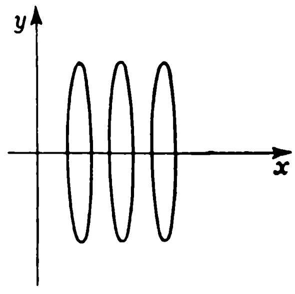
ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
IX, 4 (1968)

S. ZUBRZYCKI (Wrocław)

O WYBORZE PŁASZCZYZNY NAJLEPSZEJ DYSKRYMINACJI

W badaniach lekarskich dotyczących diagnostyki coraz częściej stosuje się metody dyskryminacji matematycznej (porównaj [2]-[6]). W ten sposób poprzez formalny aparat matematyczny umożliwiony jest wgląd w obserwacje wyrażające się wieloma liczbami i dlatego nie nadające się do poglądowego przedstawienia na płaskim rysunku. Jednak ze względu na walory poglądowego przedstawiania obserwacji ponawiane są próby uzyskania takiego przedstawienia nawet kosztem straty pewnych informacji zawartych w obserwacjach. Zazwyczaj idzie tu o rzutowanie obserwacji na odpowiednio wybraną płaszczyznę. Kluczowym sposobem w tych próbach jest wybieranie prostej czy płaszczyzny rzutowania. Może się oczywiście zdarzyć, że rzut na mniej lub bardziej arbitralnie wybraną płaszczyznę czy prostą ujawni konfigurację pozwalającą na bardzo skuteczną dyskryminację. I tak autorzy odczytu [6] podają, że przy dyskryminacji krzywych cukrowych bardzo dobrą prostą dyskryminacji okazała się prosta związana z największą wartością własną macierzy kowariancji wyrachowanej z mieszanej próbki obejmującej wszystkie rozważane typy cukrzycy. Jest zupełnie jasne, że gdyby w rozważanych grupach o elementach charakteryzowanych dwoma cechami rozkłady tych cech były odpowiednio wydłużone w niepożądanym kierunku, np. tak, jak to sugerują elipsy na rys. 1, wspomniany wyżej sposób wybierania prostej dyskryminacji kazałby dyskryminować według cechy y, zamiast, co tu byłoby właściwe, według cechy x.



Rys. 1

Chciałoby się przeto umieć wskazywać płaszczyznę rzutowania tak, żeby okazywała się ona płaszczyzną dobrej dyskryminacji, jeśli taka istnieje, nie tylko dzięki przyjaznemu zbiegowi okoliczności. Jak się wydaje, można zaproponować dość rozsądne sposoby wykorzystujące informacje o zmienności wewnątrz grup i o zmienności między grupami. Pewne kryterium wyboru płaszczyzny dyskryminacji wykorzystujące takie informacje podał J. Perkal [9] i ono jest bezpośrednim powodem do napisania tych uwag. Mianowicie kryterium wyboru zaproponowane przez Perkala ma pewne niepożądane własności. Można je wyeliminować przez odpowiednią modyfikację kryterium. Taką modyfikację omawiamy w § 6. Różni się ona tylko nieistotnie od kryterium zaproponowanego przez Rao (zobacz [10], § 9c), a także prowadzi do rozwiązania, które jest najlepsze w sensie jeszcze jednego kryterium (porównaj Wilks [12], § 18.7). Porównujemy te kryteria w § 7.

§ 1. Założenia. Rozważać będziemy K populacji, których elementy scharakteryzowane są I cechami

$$\boldsymbol{x} = [x_1, \ldots, x_I]'$$

(symboliki wektorowej będziemy używać w miarę potrzeby). O rozkładach tych cech będziemy zakładać, że mają te same nieosobliwe macierze kowariancji

(1.1) $$\boldsymbol{\Lambda} = [\lambda_{ij}]$$

macierz kowariancji wewnątrz populacji) i środki ciężkości

(1.2) $$\mu_k = [\mu_{k1}, \mu_{k2}, \ldots, \mu_{kI}]'$$

nie wszystkie równe. Oznaczmy przez

(1.3) $$\boldsymbol{M} = [m_{ij}]$$

macierz kowariancji tych środków ciężkości (macierz kowariancji między populacjami), tzn. macierz o elementach

(1.4) $$m_{ij} = \frac{1}{K}\sum_{k=1}^{K}(\mu_{ki}-\mu_{.i})(\mu_{kj}-\mu_{.j}),$$

gdzie

(1.5) $$\mu_{.i} = \frac{1}{K}\sum_{k=1}^{K}\mu_{ki}.$$

§ 2. Wybór kierunku najlepszej dyskryminacji według Fishera. Macierz kowariancji wewnątrz populacji $\boldsymbol{\Lambda}$ i macierz kowariancji między populacjami $\boldsymbol{M}$ są decydującymi elementami naszych dalszych rozważań. R. A. Fisher [7], [8] rozważał pierwotnie zagadnienie dyskryminacji dla $K = 2$ i pytał o taką liniową funkcję cech

(2.1) $$\boldsymbol{a}'\boldsymbol{x} = a_1x_1+\ldots+a_Ix_I,$$

dla której stosunek kwadratu różnicy wartości tej funkcji dla środków ciężkości rozważanych populacji do wariancji tej funkcji wewnątrz populacji jest maksymalny. Warunek ten możemy napisać następująco:

$$\frac{\{\boldsymbol{a}'(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\}^2}{\boldsymbol{a}'\boldsymbol{\Lambda}\boldsymbol{a}} = \max. \tag{2.2}$$

Okazuje się, że warunek (2.2) jest spełniony dla

$$\boldsymbol{a} = \boldsymbol{\Lambda}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2), \tag{2.3}$$

co prowadzi do znanej funkcji dyskryminacyjnej Fishera dla przypadku znanych średnich i kowariancji, w której o przynależności elementu o cechach $\boldsymbol{x}$ do pierwszej lub drugiej populacji orzeka się na podstawie wartości funkcji

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Lambda}^{-1}\boldsymbol{x}. \tag{2.4}$$

§ 3. Uogólnienie na dowolną liczbę populacji. Zauważmy, że dla dwu liczb x_1 i x_2 przy $\bar{x} = (x_1+x_2)/2$ mamy

$$(x_1-x_2)^2 = 4s^2,$$

gdzie

$$s^2 = \frac{1}{2}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2\right]$$

jest wariancją próbki złożonej z x_1 i x_2. To naprowadza na myśl, żeby w przypadku więcej niż dwu populacji licznik po lewej stronie wyrażenia (2.2) zastąpić przez wariancję próbki złożonej z wartości funkcji (2.1) dla środków ciężkości populacji, czyli z liczb

$$\boldsymbol{a}'\boldsymbol{\mu}_1, \ldots, \boldsymbol{a}'\boldsymbol{\mu}_K.$$

Łatwo sprawdzić, że wariancja ta dana jest przez

$$\boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}. \tag{3.1}$$

Tak więc przez analogię z (2.2) w przypadku więcej niż dwu populacji za najlepiej dyskryminujący kierunek możemy uznać taki wektor $\boldsymbol{a}$, dla którego

$$\frac{\boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Lambda}\boldsymbol{a}} = \max. \tag{3.2}$$

§ 4. Dyskryminacja płaska według Perkala. Perkal [9] zaproponował, żeby w przypadku więcej niż dwu populacji i więcej niż dwu cech szukać w wielowymiarowej przestrzeni cech takiej płaszczyzny, żeby rzuty populacji na tę płaszczyznę były najlepiej rozdzielone. Gdy operuje się próbkami, można wówczas sporządzić rysunek przedstawiający rzut wszystkich obserwacji na taką płaszczyznę, co daje możność tak cennego poglądowego zapoznania się z zawartością posiadanego zbioru obserwacji.

Nie jest to jeszcze formalnym postawieniem zadania. Powstaje pytanie, jaki warunek ma wyróżniać tę najlepiej rozdzielającą płaszczyznę, oraz gdy warunek zostanie już ustalony, jak tę płaszczyznę rachunkowo wyznaczyć. Perkal przedstawił dwie propozycje w tej sprawie.

Jedna dotyczy dwuzespołowej płaszczyzny rozdzielającej. Idzie tu o sytuację, kiedy zespół cech jest w pewien z góry dany sposób podzielony na dwa wyraźne podzespoły. Proponuje się wówczas dla każdego podzespołu osobno wyznaczyć według Fishera najlepiej dyskryminujący kierunek i uznać, że tak otrzymane dwie liniowe funkcje cech wyznaczają pożądaną płaszczyznę najlepszej dyskryminacji. Jest to bardzo cenna propozycja. W porównaniu z generalnym stawianiem zadania, o jakim tu szczegółowo dyskutujemy, prowadzi ona do znacznej redukcji pracy rachunkowej. Gdy zaś badacz trafnie wydzieli grupy cech niosące różne informacje o badanych indywiduach, są szanse uzyskania dobrej dyskryminacji kosztem małego trudu.

Naszą uwagę chcemy jednak skierować na drugą propozycję. Mianowicie Perkal [9] proponuje wyznaczać taką parę wektorów $\boldsymbol{a}$ i $\boldsymbol{b}$, dla której

$$P = \frac{\boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}+\boldsymbol{b}'\boldsymbol{M}\boldsymbol{b}}{\boldsymbol{a}'\boldsymbol{\Lambda}\boldsymbol{a}+\boldsymbol{b}'\boldsymbol{\Lambda}\boldsymbol{b}} = \max, \tag{4.1}$$

przy czym żąda, żeby wektory $\boldsymbol{a}$ i $\boldsymbol{b}$ spełniały jeden lub oba z dwu następujących warunków ubocznych:

$$\boldsymbol{a}'\boldsymbol{a} = 1 \quad \text{i} \quad \boldsymbol{b}'\boldsymbol{b} = 1, \tag{4.2}$$

$$\boldsymbol{a}'\boldsymbol{b} = 0. \tag{4.3}$$

Warunek (4.2) żąda, żeby wektory $\boldsymbol{a}$ i $\boldsymbol{b}$ miały długość 1, warunek (4.3) zaś, żeby były one wzajemnie prostopadłe. Maksymalną wartość lewej strony (4.1) proponuje Perkal nazywać siłą dyskryminacyjną nowych, wyróżnionych przez $\boldsymbol{a}$ i $\boldsymbol{b}$ cech, tj. cech określonych jako funkcje liniowe

$$\boldsymbol{a}'\boldsymbol{x} \quad \text{i} \quad \boldsymbol{b}'\boldsymbol{x}.$$

§ 5. Uwagi o warunku (4.1) jako o warunku wyznaczającym płaszczyznę najlepszej dyskryminacji. Chcemy tu na prostym przykładzie numerycznym zademonstrować pewne konsekwencje warunku (4.1). W przykładzie tym dla przejrzystości i prostoty przyjmiemy, że macierze $\boldsymbol{\Lambda}$ i $\boldsymbol{M}$ mają te same kierunki główne i są zadane w postaci przekątniowej. Niech więc

$$\boldsymbol{\Lambda} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{M} = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 36 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{5.1}$$

Kryterium (4.1) każe zatem wyznaczyć takie wektory

$$\boldsymbol{a} = [a_1, a_2, a_3]' \quad \text{i} \quad \boldsymbol{b} = [b_1, b_2, b_3]',$$

żeby nadać największą wartość stosunkowi

$$(5.2) \qquad P = \frac{25(a_1^2+b_1^2)+36(a_2^2+b_2^2)+2(a_3^2+b_3^2)}{5(a_1^2+b_1^2)+9(a_2^2+b_2^2)+(a_3^2+b_3^2)}.$$

Widać, że największą możliwą wartością tego stosunku jest 5 i uzyskujemy ją dla $a_1^2+b_1^2 > 0$, $a_2^2 = b_2^2 = a_3^2 = b_3^2 = 0$. Jeśli zatem odrzucimy warunki uboczne (4.2) i (4.3), to z dokładnością do współczynników proporcjonalności i przy zadaniu $\boldsymbol{a} \neq 0$ rozwiązaniem są wektory

$$(5.3) \qquad \boldsymbol{a} = [1, 0, 0]' = \boldsymbol{b}$$

lub też wektory

$$(5.4) \qquad \boldsymbol{a} = [1, 0, 0]' \quad \text{i} \quad \boldsymbol{b} = [0, 0, 0]'.$$

To drugie rozwiązanie jest wymuszone przez nałożenie tylko warunku ortogonalności (4.3), gdy żądamy $\boldsymbol{a} \neq 0$. Zauważmy, że wymuszając ponadto niezerowanie się wektora $\boldsymbol{b}$ i żądając ortogonalności, możemy tylko zmniejszyć stosunek P. Jest to na pewno niepożądana własność stosunku P jako miary siły dyskryminacyjnej, bo ta dla dwu cech powinna być chyba większa niż dla jednej z nich.

Jeśli zażądamy jednoczesnego spełnienia warunków (4.2) i (4.3), to rozwiązaniem okazuje się para wektorów

$$(5.5) \qquad \boldsymbol{a} = [1, 0, 0]' \quad \text{i} \quad \boldsymbol{b} = [0, 0, 1]',$$

która daje $P = (25+2)/(5+1) = 27/6 = 4{,}5$. Tymczasem porównując stosunki wariancji między populacjami do wariancji wewnątrz populacji dla rozpatrywanych tu trzech cech, które są równe

$$\frac{25}{5} = 5, \quad \frac{36}{9} = 4 \quad \text{i} \quad \frac{2}{1} = 2,$$

widzimy, że bardziej pożądany byłby wybór płaszczyzny wyznaczonej przez wektory $[1, 0, 0]'$ i $[0, 1, 0]'$. Przy takim wyborze mielibyśmy jednak $P = (25+36)/(5+9) = 61/14 = 4{,}357 < 4{,}5$.

§ 6. Modyfikacja. Postawmy przeto zadanie nieco inaczej. Każmy je mianowicie rozważać w postaci standaryzowanej w odpowiedni sposób. Zwróćmy najpierw uwagę na to, że wariancja wewnątrz populacji nie jest we wszystkich kierunkach jednakowa. Dlatego najpierw przy użyciu nieosobliwej macierzy A zastąpmy stare cechy przez nowe zgodnie z wzorem

$$\boldsymbol{y} = A\boldsymbol{x}$$

tak, żeby macierz kowariancji nowych cech, równa $\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}'$, okazała się macierzą jednostkową

$$\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}' = \boldsymbol{I}. \tag{6.1}$$

Macierz kowariancji środków ciężkości nowych cech $\boldsymbol{y}$ będzie teraz równa $\boldsymbol{AMA}'$.

Z kolei obróćmy nowe zmienne $\boldsymbol{y}$ za pomocą transformacji ortogonalnej

$$\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{x} \tag{6.2}$$

tak, żeby także macierz kowariancji między obróconymi środkami ciężkości przybrała postać przekątniową. Po takim obrocie macierz kowariancji wewnątrz populacji pozostanie jednostkowa, bo $\boldsymbol{CIC}' = \boldsymbol{CC}' = \boldsymbol{I}$ dla macierzy ortogonalnej $\boldsymbol{C}$. Macierz kowariancji między obróconymi środkami ciężkości, równa

$$\boldsymbol{D} = \boldsymbol{CAMA}'\boldsymbol{C}', \tag{6.3}$$

przybierze więc postać

$$\boldsymbol{D} = \begin{bmatrix} d_{11}, & 0, & \ldots, & 0 \\ 0, & d_{22}, & \ldots, & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \ldots, & d_{II} \end{bmatrix}. \tag{6.4}$$

Przez odpowiednie ponumerowanie zmiennych możemy jeszcze zapewnić zachodzenie nierówności

$$d_{11} \geqslant d_{22} \geqslant \ldots \geqslant d_{II}. \tag{6.5}$$

Zadanie proponujemy teraz stawiać następująco: wyznaczyć wektory

$$\boldsymbol{u} = [u_1, \ldots, u_I]' \quad \text{i} \quad \boldsymbol{v} = [v_1, \ldots, v_I]'$$

spełniające warunki

$$\boldsymbol{u}'\boldsymbol{u} = 1, \quad \boldsymbol{v}'\boldsymbol{v} = 1, \tag{6.6}$$

i

$$\boldsymbol{u}'\boldsymbol{v} = 0, \tag{6.7}$$

czyli ortogonalne wektory jednostkowej długości, dla których osiąga maksimum suma

$$Z = \boldsymbol{u}'\boldsymbol{D}\boldsymbol{u} + \boldsymbol{v}'\boldsymbol{D}\boldsymbol{v},$$

i za najlepiej dyskryminującą parę cech uznać $\boldsymbol{u}'\boldsymbol{z}$ i $\boldsymbol{v}'\boldsymbol{z}$, gdzie $\boldsymbol{u}$ i $\boldsymbol{v}$ spełniają (6.6) i (6.7) oraz maksymizują Z. Maksymalną wartość Z można traktować jako wskaźnik dobroci dyskryminacji.

Łatwo widzieć, że Z nie zmienia wartości, gdy wektory $\boldsymbol{u}$ i $\boldsymbol{v}$ zastąpimy przez inne wektory ortogonalne jednostkowej długości leżące w tej samej płaszczyźnie, co wektory $\boldsymbol{u}$ i $\boldsymbol{v}$. Jeśli bowiem

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

jest macierzą ortogonalną, to

$$\boldsymbol{c} = \alpha\boldsymbol{u}+\beta\boldsymbol{v} \quad \text{i} \quad \boldsymbol{d} = \gamma\boldsymbol{u}+\delta\boldsymbol{v}$$

są znów wektorami ortogonalnymi jednostkowej długości leżącymi w tej samej płaszczyźnie, co $\boldsymbol{u}$ i $\boldsymbol{v}$. Mamy zaś dla nich

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{D}\boldsymbol{c}+\boldsymbol{d}'\boldsymbol{D}\boldsymbol{d} &= \\ &= (\alpha\boldsymbol{u}'+\beta\boldsymbol{v}')\boldsymbol{D}(\alpha\boldsymbol{u}+\beta\boldsymbol{v})+(\gamma\boldsymbol{u}'+\delta\boldsymbol{v}')\boldsymbol{D}(\gamma\boldsymbol{u}+\delta\boldsymbol{v}) = \\ &= (\alpha^2+\gamma^2)\boldsymbol{u}'\boldsymbol{D}\boldsymbol{u}+(\beta^2+\delta^2)\boldsymbol{v}'\boldsymbol{D}\boldsymbol{v}+ \\ &\quad +(\alpha\beta+\gamma\delta)\boldsymbol{u}'\boldsymbol{D}\boldsymbol{v}+(\alpha\beta+\gamma\delta)\boldsymbol{v}'\boldsymbol{D}\boldsymbol{u} = \\ &= \boldsymbol{u}'\boldsymbol{D}\boldsymbol{u}+\boldsymbol{v}'\boldsymbol{D}\boldsymbol{v}. \end{aligned}$$

W tym sensie Z zależy tylko od płaszczyzny, a nie od wektorów, które ją wyznaczają.

Wobec (6.4) możemy teraz napisać

$$Z = d_{11}(u_1^2+v_1^2)+d_{22}(u_2^2+v_2^2)+\ldots+d_{II}(u_I^2+v_I^2),$$

a stąd i z (6.5) wnioskujemy, że Z osiąga największą wartość $d_{11}+d_{22}$ dla $\boldsymbol{u} = [1, 0, \ldots, 0]'$ i $\boldsymbol{v} = [0, 1, 0, \ldots, 0]'$. Rozwiązaniem jest zatem płaszczyzna wyznaczona przez wektory własne macierzy $\boldsymbol{D}$ odpowiadające jej dwu największym wartościom własnym.

Interesujące jest wyrazić warunki (6.6) i (6.7) oraz Z w starych zmiennych.

Otóż mnożąc (6.1) lewostronnie przez $\boldsymbol{A}^{-1}$ i prawostronnie przez $\boldsymbol{A}'^{-1}$, znajdujemy, że

$$\boldsymbol{\Lambda} = \boldsymbol{A}^{-1}\boldsymbol{A}'^{-1}.$$

Stąd zaś, przez przejście do macierzy odwrotnych, mamy

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{\Lambda}^{-1}.$$

Wektory $\boldsymbol{u}$ i $\boldsymbol{v}$ dające maksymalną wartość Z wyznaczaliśmy rachując w zmiennych z. Dlatego między wektorami $\boldsymbol{u}$ i $\boldsymbol{v}$ a wektorami $\boldsymbol{a}$ i $\boldsymbol{b}$ odnoszącymi się do wyjściowych zmiennych $\boldsymbol{x}$ mamy wobec (6.2) związki

$$\boldsymbol{u} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{a} \quad \text{i} \quad \boldsymbol{v} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{b}.$$

Podstawiając to do (6.6) i (6.7), otrzymujemy po redukcji

$$\boldsymbol{a}'\Lambda^{-1}\boldsymbol{a} = 1, \quad \boldsymbol{b}'\Lambda^{-1}\boldsymbol{b} = 1, \tag{6.8}$$

oraz

$$\boldsymbol{a}'\Lambda^{-1}\boldsymbol{b} = 0. \tag{6.9}$$

W (6.8) rozpoznajemy żądanie, żeby wektory $\boldsymbol{a}$ i $\boldsymbol{b}$ miały jednostkową długość w sensie metryki Mahalanobisa związanej z macierzą kowariancji Λ (porównaj [11], str. 312 i następne). Podobnie (6.9) wyraża ortogonalność w sensie metryki Mahalanobisa.

Wyrażenie Z za pomocą wektorów $\boldsymbol{a}$ i $\boldsymbol{b}$ odnoszących się do starych zmiennych jest następujące:

$$Z = \boldsymbol{a}'\Lambda^{-1}\boldsymbol{M}\Lambda^{-1}\boldsymbol{a} + \boldsymbol{b}'\Lambda^{-1}\boldsymbol{M}\Lambda^{-1}\boldsymbol{b}.$$

Najlepiej dyskryminująca para cech dana jest przez

$$\boldsymbol{a}'\Lambda^{-1}\boldsymbol{x} \quad \text{i} \quad \boldsymbol{b}'\Lambda^{-1}\boldsymbol{x}, \tag{6.10}$$

gdzie $\boldsymbol{a}$ i $\boldsymbol{b}$ spełniają (6.8) i (6.9) oraz maksymizują Z.

§ 7. Związki z kryteriami spotykanymi w literaturze. Zauważmy najpierw, że ze względu na (6.1) i (6.3) wartości własne macierzy $\boldsymbol{D}$, które są pierwiastkami równania

$$|\boldsymbol{D} - \lambda\boldsymbol{I}| = 0, \tag{7.1}$$

są takie same, jak pierwiastki równania

$$|\boldsymbol{M} - \lambda\Lambda| = 0. \tag{7.2}$$

Gdy ponadto jeszcze raz zmienimy zmienne, oznaczając

$$\mathfrak{a} = \Lambda^{-1}\boldsymbol{a} \quad \text{i} \quad \mathfrak{b} = \Lambda^{-1}\boldsymbol{b},$$

otrzymamy zadanie w następującej postaci: znaleźć wektory $\mathfrak{a}$ i $\mathfrak{b}$ spełniające warunki

$$\mathfrak{a}'\Lambda\mathfrak{a} = 1, \quad \mathfrak{b}'\Lambda\mathfrak{b} = 1, \quad \mathfrak{a}'\Lambda\mathfrak{b} = 0 \tag{7.3}$$

i maksymizujące

$$Z = \mathfrak{a}'\boldsymbol{M}\mathfrak{a} + \mathfrak{b}'\boldsymbol{M}\mathfrak{b}$$

i za najlepiej dyskryminującą parę cech uznać $\mathfrak{a}'\boldsymbol{x}$ i $\mathfrak{b}'\boldsymbol{x}$, gdzie $\mathfrak{a}$ i $\mathfrak{b}$ spełniają (7.3) i maksymizują Z. Jest to dokładnie zadanie rozpatrywane przez Rao ([10], § 9c).

Wreszcie możemy zadanie stawiać jeszcze inaczej (porównaj Wilks [12], § 18.7). Intuicyjnie mówiąc, dobrze dyskryminuje ta para cech, dla której rozrzut środków ciężkości populacji jest duży w porównaniu do rozrzutu wewnątrz populacji. Otóż weźmy za podstawę rozważań wyznaczniki z macierzy kowariancji, które, jak wiadomo, są równe iloczynowi wartości własnych tych macierzy. Zgodnie z tą myślą uznajmy za najlepiej dyskryminującą parę cech taką parę $\boldsymbol{a}'\boldsymbol{x}$ i $\boldsymbol{b}'\boldsymbol{x}$, dla której iloraz wyznacznika macierzy kowariancji wewnątrz populacji przez wyznacznik sumy macierzy kowariancji wewnątrz populacji i między środkami ciężkości populacji jest najmniejszy. Symbolicznie: idzie o znalezienie macierzy

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}' \\ \boldsymbol{b}' \end{bmatrix}, \tag{7.4}$$

dla której

$$\frac{|\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}'|}{|\boldsymbol{B}(\boldsymbol{M}+\boldsymbol{\Lambda})\boldsymbol{B}'|} = \min. \tag{7.5}$$

To ekstremum nie jest wyznaczone jednoznacznie, bo nieosobliwa liniowa transformacja cech nie zmienia ilorazu po lewej stronie (7.5). Tak zresztą być powinno — dobroć dyskryminacji powinna być niezmiennicza ze względu na nieosobliwe transformacje cech, a to dla kryterium (7.5) jest bezpośrednio widoczne. Okazuje się, że minimum w (7.5) jest w każdym razie osiągnięte dla rozwiązania według poprzedniego kryterium.

Prace cytowane

[1] T. W. Anderson, *Introduction to multivariate analysis*, New York—London 1958.

[2] T. Bogdanik, M. Warmus i J. Wartak, *Matematyczny model krzywej cukrowej*, Polskie Archiwum Medycyny Wewnętrznej 39 (1967), nr 2, str. 211-217.

[3] T. Bogdanik, J. Chlebowski, J. Rostafińska, M. Warmus, J. Wartak i A. Wasilewska, *Dyskryminacja matematyczna krzywych cukrowych w otyłości, nadciśnieniu i zawale mięśnia sercowego*, Polskie Archiwum Medycyny Wewnętrznej 39 (1967), nr 2, str. 143-150.

[4] T. Bogdanik, M. Warmus i J. Wartak, *Praktyczne zastosowania modeli matematycznych do różnicowania typów cukrzycy*, Polskie Archiwum Medycyny Wewnętrznej 39 (1967), nr 2, str. 123-130.

[5] T. Bogdanik, B. Bogdanikowa, M. Warmus i J. Wartak, *Zastosowanie metod matematycznych do różnicowania typów dysproteinemii*, w książce: *Zastosowania metod matematycznych w medycynie*, Ossolineum, Wrocław—Warszawa—Kraków 1966, str. 75-91.

[6] T. Bogdanik, M. Warmus i J. Wartak, *Zastosowanie modeli matematycznych do wszechstronnej diagnozy cukrzycy*, referat wygłoszony na sesji naukowej Polskiego Towarzystwa Biometrycznego we Wrocławiu dnia 13. II. 1967.

[7] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, Annals of Eugenics 7 (1936), str. 179-188 (cytowane za [1]).

[8] —, *The statistical utilization of multiple measurements*, Annals of Eugenics 8 (1938), str. 376-386 (cytowane za [1]).

[9] J. Perkal, *O dyskryminacji płaskiej*, Zastosow. Matem. 9 (1968), str. 315-324.

[10] C. R. Rao, *Advanced statistical methods in biometric research*, New York—London 1952.

[11] L. Schmetterer, *Mathematische Statistik*, Wien 1956.

[12] S. S. Wilks, *Mathematical statistics*, New York—London 1962.

Praca wpłynęła 5. 4. 1967

С. ЗУБЖИЦКИ (Вроцлав)

ОБ ОПРЕДЕЛЕНИИ ПЛОСКОСТИ НАИЛУЧШЕЙ ДИСКРИМИНАЦИИ

РЕЗЮМЕ

Рассматривается проблема классификации индивидов, решающей вопрос принадлежности к одной из некоторого числа популяций, на основании коллектива квантитативных признаков. Предполагаются известными средние в отдельных популяциях и общая для всех матрица моментов второго порядка. В таких условиях можно пытаться проектировать всю совокупность наблюдений на выбранную плоскость для получения наглядной картины. В статьи дается дискуссия некоторых требований, приводящих к разным определениям наилучшей плоскости проектирования.

S. ZUBRZYCKI (Wrocław)

ON THE CHOICE OF A PLANE OF BEST DISCRIMINATION

SUMMARY

The problem of classifying individuals into one of a number of populations on the basis of a collection of numerical features is considered. The knowledge of population means and of the covariance matrix (which is supposed to be common for all the populations under considerations) is assumed. In this situation one might want to get a visible picture of the situation by projecting the whole set of observations on a suitably chosen plane. In the paper comments are given on a few definitions of the best choice of such a plane.