Algorytmy eksploracji danych: Wykład 3

Wojciech Kempa

Politechnika Śląska
Wydział Matematyki Stosowanej

Metody szacowania zmienności wspólnej (1)

- Aby zbudować model analizy czynnikowej dla układu zmiennych obserwowalnych $X_1, ..., X_p$ konieczne jest oszacowanie elementów macierzy A (czyli tzw. ładunków czynnikowych), tworzących w konsekwencji tzw. **zasoby zmienności wspólnej**, a także wariancji **czynników swoistych**, czyli elementów $\Psi_1, ..., \Psi_p$.

Metody szacowania zmienności wspólnej (1)

- Aby zbudować model analizy czynnikowej dla układu zmiennych obserwowalnych $X_1, ..., X_p$ konieczne jest oszacowanie elementów macierzy A (czyli tzw. ładunków czynnikowych), tworzących w konsekwencji tzw. **zasoby zmienności wspólnej**, a także wariancji **czynników swoistych**, czyli elementów $\Psi_1, ..., \Psi_p$.
- Istotne w budowie modelu analizy czynnikowej jest, by zasób zmienności wspólnej stanowił jak największą część wariancji całkowitej danej zmiennej (wówczas zmienna ta jest „wyjaśniana" w dużym stopniu przez zestaw czynników wspólnych $F_1, ..., F_k$). Dążymy zatem do zmarginalizowania czynników swoistych każdej zmiennej.

Metody szacowania zmienności wspólnej (1)

- Aby zbudować model analizy czynnikowej dla układu zmiennych obserwowalnych $X_1, ..., X_p$ konieczne jest oszacowanie elementów macierzy A (czyli tzw. ładunków czynnikowych), tworzących w konsekwencji tzw. **zasoby zmienności wspólnej**, a także wariancji **czynników swoistych**, czyli elementów $\Psi_1, ..., \Psi_p$.
- Istotne w budowie modelu analizy czynnikowej jest, by zasób zmienności wspólnej stanowił jak największą część wariancji całkowitej danej zmiennej (wówczas zmienna ta jest „wyjaśniana" w dużym stopniu przez zestaw czynników wspólnych $F_1, ..., F_k$). Dążymy zatem do zmarginalizowania czynników swoistych każdej zmiennej.
- Ich obecność sprawia, że na głównej przekątnej macierzy współczynników korelacji R pomiędzy wyjściowymi zmiennymi (równej macierzy kowariancji S, ponieważ zmienne są zestandaryzowane) znajdują się jedynki (odpowiadające doskonałej korelacji każdej ze zmiennych ze sobą).

Metody szacowania zmienności wspólnej (2)

Gdyby nie brać pod uwagę czynników swoistych, otrzymamy tzw. **zredukowaną macierz współczynników korelacji** $\widetilde{R}$ postaci

$$\widetilde{R} \stackrel{def}{=} \begin{bmatrix} h_1^2 & r_{1,2} & \dots & r_{1,p} \\ r_{2,1} & h_2^2 & \dots & r_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p,1} & r_{p,2} & \dots & h_p^2 \end{bmatrix},$$

gdzie $r_{i,j}$ oznacza współczynnik korelacji pomiędzy zmiennymi X_i oraz X_j.

Na głównej przekątnej macierzy $\widetilde{R}$ znajdują się zasoby zmienności wspólnej, natomiast pozostałe jej elementy są równe współczynnikom korelacji pomiędzy wyjściowymi zmiennymi.

Zauważmy, że prawdziwy jest teraz następujący wzór:

$$\widetilde{R} = A \cdot A^T, \tag{1}$$

a zatem dla zredukowanej macierzy korelacji model analizy czynnikowej możemy zapisać w postaci macierzowej w następujący sposób:

$$\mathbf{X}' = A \cdot \mathbf{F},$$

gdzie $\mathbf{X}'$ jest wektorem zmiennych $X_1, ..., X_p$, dla których nie uwzględniamy wpływu czynników swoistych.
Aby próbować oszacować elementy macierzy A korzystając z równania (1), konieczna jest znajomość zredukowanej macierzy korelacji $\widetilde{R}$. Tu z kolei oszacowania wymagają elementy h_i^2 znajdujące się na głównej przekątnej (współczynniki korelacji $r_{i,j}$ estymujemy na podstawie próby losowej).

Metody szacowania zmienności wspólnej (4)

W literaturze proponuje się kilka różnych metod szacowania zasobów zmienności wspólnej h_i^2.

- **Metoda triad**: w metodzie tej przyjmujemy $h_i^2 = \frac{r_{i,j}^* r_{i,k}^*}{r_{j,k}}$, gdzie $r_{i,j}^*$ oraz $r_{i,k}^*$ to dwa największe współczynniki korelacji pomiędzy zmienną X_i, a pozostałymi zmiennymi ($i \neq j$, $i \neq k$).

Metody szacowania zmienności wspólnej (4)

W literaturze proponuje się kilka różnych metod szacowania zasobów zmienności wspólnej h_i^2.

- **Metoda triad**: w metodzie tej przyjmujemy $h_i^2 = \frac{r_{i,j}^* r_{i,k}^*}{r_{j,k}}$, gdzie $r_{i,j}^*$ oraz $r_{i,k}^*$ to dwa największe współczynniki korelacji pomiędzy zmienną X_i, a pozostałymi zmiennymi ($i \neq j$, $i \neq k$).
- **Metoda najsilniejszej korelacji**: w metodzie tej przyjmuje się $h_i^2 = \max_j |r_{i,j}|$, gdzie $i \neq j$.

Metody szacowania zmienności wspólnej (4)

W literaturze proponuje się kilka różnych metod szacowania zasobów zmienności wspólnej h_i^2.

- **Metoda triad**: w metodzie tej przyjmujemy $h_i^2 = \frac{r_{i,j}^* r_{i,k}^*}{r_{j,k}}$, gdzie $r_{i,j}^*$ oraz $r_{i,k}^*$ to dwa największe współczynniki korelacji pomiędzy zmienną X_i, a pozostałymi zmiennymi ($i \neq j$, $i \neq k$).
- **Metoda najsilniejszej korelacji**: w metodzie tej przyjmuje się $h_i^2 = \max_j |r_{i,j}|$, gdzie $i \neq j$.
- **Metoda średniej korelacji**: przyjmujemy tu $h_i^2 = \frac{1}{p-1} \sum_{k=1}^{p} r_{i,k}$.

Metody szacowania zmienności wspólnej (4)

W literaturze proponuje się kilka różnych metod szacowania zasobów zmienności wspólnej h_i^2.

- **Metoda triad**: w metodzie tej przyjmujemy $h_i^2 = \frac{r_{i,j}^* r_{i,k}^*}{r_{j,k}}$, gdzie $r_{i,j}^*$ oraz $r_{i,k}^*$ to dwa największe współczynniki korelacji pomiędzy zmienną X_i, a pozostałymi zmiennymi ($i \neq j$, $i \neq k$).
- **Metoda najsilniejszej korelacji**: w metodzie tej przyjmuje się $h_i^2 = \max_j |r_{i,j}|$, gdzie $i \neq j$.
- **Metoda średniej korelacji**: przyjmujemy tu $h_i^2 = \frac{1}{p-1} \sum_{k=1}^{p} r_{i,k}$.
- **Metoda korelacji wielorakiej**: wykorzystuje się w niej **współczynnik korelacji wielorakiej** pomiędzy ustaloną zmienną X_i, a pozostałymi zmiennymi wyjściowego układu.

Metody szacowania zmienności wspólnej (5)

- Współczynnik ten wyraża łączny wpływ zmiennych $X_1, ..., X_{i-1}, X_{i+1}, ..., X_p$ na zmienną X_i. Przyjmuje on wartości z przedziału $[0, 1]$ (im większa jego wartość, tym korelacja jest silniejsza) i jest zdefiniowany w następujący sposób:

$$R_{i.1...(i-1)(i+1)...p} \stackrel{def}{=} \sqrt{1 - \frac{|R|}{R_{i,i}}},$$

gdzie $R_{i,i}$ jest dopełnieniem algebraicznym elementu $r_{i,i}$ macierzy R.

- Współczynnik ten wyraża łączny wpływ zmiennych $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p$ na zmienną X_i. Przyjmuje on wartości z przedziału $[0, 1]$ (im większa jego wartość, tym korelacja jest silniejsza) i jest zdefiniowany w następujący sposób:

$$R_{i.1\ldots(i-1)(i+1)\ldots p} \stackrel{def}{=} \sqrt{1 - \frac{|R|}{R_{i,i}}},$$

gdzie $R_{i,i}$ jest dopełnieniem algebraicznym elementu $r_{i,i}$ macierzy R.

- W metodzie korelacji wielorakiej przyjmujemy $h_i^2 = R^2_{i.1\ldots(i-1)(i+1)\ldots p}$. Metoda ta jest najczęściej stosowana w praktyce, uwzględnia ona całą macierz współczynników korelacji (jest jedyną metodą stosowaną w pakiecie *STATISTICA*).

- **Metoda pierwszego czynnika centroidalnego:** metoda ta również uwzględnia wszystkie elementy macierzy korelacji R,

Przyjmujemy w niej $h_i^2 = \dfrac{\left(\sum\limits_{k=1}^{p} r_{i,k}\right)^2}{\sum\limits_{j=1}^{p}\sum\limits_{k=1}^{p} r_{j,k}}$.

Metody szacowania ładunków wspólnych

Wyodrębnienie ładunków wspólnych jest związane ze znalezieniem nowego układu współrzędnych, który zastąpi p-wymiarowy układ wyjściowy. Do najważniejszych metod szacowania ładunków wspólnych należą:

- **metoda centroidalna**;
- **metoda składowych głównych**;
- **metoda największej wiarygodności**.

Omówimy w szczegółach metodę centroidalną oraz metodę składowych głównych.

Metoda centroidalna (1)

- Podstawową i najstarszą zarazem metodą wyodrębniania ładunków wspólnych, zaprezentowaną przez L. Thurstone'a w 1947 roku, jest **metoda centroidalna**, zwana także **metodą środków ciężkości**.

Metoda centroidalna (1)

- Podstawową i najstarszą zarazem metodą wyodrębniania ładunków wspólnych, zaprezentowaną przez L. Thurstone'a w 1947 roku, jest **metoda centroidalna**, zwana także **metodą środków ciężkości**.
- Wektory reprezentujące wariancje zmiennych $X_1, ..., X_p$ mają swój początek w początku O układu współrzędnych (są równoległe do odpowiednich osi tego układu). Ich końce natomiast wyznaczają środek ciężkości S_1 układu wektorów.

Metoda centroidalna (1)

- Podstawową i najstarszą zarazem metodą wyodrębniania ładunków wspólnych, zaprezentowaną przez L. Thurstone'a w 1947 roku, jest **metoda centroidalna**, zwana także **metodą środków ciężkości**.
- Wektory reprezentujące wariancje zmiennych $X_1, ..., X_p$ mają swój początek w początku O układu współrzędnych (są równoległe do odpowiednich osi tego układu). Ich końce natomiast wyznaczają środek ciężkości S_1 układu wektorów.

- Pierwsza oś nowego układu współrzędnych (odpowiadająca czynnikowi F_1) będzie przechodzić przez punkt O oraz punkt S_1 będący środkiem ciężkości.

Metoda centroidalna (1)

- Podstawową i najstarszą zarazem metodą wyodrębniania ładunków wspólnych, zaprezentowaną przez L. Thurstone'a w 1947 roku, jest **metoda centroidalna**, zwana także **metodą środków ciężkości**.
- Wektory reprezentujące wariancje zmiennych $X_1, ..., X_p$ mają swój początek w początku O układu współrzędnych (są równoległe do odpowiednich osi tego układu). Ich końce natomiast wyznaczają środek ciężkości S_1 układu wektorów.
- Pierwsza oś nowego układu współrzędnych (odpowiadająca czynnikowi F_1) będzie przechodzić przez punkt O oraz punkt S_1 będący środkiem ciężkości.
- Po utworzeniu zredukowanej macierzy korelacji $\widetilde{R}$, sumujemy wartości występujące w poszczególnych kolumnach tej macierzy. Dla i-tej kolumny otrzymujemy następującą sumę:

$$T_i \stackrel{def}{=} r_{1,i} + ... + r_{i-1,i} + h_i^2 + r_{i+1,i} + ... + r_{p,i}.$$

Metoda centroidalna (2)

Niech $T \stackrel{def}{=} \sum_{i=1}^{p} T_i$. Wartość pierwszego ładunku czynnikowego dla zmiennej X_i obliczamy ze wzoru

$$a_{i,1} = \frac{T_i}{\sqrt{T}}.$$

Po obliczeniu wartości ładunków pierwszego czynnika F_1, wyodrębniamy wielkości $a_{i,2}$, czyli ładunki czynnika F_2. Definiujemy macierz R_1 pierwszych pozostałości zredukowanej macierzy korelacji $\widetilde{R}$ (tzw. **macierz rezydualną**) w następujący sposób:

$$R_1 \stackrel{def}{=} \widetilde{R} - Q_1,$$

gdzie

$$Q_1 \stackrel{def}{=} \begin{bmatrix} a_{1,1}^2 & a_{1,1}a_{2,1} & \dots & a_{1,1}a_{p,1} \\ a_{2,1}a_{1,1} & a_{2,1}^2 & \dots & a_{2,1}a_{p,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p,1}a_{1,1} & a_{p,1}a_{2,1} & \dots & a_{p,1}^2 \end{bmatrix}.$$

- Jak wiemy, obliczając ładunki dla pierwszego czynnika F_1, poprowadziliśmy oś pierwszego czynnika przez środek ciężkości wyznaczony przez końce wektorów reprezentujących wariancje zmiennych $X_1, ..., X_p$.

Metoda centroidalna (3)

- Jak wiemy, obliczając ładunki dla pierwszego czynnika F_1, poprowadziliśmy oś pierwszego czynnika przez środek ciężkości wyznaczony przez końce wektorów reprezentujących wariancje zmiennych $X_1, ..., X_p$.
- Usunięcie części wariancji wspólnej zwiazanej z pierwszym czynnikiem spowoduje przesunięcie punktu O w miejsce punktu S_1 wzdłuż osi pierwszego czynnika.

Metoda centroidalna (3)

- Jak wiemy, obliczając ładunki dla pierwszego czynnika F_1, poprowadziliśmy oś pierwszego czynnika przez środek ciężkości wyznaczony przez końce wektorów reprezentujących wariancje zmiennych $X_1, ..., X_p$.
- Usunięcie części wariancji wspólnej zwiazanej z pierwszym czynnikiem spowoduje przesunięcie punktu O w miejsce punktu S_1 wzdłuż osi pierwszego czynnika.
- Otrzymamy oś drugiego czynnika F_2, która jest prostopadła do osi pierwszego czynnika i przechodzi przez punkt $O = S_1$.

Metoda centroidalna (3)

- Jak wiemy, obliczając ładunki dla pierwszego czynnika F_1, poprowadziliśmy oś pierwszego czynnika przez środek ciężkości wyznaczony przez końce wektorów reprezentujących wariancje zmiennych $X_1, ..., X_p$.
- Usunięcie części wariancji wspólnej zwiazanej z pierwszym czynnikiem spowoduje przesunięcie punktu O w miejsce punktu S_1 wzdłuż osi pierwszego czynnika.
- Otrzymamy oś drugiego czynnika F_2, która jest prostopadła do osi pierwszego czynnika i przechodzi przez punkt $O = S_1$.
- Łączymy teraz końce początkowych wektorów z punktem S_1. Nowo powstałe wektory są krótsze od wektorów początkowych, co jest związane z usunięciem części wariancji wspólnej związanej z czynnikiem F_1.

Metoda centroidalna (4)

- Sumy współczynników występujących w każdej z kolumn macierzy R_1 są bardzo bliskie zeru. Korespondują one z odległościami końców „nowych” wektorów od początku S_1 „nowego” układu. Niektóre sumy mogą być dodatnie, inne ujemne.

Metoda centroidalna (4)

- Sumy współczynników występujących w każdej z kolumn macierzy R_1 są bardzo bliskie zeru. Korespondują one z odległościami końców „nowych" wektorów od początku S_1 „nowego" układu. Niektóre sumy mogą być dodatnie, inne ujemne.
- Aby otrzymać dodatnie sumy w kolumnach macierzy R_1, wykorzystujemy rekurencyjną, dość skomplikowaną, procedurę. Jej efektem jest zmiana zwrotu (bez zmiany długości) niektórych wektorów reprezentujących wariancje zmiennych pozostałe do wyjaśnienia po „odjęciu" części wyjaśnionej przez F_1.

Metoda centroidalna (4)

- Sumy współczynników występujących w każdej z kolumn macierzy R_1 są bardzo bliskie zeru. Korespondują one z odległościami końców „nowych" wektorów od początku S_1 „nowego" układu. Niektóre sumy mogą być dodatnie, inne ujemne.
- Aby otrzymać dodatnie sumy w kolumnach macierzy R_1, wykorzystujemy rekurencyjną, dość skomplikowaną, procedurę. Jej efektem jest zmiana zwrotu (bez zmiany długości) niektórych wektorów reprezentujących wariancje zmiennych pozostałe do wyjaśnienia po „odjęciu" części wyjaśnionej przez F_1.
- W konsekwencji wszystkie wektory znajdą się teraz po jednej i tej samej stronie osi pierwszego czynnika F_1.

Metoda centroidalna (5)

- W nowo otrzymanej w ten sposób macierzy R_1 sumujemy elementy w poszczególnych kolumnach. Otrzymujemy nową wartość T_i dla i-tej kolumny tej macierzy.

Metoda centroidalna (5)

- W nowo otrzymanej w ten sposób macierzy R_1 sumujemy elementy w poszczególnych kolumnach. Otrzymujemy nową wartość T_i dla i-tej kolumny tej macierzy.
- Przyjmując, podobnie jak poprzednio, $T = \sum_{i=1}^{p} T_i$, ładunek drugiego czynnika obliczamy ze wzoru

$$a_{i,2} = \frac{T_i}{\sqrt{T}}.$$

Metoda centroidalna (6)

- Na końcu należy jeszcze sprawdzić, jaki znak będzie miała wartość i-tego ładunku.

Metoda centroidalna (6)

- Na końcu należy jeszcze sprawdzić, jaki znak będzie miała wartość i-tego ładunku.
- Jeżeli w procedurze rekurencyjnej zmienna X_i była odwracana parzystą ilość razy (w szczególności wcale), to znak jej drugiego ładunku czynnikowego jest taki sam, jak znak poprzedniego (tu: pierwszego) ładunku.

Metoda centroidalna (6)

- Na końcu należy jeszcze sprawdzić, jaki znak będzie miała wartość i-tego ładunku.
- Jeżeli w procedurze rekurencyjnej zmienna X_i była odwracana parzystą ilość razy (w szczególności wcale), to znak jej drugiego ładunku czynnikowego jest taki sam, jak znak poprzedniego (tu: pierwszego) ładunku.
- Jeżeli zmienna ta była odwracana nieparzystą ilość razy, to jej drugi ładunek czynnikowy będzie miał przeciwny znak do ładunku pierwszego.

Metoda centroidalna (6)

- Na końcu należy jeszcze sprawdzić, jaki znak będzie miała wartość i-tego ładunku.
- Jeżeli w procedurze rekurencyjnej zmienna X_i była odwracana parzystą ilość razy (w szczególności wcale), to znak jej drugiego ładunku czynnikowego jest taki sam, jak znak poprzedniego (tu: pierwszego) ładunku.
- Jeżeli zmienna ta była odwracana nieparzystą ilość razy, to jej drugi ładunek czynnikowy będzie miał przeciwny znak do ładunku pierwszego.
- W kolejnym kroku wyznaczamy macierz pozostałości macierzy korelacji po wyodrębnieniu drugiego czynnika F_2 i wyznaczamy wartości kolejnych ładunków czynnikowych, stosując powyższą procedurę.

Metoda składowych głównych (1)

- W metodzie tej przyjmujemy, że $\widetilde{R} = R$, czyli zakładamy, że wariancje swoiste są równe zeru.

Metoda składowych głównych (1)

- W metodzie tej przyjmujemy, że $\widetilde{R} = R$, czyli zakładamy, że wariancje swoiste są równe zeru.
- Model analizy czynnikowej redukuje się wówczas do postaci

$$\mathbf{X} = A \cdot \mathbf{F}.$$

Metoda składowych głównych (1)

- W metodzie tej przyjmujemy, że $\widetilde{R} = R$, czyli zakładamy, że wariancje swoiste są równe zeru.
- Model analizy czynnikowej redukuje się wówczas do postaci

$$\mathbf{X} = A \cdot \mathbf{F}.$$

- Dalej postępujemy tak, jak w klasycznej metodzie PCA. Wyznaczamy wartości własne macierzy korelacji R, rozwiązując równanie charakterystyczne $|R - \lambda\mathbb{E}| = 0$.

Metoda składowych głównych (1)

- W metodzie tej przyjmujemy, że $\widetilde{R} = R$, czyli zakładamy, że wariancje swoiste są równe zeru.
- Model analizy czynnikowej redukuje się wówczas do postaci

$$\mathbf{X} = A \cdot \mathbf{F}.$$

- Dalej postępujemy tak, jak w klasycznej metodzie PCA. Wyznaczamy wartości własne macierzy korelacji R, rozwiązując równanie charakterystyczne $|R - \lambda \mathbb{E}| = 0$.
- Z każdą wartością własną λ_i związany jest znormalizowany wektor własny, którego współrzędne posłużą do wyznaczenia ładunków czynnikowych.

Metoda składowych głównych (2)

- Dla i-tej z kolei największej wartości własnej λ_i mamy wówczas

$$F_i = a_{i,1}X_1 + ... + a_{i,p}X_p. \tag{2}$$

gdzie współczynniki przy kolejnych zmiennych $X_1, ..., X_p$ są współrzędnymi wektora własnego odpowiadającego λ_i.

Metoda składowych głównych (2)

- Dla i-tej z kolei największej wartości własnej λ_i mamy wówczas

$$F_i = a_{i,1}X_1 + ... + a_{i,p}X_p. \tag{2}$$

  gdzie współczynniki przy kolejnych zmiennych $X_1, ..., X_p$ są współrzędnymi wektora własnego odpowiadającego λ_i.
- Po wyznaczeniu ładunków czynnikowych możemy określić, jaka część (procent) całkowitej wariancji jest wyznaczona za pomocą danego czynnika wspólnego.

Metoda składowych głównych (2)

- Dla i-tej z kolei największej wartości własnej λ_i mamy wówczas

$$F_i = a_{i,1}X_1 + ... + a_{i,p}X_p. \tag{2}$$

gdzie współczynniki przy kolejnych zmiennych $X_1, ..., X_p$ są współrzędnymi wektora własnego odpowiadającego λ_i.

- Po wyznaczeniu ładunków czynnikowych możemy określić, jaka część (procent) całkowitej wariancji jest wyznaczona za pomocą danego czynnika wspólnego.

- Dla czynnika F_i jest ona wyrażona wzorem (korzystamy tu z faktu, że suma wariancji wszystkich zmiennych znormalizowanych wynosi p)

$$\frac{1}{p}\sum_{i=1}^{p} a_{i,j}^2 \cdot 100\%.$$

Kryteria wyboru ilości czynników wspólnych (1)

Ile czynników wspólnych należy wybrać? W praktyce stosuje się tu różne kryteria, podobne do tych stosowanych w analizie składowych głównych.

- **Kryterium Kaisera**: bierzemy pod uwagę tylko te czynniki, którym odpowiadają wartości własne macierzy korelacji R większe od 1.

Kryteria wyboru ilości czynników wspólnych (1)

Ile czynników wspólnych należy wybrać? W praktyce stosuje się tu różne kryteria, podobne do tych stosowanych w analizie składowych głównych.

- **Kryterium Kaisera**: bierzemy pod uwagę tylko te czynniki, którym odpowiadają wartości własne macierzy korelacji R większe od 1.
- **Kryterium Cattella**: związane z tzw. **wykresem osypiska**, w którym bierzemy pod uwagę tylko te czynniki, którym odpowiadają wartości własne macierzy korelacji położone na lewo od punktu łagodnego spadku łamanej łączącej punkty odpowiadające kolejnym wartościom własnym (w kolejności malejącej). Sam punkt, w którym rozpoczyna się łagodny spadek możemy zaliczyć do tej grupy bądź nie.

- **Kryterium procentowe**: zakłada pozostawienie tylko takiej ilości początkowych czynników wspólnych wyznaczonych dla układu cech, by łączny udział procentowy wyjaśnianej przez nie wariancji przekroczył pewien ustalony próg, np. 75% lub 80%.

Kryteria wyboru ilości czynników wspólnych (2)

- **Kryterium procentowe**: zakłada pozostawienie tylko takiej ilości początkowych czynników wspólnych wyznaczonych dla układu cech, by łączny udział procentowy wyjaśnianej przez nie wariancji przekroczył pewien ustalony próg, np. 75% lub 80%.

- **Kryterium „połowy”**: liczba czynników nie powinna przekroczyć wartości $\frac{p}{2}$, gdzie p oznacza liczbę oryginalnych zmiennych obserwowalnych.