

1. Dane

$$X = \begin{bmatrix} 160 & 80 \\ 180 & 60 \\ 170 & 76 \end{bmatrix}$$

zawierają wzrost (w cm) i masę ciała (w kg) trzech osób.

a. Zestandardyzuj te dane.

b. Znajdź macierz korelacji tych danych.

b. Znajdź najlepszy model jednowymiarowy Y danych X .

c. Oblicz błąd względny tego modelu.

d. Przeprowadź reifikację tego modelu używając korelacji składowych głównych ze zmiennymi oryginalnymi.

2. Macierz X przedstawia dane ekonomiczne 4 państw Unii Europejskiej i Polski w momencie wprowadzenia euro w 1999 roku. D oznacza dochód narodowy na głowę (w tys \$), I inflację w %, W procent ludności z wykształceniem wyższym

$$\begin{bmatrix} & D & I & W \\ \text{Polska} & 7,5 & 9,2 & 7 \\ \text{Portugalia} & 13,8 & 2,6 & 12 \\ \text{Hiszpania} & 15,1 & 1,4 & 17 \\ \text{Irlandia} & 19,8 & 2,2 & 20 \\ \text{Niemcy} & 20,9 & 0,6 & 23 \end{bmatrix}$$

a. Przedstaw te dane w postaci standaryzowanej $Z(X)$, oblicz macierz korelacji. Dlaczego zapis w postaci standaryzowanej jest w tym przypadku adekwatny?

b. Które ze zmiennych są z sobą najbardziej związane? Skomentuj ten związek.

c. Wykonaj rzut macierzy $Z(X)$ na podprzestrzeń dochodu i procentu osób z wykształceniem wyższym (po standaryzacji).

d. Oblicz odległość macierzy $Z(X)$ i jej rzutu z punktu c. Jak możesz interpretować tę odległość?

e. Oblicz składowe główne i odpowiadające im wartości własne (możesz wykorzystać bezpłatny program Wolfram Alpha (<http://www.wolframalpha.com>)) i funkcje *eigenvalues* i *eigenvectors*.

f. Znajdź na kole korelacyjnym w układzie dwóch pierwszych składowych głównych punkty *dochód* (D), *inflacja* (I) i *wykształcenie wyższe* (W). Narysuj odpowiedni wykres. Jaka jest interpretacja praktyczna pierwszej i drugiej składowej głównej?

g. Wybierz liczbę składowych głównych d tak, aby błąd względny modelu będącego rzutem na hiperpłaszczyznę o kierunku podprzestrzeni rozpiętej na d składowych nie przekraczał 10%. Wyraź dane w bazie d składowych.

h. Oblicz odległość macierzy $Z(X)$ i jej rzutu z punktu g. Porównaj z odległością z punktu d.