

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
> getwd()
[1] "C:/Users/DC/Desktop/StatDem"
> # Wyświetlenie listy plików w katalogu roboczym.
> list.files()
[1] "data"                "Dominanta_R.txt"
[3] "Eo_World_Baza.csv"   "Eo_World_bis.xlsx"
[5] "Eo_World_Red.xlsx"  "GospDom_Polska_2002.csv"
[7] "GospDom_Polska_2002.xlsx" "GospDom_Szczecin_1900.csv"
[9] "KobietyDzieci_1980_Polska.csv" "KobietyDzieci_2011_Polska.csv"
[11] "LudnAfrykaSubsahar_2050_wiek.csv" "LudnEuropaZach_2050_wiek.csv"
[13] "LudnSwiat_2010_2100_KBR_KMR.csv" "LudnSwiat_2010_2100_Obszary.csv"
[15] "LudnSwKrzysz_1791_Wiek.csv" "Material_1.xlsx"
[17] "nupturienci_Jasienica_1861_1880.csv" "nupturienci_Jasienica_1861_1880.xlsx"
[19] "Nupturienci_kawalerowie_wiek_Jasienica_1861_1880.csv" "Nupturienci_panny_wiek_Jasienica_1861_1880.csv"
[21] "Polska_emigranci_3miesiace_2002.csv" "Polska_emigranci_3miesiace_2011.csv"
[23] "R_1_Grupowanie.docx" "R_1_Grupowanie.pdf"
[25] "R_1_Grupowanie.RData" "R_2_LiczbyRelatywne.docx"
[27] "R_2_LiczbyRelatywne.pdf" "R_2_LiczbyRelatywne.RData"
[29] "R_3_TendencjaCentralna_indywidualneDane.docx" "R_3_TendencjaCentralna_indywidualneDane.pdf"
[31] "rs_rocznik_demograficzny_2012.zip"
> # Obliczanie średniej ważonej dla pogrupowanych danych - zmienna skokowa
> (gospodarstwaDomowePolska2002 <- read.csv("GospDom_Polska_2002.csv"))
  liczbaOsob liczbaGospodDom2002
1      1      2486514
2      2      2215940
3      3      1890321
4      4      1568063
5      5       550173
6      6       170372
7      7        83080
> # Z uwagi na małą liczebność (mniej niż 5% ogólnej liczebności) ostatni wariant zmiennej "7 i więcej" zastąpiono wariantem "7".
> attach(gospodarstwaDomowePolska2002)
> (gospodarstwaDomowePolska2002Srednia <- weighted.mean(liczbaOsob,liczbaGospodDom2002))
[1] 2.589811
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
> (gospodarstwaDomoweSzczecin1900 <- read.csv("GospDom_Szczecin_1900.csv"))
```

```
  liczbaOsob liczbaGospodarstw
1         1         2984
2         2         7929
3         3         9763
4         4         9582
5         5         7444
6         6         5190
7         7         2996
8         8         1590
9         9          728
10        10          300
11        11          292
```

```
> # Z uwagi na małą liczebność (mniej niż 5% ogólnej liczebności) ostatni wariant zmiennej "11 i więcej" zastąpiono wariantem "11".
```

```
> detach(gospodarstwaDomowePolska2002)
```

```
> attach(gospodarstwaDomoweSzczecin1900)
```

```
> (gospodarstwaDomoweSzczecin1900Srednia <- weighted.mean(liczbaOsob,liczbaGospodarstw))
```

```
[1] 4.124657
```

```
> detach(gospodarstwaDomoweSzczecin1900)
```

```
> # Utworzenie własnej funkcji obliczającej średnią ważoną dla pogrupowanych danych - zmienna ciągła.
```

```
> # Założenie dla dalszych obliczeń: tabela z danymi składa się z trzech kolumn, to jest "granicy dolnej", "granicy górnej", "liczebności".
```

```
> sredniaPogrupowane <- function(dane,poPrzecinku) {round(sum((dane[,1]+dane[,2])/2*dane[,3])/sum(dane[,3]),poPrzecinku)}
```

```
> sredniaPogrupowane
```

```
function(dane,poPrzecinku) {round(sum((dane[,1]+dane[,2])/2*dane[,3])/sum(dane[,3]),poPrzecinku)}
```

```
> (emigranciPolska2011 <- read.csv("Polska_emigranci_3miesiace_2011.csv"))
```

```
  dolnaGranicaWieku gornaGranicaWieku emigranci2011
1         0         5         86800
2         5        10         74600
3        10        15         65000
4        15        20         64700
5        20        25        179200
6        25        30        382800
7        30        35        343200
8        35        40        221500
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

9	40	45	152300
10	45	50	134300
11	50	55	123000
12	55	60	87100
13	60	65	44900
14	65	70	18300
15	70	75	14400
16	75	80	10600
17	80	85	13700

```
> # Przetestowanie utworzonej funkcji na powyższych danych. Zaokrąglenie wyniku do dwóch miejsc po przecinku (po kropce).
```

```
> emigranciPolska2011Srednia <- sredniaPogrupowane(emigranciPolska2011,2)
```

```
> emigranciPolska2011Srednia
```

```
[1] 33.77
```

```
> (kawalerowieJasienica <- read.csv("Nupturienici_kawalerowie_wiek_Jasienica_1861_1880.csv"))
```

```
  dolnaGranicaWieku  gornaGranicaWieku  liczebnośćNupturientow
```

1	20	23	29
2	23	26	121
3	26	29	116
4	29	32	45
5	32	35	22
6	35	38	16

```
> (kawalerowieJasienicaSrednia <- sredniaPogrupowane(kawalerowieJasienica,2))
```

```
[1] 27.14
```

```
> # Utworzenie własnej funkcji obliczającej dominantę dla pogrupowanych danych.
```

```
> # W obliczeniach przydatna jest funkcja which.max(), która zwraca indeks (pozycję) największej wartości w wektorze danych.
```

```
> # Obliczenia dominanty dla pogrupowanych danych na podstawie wzoru:
```

```
> # dolna granica przedziału z dominantą + (liczebność przedziału z dominantą – liczebność przedziału poprzedzającego przedział z dominantą) / ((liczebność przedziału z dominantą – liczebność przedziału poprzedzającego przedział z dominantą) + (liczebność przedziału z dominantą – liczebność przedziału następnego po przedziale z dominantą)) * rozpiętość przedziału z dominantą.
```

```
> dominantaPogrupowane <- function(dane) {
```

```
+ # pozycja liczebności dominującej - pozNdom
```

```
+ pozNdom <- which.max(dane[,3])
```

```
+ # wzór:
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
+ dane[pozNdom,1]+(dane[pozNdom,3]-dane[pozNdom-1,3])/((dane[pozNdom,3]-dane[pozNdom-1,3])+(dane[pozNdom,3]-dane[pozNdom+1,3]))*(dane[pozNdom,2]-dane[pozNdom,1])
+ }
> dominantaPogrupowane
function(dane) {
# pozycja liczebności dominującej - pozNdom
pozNdom <- which.max(dane[,3])
# wzór:
dane[pozNdom,1]+(dane[pozNdom,3]-dane[pozNdom-1,3])/((dane[pozNdom,3]-dane[pozNdom-1,3])+(dane[pozNdom,3]-dane[pozNdom+1,3]))*(dane[pozNdom,2]-dane[pozNdom,1])
}
> # Przetestowanie nowej funkcji - obliczenie dominanty wieku emigrantów w 2011 roku.
> emigranciPolska2011dominanta <- dominantaPogrupowane(emigranciPolska2011)
> emigranciPolska2011dominanta
[1] 29.18586
> # Obliczenie dominanty wieku kawalerów - nupturientów.
> (kawalerowieJasienicaDominanta <- dominantaPogrupowane(kawalerowieJasienica))
[1] 25.84536
> pannyWiek <- read.csv("Nupturienci_panny_wiek_Jasienica_1861_1880.csv")
> pannyWiek
  dolnaGranicaWieku gornaGranicaWieku liczebnoscNupturientek
1          17          20          29
2          20          23         100
3          23          26         114
4          26          29          63
5          29          32          36
6          32          35          11
7          35          38           5
> # Obliczenia mediany dla pogrupowanych danych na podstawie wzoru:
> # dolna granica przedziału z medianą + (ogólna liczebność podzielona przez dwa – suma liczebności poprzedzających przedział z medianą) / liczebność przedziału z medianą * rozpiętość przedziału z medianą
> # Do utworzenia funkcji wykorzystana będzie:
> # funkcja cumsum() tworząca wektor szeregu skumulowanego
> # funkcja which() zwracająca indeks (pozycję) poszukiwanej lub poszukiwanych wartości z atrybutem TRUE (prawda).
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
> medianaPogrupowane <- function(dane) {
+ # Wartość ogólnej liczebności podzielona przez dwa - Nprzez2
+ Nprzez2 <- sum(dane[,3])/2
+ # Pozycja (wiersz) przedziału z medianą - pozMe
+ pozMe <- which(cumsum(dane[,3])>Nprzez2)[1]
+ # Wzór:
+ dane[pozMe,1]+(Nprzez2-cumsum(dane[,3])[pozMe-1])/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
+ }
> medianaPogrupowane
function(dane) {
# Wartość ogólnej liczebności podzielona przez dwa - Nprzez2
Nprzez2 <- sum(dane[,3])/2
# Pozycja (wiersz) przedziału z medianą - pozMe
pozMe <- which(cumsum(dane[,3])>Nprzez2)[1]
# Wzór:
dane[pozMe,1]+(Nprzez2-cumsum(dane[,3])[pozMe-1])/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
}
> # Przetestowanie
> medianaPogrupowane(pannyWiek)
[1] 24.31579
> # Utworzenie przykładowej ramki (a – dolna granica, b – górna granica, n – liczebność) . Mediana w pierwszym przedziale.
> a <- c(1:3)
> b <- c(2:4)
> n <- c(100,2,2)
> abn <- data.frame(a,b,n)
> abn
  a b  n
1 1 2 100
2 2 3   2
3 3 4   2
# Wprowadzenie do stosowania instrukcji warunkowej typu: if ... else ...
# if(warunek) {prawda} else {fałsz}
< if(2+2 == 4) {"To prawda"} else {"To nieprawda"}
[1] "To prawda"
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
< if(2+2 == 5) {"To prawda"} else {"To nieprawda"}
[1] "To nieprawda"
# Ciąg dalszy wprowadzenia: utworzenie własnej funkcji sprawdzającej, czy dana wartość jest liczbą z użyciem funkcji class()
# Funkcja class może zwrócić następujące wyniki:
> class(15)
[1] "numeric"
> class("dom15")
[1] "character"
> class(TRUE)
[1] "logical"
> czyLiczba <- function(dana) {if (class(dana) == "numeric") {"To jest liczba"} else {"To nie jest liczba"}}
> # Przetestowanie powyższej funkcji
> czyLiczba(10)
[1] "To jest liczba"
> czyLiczba("dom15")
[1] "To nie jest liczba"
> czyLiczba(TRUE)
[1] "To nie jest liczba"
# Modyfikacja funkcji "medianaPogrupowane" z uwzględnieniem występowania mediany w pierwszym przedziale
> medianaPogrupowane <- function(dane) {
+ # Wartość ogólnej liczebności podzielona przez dwa - Nprzez2
+ Nprzez2 <- sum(dane[,3])/2
+ # Pozycja (wiersz) przedziału z medianą - pozMe
+ pozMe <- which(cumsum(dane[,3])>Nprzez2)[1]
+ # Sprawdzenie warunku i wykonanie obliczeń
+ if(Nprzez2<dane[,3][1]) {
+ dane[pozMe,1]+Nprzez2/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
+ } else {
+ dane[pozMe,1]+(Nprzez2-cumsum(dane[,3])[pozMe-1])/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
+ }
+ }
>medianaPogrupowane
function(dane) {
# Wartość ogólnej liczebności podzielona przez dwa - Nprzez2
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

```
Nprzez2 <- sum(dane[,3])/2
# Pozycja (wiersz) przedziału z medianą - pozMe
pozMe <- which(cumsum(dane[,3])>Nprzez2)[1]
# Sprawdzenie warunku i wykonanie obliczeń
if(Nprzez2<dane[,3][1]) {
dane[pozMe,1]+Nprzez2/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
} else {
dane[pozMe,1]+(Nprzez2-cumsum(dane[,3])[pozMe-1])/dane[pozMe,3]*(dane[pozMe,2]-dane[pozMe,1])
}
}
```

```
> medianaPogrupowane(abn)
```

```
[1] 1.52
```

```
> medianaPogrupowane(pannyWiek)
```

```
[1] 24.31579
```

```
(emigranciPolska2002 <- read.csv("Polska_emigranci_3miesiace_2002.csv"))
```

```
  dolnaGranicaWieku gornaGranicaWieku emigranci2002
```

1	0	5	13100
2	5	10	19400
3	10	15	28100
4	15	20	47500
5	20	25	91300
6	25	30	116200
7	30	35	94700
8	35	40	89900
9	40	45	86200
10	45	50	71500
11	50	55	48400
12	55	60	24600
13	60	65	19500
14	65	70	14600
15	70	75	10100
16	75	80	5700
17	80	85	4500

```
> medianaPogrupowane(emigranciPolska2002)
```

dr Dariusz Chojecki, Instytut Historii i Stosunków Międzynarodowych Uniwersytetu Szczecińskiego (szkielet wykładu na podstawie programu R)

4. Podstawowe miary tendencji centralnej – dane pogrupowane

[1] 34.06811

```
> save.image("R_3_TendencjaCentralna_pogrupowaneDane.RData")
```

```
> q()
```