



**KAUNO TECHNOLOGIJOS UNIVERSITETAS
FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS
TAIKOMOSIOS MATEMATIKOS KATEDRA**

Marius Noreika

**MIRTINGUMO NUO GALVOS SMEGENŲ
INSULTO PROGNOZAVIMO MODELIAI IR
PROGRAMINĖS PRIEMONĖS**

Magistro darbas

**Vadovas
doc. dr. V. Janilionis**

**Konsultantė
habil. dr. R. Rėklaitienė
(KMU, Kardiologijos institutas)**

KAUNAS, 2007



KAUNO TECHNOLOGIJOS UNIVERSITETAS
FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS
TAIKOMOSIOS MATEMATIKOS KATEDRA

TVIRTINU
Katedros vedėjas
prof. dr. J.Rimas
2007

MIRTINGUMO NUO GALVOS SMEGENŲ
INSULTO PROGNOZAVIMO MODELIAI IR
PROGRAMINĖS PRIEMONĖS

Taikomosios matematikos magistro baigiamasis darbas

Vadovas

doc. dr. V. Janilionis

2007 05 25

Recenzentas

doc. dr. V. Šaferis

2007 06 05

Atliko

FMMM 5/1 gr. stud.

M. Noreika

2007 05 25

KAUNAS, 2007

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Leonas Saulis, profesorius (VGTU)

Sekretorius: Eimutis Valakevičius, docentas (KTU)

Nariai: Algimantas Jonas Aksomaitis, profesorius (KTU)

Arūnas Barauskas, dr., UAB „Elsis“ generalinio direktoriaus pavaduotojas

Vytautas Janilionis, docentas (KTU)

Zenonas Navickas, profesorius (KTU)

Vidmantas Povilas Pekarskas, profesorius (KTU)

Rimantas Rudzkis, habil. dr., banko „NORD/LB“ vyriausiasis analitikas

Noreika M. Forecasting models and software for mortality from stroke: Master's work in applied mathematics / supervisor dr. assoc. prof. V. Janilionis; Department of Applied mathematics, Faculty of Fundamental Sciences, Kaunas University of Technology. – Kaunas, 2007. – 104 p.

SUMMARY

Stroke is one the most important medical, biological and social problem all over the world. In general mortality from stroke is one of the most frequent causes of deaths among people in Western Europe. Estimation and forecasting of mortality from various diseases are very frequent data analysis tasks in medicine nowadays. In order to estimate expected number of deaths, probability to die from a disease or trends in mortality we should apply the most suitable statistical methods.

The main objectives of this paper were:

- to review data analysis methods applied for mortality from various diseases;
- to create statistic data analysis models for available mortality data;
- to realize created models using SAS system and SAS macro language;
- to apply created models and software for mortality from stroke data supplied by Institute of Cardiology at Kaunas University of Medicine.

Data analysis models were created using Poisson, logistic, Cox regression methods and realized in SAS macros. Created software also contains models for goodness of fit analysis, graphical visualization and prepares a report of data analysis in RTF (Rich Text Format) format. Analysis was made for mortality from stroke data among Kaunas population aged 25 to 64 during the period 1980-2004. The study contains the description of applying created data analysis models, SAS macros and received results.

According to the subject of this paper the presentation has been made at the conference and one article has been prepared for publishing.

Turinys

| | |
|--|------------|
| Turinys..... | 5 |
| Lentelių sąrašas | 6 |
| Paveikslų sąrašas | 7 |
| Įvadas | 8 |
| 1 Teorinė dalis | 9 |
| 1.1 Mirtingumo prognozavimo uždaviniai..... | 9 |
| 1.2 Mirtingumo prognozavimui taikomų metodų lyginamoji analizė | 10 |
| 1.2.1 Logistinė regresija | 11 |
| 1.2.2 Išlikimo teorija ir Kokso regresija | 16 |
| 1.2.3 Puasono regresija | 19 |
| 1.2.4 Amžiaus-periodo-kohortos modeliai | 22 |
| 1.3 Programinių priemonių pasirinkimas | 24 |
| 1.4 Darbe sprendžiami uždaviniai | 27 |
| 2 Tiriamoji dalis ir rezultatai | 28 |
| 2.1 1980-2004 m. Kauno miesto gyventojų mirtingumo duomenys | 28 |
| 2.2 Mirtingumo prognozavimo modeliai ir jų taikymo rezultatai | 29 |
| 2.2.1 Puasono regresijos modeliai | 29 |
| 2.2.1.1 Puasono regresijos modelis lytis-amžius..... | 31 |
| 2.2.1.2 Puasono regresijos modelis lytis-amžius-periodas | 35 |
| 2.2.1.3 Puasono regresijos modelis lytis-amžius-periodo dreifas | 38 |
| 2.2.1.4 Puasono regresijos modelis lytis-amžius-kohorta | 41 |
| 2.2.1.5 Puasono regresijos modelis lytis-amžius-kohortos dreifas..... | 44 |
| 2.2.1.6 Puasono regresijos modelis lytis-amžius-periodas-kohorta | 46 |
| 2.2.2 Logistinės regresijos modelis..... | 49 |
| 2.2.3 Kokso regresijos modelis..... | 53 |
| 2.3 SAS makro komandų rinkinys..... | 55 |
| Išvados | 57 |
| Literatūra | 58 |
| 1 priedas. Makro komandų rinkinio programos tekstas..... | 59 |
| 2 priedas. Modelio lytis-amžius taikymo rezultatai..... | 81 |
| 3 priedas. Modelio lytis-amžius-periodas taikymo rezultatai | 85 |
| 4 priedas. Modelio lytis-amžius-periodas dreifas taikymo rezultatai | 89 |
| 5 priedas. Modelio lytis-amžius-kohorta taikymo rezultatai | 92 |
| 6 priedas. Modelio lytis-amžius-kohortos dreifas taikymo rezultatai | 97 |
| 7 priedas. Modelio lytis-amžius-periodas-kohorta taikymo rezultatai..... | 100 |

Lentelių sąrašas

| | |
|--|-----|
| 1.1 lentelė Duomenų pavyzdys | 23 |
| 1.2 lentelė Kohortos dalinimas..... | 24 |
| 2.1 lentelė I tyrimo kintamieji..... | 28 |
| 2.2 lentelė Atrinkti II tyrimo kintamieji | 28 |
| 2.3 lentelė Puasono regresijos modelių kintamieji..... | 30 |
| 2.4 lentelė Modelio lytis-amžius regresijos lygties parametru įverčiai | 32 |
| 2.5 lentelė Modelio lytis-amžius-periodas regresijos lygties parametru įverčiai..... | 36 |
| 2.6 lentelė Modelio lytis-amžius-periodo dreifas regresijos lygties parametru įverčiai | 39 |
| 2.7 lentelė Modelio lytis-amžius-kohortos dreifas regresijos lygties parametru įverčiai..... | 45 |
| 2.8 lentelė Logistinės regresijos modelio lygties parametru įverčiai | 51 |
| 2.9 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 52 |
| 2.10 lentelė Pirsono paklaidų normalumo tikrinimas..... | 52 |
| 2.11 lentelė Kokso regresijos lygties parametru ir santykinės rizikos koeficientu įverčiai | 54 |
| 2.12 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 55 |
| 2.13 lentelė Martingalinių paklaidų normalumo tikrinimas | 55 |
| 2 priedas 1 lentelė Puasono regresijos modelio lytis-amžius rezultatai | 81 |
| 2 priedas 2 lentelė Santykinės rizikos koeficiento skirtingoms amžiaus grupėms įverčiai ir pasikliautiniai intervalai..... | 82 |
| 2 priedas 3 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 82 |
| 2 priedas 4 lentelė Pirsono paklaidų normalumo tikrinimas..... | 83 |
| 2 priedas 5 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas | 83 |
| 3 priedas 1 lentelė Puasono regresijos modelio lytis-amžius-periodas rezultatai..... | 85 |
| 3 priedas 2 lentelė Santykinės rizikos koeficiento skirtingiems periodams įverčiai ir pasikliautiniai intervalai..... | 86 |
| 3 priedas 3 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 86 |
| 3 priedas 4 lentelė Pirsono paklaidų normalumo tikrinimas..... | 86 |
| 3 priedas 5 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas | 87 |
| 4 priedas 1 lentelė Puasono regresijos modelio lytis-amžius-periodo dreifas rezultatai..... | 89 |
| 4 priedas 2 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 90 |
| 4 priedas 3 lentelė Pirsono paklaidų normalumo tikrinimas..... | 90 |
| 4 priedas 4 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas..... | 90 |
| 5 priedas 1 lentelė Puasono regresijos modelio lytis-amžius-kohorta rezultatai | 92 |
| 5 priedas 2 lentelė Modelio lytis-amžius-kohorta regresijos lygties parametru įverčiai..... | 93 |
| 5 priedas 3 lentelė Santykinės rizikos koeficiento skirtingoms kohortoms įverčiai ir pasikliautiniai intervalai..... | 94 |
| 5 priedas 4 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 95 |
| 5 priedas 5 lentelė Pirsono paklaidų normalumo tikrinimas..... | 95 |
| 5 priedas 6 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas | 95 |
| 6 priedas 1 lentelė Puasono regresijos modelio lytis-amžius-kohortos dreifas rezultatai | 97 |
| 6 priedas 2 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 98 |
| 6 priedas 3 lentelė Pirsono paklaidų normalumo tikrinimas..... | 98 |
| 6 priedas 4 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas | 98 |
| 7 priedas 1 lentelė Puasono regresijos modelio lytis-amžius-periodas-kohorta rezultatai | 100 |
| 7 priedas 2 lentelė Modelio lytis-amžius-periodas-kohorta regresijos lygties parametru įverčiai | 101 |
| 7 priedas 3 lentelė Anomaliųjų paklaidų normalumo tikrinimas | 102 |
| 7 priedas 4 lentelė Pirsono paklaidų normalumo tikrinimas..... | 102 |
| 7 priedas 5 lentelė Frymano-Tjukio paklaidų normalumo tikrinimas | 102 |
| 7 priedas 6 lentelė Puasono regresijos modelių tinkamumo analizuotiems duomenims rezultatai | 104 |

Paveikslų sąrašas

| | |
|---|----|
| 1.1 pav. Tiesinė ir logistinė regresijos funkcijos..... | 12 |
| 2.1 pav. Modelio lytis-amžius Frymano-Tjukio paklaidų grafikas..... | 33 |
| 2.2 pav. Modelio lytis-amžius tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo amžiaus.. | 34 |
| 2.3 pav. Modelio lytis-amžius vyrų tikėtinų mirtingumo dažnių (1000 gyv.) ir jų pasikliautinųjų intervalų priklausomybė nuo amžiaus | 34 |
| 2.4 pav. Modelio lytis-amžius-periodas Frymano-Tjukio paklaidų grafikas..... | 37 |
| 2.5 pav. Modelio lytis-amžius-periodas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo amžiaus..... | 37 |
| 2.6 pav. Modelio lytis-amžius-periodas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo..... | 38 |
| 2.7 pav. Modelio lytis-amžius-periodo dreifas Frymano-Tjukio paklaidų grafikas | 40 |
| 2.8 pav. Modelio lytis-amžius-periodo dreifas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo..... | 41 |
| 2.9 pav. Modelio lytis-amžius-periodo dreifas vyrų stebėtų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo..... | 41 |
| 2.10 pav. Modelio lytis-amžius-kohorta Frymano-Tjukio paklaidų grafikas..... | 43 |
| 2.11 pav. Modelio lytis-amžius-kohorta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos | 43 |
| 2.12 pav. Modelio lytis-amžius-kohortos dreifas Frymano-Tjukio paklaidų grafikas..... | 46 |
| 2.13 pav. Modelio lytis-amžius-kohortos dreifas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos | 46 |
| 2.14 pav. Modelio lytis-amžius-periodas-kohorta Frymano-Tjukio paklaidų grafikas..... | 48 |
| 2.15 pav. Modelio lytis-amžius-periodas-kohorta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) diagrama..... | 48 |
| 2.16 pav. Logistinės regresijos modelio ROC kreivė..... | 52 |
| 2.17 pav. Makro komandų rinkinio struktūra | 56 |

Įvadas

Galvos smegenų insultas (GSI) – tai medicininė, biologinė ir socialinė problema. Vakarų Europos šalyse GSI yra trečia pagal dažnį mirties priežastis. GSI sukeltos pasekmės – daugiau nei pusė persirgusiųjų lieka neįgalūs, o ketvirtadalis dėl GSI invalidais pripažįstamų asmenų būna darbingo amžiaus.

Mirtingumo nuo įvairių ligų įvertinimas ir prognozavimas pagal atlikto tyrimo duomenis – dažnas statistinės analizės uždavinys medicinoje. Juose siekiama prognozuoti tikėtiną mirčių nuo tiriamos ligos skaičių, susirgimo tam tikra liga tikimybę ar išskirti rizikos grupes, įvertinant tyrimo metu surinktų stebimos populiacijos imties kintamųjų duomenis ir nustatant, kokia priklausomybę juos sieja.

Pagrindinis šio tiriamojo darbo tikslas:

- susipažinti su statistikos metodais, taikomais mirtingumo duomenų analizei Lietuvoje ir užsienyje;
- sudaryti statistinės analizės modelius turimiems mirtingumo duomenims;
- realizuoti sudarytus modelius programiškai, panaudojant SAS sistemą ir SAS makro programavimo galimybes;
- panaudojus sukurtas programines priemones atlikti Kauno medicinos universiteto Kardiologijos instituto surinktų mirtingumo nuo galvos smegenų insulto duomenų analizę;
- įvertinti taikomų metodų tinkamumą analizuojamiems duomenims ir parengti sukurtų priemonių taikymo rekomendacijas.

Panaudojus Puasono, logistinės ir Kokso regresinės analizės metodus sudaryti mirtingumo nuo GSI prognozavimo modeliai. Sudaryti modeliai realizuoti programiškai, panaudojus SAS programavimo kalbą, SAS/IML posistemės galimybes ir SAS makro programavimo priemones. Sukurti regresinės analizės modeliai ir programines priemones panaudotos Kauno medicinos universiteto Kardiologijos instituto darbuotojų 1980-2004 metais atliktų tyrimų metu surinktų Kauno miesto 25-64 m. amžiaus gyventojų mirtingumo nuo GSI duomenų analizei atlikti.

Darbo tema buvo perskaitytas pranešimas konferencijoje “Matematika ir matematikos dėstymas – 2007” (KTU) ir parengta publikacija žurnalui „Matematika ir matematinis modeliavimas“:

Noreika M., Janilionis V., Rėklaitienė R., Tamošiūnas A. Mirtingumo nuo galvos smegenų insulto Puasono regresijos modeliai // Matematika ir matematinis modeliavimas / Kauno technologijos universitetas ISSN 1822-2757, T.3, (2007).

1 Teorinė dalis

1.1 Mirtingumo prognozavimo uždaviniai

Dauguma duomenų analizės uždavinių yra susiję su prognozavimu. Analizės metodų gausa (tiesiniai ir netiesiniai regresiniai modeliai, neuroniniai tinklai, sprendimų medžiai ir kt.) uždavinio sprendimui leidžia rinktis tinkamiausią prognozių gavimo būdą. Tinkamo metodo parinkimas turimiems duomenims yra svarbus duomenų analizės uždavinys. Regresiniai duomenų analizės metodai leidžia įvertinti stebimo populiacijos kintamojo reikšmes (pavyzdžiui, įvykio tikimybę), kaip tiesinę arba netiesinę populiacijos kintamųjų funkciją. Regresinė analizė – efektyvus eksperimentų ar tyrimų metu sukauptų duomenų (diskrečių bei tolydžių kintamųjų, cenzūruotų ir pan.) analizės būdas.

Mirtingumo nuo įvairių ligų įvertinimas ir prognozavimas pagal atlikto tyrimo duomenis – dažnas statistinės analizės uždavinys medicinoje. Literatūros [1, 3-6, 11-18] apžvalga parodė, kad mirtingumo uždavinių analizei dažniausiai naudojami regresinės analizės metodai. Sudarydami įvairius regresijos modelius galime prognozuoti tikėtiną mirčių nuo tiriamos ligos skaičių, susirgimo tam tikra liga tikimybę ar išskirti rizikos grupes, įvertindami tyrimo metu surinktų stebimos populiacijos duomenis ir nustatydami, kokia priklausomybę juos sieja. Pateiksime keletą tokių duomenų analizės uždavinių pavyzdžių:

- 1) Framingham Hearty Study tyrimo duomenys (tirti 4699 pacientai): paciento lytis, amžius, diastolinis ir sistolinis kraujo spaudimas, cholesterolio kiekis kraujyje, kūno masės indeksas ir pan. Panaudojus Kokso regresijos metodą atlikta duomenų analizė, įvertintas mirtingumas nuo širdies kraujagyslių ligų bei nustatytos atitinkamos populiacijos rizikos grupės. [3]
- 2) 1971-2000 metais tarp Kauno miesto 45-59 metų amžiaus vyrų Kauno medicinos universiteto Kardiologijos institute atliktas tyrimas, kurio tikslas – įvertinti sociodemografinių, gyvenamos rizikos veiksnių (tokių kaip išsilavinimas, šeimyninė padėtis, rūkymas, alkoholio suvartojimas ir pan.) įtaką mirtingumui nuo išorinių priežasčių (traumų, nelaimingų atsitikimų, savižudybių). Sukauptų duomenų analizei panaudoti logistinės ir Kokso regresinės analizės metodai. [18]
- 3) Mirčių nuo plaučių vėžio 1960-1979 metais tarp 30-84 metų amžiaus Škotijos vyrų tyrimas. Įvertinta amžiaus grupės, laikotarpio ir gimimo metų laikotarpio įtaka mirtingumo nuo plaučių vėžio dažnio trendui. Darbe panaudoti Puasono regresijos modeliai. [14]
- 4) Mirčių nuo plaučių vėžio dažnio 1940-1984 metais tarp 20-84 metų amžiaus moterų tyrimas atliktas Konektikute (JAV). Duomenų analizei panaudoti Puasono regresinės analizės modeliai, įvertinta amžiaus grupės, laikotarpio ir gimimo metų laikotarpio įtaka mirtingumo nuo plaučių vėžio dažnio trendui. [17]

- 5) 25-79 metų amžiaus moterų mirtingumo nuo krūties vėžio Japonijoje tyrimas, atliktas 1955-1979 metais. Duomenų analizei panaudoti Puasono regresijos modeliai, įvertinta amžiaus grupės, laikotarpio bei gimimo metų laikotarpio įtaka mirčių dažnio trendui. [16]

1.2 Mirtingumo prognozavimui taikomų metodų lyginamoji analizė

Duomenų analizės uždavinių sprendimui taikant regresinę analizę pirmiausia parenkamas regresijos modelio tipas – nusprendžiama, kokia priklausomybė sieja kintamuosius [10]. Pasirinkus atitinkamą modelį reikia:

- įvertinti nežinomus modelio parametrus;
- patikrinti, ar pasirinktasis modelis suderinamas su duomenimis;
- taikyti regresijos modelį prognozavimui.

Regresijos modelio parinkimas yra sudėtingas uždavinys. Gali būti atvejų kai nagrinėjamiems duomenims netinka joks regresijos modelis.

Duomenų analizės metodo parinkimas priklauso nuo duomenų prigimties. Pagal matavimo prigimtį kintamieji skirstomi į kiekybinius ir kokybinius [9]. Kokybinis kintamasis vadinamas kategoriniu, jei jis įgyja baigtinį reikšmių skaičių. Atliekant tyrimus dažnai analizuojami ir kiekybiniai, ir kategoriniai kintamieji. Pavyzdžiui, analizuojami kiekybiniai kintamieji ūgis, svoris, laikas, kūno-masės indeksas, testo balai bei kategoriniai kintamieji lytis, alkoholio vartojimo lygis, rūkymo lygis, amžiaus grupė, kraujo grupė ir t.t. Kategorinius kintamuosius įrašyti į regresijos modelį sudėtingiau nei kiekybinius kintamuosius. Galima sudaryti regresijos modelius atskiroms kategorinių kintamųjų kategorijoms (pvz.: kintamojo lytis reikšmėms vyras arba moteris), tačiau toks būdas neefektyvus. Regresinėje analizėje kategoriniai kintamieji į nepriklausomų kintamųjų aibę įtraukiami naudojant pseudokintamuosius (fiktyvius kintamuosius). Naudodami pseudokintamuosius gauname kiekvieną kategorijų kombinaciją atitinkančią regresijos funkciją. Visose regresinės analizės statistinėse išvadose naudojamas nežinomos paklaidos dispersijos įvertis. Pseudokintamieji leidžia įvertinti visus stebėjimus, todėl dispersijos įvertis yra tikslesnis ir gaunamos tikslesnės statistinės išvados [9, 10].

Kategorinis kintamasis į regresijos modelį įtraukiamas ne tiesiogiai, o pakeičiant jį vienu arba keliais dvireikšmiais pseudokintamaisiais [10]. Tarkime, kad kategorinis kintamasis lytis gali įgyti dvi kategorijas. Tuomet regresijos modelis papildomas vienu dvireikšmiu pseudokintamuoju \tilde{X}_1

$$\tilde{X}_1 = \begin{cases} 1, & \text{vyras,} \\ 0, & \text{moteris.} \end{cases}$$

Kai kategorinis kintamasis turi $m > 2$ kategorijų, tai jis keičiamas $m - 1$ pseudokintamaisiais. Vieną iš kategorijų kombinacijų atitinka pseudokintamųjų nulinės reikšmės, o kiekvieną iš likusių kategorijų – visų pseudokintamųjų nulinės reikšmės išskyrus vieną, kuris įgyja reikšmę 1. Į regresijos

modelį galima įtraukti kelis kategorinius kintamuosius ir tokiu atveju kiekvieną kategorinį kintamąjį pseudokintamaisiais keičiame atskirai [10].

Darbe aptarsime prognozavimo uždaviniuose taikomus mirtingumo duomenų analizės metodus, duomenų (kintamųjų) tipus bei programinės įrangos paketo SAS panaudojimo galimybes sprendžiant šiuos uždavinius. Literatūros [1, 3-6, 11-18] apžvalga parodė, kad tokių uždavinių analizei naudojami šie regresinės analizės metodai:

- Puasono regresija;
- logistinė regresija;
- Kokso regresija.

Apžvelgsime šiuos regresinės analizės metodus ir jų taikymą konkreitiems duomenims panaudojant duomenų analizės sistemos SAS galimybes.

1.2.1 Logistinė regresija

Logistinės regresijos tikslas – prognozuoti kategorinio kintamojo reikšmių tikimybes pagal nepriklausomų kintamųjų reikšmes [10]. Tokių uždavinių pavyzdžiai:

- pagal paciento svorį bei kraujo parametrus norime sužinoti, kiek tikėtina pacientui susirgti diabetu;
- žinodami rinkėjo pajamas, amžių ir išsilavinimą norime sužinoti tikimybę, kad jis rinkimuose balsuos už kairiųjų kandidatą.

Tokiais atvejais, kai priklausomi kategoriniai kintamieji yra dvireikšmiai, jų reikšmių tikimybėms prognozuoti galime taikyti logistinę regresiją. Kategorinio kintamojo reikšmių prognozavimas tam tikra prasme yra klasifikavimo uždavinys [10]. Pavyzdžiui, pagal paciento svorį bei kraujo parametrus įvertinęs susirgimo diabetu tikimybę, gydytojas paskiria (arba nepaskiria) vaistus, t.y. tokiu būdu pacientas klasifikuojamas į vieną iš dviejų grupių.

Tarkime, kad kategorinis kintamasis įgyja dvi reikšmes 1 ir 0. Užrašykime tiesinės regresijos lygtį

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in} + \varepsilon_i, \quad (1.1)$$

čia Y_i – atsitiktinis dydis, įgyjantis reikšmes 0 arba 1 su tikimybėmis $P(Y_i = 1) = \theta_i$, $P(Y_i = 0) = 1 - \theta_i$, $\beta_0, \beta_1, \dots, \beta_n$ – regresijos modelio parametrai, X_{i1}, \dots, X_{in} – regresijos modelio kintamieji, o ε_i yra atsitiktinė paklaida. Dydžio Y_i prognozėms negalima taikyti (1.1) formule apibrėžto modelio regresijos, nes [10]:

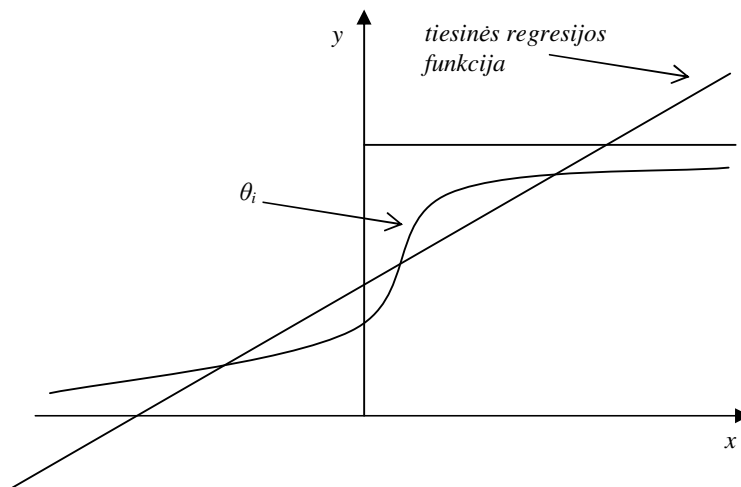
- Netenkinamos tiesinės regresijos prielaidos (Y_i įgyja tik dvi reikšmes), todėl ir ε_i taip pat gali įgyti tik dvi reikšmes (netenkinama normalumo prielaida) [10];

- Tiesinėje regresijoje prognozuojamas priklausomo kintamojo vidurkis. Kadangi $EY_i = \theta_i$, tai prognozuojame tikimybę, su kuria Y_i įgyja reikšmę 1. Bet dešinioji (1.1) lygybės pusė gali įgyti neigiamas ar didesnes už vienetą reikšmes, o tai prieštarauja tikimybės apibrėžimui.

Tikimybės θ_i priklausomybės nuo X_{i1}, \dots, X_{it} įvertinimui taikysime logistinės regresijos modelį [1, 2, 10]:

$$\theta_i = \frac{\exp\left(\beta_0 + \sum_{k=1}^t \beta_k X_{ik}\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^t \beta_k X_{ik}\right)}. \quad (1.2)$$

Tarkime, kad turime vieną nepriklausomą kintamąjį X . Tada abiejų modelių skirtumus galima pavaizduoti grafiškai (1.1 pav.).



1.1 pav. Tiesinė ir logistinė regresijos funkcijos

Iš (1.2) lygties gauname [1, 10]:

$$\frac{\theta_i}{1 - \theta_i} = \exp\left(\beta_0 + \sum_{k=1}^t \beta_k X_{ik}\right) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_t X_{it}), \quad (1.3)$$

tuomet

$$\ln\left(\frac{\theta_i}{1 - \theta_i}\right) = \beta_0 + \sum_{k=1}^t \beta_k X_{ik} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_t X_{it}. \quad (1.4)$$

Tikimybių santykis $\frac{\theta_i}{1 - \theta_i}$ vadinamas galimybe įvykti įvykiui $Y_i = 1$. Įvykio $Y_i = 1$ galimybė įvykti yra didesnė už 1 tada ir tik tada, kai $P(Y_i = 1) > P(Y_i = 0)$ [10].

Galimybės logaritmas $\ln\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \sum_{k=1}^t \beta_k X_{ik}$ nuo kintamųjų reikšmių X_{i1}, \dots, X_{it} priklauso tiesiškai. Be to (priešingai nei θ_i), jis gali įgyti reikšmes iš intervalo $(-\infty, \infty)$. Todėl šiuo atveju taikant (1.1) modelį nekyla problemos, kad prognozuojama reikšmė nepateks į leistinų reikšmių intervalą [10]. Kai $\beta_0 + \sum_{k=1}^t \beta_k X_{ik}$ įgyja mažas reikšmes, tai θ_i artimas nuliui; kai $\beta_0 + \sum_{k=1}^t \beta_k X_{ik}$ įgyja dideles reikšmes, tai θ_i artimas vienetui [10].

Logistinė regresija tinka galiojant gana bendroms prielaidoms: kintamieji X_1, \dots, X_t nebūtinai turi būti normalieji; nereikalaujama normaliai pasiskirsčiusių paklaidų; nekalbama apie priklausomo kintamojo homoskedastiškumą. Taikant tas pačias kodavimo taisykles, kaip ir daugialypėje regresijoje, į modelį galima įtraukti pseudokintamuosius [10].

Tarkime duomenis sudaro stebėjimai $(Y_i, X_{i1}, X_{i2}, \dots, X_{it})$, $i = \overline{1, n}$; čia Y_i yra 0 arba 1, o $X_{i1}, X_{i2}, \dots, X_{it}$ – intervalinių kintamųjų arba pseudokintamųjų reikšmės. Kaip ir tiesinės regresijos atveju, parametų $\beta_0, \beta_1, \dots, \beta_t$ įverčius $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t$ reikia parinkti taip, kad modelis (1.2) būtų kuo geriau suderintas su turimais duomenimis. Tam taikomas maksimalaus tikėtimumo metodas. Kiekvienam stebėjimui pagal (1.2) formulę skaičiuojama tikimybė θ_i . Parametų įverčiai $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t$ parenkami taip, kad tikėtimumo funkcijos

$$L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t) = \prod_{i=1}^n \hat{\theta}_i^{Y_i} \prod_{i=1}^n (1 - \hat{\theta}_i)^{1-Y_i} \quad (1.5)$$

reikšmė būtų maksimali, čia

$$\hat{\theta}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_t X_{it})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_t X_{it})}. \quad (1.6)$$

Parametų įverčių skaičiavimas – iteracinis procesas. Didžiausiojo tikėtimumo metodas realizuotas daugelyje statistinių paketų [10].

Apžvelgsime kaip interpretuojamos regresijos lygties parametų β_k , $k = \overline{1, t}$ reikšmės. Tiriant intervalinius kintamuosius, β_k parodo, kiek padidėja (sumažėja) $\beta_0 + \beta_1 X_1 + \dots + \beta_t X_t$ reikšmė, kai X_k padidėja vienetu, o likusieji $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_t$ yra fiksuoti. Pseudokintamųjų atveju skirtingoms β_k reikšmėms gaunamos skirtingos galimybės logaritmo (1.4 formulė) išraiškos [10]. Tačiau galimybės logaritmo pokyčius sunku interpretuoti. Todėl paprastai skaičiuojamas vadinamasis galimybių santykis (angl. odds ratio) OR [10].

Tuomet intervalinių kintamųjų atveju galimybių santykio formulė yra

$$OR = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k (X_k + 1) + \dots + \beta_t X_t)}{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \dots + \beta_t X_t)} = \exp(\beta_k). \quad (1.7)$$

Kategorinių kintamųjų atveju (kai naudojami pseudokintamieji) galimybių santykis gaunamas analogiškai [7, 10]:

$$OR = \frac{\exp(\beta_0 \cdot 1 + \beta_1 \cdot 1 + \dots + \beta_{k-1} \cdot 1 + \beta_k \cdot 1 + \beta_{k+1} \cdot 1 + \dots + \beta_t \cdot 1)}{\exp(\beta_0 \cdot 1 + \beta_1 \cdot 1 + \dots + \beta_{k-1} \cdot 1 + \beta_k \cdot 0 + \beta_{k+1} \cdot 1 + \dots + \beta_t \cdot 1)} = \exp(\beta_k), \quad (1.8)$$

čia vienu atveju k -tojo pseudokintamojo reikšmė lygi 1, kitu atveju – 0, likusių pseudokintamųjų reikšmės abiem atvejais lygios 1.

Sudarant logistinės regresijos modelį turimiems duomenims reikia įvertinti taikomo modelio tinkamumą. Šis uždavinys sprendžiamas analizuojant paklaidas [7].

Tarkime stebime s grupių ($i = 1, \dots, s$) po n_i stebimų elementų kiekvienoje grupėje. Jeigu y_i yra įvykių skaičius i -tojoje grupėje, o $\hat{\theta}_i$ pažymėsime prognozuojamą įvykio tikimybę i -tojoje grupėje, tuomet logistinės regresijos atveju Pirsono paklaida yra [1, 7]

$$e_i = \frac{y_i - n_i \hat{\theta}_i}{\sqrt{n_i \hat{\theta}_i (1 - \hat{\theta}_i)}}. \quad (1.9)$$

Pirsono paklaida įvertina skirtumą tarp stebėtų ir prognozuojamų įvykių skaičiaus. Paklaidos, kurių reikšmės nepatenka į intervalą $[-2; 2]$ priskiriamos išskirtims [1].

Logistinės regresijos modeliui skaičiuojama ir anomalioji (angl. deviance) paklaida

$$d_i = \text{sgn}(y_i - \hat{y}_i) \left[2y_i \cdot \log\left(\frac{y_i}{\hat{y}_i}\right) + 2(n_i - y_i) \cdot \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right) \right]^{\frac{1}{2}}, \quad (1.10)$$

čia $\hat{y}_i = n_i \hat{\theta}_i$.

Daugialypėje regresijoje mus domina ar visi į modelį įtraukti nepriklausomi kintamieji prognozavimui svarbūs. Aišku, jeigu prognozuojama tikimybė nuo kintamojo X_k nepriklauso, tai (1.2) modelio parametras $\beta_k = 0, k = \overline{1, t}$ [10].

Tarkime, maksimalaus tikėtimumo metodu randame parametrų įverčius $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t$, įstatome juos į maksimalaus tikėtimumo funkciją L (1.5) ir randame jos maksimumą, kuriį pažymėkime $L(\hat{\beta}) = \max_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t} L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t)$. Tarkime, pasirinkome logistinės regresijos modelį, kuriame visi $\beta_k = 0, k = \overline{1, t}$. Maksimalaus tikėtimumo funkcijos maksimumą šiuo atveju pažymėkime $L(0) = \max_{\beta_1=0, \dots, \beta_t=0} L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t)$. χ^2 kriterijus remiasi tuo, kad $L(\hat{\beta})$ mažai skiriasi nuo $L(0)$, jei visi $\beta_k = 0, k = \overline{1, t}$ [10].

$$\text{Statistinė hipotezė: } \begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_t = 0, \\ H_1 : \text{ bent vienas } \beta_k \neq 0. \end{cases}$$

$$\text{Kriterijaus statistika: } T_{\chi^2} = -2 \ln L(0) + 2 \ln L(\hat{\beta}).$$

Hipotezė H_0 atmetama su reikšmingumo lygmeniu α , jei $T_{\chi^2} > \chi_{\alpha}^2(t)$. Neatmesta nulinė hipotezė rodo, kad logistinės regresijos modelis netinka. Atmesta nulinė hipotezė H_0 rodo, kad bent vienas nepriklausomas kintamasis modelyje yra reikalingas [10].

Kitas kriterijus regresijos koeficientų reikšmingumui nustatyti yra Voldo (Wald) kriterijus. Voldo kriterijus yra Stjudento kriterijaus tiesinėje regresijoje analogas. Jis parodo ar koeficientas $\beta_k \neq 0$ [10].

$$\text{Statistinė hipotezė: } \begin{cases} H_0 : \beta_k = 0, \\ H_1 : \beta_k \neq 0. \end{cases}$$

Voldo statistika: $W = (\hat{\beta}_k / SE_k)^2$; čia SE_k yra β_k standartinio nuokrypio įvertis.

Voldo statistika turi χ^2 skirstinį su 1 laisvės laipsniu. Hipotezė H_0 atmetama ($\beta_k \neq 0$), jeigu p-reikšmė $p < \alpha$ (α - reikšmingumo lygmuo). Priešingu atveju hipotezė H_0 neatmetama [10].

Tiesinės regresijos analizėje, įvertinant nepriklausomų kintamųjų ir Y_i priklausomybę, skaičiuojamas determinacijos koeficientas r^2 . Logistinėje regresijoje yra keletas determinacijos koeficiento analogų. Pseudodeterminacijos koeficientai parodo, kurią dalį sklaidos paaiškina logistinės regresijos modelis [10].

Makfadeno (McFadden) pseudodeterminacijos koeficientas r_M^2 [10]:

$$r_M^2 = 1 - \frac{\ln L(\hat{\beta})}{\ln L(0)}. \quad (1.11)$$

Koeficientas r_M^2 įgyja reikšmes iš intervalo [0; 1]. Kokso – Snelo (Cox – Snell) pseudodeterminacijos koeficientas r_{CS}^2 [10]:

$$r_{CS}^2 = 1 - \left(\frac{L(0)}{L(\hat{\beta})} \right)^{2/n}. \quad (1.12)$$

čia n – stebėjimų skaičius ir r_{CS}^2 koeficientas mažesnis už vienatą [10].

Koeficientas r_{CS}^2 įgyja reikšmes iš intervalo [0; 1].

Taip pat naudojamas normuotasis Kokso-Snelo koeficientas (Nagelkerkės pseudo-determinacijos koeficientas) r_N^2 [10]:

$$r_N^2 = 1 - \frac{r_{CS}^2}{\max r_{CS}^2} = \frac{r_{CS}^2}{1 - (L(0))^{2/n}}. \quad (1.13)$$

Koeficientas r_N^2 įgyja reikšmes iš intervalo $[0; 1]$. Kuo labiau logistinės regresijos modelis atitinka analizuojamus duomenis, tuo labiau pseudodeterminacijos koeficientai artėja prie 1.

1.2.2 Išlikimo teorija ir Kokso regresija

Išlikimo teorija – statistiniai metodai, nagrinėjantys laiko intervalo trukmę, kurios pabaigoje įvyksta stebimas įvykis. Medicininuose tyrimuose šis įvykis dažniausiai yra apibrėžiamas kaip stebimos grupės individo mirtis. Išlikimo laiko analizės taikymo sritis – gyvenimo trukmės skirstinio modeliavimas [3-7]. Klinikiniuose tyrimuose tai yra itin naudingas įrankis, leidžiantis analizuoti nuolat papildomus tyrimo duomenis, t.y. asmenys įtraukiami į tyrimą vėliau arba pasitraukia jam dar nepasibaigus. Vienas iš svarbiausių išlikimo laiko modeliavimo uždavinių yra priklausomybės tarp išlikimo laiko ir nepriklausomų kintamųjų (kovariančių) įvertinimas. Šios užduoties sprendimui yra taikomas Kokso regresinės analizės metodas [3-7].

Tarkime, kad tiriamo individo gyvenimo trukmė T yra teigiamas atsitiktinis dydis su pasiskirstymo funkcija $F_T(t) = P(T \leq t)$. $F_T(t)$ – tikimybė mirti iki momento t [3-7].

Apibrėžkime išlikimo (patikimumo) funkciją (t.y. tikimybę, kad iki fiksuoto laiko momento t neįvyks stebimas įvykis) [3-7]:

$$S_T(t) = P(T > t) = 1 - F_T(t). \quad (1.14)$$

Kadangi $S_T(t)$ yra tikimybė, tai jos įgyjamos reikšmės yra nuo 0 iki 1 ir $S_T(0) = 1$ bei $\lim_{t \rightarrow \infty} S_T(t) = 0$ [7].

Tarkime, kad atsitiktinio dydžio T skirstinys yra absoliučiai tolydus. Tuomet galima apibrėžti tankio funkciją – tikimybę mirti tuoj pat po momento t [4-7]:

$$f_T(t) = \lim_{h \rightarrow 0} \frac{F_T(t+h) - F_T(t)}{h} = \lim_{h \rightarrow 0} \frac{P(t < T \leq t+h)}{h}. \quad (1.15)$$

Daug dažniau naudojama reikšmingesnė mirtingumo intensyvumo funkcija [4-7]:

$$\lambda_T(t) = \lim_{h \rightarrow 0} \frac{P(t < T \leq t+h | T > t)}{h}. \quad (1.16)$$

Funkcija $\lambda_T(t)$ – tikimybė, kad laiko momentu t stebimas individas dar nemiręs. Mirtingumo intensyvumo funkciją galima užrašyti panaudojus $S_T(t)$ ir $f_T(t)$ išraiškas [4-7]:

$$\lambda_T(t) = \lim_{h \rightarrow 0} \frac{F_T(t+h) - F_T(t)}{S_T(t) \cdot h} = \frac{f_T(t)}{S_T(t)}. \quad (1.17)$$

Skirtingai nuo tiesinės regresijos modelių, gyvenimo trukmės skirstiniai nėra normalieji. Išlikimo duomenys dažniausiai būna cenzūruoti, t.y. kiekvienas individo mirties momentas fiksuojamas tik tai tuo atveju, jei jis įvyksta [4]:

- ne vėliau (gal būt atsitiktinio) cenzūravimo momento C (cenzūravimas iš dešinės);
- ne anksčiau cenzūravimo momento C (cenzūravimas iš kairės);
- duotame laiko intervale (intervalinis cenzūravimas).

Priešingu atveju pats mirties momentas nežinomas, bet žinoma informacija, kad tai įvyko atitinkamai po, prieš cenzūravimo momentą C arba už žinomo intervalo ribų [4].

Cenzūravimo iš dešinės atveju žinomas mirties momentas T arba cenzūravimo momentas C . Jei T nežinomas, tai žinoma, kad $T > C$. Cenzūravimo iš kairės atveju žinomas mirties momentas T arba cenzūravimo momentas C . Jei T nežinomas, tai žinoma, kad $T < C$. Intervalinio cenzūravimo atveju žinomas mirties momentas T arba T yra nežinomas, bet žinoma, kad T nepriklauso žinomam intervalui [4].

Išlikimo teorijoje dažniausiai naudojami semiparametriniai laiko skirstinio vertinimo metodai, kai nekonkretizuojama skirstinių šeima, aprašanti gyvenimo trukmės skirstinį. Todėl paprastai vertinama kokia nors bazinė funkcija (begalinio matavimo parametras) ir regresijos lygties parametrai (jų skaičius baigtinis). Kokso regresija yra išlikimo laiko vertinimo metodas, kur modeliuojama mirtingumo intensyvumo funkcija. Šis metodas nereikalauja apibrėžti laiko trukmės skirstinį [4-7].

Kokso regresijos modelis užrašomas lygtimi

$$\lambda(t | X) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m), \quad (1.18)$$

čia β_1, \dots, β_m – regresijos lygties parametrai, $X = (X_1, \dots, X_m)$ – regresijos lygties kintamieji (kovariantės); $\lambda_0(t)$ – bazinė mirtingumo intensyvumo funkcija [6, 7].

Apžvelgsime kaip interpretuojamos Kokso regresijos lygties parametru $\beta_k, k = \overline{1, m}$ reikšmės. Parametru interpretacija analogiška kaip ir logistinės regresijos atveju, kai skaičiuojamas galimybių santykis. Kokso regresijoje skaičiuojamas santykinės rizikos koeficientas (angl. relative risk) ρ [7].

Intervalinių kintamųjų atveju (kai X_k reikšmė kinta ir likusieji kintamieji $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_m$ yra fiksuoti) santykinės rizikos koeficientas lygus

$$\rho = \frac{\lambda(t | X_k^{(2)})}{\lambda(t | X_k^{(1)})} = \frac{\lambda_0(t) \cdot \exp(\beta_1 X_1 + \dots + \beta_k X_k^{(2)} + \dots + \beta_m X_m)}{\lambda_0(t) \cdot \exp(\beta_1 X_1 + \dots + \beta_k X_k^{(1)} + \dots + \beta_m X_m)} = \exp(\beta_k (X_k^{(2)} - X_k^{(1)})). \quad (1.19)$$

Kategorinių kintamųjų atveju, kai k -tojo pseudokintamojo reikšmė keičiasi iš 1 į 0, likusių pseudokintamųjų reikšmės abiem atvejais lygios 1 santykinės rizikos koeficientas lygus

$$\rho = \frac{\lambda(t | X_k = 1)}{\lambda(t | X_k = 0)} = \frac{\lambda_0(t) \cdot \exp(\beta_1 + \dots + \beta_{k-1} + \beta_k + \beta_{k+1} + \dots + \beta_m)}{\lambda_0(t) \cdot \exp(\beta_1 + \dots + \beta_{k-1} + \beta_{k+1} + \dots + \beta_m)} = \exp(\beta_k). \quad (1.20)$$

Abiem atvejais gautas santykinės rizikos koeficientas parodo kiek kartų pasikeičia mirtingumo intensyvumas keičiantis kintamojo reikšmei [7].

Kokso regresijos lygties parametrų $\beta = (\beta_1, \dots, \beta_m)^T$ įverčiai randami maksimizuojant dalinę tikėtimumo funkciją $L(\beta) = \prod_{i=1}^m \frac{\exp(\beta^T X_i)}{\sum_{s \in R_i} \exp(\beta^T X_s)}$ (čia R_i - objektų esančių rizikos stadijoje aibė), kuri turi visas tikėtimumo funkcijos savybes [2].

Kokso regresijos modelio tinkamumas turimiems duomenims įvertinamas analizuojant paklaidas. Kokso regresijos atveju analizuojamos martingalinės ir anomaliosios paklaidos [7].

Martingalo procesas i -tajam individui apibrėžiamas formule

$$\hat{M}_i(t) = N_i(t) - \hat{E}_i(t), \quad (1.21)$$

čia $N_i(t)$ – skaičius įvykių, kurie įvyko individui i iki laiko momento t , o $\hat{E}_i(t)$ - individo i suminės rizikos iki laiko momento t įvertis. Skirtumas tarp N ir \hat{E} procesų iš tiesų yra skirtumas tarp stebėto įvykių kiekio ir tikėtino įvykių kiekio individui i . Didelis skirtumas leidžia daryti išvadą apie anomalijos arba neįprastos situacijos atvejį. Martingalinė paklaida yra pagrįsta proceso įverčiu tyrimo pabaigoje ir apibrėžiama taip: $M_i = \hat{M}_i(\infty) = N_i(\infty) - \hat{E}_i(\infty)$. Šios paklaidos turi nulinį vidurkį ir tuo pačiu M_i, M_j yra tarpusavyje nekoreliuotos. Individas, kuriam įvykis įvyksta sąlyginai anksti ir kuriam rizika yra maža, turės teigiamą martingalinę paklaidą, kuri gali įgyti maksimalią reikšmę lygi 1 (t.y.: $M_i = N_i - \hat{E}_i = 1 - 0 = 1$). Taigi maksimalias M_i reikšmes įgis netikėtai greitai tiriamo laikotarpio metu įvyki patyrę individai. Individas, kuris išgyvena išskirtinai ilgai ir įgyja maksimalią riziką, turės didelę neigiamą paklaidą (t.y.: $M_i = N_i - \hat{E}_i = 0 - \infty = -\infty$). Taigi martingalinės paklaidos patenka į intervalą: $-\infty < M_i \leq 1$. Didelės arba mažos reikšmės leidžia manyti, kad taikomas modelis yra netinkamas [7].

Kokso regresijos modeliui skaičiuojama ir anomalioji paklaida, kuri užrašoma formule [2]

$$d_i = \text{sign}(\hat{M}_i) \sqrt{2 \left(-\hat{M}_i - N_i(\infty) \ln \left(\frac{N_i(\infty) - \hat{M}_i}{N_i(\infty)} \right) \right)}. \quad (1.22)$$

Hipotezės apie parametrų β_1, \dots, β_m lygybę nuliui Kokso regresijos atveju tikrinamos analogiškai kaip ir logistinės regresijos modelyje (taikomi χ^2 suderinamumo ir Voldo kriterijai) [6].

1.2.3 Puasono regresija

Ankstesniuose skyriuose pagrindiniu kiekvieno stebėjimo vienetu buvo laikomas individualus asmuo. Kartais reikia analizuoti įvykių skaičių per laiko vienetą. Toks analizės būdas gali būti naudojamas tada, kai turime taip surinktus duomenis arba kai išgyvenamumo analizės metodu taikymas yra per daug sudėtingas. Didelių duomenų rinkinių, kurie turi daug nuo laiko priklausančių kovariančių, analizė taikant Kokso regresijos modelį reikalauja daug skaičiavimo resursų. Tokiu atveju turimų duomenų transformacija į įvykių stebimiems pacientams per laiko vienetą ženkliai palengvina duomenų analizės uždavinio sprendimą. Jeigu įvykių dažnis per laiko vienetą yra mažas, tada uždavinio sprendimui tinkamiausias Puasono regresijos modelis [3].

Tarkime stebimo įvykio dažnis I yra tikėtinas įvykių skaičius 100000 pacientų per metus. Tarkime, kad įvyko d nepriklausomų įvykių, kai stebėtų pacientų per metus skaičius yra n . Čia d yra labai mažas skaičius lyginant su n . Tokiu atveju įvykio dažnio I įvertis yra [3]

$$\hat{I} = 100000 \cdot \frac{d}{n}. \quad (1.23)$$

Įvykio dažnis taip pat gali būti išreikštas kaip tikėtinas įvykių skaičius 1000 stebėtų pacientų per mėnesį ir panašiai [3].

Tarkime turime kokio nors stebimo įvykio dažnumus I_0 ir I_1 (t.y.: $I_i = 100000 \cdot d_i / n_i$, kai $i = 0, 1$), atitinkamai dviems pacientų grupėms 0 ir 1. Tuomet įvykio santykinis rizikos koeficientas grupėje 1 lyginant su grupe 0 yra

$$\rho = I_1 / I_0. \quad (1.24)$$

Santykinės rizikos koeficiento ρ įvertis $\hat{\rho} = \hat{I}_1 / \hat{I}_0$. Įverčio $\hat{\rho}$ logaritmas $\ln \hat{\rho}$ yra apytiksliai pasiskirstęs pagal normalųjį skirstinį, kurio dispersijos įvertis $\hat{s}_{\ln(\hat{\rho})}^2 = \frac{1}{d_0} + \frac{1}{d_1}$ [3].

Tuomet santykinio rizikos koeficiento ρ 95% pasikliautinis intervalas [3]

$$PI_{0.95}(\rho) = (\hat{\rho} \cdot \exp(-1.96 \cdot \hat{s}_{\ln(\hat{\rho})}); \hat{\rho} \cdot \exp(1.96 \cdot \hat{s}_{\ln(\hat{\rho})})). \quad (1.25)$$

Tarkime, kad d yra stebėtas įvykių skaičius n pacientų imtyje. Pažymėkime π įvykio tikimybę. Tuomet atsitiktinis dydis, žymintis įvykių skaičių tarp stebimų pacientų, yra pasiskirstęs pagal binominį skirstinį su parametrais n ir π [3]

$$P(D = d) = \frac{n!}{(n-d)!d!} \pi^d (1-\pi)^{n-d}. \quad (1.26)$$

Akivaizdu, kad atsitiktinis dydis D įgyja sveikąsias reikšmes nuo 0 iki n , vidurkis $E(D) = n\pi$ ir dispersija $Var(D) = n\pi(1-\pi)$. Tuo atveju, kai n yra pakankamai didelis, o π įgyja mažas

reikšmes, tai tada binominį skirstinį galima aproksimuoti Puasono skirstiniu. Taigi, jei $n\pi \rightarrow \lambda$, o $n \rightarrow \infty$, tai (1.26) galima aproksimuoti Puasono skirstiniu:

$$P(D = d) = \frac{e^{-\lambda} \lambda^d}{d!}, \quad (1.27)$$

čia $E(D) = Var(D) = \lambda$ [3].

Sudarome Puasono regresijos modelį:

$$E(D) = \lambda = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n), \quad (1.28)$$

čia $\beta_0, \beta_1, \dots, \beta_n$ – regresijos modelio parametrai, X_1, \dots, X_n – regresijos modelio kintamieji, n – kintamųjų skaičius [3, 7]. Tarkime, kad vidurkis $E(D) = \lambda$ yra prognozuojamas tikėtinas tiriamo įvykio dažnis $\hat{\pi}$. Tada gauname Puasono regresijos lygtį

$$\hat{\pi} = \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_n X_n),$$

o prognozuojamo dažnio logaritmas

$$\ln(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_n X_n. \quad (1.29)$$

Kadangi įvykio dažnio įvertis $\hat{\pi} = \hat{D}/N$, čia \hat{D} – tikėtinų įvykių skaičius, N – populiacijos didumas, tai [3]

$$\ln(\hat{D}) = \ln(N) + \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_n X_n. \quad (1.30)$$

Tikėtino įvykių skaičiaus taškinis įvertis

$$\hat{D} = N \cdot \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_n X_n). \quad (1.31)$$

Dydis $\ln(N)$ (1.30 formulė) yra vadinamas poslinkiu, be to akivaizdu, kad prognozuojamas įvykių skaičius \hat{D} yra proporcingas dydžiui N (1.31 formulė) [7].

Mirtingumo duomenų analizėje Puasono regresijos modeliuose dažniausiai naudojami kategoriniai kintamieji. Tuomet regresijos lygties kiekvienam kategoriniam kintamajam, įgyjančiam J skirtingų reikšmių, sudaromi $j = \overline{1, J-1}$ dvireikšmiai pseudokintamieji, o paskutinė ($j = J$) kintamojo įgyjama reikšmė išreiškiama per regresijos lygties konstantą β_0 . Tarkime, turime du kategorinius regresijos kintamuosius X_1, X_2 , kur X_1 yra dvireikšmis (koduojamas vienu dvireikšmiu pseudokintamuoju $X_1^{(1)}$), o X_2 turi J skirtingų reikšmių (koduojamas $J-1$ dvireikšmiais pseudokintamaisiais $X_2^{(j)}$, $j = \overline{1, J-1}$) [3]. Tuomet Puasono regresijos modelis užrašomas lygtimi

$$\begin{aligned} \ln(\hat{D}) &= \ln(N) + \hat{\beta}_0 + \hat{\beta}_1 X_1^{(1)} + \hat{\beta}_2^{(1)} X_2^{(1)} + \hat{\beta}_2^{(2)} X_2^{(2)} \dots + \hat{\beta}_2^{(J-1)} X_2^{(J-1)} = \\ &= \ln(N) + \hat{\beta}_0 + \hat{\beta}_1 X_1^{(1)} + \sum_{j=1}^{J-1} \hat{\beta}_2^{(j)} X_2^{(j)}, \end{aligned}$$

čia pseudokintamieji $X_1^{(1)}$ ir $X_2^{(j)}$, $j = \overline{1, J-1}$ įgyja reikšmes 0 arba 1.

Puasono regresijos lygties parametrų β_k , $k = \overline{1, n}$ interpretacija analogiška logistinės bei Kokso regresijos modelių atvejui. Puasono regresijoje skaičiuojamas santykinės rizikos koeficientas (angl. relative risk) ρ [3, 7]. Medicinoje šis santykinės rizikos koeficientas dažnai vadinamas standartizuotu mirtingumo koeficientu (angl. standardized mortality rate) [7]. Santykinės rizikos koeficientas leidžia statistiškai įvertinti skirtumą tarp gautų tikėtinų įvykio dažnių (π_1 ir π_2) dviems populiacijos grupėms, kurios skiriasi tik pagal vieną iš kintamųjų [7].

Kategorinių kintamųjų atveju (naudojami pseudokintamieji) santykinės rizikos koeficientas [3, 7]:

$$\rho = \frac{\pi_2}{\pi_1} = \frac{\exp(\beta_1 + \dots + \beta_{k-1} + \beta_k^{(2)} + \beta_{k+1} + \dots + \beta_n)}{\exp(\beta_1 + \dots + \beta_{k-1} + \beta_k^{(1)} + \beta_{k+1} + \dots + \beta_n)} = \exp(\beta_k^{(2)} - \beta_k^{(1)}). \quad (1.32)$$

Puasono regresijos modelio tinkamumas turimiems duomenims įvertinamas analizuojant pakaidas. Puasono regresijos atveju analizuojamos anomaliosios [2, 3], Pirsono [2], Frymano-Tjukio [19] paklaidos. Paklaidos, kurių reikšmės nepatenka į intervalą $[-2; 2]$ yra išskirtys.

Anomalioji (angl. deviance) paklaida užrašoma formule

$$d_i = \text{sign}(y_i - \hat{y}_i) \cdot \sqrt{2y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - 2(y_i - \hat{y}_i)}, \quad (1.33)$$

čia i yra populiacijos dalies, kuriai stebimos regresijos nepriklausomų kintamųjų reikšmės yra vienodos, indeksas; y_i – stebėtas įvykių skaičius; \hat{y}_i – prognozuojamas tikėtinas įvykių skaičius [2, 3]. Puasono regresijos modelio nuokrypio statistika pasiskirsčiusi pagal chi-kvadrato skirstinį su $\Lambda - \Omega$ laisvės laipsniais (Λ – stebėjimų skaičius, Ω – regresijos lygties parametrų skaičius). Modelis idealiai tinka analizuojamiems duomenims, jeigu nuokrypio statistikos ir laisvės laipsnių santykis lygus vienetui [8].

Pirsono paklaida –

$$e_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}. \quad (1.34)$$

Frymano-Tjukio paklaida –

$$f_i = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{y}_i + 1}. \quad (1.35)$$

Frymano-Tjukio paklaida yra geresnis skirtumo tarp įvykio stebėtų ir tikėtinų dažnių matas nei sistemos SAS skaičiuojamos standartinės anomaliosios ir Pirsono paklaidos ypač tais atvejais kai analizuojami reti įvykiai, t.y. $\hat{y}_i \rightarrow 0$ [19].

1.2.4 Amžiaus-periodo-kohortos modeliai

Dažnio trendo įvertinimas nagrinėjant mirtingumą nuo tam tikros ligos yra ilgametis statistinės analizės uždavinys medicinos tyrimuose. Šio trendo analizė yra reikšminga, nes leidžia daryti prielaidas apie būsimas tendencijas ateityje bei įvertinti ir suprasti tiriamos ligos priežastis [17]. Pirmieji bandymai atlikti detalią dažnio trendo analizę apsiribojo grafiniais dažnių, apskaičiuotų amžiaus grupėms, vaizdavimais. Nors grafinis vaizdavimas yra svarbi trendo analizės dalis, tačiau regresinės analizės metodų naudojimas yra efektyvesnis laiko įtakos mirtingumui problemos sprendimo būdas [17].

Amžius yra viena svarbiausia daugelio ligų priežastis. Jei stebime grupę individų nuo jų gimimo, tai rizika susirgti tam tikra liga gali priklausyti nuo gimimo metų laikotarpio (kohortos) kaitos. Tokiu atveju skirtingoms kohortoms priklausantys individai gali priklausyti skirtingoms rizikos faktoriaus kategorijoms. Taigi labai tikėtina, kad tarp individų, priklausančių skirtingoms kohortoms, susirgimų nuo tam tikros ligos dažniai bus skirtingi, t.y. bus stebimas kohortos efektas (jį sudaro visi faktoriai, įtakojantys mirtingumo dažnį ir tiesiogiai susiję su gimimo metų laikotarpiu, t.y. su konkrečia tiriamu asmenų karta). Pavyzdžiui, rūkymas – įprotis, kuris anksčiau buvo pastebimas tik vėlyvoje paauglystėje ar net trečiame dešimtmetyje, tačiau laikui bėgant rūkančiųjų amžius jaunėja. Taigi, galima prielaida, kad tarp ligų, kurias įtakoja rūkymas, dažnumų ir kohortos bus stebima tam tikra poslinkio tendencija [17].

Sukauptoje demografinėje statistikoje duomenys apie mirtingumą dažniausiai pateikiami stebėjimai, kuriuose pateikiamas ligos diagnozės laikotarpis ir asmens amžius. Tokiu atveju vienas iš būdų įvertinti dažnio tendą yra analizuoti ligos dažnumų tam tikrais laikotarpiais tendencijas, t.y. įvertinti periodo efektą [17]. Atskirais atvejais gali būti stebimas periodo efektas, kuris įtakos didesnę riziką visiems stebimiems tame laikotarpyje individams, nepaisant amžiaus. Pavyzdžiui, padidėjęs oro ar vandens užterštumas vienam periode gali įtakoti padidėjusią riziką tam tikrų ligų atveju visoje populiacijoje. Medicinos technologijų pokyčiai taip pat gali būti įvardijami kaip periodo efektas, mažinantis mirtingumo dažnius [17].

Apibendrinant galima apibrėžti tris pagrindinius kintamuosius, kurie naudojami mirtingumo duomenų analizėje [11-17]:

- amžius – asmens amžius diagnozės metu (A);
- periodas – metai, kada diagnozuojama liga (P);
- kohorta – asmens gimimo metai (K).

Dažniausiai individualūs tyrimo duomenys būna pateikti lentelėje ir suskirstyti pagal amžiaus grupes ir periodus 5 metų intervalais (1.1 lentelė).

1.1 lentelė

Duomenų pavyzdys

| | | Periodas | | | | |
|--------|-------|-----------|-----------|-----------|-----------|-----------|
| | | 1980-1984 | 1985-1989 | 1990-1994 | 1995-1999 | 2000-2004 |
| Amžius | 25-29 | 1 | 1 | 1 | 0 | 1 |
| | 30-34 | 0 | 2 | 1 | 2 | 1 |
| | 35-39 | 3 | 3 | 3 | 5 | 6 |
| | 40-44 | 5 | 6 | 13 | 20 | 36 |
| | 45-49 | 17 | 10 | 21 | 36 | 61 |
| | 50-54 | 10 | 24 | 20 | 34 | 58 |
| | 55-59 | 18 | 22 | 28 | 38 | 72 |

Kiekviename lentelės langelyje pateiktas įvykių skaičius atitinkamoje amžiaus grupėje ir periode. Kadangi vertinami mirtingumo dažniai, analizės metu naudojami ir populiacijos dydžio duomenys – analogiška lentelė, kurios langeliuose pateikta stebimos populiacijos dalis priklausanti atitinkamai amžiaus grupei tam tikram periode. Kohortos duomenys lentelėje išdėstyti įstrižainėje (1 lentelėje skirtingų kohortų duomenys pažymėti skirtingomis spalvomis).

Jei žinomas asmens amžius ir ligos diagnozavimo periodas, tai žinoma ir gimimo metų kohorta, t.y. kintamieji A , P , K yra tiesiškai priklausomi. Formaliai šią tiesinę priklausomybę galima užrašyti išraiška: $K = P - A$ [13, 14, 17]. Turint duomenis, kurie sugrupuoti intervalais iškyla persidengiančių kohortų intervalų problema. Pavyzdžiui asmenys iš 30-34 metų amžiaus grupės 1980-1984 metų periode priklauso 1945-1954 gimimo metų kohortai. Analogiškai gauname, kad asmenys iš 35-39 metų amžiaus grupės 1980-1984 metų periode priklauso 1940-1949 gimimo metų kohortai. Sulyginę abiem atvejais gautus 10 metų intervalus gauname persidengimą [13, 14, 17].

Norint tinkamai sudaryti amžiaus-periodo-kohortos (APK) regresinės analizės modelį, kuriame būtų vertinami visi trys faktoriai, reikia įvertinti:

- tiesinę kintamųjų A , P , K priklausomybę;
- kohortų intervalų persidengimą.

Literatūroje siūloma keletas problemos sprendimo būdų. Vienas iš jų – kiekvienos iš kohortų dalinimas į du 5 metų intervalus [13, 14]. Tokiu būdu iš vienos 10 metų (pavyzdžiui 1945-1954) kohortos gauname du (1945-1949 ir 1950-1954) intervalus (1.2 lentelė).

APK modelių matematinei realizacijai naudojamas Puasono regresijos metodas. Puasono regresijos analizės metu įvertinama amžiaus, periodo ir kohortos įtaka tikėtiniems mirtingumo nuo tiriamos ligos dažniams [11-17]. Formaliai APK modelis užrašomas tikėtino dažnio logaritmo, kaip tiesinės funkcijos nuo amžiaus, periodo ir kohortos kintamųjų, išraiška:

$$\ln(\pi_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k, \quad (1.36)$$

čia μ įvykio dažnio vidurkis, α_i - amžiaus grupės įtaką ($i = 1, \dots, m, m$ - amžiaus grupių skaičius), β_j - periodo įtaką ($j = 1, \dots, n, n$ - periodų skaičius), γ_i - kohortos įtaką ($k = 1, \dots, m + n - 1$) [13, 14]. Sudarant modelį galimos įvairios naudojamų regresijos kintamųjų kombinacijos, be to galima įtraukti tokius kategorinius kintamuosius kaip lytis, gyvenamoji vieta ir pan.

1.2 lentelė

Kohortos dalinimas

| | 1980 | 1981 | 1982 | 1983 | 1984 |
|----|--|------|------|------|------|
| 35 | <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 40%; text-align: center;">vyresnieji</div> <div style="width: 20%; border-left: 1px solid black; border-right: 1px solid black; height: 100%;"></div> <div style="width: 40%; text-align: center;">jaunesnieji</div> </div> | | | | |
| 36 | | | | | |
| 37 | | | | | |
| 38 | | | | | |
| 39 | | | | | |

Viena iš siūlomų APK modelių modifikacijų: APK modelis su dreifu Δ [11, 12, 15]. Vienas iš analizuojamų regresijos kintamųjų (periodas arba kohorta) yra pakeičiamas tiesiniu trendu (dreifu). Nustatomas pradinis periodo (kohortos) atskaitos taškas P_0 (K_0) ir pereinant iš vieno periodo (kohortos) į sekantį gretimą periodą (kohortą), π_k padidėja dydžiu $\exp(\Delta)$,

$$\pi_{ij} = \exp(\mu + \alpha_i + \Delta(P_j - P_0)), \quad (1.37)$$

t.y. daroma prielaida, kad dažnis laikui bėgant kinta pastoviu dydžiu [11, 12, 15].

1.3 Programinių priemonių pasirinkimas

Regresinės duomenų analizės modelių kūrimui pasirinkta SAS programinė įranga. Sistema SAS yra statistikos paketas, kurio tiesioginė paskirtis – duomenų analizės uždavinių sprendimas. Orientavimasis vien į statistinius bei duomenų analizės uždavinius yra pagrindinis SAS skirtumas nuo kitų matematinių programinės įrangos paketų. Regresiniai duomenų analizės modeliai gali būti kuriami panaudojus SAS posistemes SAS/STAT, SAS/GRAPH, SAS/IML bei standartines SAS duomenų analizės procedūras ir SAS makro programavimo kalbos galimybes [2].

Logistinė regresija. SAS sistemoje logistinės regresijos modeliavimą galima atlikti su keliomis procedūromis:

- procedūra LOGISTIC;
- procedūra CATMOD;
- procedūra GENMOD.

Procedūra LOGISTIC yra skirta logistinės regresijos taikymui ir pateikia visą informaciją apie modelio tinkamumą tyrimo duomenims, parametrų įverčius, jų standartines paklaidas. Prognozuojamas priklausomas kintamasis bei nepriklausomi kintamieji (pagal kuriuos prognozuojama) apibrėžiami

MODEL sakiniu. Kokybiniai kintamieji, kuriems reikia sukurti pseudokintamuosius, apibrėžiami CLASS sakinyje. Procedūra LOGISTIC leidžia parinkti modelio parametrus, išvesti prognozuojamas reikšmes ir įvairias statistikas [1, 2].

Kadangi nepriklausomi kintamieji dažniausiai būna kategoriniai, tai sudarant regresijos modelį juos reikia pertvarkyti (įvedimo metu kategorinio kintamojo reikšmes koduoti skaitinėmis reikšmėmis, pvz., jei kintamasis dvireikšmis, tai koduoti 0 ir 1) arba naudoti pseudokintamuosius. Tarkime prognozuojame tikimybę θ , priklausančią nuo dviejų dvireikšmių kintamųjų X_1 ir X_2 . Pažymėkime

$Logit\theta_{ij} = \log\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right)$; čia $i=0,1$, $j=0,1$. Tuomet logistinės regresijos modelį galima užrašyti tokia

matricine forma: [1]

$$\begin{bmatrix} Logit\theta_{11} \\ Logit\theta_{12} \\ Logit\theta_{21} \\ Logit\theta_{22} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 + \beta_1 \\ \beta_0 + \beta_1 + \beta_2 \\ \beta_0 + \beta_1 + \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

Procedūra CATMOD yra universali procedūra, skirta modeliavimui, kurio metu naudojamos kategorinių kintamųjų funkcijos. Ją galima taikyti logistinės regresijos atvejui, kai visi nepriklausomi kintamieji yra kategoriniai. Procedūra CATMOD yra procedūros LOGISTIC alternatyva tam tikrais logistinės regresijos atvejais. Pagrindinis CATMOD privalumas yra tai, kad kokybiniams kintamiesiems nereikia kurti pseudokintamųjų [1, 2].

Procedūra GENMOD skirta apibendrintų tiesinių modelių analizei. Šią procedūrą galima taikyti ir logistinei regresijai. Logistinės regresijos atvejui procedūra GENMOD gali būti taikoma tada, kai duomenys pateikiami įvykiai / bandymai (events / trials) forma [1, 2].

Kokso regresija. SAS sistemoje išlikimo analizės metodų ir Kokso regresijos modeliavimą galima atlikti su trimis procedūromis:

- procedūra LIFEREG;
- procedūra LIFETEST;
- procedūra PHREG;
- procedūra TPHREG.

Procedūra LIFEREG yra parametrinė procedūra, skirta išgyvenamumo laiko pasiskirstymui kartu su susijusiais kintamaisiais modeliuoti. Ji taiko parametrinius pagreintų gedimų modelius išgyvenamumo duomenims, kurie gali būti cenzūruoti iš kairės, dešinės arba intervaliniu būdu [2].

Procedūra LIFETEST yra neparametrinė procedūra, skirta išlikimo funkcijai įvertinti. Taip pat galima palyginti dviejų ir daugiau duomenų rinkinių išlikimo kreives [2].

Procedūra PHREG skirta Kokso regresijos modeliui taikyti. Duomenys šiuo atveju gali būti cenzūruoti iš dešinės. Prognozuojamas priklausomas kintamasis, cenzūravimo kintamasis bei nepriklausomi kintamieji (pagal kuriuos prognozuojama) nurodomi MODEL sakinyje [2].

Procedūra TPHREG yra patobulinta procedūros PHREG versija. Ji skiriasi nuo PHREG tuo, kad pseudokintamieji kokybiniais kintamiesiems sukuriama automatizuotai, juos apibrėžiant CLASS sakinyje [2].

Puasono regresija. SAS sistemoje Puasono regresijos modeliavimą galima atlikti su GENMOD procedūra, kuri skirta apibendrintų tiesinių modelių analizei. Kadangi GENMOD yra universali procedūra, tai naudojant ją Puasono regresijos modeliavimui reikia parinkti atitinkamus parametrus. Regresijos modelis aprašomas sakiniu MODEL, kuriame nurodomi specialūs Puasono regresijos parametrai [2].

SAS makro programavimo kalba. Dauguma šiuolaikinių programavimo kalbų turi savybę, kuri vadinama išankstiniu programos kodo apdorojimu (angl. preprocessing). Pagrindinė šio proceso veikimo ypatybė – dalies visų programoje naudojamų simbolių srauto paėmimas ir apdorojimas į pageidaujamą simbolių rinkinį pagal iš anksto apibrėžtas taisykles. Šiuo atveju programos kodas vykdymo pradžioje yra tik patikrinamas, ar jis tinkamai suderintas su programavimo kalbos reikalavimais, tačiau neperduodamas programavimo kalbos interpretatoriui. Išankstinio programos kodo simbolių apdorojimo funkcijas atliekantis komponentas SAS programavimo sistemoje yra makro kalbos interpretatorius [2].

Naudojant sukurtas makro komandas galima išplėsti SAS sistemos galimybes. Programuojant makro komandas galima sukurti duomenų aibių apdorojimo programines priemones bei efektyviai optimizuoti SAS sistemos programavimo galimybes. Makro programos kodo rašymas yra sudėtingesnis nei standartinio SAS programavimo kalbos kodo, todėl makro komandų naudojimas paprastoms ar vienkartinėms programoms nėra itin tikslingas. Kita vertus makro komandos spartina ir supaprastina programos aprašymą, todėl tikslingas jų vartojimas efektyviai palengvina atliekamus duomenų analizės uždavinių sprendimus.

Interaktyvioji matricinė programavimo kalba (angl. Interactive Matrix Language - IML). Sistemos SAS posistemė IML (interaktyvioji matricinė programavimo kalba) yra efektyvus programavimo įrankis. Pagrindinis šios programavimo kalbos objektas – duomenų matrica: dvimatis masyvas, kurio elementai yra arba skaičiai, arba simbolinės reikšmės. Programavimas IML kalba yra dinaminis – atminties išskyrimas ir naudojamų matricių dimensijų nustatymas yra automatizuotas [2].

SAS IML posistemėje galima naudoti standartinės arba paties vartotojo sukurtas programavimo priemones. Operatorius bei procedūras galima taikyti ir matricai (skaičiuoti atvirkštine matricę, rasti tikrines matricos reikšmes ir pan.), ir atskiroms matricos reikšmėms. Be to IML posistemėje yra platus duomenų valdymo komandų pasirinkimas. Vartotojas gali nuskaityti (perkelti į matricę), sukurti naują

ar papildyti jau esamą SAS duomenų failą tiesiogiai programuodamas IML posistemėje, taip aplenkdamas standartinį SAS duomenų failo redagavimo žingsnį [2].

1.4 Darbe sprendžiami uždaviniai

Darbas atliktas bendradarbiaujant su Kauno medicinos universiteto Kardiologijos institutu.
Pagrindinis darbo tikslas:

- 1) susipažinti su statistikos metodais, taikomais mirtingumo duomenų analizei Lietuvoje ir užsienyje;
- 2) sudaryti statistinės analizės modelius turimiems mirtingumo duomenims;
- 3) realizuoti sudarytus modelius programiškai, panaudojant duomenų analizės sistemą SAS, SAS posistemės IML ir makro programavimo galimybes;
- 4) panaudojus sukurtas programines priemones atlikti Kauno medicinos universiteto Kardiologijos instituto surinktų mirtingumo nuo galvos smegenų insulto duomenų analizę;
- 5) įvertinti taikomų metodų tinkamumą analizuojamiems duomenims ir parengti sukurtų priemonių taikymo rekomendacijas.

2 Tiriamoji dalis ir rezultatai

2.1 1980-2004 m. Kauno miesto gyventojų mirtingumo duomenys

Šio darbo tikslas – sukurti modelius ir programines priemones Kauno miesto gyventojų 1980-2004 metų mirtingumo duomenų, kuriuos surinko Kauno medicinos universiteto Kardiologijos instituto darbuotojai dviejų tyrimų metu, analizei. 1980-2004 metais atlikto tyrimo metu stebėti individualūs mirčių nuo išorinių priežasčių, vėžinių susirgimų, galvos smegenų insulto, širdies kraujagyslių ligų ir kitų ligų atvejai (viso – 38822 įrašai) tarp Kauno miesto 25-64 m. amžiaus vyrų ir moterų. Stebėti kintamieji pateikti 2.1 lentelėje.

2.1 lentelė

I tyrimo kintamieji

| Kintamojo žymėjimas | Kintamasis | Kintamojo vardas SAS duomenų faile |
|---------------------|---------------------------------|------------------------------------|
| <i>L</i> | Lytis ¹ | LYTIS |
| <i>A</i> | Amžius | AMZIUS |
| <i>M</i> | Mirties metai | METAI |
| <i>Mp</i> | Mirties priežastis ² | KOD_GR |

Pastabos:

¹ kintamojo „Lytis“ galimos reikšmės: 1 – vyras, 2 – moteris;

² kintamojo „Mirties priežastis“ galimos reikšmės: 2 – vėžiniai susirgimai, 3 – išorinės priežastys, 4 – širdies kraujagyslių ligos, 5 – galvos smegenų insultai, 0 – kita;

Darbe naudojami ir gyventojų skaičiaus Kauno mieste 1980-2004 metais skirtingų amžiaus grupių (penkerių metų intervalais) duomenys.

Kito tyrimo metu 1980-2004 metais stebėti Kauno miesto 35-64 metų amžiaus gyventojų (iš viso 2440) 75 klinikinių, sociodemografinių ir gyvenamosios kintamųjų reikšmės (amžius, lytis, ūgis, svoris, kraujo spaudimas, rūkymas ir t.t.), taip pat buvo registruojama ar pacientas tiriamame laikotarpyje išgyveno, ar mirė. Kiekvienu mirties atveju yra įrašyta mirties priežastis: širdies kraujagyslių liga, galvos smegenų insultas, išorinės priežastys ir pan. Įvertinus medicinos ekspertų nuomonę, tolesnei mirtingumo duomenų analizei buvo atrinkti kintamieji, kurie pateikti 2.2 lentelėje.

2.2 lentelė

Atrinkti II tyrimo kintamieji

| Kintamojo žymėjimas | Kintamasis | Kintamojo vardas SAS duomenų faile |
|---------------------|--------------------|------------------------------------|
| X_1 | Amžius | AMZIUS |
| X_2 | Lytis ¹ | LYTIS |
| X_3 | Ūgis (cm) | UG |

2.2 lentelės tęsinys

Atrinkti II tyrimo kintamieji

| | | |
|----------|---|----------|
| X_4 | Svoris (kg) | SV |
| X_5 | Kūno masės indeksas | KMI |
| X_6 | Sistolinis AKS ² (mm Hg) | SISTOL |
| X_7 | Diastolinis AKS ² (mm Hg) | DIASTOL |
| X_8 | Rūkymas ³ | RUKONE |
| X_9 | Fizinis aktyvumas ⁴ | AKTNE |
| X_{10} | Anksčiau patirtas insultas ⁵ | INSULT |
| Y | Gyvas/miręs ⁶ | GYVIARNE |
| S | Stebėjimo laikotarpis ⁷ | METAI |
| Mp | Mirties priežastis ⁸ | MIRTPRZ |

Pastabos:

¹ kintamojo „Lytis“ reikšmės: 1 – vyras, 2 – moteris;

² AKS – arterinis kraujo spaudimas;

³ kintamojo „Rūkymas“ reikšmės: 1 – nerūko, 2 – rūko;

⁴ kintamojo „Fizinis aktyvumas“ reikšmės: 1 – aktyvus, 2 – neaktyvus;

⁵ kintamojo „Anksčiau patirtas insultas“ reikšmės: 0 – nepatirtas, 1 – patirtas;

⁶ kintamojo „Gyvas/miręs“ reikšmės: 1 – tyrimo metu pacientas nemirė, 1 – tyrimo metu pacientas mirė;

⁷ kiek laiko vyko paciento stebėjimas;

⁸ kintamojo „Mirties priežastis“ reikšmės: 0-6 – mirties priežasties kodas.

2.2 Mirtingumo prognozavimo modeliai ir jų taikymo rezultatai

Šiame skyriuje pateiksime darbe sudarytus regresinės analizės modelius, sukurtas makro komandas ir jų taikymo pavyzdžius turimiems duomenims.

Visos darbe sukurtos SAS makro komandos turi vienodą sintaksę:

```
%macro Pavadinimas (parametras_1=, parametras_2=, ..., parametras_n=);
    /* makro komandos programos kodas; */
%mend;
```

SAS programoje makro komanda iškviečiama sakiniu:

```
%Pavadinimas(parametras_1=reiksme_1, ... , parametras_n=reiksme_n);
```

2.2.1 Puasono regresijos modeliai

Puasono regresijos modeliuose naudosime šiuos žymėjimus (2.3 lentelė): L_k - lytis, A_k - amžiaus grupė, P_k - mirties ar susirgimo laikotarpis (periodas), K_k - gimimo metų grupė (kohorta),

N_k - populacijos didumas; čia k žymi stebimą populiacijos dalį $k \rightarrow (L_k, A_k, P_k, K_k)$, kuri nusakoma dvireikšmiais pseudokintamaisiais.

Kiekvienoje Puasono regresijos modelio makro komandoje realizuotas duomenų pertvarkymas į regresinei analizei būtiną duomenų struktūrą, t.y. turimi individualūs pacientų stebėjimo duomenys yra grupuojami pagal lytį, amžiaus grupę, periodą bei kohortą ir apskaičiuojamas stebėtas įvykių skaičius kiekvienoje grupėje. Puasono regresijos analizei skirtų duomenų failų formavimo programinės priemonės realizuotos panaudojus SAS interaktyviąją matricinę programavimo kalbą (IML), nes SAS/IML yra platus duomenų valdymo komandų ir procedūrų pasirinkimas.

2.3 lentelė

Puasono regresijos modelių kintamieji

| Kintamojo žymėjimas | Kintamasis | Kintamojo vardas SAS sistemoje |
|---------------------|--------------------------------|--------------------------------|
| L_k | Lytis | LYTIS |
| A_k | Amžiaus grupė | AMZIUS |
| P_k | Mirties laikotarpis (periodas) | PERIODAS |
| K_k | Gimimo metų grupė (kohorta) | KOHORTA |
| I_k | Stebėtas įvykių skaičius | ATVEJAI |
| N_k | populiacijos didumas | VISO |

Kiekvienam kategoriniam Puasono regresijos lygties kintamajam, įgyjančiam J skirtingų reikšmių, sudaromi $j = \overline{1, J-1}$ dvireikšmiai pseudokintamieji, o paskutinė ($j = J$) kintamojo įgyjama reikšmė išreiškiama per regresijos lygties konstantą:

$$L_{1k} = \begin{cases} 1, & \text{jei } k - \text{toji populiacijos dalis žymi moteris,} \\ 0, & \text{jei } k - \text{toji populiacijos dalis žymi vyrus;} \end{cases}$$

$$A_{jk} = \begin{cases} 1, & \text{jei } k - \text{toji populiacijos dalis priklauso } j - \text{tajai amžiaus grupei,} \\ 0, & \text{kitu atveju,} \end{cases} \quad j = \overline{1, J_A};$$

$$P_{jk} = \begin{cases} 1, & \text{jei } k - \text{toji populiacijos dalis priklauso } j - \text{tajam periodui,} \\ 0, & \text{kitu atveju,} \end{cases} \quad j = \overline{1, J_P};$$

$$K_{jk} = \begin{cases} 1, & \text{jei } k - \text{toji populiacijos dalis priklauso } j - \text{tajai kohortai,} \\ 0, & \text{kitu atveju,} \end{cases} \quad j = \overline{1, J_K}.$$

Darbe sudaryti šie Puasono regresijos modeliai: lytis-amžius, lytis-amžius-periodas, lytis-amžius-kohorta, lytis-amžius-periodas-kohorta, lytis-amžius-periodo dreifas, lytis-amžius-kohortos dreifas.

Atliekant duomenų analizę, Puasono regresijos modeliai iškviečiami makro komanda `%Puasono_regresija(periodo_dydis = N, kodo_nr = Mp)`, čia

- **periodo_dydis** – amžiaus, periodo, kohortos grupavimo intervalo ilgis (metais);
- **kodo_nr** – tiriamos ligos kodas.

Makro komandos atliekami veiksmai:

- 1) kiekviena makro komanda realizuoja atitinkamą Puasono regresijos modelį;
 - pvz.: makro komanda `%Puasono_regresija_1` realizuoja modelį lytis-amžius, makro komanda `%Puasono_regresija_2` realizuoja modelį lytis-amžius-periodas ir t.t.;
- 2) įvertinamas Puasono regresijos modelių tinkamumas analizuojamiems duomenims;
- 3) formuojama ataskaita, kuri įrašoma į RTF (Rich Text Format) failą.

2.2.1.1 Puasono regresijos modelis lytis-amžius

Sudarysime Puasono regresijos modelį lytis-amžius, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

- 1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį ir amžiaus grupę).

- 2) Apskaičiuoti tikėtino mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^{J_A} \hat{\gamma}_j \times A_{jk}\right) \text{ ir pasikliautinąjį intervalą } PI_{0.95}(D_k) = (P_a; P_v).$$

- 3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k = \frac{\hat{D}_k}{N_k}$ ir jo pasikliautinąjį intervalą $PI_{0.95}(\pi_k) = (P_a; P_v)$

(1000 gyv.).

- 4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas:

$$d_i = \text{sign}(y_i - \hat{y}_i) \cdot \sqrt{2y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - 2(y_i - \hat{y}_i)}, \quad e_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}, \quad f_i = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{y}_i + 1}, \quad \text{čia } i \text{ yra}$$

populiacijos dalies, kuriai stebimos regresijos nepriklausomų kintamųjų reikšmės yra vienodos, indeksas; y_i – stebėtas įvykių skaičius; \hat{y}_i – prognozuojamas tikėtinas įvykių skaičius.

- 5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.

- 6) Apskaičiuoti santykinės rizikos koeficiento ρ įvertį ir pasikliautinąjį intervalą $PI_{0.95}(\rho) = (P_a; P_v)$ pagal lytį ir skirtingoms amžiaus grupėms.

- 7) Vaizduoti modelio lytis-amžius rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede.

Makro komandos iškvietimo sakinio sintaksė:

```
%Puasono_regresija_1(
    sk = N,
    kodas = Mp,
    kintamieji = LYTIS AMZIUS ),
```

čia

- **sk** – amžiaus grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **kintamieji** – regresijos lygties kintamieji;

Sudarytą Puasono regresijos modelį lytis-amžius taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Tarkime, kad taikymo tikslas yra įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir iširti stebimų veiksnių įtaką.

Puasono regresijos modelis taikomas iškviečiant makro komandą:

```
%Puasono_regresija_1(sk = 5, kodas = 5, kintamieji = LYTIS AMZIUS ).
```

Gauti Puasono regresijos modelio lytis-amžius regresijos lygties parametrų įverčiai ir jų pasikliautiniai intervalai pateikti 2.4 lentelėje.

2.4 lentelė

Modelio lytis-amžius regresijos lygties parametrų įverčiai

| Parametras | | Laisvės laipsniai | Įvertis | Standartinė paklaida | Voldo 95% pasikliautinojo intervalo rėžiai | | X ² statistikos reikšmė | |
|------------|---------|-------------------|----------|----------------------|--|---------|------------------------------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -6.1683 | 0.0441 | -6.2548 | -6.0818 | 19533.1 | <.0001 |
| LYTIS | Moteris | 1 | -0.6650 | 0.0484 | -0.7600 | -0.5700 | 188.40 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -4.8998 | 0.3186 | -5.5242 | -4.2753 | 236.50 | <.0001 |
| AMZIUS | 30-34 | 1 | -3.6426 | 0.1784 | -3.9922 | -3.2930 | 417.06 | <.0001 |
| AMZIUS | 35-39 | 1 | -3.1283 | 0.1453 | -3.4131 | -2.8435 | 463.41 | <.0001 |
| AMZIUS | 40-44 | 1 | -2.2944 | 0.1029 | -2.4962 | -2.0927 | 496.71 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.8304 | 0.0881 | -2.0031 | -1.6578 | 432.01 | <.0001 |
| AMZIUS | 50-54 | 1 | -1.2145 | 0.0719 | -1.3554 | -1.0736 | 285.44 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.5610 | 0.0606 | -0.6798 | -0.4423 | 85.75 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

Puasono regresijos modelio lytis-amžius lygtis:

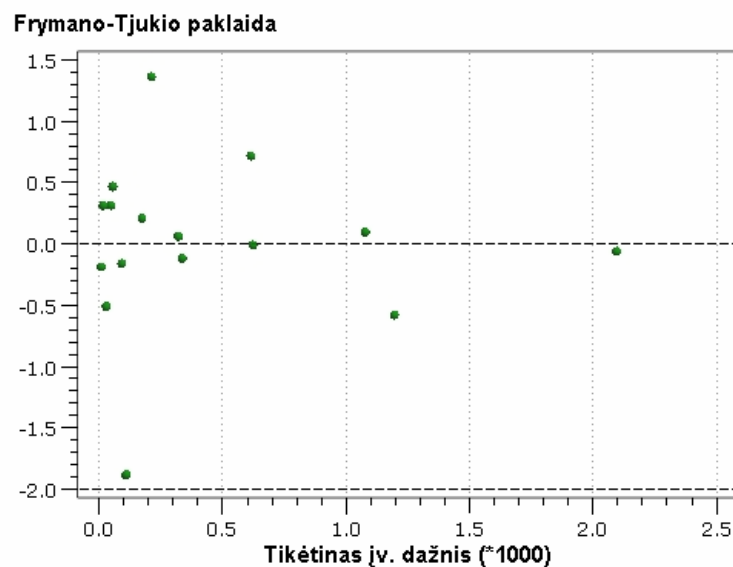
$$\hat{D}_k = N_k \cdot \exp(-6,168 - 0,665 \cdot L_{1k} - 4,9 \cdot A_{1k} - 3,643 \cdot A_{2k} - 3,128 \cdot A_{3k} - 2,294 \cdot A_{4k} - 1,83 \cdot A_{5k} - 1,215 \cdot A_{6k} - 0,561 \cdot A_{7k}).$$

Tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautinieji intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti 2 priedo 1 lentelėje.

Santykinės rizikos koeficientų įverčiai ir jų 95% pasikliautinieji intervalai skirtingoms amžiaus grupėms (lyginant su 25-29 m. amžiaus grupe) pateikti 2 priedo 2 lentelėje. Lyginant skirtingų amžiaus grupių riziką su 25-29 m. amžiaus grupe gauta:

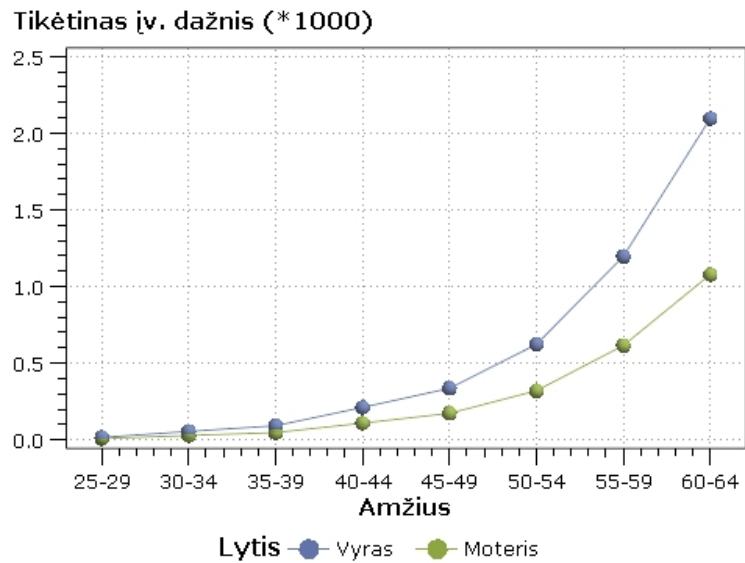
- lyginant su 45-49 m. amžiaus grupe santykinės rizikos koeficiento įvertis $\hat{\rho} = 21,528$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (11,364; 40,781)$, t.y. 45-49 m. amžiaus asmenims mirti nuo GSI yra 21,5 karto labiau tikėtina nei asmenims iš 25-29 m. grupės;
- lyginant su 60-64 m. amžiaus grupe santykinės rizikos koeficiento įvertis $\hat{\rho} = 134,26$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (71,902; 250,697)$, t.y. 60-64 m. amžiaus asmenims mirti nuo GSI yra net 134 kartus labiau tikėtina nei 25-29 m. asmenims.

Analizuojant anomalijas, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (2 priedo 3-5 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.1 paveiksle. Paklaidos patenka į intervalą $[-2; 2]$, todėl modelis tinka analizuojamiems duomenims. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 2 priede.



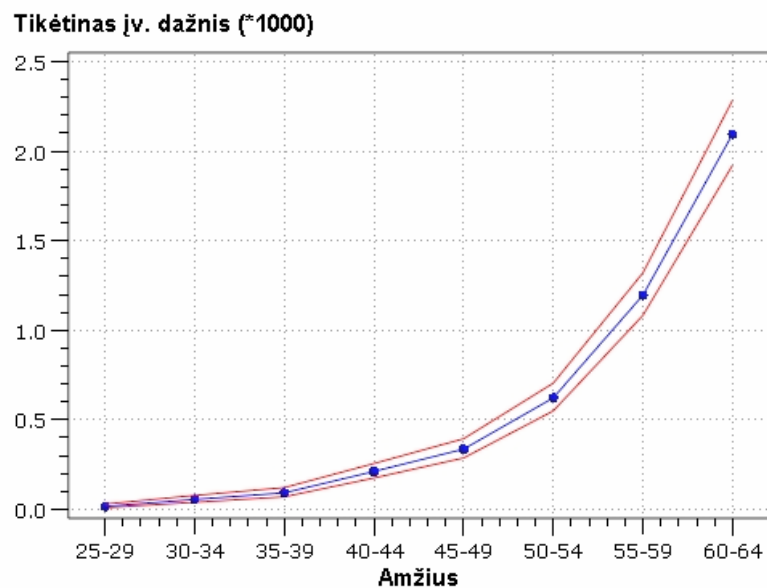
2.1 pav. Modelio lytis-amžius Frymano-Tjukio paklaidų grafikas

2.2 paveiksle pateikta modelio lytis-amžius tikėtinų mirtingumo dažnių priklausomybė nuo amžiaus.



2.2 pav. Modelio lytis-amžius tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo amžiaus

Statistiškai įvertinti skirtumą tarp vyrų ir moterų tikėtinų mirtingumo dažnių leidžia gautas santykinės rizikos koeficiento įvertis $\hat{\rho} = 1,945$, 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (1,768; 2,138)$, t.y. vyrams mirti nuo GSI su 95% garantija yra nuo 1,77 iki 2,14 karto labiau tikėtina nei moterims.



2.3 pav. Modelio lytis-amžius vyrų tikėtinų mirtingumo dažnių (1000 gyv.) ir jų pasikliautinųjų intervalų priklausomybė nuo amžiaus

2.3 paveiksle pateikta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) ir 95% pasikliautinųjų intervalų priklausomybė nuo amžiaus.

2.2.1.2 Puasono regresijos modelis lytis-amžius-periodas

Sudarysime Puasono regresijos modelį lytis-amžius-periodas, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

- 1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį, amžiaus grupę ir periodą).
- 2) Apskaičiuoti tikėtino mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^{J_A} \hat{\gamma}_j \times A_{jk} + \sum_{j=1}^{J_P} \hat{\eta}_j \times P_{jk}\right)$$

ir pasikliautinąjį intervalą $PI_{0.95}(D_k) = (P_a; P_v)$.

- 3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k$ ir jo pasikliautinąjį intervalą $PI_{0.95}(\pi_k) = (P_a; P_v)$ (1000 gyv.).
- 4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas.
- 5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.
- 6) Apskaičiuoti santykinės rizikos koeficiento ρ įvertį ir pasikliautinąjį intervalą $PI_{0.95}(\rho) = (P_a; P_v)$ skirtingiems periodams.
- 7) Vaizduoti modelio lytis-amžius-periodas rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius-periodas realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos iškvietimo sakinio sintaksė:

```
%Puasono_regresija_2(
    sk = N,
    kodas = Mp,
    kintamieji = LYTIS AMZIUS PERIODAS),
```

čia

- **sk** – amžiaus ir periodo grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **kintamieji** – regresijos lygties kintamieji;

Sudarytą Puasono regresijos modelį lytis-amžius-periodas taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir ištirti stebimų veiksnių įtaką.

Puasono regresijos modelis taikomas iškviečiant makro komandą:

```
%Puasono_regresija_2(sk = 5, kodas = 5, kintamieji = LYTIS AMZIUS PERIODAS).
```

Gauti Puasono regresijos modelio lytis-amžius-periodas regresijos lygties parametrų įverčiai ir jų pasikliautinieji intervalai pateikti 2.5 lentelėje.

2.5 lentelė

Modelio lytis-amžius-periodas regresijos lygties parametru įverčiai

| Parametras | | Laisvės laipsniai | Įvertis | Standartinė paklaida | Voldo 95% pasikliautinio intervalo rėžiai | | χ ² statistikos reikšmė | |
|------------|-----------|-------------------|----------|----------------------|---|---------|------------------------------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -6.4488 | 0.0704 | -6.5868 | -6.3108 | 8386.96 | <.0001 |
| LYTIS | Moteris | 1 | -0.6621 | 0.0485 | -0.7571 | -0.5672 | 186.74 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -4.9167 | 0.3186 | -5.5413 | -4.2922 | 238.09 | <.0001 |
| AMZIUS | 30-34 | 1 | -3.6535 | 0.1784 | -4.0032 | -3.3039 | 419.44 | <.0001 |
| AMZIUS | 35-39 | 1 | -3.1264 | 0.1454 | -3.4113 | -2.8416 | 462.64 | <.0001 |
| AMZIUS | 40-44 | 1 | -2.2948 | 0.1030 | -2.4967 | -2.0929 | 496.18 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.8380 | 0.0882 | -2.0108 | -1.6652 | 434.76 | <.0001 |
| AMZIUS | 50-54 | 1 | -1.2209 | 0.0720 | -1.3620 | -1.0798 | 287.65 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.5638 | 0.0606 | -0.6827 | -0.4450 | 86.45 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| PERIODAS | 1980-1984 | 1 | 0.2920 | 0.0838 | 0.1277 | 0.4563 | 12.13 | 0.0005 |
| PERIODAS | 1985-1989 | 1 | 0.3889 | 0.0789 | 0.2341 | 0.5436 | 24.26 | <.0001 |
| PERIODAS | 1990-1994 | 1 | 0.5240 | 0.0757 | 0.3756 | 0.6723 | 47.91 | <.0001 |
| PERIODAS | 1995-1999 | 1 | 0.1218 | 0.0821 | -0.0390 | 0.2827 | 2.20 | 0.1377 |
| PERIODAS | 2000-2004 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

Puasono regresijos modelio lytis-amžius-periodas lygtis:

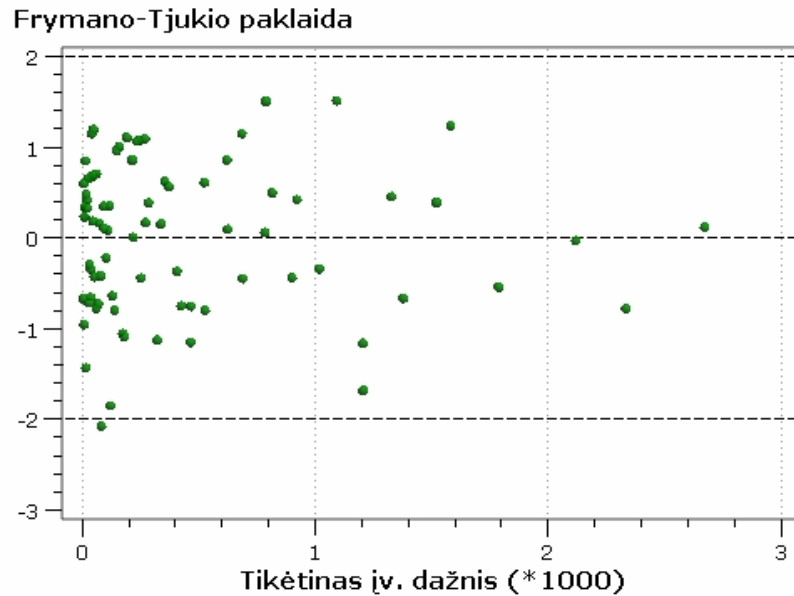
$$\hat{D}_k = N_k \cdot \exp(-6,449 - 0,662 \cdot L_{1k} - 4,917 \cdot A_{1k} - 3,654 \cdot A_{2k} - 3,126 \cdot A_{3k} - 2,295 \cdot A_{4k} - 1,838 \cdot A_{5k} - 1,221 \cdot A_{6k} - 0,564 \cdot A_{7k} + 0,292 \cdot P_{1k} + 0,389 \cdot P_{2k} + 0,524 \cdot P_{3k} + 0,122 \cdot P_{4k}).$$

Tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautinieji intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti 3 priedo 1 lentelėje.

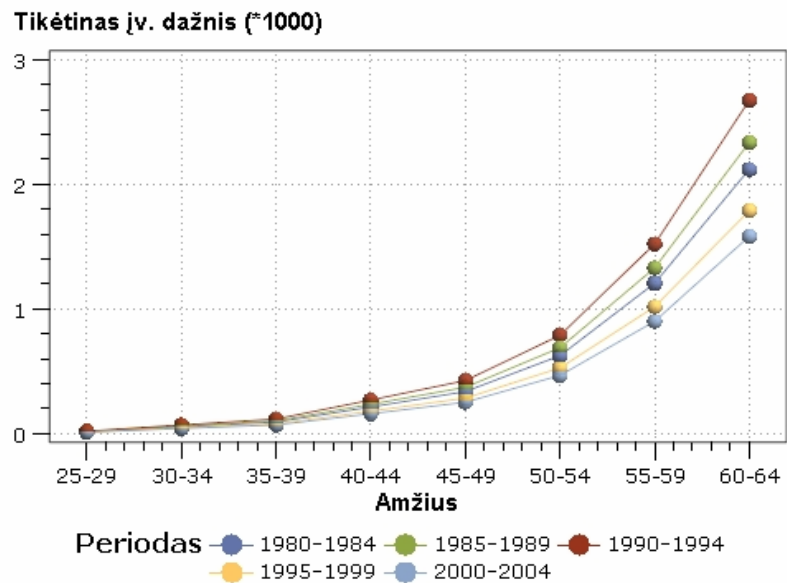
Santykinės rizikos koeficientų įverčiai ir jų 95% pasikliautinieji intervalai skirtingiems periodams (lyginant su 1980-1984m.) pateikti 3 priedo 2 lentelėje. Lyginant skirtingų periodų riziką su 1980-1984 m. gauta:

- lyginant su 1990-1994 m. santykinės rizikos koeficiento įvertis $\hat{\rho} = 1,261$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (1,091; 1,458)$, t.y. 1990-1994 m. mirčių nuo GSI dažnis buvo 1,261 karto didesnis nei 1980-1984 m.;
- lyginant su 2000-2004 m. santykinės rizikos koeficiento įvertis $\hat{\rho} = 0,747$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (0,634; 0,8801)$, t.y. 2000-2004 m. mirčių nuo GSI dažnis buvo 0,747 karto mažesnis nei 1980-1984 m.

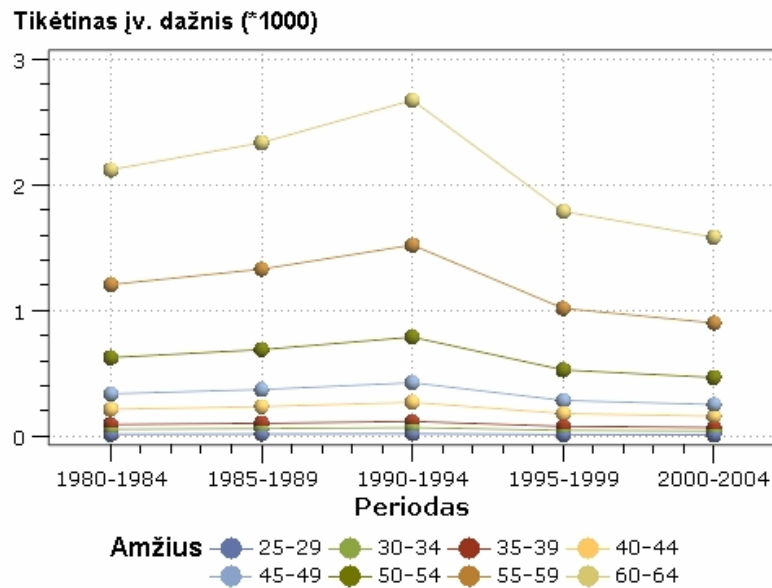
Analizuojant anomaliąsias, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (3 priedo 3-5 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.4 paveiksle. Paklaidos patenka į intervalą $[-2; 2]$, todėl modelis tinka analizuojamiems duomenims. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 3 priede.



2.4 pav. Modelio lytis-amžius-periodas Frymano-Tjukio paklaidų grafikas



2.5 pav. Modelio lytis-amžius-periodas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo amžiaus



2.6 pav. Modelio lytis-amžius-periodas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo

2.5-2.6 paveiksluose pateiktos modelio lytis-amžius-periodas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybės nuo amžiaus ir periodo. Analizuojant mirtingumo dažnius skirtingais 1980-2004 m. periodais išsiskyrė 1990-1994 m. periodas, kuriame stebėtas didžiausias mirtingumas nuo insulto. Statistiškai įvertinti skirtumus tarp skirtingų periodų tikėtinų mirtingumo dažnių leidžia gauti santykinės rizikos koeficientų įverčiai (3 priedo 2 lentelė).

2.2.1.3 Puasono regresijos modelis lytis-amžius-periodo dreifas

Sudarysime Puasono regresijos modelį lytis-amžius-periodo dreifas, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį, amžiaus grupę ir periodą).

2) Apskaičiuoti tikėtiną mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^J \hat{\gamma}_j \times A_{jk} + \Delta \cdot (P_k - P_0)\right) \quad (\text{čia } \Delta - \text{trendo parametras, } P_0 - \text{periodo atskaitos taškas})$$

ir pasikliautinąjį intervalą $PI_{0.95}(D_k) = (P_a; P_v)$.

3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k$ ir jo pasikliautinąjį intervalą $PI_{0.95}(\pi_k) = (P_a; P_v)$ (1000 gyv.).

4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas.

5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.

6) Vaizduoti modelio lytis-amžius-periodo dreifas rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius-periodo dreifas realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos iškvietimo sakinio sintaksė:

```
%Puasono_regresija_2_Drf(
    sk = N,
    kodas = Mp,
    dreifas = N1,
    P_0 = N2
    kintamieji = LYTIS AMZIUS),
```

čia

- **sk** – amžiaus ir periodo grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **dreifas** – trendo parametras;
- **P_0** – periodo atskaitos taškas;
- **kintamieji** – regresijos lygties kintamieji;

Sudarytą Puasono regresijos modelį lytis-amžius-periodo dreifas taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir ištirti stebimų veiksnių įtaką.

Puasono regresijos modelis taikomas iškviečiant makro komandą:

```
%Puasono_regresija_2_Drf(sk = 5, kodas = 5, dreifas = 0.025, P_0 = 1980,
kintamieji = LYTIS AMZIUS).
```

Gauti Puasono regresijos modelio lytis-amžius-periodo dreifas regresijos lygties parametru įverčiai ir jų pasikliautinieji intervalai pateikti 2.6 lentelėje.

Apskaičiuoti tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautinieji intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti lentelėje 4 priedo 1 lentelėje.

2.6 lentelė

Modelio lytis-amžius-periodo dreifas regresijos lygties parametru įverčiai

| Parametras | | Laisvės laipsniai | Įvertis | Standartinė paklaida | Voldo 95% pasikliautinio intervalo rėžiai | | X ² statistikos reikšmė | |
|------------------|---------|-------------------|----------|----------------------|---|---------|------------------------------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -6.4572 | 0.0441 | -6.5436 | -6.3708 | 21443.2 | <.0001 |
| LYTIS | Moteris | 1 | -0.6641 | 0.0484 | -0.7591 | -0.5692 | 187.95 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -4.8663 | 0.3186 | -5.4908 | -4.2419 | 233.28 | <.0001 |
| AMZIUS | 30-34 | 1 | -3.6214 | 0.1784 | -3.9709 | -3.2718 | 412.20 | <.0001 |

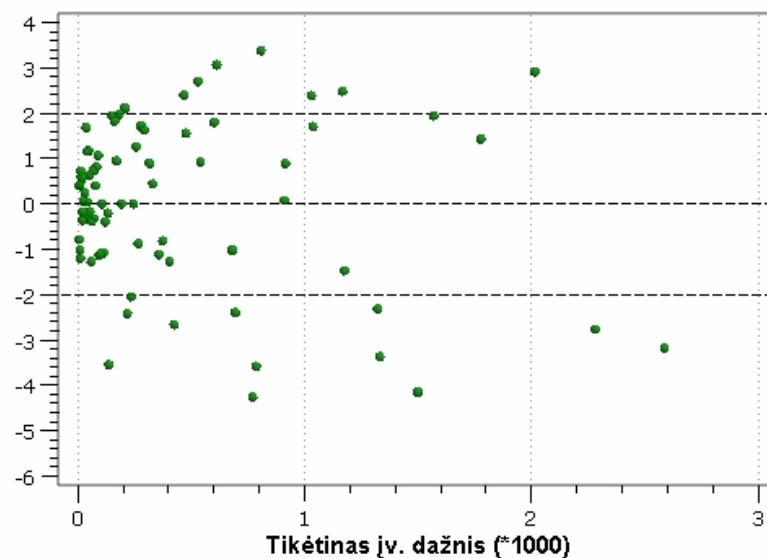
2.6 lentelės tęsinys

Modelio lytis-amžius-periodo dreifas regresijos lygties parametru įverčiai

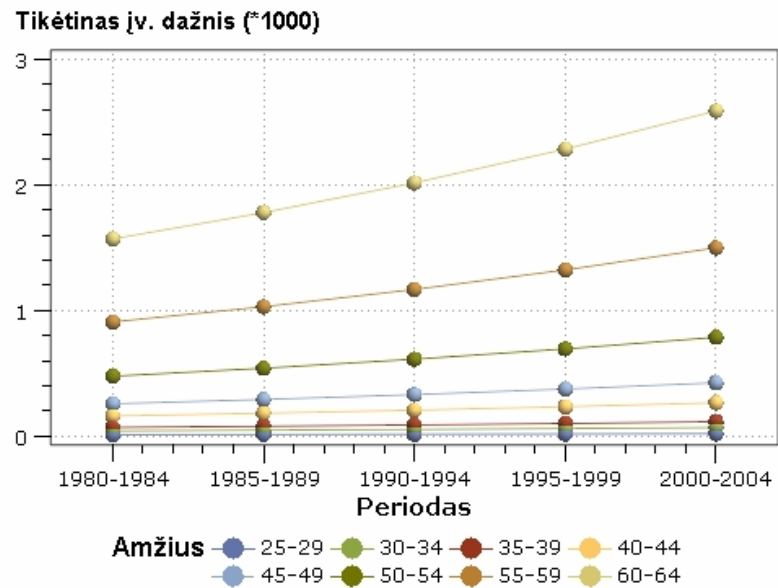
| | | | | | | | | |
|--------|-------|---|---------|--------|---------|---------|--------|--------|
| AMZIUS | 35-39 | 1 | -3.1127 | 0.1453 | -3.3975 | -2.8279 | 458.81 | <.0001 |
| AMZIUS | 40-44 | 1 | -2.2751 | 0.1029 | -2.4769 | -2.0734 | 488.43 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.8061 | 0.0881 | -1.9787 | -1.6335 | 420.62 | <.0001 |
| AMZIUS | 50-54 | 1 | -1.1905 | 0.0719 | -1.3314 | -1.0496 | 274.32 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.5460 | 0.0606 | -0.6648 | -0.4273 | 81.24 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

Analizuojant anomalijas, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (4 priedo 2-4 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.7 paveiksle. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 4 priede. Dalis anomaliųjų, Pirsono ir Frymano-Tjukio paklaidų yra išsibarstę už intervalo $[-2; 2]$ ribų. Be to modelio lytis-amžius-periodo dreifas tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybės nuo periodo grafike (2.8 paveikslas) stebimos didėjančios kreivės, bet stebėtų mirtingumo dažnių (1000 gyv.) grafike (2.9 paveikslas) 1990-1994 m. stebimos maksimalios mirtingumo dažnio reikšmės, todėl modelis lytis-amžius-periodo dreifas analizuojamiems duomenims netinka.

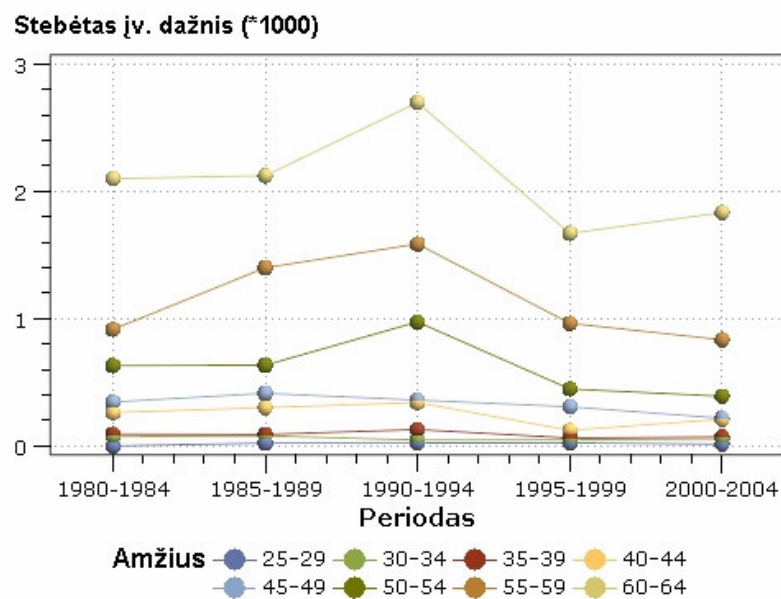
Frymano-Tjukio paklaida



2.7 pav. Modelio lytis-amžius-periodo dreifas Frymano-Tjukio paklaidų grafikas



2.8 pav. Modelio lytis-amžius-periodo dreifas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo



2.9 pav. Modelio lytis-amžius-periodo dreifas vyrų stebėtų mirtingumo dažnių (1000 gyv.) priklausomybė nuo periodo

2.2.1.4 Puasono regresijos modelis lytis-amžius-kohorta

Sudarysime Puasono regresijos modelį lytis-amžius-kohorta, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

- 1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį, amžiaus grupę ir kohortą).

2) Apskaičiuoti tikėtino mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^{J_A} \hat{\gamma}_j \times A_{jk} + \sum_{j=1}^{J_K} \hat{\lambda}_j \times K_{jk}\right)$$

ir pasikliautinąjį intervalą $PI_{0,95}(D_k) = (P_a; P_v)$.

3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k$ ir jo pasikliautinąjį intervalą $PI_{0,95}(\pi_k) = (P_a; P_v)$ (1000 gyv.).

4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas.

5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.

6) Apskaičiuoti santykinės rizikos koeficiento ρ įvertį ir pasikliautinąjį intervalą $PI_{0,95}(\rho) = (P_a; P_v)$ skirtingoms kohortoms.

7) Vaizduoti modelio lytis-amžius-kohorta rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius-kohorta realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos iškvietimo sakinio sintaksė:

```
%Puasono_regresija_3(
    sk = N,
    kodas = Mp,
    kintamieji = LYTIS AMZIUS KOHORTA),
```

čia

- **sk** – amžiaus ir kohortos grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **kintamieji** – regresijos lygties kintamieji;

Sudarytą Puasono regresijos modelį lytis-amžius-kohorta taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir iširti stebimų veiksnių įtaką.

Puasono regresijos modelis taikomas iškviečiant makro komandą:

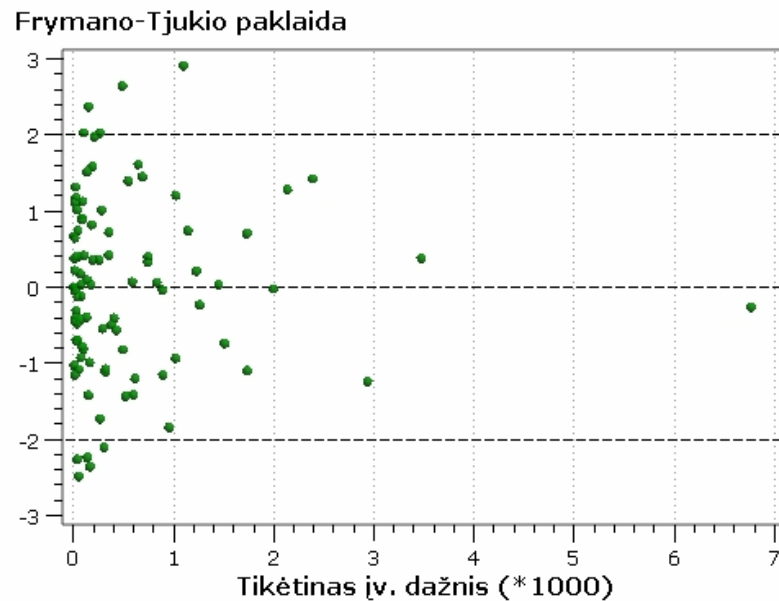
```
%Puasono_regresija_3(sk = 5, kodas = 5, kintamieji = LYTIS AMZIUS KOHORTA) .
```

Gauti Puasono regresijos modelio lytis-amžius-kohorta regresijos lygties parametrų įverčiai ir jų pasikliautinieji intervalai pateikti 5 priedo 2 lentelėje. Gauti regresijos lygties parametrų $\hat{\lambda}_j$ (prie kohortos pseudokintamųjų) įverčiai ir 95% pasikliautinieji intervalai yra didelės reikšmės, nes paskutinėje kohortoje (1976-1980 m.) neįvyko tiriamų įvykių, o kitų $\hat{\lambda}_j$ parametrų reikšmės yra išreiškiamos kaip santykinės rizikos su 1976-1980 m. kohorta logaritmai. Modelio taikymas analizuojamiems duomenims yra nekorektiškas.

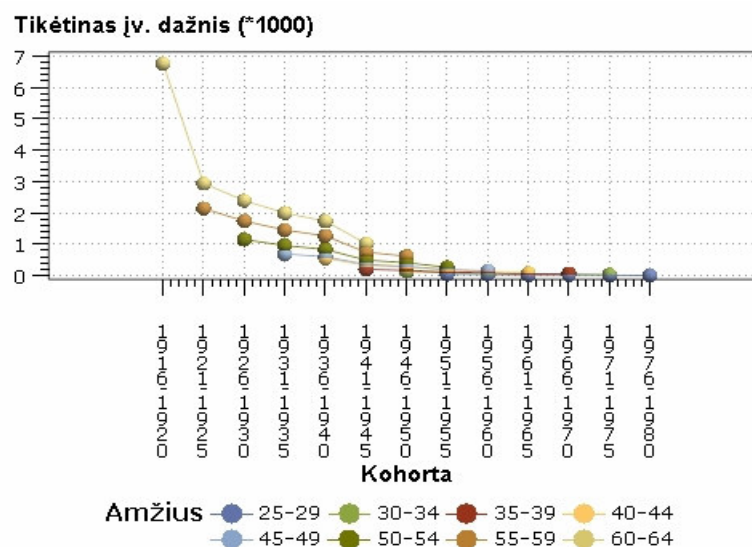
Tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautiniai intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti lentelėje 5 priedo 1 lentelėje.

Santykinės rizikos koeficientų įverčiai ir jų 95% pasikliautiniai intervalai skirtingoms kohortoms (lyginant su 1921-1925 m.) pateikti 5 priedo 3 lentelėje.

Analizuojant anomaliąsias, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (5 priedo 4-6 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.10 paveiksle. Dauguma paklaidų patenka į intervalą $[-2; 2]$. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 5 priede.



2.10 pav. Modelio lytis-amžius-kohorta Frymano-Tjukio paklaidų grafikas



2.11 pav. Modelio lytis-amžius-kohorta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos

2.11 paveiksle pateikta modelio lytis-amžius-kohorta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos. Statistiškai įvertinti skirtumus tarp skirtingų kohortų tikėtinų mirtingumo dažnių leidžia gauti santykinės rizikos koeficientų įverčiai (5 priedo 3 lentelė).

2.2.1.5 Puasono regresijos modelis lytis-amžius-kohortos dreifas

Sudarysime Puasono regresijos modelį lytis-amžius-kohortos dreifas, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį, amžiaus grupę ir kohortą).

2) Apskaičiuoti tikėtino mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^{J_A} \hat{\gamma}_j \times A_{jk} + \Delta \cdot (K_k - K_0)\right) \quad (\text{čia } \Delta - \text{trendo parametras, } K_0 - \text{kohortos}$$

atskaitos taškas) ir pasikliautinąjį intervalą $PI_{0.95}(D_k) = (P_a; P_v)$.

3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k$ ir jo pasikliautinąjį intervalą $PI_{0.95}(\pi_k) = (P_a; P_v)$ (1000 gyv.).

4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas.

5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.

6) Vaizduoti modelio lytis-amžius-kohortos dreifas rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius-kohortos dreifas realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos išskvietimo sakinio sintaksė:

```
%Puasono_regresija_3_Drf(
    sk = N,
    kodas = Mp,
    dreifas = N1,
    C_0 = N2
    kintamieji = LYTIS AMZIUS),
```

čia

- **sk** – amžiaus ir kohortos grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **dreifas** – trendo parametras;
- **C_0** – kohortos atskaitos taškas;
- **kintamieji** – regresijos lygties kintamieji;

Sudarytą Puasono regresijos modelį lytis-amžius-kohortos dreifas taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir iširti stebimų veiksnių įtaką.

Puasono regresijos modelis taikomas iškviečiant makro komandą:

```
%Puasono_regresija_3_Drf(sk = 5, kodas = 5, dreifas = -0.055, P_0 = 1916,
kintamieji = LYTIS AMZIUS).
```

Gauti Puasono regresijos modelio lytis-amžius-kohortos dreifas regresijos lygties parametru įverčiai ir jų pasikliautiniai intervalai pateikti 2.7 lentelėje.

2.7 lentelė

Modelio lytis-amžius-kohortos dreifas regresijos lygties parametru įverčiai

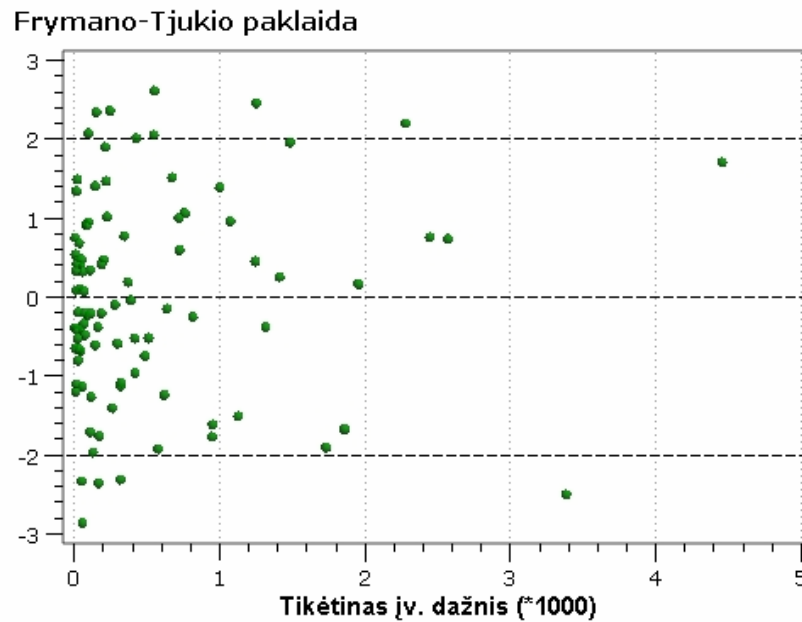
| Parametras | | Laisvės laipsniai | Įvertis | Standartinė paklaida | Voldo 95% pasikliautinio intervalo rėžiai | | χ ² statistikos reikšmė | |
|------------|---------|-------------------|----------|----------------------|---|---------|------------------------------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -5.4140 | 0.0442 | -5.5007 | -5.3273 | 14983.1 | <.0001 |
| LYTIS | Moteris | 1 | -0.6699 | 0.0485 | -0.7649 | -0.5749 | 191.05 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -3.0523 | 0.3186 | -3.6767 | -2.4278 | 91.77 | <.0001 |
| AMZIUS | 30-34 | 1 | -2.0445 | 0.1784 | -2.3941 | -1.6949 | 131.38 | <.0001 |
| AMZIUS | 35-39 | 1 | -1.7977 | 0.1453 | -2.0826 | -1.5129 | 153.03 | <.0001 |
| AMZIUS | 40-44 | 1 | -1.2503 | 0.1030 | -1.4521 | -1.0485 | 147.47 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.0680 | 0.0881 | -1.2406 | -0.8954 | 147.03 | <.0001 |
| AMZIUS | 50-54 | 1 | -0.7255 | 0.0719 | -0.8664 | -0.5846 | 101.81 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.3247 | 0.0606 | -0.4434 | -0.2059 | 28.72 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

Puasono regresijos modelio lytis-amžius-kohortos dreifas lygtis:

$$\hat{D}_k = N_k \cdot \exp(-5,414 - 0,67 \cdot L_{1k} - 3,052 \cdot A_{1k} - 2,045 \cdot A_{2k} - 1,798 \cdot A_{3k} - 1,25 \cdot A_{4k} - 1,068 \cdot A_{5k} - 0,726 \cdot A_{6k} - 0,325 \cdot A_{7k} - 0,055 \cdot (K_k - 1916)).$$

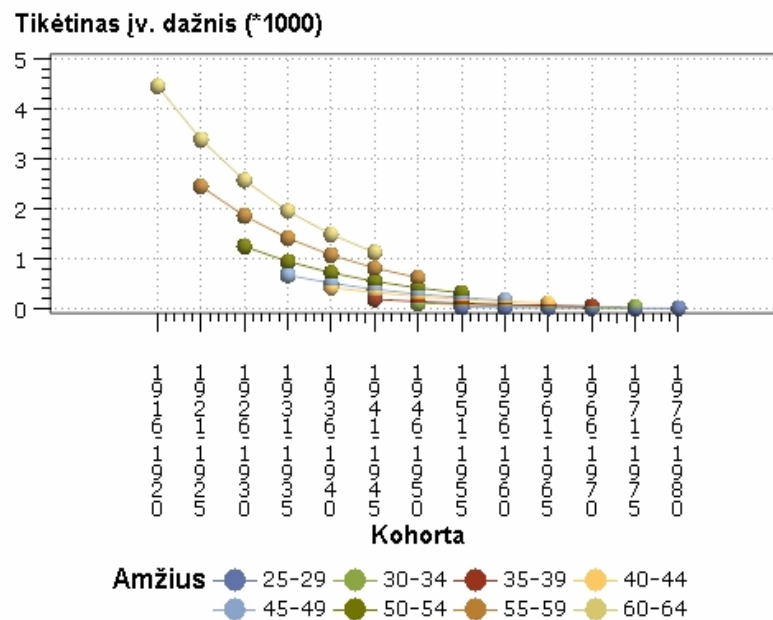
Tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautiniai intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti lentelėje 6 priede.

Analizuojant anomaliąsias, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (6 priedo 2-4 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.12 paveiksle. Dauguma paklaidų patenka į intervalą [-2; 2], todėl modelis tinka analizuojamiems duomenims. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 6 priede.



2.12 pav. Modelio lytis-amžius-kohortos dreifas Frymano-Tjukio paklaidų grafikas

Modelio lytis-amžius-kohortos dreifas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos pateikta 2.13 paveiksle.



2.13 pav. Modelio lytis-amžius-kohortos dreifas vyrų tikėtinų mirtingumo dažnių (1000 gyv.) priklausomybė nuo kohortos

2.2.1.6 Puasono regresijos modelis lytis-amžius-periodas-kohorta

Sudarysime Puasono regresijos modelį lytis-amžius-periodas-kohorta, skirtą prognozuoti tikėtiną mirties nuo nurodytos ligos įvykių skaičių, kuris turi:

- 1) Suformuoti duomenų failą, tinkamą regresinei analizei (t.y. turimi individualūs pacientų stebėjimo duomenys turi būti sugrupuoti pagal lytį, amžiaus grupę, periodą ir kohortą).
- 2) Apskaičiuoti tikėtino mirties nuo nurodytos ligos (kodas Mp) įvykių skaičiaus įvertį

$$\hat{D}_k = N_k \cdot \exp\left(\hat{\alpha} + \hat{\beta} \times L_{1k} + \sum_{j=1}^{J_A} \hat{\gamma}_j \times A_{jk} + \sum_{j=1}^{J_P} \hat{\eta}_j \times P_{jk} + \sum_{j=1}^{J_K} \hat{\lambda}_j \times K_{jk}\right)$$

ir pasikliautinąjį intervalą $PI_{0,95}(D_k) = (P_a; P_v)$.

- 3) Apskaičiuoti tikėtiną įvykio dažnį $\hat{\pi}_k$ ir jo pasikliautinąjį intervalą $PI_{0,95}(\pi_k) = (P_a; P_v)$ (1000 gyv.).
- 4) Apskaičiuoti anomalijas (d_i), Pirsono (e_i) ir Frymano-Tjukio (f_i) paklaidas.
- 5) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.
- 6) Vaizduoti modelio lytis-amžius-periodas rezultatus grafiškai.

Puasono regresijos modelis lytis-amžius-periodas-kohorta realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos išskvietimo sakinio sintaksė:

```
%Puasono_regresija_4(
    sk = N,
    kodas = Mp,
    kintamieji = LYTIS AMZIUS PERIODAS KOHORTA),
```

čia

- **sk** – amžiaus, periodo ir kohortos grupavimo intervalo ilgis (metais);
- **kodas** – tiriamos ligos kodas;
- **kintamieji** – regresijos lygties kintamieji.

Sudarytą Puasono regresijos modelį lytis-amžius-periodas-kohorta taikysime turimiems mirtingumo duomenims (I tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) dažnį ir iširti stebimų veiksnių įtaką.

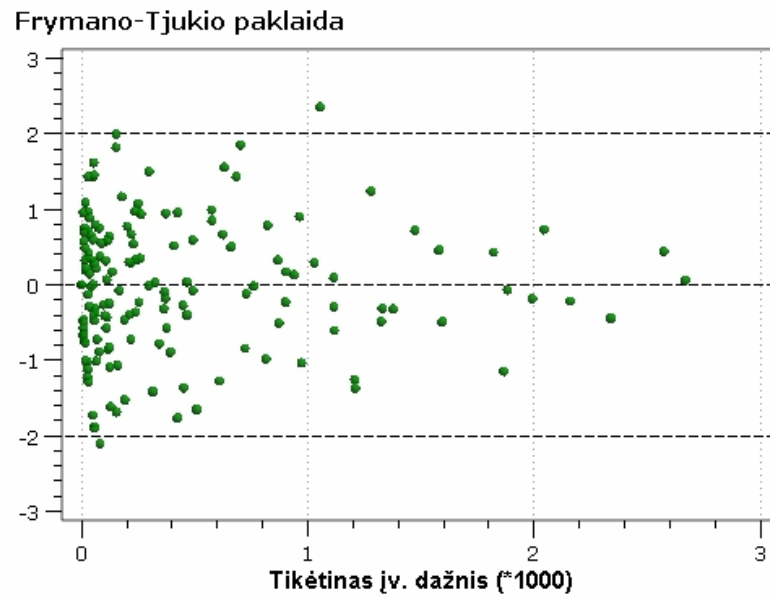
Puasono regresijos modelis taikomas išskviečiant makro komandą:

```
%Puasono_regresija_4(sk = 5, kodas = 5, kintamieji = LYTIS AMZIUS PERIODAS
KOHORTA) .
```

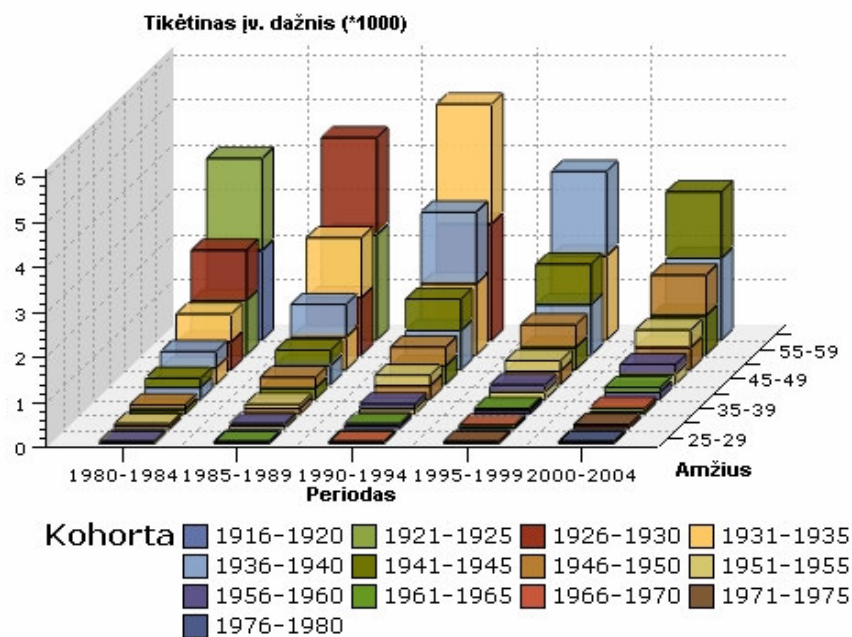
Gauti Puasono regresijos modelio lytis-amžius-periodas regresijos lygties parametų įverčiai ir jų pasikliautinieji intervalai pateikti 7 priedo 2 lentelėje. Gauti regresijos lygties parametų $\hat{\lambda}_j$ (prie kohortos pseudokintamųjų) įverčiai ir 95% pasikliautinieji intervalai yra didelės reikšmės, nes paskutinėje kohortoje (1976-1980 m.) neįvyko tiriamų įvykių, o kitų $\hat{\lambda}_j$ parametų reikšmės yra išreiškiamos kaip santykinės rizikos su 1976-1980 m. kohorta logaritmai. Modelio taikymas analizuojamiems duomenims yra nekorektiškas.

Tikėtinų mirtingumo nuo GSI dažnių (1000 gyventojų) taškiniai įverčiai ir 95% pasikliautiniai intervalai, anomaliosios, Pirsono ir Frymano-Tjukio paklaidos pateikti lentelėje 7 priedo 1 lentelėje.

Analizuojant anomaliąsias, Pirsono ir Frymano-Tjukio paklaidas gauta, kad hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra neatmetamos (7 priedo 3-5 lentelės). Frymano-Tjukio paklaidų grafikas pateiktas 2.14 paveiksle. Dauguma paklaidų patenka į intervalą $[-2; 2]$. Anomaliųjų ir Pirsono paklaidų grafikai pateikti 7 priede.



2.14 pav. Modelio lytis-amžius-periodas-kohorta Frymano-Tjukio paklaidų grafikas



2.15 pav. Modelio lytis-amžius-periodas-kohorta vyrų tikėtinų mirtingumo dažnių (1000 gyv.) diagrama

2.15 paveiksle pateikta modelio lytis-amžius-periodas-kohorta tikėtinų mirtingumo dažnių (1000 gyventojų) diagrama vyrams.

Puasono modelių tinkamumo analizuojamiems duomenims vertinimo rezultatai pateikti 7 priedo 6 lentelėje. Nuokrypio statistikos skirstinys yra chi-kvadrato su $\Lambda - \Omega$ laisvės laipsniais (Λ – stebėjimų skaičius, Ω – regresijos lygties parametru skaičius). Modelis idealiai tinka duomenims, jeigu nuokrypio statistikos ir laisvės laipsnių santykis Θ lygus vienetui.

Gauta (7 priedo 6 lentelė), kad tinkamiausi Puasono regresijos modeliai Kauno miesto 1980-2004 metų mirtingumo nuo GSI duomenims analizuoti yra lytis-amžius-periodas-kohorta ($\Theta = 0,992$) ir lytis-amžius ($\Theta = 1,011$).

2.2.2 Logistinės regresijos modelis

Sudarysime logistinės regresijos modelį, skirtą prognozuoti mirties nuo nurodytos ligos tikimybę pagal klinikinių, sociodemografinių ir gyvenamosios kintamųjų reikšmės. Modelis turi:

1) Apskaičiuoti mirtingumo nuo tiriamos ligos (kodas Mp) tikimybės $\theta_i = P(Y_i = 1 | X_i)$ įvertį

$$\hat{\theta}_i = P(Y_i = 1 | X_i) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_t X_{it})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_t X_{it})}$$

ir tikimybės θ_i pasikliautinąjį intervalą $PI_{0,95}(\theta_i) = (P_a; P_v)$.

2) Apskaičiuoti logistinės regresijos lygties parametru įverčius $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t$ ir patikrinti hipotezes

apie regresijos lygties parametru reikšmingumą $\begin{cases} H_0 : \beta_k = 0, \\ H_1 : \beta_k \neq 0. \end{cases}$ ir $\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_t = 0, \\ H_1 : bent vienas \beta_k \neq 0. \end{cases}$ taikant

Voldo ir χ^2 suderinamumo kriterijus.

3) Apskaičiuoti anomalijas (d_i) ir Pirsono (e_i) paklaidas:

$$d_i = \text{sgn}(Y_i - \hat{Y}_i) \left[2Y_i \cdot \log\left(\frac{Y_i}{\hat{Y}_i}\right) + 2(n_i - Y_i) \cdot \log\left(\frac{n_i - Y_i}{n_i - \hat{Y}_i}\right) \right]^{\frac{1}{2}}, \quad e_i = \frac{Y_i - n_i \hat{\theta}_i}{\sqrt{n_i \hat{\theta}_i (1 - \hat{\theta}_i)}}$$

skaičius i -tojoje grupėje ir $\hat{Y}_i = n_i \hat{\theta}_i$.

4) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę.

5) Apskaičiuoti Kokso–Snelo (r_{CS}^2) ir Nagelkerkės (r_N^2) pseudodeterminacijos koeficientus:

$$r_{CS}^2 = 1 - \left(\frac{L(0)}{L(\hat{\beta})} \right)^{2/n} \quad \text{ir} \quad r_N^2 = 1 - \frac{r_{CS}^2}{\max r_{CS}^2} = \frac{r_{CS}^2}{1 - (L(0))^{2/n}}, \quad \text{čia } L - \text{maksimalaus tikėtinumo funkcija, o}$$

$$L(\hat{\beta}) = \max_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t} L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t) \quad \text{ir} \quad L(0) = \max_{\beta_1=0, \dots, \beta_t=0} L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t).$$

6) Nubraižyti ROC (angl. Receiver Operating Characteristic) kreivę, kuri leidžia įvertinti logistinės regresijos modelio tinkamumą analizuojamiems duomenims. ROC kreivės parametrai: jautrumas

$$j(z) = \frac{1}{\eta_1} \sum_{i \in C_1} I(\hat{\theta}_i \geq z) \text{ ir specifiskumas } s(z) = 1 - \frac{1}{\eta_2} \sum_{i \in C_2} I(\hat{\theta}_i \geq z),$$

čia η_1 - individų, kuriems tiriamas įvykis įvyko (ši grupė žymima C_1), skaičius, η_2 - individų, kuriems tiriamas įvykis neįvyko (ši grupė žymima C_2), skaičius, $\sum_{i \in C_1} I(\hat{\theta}_i \geq z)$ - skaičius teisingai klasifikuotų individų, kuriems įvykis įvyko tiriamas

įvykis, $\sum_{i \in C_2} I(\hat{\theta}_i \geq z)$ - skaičius klaidingai klasifikuotų individų, kuriems neįvyko tiriamas įvykis.

7) Suformuoti duomenų analizės ataskaitą ir pateikti ją RTF (Rich Text Format) formato faile.

Logistinės regresijos modelis realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos iškvietimo sakinio sintaksė:

```
%Logistine_regresija(
    kodas = Mp,
    kintamieji = X1 X2 X3 ... Xt,
    kategoriniai_kintamieji = X2 X5 ... Xt-1 ),
```

čia

- **kodas** – tiriamos ligos kodas;
- **kintamieji** – visi regresijos lygties kintamieji;
- **kategoriniai_kintamieji** – kategoriniai regresijos lygties kintamieji (kuriems reikia sudaryti dvireikšmius pseudokintamuosius).

Sudarytą logistinės regresijos modelį taikysime II tyrimo duomenų matricai. Taikymo tikslas – įvertinti mirties nuo galvos smegenų insulto (GSI) tikimybę ir iširti stebimų veiksnių įtaką.

Logistinės regresijos modelis taikomas iškviečiant makro komandą:

```
%Logistine_regresija(kodas = 4,
    kintamieji = AMZIUS LYTIS UG SV KMI SISTOL DIASTOL RUKONE AKTNE INSULT,
    kategoriniai_kintamieji = LYTIS RUKONE AKTNE INSULT).
```

Pateiksime gautus logistinės regresijos modelio taikymo rezultatus. Makro komandoje panaudotas SAS procedūros LOGISTIC sakinio MODEL parametras SELECTION=Stepwise, kuris nurodo sudaryti regresijos modelį, atmetant nereikšmingus regresijos kintamuosius. Gauta, kad hipotezė $H_0 : \beta_k = 0$ kintamųjų X_2 (lytis), X_3 (ūgis), X_4 (svoris), X_5 (kūno masės indeksas), X_7 (diastolinis AKS), X_9 (fizinis aktyvumas) regresijos lygties parametrus $\beta_k : k = 2, 3, 4, 5, 7, 9$ neatmesta.

Regresijos lygties parametrų įverčiai ir jų pasikliautiniai intervalai pateikti 2.8 lentelėje. Gauta logistinės regresijos modelio lygtis

$$P(Y_i = 1 | X_i) = \frac{\exp(-10,934 + 0,102 \cdot X_{i1} + 0,026 \cdot X_{i6} + 0,357 \cdot \tilde{X}_{i8} - 0,966 \cdot \tilde{X}_{i10})}{1 + \exp(-10,934 + 0,102 \cdot X_{i1} + 0,026 \cdot X_{i6} + 0,357 \cdot \tilde{X}_{i8} - 0,966 \cdot \tilde{X}_{i10})},$$

čia \tilde{X}_{i8} (0 – pacientas nerūko; 1 – pacientas rūko) ir \tilde{X}_{i10} (0 – anksčiau insultas nepatirtas; 1 – anksčiau insultas patirtas) yra pseudokintamieji.

2.8 lentelė

Logistinės regresijos modelio lygties parametru įverčiai

| Parametras | | Įvertis | 95% pasikliautinąo intervalo rėžiai | |
|------------|--------|----------|-------------------------------------|---------|
| Parameter | | Estimate | 95% Confidence Limits | |
| Intercept | | -10.9336 | -12.9425 | -8.9248 |
| AMZIUS | | 0.1017 | 0.0696 | 0.1337 |
| SISTOL | | 0.0260 | 0.0174 | 0.0346 |
| RUKONE | Nerūko | 0.3570 | 0.0907 | 0.6233 |
| INSULT | Ne | -0.9660 | -1.4829 | -0.4491 |

Kokso–Snelo $r_{CS}^2 = 0,068$ ir Nagelkerkės $r_N^2 = 0,202$ pseudodeterminacijos koeficientai rodo, kad modelis taikomiems duomenims netinka.

Apskaičiuoti galimybių santykių įverčiai:

- galimybių santykio pagal amžių įvertis $\hat{OR} = 1,107$ ir 95% pasikliautinis intervalas $PI_{0,95}(OR) = (1,072; 1,143)$, t.y. amžiui padidėjus vieneriais metais galimybė mirti nuo GSI padidėja 1,107 karto;
- galimybių santykio pagal sistolinę AKS įvertis $\hat{OR} = 1,026$ ir 95% pasikliautinis intervalas $PI_{0,95}(OR) = (1,018; 1,035)$, t.y. sistoliniam AKS padidėjus vienu vienetu galimybė mirti nuo GSI padidės 1,026 karto;
- galimybių santykio pagal rūkymą įvertis $\hat{OR} = 2,042$ ir 95% pasikliautinis intervalas $PI_{0,95}(OR) = (1,199; 3,478)$, t.y. rūkantiems galimybė mirti nuo GSI didesnė 2,042 karto;
- galimybių santykio pagal anksčiau patirtą insultą įvertis $\hat{OR} = 0,145$ ir 95% pasikliautinis intervalas $PI_{0,95}(OR) = (0,052; 0,407)$, t.y. patyrusiems insultą galimybė mirti nuo GSI mažesnė 0,145 karto.

Analizuojant anomalijas ir Pirsono paklaidas gauta, kad visais atvejais hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra atmetamos (2.9-2.10 lentelės).

2.9 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

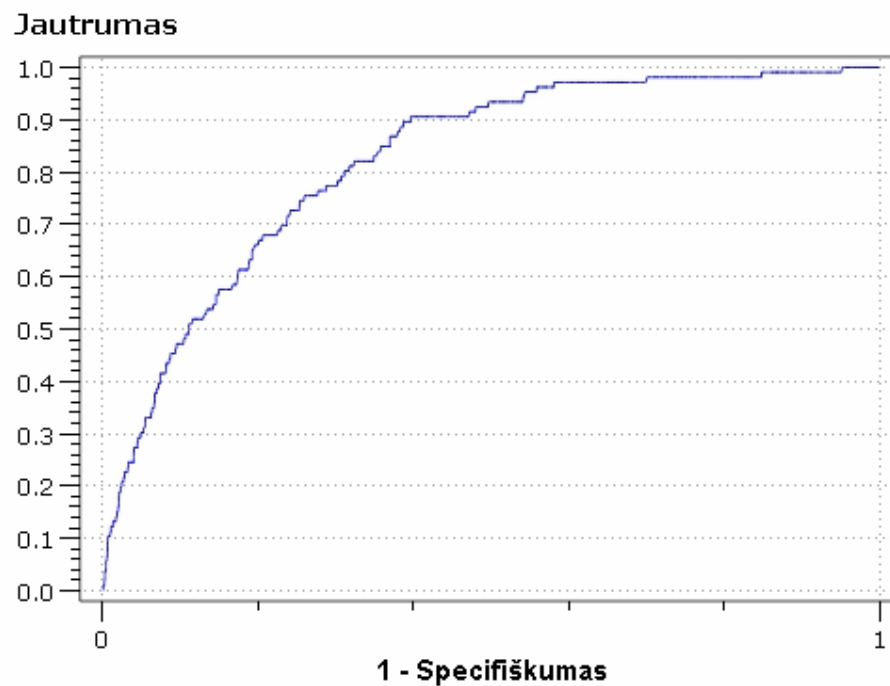
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Kolmogorov-Smirnov | D | 0.394017 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 69.43855 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 358.9368 | Pr > A-Sq | <0.0050 |

2.10 lentelė

Pirsono paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Kolmogorov-Smirnov | D | 0.464646 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 107.3739 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 513.6318 | Pr > A-Sq | <0.0050 |

Modelio tinkamumo turimiems duomenims įvertinimui nubraižyta ROC kreivė (2.16 paveikslas).



2.16 pav. Logistinės regresijos modelio ROC kreivė

Apskaičiuotas plotas po ROC kreive $c = 0,822$. Šiuo atveju didėjant specifiškumui jautrumas didėja ne taip sparčiai, todėl modelis analizuotiems duomenims netinka.

2.2.3 Kokso regresijos modelis

Sudarysime Kokso regresijos modelį, skirtą prognozuoti mirtingumo nuo nurodytos ligos intensyvumą pagal klinikinių, sociodemografinių ir gyvenamosios kintamųjų reikšmės. Modelis turi:

1) Sudaryti Kokso regresijos modelį $\lambda(t | X_i) = \lambda_0(t) \exp(\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_m X_{im})$, čia $X_i = (X_{i1}, X_{i2}, \dots, X_{im})$, $\lambda_0(t)$ – bazinė mirtingumo intensyvumo funkcija.

2) Apskaičiuoti Kokso regresijos lygties parametrų įverčius $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_t$ ir patikrinti hipotezes apie regresijos lygties parametrų reikšmingumą $\begin{cases} H_0 : \beta_k = 0, \\ H_1 : \beta_k \neq 0. \end{cases}$ ir $\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_t = 0, \\ H_1 : bent vienas \beta_k \neq 0. \end{cases}$ taikant Voldo

ir χ^2 suderinamumo kriterijus.

3) Apskaičiuoti anomalijas (d_i) ir martingalines (M_i) paklaidas:

$$d_i = \text{sign}(\hat{M}_i) \sqrt{2 \left(-\hat{M}_i - N_i(\infty) \ln \left(\frac{N_i(\infty) - \hat{M}_i}{N_i(\infty)} \right) \right)}, \quad M_i = \lim_{t \rightarrow \infty} \hat{M}_i(t) = \lim_{t \rightarrow \infty} (N_i(t) - \hat{E}_i(t)), \quad \text{čia } N_i(t) -$$

skaičius įvykių, kurie įvyko individui i iki laiko momento t , o $\hat{E}_i(t)$ - individo i suminės rizikos iki laiko momento t įvertis.

4) Atlikti paklaidų analizę, patikrinti jų skirstinio suderinamumo su standartiniu normaliuoju skirstiniu hipotezę ir grafiškai vaizduoti paklaidų sklaidą.

5) Suformuoti duomenų analizės ataskaitą ir pateikti ją RTF (Rich Text Format) formato faile.

Kokso regresijos modelis realizuotas panaudojus SAS makro programavimo priemones. Pilnas makro komandos tekstas pateiktas 1 priede. Makro komandos iškviatimo sakinio sintaksė:

```
%Kokso_regresija(
    kodas = Mp,
    kintamieji = X1 X2 X3 ... Xt,
    kategoriniai_kintamieji = X2 X5 ... Xt-1 ),
```

čia

- **kodas** – tiriamos ligos kodas;
- **kintamieji** – visi regresijos lygties kintamieji;
- **kategoriniai_kintamieji** – kategoriniai regresijos lygties kintamieji (kuriems reikia sudaryti dvireikšmius pseudokintamuosius).

Sudarytą Kokso regresijos modelį taikysime turimiems mirtingumo duomenims (II tyrimo duomenų matrica). Taikymo tikslas – įvertinti mirtingumo nuo galvos smegenų insulto (GSI) intensyvumo funkciją ir iširti stebimų veiksnių įtaką.

Kokso regresijos modelis taikomas iškviatiant makro komandą:

```
%Kokso_regresija(kodas = 4,
    kintamieji = AMZIUS LYTIS UG SV KMI SISTOL DIASTOL RUKONE AKTNE INSULT,
    kategoriniai_kintamieji = LYTIS RUKONE AKTNE INSULT).
```

Pateiksime gautus Kokso regresijos modelio taikymo rezultatus. Makro komandoje panaudotas SAS procedūros TPHREG sakinio MODEL parametras SELECTION=Stepwise, kuris nurodo sudaryti regresijos modelį, atmetant nereikšmingus regresijos kintamuosius. Gauta, kad hipotezė $H_0 : \beta_k = 0$ kintamųjų X_2 (lytis), X_3 (ūgis), X_4 (svoris), X_5 (kūno masės indeksas), X_7 (diastolinis AKS), X_9 (fizinis aktyvumas) regresijos lygties parametrams $\beta_k : k = 2, 3, 4, 5, 7, 9$ neatmetama.

Regresijos lygties parametrai įvertinti (2.11 lentelė) ir gauta Kokso regresijos modelio lygtis

$$\lambda(t | X_i) = \lambda_0(t) \exp(0,09 \cdot X_{i1} + 0,022 \cdot X_{i6} + 0,567 \cdot \tilde{X}_{i8} - 1,436 \cdot \tilde{X}_{i10}),$$

čia \tilde{X}_{i8} (1 – pacientas nerūko; 0 – pacientas rūko) ir \tilde{X}_{i10} (1 – anksčiau insultas nepatirtas; 0 – anksčiau insultas patirtas) yra pseudokintamieji.

2.11 lentelė

Kokso regresijos lygties parametrai ir santykinės rizikos koeficientų įvertiniai

| Parametras | Laisvės laipsniai | Parametro įvertis | Standartinė paklaida | | Santykinės rizikos koeficientas | Santykinės rizikos koeficiento pasikliautinio intervalo režiai | | Kintamojo žymė |
|------------|-------------------|--------------------|----------------------|------------|---------------------------------|--|-------|-----------------------------|
| Parameter | DF | Parameter Estimate | Standard Error | Pr > ChiSq | Hazard Ratio | 95% Hazard Ratio Confidence Limits | | Variable Label |
| AMZIUS | 1 | 0.09038 | 0.01544 | <.0001 | 1.095 | 1.062 | 1.128 | Amžius |
| SISTOL | 1 | 0.02160 | 0.00377 | <.0001 | 1.022 | 1.014 | 1.029 | Sistolinis kraujo spaudimas |
| RUKONE | 1 | 0.56737 | 0.25150 | 0.0241 | 1.764 | 1.077 | 2.887 | Rūkymas |
| INSULT | 1 | -1.43550 | 0.37628 | 0.0001 | 0.238 | 0.114 | 0.498 | Patirtas insultas |

Apskaičiuoti santykinės rizikos koeficientų įvertiniai ir jų pasikliautinieji intervalai (2.11 lentelė):

- santykinės rizikos koeficiento pagal amžių įvertis $\hat{\rho} = 1,095$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (1,062; 1,128)$, t.y. amžiui padidėjus vieneriais metais rizika mirti nuo GSI padidės 1,095 karto;
- santykinės rizikos koeficiento pagal sistolinį AKS įvertis $\hat{\rho} = 1,022$ ir 95% pasikliautinis intervalas $PI_{0,95}(\rho) = (1,014; 1,029)$, t.y. sistoliniam AKS padidėjus vienu vienetu rizika mirti nuo GSI padidės 1,022 karto.

Analizuojant anomalijas ir Pirsono paklaidas gauta, kad visais atvejais hipotezės apie paklaidų skirstinio suderinamumą su standartiniu normaliuoju skirstiniu yra atmetamos (2.12-2.13 lentelės).

2.12 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Kolmogorov-Smirnov | D | 0.393271 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 90.4207 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 463.2256 | Pr > A-Sq | <0.0050 |

2.13 lentelė

Martingalinių paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Kolmogorov-Smirnov | D | 0.456557 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 114.3912 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 569.2736 | Pr > A-Sq | <0.0050 |

Kokso regresijos modelis analizuotiems duomenims netinka.

2.3 SAS makro komandų rinkinys

Sudarytų logistinės, Kokso ir Puasono regresijos statistinių modelių realizacijai sukurtas makro komandų rinkinys, panaudojant sistemos SAS makro programavimo ir matricinės programavimo kalbos IML galimybes. Sukurtos duomenų analizės programinės priemonės skirtos:

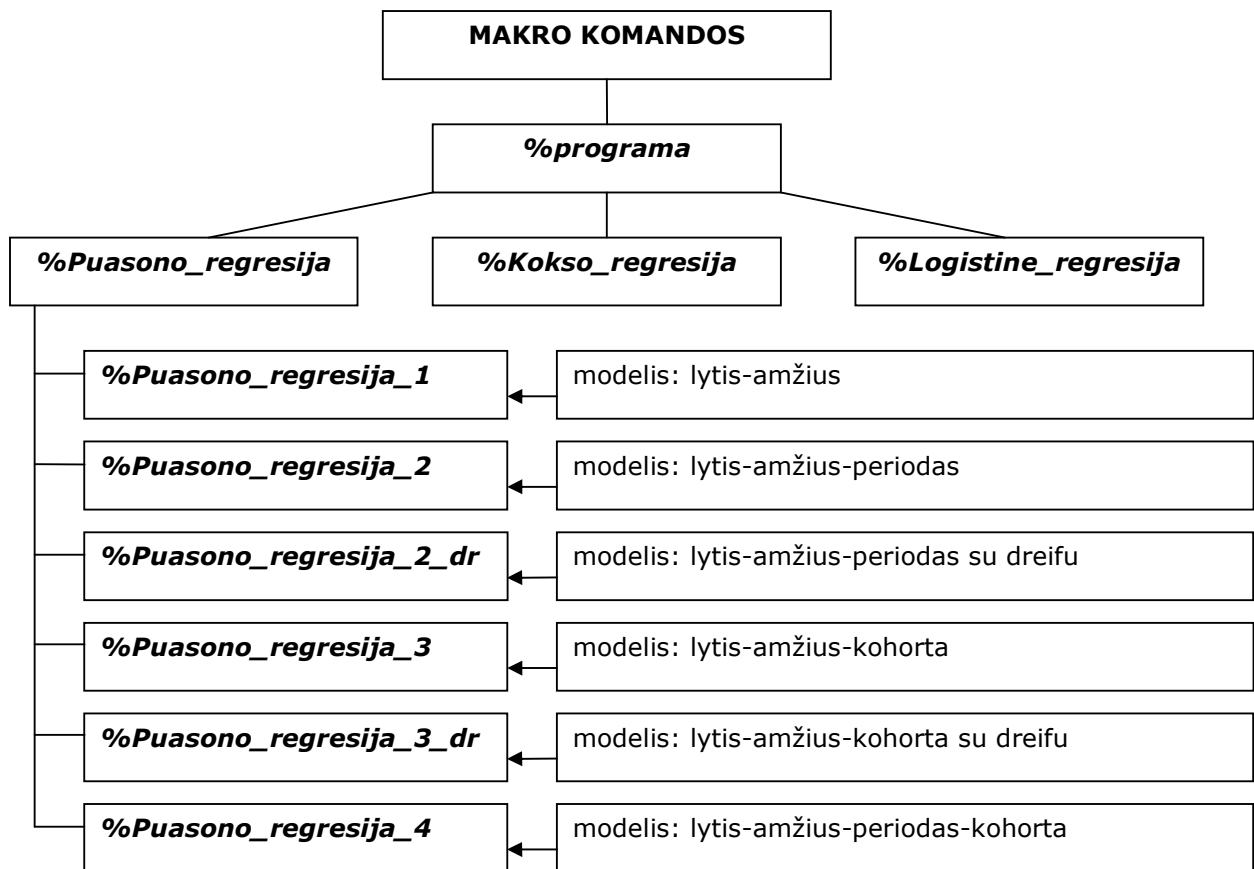
- 1) logistinės, Kokso ir Puasono regresijos statistinių modelių taikymui;
- 2) modelių tinkamumo analizuojamiems duomenims vertinimui;
- 3) gautų duomenų analizės rezultatų išvedimui ir grafiniam vaizdavimui;
- 4) atliktos duomenų analizės ataskaitos rengimo automatizavimui ir įrašymui RTF formato failuose.

Kiekvienas iš sukurtų duomenų regresinės analizės modelių yra realizuotas atskira makro komanda. Sukurtų makro komandų schema pateikta 2.17 paveiksle.

Sukurtų regresinės duomenų analizės programinių priemonių taikymui iškviečiama makro komanda:

```
%Programa ( Puasono_reg          = T,
             duom_failas_Puas_1 = Duomenys_A,
             duom_failas_Puas_2 = Duomenys_B,
             Logistine_reg      = ,
             duom_failas_Log    = Duomenys_C,
             Kokso_reg          = ,
             duom_failas_Koks   = Duomenys_D ); .
```

Visas makro komandos programos tekstas pateiktas 1 priede.



2.17 pav. Makro komandų rinkinio struktūra

Makro komandos `%Programa` parametrai:

- **Puasono_reg, Logistine_reg, Kokso_reg** – loginiai kintamieji, nurodantys kokį regresinės analizės metodą taikyti vykdant makro komandą (taikomo metodo atveju kintamajam priskiriama reikšmė „T“; bet kokios kitos reikšmės atveju metodas netaikomas);
- **duom_failas_Puas_1, duom_failas_Puas_2** – duomenys, kuriems taikomi Puasono regresijos analizės modeliai;
- **duom_failas_Log** – duomenys, kuriems taikomas logistinės regresijos analizės modelis;
- **duom_failas_Koks** – duomenys, kuriems taikomas Kokso regresijos analizės modelis.

Vykiant šią makro komandą išskviečiamos makro komandos, taikančios sukurtus Puasono (`%Puasono_regresija`), logistinės (`%Logistine_regresija`) ir Kokso (`%Kokso_regresija`) regresinės analizės modelius.

Šiame darbe atlikta mirtingumo nuo galvos smegenų insulto duomenų regresinė analizė, tačiau sukurti modeliai ir programinės priemonės gali būti taikomi ir mirtingumo nuo kitų ligų duomenų analizei. Sukurtas programinės priemonės galima nesunkiai modifikuoti, pakeičiant makro komandų parametrų reikšmes, ir atlikti turimų duomenų analizę, taikant Puasono, logistinės ir Kokso regresinės analizės modelius.

Išvados

1. Atlikta literatūros apžvalga parodė, kad mirtingumo nuo įvairių ligų duomenų analizei dažniausiai naudojami Puasono, logistinės ir Kokso regresijos modeliai.
2. Kauno medicinos universiteto Kardiologijos instituto 1980-2004 m. sukauptiems mirtingumo duomenims sudaryti mirtingumo prognozavimo modeliai, panaudojus Puasono, logistinės ir Kokso regresinės analizės metodus.
3. Sudaryti modeliai realizuoti programiškai panaudojus sistemos SAS programavimo priemones. Sukurtas SAS makro komandų rinkinys išplečia sistemos SAS galimybes. Jis leidžia atlikti mirtingumo duomenų regresinę analizę ir pateikia analizės rezultatų ataskaitą vartotojui patogiu RTF (teksto redaktoriaus Word) formatu. Sukurtų programinių priemonių apimtis ~1400 SAS programos kodo eilučių.
4. Panaudojus sukurtus modelius ir programines priemones atlikta Kardiologijos instituto surinktų 1980-2004 metų Kauno miesto 25-64 m. amžiaus gyventojų mirtingumo nuo galvos smegenų insulto duomenų analizė. Gautas mirtingumo nuo galvos smegenų insulto dažnių priklausomybės nuo amžiaus, gimimo metų ir lyties bei regresijos lygtys mirties tikimybės įvertinimui pagal klinikinius, sociodemografinius ir gyvenamosios parametrus. Geriausiai analizuojamiems duomenims tiko Puasono regresijos modeliai.
5. Sudaryti modeliai ir SAS makro komandų rinkinys gali būti pritaikyti mirtingumo nuo kitų ligų duomenų analizei.
6. Gauti rezultatai ir programinės priemonės bus naudojamos Kauno medicinos universiteto Kardiologijos institute.

Literatūra

1. Maura E. Stokes, Charles S. Davis, Gary G. Koch. Categorical Data Analysis Using the SAS® System. SAS Institute Inc., 1995, 499 p.
2. The SAS System OnlineDoc. Cary, NC, USA, SAS Institute Inc., 2006.
3. Dupont William D. Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data. West Nyack, NY, USA Cambridge University Press, 2002, 386 p.
4. Tyler Smith, Besa Smith, Margaret AK Ryan. Survival Analysis Using Cox Proportional Hazards Modeling For Single And Multiple Event Time Data. SUGI 28, paper 254, [žiūrėta 2007-02-12]. Prieiga per internetą: <http://www2.sas.com/proceedings/sugi28/254-28.pdf>
5. Tyler Smith, Besa Smith. Survival Analysis And The Application Of Cox's Proportional Hazards Modeling Using SAS. SUGI 26, paper 244, [žiūrėta 2007-02-12]. Prieiga per internetą: <http://www2.sas.com/proceedings/sugi26/p244-26.pdf>
6. Fox John, Cox Proportional-Hazards Regression for Survival Data, [žiūrėta 2007-04-16]. Prieiga per internetą: <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-cox-regression.pdf>
7. Jon Anderson's Stat 4601 biostatistics course, [žiūrėta 2007-03-02]. Prieiga per internetą: <http://cda.morris.umn.edu/~anderson/math4601/notes>
8. SAS Annotated Output Poisson Regression. UCLA Academic Technology Services, [žiūrėta 2007-03-09]. Prieiga per internetą: http://www.ats.ucla.edu/STAT/SAS/output/sas_poisson_output.htm
9. Čekanavičius V., Murauskas G. Statistika ir jos taikymai. I d. Vilnius, 2003, 238 p.
10. Čekanavičius V., Murauskas G. Statistika ir jos taikymai. II d. Vilnius, 2004, 268 p.
11. Carstensen B., Keiding N. Age-Period-Cohort models: Statistical inference in the Lexis diagram. Department of Biostatistics, University of Copenhagen, 2005, [žiūrėta 2007-02-13]. Prieiga per internetą: <http://staff.pubhealth.ku.dk/~bxc/APC/notes.pdf>
12. Carstensen B. Demography and epidemiology: Age-Period-Cohort models in the computer age. – Steno Diabetes Center, Gentofte, Denmark & Department of Biostatistics, University of Copenhagen, 2005, [žiūrėta 2007-02-13]. Prieiga per internetą: <http://pubhealth.ku.dk/upload/application/pdf/f51d6748/rr-06-1.pdf>
13. Robertson C., Boyle P. Age-Period-Cohort analysis of chronic disease rates. I: Modelling approach. Statistics in Medicine, Vol. 17, 1998, p. 1305-1323.
14. Robertson C., Boyle P. Age, period and cohort models: The use of individual records. Statistics in Medicine, Vol. 5, 1986, p. 527-538.
15. Clayton D., Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Statistics in Medicine, Vol. 6, 1987, p. 449-467.
16. Clayton D., Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. Statistics in Medicine, Vol. 6, 1987, p. 469-481.
17. Holford Theodore R. Understanding the effects of age, period and cohort on incidence and mortality rates. Annual Reviews, Public Health, 12, 1991, p. 425-457.
18. Rėklaitienė R., Tamošiūnas A., Radišauskas R., Jurėnienė K. The role of socio-demographic status and risk factors in mortality from external causes. Acta Medica Lituanica, Vol. 11, No. 4, 2004, p. 42-47.
19. Health Effects of Exposure to Low Levels of Ionizing Radiation: BEIR V. – The National Academy of Sciences, 1990, p. 227-232.

1 priedas. Makro komandų rinkinio programos tekstas

SAS failo Programa.sas programos tekstas

```

filename dir 'C:\Documents and Settings\Administrator\My Documents\Magistrinis
darbas\Programa\';
libname lib 'C:\Documents and Settings\Administrator\My Documents\Magistrinis
darbas\Programa\';
libname duom 'C:\Documents and Settings\Administrator\My Documents\Magistrinis
darbas\Programa\Duomenys';

proc iml;
  %include dir(\Puasono_regresija.sas);
  %include dir(\Logistine_regresija.sas);
  %include dir(\Kokso_regresija.sas);
quit;

/*-----*/

%macro Programa(Puasono_reg = N, duom_failas_Puas_1=, duom_failas_Puas_2=,
                Logistine_reg = N, duom_failas_log =, Kokso_reg = N,
                duom_failas_koks=);

  goptions reset=all;
  goptions device=activex; /**/

  %if %upcase(&Puasono_reg) = T %then %do;

    data lib.Duomenys_P1;
      set &duom_failas_Puas_1;
    run;

    data lib.Duomenys_P2;
      set &duom_failas_Puas_2;
    run;

    %Puasono_regresija( periodo_dydis = 5,
                      kodo_nr = 5 );

  %end;

  %if %upcase(&Logistine_reg) = T %then %do;

    data lib.Duomenys_LK;
      set &duom_failas_Log;
    run;

    %Logistine_regresija( kodas = 4,
                        kintamieji = AMZIUS LYTIS SISTOL DIASTOL UG SV KMI RUKONE AKTNE INSULT,
                        kategoriniai_kintamieji = LYTIS RUKONE AKTNE INSULT );

  %end;

  %if %upcase(&Kokso_reg) = T %then %do;

    data lib.Duomenys_LK;
      set &duom_failas_Koks;
    run;

    %Kokso_regresija( kodas = 4,
                    kintamieji = AMZIUS LYTIS SISTOL DIASTOL UG SV KMI RUKONE AKTNE INSULT,
                    kategoriniai_kintamieji = LYTIS RUKONE AKTNE INSULT );
  %end;

```

```

    %end;

goptions reset=all;

%mend;

/*-----*/
%Programa( Puasono_reg      = T,          /* taikyti Puasono regresijos metoda */
           duom_failas_Puas_1 = duom.Duomenys_P1, /* duomenų */
           duom_failas_Puas_2 = duom.Duomenys_P2, /* failai */

           Logistine_reg    = ,          /* taikyti Logistinės regresijos metoda */
           duom_failas_Log  = duom.Duomenys_3, /* duomenų failas */

           Kokso_reg        = ,          /* taikyti Kokso regresijos metoda */
           duom_failas_Koks = duom.Duomenys_3 ); /* duomenų failas */

/*-----*/

```

SAS failo Puasono_regresija.sas programos tekstas

```

/*----- Puasono regresijos modeliai -----*/

%macro Puasono_regresija_1(sk=, kodas=, kintamieji=);
proc iml;
  start prog(periodas, kod);
  reset deflib=lib;
  use Duomenys_P1;
  read all var{LYTIS AMZIUS METAI KOD_GR} into MATRICA where(METAI>1979 &
METAI<2005);
  mattrib MATRICA colname={'LYTIS' 'AMZIUS' 'METAI' 'KOD_GR'};

  use Duomenys_P2;
  read all var _ALL_ into MATRICA_2;
  mattrib MATRICA_2 colname={'METAI' 'LYTIS' 'AMZIUS' 'VISO'};

  min_1 = min(MATRICA[, 'AMZIUS']);
  max_1 = max(MATRICA[, 'AMZIUS']);
  do ii = 1 to 2;
    do kk = min_1 to max_1-1 by periodas; *lytis;
      M1 = ii||kk;
      suma = 0;
      viso = 0;
      do i = 1 to nrow(MATRICA);
        if MATRICA[i, 'LYTIS'] = ii &
           MATRICA[i, 'AMZIUS'] >= kk & MATRICA[i, 'AMZIUS'] < kk + periodas &
           MATRICA[i, 'KOD_GR'] = kod
        then suma = suma + 1;
      end;
      do i = 1 to nrow(MATRICA_2);
        if MATRICA_2[i, 'LYTIS'] = ii &
           MATRICA_2[i, 'AMZIUS'] = kk
        then viso = viso + MATRICA_2[i, 'VISO'];
      end;
      M1 = M1||suma||viso;
      M = M/M1;
    end;
  end;

  LL = nrow(M)-((ii-2)+((max_1-min_1+1)/periodas));

```

```

create lib.LL from LL [colname={'LL'}];
append from LL;
close lib.LL;

mattrib M colname={'LYTIS' 'AMZIUS' 'ATVEJAI' 'VISO'};
create lib.DUOM_IML from M [colname={'LYTIS' 'AMZIUS' 'ATVEJAI' 'VISO'}];
append from M;
close lib.duom_IML;
finish prog;
run prog(&sk, &kodas);
quit;

data lib.Duom_1;
set lib.DUOM_IML;
LN = log(VISO);
label LYTIS = "Lytis"
      AMZIUS = "Amžius"
      ATVEJAI = "Stebėtas įv. skaičius"
      VISO = "VISO";
format AMZIUS amzius.
      LYTIS lytis.;
run;

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji)";

proc genmod data = lib.Duom_1;
class &kintamieji;
model ATVEJAI = &kintamieji / dist = poisson
                        link = log
                        offset = LN
                        type3;
estimate 'Lytis' LYTIS -1 1 / exp;
estimate 'Amzius 30-34' AMZIUS -1 1 0 0 0 0 0 0 / exp;
estimate 'Amzius 35-39' AMZIUS -1 0 1 0 0 0 0 0 / exp;
estimate 'Amzius 40-44' AMZIUS -1 0 0 1 0 0 0 0 / exp;
estimate 'Amzius 45-49' AMZIUS -1 0 0 0 1 0 0 0 / exp;
estimate 'Amzius 50-54' AMZIUS -1 0 0 0 0 1 0 0 / exp;
estimate 'Amzius 55-59' AMZIUS -1 0 0 0 0 0 1 0 / exp;
estimate 'Amzius 60-64' AMZIUS -1 0 0 0 0 0 0 1 / exp;
output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

data lib.Out_1;
set lib.Out;
IR_reg = (pr/VISO)*1000;
IR = (ATVEJAI/VISO)*1000;
IR_reg_L = (lr/VISO)*1000;
IR_reg_U = (ur/VISO)*1000;
ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
label IR_reg = "Tikėtinas įv. dažnis (*1000)"
      IR = "Stebėtas įv. dažnis (*1000)"
      IR_reg_L = "Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis
(*1000)"
      IR_reg_U = "Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis
rėžis (*1000)"
      lr = "Tikėtino įv. skaičiaus pasikliautinojo intervalo apatinis rėžis"
      ur = "Tikėtino įv. skaičiaus pasikliautinojo intervalo viršutinis rėžis"
      pr = "Tikėtinas įv. skaičius"
      dev = "Anomalioji paklaida"
      chi = "Pirsono paklaida"
      ftr = "Frymano-Tjukio paklaida";
drop LN;
run;

```

```

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_1;
  plot dev*IR_reg
        chi*IR_reg
        ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
  plot pr*ATVEJAI / grid;
run;

data lib.Out2;
  set lib.Out_1 (keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

data lib.Out2;
  set lib.Out_1 (keep=LYTIS AMZIUS IR_reg IR_reg_L IR_reg_U);
run;

title font = "Times New Roman Baltic" ' Pasikliautinių intervalų grafikai ';

proc gplot data = lib.Out2;
  plot IR_reg_L*AMZIUS=1 IR_reg*AMZIUS=2 IR_reg_U*AMZIUS=3 / overlay grid
                                                vaxis = axis1
                                                haxis = axis2;

  by LYTIS;
  axis1 label= ('Tikėtinas įv. dažnis (*1000)');
  axis2 order=(25 to 60 by 5);
  symbol1 color=red value=none i=join height=0.5;
  symbol2 color=blue value=circle i=join height=0.5;
  symbol3 color=red value=none i=join height=0.5;
run;

symbol1 i=join value=dot height=0.8 color=blue;
symbol2 i=join value=dot height=0.8 color=red;

title font = "Times New Roman Baltic" ' Puasono regresijos grafikas ';

proc gplot data = lib.Out_1;
  plot IR*AMZIUS=LYTIS
        IR_reg*AMZIUS=LYTIS / haxis=axis1 grid;
  axis1 order=(25 to 60 by 5);
run;

proc iml;
  reset deflib=lib;
  use Out;
  read all var{dev} into M;
  NUOKRYPIS=SSQ(M);
  mattrib NUOKRYPIS colname={'Nuokrypis'};
  create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
  append from NUOKRYPIS;
  close lib.Lentele;
quit;

data lib.Lentele;
  set lib.Lentele;
  Modelis = "&kintamieji";

```

```

run;

data lib.Lentele;
  merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija_2(sk=, kodash=, kintamieji=);
proc iml;
  start prog(periodas, kod);
  reset deflib=lib;
  use Duomenys_P1;
  read all var{LYTIS AMZIUS METAI KOD_GR} into MATRICA where(METAI>1979 &
METAI<2005);
  mattrib MATRICA colname={'LYTIS' 'AMZIUS' 'METAI' 'KOD_GR'};

  use Duomenys_P2;
  read all var _ALL_ into MATRICA_2;
  mattrib MATRICA_2 colname={'METAI' 'LYTIS' 'AMZIUS' 'VISO'};

  min = min(MATRICA[, 'METAI']);
  max = max(MATRICA[, 'METAI']);
  min_1 = min(MATRICA[, 'AMZIUS']);
  max_1 = max(MATRICA[, 'AMZIUS']);
  do ii = 1 to 2;
    *lytis;
    do jj = min to max-1 by periodas;
      *periodas;
      do kk = min_1 to max_1-1 by periodas;
        *amziaus grupe;
        M1 = ii||jj||kk;
        suma = 0;
        viso = 0;
        do i = 1 to nrow(MATRICA);
          if MATRICA[i, 'LYTIS'] = ii &
            MATRICA[i, 'METAI'] >= jj & MATRICA[i, 'METAI'] < jj + periodas &
            MATRICA[i, 'AMZIUS'] >= kk & MATRICA[i, 'AMZIUS'] < kk + periodas &
            MATRICA[i, 'KOD_GR'] = kod
          then suma = suma + 1;
        end;
        do i = 1 to nrow(MATRICA_2);
          if MATRICA_2[i, 'LYTIS'] = ii &
            MATRICA_2[i, 'METAI'] >= jj & MATRICA_2[i, 'METAI'] < jj + periodas &
            MATRICA_2[i, 'AMZIUS'] = kk
          then viso = viso + MATRICA_2[i, 'VISO'];
        end;
        M1 = M1||suma||viso;
        M = M//M1;
      end;
    end;
  end;
  LL = nrow(M)-((ii-2)+((max-min+1)/periodas)- 1 + ((max_1-min_1+1)/periodas));
  create lib.LL from LL [colname={'LL'}];
  append from LL;
  close lib.LL;

  mattrib M colname={'LYTIS' 'PERIODAS' 'AMZIUS' 'ATVEJAI' 'VISO'};
  create lib.DUOM_IML from M [colname={'LYTIS' 'PERIODAS' 'AMZIUS' 'ATVEJAI'
'VISO'}];
  append from M;
  close lib.duom_IML;
  finish prog;
run prog(&sk, &kodash);

```

```

quit;

data lib.Duom_1;
  set lib.DUOM_IML;
  LN = log(VISO);
  label LYTIS = "Lytis"
        PERIODAS = "Periodas"
        AMZIUS = "Amžius"
        ATVEJAI = "Stebėtas įv. skaičius"
        VISO = "Viso";
  format AMZIUS amzius.
         PERIODAS periodas.
         LYTIS lytis.;
run;

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji)";

proc genmod data = lib.Duom_1;
  class &kintamieji;
  model ATVEJAI = &kintamieji / dist = poisson
                                link = log
                                offset = LN
                                type3;
  estimate 'Periodas 1985-1989' PERIODAS -1 1 0 0 0 / exp;
  estimate 'Periodas 1990-1994' PERIODAS -1 0 1 0 0 / exp;
  estimate 'Periodas 1995-1999' PERIODAS -1 0 0 1 0 / exp;
  estimate 'Periodas 2000-2004' PERIODAS -1 0 0 0 1 / exp;
  output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

data lib.Out_2;
  set lib.Out;
  IR_reg = (pr/VISO)*1000;
  IR = (ATVEJAI/VISO)*1000;
  IR_reg_L = (lr/VISO)*1000;
  IR_reg_U = (ur/VISO)*1000;
  ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
  label IR_reg = "Tikėtinas įv. dažnis (*1000)"
        IR = "Stebėtas įv. dažnis (*1000)"
        IR_reg_L = "Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis
(*1000)"
        IR_reg_U = "Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis
rėžis (*1000)"
        lr = "Tikėtino įv. skaičiaus pasikliautinojo intervalo apatinis rėžis"
        ur = "Tikėtino įv. skaičiaus pasikliautinojo intervalo viršutinis rėžis"
        pr = "Tikėtinas įv. skaičius"
        dev = "Anomalioji paklaida"
        chi = "Pirsono paklaida"
        ftr = "Frymano-Tjukio paklaida";
  drop LN;
run;

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_2;
  plot dev*IR_reg
        chi*IR_reg
        ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
  plot pr*ATVEJAI / grid;
run;

data lib.Out2;

```



```

    set lib.Out_2 (keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

symbol1 i=join value=dot height=0.8 color=black;
symbol2 i=join value=dot height=0.8 color=blue;
symbol3 i=join value=dot height=0.8 color=green;
symbol4 i=join value=dot height=0.8 color=red;
symbol5 i=join value=dot height=0.8 color=magenta;
symbol6 i=join value=dot height=0.8 color=orange;
symbol7 i=join value=dot height=0.8 color=cyan;
symbol8 i=join value=dot height=0.8 color=grey;

title font = "Times New Roman Baltic" ' Puasono regresijos grafikas ';

proc gplot data = lib.Out_2;
  plot IR*AMZIUS=PERIODAS
      IR_reg*AMZIUS=PERIODAS / haxis=axis1 grid;
  axis1 order=(25 to 60 by 5);
  plot IR*PERIODAS=AMZIUS
      IR_reg*PERIODAS=AMZIUS / haxis=axis2 grid;
  axis2 order=(1980 to 2000 by 5);
  by LYTIS;
run;

proc iml;
  reset deflib=lib;
  use Out;
  read all var{dev} into M;
  NUOKRYPIS=SSQ(M);
  mattrib NUOKRYPIS colname={'Nuokrypis'};
  create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
  append from NUOKRYPIS;
  close lib.Lentele;
quit;

data lib.Lentele;
  set lib.Lentele;
  Modelis = "&kintamieji          ";
run;

data lib.Lentele;
  merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija_2_Drf(sk= , kodas= , kintamieji= LYTIS AMZIUS, dreifas=,
P_0= );

data lib.Duom_1;
  set lib.Duom_1;
  P_P0 = PERIODAS-&P_0;
  LN = log(VISO);
  LN = LN + (P_P0*&dreifas);
  keep LYTIS AMZIUS PERIODAS ATVEJAI VISO LN P_P0;
run;

```

```

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji su
PERIODO dreifu &dreifas) ";

proc genmod data = lib.Duom_1;
  class &kintamieji;
  model ATVEJAI = &kintamieji / dist = poisson
                                link = log
                                offset = LN
                                type3;

  output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

data lib.Out_2_Drf;
  set lib.Out;
  IR_reg = (pr/VISO)*1000;
  IR = (ATVEJAI/VISO)*1000;
  IR_reg_L = (lr/VISO)*1000;
  IR_reg_U = (ur/VISO)*1000;
  ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
  label IR_reg = "Tikėtinias įv. dažnis (*1000)"
        IR = "Stebėtas įv. dažnis (*1000)"
        IR_reg_L = "Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis
(*1000)"
        IR_reg_U = "Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis
rėžis (*1000)"
        lr = "Tikėtino įv. skaičiaus pasikliautinojo intervalo apatinis rėžis"
        ur = "Tikėtino įv. skaičiaus pasikliautinojo intervalo viršutinis rėžis"
        pr = "Tikėtinias įv. skaičius"
        dev = "Anomalioji paklaida"
        chi = "Pirsono paklaida"
        ftr = "Frymano-Tjukio paklaida";
  drop LN;
run;

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_2_Drf;
  plot dev*IR_reg
        chi*IR_reg
        ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
  plot pr*ATVEJAI / grid;
run;

data lib.Out2;
  set lib.Out_2_Drf(keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

symbol1 i=join value=dot height=0.8 color=black;
symbol2 i=join value=dot height=0.8 color=blue;
symbol3 i=join value=dot height=0.8 color=green;
symbol4 i=join value=dot height=0.8 color=red;
symbol5 i=join value=dot height=0.8 color=magenta;
symbol6 i=join value=dot height=0.8 color=orange;
symbol7 i=join value=dot height=0.8 color=cyan;
symbol8 i=join value=dot height=0.8 color=grey;

```

```

title font = "Times New Roman Baltic" ' Puasono regresijos grafikas ';

proc gplot data = lib.Out_2_Drf;
  plot IR*AMZIUS=PERIODAS
        IR_reg*AMZIUS=PERIODAS / haxis=axis1 grid;
  axis1 order=(25 to 60 by 5);
  plot IR*PERIODAS=AMZIUS
        IR_reg*PERIODAS=AMZIUS / haxis=axis2 grid;
  axis2 order=(1980 to 2000 by 5);
  by LYTIS;
run;

proc iml;
  reset deflib=lib;
  use Out;
  read all var{dev} into M;
  NUOKRYPIS=SSQ(M);
  mattrib NUOKRYPIS colname={'Nuokrypis'};
  create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
  append from NUOKRYPIS;
  close lib.Lentele;
quit;

data lib.Lentele;
  set lib.Lentele;
  Modelis = "&kintamieji su PERIODO dreifu &dreifas ";
run;

data lib.LL;
  set lib.LL;
  LL = LL + &sk - 1;
run;

data lib.Lentele;
  merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija_3(sk=, kodas=, kintamieji=);

data lib.D1;
  set lib.Duomenys_P1;
  KOHORTA=METAI-AMZIUS;
  if METAI>1979 && METAI<2005;
run;

data lib.D2;
  set lib.Duomenys_P2;
  KOHORTA=METAI-AMZIUS;
  drop METAI;
run;

proc iml;
  start prog(periodas, kod);
  reset deflib=lib;
  use D1;
  read all var{LYTIS AMZIUS KOHORTA KOD_GR} into MATRICA where(METAI>1979 &
METAI<2005);
  mattrib MATRICA colname={'LYTIS' 'AMZIUS' 'KOHORTA' 'KOD_GR'};

```

```

use D2;
read all var _ALL_ into MATRICA_2;
mattrib MATRICA_2 colname={'LYTIS' 'AMZIUS' 'VISO' 'KOHORTA'};

min = min(MATRICA[, 'KOHORTA']);
max = max(MATRICA[, 'KOHORTA']);
min_1 = min(MATRICA[, 'AMZIUS']);
max_1 = max(MATRICA[, 'AMZIUS']);

do ii = 1 to 2;                                *lytis;
  do jj = min to max-1 by periodas;           *kohorta;
    do kk = min_1 to max_1-1 by periodas;    *amziaus grupe;
      M1 = ii||jj||kk;
      suma = 0;
      viso = 0;
      do i = 1 to nrow(MATRICA);
        if MATRICA[i, 'LYTIS'] = ii &
          MATRICA[i, 'KOHORTA'] >= jj & MATRICA[i, 'KOHORTA'] < jj + periodas &
          MATRICA[i, 'AMZIUS'] >= kk & MATRICA[i, 'AMZIUS'] < kk + periodas &
          MATRICA[i, 'KOD_GR'] = kod
        then suma = suma + 1;
      end;
      do i = 1 to nrow(MATRICA_2);
        if MATRICA_2[i, 'LYTIS'] = ii &
          MATRICA_2[i, 'KOHORTA'] >= jj & MATRICA_2[i, 'KOHORTA'] < jj +
periodas &
          MATRICA_2[i, 'AMZIUS'] = kk
        then viso = viso + MATRICA_2[i, 'VISO'];
      end;
      M1 = M1||suma||viso;
      if viso > 0 then M = M//M1;
    end;
  end;
end;

LL = nrow(M)-((ii-2)+((max-min+2)/periodas)- 1 + ((max_1-min_1+1)/periodas));
create lib.LL from LL [colname={'LL'}];
append from LL;
close lib.LL;

mattrib M colname={'LYTIS' 'KOHORTA' 'AMZIUS' 'ATVEJAI' 'VISO'};
create lib.DUOM_IML from M [colname={'LYTIS' 'KOHORTA' 'AMZIUS' 'ATVEJAI'
'VISO'}];
append from M;
close lib.duom_IML;
finish prog;
run prog(&sk, &kodas);
quit;

data lib.Duom_1;
set lib.DUOM_IML;
LN = log(VISO);
label LYTIS = "Lytis"
      KOHORTA = "Kohorta"
      AMZIUS = "Amžius"
      ATVEJAI = "Stebėtas įv. skaičius"
      VISO = "Viso";
format AMZIUS amzius.
      KOHORTA kohorta.
      LYTIS lytis.;

run;

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji)";

```

```

proc genmod data = lib.Duom_1;
  class &kintamieji;
  model ATVEJAI = &kintamieji / dist = poisson
                                link = log
                                offset = LN
                                type3;
  estimate '1916-1920' KOHORTA 1 -1 0 0 0 0 0 0 0 0 0 0 0 0 / exp;
  estimate '1926-1930' KOHORTA 0 -1 1 0 0 0 0 0 0 0 0 0 0 0 / exp;
  estimate '1931-1935' KOHORTA 0 -1 0 1 0 0 0 0 0 0 0 0 0 0 / exp;
  estimate '1936-1940' KOHORTA 0 -1 0 0 1 0 0 0 0 0 0 0 0 0 / exp;
  estimate '1941-1945' KOHORTA 0 -1 0 0 0 1 0 0 0 0 0 0 0 0 / exp;
  estimate '1946-1950' KOHORTA 0 -1 0 0 0 0 1 0 0 0 0 0 0 0 / exp;
  estimate '1951-1955' KOHORTA 0 -1 0 0 0 0 0 1 0 0 0 0 0 0 / exp;
  estimate '1956-1960' KOHORTA 0 -1 0 0 0 0 0 0 1 0 0 0 0 0 / exp;
  estimate '1961-1965' KOHORTA 0 -1 0 0 0 0 0 0 0 1 0 0 0 0 / exp;
  estimate '1966-1970' KOHORTA 0 -1 0 0 0 0 0 0 0 0 1 0 0 0 / exp;
  estimate '1971-1975' KOHORTA 0 -1 0 0 0 0 0 0 0 0 0 1 0 0 / exp;
  estimate '1976-1980' KOHORTA 0 -1 0 0 0 0 0 0 0 0 0 0 1 0 / exp;
  output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

data lib.Out_3;
  set lib.Out;
  IR_reg = (pr/VISO)*1000;
  IR = (ATVEJAI/VISO)*1000;
  IR_reg_L = (lr/VISO)*1000;
  IR_reg_U = (ur/VISO)*1000;
  ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
  label IR_reg = "Tikėtinasis įv. dažnis (*1000)"
        IR = "Stebėtas įv. dažnis (*1000)"
        IR_reg_L = "Tikėtinio įv. dažnio pasikliautinio intervalo apatinis rėžis
(*1000)"
        IR_reg_U = "Tikėtinio įv. dažnio pasikliautinio intervalo viršutinis
rėžis (*1000)"
        lr = "Tikėtinio įv. skaičiaus pasikliautinio intervalo apatinis rėžis"
        ur = "Tikėtinio įv. skaičiaus pasikliautinio intervalo viršutinis rėžis"
        pr = "Tikėtinasis įv. skaičius"
        dev = "Anomalioji paklaida"
        chi = "Pirsono paklaida"
        ftr = "Frymano-Tjukio paklaida";
  drop LN;
run;

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_3;
  plot dev*IR_reg
        chi*IR_reg
        ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
  plot pr*ATVEJAI / grid;
run;

data lib.Out2;
  set lib.Out_3 (keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

```

```

symbol1 i=join value=dot height=0.8 color=black;
symbol2 i=join value=dot height=0.8 color=blue;
symbol3 i=join value=dot height=0.8 color=green;
symbol4 i=join value=dot height=0.8 color=red;
symbol5 i=join value=dot height=0.8 color=magenta;
symbol6 i=join value=dot height=0.8 color=orange;
symbol7 i=join value=dot height=0.8 color=cyan;
symbol8 i=join value=dot height=0.8 color=grey;

title font = "Times New Roman Baltic" ' Puasono regresijos grafikas ';

proc gplot data = lib.Out_3;
  plot IR*KOHORTA=AMZIUS
      IR_reg*KOHORTA=AMZIUS / haxis=axis2 grid;
  axis2 value=(rotate=90) order=(1916 to 1979 by 5);
  by LYTIS;
run;

proc iml;
  reset deflib=lib;
  use Out;
  read all var{dev} into M;
  NUOKRYPIS=SSQ(M);
  mattrib NUOKRYPIS colname={'Nuokrypis'};
  create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
  append from NUOKRYPIS;
  close lib.Lentele;
quit;

data lib.Lentele;
  set lib.Lentele;
  Modelis = "&kintamieji ";
run;

data lib.Lentele;
  merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija_3_Drf(sk= , kodas= , kintamieji= LYTIS AMZIUS, dreifas=,
K_0= );

data lib.Duom_1;
  set lib.Duom_1;
  K_K0 = KOHORTA-&K_0;
  LN = log(VISO);
  LN = LN + (K_K0*&dreifas);
  keep LYTIS AMZIUS KOHORTA ATVEJAI VISO LN K_K0;
run;

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji su
KOHORTOS dreifu &dreifas) ";

proc genmod data = lib.Duom_1;
  class &kintamieji;
  model ATVEJAI = &kintamieji / dist = poisson
                                link = log
                                offset = LN
                                type3;
  output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

```

```

data lib.Out_3_Drf;
  set lib.Out;
  IR_reg = (pr/VISO)*1000;
  IR = (ATVEJAI/VISO)*1000;
  IR_reg_L = (lr/VISO)*1000;
  IR_reg_U = (ur/VISO)*1000;
  ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
  label IR_reg = "Tikėtinasis įv. dažnis (*1000)"
        IR = "Stebėtas įv. dažnis (*1000)"
        IR_reg_L = "Tikėtinio įv. dažnio pasikliautinąjo intervalo apatinis rėžis
(*1000)"
        IR_reg_U = "Tikėtinio įv. dažnio pasikliautinąjo intervalo viršutinis
rėžis (*1000)"
        lr = "Tikėtinio įv. skaičiaus pasikliautinąjo intervalo apatinis rėžis"
        ur = "Tikėtinio įv. skaičiaus pasikliautinąjo intervalo viršutinis rėžis"
        pr = "Tikėtinasis įv. skaičius"
        dev = "Anomalioji paklaida"
        chi = "Pirsono paklaida"
        ftr = "Frymano-Tjukio paklaida";
  drop LN;
run;

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_3_Drf;
  plot dev*IR_reg
        chi*IR_reg
        ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
  plot pr*ATVEJAI / grid;
run;

data lib.Out2;
  set lib.Out_3_Drf(keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

symbol1 i=join value=dot height=0.8 color=black;
symbol2 i=join value=dot height=0.8 color=blue;
symbol3 i=join value=dot height=0.8 color=green;
symbol4 i=join value=dot height=0.8 color=red;
symbol5 i=join value=dot height=0.8 color=magenta;
symbol6 i=join value=dot height=0.8 color=orange;
symbol7 i=join value=dot height=0.8 color=cyan;
symbol8 i=join value=dot height=0.8 color=grey;

title font = "Times New Roman Baltic" ' Puasono regresijos grafikas ';

proc gplot data = lib.Out_3_Drf;
  plot IR*KOHORTA=AMZIUS
        IR_reg*KOHORTA=AMZIUS / haxis=axis2 grid;
  axis2 value=(rotate=90) order=(1916 to 1979 by 5);
  by LYTIS;
run;

proc iml;
  reset deflib=lib;

```

```

    use Out;
    read all var{dev} into M;
    NUOKRYPIS=SSQ(M);
    mattrib NUOKRYPIS colname={'Nuokrypis'};
    create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
    append from NUOKRYPIS;
    close lib.Lentele;
quit;

data lib.Lentele;
    set lib.Lentele;
    Modelis = "&kintamieji su KOHORTOS dreifu &dreifas ";
run;

data lib.LL;
    set lib.LL;
    LL = LL + 12;
run;

data lib.Lentele;
    merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija_4(sk=, kodas=, kintamieji=);

proc iml;
    start prog(periodas, kod);
    reset deflib=lib;
    use Duomenys_P1;
    read all var{LYTIS AMZIUS METAI KOD_GR} into MATRICA where(METAI>1979 &
METAI<2005);
    mattrib MATRICA colname={'LYTIS' 'AMZIUS' 'METAI' 'KOD_GR'};

    use Duomenys_P2;
    read all var _ALL_ into MATRICA_2;
    mattrib MATRICA_2 colname={'METAI' 'LYTIS' 'AMZIUS' 'VISO'};

    min = min(MATRICA[, 'METAI']);
    max = max(MATRICA[, 'METAI']);
    min_1 = min(MATRICA[, 'AMZIUS']);
    max_1 = max(MATRICA[, 'AMZIUS']);
    do ii = 1 to 2;
        *lytis;
        do jj = min to max-1 by periodas;
            *periodas;
            do kk = min_1 to max_1-1 by periodas;
                *amziaus grupe;
                M1 = ii||jj||kk;
                suma_v = 0; viso_v = 0;
                suma_j = 0; viso_j = 0;
                kohorta_v = jj - kk - periodas + 1;
                kohorta_j = jj - kk + 1;
                do i = 1 to nrow(MATRICA);
                    if MATRICA[i, 'KOD_GR'] = kod & MATRICA[i, 'LYTIS'] = ii &
MATRICA[i, 'METAI'] >= jj & MATRICA[i, 'METAI'] < jj + periodas &
MATRICA[i, 'AMZIUS'] >= kk & MATRICA[i, 'AMZIUS'] < kk + periodas
                    then do;
                        skirtumas = (MATRICA[i, 'METAI']-MATRICA[i, 'AMZIUS']);
                        if skirtumas >= kohorta_v & skirtumas < kohorta_v + periodas
                            then suma_v = suma_v + 1;
                            else suma_j = suma_j + 1;
                        end;
                    end;
                end;
            end;
        end;
    end;
end;

```



```

do i = 1 to nrow(MATRICA_2);
  if MATRICA_2[i, 'LYTIS'] = ii &
    MATRICA_2[i, 'AMZIUS'] = kk &
    MATRICA_2[i, 'METAI'] = jj
  then do;
    k = 1;
    do j = 0 to periodas-1;
      viso_v = viso_v + k*MATRICA_2[i+j, 'VISO'];
      viso_j = viso_j + (1-k)*MATRICA_2[i+j, 'VISO'];
      k = k - 0.2;
    end;
    i=nrow(MATRICA_2)+1;
  end;
end;

M1_v = M1||kohorta_v||suma_v||viso_v||0;
M1_j = M1||kohorta_j||suma_j||viso_j||1;
M = M//M1_v;
M = M//M1_j;
end;
end;
end;

min_k = min(M[,4]); max_k = max(M[,4]);

LL = nrow(M)-((ii-2) + ((max-min+1)/periodas)- 1 + ((max_1-min_1+1)/periodas)-
1 + ((max_k-min_k)/periodas))- 1;
create lib.LL from LL [colname={'LL'}];
append from LL;
close lib.LL;

mattrib M colname={'LYTIS' 'PERIODAS' 'AMZIUS' 'KOHORTA' 'ATVEJAI' 'VISO'
'UD'};
create lib.DUOM_IML from M [colname={'LYTIS' 'PERIODAS' 'AMZIUS' 'KOHORTA'
'ATVEJAI' 'VISO' 'UD'}];
append from M;
close lib.duom_IML;
finish prog;
run prog(&sk, &kodas);
quit;

data lib.Duom_1;
set lib.DUOM_IML;
LN = log(VISO);
label LYTIS = "Lytis"
PERIODAS = "Periodas"
AMZIUS = "Amžius"
KOHORTA = "Kohorta"
ATVEJAI = "Stebėtas įv. skaičius"
VISO = "Viso"
UD = "Kohortos tipas";
format AMZIUS amzius.
PERIODAS periodas.
KOHORTA kohorta.
LYTIS lytis.
UD kohortos_tipas.;

run;

title font = "Times New Roman Baltic" " Puasono regresija (&kintamieji)";

proc genmod data = lib.Duom_1;
class &kintamieji;
model ATVEJAI = &kintamieji / dist = poisson
link = log
offset = LN

```

```

                                type3;
estimate 'Lytis' LYTIS -1 1 / exp;
output out = lib.Out pred=pr resdev=dev reschi=chi xbeta=xb lower=lr upper=ur;
run;

data lib.Out_4;
set lib.Out;
IR_reg = (pr/VISO)*1000;
IR = (ATVEJAI/VISO)*1000;
IR_reg_L = (lr/VISO)*1000;
IR_reg_U = (ur/VISO)*1000;
ftr = sqrt(atvejai)+sqrt(atvejai+1)-sqrt(4*pr+1);
label IR_reg = "Tikėtinias įv. dažnis (*1000)"
      IR = "Stebėtas įv. dažnis (*1000)"
      IR_reg_L = "Tikėtinio įv. dažnio pasikliautinojo intervalo apatinis rėžis
(*1000)"
      IR_reg_U = "Tikėtinio įv. dažnio pasikliautinojo intervalo viršutinis
rėžis (*1000)"
      lr = "Tikėtinio įv. skaičiaus pasikliautinojo intervalo apatinis rėžis"
      ur = "Tikėtinio įv. skaičiaus pasikliautinojo intervalo viršutinis rėžis"
      pr = "Tikėtinias įv. skaičius"
      dev = "Anomalioji paklaida"
      chi = "Pirsono paklaida"
      ftr = "Frymano-Tjukio paklaida";
drop LN;
run;

symbol1 color=green value=dot i=none height=0.5;

title font = "Times New Roman Baltic" ' Paklaidų grafikai ';

proc gplot data = lib.Out_4;
plot dev*IR_reg
      chi*IR_reg
      ftr*IR_reg / vref = -2 0 2 lvref=4 cvref=black;
plot pr*ATVEJAI / grid;
run;

data lib.Out2;
set lib.Out_4 (keep=dev chi ftr);
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';
ods select TestsForNormality;

proc univariate data=lib.Out2 normal;
run;

proc sort data=lib.Out_3;
by LYTIS;
run;

title font = "Times New Roman Baltic" ' Puasono regresijos trimatė diagrama ';

proc gchart data=lib.Out_4;
block PERIODAS / discrete
sumvar=IR_reg
subgroup=KOHORTA
group=AMZIUS;
by LYTIS;
run;
quit;

proc iml;

```

```

    reset deflib=lib;
    use Out;
    read all var{dev} into M;
    NUOKRYPIS=SSQ(M);
    mattrib NUOKRYPIS colname={'Nuokrypis'};
    create lib.Lentele from NUOKRYPIS [colname={'Nuokrypis'}];
    append from NUOKRYPIS;
    close lib.Lentele;
quit;

data lib.Lentele;
    set lib.Lentele;
    Modelis = "&kintamieji          ";
run;

data lib.Lentele;
    merge lib.Lentele lib.LL;
run;

%mend;

/*-----*/

%macro Puasono_regresija(periodo_dydis=, kodo_nr=);

data lib.Modeliai;
run;

proc format;
    value lytis 1 = 'Vyras'
              2 = 'Moteris';

    value kohortos_tipas 0 = 'Viršutinė kohorta'
                       1 = 'Apatinė kohorta';

%let eil = %str( );
%do ii = 20 %to 65 %by 5;
    %let jj = %eval(&ii+4);
    %let eil = %str (&eil &ii.="&ii.-&jj.");
%end;
value amzius &eil;

%let eil = %str( );
%do ii = 1975 %to 2005 %by 5;
    %let jj = %eval(&ii+4);
    %let eil = %str (&eil &ii.="&ii.-&jj.");
%end;
value periodas &eil;

%let eil = %str( );
%do ii = 1916 %to 1980 %by 5;
    %let jj = %eval(&ii+4);
    %let eil = %str (&eil &ii.="&ii.-&jj.");
%end;
value kohorta &eil;

run;

proc template;
    define style styles.rtfa;
    parent=styles.sasweb;

    replace body from document /
        leftmargin=1.0in

```

```

    rightmargin=0.5in
    topmargin=0.5in
    bottommargin=0.5in;
end;
run;

ods rtf file = 'Desktop\Puasono_regresija.rtf';
ods rtf style = styles.rtf;
ods rtf keepn;
ods graphics on;

%Puasono_regresija_1(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji = LYTIS
AMZIUS);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

%Puasono_regresija_2(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji = LYTIS
AMZIUS PERIODAS);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

%Puasono_regresija_2_Drf(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji =
LYTIS AMZIUS, dreifas = 0.025, P_0 = 1980);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

%Puasono_regresija_3(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji = LYTIS
AMZIUS KOHORTA);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

%Puasono_regresija_3_Drf(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji =
LYTIS AMZIUS, dreifas = -0.055, K_0 = 1916);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

%Puasono_regresija_4(sk = &periodo_dydis, kodas = &kodo_nr, kintamieji = LYTIS
AMZIUS PERIODAS KOHORTA);

data lib.Modeliai;
    set lib.Modeliai lib.Lentele;
run;

data lib.Modeliai;
    set lib.Modeliai (keep=Modelis Nuokrypis LL);
    Santykis = Nuokrypis/LL;
    label Modelis = "Modelis"
           Nuokrypis = "Nuokrypis"
           LL = "Laisvės laipsniai"
           Santykis = "Nuokrypis/Laisvės laipsniai";
run;

title font = "Times New Roman Baltic" ' Taikyti modeliai ';

```

```

data lib.Modeliai;
  set lib.Modeliai;
  if Nuokrypis > 0;
run;

proc print label noobs data=lib.Modeliai;
  var Modelis Nuokrypis LL Santykis;
run;

ods graphics off;
ods rtf close;

proc datasets library=lib;
  delete L1
        Lentele
        Out
        Out2
        Duom_1
        Duom_1ml
        D1 D2
        Duomenys_P1
        Duomenys_P2;
run;
quit;

&mend;

```

```
/*-----*/
```

SAS failo Logistine_regresija.sas programos tekstas

```
/*----- Logistinės regresijos modelis -----*/
```

```

&macro Logistine_regresija(kodas=, kintamieji=, kategoriniai_kintamieji=);

proc format;
  value lytis 1 = 'Vyras' 2 = 'Moteris';
  value rukymas 1 = 'Nerūko' 2 = 'Rūko';
  value aktyvumas 1 = 'Aktyvus' 2 = 'Neaktyvus';
  value insultas 0 = 'Ne' 1 = 'Taip';
  value gyvas 1 = 'Gyvas' 2 = 'Miręs';
run;

proc template;
  define style styles.rtfa;
  parent=styles.sasweb;

  replace body from document /
    leftmargin=1.0in
    rightmargin=0.5in
    topmargin=0.5in
    bottommargin=0.5in;
  end;
run;

ods rtf file = 'Desktop\Logistine_regresija.rtf';
ods rtf style = styles.rtfa;
ods rtf keepn;
ods graphics on;

data lib.Duom_1;
  set lib.Duomenys_LK;

```

```

if MIRTPRZ = 0 | MIRTPRZ = &kodas;
format LYTIIS lytis.
      RUKONE rukymas.
      AKTNE aktyvumas.
      INSULT insultas.
      GYVIARNE gyvas.;
run;

title font = "Times New Roman Baltic" " Logistinė regresija";

proc logistic data = lib.Duom_1;
  class &kategoriniai_kintamieji;
  model GYVIARNE(event = 'Mireš') = &kintamieji / selection=stepwise
                                          clparm=wald
                                          rsquare
                                          outroc=lib.ROC;

  output out=lib.Out_L resdev=dev reschi=chi pred=pr l=lower u=upper;
run;

data lib.Out_L;
  set lib.Out_L;
  label _LEVEL_ = "Stebimo įvykio reikšmė"
        pr = "Tikimybės įvertis"
        lower = "Pasikliautinojo intervalo apatinis režis"
        upper = "Pasikliautinojo intervalo viršutinis režis"
        dev = "Anomalioji paklaida"
        chi = "Pirsono paklaida";
run;

proc iml;
  use lib.ROC;
  read all var{_STEP_} into M;
  maksimumas = max(M);
  create lib.MAX from maksimumas;
  append from maksimumas;
  close lib.MAX;
quit;

data lib.MAX;
  set lib.MAX;
  call symput('maxm',COL1);
run;

data lib.ROC_1;
  set lib.ROC;
  if _STEP_ = &maxm;
run;

symbol1 i=join v=none c=blue;
title font = "Times New Roman Baltic" " ROC kreivė ";

proc gplot data = lib.ROC_1;
  plot _sensit_*_lmspec_ = 1 / vaxis=axis1 haxis=axis2 grid;
  axis1 label=('Jautrumas') order=(0 to 1 by .1);
  axis2 label=('1 - Specifiškumas');
run;
quit;

data lib.Out_1;
  set lib.Out_L (keep=dev chi);
run;

title font = "Times New Roman Baltic" " Paklaidų analizė ";
ods select TestsForNormality;

```

```
proc univariate data=lib.Out_1 normal;
run;
```

```
ods graphics off;
ods rtf close;
```

```
proc datasets library=lib;
  delete ROC
         ROC_1
         Out_1
         MAX
         DUOM_1
         Duomenys_LK ;
```

```
run;
quit;
```

```
%mend;
```

```
/*-----*/
```

SAS failo Kokso_regresija.sas programos tekstas

```
/*----- Kokso regresijos modelis -----*/
```

```
%macro Kokso_regresija(kodas=, kintamieji=, kategoriniai_kintamieji=);
```

```
proc format;
  value lytis 1 = 'Vyras' 2 = 'Moteris';
  value rukymas 1 = 'Nerūko' 2 = 'Rūko';
  value aktyvumas 1 = 'Aktyvus' 2 = 'Neaktyvus';
  value insultas 0 = 'Ne' 1 = 'Taip';
  value gyvas 1 = 'Gyvas' 2 = 'Miręs';
run;
```

```
proc template;
  define style styles.rtf;
  parent=styles.sasweb;

  replace body from document /
    leftmargin=1.0in
    rightmargin=0.5in
    topmargin=0.5in
    bottommargin=0.5in;
  end;
run;
```

```
ods rtf file = 'Desktop\Kokso_regresija.rtf';
ods rtf style = styles.rtf;
ods rtf keepn;
ods graphics on;
```

```
data lib.Duom_1;
  set lib.Duomenys_LK;
  if MIRTPRZ ^= &kodas then GYVIARNE = 1;
  format LYTIS lytis.
         RUKONE rukymas.
         AKTNE aktyvumas.
         INSULT insultas.
         GYVIARNE gyvas.;
run;
```

```

title font = "Times New Roman Baltic" " Kokso regresija ";

proc tphreg data = lib.Duom_1;
  class &kategoriniai_kintamieji;
  model METAI*GYVIARNE(1) = &kintamieji / selection=stepwise details rl;
  output out = lib.Out_K xbeta=xb resmart=mart resdev=dev;
run;

data lib.Out_K;
  set lib.Out_K;
  label xb = "Tiesinė prognozė"
        lower = "Pasikliautinojo intervalo apatinis rėžis"
        upper = "Pasikliautinojo intervalo viršutinis rėžis"
        dev = "Anomalioji paklaida"
        mart = "Martingalinė paklaida";
run;

title font = "Times New Roman Baltic" ' Paklaidų analizė ';

data lib.Out_1;
  set lib.Out_K (keep=dev mart);
run;

ods select TestsForNormality;

proc univariate data=lib.Out_1 normal;
run;

symbol1 color=green value=dot i=none height=0.25;

proc gplot data = lib.Out_K;
  plot (dev mart)*METAI / vref=0 grid;
run;

ods graphics off;
ods rtf close;

proc datasets library=lib;
  delete Rez Out_1
        Duom_1
        Duomenys_LK;
run;
quit;

%mend;

/*-----*/

```


2 priedas. Modelio lytis-amžius taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius rezultatai

| Lytis | Amžius | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautinąjo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautinąjo intervalo viršutinis rėžis (*1000) | Anomalių paklaida | Pirsono paklaida | Frymano-Tjukio paklaida |
|---------|--------|-----------------------------|-----------------------|------------------------------|------------------------------------|--|--|----------------------|---------------------|----------------------------|
| Vyras | 25-29 | 7 | 410778 | 6.409 | 0.01560 | 0.00839 | 0.02903 | 0.22980 | 0.23325 | 0.31299 |
| Vyras | 30-34 | 23 | 383509 | 21.036 | 0.05485 | 0.03893 | 0.07729 | 0.42183 | 0.42825 | 0.46750 |
| Vyras | 35-39 | 31 | 350449 | 32.151 | 0.09174 | 0.06957 | 0.12098 | -0.20416 | -0.20293 | -0.15969 |
| Vyras | 40-44 | 80 | 324223 | 68.477 | 0.21120 | 0.17461 | 0.25547 | 1.35596 | 1.39251 | 1.36393 |
| Vyras | 45-49 | 97 | 292983 | 98.413 | 0.33590 | 0.28646 | 0.39387 | -0.14275 | -0.14241 | -0.11747 |
| Vyras | 50-54 | 166 | 267525 | 166.372 | 0.62189 | 0.54921 | 0.70419 | -0.02885 | -0.02884 | -0.00948 |
| Vyras | 55-59 | 263 | 228146 | 272.728 | 1.19541 | 1.08254 | 1.32005 | -0.59262 | -0.58906 | -0.57875 |
| Vyras | 60-64 | 374 | 179205 | 375.414 | 2.09489 | 1.92129 | 2.28417 | -0.07304 | -0.07300 | -0.06013 |
| Moteris | 25-29 | 3 | 447458 | 3.591 | 0.00802 | 0.00430 | 0.01496 | -0.32083 | -0.31165 | -0.18740 |
| Moteris | 30-34 | 10 | 424138 | 11.964 | 0.02821 | 0.01995 | 0.03989 | -0.58456 | -0.56785 | -0.51085 |
| Moteris | 35-39 | 20 | 399522 | 18.849 | 0.04718 | 0.03563 | 0.06248 | 0.26240 | 0.26503 | 0.31415 |
| Moteris | 40-44 | 30 | 382295 | 41.523 | 0.10862 | 0.08928 | 0.13213 | -1.88230 | -1.78824 | -1.88143 |
| Moteris | 45-49 | 63 | 356526 | 61.587 | 0.17274 | 0.14638 | 0.20385 | 0.17933 | 0.18002 | 0.20992 |
| Moteris | 50-54 | 107 | 333399 | 106.628 | 0.31982 | 0.28028 | 0.36495 | 0.03601 | 0.03603 | 0.06002 |
| Moteris | 55-59 | 198 | 306251 | 188.272 | 0.61476 | 0.55228 | 0.68432 | 0.70300 | 0.70898 | 0.71733 |
| Moteris | 60-64 | 293 | 270654 | 291.586 | 1.07734 | 0.98184 | 1.18213 | 0.08276 | 0.08283 | 0.09727 |

2 lentelė

**Santykinės rizikos koeficiento skirtingoms amžiaus grupėms įverčiai ir
pasikliautiniai intervalai**

| Contrast Estimate Results | | | | | | | |
|---------------------------|----------|----------------|-------|-------------------|----------|------------|------------|
| Label | Estimate | Standard Error | Alpha | Confidence Limits | | Chi-Square | Pr > ChiSq |
| Amžius 30-34 | 1.2571 | 0.3610 | 0.05 | 0.5496 | 1.9646 | 12.13 | 0.0005 |
| Exp(Amžius 30-34) | 3.5154 | 1.2690 | 0.05 | 1.7326 | 7.1323 | | |
| Amžius 35-39 | 1.7715 | 0.3458 | 0.05 | 1.0937 | 2.4493 | 26.24 | <.0001 |
| Exp(Amžius 35-39) | 5.8796 | 2.0334 | 0.05 | 2.9852 | 11.5806 | | |
| Amžius 40-44 | 2.6053 | 0.3303 | 0.05 | 1.9580 | 3.2527 | 62.22 | <.0001 |
| Exp(Amžius 40-44) | 13.5358 | 4.4708 | 0.05 | 7.0850 | 25.8600 | | |
| Amžius 45-49 | 3.0693 | 0.3260 | 0.05 | 2.4305 | 3.7082 | 88.67 | <.0001 |
| Exp(Amžius 45-49) | 21.5275 | 7.0172 | 0.05 | 11.3641 | 40.7806 | | |
| Amžius 50-54 | 3.6853 | 0.3220 | 0.05 | 3.0542 | 4.3163 | 131.01 | <.0001 |
| Exp(Amžius 50-54) | 39.8567 | 12.8327 | 0.05 | 21.2050 | 74.9140 | | |
| Amžius 55-59 | 4.3388 | 0.3196 | 0.05 | 3.7123 | 4.9653 | 184.24 | <.0001 |
| Exp(Amžius 55-59) | 76.6129 | 24.4892 | 0.05 | 40.9465 | 143.3464 | | |
| Amžius 60-64 | 4.8998 | 0.3186 | 0.05 | 4.2753 | 5.5242 | 236.50 | <.0001 |
| Exp(Amžius 60-64) | 134.2597 | 42.7767 | 0.05 | 71.9023 | 250.6967 | | |

3 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.910898 | Pr < W | 0.1203 |
| Kolmogorov-Smirnov | D | 0.152554 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.09744 | Pr > W-Sq | 0.1135 |
| Anderson-Darling | A-Sq | 0.615225 | Pr > A-Sq | 0.0924 |

4 lentelė

Pirsono paklaidų normalumo tikrinimas

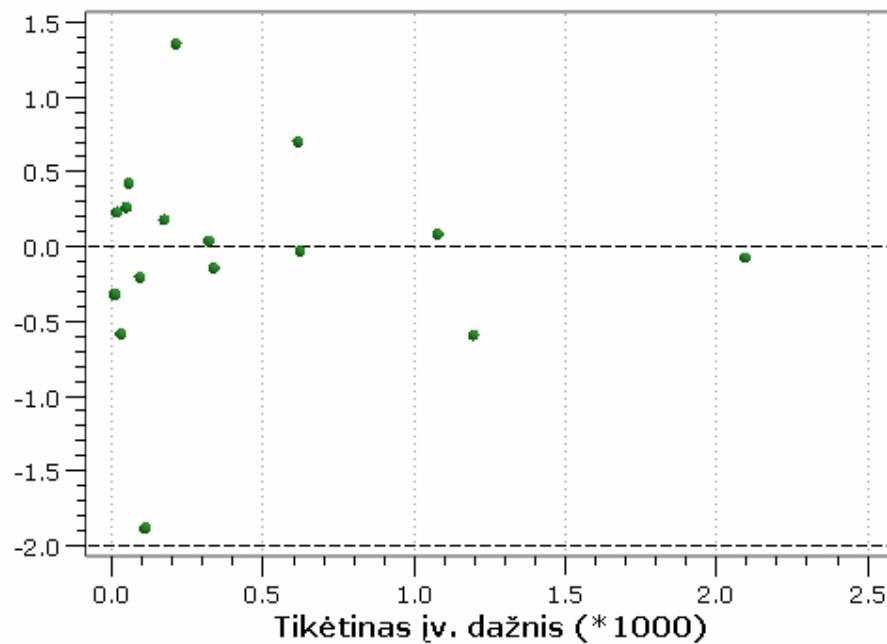
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.920406 | Pr < W | 0.1712 |
| Kolmogorov-Smirnov | D | 0.14703 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.090877 | Pr > W-Sq | 0.1396 |
| Anderson-Darling | A-Sq | 0.575284 | Pr > A-Sq | 0.1169 |

5 lentelė

Frymano-Tjukio paklaidų normalumo tikrinimas

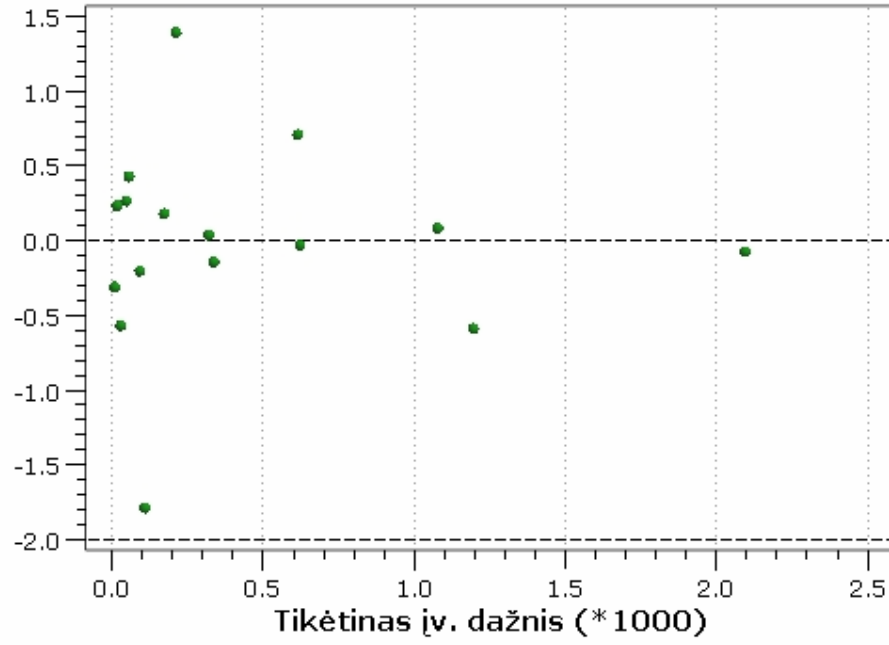
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|--------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.901646 | Pr < W | 0.0854 |
| Kolmogorov-Smirnov | D | 0.203546 | Pr > D | 0.0761 |
| Cramer-von Mises | W-Sq | 0.106624 | Pr > W-Sq | 0.0865 |
| Anderson-Darling | A-Sq | 0.663198 | Pr > A-Sq | 0.0714 |

Anomalioji paklaida



Modelio lytis-amžius anomalijų paklaidų grafikas

Pirsono paklaida



Modelio lytis-amžius Pirsono paklaidų grafikas

3 priedas. Modelio lytis-amžius-periodas taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius-periodas rezultatai

| Lytis | Amžius | Periodas | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis rėžis (*1000) | Anomalioji paklaida | Pirsono paklaida | Frymano- Tjukio paklaida |
|---------|--------|-----------|-----------------------------|-----------------------|------------------------------|------------------------------------|---|---|------------------------|---------------------|--------------------------------|
| Vyras | 25-29 | 1980-1984 | 0 | 79390 | 1.232 | 0.01552 | 0.00827 | 0.02910 | -1.56968 | -1.10993 | -1.43470 |
| Vyras | 30-34 | 1980-1984 | 5 | 67958 | 3.730 | 0.05488 | 0.03837 | 0.07851 | 0.62494 | 0.65781 | 0.69576 |
| Vyras | 35-39 | 1980-1984 | 6 | 64328 | 5.980 | 0.09297 | 0.06923 | 0.12485 | 0.00797 | 0.00798 | 0.10305 |
| Vyras | 40-44 | 1980-1984 | 17 | 64816 | 13.842 | 0.21356 | 0.17222 | 0.26482 | 0.81924 | 0.84879 | 0.85786 |
| Vyras | 45-49 | 1980-1984 | 20 | 57973 | 19.549 | 0.33721 | 0.27929 | 0.40714 | 0.10164 | 0.10203 | 0.15553 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vyras | 55-59 | 2000-2004 | 38 | 45569 | 41.032 | 0.90043 | 0.77681 | 1.04372 | -0.47930 | -0.47329 | -0.44075 |
| Vyras | 60-64 | 2000-2004 | 73 | 39834 | 63.033 | 1.58239 | 1.37840 | 1.81657 | 1.22433 | 1.25541 | 1.23622 |
| Moteris | 25-29 | 1980-1984 | 0 | 87965 | 0.704 | 0.00800 | 0.00426 | 0.01504 | -1.18658 | -0.83904 | -0.95345 |
| Moteris | 30-34 | 1980-1984 | 1 | 77680 | 2.199 | 0.02830 | 0.01973 | 0.04062 | -0.90646 | -0.80841 | -0.71545 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Moteris | 40-44 | 2000-2004 | 2 | 79941 | 6.575 | 0.08225 | 0.06574 | 0.10290 | -2.09513 | -1.78422 | -2.07870 |
| Moteris | 45-49 | 2000-2004 | 7 | 70019 | 9.093 | 0.12987 | 0.10644 | 0.15845 | -0.72376 | -0.69419 | -0.63920 |
| Moteris | 50-54 | 2000-2004 | 20 | 65172 | 15.688 | 0.24071 | 0.20263 | 0.28595 | 1.04385 | 1.08876 | 1.07031 |
| Moteris | 55-59 | 2000-2004 | 23 | 62792 | 29.160 | 0.46439 | 0.39854 | 0.54112 | -1.18492 | -1.14074 | -1.15138 |
| Moteris | 60-64 | 2000-2004 | 52 | 59691 | 48.714 | 0.81611 | 0.70827 | 0.94036 | 0.46563 | 0.47078 | 0.49634 |

2 lentelė

Santykinės rizikos koeficiento skirtingiems periodams įverčiai ir pasikliautiniai intervalai

| Contrast Estimate Results | | | | | | | |
|---------------------------|----------|----------------|-------|-------------------|---------|------------|------------|
| Label | Estimate | Standard Error | Alpha | Confidence Limits | | Chi-Square | Pr > ChiSq |
| Periodas 1985-1989 | 0.0968 | 0.0771 | 0.05 | -0.0542 | 0.2479 | 1.58 | 0.2090 |
| Exp(Periodas 1985-1989) | 1.1017 | 0.0849 | 0.05 | 0.9472 | 1.2813 | | |
| Periodas 1990-1994 | 0.2319 | 0.0739 | 0.05 | 0.0872 | 0.3767 | 9.86 | 0.0017 |
| Exp(Periodas 1990-1994) | 1.2610 | 0.0931 | 0.05 | 1.0911 | 1.4575 | | |
| Periodas 1995-1999 | -0.1702 | 0.0804 | 0.05 | -0.3277 | -0.0127 | 4.48 | 0.0342 |
| Exp(Periodas 1995-1999) | 0.8435 | 0.0678 | 0.05 | 0.7206 | 0.9874 | | |
| Periodas 2000-2004 | -0.2920 | 0.0838 | 0.05 | -0.4563 | -0.1277 | 12.13 | 0.0005 |
| Exp(Periodas 2000-2004) | 0.7468 | 0.0626 | 0.05 | 0.6336 | 0.8801 | | |

3 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|--------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.978877 | Pr < W | 0.2078 |
| Kolmogorov-Smirnov | D | 0.087251 | Pr > D | 0.1354 |
| Cramer-von Mises | W-Sq | 0.095549 | Pr > W-Sq | 0.1302 |
| Anderson-Darling | A-Sq | 0.53847 | Pr > A-Sq | 0.1690 |

4 lentelė

Pirsono paklaidų normalumo tikrinimas

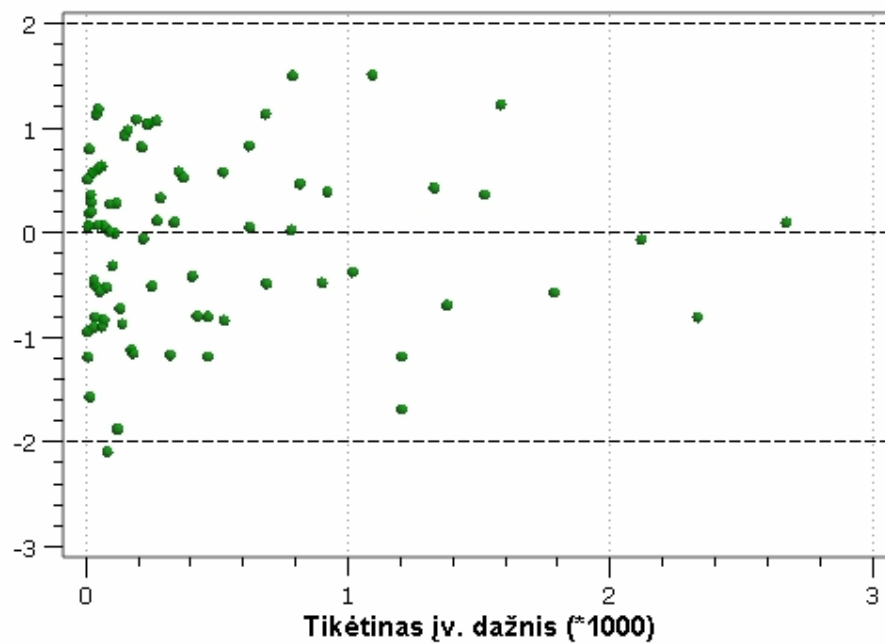
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.978028 | Pr < W | 0.1840 |
| Kolmogorov-Smirnov | D | 0.081236 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.094037 | Pr > W-Sq | 0.1361 |
| Anderson-Darling | A-Sq | 0.560458 | Pr > A-Sq | 0.1465 |

5 lentelė

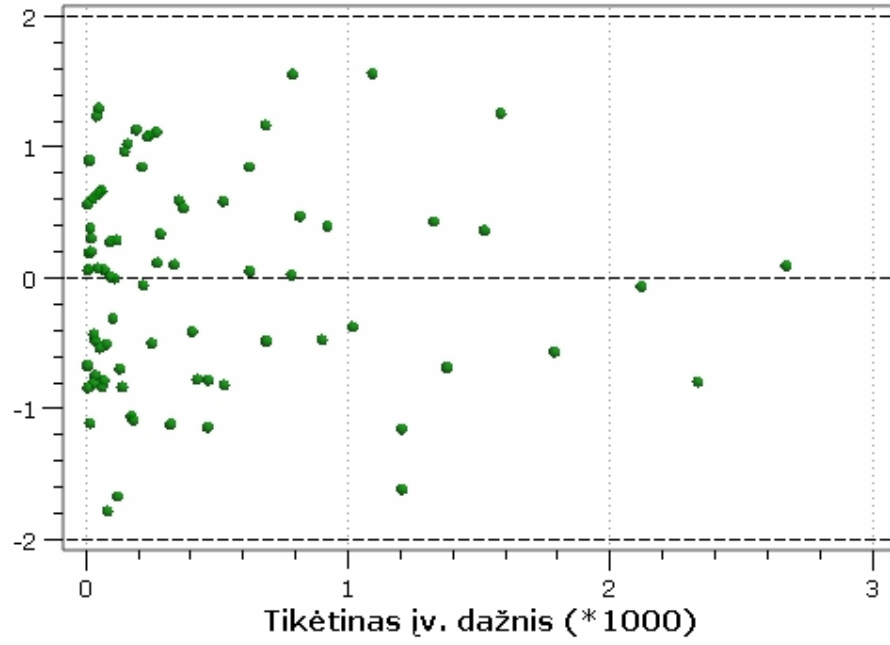
Frymano-Tjukio paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.976844 | Pr < W | 0.1550 |
| Kolmogorov-Smirnov | D | 0.084096 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.103439 | Pr > W-Sq | 0.0998 |
| Anderson-Darling | A-Sq | 0.578474 | Pr > A-Sq | 0.1335 |

Anomalioji paklaida



Modelio lytis-amžius-periodas anomalijų paklaidų grafikas

Pirsono paklaida**Modelio lytis-amžius-periodas Pirsono paklaidų grafikas**

4 priedas. Modelio lytis-amžius-periodas dreifas taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius-periodo dreifas rezultatai

| Lytis | Amžius | Periodas | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautinąjo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautinąjo intervalo viršutinis rėžis (*1000) | Anomalioji paklaida | Pirsono paklaida | Frymano- Tjukio paklaida |
|---------|--------|-----------|-----------------------------|-----------------------|------------------------------|------------------------------------|---|---|------------------------|---------------------|--------------------------------|
| Vyras | 25-29 | 1980-1984 | 0 | 79390 | 0.959 | 0.01209 | 0.00650 | 0.02248 | -1.38523 | -0.97951 | -1.19949 |
| Vyras | 30-34 | 1980-1984 | 5 | 67958 | 2.852 | 0.04197 | 0.02979 | 0.05914 | 1.14801 | 1.27179 | 1.16297 |
| Vyras | 35-39 | 1980-1984 | 6 | 64328 | 4.490 | 0.06980 | 0.05293 | 0.09205 | 0.67743 | 0.71263 | 0.74095 |
| Vyras | 40-44 | 1980-1984 | 17 | 64816 | 10.454 | 0.16128 | 0.13333 | 0.19509 | 1.85477 | 2.02470 | 1.82245 |
| Vyras | 45-49 | 1980-1984 | 20 | 57973 | 14.946 | 0.25781 | 0.21986 | 0.30231 | 1.24235 | 1.30724 | 1.25826 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vyras | 55-59 | 2000-2004 | 38 | 45569 | 68.288 | 1.49857 | 1.35706 | 1.65483 | -4.00363 | -3.66524 | -4.14817 |
| Vyras | 60-64 | 2000-2004 | 73 | 39834 | 103.055 | 2.58711 | 2.37290 | 2.82065 | -3.12547 | -2.96060 | -3.18146 |
| Moteris | 25-29 | 1980-1984 | 0 | 87965 | 0.547 | 0.00622 | 0.00334 | 0.01160 | -1.04612 | -0.73972 | -0.78570 |
| Moteris | 30-34 | 1980-1984 | 1 | 77680 | 1.678 | 0.02160 | 0.01528 | 0.03055 | -0.56645 | -0.52345 | -0.36290 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Moteris | 40-44 | 2000-2004 | 2 | 79941 | 10.941 | 0.13687 | 0.11251 | 0.16650 | -3.32946 | -2.70316 | -3.54447 |
| Moteris | 45-49 | 2000-2004 | 7 | 70019 | 15.319 | 0.21879 | 0.18541 | 0.25817 | -2.38196 | -2.12553 | -2.41741 |
| Moteris | 50-54 | 2000-2004 | 20 | 65172 | 26.389 | 0.40491 | 0.35488 | 0.46200 | -1.29977 | -1.24372 | -1.26789 |
| Moteris | 55-59 | 2000-2004 | 23 | 62792 | 48.434 | 0.77135 | 0.69296 | 0.85860 | -4.07576 | -3.65463 | -4.26002 |
| Moteris | 60-64 | 2000-2004 | 52 | 59691 | 79.487 | 1.33164 | 1.21349 | 1.46128 | -3.29264 | -3.08302 | -3.36788 |

2 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.982625 | Pr < W | 0.3498 |
| Kolmogorov-Smirnov | D | 0.064097 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.042799 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.336314 | Pr > A-Sq | >0.2500 |

3 lentelė

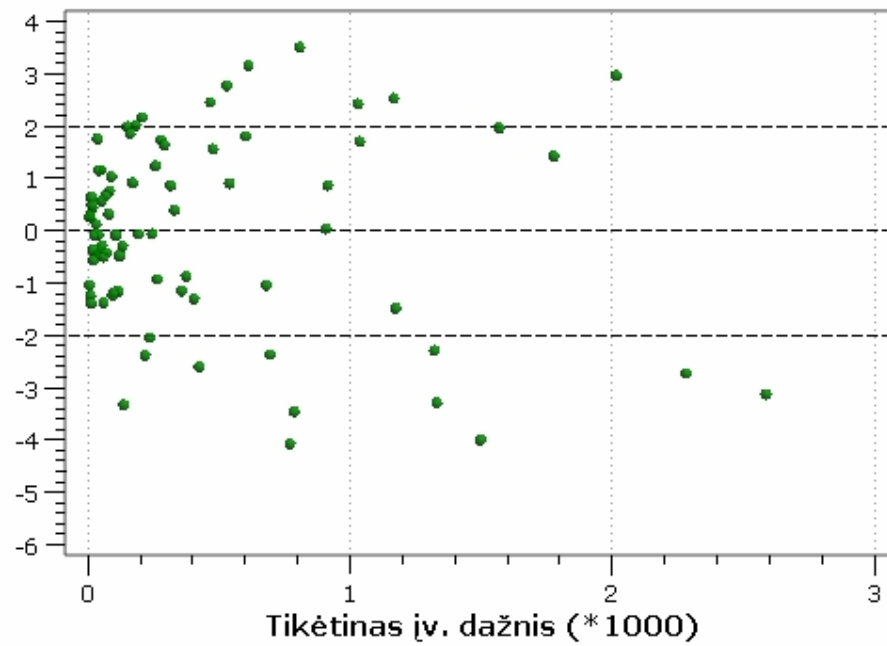
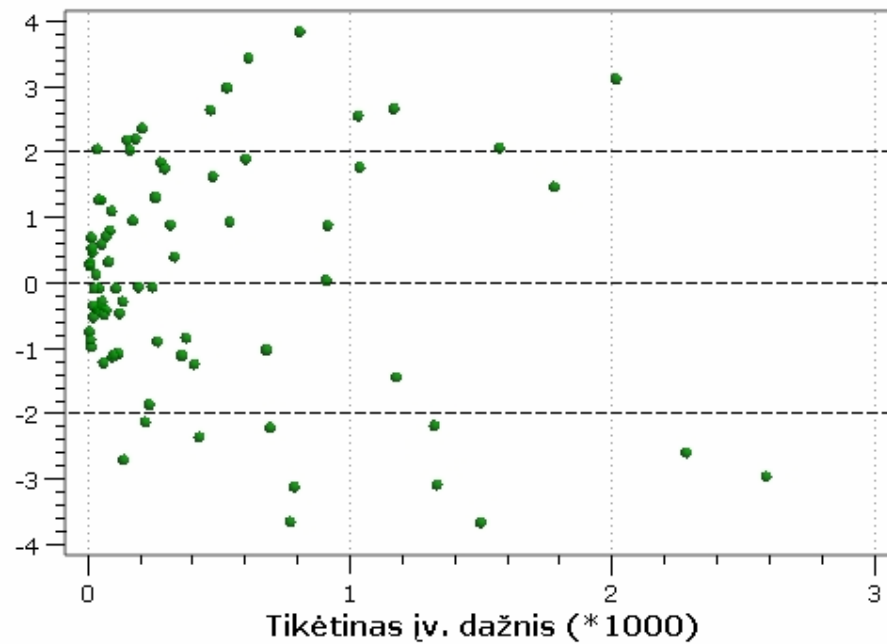
Pirsono paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.98826 | Pr < W | 0.6819 |
| Kolmogorov-Smirnov | D | 0.053694 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.028898 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.223727 | Pr > A-Sq | >0.2500 |

4 lentelė

Frymano-Tjukio paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|--------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.973196 | Pr < W | 0.0908 |
| Kolmogorov-Smirnov | D | 0.092548 | Pr > D | 0.0890 |
| Cramer-von Mises | W-Sq | 0.084115 | Pr > W-Sq | 0.1874 |
| Anderson-Darling | A-Sq | 0.588736 | Pr > A-Sq | 0.1261 |

Anomalių paklaida**Modelio lytis-amžius-periodo dreifas anomaliųjų paklaidų grafikas****Pirsono paklaida****Modelio lytis-amžius-periodo dreifas Pirsono paklaidų grafikas**

5 priedas. Modelio lytis-amžius-kohorta taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius-kohorta rezultatai

| Lytis | Amžius | Kohorta | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis rėžis (*1000) | Anomalioji paklaida | Pirsono paklaida | Frymano- Tjukio paklaida |
|---------|--------|-----------|-----------------------------|-----------------------|------------------------------|------------------------------------|---|---|------------------------|---------------------|--------------------------------|
| Vyras | 60-64 | 1916-1920 | 26 | 4084 | 27.6243 | 6.76402 | 5.12296 | 8.93077 | -0.31214 | -0.30904 | -0.26405 |
| Vyras | 55-59 | 1921-1925 | 19 | 6560 | 14.0082 | 2.13539 | 1.76534 | 2.58301 | 1.26439 | 1.33373 | 1.27904 |
| Vyras | 60-64 | 1921-1925 | 64 | 25393 | 74.5789 | 2.93699 | 2.50350 | 3.44553 | -1.25581 | -1.22499 | -1.23848 |
| Vyras | 50-54 | 1926-1930 | 14 | 10108 | 11.5585 | 1.14350 | 0.95562 | 1.36831 | 0.69485 | 0.71814 | 0.74194 |
| Vyras | 55-59 | 1926-1930 | 60 | 39817 | 69.0867 | 1.73511 | 1.50736 | 1.99726 | -1.11860 | -1.09322 | -1.09751 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vyras | 30-34 | 1971-1975 | 2 | 54997 | 1.6264 | 0.02957 | 0.01072 | 0.08161 | 0.28273 | 0.29299 | 0.40666 |
| Vyras | 25-29 | 1976-1980 | 0 | 55240 | 0.0000 | 0.00000 | 0.00000 | 0.00000 | -0.00009 | -0.00006 | -0.00000 |
| Moteris | 60-64 | 1916-1920 | 25 | 6732 | 23.3757 | 3.47233 | 2.62634 | 4.59083 | 0.33217 | 0.33595 | 0.37776 |
| Moteris | 55-59 | 1921-1925 | 23 | 10271 | 11.2592 | 1.09621 | 0.90389 | 1.32944 | 3.06213 | 3.49902 | 2.90978 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Moteris | 30-34 | 1966-1970 | 4 | 86731 | 1.8799 | 0.02167 | 0.01166 | 0.04028 | 1.34180 | 1.54630 | 1.31724 |
| Moteris | 35-39 | 1966-1970 | 1 | 59910 | 1.6020 | 0.02674 | 0.01421 | 0.05033 | -0.51134 | -0.47560 | -0.30752 |
| Moteris | 25-29 | 1971-1975 | 1 | 84132 | 0.5090 | 0.00605 | 0.00205 | 0.01785 | 0.60712 | 0.68819 | 0.67179 |
| Moteris | 30-34 | 1971-1975 | 0 | 59823 | 0.9082 | 0.01518 | 0.00549 | 0.04195 | -1.34771 | -0.95297 | -1.15235 |
| Moteris | 25-29 | 1976-1980 | 0 | 58438 | 0.0000 | 0.00000 | 0.00000 | 0.00000 | -0.00006 | -0.00005 | -0.00000 |

2 lentelė

Modelio lytis-amžius-kohorta regresijos lygties parametru įverčiai

| Analysis Of Parameter Estimates | | | | | | | | |
|---------------------------------|-----------|----|----------|----------------|----------------------------|----------|------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -27.5217 | 13080.99 | -25665.8 | 25610.75 | 0.00 | 0.9983 |
| LYTIS | Moteris | 1 | -0.6668 | 0.0485 | -0.7618 | -0.5718 | 189.11 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -2.7892 | 0.3653 | -3.5052 | -2.0733 | 58.31 | <.0001 |
| AMZIUS | 30-34 | 1 | -1.8693 | 0.2253 | -2.3108 | -1.4277 | 68.85 | <.0001 |
| AMZIUS | 35-39 | 1 | -1.6593 | 0.1753 | -2.0028 | -1.3158 | 89.64 | <.0001 |
| AMZIUS | 40-44 | 1 | -1.1528 | 0.1235 | -1.3948 | -0.9108 | 87.19 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.0657 | 0.1000 | -1.2617 | -0.8697 | 113.58 | <.0001 |
| AMZIUS | 50-54 | 1 | -0.7357 | 0.0790 | -0.8905 | -0.5809 | 86.73 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.3187 | 0.0640 | -0.4441 | -0.1934 | 24.84 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| KOHORTA | 1916-1920 | 1 | 22.5256 | 13080.99 | -25615.7 | 25660.79 | 0.00 | 0.9986 |
| KOHORTA | 1921-1925 | 1 | 21.6914 | 13080.99 | -25616.6 | 25659.96 | 0.00 | 0.9987 |
| KOHORTA | 1926-1930 | 1 | 21.4838 | 13080.99 | -25616.8 | 25659.75 | 0.00 | 0.9987 |
| KOHORTA | 1931-1935 | 1 | 21.3033 | 13080.99 | -25617.0 | 25659.57 | 0.00 | 0.9987 |
| KOHORTA | 1936-1940 | 1 | 21.1635 | 13080.99 | -25617.1 | 25659.43 | 0.00 | 0.9987 |
| KOHORTA | 1941-1945 | 1 | 20.6334 | 13080.99 | -25617.6 | 25658.90 | 0.00 | 0.9987 |
| KOHORTA | 1946-1950 | 1 | 20.4442 | 13080.99 | -25617.8 | 25658.71 | 0.00 | 0.9988 |
| KOHORTA | 1951-1955 | 1 | 20.0159 | 13080.99 | -25618.3 | 25658.28 | 0.00 | 0.9988 |
| KOHORTA | 1956-1960 | 1 | 19.7722 | 13080.99 | -25618.5 | 25658.04 | 0.00 | 0.9988 |
| KOHORTA | 1961-1965 | 1 | 19.1646 | 13080.99 | -25619.1 | 25657.43 | 0.00 | 0.9988 |
| KOHORTA | 1966-1970 | 1 | 19.3184 | 13080.99 | -25618.9 | 25657.59 | 0.00 | 0.9988 |
| KOHORTA | 1971-1975 | 1 | 18.9623 | 13080.99 | -25619.3 | 25657.23 | 0.00 | 0.9988 |
| KOHORTA | 1976-1980 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

3 lentelė

Santykinės rizikos koeficiento skirtingoms kohortoms įverčiai ir pasikliautiniai intervalai

| Contrast Estimate Results | | | | | | | |
|---------------------------|----------|----------------|-------|-------------------|----------|------------|------------|
| Label | Estimate | Standard Error | Alpha | Confidence Limits | | Chi-Square | Pr > ChiSq |
| 1916-1920 | 0.8342 | 0.1604 | 0.05 | 0.5199 | 1.1486 | 27.05 | <.0001 |
| Exp(1916-1920) | 2.3030 | 0.3694 | 0.05 | 1.6818 | 3.1538 | | |
| 1926-1930 | -0.2076 | 0.0993 | 0.05 | -0.4022 | -0.0130 | 4.37 | 0.0366 |
| Exp(1926-1930) | 0.8125 | 0.0807 | 0.05 | 0.6689 | 0.9871 | | |
| 1931-1935 | -0.3881 | 0.0978 | 0.05 | -0.5797 | -0.1965 | 15.76 | <.0001 |
| Exp(1931-1935) | 0.6784 | 0.0663 | 0.05 | 0.5601 | 0.8216 | | |
| 1936-1940 | -0.5278 | 0.0991 | 0.05 | -0.7221 | -0.3336 | 28.37 | <.0001 |
| Exp(1936-1940) | 0.5899 | 0.0585 | 0.05 | 0.4857 | 0.7163 | | |
| 1941-1945 | -1.0580 | 0.1090 | 0.05 | -1.2717 | -0.8443 | 94.17 | <.0001 |
| Exp(1941-1945) | 0.3472 | 0.0378 | 0.05 | 0.2804 | 0.4299 | | |
| 1946-1950 | -1.2471 | 0.1288 | 0.05 | -1.4995 | -0.9948 | 93.81 | <.0001 |
| Exp(1946-1950) | 0.2873 | 0.0370 | 0.05 | 0.2232 | 0.3698 | | |
| 1951-1955 | -1.6754 | 0.1620 | 0.05 | -1.9930 | -1.3579 | 106.92 | <.0001 |
| Exp(1951-1955) | 0.1872 | 0.0303 | 0.05 | 0.1363 | 0.2572 | | |
| 1956-1960 | -1.9192 | 0.1978 | 0.05 | -2.3069 | -1.5315 | 94.12 | <.0001 |
| Exp(1956-1960) | 0.1467 | 0.0290 | 0.05 | 0.0996 | 0.2162 | | |
| 1961-1965 | -2.5268 | 0.2753 | 0.05 | -3.0663 | -1.9873 | 84.27 | <.0001 |
| Exp(1961-1965) | 0.0799 | 0.0220 | 0.05 | 0.0466 | 0.1371 | | |
| 1966-1970 | -2.3729 | 0.3443 | 0.05 | -3.0478 | -1.6980 | 47.49 | <.0001 |
| Exp(1966-1970) | 0.0932 | 0.0321 | 0.05 | 0.0475 | 0.1830 | | |
| 1971-1975 | -2.7291 | 0.5526 | 0.05 | -3.8122 | -1.6459 | 24.39 | <.0001 |
| Exp(1971-1975) | 0.0653 | 0.0361 | 0.05 | 0.0221 | 0.1928 | | |
| 1976-1980 | -21.6914 | 13080.99 | 0.05 | -25660.0 | 25616.58 | 0.00 | 0.9987 |
| Exp(1976-1980) | 0.0000 | 0.0000 | 0.05 | 0.0000 | . | | |

4 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.987682 | Pr < W | 0.5151 |
| Kolmogorov-Smirnov | D | 0.058559 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.044692 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.28363 | Pr > A-Sq | >0.2500 |

5 lentelė

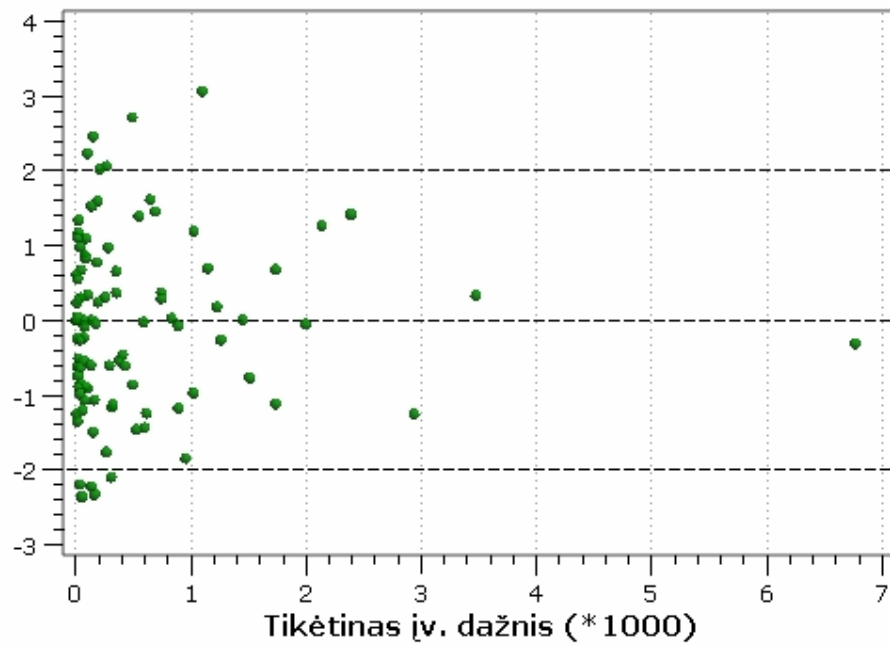
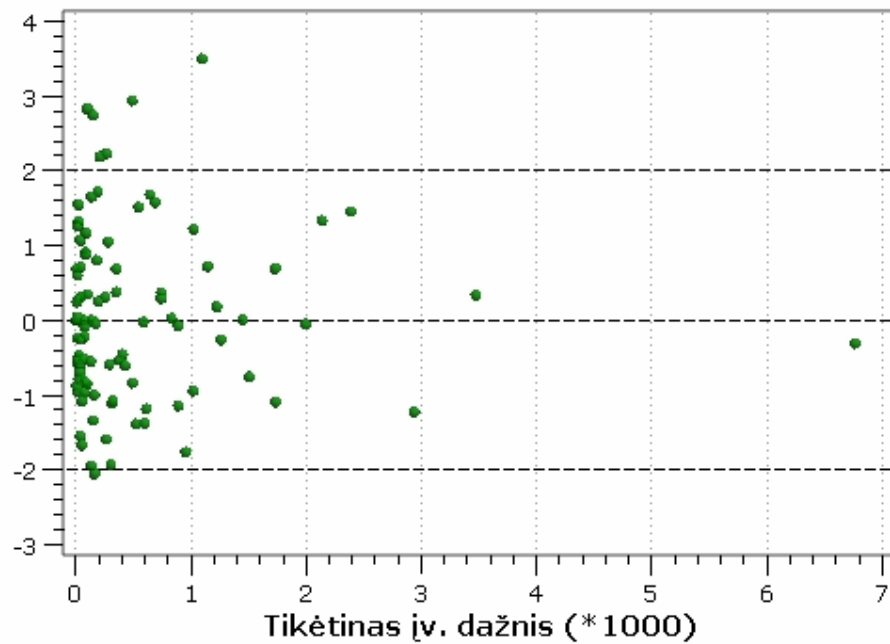
Pirsono paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|--------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.969815 | Pr < W | 0.0258 |
| Kolmogorov-Smirnov | D | 0.086611 | Pr > D | 0.0761 |
| Cramer-von Mises | W-Sq | 0.117068 | Pr > W-Sq | 0.0689 |
| Anderson-Darling | A-Sq | 0.73409 | Pr > A-Sq | 0.0550 |

6 lentelė

Frymano-Tjukio paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.992746 | Pr < W | 0.8869 |
| Kolmogorov-Smirnov | D | 0.058319 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.022873 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.158666 | Pr > A-Sq | >0.2500 |

Anomalioji paklaida**Modelio lytis-amžius-kohorta anomalijų paklaidų grafikas****Pirsono paklaida****Modelio lytis-amžius-kohorta Pirsono paklaidų grafikas**

6 priedas. Modelio lytis-amžius-kohortos dreifas taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius-kohortos dreifas rezultatai

| Lytis | Amžius | Kohorta | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautinojo intervalo viršutinis rėžis (*1000) | Anomalioji paklaida | Pirsono paklaida | Frymano- Tjukio paklaida |
|---------|--------|-----------|-----------------------------|-----------------------|------------------------------|------------------------------------|---|---|------------------------|---------------------|--------------------------------|
| Vyras | 60-64 | 1916-1920 | 26 | 4084 | 18.1895 | 4.45385 | 4.08401 | 4.85718 | 1.71933 | 1.83133 | 1.70692 |
| Vyras | 55-59 | 1921-1925 | 19 | 6560 | 16.0397 | 2.44508 | 2.21421 | 2.70001 | 0.71801 | 0.73916 | 0.75893 |
| Vyras | 60-64 | 1921-1925 | 64 | 25393 | 85.9051 | 3.38302 | 3.10210 | 3.68938 | -2.47631 | -2.36339 | -2.50169 |
| Vyras | 50-54 | 1926-1930 | 14 | 10108 | 12.5737 | 1.24393 | 1.09872 | 1.40834 | 0.39498 | 0.40225 | 0.45261 |
| Vyras | 55-59 | 1926-1930 | 60 | 39817 | 73.9486 | 1.85721 | 1.68185 | 2.05086 | -1.67754 | -1.62206 | -1.67151 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vyras | 30-34 | 1971-1975 | 2 | 54997 | 1.5397 | 0.02800 | 0.01987 | 0.03945 | 0.35448 | 0.37098 | 0.47069 |
| Vyras | 25-29 | 1976-1980 | 0 | 55240 | 0.4288 | 0.00776 | 0.00417 | 0.01444 | -0.92605 | -0.65482 | -0.64777 |
| Moteris | 60-64 | 1916-1920 | 25 | 6732 | 15.3444 | 2.27932 | 2.07770 | 2.50050 | 2.25723 | 2.46494 | 2.20108 |
| Moteris | 55-59 | 1921-1925 | 23 | 10271 | 12.8521 | 1.25130 | 1.12410 | 1.39290 | 2.54472 | 2.83066 | 2.45544 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Moteris | 30-34 | 1966-1970 | 4 | 86731 | 1.6359 | 0.01886 | 0.01334 | 0.02667 | 1.55709 | 1.84832 | 1.48949 |
| Moteris | 35-39 | 1966-1970 | 1 | 59910 | 1.4463 | 0.02414 | 0.01823 | 0.03197 | -0.39315 | -0.37108 | -0.19060 |
| Moteris | 25-29 | 1971-1975 | 1 | 84132 | 0.4400 | 0.00523 | 0.00281 | 0.00975 | 0.72248 | 0.84425 | 0.75290 |
| Moteris | 30-34 | 1971-1975 | 0 | 59823 | 0.8571 | 0.01433 | 0.01013 | 0.02026 | -1.30927 | -0.92579 | -1.10437 |
| Moteris | 25-29 | 1976-1980 | 0 | 58438 | 0.2321 | 0.00397 | 0.00213 | 0.00740 | -0.68138 | -0.48181 | -0.38873 |

2 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.983058 | Pr < W | 0.2522 |
| Kolmogorov-Smirnov | D | 0.046544 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.033875 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.294077 | Pr > A-Sq | >0.2500 |

3 lentelė

Pirsono paklaidų normalumo tikrinimas

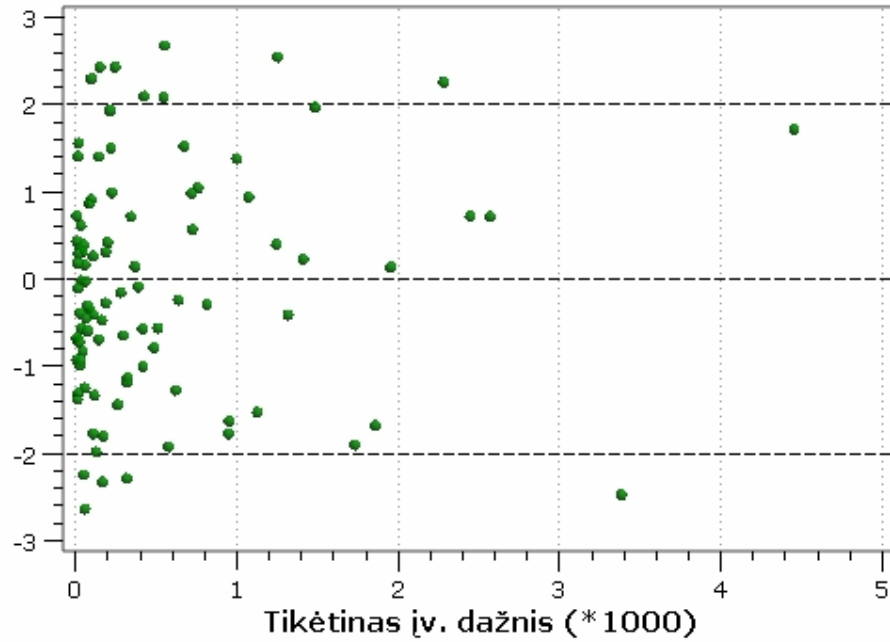
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.968899 | Pr < W | 0.0221 |
| Kolmogorov-Smirnov | D | 0.077967 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.108661 | Pr > W-Sq | 0.0881 |
| Anderson-Darling | A-Sq | 0.755763 | Pr > A-Sq | 0.0480 |

4 lentelė

Frymano-Tjukio paklaidų normalumo tikrinimas

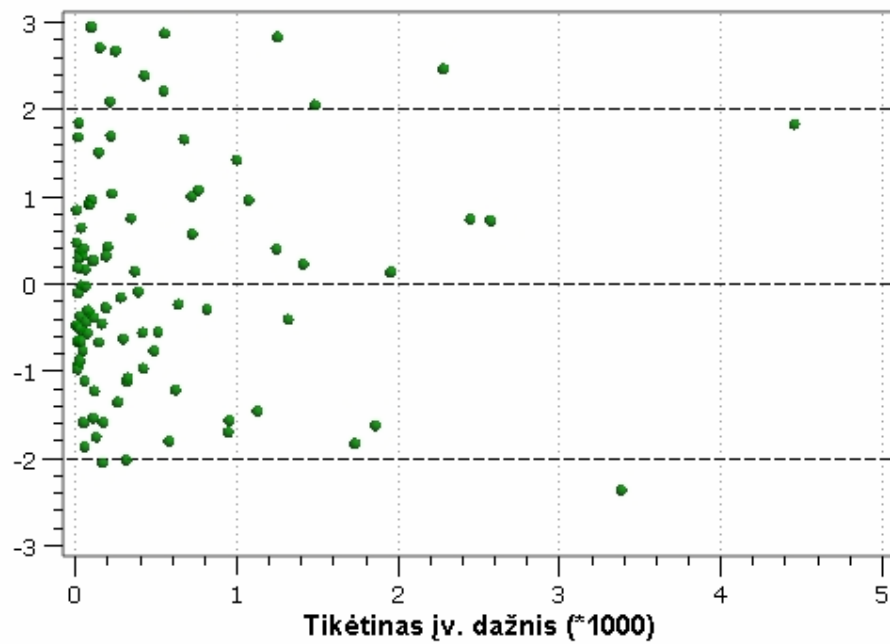
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.987751 | Pr < W | 0.5201 |
| Kolmogorov-Smirnov | D | 0.037751 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.024547 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.214658 | Pr > A-Sq | >0.2500 |

Anomalioji paklaida



Modelio lytis-amžius-kohortos dreifas anomalijų paklaidų grafikas

Pirsono paklaida



Modelio lytis-amžius-kohortos dreifas Pirsono paklaidų grafikas

7 priedas. Modelio lytis-amžius-periodas-kohorta taikymo rezultatai

1 lentelė

Puasono regresijos modelio lytis-amžius-periodas-kohorta rezultatai

| Lytis | Amžius | Periodas | Kohorta | Stebėtas įv. skaičius | Gyventojų skaičius | Tikėtinas įv. skaičius | Tikėtinas įv. dažnis (*1000) | Tikėtino įv. dažnio pasikliautiniojo intervalo apatinis rėžis (*1000) | Tikėtino įv. dažnio pasikliautiniojo intervalo viršutinis rėžis (*1000) | Anomalioji paklaida | Pirsono paklaida | Frymano- Tjukio paklaida |
|---------|--------|-----------|-----------|-----------------------------|-----------------------|------------------------------|------------------------------------|--|--|------------------------|---------------------|--------------------------------|
| Vyras | 25-29 | 1980-1984 | 1951-1955 | 0 | 45729.8 | 0.7573 | 0.01656 | 0.00796 | 0.03446 | -1.23071 | -0.87024 | -1.00731 |
| Vyras | 25-29 | 1980-1984 | 1956-1960 | 0 | 33660.2 | 0.5302 | 0.01575 | 0.00761 | 0.03259 | -1.02973 | -0.72813 | -0.76655 |
| Vyras | 30-34 | 1980-1984 | 1946-1950 | 1 | 40389.8 | 2.6821 | 0.06641 | 0.04248 | 0.10381 | -1.17942 | -1.02712 | -1.01048 |
| Vyras | 30-34 | 1980-1984 | 1951-1955 | 4 | 27568.2 | 1.4638 | 0.05310 | 0.03383 | 0.08335 | 1.72324 | 2.09620 | 1.61780 |
| Vyras | 35-39 | 1980-1984 | 1941-1945 | 3 | 39304.2 | 4.0377 | 0.10273 | 0.07223 | 0.14611 | -0.54132 | -0.51642 | -0.40930 |
| ... | ... | | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vyras | 60-64 | 2000-2004 | 1936-1940 | 46 | 23814.6 | 43.3532 | 1.82045 | 1.53561 | 2.15812 | 0.39800 | 0.40199 | 0.43144 |
| Vyras | 60-64 | 2000-2004 | 1941-1945 | 27 | 16019.4 | 23.6057 | 1.47357 | 1.22559 | 1.77172 | 0.68282 | 0.69863 | 0.71920 |
| Moteris | 25-29 | 1980-1984 | 1951-1955 | 0 | 50743.2 | 0.4338 | 0.00855 | 0.00410 | 0.01781 | -0.93143 | -0.65862 | -0.65382 |
| Moteris | 25-29 | 1980-1984 | 1956-1960 | 0 | 37221.8 | 0.3026 | 0.00813 | 0.00392 | 0.01685 | -0.77798 | -0.55012 | -0.48678 |
| ... | ... | | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Moteris | 50-54 | 2000-2004 | 1951-1955 | 7 | 25958.2 | 5.2741 | 0.20318 | 0.15196 | 0.27165 | 0.71529 | 0.75153 | 0.77351 |
| Moteris | 55-59 | 2000-2004 | 1941-1945 | 12 | 38940.6 | 17.5318 | 0.45022 | 0.37344 | 0.54279 | -1.40178 | -1.32116 | -1.36405 |
| Moteris | 55-59 | 2000-2004 | 1946-1950 | 11 | 23851.4 | 11.1198 | 0.46621 | 0.37354 | 0.58188 | -0.03600 | -0.03594 | 0.03689 |
| Moteris | 60-64 | 2000-2004 | 1936-1940 | 34 | 35611.4 | 33.4643 | 0.93971 | 0.78998 | 1.11782 | 0.09236 | 0.09261 | 0.13423 |
| Moteris | 60-64 | 2000-2004 | 1941-1945 | 18 | 24079.6 | 18.3161 | 0.76065 | 0.63078 | 0.91725 | -0.07408 | -0.07387 | -0.01615 |

2 lentelė

Modelio lytis-amžius-periodas-kohorta regresijos lygties parametru įverčiai

| Analysis Of Parameter Estimates | | | | | | | | |
|---------------------------------|-----------|----|----------|----------------|----------------------------|----------|------------|------------|
| Parameter | | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -25.2323 | 13084.33 | -25670.1 | 25619.59 | 0.00 | 0.9985 |
| LYTIS | Moteris | 1 | -0.6613 | 0.0485 | -0.7563 | -0.5663 | 186.14 | <.0001 |
| LYTIS | Vyras | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| AMZIUS | 25-29 | 1 | -4.3801 | 0.4915 | -5.3434 | -3.4168 | 79.42 | <.0001 |
| AMZIUS | 30-34 | 1 | -3.2150 | 0.3565 | -3.9137 | -2.5163 | 81.34 | <.0001 |
| AMZIUS | 35-39 | 1 | -2.7438 | 0.2893 | -3.3109 | -2.1767 | 89.92 | <.0001 |
| AMZIUS | 40-44 | 1 | -2.0249 | 0.2217 | -2.4594 | -1.5904 | 83.42 | <.0001 |
| AMZIUS | 45-49 | 1 | -1.6737 | 0.1711 | -2.0090 | -1.3384 | 95.71 | <.0001 |
| AMZIUS | 50-54 | 1 | -1.1314 | 0.1213 | -1.3691 | -0.8937 | 87.04 | <.0001 |
| AMZIUS | 55-59 | 1 | -0.5244 | 0.0781 | -0.6775 | -0.3714 | 45.11 | <.0001 |
| AMZIUS | 60-64 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| PERIODAS | 1980-1984 | 1 | 0.0804 | 0.2106 | -0.3324 | 0.4933 | 0.15 | 0.7026 |
| PERIODAS | 1985-1989 | 1 | 0.2153 | 0.1669 | -0.1117 | 0.5424 | 1.67 | 0.1968 |
| PERIODAS | 1990-1994 | 1 | 0.3913 | 0.1256 | 0.1450 | 0.6375 | 9.70 | 0.0018 |
| PERIODAS | 1995-1999 | 1 | 0.0339 | 0.0969 | -0.1560 | 0.2238 | 0.12 | 0.7262 |
| PERIODAS | 2000-2004 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| KOHORTA | 1916-1920 | 1 | 18.9334 | 13084.33 | -25625.9 | 25663.76 | 0.00 | 0.9988 |
| KOHORTA | 1921-1925 | 1 | 18.9592 | 13084.33 | -25625.9 | 25663.78 | 0.00 | 0.9988 |
| KOHORTA | 1926-1930 | 1 | 18.8789 | 13084.33 | -25625.9 | 25663.70 | 0.00 | 0.9988 |
| KOHORTA | 1931-1935 | 1 | 18.9148 | 13084.33 | -25625.9 | 25663.74 | 0.00 | 0.9988 |
| KOHORTA | 1936-1940 | 1 | 18.9236 | 13084.33 | -25625.9 | 25663.75 | 0.00 | 0.9988 |
| KOHORTA | 1941-1945 | 1 | 18.7122 | 13084.33 | -25626.1 | 25663.54 | 0.00 | 0.9989 |
| KOHORTA | 1946-1950 | 1 | 18.7471 | 13084.33 | -25626.1 | 25663.57 | 0.00 | 0.9989 |
| KOHORTA | 1951-1955 | 1 | 18.5235 | 13084.33 | -25626.3 | 25663.35 | 0.00 | 0.9989 |
| KOHORTA | 1956-1960 | 1 | 18.4734 | 13084.33 | -25626.3 | 25663.30 | 0.00 | 0.9989 |
| KOHORTA | 1961-1965 | 1 | 18.1573 | 13084.33 | -25626.7 | 25662.98 | 0.00 | 0.9989 |
| KOHORTA | 1966-1970 | 1 | 18.5606 | 13084.33 | -25626.3 | 25663.38 | 0.00 | 0.9989 |
| KOHORTA | 1971-1975 | 1 | 18.5411 | 13084.33 | -25626.3 | 25663.36 | 0.00 | 0.9989 |
| KOHORTA | 1976-1980 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

3 lentelė

Anomaliųjų paklaidų normalumo tikrinimas

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.993388 | Pr < W | 0.6810 |
| Kolmogorov-Smirnov | D | 0.038742 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.021196 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.178044 | Pr > A-Sq | >0.2500 |

4 lentelė

Pirsono paklaidų normalumo tikrinimas

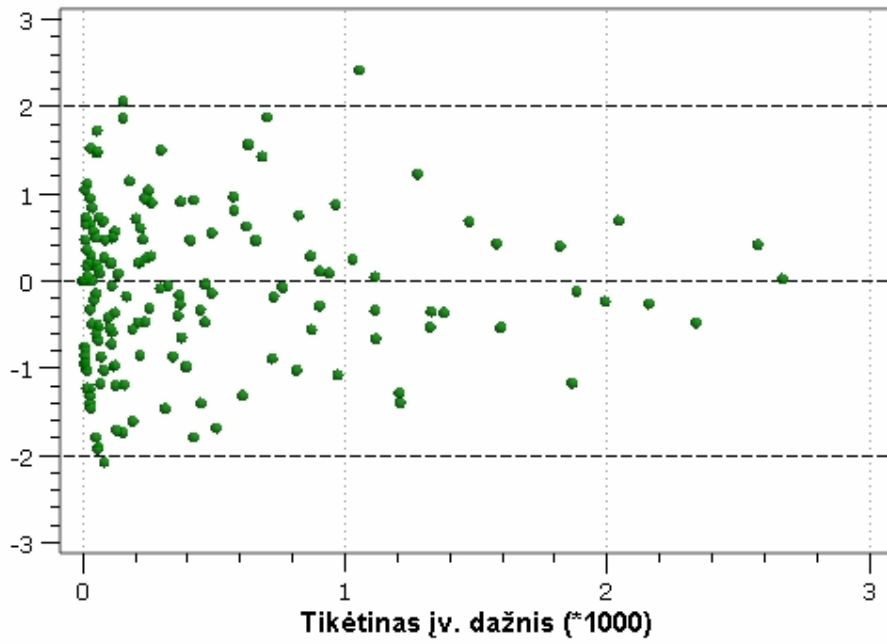
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|--------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.977483 | Pr < W | 0.0103 |
| Kolmogorov-Smirnov | D | 0.061112 | Pr > D | 0.1483 |
| Cramer-von Mises | W-Sq | 0.100535 | Pr > W-Sq | 0.1121 |
| Anderson-Darling | A-Sq | 0.715645 | Pr > A-Sq | 0.0636 |

5 lentelė

Frymano-Tjukio paklaidų normalumo tikrinimas

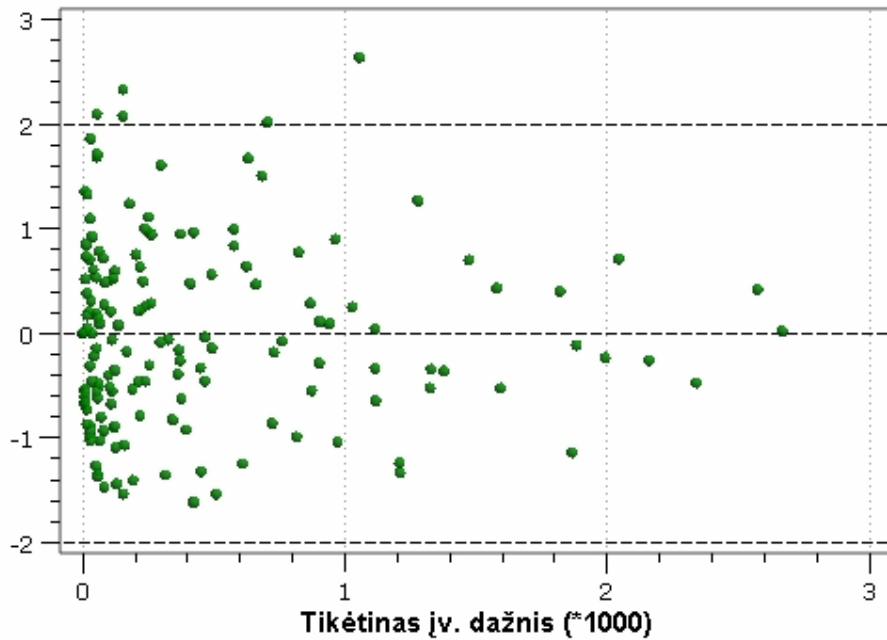
| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.99553 | Pr < W | 0.9132 |
| Kolmogorov-Smirnov | D | 0.03963 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.023264 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.163666 | Pr > A-Sq | >0.2500 |

Anomalių paklaida



Modelio lytis-amžius-periodas-kohorta anomalių paklaidų grafikas

Pirsono paklaida



Modelio lytis-amžius-periodas-kohorta Pirsono paklaidų grafikas

6 lentelė

Puasono regresijos modelių tinkamumo analizuotiems duomenims rezultatai

| Modelis | Nuokrypis | Laisvės laipsniai | Nuokrypis/Laisvės laipsniai |
|---|-----------|-------------------|-----------------------------|
| LYTIS AMZIUS | 7.080 | 7 | 1.01141 |
| LYTIS AMZIUS PERIODAS | 54.423 | 67 | 0.81229 |
| LYTIS AMZIUS su PERIODO dreifu ($\Delta = 0.025$) | 239.122 | 71 | 3.36792 |
| LYTIS AMZIUS KOHORTA | 129.501 | 75 | 1.72668 |
| LYTIS AMZIUS su KOHORTOS dreifu ($\Delta = -0.055$) | 159.063 | 87 | 1.82832 |
| LYTIS AMZIUS PERIODAS KOHORTA | 133.853 | 135 | 0.99150 |