บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

บทนี้จะกล่าวถึงทฤษฎีและหลักสถิติต่าง ๆ ที่เกี่ยวข้องกับการวิจัย ได้แก่ การแจกแจงแบบปกติของตัวแปรพหุ (multivariate normal distribution) การทดสอบแบบขั้นลดลง (Step - down Procedure) วิธีการแบบปิด (Closure Method) และ ทฤษฎี Hotelling 's T^2 รวมทั้งตัวสถิติต่าง ๆ ที่จะศึกษาในการวิจัยครั้งนี้ ซึ่งมีรายละเอียดดังนี้

การแจกแจงแบบปกติของตัวแปรพหุ

ให้ $\underset{\sim}{y}' = (y_1, y_2, \ldots, y_n)$ เป็นเวกเตอร์ที่มีการแจกแจงปกติของ n ตัวแปรสุ่ม ซึ่งสามารถเขียนได้ว่า $\underset{\sim}{y} \sim N_n(\underset{\sim}{\mu}, V)$ และฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(y_1, y_2, \ldots, y_n) = \frac{\exp\left[\frac{-1}{2}(\underset{\sim}{y} - \underset{\sim}{\mu})' V^{-1} (\underset{\sim}{y} - \underset{\sim}{\mu})\right]}{(2\pi)^{\frac{1}{2}n} |V|^{\frac{1}{2}}}$$

จะได้ว่าเวกเตอร์สุ่ม $\underset{\sim}{y}$ มีการแจกแจงปกติของ n ตัวแปรสุ่มที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ

$\underset{\sim}{\mu}' = (\mu_1, \mu_2, \ldots, \mu_K)$

และเมทริกซ์ความแปรปรวนร่วมอยู่ในรูปของ

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{bmatrix}$$

การทดสอบแบบขั้นลดลง

เป็นการทดสอบที่เริ่มจากการทดสอบสมมุติฐานร่วมทั้งหมดซึ่งมีมิติสูงสุดในขั้นที่ 1 แล้วจะทำการทดสอบต่อในขั้นต่อไปถ้ามีการปฏิเสธสมมุติฐาน ซึ่งมิติของสมมุติฐานจะลดลงขั้นละ 1 มิติ ถ้ายอมรับสมมุติฐานใด เราจะยอมรับสมมุติฐานที่มีมิติน้อยกว่าซึ่งเป็นเซตย่อยของสมมุติฐานนั้นทั้งหมด นั่นคือเราจะทดสอบสมมุติฐานก็ต่อเมื่อเราปฏิเสธสมมุติฐานที่มีมิติสูงกว่าซึ่งมีสมมุติฐานนั้นร่วมอยู่ด้วยทั้งหมด

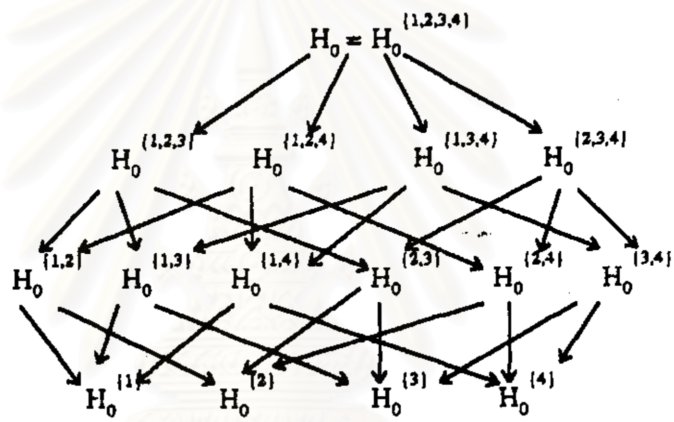
ตัวอย่างการทดสอบแบบขั้นลดลง (K = 4)

ขั้นที่ 1

ขั้นที่ 2

ขั้นที่ 3

ขั้นที่ 4



วิธีการแบบปิด

เป็นวิธีการทั่วไปที่ใช้ในการสร้างการเปรียบเทียบพหุคูณแบบขั้นลดลง ซึ่งเสนอโดย มาร์คัส เพอริทส์และแกบริล ในปี ค.ศ. 1976 วิธีที่ได้จากการประยุกต์วิธีการแบบปิดนี้เรียกว่า วิธีการทดสอบแบบปิด (Closed Testing Procedure) ซึ่งมีนิยามดังที่กล่าวแล้วในบทที่ 1 หัวข้อที่มาและความสำคัญของปัญหา ทฤษฎีบทต่อไปนี้จะแสดงให้เห็นว่า วิธีการนี้จะควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ของการทดลอง (Type I Error FWE) อย่างเข้มงวด

ทฤษฎีบท (Marcus et. al. , 1976) วิธีการทดสอบแบบปิดจะควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ของการทดลองอย่างเข้มงวด

พิสูจน์

ให้ $\{ H_i , i \in P \}$ เป็นกลุ่มของสมมุติฐานว่างที่เป็นจริง และ ให้ $H_P = \bigcap_{i \in P} H_i$ เมื่อ P เป็นเซตย่อยที่ไม่ทราบค่าของ { 1,2,...,K } ถ้า P เป็นเซตว่างเราไม่สามารถหาความคลาดเคลื่อนประเภทที่ 1 ได้ ดังนั้นจะให้ P เป็นเซตที่ไม่ว่าง ให้ A เป็นเหตุการณ์ที่ปฏิเสธสมมุติฐาน H_i อย่างน้อย 1 สมมุติฐาน และ B เป็นเหตุการณ์ที่ปฏิเสธ H_P วิธีการทดสอบแบบปิดจะปฏิเสธสมมุติฐาน H_i ก็ต่อเมื่อ เราปฏิเสธสมมุติฐานที่สร้างจาก H_i โดยเฉพาะ H_P ที่ระดับนัยสำคัญ α และการทดสอบ H_i ก็ปฏิเสธที่ระดับนัยสำคัญ α ด้วย ดังนั้น $A = A \cap B$ และภายใต้ H_P จะได้ว่า

$$FWE = Pr\{A\} = Pr\{A \cap B\} = Pr\{B\}\ Pr\{A | B\} \le \alpha$$

เนื่องจาก $Pr\{B\} = \alpha$ เมื่อ H_P เป็นจริง

ทฤษฎีบท Hotelling 's T^2

ถ้า $\underset{\sim}{y} \sim N_p\left(\underset{\sim}{0}, \Sigma\right)$ และ S เป็นตัวประมาณที่ไม่เอนเอียงของ Σ ที่มีระดับขั้นความเสรีเท่ากับ m แล้ว ตัวสถิติ Hotelling 's T^2 คือ

$$T^2 = \underset{\sim}{y}' S^{-1} \underset{\sim}{y} = \frac{mp}{m-p+1} F(p, m-p+1)$$

เมื่อ p = จำนวนสมาชิกของเวกเตอร์ $\underset{\sim}{y}$

การทดสอบสมมุติฐานโดยใช้ตัวสถิติ Hotelling 's T^2

<u>ข้อตกลงเบื้องต้น</u> คือ ตัวอย่างสุ่ม 2 ชุดซึ่งสุ่มอย่างเป็นอิสระกัน (ขนาด n_i , i = 1,2) เป็นข้อมูลที่ได้จากตัวแปร p ตัว โดยที่แต่ละชุดมีการแจกแจงปกติพหุที่มีเวกเตอร์ของค่าเฉลี่ย คือ $\underset{\sim}{\mu}_i$ (i = 1,2) และเมทริกซ์ความแปรปรวนร่วมของทั้ง 2 ชุดเท่ากันแต่ไม่ทราบค่า คือ Σ ซึ่งมีค่าลำดับชั้น (rank) เท่ากับ p ต้องการทดสอบสมมุติฐาน

$$H_0 : \underset{\sim 1}{\mu} - \underset{\sim 2}{\mu} = \underset{\sim}{0}$$
$$H_1 : \underset{\sim 1}{\mu} - \underset{\sim 2}{\mu} \neq \underset{\sim}{0}$$

เนื่องจาก $\underset{\sim ij}{y} \sim N_p\left(\underset{\sim i}{\mu}, \Sigma\right)$ $\quad i = 1,2 \; , \; j = 1,2,\ldots,n_i$

จะได้ว่า $\underset{\sim i}{\bar{y}} \sim N_p\left(\underset{\sim i}{\mu}, \frac{\Sigma}{n_i}\right)$ และ $\underset{\sim 1}{\bar{y}} - \underset{\sim 2}{\bar{y}} \sim N_p\left(\underset{\sim}{0}, \left(\frac{n_1 + n_2}{n_1 n_2}\right)\Sigma\right)$

จากทฤษฎีบท Hotelling's T^2 จะได้

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\underset{\sim 1}{\bar{y}} - \underset{\sim 2}{\bar{y}})' S^{-1}(\underset{\sim 1}{\bar{y}} - \underset{\sim 2}{\bar{y}})$$

เมื่อ S = ค่าประมาณแบบรวมของเมทริกซ์ความแปรปรวนร่วม Σ

(pooled estimate of common covariance matrix)

ซึ่งมีระดับขั้นความเสรีเท่ากับ $n_1 + n_2 - 2$

ดังนั้น

$$T^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F(p, n_1 + n_2 - p - 1)$$

เราจะยอมรับ H_0 ณ ระดับนัยสำคัญ α ถ้า $T^2 \leq \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{\alpha;p,n_1+n_2-p-1}$

ตัวสถิติที่ใช้ในการวิจัย

1. วิธีการวิเคราะห์ทีละตัวแปร

1.1 **วิธีบอนเฟอร์โรนี - โฮล์ม** เสนอโดยโฮล์ม เป็นเวอร์ชันที่ปฏิเสธตามลำดับ (sequentially rejective version) ของวิธีบอนเฟอร์โรนี มีขั้นตอนดังนี้

ขั้นที่ 1 คำนวณค่า $P_1, P_2, \ldots, P_K$ ซึ่งได้จากการทดสอบสมมุติฐานแบบข้างเดียว $H_0^{1}, H_0^{2}, \ldots, H_0^{K}$ โดยใช้ค่าสถิติจากตัวอย่าง $t_1, t_2, \ldots, t_K$ ตามลำดับ ซึ่ง t_k มีสูตรดังนี้

$$t_k = \frac{\bar{y}_{1.k} - \bar{y}_{2.k}}{\sqrt{s_p^2\left(\frac{n_1+n_2}{n_1 n_2}\right)}}$$

เมื่อ $$s_p^2 = \frac{(n_1-1)s_{1.k}^2 + (n_2-1)s_{2.k}^2}{n_1+n_2-2}$$

$$s_{i.k}^2 = \sum_{j=1}^{n_i} \frac{\left(y_{ijk} - \bar{y}_{i.k}\right)^2}{n_i - 1}$$

และ $k = 1, 2, \ldots, K$

ขั้นที่ 2 เรียงลำดับค่า P ที่ได้จากขั้นที่ 1 ดังนี้

$P_{[1]} \le P_{[2]} \le \ldots \le P_{[K]}$ สอดคล้องกับ $H_0^{[1]}, H_0^{[2]}, \ldots, H_0^{[K]}$

ขั้นที่ 3 ให้ $k = 1$

ขั้นที่ 4 ถ้า $P_{[k]} > \alpha \,/\, K - k + 1$ แล้วจะยอมรับสมมุติฐานที่เหลือทั้งหมดโดยที่ไม่ต้องทำการทดสอบ นั่นคือ ยอมรับสมมุติฐาน $H_0^{[k]}, H_0^{[k+1]}, \ldots, H_0^{[K]}$ แล้วหยุดกระบวนการทดสอบ

ขั้นที่ 5 ถ้า $P_{[k]} \le \alpha \,/\, K - k + 1$ แล้วจะปฏิเสธสมมุติฐาน $H_0^{[k]}$ จากนั้นให้ $k = k + 1$ แล้วย้อนกลับไปขั้นที่ 4

1.2 **วิธีเจมส์ - โฮล์ม** เป็นวิธีที่ปรับปรุงจากวิธีที่เสนอโดยเจมส์ ในปี ค.ศ. 1991 ซึ่งวิธีการดั้งเดิมของเจมส์นั้น มีการปรับค่า P สำหรับกรณีที่ตัวแปรมีสหสัมพันธ์กัน โดยตั้งข้อสมมติเบื้องต้นว่าสหสัมพันธ์ระหว่างตัวแปรมีค่าเท่ากัน คือ ρ และเมื่อประยุกต์วิธีการแบบปิดกับวิธีนี้ในลักษณะเดียวกับวิธีบอนเฟอร์โรนี - โฮล์ม จึงเรียกวิธีนี้ว่า วิธีเจมส์ - โฮล์ม ซึ่งมีวิธีการดังนี้

$$P_{[k]}^{adj.} = 1 - D_1(1-\rho^2) - D_2\rho^2 - D_3\rho(1-\rho) - D_4(2 - 2(1-\rho)^{1/2} - \rho - \rho^2)$$

เมื่อ $D_1 = (1 - P_{[k]})^{K-k+1}$

$D_2 = 1 - P_{[k]}$

$D_3 = [(K-k+1)/2](K-k)(1 - P_{[k]})^{K-k-1}\phi(a)^2$

$D_4 = [(K-k+1)/2](K-k)\phi(a)G(k)$

$a = \Phi^{-1}(1 - P_{[k]})$

ϕ คือ ฟังก์ชันความหนาแน่นของการแจกแจงปกติ

Φ คือ ฟังก์ชันการแจกแจงสะสมของการแจกแจงปกติ

และ G(i) เป็นค่าที่ได้จากตารางที่ 1 ในบทความของ เจมส์

เราจะปฏิเสธ $H_0^{[k]}$ ก็ต่อเมื่อ $P_{[k']}^{adj} \leq \alpha$ ทุก $k' \leq k$ นอกจากนี้ เจมส์ ยังได้เสนอวิธีการที่ง่ายในการปรับค่านี้ กรณีที่สหสัมพันธ์ไม่เท่ากัน คือ การใช้ค่าเฉลี่ยของสหสัมพันธ์ $\bar{\rho}$ แทน ρ ซึ่งคำนวณได้ดังนี้

$$\bar{\rho} = \frac{1}{m}\sum_{\tilde{k}=1}^{K-1}\sum_{\tilde{k}'=\tilde{k}+1}^{K} \rho_{(\tilde{k}\tilde{k}')}$$

เมื่อ $\rho_{(\tilde{k}\tilde{k}')}$ คือ สหสัมพันธ์ระหว่างตัวแปรที่ให้ค่า P เท่ากับ $P_{(\tilde{k})}$ และ $P_{(\tilde{k}')}$

และ $m = (K)(K-1)/2$

2. วิธีการวิเคราะห์ทุกตัวแปรพร้อมกัน

วิธีการมาตรฐานในกลุ่มนี้คือ วิธีการทดสอบ T^2 ของ Hotelling ซึ่งเป็นวิธีการทดสอบแบบอัตราส่วนภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Ratio Test) แต่ในปี ค.ศ. 1984 โอบรีนพบว่า วิธีนี้มีอำนาจการทดสอบน้อยมากในการวิเคราะห์ทางการแพทย์ ซึ่งมักจะคาดได้ล่วงหน้าว่าความแตกต่างของทุกตัวแปรน่าจะมีทิศทางเดียวกัน และได้เสนอวิธีการทดสอบตัวแปรพหุที่เหมาะสมกับการวิจัยทางการแพทย์มากกว่า และมีอำนาจของการทดสอบมากกว่า Hotelling 's T^2 คือ วิธีทดสอบแบบปิด OLS และ GLS

2.1 วิธีการทดสอบแบบปิด OLS ตัวสถิติที่ใช้ในวิธีนี้ คือ T_{OLS} ได้จากผลรวมของตัวสถิติ t สำหรับตัวแปรเดียวและหาได้ดังนี้

$$T_{OLS} = \frac{J'\underset{\sim}{t}}{(J'\hat{P}J)^{1/2}}$$

เมื่อ $J = (1,1,\ldots,1)'$

$\underset{\sim}{t}$ = เวกเตอร์ของ $t_k = \frac{\hat{\Delta}_k}{\hat{\sigma}_k}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$

โดยที่ $\hat{\Delta}_k = \bar{y}_{1.k} - \bar{y}_{2.k}$

และ $\hat{\sigma}_k = \sqrt{\frac{1}{n_1 + n_2 - 2}\left[(n_1 - 1)s_{1.k}^2 + (n_2 - 1)s_{2.k}^2\right]}$

$\hat{P} = [\hat{\rho}_{kk'}]$ $k, k' = 1,2,\ldots,K$ โดยที่ $\hat{\rho}_{kk'} = \frac{\hat{\sigma}_{kk'}}{\hat{\sigma}_k \hat{\sigma}_{k'}}$

$\hat{\sigma}_{kk'}$ คือ ความแปรปรวนร่วมตัวอย่างแบบรวม (pool) ถ้า $k \neq k'$

และ $\hat{\sigma}_k$ คือ ส่วนเบี่ยงเบนมาตรฐานตัวอย่างของ k

2.2 **วิธีการทดสอบแบบปิด GLS** ตัวสถิติที่ใช้ในวิธีนี้ คือ T_{GLS} ได้จากผลรวมเชิงเส้นแบบถ่วงน้ำหนักของตัวสถิติ t สำหรับตัวแปรเดียว หาได้ดังนี้

$$T_{GLS} = \frac{J'\hat{P}^{-1}\underset{\sim}{t}}{(J'\hat{P}^{-1}J)^{1/2}}$$

จาก T_{GLS} จะเห็นได้ว่าค่าถ่วงน้ำหนักคือ $J'\hat{P}^{-1}$ ซึ่งถ้าตัวแปรมีความสัมพันธ์กับตัวแปรอื่น ๆ น้อยกว่าก็จะมีค่าถ่วงน้ำหนักมากกว่า

ทั้ง T_{OLS} และ T_{GLS} มีการแจกแจงปกติมาตรฐานภายใต้ H_0 ดังนั้นเราจะปฏิเสธ H_0 ที่ระดับนัยสำคัญ ถ้าค่าสถิติที่ได้จากตัวอย่างมีค่ามากกว่าค่าควอนไทล์ที่ α ของการแจกแจงปกติมาตรฐาน

3. วิธีการวิเคราะห์แบบบูตสแตรป

3.1 วิธีเวสต์ฟอลและยัง

ในปี ค.ศ. 1989 เวสต์ฟอลและยังได้เสนอวิธีการทดสอบโดยใช้เทคนิคบูตสแตรปในการปรับค่า P สำหรับข้อมูลที่มีหลายตัวแปร ซึ่งเป็นวิธีที่มีประสิทธิภาพในการวิเคราะห์วิจัยทางการแพทย์ และเสนอวิธีที่ขยายจากวิธีการดั้งเดิมในปี ค.ศ.1993 ซึ่งมีขั้นตอนดังนี้

1. สุ่มตัวอย่างขนาด n_1 และ n_2 (แบบแทนที่) สำหรับกลุ่มวิธีการปฏิบัติที่ 1 และ 2 ตามลำดับจากข้อมูลดั้งเดิมที่รวมตัวอย่างของทั้ง 2 กลุ่มวิธีการปฏิบัติเข้าไว้ด้วยกัน โดยถือว่า H_0 เป็นจริง นั่นคือไม่มีความแตกต่างระหว่างกลุ่มวิธีการปฏิบัติ

2. คำนวณค่าบูตสแตรป P ได้แก่ $P_{[1]}^*$, $P_{[2]}^*$, ... , $P_{[K]}^*$ ด้วยวิธีการเดียวกับการคำนวณค่า P ที่ได้จากข้อมูลดั้งเดิม แต่ค่าบูตสแตรป P ที่คำนวณได้นี้ไม่จำเป็นจะต้องเรียงลำดับเหมือนกับค่า P ดั้งเดิม ทำเช่นเดียวกันนี้ซ้ำ ๆ กัน N ครั้ง ซึ่งจะได้เวกเตอร์ขนาด K ของค่าบูตสแตรป P ทั้งหมด N เวกเตอร์

3. เราสามารถประมาณค่า P ซึ่งปรับค่าแล้วสำหรับตัวแปรที่สอดคล้องกับสมมุติฐาน $H_0^{[1]}$ (ซึ่งคือตัวแปรที่ให้ค่า P น้อยที่สุดจากข้อมูลดั้งเดิม) ได้จากสัดส่วนของเหตุการณ์ " $\min\limits_{1\le \tilde{k}\le K} P_{[\tilde{k}]}^* \le P_{[1]}$ " แทนด้วย $\tilde{P}_{[1]}^{adj}$ สำหรับตัวแปรที่สอดคล้องกับ $H_0^{[2]}$ เราสามารถประมาณค่าบูตสแตรปที่ปรับแล้ว $\tilde{P}_{[2]}^{adj}$ ได้จากสัดส่วนของเหตุการณ์ " $\min\limits_{2\le \tilde{k}\le K} P_{[\tilde{k}]}^* \le P_{[2]}$ " และสำหรับตัวแปรที่สอดคล้องกับ $H_0^{[k]}$ เราสามารถประมาณค่าบูตสแตรปที่ปรับแล้ว $\tilde{P}_{[k]}^{adj}$ ได้จากสัดส่วนของเหตุการณ์ " $\min\limits_{k\le \tilde{k}\le K} P_{[\tilde{k}]}^* \le P_{[k]}$ "

4. หลังจากคำนวณค่า P ที่ปรับค่าโดยใช้เทคนิคบูตสแตรป $\tilde{P}_{[1]}^{adj}, \tilde{P}_{[2]}^{adj}, \ldots, \tilde{P}_{[K]}^{adj}$ (ซึ่งไม่จำเป็นต้องเรียงตามลำดับ) แล้ว เราจะปฏิเสธ H_0^{k} ก็ต่อเมื่อ $\tilde{P}_{[\tilde{k}]}^{adj} \le \alpha$ สำหรับทุก $\tilde{k} \le k$

วิธีการดั้งเดิมของเวสต์ฟอลและยังซึ่งเสนอในปี ค.ศ. 1989 นั้น ค่า P ที่ปรับค่าโดยใช้เทคนิคบูตสแตรปประมาณได้จากสัดส่วนของเหตุการณ์ " $\min\limits_{1\le \tilde{k}\le K} P_{[\tilde{k}]}^* \le P_{[k]}$ " , $k = 1,2,\ldots,K$