

บทที่ 4 ทฤษฎีสัมประสิทธิ์สหสัมพันธ์อันดับที่

ในบทนี้จะเริ่มศึกษาทฤษฎีบทของสัมประสิทธิ์สหสัมพันธ์อันดับที่ โดยมีการพิสูจน์ทฤษฎีบท ซึ่งในบทที่ 3 เป็นรวบรวมทฤษฎี, แสดงการคำนวณสัมประสิทธิ์อันดับที่ และทดสอบความมีนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์อันดับที่ ดังนั้นในบทนี้จะได้กล่าวถึงทฤษฎีบทของสัมประสิทธิ์สหสัมพันธ์อันดับที่ดังต่อไปนี้

- สัมประสิทธิ์สหสัมพันธ์อันดับที่ทั่วไป
- การแจกแจงของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลลี
- การแจกแจงของ สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน
- การพิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่ส่วนย่อยเคนคอลลี
- การพิสูจน์สูตรสัมประสิทธิ์ความสอดคล้องเคนคอลลี
- การคำนวณสัมประสิทธิ์สหสัมพันธ์อันดับที่ กรณีอันดับที่ของค่าสังเกตซ้ำกัน

4.1 สัมประสิทธิ์สหสัมพันธ์อันดับที่ทั่วไป

ในที่นี้สนใจศึกษาเกี่ยวกับลักษณะของสัมประสิทธิ์สหสัมพันธ์อันดับที่ทั่วไป และการคำนวณที่มาสูตร ของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน จากรูปแบบต่างๆ

สมมติว่ามีเซตของสิ่งที่สนใจ n สิ่ง ซึ่งมีความสัมพันธ์กับคุณลักษณะ 2 ลักษณะ ให้ตัวแปร X และ Y โดยกำหนดเป็นเลขจำนวน 1 ถึง n เพื่อวัตถุประสงค์บางอย่างหรือการเรียงอันดับ ดังนั้นจะได้ $X_1, \dots, X_n$ เป็นของตัวแปร X และ $Y_1, \dots, Y_n$ เป็นของตัวแปร Y ซึ่งค่าเหล่านี้อาจเปลี่ยนแปลงหรือเรียงลำดับได้

ในแต่ละคู่ ของ X และ Y นั่นคือที่ i และที่ j จะให้คะแนน X กำหนดโดย a_{ij} ซึ่งอยู่ภายใต้สัญลักษณ์ $a_{ij} = -a_{ji}$ ในทำนองเดียวกัน จะให้คะแนน Y เป็น b_{ij} และ $b_{ij} = -b_{ji}$ โดยนิยามสัมประสิทธิ์สหสัมพันธ์ แกมมา (Gamma ; Γ) ดังสมการต่อไปนี้

$$\Gamma = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}} \quad (4.1)$$

โดยไม่พิจารณา a_{ij} เป็น ศูนย์ เมื่อ $i \neq j$

ดังนั้นบทนี้จึงเป็นการศึกษาสัมประสิทธิ์สหสัมพันธ์ทั่วไป และจะศึกษารายละเอียดต่อไปนี้

- สัมประสิทธิ์สหสัมพันธ์อันดับที่หั่วไปกรณีที่ 1
- สัมประสิทธิ์สหสัมพันธ์อันดับที่หั่วไปกรณีที่ 2
- การพิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน
- สัมประสิทธิ์สหสัมพันธ์ในรูปผลคูณ โมเมนต์

สัมประสิทธิ์สหสัมพันธ์อันดับที่หั่วไปกรณีที่ 1 (τ)

ให้ p_i และ q_i แทนอันดับที่ i ของตัวแปร X และ Y ตามลำดับ
 p_j และ q_j แทนอันดับที่ j ของตัวแปร X และ Y ตามลำดับ
 a_{ij} คือคะแนนการเปรียบเทียบของอันดับที่ i และ j ของตัวแปร X
 b_{ij} คือคะแนนการเปรียบเทียบของอันดับที่ i และ j ของตัวแปร Y

$$\left. \begin{array}{l} \text{สมมติว่าให้ คะแนน } +1 \text{ เมื่อ } p_i < p_j \\ \text{และคะแนน } -1 \text{ เมื่อ } p_i > p_j \\ \text{นั่นคือ } a_{ij} = +1 \text{ เมื่อ } p_i < p_j \\ a_{ij} = -1 \text{ เมื่อ } p_i > p_j \end{array} \right\} \quad (4.2)$$

ในทำนองเดียวกันสำหรับคะแนน ของตัวแปร Y กำหนดให้เป็น b_{ij} ดังนั้น $\sum a_{ij}b_{ij}$ เป็น 2 เท่าของความแตกต่างระหว่างผลบวกของจำนวนอันดับที่ ที่สอดคล้องกันกับผลบวกของอันดับที่ ที่ไม่สอดคล้องกัน (S) และ $\sum a_{ij}^2 = n(n-1)$ ในทำนองเดียวกัน $\sum b_{ij}^2 = n(n-1)$ ดังนั้นจาก (4.1) Γ (แกมมา) จะเท่ากับสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ นั่นคือ

$$\Gamma = \frac{\sum a_{ij}b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}}$$

$$\tau = \frac{2S}{n(n-1)}$$

$$\Gamma = \tau$$

สัมประสิทธิ์สหสัมพันธ์อันดับที่หั่วไปกรณีที่ 2 (r_s)

กำหนดให้

$$\left. \begin{aligned} a_{ij} &= p_j - p_i \\ b_{ij} &= q_j - q_i \end{aligned} \right\} \quad (4.3)$$

p_i, p_j เป็นอันดับที่ i และ j ของตัวแปร X และ

q_i, q_j เป็นอันดับที่ i และ j ของตัวแปร Y

ทั้ง p_i และ q_i อยู่ในช่วง 1 ถึง n ดังนั้นจะได้ว่า

$$\sum (p_i - p_j)^2 = \sum (q_i - q_j)^2$$

พิสูจน์สูตร สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน

การพิสูจน์นี้ จะเป็นการพิสูจน์สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนจากแกมมาในสมการที่ 4.1 ซึ่งจะได้สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนดังรูปแบบต่อไปนี้

$$\Gamma = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

จากสมการที่ (4.1)

$$\Gamma = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}}$$

$$\Gamma = \frac{\sum (p_i - p_j)(q_i - q_j)}{\sum (p_i - p_j)^2} \quad (4.4)$$

1. พิจารณาเทอมของ $\sum_{i,j} (p_i - p_j)(q_i - q_j)$

$$\sum_{i,j} (p_i - p_j)(q_i - q_j) = \sum_i \sum_j (p_i q_i) + \sum_i \sum_j (p_j q_j) - \sum_i \sum_j (p_i q_j + p_j q_i)$$

$$= \sum_j (p_1 q_1 + p_2 q_2 + \dots + p_n q_n) + \sum_i (p_1 q_1 + p_2 q_2 + \dots + p_n q_n)$$

$$- \sum_i \sum_j p_i q_j - \sum_j \sum_i p_j q_i$$

$$= n \sum_i^n p_i q_i + n \sum_j^n p_j q_j - \left(\sum_i^n p_i \sum_j^n q_j + \sum_i^n q_i \sum_j^n p_j \right)$$

เนื่องจาก

$$\sum_i^n p_i = \sum_j^n p_j = \sum_i^n q_i = \sum_j^n q_j \text{ และ}$$

$$\sum_i^n p_i q_i = \sum_j^n p_j q_j$$

จะได้ว่า

$$\sum_{i,j}^n (p_i - p_j)(q_i - q_j) = 2n \sum_i^n p_i q_i - 2 \sum_i^n p_i \sum_j^n q_j \quad (4.5)$$

เนื่องจาก p และ q เป็นจำนวนเต็มบวก ตั้งแต่ $1, 2, 3, \dots, n$ ดังนั้นจะได้ว่า
ผลบวกของ p หรือ q เป็นดังนี้

$$\sum_{i=1}^n p_i = 1 + 2 + 3 + \dots + n \text{ จะเห็นว่าเป็นผลบวกอนุกรมเลขคณิตซึ่ง}$$

$$\sum p_i = \frac{1}{2} n(n+1) \quad (4.6)$$

ผลบวกกำลังสองของ p หรือ q เป็นดังนี้

$$\sum_{i=1}^n p_i^2 = 1^2 + 2^2 + 3^2 + \dots + n^2 \text{ จะเห็นว่าเป็นผลบวกอนุกรมซึ่ง}$$

$$\sum p_i^2 = \frac{1}{6} n(n+1)(2n+1) \quad (4.7)$$

แทนสมการที่ (4.6) ในสมการที่ (4.5) จะได้ว่า

$$\begin{aligned}\sum_{i,j}^n (p_i - p_j)(q_i - q_j) &= 2n \sum_i^n p_i q_i - 2 \left[\frac{n(n+1)}{2} \times \frac{n(n+1)}{2} \right] \\ &= 2n \sum_i^n p_i q_i - 2 \left[\frac{n^2(n+1)^2}{4} \right]\end{aligned}$$

$$\sum_{i,j}^n (p_i - p_j)(q_i - q_j) = 2n \sum_i^n p_i q_i - \frac{1}{2} n^2 (n+1)^2 \quad (4.8)$$

ขั้นตอนต่อไปพิจารณาเทอมของ $2n \sum_i^n p_i q_i$

กำหนดให้

$$\begin{aligned}\sum_{i=1}^n d_i^2 &= \sum_{i=1}^n (p_i - q_i)^2 \\ &= \sum_{i=1}^n (p_i^2 - 2p_i q_i + q_i^2) \\ &= \sum_{i=1}^n p_i^2 - 2 \sum_{i=1}^n p_i q_i + \sum_{i=1}^n q_i^2\end{aligned}$$

$$\text{จาก } \sum_{i=1}^n p_i = \sum_{i=1}^n q_i$$

จะได้

$$\sum_{i=1}^n d_i^2 = 2 \sum_{i=1}^n p_i^2 - 2 \sum_{i=1}^n p_i q_i$$

$$2 \sum_{i=1}^n p_i q_i = 2 \sum_{i=1}^n p_i^2 - \sum_{i=1}^n d_i^2 \quad (4.9)$$

n คูณ สมการที่ (4.9)

$$2n \sum_{i=1}^n p_i q_i = 2n \sum_{i=1}^n p_i^2 - n \sum_{i=1}^n d_i^2 \quad (4.10)$$

สมการที่ (4.10) แทนในสมการที่ (4.8) จะได้

$$\sum_{i,j}^n (p_i - p_j)(q_i - q_j) = 2n \sum_i^n p_i^2 - \frac{1}{2} n^2 (n+1)^2 - n \sum_{i=1}^n d_i^2 \quad (4.11)$$

จาก สมการที่ (4.7) $\sum_i^n p_i^2 = \frac{n(n+1)(2n+1)}{6}$ แทนใน สมการที่ (4.11) จะได้

$$\begin{aligned} \sum_{i,j}^n (p_i - p_j)(q_i - q_j) &= 2n \left[\frac{n(n+1)(2n+1)}{6} \right] - \frac{1}{2} n^2 (n+1)^2 - n \sum_i^n d_i^2 \\ &= \frac{n^2(n+1)(2n+1)}{3} - \frac{1}{2} n^2 (n+1)^2 - n \sum_i^n d_i^2 \\ &= \frac{2n^2(n+1)(2n+1)}{3} - \frac{3}{6} n^2 (n+1)^2 - n \sum_i^n d_i^2 \\ &= \frac{2n^2(n+1)(2n+1) - 3n^2(n+1)(n+1)}{6} - n \sum_i^n d_i^2 \\ &= \frac{2n^2(2n^2 + 3n + 1) - 3n^2(n^2 + 2n + 1)}{6} - n \sum_i^n d_i^2 \\ &= \frac{n^2(4n^2 + 6n + 2 - 3n^2 - 6n - 3)}{6} - n \sum_i^n d_i^2 \\ &= \frac{n^2(n^2 - 1)}{6} - n \sum_i^n d_i^2 \end{aligned} \quad (4.12)$$

$$2. \text{ พิจารณาเทอม } \sum_{i,j}^n (p_j - p_i)^2 = \sum_{i,j}^n p_j^2 - 2 \sum_{i,j}^n p_i p_j + \sum_{i,j}^n p_i^2$$

$$\begin{aligned}
\sum_{i,j}^n (p_j - p_i)^2 &= 2n \sum_i^n p_i^2 - 2 \sum_{i,j}^n p_i p_j \\
&= 2n \sum_i^n p_i^2 - 2 \left(\sum_{i,j}^n p_i \right)^2 \\
&= \frac{2n \times n(n+1)(2n+1)}{6} - \frac{2n^2(n+1)^2}{4} \\
&= \frac{2n^2(n+1)(2n+1) - 3n^2(n+1)^2}{6} \\
&= \frac{n^2(n+1)}{6} [2(2n+1) - 3(n+1)] \\
&= \frac{1}{6} n^2(n+1)(n-1)
\end{aligned}$$

$$\text{ดังนั้น } \sum_{i,j}^n (p_j - p_i)^2 = \frac{1}{6} n^2(n^2 - 1) \quad (4.13)$$

แทนสมการที่ (4.12) และ (4.13) ในสมการที่ (4.4) จะได้

$$\begin{aligned}
\Gamma &= \frac{\sum (p_i - p_j)(q_i - q_j)}{\sum (p_i - p_j)^2} \\
\Gamma &= \frac{\left[\frac{n^2(n^2-1)}{6} - n \sum_i^n d_i^2 \right]}{\left[\frac{1}{6} n^2(n^2-1) \right]} \\
\Gamma &= 1 - \frac{6 \sum d_i^2}{n^3 - n}
\end{aligned}$$

นั่นคือ Γ เป็นสหสัมพันธ์อันดับที่สเปียร์แมน

สหสัมพันธ์ในรูปผลคูณโมเมนต์

สหสัมพันธ์ในรูปผลคูณโมเมนต์ ก็เป็นอีกวิธีหนึ่งที่ใช้ในการการคำนวณที่มาของสูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน

กำหนดให้

$$\begin{aligned} a_{ij} &= x_j - x_i \\ b_{ij} &= y_j - y_i \end{aligned} \quad (4.14)$$

ดังนั้น

$$\frac{1}{2} \sum_{i,j} (x_j - x_i)(y_j - y_i) = n \sum_i x_i y_i - \sum_{i,j} x_i y_j \quad (4.15)$$

สมการที่ (4.15) เป็นความแปรปรวนร่วมของ X และ Y

$$\frac{1}{2} \sum_{i,j} (x_j - x_i)^2 = n \sum_i x_i^2 - (\sum_i x_i)^2 \quad (4.16)$$

สมการที่ (4.16) เป็นความแปรปรวนของ X

$$\text{จาก } \Gamma = \frac{\text{COV}(X, Y)}{\sqrt{\text{Var}X \text{ Var}Y}} = \frac{\mu_{11}}{\sqrt{\mu_2(x)\mu_2(y)}} \quad (4.17)$$

ซึ่ง Γ ก็จะกลายเป็นผลคูณโมเมนต์สหสัมพันธ์ของตัวแปร X และ Y

พิสูจน์ ผลคูณโมเมนต์สหสัมพันธ์ของตัวแปร X และ Y คือ สหสัมพันธ์อันดับที่สเปียร์แมน การพิสูจน์จะพิสูจน์จากการหาความแปรปรวน และ ความแปรปรวนร่วมโดยวิธีโมเมนต์

$$\begin{aligned} V(x) &= E(x^2) - [E(x)]^2 \\ V(p) &= E(p^2) - [E(p)]^2 \\ &= \mu_2' - (\mu_1')^2 \end{aligned} \quad (4.18)$$

1. คำนวณหาโมเมนต์ที่ 1 รอบจุดศูนย์กลางของ $p (\mu_1')$

เมื่อเซตของข้อมูลจำนวนเต็ม n จำนวน จากสมการที่ 4.6 จะได้ว่า

$$\sum p_i = \frac{1}{2}n(n+1)$$

ดังนั้น โมเมนต์ที่ 1 รอบจุดศูนย์กลางของ p คือ

$$\mu_1' = E(p')$$

$$\mu_1' = \frac{\sum p}{n}$$

$$= \frac{n(n+1)}{2n}$$

$$= \frac{n+1}{2}$$

(4.19)

2. คำนวณหาโมเมนต์ที่ 2 รอบจุดศูนย์กลางของ $p (\mu_2')$

ในทำนองเดียวกัน จากสมการที่ (4.7)

$$\sum p_i^2 = \frac{n(n+1)(2n+1)}{6}$$

โมเมนต์ที่ 2 รอบจุดศูนย์กลางของ p

$$\mu_2' = E(p'^2)$$

$$\mu_2' = \frac{\sum p^2}{n}$$

$$= \frac{n(n+1)(2n+1)}{6n}$$

$$= \frac{(n+1)(2n+1)}{6}$$

(4.20)

ดังนั้นความแปรปรวนคือ โมเมนต์ที่ 2 รอบค่าเฉลี่ย

$$\begin{aligned}
 \mu_2 &= E(p - \mu)^2 \\
 &= E(p^2) - (\mu_1')^2 \\
 &= \mu_2' - (\mu_1')^2 \\
 &= \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
 &= \frac{4(n+1)(2n+1) - 6(n+1)^2}{24} \\
 &= \frac{2(n+1)}{24} [2(2n+1) - 3(n+1)] \\
 &= \frac{1}{12} (n+1)(n-1) \\
 &= \frac{1}{12} (n^2 - 1)
 \end{aligned} \tag{4.21}$$

3. พิจารณาเทอม μ_{11}

จากสมการที่ (4.12)

$$\sum_{i,j}^n (p_i - p_j)(q_i - q_j) = \frac{n^2(n^2 - 1)}{6} - n \sum_i^n d_i^2$$

$$n \sum_i^n d_i^2 = \frac{n^2(n^2 - 1)}{6} - \sum_{i,j}^n (p_i - p_j)(q_i - q_j) \tag{4.22}$$

สมการที่ (4.22) หาคด้วย n

$$\sum_i^n d_i^2 = \frac{n(n^2-1)}{6} - \frac{1}{n} \sum_{i,j}^n (p_i - p_j)(q_i - q_j) \quad (4.23)$$

สมการที่ (4.23) หาคด้วย 2 และ

จากสมการที่ (4.5) $\sum_{i,j}^n (p_i - p_j)(q_i - q_j) = 2n \sum_i^n p_i q_i - 2 \sum_i^n p_i \sum_j^n q_j$ แทนในสมการที่ (4.23)

$$\begin{aligned} \frac{1}{2} \sum_i^n d_i^2 &= \frac{n(n^2-1)}{12} - \frac{1}{2n} [2n \sum_i^n p_i q_i - 2 \sum_i^n p_i \sum_j^n q_j] \\ &= \frac{n(n^2-1)}{12} - \frac{1}{n} [n \sum_i^n p_i q_i - \sum_i^n p_i \sum_j^n q_j] \\ &= \frac{n(n^2-1)}{12} - \frac{1}{n} \times n [\sum_i^n p_i q_i - \frac{1}{n} \sum_i^n p_i \sum_j^n q_j] \\ &= \frac{n(n^2-1)}{12} - n [\frac{1}{n} \sum_i^n p_i q_i - \frac{1}{n} \sum_i^n p_i \frac{1}{n} \sum_j^n q_j] \quad (4.24) \end{aligned}$$

สมการที่ (4.24) หาคด้วย n

$$\frac{1}{2n} \sum_i^n d_i^2 = \frac{(n^2-1)}{12} - [\frac{1}{n} \sum_i^n p_i q_i - \frac{1}{n} \sum_i^n p_i \frac{1}{n} \sum_j^n q_j]$$

$$\frac{1}{2n} \sum_i^n d_i^2 = \frac{(n^2-1)}{12} - \mu_{11}$$

$$\mu_{11} = \frac{(n^2-1)}{12} - \frac{1}{2n} \sum_i^n d_i^2 \quad (4.25)$$

แทนสมการที่ (4.21) และ (4.25) ในสมการที่ (4.17)

$$\Gamma = \frac{\text{COV}(X, Y)}{\sqrt{\text{Var}X \text{Var}Y}} = \frac{\mu_{11}}{\sqrt{\mu_2(x)\mu_2(y)}}$$

$$\mu_2(x) = \mu_2(y) = \frac{1}{12}(n^2 - 1)$$

ดังนั้น

$$\Gamma = \frac{\frac{1}{12}(n^2 - 1) - \frac{1}{2n} \sum d^2}{\frac{1}{12}(n^2 - 1)}$$

$$\Gamma = 1 - \frac{6 \sum d^2}{n^3 - n}$$

Γ ก็จะกลายเป็นผลคูณโมเมนต์สหสัมพันธ์ของตัวแปร X และ Y หรือเป็นสหสัมพันธ์อันดับที่สเปียร์แมน

การพิสูจน์สูตรสหประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน (r_s) จากสูตรสหประสิทธิ์เพียร์สัน (r)

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

กำหนดตัวแปร

- n แทน จำนวนคู่ของอันดับ
- X_i แทน อันดับของคะแนนชุด X
- Y_i แทน อันดับของคะแนนชุด Y
- $\bar{X}$ แทน ค่าเฉลี่ยอันดับของคะแนนชุด X
- $\bar{Y}$ แทน ค่าเฉลี่ยอันดับของคะแนนชุด Y

จากสูตรของ สัมประสิทธิ์สหสัมพันธ์ เพียร์สัน r

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (4.26)$$

กำหนดให้

$$x_i = X_i - \bar{X}$$

$$y_i = Y_i - \bar{Y}$$

ดังนั้น $x_i^2 = (X_i - \bar{X})^2$

ใส่ผลรวม($\sum$)จะได้

$$\begin{aligned} \sum x_i^2 &= \sum (X_i - \bar{X})^2 \\ &= \sum X_i^2 - \frac{(\sum X_i)^2}{n} \end{aligned} \quad (4.27)$$

X_i เป็นจำนวนเต็มบวก $1, 2, \dots, n$ ดังนั้น จากสมการที่ 4.6 และ 4.7

$$\sum X_i = 1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$$

$$\sum X_i^2 = 1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

แทน $\sum X_i$ และ $\sum X_i^2$ ในสมการ (4.27) จะได้

$$\begin{aligned} \sum x_i^2 &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \\ &= \frac{n^3 - n}{12} \end{aligned} \quad (4.28)$$

ในทำนองเดียวกันจะได้

$$\begin{aligned}\sum y_i^2 &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \\ &= \frac{n^3 - n}{12}\end{aligned}\quad (4.29)$$

จาก $d_i = X_i - Y_i$

$$d_i = (X_i - \bar{X}) - (Y_i - \bar{Y})$$

เนื่องจาก $\bar{X} = \bar{Y}$

ดังนั้น $d_i = x_i - y_i$

$$d_i^2 = (x_i - y_i)^2 = x_i^2 - 2x_i y_i + y_i^2$$

$$\sum d_i^2 = \sum x_i^2 + \sum y_i^2 - 2\sum x_i y_i$$

$$2\sum x_i y_i = \sum x_i^2 + \sum y_i^2 - \sum d_i^2$$

$$\sum x_i y_i = \frac{1}{2}(\sum x_i^2 + \sum y_i^2 - \sum d_i^2) \quad (4.30)$$

จากสมการ (4.26)

$$r_s = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

$$\sum x_i y_i = r_s \sqrt{\sum x_i^2 \sum y_i^2} \quad (4.31)$$

สมการ (4.30) = (4.31) จะได้

$$r_s \sqrt{\sum x_i^2 \sum y_i^2} = \frac{1}{2}(\sum x_i^2 + \sum y_i^2 - \sum d_i^2)$$

$$r_s = \frac{\sum x_i^2 + \sum y_i^2 - \sum d_i^2}{2\sqrt{\sum x_i^2 \sum y_i^2}}$$

นำค่า $\sum x_i^2, \sum y_i^2$ แทนใน สมการที่ (4.26)

$$r_s = \frac{\left(\frac{n^3-n}{12}\right) + \left(\frac{n^3-n}{12}\right) - \sum d_i^2}{2\sqrt{\left(\frac{n^3-n}{12}\right)\left(\frac{n^3-n}{12}\right)}}$$

$$= 1 - \frac{\sum d_i^2}{\left(\frac{n^3-n}{6}\right)}$$

$$= 1 - \frac{6\sum d_i^2}{n^3-n}$$

ดังนั้นจะได้ว่า

$$r_s = 1 - \frac{6\sum_{i=1}^n d_i^2}{n^3-n}$$

4.2 การแจกแจงของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์

จากการศึกษาเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ จะพบว่าเป็นการหาความสัมพันธ์ของการจัดอันดับของข้อมูลหรือตัวแปร 2 ชุด ซึ่งเป็นการศึกษาจากข้อมูลกลุ่มตัวอย่าง เพื่อที่จะนำไปอนุมานค่าพารามิเตอร์โดยการทดสอบสมมติฐานหรือประมาณค่า ดังนั้นในที่นี้จึงสนใจที่จะศึกษารายละเอียดเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ ดังต่อไปนี้

- การพิสูจน์ที่มาของสูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์
- การแจกแจงสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ กรณี n มีขนาดเล็ก
- การแจกแจงสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ กรณี n มีขนาดใหญ่
- การพิสูจน์สูตรการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์

4.2.1 พิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์

$$\tau = \frac{2S}{n(n-1)}$$

ซึ่งเป็นสูตรที่ใช้ในการหาสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์

กำหนดให้ E เป็นเหตุการณ์หรือคะแนนที่ได้จากการเปรียบเทียบในแต่ละคู่ลำดับ

จากสูตร
$$P(E) = \frac{a}{n}$$

a : น้ำหนักของคะแนนที่ได้จากการเปรียบเทียบในแต่ละคู่ลำดับ

n : คะแนนทั้งหมดหรือสมาชิกในปริภูมิตัวอย่างซึ่งมีค่าเท่ากับ $\binom{n}{2}$

ให้ S คือผลรวมของคะแนน +1 หรือ -1 ในการเปรียบเทียบลำดับที่ในแต่ละคู่อันดับ ดังนั้น จะได้สูตรในการคำนวณค่า τ

$$\tau = \frac{\text{น้ำหนักของคะแนนที่เกิดจากการเปรียบเทียบในคู่ลำดับ}}{\text{น้ำหนักของคะแนนทั้งหมดใน ปริภูมิตัวอย่าง}}$$

$$= \frac{S}{\binom{n}{2}}$$

$$= \left(\frac{S}{\frac{n(n-1)}{2}} \right)$$

$$\text{ดังนั้น } \tau = \frac{2S}{n(n-1)}$$

4.2.2 การแจกแจงสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลลีในกรณี n มีขนาดเล็ก ($n \leq 10$)

เมื่อสนใจความสัมพันธ์ของอันดับของสมาชิก n ตัว นั่นคือจะมีอันดับที่เป็นไปได้เกิดขึ้นเท่ากับ $n!$ การแจกแจงของอันดับที่เกิดขึ้นทุกอันดับจะเหมือนกัน สมมติว่าอันดับที่เกิดขึ้นเป็นอันดับตามธรรมชาติ(natural order) คืออันดับที่มีการเรียงอันดับจากน้อยไปหามาก เช่น จาก 1, 2, 3, ..., n กำหนดให้ $u(n, S)$ คือจำนวนของค่า S ในกลุ่มของอันดับที่จำนวน $n!$

พิจารณาอันดับที่ และผลของการเพิ่มสมาชิกใหม่ ($n+1$) ที่ตำแหน่งต่างๆกันของอันดับ จากตำแหน่งที่ 1 จนถึงตำแหน่งที่ n ซึ่งจะให้ผลดังนี้

1. ถ้าเพิ่มสมาชิกที่ ($n+1$) ที่ตำแหน่งที่ 1 จะเพิ่ม $-n$ ให้กับ S
 2. ถ้าเพิ่มสมาชิกที่ ($n+1$) ที่ตำแหน่งที่ 2 จะเพิ่ม $-(n-2)$ ให้กับ S
 3. ถ้าเพิ่มสมาชิกที่ ($n+1$) ที่ตำแหน่งที่ 3 จะเพิ่ม $-(n-4)$ ให้กับ S
 - $\vdots$
 - n . ถ้าเพิ่มสมาชิกที่ ($n+1$) ที่ตำแหน่งที่ n จะเพิ่ม n ให้กับ S
- นั่นคือจะได้สมการดังต่อไปนี้

$$\begin{aligned} u(n+1, S) = & u(n, S-n) + u(n, S-n+2) + u(n, S-n+4) + \dots \\ & + u(n, S-n-4) + u(n, S-n-2) + u(n, S+n) \end{aligned} \quad (4.32)$$

จากสมการนี้ทำให้ทราบถึงการแจกแจงความถี่ของ S ของ $n+1$ เมื่อทราบค่า n ดังนั้นจะสร้างอนุกรมของการแจกแจงในกรณีที่ย่างๆ เมื่อ $n = 2, 3$ หรืออื่นๆได้ ในทางปฏิบัติวิธีนี้จะง่ายเมื่อ n มีขนาดเล็ก เช่น

กรณีที่ 1 เมื่อ $n = 2$ นั่นคือ S จะมีค่าอยู่ในช่วง $-\frac{1}{2}n(n-1)$ ถึง $\frac{1}{2}n(n-1)$ หรือ -1 ถึง $+1$ พิจารณาได้จาก

$$-\frac{1}{2}n(n-1) = -\frac{1}{2}2(2-1) = -1$$

$$\frac{1}{2}n(n-1) = \frac{1}{2}2(2-1) = 1$$

$$S = \{-1, 1\}$$

ความถี่ของ S คือ $[1, 1]$

กรณีที่ 2 เมื่อ $n = 3$ S จะมีค่าอยู่ในช่วง -3 ถึง $+3$ พิจารณาได้จาก

$$-\frac{1}{2}n(n-1) = -\frac{1}{2}3(3-1) = -3$$

$$\frac{1}{2}n(n-1) = \frac{1}{2}3(3-1) = 3$$

ความถี่ของ S (กรณี $n = 3$) สามารถหาได้โดยการแสดงความถี่ของ S (กรณี $n = 2$) จำนวน 3 ครั้ง เรียงลดหลั่นกันไป โดยเลื่อนตำแหน่งไปทางขวามือ 1 ตำแหน่ง ในแต่ละครั้งดังนี้

$$\begin{array}{cccc}
 1 & & & \\
 & 1 & & \\
 & & 1 & \\
 & & & 1 \\
 \hline
 & & & 1 & 1 \\
 \hline
 \text{ผลรวม} & 1 & 2 & 2 & 1
 \end{array}$$

ผลรวมที่ได้จะเป็น ความถี่ของ S กรณีเมื่อ $n = 3$ และ S จะมีค่าอยู่ในช่วง $-\frac{1}{2}n(n-1)$ ถึง $\frac{1}{2}n(n-1)$ ดังนี้ $S = \{-3, -1, 1, 3\}$

ความถี่ของ S คือ $[1, 2, 2, 1]$

กรณีที่ 3 เมื่อ $n = 4$ S จะมีค่าอยู่ในช่วง -6 ถึง $+6$ พิจารณาได้จาก

$$-\frac{1}{2}n(n-1) = -\frac{1}{2}4(4-1) = -6$$

$$\frac{1}{2}n(n-1) = \frac{1}{2}4(4-1) = 6$$

ผลรวม 1 3 5 6 5 3 1

$\frac{1}{2}n(n-1)$ ดังนั้น $S = \{-6, -3, -1, 0, 1, 3, 6\}$

ความถี่ของ S คือ $[1, 3, 5, 6, 5, 3, 1]$

ឥរូប

- $n=3$ $S = \{-3, -1, 1, 3\}$

1	1										
2	1	1									
3	1	2	2	1							
4	1	3	5	6	5	3	1				
5	1	4	9	15	20	22	20	15	9	4	1

ชุดของตัวเลขในแถวที่ r คือผลรวมของจำนวนสมาชิก $(r-1)$ ตัวที่อยู่ทางด้านซ้ายมือและอยู่เหนือจำนวนนั้น เช่นในแถวที่ 5

$$15 = 6 + 5 + 3 + 1$$

$$20 = 5 + 6 + 5 + 3 + 1$$

$$22 = 3 + 5 + 6 + 5 + 3$$

ความถี่ของ S ที่ $n=r$ หาได้จากจำนวน r แถว เหล่านี้วิธีนี้ใช้ในการหาความถี่และความน่าจะเป็น เมื่อ $n \leq 10$ ในตาราง E ภาคผนวก

คุณสมบัติการแจกแจงของ S

1. การแจกแจงของ S จะสมมาตรสำหรับทุกอันดับ
2. ค่าของ S สอดคล้องกับสังยุคของอันดับที่คือ $-S$
3. ค่าเฉลี่ยของ S จะเท่ากับ ศูนย์
4. S มีค่าอยู่ในช่วง $-\frac{1}{2}n(n-1)$ ถึง $\frac{1}{2}n(n-1)$

4.2.3 การแจกแจงสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ กรณี n มีขนาดใหญ่ ($n > 10$)

ในการพิสูจน์ว่าสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์มีการแจกแจงปกติเมื่อ n มีขนาดเพิ่มขึ้น จำเป็นจะต้องกำหนดความแปรปรวน ดังนั้น

กำหนดให้

$$c_{ij} = a_{ij}b_{ij} \quad (4.33)$$

เมื่อ a_{ij}, b_{ij} คือ คะแนนการเปรียบเทียบของอันดับ

$$\text{และ} \quad c = \sum_{i,j=1}^n c_{ij} \quad (4.34)$$

ดังนั้น

$$c = 2S \quad (4.35)$$

และ

$$\sum_{i=1}^n a_{ii} = n+1-2i \quad (4.36)$$

เมื่อคะแนนเกี่ยวข้องกับสมาชิกที่ i และเป็น $(n-i)-(i-1)$ จะได้

$$\sum_{i=1}^n a_{ii}^2 = n(n-1) \quad (4.37)$$

a^2 เป็น $+1$ ผลบวกจำนวนนี้เป็นเพียงจำนวนวิธีที่เป็นไปได้ทั้งหมด ของการเลือกคู่จาก n

แต่ละคู่จะถูกนับ 2 ครั้ง เช่น ครั้งที่ 1 คือ AB และ ครั้งที่ 2 คือ BA

จากสมการที่ (4.36)

$$\begin{aligned}
 \sum_{i=1}^n a_{ii} &= n+1-2i \\
 \sum_{i=1}^n \sum_{l=1}^n a_{il} &= \sum_{i=1}^n n+1-2i \\
 &= \sum_{i=1}^n (n+1) - \sum_{i=1}^n 2i \\
 &= n(n+1) - 2 \sum_{i=1}^n i \\
 &= n(n+1) - 2(1+2+3+\dots+n) \\
 &= n(n+1) - 2 \frac{n(n+1)}{2}
 \end{aligned}$$

ดังนั้นจะได้ว่า $\sum_{i=1}^n \sum_{l=1}^n a_{il} = 0$ (4.38)

ในทำนองเดียวกัน

$$\sum_{i,l,t=1}^n a_{il} a_{it} = \sum_{i,l=1}^n a_{il} (n+1-2i)$$

จาก สมการที่ (4.36)

$$\sum_{i=1}^n a_{il} = (n+1-2i)$$

ดังนั้นจะได้ $\sum_{i,l,t=1}^n a_{il} a_{it} = \sum_{i=1}^n (n+1-2i)^2$

$$\begin{aligned}
 \sum_{i,l,t=1}^n a_{il} a_{it} &= \sum_{i=1}^n [(n+1)^2 - 2(n+1)(2i) + (2i)^2] \\
 &= \sum_{i=1}^n (n+1)^2 - 4(n+1) \sum_{i=1}^n i + 4 \sum_{i=1}^n i^2
 \end{aligned}$$

$$\begin{aligned}
&= n(n+1)^2 - \frac{4(n+1)n(n+1)}{2} + \frac{4n(n+1)(2n+1)}{6} \\
&= n(n+1)\left[(n+1) - \frac{4(n+1)}{2} + \frac{4(2n+1)}{6}\right] \\
&= n(n+1)\left[\frac{6(n+1) - 12(n+1) + 4(2n+1)}{6}\right] \\
&= n(n+1)\left[\frac{6n + 6 - 12n - 12 + 8n + 4}{6}\right] \\
&= n(n+1)\frac{(n-1)}{3}
\end{aligned}$$

$$\sum_{i,t=1}^n a_{it}a_{it} = \frac{1}{3}n(n^2-1) \quad (4.39)$$

จาก สมการที่ (4.33)

$$c = \sum_{i,j=1}^n a_{ij}b_{ij}$$

ค่าคาดหวังของ c

$$\begin{aligned}
E(c) &= E\sum_{i,j=1}^n a_{ij}b_{ij} \\
&= \sum_{i,j=1}^n E(a_{ij}b_{ij}) \quad (4.40)
\end{aligned}$$

เนื่องจาก a_{ij}, b_{ij} เป็นอิสระต่อกัน และจากสมการที่ (4.38) $\sum_{i,j=1}^n a_{ij} = 0$ ดังนั้นจะได้ว่า

$$E(c) = E\left(\sum_{i,j=1}^n a_{ij}\right)E\left(\sum_{i,j=1}^n b_{ij}\right) \quad (4.35)$$

$$= 0$$

นั่นคือค่าเฉลี่ยของ c หรือของ S มีค่าเป็น ศูนย์

ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลด์ $E(\tau)$

พิจารณาสมการ (4.35)

$$c = 2S$$

จะได้ว่า $E(c) = 2E(S) = 0$

นั่นคือจากสมการที่ (3.14)

$$\tau = \frac{2S}{n(n-1)}$$

ดังนั้น

$$E(\tau) = \frac{2E(S)}{n(n-1)} = 0 \quad (4.42)$$

ต่อไปจะหาค่าความแปรปรวน

จาก $V(c) = E(c^2) - [E(c)]^2$

$$\begin{aligned} E(c^2) &= E\left[\sum_{i,j=1}^n a_{ij}b_{ij}\right]^2 \\ &= E\left[\sum_{i,j=1}^n (a_{ij}^2b_{ij}^2) + \sum_{i,j=1}^n (a_{ij}b_{ij}a_{ik}b_{ik}) + \sum_{i,j=1}^n (a_{ij}b_{ij}a_{kl}b_{kl})\right]^2 \end{aligned} \quad (4.43)$$

เมื่อ $i \neq j \neq k = 1$ ซึ่งผลบวกของเทอมนี้จะบวกที่เทอมเฉพาะ แต่ทุกๆเทอมอาจจะเกิดได้

มากกว่า 1 วิธี จากการขยายเทอม $(\sum a_{ij}b_{ij})^2$

พิจารณาเทอมของ $E(c^2)$ ได้เป็น 3 กรณี

กรณีที่ 1 พิจารณากรณีเทอมที่อยู่ในรูปของ $\sum .$

การแสดงว่าเทอม $\sum .$ หายไป เป็นดังนี้

$$\begin{aligned}
\sum a_{ij}a_{kl} &= \sum 'a_{ij}a_{kl} + \sum 'a_{ij}a_{li} + \sum 'a_{ij}a_{ki} + \sum 'a_{ij}a_{jl} + \sum 'a_{ij}a_{kj} \\
&\quad + \sum 'a_{ij}a_{ij} + \sum 'a_{ij}a_{ji} \\
a_{ij} &= -a_{ji}
\end{aligned} \tag{4.44}$$

จากสมการที่ (4.38) $\sum_{i,j=1}^n a_{ij} = 0$

เมื่อ a_{ij} และ a_{kl} เป็นอิสระต่อกัน จะได้ว่า

$$E \sum a_{ij}a_{kl} = E \sum a_{ij} \sum a_{kl}$$

ดังนั้นค่าคาดหวังของสมการที่ (4.44) เป็น

$$E \sum a_{ij}a_{kl} = 0 \text{ เพราะ } \sum_{i,j}^n a_{ij} = 0$$

กรณีที่ 2 พิจารณา $E \sum_{i,j}^n a_{ij}^2 b_{ij}^2$ ซึ่งมาจากกำลัง 2 ของ $a_{ij}b_{ij}$ หรือผลคูณของเทอม $a_{ij}b_{ij}$ ดังนั้นผลรวมที่ได้ก็จะเป็นผลรวม 2 ครั้งของ $a_{ij}^2 b_{ij}^2$ ซึ่ง $i, j = 1, 2, 3, \dots, n$ และผลรวมที่ได้ก็จะมี $n(n-1)$ เทอมและค่าคาดหวังก็จะเป็น

$$\begin{aligned}
2n(n-1)E(a_{ij}^2 b_{ij}^2) &= 2n(n-1)E(a_{ij}^2)E(b_{ij}^2) \\
&= \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2
\end{aligned} \tag{4.45}$$

กรณีที่ 3 ในทำนองเดียวกันเทอมที่ 2 ทางขวามือจะได้

$$\frac{4}{n(n-1)(n-2)} \sum 'a_{ij}a_{ik} \sum 'b_{ij}b_{ik} \tag{4.46}$$

สำหรับผลคูณของ $a_{ij}, a_{ik}, b_{ij}, b_{ik}$ ทำได้ 4 วิธีเพราะว่าถ้ากำหนด i, j, k สำหรับเทอมของ a และ b จะได้ Suffix $(ij, ik) (ji, ik) (ij, ki) (ji, ki)$ และนั่นคือ $n(n-1)(n-2)$ วิธี ของการกำหนด suffix ที่แตกต่างกันของ n

$$\sum 'a_{ij}a_{ik} = \sum a_{ij}a_{ik} - \sum a_{ij}^2 \quad (4.47)$$

สมการที่ (4.45), (4.46) และ (4.47) แทนใน (4.43) จะได้

$$\begin{aligned} E(c^2) &= \frac{4}{n(n-1)(n-2)} \left\{ \sum 'a_{ij}a_{ik} \sum 'b_{ij}b_{ik} \right\} + \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \sum a_{ij}a_{ik} - \sum a_{ij}^2 \right\} \left\{ \sum b_{ij}b_{ik} - \sum b_{ij}^2 \right\} \\ &\quad + \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2 \end{aligned} \quad (4.48)$$

ความแปรปรวนสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ $V(\tau)$

จากสมการที่ (4.48)

$$\begin{aligned} E(c^2) &= \frac{4}{n(n-1)(n-2)} \left\{ \sum 'a_{ij}a_{ik} \sum 'b_{ij}b_{ik} \right\} + \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \sum a_{ij}a_{ik} - \sum a_{ij}^2 \right\} \left\{ \sum b_{ij}b_{ik} - \sum b_{ij}^2 \right\} + \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2 \end{aligned}$$

สมการที่ (4.37)

$$\sum_{i,l}^n a_{il}^2 = n(n-1)$$

และ สมการที่ (4.39)

$$\sum_{i,l,t=1}^n a_{il}a_{it} = \frac{1}{3}n(n^2-1)$$

แทนสมการที่ (4.37) และ (4.39) ในสมการที่ (4.48)

$$\begin{aligned}
 E(c^2) &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{3} n(n^2-1) - n(n-1) \right\} \left\{ \frac{1}{3} n(n^2-1) - n(n-1) \right\} \\
 &\quad + \frac{2}{n(n-1)} [n(n-1)]^2 \\
 &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{3} n(n^2-1) - n(n-1) \right\}^2 + \frac{2}{n(n-1)} [n(n-1)]^2 \\
 &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{3} n(n-1)(n+1) - n(n-1) \right\}^2 + 2n(n-1) \\
 &= \frac{4}{n(n-1)(n-2)} \left\{ n(n-1) \left[\frac{1}{3} (n+1) - 1 \right] \right\}^2 + 2n(n-1) \\
 &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{3} n(n-1)(n-2) \right\}^2 + 2n(n-1) \\
 &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{9} n^2 (n-1)^2 (n-2)^2 \right\} + 2n(n-1) \\
 &= \frac{4}{9} n(n-1)(n-2) + 2n(n-1) \\
 &= 2n(n-1) \left[\frac{2}{9} (n-2) + 1 \right] \\
 &= 2n(n-1) \left[\frac{2n-4+9}{9} \right] \\
 &= \frac{2n(n-1)(2n+5)}{9} \tag{4.49}
 \end{aligned}$$

จากสมการที่ (4.35)

$$c = 2S \text{ แทนในสมการที่ (4.49)}$$

$$E(c^2) = E(4S^2)$$

$$E(4S^2) = \frac{2n(n-1)(2n+5)}{9} \quad (4.50)$$

$$V(S) = E(S^2) = \frac{n(n-1)(2n+5)}{18} \quad (4.51)$$

จาก สมการที่ (3.14)

$$\tau = \frac{2S}{n(n-1)}$$

$$S = \frac{n(n-1)\tau}{2} \quad (4.52)$$

ความแปรปรวนของ S

$$\begin{aligned} V(S) &= V\left[\frac{n(n-1)\tau}{2}\right] \\ &= \frac{n^2(n-1)^2 V(\tau)}{4} \end{aligned} \quad (4.53)$$

$$\begin{aligned} V(\tau) &= \frac{4V(S)}{n^2(n-1)^2} \\ &= \frac{4}{n^2(n-1)^2} \times \frac{n(n-1)(2n+5)}{18} \end{aligned}$$

ดังนั้น ความแปรปรวนของ τ คือ

$$V(\tau) = \frac{2(2n+5)}{9n(n-1)} \quad (4.54)$$

ค่าความคลาดเคลื่อนมาตรฐานคือ (3.22)

$$S(\tau) = \sqrt{\frac{2(2n+5)}{9n(n-1)}}$$

ดังนั้นจะได้ว่าสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ (τ) กรณีที่ $n > 10$

มีการแจกแจงแบบปกติที่มี ค่าเฉลี่ย $E(\tau) = 0$

$$\text{และความแปรปรวน } V(\tau) = \frac{2(2n+5)}{9n(n-1)}$$

4.2.4 พิสูจน์สูตร การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลด์
กรณีที่ $n > 10$

$$Z = \frac{\tau \sqrt{9n(n-1)}}{\sqrt{2(2n+5)}}$$

กำหนดให้ τ เป็นตัวแปรสุ่ม

จาก ทฤษฎีบทที่ 2.4 ทฤษฎีลิมิตส่วนกลาง

เมื่อ $n > 10$ จะเห็นว่าขนาดตัวอย่างใหญ่เพียงพอที่จะใช้ทฤษฎีลิมิตส่วนกลางได้

ดังนั้น τ จะมีการแจกแจงเข้าสู่การแจกแจงแบบปกติไม่ว่าการแจกแจงของ τ จะเป็นอย่างไร

ค่าเฉลี่ย $E(\tau) = 0$

ความแปรปรวน $V(\tau) = \frac{2(2n+5)}{9n(n-1)}$

หรืออีกนัยหนึ่งตัวแปรสุ่ม τ จะมีการแจกแจงโดยประมาณแบบปกติมาตรฐาน $N(0, 1)$

ดังนั้นเมื่อต้องการทดสอบสมมติฐาน

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0$$

สถิติที่ใช้ในการทดสอบคือ

$$\begin{aligned} Z &= \frac{\tau - E(\tau)}{\sqrt{V(\tau)}} \\ &= \frac{\tau - 0}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \\ &= \tau \sqrt{\frac{9n(n-1)}{2(2n+5)}} \end{aligned}$$

4.3 การแจกแจงสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน

ในการทดสอบสมมติฐาน หรือการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนนั้น จะต้องทราบการแจกแจงของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน โดยการศึกษาจากข้อมูลตัวอย่างเสียก่อน จึงจะทำการทดสอบสมมติฐานหรือการประมาณค่าพารามิเตอร์ได้ ในที่นี้สนใจที่จะศึกษาสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนในรายละเอียดดังต่อไปนี้

- การแจกแจงสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน กรณี n มีขนาดเล็ก
- การแจกแจงสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน กรณี n มีขนาดใหญ่
- พิสูจน์สูตรการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน

4.3.1 การแจกแจงสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน กรณี n มีขนาดเล็ก

การแจกแจงของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนยากในการพิจารณามากกว่าการแจกแจงของสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลลี ต่อไปจะพิจารณาชุดตัวเลขต่อไปนี้

$$\begin{array}{ccccccc}
 0 & 1 & 2 & \dots & (n-3) & (n-2) & (n-1) \\
 -1 & 0 & 1 & \dots & (n-4) & (n-3) & (n-2) \\
 -2 & -1 & 0 & \dots & (n-5) & (n-4) & (n-3) \\
 & & & \vdots & & & \\
 -(n-2) & -(n-3) & -(n-4) & \dots & -1 & 0 & 1 \\
 -(n-1) & -(n-2) & -(n-3) & \dots & -2 & -1 & 0
 \end{array} \quad (4.55)$$

เซตของค่าเบี่ยงเบนระหว่างอันดับที่ $1, 2, 3, \dots, n$ และอันดับ ที่ไม่ถูกเจาะจงถูกกำหนดโดยการเลือกสมาชิกจากชุดของตัวเลข (array) เพื่อไม่ให้สมาชิก 2 ตัว ปรากฏมากกว่า 1 แถวหรือ คอลัมน์พิจารณาตัวเลขต่อไปนี้

$$\begin{array}{ccccccc}
 a^0 & a^1 & a^4 & \dots & a^{(n-3)^2} & a^{(n-2)^2} & a^{(n-1)^2} \\
 a^1 & a^0 & a^1 & \dots & a^{(n-4)^2} & a^{(n-3)^2} & a^{(n-2)^2} \\
 & & & \vdots & & & \\
 & & & & a^{(n-5)^2} & a^{(n-4)^2} & a^{(n-3)^2} \\
 & & & \vdots & & & \\
 a^{(n-2)^2} & a^{(n-3)^2} & a^{(n-4)^2} & \dots & a^1 & a^0 & a^1 \\
 a^{(n-1)^2} & a^{(n-2)^2} & a^{(n-3)^2} & \dots & a^4 & a^1 & a^0
 \end{array} \quad (4.56)$$

จะเห็นว่าเลขชี้กำลังในสมการที่ (4.56) คือกำลัง 2 ของสมการที่ (4.55) ถ้ากระจายสมการที่ (4.56) ก็จะได้ผลรวมของ $\sum d^2$ โดยการกระจาย วิธีการนี้ใช้ในการหาการแจกแจงความน่าจะเป็นเมื่อ $n \leq 25$

คุณสมบัติของการแจกแจงของ $\sum d^2$

1. ทุกค่าของ $\sum d^2$ ต้องเท่ากับเมื่อ $\sum d = 0$
2. การแจกแจงนี้จะสมมาตรสำหรับบางค่าของ $\sum d^2$ ที่สอดคล้องกับค่า $\frac{1}{3}(n^3 - n) - \sum d^2$ ซึ่งได้มาจากสังยุคของอันดับ
3. ค่าเฉลี่ยของการแจกแจงคือ $\frac{1}{6}(n^3 - n)$
4. $\sum d^2$ จะมีค่าอยู่ในช่วง 0 ถึง $\frac{1}{3}(n^3 - n)$

4.3.2 การแจกแจงสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน กรณี n มีขนาดใหญ่

ในการพิสูจน์ว่าสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนมีการแจกแจงปกติเมื่อ n มีขนาดเพิ่มขึ้น ดังนั้นจำเป็นจะต้องคำนวณ ค่าเฉลี่ย และ ความแปรปรวน ดังนี้

กำหนดให้ $a'_{ij} = j - i$ (4.57)

จะได้ว่า $\sum_{j=1}^n a'_{ij} = \frac{1}{2}n(n-1-2i)$ (4.58)

ซึ่งได้มาจากการหาผลบวกอนุกรมเลขคณิต

$$\begin{aligned} \sum_{j=1}^n a'_{ij} &= \sum_{j=1}^n (j-i) \\ &= (1-i) + (2-i) + (3-i) + \dots + (n-i) \end{aligned}$$

เป็นอนุกรมเลขคณิต ที่มี

$$a_1 = (1-i) \text{ และ}$$

$$d = (2-i) - (1-i) = 2-i-1+i = 1$$

หาผลบวกอนุกรม

$$S_n = \frac{n}{2}[2a_1 + (n-1)d]$$

$$= \frac{n}{2}[2(1-i) + (n-1)1]$$

$$= \frac{n}{2}(2-2i+n-1)$$

$$S_n = \sum_{j=1}^n a'_{ij} = \frac{n}{2}(n+1-2i)$$

$$\text{และ } \sum_{i,j=1}^n a'^2_{ij} = \sum (j-i)^2$$

$$= \sum_{i,j=1}^n (j^2 - 2ij + i^2)$$

$$= \sum_{i,j=1}^n j^2 - 2 \sum_{i,j=1}^n ij + \sum_{i,j=1}^n i^2$$

$$= \sum_{i=1}^n (1^2 + 2^2 + 3^2 + \dots + n^2) - \sum_{j=1}^n (1^2 + 2^2 + 3^2 + \dots + n^2) - 2 \sum_{i=1}^n i \sum_{j=1}^n j$$

$$(\text{และ } \sum_{i=1}^n i = \sum_{j=1}^n j)$$

$$= n \sum_{j=1}^n j^2 - n \sum_{i=1}^n i^2 - 2 \sum_{i=1}^n i \sum_{j=1}^n j$$

$$\text{จาก } \sum_{i=1}^n i = \sum_{j=1}^n j \text{ ดังนั้น } \sum_{i=1}^n i^2 = \sum_{j=1}^n j^2 \text{ ด้วย}$$

$$\begin{aligned}
\text{ดังนั้น } \sum_{i,j=1}^n a'_{ij}{}^2 &= 2n \sum_{j=1}^n j^2 - 2 \left(\sum_{j=1}^n j \right)^2 \\
&= 2n \frac{n(n+1)(2n+1)}{6} - 2 \left[\frac{n(n+1)}{2} \right]^2 \\
&= \frac{2n^2(n+1)(2n+1)}{6} - \frac{3n^2(n+1)^2}{6} \\
&= \frac{n^2(n+1)}{6} [2(2n+1) - 3(n+1)] \\
&= \frac{n^2(n+1)}{6} [4n+2-3n-3] \\
&= \frac{n^2(n+1)}{6} (n-1) \\
\sum_{i,j}^n a'_{ij}{}^2 &= \frac{n^2(n^2-1)}{6} \tag{4.59}
\end{aligned}$$

จากสมการที่ (4.58)

$$\sum_{j=1}^n a'_{ij} = \frac{1}{2} n(n+1-2i)$$

ดังนั้นจะได้

$$\begin{aligned}
\sum_{i,j,k=1}^n a'_{ij} a'_{ik} &= \sum_{i,j=1}^n a'_{ij} \left[\frac{n}{2} (n+1-2i) \right] \\
&= \sum_{i=1}^n \left[\frac{n}{2} (n+1-2i) \right]^2 \\
&= \sum_{i=1}^n \left[\frac{n^2}{4} (n+1-2i)^2 \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[\frac{n^2}{4} \{ (n+1)^2 - 4(n+1)i + 4i^2 \} \right] \\
&= \sum_{i=1}^n \left[\frac{1}{4} \{ n^2(n+1)^2 - 4n^2(n+1)i + 4n^2i^2 \} \right] \\
&= \frac{1}{4} \left[\sum n^2(n+1)^2 - \sum 4n^2(n+1)i + \sum 4n^2i^2 \right] \\
&= \frac{1}{4} \left[n^3(n+1)^2 - 4n^2(n+1) \sum i + 4n^2 \sum i^2 \right] \\
&= \frac{1}{4} \left[n^3(n+1)^2 - \frac{4n^2(n+1)n(n+1)}{2} + \frac{4n^2n(n+1)(2n+1)}{6} \right] \\
&= \frac{n^3(n+1)}{4} \left[(n+1) - \frac{4(n+1)}{2} + \frac{4(2n+1)}{6} \right] \\
&= \frac{n^3(n+1)}{4} \left[6(n+1) - \frac{12(n+1)}{6} + \frac{4(2n+1)}{6} \right] \\
&= \frac{n^3(n+1)}{4} \left[\frac{6n+6-12n-12+8n+4}{6} \right] \\
&= \frac{n^3(n+1)}{4} \times \frac{(n-1)}{3}
\end{aligned}$$

$$\sum_{i,j,k=1}^n a'_{ij} a'_{ik} = \frac{1}{12} n^3 (n^2 - 1) \quad (4.60)$$

แทน สมการที่ (4.59) และ (4.60)

$$\sum_{i,j}^n a'_{ij}{}^2 = \frac{n^2(n^2-1)}{6}$$

$$\sum_{i,j,k=1}^n a'_{ij}a'_{ik} = \frac{1}{12}n^3(n^2-1)$$

ในสมการที่ (4.48) จะได้ว่า

$$\begin{aligned} E(c^2) &= \frac{4}{n(n-1)(n-2)} \left\{ \sum a_{ij}a_{ik} - \sum a_{ij}^2 \right\} \left\{ \sum b_{ij}b_{ik} - \sum b_{ij}^2 \right\} + \frac{2}{n(n-1)} \sum a_{ij}^2 \sum b_{ij}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{12}n^3(n^2-1) \right\}^2 + \frac{2}{n(n-1)} \left\{ \frac{1}{6}n^2(n^2-1) \right\}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{12}n^3(n^2-1) - \frac{1}{6}n^2(n^2-1) \right\}^2 + \frac{2}{n(n-1)} \left\{ \frac{1}{6}n^2(n^2-1) \right\}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{6}n^2(n^2-1) \left[\frac{n}{2} - \frac{2}{2} \right] \right\}^2 + \frac{2}{n(n-1)} \left\{ \frac{1}{6}n^2(n^2-1) \right\}^2 \\ &= \frac{4}{n(n-1)(n-2)} \left\{ \frac{1}{6}n^2(n^2-1) \frac{(n-2)}{2} \right\}^2 + \frac{2}{36}n^3(n-1)(n+1)^2 \\ &= \left[\frac{4}{n(n-1)(n-2)} \right] \left[\frac{n^4(n-1)^2(n+1)^2}{36} \right] \left[\frac{(n-2)^2}{4} \right] + \frac{2}{36}n^3(n-1)(n+1)^2 \\ &= \left[\frac{n^3(n-1)(n+1)^2(n-2)}{36} \right] + \frac{2}{36}n^3(n-1)(n+1)^2 \\ &= \left[\frac{n^3(n-1)(n+1)^2}{36} \right] [(n-2) + 2] \\ &= \left[\frac{n^4(n-1)(n+1)^2}{36} \right] \end{aligned}$$

$$E(c^2) = \frac{n^4(n-1)(n+1)^2}{36} \quad (4.61)$$

จากสมการที่ (4.3) (4.33) และ (4.34) จะได้ว่า

$$c = \sum c_{ij}$$

$$c = \sum a_{ij}b_{ij}$$

$$c = \sum_{i,j} (p_j - p_i)(q_j - q_i)$$

จากสมการที่ (4.12) จะได้ว่า

$$c = \frac{1}{6}n^2(n^2-1) - n \sum d^2$$

c ยกกำลัง 2

$$c^2 = \frac{1}{36}n^4(n^2-1)^2 - n^2(\sum d^2)^2 \quad (4.62)$$

หาค่าคาดหวัง

$$E(c^2) = \frac{1}{36}n^4(n^2-1)^2 - n^2E(\sum d^2)^2 \quad (4.63)$$

ให้สมการที่ (4.61) = (4.63)

$$\frac{n^4(n-1)(n+1)^2}{36} = \frac{1}{36}n^4(n^2-1)^2 - n^2E(\sum d^2)^2$$

$$n^2E(\sum d^2)^2 = \frac{n^4(n+1)^2(n-1)^2 - n^4(n-1)(n+1)^2}{36}$$

$$E(\sum d^2)^2 = \frac{n^2(n+1)^2(n-1)^2 - n^2(n-1)(n+1)^2}{36}$$

$$= \frac{n^2(n+1)^2(n-1)}{36} [(n-1)-1]$$

$$= \frac{n^2(n+1)^2(n-1)(n-2)}{36} \quad (4.64)$$

ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์อันดับที่ของสเปียร์แมน

$$\begin{aligned}
 E(r_s) &= E\left[1 - \frac{6\sum d^2}{n(n^2-1)}\right] \\
 &= E(1) - \frac{6E(\sum d^2)}{n(n^2-1)} \quad (4.65)
 \end{aligned}$$

พิจารณาเทอมของ $E(\sum d^2)$

จาก สมการที่ (4.12)

$$c = \frac{1}{6}n^2(n^2-1) - n\sum d^2$$

และ $E(c) = 0$ จะได้ว่า

$$E(c) = E\left[\frac{1}{6}n^2(n^2-1) - n\sum d^2\right] = 0 \quad (4.66)$$

$$\frac{1}{6}n^2(n^2-1) - nE(\sum d^2) = 0$$

$$E(\sum d^2) = \frac{1n^2(n^2-1)}{6n} \quad (4.67)$$

สมการที่ (4.67) แทนในสมการที่ (4.65)

$$E(r_s) = E(1) - \frac{6\left[\frac{n^2(n^2-1)}{6n}\right]}{n(n^2-1)}$$

$$= 1 - 1$$

$$= 0$$

ค่าความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์อันดับที่สี่เพียร์แมน

จาก

$$V(X) = E(X^2) - [E(X)]^2$$

ดังนั้น

$$V(\sum d^2) = E(\sum d^2)^2 - [E(\sum d^2)]^2 \quad (4.68)$$

แทนสมการที่ (4.64) และ (4.67) ในสมการที่ (4.68) จะได้ว่า

$$\begin{aligned} V(\sum d^2) &= \frac{n^2(n-1)(n-2)(n+1)^2 - n^2(n-1)^2(n+1)^2}{36} \\ &= \frac{n^2(n-1)(n+1)^2}{36} [(n-2) - (n-1)] \\ &= \frac{n^2(n-1)(n+1)^2}{36} [(n-2-n+1)] \\ &= -\frac{n^2(n-1)(n+1)^2}{36} \quad (4.69) \end{aligned}$$

$$\text{จาก } r_s = 1 - \frac{6 \sum d^2}{n^3 - n}$$

$$\begin{aligned} V(r_s) &= V(1) - \frac{36V(\sum d^2)}{(n^3 - n)^2} \\ &= \frac{36n^2(n-1)(n+1)^2}{36(n^3 - n)^2} \end{aligned}$$

$$= \frac{n^2(n-1)(n+1)^2}{[n(n+1)(n-1)][n(n+1)(n-1)]}$$

ดังนั้น

$$V(r_s) = \frac{1}{n-1}$$

ดังนั้นจะได้ว่าสมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน กรณีที่ $n > 10$

มีการแจกแจงแบบปกติที่มี ค่าเฉลี่ย $E(r_s) = 0$

และความแปรปรวน $V(r_s) = \frac{1}{n-1}$

4.3.3 พิสูจน์สูตร ในการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน

กรณีที่ $n > 25$ ให้ใช้สูตร $Z = r_s \sqrt{n-1}$

วิธีทำ

กำหนดให้ r_s เป็นตัวแปรสุ่ม

จาก ทฤษฎีบทที่ 2.4 ทฤษฎีลิมิตสู่ส่วนกลาง

เมื่อ $n > 25$ จะเห็นว่าขนาดตัวอย่างใหญ่เพียงพอที่จะใช้ทฤษฎีลิมิตสู่ส่วนกลางได้

ดังนั้น r_s จะมีการแจกแจงเข้าสู่การแจกแจงแบบปกติไม่ว่าการแจกแจงของ r_s จะเป็นอย่างไร

ค่าเฉลี่ย $E(r_s) = 0$

ความแปรปรวน $V(r_s) = \frac{1}{n-1}$

หรืออีกนัยหนึ่งตัวแปรสุ่ม r_s จะมีการแจกแจงโดยประมาณแบบปกติมาตรฐาน $N(0, 1)$

เมื่อ สมมติฐานการทดสอบเป็น

$H_0 : \rho_s = 0$

$H_0 : \rho_s \neq 0$

สถิติทดสอบคือ

$$Z = \frac{r_s - \mu_{r_s}}{\sigma_{r_s}}$$

$$Z = \frac{r_s - 0}{\sqrt{\frac{1}{n-1}}}$$

$$Z = r_s \sqrt{n-1}$$

4.4 การพิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่ส่วนย่อยเคนคอลล์

จากหัวข้อที่ผ่านมาเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวว่ามีความสัมพันธ์กันอย่างไร โดยไม่ได้คำนึงว่าความสัมพันธ์ที่เกิดขึ้นระหว่างตัวแปรทั้ง 2 นั้นเป็นผลมาจากตัวแปรอื่นอีกหรือไม่ ในที่นี้สนใจศึกษาสถิติที่ใช้หาความสัมพันธ์ของตัวแปรมาตราเรียงอันดับ ที่สามารถนำตัวแปรอื่นที่เกี่ยวข้องมาควบคุมให้คงที่ได้ และสถิตินั้นก็คือสัมประสิทธิ์สหสัมพันธ์อันดับที่ส่วนย่อยเคนคอลล์ ซึ่งมีรูปแบบสูตรการคำนวณหลายรูปแบบดังจะได้ศึกษาต่อไปนี้

สูตรที่เกี่ยวข้อง

- $\tau_{xy.z} = \frac{\tau_{xy} - \tau_{xz}\tau_{yz}}{\sqrt{(1-\tau_{xz}^2)(1-\tau_{yz}^2)}}$
- $\tau_{xy.z} = \sqrt{\frac{\chi^2}{N}}$
- $\tau_{xy.z} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$
- $r_{ij.k} = \frac{r_{ij} - (r_{ik})(r_{jk})}{\sqrt{(1-r_{ik}^2)(1-r_{jk}^2)}}$

4.4.1 พิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่ส่วนย่อยเคนคอลล์

พิสูจน์สูตร

$$\tau_{xy.z} = \frac{\tau_{xy} - \tau_{xz}\tau_{yz}}{\sqrt{(1-\tau_{xz}^2)(1-\tau_{yz}^2)}}$$

พิจารณาทารางการจำแนกแบบ 2 ทาง ขนาด 2x2

คุณลักษณะ X	คุณลักษณะ X		รวม
	Y_1	Y_2	
X_1	a	b	a+b
X_2	c	d	c+d
รวม	a+c	b+d	$N=a+b+c+d$

จะได้ว่า

$$t_{xy} = \frac{(a+d)-(b+c)}{N} \quad (4.70)$$

$$t_{xz} = \frac{(a+b)-(c+d)}{N} \quad (4.71)$$

$$t_{yz} = \frac{(a+c)-(b+d)}{N} \quad (4.72)$$

$$N = a+b+c+d \quad (4.73)$$

จาก สมการที่ (4.70) t_{xy} ยกกำลังสอง

$$t_{xy}^2 = \left[\frac{(a+d)-(b+c)}{N} \right]^2$$

$$t_{xy}^2 = \frac{1}{N^2} [(a+d)-(b+c)]^2$$

$$t_{xy}^2 = \frac{1}{N^2} [(a+d)^2 - 2(a+d)(b+c) + (b+c)^2]$$

$$t_{xy}^2 = \frac{1}{N^2} [a^2 + 2ad + d^2 - 2(ab+ac+bd+cd) + b^2 + 2bc + c^2]$$

$$t_{xy}^2 = \frac{1}{N^2} [a^2 + b^2 + c^2 + d^2 + 2ad + 2bc - 2(ab+ac+bd+cd)]$$

$$1 - t_{xy}^2 = 1 - \left\{ \frac{1}{N^2} [a^2 + b^2 + c^2 + d^2 + 2ad + 2bc - 2(ab+ac+bd+cd)] \right\}$$

$$\left\{ \frac{[a^2 + b^2 + c^2 + d^2 + 2ad + 2bc - 2(ab+ac+bd+cd)]}{N^2} \right\} = 1 - (1 - t_{xy}^2) \quad (4.74)$$

จาก

$$N = (a+b+c+d)$$

$$N^2 = (a+b+c+d)(a+b+c+d)$$

$$N^2 = a^2 + b^2 + c^2 + d^2 + 2(ab+ac+ad+bc+bd+cd) \quad (4.75)$$

และ

$$(a+d)(b+c) = (ab+ac+bd+cd)$$

บวก $\frac{4(ab+ac+bd+cd)}{N^2}$ ทั้ง 2 ข้างของสมการ (4.74) จะได้

$$\left\{ \frac{[a^2 + b^2 + c^2 + d^2 + 2ad + 2bc - 2(ab+ac+bd+cd)]}{N^2} + \frac{4(ab+ac+bd+cd)}{N^2} \right\} \quad (4.76)$$

$$= \left\{ 1 - (1-t_{xy}^2) + \left[\frac{4(ab+ac+bd+cd)}{N^2} \right] \right\}$$

$$\frac{N^2}{N^2} = \left[1 - (1-t_{xy}^2) + \left[\frac{4(ab+ac+bd+cd)}{N^2} \right] \right]$$

$$[1 - 1 + (1-t_{xy}^2)] = \left[\frac{4(ab+ac+bd+cd)}{N^2} \right]$$

$$1 - t_{xy}^2 = \frac{4}{N^2} (a+d)(b+c) \quad (4.77)$$

ในทำนองเดียวกันก็จะได้

$$1 - t_{xz}^2 = \frac{4}{N^2} (a+b)(c+d) \quad (4.78)$$

$$1 - t_{yz}^2 = \frac{4}{N^2} (a+c)(b+d) \quad (4.78)$$

สมการ (4.70)-[(4.71) x (4.72)] ;

$$t_{xy} - t_{xz}t_{yz} = \left[\frac{(a+d) - (b+c)}{N} \right] - \left[\frac{\langle (a+b) - (c+d) \rangle \langle (a+c) - (b+d) \rangle}{N^2} \right]$$

$$t_{xy} - t_{xz}t_{yz} = \frac{1}{N^2} \left[\begin{aligned} & (a+b+c+d) \langle (a+d) - (b+c) \rangle - \\ & \langle (a+b) - (c+d) \rangle \langle (a+c) - (b+d) \rangle \end{aligned} \right]$$

$$= \frac{4}{N^2} (ad - bc) \quad (4.80)$$

สมการที่ (4.74)x(4.76);

$$(1-t_{xz}^2)(1-t_{yz}^2) = \left(\frac{4}{N^2}\right)^2 (a+b)(a+c)(b+d)(c+d)$$

$$\sqrt{(1-t_{xz}^2)(1-t_{yz}^2)} = \frac{4}{N^2} \sqrt{(a+b)(a+c)(b+d)(c+d)} \quad (4.81)$$

สมการที่ (4.80)/(4.81);

$$\frac{t_{xy} - t_{xz}t_{yz}}{\sqrt{(1-t_{xz}^2)(1-t_{yz}^2)}} = \frac{\frac{4}{N^2} (ad-bc)}{\frac{4}{N^2} \sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

$$\frac{t_{xy} - t_{xz}t_{yz}}{\sqrt{(1-t_{xz}^2)(1-t_{yz}^2)}} = \frac{(ad-bc)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (4.82)$$

จากสมการที่ (3.25)

$$\tau_{xy,z} = \frac{(ad-bc)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

ดังนั้นด้วยคุณสมบัติการถ่ายทอด จะได้ว่าสมการที่ (4.82) และ สมการ (3.25) เท่ากัน

นั่นคือ

$$\tau_{xy,z} = \frac{t_{xy} - t_{xz}t_{yz}}{\sqrt{(1-t_{xz}^2)(1-t_{yz}^2)}}$$

4.4.2 พิสูจน์สูตรสัมประสิทธิ์สหสัมพันธ์อันดับที่ส่วนย่อยเคนคอลล์สมการที่ 3.24 และ 3.25

พิสูจน์สูตร

$$\tau_{xy,z} = \sqrt{\frac{\chi^2}{N}}$$

$$\tau_{xy,z} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$$

พิจารณารางการจำแนกแบบ 2 ทาง ขนาด 2x2

คุณลักษณะ X	คุณลักษณะ X		รวม
	Y ₁	Y ₂	
X ₁	a	b	a+b
X ₂	c	d	c+d
รวม	a+c	b+d	N= a+b+c+d

$$P(AB) = P(A)P(B)$$

$$\text{ความน่าจะเป็นที่จะเกิด } X_1 : P(X_1) = \frac{a+b}{N}$$

$$\text{ความน่าจะเป็นที่จะเกิด } Y_1 : P(Y_1) = \frac{a+c}{N}$$

พิจารณาเหตุการณ์ $X_1 \cap Y_1$ จงหาความน่าจะเป็นที่เกิด $X_1 \cap Y_1$ เพื่อจะหาเกี่ยวกับ O_{11} ความสมมุติฐานที่ว่าตัวแปร X และตัวแปร Y เป็นอิสระกันจะได้ว่า

$$P(X_1 \cap Y_1) = P(X_1)P(Y_1) = \frac{(a+b)}{N} \times \frac{(a+c)}{N}$$

$$E(a) = NP(X_1 \cap Y_1) = N \times \frac{(a+b)}{N} \times \frac{(a+c)}{N}$$

$$E(a) = \frac{(a+b)(a+c)}{N} \quad (4.83)$$

ในทำนองเดียวกันก็จะได้

$$E(b) = \frac{(a+b)(b+d)}{N} \quad (4.84)$$

$$E(c) = \frac{(a+c)(c+d)}{N} \quad (4.85)$$

$$E(d) = \frac{(b+d)(c+d)}{N} \quad (4.86)$$

จาก

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4.87)$$

$$O_{ij} = \begin{matrix} & a & b & c & d \end{matrix}$$

$$E_{ij} = \begin{matrix} E(a) & E(b) & E(c) & E(d) \end{matrix}$$

ดังนั้น

$$\chi^2 = \frac{[a - E(a)]^2}{E(a)} + \frac{[b - E(b)]^2}{E(b)} + \frac{[c - E(c)]^2}{E(c)} + \frac{[d - E(d)]^2}{E(d)} \quad (4.88)$$

พิจารณากรณีที่ 1

$$(O_{ij} - E_{ij}) = [a - E(a)]$$

$$a - E(a) = a - \left[\frac{(a+b)(a+c)}{N} \right]$$

$$= \frac{aN - (a+b)(a+c)}{N}$$

$$= \frac{a(a+b+c+d) - (a+b)(a+c)}{N}$$

$$= \frac{a^2 + ab + ac + ad - (a^2 + ab + ac + bc)}{N}$$

$$= \frac{(ad - bc)}{N}$$

$$\frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{[a - E(a)]^2}{E(a)}$$

$$= \left[\frac{ad - bc}{N} \right]^2 \div \left[\frac{(a+b)(a+c)}{N} \right]$$

$$= \left[\frac{ad - bc}{N} \right]^2 \times \left[\frac{N}{(a+b)(a+c)} \right]$$

$$= \left[\frac{(ad - bc)^2}{N(a+b)(a+c)} \right] \quad (4.89)$$

พิจารณากรณีที่ 2

$$\begin{aligned}
(O_{ij} - E_{ij}) &= [b - E(b)] \\
&= b - \left[\frac{(a+b)(b+d)}{N} \right] \\
&= \frac{bN - (a+b)(b+d)}{N} \\
&= \frac{b(a+b+c+d) - (a+b)(b+d)}{N} \\
&= \frac{ab + b^2 + bc + bd - (ab + ad + b^2 + bd)}{N} \\
b - E(b) &= \frac{bc - ad}{N} \\
\frac{(O_{ij} - E_{ij})^2}{E_{ij}} &= \frac{[b - E(b)]^2}{E(b)} \\
&= \left[\frac{bc - ad}{N} \right]^2 \div \left[\frac{(a+b)(b+d)}{N} \right] \\
&= \left[\frac{bc - ad}{N} \right]^2 \times \left[\frac{N}{(a+b)(b+d)} \right] \\
&= \left[\frac{(bc - ad)^2}{N(a+b)(a+c)} \right] \quad (4.90)
\end{aligned}$$

พิจารณากรณีที่ 3

$$(O_{ij} - E_{ij}) = [c - E(c)]$$

$$c - E(c) = c - \left[\frac{(a+c)(c+d)}{N} \right]$$

$$= \frac{cN - (a+c)(c+d)}{N}$$

$$= \frac{c(a+b+c+d) - (a+c)(c+d)}{N}$$

$$= \frac{ac + bc + c^2 + cd - (ac + ad + c^2 + cd)}{N}$$

$$= \frac{(bc - ad)}{N}$$

$$\frac{(O_{ij} - E_{ij})}{E_{ij}} = \frac{(c - E(c))^2}{E(c)}$$

$$= \left[\frac{bc - ad}{N} \right]^2 \div \left[\frac{(a+c)(c+d)}{N} \right]$$

$$= \left[\frac{bc - ad}{N} \right]^2 \times \left[\frac{N}{(a+c)(c+d)} \right]$$

$$= \left[\frac{(bc - ad)^2}{N(a+c)(c+d)} \right] \quad (4.91)$$

พิจารณากรณีที่ 4

$$(O_{ij} - E_{ij}) = [d - E(d)]$$

$$d - E(d) = d - \left[\frac{(c+d)(b+d)}{N} \right]$$

$$= \frac{dN - (c+d)(b+d)}{N}$$

$$= \frac{d(a+b+c+d) - (c+d)(b+d)}{N}$$

$$= \frac{ad + bd + cd + d^2 - (bc + cd + bd + d^2)}{N}$$

$$= \frac{(ad - bc)}{N}$$

$$\frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{[b - E(b)]^2}{E(b)}$$

$$= \left[\frac{bc - ad}{N} \right]^2 \div \left[\frac{(a+c)(c+d)}{N} \right]$$

$$= \left[\frac{bc - ad}{N} \right]^2 \times \left[\frac{N}{(a+c)(c+d)} \right]$$

$$= \left[\frac{(bc - ad)^2}{N(a+c)(c+d)} \right] \quad (4.92)$$

จาก

$$\chi^2 = \sum_{i=a}^d \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

และ จากสมการที่ (4.89), (4.90), (4.91) และ (4.92) จะได้

$$\chi^2 = \left[\frac{(ad - bc)^2}{N(a+b)(a+c)} + \frac{(bc - ad)^2}{N(a+b)(b+d)} + \frac{(bc - ad)^2}{N(a+c)(c+d)} + \frac{(ad - bc)^2}{N(c+d)(b+d)} \right]$$

เนื่องจาก

$$(ad - bc)^2 = (bc - ad)^2$$

จะได้ว่า

$$\chi^2 = \frac{(ad - bc)^2}{N} \left[\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)} + \frac{1}{(a+c)(c+d)} + \frac{1}{(c+d)(b+d)} \right]$$

$$\chi^2 = \frac{(ad - bc)^2}{N} \left[\frac{(c+d)(b+d)}{(a+b)(a+c)(c+d)(b+d)} + \frac{(a+c)(c+d)}{(a+b)(a+c)(c+d)(b+d)} + \frac{(a+b)(b+d)}{(a+b)(a+c)(c+d)(b+d)} + \frac{(a+b)(a+c)}{(a+b)(a+c)(c+d)(b+d)} \right]$$

$$\chi^2 = \frac{(ad - bc)^2}{N} \left[\frac{bc + cd + bd + d^2 + ac + ad + c^2 + cd}{(a+b)(a+c)(c+d)(b+d)} + \frac{ab + ad + b^2 + bd + a^2 + ac + ab + bc}{(a+b)(a+c)(c+d)(b+d)} \right]$$

$$\chi^2 = \frac{(ad - bc)^2}{N} \left[\frac{a^2 + b^2 + c^2 + d^2 + 2bc + 2cd + 2bd + 2ac + 2ad + 2ab}{(a+b)(a+c)(c+d)(b+d)} \right]$$

จากสมการที่ (4.75) $N^2 = a^2 + b^2 + c^2 + d^2 + 2(ab + ac + ad + bc + bd + cd)$
จะได้ว่า

$$\chi^2 = \frac{(ad - bc)^2}{N} \left[\frac{N^2}{(a+b)(a+c)(c+d)(b+d)} \right]$$

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (4.93)$$

สมการที่ (4.93) หาคด้วย N

$$\frac{\chi^2}{N} = \frac{N(ad - bc)^2}{N(a+b)(a+c)(c+d)(b+d)}$$

$$\frac{\chi^2}{N} = \frac{(ad - bc)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (4.94)$$

ถอดรากที่สองของสมการที่ (4.94) จะได้

$$\sqrt{\frac{\chi^2}{N}} = \frac{(ad - bc)}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$$

ดังนั้นสัมประสิทธิ์สหสัมพันธ์ ฟิ (Phi coefficient correlation : ϕ)

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

$$= \frac{(ad - bc)}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$$

4.5 การพิสูจน์สูตร สัมประสิทธิ์ความสอดคล้องเคนคอลล์

จากที่ผ่านมา เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรมาตราเรียงอันดับ 2 ตัวแปรว่ามีความสัมพันธ์กันอย่างไร เป็นที่ทราบกันดีแล้วว่าสถิติที่ใช้หาความสัมพันธ์มีได้หลายตัว เช่น สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน หรือ สัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล์ แต่ในกรณีที่สนใจศึกษาความสัมพันธ์ระหว่างตัวแปรมาตราเรียงอันดับ ตั้งแต่ 3 ตัวแปรขึ้นไป สถิติที่ใช้หาความสัมพันธ์ของการจัดอันดับข้อมูลตั้งแต่ 3 เซตขึ้นไปก็คือสัมประสิทธิ์ความสอดคล้องเคนคอลล์ ในที่นี้จะกล่าวถึงการพิสูจน์สูตรที่เกี่ยวข้องกับสัมประสิทธิ์ความสอดคล้องเคนคอลล์ ดังรายละเอียดต่อไปนี้

สูตรที่เกี่ยวข้อง

- $\omega = \frac{12S}{m^2(n^3 - n)}$
- $R_j = \frac{mn(n+1)}{2}$
- $\chi^2 = m(n-1)\omega$

4.5.1 การพิสูจน์สูตร สูตรที่เกี่ยวข้องกับสัมประสิทธิ์ความสอดคล้องเคนคอลล์

พิสูจน์สูตร สัมประสิทธิ์ความสอดคล้องเคนคอลล์

$$\omega = \frac{12S}{m^2(n^3 - n)}$$

ให้ MaxS คือผลรวมส่วนเบี่ยงเบนกำลัง 2 ของ S ที่มีขนาดใหญ่ที่สุดเท่าที่จะเป็นไปได้ ดังนั้นในกรณีนี้ ให้

$$\omega = \frac{S}{\text{Max } S} \quad (4.95)$$

โดยมี ผู้จัดอันดับ m คน และ ผู้ถูกจัดอันดับ n คน ผลรวมของอันดับคือ R_j

$$\sum R_j = \frac{mn(n+1)}{2}$$

ค่าเฉลี่ยของผลรวมของอันดับคือ

$$\frac{\sum R_j}{n} = \frac{m(n+1)}{2}$$

ดังนั้นผลรวมกำลังสองของส่วนเบี่ยงเบน คือ

$$S = \sum_{i=1}^n \left(R_j - \frac{\sum R_j}{n} \right)^2$$

$$S = \sum_{i=1}^n \left(R_j - \frac{m(n+1)}{2} \right)^2$$

ถ้าอันดับความสอดคล้องกันอย่างสมบูรณ์ ผลรวมของอันดับจะเป็นดังนี้

$1m, 2m, 3m, \dots, nm$

$$\begin{aligned} \text{Max } S &= \sum_{i=1}^n \left(im - \frac{m(n+1)}{2} \right)^2 \\ &= \sum_{i=1}^n (im)^2 - \frac{\sum m^2 (n+1)^2}{4} \\ &= m^2 \sum i^2 - \frac{nm^2 (n+1)^2}{4} \\ &= m^2 \left[\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \right] \end{aligned}$$

$$\text{ดังนั้น } \text{Max } S = \frac{m^2(n^3 - n)}{12} \quad (4.96)$$

การวัดความสอดคล้องความคิดเห็นคือ

$$\omega = \frac{S}{\text{Max } S}$$

$$\omega = S \times \frac{12}{m^2(n^3 - n)}$$

จะได้สมการที่ (3.29)

$$\omega = \frac{12S}{m^2(n^3 - n)}$$

ω คือ สัมประสิทธิ์ความสอดคล้องเคนดอลล์

4.5.2 พิสูจน์สูตร ที่ใช้ในการทดสอบนัยสำคัญของสัมประสิทธิ์ความสอดคล้องเคนคอลด์

$$\chi^2 = m(n-1)\omega = \frac{12S}{mn(n+1)}$$

กำหนดให้ Max S คือ ผลรวมส่วนเบี่ยงเบนกำลัง 2 ของ S ที่มีขนาดใหญ่ที่สุดเท่าที่จะเป็นไปได้

S คือ ผลรวมกำลังสองของส่วนเบี่ยงเบนจากค่าเฉลี่ย ของ R_j
จากสมการที่ (4.96) และ (3.33) จะได้

$$\text{Max } S = \frac{m^2(n^3 - n)}{12} = \sigma^2$$

$$S = \sum_{j=1}^n (R_j - \frac{\sum R_j}{n})^2 = S^2$$

จากทฤษฎีบทที่ 2.3 การแจกแจงไคสแควร์อธิบายลักษณะความแปรปรวนของตัวอย่าง
จะได้ว่า

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$\text{ให้ } u = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (4.97)$$

แทนสมการที่ (4.96) และ (3.33) ในสมการที่ (4.97)

$$u = \frac{(n-1)S}{\text{Max } S}$$

$$\text{Max } S = \frac{m^2(n^3 - n)}{12}$$

$$\therefore u = \frac{(n-1)S}{\frac{m^2(n^3 - 1)}{12}}$$

$$\begin{aligned}
&= \frac{12S(n-1)}{m^2(n^3-n)} \\
&= \frac{12S(n-1)}{m^2n(n^2-1)} \\
&= \frac{12S(n-1)}{m^2n(n-1)(n+1)} \\
u &= \frac{12S}{m^2n(n+1)} = \chi_{(n-1)}^2 \quad (4.98)
\end{aligned}$$

และ

$$\chi_{(n-1)}^2 = \frac{12S}{mn(n+1)}$$

$$\text{สมการที่ (4.98)} \times \frac{m(n-1)}{m(n-1)}$$

$$\chi_{(n-1)}^2 = \frac{12Sm(n-1)}{m^2n(n-1)(n+1)}$$

$$\chi_{(n-1)}^2 = \frac{12Sm(n-1)}{m^2n(n^2-1)}$$

$$\chi_{(n-1)}^2 = m(n-1) \times \frac{12S}{m^2(n^3-n)} \quad (4.99)$$

$$\text{แทนค่า } \omega = \frac{12S}{m^2(n^3-n)} \quad \text{ในสมการที่ (4.99)}$$

จะได้ สมการที่ (3.34)

$$\chi_{(n-1)}^2 = m(n-1)\omega$$

เป็นสูตร ที่ใช้ในการทดสอบนัยสำคัญของสัมประสิทธิ์ความสอดคล้องเคนคอลล์

4.5.3 พิสูจน์สูตรผลรวมของอันดับ

$$R_j = \frac{mn(n+1)}{2}$$

เมื่อความคิดเห็นของผู้จัดอันดับจำนวน m คน สอดคล้องกันอย่างสมบูรณ์ R_j ที่ได้จะมีลักษณะเป็นอนุกรม

$$m, 2m, 3m, 4m, \dots, nm$$

จาก ผลบวก n เทอม

$$1m + 2m + 3m + \dots + nm = m(1 + 2 + 3 + \dots + n) \quad (4.100)$$

จะเห็นว่ามีลักษณะเป็นอนุกรมเลขคณิต ซึ่งมีผลบวกอนุกรมเป็นดังนี้

$$S_n = \frac{n}{2}[2a_1 + (n-1)d] \quad (4.101)$$

และ จากสมการที่ (4.100) จะได้ว่า

$$a_1 = 1$$

$$d = 1$$

แทน a_1 และ d ในสมการที่ (4.101)

$$S_n = \frac{n}{2}[2(1) + (n-1)(1)]$$

$$S_n = \frac{n}{2}(n+1)$$

เนื่องจากมีผู้จัดอันดับจำนวน m คน จึงบวก m ครั้ง ดังนั้น ผลรวมของอันดับ คือ

$$S_n = \frac{n}{2}(n+1) + \frac{n}{2}(n+1) + \dots + \frac{n}{2}(n+1)$$

$$= \frac{mn}{2}(n+1)$$

$$S_n = R_j$$

4.6 สัมประสิทธิ์สหสัมพันธ์อันดับที่ กรณีอันดับของค่าสังเกตที่ซ้ำกัน

ในการประยุกต์การหาค่าสัมประสิทธิ์สหสัมพันธ์อันดับที่ บางครั้งเกิดกรณีที่ค่าสังเกตซ้ำกันมากกว่า 2 ตัว เมื่อนำค่าสังเกตเหล่านี้มาให้อันดับที่ก็จะเกิดอันดับที่ซ้ำกันมากกว่า 2 ตัวด้วยเช่นกัน ผลของอันดับที่ของค่าสังเกตที่ซ้ำกันนี้จะมีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าลดลง ดังนั้นจึงต้องมีการปรับแก้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์เพื่อให้ค่าที่ได้มีความถูกต้องมากที่สุด วิธีที่ใช้ในการปรับแก้กันนั้นจะทำได้โดยการหาค่าเฉลี่ยของ อันดับของค่าสังเกตที่ซ้ำกัน(Mid rank) เช่น

- ถ้าอันดับของค่าสังเกตตัวที่ 3 และ 4 ซ้ำกัน ค่าเฉลี่ยอันดับที่ได้คือ $\frac{3+4}{2} = 3.5$ ซึ่ง 3.5 จะเป็นอันดับของสมาชิกตัวที่ 3 และ 4
- ถ้าอันดับของค่าสังเกตตัวที่ 2 ถึง 7 ซ้ำกันค่าเฉลี่ยอันดับที่ได้คือ $\frac{1}{6}(2+3+4+5+6+7) = 4.5$ ซึ่ง 4.5 จะเป็นอันดับของสมาชิกตัวที่ 2 ถึง 7

ส่วนค่าสังเกตที่ไม่มีตัวซ้ำก็ยังคงเรียงอันดับไปตามปกติ โดยทั่วไปถ้าอันดับของค่าสังเกตที่ซ้ำกันตั้งแต่ ค่าสังเกตที่ i และ j ค่าเฉลี่ยอันดับที่ได้คือ $\frac{i+j}{2}$ และในที่นี้สนใจพิจารณาผลของ อันดับของค่าสังเกตที่ซ้ำกัน กรณีของ สัมประสิทธิ์สหสัมพันธ์อันดับที่ของเคนคอลล, สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน และ สัมประสิทธิ์ความสอดคล้องเคนคอลล

4.6.1 การคำนวณสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล กรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน

จากบทที่ 3 ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล จะพบว่ามี การพิจารณาการจัดอันดับของผู้ให้อันดับทั้ง 2 คน โดยพิจารณาเปรียบเทียบจากอันดับซ้ายสุดไปขวาสุดทีละคู่ว่าอันดับที่จัดเรียงตามธรรมชาติ(natural order)เรียงจากน้อยไปหามากหรือกลับกัน ถ้าแต่ละคู่เรียงอันดับจากน้อยไปหามาก จะให้คะแนน +1 แต่ถ้าแต่ละคู่เรียงอันดับกลับกันจะให้คะแนน -1 แต่ในกรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน คะแนนจะเป็น 0 (ศูนย์) ซึ่งกรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน การคำนวณค่าความแตกต่างระหว่างผลบวกของจำนวนอันดับที่ ที่สอดคล้องกันกับผลบวกของอันดับที่ ที่ไม่สอดคล้องกัน(S)จะง่ายในการคำนวณ

พิจารณาการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์อันดับที่เคนคอลล (T) โดยพิจารณาจากตัวหารของ S ในที่นี้จำแนกได้ 2 กรณี

$$\Gamma = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}} = \frac{2S}{n(n-1)}$$

a) กรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน

ในการคำนวณค่า τ ตัวหารที่ใช้คือ $\frac{1}{2}n(n-1)$

b) กรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน

จะแทน $\frac{1}{2}n(n-1)$ ด้วย $\frac{1}{2}\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}$

a_{ij} คือคะแนนการเปรียบเทียบของอันดับที่ i และ j ในเซต X

b_{ij} คือคะแนนการเปรียบเทียบของอันดับที่ i และ j ในเซต Y

กรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน จะได้ว่าบางเทอมของ a_{ij}^2 ซึ่งมีค่า $n(n-1)$ จะถูกนำมา
รวมกันเพื่อให้ $\sum a_{ij}^2$ ลดจำนวนเทอมที่เป็นไปได้ลง ในทำนองเดียวกัน สำหรับ $\sum b_{ij}^2$ ดังนั้น

$\frac{1}{2}\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}$ ก็จะลดเป็น $\frac{1}{2}n(n-1)$

กรณีอันดับที่ของค่าสังเกตที่ซ้ำกันของเซต u สมาชิกทั้งหมดของคะแนนที่เพิ่มขึ้นจากบาง
คู่ จะถูกเลือกเซต u จะมีค่าเป็น 0 และต่อเนื่องกันไปดังนั้นจะมี $u(u-1)$ คู่ ผลบวกของ $\sum a_{ij}^2$ จะเป็น
ดังนี้

$$\sum a_{ij}^2 = n(n-1) - \sum u(u-1) \quad (4.102)$$

เมื่อ $\sum$ คือผลบวกทั้งหมดของคะแนนที่ซ้ำกันของเซต u

ถ้ากำหนดให้

$$U = \frac{1}{2} \sum u(u-1) \text{ เป็นอันดับที่ของค่าสังเกตที่ซ้ำกัน ของเซตที่ } 1$$

$$V = \frac{1}{2} \sum v(v-1) \text{ เป็นอันดับที่ของค่าสังเกตที่ซ้ำกัน ของเซตอื่น}$$

รูปแบบอื่นของสัมประสิทธิ์สหสัมพันธ์อันดับที่เล่นคอลด์ กรณีสัมพันธ์อันดับที่ของค่าสังเกตที่ซ้ำกัน

$$r_b = \frac{S}{\sqrt{\left[\frac{1}{2}n(n-1) - U\right]\left[\frac{1}{2}n(n-1) - V\right]}}$$

4.6.2 การคำนวณสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนกรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน

พิจารณาปัญหาในลักษณะเดียวกันกับ สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน ในที่นี้พิจารณาตัวหาร 2 ตัว และ สัมประสิทธิ์สหสัมพันธ์ 2 ค่า ซึ่งแทนด้วย r_{su} และ r_{sv} ถ้าเซตของอันดับที่ของค่าสังเกตที่ซ้ำกัน 2 เซต แทนด้วย u และ v

กำหนดให้

$$U' = \frac{1}{12} \sum (u^3 - u) \text{ เป็นอันดับที่ของค่าสังเกตที่ซ้ำกัน ของเซตที่ } i$$

$$V' = \frac{1}{12} \sum (v^3 - v) \text{ เป็นอันดับที่ของค่าสังเกตที่ซ้ำกัน ของเซต } j$$

จะได้ว่าสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน กรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน เป็นดังสมการที่ (3.6)

$$r_{sb} = \frac{\frac{1}{6}(n^3 - n) - \sum d^2 - U' - V'}{\sqrt{\left[\frac{1}{6}(n^3 - n) - 2U'\right]\left[\frac{1}{6}(n^3 - n) - 2V'\right]}}$$

การพิสูจน์สูตรสมการที่ (3.6) นั้นเป็นการนำผลที่ได้จากการพิสูจน์สูตรของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมนกรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน โดยวิธีผลคูณ โปรดักโมเมนต์ (จากหัวข้อที่ 4.1) มาช่วยในการพิสูจน์ นั่นคือ จะเห็นว่าสำหรับเซตของอันดับที่ของค่าสังเกตที่ไม่ซ้ำกัน คือ $p_1, p_2, p_3, \dots, p_n$ จะได้จากสมการที่ (4.6) ผลบวกของอันดับเป็น

$$\sum p_i = \frac{1}{2}n(n+1)$$

และจากสมการที่ (4.7) ผลบวกกำลังสองของอันดับ

$$\sum p_i^2 = \frac{1}{6}n(n+1)(2n+1)$$

ถ้าเซตของ u เป็นอันดับที่ของค่าสังเกตที่ซ้ำกันผลบวกของอันดับก็ยังคงเป็นเช่นเดิม แต่ผลบวก

กำลังสองของอันดับจะเปลี่ยนไปหลายแบบ

สมมติให้ อันดับที่ $p_k + 1, \dots, p_k + u$ เป็นอันดับที่ของค่าสังเกตที่ซ้ำกันดังนั้นผลบวกกำลังสองจะถูกลดรูปเป็นเทอมการปรับแก้ ของสัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน กรณีอันดับที่ของค่าสังเกตที่ซ้ำกันได้ดังนี้

$$(p_k + 1)^2 + (p_k + 2)^2 + \dots + (p_k + u)^2 - u\left\{p_k + \frac{1}{2}(u+1)\right\}^2 \quad (4.103)$$

$$= [(p_k + \sum p_i)^2] - u\left[p_k + \frac{1}{2}(u+1)\right]^2$$

$$= [(up_k^2 + 2p_k \sum p_i + \sum p_i^2)] - u\left[p_k^2 + 2p_k \frac{1}{2}(u+1) + \frac{1}{4}(u+1)\right]$$

$$= [(up_k^2 + 2p_k(1+2+\dots+u) + (1^2+2^2+\dots+u^2))] - u\left[p_k^2 + p_k(u+1) + \frac{1}{4}(u+1)^2\right]$$

$$= [(up_k^2 + 2p_k \frac{1}{2}u(u+1) + \frac{1}{6}u(u+1)(2u+1))] - [up_k^2 + p_k u(u+1) + \frac{1}{4}u(u+1)^2]$$

$$= up_k^2 + p_k u(u+1) + \frac{1}{6}u(u+1)(2u+1) - up_k^2 - p_k u(u+1) - \frac{1}{4}u(u+1)^2$$

$$= \frac{1}{6}u(u+1)(2u+1) - \frac{1}{4}u(u+1)^2$$

$$= u(u+1)\left[\frac{1}{6}(2u+1) - \frac{1}{4}(u+1)\right]$$

$$= u(u+1)\left[\frac{4(2u+1) - 6(u+1)}{24}\right]$$

$$\begin{aligned}
&= \frac{u(u+1)}{24} [8u + 4 - 6u - 6] \\
&= \frac{u(u+1)}{24} [2u - 2] \\
&= \frac{2}{24} u(u+1)(u-1) \\
&= \frac{1}{12} u(u^2 - 1) \quad (4.104)
\end{aligned}$$

ความแปรปรวนกรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน

$$\frac{1}{n} \sum \left\{ p_i - \frac{1}{2}(n+1) \right\}^2 = \frac{1}{n} \sum p_i^2 - \frac{1}{4}(n+1)^2 = \frac{1}{12}(n^2 - 1) \quad (4.105)$$

พิสูจน์ พิจารณาจาก

$$V(p_i) = E[p_i - E(p_i)]^2$$

$$E(p_i) = \frac{\sum p_i}{n} = \frac{1}{2n} n(n+1) = \frac{1}{2}(n+1)$$

$$V(p_i) = E[p_i - E(p_i)]^2$$

$$= \sum [p_i - E(p_i)]^2 P(p_i)$$

$$= \sum \frac{1}{n} [p_i - E(p_i)]^2$$

$$\begin{aligned}
&= \frac{1}{n} \sum \left\{ p_i - \frac{1}{2}(n+1) \right\}^2 \\
&= \frac{1}{n} \sum \left\{ p_i^2 - 2p_i \frac{1}{2}(n+1) + \frac{1}{4}(n+1)^2 \right\} \\
&= \frac{1}{n} \left[\sum p_i^2 - (n+1) \sum p_i + \frac{1}{4}n(n+1)^2 \right] \\
&= \frac{1}{n} \sum p_i^2 - \frac{(n+1)}{n} \frac{1}{2}n(n+1) + \frac{1}{4n}n(n+1)^2 \\
&= \frac{1}{6}(n+1)(2n+1) - \frac{1}{2}(n+1)^2 + \frac{1}{4}(n+1)^2 \\
&= \frac{1}{2}(n+1) \left[\frac{1}{3}(2n+1) - (n+1) + \frac{1}{2}(n+1) \right] \\
&= \frac{1}{2}(n+1) \left[\frac{2(2n+1) - 6(n+1) + 3(n+1)}{6} \right] \\
&= \frac{1}{2}(n+1) \frac{1}{6}(n-1) \\
&= \frac{1}{12}(n^2 - 1)
\end{aligned}$$

ความแปรปรวนกรณีอันดับที่ของค่าสังเกตซ้ำกัน

$$\frac{1}{12}(n^2 - 1) - \frac{1}{12n} \sum (u^3 - u) = \frac{1}{2n} \left[\frac{1}{6}(n^3 - n) - 2U' \right] \quad (4.106)$$

พิสูจน์ กรณีสันที่ของค่าสังเกตซ้ำกัน ความแปรปรวนจะต้องมีเทอมของการปรับแก้เพื่อให้ค่าที่ได้มีความถูกต้องมากที่สุด นั่นคือจะได้จากการพิจารณาความแปรปรวนกรณีสันที่ของค่าสังเกตไม่ซ้ำกัน(สมการที่ 4.105) ลบด้วยเทอมการปรับแก้จากสมการที่ (4.104) ความแปรปรวนที่ได้เป็นดังนี้

$$\begin{aligned} V(p_i) &= \frac{1}{12}(n^2 - 1) - \frac{1}{12n} \sum (u^3 - u) \\ &= \frac{1n}{12n}(n^2 - 1) - \frac{1}{12n} \sum (u^3 - u) \\ &= \frac{1}{2n} \left[\frac{1}{6}(n^3 - n) - \frac{1}{6} \sum (u^3 - u) \right] \\ &= \frac{1}{2n} \left[\frac{1}{6}(n^3 - n) - 2U' \right] \end{aligned}$$

จากสมการที่ (3.8)

$$U' = \frac{1}{12} \sum (u^3 - u)$$

$$2U' = \frac{2}{12} \sum (u^3 - u)$$

ดังนั้นในการคำนวณค่า r_{sb} จะกำหนดให้

$$r_{sb} = \frac{\text{Cov}(p, q)}{\sqrt{V(p)V(q)}} \quad (4.107)$$

$$\text{และ } V(p - q) = V(p) + V(q) - 2\text{Cov}(p, q) \quad (4.108)$$

$$\text{Cov}(p, q) = \frac{1}{2}[V(p) + V(q) - V(p - q)] \quad (4.109)$$

กรณีอันดับที่ของค่าสังเกตไม่ซ้ำกัน จากสมการที่ (4.25) จะได้ว่า

$$\begin{aligned}\mu_{11} &= \text{Cov}(p, q) \\ &= \frac{1}{12}(n^2 - 1) - \frac{1}{2n} \sum d_i^2\end{aligned}$$

กรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน การหาความแปรปรวนร่วมย่อมจะต้องมีเทอมการปรับแก้ด้วยเช่นกัน จากสมการที่ (4.25) และ (4.109) จะได้ว่า

$$\begin{aligned}\mu_{11} &= \text{Cov}(p, q) \\ \text{Cov}(p, q) &= \frac{1}{2}[V(p) + V(q) - V(p - q)] \\ &= \frac{1}{2} \left\{ \left[\frac{1}{12}(n^2 - 1) - \frac{1}{12n} \sum (u^3 - u) \right] + \left[\frac{1}{12}(n^2 - 1) - \frac{1}{12n} \sum (v^3 - v) \right] + \frac{1}{n} \sum d_i^2 \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{2n} \left[\frac{1}{6}(n^3 - 1) - 2U' \right] + \frac{1}{2n} \left[\frac{1}{6}(n^3 - 1) - 2V' \right] + \frac{2}{2n} \sum d_i^2 \right\} \\ &= \frac{1}{2} \left(\frac{1}{2n} \right) \left\{ \left[\frac{1}{6}(n^3 - 1) - 2U' \right] + \left[\frac{1}{6}(n^3 - 1) - 2V' \right] + 2 \sum d_i^2 \right\} \\ &= \frac{1}{2} \left(\frac{1}{2n} \right) \left[\frac{2}{6}(n^3 - n) - 2U' - 2V' + 2 \sum d_i^2 \right] \\ &= \frac{1}{2n} \left[\frac{1}{6}(n^3 - n) - U' - V' + \sum d_i^2 \right] \tag{4.110}\end{aligned}$$

แทนสมการที่ (4.106) และ (4.110) ในสมการที่ (4.107) จะได้

$$\begin{aligned}
r_{sb} &= \frac{\text{Cov}(p, q)}{\sqrt{V(p)V(q)}} \\
&= \frac{\frac{1}{2n} \left[\frac{1}{6} (n^3 - n) - U' - V' + \sum d_i^2 \right]}{\sqrt{\left\{ \frac{1}{2n} \left[\frac{1}{6} (n^3 - n) - 2U' \right] \right\} \left\{ \frac{1}{2n} \left[\frac{1}{6} (n^3 - n) - 2V' \right] \right\}}} \\
&= \frac{\frac{1}{2n} \left[\frac{1}{6} (n^3 - n) - U' - V' + \sum d_i^2 \right]}{\frac{1}{2n} \sqrt{\left[\frac{1}{6} (n^3 - n) - 2U' \right] \left[\frac{1}{6} (n^3 - n) - 2V' \right]}} \\
&= \frac{\left[\frac{1}{6} (n^3 - n) - U' - V' + \sum d_i^2 \right]}{\sqrt{\left[\frac{1}{6} (n^3 - n) - 2U' \right] \left[\frac{1}{6} (n^3 - n) - 2V' \right]}}
\end{aligned}$$

4.6.3 การคำนวณสัมประสิทธิ์ความสอดคล้องเคนคอลล์ กรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน

ในกรณีอันดับที่ของค่าสังเกตที่ซ้ำกัน การคำนวณค่าสัมประสิทธิ์ความสอดคล้องเคนคอลล์ ก็จะมีเทอมการปรับแก้เป็นเช่นเดียวกันกับ สัมประสิทธิ์สหสัมพันธ์อันดับที่สเปียร์แมน ดังสมการที่ (3.8)

$$U' = \frac{1}{12} \sum (u^3 - u)$$

และการคำนวณค่าสัมประสิทธิ์ความสอดคล้องเคนคอลล์จะกำหนดดังสมการที่ (3.30)

$$\omega = \frac{S}{\frac{1}{2} m^2 (n^3 - n) - m \sum U'}$$

พิจารณาตัวหารในสมการที่ (3.30) ไม่จำเป็นต้องเป็นค่าที่มากที่สุดของผลบวกกำลังสองของความคลาดเคลื่อนของผลรวมอันดับ อาจนิยาม ω ในกรณีอันดับที่ของค่าสังเกตไม่ซ้ำกันได้โดยรูปแบบทั่วไปดังนี้คือ

$$\omega = \frac{S}{mS'} \quad (4.112)$$

ซึ่ง S' คือ ผลบวกกำลังสองของความคลาดเคลื่อนของอันดับจากค่าเฉลี่ย ซึ่งคล้ายกับ S ที่นิยามไว้ในสมการที่ (3.29) เพราะว่าผลบวกกำลังสองของความคลาดเคลื่อนในทุกอันดับคือ $\frac{1}{2}(n^3 - n)$ และนั่นคือ m อันดับ การนิยามสมการที่ (3.30) จะสอดคล้องกับสมการที่ (4.112) เพราะว่าจากบทที่ 3 จะกล่าวถึงผลของอันดับที่ของค่าสังเกตที่ซ้ำกัน ในอันดับ ถ้าหากว่า U' คือการลดลงของผลบวกกำลังสองของความคลาดเคลื่อนโดย U' หรือจะกล่าวได้ว่า S' คือ ผลบวกกำลังสองของความคลาดเคลื่อนการทดลองแบบสุ่มตลอดนั่นเอง

สาเหตุที่ทำให้การประยุกต์สมการที่ (4.112) เพราะการนิยามนั้นคล้ายกับการวิเคราะห์ความแปรปรวน สมมติ เซตของอันดับซึ่งวัดมาจากค่าเฉลี่ยนั้นคือ

	กรรมวิธี			
	1	2		N
	X_{11}	X_{12}		X_{1n}
	X_{21}	X_{22}	...	X_{2n}
	$\vdots$	$\vdots$		$\vdots$
	X_{m1}	X_{m2}		X_{mn}
ผลรวม	S_1	S_2		S_n
ค่าเฉลี่ย	$\frac{S_1}{m}$	$\frac{S_2}{m}$		$\frac{S_n}{m}$

ผลรวมของแถวจะเท่ากับ	0
ซึ่งสอดคล้องกับค่าเฉลี่ยที่เท่ากับ	0
ผลรวมของคอลัมน์ก็อาจจะเขียนได้เป็น	$S_1, S_2, \dots, S_n$
ค่าเฉลี่ยของคอลัมน์ก็คือ	$\frac{S_1}{m}, \frac{S_2}{m}, \dots, \frac{S_n}{m}$
ความแปรปรวนทั้งหมดคือ	$\frac{S'}{mn}$
ความแปรปรวนของค่าเฉลี่ยคอลัมน์ก็คือ	$\frac{1}{n} \sum \left(\frac{S_i}{m} \right)^2 = \frac{S}{m^2 n}$
อัตราส่วนของความแปรปรวนของค่าเฉลี่ยคอลัมน์ ต่อความแปรปรวนทั้งหมด คือ	$\omega = \frac{S}{mS'}$