# บทที่ 3

# ทฤษฎี

## 3.1 ทฤษฎีเกี่ยวกับ Correlation

คำว่า Correlation ถ้าแปลกันตามพจนานุกรมแล้ว หมายถึง ความสัมพันธ์ ซึ่งจะต้องประกอบขึ้นมาจากสองฝ่ายเป็นอย่างน้อย ในทางสถิติเราวัดความสัมพันธ์ได้จากสูตร

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

โดยที่ $r$ = สัมประสิทธิของความสัมพันธ์ (correlation coefficient)

$x$ = ค่าของ variable ประเภทหนึ่ง

$y$ = ค่าของ variable ประเภทสอง

สูตรนี้เป็นการหาสัมประสิทธิของความสัมพันธ์ระหว่าง variables x กับ y ในวิทยานิพนธ์นี้เราจะใช้ทฤษฎีเกี่ยวกับ Correlation เพื่อหาความสัมพันธ์ระหว่างผลการศึกษาในมหาวิทยาลัยกับคะแนน ม.ศ. 5

ก่อนที่จะศึกษาเรื่อง Correlation จำเป็นที่จะต้องกล่าวถึงคำจำกัดความและทฤษฎีที่เกี่ยวเนื่องด้วย ดังจะได้อธิบายต่อไปนี้

### 3.1.1 ประชากรที่มีตัวแปรสองตัว (Bivariate Population)

คำจำกัดความที่ 3 – 1 ในการวัดค่า variable x ทุกค่า ถ้าเราทราบค่า variable y ที่สอดคล้องกัน กลุ่มของค่า variates ทั้งสองนี้ เรียกว่า ประชากรที่มีแปรสองตัว

สิ่งที่น่าสนใจเกี่ยวกับเรื่องนี้ก็คือ.–

ก. พิจารณาดูว่าในประชากรที่มีตัวแปรสองตัว ตัวแปรทั้งสองนี้มีความสัมพันธ์กันหรือไม่

ข. ถ้ามีความสัมพันธ์ เราจะหาสมการที่แสดงค่าความสัมพันธ์นั้นได้อย่างไร

เช่น ในการสำรวจหาความสัมพันธ์ระหว่าง คะแนนชีววิทยา และกลศาสตร์ เราจะหาความสัมพันธ์เพื่อทดสอบดูว่า ถ้าทราบค่าชีววิทยาแล้ว จะประมาณกลศาสตร์ได้หรือไม่

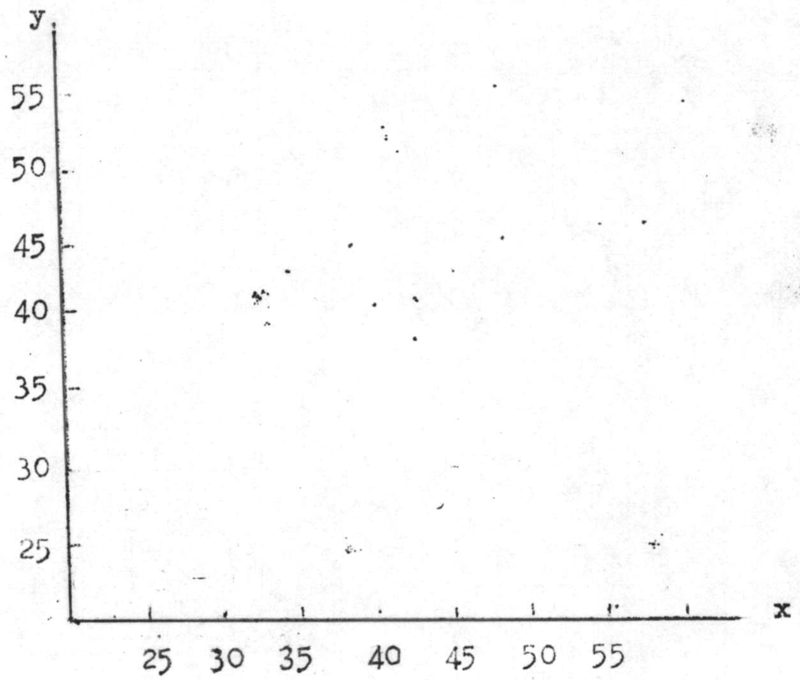
สมมติจากการสำรวจนักเรียน 10 คน เรามีข้อมูลดังตารางที่ 3 – 1

| คะแนนชีววิทยา X | คะแนนกลศาสตร์ Y |
|---|---|
| 45 | 42 |
| 42 | 40 |
| 56 | 45 |
| 48 | 55 |
| 42 | 37 |
| 35 | 42 |
| 58 | 45 |
| 40 | 40 |
| 39 | 45 |
| 50 | 45 |

ตารางที่ 3 – 1 ตัวอย่างข้อมูลที่ได้จากนักเรียน 10 คน

สิ่งที่เราต้องการจะหาคือ เมื่อทราบค่า x อื่น ๆ จะคำนวณค่า y ออกมาได้อย่างไร กล่าวคือ x เป็นตัวแปรที่ไม่ขึ้นกับค่าตัวแปรอื่น (independent variable) y คือตัวแปรซึ่งขึ้นอยู่กับค่า x (dependent variable)

จากตาราง 3 – 1 เราลองนำค่า x,y มา plot กราฟได้ดังภาพประกอบที่ 3 – 1




ภาพประกอบที่ 3 – 1 แสดงแผนผังของการกระจายตามตารางที่ 3 – 1

<u>คำจำกัดความที่ 3 – 2</u> กราฟที่แทนค่าของ x,y n คู่ใน Coordinate System เรียกว่า แผนผังของการกระจาย (Scatter diagram)

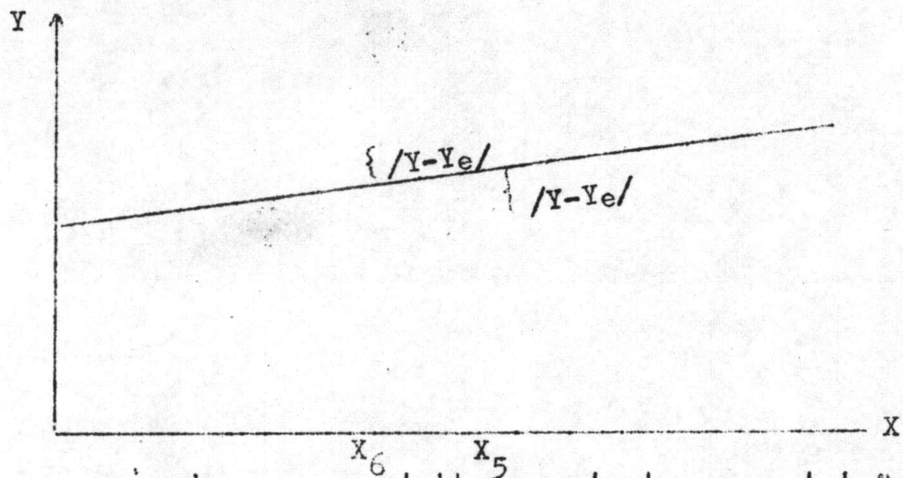
ปัญหาของเราคือจะหาสมการมาแทนแผนผังของการกระจายได้อย่างไรดีจึงจะมีค่าผิดพลาดน้อยที่สุด (least error) กล่าวคือ จากแผนผังของการกระจายตาม

ภาพประกอบที่ 3 — 1 เราสามารถแทนได้ด้วยเส้นตรงหลายเส้น และเส้นที่เราต้องการคือเส้นที่มีค่าผิดพลาดน้อยที่สุด

<u>คำจำกัดความที่ 3 — 3</u> วิธีการสร้างเส้นตรงจากค่า N ค่าของ X และ Y โดยให้ $\Sigma(Y - Y_e)^2$ มีค่าน้อยที่สุด เรียกวิธีการนี้ว่า method of least squares

โดยที่ Y = ค่าจริงที่ได้จากการสำรวจ

Ye = ค่าที่ได้จากการคำนวณ




ภาพประกอบที่ 3 — 2 แสดงค่าที่ต่างกันระหว่างค่าจริง และค่าที่ได้จากการคำนวณของ Y

ให้สมการที่แทนเส้นตรงที่ต้องการ เป็น

$$Y = a + bX \quad \ldots\ldots\ldots\ldots(1)$$

โดย Differential Calculus เราต้องการ minimize $\Sigma(Y - Y_e)^2$ จะได้ว่า a,bจะต้องเป็น solution ของสมการ

$$an + b\Sigma X = \Sigma Y \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$a\Sigma X + b\Sigma X^2 = \Sigma XY \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

ต่อไปจะหาค่า a และ b

ต่อไปจะหาค่า a และ b

(2) x $\Sigma x$

$$an\Sigma x + b(\Sigma x)^2 = \Sigma x \Sigma y \quad ..........(4)$$

(3) x n

$$an\Sigma x + bn\Sigma x^2 = n\Sigma xy \quad ..........(5)$$

(5) - (4) $b\left[n\Sigma x^2 - (\Sigma x)^2\right] = n\Sigma xy - \Sigma x\Sigma y$

$$\therefore b = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{\Sigma xy - \frac{\Sigma x\Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$= \frac{\Sigma xy - n\bar{x}\bar{y}}{x^2 - n\bar{x}^2} \quad ........(6)$$

จาก (2) $a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$

$$= \bar{y} - b\bar{x} \quad ........(7)$$

<u>คำจำกัดความที่ 3-4 :</u> เรียกสมการ (2), (3) ว่าสมการปกติ (normal equation)

<u>คำจำกัดความที่ 3-5 :</u> เส้นตรง

$$Y_e = a + bX$$

ที่หาได้จากวิธี least squares เรียกว่า line of regression of Y on X หรือสมการในการทำนายค่า Y (line of prediction for Y)

<u>การหาค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Error of Estimate)</u>

เมื่อเรามีแผนผังของการกระจาย และสามารถ fit เส้นตรงแทนจุดต่าง ๆ ได้แล้ว เราจะต้องทดสอบดูว่า เส้นตรงที่ประมาณขึ้นมาได้นั้นมีส่วนเบี่ยงเบนมาตรฐาน (standard error) เป็นเท่าไร

ในประชากรที่มีตัวแปรเพียงตัวเดียว จะหาส่วนเบี่ยงเบนมาตรฐานได้จากค่าของตัวแปรที่ต่างไปจากค่าเฉลี่ย ในกรณีที่เป็นประชากรที่มีตัวแปรสองตัว ส่วนเบี่ยงเบนมาตรฐานของค่าประมาณ (estimate) คือค่าที่จุดต่าง ๆ ห่างไปจาก regression line

<u>คำจำกัดความที่ 3-6</u> ส่วนเบี่ยงเบนมาตรฐานของ **regression line** จากตัวอย่างของ **X,Y** n คู่ คือ

$$\hat{S}_e = \sqrt{\frac{\Sigma (Y - Y_e)^2}{n}} \qquad \ldots\ldots\ldots\ldots(8)$$

จาก ( 8 ) จะเห็นว่าส่วนเบี่ยงเบนมาตรฐานนี้ จะมีค่าน้อยลงเมื่อ Y มีค่าใกล้กับ $Y_e$ และส่วนเบี่ยงเบนมาตรฐานนี้จะมีค่าเป็นศูนย์เมื่อ Y เท่ากับ $Y_e$ ทุก ๆ ค่า นั่นคือเมื่อทุก ๆ จุดของ Y อยู่บน regression line

<u>การหา correlation coefficient</u>

<u>คำจำกัดความที่ 3-7</u> correlation coefficient ของตัวอย่าง **X,Y** n คู่ คือ

$$r = \frac{\Sigma\, xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \qquad \ldots\ldots\ldots\ldots(9)$$

เมื่อ

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$



จาก ( 9 ) เราสามารถเปลี่ยนให้อยู่ในแบบที่ง่ายแก่การคำนวนเป็น

$$r = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2 \, \Sigma (Y - \bar{Y})^2}}$$

$$= \frac{\Sigma (XY - \bar{X}Y - X\bar{Y} + \bar{X}\bar{Y})}{\sqrt{\Sigma (X^2 - 2X\bar{X} + \bar{X}^2)\, \Sigma (Y^2 - 2Y\bar{Y} + \bar{Y}^2)}}$$

$$= \frac{\Sigma XY - \bar{X}\Sigma Y - \bar{Y}\Sigma X + n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - 2\bar{X}\Sigma X + n\bar{X}^2)(\Sigma Y^2 - 2\bar{Y}\Sigma Y + n\bar{Y}^2)}}$$

$$r = \frac{\Sigma XY - \bar{X}\frac{\Sigma Y}{n}.n - \bar{Y}\frac{\Sigma X}{n}.n + n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - 2\bar{X}\frac{\Sigma X}{n}.n + n\bar{X}^2)(\Sigma Y^2 - 2\bar{Y}\frac{\Sigma Y}{n}.n + n\bar{Y}^2)}}$$

$$= \frac{\Sigma XY - n\bar{X}\bar{Y} - n\bar{Y}\bar{X} + n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - 2\bar{X}^2 n + n\bar{X}^2)(\Sigma Y^2 - 2\bar{Y}^2 n + n\bar{Y}^2)}}$$

$$= \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - n\bar{X}^2)(\Sigma Y^2 - n\bar{Y}^2)}} \quad .......(10)$$

โดยปกติเวลาคำนวนจะใช้สมการ ( 10 ) เพราะสะดวกแก่การคำนวน

<u>การพิจารณาค่า limit ของ correlation coefficient</u>

<u>คำจำกัดความที่ 3-8</u> : Sum squares for regression คือ

$$S.S.R. = \Sigma(Y_e - \bar{Y})^2 \quad ..........(11)$$

เมื่อ $Y_e$ = ค่า ที่ได้จาก regression line

$\bar{Y}$ = ค่าเฉลี่ยของ Y จากตัวอย่าง n ค่า

$= \frac{\Sigma Y}{n}$

<u>คำจำกัดความที่ 3-9</u> : Sum of squares for error สำหรับ regression line คือ

$$S.S.E. = \Sigma(Y - Y_e)^2 \quad ..............(12)$$

<u>คำจำกัดความที่ 3-10</u> : Sum of squares สำหรับ regression line คือ

$$S.S. = \Sigma(Y - \bar{Y})^2 \quad ..............(13)$$

$$\frac{S.S.R.}{S.S.} = \frac{\Sigma(Y_e - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}$$

$$= \frac{\Sigma(\bar{Y} + bx - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}$$

$$= \frac{\Sigma(bx)^2}{\Sigma(y^2)}$$

$$= \frac{b^2 \Sigma x^2}{\Sigma y^2}$$

แต่ $b = \left(\frac{\Sigma xy}{\Sigma x^2}\right)$

$$\therefore \frac{S.S.R.}{S.S.} = \left(\frac{\Sigma xy}{\Sigma x^2}\right)^2 \frac{\Sigma x^2}{\Sigma y^2}$$

$$= \frac{(\Sigma xy)^2}{\Sigma x^2 \Sigma y^2}$$

แต่จาก ( 9 ) จะได้ว่า

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$\frac{S.S.R.}{S.S.} = r^2 = \frac{(\Sigma xy)^2}{\Sigma x^2 \Sigma y^2} \quad .......(14)$$

<u>คำจำกัดความที่ 3-11</u> เรียกกำลังสองของ correlation coefficient ว่า coefficient of determination

จาก Schwartz's Inequality กล่าวว่า

ถ้า $a_1, a_2, \ldots, a_n$ และ $b_1, b_2, \ldots, b_n$ เป็น real numbers ใด ๆ จะได้ว่า

$$(a_1b_1 + a_2b_2 + \ldots + a_nb_n)^2$$

$$\leq (a_1^2 + a_2^2 + \ldots + a_n^2)(b_1^2 + b_2^2 + \ldots + b_n^2)$$

หรือ

$$\Sigma(a_ib_i)^2 \leq (\Sigma a_i^2)(\Sigma b_i^2)$$

<u>พิสูจน์</u>

ให้ $\lambda$ เป็น real number ใด ๆ จะได้ว่า

$$(a_1\lambda + b_1)^2 + (a_2\lambda + b_2)^2 + \ldots + (a_n\lambda + b_n)^2 \geq 0$$

$$A^2\lambda^2 + 2C\lambda + B^2 \geq 0 \qquad \ldots\ldots\ldots(15)$$

เมื่อ $A^2 = a_1^2 + a_2^2 + \ldots + a_n^2$

$B^2 = b_1^2 + b_2^2 + \ldots + b_n^2$

$C = a_1b_1 + a_2b_2 + \ldots + a_nb_n$

จาก ( 15 )

$$\lambda^2 + \frac{2C}{A^2}\lambda + \frac{B^2}{A^2} \geq 0$$

$$\Rightarrow \quad \lambda + \frac{C}{A^2})^2 + \frac{B^2}{A^2} - \frac{C^2}{A^4} \geq 0 \qquad \ldots\ldots\ldots(16)$$

แต่ ( 16 ) จะเป็นจริงต่อเมื่อ

$$\frac{B^2}{A^2} - \frac{C^2}{A^2} \geq 0$$

$$\Rightarrow \quad C^2 \leq A^2B^2$$

$$\Rightarrow (a_1b_1 + a_2b_2 + \dots + a_nb_n)^2$$

$$\leq (a_1^2 + a_2^2 + \dots + a_n^2)(b_1^2 + b_2^2 + \dots + b_n^2)$$

$$\Rightarrow (\Sigma a_ib_i)^2 \leq (\Sigma a_i^2)(\Sigma b_i^2)$$

จาก (14) และ Schwartz's inequality จะได้ว่าเศษย่อมน้อยกว่าส่วนเสมอไป

$$r^2 \leq 1$$

$$\Rightarrow -1 \leq r \leq 1$$

ซึ่งจะสรุปคุณสมบัติของ correlation coefficient ได้ดังนี้

ก. ในกรณีที่ทุกจุดอยู่บนเส้นตรง

$$r^2 = \frac{S.S.R.}{S.S.}$$

$$= \frac{S.S.R.}{S.S.R. + S.S.E.}$$

แต่ $S.S.E. = 0$

$$r^2 = \frac{S.S.R.}{S.S.R.}$$

$$\Rightarrow r = +1 \text{ หรือ } -1$$

ข. ในกรณีที่ไม่มี linear relationship ระหว่าง X's และ Y's

$$r^2 = \frac{S.S.R.}{S.S.}$$

แต่ $S.S.R. = 0$

$$r = 0$$

ค. จาก ( 14 ) จะได้ว่า

$$0 \leq r^2 \leq 1$$

$$-1 \leq r \leq 1$$

3.2.1. ในกรณีที่มี variable หลายตัว

ในกรณีที่มี variable หลายตัวนี้ function จะอยู่ในรูป matrix

$$y = x\beta + \mu \qquad \text{........(17)}$$

โดยที่

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$x = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ B_k \end{bmatrix} \qquad U = \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_n \end{bmatrix}$$

เขียนได้เป็น $$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{ki}\beta_k + u_i$$

ในกรณีที่มี function n functions

$y_i$ = dependent variable ตัวที่ i หรือ row ที่ i ใน matrix (17)

$x_{ji}$ = independent variable ตัวที่ j ใน y ตัวที่ i หรือ column ที่ j , row ที่ i ใน matrix (17)

$\beta_j$ = regression coefficient ตัวที่ j หรือ column ที่ y ใน matrix (17)

$u_i$ = disturbance term ตัวที่ i หรือใน row ที่ i ใน matrix (17)

<u>การหาค่า Least-Squares Estimators</u>

ให้ $$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{pmatrix}$$

เป็น column vector ของ estimates ของ $\beta$

ถ้าให้ e = column vector ของ n residuals

$$= y - x\hat{\beta}$$

จะได้

$$y = x\hat{\beta} + e$$

sum of squared residuals คือ

$$\sum_{i=1}^{n} e_i^2 = e'e$$

$$= (y - x\hat{\beta})'(y - x\hat{\beta})$$

$$= (y' - (x\hat{\beta})')(y - x\hat{\beta})$$

$$= (y' - \hat{\beta}'x')(y - x\hat{\beta})$$

$$= y'y - \hat{\beta}x'y - y'x\hat{\beta} + \hat{\beta}'x'x\hat{\beta} \quad \ldots\ldots(18)$$

แต่ $\hat{\beta}'x'y$ เป็น $(1 \times k) \times (k \times n) \times (n \times 1)$

$= 1 \times 1$ vector

$=$ scalar

และ $y'x\hat{\beta}$ เป็น $(1 \times n) \times (n \times k) \times (k \times 1)$

$= 1 \times 1$ vector

$=$ scalar

$$\hat{\beta}'x'y = y'x\hat{\beta}$$

$$\therefore (18) = y'y - 2\hat{\beta}'x'y + \hat{\beta}'x'x\hat{\beta} \quad \ldots\ldots\ldots(19)$$

จะหาค่า $\beta$ ซึ่ง minimizes $\sum_{i=1}^{n} e_i^2$

$$\frac{\partial(\sum_{i=1}^{n} e_i^2)}{\partial\hat{\beta}} = -2x'y - \frac{\partial(\hat{\beta}'x'x\hat{\beta})}{\partial\hat{\beta}} \quad \ldots\ldots\ldots(20)$$

แต่จากข้อ ( 2 ) ภาคผนวกเชิงคณิตศาสตร์ จะได้ว่า

$$\frac{\partial(x'Ax)}{\partial x} = 2Ax$$

ในที่นี้ $\dfrac{\partial(\hat{\beta}'x'x\hat{\beta})}{\partial\beta} = \dfrac{\partial(\hat{\beta}'(x'x)\hat{\beta})}{\partial\beta}$

$$= 2x'x\hat{\beta}$$

∴ (20) จะได้ $\dfrac{\partial(\sum_{i=1}^{n} e_i^2)}{\partial\hat{\beta}} = -2x'y + 2x'x\beta$

ให้ $-2x'y + 2x'x\hat{\beta} = 0$

$$\Rightarrow \quad x'x\hat{\beta} = x'y$$

$$\hat{\beta} = (x'x)^{-1}(x'y) \quad \ldots\ldots\ldots\ldots(21)$$

หรือ $\hat{\beta} = ((x'x)^{-1}(x'y))'$

$$= (x'y)'((x'x)^{-1})'$$

$$= (y'x)(x'x)^{-1} \quad \ldots\ldots\ldots\ldots(22)$$

จะได้ว่า (21) เป็นสมการสำหรับ least squares estimators

<u>การหา S.D. error ของ $\hat{\beta}$</u>

$$Var(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$$

แต่ $\hat{\beta} = (x'x)^{-1}x'y \quad \ldots$ ( จาก 21 )

$$= (x'x)^{-1}x'(x\beta + U)$$

$$= (x'x)^{-1}(x'x\beta + x'U$$

$$= \beta + (x'x)^{-1}x'U$$

$$\hat{\beta} - \beta = (x'x)^{-1}x'U$$

$$\text{Var}(\hat{\beta}) = E\left[(x'x)^{-1}x'U\,U'x(x'x)^{-1}\right]$$

$$= (x'x)^{-1}x'E(UU')x(x'x)^{-1}$$

$$= (x'x)^{-1}x'\sigma^2 I_n x(x'x)^{-1}$$

$$\text{Var}(\hat{B}) = \sigma^2 I_n (x'x)^{-1}(x'x)(x'x)^{-1}$$

$$= \sigma^2 I_n (x'x)^{-1}$$

$$= \sigma^2 (x'x)^{-1} \qquad \text{..........(23)}$$

Var $(\hat{\beta}_i)$ ได้จากเทอมที่ i ของ principal diagonal ของ $(x'x)^{-1}$ คูณด้วย

Simple Correlation Coefficients

Correlation coefficients ระหว่าง dependent y independent $x_i$

$$r_{yi} = \frac{\Sigma(yx_i)}{\sqrt{(\Sigma y^2)(\Sigma x_i^2)}}$$

Multiple Correlation

Multiple correlation คือค่าที่แสดงถึงความสัมพันธ์ระหว่าง dependent variable กับ independent variables

ถ้าให้ R = multiple correlation coefficient จะได้ว่า

$$R^2 = \frac{\hat{\beta}'x'y - (1/n)(\Sigma y)^2}{y'y - (1/n)(\Sigma y)^2}$$

$$= \frac{\text{Explained SS}}{\sum_{i=1}^{n} y_i^2}$$

$$= \frac{\hat{\beta}'x'y}{y'y}$$

## Analysis of Variance

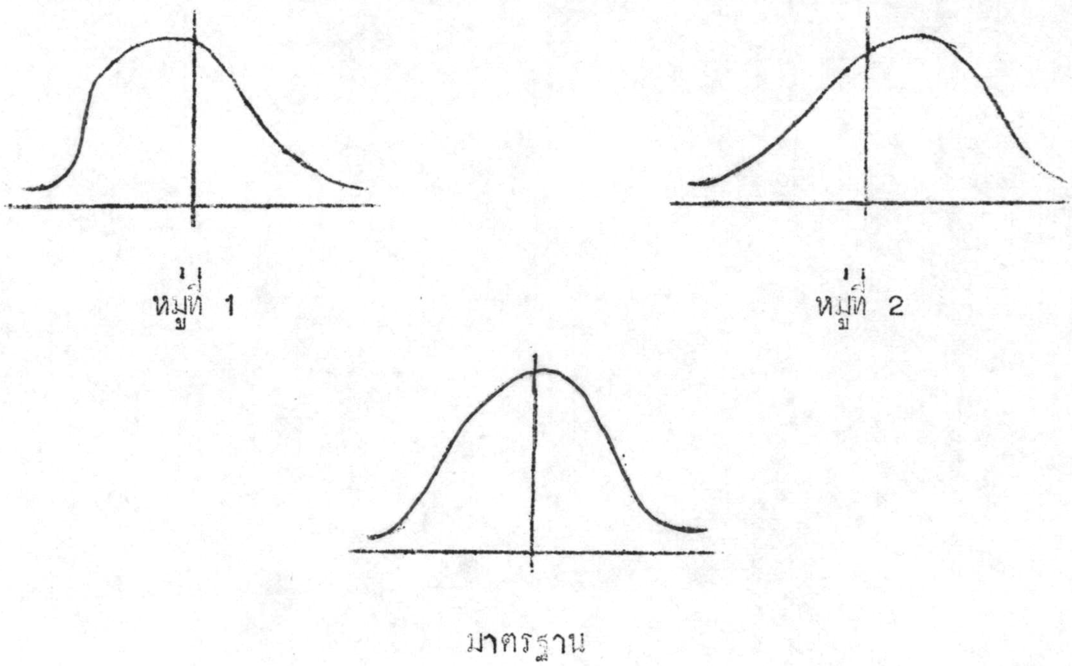
Analysis of Variance

| Source of Variation | S.S. | D.F. | Mean Sq. |
|---|---|---|---|
| $X_1, X_{2r}, X_k$ | $\beta'X'Y = Y'Y.R^2$ | $k - 1$ | $\hat{\beta}'X'Y/(k - 1)$ |
| Residual | $e'e = Y'Y(1 - R^2)$ | $n - k$ | $e'e/(n - k)$ |
| Total | $Y'Y = \sum_{i=1}^{n} Y_i^2$ | $n - 1$ | |

### 3.2. การเปลี่ยนข้อมูลให้เข้าอยู่ในแบบมาตรฐาน (Standardization)

ถ้ามีตัวแปร $x_i$'s ที่มาจากประชากรคนละหมู่ และแต่ละหมู่มีมัชฌิม (mean) และค่าความแปรปรวน (variance) ต่างกัน ก่อนจะนำ $x_i$'s จากต่างหมู่มาคำนวณหรือเปรียบเทียบกัน จะต้องเปลี่ยนให้มีการกระจาย (distribution) เหมือนกันก่อน เช่นเกี่ยวกับการบวกลบเศษส่วนต่างกัน จำเป็นจะต้องหา ค.ร.น. คือเปลี่ยนส่วนให้เท่ากันเสียก่อนจึงจะคำนวณได้

โดยทั่ว ๆ ไปในทางสถิติจะเปลี่ยนค่า $x_i$ ให้มีการกระจายแบบปกติ (Normal Distribution) คือมีค่ามัชฌิมเป็น 0 และค่าความแปรปรวนเป็น 1




ภาพประกอบที่ 3–3 แสดงการกระจายของประชากรแต่ละหมู่

สูตรในการเปลี่ยนคะแนนให้อยู่ในการกระจายแบบมาตรฐาน

$$Z = \frac{X - \bar{X}}{S}$$

เมื่อ Z = คะแนนใหม่ที่ต้องการ

X = คะแนนเดิม

$\bar{X}$ = คะแนนเฉลี่ย

S = ส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่าง นิสิตคนหนึ่งสอบได้คะแนน 80 ในวิชากฎหมายอาญา 1 ซึ่งวิชานี้มีคะแนนเฉลี่ยเป็น 70 และมีส่วนเบี่ยงเบนมาตรฐานเป็น 10 และได้คะแนน 90 ในวิชาการปกครองไทย ซึ่งวิชานี้มีคะแนนเฉลี่ยเป็น 85 และส่วนเบี่ยงเบนมาตรฐานเป็น 15 อยากทราบว่าวิชาไหนที่เรียนได้ดีกว่าเมื่อเปรียบเทียบกันแล้ว

วิธีทำ เปลี่ยนคะแนนทั้งสองวิชาให้มีการกระจายแบบมาตรฐาน คือมีคะแนนเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 คือ $X_i \frown N(0,1)$

พิจารณาคะแนนกฎหมายอาญา 1

$$\text{จาก} \quad Z = \frac{X - \bar{X}}{S}$$

$$= \frac{80 - 70}{10} = 1$$

พิจารณาคะแนนการปกครองไทย

$$Z = \frac{90 - 85}{15} = 0.34$$

จะเห็นได้ว่าเขาเรียนกฎหมายอาญา 1 ดีกว่า