statistical models

Katarzyna Tokarczuk

2024-04-17

Task 1

```
library(mosaic)
```

```
## Warning: pakiet 'mosaic' został zbudowany w wersji R 4.3.3
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
```

```
##
## Dołączanie pakietu: 'mosaic'
```

```
## Następujące obiekty zostały zakryte z 'package:dplyr':
##
##     count, do, tally
```

```
## Następujący obiekt został zakryty z 'package:Matrix':
##
##     mean
```

```
## Następujący obiekt został zakryty z 'package:ggplot2':
##
##     stat
```

```
## Następujące obiekty zostały zakryte z 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var
```

```
## Następujące obiekty zostały zakryte z 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```r
data(airquality)
```

Task 2

```r
wartosc_brakujaca <- sum(is.na(airquality))
wartosc_brakujaca
```

```
## [1] 44
```

Task 3

```r
dane <- na.omit(airquality)
```

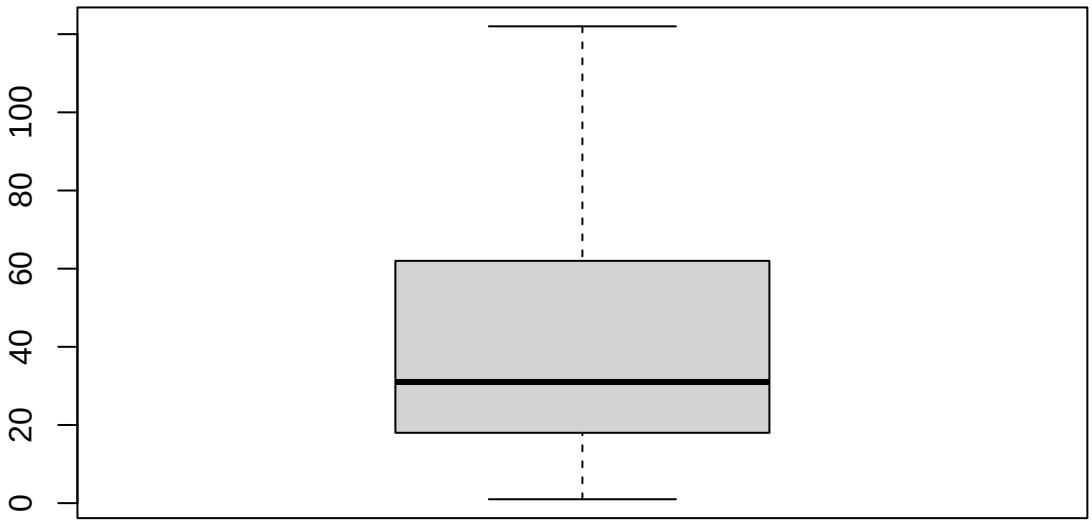
Task 4

```r
str(dane)
```

```
## 'data.frame':    111 obs. of  6 variables:
##  $ Ozone  : int  41 36 12 18 23 19 8 16 11 14 ...
##  $ Solar.R: int  190 118 149 313 299 99 19 256 290 274 ...
##  $ Wind   : num  7.4 8 12.6 11.5 8.6 13.8 20.1 9.7 9.2 10.9 ...
##  $ Temp   : int  67 72 74 62 65 59 61 69 66 68 ...
##  $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Day    : int  1 2 3 4 7 8 9 12 13 14 ...
##  - attr(*, "na.action")= 'omit' Named int [1:42] 5 6 10 11 25 26 27 32 33 34 ...
##   ..- attr(*, "names")= chr [1:42] "5" "6" "10" "11" ...
```

Task 5

```r
boxplot(dane$Ozone, outline = FALSE)
```



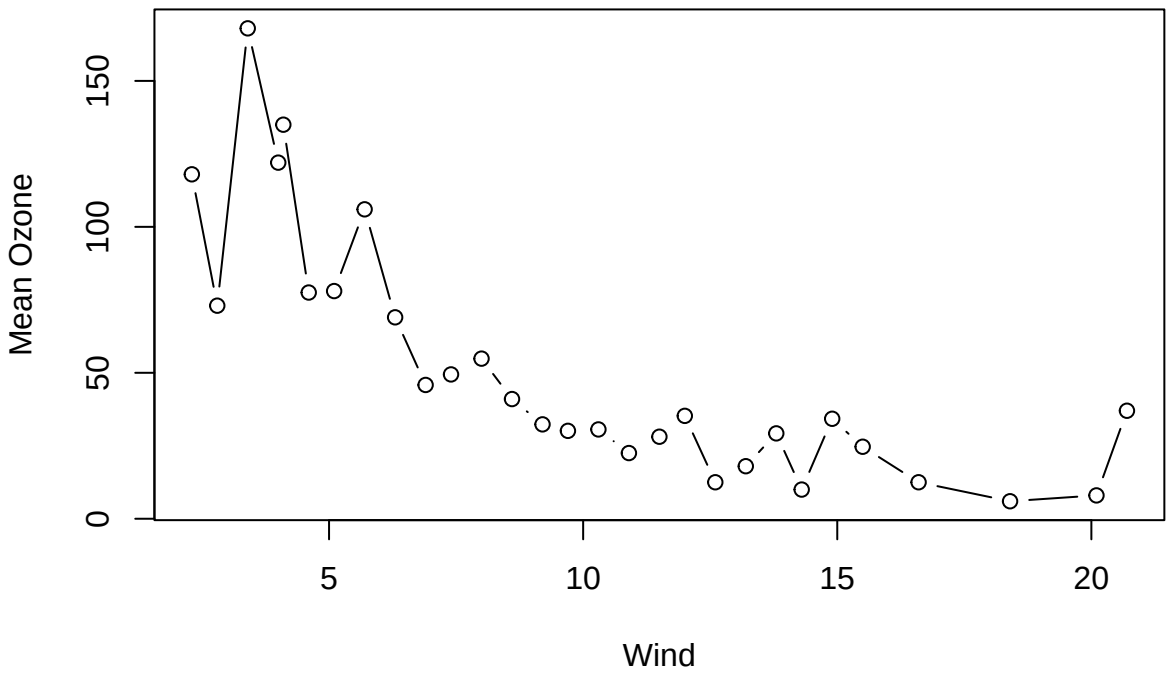
Task 6

```
mosaic::mean(Ozone ~ Wind, data = dane)
```

```
##       2.3       2.8       3.4         4       4.1       4.6       5.1       5.7
## 118.00000  73.00000 168.00000 122.00000 135.00000  77.50000  78.00000 106.00000
##       6.3       6.9       7.4         8       8.6       9.2       9.7      10.3
##  69.00000  45.83333  49.44444  54.85714  41.00000  32.33333  30.11111  30.60000
##      10.9      11.5        12      12.6      13.2      13.8      14.3      14.9
##  22.50000  28.10000  35.25000  12.50000  18.00000  29.25000  10.00000  34.25000
##      15.5      16.6      18.4      20.1      20.7
##  24.66667  12.50000   6.00000   8.00000  37.00000
```

Task 7

```
agg_srednia <- aggregate(Ozone ~ Wind, data = dane, FUN = mean)
plot(agg_srednia$Wind, agg_srednia$Ozone, type = "b", xlab = "Wind", ylab = "Mean Ozone")
```



Task 8

```
dane <-mutate(dane, Temp = (dane$Temp - 32) * 5/9)
```

Task 9

```
#a)
monthly_data <- split(dane, dane$Month)

#b)
wspol_regresji <- data.frame(Month = integer(), Intercept = numeric(), Slope = numeric(), R_squared = nu

#c)
for (i in names(monthly_data)) {
  lm_model <- lm(Temp ~ Ozone, data = monthly_data[[i]])
  Intercept <- coef(lm_model)[1]
  Slope <- coef(lm_model)[2]
  R_squared <- summary(lm_model)$r.squared
  wspol_regresji[i, ] <- c(i, Intercept, Slope, R_squared)
}

wspol_regresji
```

```
##   Month        Intercept              Slope         R_squared
## 5     5  16.764137104637 0.0986272088655536 0.375181362415946
## 6     6 20.9723967788484  0.159847321137644 0.446676429211574
## 7     7 25.4942890770753 0.0563389254655086 0.522298653555557
## 8     8  25.310913242354 0.0568148920155173 0.366616467240946
## 9     9 19.8461157034107  0.162056993349148 0.685835926788777
```

```
#d) Współczynnikiem determinacji dla każdego modelu jest zmienna R_squared
```