

Mirosława Sztemberg-Lewandowska

Uniwersytet Ekonomiczny we Wrocławiu

ANALIZA CZYNNIKOWA Z WYKORZYSTANIEM ŚRODOWISKA R

1. Wstęp

Dostępne na rynku pakiety statystyczne nie udostępniają takich możliwości, jakie niesie ze sobą środowisko R. W artykule przedstawiono wybrane pakiety i funkcje programu wykorzystywane w analizie czynnikowej. Scharakteryzowano także kolejne etapy analizy czynnikowej w języku R. Celem artykułu jest przedstawienie funkcji `factanal` oraz porównanie jej z funkcjami `factor.pa` i `princomp`.

Tabela 1. Etapy analizy czynnikowej oraz pakiety i funkcje programu R

Lp.	Etapy analizy czynnikowej	Wybrane pakiety oraz funkcje programu R
1	Selekcja zmiennych	Funkcje <code>KMO</code> , <code>MSA</code>
2	Ustalenie liczby czynników wspólnych	Pakiet <code>psych</code> (funkcje: <code>VSS.scree</code> , <code>VSS</code>) Pakiet <code>nFactors</code> (funkcje: <code>plotuScree</code> , <code>nScree</code>)
3	Wyodrębnienie czynników wspólnych	Pakiet <code>stats</code> (funkcje: <code>factanal</code> , <code>prcomp</code> , <code>princomp</code>) Pakiet <code>RScalAPACK</code> (funkcje: <code>sla.factanal</code> , <code>sla.prcomp</code> , <code>sla.princomp</code>) Pakiet <code>psych</code> (funkcja <code>factor.pa</code>) Pakiet <code>FactoMineR</code> (funkcja <code>AFDM</code>)
4	Rotacja czynników	Pakiet <code>RScalAPACK</code> (funkcje: <code>sla.varimax</code> , <code>sla.promax</code>) Pakiet <code>GRARotation</code> (funkcje: <code>GPForth</code> , <code>GPFoblg</code>)
5	Identyfikacja i interpretacja czynników	Pakiet <code>psych</code> (funkcja: <code>fa.graph</code>)
6	Ocena dopasowania modelu czynnikowego	Pakiet <code>psych</code> (funkcje: <code>factor.fit</code> , <code>factor.residuals</code>)

Źródło: opracowanie własne na podstawie dokumentacji programu R.

Procedura analizy czynnikowej polega na dokonywaniu szeregu następujących po sobie operacji obliczeniowych, w wyniku których otrzymuje się strukturę czyn-

nikową badanych zmiennych. W trakcie wyodrębniania czynników badacz zmuszony jest do podejmowania szeregu arbitralnych decyzji, które w sposób oczywisty wpływają na końcowe rezultaty. Decyzje te dotyczą (por. np. [Zakrzewska 1994, s. 7]):

- 1) selekcji zmiennych,
- 2) liczby czynników wspólnych,
- 3) metody wyodrębniania czynników wspólnych,
- 4) rotacji czynników,
- 5) identyfikacji i interpretacji czynników.

W środowisku R dostępnych jest wiele pakietów i funkcji, które są pomocne w przeprowadzaniu analizy czynnikowej (tab. 1).

Kolejne kroki analizy czynnikowej w środowisku R zobrazowano na przykładzie empirycznym dotyczącym rynku kaw. Przykład szczegółowo zaprezentowano w [Szttemberg 1999, s. 131-143]), w niniejszym artykule przytoczono tylko dane liczbowe, które wczytano z pliku o rozszerzeniu csv:

```
z <- read.csv2("K:/kawy_9.csv", header = TRUE,
strip.white = TRUE).
```

Następnie obliczono macierz korelacji między zmiennymi za pomocą polecenia `round(cor(z), digits=3)`, korelacje zaokrąglono do trzeciego miejsca po przecinku, poleceniem `options(OutDec=",")` zmieniono symbol oddzielający część całkowitą liczby z kropki na przecinek:

	x2	x3	x4	x5	x6	x7	x8	x9	x10
x2	1,000	-0,725	-0,376	-0,398	-0,563	-0,050	-0,609	-0,349	-0,312
x3	-0,725	1,000	0,236	0,196	0,463	0,300	0,637	0,253	0,399
x4	-0,376	0,236	1,000	0,832	0,519	0,162	0,389	0,865	0,327
x5	-0,398	0,196	0,832	1,000	0,704	0,336	0,500	0,868	0,435
x6	-0,563	0,463	0,519	0,704	1,000	0,497	0,686	0,711	0,691
x7	-0,050	0,300	0,162	0,336	0,497	1,000	0,437	0,369	0,595
x8	-0,609	0,637	0,389	0,500	0,686	0,437	1,000	0,502	0,540
x9	-0,349	0,253	0,865	0,868	0,711	0,369	0,502	1,000	0,456
x10	-0,312	0,399	0,327	0,435	0,691	0,595	0,540	0,456	1,000

2. Selekcja zmiennych

Przed przystąpieniem do zastosowania procedury analizy czynnikowej należy sprawdzić, czy wykorzystywane w badaniach zmienne pozostają w pewnych określonych relacjach. Jeżeli korelacje między zmiennymi są niskie, jest mało prawdopodobne, że zmienne te utworzą silne i łatwe w interpretacji czynniki wspólne. Jedną z formalnych technik, która ma na celu pomóc badaczowi w podjęciu decyzji, czy wybór modelu analizy czynnikowej jako metody analizy zebranych przez niego danych był wyborem trafnym, jest wskaźnik Kaisera-Meyera-Olkina (*KMO*) (por. np. [Zakrzewska 1994, s. 55; Gatnar 1998, s. 8-9]):

$$KMO = \frac{\sum_p \sum_{h \neq p} r_{ph}^2}{\sum_p \sum_{h \neq p} r_{ph}^2 + \sum_p \sum_{h \neq p} \hat{r}_{ph}^2}, \quad (1)$$

gdzie: r_{ij} – współczynnik korelacji pomiędzy p -tą i h -tą zmienną,

\hat{r}_{ph} – współczynnik korelacji cząstkowej (zob. np. [Zeliaś 2000, s. 84-88])
między nimi.

W języku R współczynnik korelacji cząstkowej wymaga funkcji `cor2pcor` dostępnej w bibliotece `corpcor`. Natomiast wskaźnik *KMO* można obliczyć za pomocą następującej funkcji:

```
> KMO <- function(z, cor){
+   if (cor==TRUE)
+     { l <- sum(z^2)-dim(z)[2]
+       m <- sum(z^2)+sum((cor2pcor(z))^2)-2*dim(z)[2]
+       KMO <- 1/m
+     }
+   else
+     { l <- sum((cor(z))^2)-dim(z)[2]
+       m <- sum((cor(z))^2)+sum((cor2pcor(cor(z)))^2)
+         2*dim(z)[2]
+       KMO <- 1/m
+     }
+   return(KMO)
+ }
```

W podobny sposób oblicza się miarę adekwatności doboru każdej indywidualnej zmiennej (por. np. [Zakrzewska 1994, s. 56]):

$$MSA_h = \frac{\sum_{p \neq j} r_{ph}^2}{\sum_{p \neq h} r_{ph}^2 + \sum_{p \neq h} \hat{r}_{ph}^2}. \quad (2)$$

Na podstawie wielkości MSA_h można eliminować z badań pojedyncze zmienne przed analizą (por. np. [Psychologiczne... 1993, s. 85-86]). Małe wartości MSA_h sugerują, że korelacje zmiennej h nie mogą być tłumaczone przez inne zmienne i zmienna ta powinna być wyeliminowana z badań.

Natomiast wskaźnik *MSA* dla każdej zmiennej można obliczyć w języku R, korzystając z funkcji `MSA`:

```
> MSA <- function(z, cor){
+   if (cor==TRUE)
+     { MSA <- array(2,c(1,ncol(z)))
```

```

+       for(i in (1:ncol(z))) {
+         l <- sum((z[,i])^2)-1
+         m <- sum((z[,i])^2)+sum((cor2pcor(z)[,i])^2)-2
+         MSA[i] <- l/m
+       }
+     }
+   else{
+     MSA <- array(2,c(1,ncol(z)))
+     for(i in (1:ncol(z))) {
+       l <- sum((cor(z)[,i])^2)-1
+       m <- sum((cor(z)[,i])^2)+sum((cor2pcor(cor(z))[,i])^2 )-2
+       MSA[i] <- l/m
+     }
+   }
+ }
+ return(MSA)
+ }

```

W wyniku zastosowania funkcji $KMO(z, cor=FALSE)$ i $MSA(z, cor=FALSE)$ otrzymano wskaźniki:

KMO : 0,7746156

MSA dla każdej zmiennej:

[, 1]	[, 2]	[, 3]	[, 4]	[, 5]
0,6810593	0,674342	0,7051588	0,8257298	0,7970919
[, 6]	[, 7]	[, 8]	[, 9]	
0,6881226	0,929732	0,7916745	0,8282726	

3. Ustalenie liczby czynników wspólnych

Istnieje kilka zasad określania liczby czynników głównych, stosowanych w analizie czynnikowej (por. np. [Kim, Mueller 1978]). Najważniejsze z nich to:

- kryterium bazujące na wartościach własnych,
- kryterium osypiska (*scree plot*),
- kryterium wyjaśnionej wariancji,
- kryterium istotności testu statystycznego.

Kryterium osypiska bazuje na wykresie osypiska, na którym zaznaczone są wartości własne dla kolejnych czynników. Kryterium osypiska mówi, że należy zachować tyle czynników, ile tworzy „zбочe”, natomiast zignorować te, które tworzą „osypisko”.

Wykres osypiska w języku R można uzyskać, wykorzystując np. funkcję `plotScree` (tab. 2).

Tabela 2. Składnia funkcji plotuScree w środowisku R*

plotuScree(Eigenvalue, ylab="Eigenvalue", xlab="Component", main="Scree Plot")	
Eigenvalue	wektor wartości własnych
ylab	nazwa osi Y (ylab="Wartości własne")
xlab	nazwa osi X (xlab="Składowe")
main	tytuł wykresu (main="Wykres osypiska")

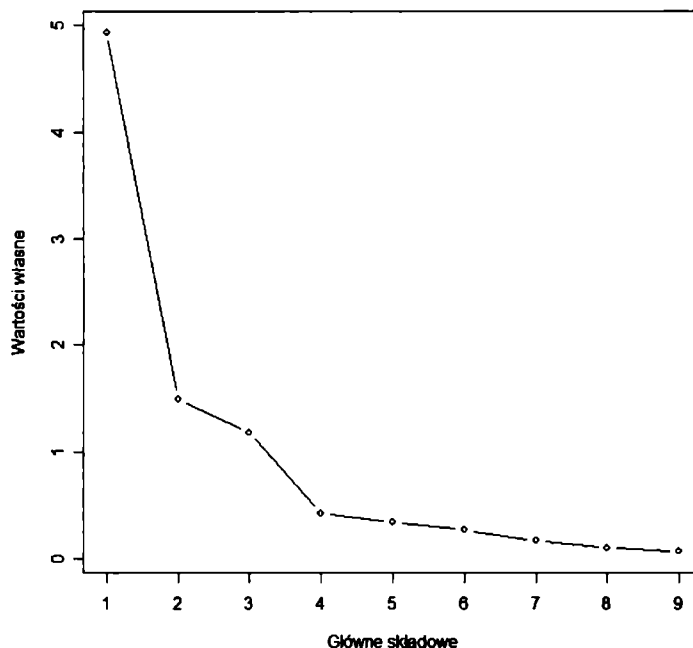
*Opis funkcji zawiera podstawowe argumenty.

Źródło: opracowanie własne na podstawie dokumentacji programu R.

W wyniku składni poleceń:

```
> library(nFactors)
> old <- par(lab=c(ncol(z),5,7))
> plotuScree(eigen(cor(z))$values, main = "", xlab =
"Główne składowe", ylab = "Wartości własne")
> par(old)
```

otrzymano wykres osypiska widoczny na rys. 1. Zjawisko osypiska występuje po trzecim czynniku.



Rys. 1. Wykres osypiska

Źródło: obliczenia własne z wykorzystaniem pakietu R.

Kryterium bazujące na wartościach własnych wymaga wyznaczenia wartości własnych macierzy korelacji i uwzględnienia tylu czynników, ile wartości własnych jest większych od jedynki. Wartości własne macierzy korelacji obliczono poprzez funkcję

```
eigen(cor(z)):
4,93374160 1,49073808 1,18258360 0,42697366 0,34590060
0,27075281 0,17251972 0,10630737 0,07048257
```

Kryterium wystarczającej proporcji wyjaśnionej wariancji mówi, że należy pozostawić tyle czynników, aby wyjaśniały założony procent wariancji, np. 80 lub 95%.

Wariancję wyjaśnioną przez kolejne składowe wyznaczono za pomocą funkcji `summary(princomp(z))`:

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	
Standard deviation	1,4852575	0,8670806	0,6234790	0,39508221	
Proportion of Variance	0,5759393	0,1962873	0,1014885	0,04075193	
Cumulative Proportion	0,5759393	0,7722266	0,8737151	0,91446702	
	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
	0,32676862	0,28801407	0,27170001	0,20344566	0,150570789
	0,02787751	0,02165712	0,01927314	0,01080613	0,005919086
	0,94234453	0,96400164	0,98327479	0,99408091	1,000000000

Wszystkie zastosowane kryteria określania liczby czynników dały trzy czynniki i tyle ich zostanie uwzględnionych w dalszej analizie.

Test chi-kwadrat jest obliczany w następnym kroku przy wywołaniu funkcji `factanal`.

4. Metody wyodrębniania czynników wspólnych

Funkcja `factanal` przeprowadza analizę czynnikową tylko metodą największej wiarygodności. W pakiecie `psych` dostępna jest jeszcze metoda osi głównych. Natomiast funkcja `princomp` wyznacza główne składowe (por. np. [Dalgaard 2002; Everitt, Hothorn 2006]).

Następnie dla przykładu liczbowego z trzema czynnikami zastosowano analizę czynnikową metodą największej wiarygodności. W celu lepszej interpretacji wyników przeprowadzono rotację *varimax*:

```
> attach(z)
> fa <- factanal(z, factors=3)
> fal <- factanal(formula(paste("~", paste(names(z),
collapse="+"))), factors=3, scores="Bartlett")$scores
> print(fa)
```

```
> print(fal)
> detach(z)
```

Tabela 3. Składnie funkcji metod wyodrębniania czynników wspólnych w środowisku R*

Metoda największej wiarygodności	
<code>factanal(x, factors, data=NULL, covmat=NULL, n.obs=NA, subset, start=NULL, scores=c("none", "regression", "Bartlett"), rotation="varimax")</code>	
<code>x</code>	macierz danych lub symboliczny opis modelu
<code>factors</code>	liczba wyodrębnionych czynników
<code>data</code>	tabela danych (gdy <code>x</code> jest symbolicznym opisem modelu)
<code>covmat</code>	macierz kowariancji (korelacji)
<code>n.obs</code>	liczba obserwacji (gdy podana została macierz kowariancji)
<code>subset</code>	wyrażenie wskazujące podzbiór obserwacji do analizy
<code>start</code>	NULL lub macierz początkowych zasobów zmienności specyficznej
<code>scores</code>	metoda wyodrębniania ładunków: "regression" – metoda Thompsona, "Bartlett" – ważona metoda najmniejszych kwadratów Bartletta
<code>rotation</code>	metoda rotacji czynników lub nazwa funkcji dokonującej rotacji
Metoda głównych składowych	
<code>princomp(x, cor=FALSE, scores=TRUE, covmat=NULL, subset=rep(TRUE, nrow(as.matrix(x))))</code>	
<code>x</code>	macierz lub tabela danych
<code>cor</code>	TRUE oznacza, że w obliczeniach wykorzystana jest macierz korelacji FALSE oznacza, że w obliczeniach wykorzystana jest macierz kowariancji
<code>scores</code>	TRUE oznacza, że dla każdej składowej dodatkowo obliczana jest: średnia, mediana, największa i najmniejsza wartość
<code>covmat</code>	macierz kowariancji
<code>subset</code>	wyrażenie wskazujące podzbiór obserwacji do analizy
Metoda osi głównych	
<code>factor.pa(r, nfactors=1, residuals=FALSE, rotate="varimax", n.obs=NULL, min.err=0.001, digits=2, max.iter=50)</code>	
<code>r</code>	macierz korelacji lub macierz danych
<code>nfactors</code>	liczba wyodrębnionych czynników (wartość domyślna 1)
<code>residuals</code>	TRUE oznacza, że wyznaczana jest macierz reszt
<code>rotate</code>	rotacja: "none", "varimax", "promax"
<code>n.obs</code>	liczba obserwacji (gdy <code>r</code> jest macierzą korelacji)
<code>min.err</code>	iteracje są powtarzane do momentu, gdy zmiany zasobów zmienności wspólnej są mniejsze od <code>min.err</code> w dwóch kolejnych iteracjach
<code>digits</code>	liczba miejsc po przecinku w arkuszu wyników
<code>max.iter</code>	maksymalna liczba iteracji

*Opis niektórych funkcji zawiera podstawowe argumenty.

Źródło: opracowanie własne na podstawie dokumentacji programu R.

Wyniki zastosowania metody największej wiarygodności z rotacją *varimax*:

```
Call:
factanal(x = z, factors = 3)
```

Uniquenesses:

	x2	x3	x4	x5	x6	x7	x8	x9	x10
	0,005	0,388	0,129	0,159	0,185	0,393	0,359	0,077	0,373

Loadings:

	Factor1	Factor2	Factor3
x2	-0,241	-0,968	
x3	0,735	0,260	
x4	0,917	0,161	
x5	0,849	0,201	0,281
x6	0,490	0,462	0,601
x7	0,122	0,769	
x8	0,291	0,559	0,493
x9	0,891	0,140	0,330
x10	0,228	0,268	0,709

	Factor1	Factor2	Factor3
SS loadings	2,811	2,162	1,958
Proportion Var	0,312	0,240	0,218
Cumulative Var	0,312	0,553	0,770

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 5,72 on 12 degrees of freedom.

The p-value is 0,929

	Factor1	Factor2	Factor3
1	-0,24057618	0,2634660	0,06643402
2	-0,64105860	-0,5616635	1,95928031
3	-0,48063939	0,3464474	0,75868747
4	-0,24057618	0,2634660	0,06643402
5	-0,38824019	1,2300862	-0,18568921
6	-1,09774052	2,3376933	1,41665902
7	-0,38719685	-0,6168142	1,16284737
8	-0,24057618	0,2634660	0,06643402
9	0,12654655	1,0893885	-0,43934525
10	1,68479518	-1,1416680	0,65191388
11	-1,61910778	-2,1794102	-0,50623614
12	-0,08528339	-0,6994475	-0,16810783
13	-0,08528339	-0,6994475	-0,16810783
14	-0,65769447	0,3563070	-3,31839800
15	1,79099292	-0,2301984	-0,39664228
16	1,72308156	-0,2273437	-0,03085721
17	1,74282570	0,6941752	-0,67426964
18	-0,90426881	-0,4885028	-0,26103672

Następnie dla tych samych danych zastosowano analizę czynnikową metodą osi głównych z rotacją *varimax*. W tym celu wykorzystano funkcję `factor.pa` dostępną w pakiecie `psych`.

```
> og <- factor.pa(z, nfactors=3, residuals=FALSE,
min.err=0.001, digits=3, max.iter=1000)
> print(og$loadings)
```

W wyniku zastosowania tej procedury otrzymano następujące ładunki czynnikowe:

	PA1	PA2	PA3
x2	-0,258	-0,916	
x3	0,779	0,273	
x4	0,888	0,179	
x5	0,881	0,166	0,266
x6	0,519	0,439	0,563
x7	0,118	0,796	
x8	0,297	0,607	0,464
x9	0,893	0,150	0,307
x10	0,251	0,272	0,696

	PA1	PA2	PA3
SS loadings	2,865	2,166	1,893
Proportion Var	0,318	0,241	0,210
Cumulative Var	0,318	0,559	0,769

\$residual	x2	x3	x4	x5	x6	x7	x8	x9	x10
x2	0,094	-0,004	0,016	-0,021	-0,034	0,019	0,019	0,016	-0,006
x3	-0,004	0,316	0,045	-0,042	-0,053	0,038	0,025	0,016	-0,013
x4	0,016	0,045	0,177	0,005	-0,052	0,001	-0,010	0,027	0,016
x5	-0,021	-0,042	0,005	0,125	0,025	0,012	0,014	-0,025	-0,016
x6	-0,034	-0,053	-0,052	0,025	0,222	-0,035	0,005	0,009	0,050
x7	0,019	0,038	0,001	0,012	-0,035	0,350	0,001	0,011	-0,004
x8	0,019	0,025	-0,010	0,014	0,005	0,001	0,328	0,004	-0,022
x9	0,016	0,016	0,027	-0,025	0,009	0,011	0,004	0,086	-0,022
x10	-0,006	-0,013	0,016	-0,016	0,050	-0,004	-0,022	-0,022	0,378

Za pomocą funkcji `princomp` przeprowadzono także analizę głównych składowych bez rotacji.

```
loadings(princomp(z, scores=TRUE, cor=TRUE))
```

Otrzymano następujące wyniki:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
x2	0,294	-0,344	-0,494	-0,164		-0,407	0,459	-0,360	0,132
x3	-0,273	0,528	0,222	-0,364	0,429		0,359	-0,384	
x4	-0,330	-0,434	0,248	-0,188	0,348	-0,260	-0,135	0,205	0,597
x5	-0,370	-0,390			-0,148	0,173	-0,335	-0,710	-0,191
x6	-0,402		-0,101	0,385	-0,268	0,407	0,514		0,420
x7	-0,244	0,115	-0,667	-0,520		0,337	-0,248	0,173	
x8	-0,358	0,281		-0,139	-0,650	-0,579			
x9	-0,376	-0,385		-0,135			0,402	0,376	-0,620
x10	-0,320	0,145	-0,427	0,591	0,419	-0,342	-0,183		-0,138

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
SS loadings	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Proportion Var	0,111	0,111	0,111	0,111	0,111	0,111	0,111	0,111	0,111
Cumulative Var	0,111	0,222	0,333	0,444	0,556	0,667	0,778	0,889	1,000

5. Rotacja czynników

W funkcji `factanal` dostępne są tylko rotacje: ortogonalna *varimax* oraz ukośna *promax*. Szereg innych rotacji dostępnych jest w pakiecie `GPARotation` (tab. 4).

Tabela 4. Składnie funkcji służących do rotacji czynników wspólnych w środowisku R*

<code>GPForth(A, Tmat=diag(ncol(A)), normalize=FALSE, eps=1e-5, maxit=1000, method="varimax")</code>	
<code>GPFoblq(A, Tmat=diag(ncol(A)), normalize=FALSE, eps=1e-5, maxit=1000, method="quartimin")</code>	
A	macierz ładunków czynnikowych przed rotacją
Tmat	początkowa macierz rotacji (wartość domyślna – macierz jednostkowa)
normalize	TRUE oznacza, że przeprowadzana jest normalizacja Kaisera
Eps	wartość kryterium stopu
maxit	maksymalna liczba iteracji
method	metoda rotacji: "quartimin", "oblmin", "oblmax", "quartimax", "varimax"

*Opis niektórych funkcji zawiera podstawowe argumenty.

Źródło: opracowanie własne na podstawie dokumentacji programu R.

Rotacją domyślną w funkcji `factanal` jest *varimax*. Chcąc uzyskać ładunki czynnikowe bez rotacji, należy wywołać polecenie:

```
>factanal(z, factors=3, rotation="none")
```

Uniquenesses:

	x2	x3	x4	x5	x6	x7	x8	x9	x10
	0,005	0,388	0,129	0,159	0,185	0,393	0,359	0,077	0,373

Loadings:

	Factor1	Factor2	Factor3
x2	-0,996		
x3	0,728		0,283
x4	0,422	0,771	-0,312
x5	0,447	0,796	
x6	0,602	0,551	0,386
x7		0,426	0,648
x8	0,634	0,306	0,381
x9	0,403	0,870	
x10	0,344	0,425	0,573

	Factor1	Factor2	Factor3
SS loadings	2,951	2,751	1,229
Proportion Var	0,328	0,306	0,137
Cumulative Var	0,328	0,634	0,770

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 5,72 on 12 degrees of freedom.

The p-value is 0,929

W celu otrzymania łatwiejszych do interpretacji ładunków czynnikowych zaprezentowano rotację wyników analizy czynnikowej metodą *quartimax* dostępną w pakiecie GPARotation.

```
> GPForth(factanal(z, factors=3, rotation="none")
$loadings, normalize=FALSE, eps=1e-5, maxit=1000,
method="quartimax")
```

W wyniku uruchomienia tej procedury otrzymano ładunki czynnikowe po rotacji *quartimax*:

	Factor1	Factor2	Factor3
x2	-0,95190	-0,296	0,0349
x3	0,72361	0,153	0,2546
x4	0,10500	0,920	-0,1127
x5	0,14311	0,898	0,1189
x6	0,41613	0,622	0,5052
x7	-0,00282	0,265	0,7328
x8	0,52824	0,412	0,4377
x9	0,07860	0,944	0,1586
x10	0,23581	0,373	0,6575

Rotating matrix:

	[,1]	[,2]	[,3]
[1,]	0,9364	0,351	-0,0120
[2,]	-0,3366	0,907	0,2515
[3,]	0,0991	-0,232	0,9678

Na podstawie wyników analizy otrzymano trzy czynniki, które są opisane przez następujące zmienne:

- pierwszy czynnik jest scharakteryzowany przez zmienne: X2, X3;
- drugi czynnik jest opisany przez zmienne: X4, X5, X6 i X9;
- trzeci czynnik – przez zmienne: X7 i X10.

Zmienna X8 nie bierze istotnego udziału w opisie czynników.

6. Podsumowanie

W środowisku R dostępnych jest wiele funkcji, które przeprowadzają analizę czynnikową. W artykule przedstawiono najważniejsze z nich. Podstawową funkcją jest `factanal`, która wyznacza czynniki metodą największej wiarygodności. Jest ona dostępna w jądrze programu w pakiecie `stats`. Funkcja `factor.pa` przeprowadza analizę czynnikową metodą osi głównych i znajduje się w pakiecie `psych`. Obie funkcje przeprowadzają rotacje ortogonalną `varimax` i ukośną `promax` lub wykonują obliczenia bez rotacji; szereg innych rotacji dostępnych jest w pakiecie `GPARotation`. W środowisku R oprogramowane są tylko te dwie metody wyodrębniania czynników, bez metody głównych składowych, która właściwie nie należy do analizy czynnikowej. Można ją wyznaczyć, korzystając z funkcji `princomp` dostępnej w pakiecie `stats`. Wywołując funkcję `factanal` lub `factor.pa`, należy podać liczbę wyodrębnianych czynników głównych, natomiast funkcja `princomp` wyznacza tyle składowych, ile jest zmiennych obserwowalnych. Dodatkowo funkcja `factor.pa` wyznacza macierz reszt między macierzą korelacji obliczoną na podstawie zmiennych obserwowalnych a macierzą korelacji obliczoną na podstawie modelu czynnikowego.

Literatura

- Dalgaard P. (2002), *Introductory Statistics with R*, Springer-Verlag, New York.
- Everitt B.S., Hothorn T. (2006), *A Handbook of Statistical Analyses Using R*, Chapman & Hall, London.
- Gatnar E. (1998), *Analiza czynnikowa. Teoria i zastosowanie*, Akademia Ekonomiczna w Katowicach (maszynopis powielony).
- Kim J.O., Mueller C.W. (1978), *Factor Analysis. Statistical Methods and Practical Issues*, Sage, Beverly Hills.

- Psychologiczne i psychometryczne problemy diagnostyki psychologicznej* (1993), red. J. Brzeziński, Poznań, UAM.
- Sztemberg M. (1999), *Wykorzystanie analizy czynnikowej do pozycjonowania kawy na rynku*, [w:] *Taksonomia 6*, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 817, AE, Wrocław, s. 131-143.
- Zakrzewska M. (1994), *Analiza czynnikowa w budowaniu i sprawdzaniu modeli psychologicznych*, UAM, Poznań.
- Zeliaś A. (2000), *Metody statystyczne*, Wydawnictwo Naukowe PWN, Warszawa.

FACTOR ANALYSIS WITH R ENVIRONMENT

Summary

Factor analysis is a data reduction technique which takes a number of different variables and attempts to identify underlying relationship which may be presented. In the factor analysis an investigator is interested in explaining the relationships within his data set in terms of the smallest number of independent summary variables. In other words you identify some hidden basic variables as combinations of the variable observed.

In this article the author presents methods of extracting initial factors, rotation and determining a number of factors using R system.