Comparisons of Type I Error Rates and Power of the Test in Post Hoc Comparison Procedures for Completely Randomized Design

Puanyanut Pinchoo [1]
Suchada Bowarnkitiwong [2]

ABSTRACT

The purposes of this research were to compare type I error rates and power of the test in completely randomized design for 14 statistics in Post Hoc comparison procedures, namely, LSD, Tukey's HSD, Bonferroni, Tukey's b, Sidak, Duncan, Scheffe's, Hochberg's GT2, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K, and Dunnett when each population had equal variances and normally distributed using $\alpha = 0.05$. *Sample sizes of concern were to be equal and unequal. Each category, data were simulated into 3 to 8 groups with small, medium, and large sample sizes. For each statistic under experimental situation, the Monte Carlo experiment was repeated 10,000 times.*

The findings were summarized as follows:

1. When the sample sizes were equal, LSD and Duncan can control the Type I error rates in every cases. There were seven methods, Sidak, Dunnett, Tukey's b, Waller-Duncan, S-N-K, Gabriel and R-E-G-WF, can control the Type I error rates in some cases. Furthermore, there were five methods, Bonferroni, Tukey's HSD, Hochberg's GT2, Scheffe's and R-E-G-WQ, cannot control Type I error rates. Moreover, when the sample sizes were unequal, there were three methods, LSD, Waller-Duncan, and Duncan, can control the Type I error rates in every cases. There were five methods, Dunnett, Tukey's b, S-N-K, Gabriel and R-E-G-WF, can control the Type I error rates in some cases. In addition, there were six methods, Bonferroni, Sidak, Tukey's HSD, Hochberg's GT2, Scheffe's and R-E-G-WQ, cannot control Type I error rates.

2. In every procedures that leads to find power of the test, power of the test would be increased according to the amount of groups and sizes of sample. LSD and Waller-Duncan methods gave the highest power of the test in every cases when having 3 -4 groups. Gabriel method gave the highest power of the test in every cases when having 5-8 groups.

[1]M.A. graduate, Department of Education Research and Psychology, Faculty of Education, Chulalongkorn University.

[2]Advisor

การเปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการเปรียบเทียบค่าเฉลี่ยรายคู่ สำหรับแผนการทดลองแบบสุ่มสมบูรณ์

ปุณยนุช พินชู[1]
สุชาดา บวรกิติวงศ์[2]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ 14 วิธี คือ LSD, Tukey's HSD, Bonferroni, Tukey's b, Sidak, Duncan, Scheffe's, Hochberg's GT2, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K และ Dunnett ในแผนการทดลองแบบสุ่มสมบูรณ์ที่ $\alpha = 0.05$ ภายใต้เงื่อนไขที่ว่าประชากร ทั้ง k กลุ่มมีค่าความแปรปรวนไม่แตกต่างกัน และมีการแจกแจงปกติ ซึ่งพิจารณาเปรียบเทียบทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน โดยกำหนดให้มีระดับทรีทเมนต์ (k) ตั้งแต่ 3 กลุ่มถึง 8 กลุ่ม แบ่งกลุ่มการทดลองเป็นกลุ่มขนาดเล็ก ขนาดกลาง และขนาดใหญ่ ข้อมูลในการวิจัยได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ซิมูเลชั่น กระทำซ้ำ 10,000 ครั้งในแต่ละขนาดการทดลอง

ผลการวิจัยสรุปได้ดังนี้

1. เมื่อกลุ่มตัวอย่างมีขนาดเท่ากัน วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ในทุกกรณี มี 2 วิธี คือ วิธี LSD และ Duncan วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เป็นบางกรณี มี 7 วิธี คือ วิธี Sidak, Dunnett, Tukey's b, Waller-Duncan, S-N-K, Gabriel และ R-E-G-WF และวิธีการทดสอบที่ไม่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้มี 5 วิธี คือ วิธี Bonferroni, Tukey's HSD, Hochberg's GT2, Scheffe's และ R-E-G-WQ เมื่อกลุ่มตัวอย่างมีขนาดไม่เท่ากัน วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ในทุกกรณี มี 3 วิธี คือ วิธี LSD, Waller-Duncan และ Duncan วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เป็นบางกรณี มี 5 วิธี คือ วิธี Dunnett, Tukey's b, S-N-K, Gabriel และ R-E-G-WF และวิธีการทดสอบที่ไม่สามารถควบคุมความคลาดเคลื่อน ประเภทที่ 1 ได้มี 6 วิธี คือ วิธี Bonferroni, Sidak, Tukey's HSD, Hochberg's GT2, Scheffe's และ R-E-G-WQ

2. ทุกวิธีการทดสอบที่นำมาคำนวณหาอำนาจการทดสอบจะมีอำนาจการทดสอบจะเพิ่มขึ้นตามจำนวนและขนาดกลุ่มตัวอย่าง เมื่อพิจารณาที่อำนาจการทดสอบ พบว่า วิธี LSD และ Waller-Duncan จะเป็น 2 วิธีที่มีค่าใกล้เคียงกันและมีอำนาจการทดสอบสูงในทุกกรณีเมื่อมีจำนวนกลุ่มตัวอย่าง 3 ถึง 4 กลุ่ม และวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุดในทุกกรณีเมื่อมีจำนวนกลุ่มตัวอย่าง 5 ถึง 8 กลุ่ม

[1] นิสิตระดับปริญญามหาบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
[2] อาจารย์ที่ปรึกษาวิทยานิพนธ์

ความเป็นมาของการวิจัย

ในการวิเคราะห์ทางสถิติโดยใช้การวิเคราะห์ความแปรปรวน (ANOVA) เพื่อเปรียบเทียบผลความแตกต่างของการให้ทรีทเมนต์ (Treatment) ต่างๆ ในกลุ่มตัวอย่าง และถ้าผลการวิเคราะห์พบว่าปฏิเสธสมมติฐานหลัก (Reject The Null Hypothesis) แสดงว่าผลของการให้ทรีทเมนต์มีความแตกต่างกัน ถ้าในการวิเคราะห์มีระดับทรีทเมนต์มากกว่า 2 ระดับแสดงว่าจะต้องมีระดับทรีทเมนต์อย่างน้อยหนึ่งคู่ที่ให้ผลแตกต่างกัน แต่เนื่องจากการวิเคราะห์ในขั้นตอนนี้ผู้วิจัยจะไม่ทราบว่าระดับทรีทเมนต์ (Treatment Level) ใดบ้างที่แตกต่างกัน ผู้วิจัยจำเป็นต้องหาความแตกต่างที่เกิดขึ้นโดยทำการเปรียบเทียบค่าเฉลี่ยรายคู่ (Post Hoc Comparison) ซึ่งเป็นการเปรียบเทียบค่าเฉลี่ยโดยการนำทุกระดับของทรีทเมนต์มาจับคู่กัน และจะสลับคู่กันครบทุกระดับทรีทเมนต์

ในปัจจุบันมีวิธีการเปรียบเทียบค่าเฉลี่ยรายคู่หลายวิธี แต่ในโปรแกรม SPSS for Windows ที่นักวิจัยนิยมนำมาใช้ทำการวิเคราะห์ความแปรปรวนจะมีวิธีการเปรียบเทียบค่าเฉลี่ยรายคู่ 2 แบบ ในแบบที่ 1 เป็นการเปรียบเทียบค่าเฉลี่ยรายคู่เมื่อทราบว่าประชากรทั้ง k กลุ่มมีค่าความแปรปรวนไม่แตกต่างกัน (equal variances assumed) ประกอบด้วย 14 วิธี ได้แก่ วิธี LSD, Tukey's HSD, Bonferroni, Tukey's b, Sidak's, Duncan, Scheffe's, Hochbergs GT2, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K และ Dunnett ในแบบที่ 2 เป็นการเปรียบเทียบค่าเฉลี่ยรายคู่ เมื่อทราบว่าประชากร k กลุ่มมีค่าความแปรปรวนแตกต่างกัน (equal variances not assumed) ประกอบด้วย 4 วิธี ได้แก่ วิธี Tamhane's T2, Games - Howell, Dunnett's T3 และ Dunnett's C

Bernhardson (1975) ทำการศึกษาเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการเปรียบเทียบพหุคูณ 5 วิธี คือ LSD, HSD, Scheffe's, S-N-K และ Duncan ด้วยขนาดของกลุ่มตัวอย่างหนึ่งขนาดคือ 15 โดยทำการเปรียบเทียบข้อมูลหลังจากทำการวิเคราะห์ความแปรปรวนด้วย F-test ที่กำหนด $\mu = 50$, $\sigma = 15$ ภายใต้การแจกแจงของประชากรแบบปกติโดยใช้สัปรูทีนโปรแกรม Gauss ผลการวิจัยพบว่า ถ้าทำการเปรียบเทียบพหุคูณหลังจากการวิเคราะห์ความแปรปรวนด้วย F- test จะทำให้ผลของอัตราความคลาดเคลื่อนของ Type I error ลดลง เนื่องจากการทำ F-test สามารถป้องกันอัตราความคลาดเคลื่อนต่อการเปรียบเทียบ

Boardman and Moffitt (1971) ทำการศึกษาเปรียบเทียบวิธีวิเคราะห์พหุคูณ 5 วิธี คือ LSD, HSD, Scheffe's, Duncan และ S-N-K ด้วยกลุ่มตัวอย่างที่มีการแจกแจงแบบปกติขนาด 5, 10 และ 15 ในระดับการทดลองตั้งแต่ 2 ถึง 11 และทำการทดลองซ้ำ 1,000 ครั้ง เปรียบเทียบอัตราความคลาดเคลื่อน 2 แบบ คือ อัตราความคลาดเคลื่อนต่อการเปรียบเทียบ และอัตราความ

คลาดเคลื่อนต่อการทดลอง ผลการวิจัยพบว่าอัตราความคลาดเคลื่อนของวิธี LSD และวิธีของ Duncan จะมีอัตราความคลาดเคลื่อนเพิ่มขึ้นตามจำนวนค่าเฉลี่ย ส่วนวิธีของ Scheffe's เป็นวิธีเปรียบเทียบพหุคูณที่มีอัตราความคลาดเคลื่อนที่คงเดิมมากสุด

Carmer and Swanson (1973; อ้างถึงใน สุญาณี จิตตะยโสธร, 2524: 19) ทำการศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 และประเภทที่ 2 ของวิธีเปรียบเทียบพหุคูณแบบต่าง ๆ ด้วยแผนการทดลองแบบ B-k โดยเปรียบเทียบค่าเฉลี่ย 10 คู่ ซึ่งมี μ =100 σ^2 = 100 ขนาดกลุ่มตัวอย่างเท่ากับ 5, 10 และ 20 ตามลำดับ ทำการทดลองซ้ำ 4,000 ครั้ง และตั้งระดับนัยสำคัญเฉพาะที่ α = 0.05 ผลการวิจัยพบว่า อัตราความคลาดเคลื่อนที่เกิดจากวิธีการเปรียบเทียบพหุคูณของ Scheffe's มีค่าน้อยกว่าวิธีอื่น ๆ วิธีของ Tukey's และวิธีของ S-N-K ให้อัตราความคลาดเคลื่อนสูงกว่าอัตราความคลาดเคลื่อนที่ระบุ และวิธีของ Tukey's จะให้อัตราความคลาดเคลื่อนน้อยกว่าวิธีของ S-N-K และจะยังใช้ได้ผลดีแม้ว่ากลุ่มตัวอย่างจะมีขนาดใหญ่ขึ้น

Dallal (2001) กล่าวว่า วิธี Tukey's HSD test จะเป็นที่ยอมรับมากกว่าวิธี Bonferroni test เนื่องจากวิธี Tukey's HSD test เป็นวิธีที่ถูกสร้างเพื่อปรับแก้ข้อด้อยของวิธี Bonferroni test และวิธี Tukey's HSD test จะปฏิเสธสมมติฐานศูนย์เมื่อมีความแตกต่างระหว่างค่าเฉลี่ยมีมากจริง ๆ นอกจากนี้ในเอกสารได้ระบุว่า วิธีการทดสอบของ Tukey's HSD test ถูกนิยมนำไปใช้มากกว่าวิธี LSD เพราะมีความชัดเจนในการเปรียบเทียบมากกว่า

Saville (1990) ทำการทดสอบเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของวิธี LSD Unrestricted, LSD, DMRT, Tukey's HSD และ Waller and Duncan ผลการวิจัยพบว่า วิธี LSD Unrestricted สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุดและให้อำนาจการทดสอบสูงสุด

บุญชม ศรีสะอาด (2538: 358) กล่าวว่า การทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลังการวิเคราะห์ความแปรปรวนแต่ละวิธีอาจให้ผลเหมือนกันหรือให้ผลแตกต่างกัน โอกาสที่จะเกิดความคลาดเคลื่อนแบบ Type I error เรียงจากน้อยไปหามากคือ วิธีของ Scheffe's วิธีของ Tukey's วิธีของ Newman-Keuls และวิธีของ Duncan ในการวิเคราะห์ข้อมูลชุดเดียวกัน วิธีของ Scheffe's มีโอกาสที่จะพบนัยสำคัญของความแตกต่างน้อยกว่าวิธีอื่น และวิธีของ Duncan มีโอกาสที่จะพบนัยสำคัญมากกว่าวิธีอื่น บางครั้งเมื่อผู้วิจัยวิเคราะห์ความแปรปรวนพบค่า F มีนัยสำคัญจึงทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ การใช้วิธีของ Scheffe's หรือวิธีของ Newman-Keuls อาจไม่พบความแตกต่างในคู่ใด ๆ แต่เมื่อใช้วิธีของ Duncan อาจพบความแตกต่างในบางคู่ก็ได้ ผู้เขียนไม่ได้กล่าวถึงวิธีของ Duncan เพราะเห็นว่าโอกาสที่จะเกิดความคลาดเคลื่อนแบบ Type I error จะมากกว่า α ที่กำหนด

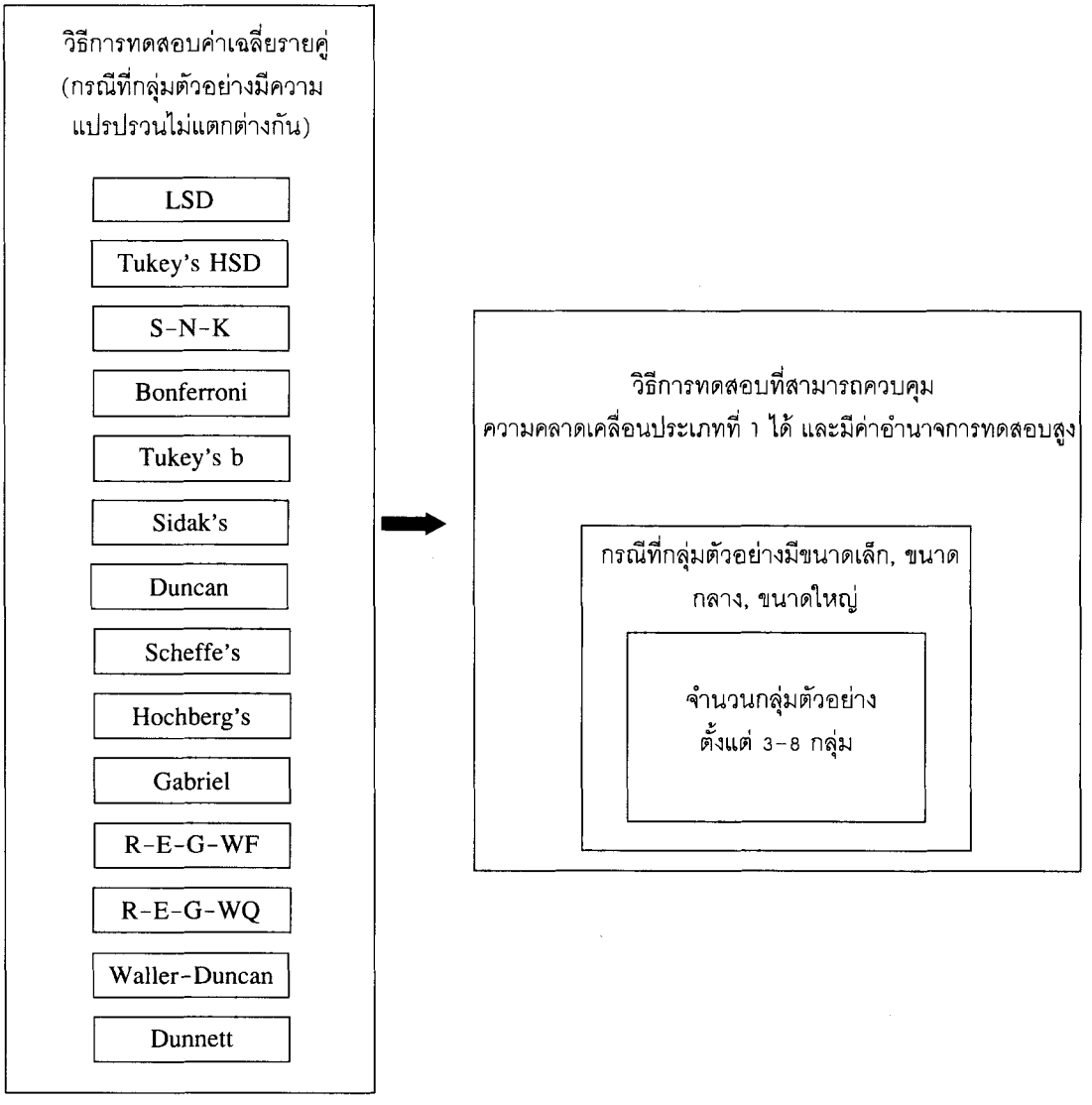
พหล ศักดิ์คะทัศน์ (2534) ทำการศึกษาเปรียบเทียบประสิทธิภาพการทดสอบของวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ย 3 วิธี คือวิธี Unrestricted LSD (U-LSD), วิธี Bonferroni (Dunn) T-test และวิธี Murphys Gap LSD (MG-LSD) โดยข้อมูลมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับ 1 ทำการทดลองด้วยเทคนิคมอนติคาร์โล ซิมูเลชั่น โดยมีจำนวนสิ่งทดลองเท่ากับ 2 ถึง 10 ทั้งแผนการทดลองแบบสุ่มสมบูรณ์และแผนการทดลองแบบบล็อกสุ่มสมบูรณ์ และจำนวนซ้ำมี 4 ระดับคือ 5, 10, 15 และ 20 ในแผนการทดลองแบบสุ่มสมบูรณ์ และจำนวนบล็อกมี 3 ระดับคือ 5, 7 และ 10 ในแผนการทดลองแบบบล็อกสุ่มสมบูรณ์ ผลการวิจัยพบว่า วิธี Murphys Gap LSD เหมาะสำหรับแผนการทดลองแบบสุ่มสมบูรณ์เมื่อจำนวนสิ่งทดลองเท่ากับ 2 ถึง 5 และเหมาะสำหรับแบบแผนการทดลองแบบบล็อกสุ่มสมบูรณ์เมื่อจำนวนสิ่งทดลองเท่ากับ 2 ถึง 8 ส่วนวิธี Unrestricted LSD จะเหมาะสำหรับแผนการทดลองแบบสุ่มสมบูรณ์เมื่อจำนวนสิ่งทดลองเท่ากับ 6 ถึง 10 แต่เมื่อแผนการทดลองเป็นแบบบล็อกสุ่มสมบูรณ์ กรณีที่สิ่งทดลองเท่ากับ 9 ถึง 10 ปรากฏว่าทั้ง 3 วิธีไม่มีวิธีใดที่เหมาะสมที่จะใช้ในการทดสอบ

สุชาดา บวรกิติวงศ์ (2548: 227-230) กล่าวว่า วิธีของ LSD จะถูกนำไปใช้ในการทดสอบมากที่สุด เนื่องจากจะสามารถค้นพบความแตกต่างได้ง่ายกว่าวิธีอื่นๆ แต่วิธีนี้จะมีจุดอ่อนคือไม่ได้ควบคุมอัตราความคลาดเคลื่อนทั้งหมดจึงส่งผลให้เกิดค่า α สูงขึ้นเมื่อจำนวนทดสอบมีหลายคู่ วิธีของ Bonferroni จะยากต่อการปฏิเสธสมมติฐานศูนย์ นอกจากคู่ที่ทดสอบจะมีค่าเฉลี่ยแตกต่างกันมากจริงๆ จึงเป็นวิธีที่นักวิจัยไม่นิยมนำมาใช้ และวิธีนี้จะมีจุดแข็งคือสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ทั้งหมด แต่วิธีนี้มีจุดอ่อนตรงที่ถ้าการทดสอบมีจำนวนคู่ในการทดสอบมากก็จะยิ่งเพิ่มโอกาสปฏิเสธสมมติฐานศูนย์มากขึ้น วิธีของ Scheffe's จะมีการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งหมด (Overall Type I error rates) จะมีคนนิยมใช้อยู่ในระดับปานกลาง วิธีของ Tukey's จะมีการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งหมด (Overall Type I error rates) จะใช้ได้ดีเมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีของ Hochberg's GT2 น่าจะให้ผลดีกว่าวิธีของ LSD, Bonferroni, Scheffe's และ Tukey's เนื่องจากการพัฒนาแต่ละครั้งจะพยายามแก้ไขจุดอ่อนของแต่ละวิธีที่มีอยู่

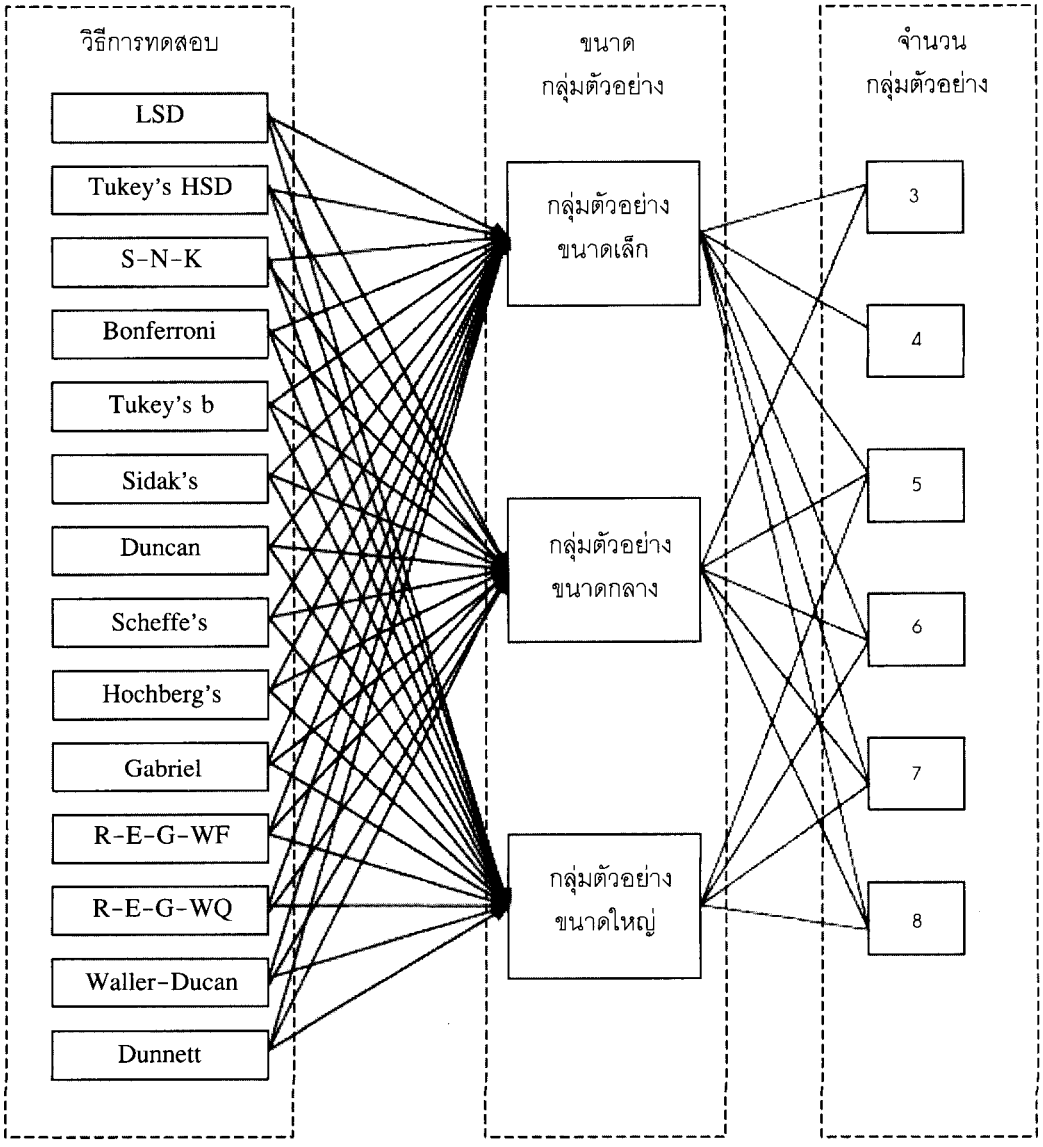
สุญาณี จิตตะยะโศธร (2524) ทำการศึกษาเปรียบเทียบความคลาดเคลื่อนประเภทที่ 1 จากข้อมูลที่ฝ่าฝืนข้อตกลงเบื้องต้นของการเปรียบเทียบพหุคูณโดยหาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการเปรียบเทียบพหุคูณ 5 วิธี คือ Tukey's, Duncan, Scheffe's, Dunnett และ S-N-K เขียนโปรแกรมคอมพิวเตอร์จำลองการทดลอง และทำการทดลอง 1,000 ครั้ง จำลองการทดลองเพื่อนับอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ด้วยการทดลองทดสอบสมมติฐานการเปรียบ-เทียบค่าเฉลี่ยเป็นรายคู่ของแต่ละวิธีจากกลุ่มตัวอย่าง 3 และ 4 กลุ่ม ที่มีขนาดกลุ่มตัวอย่าง 5, 10

และ 15 ในลักษณะการแจกแจงประชากรแบบปกติ แบบยูนิฟอร์ม และแบบเลปโตเคอร์ติคส์ กำหนดอัตราส่วนความแปรปรวนเท่ากันคือ 1:1:1 และ 1:1:1:1 สำหรับความเท่ากันของความแปรปรวนประชากรของกลุ่มตัวอย่างจำนวน 3 กลุ่ม และ 4 กลุ่ม ตามลำดับ และอัตราส่วนความแปรปรวนของประชากรเป็น 0.9:1:1.1 และ 0.8:1:1.2 สำหรับความไม่เท่ากันของความแปรปรวนประชากรของกลุ่มตัวอย่างจำนวน 3 กลุ่ม ผลการวิจัยพบว่า วิธีของทูกี้สามารถควบคุมความคลาดเคลื่อนตามที่กำหนดได้ 142 กรณี ในการทดลองทั้งสิ้น 228 กรณี ควบคุมไม่ได้ 86 กรณี ความคลาดเคลื่อนที่ควบคุมไม่ได้ส่วนใหญ่เป็นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีอัตราความคลาดเคลื่อนมากกว่าอัตราความคลาดเคลื่อนที่ระบุ วิธีของดันน์สามารถควบคุมความคลาดเคลื่อนตามที่กำหนดได้ 186 กรณีในการทดลองทั้งสิ้น 228 กรณี ควบคุมไม่ได้ 42 กรณี ความคลาดเคลื่อนที่ควบคุมไม่ได้ส่วนใหญ่เป็นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีอัตราความคลาดเคลื่อนมากกว่าอัตราความคลาดเคลื่อนที่ระบุในระดับ .05 และน้อยกว่าเมื่อระดับ .01 วิธีของเชฟเฟย์สามารถควบคุมความคลาดเคลื่อนตามที่กำหนดได้ 144 กรณีในการทดลองทั้งสิ้น 228 กรณี ควบคุมไม่ได้ 84 กรณี ความคลาดเคลื่อนที่ควบคุมไม่ได้ส่วนใหญ่เป็นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีอัตราความคลาดเคลื่อนน้อยกว่าอัตราความคลาดเคลื่อนที่ระบุ วิธีของดันเนตต์ สามารถควบคุมความคลาดเคลื่อนตามที่กำหนดได้ 31 กรณีในการทดลองทั้งสิ้น 228 กรณี ควบคุมไม่ได้ 197 กรณี ความคลาดเคลื่อนที่ควบคุมไม่ได้ส่วนใหญ่เป็นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีอัตราความคลาดเคลื่อนมากกว่าอัตราความคลาดเคลื่อนที่ระบุ วิธีของนิวแมนคูลส์สามารถควบคุมความคลาดเคลื่อนตามที่กำหนดได้ 62 กรณีในการทดลองทั้งสิ้น 228 กรณี ควบคุมไม่ได้ 166 กรณี ความคลาดเคลื่อนที่ควบคุมไม่ได้ส่วนใหญ่เป็นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองมีอัตราความคลาดเคลื่อนมากกว่าอัตราความคลาดเคลื่อนที่ระบุ

จากการศึกษาผลงานวิจัยและเอกสารที่เกี่ยวข้อง พบว่า ตัวสถิติทดสอบที่ใช้ในกรณีที่ประชากรมีความแปรปรวนเท่ากันจำนวน 14 วิธี ที่ได้กล่าวไปข้างต้นจะถูกนำไปใช้ในสถานการณ์ที่แตกต่างกันขึ้นอยู่กับขนาดตัวอย่าง จำนวนกลุ่มตัวอย่าง และวัตถุประสงค์ของนักวิจัย นอกจากนี้ในการวิจัยที่ผ่านมาจะกำหนดสถานการณ์ในการทดสอบไม่ครอบคลุมความเป็นไปได้ของเหตุการณ์ที่อาจจะเกิดขึ้นจริงในสภาพการวิจัยทางสังคมศาสตร์ ดังนั้น ผู้วิจัยจึงต้องการทำการทดสอบเปรียบเทียบ ค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ ในกรณีที่กลุ่มตัวอย่างมีความแปรปรวนเท่ากันจำนวน 14 วิธีที่มีอยู่ในโปรแกรม SPSS for Windows Version 11.0 เพื่อผลการวิจัยจะเป็นประโยชน์ต่อนักวิจัยทางสังคมศาสตร์นำไปใช้เป็นเกณฑ์ตัดสินใจในการเลือกวิธีการเปรียบเทียบค่าเฉลี่ยรายคู่ภายใต้สถานการณ์ต่าง ๆ ได้อย่างเหมาะสม ซึ่งมีกรอบแนวคิดในการวิจัยดังภาพ A และรูปแบบการทดลองตามภาพ B



ภาพ A กรอบแนวคิดในการวิจัย



ภาพ B รูปแบบการทดลอง

วัตถุประสงค์ในการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยต้องการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการเปรียบเทียบค่าเฉลี่ยรายคู่จำนวน 14 วิธี ได้แก่วิธี LSD, Tukey's HSD, Bonferroni, Tukey's b, Sidak, Duncan, Scheffe's, Hochberg's GT2, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K และ Dunnett ภายใต้เงื่อนไขประชากรทั้ง k กลุ่มมีค่าความแปรปรวนไม่แตกต่างกัน (equal variances assumed) และแต่ละกลุ่มมีการแจกแจงปกติ (Normal Distribution) ซึ่งพิจารณาเปรียบเทียบใน 2 กรณีคือ กรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันจะทำการเปรียบเทียบกลุ่มตัวอย่าง 3 ขนาด ได้แก่ กลุ่มตัวอย่างขนาดเล็ก (n=10) กลุ่มตัวอย่างขนาดกลาง (n=30) และกลุ่มตัวอย่างขนาดใหญ่ (n=60) กรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากันจะกำหนดให้จำนวนตัวอย่างเพิ่มขึ้นจาก 10 ครั้งละ 2 ในกลุ่มตัวอย่างขนาดเล็ก เพิ่มขึ้นจาก 30 ครั้งละ 4 ในกลุ่มตัวอย่างขนาดกลาง และเพิ่มขึ้นจาก 60 ครั้งละ 10 ในกลุ่มตัวอย่างขนาดใหญ่

แผนการดำเนินงาน

การวิจัยในครั้งนี้กำหนดสถานการณ์ต่างๆ สำหรับการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ โดยสร้างประชากรที่มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับหนึ่งในแผนการทดลองแบบสุ่มสมบูรณ์ที่กำหนดให้มีระดับทรีทเมนต์ 6 ระดับ คือ 3, 4, 5, 6, 7 และ 8 กลุ่มตามลำดับ สำหรับขนาดของกลุ่มตัวอย่างจะแบ่งเป็น 3 ขนาด คือ กลุ่มตัวอย่างขนาดเล็ก กลุ่มตัวอย่างขนาดกลาง และกลุ่มตัวอย่างขนาดใหญ่ ซึ่งผู้วิจัยกำหนดให้การทดลองมีสองกรณี คือ กรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน และกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ซึ่งตามปกติในงานวิจัยที่ใช้การ Simulation จะแบ่งข้อมูลเป็นขนาดต่างๆ ไม่เท่ากัน ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยจึงกำหนดให้มีกลุ่มตัวอย่าง 3 ขนาด โดยกำหนดให้ในกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันจะมีตัวอย่างขนาดเล็กเท่ากับ 10 ตัวอย่างขนาดกลางเท่ากับ 30 และตัวอย่างขนาดใหญ่เท่ากับ 60 และในกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ในกลุ่มตัวอย่างขนาดเล็กจะมีขนาดตัวอย่างเพิ่มขึ้นจาก 10 ครั้งละ 2 ในกลุ่มตัวอย่างขนาดกลางจะมีขนาดตัวอย่างเพิ่มขึ้นจาก 30 ครั้งละ 4 และในกลุ่มตัวอย่างขนาดใหญ่จะมีขนาดตัวอย่างเพิ่มขึ้นจาก 60 ครั้งละ 10 โดยใช้ระดับนัยสำคัญ α = 0.05 ดังนั้นจำนวนกลุ่มและขนาดตัวอย่างในการทดลองที่ใช้ในการวิจัยครั้งนี้สามารถอธิบายตามประเภทของการทดลองได้ดัง ตารางที่ 1 และตารางที่ 2

การสรุปอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบการทดลองต่างๆ ทำโดยการทำการทดสอบซ้ำ 10,000 ครั้งด้วยชุดตัวเลขสุ่มที่สร้างขึ้น แล้วนับจำนวนครั้งที่ผลการทดสอบปฏิเสธ

สมมติฐานศูนย์ ซึ่งพิจารณาโดยใช้เกณฑ์ของ Bradley ซึ่งผลการทดสอบที่แสดงว่าสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้จะต้องอยู่ในช่วง 0.025 ถึง 0.075 ดังนี้

$$0.025 < \frac{H}{n} < 0.075 \quad ; \quad \alpha = .05$$

เมื่อ H คือ จำนวนครั้งที่ปฏิเสธสมมติฐานศูนย์
n คือ จำนวนครั้งที่ทดสอบ

ตารางที่ 1 แบบการทดลองกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน

แบบการทดลอง	ความหมายของแบบการทดลอง	
	จำนวนกลุ่ม	ขนาดตัวอย่างในการทดลอง
(3,10)	3	10
(3,30)	3	30
(3,60)	3	60
(4,10)	4	10
(4,30)	4	30
(4,60)	4	60
(5,10)	5	10
(5,30)	5	30
(5,60)	5	60
(6,10)	6	10
(6,30)	6	30
(6,60)	6	60
(7,10)	7	10
(7,30)	7	30
(7,60)	7	60
(8,10)	8	10
(8,30)	8	30
(8,60)	8	60

ตารางที่ 2 แบบการทดลองกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน

แบบการทดลอง	ความหมายของแบบการทดลอง	
	จำนวนกลุ่ม	ขนาดตัวอย่างในการทดลอง
(3:10, 12,14)	3	10, 12, 14
(4:10, 12, 14, 16)	4	10, 12, 14, 16
(5: 10, 12, 14, 16, 18)	5	10, 12, 14, 16, 18
(6: 10, 12, 14, 16, 18, 20)	6	10, 12, 14, 16, 18, 20
(7: 10, 12, 14, 16, 18, 20, 22)	7	10, 12, 14, 16, 18, 20, 22
(8: 10, 12, 14, 16, 18, 20, 22, 24)	8	10, 12, 14, 16, 18, 20, 22, 24
(3: 30, 34, 38)	3	30, 34, 38
(4: 30, 34, 38, 42)	4	30, 34, 38, 42
(5: 30, 34, 38, 42, 46)	5	30, 34, 38, 42, 46
(6: 30, 34, 38, 42, 46, 50)	6	30, 34, 38, 42, 46, 50
(7: 30, 34, 38, 42, 46, 50, 54)	7	30, 34, 38, 42, 46, 50, 54
(8: 30, 34, 38, 42, 46, 50, 54, 58)	8	30, 34, 38, 42, 46, 50, 54, 58
(3: 60, 70, 80)	3	60, 70, 80
(4: 60, 70, 80, 90)	4	60, 70, 80, 90
(5: 60, 70, 80, 90, 100)	5	60, 70, 80, 90, 100
(6: 60, 70, 80, 90, 100, 110)	6	60, 70, 80, 90, 100, 110
(7: 60, 70, 80, 90, 100, 110, 120)	7	60, 70, 80, 90, 100, 110, 120
(8: 60, 70, 80, 90, 100, 110, 120, 130)	8	60, 70, 80, 90, 100, 110, 120, 130

เมื่อพบว่าแบบการทดลองใดสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนดผู้วิจัยจะทำการคำนวณหาอำนาจการทดสอบของสถิติทดสอบดังกล่าวด้วยการเพิ่มขนาดอิทธิพลเข้าไปในระดับทรีทเมนต์ของแผนการทดลอง เพื่อทำให้อิทธิพลในระดับทรีทเมนต์แตกต่างกัน และเพื่อดูความไว (Sensitiveness) ของตัวสถิติที่นำมาทดสอบ ดังนั้น ในงานวิจัยฉบับนี้ผู้วิจัยจึงกำหนดให้เพิ่มอิทธิพลทรีทเมนต์ใน 2 แบบ คือ แบบ A และแบบ B ซึ่งแสดงรายละเอียดตามตารางที่ 3

ตารางที่ 3 ขนาดการเพิ่มอิทธิพลทรีทเมนต์

แบบ	จำนวนกลุ่ม	กลุ่ม : ขนาดอิทธิพลที่บวกเพิ่ม							
		a	b	c	d	e	f	g	h
A	3	0.2	0.1	0.0					
	4	0.3	0.2	0.1	0.0				
	5	0.4	0.3	0.2	0.1	0.0			
	6	0.5	0.4	0.3	0.2	0.1	0.0		
	7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	
	8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
B	3	1.0	0.5	0.0					
	4	1.5	1.0	0.5	0.0				
	5	2.0	1.5	1.0	0.5	0.0			
	6	2.5	2.0	1.5	1.0	0.5	0.0		
	7	3.0	2.5	2.0	1.5	1.0	0.5	0.0	
	8	3.5	3.0	2.5	2.0	1.5	1.0	0.5	0.0

ในการวิจัยครั้งนี้ ผู้วิจัยใช้สูตรการคำนวณในการทดสอบกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน ($n_i = n_j$) และกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ($n_i \neq n_j$) ทั้ง 14 วิธี ดังแสดงในตารางที่ 4

ตารางที่ 4 สูตรการคำนวณของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ในแผนการทดลองแบบสุ่มสมบูรณ์ที่ α = 0.05 ภายใต้เงื่อนไขที่ว่าประชากรทั้ง k กลุ่มมีค่าความแปรปรวนไม่แตกต่างกัน และมีกำหนดแจกแจงปกติ

ที่	วิธีการทดสอบ	สูตรคำนวณ	
		$n_i = n_j$	$n_i \neq n_j$
1	Bonferroni test	$B = t_D\sqrt{\frac{2MSW}{n}}$	$B = t_D\sqrt{MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$
2	Duncan's	$D = q\sqrt{\frac{MSE}{n}}$	$D = q\sqrt{\frac{MSW}{2}(\frac{1}{n_i}+\frac{1}{n_j})}$
3	Dunnett's	$D = d_\alpha(k,v)\sqrt{\frac{2MSW}{n}}$	$D = d_\alpha(k,v)\sqrt{MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$
4	Gabriel' test	$G = \frac{\lvert\bar{x}_i - \bar{x}_j\rvert}{\sqrt{\frac{MSW}{2}(\frac{1}{n_i}+\frac{1}{n_j})}}$	
5	Hochberg's GT2	$GT2 = \frac{(\bar{x}_i - \bar{x}_j)}{\sqrt{MSW(\frac{1}{n_i}+\frac{1}{n_j})}}$	
6	LSD	$LSD = \sqrt{\frac{2(MSW)F}{n}}$	$LSD = \sqrt{MSW(\frac{1}{n_i}+\frac{1}{n_j})F}$
7	R-E-G-WF	$F_p = \frac{\sum_{j=1}^{5} n_j\bar{y}_j^2 - (\sum_{j=1}^{5} n_j\bar{y}_j)^2 / \sum_{j=1}^{5} n_j}{(S-1)MSW}$	
8	R-E-G-WQ	$Q_p = \frac{\max\bar{y} - \min\bar{y}}{\sqrt{\frac{MSW}{n}}}$	$Q_p = \max\left\{\frac{\lvert\bar{y}_i - \bar{y}_j\rvert}{\sqrt{\frac{MSW}{2}(\frac{1}{n_i}+\frac{1}{n_j})}}\right\}$
9	Scheffe's test	$CV_s = \sqrt{(k-1)(F^*)(MSW)(2/n)}$	
10	Sidak test	$DS = t_{DS}\sqrt{\frac{MSW}{n}}$	$DS = t_{DS}\sqrt{MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$
11	S-N-K	$SNK = q\sqrt{\frac{MSW}{n}}$	$SNK = q\sqrt{\frac{MSW}{2}(\frac{1}{n_i}+\frac{1}{n_j})}$
12	Tukey's b	$WSD = \frac{SNK + HSD}{2}, n_i = n_j$	$WSD = \frac{SNK + HSD}{2}, n_i \neq n_j$
13	Tukey's HSD	$HSD = q\sqrt{\frac{MSW}{n}}$	$HSD = q\sqrt{\frac{MSW}{2}(\frac{1}{n_i}+\frac{1}{n_j})}$
14	Waller-Duncan	$WD = t_B\sqrt{\frac{2MSW}{n}}$	$WD = t_B\sqrt{\frac{2MSW}{n_h}}\ ;\ n_h = \frac{k}{\sum_{i=1}^{k} n_i^{-1}}$

การสรุปผลจะพิจารณาจากค่าอำนาจการทดสอบร่วมด้วย เมื่อทำการทดสอบซ้ำ ๑๐,๐๐๐ ครั้ง ด้วยชุดตัวเลขสุ่มเสร็จแล้วจะนับจำนวนครั้งที่ยอมรับสมมติฐานศูนย์ เพื่อนำมาคำนวณหาอำนาจการทดสอบดังนี้

$$Power\ of\ the\ test = 1 - \frac{\chi}{n} \ ; \ \alpha = .05$$

เมื่อ χ คือ จำนวนครั้งที่ยอมรับสมมติฐานศูนย์
n คือ จำนวนครั้งที่ทดสอบ

สมมติฐานในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยในการทดลองแบบสุ่มสมบูรณ์ทุกระดับทรีทเมนต์เป็นดังนี้

$H_o : \mu_i = \mu_j$
H_1 : มีค่าเฉลี่ยอย่างน้อย ๑ คู่ที่ไม่เท่ากัน
โดยที่ $i \neq j$

ผลการวิเคราะห์ข้อมูล

๑. ผลการวิเคราะห์ค่าความคลาดเคลื่อนประเภทที่ ๑ กรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน พบว่าค่าความคลาดเคลื่อนประเภทที่ ๑ ของวิธีการทดสอบ ๑๒ วิธีจะมีค่าลดลงเมื่อจำนวนกลุ่มเพิ่มขึ้น ได้แก่ วิธี Tukey's HSD, Bonferroni, Tukey's b, Sidak, Duncan, Scheffe's, Hochberg's GT๒, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K และ Dunnett ทั้งนี้จะพบว่าวิธี Hochberg's GT๒ ที่ k=๖-๘ จะมีค่าความคลาดเคลื่อนใกล้เคียงกันเช่นเดียวกับวิธี Gabriel ที่ k=๕-๘ ก็จะมีค่าความคลาดเคลื่อนใกล้เคียงกันเช่นกัน และถ้าพิจารณาที่วิธีของ Scheffe's จะพบว่ามีค่าความคลาดเคลื่อนน้อยมาก และมีค่าเป็นศูนย์ตั้งแต่ k=๕-๘ ส่วนอีก ๒ วิธี อันได้แก่วิธี LSD และ Waller-Duncan จะพบว่า วิธี LSD จะมีค่าความคลาดเคลื่อนไม่เปลี่ยนแปลงตามจำนวนกลุ่มตัวอย่าง กล่าวคือ วิธีนี้จะมีค่าความคลาดเคลื่อนใกล้เคียงกันในทุกขนาดกลุ่มตัวอย่าง และวิธี Waller-Duncan จะมีค่าความคลาดเคลื่อนใกล้เคียงกันเมื่อ k=๓-๔ และเมื่อ k=๕-๘ จะพบว่ามีค่าความคลาดเคลื่อนสูงกว่าเกณฑ์ที่กำหนดมาก ทั้งนี้เมื่อพิจารณาวิธีการทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ ๑ ในภาพรวมทั้งหมดของการทดลองจะสามารถสรุปได้ว่า

๑.๑ วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ ๑ ได้ตามเกณฑ์ที่กำหนดทุกขนาดกลุ่มตัวอย่างตั้งแต่ k=๓ ถึง k=๘ มีด้วยกัน ๒ วิธี ได้แก่ วิธี LSD และ วิธี Duncan

๑.๒ วิธีการทดสอบที่มีอัตราความคลาดเคลื่อนประเภทที่ ๑ ต่ำกว่าเกณฑ์ที่กำหนดทุกขนาดกลุ่มตัวอย่างตั้งแต่ k=๓ ถึง k=๘ มีด้วยกัน ๕ วิธี ได้แก่ วิธี Bonferroni, วิธี Tukey's HSD, วิธี Hochbeg's, วิธี Scheffe's และ วิธี R-E-G-WQ

1.3 วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด เฉพาะกรณีที่มีจำนวนกลุ่มตัวอย่างน้อย แต่เมื่อจำนวนกลุ่มตัวอย่างมากจะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ที่กำหนดมีด้วยกัน 4 วิธี ได้แก่ วิธี Dunnett, วิธี S-N-K, วิธี Tukey's b และ วิธี R-E-G-WQ

1.4 วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด เฉพาะกรณีที่มีจำนวนกลุ่มตัวอย่างมาก แต่ถ้าจำนวนกลุ่มตัวอย่างน้อยจะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนดมีด้วยกัน 2 วิธี ได้แก่ วิธี Sidak และวิธี Gabriel

1.5 วิธีการทดสอบที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนดเมื่อจำนวนกลุ่มตัวอย่างมากขึ้นมี 1 วิธี ได้แก่ วิธี Waller-Duncan

2. ผลการวิเคราะห์ค่าความคลาดเคลื่อนประเภทที่ 1 กรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน พบว่า ค่าความคลาดเคลื่อนประเภทที่ 1 ของวิธีการทดสอบ 12 วิธีจะมีค่าลดลงเมื่อจำนวนกลุ่มเพิ่มขึ้น ได้แก่วิธี Tukey's HSD, Bonferroni, Tukey's b, Sidak, Duncan, Scheffe's, Hochberg's, R-E-G-WF, Gabriel, R-E-G-WQ, Waller-Duncan, S-N-K และ Dunnett ทั้งนี้จะพบว่าวิธี Hochberg's GT2 และวิธี Gabriel ที่ k=6-8 จะมีค่าความคลาดเคลื่อนใกล้เคียงกัน และถ้าพิจารณาที่วิธีของ Scheffe's จะพบว่ามีค่าความคลาดเคลื่อนน้อยมาก และมีค่าเป็นศูนย์ตั้งแต่ k=5-8 ส่วนอีก 2 วิธีอันได้แก่วิธี LSD และ Waller-Duncan จะพบว่า ค่าความคลาดเคลื่อนไม่เปลี่ยนแปลงตามจำนวนกลุ่มตัวอย่าง กล่าวคือ 2 วิธีนี้จะมีค่าความคลาดเคลื่อนใกล้เคียงกันในทุกขนาดกลุ่มตัวอย่าง ทั้งนี้ เมื่อพิจารณาวิธีการทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ในภาพรวมทั้งหมดของการทดลองจะสามารถสรุปได้ว่า

2.1 วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนดทุกขนาดกลุ่มตัวอย่าง ตั้งแต่ k=3 ถึง k=8 มีด้วยกัน 3 วิธี ได้แก่ วิธี LSD, วิธี Waller-Duncan และวิธี Duncan

2.2 วิธีการทดสอบที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ที่กำหนดทุกขนาดกลุ่มตัวอย่าง ตั้งแต่ k=3 ถึง k=8 มีด้วยกัน 6 วิธี ได้แก่ วิธี Bonferroni, วิธี Sidak, วิธี Tukey's HSD, วิธี Hochberg's, วิธี Scheffe's และวิธี R-E-G-WQ

2.3 วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด เฉพาะกรณีที่มีจำนวนกลุ่มตัวอย่าง 3 กลุ่ม แต่เมื่อจำนวนกลุ่มตัวอย่างมากกว่า 3 กลุ่มจะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ที่กำหนด มีด้วยกัน 3 วิธี ได้แก่ วิธี Dunnett, วิธี Tukey's b และวิธี R-E-G-WQ

2.4 วิธีการทดสอบที่ควบคุมความคลาดเคลื่อนประเภทที่ ๑ ได้ตามเกณฑ์ที่กำหนด เฉพาะกรณีที่มีจำนวนกลุ่มตัวอย่างมากมี ๑ วิธี ได้แก่ วิธี Gabriel แต่ถ้าจำนวนกลุ่มตัวอย่างน้อย จะมีอัตราความคลาดเคลื่อนประเภทที่ ๑ สูงกว่าเกณฑ์ที่กำหนด

3. ผลการวิเคราะห์อำนาจการทดสอบกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน เมื่อพิจารณาที่ความไว (Sensitiveness) ของวิธีการทดสอบ พบว่า การเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะให้ค่าอำนาจการทดสอบต่างกัน โดยการเพิ่มอิทธิพลทรีทเมนต์แบบ B จะทำให้ได้ค่าอำนาจการทดสอบมากกว่าการเพิ่มอิทธิพลทรีทเมนต์แบบ A เกิน 5 เท่าในทุกการทดสอบ และเมื่อพิจารณาที่จำนวนกลุ่มตัวอย่างในการทดสอบจะพบว่าเมื่อจำนวนกลุ่มเพิ่มขึ้นค่าอำนาจการทดสอบก็จะมากขึ้นเช่นเดียวกันกับขนาดของกลุ่มตัวอย่าง ซึ่งพบว่าในกลุ่มตัวอย่างขนาดใหญ่จะมีอำนาจการทดสอบมากกกว่ากลุ่มขนาดกลางและในกลุ่มขนาดกลางก็จะมีค่าอำนาจการทดสอบมากกว่ากลุ่มขนาดเล็ก และเมื่อพิจารณาวิธีการทดสอบที่มีค่าอำนาจการทดสอบสูงสุดในการเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะสามารถสรุปได้ดังนี้

3.1 เมื่อมีจำนวนกลุ่มตัวอย่าง 3 กลุ่ม และเพิ่มอิทธิพลทรีทเมนต์แบบ A จะพบว่า วิธี LSD และ Waller-Duncan จะมีอำนาจการทดสอบสูงสุดใกล้เคียงกันทั้ง 3 ขนาดกลุ่มตัวอย่าง แต่เมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B จะพบว่า ในกลุ่มขนาดเล็กจะมี 2 วิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน ได้แก่ วิธี Waller-Duncan และLSD เมื่อเป็นกลุ่มขนาดกลางและขนาดใหญ่จะมี 4 วิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกันได้แก่ วิธี Waller-Duncan, LSD, Duncan และ S-N-K

3.2 เมื่อมีจำนวนกลุ่มตัวอย่าง 4 กลุ่ม และเพิ่มอิทธิพลทรีทเมนต์แบบ A ในกลุ่มขนาดเล็กวิธี Gabriel และ Sidak จะเป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน แต่ในกลุ่มขนาดกลางวิธี Sidak จะเป็นวิธีที่มีอำนาจการทดสอบสูงสุด และเมื่อเป็นกลุ่มขนาดใหญ่จะมีวิธี LSD และ Waller-Duncan ที่มีค่าอำนาจการทดสอบสูงใกล้เคียงกัน เมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B จะพบว่าในกลุ่มตัวอย่างขนาดเล็กจะมีวิธี Gabriel และ Sidak เป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน และในกลุ่มขนาดกลางจะมีวิธี Sidak เป็นวิธีที่มีอำนาจการทดสอบสูงสุด แต่เมื่อเป็นกลุ่มขนาดใหญ่จะมี 4 วิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกันได้แก่ วิธี Waller-Duncan, S-N-K, Duncan และ LSD

3.3 เมื่อมีจำนวนกลุ่มตัวอย่าง 5 กลุ่ม และเพิ่มอิทธิพลทรีทเมนต์แบบ A ทั้ง 3 ขนาดกลุ่มตัวอย่างจะมีวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุด เมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B ในกลุ่มตัวอย่างขนาดเล็กจะมีวิธี Gabriel และ Sidak เป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน แต่เมื่อเป็นกลุ่มขนาดกลางและขนาดใหญ่จะมีวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุด นอกจากวิธีนี้แล้วยังพบว่าวิธีที่เหลือจะมีค่าอำนาจการทดสอบสูงใกล้เคียงกันทุกวิธี

3.4 เมื่อมีจำนวนกลุ่มตัวอย่าง 6-8 การเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะทำให้ทั้ง 3 ขนาดกลุ่มตัวอย่างมีวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุด

4. ผลการวิเคราะห์อำนาจการทดสอบกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน เมื่อพิจารณาที่ความไว (Sensitiveness) ของวิธีการทดสอบ พบว่า การเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะมีผลแตกต่างกันเช่นเดียวกับกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน นั่นคือ การเพิ่มอิทธิพลทรีทเมนต์แบบ B จะทำให้ค่าอำนาจการทดสอบของวิธีการทดสอบที่นำมาคำนวณหาอำนาจการทดสอบสูงกว่าการเพิ่มอิทธิพลทรีทเมนต์แบบ A มากกว่า 5 เท่าในทุกวิธี และเมื่อพิจารณาวิธีการทดสอบที่มีค่าอำนาจการทดสอบสูงสุดในการเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะสามารถสรุปได้ดังนี้

4.1 เมื่อมีจำนวนกลุ่มตัวอย่าง 3 กลุ่ม และเพิ่มอิทธิพลทรีทเมนต์แบบ A วิธี LSD และ Waller-Duncan เป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกันในกลุ่มตัวอย่างทั้ง 3 ขนาด และเมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B ในกลุ่มตัวอย่างขนาดเล็กจะมีวิธี LSD และ Waller-Duncan เป็นวิธีที่มีอำนาจการทดสอบสูงสุด แต่ในกลุ่มขนาดกลางและขนาดใหญ่จะมี 4 วิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน ได้แก่ วิธี S-N-K, Duncan, Waller-Duncan และ LSD

4.2 เมื่อมีจำนวนกลุ่มตัวอย่าง 4 กลุ่ม และเพิ่มอิทธิพลทรีทเมนต์แบบ A วิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุดในกลุ่มขนาดเล็ก แต่ในกลุ่มขนาดกลางและขนาดใหญ่วิธี LSD และ Waller-Duncan จะเป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน และเมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B วิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุดในกลุ่มขนาดเล็ก แต่ในกลุ่มขนาดกลางวิธี LSD, Waller-Duncan และ Duncan จะเป็นวิธีที่มีอำนาจการทดสอบสูงใกล้เคียงกัน และในกลุ่มขนาดใหญ่ทั้ง 4 วิธีที่นำมาทดสอบได้แก่ วิธี LSD, S-N-K, Duncan และ Waller-Duncan จะมีค่าอำนาจการทดสอบสูงใกล้เคียงกัน

4.3 เมื่อมีจำนวนกลุ่มตัวอย่าง 5-8 กลุ่ม การเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะมีวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุดเหมือนกันทั้ง 3 ขนาดกลุ่มตัวอย่าง และมีอีก 2 วิธีที่มีค่าอำนาจการทดสอบสูงรองลงมา ได้แก่ วิธี LSD และ Waller-Duncan และเมื่อเพิ่มอิทธิพลทรีทเมนต์แบบ B จะมีวิธี Gabriel เป็นวิธีที่มีอำนาจการทดสอบสูงสุดในกลุ่มตัวอย่างทั้ง 3 ขนาดเช่นเดียวกับการเพิ่มอิทธิพลทรีทเมนต์แบบ A

อภิปรายผลการวิจัย

ในการเลือกใช้สถิติการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ ผู้วิจัยจำเป็นต้องพิจารณาความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 และสถิติทดสอบที่มีอำนาจการทดสอบสูงสุด สำหรับงานวิจัยฉบับนี้เมื่อประชากรมีการแจกแจงปกติและประชากรทั้ง k กลุ่มมีความแปรปรวนไม่แตกต่างกันที่ระดับนัยสำคัญ 0.05 ในกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตั้งแต่ $k = 3$

กลุ่มถึง k =8 กลุ่ม พบว่า วิธี Scheffe's จะมีอัตราความคลาดเคลื่อนประเภทที่ ๑ น้อยมากทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน โดยที่ k=3 ในกลุ่มตัวอย่างทั้งสามขนาดจะมีอัตราความคลาดเคลื่อนอยู่ระหว่าง 0.005 ถึง 0.007 ที่ k=4 จะมีอัตราความคลาดเคลื่อนเท่ากับ 0.001 ในกลุ่มตัวอย่างทั้งสามขนาด และที่ k=5 ถึง k=8 จะพบว่าวิธี Schette's จะมีอัตราความคลาดเคลื่อนเป็น 0.000 ในกลุ่มตัวอย่างทั้งสามขนาดเช่นกัน แต่วิธีการทดสอบอื่นๆ จะไม่มีอัตราความคลาดเคลื่อนต่ำขนาดนี้ ซึ่งตรงกับการศึกษาของ Boardman and Moffitt ในส่วนที่ว่า "วิธีของ Scheffe's เป็นวิธีเปรียบเทียบพหุคูณที่มีอัตราความคลาดเคลื่อนที่คงเดิมมากสุด" และตรงกับการศึกษาของ Carmer and Swanson (1973) ด้วยเช่นกัน ซึ่ง Carmer and Swanson กล่าวไว้ว่า "อัตราความคลาดเคลื่อนที่เกิดจากวิธีการเปรียบเทียบพหุคูณของ Scheffe's มีค่าน้อยกว่าวิธีอื่นๆ" นอกจากนี้แล้วจากเอกสารของบุญชม ศรีสะอาด (2538) ก็กล่าวไว้ว่า "ในการวิเคราะห์ข้อมูลชุดเดียวกัน วิธีของ Scheffe's มีโอกาสที่จะพบนัยสำคัญของความแตกต่างน้อยกว่าวิธีอื่น" ดังนั้นถ้าพิจารณาอัตราความคลาดเคลื่อนโดยรวมก็สามารถอนุมานได้ว่าวิธี Scheffe's จะมีอัตราความคลาดเคลื่อนน้อยกว่าวิธีอื่นๆ และมีค่าคงที่มากที่สุด นอกจากวิธีของ Scheffe's แล้วผลการศึกษาของ Carmer and Swanson ยังพบว่า วิธีของ Tukey's จะให้อัตราความคลาดเคลื่อนน้อยกว่าวิธีของ S-N-K และจะยังใช้ได้ผลดีแม้ว่ากลุ่มตัวอย่างจะมีขนาดใหญ่ขึ้นซึ่งตรงกับผลการวิจัยในครั้งนี้เช่นกัน เพราะว่าวิธี Tukey's จะมีอัตราความคลาดเคลื่อนประเภทที่ ๑ ต่ำกว่าเกณฑ์ที่กำหนด และก็น้อยกว่าอัตราความคลาดเคลื่อนของวิธี S-N-K ในกลุ่มตัวอย่างทั้งสามขนาดและทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน

เมื่อพิจารณาที่วิธีการทดสอบ Bonferroni, Tukey's HSD, Hochberg's, R-E-G-WQ และ Scheffe's จะพบว่ามีอัตราความคลาดเคลื่อนประเภทที่ ๑ ต่ำกว่าเกณฑ์ที่กำหนดทุกขนาดการทดลอง ซึ่งสอดคล้องกับเอกสารของ สุชาดา บวรกิติวงศ์ (2548) ที่ระบุว่า "Bonferroni จะยากต่อการปฏิเสธสมมติฐานศูนย์ นอกจากคู่ที่ทดสอบจะมีค่าเฉลี่ยแตกต่างกันมากจริงๆ จึงเป็นวิธีที่นักวิจัยไม่นิยมนำมาใช้ และวิธีนี้จะมีจุดแข็งคือสามารถควบคุมความคลาดเคลื่อนประเภทที่ ๑ ทั้งหมด แต่วิธีนี้มีจุดอ่อนตรงที่ถ้าการทดสอบมีจำนวนคู่ในการทดสอบมากก็จะยิ่งเพิ่มโอกาสปฏิเสธสมมติฐานศูนย์มากขึ้น วิธีของ Scheffe's จะมีการควบคุมความคลาดเคลื่อนประเภทที่ ๑ ได้ทั้งหมด (overall Type I error rates) จะมีคนนิยมใช้อยู่ในระดับปานกลาง วิธีของ Tukey's จะมีการควบคุมความคลาดเคลื่อนประเภทที่ ๑ ได้ทั้งหมด จะใช้ได้ดีเมื่อขนาดกลุ่มตัวอย่างเท่ากัน" ดังนั้น เมื่อพิจารณาค่าความคลาดเคลื่อนประเภทที่ ๑ ของวิธี Bonferroni จะพบว่ายิ่งจำนวนกลุ่มเพิ่มขึ้นค่าความคลาดเคลื่อนประเภทที่ ๑ ก็จะยิ่งลดลงและต่ำกว่าเกณฑ์ที่กำหนดในทุกขนาดการทดลอง ทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน แสดงว่าวิธีนี้ไม่สามารถควบคุมอัตรา

ความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญ 0.05 เช่นเดียวกับวิธี Tukey's HSD, Hochberg's GT2, R-E-G-WQ และ Scheffe's และในกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากันวิธี Sidak ก็เป็นอีก 1 วิธีที่มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ที่กำหนดเช่นเดียวกับ 5 วิธีที่กล่าวไป

เมื่อพิจารณาที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของ 4 วิธี ได้แก่วิธี Tukey's HSD, Scheffe's, Duncan และ S-N-K ทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากันจะสามารถเรียงลำดับจากวิธีที่มีค่าน้อยไปมากได้เป็นวิธีของ Scheffe's, Tukey's HSD, S-N-K และ Duncan ตามลำดับ ซึ่งพบว่าตรงกับเอกสารของบุญชม ศรีสะอาด (2538) ที่กล่าวว่า "การทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลังการวิเคราะห์ความแปรปรวนแต่ละวิธีอาจให้ผลเหมือนกันหรือให้ผลแตกต่างกัน โอกาสที่จะเกิดความคลาดเคลื่อนแบบ Type I error เรียงจากน้อยไปหามากคือ วิธีของ Scheffe's วิธีของ Tukey's วิธีของ Newman-Keuls และวิธีของ Duncan" แต่ในส่วนของเอกสารที่ว่า "วิธีของ Duncan มีโอกาสที่จะพบนัยสำคัญมากกว่าวิธีอื่น" นั้นจะพบว่าไม่ตรงกับผลในการทดลองในงานวิจัยฉบับนี้เพราะว่าจากการค้นพบในครั้งนี้นั้นวิธี Duncan จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 น้อยกว่าวิธี LSD, Sidak, Waller-Duncan และ Gabriel ทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากันทุกแบบการทดลอง

เมื่อพิจารณาที่ความไว (Sensitiveness) ของตัวสถิติที่นำมาคำนวณหาอำนาจการทดสอบ พบว่า การเพิ่มอิทธิพลทรีทเมนต์แบบ A และแบบ B จะทำให้แต่ละวิธีการทดสอบได้ค่าอำนาจการทดสอบที่แตกต่างกันมากทั้งกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและไม่เท่ากัน โดยที่การเพิ่มอิทธิพลทรีทเมนต์แบบ B จะทำให้มีค่าอำนาจการทดสอบสูงกว่าการเพิ่มอิทธิพลทรีทเมนต์แบบ A มากกว่า 5 เท่าในทุกวิธีการทดสอบ ทั้งนี้เนื่องมาจากการเพิ่มอิทธิพลทรีทเมนต์แบบ B มีโอกาสทำให้ค่าเฉลี่ยของแต่ละกลุ่มแตกต่างกันมากกว่าการเพิ่มอิทธิพลทรีทเมนต์แบบ A นั่นเอง

นอกจากนี้ ในขั้นตอนการทำการทดลองผู้วิจัยได้ทำการลองทดสอบซ้ำ 1,000 ครั้ง เปรียบเทียบกับการทดสอบซ้ำ 10,000 พบว่า ในการทดลองซ้ำ 1,000 ครั้ง จะให้ผลการทดลองที่ไม่นิ่ง ในขนาดการทดลองในหลายวิธีการทดสอบ เช่น วิธี Dunnett และวิธี Tukey's b ในขนาดการทดลอง k =3, n = 10, $n_i = n_j$ พบว่า ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่นิ่ง ค่าที่ได้จะแตกต่างกันจากการให้ทดลองซ้ำ 1,000 ครั้ง ในครั้งที่ 1 กับครั้งที่ 2 และครั้งที่ 3 หรือมากกว่านั้น แต่เมื่อผู้วิจัยทำการทดสอบ โดยให้ทำการทดลองซ้ำ 10,000 ครั้ง พบว่า ผลการทดลองที่ได้จะนิ่ง คือมีค่าใกล้เคียงกันในการทดลองครั้งที่ 1 กับครั้งที่ 2 และครั้งที่ 3 หรือมากกว่านั้น

เนื่องจากในการวิจัยครั้งนี้ศึกษาเฉพาะข้อมูลเฉพาะกรณีที่ประชากรทั้ง k กลุ่มมีค่าความแปรปรวนไม่แตกต่างกันและมีการแจกแจงปกติ ที่ระดับนัยสำคัญ 0.05 แต่ในสภาพการวิจัย

บางอย่างจำเป็นต้องใช้ระดับนัยสำคัญในการทดสอบต่ำกว่า 0.05 ดังนั้นจึงควรมีการศึกษาเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ที่ระดับนัยสำคัญต่ำกว่า 0.05 นอกจากนี้ก็ควรที่จะศึกษาเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธีการทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ในแผนการทดลองแบบอื่น เช่น แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (Completely Randomized Block Design) แบบบล็อกคงที่ (Fixed-effect) และแบบบล็อกสุ่ม (Random-effect) เมื่อประชากรมีการแจกแจงปกติ และมีความแปรปรวนไม่แตกต่างกันรวมทั้งเมื่อประชากรมีการแจกแจงปกติและมีความแปรปรวนแตกต่างกัน

ตารางที่ 5 การเลือกใช้สถิติทดสอบเปรียบเทียบเฉลี่ยรายคู่ในแผนการทดลองแบบสุ่มสมบูรณ์ กรณีที่กลุ่มตัวอย่างมีความแปรปรวนไม่แตกต่างกัน (Equal variances assumed) และ $\alpha = 0.05$

กลุ่มตัวอย่างมีขนาดเท่ากัน ($n_i = n_j$; $10 \le n \le 60$)	
กลุ่ม (k)	**วิธีการทดสอบ**
3	Waller-Duncan LSD Duncan
4	Waller-Duncan LSD S-N-K Duncan
5	LSD S-N-K Gabriel Duncan
6, 7, 8	Sidak LSD Gabriel Duncan

กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ($n_i \ne n_j$; $10 \le n \le 130$)	
กลุ่ม (k)	**วิธีการทดสอบ**
3	Waller-Duncan LSD Duncan
4	Waller-Duncan LSD S-N-K Duncan
5, 6	Walter-Duncan LSD S-N-K Gabriel Duncan
7, 8	Waller-Duncan LSD Gabriel Duncan

จากการทดสอบสถิติทั้ง 14 วิธี ภายใต้เงื่อนไขต่างๆ ที่กำหนด สามารถสรุปวิธีการทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนด และมีอำนาจการทดสอบสูงเป็นข้อเสนอแนะเพื่อการเลือกใช้สถิติทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ ในแผนการทดลองแบบสุ่มสมบูรณ์ กรณีที่กลุ่มตัวอย่างมีความแปรปรวนไม่แตกต่างกัน (Equal variances assumed) ที่ระดับความเชื่อมั่น 95% (α = 0.05) และมีกลุ่มตัวอย่างตั้งแต่ 3 ถึง 8 กลุ่ม โดยผู้วิจัย พิจารณาจากวิธีที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทุกกรณี และวิธีดังกล่าวสามารถใช้ได้ทั้งในกลุ่มตัวอย่างขนาดเล็ก ($n_i = n_j =$ 10, $n_i \neq n_j$ เมื่อ $10 \leq n \leq 24$) ขนาดกลาง ($n_i = n_j =$ 30, $n_i \neq n_j$ เมื่อ $30 \leq n \leq 58$) และขนาดใหญ่ ($n_i = n_j$ 60, $n_i \neq n_j$ เมื่อ $60 \leq n \leq 130$) แบ่งเป็น 2 กรณี คือ กรณีที่กลุ่มตัวอย่างมีขนาดเท่ากันและกรณีที่กลุ่มตัวอย่างมีขนาดไม่เท่ากัน ดังแสดงในตารางที่ 5

รายการอ้างอิง

บุญชม ศรีสะอาด. (2538). *วิธีการทางสถิติสำหรับการวิจัย.* ภาควิชาพื้นฐานของการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม.

พหล ศักดิ์คะทัศน์. (2534). *การเปรียบเทียบประสิทธิภาพของวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยในการวิเคราะห์ความแปรปรวน.* วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สุชาดา บวรกิติวงศ์. (2548). *สถิติประยุกต์ทางพฤติกรรมศาสตร์.* กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุญาณี จิตตะยะโศธร. (2524). *การศึกษาโดยวิธีมอนติคาร์โล: การเปรียบเทียบความคลาดเคลื่อนประเภทที่1จากข้อมูลที่ฝ่าฝืนข้อตกลงเบื้องต้นของการเปรียบเทียบพหุคูณ.* วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

Bradley, J.V. (1968). *Distribution-free Statistical Test.* New Jersey: Prentice-Hall.

Bernhardson, C.S. (1975). Type I Error Rate When Multiple Comparison Procedures Follow a Significant F test of ANOVA. *Biometrics,* 31. (n.d.).

Dallat, G.E. (2001). PhD Scientist I, JM USDA HNRC. *Multiple Comparisons Procedures.* (n.p.).

Saville, D.J. (1990). Multiple Comparison procedure: The Practical Solution. *The American Statistician,* 44. (n.d.).

Boardman, T.J., & Moffitt, D.R. (1971). Graphical Monte Carlo Type I Error Rates for Multiple Comparison Procedures. *Biometrics,* September. (n.d.).