

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

3. Podstawowe miary tendencji centralnej – dane indywidualne

```
> getwd()
[1] "C:/Users/DC/Documents"
> setwd("C:/Users/DC/Desktop/StatDem")
> getwd()
[1] "C:/Users/DC/Desktop/StatDem"
> list.files()
 [1] "R_3_TendencjaCentralna_indywidualneDane.docx" "data" "Dominanta_R.txt"
 [4] "Eo_World_Baza.csv" "Eo_World_bis.xlsx" "Eo_World_Red.xlsx"
 [7] "KobietyDzieci_1980_Polska.csv" "KobietyDzieci_2011_Polska.csv" "LudnAfrykaSubsahar_2050_wiek.csv"
[10] "LudnEuropaZach_2050_wiek.csv" "LudnSwiat_2010_2100_KBR_KMR.csv" "LudnSwiat_2010_2100_Obszary.csv"
[13] "LudnSwKrzyz_1791_Wiek.csv" "Material_1.xlsx" "R_1_Grupowanie.docx"
[16] "R_1_Grupowanie.pdf" "R_1_Grupowanie.RData" "R_2_LiczbyRelatywne.docx"
[19] "R_2_LiczbyRelatywne.pdf" "R_2_LiczbyRelatywne.RData" "R_3_TendencjaCentralna_indywidualneDane.docx"
[22] "rs_rocznik_demograficzny_2012.zip"
> # Wczytanie danych o oczekiwanym dalszym trwaniu życia osób w określonym wieku w państwach:
> ExSwiat <- read.csv("Eo_World_Baza.csv")
> # Wyświetlenie nazw kolumn w powyższym pliku:
> names(ExSwiat)
 [1] "Kontynent" "Panstwo" "Rok" "Plec" "X0" "X5" "X10" "X15" "X20" "X25" "X30" "X35" "X40"
[14] "X45" "X50" "X55" "X60" "X65" "X70" "X75" "X80" "X85" "X90" "X95" "X100"
> # Wyświetlenie nazw kategorii i ich liczebności w kolumnie "Plec":
> attach(ExSwiat)
> table(Plec)
Plec
Female-Femmes Male-Hommes
 173 173
> # Utworzenie podzbioru danych o oczekiwanym przeciętnym trwaniu życia noworodków płci męskiej:
> E0swiatMeczyzni <- subset(X0,Plec == "Male-Hommes")
> # Utworzenie podzbioru danych o oczekiwanym przeciętnym trwaniu życia noworodków płci żeńskiej:
> E0swiatKobiety <- subset(X0,Plec == "Female-Femmes")
> # Wyświetlenie kolejno podzbiorów danych:
> E0swiatMeczyzni
 [1] 74.9 57.2 54.0 55.8 49.2 49.0 68.2 49.8 58.3 43.4 52.9 48.7 51.3 45.7 69.4 47.6 73.2 49.4 72.5 61.3 68.4 47.5 53.5 52.5 42.2 72.3 45.8 76.5 70.0 69.9 70.0 76.9
[33] 69.9 78.3 76.3 76.8 76.0 73.8 69.2 67.7 66.6 74.6 61.4 69.7 76.5 72.8 72.0 63.4 73.4 74.1 68.2 69.9 66.9 68.3 79.0 75.4 72.5 59.8 69.4 75.8 70.7 72.1 75.1 69.7
[65] 69.0 68.0 72.4 70.7 43.0 70.2 71.0 73.1 65.4 65.7 76.6 69.6 79.7 78.3 65.6 69.3 62.6 61.0 71.1 58.0 78.7 79.6 71.6 61.9 73.3 65.2 55.0 71.5 72.5 64.3 64.3 63.6
[97] 70.2 72.2 63.6 63.1 77.9 77.0 79.3 70.5 69.9 71.5 76.7 70.2 60.2 79.4 72.5 77.4 64.7 77.0 71.3 69.9 74.2 76.5 68.6 76.8 76.5 77.4 76.9 78.5 77.7 70.1 79.6 75.1
[129] 73.6 78.7 68.3 67.5 77.6 77.7 78.5 78.6 71.5 75.8 65.3 69.7 62.8 78.0 71.1 71.3 75.8 78.5 79.4 79.8 72.1 62.5 77.4 68.5 79.3 68.0 63.8 73.0 75.9 58.9 67.0 66.5
[161] 55.2 71.8 78.4 67.0 74.5 66.3 53.7 71.5 60.6 67.3 61.7 65.6 73.1
> E0swiatKobiety
```

3. Podstawowe miary tendencji centralnej – dane indywidualne

```
[1] 76.6 61.3 66.0 57.5 52.7 52.0 70.9 51.8 62.0 46.2 60.4 56.3 53.2 48.3 76.6 50.2 80.9 53.3 79.2 66.5 78.0 49.4 57.2 55.5 43.1 76.2 50.3 81.1 76.0 76.4 74.1 82.3  
[33] 78.5 83.0 83.8 81.8 80.0 78.2 75.5 73.7 71.6 81.5 67.2 75.2 82.9 77.5 79.3 68.9 78.7 81.5 70.7 75.7 72.9 73.7 77.4 80.4 80.0 63.2 77.0 81.2 77.5 78.0 80.8 73.9  
[65] 74.3 73.7 79.7 76.6 43.0 76.6 76.1 77.3 67.9 66.9 79.8 73.3 85.9 81.9 72.7 79.0 64.2 64.5 73.1 59.0 82.5 86.4 74.4 72.4 77.2 73.2 63.0 76.2 74.2 71.8 68.3 64.5  
[97] 72.9 75.7 67.6 66.7 78.1 83.8 84.1 75.3 77.6 76.1 78.8 75.6 62.0 85.1 77.3 82.9 76.4 82.7 76.7 77.1 80.1 80.8 79.2 82.3 83.1 84.3 82.3 83.3 82.8 77.9 81.3 80.3  
[129] 79.9 84.0 78.1 78.6 82.7 82.2 82.7 83.1 80.1 81.8 73.4 77.1 74.7 84.6 76.4 78.7 82.3 84.6 83.4 84.4 76.3 74.3 81.6 76.2 83.9 74.0 67.8 78.2 82.2 63.1 70.6 67.5  
[161] 57.1 80.3 82.4 76.0 79.9 72.1 54.8 74.2 61.6 73.0 65.1 69.0 75.5
```

> # Wyświetlenie rozkładów zmiennej z użyciem wykresu typu łądyga-liście:

> stem(E0swiatMezczyzni)

The decimal point is at the |

```
42 | 204  
44 | 78  
46 | 56  
48 | 70248  
50 | 3  
52 | 5957  
54 | 0028  
56 | 2  
58 | 0398  
60 | 2603479  
62 | 56814668  
64 | 337234667  
66 | 356900357  
68 | 00223345602344677799999  
70 | 00122257701133555568  
72 | 0112345555801123468  
74 | 125691148889  
76 | 035555678899004446779  
78 | 0334555677033446678
```

> stem(E0swiatKobiety)

The decimal point is at the |

```
42 | 01  
44 |  
46 | 2  
48 | 34  
50 | 238  
52 | 0723
```

3. Podstawowe miary tendencji centralnej – dane indywidualne

```
54 | 85
56 | 3125
58 | 0
60 | 436
62 | 00012
64 | 2551
66 | 057925689
68 | 390
70 | 67968
72 | 14799012347779
74 | 012233472355677
76 | 001122234446666701123345569
78 | 0011225677802237899
80 | 0011334889123556889
82 | 2233334577789901134889
84 | 01346619
86 | 4
```

```
> # Obliczenie średniej arytmetycznej dla podzbiorów:
> (E0swiatMezcyzniSrednia <- mean(E0swiatMezcyzni))
[1] 68.58902
> (E0swiatKobietySrednia <- mean(E0swiatKobiety))
[1] 73.71098
> # Obliczenie mediany dla podzbiorów:
> (E0swiatMezcyzniMediana <- median(E0swiatMezcyzni))
[1] 70.2
> (E0swiatKobietyMediana <- median(E0swiatKobiety))
[1] 76.4
> # Obliczenie kwartyli dla podzbiorów:
> (E0swiatMezcyzniKwartyle <- quantile(E0swiatMezcyzni,c(0.25,0.5,0.75)))
25% 50% 75%
64.7 70.2 75.8
> (E0swiatKobietyKwartyle <- quantile(E0swiatKobiety,c(0.25,0.5,0.75)))
25% 50% 75%
70.6 76.4 80.4
> # Obliczenie rozstępu międzykwartylowego dla podzbiorów:
> (E0swiatMezcyzniRozstepMiedzykwartylowy <- IQR(E0swiatMezcyzni))
[1] 11.1
> (E0swiatKobietyRozstepMiedzykwartylowy <- IQR(E0swiatKobiety))
```

3. Podstawowe miary tendencji centralnej – dane indywidualne

[1] 9.8

> # Obliczenie decyli dla podzbiorów:

> (E0swiatMezczyzniDecyle <- quantile(E0swiatMezczyzni,c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9)))

10% 20% 30% 40% 50% 60% 70% 80% 90%

54.20 62.68 66.56 69.16 70.20 72.10 73.92 76.50 78.24

> (E0swiatKobietyDecyle <- quantile(E0swiatKobiety,c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9)))

10% 20% 30% 40% 50% 60% 70% 80% 90%

57.80 67.02 72.90 74.64 76.40 77.92 79.84 81.56 82.90

> # Obliczenie wartości najczęściej występującej dla podzbiorów (utworzenie własnej funkcji):

> modalna <- function(dane) {names(table(dane))[max(table(dane)) == table(dane)]}

> modalna

function(dane) {names(table(dane))[max(table(dane)) == table(dane)]}

> (E0swiatMezczyzniModa <- modalna(E0swiatMezczyzni))

[1] "69.9"

> (E0swiatKobietModa <- modalna(E0swiatKobiety))

[1] "76.6" "82.3"

> # UWAGA: przy zmiennej ciągłej - jak w powyższym przypadku - wartość najczęściej występująca (dominanta - moda) powinna być obliczana z danych pogrupowanych (patrz następny wykład).

> # Rozbicie na elementy składowe powyższej funkcji na przykładzie danych dla noworodków płci żeńskiej:

> table(E0swiatKobiety)

E0swiatKobiety

```
43 43.1 46.2 48.3 49.4 50.2 50.3 51.8 52 52.7 53.2 53.3 54.8 55.5 56.3 57.1 57.2 57.5 59 60.4 61.3 61.6 62 63 63.1 63.2 64.2 64.5 65.1 66 66.5 66.7 66.9
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1
67.2 67.5 67.6 67.8 67.9 68.3 68.9 69 70.6 70.7 70.9 71.6 71.8 72.1 72.4 72.7 72.9 73 73.1 73.2 73.3 73.4 73.7 73.9 74 74.1 74.2 74.3 74.4 74.7 75.2 75.3 75.5
1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2
75.6 75.7 76 76.1 76.2 76.3 76.4 76.6 76.7 77 77.1 77.2 77.3 77.4 77.5 77.6 77.9 78 78.1 78.2 78.5 78.6 78.7 78.8 79 79.2 79.3 79.7 79.8 79.9 80 80.1 80.3
1 2 2 2 3 1 3 4 1 1 2 1 2 1 2 1 1 2 2 2 1 1 2 1 1 1 2 2 2 2
80.4 80.8 80.9 81.1 81.2 81.3 81.5 81.6 81.8 81.9 82.2 82.3 82.4 82.5 82.7 82.8 82.9 83 83.1 83.3 83.4 83.8 83.9 84 84.1 84.3 84.4 84.6 85.1 85.9 86.4
1 2 1 1 1 1 2 1 2 1 2 4 1 1 3 1 2 1 2 1 1 2 1 1 1 1 1 2 1 1 1
```

> max(table(E0swiatKobiety))

[1] 4

> max(table(E0swiatKobiety)) == table(E0swiatKobiety)

E0swiatKobiety

```
43 43.1 46.2 48.3 49.4 50.2 50.3 51.8 52 52.7 53.2 53.3 54.8 55.5 56.3 57.1 57.2 57.5 59 60.4 61.3 61.6 62 63 63.1 63.2 64.2
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE FALSE FALSE FALSE FALSE FALSE
64.5 65.1 66 66.5 66.7 66.9 67.2 67.5 67.6 67.8 67.9 68.3 68.9 69 70.6 70.7 70.9 71.6 71.8 72.1 72.4 72.7 72.9 73 73.1 73.2 73.3
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE FALSE FALSE FALSE FALSE FALSE
```

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

3. Podstawowe miary tendencji centralnej – dane indywidualne

```
73.4 73.7 73.9 74 74.1 74.2 74.3 74.4 74.7 75.2 75.3 75.5 75.6 75.7 76 76.1 76.2 76.3 76.4 76.6 76.7 77 77.1 77.2 77.3 77.4 77.5
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
FALSE FALSE FALSE FALSE FALSE FALSE
77.6 77.9 78 78.1 78.2 78.5 78.6 78.7 78.8 79 79.2 79.3 79.7 79.8 79.9 80 80.1 80.3 80.4 80.8 80.9 81.1 81.2 81.3 81.5 81.6 81.8
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE FALSE FALSE FALSE FALSE FALSE
81.9 82.2 82.3 82.4 82.5 82.7 82.8 82.9 83 83.1 83.3 83.4 83.8 83.9 84 84.1 84.3 84.4 84.6 85.1 85.9 86.4
FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
> names(table(E0swiatKobiety))
[1] "43" "43.1" "46.2" "48.3" "49.4" "50.2" "50.3" "51.8" "52" "52.7" "53.2" "53.3" "54.8" "55.5" "56.3" "57.1" "57.2" "57.5" "59" "60.4" "61.3" "61.6" "62"
[24] "63" "63.1" "63.2" "64.2" "64.5" "65.1" "66" "66.5" "66.7" "66.9" "67.2" "67.5" "67.6" "67.8" "67.9" "68.3" "68.9" "69" "70.6" "70.7" "70.9" "71.6" "71.8"
[47] "72.1" "72.4" "72.7" "72.9" "73" "73.1" "73.2" "73.3" "73.4" "73.7" "73.9" "74" "74.1" "74.2" "74.3" "74.4" "74.7" "75.2" "75.3" "75.5" "75.6" "75.7" "76"
[70] "76.1" "76.2" "76.3" "76.4" "76.6" "76.7" "77" "77.1" "77.2" "77.3" "77.4" "77.5" "77.6" "77.9" "78" "78.1" "78.2" "78.5" "78.6" "78.7" "78.8" "79" "79.2"
[93] "79.3" "79.7" "79.8" "79.9" "80" "80.1" "80.3" "80.4" "80.8" "80.9" "81.1" "81.2" "81.3" "81.5" "81.6" "81.8" "81.9" "82.2" "82.3" "82.4" "82.5" "82.7" "82.8"
[116] "82.9" "83" "83.1" "83.3" "83.4" "83.8" "83.9" "84" "84.1" "84.3" "84.4" "84.6" "85.1" "85.9" "86.4"
> # W funkcji w nawiasach kwadratowych podawana jest pozycja elementu w wektorze danych z wartością logiczną TRUE. W naszym wypadku:
> names(table(E0swiatKobiety))[74]
[1] "76.6"
> names(table(E0swiatKobiety))[111]
[1] "82.3"
> save.image("R_3_TendencjaCentralna_indywidualneDane.RData")
> q()
```