หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลอง (simulation) ขึ้นโดยใช้โปรแกรมภาษา Pascal กับเครื่อง IBM4361 โดยมีขั้นตอนแผนการทดลองและโปรแกรมที่ใช้ในการศึกษาดังต่อไปนี้

3.1 แผนการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่างๆที่ต้องการศึกษาดังนี้

3.1.1 เลือกกลุ่มตัวอย่างจากประชากรโดยกำหนดให้ประชากรมีการแจกแจงดังต่อไปนี้

ก. การแจกแจงปกติ โดยที่พารามิเตอร์ μ เท่ากับ 1 และ ϵ^2 เท่ากับ .05 , 0.10 และ 0.15

ข. การแจกแจงแบบปกติปลอมปน โดยที่สเกลแฟกเตอร์ (C) เท่ากับ 3 และ 10 เปอร์เซนต์การปลอมปน(P) เท่ากับ 5 และ 10

ค. การแจกแจงแบบลอกนอร์มอล โดยที่พารามิเตอร์ μ เท่ากับ 1 , ϵ^2 เท่ากับ 0.22 , 0.59 และ 1.00

3.1.2 ในทุกการแจกแจงของประชากร จะศึกษาในกรณีที่มีขนาดตัวอย่างเท่ากับ 30 , 50 และ 100

3.1.3 ระดับความสัมพันธ์ของตัวแปรอิสระ

เนื่องจากริดจ์รีเกรสชันเป็นวิธีหนึ่งในการวิเคราะห์สัมประสิทธิ์การถดถอยพหุ กรณีตัวแปรอิสระมีพหุสัมพันธ์กัน ดังนั้นเราจึงกำหนดระดับความสัมพันธ์ของตัวแปรอิสระดังนี้

กรณีที่มีตัวแปรอิสระเท่ากับ 3 สนใจศึกษาเมื่อระดับความสัมพันธ์เท่ากับ (0.99) ,(0.90) และ (0.70)

กรณีที่มีตัวแปรอิสระ = 5 สนใจศึกษาเมื่อระดับความสัมพันธ์เท่ากับ (0.99,0.99),(0.90,0.90) และ (0.70,0.30)[1]

3.2 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้ คือ

1. การสร้างโปรแกรมย่อย (subroutine) สำหรับค่าความคลาดเคลื่อนตามที่ต้องการศึกษา

2. การสร้างข้อมูลของตัวแปรตาม (Y) และตัวแปรอิสระ (X) เมื่อการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติและปกติปลอมปน ตัวแปรตาม (Y) จะสร้างจากตัวแปรอิสระ (X) ซึ่งเป็นค่าคงที่ และค่าความคลาดเคลื่อนที่มีการแจกแจงตามที่ต้องการศึกษา โดยที่ตัวแปรตาม (Y) มีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระ (X) สำหรับกรณีที่การแจกแจงของค่าความคลาดเคลื่อนเป็นแบบลอกนอร์มอล จะทำการแปลงค่าตัวแปรตาม (Y) ให้มีการแจกแจงตามที่ต้องการศึกษา

3. การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชัน เมื่อการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติและปกติปลอมปน โดยการกำหนดค่า ตัวประมาณริดจ์ทั้ง 3 วิธี สำหรับกรณีที่การแจกแจงของค่าความคลาดเคลื่อนเป็นแบบลอกนอร์มอลจะทำการแปลงข้อมูลโดยการแปลงที่ใช้การยกกำลังของ Box และ Cox แล้วจึงประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชัน เมื่อกำหนดค่าตัวประมาณริดจ์ทั้ง 3 วิธี

4. การหาค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุโดยใช้วิธีริดจ์รีเกรสชันของการประมาณค่าตัวประมาณริดจ์ทั้ง 3 วิธี

---

1 กรณีที่มีจำนวนตัวแปรอิสระ=3 ระดับความสัมพันธ์ (0.99),(0.90),(0.70) หมายถึง ส.ป.ส.สหสัมพันธ์อย่างง่ายของตัวแปร X_1, X_2, X_3 กรณีที่มีจำนวนตัวแปรอิสระ = 5 ระดับความสัมพันธ์ (0.99,0.99),(0.90,0.90),(0.70,0.30) ค่าแรกในวงเล็บหมายถึง ส.ป.ส.สหสัมพันธ์อย่างง่ายของตัวแปร X_1, X_2, X_3 ค่าที่สองในวงเล็บหมายถึง ส.ป.ส.สหสัมพันธ์อย่างง่าย ของตัวแปร X_4, X_5

รายละเอียดของแต่ละขั้นตอนมีดังนี้

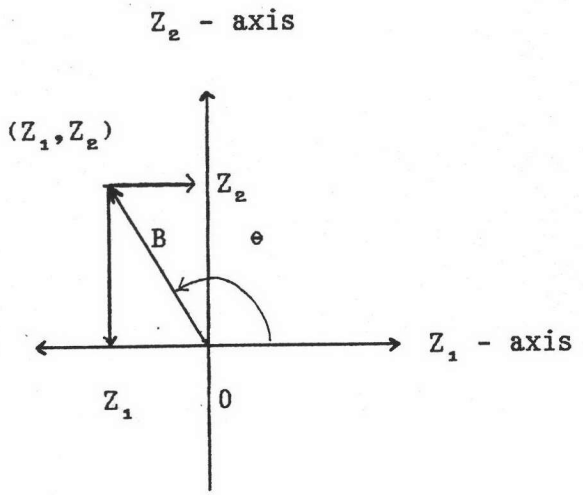
3.2.1 <u>การสร้างการแจกแจงของค่าความคลาดเคลื่อนตามที่ต้องการศึกษา</u>

การสร้างค่าความคลาดเคลื่อน ให้มีลักษณะการแจกแจงตามที่ต้องการศึกษานั้นใช้โปรแกรมภาษา Pascal กับเครื่อง IBM4361 โดยที่การสร้างลักษณะการแจกแจงต่างๆจะต้องใช้เลขสุ่ม(Random number) ซึ่งมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) เป็นพื้นฐาน

สำหรับรายละเอียดในการสร้างการแจกแจงต่างๆมีดังนี้

ก) <u>การผลิตเลขสุ่มที่มีการแจกแจงปกติ</u>

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติใช้วิธีของ Box และ Muller (ค.ศ. 1978) โดยผลิตเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน ที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 1 พร้อมๆกัน 2 ค่าดังนี้



จากรูปจะได้ว่า

(3.1) $Z_1 = B \cos(\theta)$

(3.2) $Z_2 = B \sin(\theta)$

โดยที่ $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงแบบไค-สแควร์ด้วยระดับความเป็นอิสระเท่ากับ 2 ซึ่งเทียบเท่ากับการแจกแจงแบบเอ็กซ์โปเนนเชียลที่มีค่าเฉลี่ยเท่ากับ 2 ดังนั้นจะได้รัศมี B มีค่าดังนี้

(3.3) $\quad B = (-2\ln R)^{1/2}$

โดยที่ R เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ

จากการสมมาตรของการแจกแจงแบบปกติ จะได้ว่า θ มีการแจกแจงแบบสม่ำเสมอระหว่าง 0 กับ 2π เรเดียนและรัศมี β กับ θ เป็นอิสระซึ่งกันและกัน จาก (3.1), (3.2)และ(3.3) เราสามารถสร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตราฐาน จากเลขสุ่ม 2 ชุด คือ R_1 และ R_2 กล่าวคือ

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

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

ฟังก์ชันสำหรับการจำลองแบบประชากรที่มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ คือ Procedure norm(mean.sigma) ซึ่งได้จาก

$$\text{Normal} = \mu + \sigma Z_1$$

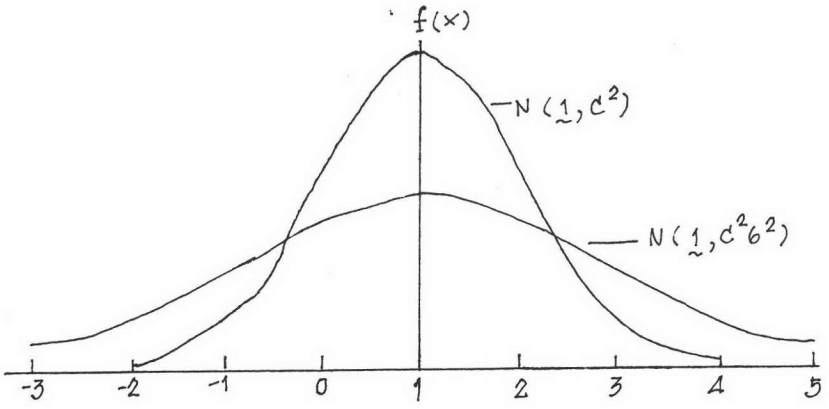
หรือ $\quad \text{Normal} = \mu + \sigma Z_2$

โดยแสดง Procedure norm(mean,sigma) ไว้ในภาคผนวก

ข) <u>การสร้างเลขสุ่มที่มีการแจกแจงปกติปลอมปน</u>

การสร้างเลขสุ่มที่มีการแจกแจงปกติปลอมปนที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตราฐานตามที่กำหนด จะใช้วิธีที่ Ramsay (ค.ศ.1977) เสนอไว้ โดยการสร้างการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูปของ

$$F(x) = (1-p)N(\mu,\sigma^2) + pN(\mu,c^2\sigma^2)$$



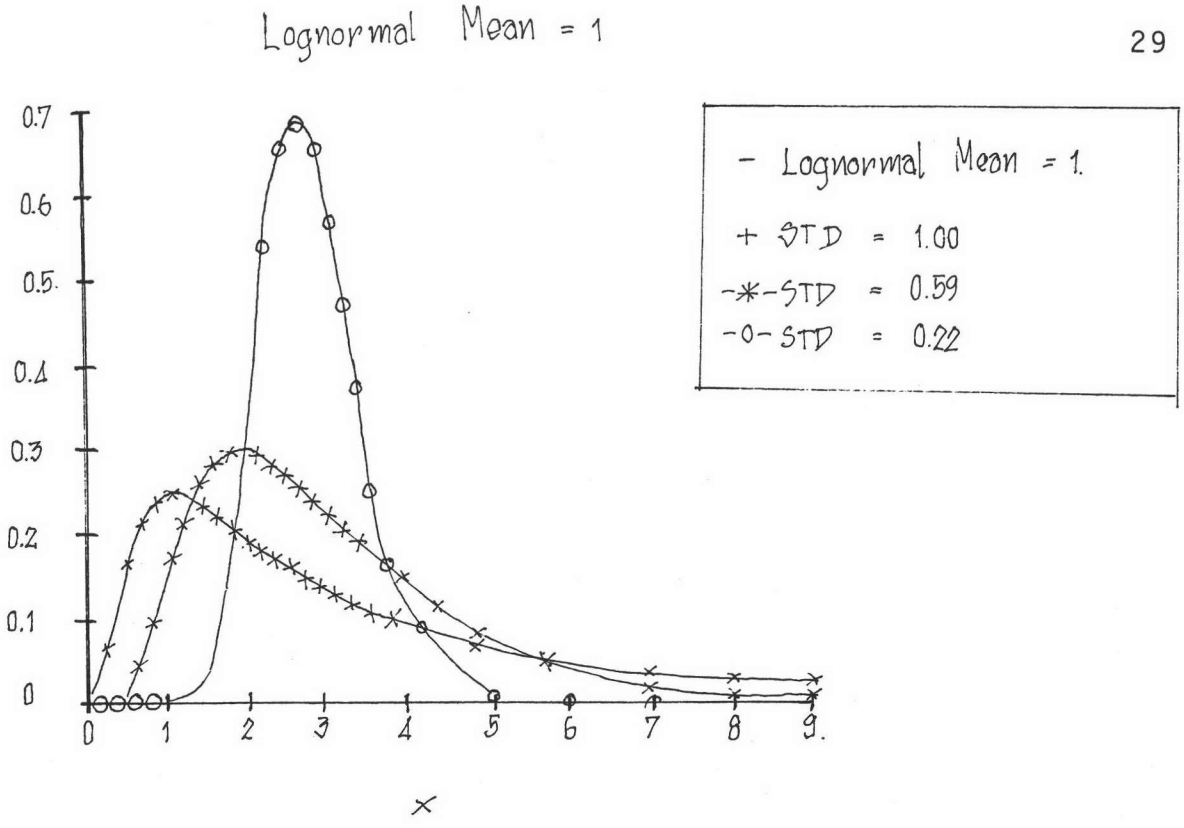
รูปที่ 3.1 แสดงเส้นโค้งการแจกแจงแบบปกติปลอมปน

โดยที่ตัวแปรสุ่ม X มาจากการแจกแจงแบบ $N(\mu,\epsilon^2)$ ด้วยความน่าจะเป็น 1-p และการแจกแจง $N(\mu,c^2\epsilon^2)$ ด้วยความน่าจะเป็น p โดยที่ μ และ ϵ^2 เป็นค่าเฉลี่ยและความแปรปรวน โดยที่ p และ c เป็นเปอร์เซนต์การปลอมปนและสเกลแฟคเตอร์ สำหรับคำสั่งในการสร้างตัวแปรให้มีการแจกแจงแบบปกติปลอมปนได้แสดงไว้ในภาคผนวก

ค) <u>การสร้างเลขสุ่มที่มีการแจกแจงแบบลอกนอร์มอล</u>

การแจกแจงแบบลอกนอร์มอลมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{1}{x\epsilon\sqrt{2\pi}} \exp\left[-\dfrac{(\ln X-\mu)^2}{2\epsilon^2}\right] & ; x>0 \ , \ \epsilon>0 \ , \ -\infty<\mu<+\infty \\ 0 & ; \text{อื่นๆ} \end{cases}$$



รูปที่ 3.2 แสดงเส้นโค้งการแจกแจงแบบลอกนอร์มอล เมื่อ μ = 1,ϵ =1.0,.59,.22

เมื่อ μ และ ϵ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y โดยที่ Y=lnx แล้ว Y จะมีการแจกแจงแบบปกติโดยมี $\exp(\epsilon^2)$ เป็น scale parameter และ μ เป็น shape parameter สำหรับค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบลอกนอร์มอล คือ $\exp(\mu + \epsilon^2/2)$ และ $\exp(2\mu + \epsilon^2).[\exp(\epsilon^2)-1]$ ตามลำดับสำหรับการสร้างเลขสุ่มที่มีการแจกแจงแบบลอกนอร์มอลได้จาก exponential ของ Procedure norm (mean.sigma) เมื่อ mean เป็นค่าเฉลี่ยและ sigma เป็นส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปกติ ดังได้แสดงไว้ในภาคผนวก

3.2.2 การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง

ในการสร้างตัวแปรอิสระ X จะสร้างให้มีความสัมพันธ์กัน ดังนี้ กรณีที่มีตัวแปรอิสระเท่ากับ 3 สนใจศึกษา ณ ระดับความสัมพันธ์ (.99),(.90),(.70) กรณีที่มีตัวแปรอิสระเท่ากับ 5 สนใจศึกษา ณ ระดับความสัมพันธ์ (.99,.99),(.90,.90),(.70,.30)

โดยการใช้วิธีการจำลองของ Wichern และ Churchill (1978:304) ทำให้สามารถสร้างตัวแปรอิสระที่มีความสัมพันธ์กันในระดับต่างๆได้

<u>ตัวอย่าง</u> ต้องการสร้างข้อมูลชุดหนึ่งให้มีตัวแปรอิสระ 5 ตัว และขนาดตัวอย่างเท่ากับ 30 ค่าของตัวแปรอิสระ ได้จากสมการ

$$X_{ij} = (1-\alpha^2)^{1/2} Z_{ij} + \alpha Z_{i6} \qquad ; j = 1,2,3 \quad i = 1,..,30$$

$$X_{ij} = (1-\alpha_*^2)^{1/2} Z_{ij} + \alpha_* Z_{i6} \qquad ; j = 4,5 \quad i = 1,...,30$$

โดยที่ $Z_{i1},..,Z_{i6}$ เป็นค่าตัวแปรอิสระที่สร้างขึ้นให้มีการแจกแจงแบบปกติที่ค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับหนึ่ง

α^2 เป็นความสัมพันธ์ระหว่างตัวแปร X_1,X_2,X_3

α_*^2 เป็นความสัมพันธ์ระหว่างตัวแปร X_4,X_5

$\alpha\alpha_*$ เป็นความสัมพันธ์ระหว่าง $X_j (j = 1,2,3)$ และ X_4 หรือ X_5

ดังนั้นหากต้องการสร้างข้อมูลชุดหนึ่งให้มีจำนวนตัวแปรอิสระเท่ากับ 5 และขนาดตัวอย่างเท่ากับ 30 โดยกำหนดให้ $\alpha^2 = 0.99$ และ $\alpha_*^2 = 0.90$ จะได้ค่าตัวแปรอิสระจากสมการ

$$X_{ij} = (1-0.99)^{1/2} Z_{ij} + (0.99)^{1/2} Z_{i6} \qquad ; j = 1,2,3 \quad i = 1,..,30$$

$$X_{ij} = (1-0.90)^{1/2} Z_{ij} + (0.90)^{1/2} Z_{i6} \qquad ; j = 4,5 \quad i = 1,..,30$$

เมื่อการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติและปกติปลอมปน เราจะสร้างข้อมูลของตัวแปรอิสระ X ซึ่งเป็นค่าคงที่ก่อนดังได้กล่าวมาแล้ว จากนั้นจึงสร้างตัวแปร Y ที่มีความสัมพันธ์เชิงเส้นกับ X ให้มีลักษณะการแจกแจงตามความคลาดเคลื่อนที่กำหนดตามรูปแบบของความสัมพันธ์เชิงเส้น คือ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่สร้างมาจากการแจกแจงแบบสม่ำเสมอ และ $\underset{\sim}{\varepsilon}$ เป็นความคลาดเคลื่อนที่มีการแจกแจงเป็นแบบต่างๆ ดังที่กล่าวมาใน (3.2.1) สำหรับการสร้าง Y นั้นจะเริ่มจากการกำหนดจำนวนตัวแปรอิสระและขนาดตัวอย่างที่ต้องการศึกษา ค่าเฉลี่ยและความแปรปรวน ลักษณะการแจกแจงของค่าความคลาดเคลื่อนพร้อมทั้งค่า $\underset{\sim}{\beta}$ ที่สร้างขึ้นมา

เมื่อการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบลอกนอร์มอล เราจะสร้างตัวแปรอิสระ X ซึ่งเป็นค่าคงที่และมีความสัมพันธ์กันดังที่กล่าวมาในตอนต้นของหัวข้อนี้ก่อนแล้วจึงสร้างตัวแปรตาม Y ให้มีการแจกแจงเป็นแบบลอกนอร์มอลตามต้องการแล้วใช้การแปลงที่ใช้การยกกำลังของ Box และ Cox เพื่อสร้างตัวแปร $Y(\lambda)$ ที่มีความสัมพันธ์เชิงเส้นกับ X ตามรูปแบบของความสัมพันธ์ คือ $Y(\lambda) = X\beta + \varepsilon$ เมื่อ β เป็นค่าพารามิเตอร์ที่สร้างขึ้นมา λ เป็นค่าพารามิเตอร์ของการแปลงที่ใช้การยกกำลังของ Box และ Cox โดยที่ ε เป็นค่าความคลาดเคลื่อนที่มีการแจกแจงเป็นแบบสมมาตรและเป็นปกติโดยประมาณ

3.3 การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชัน

3.3.1 คำนวณค่า $MSE(\beta^*(k))$ ของแต่ละวิธี โดยแทนค่า k ที่หาได้จากแต่ละวิธีลงในสมการ

$$MSE(\beta^*(k)) = \sigma^2 \sum \lambda_i/(\lambda_i+k)^2 + k^2 \sum \beta_i^2/(\lambda_i+k)^2$$

โดยค่า λ_i และ β_i เป็นค่าที่คำนวณได้จากวิธีกำลังสองน้อยที่สุด

3.3.2 เนื่องจากการวิจัยครั้งนี้กระทำซ้ำ 200 ครั้งในแต่ละสถานการณ์ ดังนั้นจึงคำนวณค่าเฉลี่ยของ MSE ของ $\beta^*(k)$ ของแต่ละวิธี นั่นคือ

$$AMSE = \sum MSE(\beta^*(k)) / 200$$

โดยที่ AMSE คือค่าเฉลี่ยของ $MSE(\beta^*(k))$ ใน 200 ครั้ง

3.3.3 หาความแตกต่างของ AMSE ของทั้ง 3 วิธี โดยจะนำการเปรียบเทียบกับวิธีที่ให้ AMSE น้อยที่สุด ดังนี้

$$DIFF = \frac{AMSE_{(i)} - AMSE_{(min)}}{AMSE_{(min)}} \times 100$$

เมื่อ DIFF(RDAMSE) คือ ความแตกต่างของ AMSE ของแต่ละวิธีเทียบกับ AMSE ของวิธีที่ให้ค่าน้อยที่สุด

$AMSE_{(min)}$ คือ AMSE ของวิธีที่ให้ค่าน้อยที่สุด

$AMSE_{(i)}$ คือ AMSE ของแต่ละวิธี

ตารางที่ 3.1 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

อันดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	ชื่อโปรแกรมที่เรียกใช้
โปรแกรมหลัก	MAIN ROUTINE	- คำนวณค่า β_i, λ_i, $MSE(\beta^*(k)), AMSE(\beta^*(k))$ RIDGE ESTIMATORS ทั้ง 4 วิธี 1. HKB 2. HK 3. TZE-SAN-LEE และ 4. BINARY SEARCH	SIM_NORM, RIDGE_METHOD VALUE_IN_MATRIC_RTN INVERSE_MATRIC_RTN CROSS_MATRIC_RTN
หมายเหตุ เมื่อความผิดพลาดมีการแจกแจงแบบปกติ, ปกติปลอมปน และ แบบลอกนอร์มอล			
PROCEDURE & FUNCTION			
1	RAND	สร้างเลขสุ่มที่การแจกแจงสม่ำเสมอ	-
2	NORMAL	สร้างเลขสุ่มที่การแจกแจงปกติ	RAND
3	SIM_NORM	สร้างเมตริกซ์ตัวแปรอิสระX	RAND
4	EIGEN	คำนวณค่า λ จากเมตริกซ์ตัวแปรอิสระ X และกำหนดค่า β	-
5	FIND_VAR_BIAS	คำนวณค่าความแปรปรวนและค่าความเอนเอียงสำหรับ RIDGE ESTIMATORS ทั้ง 4 วิธี	-
6	HKB_METHOD	คำนวณค่า k และ $MSE(\beta^*(k))$ โดยวิธี HKB	FIND_VAR_BIAS

อันดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	ชื่อโปรแกรมที่เรียกใช้
7	TZE_SAN_METHOD	คำนวณค่า k และ MSE($\hat{\beta}^*$(k)) โดยวิธี TZE	FIND_VAR_BIAS
8	HOERL_KENNARD_METHOD	คำนวณค่า k และ MSE($\hat{\beta}^*$(k)) โดยวิธี Hk	FIND_VAR_BIAS
9	BINARY_METHOD	คำนวณค่า k และ MSE($\hat{\beta}^*$(k)) โดยวิธี BINARY SEARCH	FIND_VAR_BIAS
10	RIDGE_METHOD	เรียกใช้ คำนวณค่า และแสดงผล AMSE ของ RIDGE ESTIMATORS ทั้ง 4 วิธี	HKB_METHOD TZE_SAN_METHOD HOERL_KENNARD_METHOD BINARY_METHOD
11	VALUE_IN_MATRIC_RTN	สร้างตัวแปรตาม $\underline{y}$, เมตริกซ์ X′X และ X′$\underline{y}$ 1. การแจกแจงปกติ 2. ปกติปลอมปน 3. ลอกนอร์มอล	NORM
12	DET_RTN	คำนวณค่าดีเทอร์มิแน้นท์ของเมตริกซ์	-
13	INVERSE_MATRIC_RTN	หาค่า INVERSE ของเมตริกซ์	DET_RTN
14	CROSS_MATRIC_RTN	คำนวณค่า $\hat{\beta}$ โดยวิธีกำลังสองน้อยที่สุด	-
15	PRINT_RESULT_RTN	พิมพ์ผลตารางความแปรปรวน	-

โปรแกรมทั้งหมดที่ใช้ในการวิจัยได้แสดงไว้ในภาคผนวก

แผนผังการเขียนโปรแกรม