



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

LUKAS KUZMICKAS

Studijų modulio

P160B003 Tikimybių teorija ir statistika

2 laboratorinio darbo ataskaita

Kaunas, 2022

DARBO TIKSLAS

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

1. UŽDUOTIS – I DALIS.

1. KORELIACINĖ ANALIZĖ

Iš duomenų imties „duomenys_Pirsono_koreliacijos_koeficientui.csv“ nuskaitykite savo varianto duomenis tokiu būdu:

```
variantas= N #Reikia įrašyti savo varianto numerį N
data=read.csv('duomenys_Pirsono_koreliacijos_koeficientui.csv',header=TRUE) #Failas turi būti darbiname kataloge
x1=data[,2*(N-1)+1]
y1=data[,2*N]
```

1.1. Pateikite x1 vs. y1 grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.

1.2. Apskaičiuokite Pirsono koreliacijos koeficientą tarp kintamųjų x1 ir y1. Pateikite koeficiento reikšmės interpretaciją savo duomenų imties atveju.

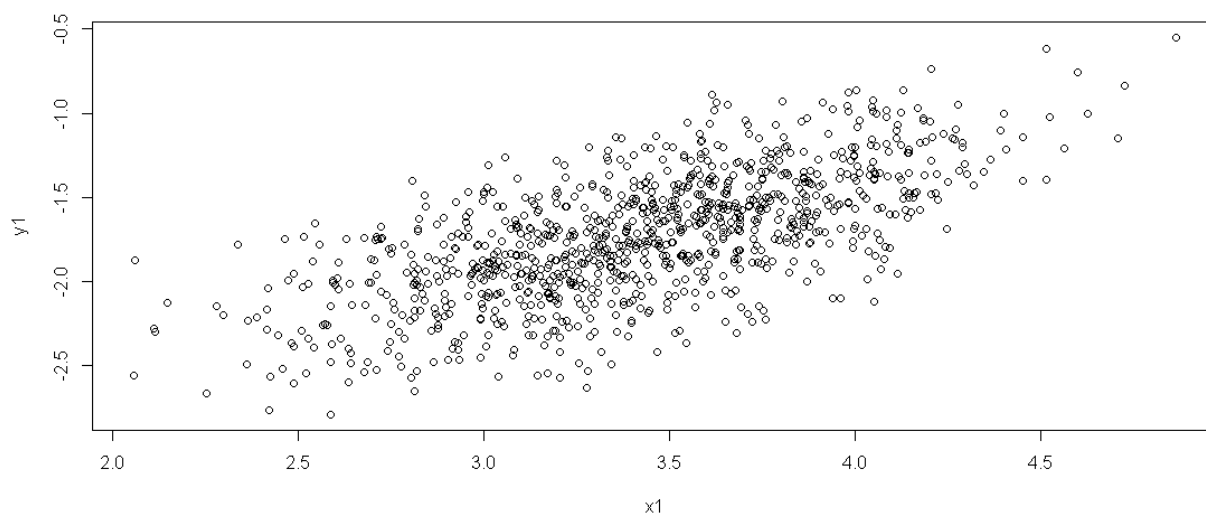
1 pav. Pirmos dalies užduotys.

1.1. R PROGRAMOS KODAS

```
#1 DALIS
#30 variantas
N = 30 #Reikia įrašyti savo varianto numerį N
data=read.csv('duomenys_Pirsono_koreliacijos_koeficientui.csv',header=TRUE) #Failas turi būti darbiname kataloge
x1=data[,2*(N-1)+1]
y1=data[,2*N]
x1
y1
#1.1. Pateikite x1 vs. y1 grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.
plot(x1,y1)
#Iš grafiko pastebime, kad x1 didėjant, y1 irgi didėja - galime teigti, jog tai tiesinė priklausomybė.
```

2 pav. 1.1. užduoties kodo fragmentas.

REZULTATAI IR IŠVADOS



3 pav. 1.1. užduoties rezultatai.

Iš 3 paveikslėlio matome, kad priklausomybė, tarp x_1 ir y_1 yra tiesinė (vienam didėjant, kitas didėja).

1.2. R PROGRAMOS KODAS

```
#1.2. Apskaičiuokite Pirsono koreliacijos koeficientą tarp kintamųjų x1 ir y1. Pateikite koeficiento reikšmės  
#interpretaciją savo duomenų imties atveju.  
cor(x1,y1, method="pearson")  
#Pirsono koreliacijos koeficientas = 0.6636055, ryšys tarp x1 ir y1 yra vidutinis.
```

4 pav. 1.2. užduoties kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #1.2. Apskaičiuokite Pirsono koreliacijos koeficientą tarp kintamųjų x1 ir y1. Pateikite koeficiento reikšmes  
> #interpretaciją savo duomenų imties atveju.  
> cor(x1,y1, method="pearson")  
[1] 0.6636055  
> #Pirsono koreliacijos koeficientas = 0.6636055, ryšys tarp x1 ir y1 yra vidutinis.
```

5 pav. 1.2. užduoties kodo rezultatas.

Apskaičiuojame Pirsono koreliacijos koeficientą ($r=0.6636055$), ir padarome išvadą, kad ryšys tarp x_1 ir y_1 yra vidutinis.

1. UŽDUOTIS – II DALIS.

Iš duomenų imties „duomenys_Spirmeno_koreliacijos_koeficientui.csv“ nuskaitykite savo varianto duomenis tokiu būdu:

```
variantas=N #Reikia įrašyti savo varianto numerį N
data=read.csv('duomenys_Spirmeno_koreliacijos_koeficientui.csv',header=TRUE) #Failas turi būti darbiname kataloge
x2=data[,2*(N-1)+1]
y2=data[,2*N]
```

1.3. Pateikite x2 vs. y2 grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.

1.4. Apskaičiuokite Spirmeno koreliacijos koeficientą tarp kintamųjų x2 ir y2. Pateikite koeficiento reikšmės interpretaciją savo duomenų imties atveju.

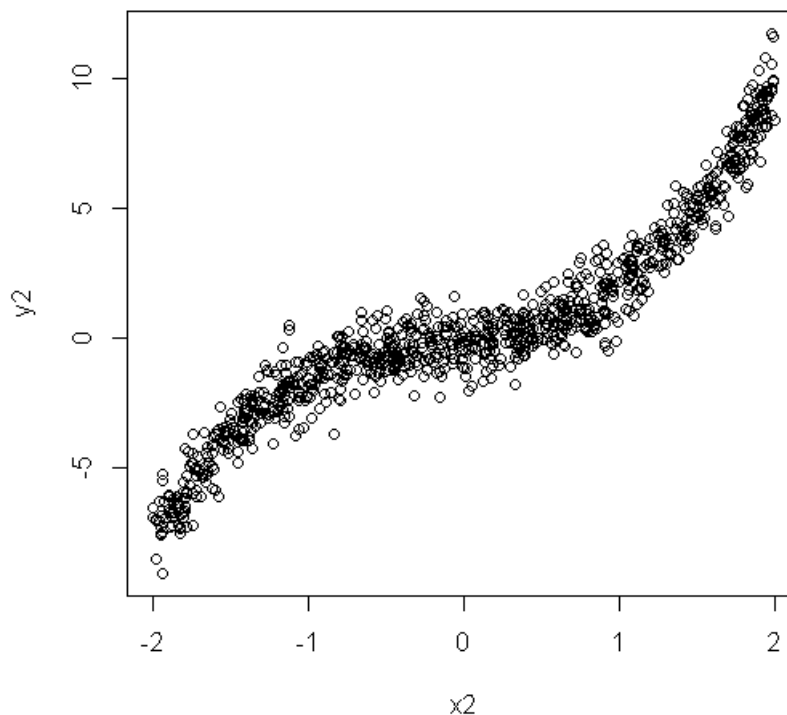
6 pav. 1 užduoties – II dalies uždaviniai.

1.3. R PROGRAMOS KODAS

```
#2 DALIS
N = 30 #Reikia įrašyti savo varianto numerį N
data=read.csv('duomenys_Spirmeno_koreliacijos_koeficientui.csv',header=TRUE) #Failas turi būti darbiname kataloge
x2=data[,2*(N-1)+1]
y2=data[,2*N]
x2
y2
#1.3. Pateikite x2 vs. y2 grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.
plot(x2, y2)
#Iš grafiko matome, kad priklausomybė yra pusiau tiesiška, priklausomybės pilnai nustatyti negalime.
```

7 pav. 1.3. užduoties kodo fragmentas.

REZULTATAI IR IŠVADOS



8 pav. 1.3. užduoties rezultatai.

Iš rezultatų, galime pasakyti, kad x_2 ir y_2 tiesiška, tačiau priklausomybės pilnai nustatyti negalime.

1.4. R PROGRAMOS KODAS

```
#1.4. Apskaičiuokite Spirmeno koreliacijos koeficientą tarp kintamųjų x2 ir y2. Pateikite koeficiento reikšmės  
#interpretaciją savo duomenų imties atveju.  
spearmancor = cor(x2, y2, method = "spearman")  
spearmancor  
#Spirmeno koreliacijos koeficientas = 0.9557525, ryšys tarp x2 ir y2 yra labai stiprus.
```

9 pav. 1.4. užduoties kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #1.4. Apskaičiuokite Spirmeno koreliacijos koeficientą tarp kintamųjų x2 ir y2. Pateikite koeficiento reikšmės  
> #interpretaciją savo duomenų imties atveju.  
> spearmancor = cor(x2, y2, method = "spearman")  
> spearmancor  
[1] 0.9557525  
> #Spirmeno koreliacijos koeficientas = 0.9557525, ryšys tarp x2 ir y2 yra labai stiprus.
```

10 pav. 1.4. užduoties rezultatai.

Apskaičiuojame Spirmeno koreliacijos koeficientą $r = 0.9557525$, padarome išvadą, kad ryšys tarp x_2 ir y_2 yra labai stiprus.

2. UŽDUOTIS.

2. REGRESINĖ ANALIZĖ

Iš duomenų imties „duomenys_regresijai.csv“ nuskaitykite savo varianto duomenis tokiu būdu:

```
variantas=18 #Reikia įrašyti savo varianto numerį  
data=read.csv('duomenys_regresijai.csv',header=TRUE) #Failas turi būti darbiname kataloge  
x=data[,2*(variantas-1)+1]  
y=data[,2*variantas]
```

- 2.1. Pateikite x vs. y grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.
- 2.2. Apskaičiuokite Pirsono koreliacijos koeficientą ir pakomentuokite jo reikšmę.
- 2.3. Apskaičiuokite tiesinės regresijos koeficientus ir juos pateikite. Parašykite vieną išvadą apie koeficientų reikšmę ir ryšį tarp kintamųjų x,y.
- 2.4. Apskaičiuokite determinacijos koeficientą tiesinės regresijos modeliui, gautam 2.3 punkte.
- 2.5. Pateikite jūsų gauto modelio prognozę vienai pasirinktai x reikšmei.
- 2.6. Pateikite modelio liekanų grafiką ir histogramą. Pakomentuokite, ar vizualiai galima teigti, jog liekanos yra pasiskirstę pagal normalųjį skirstinį su vidurkiu, lygiu nuliui?
- 2.7. Patikrinkite hipotezę ($\alpha = 0.05$), kad tiesinės regresijos liekanų skirstinys yra normalusis, kurio vidurkis lygus 0.

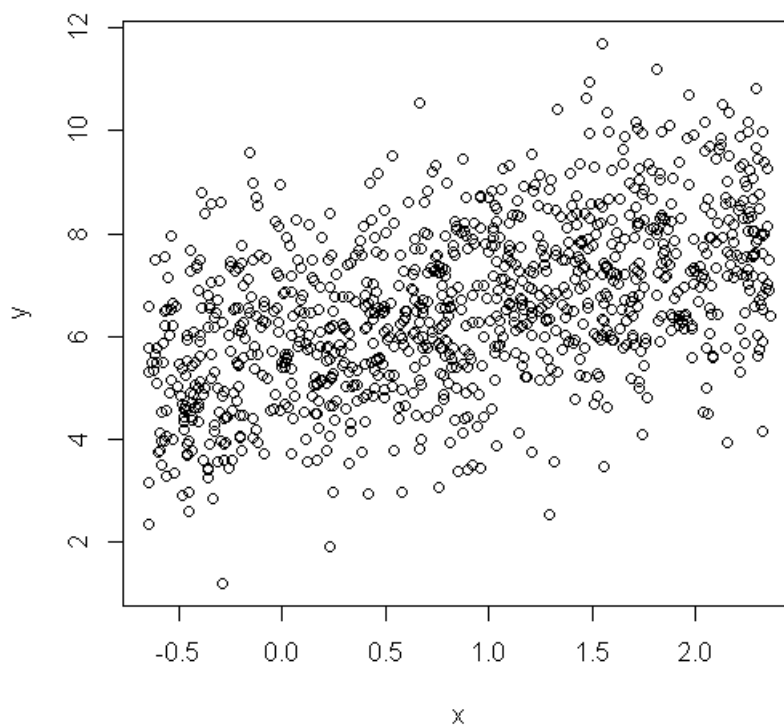
11 pav. Antros dalies uždaviniai.

2.1. R PROGRAMOS KODAS

```
#3 DALIS  
variantas= 30 #Reikia įrašyti savo varianto numerį  
data=read.csv('duomenys_regresijai.csv',header=TRUE) #Failas turi būti darbiname kataloge  
x=data[,2*(variantas-1)+1]  
y=data[,2*variantas]  
x  
y  
#2.1. Pateikite x vs. y grafiką ir pakomentuokite priklausomybę tarp šių kintamųjų.  
plot(x,y)  
#Iš grafiko matome, kad ryšys tarp duomenų yra pastovus - liniinis, išvados apie priklausomybę padaryti negalime
```

12 pav. 2.1. kodo fragmentas.

REZULTATAI IR IŠVADOS



13 pav. 2.1. uždavinio rezultatas.

Iš duomenų grafiko, galime pasakyti, kad tarp x ir y – pastovus ryšys, išvados apie priklausomybę apibrėžti negalime.

2.2. R PROGRAMOS KODAS

```
#2.2. Apskaičiuokite Pirsono koreliacijos koeficientą ir pakomentuokite jo reikšmę.  
cor(x, y, method = "pearson")  
#Pirsono koreliacijos koeficientas r = 0.5041774, ryšys tarp x ir y yra vidutinis.
```

14 pav. 2.2. kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #2.2. Apskaičiuokite Pirsono koreliacijos koeficientą ir pakomentuokite jo reikšmę.  
> cor(x, y, method = "pearson")  
[1] 0.5042791  
> #Pirsono koreliacijos koeficientas r = 0.5041774, ryšys tarp x ir y yra vidutinis.  
> |
```

15 pav. 2.2. rezultatai.

Apskaičiavome Pirsono koreliacijos koeficientą $r = 0.5041774$ ir padarome išvadą, kad ryšys tarp x ir y yra vidutinis.

2.3. R PROGRAMOS KODAS

```
#2.3. Apskaičiuokite tiesinės regresijos koeficientus ir juos pateikite. Parašykite vieną išvadą apie koeficientų  
#reikšmę ir ryšį tarp kintamųjų x,y.  
reg = lm(y~x)  
plot(x,y)  
lines(x, predict(reg))  
reg  
#Gavome a = 5.7601  
#b = 0.9374  
#Regresijos lygtis:  $Y = 5.7601 + 0.9374x$   
#Iš regresijos lygties galime daryti išvadą, kad tarp x ir y yra tiesinė priklausomybė.
```

16 pav. 2.3. kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #2.3. Apskaičiuokite tiesinės regresijos koeficientus ir juos pateikite. Parašykite vieną išvadą apie koeficientų  
> #reikšmę ir ryšį tarp kintamųjų x,y.  
> reg = lm(y~x)  
> lines(x, predict(reg))  
> reg  
  
Call:  
lm(formula = y ~ x)  
  
Coefficients:  
(Intercept)          x  
    5.7601         0.9374  
  
> #Gavome a = 5.7601  
> #b = 0.9374  
> #Regresijos lygtis:  $Y = 5.7601 + 0.9374x$ 
```

Apskaičiuojame tiesinės regresijos koeficientus ($a=5.7601$, $b=0.9374$). Ryšys tarp x ir y yra vidutinis.

17 pav. 2.3. užduoties rezultatai.

2.4. R PROGRAMOS KODAS

```
#2.4. Apskaičiuokite determinacijos koeficientą tiesinės regresijos modeliui, gautam 2.3 punkte.  
deterKoefficientas = summary(reg)  
deterKoefficientas  
#Rquared = 0.2543, determinacijos koeficientas.
```

18 pav. 2.4. kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #2.4. Apskaičiuokite determinacijos koeficientą tiesinės regresijos modeliui, gautam 2.3 punkte.  
> deterKoefficientas = summary(reg)  
> deterKoefficientas  
  
Call:  
lm(formula = y ~ x)  
  
Residuals:  
    Min       1Q   Median       3Q      Max  
-4.4486 -0.9700 -0.0393  0.9182  4.4903  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  5.76005    0.06248   92.19  <2e-16 ***  
x            0.93745    0.05082   18.45  <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 1.39 on 998 degrees of freedom  
Multiple R-squared:  0.2543,    Adjusted R-squared:  0.2536  
F-statistic: 340.3 on 1 and 998 DF,  p-value: < 2.2e-16  
  
> #Rquared = 0.2543, determinacijos koeficientas labai mažas, todėl modelis netinkamas aprašyti  
>
```

Apskaičiuojame determinacijos koeficientą regresijos modeliui (0.2543).

19 pav. 2.4. kodo rezultatai.

2.5. R PROGRAMOS KODAS

```
#2.5. Pateikite jūsų gauto modelio prognozę vienai pasirinktai x reikšmei.  
#pasirinkta x reikšmė 1  
yPrognose = predict(reg, data.frame(x=1))  
yPrognose  
#Gauname modelio prognozę y_pr = 6.6975, kai x = 1
```

20 pav. 2.5. kodo fragmentas.

REZULTATAI IR IŠVADOS

```
> #2.5. Pateikite jūsų gauto modelio prognozę vienai pasirinktai x reikšmei.  
> #pasirinkta x reikšmė 1  
> yPrognose = predict(reg, data.frame(x=1))  
> yPrognose  
      1  
6.6975
```

Gauname, kad $y = 6.6975$, kai $x = 1$

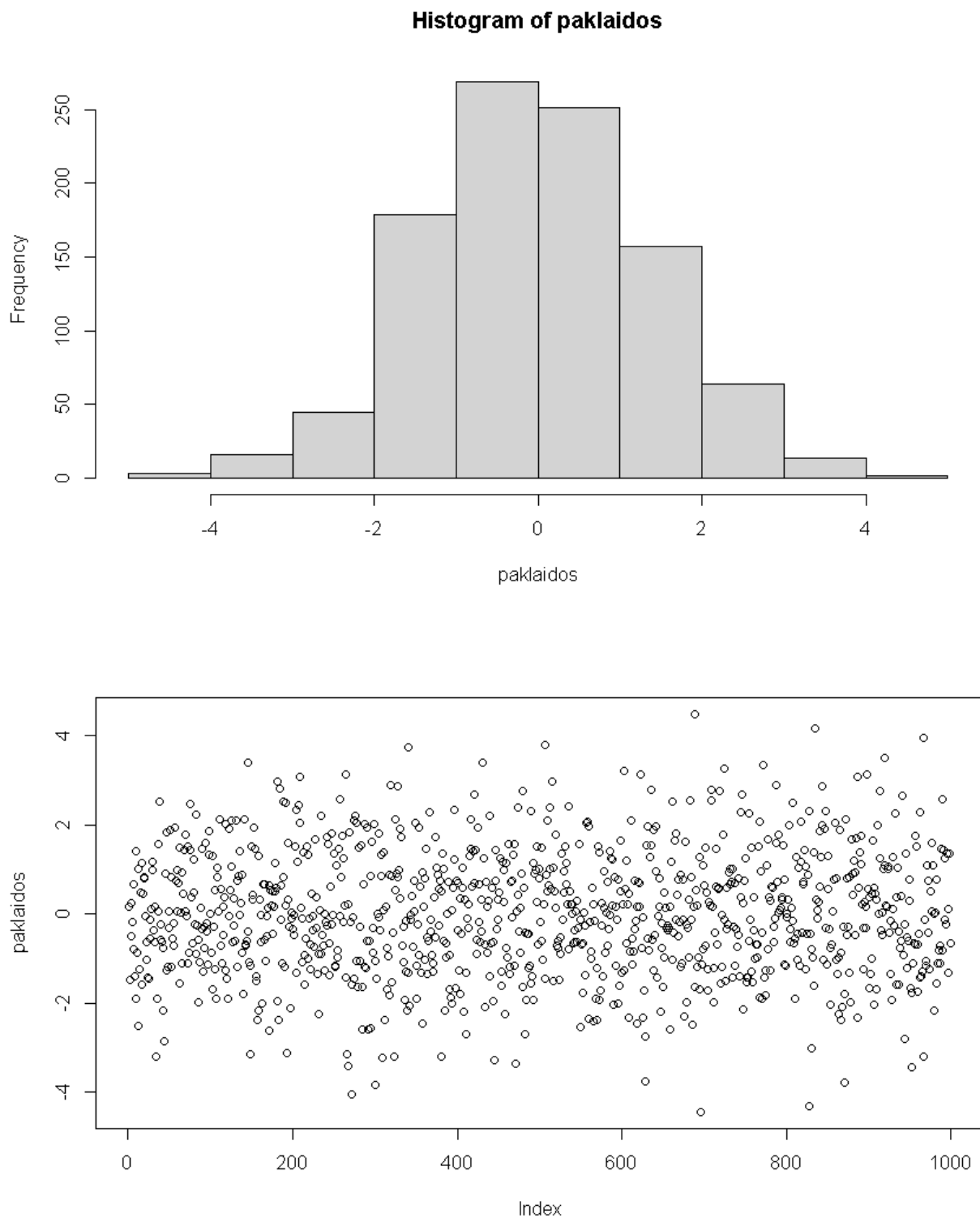
21 pav. 2.5. kodo rezultatai.

2.6. R PROGRAMOS KODAS

```
#2.6. Pateikite modelio liekanų grafiką ir histogramą. Pakomentuokite, ar vizualiai galima teigti, jog liekanos yra  
#pasiskirstę pagal normalųjį skirstinį su vidurkiu, lygiu nuliui?  
paklaidos <- residuals(reg)  
plot(paklaidos)  
mean(paklaidos)  
h <- hist.default(paklaidos)  
#Iš liekanų grafiko ir histogramos galime vizualiai teigti, jog liekanos yra pasiskirsčiusios pagal normalųjį skirstinį.
```

22 pav. 2.6. kodo fragmentas.

REZULTATAI IR IŠVADOS



23 pav. Grafikai.

Iš liekanų grafiko ir histogramos, galime teigti, kad duomenys yra pasiskirstę pagal normalųjį skirstinį.

2.7. R PROGRAMOS KODAS

```
#2.7. Patikrinkite hipotezę (alpha = 0.05), kad tiesinės regresijos liekanų skirstinys yra normalusis, kurio vidurkis
#lygus 0.
alpha = 0.05
# Kolmogorovo-Smirnovo kriterijus
install.packages('EnvStats')
library(EnvStats)
gof.list <- gofTest(paklaidos, test = "ks", distribution = "norm", param.list = list(mean = 0, sd = sd(paklaidos)), keep.data = F)
gof.list
#gauta p reikšmė = 0.8178942, p > alpha, todėl hipotezė, kad liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su vidurkiu, lygiu nuliui pasitvirtino

#Antrasis metodas
ks.test(paklaidos, "pnorm", 0, sd(paklaidos))
#gauta p reikšmė p = 0.8179, p > alpha, todėl hipotezė, kad liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su vidurkiu, lygiu nuliui pasitvirtino
```

24 pav. 2.7. kodo fragmentas.

REZULTATAI IR IŠVADOS

```
$p.value
[1] 0.8178942

$alternative
[1] "True cdf does not equal the\n          Normal(mean = 0, sd = 1.389365)\n          Distribution."

$method
[1] "Kolmogorov-Smirnov GOF"

$data.name
[1] "paklaidos"

$bad.obs
[1] 0

attr(,"class")
[1] "gof"
> #gauta p reikšmė = 0.8178942, p > alpha, todėl hipotezė, kad liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su vidurkiu, lygiu nuliui pasitvirtino
>
> #Antrasis metodas
> ks.test(paklaidos, "pnorm", 0, sd(paklaidos))
      Asymptotic one-sample Kolmogorov-Smirnov test

data:  paklaidos
D = 0.020015, p-value = 0.8179
alternative hypothesis: two-sided
> #gauta p reikšmė p = 0.8179, p > alpha, todėl hipotezė, kad liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su vidurkiu, lygiu nuliui pasitvirtino
```

25 pav. 2.7. kodo rezultatai.

Patikrinome hipotezę, gavome abiem atvejais, kad $p = 0.8179$, t.y. $p > \alpha$, tai galime teigti, kad hipotezė, kad liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su vidurkiu, lygiu nuliui pasitvirtino.