Rocznik Naukowo-Dydaktyczny WSP w Krakowie

Zeszyt 176 *Prace Psychologiczne V* 1996

Leszek Wrona

Podobieństwo psychologicznych cech jakościowych

Streszczenie

Określanie stopnia podobieństwa obiektów za pomocą zaproponowanego wskaźnika jest możliwe jedynie w przyjętej przez badacza dziedzinie podobieństwa, rozumianej jako zbiór wziętych pod uwagę możliwych cech dychotomicznych. Wspomniany wskaźnik posiada prostą interpretację trygonometryczną, da się przekształcić w miarę odległości taksonomicznej oraz w miary kontyngencji oparte o statystykę Chi^2. Statystyczne testowanie podobieństwa jest możliwe poprzez porównywanie proporcji testem t Studenta, bądź dla dużych dziedzin podobieństwa testem Chi^2. Przedstawiony wskaźnik może znaleźć zastosowanie w psychologii klinicznej, w wypadku konieczności porównywania dwóch osób lub wyników dwukrotnych badań tej samej osoby. W szczególnych przypadkach można stosować go jako miarę kontyngencji.

Rozwój statystycznych metod klasyfikowania obiektów i cech psychologicznych wymaga stosowania odpowiednich kryteriów klasyfikacji. Kryteriami tymi są często podobieństwo lub odległość taksonomiczna. Pożądane jest, by dobrane kryteria odznaczały się prostotą ułatwiającą interpretację psychologiczną zarówno ich wartości, jak i klasyfikacji na nich opartych. W szczególności, gdy dane kryterium podobieństwa da się formalnie przekształcić w miarę odległości taksonomicznej, stanowiącej określoną metrykę przestrzeni porównywanych obiektów, to jego stosowanie prowadzi do klarownych i przekonywających rezultatów.

Literatura psychologiczna i metodologiczna traktująca o powyższej kwestii zawiera interesujące rozwiązania w postaci propozycji różnorakich wskaźników podobieństwa, jak i miar odległości wraz z ich uzasadnieniem formalnym oraz interpretacją (Zakrzewska 1987).

Stopień ogólności oraz zakres interpretacji przedstawianych wskaźników jest na ogół ograniczony. Dlatego w zależności od potrzeb badawczych, psycholog wybiera taki, który uzna za najbardziej odpowiedni dla danego typu materiału oraz przyjętej metody jego analizy.

Konstrukcja jednego uniwersalnego wskaźnika, spełniającego wymogi poprawności formalnej oraz dającego szansę jednoznacznych interpretacji jest problemem niezwykle trudnym do rozwiązania, o ile takie rozwiązanie jest możliwe.

Jeśli naszym zadaniem jest zbudowanie wskaźnika podobieństwa, to musimy przede wszystkim kierować się matematyczną definicją podobieństwa, tak by utworzony wskaźnik był poprawny formalnie. Kryteria definicyjne są jednak na tyle ogólne, że nie przesądzają kształtu wyrażenia zadającego taki wskaźnik. Stąd też otwiera się możliwość budowania wielkiej liczby wyrażeń matematycznych, spełniających wprawdzie warunki definicyjne, lecz praktycznie bezużytecznych. Dodatkową komplikacją jest uzasadnione dążenie do konstrukcji takiego wskaźnika, który dałby się interpretować samoistnie, bez względu na jego dalsze zastosowania. Interpretacja taka nie jest pełna, jeśli wskaźnik nie jest przedstawiony jako statystyka, czyli nie jest znany jego rozkład. Pociąga to za sobą brak możliwości testowania hipotez statystycznych. Odkrycie i opis takiego rozkładu może nastręczać poważne trudności, jeśli analizowany wskaźnik dany jest złożoną formułą matematyczną. Niekiedy trudność tę można przezwyciężyć poprzez zastosowanie formuły przekształcającej wskaźnik w statystykę o znanym rozkładzie.

Dalsza część opracowania poświęcona jest wprowadzeniu i analizie wskaźnika podobieństwa obiektów psychologicznych posiadających jakościowe cechy dychotomiczne.

Wspólność cech jako kryterium podobieństwa

Relację podobieństwa obiektów „i" oraz „j" określa definicja orzekająca, że relacja ta jest zwrotna i symetryczna. Jeśli istnieje wskaźnik podobieństwa S, to zachodzą relacje:

1. $i\,S\,i$

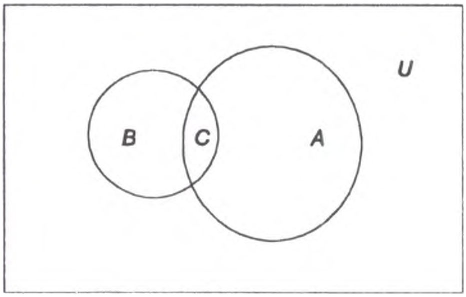
2. $i\,S\,j = j\,S\,i$

Posiadanie części wspólnej przez dwa obiekty będące zbiorami cech spełnia definicję podobieństwa (Szrejder 1975). Po pierwsze, zbiór (i) posiada sam ze sobą część wspólną równą (i), po wtóre, część wspólna zbiorów (i) oraz (j) pozostaje taka sama bez względu na kolejność porównywania tych zbiorów. Jeśli do powyższej definicji dołączymy warunek przechodniości, to relacja podobieństwa przechodzi w relację równoważności. Oznacza to, że zgodnie z przyjętym kryterium oba zbiory są równe i pokrywają się.

Dla przejrzystości wywodu zanalizujemy fikcyjny przykład badania psychologicznego, polegającego na wyborze przez podmiot badany dowolnej liczby przymiotników z przygotowanej uprzednio listy. Wybór jest dokonywany dwukrotnie. Pierwszy raz badany wybiera przymiotniki przedstawiające cechy, które chciałby posiadać (idealne), powtórnie zaś wybiera przymiotniki odnoszące się do cech aktualnie posiadanych (realnych). Można zapytać o wielkość podobieństwa realnego i idealnego obrazu jednostki. Najprostszą informacją o podobieństwie obu obrazów jest liczba cech wspólnych – c, która stanowi ilość elementów należących do zbioru cech wspólnych C (rys. 1). Zbiór cech wspólnych jest iloczynem mnogościowym zbiorów A i B, przy czym:

A jest zbiorem cech wybranych za pierwszym razem,

B jest zbiorem cech wybranych powtórnie.



Rys. 1. Dziedzina podobieństwa, dwa zbiory cech A, B

Wybory dokonywane są ze skończonej listy cech tworzącej zbiór U nazywany dalej dziedziną podobieństwa lub przestrzenią podobieństwa. Do zbioru U należy n elementów (cała lista), odpowiednio do zbiorów: A – a elementów, B – b elementów. Rozważanie liczby cech wspólnych traci sens, jeśli zbiory A i B są puste. Podobieństwo oparte na wspólnocie cech można rozważać jedynie wtedy, gdy zarówno A, jak i B posiadają co najmniej po jednym elemencie. W przeciwnym wypadku zbiór C jest zawsze pusty. Z podobnych względów dziedzina U również nie może być pusta, co więcej, musi posiadać przynajmniej dwa elementy.

Liczba cech wspólnych nie wnosi informacji o tym, jak wielka jest dziedzina podobieństwa oraz zbiory A i B. Jeżeli na przykład $c = 1$, to nie wiadomo czy badany wybrał w obu etapach jedną i tylko jedną tę samą cechę, czy też spośród wybranych cech ta jedna jest wspólna. Dlatego zatem wskaźnik podobieństwa powinien informować o relacji między zbiorami A, B, U. Jeśli wymóg ten nie jest spełniony, to wypadałoby uznać, że ktoś kto wybrał jedną i tylko jedną identyczną cechę w obu etapach, charakteryzuje się takim samym podobieństwem swych obrazów, idealnego i realnego, jak ten, który wybrał za każdym razem po 50 cech, z czego tylko jedna jest wspólna. Z psychologicznego punktu widzenia właśnie u drugiej osoby podobieństwo obu obrazów powinno być mniejsze.

Unormowany wskaźnik podobieństwa

Dwukrotny wybór przymiotników z podanej listy może prowadzić do dwóch różnych rezultatów, które trzeba przeanalizować osobno. Po pierwsze, łączna liczba dwukrotnie wybranych przymiotników może być co najwyżej równa n przy uwzględnieniu części wspólnej ($a + b \leq n$), po drugie, łączna ilość przymiotników wybranych w tenże sposób może przekraczać n ($a + b > n$).

Weźmy na początek pod uwagę przypadek pierwszy. Dane przedstawione na rysunku 1 można ująć w tablicy kontyngencyjnej (tabl. 1).

Tablica 1. Kontyngencja cech

		j		
		B	$\bar{B}$	
i	A	c	$a-c$	a
	$\bar{A}$	$b-c$	$n-a-b+c$	$n-a$
		b	$n-b$	n

Tablicę tę można traktować jako macierz $[ij]$, jeśli pominiemy sumy brzegowe. Macierz tę przekształcamy poprzez dodanie pierwszego wiersza do drugiego, a następnie pierwszej kolumny do drugiej. Przekształcenie to przedstawia się następująco:

$$1.\quad \begin{array}{cc} & j \\ i & \begin{bmatrix} c & a-c \\ b-c & n-a-b+c \end{bmatrix} \end{array} = \begin{array}{cc} & j \\ i & \begin{bmatrix} c & a-c \\ b & n-b \end{bmatrix} \end{array} = \begin{array}{cc} & j \\ i & \begin{bmatrix} c & a \\ b & n \end{bmatrix} \end{array}$$

Iloczyn krzyżowy, czyli wyznacznik $[ij]$ ostatniej macierzy jest równy $nc - ab$. Jeśli wyznacznik ten pomnożymy przez $4/n^2$, to otrzymamy:

$$1.1 \qquad \frac{4}{n^2}|ij| = \frac{4}{n^2}(nc - ab) = \frac{c - \frac{ab}{n}}{\frac{n}{4}} = \frac{c - E(c)}{max\,E(c)} = S_{ij}$$

Interpretacja tak zdefiniowanego wskaźnika podobieństwa S_{ij} jest bardzo prosta. Wyrażenie *ab/n* oznacza wartość oczekiwaną *E/c/* części wspólnej zbiorów *A* i *B*, natomiast *n/4* jest maksymalną wartością oczekiwaną części wspólnej (*maxE/c/*). Obie wartości oczekiwane dane są z rozkładu hipergeometrycznego (Feller 1981). Wskaźnik ten jest więc stosunkiem ilości cech wspólnych powyżej, bądź poniżej ilości generowanej losowo do maksymalnej liczby oczekiwanej cech wspólnych dla danej dziedziny podobieństwa. Jeżeli zatem *c* jest równe *E/c/* przy określonych *a, b, c*, to wskaźnik przyjmuje wartość zero. Oznacza to, że zarówno podobieństwo, jak i zróżnicowanie obiektów wynika z przypadku. Gdy natomiast *c* jest mniejsze od *E/c/*, to wskaźnik jest liczbą ujemną, co świadczy o przewadze zróżnicowania nad podobieństwem. Odpowiednio, gdy *c* jest większe od *E/c/*, wskaźnik staje się dodatni i ukazuje przewagę podobieństwa nad zróżnicowaniem. Spełnienie warunku unormowania i uwzględniania relacji między *A, B, U*, powoduje, iż skrajnymi wartościami wskaźnika są –1 i 1. Wartości te pojawiają się w dwóch przypadkach:

+1; gdy liczby *a, b* i *c* są wszystkie równe *n*/2,

–1; gdy *a* i b są równe n/2 oraz c jest równe zero.

Skrajne podobieństwo czyli nierozróżnialność obiektów w dziedzinie *U* zachodzi wtedy, gdy *a, b, c* są sobie równe i wszystkie dwukrotnie przekraczają maksymalną wartość oczekiwaną ilości cech wspólnych w tej dziedzinie. Skrajne niepodobieństwo, czyli odmienność ma natomiast miejsce wtedy, gdy zarówno *a*, jak i *b* dwukrotnie przekraczają wspomnianą wartość oczekiwaną, natomiast liczba cech wspólnych wynosi zero.

Drugi przypadek, czyli przekroczenie przez łączną ilość wybranych cech liczby *n*, wymaga uwzględnienia we wzorze 1 poprawki z uwagi na możliwość wymuszenia wyboru pełnej liczby cech wspólnych przez ograniczoną wielkość dziedziny podobieństwa. Wymuszenie *„w"* pojawia się na skutek tendencji badanego do wybierania dużej ilości przymiotników. Jeśli tendencja ta wystąpi w obu wyborach, to z konieczności niektóre przymiotniki muszą się powtórzyć. Oczywiście nie sposób rozstrzygnąć, czy u konkretnego badanego tendencja taka wystąpiła, lecz ostrożność wymaga odpowiedzi na pytanie: jak kształtowałby się wskaźnik podobieństwa, gdyby wyeliminować wymuszenie?

Wymuszenie w trzeba uwzględnić zarówno w części wspólnej zbioru C, jak i wartości oczekiwanej $E/c/$ przy danych a i b. Łatwo to uzyskać, wiedząc że „w" jest równe $n - a + b$. Wtedy korekta c polega na odjęciu od tej liczby wartości „w". Otrzymanie skorygowanej wartości a_w i b_w wymaga proporcjonalnego zmniejszenia a i b o takie wielkości, aby suma a_w i b_w była równa n. Zgodnie więc z przyjętym wymogiem

1.2 $$a_w = a - \frac{a \cdot w}{a+b} = \frac{a\,n}{a+b}$$

1.3 $$b_w = b - \frac{a \cdot w}{a+b} = \frac{b\,n}{a+b}$$

Wtedy skorygowany wskaźnik podobieństwa S_{ijw} wyraża się formułą:

2. $$S_{ijw} = \frac{h}{n^2} \times \begin{vmatrix} c - w & \frac{a\,n}{a+b} \\ \frac{b\,n}{a+b} & n \end{vmatrix} = \frac{c - w - \frac{a\,b\,n}{(a+b)^2}}{\frac{n}{4}} = \frac{c - w - E_w(c)}{max\,E(c)}$$

Interpretacja skorygowanego wskaźnika, jak i granice jego zmienności pozostają tu takie same.

Relatywizacja wskaźnika podobieństwa

Niekiedy wielkość wskaźnika podobieństwa jest trudna do interpretacji, szczególnie wówczas, gdy a oraz b są małe w stosunku do n. Można postawić pytanie, jak zmieniłaby się wysokość S_{ij}, gdyby porównać ją z wielkościami ekstremalnymi dla określonych a, b, n. Wielkość ekstremalna jest podstawą do określania wysokości S_{ij} względem wysokości możliwej w zastanym układzie danych. Jeśli S_{ij} jest dodatni, to przyjmuje maksymalną wartość dla $a = b = c = a + b/2$, natomiast gdy jest mniejszy od zera, to jego wartość minimalna wyznaczona jest przez warunek: $a = b = a + b/2$ oraz $c = 0$.

Zatem maksymalny S_{ij} dany jest wzorem:

2.1 $$S_{ij\,max} = \frac{4}{n^2} \times k \begin{vmatrix} \frac{a+b}{2} & l & \frac{a+b}{2} \\ \frac{a+b}{2} & & n \end{vmatrix} = \frac{4}{n^2} | kl |$$

natomiast minimalny:

2.2 $$S_{ij\,min} = \frac{4}{n^2} \times k \begin{vmatrix} 0 & l' & \frac{a+b}{2} \\ \frac{a+b}{2} & & n \end{vmatrix} = \frac{4}{n^2} | k'l' |$$

Ostatecznie zrelatywizowany dodatni wskaźnik podobieństwa ma postać:

3. $$S_{ijr} = \frac{S_{ij}}{S_{ij\,max}} = \frac{| ij |}{| kl |}$$

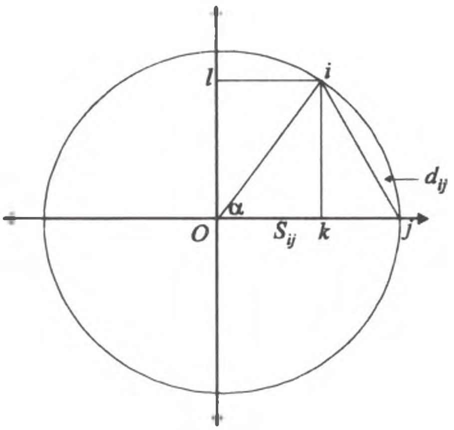
z kolei ujemny jest wyrażany wzorem:

4. $$S_{ijr} = \frac{S_{ij}}{\left| S_{ij\,min} \right|} = \frac{| ij |}{\| k'l' \|}$$

Jak widać z wzorów 3 i 4, zrelatywizowane wskaźniki podobieństwa otrzymujemy, dzieląc przez siebie odpowiednie wyznaczniki, przy czym dla zachowania znaku, w mianowniku wzoru 4 występuje wartość bezwzględna. Interpretacja tak zdefiniowanych wskaźników jest prosta, lecz gdy przestrzeń podobieństwa jest mało liczna lepiej ich nie stosować.

Wskaźnik podobieństwa, a miara odległości

Proponowana postać wskaźnika podobieństwa ma kształt proporcji, wobec czego łatwo tu o interpretację geometryczną (rys. 2). Niech zatem w trójkącie *Oij* cosinus kąta alfa, czyli odcinek *Ok* przedstawia wskaźnik podobieństwa. Odległość punktu *i* od *j* /d_{ij}/ można obliczyć stosując przekształcenie:



Rys. 2. Interpretacja geometryczna podobieństwa

5. $$d_{ij} = 2r\sin\frac{\alpha}{2} = 2\sin\frac{\arccos S_{ij}}{2}$$

przy założeniu, że promień *r* okręgu o środku O jest równy jedności. Tak określona wielkość jest odległością taksonomiczną (metryką) w przestrzeni porównywanych obiektów.

Bez trudu można sprawdzić, że metryka ta spełnia warunki:

– minimalnej odległości; $d_{ii} = 0$,

– symetrii; $d_{ij} = d_{ji}$,

– nierówności trójkąta; $d_{ij} + d_{jk} \geq d_{ik}$.

Wynika stąd, że zbiór porównywanych obiektów z określoną w nim miarą d_{ij} stanowi przestrzeń metryczną z metryką d_{ij}. Okazuje się, że jest to przestrzeń euklidesowa, co bez trudu da się wykazać, zauważając że odcinek *ij* ma współrzędne: 1,0 kl. Wobec tego długość odcinka $ij/d_{ij}/$ dana jest z twierdzenia Pitagorasa:

5.1 $$d_{ij} = \sqrt{(1-k)^2 + (0-l)^2} = \sqrt{(1-\cos\alpha)^2 + \sin^2\alpha}$$

W zastosowaniach praktycznych miarę d_{ij} można normalizować dzieląc ją przez 2. W ten sposób otrzymamy wartości metryki nie przekraczające przedziału (0,1).

Wskaźnik podobieństwa a miary kontyngencji

Popularny w zastosowaniach współczynnik kontyngencji α Yule'a dany jest wzorem:

6. $$\phi = \frac{|ij|}{\sqrt{ab(n-a)(n-b)}}$$

przy założeniu, że $a + b = n$ bądź $a_w + b_w = n$. Stąd po oczywistych przekształceniach otrzymujemy:

7. $$\phi = \frac{n^2}{4} \times \frac{S_{ij}}{\sqrt{ab(n-a)(n-b)}}$$

i odpowiednio:

7.1 $$S_{ij} = \frac{4}{n^2} \times \phi \sqrt{ab(n-a)(n-b)}$$

Można wykazać, że wskaźnik podobieństwa przyjmuje wartość bezwzględną mniejszą niż współczynnik Yule'a. Wystarczy zauważyć, że wyrażenie pod kreską ułamkową we wzorze 8 może przyjąć maksymalną wartość $n^2/4$ jedynie wtedy, gdy $a = b = n/2$. Oczywiście $n^2/4 \times 4/n^2 = 1$, więc $S_{ij} = 1 \cdot \phi$. Jeśli jednak a jest różne od b, to wyrażenie owo jest mniejsze niż $n^2/4$. Wynika stąd nierówność $S_{ij} < \phi$.

Różnica między wskaźnikiem podobieństwa a współczynnikiem kontyngencji Yule'a ujawnia się w pełni, gdy suma a i b jest większa od n. Przykład przedstawiony w tabeli 3 pokazuje, że nieuwzględnienie poprawki na wymuszenie zmienia znak współczynnika kontyngencji i zaniża jego wartość. Tam zatem, gdzie początkowo zachodziło podejrzenie pozytywnej zbieżności, jest ona zgodnie z wzorem negatywna.

$$\begin{bmatrix} 60 & 70 \\ 80 & 100 \end{bmatrix}$$

Tablica 2. Dane bez poprawki $\phi = 0.218$

Tablica 4 przedstawia ten sam układ danych po eliminacji wymuszenia. Okazuje się, że odpowiednio zastosowane poprawki zmieniają obraz kontyngencji. Nie można stąd jednak wyciągać zbyt daleko idących wniosków. Kontyngencja mierzona klasycznym wzorem Yule'a oznacza związek między różnymi cechami w niewielkiej dziedzinie podobieństwa, a nie jak w naszym wypadku, występowanie cech wspólnych. Stosowanie zatem poprawek usuwających wymuszenie nie jest tu uzasadnione.

$$\begin{bmatrix} 10 & 46.666... \\ 53.33... & 100 \end{bmatrix}$$

Tablica 3. Dane z uwzględnieniem poprawki

$S_{ij} = -0{,}598$
$\phi = -0{,}598$

Przekształcenie wskaźnika podobieństwa w miary kontyngencji oparte o statystykę X^2 nie stwarza większych problemów. Skoro $X^2 = n\,\phi^2$ to:

8. $$\chi^2 = \frac{n^5}{16} \times \frac{S_{ij}^2}{ab(n-a)(n-b)} = \frac{n^3\left(c - \frac{ab}{n}\right)^2}{ab(n-a)(n-b)}$$

Współczynniki zbieżności obliczane z tej statystyki są powszechnie znane, dlatego nie ma potrzeby ich przytaczania (Blalock 1975). Zwróćmy raczej uwagę na możliwość testowania wskaźnika podobieństwa poprzez statystykę X^2 przy wartościach krytycznych ustalonych dla jednego stopnia swobody. X^2 oblicza się jedynie wtedy, gdy dziedzina podobieństwa jest dostatecznie duża. W praktyce testowanie hipotezy zerowej ($S_{ij} = 0$) jest możliwe dla *n* co najmniej równego 50.

Testowanie wskaźnika podobieństwa w małych dziedzinach

Rozkład przedstawionego wskaźnika podobieństwa jest złożony. Jego odchylenie od rozkładu normalnego jest istotne nawet dla *n* równego 300. Nie wydaje się zatem sensowne dokonywanie skomplikowanych przekształceń, aby rozkład ten znormalizować. Wygodniej dla małych *n* badać istotność wskaźnika statystyką t Studenta, stosując ją jak w wypadku porównywania proporcji otrzymanej ze standardem równym 0. Jeśli pominiemy znak algebraiczny współczynnika i posłużymy się jego wartością bezwzględną, to możemy skorzystać z przekształcenia:

9. $$X_i = 2\,arc\,sin\sqrt{|S_{ij}|}$$

10. $$X_O = 2\,arc\,sin\sqrt{O} = O$$

Statystyka t dana jest formułą:

11. $$t = \frac{X_i - O}{S_x} \quad \text{gdzie} \quad S_x = \frac{2}{\sqrt{n}}$$

Po dokonaniu podstawień i prostych przekształceniach otrzymujemy:

12. $$t = \sqrt{n}\ arc\ sin \sqrt{\left| S_{ij} \right|}$$

Wartość krytyczną odczytujemy z tablic (Góralski 1987, s.325).

***Dodatek 1.* Przeliczenie wskaźnika podobieństwa na miarę odległości**

$$d_{ij} = sin \frac{arc\ cos\ S_{ij}}{2}$$

S_{ij}	.00	–.01	–.02	–.03	–.04	–.05	–.06	–.07	–.08	–.09	–.10
						d_{ij}					
–.90	.975	.977	.982	.982	.985	.987	.990	.992	.985	.997	1.00
–.80	.949	.951	.955	.957	.960	.962	.964	.967	.970	.972	
–.70	.922	.925	.927	.930	.933	.935	.938	.941	.943	.946	
–.60	.894	.897	.900	.903	.906	.908	.911	.914	.917	.919	
–.50	.866	.869	.872	.875	.877	.880	.883	.886	.889	.892	
–.40	.837	.840	.843	.846	.849	.851	.854	.857	.860	.863	
–.30	.806	.809	.812	.815	.819	.822	.825	.828	.831	.834	
–.20	.775	.778	.781	.784	.787	.791	.794	.797	.800	.803	
–.10	.742	.745	.748	.752	.755	.758	.762	.765	.768	.771	
.00	.707	.711	.714	.718	.721	.725	.728	.731	.735	.738	
S_{ij}	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10
.00	.707	.704	.700	.696	.693	.689	.686	.682	.678	.675	
.10	.670	.667	.663	.660	.656	.652	648	.644	.640	.636	
.20	.632	.628	.624	.620	.616	.612	.608	.604	.600	.596	
.30	.592	.587	.583	.579	.574	.570	.566	.561	.557	.552	
.40	.548	.543	.539	.534	.530	.524	.520	.515	.510	.505	
.50	.500	.495	.490	.485	.480	.474	.469	.464	.458	.453	
.60	.447	.442	.436	.430	.424	.418	.412	.406	.400	.394	
.70	.387	.381	.374	.367	.361	.354	.346	.339	.332	.324	
.80	.316	.308	.300	.292	.283	.274	.265	.255	.245	.235	
.90	.224	.212	.200	.187	.173	.158	.141	.122	.100	.071	.00

1. *Spośród 300 przymiotników (nazw cech osobowości) badany wybrał 70 jako nazwy cech realnych, 80 jako nazwy cech idealnych, przy czym 30 pozycji wybrał dwukrotnie. Zbadać podobieństwo obrazu idealnego i realnego oraz odległość między nimi.*

Rozwiązanie: Przestrzeń podobieństwa posiada $n = 300$ cech, natomiast liczby cech w kolejnych wyborach wynoszą: $a = 70$, $b = 80$ przy czym $c = 30$ to liczba pozycji powtarzających się. Skoro $a + b < n$, stosujemy wzór 1.1 bez poprawki na wymuszenie. Otrzymujemy więc:

$$S_{ij} = \frac{4}{300^2}(300 \cdot 30 - 70 \cdot 80) = 0{,}151$$

Badamy istotność wskaźnika testem t Studenta danym z wzoru 12.

$$t = \sqrt{300} \cdot arc\ sin\sqrt{0{,}151} = 6{,}93$$

Odległość obrazów odczytujemy z tablicy zamieszczonej w dodatku 1. Wynosi ona 0,652. Podobieństwo obu obrazów jest zatem pozytywne (przewaga podobieństwa nad zróżnicowaniem) i bardzo istotne ($p < 0{,}001$), choć słabe. Odległość obu obrazów od siebie jest znaczna. Jeśli interesuje nas zrelatywizowany wskaźnik podobieństwa, to stosujemy wzór 3.

$$S_{ij} = \frac{\begin{vmatrix} 30 & 70 \\ 80 & 300 \end{vmatrix}}{\begin{vmatrix} 75 & 75 \\ 75 & 300 \end{vmatrix}} = \frac{30 \cdot 300 - 70 \cdot 80}{75 \cdot 300 - 75 \cdot 75} = 0{,}201$$

Okazuje się więc, że relatywizacja tego wskaźnika do jego wartości maksymalnej dla przyjętych danych podwyższa go w niewielkim stopniu.

2. *Badany dokonał wyborów z tej samej listy w taki sposób, że* $a = 200$, $b = 150$, $c = 100$. *Zbadać podobieństwo obu obrazów.*

Rozwiązanie: Wymuszenie $w = 350 - 300 = 50$. Posłużmy się wzorem 2.

$$S_{ijw} = \frac{150 - 50 - \dfrac{100 \cdot 150 \cdot 30}{(150 + 200)^2}}{\dfrac{300}{4}} = 0{,}177$$

Tak jak w poprzednim przykładzie, obliczamy t Studenta i odczytujemy odległość z tablicy. Wartość *t* wynosi 7,52, a odległość 0,640. Wyniki świadczą o pozytywnym, bardzo istotnym statystycznie, lecz słabym podobieństwie. Odległość między obrazem idealnym a realnym jest znaczna.

Łatwo zauważyć, że procedura badania podobieństwa jest bardzo prosta. Jej istotę stanowi zastosowanie odpowiedniego wzoru. Jeśli z danych wynika, że wymuszenie nie istnieje, to stosuje się wzór 1.1, w przeciwnym wypadku musimy posłużyć się wzorem 2. Wskaźnik zrelatywizowany obliczamy jedynie wtedy, gdy nie ma wymuszenia, ponieważ dla danych z wymuszeniem maksymalna wartość wskaźnika podobieństwa równa jest jedności.

Uwagi końcowe

Przedstawiony w niniejszym opracowaniu wskaźnik podobieństwa można stosować dla par obiektów z wyodrębnionymi cechami dychotomicznymi w określonej dziedzinie podobieństwa. Jeśli wspomniana dziedzina posiada więcej cech niż suma pochodząca z danej pary to postać, jak i interpretacja wskaźnika jest prosta. Dla niewielkich *n* szansa wyczerpania całej dziedziny podobieństwa rośnie, a w szczególnym wypadku ilość dwukrotnie wybranych cech może przekroczyć wielkość dziedziny. Stosowanie poprawek usuwających zachodzące wówczas wymuszenie może budzić wątpliwości interpretacyjne. Nie istnieje bowiem reguła, wedle której można byłoby rozstrzygnąć, czy wymuszenie zachodzi realnie. Lepiej zatem, w razie wątpliwości, rozszerzyć dziedzinę podobieństwa, niż stosować zabiegi formalne.

Literatura

Blalock H.M., *Statystyka dla socjologów*. Warszawa 1975

Feller W., *Wstęp do rachunku prawdopodobieństwa*. Warszawa 1981

Góralski A., *Metody opisu i wnioskowania statystycznego w psychologii i pedagogice*. Warszawa 1987

Lazarsfeld P., *Algebra systemów dychotomicznych*, Warszawa 1968

Marek T., Noworol C., *Wprowadzenie do analizy skupień*. Kraków 1983

Marek T., Noworol C., *Zarys analizy skupień*, w: Brzeziński J. (red.) *Wielozmiennowe modele statystyczne w badaniach psychologicznych*. Warszawa – Poznań 1987

Szrejder J.A., *Równość, podobieństwo, porządek*, Warszawa 1975

Zakrzewska M., *O miarach podobieństwa obiektów i cech przydatnych w psychologicznych zastosowaniach analizy skupień*, w: Brzeziński J. (red.) *Wielozmiennowe modele...*, op.cit.

Leszek Wrona

Similarity of Psychlogical Qualitative Features

Summary

Describing the degree of similarity of objects with use of the proposed factor is possible only within the domain of similarity defined by the researcher and understood as a set of possible dichotomic features under consideration. The above mentioned factor has a simple trigonometric interpretation and can be transformed into the measure of taxonomic distance and measures of contingency based of the Chi^2 statistics. It is possible to test similarity statistically by comparing proportions with use of t – Student's test or in case of large domains of similarity with the Chi^2 test. The described factor may be employed in clinical psychology when the necesity of comparing two people or results of two tests of the a person. In particular cases it may by used as a measure of contingency.