บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในปัจจุบัน วิธีการทางสถิติได้เข้ามามีบทบาทอย่างยิ่งต่อการวิจัยและการพัฒนาในงานทุกสาขา โดยข้อมูลที่นำมาวิเคราะห์ได้มาจากการทดลองหรือการสำรวจ ซึ่งก่อนทำการทดลองหรือสำรวจนั้นจะต้องมีการออกแบบและวางแผนเพื่อทำให้การเก็บรวบรวมข้อมูลนั้นมีประสิทธิภาพและนำไปวิเคราะห์ได้สะดวกและถูกต้องเป็นไปตามแผนการทดลองหรือแผนการสำรวจที่ได้วางไว้ นอกจากนี้ยังสามารถตอบปัญหาที่สนใจศึกษาได้อย่างถูกต้อง การวิเคราะห์ความแปรปรวนนั้นเป็นอีกวิธีการหนึ่งทางสถิติที่นิยมใช้กันแพร่หลาย เพื่อใช้ในการศึกษาเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์

ผลสรุปที่ได้จากการวิเคราะห์ความแปรปรวนนั้นจะบอกผู้วิจัยได้เพียงแค่ค่าเฉลี่ยของทรีตเมนต์นั้นมีค่าเท่ากันหรือไม่เท่ากัน แต่ไม่สามารถบอกได้ว่า มีทรีตเมนต์ใดบ้างที่มีค่าเฉลี่ยแตกต่างกันหรือไม่สามารถบอกได้ว่าทรีตเมนต์เหล่านั้นแตกต่างกันในลักษณะใด โดยทั่วไปแล้ววิธีการศึกษาหลังจากการวิเคราะห์ความแปรปรวนจะพิจารณาว่าอิทธิพลของทรีตเมนต์เป็นค่าคงที่หรือตัวแปรสุ่ม ถ้าอิทธิพลของทรีตเมนต์เป็นตัวแปรสุ่มจะทำการวิเคราะห์ส่วนประกอบของความแปรปรวน แต่ถ้าอิทธิพลของทรีตเมนต์เป็นค่าคงที่และทรีตเมนต์หรือระดับของปัจจัยนั้นเป็นข้อมูลเชิงคุณภาพหรือเชิงปริมาณ ถ้าทรีตเมนต์เป็นข้อมูลเชิงปริมาณจะศึกษารูปแบบความสัมพันธ์ระหว่างทรีตเมนต์กับตัวแปรตามหรือตัวแปรตอบสนองโดยใช้วิธีการเดียวกับการวิเคราะห์การถดถอย รวมทั้งเขียนกราฟของความสัมพันธ์ด้วยและถ้าอิทธิพลของทรีตเมนต์เป็นค่าคงที่ และทรีตเมนต์เป็นข้อมูลเชิงคุณภาพจะทำการเปรียบเทียบทรีตเมนต์หรือทดสอบค่าเฉลี่ยของทรีตเมนต์

เมื่อผู้วิจัยต้องการทำการทดสอบสมมติฐานหลายสมมติฐานที่มีความเกี่ยวข้องกันหรือทำการประมาณค่าพารามิเตอร์แบบช่วงหลายๆ ช่วงที่มีความเกี่ยวข้องกัน โดยต้องการทดสอบพร้อมๆ กัน จึงมีความจำเป็นที่ต้องใช้วิธีที่เรียกว่า “การเปรียบเทียบพหุคูณ (multiple comparison)” ซึ่งเป็นวิธีที่ใช้ในการเปรียบเทียบทรีตเมนต์ที่ไม่ได้ระบุว่าต้องการเปรียบเทียบทรีตเมนต์หรือค่าเฉลี่ยของทรีตเมนต์คู่ใดจนกว่าจะได้ทำการวิเคราะห์ความแปรปรวน โดยวิธีดังกล่าวนี้จะต้องพิจารณาเรื่องของ “อัตราความผิดพลาด (error rate)” ที่จะเกิดขึ้นด้วย

ในบางสถานการณ์ ผู้วิจัยต้องการทำการเปรียบเทียบระหว่างทรีตเมนต์อื่นๆ กับทรีตเมนต์ควบคุมหรือทรีตเมนต์อื่นๆ กับทรีตเมนต์ที่ให้ผลดีที่สุด ซึ่งทรีตเมนต์ที่ดีที่สุดนั้นอาจเป็นทรีตเมนต์ที่ให้ค่าเฉลี่ยสูงสุดหรือต่ำสุดก็ได้ ในการทดลองบางประเภทอาจมีทรีตเมนต์หนึ่งที่ใช้เป็นทรีตเมนต์ควบคุม และสำหรับบางทรีตเมนต์หรือทรีตเมนต์อื่นๆ ที่เหลือทั้งหมดเป็นทรีตเมนต์ในการทดลองที่ต้องการปรียบเทียบกับทรีตเมนต์ควบคุม นั่นคือการเปรียบเทียบจะเป็นการเปรียบเทียบความแตกต่างระหว่างทรีตเมนต์อื่นๆ กับทรีตเมนต์ที่ใช้เป็นทรีตเมนต์ควบคุม วิธีการที่ใช้ในกรณีนี้ เช่น วิธีการของดันเนตต์ (Dunnett) ซึ่งวิธีการดังกล่าวนี้จะควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบชนิดเข้ม (Strong Familywise Error Rate) (กมล บุษบา, 2551)

ในการทดสอบสมมติฐานแต่ละการทดสอบนั้นจะมีอัตราการเกิดความผิดพลาดประเภทที่ 1 (Type I error) ของการทดสอบ ผู้วิจัยสามารถควบคุมอัตราการเกิดความผิดพลาดเหล่านี้ได้ โดยจุดประสงค์ของวิธีการเปรียบเทียบพหุคูณจึงใช้สำหรับควบคุม ค่าที่มากที่สุดของอัตราการเกิดความผิดพลาดประเภทที่ 1 (Maximum overall Type I error rate) หมายถึง ค่าที่มากที่สุดของความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างอย่างน้อยหนึ่งสมมติฐาน เมื่อสมมติฐานว่างนั้นเป็นจริง โดยเรียกอัตราความผิดพลาดดังกล่าวว่า “อัตราความผิดพลาดสูงสุดต่อการทดลอง (Maximum Experimentwise Error Rate: MEER)” หรือ “อัตราความผิดพลาดสูงสุดต่อวงศ์การทดสอบ (Maximum Familywise Error Rate: FWE)” (Westfall and Wolfinger, 2000)

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 โดยไม่คำนึงถึงว่าจริงๆ แล้วมีสมมติฐานใดบ้างในเซตของการทดสอบที่เป็นจริง การเปรียบเทียบพหุคูณนั้นจะควบคุม FWE ได้อย่างเข้มงวด สำหรับการทดสอบแบบขั้นบันไดที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ วิธี ความแตกต่างที่มีนัยสำคัญน้อยที่สุด (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 (ปิยพงศ์ ขวัญเลิศ, 2550)

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 เช่น Step-up Hochberg adjusted p-values และ Rom’s method ส่วนวิธีที่ใช้ในวิธีสเต็ปดาวน์ เช่น Hokm’s procedure, step-down Šidák adjusted p-values, Simes modified Bonferroni, Westfall and Young step-down min P adjusted p-values และ step-down max T adjusted p-values (Dudoit et al, 2003)

ในหลายๆ สถานการณ์ที่เราไม่ทราบการแจกแจงร่วมของสถิติทดสอบ เราจะอาศัยวิธีการสุมซ้ำ เช่น วิธีการสุมซ้ำด้วยวิธีการเปลี่ยนลำดับ (permutation resampling), วิธีการสุมซ้ำด้วยวิธีการบูทสแตรป (bootstrap resampling) สามารถใช้ในการประมาณค่าพีที่ไม่ได้ปรับค่า (unadjusted p-values) และค่าพีที่ปรับค่า (adjusted p-values) โดยไม่ต้องคำนึงถึงข้อตกลงเกี่ยวกับค่าพารามิเตอร์เกี่ยวกับแจกแจงร่วมของสถิติทดสอบ

คำสั่ง PROC MULTTEST ใน SAS® ถูกเขียนขึ้นเพื่อจุดประสงค์ในการทดสอบพหุคูณและการเปรียบเทียบพหุคูณโดยตรง ซึ่งเสนอวิธีการปรับค่าพี โดยใช้ บอนเฟอร์โรนี , Šidák, การสุ่มซ้ำด้วยวิธีการบูทสแตรป และการสุ่มซ้ำด้วยวิธีการเปลี่ยนลำดับ ซึ่งการปรับค่าตามวิธีดังกล่าวใช้กับวิธีการทดสอบแบบขั้นตอนเดียว และวิธีสเต็ปดาวน์ ซึ่งในคำสั่ง PROC MULTTEST นี้ถูกออกแบบมาให้สามารถปรับใช้ได้ในกรณีที่ข้อมูลไม่เป็นไปตามข้อตกลง โดยใช้การสุ่มซ้ำด้วยวิธีการบูทสแตรป และการสุ่มซ้ำด้วยวิธีการเปลี่ยนลำดับ (Westfall and Tobias, 1999; SAS® Institude Inc. 2002)

ในปี 2551 ยุทธสิทธิ์ (Yutasit) ได้ศึกษาวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) สำหรับการเปรียบเทียบค่าเฉลี่ยกรณีที่มีทรีตเมนต์ควบคุม เปรียบเทียบกับวิธีการของดันเนตต์ (Dunnett) โดยศึกษาภายใต้ข้อมูลที่มีการแจกแจงปรกติและขนาดตัวอย่างเท่ากัน โดยกำหนดให้มี 3 ทรีตเมนต์ และมี 1 ทรีตเมนต์ควบคุม ซึ่งผลการศึกษา พบว่า วิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี มีประสิทธิภาพใกล้เคียงกับสถิติทดสอบ ของดันเนตต์ (Komonniramit, Y., 2008)

ในการวิเคราะห์ทางสถิติหรือการทดสอบต่างๆ ข้อมูลหรือประชากรที่นำมาพิจารณาส่วนใหญ่ต้องเป็นไปตามข้อตกลงเบื้องต้น ในการทดสอบความแตกต่างของค่าเฉลี่ยระหว่างประชากรหลายกลุ่มนั้น มีข้อตกลงเบื้องต้นว่าด้วยประชากรต้องได้มาอย่างสุ่ม มีการแจกแจงปรกติและความแปรปรวนของแต่ละกลุ่มเท่ากันและประชากรแต่ละกลุ่มต้องเป็นอิสระกัน อย่างไรก็ตามถ้าข้อมูลในการทดสอบดังกล่าวไม่เป็นไปตามข้อตกลงเบื้องต้น ผลลัพธ์ที่ได้จะมีผลต่อระดับนัยสำคัญและกำลังการทดสอบอย่างไร ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาความแกร่งของวิธีสเต็ปดาวน์ บูทสแตรป มิน พี ภายใต้ลักษณะข้อมูลที่ไม่เป็นไปตามข้อตกลงเบื้องต้น โดยศึกษาภายใต้ลักษณะข้อมูลที่มีการแจกแจงปรกติ ที่มีความแปรปรวนในแต่ละทรีตเมนต์ไม่เท่ากัน และศึกษาภายใต้ลักษณะข้อมูลที่มีการแจกแจงล็อกนอร์มัล ที่มีความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน ซึ่งเป็นข้อมูลที่พบกันอยู่มากในการศึกษาทดลอง เช่น ข้อมูลปริมาณน้ำฝนในแต่ละเดือน ระยะเวลาที่พนักงานให้บริการลูกค้าต่อคน หรือระยะเวลาที่ลูกค้ารอจนกว่าจะได้รับบริการ เป็นต้น และการแจกแจงล็อกนอร์มัล ยังนำไปประยุกต์ใช้กันมากในทางวิศวกรรม เศรษฐศาสตร์ และการประกันภัย

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความแกร่งของวิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี (Step-down Independent Bootstrap min P) และวิธีสเต็ปดาวน์ดิเพนเดนท์ บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) ที่ใช้สำหรับการเปรียบเทียบค่าเฉลี่ย กรณีที่มีทรีตเมนต์ควบคุม

1.3 สมมติฐานการวิจัย

สมมติฐานในการวิจัยครั้งนี้ มีดังต่อไปนี้

1.3.1 ภายใต้ประชากรที่มีการแจกแจงปรกติ ที่มีความแปรปรวนในแต่ละทรีตเมนต์ไม่เท่ากันและจำนวนซ้ำในแต่ละทรีตเมนต์เท่ากัน ผู้วิจัยคาดว่า วิธีสเต็ปดาวน์อินดิเพนเดนท์ บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี สามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ และมีกำลังการทดสอบสูงกว่าสถิติทดสอบของดันเนตต์

1.3.2 ภายใต้ประชากรที่มีการแจกแจงเป็นแบบล็อกนอร์มัล ที่มีความแปรปรวนในแต่ละทรีตเมนต์เท่ากันและจำนวนซ้ำในแต่ละทรีตเมนต์เท่ากัน ผู้วิจัยคาดว่า วิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี สามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ และมีกำลังการทดสอบสูงกว่าสถิติทดสอบของดันเนตต์

1.4 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดขอบเขตของการวิจัย ดังนี้

1.4.1 ในการศึกษาครั้งนี้กำหนดให้ทรีตเมนต์ที่ทำการศึกษามี 3 ทรีตเมนต์ และมีทรีตเมนต์ควบคุม 1 ทรีตเมนต์ โดยกำหนดให้ทรีตเมนต์ที่ 1 เป็นทรีตเมนต์ควบคุม

1.4.2 ขนาดตัวอย่างในแต่ละทรีตเมนต์ (จำนวนซ้ำ) กำหนดให้แต่ละทรีตเมนต์มีจำนวนซ้ำเท่ากัน โดยกำหนดขนาดตัวอย่างดังนี้

(n_1, n_2, n_3, n_4) = (3, 3, 3, 3)

(n_1, n_2, n_3, n_4) = (5, 5, 5, 5)

(n_1, n_2, n_3, n_4) = (7, 7, 7, 7)

(n_1, n_2, n_3, n_4) = (10, 10, 10, 10)

(n_1, n_2, n_3, n_4) = (15, 15, 15, 15)

1.4.3 กรณีที่ข้อมูลมีการแจกแจงปรกติ และความแปรปรวนในแต่ละทรีตเมนต์ไม่เท่ากัน แบ่งเป็น 2 กรณี คือ

1.4.3.1 กำหนดให้ความแปรปรวนในทรีตเมนต์ควบคุมแตกต่างกับความแปรปรวนในทรีตเมนต์อื่นๆ โดยที่ทรีตเมนต์อื่นๆ มีความแปรปรวนเท่ากัน คือ

สถานการณ์ที่ 1 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับปานกลาง

(σ_1^2 = 25 , σ_2^2 = 50 , σ_3^2 = 50 , σ_4^2 = 50)

สถานการณ์ที่ 2 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับปานกลาง

(σ_1^2 = 50 , σ_2^2 = 25 , σ_3^2 = 25 , σ_4^2 = 25)

สถานการณ์ที่ 3 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับมาก

(σ_1^2 = 25 , σ_2^2 = 75 , σ_3^2 = 75 , σ_4^2 = 75)

สถานการณ์ที่ 4 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับมาก

(σ_1^2 = 75 , σ_2^2 = 25 , σ_3^2 = 25 , σ_4^2 = 25)

สถานการณ์ที่ 5 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับมาก

(σ_1^2 = 25 , σ_2^2 = 100 , σ_3^2 = 100 , σ_4^2 = 100)

สถานการณ์ที่ 6 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับมาก

(σ_1^2 = 100 , σ_2^2 = 25 , σ_3^2 = 25 , σ_4^2 = 25)

- การประมาณค่าความผิดพลาดประเภทที่ 1 กำหนดให้ค่าเฉลี่ยในแต่ละทรีตเมนต์เท่ากัน ซึ่งมีค่าเท่ากับ 100 ($\mu_1 = \mu_2 = \mu_3 = \mu_4$ =100) และความแปรปรวนในแต่ละทรีตเมนต์ตามสถานการณ์ที่ 1 - 6
- การประมาณค่ากำลังการทดสอบ กำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมมีค่าแตกต่างกับทรีตเมนต์อื่นๆ โดยกำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมเท่ากับ 90 และค่าเฉลี่ยในทรีตเมนต์อื่นๆ เท่ากันซึ่งมีค่าเท่ากับ 100 (μ_1 = 90, $\mu_2 = \mu_3 = \mu_4$ =100) และความแปรปรวนในแต่ละทรีตเมนต์ตามสถานการณ์ที่ 1 - 6

1.4.3.2 กำหนดให้ความแปรปรวนแตกต่างกันหมดในทุกๆ ทรีตเมนต์ โดยกำหนดเป็นอัตราส่วน 1:2:3:4 คือ

สถานการณ์ที่ 7 อัตราส่วนของความแปรปรวนแตกต่างกัน ระดับมาก

(σ_1^2 = 25, σ_2^2 = 50, σ_3^2 = 75, σ_4^2 = 100)

- การประมาณค่าความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 กำหนดให้ค่าเฉลี่ยในแต่ละทรีตเมนต์เท่ากัน ซึ่งมีค่าเท่ากับ 100 ($\mu_1 = \mu_2 = \mu_3 = \mu_4$ =100) และความแปรปรวนในแต่ละทรีตเมนต์ตามสถานการณ์ที่ 7
- การประมาณค่ากำลังการทดสอบ กำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมมีค่าแตกต่างกับทรีตเมนต์อื่นๆ โดยกำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมเท่ากับ 90 และค่าเฉลี่ยในทรีตเมนต์อื่นๆ เท่ากันซึ่งมีค่าเท่ากับ 100 (μ_1 = 90, $\mu_2 = \mu_3 = \mu_4$ =100) และความแปรปรวนในแต่ละทรีตเมนต์ตามสถานการณ์ที่ 7

1.4.4 กรณีที่ประชากรมีการแจกแจงล็อกนอร์มอล (lognormal distribution) กำหนดให้ความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน คือ

สถานการณ์ที่ 8 ($\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.5$, $\sigma_3^2 = 0.5$, $\sigma_4^2 = 0.5$)

สถานการณ์ที่ 9 ($\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\sigma_3^2 = 1$, $\sigma_4^2 = 1$)

สถานการณ์ที่ 10 ($\sigma_1^2 = 1.5$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 1.5$, $\sigma_4^2 = 1.5$)

สถานการณ์ที่ 11 ($\sigma_1^2 = 2$, $\sigma_2^2 = 2$, $\sigma_3^2 = 2$, $\sigma_4^2 = 2$)

- การประมาณค่าความผิดพลาดประเภทที่ 1 กำหนดให้ค่าเฉลี่ยในแต่ละทรีตเมนต์เท่ากัน ซึ่งมีค่าเท่ากับ 0 ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$) และความแปรปรวนในแต่ละทรีตเมนต์เท่ากัน ตามสถานการณ์ที่ 8 - 11
- การประมาณค่ากำลังการทดสอบ กำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมมีค่าแตกต่างกับทรีตเมนต์อื่นๆ โดยกำหนดให้ค่าเฉลี่ยในทรีตเมนต์ควบคุมเท่ากับ 2 และค่าเฉลี่ยในทรีตเมนต์อื่นๆ เท่ากันซึ่งมีค่าเท่ากับ 0 ($\mu_1 = 2$, $\mu_2 = \mu_3 = \mu_4 = 0$) และความแปรปรวนในแต่ละทรีตเมนต์ตามสถานการณ์ที่ 8 - 11

1.4.5 กำหนดระดับนัยสำคัญ 0.05

1.4.6 ในการศึกษาวิธีสเต็ปดาวน์อินดิเพนเดนท์บูทสแตรป มิน พี (Step-down Independent Bootstrap min P) จะกำหนดจำนวนรอบของการสุ่มซ้ำแบบบูทสแตรปที่เป็นอิสระต่อกัน (Independent Bootstrap Resampling) เท่ากับ 1,000 รอบ และวิธีสเต็ปดาวน์ดิเพนเดนท์บูทสแตรป มิน พี (Step-down Dependent Bootstrap min P) จะกำหนดจำนวนการคัดลอก (copy) ชุดข้อมูลตัวอย่างสุ่มสำหรับวิธีการสุ่มซ้ำแบบบูทสแตรปที่ไม่เป็นอิสระต่อกัน (Dependent Bootstrap Resampling) เท่ากับ 2 และ 4 ชุดตามลำดับ

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

1.5 เกณฑ์ที่ใช้ในการเปรียบเทียบ

ในการเปรียบเทียบสถิติทดสอบทั้ง 3 วิธี ใช้กำลังการทดสอบ (Power of The Test) ของสถิติทดสอบเป็นเกณฑ์ในการเปรียบเทียบ โดยที่สถิติทดสอบนั้นต้องสามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ได้ ก่อนที่จะนำไปเปรียบเทียบกำลังการทดสอบ เกณฑ์ที่ใช้ในการพิจารณาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 จากการทดลอง ของการทดสอบที่ระดับนัยสำคัญต่างๆ คือ เกณฑ์ที่ใช้ในการทดสอบสมมติฐานว่า ค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ที่ได้จากการทดลองจะเท่ากับค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ที่กำหนดไว้หรือไม่ โดยใช้การทดสอบทวินาม ที่ระดับนัยสำคัญ 0.05 ถ้าค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 จากผลการทดลองมีค่าน้อยกว่าหรือเท่ากับ 0.061 จะสรุปว่า สถิติทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ได้ แต่ถ้าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 จากผลการทดลองมีค่ามากกว่ากว่า 0.061 จะสรุปว่า สถิติทดสอบนั้นไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ไว้ได้

เมื่อสถิติทดสอบสามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ได้แล้วนั้น จึงจะเปรียบเทียบกำลังการทดสอบของสถิติทดสอบ เพื่อเลือกสถิติทดสอบที่มีกำลังการทดสอบสูงสุด โดยพิจารณาว่าสถิติทดสอบใดมีค่ากำลังการทดสอบสูงสุด จะเรียกว่า สถิติทดสอบนั้นมีประสิทธิภาพสูงสุด ในสถานการณ์นั้นๆ

1.6 ประโยชน์ที่ได้จากการวิจัย

เป็นแนวทางในการเลือกใช้สถิติทดสอบสำหรับการเปรียบเทียบค่าเฉลี่ยกรณีที่มีทรีตเมนต์ควบคุม ที่เหมาะสมกับสถานการณ์ต่างๆ

1.7 คำจำกัดความ

ในการวิจัยครั้งนี้มีการจำกัดความหมายและอธิบายคำศัพท์ไว้ดังนี้

1.7.1 ความผิดพลาดประเภทที่ 1 (Type I error) หมายถึง ความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นจริง และความน่าจะเป็นของความผิดพลาดประเภทนี้คือระดับนัยสำคัญ (Significance level) ของการทดสอบ แทนด้วย α

1.7.2 ความผิดพลาดประเภทที่ 2 (Type II error) หมายถึง ความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นเท็จ และความน่าจะเป็นของความผิดพลาดประเภทนี้ แทนด้วย β

1.7.3 กำลังการทดสอบ (Power of the test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นเท็จ ซึ่งมีค่าเท่ากับ $1-\beta$ เมื่อ β คือ ความน่าจะเป็นของความผิดพลาดประเภทที่ 2

1.7.4 ความแกร่ง (Robustness) หมายถึง คุณสมบัติของการทดสอบที่ไม่แสดงความไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น กรณีที่ลักษณะของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของการทดสอบ ซึ่งจะมีผลต่อความน่าจะเป็นของความผิดพลาดประเภทที่ 1 (α) และความน่าจะเป็นของความผิดพลาดประเภทที่ 2 (β)

1.7.5 ทรีตเมนต์ (Treatment) หมายถึง กลุ่ม พวก หรือระดับต่างๆ ของปัจจัยที่นำมาเปรียบเทียบ

1.7.6 อัตราการค้นพบที่ไม่ถูกต้อง (False Discovery Rate : FDR) คำว่า “การค้นพบ” นั้นในทางสถิติหมายถึง การปฏิเสธสมมติฐานว่างหนึ่งๆ ดังนั้น สัดส่วนของการค้นพบที่ไม่ถูกต้องจึงหมายถึง จำนวนของการค้นพบที่ไม่ถูกต้อง (การเกิดความผิดพลาดประเภทที่ 1) หารด้วยจำนวนการค้นพบทั้งหมด ทำให้นิยามอัตราการค้นพบที่ไม่ถูกต้องได้ว่าเป็นค่าคาดหมายของสัดส่วนการค้นพบที่ไม่ถูกต้อง ในกรณีที่สมมติฐานว่างเป็นจริงนั่นหมายความว่าการค้นพบทั้งหมดนั้นไม่เป็นจริง และทำให้ได้ว่า FDR ก็คือ ความผิดพลาดชนิดต่อการทดลอง การควบคุมค่า FDR ให้มีค่าเป็น α_D นั้นจะเป็นการควบคุมความผิดพลาดชนิดต่อการทดลองด้วย อย่างไรก็ตามการควบคุม FDR นั้นเป็นการควบคุมค่าคาดหมายของสัดส่วนของการปฏิเสธสมมติฐานที่เป็นความผิดพลาดประเภทที่ 1 เมื่อสมมติฐาน H_{0i} บางสมมติฐานเป็นจริง และบางสมมติฐานไม่เป็นจริง ซึ่งเป็นการควบคุมในลักษณะที่ความผิดพลาดชนิดต่อการทดลองไม่สามารถทำได้ วิธีการทดสอบที่ควบคุม FDR จะเป็นวิธีที่มีความอนุรักษ์ (conservative) น้อยกว่า แต่มี

กำลังการทดสอบมากกว่าวิธีที่ควบคุมอัตราความผิดพลาดชนิดต่อการทดลอง (http://en.wikipedia.org/wiki/False_discovery_rate) ถ้ากำหนดให้ FDR เท่ากับ 0.05 หมายความว่า ถ้ามีการปฏิเสธสมมติฐาน 20 ครั้ง จะมีเพียง 1 ครั้งที่เป็นการปฏิเสธสมมติฐานที่ไม่ถูกต้อง และ 19 ครั้งที่เป็นการปฏิเสธสมมติฐานที่ถูกต้อง (เนื่องจาก 1/20 = 0.05) (กมล บุษบา, 2551)

1.7.7 อัตราความผิดพลาดต่อวงศ์การทดสอบชนิดเข้ม (Strong Familywise Error Rate) คือ ความน่าจะเป็นของการเกิดการค้นพบที่ไม่ถูกต้อง หรือคือ ความน่าจะเป็นที่สัดส่วนของการค้นพบที่ไม่ถูกต้องมีค่ามากกว่าศูนย์ การควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบชนิดเข้มให้มีค่า α_s หมายถึงการควบคุมความน่าจะเป็นของการเกิดการปฏิเสธสมมติฐานที่ไม่ถูกต้องใดๆ มีค่าน้อยกว่าหรือเท่ากับ α_s โดยไม่สนใจว่าการปฏิเสธที่ถูกต้องจะมีจำนวนเท่าใด และการควบคุมอัตราความผิดพลาดต่อวงศ์การทดสอบชนิดเข้มให้มีค่าเป็น α_s นั้นจะเป็นการควบคุม FDR ให้มีค่าไม่เกิน α_s ด้วย (กมล บุษบา, 2551)