

# Programowanie w języku R. Wprowadzenie

Aleksander Denisiuk  
Uniwersytet Warmińsko-Mazurski  
Olsztyn, ul. Słoneczna 54  
[denisjuk@matman.uwm.edu.pl](mailto:denisjuk@matman.uwm.edu.pl)

21 lutego 2022

# *Wprowadzenie*

Analiza danych

Jak opracować dane

Historia R

Najnowsza wersja tego dokumentu dostępna jest pod adresem

<http://wmii.uwm.edu.pl/~denisjuk/uwm>

## Analiza danych

- ❖ Dane
- ❖ Gromadzenie danych
- ❖ Wyniki

Jak opracować dane

Historia R

# Analiza danych

# Pozyskiwanie danych

Analiza danych

❖ Dane

❖ Gromadzenie danych

❖ Wyniki

Jak opracować dane

Historia R

- Obserwacja
  - ◆ zminimalizowane oddziaływanie obserwatora na obiekt
    - czym się żywi rzadki gatunek ssaków
    - exit poll
- Eksperyment
  - ◆ określone oddziaływanie obserwatora na obiekt
    - efektywność leku

# Populacja generalna a próba

Analiza danych

❖ Dane

❖ Gromadzenie danych

❖ Wyniki

Jak opracować dane

Historia R

- Populacja generalna vs. próba
  - ◆ jaki rodzaj lodów na większy popyt?
  - ◆ spis powszechny Imperium Rosyjskiego 1897

# Powtórzenie

Analiza danych

❖ Dane

❖ Gromadzenie danych

❖ Wyniki

Jak opracować dane

Historia R

- wiele różnych obiektów
- powtórzenia niezależne
  - ◆ średnia długość żabek, wyłowionych siatką ze stawu
  - ◆ pokazywanie przedmiotu na 1/10 sekundy
    - 10 osób po 5 doświadczeń — ile powtórzeń?
  - ◆ jakość telewizorów na taśmie produkcyjnej
    - 10 telewizorów pod rząd?
  - ◆ liczebność próby:
    - jak najwięcej
    - 30

# Losowość

Analiza danych

❖ Dane

❖ Gromadzenie danych

❖ Wyniki

Jak opracować dane

Historia R

- Każdy obiekt populacji generalnej powinien mieć takie same prawdopodobieństwo trafić do próby
  - ◆ wybrać w lesie 100 drzew, aby oszacować stężenie metali ciężkich
    - wprowadzić porządek, którego początkowo nie ma
  - ◆ badanie skuteczności insektycydów
    - pierwszy owad z szelki Petriego (?)
- Trzymać się populacji generalnej
  - ◆ klienci sklepu
  - ◆ zwierzęta w zoo
- Generalized randomized block design
- Obserwacja vs. eksperyment
  - ◆ skuteczność leku

# Wnioski z analizy danych

Analiza danych

❖ Dane

❖ Gromadzenie danych

❖ Wyniki

Jak opracować dane

Historia R

- Weryfikacja *oczywistych* hipotez
  - ◆ ćma kukurydziana
- Czynniki statystyczne
  - ◆ średnia, wariancja
- Hipotezy statystyczne
  - ◆ istotność różnicy między próbami
- Analiza korelacji
  - ◆ zależność między przyczyną a skutkiem
- Analiza strukturalna
  - ◆ klasyfikacja

Analiza danych

**Jak opracować dane**

- ❖ Ręcznie
- ❖ Komputerowo

Historia R

# Jak opracować dane

# Ręcznie

Analiza danych

Jak opracować dane

❖ Ręcznie

❖ Komputerowo

Historia R

- No comments

# *Programy biurowe*

Analiza danych

Jak opracować dane

❖ Ręcznie

❖ **Komputerowo**

Historia R

- Kalkulator
- Arkusz kalkulacyjny

# Programy okienkowe

Analiza danych

Jak opracować dane

❖ Ręcznie

❖ Komputerowo

Historia R

- STATISTICA
- SPSS
- Stata
  - ◆ Zamkniętość
  - ◆ Nieelastyczność
  - ◆ Niewiarygodność

# Śrdowiska statystyczne

Analiza danych

Jak opracować dane

❖ Ręcznie

❖ Komputerowo

Historia R

- Systemy interakcyjne
- Brak możliwości wyboru działania z menu
- Programowane
- Pełna kontrola nad obliczeniami
  - ◆ SAS
    - 193 570 PLN
  - ◆ R
    - wolna implementacja języka S

Analiza danych

Jak opracować dane

Historia R

❖ S

❖ R

# Historia R

# S

Analiza danych

Jak opracować dane

Historia R

❖ S

❖ R

- 1976: Bell Labs, FORTRAN
- 1980: UNIX
- 1988: S-Plus (około \$ 6 500)

# R

Analiza danych

Jak opracować dane

Historia R

❖ S

❖ R

- 1993, Ross Ihaka, Robert Gentleman (Auckland University of Technology)
  - ◆ alternatywna implementacja języka S
    - nowa gałąź (wpływ języka Scheme)
  - ◆ GNU GPL v2
  - ◆ CRAN
    - około 3 500 pakietów
  - ◆ Powszechnie używany
    - między innymi Twitter, Google, Boeing

# Wady i zalety

Analiza danych

Jak opracować dane

Historia R

❖ S

❖ R

- Elastyczność
- Wolność
- Trudność
  - ◆ możliwość pisania skryptów
- Wolność
  - ◆ możliwość implementacji funkcji w C, w FORTRANie, w Pythonie