# Aktywność obywatelska - skala mokkena

*Szymon Talaga*

*11.12.2014*

Raport przedstawia procedure skalowania aktywności obywatelskiej/społecznej (socact1 do socact14) w oparciu o jednowymiarowy nieparametryczny model Mokkena. Tak samo jak w przypadku innych raportów i skryptów poprawne wykonanie wszystkich przedstawionych tu procedur jest uzależnione od poprawnej struktury systemu plików (katalogów).

Opisy zmiennych znaleźć można w pliku Codebook.pdf w folderze MainData.

Wczytanie pakietu 'mokken', 'knitr' i dodatkowych funkcji własnych:

```
library(mokken)
```

```
## Loading required package: poLCA
## Loading required package: scatterplot3d
## Loading required package: MASS
```

```
library(knitr)
library(lattice)
source("../Scaling/scaling-Helper.R")
```

Wczytanie zbioru danych ("MainData1.RData") i ograniczenie go do zestawu zmiennych socact.

```
load("../MainData/MainData1.RData") # load the main dataset
D.back = D # backup copy of the full dataset
D = D[,which(names(D)=="socact1"):which(names(D)=="socact14")]
```

Pierwsze sześć obserwacji zbioru danych:

| socact1 | socact2 | socact3 | socact4 | socact5 | socact6 | socact7 | socact8 | socact9 | socact10 | socact11 | socact12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Zmienne posortowane pod względem poziomu trudności:

| | |
|---|---|
| socact11 | 0.017 |
| socact7 | 0.022 |
| socact8 | 0.035 |

| | |
|---|---|
| socact10 | 0.035 |
| socact2 | 0.039 |
| socact12 | 0.065 |
| socact3 | 0.152 |
| socact14 | 0.156 |
| socact9 | 0.190 |
| socact4 | 0.212 |
| socact13 | 0.268 |
| socact5 | 0.273 |
| socact6 | 0.459 |
| socact1 | 0.593 |

Procedura testowania dopasawania danych do skali Mokkena oparta jest na: van Schuur, 2003; van der Ark, 2007, 2012. Ogólnie rzecz biorąc skala Mokkena jest skalą kumulatywną, co oznacza, że zakłada się, iż pozytywna odpowiedź na pytanie trudniejsze powinna pociągać za sobą pozytywne odpowiedzi na pytania łatwiejsze. Można zatem powiedzieć, że jest ona stochastycznym uogólnieniem deterministycznego skalogramu Guttmana. W związku z tym ocena dopasowania danych do modelu oparta jest na zliczaniu tzw. błędów gutmannowskich, czyli przypadków, w których respondent udziela odpowiedzi pozytywnej na pytanie trudniejsze i negatywnej na pytanie łatwiejsze. Należy przy tym pamiętać, że w kontekście skali Mokkena prosta średnia arytmetyczna jest wystarczającym estymatorem trudności pozycji testowej, zatem trudność pytania to po prostu odsetek pozytywnych odpowiedzi.

Kumulatywność skali Mokkena objawia się poprzez jej dwie właściwości - homogeniczną monotoniczność i podwójną monotoniczność. Homogeniczna monotoniczność oznacza, że dla każdego badanego prawdopodobieństwo pozytywnej odpowiedzi na pytanie jest dodatnio związane z występującym u niego natężeniem badanej cechy - w tym przypadku aktywności obywatelskiej. W przypadku, gdy homogeniczna monotoniczność faktycznie występuje, to gwarantuje to poprawny pomiar cechy na skali porządkowej, w której ilość pozytywnych odpowiedzi jest wystarczającym estymatorem badanej cechy. Spełnienie dodatkowego wymogu podwójnej homogeniczności - co oznacza, że dla każdej osoby wszystkie pytania można ułożyć w taki sam sposób od najłatwiejszego do najtrudniejszego - prowadzi do tego, że odsetek pozytywnych odpowiedzi jest wystarczającym estymatorem trudności pozycji testowej.

Ponadto skala Mokkena jest oparta również na dwóch dodatkowych założeniach - jednowymiarowości (która zostanie sprawdzona na koniec) oraz lokalnej stochastycznej niezależności. Niezależność odnosi się do faktu, że zakłada się, iż odpowiedzi udzielane przez badanych są zależne jedynie od występującego u nich natężenia badanej cechy. Założenie to nie będzie weryfikowane jednak nie jest to konieczne - tak naprawdę jest to ciche założenie stojące za właściwie każdym pomiarem kwestionariuszowym i jako takie zostanie uznane za spełnione.

We wszystkich tabelach, które zostaną pokazane najważniejsze są kolumny #ac, #vi, #zsig i crit. Pierwsza odnosi się do maksymalnej możliwej ilości naruszeń modelu dla danej zmiennej; #vi to ilość faktycznie występujących naruszeń; #zsig to ilość statystyczne istotnych naruszeń; crit to specjalna statystyka, która jest miarą krytycznościtych naruszeń - wartości powyżej 40 to umiarkowane naruszenia, zaś wartości powyżej 80 to naruszenia bardzo silne.

Zgodność z założoniem homogenicznej monotoniczności / ukrytej monotoniczności (monotonous homogeneity / latent homogeneity)

```
mono = check.monotonicity(D)
kable(summary(mono))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.71 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact7 | 0.59 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact8 | 0.46 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact10 | 0.51 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact2 | 0.56 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact12 | 0.40 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact3 | 0.34 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact14 | 0.27 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact9 | 0.43 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact4 | 0.38 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact13 | 0.37 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact5 | 0.43 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact6 | 0.35 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact1 | 0.51 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Jak widać dane są doskonale zgodne z modelem homogenicznej monotoniczności.

Zgodnośc z założeniem podwójnej monotoniczności. Metoda macierzy $P$:

```
pmat = check.pmatrix(D)
kable(summary(pmat))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.71 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact7 | 0.59 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact8 | 0.46 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact10 | 0.51 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact2 | 0.56 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact12 | 0.40 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact3 | 0.34 | 156 | 1 | 0.01 | 0.03 | 0.03 | 2e-04 | 1.77 | 1 | 24 |
| socact14 | 0.27 | 156 | 1 | 0.01 | 0.05 | 0.05 | 3e-04 | 2.11 | 1 | 31 |
| socact9 | 0.43 | 156 | 2 | 0.01 | 0.05 | 0.08 | 5e-04 | 2.11 | 2 | 30 |
| socact4 | 0.38 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact13 | 0.37 | 156 | 2 | 0.01 | 0.04 | 0.07 | 5e-04 | 0.00 | 0 | 6 |
| socact5 | 0.43 | 156 | 2 | 0.01 | 0.04 | 0.07 | 5e-04 | 0.00 | 0 | 3 |
| socact6 | 0.35 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact1 | 0.51 | 156 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |

Są pewne naruszenia modelu, ale są one stosunkowo nieznaczne. Widać, że narusza je najsilniej pozycja socact14 (działalność w stowarzyszeniu / klubie sportowym).

Zgodność z założeniem podwójnej monotoniczności. Metoda wyników resztowych (restscores):

```
rest = check.restscore(D)
kable(summary(rest))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.71 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact7 | 0.59 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact8 | 0.46 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact10 | 0.51 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact2 | 0.56 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact12 | 0.40 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact3 | 0.34 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact14 | 0.27 | 13 | 2 | 0.15 | 0.05 | 0.09 | 0.0066 | 1.11 | 0 | 38 |
| socact9 | 0.43 | 13 | 1 | 0.08 | 0.05 | 0.05 | 0.0037 | 1.11 | 0 | 18 |
| socact4 | 0.38 | 13 | 1 | 0.08 | 0.04 | 0.04 | 0.0030 | 0.86 | 0 | 17 |
| socact13 | 0.37 | 13 | 1 | 0.08 | 0.05 | 0.05 | 0.0041 | 1.18 | 0 | 22 |
| socact5 | 0.43 | 13 | 1 | 0.08 | 0.05 | 0.05 | 0.0041 | 1.18 | 0 | 19 |
| socact6 | 0.35 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact1 | 0.51 | 13 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |

Znowu są pojawiają się lekkie niezgodności z modelem i pozycja socact14 jest najgorzej dopasowana.

W tym momencie należy również zwrócić uwagę na kolumnę itemH w powyższej tabeli, która przedstawia parametry skalowania dla poszczegónych pozycji. Jest to specyficzna miara korelacji każdej pozycji z pozostałymi pozycjami oparta na ilości błędów gutmannowskich. Wartości poniżej 0,3 oznaczają kiepskie dopasowanie (van Schuur, 2003). Jak widać oznacza, to że pozycja socact14 powinna zostać usunięta ze skali. Ma to również sens intuicyjny, ponieważ określenie stowarzyszenie / klub sportowy jest zdecydowanie nieostre i może się nakładać na inne pytanie, co w oczywisty sposób może prowadzić do wystąpienia dużej ilości wzorców odpowiedzi niezgodnych z ideałem skalogramu guttmanowskiego.

Zatem teraz przedstawione zostanie dopasowanie danych do modelu skali Mokkena z wyłączeniem pytania socact14.

Homogeniczna monotoniczność:

```
mono = check.monotonicity(D)
kable(summary(mono))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.69 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact7 | 0.62 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact8 | 0.46 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact10 | 0.55 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact2 | 0.58 | 0 | 0 | NaN | 0 | 0 | NaN | 0 | 0 | 0 |
| socact12 | 0.42 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact3 | 0.37 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact9 | 0.47 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact4 | 0.40 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact13 | 0.36 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact5 | 0.44 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact6 | 0.39 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| socact1 | 0.52 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Znowu występuje pełna zgodność z założeniem homogenicznej monotoniczności.

Podwójna monotoniczność - metoda macierzy $P$:

```
pmat = check.pmatrix(D)
kable(summary(pmat))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.69 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact7 | 0.62 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact8 | 0.46 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact10 | 0.55 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact2 | 0.58 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact12 | 0.42 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact3 | 0.37 | 132 | 1 | 0.01 | 0.03 | 0.03 | 3e-04 | 1.77 | 1 | 23 |
| socact9 | 0.47 | 132 | 1 | 0.01 | 0.03 | 0.03 | 3e-04 | 1.77 | 1 | 18 |
| socact4 | 0.40 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact13 | 0.36 | 132 | 2 | 0.02 | 0.04 | 0.07 | 6e-04 | 0.00 | 0 | 7 |
| socact5 | 0.44 | 132 | 2 | 0.02 | 0.04 | 0.07 | 6e-04 | 0.00 | 0 | 3 |
| socact6 | 0.39 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |
| socact1 | 0.52 | 132 | 0 | 0.00 | 0.00 | 0.00 | 0e+00 | 0.00 | 0 | 0 |

Są lekkie odchylenia, ale żadna z wartości crit nie przekracza 40. Ponadto wszystkie pozycje charakteryzują się współczynnikami skalowania ($H$) większymi od 0,3.

Podwójna monotoniczność - metoda wartości resztowych:

```r
rest = check.restscore(D)
kable(summary(rest))
```

| | ItemH | #ac | #vi | #vi/#ac | maxvi | sum | sum/#ac | zmax | #zsig | crit |
|---|---|---|---|---|---|---|---|---|---|---|
| socact11 | 0.69 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact7 | 0.62 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact8 | 0.46 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact10 | 0.55 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact2 | 0.58 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact12 | 0.42 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact3 | 0.37 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact9 | 0.47 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact4 | 0.40 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact13 | 0.36 | 12 | 1 | 0.08 | 0.05 | 0.05 | 0.0043 | 1.12 | 0 | 23 |
| socact5 | 0.44 | 12 | 1 | 0.08 | 0.05 | 0.05 | 0.0043 | 1.12 | 0 | 19 |
| socact6 | 0.39 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |
| socact1 | 0.52 | 12 | 0 | 0.00 | 0.00 | 0.00 | 0.0000 | 0.00 | 0 | 0 |

Ponownie nie ma żadnych silnych naruszeń modelu - dodatkowo żadne z odchyleń nie jest istotne statystycznie. Oznacza to, że dane są dobrze dopasowane do modelu podwójnej monotoniczności, zatem została osiągnięta pełna zgodność z modelem skali Mokkena.

Globalby parametr skalowania dla całej skali wynosi 0.442, (0.055) (w nawiasach podano błąd standrdowy), co oznacza, że jego 95% przedział ufności to: 0.3342, 0.5498. Można zatem powiedzieć, że opracowana skala jest dosyć dobrej jakości.

Rzetelność skali mierzona współczynnikie $\alpha$-Cronbacha wyniosła 0,76, co wskazuje na stosunkowo dobrą precyzję pomiaru. Współczynnik MS będący miarą stworzoną specjalnie dla skali Mokkena (van der Ark, 2012) wskazuje na nawet trochę wyższą rzetelność, ponieważ wyniósł on 0,77.

Na koniec zostanie sprawdzona jednowymiarowość skali. Testem będzie automatyczna procedura grupowania pozycji testowych w jednowymiarowe skale. Jeżeli wszystkie pozycje zostaną zgrupowane w jeden wymiar/skale, to będzie to oznaczało, że opracowana skala jest w istocie jednowymiarowa. Tradycyjnie do grupowania pozycji testowych stosuje się opracowany przez Mokkena algorytm oparty o hierarchiczną analizę skupień. Jednak jest to procedura, która ma niejednoznaczną funckję celu i jako taka nie może być uznana za optymalną. Van der Ark zaproponował lepszą metodę, opartą o algorytm genetyczny i w pełni zgodną z duchem oryginalnej idei Mokkena, która posiada jednoznaczną funckję celu i jako taka daje bardziej rzetelne wyniki.

```r
scale = aisp(D, search="ga")
kable(scale)
```

| | Scale |
|---|---|
| socact11 | 1 |
| socact7 | 1 |

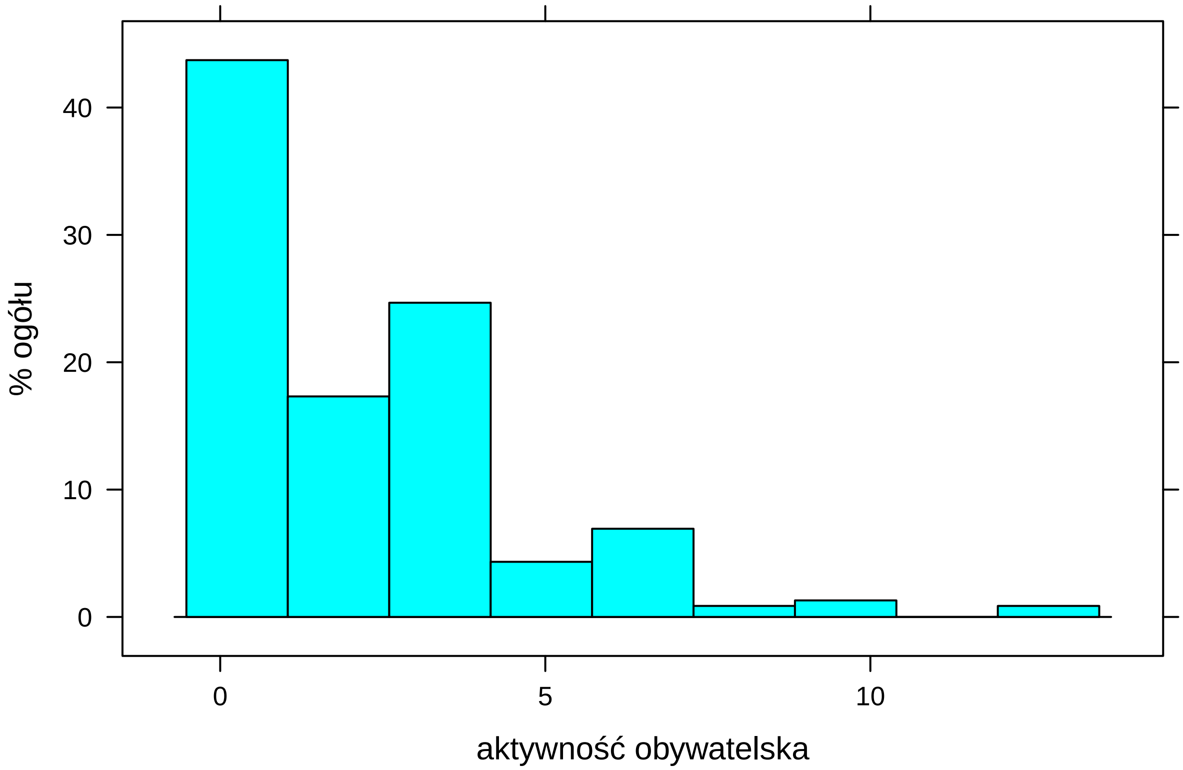
| | Scale |
|---|---|
| socact8 | 1 |
| socact10 | 1 |
| socact2 | 1 |
| socact12 | 1 |
| socact3 | 1 |
| socact9 | 1 |
| socact4 | 1 |
| socact13 | 1 |
| socact5 | 1 |
| socact6 | 1 |
| socact1 | 1 |

Jak widać algorytm przydzielił wszystkie pozycje do jednej skali, co oznacza, że jest ona w pełni jednowymiarowa. Należy zatem powiedzieć, że opracowana skala jest stosunkowo wysokiej jakości i może być uznana za rzetelny wskaźnik aktywności obywatelskiej.

Rozkład aktywności obywatelskiej:

```r
civic = apply(D, 1, sum)
summary(civic)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   2.000   2.359   3.000  13.000
```




Rozkład pokazuje, że skala nie jest idealna, ponieważ jest odrobinę zbyt trudna - dobrze różnicuje pomiędzy osobami o wysokiej aktywności, ale nie różnicuje tych mało aktywnych.

Na koniec należy zaznaczyć, że chociaż formalnie oparta o model Mokkena skala ma charakter porządkowy, to praktyka badawcza pokazuje, że z powodzeniem może ona być używana w charakterze skali przedziałowej (van der Gaag, Snijders, 2005).

**Koniec raportu**