

**Andrzej Bąk**

Uniwersytet Ekonomiczny we Wrocławiu

## **METODY SYMULACYJNE W BADANIACH PREFERENCJI Z WYKORZYSTANIEM PROGRAMU R**

### **1. Wstęp**

Na gruncie ekonomii metody symulacyjne znajdują zastosowanie w tych dziedzinach, w których kluczową rolę odgrywa proces podejmowania decyzji, zwłaszcza w warunkach niepewności, ryzyka, niepełnej informacji oraz wielowariantowości rozwiązań. Z tych też względów metody symulacyjne są bardzo użytecznym i często wykorzystywanym narzędziem do tworzenia modeli symulacyjnych firm, rynku, zarządzania, marketingu, produkcji, zapasów, teorii kolejek itp. Modelowanie symulacyjne może być użyteczne w różnych sferach aktywności ekonomicznej przedsiębiorstw, także w badaniach marketingowych i analizie rynku, w tym również w badaniach preferencji konsumentów.

Wraz z rozwojem techniki elektronicznej i oprogramowania komputerowego zakres i skala zastosowań metod symulacyjnych na gruncie badań marketingowych i badań preferencji ciągle wzrastają. Celem artykułu jest prezentacja możliwości wykorzystania metod symulacji komputerowej z wykorzystaniem programu R i w połączeniu z dekompozycyjnymi metodami badania preferencji, szczególnie *conjoint analysis*.

Program R nie oferuje pakietu bezpośrednio wspierającego badania preferencji za pomocą metod dekompozycyjnych, do których zalicza się dwie grupy metod: *conjoint analysis* i metody wyborów dyskretnych. Procedury obliczeniowe zawarte w różnych pakietach mogą być jednak wykorzystane w badaniach preferencji oraz w analizach symulacyjnych z zastosowaniem metod dekompozycyjnych.

W artykule scharakteryzowano dwie grupy zagadnień:

- pakiety i procedury obliczeniowe dostępne w programie R, które mogą znaleźć zastosowanie w badaniach preferencji za pomocą metod dekompozycyjnych (*conjoint analysis*),
- możliwości wykorzystania metod symulacyjnych w analizie danych o preferencjach z wykorzystaniem pakietów i funkcji programu R.

## 2. Dekompozycyjne metody badania preferencji

Pojęcie preferencji (relacji preferencji) odgrywa w ekonomii, zwłaszcza w teorii zachowań konsumentów i teoriach użyteczności, bardzo ważną rolę. Relacja preferencji jest podstawą badań indywidualnych decyzji (wyborów) konsumentów (popytu indywidualnego), co umożliwia z kolei analizę popytu rynkowego. Bezpośredni pomiar użyteczności, rozumianej jako poziom zadowolenia czy satysfakcji konsumenta, jest zadaniem trudnym. W celu jego rozwiązania stosuje się metody zmierzające do kwantyfikacji użyteczności oparte na kategorii preferencji. Preferencje są relacją binarną odnoszoną do wektorów opisujących wielowymiarowe obiekty, tzw. profile dóbr lub usług, których zbiory tworzą koszyki towarów lub plany konsumpcji. Relacja preferencji umożliwia przypisanie konsumentowi indywidualnej skali preferencji, na której można wartościować profile dóbr lub usług i optymalizować wybory rynkowe.

W badaniach preferencji konsumentów wykorzystuje się metody umożliwiające pomiar (kwantyfikację) preferencji oraz modele odwzorowujące zachowania rynkowe konsumentów. Ważną grupę narzędzi badawczych w tym obszarze stanowią tzw. metody dekompozycyjne, do których zalicza się dwie grupy metod: *conjoint analysis* i wyborów dyskretnych (zob. [Bąk 2004]).

Podstawą analizy danych o preferencjach w podejściu dekompozycyjnym jest model dekompozycyjny, reprezentujący zależność między preferencjami konsumentów a atrybutami profilów, którą w sposób ogólny można przedstawić w postaci równania:

$$U_{is} = f_s(\mathbf{X}, \boldsymbol{\beta}, \varepsilon_{is}) \quad (1)$$

gdzie:  $U_{is}$  – użyteczność całkowita  $i$ -tego profilu dla  $s$ -tego konsumenta wyrażająca jego preferencje,

$f_s$  – funkcja preferencji  $s$ -tego konsumenta,

$\mathbf{X}$  – macierz zawierająca realizacje zmiennych objaśniających opisujących profile (poziomy atrybutów lub realizacje zmiennych sztucznych reprezentujących układ czynnikowy),

$\boldsymbol{\beta}$  – macierz parametrów (użyteczności cząstkowych),

$\varepsilon_{is}$  – składnik losowy modelu.

Wartości zmiennej objaśnianej  $U_{is}$  reprezentujące preferencje empiryczne mogą być mierzone na skalach niemetrycznych lub metrycznych, natomiast wartości zmiennych objaśniających  $\mathbf{X}$  są mierzone najczęściej na skalach niemetrycznych. Wartości zmiennej objaśnianej są wynikiem bezpośrednich ocen respondentów i reprezentują ich preferencje. Wartości zmiennych objaśniających reprezentują natomiast poziomy atrybutów opisujących oceniane profile.

Na podstawie zgromadzonych ocen respondentów (użyteczności całkowitych profilów) dokonuje się podziału (dekompozycji) preferencji całkowitych w celu oszacowania tzw. użyteczności cząstkowych poziomów atrybutów oraz obliczenia udziałów poszczególnych atrybutów w całkowitej użyteczności każdego profilu (por. [Green, Wind 1975]).

### 3. Metody symulacyjne w badaniach preferencji

Metody symulacyjne mogą być stosowane w połączeniu z metodami dekompozycyjnymi w analizach następujących grup zagadnień o charakterze teoretycznym i aplikacyjnym z zakresu badań preferencji [Bąk 2004]:

- symulacyjne badanie własności formalnych i aplikacyjnych poszczególnych procedur dekompozycyjnych (*conjoint analysis* i wyborów dyskretnych),
- symulacyjne generowanie danych wykorzystywanych w badaniach formalnych i aplikacyjnych własności algorytmów dekompozycyjnych,
- wykorzystanie metod symulacyjnych w prognozowaniu preferencji konsumentów na podstawie danych zgromadzonych metodą *conjoint analysis*,
- symulacyjne modelowanie rynku na podstawie danych zgromadzonych metodą *conjoint analysis*.

Na gruncie badań preferencji metody symulacyjne mogą być wykorzystywane przede wszystkim do generowania danych losowych dotyczących preferencji konsumentów, do generowania podzbiorów profilów oraz w celu prowadzenia analiz porównawczych różnych modeli i algorytmów *conjoint analysis*.

### 4. Analiza *conjoint* w programie R

Aktualnie nie jest dostępny pakiet programu R realizujący w sposób kompleksowy procedurę *conjoint analysis*. Niemniej jednak różne pakiety zawierają funkcje, które mogą być wykorzystane do realizacji zadań obliczeniowych na wybranych etapach tej procedury.

Standardowa procedura pomiaru preferencji metodą *conjoint analysis* obejmuje następujące etapy [Green, Srinivasan 1978; Walesiak, Bąk 2000; Bąk 2004]:

1. Specyfikacja zadania badawczego (atrybuty profilów, zmienna objaśniana).
2. Wybór skali pomiaru preferencji (skale niemetryczne i metryczne).
3. Wybór metody gromadzenia danych o preferencjach (metoda pełnych profilów, porównania profilów parami).
4. Generowanie profilów do badań (układy czynnikowe, dane symulacyjne).
5. Wybór modelu (model regresji wielorakiej, model monotonicznej analizy wariancji).
6. Estymacja użyteczności cząstkowych (metoda najmniejszych kwadratów, algorytm monotonicznej analizy wariancji).
7. Ocena wiarygodności pomiaru.

8. Wykorzystanie wyników (ocena ważności atrybutów, segmentacja, udziały w rynku).

Nie wszystkie etapy procedury badawczej *conjoint analysis* wymagają stosowania metod obliczeniowych i korzystania z oprogramowania komputerowego. W tab. 1 zestawiono te etapy procedury, w których wykorzystanie metod obliczeniowych i algorytmów komputerowych jest niezbędne oraz wskazano na funkcje i pakiety programu R, które można w tych celach wykorzystać [Everitt 2004; Everitt, Dunn 2001]. Z zestawienia wynika, że ocena ważności atrybutów oraz symulacja udziałów w rynku (realizowane na 8 etapie procedury) nie są w aktualnie dostępnych pakietach programu R oprogramowane.

Tabela 1. Etapy procedury *conjoint analysis* w programie R

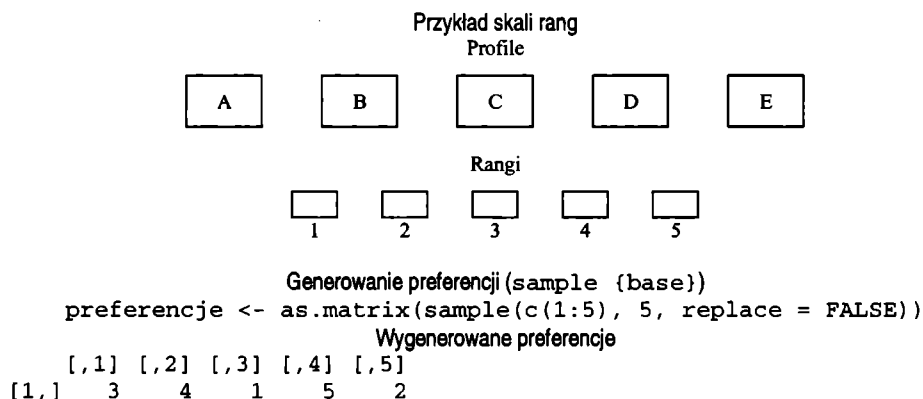
Etap procedury [nr]	Funkcja {pakiet}
Generowanie profili do badań – układy czynnikowe [4]	gen.factorial {AlgDesign}, optFederov {AlgDesign}
Generowanie preferencji symulacyjnych	sample {base}, runif {stats}, rbinom {stats}, rmultinom {stats}, rmulti {poLCA}, poLCA.simdata {poLCA}
Generowanie profili do badań – dane symulacyjne [4]	Mvnorm {mvtnorm}, rmulti {poLCA}, poLCA.simdata {poLCA}
Estymacja użyteczności częściowych (model regresji) [6]	lm {stats}, factor {base}, contrasts {stats}
Ocena ważności atrybutów [8]	brak
Symulacja udziałów w rynku [8]	brak
Segmentacja [8]	agnes {cluster}, kmeans {stats}

Źródło: opracowanie własne.

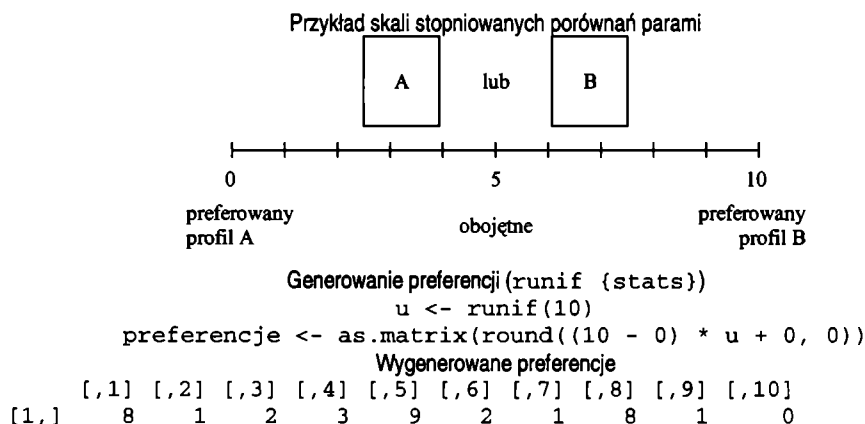
W literaturze przedmiotu brakuje przykładów badań preferencji metodami *conjoint analysis* z wykorzystaniem programu R. Dostępne funkcje, które mogą być przydatne w badaniach preferencji metodami *conjoint analysis*, wymagają weryfikacji empirycznej oraz analiz porównawczych.

W dalszej części artykułu przedstawiono przykłady wykorzystania technik symulacyjnych i funkcji programu R do generowania danych o preferencjach, generowania zbiorów profili oraz w celu oceny jakości (dopasowania do danych) modeli *conjoint analysis*.

**Generowanie losowych danych o preferencjach.** Symulacyjne dane o preferencjach mogą być przydatne w fazie przygotowywania procedur obliczeniowych oraz badania własności formalnych i aplikacyjnych różnych wariantów *conjoint analysis*. W sposób losowy mogą być generowane dane o preferencjach na różnych skalach pomiaru (niemetrycznych i metrycznych), wykorzystywane w badaniach symulacyjnych własności metod *conjoint analysis* (np. za pomocą funkcji `sample {base}`, `runif {stats}`, `rbinom {stats}`, `rmultinom {stats}`). Na rys. 1 i 2 przedstawiono przykłady generowania danych symulacyjnych o preferencjach mierzonych na skali niemetrycznej (skala rang) i metrycznej (skala stopniowanych porównań parami).



Rys. 1. Generowanie danych symulacyjnych o preferencjach na skali rang  
Źródło: opracowanie własne.



Rys. 2. Generowanie danych symulacyjnych o preferencjach na skali stopniowanych porównań parami  
Źródło: opracowanie własne.

**Generowanie zbioru profilów.** W procedurze badawczej *conjoint analysis* podstawową rolę odgrywa etap generowania zbioru profilów. W tym celu literatura przedmiotu rekomenduje układy czynnikowe (cząstkowe plany ortogonalne lub efektywne) albo metody symulacyjne [Green, Srinivasan 1978].

W programie R dostępne są funkcje generujące układy czynnikowe pełne i cząstkowe (`gen.factorial` {AlgDesign}, `optFederov` {AlgDesign}) oraz funkcje umożliwiające wygenerowanie zbioru profilów w sposób symulacyjny (np. `Mvnorm` {mvtnorm}, `poLCA.simdata` {poLCA}). Na rys. 3 przedstawiono przykład generowania cząstkowego układu czynnikowego dla trzech atrybutów (jednego dwupoziomowego i dwóch trypoziomowych) [Orme 2002].

Atrybuty i poziomy charakteryzujące przedmiot badań

Atrybut	Poziomy
marka	a, b, c
kolor	niebieski, czerwony
cena	niska, średnia, wysoka

```

Generowanie cząstkowego układu czynnikowego (optFederov {AlgDesign})
library(AlgDesign)
runif(1)
pelny = gen.factorial(c(3, 2, 3), factors = "all")
uklad = optFederov(-., pelny)
Wygenerowany podzbiór profilów (układ cząstkowy)
  1  2  5  6  9  10 11 13 14 16 18
X1 "1" "2" "2" "3" "3" "1" "2" "1" "2" "1" "3"
X2 "1" "1" "2" "2" "1" "2" "2" "1" "1" "2" "2"
X3 "1" "1" "1" "1" "2" "2" "2" "3" "3" "3" "3"

```

Rys. 3. Generowanie cząstkowego układu czynnikowego

Źródło: opracowanie własne.

Na rys. 4 zaprezentowano natomiast zbiór profilów wygenerowany w sposób losowy (za pomocą funkcji `poLCA.simdata()` z pakietu `poLCA`) dla tego samego zestawu atrybutów i poziomów. W obu przypadkach uzyskano podzbiory liczące po 11 profilów (pełny układ czynnikowy liczy 18 profilów).

Cząstkowy układ czynnikowy lub losowo wygenerowany podzbiór profilów stanowią podstawę gromadzenia danych o preferencjach respondentów. Dane o preferencjach respondentów można także, na potrzeby badań porównawczych, wygenerować losowo w sposób zilustrowany na rys. 1 i 2 (z wykorzystaniem funkcji `sample()` z pakietu `base` i `runif()` z pakietu `stats`).

```

Generowanie podzbioru profilów - wybór losowy (poLCA.simdata {poLCA})
library(poLCA)
probs <- list(matrix(c(0.6,0.1,0.3, 0.6,0.3,0.1, 0.3,0.1,0.6 ), ncol=3,byrow=TRUE),
             matrix(c(0.2,0.8, 0.7,0.3, 0.3,0.7 ), ncol=2,byrow=TRUE),
             matrix(c(0.3,0.6,0.1, 0.1,0.3,0.6, 0.3,0.6,0.1 ), ncol=3,byrow=TRUE))
podzbiór <- poLCA.simdata(N=11, probs, classdist=c(0.2,0.3,0.5))
Wygenerowany podzbiór profilów
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
X1  1  1  1  1  1  1  2  1  3  1  3
X2  2  1  2  2  2  1  1  2  2  1  2
X3  2  3  3  2  1  3  2  2  2  3  2

```

Rys. 4. Generowanie losowego podzbioru profilów

Źródło: opracowanie własne.

**Estymacja użyteczności cząstkowych.** W celu estymacji parametrów wykorzystuje się liniowy model regresji wielorakiej ze zmiennymi sztucznymi. Sposób kodowania atrybutów za pomocą zmiennych sztucznych (co ma wpływ na wartości szacowanych parametrów) można ustalić, korzystając z funkcji `contrasts()` z

pakietu `stats`, natomiast do szacowania modelu wykorzystuje się funkcję `lm()` z pakietu `stats` (zob. rys. 5).

```
options(contrasts = c("contr.sum", "contr.poly"))
m = lm(y$preferencje ~ factor(x$Marka) + factor(x$Kolor) + factor(x$Cena))
```

Wyniki estymacji na podstawie cząstkowego układu czynnikowego

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.0235	1.0064	5.985	0.00187 **
factor(x\$Marka)2	1.5059	0.9607	1.567	0.17778
factor(x\$Marka)3	2.6000	0.9028	2.880	0.03459 *
factor(x\$Kolor)2	0.9118	0.7644	1.193	0.28647
factor(x\$Cena)2	-1.9059	0.9607	-1.984	0.10406
factor(x\$Cena)3	-4.5059	0.9607	-4.690	0.00538 **

Residual standard error: 1.236 on 5 degrees of freedom

Multiple R-Squared: 0.8888, Adjusted R-squared: 0.7776

F-statistic: 7.994 on 5 and 5 DF, p-value: 0.01978

Wyniki estymacji na podstawie układu wygenerowanego losowo

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.4118	1.4271	3.091	0.02712 *
factor(x\$Marka)2	2.0588	1.0229	2.013	0.10030
factor(x\$Marka)3	3.5882	0.6697	5.358	0.00304 **
factor(x\$Kolor)2	1.0000	1.1272	0.887	0.41563
factor(x\$Cena)2	-1.4706	0.7611	-1.932	0.11117
factor(x\$Cena)3	-4.0588	1.4140	-2.871	0.03497 *

Residual standard error: 0.7971 on 5 degrees of freedom

Multiple R-Squared: 0.9725, Adjusted R-squared: 0.9451

F-statistic: 35.4 on 5 and 5 DF, p-value: 0.0006595

Rys. 5. Estymacja modelu *conjoint analysis*

Źródło: opracowanie własne.

Wyniki estymacji zamieszczone na rys. 5 wskazują, że dopasowanie modelu do danych jest lepsze w przypadku układu wygenerowanego losowo ( $R^2 = 0,97$ ) niż w przypadku cząstkowego efektywnego układu czynnikowego ( $R^2 = 0,89$ ). Wskazują na to także wyniki przebiegów symulacyjnych, których nie zamieszczono w artykule. Zagadnienie to wymaga jednak dalszych badań.

## 5. Podsumowanie

Zagadnienia scharakteryzowane w artykule oraz wstępne wyniki przeprowadzonych badań prowadzą do następujących wniosków:

- modele *conjoint analysis* szacowane na podstawie układów czynnikowych i danych symulacyjnych (profilów wybieranych losowo) charakteryzują się różną jakością dopasowania do danych – zagadnienie to wymaga dalszych badań,
- generatory profilów wybieranych losowo wymagają badań porównawczych,

- funkcje programu R oferowane w różnych pakietach mogą być wykorzystane do estymacji modeli *conjoint analysis* oraz w szerzej rozumianych badaniach preferencji,
- program R nie zawiera pakietu realizującego całą procedurę *conjoint analysis*, w związku z czym istnieje potrzeba opracowania takiego pakietu.

## Literatura

- Bąk A. (2004), *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1013, AE, Wrocław.
- Everitt B.S. (2004), *An R and S-PLUS Companion to Multivariate Analysis*, Springer-Verlag, London.
- Everitt B.S., Dunn G. (2001), *Applied Multivariate Data Analysis*, Arnold, London.
- Green P.E., Srinivasan V. (1978), *Conjoint Analysis in Consumer Research: Issues and Outlook*, „Journal of Consumer Research”, 5, September, s. 103-123.
- Green P.E., Wind Y. (1975), *New Way to Measure Consumers' Judgments*, „Harvard Business Review”, 53, July-August, s. 107-117.
- Orme B. (2002), *Analysis of Traditional Conjoint Using Microsoft Excel™: An Introductory Example*, (URL:) <http://www.sawtoothsoftware.com>.
- Walesiak M., Bąk A. (2000), *Conjoint analysis w badaniach marketingowych*, AE, Wrocław.

## SIMULATION METHODS IN THE RESEARCH OF PREFERENCES USING R COMPUTER PROGRAMME

### Summary

Simulation methods may be used in the research of consumer preferences linked with conjoint analysis. R computer programme does not offers package containing whole conjoint analysis procedure. Nevertheless various packages may be used in the research of consumer preferences and simulation analysis connection with conjoint analysis.

The paper presents R packages and procedures which may be used in the research of preferences using conjoint analysis and simulation methods.