

1. Grupowanie

```
> # Zglądamy na www.r-project.org i/lub pobieramy program, na przykład z http://r.meteo.uni.wroc.pl/
> # Literatura:
> # - John M. Quick, Analiza statystyczna w środowisku R dla początkujących, Gliwice 2012, ss. 280.
> # - Łukasz Komsta, Wprowadzenie do środowiska R, ss. 39, 21 sierpnia 2004 (www.cran.r-project.org/doc/contrib/Komsta-Wprowadzenie.pdf).
> # - Przemysław Biecek, Przewodnik po pakiecie R, wyd. 2, Wrocław 2011, ss. 409 (część w sieci: www.biecek.pl/R/Rwydanie2.pdf).
> # - (red.) Marek Walesiak, Eugeniusz Gatnar, Statystyczna analiza danych z wykorzystaniem programu R, Warszawa 2012, ss. 468.
> # Sprawdzenie ścieżki do katalogu roboczego.
> getwd()
[1] "C:/Users/DC/Documents"
> # Ustawienie ścieżki do własnego katalogu roboczego utworzonego na pulpicie o nazwie StatDem.
> setwd("C:/Users/DC/Desktop/StatDem/")
> getwd()
[1] "C:/Users/DC/Desktop/StatDem"
> # Utworzenie wektora danych katagorycznych i przypisanie mu nazwy: pochodzenieSzczecin.
> pochodzenieSzczecin <- c("t","n","t","n","n","n","t","n","t","t","n","t","t","n","n","n","t","t","n","n","t","n","n","n","t","n","n")
> # Wyświetlenie wektora (nazywanego w programowaniu zmienną – w tym wypadku w statystyce mowa o zbiorowości).
> pochodzenieSzczecin
[1] "t" "n" "t" "n" "n" "n" "t" "n" "t" "t" "n" "t" "t" "n" "n" "n" "t" "t" "n" "n" "t" "n" "n" "n" "t" "n" "n"
> # Wyznaczenie liczebności (długość łańcucha danych).
> pochodzenieSzczecinLiczebnosć <- length(pochodzenieSzczecin)
> pochodzenieSzczecinLiczebnosć
[1] 30
> # Wyświetlenie fatora, tu: podanie poszczególnych wartości elementów zbioru i ich kategorii uporządkowanych alfabetycznie (Levels).
> pochodzenieSzczecinFactor <- factor(pochodzenieSzczecin)
> pochodzenieSzczecinFactor
[1] t n t n n n t n t t n t t n n n t t n t t n n n t n n
Levels: n t
> # Wyświetlenie wyników grupowania prostego (liczebności poszczególnych kategorii uporządkowanych alfabetycznie)
> pochodzenieSzczecinGrupowanie <- table(pochodzenieSzczecin)
> pochodzenieSzczecinGrupowanie
pochodzenieSzczecin
 n t
17 13
> # Utworzenie nowego wektora – zmiennej i przypisanie mu nazwy: rodzenstwo.
```

1. Grupowanie

```
> rodzenstwo <- c(3,1,0,1,2,0,0,3,0,2,1,0,1,2,0,1,1,0,0,4,1,2,2,2,1,1,3,0,3,4)
> rodzenstwo
[1] 3 1 0 1 2 0 0 3 0 2 1 0 1 2 0 1 1 0 0 4 1 2 2 2 1 1 3 0 3 4
> rodzenstwoLiczebność <- length(rodzenstwo)
> rodzenstwoLiczebność
[1] 30
> # Wyświetlenie faktora dla powyższej zmiennej, tu: podanie poszczególnych wartości elementów zbioru i ich wariantów uporządkowanych od najmniejszego do największego (Levels).
> rodzenstwoFactor <- factor(rodzenstwo)
> rodzenstwoFactor
[1] 3 1 0 1 2 0 0 3 0 2 1 0 1 2 0 1 1 0 0 4 1 2 2 2 1 1 3 0 3 4
Levels: 0 1 2 3 4
> rodzenstwoGrupowanie <- table(rodzenstwo)
> rodzenstwoGrupowanie
rodzenstwo
0 1 2 3 4
9 9 6 4 2
> # Grupowanie złożone (jednocześnie dwie zmienne: rodzeństwo i pochodzenieSzczecin)
> rodzenstwoPochodzenieSzczecinGrupowanie <- table(rodzenstwo,pochodzenieSzczecin)
> rodzenstwoPochodzenieSzczecinGrupowanie
      pochodzenieSzczecin
rodzenstwo n t
      0 3 6
      1 6 3
      2 4 2
      3 3 1
      4 1 1
> # Utworzenie ramki danych (kolumny danych indywidualnych).
> rodzenstwoPochodzenieSzczecinRamka <- data.frame(rodzenstwo,pochodzenieSzczecin)
> # Wyświetlenie ramki.
> rodzenstwoPochodzenieSzczecinRamka
  rodzenstwo pochodzenieSzczecin
1          3                   t
2          1                   n
```

1. Grupowanie

3	0	t
4	1	n
5	2	n
6	0	n
7	0	t
8	3	n
9	0	t
10	2	t
11	1	n
12	0	t
13	1	t
14	2	n
15	0	n
16	1	n
17	1	t
18	0	t
19	0	n
20	4	t
21	1	t
22	2	n
23	2	n
24	2	t
25	1	n
26	1	n
27	3	n
28	0	t
29	3	n
30	4	n

> # Wyświetlenie wyników grupowania na podstawie odwołań do ramki z użyciem znaku specjalnego – \$ – wskazującego drogę do kolumn ramki (wektorów).

> rodzenstwoPochodzenieSzczecinGrupowanieRamka <-

table(rodzenstwoPochodzenieSzczecinRamka\$rodzenstwo,rodzenstwoPochodzenieSzczecinRamka\$pochodzenieSzczecin)

> rodzenstwoPochodzenieSzczecinGrupowanieRamka

n t

0 3 6

1. Grupowanie

```
1 6 3
```

```
2 4 2
```

```
3 3 1
```

```
4 1 1
```

```
> Utworzenie nowej zmiennej: czasDojazduH (w godzinach).
```

```
> czasDojazduH <- c(0.4,1,0.5,0.7,0.3,1.3,0.1,1.8,0.2,1.2,1.9,1,1.3,1.6,0.9,0.8,0.2,0.1,0.6,0.5,0.7,0.5,1.2,1.4,0.3,0.5,1,0.2,1.5,1.2)
```

```
> czasDojazduH
```

```
[1] 0.4 1.0 0.5 0.7 0.3 1.3 0.1 1.8 0.2 1.2 1.9 1.0 1.3 1.6 0.9 0.8 0.2 0.1 0.6 0.5 0.7 0.5 1.2 1.4 0.3 0.5 1.0 0.2 1.5 1.2
```

```
> czasDojazduHliczebność <- length(czasDojazduH)
```

```
> czasDojazduHliczebność
```

```
[1] 30
```

```
> czasDojazduHfactor <- factor(czasDojazduH)
```

```
> czasDojazduHfactor
```

```
[1] 0.4 1 0.5 0.7 0.3 1.3 0.1 1.8 0.2 1.2 1.9 1 1.3 1.6 0.9 0.8 0.2 0.1 0.6 0.5 0.7 0.5 1.2 1.4 0.3 0.5 1 0.2 1.5 1.2
```

```
Levels: 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.2 1.3 1.4 1.5 1.6 1.8 1.9
```

```
> # Użycie funkcji cut() do wydzielenia grup nazywanych dawniej przedziałami klasowymi i określenie ich granic dolnych i górnych wektorem c (standardowo przedziały domknięte z góry).
```

```
> czasDojazduHprzedziały <- cut(czasDojazduH,c(0,0.5,1,1.5,2))
```

```
> czasDojazduHprzedziały
```

```
[1] (0,0.5] (0.5,1] (0,0.5] (0.5,1] (0,0.5] (1,1.5] (0,0.5] (1.5,2] (0,0.5] (1,1.5] (1.5,2] (0.5,1] (1,1.5] (1.5,2] (0.5,1] (0.5,1] (0,0.5] (0,0.5] (0.5,1] (0,0.5]
```

```
[21] (0.5,1] (0,0.5] (1,1.5] (1,1.5] (0,0.5] (0,0.5] (0.5,1] (0,0.5] (1,1.5] (1,1.5]
```

```
Levels: (0,0.5] (0.5,1] (1,1.5] (1.5,2]
```

```
> # Pierwszy zapis oznacza, że osoba miała czas dojazdu na uczelnię nie dłuższy jak pół godziny (dosłownie 30 i mniej minut), drugi zaś – więcej od 30 minut a do godziny włącznie.
```

```
> # Wykonanie grupowania statystycznego na podstawie przynależności danych indywidualnych do poszczególnych, ogólniejszych grup.
```

```
> czasDojazduHprzedziałyGrupowanie <- table(czasDojazduHprzedziały)
```

```
> czasDojazduHprzedziałyGrupowanie
```

```
czasDojazduHprzedziały
```

```
(0,0.5] (0.5,1] (1,1.5] (1.5,2]
```

```
12 8 7 3
```

```
> # Wyświetlenie ramki dla trzech zmiennych: rodzeństwo, pochodzenieSzczecin, czasDojazduH.
```

```
> rodzeństwoPochodzenieSzczecinCzasDojazduHRamka <- data.frame(rodzeństwo,pochodzenieSzczecin,czasDojazduH)
```

```
> rodzeństwoPochodzenieSzczecinCzasDojazduHRamka
```

```
rodzeństwo pochodzenieSzczecin czasDojazduH
```

1. Grupowanie

1	3	t	0.4
2	1	n	1.0
3	0	t	0.5
4	1	n	0.7
5	2	n	0.3
6	0	n	1.3
7	0	t	0.1
8	3	n	1.8
9	0	t	0.2
10	2	t	1.2
11	1	n	1.9
12	0	t	1.0
13	1	t	1.3
14	2	n	1.6
15	0	n	0.9
16	1	n	0.8
17	1	t	0.2
18	0	t	0.1
19	0	n	0.6
20	4	t	0.5
21	1	t	0.7
22	2	n	0.5
23	2	n	1.2
24	2	t	1.4
25	1	n	0.3
26	1	n	0.5
27	3	n	1.0
28	0	t	0.2
29	3	n	1.5
30	4	n	1.2

```
<# Zapisanie wyników prac – zmiennych w pliku o nazwie: R_1_Grupowanie.RData (ścieżka do katalogu określona na samym początku prac).
```

```
> save.image("R_1_Grupowanie.RData")
```

```
<# Wyjście z programu.
```

```
> q()
```