บทที่ 2

ทฤษฎี Regression ประยุกต์กับการสร้างดัชนีราคา

ดังที่กล่าวมาแล้วว่า การสร้างดัชนีราคานั้นมีความคลาดเคลื่อนอันเนื่องมาจากการเปลี่ยนแปลงทางด้านคุณภาพของสินค้า ทั้งนี้เพราะสินค้ามีการเปลี่ยนแปลงลักษณะจำเพาะอยู่เสมอ ๆ ในการแก้ปัญหานี้กับการสร้างดัชนีราคา เช่น ดัชนีราคาผู้บริโภค ดัชนีราคาขายส่งและดัชนีราคาขายปลีก ฯลฯ การทดลองศึกษาหาความสัมพันธ์เชิงฟังก์ชันระหว่างราคาสินค้ากับลักษณะจำเพาะของสินค้า แล้วนำทฤษฎี Regression มาประยุกต์กับการสร้างดัชนีราคา โดยทำการวิเคราะห์ว่าราคาสินค้าจะมีความสัมพันธ์กับลักษณะจำเพาะลักษณะใดบ้าง ผลที่ได้นี้อาจนำไปประมาณราคาเฉลี่ยของสินค้าเพื่อสร้างดัชนีราคาตามที่ต้องการได้ อย่างไรก็ดีจะใช้ทฤษฎี Regression มาประยุกต์กับการสร้างดัชนีราคานั้นได้ก็ต่อเมื่อมีข้อมูลที่เกี่ยวกับราคาสินค้าและลักษณะจำเพาะของสินค้านั้น ๆ แล้ว วิธีที่จะได้ข้อมูลดังกล่าวก็โดยการสืบราคาสินค้าที่ตรงตามลักษณะจำเพาะของสินค้านั้น เพื่อให้การหาความสัมพันธ์เชิงฟังก์ชันระหว่างราคาสินค้ากับลักษณะจำเพาะของสินค้าเป็นไปโดยสะดวก ในทางปฏิบัติจะต้องกำหนดลักษณะจำเพาะของสินค้าขึ้นเพื่อใช้ได้กับในทางทฤษฎี Regression เพราะลักษณะจำเพาะที่เลือกขึ้นมาใช้สร้างความสัมพันธ์จะต้องมีความหมายในการเปลี่ยนแปลงทางคุณภาพของสินค้า ดังนั้นก่อนกำหนดลักษณะจำเพาะของสินค้าจำเป็นจะต้องทราบถึงคุณภาพของสินค้า หลักเกณฑ์ในการพิจารณาแยกคุณภาพของสินค้า การเลือกตัวแปรเพื่อศึกษาคุณภาพของสินค้า และการวัดคุณภาพของสินค้า ตามความหมายของการใช้ทฤษฎี Regression ที่ใช้เสียก่อน

<u>คุณภาพของสินค้าในความหมายของการใช้ทฤษฎี Regression</u>

นักเศรษฐศาสตร์อธิบายคำว่า Quality หรือ "คุณภาพ" ไว้ว่า คือ Property หรือ attribute หรือ Characteristic แต่คุณภาพของสินค้าในความหมายของการใช้ทฤษฎี Regression จะหมายถึงสินค้าซึ่งมี Property หรือ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

attribute หรือ Characteristic ที่ต่างกันเท่านั้น เช่น ไม้สัก กับไม้แดง มี attribute ต่างกัน ตู้เย็นขนาด 10 ลูกบาศก์ฟุตกับตู้เย็นขนาด 12 ลูกบาศก์ฟุต มี Property ต่างกัน ผงซักฟอกกับสบู่มี Characteristic ต่างกัน

โดยทั่วไปสินค้าบางชนิดอาจเห็นความแตกต่างในทางคุณภาพที่เด่นชัด คือ สามารถแยก ranking of products ในรูปของ Grade, desireability หรือ Usefullness และ degree of excellence ได้ เช่น ข้าวสาร แบ่งออกเป็น ข้าวสาร 100% ข้าวสาร 5% ข้าวสาร 10% และข้าวสาร 15% แต่สินค้าบางชนิดอาจแสดงให้เห็นความแตกต่างทางคุณภาพโดยทาง listing set of attribute หรือลักษณะซึ่งแยกโดยการผลิต เช่นการผลิตหลอดไฟฟ้าชนิดต่าง ๆ อาจเห็นความแตกต่างในขนาด รูปร่าง สี หรือลักษณะที่ปรากฏทางรูปร่างอื่น ๆ รวมทั้งเทคนิคที่ทำด้วย นอกจากนี้สินค้าบางชนิดอาจเห็นความแตกต่างที่เกี่ยวกับจำนวนหรือปริมาณของลักษณะที่ต่างกัน เช่น กระดาษ 80 แกรม กับ กระดาษ 70 แกรม เป็นต้น

<u>หลักเกณฑ์ในการพิจารณาแยกคุณภาพของสินค้า</u>

โดยทั่วไปสินค้าซึ่งมีคุณภาพตามลักษณะดังกล่าวข้างต้นจะสามารถแยกออกเป็นส่วน ๆ ได้โดยที่แต่ละส่วนจะเป็นอิสสระซึ่งกันและกัน โดยเรียกรวม ๆ ว่ามี Characteristics ที่ต่างกัน บางครั้งมีผู้เข้าใจว่า การแยกคุณภาพของสินค้าอาจแยกได้ใน 2 กรณี คือในด้านการผลิตหมายถึงต้นทุนการผลิต (Production Cost) และทางด้านการประเมินของผู้ใช้สินค้าบางชนิด เช่น สินค้าที่มีสภาพไม่คงทน (Non-durable goods)มักจะต้องพิจารณาจากการประเมินของผู้ใช้ซึ่งขึ้นอยู่กับความพอใจของผู้บริโภค (Consumers Preference)คือศึกษาในทาง Utility Function แต่สินค้าบางชนิดที่เป็นสินค้าคงทน (durable goods) มักจะพิจารณากันในด้านต้นทุนการผลิต(Production Cost)มากกว่า

ในการวัดคุณภาพของสินค้าตามวิธีของทฤษฎี Regression จะต้องพิจารณาแยกสินค้าออกเป็นส่วน ๆ ที่เห็นว่าเกี่ยวกับคุณภาพหรือปริมาณของสินค้านั้น หรืออาจจะเรียกว่าเป็น

ดังกล่าว บางกรณีเกิดปัญหาที่เรียกว่า Multi-collinearity และauto-correlation ซึ่งจะได้กล่าวถึงในภายหลัง

การวัดคุณภาพตามวิธีของการใช้ทฤษฎี Regression

การวัดคุณภาพตามวิธีของการใช้ทฤษฎี Regression ควรจะต้องกำหนดเงื่อนไขไว้ว่า "ในเวลาหนึ่งเวลาใดสินค้าชนิดเดียวกันที่มีคุณภาพต่ำกว่าจะไม่ถูกซื้อในราคาเดียวกันหรือสูงกว่าสินค้าที่มีคุณภาพสูงกว่า" เช่น ในเดือนมกราคม ข้าวสาร 100% ราคาถังละ 55 บาท ข้าวสาร 15% ราคาถังละ 60 บาท ใช้วัดคุณภาพไม่ได้ เนื่องจากข้าวสาร 15% เป็นสินค้าที่มีคุณภาพต่ำกว่าข้าวสาร 100% และราคาไม่ควรจะมากกว่าหรือเท่ากับ 55 บาท และ "สินค้าที่เกือบจะไม่มีคุณภาพ หรือคุณภาพต่ำ" เช่น วิทยุซึ่งผลิตและเก็บไว้นานจนเสื่อมคุณภาพหรือในการผลิตคราวเดียวกัน มีอุปกรณ์ภายในเครื่องชำรุดเสียหาย ก็จะไม่นำมาพิจารณาในการหาความสัมพันธ์เชิงฟังก์ชั่นระหว่างราคาสินค้ากับลักษณะจำเพาะที่เกี่ยวกับคุณภาพ เพราะสินค้าดังกล่าวย่อมไม่เกิดประโยชน์ในการที่จะนำมาใช้วัดคุณภาพตามวิธีของการใช้ทฤษฎี Regression เพราะราคาสินค้าเหล่านั้นไม่ขึ้นอยู่กับลักษณะจำเพาะของสินค้าที่เลือกขึ้นกล่าวคือ เป็นราคาสินค้าที่ไม่มีคุณภาพ

ความสัมพันธ์เชิงฟังก์ชั่นระหว่างราคาและลักษณะจำเพาะของสินค้า

เมื่อกำหนดลักษณะจำเพาะของสินค้าตรงตามคุณภาพของสินค้า หลักเกณฑ์ในการพิจารณาแยกคุณภาพของสินค้า การเลือกตัวแปรเพื่อศึกษาคุณภาพของสินค้า และการวัดคุณภาพของสินค้าตามความหมายของการใช้ทฤษฎี Regression ได้แล้ว ก็จะสามารถหาข้อมูลของราคาสินค้าและลักษณะจำเพาะของสินค้านั้น ๆ มาศึกษาหาความสัมพันธ์เชิงฟังก์ชั่นได้

ในการหาความสัมพันธ์เชิงฟังก์ชั่นระหว่างราคาสินค้ากับลักษณะจำเพาะของสินค้านี้จะพบว่า สามารถสร้างความสัมพันธ์เชิงฟังก์ชั่นในรูปสมการต่าง ๆ ได้ตามความเหมาะสมของสินค้านั้น ๆ เช่น

1. $P = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_k X_k + e$

2. $P = \beta_o + \beta_1 X + \beta_2 X^2 + e$

3. $P = X^{\beta} + e$

4. $P = \beta_o + \beta_1 X^2 + e$

5. $P = \beta X + e$

ฯลฯ

โดยทั่ว ๆ ไปความสัมพันธ์ระหว่างราคาสินค้าและลักษณะจำเพาะต่าง ๆ ของสินค้ามักอยู่ในรูป polynomial ดีกรีต่าง ๆ กล่าวคือ

$$P_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots\cdots + \beta_k X_{ik} + e_i$$

เมื่อ P_i เป็นตัวแปรตาม (dependent Variable) ซึ่งแทนราคาสินค้าที่เก็บรวบรวมได้รายการที่ i และ $X_{i1}, X_{i2}, \cdots\cdots, X_{ik}$ เป็นตัวแปรอิสสระ (Independent Variable) ที่เลือกขึ้นแทนลักษณะจำเพาะ (specification) k ลักษณะที่เป็นตัวกำหนดคุณภาพของสินค้ารายการที่ i คิดเป็นราคาต่อหน่วย

β_o เป็น P-intercept หรือราคาของสินค้าที่เกือบจะไม่มีคุณภาพ

β_j เป็น implicit price หรือราคาที่เปลี่ยนไปเมื่อลักษณะจำเพาะ X_{ij} เปลี่ยนไป 1 หน่วย และลักษณะจำเพาะอื่น ๆ ที่เหลือ $X_{i1}, X_{i2}, \cdots\cdots, X_{i(j-1)}, X_{i(j+1)}, \cdots\cdots\cdots, X_{ik}$ คงที่

e_i เป็นความคลาดเคลื่อนของราคาสินค้ารายการที่ i และมีคุณสมบัติดังนี้

1. $E(e_i) = 0$

2. $E(e_i e_j) = 0$

3. $E(e_i^2) = \sigma^2$

4. $E(e_i X_i) = 0$

<u>การประมาณ implicit price (β_j)</u>

เนื่องจาก $P_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + e_i$

ดังนั้น $e_i = P_i - \beta_o - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik}$

ถ้า $e_i^2 = \left[P_i - \beta_o - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik}\right]^2$

ถ้าใช้ตัวอย่างขนาด n หรือทำการสืบราคาสินค้าชนิดนั้น n ร้าน โดยที่ $n \geq k$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (P_i - \beta_o - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik})^2$$

ในการประมาณราคาสินค้าให้มีความคลาดเคลื่อนจากราคาจริงน้อยที่สุดนี้อาจใช้วิธีกำลังสองน้อยที่สุด (Method of Least Square Estimation) หรือวิธีความน่าจะเป็นมากที่สุด (Method of Maximum Likelihood Estimation) จะได้ $\hat{\beta}_o$ เป็นค่าประมาณของ P-intercept และ $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ เป็นค่าประมาณของ implicit price คือ

$$\hat{\beta}_o = \bar{P} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \cdots - \beta_k \bar{X}_k$$

$$\hat{\beta}_p = \sum_{i=1}^{n} X_{ip}P_i - \beta_o \sum_{i=1}^{n} X_{ip} - \beta_1 \sum_{i=1}^{n} X_{i1}X_{ip} - \beta_2 \sum_{i=1}^{n} X_{i2}X_{ip}$$

$$- \cdots\cdots - \beta_{p-1} \sum_{i=1}^{n} X_{ip-1}X_{ip} - \cdots - \beta_k \sum_{i=1}^{n} X_{ik}X_{ip} \Big/ \sum_{i=1}^{n} X_{ip}^2$$

เมื่อ $p = 1, 2, 3, \ldots\ldots k$

ในกรณีที่ลักษณะจำเพาะของสินค้าชนิดเดียวกันมีเป็นจำนวนมาก การแก้สมการปกติ (Normal Equation) หรือหาค่า $\hat{\beta}_o$, $\hat{\beta}_j$ โดยตรงจะทำได้ยากหรือเสียเวลามาก ดังนั้นอาจจะใช้ทฤษฎีของ Matrix ช่วยในการประมาณ implicit price ได้ดังนี้

สมการแสดงความสัมพันธ์ระหว่างราคาสินค้าและลักษณะจำเพาะต่าง ๆ ของสินค้าคือ

$$\underset{\sim}{P} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{e}$$

เมื่อ $\underset{\sim}{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix}$, $\underset{\sim}{X} = \begin{bmatrix} X_{11}X_{12}X_{13} \cdots\cdots X_{1k} \\ X_{21}X_{22}X_{23} \cdots\cdots X_{2k} \\ \vdots\ \vdots\ \vdots \quad \cdots\cdots \quad \vdots \\ X_{n1}X_{n2}X_{n3} \quad\quad X_{nk} \end{bmatrix}$

$\underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, $\underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ค่าประมาณของ implicit price, $\hat{\beta}_j$ เขียนได้อยู่ในรูป

$$\underline{\hat{\beta}} = \left[\underline{X}'\underline{X}\right]^{-1} \underline{X}'\underline{P}$$

เมื่อ $\underline{X}'\underline{X} =$

$$\begin{bmatrix}
\sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i3} & \cdots\cdots & \sum_{i=1}^{n} X_{i1}X_{ik} \\
\sum_{i=1}^{n} X_{i2}X_{i1} & \sum_{i=1}^{n} X_{i2}^2 & \sum_{i=1}^{n} X_{i2}X_{i3} & \cdots\cdots & \sum_{i=1}^{n} X_{i2}X_{ik} \\
\sum_{i=1}^{n} X_{i3}X_{i1} & \sum_{i=1}^{n} X_{i3}X_{i2} & \sum_{i=1}^{n} X_{i3}^2 & \cdots\cdots & \sum_{i=1}^{n} X_{i3}X_{ik} \\
\vdots & \vdots & \vdots & & \vdots \\
\sum_{i=1}^{n} X_{ik}X_{i1} & \sum_{i=1}^{n} X_{ik}X_{i2} & \sum_{i=1}^{n} X_{ik}X_{i3} & \cdots\cdots & \sum_{i=1}^{n} X_{ik}^2
\end{bmatrix}$$

$\underline{X}$ ในที่นี้คือ Matrix of Specification

$\underline{X}'\underline{X}$ คือ Coefficient Matrix of Specification

$\left[\underline{X}'\underline{X}\right]^{-1}$ คือ Variance Matrix of Specification

เมื่อประมาณค่าของ implicit price $\hat{\beta}_j$ ได้แล้ว จะทราบว่าความสัมพันธ์ระหว่างราคาสินค้ากับลักษณะจำเพาะของสินค้าเป็นไปในลักษณะใด กล่าวคือ ราคาสินค้าจะเพิ่มขึ้นหรือลดลงขึ้นอยู่กับค่าของ implicit price $\hat{\beta}_j$ นี้ อย่างไรก็ดีค่าของ $\hat{\beta}_j$ แต่ละตัวซึ่งประมาณโดยใช้วิธีดังกล่าวจะมีความสำคัญต่อลักษณะจำเพาะ X_{ij} ที่เลือกขึ้นและมีความเชื่อถือได้มากน้อยเพียงไรนั้น จะต้องนำทฤษฎีวิเคราะห์ความถดถอยและสหสัมพันธ์มาใช้ในการทดสอบ หรือกล่าวอีกอย่างหนึ่งว่า ในการทดสอบโดยใช้ทฤษฎีวิเคราะห์ความถดถอยและสหสัมพันธ์เพื่อทดสอบดูว่ามีลักษณะจำเพาะใดบ้างที่มีอิทธิพลต่อราคาสินค้า อันเนื่องมาจากการเปลี่ยนแปลงในทางคุณภาพและลักษณะจำเพาะใดบ้างที่ไม่มีอิทธิพลหรือมีอิทธิพลน้อยต่อราคาสินค้าอันเนื่องมาจากการเปลี่ยนแปลงในทางคุณภาพก็จะไม่นำมาพิจารณา

ในการทดสอบโดยใช้ทฤษฎีวิเคราะห์ความถดถอยและสหสัมพันธ์นี้มักทำการทดสอบเป็นขั้น ๆ ไปดังนี้

ขั้นที่ 1. การทดสอบโดยใช้ F - test

ขั้นที่ 2. การทดสอบโดยใช้ t - test

ขั้นที่ 3. การทดสอบ Multicollinearity โดยใช้ Partial และ Multiple Correlation

ขั้นที่ 4. การทดสอบ Auto-correlation โดยใช้ Durbin-Watson test

1. <u>การทดสอบโดยใช้ F - test</u>

การทดสอบโดยใช้ F - test คือการทดสอบเพื่อดูว่าลักษณะจำเพาะที่นำมาใช้ในการกำหนดนั้นโดยทั่ว ๆ ไปมีความสัมพันธ์ต่อราคาสินค้าอย่างมีนัยสำคัญหรือไม่ วิธีนี้จะต้องใช้ทฤษฎี Multiple Linear Regression มาประยุกต์ กล่าวคือ

ความคลาดเคลื่อนของ P มีการกระจายแบบปกติ ซึ่งมีค่าเฉลี่ยเลขคณิต =0, ความแปรปรวน = σ^2

$e \sim N\ (0, \sigma^2)$ และ Unbiased Estimate ของ σ^2 คือ $S^2 = \sum_{i=1}^{n} (P_i - \hat{P}_i)^2 / n - k$

เมื่อ n เป็นจำนวนราคาหรือจำนวนร้านค้าตัวอย่างที่เก็บรวบรวมข้อมูลมาได้

k เป็นจำนวนลักษณะจำเพาะที่นำมาใช้ในการคำนวณ

ถ้าให้ $e_i = (P_i - \hat{P}_i)$ เป็นความคลาดเคลื่อนระหว่างราคาที่เก็บรวบรวมมาได้กับราคาที่ประมาณโดยใช้ลักษณะจำเพาะ k ตัว

และ $\hat{p}_i = \hat{P}_i - \bar{P}$ คือความคลาดเคลื่อนของราคาที่ประมาณได้จากค่าเฉลี่ยเลขคณิตของราคา

$p_i = P_i - \bar{P}$ คือความคลาดเคลื่อนของราคาที่เก็บรวบรวมได้จากค่าเฉลี่ยเลขคณิตของราคา

ดังนั้น $p_i = (\hat{P}_i - \bar{P}) + (\bar{P} - \hat{P}_i)$

และ $\sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} \hat{p}_i^2 + \sum_{i=1}^{n} e_i^2$

จะเห็นได้ว่าผลรวมกำลังสองของความคลาดเคลื่อนระหว่างราคาที่เก็บรวบรวมได้จากค่าเฉลี่ยเลขคณิตของราคา ประกอบด้วย

1. $\sum_{i=1}^{n} \hat{p}_i^2$ คือ ผลรวมกำลังสองของความคลาดเคลื่อนของราคาที่ประมาณได้จากค่าเฉลี่ยเลขคณิตของราคา

$$\sum_{i=1}^{n} \hat{p}_i^2 = \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}p_i + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}p_i + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}p_i$$

2. $\sum_{i=1}^{n} e_i^2$ คือ ผลรวมกำลังสองของความคลาดเคลื่อนระหว่างราคาที่รวบรวมได้กับราคาที่ประมาณ

โดยการทดสอบ $H_o: \beta_j = 0 \; ; \; j = 1, 2, 3, \ldots, k$

$$F = \frac{\sum_{i=1}^{n} \hat{p}_i^2}{k} \Big/ \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$$

เมื่อเปรียบเทียบค่า F ที่คำนวณได้กับ $F_{k,n-k-1}$ จากตาราง F-distribution โดยกำหนดนัยสำคัญทางสถิติให้เท่ากับ α แล้วจะสรุปได้ว่าโดยทั่วไปลักษณะจำเพาะ $X_1, X_2, \ldots, X_k$ มีความสัมพันธ์กับราคาสินค้า P อย่างมีนัยสำคัญหรือไม่

2. <u>การทดสอบโดยใช้ t - test</u>

การทดสอบโดยใช้ t-test เพื่อต้องการดูว่ามีลักษณะจำเพาะใดบ้างที่มีความสัมพันธ์กับราคาสินค้าอย่างมีนัยสำคัญ หลังจากที่ทราบจากการทดสอบโดยใช้ F-test แล้วว่าโดยทั่ว ๆ ไป ลักษณะจำเพาะ**ที่นำมาใช้สร้าง**ความสัมพันธ์มีความสัมพันธ์กับราคาสินค้า

โดยทดสอบ $H_o : \beta_j = 0 \; ; \; j = 1, 2, \ldots, k$

ใช้ Student't test

$$t = \frac{\hat{\beta}_j - \beta_j}{S_{\hat{\beta}_j}} = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}}$$

เมื่อ $S_{\hat{\beta}_j} = \sqrt{c_{ii} s_P^2}$

$$s_P^2 = \sum_{i=1}^{n} e_i^2 / n-k-1$$

c_{ii} เป็นค่าของ diagonal ของ Variance Matrix of Specification

เมื่อเปรียบเทียบค่า t ที่คำนวณได้กับ t (n-k-1) จากตาราง t-distribution โดยกำหนดนัยสำคัญทางสถิติให้เท่ากับ α แล้ว จะสรุปได้ว่าลักษณะจำเพาะใดบ้างที่มีความสัมพันธ์กับราคาสินค้า P อย่างมีนัยสำคัญหรือไม่

3. <u>การทดสอบ Multicollinearity โดยใช้ Partial และ Multiple Correlation</u>

การทดสอบโดยใช้ Partial และ Multiple Correlation คือการทดสอบเพื่อหาความสัมพันธ์ระหว่างราคา P กับลักษณะจำเพาะ X_j เมื่อ $j = 1,2,\ldots,k$ หรือดูความสัมพันธ์ระหว่างลักษณะจำเพาะ X_j กับ X_1 เมื่อ $j \neq 1$ ว่ามีความสัมพันธ์กันหรือไม่ โดยคำนวณจากสูตร Partial Correlation Coefficient

$$r_{X_j X_1} = \frac{\text{cov}\,(X_j, X_1)}{\sqrt{V(X_j)\, V(X_1)}}$$

เปรียบเทียบกับ Multiple Correlation Coefficient $R = r_{P\hat{P}}$ คือการทดสอบเพื่อหาความสัมพันธ์ระหว่างราคาที่รวบรวมได้กับราคาซึ่งประมาณโดยใช้ลักษณะจำเพาะทั้ง k ลักษณะว่ามีความสัมพันธ์กันหรือไม่ โดยคำนวณจากสูตร

$$R = r_{P\hat{P}} = \frac{\text{cov}\,(P,\hat{P})}{\sqrt{V(P)\, V(\hat{P})}}, \quad P = \hat{P} + e, \quad 0 \leq R \leq 1$$

โดยที่ค่าของ $r_{P\hat{P}} > r_{X_j X_1}$ เสมอ

ในการทดสอบโดยวิธีนี้จะพบว่ามีลักษณะจำเพาะบางลักษณะซึ่งมีความสัมพันธ์กันในระดับสูงมาก (highly intercorrelated) เช่น $r_{X_j X_1}$ มีค่า .9

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หมายความว่า x_j กับ x_1 มีความสัมพันธ์กันในระดับสูงมาก x_j' อาจเขียนเป็นผลรวมของ x_j กับ x_1 ได้ นั่นคือ ลักษณะจำเพาะ(x_j, x_1) ที่เลือกมานั้นมีความสัมพันธ์ซึ่งกันและกัน (คือเป็น linear dependent) ซึ่งในทางทฤษฎีเรียกว่าเกิดกรณี Multicollinearity

ผลของการเกิด Multicollinearity นี้จะทำให้ความแปรปรวนของราคาสินค้าที่ประมาณได้มีค่ามาก ทั้งนี้เนื่องมาจากการเปลี่ยนแปลงของลักษณะจำเพาะหนึ่งเป็นผลเนื่องมาจากลักษณะจำเพาะอื่นที่มีความสัมพันธ์กัน เช่น ความสัมพันธ์ระหว่างราคากระเป๋าถือกับราคาหนังสัตว์ ถ้าราคาหนังสัตว์ต่อหน่วยเปลี่ยนแปลงไป จะทำให้ราคากระเป๋าถือต่อหน่วยเปลี่ยนแปลงไปด้วย แต่การที่ราคาต่อหน่วยเปลี่ยนแปลงไปนี้จะต้องพิจารณาความแปรปรวนของราคาทั้งสองด้วย กล่าวคือถ้าจะประมาณราคาสินค้าหมวดผลิตภัณฑ์หนังสัตว์ โดยหาความสัมพันธ์ระหว่างราคาผลิตภัณฑ์หนังสัตว์กับราคากระเป๋าถือ และราคาหนังสัตว์ ถ้าการคำนวณพบว่าความแปรปรวนของราคาผลิตภัณฑ์หนังสัตว์มีค่ามาก ค่าของ R^2 จะลดน้อยลง แต่ถ้านำราคาหนังสัตว์ออกจากสมการความสัมพันธ์แล้ว ความแปรปรวนของราคาผลิตภัณฑ์หนังสัตว์จะไม่เปลี่ยนจากเดิมมาก ค่าของ R^2 จะไม่เปลี่ยนไปจากเดิมมาก ดังนั้นถ้าหาความสัมพันธ์ระหว่างราคากระเป๋าถือและราคาหนังสัตว์ และตัดราคาหนังสัตว์ออกจากสมการ จะทำให้ความแปรปรวนของราคาผลิตภัณฑ์หนังสัตว์ลดน้อยลง และค่าของ R^2 จะสูงขึ้น โดยทั่วไปในการแก้ปัญหานี้คือหาความสัมพันธ์ระหว่างลักษณะจำเพาะนั้นในรูป Linear regression ก่อน แล้วตัดลักษณะจำเพาะลักษณะหนึ่งทิ้งไปหรืออาจจะใช้ Factor Analysis เข้ามาช่วยก็ได้

Factor analysis

การใช้ Factor analysis จะเป็นวิธีหนึ่งที่ช่วยแก้ปัญหากรณีMulticollinearity ได้ ในกรณีที่สินค้าบางชนิดมีลักษณะจำเพาะบางลักษณะซึ่งมีความสัมพันธ์ซึ่งกันและกันในทางคุณภาพจนไม่สามารถจะเลือกลักษณะจำเพาะที่เป็นอิสสระซึ่งกันและกันได้โดยง่าย การใช้Factor analysis จะช่วยลดจำนวนของลักษณะจำเพาะบางลักษณะซึ่งมีความสัมพันธ์กับลักษณะจำเพาะอื่น ๆ ในทางคุณภาพได้ โดยที่จะทำให้การสร้างความสัมพันธ์ระหว่างราคาและลักษณะจำเพาะต่าง ๆ ทำได้ง่ายและมีความเชื่อถือได้มากขึ้น

ในการวิเคราะห์ทาง factor นี้ ให้ P เป็นราคาสินค้าซึ่งเป็น function ของ linear combination ของลักษณะจำเพาะ $F_1, F_2, \ldots, F_m$ ซึ่งมีความสัมพันธ์กันหรือเขียนได้เป็น

$P = f(F_1, F_2, \ldots\ldots, F_m)$ เมื่อ F_i; $i = 1, 2, \ldots\ldots, m$ เป็น linear combination ของลักษณะจำเพาะซึ่งมีความสัมพันธ์ซึ่งกันและกัน k ตัว ดังนี้

$$F_1 = a_{11}x_1 + a_{12}x_2 + \cdots\cdots + a_{1k}x_k \cdots\cdots (1)$$

$$F_2 = a_{21}x_1 + a_{22}x_2 + \cdots\cdots + a_{2k}x_k \cdots\cdots (2)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$F_m = a_{m1}x_1 + a_{m2}x_2 + \cdots\cdots + a_{mk}x_k \cdots\cdots (m)$$

เมื่อ $k > m$ แล้วประมาณค่าของ a_{ij} โดยหา latent root เพื่อรวมลักษณะจำเพาะเหล่านี้เข้าด้วยกัน แล้วใช้ค่าของ $F_1, F_2, \ldots\ldots, F_m$ ในการหาความสัมพันธ์เชิงฟังก์ชันระหว่างราคากับลักษณะจำเพาะของสินค้าในรูปสมการ

$$P_i' = \beta_o' + \beta_1' F_1 + \beta_2' F_2 + \cdots\cdots + \beta_m' F_m + e_i$$

สมการนี้จะเป็นสมการซึ่งได้ลดลักษณะจำเพาะ (k ตัว) ที่มีความสัมพันธ์ซึ่งกันและกันออกไปแล้ว จึงได้ลักษณะจำเพาะที่เหลือ (m ตัว) ซึ่งเป็นอิสสระซึ่งกันและกัน (สำหรับรายละเอียดเกี่ยวกับการใช้ Factor Analysis นี้ ผู้เขียนจะไม่ขอนำมากล่าวในที่นี้ ถ้าผู้ใดสนใจเกี่ยวกับรายละเอียดในเรื่องนี้จะศึกษาได้จากในหนังสือ Modern Factor Analysis ของ Hyman, H.m, University of Chicago Press, Revised Edition.)

4. <u>การทดสอบ Auto Correlation โดยใช้ Durbin-Watson test</u>

เมื่อทำการทดสอบความสัมพันธ์ระหว่างราคากับลักษณะจำเพาะของสินค้าทั้ง 3 วิธีแล้ว จะต้องทดสอบดูว่าความคลาดเคลื่อนระหว่างราคาที่เก็บรวบรวมได้กับราคาที่ประมาณโดยลักษณะจำเพาะนั้น จะเป็นอิสสระซึ่งกันและกันหรือไม่ ในบางกรณีความคลาดเคลื่อนแต่ละตัวอาจไม่เป็นอิสสระซึ่งกันและกัน ซึ่งในทางทฤษฎีเรียกว่าเกิด auto-correlation ใน disturbance term (e)

นั่นคือ $E(e_i e_j) \neq 0$

ในทางปฏิบัติจะมีวิธีทดสอบโดยใช้ Durbin-Watson Statistic-test โดยที่

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

เปรียบเทียบค่า d จากตาราง Durbin - Watson testจะมีค่า d_L (คือ d ของ lower limit) และ d_U (คือ d ของupper limit) ค่า d ที่คำนวณได้จะเป็นไปดังนี้

1. ถ้า d ที่คำนวณได้ $< d_L$ แสดงว่ามี auto-correlation
2. ถ้า d ที่คำนวณได้ $> d_U$ แสดงว่าไม่มี auto-correlation

ในกรณีที่ auto-correlation เกิดขึ้นจะหมายความว่าสมการที่ใช้หาความสัมพันธ์ระหว่างราคาและลักษณะจำเพาะของสินค้านั้นยังไม่น่าเชื่อถือ ควรที่จะได้แก้ไขเสียก่อน ซึ่งตามปกติมีวิธีแก้ไขได้ดังนี้

ให้ $e_i = \rho e_{i-1} + e'_i \quad ; \quad 0 < \rho \leq 1$

โดยที่

$$E(e'_i) = 0$$

$$E(e'_i e'_{i-1}) = 0$$

$$E(e'^2_i) = \sigma^2$$

$$E(e'_i e_i) = 0$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย