บทที่ 2

ตัวสถิติที่ใช้ในการวิจัย

2.1 การวิเคราะห์ปัจจัย(Factor Analysis)

การวิเคราะห์ปัจจัย (Factor Analysis) ในขั้นตอนการสกัดปัจจัย (Factor Extraction) จะใช้วิธีองค์ประกอบหลัก (Principal Component) ซึ่งเป็นวิธีวิเคราะห์เพื่อวัด"ขนาด"และอิทธิพลของปัจจัยต่างๆซึ่งมีผลต่อ"ขนาด"ของตัวแปรหรือลักษณะที่เราสนใจศึกษา แต่ทั้งนี้ "ขนาด" ของตัวแปรหรือลักษณะที่สนใจศึกษาซึ่งวัดได้เป็นเพียง "ขนาด" เปรียบเทียบเท่านั้นไม่ได้เป็น"ขนาด" ที่แท้จริง ซึ่งอาศัยหลักความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูล มีวิธีการดังต่อไปนี้

1. กำหนดปัจจัยที่คาดว่าจะมีผลหรืออิทธิพลต่อตัวแปรหรือลักษณะที่เราสนใจศึกษา (x_i)
2. ตรวจสอบปัจจัยแต่ละตัวเกี่ยวกับลักษณะต่าง ๆ
   - 2.1 เป็นตัวแปรเชิงปริมาณชนิดต่อเนื่องหรือไม่
   - 2.2 หน่วยของตัวแปรเป็นหน่วยเดียวกันหรือไม่
   - 2.3 ค่าของตัวแปรที่นำมาใช้เปรียบเทียบกันได้หรือไม่
3. หา Dispersion Matrix ของปัจจัยข้างต้น (S)
4. หา Characteristic root ของ S คือ $\lambda_1, \lambda_2, \ldots, \lambda_p$
5. นำ Characteristic root ของ S ที่มีค่าสูงสุด (λ_1) มาหา Characteristic vector เรียกว่า $\underline{a_1}$
6. Normalized Characteristic vector $\underline{a_1}$ เรียกว่า $\underline{a_1^*}$ จะได้ Component ที่ 1 อยู่ในรูป $Y_1 = \underline{a_1^*}\underline{X} = \sum_{j=1}^{p} a_{ij}X_j$

ซึ่งใช้สำหรับวัด "ขนาด" หรือ " SIZE" ของตัวแปรหรือลักษณะที่สนใจ

7. หา Component อื่น ๆ ที่เหลือจาก Characteristic root อื่น ๆ ที่มีค่าสูงรองลงมาจาก λ_1 ตามลำดับ คือ $\lambda_2, \lambda_3, \ldots, \lambda_p$ โดยวิธีเดียวกัน เพื่อนำมาใช้ในการทดสอบสมมติฐานเกี่ยวกับปัจจัยต่าง ๆ (Orthogonal Comparison)

ในกรณีที่นำตัวแปรมาวิเคราะห์ p ตัว และใช้ตัวอย่างขนาด n

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots\cdots & X_{1p} \\ X_{21} & X_{22} & \cdots\cdots & X_{2p} \\ \vdots & & & \\ \vdots & & & \\ X_{n1} & X_{n2} & \cdots\cdots & X_{np} \end{bmatrix}_{n\times p}$$

ให้ Y เป็นตัวแปรตัวใหม่ที่จะใช้แทนปัจจัยหรือตัวแปรที่นำมาวิเคราะห์ทั้ง p ตัว $(X_1, X_2, \ldots\ldots, X_p)$

$$Y = \underset{1\times p}{\underline{a}'}\ \underset{1\times p}{\underline{X}} \qquad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

เพื่อที่จะสามารถหาค่า $\underline{a}$ ที่ประกอบด้วยค่า $a_1, a_2, \ldots, a_p$ ได้ จึงต้องมีสมการแสดงความสัมพันธ์ระหว่าง Y กับ X ทั้งหมด p สมการ

ให้ $\underline{Y_i} = \underline{a_i'}\,\underline{X}$ เมื่อ $\underline{a_i} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}_{p\times 1}$

การหาค่า vector $\underline{a_i}$

ให้ S แทน Dispersion matrix ของ $\underline{X}$

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots\cdots & S_{1p} \\ S_{21} & S_{22} & \cdots\cdots & S_{2p} \\ \vdots & & & \\ \vdots & & & \\ S_{p1} & S_{p2} & \cdots\cdots & S_{pp} \end{bmatrix}_{p\times p}$$

Characteristic root ของ S คือ $\lambda_1 > \lambda_2 > \lambda_3 > \ldots\ldots > \lambda_p$ หาจาก $|s - \lambda I| = 0$

Characteristic vector $\underline{a_1}$ หาจาก $(s - \lambda_i I)\underline{a_i} = \underline{0}$

Normalized Characteristic vector $\underline{a_1}$ หาจาก

$$\begin{bmatrix} a_{i1} / \sqrt{a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2} \\ a_{i2} / \sqrt{a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2} \\ \vdots \\ \vdots \\ a_{ip} / \sqrt{a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2} \end{bmatrix}$$

2.2 วิธี Cumulative $\sqrt{f(y)}$

วิธี Cumulative $\sqrt{f(y)}$ เป็นวิธีที่ใช้หาขอบเขตของกลุ่ม(Construction of Strata) เพื่อแบ่งหน่วยทดลองที่เราสนใจออกเป็นกลุ่มย่อย โดยพิจารณาจากการสะสมของ $\sqrt{f(y)}$ เมื่อ

- f(y) เป็นจำนวนผู้เอาประกันภัยในแต่ละกลุ่มของปัจจัยที่มีอิทธิพลต่อการคิดอัตราเบี้ยประกันภัย

- y เป็นค่าน้ำหนักที่ได้จากการแทนตัวแปรในแบบจำลอง(Principal Component) ซึ่งแสดงค่าปัจจัยที่มีอิทธิพลต่อการคิดอัตราเบี้ยประกันภัยของการประกันภัยรถยนต์

ขั้นตอนการจำแนกกลุ่ม

1. เรียงลำดับคะแนนความสำคัญของปัจจัยที่มีอิทธิพลต่อการคิดอัตราเบี้ยประกันภัยแล้วแบ่งออกเป็นช่วงย่อยๆ โดยให้แต่ละช่วงมีขนาดพอสมควร

2. หาความถี่f(y) ของผู้เอาประกันภัยในแต่ละช่วงย่อยของปัจจัยที่มีอิทธิพลต่อการคิดอัตราเบี้ยประกันภัย

3. หาค่าสะสมของความถี่ของผู้เอาประกันภัย {Cumulative$\sqrt{f(y)}$}

4. แบ่งยอดรวมของค่าสะสมของความถี่ของผู้เอาประกันภัย {Cumulative$\sqrt{f(y)}$}โดยการหารด้วยจำนวนกลุ่มที่ต้องการ คือ 5 กลุ่ม

ผลหารที่ได้เป็นตัวแปรค่าสะสม $\sqrt{f(y)}$ ซึ่งแสดงถึงขอบเขตกลุ่มผู้เอาประกันภัยในแต่ละน้ำหนักของปัจจัยที่มีอิทธิพลต่อการคิดอัตราเบี้ยประกันภัย