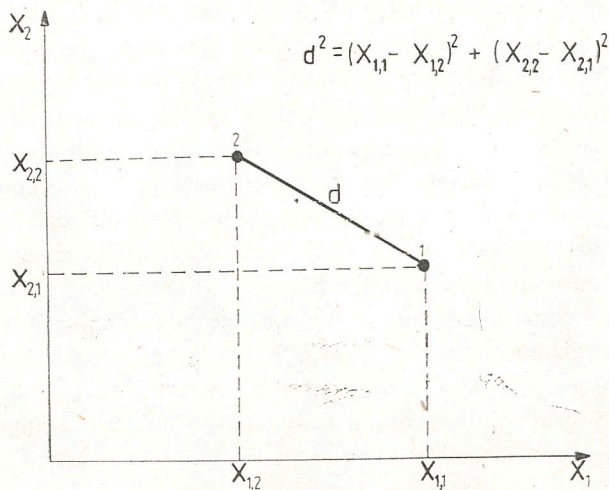


MACIEJ HENNEBERG

WIARYGODNOŚĆ WIELOCECHOWYCH MIAR ODLEGŁOŚCI
GEOMETRYCZNEJ ZE SZCZEGÓLNYM UWZGLĘDNIENIEM
ODLEGŁOŚCI PENROSE'A

Dwa obiekty scharakteryzowane za pomocą wartości jednej zmiennej odległe są od siebie o wielkość równą różnicy pomiędzy wartością zmiennej charakteryzującą pierwszy obiekt i wartością charakteryzującą drugi obiekt. Jeśli opis obiektów opiera się na wartościach dwu zmiennych, wówczas najmniejsza, wzdłuż linii prostej, odległość pomiędzy nimi równa się pierwiastkowi sumy kwadratów różnic wartości jednej i drugiej zmiennej charakteryzującej te obiekty. Ilustruje to rysunek 1, na którym opis ma geometryczną postać dwuwymiarową w klasycznym prostokątnym układzie współrzędnych. Jak widać, idea odległości geometrycznej wywodzi się z twierdzenia Pitagorasa.

Przy charakteryzowaniu obiektów za pomocą większej liczby zmiennych postępuje się podobnie, tyle że obraz geometryczny leży w przestrzeni o takiej liczbie wymiarów, jaka jest liczba zmiennych używanych dla scharakteryzowania obiektów. Układy zmiennych prostokątnych o więcej niż trzech wymiarach trudno sobie wyobrazić, niemniej logika postępowania jest taka sama jak przy określaniu odległości w ukła-



Rys. 1. Geometryczny obraz odległości pomiędzy dwoma obiektami (1 i 2), z których każdy scharakteryzowany jest wartościami dwóch zmiennych (x_1 i x_2)

dzie dwu- czy trójwymiarowym. Zatem ogólna postać odległości geometrycznej jest następująca:

$$d_{1,2}^2 = (x_{1,1} - x_{1,2})^2 + \dots + (x_{i,1} - x_{i,2})^2 + \dots + (x_{m,1} - x_{m,2})^2 \quad (1)$$

gdzie: $d_{1,2}$ – odległość geometryczna pomiędzy obiektami 1 i 2, $x_{1,1}$ – pierwsza cecha pierwszego obiektu, $x_{i,2}$ – i -ta cecha drugiego obiektu, m – liczba cech.

Można to przedstawić krócej dla odległości pomiędzy dowolnymi obiektami a i b :

$$d_{a,b}^2 = \sum_{i=1}^m (x_{i,a} - x_{i,b})^2 \quad (1a)$$

Powyższa zależność jest spełniona, jeśli poszczególne zmienne są względem siebie całkowicie ortogonalne, to jest, gdy ich wartości nie korelują z sobą. W praktyce często napotyka się związki powodujące nieortogonalność, co wymaga „prostowania” osi reprezentujących wartości poszczególnych zmiennych za pomocą odpowiednich zabiegów rachunkowych (np. odwracanie macierzy kowariancji w metodzie Mahalanobisa – Anderson [1958]). Wówczas obliczenie odległości jest obciążone dodatkowymi komplikacjami, jest to jednak kwestia odrębna, tutaj rozpatrzmy jedynie „czyste” odległości geometryczne, zakładając, że warunek ortogonalności zmiennych jest spełniony.

W szeregu dyscyplin naukowych wyznacza się odległości pomiędzy obiektami zbiorowymi, składającymi się z licznych pojedynczych elementów. Bardzo często postępuje się tak właśnie w biologii, toteż ograniczę się tutaj do przykładów z tej dziedziny, jakkolwiek identyczne formalnie postępowanie dotyczy badań ekonomicznych, demograficznych, geomorfologicznych, psychologicznych itp.

Wartości zmiennych charakteryzujących obiekty zbiorowe używane do obliczania odległości stanowią na ogół oszacowania miar położenia z prób o skończonej liczebności. Taki sposób charakteryzowania obciążony jest, jak wiadomo, błędem wynikającym z rozproszenia wartości jednostkowych wokół miary położenia oraz z wielkości próby – liczby pomierzonych jednostek. Porównywanie takich charakterystyk liczbowych obiektów zbiorowych celem ustalenia odległości geometrycznej pomiędzy obiektami obciążone jest więc nieuchronnie błędem. Odległości pomiędzy obiektami zbiorowymi, oparte na oszacowaniach wartości poszczególnych zmiennych z prób, nie są wielkościami pewnymi, ale również oszacowaniami obciążonymi określonym błędem. Dokładność takiego oszacowania, a zatem i wielkość jego błędu, zależy od rzetelności oszacowania miar położenia. Ma to szczególne znaczenie wówczas, gdy obliczamy odległości pomiędzy szeregiem obiektów celem ich uporządkowania pod względem wzajemnego podobieństwa. Klasycznym przykładem może tu być ustalanie odległości „biologicznych”, opartych na szeregu cech metrycznych organizmów, pomiędzy grupami ludzkimi, dla ustalenia stopnia wzajemnych podobieństw i różnic pomiędzy tymi grupami. Stosowanie w tym celu odległości geometrycznej wprowadzili J. Czekanowski i jego uczniowie zgrupowani w tzw. lwowskiej szkole antropologicznej. Podlegała ona licznym modyfikacjom i ulepszeniom między innymi przez L. S. Penrose'a [1954] i A. Wankego [1955], została też znacznie

przekształcona przez P. C. Mahalanobisa [1930] w formę tzw. odległości statystycznej. Szerzej zagadnienia te opisują T. S. Constandse-Westermann [1972] i R. Knusmann [1962].

Postępowanie polega tu na pomierzeniu wartości tego samego zestawu cech (np. 10 cech czaszki) w zbiorach osobników należących do kilku (kilkunastu, kilkadziesiątu itd.) grup, obliczeniu średnich arytmetycznych każdej cechy w każdej z grup, ewentualnym ich ustandaryzowaniu na wielkości odpowiednich odchyłeń standardowych, i wyliczeniu odległości pomiędzy średnimi. Odległości te interpretuje się następnie w ten sposób, że przypisuje się największe wzajemne podobieństwo tym grupom, dla których otrzymano z rachunków najmniejszą wartość odległości, a najmniej podobne są do siebie grupy najbardziej „odległe”. Wielkość odległości traktowana jest bezkrytycznie, jako wartość idealna, absolutnie prawdziwa i rzetelna. Dla jednej tylko metody określania odległości geometrycznej – metody Penrose'a – starano się wprowadzić statystyczny test istotności pozwalający ocenić, czy obliczona odległość różni się istotnie od zera czy nie. Test ten opracowany został przez N. A. Rahmana [1962] a spopularyzowany przez autorów niemieckich [Knusmann 1967; Schwidetzky 1967; Schwidetzky, Rösing 1975; Rösing 1975; Rösing, Schwidetzky 1977], w Polsce stosowali go np. J. Piontek i M. Kaczmarek [1981]. Wspomniany test jest oparty na statystyce nieparametrycznej (Chi-kwadrat) i nie uwzględnia liczebności prób, dla których określano odległości. Jego postać jest następująca (oznaczenia według Rösinga [1975]):

$$(2) \quad D_Q = C_Q^2 \frac{m}{1-R+mR} \quad \text{por. } D_Q \text{ z wartością krytyczną } \chi^2 \text{ przy jednym stopniu swobody}$$

$$(3) \quad D_P = C_P^2 \frac{m}{1-R} \quad \text{por. } D_P \text{ z wartością krytyczną } \chi^2 \text{ przy } m-1 \text{ stopni swobody}$$

$$(4) \quad D_R = D_Q + D_P \quad \text{przy } m \text{ stopniach swobody}$$

gdzie: D_Q – istotność odległości wielkości, D_P – odległości ogólnej bez poprawki na interkorelację, D_R – z poprawką na interkorelację, C_Q^2 – odległość wielkości, C_P^2 – odległość ogólna – różnica pomiędzy odległością kształtu (C_H^2) i odległością wielkości, m – liczba cech, R – współczynnik interkorelacji pomiędzy cechami. Jak widać, przy tym samym zestawie cech wartości testu będą zawsze identyczne niezależnie od liczebności porównywanych prób. R. Knusmann [1967] zdawał sobie sprawę z tej wady, mimo to jednak zalecał „ostrożne” stosowanie przedstawionego wyżej testu (s. 138). I. Schwidetzky [1967] zbadała empirycznie wpływ liczebności próby na wielkość obliczonej odległości Penrose'a w ten sposób, że z tych samych serii czaszek brała po 1, 2, 3... itd. osobniki i na tak powstałych małych próbach obliczała odległości. Stwierdziła ona, że wielkość próby rzędu 7, a czasem mniej osobników (przy 10 - 11 cechach) jest wystarczająca dla uzyskania zadowalająco, według autorki, dokładnej oceny odległości, ponieważ podwyższanie liczebności próby do 8 i więcej osobników powoduje nieznaczne już tylko zmiany wartości obliczanych odległości. W ten sposób kwestię wpływu liczebności prób uznano za rozwiązaną i przyjęto stałe, „magiczne” wielkości istotnych statystycznie, na podsta-

wie testu Rahmana, odległości (np. 0,21, 0,33 — Roth-Lutra [1971]). Inaczej ma się sytuacja w odniesieniu do odległości statystycznej Mahalanobisa, gdzie określenie istotności uwzględnia zarówno liczbę cech jak i liczebności porównywanych prób [Anderson 1958].

W obu jednak przypadkach nie stosuje się metod wnioskowania statystycznego przy porównywaniu odległości pomiędzy sobą; omówione wyżej określanie istotności statystycznej odległości ma na celu jedynie stwierdzenie, czy dana odległość jest istotnie różna od zera. Celem niniejszej pracy jest wprowadzenie metod oceny rzetelności wartości odległości geometrycznych obliczanych z prób w taki sposób, który umożliwiłaby również porównywanie odległości pomiędzy sobą.

Zajmiemy się tutaj przede wszystkim klasyczną odległością geometryczną, nawiązując do oznaczeń używanych w metodzie Penrose'a, jakkolwiek większość przedstawianych poniżej uwag odnosi się do wszystkich metod określania odległości „biologicznych”.

Przed przystąpieniem do zasadniczego wywodu celowe jest podanie listy stosowanych dalej oznaczeń:

- \hat{C}_H^2 — oszacowana z prób odległość geometryczna
- \hat{C}_Q^2 — oszacowana z prób odległość „wielkości” Penrose'a
- \hat{C}_P^2 — oszacowana z prób odległość „ogólna” Penrose'a
- \hat{C}_R^2 — oszacowana z prób odległość „uogólniona” z poprawką na interkorelację cech
- $C_H^2, C_Q^2, C_P^2, C_R^2$ — wartości tych samych odległości po usunięciu wpływu liczebności prób
- $\zeta_H^2, \zeta_Q^2, \zeta_P^2, \zeta_R^2$ — „idealne” wartości tych samych odległości, tj. po usunięciu wpływu liczebności prób oraz liczby cech
- \bar{x} — średnia z próby
- s — odchylenie standardowe z próby, tu traktowane też jako oszacowanie odchylenia standardowego w populacji generalnej σ
- μ — średnia w populacji generalnej
- E — błąd losowy — odchylenie standardowe wartości danego parametru (wskazanego w indeksie przy E) z prób wokół wartości w populacji generalnej
- N — liczebność próby — liczba obiektów jednostkowych
- m — liczba cech

Klasyczna odległość geometryczna ma postać średniego kwadratu różnic pomiędzy znormalizowanymi na odchylenie standardowe wartościami średnich arytmetycznych cech metrycznych charakteryzujących dwa zbiory osobników

$$\hat{C}_H^2 = \frac{1}{m} \sum_{i=1}^m \left(\frac{x_{i,1} - x_{i,2}}{s_i} \right)^2 \quad (5)$$

Taki sposób obliczania wymaga założenia, że rozproszenie wartości porównywanych cech w obu zbiorach jest takie samo, równe rozproszeniu w jednym ze zbiorów, rozproszeniu w jakimś arbitralnie przyjętym zbiorze „wzorcowym”, lub średniej miar rozproszenia w obu zbiorach. Proponowano też [Thoma 1978], by do normali-

zacji używać międzygrupowego odchylenia standardowego. Po normalizacji wartość miary rozproszenia dla każdej z cech wynosi teoretycznie 1 ($s^2=1$).

Jak wiadomo, rzetelność oszacowania średniej arytmetycznej ocenia się za pomocą błędu będącego odchyleniem standardowym średnich z prób o danej liczebności wokół średniej w populacji generalnej

$$E_{\bar{x}} = \frac{s}{\sqrt{N}} \quad (6)$$

Rzetelność zaś oceny różnicy pomiędzy dwoma oszacowaniami średnich z prób wyraża się błędem różnicy średnich

$$E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} \quad (7)$$

Jest to pierwiastek wariancji różnic pomiędzy średnimi z prób wokół różnicy średnich populacji generalnych. Gdy odchylenia standardowe i liczebności obu prób są takie same ($s_1=s_2$; $N_1=N_2$), otrzymujemy

$$E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{2s^2}{N}} = s\sqrt{2/N} \quad (8)$$

zaś w przypadku wartości ustandaryzowanych ($s_1=s_2=1$) przy różnych liczebnościach ($N_1 \neq N_2$) prób

$$E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \quad (9)$$

a przy jednakowych liczebnościach prób ($N_1=N_2=N$)

$$E_{\bar{x}_1 - \bar{x}_2} = \sqrt{2/N} \quad (10)$$

Odległość geometryczną można traktować jako wariancję wartości średnich jednego z obiektów wokół wartości średnich drugiego z obiektów. Wariancja ta ma dwa składniki. Jeden to różnice pomiędzy średnimi populacji generalnych, z których pobrano badane próby, a drugi to wariancja błędu losowego. Odległość geometryczną obliczoną z próby można przedstawić jako sumę

$$\hat{C}_H^2 = E_{\bar{x}_1 - \bar{x}_2}^2 + C_H^2 = \frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{m} \sum_{i=1}^m \left(\frac{\mu_{i,1} - \mu_{i,2}}{\sigma_i} \right)^2 \quad (11)$$

Ponieważ przy ustalaniu odległości chodzi o uchwycenie różnic pomiędzy populacjami generalnymi, z których pochodzą badane próby, od wartości odległości oznaczonych z prób należy odjąć wielkość błędu losowego celem oznaczenia „prawdziwej” odległości

$$C_H^2 = \hat{C}_H^2 - \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \quad (12)$$

Widzimy zatem, że w przypadku odległości geometrycznej, która jest *de facto* wariancją, da się nie tylko ocenić wielkość błędu oszacowania, ale także, korzystając z właściwości zwanej addytywnością wariancji, ocenić rzeczywistą wielkość odległości geometrycznej. Spełnione być musi jedynie założenie o tym, że wariancje wszystkich cech w obu porównywanych zespołach obiektów są równe 1. Jest to już jednak odrębna kwestia jakości szacunków odchyłeń standardowych używanych do normalizacji danych.

Badanie podobieństw i różnic obiektów biologicznych za pomocą odległości geometrycznej ma na celu stwierdzenie ogólnego „pokrewieństwa” tych obiektów, a więc podobieństwa, lub różnicy, we wszystkich możliwych cechach. W praktyce nie bada się jednak absolutnie wszystkich cech, a jedynie pewien ich zestaw (m), który można traktować jako próbę pobraną spośród wszystkich cech. Zatem wartość odległości geometrycznej da się traktować jako wariancję z próby m cech. Ponieważ błąd standardowy oszacowania wariancji jest funkcją jej wielkości oraz liczebności próby [Jasicki i in. 1962] w naszym przypadku wynosi on:

$$E_{c_H^2} = \frac{C_H^2 \sqrt{2(m-1)}}{m} \quad (13)$$

Wielkość tego błędu stanowi podstawę oszacowania przedziału ufności, w którym mieści się „idealna” wartość ζ_H^2 .

Z powyższych rozważań wynika też, że porównanie pomiędzy sobą parami wartości odległości C_H^2 (po usunięciu wpływu liczebności prób) oprócz można na teście F Snedecora służącym do porównywania ze sobą wariancji. Wartość funkcji testowej przedstawia się następująco:

$$\bar{F}^0 = \frac{C_{H_1}^2}{C_{H_2}^2} \cdot \frac{m_1(m_2-1)}{m_2(m_1-1)} \quad (14)$$

Jak zawsze w tym teście wartość wariancji w liczniku musi być większa od wartości w mianowniku, a wartość testową porównuje się z wartością krytyczną dla m_1-1 i m_2-1 stopni swobody.

Usuwanie wpływu liczebności prób na wielkość \hat{C}_H^2 komplikuje się, gdy liczebności obserwacji poszczególnych cech różnią się w obrębie każdego ze zbiorów, a dzieje się tak często w praktyce badań materiałów szkieletowych o różnej kompletności zachowania poszczególnych części czaszek. Wówczas należałoby ustalać osobno wariancję błędu losowego wartości każdej z cech, można jednak zastosować obliczenie średniej, dla danego zbioru liczebności, przez analogię do liczebności średniej stosowanej w rachunkach powtarzalności (korelacji wewnątrzklasowej) [Sváb 1978]

$$\bar{N} = \frac{1}{m-1} \left(\sum_{i=1}^m n_i - \frac{\sum_{i=1}^m n_i^2}{\sum_{i=1}^m n_i} \right) \quad (15)$$

n_i – liczba obserwacji i -tej cechy. Wartości \bar{N} podstawia się odpowiednio do (11) i (12).

Pierwsze, wprowadzone przez J. Czekanowskiego [1909] określenia odległości biologicznej opierały się na obliczaniu nie odległości geometrycznej, a przeciętnej różnicy pomiędzy porównywanymi obiektami. Do dziś używa się tego sposobu. W metodzie Penrose'a, poza wartością \hat{C}_H^2 oblicza się wartość \hat{C}_Q^2 zwaną „odległością wielkości”. Jest ona po prostu kwadratem przeciętnej różnicy pomiędzy wartościami średnimi unormowanymi na odchylenia standardowe

$$\hat{C}_Q^2 = \left(\frac{1}{m} \sum \frac{\bar{x}_{i,1} - \bar{x}_{i,2}}{s_i} \right)^2 \quad (16)$$

Ponieważ \hat{C}_Q jest po prostu średnią wartością różnicy średnich z prób, a wariancja błędów tych różnic jest, jak wiadomo z (9), równa sumie odwrotności liczebności prób, wariancja błędu losowego \hat{C}_Q wyniesie

$$E_{\hat{C}_Q}^2 = \frac{1}{m} \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \quad (17)$$

W przypadku średnich, odmiennie niż w przypadku wariancji, błędy losowe nie sumują się w jednym kierunku, a częściowo się znoszą, gdyż obarczone błędami oszacowania z prób dodawane są do siebie wraz z ich znakami. Stąd odwrotnie proporcjonalna zależność wielkości błędu od liczby branych pod uwagę cech. Dla miary położenia, mimo znajomości wielkości błędu losowego nie można wyznaczyć dokładnej, nieobciążonej błędem wartości, a jedynie można podać przedział ufności. Wyznacza się go dodając do wartości wyznaczonej z próby i odejmując od niej odpowiednie do założonej szerokości przedziału wielokrotności wartości błędu losowego. Dla 95% przedziału ufności C_Q granice kształtują się następująco:

górna

$$C_Q = \hat{C}_Q + 1,96 \sqrt{\frac{1}{m} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)} \quad (18)$$

i dolna

$$C_Q = \hat{C}_Q - 1,96 \sqrt{\frac{1}{m} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)} \quad (18a)$$

W ten sposób wyznacza się granice przedziału ufności wartości C_Q ; dla częściej używanej w zastosowaniach wartości C_Q^2 obliczone wielkości granic przedziałów ufności należy podnieść do kwadratu.

Przedstawione postępowanie jest słuszne przy założeniu, że badamy odległość pomiędzy zbiorami obiektów pod względem określonego zestawu cech (m). Chcąc natomiast wyznaczyć granice przedziału ufności dla wartości ζ_Q^2 mającej wyrażać ogólne podobieństwo dwóch zbiorów we wszystkich ich cechach, a oszacowane tylko na podstawie zestawu m cech musimy zająć się tzw. „odległością ogólną” – \hat{C}_P^2 . Otrzymuje się ją przez odjęcie wartości \hat{C}_Q^2 od wartości \hat{C}_H^2

$$\hat{C}_P^2 = \hat{C}_H^2 - \hat{C}_Q^2 \quad (19)$$

Ta odległość ogólna jest niczym innym jak wariancją różnic poszczególnych cech wokół różnicy przeciętnej C_Q . Błąd średniej, jak łatwo zauważyć, wynosi w tym przy-

padku:

$$E_{\zeta_Q} = \sqrt{\frac{\hat{C}_P^2 m}{m(m-1)}} = \sqrt{\frac{\hat{C}_P^2}{m-1}} \quad (20)$$

Błędu tego, podobnie jak i poprzednio, nie da się wytrącić z wartości \hat{C}_Q a jedynie można na jego podstawie wyznaczyć przedział ufności dla dokładnej wartości ζ_Q . Dla 95% takiego przedziału granice wynoszą:

górna

$$\zeta_Q = \hat{C}_Q + 1,96E_{\zeta_Q} \quad (21)$$

i dolna

$$\zeta_Q = \hat{C}_Q - 1,96E_{\zeta_Q} \quad (21a)$$

Dla wartości ζ_Q^2 obliczone wartości granic należy oczywiście podnieść do kwadratu. Nie ma sensu podnoszenie do kwadratu wartości granicy dolnej, gdy okazuje się, iż jest ona mniejsza od zera. W takim wypadku przyjmujemy że przedział ufności, niezależnie od wyrachowanej bezwzględnej wartości liczbowej dolnej granicy, rozciąga się od zera do wartości górnej granicy podniesionej do kwadratu. Powyższe oszacowanie ma charakter przybliżony, ujmując wpływ liczebności prób tylko w sposób pośredni, poprzez uwzględnienie surowej wartości \hat{C}_H^2 , która jest zawsze większa przy mniejszej liczebności prób, a zatem odzwierciedla liczebność prób w wartości \hat{C}_P^2 . Dokładne wyznaczenie granic przedziału ufności wartości ζ_Q^2 wymagałoby uprzedniego wyznaczenia granic przedziału ufności dla ζ_H^2 i ζ_P^2 , co znacznie komplikuje rachunki a nie przynosi już znaczącego przyrostu informacji (uszczerbowienia) o zakresie, w jakim może się mieścić faktyczna wartość „odległość wielkości” ζ_Q^2 .

Przedział wyznaczany na podstawie równań (21) i (21a) jest już i tak, szczególnie po podniesieniu do kwadratu, wystarczająco szeroki, by stwierdzić, że wyliczana przez badaczy wielkość liczbowo \hat{C}_Q^2 nie ma większej rzeczywistej wartości poznawczej w przypadku wnioskowania na jej podstawie o ogólnym podobieństwie biologicznym porównywanych zbiorów.

Określenie wiarygodności statystycznej dwóch miar odległości uogólnionej, tj. \hat{C}_P^2 (por. (19)) oraz poprawionej ze względu na interkorelację cech wartości \hat{C}_R^2 [Knussmann 1967]:

$$\hat{C}_R^2 = \hat{C}_H^2 - K \cdot \hat{C}_Q^2; \quad K = \frac{mR}{1 - R + mR} \quad (22)$$

(gdzie R – współczynnik interkorelacji m cech) polega na wyznaczeniu wartości C_H^2 według (12) oraz przedziału ufności C_Q^2 i następnie podstawieniu, w przypadku obliczania górnych granic przedziałów ufności, odpowiednio dolnych, a w przypadku dolnej granicy – górnych granic przedziałów ufności wartości C_Q^2 . Można by jeszcze rozważać wiarygodność poprawki K w zależności od liczby branych pod uwagę cech i sposobu wyznaczania interkorelacji, wydaje się to jednak niekonieczne gdyż zwykle współczynnik R przyjmuje się jako obowiązującą, daną *a priori* wartość, a liczba cech jest wartością całkowicie pewną nie podlegającą oszacowaniu.

Jeżeli celem jest wnioskowanie o podobieństwie w obrębie zestawu m cech wówczas granice przedziału ufności wartości C_R^2 przedstawiają się następująco:
górna

$$C_R^2 = C_H^2 - K(\hat{C}_Q - 1,96E_{C_Q})^2 \quad (23)$$

i dolna

$$C_R^2 = C_H^2 - K(\hat{C}_Q + 1,96E_{C_Q})^2 \quad (23a)$$

W przypadku C_P^2 postępowanie jest takie samo, tylko do równań (23) i (23a) podstawia się $K=1$.

Dla znalezienia granic przedziału ufności wartości dokładnych ζ_P^2 i ζ_R^2 , służących do wnioskowania o ogólnym podobieństwie biologicznym, należy oznaczyć granice przedziału ufności dla ζ_H^2 według wzoru (13) i następnie podstawić do (23) zamiast C_H^2 wartość górnej granicy tego przedziału a do (23a) dolnej. Należy jednak zauważyć, że na skutek kombinacji prawdopodobieństw tak wyznaczone granice są właściwe dla przedziału ufności 99,75%. Aby otrzymać granice 95% przedziału ufności odległości „uogólnionych”, należy wartości błędów losowych ζ_H^2 i ζ_Q^2 mnożyć przez 0,76, co odpowiada prawdopodobieństwu popełnienia błędu dla każdej z nich równemu 0,22, a zatem kombinacji prawdopodobieństw równej 5%. Dla przybliżonej oceny można też uznać \hat{C}_P^2 lub \hat{C}_R^2 za wariancję i skorzystać z wzoru (13) podstawiając je zamiast C_H^2 .

Z powyższych rozważań wynika, że wartości odległości biologicznych szczególnie zaś „uogólnionych” odległości Penrose’a obarczone są znacznymi błędami losowymi. Są one największe w tych sytuacjach, w których uznaje się je za mierniki generalnego podobieństwa biologicznego, a nie za miary różnic pomiędzy zestawami cech. Rachunkowa komplikacja miar typu C_P^2 czy C_R^2 w stosunku do prostej odległości geometrycznej nie podnosi wcale dokładności uzyskiwanych wyników, a wprost przeciwnie, zwiększa ją zaciemniając ocenę wiarygodności statystycznej, a co za tym idzie pozostawiając większą dowolność interpretacji wyników zgodnie z uprzedniami w stosunku do badań ilościowych intuicjami czy wręcz przekonaniem badacza o relacjach pomiędzy grupami. Wydaje się, że optymalną miarą pozostaje odległość D^2 Mahalanobisa uwolniona od wpływu korelacji pomiędzy cechami i o dającej się łatwo oceniać wiarygodności statystycznej [Anderson 1958, Caliński i Kaczmarek 1976]. Jeśli zaś nie można jej obliczyć z jakichkolwiek względów technicznych, najlepiej stosować prostą odległość geometryczną typu C_H^2 lub DD^2 Czekańskiego wraz z poprawką na liczebność prób.

Jednym z paradoksów związanych z omawianymi metodami jest to, że wraz ze spadkiem liczebności jednostek obserwacyjnych w porównywanych zbiorach, nieuchronnie wzrastają przeciętne odległości pomiędzy nimi. Rośnie też wariancja odległości liczonych dla różnych zbiorów, w związku z czym otrzymuje się obraz pozornie „jaśniejszy”, o wyraźniej rysujących się różnicach międzygrupowych. Jest to jednak artefakt wynikający z wpływu liczebności prób.

I. Schwidetzky [1967] i R. Knussmann [1967] w sposób „empiryczny”, wykonując rachunki dla rozmaicie dobranych prób, starali się wykazać przydatność metody Penrose’a do badań podobieństw i różnic pomiędzy grupami ludzkimi.

I. Schwidetzky, na przykład, obserwowała, jak zmieniają się odległości pomiędzy tymi samymi zbiorami czaszek, jeśli do obliczeń weźmie się kolejno dane o jednej, dwóch, trzech itd. czaszkach z każdego zbioru, stwierdzając, że liczebności rzędu 7 i więcej obiektów z każdego zbioru są wystarczające dla „ustabilizowania” wartości odległości otrzymywanych z rachunków. R. Knussmann stwierdził wysoką korelację pomiędzy odległościami Penrose’a i Mahalanobisa w zbiorze 14 serii czaszkowych o liczebności 10 do 88 osobników ($r=0,93$), natomiast dla serii o mniejszej liczebności (5 - 9 osobników) korelacja była bardzo niska ($r=0,21$) i nieistotna statystycznie. Dla połączonego zbioru wszystkich, mniejszych i większych serii (w sumie 210 odległości) korelacja dwu omawianych miar okazała się także mała ($r=0,33$). Fakt ten nie zniechęcił jednak licznych badaczy do stosowania na szeroką skalę odległości Penrose’a a nawet tendencyjnego powoływania się jedynie na „dobry” wynik R. Knussmanna dla serii o większej liczebności jako na dowód znacznej wartości poznawczej odległości Penrose’a (por. np. Piontek [1979]).

Liczni badacze uzyskiwali za pomocą metody Penrose’a wyniki zgodne z „rzeczywistością etniczną” czy „intuicją historyczną” dotyczącą zróżnicowania grup ludzkich. Skrajnym przykładem takiego postępowania jest praca J. Piontka i M. Kaczmarek [1981], której autorzy uwierzywszy w moc metody Penrose’a i testu Rahmana, dobierając tendencyjnie materiał słowiański o różnej liczebności „udowodnili”, że odległości biologiczne nie odzwierciedlają tego, co według nich stanowi różnice etniczne. Jest to całkowicie niepoprawne tak z punktu widzenia logiki jak i metodyki badań statystycznych.

Można empirycznie wykazać, w jakim stopniu liczebność prób wpływa na wyniki osiągnięte za pomocą obliczeń odległości geometrycznej i jej pochodnych. Nim przejdziemy do dowodów *stricte* empirycznych, przyjrzyjmy się przykładom rachunkowym zastosowania wniosku statystycznego proponowanego w niniejszej pracy.

PRZYKŁADY RACHUNKOWE

Przykład I. Porównujemy z sobą dwie serie czaszek biorąc pod uwagę „standardowy” zestaw 10 cech puszeki mózgowej i twarzy. Pierwsza seria liczy 25 czaszek, druga 32 czaszki. Z rachunków otrzymujemy: $\hat{C}_H^2=0,45$, $\hat{C}_Q^2=0,12$, $\hat{C}_P^2=0,33$ i $\hat{C}_R^2=0,36$. Według wzoru (12) C_H^2 wynosi $0,45 - 0,07 = 0,39$, zaś granice 95% przedziału ufności dla ζ_H^2 z (13): górna $0,39 + 1,96 \cdot 0,17 = 0,72$ i dolna $0,39 - 1,96 \cdot 0,17 = 0,06$. Przedział ufności dla C_Q^2 określamy korzystając z wzorów (18) i (18a) na $[0,35 + 1,96\sqrt{0,1(0,03+0,04)}]^2 = 0,26$ (górna granica) i $(0,35 - 0,16)^2 = 0,04$ (dolna granica). Wyznaczenie granic przedziału ufności dla ζ_Q^2 wymaga z kolei zastosowania wzorów (20) (21) i (21a), skąd otrzymujemy $E_{C_Q}^2 = 0,19$, a co za tym idzie ζ_Q^2 mieści się w granicach od zera, bo $0,35 - 0,19 \cdot 1,96 < 0$, do $(0,35 + 0,37)^2 = 0,52$. Ponieważ przedział ufności obejmuje zero, wartość ta nie rozstrzyga o biologicznej różnicy pomiędzy badanymi grupami.

Określenie przedziału ufności dla C_P^2 przebiega według (23) i (23a). Korzystamy tu z uprzednio obliczonych granic przedziału ufności C_Q^2 , które wynoszą 0,04 i 0,26

oraz wartości C_H^2 wynoszącej w niniejszym przykładzie 0,39. Górna granica przedziału ufności C_P^2 wynosi zatem $0,39 - 0,04 = 0,35$, a dolna $0,39 - 0,26 = 0,13$. Odpowiednie rachunki dla C_R^2 wymagają pomnożenia 0,04 i 0,26 przez poprawkę K uwzględniającą liczbę cech (tu 10) i interkorelację pomiędzy nimi. Wynosi ona dla standardowego zestawu cech czaszki 0,233 [Rösing 1975]. Zatem dla 10 cech $K = 0,75$. Uwzględniając podane poprzednio wartości liczbowe oraz tę poprawkę, z wzorów (23) i (23a) otrzymujemy granice 95% przedziału ufności dla C_R^2 równe: $0,39 - 0,75 \cdot 0,04 = 0,36$ (górną) i $0,39 - 0,75 \cdot 0,26 = 0,20$ (dolną).

Szacunki granic przedziałów ufności dla wartości „dokładnych” pozwalających uogólniać wnioski na całość cech biologicznych wymagają uwzględnienia kombinacji prawdopodobieństw, a zatem pomnożenia odpowiednich wartości błędów przez 0,76. Dla odległości ζ_Q^2 tak uzyskany 78% przedział ufności rozciąga się od $(0,35 - 0,76 \cdot 0,19)^2 = 0,04$ do $(0,35 + 0,76 \cdot 0,19)^2 = 0,24$ zaś górna granica przedziału ufności ζ_H^2 wynosi $0,39 + 0,76 \cdot 0,17 = 0,52$ a dolna $0,39 - 0,76 \cdot 0,17 = 0,26$ przy prawdopodobieństwie popełnienia błędu równym 22%. Zatem cały 95% przedział ufności dla ζ_P^2 ma granice: górną $0,52 - 0,04 = 0,48$ i dolną $0,26 - 0,24 = 0,02$. Taki sam przedział dla ζ_R^2 po wykonaniu odpowiedniego mnożenia przez $K = 0,75$ rozciąga się od $0,26 - 0,75 \cdot 0,24 = 0,08$ do $0,52 - 0,75 \cdot 0,04 = 0,49$. Widać, że przedziały dla wartości odległości ogólnych są bardzo szerokie mimo uwzględnienia 10 cech w seriach czaszkowych o nie najmniejszej, jak na tego typu materiały, liczebności. Zatem wnioskowanie o ich wzajemnym podobieństwie biologicznym nie jest zbyt dokładne. Niemożliwe będzie też uprawnione matematycznie stwierdzenie, czy odległość pomiędzy tymi seriami jest większa czy mniejsza od odległości pomiędzy innymi dwoma seriami, dla których wyrachowano np. $\hat{C}_P^2 = 0,21$ i $\hat{C}_R^2 = 0,23$.

Przykład II. Porównajmy z sobą nieliczne, co się często zdarza, zbiory czaszek neolitycznych ($N_1 = 7$, $N_2 = 9$), charakteryzowane przez 6 cech każdy. Wyrachowane z danych odległości wynoszą: $\hat{C}_H^2 = 1,75$, $\hat{C}_Q^2 = 0,69$, $\hat{C}_P^2 = 1,06$, $\hat{C}_R^2 = 1,23$. Obliczamy $C_H^2 = 1,75 - 0,25 = 1,50$, a 95% przedział ufności $\zeta_H^2 = 1,50 \pm 1,96 \cdot 0,79$ czyli od $-0,05$ do $3,05$, a zatem odległość jest formalnie nieistotnie różna od zera. Przedział ufności C_Q^2 wynosi $(0,83 \pm 1,96 \cdot 0,21)^2$ czyli od 0,18 do 1,52. Stąd 95% przedział ufności C_P^2 ma dolną granicę równą $1,50 - 1,54 = -0,04$ a górną $1,50 - 0,18 = 1,32$ obejmując tym samym zero a zatem wskazując na brak istotności statystycznej wartości \hat{C}_P^2 . Odpowiedni przedział ufności dla C_R^2 rozciąga się od 0,35 do 1,37 zając tym samym z odpowiednim przedziałem dla poprzedniego przykładu. Można więc wnioskować, że wartość \hat{C}_R^2 z poprzedniego przykładu równa 0,36 nie różni się istotnie od wyliczonej w niniejszym przykładzie równej 1,23, mimo iż „na oko” różnica pomiędzy nimi jest duża.

Przedziały ufności dla wartości ζ_Q^2 , ζ_P^2 i ζ_R^2 są jeszcze znacznie szersze i odpowiednie ich granice wynoszą kolejno: od 0,17 do 1,56, od $-0,49$ do 1,87 oraz od $-0,14$ do 1,93. Zatem odległości „uogólnione” mają przedziały obejmujące zero, a wszystkie wyliczone przedziały ufności zająć się z przedziałami ufności dla wartości z poprzedniego przykładu. Niewiele więc można wnioskować o „wzajemnym ogólnym

podobieństwie biologicznym” porównywanych tu serii ani też jego stosunku do podobieństwa serii porównywanych w poprzednim przykładzie. Jediną nadającą się tu do rzetelnego wnioskowania wartością jest C_H^2 .

Analizując wyniki otrzymane w obydwu przykładach warto zwrócić jeszcze uwagę na to, iż „surowe” wartości odległości \hat{C}_P^2 i \hat{C}_R^2 leżą nie w środku właściwych im przedziałów ufności, ale w pobliżu ich górnych granic, co wynika ze stosowania odpowiednio poprawionej na liczebność prób wartości C_H^2 do ich obliczania. Widać więc wyraźnie, że liczebność prób porównywanych wywiera na te odległości wpływ jednostronny. Zmniejszanie liczebności prób prowadzi do powiększania odległości i *vice versa*, oczywiście w sensie statystycznym, większego prawdopodobieństwa.

UZASADNIENIE EMPIRYCZNE

Na podstawie danych i wyników przedstawionych przez autorów pracujących metodą Penrose’a można okazać wpływ liczebności prób na wartości miar odległości. Ponieważ najszerzej używano odległości \hat{C}_R^2 jako „najogólniejszej” miary, ona właśnie najczęściej podawana jest w opracowaniach. Autorzy nie przytaczają innych „pośrednich” według nich miar. Sytuacja taka utrudnia postępowanie mające na celu okazanie wpływu liczebności prób na wielkość odległości gdyż, jak widać w poprzednich częściach niniejszej pracy, najwyraźniejszy wpływ wywiera ona na odległość geometryczną typu \hat{C}_H^2 , natomiast w przypadku miary „ogólnej” liczebność prób po części powiększa tylko losowe, obustronne (w górę i w dół) wahania wartości \hat{C}_Q^2 , a poprzez nią także \hat{C}_R^2 .

Postępowanie empiryczne polegało na wybraniu z literatury danych dotyczących liczebności serii czaszek oraz odległości \hat{C}_R^2 pomiędzy nimi, a następnie zbadaniu za pomocą analizy wariancji związku, jaki zachodzi pomiędzy sumami odwrotności liczebności porównywanych serii (według 12) a wielkościami odległości Penrose’a. Postępowanie to zastosowano do trzech zespołów danych. Na pierwszy z nich składają się dane o 22 seriach z terenu Europy datowanych na VII - VIII wiek n.e. [Roth-Lutra 1970, 1971]. Drugi zespół stanowią dane I. Schwidetzky [1967] o 22 seriach neolitycznych (4 - 3 tys. lat p.n.e.) z terenu Europy. W tym przypadku wzięte zostały pod uwagę tylko dane dotyczące serii charakteryzowanych kompletem cech; wyłączono z analizy serie charakteryzowane tylko przez 7 cech.

Na trzeci zespół składa się 14 żeńskich serii czaszkowych z terenu Polski datowanych na wczesne średniowiecze [Piontek 1979]. W zespole tym jedna z odległości (Wiślica-Tomice) została w druku błędnie podana jako 0,01, po przeliczeniu z danych opublikowanych w cytowanej pracy J. Piontka okazało się, iż odległość ta wynosi 0,24. Taką też wartość włączono do obliczeń.

Tabela 1 zawiera dane liczbowe charakteryzujące związek liczebności z odległościami \hat{C}_R^2 . We wszystkich trzech przypadkach zależność jest istotna, a jej kierunek zgodny z przewidywaniami. Jedinym wyjątkiem stanowi klasa największej liczebności porównywanych serii (do 0,1) dla neolitycznych serii I. Schwidetzky’ej [1967].

Tabela 1. Zależność pomiędzy liczebnościami porównywanych serii czaszkowych (N_1 i N_2) i odległościami \hat{C}_R^2 Penrose'a

$\frac{1}{N_1} + \frac{1}{N_2}$		do 0,1	do 0,2	do 0,3	do 0,4	ponad 0,4	F°, p, η
Autor i rodzaj danych							
Roth-Lutra [1970, 1971] męskie, VII - VIII w.n.e. Europa	N	51	65	38	33	45	$F^\circ = 15,46$
	\bar{C}_R^2	0,29	0,34	0,42	0,59	0,56	$p = 3 \cdot 10^{-11}$
	s^2	0,03	0,04	0,04	0,10	0,06	$\eta = 0,46$
Schwidetzky [1967] męskie, 4 - 3 tys. lat p.n.e. Europa	N	28	78	53	31	29	$F^\circ = 2,63$
	\bar{C}_R^2	0,88	0,79	0,79	1,06	1,16	$p = 0,04$
	s^2	0,43	0,38	0,52	0,32	0,30	$\eta = 0,22$
Piontek [1979] kobiece, wczesne średniowiecze, Polska	N	13	66	10	—	—	$F^\circ = 3,51$
	\bar{C}_R^2	0,26	0,40	0,41	—	—	$p = 0,03$
	s^2	0,01	0,04	0,02	—	—	$\eta = 0,28$

Jednak liczba tych odległości jest mała i wahnięcie to może mieć charakter przypadkowy.

Fakt stwierdzenia istotnych zależności jest szczególnie znaczący w świetle poprzednio już wyrażonych zastrzeżeń co do wartości \hat{C}_R^2 jako miary niejednokierunkowo zależnej od liczebności, oraz tego, że liczebności obserwacji dla poszczególnych cech w każdej z serii były różne, co zmusiło do używania w analizie wariancji przybliżonych, uśrednionych liczebności obserwacji dla poszczególnych serii. Uzasadnione zatem wydaje się wyrażenie przekonania, że gdyby zastosować miarę \hat{C}_H^2 oraz dokładne liczebności obserwacji dla każdej z serii, zależność zarysowałaby się jeszcze wyraźniej.

Tabela 2 zawiera podobne dane dotyczące odległości D^2 Mahalanobisa pochodzące z własnych niepublikowanych jeszcze obliczeń dla 22 serii męskich z terenu Europy Środkowej datowanych na wczesne średniowiecze. Zależność jest nieistotna statystycznie, przyrost wartości średnich wraz ze spadkiem liczebności badanych serii rysuje się słabo.

Tabela 2. Zależność pomiędzy liczebnościami porównywanych serii czaszkowych (N_1 i N_2) i odległościami D^2 Mahalanobisa

$\frac{1}{N_1} + \frac{1}{N_2}$		do 0,1	do 0,2	ponad 0,2	F°, p, η
Rodzaj danych					
dane własne, męskie, wczesne średn. Europa Środkowa	N	87	120	27	$F^\circ = 1,40$
	\bar{D}^2	4,37	4,63	5,63	$p = 0,25$
	s^2	15,21	9,12	10,63	$\eta = 0,11$

Innym jeszcze sposobem postępowania zmierzającym do okazania zależności pomiędzy liczebnością badanych serii a wielkościami odległości Penrose'a było obliczenie przeciętnych liczebności zespołów serii czaszek i zestawienie ich z przeciętnymi odległościami \hat{C}_R^2 obliczonymi dla tych samych zespołów serii. Skorzysta-

Tabela 3. Charakterystyki rozkładów liczebności serii z Europy używanych do obliczania średnich odległości \bar{C}_R^2 Penrose'a przez I. Schwidetzky [1967, 1972] — do końca okresu p.n.e. oraz J. Piontka [1979] — dla wczesnego średniowiecza. \bar{g} — średnia geometryczna, \bar{h} — średnia harmoniczna

Okres	N serii	\bar{x}	\bar{g}	\bar{h}	s	\bar{C}_R^2
do 4 tys. lat p.n.e.	9	8,1	6,8	5,6	4,5	0,93
4 - 3 tys. lat p.n.e.	13	29,7	19,1	13,0	29,1	0,63
3 - 2,5 tys. lat p.n.e.	15	23,2	19,1	16,6	16,9	0,59
2,5 - 2 tys. lat p.n.e.	34	22,9	17,0	13,4	21,7	0,73
2 - 1 tys. lat p.n.e.	31	24,7	20,3	15,9	15,2	0,58
1000 - 0 lat p.n.e.	71	42,2	23,5	17,8	102,3	0,50*
wczesne średniowiecze	40	37,9	24,3	17,8	46,7	0,37

* — wartość uśredniona z podanych przez autorkę 0,54 i 0,45.

liśmy w tym przypadku z danych I. Schwidetzky'ej [1967, 1972] oraz J. Piontka [1979]. Obliczyliśmy średnie arytmetyczne, geometryczne i harmoniczne liczebności serii, dla których autorzy ci określali przeciętne odległości Penrose'a. Okazuje się (tab. 3 i 4), że związki przeciętnej liczebności serii z przeciętnymi odległościami Penrose'a są bardzo wysokie, współczynniki korelacji nieistotnie odbiegają od idealnej wartości 1,0. Według podanych stałych równań regresji można z dużą pewnością przewidywać przeciętną odległość serii na podstawie znanej ich liczebności.

Tabela 4. Obliczona z danych tabeli 3 korelacja pomiędzy przeciętną liczebnością porównywanych serii a przeciętną odległością \bar{C}_R^2 pomiędzy nimi oraz parametry równań regresji postaci $\bar{C}_R^2 = a + b\bar{N}$, \bar{N} oznacza przeciętną liczebność

Średnia	r	a	b
arytmetyczna \bar{x}	-0,89	1,00	-0,014
geometryczna \bar{g}	-0,96	1,16	-0,029
harmoniczna \bar{h}	-0,93	1,17	-0,038

Ten ostatni wynik stawia pod znakiem zapytania interpretację jaką wymienieni autorzy stosowali wobec obserwowanych zmian przeciętnej odległości Penrose'a w czasie. Uważali oni te zmiany za odzwierciedlenie rzeczywistego spadku zróżnicowania międzypopulacyjnego z upływem czasu. Tymczasem przed wyciągnięciem takiego wniosku należałoby usunąć z wartości przeciętnej odległości tę jej część, która wynika z prostego faktu wzrostu przeciętnej liczebności serii czaszek wraz z upływem czasu historycznego.

Podsumowując należy stwierdzić małą przydatność złożonych „uogólnionych” odległości do badań zróżnicowania międzypopulacyjnego człowieka, a w szczególności do orzekania o stopniu ich wzajemnego „podobieństwa biologicznego”. Do tych celów lepiej jest stosować proste odległości geometryczne lub odległości statystyczne Mahalanobisa uwzględniając fakt wpływu liczebności porównywanych prób na ich wartości i stosując odpowiednie metody wnioskowania statystycznego.

PIŚMIENNICTWO

- Anderson T. W., 1958, *Introduction to Multivariate Statistical Analysis*, Nowy Jork.
- Caliński T., Z. Kaczmarek, 1976, *A note on the calculation and use of the generalized distance between multivariate samples*, Zesz. Nauk. UAM, Geografia, 8, 7.
- Constandse-Westermann T. S., 1972, *Coefficients of Biological Distance*, Anthropol. Publ. Oosterhout.
- Czekanowski J., 1909, *Zur Differentialdiagnose der Neandertalgruppe*, Korrespondenzblatt der Deutsch. Ges. für Anthrop., 40, 44.
- Jasicki B., S. Panek, P. Sikora, E. Stolyhwo, 1962, *Zarys antropologii*, Warszawa.
- Knussmann R., 1962, *Moderne statistische Verfahren in der Rassenkunde*, [w:] *Die Neue Rassenkunde* (red. I. Schwidetzky), Stuttgart, 233.
- Knussmann R., 1967, *Penrose-Abstand und Diskriminanzanalyse*, Homo, 18, 134.
- Mahalanobis P. C., 1930, *On tests and measures of group divergence*, J. Asiat. Soc. Beng., 26, 541.
- Penrose L. S., 1954, *Distance size and shape*, Ann. Eug., 18, 337.
- Piontek J., 1979, *Procesy mikroewolucyjne w europejskich populacjach ludzkich*, Poznań.
- Piontek J., M. Kaczmarek, 1981, *Badania etnograficzne w antropologii: Próba nowego spojrzenia*, Przegł. Antrop., 47, 129.
- Rahman N. A., 1962, *On the sampling distribution of the studentized Penrose measure of distance*, Ann. Hum. Genet., 26, 97.
- Rösing F. W., 1975, *Die Fränkische Bevölkerung von Mannheim-Vogelstang (6.-7. Jh.) und die Merowingerzeitlichen Germanengruppen Europas*, Hamburg.
- Rösing F. W., I. Schwidetzky, 1977, *Vergleichend-statistische Untersuchungen zur Anthropologie des frühen Mittelalters (500 - 1000 n.d.Z.)*, Homo, 28, 65.
- Roth-Lutra K. H., 1970, *Vergleichend-statistische Untersuchungen zur Anthropologie des Früh- und Hochmittelalters in Europa I.*, Homo, 21, 104.
- Roth-Lutra K. H., 1971, *Vergleichend-statistische Untersuchungen zur Anthropologie des Früh- und Hochmittelalters in Europa II.*, Homo, 22, 84.
- Schwidetzky I., 1967a, *Erfahrungen mit dem Penrose-Abstand*, Homo, 18, 140.
- Schwidetzky I., 1967b, *Ergebnisse der Penrose-Analyse: Das Gesamtmaterial*, Homo, 18, 174.
- Schwidetzky I., 1972, *Vergleichend-statistische Untersuchungen zur Anthropologie der Eisenzeit*, Homo, 23, 245.
- Schwidetzky I., F. W. Rösing, 1975, *Vergleichend-statistische Untersuchungen zur Anthropologie der Römerzeit (0 - 500 u.Z.)*, Homo, 26, 193.
- Sváb J., 1978, *Genotyka populacji*, Warszawa.
- Thoma A., 1978, *Distance de forme entre groupes*, Bull. et Mém. de la Soc. d'Anthrop. de Paris, 5, XIII, 15.
- Wanke A., 1955, *Indywidualne określanie taksonomiczne*, Przegł. Antrop., 21, 968.

ul. Polna 31/5a
60-535 Poznań

RELIABILITY OF MULTIVARIATE MEASURES OF GEOMETRIC DISTANCE WITH SPECIAL REFERENCE TO THE PENROSE DISTANCE

by MACIEJ HENNEBERG

The author briefly describes the idea of multivariate geometric distance (eq. (1), (1a)) and states that it may be, for comparisons between groups, treated formally as a variance of one set of averages around the other. Since the averages of variables characterizing groups are computed

for a limited number of observations, they are in fact only estimates with a certain degree of error. According to well known statistical theorems variance of sample means around the mean of the general population is a function of standard deviation and sample size (eq. (6 - 10)). While we deal with normalized values (S.D. = 1) this variance which square root is called a standard error becomes simply proportional to inverse sample size. Therefore a multivariate geometrical distance computed from sample averages (of C_H^2 type) can be perceived as consisting of two parts: variance between averages of general populations from which the samples are respectively taken and variance of random error (eq. (11)). This allows corrections of geometrical distance values for sample sizes to be made (eq. (12)). Furthermore, if one is looking for general similarity or dissimilarity between two populations it must be taken into account that a measure of distance is computed on a limited number of variables (m) only. Appropriate correction for number of variables may be obtained in a form of confidence limits computation (eq. (13) and following). „Distances” estimated directly from samples are labelled \hat{C}^2 , corrected for sample size only C^2 and corrected for both sample size and number of variables ζ^2 with appropriate indices: H - „shape distance”, Q - „size distance”, P - „generalized distance” and R - „generalized distance” corrected for intercorrelation between variables.

Numerical examples of statistical evaluation are given. When computing confidence interval limits for „generalized” measures it should be taken into account that combination of probabilities of error for their constituent parts occurs and thus appropriate error estimates should not be multiplied eg. by 1,96 for 95 % confidence interval but by 0,76 (see Przykład I and Przykład II).

From general considerations and numerical results it may be concluded that for the range of sample sizes and variable numbers most commonly encountered in physical anthropology studies confidence intervals, of „generalized” measures of distance are quite wide and hence their reliability low, putting their practical usefulness to serious doubt. More reliable and easier testable turns out to be a simple geometrical distance.

The third part of this paper deals with empirical proofs of sample size influence on distance values. First (table 1) are presented relationships between sample sizes and individual \hat{C}_R^2 distances computed by other authors. It may be seen that relationship is clear-cut and statistically significant. The same sort of relationship for D^2 of Mahalanobis (table 2) turned out to be insignificant. The other kind of proof is provided by correlation between average sample size (arithmetic, geometric and harmonic means) and average distance calculated for the same set of sample groupings (table 3 and 4). This last result puts in doubt claims of some authors that a decrease of biological distance between groups occurred through time. Before arriving at such a conclusion they should correct their results for effects of sample sizes.

The overall conclusion of the paper is that tests of statistical significance should be applied to various measures of distance and that the best of these measures is D^2 of Mahalanobis. If this measure cannot be used for whatever reasons it is most advisable to use simple measure of geometrical distance corrected for sample sizes.