

Porównanie wyników porządkowania, ze względu na różne sposoby wyznaczania zmiennej syntetycznej

Wstęp

Zbiór zawierający 8 najbardziej różniących się obiektów, został poddany porządkowaniu, przy wykorzystaniu metody opartej na zmiennej syntetycznej, sztuczna funkcja, która każdemu obiektu przyporządkowuje pewną wartość zmiennych po ich przekształceniu. Poniżej zostaną zaprezentowane zestawienie wyników porządkowania ze względu na zastosowanie różnych sposobów wyliczenia zmiennej syntetycznej, m.in. przy wykorzystaniu średniej arytmetycznej, średniej geometrycznej, średniej harmonicznej. Do porządkowania wybrano metodę bezwzorcową - metodę rang. Na wstępie należy dodać, że wszystkie kroki porządkowania za wyjątkiem sposobu liczenia zmiennej syntetycznej, są identyczne w każdym podejściu (stymulacja, normalizacja zmiennych, przypisywanie rang).

Przygotowanie zbioru

W pierwszej kolejności należy zaimportować zbiór obiektów. Porządkowanie będzie przeprowadzone na podstawie wartości zmiennych: cena, moc, pojemność, rok produkcji, przebieg

```
library(readxl)
zbior_danych <- read_excel("~/Praca licencjacka/Moje dane_zrodlo/8_Rozniacych_sie_obiektow.xlsx",
  sheet = "Arkusz1", col_types = c("numeric","text",
    "text", "text", "text", "text",
    "numeric", "numeric", "text", "numeric",
    "text", "text", "text", "text","text",
    "text", "numeric", "text", "text",
    "text", "numeric", "numeric",
    "numeric", "numeric", "numeric"))
```

`col_type = "blank"` deprecated. Use "skip" instead.

Podgląd danych:

```
head(zbior_danych)

## # A tibble: 6 x 29
##   Nr MARKA MODEL WERSJA TYP WOJEWODZTWO `CENA.BRUTTO_[p~`MOC_[km]`
##   <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 1.00 Mazda 3 II komp~ zachodniop~ 25200 105
## 2 2.00 Jaguar XF X260 kombi dolnoslask~ 323600 240
## 3 3.00 Subaru B9 Tr~ <NA> suv malopolskie 38900 245
## 4 4.00 Volks~ Golf VII kombi lodzkie 113900 150
## 5 5.00 Peuge~ 508 <NA> kombi slaskie 42500 115
## 6 6.00 Opel Antara <NA> suv lodzkie 24000 150
## # ... with 21 more variables: `POJEMNOSC.SKOKOWA_[cm3]` <dbl>,
## # ROK.PRODUKCJI <dbl>, `PRZEBIEG_[km]` <dbl>, KOLOR <chr>, L.DZRZWI
## # <dbl>, RODZAJ.PALIWA <chr>, SKRZYNIA.BIEGOW <chr>, NAPED <chr>,
## # KRAJ.AKTUALNEJ.REJESTRACJI <chr>, KRAJ.POCHODZENIA <chr>,
## # STATUS.POJAZDU.SPROWADZONEGO <chr>, PIERWSZY.WLASCICIEL <dbl>,
## # KTO.SPRZEDAJE <chr>, STAN <chr>, SERWISOWANY <chr>, ABS <dbl>,
## # KOMPUTER.POKLADOWY <dbl>, ESP <dbl>, KLIMATYZAJCA <dbl>, BEZWYPADKOWY
```

```
## # <dbl>, USZKODZONY <dbl>
```

Przypisanie innej nazwy dla danych, aby się wygodniej pracowało na nich:

```
dane2<-zbior_danych
```

Aby każda zmienna, tj. cecha miała ten sam charakter należy przeprowadzić stymulację zmiennej przebieg - uczynić z niej stymulantę, w tym celu skorzystam z przekształcenia ilorazowego

$$x_{ij}^S = [x_{ij}^D]^{-1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.$$

Kod w R:

```
for (i in 1:nrow(dane2)){  
  dane2[i,11]=1/dane2[i,11]  
}
```

W celu uzyskania porównywalności między zmiennymi, zostały one poddane transformacji normalizacyjnej - unitaryzacji, dzięki temu wartości każdej zmiennej należą do przedziału [0,1]

$$z_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.$$

Kod w R

```
maksi=0  
minim=0  
for (j in 7:11){  
  maksi[j]=max(dane2[j])  
  minim[j]=min(dane2[j])  
  for (i in 1:nrow(dane2)){  
    dane2[i,j]=(dane2[i,j]-minim[j])/(maksi[j]-minim[j])  
  }  
}  
  
dane2$`CENA.BRUTTO_[pln]` #podłód danych po transformacji normalizacyjnej
```

```
## [1] 0.07127584 1.00000000 0.11391499 0.34734097 0.12511944 0.06754103  
## [7] 0.00000000 0.02801423
```

W kolejnym kroku dla każdego obiektu, w każdej zmiennej opisującej obiekt przyporządkowano rangę na podstawie wartości wartości tej zmiennej w odniesieniu do pozostałych obiektów. Gdzie rangę=1 uzyskał obiekt, którego znormalizowana wartość j-tej cechy była największa. W tym celu stworzono pomocnicze kolumny, oraz zastosowano funkcję rank(-x)

```
dane2$range1_cena=rank(-dane2$`CENA.BRUTTO_[pln]`)  
dane2$range2_moc=rank(-dane2$`MOC_[km]`)  
dane2$range3_pojemnosc=rank(-dane2$`POJEMNOSC.SKOKOWA_[cm3]`)  
dane2$range4_rok=rank(-dane2$`ROK.PRODUKCJI`)  
dane2$range5_przebieg=rank(-dane2$`PRZEBIEG_[km]`)
```

Wyznaczanie zmiennej syntetycznej

Wyznaczanie zmiennej syntetycznej przy wykorzystaniu średniej arytmetycznej

Zmienna syntetyczna jest postaci:

$$s_i = \frac{1}{m} \sum_{j=1}^m z_{ij}, \quad i = 1, 2, \dots, n$$

gdzie m - to liczba zmiennych opisujących obiekty,

```
dane2[, "zmienna_syntetyczna_ar"] <- NA #zainicjowanie pustej kolumny

for (i in 1:nrow(dane2)){
  dane2[i, ncol(dane2)] = (dane2[i, 30] + dane2[i, 31] + dane2[i, 32] + dane2[i, 33] + dane2[i, 34]) / 5
}
```

Wyznaczanie zmiennej syntetycznej przy wykorzystaniu średniej geometrycznej

Zmienna syntetyczna jest postaci:

$$s_i = \prod_{j=1}^m (z_{ij})^{\frac{1}{m}}, \quad i = 1, 2, \dots, n,$$

gdzie m - to liczba zmiennych opisujących obiekty,

```
dane2[, "zmienna_syntetyczna_ge"] <- NA #zainicjowanie pustej kolumny

for (i in 1:nrow(dane2)){
  dane2[i, ncol(dane2)] = (dane2[i, 30] * dane2[i, 31] * dane2[i, 32] * dane2[i, 33] * dane2[i, 34]) ^ (1/5)
}
```

Wyznaczanie zmiennej syntetycznej przy wykorzystaniu średniej harmonicznej

Zmienna syntetyczna jest postaci:

$$s_i = \left[\sum_{j=1}^m \frac{1}{z_{ij}} \right]^{-1} \cdot m, \quad i = 1, 2, \dots, n,$$

gdzie m - to liczba zmiennych opisujących obiekty,

```
dane2[, "zmienna_syntetyczna_ha"] <- NA #zainicjowanie pustej kolumny

for (i in 1:nrow(dane2)){
  dane2[i, ncol(dane2)] = 5 * 1 / ((dane2[i, 30]) ^ (-1) + (dane2[i, 31]) ^ (-1) + (dane2[i, 32]) ^ (-1) +
    (dane2[i, 33]) ^ (-1) + (dane2[i, 34]) ^ (-1))
}
```

Podgląd fragmentu tabeli, po wyliczeniu zmiennych syntetycznych, gdzie kolumna “Nr” oznacza indeksy obiektów poddanych porządkowaniu, a kolumny “zmienna_syn_ar”, “zmienna_syntetyczna_ge”, “zmienna_syntetyczna_ha”, oznaczają kolejno: zmienną syntetyczną wyznaczoną za pomocą średniej arytmetycznej, średniej geometrycznej, średniej harmonicznej.

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

knitr::kable(select(dane2, c("Nr", "zmienna_syntetyczna_ar", "zmienna_syntetyczna_ge",
                             "zmienna_syntetyczna_ha")))
```

Nr	zmienna_syntetyczna_ar	zmienna_syntetyczna_ge	zmienna_syntetyczna_ha
1	5.0	4.934296	4.872970
2	1.7	1.643752	1.578947
3	3.1	2.448480	1.864407
4	3.0	2.361878	1.926605
5	4.8	4.521602	4.251012
6	4.8	4.608620	4.408397
7	7.0	6.575007	6.000000
8	6.6	6.484588	6.354995

Uporządkowanie obiektów ze względu na wartości wyznaczonych zmiennych syntetycznych:

Następnie tabela “dane2” została uporządkowana ze względu na kolumnę “zmienna_syntetyczna_ar”, a w kolejnym kroku została stworzona tabela “dane_arytmetyczna”, zawiera obiektu uporządkowane ze względu na kolumnę “zmienna_syntetyczna_ar”, oraz kolumny dotyczące indeksu, marki, modelu i zmiennej syntetycznej wyznaczonej za pomocą średniej harmonicznej, średniej geometrycznej. Analogiczne uporządkowanie oraz stworzenie nowej kolumny zostało przeprowadzone ze względu na kolumny “zmienna_syntetyczna_ge”, “zmienna_syntetyczna_ha”, co zostało zamieszczone poniżej.

```
dane2<-dane2[order(dane2$zmienna_syntetyczna_ar),]

library(dplyr)

dane_arytmetyczna<-select(dane2, c("Nr", "MARKA", "MODEL",
                                   "zmienna_syntetyczna_ar", "zmienna_syntetyczna_ge",
                                   "zmienna_syntetyczna_ha"))

dane2<-dane2[order(dane2$zmienna_syntetyczna_ge),]
dane_geometryczna<-select(dane2, c("Nr", "MARKA", "MODEL",
                                   "zmienna_syntetyczna_ge", "zmienna_syntetyczna_ar",
                                   "zmienna_syntetyczna_ha"))
```

```
dane2<-dane2[order(dane2$zmienna_syntetyczna_ha),]
dane_harmoniczna<-select(dane2, c("Nr", "MARKA", "MODEL",
                                "zmienna_syntetyczna_ha", "zmienna_syntetyczna_ar",
                                "zmienna_syntetyczna_ge"))
```

Podgląd tabel: “dane_arytmetyczna”, “dane_geometryczna”, “dane_harmoniczna”

```
#dane_arytmetyczna
```

```
knitr::kable(dane_arytmetyczna[])
```

Nr	MARKA	MODEL	zmienna_syntetyczna_ar	zmienna_syntetyczna_ge	zmienna_syntetyczna_ha
2	Jaguar	XF	1.7	1.643752	1.578947
4	Volkswagen	Golf	3.0	2.361878	1.926605
3	Subaru	B9 Tribeca	3.1	2.448480	1.864407
5	Peugeot	508	4.8	4.521602	4.251012
6	Opel	Antara	4.8	4.608620	4.408397
1	Mazda	3	5.0	4.934296	4.872970
8	Seat	Toledo	6.6	6.484588	6.354995
7	Fiat	Seicento	7.0	6.575007	6.000000

```
#dane_geometryczna
```

```
knitr::kable(dane_geometryczna[])
```

Nr	MARKA	MODEL	zmienna_syntetyczna_ge	zmienna_syntetyczna_ar	zmienna_syntetyczna_ha
2	Jaguar	XF	1.643752	1.7	1.578947
4	Volkswagen	Golf	2.361878	3.0	1.926605
3	Subaru	B9 Tribeca	2.448480	3.1	1.864407
5	Peugeot	508	4.521602	4.8	4.251012
6	Opel	Antara	4.608620	4.8	4.408397
1	Mazda	3	4.934296	5.0	4.872970
8	Seat	Toledo	6.484588	6.6	6.354995
7	Fiat	Seicento	6.575007	7.0	6.000000

```
#dane_harmoniczna
```

```
knitr::kable(dane_harmoniczna[])
```

Nr	MARKA	MODEL	zmienna_syntetyczna_ha	zmienna_syntetyczna_ar	zmienna_syntetyczna_ge
2	Jaguar	XF	1.578947	1.7	1.643752
3	Subaru	B9 Tribeca	1.864407	3.1	2.448480
4	Volkswagen	Golf	1.926605	3.0	2.361878
5	Peugeot	508	4.251012	4.8	4.521602
6	Opel	Antara	4.408397	4.8	4.608620
1	Mazda	3	4.872970	5.0	4.934296
7	Fiat	Seicento	6.000000	7.0	6.575007
8	Seat	Toledo	6.354995	6.6	6.484588

Wnioski

Na podstawie powyższych tabel widać, iż różnice w wynikach uporządkowania są niezauważalne dla obiektów uporządkowanych ze względu na zmienną syntetyczną wyznaczoną za pomocą średniej arytmetycznej oraz średniej geometrycznej. Być może gdyby zostały dobrane obiekty, których zmienne byłyby bardziej zróżnicowane w obiektach to można by zauważyć różnice po przeprowadzeniu porządku. Z kolei w przypadku obiektów dla których zmienna syntetyczna została obliczona z wykorzystaniem średniej harmonicznej, widać iż zmianie uległa pozycja 2 oraz ostatnia. W wyniku porządku, na miejscu 2 znalazło się Subaru, a na ostatnim miejscu Seat.