บทที่ 2

สถิติที่ใช้ในการวิจัย

ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับชนิดของข้อมูลถูกตัดทิ้ง (type of censored data) ลักษณะข้อมูลจริงที่มีการแจกแจงปกติทวิ (Bivariate Normal Distribution) และการแจกแจงแกมมาทวิ (Bivariate Gamma Distribution) รวมทั้งตัวสถิติที่ใช้ทดสอบค่าสัมประสิทธิ์สหสัมพันธ์พร้อมทั้งตัวอย่างการใช้ตัวสถิติซึ่งมีรายละเอียดต่างๆ ดังนี้

ชนิดของข้อมูลที่มีค่าถูกตัดทิ้ง (Type of Data Censoring)

1. ข้อมูลถูกตัดทิ้งประเภทที่ 1 (Type I Censoring)

ข้อมูลถูกตัดทิ้งประเภทนี้จะกำหนดระยะเวลาของการเกิดข้อมูลถูกตัดทิ้งไว้ล่วงหน้า เช่น การศึกษาอายุการใช้งานของเครื่องจักรจำนวนหนึ่งซึ่งมีกำหนดอายุการใช้งานของเครื่องจักรคือ 50 ชั่วโมง ถ้าเครื่องจักรทำงานเกิน 50 ชั่วโมงถือว่า อายุการใช้งานของเครื่องจักรที่เกินเป็นข้อมูลถูกตัดทิ้ง เพราะเราไม่สามารถบันทึกข้อมูลได้

2. ข้อมูลถูกตัดทิ้งประเภทที่ 2 (Type II Censoring)

ในบางครั้งผู้วิจัยไม่สามารถจะกำหนดระยะเวลาการเกิดข้อมูลถูกตัดทิ้งที่เหมาะสมได้จึงกำหนดจำนวนข้อมูลที่ถูกตัดทิ้งไว้ล่วงหน้าและจะหยุดทำการทดลอง เมื่อข้อมูลที่ถูกตัดทิ้งเกิดขึ้นครบตามจำนวนที่กำหนดแทนไว้ เช่น การศึกษาความสัมพันธ์ระหว่างโรคเส้นโลหิตตีบ(Arteriosclerosis)กับระยะเวลาที่มีชีวิตอยู่ของสัตว์ชนิดหนึ่ง ผู้ทดลองได้ทำการตรวจสอบซากศพหลังจากสัตว์ตายแล้ว แต่เนื่องจากต้องใช้เวลานานเพื่อรอให้หน่วยทดลองตาย ผู้ทดลองจึงได้กำหนดจำนวนข้อมูลที่ถูกตัดทิ้งไว้ล่วงหน้า

3. ข้อมูลที่ถูกตัดทิ้งแบบสุ่ม (Random Censoring)

ข้อมูลที่ถูกตัดทิ้งชนิดนี้ส่วนใหญ่จะพบในข้อมูลทางการแพทย์ เช่น คนไข้หลังจากได้รับการรักษาแล้วเราไม่สามารถเก็บข้อมูลของคนไข้บางรายได้ครบตามจำนวนที่กำหนดไว้

ในการวิจัยครั้งนี้ผู้วิจัยจะทำการศึกษาทั้งในกรณีที่ข้อมูลสมบูรณ์และข้อมูลที่มีค่าถูกตัดทิ้งทางขวา ข้อมูลที่มีค่าถูกตัดทิ้งทางขวาเป็นกรณีเฉพาะของการมีข้อมูลถูกตัดทิ้งในลักษณะดังกล่างข้างต้น การทดลองที่ทำให้ข้อมูลมีค่าถูกตัดทิ้งทางขวาได้แก่ การทดลองเกี่ยวกับอายุการใช้งานของเครื่องจักรที่ใช้ในการผลิตในโรงงานอุตสาหกรรม แล้วบันทึกระยะเวลาหรือจำนวนชั่วโมงที่เครื่องจักรนั้นจะเสียหรือทำงานไม่ได้ ในระหว่างที่ดำเนินการทดลองอยู่เครื่องจักรที่เสียหรือทำงานไม่ได้นั้นจะเป็นข้อมูลไม่ถูกตัดทิ้ง (uncensored data) เมื่อสิ้นสุดการทดลองเครื่องจักรที่ยังคงอยู่ในสภาพที่ใช้งานได้ดีจะเป็นเครื่องจักรที่ไม่ทราบอายุการใช้งานที่แน่นอน ข้อมูลนี้จะเป็นข้อมูลที่มีค่าถูกตัดทิ้งทางขวา (right censored data)

ชนิดของการแจกแจงข้อมูล

1. การแจกแจงสองตัวแปรซึ่งมีความสัมพันธ์กันเชิงเส้นและมีการแจกแจงปกติทวิ

ถ้า $(X,Y)'$ เป็นเวกเตอร์สุ่มที่มีการแจกแจงปกติทวิ ฟังก์ชันความน่าจะเป็นร่วมของตัวแปรทั้งสองจะอยู่ในรูปของ

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{(1-\rho^2)}}\exp\left[\frac{1}{-2(1-\rho^2)}\left\{(x-\mu_X)^2 - 2\rho(x-\mu_X)(y-\mu_Y) + (y-\mu_Y)^2\right\}\right] \quad (2.1)$$

เราเรียก $(X,Y)'$ ว่าเป็นเวกเตอร์สุ่มที่มีการแจกแจงปกติทวิที่มีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{\mu} = \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}$

และมีเมตริกซ์ความแปรปรวนร่วม $\Sigma = \begin{bmatrix}\sigma_X^{\,2} & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^{\,2}\end{bmatrix}$

เมื่อ ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสุ่ม X และ Y

μ_X คือ ค่าเฉลี่ยของตัวแปรสุ่ม X

μ_Y คือ ค่าเฉลี่ยของตัวแปรสุ่ม Y

$\sigma_X^{\,2}$ คือ ค่าความแปรปรวนของตัวแปรสุ่ม X

$\sigma_Y^{\,2}$ คือ ค่าความแปรปรวนของตัวแปรสุ่ม Y

2. การแจกแจงสองตัวแปรซึ่งมีความสัมพันธ์กันเชิงเส้นและมีการแจกแจงแกมมาทวิ

ถ้า $(X_1, X_2)'$ เป็นเวกเตอร์สุ่มที่มีการแจกแจงแกมมาทวิ ฟังก์ชันความน่าจะเป็นร่วมของตัวแปรทั้งสองจะแบ่งออกเป็น 2 กรณี

<u>กรณีที่ 1</u> $\alpha_1 = \alpha_2$ และ $\beta_1 = \beta_2 = 1$

$$\overset{*}{f}(x_1, x_2) = \prod_{j=1}^{2}\left[\left(\frac{1}{\Gamma(\alpha_j)}\right) x_j^{\alpha-1} e^{-x_j}\right]\left[1 + \sum_{j=1}^{\infty} \rho^{j} L_j^{\alpha-1}(x_1) L_j^{\alpha-1}(x_2)\right]$$

$$; \ \alpha > 0, 0 \leq \rho < 1, x_1 \text{ และ } x_2 > 0$$

<u>กรณีที่ 2</u> $\alpha_1 > \alpha_2$ และ $\beta_1 = \beta_2 = 1$

$$\overset{**}{f}(x_1, x_2) = \prod_{j=1}^{2}\left[\left(\frac{1}{\Gamma(\alpha_j)}\right) x_j^{\alpha_j-1} e^{-x_j}\right]\left[1 + \sum_{j=1}^{2} a_j L_j^{\alpha_1-1}(x_1) L_j^{\alpha_2-1}(x_2)\right]$$

$$; \ x_1 \text{ และ } x_2 > 0$$

เมื่อ α คือ พารามิเตอร์แสดงสเกล (scale parametor)

β คือ พารามิเตอร์แสดงรูปร่าง (shape parametor)

$L_j^{\alpha-1}(x)$ คือ ลาแกรพหุนาม (laguerre polynomial)

a_j คือ สหสัมพันธ์ระหว่าง x_1 กับ x_2

ตัวสถิติที่ใช้ในการวิจัย

พื้นฐานของการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของตัวสถิติที่ใช้ในการวิจัยครั้งนี้จะเป็นตัวประมาณภาวะน่าจะเป็นสูงสุดแก้ไข (modified maximum likelihood estimator

---

* มีชื่อว่า a symmetric gamma distribution สร้างโดย Sarmanov

** มีชื่อว่า asymmetric bivariate gamma distribution สร้างโดย Sarmanov

: MML) จะคล้ายคลึงกับตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimator : ML) การประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของตัวสถิตินั้น ในกรณีของตัวประมาณ ML ข้อมูลจะต้องมีการแจกแจงปกติเท่านั้น แต่ถ้าเป็นตัวประมาณ MML ข้อมูลที่ใช้จะเบี่ยงเบนจากการแจกแจงปกติได้เล็กน้อยซึ่งทำให้สะดวกในทางปฏิบัติ

1. ตัวสถิติทดสอบ Z_f (Fisher Statistics)

ในปี ค.ศ. 1921 Fisher เสนอตัวสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ด้วยตัวสถิติทดสอบ Z_f ซึ่งเราจะพิจารณาตัวสถิติทดสอบ Z_f จากค่าอัตราส่วนของการแปลงข้อมูลเดิม r โดยใช้ ln กับค่ารากที่สองของความแปรปรวน

ให้ $X_{(a)},..X_{(i)}..,X_{(b)}$ เป็นตัวสถิติอันดับของ X , $i = (a = r_1 + 1,..., b = n - r_2)$
และให้ $Y_{(a)},..Y_{(i)}..,Y_{(b)}$ เป็นตัวสถิติอันดับของ Y

โดยค่าตัวแปร X จะสัมพันธ์กับค่าตัวแปร Y ที่เรียงลำดับ (concomitant order statistics)
ซึ่ง X_{r+1} = ค่าสังเกตที่เล็กที่สุด , X_{n-r} = ค่าสังเกตที่ใหญ่ที่สุด โดยที่ r_1 = จำนวนข้อมูลขาดหายทางซ้าย และ r_2 = จำนวนข้อมูลขาดหายทางขวา

ตัวประมาณ MML ของ ρ จะอยู่ในรูปของ

$$\hat{\rho} = \frac{s_{12}\,\hat{\sigma}_2}{s_2^{\,2}\,\hat{\sigma}_1}$$

เมื่อ $s_1^{\,2}$ = ความแปรปรวนของตัวแปร X

$s_2^{\,2}$ = ความแปรปรวนตัวอย่างของตัวแปร Y

$$s_{12} = \frac{\sum_{i=a}^{b}[x_i - \bar{x}][y_i - \bar{y}]}{(A-1)} \qquad (2.2)$$

$$\hat{\sigma}_1 = \left[s_1^{\,2} + \frac{s_{12}^{\,2}}{s_2^{\,2}}\left[\frac{\hat{\sigma}_2^{\,2}}{s_2^{\,2}} - 1\right]\right]^{1/2} \qquad (2.3)$$

(2.4) $$\hat{\sigma}_2 = \frac{\left[B + \sqrt{B^2 + 4AC}\right]}{\left[2\sqrt{A(A-1)}\right]}$$

และค่า A , B และ C สามารถคำนวณหาได้ดังนี้

กำหนดให้ q_1 คือ อัตราส่วนของข้อมูลที่ถูกตัดทิ้งทางซ้าย

q_2 คือ อัตราส่วนของข้อมูลที่ถูกตัดทิ้งทางขวา

$f(t_i)$ คือ ฟังก์ชันความหนาแน่นของ t_i , $i = 1,2$ ซึ่ง t_1 และ $t_2 \sim N(0,1)$

และ $F(t_i)$ คือ ฟังก์ชันการแจกแจงสะสมของ t_i

ซึ่งค่า $F(t_1)$ และ $F(t_2)$ ในที่นี้เราสามารถหาค่าได้จาก

(2.5) $$F(t_1) = q_1$$

(2.6) $$F(t_2) = 1 - q_2$$

จากสมการที่ (2.5) และ (2.6) เราจะหาค่าของ t_1 และ t_2 ภายใต้วิธีการอินทิเกรตของซิมป์ซัน (Simpson's integration method) แล้วหาค่า $f(t_1)$ และ $f(t_2)$ ไปแทนค่าลงในสมการที่ (2.7) (2.8) (2.9) และ (2.10)

(2.7) $$\beta_1 = -f(t_1)\frac{\left[t_1 + \frac{f(t_1)}{q_1}\right]}{q_1}$$

(2.8) $$\beta_2 = -f(t_2)\frac{\left[t_2 - \frac{f(t_2)}{q_2}\right]}{q_2}$$

(2.9) $$\alpha_1 = \frac{f(t_1)}{q_1} - \beta_1 t_1$$

(2.10) $$\alpha_2 = \frac{f(t_2)}{q_2} - \beta_2 t_2$$

เรานำค่า β_1 , β_2 , α_1 และ α_2 แทนค่าลงในสมการที่ (2.11) และ (2.12)

(2.11) $$m = n - r_1 - r_2 + r_1\beta_1 + r_2\beta_2$$

(2.12) $$K = \frac{\left[\sum_{i=a}^{b} y_i + r_1\beta_1 y_a + r_2\beta_2 y_b\right]}{m}$$

ดังนั้นเราจะได้ค่า A B และ C ที่เกิดจากการแทนค่าที่ได้จากสมการที่ (2.7) ถึง(2.12) ดังนี้

(2.13) $$A = n - r_1 - r_2$$

(2.14) $$B = r_2\alpha_2\{y_b - K\} - r_1\alpha_1\{y_a - K\}$$

(2.15) $$C = \sum_{i=a}^{b} y_i^2 + r_1\beta_1 y_a^2 + r_2\beta_2 y_b^2 - mK^2$$

1.1 เราต้องการทดสอบ $H_o: \rho = 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_f ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนดค่า ρ ในสมมุติฐานว่าง(H_0) เป็นดังนี้

$$V(\hat{\rho}/H_o) = \frac{1}{\left(q - \frac{g^2}{A} - 3\right)}$$ ซึ่ง $q - \frac{g^2}{A} = n$ ถ้าเราวิเคราะห์ข้อมูลทั้งหมด

เนื่องจาก $X_i, i = 1,...n$ เป็นอิสระซึ่งกันและกันและมีการแจกแจงเหมือนกัน โดยต่างก็เป็นตัวแปรสุ่มที่ต่อเนื่องซึ่งมีฟังก์ชันความหนาแน่นที่ต่อเนื่องเป็นช่วงๆ(piecewise continuous) f และฟังก์ชันการแจกแจงสะสม F กล่าวคือ $Y_1 \leq Y_2 \leq ... \leq Y_n$

และกำหนดให้ $Z_i = \frac{Y_i - \mu_2}{\sigma_2}$ จะได้ว่า $\forall$, $1 \leq r \leq n$, Z_r จะมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$g_r(z_r) = n\binom{n-1}{r-1} f(z_r) F(z_r)^{r-1} \left(1 - F(z_r)\right)^{n-r}$$

เมื่อ $g = \sum_{i=a}^{b} E(Z_i)$, $q = \sum_{i=a}^{b} E(Z_i^2)$

จะได้ว่าตัวสถิติ Z_f อยู่ในรูปของ

$$Z_f = 0.5\left(q - \frac{g^2}{A} - 3\right)^{1/2} \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)$$

1.2 เราต้องการทดสอบ $H_o: \rho = \rho_o$, $\rho_o \neq 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_f ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนดค่า ρ ในสมมุติฐานว่าง(H_0)เป็นดังนี้

$$V(\hat{\rho} / H_o) = \frac{1}{\left(q - \frac{g^2}{A} - 3\right)}$$

จะได้ว่าตัวสถิติ Z_f อยู่ในรูปของ

$$Z_f = \left(q - \frac{g^2}{A} - 3\right)^{1/2} \left\{0.5 \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) - 0.5 \ln\left(\frac{1+\rho_0}{1-\rho_o}\right)\right\}$$

ซึ่งเราจะปฏิเสธสมมุติฐาน $H_o: \rho = 0$ ถ้า $|Z_f| > Z^*_{\alpha/2}$ เมื่อ $Z_f \sim N(0,1)$

และเราจะปฏิเสธสมมุติฐาน $H_o: \rho = \rho_o$, $\rho_o \neq 0$ ถ้า $|Z_f| > Z^{**}_{\alpha/2}$ เมื่อ $Z_f \sim N(0,1)$

การคำนวณค่าสถิติของตัวสถิติทดสอบ Z_f แสดงได้ดังตัวอย่างที่ 1

ตัวอย่างที่ 1 แพทย์ต้องการตรวจสอบความสัมพันธ์ระหว่างเพศหญิงและเพศชายที่สมรสแล้ว โดยที่คู่สมรสจะต้องเป็นความดันโลหิตสูงที่เกิดขึ้นหลังระยะการบีบตัวของห้องปลายหัวใจ (systolic blood pressure) แพทย์ได้ทำการเลือกคู่สมรสซึ่งมีอายุอยู่ระหว่าง 25-34 ปีจำนวน 20 คู่ ข้อมูลที่เก็บได้เป็นความค่าความดันของเลือดมีดังนี้

เพศหญิง (x) : 136 121 128 100 110 116 127 150 180 172 156 98 132 142 138 126 124 137 160 125

เพศชาย (y) : 110 112 128 106 127 100 98 142 143 150 135 115 125 130 132 146 127 128 135 110

เราต้องการทดสอบ H_o: ไม่มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

เทียบกับ H_1: มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

จากข้อมูล n = A =20 , $r_1 = 0$และ$r_2 = 0$ ดังนั้นค่าของ $\hat{\rho} = \frac{s_{12}\,\hat{\sigma}_2}{s_2^{\,2}\,\hat{\sigma}_1}$ เราสามารถคำนวณได้ดังนี้

ตารางที่ 1.1 เรียงค่าข้อมูลตัวแปรตาม y จากน้อยไปหามาก

i	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	127	116	47.61	726.3025
2	116	100	320.41	622.5025
3	100	106	1149.21	359.1025
4	136	110	4.41	223.5025
5	125	110	79.21	223.5025
6	121	112	166.41	167.7025
7	98	115	1288.81	99.0025
8	132	125	3.61	0.0025
9	110	127	571.21	4.2025
10	124	127	98.21	4.2025
11	137	128	9.61	90.3025
12	128	128	34.81	9.3025
13	142	130	8.10	25.5025
14	138	132	16.81	49.7025
15	160	135	681.21	101.0025
16	156	135	488.41	101.0025
17	150	142	259.21	290.7025

i	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
18	180	143	2125.21	320.8025
19	126	146	65.41	443.1025
20	172	150	1451.61	627.5025
รวม	2678	2499	8923.80	4412.949

ถ้าเราแทนค่า $q_1 = 0$, $q_2 = 0$, $\alpha_1 = \alpha_2 = 0$ และ $\beta_1 = \beta_2 = 0$ ลงในสมการที่ (2.11) และ (2.12) จะได้ว่า

$$m = 20 - 0 = 20$$

และ $$K = \frac{1}{20}\{(2499) + 0(98) + 0(150)\} = 124.95$$

นำค่า α_1 , α_2 , β_1 , β_2 , m และ K แทนค่าลงในสมการที่ (2.13) (2.14) และ (2.15) จะได้ค่า A , B และ C ดังนี้

$$A = 20 - 0 = 20 \ , \ B = 0$$

และ $$C = \sum_{i=1}^{20} y_i^2 + r_1\beta_1 y_1^2 + r_2\beta_2 y_{20}^2 - mK^2$$

$$= 316663 + 0(98)^2 + 0(150)^2 - 20(124.95/20)^2 = 4412.95$$

หลังจากนั้นเราจะนำค่า m , K , A , B และ C แทนค่าลงในสมการที่ (2.2) (2.4) และ (2.3) จะได้ค่า s_{12}^2 $\hat{\sigma}_2$ $\hat{\sigma}_1$ s_1^2 และ s_2^2 เป็นดังนี้

$$s_{12} = \frac{\sum_{i=1}^{20}[x_i - 133.9][y_i - 124.95]}{(20-1)} = 224.836$$

$$\hat{\sigma}_2 = \frac{\left[0 + \sqrt{0 + 4(20)(4412.95)}\right]}{\left[2\sqrt{20(20-1)}\right]} = 15.24$$

$$\hat{\sigma_1} = \left[469.673 + \frac{(224.836)^2}{(232.26)^2}\left[\frac{(15.24)^2}{(232.26)} - 1\right]\right]^{1/2} = 21.67$$

$$s_1^2 = \frac{\sum_{i=1}^{20}\left[x_i - 133.9\right]^2}{(20-1)} = 469.673$$

และ $$s_2^2 = \frac{\sum_{i=1}^{20}\left[y_i - 124.95\right]^2}{(20-1)} = 232.260$$

ดังนั้น

$$\hat{\rho} = \frac{(224.836)(15.24)}{(232.26)(21.67)} = 0.68$$

จากโจทย์ตัวอย่าง จำนวนข้อมูลครบถ้วนเราจะได้ค่าความแปรปรวนของตัวสถิติทดสอบ Z_f มีค่าเท่ากับ $\frac{1}{(20-3)}$

จะได้ค่าตัวสถิติทดสอบ Z_f ดังนี้

$$Z_f = 0.5\,(20-3)^{1/2}\,\ln\left(\frac{1+0.68}{1-0.68}\right) = 3.418524$$

เราจะนำค่าตัวสถิติทดสอบ Z_f ที่ได้จากการเปิดตารางการแจกแจงปกติมาตรฐานมาเปรียบเทียบกับค่าตัวสถิติทดสอบ Z_f ที่คำนวณได้ จะพบว่า $Z_{0.05} = 1.96$ น้อยกว่า 3.418524

เพราะฉะนั้น เราจะปฏิเสธสมมุติฐานที่ว่า การเกิดความดันโลหิตสูงหลังระยะการบีบตัวของห้องปลายหัวใจของคู่สมรสไม่มีความสัมพันธ์กันระหว่างเพศชายและเพศหญิง

2. ตัวสถิติทดสอบ Z_k (Konishi Statistics)

ในปี ค.ศ. 1978 Konishi ได้เสนอตัวสถิติทดสอบ Z_k ซึ่งพัฒนามาจากตัวสถิติทดสอบ Z_f โดยที่ความแปรปรวนของตัวสถิติทดสอบ Z_k จะปรับปรุงมาจากค่าความ

แปรปรวนของตัวสถิติทดสอบ Z_f ในส่วนของความแปรปรวนของตัวสถิติทดสอบ Z_k ได้แบ่งออกเป็นสองส่วนคือส่วนของจำนวนข้อมูลกับส่วนของเทอมถูกต้อง(correction term) ที่ขึ้นอยู่กับค่า $\hat{\rho}$ ของข้อมูล การแจกแจงของค่า r ที่ใช้ในตัวสถิติทดสอบ Z_k นั้นจะมีความถูกต้องในการประมาณค่าการแจกแจงของค่า r มากขึ้นและมีคุณสมบัติของฟังก์ชันการแจกแจงเมื่อใกล้อนันต์ ดังนั้นถ้าขนาดตัวอย่างมีจำนวนน้อยการประมาณค่า $\hat{\rho}$ จะถูกต้องมากขึ้น Konishi จึงเสนอตัวสถิติทดสอบ Z_k ซึ่งพิจารณาจากค่าอัตราส่วนของการแปลงข้อมูลกับค่ารากที่สองของความแปรปรวนซึ่งขึ้นอยู่กับค่า $\hat{\rho}$

ให้ $X_{(a)},..X_{(i)}..,X_{(b)}$ เป็นตัวสถิติอันดับของ X , $i = (a = r_1 + 1,..., b = n - r_2)$

และให้ $Y_{(a)},..Y_{(i)}..,Y_{(b)}$ เป็นตัวสถิติอันดับของ Y โดยค่าตัวแปร X จะสัมพันธ์กับค่าตัวแปร Y ที่เรียงลำดับ (concomitant order statistics)

2.1 เราต้องทดสอบ $H_o: \rho = 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_k ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนดค่า ρ ในสมมุติฐานว่าง(H_0)เป็นดังนี้

$$V(\hat{\rho}/H_o)=\frac{1}{\left(q-\frac{g^2}{A}-2.5+0.25\hat{\rho}^2\right)}=\frac{1}{M}$$ ซึ่ง $q-\frac{g^2}{A}=n$ ถ้าเราวิเคราะห์ข้อมูลทั้งหมด

จะได้ว่าตัวสถิติ Z_k อยู่ในรูปของ

$$Z_k = 0.5\, M^{1/2} \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)$$

2.2 เมื่อทดสอบ $H_o: \rho = \rho_o\ ,\ \rho_o \neq 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_k ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนดค่า ρ ในสมมุติฐานว่าง(H_0)เป็นดังนี้

$$V(\hat{\rho}/H_o)=\frac{\left(1-\hat{\rho}^2+\hat{\rho}^2\left(\frac{q}{A}-\frac{g^2}{A^2}\right)\right)}{M}$$

จะได้ว่าตัวสถิติ Z_k อยู่ในรูปของ

$$Z_k=\frac{M^{1/2}\left[0.5\ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)-0.5\ln\left(\frac{1+\rho_o}{1-\rho_o}\right)\right]}{\left(1-\hat{\rho}^2+\hat{\rho}^2\left(\frac{q}{A}-\frac{g^2}{A^2}\right)\right)^{1/2}}$$

ซึ่งเราจะปฏิเสธสมมุติฐาน $H_o: \rho=0$ ถ้า $|Z_k(A)|> K^*_{\alpha/2}(A)$ เมื่อ $Z_k(A) \sim$ adjusted N(0,1) distribution

และเราจะปฏิเสธสมมุติฐาน $H_o: \rho=\rho_o$, $\rho_o \neq 0$ ถ้า $|Z_k(A)|> K^{**}_{\alpha/2}(A)$ เมื่อ $Z_k(A) \sim$ adjusted N(0,1) distribution

การคำนวณค่าสถิติของตัวสถิติทดสอบ Z_k แสดงได้ดังตัวอย่างที่ 2

ตัวอย่างที่ 2 จากข้อมูลในโจทย์ตัวอย่างที่ 1 เราจะได้ว่าค่า $n = A = 20$ $s_1^2 = 469.93$ $s_2^2 = 232.260$ $s_{12}^2 = 224.836$ $\hat{\sigma}_2 = 15.24$ และ $\hat{\sigma}_1 = 21.67$

เราต้องการทดสอบ H_o: ไม่มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

เทียบกับ H_1: มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

ดังนั้น

$$\hat{\rho}=\frac{s_{12}\hat{\sigma}_2}{s_2^2\hat{\sigma}_1}=0.68$$

จากโจทย์ตัวอย่างจำนวนข้อมูลครบถ้วน เราจะได้ค่าความแปรปรวนของตัวสถิติทดสอบ Z_k มีค่าเท่ากับ $\frac{1}{(20-2.5+0.25(0.68)^2}=\frac{1}{17.61556}$

จะได้ค่าตัวสถิติทดสอบ Z_k ดังนี้

$$Z_k = 0.5\,(17.61556)^{1/2}\,\ln\left(\frac{1+0.68}{1-0.68}\right) = 3.47986$$

เราจะนำค่าตัวสถิติทดสอบ Z_k ที่ได้จากตาราง adjusted standard normal distribution มาเปรียบเทียบกับค่าตัวสถิติทดสอบ Z_k ที่คำนวณได้พบว่า $K_{0.05}(20) = 1.99757$ น้อยกว่า 3.47986

เพราะฉะนั้น เราจะปฏิเสธสมมุติฐานที่ว่า การเกิดความดันโลหิตสูงหลังระยะการบีบตัวของห้องปลายหัวใจของคู่สมรสไม่มีความสัมพันธ์กันระหว่างเพศชายและเพศหญิง

3. ตัวสถิติทดสอบ Z_v (Vaughan Statistics)

ในปี ค.ศ. 1975 Tiku และ ค.ศ. 1985 Bhattacharyya เสนอการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์มีการแจกแจงแบบปกติเมื่อจำนวนข้อมูลมากขึ้น กำหนดค่า r_1 และ r_2 คงที่ ดังนั้นการแจกแจงของ ρ จะเป็นการแจกแจงปกติ เราจะพิจารณาตัวสถิติทดสอบ Z_v จากค่าอัตราส่วนตัวประมาณ MML ของ ρ คือ ρ กับค่ารากที่สองของความแปรปรวน

ให้ $X_{(a)},..X_{(i)}..,X_{(b)}$ เป็นตัวสถิติอันดับของ X , $i = (a = r_1 + 1,..., b = n - r_2)$

และให้ $Y_{(a)},..Y_{(i)}..,Y_{(b)}$ เป็นตัวสถิติอันดับของ Y โดยค่าตัวแปร X จะสัมพันธ์กับค่าตัวแปร Y ที่เรียงลำดับ (concomitant order statistics)

3.1 เราต้องการทดสอบ $H_o: \rho = 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_v ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนด ρ ในสมมุติฐานว่าง(H_0)เป็นดังนี้

$$V(\hat{\rho}/H_o) = \frac{1}{q - \frac{g^2}{A}} \text{ ซึ่ง } q - \frac{g^2}{A} = n \text{ ถ้าเราวิเคราะห์ข้อมูลทั้งหมด}$$

จะได้ตัวสถิติ Z_v อยู่ในรูปของ

$$Z_v = \hat{\rho}\sqrt{q - \frac{g^2}{A}}$$

3.2 เราต้องการทดสอบ $H_o: \rho = \rho_o , \rho_o \neq 0$

เทียบกับ $H_1: \rho \neq 0$

ค่าความแปรปรวนของข้อมูลของตัวสถิติทดสอบ Z_v ที่มีเงื่อนไขอยู่ในรูป $\hat{\rho}$ เมื่อเรากำหนด ρ ในสมมุติฐานว่าง(H_0)เป็นดังนี้

$$V(\hat{\rho}/H_o) = \frac{(1-\rho_o^2)^2\left(1-\rho_o^2+\rho_o^2\left(\frac{q}{A}-\frac{g^2}{A^2}\right)\right)}{\left(q-\frac{g^2}{A}\right)}$$

จะได้ตัวสถิติ Z_v อยู่ในรูปของ

$$Z_v = \frac{(\hat{\rho}-\rho_o)\left(q-\frac{g^2}{A}\right)^{1/2}}{(1-\rho_o^2)\left[1-\rho_o^2+\rho_o^2\left(\frac{q}{A}-\frac{g^2}{A^2}\right)\right]^{1/2}}$$

ซึ่งเราจะปฏิเสธสมมุติฐาน $H_o: \rho = 0$ ถ้า $|Z_v| > Z^*_{\alpha/2}$ เมื่อ $Z_v \sim N(0,1)$

และเราจะปฏิเสธสมมุติฐาน $H_o: \rho = \rho_o , \rho_o \neq 0$ ถ้า $|Z_v| > Z^{**}_{\alpha/2}$ เมื่อ $Z_v \sim N(0,1)$

การคำนวณค่าสถิติของตัวสถิติทดสอบ Z_v แสดงได้ดังตัวอย่างที่ 3

ตัวอย่างที่ 3 จากข้อมูลในโจทย์ตัวอย่างที่ 1 เราจะได้ว่าค่า $n = A = 20$, $s_1^2 = 469.93$ $s_2^2 = 232.260$, $s_{12}^2 = 224.836$, $\hat{\sigma}_2 = 15.24$ และ $\hat{\sigma}_1 = 21.67$

เราต้องการทดสอบ H_o: ไม่มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

เทียบกับ H_1: มีสหสัมพันธ์ระหว่างเพศชายและเพศหญิง

ดังนั้น

$$\hat{\rho} = \frac{s_{12}\hat{\sigma_2}}{s_2^2\hat{\sigma_1}} = 0.68$$

จากโจทย์ตัวอย่างจำนวนข้อมูลครบถ้วน เราจะได้ค่าความแปรปรวนของตัวสถิติทดสอบ Z_v มีค่าเท่ากับ $\frac{1}{20}$

จะได้ค่าตัวสถิติทดสอบ Z_v ดังนี้

$$Z_v = \sqrt{20}(0.68) = 3.04104$$

เราจะนำค่าตัวสถิติทดสอบ Z_v ที่ได้จากการเปิดตารางการแจกแจงปกติมาตรฐานนำไปเปรียบเทียบกับค่าตัวสถิติทดสอบ Z_v ที่คำนวณได้ พบว่า $Z_{0.05} = 1.96$ น้อยกว่า 3.04104

เพราะฉะนั้น เราจะปฏิเสธสมมุติฐานที่ว่า การเกิดความดันโลหิตสูงหลังระยะการบีบ-ตัวของห้องปลายหัวใจของคู่สมรสไม่มีความสัมพันธ์กันระหว่างเพศชายและเพศหญิง