Nauka dla społeczeństwa

Katedra i Zakład Informatyki i Statystyki

dr Justyna Marcinkowska

data przygotowania: 2024-05-05

Zalecenia dotyczące warsztatów – część 5 – kod w R cd.:

*Ilustracja graficzna*

*w publikacji naukowej*

*(język R)*

# Ćwiczenia praktyczne na warsztatach:

• Omówienie podstawowej składni języka R,

• Przygotowanie do pracy z danymi: import, tworzenie data frame,

• Wstęp do analizy danych i graficznej wizualizacji:

* przetwarzanie i analiza danych (biblioteka dplyr),
* tworzenie bogatych graficznie, typowych i nietypowych wykresów na podstawie warstw (biblioteka ggplot2)

• Przygotowanie wykresów pod wymogi czasopism naukowych, eksport wykresów wysokiej rozdzielczości



Wstęp do R cd.

# Ładujemy potrzebne nam pakiety:

Uruchom poniższą wstawkę

# Sprawdź, czy pakiety są zainstalowane
if (!requireNamespace("rstatix", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("rstatix")
}
if (!requireNamespace("writexl", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("writexl")
}
if (!requireNamespace("dplyr", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("dplyr")
}
if (!requireNamespace("ggplot2", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("ggplot2")
}
if (!requireNamespace("knitr", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("knitr")
}
if (!requireNamespace("readxl", quietly = TRUE)) {
 # Jeśli nie jest, zainstaluj pakiet
 install.packages("readxl")
}
library(rstatix)
library(writexl)
library(dplyr)
library(ggplot2)
library(readxl)
library(knitr)

# zad 4.

Ładujemy bazę sodium.xls (arkusz ***data***) przy pomocy funkcji ***read\_excel()*** z pakietu readxl

sodium<-read\_excel("sodium.xlsx",sheet="data")

opisówka… podaj przykłady…

max(sodium$age)

[1] 107

Aby uprościć nasze zapisy - najwygodniej używa się funkcji z pakietu dplyr

Potokowanie: operator **%>%** pochodzi z pakietu automatycznie ładowanego przez **dplyr**.

# Przykład - użycie potoku:

Sprawdźmy, jakie kolumny ma ta baza?

sodium %>%
 colnames()

[1] "nr" "sex"
[3] "age" "medical\_transport"
[5] "hospitalization" "Stay\_in\_the\_Emergency\_Department"
[7] "department" "sodium\_concentration"

# zad 5.

Sprawdź (korzystając z potokowania):

1. ile jest kolumn (funkcja ncol()) w baziesodium?
2. ile jest wierszy?

#wpisz kod:

# zad 6.

korzystając z funkcji filter() (z pakietu dplyr) wyświetl:

ile jest w bazie sodium wierszy (osób), które były hospitalizowane?

#wpisz kod:

# zad 7.

W zwykłym działaniu na wektorach mieliśmy przykład::

#wyświetlamy BMI dla Michałów: baza\_grupy[imiona==“Michał”,“BMI”]

#wszystkie kolumny dla wiersazy opisanych warunkiem: wzrostu>185 baza\_grupy[baza\_grupy$wzrost>185,]

zadanie: zrób to samo przy pomocy pakietu dyplyr czyli funkcji filer i select

#wpisz kod:

# zad 8

Zlicz (funkcja ***summarise(n())***) ile osób było hospitalizowanych na poszczególnych oddziałach. Wykorzystaj w potokowaniu funkcję ***group\_by()*** oraz ***filter()***

#wpisz kod:

# Przykład - tworzymy ramkę z wynikami

wynik\_grupujacy <- sodium %>%
 group\_by(department)%>%
 filter(hospitalization==1)%>%
 summarise(
 liczność=n(),
 'średni wiek'= round(mean(age),2),
 'maksymalny poziom sodu'=max(sodium\_concentration),
 'średnia długość pobytu na '=
 round(mean(Stay\_in\_the\_Emergency\_Department),2)
 )
print(wynik\_grupujacy)

# A tibble: 4 × 5
 department liczność `średni wiek` `maksymalny poziom sodu`
 <chr> <int> <dbl> <dbl>
1 Chir 448 58.6 165
2 IOM 33 64.9 157
3 Interna 931 68.3 165
4 Neuro 291 60.2 163
# ℹ 1 more variable: `średnia długość pobytu na ` <dbl>

# Wyświetlenie ładnej tabeli
kable(wynik\_grupujacy, "html")

| department | liczność | średni wiek | maksymalny poziom sodu | średnia długość pobytu na |
| --- | --- | --- | --- | --- |
| Chir | 448 | 58.56 | 165 | 3.55 |
| IOM | 33 | 64.94 | 157 | 2.63 |
| Interna | 931 | 68.30 | 165 | 5.47 |
| Neuro | 291 | 60.22 | 163 | 3.94 |

# zad 9.

Zliczamy pacjentów o skrajnie niskim (<129) lub skrajnie wysokim (>150) stężeniu sodu.

Wskazówka: lub oznacza or, czyli sumę zbiorów

#wpisz kod:

# przykład - przekodowanie - funlcja cut():

Utwórz nową kolumnę sodium\_category i przekoduj w niej wartości poziomu sodu na 5 grup wg. punktów odcięcia: 129, 135, 144,150,170 o nazwach: “very low values”, “below standard”, “standard”, “above normal”,“very high values”

# Utwórz przedziały
breaks <- c( 0, 129, 135, 144, 150, 170, Inf)

# Etykiety dla przedziałów
labels <- c("very low values", "below standard", "standard", "above normal", "very high values", "extremely high values")

# Utwórz nową kolumnę z kategoriami
sodium$sodium\_category <- cut(sodium$sodium\_concentration, breaks = breaks, labels = labels, include.lowest = TRUE)

# przykład - przekodowanie case\_when()

# Utwórz nową kolumnę z wykorzystaniem case\_when
sodium <- sodium %>%
 mutate(category = case\_when(
 sodium\_concentration <= 129 ~ "very low values",
 sodium\_concentration <= 135 ~ "below standard",
 sodium\_concentration <= 144 ~ "standard",
 sodium\_concentration <= 150 ~ "above normal",
 TRUE ~ "very high values" # Jeśli żaden warunek nie jest spełniony
 ))

# zad 10.

Stwórz nową kolumnę grupa\_wieku i przekoduj do niej wiek dzieląc na 3 grupy młodsza, starsza, najstarsza. Jako punkty graniczne weźmy 40 lat i 70 lat

#wpisz kod: