PRACE NAUKOWE Wyższej Szkoły Pedagogicznej w Częstochowie
Nr 3/1980 Seria Matematyczno-Przyrodnicza

Tomasz M. ZIMNY

W sprawie ogólnej systematyki miar współuporządkowania

Wstęp

Spostrzeganie i porównywanie własności elementów materialnych prowadzi do ich mierzenia w szerokim tego słowa znaczeniu, a więc do ich rozróżniania i budowy skal klasyfikacyjnych, do ich porządkowania i budowy skal porządkowych, wreszcie do ich mierzenia w ścisłym tego słowa znaczeniu przy użyciu umownych jednostek miary i budowy skal przedziałowych i stosunkowych. Elementy materialne można porównywać ze względu na wiele ich właściwości i można badać ścisłość związku między wyróżnionymi ich właściwościami. Jeśli mierzymy dwie właściwości pewnego zbioru elementów na skalach porządkowych mocnych lub słabszych, to ścisłość związku między tymi właściwościami określamy zwykle za pomocą dwu różnych miar:

1. współczynnika ϱ K. Spearmana (1906)
2. współczynnika τ M. Kendalla (1938) w oparciu o wcześniejsze osiągnięcia Greinera (1909) oraz Esschera (1924).

1. Współczynnik ϱ K. Spearmana. Jego geneza i znaczenie

Współczynnik ϱ K. Spearmana nie jest klasyczną miarą współuporządkowania dwu zbiorów elementów, lecz miarą współzależności rang poszczególnych elementów tych zbiorów. Jest współczynnikiem r Pearsona

$$r = \frac{\mathrm{cov}(x,y)}{s(x)\, s(y)} \tag{1.1}$$

gdzie cov(x,y) — kowariancja zmiennych x oraz y.

s(x), s(y) — odchylenie standardowe odpowiednio zmiennej x, y

stosowanym przy założeniu, że pomiary własności elementów są równe ich rangom, czyli numerom miejsc zajmowanych w zbiorze uporządkowanym według niemalejącego (rosnącego) albo nierosnącego (malejącego) natężenia badanej właściwości. Wyprowadzenie wzoru współczynnika ϱ K. Spearmana ze współczynnika korelacji liniowej r Pearsona przy powyższym założeniu dokonam najpierw jako wzoru ϱ_s dla zbiorów silnie uporządkowanych, a więc nie zawierających elementów równych pod względem natężenia badanej właściwości, a następnie jako wzoru ϱ_w dla zbiorów słabo uporządkowanych, czyli zawierających elementy równe ze względu na badaną właściwość (ściślej: nieodróżnialne albo różniące się nieistotnie mało). Oczywiste jest, że współczynnik ϱ_s jest szczególnym przypadkiem ϱ_w.

1.1. Wyprowadzenie współczynnika ϱ_s dla uporządkowań silnych

W dwu zbiorach własności elementów (x), (y) silnie uporządkowanych zachodzi:

$$\bar{x}=\bar{y}=\frac{1}{2}(n+1) \tag{1.2.}$$

$$s^2(x)=s^2(y)=\frac{1}{n}\sum_{i=1}^{n}\left[x_i-\frac{1}{2}(n+1)\right]=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\frac{1}{4}(n+1)^2 \tag{1.3.}$$

$$\mathrm{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\left[x_i-\frac{1}{2}(n+1)\right]\left[y_i-\frac{1}{2}(n+1)\right]=$$

$$=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\frac{1}{4}(n+1)^2 \tag{1.4.}$$

We wzorze (1.3.) pozostaje dalej do przekształcenia według przyjętego założenia wyrażenie $\sum_{i=1}^{n}x_i^2$ a we wzorze (1.4.) wyrażenie $\sum_{i=1}^{n}x_iy_i$, przystępując więc, najpierw do przekształcenia wyrażenia $\sum_{i=1}^{n}x_i^2$ Przedstawiam $\sum_{i=1}^{n}x_i^s$ dla s=1, 2, 3, ..., m w postaci wielomianu

$$\sum_{i=1}^{n}a_in^i \qquad a_i \in R^n$$

$$\sum_{i=1}^{n}x_i^s=\sum_{i=1}^{n}a_in^i \tag{1.3.1.}$$

Stosuję metodę indukcji matematycznej

$$\sum_{i=1}^{n+1}x_i^s=\sum_{i=0}^{n}a_i(n+1)^i \tag{1.3.2.}$$

przekształcam

$$\sum_{i=1}^{k}a_i(n+1)^i=\sum_{i=0}^{k+1}\sum_{j=1}^{i}a_i\binom{i}{j}n^{i-j}=\sum_{h=0}^{k}\sum_{i=h}^{k+1}a_i\binom{i}{i-h}n^h \tag{1.3.3.}$$

Podstawiając (1.3.3.) do (1.3.2.) i odejmując stronami (1.3.1) otrzymuję:

$$(n+1)^s = \sum_{h=o}^{k} \left[\sum_{i=h}^{k+1} a_i \binom{i}{i-h} n^h - a_h n^h \right] \qquad (1.3.4.)$$

czyli

$$\sum_{h=o}^{s} \binom{s}{h} n^h = \sum_{h=o}^{k} \left[\sum_{i=h+1}^{k+1} a_i \binom{i}{i-h} \right] n^h \qquad (1.3.5.)$$

Dwa wielomiany są równe wtedy i tylko wtedy, gdy współczynniki przy tych samych potęgach są równe. Wielomiany (1.3.5.) są równe, stąd $k=s$. Otrzymujemy następujący układ równań:

$$\binom{s}{h} = \sum_{i=h+1}^{s+1} a_i \binom{i}{i-h} \quad \text{gdzie } h=0,1,\ldots,s \qquad (1.3.6.)$$

Jest to układ $s+1$ równań schodkowych z $s+1$ niewiadomymi taki, że w $s+1$-szym równaniu występuje 1 niawiadoma, a w 1-szym równaniu występuje $s+1$ niewiadomych.

Po rozwiązaniu tego układu równań dla $s=2$ otrzymuję:

$$a_1 = \frac{1}{6}, \; a_2 = \frac{1}{2}, \; a_3 = \frac{1}{3}$$

czyli

$$\sum_{i=1}^{n} x_i^2 = \frac{1}{6} n + \frac{1}{2} n^2 + \frac{1}{3} n^3 = \frac{1}{6} n(n+1)(2n+1) \qquad (1.3.7.)$$

Po podstawieniu (1.3.7.) do (1.3.) otrzymuję:

$$s^2(x) = s^2(y) = \frac{1}{12}(n^2-1) \qquad (1.3.8.)$$

Z kolei przyjępuję do przekształcenia wyrażenia $\sum_{i=1}^{n} x_i y_i$

$$\sum_{i=1}^{n} x_i y_i = \frac{1}{2} \sum_{i=1}^{n} x_i^2 + \frac{1}{2} \sum_{i=1}^{n} y_i^2 - \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (1.4.1.)$$

podstawiając (1.3.7.) otrzymuję:

$$\sum_{i=1}^{n} x_i y_i = \frac{1}{6}(n+1)(2n+1) - \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (1.4.2.)$$

Podstawiając (1.4.2.) do (1.4.) otrzymuję:

$$\operatorname{cov}(x,y) = \frac{1}{12}(n^2-1) - \frac{1}{2n} \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (1.4.3.)$$

Podstawiając (1.3.8.) oraz (1.4.3.) do (1.1.) i przyjmując, że $x_i - y_i = d_i$ otrzymuję:

$$\varrho_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (1.5.)$$

1.2. Wyprowadzenie współczynnika ϱ_w Spearmana dla uporządkowania słabych

Dane są dwa zbiory własności elementów słabo uporządkowane, a więc zawierające elementy równe pod względem obserwowanych właściwości, a w nich p_k grup elementów równych w k-tym zbiorze ($k=1,2$) i każda grupa elementów równych zawiera t_i elementów ($i=1,2, \ldots, p_k$). Suma kwadratów t_i rang elementów różnych począwszy od $k+1$, a skończywszy na $k+t_i$ jest równa:

$$\sum_{l=1}^{t_i}(k+l)^2=\frac{1}{6}(2k+2t_i+1)(k+t_i+1)(k+t_i)-\frac{1}{6}(2k+1)(k+1)k \quad (1.6.)$$

po przekształceniu (1.6.) otrzymuję:

$$\sum_{l=1}^{t_i}(k+l)^2=k^2t_i+kt_i(t_i+1)+\frac{1}{6}t_i(t_i+1)(2t_i+1) \quad (1.6.1.)$$

Natomiast suma kwadratów t_i rang elementów równych jest równa:

$$\sum_{l=1}^{t_i}\left[k+\frac{1}{2}(t_i+1)\right]^2=k^2t_i+kt_i(t_i+1)+\frac{1}{4}t_i(t_i+1)^2 \quad (1.7.)$$

Różnica między (1.6.) i (1.7.) wynosi zatem:

$$\frac{1}{6}(2t_i+1)(t_i+1)t_i-\frac{1}{4}(t_i+1)^2t_i=\frac{1}{12}(t_i^2-1)t_i \quad (1.8.)$$

Stąd: wariancja $s^2(x)$ zmniejszy się o

$$\frac{1}{12n}\sum_{t_i=1}^{P_1}(t_i^2-1)t_i=\frac{1}{n}T_x \quad 1.9a.)$$

wariancja $s^2(y)$ zmniejszy się o

$$\frac{1}{12n}\sum_{t_i=1}^{P_2}(t_i^2-1)t_i=\frac{1}{n}T_y \quad (1.9b.)$$

a kowariancja cov(x,y) zmniejszy się o

$$\frac{1}{2n}(T_x+T_y) \quad (1.9.c.)$$

W rezultacie otrzymuję:

$$s^2(x)=\frac{1}{12}(n^2-1)-\frac{1}{n}T_x \quad (1.10.1.)$$

$$s^2(x)=\frac{1}{12}(n^2-1)-\frac{1}{n}T_y \quad 1.10.2.)$$

$$\mathrm{cov}(x,y)=\frac{1}{12}(n^2-1)-\frac{1}{2n}(T_x+T_y)-\frac{1}{2n}\sum_{i=1}^{n}d_i^2 \quad (1.10.3.)$$

Podstawiając (1.10.1.), (1.10.2.) oraz (1.10.3.) do 1.1) otrzymuję:

$$\varrho_w=\frac{n(n^2-1)-6(T_x+T_y)-6\sum_{i=1}^{n}d_i^2}{\{[n(n^2-1)-12T_x][n(n^2-1)-12T_y]\}^{1/2}}$$

Wzór ϱ_w (1.11.) przybiera postać wzoru ϱ_s dla wartości $T_x=0$ oraz $T_y=0$, tzn .wtedy gdy $p_1=p_2=0$. Wzór ϱ_s stanowi więc szczególny przypadek wzoru ϱ_w.

2. Współczynnik τ M. Kendalla, jego geneza i znaczenie.

Współczynnik τ M. Kendalla bazuje na liczbie inwersji (przestawień) między dwoma zbiorami uporządkowanymi. Jest więc klasycznym współczynnikiem współuporządkowania dwu zbiorów elementów według badanych właściwości (zmiennych). Podobnie jak poprzednio przyjrzymy się najpierw współczynnikowi τ M. Kendalla dla uporządkowań silnych, a następnie jego uogólnieniu na uporządkowania słabe.

2.1. Współczynnik τ_s M. Kendalla dla uporządkowań silnych

M. Kendall definiuje współczynnik τ_s [5] w następujący sposób:

$$\tau_s = \frac{4R}{n(n-1)} - 1 \tag{2.1.}$$

gdzie n — liczba elementów w próbie

R — liczba przypadków dla których zachodzi równocześnie

$(x_i < x_j)$ oraz $(y_i < y_j)$ bądź $(x_i > x_j)$ oraz $(y_i > y_j)$

Obliczając maksymalną liczbę inwersji I_{max}

$$I_{max} = \binom{n}{2} = \frac{1}{2} n(n-1) \tag{2.2.}$$

oraz faktyczną liczbę inwersji I łatwo wykazać, że

$$R = I_{max} - I \tag{2.3.}$$

Podstawiając (2.3.) oraz (2.2.) do (2.1.) łatwo otrzymać postać wprowadzoną przez Z. Zimnego [12]

$$\tau_s = 1 - \frac{2I}{I_{max}} \tag{2.1a}$$

2.2. Współczynnik τ_w M. Kendalla uogólniony na uporządkania słabe

Gdy zbiory badanych elementów dają się uporządkować jedynie słabo, tj. gdy zawierają elementy równe pod względem właściwości badanej, M. Kendall proponuje następującą poprawkę uogólniającą. Przyjmijmy, że w k-tym zbiorze uporządkowanym dla $k=1,2$ mamy p_k grup jednakowych elementów i każda z tych grup ma t_i elementów gdzie $i=1,\dots, p_k$. Maksymalna liczba inwersji zmniejszy się dla k-tego zbioru o $\sum_{t_i=1}^{p_k} \binom{t_i}{2}$.

W tym miejscu M. Kendall zrobił krok teoretycznie niepoprawny aczkolwiek numerycznie dający dobre przybliżenie: Zamiast wziąć pod uwagę średnią arytmetyczną zmniejszonych sum inwersji obydwu słabych uporządkowań wziął ich średnią geometryczną. A przecież operacją charakterystyczną dla współczynnika τ jest dodawanie, a nie mnożenie, jak dla r oraz ϱ. Wobec tego maksymalna liczba inwersji nie jest równa jak u M. Kendalla

$$I_{max}=\left\{\left[\binom{n}{2}-\sum_{t_i=1}^{P_1}\binom{t_i}{2}\right]\left[\binom{n}{2}-\sum_{t_i=1}^{P_2}\binom{t_i}{2}\right]\right\}^{\frac{1}{2}} \tag{2.4.}$$

lecz

$$I_{max}=\frac{1}{2}\left\{\left[\binom{n}{2}-\sum_{t_i=1}^{P_1}\binom{t_i}{2}\right]+\left[\binom{n}{2}-\sum_{t_i=1}^{P_2}\binom{t_i}{2}\right]\right\}=$$

$$=\binom{n}{2}-\frac{1}{2}\left[\sum_{t_i=1}^{P_1}\binom{t_i}{2}+\sum_{t_i=1}^{P_2}\binom{t_i}{2}\right] \tag{2.5.}$$

jak to przyjął Z. Zimny (13).

Ostatecznie M. Kendall podaje następujący wzór uogólniony na uporządkowania słabe.

$$\tau_W=\frac{S}{I_{max}} \tag{2.6.}$$

gdzie $S=R-I$, natomiast I jest liczbą inwersji, czyli liczbą przypadków, dla których zachodzi jednocześnie $(x_i < x_j)$ i $(y_i > y_j)$ bądź $(x_i > x_j)$ i $(y_i < y_j)$.

2.3. Współczynnik τ^* Z. Zimnego uogólniony na uporządkowania słabe

Z. Zimny uznał koncepcję współczynnika τ M. Kendalla za klasyczną koncepcję miary współuporządkowania opartą na liczbie inwersji w badanych uporządkowaniach. Stwierdził jednak, że dla określenia liczby inwersji w uporządkowaniach słabych zachodzi potrzeba wprowadzenia pojęcia półinwersji (semiinwersji). Półinwersje mogą występować w obydwu uporządkowaniach gdy:

1a) dla $x_i < x_j$ $\quad y_i = y_j$
1b) dla $x_i > x_j$ $\quad y_i = y_j$
2a) dla $x_i = x_j$ $\quad y_i < y_j$
2b) dla $x_i = x_j$ $\quad y_i > y_j$

Przy sumowaniu inwersji, jeśli jednej inwersji przypisujemy wartość 1, to półinwersji przypisujemy wartość $\frac{1}{2}$, tak że dwie półinwersje są równe jednej inwersji, co jest oczywiste, np.:

	$y_i<y_j$	$y_i=y_j$	$y_i>y_j$
$x_i<x_j$	0	$\frac{1}{2}$	1
$x_i=x_j$	$\frac{1}{2}$	0	$\frac{1}{2}$
$x_i>x_j$	1	$\frac{1}{2}$	0

Transformując wzór M. Kendalla w sposób następujący

$$\tau_w=1-\frac{2I_w}{I_{wmax}} \qquad (2.1a.)$$

Z. Zimny zdefiniował:

$$I_w=a+\frac{1}{2}(b+c) \qquad (2.7)$$

gdzie a — liczba przypadków dla których zachodzi jednocześnie $(x_i<x_j)$ i $(y_i>y_j)$ bądź $(x_i>x_j)$ i $(y_i<y_j)$ czyli a=I ze wzoru (2.6.1.)

b — liczba przypadków, dla których zachodzi jednocześnie $(x_i=x_j)$ i $(y_i<y_j)$ bądź $(x_i=x_j)$ i $(y_i>y_j)$

c — liczba przypadków, dla których zachodzi jednocześnie $(x_i<x_j)$ i $(y_i=y_j)$ bądź $(x_i>x_j)$ i $(y_i=y_j)$

czyli b+c jest liczbą półinwersji.

Podstawiając (2.5.) oraz (2.7.) do (2.1a.) otrzymujemy:

$$\tau^*=1-\frac{2a+b+c}{\binom{n}{2}-\frac{1}{2}\left[\sum_{t_i=1}^{P_1}\binom{t_i}{2}+\sum_{t_j=1}^{P_2}\binom{t_i}{2}\right]} \qquad (2.8.)$$

3. Uogólnione współczynniki korelacji

3.1. Uogólniony współczynnik Γ korelacji M. Kendalla

Koncepcja obliczenia I_{max} przez Kendalla wzięła się prawdopodobnie z jego wzoru na uogólniony współczynnik Γ, który miał obejmować współczynniki r, ρ i τ jako przypadki szczególne, i wg którego współczynnik τ można przedstawić także w postaci momentu iloczynowego:

$$\Gamma=\frac{\sum_{i=1}^{n}\sum_{j>i}a_{ij}b_{ij}}{\left(\sum_{i=1}^{n}\sum_{j>i}a_{ij}\sum_{i=1}^{n}\sum_{j>i}b_{ij}\right)^{1/2}} \qquad (3.1.)$$

gdzie a_{ij} — jest oceną obserwacji należącej do zbioru X.
b_{ij} — jest oceną obserwacji należącej do zbioru Y.

Przyjmując $a_{ij}=\begin{cases} 1 \text{ dla } x_i > x_j \\ -1 \text{ dla } x_i < x_j \end{cases}$ $b_{ij}=\begin{cases} 1 \text{ dla } y_i > y_j \\ -1 \text{ dla } y_i < y_j \end{cases}$

otrzymujemy że $\Gamma=\tau$, lecz tylko dla uporządkowań silnych.

Licznik $\sum_{i=1}^{n} \sum_{j>i} a_{ij}b_{ij}=R-I$,

a mianownik $\sum_{i=1}^{n} \sum_{j>i} a_{ij}=\sum_{i=1}^{n} \sum_{j>i} b_{ij}=\binom{n}{2}=I_{max}$.

w rezultacie otrzymujemy, że $\Gamma=\dfrac{R-I}{I_{max}}=\tau$

Jednakże współczynnik τ jest z natury inny od r i ϱ i wszelkie próby przedstawienia jednego w postaci drugiego lub odwrotnie zdają się być skazane na niepowodzenie.

3.2. Uogólniony współczynnik korelacji (monotoniczności) μ_i L. Guttmana

Dużo ciekawszą propozycję uogólnionego współczynnika korelacji przedstawił L. Guttman [9], rozwijając koncepcję współczynnika Γ M. Kendalla. L. Guttman konstruuje swój współczynnik w następujący sposób:

niech (x_i,y_i) dla $i=1,2, ..., n$ będą parami obserwacji

niech $\alpha_{ij}=\begin{cases} 1 \text{ dla } x_i > x_j \\ 0 \text{ dla } x_i = x_j \\ -1 \text{ dla } x_i < x_j \end{cases}$ $\beta_{ij}=\begin{cases} 1 \text{ dla } y_i > y_j \\ 0 \text{ dla } y_i = y_j \\ -1 \text{ dla } y_i < y_j \end{cases}$

niech $\Theta_{ij}^{(1)}=|\alpha_{ij}\beta_{jj}|$ $\Theta_{ij}^{(2)}=|\alpha_{ij}|$ $\Theta_{ij}^{(3)}=|\beta_{ij}|$ $\Theta_{ij}^{(4)}=|\alpha_{ij}|+|\beta_{ij}|-|\alpha_{ij}\beta_{ij}|$

oraz niech $w_{ij}^{(m)}$ będzie nie ujemną liczbą różną od zera dla wszystkich $\Theta_{ij}^{(m)}$ różnych od zera, m=1, 2, 3, 4. Wtedy ogólny współczynnik korelacji będzie równy:

$$\mu^{(m)}=\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{(m)}\alpha_{ij}\beta_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{(m)}\Theta_{ij}^{(m)}} \tag{3.2.}$$

L. Guttman wyróżnia trzy rodziny współczynników $\mu^{(m)}$

1. $\mu_1^{(m)}$ dla których $w_{ij}^{(m)} < w_{ik}^{(m)}$ jeżeli bądź $x_i < x_j < x_k$,
bądź $y_i < y_j < y_k$
2. $\mu_2^{(m)}$ dla których $w_{ij}^{(m)} < w_{ik}^{(m)}$ jeżeli $x_i < x_j < x_k$, oraz $y_i < y_j < y_k$
3. $\mu_o^{(m)}$ dla których $w_{ij}^{(m)} = \Theta_{ij}^{(m)}$ bądź $w_{ij}^{(m)}=\text{const}$.

4. Porównanie klasycznych miar współuporządkowania

4.1. Porównanie współczynnika $\mu_o^{(4)}$ L. Guttmana oraz τ M. Kendalla dla uporządkowań silnych

W obecnych rozważaniach interesuje mnie podana przez L. Guttmana równość $\mu_o^{(4)} = \tau$. Dla uporządkowań silnych równość ta zachodzi faktycznie. Łatwo to wykazać na podstawie wzoru:

$$\mu_o^{(4)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\beta_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\Theta_{ij}^{(4)}} \tag{4.1.}$$

który we wzorze (4.1): otrzymuje z (3.2) gdy w_{ij}=const.

— licznik $\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\beta_{ij}$ jest różnicą podwojonej liczby przypadków, dla których zachodzi $x_i > x_j$ oraz $y_i > y_j$, bądź $x_i < x_j$ oraz $y_i < y_j$ i podwojonej liczby przypadków, dla których zachodzi $x_i < x_j$ oraz $y_i > y_j$, bądź $x_i > x_j$ oraz $y_i < y_j$, gdzie odjemna (czyli podwojona liczba przypadków $x_i > x_j$ oraz $y_i > y_j$ bądź $x_i < x_j$ oraz $y_i < y_j$) stanowi podwojoną liczbę par tak samo uporządkowanych w obydwu ciągach i jest równa 2R. Natomiast odjemnik (czyli podwojona liczba przypadków $x_i < x_j$ oraz $y_i > y_j$, bądź $x_i > x_j$ oraz $y_i < y_j$) stanowi podwojoną liczbę par odwrotnie uporządkowanych w obydwu ciągach, czyli jest równa 2I, stąd

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\beta_{ij} = 2R - 2I = 2(I_{max} - 2I) \tag{4.2.}$$

— mianownik zaś jest równy $2\binom{n}{2} = 2I_{max}$,

czyli
$$\sum_{i=1}^{n}\sum_{j=1}^{n}\Theta_{ij}^{(4)} = 2I_{max} \tag{4.3}$$

Podstawiając (4.2.) i (4.3.) do (4.1.) otrzymuję:

$$\mu_n^{(4)} = \frac{2(I_{max} - 2I)}{2I_{max}} = 1 - \frac{2I}{I_{max}} = \tau$$

Warto tu zauważyć, że dla uporządkowań silnych w żadnej z macierzy nie występuje zero poza główną przekątną.

4.2. Porównanie współczynnika $\mu_o^{(4)}$ L. Guttmana i τ M. Kedalla dla uporządkowań słabych

Dla dokonania tego porównania wyróżniłem 9 następujących przypadków:

Lp.	przypadek	α_{ij}	β_{ij}	$\alpha_{ij}\beta_{ij}$	$\Theta_{ij}^{(1)}$	$\Theta_{ij}^{(2)}$	$\Theta_{ij}^{(3)}$	$\Theta_{ij}^{(4)}$
(1)	$x_i > x_j$ oraz $y_i > y_j$	1	1	1	1	1	1	1
(2)	$x_i > x_j$ oraz $y_i = y_j$	1	0	0	0	1	0	1
(3)	$x_i > x_j$ oraz $y_i < y_j$	1	—1	—1	1	1	1	1
(4)	$x_i = x_j$ oraz $y_i > y_j$	0	1	0	0	0	1	1
(5)	$x_i = x_j$ oraz $y_i = y_j$	0	0	0	0	0	0	0
(6)	$x_i = x_j$ oraz $y_i < y_j$	0	—1	0	0	0	1	1
(7)	$x_i < x_j$ oraz $y_i > y_j$	—1	1	—1	1	1	1	1
(8)	$x_i < x_j$ oraz $y_i = y_j$	—1	0	0	0	1	0	1
(9)	$x_i < x_j$ oraz $y_i < y_j$	—1	—1	1	1	1	1	1

Liczbę wszystkich przypadków (1) oraz (9) oznaczamy przez R.
Liczbę wszystkich przypadków (2) oraz (8) oznaczamy przez c.
Liczbę wszystkich przypadków (3) oraz (7) oznaczamy przez a.
Liczbę wszystkich przypadków (4) oraz (6) oznaczamy przez b.
Liczbę wszystkich przypadków (5) oznaczamy przez f.
Powyższe oznaczenia są zgodne z oznaczeniami stosowanymi wcześniej.
Otrzymujemy, że

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij}\beta_{ij} = 2(R - a) \tag{4.4.1.}$$

oraz

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \Theta_{ij}^{(4)} = 2(R + a + b + c) \tag{4.4.2.}$$

należy zauważyć, że suma $R+a+b+c+f = \binom{n}{2}$ (4.4.3.)

Podstawiając (4.4.3.) otrzymuję:

$$\mu_o^{(4)} = \frac{R - a}{R + a + b + c} \tag{4.4.}$$

bądź uwzględniając (4.4.3.) otrzymuję:

$$\mu_o^{(4)} = \frac{R - a}{\binom{n}{2} - f} \tag{4.4a.}$$

Wzór (4.4a.) jest najwygodniejszy do obliczania wartości współczynnika $\mu_o^{(4)}$. Jaka jest różnica między współczynnikami τ M. Kendalla, a $\mu_o^{(4)}$ L. Guttmana dla słabych uporządkowań.

$$\tau = \frac{R - a}{\left\{ \left(\binom{n}{2} - f - c\right)\left(\binom{n}{2} - f - b\right) \right\}^{1/2}} \tag{2.6a.}$$

Liczniki obydwu wzorów są takie same. Natomiast różność mianowników powoduje, że jeżeli $c \neq 0$ lub $b \neq 0$, to $\tau \neq \mu_o^{(4)}$ (przy czym τ jest wtedy najczęściej liczbą niewymierną). Współczynnik $\mu_o^{(4)}$ nie uwzględnia liczb półinwersji $b+c$ wogóle, natomiast współczynnik τ uwzględnia je tylko w mianowniku i wiąże dla obu zmianach multiplikatywnie.

4.3. Porównanie współczynnika $\mu_o^{(4)}$ L. Guttmana oraz τ^* Z. Zimnego dla uporządkowania słabych

Stosując powyższą analizę do współczynnika τ^* Z. Zimnego dla słabych uporządkowań mamy:

$$I_w = a + \frac{1}{2}(b+c) \tag{2.7.}$$

oraz

$$I_{max} = \binom{2}{n} - \frac{1}{2}(c+f) - \frac{1}{2}(b+f)$$

czyli

$$I_{max} = R + a + \frac{1}{2}(b+c) \tag{4.5.}$$

Podstawiając (2.7.) oraz (4.5.) do (2.1a.) otrzymuję:

$$=1 - \frac{2a+b+c}{R+a+\frac{1}{2}(b+c)} = \frac{R - [a+\frac{1}{2}(b+c)]}{R+[a+\frac{1}{2}(b+c)]} \tag{4.6.}$$

Dla $b=0$ oraz $c=0$ $\qquad \tau = \tau^* = \mu_o^{(4)} = \dfrac{R-a}{R+a}$

Dla $b \neq 0$ oraz $c \neq 0$ $\qquad \tau \neq \tau^* \neq \mu_o^{(4)} \neq \dfrac{R-a}{R+a}$

Współczynnik τ^* przyjmuje wartość zero, gdy $R = a + \frac{1}{2}(b+c)$,

a współczynnik τ oraz $\mu_o^{(4)}$, gdy $R=a$. Współczynnik τ, τ^* oraz $\mu_o^{(4)}$ przyjmują wartości równe wtedy i tylko wtedy, gdy $b=c=0$.
Warto zauważyć, że — wszystkie te trzy współczynniki: mogą przyjmować wartość 1 albo —1 tylko wtedy gdy $b=c=0$; — są mimo słabego uporządkowania tylko wtedy, gdy $f \neq 0$ i $b=0$ i $c=0$.

5. Systematyka miar współuporządkowania.

Systematyka miar współuporządkowania wymaga widzenia jej na tle ogólnego modelu miary siły związku między badanymi zmiennymi. Spośród prób konstrukcji ogólnego modelu miary siły związku między dwoma zmiennymi omówiłem dotąd współczynnik Γ M.G. Kendalla oraz $\eta_i^{(4)}$ L. Guttmana.

Analiza toku rozumowania obu tych autorów przy budowie wymienionych modeli pozwoliła mi przyjąć za podstawę do własnej analizy zbiór możliwych uporządkowań par elementów według dwu zmiennych. Z opartej na tym zbiorze analizy porównawczej współczynników τ M.G. Kendalla i τ^* Z.M. Zimnego w odniesieniu do uporządkowań słabych wynika między innymi, że: współczynniki τ i τ^* podobnie jak np. współczynnik r K. Pearsona i wiele innych (współczynnik $\mu_i^{(m)}$ L. Guttmana nie spełnia tego warunku) są unormowane przez średnią z miar zmienności dwu zmiennych, oraz wynika dalej, że M.G. Kendall przyjmuje jako tę średnią, średnią geometryczną, natomiast Z.M. Zimny — średnią arytmetyczną. Porównanie innych współczynników potwierdza istnienie dwu klas współczynników siły związku, a mianowicie klasy współczynników:

1. z miarą unormowaną przez średnią arytmetyczną miar zmienności obu zmiennych, a więc z addytywnym połączeniem obu miar zmienności,
2. z miarą unormowaną przez średnią geometryczną miar zmienności obu zmiennych, a więc z multiplikatywnym połączeniem obu miar zmienności.

Do pierwszej klasy współczynników zaliczamy:

— dla skal klasyfikacyjnych (dwu zmiennych dychotomicznych)

a) współczynnik B. L. Jaxy Bykowskiego

$$B = \frac{(a+d) - (b+)}{(a+d)+(b+c)} \tag{5.1.}$$

b) współczynnik Q U. Yule'a

$$Q = \frac{ad - bc}{ad+bc} \tag{5.2.}$$

c) współczynnik H_{AB} J. Lutyńskiego

$$H_{AB} = \frac{(a+b)(b+c)+(a+c)(c+d)}{2(ad - bc)} \tag{5.3.}$$

gdzie — a, b, c, d, A, B są odpowiednimi oznaczeniami z czteropólówki

	A	$\bar{A}$	
B	a	b	a+b
B	c	d	c+d
	a+c	b+d	N=a+b+c+d

— dla skal porządkowych

a) współczynnik γ Kruskalla-Goodmana

$$\gamma = \frac{R - a}{R+a} \tag{5.4.}$$

gdzie R, a mają takie samo znaczenie jak we wzorach (4.4.), (4.6.) (2.6a.).

Współczynnik γ jest uogólnieniem współczynnika Q i może być stosowany także dla skal klasyfikacyjnych dla zmiennych dwu i wielowartościowych,

b) współczynnik τ^* Z. M. Zimnego

$$\tau^* = 1 - \frac{2I}{I_{max}} \qquad (2.1a.)$$

Do drugiej klasy współczynników zaliczamy:

— dla skal klasyfikacyjnych (dwu zmiennych dychotomicznych)

a) współczynnik φ

$$\varphi = \frac{ad - bc}{[(a+b)(a+c)(b+d)(c+d)]^{1/2}} \qquad (5.5.)$$

gdzie a, b, c, d mają takie samo znaczenie jak we wzorach (5.1.), (5.2.) oraz (5.3.)

— dla skal porządkowych

a) współczynnik τ M.G. Kendalla

$$\tau = \frac{S}{\sqrt{\left[\binom{n}{2} - \sum_{t_i=1}^{P_2}\binom{t_i}{2}\right]\left[\binom{n}{2} - \sum_{t_i=1}^{P_1}\binom{2}{t_i}\right]}} \qquad (5.6.)$$

oznaczenia jak we wzorze (2.4.) i (2.6.)

Współczynnik τ M.G. Kendalla jest uszczegółowieniem współczynnika φ dla skale porządkowej, tzn. że dla czteropolówki $\tau = \varphi$.

b) współczynnik ϱ K. Spearmana

$$\varrho = \frac{n(n^2 - 1) - 6(T_x + T_y) - 6\sum_{i=1}^{n} d_i^2}{\{[n(n^2 - 1) - 12T_x][n(n^2 - 1) - 12T_y]\}^{1/2}} \qquad (1.11.)$$

— dla skal przedziałowych i stosunkowych

a) współczynnik korelacji liniowej r K. Pearsona

$$r = \frac{cov(x,y)}{s(x)\ s(y)} \qquad (1.1)$$

Mianownik takiego ogólnego modelu jest zatem znany, jest to średnia M (arytmetyczna lub geometryczna) miar zmienności, z, dwu zmiennych x oraz y. Natomiast licznik jest różnicą między sumą, A, odległości w parach o dodatnim zwązizku, a sumą, B, odległości w parach o ujemnym związku. Oczywiście odległości te będziemy mierzyć różnie w zależności od skali pomiaru i tak:

1. na skalach stosunkowej i przedziałowej — w umownych jednostkach miary, w których mierzymy natężenie badanej właściwości w elementach.
2. na skali porządkowej jest to odległość między rangami.

3. na skali klasyfikacyjnej natomiast przyjmujemy, że odległość ta jest stała i równa 1, stąd bierzemy sumę par w których współwystępują dane własności badanych właściwości.

Model ten przyjmuje zatem postać:

$$\zeta = \frac{A - B}{M(z(x), z(y))} \qquad (5.7.)$$

Na niższym szczeblu ogólności można pokazać specyficzne modele dla skal klasyfikacyjnych i porządkowej oraz dla skal przedziałowej i stosunkowej.

Jako model dla skal klasyfikacyjnej i porządkowej możemy przyjąć:
— współczynnik τ^* Z. Zimnego wzór (4.6) (ze średnią arytmetyczną),
— współczynnik τ M.G. Kendalla (ze średnią geometryczną).

Jako model dla skal przedziałowej i stosunkowej możemy przyjąć:
— współczynnik korelacji liniowej r K. Pearsona.

skala	klasa 1	klasa 2
Klasyfikacyjna	H_{AB} B Q	φ
porządkowa	τ^* γ	τ
przedziałowa		r
stosunkowa		r

Współczynnik Γ M. Kendalla obejmuje tylko r, ϱ, τ (dla silnych uporządkowań) oraz γ natomiast $\mu_i^{(m)}$ L. Guttmana obejmuje Q, γ, oraz τ dla silnych uporządkowań.

W tej sytuacji powstaje pytanie:

Czy i przy jakich założeniach teoretycznych te dwa różne połączenia są zasadne?

Przykład 1. Dla skal silnie uporządkowanych.

x	1	2	3	4	5	6	7	8	9	10	
y	2	5	8	3	1	4	9	10	7	6	n=10

A. Współczynnik τ liczony wg wzoru (2.1.)

$$R = 8+5+2+5+5+4+1+0+0 = 30$$

$$\tau = \frac{4 \cdot 30}{10(10-1)} - 1 = \frac{1}{3}$$

B. Współczynnik τ^* liczony wg wzoru (2.1a.)

$$I = 1+3+5+1+0+0+2+2+1 = 15$$

$$I_{max} = \binom{10}{2} = 45$$

$$\tau^* = 1 - \frac{30}{45} = \frac{1}{3}$$

C. Współczynnik τ liczony wg wzoru (4.1.)

$$[\alpha_{ij}] = \begin{bmatrix} 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$[\beta_{ij}] = \begin{bmatrix} 0 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 0 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

$$[\alpha_{ij}\ \beta_{ij}] = \begin{bmatrix} 0 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 0 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 0 & -1 \\ 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

$$[\Theta^{4}_{ij}] = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \Theta^{ij}_{(4)} = 90 \qquad \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij}\beta_{ij} = 30 \qquad \mu^{(4)}_{o} = \frac{30}{90} = \frac{1}{3}$$

Przykład 2. Dla uporządkowań słabych. Przykład jest liczony wg wzorów wynikających z analizy współczynników współuporządkowania.

lp.	1	2	3	4	5	6	7	8	9	10
X	1	3	3	3	5,5	5,5	8	8	8	10
Y	3	1,5	1,5	4,5	4,5	7,5	7,5	9,5	9,5	6

$R=7+6+6+5+4+2+0+0+0+0=30$

$a=2+0+0+0+0+1+1+1+1+0=6$

$b=0+1+1+0+1+0+2+0+0+0=5$

$c=0+0+0+1+0+1+0+0+0+0=2$

$f=0+1+0+0+0+0+0+1+0+0=2$

A. Współczynnik τ liczony wg wzoru (2.6a.)

$$\tau=\frac{30-6}{\sqrt{(45-2-2)(45-2-5)}}=0{,}6085$$

B. Współczynnik τ^* liczony ze wzoru (4.6.)

$$\tau^*=\frac{30-6-0{,}5(5+2)}{30+6+0{,}5(5+2)}=0{,}5190$$

C. Współczynnik $\mu_o^{(4)}$ liczony ze wzoru (4.4.)

$$\mu_o^{(4)}=\frac{30-6}{30+6+5+2}=0{,}5580$$

D. Współczynnik γ liczony wg wzoru (5.4.)

$$\gamma=\frac{30-6}{30+6}=0{,}6667$$

LITERATURA

1 Anderson T. W. The statistical Analysis of Time-Series, London 1971.

2 David F. N. Tables of Correlation Coeficient, London 1938.

3 Góralski A. Metody opisu i wnioskowania statystycznego w psychologii, Warszawa 1974.

4 Guilford J. P. Podstawowe metody statystyczne w psychologii i pedagogice, Warszawa 1960.

5 Kendall M. G. Rank Correlation Methods, London 1955.

6 Kendall M. G. Buckland W. R. Słownik terminów statystycznych, Warszawa 1975.

7 Metody statystyczne w socjologii praca zb. pod red. K. Szaniawskiego, Warszawa 1968.

8 Pawłowski Z. Statystyka matematyczna, Warszawa 1976.

9 Shye S. Theory Construction and Data — Analysis in the Social Sciences, San Franciska 1978.

10 Siegel S. Nonparametric Statistics for the Behavioral Sciences, Nowy York 1956.

11 Yule U. G. Kendall M. G. An Introduction to theTheory of Statistics, London 1958.

12 Zimny Z. M. Wprowadzenie do psychologii, Katowice 1974.

13 Zimny Z. M. Semiinversions and Their Applications, Częstochowa 1979.

T. M. Zimny

IN THE MATTER OF GENERAL SYSTEMATICS OF THE MEASURES OF COORDINATION

Summary

The comparative analysis of some coordination coefficients indicated that the Kendall's τ, the Zimny's τ^* and Guttman's $\mu_0^{(4)}$ are not reduceable to the Spearman's ϱ.

From among the three coeficients τ, τ^* and $\mu_0^{(4)}$ only τ^* is a wright one, because it is consequently based on semiinversions.

In comparing some coefficients of general closeness of relationship specially the coefficient Kendall's Γ and Guttman's $\mu_i^{(m)}$ (where $m = 1, 2, 3, 4$ and $i = 0, 1, 2$) the author showed that none of them is a universal one and proposed a universal model of such a coefficient i.e. the coefficient

$$\zeta = \frac{A - B}{M[z(x),\ z(y)]}$$

where: A — the sum of distanses between two positive related elements in each pair

B — the sum of distanses between two negative related elements in each pair

$M[z(x), z(y)]$ — mean of the measures of variability $z(x)$, $z(y)$ of two variables x and y.

which is more general than the two former ones Γ and $\mu_i^{(m)}$.