บทที่ 3

วิธีดำเนินการวิจัย

ในการศึกษาเพื่อวิเคราะห์ปัจจัยที่มีต่อการกำหนดอัตราเบี้ยประกันสุขภาพ มีขั้นตอนการวิจัยมีดังนี้

1.คำนวณหาขนาดของปัจจัยที่มีอิทธิพลต่อการคำนวณเบี้ยประกันสุขภาพ

2.คำนวณอัตราเบี้ยประกันสุขภาพ

3.เปรียบเทียบอัตราเบี้ยประกันสุขภาพที่คำนวณได้จากวิธีต่าง ๆ

การคำนวณหาขนาดของปัจจัยที่มีอิทธิพลต่อการคำนวณเบี้ยประกันสุขภาพ

การคำนวณหาขนาดของปัจจัยที่มีอิทธิพลต่อการคำนวณเบี้ยประกันสุขภาพ ในการวิจัยครั้งนี้จะเสนอวิธีการ 3 วิธีคือ

1. วิธีวิเคราะห์ความแปรปรวนและทดสอบความแตกต่างระหว่างค่าเฉลี่ยภายหลังการทดสอบความแปรปรวน

2. วิธีวิเคราะห์ปัจจัย

3. วิธีวิเคราะห์การถดถอยพหุคูณ

กำหนดตัวแปรต่าง ๆ สำหรับการวิเคราะห์ข้อมูล ดังนี้

Y หมายถึง ค่าใช้จ่ายในการรักษาพยาบาล

X_1 หมายถึง ช่วงอายุ 15-35 ปี

X_2 หมายถึง ช่วงอายุ 36-40 ปี

X_3 หมายถึง ช่วงอายุ 41-45 ปี

X_4 หมายถึง ช่วงอายุ 46-50 ปี

X_5 หมายถึง ช่วงอายุ 51-55 ปี

X_6 หมายถึง ช่วงอายุ 56-59 ปี

X_7 หมายถึง อาชีพชั้น 1

X_8 หมายถึง อาชีพชั้น 2

X_9 หมายถึง อาชีพชั้น 3

X_{10} หมายถึง เพศหญิง

X_{11} หมายถึง เพศชาย

SEX หมายถึง ปัจจัยเพศ

CLASS หมายถึง ปัจจัยอาชีพ

AGE หมายถึง ปัจจัยอายุ

โดยที่ลักษณะของตัวแปรดังกล่าวข้างต้น มีลักษณะเป็นข้อมูลเชิงคุณภาพ (qualitative)

-การกำหนดค่าสำหรับตัวแปร $X_1 - X_{11}$

จะกำหนดค่าหนึ่งของตัวแปรนั้นเป็น 0 และอีกค่าหนึ่งกำหนดให้เป็น 1 เช่น ถ้าผู้เอาประกันภัย เป็นเพศชาย อาชีพชั้น 1 อายุในช่วง 15-35 ปี ค่าของตัวแปร $X_1 = 1$ $X_7 = 1$ และ $X_{11} = 1$ ส่วนตัวแปรอื่น ๆ จะมีค่าเป็น 0 ถ้าผู้เอาประกันภัยเป็นเพศหญิงอาชีพชั้น 2 อายุในช่วง 55-59 ปี ค่าของตัวแปร $X_6 = 1$ $X_8 = 1$ และ $X_{10} = 1$ ส่วนตัวแปรอื่น ๆ จะมีค่าเป็น 0 เป็นต้น

-การกำหนดค่าสำหรับตัวแปร SEX CLASS และ AGE

SEX : กำหนดให้ เพศหญิง มีค่าเป็น 0

เพศชาย มีค่าเป็น 1

CLASS : กำหนดให้ อาชีพชั้น 1 มีค่าเป็น 1

อาชีพชั้น 2 มีค่าเป็น 2

อาชีพชั้น 3 มีค่าเป็น 3

AGE : กำหนดให้ อายุ 15-35 ปี มีค่าเป็น 1

อายุ 36-40 ปี มีค่าเป็น 2

อายุ 41-45 ปี มีค่าเป็น 3

อายุ 46-50 ปี มีค่าเป็น 4

อายุ 51-55 ปี มีค่าเป็น 5

อายุ 56-59 ปี มีค่าเป็น 6

1. <u>วิธีวิเคราะห์ความแปรปรวนและทดสอบความแตกต่างระหว่างค่าเฉลี่ยภายหลังการทดสอบความแปรปรวน</u>

การวิเคราะห์ความแปรปรวน เป็นวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ย (Mean)

ของกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไป ซึ่งเป็นการสรุปว่ากลุ่มตัวอย่างต่าง ๆ มีค่าเฉลี่ยต่างกันหรือไม่ ในการวิจัยครั้งนี้จะใช้วิธีวิเคราะห์ความแปรปรวนจำแนกทางเดียว (One-way Analysis of variance) มีขั้นตอนดังนี้

1.1 จำแนกข้อมูลให้อยู่ในรูปต่อไปนี้

กลุ่มที่ 1	กลุ่มที่ 2	กลุ่มที่ 3	....	กลุ่มที่ k
x_{11}	x_{21}	x_{31}	....	x_{k1}
x_{12}	x_{22}	x_{32}	....	x_{k2}
x_{13}	x_{23}	x_{33}	....	x_{k3}
:	:	:	:	:
x_{1n_1}	x_{2n_2}	x_{3n_3}	....	x_{kn_k}

โดยที่

x_{ij} แทน ค่าสังเกตของสิ่งทดลอง (treatment) ที่ i สำหรับตัวอย่างที่ j ซึ่งในที่นี้หมายถึงค่าใช้จ่ายในการรักษาพยาบาลเนื่องจากการประกันสุขภาพของสิ่งทดลองที่ i ผู้เอาประกันภัยคนที่ j

i แทน สิ่งทดลอง หรือ กลุ่ม (treatment) ในที่นี้คือคุณลักษณะของผู้เอาประกันภัยประกอบด้วย เพศ อาชีพ และอายุ โดยจำแนก

ปัจจัยเพศ เป็น 2 ระดับ คือ เพศชายและเพศหญิง

ปัจจัยอาชีพ เป็น 3 ระดับ คือ อาชีพชั้น 1 อาชีพชั้น 2 และอาชีพชั้น 3

ปัจจัยอายุ เป็น 6 ระดับ คือ อายุ 15-35 ปี อายุ 36-40 ปี อายุ 41-45 ปี อายุ 46-50 ปี อายุ 51-55 ปี และอายุ 56-59 ปี

กลุ่มทั้งหมดซึ่งเกิดจากการแบ่งระดับของแต่ละคุณลักษณะทั้ง 3 ข้างต้นจะเท่ากับ 36 กลุ่ม คือ

กลุ่ม 1 (G1) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 2 (G2) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 3 (G3) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 4 (G4) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 5 (G5) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 6 (G6) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 1 อายุอยู่ในช่วง 56 - 59 ปี

กลุ่ม 7 (G7) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 8 (G8) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 9 (G9) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 10 (G10) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 11 (G11) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 12 (G12) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 2 อายุอยู่ในช่วง 56 - 59 ปี

กลุ่ม 13 (G13) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 14 (G14) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 15 (G15) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 16 (G16) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 17 (G17) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 18 (G18) แทน ผู้เอาประกันภัยเพศหญิงอาชีพชั้น 3 อายุอยู่ในช่วง 56 - 59 ปี

กลุ่ม 19 (G19) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 20 (G20) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 21 (G21) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 22 (G22) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 23 (G23) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 24 (G24) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 1 อายุอยู่ในช่วง 56 - 59 ปี

กลุ่ม 25 (G25) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 26 (G26) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 27 (G27) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 28 (G28) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 29 (G29) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 30 (G30) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 2 อายุอยู่ในช่วง 56 - 59 ปี

กลุ่ม 31 (G31) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 15 - 35 ปี

กลุ่ม 32 (G32) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 36 - 40 ปี

กลุ่ม 33 (G33) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 41 - 45 ปี

กลุ่ม 34 (G34) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 46 - 50 ปี

กลุ่ม 35 (G35) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 51 - 55 ปี

กลุ่ม 36 (G36) แทน ผู้เอาประกันภัยเพศชายอาชีพชั้น 3 อายุอยู่ในช่วง 56 - 59 ปี

j แทน ค่าสังเกตที่ 1,2,3,...,n_i

n_i แทน จำนวนตัวอย่างที่นำมาในแต่ละกลุ่มทดลอง ในที่นี้ คือ จำนวนผู้เอาประกันภัยในกลุ่มที่ i

N แทน จำนวนตัวอย่างทั้งหมดในกลุ่มทดลองคือ $N = n_1 + n_2 + n_3 + ... + n_k$ ในที่นี้คือ จำนวนผู้เอาประกันภัยทั้งหมด

1.1.1 ตัวแบบเชิงคณิตศาสตร์

ค่า x_{ij} สามารถแสดงในรูปของผลบวกขององค์ประกอบต่าง ๆ ดังนี้

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1, 2, ..., k$$
$$j = 1, 2, ..., n_i$$

μ แทน ค่าเฉลี่ยของประชากรทั้งหมด เรียกว่า Grand mean

τ_i แทน ผลกระทบจากสิ่งทดลองที่แตกต่างกัน (treatment effect)

ε_{ij} แทน ความคลาดเคลื่อนเชิงสุ่ม (random error) ที่เกิดจากสิ่งทดลองที่ i หน่วยทดลองที่ j

μ_i แทน ค่าเฉลี่ยจริงของสิ่งทดลองที่ i ทั้งหมด ซึ่ง $\mu_i = \mu + \tau_i$ ดังนั้น ตัวแบบเชิงคณิตศาสตร์อาจกำหนดได้อีกแบบหนึ่งดังนี้ คือ

$$x_{ij} = \mu_i + \varepsilon_{ij}$$

1.1.2 ข้อสมมติ

1. ε_{ij} มีการแจกแจงเป็นอิสระต่อกันและมีการแจกแจงแบบปกติด้วย ค่าเฉลี่ย 0 ความแปรปรวน σ_ε^2 (หรือ $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$)

2. τ_j เป็นผลกระทบชนิดคงที่ (fixed effect) และ $\sum_{i=1}^{k} \tau_i = 0$ กล่าวคือ กำหนดว่าสิ่งทดลองทั้งหมดที่ศึกษาแต่ละสิ่งทดลองประกอบกันเป็นประชากรที่สนใจ

1.1.3 การทดสอบสมมติฐาน

จากสมมติฐานที่กำหนดไว้ว่า

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

ถ้าผลสรุปได้ว่ายอมรับสมมติฐานนี้แสดงว่าค่าเฉลี่ย μ_i แต่ละตัวไม่แตกต่างกันและต่างมีค่าเท่ากับ μ ถ้าพิจารณาจากสมการ $\mu_i = \mu + \tau_i$ แสดงว่า treatment แต่ละ treatment ไม่มีผลต่อค่าสังเกต นั่นคือ สมมติฐานอาจจะกำหนดได้อีกรูปแบบหนึ่งคือ

$H_0 : \tau_1 = \tau_2 = ... = \tau_k = 0$

$H_1 : \tau_i$ อย่างน้อย 1 ตัวไม่เป็น 0

1.2 การวิเคราะห์ความแปรปรวน

กำหนดให้

$$x_{i.} = \sum_{j=1}^{n_i} x_{ij}$$

$$x_{..} = \sum_{i=1}^{k} x_{i.} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$$

$$\bar{x}_{i.} = \frac{x_{i.}}{n_i}$$

$$\bar{x}_{..} = \frac{x_{..}}{\sum_{i=1}^{k} n_i}$$

ฉะนั้น ค่าความแปรปรวนที่เป็นผลมาจากความแตกต่างระหว่างสิ่งทดลอง จึงมีค่าเท่ากับ $\sum_{i=1}^{k}(\bar{x}_{i.} - \bar{x}_{..})^2 / (k-1)$ และเมื่อคูณด้วยจำนวนข้อมูลในแต่ละสิ่งทดลองก็จะให้ค่าประมาณของ σ^2 วิธีหนึ่ง คือ $\sum_{i=1}^{k} n_i(\bar{x}_{i.} - \bar{x}_{..})^2 / (k-1)$ ในทำนองเดียวกัน ค่าประมาณของ σ^2 ที่ได้จากค่าความแปรปรวนของค่าสังเกตโดยตรงจะคำนวณได้ดังนี้ คือ $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2 / \sum_{i=1}^{k}(n_i - 1)$

ผลการแยกความแปรปรวนทั้งสองของ x_{ij} ออกเป็นส่วน ๆ ตามแหล่งที่มานี้ เพื่อความสะดวกในการวิเคราะห์จะแสดงในรูปตารางการวิเคราะห์ความแปรปรวน (ANOVA) ดังนี้

ตารางการวิเคราะห์ความแปรปรวน

แหล่งของความแปรปรวน (source of variation =SV)	องศาความเป็นอิสระ(degree of freedom = df)	ผลบวกกำลังสอง (sum of square = SS)	ค่าเฉลี่ยกำลังสอง (mean square = MS)	F
สิ่งทดลอง (treatment)	k-1	$\sum_{i=1}^{k} n_i(\bar{x}_{i.} - \bar{x}_{..})^2$	$MS(T) = \frac{SS(T)}{k-1}$	$\frac{MS(T)}{MS(B)}$
ความคลาดเคลื่อน (error)	$\sum_{i=1}^{k}(n_i - 1)$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{ij} - \bar{x}_{i.})^2$	$MS(B) = \frac{SS(B)}{\sum_{i=1}^{k}(n_i - 1)}$	
ผลรวม (total)	$\sum_{i=1}^{k} n_i - 1$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{ij} - \bar{x}_{..})^2$		

สมมติฐานที่ใช้ในการทดสอบ คือ

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

$H_1 : \mu_i \neq \mu_j \quad (i \neq j) \quad I, j = 1, 2, \ldots, k$

หรือ

H_0 : ไม่มีความแตกต่างระหว่างค่าเฉลี่ยของแต่ละประชากร

H_1 : มีความแตกต่างระหว่างค่าเฉลี่ยของแต่ละประชากรอย่างน้อย 2 ประชากร

การตัดสินใจขึ้นอยู่กับค่าสัดส่วน F ที่คำนวณจาก (MS(T)/MS(E)) กับ อัตราส่วน F จากตารางมาตรฐาน กล่าวคือ ถ้า F ที่คำนวณได้มีค่ามากกว่า F จากตารางที่ระดับนัยสำคัญ α และ $df = \sum_{i=1}^{k} n_i - 1, \sum_{i=1}^{k} (n_i - 1)$ จะปฏิเสธสมมติฐาน H_0

1.3 ทดสอบค่าเฉลี่ยภายหลังการปฏิเสธสมมติฐาน

สำหรับกรณีที่การตัดสินใจสรุปว่าได้มีการปฏิเสธสมมติฐาน H_0 นั้น จะสรุปได้เพียงว่ามีความแตกต่างกันระหว่างค่าเฉลี่ยของประชากรอย่างน้อย 2 ประชากร แต่ไม่สามารถบอกได้ว่าประชากรคู่ใดที่มีค่าเฉลี่ยแตกต่างกันก็อาจทำได้โดยการทดสอบแบบคู่ (Pairwise Test) หรือการเปรียบเทียบพหุคูณ (Multiple Comparisons) ซึ่งการทดสอบดังกล่าวจะมีการคำนวณหาค่าผลต่างสัมบูรณ์ (Absolute Difference) ของค่าเฉลี่ยจากตัวอย่างทุกคู่ที่เป็นไปได้คือ $\left|\bar{x}_i - \bar{x}_j\right|$ และจะนำค่าที่ได้แต่ละค่าเปรียบเทียบกับค่าสถิติ ซึ่งในการวิจัยครั้งนี้จะใช้วิธี Least Significant Difference (LSD)

วิธี LSD เป็นวิธีการทดสอบความแตกต่าง ระหว่างค่าเฉลี่ยที่เรียกว่าผลต่างนัยสำคัญ (Least Significant Difference) ซึ่งมีค่าสถิติ LSDที่คำนวณได้ดังนี้

$$LSD(\alpha) = t_{\alpha,\gamma} \cdot s_{\bar{d}}$$

$$s_{\bar{d}} = \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

โดยที่ $\gamma = n - k$

$n_i \neq n_j$

$LSD(\alpha)$ คือ ค่าผลต่างนัยสำคัญที่คำนวณได้สำหรับประชากรกลุ่มที่ i และ j

MSE คือ ค่า Mean Square Error จากตารางวิเคราะห์ความแปรปรวน

$t_{\alpha,\gamma}$ คือ ค่าสถิติ t จากตาราง t ที่ระดับนัยสำคัญ α และ $df = \sum_{i=1}^{k}(n_i - 1)$

$LSD(\alpha) = t_{\alpha,\gamma} \cdot s_{\bar{d}}$

การทดสอบสมมติฐานสำหรับผลต่างของค่าเฉลี่ยทุกคู่มีข้อกำหนดดังนี้

สำหรับสมมติฐาน $H_0: \mu_i = \mu_j$

$H_1: \mu_i \neq \mu_j \quad (i \neq j) \quad i,j = 1, 2, ..., k$

จะปฏิเสธสมมติฐานเมื่อ $\left|\bar{x}_i - \bar{x}_j\right| > LSD(\alpha)$ สำหรับประชากรที่ i และ j

1.4 พิจารณาค่าใช้จ่ายในการรักษาพยาบาลของการประกันสุขภาพเฉลี่ยจำแนกตามปัจจัยที่ได้จากข้อ 1.3 แล้วคำนวณหาสัดส่วนเพื่อหาค่าอิทธิพล (ค่าน้ำหนัก) ในแต่ละปัจจัย โดยใช้ข้อมูลปัจจัยใดปัจจัยหนึ่งเพียงปัจจัยเดียวเป็นตัวฐาน (ในที่นี้จะใช้กลุ่มผู้เอาประกันภัยเพศชาย อาชีพชั้น 1 อายุในช่วง 15-35 ปี เนื่องจากกลุ่มดังกล่าวเป็นกลุ่มที่ได้มีการตั้งสมมติฐานว่าจะมีค่าอิทธิพลของปัจจัยต่ำสุด)

2. <u>วิธีวิเคราะห์ปัจจัย</u>

วิธีวิเคราะห์ปัจจัยโดยใช้หลักการวิเคราะห์องค์ประกอบหลักเป็นวิธีที่ให้แบบแผนความสัมพันธ์ระหว่างปัจจัยต่างๆ และแยกปัจจัยเหล่านั้นเป็นกลุ่มๆ ตามความสัมพันธ์ที่เกี่ยวข้อง ตัวประกอบจะให้ค่าถ่วงน้ำหนักซึ่งสามารถแปลงคะแนนดิบของปัจจัยนั้นๆ ให้อยู่ในรูปค่าถ่วงน้ำหนัก เพื่อจัดลำดับความสัมพันธ์ให้ชัดเจนถูกต้องยิ่งขึ้น

<u>รูปแบบความสัมพันธ์ของตัวแปรในองค์ประกอบ</u>

หลักการของวิธีวิเคราะห์องค์ประกอบโดยทั่วไป คือ หาตัวแปรใหม่ที่แสดงความสัมพันธ์เชิงเส้นตรง (Linear Combination) ของตัวแปรเดิมหลายๆ ตัวแปร

ขั้นตอนการคำนวณหารูปแบบของตัวแปรในองค์ประกอบ (Component) ต่างๆ

1) กำหนดตัวแปรที่ใช้ในการศึกษาวัดความสำคัญของปัจจัยที่มีผลต่อการคำนวณอัตราเบี้ยประกันสุขภาพ ให้ X เป็นตัวแปรต่าง ๆ ที่ใช้วัด

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots\cdots\cdots & x_{1k} \\ x_{21} & x_{22} & \cdots\cdots\cdots & x_{2k} \\ : & : & \cdots\cdots\cdots & : \\ : & : & \cdots\cdots\cdots & : \\ x_{n1} & x_{n2} & \cdots\cdots\cdots & x_{nk} \end{bmatrix}$$

โดยที่ x_{ij} แทนค่าสังเกตของตัวแปรที่ i สำหรับตัวแปรที่ j

2) หาค่าความแปรปรวนร่วม (Covariance) ของตัวแปรทั้งหมด

S = E [(X-E(X)) (X-E(X))]

S = covariance matrix ของ X

3) หาค่าคาแรคเตอรีสติครูท (Characteristic root; λ)

$$| S-\lambda I | = 0$$

I = Identity matrix

$$\lambda = [\lambda_1 \quad \lambda_2 \quad .....\lambda_k \ ]$$

λ_i = คาแรคเตอรีสติครูทที่ i ของ s

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 .... \geq \lambda_k$$

4) หาค่าคาแรคเตอรีสติคเวกเตอร์ (Characterist Vectors ; $\underset{\sim}{a}$)

$$S\underset{\sim}{a} = \lambda \underset{\sim}{a}$$

a_i = คาแรกเตอรีสติคเวกเตอร์ที่ i ที่สอดคล้องกับคาแรคเตอรีสติครูท λ_i

$$a_i = \begin{bmatrix} a_1 \\ a_2 \\ : \\ : \\ a_k \end{bmatrix}$$

5) หา normalized ของคาแรคเตอริสติกเวคเตอร์

$$normalized = \begin{bmatrix} a_1 \big/ \sqrt{lenght} \\ a_2 \big/ \sqrt{lenght} \\ : \\ : \\ a_k \big/ \sqrt{lenght} \end{bmatrix}$$

$$lenght = a_1^2 + a_2^2 + \dots\dots\dots\dots + a_k^2$$

นำค่า normalized ของคาแรคเตอริสติกเวคเตอร์ มาเรียงกันจะได้พรินสิปอลคอมโพแนนท์เวคเตอร์ (Principal Component Vector)

6) หมุนแกนองค์ประกอบให้แกนตั้งฉากซึ่งกันและกัน (Orthogonal) จะได้องค์ประกอบที่เป็นอิสระต่อกันค่าที่ได้จากการหมุนแกนคือค่าองค์ประกอบ (Principal Component) ที่จะนำไปใช้ในการวิเคราะห์เพื่อหาตัวแปรที่สำคัญโดยพิจารณาจากองค์ประกอบที่มีค่าความแปรปรวนสูงสุด

องค์ประกอบที่ i (Principal Component ที่ i)

$$Y_i = \underset{\sim}{a}_i' X_i = \sum_{j=1}^{k} a_{ij} X_j$$

$$Y_i = a_{i1} X_1 + a_{i2} X_2 + \dots\dots\dots\dots\dots + a_{ik} X_k$$

จะเลือกองค์ประกอบที่ให้ความแปรปรวนสูงสุด โดยพิจารณา

$$\underset{\sim}{a}_i' \underset{\sim}{a}_i = 1$$

$$\underset{\sim}{a}_i' \underset{\sim}{a}_j = 0 \quad ; \quad i \neq j$$

Y_1 จะมีค่าวาเรียนซ์สูงที่สุด

Y_2 จะมีค่าวาเรียนซ์รองลงมา

.

.

Y_k จะมีค่าวาเรียนซ์ต่ำสุด

ดังนั้น จึงใช้องค์ประกอบที่ 1 (Principal Component ที่ 1)

รูปแบบที่ใช้ $Y = a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k$

หรือ $Y = a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k$

ซึ่งค่า $a_1, a_2, \dots, a_k$ ที่ได้จะแสดงค่าน้ำหนักของอิทธิพลในแต่ละปัจจัย

3. <u>วิธีวิเคราะห์การถดถอยพหุคูณ</u>

3.1 หารูปแบบความสัมพันธ์ระหว่างตัวแปรต่าง ๆ

เนื่องจากอัตราเบี้ยประกันสุขภาพเกี่ยวข้องกับค่าใช้จ่ายในการรักษาพยาบาล ดังนั้นจึงจะกำหนดน้ำหนักของตัวแปรต่าง ๆ ที่จะนำมาสร้างแบบจำลองด้วยการใช้วิธีวิเคราะห์การถดถอยพหุคูณ โดยให้ค่ารักษาพยาบาลในแต่ละปีของข้อมูลเป็นตัวแปรตาม ซึ่งมีรูปแบบดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

โดยที่ β_i เป็นค่าสัมประสิทธิ์การถดถอยพหุคูณไม่ทราบค่าตัวแบบ $i = 0,1,\dots,k$ และถ้าหากมีจำนวนข้อมูล n ตัว จะได้รูปของข้อมูลเป็นดังนี้

ข้อมูล	ตัวแปรตาม	ตัวแปรอิสระที่ 1	ตัวแปรอิสระที่ 2	......	ตัวแปรอิสระที่ k
1	y_1	x_{11}	x_{21}	......	x_{k1}
2	y_2	x_{12}	x_{22}	......	x_{k2}
3	y_3	x_{13}	x_{23}	......	x_{k3}
:	:	:	:	......	:
n	y_n	x_{1n}	x_{2n}	......	x_{kn}

จากข้อมูลข้างต้น นำมาเขียนในรูปของสมการเมตริกซ์จะได้เป็น

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}$$

หรือ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

โดยที่ $\underset{\sim}{Y}$ คือ ข้อมูลของตัวแปรตาม โดยมีเวกเตอร์อันดับที่ n×1

X คือ ข้อมูลของตัวแปรอิสระ โดยมีเวกเตอร์อันดับที่ n× (k+1)

$\underset{\sim}{\beta}$ คือ สัมประสิทธิ์ของการถดถอย โดยมีเวกเตอร์อันดับที่ (k+1)×1

$\underset{\sim}{\varepsilon}$ คือ ค่าความคลาดเคลื่อนของค่าประมาณจากค่าจริง โดยมีเวกเตอร์อันดับที่ n×1

ภายใต้ข้อสมมติเกี่ยวกับการแจกแจงความน่าจะเป็นของความคลาดเคลื่อน (ε) ดังนี้

1. การแจกแจงของ ε_i เป็นแบบปกติ โดยมีค่าเฉลี่ยเป็น 0 คือ $E(\varepsilon_i) = 0$
2. ความแปรปรวนของ ε_i มีค่าคงที่และเท่ากับค่าความแปรปรวนของ Y_i นั่นคือ $V(\varepsilon_i) = V(Y_i) = \sigma^2$
3. ε_i และ ε_j เป็นอิสระแก่กัน นั่นคือ $COV(\varepsilon_i,\varepsilon_j) = 0$ เมื่อ $i \neq j$

ภายใต้สมมติฐานข้างต้นและด้วยวิธีกำลังสองน้อยที่สุด (Least-square Method) ในการประมาณค่าพารามิเตอร์ β_0 และ β_i (i = 1,2,...,k) จะได้ชุดของสมการปกติจำนวน k+1 สมการ คือ

$$
\begin{aligned}
&\sum y_i = b_0 n + b_1 x_{1i} + b_2 \sum x_{2i} + \ldots\ldots\ldots\ldots + b_k \sum x_{ki} \\
&\sum x_{1i} y_i = b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} + \ldots\ldots\ldots\ldots + b_k \sum x_{1i} x_{ki} \\
&\sum x_{2i} y_i = b_0 \sum x_{2i} + b_1 \sum x_{1i} x_{2i} + b_2 \sum x_{2i}^2 + \ldots\ldots\ldots\ldots + b_k \sum x_{2i} \sum x_{ki} \\
&\vdots \\
&\vdots \\
&\sum x_{ki} y_i = b_0 \sum x_{ki} + b_1 \sum x_{1i} x_{ki} + b_2 \sum x_{2i} x_{ki} + \ldots\ldots\ldots + b_k \sum x_{ki}^2
\end{aligned}
$$

โดยที่ b_i (i = 0, 1, 2, ..., k) คือ ค่าประมาณพารามิเตอร์ β_i (i = 0,1,2,......,k) ตามลำดับ ค่าประมาณเหล่านี้ จะคำนวณได้โดยการแก้สมการข้างต้น

สมการที่ได้ $Y = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots + b_kX_k$ ซึ่งเป็นสมการในรูปของคะแนนดิบ แต่ในการวิจัยครั้งนี้จะพิจารณาค่าสัมประสิทธิ์การถดถอยที่อยู่ในรูปคะแนนมาตรฐาน ซึ่งแสดง

ค่าน้ำหนักของแต่ละปัจจัยได้ รูปแบบของสมการคือ

$$Z_Y = B_1 Z_{X1} + B_2 Z_{X2} + \dots + B_k Z_{Xk}$$

โดยที่ Z_Y หมายถึง ตัวแปรตาม Y ในรูปคะแนนมาตรฐาน

Z_{Xi} หมายถึง ตัวแปรอิสระ X_i ในรูปคะแนนมาตรฐาน

B_i หมายถึง ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระในรูปคะแนนมาตรฐาน

ในการคำนวณหาค่า B_j , $j = 1, 2, \dots, k$ จะใช้หลักของเมตริกซ์ซึ่งพิจารณาจากความสัมพันธ์ดังต่อไปนี้ $R_{ij}\underset{\cdot}{B} = \underset{\cdot}{R}_y$ i แทนแถว , j แทนคอลัมภ์

เมื่อ R_{ij} เป็นเมตริกซ์ ของค่าสหสัมพันธ์เชิงเดี่ยวระหว่างตัวแปรอิสระด้วยกัน (Intercorrelation Matrix)

$\underset{\cdot}{B}$ เป็นเมตริกซ์ของค่าสัมประสิทธิ์การถดถอย หรือ Beta weight

$\underset{\cdot}{R}_y$ เป็นเมตริกซ์ของค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว ซึ่งสามารถเขียนได้ดังนี้

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2k} \\ r_{31} & r_{32} & r_{33} & \dots & r_{3k} \\ : & : & : & \dots & : \\ r_{k1} & r_{k2} & r_{k3} & \dots & r_{kk} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ : \\ B_k \end{bmatrix} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ r_{y3} \\ : \\ r_{yk} \end{bmatrix}$$

โดยที่

$$r_{ij} = \frac{N\sum_{k=1}^{N} x_{ki}x_{kj} - \left(\sum_{k=1}^{N} x_{ki}\right)\left(\sum_{k=1}^{N} x_{kj}\right)}{\sqrt{\left\{N\sum_{k=1}^{N} x_{ki}^2 - \left(\sum_{k=1}^{N} x_{ki}^2\right)\right\}\left\{N\sum_{k=1}^{N} x_{kj}^2 - \left(\sum_{k=1}^{N} x_{kj}^2\right)\right\}}}$$

$$r_{yj} = \frac{N\sum_{k=1}^{N} y_k x_{kj} - \left(\sum_{k=1}^{N} y_k\right)\left(\sum_{k=1}^{N} x_{kj}\right)}{\sqrt{\left\{N\sum_{k=1}^{N} y_k^2 - \left(\sum_{k=1}^{N} y_k^2\right)\right\}\left\{N\sum_{k=1}^{N} x_{kj}^2 - \left(\sum_{k=1}^{N} x_{kj}^2\right)\right\}}}$$

เมื่อ N คือ จำนวนข้อมูลทั้งหมด

โดยใช้วิธีการคูณเมตริกซ์สามารถแก้สมการหาค่า $B_1, B_2, B_3, \ldots, B_k$ ได้ทั้งหมด

ในการคัดเลือกสมการถดถอยจะใช้วิธี Stepwise Regression ขั้นแรกจะเป็นการคัดเลือกตัวแปรอิสระเข้าไปในสมการทีละตัว โดยพิจารณาเลือกตัวแปรอิสระที่มีความสัมพันธ์สูงสุดกับตัวแปรตามเข้าไปในสมการ แล้วจึงพิจารณาตัวแปรอิสระที่มีความสัมพันธ์เชิงส่วน (ความสัมพันธ์ที่กำหนดให้ตัวแปรอิสระที่เข้าไปในสมการแล้วคงที่) กับตัวแปรตามมากที่สุดเข้าไปในสมการและพร้อมกันนั้นก็พิจารณาว่าตัวแปรอิสระที่เข้าไปในสมการก่อนหน้านั้นทุกตัวแปรยังควรจะอยู่ในสมการอีกหรือไม่ ถ้าไม่ควรอยู่ก็ตัดออกและดำเนินการคัดเลือกตัวแปรอิสระใหม่ ถ้าควรอยู่ก็ดำเนิน การคัดเลือกตัวแปรอิสระใหม่ เพราะวิธี Stepwise Regression ถือว่าเมื่อตัวแปรอิสระอยู่ในสมการแล้วอาจจะมีผลต่อสมการแบบหนึ่ง และเมื่อเพิ่มตัวแปรใหม่เข้าไป ตัวแปรอิสระนั้นอาจจะมีผลต่อสมการแตกต่างจากเดิมก็ได้ การคัดเลือกตัวแปรอิสระจะดำเนินการไปจนกระทั่งไม่มีตัวแปรอิสระใดเข้าไปหรือถูกตัดออกจากสมการได้อีก วิธีการคัดเลือกตัวแปรอิสระโดยวิธี Stepwise Regression นั้นอาจกำหนดเป็นขั้นตอนได้ดังนี้

(1) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ r ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว

(2) เลือกตัวแปรอิสระที่มีค่า r สูงสุดแล้วจึงกำหนดรูปแบบของสมการดังนี้

$Y = \beta_0 + \beta_p X_p + \varepsilon$ (X_p คือ ตัวแปรอิสระที่ถูกเลือกเข้าไป)

(3) ทดสอบสมมติฐาน $H_0 : \beta_p = 0$ โดยค่าสถิติ F-test

ถ้ายอมรับสมมติฐานให้หยุดและถือว่าไม่มีตัวแปรอิสระใดถูกเลือก ถ้าปฏิเสธสมมติฐานให้ดำเนินการคัดเลือกตัวแปรอิสระต่อไป โดย

(4) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วน (Partial Correlation Coefficient) $r_{y1.23\ldots k}$

($r_{y1.23\ldots k} = \dfrac{r_{y1.23\ldots(k-1)} - \{r_{yk.23\ldots(k-1)}\}\{r_{1k.23\ldots(k-1)}\}}{\sqrt{\{1 - r^2_{yk.23\ldots(k-1)}\}\{1 - r^2_{1k.23\ldots(k-1)}\}}}$) ระหว่างตัวแปรตามกับตัวแปรอิสระที่เหลือแต่ละตัวโดยกำหนดให้ตัวแปรอิสระที่อยู่ในสมการมีค่าคงที่

3.2 หาค่า factor ของปัจจัยต่าง ๆ โดยการเปรียบเทียบค่าสัมประสิทธิ์ที่ได้จากข้อ 3.1

การคำนวณอัตราเบี้ยประกันสุขภาพ

จากที่ได้กล่าวมาแล้วในบทที่ 1 ถึงปัจจัยที่ใช้ในการคำนวณเบี้ยประกันภัยสุขภาพในที่นี้จะพิจารณาเพียง 4 ปัจจัย ได้แก่

1. ค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปี

2. อัตราดอกเบี้ย

3. ค่าใช้จ่าย

4. เงินสำรองเพื่อเหตุฉุกเฉิน

โดยจะรวมปัจจัยที่ 3 และ 4 เข้าด้วยกัน

การกำหนดสมการในการคำนวณเบี้ยประกันภัย

ค่าปัจจุบันของเบี้ยประกันภัยรับ = ค่าปัจจุบันของเงินผลประโยชน์จ่าย

$$P = X.v^{0.5}$$

$$G = P+E.G$$

$$G-E.G = X.v^{0.5}$$

$$G = \frac{X.v^{0.5}}{1-E}$$

$$= \frac{X.(1+i)^{-0.5}}{1-E}$$

สมมติฐาน ให้การเบิกค่ารักษาพยาบาลเกิดขึ้นสม่ำเสมอตลอด ทั้งนี้จึงสมมติให้การเบิกค่ารักษาพยาบาลเกิดขึ้น ณ จุดกึ่งกลางปี

ให้ i เป็นอัตราดอกเบี้ยในการลงทุนต่อปี

$v = \frac{1}{1+i}$ เป็นฟังก์ชั่นที่ใช้ในการหาค่าปัจจุบัน

P เป็นเบี้ยประกันภัยสุทธิ (Net Premium) ต่อปี

G เป็นเบี้ยประกันภัยรวม (Gross Premium) ได้แก่ เบี้ยประกันภัย สุทธิที่รวมค่าใช้จ่าย และเงินสำรองเพื่อเหตุฉุกเฉิน

E เป็นค่าใช้จ่ายและเงินสำรองเพื่อเหตุฉุกเฉิน กำหนดเป็นร้อยละของเบี้ยประกันภัยรวม

X เป็นค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปี

ค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปี สามารถจำแนกได้เป็น 2 ประเภท คือ

1. ค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปีของคนไข้ใน

ให้ A เป็นค่ารักษาพยาบาลที่เกิดขึ้นภายใน 1 ปีเฉลี่ยต่อวันพักรักษาตัว

B เป็นจำนวนวันในการพักรักษาตัวที่เกิดขึ้นภายใน 1 ปีเฉลี่ยต่อการรักษาพยาบาล 1 ครั้ง

C เป็นจำนวนครั้งที่เข้ารับการรักษาพยาบาลที่เกิดขึ้นภายใน 1 ปีเฉลี่ยต่อผู้เอาประกันภัย 1 คน

X เป็นค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปี

$$X = A \times B \times C$$

2. ค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปีของคนไข้นอก

ให้ A เป็นค่ารักษาพยาบาลที่เกิดขึ้นภายใน 1 ปีเฉลี่ยต่อครั้งที่เข้ารับการรักษาพยาบาล

B เป็นจำนวนครั้งที่เข้ารับการรักษาพยาบาลภายใน 1 ปีเฉลี่ยต่อผู้เอาประกันภัย 1 คน

X เป็นค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปี

$$X = A \times B$$

โดยความเป็นจริงแล้วการประกันสุขภาพเป็นการประกันภัยที่ชดเชยค่ารักษาพยาบาลสำหรับคนไข้ใน แต่ในการวิจัยครั้งนี้ข้อมูลที่จัดเก็บ ไม่มีรายละเอียดของจำนวนวันที่เข้ารักษาตัว ดังนั้น ในการคำนวณค่ารักษาพยาบาลเฉลี่ยต่อคนต่อปีจึงคำนวณแบบคนไข้นอก

เนื่องจากค่ารักษาพยาบาลที่จ่ายทดแทนให้แก่ผู้เอาประกันภัยเป็นค่ารักษาพยาบาลที่เกิดขึ้นจริงแต่สูงสุดไม่เกินผลประโยชน์ความคุ้มครองที่กำหนด ดังนั้นในการคำนวณค่าใช้จ่ายในการรักษาพยาบาลจะคำนวณต่อผลประโยชน์ความคุ้มครอง 1 บาท โดยที่

ค่าใช้จ่ายในการรักษาพยาบาลเฉลี่ยต่อคนต่อปีต่อผลประโยชน์ความคุ้มครอง 1 บาท

$$= \frac{\text{ค่าใช้จ่ายในการรักษาพยาบาลทั้งหมด}}{\text{ผลรวมของผลประโยชน์ความคุ้มครองสูงสุดของผู้เอาประกันภัย}}$$