บทที่ 5
การศึกษาถึงความสัมพันธ์ระหว่างตัวแปรสุ่ม
(Study of Relationship Between Random Variables)

หน้า

บทที่ 5
การศึกษาถึงความสัมพันธ์ระหว่างตัวแปรเชิงสุ่ม
(Study of Relationship between Random Variables)

5.1 Wishart Matrix (E matrix)

ในการทำการวิเคราะห์สหสัมพันธ์ (Correlation analysis) เราต้องมี matrix **E** หรือ matrix ของ sums of squares และ sums of products for errors เราเรียก matrix **E** ว่า **"Wishart matrix"** matrix **E** เป็น symmetric positive definite matrix

5.1.1 Single Sample

พิจารณา data matrix

$$\begin{vmatrix} y_{11} & \cdots & y_{p1} \\ \vdots & \vdots & \vdots \\ y_{1j} & \cdots & y_{pj} \\ \vdots & & \vdots \\ y_{1n} & \cdots & y_{pn} \end{vmatrix} \qquad \begin{vmatrix} \mathbf{y}_1 \\ \\ \mathbf{y}_j \\ \\ \mathbf{y}_n' \end{vmatrix} \quad \text{n observation vectors}$$

n คือ sample size

$\mathbf{y}_j'$ เป็น observation vector ของ random variable Y ใน p-dimensional space, j = 1, ... n

$$\text{โดยที่ } Y' = [Y_1 \ldots Y_p]$$

$$Y \sim N_p(\underline{\mu}, \Sigma)$$

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$$

$$\bar{\mathbf{y}}' = [\bar{y}_1 \; \bar{y}_2 \; \ldots \; \bar{y}_p]$$

$$\mathbf{E} = \sum_{j=1}^{n} (\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})'$$

$$= \sum_{j=1}^{n} \mathbf{y}_j\mathbf{y}_j' - n\bar{\mathbf{y}}\bar{\mathbf{y}}'$$

$$\mathbf{y}_j\mathbf{y}_j' = \begin{vmatrix} y_{1j} \\ y_{2j} \\ \\ y_{pj} \end{vmatrix} [y_{1j}\ y_{2j}\ \dots\ y_{pj}] = \begin{vmatrix} y_{1j}^2 & y_{1j}y_{2j} & \dots & y_{1j}y_{pj} \\ y_{1j}y_{2j} & y_{21}^2 & \dots & y_{2j}y_{pj} \\ \\ y_{1j}y_{pj} & y_{2j}y_{pj} & \dots & y_{pj}^2 \end{vmatrix}$$

$$\bar{\mathbf{y}}\bar{\mathbf{y}}' = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_p \end{bmatrix} [\bar{y}_1\ \bar{y}_2\ \dots\ \bar{y}_p] = \begin{vmatrix} \bar{y}_1^2 & \bar{y}_1\bar{y}_2 & \dots & \bar{y}_1\bar{y}_p \\ \bar{y}_1\bar{y}_2 & \bar{y}_2^2 & \dots & \bar{y}_2\bar{y}_p \\ \vdots \\ \bar{y}_1\bar{y}_p & \bar{y}_2\bar{y}_p & \dots & \bar{y}_p^2 \end{vmatrix}$$

$$\mathrm{E} = \begin{vmatrix} \sum_{j=1}^{n} y_{1j}^2 - n\bar{y}_1^2 & \sum_{j=1}^{n} y_{1j}y_{2j} - n\bar{y}_1\bar{y}_2 & \dots \sum_{j=1}^{n} y_{1j}y_{pj} - n\bar{y}_1\bar{y}_p \\ \sum_{j=1}^{n} y_{1j}y_{2j} - n\bar{y}_1\bar{y}_2 & \sum_{j=1}^{n} y_{2j}^2 - n\bar{y}_2^2 & \dots \sum_{j=1}^{n} y_{2j}y_{pj} - n\bar{y}_2\bar{y}_p \\ \\ \sum_{j=1}^{n} y_{1j}y_{2j} - n\bar{y}_1\bar{y}_p & \sum_{j=1}^{n} y_{2j}y_{pj} - n\bar{y}_2\bar{y}_p & \dots \sum_{j=1}^{n} y_{pj}^2 - n\bar{y}_p^2 \end{vmatrix}_{p\times p}$$

$$= \begin{vmatrix} SSy_1 & SPy_1y_2 & \dots & SPy_1y_p \\ SPy_1y_2 & SSy_2 & \dots & SPy_2y_p \\ \\ SPy_1y_p & SPy_2y_p & \dots & SSy_p \end{vmatrix}$$

$\boxed{\mathbf{S} = \frac{1}{(n-1)}\mathbf{E}}$ เป็น unbiased estimate ของ Σ

S มีชื่อว่า Sample Variance - Covariance matrix

5.1.2 k Independent Groups

ถ้าค่าสังเกตมาจาก k independent groups of sampling units

จากกลุ่มที่ h $\mathbf{Y}_h \sim N_p(\mu_h, \Sigma)$, $h = 1, \ldots, k$

$$\mathbf{Y}_h = [Y_{1h}\ Y_{2h} \ldots Y_{ph}]$$

Σ คือ common variance-covariance matrix

Data matrix ของกลุ่มที่ h คือ (n_h ขนาดของตัวอย่างสำหรับกลุ่มที่ h)

$$\begin{vmatrix} y_{11h} & \cdots & y_{p1h} \\ \cdot & & \cdot \\ y_{1jh} & \cdots & y_{pjh} \\ \vdots & & \vdots \\ y_{1n_hh} & \cdots & y_{pn_hh} \end{vmatrix} = \begin{vmatrix} y_{1h}' \\ \\ y_{jh}' \\ \\ y_{n_hh}' \end{vmatrix}$$

$$\hat{\underline{\mu}}_h = \bar{\mathbf{y}}_h = [\bar{y}_{1h}\ \bar{y}_{2h} \ldots \bar{y}_{ph}]$$

y_{ijh} = ค่าสังเกตตัวที่ j จากกลุ่มที่ h ของตัวแปรตัวที่ i

$i = 1, \ldots, p,\ j = 1, \ldots, n_h,\ h = 1, \ldots, k$

Unbiased estimate ของ Σ คือ S

$$\mathbf{S} = \frac{1}{(n-k)} \sum_{h=1}^{k} \sum_{j=1}^{n_h} (\mathbf{y}_{jh} - \bar{\mathbf{y}}_h)(\mathbf{y}_{jh} - \bar{\mathbf{y}}_h)'$$

$$= \frac{1}{(n-k)} \sum_{h=1}^{k} \mathbf{E}_h$$

โดยที่ $n = n_1 + n_2 + \ldots 4\ n_k = \sum_{h=1}^{k} n_h$

และ $\mathbf{E}_h$ คือ matrix of sums of squares and cross products ใน h th group

5.2 Wishart Distribution

เราอาจเขียน matrix **E** และ $\sum_{h=1}^{k} \mathbf{E}_h$ ให้อยู่ในรูปของ sum of products ของ (n-l) และ (n – k) independent p-dimensional random vectors ซึ่งมี Common distribution $N(\mathbf{0}, \Sigma)$

โดยทั่ว ๆ ไป symmetric positive delinite matrix **E** ของ quadratic และ bilinear forms อาจเขียนในรูปผลบวก

$$\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'$$

โดยที่

$$\mathbf{X}_i' = [X_1 \ldots, X_p]$$

และ $\mathbf{X}_i \sim NID_p(0, \Sigma)$ (NID = Normally independenly distributed)

E มี การแจกแจง เป็น **Wishart distribution**

Density function ของ symmetric positive definite matrix $\mathbf{E}_{pxp}$ คือ

$$w(\mathbf{E};\Sigma,n_e) = \frac{|\mathbf{E}|^{\frac{1}{2}(n_e-p-1)} \exp\left(-\frac{1}{2}\operatorname{tr}\mathbf{E}\,\Sigma^{-1}\right)}{\left(2^{\frac{n_e p}{2}} \pi^{\frac{1}{4}(p(p-1))}\right)|\Sigma|^{\frac{n_e}{2}} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n_e+1-i)\right]} \text{, } \mathbf{E} \text{ positive definite} \quad \ldots(1)$$

$$= 0 \quad \text{,elsewhere}$$

สำหรับ single-sample matrix **E** : $n_e = (n-1)$

สำหรับ k-sample within groups matrix : $n_e = (n-k)$

n_e คือ degrees of freedom parameter ของ **Wishart** distribution

ถ้า $p = 1$ และ $\Sigma = 1$ **Wishart** distribution กลายเป็น χ^2 – dist"

ซึ่งมี n_e df ($\Sigma = 1 \Rightarrow \sigma_Y^2 = 1$)

พิจารณา Chi-square distribution

$$\boxed{f(\chi^2) = \frac{2^{-\frac{n_e}{2}}}{\Gamma\left(\frac{n_e}{2}\right)}(\chi^2)^{\frac{n_e-2}{2}} \exp\left(-\frac{1}{2}\chi^2\right), \quad 0 \le \chi^2 < \infty \Rightarrow \chi^2_{n_e}}$$

จาก ①

$$w(E; 1, n_e) = \frac{E^{\frac{n_e-2}{2}} \exp\left(-\frac{1}{2}E\right)}{2^{\frac{n_e}{2}}\Gamma\left(\frac{n_e}{2}\right)}$$

$$= \frac{2^{-\frac{n_e}{2}}}{\Gamma\left(\frac{n_e}{2}\right)} E^{\left(\frac{n_e-2}{2}\right)} \exp\left(-\frac{1}{2}E\right), \quad 0 \le E < \infty, \; E = e_{11}$$

ซึ่งคือ $\chi^2_{n_c}$ นั่นเอง

$$\left| \begin{aligned} e_{11} &= \sum_{j=1}^{n} (y_{1j} - \bar{y}_1)^2 \ (p=1,\ y_1 = y) \\ &= \sum_{j=1}^{n} (y_j - \bar{y})^2 \\ &= \frac{1}{\sigma_Y^2} \sum_{j=1}^{n} (y_j - \bar{y})^2 = \frac{(n-1)s_Y^2}{\sigma_Y^2} \text{ เป็นค่าของ } X^2 \text{r.v} \sim \chi^2_{n-1} = \chi^2_{n_c} \end{aligned} \right|$$

5.3 Inferences for Variance – Covariance Matrices

5.3.1 การทดสอบสมมุติฐาน $H_0: \Sigma = \Sigma_0$

จากข้อมูลที่มีลักษณะใน 5.1.1

$$H_0: \underset{p \times p}{\Sigma} = \Sigma_0$$

ตัวสถิติทดสอบ $\boxed{L = \nu(\ln|\Sigma_0| - \ln|S| + \operatorname{tr} S\,\Sigma_0^{-1} - p)}$

โดยที่ ν = df ของ S

ถ้า n ใหญ่ $L \sim \chi^2_{\frac{p(p+1)}{2}}$ ถ้า H_0 จริง

Reject H_0 ถ้า $L_c > \chi^2_{\frac{p(p+1)}{2}, \alpha}$

ถ้า n มีขนาดปานกลาง ใช้ scaled statistic

$$L' = [1 - \frac{1}{6(n-1)}(2p + 1 - \frac{2}{(p+1)})]\, L$$

(เพื่อทำให้การประมาณ โดยใช้ χ^2 ดีขึ้น)

Reject H_0 ถ้า $L'_c > \chi^2_{\frac{p(p+1)}{2}, \alpha}$

ตัวอย่างที่ 5.1 จากตัวอย่างขนาด n = 20 จงทดสอบ $H_0: \Sigma = \begin{vmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{vmatrix}$ ที่ α = .05

ถ้า Sample variance-covariance matirx คือ $S = \begin{bmatrix} 3.42 & 2.60 & 1.89 \\ 2.60 & 8.00 & 6.51 \\ 1.89 & 6.51 & 9.62 \end{bmatrix}_{p \times p}$

$|\Sigma_0| = 86$, $|S| = 88.6355$, $\mathrm{tr}\, S\Sigma_0^{-1} = 3.2222$

$\nu = n - 1 = 19$, $p = 3$, $\frac{p(p+1)}{2} = 6$

$L = 19\,(\ln 86 \quad \ln 88.6355 + 3.2222 - 3)$

$= 3.65$

$$L' = [1 - \frac{1}{6\,(20-1)}(2 \times 3 + 1) - \frac{2}{(3+1)}]\ 3.65 = (1 - .0570175)\ 3.65$$

$= 3.44$

$\chi^2_{6,\ .05} = 12.6$

$L' < \chi^2_{6,\ .05} = 12.6$ ดังนั้นเราไม่ปฏิเสธ H_0 .

5.3.2 การทดสอบสมมติฐาน $H_0: \Sigma_1 = \ldots = \Sigma_k = \Sigma$

ในหัวข้อ 4.14 ได้กล่าวถึงวิธีการและตัวอย่างในการทดสอบ $H_0: \Sigma_1 = \ldots = \Sigma_k$ มาแล้ว ในที่นี้จะกล่าวอย่างสรุป

ถ้าให้ S_i เป็น unbiased estimate ของ Σ_i

S_i มี df $= n_{ei}$

ถ้าเราสุ่ม n_i observation vectors จาก population ที่ i ในกรณีปกติคือ
สุ่ม 1 ตัวอย่างจากประชากรแต่ละประชากร $n_{ei} = n_i - 1$

ถ้า H_o จริง $S = \left(\frac{1}{\sum_{i=1}^{k}(n_i - 1)}\right) \sum_{i=1}^{k}(n_i - 1)\, S_i$ จะเป็น pooled estimate ของ Σ (common variance - covariance matrix)

ตัวสถิติทดสอบคือ $\boxed{X^2 = MC}$

$$\text{โดยที่ } M = \sum_{i=1}^{k} (n_i - 1) \ln |S| - \sum_{i=1}^{k} (n_i - 1) \ln |S_i|$$

$$\text{หรือ } M = 2.3026 \left[\sum_{i=1}^{k} (n_i - 1) \log |S| - \sum_{i=1}^{k} (n_i - 1) \log |S_i|\right]$$

$$\text{และ } c = 1 - \left(\frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right) \left(\sum_{i=1}^{k} \left(\frac{1}{n_i - 1}\right) - \frac{1}{\sum_{i=1}^{k} (n_i - 1)}\right)$$

ถ้า H_0 จริง เราประมาณการกระจายของ $X^2 = MC$ ด้วย $\boxed{\chi^2_{\frac{1}{2}(k-1)p(p+1)}}$

เมื่อ $(n_i - 1)$ มีขนาดใหญ่

เรา Reject H_0 ถ้า $\chi_c^2 > \chi^2_{\frac{1}{2}(k-1)p(p+1),\, \alpha}$

ถ้า $(n_i - 1)$ เท่ากันหมดคือเท่ากับ $(n - 1)$

$$C = 1 - \frac{(2p^2 + 3p - 1)(k + 1)}{6(p + 1)\, k\, (n - 1)}$$

หมายเหตุ การประมาณข้างต้นจะดีถ้าทั้ง k และ p ไม่เกิน 4 หรือ 5 และ $(n_i - 1) \geq 20$

ตัวอย่างที่ 5.2 จงทดสอบ $H_0: \Sigma_1 = \Sigma_2$ เมื่อ $n_1 = 32$, $n_2 = 32$ และ

$$S_1 = \begin{bmatrix} 4.32 & 1.88 \\ 1.88 & 9.18 \end{bmatrix} \quad \text{และ} \quad S_2 = \begin{vmatrix} 2.52 & 1.90 \\ 1.90 & 10.06 \end{vmatrix}$$

$$\therefore S = \begin{bmatrix} 3.42 & 1.89 \\ 1.89 & 9.62 \end{bmatrix}$$

$p = 2$, $k = 2$, $n_1 = n_2 = 32 = n$ $\therefore n = 32$

$$\frac{1}{2}(k-1)\, p(p+1) = \frac{1}{2}(2-1)\, 2(2+1) = 3$$

$$M = \left\{(n_1 - 1) + (n_2 - 1)\right\} \ln|S| - (n_1 - 1)\ln|S_1| - (n_2 - 1)\ln|S_2|$$

$$= 62 \ln 29.328 - 31 \ln 36.123 - 31 \ln 21.741$$

$$= 2.82$$

$$C = 1 - \frac{(8 + 6 - 1)(2 + 1)}{6(2 + 1)\ 2(32 - 1)}$$

$$1 \quad \frac{(13)(3)}{6(3)(2)(31)} = 1 - .035 = 0.965$$

$$\chi^2_c = MC = 2.7213$$

$$\boxed{CR : \chi^2 > \chi^2_{3,\ .05} = 7.81}$$

$\chi^2_c < 7.81$ ดังนั้นเราไม่ปฏิเสธ H_0 ที่ $\alpha = .05$ นั่นคือยอมรับว่า $\Sigma_1 = \Sigma_2$

5.4 สหสัมพันธ์ (Correlations)

5.4.1 Simple Correlation (ϱ_{yz})

เราหาความสัมพันธ์ระหว่างตัวแปรเชิงสุ่ม Y และตัวแปรเชิงสุ่ม Z ได้โดยสูตร

$$\text{Corr}(Y, Z) = \varrho = \frac{\text{Cov}(Y, Z)}{\sqrt{\text{Var}(Y)\ \text{Var}(Z)}}$$

สุ่มตัวอย่างขนาด n จาก bivariate nomal distribution

Observations : (y_i, z_i), $i = 1, \ldots, n$

ให้ e_{11} = sum of squares of errors for Y

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2$$

e_{22} = sum of squares of errors for Z

$$= \sum_{i=1}^{n} (z_i - \bar{z})^2$$

e_{12} = sum of products of errors for Y and Z

$$= \sum_{i=1}^{n} (y_i - \bar{y})(z_i - \bar{z})$$

$$s_{11} = \frac{1}{(n-1)} e_{11} ,$$

$$s_{22} = \frac{1}{(n-1)} e_{22}$$

และ $s_{12} = \frac{1}{(n-1)} e_{12}$

นั่นคือ

$s_{11} = \widehat{Var(Y)}$ $\quad s_{22} = \widehat{Var(Z)}$

และ $s_{12} = \widehat{Cov(Y, Z)}$

Maximum likelihood estimate ของ ρ คือ r

โดยที่

$$r = \frac{e_{12}}{\sqrt{e_{11} e_{22}}}$$

หรือ

$$r = \frac{s_{12}}{\sqrt{s_{11} s_{22}}}$$

$$E = \begin{bmatrix} e_{11} & e_{12} \\ e_{12} & e_{22} \end{bmatrix}$$ คือ Wishart matrix

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}$$ คือ sample variance-covariance matrix ของ Y และ Z

$$= \begin{bmatrix} \widehat{Var(Y)} & \widehat{Cov(Y, Z)} \\ \widehat{Cov(Y, Z)} & \widehat{Var(Z)} \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$$ คือ correlation matrix

โดยที่

$$r_{12} = \frac{e_{12}}{\sqrt{e_{11}\, e_{22}}}$$

หรือ

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}\, s_{22}}}$$

การทดสอบสมมติฐาน $H_o : \varrho = 0$ (Test of independence)

$H_o : \varrho = 0$

$$t_c = \frac{r\sqrt{n_e - 1}}{\sqrt{1-r^2}}$$

ถ้า H_o จริง t_c จะเป็นค่าของตัวแปรเชิงสุ่ม $T \sim t_{n_e-1}$

ถ้าสุ่มตัวอย่างจาก 1 ประชากร $n_e = n-1$ โดยที่ n = ขนาดของตัวอย่าง

ดังนั้น

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

5.4.2 **Multiple Correlation** และการทดสอบ $H_o : \varrho^2_{(Y\, Z)} = 0$

เราหาความสัมพันธ์ของชุดของตัวแปรเชิงสุ่ม คือ $\mathbf{Y}' = [\, Y_1\, Y_2 \ldots Y_q \,]$ กับตัวแปรเชิงสุ่ม Z โดยการหาค่า Multiple Correlation : $\varrho_{(\underset{\sim}{Y}, Z)}$

ให้ $U = a' Y$ โดยที่เลือก $\mathbf{a}' = [\, a_1 \ldots a_q \,]$ = vector of weights ที่ทำให้ค่าสัมบูรณ์ของ simple correlation ระหว่าง U และ Z มีค่าสูงที่สุด

5.4.2.1 คำจำกัดความ simple correlation ระหว่าง $U = \mathbf{a}' Y$ และ Z ซึ่งมีค่าสูงสุดนั้น เราเรียกว่า multiple correlation ระหว่าง set ของตัวแปรเชิงสุ่ม q ตัว (Set **Y**) และตัวแปรเชิงสุ่ม Z

ให้ $\text{Var}(\underset{\sim}{Y}) = \Sigma_{11\,(q \times q)}$

$\text{Cov}(\underset{\sim}{Y}, Z) = \underset{\sim}{\sigma}_{12\,(q \times 1)}$

$\text{Var}(Z) = \sigma_{22}$

$$
\therefore\ \Sigma = \begin{array}{c} \\ \end{array}\begin{array}{cc} (q) & (1) \\ \end{array}
$$

$$
\Sigma = \begin{bmatrix} \Sigma_{11} & \sigma_{12} \\ \sigma_{12}' & \sigma_{22} \end{bmatrix} \begin{array}{c} (q) \\ (1) \end{array}
$$

$$
\begin{array}{c}
\\
5
\end{array}
\begin{array}{ccccc|c}
Y_1 & Y_2 & \cdots & Y_q & Z & \\
\hline
\mathrm{Var}(Y_1) & \mathrm{cov}(Y_1, Y_2) & \cdots & \mathrm{cov}(Y_1, Y_q) & \mathrm{Cov}(Y_1, Z) & Y_1 \\
\mathrm{Cov}(Y_1, Y_2) & \mathrm{Var}(Y_2) & \cdots & \mathrm{cov}(Y_2, Y_q) & \mathrm{Cov}(Y_2, Z) & Y_2 \\
\vdots & & & \vdots & \vdots & \vdots \\
\mathrm{Cov}(Y_1, Y_q) & \mathrm{Cov}(Y_2, Y_q) & \cdots & \mathrm{Var}(Y_q) & \mathrm{Cov}(Y_q, Z) & Y_q \\
\hline
\mathrm{Cov}(Y_1, Z) & \mathrm{cov}(Y_2, Z) & \cdots & \mathrm{Cov}(Y_q, Z) & \mathrm{Var}(Z) & Z
\end{array}
$$

แล้ว
$$
\varrho^2_{(\underset{\sim}{Y}, Z)} = \frac{1}{\sigma_{22}}\, \sigma_{12}' \left(\Sigma_{11}^{-1} \sigma_{12}\right) = \frac{1}{\sigma_{22}}\, \sigma_{12}'\, x
$$

โดยที่ x เป็น solution ของสมการ $\Sigma_{11}\, x = \underset{\sim}{\sigma}_{12}$

นั่นคือ $x = \Sigma_{11}^{-1}\, \underset{\sim}{\sigma}_{12}$

Maximum likelihood estimate ของ $\varrho^2_{(Y, Z)}$ คือ
$$
\frac{s_{12}'\, S_{11}^{-1}\, s_{12}}{s_{22}} = r^2_{(\underset{\sim}{Y}, Z)}
$$

หรือ
$$
r^2_{(Y, Z)} = \frac{1}{e_{22}}\, e_{12}'\, E_{11}^{-1}\, e_{12} = \frac{1}{e_{22}}\, e_{12}'\, x
$$

หรือ
$$
r^2_{(y, z)} = r_{12}'\, R_{11}^{-1}\, r_{12}
$$

$$
r_{22} = r_{zz} = 1
$$

โดยที่

$$
E = \begin{array}{c} \\ \\ \end{array}
\begin{array}{cc} (q) & (1) \\ \left[\begin{array}{cc} E_{11} & e_{12} \\ e_{12}' & e_{22} \end{array}\right] \end{array}
\begin{array}{c} (q) \\ (1) \end{array}
$$

$$
R = \begin{array}{c} \\ \end{array}
\begin{array}{c} \begin{array}{cc} Y & Z \end{array} \\ \left[\begin{array}{cc} R_{11} & r_{12} \\ r_{12}' & r_{zz=1} \end{array}\right] \end{array}
\begin{array}{c} Y \\ Z \end{array}
$$

5.4.2.2 การทดสอบสมมติฐาน $H_0 : \sigma_{12} = \mathbf{O} \quad \leftrightarrow \quad \varrho^2(\mathbf{Y}, Z) = 0$

ตัวสถิติทดสอบคือ F

$fc = \dfrac{r^2 \; (n_e - q)}{4 \; (1 \quad r^2)}$ ซึ่งถ้า H_o จริง f_c จะเป็นค่าของ

$F \sim F_{(q, \, n_e \, - \, q)}$

โดยที่ q = จำนวนตัวแปรเชิงสุ่มใน set − **Y**

$(q \times q)$ คือ dimension ของ $\mathbf{E}_{11}$ นั่นเอง

$n_e = n - 1$ ถ้าสุ่มตัวอย่าง 1 ตัวอย่างจากประชากรเดียว

5.4.3 สหสัมพันธ์แคนนอนนิคัล (Canonical Correlation) และการทดสอบความเป็นอิสระต่อกันของตัวแปร 2 ชุด ($H_o : \Sigma_{12} = \mathbf{O}$)

เราหาความสัมพันธ์ของตัวแปรเชิงสุ่ม 2 ชุด คือ $\mathbf{Y}' = [\, \mathbf{Y}_1 \; \mathbf{Y}_2 \dots \mathbf{Y}_p \,]$ และ $\mathbf{Z}' = [\, Z_1 \; Z_2 \dots Z_q \,]$ โดยการหาค่าสหสัมพันธ์แคนนอนนิคัล : $\varrho_{(\mathbf{Y}, \, \mathbf{Z})}$

ให้ $U = \mathbf{a}' \, \mathbf{Y}$

และ $V = \mathbf{b}' \, \mathbf{Z}$

เลือก **a** (“regression-like” parameters) และ

b (“most predictable” criterion) ซึ่งจะทำให้ค่าสัมบูรณ์ของ simple correlation ระหว่าง u และ v มีค่าสูงสุด

5.4.3.1 **ค่าจำกัดความ** simple correlation **ระหว่าง** set **-Y และ** set **-Z ดังกล่าวข้างต้น คือ** (largest) canonical correlation **ระหว่าง** set **-Y และ** set **-Z**

ให้ $Var(\mathbf{Y}) = \Sigma_{11} \rightarrow Var(U) = \mathbf{a}'\Sigma_{11}\mathbf{a}$

$Var(\mathbf{Z}) = \Sigma_{22} \rightarrow Var(V) = \mathbf{b}'\Sigma_{22}\mathbf{b}$

C O " $(\mathbf{Y},\mathbf{Z}') = \Sigma_{12} \rightarrow Cov(u, V) = \mathbf{a}'\Sigma_{12}\mathbf{b}$

$$\therefore \Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} & \begin{array}{c} (P) \\ (q) \end{array} \\ \begin{array}{cc} (P) & (q) \end{array} & \end{array}$$

$\varrho^2_{(y,z)}$ = largest characteristic root **ของ** $(\Sigma_{11}^{-1}\Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{12}')$ Maximum likelihood estimate

ของ $\varrho^2_{(y,z)}$ **คือ** $r^2_{(y,z)}$

$r^2_{(y,z)} = Char_{max}\,(\mathbf{E}_{11}^{-1}\mathbf{E}_{12}\,\mathbf{E}_{22}^{-1}\mathbf{E}_{12}')$ (หาจาก E matrix)

หรือ $= Char_{max}\,(\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\,\mathbf{S}_{22}^{-1}\mathbf{S}_{12}')$ (หาจาก S matrix)

หรือ $= Char_{max}\,(\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\,\mathbf{R}_{22}^{-1}\mathbf{R}_{12}')$ (หาจาก R matrix)

5.4.3.2 **ขั้นตอนการหาค่าของ $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ และ $r^2_{(y,z)}$**

การหาค่าประมาณของ $\mathbf{a}$, $\mathbf{b}$ และ $\varrho^2_{(y,z)}$ คือหา $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ และ $r^2_{(y,z)}$ ให้ทำดังนี้

ให้เรียก set ของตัวแปรเชิงสุ่มซึ่งมีจำนวนตัวแปร**น้อยกว่า** ว่า set **- Y** และเรียก set ของตัวแปรเชิงสุ่มซึ่งมีจำนวนตัวแปร**มากกว่า** ว่า set **- Z**

หา matrix $\underline{E}$ หรือ $\underline{S}$ หรือ $\underline{R}$

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{12}' & \mathbf{E}_{22} \end{bmatrix}$$

โดยที่ $\mathbf{E}_{11}$, $\mathbf{E}_{22}$ และ $\mathbf{E}_{12}$ คือ matrices ของ SS และ SP ของ errors สำหรับ set - **Y** set - **Z** และ **(Y, Z)**

$\mathbf{S} = \frac{1}{n_e}\mathbf{E}$ (Unbiased estimate ของ Σ)

ในที่นี้จะใช้ matrix **R** (การหาจาก **matrix E** หรือ matrix **S** นั้นทำได้โดยการแทน $\mathbf{R}_{ij}$ ด้วย E_{ij} หรือ S_{ij} ในขั้นต่าง ๆ)

R = Sample Correlation matrix

ขั้นที่ 1 **a)** หา $\mathbf{R}_{11}^{-1}$

ขั้นที่ 1 **b)** หา $\mathbf{R}_{22}^{-1}$

ขั้นที่ 2 หา $\mathbf{B} = \mathbf{R}_{22}^{-1}\mathbf{R}_{12}'$

ขั้นที่ 3 หา $C = \mathbf{R}_{12}\,\mathbf{B} = \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{12}'$

(ผลพลอยได้จากขั้นนี้ : ถ้าเราหารสมาชิกบน diagonal ของ matrix **C** ด้วยสมาชิกบน diagonal ของ $\mathbf{R}_{11}$ ที่สมนัยกันเราจะได้กำลังสองของค่าประมาณของ multiple correlation ระหว่าง Y_i กับ set - **Z**)

ขั้นที่ 4 หา $F = \mathbf{R}_{11}^{-1}C = \mathbf{R}_{11}^{-1}\,\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{12}'$

ขั้นที่ 5 หา Char_{max} **(F)** และ eigenvector ของ **F**

$\text{Char}_{max}(\mathbf{F}) = r^2_{(\mathbf{Y},\mathbf{Z})}$ ค่านี้ต้องมากกว่ากำลังสองของสมาชิกตัวใด ๆ ของ $\mathbf{R}_{12}$ ($\sqrt{r^2_{(\mathbf{Y},\mathbf{Z})}}$ > ค่าสัมบูรณ์ของสมาชิกตัวใด ๆ ของ $\mathbf{R}_{12}$)

ขั้นที่ 6 สมาชิกของ eigenvector คูณด้วยค่าคงที่ใด ๆ คือค่าประมาณของ weights

$a \rightarrow \hat{\mathbf{a}} = [\,\hat{a}_1\, \hat{a}_2 \ldots \hat{a}_p\,]$

ดังนั้น $\hat{U} = \hat{a}_1Y_1 + \hat{a}_2Y_2 + \ldots + \hat{a}_pY_p \rightarrow$ **best** predictor

ขั้นที่ 7 หา $B\,\hat{\mathbf{a}} = \hat{\mathbf{b}}$ (B จากขั้นที่ 2, $\hat{\mathbf{a}}$ จากขั้นที่ 6)

สมาชิกของ $\hat{\mathbf{b}}$ คูณด้วยค่าคงที่ใด ๆ คือค่าประมาณของ weights **b**

$\hat{\mathbf{b}} = [\,\hat{b}_1\; \hat{b}_2 \ldots \hat{b}_q\,]$

ดังนั้น $\hat{V} = \hat{b}_1Z_1 + \hat{b}_2Z_2 + \ldots + \hat{b}_qZ_q \rightarrow$ best predictable criterion

5.4.3.3 การทดสอบสมมติฐาน $H_o : \Sigma_{12} = O$ (ทดสอบความเป็นอิสระต่อกันของตัวแปรเชิงสุ่ม 2 ชุด)

ตัวสถิติทดสอบคือ $r^2_{(y, z)}$ จากขั้นที่ 5 ในหัวข้อ 5.4.3.2

$$r^2_{(Y,Z)} = \text{Char}_{\max} (\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\ \mathbf{R}_{22}^{-1}\mathbf{R}'_{12})$$

ถ้า $r^2_{(y, z)}$ มีค่ามากกว่าค่าที่อ่านจาก Heck's chart ในตาราง A2 เราจะปฏิเสธ H_o parameters สำหรับการอ่านค่าจาก Heck's Chart คือ

$$s = \min (p, q)$$

$$m = \frac{|p - q| - 1}{2}$$

$$n = \frac{n_e - p - q - 1}{2}$$ (n ที่ใช้ที่นี่ไม่ใช่ sample size)

$n_e = n - 1$ ถ้าสุ่มตัวอย่างขนาด n จากประชากรเดียว (single population)

5.4.4 Partial Correlation

5.4.41 Simplest Case :

สหสัมพันธ์ระหว่างตัวแปรเชิงสุ่ม Y และตัวแปรเชิงสุ่ม Z หลังจากได้ควบคุม (ตัด) อิทธิพลของตัวแปรเชิงสุ่ม U ออกแล้ว ("eliminating" or "partialing out" by "regression" the effect of a random variable U)

$$\varrho_{(Y, Z|U)} = \text{Corr } (Y, Z|U) = \frac{\varrho_{YZ} \ \varrho_{YU} \ \varrho_{ZU}}{\sqrt{1 - \varrho^2_{YU}}\sqrt{1 - \varrho^2_{ZU}}}$$

การทดสอบ $H_o : \varrho_{(Y,Z/U)} = 0$ (partial independance between Y และ Z given U)

$\varrho_{(Y, Z|U)} = r_{(Y, Z|U)}$ คือ first - order partial correlation

$$r_{(Y, Z|U)} = \frac{r_{YZ} - r_{YU}\, r_{ZU}}{\sqrt{1 - r^2_{YU}}\sqrt{1 - r^2_{ZU}}}$$

คำนวน $t_c = \dfrac{r_{(Y,Z|U)}\sqrt{n_e - 2}}{\sqrt{1 - r^2_{(Y,Z|U)}}}$

ซึ่งเป็นสูตรเดียวกับสูตรใน 5.4.1 แต่แทน r ด้วย $r_{(Y,Z|U)}$ และแทน n_e ด้วย $n_e - 1$

pth -order partial correlation คือ $r_{(Y,Z|U_1,U_2, ..., up)}$

2nd - order partial correlation คือ $r_{(Y,Z|U_1,U_2)}$

$$r_{(Y,Z|U_1,U_2)} = \frac{r_{(Y,Z|U_1)} - r_{(Y,U_2|U_1)}r_{(Z,U_2|U_1)}}{\sqrt{1 - r^2_{(Y,U_2|U_1)}}\sqrt{1 - r^2_{(Z,U_2|U_1)}}}$$

$$\text{หรือ} = \frac{r_{(Y,Z|U_2)} - r_{(Y,U_1|U_2)}\, r_{(Z,U_1|U_2)}}{\sqrt{1 - r^2_{(Y,U_1|U_2)}}\sqrt{1 - r^2_{(Z,U_1|U_2)}}}$$

5.4.4.2 General Case

ตัวแปรเชิงสุ่ม 3 ชุดคือ set - Y set - z และ set - u ซึ่งจะถูกตัดออก

โดยที่ $Y' = [Y_1 \dots Y_p]$

$Z' = [Z_1 \dots Z_q]$

$U' = [U_1 \dots U_r]$

$$\Sigma = \begin{array}{c} \begin{array}{ccc} (p) & (q) & (r) \end{array} \\ \left[\begin{array}{c|c|c} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \hline \Sigma'_{12} & \Sigma_{22} & \Sigma_{23} \\ \hline \Sigma'_{13} & \Sigma'_{23} & \Sigma_{33} \end{array}\right] \end{array} \begin{array}{c} (p) \\ (q) \\ (r) \end{array}$$

ในการคำนวณค่าประมาณของ $\varrho^2_{(Y,Z|U)} = r^2_{(Y,Z|U)}$ อาจหาจากเมตริกซ์ **E** หรือ **S** หรือ **R** โดยใช้สูตรต่าง ๆ ใน 5.4.2 และ 5.4.3

$$\Sigma^*_{11} = \Sigma_{11} - \Sigma_{13}\Sigma^{-1}_{33}\Sigma'_{13} = \text{Var}(\mathbf{Y}|\mathbf{U} = \mathbf{u}_0)$$

$$\Sigma^*_{22} = \Sigma_{22} - \Sigma_{23}\Sigma^{-1}_{33}\Sigma'_{23} = \text{Var}(\mathbf{Z}|\mathbf{U} = \mathbf{u}_0)$$

$$\Sigma^*_{12} = \Sigma_{12} - \Sigma_{13}\Sigma^{-1}_{33}\Sigma'_{23} = \text{Cov}(\mathbf{Y}, \mathbf{Z}'|\mathbf{U} = \mathbf{u}_0)$$

โปรดสังเกต ถ้า Σ_{11} เป็น scalar, Σ_{12}เป็น vector $\rightarrow r_{(Y,Z|U)}$= multiple - partial correlation

ถ้าทั้งหมดเป็น matrices $\rightarrow r_{(Y,Z|U)}$ = canonical - partial correlation

- การคำนวณทำโดยใช้สูตรจาก 5.4.2 และ 5.4.3 แต่ให้เติม * ที่ matrices ทุกตัวในตัวสถิติทดสอบ และ df ทุกตัวแทน n_e ด้วย $(n_e - r)$ (r = # สมาชิกใน set – U)

5.4.4.3 การทดสอบ $H_o: \rho^2_{(\underline{Y}, \underline{Z}|\underline{U})} = 0$

คำนวณ $r^2_{(\underline{Y}, \underline{Z}|\underline{U})}$ แล้วเทียบกับค่าจาก Heck's chart โดยที่ parameters คือ

$s = \min(p, q)$

$$m = \frac{|p - q| - 1}{2}$$

$$n = \frac{(n_e - r) - p - q - 1}{2}$$

ถ้า $r^2_{(\underline{Y}, \underline{Z}|\underline{U})} <$ ค่าจาก chart เราจะไม่ปฏิเสธ H_o แสดงว่า set – **Y** และ set – **Z** เป็นอิสระต่อกันหลังจากได้ partial out set – U ออกแล้ว

ถ้า
$$\mathbf{R}^*_{11} = \mathbf{R}_{11} - \mathbf{R}_{13}\mathbf{R}^{-1}_{33}\mathbf{R}'_{13}$$

$$\mathbf{R}^*_{22} = \mathbf{R}_{22} - \mathbf{R}_{23}\mathbf{R}^{-1}_{33}\mathbf{R}'_{23}$$

$$\mathbf{R}^*_{12} = \mathbf{R}_{12} - \mathbf{R}_{13}\mathbf{R}^{-1}_{33}\mathbf{R}'_{23}$$

$$\boxed{r^2_{(\underline{Y}, \underline{Z}|\underline{U})} = \text{Char}_{\max}(\mathbf{R}^*_{11})^{-1}\mathbf{R}^*_{12}(\mathbf{R}^*_{22})^{-1}\mathbf{R}^{*\prime}_{12}}$$

5.4.5 **ตัวอย่างเรื่อง** Correlations

ตัวอย่างที่ 5.1 **จาก** sample correlation matrix **ของ** Y, Z **และ** U **โดยที่**

$\mathbf{Y} = [\ Y_1\ Y_2\ ] \quad \rightarrow \quad p = 2$

$\mathbf{Z} = [\ Z_1\ Z_2\ Z_3\ ] \quad \rightarrow \quad q = 3$

$\mathbf{U} = [\ U_1\ U_2\ U_3\ ] \quad \rightarrow \quad r = 3$

จงหา 1) multiple correlations **ระหว่าง** Y_i, i = 1, 2 **กับ** Z

2) canonical correlation **ระหว่าง** Y **กับ** Z **และทดสอบ** $H_o : \Sigma_{12} = \underset{\sim}{0}$ **ที่** $\alpha = .05$

3) canonical partial correlation **ระหว่าง** Y **และ** Z **เมื่อ** partialed out U **และทดสอบ** $H_o : \varrho^2_{(Y,Z|U)} = 0$ **ที่** $\alpha = .05$

ข้อมูลได้จากตัวอย่างจากประชากรหนึ่งประชากรโดยที่ n = 110

$$\mathbf{R} = \begin{array}{c} \\ \left[\begin{array}{cc|ccc|ccc} 1.00 & 0.58 & -.06 & 0.41 & 0.42 & 0.09 & 0.58 & 0.70 \\ 0.58 & 1.00 & -.20 & 0.59 & 0.53 & -.17 & 0.80 & 0.77 \\ \hline -.06 & -.20 & 1.00 & -.03 & -.01 & 0.67 & -.42 & -.03 \\ 0.41 & 0.59 & -.03 & 1.00 & 0.48 & -.10 & 0.62 & 0.56 \\ 0.42 & 0.53 & -.01 & 0.48 & 1.00 & -0.12 & 0.48 & 0.64 \\ \hline 0.09 & -.17 & 0.67 & -.10 & 0.12 & 1.00 & -.48 & 0.09 \\ 0.58 & 0.80 & -.42 & 0.62 & 0.48 & -.48 & 1.00 & 0.68 \\ 0.70 & 0.77 & -.03 & 0.56 & 0.64 & 0.09 & 0.68 & 1.00 \end{array}\right] \end{array} \begin{array}{c} Y \\ \\ Z \\ \\ U \end{array} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R'_{12} & R_{22} & R_{23} \\ R'_{13} & R'_{23} & R_{33} \end{bmatrix}$$

(columns: Y | Z | U)

กำหนด $\mathbf{R}_{22}^{-1} = \mathbf{R}_{zz}^{-1} = \begin{bmatrix} 1.000926 & .032775 & -.005723 \\ 0.032775 & 1.300449 & -.623888 \\ -0.005723 & -.623888 & 1.299409 \end{bmatrix}$

$\mathbf{R}_{22}^{*-1} = \mathbf{R}_{zz}^{*-1} = \begin{bmatrix} & -.399681 & .135049 \\ -.399681 & 1.899359 & -.286754 \\ .135049 & -.286754 & 1.829402 \end{bmatrix}$

ขั้นที่ 1a หา $\mathbf{R}_{11}^{-1}$

$$\mathbf{R}_{11} = \begin{bmatrix} 1 & 0.58 \\ 0.58 & 1 \end{bmatrix}, \quad |R_{11}| = 1 - (.58)^2 = .6636$$

$$\mathbf{R}_{11}^{-1} = \frac{1}{.6636}\begin{bmatrix} 1 & -0.58 \\ -0.58 & 1 \end{bmatrix} = \begin{bmatrix} 1.506932 & -.874020 \\ -.874020 & 1.506932 \end{bmatrix}$$

ขั้นที่ 1b) หา $\mathbf{R}_{22}^{-1}$

โจทย์กำหนดให้ $\mathbf{R}_{22}^{-1} = \begin{bmatrix} -.00.032775 & -.623888 & -.04-.623888 \end{bmatrix}$

โดยที่ $\mathbf{R}_{22} = \begin{bmatrix} 1 & -.03 & -.01 \\ -.03 & 1 & .48 \\ -.01 & .48 & 1 \end{bmatrix}$

ขั้นที่ 2 หา $\mathbf{B} = \mathbf{R}_{22}^{-1}\,\mathbf{R}'_{12}$

$$= \mathbf{R}_{22}^{-1}\begin{bmatrix} -.06 & -.20 \\ .41 & .59 \\ .42 & .53 \end{bmatrix}$$

ขั้นที่ 3 หา $\mathbf{C} = \mathbf{R}_{12}\,\mathbf{B} = \mathbf{R}_{12}\,\mathbf{R}_{22}^{-1}\,\mathbf{R}'_{12}$

$$= \begin{bmatrix} -.20 & .59 & .53 \\ -.06 & .41 & .42 \end{bmatrix}\begin{bmatrix} -.04902147 & \\ .26918463 & .43004927 \\ .29030108 & .32173745 \\ & -.18388114 \end{bmatrix}$$

$$= \begin{bmatrix} .235233 & .322483 \\ .322483 & .461026 \end{bmatrix}$$

1) หา multiple correlation ระหว่าง Y_i, i= 1, 2 กับ set – Z

จาก matrix C ถ้าเราหารสมาชิกบน diagonal ของ C ด้วยสมาชิกของ $\mathbf{R}_{11}$ ในตำแหน่งที่สมนัยกัน เราจะได้ค่าประมาณของ multiple correlation ของ Y_i กับ set – Z

$$\therefore r^2_{(Y_1, Z)} = \frac{.235233}{1} = .235233 \rightarrow \sqrt{r^2_{(Y_1, Z)}} = .4850$$

$$r^2_{(Y_2, Z)} = \frac{.461021}{1} = .461026 \rightarrow \sqrt{r^2_{(Y_2, Z)}} = .6790$$

ขั้นที่ 4 หา $F = \mathbf{R}_{11}^{-1}\mathbf{C} = \mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{12}'$ ($\mathbf{R}_{11}^{-1}$ จากขั้นที่ 1a)

$$F = \begin{bmatrix} -.874020 & 1.506932 \\ 1.506932 & -.874020 \end{bmatrix} \begin{bmatrix} .322483 & .461026 \\ .235233 & .322483 \end{bmatrix}$$

$$= \begin{bmatrix} .280362 & .412878 \\ .072624 & .083014 \end{bmatrix}$$

ขั้นที่ 5 หา $\text{Char}_{\max}(\mathbf{F})$

$|\mathbf{F}| = .0067108808$

$\text{tr } F = .072624 + .412878 = .485502$

$\lambda^2 - (\text{tr } F)\ \lambda + |\mathbf{F}| = 0$

$\lambda^2 - .485502X + .0067108808 = 0$

$$A = \frac{-(-.485502) \pm \sqrt{(-.485502)^2 - 4\ (.0067108808)}}{2}$$

$\therefore \lambda_1 = .4713,\ \lambda_2 = .0142$

$$\therefore \text{Char}_{\max}(\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{12}') = .4713 = r^2_{(Y, Z)}$$

$\therefore \sqrt{r^2_{(Y,Z)}} = .6865$ ซึ่งมีค่ามากกว่าค่าสัมบูรณ์ของสมาชิกตัวใด ๆ ของ $\mathbf{R}_{12}$

$$\mathbf{R}_{12} = \begin{bmatrix} r_{(Y_1, Z_1)} & r_{(Y_1, Z_2)} & r_{(Y_1, Z_3)} \\ r_{(Y_2, Z_1)} & r_{(Y_2, Z_2)} & r_{(Y_2, Z_3)} \end{bmatrix}$$

$$= \begin{bmatrix} -.06 & .41 & .42 \\ -.20 & .59 & .53 \end{bmatrix}$$

ขั้นที่ 6 และ 7 (ทำเพิ่มจากโจทย์ตัวอย่าง)

ขั้นที่ 6 หา Characteristic vector หรือ eigenvector สำหรับ $\lambda_1 = \text{Char}_{max}$ **(F)**

$$\mathbf{Fx} = \lambda_1 \mathbf{x} \qquad , x = \text{eigenvector}$$

$$\mathbf{Fx} = .4713\ \mathbf{x}$$

ให้ $x_1 = 1$

$$\begin{bmatrix} .072624 & .083014 \\ .280362 & .412878 \end{bmatrix}\begin{bmatrix} 1 \\ x_2 \end{bmatrix} = .4713 \begin{bmatrix} 1 \\ x_2 \end{bmatrix}$$

$$.072624 + 0.83014\, x_2 = 4713$$

$$x_2 = 4.802515234$$

$$\therefore \mathbf{x} = \begin{bmatrix} 1 \\ 4.8 \end{bmatrix}$$

$$\mathbf{\hat{a}'} = [\,\hat{a}_1\ \hat{a}_2\,] = \text{Constant}\ [\,x_1 \quad x_2\,]$$

$$= 10\ [\,x_1\ x_2\,]$$

$$= [\,1\,0 \quad 48\,]$$

$$\hat{U} = \hat{a}_1\ Y_1 + \hat{a}_2 Y_2$$

$\therefore \hat{U} = 10Y_1 + 48Y_2$ เป็น estimator ของ the best predictable criterion (หรือเป็น the best predictor ก็ได้ทั้งนี้แล้วแต่การกำหนดตัวแปร)

ขั้นที่ 7 หา $\hat{b}$ จาก $\mathbf{B}\hat{a} = \hat{b}$

B จากขั้นที่ 2 $\quad$ **â** จากขั้นที่ 6

$$\begin{bmatrix} -.04902147 & -.18388114 \\ .26918463 & .43004927 \\ .29030108 & .32173745 \end{bmatrix}\begin{bmatrix} 1 \\ \\ 4.802515234 \end{bmatrix} = \begin{bmatrix} -.932113 \\ 2.334503 \\ 1.83545 \end{bmatrix}$$

$$= \begin{bmatrix} -.9 \\ 2.3 \\ 1.8 \end{bmatrix}$$

$\therefore \hat{V} = -9Z_1 + 23Z_2 + 18Z_3$ เป็น estimator ของ the best predictor

2) ค่าประมาณของค่าสัมบูรณ์ของ Canonical correlation $\sqrt{r^2_{(Y,Z)}} = .6865$ (จากขั้นที่ 5)

$$r^2_{(Y,Z)} = .4713$$

การทดสอบ $H_0 : \Sigma_{12} = 0 \leftrightarrow \varrho^2_{(Y,Z)} = 0$ da = .05

ตัวสถิติทดสอบ คือ $r^2_{(Y,Z)} = .4713$

parameters เพื่อเปิดค่าจาก Heck's chart คือ

$s = \min(p, q) = \min(2, 3) = 2$

$$m = \frac{|p-q|-1}{2} = \frac{|2-3|-1}{2} = 0$$

$$n = \frac{n_e - p - q - 1}{2} \quad (n_e = n - 1 = 110\text{-}1 = 109)$$

$$= \frac{109-2-3-1}{2} = 51.5$$

จาก Heck's chart [s = 2, m = 0, n = 51.5] $\alpha = .05$

จุดวิกฤตคือ .095

$$\because r^2_{(Y,Z)} = .4713 > .095 \text{ เราปฏิเสธ } H_0 \text{ ที่ } \alpha = .05$$

สรุปว่ามีความสัมพันธ์ระหว่าง set - Y และ set - Z

3) หาค่าประมาณของค่าสัมบูรณ์ของ Canonical Partial correlation ระหว่าง set - Y และ set - Z เมื่อ partialed out set - U แล้ว

ค่าประมาณของค่าสัมบูรณ์ของ Canonical partial correlation = $r_{(Y,Z|U)}$

$$r^2_{(Y,Z|U)} = \text{Char}_{\max}\left[(\underset{\sim}{R}^*_{11})^{-1} (\underset{\sim}{R}^*_{12}) (\underset{\sim}{R}^*_{22})^{-1} (\underset{\sim}{R}^*_{12})' \right]$$

$$\underset{\sim}{R}^*_{11} = \underset{\sim}{R}_{11} - \underset{\sim}{R}_{13} \underset{\sim}{R}^{-1}_{33} \underset{\sim}{R}'_{13}$$

$$\underset{\sim}{R}^*_{22} = \underset{\sim}{R}_{22} - \underset{\sim}{R}_{23} \underset{\sim}{R}^{-1}_{33} \underset{\sim}{R}'_{23}$$

$$\underset{\sim}{R}^*_{12} = \underset{\sim}{R}_{12} - \underset{\sim}{R}_{13} \underset{\sim}{R}^{-1}_{33} \underset{\sim}{R}'_{23}$$

$$\mathbf{R}_{33} = \begin{bmatrix} 1.00 & -.48 & .09 \\ -.48 & 1.00 & .68 \\ .09 & .68 & 1.00 \end{bmatrix}, \quad |\mathbf{R}_{33}| = .240348$$

$$\mathbf{R}_{33}^{-1} = \frac{1}{.240348} \begin{bmatrix} \begin{vmatrix} 1 & .68 \\ .68 & 1 \end{vmatrix} & -\begin{vmatrix} -.48 & .68 \\ .09 & 1 \end{vmatrix} & \begin{vmatrix} -.48 & 1 \\ .09 & .68 \end{vmatrix} \\ \begin{vmatrix} .48 & .09 \\ .68 & 1 \end{vmatrix} & \begin{vmatrix} 1 & .09 \\ .09 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & -.48 \\ .09 & .68 \end{vmatrix} \\ \begin{vmatrix} -.48 & .09 \\ 1 & .68 \end{vmatrix} & -\begin{vmatrix} 1 & .09 \\ .48 & .68 \end{vmatrix} & \begin{vmatrix} 1 & -.48 \\ -.48 & 1 \end{vmatrix} \end{bmatrix}$$

$$= \frac{1}{.240348} \begin{bmatrix} .5376 & .5412 & -.4164 \\ .5412 & .9919 & -.7232 \\ -.4164 & -.7232 & .7696 \end{bmatrix}$$

$$= \begin{bmatrix} 2.236757 & 2.251735 & -1.732488 \\ 2.251735 & 4.126933 & -3.008970 \\ -1.732488 & -3.008970 & 3.202024 \end{bmatrix}$$

หา $\mathbf{R}^*_{11} = \mathbf{R}_{11} - \mathbf{R}_{13}\,\mathbf{R}_{33}^{-1}$ R’,,

$$\mathbf{B}_1 = \mathbf{R}_{33}^{-1}\,\mathbf{R}'_{13} = \begin{bmatrix} 2.9457283 & .08712355 \\ .48999829 & .60184455 \\ .34029028 & .35290544 \end{bmatrix}$$

$$\text{C,} = \mathbf{R}_{13}\,\mathbf{B}_1 = \begin{bmatrix} .548914 & .603945 \\ .603945 & .738402 \end{bmatrix}$$

$$\mathbf{R}^*_{11} = \text{R,,} - \mathbf{C}_1 = \begin{bmatrix} .451086 & -.023945 \\ -.023945 & .261598 \end{bmatrix}$$

$= .1174298324$

$$(R^*_{11})^{-1} = \frac{|R^*_{11}|}{} = \begin{bmatrix} .023945 & .451086 \end{bmatrix}$$

$$\begin{bmatrix} 2.227696273 & .2039090026 \\ .2039090026 & 3.841323715 \end{bmatrix}$$

$$R^*_{22} = \begin{bmatrix} 538503 & .109917 & -.022524 \\ .109917 & .561690 & .079929 \\ -.022524 & .079929 & .560818 \end{bmatrix}$$

$(R^*_{22})^{-1}$ กำหนดให้จากโจทย์

$$R^*_{12} = \begin{bmatrix} -.041356 & -.054904 & -.068334 \\ .004989 & .027942 & .0048 \end{bmatrix}$$

HI $D = (R_{11})^{-1}\ R^*_{12}\ (R^*_{22})^{-1}\ (R^*_{12}) = \begin{bmatrix} .031487 & -.005938 \\ -.007795 & .004778 \end{bmatrix}$

Ml $\text{Char}_{max}\ (\mathbf{D}) = r^2_{(\underline{Y},\ \underline{Z}\,|\,\underline{U})}$

$\text{tr}\ \mathbf{D} = .036265$

$|\mathbf{D}| = .0001041582$

$\lambda^2 - .036265X + .0001041582 = 0$

$\lambda_1 = .03312,\ \lambda_2 = .003145$

$\text{Char}_{max}\ (\mathbf{D}) = r^2_{(\underline{Y}\ \underline{Z}\,|\,\underline{U})} = .03312$

$\sqrt{r^2_{(\underline{Y},\ \underline{Z}\,|\,\underline{U})}} = .181989$ = ค่าประมาณของค่าสัมบูรณ์ ของ canonical partial correlation

การทดสอบ $H_o : \varrho^2_{(Y, Z|U)} = 0$ ที่ $\alpha = .05$

$r = 3$ (จำนวนตัวแปรเชิงสุ่มใน set - U)

$s = 2$

$m = 0$

$$n = \frac{(n_e - r) - p - q - 1}{2} = \frac{(109 - 3) - 2 - 3 - 1}{2} = 50$$

จาก Heck's Chart ใน A2 $s = 2$, $m = 0$, $n = 50$, $\alpha = .05$ จุดวิกฤต คือ .1

$\because r^2_{(Y, Z|U)} = .03312 < .1$

เราไม่ปฏิเสธ H_0 ที่ $\alpha = .05$ สรุปว่า set - **Y** และ set **Z** เป็นอิสระต่อกันหลังจากตัด set - **U** ออกแล้ว

หมายเหตุ ใน 2) การทดสอบ $H_o : \varrho^2_{(Y, Z)} = 0$ นั้นเราปฏิเสธ H_o และสรุปว่า set-**Y** และ set-**Z** มีความสัมพันธ์กัน ทั้งนี้เพราะเราไม่ได้เอา set-**U** ซึ่งมีอิทธิพลต่อ set-**Y** และ set-**Z** มาพิจารณาด้วย นั่นคือถือว่า set-**U** เป็นค่าคงที่

แบบฝึกหัดที่ 5

5.1 ในการศึกษาถึงความสัมพันธ์ระหว่าง Wechsler Adult Intelligence Scale (WAIS) กับอายุ

Correlation matrix ของ Y_1, Y_2, Z_1, Z_2 คือ

$$
\mathbf{R} = \begin{bmatrix} 1 & .45 & -.19 & .43 \\ .45 & 1 & -.02 & .62 \\ -.19 & -.02 & 1 & -.29 \\ .43 & .62 & -.29 & 1 \end{bmatrix}
$$

โดยที่

$$
\left.\begin{array}{lcl} Y_1 & = & \text{คะแนนคณิตศาสตร์} \\ Y_2 & = & \text{คะแนนทดสอบศัพท์} \end{array}\right\} \text{จาก WAIS}
$$

$$
\begin{array}{lcl} Z_1 & = & \text{อายุตามปฏิทิน} \\ Z_2 & = & \text{ปีที่เรียนในโรงเรียน} \end{array}
$$

ข้อมูลได้จากตัวอย่างประกอบด้วยชายและหญิงอายุระหว่าง 24-64 ปี จำนวน 933 คน (n = 933)

จงทดสอบที่ $\alpha = .05$ ว่า $\mathbf{Y} = [\,Y_1\ Y_2\,]$ และ $\mathbf{Z} = [\,Z_1\ Z_2\,]$ เป็นอิสระต่อกันนั่นคือทดสอบ $H_o : \rho^2_{(Y, Z)} = 0$ (คำตอบ $r^2_{(Y, Z)} = .4285$)

5.2 จากโจทย์ข้อ 5.1 จงหา $\hat{a}$ และ $\hat{b}$ เพื่อหา $\hat{U} = \hat{a}_1Y_1 + \hat{a}_2Y_2$ และ $\hat{V} = \hat{b}_1Z_1 + \hat{b}_2Z_2$ ซึ่งเป็น the best predictor และthe best predictable criterion ตามลำดับ

5.3 หา multiple correlations ระหว่าง $Y_i,\ i = 1, 2$ กับ set-Z