Wykład 4 – notatki

Maciej Beręsewicz

Na dzisiejszych zajęciach zajmiemy się analizą spółek giełdowych, wśród których znajdowały się spółki będące w stanie upadłości. Zadaniem naszym jest zweryfikowanie, które zmienne mogą być użyte do przewidywania oraz czy są jakieś graniczne wartości wykorzystanych zmiennych, po których można takie spółki rozróżnić.

Aby wstawić miejsce, w którym możemy pisać kody należy użyć skrótu CTRL + ALT + I.

library(readxl)  
#install.packages("tidyverse") ## kolekcja pakietów do analizy danych  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 3.6.2

## Warning: package 'tibble' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'readr' was built under R version 3.6.2

## Warning: package 'purrr' was built under R version 3.6.2

## Warning: package 'dplyr' was built under R version 3.6.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Zacznijmy od wczytania danych o spółkach

spolki <- read\_excel(path = "../data-raw/spolki\_akcyjne\_bankrut.xlsx",  
 col\_names = c("id", "roa", "zysk\_bilans", "wsk\_bie\_ply", "zob\_akt", "rotacja", "grupa"),   
 skip = 1)  
spolki

## # A tibble: 180 x 7  
## id roa zysk\_bilans wsk\_bie\_ply zob\_akt rotacja grupa   
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 1 0.0210 0.0590 0.941 0.665 132. bankrut  
## 2 2 -0.247 -0.0760 0.508 0.933 507. bankrut  
## 3 3 -0.0159 0.0236 1.21 0.601 367. bankrut  
## 4 4 0.174 0.151 1.71 0.480 170. bankrut  
## 5 5 0.0502 0.0676 1.04 0.739 219. bankrut  
## 6 6 0.0192 0.0869 1.13 0.643 363. bankrut  
## 7 7 -0.136 -0.112 1.28 0.457 268. bankrut  
## 8 8 -0.0678 -0.0109 0.861 0.801 124. bankrut  
## 9 9 -0.206 -0.173 0.943 0.817 338. bankrut  
## 10 10 0.0118 0.0699 1.16 0.733 121. bankrut  
## # … with 170 more rows

Robimy proste podsumowanie aby sprawdzić jak dużo mamy spółek zdrowych i chorych. Zrobimy to na dwa sposoby:

1. z wykorzystaniem funkcji table z “czystego R”

table(spolki$grupa)

##   
## bankrut zdrowa   
## 90 90

1. z wykorzystaniem pakietu dplyr i funkcji count

count(spolki, grupa, name = "liczba")

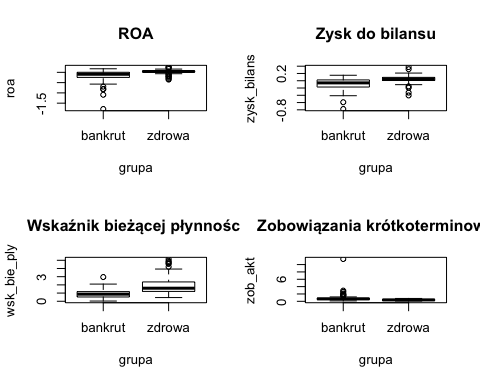
## # A tibble: 2 x 2  
## grupa liczba  
## <chr> <int>  
## 1 bankrut 90  
## 2 zdrowa 90

Mamy następujące hipotezy, które będziemy weryfikowali

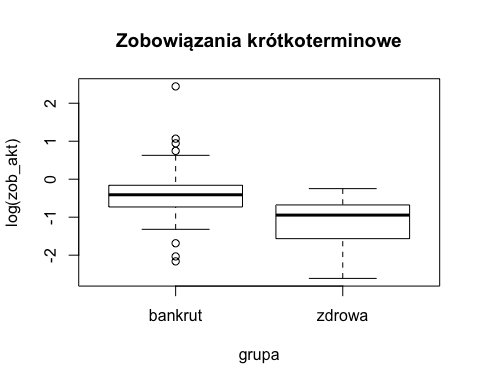
1. wsk\_bie\_ply jeżeli jest < 1.2 -> możliwy wskaźnik bankructwa
2. zob\_akt im wyższa wartość tym “większe problemy”
3. roa i zysk\_bilans < 0 -> może nas niepokoić

Jeżeli chcemy sprawdzić czy dwie lub więcej grup różnią się ze względu na jakieś zmienne, najlepiej użyć wykres pudełkowy (boxplot)

par(mfrow = c(2, 2))  
boxplot(roa ~ grupa, data = spolki, main = "ROA")  
boxplot(zysk\_bilans ~ grupa, data = spolki, main = "Zysk do bilansu")  
boxplot(wsk\_bie\_ply ~ grupa, data = spolki, main = "Wskaźnik bieżącej płynności")  
boxplot(zob\_akt ~ grupa, data = spolki, main = "Zobowiązania krótkoterminowe")

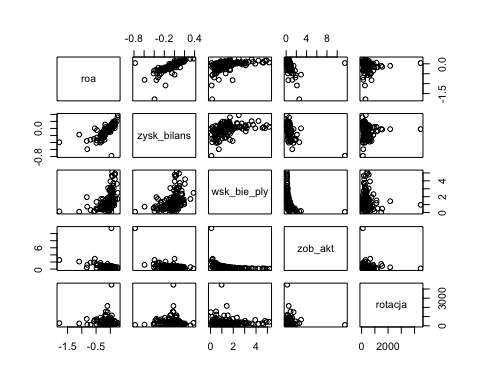


boxplot(log(zob\_akt) ~ grupa, data = spolki, main = "Zobowiązania krótkoterminowe")



Sprawdźmy jak wyglądają zależności między danymi wykorzystując wykres rozrzutu

pairs(spolki[, 2:6])



round(cor(spolki[, 2:6], method = "spearman"), 2)

## roa zysk\_bilans wsk\_bie\_ply zob\_akt rotacja  
## roa 1.00 0.85 0.58 -0.44 -0.28  
## zysk\_bilans 0.85 1.00 0.48 -0.38 -0.28  
## wsk\_bie\_ply 0.58 0.48 1.00 -0.71 -0.06  
## zob\_akt -0.44 -0.38 -0.71 1.00 -0.20  
## rotacja -0.28 -0.28 -0.06 -0.20 1.00