บทที่ 3

ระเบียบวิธีวิจัย

3.1 การสร้างเลขสุ่ม

3.1.1 การสร้างเลขสุ่มให้มีการแจกแจงแบบสม่ำเสมอ (UNIFORM DISTRIBUTION)

การสร้างเลขสุ่มโดยใช้วิธี MULTIPLICATIVE CONGRUENTIAL กล่าวคือ กำหนดให้

$$R_i = R_{i-1} (2^P + K) \bmod (2^{31}) \qquad (3.1.1)$$

เมื่อ P เป็นค่าตัวเลขตั้งแต่ 3 ถึง 30

K เป็นค่าจำนวนเต็มคี่

R_o เป็นค่าเริ่มต้นซึ่งเป็นค่าจำนวนเต็มคี่

ตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ ซึ่งมีค่าระหว่าง 0 และ 1 หาได้จาก

$$U_i = \frac{R_i}{(2^{31}-1)} \qquad (3.1.2)$$

SUBROUTINE ที่ใช้ในการสร้างตัวเลขสุ่มนี้คือ SUBROUTINE RANDU

```
      SUBROUTINE RANDU (IX, IY, YFL)
      IY  =  IX * 65539
      IF(IY) 5, 6, 6
5     IY  =  IY + 2147483647 + 1
6     YFL =  IY
      YFL =  YFL/2147483647
      RETURN
      END
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.1.2 <u>การสร้างตัวเลขสุ่มให้มีการแจกแจงแบบปกติ (UNIVARIATE NORMAL) ที่มีค่าเฉลี่ย 0 และความแปรปรวน 1</u>

สำหรับวิธีการสร้างตัวเลขสุ่มให้มีการแจกแจงแบบปกติมีหลายวิธี ในที่นี้จะเลือกใช้วิธีการแจกแจงแบบเกาซ์เซียน (GAUSSIAN DISTRIBUTION) เพราะว่าเป็นวิธีการที่ง่าย วิธีนี้อาศัยทฤษฎีลิมิตส่วนกลาง (CENTRAL LIMIT THEOREM) โดยหาผลรวมของตัวเลขสุ่ม n ค่า ก็จะได้ตัวเลขสุ่มที่มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ย μ และความแปรปรวน σ^2 โดยทั่ว ๆ ไปจะถือขนาด n ที่เหมาะสม คือค่า 10 หรือ 12 ซึ่งการวิจัยเรื่องนี้เลือก $n = 12$

SUBROUTINE ที่ใช้ในการสร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย μ และความแปรปรวน σ^2 คือ SUBROUTINE GAUSS แต่ในที่นี้ให้ชื่อใหม่เป็น SUBROUTINE NORMAL เนื่องจาก SUBROUTINE GAUSS ให้ค่า $X \sim N(\mu, S^2)$ แต่ SUBROUTINE NORMAL ให้ค่า $X \sim N(0,1)$

```
      SUBROUTINE NORMAL (XX, IX)
      A  =  0.0
      DO  400  I = 1, 12
      CALL RANDU(IX, IY, Y)
      IX  =  IY
400   A   =  A + Y
      XX  =  A - 6.0
      RETURN
      END
```

3.1.3 <u>การสร้างตัวเลขสุ่มของตัวแปรพหุให้มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย $\underline{\mu}$ และความแปรปรวนเป็น Σ</u>

จาก $Y \sim N(O, I_n)$

$Y = (Y_1, Y_2, \ldots, Y_p)'$ สร้างตัวเลขสุ่ม $Y_1, Y_2, \ldots, Y_p$ ให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน 1 โดย SUBROUTINE NORMAL จากนั้นสร้างตัวเลขสุ่มของตัวแปรพหุให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย $\underline{0}$ และความแปรปรวนร่วมเป็น Σ ได้โดย $X = CY$ ดังนั้นหา C จาก $\Sigma = CC'$ เนื่องจาก $\Sigma = R = [\rho_{ij}]$ โดยที่

$$\rho_{ij} = \begin{cases} \rho & \text{เมื่อ } i \neq j \\ 1 & \text{เมื่อ } i = j \end{cases}$$

เมื่อนำค่า Σ ไปหาค่า C จาก SUBROUTINE SETUP มีข้อแม้ว่า Σ จะต้องไม่เป็น SINGULAR METRIX โดยถือเกณฑ์ว่า ถ้า DETERMINANT ของ Σ มีค่าน้อยกว่า 0.0001 ถือว่า Σ เป็น SINGULAR METRIX จะมีผลทำให้ Σ^{-1} มีค่ามากเกินไป และจะมีผลต่อ C ซึ่งเป็น TRIANGULAR METRIX ที่ต้องการด้วย

3.2 <u>โปรแกรมที่ใช้ในการประมาณค่าสูญหายโดยวิธีต่าง ๆ</u>

เนื่องจากวิธีการประมาณที่สำคัญในการเปรียบเทียบมี 4 วิธีการ คือ วิธีที่ใช้ค่าเฉลี่ย วิธีวิเคราะห์ความถดถอยพหุเชิงเส้น วิธีวิเคราะห์ความถดถอยพหุเชิงเส้นดัดแปลง และวิธีวิเคราะห์ส่วนประกอบหลัก ซึ่ง 3 วิธีการหลังนี้เป็นวิเคราะห์ที่ค่อนข้างยุ่งยาก ดังนั้นเพื่อความถูกต้องแม่นยำ ผู้วิจัยจึงเลือกใช้ SUBROUTINE REGRSS และ SUBROUTINE FACTOR ซึ่งเป็นโปรแกรมสำเร็จรูปของ IBM (INTERNATIONAL BUSINESS MACHINE CORPORATION) ซึ่งเป็นที่ยอมรับโดยทั่ว ๆ ไปว่ามีความเชื่อถือได้ สำหรับวิธีประมาณค่าสูญหายโดยใช้ค่าเฉลี่ยนั้นได้พัฒนาโปรแกรมขึ้นมาเอง โดยให้ชื่อว่า SUBROUTINE MEAN

ดังนั้น SUBROUTINE ที่สำคัญที่ใช้ในการวิจัยครั้งนี้คือ

SUBROUTINE MULNOR ใช้สร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติของตัวแปรพหุ

SUBROUTINE MISSNG ใช้สุ่มตำแหน่งที่หายของข้อมูลในแต่ละตัวแปร ซึ่งถือว่าการหายเป็นไปโดยสุ่ม

SUBROUTINE MEAN ใช้ค่าเฉลี่ยของข้อมูลที่ไม่สูญหายของตัวแปรไปแทนค่าสูญหายของตัวแปรนั้น และคำนวณค่าสถิติที่ต้องการ

SUBROUTINE REGR ใช้ประมาณค่าสูญหายโดยวิธีวิเคราะห์การถดถอยพหุเชิงเส้น และคำนวณค่าสถิติที่ต้องการ

SUBROUTINE REGRSS ใช้หาสมการถดถอย เพื่อจะได้นำค่าสัมประสิทธิ์การถดถอยไปประมาณค่าที่สูญหายนั้น

SUBROUTINE FACT ใช้ประมาณค่าสูญหายโดยวิธีวิเคราะห์ส่วนประกอบหลัก และคำนวณค่าสถิติที่ต้องการ

SUBROUTINE FACTOR ใช้หาคอมโปเนนท์แรก เพื่อจะได้นำค่าสัมประสิทธิ์จากการแยกส่วนประกอบหลักไปประมาณค่าสูญหายนั้น

นอกจากนี้ยังมี SUBROUTINE อื่น ๆ ที่เกี่ยวข้อง ได้แก่

SUBROUTINE RANDU ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงสม่ำเสมอ U(0, 1)

SUBROUTINE NORMAL ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติ N(0, 1)

SUBROUTINE SETUP ใช้หา TRIANGULAR MATRIX (C)

SUBROUTINE MINV ใช้หา INVERSE ของ Σ

SUBROUTINE DUPL ใช้ตรวจสอบตำแหน่งที่หายไม่ให้ซ้ำกันในแต่ละตัวแปร

SUBROUTINE MULTR ใช้หาสัมประสิทธิ์การถดถอย

SUBROUTINE CORRE ใช้หาเมตริกซ์ของความสัมพันธ์ (R)

SUBROUTINE ORDER ใช้คัดลอกเมตริกซ์ความสัมพันธ์ไว้

SUBROUTINE TRACE ใช้คำนวณ cumulative percentage eigen value

SUBROUTINE EIGEN ใช้หาค่าไอเกน (eigen value)

SUBROUTINE RPTGEN ใช้หาผลรวมของค่าสถิติ

3.3 ขั้นตอนการวิจัย

3.3.1 การเตรียมข้อมูลสำหรับการวิจัย

การวิจัยครั้งนี้ได้จำลองการทดลองตามสถานการณ์ต่าง ๆ ขึ้นในเครื่องคอมพิวเตอร์ IBM 3031 ด้วยเทคนิค MONTE CARLO SIMULATION โดยจะกระทำซ้ำ ๆ กัน 1000 ครั้งในแต่ละสถานการณ์ ตามขั้นตอนนี้ คือ

3.3.1.1 สร้างข้อมูลหลายตัวแปรตามสถานการณ์ที่กำหนดโดยเรียกใช้ SUBROUTINE MULNOR

3.3.1.2 ทำการสุ่มตัวเลขตำแหน่งของข้อมูลที่จะหายในแต่ละตัวแปรในสัดส่วนที่ใกล้เคียงกันคือ 10% โดยเรียกใช้ SUBROUTINE MISSNG

3.3.1.3 สร้างสัญญาณให้ข้อมูลทุก ๆ ค่าสังเกต และข้อมูลทุก ๆ ตัว คือ IPLAG(J) = I เมื่อ I คือจำนวนข้อมูลสูญหายในค่าสังเกตที่ j ; j = 1, 2, ... , N

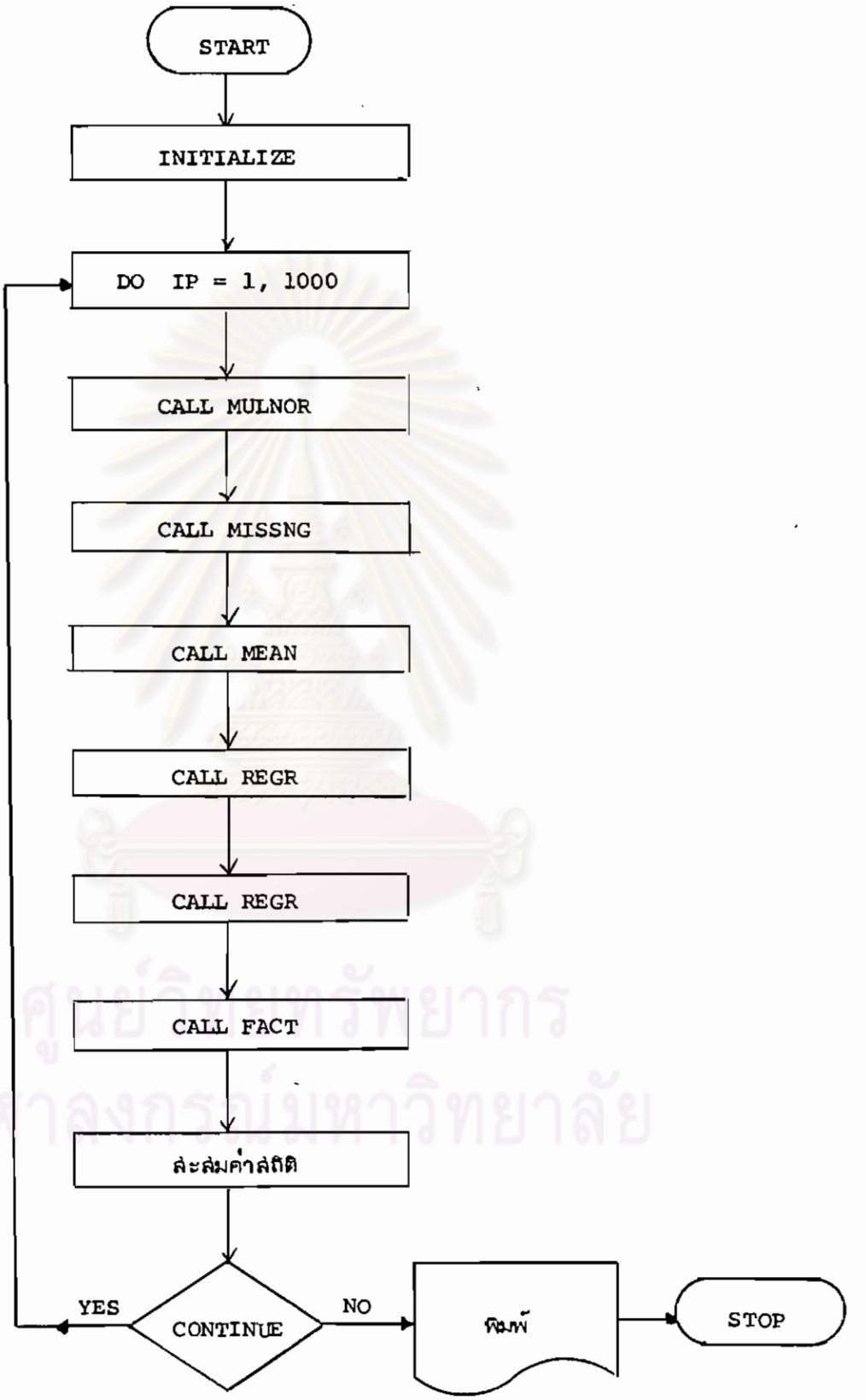
ค่าที่เป็นไปได้ของ I คือ 0 หรือ 1 หรือ 2.....หรือ P

PB(I, J) มี 2 มิติ ซึ่งขนาดเท่ากับตัวแปรสุ่มที่สร้างขึ้นคือขนาด n x p

$$PB(I, J) = \begin{cases} 0 & \text{เมื่อค่าสังเกตที่ j ของตัวแปรที่ i ไม่สูญหาย} \\ 1 & \text{เมื่อค่าสังเกตที่ j ของตัวแปรที่ i สูญหายไป} \end{cases}$$

3.3.1.4 ประมาณค่าสูญหายโดยวิธีการทั้ง 4 วิธีตามลำดับ และหาค่าสถิติที่ต้องการ ซึ่งวิธีการเหล่านี้ สามารถนำมาเขียนเป็นฝังโปรแกรมได้ดังนี้คือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.3.2 การคำนวณค่าสถิติที่ต้องการ

เนื่องจากทุก ๆ ตัวแปรจะต้องมีการหายของตัวเลขโดยสุ่มประมาณ 10% ดังนั้นจะต้องมีการประมาณค่าทุก ๆ ค่าที่สูญหาย เพื่อให้ผู้อ่านเข้าใจง่ายขึ้น จึงจะยอกตัวอย่าง เมื่อ p = 3 n = 10 และสัดส่วนการสูญหายประมาณ 30% ดังนี้

จาก (3.3.1.1) ได้ข้อมูลดังตารางข้างล่างนี้ โดยให้ ▱ เป็นตำแหน่งของข้อมูลที่สูญหายจะสามารถหา IPLAG ได้ดังนี้คือ

	ตัวแปร 1	ตัวแปร 2	ตัวแปร 3	IPLAG(J)
1	▱			1
2		▱	▱	2
3			▱	1
4				0
5				0
6	▱	▱		2
7				0
8				0
9	▱	▱	▱	3
10				0

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมตริกซ์ของ PB (I, J) จะได้

1	1	0	0
2	0	1	1
3	0	0	1
4	0	0	0
5	0	0	0
6	1	1	0
7	0	0	0
8	0	0	0
9	1	1	1
10	0	0	0

เมื่อพิจารณาจากข้อมูลจะเห็นว่ามีค่าสังเกตที่ทุก ๆ ตัวแปรหาย คือค่าสังเกตที่ 9 ดังนั้นค่าสังเกตนี้จะไม่นำมาพิจารณา เพราะฉะนั้นจะเหลือ 9 ค่าสังเกตเท่านั้น

<u>การคำนวณค่าความคลาดเคลื่อนเฉลี่ยเมื่อประมาณค่าสูญหายโดยวิธีที่ใช้ค่าเฉลี่ย (วิธีที่ 1)</u>

ประมาณค่าสูญหายโดยหาค่าเฉลี่ยของแต่ละตัวแปรได้ทั้งหมด 3 ค่าคือ ตัวแปรที่ 1 หาค่าเฉลี่ยค่าสังเกตที่ 1 ถึง 10 ยกเว้นค่าสังเกตที่ 1 6 9 ทั้งหมด 7 ค่าสังเกต แล้วนำค่าที่ได้ไปแทนในตัวแปรที่ 1 ค่าสังเกตที่ 1 6 เท่านั้น สำหรับตัวแปรที่ 2 3 ก็ในทำนองเดียวกัน จากนั้นก็คำนวณค่าสถิติที่ต้องการ

<u>การคำนวณค่าความคลาดเคลื่อนเฉลี่ยเมื่อประมาณค่าสูญหายโดยวิธีวิเคราะห์ความถดถอยพหุคูณเชิงเส้น (วิธีที่ 2)</u>

ขั้นที่ 1 ประมาณค่าสูญหายในค่าสังเกตที่มีค่าสูญหาย 1 ค่า คือ นำค่าสังเกตที่ 4 5 7 8 10 มาหาสมการถดถอย 3 สมการคือ สมการที่ 1 ให้ตัวแปรที่ 1 เป็นตัวแปรตาม ตัวแปรที่ 2 3 เป็นตัวแปรอิสระ สมการที่ 2 ให้ตัวแปรที่ 3 เป็นตัวแปรตาม ตัวแปรที่ 1 2 เป็นตัวแปรอิสระ และสมการที่ 3 ให้ตัวแปรที่ 2 เป็นตัวแปรตาม ตัวแปรที่ 1 3 เป็นตัวแปรอิสระ จากนั้นนำสมการที่ได้นี้ไปประมาณค่าสูญหายในค่าสังเกตที่ 1 3 โดยเลือกใช้สมการที่ 1 2 ตามลำดับ โดยสมมติให้ค่าเริ่มต้นสำหรับค่าที่หายไปนี้เป็นศูนย์ แล้วหาผลต่างระหว่างค่าศูนย์กับค่าที่ประมาณนี้ ถ้าผลต่างกัน > 0.1 ก็จะนำไปหาสมการถดถอยอีก 3 ชุดโดยที่มีค่าสังเกตเพิ่มขึ้นอีก 2 ค่า เป็น

ค่าสังเกตที่ 1 3 4 5 7 8 10 แล้วประมาณค่าสังเกตที่ 1 3 อีกครั้งหนึ่ง แล้วหาผลต่างระหว่างค่าประมาณก่อนหน้าและค่าประมาณที่ได้ ถ้าผลต่าง > 0.1 ก็จะมีการทำซ้ำไปเรื่อย ๆ จนกระทั่งผลต่าง ≤ 0.1

ขั้นที่ 2 ประมาณค่าสูญหายในค่าสังเกตที่มีค่าสูญหาย 2 ค่า คือ นำค่าสังเกตที่ 1 3 4 5 7 8 10 ไปหาสมการถดถอยอีก 3 สมการ เพื่อประมาณค่าสังเกตที่ 2 6 โดยมีการทำซ้ำ เช่นเดียวกับขั้นที่ 1

ขั้นที่ 3 คำนวณค่าสถิติที่ต้องการ

<u>การคำนวณค่าความคลาดเคลื่อนเฉลี่ย เมื่อประมาณค่าสูญหายโดยวิธีวิเคราะห์ความถดถอยพหุคูณเชิงเส้นตัดแปลง</u> (วิธีที่ 3)

วิธีการนี้จะคล้ายกับวิธีการที่ 2 แต่จะต่างกันที่ลักษณะข้อมูลเริ่มต้นคือ ก่อนที่จะมีการหาสมการถดถอยทั้ง 3 สมการนั้น นำค่าเฉลี่ยของแต่ละตัวแปรไปแทนค่าสูญหายในแต่ละตัวแปร จากนั้นจะมีค่าสังเกตที่สมบูรณ์ทั้งหมด 9 ค่า คือ 1 2 3 4 5 6 7 8 10 จากนั้นก็ทำเหมือนวิธีที่ 2 แล้วคำนวณค่าสถิติที่ต้องการ

<u>การคำนวณค่าความคลาดเคลื่อนเฉลี่ย เมื่อประมาณค่าสูญหายโดยวิธีวิเคราะห์ส่วนประกอบหลัก</u> (วิธีที่ 4)

ขั้นที่ 1 ประมาณค่าสูญหายในค่าสังเกตที่มีค่าสูญหาย 1 ค่าคือค่าสังเกต 1 3 โดยนำค่าข้อมูลไปทำให้มีการแจกแจงแบบปกติ $N(0, 1)$ หรือเรียกว่าการ STANDARDIZE นั่นเอง สำหรับตำแหน่งที่มีการสูญหายให้แทนด้วยค่า 0

ขั้นที่ 2 จากขั้นที่ 1 จะมีค่าสังเกตที่สมบูรณ์ 9 ค่าคือ 1 2 3 4 5 6 7 8 10 นำค่าสังเกตนี้ไปหาสัมประสิทธิ์จากการแยกตัวประกอบโดย SUBROUTINE FACT ได้ค่าสัมประสิทธิ์จากส่วนประกอบที่ 1 p ค่าคือ $q = [\, q_1\ q_2\ q_3\ \ldots q_p \,]'$

ขั้นที่ 3 ประมาณค่าสูญหายด้วย $a_r q_j$ โดยที่ $a_r = \sum_{j=1}^{p} y_{jr}\, q_j$; $r = 1, 2, \ldots, n$

ขั้นที่ 4 จากนั้นแปลงค่าที่ STANDARDIZE นี้เป็นค่าข้อมูลเพื่อประมาณค่าที่สูญหาย และจะมีการทำซ้ำถ้าผลต่างของค่าประมาณก่อนหน้าและค่าประมาณที่ได้นี้ > 0.1

ขั้นที่ 5 ประมาณค่าสูญหายในค่าสังเกตที่มีค่าสูญหาย 2 ค่าคือ ค่าสังเกตที่ 2 6 ซึ่งวิธีการจะคล้ายกับขั้นตอนที่ 1 ถึงขั้นตอนที่ 4

ขั้นที่ 6 คำนวณค่าสถิติที่ต้องการ

3.4 <u>การวิเคราะห์ความแปรปรวนกรณีมีตัวแปรตามหลายตัว</u>

นำค่า MSE จากสถานการณ์ต่าง ๆ มาทำการทดสอบผลของ n, P และ RHO ที่แตกต่างกันมีผลต่อค่า MSE ของวิธีการประมาณค่าสูญหายแต่ละวิธีหรือไม่ และทดสอบว่าวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ค่าความคลาดเคลื่อนเฉลี่ยแตกต่างกันหรือไม่ โดยใช้โปรแกรมสำเร็จรูป SPSSX คำสั่งที่ใช้คือ MANOVA ตัวอย่างโปรแกรมการเรียกใช้คำสั่ง MANOVA อยู่ในภาคผนวก

ในการวิเคราะห์ความแปรปรวนทดสอบผลของปัจจัยต่าง ๆ ว่ามีผลต่อค่าความคลาดเคลื่อนเฉลี่ยของแต่ละวิธี ถ้าหากผลการวิเคราะห์มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ ต้องทำการทดสอบต่อไปว่าวิธีการประมาณค่าสูญหายวิธีใดที่ให้ค่าความคลาดเคลื่อนเฉลี่ยที่แตกต่างกันโดยใช้ UNIVARIATE F-TEST และในการวิเคราะห์ความแปรปรวนทดสอบว่าวิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ค่าความคลาดเคลื่อนเฉลี่ยแตกต่างกันหรือไม่ ถ้าหากผลการวิเคราะห์มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ ต้องทำการทดสอบต่อไปว่าวิธีการใดที่แตกต่างกันบ้างโดยใช้ LSD (Least significant difference)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย