

Laboratorium Analizy Procesów Ucznienia.

Data wykonania ćwiczenia:

23.02.2024

Rok studiów:

1

Semestr:

1

Grupa studencka:

1b

Grupa laboratoryjna:

-

Ćwiczenie nr

1

Temat: Podstawy języka R.

Osoby wykonujące ćwiczenia:

1. Gracjan Wackermann

Katedra Informatyki i Automatyki

1. Cel ćwiczenia:

Celem ćwiczenia było nabycie podstawowej znajomości języka R rozwiązując zadanie tworzenia i wyświetlenia ramki danych odpowiednio do określonego wariantu.

2. Zadanie do wykonania:

Zadanie dotyczy tworzenia danych, które będą wykorzystywane na kolejnych zajęciach w celu podejmowania decyzji przy kupowaniu urządzeń RTV AGD. Sprawozdanie sporządzić zgodnie ze wzorem i odesłać przez system e-uczelnia.

Pliki w postaci:

1. Plik .R
2. Wyniki z konsoli (dowolny plik tekstowy)
3. Plik .csv

- Wariant nr. 3 -

- (a) Do zmiennej `a` podstaw wartość wyrażenia $5/4^3$. Do zmiennej `b` podstaw podwójną wartość zmiennej `a`. Wywołaj funkcję sprawdzającą, która z wartości zmiennych jest mniejsza.

```
a <- 5/4^3
b <- 2 * a
if (a < b) {
  print("a jest mniejsze")
} else {
  print("a jest większe lub równe")
}
```

- (b) Uruchom i poczytaj dokumentację dla funkcji `min()`.

```
?min
```

- (c) Stwórz wektor `a` zawierający liczby od 50 do 75. Policz średnią kwadratów liczb zawartych w wektorze.

```
a <- 50:75
srednia_kwadratow <- mean(a^2)
print(srednia_kwadratow)
```

- (d) Wyświetl wszystkie funkcje zawierające frazę `min` w swojej nazwie.

```
apropos("min")
```

- (e) Ustaw dowolny katalog roboczy. Następnie stwórz zmienną `a` zawierającą łańcuch znaków “aparat z wymienną optyką”. Zapisz zmienną `a` z obszaru roboczego do pliku w katalogu roboczym. Następnie usuń zmienną `a`. Sprawdź wartość zmiennej `a` (powinno jej brakować). Na końcu wczytaj plik ze zmienną `a` i sprawdź jej wartość.

```
setwd("G:/Mój dysk/InformatykaMGR 2024 2025/3. Lista
Przedmiotów - 1 semestr/7. APU [E] Analiza procesów uczenia/1.
Laboratoria/1. 23.02 - 1 sprawko/Sprawko1Gracjan/skrypt")
a <- "aparat z wymienną optyką"
save(a, file="SkryptPodpunktE.Rdata")
rm(a)
print(a)
load("SkryptPodpunktE.Rdata")
print(a)
```

- (f) Zainstaluj i załaduj pakiet `gridExtra`, który umożliwia m.in. ładną wizualizację danych tabelarycznych. Następnie przy pomocy dokumentacji pakietu znajdź funkcję do wizualizacji danych tabelarycznych. Użyj jej na pierwszych 10 wierszach zbioru danych `Seatbelts`.

```
install.packages("gridExtra")
library(gridExtra)
grid.table(head(Seatbelts, 10))
```

(g) Stwórz wektor zawierający ciąg liczb 1000, 995, 990, ... 800.

```
a <- seq(1000, 800, by=-5)
```

(h) Stwórz wektora a z liczbami od 29 do 5 oraz wektor b z liczbami od 21 do 33. Utwórz nowy wektory d będący połączeniem wektora b i a (w takiej kolejności). Wyświetl go.

```
a <- 29:5  
b <- 21:33  
d <- c(b, a)  
print(d)
```

(i) Stwórz wektor nazwa zawierający nazwy 10 aparatów z wyminną optyką. Potem stwórz wektory rozdzielczość, zakres_czułości, cena, liczba_opinii zawierające kolejno dane 10 aparatów. Następnie stwórz ramkę danych aparaty złożoną z wektorów nazwa, rozdzielczość, zakres_czułości, cena, liczba_opinii. Wylicz średnią cenę aparatów.

```
nazwa <- c("aparat1", "aparat2", "aparat3", "aparat4",  
"aparat5", "aparat6", "aparat7", "aparat8", "aparat9",  
"aparat10")  
rozdzielczosc <- c(1, 2, 3, 4, 5, 10, 15, 20, 25, 30)  
zakres_czu_losci <- c(100, 150, 200, 250, 300, 10, 20, 30, 40,  
50)  
cena <- c(500, 600, 700, 800, 900, 1000, 100, 200, 300, 400)  
liczba_opinii <- c(50, 60, 70, 80, 90, 100, 2, 3, 4, 5)  
aparaty <- data.frame(nazwa, rozdzielczosc, zakres_czu_losci,  
cena, liczba_opinii)  
mean(aparaty$cena)
```

- (j) Do stworzonej w poprzednim zadaniu ramki danych aparatów dodaj wpis zawierający dane nowego aparatu. Wylicz średnią ceny ponownie.

```
nowy_aparat <- data.frame(nazwa = "nowy_aparat", rozdzielczosc = 10, zakres_czu_losci = 100, cena = 1000, liczba_opinii = 10000)
aparaty <- rbind(aparaty, nowy_aparat)
mean(aparaty$cena)
```

- (k) Korzystając z ramki danych aparaty dodaj nową kolumnę określając ocenę klientów. Wpisz do kolumny odpowiednio oceny w skali od 0 do 5 krok 0.5. Dodana kolumna powinna się automatycznie przekonwertować do cech jakościowych (tzw. factors). Wylicz średnią ceny każdej oceny.

```
aparaty$ocena_klientow <- seq(0, 5, by=0.5)
aparaty$ocena_klientow <- as.factor(aparaty$ocena_klientow)
srednia_cena_ocena <- tapply(aparaty$cena,
  aparaty$ocena_klientow, mean)
print(srednia_cena_ocena)
```

- (l) Do ramki danych aparaty dodaj kolejne 4 aparaty. Narysuj na wykresie słupkowym liczebność reprezentantów każdej z ocen klientów.

```
nowe_aparaty <- data.frame(nazwa = c("aparat11", "aparat12",
  "aparat13", "aparat14"),
  rozdzielczosc = c(11, 12, 13, 14),
  zakres_czu_losci = c(100, 110, 120,
  130),
  cena = c(600, 700, 800, 900),
  liczba_opinii = c(90, 100, 110,
  120))
library(ggplot2)
ggplot(aparaty, aes(x=ocena_klientow)) +
  geom_bar(stat="count")
```

- (m) Wykorzystując ramkę danych aparaty pokaż procentowy udział każdej oceny przy pomocy wykresu kołowego oraz wachlarzowego.

```
pie(table(aparaty$ocena_klientow))  
plot(table(aparaty$ocena_klientow))
```

- (n) Do ramki danych aparaty dodaj nową kolumnę status_opinii z wartościami: "nie ma", "mniej 50 opinii", "50-100 opinii", "więcej 100 opinii" w zależności od liczby opinii. Zamień dodaną kolumnę na cechy jakościowe. Następnie przy pomocy wykresu kołowego wyrysuj procentowy udział aparatów o konkretnym statusie opinii.

```
aparaty$status_opinii <- cut(aparaty$liczba_opinii, breaks =  
c(-Inf, 0, 50, 100, Inf), labels = c("nie ma", "mniej 50  
opinii", "50-100 opinii", "więcej 100 opinii"))  
aparaty$status_opinii <- as.factor(aparaty$status_opinii)  
pie(table(aparaty$status_opinii))
```

- (o) Wykorzystując ramkę danych aparaty stwórz zdanie o każdym z aparatów postaci: nazwa + " ma ocenę klientów " + ocena_klientów + " bo ma liczbę opinii " + liczba_opinii. Plus oznacza konkatenację łańcuchów i wartości.

```
wyniki <- paste(aparaty$nazwa, " ma ocenę klientów ",  
aparaty$ocena_klientow,  
" bo ma liczbę opinii ", aparaty$liczba_opinii)  
print(wyniki)
```

- (p) Zachować ramkę danych w pliku .csv. Załadować ramkę danych z pliku .csv
Dane (15 aparatów) pobrać ze strony <http://www.euro.com.pl>

```
write.csv(aparaty, file = "aparaty.csv", row.names = FALSE)  
read.csv("aparaty.csv")
```

3. Wnioski:

- Język R pozwala na kompleksową analizę danych, w tym na statystyki opisowe, wizualizacje, modelowanie i uczenie maszynowe.
- Wykonanie ćwiczenia pozwoliło na zapoznanie się z podstawową strukturą języka i funkcjami jego stosowania.
- Przygotowany program posłuży nam dalej w kolejnych ćwiczeniach.