

Narzędzia informatyczne. R

Aleksander Denisiuk
Uniwersytet Warmińsko-Mazurski
Olsztyn, ul. Słoneczna 54
denisjuk@matman.uwm.edu.pl

R

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

Najnowsza wersja tego dokumentu dostępna jest pod adresem

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

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

Wprowadzenie

R

Wprowadzenie

❖ R

- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Legalny system do profesjonalnego przetwarzania danych
 - ◆ statystyka, finanse, bioinformatyka, big data, etc
 - ◆ język programowania, pakiety
 - ◆ grafika
- 1993, Ross Ihaka, Robert Gentleman (Auckland University of Technology)
 - ◆ alternatywna implementacja języka S
 - ◆ CRAN
- Dwa tryby
 - ◆ interaktywny
 - ◆ wsadowy

```
R CMD BATCH file.r  
Rscript file.r
```
- Środowiska graficzne

Dane

Wprowadzenie

❖ R

❖ Dane

❖ Wektory

❖ Macierze

❖ Dane nominalne

❖ Listy

❖ Ramki

❖ Import

❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- numeric, logical, character
- Obiekty, zmienne
- Wyrażenia

```
x = 1/3
```

```
eq.delta <- 2*2-4*2*3
```

```
factorial(7) -> X
```

Wektory

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Konkatencja

```
x <- c(1, 2, 3, 4, 5)
```

- Z klawiatury

```
y <- scan()
```

- Ciąg

```
s <- seq(1, 7)
```

```
s <- 1:7
```

```
S <- seq(from = 1, to = 5, by = 0.5)
```

- Powtórzenie

```
header <- rep("tekst", 7)
```

Działania na wektorach

Wprowadzenie

❖ R

❖ Dane

❖ Wektory

❖ Macierze

❖ Dane nominalne

❖ Listy

❖ Ramki

❖ Import

❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

● Konkatenacja

```
v <- c(v1, v2)
```

◆ jeżeli jeden z wektorów jest tekstowy, to drugi zostanie przekowertowany

● Dostęp do elementów

```
y[3]; y[3:5]
```

```
y[y>2]; y[c(1, 4)]
```

● Usuwanie elementów

```
y[-c(1, 4)]
```

● Sortowanie

```
sort(z)
```

```
sort(z, decreasing = TRUE)
```

Macierze

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ **Macierze**
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

● Kolumnami

```
my.macierz <- matrix(seq(1, 16),  
                    nrow = 4, ncol = 4)
```

● Wierszami

```
my.macierz <- matrix(seq(1, 16), nrow = 4,  
                    ncol = 4, byrow = TRUE)
```

● Nagłówki

```
rownames(macierz) <- c("1", "2", "3", "4")  
colnames(macierz) <- c("A", "B", "C", "D")
```

● Cykliczne uzupełnienie

```
macierz2 <- matrix(seq(1, 12), nrow = 4,  
                  ncol = 4, byrow = TRUE)
```


Macierze a wektory

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ **Macierze**
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

● Macierz z wektora

```
macierz <- 1:16  
dim(macierz) <- c(4, 4)  
dim(macierz)
```

● Macierz z wektorów

```
a <- c(1, 2, 3, 4)  
b <- c(5, 6, 7, 8)  
d <- c(9, 10, 11, 12)  
e <- c(13, 14, 15, 16)  
cbind(a, b, d, e)  
rbind(a, b, d, e)
```

Działania na macierzach

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ **Macierze**
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Działania wektorowe

```
macierz[13]
```

```
macierz[2, 3]
```

```
macierz[, 3]
```

```
macierz[1, ]
```

- Mnożenie (każdy przez każdy, wynik — wektor)

```
macierz[1, ] * macierz[4, ]
```

- Transponowanie

```
t(macierz)
```

Dane nominalne

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Kodowanie liczbami

```
color <- c(0, 0, 1, 1, 1, 2)
```

- Zmiennymi wyliczeniowymi

```
color <- factor(color, levels = c(0, 1, 2))
```

- Etykiety

```
colors <- c("blue", "blue",  
           "brown", "brown", "brown", "gray")
```

```
eyes <- factor(colors)
```

- Uporządkowane

```
data <- c("small", "medium", "small",  
         "large", "medium", "large")
```

```
cubes <- factor(data,  
               levels=c("small", "medium", "large"),  
               ordered=TRUE)
```

Listy

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Elementy różnych typów

```
student <- list(name = "Jan",  
                surname = "Kowalski",  
                birthyear = 1995,  
                male = TRUE)
```

- Dostęp

```
student$name  
student[[1]]  
student[[1]][2]
```

- Struktura listy

```
str(student)
```

Ramki danych

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ **Ramki**
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Podobne do arkusza kalkulacyjnego

```
city <- c("Olsztyn", "Olsztyn", "Elbląg",  
         "Elbląg", "Gdańsk", "Gdańsk")  
gender <- c("Male", "Female", "Male",  
           "Female", "Male", "Female")  
number <- c(81505, 94205, 59037,  
           64234, 218370, 242445)  
CITY <- data.frame(City = city,  
                  Gender = gender, Number = number)
```

```
CITY  
  
      City Gender Number  
1 Olsztyn   Male  81505  
2 Olsztyn Female  94205  
3  Elbląg   Male  59037  
4  Elbląg Female  64234  
5  Gdańsk   Male 218370  
6  Gdańsk Female 242445
```

Dostęp

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ **Ramki**
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

● Do kolumn

```
CITY$Gender  
CITY[, 2]  
CITY[[2]]  
CITY["Gender"]
```

● Do komórek

```
CITY$Gender[4]  
CITY$Gender[1:3]  
CITY$City[CITY$Number > 100000]  
CITY$Number[CITY$Gender == "Male"]  
CITY$City[CITY$Gender == "Male"  
          & CITY$Number < 100000]
```

● Do nazw kolumn

```
names(CITY)
```

Działania

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ **Ramki**
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Początek i koniec

```
head(CITY, n = 3)
```

```
tail(CITY, n = 2)
```

- Tekstowe przedstawienie

```
str(CITY)
```

- Sortowanie

```
CITY2 <- CITY[order(CITY$Gender,  
-CITY$Number), ]
```

- etc

Import danych w R

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- Najlepiej z pliku tekstowego
 - ◆ `.txt` — dane rozdzielone tabulacją
 - ◆ `.csv` — rozdzielone przecinkiem
 - pakiet `foreign` — import z rozmaitych formatów
- Puste dane trzeba zastąpić przez `NA`
- Pierwszy wiersz — nazwy kolumn
 - ◆ pierwszy element w wierszach może być nazwą wiersza

```
chem <- read.table(file = "hydro_chem.txt",  
                  header = TRUE)  
chem <- read.table(file = "hydro_chem.csv",  
                  header = TRUE, sep = ",")  
chem <- read.csv(file = "hydro_chem.csv",  
                header = TRUE)
```


Testowe dane w R

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ **Import**
- ❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- R ma kilka przykładowych zestawów danych
- Wyliczenie dostępnych danych

```
data()
```

- **Wczytywanie danych**

```
data(AirPassengers)
```

```
data(Indometh)
```

```
data(mtcars)
```

Obliczenia

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ **Obliczenia**

Wizualizacja

Statystyka opisowa

Studium Przypadku

● Obliczenia wektorowe

```
x <- 1:10
```

```
sqrt(x)
```

```
[1] 1.000000 1.414214 1.732051 ...
```

apply

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ **Obliczenia**

Wizualizacja

Statystyka opisowa

Studium Przypadku

```
M <- matrix(seq(1,16), 4, 4)
```

```
apply(M, 1, min)
```

```
[1] 1 2 3 4
```

```
apply(M, 2, max)
```

```
[1] 4 8 12 16
```

```
M <- array(seq(32), dim = c(4,4,2))
```

```
apply(M, 1, sum)
```

```
[1] 120 128 136 144
```

```
apply(M, c(1,2), sum)
```

```
[,1]    [,2]    [,3]    [,4]
```

```
[1,]    18     26     34     42
```

```
[2,]    20     28     36     44
```

```
[3,]    22     30     38     46
```

```
[4,]    24     32     40     48
```

lapply

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ **Obliczenia**

Wizualizacja

Statystyka opisowa

Studium Przypadku

```
x <- list(a = 1, b = 1:3, c = 10:100)
```

```
lapply(x, FUN = length)
```

```
$a
```

```
[1] 1
```

```
$b
```

```
[1] 3
```

```
$c
```

```
[1] 91
```

```
lapply(x, FUN = sum)
```

```
$a
```

```
[1] 1
```

```
$b
```

```
[1] 6
```

```
$c
```

```
[1] 5005
```

sapply

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import
- ❖ **Obliczenia**

Wizualizacja

Statystyka opisowa

Studium Przypadku

```
x <- list(a = 1, b = 1:3, c = 10:100)
```

```
sapply(x, FUN = length)
```

a	b	c
1	3	91

```
sapply(x, FUN = sum)
```

a	b	c
1	6	5005

tapply

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import

❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

```
x <- sample(1:4, size=50, replace=TRUE)
gr <- as.factor(sample(c("A", "B", "C", "D"),
                      size=50, replace=T))
```

```
tapply(x, gr, sum)
```

A	B	C	D
25	35	41	27

- Oblicza sumę x dla każdego elementu grupy gr

Inne

Wprowadzenie

- ❖ R
- ❖ Dane
- ❖ Wektory
- ❖ Macierze
- ❖ Dane nominalne
- ❖ Listy
- ❖ Ramki
- ❖ Import

❖ Obliczenia

Wizualizacja

Statystyka opisowa

Studium Przypadku

- replicate
- vapply
- mapply
- rapply
- by
- ...

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

Wizualizacja

Prędkość eliminacji indometacyny

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

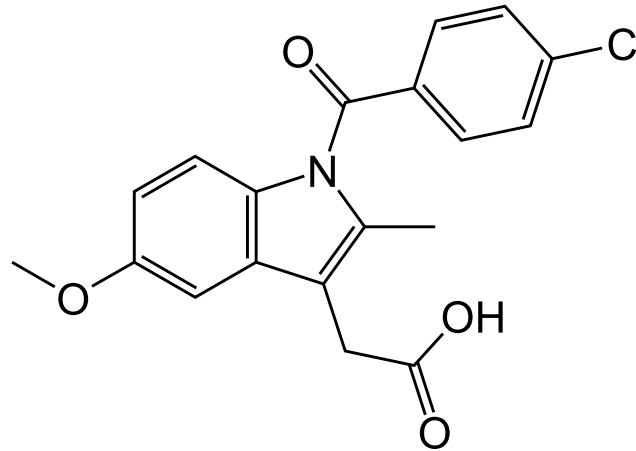
❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku



```
data (Indometh)
```

```
names (Indometh)
```

```
[1] "Subject" "time"      "conc"
```

```
attach (Indometh)
```

```
plot (time, conc)
```

Wynik

[Wprowadzenie](#)

[Wizualizacja](#)

❖ **Indometacyna**

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

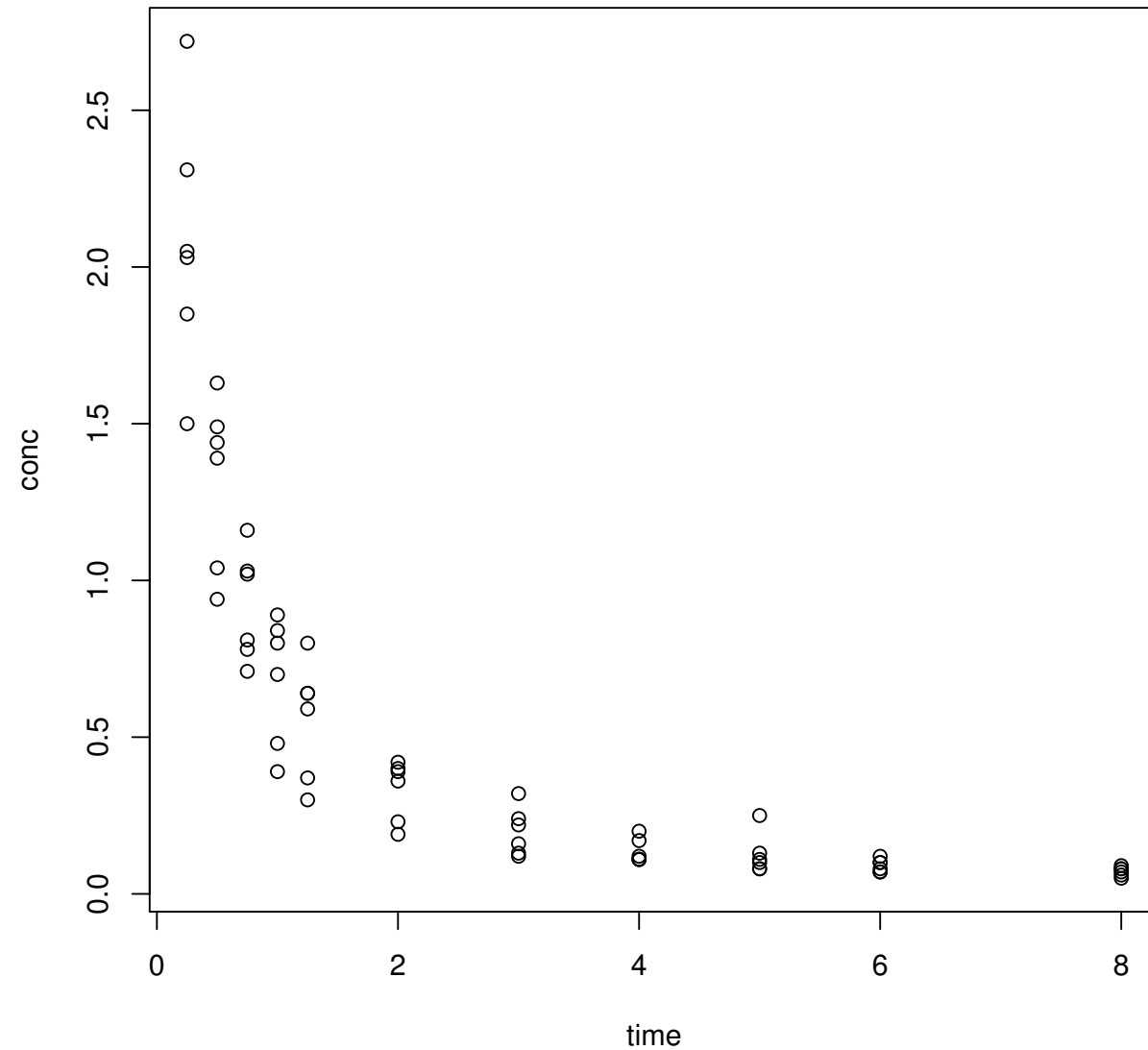
❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

[Statystyka opisowa](#)

[Studium Przypadku](#)



Średnia koncentracja

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
means <- tapply(conc, time, mean)
indom.times <- as.numeric(names(means))
plot(indom.times, means, xlab = "Czas",
      ylab = "Koncentracja",
      type="b",
      main = "Prędkość eliminacji indometacyny",
      )
```

[Wprowadzenie](#)

[Wizualizacja](#)

❖ **Indometacyna**

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

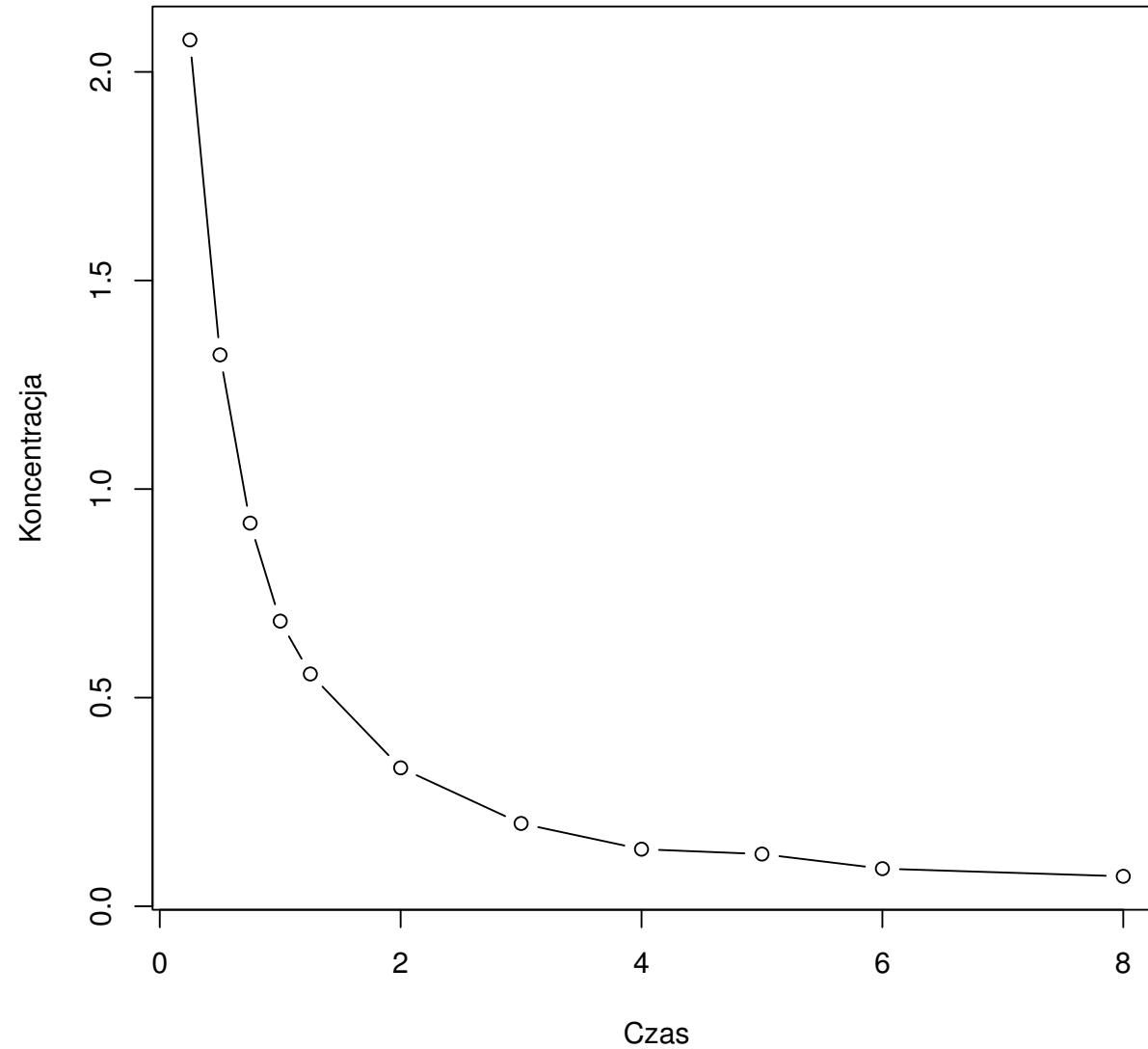
❖ Wykres kołowy

❖ Diagram słupkowy

[Statystyka opisowa](#)

[Studium Przypadku](#)

Prędkość eliminacji indometacyny



Histogramy

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ **Histogramy**

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

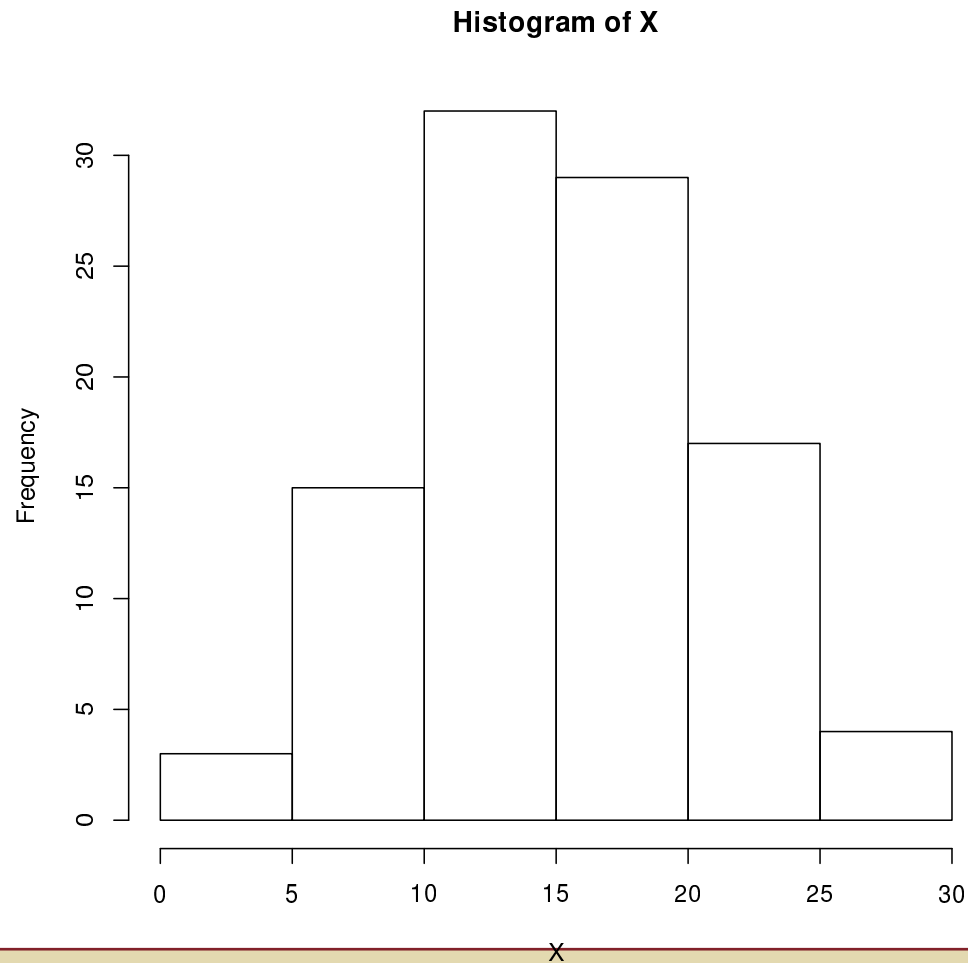
❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
X <- rnorm(n = 100,  
           mean = 15, sd = 5)  
hist(X)
```



Gęstość

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ **Histogramy**

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

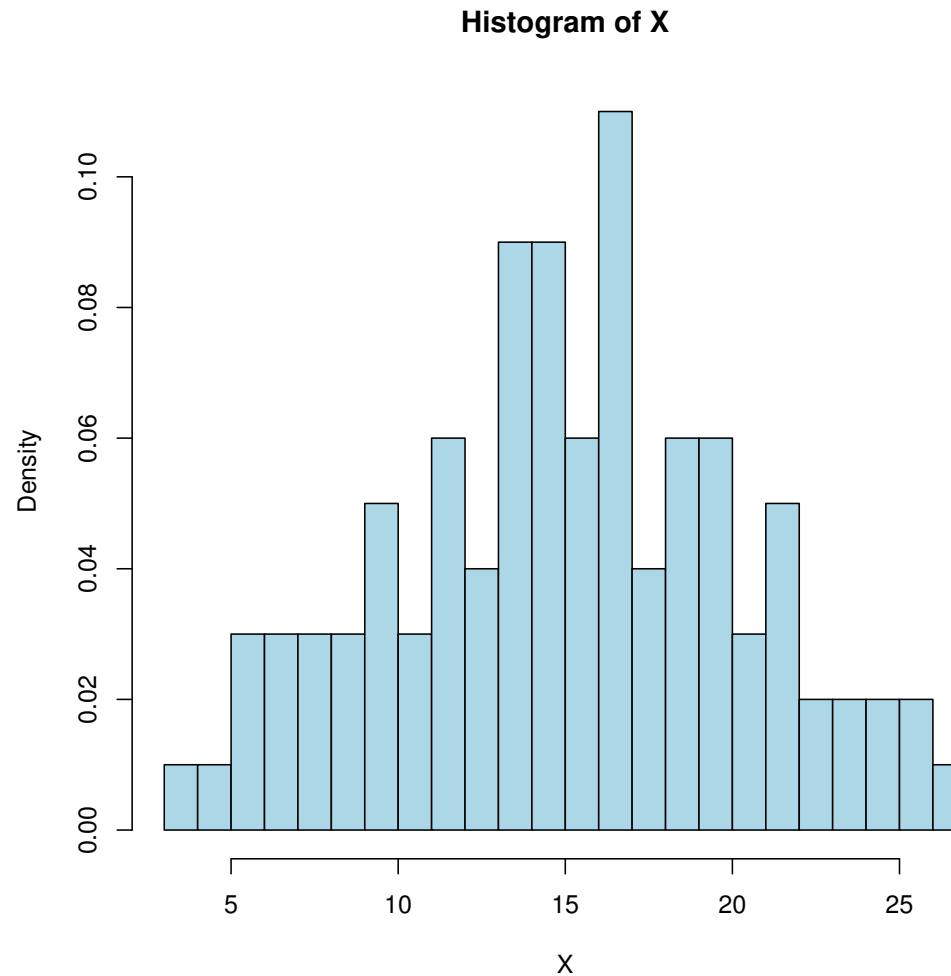
❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
hist(X, breaks=20, col="lightblue", freq=FALSE)
```



Gęstość

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ **Histogramy**

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
hist(X, breaks=20, freq=FALSE, col="lightblue",  
xlab = "Zmienna X",  
ylab = "Gęstość prawdopodobieństwa",  
main = "Histogram, połączony z gęstością")  
lines(density(X), col = "red", lwd = 2)  
lines(density(X, bw = 0.8), col = "blue",  
      lwd = 2)
```

- bw — *bandwidth*, parameter wygładzania

[Wprowadzenie](#)

[Wizualizacja](#)

❖ [Indometacyna](#)

❖ **Historamy**

❖ [Gęstości](#)

❖ [Przedział typowy](#)

❖ [Gęstość 2W](#)

❖ [Dane nominalne](#)

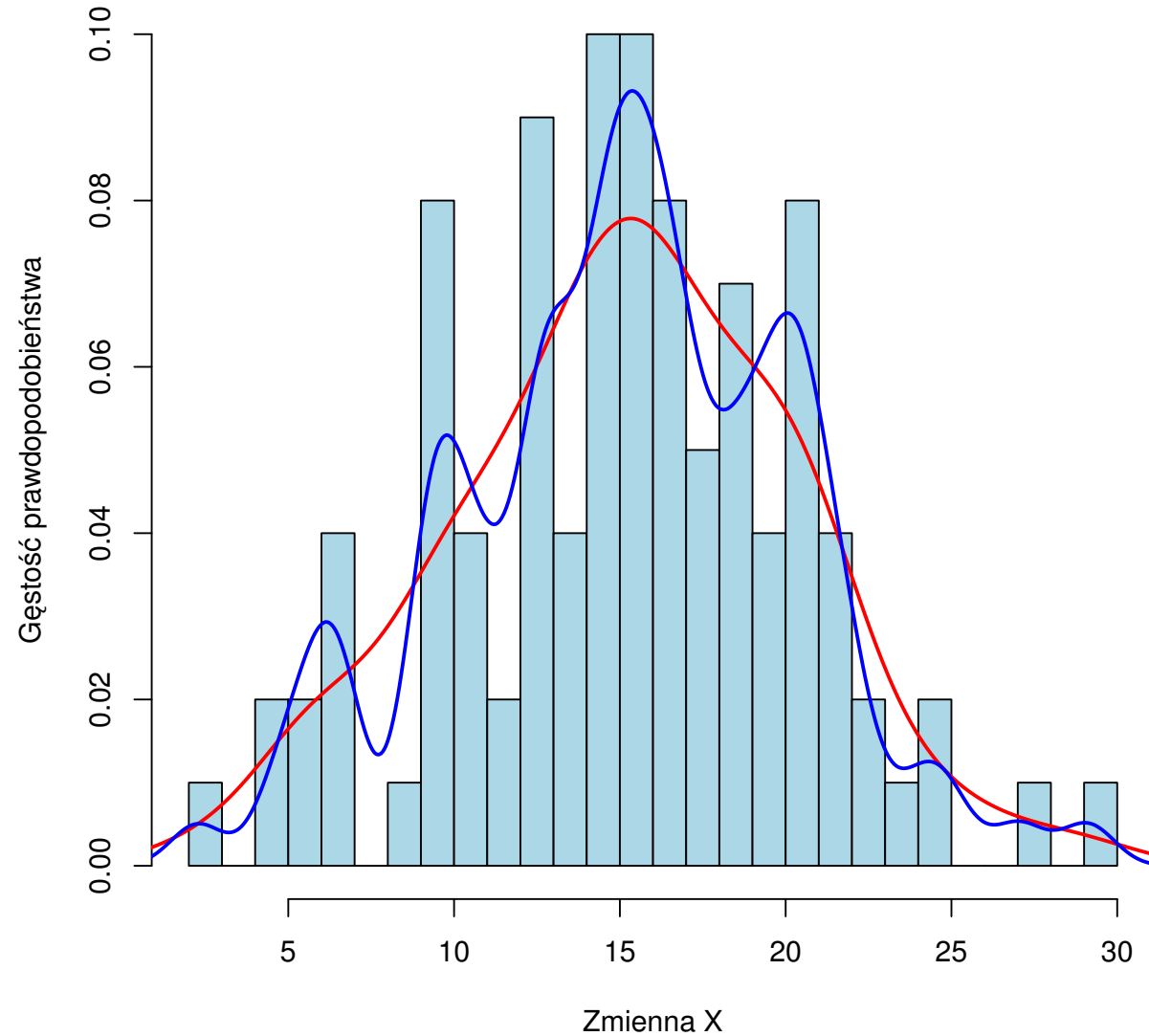
❖ [Wykres kołowy](#)

❖ [Diagram słupkowy](#)

[Statystyka opisowa](#)

[Studium Przypadku](#)

Histogram, połączone z gęstością



Dane

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ **Gęstości**

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

- Pakiet `sm`
 - ◆ instalacja: `install.packages("sm")`
- Dane: `InsectSprays`
 - ◆ 6 sprayów
 - ◆ 12 roślin

```
head(InsectSprays)
```

```
      count spray
1         10     A
2          7     A
3         20     A
4         14     A
5         14     A
6         12     A
```

Porównywanie

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ **Gęstości**

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
attach (InsectSprays)
```

```
library (sm)
```

```
sm.density.compare(count, spray,  
                    lwd = 2, xlab="Ilość owadów")  
title(main="Krzywe rozkładów gęstości")
```

```
colfill<-c(2:(2+length(levels(spray))))  
legend("topright",  
       levels(spray), fill=colfill)
```

```
# Legenda w miejscu kliknięcia myszką  
# legend(locator(1), levels(spray),  
        fill=colfill)
```

[Wprowadzenie](#)

[Wizualizacja](#)

❖ Indometacyna

❖ Histogramy

❖ **Gęstości**

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

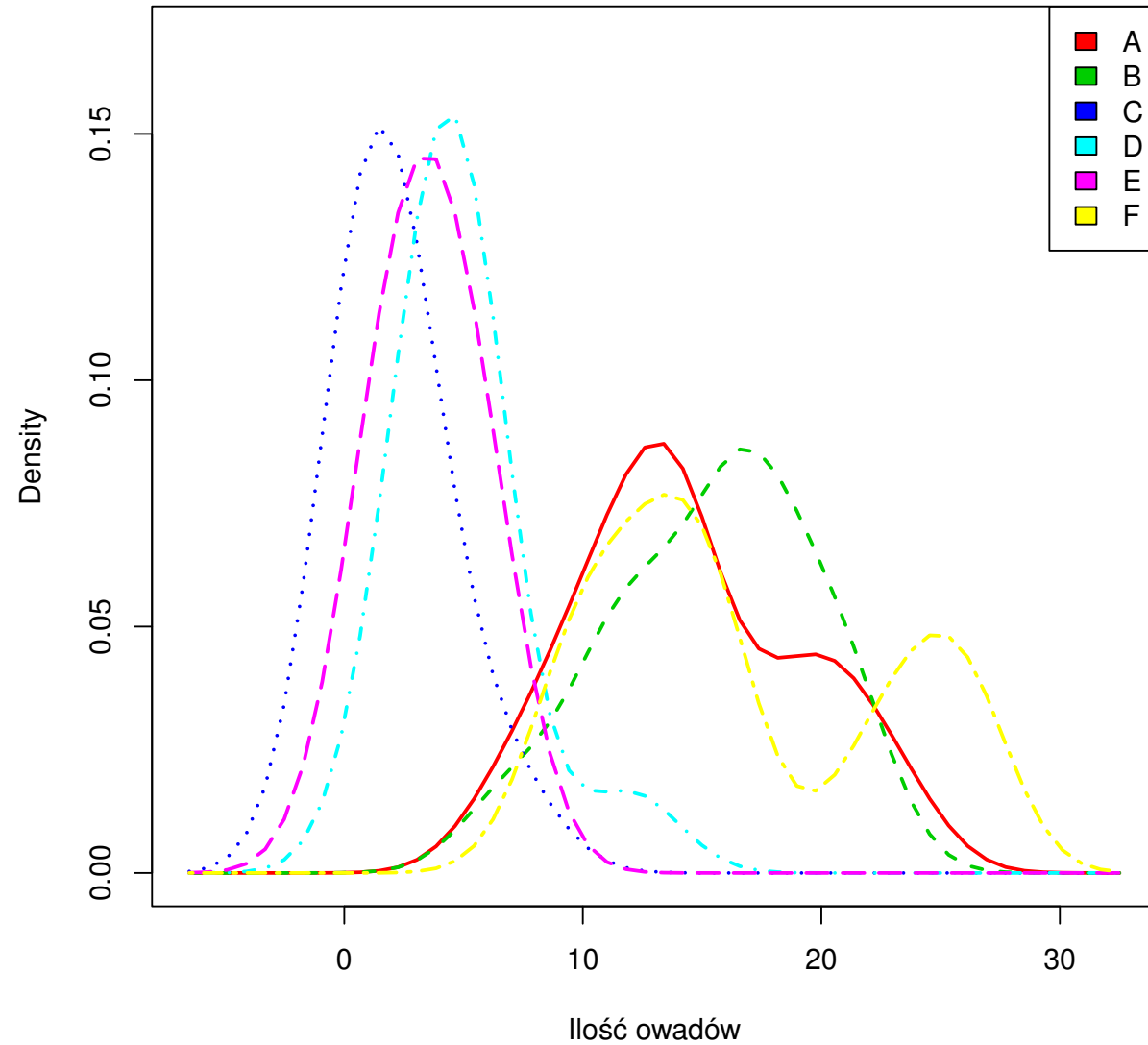
❖ Wykres kołowy

❖ Diagram słupkowy

[Statystyka opisowa](#)

[Studium Przypadku](#)

Krzywe rozkładów gęstości



Wykres z przedziałem typowym

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ **Przedział typowy**

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
boxplot(count ~ spray, data = InsectSprays,  
        xlab = "Insektycydy",  
        ylab = "Ilość owadów",  
        main = "Efektywność Insektycydów",  
        col = "coral")
```

- Mediana, kwartyle, obserwacje odstające

[Wprowadzenie](#)

[Wizualizacja](#)

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ **Przedział typowy**

❖ Gęstość 2W

❖ Dane nominalne

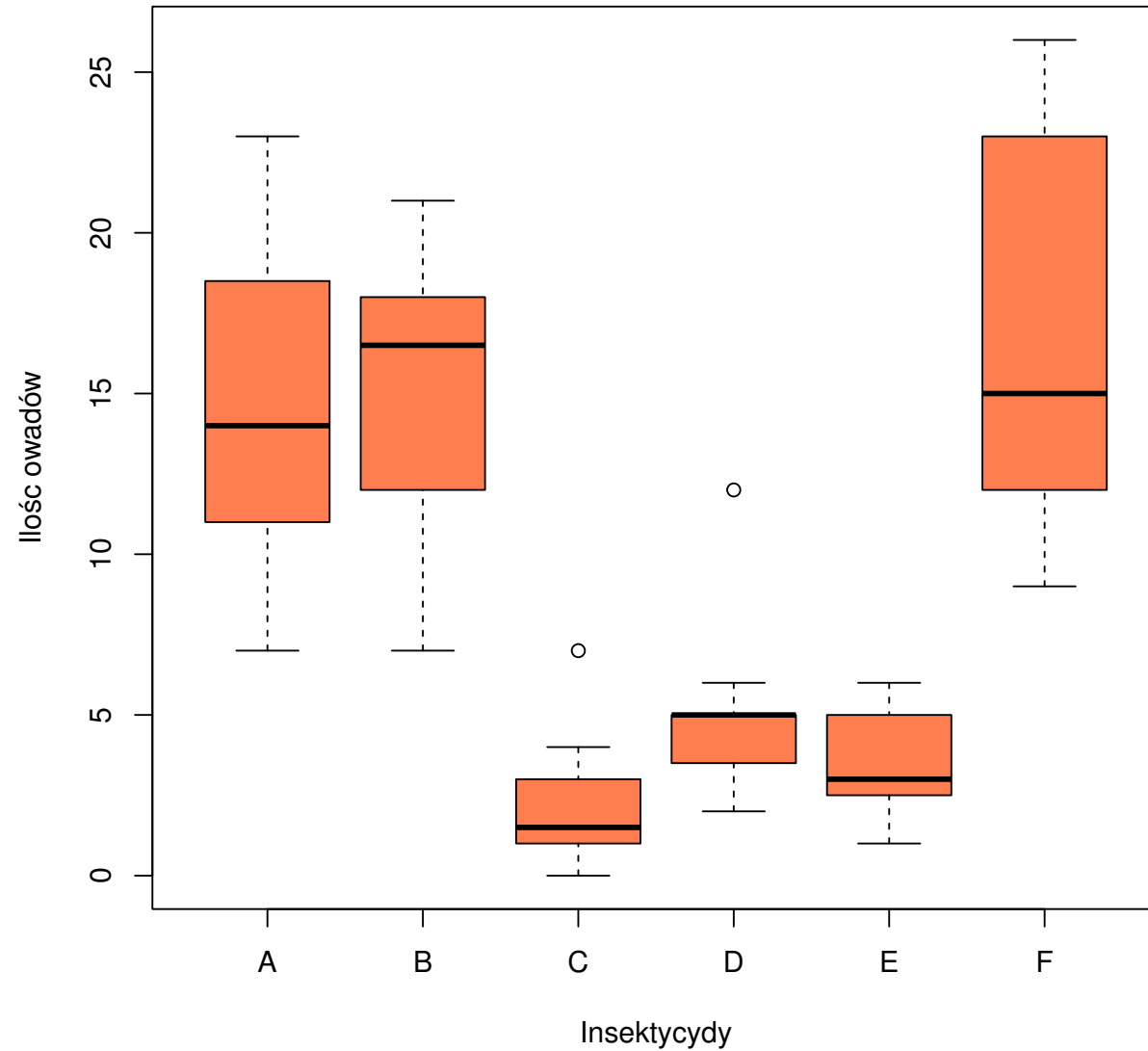
❖ Wykres kołowy

❖ Diagram słupkowy

[Statystyka opisowa](#)

[Studium Przypadku](#)

Efetywność Insektycydów



Gęstość rozkładu dwóch zmiennych

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
data (Indometh) ; attach (Indometh)
library (MASS)
f <- kde2d (time, conc)
image (f, xlab="Czas",
       ylab="Koncentracja indometacyny")
contour (f, add=TRUE)
```

Wynik

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ **Gęstość 2W**

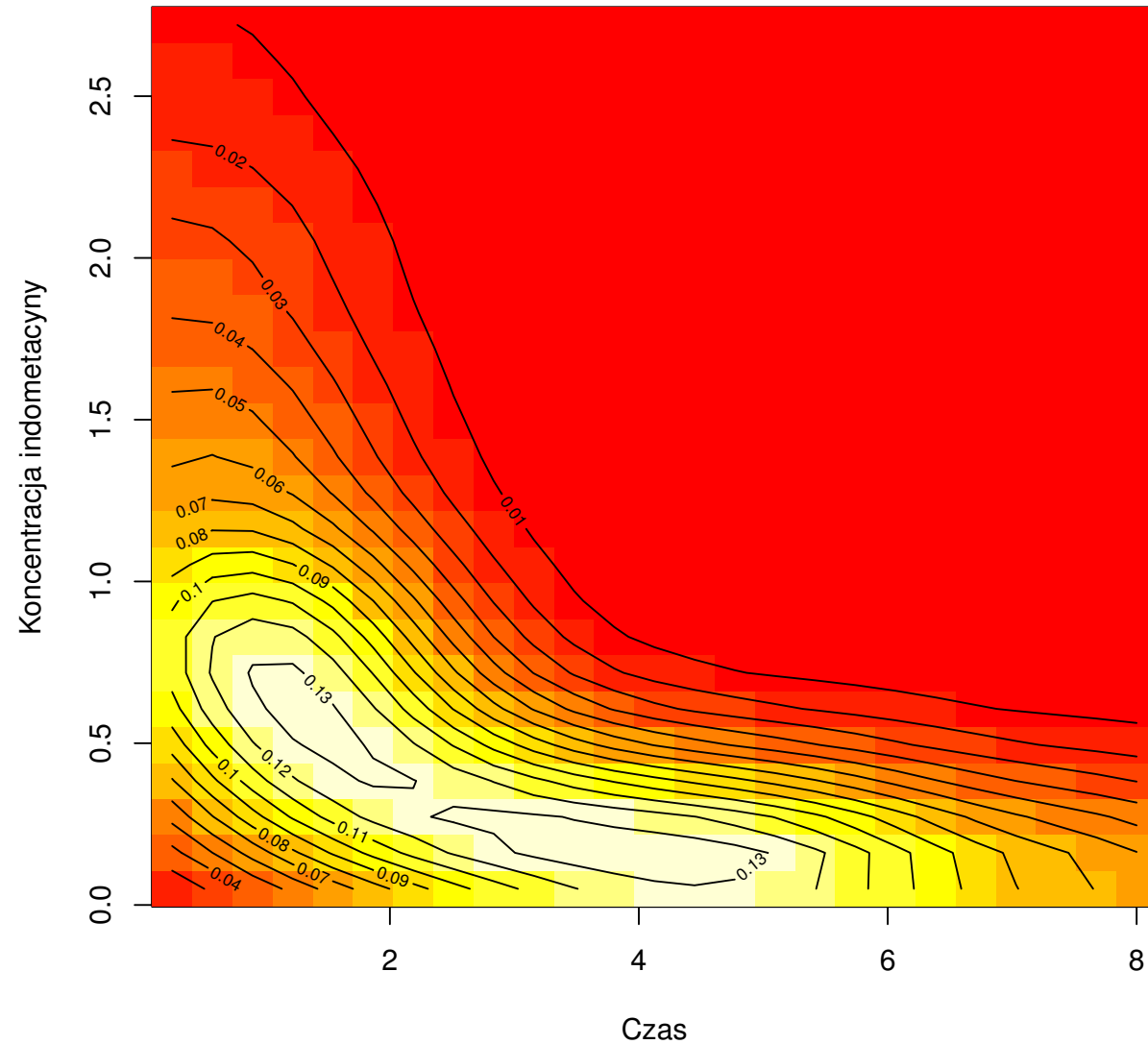
❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku



Dane

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

- Pakiet HSAUR2 (An Introduction to Applied Multivariate Analysis with R, Brian Everitt, Torsten Hothorn, Springer, 2011))
 - ◆ instalacja: `install.packages("HSAUR2")`
- Dane: plasma
 - `library(HSAUR2)`
 - `data(plasma)`

Zagadnienie

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W

❖ Dane nominalne

- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

summary (plasma)

	fibrinogen	globulin	ESR
Min.	:2.090	Min. :28.00	ESR < 20:26
1st Qu.	:2.290	1st Qu.:31.75	ESR > 20: 6
Median	:2.600	Median :36.00	
Mean	:2.789	Mean :35.66	
3rd Qu.	:3.167	3rd Qu.:38.00	
Max.	:5.060	Max. :46.00	

- ESR — *Odczyn Biernackiego (OB)*, wskaźnik opadania erytrocytów, dane nominalne
- Czy istnieje zależność między OB a koncentracją fibrynogenu oraz globulin

Funkcja `cdplot`

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ **Dane nominalne**
- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

● Wykres gęstości warunkowej

```
layout(matrix(1:2, ncol = 2))
cdplot(ESR ~ fibrinogen, data = plasma)
cdplot(ESR ~ globulin, data = plasma)
```

Wynik

Wprowadzenie

Wizualizacja

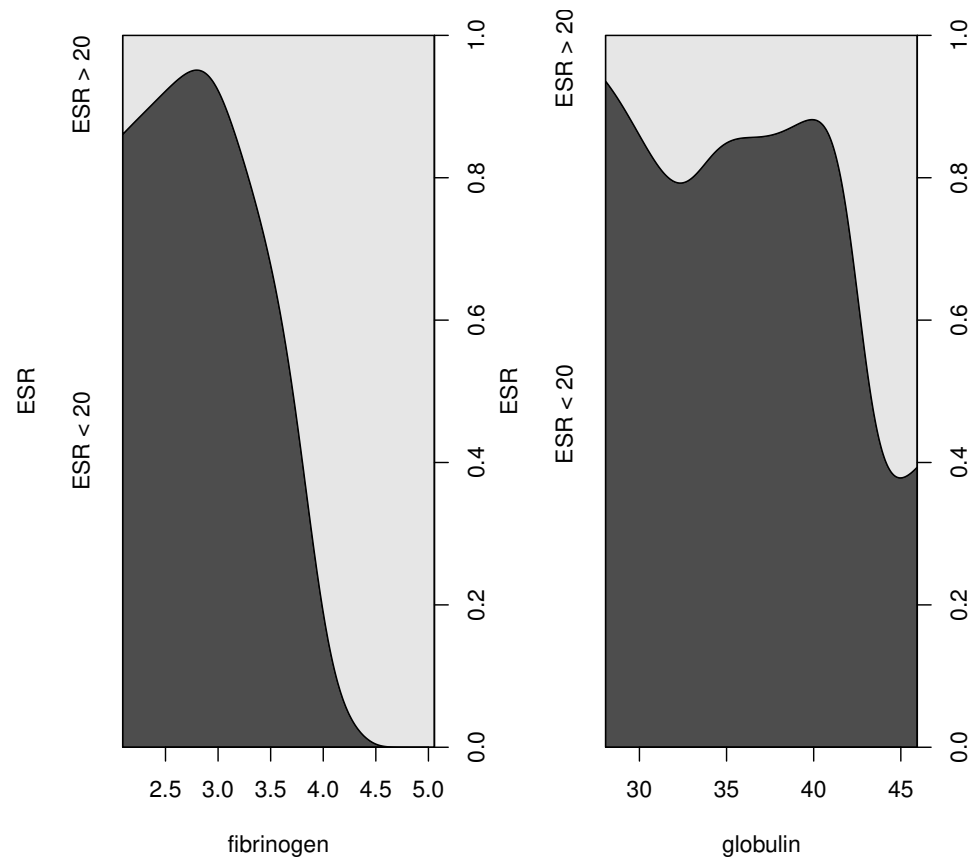
- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W

❖ Dane nominalne

- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku



- Zwiększenie koncentracji białka powoduje zwiększenie prawdopodobieństwa $OB > 20$

Dane

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

● Wybory prezydenckie 2015

```
votes <- c(34.76, 33.77, 20.80, 3.26,  
          2.38, 1.60, 1.42,  
          0.83, 0.52, 0.46, 0.20)  
percent.voted <- c(48.96, 51.04)
```

Etykiety

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

❖ Wykres kołowy

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
names (votes) <- c ("Andrzej Duda",  
  "Bronisław Komorowski",  
  "Paweł Kukiz",  
  "Janusz Korwin-Mikke",  
  "Magdalena Ogórek",  
  "Adam Jarubas",  
  "Janusz Palikot",  
  "Grzegorz Braun",  
  "Marian Kowalski",  
  "Jacek Wilk",  
  "Paweł Tanajno")  
names (percent.voted) <- c ("zagłosowało",  
  "nie głosowało")
```

Wykres

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
pielabels <- paste(votes, "%", sep=" ")
par(mfrow=c(1,2))
pie(percent.voted, radius = 0.9, cex = 0.6,
     main = "Frekwencja",
     col = c("black", "gray80"))
pie(votes, cex = 0.6, radius = 0.9,
     init.angle = -10, main = "Wyniki wyborów",
     col = c(2:8), labels=pielabels)
legend("top", names(votes),
      cex=0.6, fill= c(2:8))
```

Wprowadzenie

Wizualizacja

❖ Indometacyna

❖ Histogramy

❖ Gęstości

❖ Przedział typowy

❖ Gęstość 2W

❖ Dane nominalne

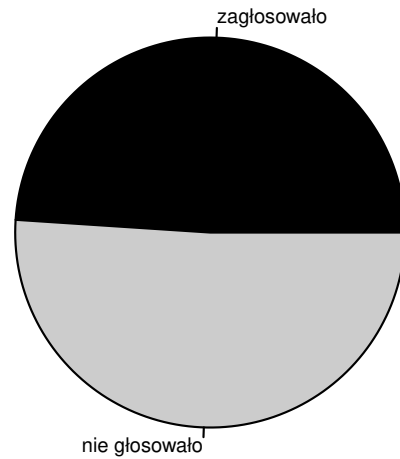
❖ **Wykres kołowy**

❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

Frekwencja



Wyniki wyborów

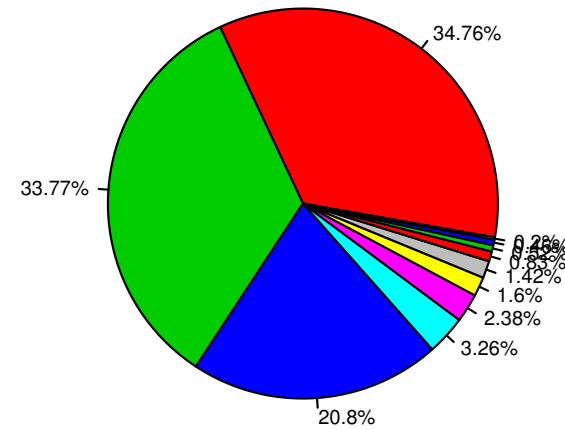


Diagram słupkowy

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ **Diagram słupkowy**

Statystyka opisowa

Studium Przypadku

```
attach (InsectSprays)  
Means <- tapply (count, spray, mean)  
barplot (Means,  
  col = "steelblue",  
  xlab = "Insektycyd",  
  ylab = "Ilość owadów",  
  border = "red",  
  width = sqrt (Means) )
```


Wynik

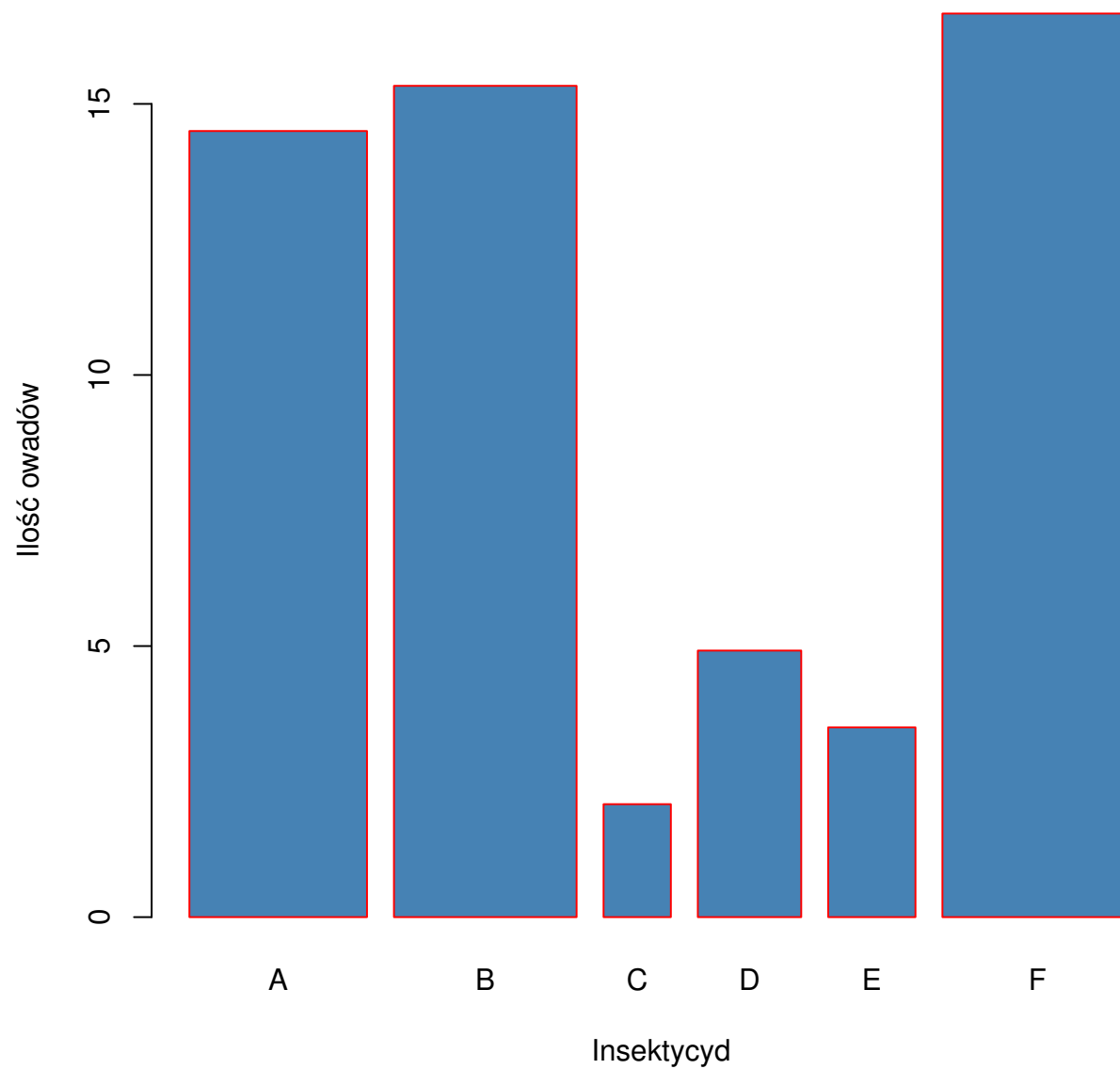
Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy**

Statystyka opisowa

Studium Przypadku



Poziomo

Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy

Statystyka opisowa

Studium Przypadku

```
attach(InsectSprays)
```

```
Means <- tapply(count, spray, mean)  
barplot(Means, density = 20,  
        angle = -45, space = 2,  
        col = "red", horiz = TRUE, las = 1,  
        ylab = "Insektycyd",  
        xlab = "Ilość owadów")
```

Wynik

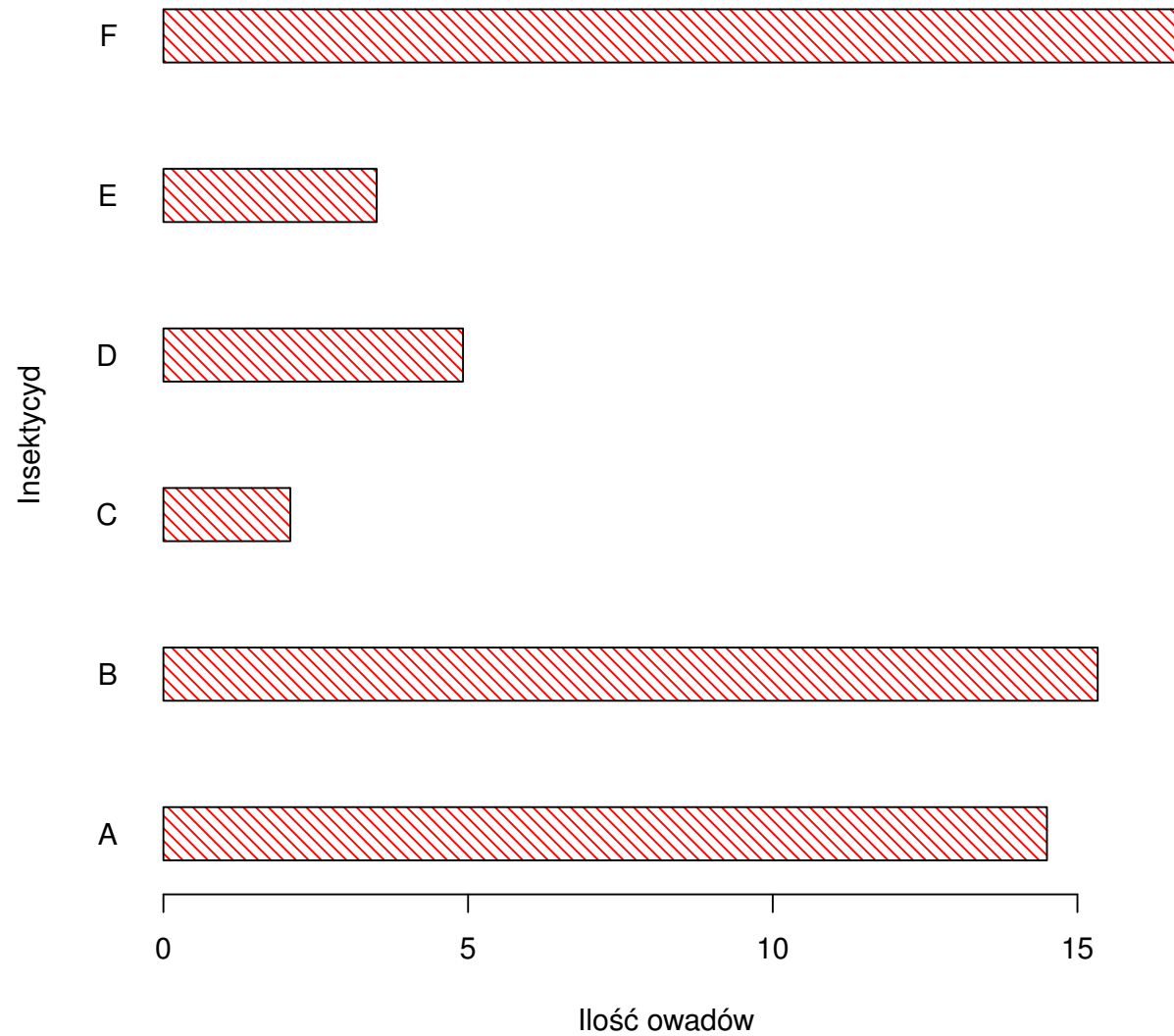
Wprowadzenie

Wizualizacja

- ❖ Indometacyna
- ❖ Histogramy
- ❖ Gęstości
- ❖ Przedział typowy
- ❖ Gęstość 2W
- ❖ Dane nominalne
- ❖ Wykres kołowy
- ❖ Diagram słupkowy**

Statystyka opisowa

Studium Przypadku



Wprowadzenie

Wizualizacja

Statystyka opisowa

- ❖ Dane
- ❖ Wskaźniki statystyczne
- ❖ Funkcja `summary`
- ❖ Badanie asymetrii i koncentracji

Studium Przypadku

Statystyka opisowa

Dane

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane

❖ Wskaźniki
statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii
i koncentracji

Studium Przypadku

```
data (mtcars)
```

```
head (mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Podstawowe parametry statystyczne

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane

❖ Wskaźniki statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii i koncentracji

Studium Przypadku

```
# Średnia:  
mean (mtcars$mpg)
```

```
[1] 20.1
```

```
# Mediana
```

```
median (mtcars$mpg)
```

```
[1] 19.2
```

```
# Wariancja:
```

```
var (mtcars$mpg)
```

```
[1] 36.3
```

```
# Kwartyle
```

```
quantile (mtcars$mpg)
```

```
      0%      25%      50%      75%     100%  
10.400 15.425 19.200 22.800 33.900
```

```
# Odchylenie standardowe:
```

```
sd (mtcars$mpg)
```

```
[1] 6.0
```

```
# Wartość minimalna:
```

```
min (mtcars$mpg)
```

```
[1] 10.4
```

```
# Wartość maksymalna:
```

```
max (mtcars$mpg)
```

```
[1] 33.9
```

Znalezienie minimów i maksymów

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane

❖ Wskaźniki
statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii
i koncentracji

Studium Przypadku

```
which.min (mtcars$mpg)
```

```
[1] 15
```

```
which.max (mtcars$mpg)
```

```
[1] 20
```

```
rownames (mtcars) [which.min (mtcars$mpg) ]
```

```
[1] "Cadillac Fleetwood"
```

```
rownames (mtcars) [which.max (mtcars$mpg) ]
```

```
[1] "Toyota Corolla"
```

Funkcja *summary*

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane

❖ Wskaźniki
statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii
i koncentracji

Studium Przypadku

summary (mtcars\$mpg)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.40	15.42	19.20	20.09	22.80	33.90

summary (mtcars)

Funkcja *summary* dla ramki

Wprowadzenie

Wizualizacja

Statystyka opisowa

- ❖ Dane
- ❖ Wskaźniki statystyczne
- ❖ Funkcja `summary`
- ❖ Badanie asymetrii i koncentracji

Studium Przypadku

summary (mtcars)

```
      mpg          cyl          disp          hp
Min.   :10.40    Min.   :4.000    Min.   : 71.1    Min.   : 52.0
1st Qu.:15.43    1st Qu.:4.000    1st Qu.:120.8    1st Qu.: 96.5
Median :19.20    Median :6.000    Median :196.3    Median :123.0
Mean   :20.09    Mean   :6.188    Mean   :230.7    Mean   :146.7
3rd Qu.:22.80    3rd Qu.:8.000    3rd Qu.:326.0    3rd Qu.:180.0
Max.   :33.90    Max.   :8.000    Max.   :472.0    Max.   :335.0

      drat          wt          qsec          vs
Min.   :2.760    Min.   :1.513    Min.   :14.50    Min.   :0.0000
1st Qu.:3.080    1st Qu.:2.581    1st Qu.:16.89    1st Qu.:0.0000
Median :3.695    Median :3.325    Median :17.71    Median :0.0000
Mean   :3.597    Mean   :3.217    Mean   :17.85    Mean   :0.4375
3rd Qu.:3.920    3rd Qu.:3.610    3rd Qu.:18.90    3rd Qu.:1.0000
Max.   :4.930    Max.   :5.424    Max.   :22.90    Max.   :1.0000

      am          gear          carb
Min.   :0.0000    Min.   :3.000    Min.   :1.000
1st Qu.:0.0000    1st Qu.:3.000    1st Qu.:2.000
Median :0.0000    Median :4.000    Median :2.000
Mean   :0.4062    Mean   :3.688    Mean   :2.812
3rd Qu.:1.0000    3rd Qu.:4.000    3rd Qu.:4.000
Max.   :1.0000    Max.   :5.000    Max.   :8.000
```

Poprawki dla danych nominalnych

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane
❖ Wskaźniki
statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii
i koncentracji

Studium Przypadku

```
mtcars$vs <- as.factor(mtcars$vs)
```

```
mtcars$am <- as.factor(mtcars$am)
```

```
summary(mtcars$vs)
```

```
 0  1
```

```
18 14
```

```
summary(mtcars$am)
```

```
 0  1
```

```
19 13
```

Badanie asymetrii i koncentracji

Wprowadzenie

Wizualizacja

Statystyka opisowa

❖ Dane

❖ Wskaźniki
statystyczne

❖ Funkcja `summary`

❖ Badanie asymetrii
i koncentracji

Studium Przypadku

```
install.packages("moments")
library(moments)
# Wskaźnik skosności
skewness(mtcars$mpg)
[1] 0.6404399
# Kurtoza
kurtosis(mtcars$mpg)
[1] 2.799467
```

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

Studium Przypadku

Badanie hipotez statystycznych

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

- Mamy kilka grup badanych jednostek
- Pytanie: czy różnice tych grup są statystycznie istotne?
 - ◆ hipoteza zerowa: różnice są przypadkowe
 - ◆ jakie jest prawdopodobieństwo otrzymać takie wyniki losowo?
 - Duże: hipoteza zerowa jest przyjmowana
 - Małe: hipoteza zerowa jest odrzucana
- Metoda: analiza wariancji

Losowe próbki z rozkładu normalnego

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

- W miasteczku (200 mieszkańców) badano wpływ diety na rzut serca
 - ◆ Wybrano losowo 28 mieszkańców
 - ◆ Podzielono na 4 grupy po 7 osób
 - Pierwsza grupa nie zmieniała diety
 - Druga jadła same makarony
 - Trzecia — mięso
 - Czwarta — owoce
 - ◆ po miesiącu zmierzono rzut serca

Wyniki

Wprowadzenie

Wizualizacja

Statystyka opisowa

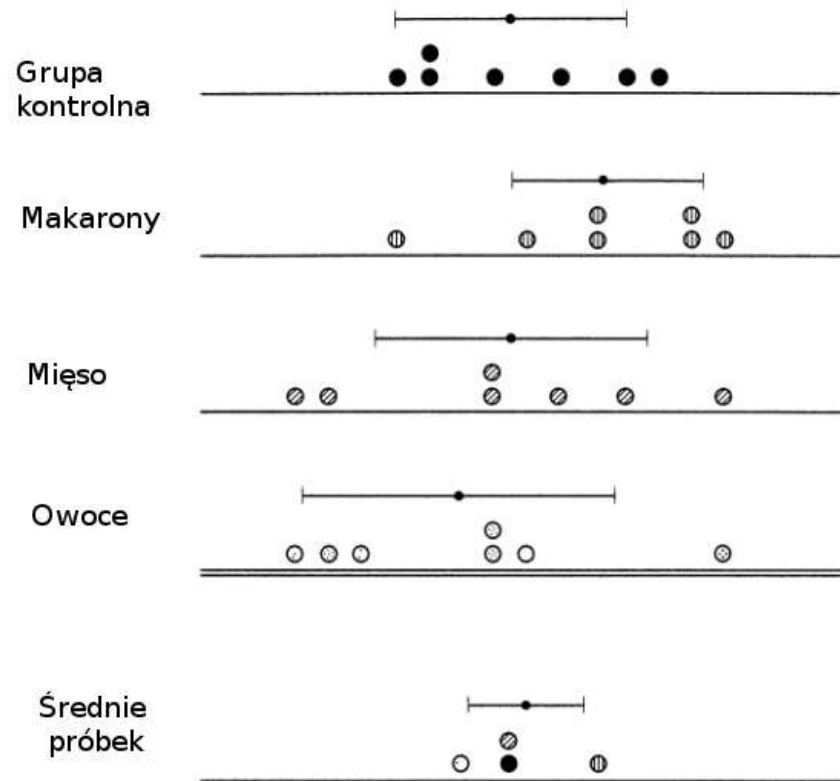
Studium Przypadku

❖ Hipotezy statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Czy dieta ma wpływ na rzut serca?
- Hipoteza zerowa: nie ma wpływu.
- Jakie jest prawdopodobieństwo wylosować takie wyniki?

Całe miasteczko

Wprowadzenie

Wizualizacja

Statystyka opisowa

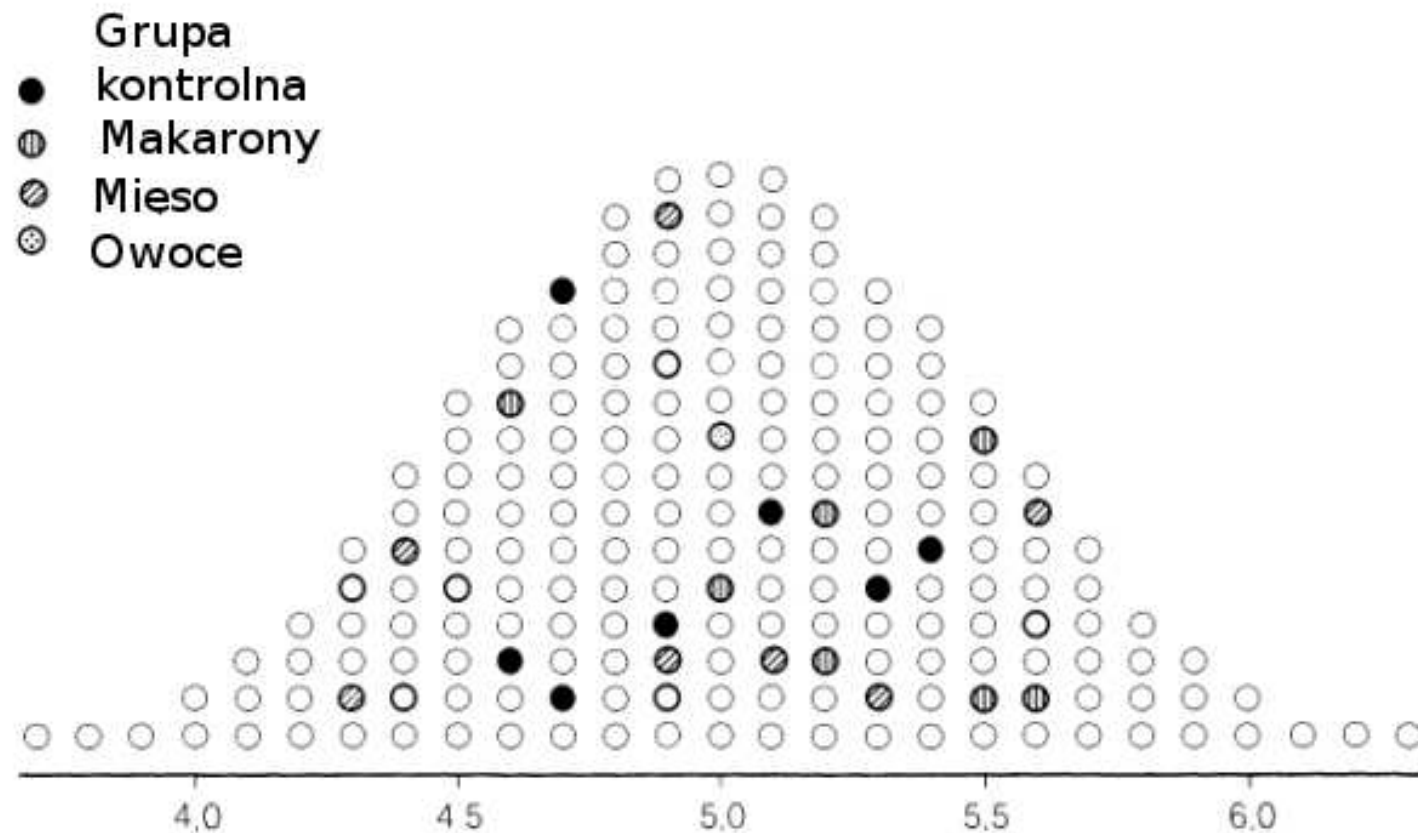
Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Nasze grupy są losowymi próbkami z całości

Czy „poprawione” wyniki się różnią?

Wprowadzenie

Wizualizacja

Statystyka opisowa

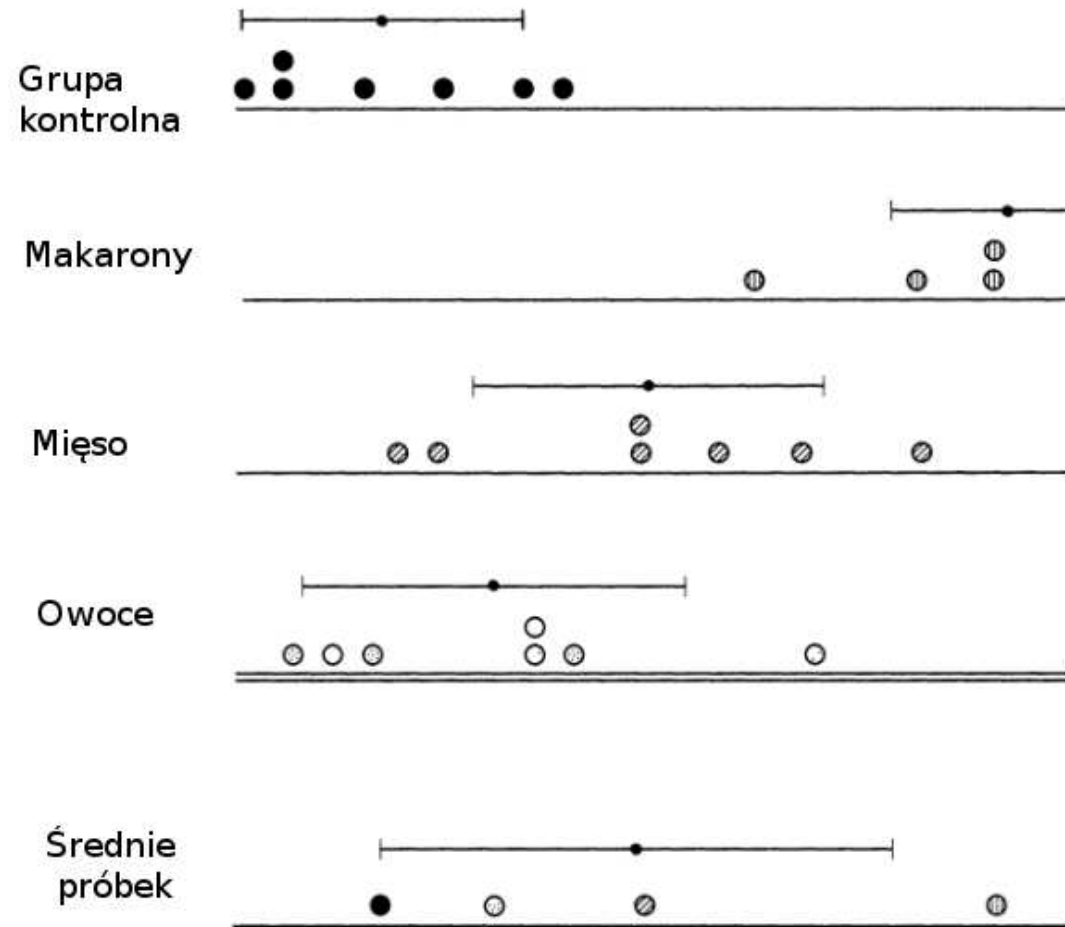
Studium Przypadku

❖ Hipotezy statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Czyba się różnią?

Porównywanie

Wprowadzenie

Wizualizacja

Statystyka opisowa

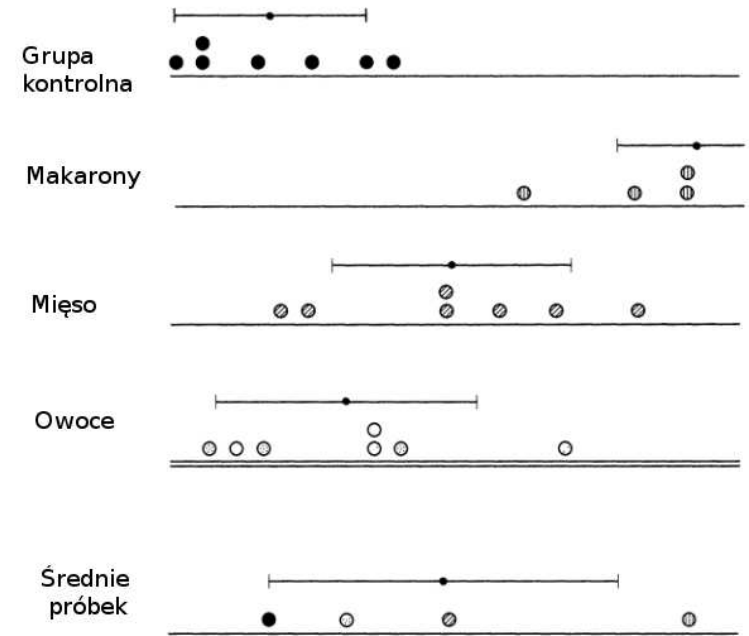
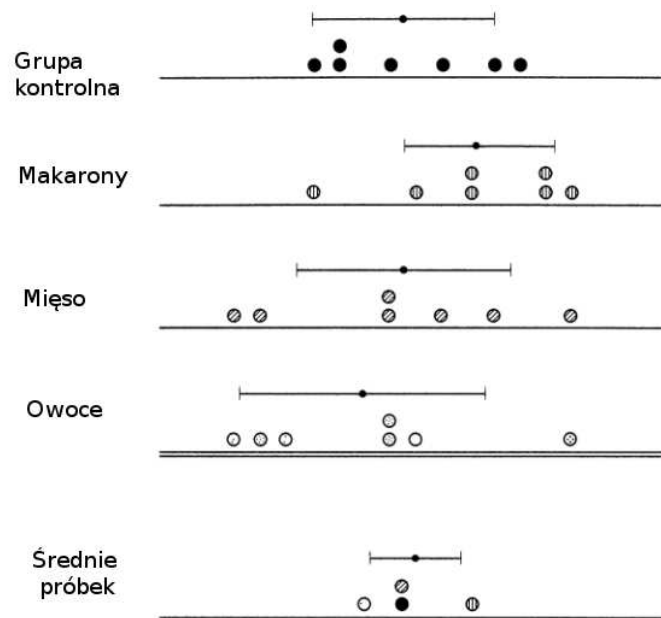
Studium Przypadku

❖ Hipotezy statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Czemu uważamy, że różnice w drugim przypadku są istotne?

Analiza wariancji

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

- Porównujemy wariancje
- Założenie: rozkład normalny
- Kryterium parametryczne

Dwa sposoby na obliczenie oceny wariancji

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ **Analiza wariancji**

❖ Test f

❖ R

- Na podstawie wariancji grup (wariancja wewnętrzna)

$$s_{\text{wewn}}^2 = \frac{1}{4} \left(s_{\text{kontr}}^2 + s_{\text{makar}}^2 + s_{\text{mięso}}^2 + s_{\text{owoce}}^2 \right)$$

- Na podstawie wariancji oceny średnich (wariancja międzygrupowa)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \Rightarrow s_{\text{mg}}^2 = N s_{\bar{X}}^2$$

- Jeżeli grupy są losowymi podzbiorami rozkładu normalnego, to obydwie oceny dają przybliżono ten sam wynik.

Test f (test Fishera)

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

- $F = \frac{s_{mg}^2}{s_{wewn}^2}$
 - ◆ jeżeli $F \approx 1$, to hipotezę zerową należy przyjąć
 - ◆ jeżeli $F \gg 1$ (znacznie więcej), to hipotezę zerową należy odrzucić
 - Dla danych eksperymentu F jest blisko do jedynki
 - Dla danych „poprawionych” $F = 68,0$.

Wartość krytyczna F

Wprowadzenie

Wizualizacja

Statystyka opisowa

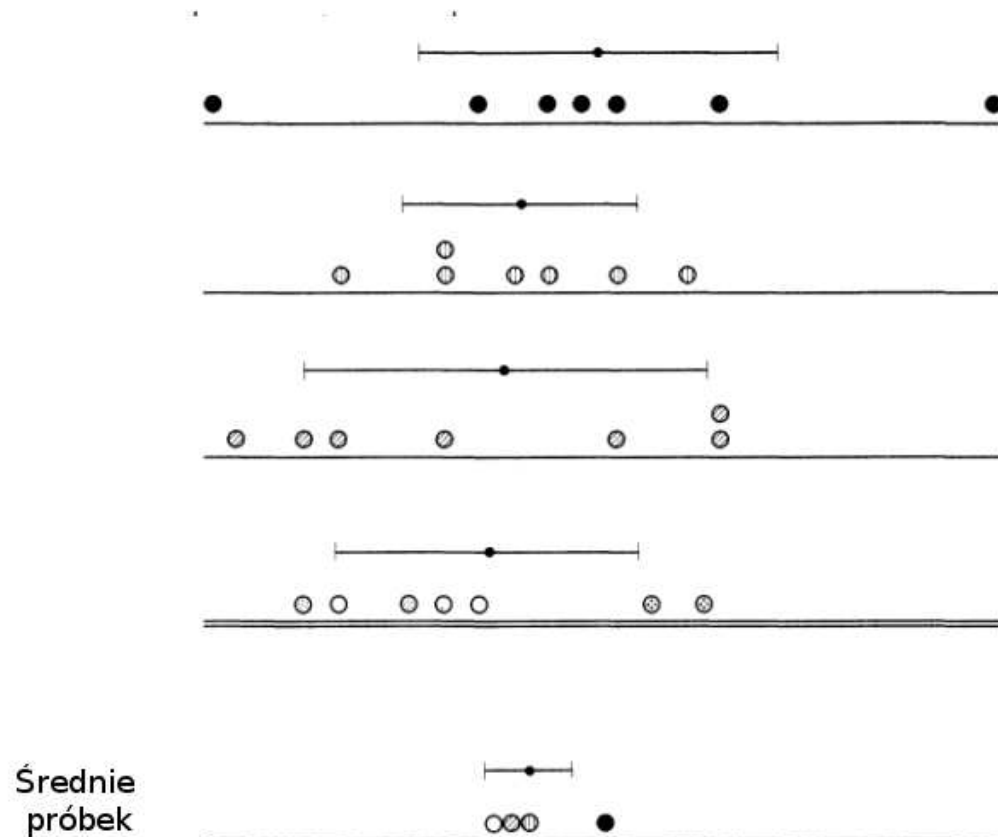
Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Wartości F dla losowych próbek będą się różnić
- Od jakiej wartości F należy uważać za dużą?

Obliczymy F dla 200 próbek

Wprowadzenie

Wizualizacja

Statystyka opisowa

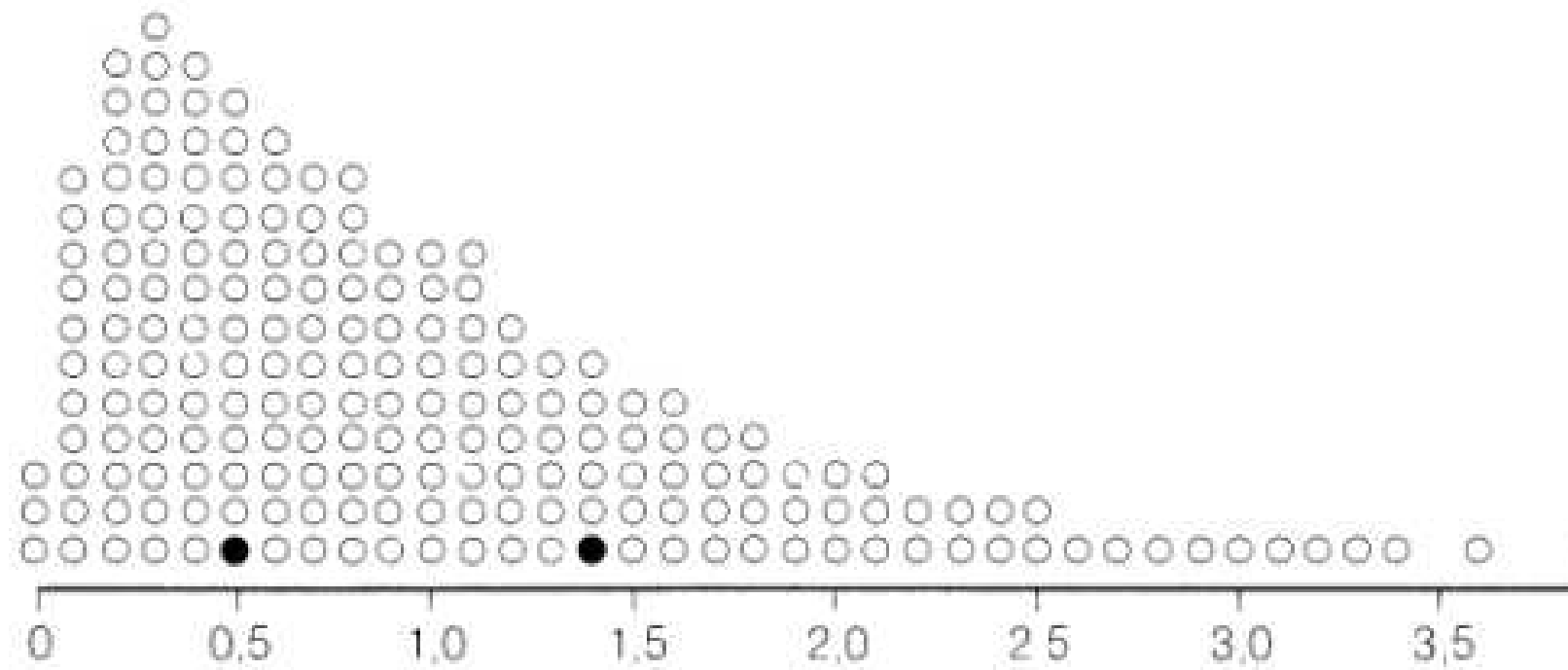
Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Dla większości próbek F jest około jedynki (od 0 do 2)

Obliczymy F dla 200 próbek

Wprowadzenie

Wizualizacja

Statystyka opisowa

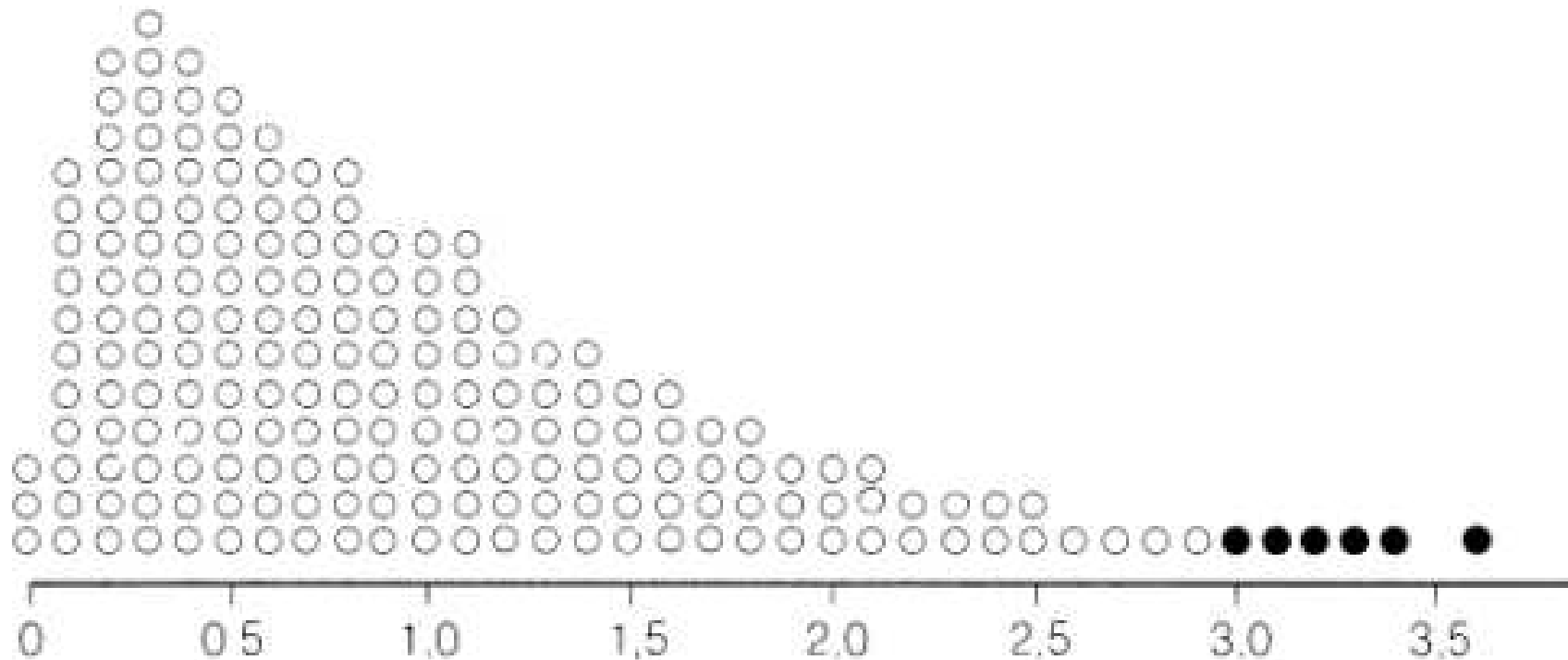
Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Dla dziesięciu próbek $F \geq 3$
- Jeżeli będziemy odrzucać hipotezę zerową dla $F \geq 3$, to pomylimy się tylko w 5% przypadków
- $F = 3$ nazywa się *wartością krytyczną*
- $\alpha = 0,05$ (5%) nazywa się *poziomem ufności*

Obliczymy F dla wszystkich 7-osobowych próbek

Wprowadzenie

Wizualizacja

Statystyka opisowa

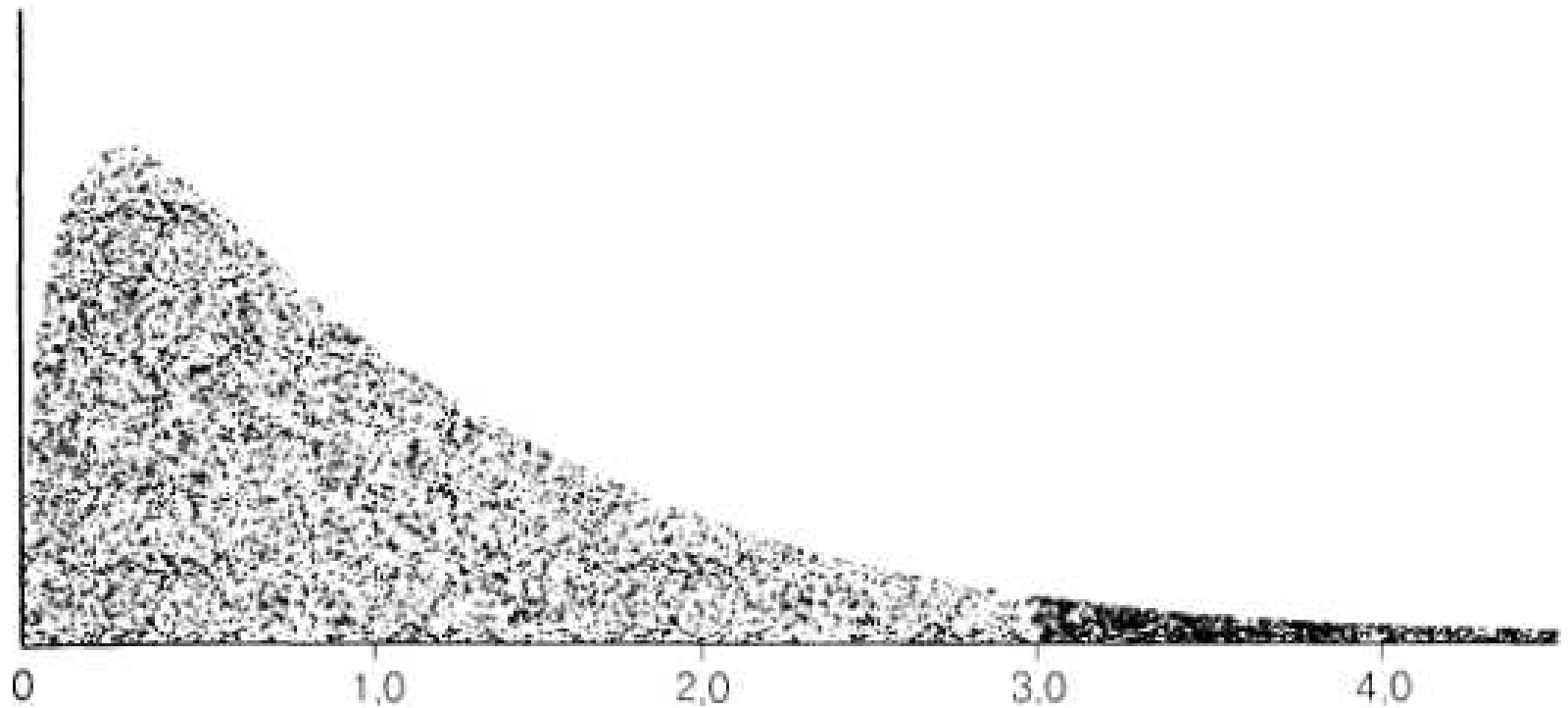
Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R



- Wartość krytyczna wynosi około 3,01 na poziomie ufności 5%

Gdyby liczebność całej populacji dążyła do ∞

Wprowadzenie

Wizualizacja

Statystyka opisowa

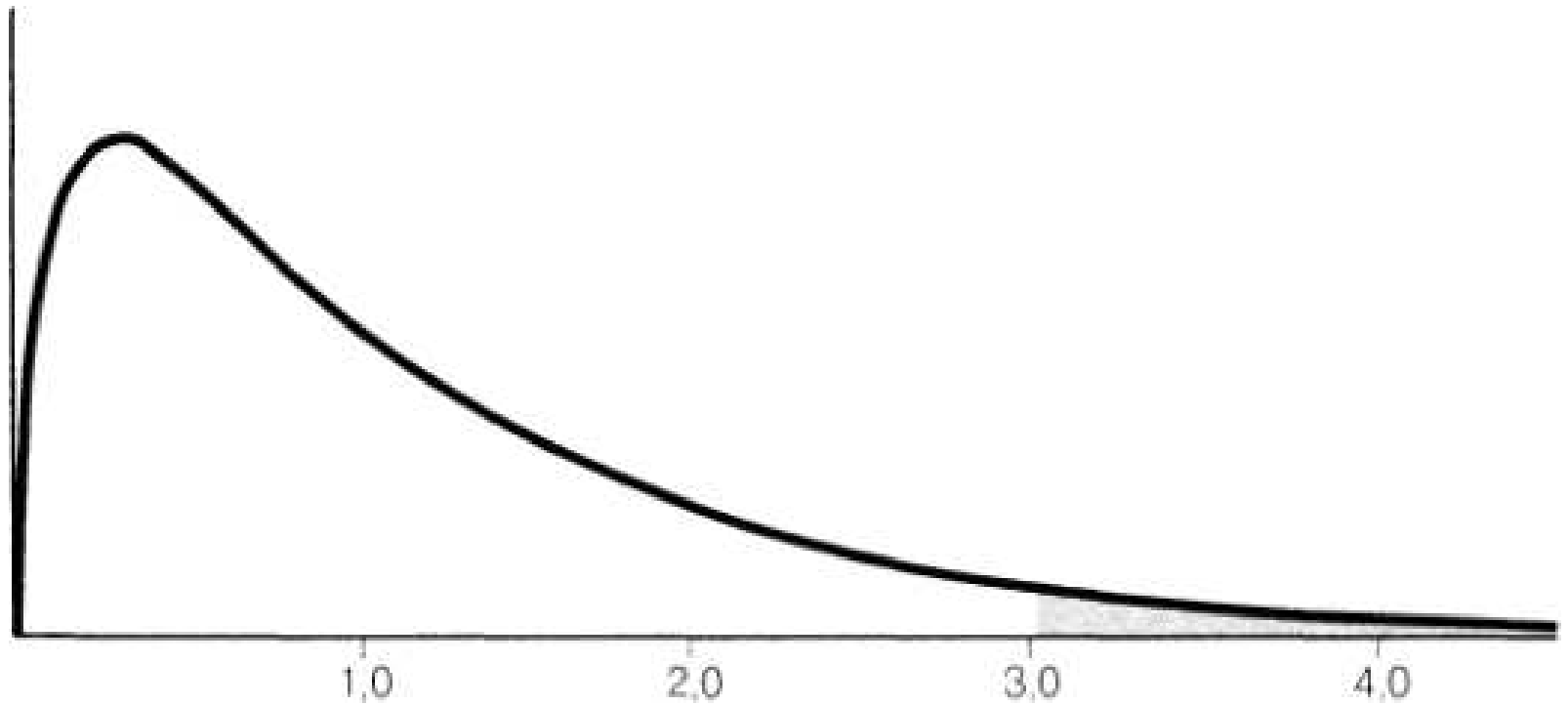
Studium Przypadku

❖ Hipotezy statystyczne

❖ Analiza wariancji

❖ **Test f**

❖ R



- Wartość krytyczna wynosi 3,01 na poziomie ufności 5%

Ogólny przypadek

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

- Wartość krytyczna zależy od
 - ◆ poziomu ufności (5%, 1%)
 - ◆ liczby stopni swobody $\nu_1 = m - 1$, $\nu_2 = m(n - 1)$,
gdzie m jest ilością grup, n jest liczebnością każdej grupy
- Wartość krytyczna może być znaleziona z tablic $F_\alpha(\nu_1, \nu_2)$.
- Przykład tablicy:
[NIST/SEMATECH e-Handbook of Statistical Methods](#)
- W naszym przykładzie 4 grupy po 7 osób

Obliczenia za pomocą R. Dane

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

```
cardiac <-  
data.frame(output=  
c(4.6, 4.7, 4.9, 4.6, 5.4, 5.1, 5.3, # kontrolna  
4.6, 5.0, 5.2, 5.2, 5.6, 5.5, 5.5, # makarony  
5.6, 5.3, 4.9, 5.2, 4.9, 4.4, 4.3, # mięso  
4.9, 4.4, 4.8, 4.5, 4.9, 4.8, 5.6), # owoce  
diet = rep(c("Grupa kontrolna",  
"Makarony", "Mięso", "Owoce"),  
           c(7, 7, 7, 7)))
```

Obliczenia i analiza

Wprowadzenie

Wizualizacja

Statystyka opisowa

Studium Przypadku

❖ Hipotezy
statystyczne

❖ Analiza wariancji

❖ Test f

❖ R

```
summary(aov(output ~ diet, data = cardiac))
              Df Sum Sq Mean Sq F value Pr(>F)
diet           3  0.581   0.1937   1.282  0.303
Residuals    24  3.626   0.1511
```

- Wartość testu f wynosi 1,282
- Przewidywanie wylosować taki wynik wynosi 0,303, i.e. ponad 30%
- Hipotezę zerową należy przyjąć:
 - ◆ dieta nie ma wpływu na rzut serca