วารสารศึกษาศาสตร์ ปีที่ ๔ ฉบับที่ ๑ มกราคม - ธันวาคม ๒๕๔๗

Journal of Education, Volume 4, Number 1 January - December 2004

การวิเคราะห์ความแปรปรวนตัวแปรพหุ
(Multivariate Analysis of Variance : MANOVA)

*เสริม ทัศศรี**

บทนำ

การวิจัยที่มีวัตถุประสงค์เพื่อต้องการตรวจสอบค่าเฉลี่ยของประชากรที่มี 3 กลุ่มขึ้นไปว่ามีความแตกต่างกันหรือไม่ และถ้าพบว่าแตกต่างกันก็จะต้องหาว่าค่าเฉลี่ยของกลุ่มใดแตกต่างกัน การวิเคราะห์ข้อมูลดังกล่าวจะใช้สถิติการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) คู่กับการเปรียบเทียบพหุคูณ (Multiple comparisons) และการวิเคราะห์ความแปรปรวนอาจจำแนกแบบได้ตามปัจจัยที่ทำให้เกิดความผันแปรของผลการวัด เช่น ถ้ามีปัจจัยเดียวจะเรียกว่าการวิเคราะห์ความแปรปรวนแบบทางเดียว (One – Way Analysis of Variance) และการวิเคราะห์ความแปรปรวนแบบสองทาง (Two – Way Analysis of Variance) เป็นต้น การวิเคราะห์ดังกล่าวจะใช้กับตัวแปรตามแบบเดี่ยว (Univariate) ในการวิจัยบางปัญหาที่มีเงื่อนไขเหมือนกับ ANOVA แต่มีตัวแปรตามสองตัวหรือมากกว่าและมีการกระทำพร้อมกัน (Simultanous) และตัวแปรตามอาจมีความสัมพันธ์กัน การวิเคราะห์เพื่อตรวจสอบค่าเฉลี่ยของประชากรว่าแตกต่างกันหรือไม่โดยกระทำพร้อมกันทุกตัวแปรจะใช้การวิเคราะห์ความแปรปรวนตัวแปรพหุ (MANOVA) เพื่อให้สะดวกในการทำความเข้าใจ MANOVA ผู้เขียนจึงนำเสนอรายละเอียดเนื้อตามลำดับหัวข้อต่อไปนี้คือ ลักษณะของข้อมูล รูปแบบ (Model) ทางสถิติของ MANOVA ข้อตกลงเบื้องต้นเกี่ยวกับโครงสร้างข้อมูล การหา sum of squares and cross – product (SSCP) สมมติฐานและตาราง MANOVA การทดสอบค่าสถิติ ตัวอย่างการวิเคราะห์และการแปลผล

ลักษณะของข้อมูล

ข้อมูลที่จะนำมาวิเคราะห์ MANOVA เริ่มจากหน่วยตัวอย่างแต่ละหน่วยจะมีค่าสังเกตเป็นชุด มีจำนวนเท่ากับจำนวนตัวแปรตามหรือมิติของการวัด (dimension) ซึ่งแทนด้วย p เช่น ถ้ามีตัวแปรตาม 3 ตัว p = 3 ค่าสังเกตจะอยู่ในรูปของเวคเตอร์ (vector) สมมติมีประชากร g กลุ่ม แต่ละกลุ่มมีขนาดเป็น n_1, n_2, ..., n_g ค่าสังเกตของตัวแปรตามในประชากรแต่ละกลุ่มจะมีโครงสร้างดังนี้

* ผู้ช่วยศาสตราจารย์ ประจำภาควิชาการประเมินผลและวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ

ประชากร 1	:	$\mathbf{X}_{11}$	, $\mathbf{X}_{12}$	, . . .	, $\mathbf{X}_{1n1}$
ประชากร 2	:	$\mathbf{X}_{21}$	, $\mathbf{X}_{22}$	, . . .	, $\mathbf{X}_{2n2}$
⋮	:	⋮	, ⋮	, . . .	, ⋮
ประชากร g	:	$\mathbf{X}_{g1}$	, $\mathbf{X}_{g2}$	, . . .	, $\mathbf{X}_{gng}$

ค่าสังเกตของหน่วยตัวอย่างใด ๆ แทนด้วย X_{ℓ_j} โดยที่ ℓ = 1 , 2 , ... , g และ j = 1 , 2 , ... , n_ℓ และเวคเตอร์ของค่าสังเกตมีมิติเป็น PX1 ฉะนั้นค่าเฉลี่ยที่จะนำมาเปรียบเทียบกันก็จะเป็น เวคเตอร์ที่มีมิติเหมือนกับค่าสังเกต คือ มีมิติเป็น PX1

ข้อตกลงเบื้องต้นเกี่ยวกับโครงสร้างข้อมูล

1. $X_{\ell 1}$, $X_{\ell 2}$, ..., $X_{\ell n}$ เป็นตัวอย่างสุ่มขนาด n_ℓ มาจากประชากรที่มีค่าเฉลี่ย $_\ell$; ℓ = 1, 2, ..., g และตัวอย่างแต่ละกลุ่มเป็นอิสระจากกัน
2. ประชากรทุกกลุ่มมีเมทริกซ์ความแปรปรวนร่วมกัน (Common Covariance Matrix : Σ)
3. ประชากรแต่ละกลุ่มเป็นแบบปกติตัวแปรพหุ (Multivariate Normal) กรณีขนาดตัวอย่างใหญ่ เงื่อนไขข้อนี้ไม่เข้มงวดนัก

รูปแบบ (Model) ทางสถิติของ MANOVA

รูปแบบทางสถิติของ MANOVA สำหรับเปรียบเทียบเวคเตอร์ค่าเฉลี่ย (mean vactor) ของประชากร g กลุ่ม เหมือนกับรูปแบบของ ANOVA แต่ค่าสังเกตในรูปแบบของ MANOVA จะเป็นเวคเตอร์ คือ

$$_{\ell j} = \quad + \quad _{\ell} + \mathbf{e}_{\ell j} \quad , \quad \ell = 1, 2, ..., g \text{ และ } j = 1, 2, ..., n_\ell$$

เมื่อ $\mathbf{e}_{\ell_j}$ เป็นตัวแปรที่มีการแจกแจงแบบปกติแบบ **Np (O , Σ)**

(การแจกแจงแบบปกติตัวแปรพหุที่มีมิติเป็น p มีเวคเตอร์ค่าเฉลี่ยเป็น **O** และเมริกซ์ความแปรปรวนร่วมเป็น $\sum$) μ เป็นเวคเตอร์ค่าเฉลี่ยของประชากรทั้งหมดและ τ_ℓ เป็นเวคเตอร์ผลของการจัดกระทำ (treament effect) $\sum_{\ell=1}^{g} n_\ell \; _\ell = 0$ ซึ่งส่วนประกอบแต่ละส่วนของเวคเตอร์ค่าสังเกต $_{\ell_j}$ จะสอดคล้องกับรูปแบบของตัวแปรเดี่ยว

การหา SSCP (sum of squares and cross – product)

ส่วนประกอบของเวคเตอร์ค่าสังเกตสามารถเขียนได้ในรูปแบบดังนี้

$$_{\ell j} \quad = \quad \bar{\ } \quad + \quad (\bar{\ }_{\ell} - \bar{\ }) \quad + \quad (_{\ell j} - \bar{\ })$$

(ค่าสังเกต) (ค่าเฉลี่ยของกลุ่มตัวอย่างทั้งหมด , $\hat{\mu}$) (ค่าประมาณของผลการจัดกระทำ $\hat{\tau}_{\ell}$) (ค่าประมาณของความคลาดเคลื่อน , $\hat{e}_{\ell j}$)

ในการวิเคราะห์ความแปรปรวนแบบทางเดียวจะมีการหาค่า SS (sum of squares) $(_{\ell j} - \bar{\ })(_{\ell j} - \bar{X})'$ และแยกค่า SSt (sum of squares total) ออกเป็น SSb (sum of squares between) และ SSw (sum of squares within) การวิเคราะห์ความแปรปรวนตัวแปรพหุก็อาศัยเหตุผลดังกล่าวมาแยกค่า SS ในทำนองเดียวกัน แต่เนื่องจากส่วนประกอบของรูปแบบใน MANOVA เป็นเวคเตอร์ ดังนั้น SS จึงเป็นผลคูณของเวคเตอร์และทรานโพสต์ (transpose) ได้ผลคูณออกมาในรูปของ SSCP เช่น กำลังสองของจำนวน เขียนในรูปของ $_{\ell j} - \bar{\ }$ $(_{\ell j} - \bar{\ })^2$ แต่สำหรับใน MANOVA $_{\ell j} - \bar{\ }$ จะเป็นเวคเตอร์เขียนในรูปของการคูณไขว้ (cross – product : CP) ดังนั้น จากสมการ ❶ สามารถเขียนในรูปของ CP ได้ดังนี้

$$
\begin{aligned}
(_{\ell j} - \bar{\ })(_{\ell j} - \bar{\ })' &= \left[(_{\ell j} - \bar{\ }_{\ell}) + (\bar{\ }_{\ell} - \bar{\ })\right]\left[(X_{\ell j} - \bar{\ }_{\ell}) + (\bar{\ }_{\ell} - \bar{\ })\right]' \\
&= (_{\ell j} - \bar{\ }_{\ell})(_{\ell j} - \bar{\ }_{\ell})' + (X_{\ell j} - \bar{\ }_{\ell})(\bar{\ }_{\ell} - \bar{\ })' + \ell \\
&\quad (\bar{\ }_{\ell} - \bar{\ })(_{\ell j} - \bar{\ }_{\ell})' + (\bar{\ }_{\ell} - \bar{\ })(\bar{\ }_{\ell} - \bar{\ })'
\end{aligned}
$$

เมื่อรวมตลอดค่าของ j ผลคูณของเวคเตอร์ตรงกลาง 2 ชุด เป็นเมทริกซ์ศูนย์ทั้งคู่ เพราะว่า $\sum_{j=1}^{n_\ell}(_{\ell j} - \bar{\ }_{\ell}) = 0$ และรวมตลอดค่าของ g ดังนั้นจะได้ SSCP ดังนี้

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(_{\ell j} - \bar{\ })(_{\ell j} - \bar{\ })' = \sum_{\ell=1}^{g} n_\ell(\bar{\ }_{\ell} - \bar{\ })(\bar{\ }_{\ell} - \bar{\ })' + \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(_{\ell j} - \bar{\ }_{\ell})(_{\ell j} - \bar{\ }_{\ell})'$$

(totalSSCP (corrected)) (treatmentSSCP (between)) (residualSSCP (within))

เมทริกซ์ของ within SSCP (W) เขียนได้ในรูปต่อไปนี้

$$W = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\ _{\ell j} - \overline{\ }_{\ell})(\ _{\ell j} - \overline{\ }_{\ell})'$$

$$= (n_1\text{-}1)\mathbf{S_1} + (n_2\text{-}1)\mathbf{S_2} + \ \ldots + (n_g\text{-}1)\mathbf{S_g} \qquad \text{- - - ❷}$$

เมื่อ S_ℓ เป็นเมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่างที่ ℓ และเมทริกซ์นี้เป็นกรณีทั่วไปของ $(n_1 + n_2 - 2)\mathbf{S}_{pooled}$ ซึ่งมีบทบาทสำคัญในการทดสอบผลของการจัดกระทำในกรณีที่มีตัวอย่าง 2 กลุ่ม

สมมติฐานและตาราง MANOVA

สมมติฐานเป็นกลาง ($\mathbf{H_0}$) ที่ใช้ทดสอบผลของการจัดกระทำเป็นลักษณะเดียวกับตัวแปรเดี่ยว แต่อยู่ในรูปของเวคเตอร์ คือ

$$\mathbf{H_0} \ : \ \tau_1 = \tau_2 = \ldots = \tau_g = 0$$

ในการทดสอบ $\mathbf{H_0}$ ใช้การพิจารณาจากขนาดสัมพัทธ์ของ SSCP ระหว่างการจัดกระทำ (treatment) และส่วนที่เหลือ (residual) และพิจารณาขนาดสัมพัทธ์ของ SSCP ระหว่างส่วนที่เหลือและผลรวมปรับแก้ โดยมี df เหมือนกับการวิเคราะห์แบบตัวแปรเดี่ยว สำหรับตารางวิเคราะห์จะมีส่วนประกอบเหมือนกับการวิเคราะห์ความแปรปรวน แต่ SS จะเป็นผลรวมของผลคูณของเวคเตอร์และทรานโพสต์แทนที่จะเป็นผลรวมของกำลังสองในแต่ละส่วน ดังตาราง

ตาราง MANOVA สำหรับเปรียบเทียบเวคเตอร์ค่าเฉลี่ยของประชากร

Source of variation	Matrix of sum of squares and cross – products (SSCP)	Degree of freedom (df)
Treatment	$\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\overline{X}_\ell - \overline{X})(\overline{X}_\ell - X)'$	g – 1
Residual (Error)	$\mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(X_{\ell j} - \overline{X}_\ell)(X_{\ell j} - \overline{X}_\ell)'$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (correct for the mean)	$\mathbf{B + W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(X_{\ell j} - \overline{X})(X_{\ell j} - \overline{X})'$	$\sum_{\ell=1}^{g} n_\ell - 1$

การทดสอบค่าสถิติ

การทดสอบค่าสถิติเพื่อปฏิเสธหรือยอมรับ H_0 ซึ่งพิจารณาได้จากค่า $\wedge^*$ (Wilks' Lambda) กล่าวคือ จะปฏิเสธ H_0 ถ้า $\wedge^*$ มีค่าเล็กมาก สำหรับค่า $\wedge^*$ หาได้โดยใช้สัดส่วนของ $|\mathbf{W}|$ และ $|\mathbf{B}+\mathbf{W}|$ คือ

$$\wedge^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} = \frac{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(X_{\ell j}-\overline{X}_\ell)(X_{\ell j}-\overline{X}_\ell)'\right|}{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(X_{\ell j}-\overline{X})(X_{\ell j}-\overline{X})'\right|}$$

ค่า $\wedge^*$ มีรูปแบบที่สมนัยกับการแจกแจงแบบเอฟ (F–distribution) ที่ใช้ทดสอบ H_0 ในตัวแปรเดี่ยว สำหรับกรณีตัวแปรเดี่ยวถือเป็นกรณีเฉพาะของการแจกแจงของ $\wedge^*$ ส่วนกรณีทั่วไปซึ่งเป็นตัวแปรพหุซึ่งจำนวนกลุ่มตัวอย่างมีหลากหลายและกลุ่มตัวอย่างขนาดใหญ่ สามารถปรับขยายการแจกแจงของ $\wedge^*$ ที่สมนัยกับการแจกแจงแบบเอฟ เพื่อใช้ทดสอบ H_0 ให้สอดคล้องกับเงื่อนไขจำนวนตัวแปรตามและจำนวนกลุ่มตัวอย่างได้ ดังตาราง

ตารางการแจกแจง $\wedge^*$ ของ Wilks' เมื่อ $\wedge^* = |\mathbf{W}|/|\mathbf{B}+\mathbf{W}|$ - - - ❸

จำนวนตัวแปร	จำนวนกลุ่ม	การแจกแจงแบบสุ่มสำหรับข้อมูลตัวแปรพหุปกติ
$p = 1$	$g \geq 2$	$\left(\frac{\sum n_\ell - g}{g-1}\right)\left(\frac{1-\wedge^*}{\wedge^*}\right) \sim F_{g-1,\sum n_\ell - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\sum n_\ell - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right) \sim F_{2(g-1),2(\sum n_\ell - g - 1)}$
$p \geq 3$	$g = 2$	$\left(\frac{\sum n_\ell - p - 1}{p}\right)\left(\frac{1-\wedge^*}{\wedge^*}\right) \sim F_{p,\sum n_\ell - p - 1)}$
$p \geq 1$	$g = 3$	$\left(\frac{\sum n_\ell - p - 2}{p}\right)\left(\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right) \sim F_{2p,2(\sum n_\ell - p - 2)}$

ถ้า H_0 เป็นจริง และ $\sum n_\ell = n$ มีขนาดใหญ่ ค่าสถิติข้างล่าง คือ

$$-\left(n-1-\left(\frac{p+g}{2}\right)\right)\ln \wedge^* = -\left(n-1-\left(\frac{p+g}{2}\right)\right)\ln\left(\frac{|W|}{|B+W|}\right)$$

จะมีลักษณะการแจกแจงคล้ายไคสแควร์ โดยมี df เท่ากับ p(g-1) และจะปฏิเสธ $\mathbf{H_0}$ ที่ระดับนัยสำคัญ α ถ้าค่าสถิติ

$$-\left(n-1-\left(\frac{p+g}{2}\right)\right)\ln\left(\frac{|W|}{|B+W|}\right) > {}^{2}_{p(g-1)}$$ - - - ❹

ตัวอย่างการวิเคราะห์และการแปลผล

ตัวอย่าง 1

สมมติมีตัวแปรตามที่สนใจศึกษา 2 ตัวแปร ($p = 2$) ทำการสังเกตตัวแปรตาม 2 ตัวพร้อมกัน จากกลุ่มตัวอย่าง 3 กลุ่ม ($g = 3$) ที่มีขนาดเป็น $n_1 = 3$, $n_2 = 2$ และ $n_3 = 3$ ได้ค่าสังเกตเป็นคู่ ℓj จัดเป็นแถวได้ดังนี้

$$\left(\begin{array}{ccc} \begin{bmatrix}9\\3\end{bmatrix} & \begin{bmatrix}6\\2\end{bmatrix} & \begin{bmatrix}7\\9\end{bmatrix} \\ \begin{bmatrix}0\\4\end{bmatrix} & \begin{bmatrix}2\\0\end{bmatrix} & \\ \begin{bmatrix}3\\6\end{bmatrix} & \begin{bmatrix}1\\9\end{bmatrix} & \begin{bmatrix}2\\7\end{bmatrix} \end{array}\right)$$

โดย $\overline{X}_1 = \begin{bmatrix}8\\4\end{bmatrix}$, $\overline{X}_2 = \begin{bmatrix}1\\2\end{bmatrix}$, $\overline{X}_3 = \begin{bmatrix}2\\8\end{bmatrix}$

และ $\overline{\mathbf{X}} = \begin{bmatrix}4\\5\end{bmatrix}$

ข้อมูลจากการสังเกตในตัวแปรแรกข้างบน นำมาแสดงในรูปของผลรวมค่าเฉลี่ยทั้งหมด ผลของการจัดกระทำและส่วนที่เหลือโดยใช้ ANOVA แบบตัวแปรเดี่ยว จะได้ว่า

$$\begin{pmatrix}9 & 6 & 7\\0 & 2 & \\3 & 1 & 2\end{pmatrix} = \begin{pmatrix}4 & 4 & 4\\4 & 4 & \\4 & 4 & 4\end{pmatrix} + \begin{pmatrix}4 & 4 & 4\\-3 & -3 & \\-2 & -2 & -2\end{pmatrix} + \begin{pmatrix}1 & -2 & -1\\-1 & 1 & \\1 & -1 & 0\end{pmatrix}$$

ค่าสังเกต ค่าเฉลี่ยทั้งหมด ผลของการจัดกระทำ ส่วนที่เหลือ

และ SS คือ

$$SS_{obs} = SS_{mean} + SS_{tr} + SS_{res}$$

$$216 = 128 + 78 + 10$$

$$\text{TotalSS (corrected)} = SS_{obs} - SS_{mean}$$

$$= 216 - 128$$

$$= 88$$

สำหรับตัวแปรที่ 2 กระทำในลักษณะเดียวกัน จะได้ว่า

$$\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & \\ 5 & 5 & 5 \end{pmatrix} + \begin{pmatrix} -1 & -1 & -1 \\ -3 & -3 & \\ 3 & 3 & 3 \end{pmatrix} + \begin{pmatrix} -1 & -2 & 3 \\ 2 & -2 & \\ 0 & 1 & -1 \end{pmatrix}$$

ค่าสังเกต ค่าเฉลี่ยทั้งหมด ผลของการจัดกระทำ ส่วนที่เหลือ

และ

$$\begin{aligned} SS_{obs} &= SS_{mean} + SS_{tr} + SS_{res} \\ 272 &= 200 + 48 + 24 \\ \text{TotalSS (corrected)} &= SS_{obs} - SS_{mean} \\ &= 272 - 200 \\ &= 72 \end{aligned}$$

สำหรับวิธีคำนวณ $\overline{X}$ ผลของการจัดกระทำ และส่วนที่เหลือ หาได้ดังนี้

$$\overline{X}_1 = \begin{bmatrix} (9+6+9)/3 \\ (3+2+7)/3 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

$$\overline{X}_2 = \begin{bmatrix} (0+2)/2 \\ (4+0)/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\overline{X}_3 = \begin{bmatrix} (3+1+2)/3 \\ (8+9+7)/3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$

$$\overline{X} = \begin{bmatrix} (9+6+9+0+2+3+1+2)/9 \\ (3+2+7+4+0+8+9+7)/9 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

ผลของการจัดกระทำ $\left(\hat{\tau}_\ell\right) = \left(\overline{}_\ell - \overline{}\right)$

ดังนั้น $\frown_1 = 8 - 4 = 4$, $\frown_2 = 1 - 4 = -3$ และ $\frown_3 = 2 - 4 = -2$

$$\hat{\tau}_\ell = \begin{pmatrix} 8-4 & 8-4 & 8-4 \\ 1-4 & 1-4 & \\ 2-4 & 2-4 & 2-4 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix}$$

ส่วนที่เหลือ $\hat{e}_{\ell j} = (X_{\ell j} - \overline{X}_{\ell})$

$$\hat{e}_{\ell j} = \begin{pmatrix} 9-8 & 6-8 & 9-8 \\ 0-1 & 2-1 & \\ 3-2 & 1-2 & 2-2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}$$

สำหรับตัวแปรที่ 2 สามารถหาผลของการจัดกระทำและส่วนที่เหลือใช้วิธีการคำนวณเช่นเดียวกับตัวแปรที่ 1

การวิเคราะห์ทั้งสองส่วนประกอบกันเพื่อหาผลรวมของ CP กระทำโดยรวมผลคูณของค่าตัวแปรทั้งสองที่อยู่ในลำดับเดียวกัน แถวต่อแถวจะได้ CP เพื่อทำให้ได้ตาราง MANOVA มีความสมบูรณ์ CP ที่ได้ประกอบด้วย

ค่าเฉลี่ย : 4(5) + 4(5) + ... + 4(5) = 8(4)(5) = 160

ผลของการจัดกระทำ : 3(4)(-1) + 2(-3)(-3) + 3(-2)(3) = -12

ส่วนที่เหลือ : 1(-1) + (-2)(-2) + 1(3) + (-1)(2) + ... + 0(-1) = 1

ผลรวม : 9(3) + 6(2) + ... + 2(7) = 149

ผลรวม CP (ปรับแก้) = ผลรวม CP - ค่าเฉลี่ย CP

= 149 - 160

= -11

ดังนั้นตาราง MANOVA จะมีรูปแบบดังนี้

Source of vatiation	Matrix of sum of squares and cross – products (SSCP)	Df
Treatment (B)	$\begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$	3 – 1 = 2
Residual (W)	$\begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$	3 + 2 + 3 – 3 = 5
Total (corrected) (B + W)	$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$	7

ตรวจสอบสมการได้ดังนี้

$$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix} = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix} + \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$$

หาค่า $\wedge^*$ โดยใช้สมการ ❸ จะได้ว่า

$$\wedge^* = \frac{|W|}{|B+W|} = \frac{\begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}}{\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}} = \frac{10(24) - (1)^2}{88(72) - (11)^2}$$

$$= \frac{239}{6,215} = 0.0385$$

เนื่องจาก p = 2 และ g = 3 ในตารางการแจกแจง $\wedge^*$ (สมมติความเป็นปกติของการแจกแจงและความเท่ากันของเมทริกซ์ความแปรปรวนร่วม) ชี้ให้เห็นถึงความเฉพาะในการทดสอบ H_0 (p = 2, g = 3) โดยสมมติฐานที่ทดสอบ คือ $\mathbf{H_0}$: ${}_1 = \tau_2 = \ldots = {}_g = 0$ (ไม่มีผลของการจัดกระทำ) และ $\mathbf{H_1}$: ${}_\ell \neq 0$ (ผลของการจัดกระทำมีความแตกต่างกันอย่างน้อย 1 คู่) เพื่อที่จะหาผลการทดสอบทางสถิติ จึงคำนวณค่าสถิติที่ทดสอบแล้วนำไปเปรียบเทียบกับค่าสถิติ F ที่มี $v_1 = 2(g-1)$, $v_2 = 2(\Sigma n_e - g - 1)$ (จากตารางการแจกแจง $\wedge^*$ เมื่อ p = 2, g $\geq$ 2) ค่าสถิติที่ทดสอบ คือ

$$\left(\left(\frac{\sum n_\ell - g - 1}{(g-1)}\right)\left(\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right) = \left(\frac{8-3-1}{3-1}\right)\left(\frac{1-\sqrt{0.0385}}{\sqrt{0.0385}}\right) = 8.19 \right)$$

ค่า F จากตารางการแจกแจงแบบ F ที่ $\alpha = .01$, $v_1 = 4$, $v_2 = 8$ คือ $F_{4,8\,(.01)} = 7.01$ ซึ่ง $8.19 > F_{4,8\,(.01)}$ ดังนั้นจึงปฏิเสธ H_0 ที่มี $\alpha = .01$ ซึ่งสรุปได้ว่า ผลของการจัดกระทำมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ในกรณีที่ p มีขนาดใหญ่ ข้อมูลก็จะมีมากขึ้นตามไปด้วย ทำให้ไม่สะดวกในการหาเมทริกซ์ B และ W ด้วยเครื่องคำนวณธรรมดา ดังนั้นในการสร้างตาราง MANOVA จึงใช้คอมพิวเตอร์หาเมทริกซ์ B และ W

จากตัวอย่าง 1 ทำให้ทราบแนวทางการคิดคำนวณจากข้อมูลที่สังเกต แต่หากข้อมูลที่สังเกต ถูกนำมากระทำในระดับหนึ่งแล้ว คือ หาค่าเวคเตอร์ค่าเฉลี่ยของตัวแปรและเมทริกซ์ความแปรปรวนร่วมของตัวแปรในแต่ละกลุ่มไว้ชั้นหนึ่งแล้ว สามารถทดสอบ $\mathbf{H_0}$ ได้ตามตัวอย่าง 2

ตัวอย่าง 2

เทศบาลมหานครแห่งหนึ่งต้องการศึกษาค่าใช้จ่ายรายเดือนของนักเรียนนักศึกษาในแต่ละระดับว่ามีค่าใช้จ่ายในแต่ละหมวดแตกต่างกันหรือไม่ และค่าใช้จ่ายหมวดใดของนักเรียนนักศึกษามีความแตกต่างกัน จึงศึกษาจากนักเรียนมัธยมศึกษา นักเรียนอาชีวศึกษาและนักศึกษาระดับอุดมศึกษา ที่ตั้งอยู่ในเขตเทศบาลมหานคร โดยศึกษาค่าใช้จ่ายรายเดือนในหมวดค่าอาหาร ค่าอุปกรณ์การเรียน ค่าสันทนาการ และค่าติดต่อสื่อสาร โดยคิดค่าใช้จ่ายรายเดือนมีหน่วยเป็นพันบาท ได้เวคเตอร์ค่าเฉลี่ยของค่าใช้จ่ายและเมทริกซ์ความแปรปรวนร่วมของค่าใช้จ่ายระหว่างหมวดต่าง ๆ แต่ละระดับดังนี้

ระดับ	จำนวน
มัธยมศึกษา	$n_1 = 271$
อาชีวศึกษา	$n_2 = 138$
อุดมศึกษา	$n_3 = 107$

$$\sum_{\ell=1}^{3} n_\ell = 516$$

เวคเตอร์ค่าเฉลี่ยของกลุ่มตัวอย่าง

$$\overline{X}_1 = \begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix} \quad ; \quad \overline{X}_2 = \begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix} \quad ; \quad \overline{X}_3 = \begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix}$$

เมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่าง

$$S_1 = \begin{pmatrix} .291 & & & \\ -.001 & .011 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{pmatrix}$$

ดัดแปลงจาก : Richard A. Johnson and Dean W. Wichern, 1988.

ตัวอย่างนี้ g = 3, p = 4 จากสมการ ❷ หาค่า **W** ได้ดังนี้

$$\mathbf{W} = (n_1\text{-}1)\mathbf{S_1} + (n_2\text{-}1)\mathbf{S_2} + (n_3\text{-}1)\mathbf{S_3}$$

$$\mathbf{W} = 270\begin{pmatrix} .291 & & & \\ -.001 & .011 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{pmatrix} + 137\begin{pmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{pmatrix} +$$

$$106\begin{pmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{pmatrix}$$

$$= \begin{pmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.581 & 2.428 & .394 & 6.538 \end{pmatrix}$$

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1 + n_2 + n_3}$$

$$\bar{X} = \frac{271}{561}\begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix} + \frac{138}{561}\begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix} + \frac{107}{561}\begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix} = \begin{bmatrix} 2.136 \\ .519 \\ .102 \\ .380 \end{bmatrix}$$

เนื่องจากกลุ่มตัวอย่างมีขนาดใหญ่จึงมีเหตุผลที่จะสรุปว่า $\sum_1 = \sum_2 = \sum_3$ นั่นคือ กลุ่มตัวอย่างมีเมทริกซ์ความแปรปรวนร่วมกัน **(S_{pooled})**

การทดสอบสมมติฐานที่ว่า ค่าใช้จ่ายรายเดือนในหมวดต่าง ๆ ของนักเรียนนักศึกษาระดับมัธยมศึกษา อาชีวศึกษา และอุดมศึกษา แตกต่างกันหรือไม่ กำหนด $\mathbf{H_0}$ ดังนี้

$$\mathbf{H_0} : \ _1 = \ _2 = \ _3$$

การทดสอบ $\mathbf{H_0}$ สามารถใช้ตารางการแจกแจง $\wedge^*$ ที่ $p \geq 1$ และ $g = 3$ (เนื่องจาก $p = 4$) และใช้คอมพิวเตอร์คำนวณ $\wedge^*$ ได้ค่าดังนี้

$$\wedge^* = \frac{|W|}{|B + W|} = .7714$$

และ $$\left(\frac{\sum n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right) = \left(\frac{516 - 4 - 2}{4}\right)\left(\frac{1 - \sqrt{.7714}}{\sqrt{.7714}}\right)$$

$$= 17.67$$

กำหนด $\alpha = .01$ ดังนั้น $F_{2(4),2(510)(.01)} = \chi^2_8(.01)/8 = 2.51$ เนื่องจาก $17.67 > F_{8,1020(.01)}$ จึงปฏิเสธ H_0 ที่ระดับ .01 ดังนั้นค่าเฉลี่ยของค่าใช้จ่ายรายเดือนใน หมวดต่าง ๆ ของนักเรียนนักศึกษาระดับมัธยมศึกษา อาชีวศึกษา และอุดมศึกษา จึงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ถ้าเปรียบเทียบกับการใช้ χ^2 จากสมการ ❹ ทดสอบ ซึ่งกลุ่มตัวอย่างมีขนาดใหญ่ (516) ก็ให้ผลการทดสอบเช่นเดียวกัน

การเปรียบเทียบพหุคูณ

ในการทดสอบสมมติฐานเมื่อ H_0 ถูกปฏิเสธ แสดงว่า ผลของการจัดกระทำอย่างน้อย 1 คู่ที่มีผลแตกต่างกัน จึงต้องเปรียบเทียบว่าผลของการจัดกระทำในตัวแปรใดในกลุ่มต่าง ๆ ที่เปรียบเทียบ g กลุ่ม และเปรียบเทียบเป็นรายคู่จะได้ g(g-1)/2 คู่ ผลของการจัดกระทำที่ k (k = 1, 2, ..., g) โดยประมาณคือ $\hat{}_k = \bar{X}_k - \bar{X}$ และค่าประมาณผลของการจัดกระทำที่ k ตัวแปรที่ i (i = 1, 2, ..., p)คือ $\hat{}_{ki} = \bar{}_{ki} - \bar{}_i$ ค่าประมาณของการจัดกระทำที่ ℓ ตัวแปรที่ i คือ $\hat{\tau}_{\ell i} = \bar{X}_{\ell i} - \bar{X}_i$ และผลต่างของ $\hat{\tau}_{ki}$ กับ $\hat{\tau}_{\ell i}$ จะเป็นผลต่างของค่าเฉลี่ยของตัวอย่างที่เป็นอิสระต่อกันสองกลุ่ม ซึ่งเป็นไปตามลักษณะของเบื้องต้นของการทดสอบที (t – test) โดยมีความแปรปรวนของผลต่างของค่าเฉลี่ยดังนี้

$$\mathrm{Var}(\hat{}_{ki} - \hat{}_{\ell i}) = \mathrm{Var}(\bar{}_{ki} - \bar{}_{\ell i}) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\sigma_{ii}$$

โดยที่ σ_{ii} เป็นสมาชิกในแนวทะแยงของ $\boldsymbol{\Sigma}$ ลำดับที่ i และค่าประมาณของ $\text{Var}(\bar{\ }_{ki} - \bar{\ }_{\ell i})$ หาได้โดยแทน σ_{ii} ด้วย W_{ii} และหารด้วย df ของ W โดยที่ W_{ii} เป็นสมาชิกในแนวทะแยงของ **W** ลำดับที่ i และ $n = n_1 + n_2 + \ldots + n_g$ นั่นคือ

$$\widehat{\text{Var}}(\bar{\ }_{ki} - \bar{\ }_{\ell i}) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\frac{W_{ii}}{n-g}$$

ด้วยเหตุที่ตัวแปรตามมี p ตัว และมีกลุ่มที่เปรียบเทียบกันมีจำนวน g(g - 1)/2 คู่ ในการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มที่กระทำพร้อมกันทุกตัวแปร ค่าวิกฤตในการเปรียบเทียบจึงใช้ค่า $t_{n-g\,(\alpha/2m)}$ เนื่องจากมีกลุ่มที่เปรียบเทียบ m คู่ (m = pg(g – 1)/2)

สมมติฐานในการเปรียบเทียบพหุคูณและการทดสอบค่าสถิติ

เพื่อที่จะเปรียบเทียบว่าค่าเฉลี่ยของตัวแปรใด ๆ ในกลุ่มที่ k และ ℓ ว่าแตกต่างกันหรือไม่ สมมติฐานในการทดสอบจะเป็นดังนี้

$$\mathbf{H_0} : \ _{\mathbf{k}} = \ _{\ell}$$

$_{\mathbf{k}}$ และ $_{\ell}$ หมายถึง เวคเตอร์ค่าเฉลี่ยของผลการจัดกระทำกลุ่มที่ k และ ℓ และให้ $\hat{\tau}_{ki}$ และ $\hat{\tau}_{\ell i}$ เป็นค่าเฉลี่ยของผลการจัดกระทำกลุ่มที่ k และ ℓ ตัวแปรที่ i ซึ่งเป็นส่วนประกอบของ $_{\mathbf{k}}$ และ $_{\ell}$ สมมติฐานในการทดสอบจะเป็นดังนี้

$$H_0 : \tau_{ki} = \tau_{\ell i}$$

การทดสอบค่าสถิติใช้การทดสอบที มีสูตรต่อไปนี้

$$t = \frac{\hat{\ }_{ki} - \hat{\ }_{\ell i}}{\sqrt{\left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\frac{W_{ii}}{n-g}}} \quad ; \ df = n - g \qquad ❺$$

ค่าวิกฤตในการเปรียบเทียบ คือ $t_{n-g(\alpha/2m)}$; m = pg(g – 1)/ 2 ถ้าค่าสถิติที > ค่าวิกฤต สรุปได้ว่า ค่าเฉลี่ยของตัวแปรที่ i ระหว่างกลุ่ม k และ ℓ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ตัวอย่าง 3

จากตัวอย่าง 2 สรุปว่า ค่าใช้จ่ายต่อเดือนของนักเรียนนักศึกษามีความแตกต่างกันในแต่ละระดับ เราสามารถใช้ ❺ เพื่อเปรียบเทียบว่าตัวแปรที่ 4 (X_4 : ค่าใช้จ่ายในการติดต่อสื่อสาร) ระหว่างนักเรียนมัธยมศึกษา นักเรียนอาชีวศึกษาและอุดมศึกษาแตกต่างกันหรือไม่ได้ดังนี้

ข้อมูลจากตัวอย่าง 2

$$\hat{\ }_{1} = (\overline{\mathbf{X}}_{1} - \overline{\mathbf{X}}) = \begin{bmatrix} .070 \\ -.039 \\ -.020 \\ -.020 \end{bmatrix} , \quad \hat{\ }_{2} = \begin{bmatrix} .031 \\ .077 \\ .022 \\ .038 \end{bmatrix} , \quad \hat{\ }_{3} = \begin{bmatrix} .1 \\ .002 \\ .023 \\ .003 \end{bmatrix} ,$$

$$\mathbf{W} = \begin{pmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.581 & 2.428 & .394 & 6.538 \end{pmatrix}$$

$\hat{\ }_{14} - \hat{\ }_{24}$ = -.020 - .038 = -.058

$\hat{\ }_{14} - \hat{\ }_{3}$ = -.020 - .003 = -.023

$\hat{\ }_{24} - \hat{\ }_{3}$ = .038 - .003 = -.035

และ n = 271 + 138 + 107 = 516

ดังนั้น $\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{W_{44}}{n-g}} = \sqrt{\left(\frac{1}{271} + \frac{1}{138}\right)\frac{6.538}{516-3}}$ = .0118

เมื่อ p = 4 และ g = 3 ที่ α = .01, n – g = 513 ค่าวิกฤตของสถิติที่ใช้ทดสอบค่าเฉลี่ยโดยพร้อมกันทุกตัวแปร และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ หาค่าได้จากตารางการแจกแจงแบบปกติ คือ

$$t_{513\,(.01/4(3)2)} = 3.29$$

สมมติฐานที่ทดสอบ คือ $H_0 : {}_{14} = {}_{24}$

ค่าสถิติที่ทดสอบ
$$t = \frac{\hat{\ }_{14} - \hat{\ }_{24}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{W_{44}}{n-g}}}$$
$$= \frac{-.058}{.0118}$$
$$= -4.9152$$

ซึ่งค่า $|t| > t_{513\ (.01/4\ (3)\ 2)}$ จึงสรุปได้ว่า ค่าใช้จ่ายเฉลี่ยต่อเดือนเกี่ยวกับการติดต่อสื่อสารของนักเรียนมัธยมศึกษาและนักเรียนอาชีวศึกษาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ค่าสถิติที่ทดสอบ $H_0 : {}_{14} = {}_{3}$ มีค่า -1.7829 และ

ค่าสถิติที่ทดสอบ $H_0 : {}_{24} = {}_{3}$ มีค่า 2.4138

ค่าสถิติที่ทดสอบทั้งสองค่า $< t_{513\,(.01/4(3)2)}$

จึงสรุปได้ว่า ค่าใช้จ่ายต่อเดือนเกี่ยวกับการติดต่อสื่อสารของนักเรียนมัธยมศึกษาและนักศึกษาอุดมศึกษาไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทำนองเดียวกันค่าใช้จ่ายดังกล่าวของนักเรียนอาชีวศึกษาและนักศึกษาอุดมศึกษาก็ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

บทสรุป

การวิเคราะห์ความแปรปรวนตัวแปรพหุ ใช้สำหรับการวิเคราะห์เพื่อตรวจสอบความแตกต่างของเวคเตอร์ค่าเฉลี่ยของประชากรที่มีมากกว่า 2 กลุ่ม และตัวแปรตามที่นำมาวิเคราะห์มีความสัมพันธ์กัน ในการวิเคราะห์เมื่อพบว่าเวคเตอร์ค่าเฉลี่ยของประชากรแตกต่างกัน ก็จะทำการตรวจสอบความแตกต่างเป็นรายตัวแปร โดยมีหลักที่ว่าการเปรียบเทียบรายตัวแปรอยู่ภายใต้การกระทำโดยพร้อมกัน อย่างไรก็ตามการใช้ MANOVA วิเคราะห์ข้อมูลควรมีการตรวจสอบข้อตกลงเบื้องต้น เช่น ความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของประชากร และมีหลักฐานหรือเหตุผลที่แสดงว่าตัวแปรตามมีความสัมพันธ์กัน แต่หากตัวแปรตามที่นำมาวิเคราะห์ไม่มีความสัมพันธ์ก็อาจใช้ ANOVA วิเคราะห์จะสะดวกกว่า ถ้าข้อมูลเป็นไปตามข้อตกลงเบื้องต้น และเมื่อมีตัวแปรตามเกิน 3 ตัว การตรวจสอบข้อตกลงเบื้องต้นของข้อมูลก็ดี การตรวจสอบความสัมพันธ์ระหว่างตัวแปรตามก็ดี ตลอดจนการหาเมทริกซ์ความแปรปรวนร่วมและการหาค่าดีเทอร์มิเนนท์ของเมทริกซ์ จะเป็นข้อจำกัดในการวิเคราะห์ด้วยเครื่องคำนวณธรรมดา ซึ่งสามารถแก้ไขได้โดยการใช้คอมพิวเตอร์วิเคราะห์ด้วยโปรแกรมสำเร็จรูป

บรรณานุกรม

ฉัตรศิริ ปิยะพิมลสิทธิ์. (2547). ***การใช้โปรแกรมคอมพิวเตอร์วิเคราะห์ความแปรปรวนพหุและการจำแนกประเภท***. (Online) Available : http://www.watpon.com. เข้าถึงเมื่อ มีนาคม 2547.

Anderson, T.W. (1974). ***An Introduction to Multivariate Statistical Analysis***. New Delhi : Wiley Eastern.

Johnson, R. A. and D. W. Wichern. (1986). ***Applied Multivariate Statistical Analysis***. (2^{nd} ed.), New Jersey : Prentice – Hall.

Munro,B. H. (2001). ***Statistical Methods for Health Care Research***. (4^{th} ed.), Philadelphia : Lippincott Williams & Wilkins.

Searle, S. R. (1982). ***Matrix Algebra useful for statistics***. New York : John Wiley and Sons.