การศึกษาความแกร่งของวิธีสเต็ปดาวน์ บูทสแตรป มิน พี สำหรับการเปรียบเทียบค่าเฉลี่ย เมื่อมีทรีตเมนต์ควบคุม

Robustness Study of Step-down Bootstrap min P for Comparing Several Means with a Control

กมล บุษบา และอริญชย์ เพชรรัตน์

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความแกร่งของวิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี (Step-down Independent Bootstrap min P) และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) ที่ใช้สำหรับการเปรียบเทียบค่าเฉลี่ย กรณีที่มีทรีตเมนต์ควบคุม เปรียบเทียบกับสถิติทดสอบของดันเนตต์ โดยศึกษาภายใต้ลักษณะข้อมูลที่มีการแจกแจงล็อกนอร์มัล (Lognormal Distribution) ที่มีความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน กำหนดจำนวนทรีตเมนต์ที่ทำการศึกษามี 3 ทรีตเมนต์ และมีทรีตเมนต์ควบคุม 1 ทรีตเมนต์ ในแต่ละทรีตเมนต์มีจำนวนหน่วยทดลองเท่ากัน คือ 3, 5, 7, 10 และ 15 กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05 กำหนดจำนวนรอบของการสุ่มซ้ำแบบบูทสแตรปเท่ากับ 1,000 รอบ และกำหนดจำนวนการคัดลอก (copy) ชุดข้อมูลตัวอย่างสุ่มสำหรับวิธีการสุ่มซ้ำแบบบูทสแตรปที่ไม่เป็นอิสระต่อกัน (Dependent Bootstrap Resampling) เท่ากับ 2 และ 4 ชุด ตามลำดับ ทำการจำลองข้อมูลด้วยวิธีมอนติคาร์โล (Monte Carlo Method) ในแต่ละสถานการณ์จำนวน 1,000 ครั้ง

ผลการศึกษา พบว่า ภายใต้ข้อมูลที่มีการแจกแจงล็อกนอร์มัลที่มีความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน สถิติทดสอบทุกตัว สามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 ได้ ในทุกขนาดตัวอย่างที่ทำการศึกษา และพบว่า วิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีกำลังการทดสอบสูงกว่าสถิติทดสอบดันเนตต์ ยกเว้นที่ขนาดตัวอย่างน้อยมากซึ่งในที่นี้คือ 3

คำสำคัญ: สถิติทดสอบของดันเนตต์ สเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี สเต็ปดาวน์ดิเพนเดนท์บูทส-แตรป มิน พี การทดสอบแบบปิด

Abstract

The purpose of this research is to study the robustness of Step – down independent bootstrap min P procedure and Step – down dependent bootstrap min P procedure for comparing several means with a control. The procedures are further compared with the Dunnett's statistic. The population distribution of this study is lognormal distribution with equal variances for all treatments. In this study, 3 treatment groups and 1 control group are considered. The sample size for each testing group is equally distributed at 3, 5, 7, 10 and 15 respectively. The significance level (α) of the test is set at 0.05. Furthermore, the number of bootstrap resampling is 1,000 replicates. The number of copies of the sample data for the dependent bootstrap is being investigated at 2 and 4 copies. Monte Carlo simulation with 1,000 repetitions is currently a realization.

Under lognormal distribution with equal variances for all treatments, the probability of type I error of all test statistics can be controlled for all cases and the empirical power of Step – down independent bootstrap min P procedure and Step – down dependent bootstrap min P procedure are higher than Dunnett's statistic except for the cases where the sample size is equal to 3.

Keywords: Dunnett's test statistic two-sided, step-down independent bootstrap min P, step-down dependent bootstrap min P, closed testing

1. บทนำ

ในปัจจุบันวิธีการทางสถิติได้เข้ามามีบทบาทอย่างยิ่งต่อการวิจัยและการพัฒนาในงานทุกสาขา โดยข้อมูลที่นำมาวิเคราะห์ได้มาจากการทดลองหรือการสำรวจ ซึ่งก่อนทำการทดลองหรือสำรวจนั้นจะต้องมีการออกแบบและวางแผนเพื่อทำให้การเก็บรวบรวมข้อมูลนั้นมีประสิทธิภาพและนำไปวิเคราะห์ได้สะดวกและถูกต้องเป็นไปตามแผนการทดลองหรือแผนการสำรวจที่ได้วางไว้ การวิเคราะห์ความแปรปรวนนั้นเป็นอีกวิธีการหนึ่งทางสถิติที่นิยมใช้กันแพร่หลาย เพื่อใช้ในการศึกษาเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์

ในกรณีที่ต้องการทำการเปรียบเทียบระหว่าง ทรีตเมนต์อื่น ๆ กับทรีตเมนต์ควบคุม หรือทรีตเมนต์อื่น ๆ กับทรีตเมนต์ที่ให้ผลดีที่สุด ในการทดลองบางประเภทอาจมีทรีตเมนต์หนึ่งที่ใช้เป็นทรีตเมนต์ควบคุม ซึ่งการเปรียบเทียบความแตกต่างระหว่างทรีตเมนต์อื่น ๆ กับทรีตเมนต์ที่ใช้เป็นทรีตเมนต์ควบคุม ซึ่งวิธีการที่ใช้ในกรณีนี้ได้แก่ สถิติทดสอบของดันเนตต์ ซึ่งวิธีนี้จะควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบชนิดเข้ม (Strong Familywise Error Rate) [1]

ในการทดสอบสมมติฐานแต่ละการทดสอบนั้นจะมีอัตราการเกิดความผิดพลาดประเภทที่ 1 (Type I error) เกิดขึ้น ซึ่งผู้วิจัยสามารถควบคุมอัตราการเกิดความผิดพลาดเหล่านี้ได้ ดังนั้นจุดประสงค์ของการเปรียบเทียบพหุคูณเป็นการควบคุมค่าที่มากที่สุดของอัตราการเกิดความผิดพลาดประเภทที่ 1 (Maximum overall Type I error rate) หมายถึง ค่าที่มากที่สุดของความน่าจะเป็นที่จะปฏิเสธสมมติฐาน

ว่างอย่างน้อยหนึ่งสมมติฐาน เมื่อสมมติฐานว่างนั้นเป็นจริง โดยเรียกอัตราความผิดพลาดดังกล่าวว่า “อัตราความผิดพลาดสูงสุดต่อการทดลอง (Maximum Experimentwise Error Rate: MEER)” หรือ “อัตราความผิดพลาดสูงสุดต่อวงศ์การทดสอบ (Maximum Familywise Error Rate: FWE)” [2]

ในปัจจุบันการเปรียบเทียบพหุคูณนั้นได้นำ “การทดสอบแบบปิด (Closed testing)” เข้ามาประยุกต์ใช้ ซึ่งมีการทดสอบแบบขั้นบันได (Step-wise method) และการทดสอบแบบขั้นตอนเดียว (Single-step method) เมื่อกำหนดให้ $H = \{H_{01}, H_{02}, ..., H_{0k}\}$ คือ เซตของสมมติฐานว่างที่เราต้องการทดสอบจำนวน k การทดสอบ ถ้าความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 ในการทดสอบอย่างน้อยหนึ่งการทดสอบ มีค่าไม่เกินระดับนัยสำคัญ α โดยไม่คำนึงถึงว่าจริง ๆ แล้วมีสมมติฐานใดบ้างในเซตของการทดสอบที่เป็นจริง การเปรียบเทียบพหุคูณนั้นจะควบคุม FWE ได้อย่างเข้มงวด สำหรับการทดสอบแบบขั้นบันไดที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ วิธีความแตกต่างที่มีนัยสำคัญน้อยที่สุด (Least Significant Difference - LSD) ของ ฟิชเชอร์ (Fisher) ซึ่งจะทำการทดสอบที (t-test) เพื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ในแต่ละสมมติฐาน ที่ระดับนัยสำคัญ α หลังจากที่เราปฏิเสธสมมติฐานว่างในการวิเคราะห์ความแปรปรวนด้วยสถิติเอฟ จะเห็นได้ว่าวิธีการนี้จะควบคุม FWE ได้ไม่เข้มงวด เพราะจะควบคุมได้เฉพาะกรณีที่ทุกการทดสอบใน H เป็นจริงเท่านั้น และสำหรับการทดสอบแบบขั้นตอนเดียวที่นิยมใช้กันวิธีหนึ่ง คือ วิธีบอนเฟอร์โรนี (Bonferroni) ซึ่งจะทำการทดสอบที (t-test) เพื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ในแต่ละสมมติฐานใน H ที่ระดับนัยสำคัญ α / k โดยไม่ต้องพิจารณา ผลการทดสอบจากการวิเคราะห์ความแปรปรวนด้วยสถิติเอฟ (F-test) ซึ่งจะเห็นได้ว่าวิธีการนี้จะควบคุม FWE ได้อย่างเข้มงวด [3]

การทดสอบแบบขั้นบันได (Step-wise) สามารถแบ่งออกเป็น 2 ขั้น คือ วิธีสเต็ปอัพ (Step-up procedure) และวิธีสเต็ปดาวน์ (Step-down procedure) ซึ่งในการทดสอบสมมติฐานนั้นจะมีขั้นตอนการพิจารณาค่า p-value แตกต่างกัน สำหรับวิธีที่ใช้ในวิธีสเต็ปอัพ ได้ประยุกต์ใช้ “Simes inequality” ในการควบคุม FWE [4]

ในบางสถานการณ์ที่เราไม่ทราบการแจกแจงร่วมของสถิติทดสอบ เราจะอาศัยวิธีการสุ่มซ้ำ เช่น วิธีการสุ่มซ้ำด้วยวิธีการเปลี่ยนลำดับ (permutation resampling), วิธีการสุ่มซ้ำด้วยวิธีการบูทสแตรป (bootstrap resampling) สามารถใช้ในการประมาณค่า p-value ที่ไม่ได้ปรับค่า (unadjusted p-values) และค่า p-value ที่ปรับค่า (adjusted p-values) โดยไม่ต้องคำนึงถึงข้อตกลงเกี่ยวกับค่าพารามิเตอร์ของการแจกแจงร่วมของสถิติทดสอบ

คำสั่ง PROC MULTTEST ใน SAS® ถูกเขียนขึ้นเพื่อจุดประสงค์ในการทดสอบพหุคูณและการเปรียบเทียบพหุคูณโดยตรง ซึ่งเสนอวิธีการปรับค่าพี โดยใช้ บอนเฟอร์โรนี , Šidák, การสุ่มซ้ำด้วยวิธีการบูทสแตรป และการสุ่มซ้ำด้วยวิธีการเปลี่ยนลำดับ ซึ่งการปรับค่าตามวิธีดังกล่าวใช้กับวิธีการทดสอบแบบขั้นตอนเดียว และวิธี สเต็ปดาวน์ ซึ่งในคำสั่ง PROC MULTTEST นี้ถูกออกแบบมาให้สามารถปรับใช้ได้ในกรณีที่ข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น โดยใช้การสุ่มซ้ำด้วยวิธีการบูทสแตรป และการสุ่มซ้ำด้วยวิธีการเปลี่ยนลำดับ [5]

ในปี ค.ศ. 2008 Komonniramit ได้ศึกษาวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) สำหรับการเปรียบเทียบค่าเฉลี่ยกรณีที่มีทรีตเมนต์ควบคุมเปรียบเทียบกับสถิติทดสอบของดันเนตต์ โดยศึกษาภายใต้ข้อมูลที่มีการแจกแจงปรกติและขนาดตัวอย่างเท่ากัน โดยกำหนดให้มี 3 ทรีตเมนต์ และมี 1 ทรีตเมนต์ควบคุม ซึ่งผลการศึกษา พบว่า วิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีประสิทธิภาพใกล้เคียงกับสถิติทดสอบของดันเนตต์ [6]

ในการวิเคราะห์ทางสถิติหรือการทดสอบต่าง ๆ ข้อมูลหรือประชากรที่นำมาพิจารณาส่วนใหญ่ต้องเป็น ไปตามข้อตกลงเบื้องต้น ในการทดสอบความแตกต่างของค่าเฉลี่ยระหว่างประชากรหลายกลุ่มนั้น มีข้อตกลงเบื้องต้นว่าด้วยประชากรต้องได้มาอย่างสุ่ม มีการแจกแจงปรกติและความแปรปรวนของแต่ละกลุ่มเท่ากันและประชากรแต่ละกลุ่มต้องเป็นอิสระกัน อย่างไรก็ตามถ้าข้อมูลในการทดสอบดังกล่าวไม่เป็นไปตามข้อตกลงเบื้องต้น ผลลัพธ์ที่ได้จะมีผลต่อระดับนัยสำคัญและกำลังการทดสอบอย่างไร ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาความแกร่งของวิธีสเต็ปดาวน์ บูทสแตรป มิน พี ภายใต้ลักษณะข้อมูลที่ไม่เป็นไปตามข้อตกลงเบื้องต้น โดยศึกษาภายใต้ลักษณะข้อมูลที่มีการแจกแจงล็อกนอร์มัล ที่มีความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน ซึ่งเป็นข้อมูลที่พบกันอยู่มากในการศึกษาทดลอง เช่น ข้อมูลปริมาณน้ำฝนในแต่ละเดือน ระยะเวลาที่พนักงานให้บริการลูกค้าต่อคน หรือระยะเวลาที่ลูกค้ารอจนกว่าจะได้รับบริการ เป็นต้น และการแจกแจงล็อกนอร์มัล ยังนำไปประยุกต์ใช้กันมากในทางวิศวกรรม เศรษฐศาสตร์ และการประกันภัย

2. สถิติทดสอบที่ใช้ในการวิจัย

ในการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์อื่น ๆ กับทรีตเมนต์ควบคุมนั้น สถิติทดสอบที่นำมาศึกษาในการวิจัยครั้งนี้ ได้แก่

2.1 สถิติทดสอบของดันเนตต์ (Dunnett's two sided test statistic) [7]

แผนการทดลองทั่วไปแบ่งทรีตเมนต์ออกเป็น 2 ประเภท คือ (1) ทรีตเมนต์ควบคุม (specified control treatment) และ (2) ทรีตเมนต์อื่น ๆ ในการเปรียบเทียบพหุคูณโดยสถิติทดสอบของดันเนตต์นั้น มีวัตถุประสงค์เพื่อต้องการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์ต่าง ๆ กับ ทรีตเมนต์ควบคุม โดยมีวิธีการดังนี้

สมมติฐานการทดสอบ คือ

$$H_{0i} : \mu_c = \mu_i$$
$$H_{1i} : \mu_c \neq \mu_i$$

สถิติทดสอบ คือ

$$D_i = \frac{|\bar{y}_c - \bar{y}_i|}{\sqrt{MSE(1/n_c + 1/n_i)}}$$

โดยที่ $$MSE = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{k(n-1)}$$

กำหนดให้ y_{ij} แทน ค่าสังเกตที่ j จากทรีตเมนต์ที่ i ($i = 1, 2, 3, \ldots, k$ ทรีตเมนต์ และ $j = 1, 2, 3, \ldots, n$ ซ้ำ)

$\bar{y}_c$ แทน ค่าเฉลี่ยของทรีตเมนต์ควบคุม

$\bar{y}_i$ แทน ค่าเฉลี่ยของทรีตเมนต์กลุ่มที่ i โดยที่ $i = 1, 2, \ldots, k$-1

n_c แทน ขนาดตัวอย่าง(จำนวนซ้ำ) ของทรีตเมนต์ควบคุม

n_i แทน ขนาดตัวอย่าง(จำนวนซ้ำ) ของทรีตเมนต์กลุ่มที่ i

MSE แทน ความคลาดเคลื่อนจากการวิเคราะห์ความแปรปรวน

2.2 วิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี (Step-down Independent Bootstrap min P)

สูตรที่ใช้ในการหาปรับค่า p-value สำหรับวิธี สเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี คือ

$$\tilde{p}_{(i)} = \max_{m=1,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,k\}} P_l \le p_{(m)} \mid H_0\right)\right\}$$

ในการศึกษาครั้งนี้ กำหนดให้ทรีตเมนต์ที่ทำการศึกษามี 3 ทรีตเมนต์ และมีทรีตเมนต์ควบคุม 1 ทรีตเมนต์ เมื่อกำหนดให้ทรีตเมนต์ที่ 1 เป็นทรีตเมนต์ควบคุม ($\mu_c = \mu_1$) ดังนั้นสมมติฐานการทดสอบทั้งหมดในการเปรียบเทียบค่าเฉลี่ยรายคู่ระหว่างทรีตเมนต์ควบคุมกับทรีตเมนต์อื่น ๆ มีดังนี้

$$H_{02}: \mu_1 = \mu_2 \quad \text{vs}$$
$$H_{12}: \mu_1 \neq \mu_2$$
$$H_{03}: \mu_1 = \mu_3 \quad \text{vs}$$
$$H_{13}: \mu_1 \neq \mu_3$$
$$H_{04}: \mu_1 = \mu_4 \quad \text{vs}$$
$$H_{14}: \mu_1 \neq \mu_4$$

สถิติทดสอบ

$$t_i = \frac{\bar{y}_1 - \bar{y}_i}{\sqrt{MSE(2/n)}} \text{ , โดยที่ } i = 2, 3, 4$$

ทำการปรับค่า p-value จากสูตร

$$\tilde{p}_{(i)} = \max_{m=2,\ldots,i}\left\{\Pr\left(\min_{l\in\{m,\ldots,4\}} P_l \le p_{(m)} \mid H_0\right)\right\}$$

โดยพิจารณาดังนี้

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \le p_{(2)}\right)\right\} = \Pr\left(\min(P_2, P_3, P_4) \le p_{(2)}\right)$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \le p_{(2)}\right), \Pr\left(\min_{l\in\{3,4\}} P_l \le p_{(3)}\right)\right\}$$

$$\tilde{p}_{(4)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \le p_{(2)}\right), \Pr\left(\min_{l\in\{3,4\}} P_l \le p_{(3)}\right), \Pr\left(\min_{l\in\{4\}} P_l \le p_{(4)}\right)\right\}$$

โดยที่ $p_{(m)}$; m = 2,3,4 คือ ค่า p-value ที่คำนวณได้จากการทดสอบสมมติฐานข้างต้นจำนวน 3 สมมติฐาน โดยคำนวณจากข้อมูลเริ่มต้นที่ได้ทำการจำลองโดยเรียงค่าจากน้อยไปมาก เพื่อใช้ในการพิจารณาปรับค่า p-value

P_l คือ ค่า p-value ที่คำนวณได้จากการทดสอบสมมติฐานข้างต้นจำนวน 3 สมมติฐาน หลังจากการสุ่มซ้ำแบบคืนที่โดยวิธีบูทสแตรป

2.3 วิธีสเต็ปดาวน์ดิเพนเดนท์ บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P)

สูตรที่ใช้ในการหาค่าสถิติทดสอบสำหรับวิธีสเต็ปดาวน์ดิเพนเดนท์ บูทสแตรป มิน พี นั้นพิจารณาเหมือนกับวิธีการสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี แต่จะแตกต่างกันที่การพิจารณาหาค่า P_l ซึ่งจะคำนวณค่า p-value ที่ได้จากการทดสอบสมมติฐานข้างต้นจำนวน 3 สมมติฐาน หลังจากการสุ่มซ้ำแบบไม่คืนที่โดยวิธีบูทสแตรป ที่ได้จากการคัดลอกข้อมูลที่ได้จากการจำลองขั้นแรก ตามจำนวนการคัดลอกที่กำหนด

2.4 ตัวอย่างการปรับค่า p-value

สมมุติว่าข้อมูลจากการทดลองหนึ่งที่มีทรีตเมนต์ 3 ทรีตเมนต์และแต่ละทรีตเมนต์มีจำนวน 3 ซ้ำ แสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลสมมติจากการทดลองที่มี 4 ทรีตเมนต์ และมีจำนวนซ้ำ 4 ซ้ำ

จำนวนซ้ำ	ทรีตเมนต์			
	1	**2**	**3**	**4**
1	89.8	84.4	64.4	75.2
2	93.8	116.0	79.8	62.4

3	88.4	84.0	88.0	62.4
4	112.6	68.6	69.4	73.8
ค่าเฉลี่ย	**96.2**	**88.3**	**75.4**	**68.5**

สมมติให้ทรีตเมนต์ที่ 1 เป็นทรีตเมนต์ควบคุม คำนวณค่า $p_{(m)}$ จากสูตร $t_i = \frac{\bar{y}_1 - \bar{y}_i}{\sqrt{MSE(2/n)}}$, โดยที่ $i = 2, 3, 4$ ได้ค่าดังนี้

$H_{02} : \mu_1 = \mu_2$ vs $H_{12} : \mu_1 \neq \mu_2$ ให้ค่า p_{02} = 0.4092

$H_{03} : \mu_1 = \mu_3$ vs $H_{13} : \mu_1 \neq \mu_3$ ให้ค่า p_{03} = 0.0443

$H_{04} : \mu_1 = \mu_4$ vs $H_{14} : \mu_1 \neq \mu_4$ ให้ค่า p_{04} = 0.0111

จากนั้นเรียงค่า p-value ที่คำนวณได้ จากน้อยไปมาก ทำให้ได้ว่า

$p_{(2)} = p_{04} = 0.0111$

$p_{(3)} = p_{03} = 0.0443$

$p_{(4)} = p_{02} = 0.4092$

จากนั้นทำการสุ่มข้อมูลใหม่ภายใต้สมมติฐาน $H : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ซึ่งในที่นี้สุ่มซ้ำตามวิธีของบูทสแตรปที่เป็นอิสระต่อกัน (Independent Bootstrap Resampling)

ตารางที่ 2 ตัวอย่างข้อมูลที่ได้จากการสุ่มซ้ำรอบที่ 1

จำนวนซ้ำ	**ทรีตเมนต์**			
	1	**2**	**3**	**4**
1	68.6	64.4	79.8	62.4
2	89.8	69.4	116.0	64.4
3	89.8	68.6	62.4	84.4
4	84.0	79.8	79.8	79.8
ค่าเฉลี่ย	**83.1**	**70.6**	**84.5**	**72.8**

คำนวณค่า P_l จากสูตร $t_i = \frac{\bar{y}_1 - \bar{y}_i}{\sqrt{MSE(2/n)}}$, โดยที่ $i = 2, 3, 4$ ได้ค่าดังนี้

$H_{02} : \mu_1 = \mu_2$ vs $H_{12} : \mu_1 \neq \mu_2$ ให้ค่า P_l = 0.2273

$H_{03} : \mu_1 = \mu_3$ vs $H_{13} : \mu_1 \neq \mu_3$ ให้ค่า P_l = 0.8851

$H_{04} : \mu_1 = \mu_4$ vs $H_{14} : \mu_1 \neq \mu_4$ ให้ค่า P_l = 0.3151

หลังจากนั้นทำการสุ่มข้อมูลใหม่ภายใต้สมมติฐาน $H : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ในการสุ่มซ้ำจนครบจำนวนรอบของการสุ่มซ้ำ ซึ่งในที่นี้ยกตัวอย่างการสุ่มซ้ำ 3 รอบ ซึ่งได้ค่า $p_{(m)}$ และ P_l ของการสุ่มซ้ำแต่ละรอบดังตารางที่ 3

ตารางที่ 3 การปรับค่า p-value

รายละเอียด	**เรียงลำดับสมมติฐานตามค่า $p_{(m)}$**		
	$p_{(2)}$	$p_{(3)}$	$p_{(4)}$
สมมติฐาน	$H_{04} : \mu_1 = \mu_4$	$H_{03} : \mu_1 = \mu_3$	$H_{02} : \mu_1 = \mu_2$
ค่า $p_{(m)}$	0.0111	0.0443	0.4092
ค่า P_l รอบที่	0.3151	0.8851	0.2273
ค่า P_l รอบที่	0.3350	0.3869	0.8605
ค่า P_l รอบที่	0.9124	0.6841	0.4247
ค่า p-value ที่ปรับค่า	$\tilde{p}_{(2)} = 0$	$\tilde{p}_{(3)} = 0$	$\tilde{p}_{(4)} = 0.3333$

จากตารางที่ 3 ค่า p-value ที่ปรับค่าแล้ว คือ

$$\tilde{p}_{(2)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \leq p_{(2)}\right)\right\} = \Pr\left(\min(P_2, P_3, P_4) \leq p_{(2)}\right) = 0$$

$$\tilde{p}_{(3)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \leq p_{(2)}\right), \Pr\left(\min_{l\in\{3,4\}} P_l \leq p_{(3)}\right)\right\}$$

$$\tilde{p}_{(4)} = \max\left\{\Pr\left(\min_{l\in\{2,3,4\}} P_l \leq p_{(2)}\right), \Pr\left(\min_{l\in\{3,4\}} P_l \leq p_{(3)}\right), \Pr\left(\min_{l\in\{4\}} P_l \leq p_{(4)}\right)\right\}$$

$$= \max(0,0,\frac{1}{3})$$

$$= \frac{1}{3}$$

$$= 0.3333$$

3. วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาเปรียบเทียบความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 และกำลังการทดสอบของสถิติทดสอบ ที่ใช้ในการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์ กรณีมีทรีตเมนต์ควบคุม เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์กับทรีตเมนต์ควบคุม โดยทำการจำลองและประมวลผลด้วยโปรแกรม R เวอร์ชัน 2.9.0 โดยมีขั้นตอนดังนี้

- กำหนดให้ทรีตเมนต์ที่ทำการศึกษามีจำนวน 3 ทรีตเมนต์ และมีทรีตเมนต์ควบคุม 1 ทรีตเมนต์ โดยกำหนดให้ทรีตเมนต์ที่ 1 เป็นทรีตเมนต์ควบคุม

- กำหนดให้ขนาดตัวอย่างในแต่ละทรีตเมนต์ (จำนวนซ้ำ) มีจำนวนซ้ำเท่ากันคือ 3, 5, 7, 10 และ 15

- ประชากรที่ใช้ในการศึกษามีการแจกแจงล็อกนอร์มัล โดยมีค่าความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน คือ 0.5, 1, 1.5, และ 2

- การประมาณค่าความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 กำหนดให้ค่าเฉลี่ยในแต่ละทรีตเมนต์เท่ากัน ซึ่งมีค่าเท่ากับ 0 ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$) และความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน โดยกำหนดให้เป็น $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 0.5, 1, 1.5,$ และ 2

- การประมาณค่ากำลังการทดสอบ กำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมมีค่าแตกต่างกับค่าเฉลี่ยใน ทรีตเมนต์อื่น ๆ โดยกำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมเท่ากับ 2 และค่าเฉลี่ยในทรีตเมนต์อื่นๆ เท่ากัน ซึ่งมีค่าเท่ากับ 0 ($\mu_1 = 2$, $\mu_2 = \mu_3 = \mu_4 = 0$) และความแปรปรวนในแต่ละทรีตเมนต์เท่ากันโดยกำหนดให้เป็น $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 0.5, 1, 1.5,$ และ 2

- กำหนดระดับนัยสำคัญเท่ากับ 0.05

- ในการศึกษาวิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี (Step-down Independent Bootstrap min P) จะกำหนดจำนวนรอบของการสุ่มซ้ำแบบบูทสแตรปที่เป็นอิสระต่อกัน (Independent Bootstrap Resampling) เท่ากับ 1000 รอบ และวิธีสเต็ปดาวน์ดิเพนเดนท์ บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) จะกำหนดจำนวนการคัดลอก (copy) ชุดข้อมูลตัวอย่างสุ่มสำหรับวิธีการสุ่มซ้ำแบบบูทสแตรปที่ไม่เป็นอิสระต่อกัน (Dependent Bootstrap Resampling) เท่ากับ 2 และ 4 ชุดตามลำดับ

- ในการศึกษาครั้งนี้ทำการจำลองข้อมูลด้วยวิธีมอนติคาร์โล (Monte Carlo Method) ในแต่ละสถานการณ์จำนวน 1,000 ครั้ง

4. ผลการวิจัย

ผู้วิจัยจะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่าง ๆ ในตาราง

Dunnett หมายถึง สถิติทดสอบของดันเนตต์ (Dunnett's two sided test statistic)

IB หมายถึง สถิติทดสอบของวิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี (Step-down Independent Bootstrap min P)

DB2 หมายถึง สถิติทดสอบของวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down

Dependent Bootstrap min P) เมื่อทำการคัดลอกชุดข้อมูลจำนวน 2 ชุด

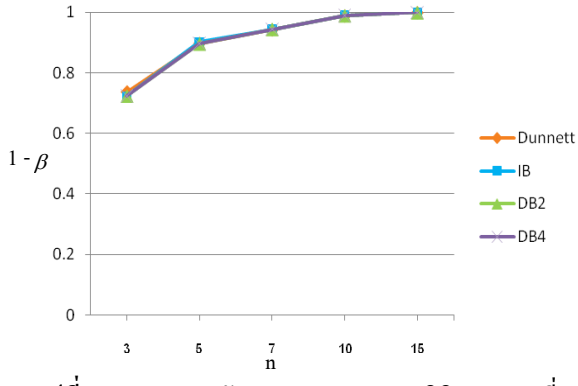
DB4 หมายถึง สถิติทดสอบของวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) เมื่อทำการคัดลอกชุดข้อมูลจำนวน 4 ชุด

ตารางที่ 4 แสดงค่าความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 และกำลังการทดสอบของสถิติทดสอบที่ระดับนัยสำคัญเท่ากับ 0.05

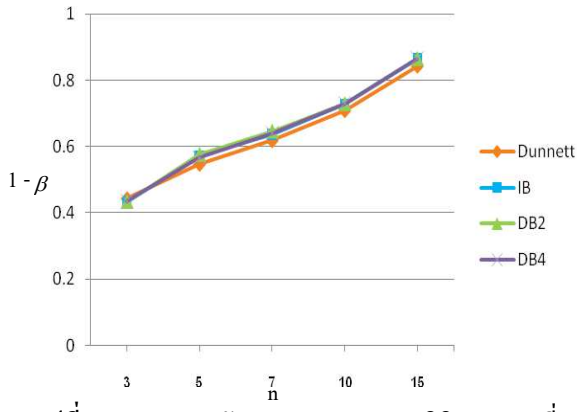
σ^2	n	ความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1				กำลังการทดสอบ			
		Dunnett	IB	DB2	DB4	Dunnett	IB	DB2	DB4
0.5	3	0.045	0.048	0.047	0.047	0.736 **	0.722	0.724	0.724
	5	0.035	0.039	0.039	0.039	0.897	0.901 **	0.896	0.896
	7	0.031	0.032	0.031	0.035	0.942	0.943	0.944 **	0.944 **
	10	0.032	0.035	0.040	0.037	0.988	0.990 **	0.989	0.988
	15	0.025	0.027	0.030	0.028	0.999	0.999	0.999	0.999
1	3	0.044	0.049	0.047	0.049	0.444 **	0.433	0.433	0.435
	5	0.025	0.031	0.028	0.029	0.547	0.572	0.576 **	0.566
	7	0.019	0.022	0.020	0.023	0.618	0.637	0.646 **	0.640
	10	0.011	0.017	0.017	0.015	0.708	0.728	0.729	0.730 **
	15	0.009	0.014	0.014	0.014	0.842	0.865	0.866 **	0.866 **
1.5	3	0.046	0.048	0.051	0.05	0.316 **	0.316 **	0.311	0.304
	5	0.019	0.029	0.03	0.03	0.316	0.329	0.326	0.330 **
	7	0.013	0.016	0.015	0.016	0.367	0.387	0.393 **	0.387
	10	0.014	0.015	0.015	0.016	0.399	0.417	0.421	0.429 **
	15	0.003	0.007	0.008	0.008	0.498	0.530	0.530	0.533 **
2	3	0.040	0.047	0.049	0.049	0.243	0.244	0.241	0.249 **
	5	0.018	0.025	0.023	0.026	0.198	0.224 **	0.217	0.215
	7	0.013	0.017	0.017	0.016	0.194	0.218	0.224 **	0.220
	10	0.002	0.005	0.005	0.003	0.231	0.259	0.264 **	0.261
	15	0.003	0.002	0.002	0.003	0.265	0.294	0.304 **	0.294

* หมายถึง สถิติทดสอบไม่สามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 ได้

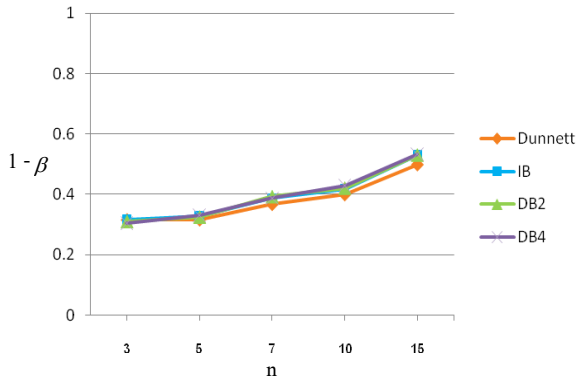
** หมายถึง สถิติทดสอบมีกำลังการทดสอบสูงสุด (พิจารณาภายใต้ความแปรปรวนและขนาดตัวอย่างเดียวกัน)



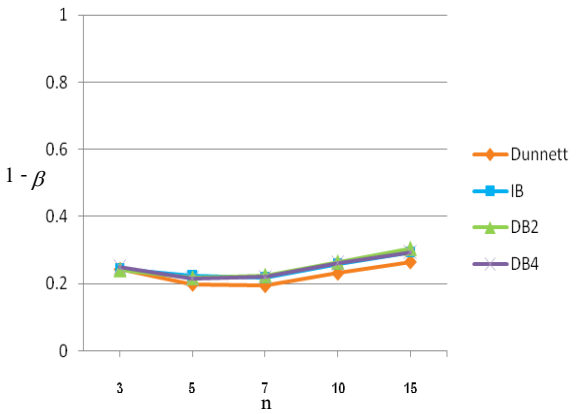
รูปที่ 1 แสดงกำลังการทดสอบของสถิติทดสอบ ที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากับ 0.5



รูปที่ 2 แสดงกำลังการทดสอบของสถิติทดสอบ ที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากับ 1



รูปที่ 3 แสดงกำลังการทดสอบของสถิติทดสอบ ที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากับ 1.5



รูปที่ 4 แสดงกำลังการทดสอบของสถิติทดสอบ ที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากับ 2

จากตารางที่ 1 พบว่า สถิติทดสอบทุกตัวสามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 ได้สำหรับทุกระดับความแปรปรวนที่ศึกษาและในทุกขนาดตัวอย่างที่ทำการศึกษา เมื่อพิจารณาที่กำลังการทดสอบกรณีที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากับ 0.5 พบว่าสถิติทดสอบทุกตัวมีกำลังการทดสอบที่ใกล้เคียงกัน และเมื่อพิจารณาที่ขนาดตัวอย่างเท่ากับ 15 พบว่าสถิติทดสอบทุกตัวมีกำลังการทดสอบสูง โดยไม่สามารถจำแนกได้ว่าสถิติทดสอบตัวใดมีกำลังการทดสอบสูงสุด สังเกตได้จากรูปที่ 1

เมื่อพิจารณากำลังการทดสอบในกรณีที่ความแปรปรวนในแต่ละทรีทเมนต์มากขึ้น พบว่าโดยภาพรวมแล้ววิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีกำลังการทดสอบสูงกว่าสถิติทดสอบของดันเนตต์ และค่อนข้างสูงกว่าอย่างชัดเจน ที่ความแปรปรวนเพิ่มขึ้น โดยสังเกตได้จากรูปที่ 2 - รูปที่ 4

5. สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

1. ภายใต้ประชากรที่มีการแจกแจงเป็นแบบล็อกนอร์มัล โดยที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากันและจำนวนซ้ำในแต่ละทรีตเมนต์เท่ากัน

- ที่ระดับความแปรปรวนเท่ากับ 0.5 และ 1 วิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีประสิทธิภาพใกล้เคียงกับสถิติทดสอบของดันเนตต์

- ที่ระดับความแปรปรวน1.5 และ 2 วิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีประสิทธิภาพสูงกว่ากับสถิติทดสอบของดันเนตต์เล็กน้อย

2. วิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี นั้นมีความแกร่ง สามารถใช้ได้กับลักษณะของข้อมูลที่บ่ายเบนไปจากการแจกแจงแบบปกติ ในที่นี้คือการแจกแจงแบบ ล็อกนอร์มัล โดยที่ความแปรปรวนในแต่ละทรีตเมนต์เท่ากันและจำนวนซ้ำในแต่ละทรีตเมนต์เท่ากัน

3. กำลังการทดสอบของสถิติทดสอบทุกตัวจะแปรผันตามขนาดตัวอย่าง กล่าวคือ เมื่อขนาดตัวอย่างมากขึ้น กำลังการทดสอบของตัวสถิติทดสอบจะสูงขึ้นและกำลังการทดสอบของสถิติทดสอบทุกตัวจะแปรผกผันกับระดับความแปรปรวน กล่าวคือ เมื่อความแปรปรวนมากขึ้น สถิติทดสอบจะมีกำลังการทดสอบลดลง

5.2 ข้อเสนอแนะ

1. ในการจำลองข้อมูลไม่ควรเลือกขนาดตัวอย่างมากเกินไป เนื่องจากจะทำให้กำลังการทดสอบมีค่าเข้าใกล้ 1 ซึ่งจะไม่มีประโยชน์ในการเปรียบเทียบกำลังการทดสอบ

2. อาจทำการศึกษาการแจกแจงรูปแบบอื่นที่ต่างไปจากการแจกแจงล็อกนอร์มัล

6. เอกสารอ้างอิง

[1] กมล บุษบา, การวิเคราะห์การทดลองเชิงสถิติด้วย SPSS : วิเคราะห์เชิงลึกพร้อมตัวอย่างจากปัญหาจริง, ปทุมธานี, 2551.

[2] Westfall, P. H., & Wolfinger, R. D. (2000). Close Multiple Testing Procedures and PROC MULLTEST. SAS® Observations, from http://ftp.sas.com/techsup/download/observations/obswww23/.

[3] ปิยพงศ์ ขวัญเลิศ, การเปรียบเทียบความแกร่งของวิธีการทดสอบพหุคูณแบบปิดสำหรับค่าเฉลี่ยประชากร. วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, ภาควิชาคณิตศาสตร์และสถิติ, 2551.

[4] Dudoit, S., Shaffer, J. P., & Boldrick, J. C., Multiple Hypothesis Testing in Microarray Experiments, Statistical Science, Vol. 18, No. 1, pp. 71-103, 2003.

[5] Westfall, P.H., Tobias, R.D., Rom, D., Wolfinger, R.D., and Hochberg, Y., *Multiple Comparisons and Multiple Tests Using the SAS System*, Cary, NC: SAS Institute Inc., 1999.

[6] Komonniramit, Y., Step-down Dependent Bootstrap min P for Comparison Several Means with a Control, Unpublished masrer's

thesis, Thammasat University, Faculty of Science and Technology, 2008.

[7] Dunnett, C. W., A Multiple Comparison Procedure for Comparing Several Treatments with a Control, Journal of the American Statistical Association, Vol. 50, No. 272, pp. 1096-1121, 1955.