บทที่ 3

การวิเคราะห์ทางสถิติ

การวิเคราะห์ข้อมูลที่ได้จากแบบสอบถาม และจากการเก็บตัวอย่างของตัวแปรต่างๆ ในการวิจัยนี้จะแบ่งการวิเคราะห์ข้อมูลทางสถิติออกเป็นดังนี้

(1) การวิเคราะห์ด้านความสบายเชิงความร้อน

(1.1) วิเคราะห์หาอุณหภูมิที่ทำให้เกิดความสบาย จากสมการการโหวตเฉลี่ยของ คุณ ประพนธ์ วงษ์ท่าเรือ(1) คือ

$$Y = -4.0435 + 0.0127X1 + 8.544X2 + 0.2725X3$$

เมื่อ

Y = ค่าการโหวต

X1 = ระดับกิจกรรม(Activity Level), W/m^2

X2 = ค่าความต้านทานการถ่ายเทความร้อนของเสื้อผ้า (Clothing Thermal Resistance), m^2 °C/W

X3 = อุณหภูมิของอากาศภายในห้อง, ° C

(2) การวิเคราะห์ด้านแสงสว่าง

(2.1) วิเคราะห์หาสมการถดถอย ระหว่างการโหวตเฉลี่ยกับตัวแปรต่างๆของสภาพแวดล้อมที่มีอิทธิพลต่อการส่องสว่าง ตามสมการดังนี้

$$Y = a+b1X1+b2X2+b3X3+b4X4+b5X5+b6X6+b7X7$$

I171A0006

เมื่อ

Y = ค่าการโหวต

X1 = ระดับความส่องสว่างจากดวงโคมโดยตรง

X2,X3,X4 = สัมประสิทธิ์การสะท้อนแสงของเพดาน ผนัง และพื้น

X5 = สมรรถภาพในการมองของตา

X6 = อายุ

X7 = เพศ

(2.2) วิเคราะห์หาระดับความสว่างที่ทำให้เกิดความสบายต่อการมอง โดยพิจารณาออกเป็นช่วงอายุ และที่ค่าต่างๆกันของตัวแปรอิสระที่มีผลต่อการโหวต

(3) การวิเคราะห์ด้านเสียง

(3.1) วิเคราะห์หาสมการถดถอย ระหว่างการโหวตเฉลี่ยกับตัวแปรต่างๆของสภาพแวดล้อมที่มีอิทธิพลต่อระดับเสียงรบกวนภายในห้อง ดังนี้

$$Y = a+b1X1+b2X2+b3X3+b4X4+b5X5$$

เมื่อ

Y = ค่าการโหวต

X1 = ระดับเสียงภายในห้อง

X2 = ระดับเสียงขณะใดขณะหนึ่ง

X3 = สมรรถภาพในการได้ยินของหู

X4 = อายุ

X5 = เพศ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

(c.2) วิเคราะห์หาระดับเสียงรบกวนภายในที่ไม่มีผลทำให้เกิดความรำคาญแก่ผู้ปฏิบัติงาน
โดยพิจารณาออกเป็นช่วงอายุ และที่ค่าต่างๆกันของตัวแปรอิสระที่มีผลต่อการโหวต

3.1 การวิเคราะห์การถดถอย

เนื่องจากการวิจัยนี้ ต้องใช้การประเมินผลทางสถิติ โดยวิธีการถดถอยและสหพันธ์(Regression and Correlation) ในการวิเคราะห์ข้อมูลที่เก็บจากสถานที่ทำงานจริง เพื่อหาสมการการโหวตเฉลี่ยของคนไทย (ซึ่งในที่นี้เป็นตัวแปรตาม) ว่ามีตัวแปรอะไรบ้างที่มีนัยสำคัญ(Significance) ดังนั้นจะขอกล่าวถึงการถดถอยอย่างง่ายและพหุคูณ ดังหัวข้อย่อยต่อไป

3.1.1 การถดถอยและสหสัมพันธ์(Regression and Correlation)

หมายถึง วิธีการอย่างหนึ่งที่ใช้ในการศึกษาตัวแปรหนึ่ง เมื่อมีความสัมพันธ์หรือได้รับอิทธิพลจากตัวแปรอื่นๆ ตัวแปรที่เกี่ยวข้องกับตัวแปรที่ต้องการศึกษาอาจมีตัวเดียว การวิเคราะห์ลักษณะนี้เรียกว่า การวิเคราะห์ถดถอยและสหพันธ์อย่างง่าย(Simple Regression and Correlation Analysis) แต่ในบางกรณีตัวแปรที่เกี่ยวข้องกับตัวแปรที่ต้องการศึกษา อาจมีตั้งแต่ 2 ตัวขึ้นไป การวิเคราะห์การถดถอย ลักษณะนี้เรียกว่า การถดถอยพหุคูณ ไม่ว่าจะเป็นการวิเคราะห์อย่างง่าย หรือการถดถอยพหุคูณ วิธีการที่ใช้จะคล้ายคลึงกัน เพียงแต่การวิเคราะห์การถดถอยแบบพหุคูณ จะมีความยุ่งยากมากขึ้น เนื่องจากตัวแปรที่เกี่ยวข้องมีมากกว่า 1 ตัวยิ่งถ้ามีตัวแปรที่เกี่ยวข้องมีมากเท่าใด ความยุ่งยากก็จะมากขึ้นเท่านั้น

ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

การวิเคราะห์การถดถอยอย่างง่ายมีข้อตกลงเบื้องต้นดังนี้

(ก) ตัวแปรอิสระ X และตัวแปรตาม Y จะต้องมีความสัมพันธ์กันแบบเส้นตรง

$$Y = a + bX + e$$

เมื่อค่า e คือความคาดเคลื่อนสุ่มที่เกิดขึ้นในค่าที่สังเกตุได้ของตัวแปร Y อันเนื่องมาจากตัวอย่างที่สุ่มมา

(ข) ตัวแปรตามจะต้องเป็นตัวแปรสุ่มชนิดต่อเนื่อง ในขณะที่ตัวแปรอิสระเป็นกลุ่มของค่าต่างๆ ที่เลือกกำหนดได้

(ค) ความแปรปวนของ Y สำหรับแต่ละค่าของ X ที่กำหนดจะต้องมีค่าเท่ากัน หรือเรียกว่า Homoscedasticity สามารถพิจารณาอย่างคร่าวๆ ว่าข้อมูลที่ได้มาสอดคล้องกับคุณสมบัติข้อนี้หรือไม่ อาจทำได้โดยการพิจารณาจากแผนภาพการกระจายของข้อมูล จากกลุ่มตัวอย่าง ถ้าการกระจัดกระจายของจุดไปจากเส้นแนวโน้มมีพอๆกัน สำหรับค่าต่างๆของ X ที่สุ่มมา แสดงว่าที่ได้มาสอดคล้องกับคุณสมบัติข้อนี้

(ง) ค่าที่สังเกตุได้ในแต่ละค่าต้องไม่มีความสัมพันธ์กัน ถ้าเป็นตัวอย่างที่สุ่มมาก็จะสอดคล้องกับคุณสมบัติข้อนี้

3.1.2 การถดถอยเชิงเส้นอย่างง่าย(Sample Linear Regression)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

มีสมการพื้นฐานดังนี้

$$Y' = a + bX \qquad (3.1)$$

เมื่อ

Y′ คือ ค่าทำนายตัวแปรตาม Y

X คือ ค่าตัวแปรอิสระ

a คือ ค่าคงที่ หรือจุดตัดบนแกน y

b คือ สัมประสิทธิ์การถดถอย

โดยที่ค่า a หาได้จากสูตร

$$a = \bar{Y} - b\bar{X} \qquad (3.2)$$

และ b หาได้จากสูตร

$$b = \Sigma xy / \Sigma x^2 \qquad (3.3)$$

เมื่อ x, y คือค่าเบี่ยงเบนจากค่าเฉลี่ยของ X, Y หรือ $x = X-\bar{X}$, $y = Y-\bar{Y}$ และ $\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/N$, $\Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/N$, $\Sigma xy = \Sigma XY - (\Sigma X)(\Sigma Y)/N$ และ N คือ จำนวนข้อมูล
ส่วนวิธีการลากเส้นถดถอย มีกฎอยู่ว่าต้องลากผ่านค่าคงที่ หรือจุดตัดบนแกน y และจุด X, Y

3.1.3 ความหมายและการหาค่าสัมประสิทธิ์สหสัมพันธ์, r

ในการวิเคราะห์การถดถอย จะมีสัมประสิทธิ์ตัวหนึ่งที่เป็นดัชนีแสดงขนาด และทิศทางของความสัมพันธ์ระหว่างตัวแปร ดัชนีตัวนี้ได้แก่ สัมประสิทธิ์สหพันธ์(r) ซึ่งค่า r มีค่าอยู่ระหว่าง $-1 < R < +1$ ถ้า $r=-1.0$ เมื่อนำมาพล็อตกราฟ ค่า X, Y ทุกค่าจะอยู่บนเส้นถดถอย แต่ถ้า r ยิ่งเข้าใกล้ 0 จุด X, Y จะยิ่งกระจายห่างจากเส้นถดถอยมากขึ้น เครื่องหมายบวกหรือลบที่ติดอยู่ แสดงว่าความชันของเส้นถดถอย เป็นบวกหรือลบ ส่วนการหาค่าสัมประสิทธิ์ตัวนี้มีหลายวิธี ซึ่งแต่ละวิธีจะให้ค่าเท่ากันเช่น

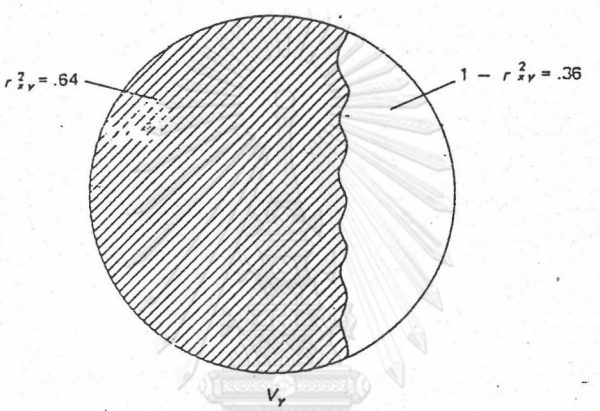
$$r_{xy} = \Sigma xy / (\Sigma x^2 \Sigma y^2)0.5 \qquad (3.4)$$

$$r_{xy} = \Sigma xy / N s_x s_y \qquad (3.5)$$

เมื่อ s_x คือค่าเบี่ยงเบนมาตรฐานของ X

3.1.4 ความหมายและการหาค่าสัมประสิทธิ์การตัดสินใจ, r^2

เป็นสัมประสิทธิ์ที่จะบอกให้ทราบถึงความแปรปรวนของ Y ที่สามารถอธิบายด้วยตัวแปร X ว่ามีได้มากน้อยเท่าไร เช่น จากรูปที่ 3.1 พื้นที่แรเงาคือสัดส่วนของความแปรปรวน Y ที่สามารถอธิบายได้ด้วยตัวแปร X=0.64 ไม่สามารถอธิบายได้ด้วยตัวแปร X=0.36 เมื่อพื้นที่วงกลม(V_y) คือความแปรปรวนทั้งหมด Y=1



รูปที่ 3.1 สัดส่วนความแปรปรวนของ Y

3.1.5 การทดสอบนัยสำคัญของการถดถอยอย่างง่าย

เนื่องจากเส้นถดถอย สำหรับชุดข้อมูลในแต่ละชุดเป็นเพียงเส้นหนึ่งที่ถูกเลือกขึ้นมา ดังนั้นความคลาดเคลื่อนที่เกิดขึ้นจึงเป็นแบบสุ่ม จึงจำเป็นต้องมีการทดสอบนัยสำคัญของสัมประสิทธิ์ถดถอย โดยการตั้งสมมติฐานเมื่อนำมาใช้กับการวิเคราะห์ การถดถอย ซึ่งมีรูปสมการดังนี้

$$SS_t = SS_{reg} + SS_{res} \qquad (3.6)$$

เมื่อ

SS_t คือ ผลรวมทั้งหมดกำลังสองของ Y

$= \Sigma y_t = \Sigma Y^2 - (\Sigma Y)^2/N$

SS_{reg} คือ ผลรวมกำลังสองของ Y เนื่องจากการถดถอย

$= \Sigma y'^2 = \Sigma Y'^2 - (\Sigma Y')^2/N$

SS_{res} คือ ผลรวมกำลังสองของ Residuals

$= \Sigma d^2 = \Sigma (Y-Y')^2$

และการทดสอบหานัยสำคัญจะใช้ F-ratio ดังสูตร

$$F = (SS_{reg}/df1)/(SS_{res}/df2) \qquad (3.7)$$

เมื่อ

df1 คือ องศาอิสระ(Degree of Freedom) ของการถดถอย(=K)

df2 คือ องศาอิสระของตัวแปรที่เหลือ(Residuals=N-K-1)

K คือ จำนวนตัวแปรอิสระ

เมื่อหาค่า F ได้แล้วจะเปิดตาราง F-distribution ในตาราง ข.3(ภาคผนวก ข) เพื่อตรวจสอบดูว่ามีนัยสำคัญหรือไม่ ถ้าค่า F ที่คำนวณจากสมการ(3.7) มากกว่าค่า F ในตาราง ก็แสดงว่าตัวแปรนั้นมีนัยสำคัญ

3.1.6 การวิเคราะห์การถดถอยพหุคูณ(Multiple Regression Anlysis)

การวิเคราะห์ที่จะกล่าวต่อไปนี้ เป็นวิธีการทั่วๆไปของการวิเคราะห์ที่มีตัวแปรอิสระกี่ตัวก็ได้ สามารถใช้ได้กับการถดถอยทุกรูปแบบ ไม่ว่าตัวแปรจะเป็นแบบต่อเนื่อง ลำดับชั้น หรือเป็นแบบระหัส การวิเคราะห์สมการที่มีตัวแปรตั้งแต่ 3 ตัวขึ้นไป จำเป็นจะต้องใช้เมตริกซ์ หา ส.ป.ส.การถดถอย รูปแบบของสมการการถดถอยพหุคูณทั่วๆไปมีดังนี้

$$Y' = a + b1X1 + b2X2 + \ldots + bkXk \qquad (3.8)$$

ค่า a, b1...k เป็นค่าที่ต้องการหาเพื่อให้สมการมีคำตอบที่ผิดเพี้ยนไปน้อยที่สุดจากข้อมูลดิบ

3.1.7 ข้อสมมติฐานในการวิเคราะห์การถดถอยพหุคูณ

ในการวิเคราะห์การถดถอยพหุคูณเชิงเส้นจะมีสมมติฐานดังนี้

1. ตัวแปรอิสระและตัวแปรตามจะมีความสัมพันธ์กันในแบบเส้นตรง สมการที่มีตัวแปรอิสระ k ตัว จะอยู่ในรูป

$$Y = a + b1X1 + b2X2 + \ldots + bkXk + e \qquad (3.9)$$

เมื่อ e คือความคลาดเคลื่อนสุ่มซึ่งเกิดจากการสุ่มตัวอย่าง

2. ตัวแปรตามเป็นตัวแปรสุ่มต่อเนื่อง ในขณะที่ตัวแปรอิสระเป็นตัวแปรควบคุมและเป็นกลุ่มของค่าที่ไม่ใช่ค่าสุ่ม ดังนั้นค่า e ในรูปแบบเชิงเส้นจึงเป็นความคลาดเคลื่อนของตัวแปรอันเกิดจากการสุ่มตัวอย่าง การวิเคราะห์การถดถอยพหุคูณ สามารถนำมาใช้ได้แม้ว่าตัวแปรอิสระจะเป็นตัวแปรสุ่มเช่นเดียวกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3. ความแปรปรวนของต้วตาม เมื่อกำหนดค่าต่างๆ ของต้วแปรอิสระจะมีค่าเท่าก้นหรือเรียกว่ามีคุณสมบัติของ Homoscedasticity
4. ค่าของต้วแปรสุ่มต้วที่อยู่ติดก้นจะต้องไม่สัมพันธ์ก้น ถ้าค่าของต้วแปรสุ่มไม่สอดคล้องกับคุณสมบัติของนี้ เรียกว่าข้อมูลมีความสัมพันธ์ซึ่งก้นและก้นอยู่ในต้ว
5. ต้วแปรอิสระแต่ละต้วจะต้องไม่มีความสัมพันธ์ก้น

ถ้าจะกล่าวก้นโดยแท้จริงแล้ว ข้อสมมติฐานทั้ง 5 เป็นสิ่งที่เกิดขึ้นอย่างสมบรูณ์ได้ยาก หากข้อสมมติฐานดังกล่าวเป็นการใกล้เคียงก้บความจริง แบบแผนของความสัมพันธ์ก็จะเป็นสิ่งที่มีประโยชน์มาก

3.1.8 คำตอบทั่วไปของสมการการถดถอย

จากสมการ(3.8) เราจะต้องใช้สมการ k สมการในการหาค่า b จากแคลคูลัส จะใช้สมการพีชคณิตเชิงเส้น(Normal Equation) สำหรับต้วแปรอิสระ 3 ต้วดังนี้

$$\begin{aligned} \beta 1 + r12\,\beta 2 + r13\beta 3 &= ry1 \\ r21\,\beta 1 + \beta 2 + r23\beta 3 &= ry2 \\ r31\,\beta 1 + r32\beta 2 + \beta 3 &= ry3 \end{aligned} \qquad (3.10)$$

จากสมการ(3.10) สามารถเขียนเป็นเมตริกซ์ ได้ดังนี้

$$\begin{bmatrix} r11 & r12 & r13 \\ r21 & r22 & r23 \\ r31 & r32 & r33 \end{bmatrix} \begin{bmatrix} \beta 1 \\ \beta 2 \\ \beta 3 \end{bmatrix} = \begin{bmatrix} ry1 \\ ry2 \\ ry3 \end{bmatrix} \qquad (3.11)$$

$$[\,Rij\,]\,[\,\beta j\,] = [\,Ryj\,]$$

เป็นที่ทราบมาแล้วว่าเมตริกซ์ สามารถ บวก, ลบ, คูณ ได้แต่ไม่สามารถนำมาหารได้แต่จะใช้อินเวอร์สเมตริกซ์ แทนจาก (3.11) จะได้

$$[\beta j] = [Rij]^{-1}[Ry1] \qquad (3.12)$$

เมื่อ $[Rij]^{-1}$ คือ อินเวอร์สเมตริกซ์ของ [Rij]

จากสมการ(3.12) จะหาค่า b จากสูตร

$$b_j = \beta j(s_y/s_j) \qquad (3.13)$$

เมื่อ

b คือ regression weight ที่ j= 1, 2, 3,...

s_y คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรตาม Y ที่ j= 1, 2, 3,...

s_j คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรอิสระ ที่ j= 1, 2, 3,...

3.1.9 ความหมายและการหาค่าสัมประสิทธิ์การตัดสินใจพหุคูณ, R^2

คือสัมประสิทธิ์ที่บอกถึงสัดส่วนของความแปรปรวน ในตัวแปรตาม Y ที่อธิบายได้ด้วยตัวแปรอิสระต่างๆในเส้นถดถอย ซึ่งหาได้จากสูตร

$$R^2 = SS_{reg}/SS_t \qquad (3.14)$$

เมื่อ

SS_{reg} หาได้จาก $\Sigma y'^2$

SS_{res} หาได้จาก Σd^2

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.1.10 ความหมายและการหาค่าสัมประสิทธิ์สหพันธ์พหุคูณ, R

คือ สัมประสิทธิ์สหพันธ์เชิงพหุคูณ ที่อยู่ในรูปสัมบูรณ์หรือไม่มีเครื่องหมาย กำกับเพราะอาจเป็นไปได้ที่ ตัวแปรตัวหนึ่งมีความสัมพันธ์ในทางตรงกันข้ามกับตัวแปรตามในขณะที่ตัวแปรตัวอื่นๆ มีความสัมพันธ์แบบผกผันกับตัวแปรตาม ดังนั้นความสัมพันธ์ระหว่างตัวแปรทั้งหมดจึงไม่สามารถบอกได้ชัดเจน ว่าเป็นบวกหรือลบ ซึ่งจะหาได้จากสูตร

$$R = (R^2)^{0.5} \qquad (3.15)$$

เช่น $R^2 = 0.66$

จะได้ $R = 0.81$

3.1.11 การทดสอบนัยสำคัญของการถดถอยพหุคูณ

การทดสอบนัยสำคัญด้วย F-ratio เพื่อทดสอบสมมติฐานว่าตัวแปรอิสระ X1...Xk มีความสัมพันธ์กับ Y ตามรูปแบบหรือไม่ ดังนี้

$H_o = \beta 1 = \beta 2 = \beta k = 0$

$H_1 = \beta j \neq 0$ อย่างน้อยหนึ่งตัว

หลังจากการตั้งสมมติฐานแล้ว จะใช้การทดสอบ F-ratio ตรวจสอบการมีนัยสำคัญ โดยใช้สูตร

$$F = \frac{R^2/K}{(1-R^2)/(N-K-1)} \qquad (3.16)$$

เมื่อ K = จำนวนตัวแปรอิสระ

เมื่อหาค่า F ได้แล้วจะเปิดตาราง F-distribution ในภาคผนวก ข เพื่อตรวจสอบว่ามีนัยสำคัญหรือไม่ ถ้าค่า F ที่คำนวณจากสมการ(3.16) มากกว่าค่า F ในตาราง ก็แสดงว่าตัวแปรนั้นมีนัยสำคัญ

3.1.12 นัยสำคัญทางสถิติของตัวแปรที่เพิ่มเข้ามาในสมการถดถอย

จากหัวข้อ 3.1.11 เรายังไม่ทราบว่า X ตัวไหนมีอิทธิพลต่อ Y มากหรือน้อยเพียงใด(เนื่องจากมี X หลายตัว) ดังนั้นจึงต้องมีการทดสอบว่า X ใดมีนัยสำคัญบ้างโดยลดตัวแปรอิสระลงทีละตัว แล้วตรวจสอบนัยสำคัญของตัวแปรนั้นว่ามีหรือไม่ ถ้าไม่มีแสดงว่าเราสามารถตัดตัวแปรนั้นออกไปจากสมการได้ การทดสอบการมีนัยสำคัญของตัวแปร โดยการเปรียบเทียบค่า F และค่า T ที่ได้จากการคำนวณ กับค่าวิกฤติ F และ T ในตาราง ข.3 และตาราง ข.4 ตามลำดับ ถ้าค่าที่คำนวณได้มากกว่าค่าวิกฤติ ก็แสดงว่าตัวแปรนั้น มีนัยสำคัญ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY