COVID-19 plitimas valstybėse

Laboratorinis darbas

Antanas Užpelkis, Matas Amšiejus

Vilniaus Universitetas

2020-04-03

Duomenys yra paimti iš EU Open Data Portal apie koronaviruso plitimą skirtingose pasaulio valstybėse. Duomenys kaupti nuo 2019 gruodžio 31 iki dabar, tačiau mes failą parsisiuntėme gegužės 10 dieną. Duomenys yra tokie: data(pilna), diena, mėnuo, metai, užsikrėtimų tą dieną skaičius vienetais, mirusių nuo koronaviruso skaičius tą dieną (vienetais), valstybė, valstybės geo ID, valstybės kodas, pupuliacija 2018 metų duomenimis, žemynas (pastaba: Š ir P Amerikos čia imamos kaip vienas žemynas Amerika). Duomenų eilučių yra 16741 (mūsų faile) ir kasdien pildoma.

1. Kiekybinė imtis

Analizavome vienos konkrečios dienos (2020 m. gegužės 10 d.) duomenis. Duomenis analizavome paėmę ligos atvejų skaičių 1000 gyventojų (atvejų skaičius / populiacija \* 1000).

*Duomenų nuskaitymas:*

*Pastaba:atsakymai, kuriuose turėtų būti rodomi visi imties duomenys bus užrašomi 10 mažiausių ir 10 didžiausių reikšmių(duomenys kaip variacinė eilutė, duomenų rušiavimas, rangai it t.t.).*

Atsisiuntėme duomenis .csv formatu.

duom<-read.csv("D:/Users/User/Desktop/studijos/pirmas kursas/TiRiamoji duomenu analize/covid.txt") #duom- nekeisti, pagrindiniai duomenys

duomenys<-subset(duom,month==5&day==10,select = c(cases, deaths, countriesAndTerritories, continentExp, popData2018))

#Atsirenkame tik gegužės 10 dieną su mums reikalingais stulpeliais

names(duomenys)<-c("susirgimai","mirtys", "salis", "zemynas", "gyventojai")

#Pakeičiame stulpelių pavadinimus į lietuviškus

duomenys<-duomenys[complete.cases(duomenys), ]

#Atmetame visas tuščias reikšmes, atsispausdiname tai, ką pasilikome:

susirgimai mirtys salis zemynas gyventojai

4 255 6 Afghanistan Asia 37172386

129 6 0 Albania Europe 2866376

195 189 6 Algeria Africa 42228429

325 2 1 Andorra Europe 77006

386 0 0 Angola Africa 30809762

487 0 0 Antigua\_and\_Barbuda America 96286

duomenys3<-duomenys

duomenys3$susirgimai<-round(duomenys3$susirgimai/duomenys3$gyventojai\*1000, digits = 3)

#pakeičiame susirgimų reikšmes į susirgimus tūkstančiui gyventojų

susirgimai mirtys salis zemynas gyventojai

4 0.007 6 Afghanistan Asia 37172386

129 0.002 0 Albania Europe 2866376

195 0.004 6 Algeria Africa 42228429

325 0.026 1 Andorra Europe 77006

386 0.000 0 Angola Africa 30809762

487 0.000 0 Antigua\_and\_Barbuda America 96286

2.

min(duomenys3$susirgimai);max(duomenys3$susirgimai)

# Randame didžiausią ir mažiausią susirgimų reikšmę 1000 gyv.

[1] 0

[1] 0.414

surusiuoti3<-sort(duomenys3$susirgimai);head(surusiuoti3, 10);tail(surusiuoti3, 10)

#sudarome variacinę eilutę

[1] 0 0 0 0 0 0 0 0 0 0

[1] 0.075 0.078 0.099 0.100 0.100 0.134 0.210 0.283 0.406 0.414

mazeja3<-sort(duomenys3$susirgimai, decreasing = T);head(mazeja3, 10);tail(mazeja3, 10)

# susirgimai mažėjimo tvarka

[1] 0.414 0.406 0.283 0.210 0.134 0.100 0.100 0.099 0.078 0.075

[1] 0 0 0 0 0 0 0 0 0 0

rangai3<-rank(duomenys3$susirgimai);head(rangai3, 10);tail(rangai3, 10) # rangai

[1] 141.5 103.5 122.0 178.0 38.5 38.5 122.0 184.0 38.5 86.5

[1] 38.5 195.0 147.5 103.5 103.5 38.5 38.5 38.5 131.5 38.5

3.

mean(duomenys3$susirgimai);mean(duomenys3$susirgimai,trim=0.1) # vidurkis ir nupjautinis vidurkis

[1] 0.0166601

[1] 0.005638037

median(duomenys3$susirgimai) #mediana

[1] 0.002

quantile(duomenys3$susirgimai, type=1) #kvantiliai

0% 25% 50% 75% 100%

0.000 0.000 0.002 0.011 0.414

summary(duomenys3$susirgimai) #kvantiliai, mediana, vidurkis naudojant funkciją summary

Min. 1st Qu. Median Mean 3rd Qu. Max.

1. 0.00000 0.00200 0.01666 0.01100 0.41400

4.

range(duomenys3$susirgimai) #imties plotis

[1] 0.000 0.414

var(duomenys3$susirgimai) #dispersija

[1] 0.002539829

sd(duomenys3$susirgimai) #standartinis nuokrypis

[1] 0.05039672

IQR(duomenys3$susirgimai) #tarpkvantilinis plotis

[1] 0.011

mean(abs(duomenys3$susirgimai-mean(duomenys3$susirgimai))) #vidutinis absoliutinis nuokrypis

[1] 0.02250499

mean(abs(be3-mean(be3))) # vidutinis absoliutinis nuokrypis be išskirčių (be3 bus randama žemiau(5.))

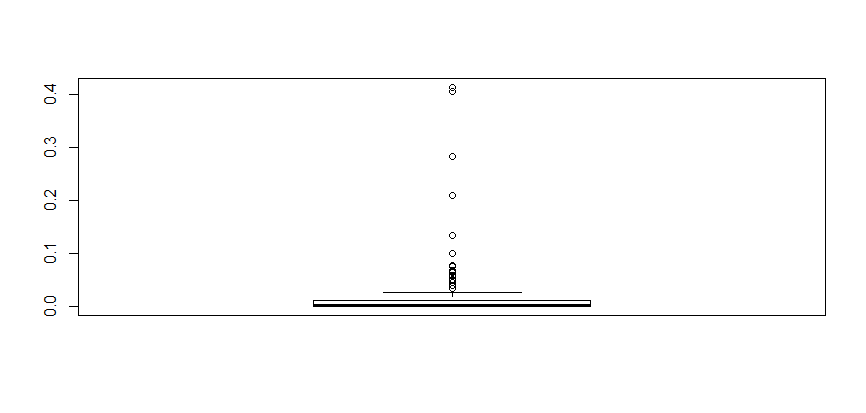
[1] 0.004740239

median(abs(duomenys3$susirgimai-median(duomenys3$susirgimai))) #medianinis absoliutinis nuokrypis

[1] 0.002

median(abs(be3-median(be3))) #medianinis absoliutinis nuokrypis be išskirčių

[1] 0.001

5.

isskirtys3 = boxplot(duomenys3$susirgimai)$out

which(duomenys3$susirgimai %in% isskirtys3) #išskirčių indeksai

[1] 8 14 17 18 26 34 39 53 55 92 103 116 122 144 147 152 154 159 161 166 178 192 193 195

be3<-duomenys3$susirgimai[!(duomenys3$susirgimai %in% isskirtys3)] #duomenys be pradinių išskirčių

[1] 0.007 0.002 0.004 0.026 0.000 0.000 0.004 0.000 0.001 0.005

6.

n<-nrow(duomenys3) #duomenų aibės dydis

[1] 203

h2<-max(duomenys3$susirgimai)/10 #gauname atstumą tarp 10 intervalų, kuriuos turėsime

[1] 0.0414

cc2<-seq(from=0,to=h2\*10,h2) #intervalų grupavimo galai

[1] 0.0000 0.0414 0.0828 0.1242 0.1656 0.2070 0.2484 0.2898 0.3312 0.3726 0.4140

v2<-cc2[-1]-h2/2 #intervalų vidurio taškai

[1] 0.0207 0.0621 0.1035 0.1449 0.1863 0.2277 0.2691 0.3105 0.3519 0.3933

nn<-table(cut(duomenys3$susirgimai, b=cc2, include.lowest = TRUE)) #intervalų dažnių lentelė

[0,0.0414] (0.0414,0.0828] (0.0828,0.124] (0.124,0.166] (0.166,0.207] (0.207,0.248] (0.248,0.29]

181 14 3 1 0 1 1

(0.29,0.331] (0.331,0.373] (0.373,0.414]

0 0 2

nnn<-cumsum(nn) #sukauptųjų dažnių lentelė

[0,0.0414] (0.0414,0.0828] (0.0828,0.124] (0.124,0.166] (0.166,0.207] (0.207,0.248] (0.248,0.29]

181 195 198 199 199 200 201

(0.29,0.331] (0.331,0.373] (0.373,0.414]

201 201 203

ff3<-nn3/n3 #santykinių dažnių lentelė

[0,0.0414] (0.0414,0.0828] (0.0828,0.124] (0.124,0.166] (0.166,0.207] (0.207,0.248]

0.891625616 0.068965517 0.014778325 0.004926108 0.000000000 0.004926108

(0.248,0.29] (0.29,0.331] (0.331,0.373] (0.373,0.414]

0.004926108 0.000000000 0.000000000 0.009852217

fff3<-cumsum(ff3) #sukauptųjų santykinių dažnių lentelė

[0,0.0414] (0.0414,0.0828] (0.0828,0.124] (0.124,0.166] (0.166,0.207] (0.207,0.248] (0.248,0.29]

0.8916256 0.9605911 0.9753695 0.9802956 0.9802956 0.9852217 0.9901478

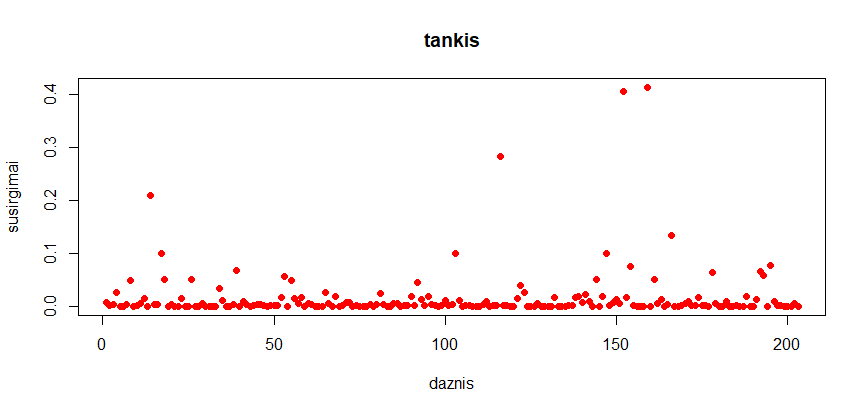
(0.29,0.331] (0.331,0.373] (0.373,0.414]

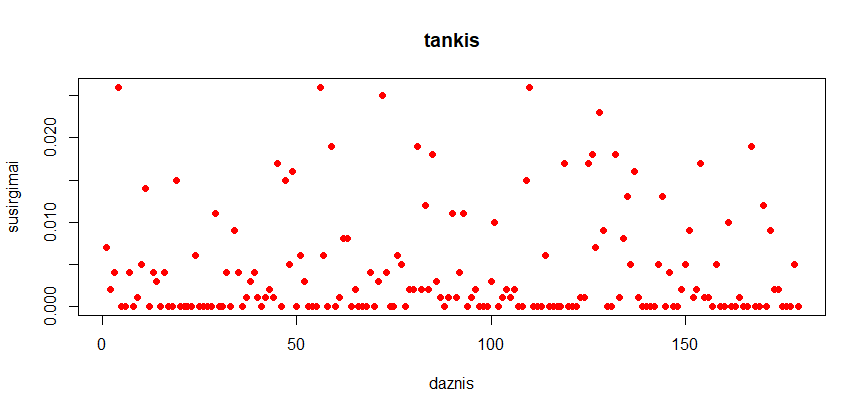
0.9901478 0.9901478 1.0000000

7.

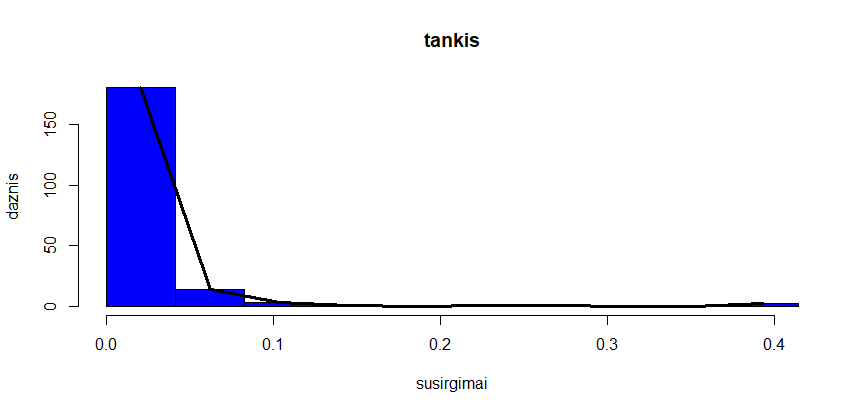
plot(x=duomenys3$susirgimai, main = "tankis", ylab = "susirgimai" , xlab = "daznis", col="red", pch = 19)

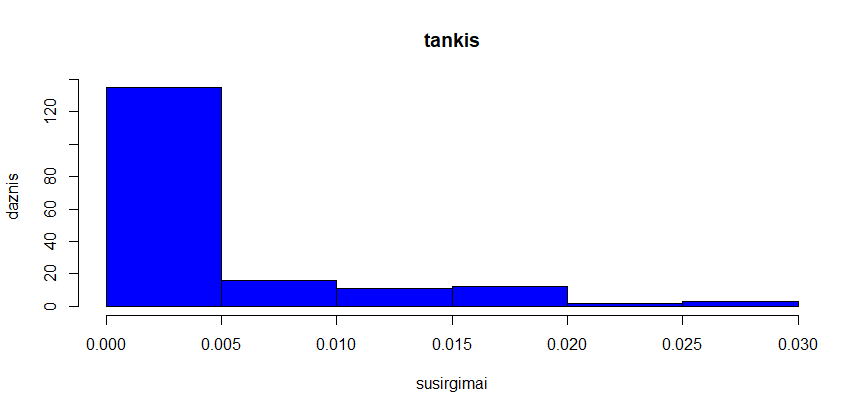
#empirinė dažnių lentelė

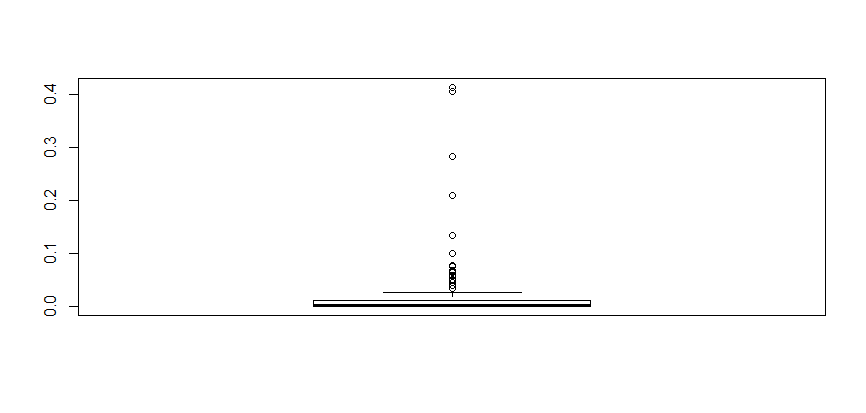


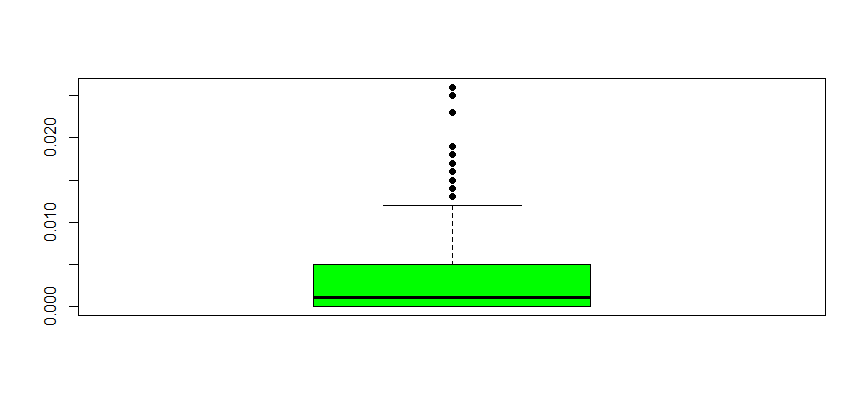
plot(be3, main = "tankis", ylab = "susirgimai" , xlab = "daznis", col="red", pch = 19) #empirinė dažnių lentelė be išskirčių

hist(duomenys3$susirgimai,breaks = cc2, col="blue", main = "tankis", ylab = "daznis" , xlab = "susirgimai");lines(v2,nn3,lwd=3) #histograma (pasiskirstymo dažnis)

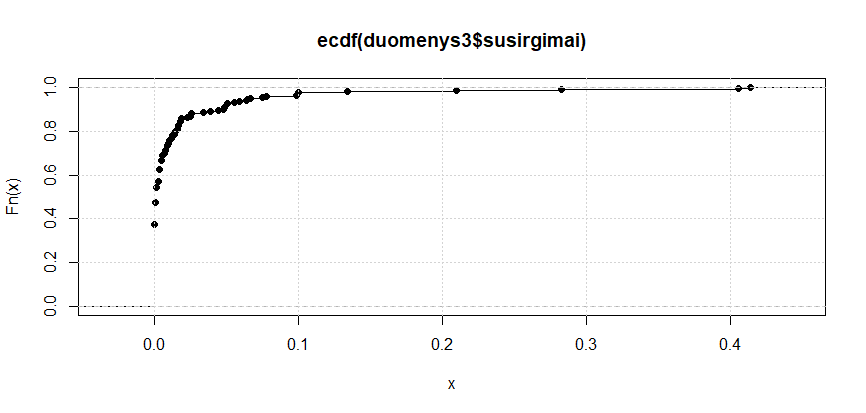


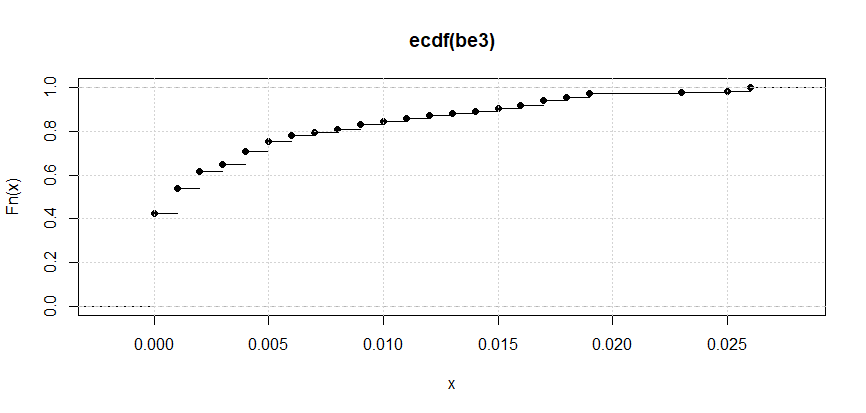
hist(be3, col="blue", main= "tankis", ylab = "daznis" , xlab = "susirgimai") #pasiskirstymo dažnis (be išskirčių)

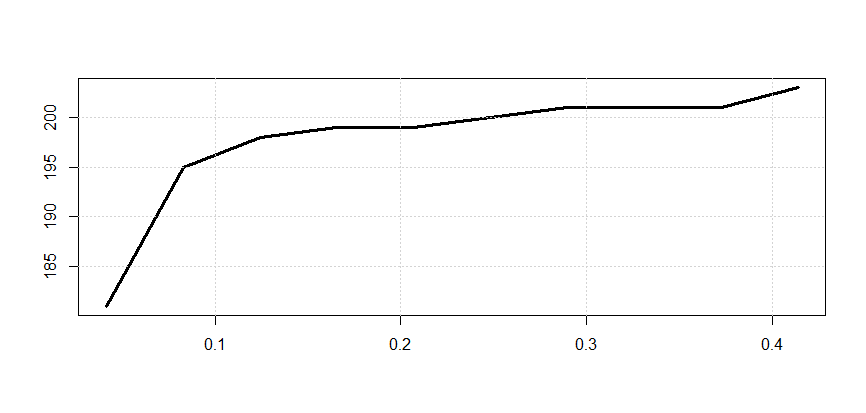
boxplot(duomenys3$susirgimai) #stačiakampė diagrama

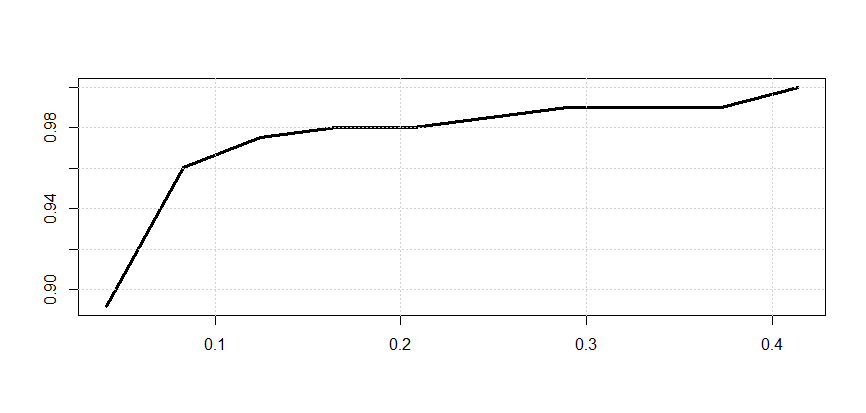


boplot(be3, col="green", pch = 19) #stačiakampė diagrama (be išskirčių)

plot(ecdf(duomenys3$susirgimai));grid() #empirinė dažnių pasiskirstymo funkcija

plot(ecdf(be3));grid() #e empirinė dažnių pasiskirstymo funkcija (be išskirčių)

plot(cc2[-1],nnn3,type='l',lwd=3,xlab='',ylab=''); grid() #sukauptųjų dažnių laužtė

plot(cc2[-1],fff3,type='l',lwd=3,xlab='',ylab=''); grid() #sukauptųjų santykinių dažnių laužtė