การเปรียบเทียบตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่ม ในตัวแบบความถดถอยเชิงเส้น

นางสาวณัฐฐิกาญจน์ วิรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

A COMPARISON OF BOOTSTRAPPED ANDERSON-DARLING'S STATISTIC AND BOOTSTRAPPED SHAPIRO-WILK'S STATISTIC FOR NON-NORMALITY OF RANDOM ERROR IN LINEAR REGRESSION MODEL

Miss Nattigan Wirat

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2007

หัวข้อวิทยานิพนธ์	**การเปรียบเทียบตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับ ตัวสถิติบูทสแตรปแชปิโร-วิลค์สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้น**
โดย	นางสาวณัฐฐิกาญจน์ วิรัตน์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ผกาวดี ศิริรังษี)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์

(รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา)

.... กรรมการ

(อาจารย์ ดร.อรุณี กำลัง)

ณัฐฐิกาญจน์ วิรัตน์ : การเปรียบเทียบตัวสถิติบูทสแตรปแอนเดอร์สัน-คาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้น (A COMPARISON OF BOOTSTRAPPED ANDERSON-DARLING'S STATISTIC AND BOOTSTRAPPED SHAPIRO-WILK'S STATISTIC FOR NON-NORMALITY OF RANDOM ERROR IN LINEAR REGRESSION MODEL.) อ. ที่ปรึกษา : รศ. ดร. สุพล ดุรงค์วัฒนา, 108 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบตัวสถิติที่ใช้ในการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้น ในที่นี้ทำการศึกษาตัวสถิติ คือ ตัวสถิติบูทสแตรปแอนเดอร์สัน-คาร์ลิง กับ ตัวสถิติบูทสแตรปแชปิโร-วิลค์ ซึ่งเกณฑ์ที่ใช้ในการเปรียบเทียบตัวสถิติ คือ ค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ ภายใต้สถานการณ์ต่างๆ ที่ศึกษาดังนี้

1. จำนวนตัวแปรอิสระเท่ากับ 3 และ 5
2. ขนาดตัวอย่างเท่ากับ 10 20 30 และ 50 ตามลำดับ
3. การแจกแจงของความคลาดเคลื่อนที่ศึกษามี 3 การแจกแจง คือ การแจกแจงปกติ การแจกแจงโลจิสติคและการแจกแจงดับเบิ้ลเอ็กซ์โปเนนเชียล โดยกำหนดให้ทุกการแจกแจงมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 5 และ 10 ทำการศึกษาที่ระดับนัยสำคัญ 0.01 0.05 และ 0.10

ในการวิจัยครั้งนี้ ได้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลซึ่งกระทำซ้ำ 600 ครั้งในแต่ละสถานการณ์ ด้วยโปรแกรม S-PLUS 2000 ผลการวิจัยสรุปได้ดังนี้

1.ค่าความคลาดเคลื่อนประเภทที่ 1

ผลการทดสอบพบว่าตัวสถิติบูทสแตรปแอนเดอร์สัน – คาร์ลิง และตัวสถิติบูทสแตรปแชปิโร-วิลค์ สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทุกสถานการณ์เมื่อขนาดตัวอย่างมีค่าเป็น 30 และ 50 แต่ในกรณีที่ขนาดตัวอย่างมีค่าเป็น 10 และ 20 ตัวสถิติทั้งสองไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ในบางสถานการณ์

2.อำนาจการทดสอบ

ตัวสถิติบูทสแตรปแชปิโร - วิลค์ จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแอนเดอร์สัน - คาร์ลิงเมื่อขนาดตัวอย่างมีค่าเท่ากับ 10 และ 20 แต่ในกรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 30 และ 50 ตัวสถิติบูทสแตรปแอนเดอร์สัน - คาร์ลิง จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแชปิโร - วิลค์ ทุกการแจกแจง ทุกระดับของจำนวนตัวแปรอิสระ ทุกระดับของขนาดตัวอย่าง ทุกระดับของส่วนเบี่ยงเบนมาตรฐานและทุกระดับนัยสำคัญที่ทำการทดสอบ

โดยทั่วไปอำนาจการทดสอบจะแปรผันตามปัจจัยดังต่อไปนี้ จำนวนตัวแปรอิสระ ขนาดตัวอย่าง และระดับนัยสำคัญ แต่จะแปรผกผันกับส่วนเบี่ยงเบนมาตรฐาน

ภาควิชา ....สถิติ........	ลายมือชื่อนิสิต ....ณัฐฐิกาญจน์ วิรัตน์....
สาขาวิชา ....สถิติ........	ลายมือชื่ออาจารย์ที่ปรึกษา ....
ปีการศึกษา ....2550........	

#4882182926: MAJOR STATISTICS

KEYWORD : BOOTSTRAPPED ANDERSON-DARLING'S STATISTIC / BOOTSTRAPPED SHAPIRO-WILK'S STATISTIC / LINEAR REGRESSION MODEL

NATTIGAN WIRAT: A COMPARISON OF BOOTSTRAPPED ANDERSON – DARLING 'S STATISTIC AND BOOTSTRAPPED SHAPIRO-WILK'S STATISTIC FOR NON-NORMALITY OF RANDOM ERROR IN LINEAR REGRESSION MODEL. THESIS ADVISOR : ASSOC. PROF. SUPOL DURONGWATANA, Ph.D., 108 pp.

The objective of this research is to study and compare statistic for non-normality of random error in linear regression model. Bootstrapped Anderson-Darling's Statistic and Bootstrapped Shapiro-Wilk's Statistic are considered in this study. The probability of type I error and the power of the test are two criteria using for comparison. The comparison are done under several situations which are as follows:

1. The number of independent variables are 3 and 5

2. The sample sizes are 10 20 30 and 50

3. This study used three different distributions of errors. They are normal distribution, logistic distribution and double exponential distribution. For all of the distributions, the means of 0, the standard deviations of 1, 5 and 10 are used. The significance levels for this study are at 0.01 0.05 and 0.10 level.

The data for this research is simulated by using the Monte-Carlo simulation technique with 600 repetitions for each situation by S-PLUS 2000 program. The results of this research can be summarized as follows:

1.Probability of type I error

Bootstrapped Anderson-Darling's Statistic and Bootstrapped Shapiro-Wilk's Statistic can control probability of type I error at all of the cases when the sample sizes are 30 and 50 but when the sample sizes are 10 and 20 the two statistics can control probability of type I error at some of the cases.

2.Power of the test

Bootstrapped Shapiro – Wilk 's Statistic gives the higher power of the test than Bootstrapped Anderson – Darling's Statistic when the sample sizes are 10 and 20 but when the sample sizes are 30 and 50 Bootstrapped Anderson-Darling' s Statistic gives the higher power of the test than Bootstrapped Shapiro – Wilk 's Statistic all of the distributions, the number of independent variables, sample sizes, standard deviations and significance levels.

In general, power of the test varies with, the number of independent variables, sample sizes and significance levels but converse to standard deviations.

Department :	Statistics	Student's Signature	Nattigan Wirat
Field of Study :	Statistics	Advisor's Signature	
Academic Year :	2007		

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี จากความอนุเคราะห์ของบุคคลหลายฝ่ายด้วยกัน ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำปรึกษา แนะนำ ตลอดจนดูแลแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด รองศาสตราจารย์ ผกาวดี ศิริรังษี ในฐานะประธานกรรมการสอบวิทยานิพนธ์และอาจารย์ ดร.อรุณี กำลัง ในฐานะกรรมการสอบวิทยานิพนธ์ ที่ได้ช่วยกรุณาตรวจและแก้ไขให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ถูกต้องยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษาและประสิทธิ์ประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา นายบำรุงศักดิ์ เผื่อนอารีย์ นิสิตปริญญาโทภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำปรึกษาและความช่วยเหลือเกี่ยวกับการเขียนโปรแกรม และการใช้โปรแกรม S-PLUS 2000

ท้ายนี้ผู้วิจัยขอกราบขอบพระคุณบิดา มารดา ซึ่งสนับสนุนด้านการศึกษาและให้กำลังใจแก่ผู้วิจัยเสมอมาจนกระทั่งสำเร็จการศึกษา และขอบคุณรุ่นพี่ รุ่นน้อง รวมถึงเพื่อนๆทุกคนที่เป็นกำลังใจและให้ความช่วยเหลือด้วยดีเสมอมา จึงขอกราบขอบพระคุณมา ณ ที่นี้

สารบัญ

หน้า

สารบัญตาราง

ตารางที่ หน้า

ตารางที่ หน้า

สารบัญรูป

รูปที่ หน้า

รูปที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

สำหรับการวิจัยเพื่อหาผลสรุปโดยทั่ว ๆ ไปนอกจากผู้วิจัยจะต้องมีความรู้ความเข้าใจในเรื่องที่ศึกษาเป็นอย่างดีแล้วจำเป็นต้องอาศัยความรู้ทางด้านสถิติ ปัจจุบันข้อมูลข่าวสารมีความสำคัญและจำเป็นต่อการดำเนินงานในหลาย ๆ ด้านการตัดสินใจหรือวางแผนในกิจกรรมต่างๆ จะต้องใช้ข้อมูลประกอบ เช่น ทราบถึงข้อจำกัดและความเหมาะสมของแต่ละวิธีที่จะนำมาใช้กับข้อมูล ตลอดจนความสอดคล้องระหว่างสถิติที่ใช้กับเป้าหมายของการวิเคราะห์ โดยเฉพาะอย่างยิ่งในกระบวนการวิเคราะห์ข้อมูล เพื่อที่จะให้ได้ผลสรุปที่ถูกต้องและสมเหตุสมผล ซึ่งเมื่อพิจารณาในแง่ของการวิเคราะห์และสรุปผล อาจแบ่งวิธีการทางสถิติได้เป็น 2 ประเภทตามลักษณะของข้อมูลคือ วิธีการพาราเมตริกซ์ (Parametric Methods) และวิธีการนอนพาราเมตริกซ์ (Nonparametric Methods) ซึ่งในทางปฏิบัติพบว่าวิธีการนอนพาราเมตริกซ์มีใช้กันอย่างกว้างขวางในงานวิจัย ทั้งนี้เพราะสามารถคำนวณได้ง่ายและรวดเร็ว อีกทั้งข้อกำหนดบางอย่างเกี่ยวกับข้อมูลที่ศึกษาก็มีไม่มากนัก และไม่เข้มงวดเท่ากับวิธีการทางพาราเมตริกซ์จึงทำให้สะดวกต่อการนำไปใช้

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติวิธีหนึ่งซึ่งมีผู้นิยมใช้กันมากในงานหลายสาขา เช่น เศรษฐศาสตร์ บริหารธุรกิจ สังคมศาสตร์ เกษตรศาสตร์ วิศวกรรมศาสตร์ เป็นต้น เพื่อประโยชน์ในการอธิบายความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ความเข้าใจเกี่ยวกับเรื่องการวิเคราะห์การถดถอยเป็นพื้นฐานที่สำคัญต่อการวิเคราะห์เชิงปริมาณหลายด้าน สำหรับการวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการทางสถิติที่นำใช้คาดคะเนหรือพยากรณ์ค่าของตัวแปรที่สนใจศึกษาซึ่งเรียกว่า ตัวแปรตาม (Dependent Variable) ใช้สัญลักษณ์แทนด้วย Y โดยที่ค่าของตัวแปรตามต้องอาศัยความสัมพันธ์กับตัวแปรอื่นที่เกี่ยวข้องซึ่งมีอิทธิพลต่อตัวแปรตามที่เรียกว่า ตัวแปรอิสระ (Independent Variable) ใช้สัญลักษณ์แทนด้วย X เมื่อพิจาณาความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัวกับตัวแปรอิสระ 1 ตัว เรียกวิธีวิเคราะห์นี้ว่าการวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression Analysis) ในสภาพความเป็นจริงนั้นตัวแปรอิสระที่ศึกษาอาจจะมีมากกว่า 1 ตัวแปรก็เป็นไปได้ เราเรียกการวิเคราะห์ในกรณีนี้ว่า การวิเคราะห์ความถดถอยเชิงพหุ (Multiple Regression Analysis) ทั้งนี้เพราะตัวแบบความถดถอยอย่างง่ายจากตัวแปรอิสระเพียงตัวเดียวอาจไม่เหมาะสมในทางทฤษฎีและในทางปฏิบัติ และถ้าตัวแปรอิสระแต่ละตัวกับตัวแปรตามมีลักษณะ

ความสัมพันธ์ในรูปแบบเชิงเส้นจะเรียกการวิเคราะห์ความถดถอยดังกล่าวว่า การวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ดังนั้นเพื่อการวิเคราะห์ที่อยู่ในรูปแบบที่ใกล้เคียงกับความเป็นจริงผู้วิจัยจึงใช้ตัวแบบความถดถอยเชิงเส้นพหุคูณสำหรับการวิจัยในครั้งนี้ ซึ่งสามารถแสดงในรูปของสมการได้ดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_p x_{ip} + \varepsilon_i \quad ; \ \mathrm{i} = 1,2,3,\ldots,\mathrm{n}$$

หรือเขียนในรูปเมตริกซ์ได้เป็น

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของค่าพารามิเตอร์ที่ไม่ทราบค่าขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

n เป็นขนาดของข้อมูลตัวอย่าง

p เป็นจำนวนตัวแปรอิสระที่ใช้ในสมการถดถอยเชิงเส้น

โดยมีข้อตกลงเบื้องต้นดังนี้

1. กำหนดให้ตัวแปรสุ่มความคลาดเคลื่อน $\varepsilon_i ; i = 1,2,\ldots,n$ มีการแจกแจงความน่าจะเป็นแบบปกติ นั่นคือ $\varepsilon_i \sim$ Normal $; i = 1,2,\ldots,n$ เรียกคุณสมบัตินี้ว่าการมีการแจกแจงแบบปกติ (Normality)

2. กำหนดให้ค่าเฉลี่ยของตัวแปรสุ่มความคลาดเคลื่อน $\varepsilon_i ; i = 1,2,\ldots,n$ มีค่าเป็นศูนย์ กล่าวคือ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$

3. กำหนดให้ความแปรปรวนของตัวแปรสุ่มความคลาดเคลื่อน $\varepsilon_i ; i = 1,2,\ldots,n$ $Var(\varepsilon_i) = \sigma^2$ เรียกคุณสมบัติดังกล่าวนี้ว่า ความคงที่ของความแปรปรวน (Homoscedasticity) ถ้าขาดคุณสมบัติดังกล่าว จะเรียกว่าความไม่คงที่ของความแปรปรวน (Heteroscedasticity) และเรียกพารามิเตอร์ σ^2 ในข้อกำหนดนี้ว่าพารามิเตอร์ความแปรปรวนของความคลาดเคลื่อนสุ่ม และมักนิยมเรียกว่าพารามิเตอร์รบกวน (Nuisance Parameter)

4. กำหนดให้ตัวแปรสุ่มความคลาดเคลื่อน ε_i และ ε_j สำหรับ $i \neq j$ มีการแจกแจงที่เป็นอิสระกัน (Independence) จะทำให้ $\operatorname{cov}(\varepsilon_i , \varepsilon_j) = 0$ สำหรับ $i \neq j = 1,2,\ldots,n$

5. ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (multicollinearity)

อย่างไรก็ตามในทางปฏิบัติ บ่อยครั้งที่เราพบว่าข้อมูลที่นำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว โดยตัวแปรสุ่มความคลาดเคลื่อน $\varepsilon_i; i = 1,2,...,n$ มีการแจกแจงไม่เป็นแบบปกติ ซึ่งจะส่งผลกระทบกับการวิเคราะห์ความถดถอย

ซึ่งผลกระทบจากการที่ข้อมูลไม่ได้มีคุณสมบัติเป็นไปตามข้อกำหนดเบื้องต้นรุนแรงมาก และมีผลกระทบโดยตรงต่อการอนุมานความถดถอย ดังนั้น การตรวจสอบว่าข้อมูลมีคุณสมบัติเป็นไปตามข้อกำหนดหรือไม่นั้นจึงเป็นความจำเป็นอย่างยิ่ง โดยวิธีการตรวจสอบข้อกำหนดเบื้องต้นดังกล่าวของข้อมูลนั้น ไม่ว่าจะเป็นข้อกำหนดเบื้องต้นใด ข้อมูลที่จะนำมาตรวจสอบคือ เศษเหลือ (Residuals) สำหรับวิธีการตรวจสอบข้อกำหนดเบื้องต้นของข้อมูลนั้นสามารถแบ่งออกเป็น 2 ประเภท ดังนี้

- วิธีตรวจสอบด้วยกราฟ (Graphical Method)
- วิธีตรวจสอบด้วยการหานัยสำคัญทางสถิติ (Statistical Significance Method)

สำหรับการตรวจสอบด้วยกราฟ จะใช้ข้อมูลเศษเหลือมาเขียนกราฟ และกราฟที่ใช้ในการตรวจสอบมีมากมายหลายชนิดด้วยกัน ซึ่งกราฟที่นิยมใช้มีดังนี้

- กรณีข้อมูลน้อย $(n < 100)$
  - Quantile-Quantile Plot (Q-Q Plot)
  - Percentile-Percentile Plot (P-P Plot)
  - Normal Probability Plot (NPP Plot)
- กรณีข้อมูลมาก $(n \geq 100)$
  - Histogram
  - Stem-and-leaf-Plot

เช่นเดียวกันกับการตรวจสอบด้วยกราฟ การตรวจสอบหานัยสำคัญสามารถแบ่งออกเป็น 2 ส่วน ดังนี้คือ

- กรณีข้อมูลน้อย $(n < 100)$
  - ตัวสถิติของ Shapiro และ Wilk
  - ตัวสถิติของ Anderson และ Darling
- กรณีข้อมูลมาก $(n \geq 100)$
  - ตัวสถิติของ Kolmogorov และ Smirnov
  - ตัวสถิติไคสแควร์เพื่อการทดสอบสารูปสนิทดี (Goodness-of-Fit Test)

นอกจากนี้ยังมีตัวสถิติทดสอบต่าง ๆ อีกเช่น ตัวสถิติของ Cramer-von Mises, Watson, Kuiper, Durbin, Jarque-Bera, Shapiro-Francia, Weisberg-Bingham, D'Agostino และ Filliben ซึ่งจะมีกระบวนการในการทดสอบที่แตกต่างกันไป

ในการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้นพหุคูณก่อนอื่นเราจะต้องมีกระบวนการการประมาณค่าสัมประสิทธิ์ในสมการถดถอย หลังจากนั้นจึงทำการหาเศษเหลือและนำค่าเศษเหลือที่ได้ไปทดสอบโดยใช้ตัวสถิติทดสอบต่าง ๆ ที่ใช้สำหรับตรวจสอบการแจกแจงไม่เป็นปกติ

วิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์ในสมการถดถอยเชิงเส้นมีอยู่ด้วยกันหลายวิธี แต่ผู้วิจัยควรเลือกใช้วิธีที่เหมาะสมกับลักษณะของข้อมูล และเป็นไปตามข้อตกลงเบื้องต้น (Assumption) ของแต่ละวิธีเพราะจะทำให้ได้ตัวประมาณที่ดีมีประสิทธิภาพ

วิธีที่นิยมใช้กันมากในการประมาณค่าสัมประสิทธิ์ความถดถอย คือ วิธีกำลังสองน้อยสุด (Ordinary Least Squares Method) ซึ่งจะได้ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ $\underset{\sim}{\beta}$ และให้ค่าความแปรปรวนต่ำสุด $= \sigma^2(X'X)^{-1}$ นั่นก็คือ $\hat{\underset{\sim}{\beta}}$ มีคุณสมบัติเป็นตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง (Best Linear Unbiased Estimator (BLUE)) ตามทฤษฎีของ Gauss-Markov และจะประมาณค่า σ^2 ด้วย $\hat{\sigma}^2$ ซึ่ง $\hat{\sigma}^2 = \frac{1}{n-p}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)$ เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ σ^2 แต่ $\hat{\underset{\sim}{\beta}}$ และ $\hat{\sigma}^2$ จะเป็นตัวประมาณที่มีประสิทธิภาพ (Efficiency Estimator) คือมีค่าความแปรปรวนต่ำสุดเพียงตัวเดียวในบรรดาตัวประมาณที่ไม่เอนเอียง (Uniformly Minimum Variance Unbiased Estimator (UMVUE)) ก็ต่อเมื่อความคลาดเคลื่อน (Error) มีการแจกแจงแบบปกติและมีคุณสมบัติตามข้อกำหนดข้างต้น ในกรณีที่ไม่ทราบการแจกแจงของความคลาดเคลื่อน (Error) จึงควรพิจารณาหาวิธีการอื่นในการประมาณค่าของ $\underset{\sim}{\beta}$ และ σ^2 ที่ดีกว่าตัวประมาณที่ได้จากวิธีกำลังสองน้อยสุด โดยไม่จำเป็นต้องมีข้อตกลงเบื้องต้นเกี่ยวกับลักษณะการแจกแจงของความคลาดเคลื่อน

ในบรรดาวิธีการทางนอนพาราเมตริกซ์ ได้มีผู้ศึกษาวิธีการประมาณค่า Standard Error ในกรณีที่ไม่ทราบลักษณะการแจกแจงของประชากรและไม่สามารถหาได้จากสูตรทั่วไป โดยใช้เทคนิคของการสุ่มตัวอย่างซ้ำ (Resampling) ซึ่งมีอยู่ด้วยกันหลายวิธี ได้แก่

- The jackknife
- The bootstrap

- Half – sampling
- Subsampling
- Balanced repeated replication
- The infinistesimal jackknife
- Influence function techniques
- The delta method

ซึ่งแต่ละวิธีมาจากแนวความคิดพื้นฐานคล้ายกันคือ หาค่าประมาณของ Standard Error โดยการสุ่มตัวอย่างซ้ำจากข้อมูลที่เก็บรวบรวมมา และพบว่าวิธีบูทสแตรป(Bootstrap) เป็นวิธีที่ให้ผลดีที่สุดเพราะว่าการหาค่าประมาณโดยวิธีนี้เป็น Nonparametric Maximum Likelihood Estimate ทำให้ตัวประมาณที่ได้เป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator (MLE))[1] นอกจากนี้วิธีบูทสแตรปยังสามารถนำไปใช้ในการประมาณค่าพารามิเตอร์อื่นๆที่สนใจ เมื่อไม่ทราบลักษณะการแจกแจงของประชากร โดยที่หลักการของวิธีบูทสแตรป คือ เป็นการสุ่มตัวอย่างจากประชากรแบบใส่คืน (With Replacement) นั่นคือเป็นการสุ่มตัวอย่างที่ยอมให้มีหน่วยตัวอย่างซ้ำกันได้ โดยที่แต่ละหน่วยตัวอย่างมีโอกาส (Probability) ในการถูกสุ่มเท่ากัน ทำการสุ่มตัวอย่างด้วยจำนวนครั้งที่มากพอ เพื่อสร้างการแจกแจงของตัวสถิติตัวอย่างจากนั้นประมาณค่าพารามิเตอร์ที่สนใจ

สำหรับงานวิจัยที่เกี่ยวข้องนั้น ได้มีนักสถิติจำนวนมากทำการศึกษาเกี่ยวกับอำนาจการทดสอบของตัวสถิติที่ใช้ในการในการทดสอบการแจกแจงแบบปกติ ซึ่งในส่วนนี้จะเสนอเฉพาะผลงานบางส่วนที่สำคัญและเกี่ยวข้องกับวิทยานิพนธ์เรื่องนี้เท่านั้น

Shapiro และคณะ (1968) เป็นผู้ริเริ่มในการศึกษาอำนาจการทดสอบของตัวสถิติต่าง ๆ ในการทดสอบการแจกแจงแบบปกติ โดยทำการศึกษาตัวสถิติ 9 ตัว คือ Shapiro-Wilk Statistic (W) $\sqrt{b_1}$ b_2 Kolmogorov-Smirnov Test (K) Cramer-von Mises (W^2) Anderson-Darling (A^2) Durbin (D) Chi-square Test (χ^2) และ Studentized Rank Test (U) ภายใต้การแจกแจง 12 การแจกแจง ซึ่งมีค่าพารามิเตอร์ที่แตกต่างกันรวมเป็น 45 การแจกแจง ได้ผลสรุปดังนี้

1. Shapiro-Wilk Statistic ใช้ได้ดีในการทดสอบทั่วไป
2. การทดสอบโดยใช้ Empirical Distribution Function ได้อำนาจการทดสอบต่ำ

---

[1] Efron, B. “The Bootstrap” <u>The Jackknife, the Bootstraps and Other Resampling plans</u> (1982):27

3. Studentized Rank Test มีอำนาจการทดสอบสูงเมื่อประชากรมีการแจกแจงแบบ Symmetric Short-Tailed และมีอำนาจการทดสอบต่ำเมื่อประชากรมีการแจกแจงแบบ Asymmetric Short-Tailed และ Asymmetric Long-Tailed

4. $\sqrt{b_1}$ และ b_2 ใช้ในการทดสอบได้ดี แต่มีอำนาจการทดสอบต่ำกว่า Shapiro-Wilk Statistic

Vigliohne ได้เปรียบเทียบการทดสอบระหว่างตัวสถิติ Hosking and Willis เมื่อสัมประสิทธิ์ความเบ้มีค่าต่ำกับตัวสถิติ Bootstrapped Anderson-Darling โดยจะพิจารณาถึงค่าของอำนาจในการทดสอบและความผิดพลาดชนิดที่ 1 ได้ผลสรุปดังนี้

- ถ้าสัมประสิทธิ์ความเบ้มีค่าต่ำกว่า 0.23 ควรใช้ตัวสถิติ Hosking and Willis
- ถ้าสัมประสิทธิ์ความเบ้มีค่าสูงกว่า 0.23 ควรใช้ตัวสถิติ Bootstrapped Anderson-Darling

Pranab Kumar Sen และคณะ ได้ทำการศึกษาตัวสถิติ Shapiro-Wilk ในสถานการณ์ที่มีตัวแปรรบกวนในตัวแบบความถดถอยโดยใช้กระบวนการของ Monte Carlo เข้ามาช่วย เปรียบเทียบกับตัวสถิติ Jareckova ผลการทดสอบพบว่า เมื่อใช้ตัวสถิติ Shapiro-Wilk จะทำให้ง่ายต่อการคิดค่าต่างๆ ทางสถิติและมีอำนาจในการทดสอบสูง และในกรณีที่ขนาดตัวอย่างมีจำนวนจำกัด อำนาจการทดสอบของตัวสถิติ Shapiro-Wilk จะสูงกว่าการทดสอบของตัวสถิติ Jareckova ในขณะที่ขนาดตัวอย่างมีขนาดใหญ่ผลที่ได้จะตรงกันข้ามกัน การทดสอบของตัวสิถิติ Jareckova จะใช้ได้ดีเมื่อขนาดตัวอย่างมีขนาดใหญ่

เกตุจันทร์ พัชรินทร์ศักดิ์ ได้ทำการเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบปกติ 6 ตัวคือ ตัวสถิติของ Shapiro-Wilk ตัวสถิติของ Cramer-von Mises ตัวสถิติของ Anderson-Darling ตัวสถิติของ Watson ตัวสถิติของ Kuiper และตัวสถิติของ Durbin โดยกำหนดการแจกแจงของประชากรเป็นแบบปกติและแบบเบ้ โดยใช้ขนาดตัวอย่างเท่ากับ 10 20 30 50 และ 100 ที่ระดับนัยสำคัญ 0.05 และ 0.10 โดยใช้เกณฑ์ของความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบ ในการพิจารณาเลือกสถิติทดสอบ ผลการทดสอบได้ผลสรุปดังนี้

1. ตัวสถิติ Anderson-Darling
   - ที่ระดับนัยสำคัญ 0.01

     เหมาะสำหรับการทดสอบที่ประชากรมีการแจกแจงแบบ Symmetric Short-Tailed และขนาดตัวอย่างมากกว่า 30

เหมาะสำหรับการทดสอบที่ประชากรมีการแจกแจงแบบ Near Normal Symmetric Long-Tailed, Asymmetric Short-Tailed, Asymmetric Long-Tailed และขนาดตัวอย่างมากกว่า 50

- ที่ระดับนัยสำคัญ 0.05 และ 0.01

เหมาะสำหรับการทดสอบที่ประชากรมีการแจกแจงแบบ Symmetric Short-Tailed Near Normal, Symmetric Long-Tailed, Asymmetric Short-Tailed, Asymmetric Long-Tailed และขนาดตัวอย่างทุกระดับ

2. ตัวสถิติ Shapiro-Wilk

- ที่ระดับนัยสำคัญ 0.01

เหมาะสำหรับการทดสอบที่ประชากรมีการแจกแจงแบบ Near Normal Symmetric Long-Tailed, Asymmetric Short-Tailed, Asymmetric Long-Tailed และขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 50

3. ตัวสถิติ Kuiper

- ที่ระดับนัยสำคัญ 0.01

เหมาะสำหรับการทดสอบที่ประชากรมีการแจกแจงแบบ Symmetric Short-Tailed และขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 30

จากปัญหาดังกล่าวและผลงานวิจัยที่ผ่านมา จึงสนใจที่จะทำการศึกษาเปรียบเทียบการทดสอบตัวสถิติสำหรับการแจกแจงของความคลาดเคลื่อนว่ามีการแจกแจงแบบปกติหรือไม่ ซึ่งได้ใช้หลักการของบูทสแตรปเข้ามาช่วย โดยจะใช้เทคนิคการจำลองแบบที่เรียกว่า เทคนิคมอนติคาร์โลซีมูเลชัน (Monte Carlo Simulation Technique) ในการสร้าง (Generate) ข้อมูลตามขนาดและลักษณะที่ต้องการ ซึ่งเกณฑ์ที่ใช้ในการเลือกตัวสถิติทดสอบจะพิจารณาจาก ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

1.2 วัตถุประสงค์การวิจัย

ในการวิจัยครั้งนี้ ต้องการเปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบปกติ 2 ตัว คือ

1. Bootstrapped Anderson-Darling test
2. Bootstrapped Shapiro-Wilk test

1.3 สมมติฐานการวิจัย

ตัวสถิติ Bootstrapped Anderson-Darling มีความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และให้อำนาจการทดสอบสูงกว่าตัวสถิติของ Bootstrapped Shapiro-Wilk

1.4 ข้อตกลงเบื้องต้น

1.4.1 รูปแบบ(model) ที่ใช้คือ

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} ... + \beta_p x_{ip} + \varepsilon_i \quad ; \quad i = 1,2,3,\ldots,n$$

1.4.2 ความคลาดเคลื่อนเป็นตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบเดียวกันและเป็นอิสระซึ่งกันและกัน (Identically Independent Distribution)

นั่นคือ $\varepsilon_i \sim$ iid F $\quad i = 1,2,\ldots,n$

F เป็น Probability Distribution ที่ไม่ทราบ

$$E(\varepsilon_i) = 0 \quad ; \; i = 1,2,\ldots,n$$

$$E(\varepsilon_i \varepsilon_j) = 0 \quad ; \; i \neq j = 1,2,\ldots,n$$

$$Var(\varepsilon_i) = \sigma^2 \quad ; \; \sigma^2 > 0 \quad \text{และไม่ทราบค่า}$$

1.4.3 ตัวแปรอิสระเป็นอิสระแก่กันและอิสระจากความคลาดเคลื่อน

1.4.4 ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่ที่เป็นอิสระซึ่งกันและกัน

1.4.5 จำนวนข้อมูลตัวอย่างต้องมากกว่าจำนวนตัวแปรอิสระ ($n > p$)

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติ 2 ตัวต่อไปนี้

1.5.1.1) Bootstrapped Anderson-Darling test

1.5.1.2) Bootstrapped Shapiro-Wilk test

1.5.2 ขนาดตัวอย่างที่ทำการศึกษา(n)มี 4 ระดับคือ 10 20 30 และ 50

1.5.3 กำหนดจำนวนตัวแปรอิสระ(p)เป็น 3 และ 5

1.5.4 กำหนดระดับนัยสำคัญของการทดสอบ (Level of Significance) 3 ระดับคือ 0.01 0.05 และ 0.10

1.5.5 ลักษณะการแจกแจงของความคลาดเคลื่อนที่ศึกษาคือ

- การแจกแจงแบบปกติ (Normal Distribution)
- การแจกแจงแบบโลจิสติค (Logistic Distribution)
- การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution)

1.5.6 ทุกลักษณะการแจกแจงจะมีค่าเฉลี่ยของความคลาดเคลื่อนเป็น 0 นั่นคือ $E(\varepsilon_i) = 0$; $i = 1,2,...,n$

1.5.7 กำหนดให้ข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: σ) ในระดับต่างๆกันคือ 1 5 และ 10

1.5.8 ในการวิจัยครั้งนี้สร้างแบบการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo simulation) เขียนด้วยโปรแกรม S-PLUS 2000

1.5.9 การจำลองในแต่ละสถานการณ์ของการทดลองการทำซ้ำ 600 รอบ

1.5.10 การสร้างตัวอย่างสุ่มในวิธีการทดสอบตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์จะกระทำซ้ำ 200 รอบ

1.5.11 แต่ละลักษณะการแจกแจงจะมีค่าฟังก์ชันความน่าจะเป็น ค่าคาดหวัง ค่าความแปรปรวน ดังนี้

1.5.11.1 <u>การแจกแจงแบบปกติ</u> (Normal Distribution)

ฟังก์ชันความน่าจะเป็น คือ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad ;-\infty < x < \infty$$

ค่าคาดหวัง $E(x) = \mu$

ค่าความแปรปรวน $V(x) = \sigma^2$

1.5.11.2 <u>การแจกแจงแบบโลจิสติค</u> (Logistic Distribution)

ฟังก์ชันความน่าจะเป็น คือ

$$f(x) = \frac{1}{\beta} . \frac{e^{-\left(\frac{x-\alpha}{\beta}\right)}}{\left(1+e^{-\left(\frac{x-\alpha}{\beta}\right)}\right)^2} \quad ;-\infty < x < \infty$$

ค่าคาดหวัง $E(x) = \alpha$

ค่าความแปรปรวน $V(x) = \frac{1}{3}\pi^2\beta^2$

1.5.11.3 <u>การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล</u> (Double Exponential Distribution)

ฟังก์ชันความน่าจะเป็น คือ

$$f(x) = \frac{1}{2\beta}.e^{-\left|\frac{x-\alpha}{\beta}\right|} \qquad ; -\infty < x < \infty$$

ค่าคาดหวัง $E(x) = \alpha$

ค่าความแปรปรวน $V(x) = 2\beta^2$

1.6 เกณฑ์ที่ใช้ในการตัดสินใจ

ในการวิจัยครั้งนี้จะใช้ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ (Power of the Test) เป็นเกณฑ์ในการเปรียบเทียบตัวสถิติทดสอบ

1.6.1 ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ซึ่งวัดจากสัดส่วนความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นเป็นจริง โดยใช้เกณฑ์ของ Bradley (ค.ศ. 1978, 144-52) ซึ่งจะพิจารณาว่าถ้าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองอยู่ในช่วง ซึ่งจำแนกได้ดังนี้

เมื่อทดสอบที่ระดับนัยสำคัญ 0.01 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.005,0.015]

เมื่อทดสอบที่ระดับนัยสำคัญ 0.05 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.025,0.075]

เมื่อทดสอบที่ระดับนัยสำคัญ 0.10 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.050,0.150]

จากผลการทดลองของแต่ละสถานการณ์ที่จำลองขึ้นมา ถ้าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบใดอยู่นอกขอบเขตที่ระบุสำหรับเกณฑ์ที่กำหนดแสดงว่าการทดสอบนั้นไม่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ซึ่งสามารถแบ่งออกเป็น 2 กรณีคือ

1.6.1.1 กรณีที่ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดลองมากกว่าขอบเขตบนของเกณฑ์ที่ใช้พิจารณา แสดงว่าการทดสอบนั้นมีค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มากกว่าค่าระดับนัยสำคัญที่กำหนด

1.6.1.2 กรณีที่ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดลองน้อยกว่าขอบเขตล่างของเกณฑ์ที่ใช้พิจารณา แสดงว่าการทดสอบนั้นมีค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 น้อยกว่าค่าระดับนัยสำคัญที่กำหนด

ในกรณีที่ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่ระบุสำหรับแต่ละเกณฑ์ที่กำหนด แสดงว่าการทดสอบนั้นมีค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เท่ากับค่าระดับนัยสำคัญที่กำหนด และสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

1.6.2 อำนาจการทดสอบ (Power of the Test) ซึ่งวัดจากสัดส่วนความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างไม่จริง โดยจะพิจารณาว่าสถิติทดสอบตัวใดมีอำนาจในการทดสอบสูงกว่ากัน

1.7 วิธีดำเนินการวิจัย

งานวิจัยนี้ได้แบ่งวิธีดำเนินการออกเป็นขั้นตอนหลักๆ ดังนี้

1.7.1 ศึกษาวิธีการและทำความเข้าใจในทฤษฎีที่เกี่ยวข้องกับตัวสถิติที่ใช้สำหรับตรวจสอบการแจกแจงไม่เป็นปกติของแต่ละวิธี คือ
- บูทสแตรปแอนเดอร์สัน – ดาร์ลิง
- บูทสแตรปแชปิโร – วิลค์

1.7.2 สร้างโปรแกรมย่อยสำหรับการแจกแจงของความคลาดเคลื่อนตามที่กำหนดไว้ในขอบเขตของการวิจัย

1.7.3 สร้างข้อมูลเพื่อใช้ในการวิจัย โดยในงานวิจัยครั้งนี้ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล กำหนดจำนวนรอบทำซ้ำ 600 รอบในแต่ละสถานการณ์

1.7.4 คำนวณหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อประชากรมีการแจกแจงแบบปกติ และคำนวณหาค่าอำนาจการทดสอบเมื่อประชากรมีการแจกแจงแบบโลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล

1.7.5 เปรียบเทียบตัวสถิติดังกล่าวกับเกณฑ์ที่กำหนด พร้อมทั้งบันทึกผลเพื่อสรุปการวิจัย

1.8 คำจำกัดความ

1.8.1 ความคลาดเคลื่อนชนิดที่ 1 (Type I error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานว่าง (Null Hypothesis) เมื่อสมมติฐานว่างเป็นจริง

1.8.2 อำนาจการทดสอบ (Power of The Test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง(Null Hypothesis) เมื่อสมมติฐานว่างไม่เป็นจริง

1.8.3 ความแปรปรวน (Variance) ของตัวประมาณ คือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว ความแปรปรวนของ $\hat{\theta}$ คือ $E(\hat{\theta} - E(\hat{\theta}))^2$

1.8.4 <u>ความไม่เอนเอียง</u> (Unbiased) ของตัวประมาณ คือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว จะถือว่า $\hat{\theta}$ เป็นตัวประมาณที่ไม่เอนเอียงของ θ ก็ต่อเมื่อ $E(\hat{\theta}) = \theta$

1.8.5 <u>ตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง</u> (Best Linear Unbiased Estimator) หรือ BLUE เป็นคุณสมบัติหนึ่งของตัวประมาณ โดยตัวประมาณ θ จะมีคุณสมบัติเป็น BLUE ของพารามิเตอร์ θ ถ้า $\hat{\theta}$ มีคุณสมบัติครบ 3 ข้อดังกล่าวต่อไปนี้

1.8.5.1 เป็นฟังก์ชันเชิงเส้นของตัวอย่างสุ่ม

1.8.5.2 เป็นตัวประมาณที่ไม่เอนเอียง

1.8.5.3 เป็นตัวประมาณที่มีค่าความแปรปรวนต่ำสุด

1.8.6 <u>ตัวประมาณที่มีค่าความแปรปรวนต่ำสุดเพียงตัวเดียวในบรรดาตัวประมาณที่ไม่เอนเอียง</u> (Uniformly Minimum Variance Estimator (UMVUE)) เป็นคุณสมบัติของตัวประมาณที่ผู้วิจัยต้องการ คือถ้ามีตัวประมาณที่ไม่เอนเอียงซึ่งเป็นสถิติที่พอเพียงสำหรับ θ และมีค่าความแปรปรวนต่ำกว่าค่าความแปรปรวนของตัวประมาณอื่น ๆ ที่ไม่เอนเอียงสำหรับ θ แล้ว ตัวประมาณดังกล่าว จะมีคุณสมบัติเป็นตัวประมาณที่ดีที่สุดหรือมีค่าความแปรปรวนต่ำสุดเพียงตัวเดียวในบรรดาตัวประมาณที่ไม่เอนเอียง

1.8.7 <u>สัมประสิทธิ์ความถดถอย</u> (Regression coefficient(β)) หมายถึงค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของตัวแปรตาม (Y) เมื่อตัวแปรอิสระ (X) เปลี่ยนไปหนึ่งหน่วย หรือก็คือความชัน (slope) ของเส้นตรงนั่นเอง

1.9 ประโยชน์ของการวิจัย

เพื่อเป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้ตัวสถิติทดสอบ สำหรับทดสอบการแจกแจงของความคลาดเคลื่อนได้อย่างเหมาะสม โดยการนำวิธีการของบูทสแตรปเข้ามาช่วยเมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น

บทที่ 2

ทฤษฎีและสถิติที่ใช้ในการวิจัย

การวิจัยในครั้งนี้เป็นการศึกษาเปรียบเทียบตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติ ของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ในกรณีที่มีจำนวนตัวแปรอิสระมากกว่า 1 ตัว ในขั้นต้นจะกล่าวถึงรายละเอียดของวิธีการประมาณค่าพารามิเตอร์และคุณสมบัติของตัวประมาณที่ได้ ซึ่งจะแสดงรูปของเวคเตอร์และเมตริกซ์ สำหรับรายละเอียดเกี่ยวกับกระบวนการหาเศษเหลือซึ่งมีการนำวิธีการของบูทสแตรปมาใช้ ตัวสถิติทดสอบบูทสแตรปแชปิโร-วิลค์และตัวสถิติทดสอบบูทสแตรปแอนเดอร์สัน-ดาร์ลิง จะกล่าวถึงในหัวข้อถัดไป

2.1 การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองต่ำสุด (Ordinary Least Square Estimation)[2]

ตัวสถิติที่ใช้ในการประมาณค่าพารามิเตอร์ในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติ คือ ตัวประมาณที่ได้จากวิธีกำลังสองต่ำสุด

2.1.1 วิธีกำลังสองต่ำสุด (Ordinary Least Square Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) เป็นวิธีการที่คิดขึ้นโดย คาร์ล เฟรดริก เกาส์ (Karl Friedrich Gauss 1777-1855) และ อังเดร แอนดรีวิช มาร์คอฟ (Andrel Andreevich Markov 1856-1922) ซึ่งมีหลักเกณฑ์ดังนี้คือ ให้หาตัวประมาณของพารามิเตอร์ที่ทำให้ผลบวกของกำลังสองของความคลาดเคลื่อน (Sum Square Errors (SSE)) มีค่าต่ำสุด ซึ่งรายละเอียดในการหาแสดงในรูปของเวกเตอร์และเมตริกซ์ได้ดังนี้

2.1.1.1 การหาตัวประมาณกำลังสองต่ำสุด

ในที่นี้จะพิจารณาตัวแบบเชิงเส้น (Linear model) ในรูป

$$\underset{\sim}{y} = X\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

[2] มาลี ตระการศิรินนท์ "การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการความถดถอยเชิงเส้นด้วยวิธีกำลังสองต่ำสุดและวิธีบูทสแตรป",(วิทยานิพนธ์ปริญญามหาบัณฑิตภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย 2532),หน้า 13-15

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมตริกซ์ของตัวแปรอิสระหรือค่าคงที่ที่ทราบค่าขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของค่าพารามิเตอร์ซึ่งเป็นค่าสัมประสิทธิ์ความถดถอยขนาด $(p+1) \times 1$

n เป็นจำนวนค่าสังเกต

p เป็นจำนวนตัวแปรอิสระ

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

ซึ่ง $E\left(\underset{\sim}{\varepsilon}\right) = 0$

$E\left(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}'\right) = \sigma^2 I$ โดยที่ I เป็นเมตริกซ์เอกลักษณ์ (single matrix)

จากหลักเกณฑ์ดังกล่าวข้างต้น จะทำการหา $\hat{\underset{\sim}{\beta}}$ ซึ่งเป็นตัวประมาณกำลังสองต่ำสุดของ $\underset{\sim}{\beta}$ ที่ทำให้

$$\text{SSE} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\underset{\sim}{\varepsilon}}'\,\hat{\underset{\sim}{\varepsilon}}$$

มีค่าต่ำสุด โดยการหาอนุพันธ์ของ SSE เทียบกับ $\hat{\underset{\sim}{\beta}}$ แล้วกำหนดให้เท่ากับ 0 นั่นคือ

$$\begin{aligned}
\text{SSE} &= \hat{\underset{\sim}{\varepsilon}}'\,\hat{\underset{\sim}{\varepsilon}} = \left(\underset{\sim}{y} - X\,\hat{\underset{\sim}{\beta}}\right)'\left(\underset{\sim}{y} - X\,\hat{\underset{\sim}{\beta}}\right) \\
&= \left(\underset{\sim}{y}' - \hat{\underset{\sim}{\beta}}'X'\right)\left(\underset{\sim}{y} - X\,\hat{\underset{\sim}{\beta}}\right) \\
&= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'X\,\hat{\underset{\sim}{\beta}} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\,\hat{\underset{\sim}{\beta}} \\
&= \underset{\sim}{y}'\underset{\sim}{y} - 2\,\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\,\hat{\underset{\sim}{\beta}}
\end{aligned}$$

$$\frac{\partial}{\partial \hat{\underset{\sim}{\beta}}}\,\text{SSE} = -2X'\underset{\sim}{y} + 2X'X\,\hat{\underset{\sim}{\beta}} = 0$$

$X'X\,\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{y}$ ซึ่งเป็นสมการปกติ (Normal Equation)

ดังนั้นจะได้ตัวประมาณกำลังสองน้อยสุดคือ

$$\hat{\underset{\sim}{\beta}} = \left(X'X\right)^{-1}X'\underset{\sim}{y}$$

เมื่อ $X'X$ ไม่เป็นเมตริกซ์เอกเทศ (Singular matrix)

2.2 คุณสมบัติที่สำคัญของตัวประมาณกำลังสองต่ำสุด[3]

2.2.1 ความไม่เอนเอียง (Unbiasedness)

เมื่อ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ และ $E\left(\underset{\sim}{\varepsilon}\right) = 0$ ตัวประมาณกำลังสองต่ำสุดของ $\underset{\sim}{\beta}$ เป็นตัวประมาณที่ไม่เอนเอียง นั่นคือ $E\left(\underset{\sim}{\hat{\beta}}\right) = \underset{\sim}{\beta}$

พิสูจน์ จาก

$$\underset{\sim}{\hat{\beta}} = \left(X'X\right)^{-1}X'\underset{\sim}{y}$$

$$= \left(X'X\right)^{-1}X'\left(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}\right)$$

$$= \left(X'X\right)^{-1}X'X\underset{\sim}{\beta} + \left(X'X\right)^{-1}X'\underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{\beta} + \left(X'X\right)^{-1}X'\underset{\sim}{\varepsilon}$$

$$E\left(\underset{\sim}{\hat{\beta}}\right) = E\left(\underset{\sim}{\beta}\right) + E\left(X'X\right)^{-1}X'\underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{\beta} + \left(X'X\right)^{-1}X'E\left(\underset{\sim}{\varepsilon}\right)$$

$$= \underset{\sim}{\beta} \quad ; \quad E\left(\underset{\sim}{\varepsilon}\right) = 0$$

$\therefore \underset{\sim}{\hat{\beta}}$ เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ $\underset{\sim}{\beta}$

2.2.2 ความแปรปรวนของ $\underset{\sim}{\hat{\beta}} = \sigma^2\left(X'X\right)^{-1}$ เมื่อ $E\left(\underset{\sim}{\varepsilon}\right) = 0$, $V\left(\underset{\sim}{\varepsilon}\right) = \sigma^2 I$

พิสูจน์

จาก $\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta} = \left(X'X\right)^{-1}X'\underset{\sim}{\varepsilon}$

$$V\left(\underset{\sim}{\hat{\beta}}\right) = E\left(\underset{\sim}{\hat{\beta}} - E\left(\underset{\sim}{\hat{\beta}}\right)\right)\left(\underset{\sim}{\hat{\beta}} - E\left(\underset{\sim}{\hat{\beta}}\right)\right)'$$

$$= E\left(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta}\right)\left(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta}\right)'$$

$$= E[\left(X'X\right)^{-1}X'\underset{\sim}{\varepsilon}.\underset{\sim}{\varepsilon}'X\left(X'X\right)^{-1}]$$

$$= \left(X'X\right)^{-1}X'E\left(\underset{\sim}{\varepsilon}.\underset{\sim}{\varepsilon}'\right)X\left(X'X\right)^{-1}$$

$$\therefore V\left(\underset{\sim}{\hat{\beta}}\right) = \sigma^2\left(X'X\right)^{-1}$$

---

[3] มาลี ตระการศิรินนท์ “การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการความถดถอยเชิงเส้นด้วยวิธีกำลังสองต่ำสุด และวิธีบูทสแตรป”,(วิทยานิพนธ์ปริญญามหาบัณฑิตภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย 2532),หน้า 15-17

<u>2.2.3 ค่าประมาณความแปรปรวนของ</u> $\underset{\sim}{\hat{\beta}} = \hat{\sigma}^2 \left(X^{'}X\right)^{-1}$ เมื่อ $\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\varepsilon_i^2$

จาก $$V\left(\underset{\sim}{\hat{\beta}}\right) = \sigma^2 \left(X^{'}X\right)^{-1}$$

ในกรณีที่ไม่ทราบค่า σ^2 จึงต้องทำการประมาณ

จาก $$\underset{\sim}{\hat{\beta}} = \underset{\sim}{\beta} + \left(X^{'}X\right)^{-1} X^{'} \underset{\sim}{\varepsilon}$$

<u>พิจารณา</u>

$$\begin{aligned}
\underset{\sim}{\hat{\varepsilon}} &= \underset{\sim}{y} - \underset{\sim}{\hat{y}} \\
&= \underset{\sim}{y} - X\underset{\sim}{\hat{\beta}} \\
&= X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} - X[\underset{\sim}{\beta} + \left(X^{'}X\right)^{-1} X^{'} \underset{\sim}{\varepsilon}] \\
&= X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} - X\underset{\sim}{\beta} - X\left(X^{'}X\right)^{-1} X^{'} \underset{\sim}{\varepsilon} \\
&= \underset{\sim}{\varepsilon} - X\left(X^{'}X\right)^{-1} X^{'} \underset{\sim}{\varepsilon} \\
&= [I_n - X\left(X^{'}X\right)^{-1} X^{'}]\underset{\sim}{\varepsilon} \\
&= M\underset{\sim}{\varepsilon}
\end{aligned}$$

เมื่อ $M = I_n - X\left(X^{'}X\right)^{-1} X^{'}$ เป็น symmetric matrix และเป็น Idempotent matrix นั่นคือ $M^2 = M$ และ $M^{'} = M$

<u>พิสูจน์</u>

$$\begin{aligned}
MM &= [I_n - X\left(X^{'}X\right)^{-1} X^{'}][I_n - X\left(X^{'}X\right)^{-1} X^{'}] \\
&= I_n - X\left(X^{'}X\right)^{-1} X^{'} - X\left(X^{'}X\right)^{-1} X^{'} + X\left(X^{'}X\right)^{-1} X^{'}X\left(X^{'}X\right)^{-1} X^{'} \\
&= I_n - X\left(X^{'}X\right)^{-1} X^{'} - X\left(X^{'}X\right)^{-1} X^{'} + X\left(X^{'}X\right)^{-1} X^{'} \\
&= I_n - X\left(X^{'}X\right)^{-1} X^{'} \\
&= M
\end{aligned}$$

<u>พิสูจน์</u>

$$\begin{aligned}
M^{'} &= [I_n - X\left(X^{'}X\right)^{-1} X]^{'} \\
&= [I_n - X\left(X^{'}X\right)^{-1} X^{'}] \\
&= M
\end{aligned}$$

M เป็น Idempotent matrix จริง

จาก

$$\begin{aligned}
E\left(\underset{\sim}{\hat{\varepsilon}}^{'}\underset{\sim}{\varepsilon}\right) &= E\left(\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right) \\
&= E\left(\underset{\sim}{\hat{\varepsilon}} - E\left(\underset{\sim}{\hat{\varepsilon}}\right)\right)^{'}\left(\underset{\sim}{\hat{\varepsilon}} - E\left(\underset{\sim}{\hat{\varepsilon}}\right)\right) \\
&= E\left(\underset{\sim}{\hat{\varepsilon}}^{'}\underset{\sim}{\hat{\varepsilon}}\right) \\
&= E\left(\underset{\sim}{\varepsilon}^{'}M^{'}M\underset{\sim}{\varepsilon}\right) \\
&= E\left(\underset{\sim}{\varepsilon}^{'}M\underset{\sim}{\varepsilon}\right)
\end{aligned}$$

$$= E\left(tr\,\underset{\sim}{\varepsilon}^{'}\,M\,\underset{\sim}{\varepsilon}\right)$$
$$= E\left(trM\,\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}^{'}\right)$$
$$= trE\left(M\,\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}^{'}\right)$$
$$= trME\left(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}^{'}\right)$$
$$= \sigma^2 tr(M)$$
$$= \sigma^2[tr\left(I_n - X\left(X^{'}X\right)^{-1}X\right)]$$
$$= \sigma^2[trI_n - tr\left(X\left(X^{'}X\right)^{-1}X^{'}\right)]$$
$$= \sigma^2[trI_n - trX^{'}X\left(X^{'}X\right)^{-1}]$$
$$= \sigma^2\left(trI_n - trI_p\right)$$
$$= \sigma^2(n-p)$$

$$E\left(\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-p}\right) = \sigma^2$$

$\sum_{i=1}^{n}\frac{\hat{\varepsilon}_i^2}{n-p}$ เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ σ^2

ค่าประมาณความแปรปรวนของ $\hat{\underset{\sim}{\beta}} = \hat{\sigma}^2\left(X^{'}X\right)^{-1}$ เมื่อ $\hat{\sigma}^2 = \sum_{i=1}^{n}\frac{\hat{\varepsilon}_i^2}{n-p}$

2.3 การทดสอบตัวสถิติบูทสแตรปแชปิโร-วิลค์และตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง[4]

วิธีการของบูทสแตรปเป็นการหาค่าประมาณที่ไม่อิงพารามิเตอร์ (Nonparametric Maximum Likelihood Estimate) ทำให้ตัวประมาณที่ได้เป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimate(MLE)) นอกจากนี้วิธีบูทสแตรปยังสามารถนำไปใช้ในการประมาณค่าต่างๆ ที่สนใจ เมื่อไม่ทราบลักษณะการแจกแจงของประชากร โดยมีหลักเกณฑ์ดังนี้คือ ทำการสุ่มตัวอย่างจากข้อมูลที่เก็บรวบรวมมาแบบใส่คืน (with replacement) ขนาดเท่ากับจำนวนตัวอย่างหรือข้อมูลที่มีอยู่แล้วนั้น เพื่อสร้างข้อมูลชุดใหม่แล้วนำมาใช้ในการประมาณค่าสถิติหรือค่าพารามิเตอร์ที่สนใจ

สำหรับจำนวนครั้งที่ทำการสุ่มตัวอย่าง (Bootstrap sampling) นั้นควรจะอยู่ในช่วง 50 – 200 ก็เพียงพอที่จะทำให้ได้ตัวประมาณที่ดี โดยมีขั้นตอนการทำงานดังนี้

---

[4] Natalie Neumeyer, Holger Dette, and Eva-Renate Nagel, “Bootstrap tests for the error distribution in linear and nonparametric regression models”,Ruhr-University Bochum Germany (June 4,2004):1-27

1. สร้างตัวเลขสุ่ม เพื่อนำไปใช้ในการสุ่มตัวอย่างแบบใส่คืน
2. จากตัวอย่างที่ได้แต่ละชุด นำมาหาค่าสถิติหรือค่าประมาณของพารามิเตอร์ที่สนใจ

การศึกษาครั้งนี้ได้นำเอาวิธีการของบูทสแตรปมาใช้ในการเลือกตัวสถิติทดสอบสำหรับทดสอบการแจกแจงของความคลาดเคลื่อน เมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น ซึ่งมีรายละเอียดในการหาแสดงในรูปของเวกเตอร์และเมตริกซ์ได้ดังนี้

2.3.1 การหาค่าตัวสถิติแชปิโร-วิลค์และบูทสแตรปแชปิโร-วิลค์

จาก $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

เมื่อ $E(\underset{\sim}{\varepsilon}) = 0$

$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 I$

$\underset{\sim}{\varepsilon_i}$ ~ iid F ; ไม่ทราบการแจกแจง F

การหาตัวประมาณกระทำตามขั้นตอนต่อไปนี้

2.3.1.1 คำนวณหาค่า $\underset{\sim}{\hat{\beta}}$ โดยวิธีกำลังสองต่ำสุด

ได้ $\underset{\sim}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$

และ $\underset{\sim}{\hat{y}} = X\underset{\sim}{\hat{\beta}}$

$\underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{y} - \underset{\sim}{\hat{y}}$ ซึ่ง $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$

ให้ F เป็นการแจกแจงเชิงประจักษ์ (empirical distribution) ของ $\hat{\varepsilon}_i$

2.3.1.2 คำนวณหาค่าตัวสถิติแชปิโร-วิลค์ โดยนำค่าของ $\underset{\sim}{\hat{\varepsilon}}$ ที่คำนวณได้แทนในสูตรสถิติดังต่อไปนี้

การทดสอบของตัวสถิติแชปิโร-วิลค์ (Shapiro-Wilk method)[5]

วิธีทดสอบของตัวสถิติแชปิโร-วิลค์ เป็นวิธีทดสอบที่ค่อนข้างแกร่ง เมื่อเปรียบเทียบกับวิธีทดสอบอื่นๆ และเหตุผลที่สำคัญอีกประการหนึ่งคือ วิธีทดสอบนี้ยังสามารถใช้ได้กับขนาดตัวอย่างขนาดเล็กได้ (n < 30) สำหรับสถิติทดสอบความปกติของตัวสถิติแชปิโร-วิลค์ เป็นดังนี้

---

[5]สุพล ดุรงค์วัฒนา, การวางแผนการทดลองเพื่อการวิจัยขั้นสูง, (ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย 2549), หน้า 190-191

$$SW = \frac{\left(\sum_{i=1}^{n} a_i \hat{\varepsilon}_{(i)}\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}$$

$\hat{\varepsilon}_{(i)}$ คือ ค่าเศษเหลืออันดับของเศษเหลือจากข้อมูล

a_i คือ ค่าคงที่ที่ถูกสร้างขึ้นจากค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วมของตัวสถิติลำดับของข้อมูลเศษเหลือ $\hat{\varepsilon}_i$ ขนาด n จากประชากรหรือตัวแปรสุ่มที่มีการแจกแจงแบบปกติโดยสามารถเปิดได้จากตารางค่าสัมประสิทธิ์ของตัวสถิติทดสอบแชปิโร-วิลค์

$\hat{\varepsilon}_i$ คือ ค่าเศษเหลือที่ไม่ได้จัดลำดับ

n คือ ขนาดตัวอย่าง

2.3.1.3 สุ่ม $\hat{\varepsilon}_i$ แบบใส่คืน (with replacement) ขนาด n

ได้ $\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$

$\varepsilon_i^* \sim$ iid F

2.3.1.4 นำค่า ε_i^* มาพิจารณาโดยรวมไว้ในสมการ

จะได้ $\underset{\sim}{y}^* = X \underset{\sim}{\hat{\beta}} + \underset{\sim}{\varepsilon}^* = \underset{\sim}{\hat{y}} + \underset{\sim}{\varepsilon}^*$

เมื่อ $\underset{\sim}{\varepsilon}^* = \varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$

คำนวณหาค่า $\underset{\sim}{\hat{\beta}}^*$ โดยวิธีกำลังสองต่ำสุด เพื่อที่จะนำไปหาค่าของ $\hat{\varepsilon}_i^*$ ต่อไปโดยยึดหลักเกณฑ์เดียวกัน คือทำการหา $\underset{\sim}{\hat{\beta}}^*$ ซึ่งเป็นตัวประมาณกำลังสองต่ำสุดของ $\underset{\sim}{\hat{\beta}}$ ที่ทำให้

$$SSE^* = \sum_{i=1}^{n} \hat{\varepsilon}_i^{*2} = \underset{\sim}{\hat{\varepsilon}}^{*'} \underset{\sim}{\hat{\varepsilon}}^* \quad \text{มีค่าต่ำสุด}$$

โดยการหาอนุพันธ์ของ SSE^* เทียบกับ $\underset{\sim}{\beta}$ แล้วกำหนดให้เท่ากับ 0 นั่นคือ

$$\begin{aligned} SSE^* &= \underset{\sim}{\hat{\varepsilon}}^{*'} \underset{\sim}{\hat{\varepsilon}}^* = \left(\underset{\sim}{y}^* - X \underset{\sim}{\hat{\beta}}^*\right)' \left(\underset{\sim}{y}^* - X \underset{\sim}{\hat{\beta}}^*\right) \\ &= \left(\underset{\sim}{y}^{*\prime} - \underset{\sim}{\hat{\beta}}^{*\prime} X'\right) \left(\underset{\sim}{y}^* - X \underset{\sim}{\hat{\beta}}^*\right) \\ &= \underset{\sim}{y}^{*'} \underset{\sim}{y}^* - \underset{\sim}{y}^{*'} X \underset{\sim}{\hat{\beta}}^* - \underset{\sim}{\hat{\beta}}^{*'} X' \underset{\sim}{y}^* + \underset{\sim}{\hat{\beta}}^{*'} X' X \underset{\sim}{\hat{\beta}}^* \end{aligned}$$

$$= \underset{\sim}{y}^{*'} \underset{\sim}{y}^{*} - 2\underset{\sim}{\hat{\beta}}^{*'} X^{'} \underset{\sim}{y}^{*} + \underset{\sim}{\hat{\beta}}^{*'} X^{'} X \underset{\sim}{\hat{\beta}}^{*}$$

$$\frac{\partial}{\partial \underset{\sim}{\hat{\beta}}^{*}} \text{SSE} = -2X^{'} \underset{\sim}{y}^{*} + 2X^{'} X \underset{\sim}{\hat{\beta}}^{*} = 0$$

$$\therefore X^{'} X \underset{\sim}{\hat{\beta}}^{*} = X^{'} \underset{\sim}{y}^{*}$$

$$\therefore \underset{\sim}{\hat{\beta}}^{*} = \left(X^{'} X\right)^{-1} X^{'} \underset{\sim}{y}^{*}$$

จะได้ว่า $\underset{\sim}{\hat{y}}^{*} = X \underset{\sim}{\hat{\beta}}^{*}$

และ $\underset{\sim}{\hat{\varepsilon}}^{*} = \underset{\sim}{y}^{*} - \underset{\sim}{\hat{y}}^{*}$

2.3.1.5 กระทำตามขั้นตอนในข้อ 2.3.1.3 - 2.3.1.4 ซ้ำ 200 ครั้ง จะได้ $\hat{\varepsilon}^{*1}, \hat{\varepsilon}^{*2}, ..., \hat{\varepsilon}^{*200}$ จำนวน 200 ค่า (200 คือ จำนวนรอบของการทำซ้ำในกระบวนการบูทสแตรป)

2.3.1.6 คำนวณหาค่าตัวสถิติบูทสแตรปแชปิโร-วิลค์ โดยนำค่าของ $\underset{\sim}{\hat{\varepsilon}}^{*}$ ที่คำนวณได้แทนในสูตรสถิติดังต่อไปนี้

<u>การทดสอบของตัวสถิติบูทสแตรปแชปิโร-วิลค์</u>

$$BSW = \frac{\left(\sum_{i=1}^{n} a_i \hat{\varepsilon}_{(i)}^{*}\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^{*2}}$$

$\hat{\varepsilon}_{(i)}^{*}$ คือ ค่าเศษเหลืออันดับของเศษเหลือจากข้อมูลที่ได้จากกระบวนการบูทสแตรป

a_i คือ ค่าคงที่ที่ถูกสร้างขึ้นจากค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วมของตัวสถิติลำดับของข้อมูลเศษเหลือ $\hat{\varepsilon}_i^{*}$ ขนาด n จากประชากรหรือตัวแปรสุ่มที่มีการแจกแจงแบบปกติโดยสามารถเปิดได้จากตารางค่าสัมประสิทธิ์ของตัวสถิติทดสอบแชปิโร-วิลค์

$\hat{\varepsilon}_i^{*}$ คือ ค่าเศษเหลือที่ได้จากกระบวนการบูทสแตรปที่ไม่ได้จัดลำดับ

n คือ ขนาดตัวอย่าง

ซึ่งจะได้ค่าสถิติจำนวน 200 ค่าเช่นกัน แสดงดังตารางต่อไปนี้

จำนวนรอบ	ค่าสถิติบูทสแตรปแชปิโร-วิลค์
1	BSW_1
2	BSW_2
3	BSW_3
.	.
.	.
.	.
i	BSW_i
.	.
.	.
.	.
200	BSW_{200}

2.3.1.7 คำนวณค่าอาณาเขตวิกฤตจากค่าสถิติที่ได้จากกระบวนการบูทสแตรป[6]

สำหรับวิธีการคำนวณค่าอาณาเขตวิกฤตของตัวสถิติที่ได้มาจากกระบวนการบูทสแตรปมีอยู่ด้วยกันหลายวิธีในการวิจัยครั้งนี้ผู้วิจัยได้เลือกวิธีการ Bias-Corrected[7] หรือ BC ซึ่งมีขั้นตอนดังนี้

<u>ขั้นตอนที่ 1</u> คำนวณหาค่าของ Z_0

ซึ่ง Z_0 คือค่าของ Z ที่มีค่าตรงกับสัดส่วนของค่าสถิติที่ได้จากกระบวนการบูทสแตรปที่น้อยกว่าหรือเท่ากับค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรป โดยสามารถหาค่าได้จาก

$$Z_0 = \Phi^{-1}\{pr(BSW \le SW)\}$$

เมื่อ Φ คือฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติ

BSW คือค่าสถิติที่ได้จากกระบวนการบูทสแตรป

SW คือค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรป

[6] Christopher Z. Mooney and Robert D. Dual, <u>Bootstrapping: A Nonparametric Approach to Statistical Inference</u>, Sage Publications(1993), p.37-40

[7] วิธีการ Bias corrected หรือ BC จะไม่มีข้อจำกัดเกี่ยวกับการแจกแจงของความคลาดเคลื่อนและจำนวนรอบของการทำซ้ำ

ขั้นตอนที่ 2 นำค่าสถิติที่ได้จากกระบวนการบูทสแตรปมาเรียงค่าจากน้อยไปมาก

ขั้นตอนที่ 3 คำนวณหาค่าอาณาเขตวิกฤต

ค่าอาณาเขตวิกฤตสามารถหาได้จากค่าของตัวสถิติที่ได้จากกระบวนการบูทสแตรปที่อยู่ในตำแหน่งดังนี้

$[\{\Phi(2Z_0 + Z_\alpha)\}\times 100]$ percentile หรือ $[\{\Phi(2Z_0 - Z_\alpha)\}\times 100]$percentile

เช่น กำหนดให้ $\{pr(BSW \le SW)\} = 0.65$ ระดับนัยสำคัญคือ 0.05
จะได้ว่า

ขั้นตอนที่ 1 คำนวณหาค่าของ Z_0

จาก $Z_0 = \Phi^{-1}\{pr(BSW \le SW)\}$

ดังนั้น $Z_0 = \Phi^{-1}(0.65)$

$= 0.39$

ขั้นตอนที่ 2 นำค่าสถิติที่ได้จากกระบวนการบูทสแตรปมาเรียงค่าจากน้อยไปมาก

ในที่นี้กำหนดให้จำนวนกระทำซ้ำในกระบวนการบูทสแตรปคือ 200 รอบ ดังนั้นค่าสถิติจากกระบวนการบูทสแตรปที่ได้จากแต่ละสถานการณ์จะมีทั้งหมด 200 ค่า นำค่าสถิติทั้ง 200 ค่ามาเรียงลำดับจากน้อยไปหามาก

ขั้นตอนที่ 3 คำนวณหาค่าอาณาเขตวิกฤต

ตำแหน่งของอาณาเขตวิกฤตคือ $[\{\Phi(2Z_0 + Z_\alpha)\}\times 100]$percentile

หรือ $[\{\Phi(2Z_0 - Z_\alpha)\}\times 100]$percentile

ในที่นี้สมมติให้ปฏิเสธสมมติฐานหลักเมื่อ $SW_{cal} < SW_{table}$

โดยที่ค่าของ SW_{cal} คือ ค่าสถิติแชปิโร-วิลค์ที่ได้จากการคำนวณซึ่งไม่ผ่านกระบวนการบูทสแตรป

SW_{table} คือ ค่าวิกฤตที่ได้จากตารางสถิติของแชปิโร-วิลค์

ดังนั้นตำแหน่งของอาณาเขตวิกฤตคือ $[\{\Phi(2Z_0 - Z_\alpha)\}\times 100]$percentile

$= [\{\Phi(2(0.39) - 1.645)\}\times 100]$percentile

$$=\Phi(-0.865)\times 100$$

$=19.35^{th}$ percentile ของค่าสถิติที่ได้จาก

กระบวนการบูทสแตรปที่มีการเรียงค่าแล้วจากน้อยไปมาก (ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 19.35 ของข้อมูลที่มีการเรียงค่าแล้วจากน้อยไปมาก)

2.3.1.8 พิจารณาค่าอาณาเขตวิกฤตที่ได้เทียบกับค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรป

ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 = $\dfrac{\text{จำนวนครั้ง}\,(SW < BSW^{*})^{*}}{\text{จำนวนรอบภายนอกที่ทำการทดลอง}}$

อำนาจการทดสอบ = $\dfrac{\text{จำนวนครั้ง}\,(SW < BSW^{*})^{**}}{\text{จำนวนรอบภายนอกที่ทำการทดลอง}}$

เมื่อ SW คือ ค่าสถิติแชปิโร-วิลค์ที่ได้จากการคำนวณซึ่งไม่ผ่านกระบวนการบูทสแตรป

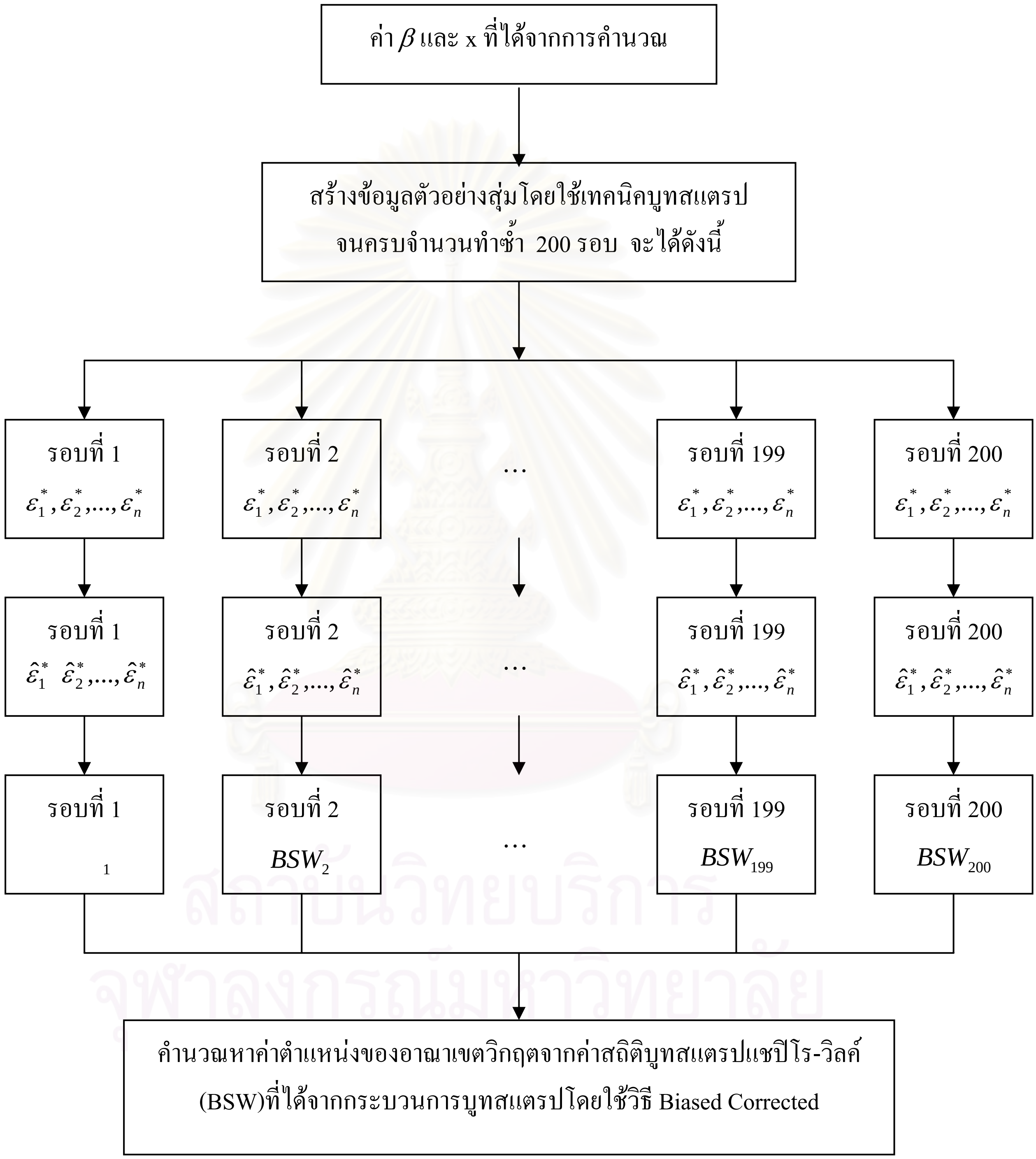
BSW^{*} คือ ค่าวิกฤตที่ได้จากกระบวนการบูทสแตรป

---

[*] เมื่อกำหนดชุดข้อมูลให้สอดคล้องกับสมมติฐานว่าง

[**] เมื่อกำหนดชุดข้อมูลให้สอดคล้องกับสมมติฐานทางเลือก

รูปที่ 2.1 แผนผังขั้นตอนการหาค่าอาณาเขตวิกฤตของตัวสถิติบูทสแตรปแชปิโร-วิลค์ (BSW)ที่ได้จากกระบวนการบูทสแตรป



2.3.2 การหาค่าตัวสถิติแอนเดอร์สัน-ดาร์ลิงและตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง

จาก $\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

เมื่อ $E\left(\underset{\sim}{\varepsilon}\right) = 0$

$E\left(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}'\right) = \sigma^2 I$

$\underset{\sim}{\varepsilon_i}$ ~ iid F ; ไม่ทราบการแจกแจง F

การหาตัวประมาณกระทำตามขั้นตอนต่อไปนี้

2.3.2.1 คำนวณหาค่า $\hat{\underset{\sim}{\beta}}$ โดยวิธีกำลังสองต่ำสุด

ได้ $\hat{\underset{\sim}{\beta}} = \left(X'X\right)^{-1} X' \underset{\sim}{y}$

และ $\hat{\underset{\sim}{y}} = X \hat{\underset{\sim}{\beta}}$

$\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$ ซึ่ง $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$

ให้ F เป็นการแจกแจงเชิงประจักษ์ (empirical distribution) ของ $\hat{\varepsilon}_i$

2.3.2.2 คำนวณหาค่าตัวสถิติแอนเดอร์สัน-ดาร์ลิง โดยนำค่าของ $\hat{\underset{\sim}{\varepsilon}}$ ที่คำนวณได้แทนในสูตรสถิติดังต่อไปนี้

<u>การทดสอบของแอนเดอร์สัน-ดาร์ลิง</u> (Anderson-Darling Method)[8]

วิธีการทดสอบของแอนเดอร์สัน-ดาร์ลิง (Stevens, 1974) เป็นสถิติทดสอบว่าข้อมูลตัวอย่างมาจากข้อมูลตัวแปรสุ่มหรือจากประชากรที่มีการแจกแจงเฉพาะเจาะจงอย่างใดอย่างหนึ่งหรือไม่ สถิติทดสอบนี้ได้รับการพัฒนามาจากสถิติทดสอบของโคโมโกลอฟกับสเมียร์นอฟ (Kolmogorov-Smirnov Test) สำหรับสถิติทดสอบของโคโมโกลอฟกับสเมียร์นอฟนั้นมีข้อได้เปรียบตรงที่ว่าไม่ขึ้นกับการแจกแจงของประชากรหรือข้อมูลตัวแปรสุ่มนั้นๆ (Distribution - Free) ซึ่งตรงข้ามกับสถิติทดสอบของแอนเดอร์สันกับดาร์ลิงที่ไวกับการแจกแจงของประชากรหรือตัวแปรสุ่มนั้นๆ นั่นหมายความว่าถ้าข้อมูลมาจากการแจกแจงอื่นๆ ที่ไม่ใช่การแจกแจงแบบปกติ สถิติของโคโมโกลอฟกับสเมียร์นอฟจะไม่ไวต่อการเบี่ยงเบนไปจากการแจกแจงแบบปกติ ในขณะที่สถิติของแอนเดอร์สันกับดาร์ลิงจะไวกว่า

---

[8] สุพล ดุรงค์วัฒนา, <u>การวางแผนการทดลองเพื่อการวิจัยขั้นสูง,</u> (ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย 2549), หน้า 189-190

และสถิติทดสอบของแอนเดอร์สันกับดาร์ลิง AD กำหนดได้ดังนี้

$$AD = -(n) - S$$

$$S = \sum_{i=1}^{n}\left[\frac{(2i-1)}{n}\right]\left[\ln F\left(\hat{\varepsilon}_{(i)}\right) + \ln\left(1 - F\left(\hat{\varepsilon}_{(n+1-i)}\right)\right)\right]$$

$F(.)$ คือ ค่าฟังก์ชันความถี่สะสมของการแจกแจงแบบปกติ

$\hat{\varepsilon}_{(i)}$ คือ ค่าเศษเหลืออันดับของเศษเหลือจากข้อมูล

n คือ ขนาดตัวอย่าง

2.3.2.3 สุ่ม $\hat{\varepsilon}_i$ แบบใส่คืน (with replacement) ขนาด n

ได้ $\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$

$\varepsilon_i^* \sim$ iid F

2.3.2.4 นำค่า ε_i^* มาพิจารณาโดยรวมไว้ในสมการ

จะได้ $\underset{\sim}{y}^* = X\,\underset{\sim}{\hat{\beta}} + \underset{\sim}{\varepsilon}^* = \underset{\sim}{\hat{y}} + \underset{\sim}{\varepsilon}^*$

เมื่อ $\underset{\sim}{\varepsilon}^* = \varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$

คำนวณหาค่า $\underset{\sim}{\hat{\beta}}^*$ โดยวิธีกำลังสองต่ำสุด เพื่อที่จะนำไปหาค่าของ $\hat{\varepsilon}_i^*$ ต่อไปโดยยึดหลักเกณฑ์เดียวกัน คือทำการหา $\underset{\sim}{\hat{\beta}}^*$ ซึ่งเป็นตัวประมาณกำลังสองต่ำสุดของ $\underset{\sim}{\hat{\beta}}$ ที่ทำให้

$$SSE^* = \sum_{i=1}^{n}\hat{\varepsilon}_i^{*2} = \underset{\sim}{\hat{\varepsilon}}^{*'}\,\underset{\sim}{\hat{\varepsilon}}^* \quad \text{มีค่าต่ำสุด}$$

โดยการหาอนุพันธ์ของ SSE^* เทียบกับ $\underset{\sim}{\beta}$ แล้วกำหนดให้เท่ากับ 0 นั่นคือ

$$\begin{aligned} SSE^* = \underset{\sim}{\hat{\varepsilon}}^{*'}\,\underset{\sim}{\hat{\varepsilon}}^* &= \left(\underset{\sim}{y}^* - X\,\underset{\sim}{\hat{\beta}}^*\right)'\left(\underset{\sim}{y}^* - X\,\underset{\sim}{\hat{\beta}}^*\right) \\ &= \left(\underset{\sim}{y}^{*'} - \underset{\sim}{\hat{\beta}}^{*'}X'\right)'\left(\underset{\sim}{y}^* - X\,\underset{\sim}{\hat{\beta}}^*\right) \\ &= \underset{\sim}{y}^{*'}\,\underset{\sim}{y}^* - \underset{\sim}{y}^{*'}X\,\underset{\sim}{\hat{\beta}}^* - \underset{\sim}{\hat{\beta}}^{*'}X'\,\underset{\sim}{y}^* + \underset{\sim}{\hat{\beta}}^{*'}X'X\,\underset{\sim}{\hat{\beta}}^* \end{aligned}$$

$$= \underset{\sim}{y}^{*'} \underset{\sim}{y}^{*} - 2\underset{\sim}{\hat{\beta}}^{*'} X' \underset{\sim}{y}^{*} + \underset{\sim}{\hat{\beta}}^{*'} X'X \underset{\sim}{\hat{\beta}}^{*}$$

$$\frac{\partial}{\partial \underset{\sim}{\hat{\beta}}^{*}} \text{SSE} = -2X' \underset{\sim}{y}^{*} + 2X'X \underset{\sim}{\hat{\beta}}^{*} = 0$$

$$\therefore X'X \underset{\sim}{\hat{\beta}}^{*} = X' \underset{\sim}{y}^{*}$$

$$\therefore \underset{\sim}{\hat{\beta}}^{*} = (X'X)^{-1} X' \underset{\sim}{y}^{*}$$

จะได้ว่า $\underset{\sim}{\hat{y}}^{*} = X \underset{\sim}{\hat{\beta}}^{*}$

และ $\underset{\sim}{\hat{\varepsilon}}^{*} = \underset{\sim}{y}^{*} - \underset{\sim}{\hat{y}}^{*}$

2.3.2.5 กระทำตามขั้นตอนในข้อ 2.3.2.3-2.3.2.4 ซ้ำ 200 ครั้ง จะได้ $\hat{\varepsilon}^{*1}, \hat{\varepsilon}^{*2}, \ldots, \hat{\varepsilon}^{*B}$ จำนวน 200 ค่า (200 คือ จำนวนรอบของการทำซ้ำในกระบวนการบูทสแตรป)

2.3.2.6 คำนวณหาค่าตัวสถิติบูทสแตรป แอนเดอร์สัน-ดาร์ลิง โดยนำค่าของ $\underset{\sim}{\hat{\varepsilon}}^{*}$ ที่คำนวณได้แทนในสูตรสถิติดังต่อไปนี้

<u>การทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง</u>

$$BAD = -(n) - S$$

$$S = \sum_{i=1}^{n} \left[\frac{(2i-1)}{n} \right] \left[\ln F\left(\hat{\varepsilon}^{*}_{(i)}\right) + \ln\left(1 - F\left(\hat{\varepsilon}^{*}_{(n+1-i)}\right)\right) \right]$$

$F(.)$ คือ ค่าฟังก์ชันความถี่สะสมของการแจกแจงแบบปกติ

$\hat{\varepsilon}^{*}_{(i)}$ คือ ค่าเศษเหลืออันดับของเศษเหลือจากข้อมูลที่ได้จากกระบวนการบูทสแตรป

n คือ ขนาดตัวอย่าง

ซึ่งจะได้ค่าสถิติจำนวน 200 ค่าเช่นกัน แสดงดังตารางต่อไปนี้

จำนวนรอบ	ค่าสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง
1	BAD_1
2	BAD_2
3	BAD_3
.	.
.	.
.	.
i	BAD_i
.	.
.	.
.	.
200	BAD_{200}

2.3.2.7 คำนวณค่าอาณาเขตวิกฤตจากค่าสถิติที่ได้จากกระบวนการบูทสแตรป

สำหรับวิธีการคำนวณค่าอาณาเขตวิกฤตของตัวสถิติที่ได้มาจากกระบวนการบูทสแตรปมีอยู่ด้วยกันหลายวิธีในการวิจัยครั้งนี้ผู้วิจัยได้เลือกวิธีการ Bias-Corrected หรือ BC ซึ่งมีขั้นตอนดังนี้

<u>ขั้นตอนที่ 1</u> คำนวณหาค่าของ Z_0

ซึ่ง Z_0 คือค่าของ Z ที่มีค่าตรงกับสัดส่วนของค่าสถิติที่ได้จากกระบวนการบูทสแตรปที่น้อยกว่าหรือเท่ากับค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรปสามารถหาค่าได้จาก

$$Z_0 = \Phi^{-1}\{pr(BAD \le AD)\}$$

เมื่อ Φ คือฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติ

BAD คือค่าสถิติที่ได้จากกระบวนการบูทสแตรป

AD คือค่าสถิติที่ไม่ผ่านกระบวนการบูทสแตรป

<u>ขั้นตอนที่ 2</u> นำค่าสถิติที่ได้จากกระบวนการบูทสแตรปมาเรียงค่าจากน้อยไปมาก

<u>ขั้นตอนที่ 3</u> คำนวณหาค่าอาณาเขตวิฤต

ค่าอาณาเขตวิกฤตสามารถหาได้จากค่าของตัวสถิติที่ได้จากกระบวนการบูทสแตรปที่อยู่ในตำแหน่งดังนี้

$\left[\left\{\Phi\left(2Z_0 + Z_\alpha\right)\right\}\times 100\right]$ percentile หรือ $\left[\left\{\Phi\left(2Z_0 - Z_\alpha\right)\right\}\times 100\right]$percentile

เช่น กำหนดให้ $\left\{pr\left(BAD \le AD\right)\right\} = 0.65$ ระดับนัยสำคัญคือ 0.05
จะได้ว่า

<u>ขั้นตอนที่ 1</u> คำนวณหาค่าของ Z_0

จาก $Z_0 = \Phi^{-1}\left\{pr\left(BAD \le AD\right)\right\}$

ดังนั้น $Z_0 = \Phi^{-1}(0.65)$

$= 0.39$

<u>ขั้นตอนที่ 2</u> นำค่าสถิติที่ได้จากกระบวนการบูทสแตรปมาเรียงค่าจากน้อยไปมาก

ในที่นี้กำหนดให้จำนวนกระทำซ้ำในกระบวนการบูทสแตรปคือ 200 รอบ ดังนั้นค่าสถิติจากกระบวนการบูทสแตรปที่ได้จากแต่ละสถานการณ์จะมีทั้งหมด 200 ค่า นำค่าสถิติทั้ง 200 ค่ามาเรียงลำดับจากน้อยไปหามาก

<u>ขั้นตอนที่ 3</u> คำนวณหาค่าอาณาเขตวิกฤต

ตำแหน่งของอาณาเขตวิกฤตคือ $\left[\left\{\Phi\left(2Z_0 + Z_\alpha\right)\right\}\times 100\right]$percentile

หรือ $\left[\left\{\Phi\left(2Z_0 - Z_\alpha\right)\right\}\times 100\right]$percentile

ในที่นี้สมมติให้ปฏิเสธสมมติฐานหลักเมื่อ $AD_{cal} > AD_{table}$

โดยที่ค่าของ AD_{cal} คือ ค่าสถิติแอนเดอร์สัน-ดาร์ลิงที่ได้จากการคำนวณซึ่งไม่ผ่านกระบวนการบูทสแตรป

AD_{table} คือ ค่าวิกฤตที่ได้จากตารางสถิติของแอนเดอร์สัน-ดาร์ลิง

ดังนั้นตำแหน่งของอาณาเขตวิกฤตคือ $\left[\left\{\Phi\left(2Z_0 + Z_\alpha\right)\right\}\times 100\right]$percentile

$= \left[\left\{\Phi\left(2(0.39) + 1.645\right)\right\}\times 100\right]$percentile

$= \Phi\left(2.425\right)\times 100$

$= 99.23^{th}$ percentile ของค่าสถิติที่ได้จากกระบวนการบูทสแตรปที่มีการเรียงค่าแล้วจากน้อยไปมาก (ตำแหน่ง

เปอร์เซ็นต์ไทล์ที่ 99.23 ของข้อมูลที่มีการเรียงค่าแล้วจากน้อยไปมาก)

2.3.2.8 พิจารณาค่าอาณาเขตวิกฤตที่ได้เทียบกับค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรป

$$\text{ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1} = \frac{\text{จำนวนครั้ง}\,(AD > BAD^{*})^{*}}{\text{จำนวนรอบภายนอกที่ทำการทดลอง}}$$

$$\text{อำนาจการทดสอบ} = \frac{\text{จำนวนครั้ง}\,(AD > BAD^{*})^{**}}{\text{จำนวนรอบภายนอกที่ทำการทดลอง}}$$

เมื่อ AD คือ ค่าสถิติแอนเดอร์สัน-ดาร์ลิงที่ได้จากการคำนวณซึ่งไม่ผ่านกระบวนการบูทสแตรป

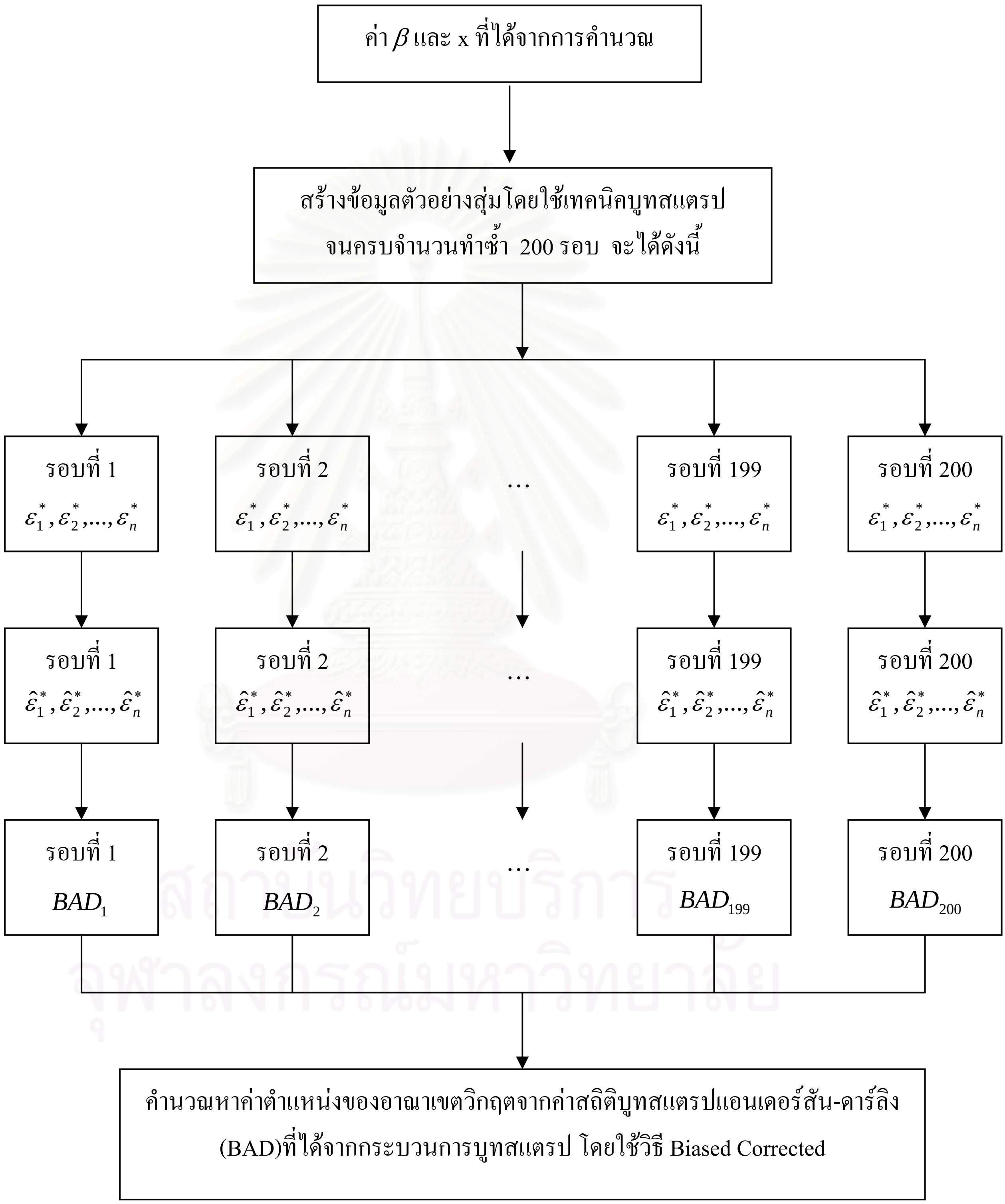
BAD^{*} คือ ค่าวิกฤตที่ได้จากกระบวนการบูทสแตรป

---

* เมื่อกำหนดชุดข้อมูลให้สอดคล้องกับสมมติฐานว่าง

** เมื่อกำหนดชุดข้อมูลให้สอดคล้องกับสมมติฐานทางเลือก

รูปที่ 2.2 แผนผังขั้นตอนการหาค่าอาณาเขตวิกฤตของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง (BAD) ที่ได้จากกระบวนการบูทสแตรป



ในที่นี้จะเห็นว่า ε_i^* เป็นตัวอย่างสุ่มที่ได้โดยวิธีบูทสแตรป (Bootstrap Sample) ซึ่งมีคุณสมบัติดังนี้คือ

1. $E\left(\varepsilon_i^*\right)=0$; i = 1,2,…,n
2. $\text{cov}\left(\varepsilon_i^*\varepsilon_j^*\right)=0$
3. $V\left(\varepsilon_i^*\right)=\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$

<u>พิสูจน์</u> $E\left(\varepsilon_i^*\right)=0$

เมื่อ ε_i^* เป็นการสุ่มตัวอย่างแบบใส่คืนจาก $\hat{\varepsilon}_1,\hat{\varepsilon}_2,...,\hat{\varepsilon}_n$

ให้ b_i = จำนวนครั้งที่หน่วย i ตกอยู่ในตัวอย่าง

b_i Binomial $(n, p=\frac{1}{n})$

$$E(b_i)=\frac{n}{n}=1$$

$$V(b_i)=n.\frac{1}{n}(1-\frac{1}{n})=\frac{n-1}{n}$$

$$\text{cov}(b_i,b_j)=-n.p_i.p_j=-n.\frac{1}{n}.\frac{1}{n}=-\frac{1}{n}$$

$$\begin{aligned}\mathrm{E}(\varepsilon_i^*) &= E\left\{E(\varepsilon_i^*/\hat{\varepsilon}_i,...,\hat{\varepsilon}_n)\right\}\\ &= E\left\{\sum_{i=1}^{n}p_i\varepsilon_i^*\right\}\\ &= E\left\{\sum_{i=1}^{n}\frac{1}{n}\varepsilon_i^*\right\}\\ &= E\left\{\frac{1}{n}\sum_{i=1}^{n}b_i\hat{\varepsilon}_i\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_iE(b_i)\end{aligned}$$

$$E\left(\varepsilon_i^*\right)=0 \qquad \sum_{i=1}^{n}\hat{\varepsilon}_i=0$$

<u>พิสูจน์</u> $\text{cov}\left(\varepsilon_i^*\varepsilon_j^*\right)=0$

$$\text{cov}\left(\varepsilon_i^*\varepsilon_j^*\right)=0$$

เพราะว่า ε_i^* และ ε_j^* เป็นอิสระแก่กัน เนื่องจากการสุ่มแบบใส่คืนจาก $\hat{\varepsilon}_1,\hat{\varepsilon}_2,...,\hat{\varepsilon}_n$

<u>พิสูจน์</u> $V\left(\varepsilon_i^*\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$

จาก $V\left(\varepsilon_i^*\right) = E(\varepsilon_i^{*2}) - \left\{E(\varepsilon_i^*)\right\}^2$

$$= E(\varepsilon_i^{*2})$$

$$= E\left\{\sum_{i=1}^{n} p_i \varepsilon_i^*\right\}$$

$$= E\left\{\sum_{i=1}^{n} \frac{1}{n} \varepsilon_i^{*2}\right\}$$

$$= E\left\{\sum_{i=1}^{n} \frac{1}{n} b_i \hat{\varepsilon}_i^2\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 E(b_i)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \ .1$$

$$= \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติ ของความคลาดเคลื่อนสุ่มในตัวแบบความถดถอยเชิงเส้นแบบพหุคูณ โดยจะศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติ ซึ่งจะสร้าง(generate)ข้อมูลให้ความคลาดเคลื่อน (ε) มีการแจกแจงแบบปกติ (Normal Distribution) แบบโลจิสติค (Logistic Distribution) และแบบดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution) ทั้งนี้เนื่องจากการแจกแจงแบบต่างๆ ดังกล่าว ยกเว้นการแจกแจงแบบปกติเป็นการแจกแจงที่มีลักษณะการกระจายไปทางหางมากหรือมีหางยาวกว่าปกติ ซึ่งเป็นลักษณะที่สนใจศึกษา และที่สนใจศึกษากรณีที่ ε มีการแจกแจงแบบปกติด้วย เพราะว่าการประมาณค่าพารามิเตอร์โดยวิธีบูทสแตรปไม่จำเป็นต้องมีข้อกำหนดเกี่ยวกับการแจกแจงของ ε คือ ε อาจมีลักษณะการแจกแจงแบบใดก็ได้ที่ไม่ทราบ ซึ่งอาจเป็นได้ทั้งแบบปกติและไม่ปกติ

ทั้งนี้เทคนิคที่ใช้ในการหาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ คือ วิธีมอนติคาร์โล ซึ่งเป็นเทคนิคอย่างหนึ่งที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์ โดยเฉพาะในกรณีที่ไม่สามารถหาคำตอบโดยวิธีทางทฤษฎีได้

เนื่องจากวิธีมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยครั้งนี้ ดังนั้นในตอนแรกของบทนี้จะกล่าวถึงวิธีมอนติคาร์โลก่อน แล้วจึงแสดงรายละเอียดของแผนการดำเนินการวิจัยและขั้นตอนในการดำเนินการวิจัยตามลำดับ ซึ่งรายละเอียดต่างๆ เป็นดังนี้

3.1 การผลิตเลขสุ่มจากรูปแบบการแจกแจง[9]

การวิจัยเรื่องนี้มีลักษณะเป็นการวิจัยเชิงทดลอง ซึ่งจำลองขึ้นด้วยการทำงานของเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น เพื่อหาผลสรุปในการเปรียบเทียบความสามารถ

---

[9] วโรภาส ประดิษฐกิจจรชัย “การเปรียบเทียบการทดสอบอัตราส่วนความควรจะเป็นแบบปกติและการทดสอบอัตราส่วนความควรจะเป็นแบบมอนติคาร์โลสำหรับการวิเคราะห์ความถดถอยเชิงพหุ”,(วิทยานิพนธ์ปริญญามหาบัณฑิตภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย 2547), หน้า 22-23

ในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบการแจกแจงแบบปกติของตัวสถิตินอนพาราเมตริกซ์ 2 ตัว

เทคนิคมอนติคาร์โลซิมูเลชัน มีหลักที่สำคัญคือ การใช้เลขสุ่ม (Random Number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา เทคนิคนี้ถูกคิดขึ้นครั้งแรกในกลางศตวรรษที่ 19 และมีผู้นำไปใช้อย่างแพร่หลายต่อมาเพื่อนำมาตอบปัญหาต่างๆที่ยังขัดแย้งกันอยู่ ในปัจจุบันวิธีมอนติคาร์โลเป็นวิธีที่นิยมใช้กันมาก โดยเฉพาะอย่างยิ่งวิธีนี้จะถูกนำมาใช้เมื่อคิดว่าเป็นวิธีที่ดีที่สุดหรือเป็นวิธีเดียวที่จะใช้ศึกษาได้ เมื่อการพัฒนาทางด้านวิชาการมีมากขึ้น ปัญหาต่างๆที่เกิดขึ้นในทางปฏิบัติ ซึ่งไม่สามารถหาคำตอบได้โดยวิธีทางทฤษฎีก็จะยิ่งมีมากขึ้น ทำให้เทคนิคมอนติคาร์โลมีความจำเป็นมากขึ้น

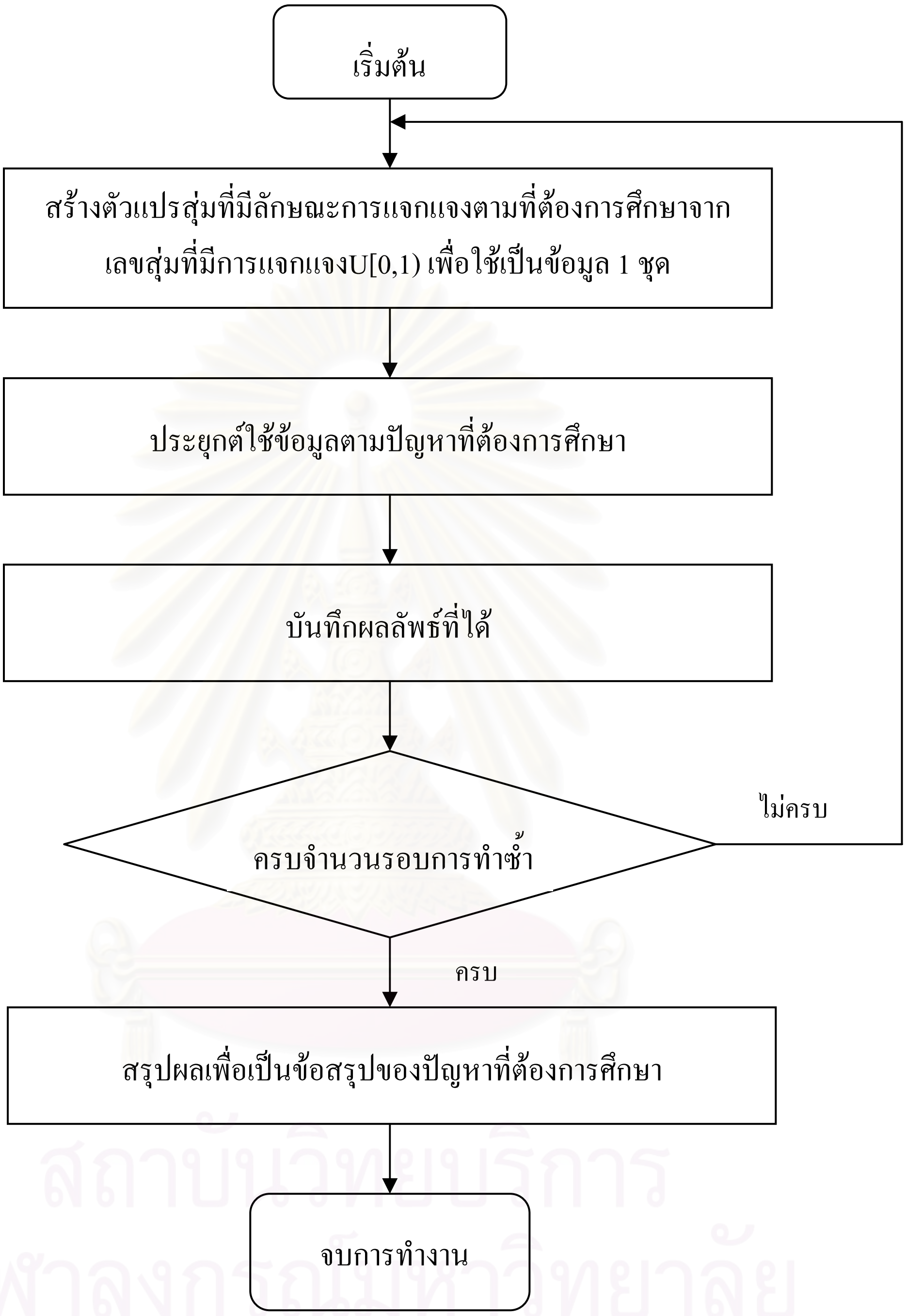
ขั้นตอนของวิธีมอนติคาร์โลแบ่งได้ 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างเลขสุ่ม (Generate Random Number) การใช้เลขสุ่มเป็นสิ่งที่สำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะหลักการของวิธีมอนติคาร์โลนั้นจะใช้เลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์มในช่วง [0,1) สำหรับการสร้างเลขสุ่มมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่เกิดขึ้นจะต้องมีการแจกแจงแบบยูนิฟอร์มในช่วง [0,1) และเป็นอิสระแก่กัน

ขั้นตอนที่ 2 ประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับปัญหาที่ต้องการศึกษา บางปัญหาอาจจะใช้เลขสุ่มได้โดยตรง ในขณะที่บางปัญหาอาจต้องใช้ขั้นตอนอื่นอีกหลายขั้นตอน โดยที่มีการใช้ตัวเลขสุ่มในบางขั้นตอนเท่านั้น

ขั้นตอนที่ 3 การทดลองกระทำ เมื่อประยุกต์ปัญหาที่สนใจให้ใช้เลขสุ่มได้แล้วขั้นต่อไปก็คือ การทดลอง โดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆกัน (Replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

รูปที่ 3.1 แสดงผังงานสำหรับขั้นตอนของวิธีมอนติคาร์โล



เลขสุ่มที่ผลิตได้จากเทคนิคมอนติคาร์โลมีคุณสมบัติดังนี้

1. ตัวเลขที่ได้มีการกระจายความน่าจะเป็นแบบสม่ำเสมอและเป็นอิสระซึ่งกันและกัน
2. อนุกรมของตัวเลขที่ได้สามารถสร้างซ้ำเดิมได้
3. อนุกรมของตัวเลขไม่ซ้ำเดิมในช่วงที่ต้องการใช้ตัวเลขแบบสุ่ม
4. ใช้ระยะเวลาสั้นๆ ในการสร้างตัวเลขแบบสุ่ม
5. ใช้หน่วยความจำของคอมพิวเตอร์น้อย

3.2 แผนการดำเนินการวิจัย

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดสถานการณ์ต่างๆ ที่จะทำการศึกษาดังนี้

3.2.1 ตัวแบบความถดถอยเป็นตัวแบบที่อยู่ภายใต้ข้อตกลงเบื้องต้น

3.2.2 จำนวนตัวแปรอิสระ(p)ที่ใช้ในการศึกษาเท่ากับ 3 และ 5

3.2.3 กำหนดจำนวนตัวอย่าง(n)ที่ใช้ในการศึกษาเท่ากับ 10 20 30 และ 50

3.2.4 ลักษณะการแจกแจงของค่าความคลาดเคลื่อนที่จะศึกษามีดังนี้

3.2.4.1 การแจกแจงแบบปกติ

3.2.4.2 การแจกแจงแบบโลจิสติค

3.2.4.3 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

3.2.5 กำหนดให้ข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 1 5 และ 10

3.2.6 ระดับนัยสำคัญของการทดสอบ (α) ที่ศึกษาคือ 0.01 0.05 และ 0.10

3.2.7 ในการวิจัยครั้งนี้สร้างแบบการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo simulation) เขียนด้วยโปรแกรม S-PLUS 2000

3.2.8 การจำลองในแต่ละสถานการณ์ของการทดลองการทำซ้ำ 600 รอบ

3.2.9 การสร้างตัวอย่างสุ่มในวิธีการทดสอบตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ จะกระทำซ้ำ 200 รอบ

3.3 ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัยมีดังนี้

3.3.1 สร้างข้อมูลของตัวแปรอิสระ

3.3.2 สร้างข้อมูลของความคลาดเคลื่อนให้มีลักษณะตามที่ต้องการศึกษา

3.3.3 สร้างข้อมูลตัวแปรตาม

3.3.4 คำนวณค่าสถิติทดสอบทั้ง 2 วิธี

3.3.5 หาค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

3.3.6 เปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 กับเกณฑ์ที่กำหนดไว้และอำนาจการทดสอบระหว่างตัวสถิติทั้ง 2 วิธี

ซึ่งรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

3.3.1 สร้างข้อมูลของตัวแปรอิสระ

สร้างข้อมูลของตัวแปรอิสระ (X) โดยกำหนดให้ข้อมูลของแต่ละตัวแปรอิสระมีการแจกแจงแบบปกติโดยค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และไม่มีความสัมพันธ์กันเองระหว่างตัวแปรอิสระ

3.3.2 สร้างข้อมูลของความคลาดเคลื่อนให้มีลักษณะตามที่ต้องการศึกษา

ในการวิจัยครั้งนี้ ได้ทำการสร้างการแจกแจงของความคลาดเคลื่อนแบบปกติ โลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล สำหรับโปรแกรม S-PLUS 2000 จะใช้ฟังก์ชันดังนี้

3.3.2.1 ความคลาดเคลื่อนมีการแจกแจงแบบปกติใช้ฟังก์ชัน rnorm (n, μ ,sd) โดย n แทนขนาดตัวอย่าง μ แทนค่าเฉลี่ยที่กำหนดให้เป็น 0 และ sd แทนส่วนเบี่ยงเบนมาตรฐานที่กำหนดให้เป็น 1 5 และ 10

3.3.2.2 ความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติคใช้ฟังก์ชันrlogis (n, α , β) โดย n แทนขนาดตัวอย่าง α แทนค่าเฉลี่ยที่กำหนดให้เป็น 0 และ β เป็นค่าพารามิเตอร์ที่กำหนดขึ้นเพื่อให้ได้ส่วนเบี่ยงเบนมาตรฐานเป็น 1 5 และ 10 โดยที่การแจกแจงแบบโลจิสติคมีค่าความแปรปรวนเท่ากับ $\frac{1}{3}\pi^2\beta^2$ ดังนั้นกรณีที่

- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 1 จะได้ว่า β มีค่าเท่ากับ 0.5511
- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 5 จะได้ว่า β มีค่าเท่ากับ 2.7756
- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 10 จะได้ว่า β มีค่าเท่ากับ 5.511

3.3.2.3 ความคลาดเคลื่อนมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล จะไม่มีคำสั่งสำเร็จรูปที่เรียกใช้ได้เลย ซึ่งในขั้นแรกต้องสร้างข้อมูลให้มีการแจกแจงแบบยูนิฟอร์มก่อน จากนั้นกำหนดค่า α และ β เป็นค่าพารามิเตอร์ที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 5 และ 10 โดยที่การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียลมีค่าความแปรปรวนเท่ากับ $2\beta^2$ ดังนั้นกรณีที่

- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 1 จะได้ว่า β มีค่าเท่ากับ 0.7071
- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 5 จะได้ว่า β มีค่าเท่ากับ 3.5355
- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 10 จะได้ว่า β มีค่าเท่ากับ 7.071

3.3.3 สร้างข้อมูลตัวแปรตาม

จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ กำหนดให้ $\underset{\sim}{\beta}$ เป็นค่าคงที่โดยในกรณีที่ตัวแปรอิสระเท่ากับ 3 กำหนดให้ $\underset{\sim}{\beta} = (1,1,1,1)'$ และในกรณีที่ตัวแปรอิสระเท่ากับ 5 กำหนดให้ $\underset{\sim}{\beta} = (1,1,1,1,1,1)'$ ซึ่งค่าความคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ มีการแจกแจงตามที่กำหนดไว้ในขอบเขตของการวิจัย

3.3.4 คำนวณค่าสถิติทดสอบทั้ง 2 วิธี

กำหนดจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน แล้วทำการสร้างชุดข้อมูลสุ่มตามลักษณะการแจกแจงของความคลาดเคลื่อนที่กำหนด และนำข้อมูลที่ได้ไปคำนวณค่าต่าง ๆ ตามสูตรการทดสอบทั้ง 2 วิธี คือ

3.3.4.1 สถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง

3.3.4.2 สถิติบูทสแตรปแชปิโร-วิลค์

3.3.5 การหาค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

คำนวณค่าสถิติทดสอบและคำนวณหาค่าอาณาเขตวิกฤตทั้ง 2 วิธี และเปรียบเทียบค่าอาณาเขตวิกฤตกับค่าสถิติที่ไม่ได้ผ่านกระบวนการบูทสแตรป จากนั้นทำการหาความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ สามารถสรุปเป็นขั้นตอนได้ดังนี้

3.3.5.1 สร้างการแจกแจงของความคลาดเคลื่อนให้มีการแจกแจงแบบปกติสำหรับการหาค่าความคลาดเคลื่อนประเภทที่ 1 และให้ความคลาดเคลื่อนมีการแจกแจงแบบโลจิติคและดับเบิ้ลเอ็กซ์โปเนนเชียลเมื่อพิจารณาหาอำนาจการทดสอบ

3.3.5.2 คำนวณค่าความคลาดเคลื่อนประเภทที่ 1 เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ และคำนวณค่าอำนาจการทดสอบเมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิติคและดับเบิ้ลเอ็กซ์โปเนนเชียล

3.3.5.3 เปลี่ยนค่าส่วนเบี่ยงเบนมาตรฐาน จนกระทั่งครบทุกสถานการณ์ โดยในแต่ละสถานการณ์จะกระทำซ้ำกัน 600 รอบ

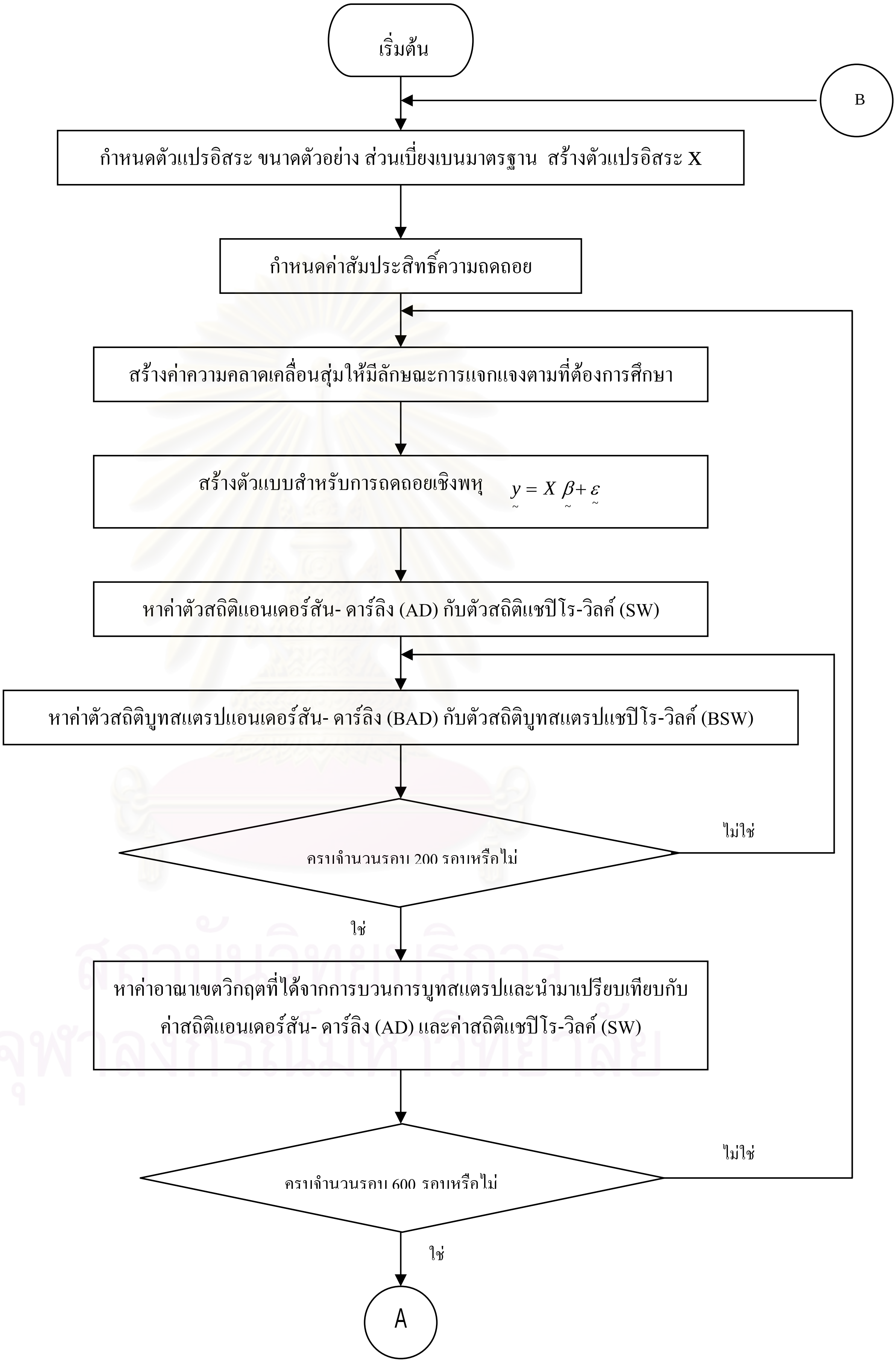
3.3.6 เปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

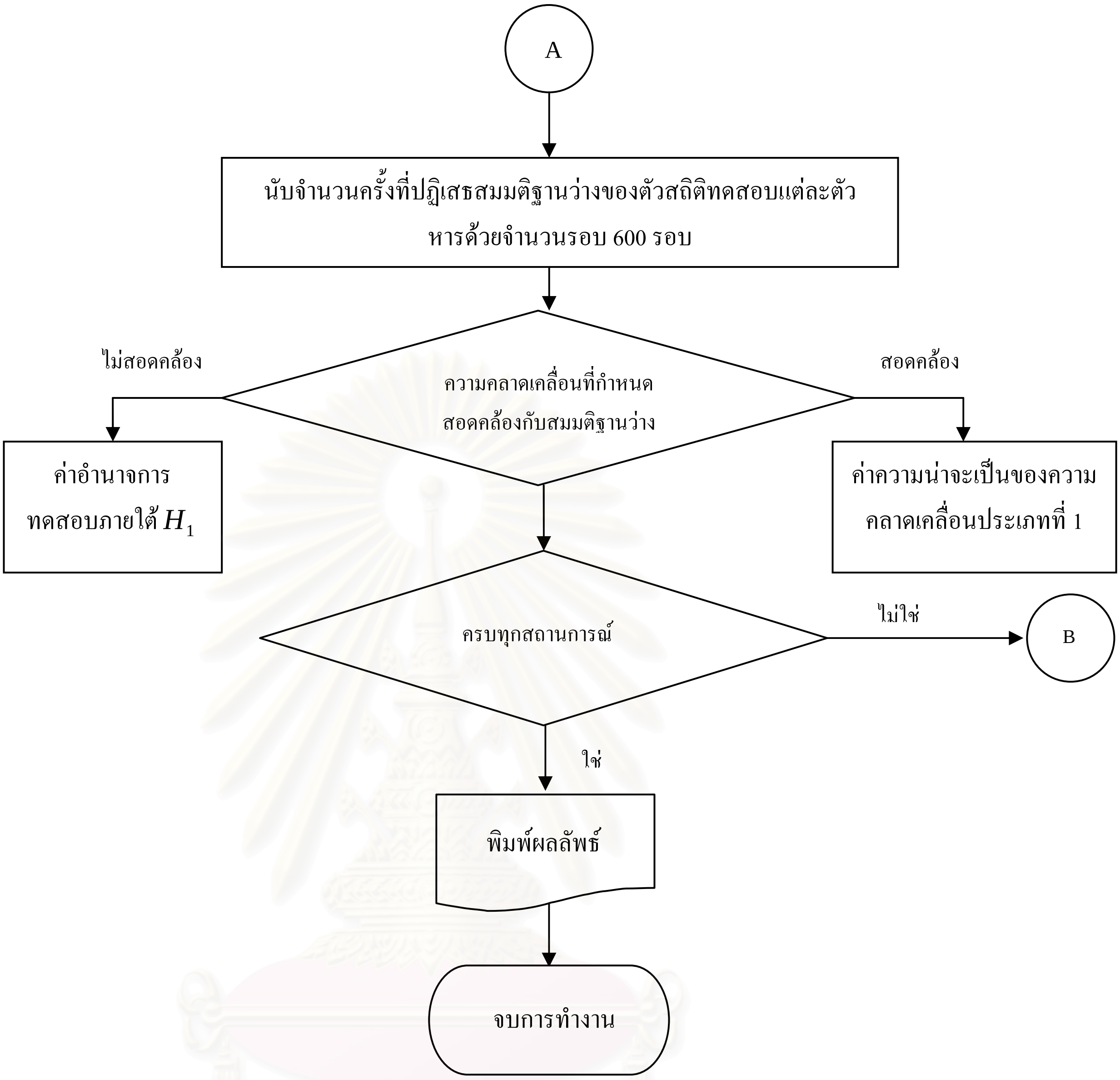
เปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 ว่าตัวสถิติทดสอบวิธีใดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้มากกว่ากันและให้อำนาจการทดสอบสูงสุด ก็จะเป็นตัวสถิติที่เหมาะสม

3.4 แผนผังขั้นตอนการทำงานของโปรแกรม

ลักษณะการทำงานของโปรแกรมในการวิจัยครั้งนี้ใช้ S-PLUS 2000 ในการประมวลผลข้อมูล โดยมีขั้นตอนการทำงานดังรูป 3.2 ดังนี้

รูปที่ 3.2 แสดงผังงานสำหรับขั้นตอนการวิจัย




A
นับจำนวนครั้งที่ปฏิเสธสมมติฐานว่างของตัวสถิติทดสอบแต่ละตัว
หารด้วยจำนวนรอบ 600 รอบ
ไม่สอดคล้อง
ความคลาดเคลื่อนที่กำหนด
สอดคล้องกับสมมติฐานว่าง
สอดคล้อง
ค่าอำนาจการ
ทดสอบภายใต้ H_1
ค่าความน่าจะเป็นของความ
คลาดเคลื่อนประเภทที่ 1
ครบทุกสถานการณ์
ไม่ใช่
B
ใช่
พิมพ์ผลลัพธ์
จบการทำงาน

3.5 ขั้นตอนในการหาค่าของตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับค่าตัวสถิติบูทสแตรปแชปิโร-วิลค์

ในการหาค่าตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ มีขั้นตอนดังต่อไปนี้

3.5.1 คำนวณหาค่าตัวประมาณกำลังสองน้อยสุดจากสูตร $\hat{\underset{\sim}{\beta}} = \left(X'X\right)^{-1} X' \underset{\sim}{y}$

3.5.2 หาค่าประมาณของความคลาดเคลื่อนจาก $\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$

3.5.3 จากค่า $\hat{\varepsilon}_i$ ทำการสุ่มตัวอย่างขนาด n แบบใส่คืน (with replacement) จะได้ค่า $\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$ เป็นสมาชิกเก็บไว้ในเวกเตอร์ $\underset{\sim}{\varepsilon}^*$ เมื่อ ε_i^* คือตัวอย่างสุ่มที่ได้ตัวที่ i จาก $\hat{\varepsilon}_i^*$

3.5.4 นำค่า ε_i^* มารวมไว้ในสมการ $\underset{\sim}{y}^* = X \hat{\underset{\sim}{\beta}} + \underset{\sim}{\varepsilon}^*$

จะได้ $\underset{\sim}{y}^* = \hat{\underset{\sim}{y}} + \underset{\sim}{\varepsilon}^*$

3.5.5 คำนวณหาค่า $\hat{\underset{\sim}{\beta}}^*$ โดยวิธีกำลังสองต่ำสุด เพื่อที่จะนำไปหาค่าของ $\hat{\varepsilon}_i^*$ ต่อไปจาก

$$\hat{\underset{\sim}{\beta}}^* = \left(X'X\right)^{-1} X' \underset{\sim}{y}^*$$

จะได้ว่า $\hat{\underset{\sim}{y}}^* = X \hat{\underset{\sim}{\beta}}^*$

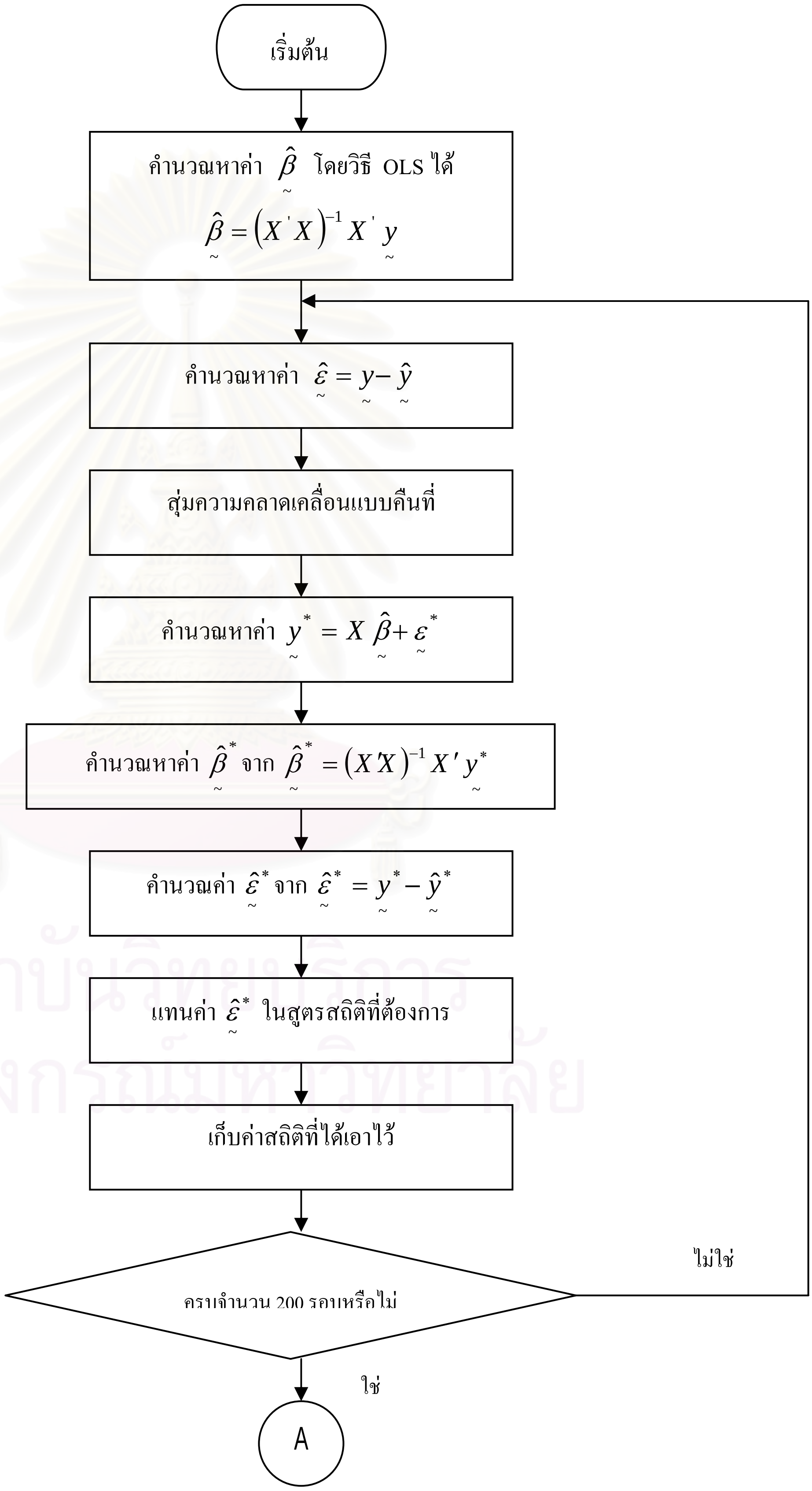
และ $\hat{\underset{\sim}{\varepsilon}}^* = \underset{\sim}{y}^* - \hat{\underset{\sim}{y}}^*$

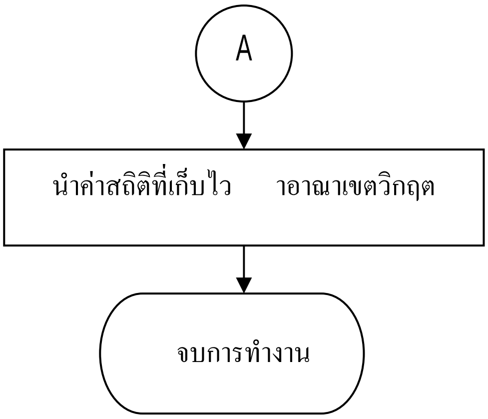
3.5.6 กลับไปทำขั้นตอนที่ 3.5.2 ใหม่ซ้ำ 200 ครั้ง จะได้ $\hat{\varepsilon}^{*1}, \hat{\varepsilon}^{*2}, ..., \hat{\varepsilon}^{*200}$ นำไปแทนค่าในสูตรสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ จะได้ค่าของตัวสถิติทั้งหมดอย่างละ 200 ค่า (จำนวนรอบของการประมาณด้วยวิธีบูทสแตรป)

3.5.7 นำค่าสถิติทั้ง 200 ค่าของแต่ละตัวสถิติมาหาอาณาเขตวิกฤต

จากขั้นตอนดังกล่าวสามารถสรุปอยู่ในรูปของผังงานได้ดังนี้

รูปที่ 3.3 แสดงผังงานขั้นตอนในการหาค่าตัวสถิติบูทสแตรปแอนเดอร์สัน- ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์





3.6 โปรแกรมที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้โปรแกรมที่ใช้ในการวิจัยทั้งหมดเขียนด้วยโปรแกรม S-PLUS 2000 ซึ่งในแต่ละสถานการณ์จะมีลักษณะการทำงานของโปรแกรมเหมือนกัน โดยรายละเอียดของโปรแกรม คำอธิบายแต่ละคำสั่งและขั้นตอนในการทำงานของโปรแกรมแต่ละวิธีจะแสดงไว้ในภาคผนวก ข

บทที่ 4

ผลการวิจัย

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบตัวสถิติที่ใช้สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่ม สำหรับการวิเคราะห์ความถดถอยเชิงพหุด้วยสถิติทดสอบ 2 วิธี คือ ตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ โดยศึกษาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบ ภายใต้การแจกแจงของความคลาดเคลื่อน (ε) แบบปกติ โลจิสติค และ ดับเบิ้ลเอ็กซ์โปเนนเชียล ในสถานการณ์ต่างๆ คือ ทำการศึกษาในสถานการณ์ที่จำนวนตัวแปรอิสระเท่ากับ 3 และ 5 ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 10 20 30 และ 50 ที่ระดับนัยสำคัญ 0.01 0.05 และ 0.10 ซึ่งผู้วิจัยได้ทำการจำลองข้อมูล ให้มีความคลาดเคลื่อนมาตรฐาน (SD) 3 ระดับ คือ 1 5 และ 10 โดยวิธีการจำลองข้อมูลนั้นจะอาศัยเทคนิคมอนติคาร์โลซีมูเลชัน (Monte Carlo simulation) จะกระทำซ้ำในแต่ละสถานการณ์จำนวน 600 รอบ และสำหรับการทดสอบตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ จะกระทำการสร้างตัวอย่างสุ่มจำนวน 200 รอบ

ในการศึกษาวิจัยครั้งนี้จะพิจารณาความเหมาะสมของตัวสถิติที่ใช้ในการทดสอบ โดยพิจารณาจากการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ซึ่งได้จากการนับจำนวนครั้งในการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นเป็นจริงต่อชุดข้อมูลทั้งหมด และการเปรียบเทียบอำนาจการทดสอบของ ตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ ได้จากการนับจำนวนครั้งในการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นเป็นเท็จต่อชุดข้อมูลทั้งหมด โดยสมมติฐานในการวิจัยครั้งนี้คือ

H_0 : ความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

และเพื่อความสะดวกในการนำเสนอผลการวิจัยในครั้งนี้ จึงใช้สัญลักษณ์ต่อไปนี้แสดงในตาราง โดยแทนความหมายต่างๆ ดังนี้

p แทน จำนวนตัวแปรอิสระ

n แทน ขนาดตัวอย่าง

SD แทน ค่าส่วนเบี่ยงเบนมาตรฐาน

BSW แทน ตัวสถิติบูทสแตรปแชปิโร-วิลค์

BAD แทน ตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง

α แทน ระดับนัยสำคัญ

การนำเสนอผลการวิจัยของการเปรียบเทียบตัวสถิติที่ใช้ในการทดสอบ 2 วิธี จะแบ่งการนำเสนอผลการวิจัยออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 4.1 ผลการวิจัยของการเปรียบเทียบตัวสถิติที่ใช้ในการทดสอบโดยการพิจารณาจากค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 แสดงดังตาราง 4.1-4.12

ส่วนที่ 4.2 ผลการวิจัยของการเปรียบเทียบตัวสถิติที่ใช้ในการทดสอบโดยการพิจารณาจากค่าอำนาจการทดสอบ แสดงดังตาราง 4.13-4.18 และรูปที่ 4.1-4.12 ตามลำดับ

การนำเสนอผลการวิจัยในรายละเอียดของค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ในแต่ละกรณี ได้ทำการพิจารณาความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 โดยใช้เกณฑ์ของ Bradley ซึ่งจำแนกได้ดังนี้

เมื่อทดสอบที่ระดับนัยสำคัญ 0.01 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.005,0.015]

เมื่อทดสอบที่ระดับนัยสำคัญ 0.05 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.025,0.075]

เมื่อทดสอบที่ระดับนัยสำคัญ 0.10 จะสามารถควบคุมความผิดพลาดประเภทที่ 1 ของตัวสถิติได้เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง [0.050,0.150]

จากผลการวิจัยพบว่ากรณีศึกษาส่วนใหญ่สามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ และต่อไปนี้จะกล่าวถึงรายละเอียดในส่วนต่างๆ ของผลการวิจัยต่อไป

ส่วนที่ 4.1ผลการวิจัยของการเปรียบเทียบตัวสถิติที่ใช้ในการทดสอบ โดยการพิจารณาจากค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

ผลการวิจัยพบว่า กรณีศึกษาส่วนใหญ่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

ตาราง 4.1 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.01$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.01$	
			BSW	BAD
P = 3	1	10	0.0105	0.0085
		20	0.0100	0.0067
		30	0.0075	0.0055
		50	0.0050	0.0050
	5	10	0.0117	0.0110
		20	0.0101	0.0102
		30	0.0095	0.0057
		50	0.0073	0.0055
	10	10	0.0133	0.0147
		20	0.0110	0.0115
		30	0.0098	0.0100
		50	0.0083	0.0095

จากตารางที่ 4.1 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.01$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

2. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.2 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.05$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.05$	
			BSW	BAD
P = 3	1	10	0.0267	0.0583
		20	0.0256	0.0317
		30	0.0252	0.0269
		50	0.0250	0.0250
	5	10	0.0289	0.0667
		20	0.0278	0.0633
		30	0.0265	0.0367
		50	0.0250	0.0333
	10	10	0.0433	0.0717
		20	0.0350	0.0707
		30	0.0300	0.0397
		50	0.0257	0.0360

จากตารางที่ 4.2 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.05$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

2.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.3 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.10$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.10$	
			BSW	BAD
P = 3	1	10	0.0817	0.0900
		20	0.0700	0.0517
		30	0.0637	0.0515
		50	0.0503	0.0500
	5	10	0.1000	0.1250
		20	0.0950	0.1200
		30	0.0883	0.0517
		50	0.0800	0.0504
	10	10	0.1250	0.1300
		20	0.0967	0.1255
		30	0.0883	0.1100
		50	0.0800	0.0983

จากตารางที่ 4.3 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.10$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

2. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.4 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.01$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.01$	
			BSW	BAD
P = 5	1	10	0.0115	0.0143
		20	0.0102	0.0083
		30	0.0100	0.0083
		50	0.0067	0.0067
	5	10	0.0150	0.0185*
		20	0.0147	0.0155*
		30	0.0132	0.0145
		50	0.0110	0.0100
	10	10	0.0165*	0.0367*
		20	0.0150	0.0250*
		30	0.0150	0.0150
		50	0.0133	0.0125

* หมายถึงไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้

จากตารางที่ 4.4 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.01$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

2.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ตัวสถิติที่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้เมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 คือ BAD ส่วนขนาดตัวอย่างระดับอื่นๆ สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ตัวสถิติที่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้เมื่อขนาดตัวอย่างเท่ากับ 10 คือ BSW และ BAD และเมื่อขนาดตัวอย่างเท่ากับ 20 คือ BAD ส่วนขนาดตัวอย่างระดับอื่นๆ สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.5 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.05$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.05$	
			BSW	BAD
P = 5	1	10	0.075	0.0750
		20	0.0617	0.0683
		30	0.0601	0.0417
		50	0.0383	0.0333
	5	10	0.0750	0.1250*
		20	0.0633	0.0697
		30	0.0615	0.0633
		50	0.0605	0.0417
	10	10	0.0750	0.1317*
		20	0.0713	0.0767
		30	0.0683	0.0733
		50	0.0613	0.0700

* หมายถึงไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้

จากตารางที่ 4.5 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.05$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

2.ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ตัวสถิติที่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้เมื่อขนาดตัวอย่างเท่ากับ 10 คือ BAD ส่วนขนาดตัวอย่างระดับอื่นๆ สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ตัวสถิติที่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้เมื่อขนาดตัวอย่างเท่ากับ 10 คือ BAD ส่วนขนาดตัวอย่างระดับอื่นๆ สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.6 แสดงการเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.10$

จำนวนตัวแปรอิสระ (p)	SD	ขนาดตัวอย่าง	$\alpha = 0.10$	
			BSW	BAD
P = 5	1	10	0.1000	0.1305
		20	0.0900	0.0783
		30	0.0875	0.0777
		50	0.0535	0.0700
	5	10	0.1367	0.1413
		20	0.1333	0.1270
		30	0.1283	0.0833
		50	0.1165	0.0767
	10	10	0.1380	0.1500
		20	0.1344	0.1500
		30	0.1283	0.1095
		50	0.1175	0.0993

จากตารางที่ 4.6 ผลการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงและตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อจำนวนตัวแปรอิสระที่ใช้ทดสอบคือ p = 5 และระดับนัยสำคัญ $\alpha = 0.10$ จำแนกตามส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

1. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

3. ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

ทั้งตัวสถิติ BSW และ BAD สามารถสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมด

ตาราง 4.7 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

α \ **n**	**10**	**20**	**30**	**50**
0.01	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.05	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ตาราง 4.8 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

α \ n	10	20	30	50
0.01	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.05	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ตาราง 4.9 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

n / α	**10**	**20**	**30**	**50**
0.01	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.05	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ตาราง 4.10 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

n / α	10	20	30	50
0.01	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.05	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ตาราง 4.11 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

α \ n	10	20	30	50
0.01	BSW	BSW	BSW,BAD	BSW,BAD
0.05	BSW	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ตาราง 4.12 ตัวสถิติที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้โดยจำแนกตามขนาดตัวอย่างและระดับนัยสำคัญ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

n / α	**10**	**20**	**30**	**50**
0.01	-	BSW	BSW,BAD	BSW,BAD
0.05	BSW	BSW,BAD	BSW,BAD	BSW,BAD
0.10	BSW,BAD	BSW,BAD	BSW,BAD	BSW,BAD

ส่วนที่ 4.2 ผลการวิจัยของการเปรียบเทียบตัวสถิติที่ใช้ในการทดสอบ โดยการพิจารณาจากค่าอำนาจการทดสอบ

4.2.1 กรณีเปรียบเทียบตัวแปรอิสระ 3 ตัว ดังตาราง 4.13-4.15 และรูปที่ 4.1-4.6 ที่ระดับนัยสำคัญ 0.01 0.05 และ 0.10

เมื่อข้อมูลมีการแจกแจงแบบโลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล พบว่ากรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 10 และ 20 ตัวสถิติบูทสแตรปแชปิโร - วิลค์ จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิงแต่ในกรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 30 และ 50 ตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแชปิโร - วิลค์

4.2.2 กรณีเปรียบเทียบตัวแปรอิสระ 5 ตัว ดังตาราง 4.16-4.18 และรูปที่ 4.7-4.12 ที่ระดับนัยสำคัญ 0.01 0.05 และ 0.10

เมื่อข้อมูลมีการแจกแจงแบบโลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล พบว่ากรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 10 และ 20 ตัวสถิติบูทสแตรปแชปิโร - วิลค์ จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง แต่ในกรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 30 และ 50 ตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแชปิโร - วิลค์

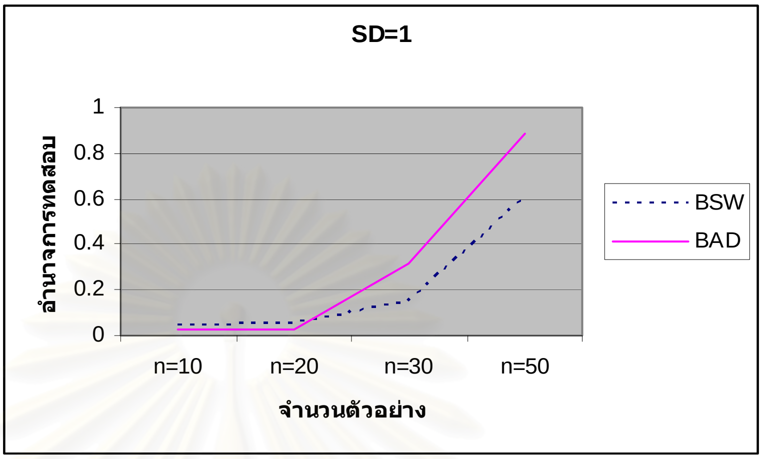
จากผลการวิจัยพบว่าเมื่อขนาดตัวอย่างและจำนวนตัวแปรอิสระเพิ่มขึ้นพบว่า ค่าอำนาจการทดสอบจะมีแนวโน้มสูงขึ้น แต่เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมากขึ้น พบว่าค่าอำนาจการทดสอบมีแนวโน้มลดลงและที่ระดับนัยสำคัญ 0.10 จะให้ค่าอำนาจการทดสอบสูงกว่าที่ระดับนัยสำคัญ 0.05 และ0.01 ตามลำดับ

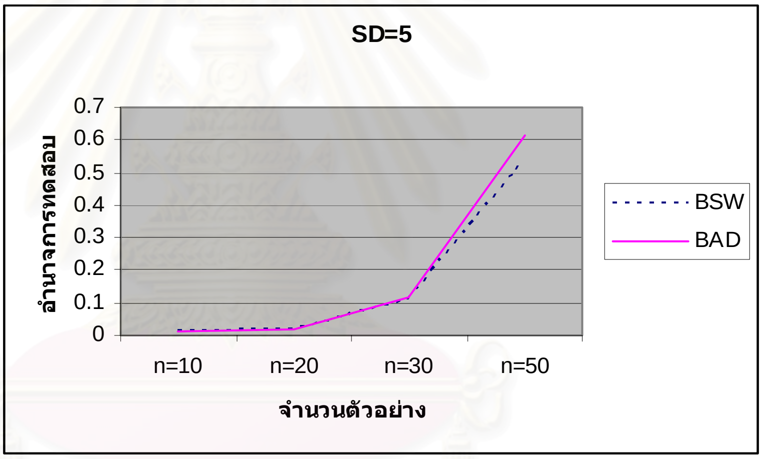
จากตารางที่ 4.13-4.18 โดยสรุปพบว่าที่ระดับนัยสำคัญ 0.01 0.05 และ0.10 เมื่อข้อมูลมีการแจกแจงแบบโลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล ตัวสถิติบูทสแตรปแชปิโร - วิลค์ จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง ในกรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 10 และ 20 ทุกระดับของจำนวนตัวแปรอิสระและทุกระดับของค่าส่วนเบี่ยงเบนมาตรฐาน โดยที่ตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแชปิโร - วิลค์ เมื่อขนาดตัวอย่างมีค่าเท่ากับ 30 และ 50 ทุกระดับของจำนวนตัวแปรอิสระและทุกระดับของค่าส่วนเบี่ยงเบนมาตรฐาน

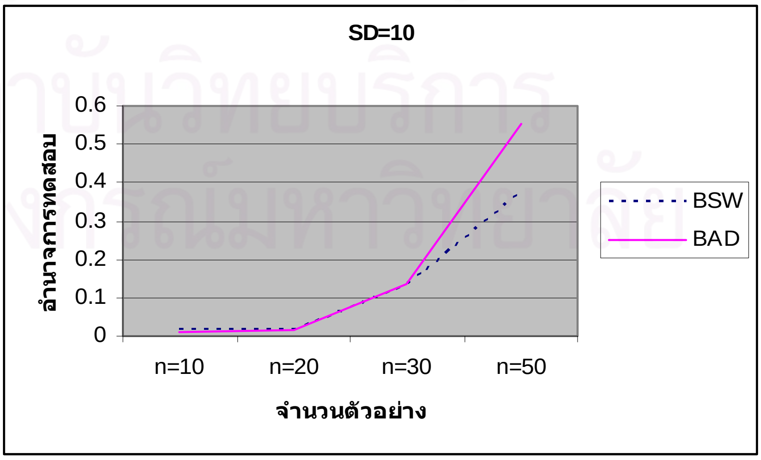
ตาราง 4.13 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระคือ p = 3 และระดับนัยสำคัญ $\alpha = 0.01$

รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติก	BSW	0.417	0.0500	0.1483	0.6033	0.0150	0.0177	0.0183	0.5400	0.0133	0.0167	0.1283	0.3767
	BAD	0.0233	0.0300	0.3167	0.8883	0.0135	0.0159	0.1183	0.6150	0.0117	0.0153	0.1333	0.5517
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.0200	0.0867	0.2200	0.7483	0.0200	0.0317	0.1983	0.7133	0.0183	0.0583	0.0867	0.7230
	BAD	0.0133	0.0217	0.3490	0.9867	0.0183	0.0200	0.3083	0.9000	0.0133	0.0150	0.2183	0.7250

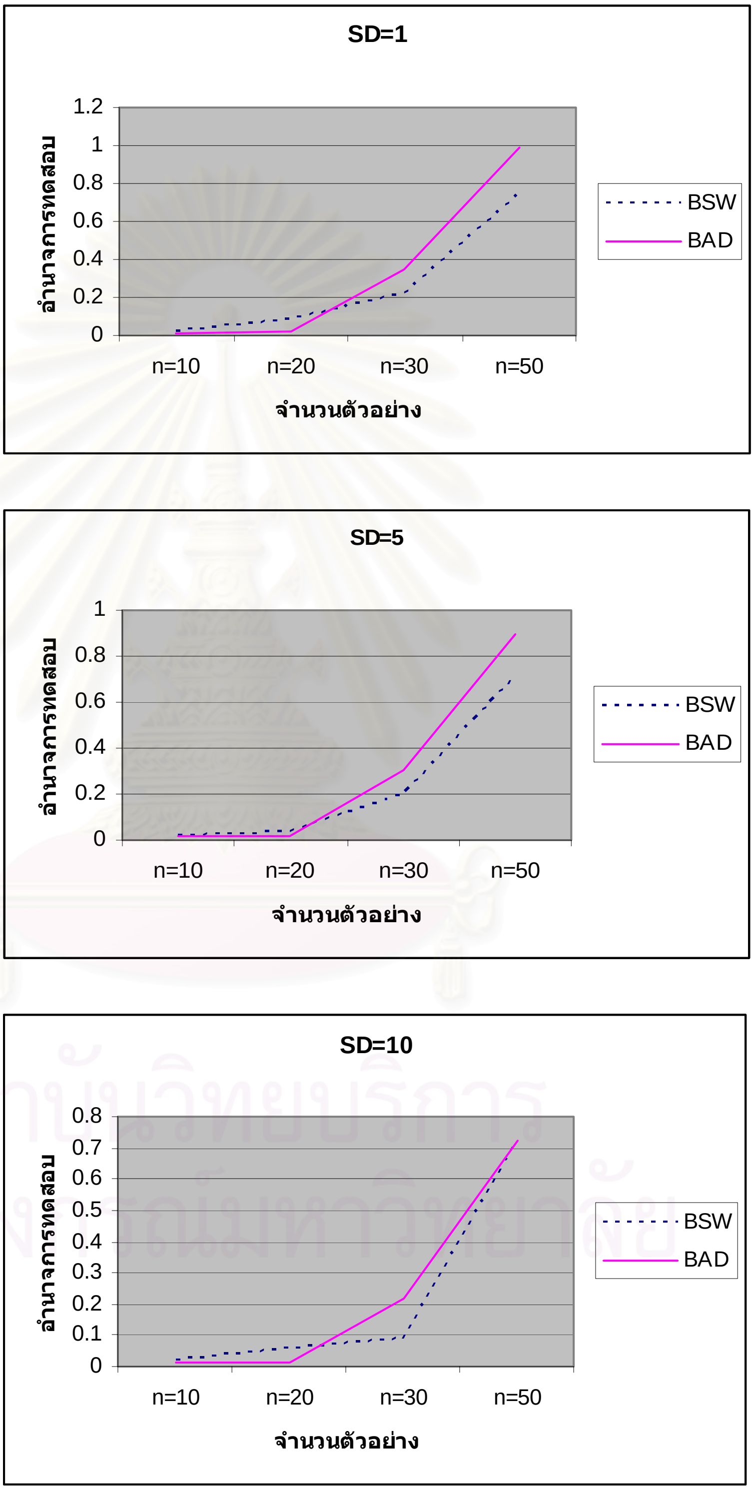
รูปที่4.1 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบโลจิสติค และระดับนัยสำคัญ $\alpha = 0.01$







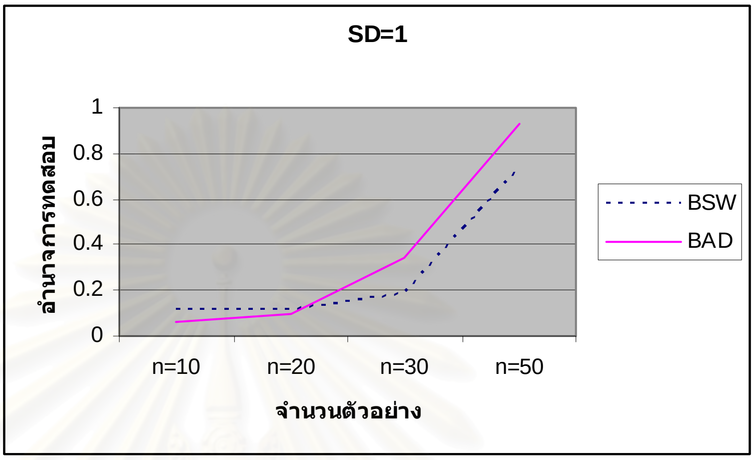
รูปที่ 4.2 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.01$

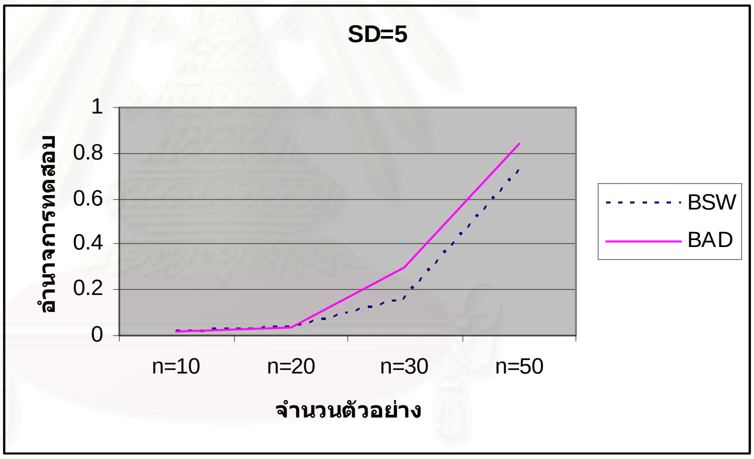


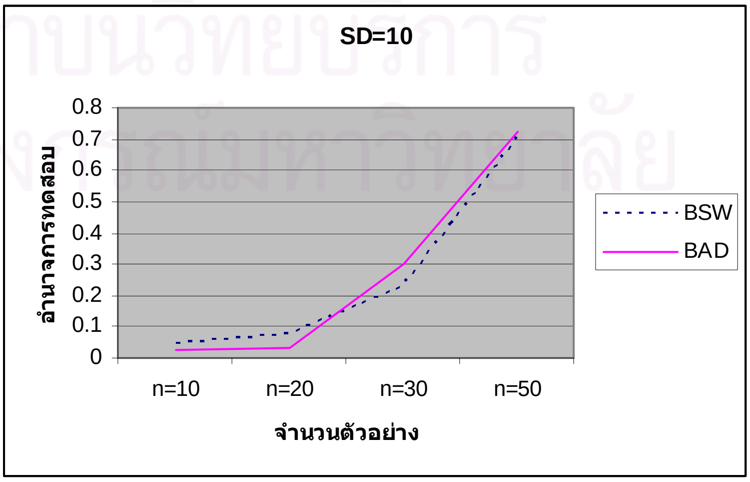
ตาราง 4.14 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระ คือ p = 3 และระดับนัยสำคัญ $\alpha = 0.05$

รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติค	BSW	0.1100	0.1133	0.1867	0.7333	0.0183	0.0383	0.1557	0.7307	0.0433	0.0750	0.2312	0.7050
	BAD	0.0600	0.0950	0.3450	0.9317	0.0167	0.0350	0.3010	0.8450	0.0267	0.0317	0.3000	0.7233
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.0550	0.1767	0.3801	0.9201	0.0517	0.1383	0.3750	0.9100	0.0317	0.1083	0.1917	0.8517
	BAD	0.0317	0.0670	0.4100	0.9933	0.0500	0.0633	0.4017	0.9433	0.0183	0.0283	0.3233	0.8800

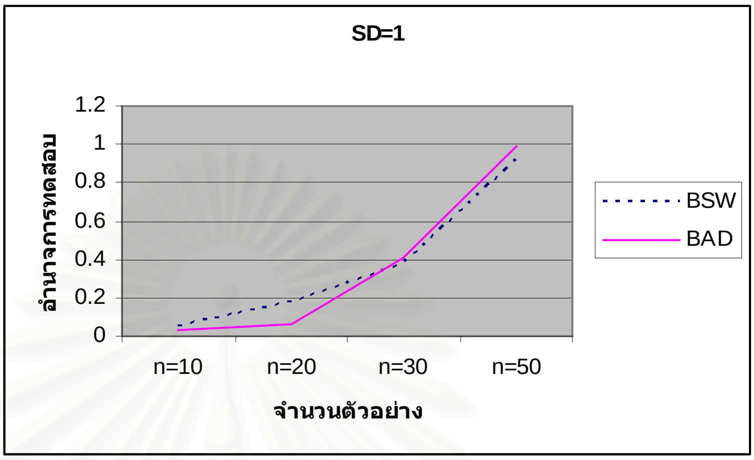
รูปที่ 4.3 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน – ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร - วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบโลจิสติคและระดับนัยสำคัญ $\alpha = 0.05$

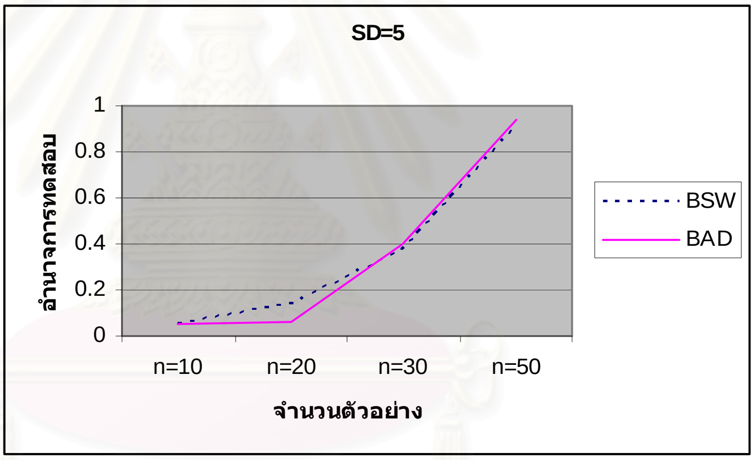


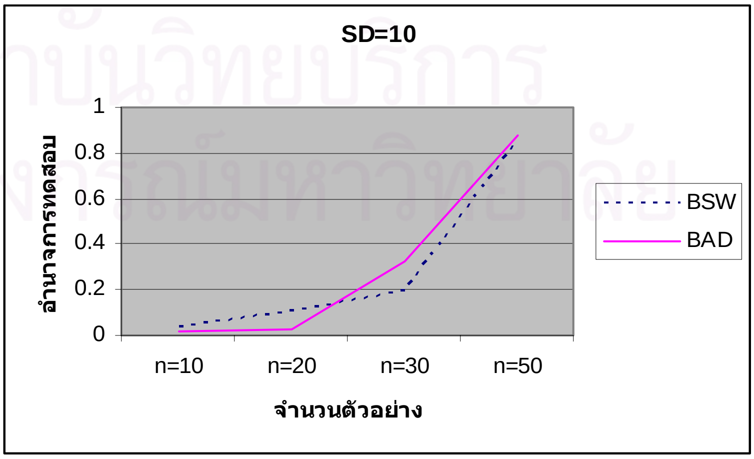




รูปที่ 4.4 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน – ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.05$



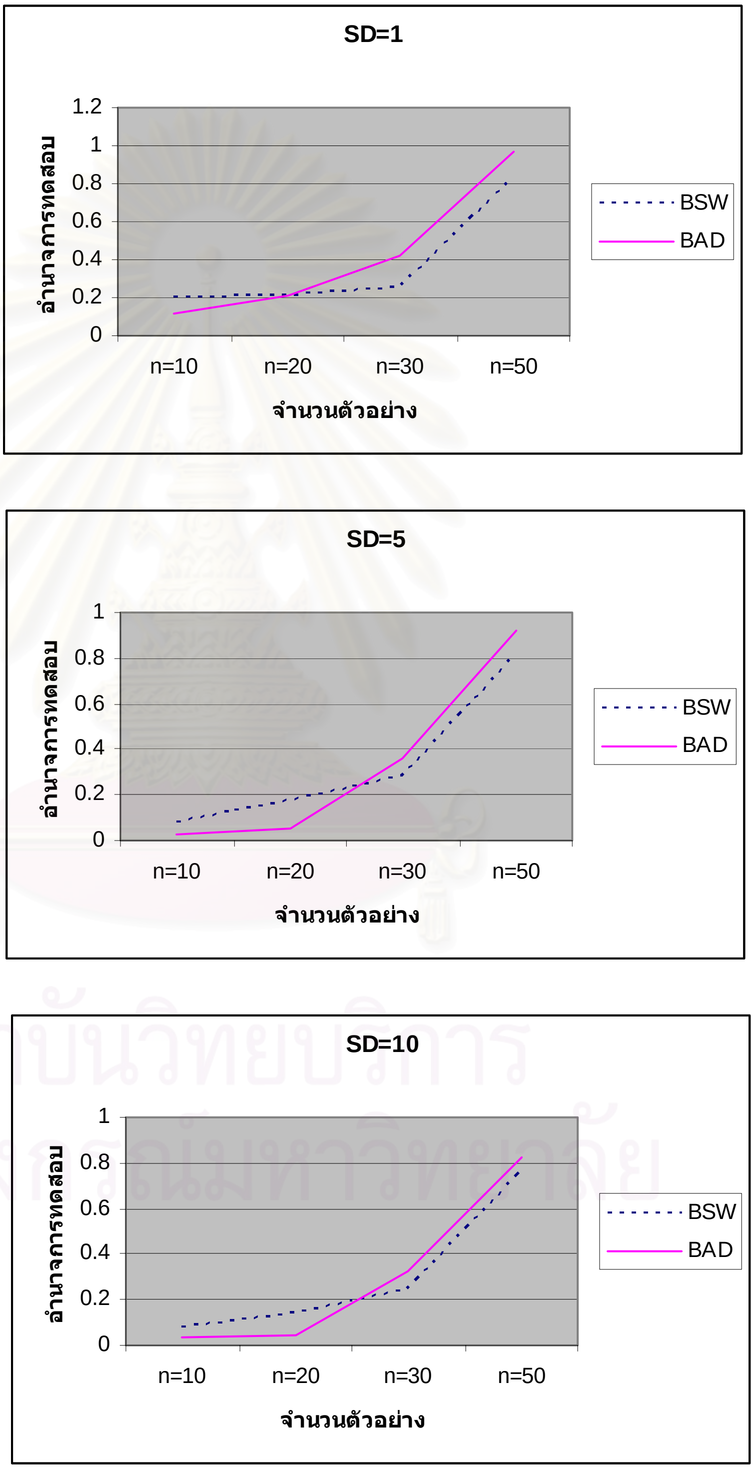




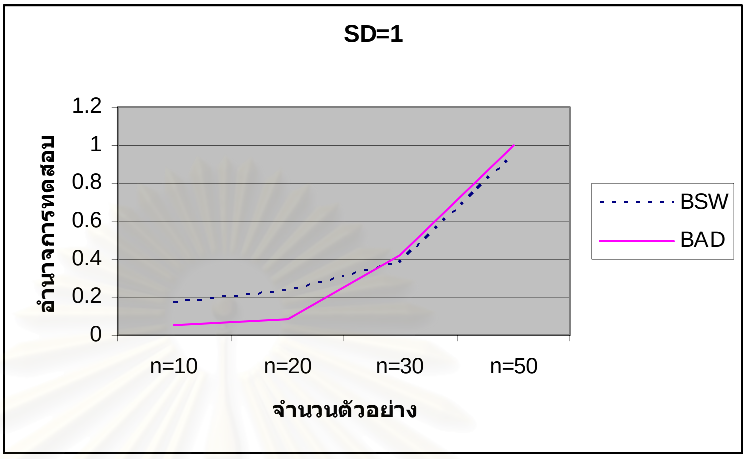
ตาราง 4.15 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระ คือ $p = 3$ และระดับนัยสำคัญ $\alpha = 0.10$

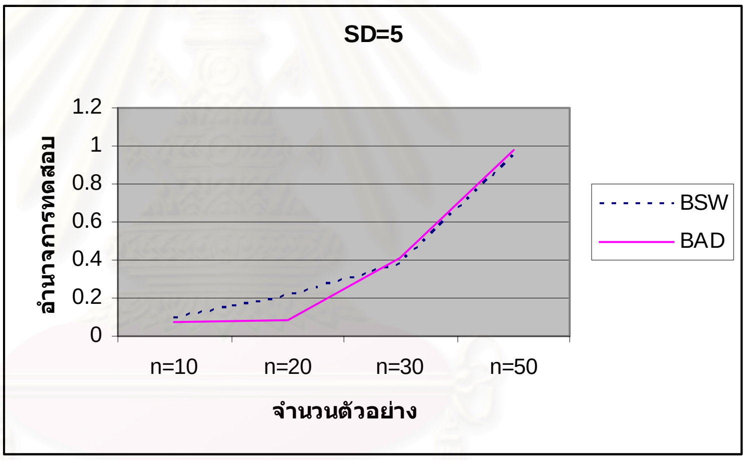
รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติค	BSW	0.1950	0.2150	0.2567	0.8283	0.0800	0.1783	0.2817	0.8124	0.0800	0.1400	0.2417	0.7650
	BAD	0.1117	0.2117	0.4200	0.9717	0.0267	0.0517	0.3601	0.9167	0.0367	0.0395	0.3222	0.8267
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.1683	0.2367	0.3830	0.9508	0.0917	0.2083	0.3803	0.9433	0.0883	0.1917	0.3283	0.8750
	BAD	0.0500	0.0850	0.4220	0.9950	0.0700	0.0817	0.4123	0.9810	0.0317	0.0867	0.3807	0.9483

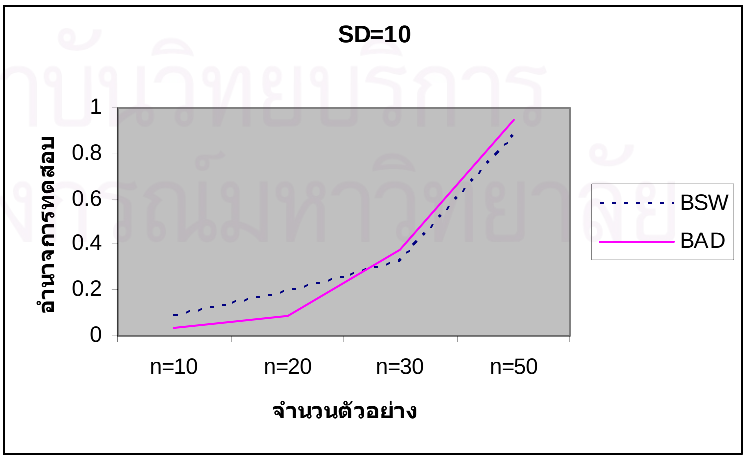
รูปที่ 4.5 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบโลจิสติคและระดับนัยสำคัญ $\alpha = 0.10$



รูปที่ 4.6 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.10$



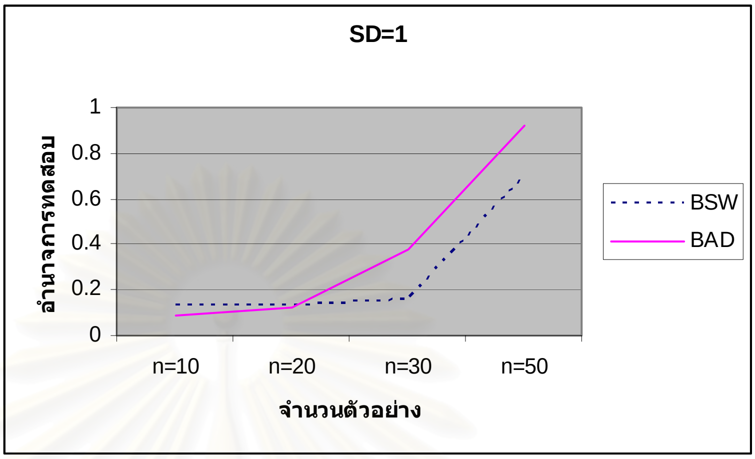


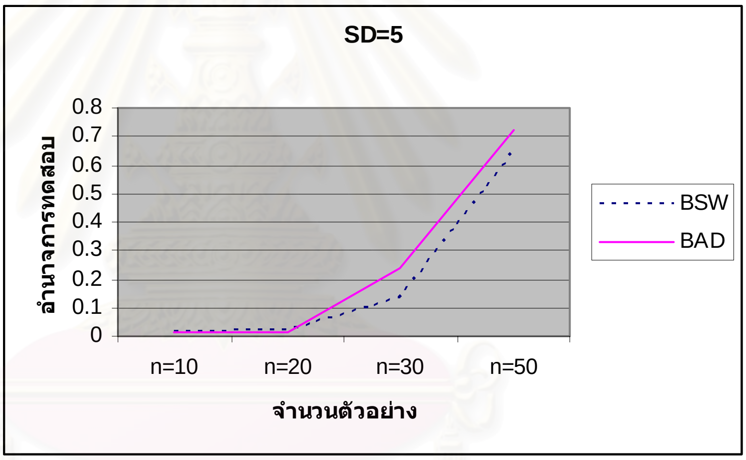


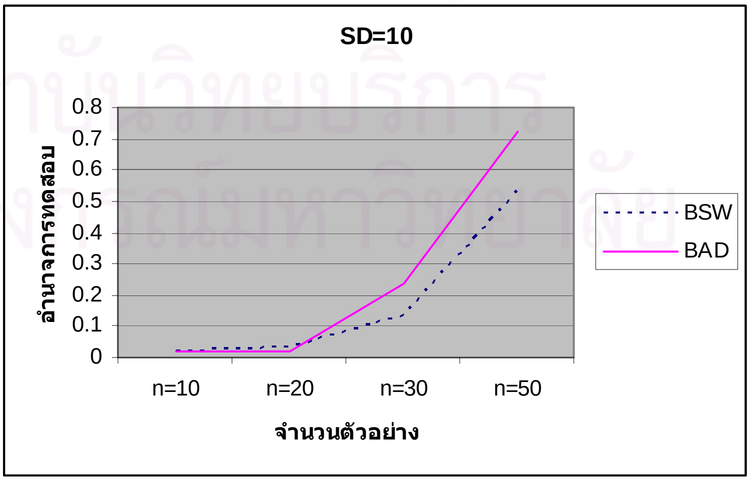
ตาราง 4.16 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระ คือ p = 5 และระดับนัยสำคัญ $\alpha = 0.01$

รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติค	BSW	0.1283	0.1317	0.1566	0.7000	0.0167	0.0183	0.1333	0.6483	0.0200	0.0317	0.1333	0.5333
	BAD	0.0850	0.1200	0.3750	0.9183	0.0165	0.0167	0.2417	0.7233	0.0163	0.0183	0.2350	0.7217
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.0201	0.0933	0.3117	0.9300	0.0467	0.0850	0.2450	0.9100	0.0267	0.0617	0.1167	0.8033
	BAD	0.0167	0.0317	0.4150	0.9870	0.0233	0.0383	0.3397	0.9333	0.0252	0.0317	0.2983	0.8917

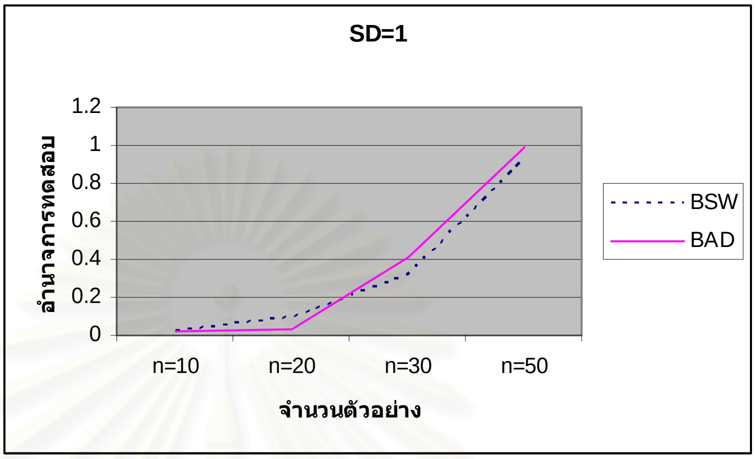
รูปที่ 4.7 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบโลจิสติคและระดับนัยสำคัญ $\alpha = 0.01$

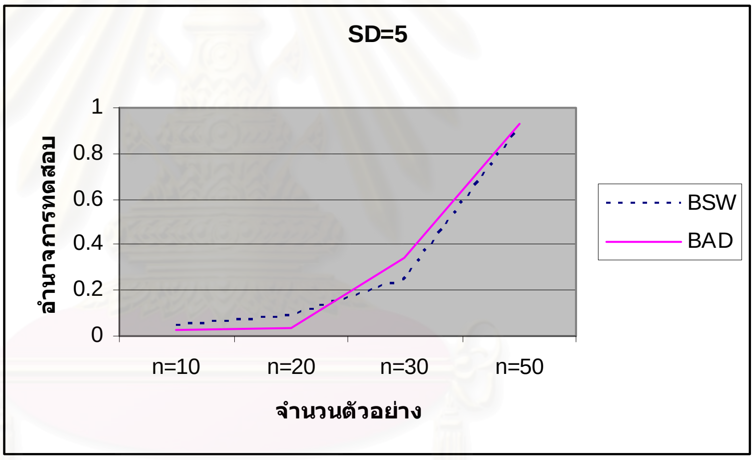


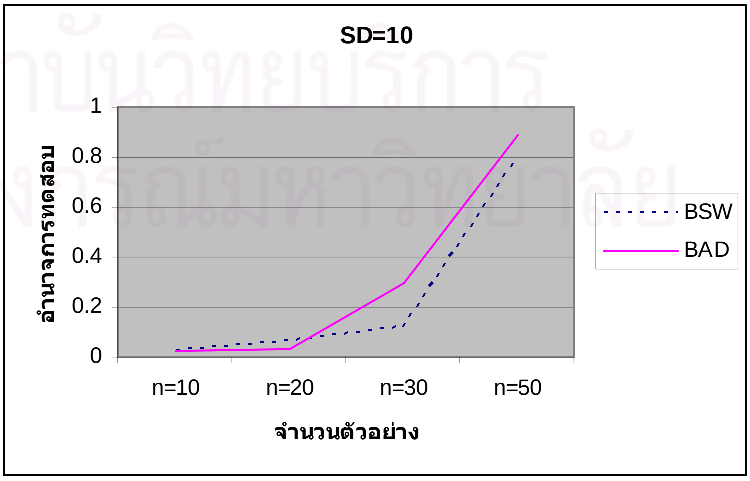




รูปที่ 4.8 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.01$



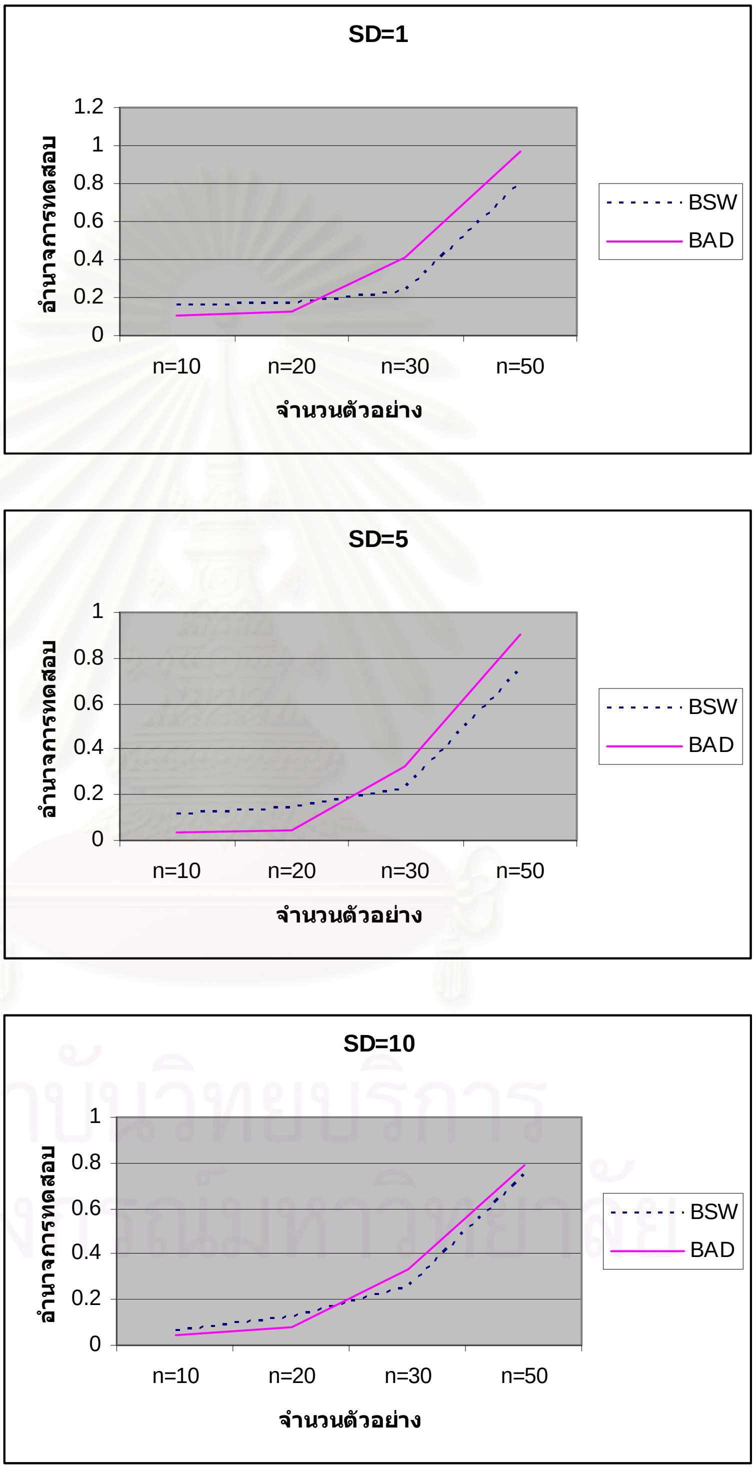




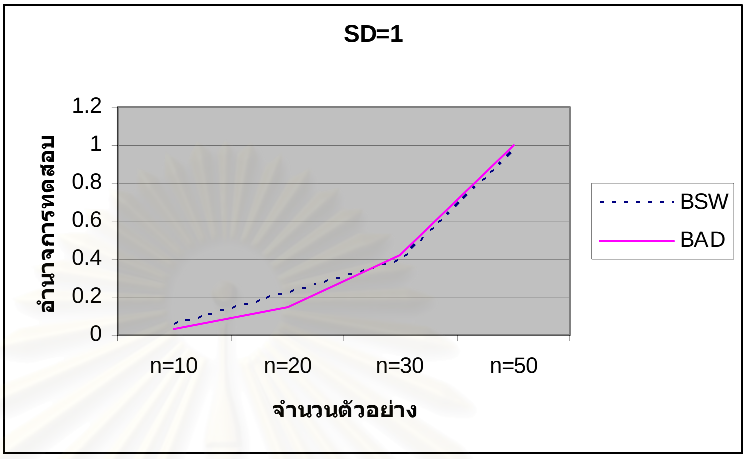
ตาราง 4.17 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระ คือ p = 5 และระดับนัยสำคัญ $\alpha = 0.05$

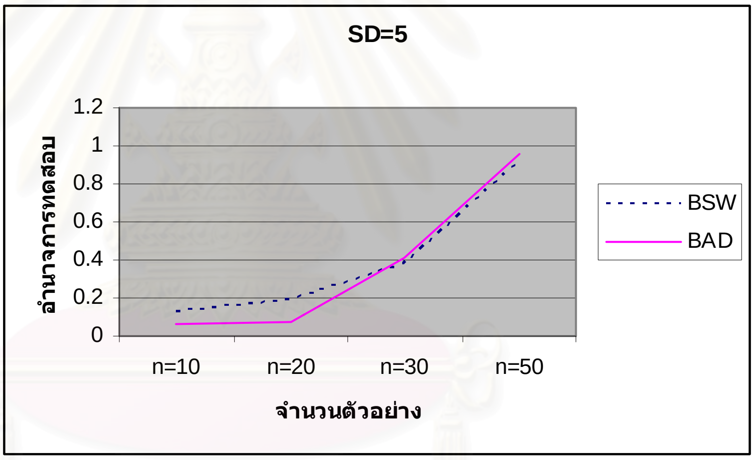
รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติค	BSW	0.1533	0.1733	0.2350	0.8033	0.1100	0.1433	0.2317	0.7500	0.0583	0.1250	0.2517	0.7500
	BAD	0.1033	0.1250	0.4152	0.9700	0.0333	0.0400	0.3219	0.9050	0.0433	0.0750	0.3350	0.7933
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.0567	0.2200	0.3880	0.9717	0.1283	0.1883	0.3780	0.9167	0.0483	0.1267	0.2183	0.9113
	BAD	0.0333	0.1500	0.4217	0.9978	0.0667	0.0767	0.4090	0.9600	0.0305	0.0733	0.3533	0.9417

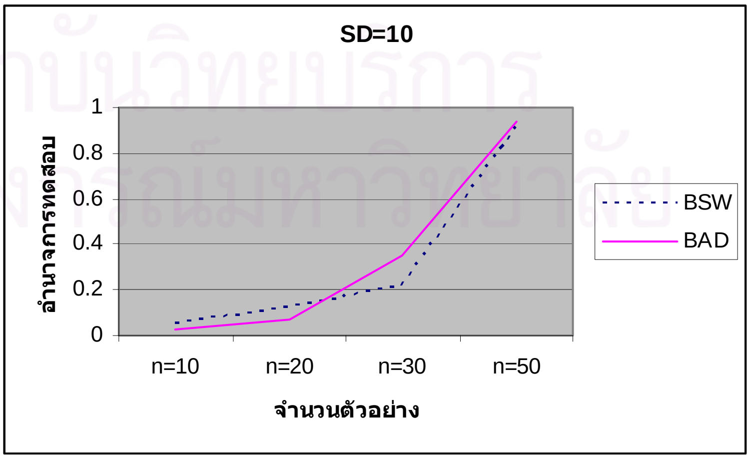
รูปที่ 4.9 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน– ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบโลจิสติคและระดับนัยสำคัญ $\alpha = 0.05$



รูปที่ 4.10 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.05$



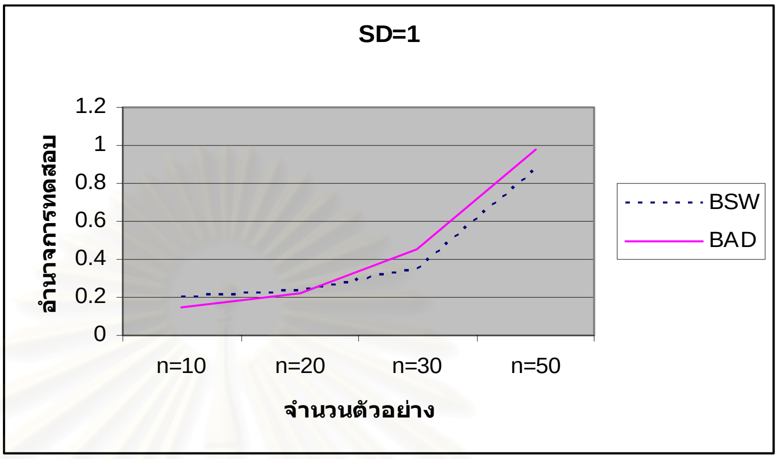


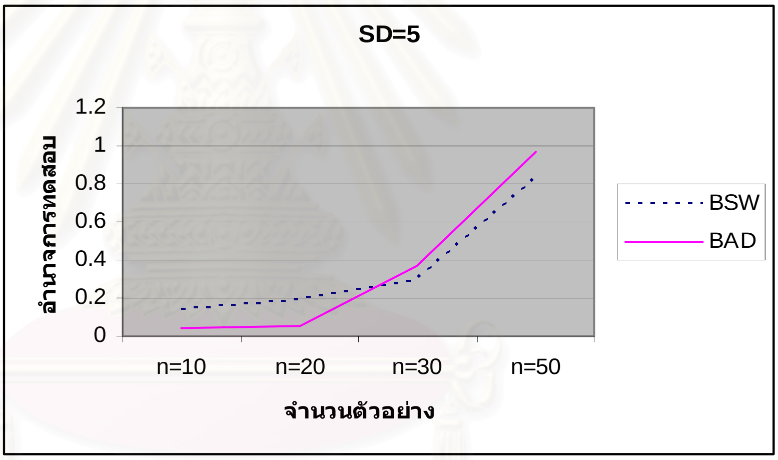


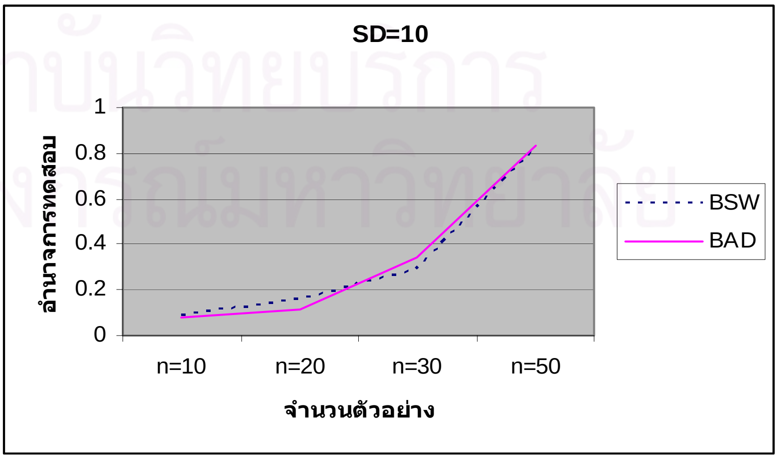
ตาราง 4.18 แสดงการเปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระ คือ $p = 5$ และระดับนัยสำคัญ $\alpha = 0.10$

รูปแบบการแจกแจง	สถิติทดสอบ	SD = 1				SD = 5				SD = 10			
		n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n = 30	n = 50	n = 10	n = 20	n =30	n = 50
โลจิสติค	BSW	0.1967	0.2267	0.3517	0.8717	0.1383	0.1901	0.2933	0.8283	0.0883	0.1617	0.2917	0.8247
	BAD	0.1505	0.2217	0.4524	0.9800	0.0473	0.0533	0.3670	0.9700	0.0767	0.1117	0.3443	0.8333
ดับเบิ้ลเอ็กซ์โปเนนเชียล	BSW	0.1743	0.2600	0.4083	0.9800	0.1767	0.2383	0.3867	0.9617	0.0933	0.1967	0.3550	0.9467
	BAD	0.1283	0.2267	0.4633	0.9983	0.0965	0.1117	0.4210	0.9900	0.0448	0.1400	0.4055	0.9791

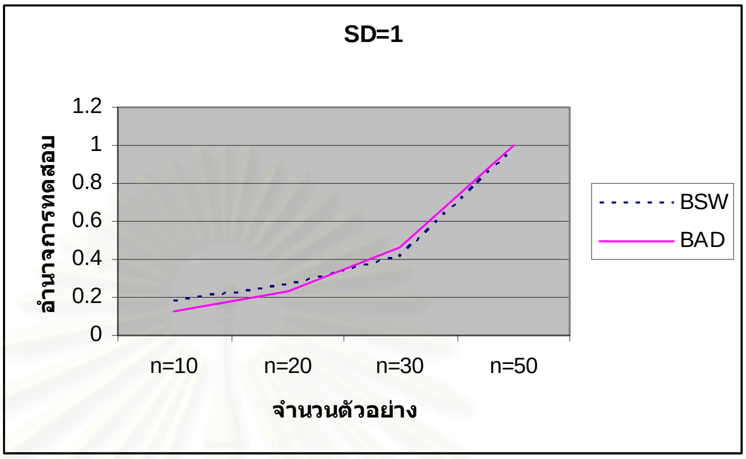
รูปที่4.11 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบโลจิสติคและระดับนัยสำคัญ $\alpha = 0.10$

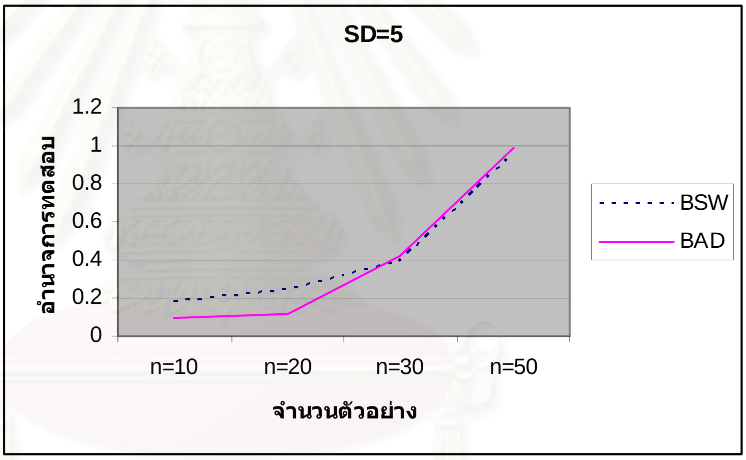


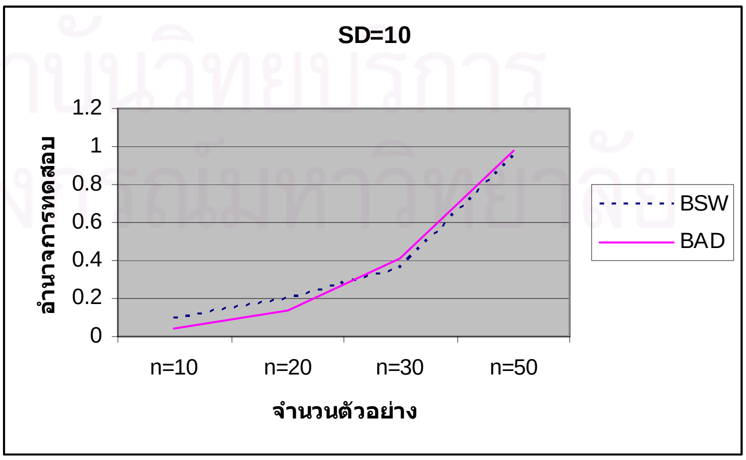




รูปที่ 4.12 เปรียบเทียบอำนาจการทดสอบของตัวสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ข้อมูลมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล และระดับนัยสำคัญ $\alpha = 0.10$







บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้จัดทำขึ้นโดยมีวัตถุประสงค์เพื่อ ศึกษาเปรียบเทียบตัวสถิติที่ใช้สำหรับการตรวจสอบการแจกแจงไม่เป็นปกติของความคลาดเคลื่อนสุ่ม ด้วยสถิติทดสอบ 2 วิธี คือ ตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ ดังนั้นเพื่อหาข้อสรุปว่าตัวสถิติทดสอบของวิธีใดมีความเหมาะสมที่จะใช้ทดสอบการแจกแจงของความคลาดเคลื่อนสุ่ม ผู้วิจัยจึงสนใจทำการเปรียบเทียบตัวสถิติทดสอบดังกล่าวข้างต้น โดยพิจารณาจากค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (Type I error) และอำนาจการทดสอบ (Power of the test) ที่แสดงไว้ในบทที่ 4 ที่กล่าวมา

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นเชิงพหุในสถานการณ์ต่างๆ ที่กำหนดขึ้นดังนี้

- ตัวแบบความถดถอยเป็นตัวแบบที่อยู่ภายใต้ข้อตกลงเบื้องต้น

- ประชากรที่ศึกษาสร้างมาจากตัวแบบ

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

- กำหนดตัวแปรอิสระที่ทำการศึกษา (X) = 3,5

- กำหนดจำนวนตัวอย่างของแต่ละชุดค่าสังเกต (n) = 10,20,30,50

- กำหนดลักษณะการแจกแจงของความคลาดเคลื่อนคือ การแจกแจงแบบปกติ การแจกแจงแบบโลจิสติคและการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

- กำหนดให้ข้อมูลมีค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : σ) ในระดับต่างๆ กันคือ 1 5 และ10

- ระดับนัยสำคัญของการทดสอบ (α) ที่ศึกษาคือ 0.01 0.05 และ 0.10

ในการพิจารณาเปรียบเทียบตัวสถิติทดสอบทั้ง 2 วิธีนั้น พิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ ซึ่งในบทนี้มีการสรุปผลการวิจัยออกเป็น 2 ส่วน โดยแต่ละส่วนจะกล่าวถึง ความสามารถในการควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และผลการเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบ และข้อเสนอแนะ ซึ่งรายละเอียดมีดังต่อไปนี้

5.1 สรุปผลการวิจัย

5.1.1 ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

เปรียบเทียบความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิง กับตัวสถิติบูทสแตรปแชปิโร-วิลค์ ที่ระดับนัยสำคัญ 0.01 0.05 และ 0.10 โดยใช้เกณฑ์พิจารณาความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของ Bradley

5.1.1.1 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ

5.1.1.2 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ

5.1.1.3 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ

5.1.1.4 เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ

5.1.1.5 เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทั้งหมดที่ระดับนัยสำคัญทั้ง 3 ระดับ
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดีเมื่อระดับนัยสำคัญของการทดสอบเป็น 0.10 ยกเว้นเมื่อใช้ขนาดตัวอย่างเป็น 10 ที่ระดับนัยสำคัญเป็น 0.01 และ 0.05 และเมื่อขนาดตัวอย่างเป็น 20 ที่ระดับนัยสำคัญเป็น 0.01

5.1.1.6 เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

- ตัวสถิติ BSW สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดีเมื่อระดับนัยสำคัญของการทดสอบเป็น 0.05 และ 0.10 ยกเว้นเมื่อใช้ขนาดตัวอย่างเป็น 10 ที่ระดับนัยสำคัญเป็น 0.01
- ตัวสถิติ BAD สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดีเมื่อระดับนัยสำคัญของการทดสอบเป็น 0.10 ยกเว้นเมื่อใช้ขนาดตัวอย่างเป็น 10 ที่ระดับนัยสำคัญเป็น 0.01 และ 0.05 และเมื่อขนาดตัวอย่างเป็น 20 ที่ระดับนัยสำคัญเป็น 0.01

ดังนั้นกล่าวโดยสรุปได้ว่าตัวสถิติบูทสแตรปแอนเดอร์สัน – ดาร์ลิง และตัวสถิติบูทสแตรปแชปิโร-วิลค์ สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ทุกสถานการณ์เมื่อขนาดตัวอย่างมีค่าเป็น 30 และ 50 แต่ในกรณีที่ขนาดตัวอย่างมีค่าเป็น 10 และ 20 ตัวสถิติทั้งสองไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ในบางสถานการณ์ทั้งนี้จะต้องพิจารณาถึงปัจจัยอื่นๆที่อาจจะมีผลต่อความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ด้วย เช่น จำนวนตัวแปรอิสระ ส่วนเบี่ยงเบนมาตรฐานและระดับนัยสำคัญ

5.1.2 ค่าอำนาจการทดสอบ

ปัจจัยที่มีผลต่อค่าอำนาจการทดสอบ

5.1.2.1 จำนวนตัวแปรอิสระ

ในทุกกรณีศึกษาจำนวนตัวแปรอิสระจะแปรผันตามกับค่าอำนาจการทดสอบ ทั้งนี้เพราะในการเพิ่มตัวแปรอธิบายของตัวแบบความถดถอยจะทำให้ค่า MSE เพิ่มขึ้นซึ่งมีผลทำให้การแจกแจงของข้อมูลเบี่ยงเบนไปจากเดิม และทำให้ตัวสถิติสามารถตรวจสอบได้ดีขึ้น เมื่อตัวสถิติสามารถตรวจสอบได้ดีขึ้นดังนั้นจึงเป็นการเพิ่มอำนาจการทดสอบ

5.1.2.2 ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน

ในกรณีส่วนใหญ่ส่วนเบี่ยงเบนมาตรฐานจะแปรผกผันกับค่าอำนาจการทดสอบ ทั้งนี้เพราะในการเพิ่มความเบี่ยงเบนของข้อมูลจะทำให้ข้อมูลมีความแตกต่างกันมากขึ้น ดังนั้นค่าสถิติที่ได้จะเกิดความคลาดเคลื่อนขึ้นจึงเป็นสาเหตุที่ทำให้อำนาจการทดสอบลดลง

5.1.2.3 ขนาดตัวอย่าง

ในทุกกรณีศึกษาขนาดตัวอย่างจะแปรผันตามกับค่าอำนาจการทดสอบ ทั้งนี้เพราะในการเพิ่มขนาดตัวอย่างให้มากขึ้นจะทำให้ค่าที่ได้มีค่าใกล้เคียงค่าจริงมากขึ้นซึ่งจะเป็นการลดความเบี่ยงเบนของข้อมูลในตัวแบบความถดถอย จึงทำให้อำนาจการทดสอบมีค่าเพิ่มขึ้น

5.1.2.4 ระดับนัยสำคัญ

ในทุกกรณีศึกษาระดับนัยสำคัญจะแปรผันตามกับค่าอำนาจการทดสอบ ทั้งนี้เพราะเมื่อความผิดพลาดที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างจริง (α) มีมากขึ้นจะลดความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง (β) จึงทำให้ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นไม่จริงหรือค่าอำนาจการทดสอบ ($1-\beta$) มีค่าที่เพิ่มขึ้น

5.2 ข้อเสนอแนะ

5.2.1 ด้านการนำไปใช้

5.2.1.1 จากผลการวิจัย ในการเปรียบเทียบตัวสถิติทดสอบทั้ง 2 วิธีพบว่ากรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 10 และ 20 ตัวสถิติบูทสแตรปแชปิโร - วิลค์ จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง แต่ในกรณีที่ขนาดตัวอย่างมีค่าเท่ากับ 30 และ 50 ตัวสถิติบูทสแตรปแอนเดอร์สัน - ดาร์ลิง จะให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติบูทสแตรปแชปิโร - วิลค์

5.2.2 ด้านการศึกษาวิจัย

5.2.2.1 การศึกษาวิจัยในครั้งนี้ ศึกษาเฉพาะสมการถดถอยเชิงเส้นเท่านั้น สำหรับการวิจัยในครั้งต่อไป อาจทำการศึกษาในการวิเคราะห์ความถดถอยรูปแบบอื่นๆ

5.2.2.2 ในการศึกษาวิจัยครั้งนี้ ศึกษาในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติ โลจิสติคและดับเบิ้ลเอ็กซ์โปเนนเชียล เท่านั้น ในการศึกษาวิจัยครั้งต่อไปอาจศึกษากรณีที่ความคลาดเคลื่อนมีการแจกแจงรูปแบบอื่นๆ

5.2.2.3 สำหรับการวิจัยในครั้งนี้ใช้โปรแกรม S-PLUS 2000 ซึ่งใช้ระยะเวลาในการประมวลผลนานมาก ดังนั้นเพื่อความสะดวกในการศึกษาวิจัยครั้งต่อไปอาจใช้โปรแกรมอื่นสำหรับการวิจัยแทน เช่น โปรแกรม BORLAND DELPHI 6 เป็นต้น

รายการอ้างอิง

ภาษาไทย

เกตุจันทร์ พัชรินทร์ศักดิ์. การเปรียบเทียบวิธีการนอนพาราเมตริกซ์สำหรับการทดสอบการแจกแจงแบบปกติ. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2534.

ทรงศิริ แต้สมบัติ. การวิเคราะห์การถดถอย. กรุงเทพมหานคร : ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, 2541.

มานพ วราภักดิ์. การจำลองเบื้องต้น. กรุงเทพมหานคร : ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2547.

มาลี ตระการศิรินันท์. การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการความถดถอยเชิงเส้นด้วยวิธีกำลังสองต่ำสุดและวิธีบูทสแตรป. วิทยานิพนธ์ปริญญามหาบัณฑิต,สาขาวิชาสถิติ ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2532.

วิชัย สุรเชิดเกียรติ. การจำลอง. พิมพ์ครั้งที่ 2. กรุงเทพหานคร: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2542.

สุพล ดุรงค์วัฒนา. การวิเคราะห์เชิงสถิติ:การวิเคราะห์ความถดถอย. กรุงเทพมหานคร: ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย, 2537.

สุพล ดุรงค์วัฒนา. การวางแผนการทดลองเพื่อการวิจัยขั้นสูง. กรุงเทพมหานคร: ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง. กรุงเทพมหานคร: ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

ภาษาอังกฤษ

Alberto Viglione: Non-supervised Regional Frequency Analysis. Stats (December 2006).

Christopher Z. Mooney and Robert D. Duval. Bootstrapping : A Nonparametric Approach to Statistical Inference. (37-40): Sage Publications, 1993.

Jean-Marie Dufour, Abdeljelil Farhat, Lucien Gardiol, and Lynda Khalaf. Simulation-based finite sample normality tests in linear regressions: Econometrics Journal 1 (1998): c154-c173.

Michael R.Chernick. Bootstrap Methods : A Practitionner's Guide. (54-57): A Wiley - Interscience Publication ,1999.

Natalie Neumeyer, Holger Dette, and Eva-Renate Nagel. Bootstrap tests for the error distribution in linear and nonparametric regression models: Ruhr-University Bochum Germany, 2004.

Pranab Kumar Sen, Jana Jureckova, and Jan Picek. Goodness - of - fit Test of Shapiro – Wilk Type with Nuisance Regression and Scale: Austrian Journal of Statistics 32 (2003): c163-c176.

ภาคผนวก ก

การจำลองการแจกแจงของความคลาดเคลื่อน

1.โปรแกรมสำหรับสร้างตัวเลขสุ่ม[1]

ตัวเลขสุ่มมีความจำเป็นอย่างมากในการจำลองปัญหาเกือบทั้งหมด ในภาษาคอมพิวเตอร์เกือบทุกภาษามักจะมีโปรแกรมย่อยหรือฟังก์ชันในการสร้างตัวเลขสุ่มโดยตัวเลขสุ่มที่สร้างขึ้นมานี้ มาจากความสัมพันธ์ที่มีรูปแบบซ้ำๆ (Recurrence relation) กล่าวคือต่อไปเกิดจากกระบวนการทางคณิตศาสตร์และตรรกศาสตร์ของตัวเลขปัจจุบันหรือตัวเลขในอดีต ลำดับ(Sequence) ของตัวเลขซึ่งผลิตได้จึงเป็นลำดับของเลขสุ่มในความหมายที่ไม่แท้จริงซึ่งมีคาบ (Period) เกิดขึ้น ถ้าลำดับตัวเลขดังกล่าวผ่านการทดสอบคุณสมบัติทางสถิติที่สำคัญของตัวเลขสุ่มคือมีความสม่ำเสมอ (Uniformity)และความเป็นอิสระ(Independence) เราจะเรียกลำดับตัวเลขนี้ว่า“ตัวเลขคล้ายสุ่ม” (Pseudo random number)

โปรแกรมสำหรับสร้างตัวเลขสุ่มที่ดีจะต้องคำนึงถึงสิ่งสำคัญต่อไปนี้

1.โปรแกรมย่อยหรือฟังก์ชันในการสร้างตัวเลขคล้ายสุ่มต้องทำงานได้เร็ว

2. ต้องใช้เนื้อที่ในหน่วยความจำน้อย

3. ต้องมีวัฏจักรที่ยาว

4. ต้องมีความสม่ำเสมอและเป็นอิสระกัน

5. ต้องสามารถนำกลับมาใช้ได้อีก เพื่อใช้เปรียบเทียบความสามารถระหว่างระบบงาน

การสร้างตัวเลขสุ่มโดยใช้โปรแกรมมีข้อดีหลายประการคือ สามารถสร้างได้เร็วและสามารถสร้างลำดับของตัวเลขชุดเดิมออกมาได้ ซึ่งเป็นสิ่งจำเป็นอย่างยิ่งในกรณีที่ใช้การทดสอบแบบจำลองและต้องการจะทบทวนการคำนวณโปรแกรม ส่วนข้อเสียในการสร้างนี้ก็คือ การสร้างลำดับของตัวเลขเป็นลำดับที่มีคาบ และการทำให้ตัวเลขสุ่มมีคุณสมบัติเชิงสถิติทางทฤษฎีทำได้ยากพอควร

การสร้างตัวเลขสุ่มโดยการใช้โปรแกรมได้พัฒนาไปอย่างรวดเร็ว ในปี พ.ศ. 2489 วอนนิวแมน (Von Neuman) และเมโทรโพลิส (Metropolis) ได้เสนอวิธีตัวกลางกำลังสอง (Mid-square method) ต่อมาได้มีการนำวิธีตัวกลางกำลังสองไปพัฒนาโดยฟอซิส (Forsythe) และในปี พ.ศ.2494 ลาเมอร์ (Lehmer) ได้เสนอวิธีสร้างตัวเลขสุ่มด้วยการใช้เศษจากการหารผลคูณ (Multiplicative congruential method) ซึ่งเป็นวิธีที่นิยมใช้ในปัจจุบัน

---

[1] วิชัย สุรเชิดเกียรติ, การจำลอง, (คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ 2542), หน้า 102-103

วิธีการใช้เศษจากการหารผลคูณ จะหาเลขสุ่มโดยทำการคำนวณจากสมการ

$$X_{i+1} = (X_i.a) \bmod m \qquad \text{...(1)}$$

เมื่อ X_i คือเลขคล้ายสุ่มตัวที่ i ; i = 1,2,3,…

X_{i+1} คือเลขคล้ายสุ่มตัวที่ i +1

a คือตัวคูณคงที่ (Constant Multiplier)

$\bmod m$ คือค่า $(X_i.a)$ ถูกหารด้วย m จนกระทั่งเหลือเศษน้อยกว่าค่า m ตัวเลขที่เหลือเศษจึงเป็นเลขคล้ายสุ่มตัวต่อไปคือ X_{i+1}

การหาตัวเลขคล้ายสุ่มด้วยวิธีนี้ เริ่มต้นจากการกำหนดค่าเริ่มต้น(Initial value or seed) X_0 เป็นเลขจำนวนเต็มค่าใดค่าหนึ่งในช่วง [0 , m-1] แทนค่าในสมการที่ 1 จะได้ตัวเลขคล้ายสุ่มมา จากนั้นจึงนำตัวเลขคล้ายสุ่มนี้ไปสร้างตัวเลขคล้ายสุ่มตัวต่อไป การเลือกค่า m,a และ X_0 จึงมีความสำคัญในการผลิตเลขคล้ายสุ่มที่มีคาบใกล้เคียงกับ m มากที่สุด

ลาเมอร์ได้มีการทดลองเลือกใช้ค่า m , a และ X_0 ที่จับต่างๆ กันเพื่อใช้ผลิตเลขคล้ายสุ่มตามสมการที่ 1 พบว่า ถ้าเลือก X_0 เป็นเลขคี่ และ $m = 2^r$ เมื่อ $r > 2$ และ (เมื่อ k เป็นจำนวนเต็มบวกใดๆ) จะได้คาบของตัวเลขคล้ายสุ่มมากที่สุด และเท่ากับ 2^{r-2} วิธีดังกล่าวเป็นวิธีการเลือกค่าพารามิเตอร์ทั้ง 3 ตัว เพื่อจะได้กลุ่มของเลขคล้ายสุ่มที่ดีและมีการแจกแจงสม่ำเสมอในช่วง [0,1]

2.การผลิตเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ[2]

จากสมการการผลิตเลขสุ่มสามารถผลิตเลขสุ่มที่มีการแจกแจงสม่ำเสมอได้โดยตรงคือ

1. ค่า m เป็นจำนวนเต็มที่ใหญ่ที่สุด (largest integer) และเป็นเลขคี่ที่สามารถคำนวณได้ จากเครื่องคอมพิวเตอร์จาก $m = 2^b$ เมื่อ b เป็นค่าความยาว 1 คำ(word) หรือจำนวนบิท (bit) ใน 1 คำ ในการศึกษาครั้งนี้ใช้เครื่องคอมพิวเตอร์ที่มี 32 บิท ซึ่งบิทสุดท้ายใช้สำหรับ

[2] วิชัย สุรเชิดเกียรติ, การจำลอง, (คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ 2542), หน้า 102-103

แสดงเครื่องหมาย ดังนั้นเลขจำนวนเต็มที่ใหญ่ที่สุดใน 1 คำ และเป็นเลขคี่ที่คอมพิวเตอร์ได้รับคือ $2^{b-1}-1$ เท่ากับ $2^{31-1}-1=2147483647$ นั่นคือค่า m ควรมีค่า $=2147483647$

2. ค่า seed(X_0) ควรมีค่าที่เป็นจำนวนเฉพาะสัมพัทธ์กับค่า m (relatively prime to m) เมื่อ m เป็นค่ากำลังของสอง (จาก $m=2^b$) ดังนั้น X จึงควรมีค่าเป็นเลขจำนวนเต็มบวกที่เป็นเลขคี่ (ในกรณีที่ใช้ X_0 เป็นเลขคู่จะพบว่าทุก ๆ ค่า X_I ต่อไปจะเป็นเลขคู่เสมอ จึงไม่มีคุณสมบัติเป็นเลขสุ่ม)

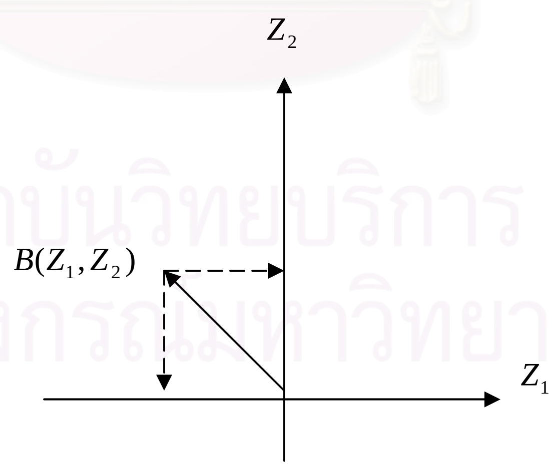
3. ค่าคงที่ที่ใช้เป็นคูณ a (Constant Multiplier) ควรมีค่าเป็นจำนวนเฉพาะสัมพัทธ์กับ m ด้วย นั่นคือ a ต้องเป็นเลขคี่ พบว่าวิธีเลือกที่ดีที่สุดสำหรับค่า a เมื่อใช้ความสัมพัทธ์เป็น $\pm 3(\bmod m)$ หรือ $a=8t\pm 3$ เมื่อ t เป็นค่าบวกใดๆ a จะมีค่าเข้าใกล้ $2^{b/2}$ ซึ่ง a จะเป็นเลขอันดับแรกของความสัมพันธ์ระหว่างเลขคล้ายสุ่ม ดังนั้นเราจึงเลือกใช้ค่า

$a=2^{16}+3=65539$

3.การผลิตเลขสุ่มที่มีการแจกแจงเป็นปกติ[3]

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติจะใช้วิธีของบอกซ์และมุลเลอร์(Box and Muller) ซึ่งผลิตเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ยเป็นศูนย์และมีความแปรปรวนเป็นหนึ่งพร้อมกัน 2 ค่า โดยใช้ตัวผลิต (generator) Z_1 และ Z_2 ดังรูปต่อไปนี้

รูปที่ ก1 แสดงการผลิตเลขสุ่มที่มีการแจกแจงแบบปกติโดยใช้ตัวผลิต



[3] วิชัย สุรเชิดเกียรติ, การจำลอง, (คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ 2542), หน้า 102-103

จากรูปเราจะได้ว่า $Z_1 = B\cos\theta$

และ $Z_2 = B\sin\theta$

เมื่อ θ เป็นมุมที่วัดจากแกน Z_1 ไปยังแกนของค่า B ในทิศทวนเข็มนาฬิกา

เนื่องจาก $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงแบบไคกำลังสอง (Chi-square distribution) ด้วยระดับความเป็นเสรี (Degree of freedom) เท่ากับ 2 ซึ่งเทียบเท่ากับการแจกแจงแบบชี้กำลัง (exponential distribution) ที่มีค่าเฉลี่ยเท่ากับ 2 เราสามารถใช้วิธีการแปลงผกผัน (Inverse transformation) สร้างเลขสุ่มที่มีการแจกแจงดังต่อไปนี้

$$B = \left(-2\ln(R)\right)^{1/2}$$

เมื่อ R เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ

จากการสมมาตรของการแจกแจงปกติ เราจะได้ว่า θ มีการแจกแจงสม่ำเสมอระหว่าง 0 ถึง 2π เรเดียน และค่า B และ θ เป็นอิสระกัน (Mutually independent)

$$Z_1 = \left(-2\ln(R_1)\right)^{1/2}\cos(2\pi R_2)$$

$$Z_2 = \left(-2\ln(R_1)\right)^{1/2}\sin(2\pi R_2)$$

สำหรับการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 เราจะทำการแปลงค่า Z_1 และ Z_2 ดังกล่าวโดยใช้สมการ

$$NORMAL_1 = \mu + \sigma Z_1$$

$$NORMAL_2 = \mu + \sigma Z_2$$

เราจะได้ว่า $NORMAL_1$ และ $NORMAL_2$ มีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2

4.การผลิตเลขสุ่มที่มีการแจกแจงแบบโลจิสติค[4]

การแจกแจงแบบโลจิสติค (logistic distribution) มีฟังก์ชันความหนาแน่น

---

[4] มานพ วราภักดิ์, การจำลองเบื้องต้น, (คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย 2547), หน้า 156

$$f(x)=\frac{\exp[-(x-\mu)/\beta]}{\beta\{1+\exp[-(x-\mu)/\beta]\}^2}, -\infty < x < \infty; -\infty < \mu < \infty, \beta > 0$$

โดยมี μ และ β เป็นพารามิเตอร์ ถ้า X มีการแจกแจงโลจิสติคด้วยพารามิเตอร์ μ,β จะเขียนแทนด้วย $X \sim Lo(\mu,\beta)$ ตัวอย่าง X เช่น X แทนขนาดของความคลาดเคลื่อนสุ่ม และใช้ประโยชน์ในการจำลองค่าผิดปกติ เป็นต้น ถ้า $X \sim Lo(\mu,\beta)$ พิสูจน์ได้ว่า

$$E(X)=\mu$$

$$Var(X)=\frac{(\pi\beta)^2}{3}$$

$$\gamma_1 = 0$$

$$\gamma_2 = 4.2$$

การจำลองตัวแปรสุ่มโลจิสติค ใช้วิธีการแปลงผกผันได้ไม่ยาก เริ่มด้วยการหาฟังก์ชันการแจกแจงสะสม F ของ X

$$F(x)=\frac{1}{\beta}\int_{-\infty}^{x}\frac{e^{-(y-\mu)/\beta}}{\left[1+e^{-(y-\mu)/\beta}\right]^2}dy$$

$$=\frac{1}{1+e^{-(x-\mu)/\beta}}$$

จากนี้หา x ในเทอมของ r $(0 \le r \le 1)$ จากสมการ $F(x)=r$ ได้ตัวแบบจำลองสำหรับ $X \sim Lo(\mu,\beta)$ ดังนี้

$$X=\mu-\beta\ln\left(\frac{1}{R}-1\right), \ \mathrm{R} \sim \mathrm{U}(0,1)$$

5. การผลิตเลขสุ่มที่มีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล[5]

ฟังก์ชันการแจกแจงคือ

$$f(x)=\left(\frac{1}{2\beta}\right)e^{\left(-\frac{|x-\alpha|}{\beta}\right)} \qquad ; -\infty < \alpha < \infty, \beta > 0$$

---

[5] มานพ วราภักดิ์, การจำลองเบื้องต้น, (คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย 2547), หน้า 156

การผลิตเลขสุ่มที่มีการแจกแจงดับเบิ้ลเอ็กซ์โปเนนเชียลจะใช้วิธีการแปลงผกผัน โดยเริ่มด้วยการหาฟังก์ชันการแจกแจงสะสมดังนี้

$$F(x)=\begin{cases}\frac{1}{2}e^{\frac{(x-\alpha)}{\beta}}, -\infty < x < \alpha \\ 1-\frac{1}{2}e^{\frac{-(x-\alpha)}{\beta}}, \alpha \le x < \alpha\end{cases}$$

จากนั้นแก้สมการหา x ในเทอมของ r จาก $F(x)=r, 0 \le r \le 1$ จะได้

$\frac{1}{2}e^{\frac{(x-\alpha)}{\beta}} = r$ ก็ต่อเมื่อ $x=\alpha+\beta\ln(2r)$ สำหรับ $0<r<\frac{1}{2}$ และ

$1-\frac{1}{2}e^{-\frac{(x-\alpha)}{\beta}} = r$ ก็ต่อเมื่อ $x=\alpha-\beta\ln(2(1-r))$ สำหรับ $\frac{1}{2} \le r < 1$

ดังนั้นจะได้ตัวแบบจำลองตัวแปรสุ่มเป็นดังนี้

$$X=\begin{cases}\alpha+\beta\ln(2R), 0<R<\frac{1}{2} \\ \alpha-\beta\ln(2(1-R)), \frac{1}{2} \le R < 1\end{cases}$$

ภาคผนวก ข

โปรแกรมสำหรับการดำเนินการวิจัย

ตาราง ข1 แสดงฟังก์ชันการทำงานของโปรแกรม S-PLUS 2000 ทั้งหมดที่ใช้ในการวิจัย

ลำดับที่	ชื่อฟังก์ชัน	หน้าที่การทำงาน
1	dim	กำหนดมิติของตัวแปรสำหรับเก็บข้อมูล
2	array(c(),dim)	ตัวแปรสำหรับการเก็บข้อมูล
3	rnorm(n,mean,sd)	สร้างตัวเลขสุ่มให้มีการแจกแจงปกติ n ค่าที่มีค่าเฉลี่ยเท่ากับ mean และส่วนเบี่ยงเบนมาตรฐานเท่ากับ sd
4	rlogis(n,a,b)	สร้างตัวเลขสุ่มให้มีการแจกแจงโลจิสติค n ค่า ที่มีพารามิเตอร์แอลฟาเท่ากับ a และพารามิเตอร์เบต้าเท่ากับb
5	runif(n,a,b)	สร้างตัวเลขสุ่มให้มีการแจกแจงยูนิฟอร์ม n ค่าที่มีค่าอยู่ในช่วง a ถึง b
6	matrix(nrow,ncol)	ทำการเก็บข้อมูลในรูปของเมตริกซ์โดย nrow แทนจำนวนแถวที่ต้องการและ ncol แทนจำนวนสดมภ์ที่ต้องการ
7	ginverse()	คำนวณหาค่าเมตริกซ์ผกผัน
8	sort()	เรียงค่าของข้อมูลจากน้อยไปมาก
9	mean()	คำนวณหาค่าเฉลี่ยเลขคณิตของข้อมูล
10	sum()	คำนวณหาค่าผลรวมของข้อมูล
11	pnorm	แสดงค่าความน่าจะเป็นสะสมของการแจกแจงแบบปกติ
12	qnorm	แสดงค่าอินเวอร์สของความน่าจะเป็นสะสมของการแจกแจงแบบปกติ
13	ifelse	การเลือกชุดข้อมูลตามเงื่อนไขที่กำหนดไว้
14	quantile	ค่าตำแหน่งควอนไทล์ของข้อมูล
15	round(a,dig)	การปัดเศษของตัวเลข a โดยจะใช้คู่กับ dig
16	dig	การกำหนดตำแหน่งของทศนิยมที่ต้องการ

ตาราง ข2 แสดงความหมายของสัญลักษณ์ต่างๆของโปรแกรม S-PLUS 2000

ลำดับที่	สัญลักษณ์	ความหมาย
1	p	จำนวนตัวแปรอิสระที่ในการทดลอง
2	n	ขนาดตัวอย่าง
3	sd	ค่าเบี่ยงเบนมาตรฐาน
4	bs	จำนวนรอบของการทำซ้ำในกระบวนการบูทสแตรป
5	loops	จำนวนรอบของการทำซ้ำในแต่ละสถานการณ์
6	error	ตัวแปรเก็บความคลาดเคลื่อนที่สร้างขึ้น
7	probSW0.01, probSW0.05, probSW0.10	ค่าความน่าจะเป็นของการปฏิเสธสมมติฐานว่าง/ค่าอำนาจการทดสอบจากตัวสถิติบูทสแตรปแชปิโร-วิลค์ที่ระดับนัยสำคัญ 0.01 0.05 0.10
8	probAD0.01, probAD0.05, probAD0.10	ค่าความน่าจะเป็นของการปฏิเสธสมมติฐานว่าง/ค่าอำนาจการทดสอบจากตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงที่ระดับนัยสำคัญ 0.01 0.05 0.10

ตัวอย่างโปรแกรมสำหรับการคำนวณค่าสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิงกับตัวสถิติบูทสแตรปแชปิโร-วิลค์

(*กำหนดค่าต่างๆตามขอบเขตการวิจัย***)**

```
p_3
n_10
sd_1
bs_200
loops_600
air_array(0,c(2))
airr_array(0,c(2))
```

(*ประกาศตัวแปรสำหรับการเก็บผลการวิจัยตัวสถิติบูทสแตรปแชปิโร-วิลค์***)**

```
sort.data_array(,dim=c(bs))
sort.mb1_array(,dim=c(bs))
m_array(,dim=c(bs))
s_array(,dim=c(bs))
ss_array(,dim=c(n))
h_array(,dim=c(n/2))
hh_array(,dim=c(bs))
hh2_array(,dim=c(bs))
wb_array(,dim=c(bs))
countno0.01_array(,dim=c(1,loops))
countno0.05_array(,dim=c(1,loops))
countno0.10_array(,dim=c(1,loops))
```

(*ประกาศตัวแปรสำหรับการเก็บผลการวิจัยตัวสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิง***)**

```
sort.mb2_array(,dim=c(bs))
Fxb_array(,dim=c(n))
```

```
bb_array(,dim=c(n))
spb_array(,dim=c(n))
ADb_array(,dim=c(bs))
bnbb_array(,dim=c(bs))
countnoan0.01_array(,dim=c(1,loops))
countnoan0.05_array(,dim=c(1,loops))
countnoan0.10_array(,dim=c(1,loops))
```

(*เริ่มต้นการทำงานของโปรแกรม***)**

#สร้างตัวแปรอิสระจำนวน 3 ตัว

```
x1_array(rnorm(n,0,1),dim=c(n,1))
x2_array(rnorm(n,0,1),dim=c(n,1))
x3_array(rnorm(n,0,1),dim=c(n,1))
```

#สร้างตัวแปรอิสระจำนวน 5 ตัว

```
x1_array(rnorm(n,0,1),dim=c(n,1))
x2_array(rnorm(n,0,1),dim=c(n,1))
x3_array(rnorm(n,0,1),dim=c(n,1))
x4_array(rnorm(n,0,1),dim=c(n,1))
x5_array(rnorm(n,0,1),dim=c(n,1))
```

```
for (d in 1:loops)
{#loop
```

#สร้างความคลาดเคลื่อนให้มีการแจกแจงแบบปกติ

```
error_array(rnorm(n,0,sd),dim=c(n,1))
```

#สร้างความคลาดเคลื่อนให้มีการแจกแจงแบบโลจิสติค

```
error_array(rlogis(n,0,0.5511),dim=c(n,1))
```

#สร้างความคลาดเคลื่อนให้มีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

```
alpha_0
beta_0.7071
error_array(,dim=c(n,1))
r_array(runif(n,0,1),dim=c(n,1))
for(i in 1:n)
{
    if (r[i]<0.5)
        error[i]_beta*log(2*r[i])+alpha
    else
        error[i]_-beta*log(2*(1-r[i]))+alpha
}
```

#สร้างข้อมูลของตัวแปรตาม (y) และหาค่าเศษเหลือ

```
x0_matrix(c(1),n,1)
X_matrix(c(x0,x1,x2,x3),n,p+1)
B_matrix(c(1),p+1,1)
Y_(X%*%B)+error
m1_ginverse(t(X)%*%X)
m2_t(X)%*%Y
B.hat_m1%*%m2
Y.hat_X%*%B.hat
er.hat_Y-Y.hat
```

#คำนวณค่าสถิติแชปิโร-วิลค์

```
b_array(,dim=c(5))
a_c(0.5739,0.3291,0.2141,0.1224,0.0399)
sort.nb_array(sort(er.hat),dim=c(1,n))
snb_sum((sort.nb-mean(er.hat))^2)
```

```
for(k in 1:5)
{
     b[k]_a[k]*(sort.nb[n-k+1]-sort.nb[k])
}
bnb_sum(b)^2
wnb_bnb/snb

#คำนวณค่าสถิติแอนเดอร์สัน-ดาร์ลิง
Fx_array(,dim=c(n))
bc_array(,dim=c(n))
sp_array(,dim=c(n))
for(j in 1:n)
{
     Fx[j]_pnorm(sort.nb[j],0,sd)
}
for(j in 1:n)
{
     sp[j]_(2*j)-1
     bc[j]_sp[j]*(log(Fx[j])+log(1-Fx[n+1-j]))
}
bnbc_sum(bc)
AD_((-1)*bnbc/n)-n

#การสุ่มตัวอย่างโดยใช้วิธีบูทสแตรป
temp_matrix(c(0),p+1,1)
for(i in 1:bs)
{
     set.x_c(1:n)
     prob.x_c(1/n)
```

```
ran_sample(set.x,size=n,prob.x)
er.hat.bo_array(,dim=c(n,1))
for(t in 1:n)
{
    er.hat.bo[t]_er.hat[ran[t]]
}
Y.hat.bo_Y.hat+er.hat.bo
m3_t(X)%*%Y.hat.bo
temp_m1%*%m3
Y.hat.star_X%*%temp
er.hat.star_Y.hat.bo-Y.hat.star

#คำนวณค่าสถิติบูทสแตรปแชปิโร-วิลค์
sort.data_sort(er.hat.star)
for(q in 1:n)
{
    ss[q]_((sort.data[q]-mean(sort.data))^2)
}
s_sum(ss)
m[i]_mean(sort.data)
for(j in 1:5)
{
    h[j]_a[j]*(sort.data[n-j+1]-sort.data[j])
}
hh[i]_sum(h)
hh2_sum(h)^2
air[1]_air[1]+1
air[2]_air[2]+1
wb[i]_hh2/s
```

```
write(wb[i],file="ao1.dat")
if(air[1]==1)
     write(wb[i],file="aoo1.dat")
else
     write(wb[i],file="aoo1.dat",append=T)
```

#คำนวณค่าสถิติบูทสแตรปแอนเดอร์สัน–ดาร์ลิง

```
for(ja in 1:n)
{
     Fxb[ja]_pnorm(sort.data[ja],0,sd)
}
     for(j in 1:n)
{
     spb[j]_(2*j)-1
     bb[j]_spb[j]*(log(Fxb[j])+log(1-Fxb[n+1-j]))
}
bnbb_sum(bb)
airr[1]_airr[1]+1
airr[2]_airr[2]+1
 write( ADb[i]_((-1)*bnbb/n)-n,file="bo1.dat")
if(airr[1]==1)
     write(ADb[i],file="boo1.dat")
else
     write(ADb[i],file="boo1.dat",append=T)
}
```

#คำนวณหาค่าอาณาเขตวิกฤตของตัวสถิติบูทสแตรปแชปิโร-วิลค์ที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
kp_array(,dim=c(bs))
b11_scan("aoo1.dat")
```

```
mb11_matrix(b11)
for(ww in 1:bs)
{
    kp[ww]_ifelse(mb11[ww,1]<=wnb,1,0)
}
sum.kp_sum(kp)
prob.kp_sum.kp/bs
Z0_qnorm(prob.kp)
lw1_pnorm((2*Z0)-2.326)
lw2_pnorm((2*Z0)-1.645)
lw3_pnorm((2*Z0)-1.282)
sort.mb1_sort(mb11)
no0.01_quantile(sort.mb1,lw1)
no0.05_quantile(sort.mb1,lw2)
no0.10_quantile(sort.mb1,lw3)
```

#นับจำนวนครั้งที่ปฏิเสธสมมติฐานหลักของตัวสถิติบูทสแตรปแชปิโร-วิลค์ที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
countno0.01[d]_ifelse(wnb<no0.01,1,0)
countno0.05[d]_ifelse(wnb<no0.05,1,0)
countno0.10[d]_ifelse(wnb<no0.10,1,0)
```

#คำนวณหาค่าอาณาเขตวิกฤตของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
kpp_array(,dim=c(bs))
b12_scan("boo1.dat")
mb12_matrix(b12)
for(www in 1:bs)
{
```

```
        kpp[www]_ifelse(mb12[www,1]<=AD,1,0)
    }
    sum.kpp_sum(kpp)
    prob.kpp_sum.kpp/bs
    Z00_qnorm(prob.kpp)
    lw11_pnorm((2*Z00)+2.326)
    lw33_pnorm((2*Z00)+1.645)
    lw44_pnorm((2*Z00)+1.282)
    sort.mb2_sort(mb12)
    noan0.01_quantile(sort.mb2,lw11)
    noan0.05_quantile(sort.mb2,lw22)
    noan0.10_quantile(sort.mb2,lw33)
```

#นับจำนวนครั้งที่ปฏิเสธสมมติฐานหลักของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
    countnoan0.01[d]_ifelse(AD>noan0.01,1,0)
    countnoan0.05[d]_ifelse(AD>noan0.05,1,0)
    countnoan0.10[d]_ifelse(AD>noan0.10,1,0)
  print("d")
  print(d)
}#loop
```

#คำนวณหาค่าความน่าจะเป็นที่จะปฏิเสธสมมติฐานหลักของตัวสถิติบูทสแตรปแชปิโร-วิลค์ที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
sum.total0.01_sum(countno0.01)
sum.total0.05_sum(countno0.05)
sum.total0.10_sum(countno0.10)
probSW0.01_round(sum.total0.01/loops,dig=5)
probSW0.01
```

```
probSW0.05_round(sum.total0.05/loops,dig=5)
probSW0.05
probSW0.10_round(sum.total0.10/loops,dig=5)
probSW0.10
```

#คำนวณหาค่าความน่าจะเป็นที่จะปฏิเสธสมมติฐานหลักของตัวสถิติบูทสแตรปแอนเดอร์สัน-ดาร์ลิงที่ระดับนัยสำคัญ 0.01, 0.05, 0.10

```
sum.antotal0.01_sum(countnoan0.01)
sum.antotal0.05_sum(countnoan0.05)
sum.antotal0.10_sum(countnoan0.10)
probAD0.01_round(sum.antotal0.01/loops,dig=5)
probAD0.01
probAD0.05_round(sum.antotal0.05/loops,dig=5)
probAD0.05
probAD0.10_round(sum.antotal0.10/loops,dig=5)
probAD0.10
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวณัฐฐิกาญจน์ วิรัตน์ เกิดเมื่อวันที่ 21 พฤศจิกายน พ.ศ. 2525 ที่จังหวัดนครศรีธรรมราช สำเร็จการศึกษาปริญญาตรีสถิติศาสตรบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เมื่อปี พ.ศ. 2547 เข้าศึกษาต่อในระดับปริญญามหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ. 2548