Vilniaus Universitetas

Cenzuruotų imčių analizė

Laboratorinis darbas

Darbą atliko:

Matas Gaulia, Matas Kamarauskas

Duomenų Mokslas

4 kursas 1 gr.

Vilnius, 2022

# Naudoti metodai

Darbas atliktas naudojant R.

Naudoti R paketai:

*tidyverse*

*dplyr  
survival  
ggfortify  
ggplot2  
muhaz*

# Duomenys ir jų šaltiniai

Duomenys apie žmones sergančius COVID-19 ir jų išgyvenamumą.

Duomenų šaltinis - Kaggle. Prieiga per internetą: [Lung-Survival Dataset | Kaggle](https://www.kaggle.com/datasets/aysuncag/covid19-survival-dataset)

„time“ – laikas iki mirties dienomis.

„sex“ – lytis, 1 – vyras, 2 - moteris

„status“ – statusas, 1 – cenzūruotas, 2 - mirtis

„age“ - amžius

„ph.ecog“ – Neįgalumo statusas

0 – pilnos judėjimo galimybės

1 – galimi tik fiziškai lengvi darbai

2 – gali pasirūpinti savimi bet neįgalus dirbti

3 – ne pilnai gali savimi pasirūpinti

4 – pilnas neįgalumas

# Tikslas ir uždaviniai

Tikslas: Ištirti COVID-19 išgyvenamumą ir patikrinti homogeniškumo hipotezes.

Uždaviniai:

Atlikti pirminę duomenų analizę

Ištriti gyvenamumą pagal lytį ir neįgalumo statusą

Patikrinti homogeniškumo hipotezes

# Atliktos analizės aprašymas

Duomenų aibę sudaro duomenys 227 stebėjimai  
Lyčių pasiskirstymas: 137 vyrai ir 90 moterys.  
Cenzūravimo pasiskirstymas: 63 cenzūruoti, 164 – ne.  
Neįgalumo pasiskirstymas:   
0 – 63 stebėjimai  
1 – 113 stebėjimai  
2 – 50 stebėjimai  
3 – 1 stebėjimai  
4 – 0 stebėjimų

d <- read\_csv("SurvivalCovid.csv") %>%   
 select("time", "sex", "ph.ecog", "status") %>%   
 drop\_na()

d$time <- as.integer(d$time)

d$ph.ecog <- as.factor(d$ph.ecog)

d$sex <- as.factor(d$sex)

d$status <- as.integer(d$status)

#perkoduojame stulpeli (buvo 1-cenzuruota, dabar 0-cenzuruota)

d$status <- d$status - 1

NROW(d)  
227

summary(as.factor(d$status))

0 63

1 164

summary(as.factor(d$ph.ecog))0 63

1 113

2 50

3 1

# Išgyvenamumas

s <- Surv(d$time, d$status)

autoplot( survfit(s ~ 1),   
 censor.shape = '\*',   
 censor.size = 10,   
 surv.colour = 'red',   
 censor.colour = 'blue')*Chart

Description automatically generated*

# Išgyvenamumas pagal lytį

autoplot(survfit(s ~ d$sex),   
 censor.shape = '\*',   
 censor.size = 10) +

scale\_color\_manual(values = c("blue","red")) +

scale\_fill\_manual(values = c("blue","red"))  
*Chart

Description automatically generated*

Matome kad vyrų išgyvenamumas mažesnis nei moterų bet kuriuo laiko momentu.

# Sukauptos rizikos įvertis

Iverciai <- summary(survfit(s ~ 1, type="kaplan-meier"))

Lambda\_hat <- (-1)\*log(Iverciai$surv)

Lambda\_hat <- c(Lambda\_hat, tail(Lambda\_hat, 1))

plot(c(Iverciai$time, 900), Lambda\_hat,

xlab="time",

ylab="Lambda\_hat",

main="Cumulative hazard",   
 ylim=range(Lambda\_hat),

type="s") *Chart, line chart

Description automatically generated*

Matome kad rizika gana pastoviai auga.

print(survfit(s ~ 1), print.rmean=TRUE)

Call: survfit(formula = s ~ 1)

n events rmean\* se(rmean) median 0.95LCL 0.95UCL

[1,] 227 164 378 19.7 310 285 363

\* restricted mean with upper limit = 1022

# Branduolinis rizikos įvertis

result.simple <- muhaz(d$time,   
 d$status,   
 max.time=900,   
 bw.method="global",   
 b.cor="none")

plot(result.simple)

*Chart, line chart, histogram

Description automatically generated*Išgyvenamumas pagal neigalumą

ecog <- s~d$ph.ecog

colors <- c("orange", "green", "red", "blue", "black")

plot(survfit(ecog),

xlab="Time in days",

ylab="Survival probability",

col=colors,

lwd =2)

legend("topright",

legend=c("Good-0", "1", "2", "3", "Bad-4"),

col=colors, lwd =2)

Chart, line chart, histogram

Description automatically generated

Matome kad gerokai skiriasi išgyvenamumas pagal neįgalumo grupes, blogiausiai grupei (pilnas neįgalumas) nepriklausė nei vienas pacientas, o trečiąjai grupei priklausė tik vienas pacientas.

# Homogeniškumo hipotezių tikrinimas

Lograginis kriterijus:

survdiff(ecog, rho=0)

Table

Description automatically generated

Matome, kad p-reikšmė yra mažiau nei 0.05, tad galime teigti kad egzistuoja statistiškai skirtingas skirtumas tarp grupių išgyvenamumo.

Gehan-Wilcoxon kriterijaus Peto ir Peto modifikacija:

survdiff(ecog, rho=1)

Table

Description automatically generated with medium confidence

Matome, kad p-reikšmė yra mažiau nei 0.05, tad galime teigti kad egzistuoja statistiškai skirtingas skirtumas tarp grupių išgyvenamumo.

# Rezultatai

Remiantis atliktu tyrimu galime teigti kad lytis ir neįgalumo statusas turi įtakos žmogaus išgyvenamume nuo COVID-19 ligos. Gauti rezultatai kad vyrų mirtingumas didesnis nei moterų, taip pat didesnis neįgalumas daro neigiamą įtaką neįgalumui.