลง

อภิปรายผล

สถิติทดสอบเอฟมีความแข็งแกร่งเมื่ออัตราส่วนความแปรปรวนของประชากรแต่ละกลุ่มเท่ากัน แม้ว่า จะเป็นตัวอย่างขนาดเล็ก กลาง หรือใหญ่ และตัวอย่างเท่ากันหรือไม่เท่ากัน ทั้งที่ระดับนัยสำคัญเท่ากับ .05 และ .01 นับเป็นคุณสมบัติที่ดีของสถิติทดสอบเอฟ ซึ่งสอดคล้องกับงานวิจัยของ Suboonson (1998), Bamrungsawas (1992) & Youngyootmeechai (1996) พบว่า สถิติทดสอบเอฟควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งระดับนัยสำคัญ.05 และ .01 เฉพาะเมื่อความแปรปรวนของประชากรเท่ากัน ใช้ได้กับตัวอย่างทั้งขนาดเล็กและใหญ่ ที่เท่ากันหรือไม่เท่ากัน แต่ในกรณีที่อัตราส่วนความแปรปรวนของประชากรแตกต่างกัน สถิติทดสอบเอฟไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ซึ่งสอดคล้องบางส่วนกับงานวิจัยของ Rogan and Keselman (1977) กล่าวว่า สถิติทดสอบเอฟไม่แกร่งต่อความไม่เท่ากันของความแปรปรวนทุกระดับ งานวิจัยนี้มีผลเฉพาะตัวอย่างขนาดเล็กเท่านั้น นอกจากนี้ผลการวิจัยนี้ยังพบอีกว่าการเพิ่มขนาดตัวอย่างและเพิ่มอัตราส่วนค่าเฉลี่ยของประชากรจะมีผลทำให้อำนาจการทดสอบของสถิติเอฟมีค่ามากขึ้น แต่การเพิ่มอัตราส่วนความแปรปรวนของประชากรมากขึ้น ในตัวอย่างขนาดเล็ก

จะมีผลทำให้ให้อำนาจการทดสอบของสถิติเอฟมีค่าน้อย นั่นคือ ความแตกต่างของความแปรปรวนของประชากรยิ่งมากขึ้น ค่าความคลาดเคลื่อนก็ยิ่งมากขึ้น และยิ่งขนาดตัวอย่างใหญ่ขึ้นจะกระทบต่อความไม่เท่ากันของความแปรปรวนน้อยลง

ข้อเสนอแนะ

เมื่อความแปรปรวนของประชากรแตกต่างกันมาก ควรใช้ตัวอย่างขนาดกลางหรือขนาดใหญ่ เพราะจะสามารถควบคุมอัตราความเคลื่อนที่ประเภทที่ 1 แต่ถ้าตัวอย่างมีขนาดเล็ก ควรเลือกใช้สถิติทดสอบตัวอื่นเช่น สถิตินอนพาราเมตริกซ์

References

- กิ่งทอง ยงยุทธมิชัย. (2539). *การเปรียบเทียบตัวสถิติทดสอบค่าเฉลี่ยของประชากรที่มีความแปรปรวนไม่เท่ากัน กรณีศึกษาสำหรับแผนการทดลองแบบสุ่มตลอด* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- นันทวัน บำรุงสวัสดิ์. (2534). *การเปรียบเทียบวิธีทดสอบความเท่ากันของค่าเฉลี่ย เมื่อความแปรปรวนของประชากรไม่เท่ากัน* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ระพีพันธ์ โพธิ์ศรี. (2551). *สถิติเพื่อการวิจัย*. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สมทรง สุกุณสันต์. (2531). *การเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบเอฟ เอฟสตาห์ และยู เมื่อความแปรปรวนของประชากรไม่เท่ากัน* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- สุชาติ บวรกิตวงศ์. (2548). *สถิติประยุกต์ทางพฤติกรรมศาสตร์*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุพัตรา ชมะบุรณ์. (2546). *การเปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของสถิติทดสอบเอฟ สถิติทดสอบฟรیدแมน และสถิติทดสอบนอร์มอล-สกอว์ สำหรับแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- อนันต์ชัย เชื้อนธรรม. (2549). *วิธีการทางสถิติและการวิเคราะห์ข้อมูล*. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- Alexander, R.A., & Govern, D.M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of educational statistics*, 19, 91-101.

- Clinch, J.J., & Keselman, H.J. (1982). Parametric alternatives to the analysis of variance. *Journal of educational statistics*, 7, 207-214.
- Cochran, W.G., & Cox, G.M. (1976). *Experimental Design*. New York: John Hiley and Sons.
- Feir, B.J., & Toothaker, L.E. (1974). The ANOVA F – test versus the Kruskal – Wallis test: A robustness study. *American educational research association*, 59, 226-234.
- Guo, J.H., & Luh, W.M. (2008). Approximate sample size formulas for testing group mean differences when variances are unequal in one – way ANOVA. *Educational and psychological measurement*, 68, 959-971.
- Rogan, J.C., & Keseiman, H. J. (1977). Is the ANOVA F – test robust to variance heterogeneity when sample size are equal? : An investigation via a coefficient of variation. *American Educational Research Journal*, 14, 493 – 498.
- Scheffe, H. (1970). *The Analysis of Variance (6 th ed)*. New York: John Wiley and Sons.
- Tomarken, A.J., & Serlin, R.C. (1986). Comparison of ANOVA Alternatives Under Variance Heterogeneity and Specific Noncentrality Stuctures. *Psychological Bulletin*, 99, 90-99.

Translated Thai References

- Bamrungsawas, N. (1992). *A Comparison on Method for Testing equality of means when population variances are unequal* (Master's thesis). Bangkok: Chulalongkorn University Press. (in Thai)
- Bowarnkitiwong, S. (2005). *Statistics applied of behavioral sciences* (Master's thesis). Bangkok: Chulalongkorn University Press. (in Thai)
- Khuantham, A. (2006). *Statistical methods and analysis*. Bangkok: Kasetsart University Press. (in thai)
- Phosi, R. (2007). *Statistics for Research*. Bangkok: Chulalongkorn University Press. (in Thai)
- Shamaboon, S. (2003). *Comparisons of type I error rates and power of the test among F-test, Friedman test and Normal-Scores tests for completely randomized block design* (Master's thesis). Bangkok: Chulalongkorn University. (in Thai)

- Suboonson, S. (1988). *The comparison of ability to control type I error among F-test, F*-test, and U-test with unequal variances of populations* (Master's thesis). Bangkok: Chulalongkorn University. (in Thai)
- Youngyootmeechai, K. (1996). *A comparison on test statistics for population means with unequal variances : a case study of completely randomized design* (Master's thesis). Bangkok: Chulalongkorn University. (in Thai)

คณะผู้เขียน

รองศาสตราจารย์ ดร.สุชาดา บวรกิตติวงศ์

คณะครุศาสตร์/จุฬาลงกรณ์มหาวิทยาลัย

254 ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330

e-mail: bsuchada@chula.ac.th

นางสาวกณทิรา อารีกุล

คณะครุศาสตร์/จุฬาลงกรณ์มหาวิทยาลัย

254 ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330

e-mail: kate_happy1@hotmail.com