

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**

**INFORMATIKOS FAKULTETAS**

**Simona Ragauskaitė**

Studijų modulio

**P160B003 Tikimybių teorija ir statistika**

**2 laboratorinio darbo ataskaita**

Kaunas, 2021

Turinys

[DARBO TIKSLAS 3](#_Toc90378683)

[1. UŽDUOTIS 3](#_Toc90378684)

[1.1. R PROGRAMOS KODAS 3](#_Toc90378685)

[2.2. REZULTATAI IR IŠVADOS 4](#_Toc90378688)

[2. UŽDUOTIS. 5](#_Toc90378689)

[2.1. R PROGRAMOS KODAS 5](#_Toc90378690)

[2.2. REZULTATAI IR IŠVADOS 6](#_Toc90378691)

[Literatūra 10](#_Toc90378692)

DARBO TIKSLAS

Ugdyti gebėjimus taikyti teorinias staistikos žinias praktikoje, programuoti statistikos uždavinius panaudojant R programavimo kalbą, atlikti realių duomenų koreliacinę ir regresinę analizę, pavaizduoti duomenis grafiškai, tikrinti hipotezes apie modelio koeficientus, užrašyti tiesinės regresijos modelio išraišką, interpretuoti gautus rezultatus, formuluoti išvadas, rengti ataskaitą.

# UŽDUOTIS

## R PROGRAMOS KODAS

> library(readxl)

> air\_quality <- read\_excel("air\_quality.xlsx")

> variantas=12

> kintamuju\_kombinacijos=expand.grid(1:10,1:12)

>kintamuju\_kombinacijos=kintamuju\_kombinacijos[kintamuju\_kombinacijos$Var1!=kintamuju\_kombinacijos$Var2,]

> varianto\_kintamieji=kintamuju\_kombinacijos[variantas,]

>varianto\_duomenys=air\_quality[((variantas-1)\*100+1):(variantas\*100),c(varianto\_kintamieji$Var1,varianto\_kintamieji$Var2)]

> x<-varianto\_duomenys$Titan

> y<-varianto\_duomenys$tin\_oxide

> plot(x,y)

> r<-cor(x,y,method = "pearson")

> Rs<-cor(x,y,method = "spearman")



## REZULTATAI IR IŠVADOS

Chart, scatter chart

Description automatically generated

Nr. 1 duomenys – kintamųjų taųkų sklaidos diagrama

A picture containing text

Description automatically generated

Nr. 2 duomenys – Pisono ir Spirmano koreliacijos koeficientai tarp kintamųjų

Mano variantas yra 12, o varianto kintamieji 4 ir 5, jie tyrime reprezentuoja iš duomenų failo „air\_quality.csv“ duomenų stulpelius „Titan“ ir „NOx“.

Nubraižiau sklaidos diagramą (*Nr. 2 duomenys*), gavau tiesinę priklausomybę, nes būtų galima nubrėžti tiese pagal taškų išsidėstymą diagramoje.

Pirsono koreliacijos koeficientas – egzituoja stiprus teigiamas ryšys.

Spirmano koreliacijos koeficientas – egzistuoja stiprus teigiamas ryšys ( yra rangas ( pozicija X variacinėje eilutėje), o – rangas).

# UŽDUOTIS.

## R PROGRAMOS KODAS

> variantas=12

> data = read.csv("air\_quality.csv")

> kintamuju\_kombinacijos=expand.grid(1:9,1:9)

> kintamuju\_kombinacijos=kintamuju\_kombinacijos[kintamuju\_kombinacijos$Var1!=kintamuju\_kombinacijos$Var2,]

> varianto\_kintamieji=kintamuju\_kombinacijos[variantas%%72,]

>varianto\_duomenys=data[((variantas-1)\*100+1):(variantas\*100),c(varianto\_kintamieji$Var1,varianto\_kintamieji$Var2)]

> View(varianto\_kintamieji)

> View(varianto\_duomenys)

> x<-varianto\_duomenys$NOx

> y<-varianto\_duomenys$tin\_oxide

> plot(x,y)

> r<-cor(x,y,method = "pearson")

> Rs<-cor(x,y,method = "spearman")

> reg=lm(y~x)

> reg

> lines(x, predict(reg))

> summary(reg)

> predict(reg,data.frame(x=400))

> e<-residuals(reg)

> h<-hist.default(e)

> plot(e)

> gof.list<-gofTest(e, test = "ks", distribution = "norm", param.list = list(mean=0, sd=sd(e)), keep.data = F)

> gof.list

## REZULTATAI IR IŠVADOS

Chart, scatter chart

Description automatically generated

Nr. duomenys 3 – sklaidos diagrama

Graphical user interface, text, application, email

Description automatically generated

Nr. 4 duomenys – Pirsono ir Spirmeno koeficientai

Chart, scatter chart

Description automatically generated

Nr. 5 duomenys – sklaidos diagrama su nubrėžta tiese

A screenshot of a computer

Description automatically generated with medium confidence

Nr. 6 duomenys – determinacijos koeficientas

**Text, letter

Description automatically generated**

Nr. 7 duomenys – regresinės tiesės koeficientai

A picture containing logo

Description automatically generated

Nr. 8 duomenys – modelio prognozė

Chart, scatter chart

Description automatically generated

Nr. 9 duomenys – modelio liekanų (paklaidų) grafikas

Chart, histogram

Description automatically generated

Nr. 10 duomenys – histograma

Graphical user interface, text, application, email

Description automatically generated

Nr. 11 duomenys – tikrinu hipotezę su reikšmingumo lygmeniu

Mano variantas yra 12, o varianto kintamieji 5 ir 2, jie tyrime reprezentuoja iš duomenų failo „air\_quality.csv“ duomenų stulpelius „NOx“ ir „tin\_oxide“.

Nubraižiau sklaidos diagramą (*Nr. 3 duomenys*), gavau tiesinę priklausomybę, nes būtų galima nubrėžti tiese pagal taškų išsidėstymą diagramoje.

Apskaičiavau koleriacijos koeficientus (*Nr. 4 duomenys*):

* Pirsono koreliacijos koeficientas – egzituoja vidutinis teigiamas ryšys.
* Spirmano koreliacijos koeficientas – egzistuoja silpnas teigiamas ryšys.

Pirsono koeficientas yra tinkamas mano naudojamų kintamųjų priklausomybei įvertinti, nes iš sklaidos diagramos atrodo kaip tiesinis ryšys, taškai pasiskirstę nuo apatinio kairiojo kampo iki viršutinio dešiniojo.

Regresijos tiesės lygtis , apskaičiavau koeficientus ir (*Nr. 7 duomenys*), iš to seka, kad gaunu regresinę tiesę tokią: , ją nubrėžiu sklaidos diagramoje (*Nr. 5 duomenys*).

Apskaičiavau determinacijos koeficientą tiesinės regresijos modeliui (*Nr. 6 duomenys*) (Skaitiklis *SSR* yra vadinamas regresijos kvadratų suma, o vardiklis *SST* – visų kvadratų suma).

Mano modelio prognozei tikrinti pasirinkau, kai . Gaunu, kad modelio reikšmė yra lygi: (*Nr. 8 duomenys*).

Nubraižiau modelio liekanų (paklaidų) grafiką (*Nr. 9 duomenys*) ir histogramą (*Nr. 10 duomenys*). Tikrinu hipotezę (reikšmingumo lygmuo ), kad paklaidos yra pasiskirsčiusios pagal normalųjį normalųjį skirstinį su vidurkiu .

Kadangi (*Nr. 11 duomenys*), tai priimame hipotezę, kad paklaidos pasiskirsčiusios pagal normalųjį dėsnį .

Galutinė išvada, apie sudaryto modelio tinkamumą:

* Koreliacinio dydžio stiprumas – Pirsono – vidutinis, Spirmano – silpnas.
* Determinacijos koeficientas , tai tiesinės regresijos modelis yra tinkamas turimiems duomenims.
* Sklaidos diagrama – pasiskirsčiusios reikšmės homoskedastiškai.
* Hipotezė dėl paklaidų pasiskirstymo yra priimta.

Literatūra

1. <https://moodle.ktu.edu/course/view.php?id=6858> [žiūrėta 2021-12-08]