

VILNIAUS UNIVERSITETO

KAUNO FAKULTETAS

SOCIALINIŲ MOKSLŲ IR TAIKOMOSIOS INFORMATIKOS INSTITUTAS

Informacinių sistemų ir kibernetinės saugos studijų programa

IRMANTAS VASILJEVAS

STATISTIKOS STUDIJŲ DALYKAS

SAVARANKIŠKAS DARBAS No. 3

KAUNAS 2019

TURINYS

PAVEIKSLŲ SĄRAŠAS	4
LENTELIŲ SĄRAŠAS	4
1. PIRMOJI UŽDUOTIS	5
1.1 Ekloginės priklausomybės grafikas	5
1.2 Hiperbolinės priklausomybės grafikas	5
1.3 Loginės priklausomybės grafikas	6
2. ANTROJI UŽDUOTIS	8
2.1 Duomenų pasirinkimo situacija	8
2.2 Geriausios regresinės lygties paieška	8
2.3 Determinacijos koeficientai	10
3. TREČIOJI UŽDUOTIS	11
3.1 Duomenų pasirinkimo situacija	11
ŠALTINIAI	13

PAVEIKSLŲ SĄRAŠAS

pav 1	Ekologinė priklausomybė	5
pav 2	Hiperbolinė priklausomybė.....	6
pav 3	Loginė priklausomybė	6
pav 4	Regresinės lygties paieška. Tiesinė	8
pav 5	Regresinės lygties paieška. Loginė	9
pav 6	Regresinės lygties paieška. Logaritminė	9
pav 7	Regresinės lygties paieška. Eksponentinė	10
pav 8	Determinacijos koeficientai	10
pav 9	Trendo lygties paieška. Ekologinė regresija	11
pav 10	Trendo lygties paieška. Trupmeninė racionalioji regresija.....	12
pav 11	autokoreliacija 1	14
pav 12	autokoreliacija 2	14

LENTELIŲ SĄRAŠAS

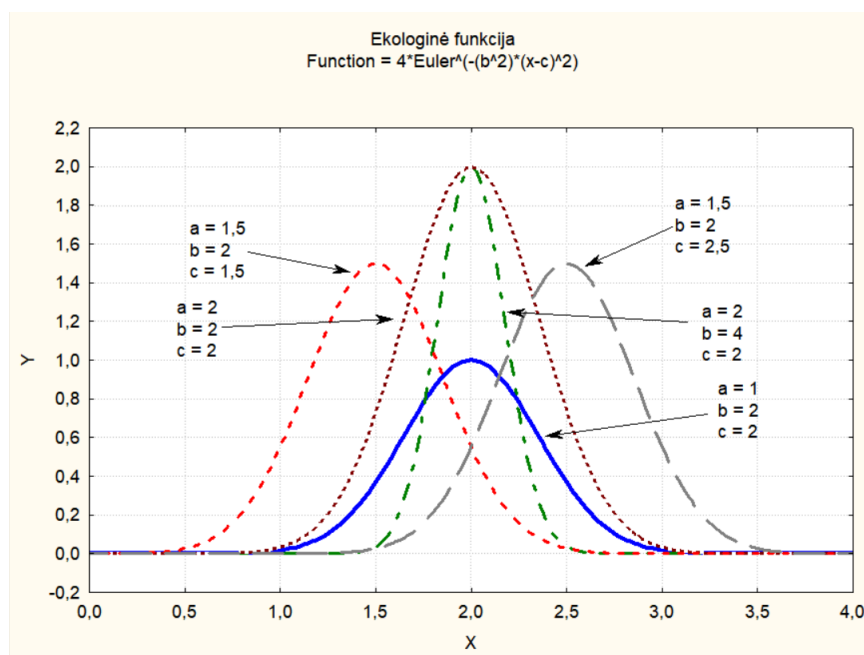
lentelė 1	Ilgalaikio gydymo prognozės duomenys	8
lentelė 2	Reikšmių prognozavimas	12

1. PIRMOJI UŽDUOTIS

Nubrėžti trijų standartinių regresinių kreivių grafikus pagal knygos 156 p. lentelę, pradedant nuo kreivės atitinkančios asmeninį eilės Nr. Paaiškinti kreivių kitimą.

1.1 Ekloginės priklausomybės grafikas

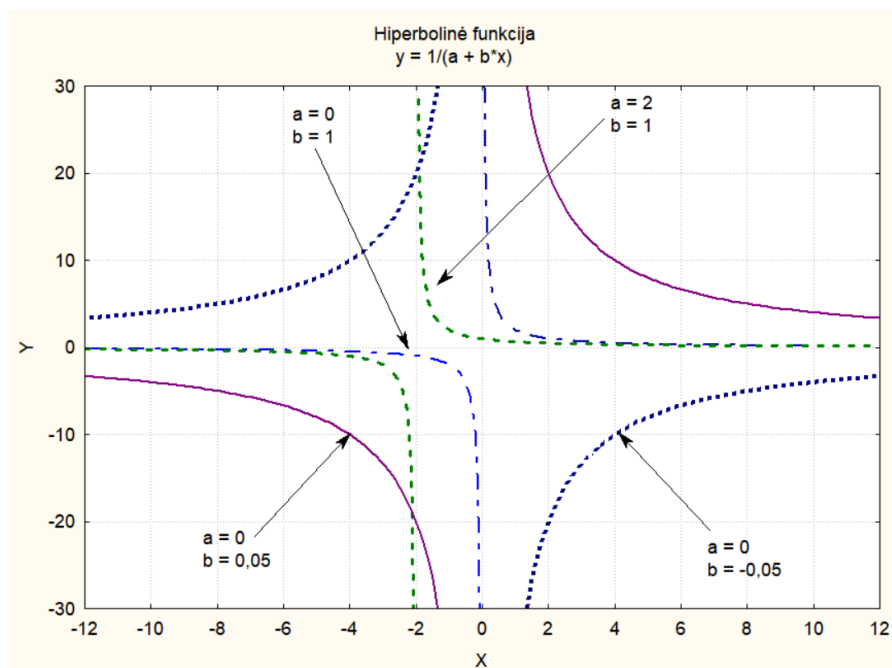
Ekloginės priklausomybės skaičiavime yra naudojami 3 parametrai - a, b, c. Didėjant parametro a reikšmei grafiko viršūnė kyla Y ašies atžvilgiu. Didėjant b parametru grafiko šakos siaurėja. Parametras c lemia grafiko transformaciją palei X ašį. Didėjant c parametro reikšmės vyksta transformacija į teigiamą X ašies pusę link plus begalybės.



pav 1 Ekloginė priklausomybė

1.2 Hiperbolinės priklausomybės grafikas

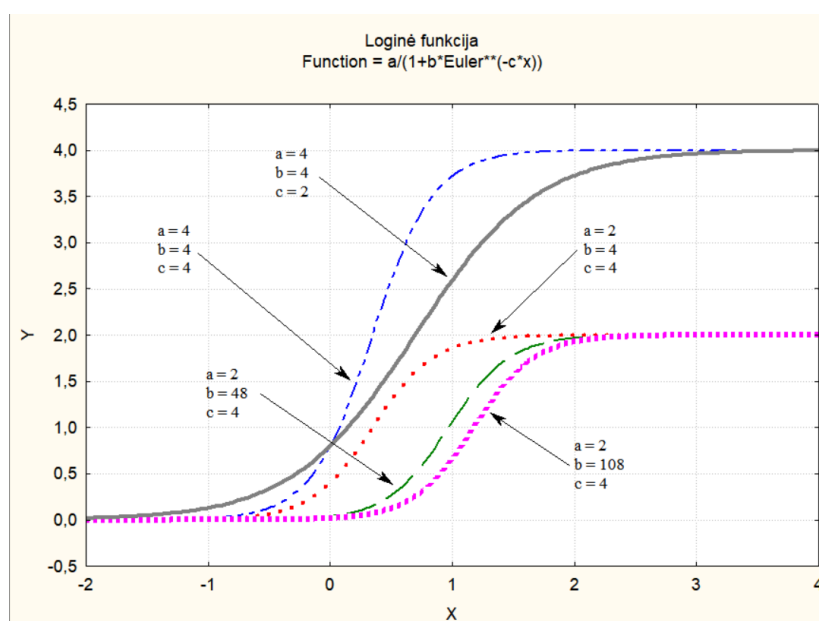
Hiperbolinės priklausomybės skaičiavime yra naudojami 2 parametrai - a ir b. Didėjant parametro a reikšmei vyksta grafiko transformacija dešinę palei X ašį. Mažėjant b parametro reikšmei hiperbolės šakos tolsta viena nuo kitos.



pav 2 Hiperbolinė priklausomybė

1.3 Loginės priklausomybės grafikas

Hiperbolinės priklausomybės skaičiavime yra naudojami 2 parametrai - a, b ir c. Didėjant parametro a reikšmei grafikas plėtėja Y ašies atžvilgiu. Didėjant b parametro reikšmei vyksta grafiko transformacija link plus begalybės palei X ašį. Mažėjant c parametro reikšmei mažėja kampas tarp grafiko ir X ašies.



pav 3 Loginė priklausomybė

2. ANTROJI UŽDUOTIS

Surasti internete duomenis kreivinės regresijos taikymui. Surasti geriausiai tinkančią kreivinės regresinės lygtį. Palyginimui pateikti, bent tris skirtingas kreives. Viena iš jų turi būti brėžta 1 užduotyje. Suskaičiuoti determinacijos koeficientus ir pateikti rezultatų analizę.

2.1 Duomenų pasirinkimo situacija

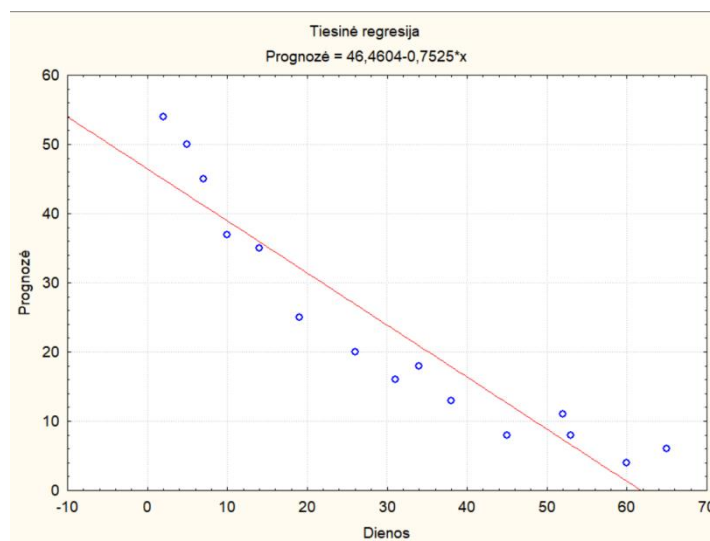
Tiriama 15 žmonių, kurie buvo paguldyti į ligoninę dėl traumų. Vienas duomenų stulpelis rodo dienų skaičių, kurį žmonės praleido ligoninėje. Kitas ilgalaikio gydymo prognozę po pirminės apžiūros [4].

lentelė 1 Ilgalaikio gydymo prognozės duomenys

Dienų skaičius	2	5	7	10	14	19	26	31	34	38	45	52	53	60	65
Prognozė	54	50	45	37	35	25	20	16	18	13	8	11	8	4	6

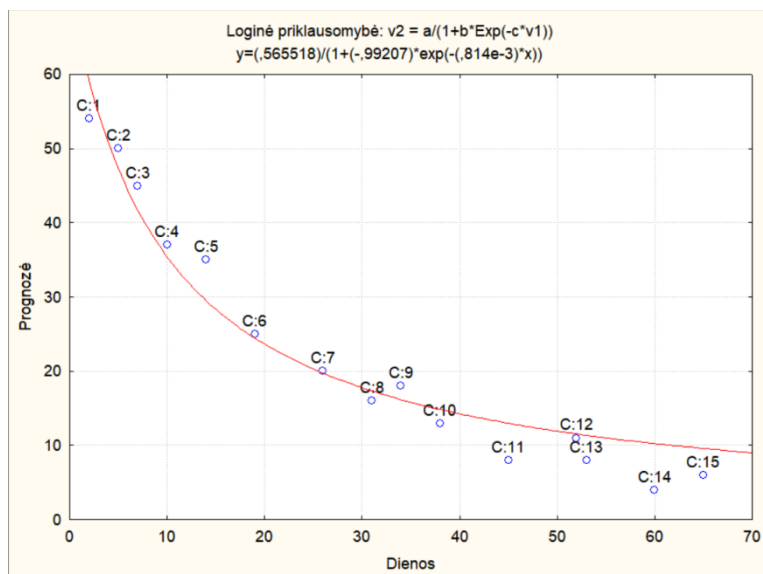
2.2 Geriausios regresinės lygties paieška

Geriausios regresinės lygties ieškome tikrindami skirtingas priklausomybes. Iš žemiau pavaizduoto grafiko matome, jog tiesinė priklausomybė netinka.



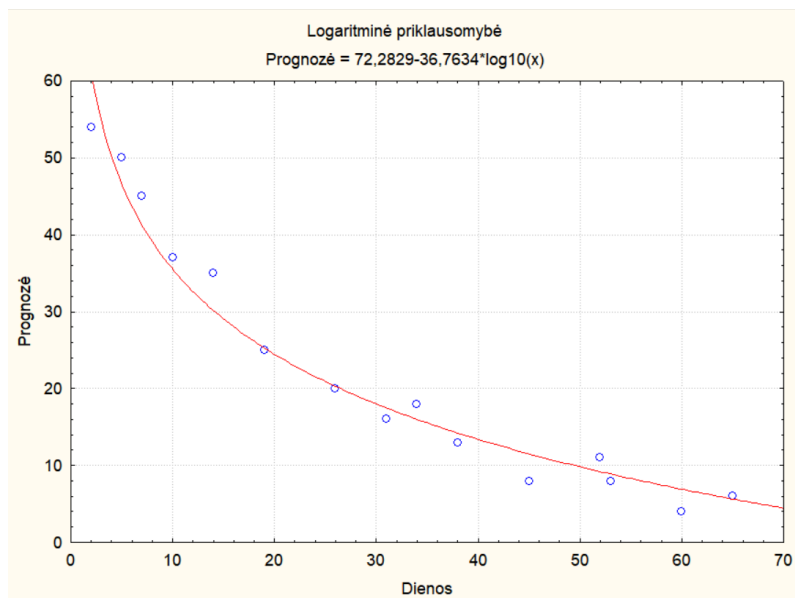
pav 4 Regresinės lygties paieška. Tiesinė

Loginė priklausomybė pakankami gerai atvaizduoja reikšmių pasiskirstymą prognozuojant pusę reikšmių. Kitoje pusėje atsiranda netikslumų.



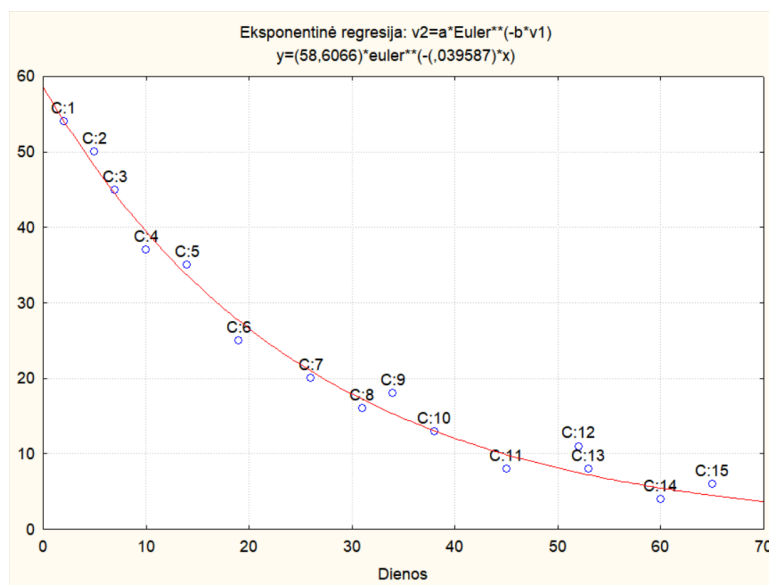
pav 5 Regresinės lygties paieška. Loginė

Logaritminė priklausomybė yra visai tinkama, tik nepakankamai tiksliai vaizduoja grafiko leidimąsi pirmosiomis dienomis.



pav 6 Regresinės lygties paieška. Logaritminė

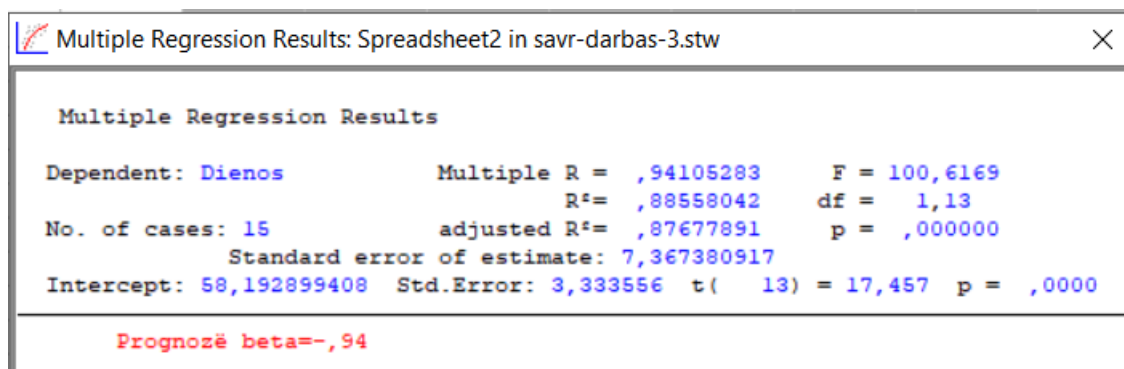
Iš žemiau esančio grafiko aiškiai matosi, jog eksponentinė priklausomybė yra tinkamiausia regresinė lygtis pasirinktiems duomenims. Šios lygties išraiška $y = a \times e^{-bx}$, o šių duomenų atžvilgiu - $y = 58,6066 \times e^{-0,039587x}$.



pav 7 Regresinės lygties paieška. Eksponentinė

2.3 Determinacijos koeficientai

Iš žemiau pavaizduoto lentelės matome, jog koreliacijos koeficientas – 0.94, determinacijos koeficientas – 0.88, o standartinė įvertinimo klaida – 7.36. Rezultatai rodo, kad yra stipri priklausomybė tarp dienų, praleistų ligoninėje ir gydymo prognozės.



pav 8 Determinacijos koeficientai

3. TREČIOJI UŽDUOTIS

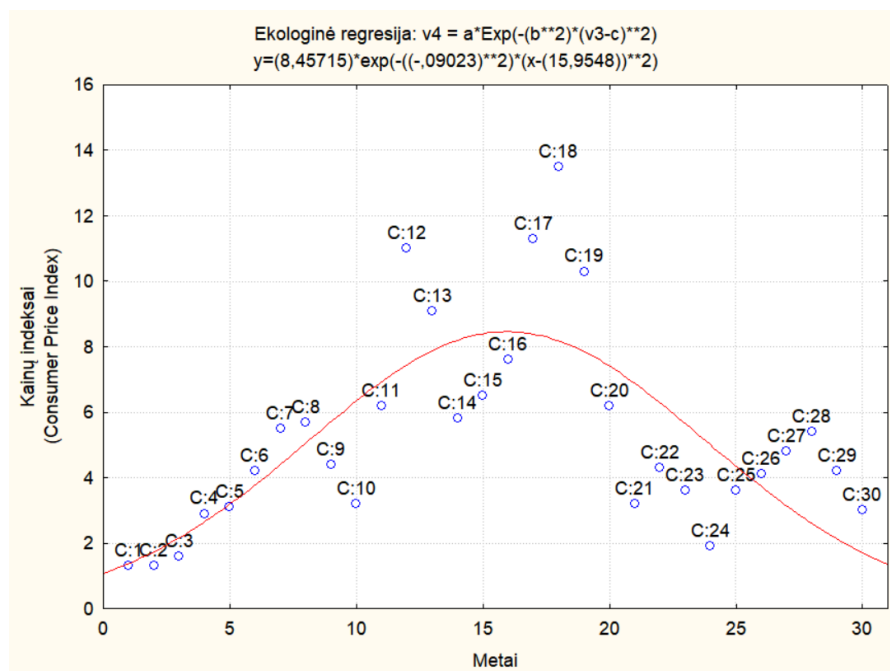
Pirma laiko eilutė turi būti be sezoniškumo komponentės. Reikia nustatyti geriausio trendo lygtį (pabandyti 2 regresines kreives) ir atlikti paprasčiausią prognozę sekantiems dviem laiko periodams naudojant geriausio trendo lygtį.

3.1 Duomenų pasirinkimo situacija

Kasmet JAV buvo skaičiuojamas kainų indekso (angl. Customer Price Index) pokytis [3]. Šiuose duomenyse nustatysime geriausio trendo lygtį.

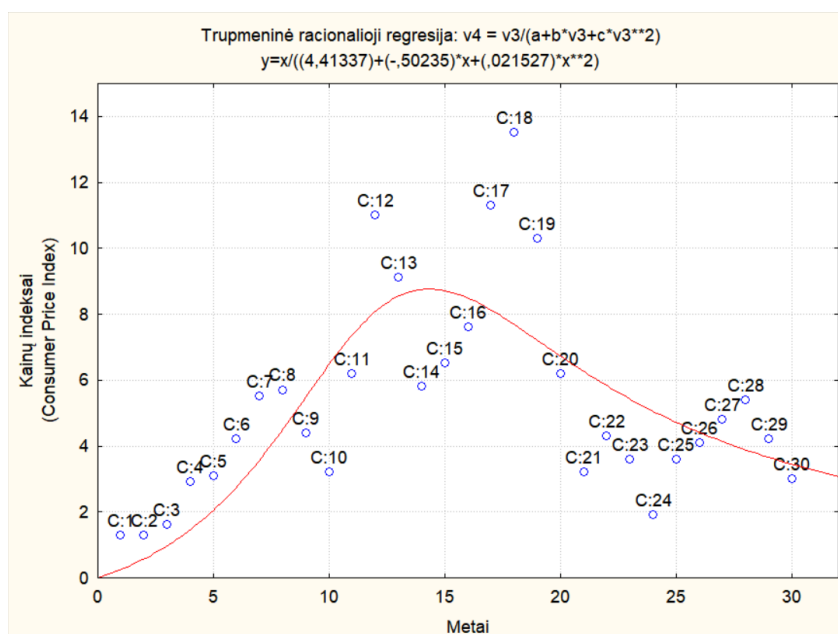
Kadangi pavaizdavus duomenis koordinačių plokštumoje jų pasiskirstymas pasirodė sąlyginai panašus į parabolę, tai natūraliai buvo prieita prie išvados, jog trendo lygtis bus kreivė, savo forma primenanti parabolę. Dėl to tikrinsime ekologinę bei trupmeninę racionalinę priklausomybes.

Ekologinės priklausomybės trūkumas yra tas, jog ji menkai apima naujausių metų duomenis.



pav 9 Trendo lygties paieška. Ekologinė regresija

Trupmeninė racionalioji priklausomybė geriau apima naujausių metų reikšmes, tad darome prielaidą, jog ji yra tikslesnė prognozavimo atžvilgiu nei ekologinė priklausomybė.



pav 10 Trendo lygties paieška. Trupmeninė racionalioji regresija

Dabar remdamiesi trupmeninės racionaliosios priklausomybės formule apskaičiuosime galima sekančių dviejų metų reikšmes pagal formulę $y = \frac{x}{a + bx + cx^2}$. Gauname šiuose rezultatus: 31-ame laiko periode kainos indeksas bus 3,254, o 32-ame - 3,082.

lentelė 2 Reikšmių prognozavimas

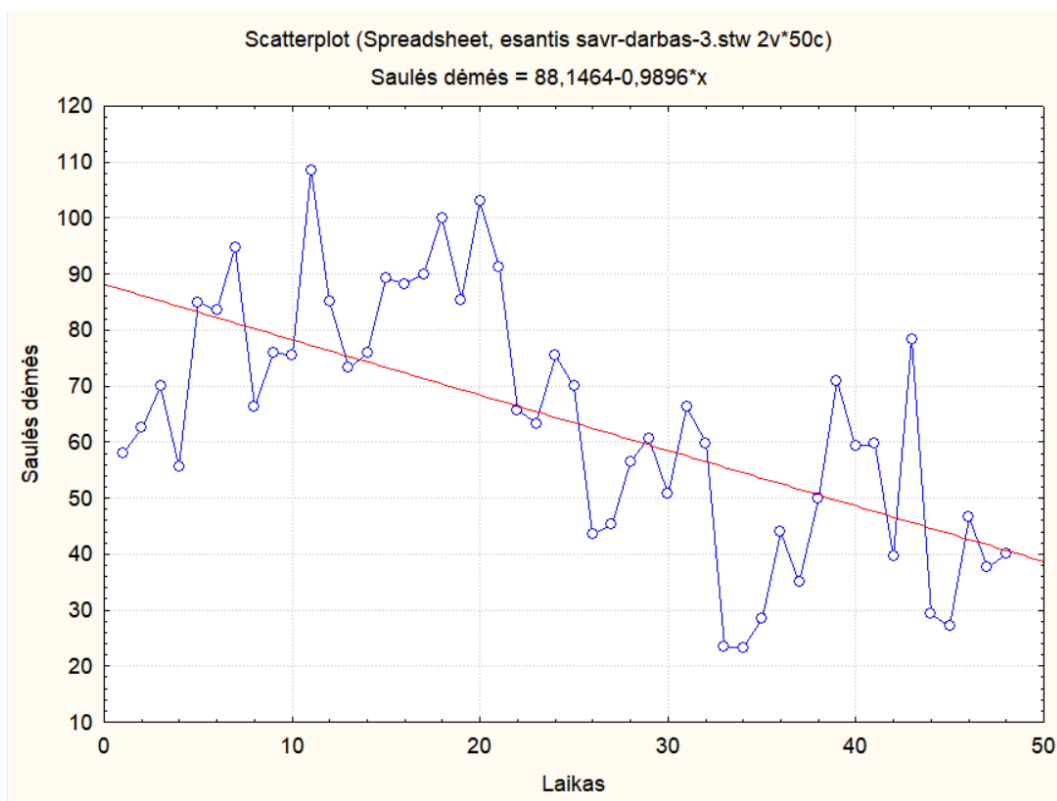
Kintamieji	Reikšmės
a	4,41337
b	-0,50235
c	0,021527
x ₃₁	31
x ₃₂	32
res ₃₁	3,253
res ₃₂	3,082

4. KETVIRTOJI UŽDUOTIS

Antroje laiko eilutėje nustatyti trendą ir suskaičiuoti sezoniškumo indeksus. Atlikti paprasčiausią prognozę sekantiems dviem laiko periodams naudojant trendą ir sezoniškumo indeksus.

4.1 Duomenų pasirinkimo situacija

Nuo 1749 iki 1983 metų buvo stebimas saulės dėmių kiekis. Analizėje naudojame keturių pirmųjų metų duomenis. [5]



Analizuojamų duomenų trendo formulė yra $y = 88,1464 - 0,9896x$

5. PENKTOJI UŽDUOTIS

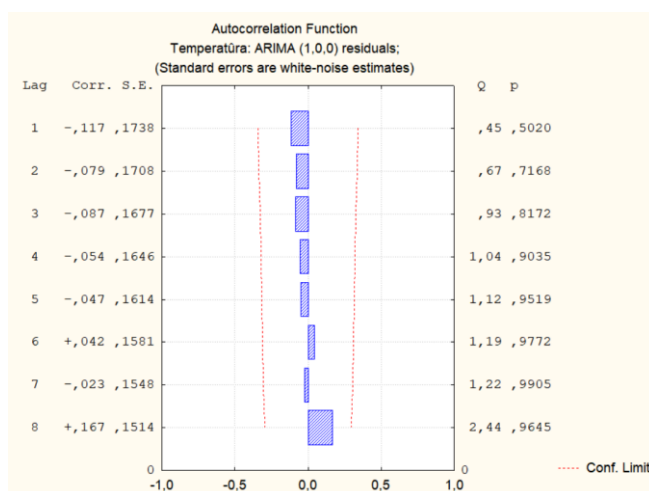
Trečioje eilutėje neturinčioje trendo reikia atlikti dviejų laiko periodų prognozę naudojant bent tris skirtingus prognozavimo metodus. Palyginti gautas prognozes naudojant MSE įvertinimą, nurodyti kuris metodas duoda geriausius rezultatus.

5.1 Duomenų parinkimo situacija

Nagrinėsime duomenis, kurie rodo mažiausią oro temperatūrą 1981 metų sausį [6].

Naudodami autokoreliaciją nusistatome koeficientus slenkamųjų metodų vidurkių taikymui. Jų prireiks prognozuojant.

Iš grafiko matome, jog tinkamiausi koeficientai yra 1, 3 ir 8 pozicijose.



pav 11 autokoreliacija 1

Pasitikriname koeficientus lentelėje ir gauname patvirtinimą, jog tinkamiausi yra 1, 3 ir 8 pozicijose.

Autocorrelation Function (Spreadsheet16 in savr-darbas-3.stw)				
Temperatūra: ARIMA (1,0,0) residuals;				
(Standard errors are white-noise estimates)				
Lag	Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	-0,116678	0,173805	0,450661	0,502026
2	-0,079238	0,170783	0,665928	0,716799
3	-0,086809	0,167705	0,933868	0,817248
4	-0,053859	0,164570	1,040975	0,903518
5	-0,046722	0,161374	1,124802	0,951882
6	0,041655	0,158114	1,194208	0,977167
7	-0,023344	0,154785	1,216954	0,990530
8	0,167422	0,151383	2,440080	0,964460

pav 12 autokoreliacija 2

ŠALTINIAI

1. Virgilijus Sakalauskas. *Statistikos paskaitų medžiaga*. Žiūrėta 2019-12-19 per internetą: VU informacijos šaltiniai.

2. Cengage Learning. *Data Sets, Percent Change In Consumer Price Index*. Žiūrėta 2019-12-19 per internetą:
<https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/tssc/frames/frame.html>.

3. Cengage Learning. *Data Sets, Percent Change In Consumer Price Index*. Žiūrėta 2019-12-19 per internetą:
<https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/tssc/frames/frame.html>.

4. Statsoft Documentation. *Fixed Nonlinear Regression - Example*. Žiūrėta 2019-12-19 per internetą:
<<https://documentation.statsoft.com/STATISTICAHelp.aspx?path=NonlinearRegression/FixedNonLinear/Examples/FixedNonlinearRegressionExamples>>.

5. Machine Learning Mastery. *7 Time Series Datasets for Machine Learning*. Žiūrėta 2019-12-19 per internetą: <<https://machinelearningmastery.com/time-series-datasets-for-machine-learning/>>.<<https://raw.githubusercontent.com/jbrownlee/Datasets/master/monthly-sunspots.csv>>.

6. Machine Learning Mastery. *7 Time Series Datasets for Machine Learning*. Žiūrėta 2019-12-19 per internetą: <<https://machinelearningmastery.com/time-series-datasets-for-machine-learning/>>.<<https://raw.githubusercontent.com/jbrownlee/Datasets/master/daily-min-temperatures.csv>>.