“การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม”

โดย

คณาวุฒิ เจียมวัฒนพงศ์

นิศานาถ อิงคดาภา

สนับสนุนงบประมาณโดย

มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

ประจำปีงบประมาณ 2562

Tests for Equality of Covariance Matrices

By

KNAVOOT JIAMWATTANAPONG

NISANAD INGADAPA

Granted by

Rajamangala University of Technology Rattanakosin

Fiscal Year 2018

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือและกำลังใจจากกัลยาณมิตรทุกท่านทั้งที่มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์และที่มหาวิทยาลัยเชียงใหม่

ผู้วิจัยขอกราบขอบพระคุณศาสตราจารย์สำรวม จงเจริญ ผู้ให้คำแนะนำและเป็นที่ปรึกษาให้กับผู้วิจัยมาโดยตลอด

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ที่ให้การสนับสนุนงบประมาณสำหรับดำเนินการวิจัยในครั้งนี้

คณาวุฒิ เจียมวัฒนพงศ์และคณะ

สิงหาคม 2562

บทคัดย่อ

รหัสโครงการ : C-28/2562
ชื่อโครงการ : การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม
ชื่อนักวิจัย : ผศ.ดร.คณาวุฒิ เจียมวัฒนพงศ์ และ นางนิศานาถ อิงคดาภา

การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมเป็นหนึ่งในข้อตกลงเบื้องต้นที่สำคัญของการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) และการวิเคราะห์จำแนกกลุ่ม (Discriminant analysis) วิธี Box'M Test เป็นวิธีทดสอบที่ได้รับความนิยมในการทดสอบดังกล่าว และเป็นวิธีที่ปรากฏในโปรแกรมสำเร็จรูปทางสถิติอย่าง SPSS สถิติ Box'M สามารถแปลงไปยังสถิติทดสอบที่ประมาณการแจกแจงด้วยการแจกแจงไคกำลังสอง(Chi-squared distribution) และการแจกแจงเอฟ($F-$distribution) ได้ งานวิจัยนี้มุ่งที่จะศึกษาประสิทธิภาพของวิธี Box'M Test และเปรียบเทียบวิธีดังกล่าวกับวิธีทดสอบโดยอาศัยการประมาณการแจกแจงด้วยการแจกแจงไคกำลังสองและการแจกแจงเอฟภายใต้ข้อมูลแจกแจงแบบปรกติหลายตัวแปร ส่วนในกรณีที่ข้อมูลไม่ได้แจกแจงปรกติได้ศึกษาวิธีทดสอบแบบไม่อิงพารามิเตอร์ โดยวิธี Bootstrap

ผลการวิจัย พบว่า

1. เมื่อข้อมูลมีการแจกแจงปรกติหลายตัวแปร(Multivariate normality) และขนาดตัวอย่างแต่ละชุดใกล้เคียงกัน วิธี Box'M Test อีกทั้งวิธีทดสอบที่ใช้การแจกแจงไคกำลังสองและการแจกแจงเอฟในการประมาณการแจกแจงของสถิติทดสอบล้วนมีประสิทธิภาพดีภายใต้สถานการณ์ที่ศึกษา นอกจากนี้จำนวนชุดของประชากร จำนวนตัวแปร และขนาดตัวอย่าง มีผลต่อประสิทธิภาพของวิธีทดสอบทั้งสามวิธี

2. สำหรับข้อมูลที่มิได้มีการแจกแจงแบบปรกติ(Non-normality) วิธีทดสอบที่ไม่อิงพารามิเตอร์โดยใช้วิธี Bootstrap และวิธีที่ประมาณการแจกแจงของสถิติทดสอบด้วยการแจกแจงไคกำลังสองและการแจกแจงเอฟยังให้ผลไม่เป็นที่น่าพอใจ

3. คำแนะนำสำหรับนักวิจัยที่ใช้วิธี Box'M Test ในโปรแกรม SPSS คือ ควรเลือกใช้ระดับนัยสำคัญที่ 0.001 และควรคำนึงถึงการแจกแจงของข้อมูลรวมทั้งขนาดของตัวอย่างที่ใช้ด้วย

คำสำคัญ : การทดสอบเมทริกซ์ความแปรปรวน, การทดสอบ Box'M, ความแปรปรวนไม่เท่ากัน

E-mail Address : knavoot.jia@rmutr.ac.th

ระยะเวลาโครงการ : ตุลาคม 2561 – กันยายน 2562

Abstract

Code of project : C-28/2562

Project name : Tests for Equality of Covariance Matrices

Researcher name : Asst. Dr. Knavoot Jiamwattanapong and Mrs. Nisanad Ingadapa

Homogeneity of covariance matrices is one of the important assumptions of the multivariate analysis of variance (or MANOVA) and also the discriminant analysis. Box'M test is a commonly used method to check whether the covariance matrices across groups are equal. This test is included in many statistical packages such as SPSS. The Box'M statistic can be transformed to the test statistics and their distributions are approximated using chi-squared and F distributions. This study aims to evaluate the performance of Box'M test and compare to the other two related tests under multivariate normality. In the case of non-normality, the nonparametric method using bootstrap is studied.

The results were as follows:

1. Under multivariate normality and the sample sizes of all groups are quite the same, the Box'M test and also the other two tests using chi-squared and $F-$ distributions performs well in the situations studied. Moreover, the number of populations, the number of variables, and the sample size affect the performance of all of the three tests.

2. Under non-normality, the bootstrap method used in this study and the two tests based on Box'M statistic including chi-squared and F distributions are unsatisfactory.

3. To apply Box'M tests in SPSS, the significance level of 0.001 is recommended and the distribution of data and the sample size should be taken into account.

<u>**Keywords: Tests for covariance matrices, Box'M test, Heterogeneity of variances**</u>

E-mail Address : **knavoot.jia@rmutr.ac.th**

Period of project : October 2018 – September 2019

สารบัญ

สารบัญ (ต่อ)

สารบัญตาราง

บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

ความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของประชากรเป็นหนึ่งในข้อสมมติเบื้องต้น (Assumption) ที่สำคัญของการวิเคราะห์ความแปรปรวนหลายตัวแปร(Multivariate analysis of variance หรือ MANOVA) และการวิเคราะห์จำแนกกลุ่ม(Discriminant analysis) ซึ่งเป็นวิธีการวิเคราะห์ข้อมูลหลายตัวแปร(Multivariate data) ที่สำคัญและพบได้บ่อยครั้งในสาขาวิชาต่างๆ นับตั้งแต่ด้านการแพทย์ ด้านวิทยาศาสตร์และวิศวกรรมศาสตร์ ด้านสังคมศาสตร์ ด้านบริหารธุรกิจ ด้านการศึกษา ด้านรัฐศาสตร์ ตลอดจนด้านการบริหารทรัพยากรมนุษย์

ถึงแม้เทคนิคในการวิเคราะห์ข้อมูลหลายตัวแปรจะได้รับความนิยม มีการนำไปใช้กันอย่างแพร่หลายในปัจจุบันแล้วก็ตาม การละเมิดข้อสมมติเบื้องต้นอาจส่งผลกระทบต่อการประมาณค่าในตัวแบบที่ศึกษาได้(Lix & Keselman, 1996) ตัวอย่างเช่น เมื่อต้องการทดสอบสมมติฐานเพื่อเปรียบเทียบเวคเตอร์ของค่าเฉลี่ย(Mean vectors) แล้วให้ข้อสรุปที่เป็นการปฏิเสธสมมติฐานหลัก ข้อสรุปดังกล่าวอาจเป็นผลมาจากความไม่เท่ากันของเมทริกซ์ความแปรปรวนร่วม(Variance-covariance matrices) โดยมิได้เป็นผลมาจากค่าเฉลี่ยก็ได้

วิธีการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมสำหรับประชากร k ชุดในการวิเคราะห์ข้อมูลหลายตัวแปรเป็นที่สนใจและได้มีการพัฒนาวิธีการทดสอบเพื่อใช้งานเป็นจำนวนมาก แนวทางในการพัฒนาวิธีการทดสอบดังกล่าวอาจจำแนกได้เป็น 3 แนวทาง(Gupta & Bodna, 2014) คือ (1)การทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) (2)การทดสอบที่อาศัยระยะทางเอมไพริคัล(empirical distance) ซึ่งการทดสอบที่ใช้แนวทางนี้มักสามารถใช้ในการทดสอบสมมติฐานกรณีของ high-dimensional data (ข้อมูลที่มีจำนวนตัวแปรมากกว่าขนาดตัวอย่าง หรือ $p > n$)ได้ เช่น งานของ Ledoit & Wolf(2002) และ (3)การทดสอบที่อาศัยการแจกแจงของค่าสูงสุดของค่าไอเกน(Largest eigenvalue distribution) หรือที่อาศัยวิธีการใน Random matrix theory(ดู Cai & Jiang, 2011) วิธีการทดสอบส่วนใหญ่ที่พัฒนาขึ้นอยู่บนพื้นฐานของการแจกแจงปรกติหลายตัวแปร(Multivariate normal distribution) และหากข้อมูลไม่ได้แจกแจงปรกติแล้ว การใช้วิธีทดสอบไม่อิงพารามิเตอร์(Nonparametric methods) เช่น วิธี bootstrap ก็สามารถนำมาปรับใช้ได้(Srivastava, 2002)

ในบรรดาวิธีการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมนี้ อาจกล่าวได้ว่าวิธีการทดสอบของ Box(1949, 1950) หรือที่เรียกว่า **Box's M Test** นั้นเป็นวิธีทดสอบที่ใช้กัน

แพร่หลายและได้บรรจุไว้ในโปรแกรมสำเร็จรูปทางสถิติทั่วไป เช่น SPSS เป็นต้น วิธี Box’s M Test นี้พัฒนาขึ้นจากเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) โดยมีข้อสมมติเบื้องต้นของข้อมูลที่มีการแจกแจงปรกติหลายตัวแปร โดยวิธีนี้จะสามารถหาค่าวิกฤติของการทดสอบแม่นตรง (Exact test) ได้ เฉพาะกรณีที่จำนวนตัวแปร(p) เท่ากับ 2, 3, 4 และ 5 โดยมีจำนวนชุดของประชากร(k) ไม่เกิน 10 ชุด วิธีทดสอบดังกล่าวยังสามารถใช้การแจกแจงไคกำลังสอง(Chi-square distribution) และการแจกแจงเอฟ($F-$distribution) ประมาณการแจกแจงของสถิติทดสอบได้ด้วย หากแต่เงื่อนไขที่เหมาะสมในการประมาณการแจกแจงดังกล่าวอาจแตกต่างกันอยู่บ้าง การศึกษาประสิทธิภาพของวิธี Box’M Test ภายใต้สถานการณ์ที่จำนวนตัวแปร, จำนวนกลุ่มของประชากรที่ต้องการทดสอบ และขนาดของตัวอย่างแต่ละชุด(n_i) มีการเปลี่ยนแปลง

จากที่กล่าวข้างต้น จึงเป็นที่น่าสนใจที่จะศึกษาวิธีการทดสอบที่ใช้กันอย่างแพร่หลาย คือ วิธี Box’s M Test ว่ามีประสิทธิภาพเพียงใดและควรใช้การประมาณการแจกแจงในเงื่อนไขใดเพื่อให้ได้ผลสรุปที่น่าเชื่อถือในการตรวจสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม ก่อนที่นักวิจัยหรือผู้วิเคราะห์ข้อมูลจะได้ดำเนินการวิเคราะห์ข้อมูลด้วยวิธีการอื่นต่อไป

2. วัตถุประสงค์การวิจัย

1) เพื่อศึกษาประสิทธิภาพของวิธี Box’s M Test ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมเมื่อจำนวนชุดของประชากร จำนวนตัวแปรและขนาดตัวอย่างมีการเปลี่ยนแปลง

2) เพื่อเปรียบเทียบวิธีการทดสอบไม่อิงพารามิเตอร์กับวิธี Box’s M Test ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมสำหรับกรณีประชากร 2 ชุด

3) เพื่อให้คำแนะนำสำหรับผู้วิเคราะห์ข้อมูลและนักวิจัยที่ต้องการตรวจสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม

3. คำถามการวิจัย

1) ประสิทธิภาพของการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนด้วยวิธี Box’M Test เป็นอย่างไร เมื่อจำนวนชุดของประชากร จำนวนตัวแปรและขนาดตัวอย่างมีการเปลี่ยนแปลง

2) การทดสอบแบบไม่อิงพารามิเตอร์สำหรับทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมมีประสิทธิภาพอย่างไรเมื่อเปรียบเทียบกับวิธี Box’s M Test ภายใต้การแจกแจงที่ไม่ใช่การแจกแจงปรกติ

3) เมื่อนักวิจัยทางสังคมศาสตร์ต้องใช้วิธีทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม จะมีแนวทางในการดำเนินการอย่างไร

4. ขอบเขตการวิจัย

1) สำหรับทุกวิธีทดสอบที่เป็นการทดสอบแบบอิงพารามิเตอร์จะกำหนดให้ประชากรมีการแจกแจงปรกติหลายตัวแปร(Multivariate normal distribution) ส่วนการทดสอบไม่อิงพารามิเตอร์จะใช้วิธี Bootstrap และกำหนดให้ประชากรมีการแจกแจงแบบทีหลายตัวแปร(Multivariate $t-$distribution) ที่องศาเสรีเท่ากับ 5, 10, 15 และ 20

2) จำนวนชุดของประชากรที่ศึกษา(k) เท่ากับ 2 และ 6 กลุ่ม

3) จำนวนตัวแปร(p) เท่ากับ 2 และ 5

4) ไม่มีค่านอกเกณฑ์(no outliers)

บทที่ 2
การทบทวนวรรณกรรมที่เกี่ยวข้อง

ความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของประชากรทุกชุดเป็นหนึ่งในข้อสมมติเบื้องต้น(Assumption) ที่สำคัญในการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) และการวิเคราะห์จำแนกกลุ่ม(Discriminant analysis) วิธีในการทดสอบความเท่ากันเมทริกซ์ความแปรปรวนร่วมจึงได้รับการพัฒนาขึ้นเป็นจำนวนมากและอาจจำแนกตามแนวทางที่ใช้ในการพัฒนาตัวสถิติทดสอบได้เป็น 3 แนวทาง(Gupta & Bodna, 2014) คือ (1) การทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) ตัวอย่างเช่น การทดสอบของ Mauchly(1940) และการทดสอบของ Box(1949, 1950) (2) การทดสอบที่อาศัยระยะทางเอมไพริคัล(Empirical distance) ซึ่งการทดสอบที่ใช้แนวทางนี้มักสามารถใช้ในการทดสอบสมมติฐานกรณีของ high-dimensional data(ข้อมูลที่จำนวนตัวแปรมากกว่าขนาดตัวอย่าง หรือ $p > n$) ได้ เช่น การทดสอบของ John(1971) และ การทดสอบของ Nagao(1973) และ (3) การทดสอบที่อาศัยการแจกแจงของค่าสูงสุดของค่าไอเกน(Largest eigenvalue distribution) หรือที่อาศัยวิธีการใน Random matrix theory เช่น การทดสอบของ Cai และ Jiang(2011)

ในบรรดาวิธีการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของประชากร k ชุด อาจกล่าวได้ว่า วิธีการทดสอบของ Box หรือที่เรียกว่าวิธี Box’s M Test นั้นเป็นวิธีการที่ได้รับความนิยมและสามารถพบได้ในโปรแกรมสำเร็จรูปทางสถิติ เช่น SPSS เป็นต้น การทดสอบของ Box เป็นวิธีที่พัฒนาขึ้นโดยอาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) ซึ่งเป็นแนวทางเดียวกับที่ Bartlett ใช้ในการพัฒนาวิธีการทดสอบความเท่ากันของความแปรปรวนสำหรับข้อมูลตัวแปรเดียว(Bartlett, 1937) การทบทวนวรรณกรรมนี้ จึงได้นำเสนอวิธีการทดสอบความเท่ากันของความแปรปรวนในกรณีของข้อมูลตัวแปรเดียวซึ่งเป็นวิธีของ Bartlett ไว้ในช่วงต้นด้วย

1. การทดสอบความเท่ากันของความแปรปรวนกรณีตัวแปรเดียว

ในการวิเคราะห์ข้อมูลตัวแปรเดียว(Univariate analysis) การทดสอบความเท่ากันของความแปรปรวน(Homogeneity of variance) โดยวิธีของ Bartlett (Bartlett’s test) อาศัยข้อสมมติเบื้องต้นของการแจกแจงปรกติ(Normality) และความแปรปรวนของตัวอย่างทุกชุดต้องเป็นอิสระกัน

สมมติฐานหลักของ Bartlett’s test คือ ความแปรปรวนของประชากรทุกชุดเท่ากัน เทียบกับ สมมติฐานรองคือ ประชากรอย่างน้อย 2 ชุดมีความแปรปรวนต่างกัน ดังนี้

$$\text{H}: \ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

$$\text{A}: \ \sigma_i^2 \ \neq \ \sigma_j^2 \ \text{อย่างน้อย 1 คู่ที่} \ i \neq j, \ i, j = 1, 2, ..., k \qquad (1)$$

โดยที่ σ_i^2 แทน ความแปรปรวนของประชากรกลุ่มที่ i

k แทน จำนวนชุดของประชากร

สถิติทดสอบของ Bartlett ได้แก่ $B = m / c$ โดยค่า c และ m คำนวณได้ดังนี้

$$c = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu} \right) \qquad (2)$$

และ $$m = \nu \ln s_p^2 - \sum_{i=1}^{k} \nu_i \ln s_i^2 \qquad (3)$$

โดยที่ $$s_i^2 = \frac{1}{\nu_i} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i^2), \qquad \overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

$$s_p^2 = \frac{\sum_{i=1}^{k} \nu_i s_i^2}{\nu}, \qquad \nu_i = n_i - 1 \ \text{และ} \ \nu = \sum_{i=1}^{k} \nu_i$$

สถิติทดสอบ B มีการแจกแจงโดยประมาณแบบ χ_{k-1}^2 และจะปฏิเสธสมมติฐานหลักเมื่อ $B > \chi_{\alpha,\, k-1}^2$

นอกจากจะประมาณการแจกแจงสถิติทดสอบ B โดยใช้การแจกแจงกำลังสอง(Chi-squared distribution) แล้ว สามารถประมาณการแจกแจงด้วยการแจกแจง F ($F-$distribution) ที่องศาเสรี a_1 และ a_2 ดังนี้

$$F = \frac{a_2 m}{a_1 (b - m)} \qquad (4)$$

โดยที่ $a_1 = k - 1$, $a_2 = (k+1)/(c-1)^2$ และ $b = a_2 / (2 - c + 2/a_2)$

และจะปฏิเสธสมมติฐานหลักเมื่อ $F > F_{\alpha, a_1, a_2}$ เมื่อ α คือ ระดับนัยสำคัญ

อย่างไรก็ตาม เนื่องจากการทดสอบของ Bartlett อาศัยข้อสมมติเบื้องต้นที่ว่าความแปรปรวนของตัวอย่างทุกชุด $s_1^2, s_2^2, \ldots, s_k^2$ ต้องเป็นอิสระกัน แต่เนื่องจากสมาชิกในแนวทแยงมุมหลัก $s_{11}, s_{22}, \ldots, s_{pp}$ ของเมทริกซ์ความแปรปรวนร่วมของตัวอย่างมักมีความสัมพันธ์กัน ดังนั้น วิธีการทดสอบของ Bartlett จึงไม่เหมาะที่จะใช้กับในกรณีตัวแปรหลายตัว (Rencher, 2002 pp. 255)

2. การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม

กรณีของข้อมูลหลายตัวแปร(Multivariate case) การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม(Tests for equality of covariance matrices) สำหรับประชากร k ชุดได้รับความสนใจและมีการพัฒนาวิธีการทดสอบขึ้นเป็นจำนวนมาก ในจำนวนนี้วิธีการทดสอบที่ได้รับการยอมรับและใช้กันอย่างแพร่หลายเพื่อตรวจสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม คือ การทดสอบของ Box(1949, 1950) หรือที่เรียกว่า Box’s M test วิธีการนี้พัฒนามาจากเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) และอาศัยข้อตกลงเบื้องต้นของการแจกแจงปรกติหลายตัวแปร(Multivariate normality) ในที่นี้จึงได้นำเสนอการทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็นไว้ในช่วงต้น และจะกล่าวถึงวิธีทดสอบของ Box ในลำดับต่อไป ในส่วนท้ายของการทบทวนวรรณกรรมได้นำเสนอวิธีการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมด้วยวิธีการทางสถิติที่ไม่อิงพารามิเตอร์(Nonparametric method) ไว้ด้วย

กำหนดให้ $\mathbf{x}_{ij}$ มีการแจกแจงแบบ $\text{iid}\, N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),\ i = 1, 2, \ldots, k,\ \ j = 1, 2, \ldots, n_i$ จากประชากรชุดที่ i

สมมติฐานหลัก(H) และสมมติฐานรอง(A) ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม เป็นดังนี้

$$\mathrm{H}: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_k$$

$$\mathrm{A}: \ \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j \text{ อย่างน้อย 1 คู่ที่ } i \neq j,\ i, j = 1, 2, \ldots, k \tag{5}$$

กำหนดสัญลักษณ์ดังนี้

$$\mathbf{V}_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^t \ , \quad \overline{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$\mathbf{V} = \sum_{i=1}^{k} \mathbf{V}_i \ , \qquad \mathbf{S}_i = \frac{1}{n_i - 1} \sum_{i=1}^{k} \mathbf{V}_i \ , \qquad \mathbf{S} = \nu \mathbf{V} \ ,$$

$$\nu_i = n_i - 1 \ , \ \nu = \sum_{i=1}^{k} \nu_i \ , \quad n = \sum_{i=1}^{k} n_i$$

2.1 การทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น

การทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น(Likelihood ratio criterion) เป็นการทดสอบแรกที่เสนอโดย Mauchly(1940) วิธีการนี้อาศัย generalized variance และ trace ของเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง($\mathbf{S}$) โดยมีข้อจำกัดที่ว่า เมทริกซ์ความแปรปรวนร่วมของตัวอย่างแต่ละชุดจะต้องเป็นเมทริกซ์ไม่เอกฐาน(Non –singular matrix)

การทดสอบที่ใช้เกณฑ์อัตราส่วนน่าจะเป็น(Likelihood ratio criterion) อาศัยสถิติ λ ดังนี้

$$\lambda = \left(\prod_{i=1}^{k} |\mathrm{V}_i|^{n_i/2} / |\mathrm{V}|^{n/2} \right) \left(n^{pn/2} / \prod_{i=1}^{k} n_i^{pn_i/2} \right) \qquad (6)$$

สามารถปรับสถิติ λ ให้เป็นสถิติ M ได้ดังนี้

$$M = \frac{|\mathbf{S}_1|^{\nu_1/2} |\mathbf{S}_2|^{\nu_2/2} \cdots |\mathbf{S}_k|^{\nu_k/2}}{|\mathbf{S}_p|^{\nu/2}} \qquad (7)$$

โดยที่ $\mathbf{S}_p$ แทน Pooled sample covariance matrix ซึ่งหาได้จาก

$$\mathbf{S}_p = \frac{\sum_i \nu_i \mathbf{S}_i}{\sum_i \nu_i} \qquad (8)$$

ค่าสถิติ M นี้มีค่าอยู่ในช่วง 0 ถึง 1 ยิ่งค่าเข้าใกล้ 1 มากขึ้น โอกาสที่จะยอมรับสมมติฐานหลักที่ว่าเมทริกซ์ความแปรปรวนร่วมของประชากรทุกชุดเท่ากันก็จะยิ่งมากขึ้น และหากค่าสถิติ M มีค่าเข้าใกล้ 0 โอกาสที่จะปฏิเสธสมมติฐานหลักก็จะมีมากขึ้น นอกจากนี้หาก $|\mathbf{S}_1|,|\mathbf{S}_2|,...,|\mathbf{S}_k|$ มีค่าต่างกันมากขึ้น ค่าสถิติ M จะมีค่าเข้าใกล้ 0 (Rencher, 2002, pp. 256) ข้อจำกัดที่สำคัญประการหนึ่งของวิธีการทดสอบที่ใช้เกณฑ์อัตราส่วนภาวะน่าจะเป็นนี้ คือ ทุกๆ ค่า ν_i ต้องมีค่ามากกว่า p มิฉะนั้นจะทำให้มีบางค่า i ที่ทำให้ $|\mathbf{S}_i| = 0$ ซึ่งส่งผลให้ $M = 0$ ด้วย ทำให้วิธีการนี้ไม่สามารถนำไปประยุกต์ใช้ในกรณีของ High-dimensional data ได้

เมื่อแปลงสถิติ M เป็น $-2\ln M$ จะได้ สถิติทดสอบ T_1

$$T_1 = -2\ln M = \nu(k\ln|\mathbf{S}_p| - \textstyle\sum_i \ln|\mathbf{S}_i|) \qquad (9)$$

เรียกสถิติ T_1 (หรือ $-2\ln M$) นี้ว่า Box's M statistic และจะได้ว่าสถิติ T_1 นี้ในกรณีเฉพาะเมื่อ $p = 2, 3, 4, 5$ และ $k = 2, 3, ..., 10$ สามารถหาค่าวิกฤติของการทดสอบแม่นตรง(Exact test) ได้จาก Rencher (2002, ตาราง A.14) และ Lee, Chiang, and Krishnaiah (1977) และเมื่อแปลงสถิติ M ไปเป็นสถิติทดสอบ T_2 ที่มีการแจกแจงไคกำลังสองโดยประมาณ(เมื่อ $n_i \to \infty$) สามารถทำได้ดังนี้

$$T_2 = -2(1-c_1)\ln M \qquad (10)$$

โดยที่ $$c_1 = \frac{2p^2+3p-1}{6(p+1)(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n-k}\right) \qquad (11)$$

และ $$\ln M = \frac{1}{2}\sum_{i=1}^{k}\nu_i \ln|\mathbf{S}_i| - \frac{1}{2}\left(\sum_{i=1}^{k}\nu_i\right)\ln|\mathbf{S}_p| \qquad (12)$$

สถิติทดสอบ T_2 มีการแจกแจงโดยประมาณ คือ การแจกแจงไคกำลังสองด้วยองศาเสรี $\frac{1}{2}(k-1)p(p+1)$ และจะปฏิเสธสมมติฐานหลักเมื่อ $T_M > \chi^2_{\frac{1}{2}(k-1)p(p+1)}$

กรณีที่ขนาดตัวอย่างทุกชุดเท่ากัน จะได้ว่า $\nu_1 = \nu_2 = \cdots = \nu_k = \nu$ ค่า c_1 ใน (11) สามารถลดรูปได้เป็น

$$c_1 = \frac{(2p^2 + 3p - 1)(k+1)}{6k\nu(p+1)} \tag{13}$$

นอกจากนี้สถิติ M สามารถแปลงให้มีการแจกแจงโดยประมาณเป็นการแจกแจงเอฟได้ ดังนี้

c_1 หาได้จาก (13)

$$c_2 = \frac{(p-1(p+1)}{6(k-1)}\left[\sum_{i=1}^{k}(1/\nu_i^2) - 1/\nu^2\right]$$

$$a_1 = \frac{1}{2}(k-1)p(p+1)\ , \qquad a_2 = \frac{a_1 + 2}{\left|c_2 - c_1^2\right|}$$

$$b_1 = \frac{1 - c_1 - a_1/a_2}{a_1}\ , \qquad b_2 = \frac{1 - c_1 + 2/a_2}{a_2}$$

กรณี $c_2 > c_1^2$ จะได้ว่าสถิติทดสอบ $F = -2b_1 \ln M$ มีการแจกแจงโดยประมาณคือ F_{a_1, a_2} และกรณี $c_2 < c_1^2$ จะได้ว่าสถิติทดสอบ $F = \frac{-2a_2b_2 \ln M}{a_1(1 + 2b_2 \ln M)}$ มีการแจกแจงโดยประมาณคือ F_{a_1, a_2} ในที่นี้จะกำหนดให้วิธีการทดสอบนี้เป็น T_3 นั่นคือ

$$T_3 = \begin{cases} = -2b_1 \ln M & ,c_2 > c_1^2 \\ \dfrac{-2a_2b_2 \ln M}{a_1(1 + 2b_2 \ln M)} & ,c_2 < c_1^2 \end{cases} \tag{14}$$

ทั้งสองกรณีข้างต้น หากขนาดตัวอย่างทุกชุดเท่ากัน ค่า c_2 สามารถลดรูปได้ดังนี้

$$c_2 = \frac{(p-1)(p+2)(k^2 + k + 1)}{6k^2\nu^2}$$

วิธี Box’M Test นี้มีความไว(Sensitive) ต่อข้อมูลที่ไม่ได้แจกแจงปรกติ ดังนั้น หากข้อมูลไม่ได้มีการแจกแจงปรกติ ไม่ควรใช้วิธีนี้อีกทั้งวิธีนี้ยังให้กำลังของการทดสอบต่ำเมื่อตัวอย่างมีขนาดเล็กด้วย กล่าวคือ หากทดสอบด้วยวิธี Box’M แล้วการทดสอบไม่มีนัยสำคัญ ไม่จำเป็นที่เมทริกซ์ความแปรปรวนร่วมนั้นจะเท่ากันทุกกลุ่ม**(Cohen, 2008)** นอกจากนี้วิธี Box’M Test ยังไวเกินไปในกรณีของข้อมูลขนาดใหญ่ด้วยซึ่งในกรณีนี้ **Hahs-Vaughn(2016)** ได้แนะนำให้เลือกใช้ระดับนัยสำคัญในการทดสอบ(α) ที่ 0.001

2.2 การทดสอบที่อาศัยวิธีการของสถิติไม่อิงพารามิเตอร์

กรณีประชากร 2 ชุด กำหนดให้ $\mathbf{x}_{ij}$ มีการแจกแจงแบบ iid $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1,2$ $j = 1,2,\dots,n_i$ สมมติฐานหลักการทดสอบ คือ $\mathrm{H}: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ ภายใต้แจกแจงปรกติหลายตัวแปร (Multivariate normal distribution) การทดสอบที่อาศัยเกณฑ์อัตราส่วนภาวะน่าจะเป็น คือ

$$T_4 = \frac{|\mathbf{S}_1|^{\nu_1/2} |\mathbf{S}_2|^{\nu_2/2}}{|\mathbf{S}_p|^{(\nu_1+\nu_2)/2}} \qquad (15)$$

โดยที่ $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$,

และ $(\nu_1 + \nu_2)\mathbf{S}_p = \nu_1 \mathbf{S}_1 + \nu_2 \mathbf{S}_2$

เมื่อข้อมูลไม่ได้แจกแจงปรกติ(Non-normal distribution) สถิติทดสอบ T ใน (15) ยังคงใช้ในการทดสอบได้เพียงแต่ใช้วิธี Bootstrap ในการหาค่าวิกฤติ ตามขั้นตอนดังนี้ (Srivastava, 2002, pp. 600)

ขั้นตอนของวิธี Bootstrap

1) กำหนด $\mathbf{u}_j$ และ $\mathbf{v}_j$ (Standardized residuals)

$$\mathbf{u}_j = \mathbf{S}^{1/2}\mathbf{S}_1^{-1/2}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1), \quad j = 1,2,\dots,n_1$$

$$\mathbf{v}_j = \mathbf{S}^{1/2}\mathbf{S}_2^{-1/2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2), \quad j = 1,2,\dots,n_2$$

2) กำหนดตัวอย่างจากวิธี Bootstrap ดังนี้

$$\mathbf{u}_{11}^*,\ldots,\mathbf{u}_{1n_1}^*,\ldots,\mathbf{u}_{B1}^*,\ldots,\mathbf{u}_{Bn_1}^*$$

$$\mathbf{v}_{11}^*,\ldots,\mathbf{v}_{1n_1}^*,\ldots,\mathbf{v}_{B1}^*,\ldots,\mathbf{v}_{Bn_1}^*$$

3) คำนวณ

$$\nu_1\mathbf{S}_{i1}^* = \sum_{j=1}^{n_1}(\mathbf{u}_{ij}^*-\bar{\mathbf{u}}_i^*)(\mathbf{u}_{ij}^*-\bar{\mathbf{u}}_i^*)^t \;,\; i=1,2,\ldots,B$$

$$\nu_2\mathbf{S}_{i2}^* = \sum_{j=1}^{n_2}(\mathbf{v}_{ij}^*-\bar{\mathbf{v}}_i^*)(\mathbf{v}_{ij}^*-\bar{\mathbf{v}}_i^*)^t \;,\; i=1,2,\ldots,B$$

$$(\nu_1+\nu_2)\mathbf{S}_i^* = \nu_1\mathbf{S}_{i1}^* + \nu_2\mathbf{S}_{i2}^* \;,\; i=1,2,\ldots,B$$

$$T_i^* = \frac{\left|\mathbf{S}_{i1}^*\right|^{\nu_1/2}\left|\mathbf{S}_{i2}^*\right|^{\nu_2/2}}{\left|\mathbf{S}_i^*\right|^{(\nu_1+\nu_2)/2}} \;,\; i=1,2,\ldots,B$$

4) เรียงลำดับ T_i^* เพื่อหาค่าวิกฤติ(Significance point) T_α^* จาก Empirical distribution

5) ปฏิเสธสมมติฐานหลักของความเท่ากันของเมทริกซ์ความแปรปรวนร่วม 2 เมทริกซ์เมื่อ $T_4 > T_\alpha^*$

3. โปรแกรม SPSS กับการทดสอบความแปรปรวนร่วม

จากการศึกษาโปรแกรม IBM SPSS Statistics 20 (หรือเรียกสั้นๆว่าโปรแกรม SPSS) พบว่าสถิติ Box’M ในผลลัพธ์ของโปรแกรมคือ ค่า $-2\ln M$ ใน (9) และแปลงเป็นสถิติทดสอบ F ซึ่งเท่ากับ T_3 ใน (14)

โปรแกรม SPSS วิธี Box’M Test ปรากฏอยู่ใน 2 วิธีการ(Procedures) คือ **DISCRIMINANT** และ **MANOVA** เพื่อเป็นการแสดงผลลัพธ์จากการใช้วิธี Box’M Test ในที่นี้จะกล่าวถึงการใช้วิธี Box’M Test ภายใต้การวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) โดยจะแสดงคำสั่งที่ใช้และผลลัพธ์ที่ได้ ข้อมูลที่ใช้แสดงเป็นตัวอย่างนี้เป็นข้อมูลจากการทดลองปลูกข้าว 4 ชนิด (A, B, C และ D) ในแปลง 20 แปลง ข้าวแต่ละชนิดจะถูกสุ่มปลูกลงในแปลง 5 แปลง เมื่อเวลาผ่านไป 6

สัปดาห์ เก็บรวบรวมข้อมูลความสูงของต้น(X1) และ จำนวนต้นต่อกอ(X2) ได้ข้อมูลดังนี้ (Srivastava, 2002 pp.182)

Rice Varieties							
A		B		C		D	
X1	X2	X1	X2	X1	X2	X1	X2
58	4	49	7	54	6	58	4
62	6	55	6	52	6	55	4
60	7	48	4	51	7	51	5
54	6	52	4	58	5	49	6
58	6	49	7	61	6	52	4

คำสั่งที่ใช้ Analyze > General Linear Model > Multivariate...
ใน Options ให้เลือก Homogeneity tests

ผลของการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมโดยวิธี Box'M เป็นดังนี้

Box's Test of Equality of Covariance Matrices[a]

Box's M	7.973
F	.683
df1	9
df2	2933.711
Sig.	.725

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a. Design: Intercept + rice

จากผลลัพธ์ข้างต้น พบว่า ค่าสถิติ Box'M เท่ากับ 7.973 เมื่อแปลงเป็นสถิติทดสอบ F จะเท่ากับ 0.683 และค่าพีของการทดสอบสมมติฐาน(คำนวณจากค่าสถิติทดสอบ F) เท่ากับ 0.725 ซึ่ง

เมื่อกำหนดระดับนัยสำคัญไว้ที่ 0.001 จะถือว่าการทดสอบไม่มีนัยสำคัญ สรุปได้ว่าเมทริกซ์ความแปรปรวนร่วมของข้าว 4 พันธุ์นี้ไม่แตกต่างกันอย่างมีนัยสำคัญ

หากเป็นกรณีที่วิธี Box'M Test มีนัยสำคัญแล้ว อาจใช้วิธี Levene's Test(Stevens, 2009) ในการทดสอบความเท่ากันของความแปรปรวนทีละตัวแปรซึ่งโปรแกรม SPSS ได้ให้ผลลัพธ์ไว้ด้วยแล้ว ดังนี้

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
plant height	.681	3	16	.576
number of tillers	2.132	3	16	.136

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + rice

จากผลลัพธ์ของ Levene's Test ข้างต้นของการทดสอบความเท่ากันของความแปรปรวนของแต่ละตัวแปร จะเห็นว่า ผลการทดสอบความแปรปรวนของตัวแปรทั้งสอง คือ plant height (ความสูงของต้น) และ number of tillers(จำนวนต้นต่อกอ) มีค่าพีเท่ากับ 0.576 และ 0.136 ตามลำดับ(ซึ่งมากกว่า 0.001) จึงไม่มีนัยสำคัญ(หรือกล่าวได้ว่าความแปรปรวนของทุกกลุ่มในแต่ละตัวแปรเท่ากัน)

กรณีที่ตัวแปรตัวใดตัวหนึ่งจากผลการทดสอบ Levene's Test มีนัยสำคัญแสดงว่ามีอย่างน้อยหนึ่งกลุ่มของตัวแปรนั้นที่มีความแปรปรวนต่างไปจากกลุ่ม ให้พิจารณาการแปลงข้อมูลของตัวแปรนั้นเสียก่อน จากนั้นจึงใช้วิธี Box'M Test ต่อไป

กรณีของการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) หากใช้วิธี Box'M Test ในการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมแล้ว(และได้ทำการแปลงข้อมูลแล้ว) พบว่า ผลการทดสอบ Box'M ยังมีนัยสำคัญที่ระดับนัยสำคัญ 0.001 แสดงว่า ความแปรปรวนอย่างน้อยหนึ่งกลุ่มต่างไปจากกลุ่มอื่นมาก Tabachnick และ Fidell (2007) แนะนำให้เลือกใช้ผลจากสถิติทดสอบ Pillai's trace แทนการใช้ Wilk's lambda(เนื่องจากสถิติทดสอบ Wilk's lambda จะทำให้ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่กำหนดไว้มาก) ซึ่งโปรแกรม SPSS ได้ให้ผลลัพธ์ของสถิติทดสอบ Pillai's trace ไว้ด้วยแล้วในตารางเดียวกัน ดังตัวอย่างข้างล่างนี้

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.998	3160.057[b]	2.000	15.000	.000
	Wilks' Lambda	.002	3160.057[b]	2.000	15.000	.000
	Hotelling's Trace	421.341	3160.057[b]	2.000	15.000	.000
	Roy's Largest Root	421.341	3160.057[b]	2.000	15.000	.000
rice	Pillai's Trace	.707	2.916	6.000	32.000	.022
	Wilks' Lambda	.384	3.070[b]	6.000	30.000	.018
	Hotelling's Trace	1.369	3.194	6.000	28.000	.016
	Roy's Largest Root	1.166	6.217[c]	3.000	16.000	.005

a. Design: Intercept + rice

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

เนื่องจากกรณีนี้เป็นกรณีที่ผลการทดสอบ Box’M ไม่มีนัยสำคัญ ดังนั้นจึงใช้สถิติทดสอบ Wilk’s lambda ซึ่งให้ค่าพีเท่ากับ 0.018 แต่หากผลการทดสอบ Box’M มีนัยสำคัญที่ระดับ 0.001 ควรเลือกใช้สถิติทดสอบ Pillai’s trace ซึ่งให้ค่าพีเท่ากับ 0.022 แทน

บทที่ 3
ระเบียบวิธีการวิจัย

การวิจัยเรื่องการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม มีจุดประสงค์หลักในการศึกษาประสิทธิภาพของวิธี Box’M Test ซึ่งรวมถึงตัวสถิติทดสอบที่ใช้การแจกแจงไคกำลังสองและการแจกแจงเอฟในการประมาณการแจกแจงของตัวสถิติทดสอบดังกล่าวด้วย นอกจากนี้ยังได้ศึกษาการใช้โปรแกรม SPSS ในการทดสอบ Box’M ด้วย ดังนี้

1. การจำลองข้อมูล

การจำลองข้อมูลในการศึกษานี้ใช้โปรแกรม R version 3.6.1 เพื่อทำการจำลองข้อมูลภายใต้สมมติฐานหลักและสมมติฐานรอง โดยแบ่งการศึกษาออกเป็น 3 ตอน คือ (1) เป็นการศึกษาเพื่อเปรียบเทียบวิธีทดสอบ 3 วิธี คือ T_1, T_2 และ T_3 ภายใต้การแจกแจงแบบปรกติหลายตัวแปรและขนาดตัวอย่างเท่ากัน (2) เป็นการศึกษาเพื่อเปรียบเทียบวิธีการทดสอบ 2 วิธี คือ T_2 และ T_3 ภายใต้การแจกแจงแบบปรกติหลายตัวแปรและขนาดตัวอย่างไม่เท่ากัน และ(3) เป็นการศึกษาเพื่อเปรียบเทียบวิธีการทดสอบ 3 วิธี คือ T_2, T_3 และ T_4 ภายใต้การแจกแจงที่ไม่ใช่การแจกแจงปรกติ (Non-normality) โดย T_4 ใช้วิธีการ Bootstrap

1.1 การศึกษาประสิทธิภาพของวิธี Box’s M Test โดยเปรียบเทียบสถิติทดสอบ 3 ตัวดังนี้

(1) T_1 (Exact Test)

(2) T_2 (Approximate chi-square distribution)

(3) T_3 (Approximate F-distribution)

การศึกษาในส่วนนี้อาศัยการจำลองข้อมูลภายใต้การแจกแจงปรกติหลายตัวแปร (Multivariate normality, MVN) ที่มีค่าเฉลี่ยของทุกตัวแปรเท่ากับ 1 และเมทริกซ์ความแปรปรวน C_0 ภายใต้สมมติฐานหลัก และ Σ_k ภายใต้สมมติฐานรอง โดยมีสมมติฐานหลักคือ $\mathrm{H}:\Sigma_1=\Sigma_2=C_0$ เมื่อ $C_0=0.9I_p+0.1J_pJ_p'$, $J_p=(1,1,\ldots,1)'$ และสมมติฐานรองคือ $\mathrm{A}:\Sigma_1=\Sigma_2\ldots=\Sigma_{k-1}=C_0\neq\Sigma_k$ โดย Σ_k มี 3 ระดับ ได้แก่ $\Sigma_k^{(1)}=0.5I_p+0.5J_pJ_p'$, $\Sigma_k^{(2)}=0.3I_p+0.7J_pJ_p'$ และ $\Sigma_k^{(3)}=0.1I_p+0.9J_pJ_p'$ นั่นคือ $\Sigma_k^{(3)}$ จะแตกต่างไปจาก C_0 มากสุด รองลงไปคือ $\Sigma_k^{(2)}$ และ $\Sigma_k^{(1)}$ ตามลำดับ, จำนวนประชากร(k) เท่ากับ 2 และ 6 ชุด, จำนวนตัวแปร (p) เท่ากับ 2 และ 5, ขนาดตัวอย่างแต่ละชุด(n_i) เท่ากัน ได้แก่ $n_i=\{10,20\}$, ระดับนัยสำคัญ (α) เท่ากับ 0.05 และแต่ละเงื่อนไขดำเนินการทำซ้ำจำนวน 10,000 รอบ

1.2 การศึกษาประสิทธิภาพของวิธี Box's M Test โดยกำหนดขนาดตัวอย่างต่างกันและเปรียบเทียบสถิติทดสอบ 2 ตัว คือ

(1) T_2 (Approximate chi-square distribution)

(2) T_3 (Approximate F-distribution)

การศึกษาในส่วนนี้อาศัยการจำลองข้อมูลภายใต้การแจกแจงปรกติหลายตัวแปร โดยมีสมมติฐานเช่นเดียวกับใน 1.1, $k=\{2,6\}$, $p=\{2,5\}$, ขนาดตัวอย่าง ได้แก่ $(n_1,n_2)=(10,20)$ และ $(n_1,...,n_6)=(10,10,20,20,30,30)$, $\alpha=\{0.05, 0.01, 0.001\}$) และแต่ละเงื่อนไขดำเนินการทำซ้ำจำนวน 10,000 รอบ

1.3 การศึกษาประสิทธิภาพของวิธีการทดสอบแบบไม่อิงพารามิเตอร์และวิธี Box's M Test โดยเปรียบเทียบสถิติทดสอบ 3 ตัว คือ

(1) T_2 (Approximate chi-square distribution)

(2) T_3 (Approximate F-distribution)

(3) T_4 (Bootstrap Method)

การศึกษาส่วนนี้ได้จำลองข้อมูลภายใต้ข้อมูลที่มีการแจกแจงแบบทีหลายตัวแปร (Multivariate $t-$distribution) โดยใช้ฟังก์ชัน rmvt in Package mvtnorm และกำหนดองศาเสรี (df) เท่ากับ 5, 10, 15 และ 20, $k=2$, $p=\{2,5\}$, ขนาดตัวอย่างแต่ละชุด(n_i) เท่ากัน ได้แก่ $n_i=\{10,20\}$ และแต่ละเงื่อนไขดำเนินการทำซ้ำจำนวน 10,000 รอบ

2. ประสิทธิภาพของวิธีทดสอบ

ประสิทธิภาพของวิธีทดสอบ หมายถึง เกณฑ์ในการตัดสินว่า วิธีทดสอบใดที่ให้ค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1(Attained significance level, ASL) ใกล้เคียงกับค่าที่กำหนดไว้(Nominal value) และหากวิธีการทดสอบมีค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1(ASL) อยู่ในเกณฑ์ที่ยอมรับได้แล้ว วิธีทดสอบที่มีกำลังการทดสอบ(Empirical power) สูงกว่า จะถือว่ามีประสิทธิภาพดีกว่า

ในการศึกษาครั้งนี้ จะนิยามค่า ASL ไว้ดังนี้

$$\mathrm{ASL} = \frac{\text{number of } t_H > c}{m},$$

เมื่อ t_H แทนค่าสถิติทดสอบ T ที่ได้จากการจำลองข้อมูลภายใต้สมมติฐานหลัก

c แทนค่าวิกฤตของการทดสอบ

และ m แทนจำนวนรอบในการทำซ้ำของการจำลองข้อมูล ในที่นี้กำหนด $m = 10,000$

ที่ระดับนัยสำคัญ $\alpha = 0.05$ ค่า ASL มีการแจกแจงแบบทวินาม $b(10000, 0.05)$ และค่าส่วนเบี่ยงเบนมาตรฐานสามารถประมาณด้วย $\widehat{se}(\mathrm{ASL}) = \sqrt{0.05(0.95)/10,000} \simeq 0.00218$ เกณฑ์ที่ใช้ในการพิจารณาค่า ASL คือ $\alpha \pm 3\widehat{se}(\mathrm{ASL})$ ทำให้ได้ช่วงสำหรับพิจารณาค่า ASL คือ (0.043, 0.057) สำหรับที่ $\alpha = 0.01$ สามารถหาช่วงที่ใช้ในการพิจารณายอมรับค่า ASL ได้ในทำนองเดียวกันจะได้ช่วงพิจารณายอมรับค่า ASL คือ (0.007, 0.013) และที่ $\alpha = 0.001$ สามารถหาช่วงที่ใช้ในการพิจารณายอมรับค่า ASL ได้ในทำนองเดียวกันจะได้ช่วงพิจารณายอมรับค่า ASL คือ (0.000, 0.002) ทั้งนี้เกณฑ์ที่ใช้ข้างต้นยังใกล้เคียงกับเกณฑ์ของ Cochran (1954) ด้วย

กำลังของการทดสอบ(Empirical power) นิยามดังนี้

$$\text{Empirical power} = \frac{\text{number of } t_K > c}{m},$$

โดยที่ t_K แทนค่าสถิติทดสอบ T ที่ได้จากการจำลองข้อมูลภายใต้สมมติฐานรอง

c แทนค่าวิกฤตของการทดสอบ

และ m แทนจำนวนรอบในการทำซ้ำของการจำลองข้อมูล ในที่นี้กำหนด $m = 10,000$

บทที่ 4
ผลการวิจัย

ผลการศึกษา แบ่งเป็น 4 ตอน ดังนี้

ตอนที่ 1 เป็นการเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 3 วิธีในกรณีที่ขนาดของตัวอย่างทุกชุดเท่ากันที่ระดับนัยสำคัญ 0.05 ได้แก่ T_1 (Exact test), T_2 (Approximated with chi-squared distribution) และ T_3 (Approximated with $F-$distribution)

ตอนที่ 2 เป็นการเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 2 วิธี คือ T_2 และ T_3 ในกรณีที่ขนาดของตัวอย่างไม่เท่ากันที่ระดับนัยสำคัญ 0.05, 0.01 และ 0.001

ตอนที่ 3 เป็นการเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 3 วิธี คือ T_2 , T_3 และ T_4 ในกรณีจำนวนประชากร 2 ชุด, ขนาดของตัวอย่างเท่ากัน, การแจกแจงของประชากรคือ การแจกแจงแบบทีหลายตัวแปร(Multivariate t-distribution) ที่องศาเสรีต่างกัน ผลการศึกษาเป็นดังตารางที่ 8

ตอนที่ 4 เป็นการศึกษาการใช้โปรแกรม SPSS ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม

ตอนที่ 1 การศึกษาเพื่อเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 3 วิธี ได้แก่ T_1 (Exact test), T_2 (Approximated with chi-squared distribution) และ T_3 (Approximated with $F-$distribution) ในกรณีที่ขนาดของตัวอย่างทุกชุดเท่ากันและกำหนดระดับนัยสำคัญ(α) เท่ากับ 0.05 ผลการศึกษาเป็นดังตารางที่ 1 และ 2

ตารางที่ 1 ค่า ASL ของวิธีทดสอบ 3 วิธีเมื่อขนาดตัวอย่างเท่ากันและ $\alpha = 0.05$

k	p	n_i	T_1	T_2	T_3
2	2	10	0.0513	0.0513	0.0513
		20	0.0511	0.0508	0.0508
	5	10	0.0478	0.0580	0.0507
		20	0.0477	0.0490	0.0480
6	2	10	0.0507	0.0514	0.0508
		20	0.0504	0.0504	0.0504
	5	10	0.0488	0.0746	0.0567
		20	0.0514	0.0557	0.0518

ตารางที่ 1 เป็นผลการศึกษาภายใต้สมมติฐานหลัก $\mathrm{H}:\Sigma_1 = \ldots = \Sigma_k$ เป็นจริง กำหนดระดับนัยสำคัญ(α) เท่ากับ 0.05 จะเห็นว่า โดยภาพรวมทุกวิธีการทดสอบอยู่ในเกณฑ์ดี(เนื่องจากค่า ASL มีค่าใกล้เคียงระดับนัยสำคัญที่กำหนดไว้)

จากตารางที่ 1 เมื่อเพิ่มขนาดตัวอย่าง(n_i) จาก 10 เป็น 20 พบว่า ค่า ASL เปลี่ยนแปลงโดยลดลงเล็กน้อย และวิธีทดสอบทั้ง 3 วิธีนี้ให้ผลดีในกรณีที่จำนวนประชากร(k)เท่ากับ 2 ไม่ว่าจำนวนตัวแปร(p) จะเท่ากับ 2 หรือ 5 ก็ตาม ส่วนในกรณีที่ประชากรมีจำนวน 6 ชุดนั้น พบว่า T_1 และ T_3 ยังคงให้ผลดีเช่นเดียวกับกรณีประชากร 2 ชุด ส่วน T_2 จะให้ผลดีในกรณีของ 2 ตัวแปรแต่หากเพิ่มจำนวนตัวแปรเป็น 5 จะเห็นว่า T_2 ให้ผลด้อยกว่าอีก 2 วิธี

ผลการจำลองข้อมูลภายใต้สมมติฐานรองเพื่อศึกษากำลังของการทดสอบเป็นไปดังตารางที่ 2 โดยสมมติฐานรอง คือ $\mathrm{A}:\Sigma_1 = \ldots = \Sigma_{k-1} = C_0 \neq \Sigma_k = \Sigma_k^{(i)}$, $i = 1, 2, 3$ โดย $\Sigma_k^{(i)}$ มีสมาชิกในแนวทแยงมุมหลัก(diagonal elements) เท่ากับ 1(เช่นเดียวกับ $\Sigma_1,\ldots,\Sigma_{k-1}$) แต่สมาชิกนอกแนวทแยงมุมหลัก(off-diagonal elements) ของ $\Sigma_k^{(1)},\Sigma_k^{(2)}$ และ$\Sigma_k^{(3)}$ มีค่าเท่ากับ 0.5, 0.7 และ 0.9 ตามลำดับ ซึ่งต่างไปจาก C_0 ซึ่งมีสมาชิกนอกแนวทแยงมุมหลักเท่ากับ 0.1

เมื่อพิจารณากำลังของการทดสอบ 3 วิธี คือ T_1, T_2 และ T_3 ในตารางที่ 2 พบว่า ยิ่งความแตกต่างมากของเมทริกซ์ความแปรปรวนร่วมเพิ่มขึ้น กำลังของการทดสอบทั้ง 3 วิธีจะเพิ่มขึ้น

เมื่อเพิ่มขนาดตัวอย่างจาก 10 เป็น 20 พบว่า กำลังของการทดสอบทุกวิธีเพิ่มขึ้น นอกจากนี้เมื่อเพิ่มจำนวนตัวแปร(p) จาก 2 เป็น 5 พบว่า กำลังของการทดสอบทุกวิธีเพิ่มขึ้น แต่เมื่อจำนวนตัวแปรและขนาดตัวอย่างคงที่ กำลังของการทดสอบจะลดลงเมื่อจำนวนชุดของประชากรเพิ่มขึ้นจาก 2 เป็น 6 ชุด

โดยภาพรวมเมื่อขนาดตัวอย่างเท่ากัน พบว่า T_2 ให้กำลังของการทดสอบสูงกว่าวิธีอื่น รองลงไปคือ T_3 ซึ่งให้กำลังของการทดสอบน้อยกว่า T_2 เล็กน้อย และ T_1 ให้กำลังของการทดสอบต่ำกว่าวิธีอื่น

ตารางที่ 2 กำลังการทดสอบของวิธีทดสอบ 3 วิธีเมื่อขนาดตัวอย่างเท่ากันและ $\alpha = 0.05$

p	n_i	Σ_k	T_1	T_2	T_3
$k=2$					
2	10	$\Sigma_k^{(1)}$	0.0966	0.0966	0.0966
		$\Sigma_k^{(2)}$	0.2268	0.2267	0.2267
		$\Sigma_k^{(3)}$	0.7337	0.7335	0.7335
	20	$\Sigma_k^{(1)}$	0.1547	0.1808	0.1809
		$\Sigma_k^{(2)}$	0.4797	0.5249	0.5249
		$\Sigma_k^{(3)}$	0.9830	0.9865	0.9865
5	10	$\Sigma_k^{(1)}$	0.1111	0.1333	0.1188
		$\Sigma_k^{(2)}$	0.3142	0.3549	0.3298
		$\Sigma_k^{(3)}$	0.9526	0.9620	0.9559
	20	$\Sigma_k^{(1)}$	0.3291	0.3341	0.3297
		$\Sigma_k^{(2)}$	0.8827	0.8859	0.8831
		$\Sigma_k^{(3)}$	1.0000	1.0000	1.0000
$k=6$					
2	10	$\Sigma_k^{(1)}$	0.0807	0.0812	0.0808
		$\Sigma_k^{(2)}$	0.1556	0.1569	0.1556
		$\Sigma_k^{(3)}$	0.5317	0.5343	0.5319
	20	$\Sigma_k^{(1)}$	0.1262	0.1363	0.1363
		$\Sigma_k^{(2)}$	0.3613	0.3848	0.3842
		$\Sigma_k^{(3)}$	0.9686	0.9731	0.9731
5	10	$\Sigma_k^{(1)}$	0.1093	0.1586	0.1267
		$\Sigma_k^{(2)}$	0.2774	0.3574	0.3026
		$\Sigma_k^{(3)}$	0.8703	0.9127	0.8849
	20	$\Sigma_k^{(1)}$	0.2703	0.2820	0.2722
		$\Sigma_k^{(2)}$	0.8005	0.8104	0.8019
		$\Sigma_k^{(3)}$	1.0000	1.0000	1.0000

ตอนที่ 2 การศึกษาเพื่อการเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 2 วิธี คือ T_2 และ T_3 ในกรณีที่ขนาดของตัวอย่างไม่เท่ากันที่ระดับนัยสำคัญ 0.05, 0.01 และ 0.001 ผลการศึกษาเป็นดังตารางที่ 3-6

ในกรณีที่ขนาดตัวอย่างไม่เท่ากัน(ขนาดตัวอย่างไม่ได้เท่ากันทุกชุด) การศึกษาเพื่อเปรียบเทียบวิธีการทดสอบ 2 วิธี คือ T_2 และ T_3 ภายใต้สมมติฐานหลักเป็นจริงที่ระดับนัยสำคัญต่างๆ ผลปรากฏดังตารางที่ 3

ตารางที่ 3 ค่า ASL ของวิธีทดสอบ 2 วิธีที่ α เท่ากับ 0.05, 0.01 และ 0.001

k	p	ขนาดตัวอย่าง $(n_1,\ldots,n_k)$	T_2	T_3
			$\alpha = 0.05$	
2	2	(10, 20)	0.0455	0.0455
	5	(10, 20)	0.0606	0.0533
6	2	(10, 10, 20, 20, 30, 30)	0.0518	0.0507
	5	(10, 10, 20, 20, 30, 30)	0.0642	0.0528
			$\alpha = 0.01$	
2	2	(10, 20)	0.0105	0.0105
	5	(10, 20)	0.0107	0.0088
6	2	(10, 10, 20, 20, 30, 30)	0.0113	0.0111
	5	(10, 10, 20, 20, 30, 30)	0.0163	0.0125
			$\alpha = 0.001$	
2	2	(10, 20)	0.0011	0.0011
	5	(10, 20)	0.0018	0.0012
6	2	(10, 10, 20, 20, 30, 30)	0.0010	0.0009
	5	(10, 10, 20, 20, 30, 30)	0.0019	0.0013

จากตารางที่ 3 เป็นการศึกษากรณีที่ขนาดตัวอย่างไม่ได้เท่ากันทุกกลุ่ม พบว่า ค่า ASL ของวิธี T_3 ดีกว่าวิธี T_2 แต่ทั้งสองวิธีล้วนมีค่า ASL อยู่ในเกณฑ์ที่ยอมรับได้ทั้งในกรณีที่ α เท่ากับ 0.05, 0.01 และ 0.001

กำลังของการทดสอบของวิธี T_2 และ T_3 เป็นดังตารางที่ 4

ตารางที่ 4 กำลังการทดสอบของวิธีทดสอบ 2 วิธีเมื่อขนาดตัวอย่างไม่เท่ากันและ $\alpha = 0.05$

k	p	ขนาดตัวอย่าง $(n_1,...,n_k)$	T_2	T_3
$\Sigma_k = \Sigma_k^{(1)}$				
2	2	(10, 20)	0.1256	0.1254
	5	(10, 20)	0.1975	0.1805
6	2	(10, 10, 20, 20, 30, 30)	0.1843	0.1831
	5	(10, 10, 20, 20, 30, 30)	0.4207	0.3853
$\Sigma_k = \Sigma_k^{(2)}$				
2	2	(10, 20)	0.3547	0.3543
	5	(10, 20)	0.6036	0.5787
6	2	(10, 10, 20, 20, 30, 30)	0.5660	0.5647
	5	(10, 10, 20, 20, 30, 30)	0.9503	0.9387
$\Sigma_k = \Sigma_k^{(3)}$				
2	2	(10, 20)	0.8995	0.8993
	5	(10, 20)	0.9993	0.9991
6	2	(10, 10, 20, 20, 30, 30)	0.9990	0.9990
	5	(10, 10, 20, 20, 30, 30)	1.0000	1.0000

จากตารางที่ 4 พบว่า วิธี T_2 มีกำลังของการทดสอบสูงกว่าวิธี T_3 เล็กน้อย และเมื่อเมทริกซ์ความแปรปรวนร่วมมีความแตกต่างกันมากขึ้น ค่ากำลังของการทดสอบของทั้งสองวิธีจะเพิ่มขึ้นด้วย

เมื่อจำนวนตัวแปร(p) เพิ่มขึ้นจาก 2 เป็น 5 เห็นได้ชัดว่า กำลังของการทดสอบของทั้งสองวิธีจะเพิ่มขึ้นด้วย

ตารางที่ 4 เป็นการศึกษาเมื่อกำหนดระดับนัยสำคัญ 0.05 และเมื่อเปลี่ยนระดับนัยสำคัญที่ศึกษาเป็น 0.01 ผลปรากฏดังตารางที่ 5

ตารางที่ 5 กำลังการทดสอบของวิธีทดสอบ 2 วิธีเมื่อขนาดตัวอย่างไม่เท่ากันและ $\alpha = 0.01$

k	p	$(n_1,...,n_k)$	T_2	T_3
$\Sigma_k = \Sigma_k^{(1)}$				
2	2	(10, 20)	0.0398	0.0397
	5	(10, 20)	0.0605	0.0521
6	2	(10, 10, 20, 20, 30, 30)	0.0587	0.0583
	5	(10, 10, 20, 20, 30, 30)	0.2008	0.1736
$\Sigma_k = \Sigma_k^{(2)}$				
2	2	(10, 20)	0.1578	0.1571
	5	(10, 20)	0.3406	0.3122
6	2	(10, 10, 20, 20, 30, 30)	0.2939	0.2921
	5	(10, 10, 20, 20, 30, 30)	0.8357	0.8072
$\Sigma_k = \Sigma_k^{(3)}$				
2	2	(10, 20)	0.7754	0.7746
	5	(10, 20)	0.9955	0.9948
6	2	(10, 10, 20, 20, 30, 30)	0.9878	0.9875
	5	(10, 10, 20, 20, 30, 30)	1.0000	1.0000

จากตารางที่ 5 พบว่า เมื่อเปลี่ยนระดับนัยสำคัญจาก 0.05 เป็น 0.01 ค่ากำลังของการทดสอบยังคงให้ผลในลักษณะเดิม กล่าวคือ วิธี T_2 มีกำลังของการทดสอบสูงกว่า T_3 เล็กน้อย และเมื่อจำนวนตัวแปร(p) เพิ่มขึ้นจาก 2 เป็น 5 เห็นได้ชัดว่า กำลังของการทดสอบของทั้งสองวิธีจะเพิ่มขึ้นด้วย

เมื่อกำหนดระดับนัยสำคัญ(α) เท่ากับ 0.001 ผลปรากฏดังตารางที่ 6 ซึ่งพบว่า ค่ากำลังของการทดสอบยังคงให้ผลในลักษณะเดียวกันกับที่ α เท่ากับ 0.05 และ 0.01

ตารางที่ 6 กำลังการทดสอบของวิธีทดสอบ 2 วิธีเมื่อขนาดตัวอย่างไม่เท่ากันและ $\alpha = 0.001$

k	p	$(n_1,\ldots,n_k)$	T_2	T_3
$\Sigma_k = \Sigma_k^{(1)}$				
2	2	(10, 20)	0.0052	0.0052
	5	(10, 20)	0.0099	0.0081
6	2	(10, 10, 20, 20, 30, 30)	0.0093	0.0093
	5	(10, 10, 20, 20, 30, 30)	0.0547	0.0440
$\Sigma_k = \Sigma_k^{(2)}$				
2	2	(10, 20)	0.0375	0.0375
	5	(10, 20)	0.1172	0.1001
6	2	(10, 10, 20, 20, 30, 30)	0.0991	0.0976
	5	(10, 10, 20, 20, 30, 30)	0.5966	0.5505
$\Sigma_k = \Sigma_k^{(3)}$				
2	2	(10, 20)	0.5316	0.5309
	5	(10, 20)	0.9673	0.9605
6	2	(10, 10, 20, 20, 30, 30)	0.9181	0.9166
	5	(10, 10, 20, 20, 30, 30)	1.0000	1.0000

ตอนที่ 3 การศึกษาเพื่อเปรียบเทียบประสิทธิภาพของวิธีการทดสอบ 3 วิธี คือ T_2 , T_3 และ T_4 ในกรณีจำนวนประชากร 2 ชุด, ขนาดตัวอย่างแต่ละชุดเท่ากัน, กำหนดระดับนัยสำคัญเท่ากับ 0.05, การแจกแจงของประชากร คือ การแจกแจงแบบทีหลายตัวแปร(Multivariate $t-$distribution) ที่องศาเสรีต่างกัน ผลการศึกษาเป็นดังตารางที่ 7

ตารางที่ 7 ค่า ASL ของวิธีทดสอบ 3 วิธีภายใต้การแจกแจงแบบทีหลายตัวแปรที่องศาเสรี(df) ต่างกันเมื่อ $k = 2$ และ $\alpha = 0.05$

p	n_i	T_2	T_3	T_4
df = 5				
2	10	0.1519	0.1520	0.0243
	20	0.2240	0.2242	0.0306
5	10	0.2147	0.1963	0.0217
	20	0.3566	0.3519	0.0087
df = 10				
2	10	0.0826	0.0826	0.0278
	20	0.1077	0.1077	0.0391
5	10	0.1170	0.1043	0.0328
	20	0.1420	0.1388	0.0171
df = 15				
2	10	0.0710	0.0710	0.0315
	20	0.0751	0.0751	0.0436
5	10	0.0924	0.0821	0.0366
	20	0.1000	0.0981	0.0210
df = 20				
2	10	0.0602	0.0602	0.0356
	20	0.0707	0.0707	0.0469
5	10	0.0817	0.0716	0.0418
	20	0.0853	0.0833	0.0251

จากตารางที่ 7 จะเห็นว่า โดยภาพรวมทุกวิธี คือ T_2, T_3 และ T_4 ให้ผลอยู่ในเกณฑ์ที่ยอมรับไม่ได้ ทั้งนี้ T_2 และ T_3 ให้ค่า ASL สูงกว่าระดับนัยสำคัญที่กำหนดขณะที่ T_4 ให้ค่าต่ำเกินไป และอาจมีเฉพาะกรณีที่องศาเสรีของทีเท่ากับ 20, จำนวนตัวแปรเท่ากับ 2 และขนาดตัวอย่างเท่ากับ 20 เท่านั้นที่ T_4 ให้ค่า ASL อยู่ในเกณฑ์ที่ยอมรับได้ อย่างไรก็ตามโดยภาพรวมแล้วกรณีนี้ถือได้ว่ายัง

ไม่มีวิธีใดที่ให้ค่า ASL อยู่ในเกณฑ์ที่ยอมรับได้ ดังนั้น จึงไม่มีความจำเป็นในการพิจารณากำลังของการทดสอบต่อไป

ตอนที่ 4 เป็นการศึกษาการใช้โปรแกรม SPSS ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม

จากการศึกษาโปรแกรม SPSS ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมด้วยวิธี Box'M Test พบว่า วิธี Box'M Test ปรากฏอยู่ใน 2 วิธีการ(Procedures) คือ **MANOVA** และ **DISCRIMINANT** ดังนี้

คำสั่งที่ใช้ใน MANOVA

Analyze > General Linear Model > Multivariate

ใน **Options** ให้เลือก **Homogeneity tests**

คำสั่งที่ใช้ใน DISCRIMINANT

Analyze > Classify > Discriminant

ใน **Statistics** ให้เลือก **Box'M**

สถิติ Box'M ในผลลัพธ์(Output) ของโปรแกรม SPSS คือ ค่า $-2\ln M$ ใน (9) และแปลงเป็นสถิติทดสอบ F ซึ่งเท่ากับ T_3 ใน (14)

วิธี Box'M Test มีข้อเสียบางประการ ได้แก่ วิธี Box'M Test มีความไว(Sensitive) ต่อข้อมูลที่ไม่ได้แจกแจงปรกติ, มีกำลังของการทดสอบต่ำเมื่อตัวอย่างมีขนาดเล็ก และมีความไวเกินไปสำหรับชุดข้อมูลขนาดใหญ่ ดังนั้น ในการใช้วิธี Box'M Test จึงควรกำหนดระดับนัยสำคัญไว้ที่ 0.001(แทนที่จะกำหนดเป็น 0.05)

กล่าวโดยสรุป คือ วิธี Box'M Test เหมาะสมกับข้อมูลที่มีการแจกแจงปรกติหลายตัวแปร (Multivariate normality) และชุดข้อมูลไม่ควรมีขนาดใหญ่มาก ส่วนขนาดตัวอย่างของแต่ละกลุ่มควรเท่าหรือใกล้เคียงกัน ดังผลที่ได้ปรากฏในตอนที่ 1 ถึง 3

หากผลการทดสอบ Box'M Test มีนัยสำคัญ จำเป็นต้องทดสอบความเท่ากันของความแปรปรวนทีละตัวแปร วิธีที่สามารถใช้ในการทดสอบความเท่ากันของความแปรปรวนทีละตัวแปรได้แก่ วิธี Levene's Test ซึ่งจะทำให้บอกได้ว่า ตัวแปรใดบ้างที่มีความแปรปรวนระหว่างกลุ่มแตกต่างกัน และเมื่อตรวจพบตัวแปรที่มีปัญหาความแปรปรวนไม่เท่ากัน(Heterogeneity of variance) แล้ว อาจแก้ปัญหาโดยใช้วิธีการแปลงข้อมูลให้มีความแปรปรวนเท่ากันเสียก่อน จากนั้นจึงใช้วิธี Box’M Test ทดสอบอีกครั้งหนึ่ง

สำหรับการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) หากใช้วิธี Box’M Test ในการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมแล้ว(และได้ทำการแปลงข้อมูลแล้ว) พบว่า ผลการทดสอบ Box’M ยังมีนัยสำคัญที่ระดับนัยสำคัญ 0.001 แสดงว่า ความแปรปรวนอย่างน้อยหนึ่งกลุ่มต่างไปจากกลุ่มอื่นมาก แนะนำให้เลือกใช้ผลจากสถิติทดสอบ Pillai’s trace แทนการใช้ Wilk’s lambda เนื่องจากสถิติทดสอบ Wilk’s lambda จะทำให้ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่กำหนดไว้มาก(Tabachnick และ Fidell, 2007) ซึ่งโปรแกรม SPSS ได้ให้ผลลัพธ์ของสถิติทดสอบ Pillai’s trace ไว้ด้วยแล้ว ส่วนในการวิเคราะห์จำแนกกลุ่ม (Discriminant analysis) นั้นหากไม่มีค่านอกเกณฑ์(Outliers)ในชุดข้อมูล สามารถใช้ Discriminant function analysis ได้เนื่องจากวิธีการดังกล่าวมีความแกร่งต่อความไม่เท่ากันของความแปรปรวน

บทที่ 5
สรุปผล อภิปรายผลและข้อเสนอแนะ

การวิจัยเรื่องการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของวิธี Box'M Test ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม โดยการศึกษาครอบคลุมทั้งสถิติทดสอบ Box'M และ การประมาณการแจกแจงด้วยการแจกแจงไคกำลังสองและการแจกแจงเอฟ และเพื่อให้คำแนะนำสำหรับนักวิจัยที่ใช้โปรแกรม SPSS ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมด้วยวิธี Box'M Test

1. สรุปผลการวิจัย

1) เมื่อข้อมูลมีการแจกแจงปรกติหลายตัวแปร(Multivariate normality) และขนาดตัวอย่างแต่ละชุดเท่าหรือใกล้เคียงกัน วิธี Box'M Test(T_1) ในการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมมีประสิทธิภาพดี และเมื่อแปลงสถิติ Box'M ไปเป็นสถิติทดสอบ T_2 ซึ่งประมาณด้วยการแจกแจงไคกำลังสอง(Chi-squared distribution) หรือแปลงไปเป็นสถิติทดสอบ T_3 ซึ่งประมาณด้วยการแจกแจงเอฟ($F-$distribution) ก็ตาม จะมีประสิทธิภาพดีเช่นกัน

จำนวนชุดของประชากร จำนวนตัวแปร และขนาดตัวอย่าง ล้วนมีผลต่อประสิทธิภาพของวิธี T_1, T_2 และ T_3 โดยประสิทธิภาพของวิธีทดสอบจะลดลงเมื่อจำนวนชุดของประชากรเพิ่มขึ้น (จากประชากร 2 ชุด เป็น 6 ชุด) ประสิทธิภาพของวิธีทดสอบจะเพิ่มขึ้นเมื่อเพิ่มจำนวนตัวแปร(จาก 2 เป็น 5 ตัวแปร) และเมื่อเพิ่มขนาดตัวอย่าง(จาก 10 เป็น 20) ประสิทธิภาพของวิธีทดสอบจะเพิ่มขึ้นเช่นกัน

ในกรณีที่ขนาดตัวอย่างไม่เท่ากัน(ขนาดตัวอย่างไม่ได้เท่ากันทุกชุด) จะไม่สามารถหาค่าวิกฤติของการทดสอบแม่นตรง(Exact test) ได้ อย่างไรก็ตาม การใช้วิธี T_2 และ T_3 ให้ผลได้ดีเช่นกันและถึงแม้ T_3 จะให้กำลังของการทดสอบต่ำกว่า T_2 แต่ก็ไม่ได้ต่างกันมากนักในสถานการณ์ที่ศึกษา

2) เมื่อข้อมูลมีการแจกแจงแบบทีหลายตัวแปร(Multivariate $t-$distribution) วิธีทดสอบ T_4 ซึ่งเป็นวิธีทดสอบที่ไม่อิงพารามิเตอร์โดยใช้วิธี Bootstrap ในการหาค่าวิกฤตินั้นให้ผลไม่อยู่ในเกณฑ์ที่ยอมรับได้ และวิธีที่อิงพามิเตอร์ ได้แก่ วิธี T_2 และ T_3 และ T_4 ในสถานการณ์นี้ก็ให้ผลไม่อยู่ในเกณฑ์ที่ยอมรับได้เช่นกัน ดังนั้น สรุปได้ว่า จากวิธีการที่ศึกษาในงานวิจัยนี้ ยังไม่มีวิธีการที่เหมาะสมสำหรับทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมเมื่อข้อมูลไม่ได้แจกแจงปรกติและยังต้องอาศัยการพัฒนาวิธีการทดสอบซึ่งอาจเป็นวิธีการทดสอบที่ไม่อิงพารามิเตอร์ต่อไป

3) จากการศึกษาวิธีทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วมโดยวิธี Box’M Test ในโปรแกรม SPSS พบว่า วิธี Box’M Test ปรากฏอยู่ใน 2 วิธีการ(Procedures) คือ **MANOVA** และ **DISCRIMINANT** ดังนี้

คำสั่งที่ใช้ใน MANOVA

Analyze > General Linear Model > Multivariate

ใน **Options** ให้เลือก **Homogeneity tests**

คำสั่งที่ใช้ใน DISCRIMINANT

Analyze > Classify > Discriminant

ใน **Statistics** ให้เลือก **Box’M**

สถิติ Box’M ในผลลัพธ์(Output) ของโปรแกรม SPSS คือ ค่า $-2\ln M$ ซึ่งเท่ากับ T_1 ใน (9) และแปลงไปเป็นสถิติทดสอบ F ซึ่งเท่ากับ T_3 ใน (14) ค่าพี(P-valuue) ในผลลัพธ์จึงเป็นของสถิติทดสอบ F ดังกล่าว

เนื่องจากวิธี Box’M Test มีข้อเสียบางประการ ได้แก่ วิธี Box’M Test มีความไว (Sensitive) ต่อข้อมูลที่ไม่ได้แจกแจงปรกติ, มีกำลังของการทดสอบต่ำเมื่อตัวอย่างมีขนาดเล็ก และมีความไวเกินไปสำหรับชุดข้อมูลขนาดใหญ่ ดังนั้น ในการเลือกใช้วิธี Box’M Test จึงควรกำหนดระดับนัยสำคัญไว้ที่ 0.001(แทนที่จะกำหนดเป็น 0.05)

หากผลการทดสอบ Box’M Test ใน SPSS มีนัยสำคัญ จำเป็นต้องทดสอบความเท่ากันของความแปรปรวนทีละตัวแปร วิธีที่สามารถใช้ในการทดสอบความเท่ากันของความแปรปรวนทีละตัวแปรและโปรแกรม SPSS ได้ให้ผลลัพธ์ไว้แล้ว ได้แก่ วิธี Levene’s Test ซึ่งจะทำให้บอกได้ว่าตัวแปรใดบ้างที่มีความแปรปรวนระหว่างกลุ่มแตกต่างกัน และเมื่อตรวจพบตัวแปรที่มีปัญหาความแปรปรวนไม่เท่ากัน(Heterogeneity of variance) แล้ว อาจแก้ปัญหาโดยใช้วิธีการแปลงข้อมูลให้มีความแปรปรวนเท่ากันเสียก่อน จากนั้นจึงใช้วิธี Box’M Test ทดสอบอีกครั้งหนึ่ง

สำหรับการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) หากใช้วิธี Box’M Test ในการตรวจสอบข้อสมมติเบื้องต้นเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมแล้ว(และได้ทำการแปลงข้อมูลแล้ว) พบว่า ผลการทดสอบ Box’M ยังมีนัยสำคัญที่ระดับนัยสำคัญ 0.001 แสดงว่า ความแปรปรวนอย่างน้อยหนึ่งกลุ่มต่างไปจากกลุ่มอื่นมาก แนะนำให้เลือกใช้ผลจากสถิติทดสอบ Pillai’s trace แทนการใช้ Wilk’s lambda เนื่องจากสถิติทดสอบ Wilk’s lambda จะทำให้ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่กำหนดไว้มาก(Tabachnick และ Fidell, 2007) ซึ่งโปรแกรม

SPSS ได้ให้ผลลัพธ์ของสถิติทดสอบ Pillai’s trace ไว้ด้วยแล้ว ส่วนในการวิเคราะห์จำแนกกลุ่ม (Discriminant analysis) นั้นหากไม่มีค่านอกเกณฑ์(Outliers)ในชุดข้อมูล สามารถใช้ Discriminant function analysis ได้เนื่องจากวิธีการดังกล่าวมีความแกร่งต่อความไม่เท่ากันของความแปรปรวน

2. อภิปรายผล

1) วิธี Box’M Test เป็นวิธีที่ให้กำลังของการทดสอบต่ำเมื่อขนาดตัวอย่างน้อยซึ่งเห็นได้จากการศึกษาครั้งนี้ที่กำหนดขนาดตัวอย่างเท่ากับ 10 และเมทริกซ์ความแปรปรวนร่วมมีความแตกต่างกันน้อย ผลที่พบในการศึกษาครั้งนี้สอดคล้องกับงานของ Cohen(2008) แต่เมื่อกำหนดขนาดตัวอย่างเท่ากับ 20 วิธี Box’M Test มีความไวต่อข้อมูลอย่างเห็นได้ชัดซึ่งสอดคล้องกับคำแนะนำของ Hahs-Vaughn(2016) ที่แนะนำให้ใช้ระดับนัยสำคัญที่ 0.001 ในการทดสอบด้วยวิธี Box’M Test

2) กรณีที่ข้อมูลไม่ได้แจกแจงปรกตินั้น วิธีการทดสอบที่ศึกษาในงานวิจัยนี้ยังให้ผลไม่ดีเท่าที่ควร และต้องการการพัฒนาต่อไป แนวทางของการพัฒนาที่อาจใช้ได้แก่ วิธีการที่ไม่อิงพารามิเตอร์ เช่น งานของ Zhang และ Boos(1989) ซึ่งได้ศึกษาวิธีการ Bootstrap เป็นต้น

3. ข้อเสนอแนะ

สำหรับงานวิจัยในครั้งต่อไป เรื่องที่น่าสนใจได้แก่ การพัฒนาวิธีทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม กรณีที่ข้อมูลไม่ได้แจกแจงปรกติ ซึ่งอาจใช้แนวทางของวิธีการทดสอบที่ไม่อิงพารามิเตอร์ เช่น วิธี Bootstrap

นอกจากนี้ยังอาจศึกษาการวิเคราะห์ความแปรปรวนหลายตัวแปร(MANOVA) เพื่อหาวิธีทดสอบที่มีความแกร่ง(Robustness) ต่อข้อมูลที่ไม่ได้แจกแจงปรกติ โดยอาจศึกษาเปรียบเทียบวิธีทดสอบที่โปรแกรม SPSS ให้ไว้ ได้แก่ วิธี Pillai’s Trace, Wilk’s Lambda, Hotelling’s Trace และ Roy’s Largest Root เพื่อจะได้ให้ข้อสรุปหรือคำแนะนำที่เป็นประโยชน์ต่อการเลือกวิธีการทดสอบสำหรับการวิเคราะห์ข้อมูลในงานวิจัยในปัจจุบันที่มีความหลากหลายมากขึ้น

บรรณานุกรม

Bartlett, M.S. (1937, May). Properties of sufficiency and statistical tests. In Proc. R. Soc. Lond. A (Vol. 160, No. 901, pp. 268-282). The Royal Society.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36: 317–346.

Box, G.E.P. (1950). Problems in the analysis of growth and wear curves. Biometrics, 6(4), 362-389.

Cai, T.T., & Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. The Annals of Statistics, 39(3), 1496-1525.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. Biometrics, 10(4), 417-451.

Cohen, B.H. (2008). Explaining psychological statistics. John Wiley & Sons.

Gupta, A.K., & Bodnar, T. (2014). An exact test about the covariance matrix. Journal of Multivariate Analysis, 125, 176-189.

Hahs-Vaughn, D. (2016). Applied Multivariate Statistical Concepts. Taylor & Francis.

John, S. (1971). Some optimal multivariate tests. Biometrika, 58(1), 123-127.

Ledoit, O., & Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. The Annals of statistics, 30(4), 1081-1102.

Lee, J.C., Chang, T.C., & Krishnaiah, P. R. (1975). Approximation to the distributions of the likelihood ratio statistics for testing certain structures on the covariance matrices of real multivariate normal populations. In: Krishnaiah, P. R. (Ed.). Multivariate Analysis. Amsterdam: North-Holland, 1977. v.4, p.105-118.

Lix, L.M., Keselman, J.C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. Review of educational research, 66(4), 579-619.

Mauchly, J.W. (1940). Significance test for sphericity of a normal n-variate distribution. The Annals of Mathematical Statistics, 11(2), 204-209.

Nagao, H. (1973). On some test criteria for covariance matrix. The Annals of Statistics, 700-709.

Rencher, A.C. (2002). Methods of Multivariate Analysis. Wiley Series in Probability and Statistics.

Srivastava, M.S. (2002). Methods of multivariate statistics. Wiley Series in Probability and Statistics.

Stevens, J.P. (2009). Applied multivariate statistics for the social sciences (5th ed.). New York, NY, US: Routledge/Taylor & Francis Group.

Tabachnick, B.G., and Fidell, L.S. (2007). Using Multivariate statistics (5th ed.). Boston, MA: Allyn and Bacon.

ประวัติผู้วิจัย

ประวัติผู้วิจัย

หัวหน้าโครงการ

1. ชื่อ สกุล ผศ.ดร.คณาวุฒิ เจียมวัฒนพงศ์

2. ตำแหน่งปัจจุบัน ผู้ช่วยศาสตราจารย์ประจำคณะศิลปศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

3. หน่วยงานที่สามารถติดต่อได้

คณะศิลปศาสตร์ ห้อง 6206
มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ (ศาลายา)
96 หมู่ 3 ตำบลศาลายา อำเภอพุทธมณฑล นครปฐม 73170
โทร. 086-556-5941
Email: knavoot.jia@rmutr.ac.th, tuistat10@hotmail.com

4. ประวัติการศึกษา

ปริญญาเอก สถาบันบัณฑิตพัฒนบริหารศาสตร์ Ph.D.(Statistics) 2559
ปริญญาโท สถาบันบัณฑิตพัฒนบริหารศาสตร์ วท.ม.(สถิติประยุกต์) 2545
ปริญญาตรี มหาวิทยาลัยมหิดล วศ.บ.(วิศวกรรมเคมี) 2537

5. สาขาวิชาการที่มีคามชำนาญพิเศษ

High-dimensional Data Analysis, Experimental Design, Statistics Education

6. ประสบการณ์ที่เกี่ยวข้องกับการบริหารงานวิจัย

(1) การพัฒนาแผนการสอนวิชาสถิติที่เสริมทักษะการอ่านภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
ปีที่ได้รับการสนับสนุน 2551
หน่วยงานที่ให้ทุน มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

(2) ปัญหาการเรียนการสอนวิชาสถิติระดับปริญญาตรี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
ปีที่ได้รับการสนับสนุน 2552
หน่วยงานที่ให้ทุน มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

(3) การทดสอบค่าเฉลี่ยประชากร k ชุดเมื่อความแปรปรวนของประชากรไม่เท่ากัน
ปีที่ได้รับการสนับสนุน 2561
หน่วยงานที่ให้ทุน มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

ผู้ร่วมวิจัย

1. ชื่อ สกุล นางนิศานาถ อิงคดาภา

2. ตำแหน่งปัจจุบัน อาจารย์ประจำคณะศิลปศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

3. หน่วยงานที่สามารถติดต่อได้

คณะศิลปศาสตร์ ห้อง 6206
มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ (ศาลายา)
96 หมู่ 3 ตำบลศาลายา อำเภอพุทธมณฑล นครปฐม 73170
โทร. 082-087-4073
Email: nisanad.ing@rmutr.ac.th, nisanad@hotmail.com

4. ประวัติการศึกษา

ปริญญาโท Syracuse University, New York, U.S.A. M.S.(Information Science 2543
ปริญญาโท มหาวิทยาลัยธรรมศาสตร์ ศศม.(ภาษาศาสตร์) 2536
ปริญญาตรี มหาวิทยาลัยธรรมศาสตร์ สาขานิติศาสตร์ 2548
ปริญญาตรี จุฬาลงกรณ์มหาวิทยาลัย อบ.(ภาษาอังกฤษ) 2531

5. สาขาวิชาการที่มีคามชำนาญพิเศษ

การประมวลผลภาษาธรรมชาติ ภาษาศาสตร์ และ ภาษาอังกฤษ

6. ประสบการณ์ที่เกี่ยวข้องกับการบริหารงานวิจัย

(1) ระบบถามตอบปัญหาอัตโนมัติเรื่องกฎและระเบียบ
ปีที่ได้รับการสนับสนุน 2552
หน่วยงานที่ให้ทุน สำนักงานคณะกรรมการวิจัยแห่งชาติ

(2) โปรแกรมช่วยเขียนจดหมายภาษาอังกฤษทางธุรกิจ
ปีที่ได้รับการสนับสนุน 2558
หน่วยงานที่ให้ทุน มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์