PBRU SCIENCE JOURNAL
2026.23(1) : 17-29

การเปรียบเทียบประสิทธิภาพของสถิติทดสอบที่อิงวิธีการสุ่มตัวอย่างซ้ำสำหรับแผนแบบบล็อกสมบูรณ์เชิงสุ่มภายใต้ข้อมูลที่มีการแจกแจงล็อกนอร์มัล

A Comparison of the Efficiency of Resampling-Based Test Statistics for Randomized Complete Block Designs under Log-Normal Distributed Data

ธนพร เพ็งรักษ์ และ บำรุงศักดิ์ เผื่อนอารีย์*

Thanaporn Pengrak and Bumrungsak Phuenaree*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

***Corresponding author; E-mail:** bumrungsak@buu.ac.th

Received: 04 December 2025 /Revised: 07 January 2026 /Accepted: 09 February 2026

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของสถิติทดสอบที่อิงวิธีการสุ่มตัวอย่างซ้ำสำหรับแผนแบบบล็อกสมบูรณ์เชิงสุ่ม ได้แก่ สถิติทดสอบเอฟแบบบูตสแทร็ป และสถิติทดสอบเอฟแบบการเรียงสับเปลี่ยน โดยเปรียบเทียบกับวิธีพื้นฐานอีกสองวิธี ได้แก่ สถิติทดสอบเอฟ และ สถิติทดสอบฟรายด์แมน ทำการศึกษาภายใต้ข้อมูลที่มีการแจกแจงล็อกนอร์มัล กำหนดจำนวนทรีตเมนต์ที่ทำการศึกษา คือ 3 ทรีตเมนต์ จำนวนบล็อกที่ทำการศึกษา คือ 3, 5, 7 และ 10 บล็อก ระดับนัยสำคัญในการทดสอบเท่ากับ 0.05 ผลการศึกษาพบว่า สถิติทดสอบเอฟ สถิติทดสอบฟรายด์แมน และสถิติทดสอบเอฟแบบการเรียงสับเปลี่ยนสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกกรณีที่ทำการศึกษา แต่สถิติทดสอบเอฟแบบบูตสแทร็ปสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อ จำนวนบล็อกตั้งแต่ 5 ขึ้นไป เมื่อพิจารณากำลังการทดสอบพบว่า สถิติทดสอบเอฟมีกำลังการทดสอบสูง เมื่อจำนวนบล็อกเท่ากับ 3 แต่เมื่อจำนวนบล็อกเพิ่มมากขึ้น สถิติทดสอบเอฟแบบบูตสแทร็ป มีกำลังการทดสอบสูงที่สุดเกือบทุกกรณี

คำสำคัญ: การจำลองแบบมอนติคาร์โล แผนการทดลองแบบบล็อกสมบูรณ์เชิงสุ่ม การควบคุมความผิดพลาดแบบที่หนึ่ง กำลังการทดสอบ

Abstract

This study aimed to evaluate the performance of resampling-based test statistics for the randomized complete block design. The tests considered were the bootstrap F-test and the permutation F-test, and they were compared with two traditional methods: the F-test and the Friedman test. The experiments were conducted under log-normal distributed data. The design included three treatments, and the numbers of blocks considered were 3, 5, 7, and 10. The significance level was set at 0.05. The results showed that the F-test, the Friedman test, and the permutation F-test controlled the Type I error in all scenarios, while the bootstrap F-test controlled the Type I error only when there were at least five blocks. For statistical power, the F-test had the highest power when there were three blocks. However, when the number of blocks increased, the bootstrap F-test provided the highest power in most cases.

Keywords: Monte Carlo simulation, Randomized complete block designs, Type I error control, Power of the test.

บทนำ

ในการวิจัยเชิงปริมาณ การเลือกใช้สถิติทดสอบให้เหมาะสมกับลักษณะข้อมูลเป็นขั้นตอนที่มีความสำคัญอย่างยิ่ง เนื่องจากจะส่งผลโดยตรงต่อความถูกต้อง และความน่าเชื่อถือของผลการวิจัย การทดสอบสมมติฐานสำหรับการเปรียบเทียบค่าเฉลี่ยประชากรมากกว่าสองกลุ่ม มักนิยมใช้วิธีการทางสถิติที่เรียกว่า การวิเคราะห์ความแปรปรวน ด้วยสถิติทดสอบเอฟ ซึ่งเป็นการทดสอบแบบอิงพารามิเตอร์ที่มีหนึ่งในข้อสมมุติเบื้องต้นที่สำคัญ คือ ข้อมูลต้องมีการแจกแจงปรกติ อย่างไรก็ตาม ในงานวิจัยบางเรื่อง ข้อมูลที่นำมาวิเคราะห์อาจไม่ได้เป็นไปตามข้อสมมุติเบื้องต้นดังกล่าว

การวิเคราะห์ความแปรปรวนสำหรับแผนการทดลองแบบบล็อกสมบูรณ์เชิงสุ่ม (Randomized complete block design: RCBD) ซึ่งเป็นแผนการทดลองที่นิยมใช้เมื่อมีความแตกต่างของหน่วยทดลอง โดยออกแบบให้หน่วยทดลองที่คล้ายคลึงกันถูกจัดรวมเป็นบล็อก อย่างไรก็ตาม แม้แผนการทดลองจะถูกออกแบบอย่างเหมาะสม ข้อมูลที่ได้ก็อาจไม่สอดคล้องกับข้อสมมุติเบื้องต้นในเรื่องของการแจกแจงปรกติของข้อมูลที่ได้ ตัวอย่างเช่น ในงานวิจัยของ Nedelkov et al.[1] ได้ศึกษาพฤติกรรมการกินอาหารของสัตว์ และใช้แผนการทดลองแบบบล็อกสมบูรณ์เชิงสุ่ม โดยที่ตัวแปรตามบางตัวมีการแจกแจงแบบล็อกนอร์มัล ได้แก่ ข้อมูลปริมาณอาหารที่สัตว์กิน (กรัม) อัตราการกิน (กรัม/นาที) เป็นต้น นอกจากนี้ยังมีข้อมูลที่ไม่ได้เกิดจากการทดลองแต่เป็นข้อมูลที่มีความสำคัญ ซึ่งเมื่อนำมาศึกษาแล้วพบว่า ไม่มีการแจกแจงปรกติ ตัวอย่างเช่น ข้อมูลเกี่ยวกับมลพิษทางอากาศ ได้แก่ ฝุ่นละอองขนาดเล็ก (Particulate matter) PM2.5, PM10 เป็นต้น โดยข้อมูลดังกล่าวสำหรับในบางพื้นที่มีการแจกแจงล็อกนอร์มัล[2-4] ซึ่งเป็นตัวอย่างของข้อมูลที่ทำให้การใช้สถิติอิงพารามิเตอร์ไม่ใช่วิธีการที่เหมาะสม

จากปัญหาดังกล่าว ทำให้เกิดความสนใจในการใช้สถิติแบบไม่อิงพารามิเตอร์มากขึ้น เนื่องจากไม่มีข้อสมมุติเบื้องต้นเกี่ยวกับการแจกแจงของข้อมูล สำหรับการทดสอบความแตกต่างระหว่างค่ากลางมากกว่า 2 กลุ่มในกรณีนี้ วิธีการทดสอบที่ใช้แพร่หลาย คือ การทดสอบฟรายด์แมน (Friedman test) ซึ่งใช้สำหรับกรณีที่ตัวอย่างไม่เป็นอิสระต่อกันและอยู่ในมาตราอันดับ นอกจากการทดสอบฟรายด์แมนยังมีงานวิจัยของ O'Shaughnessy et al.[5] ได้ประยุกต์ใช้วิธีบูตสแทร็ปร่วมกับสถิติทดสอบเอฟ ซึ่งวิธีบูตสแทร็ปนี้ช่วยให้สามารถประมาณการค่าสถิติ เช่น ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน หรือค่าพี (P-value) ที่ใช้ในการทดสอบสมมติฐาน ของประชากรจากข้อมูลตัวอย่างได้โดยไม่ต้องอ้างอิงถึงรูปแบบการแจกแจงของประชากร นอกจากนั้น บำรุงศักดิ์ เผื่อนอารีย์ และ โชคศรัญญ์ สิริลีพันธุ์[6] ได้ใช้สถิติทดสอบการเรียงสับเปลี่ยนสำหรับทดสอบความแตกต่างระหว่างค่ากลางมากกว่า 2 ประชากร ซึ่งเป็นอีกหนึ่งวิธีที่มีประสิทธิภาพในการวิเคราะห์ข้อมูลเชิงสถิติในกรณีที่ข้อมูลไม่มีการแจกแจงปรกติ การใช้วิธีบูตสแทร็ป และการเรียงสับเปลี่ยนนี้ช่วยให้นักวิจัยสามารถทำการทดสอบสมมติฐานและวิเคราะห์ข้อมูลได้อย่างยืดหยุ่นโดยไม่ต้องคำนึงถึงการแจกแจงที่เฉพาะเจาะจง ซึ่งเป็นเหตุผลที่วิธีเหล่านี้ได้รับความนิยมในการวิเคราะห์ข้อมูลทางสถิติโดยเฉพาะในกรณีที่ข้อมูลไม่ตรงตามสมมติฐานเบื้องต้น และจากงานวิจัยของ วนิดา พงษ์ศักดิ์ชาติ และ นพรัตน์ แป้นงาม[7] ได้ทำการเปรียบเทียบประสิทธิภาพการทดสอบความเท่ากันของค่ากลางของประชากรมากกว่า 2 กลุ่ม เมื่อข้อมูลอยู่ในมาตราอันดับและมีความสัมพันธ์กัน พบว่าการทดสอบเอฟเป็นวิธีทดสอบที่มีความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ได้ดีกว่าวิธีการทดสอบฟรายด์แมนเมื่อกลุ่มตัวอย่างมีขนาดเล็ก แต่เมื่อตัวอย่างมีขนาดใหญ่ขึ้นพบว่าการทดสอบเอฟและการทดสอบฟรายด์แมนมีความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ใกล้เคียงกันในการแจกแจงที่ต่างกัน และงานวิจัยของ จิรภา โคมเดือน และอัชฌา อระวีพร[8] พบว่าวิธีบูตสแทร็ปเป็นวิธีที่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีเมื่อขนาดตัวอย่างใกล้เคียง 30

ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาการเปรียบเทียบความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบของการทดสอบเอฟ การทดสอบ

ฟรายด์แมน การทดสอบเอฟแบบบูตสแทร็ป และการทดสอบเอฟแบบเรียงสับเปลี่ยนภายใต้แผนแบบบล็อกสมบูรณ์เชิงสุ่ม เมื่อข้อมูลมีการแจกแจงล็อกนอร์มัลที่มีขนาดตัวอย่างต่าง ๆ ซึ่งงานวิจัยในครั้งนี้ได้ศึกษาโดยใช้เทคนิคมอนติคาร์โลในการจำลองข้อมูล

เอกสารที่เกี่ยวข้อง

1. **ตัวแบบเชิงสถิติ**

ตัวแบบเชิงสถิติสำหรับค่าสังเกตที่ได้จากการทดลองที่ใช้แผนแบบบล็อกสมบูรณ์เชิงสุ่ม คือ

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij};$$
$$i = 1,2,...,t;\ j = 1,2,...b$$

โดยที่

y_{ij} คือค่าสังเกตที่ได้จากหน่วยทดลองในบล็อกที่ j ที่ได้รับทรีตเมนต์ที่ i

μ คือค่าเฉลี่ยรวมของประชากร

τ_i คือ อิทธิพลของทรีตเมนต์ที่ i

β_j คือ อิทธิพลบล็อกที่ j

ε_{ij} คือ ความคลาดเคลื่อนจากการทดลอง โดยมีค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และ 3

t คือ จำนวนทรีตเมนต์

b คือ จำนวนบล็อก

สำหรับงานวิจัยนี้จะกำหนดให้ปัจจัยที่สนใจศึกษาและปัจจัยรบกวนเป็นปัจจัยที่มีอิทธิพลตรึง (Fixed effect)

โดยการพิจารณา

$\sum_{i=1}^{t} \tau_i = 0$ และ $\sum_{j=1}^{b} \beta_j = 0$ ซึ่งกำหนดระดับความแตกต่างระหว่างอิทธิพลของทรีตเมนต์และอิทธิพลของบล็อก

โดยกำหนดให้

$$\Phi_\tau = \frac{\sqrt{\sum_{i=1}^{t} \tau_i^2}}{\sigma} \text{ และ } \Phi_\beta = \frac{\sqrt{\sum_{j=1}^{b} \beta_j^2}}{\sigma}$$

เมื่อ Φ_τ, Φ_β คือ ค่าสัมประสิทธิ์ความเบี่ยงเบนของอิทธิพลทรีตเมนต์และอิทธิพลบล็อก ตามลำดับ และ σ คือค่าเบี่ยงเบนมาตรฐานของข้อมูล โดยกำหนดค่าสัมประสิทธิ์ความเบี่ยงเบนระหว่างอิทธิพลบล็อกเท่ากับ 7 เพื่อให้แต่ละบล็อกแตกต่างกันอย่างชัดเจน และสามารถกำหนดค่าอิทธิพลบล็อกได้ดัง Table 1

สำหรับการกำหนดค่าสัมประสิทธิ์ความเบี่ยงเบนของทรีตเมนต์จะแบ่งออกเป็น 3 ระดับ คือ 1, 2 และ 3 แทนระดับความแตกต่างน้อย ปานกลาง และมาก ตามลำดับ และสามารถกำหนดค่าอิทธิพลของทรีตเมนต์ได้ดัง Table 2-3 โดยแบ่งเป็น 2 กรณีคือ ทรีตเมนต์แตกต่างกันหนึ่งกลุ่มและทรีตเมนต์แตกต่างกันทุกกลุ่ม โดยทั้งสองกรณีจะใช้จำลองข้อมูลเพื่อศึกษากำลังการทดสอบของสถิติทดสอบ

Table 1. Block effect

β_j	Number of blocks			
	3	5	7	10
β_1	-9.8995	-8.8544	-7.9373	-6.9361
β_2	0	4.4272	-5.2915	-5.3947
β_3	9.8995	0	-2.6458	-3.8534
β_4	-	4.4272	0	-2.3120
β_5	-	8.8544	2.6458	-0.7707
β_6	-	-	5.2915	0.7707
β_7	-	-	7.9373	2.3120
β_8	-	-	-	3.8534
β_9	-	-	-	5.3947
β_{10}	-	-	-	6.9361

Table 2. Treatment effect (one-group difference)

σ	Φ_τ	τ_1	τ_2	τ_3
1	1	-0.8165	0.4082	0.4082
	2	-1.6330	0.8165	0.8165
	3	-2.4495	1.2247	1.2247
3	1	-2.4495	1.2247	1.2247
	2	-4.8990	2.4495	2.4495
	3	-7.3485	3.6742	3.6742

Table 3. Treatment effect (across all groups)

σ	Φ_τ	τ_1	τ_2	τ_3
1	1	-0.7071	0	0.7071
	2	-1.4142	0	1.4142
	3	-2.1213	0	2.1213
3	1	-2.1213	0	2.1213
	2	-4.2426	0	4.2426
	3	-6.3640	0	6.3640

2.สถิติทดสอบที่ศึกษามีรายละเอียดดังนี้

2.1 สถิติทดสอบเอฟ

$$F = \frac{\left[\frac{1}{b}\sum_{i=1}^{t} y_{i.}^2 - \frac{y_{..}^2}{tb}\right] \Big/ (t-1)}{\left[\sum_{i=1}^{t}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{b}\sum_{i=1}^{t} y_{i.}^2 - \frac{1}{t}\sum_{j=1}^{b} y_{.j}^2 + \frac{y_{..}^2}{tb}\right] \Big/ (t-1)(b-1)} \qquad (1)$$

โดยที่ t คือ จำนวนทรีตเมนต์,
b คือ จำนวนบล็อก,
y_{ij} คือ ค่าสังเกตที่ได้จากหน่วยทดลองในบล็อกที่ j ที่ได้รับทรีตเมนต์ที่ i,
$y_{i.}$ คือ ผลรวมของค่าสังเกตในทรีตเมนต์ที่ i,
$y_{.j}$ คือ ผลรวมของค่าสังเกตในบล็อกที่ j และ
$y_{..}$ คือผลรวมของค่าสังเกตทั้งหมด

2.2 สถิติทดสอบฟรายด์แมน

$$Q=\left[\frac{12}{bt(t+1)}\sum_{i=1}^{t}R_i^2\right]-3b(t+1) \qquad (2)$$

โดยที่ t คือ จำนวนทรีตเมนต์,
b คือ จำนวนบล็อก,
R_i คือ ผลรวมของลำดับที่ของทรีตเมนต์ที่ i

2.3 สถิติทดสอบเอฟแบบการเรียงสับเปลี่ยน

วิธีการเรียงสับเปลี่ยนถูกนำเสนอโดย Fisher[9] มีหลักการคือ สมมุติตัวอย่างประกอบด้วย b บล็อก และ t ทรีตเมนต์ ดังนั้นข้อมูลที่ได้จะแทนด้วย y_{ij} ซึ่งเป็นค่าตอบสนองของทรีตเมนต์ที่ i ในบล็อกที่ j โดยที่ในแต่ละบล็อกจะทำการสุ่มตัวอย่างข้อมูลจากแต่ละทรีตเมนต์ภายในบล็อกนั้น ๆ แบบไม่ใส่คืน ทำซ้ำจนครบทุกบล็อก เพื่อสร้างชุดข้อมูลจำลองที่มีขนาดเท่ากับชุดข้อมูลเดิม หลังจากได้ข้อมูลตัวอย่างชุดใหม่แล้ว จะคำนวณค่าสถิติทดสอบ F_P ตามสมการที่ (1) และคำนวณค่าพี (p-value) โดยใช้สัดส่วนของค่าสถิติทดสอบในแต่ละครั้งที่มากกว่าค่าสถิติทดสอบของข้อมูลตั้งต้น สรุปเป็นขั้นตอนได้ดังนี้

2.3.1 จำลองข้อมูลตามสถานการณ์ที่กำหนดและคำนวณสถิติทดสอบตั้งต้น F

2.3.2 เริ่มจากบล็อกที่ 1 สุ่มตัวอย่างซ้ำแบบไม่ใส่คืน ภายในบล็อกจนครบทุกทรีตเมนต์

2.3.3 ในบล็อกถัดมาสุ่มตัวอย่างซ้ำแบบไม่ใส่คืน ภายในบล็อกจนครบทุกทรีตเมนต์เช่นเดียวกัน ทำไปเรื่อย ๆ จนครบทุกบล็อกตามสถานการณ์ที่ศึกษา จะได้ข้อมูลการเรียงสับเปลี่ยนชุดที่ 1

2.3.4 คำนวณสถิติทดสอบจากข้อมูลชุดใหม่ได้เป็น F_P^1

2.3.5 ทำซ้ำในขั้นตอนที่ 2.3.2 – 2.3.4 จำนวน 500 ครั้ง จะได้ค่าสถิติทดสอบ $F_P^1, F_P^2, ..., F_P^{500}$

2.3.6 คำนวณค่าพีโดยใช้สัดส่วนของ $F_P > F$

2.4 สถิติทดสอบเอฟแบบบูตสแทร็ป

วิธีบูตสแทร็ปถูกพัฒนาขึ้นโดย Efron[10] ซึ่งเป็นวิธีการสุ่มตัวอย่างแบบใส่คืน (Resampling with replacement) การประยุกต์ใช้สำหรับแผนแบบบล็อกสมบูรณ์เชิงสุ่มนั้น สมมุติตัวอย่างที่ประกอบด้วย b บล็อก และ t ทรีตเมนต์ ดังนั้นข้อมูลที่ได้จะแทนด้วย y_{ij} ซึ่งเป็นค่าตอบสนองของทรีตเมนต์ที่ i ในบล็อกที่ j โดยที่ในแต่ละบล็อกจะทำการสุ่มตัวอย่างข้อมูลจากแต่ละทรีตเมนต์ภายในบล็อกนั้น ๆ แบบใส่คืน ทำซ้ำจนครบทุกบล็อกเพื่อสร้างชุดข้อมูลจำลองที่มีขนาดเท่ากับชุดข้อมูลเดิม หลังจากได้ข้อมูลตัวอย่างชุดใหม่แล้ว จะคำนวณค่าสถิติทดสอบ F_B^1 ตามสมการที่ (1) และคำนวณค่าพีโดยใช้สัดส่วนของค่าสถิติทดสอบในแต่ละครั้งที่มากกว่าค่าสถิติทดสอบของข้อมูลตั้งต้น สรุปเป็นขั้นตอนได้ดังนี้

2.4.1 จำลองข้อมูลตามสถานการณ์ที่กำหนดและคำนวณสถิติทดสอบตั้งต้น F

2.4.2 เริ่มจากบล็อกที่ 1 สุ่มตัวอย่างซ้ำแบบใส่คืน ภายในบล็อกจนครบทุกทรีตเมนต์

2.4.3 ในบล็อกถัดมาสุ่มตัวอย่างซ้ำแบบใส่คืนภายในบล็อกจนครบทุกทรีตเมนต์ เช่นเดียวกันทำไปเรื่อย ๆ จนครบทุกบล็อกตามสถานการณ์ที่ศึกษา จะได้ข้อมูลบูตสแทร็ปชุดที่ 1

2.4.4 คำนวณสถิติทดสอบจากข้อมูลชุดใหม่ได้เป็น F_B^1

2.4.5 ทำซ้ำในขั้นตอนที่ 2.4.2 - 2.4.4 จำนวน 500 ครั้ง จะได้ค่าสถิติทดสอบ $F_B^1, F_B^2, ..., F_B^{500}$

2.4.6 คำนวณค่าพีโดยใช้สัดส่วนของ $F_B > F$

2.5 การแจกแจงของค่าคลาดเคลื่อนที่ศึกษา

การแจกแจงล็อกนอร์มัล (Lognormal distribution) มีฟังก์ชันความน่าจะเป็นของการแจกแจงล็อกนอร์มัล

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right);x\geq 0,-\infty<\mu<\infty,\sigma>0 \quad (3)$$

โดยที่ μ คือ พารามิเตอร์บ่งตำแหน่ง และ σ^2 คือ พารามิเตอร์บ่งขนาด สำหรับในงานวิจัยนี้ กำหนดให้พารามิเตอร์ (μ,σ^2) มีค่าเท่ากับ (1 , 0.1139) และ (1, 0.5374) และปรับข้อมูลที่ได้ให้มีค่าเฉลี่ยเท่ากับ 0 ตามคุณสมบัติของค่าคลาดเคลื่อน อย่างไรก็ตามเมื่อใช้วิธีนี้ในการจำลอง ค่าคลาดเคลื่อนที่ได้ยังคงมีการแจกแจงล็อกนอร์มัล เพียงแต่เป็นการเปลี่ยนตำแหน่งพารามิเตอร์ของการแจกแจง (Shifted parameter) เพื่อให้เป็นไปตามคุณสมบัติของค่าคลาดเคลื่อนเท่านั้น ซึ่งการจำลองข้อมูลลักษณะนี้เพื่อต้องการให้ข้อมูลที่ได้สอดคล้องกับข้อมูลจริงที่เกิดขึ้นในทางปฏิบัติมากที่สุด

2.6 เกณฑ์ในการพิจารณา

2.6.1 ความสามารถในการควบคุมความผิดพลาดแบบที่ 1 พิจารณาจากค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ซึ่งคำนวณจาก สัดส่วนของการปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นจริง $(\hat{\alpha})$ และนำมาเปรียบเทียบกับเกณฑ์ของ Bradley[11] หาก $\hat{\alpha}\in$ [0.025, 0.075] แสดงว่าสถิติทดสอบดังกล่าวสามารถควบคุมความผิดพลาดแบบที่ 1 ได้

2.6.2 กำลังการทดสอบ คำนวณจาก สัดส่วนของการปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นเท็จ ซึ่งจะพิจารณาเฉพาะสถิติทดสอบที่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้เท่านั้น โดยที่สถิติทดสอบที่มีค่ากำลังการทดสอบสูงที่สุดจะถือว่ามีประสิทธิภาพมากที่สุด

วิธีดำเนินการวิจัย

วัตถุประสงค์ของงานวิจัยนี้เพื่อเปรียบเทียบประสิทธิภาพของสถิติทดสอบที่อิงวิธีการสุ่มตัวอย่างซ้ำสองวิธี ได้แก่ สถิติทดสอบเอฟแบบการเรียงสับเปลี่ยนและสถิติทดสอบเอฟแบบบูตสแทร็ป โดยเปรียบเทียบกับวิธีพื้นฐานอีกสองวิธี ได้แก่ สถิติทดสอบเอฟ สถิติทดสอบฟรายด์แมน ซึ่งใช้ทดสอบค่ากลางสำหรับแผนแบบบล็อกสมบูรณ์เชิงสุ่มภายใต้ข้อมูลที่มีการแจกแจงล็อกนอร์มัล กำหนดค่าเฉลี่ยรวมของประชากร (μ) เท่ากับ 10 และค่าส่วนเบี่ยงเบนมาตรฐานของประชากร เท่ากับ 1 และ 3 ตามลำดับ โดยใช้ระดับนัยสำคัญเท่ากับ 0.05 ทำซ้ำด้วยเทคนิคมอนติคาร์โล 5,000 รอบ โดยกำหนดจำนวนรอบบูตสแทร็ปและรอบการเรียงสับเปลี่ยนอย่างละ 500 รอบ ในการศึกษานี้กำหนดจำนวนทรีตเมนต์เท่ากับ 3 ทรีตเมนต์ จำนวนบล็อกเท่ากับ 3, 5, 7 และ 10 จำลองข้อมูลโดยให้ไม่มีความแตกต่างกันของข้อมูลระหว่างทรีตเมนต์ แล้วคำนวณค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สำหรับการคำนวณค่าประมาณกำลังการทดสอบจะต้องจำลองข้อมูลให้มีความแตกต่างกันระหว่างทรีตเมนต์ ตาม Table 2 - 3 เพื่อให้ครอบคลุมสถานการณ์ที่เป็นไปได้ทุกรูปแบบ จากนั้นพิจารณาประสิทธิภาพของสถิติทดสอบจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบ สำหรับขั้นตอนการจำลองข้อมูลมีดังนี้

1. สร้างความคลาดเคลื่อน (ε_{ij}) ให้มีการแจกแจงล็อกนอร์มัลตามที่กำหนด

2. สร้างอิทธิพลของทรีตเมนต์ (τ_i) ตามสมมุติฐานหลักซึ่งแบ่งออกเป็น 2 กรณีย่อย ได้แก่

2.1 การประมาณค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 จะสร้างอิทธิพลของทรีตเมนต์ตามสมมุติฐานหลัก นั่นคือ ค่าเฉลี่ยของทรีตเมนต์ไม่มีความแตกต่างกัน ซึ่งในงานวิจัยนี้กำหนดให้ τ_1, τ_2, τ_3 มีค่าเท่ากับ 0

2.2 การประมาณค่ากำลังการทดสอบ จะสร้างอิทธิพลของทรีตเมนต์ตามสมมุติฐานทางเลือก กล่าวคือ ค่าเฉลี่ยของทรีตเมนต์แตกต่างกัน โดยกำหนดค่า τ_1, τ_2, τ_3 ตาม Table 2 - 3

3. สร้างอิทธิพลของบล็อก (β_j) ให้แตกต่างกันโดยกำหนดค่า β_j ตาม Table 1

4. สร้างข้อมูลตามแผนการทดลองบล็อกสมบูรณ์เชิงสุ่ม

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij};$$
$$i = 1, 2, ..., t;\ j = 1, 2, ...b$$

5. คำนวณค่าพีจากสถิติทดสอบที่ศึกษาทุกวิธี

6. เปรียบเทียบค่าพีที่ได้กับระดับนัยสำคัญที่กำหนด หากค่าพีน้อยกว่า 0.05 จะปฏิเสธสมมุติฐานหลัก

7. ทำซ้ำขั้นตอนที่ 1 – 3 จำนวน 5,000 ครั้ง

8. คำนวณค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 หรือค่าประมาณกำลังการทดสอบแล้วแต่กรณีที่จำลองข้อมูลในข้อ 2.1 - 2.2

ผลการวิจัย

ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 และ ค่าประมาณกำลังการทดสอบ แสดงใน Table 4-6 โดยกำหนดสัญลักษณ์ของสถิติทดสอบแต่ละวิธีคือ สถิติทดสอบเอฟ (F-test: F) สถิติทดสอบฟรายด์แมน (Friedman test: FM) สถิติทดสอบเอฟแบบบูตสแทร็ป (Bootstrap F-test: FB) และ สถิติทดสอบเอฟแบบการเรียงสับเปลี่ยน (Permutation F-test: FP) ตามลำดับ

จากการศึกษาความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ของสถิติทดสอบ แสดงใน Table 4 พบว่า สถิติทดสอบเอฟ สถิติทดสอบฟรายด์แมน และสถิติทดสอบเอฟแบบการเรียงสับเปลี่ยน สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์ของ Bradley ทุกกรณี ในขณะที่สถิติทดสอบเอฟแบบบูตสแทร็ปสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เฉพาะกรณีที่จำนวนบล็อกไม่น้อยกว่า 5 เท่านั้น

Table 4. Type I error rates

σ	Number of blocks	Test statistics			
		F	FM	FB	FP
1	3	0.0478	0.0256	0.0992**	0.0478
	5	0.0462	0.0390	0.0740	0.0510
	7	0.0522	0.0556	0.0624	0.0516
	10	0.0498	0.0484	0.0598	0.0524
3	3	0.0376	0.0282	0.0906**	0.0308
	5	0.0332	0.0374	0.0720	0.0496
	7	0.0368	0.0462	0.0640	0.0506
	10	0.0372	0.0448	0.0562	0.0480

** indicates failure to control the type I error rate under Bradley's criterion

สำหรับการศึกษาประสิทธิภาพของการทดสอบด้วยกำลังการทดสอบ กรณีข้อมูลแตกต่างกันหนึ่งกลุ่ม แสดงใน Table 5 และ Figure 1 – 2 สถิติทดสอบเอฟจะมีกำลังการทดสอบสูงที่สุดเมื่อจำนวนบล็อกมีขนาดเล็ก $(b=3)$ แต่เมื่อจำนวนบล็อกเพิ่มมากขึ้น $(b\geq 5)$ สถานการณ์ส่วนใหญ่สถิติทดสอบเอฟแบบบูตสแทร็ปมีกำลังการทดสอบสูงที่สุด อย่างไรก็ตามเมื่ออิทธิพลของบล็อกมีขนาดเพิ่มขึ้นสถิติทดสอบเอฟทั้งสามวิธีคือ สถิติทดสอบเอฟ สถิติทดสอบเอฟแบบเรียงสับเปลี่ยน และสถิติทดสอบเอฟแบบบูตสแทร็ป มีกำลังการทดสอบสูงใกล้เคียงกัน

Table 5. Estimated power of the test for one-group difference

σ	Φ_τ	Number of blocks	Test statistics			
			F	FM	FB	FP
1	1	3	0.1918*	0.1102	-	0.1676
		5	0.3864	0.3030	0.4718*	0.3880
		7	0.5632	0.4692	0.6112*	0.5692
		10	0.7452	0.6308	0.7674*	0.7480
	2	3	0.5724*	0.2136	-	0.3334
		5	0.9030	0.7740	0.9330*	0.8974
		7	0.9830	0.9010	0.9860*	0.9824
		10	0.9972	0.9784	0.9978*	0.9974
	3	3	0.8578*	0.2456	-	0.4068
		5	0.9936	0.9522	0.9962*	0.9920
		7	0.9990*	0.9838	0.9988	0.9990*
		10	1.0000*	0.9994	0.9998	1.0000*
3	1	3	0.2862*	0.1518	-	0.2268
		5	0.5078	0.4652	0.5952*	0.5410
		7	0.6466	0.6056	0.6988*	0.6676
		10	0.7830	0.7840	0.8098*	0.7996
	2	3	0.8028*	0.6776	-	0.3614
		5	0.8952	0.8054	0.9200*	0.9004
		7	0.9564	0.8924	0.9626*	0.9586
		10	0.9882	0.9794	0.9886	0.9888*
	3	3	0.9390*	0.8762	-	0.4056
		5	0.9770	0.9266	0.9830*	0.9780
		7	0.9904	0.9624	0.9910*	0.9898
		10	0.9982	0.9976	0.9986*	0.9986*

* indicates the highest estimated power of the test

- indicates failure to control the type I error

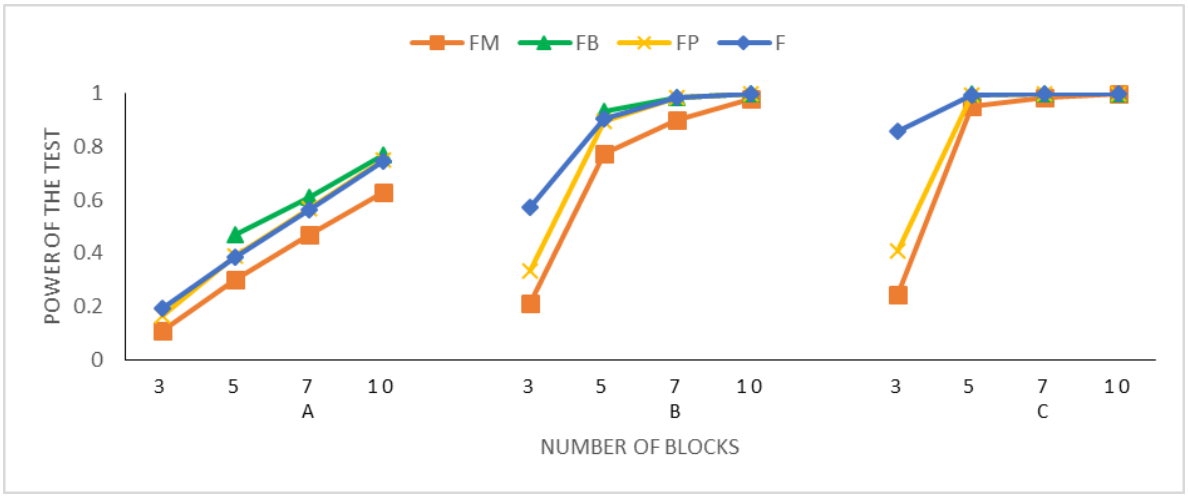

Figure 1. Estimated power of the test for single-group difference under σ = 1 for the treatment effect Coefficient: A : Φ_τ =1, B : Φ_τ = 2, C : Φ_τ = 3

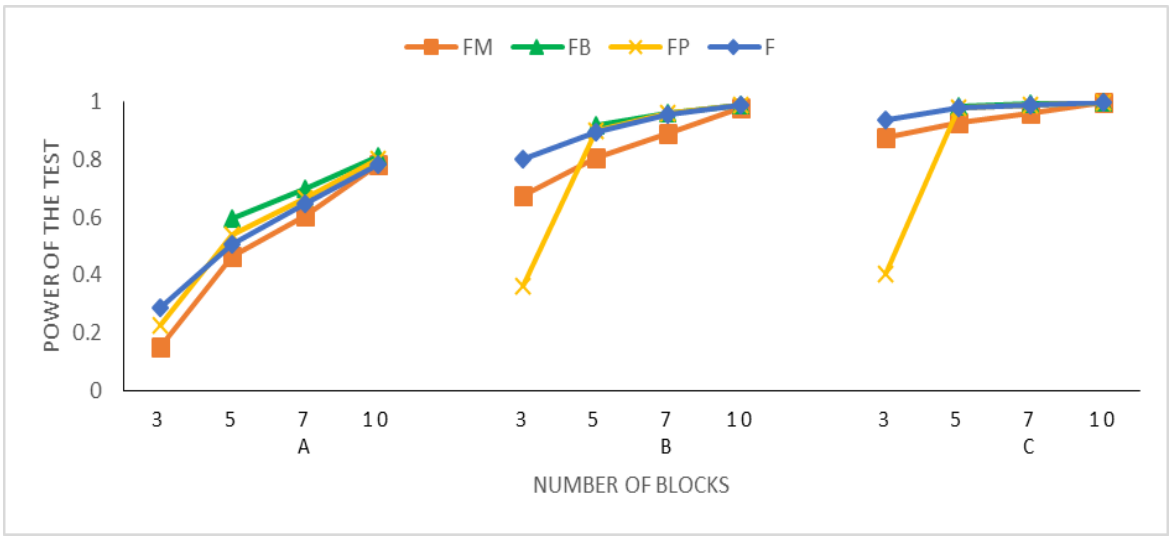

Figure 2. Estimated power of the test for single-group difference under σ = 3 for the treatment effect Coefficient: A : Φ_τ =1, B : Φ_τ = 2, C : Φ_τ = 3

Table 6. Estimated power of the test for across all groups

σ	Φ_τ	Number of blocks	Test statistics			
			F	FM	FB	FP
1	1	3	0.1972*	0.1116	-	0.1666
		5	0.3924	0.3018	0.4716*	0.3958
		7	0.5644	0.4918	0.6084*	0.5636
		10	0.7470	0.6554	0.7644*	0.7492
	2	3	0.5832*	0.3710	-	0.4944
		5	0.9002	0.7936	0.9260*	0.8910
		7	0.9804	0.9532	0.9842*	0.9800
		10	0.9984*	0.9942	0.9982	0.9984*
	3	3	0.8708*	0.6764	-	0.7914
		5	0.9958	0.9664	0.9966*	0.9932
		7	0.9994*	0.9992	0.9994*	0.9984
		10	1.0000*	1.0000*	1.0000*	1.0000*
3	1	3	0.2838*	0.1940	-	0.2694
		5	0.4988	0.4700	0.5820*	0.5204
		7	0.5078	0.4750	0.5898*	0.5276
		10	0.7888	0.8502*	0.8130	0.7970
	2	3	0.6764*	0.5088	-	0.6266
		5	0.9024	0.8446	0.9224*	0.9002
		7	0.9002	0.8360	0.9138*	0.9100
		10	0.9882	0.9968*	0.9898	0.9888
	3	3	0.8832*	0.7304	-	0.8338
		5	0.9774	0.9530	0.9800*	0.9754
		7	0.9800	0.9510	0.9824*	0.9764
		10	0.9996	1.0000*	0.9998	0.9994

* indicates the highest estimated power of the test

- indicates failure to control the type I error

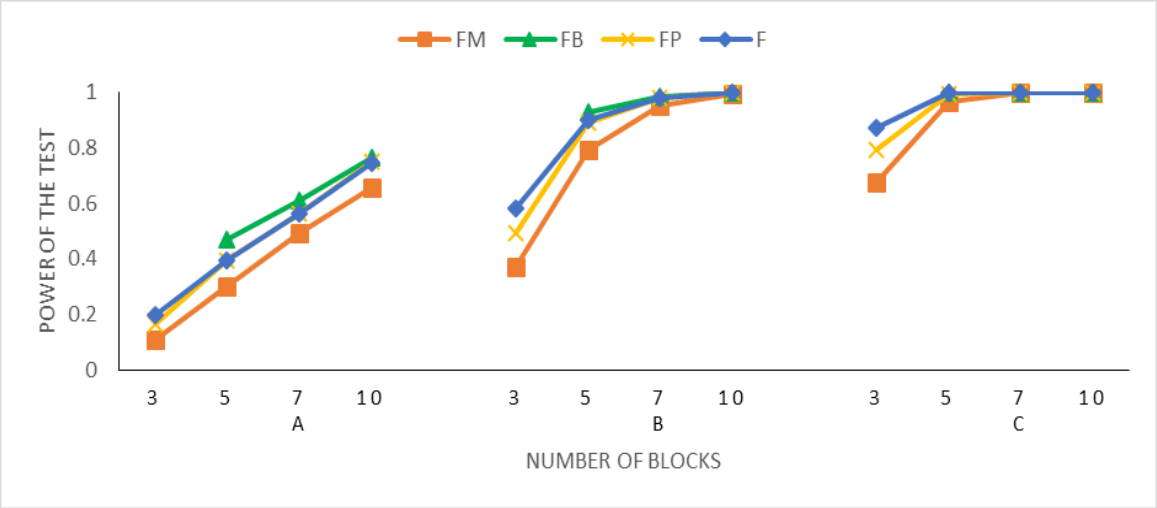

Figure 3. Estimated power of the test for across all groups under σ = 1 for the treatment effect Coefficient: A : Φ_τ =1, B : Φ_τ = 2, C : Φ_τ = 3

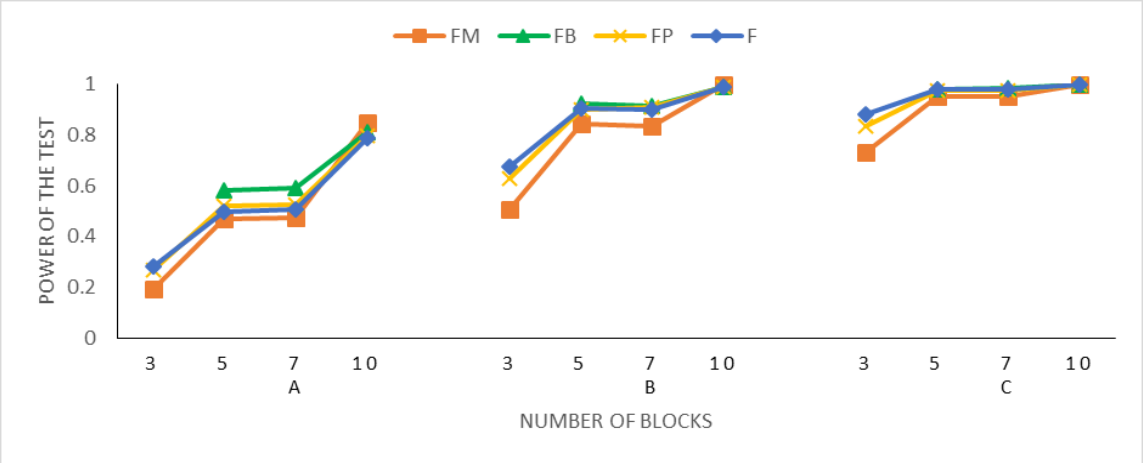

Figure 4. Estimated power of the test for across all groups under σ = 3 for the treatment effect Coefficient: A : Φ_τ =1, B : Φ_τ = 2, C : Φ_τ = 3

ในส่วนของกรณีข้อมูลแตกต่างกันทุกกลุ่มแสดงใน Table 6 และ Figure 3 - 4 แบ่งการสรุปผลเป็น 2 ส่วน คือ ในส่วนแรกข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 พบว่าสถิติทดสอบเอฟจะมีกำลังการทดสอบสูงที่สุดเมื่อจำนวนบล็อกมีขนาดเล็ก $(b=3)$ แต่เมื่อจำนวนบล็อกเพิ่มมากขึ้น $(b \geq 5)$ สถานการณ์ส่วนใหญ่สถิติทดสอบเอฟแบบบูตสแทร็ปมีกำลังการทดสอบสูงที่สุด อย่างไรก็ตามเมื่ออิทธิพลของบล็อกมีขนาดเพิ่มขึ้น สถิติทดสอบเอฟทั้งสามวิธีคือ สถิติทดสอบเอฟ สถิติทดสอบเอฟแบบเรียงสับเปลี่ยน และสถิติทดสอบเอฟแบบบูตสแทร็ป มีกำลังการทดสอบสูงใกล้เคียงกัน

สำหรับส่วนที่สองข้อมูลค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 พบว่าสถิติทดสอบเอฟจะมีกำลังการทดสอบสูงที่สุดเมื่อจำนวนบล็อกมีขนาดเล็ก $(b=3)$ ในขณะที่สถิติทดสอบเอฟแบบบูตสแทร็ปมีกำลังการทดสอบสูงที่สุด เมื่อจำนวนบล็อกเท่ากับ 5 และ 7 นอกจากนี้ยังพบว่าเมื่อจำนวนบล็อกเท่ากับ 10 สถิติทดสอบสถิติทดสอบฟรายด์แมนจะมีกำลังการทดสอบสูงที่สุด อย่างไรก็ตามเมื่ออิทธิพลของบล็อกมีขนาดเพิ่มขึ้น สถิติทดสอบทุกวิธีมีกำลังการทดสอบใกล้เคียงกัน

สรุปผลการวิจัยและอภิปรายผลการวิจัย

จากผลการศึกษาพบว่าสถิติทดสอบเอฟ สถิติทดสอบฟรายด์แมน และสถิติทดสอบเอฟแบบการเรียงสับเปลี่ยน สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ตามเกณฑ์ของ Bradley ทุกกรณี ในขณะที่สถิติทดสอบเอฟแบบบูตสแทร็ป สามารถควบคุมความผิดพลาดแบบที่ 1 ได้เฉพาะกรณีที่จำนวนบล็อกไม่น้อยกว่า 5 เนื่องจากวิธีดังกล่าวเป็นการสุ่มตัวอย่างซ้ำแบบใส่คืน ซึ่งหากข้อมูลมีขนาดเล็กเกินไป (จำนวนบล็อกน้อย) จะทำให้ได้ตัวอย่างซ้ำแบบเดิม ๆ ไม่หลากหลาย จึงทำให้ไม่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ โดยผลจากการทดลองที่ได้แสดงถึงการปฏิเสธสมมติฐานหลัก (ตัดสินใจผิดพลาด) มากเกินกว่าที่กำหนดไว้ในงานวิจัย ส่งผลให้เกิดข้อสรุปที่คลาดเคลื่อนและความน่าเชื่อถือลดลง ดังนั้น สถิติทดสอบเอฟแบบบูตสแทร็ปจึงไม่มีประสิทธิภาพในกรณีที่กล่าวมา

ผลการพิจารณากำลังการทดสอบ พบว่า เมื่อจำนวนบล็อกมีขนาดเล็ก สถิติทดสอบเอฟมีกำลังการทดสอบสูงที่สุดทุกกรณี และเมื่อจำนวนบล็อกขนาดใหญ่ขึ้น สถานการณ์ส่วนใหญ่สถิติทดสอบเอฟแบบบูตสแทร็ปจะมีกำลังการทดสอบสูงที่สุด อย่างไรก็ตามเมื่ออิทธิพลของบล็อกมีขนาดเพิ่มขึ้นสถิติทดสอบเอฟทั้งสามวิธีคือ สถิติทดสอบเอฟ สถิติทดสอบเอฟแบบเรียงสับเปลี่ยน และสถิติทดสอบเอฟแบบบูตสแทร็ป มีกำลังการทดสอบสูงใกล้เคียงกัน สำหรับสถิติทดสอบฟรายด์แมนจะมีกำลังการทดสอบสูงที่สุดเมื่อข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐานมากและจำนวนบล็อกมีขนาดใหญ่เท่านั้น

นอกจากนี้ยังพบว่าค่ากำลังการทดสอบแปรผันตามจำนวนบล็อก อิทธิพลของบล็อก และค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูล กล่าวคือ เมื่อจำนวนบล็อก อิทธิพลของบล็อก และค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูล มีค่ามากขึ้น กำลังการทดสอบของทุกวิธีการจะเพิ่มขึ้นด้วย

สำหรับการนำไปใช้ในทางปฏิบัติ หากการทดลองมีจำนวนบล็อกน้อย สามารถใช้สถิติทดสอบเอฟได้ถึงแม้ว่าข้อมูลจะไม่ได้มีการแจกแจงปรกติซึ่งในงานวิจัยนี้คือข้อมูลมีการแจกแจงล็อกนอร์มัล เนื่องจากผลการทดสอบข้างต้นเห็นได้ว่าสถิติทดสอบเอฟมีความแกร่งและมีกำลังการทดสอบสูงที่สุด แต่ในกรณีที่จำนวนบล็อกขนาดปานกลาง สามารถใช้สถิติทดสอบเอฟแบบบูตสแทร็ป ได้ดี และหากจำนวนบล็อกมีขนาดใหญ่ สามารถเลือกใช้สถิติทดสอบได้ทุกวิธีเนื่องจากมีประสิทธิภาพใกล้เคียงกัน

ผลการทดลองที่ได้สามารถใช้เป็นแนวทางในการศึกษาเพิ่มเติมได้ กล่าวคือ อาจศึกษาข้อมูลที่มีการแจกแจงอื่น ๆ เพิ่มเติม เช่น การแจกแจงแกมมา การแจกแจงไวบูล เป็นต้น อย่างไรก็ตามงานวิจัยนี้ศึกษาเฉพาะกรณีทรีตเมนต์เท่ากับ 3 กลุ่ม ดังนั้น อาจศึกษาเพิ่มเติมในกรณีที่มีจำนวนทรีตเมนต์มากกว่า 3 กลุ่ม เพื่อให้ผลการทดลองที่ได้สามารถนำไปประยุกต์ใช้กับข้อมูลจริงอย่างครอบคลุมมากยิ่งขึ้น

เอกสารอ้างอิง

1. Nedelkov K, Harper MT, Melgar A, Chen X, Räisänen S, Martin CMMR, et al. Acceptance of flavored concentrate premixes by young ruminants following a short-term exposure. J Dairy Sci 2019;102:388-94.
2. Tharu NK, Baidhya S. Modeling PM2.5 and TSP concentrations: a comparison of Weibull, lognormal, and gamma distributions. Patan Prospect J 2024;4:69-78.
3. Hamid HA, Yahaya AS, Ramli NA, Ul-Saufie AZ. Finding the best statistical distribution model in PM_{10} concentration modeling by using lognormal distribution. J Appl Sci 2013;13:294–300.
4. Pongsakchat V, Kidpholjaroen P. The statistical distributions of $PM_{2.5}$ in Rayong and Chonburi provinces, Thailand. Asian J Appl Sci 2020;8:172-7.
5. O'Shaughnessy PY, Hui FKC, Müller S, Welsh AH. Bootstrapping F test for testing random effects in linear mixed models. [Internet]. 2018 [cited 2024 Oct 5]. Available from:

https://www.researchgate.net/publication/329568773.

6. บำรุงศักดิ์ เผื่อนอารีย์, โชคศรัญญ์ สิริลีพันธุ์. การเปรียบเทียบประสิทธิภาพของสถิติทดสอบสำหรับค่าเฉลี่ยประชากรมากกว่า 2 กลุ่ม ภายใต้การแจกแจงแบบสมมาตร. วารสารวิทยาศาสตร์บูรพา 2561;23:510-9.
7. วนิดา พงษ์ศักดิ์ชาติ, นพรัตน์ แป้นงาม. การเปรียบเทียบประสิทธิภาพการทดสอบความเท่ากันของค่ากลางเมื่อข้อมูลอยู่ในมาตราอันดับและมีความสัมพันธ์กัน. วารสารวิทยาศาสตร์บูรพา 2561;23:1070-83.
8. จิรภา โคมเดือน, อัชฌา อระวีพร. การเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้ทดสอบค่าเฉลี่ยของ 3 ประชากรในกรณีที่ความแปรปรวนไม่เท่ากัน. วารสารวิทยาศาสตร์และเทคโนโลยี 2560;25:918-28.
9. Fisher RA. The design of experiments. New York: Hafner Publishing Company; 1971.
10. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
11. Bradley JV. Robustness? Br J Math Stat Psychol 1978;31:144–52.