

การแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากแผนแบบการทดลองสุ่มตลอด



นายสุทธิศักดิ์ สีลาภิรักษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-53-2697-6

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

OUTLIER PROBLEM RESOLUTION FOR DATA ANALYSIS OBTAINED FROM COMPLETELY  
RANDOMIZED DESIGN

Mr.Suttisak Seelarak

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 2005

ISBN 974-53-2697-6



ศุทธิศักดิ์ สีสารักษ์ : การแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากแผนแบบการทดลองสุ่มตลอด.  
(OUTLIER PROBLEM RESOLUTION FOR DATA ANALYSIS OBTAINED FROM COMPLETELY  
RANDOMIZED DESIGN) อ. ที่ปรึกษา : รศ.ดร.สุพล คุรงค์วัฒนา, 199 หน้า. ISBN 974-53-2697-6.

วัตถุประสงค์ของการวิจัยครั้งนี้เพื่อแก้ปัญหาข้อมูลผิดปกติในการวิเคราะห์ข้อมูลที่ได้จากแผนแบบการทดลองสุ่มตลอด โดยที่ตัวแบบมีรูปแบบดังนี้  $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$  เมื่อ  $i = 1, \dots, k$  และ  $j = 1, \dots, n$  โดยที่  $Y_{ij}$  แทนข้อมูลตอบสนองที่  $j$  ได้รับวิธีทดลองที่  $i$   $\mu$  แทนค่าเฉลี่ยรวม  $\tau_i$  แทนอิทธิพลของวิธีทดลองที่  $i$   $\varepsilon_{ij}$  แทนความคลาดเคลื่อนของข้อมูลตอบสนองที่  $j$  ซึ่งได้รับวิธีทดลองที่  $i$  และ  $\varepsilon_{ij}$  มีการแจกแจงแบบปกติโดยเป็นอิสระซึ่งกันและกัน มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น  $\sigma^2$   $k$  แทนจำนวนวิธีทดลอง และ  $n$  แทนจำนวนซ้ำในแต่ละวิธีทดลอง ในการวิจัยครั้งนี้ได้ทำการสร้างข้อมูลตามขอบเขตการวิจัยด้วยโปรแกรม S-PLUS 2000 โดยกำหนดให้จำนวนวิธีทดลองเท่ากับ 3, 4 และ 5 จำนวนซ้ำในการทดลองเท่ากับ 5, 6, 7 และ 8 ให้สัมประสิทธิ์การแปรผันมีความแตกต่างกัน 3 ระดับ ได้แก่ 30% 40% และ 50% สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการแปลงข้อมูลที่เหมาะสมในการแก้ปัญหาดังกล่าวข้างต้นคือ ค่าสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูล ค่าสัดส่วนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูล ค่าสัดส่วนของการปฏิเสธสมมติฐานว่างและอำนาจการทดสอบของการทดสอบเอฟที่ระดับนัยสำคัญ 0.01 และ 0.05 ผลการศึกษาจะสรุปได้ดังนี้

1. การแปลงข้อมูลด้วยค่า  $\lambda = 0.5$  ให้ค่าสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูลสูงที่สุด และพบว่าการแปลงข้อมูลด้วย  $\lambda = 0.5$  ยังให้ค่าสัดส่วนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูลสูงที่สุด

2. การแปลงข้อมูลด้วยค่า  $\lambda = 0.5$  และ  $\lambda = 1.5$  เป็นวิธีการแปลงข้อมูลที่เหมาะสมในการแก้ปัญหาข้อมูลผิดปกติที่ได้จากแผนแบบการทดลองสุ่มตลอด และพึงควรระวังปัญหาการแจกแจงของข้อมูลตอบสนองอาจไม่เป็นการแจกแจงแบบปกติและความแปรปรวนไม่เท่ากันภายหลังการแปลงข้อมูลด้วย  $\lambda = 1.5$

3. เมื่อจำนวนค่าผิดปกติมากขึ้นพบว่า ในแต่ละสถานการณ์ค่าสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูลจะมีแนวโน้มสูงขึ้น

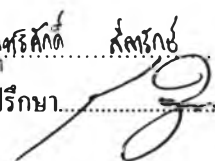
ภาควิชา ..... สถิติ.....

สาขาวิชา ..... สถิติ.....

ปีการศึกษา.....2548.....

ลายมือชื่อนิติ..... ศุทธิศักดิ์ สีสารักษ์.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....



# # 4682471026 : MAJOR STATISTICS

KEY WORD: BOX-COX TRANSFORMATION / OUTLIER / COMPLETELY RANDOMIZED DESIGN

SUTTISAK SEELARAK : OUTLIER PROBLEM RESOLUTION FOR DATA ANALYSIS  
OBTAINED FROM COMPLETELY RANDOMIZED DESIGN. THESIS ADVISOR : ASSOC. PROF.  
SUPOL DURONGWATANA, Ph.D., 199 pp. ISBN 974-53-2697-6.

The objective of this study is to solve outlier problem of the observation from completely randomized design by transformation. The fixed-effect completely randomized design model is  $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$  when  $i = 1, \dots, k$  and  $j = 1, \dots, n$  where  $Y_{ij}$  is the  $j^{\text{th}}$  response variable for the  $i^{\text{th}}$  treatment,  $\mu$  is the grand mean,  $\tau_i$  is the  $i^{\text{th}}$  treatment effect and  $\varepsilon_{ij}$  is a random error of the  $j^{\text{th}}$  response variable for the  $i^{\text{th}}$  treatments. The  $\varepsilon_{ij}$  is independently and normally distribution with mean 0 and variance  $\sigma^2$  when k is the number of treatment. In this study, the generation of response observations is done by using S-PLUS 2000 package. The data are generated with 3, 4 and 5 treatments. The number of replications for each treatment is 5, 6, 7 and 8 respectively. The coefficient of variation is the specified at 30%, 40% and 50%. The criterion used for success of outlier correction is the proportion of outlier case that reduced after transformation, the proportion of normal-distribution case data and homogeneity of variances after transformation, the proportion of null hypothesis rejection and the power of the test at 0.01 and 0.05 significance level. These criterion are measures for comparison about transformed method. The result of this study can be summarized as follow:

1. The transformation of data with  $\lambda = 0.5$  has the highest proportion of outlier case that are reduced after transforming. Moreover, the transformed data with  $\lambda = 0.5$  still sometime normally distribution and homogeneity of variances with high proportion.
2. The transformation of data with  $\lambda = 0.5$  and  $\lambda = 1.5$  are suitable for solving outlier data . However, the response data may be non-normality and heterogeneity when transformed data with  $\lambda = 1.5$
3. When the number of outliers increase, the proportion of case that are reduced after transformation increase.

Department .....Statistics.....  
Field of study....Statistics.....  
Academic year...2005.....

Student's signature.....  
Advisor's signature.....

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของรองศาสตราจารย์ ดร.สุพล คุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำปรึกษาตลอดจนแก้ไขข้อบกพร่องต่างๆด้วยดีเสมอมา จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณ เป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.กัลยา วาณิชย์บัญชา ในฐานะประธานกรรมการ และรองศาสตราจารย์ วัชรภรณ์ สุริยาภรณ์ กรรมการสอบวิทยานิพนธ์ที่กรุณาตรวจแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติ ที่ให้โอกาสทางการศึกษาและประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดาและมารดา ซึ่งสนับสนุนด้านการเงินและให้กำลังใจแก่ผู้วิจัยมาจนกระทั่งสำเร็จการศึกษาและขอขอบคุณพี่ๆที่ให้กำลังใจด้วยดีเสมอมา

ท้ายสุดนี้ ผู้วิจัยขอขอบคุณนายพรพล คงอิม เพื่อนที่คอยให้คำปรึกษาในการเขียนโปรแกรมจึงทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี และขอขอบคุณเพื่อนๆ ที่คอยให้กำลังใจด้วยดีเสมอมา นอกจากนี้ยังได้รับการสนับสนุนจากทุนอุดหนุนการวิจัยของบัณฑิตวิทยาลัยและความช่วยเหลือจากเจ้าหน้าที่จากธุรการภาควิชาสถิติ จึงขอกราบขอบพระคุณมา ณ ที่นี้

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ฐ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ข้อยกเว้นเบื้องต้น.....	3
1.4 ขอบเขตการวิจัย.....	5
1.5 เกณฑ์ในการตัดสินใจ.....	6
1.6 คำจำกัดความที่ใช้ในการวิจัย.....	6
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
1.8 วิธีดำเนินการวิจัย.....	8
2 แนวคิดและทฤษฎี.....	9
2.1 แผนแบบการทดลองสุ่มตลอด.....	9
2.2 การวิเคราะห์ความแปรปรวนหรือการทดสอบเอฟสำหรับ แผนแบบการทดลองสุ่มตลอด.....	10
2.3 การแปลงข้อมูลโดยใช้เลขยกกำลัง.....	12
2.4 การตรวจสอบข้อมูลผิดปกติในแผนแบบการทดลอง.....	13
2.5 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการแปลงข้อมูล.....	14
3 วิธีดำเนินการวิจัย.....	15
3.1 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล.....	15
3.2 แผนการดำเนินการวิจัย.....	16
3.3 ขั้นตอนในการดำเนินการวิจัย.....	17
3.3.1 สร้างความคลาดเคลื่อนให้มีการแจกแจงตามที่กำหนดไว้.....	17
3.3.2 สร้างอิทธิพลของวิธีทดลองให้มีความแตกต่างกัน.....	17

บทที่	หน้า
3.3.3 การสร้างข้อมูลตามแผนการทดลองแบบสุ่มตลอด.....	19
3.3.4 การสร้างข้อมูลตอบสนองให้เป็นค่าผิดปกติ.....	19
3.3.5 การแปลงข้อมูลจากข้อมูลที่สร้างขึ้น.....	19
3.3.6 การหาค่าสัดส่วนของจำนวนค่าผิดปกติที่ลดลง ค่าสัดส่วนการปฏิเสธ สมมติฐานว่างและอำนาจการทดสอบของการทดสอบเอฟและการ เปรียบเทียบ.....	20
3.4 แผนผังแสดงขั้นตอนการทำงานของโปรแกรม.....	21
4 ผลการวิจัย.....	24
4.1 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของจำนวนค่า ผิดปกติที่ลดลงภายหลังการแปลงข้อมูล.....	27
4.2 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของข้อมูลที่มีการ แจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูล.....	34
4.3 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธ สมมติฐานว่างของการทดสอบเอฟ.....	42
4.4 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากค่าอำนาจการทดสอบ ของการทดสอบเอฟ.....	51
5 สรุปผลการวิจัยและข้อเสนอแนะ.....	141
5.1 สรุปผลการวิจัย.....	142
5.2 ข้อเสนอแนะ.....	146
รายการอ้างอิง.....	147
ภาคผนวก .....	149
ภาคผนวก ก.....	150
ภาคผนวก ข.....	153
ภาคผนวก ค.....	178
ภาคผนวก ง.....	193
ประวัติผู้เขียนวิทยานิพนธ์.....	199



ตาราง	หน้า
1.1 ลักษณะของข้อมูลจากแผนการทดลองแบบสุ่มตลอด.....	4
2.1 ตารางวิเคราะห์ความแปรปรวนสำหรับแผนแบบทดลองสุ่มตลอดปัจจัยคงที่เมื่อไม่มีหน่วยตัวอย่างย่อยและจำนวนซ้ำของแต่ละวิธีทดลองเท่ากัน.....	10
4.1 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูลเมื่อ $k=3$ .....	31
4.2 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูลเมื่อ $k=4$ .....	32
4.3 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของจำนวนค่าผิดปกติที่ลดลงภายหลังการแปลงข้อมูลเมื่อ $k=5$ .....	33
4.4 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูลเมื่อ $k=3$ .....	39
4.5 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูลเมื่อ $k=4$ .....	40
4.6 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลังการแปลงข้อมูลเมื่อ $k=5$ .....	41
4.7 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=3$ และ ระดับนัยสำคัญ 0.01.....	45
4.8 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=3$ และ ระดับนัยสำคัญ 0.05.....	46
4.9 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=4$ และ ระดับนัยสำคัญ 0.01.....	47
4.10 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=4$ และ ระดับนัยสำคัญ 0.05.....	48
4.11 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=5$ และ ระดับนัยสำคัญ 0.01.....	49
4.12 การเปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากสัดส่วนของการปฏิเสธสมมติฐานว่างของการทดสอบเอฟเมื่อ $k=5$ และ ระดับนัยสำคัญ 0.05.....	50
4.13 การเปรียบเทียบวิธีการแปลงข้อมูลโดยการพิจารณาจากค่าอำนาจการทดสอบของการทดสอบเอฟเมื่อ $k=3$ ระดับนัยสำคัญ 0.01 และจำนวนค่าผิดปกติ 5%.....	57
4.14 การเปรียบเทียบวิธีการแปลงข้อมูลโดยการพิจารณาจากค่าอำนาจการทดสอบของการทดสอบเอฟเมื่อ $k=3$ ระดับนัยสำคัญ 0.05 และจำนวนค่าผิดปกติ 5%.....	64





ตาราง	หน้า
ข21 แสดงจำนวนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลัง การแปลงข้อมูลในการแปลงข้อมูลแต่ละวิธีเมื่อ $k=5$ $n=7$ และจำนวนค่าผิดปกติ 5%	174
ข22 แสดงจำนวนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลัง การแปลงข้อมูลในการแปลงข้อมูลแต่ละวิธีเมื่อ $k=5$ $n=7$ และจำนวนค่าผิดปกติ 10%	175
ข23 แสดงจำนวนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลัง การแปลงข้อมูลในการแปลงข้อมูลแต่ละวิธีเมื่อ $k=5$ $n=8$ และจำนวนค่าผิดปกติ 5%	176
ข24 แสดงจำนวนของข้อมูลที่มีการแจกแจงแบบปกติและความแปรปรวนเท่ากันภายหลัง การแปลงข้อมูลในการแปลงข้อมูลแต่ละวิธีเมื่อ $k=5$ $n=8$ และจำนวนค่าผิดปกติ 10%	177







รูปที่		หน้า
4.45	เปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากอำนาจการทดสอบของการทดสอบ เอฟเมื่อ $k=5$ $n=5$ ระดับนัยสำคัญ 0.05 และจำนวนค่าผิดปกติ 10%.....	137
4.46	เปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากอำนาจการทดสอบของการทดสอบ เอฟเมื่อ $k=5$ $n=6$ ระดับนัยสำคัญ 0.05 และจำนวนค่าผิดปกติ 10%.....	138
4.47	เปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากอำนาจการทดสอบของการทดสอบ เอฟเมื่อ $k=5$ $n=7$ ระดับนัยสำคัญ 0.05 และจำนวนค่าผิดปกติ 10%.....	139
4.48	เปรียบเทียบวิธีการแปลงข้อมูลโดยพิจารณาจากอำนาจการทดสอบของการทดสอบ เอฟเมื่อ $k=5$ $n=8$ ระดับนัยสำคัญ 0.05 และจำนวนค่าผิดปกติ 10%.....	140