

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**

**INFORMATIKOS FAKULTETAS**

**Simona Ragauskaitė**

Studijų modulio

**P160B003 Tikimybių teorija ir statistika**

**1 laboratorinio darbo ataskaita**

Kaunas, 2021

Turinys

[DARBO TIKSLAS 3](#_Toc89706972)

[1. UŽDUOTIS 3](#_Toc89706973)

[1.1. R PROGRAMOS KODAS 3](#_Toc89706974)

[1.2. Rezultatai ir išvados 4](#_Toc89706975)

[2. UŽDUOTIS. 6](#_Toc89706978)

[2.1. R PROGRAMOS KODAS 6](#_Toc89706979)

[2.2. REZULTATAI IR IŠVADOS 7](#_Toc89706980)

[Literatūra 11](#_Toc89706981)

DARBO TIKSLAS

Ugdyti gebėjimus taikyti teorinias staistikos žinias praktikoje, programuoti statistikos uždavinius panaudojant R programavimo kalbą, atlikti tiriamąją duomenų analizę, tikrinti hipotezes, apskaičiuoti parametrų pasikliautinuosius intervalus, interpretuoti gautus rezultatus , formuluoti išvadas rengti ataskaitas.

# UŽDUOTIS

Aprašomoji statistika ir neparametrinės hipotezės.

## R PROGRAMOS KODAS

> spotify\_data <- read\_excel("spotify\_data.xlsx")

> variantas=12

> varianto\_kintamasis=variantas%%10+1

> print(paste("Jusu varianto kintamojo vardas yra: ",colnames(data)[varianto\_kintamasis]))

[1] "Jusu varianto kintamojo vardas yra: "

> varianto\_duomenys=spotify\_data[((variantas-1)\*100+1):(variantas\*100),varianto\_kintamasis]

> colnames(varianto\_duomenys)=colnames(spotify\_data)[varianto\_kintamasis]

> x<-varianto\_duomenys$duration\_ms

> imties\_dydis<-length(x)

> minimumas<-min(x)

> maksimumas<-max(x)

> imties\_plotis<- maksimumas-minimumas

> x<-sort(x)

> Q1<-x[(imties\_dydis\*1/4)+1]

> Q2<-x[(imties\_dydis\*1/2)+1]

> Q3<-x[(imties\_dydis\*3/4)+1]

> IQR<-Q3-Q1

> empirinis\_vidurkis<-mean(x)

> nepaslinktaja\_dispersija<-var(x)

> nepaslinktaji\_standartini\_nuokrypi<-sd(x)

> histograma<-hist.default(x)

> staciakampe\_diagrama<-boxplot(x)

> gof.list<-gofTest(x,test="chisq", distribution = "norm", keep.data = F)

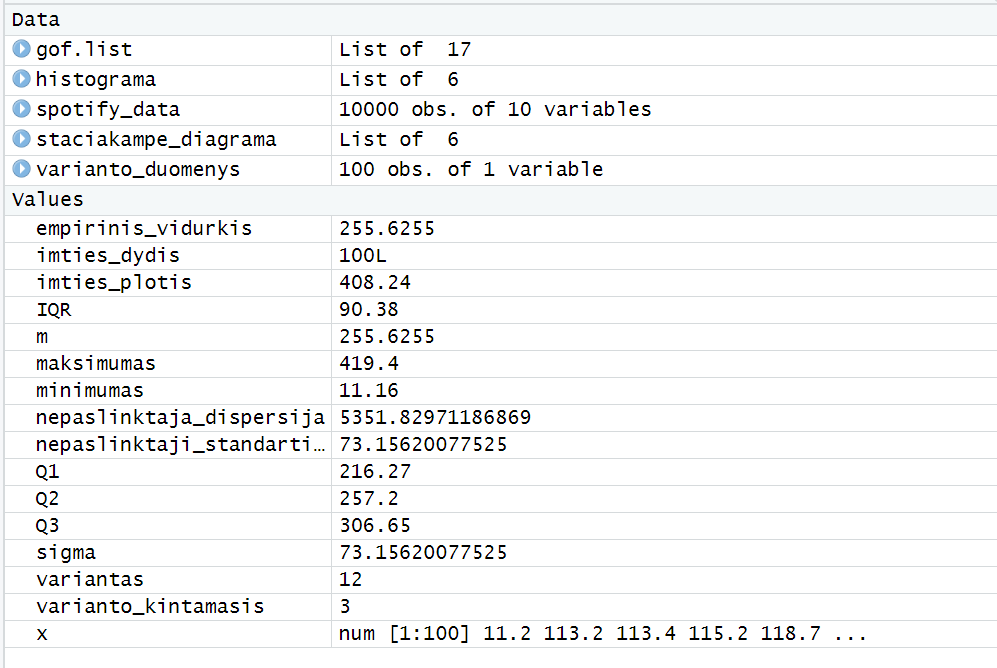
> gof.list

> m<-empirinis\_vidurkis

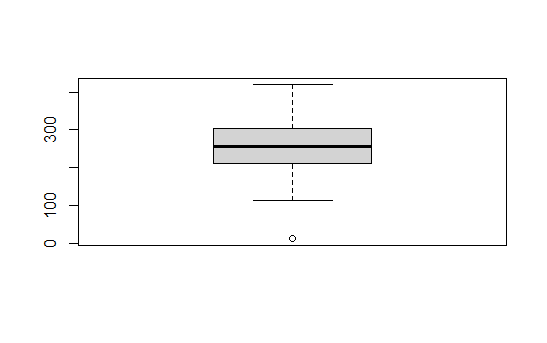
> sigma<-nepaslinktaji\_standartini\_nuokrypi

## Rezultatai ir išvados

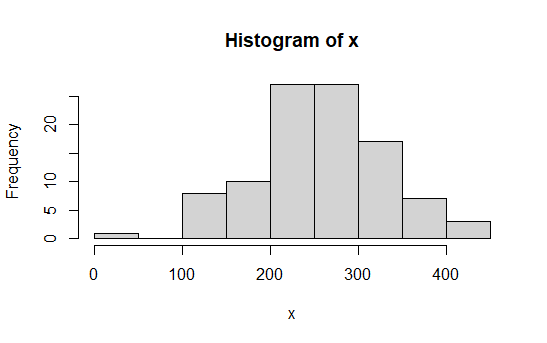




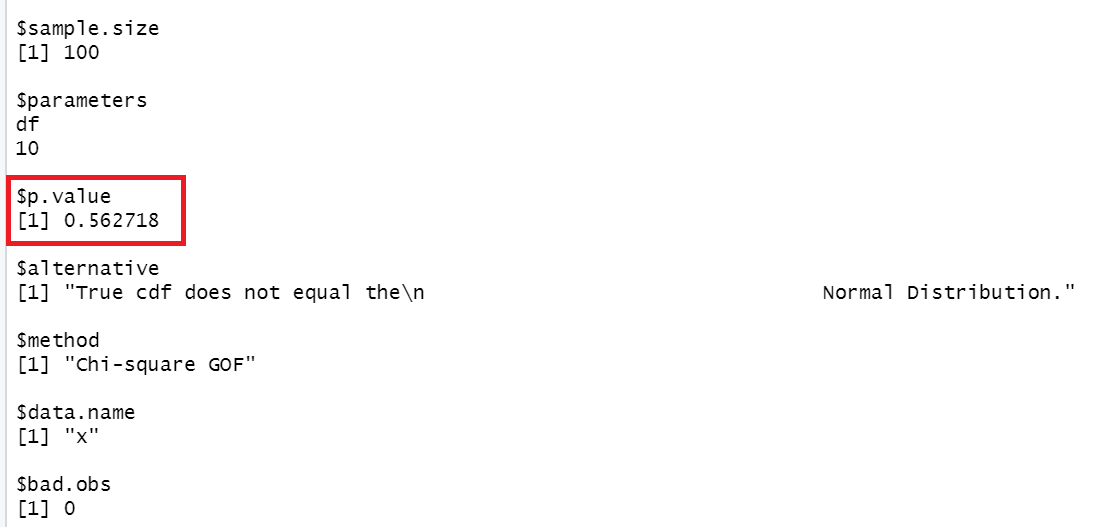
Nr. 1 duomenys – skaitinės imties charakteristikos, taškiniai įverčiai



Nr. 2 duomenys – stačiakampė diragrama



Nr. 3 duomenys – histograma



Nr. 4 duomenys – reikšmė, su kuria lyginame reikšmingumo lygmenį

Mano variantas yra 12, o varianto kintamasis – 3, jis reprezentuoja „spotify\_data.csv“ duomenų imties failo trečiajį duomenų – imčių stulpelį.

Naudodamasi „RStudio“ programa suradau skaitinės imties charakteristikas (*Nr. 5 duomenys*):

* Imties dydis, imčių kiekis buvo lygus – 100.
* Minimumas – 11,16.
* Maksimumas – 419,4.
* Imties plotis, buvo apskaičiuojamas iš maksimumo atėmus minimumą. Imties\_plotis – 408,24.
* Kvartiliai (Q=X([n\*p]+1), n – imties dydis, žinoma, dar prieš tai sudarius variacinę eilutę, kad gautume teisingas kvartilių reikšmes):
  + Kai p = ¼, Q1 = X([100\*1/4]+1) = X(26) = 216,27;
  + Kai p = ½, Q2 = X([100\*1/2]+1) = X(51) = 257,2;
  + Kai p = ¾, Q3 = X([100\*3/4]+1) = X(76) = 306,65.
* Kvartinių skirtumą, IQR = Q3 – Q1 = 90.38.
* Empirinį vidurkį: .
* Dispersiją: .
* Standartinį nuokrypį: .

Nubraižiau su programine įranga stačiakampę diagramą(*Nr. 2 duomenys*) ir histogramą(*Nr. 3 duomenys*). Darome prielaidą, kad a.d. X pasiskirstęs pagal normalųjį dėsnį arba kitaip vadinamą Gauso pasiskirstymą, t.y.:

Parametro taškinis įvertis: .

Parametro taškinis įvertis:.

Pagal histogramą patikrinu ar yra normalusis pasiskirstymas. Su reikšmingumo lygmeniu patikriname suderinamumo hipotezę:

( – normaliojo a.d. X pasiskirstymo funkcija).

Kadangi reikšmingumo lygmenis yra mažesnis už apskaičiuotąją p reikšmę < , (*Nr. 4 duomenys*) tai suderinamumo hipotezę primame.

# UŽDUOTIS.

Grafinė analizė ir parametrinės hipotezės.

## R PROGRAMOS KODAS

> variantas=12

> varianto\_kintamasis=variantas%%12+1

> print(paste("Jusu varianto kintamojo vardas yra: ",

+ colnames(data)[varianto\_kintamasis]))

> varianto\_duomenys=air\_quality[((variantas-1)\*100+1):(variantas\*100),varianto\_kintamasis]

> colnames(varianto\_duomenys)=colnames(air\_quality)[varianto\_kintamasis]

> x<-varianto\_duomenys$CO

> histograma<-hist.default(x)

> staciakampe\_diagrama<-boxplot(x)

> imties\_dydis<-length(x)

> x<-sort(x)

> Q1<-x[(imties\_dydis\*1/4)+1]

> Q2<-x[(imties\_dydis\*1/2)+1]

> Q3<-x[(imties\_dydis\*3/4)+1]

> gof.list<-gofTest(x,test="chisq", distribution = "norm", keep.data = F)

> gof.list

> Q\_Q<-qqnorm(x)

> vidurkis<-mean(x)

> disp<-var(x)

> vidurkio\_taskinis\_ivertis<-vidurkis

> dispersijos\_taskinis\_ivertis\_sigma2<-disp

> t.test(x,conf.level = 0.95)$conf.int

> t.test(x,mu=0,alternative = "two.sided")

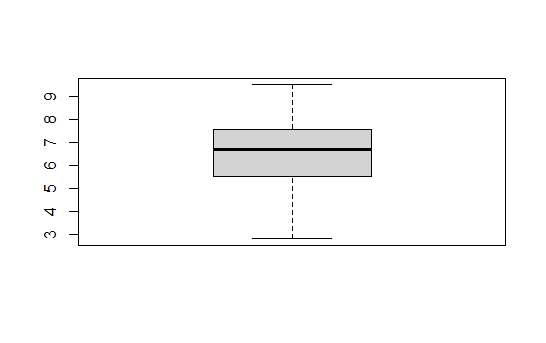
> interval\_var1(x, mu=Inf,alpha=0.05)

## REZULTATAI IR IŠVADOS

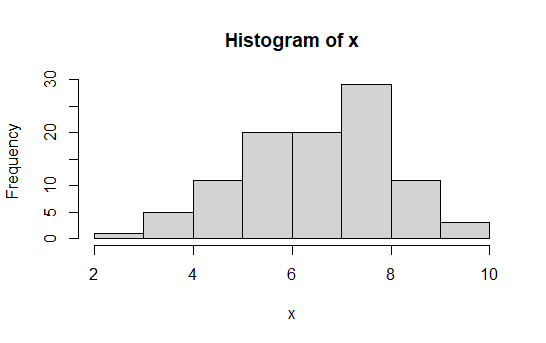
Graphical user interface, table

Description automatically generated

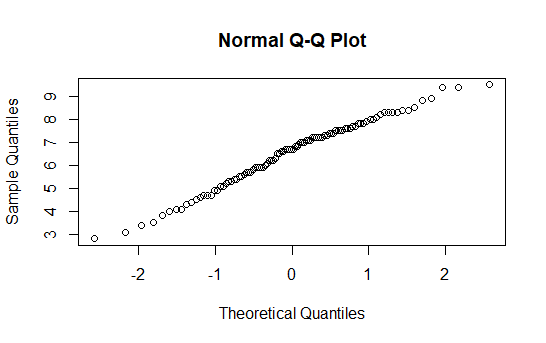
Nr. 6 duomenys – taškiniai vidurkio ir dispersijos įverčiai



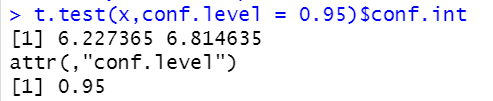
Nr. 7 duomenys – stačiakanpė diagrama



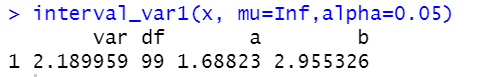
Nr. 8 duomenys – histograma



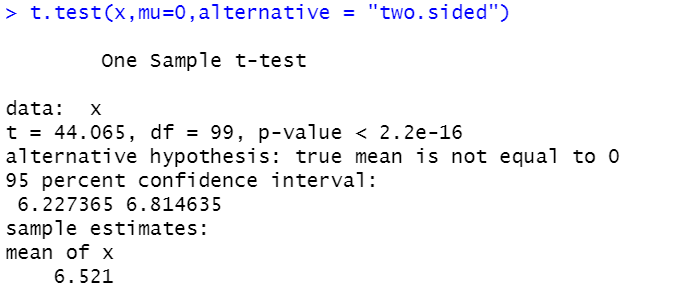
Nr. 9 duomenys – kvantilinis grafikas



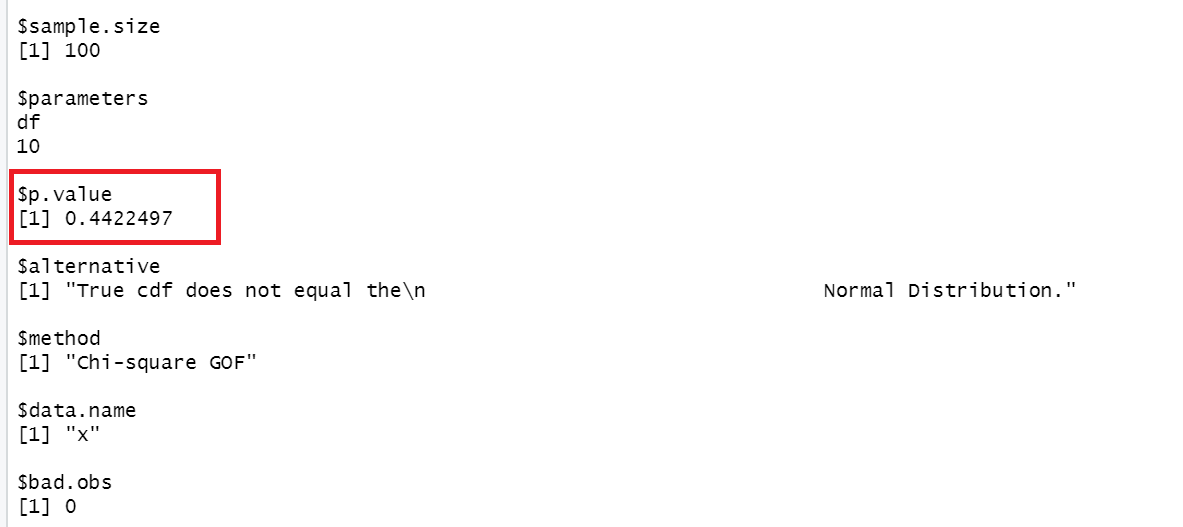
Nr. 10 duomenys – vidurkio pasikliovimo intervalas



Nr. 11 duomenys – dispersijos pasikliovimo intervalas



Nr. 12 duomenys – hipotezės patikrinimas apie vidurkio lygybę skaičiui µ0 su alternatyviąja hipoteze



Nr. duomenys 13– reikšmė, su kuria lyginame reikšmingumo lygmenį

Mano variantas yra 12, o varianto kintamasis – 1, jis reprezentuoja „air\_quality.csv“ duomenų imties failo trečiajį duomenų – imčių stulpelį.

Nubraižiau histogramą (*Nr. 7 duomenys*), stačiakampę diagramą (*Nr. 6 duomenys*) ir kvantilinį grafiką (*Nr. 8 duomenys*). Iš histrogramos galime teigti, jog yra normalusis pasiskirstymas, nes apskaičiavus su reikšmingumo lygmeniu patikriname ir priimu suderinamumo hipotezę < (*Nr. 13 duomenys*). Iš stačiakampės diagramos, galime matyti stačiakampio pradžioje pirmąjį kvantilių, o pabaigoje – trečiąjį kvantilių. Horizontali linija, dalijanti stačiakampį – mediana (antrasis kvantilius). Pasiskirstymas yra normalusis, nes stačiakampis susidaro viduryje koordinačių plokštumoje. Iš kvantilinio grafiko galime matyti, kad visi taškai išskyrūs kraštiniai yra su nedidelėmis paklaidomis grafike, jei nubrėžtume tiesę, ko pasekoje a. d. X pasiskirstymas yra normalusis.

Taškinis vidurkio įvertis (*Nr. 5 duomenys*): .

Taškinis dispersijos įvertis (*Nr. 5 duomenys*): .

Darom prielaidą, kad .

Vidurkio pasikliovimo intervalas (*Nr. 9 duomenys*) yra – (6,227365;6,814635).

Dispersijos pasikliovimo intervalas (*Nr. 10 duomenys*) yra – (1.68823;2.955326).

Su reikšmingumo lygmeniu tikrinu hipotezę apie vidurkio reikšmę( ir reikšmingumo lygmeniu , kai varianto numeris lyginis):

– pagrindinė hipotezė

– alternatyvi hipotezė

Kadangi reikšmingumo lygmuo yra didesnis už apskaičiuotąją p reikšmę (*Nr. 11 duomenys*), tai pagrindinę hipotezę atmetame ir priimame alternatyvią hipotezę, kad .

Literatūra

1. <https://moodle.ktu.edu/course/view.php?id=6858> [žiūrėta 2021-11-30]
2. <http://web.vu.lt/mif/a.reklaite/files/2013/02/2010.01-Pats-trumpiausias-su-R.pdf> [žiūrėta 2021-11-30]