วิธีมอนติคาร์โลเป็นสาขาหนึ่งของคณิตศาสตร์เชิงทดลอง ซึ่งหลักการของวิธีมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่ม (Random Number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

วิธีมอนติคาร์โลแบ่งเป็น 3 ขั้นตอน ดังนี้

ก. การสร้างเลขสุ่ม วิธีการสร้างตัวเลขสุ่มมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่เกิดขึ้นจะต้องมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1)และเป็นอิสระต่อกัน

ข. การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจจะใช้ในบางขั้นตอนย่อย

ค. การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มแล้วขั้นตอนต่อไป คือ การทดลอง โดยใช้กระบวนของการสุ่ม (Random Process) มากระทำซ้ำ ๆ กัน (Replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

การสร้างตัวเลขสุ่ม (Random Number)

ในการสร้างลักษณะการแจกแจงแบบต่าง ๆ จะต้องใช้เลขสุ่มเป็นพื้นฐานในการสร้าง สำหรับวิธีการสร้างตัวเลขสุ่มมีหลายวิธี แต่ในการทำวิจัยครั้งนี้จะใช้วิธีการสร้างเลขสุ่มตามวิธีของ White และ Schmidt (1975 : 421) ซึ่ง White และ Schmidt สร้างตัวเลขสุ่มโดยอาศัยหลักการเดียวกับวิธีของ Shanon (1975 : 352-356) แต่ระบุค่าบางค่าลงไปเลย ดังนี้

1. เลือกตัวเลขคี่บางตัวซึ่งน้อยกว่า 9 หลักเป็นค่าเริ่มต้น กล่าวคือ ให้ค่าเริ่มต้น IX = 973253

2. คูณตัวเลขที่กำหนดเป็นค่าเริ่มต้นด้วยค่า a ซึ่งเป็นเลขจำนวนเต็มอย่างน้อย 5 หลัก โดยกำหนดให้ a = 65539

3. คูณผลลัพธ์ในขั้นตอนที่สองด้วย 1/m โดยที่ m = 2147483647

4. จากขั้นตอนที่ 3 ก็จะได้ค่าตัวเลขสุ่มซึ่งมีค่าในช่วง (0,1)

5. กำหนดค่าเริ่มต้นของตัวแปรสุ่มถัดไปให้เท่ากับผลคูณในขั้นที่ 2

6. กระทำซ้ำ ๆ กันจากขั้นที่ 2 ถึง 5 จนกระทั่งได้ค่าตัวเลขสุ่มครบตามต้องการ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากขั้นตอนทั้ง 6 ขั้นตอนนี้สามารถสรุปเป็นโปรแกรมย่อย (Subroutine) ซึ่งเขียนเป็นภาษาฟอร์แทน IV ดังแสดงในภาคผนวก ก.

ขบวนการทั้ง 6 ขั้นตอนนี้เป็นขบวนการสร้างเลขสุ่มเพื่อที่จะนำมาสร้างตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (uniform distribution) โดยวิธี multiplicative congruential กล่าวคือ กำหนดให้

$$R_i = R_{i-1}(2^P + K) \bmod (2^{31})$$

เมื่อ P เป็นเลขจำนวนเต็มบวกที่มีค่าตั้งแต่ 3 ถึง 30

K เป็นเลขจำนวนเต็มคี่

R_0 เป็นค่าเริ่มต้น ซึ่งเป็นเลขจำนวนเต็มคี่

ตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม มีค่าระหว่าง 0 และ 1 หาได้จาก

$$U_i = \frac{R_i}{2^{31} - 1}$$

<u>ขบวนการในการสร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติ</u>

<u>ขั้นตอนที่ 1 การสร้างเลขสุ่มให้มีการแจกแจงปกติมาตรฐาน</u> (Standard Normal Distribution) ที่มีค่าเฉลี่ย = 0 และความแปรปรวน = 1 โดยใช้วิธี Direct Transformation ของ Box กับ Muller (1958) โดยจะทำการสร้างตัวแปรสุ่มครั้งละสองตัวแปร

ZTWO = SQRT (-2 X ALOG (RONE)) X SIN (2 X PI X RTWO)

ZONE = SQRT (-2 X ALOG (RONE)) X COS (2 X PI X RTWO)

โดยที่ RONE และ RTWO คือตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นตอนที่ 2 การสร้างค่าสังเกตให้มีการแจกแจงปกติ (Normal Distribution) โดยอาศัยตัวเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานที่สร้างมาจากขั้นตอนที่ 1 โดยกำหนดให้ข้อมูลมีการกระจายน้อย การกระจายปานกลาง และการกระจายมาก ซึ่งกำหนดให้ C.V. มีค่าเท่ากับ 0.05, 0.20 และ 1.00 เป็นตัวกำหนดการกระจาย SUBROUTINE ที่ใช้ในการสร้างเลขสุ่มคือ SUBROUTINE INIT โดยจะสร้างค่าสังเกตจากฟังชั่นที่ชื่อ FUNCTION NORMAL ซึ่งแสดงไว้ในภาคผนวก ก.

3.2 ขั้นตอนการวิจัย

การวิจัยครั้งนี้ได้จำลองการทดลองตามสถานการณ์ต่างๆ ขึ้นมาด้วยเครื่องคอมพิวเตอร์ IBM 4381 โดยอาศัยเทคนิคมอนติคาร์โล โดยจะกระทำซ้ำ ๆ กัน 700 ครั้ง ในแต่ละสถานการณ์ตามขั้นตอนดังนี้

3.2.1 การสร้างข้อมูล

3.2.1.1 ในการวิจัยครั้งนี้การสร้างข้อมูลตัวแปรตาม (Y) ตัวแปรอิสระ (X) และค่าความคลาดเคลื่อนโดยอาศัยรูปแบบ Regression Model ดังนี้

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad ; \underset{\sim}{\varepsilon} \sim N(0, \sigma^2)$$

โดยที่ $\underset{\sim}{Y}$ คือ เมตริกซ์ของตัวแปรตาม

X คือ เมตริกซ์ของตัวแปรอิสระ

$\underset{\sim}{\beta}$ คือ เมตริกซ์ของสัมประสิทธิ์ความถดถอยเชิงเส้น

$\underset{\sim}{\varepsilon}$ คือ เมตริกซ์ของค่าคลาดเคลื่อน

โดยจะสร้างตัวแปรอิสระ X ซึ่งเป็นค่าคงที่ก่อนแล้วจึงสร้างตัวแปร Y ที่มีความสัมพันธ์เชิงเส้นกับ X ให้มีลักษณะการแจกแจงตามความผิดผลาดที่กำหนดตามรูปแบบความสัมพันธ์คือ $Y = X\beta + \varepsilon$ เมื่อ β เป็นค่าพารามิเตอร์ที่กำหนดขึ้นมา ซึ่งการกำหนดค่า β เพื่อให้เกิดค่าคลาดเคลื่อนของตัวประมาณน้อยที่สุดค่า βควรจะมีค่าระหว่าง -1 ถึง 1 (Bolch and

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Huang) และค่า β ชุดดังกล่าวข้างล่างนี้มีหลักเกณฑ์ว่าไม่ว่าขนาดตัวแปร X เพิ่มขึ้นเท่าไรก็ตามค่า β ที่ได้จากการทดลองจะเป็นค่า β ชุดเดิมและมีค่าสหสัมพันธ์สูง โดยกำหนดให้ β มีค่าดังนี้

$\beta(1) = 0.0$

$\beta(2) = 0.5622$

$\beta(3) = -0.0935$

$\beta(4) = 0.0286$

$\beta(5) = 0.0267$

$\beta(6) = 0.0405$

$\beta(7) = -0.0703$

ซึ่ง SUBROUTINE ที่ใช้ในการสร้างข้อมูลมีดังนี้

SUBROUTINE INIT ในการสร้างข้อมูลตัวแปรอิสระ และ

SUBROUTINE DATA ในการสร้างข้อมูลตัวแปรตามและค่าความคลาดเคลื่อน

3.2.1.2 ทำการสุ่มตำแหน่งของข้อมูลที่จะสูญหายในแต่ละตัวแปรในสัดส่วนที่เท่ากันของทุกตัวแปร 5%, 10%, 15% โดยเรียกใช้

SUBROUTINE MISSNG สำหรับการสูญหายของตัวแปรอิสระ และ

SUBROUTINIS MISSY สำหรับการสูญหายของตัวแปรตาม

3.2.1.3 สร้างรหัสให้ข้อมูลทุก ๆ ตัวของทุก ๆ ตัวแปร กล่าวคือ

ก. IPLAG (J) = I

เมื่อ IPLAG (J) คือ ข้อมูลสูญหายในค่าสังเกตที่ J

I คือ จำนวนข้อมูลสูญหายในค่าสังเกตที่ J

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อ J = 1, 2, ... , N และค่าที่เป็นไปได้ของ I คือ 0 หรือ 1 หรือ 2, ... หรือ M เมื่อ M < N

ข. $\{PB_{I,J}\}$ คือ เมตริกซ์ขนาด N X M ที่แสดงตำแหน่งของข้อมูลสูญหาย

$$PB_{I,J} = \begin{cases} 0 & \text{เมื่อค่าสังเกตที่ J ของตัวแปรที่ I ไม่สูญหาย} \\ 1 & \text{เมื่อค่าสังเกตที่ J ของตัวแปรที่ I สูญหาย} \end{cases}$$

3.2.1.4 คำนวณหาจำนวนและตำแหน่งที่สูญหายของข้อมูล

เนื่องจากทุก ๆ ตัวแปรอิสระจะมีการสูญหายของข้อมูลโดยสุ่มในสัดส่วนที่เท่ากันทุกตัวแปร ดังนั้นจะต้องมีการประมาณค่าทุก ๆ ค่าที่สูญหายจึงขอยกตัวอย่าง เมื่อ M = 2 N = 5 และสัดส่วนการสูญหาย (MM) = 5% ได้ข้อมูลดังตารางข้างล่างนี้ โดยให้ ▭ เป็นตำแหน่งของข้อมูลที่สูญหาย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จำนวนข้อมูลสูญหายในแต่ละตัวแปร คือ IPLAG (J)

$$\text{โดยที่ IPLAG (J)} = \frac{\text{จำนวนข้อมูล X \% การสูญหาย}}{100}$$

ถ้าค่าที่คำนวณได้เป็นเลขทศนิยมจะใช้เลขจำนวนเต็มที่น้อยที่สุดที่มีค่ามากกว่าค่านั้น เช่น

$$\text{IPLAG (J)} = \frac{5 \text{ X } 5}{100} = 0.25$$

จะให้ IPLAG (J) = 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้นตารางที่ได้คือ

	ตัวแปรที่ 1	ตัวแปรที่ 2	IPLAG(J)
1			0
2			0
3	▭		1
4		▭	1
5			0

เมตริกซ์ที่ใช้กำหนดตำแหน่งของข้อมูลสูญหาย PB(I,J) คือ

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

3.2.2 คำนวณค่าสถิติด้วยวิธีต่าง ๆ ดังนี้

วิธีที่ 1. วิธีที่ไม่มีข้อมูลสูญหาย โดยจะประมาณสมการถดถอยและคำนวณค่าสถิติที่ต้องการ

วิธีที่ 2. วิธีที่มีข้อมูลสูญหาย โดยตัดชุดข้อมูลที่มีค่าสูญหายออกแล้วประมาณสมการถดถอยและคำนวณค่าสถิติที่ต้องการ

วิธีที่ 3. วิธีที่ใช้สมการถดถอยเชิงเส้น โดยประมาณค่าที่สูญหายจากสมการถดถอยเชิงเส้นเมื่อตัดข้อมูลสูญหายทั้งดังที่กล่าวแล้วในบทที่ 2 แล้วนำค่าประมาณไปแทนในค่าของข้อมูลที่สูญหายแล้วประมาณสมการถดถอย และคำนวณค่าสถิติที่ต้องการ

วิธีที่ 4. วิธีที่ใช้ MLE ประมาณค่าสูญหายโดยหาค่าประมาณ maximum likelihood ของแต่ละตัวแปรแล้วนำไปแทนในค่าสูญหายของแต่ละตัวแปรเพื่อประมาณสมการ

ถดถอยและคำนวณค่าสถิติที่ต้องการ

วิธีที่ 5. วิธีที่ใช้ค่าเฉลี่ย ประมาณค่าสูญหายโดยหาค่าเฉลี่ยของแต่ละตัวแปร แล้วนำค่าที่ได้ไปแทนในค่าสูญหายของแต่ละตัวแปร จากนั้นประมาณสมการถดถอยและคำนวณค่าสถิติที่ต้องการ

วิธีที่ 6. วิธีที่ใช้มัธยฐาน โดยหาค่ามัธยฐานของแต่ละตัวแปร แล้วนำค่าที่ได้แทนในค่าสูญหายของแต่ละตัวแปร จากนั้นประมาณสมการถดถอยและคำนวณค่าสถิติที่ต้องการ

3.2.3 ในการสร้างข้อมูลดังกล่าวข้างต้นได้ทำการทดสอบความเหมาะสมของตัวแปรอิสระ (X) โดยทดสอบ multicollinearity ไปด้วย

3.3 โปรแกรมที่ใช้ในการวิเคราะห์การถดถอยเชิงซ้อนโดยวิธีต่าง ๆ

SUBROUTINE ที่สำคัญที่ใช้ในการวิจัยครั้งนี้ คือ

SUBROUTINE INIT ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงปกติของตัวแปรอิสระ

SUBROUTINE DATA ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงปกติของตัวแปรตามและค่าคลาดเคลื่อน

SUBROUTINE MISSNG ใช้สุ่มตำแหน่งที่สูญหายของข้อมูลในแต่ละตัวแปรอิสระ ซึ่งถือว่าการสูญหายของข้อมูลเป็นไปอย่างสุ่ม

SUBROUTINE MISSY ใช้สุ่มตำแหน่งที่สูญหายของข้อมูลในตัวแปรตาม ซึ่งถือว่าการสูญหายของข้อมูลเป็นไปอย่างสุ่มโดยตำแหน่งการสูญหายจะเป็นตำแหน่งเดียวกับการสูญหายของตัวแปรอิสระ

SUBROUTINE REGESS ใช้ประมาณค่าสูญหายด้วยสมการถดถอยเชิงเส้น และประมาณสัมประสิทธิ์ความถดถอยและคำนวณค่าสถิติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

SUBROUTINE MLE ใช้ประมาณสัมประสิทธิ์ความถดถอยและคำนวณค่าสถิติ

SUBROUTINE MEDIAN ใช้ประมาณค่าสูญหายด้วยค่ากลางของข้อมูล และประมาณสัมประสิทธิ์ความถดถอยและคำนวณค่าสถิติ

SUBROUTINE MEAN ใช้ประมาณค่าสูญหายด้วยค่าเฉลี่ยและประมาณสัมประสิทธิ์ความถดถอยและคำนวณค่าสถิติ

นอกจากนี้ยังมี SUBROUTINE อื่น ๆ ที่เกี่ยวข้องได้แก่

SUBROUTINE RAND ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (0, 1)

SUBROUTINE NORMAL ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติ N (μ, Σ)

SUBROUTINE INVS ใช้หา INVERSE ของ MATRIX

SUBROUTINE DUPL ใช้ตรวจสอบตำแหน่งที่หายไม่ให้ซ้ำกันในแต่ละตัวแปรอิสระ

SUBROTUINE SORT ใช้ในการเรียงข้อมูลจากน้อยไปหามาก

3.4 ผังงานแสดงการประมาณสมการถดถอยเชิงซ้อนและการคำนวณค่าสถิติที่ต้องการด้วยวิธีการต่าง ๆ ดังกล่าวข้างต้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย