PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

ROBERT KAPŁON

MODELE ANALIZY CZYNNIKOWEJ Z DWOMA ZMIENNYMI UKRYTYMI

1. WPROWADZENIE

Model analizy czynnikowej jest często wykorzystywaną techniką analizy danych wielowymiarowych. W zależności od postawionego celu można (por. [16]):wyodrębnić ukrytą, bezpośrednio nieobserwowalną strukturę zmiennych, zredukować wymiar przestrzeni zmiennych wraz z graficzną ich prezentacją.

W metodzie tej, jak i wielu innych, przyjmuje się *implicite* założenie o jednorodności obserwacji. Jeśli istnieją przesłanki poddające w wątpliwość takie założenie, wtedy należy poszukiwać pewnych modyfikacji modelu, aby jak najdokładniej odzwierciedlić strukturę danych. Jedno z podejść oparte jest na mieszaninie rozkładów. Jego istota sprowadza się do włączenia dodatkowej zmiennej ukrytej, która dzieli badaną zbiorowość na klasy. W konsekwencji możliwa jest jednoczesna estymacja parametrów w każdej klasie.

W pracy [7]zaproponowano taki model wraz z procedurą estymacji parametrów. Uwzględniono tam tylko przypadek ogólny, tzn. założono różne macierze ładunków czynnikowych oraz wariancji specyficznej. Jeśli nałoży się na nie pewne ograniczenia, to można otrzymać modele prostsze, które z jednej strony – mogą z prawie jednakową precyzją odwzorowywać strukturę obserwacji co modele bardziej złożone, z drugiej natomiast – mogą wykazywać większą stabilność estymacji. O tej stabilności wspomina się przy okazji szacowania parametrów mieszanki wielowymiarowych rozkładów normalnych (por. [13]).

Warto nadmienić, że w tzw. klasyfikacji opartej na modelach reprezentujących zadaną klasę (*model-basedclustering*), dokonuje się pewnych uproszczeń macierzy kowariancji, jeśli pochodzi ona z wielowymiarowego rozkładu normalnego. Propozycję taką przedstawili Banfield iRafteryw pracy [3], dokonując dekompozycji spektralnej macierzy. W ten sposób dokonali jej reparametryzacjiwyróżniając: orientację, kształt i rozmiar klasy. To pozwoliło na specyfikację 8 modeli. Bazując na tych ustaleniach Celeux i Govaert[5] rozszerzyli tę liczbę do 14.

Powyższe uwagi składają się na cel niniejszego opracowania jakim jest identyfikacja i budowa modeli analizy czynnikowej z dwoma zmiennymi ukrytymi, zależna od postaci macierzy ładunków czynnikowych i macierzy wariancji specyficznej. Ponadto zostanie zaproponowana procedura estymacji dla każdego modelu oraz kryteria, jakimi można się posłużyć, przy wyborze jednego z nich.

2. MODELE ANALIZY CZYNNIKOWEJ

Model analizy czynnikowej uwzględniający niejednorodność obserwacji można przedstawić wpostaci [7]:

$$\mathbf{y}_i = \boldsymbol{\mu}_c + \boldsymbol{\Lambda}_c \mathbf{b}_i + \mathbf{e}_i,$$

gdzie: $[\mathbf{y}_i]_{px1}$ – wektor obserwacji, $[\boldsymbol{\mu}_c]_{px1}$– wektor wartości przeciętnych,$[\boldsymbol{\Lambda}_c]_{pxq}$ – macierz ładunków czynnikowych, $[\mathbf{b}_i]_{qx1}$ – zmienna ukryta, $[\mathbf{e}_i]_{px1}$ – składnik losowy. Przy założeniu, że obserwacja i należy do klasy $K_c (c = 1, \ldots, C)$, rozkłady warunkowe mają postać:

$$\begin{aligned}
(\mathbf{b}_i | i \in K_c) &\sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \\
(\mathbf{e}_i | i \in K_c) &\sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi}_c), \\
(\mathbf{a}_i | \mathbf{b}_i; i \in K_c) &\sim \mathcal{N}_p(\boldsymbol{\mu}_c + \boldsymbol{\Lambda}_c \mathbf{b}_i, \boldsymbol{\Psi}_c),
\end{aligned}$$

$$gdzie : \boldsymbol{\Psi}_c = diag(\sigma_{1c}^2, \ldots, \sigma_{pc}^2).$$

Niech $p(\pi_c)$ będzie prawdopodobieństwem przynależności obserwacji i do klasy K_c. Wtedy rozkład wektora zaobserwowanych odpowiedzi można zapisać ogólnie:

$$f(\mathbf{y}_i) = \sum_{c=1}^{C} \int f(y_i | \mathbf{b}_i, \pi_c) f(\mathbf{b}_i | \pi_c) p(\pi_c) \mathrm{d}\mathbf{b}_i,$$

$$f(\mathbf{y}_i) = \sum_{C}^{\sum} c = 1, \pi_c) f(\mathbf{b}_i | \pi_c) p(\pi_c) \mathrm{d}\mathbf{b}_i$$

lub, przy uwzględnieniu warunkowych rozkładów normalnych, w postaci:

$$f(\mathbf{y}_i) \propto \sum_{c=1}^{C} p(\pi_c) |\boldsymbol{\Gamma}_c|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Gamma}_c^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_c) \right],$$

$$\boldsymbol{\Gamma}_c = \boldsymbol{\Lambda}_c \boldsymbol{\Lambda}_c^T + \boldsymbol{\Psi}_c.$$

W zależności od ograniczeń, jakie nakłada się na składowe macierzy kowariancji $\boldsymbol{\Gamma}_c$, wyróżnia się osiem modeli. W każdym z nich macierz ładunków czynnikowych może być taka sama ($\boldsymbol{\Lambda}_c = \boldsymbol{\Lambda}$) lub różnić się między klasami. W wypadku macierzy wariancji specyficznej, oprócz różnic miedzy klasami, uwzględnić można różnice wewnątrz klas. W konsekwencji pojawiają się 4 możliwości: brak jakichkolwiek ograniczeń – $\boldsymbol{\Psi}_c$, ograniczenia dotyczą wyłącznie różnicwewnątrz klas – $\boldsymbol{\Psi}_c = \boldsymbol{\Psi}$ lub różnic między klasami – $\boldsymbol{\Psi}_c = \sigma_c^2 \mathbf{I}$, wariancja we wszystkich klasach i wewnątrz nich jest identyczna $\boldsymbol{\Psi}_c = \sigma^2 \mathbf{I}$. W tabeli 1 przedstawiono w sposób syntetyczny, na jakie parametry nałożono ograniczenia.

Tabela 1.

Ograniczenia macierzy ładunków i wariancji

Ograniczenia na $\mathbf{\Lambda}_c$	Ograniczenia na $\mathbf{\Psi}_c$	
	Między klasami	**Wewnątrz klas**
U	U	U
C	U	U
U	C	U
C	C	U
U	U	C
C	U	C
U	C	C
C	C	C

C – ograniczenie, U – brak ograniczenia.
Źródło: opracowanie własne.

3. ESTYMACJA PARAMETRÓW MODELI

W wypadku mieszaniny rozkładów, z jaką mamy tutaj do czynienia, do estymacji parametrów wykorzystuje się najczęściej algorytm **EM.** Jest to iteracyjna procedura, w której wyróżnia się dwa zasadnicze kroki [11]. W pierwszym oblicza się warunkową wartość oczekiwaną logarytmu funkcji wiarygodności dla kompletnego zbioru danych:

$$Q(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = \mathbb{E}\left[\log L_c\left(\mathbf{\Theta}|\mathbf{Y},\mathbf{B},\mathbf{Z}\right)\Big|\mathbf{Y},\mathbf{\Theta}^{(t)}\right].$$

Zbiór ten nazywamy kompletnym, po wprowadzeniu dodatkowej zmiennej z_{ic}, która ma rozkład wielomianowy indeksowany parametrem $p(\boldsymbol{\pi}) = (p(\pi_c), \ldots, p(\pi_c))$. Przyjmuje ona wartość 1, gdy obserwacja i pochodzi z klasy K_c, lub 0 w przeciwnym wypadku.

Dla modelu analizy czynnikowej odpowiednie funkcje Q mają postać (por. [7]):

$$Q_{p(\boldsymbol{\pi})}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\,p(\pi_c) \tag{1}$$

$$Q_{\boldsymbol{\mu}}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\left[\boldsymbol{\mu}_c^T\mathbf{\Psi}_c^{-1}\mathbf{y}_i - \frac{1}{2}\boldsymbol{\mu}_c^T\mathbf{\Psi}_c^{-1}\boldsymbol{\mu}_c - \boldsymbol{\mu}_c^T\mathbf{\Psi}_c^{-1}\mathbf{\Lambda}_c\mathbf{v}_{ic}^{(t)})\right] \tag{2}$$

$$Q_{\mathbf{\Lambda}}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\mathbf{\Psi}_c^{-1}\mathbf{\Lambda}_c\mathbf{v}_{ic}^{(t)} - \frac{1}{2}tr(\mathbf{\Lambda}_c^T\mathbf{\Psi}_c^{-1}\mathbf{\Lambda}_c\mathbf{W}_{ic}^{(t)})\right] \tag{3}$$

$$Q_{\boldsymbol{\Psi}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\Big[-\log|\boldsymbol{\Psi}_c^{-1}| + (\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c) - \\ -2(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)} + tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})\Big] \tag{4}$$

gdzie:

$$\tau_{ic}^{(t)} = \frac{p^{(t)}(\pi_c)|\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}M_{ic}^{(t)}\right]}{\sum_c p^{(t)}(\pi_c)|\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}M_{ic}^{(t)}\right]},$$
$$M_{ic}^{(t)} = \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t)}\right)^T(\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)})^{-1}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t)}\right),$$

natomiast $[\mathbf{v}_{ic}^{(t)}]_{qx1} = \mathbb{E}(\mathbf{b}_i|\mathbf{y}_i, z_{ic})$ oraz $[\mathbf{W}_{ic}^{(t)}]_{qxq} = \mathbb{E}(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{y}_i, z_{ic})$.

W drugim kroku maksymalizuje się funkcję Q ze względu na nieznane parametry $\boldsymbol{\Theta}$,

$$\boldsymbol{\Theta}^{(t+1)} = \arg\max_{\boldsymbol{\Theta}} Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}\right), \tag{5}$$

przyjmując, że $\boldsymbol{\Theta}^{(t)}$ są znane, gdyż zostały oszacowane w iteracji (t).

Rozwiązanie (5) będzie różniło się miedzy modelami. Z tego też względu każdy z nich zostanie rozpatrzony osobno. Ponieważ nie wprowadza się żadnych ograniczeń ze względu na prawdopodobieństwa przynależności do klas oraz wektor wartości średnich, więc proces ich szacowania sprowadza się do jednego, ogólnego przypadku.

Wprowadzając mnożnik Lagrange'a do funkcji (1) nietrudno pokazać, że

$$p(\pi_c)^{(t+1)} = \frac{\sum_{i=1}^{n}\tau_{ic}^{(t)}}{n}. \tag{6}$$

W wypadku wartości przeciętnych należy zróżniczkować funkcję (2). Ponieważ

$$\frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\mathbf{y}_i)}{\boldsymbol{\mu}_c} = \boldsymbol{\Psi}_c^{-1}\mathbf{y}_i, \quad \frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c)}{\partial\boldsymbol{\mu}_c} = 2\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c, \quad \frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)})}{\boldsymbol{\mu}_c} = \boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)},$$

i z założenia, że $\partial Q_{\boldsymbol{\mu}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)})/\partial\boldsymbol{\mu}_c = 0$, więc estymator wartości średnich w iteracji $(t+1)$ ma postać:

$$\boldsymbol{\mu}_c^{(t+1)} = \sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\Lambda}_c^{(t)}\mathbf{v}_{ic}^{(t)}\right)\left(np(\pi_c)^{(t+1)}\right)^{-1}. \tag{7}$$

Oczywiście, nałożone ograniczenia na macierze ładunków czynnikowych i wariancji specyficznej należy uwzględnić we wzorze (6) i (7).

Model UUU:

w modelu tym nie przyjmuje się żadnych ograniczeń na macierze ładunków czynnikowych i wariancji specyficznej. Odpowiednie składowe pochodnej funkcji (3) mają postać:

$$\frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)}}{\partial\boldsymbol{\Lambda}_c} = \boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T}, \quad \frac{\partial tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})}{\partial\boldsymbol{\Lambda}_c} = 2\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)}. \quad (8)$$

Rozwiązując równanie $\partial Q_{\boldsymbol{\Lambda}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)})/\partial\boldsymbol{\Lambda}_c = 0$, otrzymuje się w iteracji $(t+1)$:

$$\boldsymbol{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\mathbf{v}_{ic}^{(t)T}\left(\sum_{i=1}^{n}\tau_{ic}^{(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}$$

Różniczkując z kolei składowe funkcji (4) ze względu na odwrotną macierz wariancji specyficznej otrzymano:

$$\frac{\partial\log|\boldsymbol{\Psi}_c^{-1}|}{\partial\boldsymbol{\Psi}_c^{-1}} = \boldsymbol{\Psi}_c, \quad \frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c)}{\partial\boldsymbol{\Psi}_c^{-1}} = (\mathbf{y}_i - \boldsymbol{\mu}_c)(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T),$$
$$\frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)}}{\partial\boldsymbol{\Psi}_c^{-1}} = (\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^T, \quad \frac{\partial tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})}{\partial\boldsymbol{\Psi}_c^{-1}} = \boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^T.$$

Po rozwiązaniu odpowiedniego równania estymator największej wiarygodności w iteracji $(t+1)$ można zapisać w postaci:

$$\boldsymbol{\Psi}_c^{(t+1)} = \frac{1}{np(\pi_c)^{(t+1)}}diag\left[\sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^{(t+1)T}\right)\right]$$

Model CUU:

nakłada się graniczenie na macierz ładunków czynnikowych, czyli $\boldsymbol{\Lambda}_c = \boldsymbol{\Lambda}$. Wykorzystując funkcję (3) oraz obliczone wcześnie pochodne (8) otrzymuje się równanie, które jest podstawą do oszacowania ładunków czynnikowych:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\boldsymbol{\Psi}_c^{-1}\mathbf{P}_{ic}^{(t)} = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}\mathrm{W}_{ic}^{(t)},$$
$$\mathbf{P}_{ic}^{(t)} = (\mathbf{y}_i - \boldsymbol{\mu}_c)\,\mathbf{v}_{ic}^{(t)T}.$$

Niech wiersz j macierzy $\mathbf{P}_{ic}^{(t)}$ zostanie oznaczony w postaci $\mathbf{P}_{ic}^{(t)}[j,]$, wtedy powyższe równanie, dla wiersza j, można zapisać:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2}\mathbf{P}_{ic}^{(t)}[j,] = \boldsymbol{\Lambda}[j,]\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2}\mathbf{W}_{ic}^{(t)},$$

i tym samym wyznaczyć w kolejnej iteracji wartości:

$$\mathbf{\Lambda}^{(t+1)}[j,] = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2(t)}\mathbf{P}_{ic}^{(t)}[j,]\left(\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

W wypadku estymatora wariancji specyficznej można bezpośrednio, po niewielkiej korekcie ze względu na macierz ładunków, wykorzystać wynik otrzymany dla modelu *UUU*:

$$\mathbf{\Psi}_c^{(t+1)} = \frac{1}{np(\pi_c)^{(t+1)}}diag\left[\sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}^{(t+1)T}\right)\right]$$

Model CUC:

konsekwencją nałożonych ograniczeń jest $\mathbf{\Lambda}_c = \mathbf{\Lambda}$ oraz $\mathbf{\Psi}_c = \sigma_c^2\mathbf{I}$. Wykorzystując (3) i (8) otrzymano równanie:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2}(\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T} = \mathbf{\Lambda}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2}\mathbf{W}_{ic}^{(t)},$$

z którego wyznaczono:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2(t)}(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)})\mathbf{v}_{ic}^{(t)T}\left(\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

W wypadku wariancji przekształcono funkcję (4) do postaci:

$$\begin{aligned} Q_{\mathbf{\Psi}}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(k)}\Big[-p\log\sigma_c^{-2} + \sigma_c^{-2}(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)(\mathbf{y}_i - \boldsymbol{\mu}_c) - \\ -2\sigma_c^{-2}(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\mathbf{\Lambda}_c\mathbf{v}_{ic}^{(t)} + \sigma_c^{-2}tr(\mathbf{\Lambda}_c^T\mathbf{\Lambda}_c\mathbf{W}_{ic}^{(t)})\Big]. \end{aligned} \tag{9}$$

Ponieważ $\partial Q_{\mathbf{\Psi}_c}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)})/\partial\sigma_c^{-2} = 0$, więc w iteracji $(t+1)$ estymator wariancji ma postać:

$$\sigma_c^{2(t+1)} = \frac{1}{p\cdot n\cdot p(\pi_c)^{(t+1)}}\sum_{i=1}^{n}\tau_{ic}^{(t)}\left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T})\left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}^{(t+1)}\mathbf{v}_{ic}^{(t)}\right) + tr(\mathbf{\Lambda}^{(t+1)T}\mathbf{\Lambda}^{(t+1)}\mathbf{W}_{ic}^{(t)})\right].$$

Model CCU:

ograniczenia dotyczą obu macierzy, tzn. $\mathbf{\Lambda}_c = \mathbf{\Lambda}$ oraz $\mathbf{\Psi}_c = \mathbf{\Psi}$. Bazując na wynikach otrzymanych dla modelu bez ograniczeń (*UUU*), macierz ładunków czynnikowych ma postać:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1},$$

podczas gdy macierz wariancjiprzedstawia się następująco:

$$\mathbf{\Psi}^{(t+1)} = \frac{1}{n} diag\left[\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}^{(t+1)T}\right)\right].$$

Model UCU:

macierz wariancji specyficznej jest taka sama w każdej klasie, a więc $\mathbf{\Psi}_c = \mathbf{\Psi}$, natomiast na macierz ładunków czynnikowych nie nakłada się żadnych ograniczeń. Z tego względu postać estymatora ładunków czynnikowych będzie taka sama jak w modelu *UUU*:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

Z kolei estymator wariancji specyficznej:

$$\mathbf{\Psi}^{(t+1)} = \frac{1}{n} diag\left[\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}_c^{(t+1)T}\right)\right].$$

Model UUC:

ograniczenie w tym modelu dotyczy tylko macierzy wariancji specyficznej: $\mathbf{\Psi}_c = \sigma_c^2\mathbf{I}$. Ze względu na podobieństwo do szacowania parametrów modelu *CUC*, wystarczy drobna modyfikacja, by zaadoptować tamto rozwiązanie i otrzymać:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} (\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1},$$

$$\sigma_c^{2(t+1)} = \frac{1}{p \cdot np(\pi_c)^{(t+1)}} \sum_{i=1}^{n} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T})\left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}_c^{(t+1)}\mathbf{v}_{ic}^{(t)}\right) + tr(\mathbf{\Lambda}_c^{(t+1)T}\mathbf{\Lambda}_c^{(t+1)}\mathbf{W}_{ic}^{(t)})\right].$$

Model UCC:

jest to model podobny do wcześniejszego (*UUC*) z tą różnicą, że wariancja specyficzna jest taka sama wewnątrz jak i między klasami: $\mathbf{\Psi}_c = \sigma^2\mathbf{I}$. Wzór na oszacowanie ładunków czynnikowych będzie więc identyczny jak w modelu *UUU*:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} (\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)} \right)^{-1},$$

natomiast estymator wariancji, po koniecznej modyfikacji ma postać:

$$\sigma^{2(t+1)} = \frac{1}{p \cdot n} \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T}) \left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}_c^{(t+1)} \mathbf{v}_{ic}^{(t)} \right) + tr(\mathbf{\Lambda}_c^{(t+1)T} \mathbf{\Lambda}_c^{(t+1)} \mathbf{W}_{ic}^{(t)}) \right].$$

Model CCC:

ostatni z rozważanych modeli jest najmniej rozbudowany, w sensie liczby parametrów, gdyż nałożone zostały wszystkie możliwe ograniczenia: $\mathbf{\Lambda}_c = \mathbf{\Lambda}, \mathbf{\Psi}_c = \sigma^2\mathbf{I}$. Wykorzystującwyniki otrzymane dla poprzednich modeli otrzymano:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)} \right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)} \right)^{-1},$$

$$\sigma^{2(t+1)} = \frac{1}{p \cdot n} \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T}) \left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}^{(t+1)} \mathbf{v}_{ic}^{(t)} \right) + tr(\mathbf{\Lambda}^{(t+1)T} \mathbf{\Lambda}^{(t+1)} \mathbf{W}_{ic}^{(t)}) \right].$$

Iteracyjny charakter procedury estymacji powoduje, że w każdym kolejnym kroku t zwiększa się o 1. Ponieważ takiego postępowania nie można prowadzić w nieskończoność, dlatego należy zaproponować kryterium stopu, kiedy to dodatkowy krok estymacji nie przynosi istotnych korzyści. Wśród propozycji można wskazać podejścia bazujące na różnicy w wartościach estymowanych parametrów lub wartościach samej funkcji. Jeśli owe różnice będą mniejsze od założonego progu, wtedy procedurę należy przerwać (por. [10]).

4. WYBÓR MODELU

Podstawowe pytanie dotyczące kwestii wyboru modelu sprowadza się do rozstrzygnięcia, czy model w którym nałożono pewne ograniczenia na parametry nie będzie znacznie odbiegał – w sensie dopasowania do danych – od modelu, w którym tych ograniczeń nie ma, lub jest ich mniej. W przypadku tzw. modeli hierarchicznych, można wykorzystać test oparty na ilorazie funkcji wiarygodności. Gdy nie są one hierarchiczne, to wobec niespełnienia warunków regularności, takie podejście jest niewłaściwe, gdyż prowadzi do nieznajomości rozkładu statystyki testowej (por. [12]).

Można próbować aproksymować ten rozkład wykorzystując podejście bootstrapowe [14], lecz czasochłonność obliczeń sprawia, że poszukuje się innych rozwiązań.

Wśród nich znajdują się kryteria informacyjne, które można sklasyfikować w następujące grupy: kryteria ze szkoły Akaike, kryteria bayesowskie i kryteria klasyfikacyjne. W zależności od wykorzystywanej klasy weryfikowanych modeli, ich wiarygodność się zmienia. Przykładowo, w wielomianowym modelu logitowym opartym na mieszkankach rozkładów kryterium informacyjne Akaike *AIC* jest bardziej wiarygodne od kryterium bayesowskiego *BIC* [2], natomiast dla mieszkanek rozkładów normalnych jest odwrotnie [12].

W pracy [8] przeprowadzonotakie badania symulacyjne dotyczące analizy czynnikowej z dwoma zmiennymi ukrytymi. Okazało się, że największą trafnością co do wyboru właściwego modelu odznaczały się dwa kryteria: *ICL BIC* oraz *BIC*. Trochę gorzej wypadło kryterium zgodne *AIC* i kryterium *AIC*. Pozostałe z rozważanych kryteriów (*NEC, CLC*) wykazały się dość mała trafnością. Choć badania te dotyczyły modelu *UCU*, to jednak warto skorzystać z tych wyników i skupić się tylko na czterech kryteriach:

- kryterium Akaike [1]:

$$AIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + 2r,$$

- zgodne kryterium Akaike [4]:

$$CAIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + r\log n + r,$$

- kryterium bayesowskie [9]:

$$BIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + r\log n,$$

- łączne kryterium bayesowskie i klasyfikacyjne [12]:

$$ICLBIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + 2EN\left(\hat{\boldsymbol{\tau}}\right) + r\log n,$$

$$EN\left(\hat{\boldsymbol{\tau}}\right) = -\sum_{i=1}^{n}\sum_{c=1}^{C}\hat{\boldsymbol{\tau}}_{ic}\log\hat{\boldsymbol{\tau}}_{ic},$$

gdzie r jest liczbą szacowanych parametrów, natomiast $L\left(\hat{\boldsymbol{\Theta}}\right)$ funkcją wiarygodności. Spośród konkurencyjnych modeli wybiera się ten, dla którego obliczone kryterium informacyjne ma najmniejszą wartość. Należy odnotować, że opisywane modele mogą mieć różne warianty zależnie od liczby klas C oraz liczby czynników q. Nie zmienia to jednak procedury postępowania przy wyborze.

5. PRZYKŁAD EMPIRYCZNY – DANE O SATYSFAKCJI

Aby zilustrować użyteczność omawianych modeli, wykorzystano rzeczywiste dane – zaczerpnięte z pracy [6] – odnoszące się do opinii klientów wyrażonej na temat

1. Jakość produktów
2. Aktywność e-commerce
3. Wsparcie techniczne
4. Rozwiązywanie skarg
5. Reklama
6. Linia produktowa
7. Wizerunek
8. Konkurencyjność cen
9. Gwarancje
10. Nowe produkty
11. Zamówienia i faktury
12. Elastyczność cen
13. Szybkość dostawy

współpracy z pewną firmą działającą w branży przemysłowej. Ocenie poddano 13 zmiennych, wykorzystując jedenastostopniową skalę.

Do oceny porównawczej, rozważanych w części teoretycznej modeli, włączono dodatkowo wariant model *CCU*, który powstaje poprzez nałożenie ograniczeń na wartości średnie. W konsekwencji średnie nie będą różniły się między klasami. Tak sformułowany model można uznać za klasyczną postać modelu analizy czynnikowej. Jest to więc dobry punkt odniesienia, gdyż taki model jest zaimplementowany w każdym pakiecie statystycznym (np. SPSS, SAS, STATISTICA, R), co z kolei przekłada się na częste jego wykorzystanie. Wydaje się zasadne, z punktu widzenia praktycznych zastosowań, porównanie takiego modelu z modelami rozważanymi w pracy.

Procedura estymacji każdego modelu wymaga określenia *ex ante* liczby klas oraz liczby ładunków czynnikowych. Ponieważ liczba estymowanych parametrów jest ściśle powiązana z liczbą klas, a rozmiar próby jest niezbyt duży (200 obserwacji), dlatego przyjęto maksymalną liczbę klas na poziomie dwóch. W wypadku ładunków czynnikowych rozważano modele z 2, 3 i 4 czynnikami. Odpowiednie programy szacujące parametry modeli napisano w środowisku **R** [15].

W tabeli 2 podano wartości logarytmów funkcji wiarygodności oraz kryteriów informacyjnych. Dzięki temu można dokonać formalnego rozstrzygnięcia co do postaci modelu. Nie zamieszczono dwóch modeli: *CUC* i *UCC*. Wynika to z prostej obserwacji, że model bardziej rozbudowany*UUC* jest jednym z najgorszych w sensie wartości kryteriów informacyjnych.

Pierwsze spostrzeżenie jakie nasuwa się w odniesieniu do tabeli 2 to bardzo wysokie wartości kryteriów informacyjnych modelu klasycznego. Również i logarytm funkcji wiarygodności znacznie odstaje od pozostałych. To utwierdza w przekonaniu, że model ten jest nieodpowiedni, dlatego wyboru należy poszukiwać wśród modeli bardziej złożonych.

Przyjęcie ograniczenia na wartości wariancji specyficznej wewnątrz klas – skutkującetym, że każda zmienna ma taką samą wariancję (choć może się ona różnić między klasami) – powoduje gorsze dopasowanie takich modeli do danych w stosunku do pozostałych modeli. Szczególnie może zaskakiwać model *UUC*, w którym wartości średnie i ładunki czynnikowe szacowane są dla każdej z dwóch klas; po-

Tabela 2.

Wartości funkcji wiarygodności i kryteriów informacyjnych w zależności od modelu

Klasyczny	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-4690	-4605	-4598
AIC	9417	9271	9267
BIC	9581	9538	9577
CAIC	9619	9600	9649
ICL BIC	-	-	-

UCU	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3156	-3045	-2891
AIC	6401	6202	5915
BIC	6788	6683	6482
CAIC	6878	6795	6614
ICL BIC	6807	6689	6485

CCC	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3711	-3606	-3489
AIC	7475	7275	7051
BIC	7703	7551	7369
CAIC	7756	7615	7443
ICLBIC	7717	7559	7374

UUC	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3670	-3540	-3418
AIC	7418	7182	6958
BIC	7758	7616	7478
CAIC	7837	7717	7599
ICLBIC	7765	7620	7482

CCU	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3207	-3084	-2957
AIC	6478	6244	6000
BIC	6758	6571	6370
CAIC	6823	6647	6456
ICL BIC	6773	6585	6374

CUU	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3204	-3069	-2955
AIC	6487	6226	6009
BIC	6822	6609	6435
CAIC	6900	6698	6534
ICL BIC	6883	6611	6438

UUU	Liczba ładunków czynnikowych		
	2	3	4
log*L*	-3126	-2974	-2865
AIC	6355	6073	5875
BIC	6798	6611	6498
CAIC	6901	6736	6643
ICL BIC	6814	6617	6499

Źródło: Opracowanie własne.

dobnie wariancja. Pomimo swojej elastyczności wartości kryteriów informacyjnych jak i logarytmy funkcji wiarygodności znacznie odstają od pozostałych modeli.
Najbardziej ogólny model *UUU*, na który nie nałożono żadnych ograniczeń, cechuje się najniższą wartością kryterium *AIC*. Biorąc pod uwagę pozostałe kryteria – pamiętając jednocześnie o tendencji kryterium *AIC* do wyboru modeli zbytnio rozbudowanych oraz wynikach badań symulacyjnych (wspomnianych w pkt. 4) – modelem najbardziej preferowanym byłby model *CCU* o czterech czynnikach wspólnych.
Wartości oszacowanych ładunków czynnikowych zamieszczono w tabeli 3. Kierując się rekomendacjami w zakresie ich praktycznej przydatności (por. [6]) przyjęto, że minimalna wartość ładunku dla danej zmiennej powinna być nie mniejsza niż 0,4. Dopiero wtedy taka zmienna może współtworzyć dany czynnik.

Tabela 3.

Wartości ładunków czynnikowych

	Czynnik			
Zmienna	**1**	**2**	**3**	**4**
Linia produktowa	0,96	0,17	-0,08	0,10
Szybkość dostawy	0,83	0,13	0,52	0,10
Rozwiązywanie skarg	0,74	0,12	0,46	0,10
Zamówienia i faktury	0,59	0,16	0,47	0,14
Elastyczność cen	0,32	-0,03	0,94	0,06
Reklama	0,24	0,58	0,14	-0,02
Wizerunek	0,20	0,97	0,06	0,10
Aktywność e-commerce	0,16	0,77	0,05	0,04
Gwarancje	0,09	0,10	-0,02	0,99
Nowe produkty	0,06	0,03	0,17	0,01
Wsparcie techniczne	0,05	0,03	0,00	0,84
Konkurencyjność cen	-0,01	-0,01	-0,03	0,01
Jakość produktów	-0,04	0,19	0,09	0,00

Źródło: Opracowanie własne.

Interpretacja czynnika 2 i 4 nie nastręcza większych problemów, gdyż można byłoby te wymiary opisać odpowiednio jako: marketing i serwis. Problematyczne staje się jednak nazwanie czynnika 1 i 3, gdyż trzy zmienne mają wysokie wartości ładunków na każdym z tych czynników. Zmienne które wyraźnie różnicują czynniki to linia produktowa i elastyczność cen; tutaj też ładunki są najwyższe. Być może

pewnym rozwiązaniem byłoby utożsamienie czynników z tymi zmiennymi. Jest to jeden ze sposobów interpretacji czynników (por. [6]).

6. PODSUMOWANIE

Przedstawione w pracy modele analizy czynnikowej wraz z procedurą estymacji parametrów pozwalają, przynajmniej w założeniu, rozszerzyć instrumentarium badawcze o dodatkowe modele. Modele te cechuje większa elastyczność, co może być szczególnie przydatne w analizie niejednorodnych zbiorów danych.
Jak można było zaobserwować na przykładzie danych o satysfakcji, klasyczny model analizy czynnikowej nie był odpowiedni, jeśli decyzję oprzeć na kryteriach informacyjnych. Zauważono również, znaczne obniżenie wartości informacyjnej modelu, jeśli przyjęto (w ramach rozważanych 8 modeli) równą wariancję specyficzną dla każdej zmiennej. Sytuacja nie uległa poprawie, jeśli wariancje różniły się między klasami.
Wydaje się, że egzemplifikacja proponowanych modeli z wykorzystaniem większych zbiorów danych, cechujących się większą niejednorodnością powinna stanowić podstawę dalszych badań.

Politechnika Wrocławska

LITERATURA

[1] Akaike H. (1973), *Information theory and anextension of the maximum likelihoodprinciple,* [w:] Petrov B.N., Csaki F. (Eds.), Second international symposium on information theory (pp.). Budapest: AcademiaiKiado, s. 267-281.
[2] Andrews L., Currim I.S. (2003), *A Comparison of segment retention criteria for finite mixture logit models*, Journal of Marketing Research, **40**(2), s. 235-243.
[3] Banfield, J.D.,Raftery, A.E. (1993). *Model-based Gaussian and non-Gaussian clustering,* Biometrics, **49**, s. 803-821.
[4] Bozdogan, H. (1987), *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*. Psychometrika, **52**, s. 345-370.
[5] Celeux G., Govaert G. (1995), *Gaussian Parsimonious Clustering Models*, Pattern Recognition, **28**(5), s. 781-793.
[6] Hair J.F., Black W.C., Babin B.J., Anderson R.E., (2010), *Multivariate Data Analysis,* Prentice Hall, New York.
[7] Kapłon R. (2004), *Estymacja parametrów modelu czynnikowego wykorzystującego klasy ukryte*, [w:] Jajuga K., Walesiak M. „Klasyfikacja i analiza danych – teoria i zastosowania". Taksonomia nr 11, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, s. 204-211.
[8] Kapłon R. (2007), *O liczbie klas w modelu analizy czynnikowej z dwoma zmiennymi ukrytymi,* [w:] Jajuga K., Walesiak M. „Klasyfikacja i analiza danych – teoria i zastosowania". Taksonomia nr 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, s. 253-260.
[9] Kass R. E., Raftery A. E. (1995), *Bayes factors,*Journal of the American Statistical Association, **90**, s. 773-795.
[10] Luenberger D.G., Ye Y. (2008), *Linear and Nonlinear Programming*, Springer.
[11] McLachlan G.J., Krishnan T. (1997), *The EM Algorithm and Extensions*. New York: Wiley.

[12] McLachlan G.J., Peel. D. (2000), *Finite Mixture Models*, New York: Wiley.
[13] McLachlan, G.J, Basford K. (1988),*Mixturemodels*. New York: Marcel Dekker.
[14] McLachlan, G.J. (1987), *On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture,Journal of the Royal Statistical Society Series C (Applied Statistics),* **36**, s. 318-324.
[15] R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[16] Walesiak M. (1996), *Metody analizy danych marketingowych,* Wydawnictwo Naukowe PWN, Warszawa.

MODELE ANALIZY CZYNNIKOWEJ Z DWOMA ZMIENNYMI UKRYTYMI

Streszczenie

Celem analizy czynnikowej jest redukcja zmiennych poprzez ich zastąpienie mniejszą liczbą czynników, które traktowane są jako konstrukty lub zmienne ukryte. Niestety, jeśli mamy do czynienia z niejednorodnym zbiorem danych, estymacja jednego zbioru wartości średnich, ładunków czynnikowych czy wariancji specyficznych może prowadzić do błędnych wniosków. Jednym ze sposobów radzenia sobie z tą niejednorodnością jest wprowadzenie dodatkowej zmiennej ukrytej do modelu analizy czynnikowej. W konsekwencji zakłada się, że obserwacje pochodzą z dwóch lub więcej podpopulacji, których struktura jest nieznana.
W zależności od ograniczeń nałożonych na macierze ładunków czynnikowych i wariancji specyficznej, można otrzymać różne warianty modelu. W pracy przedstawiono 8 modeli analizy czynnikowej, wraz z propozycją procedury estymacji parametrów algorytmem **EM**. Zwrócono również uwagę na problem w wykorzystaniu testów statystycznych, opartych na ilorazie wiarygodności, wskazując na kryteria informacyjne jako alternatywne podejście. Proponowane podejście zilustrowano przykładem, wykorzystując w tym celu rzeczywiste dane dotyczące satysfakcji.

FACTOR ANALYSIS MODELS WITH TWO LATENT VARIABLES

Abstrakt

The goal of factor analysis is to reduce the redundancy among variables by using smaller number of factors that are treated as constructs or latent variables. Unfortunately, if we face with data heterogeneity, the estimates of a single set of means, factor loadings and specific variances may be misleading.One way of accounting for unobserved heterogeneity is to include another latent variable in a factor analysis model. As a consequence, the observations in a samples are assumed to arise from two or more subpopulations that are mixed in unknown proportions.
Since putting some restrictions on parameters such as factor loadings and specific variancesone can get more parsimonious models. Therefore, the purpose of this paper is to present the eight factor analysis models. Methods of optimization to derive the maximum likelihood estimates based on EM algorithm as well as model selection procedure are considered. Proposed approach is illustrated by using a set of data referring to preferences.