

ŠIAULIŲ UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATIKOS KATEDRA

Justina Urbonavičiūtė

Gyventojų mirtingumo ir sveikatos apsaugos sistemos
tarpusavio sąryšio tyrimas

Bakalauro darbas

Darbo vadovė:
prof. dr. R. Kačinskaitė

ŠIAULIAI, 2013

TURINYS

ĮVADAS	4
I. TEORINĖ DALIS	6
1. PAGRINDINĖS EMPIRINĖS CHARAKTERISTIKOS	6
2. HIPOTEZĖS	12
2.1. NORMALUSIS SKIRSTINYS	12
2.2. HIPOTEZĖ APIE VIDURKIO LYGYBĘ SKAIČIUI, KAI DISPERSIJA YRA ŽINOMA.....	14
2.3. HIPOTEZĖ APIE VIDURKIŲ LYGYBĘ DVIEMS IMTIMS.....	15
2.4. NEPARAMETRINĖS HIPOTEZĖS	16
2.5. χ^2 KRITERIJUS. POŽYMIŲ NEPRIKL AUSOMUMAS	17
2.6. VILKOKSONO KRITERIJUS.....	19
3. KORELIACINĖ ANALIZĖ	21
3.1. KORELIACIJOS KOEFICIENTAS.....	21
3.2. HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTO LYGYBĘ NULIUI.....	23
4. TIESINĖ REGRESIJA	26
4.1. DETERMINACIJOS KOEFICIENTAS.....	28
4.2. DAUGIALYPĖ REGRESIJA.....	29
II. PRAKTINĖ DALIS	30
1. DUOMENYS.....	30
2. DUOMENŲ ANALIZĖ.....	31
2.1. PRADINĖ ANALIZĖ.....	31
2.2. DUOMENŲ NORMALIŠKUMO TIKRINIMAS	45
3. PARAMETRINIŲ HIPOTEZIŲ TIKRINIMAS.....	47
3.1. PARAMETRINĖS HIPOTEZĖS VIENAI IMČIAI.....	47
3.2. PARAMETRINĖS HIPOTEZĖS APIE DVIEJŲ IMČIŲ PARAMETRŲ LYGYBĘ	51

4. HIPOTEZĖ APIE SKIRSTINIŲ LYGUMĄ.....	59
5. NEPRIKLAUSOMUMO KRITERIJAUŠ TIKRINIMAS	61
6. HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTO LYGYBĘ NULIUI.....	64
7. REGRESINĖ ANALIZĖ	70
7.1. PAPRASTOJI TIESINĖ REGRESIJA	70
7.2. DAUGIALYPĖ REGRESIJA.....	75
8. PROGNOZAVIMAS LAIKO EILUTĖMIS	76
IŠVADOS	78
SUMMARY	79
LITERATŪRA	80
PRIEDAI.....	82

IVADAS

Viena iš aktualiausių šiandieninės Lietuvos problemų – gyventojų skaičiaus mažėjimas. Jis yra susijęs su emigracija, nedideliu gimstamumu, kasmet didėjančiu mirusių gyventojų skaičiumi bei kitomis priežastimis. Statistika yra viena iš priemonių, padedančių nustatyti šių reiškinių tarpusavio ryšius ir dėsningumus.

Yra nemažai mokslinių darbų, kuriuose nagrinėjami veiksniai, lemiantys gyventojų mirtingumo priežastis. D. Lukaševičiūtė nagrinėjo ligonių, sergančių reumatoidiniu artritu mirtingumą [10], M. Noreika pateikė mirtingumo nuo galvos smegenų insulto prognozavimo modelius [12]. Tačiau lieka neaišku, kaip ir kodėl Lietuvoje kasmet kinta gyventojų mirtingumas. Žurnalo „Filosofija. Sociologija“ straipsnyje *Lietuvos gyventojų mirtingumo lygio ir mirties priežasčių struktūros prieštaringi pokyčiai* [14] aptariami Lietuvos gyventojų mirtingumo pokyčių nukrypimai nuo teorinių modelių ir išsivysčiusių šalių mirtingumo raidos dėsningumų.

2010 m. portale *15min.lt* skelbiama, kad „Lietuvos gyventojų mirtingumas yra vienas didžiausių tarp Europos Sąjungos gyventojų, lietuviai taip pat yra tarp mažiausiai besirūpinančių sveika gyvensena europiečių“ [15]. Tačiau iš tiesų pastebima, kad vis daugiau žmonės rūpinasi savo mityba, gyvenimo kokybe, sveikata, bet dėl to jų gyvenimo trukmė neilgėja, mirtingumas Lietuvos statistikos departamento duomenimis taip pat stipriai nesikeičia. Kyla klausimas – galbūt dėl nemažėjančio mirtingumo kalti ne patys gyventojai, o valstybės dėmesys, skiriamas jų sveikatos apsaugai. Ši problema nagrinėjama bakalauro darbe.

Bakalauro darbo tema – *Gyventojų mirtingumo ir sveikatos apsaugos sistemos tarpusavio sąryšio tyrimas*.

Darbo tikslas – nustatyti ir išanalizuoti gydytojų (sveikatos priežiūros specialistų), išlaidų sveikatos priežiūrai, lovų skaičiaus ligoninėse, gyventojų gimstamumo ir mirtingumo tarpusavio ryšius ir tendencijas.

Darbo uždaviniai:

- Nustatyti ar egzistuoja koreliacija tarp mirusių gyventojų, gydytojų (sveikatos priežiūros specialistų), išlaidų sveikatos priežiūrai ir lovų skaičiaus ligoninėse bei nustatyti šias priklausomybes nusakančius regresijos modelius.
- Patikrinti įvairias hipotezes apie Lietuvos gyventojų gimstamumo ir mirtingumo kitimą.
- Prognozuoti mirusių gyventojų skaičių ateityje naudojantis laiko eilučių teorijos žiniomis.
- Apibendrinti gautus rezultatus.

Darbas sudarytas iš dviejų dalių – teorinės ir praktinės. Teorinė dalis paruošta naudojantis [2], [4]–[7], [9], [11], [13] literatūros šaltiniais. Šioje dalyje aprašoma, kaip randamos pagrindinės duomenų empirinės charakteristikos, kaip tikrinami kriterijai, hipotezės, koreliacijos koeficiento lygybė nuliui bei regresijos tiesės radimas.

Praktinėje darbo dalyje atliekami skaičiavimai remiantis teorine darbo dalimi, tikrinami kriterijai, hipotezės, ieškomas koreliacinis ryšys bei surandamos ryši nusakančios regresijos lygtys, nubrėžiamos tiesės. Darbe naudojami Lietuvos statistikos departamento duomenys (www.stat.gov.lt).

Darbo pabaigoje formuluojamos išvados, gautos tyrimo metu. Prieduose pateikiami duomenys ir testai, atlikti taikant kompiuterines programas.

Statistiniams skaičiavimams buvo naudotos kompiuterinės programos MS Excel, Mathcad ir SPSS.

I. TEORINĖ DALIS

1. PAGRINDINĖS EMPIRINĖS CHARAKTERISTIKOS

Tarkime, kad turime imtį $X_1, \dots, X_n$, kurioje $X_i \sim F(x)$, $i = 1, \dots, n$, t. y. X_i turi tą pačią pasiskirstymo funkciją. Mūsų tikslas – remiantis imtimi $X_1, \dots, X_n$, įvertinti pasiskirstymo funkciją $F(x)$, kuri nėra žinoma. Tai geriausia padaryti įvedant empirinio skirstinio apibrėžimą.

1.1 apibrėžimas. Empiriniu skirstiniu vadiname sąlyginį skirstinį diskretaus atsitiktinio dydžio X , įgyjančio imties reikšmes $X_1, \dots, X_n$, su tikimybe $\frac{1}{n}$, esant sąlygai, kad pasirodė $X_1, \dots, X_n$.

Toks apibrėžimas pakankamai logiškas, kadangi, neturėdami papildomos informacijos, laikome imties reikšmių $X_1, \dots, X_n$ pasirodymą vienodai galimu. Imties empirinį skirstinį patogų pavaizduoti skirstinio lentelę.

X	X_1	X_2	$\dots$	X_n
P	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

1.2 apibrėžimas. Pasiskirstymo funkcija, apibrėžta pagal empirinį skirstinį, vadinama empirine pasiskirstymo funkcija ir žymima $F_n(x)$. Jos pavidalas

$$F_n(x) = \sum_{\{k: X_k \leq x\}} \frac{1}{n} = \frac{n_x}{n}; \quad (1.1)$$

čia n_x yra skaičius imties $X_1, \dots, X_n$ elementų, mažesnių arba lygių x .

Tegu $(x_1, \dots, x_n)$ yra konkreči imties $(X_1, \dots, X_n)$ realizacija. Tada (1.1) formulėje vietoje imties įrašę jos realizaciją, gauname funkciją, žymimą

$$F_n(x) = \sum_{\{k: x_k \leq x\}} \frac{1}{n} = \frac{n_x}{n}. \quad (1.2)$$

Sakykime, kad turime imtį $(X_1, \dots, X_n)$ su $X_i \sim F(x)$. Ją atitinka empirinis skirstinys

X	X_1	X_2	$\dots$	X_n
P	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

1.3 apibrėžimas. Empirinio skirstinio skaitinės charakteristikos (vidurkis, dispersija, pradiniai ir antriniai momentai ir t. t.) vadinamos empirinėmis imties charakteristikomis.

Diskretaus atsitiktinio dydžio vidurkio formulė yra

$$MX = \sum_{i=1}^n x_i p_i, \quad (1.3)$$

jei dydis X duotas skirstinio lentele

X	X_1	X_2	$\dots$	X_n
P	p_1	p_2	$\dots$	p_n

Pagal empirinį skirstinį, į (1.3) formulę vietoj p_i įrašę $\frac{1}{n}$, o vietoj x_i dydį X_i , gauname empirinį vidurkį, kurį žymėsime $\bar{X}$:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.4)$$

Analogiškai, pritaikę diskretaus atsitiktinio dydžio dispersijos formulę

$$DX = \sum_{i=1}^n (x_i - MX)^2 p_i, \quad (1.5)$$

gauname empirinės dispersijos formulę:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2. \quad (1.6)$$

1.4 apibrėžimas. Empirine moda, žymima M_0 , vadinamas dažniausiai pasikartojantis imties $(X_1, \dots, X_n)$ elementas, jei X_i yra diskretus dydis. Kai X_i yra tolydus dydis, M_0 yra grupuotosios imties didžiausio dažnio intervalo vidurys.

1.5 apibrėžimas. Empirine mediana, žymima M_e , vadinamas tas imties elementas, kuris yra variacinės eilės viduryje, t. y.

$$M_e = \begin{cases} X_{(l+1)}, & \text{jei } n = 2l + 1, \\ \frac{1}{2}(X_{(l)} + X_{(l+1)}), & \text{jei } n = 2l. \end{cases} \quad (1.7)$$

Irašius konkrečią imties realizaciją $(x_1, \dots, x_n)$ į šias formules, gausime imties empirinių charakteristikų realizacijas, kurias žymėsime $\bar{x}, s^2, a_k$ (pradinis momentas), m_k (centrinis momentas), g_1 (asimetrijos koeficientas), g_2 (aksceso koeficientas), m_0, m_e .

Imties elementus surašius didėjimo tvarka

$$X_1 \leq X_2 \leq \dots \leq X_n \quad (1.8)$$

gauname seką, kurią vadiname variacine imties eilute, o jos elementai $X_{(k)} (k = \overline{1, n})$ bus vadinami pozicinėmis statistikomis. Kraštiniai variacinės eilutės nariai (ekstremaliosios reikšmės) žymimos šitaip

$$X_{(1)} = \min(X_1, X_2, \dots, X_n), \quad X_{(n)} = \max(X_1, X_2, \dots, X_n). \quad (1.9)$$

Šie nariai apibrėžia imties plotį $X_{(n)} - X_{(1)}$. Ši statistika apibūdina imties sklaidą. Sakyme, imtyje $(X_1, \dots, X_n)$ yra tik $s (s \leq n)$ skirtingų elementų $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_s$. Jų pasikartojimo

dažnius pažymėkime $k_1, k_2, \dots, k_s$, o santykinius dažnius – $W_1 = \frac{k_1}{n}, \dots, W_s = \frac{k_s}{n}$. Kai imties elementai nepasikartoja ($s = n$), tai $W_1 = W_2 = \dots = W_n = \frac{1}{n}$. Empiriniu skirstiniu vadiname šią reikšmių $(\tilde{X}_i, W_i)$, kai $i = \overline{1, s}$, aibę:

$\tilde{X}_i$	$\tilde{X}_1$	$\tilde{X}_2$	...	$\tilde{X}_n$
W_i	W_1	W_2		W_s

Žinome, kad $W_1 + W_2 + \dots + W_s = 1$.

Teorinės pasiskirstymo funkcijos $F(x)$ statistinis įvertis yra empirinė pasiskirstymo funkcija. Empirine pasiskirstymo funkcija $\hat{F}_n$ vadiname įvykio $(X < x)$ santykinį dažnį

$$\hat{F}_n(x) = \frac{k_x}{n} = \begin{cases} 0, & \text{kai } x \leq \tilde{X}_{(1)}, \\ \sum_{x_i < x} W_i, & \text{kai } \tilde{X}_{(i)} < x \leq \tilde{X}_{(i+1)}, \\ 1, & \text{kai } x > \tilde{X}_{(s)}. \end{cases} \quad (1.10)$$

čia k_x – atsitiktinės imties dydžių X_i , mažesnių už x , dažnis; $k_x = |\{i : X_i < x\}|$.

Jeigu imties elementai nesikartoja ($s = n$), tai $W_1 = W_2 = \dots = W_n = \frac{1}{n}$ ir empirinė pasiskirstymo funkcija

$$\tilde{F}_n(x) = \frac{k}{n}, \text{ kai } X_{(k)} < x \leq X_{(k+1)}, \quad k = \overline{1, n-1}, \quad (1.11)$$

o k yra įvykio $(X < x)$ dažnis imtyje. Šiuo atveju funkcijos $\hat{F}_n(x)$ šuolių dydis yra $\frac{1}{n}$.

Empirinė pasiskirstymo funkcija $\hat{F}_n(x)$ yra statistika. Dar daugiau – tai atsitiktinis procesas. Šiai funkcijai būdingos visos diskrečiojo atsitiktinio dydžio teorinės pasiskirstymo

funkcijos savybės. Būtina pabrėžti, kad funkcija $\hat{F}_n(x)$ yra ne tikimybė, bet įvykio $(X < x)$ santykinis dažnis (atsitiktinių dydžių X_i , mažesnių už x , skaičiaus bei imties tūrio santykis).

Dažnio k skirstinys yra binominis, o jo parametrai – n ir $p = P(X < x) = F(x)$. Todėl statistikos $\hat{F}_n(x)$ vidurkis

$$M\hat{F}_n(x) = \frac{np}{n} = F(x), \quad (1.12)$$

o dispersija

$$D\hat{F}_n(x) = \frac{np(1-p)}{n^2} = \frac{F(x)(1-F(x))}{n}. \quad (1.13)$$

Be to, iš Bernulio didžiųjų skaičių dėsnio išplaukia, kad $\hat{F}_n(x) \xrightarrow{p} F(x)$, kai $n \rightarrow \infty$. Vadinasi, kai imties tūris n didelis, empirinė pasiskirstymo funkcija $\hat{F}_n(x)$ mažai skiriasi nuo teorinės $F(x)$. Taigi galima teigti, kad $\hat{F}_n(x)$ yra “kokybiškas” teorinės pasiskirstymo funkcijos $F(x)$ įvertis.

Praktikoje didelė imtis iš pradžių grupuojama ir tik po to skaičiuojami empiriniai skirstiniai bei kitos charakteristikos. Be to, grupuota imtimi apibūdinamas teorinio tankio statistinis analogas. Šiuo atveju visą dydžio X imtį suskaidome į s vienodo ilgio intervalų taškais

$$X_{(1)} = \alpha_0 < \alpha_1 < \alpha_2 = X_{(n)}. \quad (1.14)$$

Tada kiekvieno intervalo ilgis $\Delta = \frac{X_{(n)} - X_{(1)}}{s}$. Kai imties tūris yra šimtų eilės, intervalų skaičius s paprastai būna apibūdinamas nelygybėmis $6 \leq s \leq 20$. Dar siūloma formulė $s = [1 + 3,32 \lg n]$. Sakykime, į intervalus $[\alpha_0, \alpha_1), [\alpha_1, \alpha_2), \dots, [\alpha_{s-1}, \alpha_s]$ pateko atitinkamai $k_1, k_2, \dots, k_s$ imties reikšmių. Empirinę grupuotą skirstinį užrašome lentelė

Intervalai	$[\alpha_0, \alpha_1)$	$[\alpha_1, \alpha_2)$	...	$[\alpha_{s-1}, \alpha_s]$
Vidurio taškai	$\tilde{X}_1$	$\tilde{X}_2$	...	$\tilde{X}_s$
Santykiniai dažniai	$\frac{k_1}{n}$	$\frac{k_2}{n}$	...	$\frac{k_s}{n}$

Kaip ir anksčiau, empirinė pasiskirstymo funkcija

$$\hat{F}_n(x) = W(X < x) = \frac{k}{n} = \frac{k_1 + k_2 + \dots + k_i}{n}; \text{ čia } \alpha_1 < x \leq \alpha_{i+1}, i = \overline{1, s-1}. \quad (1.15)$$

Šis skyrius paruoštas naudojantis [7] literatūros šaltiniu.

2. HIPOTEZĖS

Turint tam tikrus duomenis, suformuluojama hipotezė apie tiriamą savybę (arba hipotezė ir jai alternatyva) ir, pasinaudojant vienu ar kitu hipotezių priėmimo statistiniu kriterijumi, hipotezė priimama arba atmetama (arba atmetama hipotezė ir priimama alternatyva). Hipotezių priėmimas ir atmetimas atliekamas pagal tam tikrą schemą. Galimi priėmimo ir atmetimo atvejai:

- hipotezė teisinga ir, remiantis hipotezių priėmimo statistiniu kriterijumi, hipotezė priimama;
- hipotezė teisinga ir, remiantis hipotezių priėmimo statistiniu kriterijumi, hipotezė atmetama (šis atvejis įvardijamas kaip pirmos rūšies klaida);
- hipotezė neteisinga ir, remiantis hipotezių priėmimo statistiniu kriterijumi, hipotezė priimama (šis atvejis įvardijamas kaip antros rūšies klaida).

2.1. NORMALUSIS SKIRSTINYS

Atsitiktinis dydis yra kintamasis, kurio reikšmės prikauso nuo atsitiktinumo. Vieni atsitiktiniai dydžiai gali įgyti tik izoliuotas reikšmes – tai diskretieji atsitiktiniai dydžiai. Kiti gali įgyti bet kurią reikšmę iš intervalo ar net bet kurį skaičių iš visos skaičių tiesės – tai tolydieji atsitiktiniai dydžiai.

Funkcijas vieną nuo kitos atskiriame pagal tai, kokie veiksmai atliekami su argumentu. Atsitiktinį dydį X apibrėžiame tikimybės tankio funkcija. Tai yra tokia funkcija $f(x)$, kurios grafikas, Ox ašis ir taške x iškelta ordinatė aprėžia kairėje nuo ordinačės plotą, kuris lygus tikimybei $F(x) = P\{X < x\}$.

Funkcija $F(x)$ vadinama atsitiktinio dydžio X skirstinio funkcija. Joje yra visa tikimybė informacija apie atsitiktinį dydį.

Tolydusis atsitiktinis dydis X , kurio tikimybės tankio funkcija yra

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \sigma > 0, \quad (2.1)$$

vadinamas normaliuoju.

Tai yra dažniausiai pasitaikantis atsitiktinis dydis. Parametrai a ir σ turi tokią prasmę:

$$EX = a, DX = \sigma^2.$$

Dydis σ vadinamas standartiniu nuokrypiu.

Tikimybės X reikšmei patekti į intervalus $(a - \sigma, a + \sigma)$, $(a - 2\sigma, a + 2\sigma)$ ir $(a - 3\sigma, a + 3\sigma)$ yra tokios:

$$P\{a - \sigma < X < a + \sigma\} \approx 0,683 = 68,3\%,$$

$$P\{a - 2\sigma < X < a + 2\sigma\} \approx 0,955 = 95,5\%,$$

$$P\{a - 3\sigma < X < a + 3\sigma\} \approx 0,997 = 99,7\%.$$

Kadangi tankio grafikas yra simetriškas ordinatės, išvestos iš vidurkio taško, atžvilgiu, pusė viso ploto yra kairėje nuo tos ordinatės. Todėl vidurkis yra 50% kvantilis. Nesunku pastebėti, kad $a - \sigma$ yra

$$(100 - 68,3) / 2 \approx 15,9(\%) \text{ kvantilis.}$$

Statistikoje yra svarbi tokia lygybė :

$$P\{a - t_{\frac{1+Q}{2}}\sigma < X < a + t_{\frac{1+Q}{2}}\sigma\} = Q. \quad (2.2)$$

Pagal tikimybės Q dydį galime rasti $t_{\frac{1+Q}{2}}$. Dažniausiai taikomos tokios reikšmės:

Q	0,65	0,9	0,95	0,99
$t_{\frac{1+Q}{2}}$	0,935	1,645	1,960	2,576

Iš čia pagal priešingo atsitiktinio įvykio tikimybės formulę gauname:

$$P\{(X \leq a - t_{\frac{1+Q}{2}}\sigma) \cup (X \geq a + t_{\frac{1+Q}{2}}\sigma)\} = 1 - Q. \quad (2.3)$$

Šią formulę žodžiais galime nusakyti šitaip: tikimybė, kad atsitiktinis dydis X įgis reikšmę, ne didesnę už $a - t_{\frac{1+Q}{2}}\sigma$ arba ne mažesnę už $a + t_{\frac{1+Q}{2}}\sigma$, lygi $1 - Q$.

2.2. HIPOTEZĖ APIE VIDURKIO LYGYBĘ SKAIČIUI, KAI DISPERSIJA YRA ŽINOMA

Tarkime, kad stebime normalųjį atsitiktinį dydį X , kurio vidurkis lygus μ , o dispersija σ^2 , t. y. $X \sim N(\mu, \sigma^2)$. Populiacijos dispersija σ^2 žinoma, o vidurkis μ – nežinomas. Norime patikrinti hipotezę apie vidurkio lygybę skaičiui, t. y.

$$H_0 : \mu = a. \quad (2.4)$$

Sprendimui priimti su tam tikru reikšmingumo lygmeniu α turime parinkti tinkamą statistiką ir sukonstruoti kritinę sritį. Kritinei sričiai sudaryti naudojamas paprasčiausias vidurkio įvertis – statistika $\bar{X}$. Kritinė sritis (t. y. tikimybė atmesti hipotezę, kai ji yra teisinga) sudaroma remiantis tuo, kad normuotas atsitiktinis dydis

$$Z = \frac{\bar{X} - a}{\frac{\sigma}{\sqrt{n}}} \quad (2.5)$$

turi standartinį normalųjį skirstinį su parametrais $\mu = 0$, $\sigma^2 = 1$.

Norėdami priimti sprendimą apie hipotezės teisingumą turėsime apskaičiuoti kriterijaus statistikos reikšmę

$$Z = \frac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}}. \quad (2.6)$$

Sprendimus apie hipotezę su skirtingomis alternatyvomis priimame pagal tokią lentelę:

Alternatyva H_1	H_0 atmetama, jei	H_0 neatmetama, jei
$\mu \neq a$	$ Z > z_{\frac{\alpha}{2}}$	$ Z \leq z_{\frac{\alpha}{2}}$
$\mu > a$	$Z > z_{\alpha}$	$Z \leq z_{\alpha}$
$\mu < a$	$Z < -z_{\alpha}$	$Z \geq -z_{\alpha}$

Lentelėje z_{α} yra standartinio normaliojo skirstinio α lygmens kritinė reikšmė.

2.3. HIPOTEZĖ APIE VIDURKIŲ LYGYBĘ DVIEMS IMTIMS

Turime susijusius duomenis – poras.

Tarkime, kad intervalinių duomenų poros $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ gautos matuojant du priklausomuosius atsitiktinius dydžius $X \sim N(\mu_1, \sigma_1^2)$ ir $Y \sim N(\mu_2, \sigma_2^2)$, kurių vidurkliai μ_1, μ_2 ir dispersijos – nežinomi. Tikriname hipotezę

$$H_0 : \mu_1 = \mu_2. \quad (2.7)$$

Kriterijaus statistikos T , kuri turi Stjudento skirstinį, reikšmę t skaičiuojame šitaip:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_d^2}{n}}} = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}, \quad (2.8)$$

čia $\bar{d} = \frac{d_1 + d_2 + \dots + d_n}{n}$, $s_d^2 = \frac{d_1^2 + d_2^2 + \dots + d_n^2 - n\bar{d}^2}{n-1}$, $d_i = x_i - y_i$, $i = 1, 2, \dots, n$.

Sprendimai apie hipotezę su skirtingomis alternatyvomis priimami remiantis lentele ($t_\alpha(n-1)$ yra Stjudento skirstinio su $(n-1)$ laisvės laipsnių α lygmens kritinė reikšmė):

Alternatyva H_1	H_0 atmetama, jei	H_0 neatmetama, jei
$\mu_1 \neq \mu_2$	$ t > t_{\frac{\alpha}{2}}(n-1)$	$ t \leq t_{\frac{\alpha}{2}}(n-1)$
$\mu_1 > \mu_2$	$t > t_\alpha(n-1)$	$t \leq t_\alpha(n-1)$
$\mu_1 < \mu_2$	$t < -t_\alpha(n-1)$	$t \geq -t_\alpha(n-1)$

2.4. NEPARAMETRINĖS HIPOTEZĖS

Neparametrinė hipotezė vadinsime prielaidą apie stebimo atsitiktinio dydžio X skirstinio funkcijos $F(x)$ analizinę išraišką. Tokia hipotezė formuluojama šitaip:

$$H_0 : F(x) \equiv F_0(x). \quad (2.9)$$

Jei $F_0(x)$ yra visiškai nusakyta skirstinio funkcija, tai tokią hipotezę vadinsime paprastąja suderinamumo hipoteze. Reikia pažymėti, kad praktikoje skirstinio funkcija $F_0(x)$ labai retai būna visiškai nusakyta. Paprastai ji priklauso nuo nežinomų parametrų $\theta_1, \dots, \theta_s$. Jei $F_0(x)$ nėra visiškai nusakyta, tai tokią hipotezę vadinsime sudėtingąja suderinamumo hipoteze.

Tarkime, kad turime imtį $X^n = (X_1, \dots, X_n)$. Reikia patikrinti hipotezę $H_0 : X_i \sim F(x)$, kuri teigia, kad stebimi dydžiai X_i turi pasiskirstymo funkciją $F(x)$, kuri nusako tam tikrą skirstinių klasę, pavyzdžiui, normaliųjų skirstinių ir t. t. Čia nėra svarbūs galimi pasiskirstymo funkcijos parametrai. Svarbu tik žinoti, ar X_i pasiskirstymo funkcija yra iš tos skirstinių klasės, ar ne. Tokios hipotezės vadinamos neparametrinėmis, o kriterijai, skirti joms tikrinti, vadinami suderinamumo kriterijais. Svarbiausi jų – χ^2 kriterijus požymių nepriklausomumui tikrinti ir Vilkoksono kriterijus.

2.5. χ^2 KRITERIJUS. POŽYMIŲ NEPRIKLAUSOMUMAS

Tarkime, kad turime porinių stebėjimų imtį $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, gautą stebint kategorinių kintamųjų porą (X, Y) . Pirmiausia duomenų aibę užrašykime porine dažnių lentele:

	y_1	y_2	$\dots$	y_c	Σ
x_1	o_{11}	o_{12}	$\dots$	o_{1c}	$n_{1\cdot}$
x_2	o_{21}	o_{22}	$\dots$	o_{2c}	$n_{2\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_r	o_{r1}	o_{r2}	$\dots$	o_{rc}	$n_{r\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	$\dots$	$n_{\cdot c}$	n

Tarkime, kad $p_{ij} = P(X = x_i, Y = y_j)$ – populiacijos dalis, kuriai matuojamų požymių pora (X, Y) įgyja reikšmę (x_i, y_j) ; $p_i = P(X = x_i)$ – populiacijos dalis, kuriai požymis X įgyja reikšmę x_i ; $q_j = P(Y = y_j)$ – populiacijos dalis, kuriai požymis Y įgyja reikšmę y_j . Aišku, kad

$$\sum_{i=1}^r p_i = \sum_{j=1}^c q_j = 1. \quad (2.10)$$

Iš tikimybių teorijos žinoma, kad kintamieji nepriklausomi, jei $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ su visais $i = 1, \dots, r$, $j = 1, \dots, c$. Taigi reikia patikrinti statistinę hipotezę

$$\begin{cases} H_0 : p_{ij} = p_i q_j, \text{ su visais } i = 1, \dots, r, j = 1, \dots, c ; \\ H_1 : p_{ij} \neq p_i q_j, \text{ bent vienai porai } (i, j). \end{cases} \quad (2.11)$$

Jei hipotezė teisinga, tai p_{ij} yra nežinomų parametrų p_i, q_j , $i = 1, \dots, r$, $j = 1, \dots, c$, funkcijos. Atsižvelgiant į tai, kad kiekvieną įvykį apibūdina du indeksai i ir j , kriterijaus statistiką galima užrašyti taip:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - n\hat{p}_{ij})^2}{\hat{p}_{ij}}. \quad (2.12)$$

Nežinomų parametrų įverčiai, gauti χ^2 minimumo metodu, yra:

$$\hat{p}_i = \frac{n_{i.}}{n}, i = 1, \dots, r; \quad \hat{q}_j = \frac{n_{.j}}{n}, j = 1, \dots, c. \quad (2.13)$$

Tikėtinieji dažniai E_{ij} apskaičiuojami pagal formulę

$$E_{ij} = n\hat{p}_{ij} = \hat{p}_i \cdot \hat{q}_j = \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} = \frac{n_{i.}n_{.j}}{n}. \quad (2.14)$$

Iš čia išplaukia, kad esant teisingai hipotezei H_0 statistika

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n} = n \left(\sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij}^2}{n_{i.}n_{.j}} - 1 \right) \quad (2.15)$$

yra asimptotiškai pasiskirsčiusi pagal χ^2 dėsnį su

$$rc - [(r-1) + (c-1)] - 1 = rc - r - c + 1 = (r-1)(c-1)$$

laisvės laipsnių. Hipotezė H_0 apie kintamųjų nepriklausomumą yra atmetama, kai apskaičiuotos statistikos χ^2 reikšmė yra didesnė už χ^2 skirstinio su $(r-1)(c-1)$ laisvės laipsnių α lygmens kritinę reikšmę (α – pasirinktas reikšmingumo lygmuo).

2.6. VILKOKSONO KRITERIJUS

Vilkoksono kriterijus taikomas neparametrinių ir nepriklausomų imčių lyginimo testams. Šis testas veiksmingiausias, kai kintamųjų skirstiniai skiriasi tik postūmio parametru (t. y. X skirstinys sutampa su $Y + a$ skirstiniu). Testas tikrina, ar dviejų imčių persiklojimas nėra mažesnis, negu turėtų būti tuo atveju, kai abi imtys visiškai atsitiktinai paimtos iš tos pačios generalinės aibės.

Mažų imčių ($n \leq 25$) atveju stebima dviejų tolydžiųjų nepriklausomų kintamųjų X ir Y pora. Duomenys $(x_1, x_2, \dots, x_n)$ ir $(y_1, y_2, \dots, y_n)$ yra gauti matavimams naudojant santykių, rangų arba intervalų matavimų skales. Abi imtys sujungiamos į vieną išdėstant didėjančia tvarka (sudaroma bendra variacinė eilutė). Nariams priskiriami rangai ir apskaičiuojamos statistikos:

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1, \quad U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2; \quad (2.16)$$

čia R_1 ir R_2 – pirmosios ir antrosios imčių rangų sumos.

Statistika U_1 rodo, kiek pirmos imties narių yra kairiau už kiekvieną antros imties narį. Kai $U_1 = 0$, tai visi pirmos imties nariai mažesni už antrosios imties narius. Didelės ar mažos

U_1 reikšmės kelia abejonių, ar stebimų požymių skirstiniai vienodi. Žinant statistikas U_1 ir U_2 priimamas sprendimas, ar imčių skirtumas yra patikimas.

Didelių imčių atveju (t. y. kai $n > 25$) kriterijaus statistika yra apskaičiuojama pagal formulę

$$Z = \frac{U_1 - \mu}{\sigma}; \quad (2.17)$$

$$\text{čia } \mu = \frac{n_1 \cdot n_2}{2}, \quad \sigma = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}.$$

Su reikšmingumo lygmeniu α , kai $|Z| > z_{\alpha/2}$ ($z_{\alpha/2}$ – standartinio normaliojo skirstinio $\alpha/2$ lygmens kritinė reikšmė), tai hipotezę H_0 (skirstiniai statistiškai reikšmingai nesiskiria) atmetame ir priimame hipotezę H_1 (skirstiniai statistiškai reikšmingai skiriasi). Tai reiškia, kad skirstiniai skiriasi statistiškai reikšmingai.

Šio skyriaus teorinė medžiaga paruošta naudojantis [2], [4], [5], [9] ir [11] literatūros šaltiniais.

3. KORELIACINĖ ANALIZĖ

3.1. KORELIACIJOS KOEFICIENTAS

Du atsitiktiniai dydžiai X ir Y vadinami susietais statistiškai, jei kiekvienai vieno dydžio reikšmei atitinka apibrėžtas kito dydžio skirstinys. Šių dydžių statistinė priklausomybė vadinama koreliacine priklausomybe, kai, kintant vienam dydžiui, kinta kito dydžio vidutinė reikšmė. Kai du kintamieji koreliuoja – didėjant vieno kintamojo reikšmei, didėja kito kintamojo reikšmė – toks jų tarpusavio ryšys vadinamas teigiamąja koreliacija. Grafiškai šių dviejų kintamųjų tarpusavio priklausomybė gali būti pavaizduota kaip visuma taškų, išsidėsčiusių netoli tiesės, kylančios aukštyn iš kairės į dešinę. Toks grafikas vadinamas taškine diagrama. Jei didėjant vieno kintamojo reikšmei kito kintamojo reikšmė mažėja, tai sakoma, kad šie kintamieji yra neigiamai koreliuoti. Šiuos du kintamuosius siejančiame grafike matytume, kad taškai diagramoje išsidėstę netoli tiesės, kuri leidžiasi žemyn iš kairės į dešinę. Jeigu kintamųjų pora nėra koreliuota, tuomet tokiu atveju taškine diagrama atrodytų kaip taškų visuma, kurioje negalima įžvelgti taškų išsidėstymo netoli tiesės tendencijos.

Esant pakankamai stipriam koreliaciniam ryšiui, galima daryti išvadas apie vieno iš kintamųjų reikšmes, kai žinomos kito kintamojo reikšmės, t. y. numatyti būsimus įvykius, juos prognozuoti. Daugiau apie įvykių prognozavimą – tiesinės regresijos skyrelyje.

Tiesinė kintamųjų priklausomybė yra paprasčiausia funkcinės priklausomybės forma. Be tiesinės priklausomybės dar yra ir sudėtingesnių sąryšio formų – hiperbolinė priklausomybė, logaritminė priklausomybė, parabolinė priklausomybė ir eksponentinė priklausomybė.

Koreliacinės analizės tikslas – nustatyti stochastinio ryšio tarp veiksmų X ir Y egzistavimą. Tai daroma pagal turimus statistinius duomenis skaičiuojant koreliacijos koeficientą ir įvertinant jo reikšmingumą. Jei koreliacijos koeficiento dydis yra reikšmingas,

tai daroma išvada apie stochastinio ryšio egzistavimą. Prielaidos koreliacinei analizei atlikti yra tokios:

- abu veiksniai X ir Y yra atsitiktiniai dydžiai;
- abu atsitiktiniai dydžiai X ir Y yra pasiskirstę pagal normalųjį dėsnį.

Sakykime, kad turime dvimatę atsitiktinę imtį $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, kurios elementai (X_i, Y_i) , turintys tą patį dvimatį skirstinį $P_{\theta}^{(2)}$, $\theta \in \Theta$, paimti iš nepriklausomų dvimačių atsitiktinių vektorių sekos $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n), \dots$. Šią imtį pirmiausia charakterizuoja išsibarstymo diagrama ir dvimatis empirinis skirstinys.

3.1 apibrėžimas. Dvimatės imties išsisklaidymo diagrama vadiname imties taškų (X_i, Y_i) , $i = 1, \dots, n$, atidėtų stačiakampėje koordinačių sistemoje, visumą.

3.2 apibrėžimas. Dvimatės imties empiriniu skirstiniu vadiname sąlyginį skirstinį diskretaus atsitiktinio vektoriaus (X, Y) , įgyjančio reikšmes (X_i, Y_i) , $i = 1, 2, \dots, n$, su tikimybe $\frac{1}{n}$, esant sąlygai, kad pasirodė $(X_1, Y_1), \dots, (X_n, Y_n)$.

To vektoriaus skirstinio skaitinės charakteristikos vadinamos dvimatės imties empirinėmis charakteristikomis.

Vadinasi, teorinio dvimatės atsitiktinės imties $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ koreliacijos koeficiento

$$\rho(X, Y) = \frac{M(X_i - MX_i)(Y_i - MY_i)}{\sqrt{DX_i DY_i}} \quad (3.1)$$

empirinis analogas yra empirinis koreliacijos koeficientas, apskaičiuojamas pagal formulę

$$R = \frac{m_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} ; \quad (3.2)$$

čia

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad (3.3)$$

$$S_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad S_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (3.4)$$

$$m_{xy} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}). \quad (3.5)$$

Koreliacijos koeficientas ρ apibūdina tiesinę priklausomybę tarp dydžių X_i ir Y_i . Kadangi teorinio (tikrojo) koreliacijos koeficiento ρ tarp X_i ir Y_i nežinome, tai vietoje jo naudojame jo įvertį – empirinį koreliacijos koeficientą R .

3.2. HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTO LYGYBĘ NULIUI

Tarkime, kad stebime intervalinių kintamųjų porą (X, Y) , gautą matuojant dvimatį normalųjį atsitiktinį dydį. Atsitiktinę imtį sudaro poros $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Norime nustatyti, ar kintamieji X ir Y koreliuoja. Kadangi iš imties negalime spręsti apie tikrąjį koreliacijos koeficientą, todėl lieka neaišku, ar koreliacija statistiškai reikšmingai skiriasi nuo nulio.

Konstruojant kritines sritis remiamasi tuo, kad

$$T = R \sqrt{\frac{n-2}{1-R^2}} \quad (3.6)$$

turi Stjudento skirstinį su $(n - 2)$ laisvės laipsnių, jeigu $\rho = 0$.

Pirmiausia, matuojant dvimatį atsitiktinį dydį (X, Y) , gauname intervalinių duomenų porinę imtį $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$. Tuomet iškeliami statistinė hipotezė:

$$\begin{cases} H_0 : \rho = 0, \\ H_1 : \rho \neq 0. \end{cases} \quad (3.7)$$

Randomame kriterijaus statistiką:

$$T = R \sqrt{\frac{n-2}{1-R^2}}, \quad (3.8)$$

čia R yra koreliacijos koeficiento realizacija, skaičiuojama pagal formulę

$$R = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}. \quad (3.9)$$

Tuomet, pagal toliau išdėstytas sąlygas priimamas arba atmetamas sprendimas. Tarkime, kad reikšmingumo lygmuo yra α . Hipotezė H_0 atmetama (X ir Y statistiškai reikšmingai koreliuoja), jeigu $|T| > t_{\alpha/2}(n-2)$. Čia $t_{\alpha/2}(n-2)$ yra Stjudento skirstinio su $(n-2)$ laisvės laipsnių $\alpha/2$ lygmens kritinė reikšmė. Hipotezė H_0 neatmetama, jeigu $|T| \leq t_{\alpha/2}(n-2)$.

Vienpusėms alternatyvoms naudojama ta pati statistika T , apibrėžiama (3.6) formule. Sprendimo taisyklės, esant skirtingoms alternatyvoms, pateikiamos lentelėje (kai $H_0 : \rho = 0$):

Alternatyva H_1	H_0 atmetama, jeigu	H_0 neatmetama, jeigu
$\rho \neq 0$	$ T > t_{\alpha/2}(n-2)$	$ T \leq t_{\alpha/2}(n-2)$
$\rho > 0$	$T > t_{\alpha}(n-2)$	$T \leq t_{\alpha}(n-2)$
$\rho < 0$	$T < -t_{\alpha}(n-2)$	$T \geq -t_{\alpha}(n-2)$

Šio skyriaus teorinė medžiaga paruošta naudojantis [6], [7], [9] ir [13] literatūros šaltiniais.

4. TIESINĖ REGRESIJA

Regresijos modelis – statistinis modelis, leidžiantis vieno kintamojo reikšmės prognozuoti pagal kito kintamojo reikšmes. Statistiniai metodai, skirti regresijos modeliui sudaryti, patikrinti, ar jis tinkamas, ir taikyti prognozes, vadinama regresine analize.

Svarbiausias tiesinės regresinės analizės privalumas yra tai, kad sukuriami kintamuosius siejanti funkcija (sudaromas jos modelis). Kintamasis, kurio reikšmės norima prognozuoti, vadinamas priklausomu kintamuoju. Kintamasis, pagal kurio reikšmes norima prognozuoti priklausomo kintamojo reikšmės, vadinamas nepriklausomu kintamuoju. Regresijos modelio kintamuosius gali sieti vienaspusė arba dvipusė priklausomybė. Dvipusės priklausomybės atveju, kurį kintamąjį laikyti priklausomu kintamuoju, lemia tiriamoji problema. Tačiau nei vienaspusė, nei dvipusė regresijos modelio kintamųjų priklausomybė dar nereiškia jų priežastinio ryšio.

Bendriausias tiesinis tikimybinis modelis, siejantis intervalinius kintamuosius X ir Y , atrodo šitaip:

$$y_i = a + bx_i + e_i. \quad (4.1)$$

Regresiją galima sėkmingai taikyti, kai:

- paklaidos e_i – normaliai pasiskirstę atsitiktiniai dydžiai;
- visų e_i vidurkiai lygūs nuliui;
- visų e_i dispersijos yra lygios;
- visi e_i nepriklausomi.

Regresijos koeficientas a ir regresijos konstanta b yra apskaičiuojami

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}, \quad (4.2)$$

$$b = \frac{\sum y_i}{n} - a \frac{\sum x_i}{n}. \quad (4.3)$$

Tariame, kad duomenis sudaro intervalinių kintamųjų poriniai stebėjimai $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Turime atkreipti dėmesį, kad tarp $x_1, \dots, x_n$ gali būti ir sutampančių reikšmių, be to, jei $x_i = x_j$, tai nebūtinai $y_i = y_j$, t. y. vienodas x reikšmės nebūtinai atitiks vienodos y reikšmės. Pagrindinis tikslas, sudarant regresijos tiesę, parametrų a ir b įverčius $\hat{a}$ ir $\hat{b}$ parinkti taip, kad funkcijos $\hat{y}(x) = \hat{a} + \hat{b}x$ reikšmės taškuose x_i kuo mažiau skirtųsi nuo y_i . Gautoji funkcija bus naudojama priklausomojo kintamojo reikšmėms prognozuoti.

Kiekvieną x_i atitinka jo porinis stebėjimas y_i ir funkcijos $\hat{y}(x_i) = \hat{a} + \hat{b}x_i$ reikšmė. Geriausiai tinkanti funkcija yra tokia, kurios skirtumai $\hat{e}_i = y_i - \hat{y}(x_i), i = 1, 2, \dots, n$, mažiausi. $\hat{e}_i = y_i - \hat{y}(x_i) = y_i - (\hat{a} + \hat{b}x_i)$, esant $i = 1, 2, \dots, n$, vadinama i -tąja liekamąja paklaida.

Įverčiai $\hat{a}$ ir $\hat{b}$ randami minimizuojant

$$SSE = \sum_{i=1}^n (y_i - a - bx_i)^2. \quad (4.4)$$

Sumą SSE minimizuoja

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i \sum_{i=1}^n y_i) / n}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 / n}, \quad (4.5)$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \quad (4.6)$$

čia $\bar{y} = \sum_{i=1}^n y_i / n$, $\bar{x} = \sum_{i=1}^n x_i / n$. Aprašytasis funkcijos $\hat{y}(x)$ parametrų parinkimo metodas vadinamas mažiausių kvadratų metodu. Šiuo metodu gauti įverčiai yra nepaslinktieji ir turi mažiausias dispersijas.

4.1. DETERMINACIJOS KOEFICIENTAS

Regresinėje analizėje abu kintamieji yra intervaliniai. Tarkime, kad duomenis sudaro intervalinių kintamųjų poriniai stebėjimai: $(x_1, y_1), \dots, (x_n, y_n)$.

Turime gautą lygybę

$$SST = SSR + SSE, \quad (4.7)$$

čia $SST = \sum_{i=1}^n (y_i - \bar{y})^2$, $SSR = \sum_{i=1}^n (\hat{y}(x_i) - \bar{y})^2$, $SSE = \sum_{i=1}^n \hat{e}_i^2$, SST – visa kvadratų suma,

SSR – regresijos kvadratų suma, SSE – liekamųjų paklaidų kvadratų suma.

Santykis $\frac{SSR}{SST}$ vadinamas determinacijos koeficientu ir žymimas R^2 :

$$R^2 = \frac{SSR}{SST}. \quad (4.8)$$

Determinacijos koeficientas glaudžiai siejasi su Pirsono koreliacijos koeficientu – paprastosios tiesinės regresijos atveju determinacijos koeficientas sutampa su empirinio Pirsono koreliacijos koeficiento kvadratu (determinacijos koeficientas $= R^2$).

Svarbu paminėti, kad taip yra tik paprastosios tiesinės regresijos (vieno nepriklausomo kintamojo) atveju. Determinacijos koeficientas plačiai naudojamas kaip regresijos modelio tinkamumo indikatorius.

Didesnis determinacijos koeficientas reiškia, kad stebėjimai yra labiau koncentruoti apie mažiausių kvadratų metodu gautą tiesę. Remiantis vien tik determinacijos koeficientu, negalima pasakyti, ar tiesinės regresijos modelis turimiems duomenims tinka. Tačiau, praktiškai taikant regresinę analizę, dažniausiai reikalaujama, kad $R^2 \geq 0,25$. Jeigu $R^2 < 0,25$, tuomet labai abejotina, ar tiesinės regresijos modelis yra tinkamas.

4.2. DAUGIALYPĖ REGRESIJA

Daugialypės regresijos modelis taikomas tuomet, kai nepriklausomų intervalinių kintamųjų yra daugiau nei vienas. Analogiškai, kaip ir tiesinėje regresijoje, sudaromi daugialypės regresijos modeliai. Daugialypės tiesinės regresijos modelis užrašomas šitaip:

$$y_i = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + e_i. \quad (4.9)$$

Šioje lygtyje a ir b_i – nežinomos konstantos, e_i – atsitiktinė paklaida, x_i – nepriklausomi kintamieji, y_i – priklausomas kintamasis. Regresijos modelio prielaidos išlieka tos pačios, kaip ir tiesinės regresijos:

- paklaidos e_i - normaliai pasiskirstę atsitiktiniai dydžiai;
- visų e_i vidurkiai lygūs nuliui;
- visų e_i dispersijos yra lygios;
- visi e_i nepriklausomi.

Šio skyriaus teorinė medžiaga paruošta naudojantis [5], [6], [9] ir [13] literatūros šaltiniais.

II. PRAKTINĖ DALIS

1. DUOMENYS

Darbe tiriama tam tikrų sveikatos apsaugos sistemos faktorių įtaka Lietuvos gyventojų mirtingumui 1996–2011 metais. Tai – gydytojų (sveikatos priežiūros specialistų) bei ligoninėse esančių lovų skaičius 1996–2011 metais ir išlaidos (mln. Lt), skirtos sveikatos priežiūrai 2008–2011 metais. Tyrimui atlikti naudojami duomenys iš Lietuvos statistikos departamento svetainės (www.stat.gov.lt).

Praktinėje darbo dalyje atliktai duomenų analizei, kriterijų bei hipotezių tikrinimui naudotos programos – Microsoft Office Excel, Mathcad ir SPSS. Taip pat remtasi [1], [4], [9] ir [13] literatūros šaltiniais.

2. DUOMENŲ ANALIZĖ

2.1. PRADINĖ ANALIZĖ

Norint analizuoti turimus duomenis, pirmiausia turime apskaičiuoti visų tyrimui naudojamų duomenų grupių empirines skaitines charakteristikas: vidurkį, dispersiją, standartinį nuokrypį. Kadangi duomenys nėra pasikartojantys arba pasikartoja labai retai (kasmet miršta ir gimsta skirtingas skaičius žmonių, gydytojų ar lovų skaičius ligoninėse taip pat nėra kasmet vienodas), nėra prasmės ieškoti jų modos, medianos ar juos grupuoti ir sudaryti dažnių lentelę.

Pirmiausia rasime 2002–2011 metais vidutinio per parą gimusių kūdikių ir vidutinio per parą mirusių žmonių skaičiaus bei gyventojų skaičiaus metų pradžioje (1 priedas) empirines skaitines charakteristikas.

Vidutinio per parą gimusių kūdikių skaičiaus vidurkis (apskaičiuota pagal (1.4) formulę):

$$\bar{X} = \frac{82+84+83+84+86+89+96+100+98+94}{10} = \frac{896}{10} = 89,6;$$

vidutinio per parą gimusių kūdikių skaičiaus dispersija (apskaičiuota imčiai pagal (1.6) formulę):

$$\begin{aligned} S^2 &= \frac{1}{9}[(82-89,6)^2 + (84-89,6)^2 + (83-89,6)^2 + (84-89,6)^2 + (86-89,6)^2 + \\ &\quad + (89-89,6)^2 + (96-89,6)^2 + (100-89,6)^2 + (98-89,6)^2 + (94-89,6)^2] = \\ &= \frac{1}{9}(57,76 + 31,36 + 43,56 + 31,36 + 12,96 + 0,36 + 40,96 + 108,16 + 70,56 + \\ &\quad + 19,36) = \frac{1}{9} \cdot 416,4 = 46,26; \end{aligned}$$

vidutinio per parą gimusių kūdikių skaičiaus standartinis nuokrypis:

$$S = \sqrt{46,26} = 6,8.$$

Vidutinio per parą mirusių žmonių skaičiaus vidukis (apskaičiuota pagal (1.4) formulę):

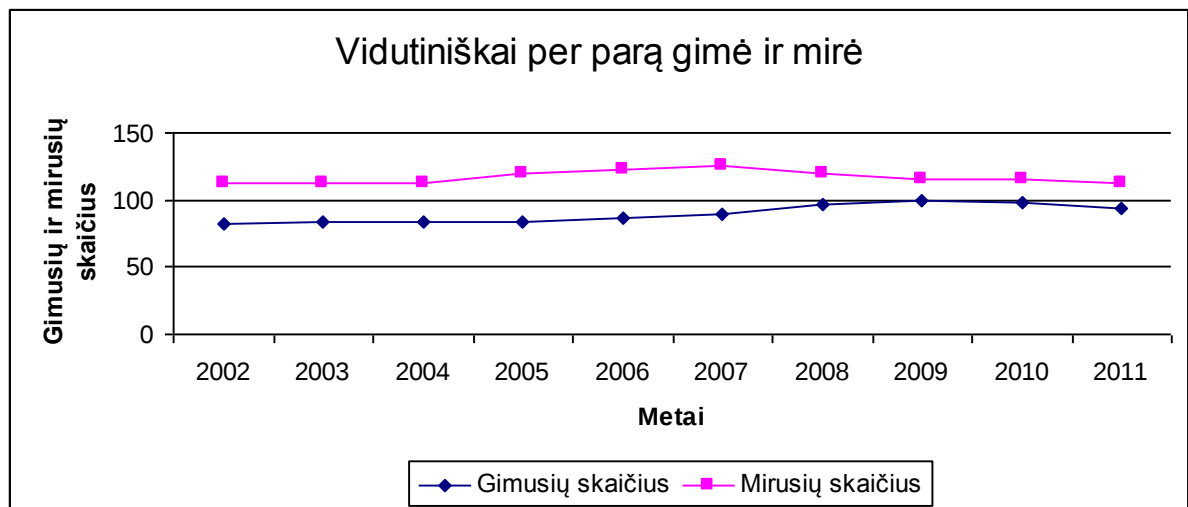
$$\bar{X} = \frac{113+112+113+120+123+125+120+115+115+112}{10} = \frac{1168}{10} = 116,8;$$

vidutinio per parą mirusių žmonių skaičiaus dispersija (apskaičiuota imčiai pagal (1.6) formulę):

$$\begin{aligned} S^2 &= \frac{1}{9}[(113-116,8)^2 + (112-116,8)^2 + (113-116,8)^2 + (120-116,8)^2 + \\ &+ (123-116,8)^2 + (125-116,8)^2 + (120-116,8)^2 + (115-116,8)^2 + \\ &+ (115-116,8)^2 + (112-116,8)^2] = \frac{1}{9}(14,44 + 23,04 + 14,44 + 10,24 + \\ &+ 38,44 + 67,24 + 10,24 + 3,24 + 3,24 + 23,04) = \frac{1}{9} \cdot 207,6 = 23,06; \end{aligned}$$

vidutinio per parą mirusių žmonių skaičiaus standartinis nuokrypis:

$$S = \sqrt{23,06} = 4,8.$$



2.1 pav.

2.1 pav. matome, kiek vidutiniškai 2002–2011 metais per parą gimė kūdikių ir kiek tuo pačiu periodu mirė gyventojų. Pagal atliktus skaičiavimus matome, kad vidutiniškai per parą 2002–2011 metais gimusių kūdikių vidurkis yra 89,6, o vidutiniškai tuo pačiu periodu per

parą mirusių gyventojų skaičiaus vidurkis yra 116,8. Apskaičiavus šiuos vidurkius ir remiantis 2.1 pav. galima teigti, kad mirtingumas yra gerokai didesnis, nei gimstamumas. Tolimesniame skyrelyje bus nagrinėjama, ar vidutinis gimstamumas ir mirtingumas per parą Lietuvoje skiriasi statistiškai reikšmingai.

Gyventojų skaičiaus metų pradžioje (1 priedas) vidurkis (apskaičiuota pagal (1.4) formulę):

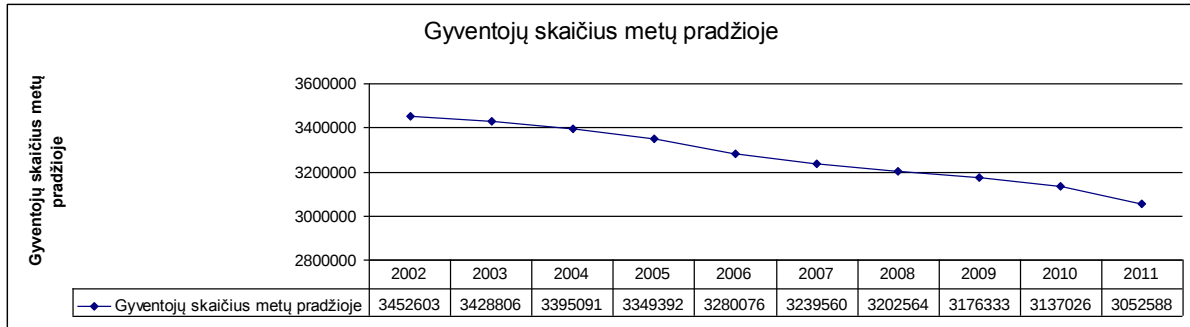
$$\bar{X} = \frac{3452603+3428806+3395091+3349392+3280076+3239560+3202564+3176333+3137026+3052588}{10} = \frac{32714039}{10} = 3271403,9;$$

gyventojų skaičiaus metų pradžioje dispersija (apskaičiuota imčiai pagal (1.6) formulę):

$$\begin{aligned} S^2 &= \frac{1}{9}[(3452603-3271403,9)^2 + (3428806-3271403,9)^2 + (3395091-3271403,9)^2 + \\ &+ (3349392-3271403,9)^2 + (3280076-3271403,9)^2 + (3239560-3271403,9)^2 + \\ &+ (3202564-3271403,9)^2 + (3176333-3271403,9)^2 + (3137026-3271403,9)^2 + \\ &+ (3052588-3271403,9)^2] = \frac{1}{10}(32833113840,81 + 24775421084,41 + \\ &+ 15298498706,41 + 6082143741,61 + 75205318,41 + 1014033967,21 + \\ &+ 4738931832,01 + 9038476026,81 + 18057420008,41 + 47880398092,81) = \\ &= \frac{1}{9} \cdot 159793642618,89 = 17754849179,87; \end{aligned}$$

gyventojų skaičiaus metų pradžioje standartinis nuokrypis:

$$S = \sqrt{17754849179,877777} = 133247,32.$$



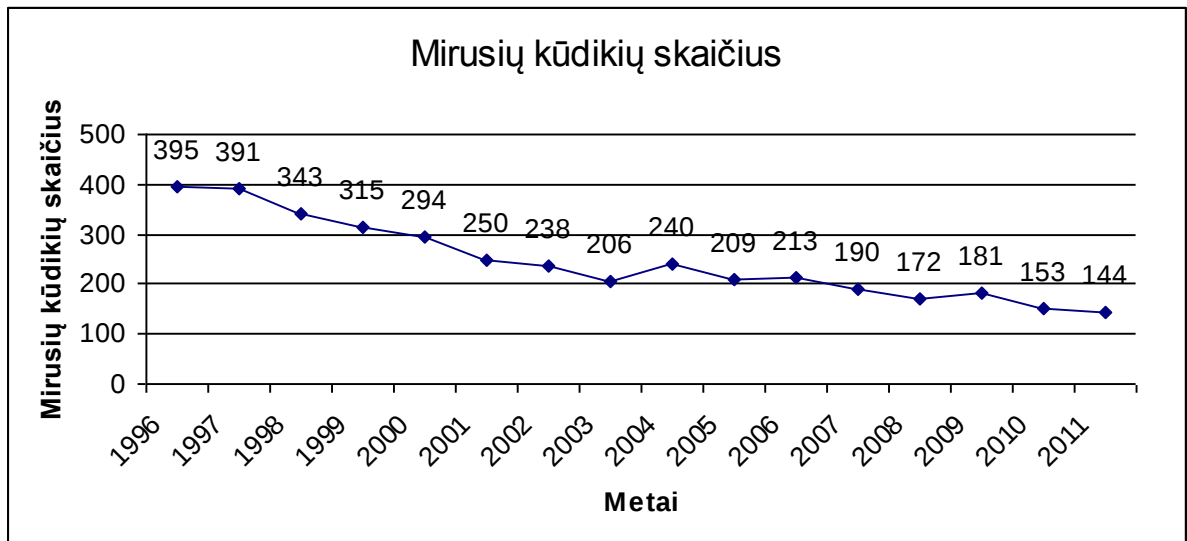
2.2 pav.

2.2 pav. matome, kaip kito gyventojų skaičius Lietuvoje 2002–2011 metais. Nurodytas gyventojų skaičius – tai gyventojų skaičius Lietuvoje atitinkamų metų pradžioje. Lietuvos statistikos departamento duomenimis, 2012 metų pradžioje Lietuvoje gyveno 3 007 758, 2012 metų kovo mėnesio pradžioje – 3 001 245, o 2013 metų kovo mėnesio pradžioje – 2 972 949 gyventojai. Pagal svetainėje pateiktus duomenis matome, kad gyventojų skaičius Lietuvoje kiekvieną mėnesį vis mažėja. Pagal atliktus skaičiavimus matome, kad vidutinis Lietuvos gyventojų skaičius 2002–2011 metų pradžioje buvo 3 271 403,9, tačiau vėliau bus nagrinėjama, ar vidutinis gyventojų skaičius metų pradžioje yra statistiškai reikšmingai didesnis nei 3 mln. gyventojų.

Toliau (2.1 lentelė) pateikiamos visų tyrime naudojamų duomenų empirinės skaitinės charakteristikos. Visi duomenys pateikti 1–5 prieduose.

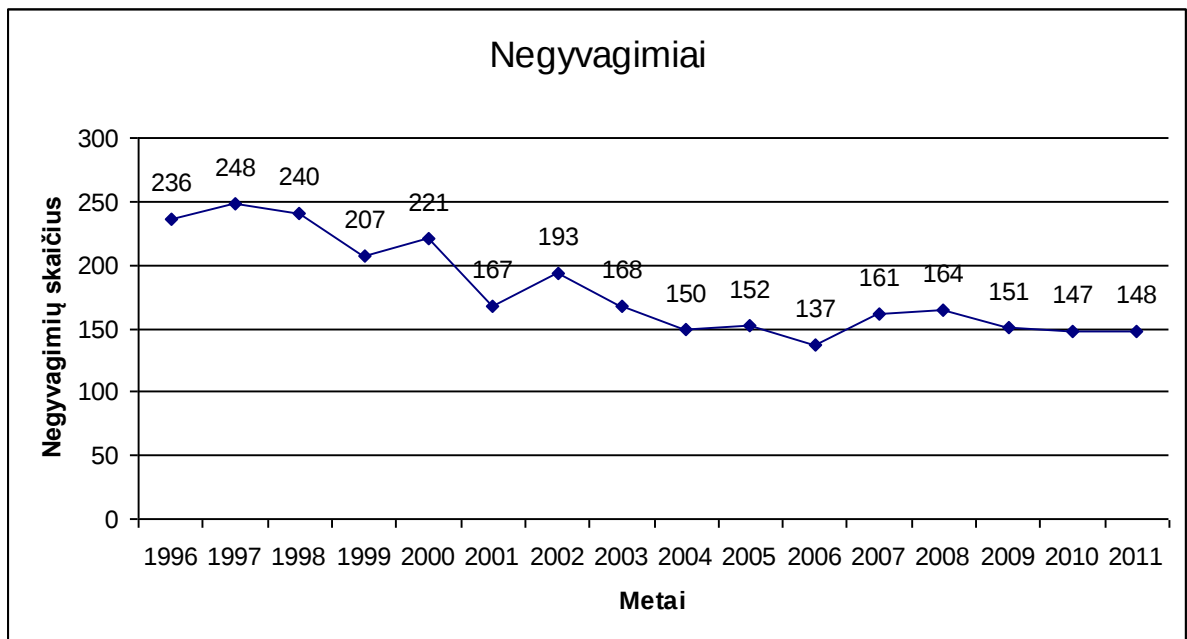
2.1 lentelė

Gimė per parą	Vidurkis	89,60	Natūrali gyventojų kaita	Vidurkis	-7989,25
	Dispersija	46,26		Dispersija	13970857,26
	Standartinis nuokrypis	6,80		Standartinis nuokrypis	3737,76
Mirė per parą	Vidurkis	116,80	Gydytojas. Sveikatos priežiūros specialistas	Vidurkis	13859,69
	Dispersija	23,06		Dispersija	301636,49
	Standartinis nuokrypis	4,80		Standartinis nuokrypis	549,21
Gyventojų skaičius metų pradžioje	Vidurkis	3271403,90	Gydytojas. Praktikuojantis sveikatos priežiūros specialistas	Vidurkis	12357,60
	Dispersija	17754849179,88		Dispersija	21996,80
	Standartinis nuokrypis	133247,32		Standartinis nuokrypis	148,31
Mirę kūdikiai	Vidurkis	245,87	Gydytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	Vidurkis	40,05
	Dispersija	6384,25		Dispersija	0,39
	Standartinis nuokrypis	79,90		Standartinis nuokrypis	0,63
Negyvagimiai	Vidurkis	180,62	Gydytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	Vidurkis	37,50
	Dispersija	1423,32		Dispersija	0,69
	Standartinis nuokrypis	37,73		Standartinis nuokrypis	0,83
Negyvagimiai vyrai	Vidurkis	101,94	Slaugytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	Vidurkis	6,74
	Dispersija	571,26		Dispersija	0,66
	Standartinis nuokrypis	23,90		Standartinis nuokrypis	0,81
Negyvagimės moterys	Vidurkis	78,69	Slaugytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	Vidurkis	7,20
	Dispersija	509,69		Dispersija	0,25
	Standartinis nuokrypis	22,58		Standartinis nuokrypis	0,40
Mirusieji	Vidurkis	41923,50	Lovų skaičius ligoninėse	Vidurkis	30812,75
	Dispersija	3328739,73		Dispersija	16390168,33
	Standartinis nuokrypis	1824,48		Standartinis nuokrypis	4048,48
1000 gyventojų tenka mirusių	Vidurkis	12,25	10 000 gyventojų tenka lovų ligoninėse	Vidurkis	88,79
	Dispersija	0,65		Dispersija	59,74
	Standartinis nuokrypis	0,81		Standartinis nuokrypis	7,73
Gimusiai	Vidurkis	33934,25	Lovų skaičius ligoninėse (be slaugos lovų)	Vidurkis	22729,80
	Dispersija	8980191,13		Dispersija	157368,70
	Standartinis nuokrypis	2996,69		Standartinis nuokrypis	396,69
1000 gyventojų tenka gimusių	Vidurkis	9,91	10000 gyventojų tenka lovų ligoninėse (be slaugos lovų)	Vidurkis	68,92
	Dispersija	0,81		Dispersija	0,88
	Standartinis nuokrypis	0,90		Standartinis nuokrypis	0,93
1000 gimusių tenka mirusių kūdikių	Vidurkis	7,20	Išlaidos sveikatos priežiūrai (mln. Lt)	Vidurkis	24323,25
	Dispersija	3,75		Dispersija	764168,52
	Standartinis nuokrypis	1,94		Standartinis nuokrypis	874,17



2.3 pav.

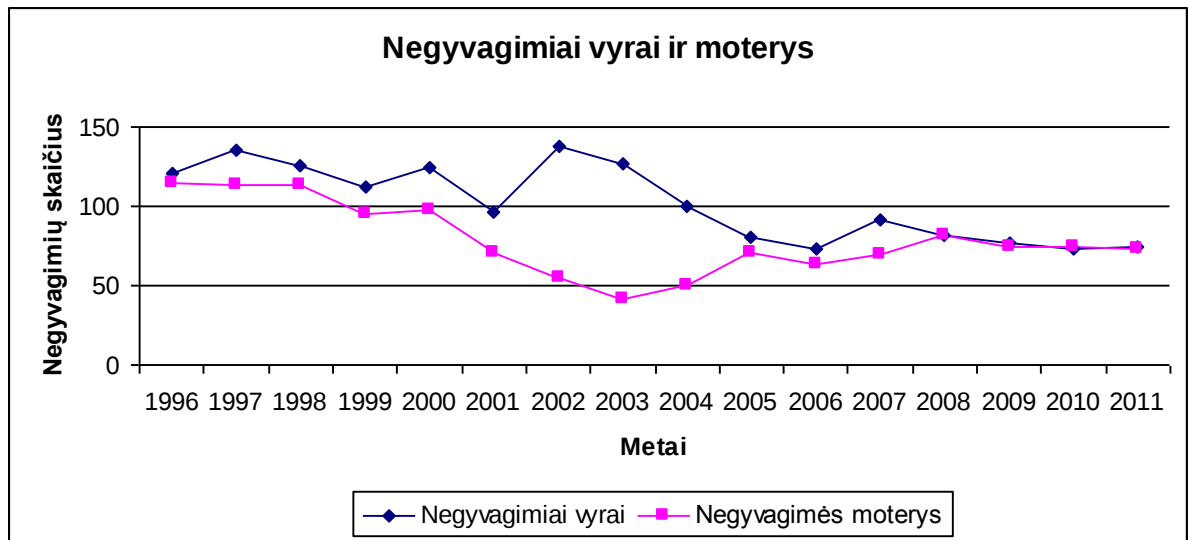
2.3 pav. pavaizduota, kiek kasmet Lietuvoje miršta kūdikių. Matome, kad nuo 1996 iki 2011 metų kūdikių mirtingumas vis mažėja ir per 15 metų jis sumažėjo beveik 3 kartus. Žinome, kad vidutiniškai 1996–2011 metais Lietuvoje kasmet mirdavo po 245,87 kūdikio (2.1 lentelė).



2.4 pav.

Priešingai nei mirusių kūdikių, negyvagimių skaičius (2.4 pav.) nėra pastoviai mažėjantis, tačiau laikui bėgant jis neviršija 1997 metais buvusio didžiausio negyvagimių skaičiaus per minėtą laikotarpį – 248. Vidutinis 1996–2011 metais gimusių negyvų kūdikių skaičius Lietuvoje – 180,62 (2.1 lentelė).

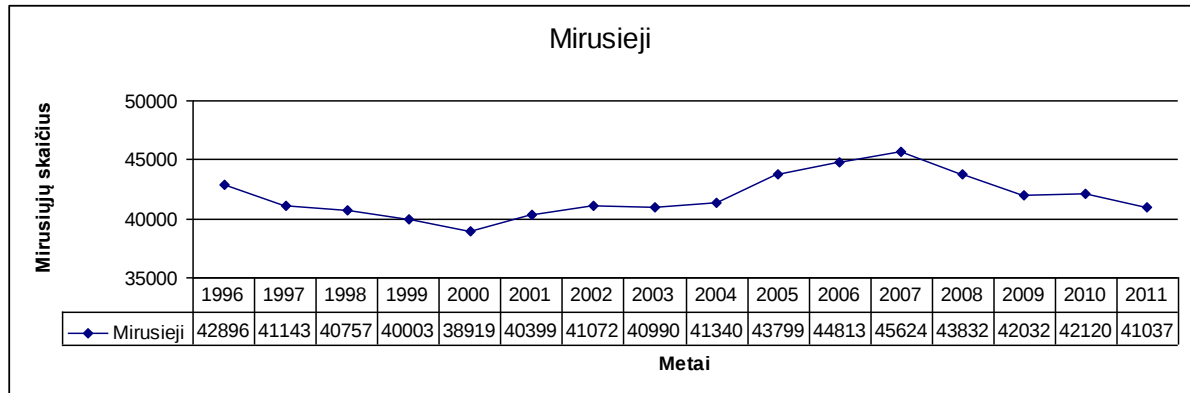
Toliau pateiktame 2.5 pav. galima stebėti, kaip skiriasi negyvagimių vyrų ir negyvagimių moterų skaičius.



2.5 pav.

Pagal pavaizduotus duomenis galime teigti, kad 1996 metais negyvagimių vyrų ir moterų skaičius Lietuvoje buvo beveik toks pat, iki 2001 metų negyvagimių vyrų būdavo daugiau, nei negyvagimių moterų, 2002–2004 metų laikotarpiu negyvagimių vyrų buvo gerokai daugiau, nei negyvagimių moterų, o nuo 2005 iki 2011 metų abiejų lyčių negyvagimių skaičius mažai skiriasi bei yra mažai kintantis.

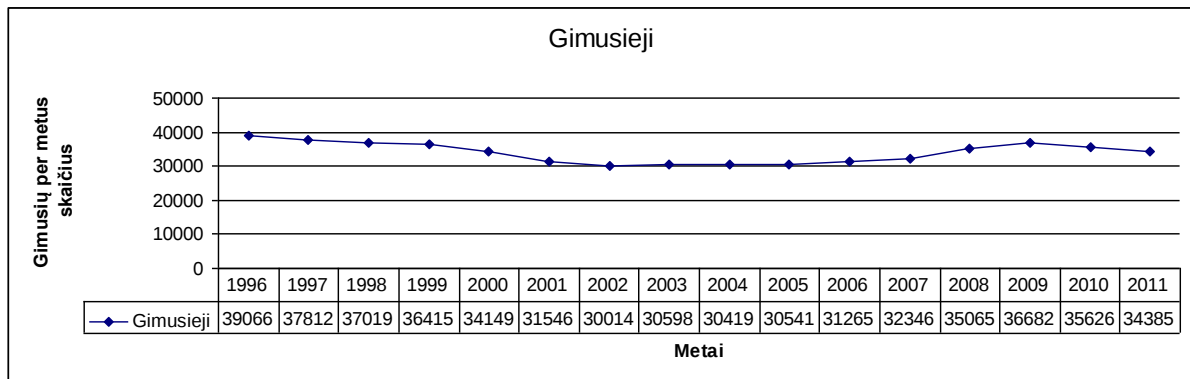
Anksčiau pateiktoje lentelėje (2.1 lent.) matome, kad negyvagimių vyrų vidurkis 1996–2011 metais yra 101,94, o tuo pačiu periodu gimusių negyvų moteriškos lyties kūdikių – 78,69. Pagal šiuos skaičiavimus ir pavaizduotą poligoną galime teigti, kad negyvagimių vyrų vidurkis yra gerokai didesnis, nei moterų, tačiau tolimesniame skyrelyje nagrinėsime, ar šie vidurkiai iš tiesų skiriasi statistiškai reikšmingai.



2.6 pav.

2.6 pav. pavaizduota, kaip kasmet kito mirusių gyventojų skaičius Lietuvoje. Matome, kad šis skaičius nuolat svyruoja tarp 40 ir 45 tūkstančių ribų bei vieną kartą nukrenta žemiau 40000 (2000-aisiais metais Lietuvoje mirė 38919 gyventojai) ir taip pat tik vieną kartą pakyla virš 45000 (2007-aisiais metais Lietuvoje mirė 45624 gyventojai) ribos.

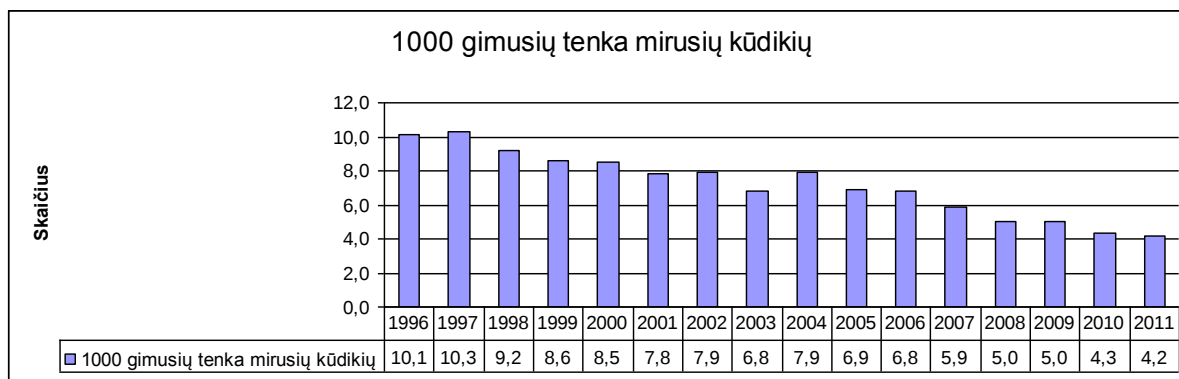
Kasmet dėl mirties 1996–2011 metais vidutiniškai Lietuva netekdavo 41923,5 gyventojų.



2.7 pav.

Analogiškai mirtingumui, gimstamumas kasmet (2.7 pav.) taip pat kinta nežymiai – 1996–2011 metų periodu gimusių kūdikių skaičius svyruoja tarp 30000 ir 40000 ribų, tačiau jis niekada nebuvo mažesnis nei 30000 (mažiausias gimstamumas buvo užfiksuotas 2002 metais – 30014 kūdikių) ir niekada nebuvo didesnis nei 40000 (didžiausias gimstamumas – 1996 metais – 39066 kūdikiai) kūdikių.

Vidutiniškai Lietuvoje 1996–2011 metais kasmet gimė 33934,25 kūdikiai.

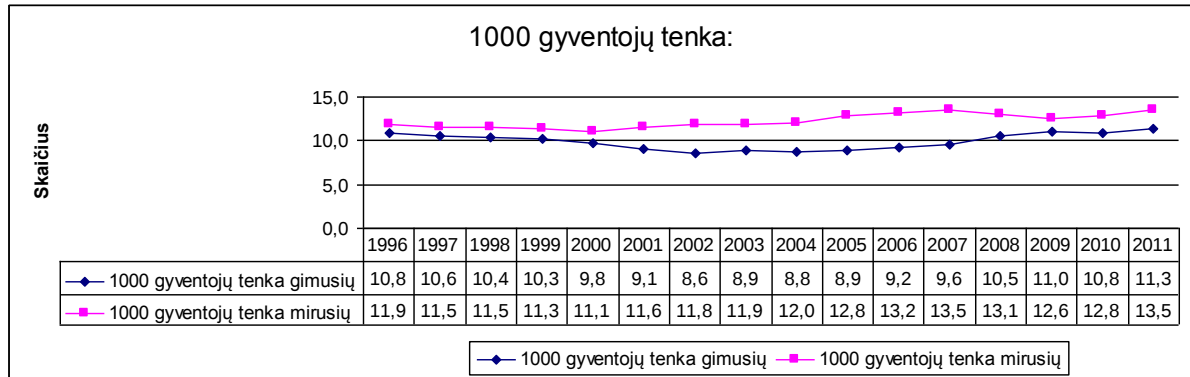


2.8 pav.

2.3 pav. pateikta, kaip 1996–2011 metais kasmet kinta mirusių kūdikių skaičius, o 2.7 pav. pavaizduotas to paties periodo gimusių kūdikių skaičiaus kasmetinis kitimas. Mirusių kūdikių skaičius svyruoja nuo 144 (2011 metais) iki 395 (1996 metais), gimusių kūdikių skaičius – nuo 30014 (2002 metais) iki 39066 (1996 metais). Matome, kad 1996 metais ir kūdikių gimstamumas, ir mirtingumas buvo didžiausi. Minėtais metais 1000 gimusių kūdikių teko 10,1 mirusio kūdikio (2.8 pav.).

Vėliau šis skaičius kito įvairiai – 1997 buvo didesnis, nei praėjusiais metais, vėliau ėmė sparčiai mažėti, 2001–2005 metais beveik nekito, o nuo 2006 metų sparčiai mažėjo, kol pasiekė mažiausią skaičių nagrinėjamame periode – 1000 gimusių kūdikių teko vos 4,2 mirusio kūdikio (2011 metais).

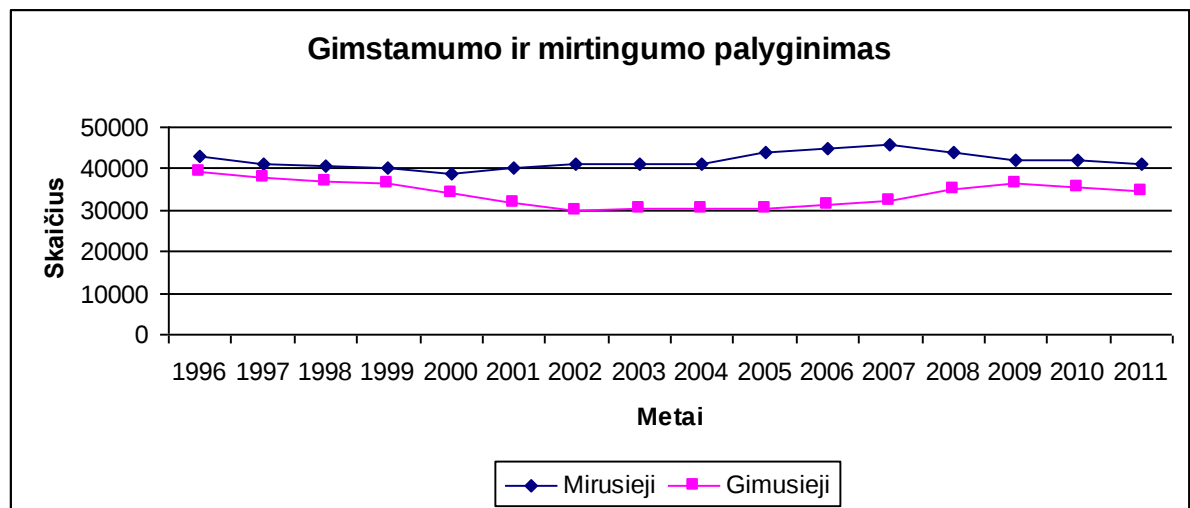
Vidutiniškai kasmet 1000 gimusių kūdikių teko 7,2 mirusio kūdikio (2.1 lent.).



2.9 pav.

Be mirusių kūdikių, tenkančių 1000 gimusių kūdikių, taip pat galime panagrinėti, kiek 1000 Lietuvos gyventojų kasmet tenka gimusių kūdikių ir kiek 1000 gyventojų tenka mirusių Lietuvos gyventojų.

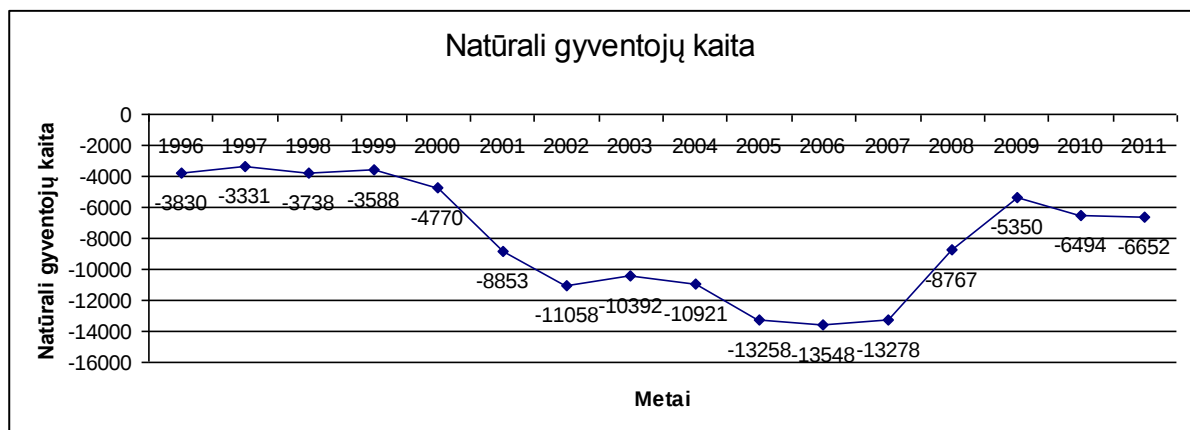
2.9 pav. matome, kad šios dvi kreivės kinta labai panašiai, o didžiausias atotrūkis tarp jų pastebimas 2001–2008 metais. Mažiausias skirtumas tarp gimusių kūdikių ir mirusių gyventojų skaičiaus, tenkančio 1000–čiui Lietuvos gyventojų, buvo 1997 metais – 0,9, kai gimstamumas buvo vienas didžiausių nagrinėjamame periode.



2.10 pav.

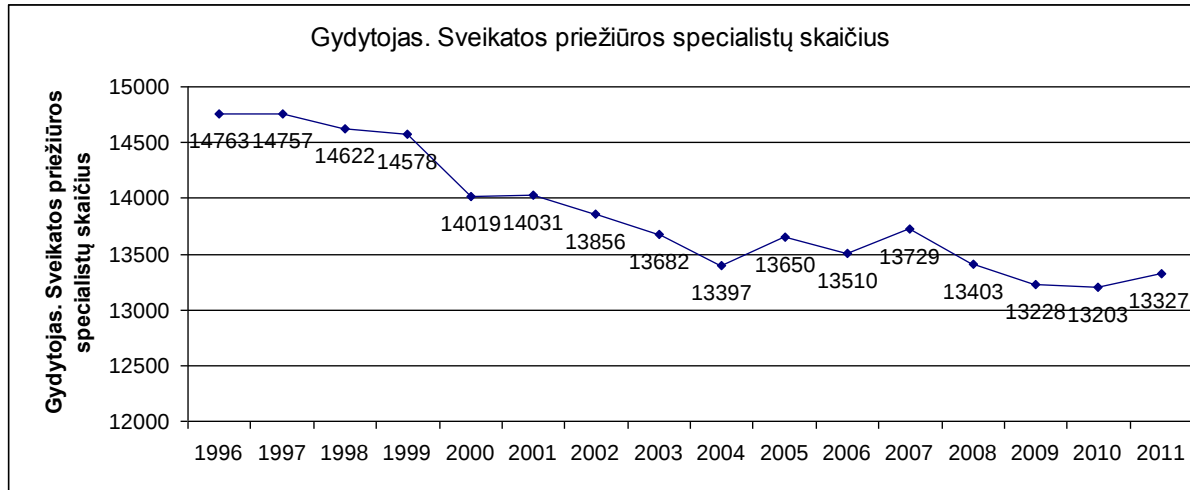
2.10 pav. matome, kaip skiriasi 1996–2011 metais mirusių gyventojų ir gimusių kūdikių skaičius. Didžiausias atotrūkis tarp šių dviejų kreivių pastebimas 2001–2008 metais, kai mirtingumas buvo didžiausias, o gimstamumas mažiausias. Dar vaizdžiau šis skirtumas atspindi natūralios gyventojų kaitos poligone (2.11 pav.). Natūrali gyventojų kaita – tai gyventojų skaičiaus pakitimai, susiję gimstamumu ir mirtingumu. Teigiamą natūrali gyventojų kaitą šalyje būna tuomet, kai gimstamumas būna didesnis nei mirtingumas. Stebint natūralią Lietuvos gyventojų kaitą matome, kad ji tiriamu laikotarpiu (1996–2011 metais) visuomet buvo neigiama, t. y. gimstamumas niekada neviršijo mirtingumo.

Tolesniame skyrelyje bus nagrinėjama, ar vidutinė natūrali gyventojų kaita Lietuvoje statistiškai reikšmingai skiriasi nuo –7000. Žinome, kad nagrinėjamu periodu (1996–2011 metais) vidutinė gyventojų kaita buvo –7989,25 (2.1 lent.).



2.11 pav.

2.12 pav. matome, kaip 1996–2011 metais Lietuvoje kito gydytojų (sveikatos priežiūros specialistų) skaičius. Tiriamu laikotarpiu jų skaičius niekada nebuvo didesnis nei pirmaisiais metais. 2005-aisiais, 2007-aisiais ir 2011-aisiais metais gydytojų (sveikatos priežiūros specialistų) skaičius buvo nežymiai padidėjęs, tačiau lyginant pirmuosius ir paskutiniuosius tyrimo metus matome, kad jis sumažėjęs net 1436. Vidutiniškai 1996–2011 metais Lietuvoje dirbo 13859,69 gydytojai.

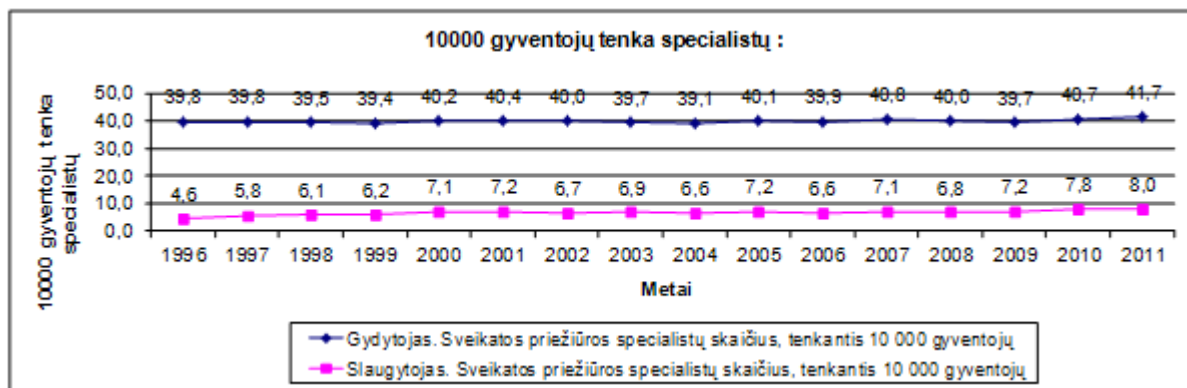


2.12 pav.

2.13 pav. pavaizduota, kiek tiriamu laikotarpiu 10000 gyventojų tenka specialistų – gydytojų ir slaugytojų. Nuo 1996 iki 2011 metų pastebimas specialistų skaičiaus, tenkančio 10000 Lietuvos gyventojų, didėjimas, kuris prieštarauja 2.12 pav. pavaizduotai kreivei (matome, kad specialistų skaičius iš tiesų mažėjantis).

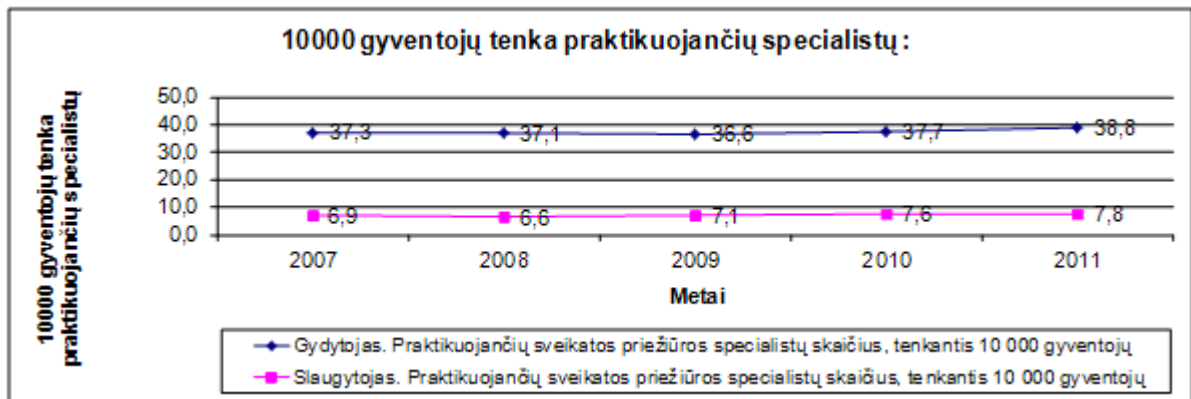
Tačiau tai nesunku paaiškinti – Lietuvos gyventojų skaičius mažėjo gerokai sparčiau nei specialistų skaičius, tenkantis 10000 gyventojų.

Vidutiniškai kasmet 10000 Lietuvos gyventojų teko 40,05 gydytojo ir 6,74 slaugytojo.



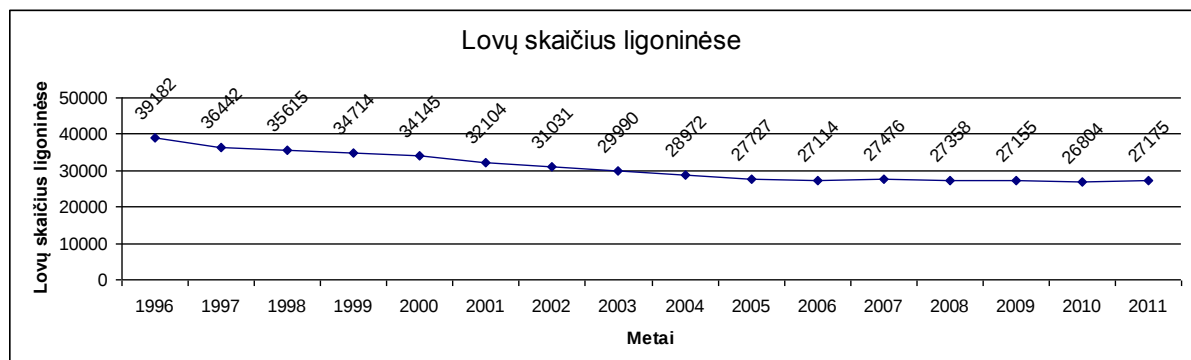
2.13 pav.

2.14 pav. pavaizduota, kaip 2007–2011 metais kito praktikuojančių sveikatos priežiūros specialistų (gydytojų ir slaugytojų) skaičius, tenkantis 10000 Lietuvos gyventojų. Vidutiniškai 10000 gyventojų kasmet teko 37,5 gydytojo ir 7,2 slaugytojo. Pagal 2.13 pav. analizę matome, kad dirbančių ir praktikuojančių sveikatos priežiūros specialistų vidurkis mažai skiriasi.

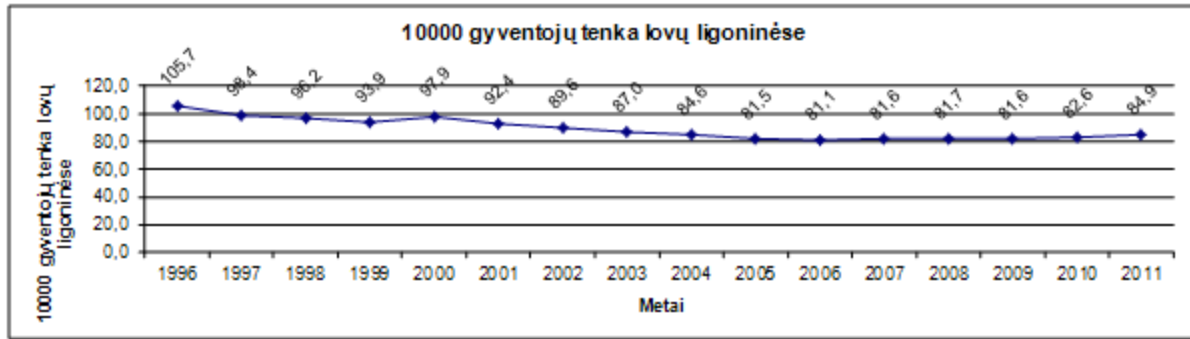


2.14 pav.

2.15 pav. ir 2.16 pav. pavaizduota, kaip tiriamu laikotarpiu kito lovų skaičius ligoninėse ir lovų skaičius ligoninėse, tenkantis 10000 Lietuvos gyventojų. Matome, kad abi kreivės leidžiasi žemyn, vadinasi tiek lovų skaičius, tiek gyventojų skaičius šalyje mažėjo beveik proporcingai. Kasmet Lietuvoje vidutiniškai buvo 30812,75 lovų, o tuo pačiu laikotarpiu 10000 gyventojų vidutiniškai teko po 88,79 lovos.

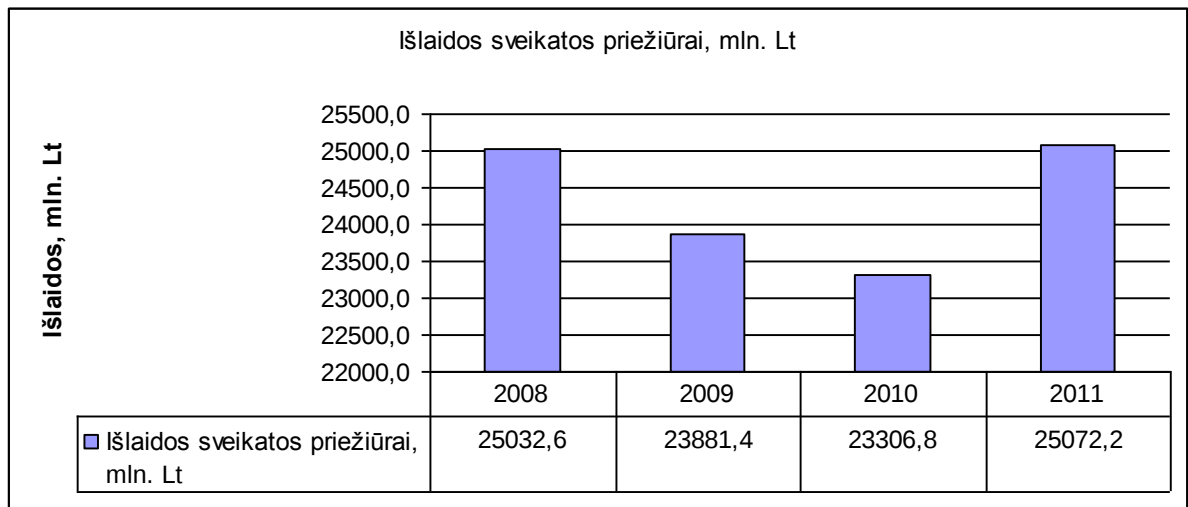


2.15 pav.



2.16 pav.

2.17 pav. matome, kaip 2008–2011 metais kito išlaidos (mln. Lt) sveikatos priežiūrai. Mažiausiai išlaidų sveikatos buvo 2010 metais, daugiausia – 2011 metais. Vidutiniškai šalyje kasmet skiriama 24323,25 mln. Lt skiriama sveikatos priežiūrai.



2.17 pav.

2.2. DUOMENŲ NORMALIŠKUMO TIKRINIMAS

Norėdami atlikti tyrimą, turime įsitikinti, kad visi naudojami duomenys yra pasiskirstę pagal normalųjį skirstinį. Pirmiausia tikrinamas per metus mirusių kūdikių skaičius. Iškeliamo dvi hipotezes :

H_0 : mirusių per metus kūdikių skaičiaus skirstinys statistiškai reikšmingai nesiskiria nuo normaliojo skirstinio.

H_1 : mirusių per metus kūdikių skaičiaus skirstinys statistiškai reikšmingai skiriasi nuo normaliojo skirstinio.

Naudojantis programa SPSS atliekame Kolmogorovo-Smirnov testą (2.2 lentelė).

2.2 lentelė

One-Sample Kolmogorov-Smirnov Test		
		Mirusiu kūdikiu skaicius
N		16
Normal Parameters ^a	Mean	245.88
	Std. Deviation	79.902
Most Extreme Differences	Absolute	.167
	Positive	.167
	Negative	-.101
Kolmogorov-Smirnov Z		.668
Asymp. Sig. (2-tailed)		.764
a. Test distribution is Normal.		

Asymp.Sig.(2-tailed) eilutėje matome, kad stebimasis reikšmingumo lygmuo $p = 0,764$. Tai reiškia, kad $p > \alpha$ (reikšmingumo lygmuo p didesnis už statistinio kriterijaus reikšmingumo lygmenį $\alpha = 0,05$). Tuomet turime priimti hipotezę H_0 ir sakome, kad nėra pagrindo ją atmesti. Tai reiškia, kad tiriamas skirstinys (mirusių kūdikių per metus skaičius) statistiškai reikšmingai nesiskiria nuo normaliojo skirstinio.

Kitiems duomenims taip pat iškeliami tokios hipotezės ir tikriname taikant Kolmogorovo-Smirnovo testą (tikrinimas pateiktas 6 priede). Toliau pateikti visų tyrimui naudojamų duomenų skirstinių reikšmingumo lygmenys rodo (2.3 lentelė), kad jie yra didesni už statistinio kriterijaus reikšmingumo lygmenį ($p > \alpha$, kai $\alpha = 0,05$):

2.3 lentelė

Gimė per parą	p=0,811	Natūrali gyventojų kaita	p=0,914
Mirė per parą	p=0,580	Gydytojas. Sveikatos priežiūros specialistų skaičius	p=0,828
Gyventojų skaičius metų pradžioje	p=0,998	Gydytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius	p=0,953
Mirę kūdikiai	p=0,764	Gydytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10 000 gyventojų	p=0,826
Negyvagimiai	p=0,245	Gydytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10 000 gyventojų	p=0,985
Negyvagimiai vyrai	p=0,725	Slaugytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10 000 gyventojų	p=0,681
Negyvagimės moterys	p=0,498	Slaugytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10 000 gyventojų	p=0,993
Mirusieji	p=0,624	Lovų skaičius ligoninėse	p=0,453
1000 gyventojų tenka mirusių	p=0,630	10 000 gyventojų tenka lovų ligoninėse	p=0,592
Gimusieji	p=0,793	Lovų skaičius ligoninėse (be slaugos lovų)	p=1,000
1000 gyventojų tenka gimusių	p=0,792	10 000 gyventojų tenka lovų ligoninėse (be slaugos lovų)	p=0,847
1000 gimusių tenka mirusių kūdikių	p=0,971	Išlaidos sveikatos priežiūrai (mln. Lt)	p=0,886

Tuomet kiekvienam duomenų skirstiniui turime priimti hipotezę H_0 ir sakome, kad nėra pagrindo ją atmesti. Tai reiškia, kad visi tiriami skirstiniai statistiškai reikšmingai nesiskiria nuo normaliojo skirstinio. Galime atlikti tolimesnius skaičiavimus bei tikrinti hipotezes.

3. PARAMETRINIŲ HIPOTEZIŲ TIKRINIMAS

3.1. PARAMETRINĖS HIPOTEZĖS VIENAI IMČIAI

Pirmiausia tikrinamos hipotezės vienai imčiai. Reikšmingumo lygmenį parenkame $\alpha = 0,05$ (visoms tiriamoms imtims vienodas).

Imčių parametrus tirti suformuluojami trys klausimai:

1. Ar vidutinis gyventojų skaičius metų pradžioje Lietuvoje statistiškai reikšmingai didesnis nei 3 mln. gyventojų?

2. Ar vidutinė natūrali gyventojų kaita Lietuvoje statistiškai reikšmingai skiriasi nuo -7000 ?

3. Ar vidutinis kasmet mirstančių gyventojų skaičius statistiškai reikšmingai didesnis nei 40000?

Iškeliamos hipotezės pirmajam klausimui.

$H_0: \mu = 3 \text{ mln. gyventojų.}$

$H_1: \mu > 3 \text{ mln. gyventojų.}$

Apskaičiuojame (pagal (2.6) formulę) kriterijaus statistikos reikšmę:

$$Z = \frac{3271403,90 - 3000000}{\frac{126409,51}{\sqrt{10}}} = \frac{271403,9}{39974,1969} = 6,789.$$

Pagal gautą rezultatą matome, kad $Z > z_{\alpha}$ ($6,789 > 1,644$), todėl hipotezę H_0 yra atmetama.

Su pasirinktu statistinio kriterijaus reikšmingumo lygmeniu $\alpha = 0,05$, naudojantis programa SPSS, dar kartą patikriname šias hipotezes (3.1 lentelė).

3.1 lentelė

One-Sample Test						
	Test Value = 3000000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Gyventojų skaičius metu pradžioje	6.441	9	.000	271403.900	176084.51	366723.29

Matome, kad reikšmingumo lygmuo $p = 0,000$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad vidutinis gyventojų skaičius Lietuvoje tiriamu laikotarpiu statistiškai reikšmingai didesnis nei 3 mln. gyventojų.

Iškeliamo hipotezes antrajam klausimui.

$$H_0: \mu = -7000.$$

$$H_1: \mu \neq -7000.$$

Apskaičiuojame (pagal (2.6) formulę) kriterijaus statistikos reikšmę:

$$Z = \frac{-7989,25 - (-7000)}{\frac{3737,76}{\sqrt{16}}} = \frac{-989,25}{934,44} = -1,059.$$

Pagal gautą rezultatą matome, kad $Z \leq z_{\alpha}$ ($-1,059 \leq 1,644$). Nėra pagrindo atmesti hipotezę H_0 , todėl ji yra priimama.

3.2 lentelė

One-Sample Test						
	Test Value = -7000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Natūrali gyventojų kaita	-1.059	15	.307	-989.250	-2980.96	1002.46

Matome (3.2 lentelė), kad reikšmingumo lygmuo $p = 0,307$. Tai reiškia, kad $p > \alpha$, todėl hipotezę H_0 priimame, nes nėra pagrindo ją atmesti. Sakome, kad vidutinė natūrali gyventojų kaita Lietuvoje tiriamu laikotarpiu statistiškai reikšmingai nesiskiria nuo -7000 .

Iškeliamos hipotezės trečiajam klausimui.

$H_0: \mu = 40000$ mln. gyventojų.

$H_1: \mu > 40000$ mln. gyventojų.

Apskaičiuojame (pagal (2.6) formulę) kriterijaus statistikos reikšmę:

$$Z = \frac{41923,5 - 40000}{\frac{1824,48}{\sqrt{16}}} = \frac{1923,5}{456,12} = 4,217.$$

Pagal gautą rezultatą matome, kad $Z > z_{\alpha}$ ($4,217 > 1,644$), todėl hipotezę H_0 yra atmetama.

Su pasirinktu statistinio kriterijaus reikšmingumo lygmeniu $\alpha = 0,05$, naudojantis programa SPSS, taip pat patikriname šias hipotezes.

3.3 lentelė

One-Sample Test						
	Test Value = 40000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Mirusieji	4.217	15	.001	1923.500	951.30	2895.70

Matome (3.3 lentelė), kad reikšmingumo lygmuo $p = 0,001$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad vidutinis mirštančių gyventojų skaičius Lietuvoje tiriamu laikotarpiu yra statistiškai reikšmingai didesnis nei 40000 gyventojų.

3.2. PARAMETRINĖS HIPOTEZĖS APIE DVIEJŲ IMČIŲ PARAMETRŲ LYGYBĘ

Šiame skyrelyje bus iškeliamos ir tikrinamos hipotezės dviems imtims, t. y. tikrinama, ar imčių vidurkiai tarpusavyje skiriasi statistiškai reikšmingai.

Parenkame (visoms 5 tikrinamoms imčių poroms) reikšmingumo lygmenį $\alpha = 0,05$.

1. Ar vidutinis gimstamumas ir mirtingumas 2002–2011 metais per parą Lietuvoje skiriasi statistiškai reikšmingai?

3.4 lentelė

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Per para gime	89.60	10	6.802	2.151
	Per para mire	116.80	10	4.803	1.519

Matome (3.4 lentelė), kad per parą gimusių kūdikių vidurkis yra 89,60, o per parą mirusių gyventojų skaičiaus vidurkis – 116,80. Iškeliame hipotezes:

$H_0: \mu_{pgg} = \mu_{ppm}$ (per parą gimusių kūdikių ir mirusių gyventojų skaičiaus vidurkiai statistiškai reikšmingai nesiskiria).

$H_1: \mu_{pgg} \neq \mu_{ppm}$ (per parą gimusių kūdikių ir mirusių gyventojų skaičiaus vidurkiai skiriasi statistiškai reikšmingai).

Apskaičiuojame imčių reikšmių skirtumus $d_i = x_i - y_i$, kai $i = 1, 2, \dots, 10$:

$$d_1 = x_1 - y_1 = 82 - 113 = -31,$$

$$d_2 = x_2 - y_2 = 84 - 112 = -28,$$

$$\begin{aligned}
d_3 &= x_3 - y_3 = 83 - 113 = -30, \\
d_4 &= x_4 - y_4 = 84 - 120 = -36, \\
d_5 &= x_5 - y_5 = 86 - 123 = -37, \\
d_6 &= x_6 - y_6 = 89 - 125 = -36, \\
d_7 &= x_7 - y_7 = 96 - 120 = -24, \\
d_8 &= x_8 - y_8 = 100 - 115 = -15, \\
d_9 &= x_9 - y_9 = 98 - 115 = -17, \\
d_{10} &= x_{10} - y_{10} = 94 - 112 = -18.
\end{aligned}$$

Tuomet apskaičiuojame imčių reikšmių skirtumų vidurkį $\bar{d}$ ir dispersiją s_d^2 :

$$\begin{aligned}
\bar{d} &= \frac{(-31) + (-28) + (-30) + (-36) + (-37) + (-36) + (-24) + (-15) + (-17) + (-18)}{10} = \\
&= -27,2, \\
s_d^2 &= \frac{961 + 784 + 900 + 1296 + 1369 + 1296 + 576 + 225 + 289 + 324 - (10 \cdot 739,84)}{9} = \\
&= 69,067
\end{aligned}$$

ir surandame (pagal (2.8) formulę) t reikšmę:

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}} = \frac{-27,2}{\sqrt{\frac{69,067}{10}}} = -10,35.$$

Gautos t reikšmės modulis yra didesnis už $t_{\frac{\alpha}{2}}(n-1)$, t.y. $|-10,35| > 2,262$. Todėl,

naudojantis lentele hipotezė H_0 turi būti atmesta.

3.5 lentelė

Paired Samples Test

Atlikus testą matome, kad reikšmingumo lygmuo (Sig.(2-tailed) eilutė) $p = 0,000$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad vidutinis per parą gimusių kūdikių skaičiaus vidurkis statistiškai reikšmingai skiriasi nuo per parą mirusių gyventojų skaičiaus vidurkio.

3.6 lentelė

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Negyvagimiai vyrai	101.94	16	23.901	5.975
	Negyvagimes moterys	78.69	16	22.576	5.644

53

$H_0: \mu_{ng} = \mu_{nm}$ (negyvagimių vyrų ir negyvagimių moterų vidurkiai statistiškai reikšmingai nesiskiria).

$H_1: \mu_{ng} \neq \mu_{nm}$ (negyvagimių vyrų ir negyvagimių moterų vidurkiai skiriasi statistiškai reikšmingai).

3.7 lentelė

Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Negyvagimiai vyrai - Negyvagimes moterys	23.250	27.177	6.794	8.768	37.732	3.42	15	.004

Atlikus testą matome (3.7 lentelė), kad reikšmingumo lygmuo (Sig.(2-tailed) eilutė) $p = 0,004$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad negyvagimių vyrų ir negyvagimių moterų skaičiaus vidurkiai skiriasi statistiškai reikšmingai.

3. Ar galima teigti, kad vidutiniškai 1996–2011 metais 1000 gyventojų tenka tiek pat mirusiųjų, kiek 1000 gimusių kūdikių tenka mirusių kūdikių?

3.8 lentelė

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 1000 gyventojų tenka mirusiu	12.256	16	.8066	.2017
1000 gimusių tenka mirusiu kūdikiu	7.200	16	1.9356	.4839

Matome (3.8 lentelė), kad 1000 gyventojų tenkančių mirusių skaičiaus vidurkis yra 12,256, o 1000 gimusių tenkančių mirusių kūdikių skaičiaus vidurkis – 7,2. Programos SPSS pagalba patikrinsime, ar šie vidurkiai skiriasi statistiškai reikšmingai. Iškeliame hipotezes :

$H_0: \mu_{gtm} = \mu_{gtmk}$ (1000 gyventojų tenkančių mirusių skaičiaus ir 1000 gimusių tenkančių mirusių kūdikių skaičiaus vidurkiai statistiškai reikšmingai nesiskiria).

$H_1: \mu_{gtm} \neq \mu_{gtmk}$ (1000 gyventojų tenkančių mirusių skaičiaus ir 1000 gimusių tenkančių mirusių kūdikių skaičiaus vidurkiai skiriasi statistiškai reikšmingai).

3.9 lentelė

Paired Samples Test

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 1000 gyventoju tenka mirusiu - 1000 gimusiu tenka mirusiu kudikiu	5.0562	2.6214	.6554	3.6594	6.4531	7.71	15	.000

Atlikus testą matome (3.9 lentelė), kad reikšmingumo lygmuo (Sig.(2-tailed) eilutė) $p = 0.000$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad 1000 gyventojų tenkančių mirusių skaičiaus ir 1000 gimusių tenkančių mirusių kūdikių skaičiaus vidurkiai skiriasi statistiškai reikšmingai

4. Ar tiesa, kad vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) ir slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius skiriasi statistiškai reikšmingai?

3.10 lentelė

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Gydytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	40.050	16	.6261	.1565
	10000 gyventojų tenka lovų ligoninėse	88.794	16	7.7294	1.9323

Matome (3.10 lentelė), kad vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius yra 40,05, o vidutinis slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius – 88,794. Programos SPSS pagalba patikrinsime, ar šie vidurkiai skiriasi statistiškai reikšmingai. Iškeliame hipotezes:

$H_0: \mu_{g10000} = \mu_{sl10000}$ (vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) ir vidutinis slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius statistiškai reikšmingai nesiskiria).

$H_1: \mu_{g10000} \neq \mu_{sl10000}$ (vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) ir vidutinis slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius skiriasi statistiškai reikšmingai).

3.11 lentelė

Paired Samples Test

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 Gydytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų - 10000 gyventojų tenkantiu ligoninėse	-48.7437	7.9164	1.9791	-52.9621	-44.5254	-24.62	15	.000

Atlikus testą matome (3.11 lentelė), kad reikšmingumo lygmuo (Sig.(2-tailed) eilutė) $p = 0,000$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) ir vidutinis slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius skiriasi statistiškai reikšmingai.

5. Ar vidutinis gimstamumas ir mirtingumas 1996–2011 metais Lietuvoje skiriasi statistiškai reikšmingai?

3.12 lentelė

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Mirusieji	41923.50	16	1824.483	456.121
Gimusieji	33934.25	16	2996.697	749.174

Matome (3.12 lentelė), kad mirusiųjų skaičiaus vidurkis yra 41923,5, o gimusiųjų skaičiaus vidurkis – 33934,25. Programos SPSS pagalba patikrinsime, ar šie vidurkiai skiriasi statistiškai reikšmingai. Iškeliame hipotezes:

$H_0: \mu_m = \mu_g$ (mirtingumo ir gimstamumo kasmetiniai vidurkiai statistiškai reikšmingai nesiskiria).

$H_1: \mu_m \neq \mu_g$ (mirtingumo ir gimstamumo kasmetiniai vidurkiai skiriasi statistiškai reikšmingai).

3.13 lentelė

Paired Samples Test								
		Paired Differences						
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df
					Lower	Upper		
Pair 1	Mirusieji – Gimusieji	7989.250	3737.761	934.440	5997.538	9980.962	8.55	15
								Sig. (2-tailed)
								.000

Atlikus testą matome (3.13 lentelė), kad reikšmingumo lygmuo (Sig.(2-tailed) eilutė) $p = 0,000$. Tai reiškia, kad $p < \alpha$, todėl hipotezę H_0 turime atmesti ir priimame hipotezę H_1 . Sakome, kad gimstamumo ir mirtingumo kasmetiniai vidurkiai skiriasi statistiškai reikšmingai.

4. HIPOTEZĖ APIE SKIRSTINIŲ LYGUMĄ

Naudojantis Vilkoksono kriterijumi patikrinsime hipotezes apie skirstinių lygumą.

Pirmiausia tarpusavyje tikrinami šie skirstiniai – išlaidos sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičius (2008–2011 metais). Iškeliame hipotezės:

H_0 : skirstiniai yra lygūs.

H_1 : skirstiniai nėra lygūs.

Naudojantis programa SPSS atliekamas testas (4.1 lentelė).

4.1 lentelė

Test Statistics ^a	
	Sveikatos priežiūros išlaidos - Mirusieji 2008 - 2011 metais
Z	-1.826 ^a
Asymp. Sig. (2-tailed)	.068

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Asymp.Sig.(2-tailed) eilutėje esanti p reikšmė yra didesnė už pasirinktą reikšmingumo lygmenį $\alpha = 0,05$, t.y. $0,068 > 0,05$. Tai reiškia, kad nėra pagrindo atmesti hipotezę H_0 , todėl ją priimame ir sakome, kad šie skirstiniai yra lygūs.

Toliau tarpusavyje tikrinami lovų skaičiaus ligoninėse ir mirusiųjų skaičiaus (1996–2011 metais) skirstiniai. Iškeliame hipotezes:

H_0 : skirstiniai yra lygūs.

H_1 : skirstiniai nėra lygūs.

4.2 lentelė

Test Statistics ^a	
	Lovu skaičius ligoninėse - Mirusieji
Z	-3.516 ^a
Asymp. Sig. (2-tailed)	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Atlikus testą matome (4.2 lentelė), kad Asymp.Sig.(2-tailed) eilutėje esanti p reikšmė yra lygi nuliui, t.y. $p < 0,05$. Tai reiškia, kad turime atmesti hipotezę H_0 ir priimti alternatyvią hipotezę H_1 . Sakome, kad skirstiniai nėra lygūs.

Tikriname gydytojų (sveikatos priežiūros specialistų) ir mirusiųjų (1996–2011 metais) skaičių skirstinius. Iškeliame hipotezes:

H_0 : skirstiniai yra lygūs.

H_1 : skirstiniai nėra lygūs.

4.3 lentelė

Test Statistics ^a	
	Gydytojas. Sveikatos priežiūros specialistų skaičius - Mirusieji
Z	-3.516 ^a
Asymp. Sig. (2-tailed)	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Asymp.Sig.(2-tailed) eilutėje (4.3 lentelė) esanti p reikšmė yra lygi nuliui ($p < \alpha$). Tai reiškia, kad turime atmesti hipotezę H_0 ir priimti hipotezę H_1 . Sakome, kad skirstiniai nėra lygūs.

5. NEPRIKLAUSOMUMO KRITERIJAUŠ TIKRINIMAS

Prieš tiriant priklausomybes ir ieškant koreliacijos tarp dviejų kintamųjų, patikrinsime skirstinių požymių nepriklausomumą. Reikšmingumo lygmenį pasirenkame $\alpha = 0,05$.

Iškeliamos hipotezės.

H_0 : išlaidos sveikatos priežiūrai (mln. Lt) neturi įtakos mirusiųjų skaičiui 2008–2011 metais.

H_1 : išlaidos sveikatos priežiūrai (mln. Lt) turi įtakos mirusiųjų skaičiui 2008–2011 metais.

5.1 lentelė

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.000 ^a	9	.213
Likelihood Ratio	11.090	9	.270
Linear-by-Linear Association	.063	1	.801
N of Valid Cases	4		

a. 16 cells (100.0%) have expected count less than 5. The minimum expected count is .25.

Pagal lentelėje (5.1 lentelė) turimus duomenis (Asymp.Sig.(2-tailed) eilutė) matome, kad $p = 0,213$. Tai reiškia, kad $p > \alpha$, tuomet nėra pagrindo atmesti H_0 ir ją priimame. Sakome, kad išlaidos sveikatos priežiūrai (mln. Lt) neturi įtakos mirusiųjų skaičiui 2008–2011 metais.

Iškeliamos hipotezės:

H_0 : lovų skaičius ligoninėse neturi įtakos gyventojų mirtingumui 1996–2011 metais.

H_1 : lovų skaičius ligoninėse turi įtakos gyventojų mirtingumui 1996–2011 metais.

5.2 lentelė

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.400E2	225	.235
Likelihood Ratio	88.723	225	1.000
Linear-by-Linear Association	3.439	1	.064
N of Valid Cases	16		

a. 256 cells (100.0%) have expected count less than 5. The minimum expected count is .06.

Pagal lentelėje (5.2 lentelė) turimus duomenis (Asymp.Sig.(2-tailed) eilutė) matome, kad $p = 0,235$. Tai reiškia, kad $p > \alpha$, tuomet nėra pagrindo atmesti H_0 ir ją priimame. Sakome, kad lovų skaičius ligoninėse neturi įtakos gyventojų mirtingumui 1996–2011 metais.

Iškeliamos hipotezės:

H_0 : gydytojų (sveikatos priežiūros specialistų) skaičius neturi įtakos gyventojų mirtingumui 1996–2011 metais.

H_1 : gydytojų (sveikatos priežiūros specialistų) skaičius turi įtakos gyventojų mirtingumui 1996–2011 metais.

5.3 lentelė

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.400E2	225	.235
Likelihood Ratio	88.723	225	1.000
Linear-by-Linear Association	1.578	1	.209
N of Valid Cases	16		

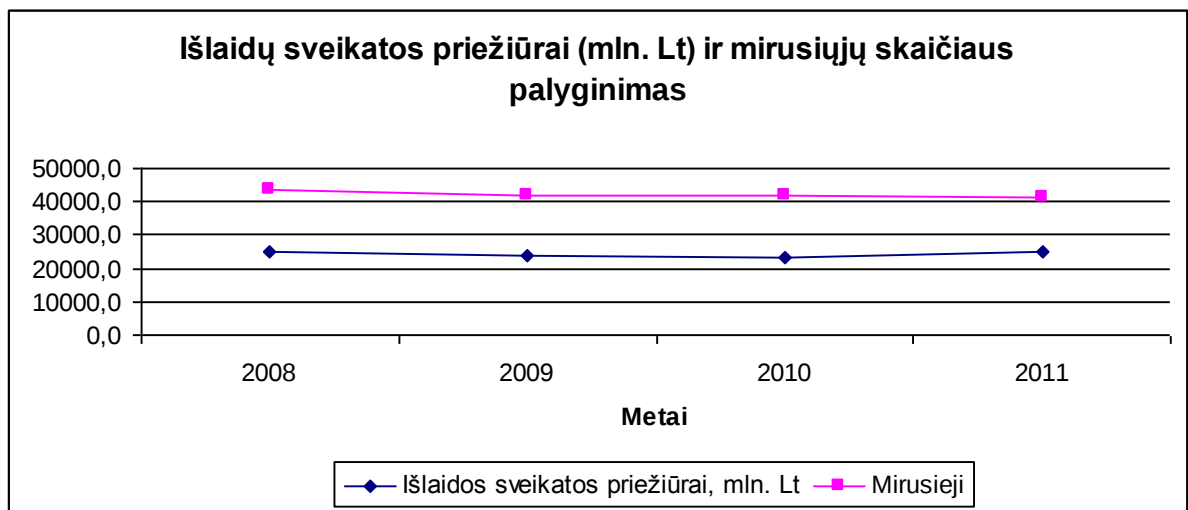
a. 256 cells (100.0%) have expected count less than 5. The minimum expected count is .06.

Pagal lentelėje (5.3 lentelė) turimus duomenis (Asymp.Sig.(2-tailed) eilutė) matome, kad $p = 0,235$. Tai reiškia, kad $p > \alpha$, tuomet nėra pagrindo atmesti H_0 ir ją priimame. Sakome, kad gydytojų (sveikatos priežiūros specialistų) skaičius neturi įtakos gyventojų mirtingumui 1996–2011 metais.

6. HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTO LYGYBĘ NULIUI

Svarbiausia šio tyrimo užduotis – rasti, kas darė didžiausią įtaką Lietuvos gyventojų mirtingumui tam tikru periodu. Tam, kad tai padarytume, reikia ištirti įvairias priklausomybes, t. y. patikrinti, ar tarp dviejų tiriamų kintamųjų egzistuoja koreliacija ir ar ji yra statistiškai reikšminga.

Pirmiausia patikrinsime, ar egzistuoja koreliacija (ir ar ji yra statistiškai reikšminga) tarp išlaidų sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičiaus 2008–2011 metais. Šių dviejų kintamųjų kitimas tiriamu periodu pavaizduotas 6.1 pav.



6.1 pav.

Iškeliamos dvi hipotezės:

H_0 : koreliacijos koeficientas nėra statistiškai reikšmingas.

H_1 : koreliacijos koeficientas yra statistiškai reikšmingas.

Pirmiausia apskaičiuojame (pagal (3.2) formulę) koreliacijos koeficientą:

$$\sum_{i=1}^4 x_i y_i = (25032,6 \cdot 43832) + (23881,4 \cdot 42032) + (23306,8 \cdot 42120) + (25072,2 \cdot 41037) = 4111582215,4;$$

$$\sum_{i=1}^4 x_i = 25032,6 + 23881,4 + 23306,8 + 25072,2 = 97293;$$

$$\sum_{i=1}^4 y_i = 43832 + 42032 + 42120 + 41037 = 169021 ;$$

$$\sum_{i=1}^4 x_i^2 = (25032,6)^2 + (23881,4)^2 + (23306,8)^2 + (25072,2)^2 = 2368774467,8 ;$$

$$\left(\sum_{i=1}^4 x_i \right)^2 = 97293^2 = 9465927849;$$

$$\sum_{i=1}^4 y_i^2 = (43832)^2 + (42032)^2 + (42120)^2 + (41037)^2 = 7146063017 ;$$

$$\left(\sum_{i=1}^4 y_i \right)^2 = 169021^2 = 28568098441;$$

$$R = \frac{(4 \cdot 4111582215,4) - (97293) \cdot (169021)}{\sqrt{((4 \cdot 2368774467,8) - 9465927849)((4 \cdot 7146063017) - 28568098441)}} = 0,145.$$

Matome, kad koreliacijos koeficientas $R = 0,145$. Tai rodo, kad ryšys tarp kintamųjų yra labai silpnas arba jo iš viso nėra. Randame (pagal (3.6) formulę) kriterijaus statistikos T stebėtają reikšmę:

$$t = R \sqrt{\frac{n-2}{1-R^2}} = 0,145 \cdot \sqrt{\frac{4-2}{1-0,145^2}} = 0,207.$$

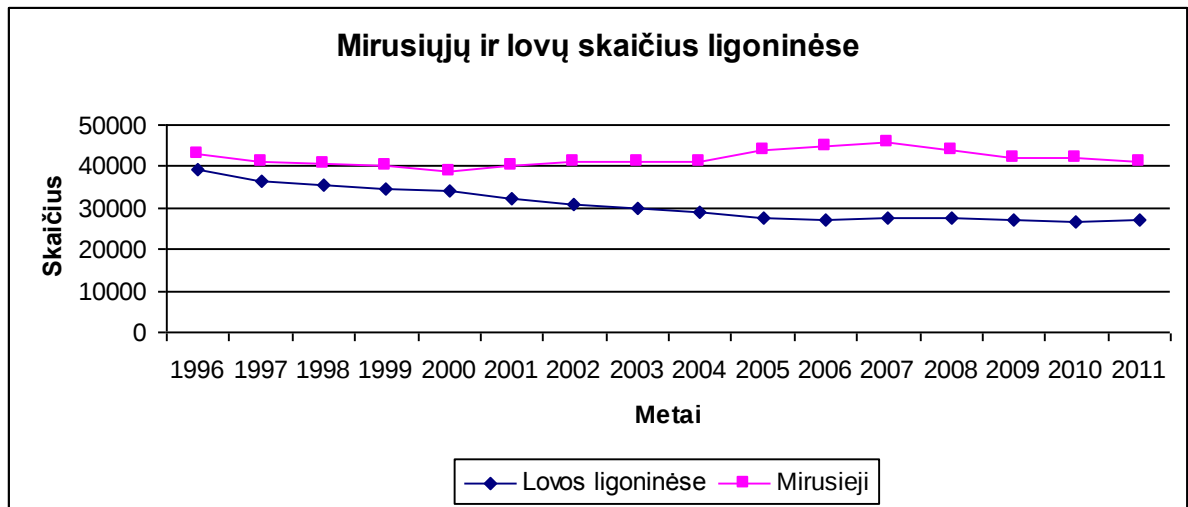
Gautos t reikšmės modulis yra mažesnis už $t_{\frac{\alpha}{2}}(n-2)$, t. y. $|0,207| < 2,306$. Todėl, naudojantis lentelę, matome, kad nėra pagrindo atmesti hipotezę H_0 , turime ją priimti. Sakome, kad koreliacijos koeficientas nėra statistiškai reikšmingas.

6.1 lentelė

Correlations			
		Sveikatos prižiūros islaidos	Mirusieji 2008 - 2011 metais
Sveikatos prižiūros islaidos	Pearson Correlation	1.000	.145
	Sig. (2-tailed)		.855
	N	4.000	4
Mirusieji 2008 - 2011 metais	Pearson Correlation	.145	1.000
	Sig. (2-tailed)	.855	
	N	4	4.000

Naudodamiesi programa SPSS atliekame testą (6.1 lentelė) ir pagal lentelės duomenis (Sig.(2-tailed) eilutė) matome, kad $p = 0,855$. Kadangi reikšmingumo lygmuo $\alpha = 0,01$, tuomet $p > \alpha$. Vadinasi, nėra pagrindo atmesti H_0 . Priimame hipotezę H_0 ir sakome, kad gautasis koreliacijos koeficientas 0,145 nėra statistiškai reikšmingas.

Toliau tikriname, ar egzistuoja koreliacija (ir ar ji yra statistiškai reikšminga) tarp mirusių gyventojų ir lovų ligoninėse skaičiaus 1996–2011 metais. Šių dviejų kintamųjų kitimas tiriamu periodu pavaizduotas 6.2 pav.



6.2 pav.

Vėl iškeliame dvi hipotezes:

H_0 : koreliacijos koeficientas nėra statistiškai reikšmingas.

H_1 : koreliacijos koeficientas yra statistiškai reikšmingas.

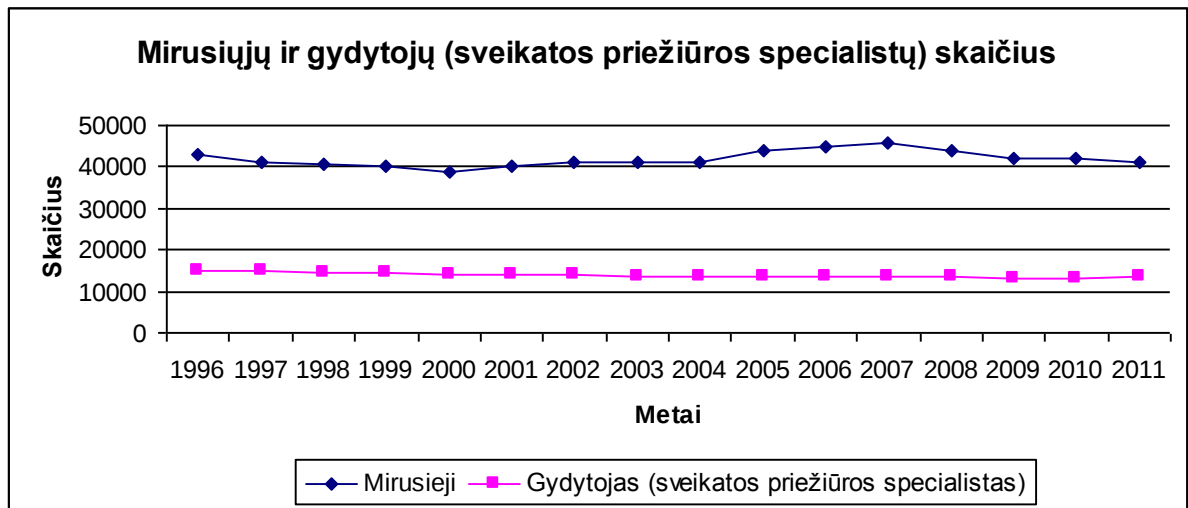
6.2 lentelė

Correlations			
		Mirusieji	Lovų skaičius ligoninėse
Mirusieji	Pearson Correlation	1.000	-.479
	Sig. (2-tailed)		.061
	N	16.000	16
Lovų skaičius ligoninėse	Pearson Correlation	-.479	1.000
	Sig. (2-tailed)	.061	
	N	16	16.000

Naudodamiesi programa SPSS atliekame testą (6.2 lentelė) ir pagal lentelės duomenis (Sig.(2-tailed) eilutė) matome, koreliacijos koeficientas tarp šių kintamųjų yra $-0,479$. Tai reiškia, kad ryšys tarp šių kintamųjų yra silpnas. Matome, kad $p = 0,061$. Kadangi reikšmingumo lygmuo $\alpha = 0,01$, tuomet $p > \alpha$. Vadinasi, nėra pagrindo atmesti H_0 .

Priimame hipotezę H_0 ir sakome, kad gautasis koreliacijos koeficientas $-0,479$ nėra statistiškai reikšmingas.

Taip pat svarbu ištirti, ar priklausomybė tarp mirusiųjų ir gydytojų (sveikatos priežiūros specialistų) skaičiaus, nusakoma koreliacija, yra statistiškai reikšminga? Duomenų kitimas 1996–2011 metais pavaizduotas 6.3 pav.



6.3 pav.

Iškeliamos dvi hipotezės:

H_0 : koreliacijos koeficientas nėra statistiškai reikšmingas.

H_1 : koreliacijos koeficientas yra statistiškai reikšmingas.

6.3 lentelė

Correlations			
		Mirusieji	Gydytojas. Sveikatos prižiūros specialistu skaicius
Mirusieji	Pearson Correlation	1.000	-.324
	Sig. (2-tailed)		.220
	N	16.000	16
Gydytojas. Sveikatos prižiūros specialistu skaicius	Pearson Correlation	-.324	1.000
	Sig. (2-tailed)	.220	
	N	16	16.000

Atlikus testą matome (6.3 lentelė), kad koreliacijos koeficientas tarp šių kintamųjų yra $-0,324$. Taip pat matome, kad $p = 0,220$. Kadangi reikšmingumo lygmuo $\alpha = 0,01$, tuomet $p > \alpha$. Vadinasi, nėra pagrindo atmesti H_0 . Priimame hipotezę H_0 ir sakome, kad gautasis koreliacijos koeficientas $-0,324$ nėra statistiškai reikšmingas.

7. REGRESINĖ ANALIZĖ

7.1. PAPRASTOJI TIESINĖ REGRESIJA

Praėjusiame skyrelyje tyrėme koreliaciją tarp šių kintamųjų porų:

1. Išlaidos sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičius 2008–2011 metais.
2. Mirusieji gyventojai ir lovų skaičius ligoninėse 1996–2011 metais.
3. Mirusieji gyventojai ir gydytojų (sveikatos priežiūros specialistų) skaičius 1996–2011 metais.

Atlikus testus (6.1, 6.2, 6.3 lentelės) programos SPSS pagalba matome, kad gautieji koreliacijos koeficientai nėra statistiškai reikšmingi. Tačiau vis dėlto bandysime rasti regresijos tieses ir jas pavaizduoti.

Analizuojama pirmoji priklausomybė – išlaidos sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičius 2008–2011 metais. Kintamasis, kurio reikšmės bus prognozuojamos (mirusiųjų gyventojų skaičius), vadinamas priklausomu kintamuoju, o kintamasis, pagal kurio reikšmes pronozuosime (išlaidos sveikatos priežiūrai, mln. Lt) – nepriklausomu kintamuoju.

Rasime regresijos tiesei reikalingus koeficientą a (naudojantis (4.2) formule) ir regresijos konstantą b (naudojantis (4.3) formule):

$$\sum_{i=1}^4 x_i = 25032,6 + 23881,4 + 23306,8 + 25072,2 = 97293 ;$$

$$\sum_{i=1}^4 y_i = 43832 + 42032 + 42120 + 41037 = 169021 ;$$

$$\sum_{i=1}^4 x_i y_i = (25032,6 \cdot 43832) + (23881,4 \cdot 42032) + (23306,8 \cdot 42120) + (25072,2 \cdot 41037) = 4111582215,4;$$

$$\sum_{i=1}^4 x_i^2 = (25032,6)^2 + (23881,4)^2 + (23306,8)^2 + (25072,2)^2 = 2368774467,8;$$

$$\left(\sum_{i=1}^4 x_i \right)^2 = 97293^2 = 9465927849.$$

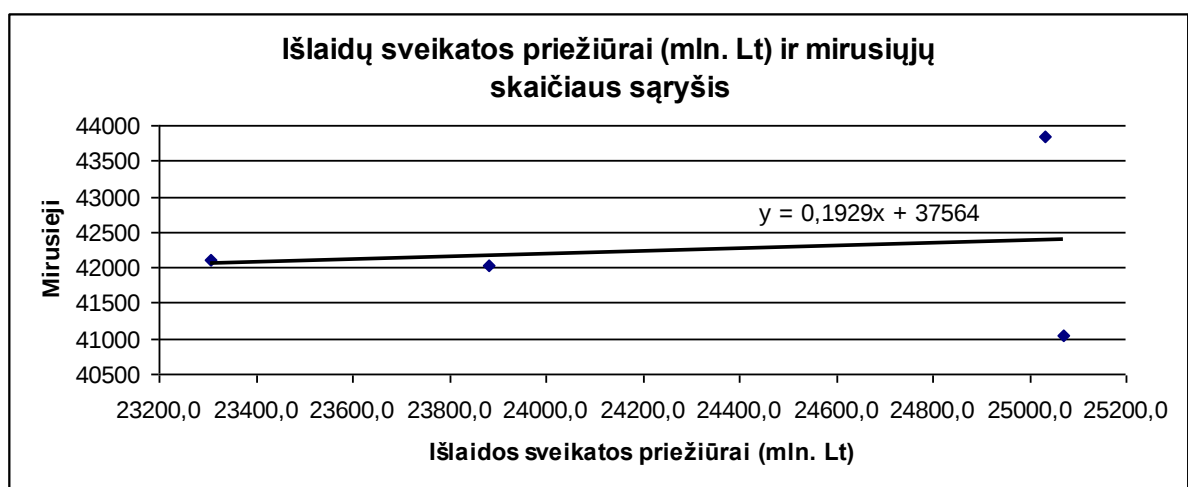
$$a = \frac{(4 \cdot 4111582215,4) - (97293) \cdot (169021)}{(4 \cdot 2368774467,8) - 9465927849} = 0,193 ;$$

$$b = \frac{169021}{4} - 0,193 \cdot \frac{97293}{4} = 37560,863.$$

Galime parašyti regresijos tiesės lygtį (pagal (4.1))

$$y = 0,193x + 37560,863$$

ir ją nubrėžti (7.1 pav.).



7.1 pav.

Apskaičiuojame determinacijos koeficientą (koreliacijos koeficiento kvadratas):

$$R^2 = 0,145^2 = 0,021 .$$

Jei determinacijos koeficientas yra didesnis nei 0,25, tai regresijos modelis yra laikomas tinkamu. Nagrinėjamu atveju matome, kad $R^2 = 0,021$, t. y. $0,021 < 0,25$, todėl šiuo regresijos modeliu pasikliauti negalima.

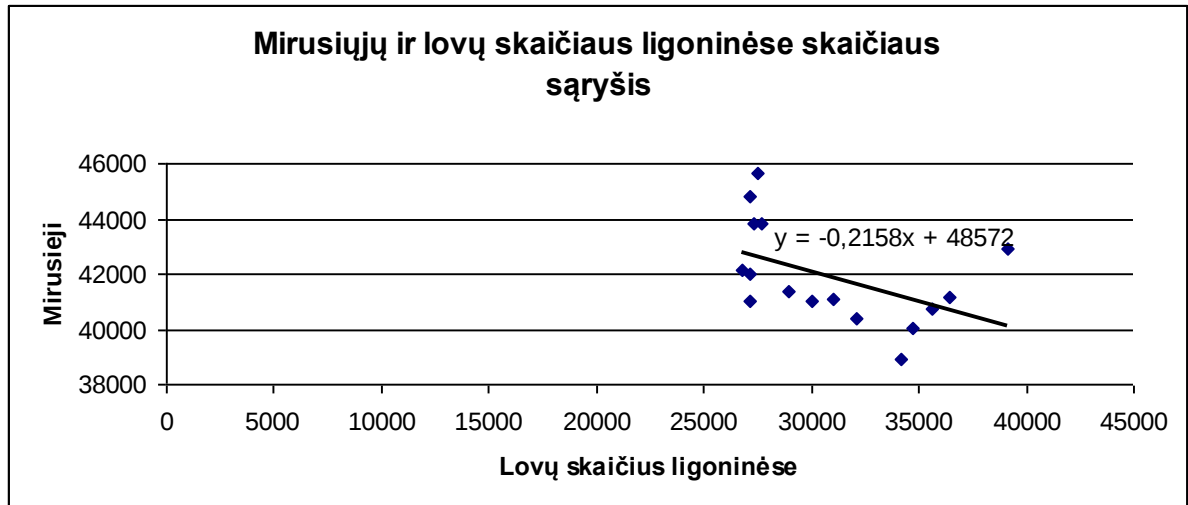
Antroji priklausomybė – lovų skaičius ligoninėse ir mirusiųjų skaičius 1996–2011 metais. Kintamasis, kurio reikšmės bus prognozuojamos (mirusiųjų gyventojų skaičius), vadinamas priklausomu kintamuoju, o kintamasis, pagal kurio reikšmes pronažuosime (lovų skaičius ligoninėse) – nepriklausomu kintamuoju.

Pagal SPSS programos pagalba gautus duomenis matome (8 priedas), kad koeficientas $a = -0,216$, regresijos kostanta $b = 48572,056$.

Galime parašyti regresijos tiesės lygtį (pagal (4.1))

$$y = -0,216x + 48572,056$$

ir ją nubrėžti (7.2 pav.).



7.2 pav.

Apskaičiuojame determinacijos koeficientą (koreliacijos koeficiento kvadratas):

$$R^2 = 0,479^2 = 0,229 .$$

Nagrinėjamu atveju matome, kad $R^2 = 0,229$, t. y. $0,229 < 0,25$, todėl šiuo regresijos modeliu pasikliauti negalima.

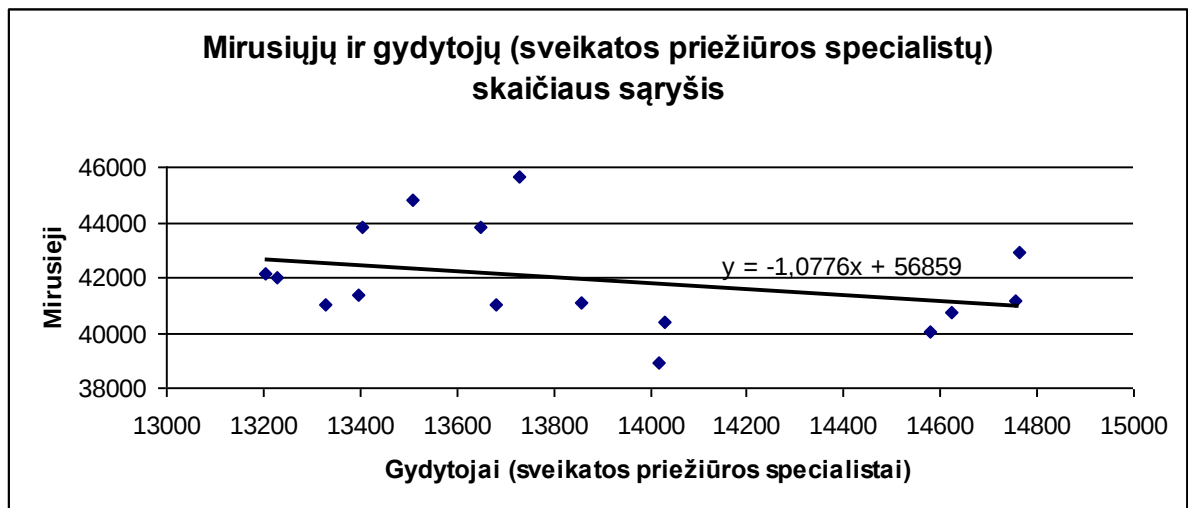
Trečioji priklausomybė – gydytojų (sveikatos priežiūros specialistų) skaičius ir mirusiųjų skaičius 1996–2011 metais. Kintamasis, kurio reikšmės bus prognozuojamos (mirusiųjų gyventojų skaičius), vadinamas priklausomu kintamuoju, o kintamasis, pagal kurio reikšmes pronažuosime (gydytojų (sveikatos priežiūros specialistų) skaičius) – nepriklausomu kintamuoju.

Pagal SPSS programos pagalba gautus duomenis matome (9 priedas), kad $a = -1,078$, $b = 56858,865$.

Galime parašyti regresijos tiesės lygtį (pagal (4.1))

$$y = -1,078x + 56858,865$$

ir ją nubrėžti (7.3 pav.).



7.3 pav.

Apskaičiuojame determinacijos koeficientą (koreliacijos koeficiento kvadratas):

$$R^2 = 0,324^2 = 0,105 .$$

Nagrinėjamu atveju matome, kad $R^2 = 0,105$, t. y. $0,105 < 0,25$, todėl šiuo regresijos modeliu pasikliauti negalima.

7.2. DAUGIALYPĖ REGRESIJA

Taipogi galima parašyti ir daugialypės regresijos modelį. Atlikus testą SPSS programa, gauname koeficientus, reikalingus modeliui, kuris nurodo mirusiųjų gyventojų skaičiaus priklausomybę nuo dviejų nepriklausomų kintamųjų – gydytojų (sveikatos priežiūros specialistų) skaičiaus ir lovų skaičiaus ligoninėse.

Regresijos konstanta $a = 5381,836$, koeficientas prie pirmo nepriklausomo kintamojo (gydytojų (sveikatos priežiūros specialistų) skaičiaus) $b_1 = 4,366$, koeficientas prie antro nepriklausomo kintamojo (lovų skaičiaus ligoninėse) $b_2 = -0,778$ (duomenys gauti programa SPSS, 10 priedas).

Daugialypės regresijos lygtis (pagal (4.9)), kur x_1 – gydytojų (sveikatos priežiūros specialistų) skaičius, o x_2 – lovų skaičius ligoninėse, atrodo šitaip:

$$y = 5381,836 + 4,366 x_1 - 0,778 x_2.$$

8. PROGNOZAVIMAS LAIKO EILUTĖMIS

Tyrime naudojamas mirusiųjų gyventojų skaičius 1996–2011 metais. Galima bandyti prognozuoti, koks bus mirusiųjų skaičius 2013 metais. Pirmiausia patikrinsime mirusiųjų skaičiaus prognozę 2012 metams ir palyginsime ją su Lietuvos statistikos departamente nurodytu skaičiumi.

Pasižymime duomenis:

X_t – mirusiųjų gyventojų skaičius momentu t .

t – momentai (1996–2011 metai).

Apsiskaičiuojame $\bar{X}_t$, $\bar{t}$, $\bar{t}^2$, $\overline{X_t t}$:

$$\begin{aligned}\bar{X}_t &= (42896 + 41143 + 40757 + 40003 + 38919 + 40399 + 41072 + 40990 + 41340 + \\ &+ 43799 + 44813 + 45624 + 43832 + 42032 + 42120 + 41037) \cdot \frac{1}{16} = 41923,5;\end{aligned}$$

$$\bar{t} = (1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14 + 15 + 16) \cdot \frac{1}{16} = 8,5;$$

$$\begin{aligned}\bar{t}^2 &= (1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 + 100 + 121 + 144 + 169 + 196 + 225 + \\ &+ 256) \cdot \frac{1}{16} = 93,5;\end{aligned}$$

$$\begin{aligned}\overline{X_t t} &= ((42896 \cdot 1) + (41143 \cdot 2) + (40757 \cdot 3) + (40003 \cdot 4) + (38919 \cdot 5) + (40399 \cdot 6) + \\ &+ (41072 \cdot 7) + (40990 \cdot 8) + (41340 \cdot 9) + (43799 \cdot 10) + (44813 \cdot 11) + (45624 \cdot 12) + \\ &+ (43832 \cdot 13) + (42032 \cdot 14) + (42120 \cdot 15) + (41037 \cdot 16)) \cdot \frac{1}{16} = 359813,44.\end{aligned}$$

Apskaičiuojame Δ , Δa , Δb :

$$\Delta = \begin{vmatrix} \overline{t^2} & \bar{t} \\ \bar{t} & 1 \end{vmatrix} = \begin{vmatrix} 93,5 & 8,5 \\ 8,5 & 1 \end{vmatrix} = 93,5 \cdot 1 - 8,5 \cdot 8,5 = 21,25;$$

$$\Delta a = \begin{vmatrix} \overline{X_t t} & \bar{t} \\ \bar{X}_t & 1 \end{vmatrix} = \begin{vmatrix} 359813,44 & 8,5 \\ 41923,5 & 1 \end{vmatrix} = 359813,44 \cdot 1 - 8,5 \cdot 41923,5 = 3463,69;$$

$$\Delta b = \begin{vmatrix} \overline{t^2} & \overline{X_t t} \\ \bar{t} & \bar{X}_t \end{vmatrix} = \begin{vmatrix} 93,5 & 359813,44 \\ 8,5 & 41923,5 \end{vmatrix} = 93,5 \cdot 41923,5 - 359813,44 \cdot 8,5 = 861433,01.$$

Tuomet galime rasti įverčius $\hat{a}$ ir $\hat{b}$:

$$\hat{a} = \frac{\Delta a}{\Delta} = \frac{3463,69}{21,25} = 162,9972;$$

$$\hat{b} = \frac{\Delta b}{\Delta} = \frac{861433,01}{21,25} = 40538,02.$$

Turėdami įverčius galime parašyti tiesinę prognozavimo funkciją:

$$\hat{m}_t = \hat{a}t + \hat{b}.$$

Imdami $t = 17$, apskaičiuojame mirusiųjų gyventojų skaičių 2012 metams:

$$\hat{m}_{17} = 162,9972 \cdot 17 + 40538,02 = 43308.$$

Matome, kad Lietuvos statistikos departamento svetainėje nurodytas 2012 metais mirusių gyventojų skaičius yra 40938. Skirtumas tarp tikrojo skaičiaus ir rastosios reikšmės – 2370.

Vis dėlto galime apskaičiuoti, kiek apytiksliai mirusių gyventojų bus 2013 metais.

$$\hat{m}_{18} = 162,9972 \cdot 18 + 40538,02 = 43471.$$

IŠVADOS

Atlikus tyrimą ir išnagrinėjus tris priklausomybes, galime padaryti tokias išvadas:

- Vidutinis gyventojų skaičius Lietuvoje 2002–2011 m. statistiškai reikšmingai didesnis nei 3 mln. gyventojų.
- Vidutinė natūrali gyventojų kaita Lietuvoje 1996–2011 m. statistiškai reikšmingai nesiskiria nuo –7000.
- Vidutinis mirštančių gyventojų skaičius Lietuvoje 1996–2011 m. yra statistiškai reikšmingai didesnis nei 40000 gyventojų.
- Vidutinis per parą gimusių kūdikių skaičiaus vidurkis statistiškai reikšmingai skiriasi nuo per parą mirusių gyventojų skaičiaus vidurkio.
- Negyvagimių vyrų ir negyvagimių moterų skaičiaus vidurkiai skiriasi statistiškai reikšmingai.
- 1000 gyventojų tenkančių mirusių skaičiaus ir 1000 gimusių tenkančių mirusių kūdikių skaičiaus vidurkiai skiriasi statistiškai reikšmingai.
- Vidutinis gydytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) ir vidutinis slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius skiriasi statistiškai reikšmingai.
- Gimstamumo ir mirtingumo kasmetiniai vidurkiai skiriasi statistiškai reikšmingai.
- Nėra priklausomybės tarp išlaidų sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičiaus 2008–2011 metais.
- Egzistuoja silpna neigiama priklausomybė tarp lovų skaičiaus ligoninėse ir mirusiųjų gyventojų skaičiaus 1996–2011 metais.
- Egzistuoja silpna neigiama priklausomybė tarp mirusiųjų ir gydytojų (sveikatos priežiūros specialistų) skaičiaus 1996–2011 metais.
- Sudaryti tiesinės regresijos modeliai nėra reikšmingi, taigi galime teigti, kad mirtingumas Lietuvoje yra natūralus procesas, kuriam niekas neturi įtakos.

SUMMARY

Investigation of relations between people mortality and health protection system

People mortality in Lithuania is increasing every year, so we are going to try to identify, what things make the biggest influence. Investigated factors of health protection system are: expenses on health care, quantity of doctors (health care professionals) and number of hospital beds in Lithuania. In Bachelor thesis, we are searching for a dependence between people mortality and those 3 factors of health protection system mentioned before. This investigation is based on data from 1996–2011 (www.stat.gov.lt). In this work we are testing several hypothesis and for investigation we use correlation analysis and linear regression. Also trying to predict people mortality in the near future.

LITERATŪRA

- [1] J. Bagdanavičius, V. Šiaudytis, A. Vaitelienė, *Statistikos metodai socialiniuose – ekonominiuose tyrimuose*, Vilnius: Vilniaus pedagoginio universiteto leidykla, 2007.
- [2] A. Bakštyš, T. Tamošiūnas, *Tikimybė ir statistika*, Šiauliai: Šiaulių universitetas, 1997.
- [3] K. Bučys, *Tikimybių teorijos ir matematinės statistikos praktikumas*, Klaipėda: Klaipėdos universitetas, 2003.
- [4] V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai*, I, Vilnius: TEV, 2003.
- [5] V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai*, II, Vilnius: TEV, 2004.
- [6] R. Eidukevičius, D. Juknevičienė, N. Kosareva, S. Pamerneckis, *Matematinė statistika istorijoje*, Vilnius: Vilniaus universiteto leidykla, 1998.
- [7] V. Kanišauskas, *Tikimybių teorijos ir matematinės statistikos pagrindai*, Šiauliai: Šiaulių universiteto leidykla, 2000.
- [8] J. Kruopis, *Matematinė statistika*, Vilnius: Mokslo ir enciklopedijų leidykla, 1993.
- [9] T. Leonavičienė, *SPSS programų paketo taikymas statistiniuose tyrimuose*, Vilnius: Vilniaus pedagoginio universiteto leidykla, 2007.
- [10] D. Lukaševičiūtė, *Regresiniai modeliai išgyvenamumo analizėje ir jų taikymas ligonių, sergančių reumatoidiniu artritu, mirtingumo analizei*, Magistro baigiamasis darbas, VU, Vilnius, 2009.
- [11] H. Markšaitis, O. Navickienė, *Tikimybių teorija ir matematinė statistika*, Vilnius: Mykolo Romerio universitetas, 2012.
- [12] M. Noreika, *Mirtingumo nuo galvos smegenų insulto prognozavimo modeliai ir programinės priemonės*, Magistro darbas, KTU, Kaunas, 2007.
- [13] A. Pabedinskaitė, *Kiekybiniai sprendimų metodai I dalis*, Vilnius: Technika, 2005.
- [14] VL. Stankūnienė, D. Jasilionis, *Lietuvos gyventojų mirtingumo lygio ir mirties priežasčių struktūros prieštaringi pokyčiai*, Filosofija. Sociologija. T. 22. Nr. 4, p. 334–347, 2011.

[15] *Tyrimas: Lietuvoje mirtingumas didžiausias, rūpinimasis sveikata – mažiausias*
[interaktyvus] [žiūrėta 2013 kovo 15 d.]. Prieiga per internetą:
<http://zebra.15min.lt/lt/laikas_sau/sveikata/tyrimas-lietuvoje-mirtingumas-didziausias-rupinimasis-sveikata-maziausias-220672.html>

PRIEDAI

1 priedas. Per parą gimusių kūdikių ir mirusių žmonių skaičius Lietuvoje bei gyventojų skaičius šalyje metų pradžioje (2002–2011 metais).

Metai	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Gimė per parą	82	84	83	84	86	89	96	100	98	94
Mirė per parą	113	112	113	120	123	125	120	115	115	112
Viso gyventojų metų pradžioje	3452603	3428806	3395091	3349392	3280076	3239560	3202564	3176333	3137026	3052588

2 priedas. Mirusių kūdikių, negyvagimių (vyrų ir moterų) bei mirusiųjų skaičius Lietuvoje ir natūrali gyventojų kaita šalyje (1996–2011 metais).

Metai	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Mirusių kūdikių skaičius	395	391	343	315	294	250	238	206	240	209	213	190	172	181	153	144
Negyvagimiai	236	248	240	207	221	167	193	168	150	152	137	161	164	151	147	148
Vyrai	121	135	126	112	124	96	138	127	100	81	73	91	82	77	73	75
Moterys	115	113	114	95	97	71	55	41	50	71	64	70	82	74	74	73
Mirusieji	42896	41143	40757	40003	38919	40399	41072	40990	41340	43799	44813	45624	43832	42032	42120	41037
1000 gyventojų tenka mirusių	11,9	11,5	11,5	11,3	11,1	11,6	11,8	11,9	12,0	12,8	13,2	13,5	13,1	12,6	12,8	13,5
Gimusiai	39066	37812	37019	36415	34149	31546	30014	30598	30419	30541	31265	32346	35065	36682	35626	34385
1000 gyventojų tenka gimusių	10,8	10,6	10,4	10,3	9,8	9,1	8,6	8,9	8,8	8,9	9,2	9,6	10,5	11,0	10,8	11,3
1000 gimusių tenka mirusių kūdikių	10,1	10,3	9,2	8,6	8,5	7,8	7,9	6,8	7,9	6,9	6,8	5,9	5,0	5,0	4,3	4,2
Natūrali gyventojų kaita	-3830	-3331	-3738	-3588	-4770	-8853	-11058	-10392	-10921	-13258	-13548	-13278	-8767	-5350	-6494	-6652

3 priedas. Gydytojų (sveikatos priežiūros specialistų), gydytojų (praktikuojančių sveikatos priežiūros specialistų) bei slaugytojų (sveikatos priežiūros specialistų, tenkančių 10000 gyventojų) skaičius Lietuvoje (1996–2011 metais).

Metai	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Gydytojas. Sveikatos priežiūros specialistų skaičius	14763	14757	14622	14578	14019	14031	13856	13682	13397	13650	13510	13729	13403	13228	13203	13327
Gydytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius												12551	12413	12191	12226	12407
Gydytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	39,8	39,8	39,5	39,4	40,2	40,4	40,0	39,7	39,1	40,1	39,9	40,8	40,0	39,7	40,7	41,7
Gydytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų												37,3	37,1	36,6	37,7	38,8
Slaugytojas. Sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų	4,6	5,8	6,1	6,2	7,1	7,2	6,7	6,9	6,6	7,2	6,6	7,1	6,8	7,2	7,8	8,0
Slaugytojas. Praktikuojančių sveikatos priežiūros specialistų skaičius, tenkantis 10000 gyventojų												6,9	6,6	7,1	7,6	7,8

4 priedas. Lovų skaičius ligoninėse Lietuvoje (1996–2011 metais).

Metai	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Lovų skaičius ligoninėse	39182	36442	35615	34714	34145	32104	31031	29990	28972	27727	27114	27476	27358	27155	26804	27175
10 000 gyventojų tenka lovų ligoninėse	105,7	98,4	96,2	93,9	97,9	92,4	89,6	87,0	84,6	81,5	81,1	81,6	81,7	81,6	82,6	84,9
Lovų skaičius ligoninėse (be slaugos lovų)												23233	22958	22719	22190	22549
10 000 gyventojų tenka lovų ligoninėse (be slaugos lovų)												69,0	68,5	68,2	68,4	70,5

5 priedas. Išlaidos sveikatos priežiūrai, mln. Lt (2008–2011 metais).

Metai	2008	2009	2010	2011
Gydymo paslaugos	3476,9	3317,8	3217,1	3477,0
Reabilitacinė priežiūra	253,5	237,5	220,0	296,3
Ilgalaikės slaugos paslaugos	520,3	625,3	589,9	564,1
Pagalbinės sveikatos priežiūros paslaugos	433,8	412,4	383,3	414,8
Medicinos prekės ambulatoriniams ligoniams	2103,7	2027,6	1986,8	2035,5
Prevencija ir visuomenės sveikatos paslaugos	86,8	78,4	56,7	77,8
Sveikatos priežiūros administravimas ir sveikatos draudimas	222,8	141,9	141,9	148,2
Einamosios sveikatos priežiūros išlaidos	7097,9	6840,9	6595,8	7013,6
Su sveikata susijusios funkcijos	3143,0	3175,6	3150,1	3404,9
Sveikatos priežiūros paslaugas teikiančių institucijų kapitalo formavimas	298,0	91,6	184,7	313,2
Bendrosios sveikatos priežiūros išlaidos	7395,9	6932,4	6780,5	7326,8
Viso išleista:	25032,6	23881,4	23306,8	25072,2

6 priedas. Duomenų normališkumo tikrinimas programa SPSS.

One-Sample Kolmogorov-Smirnov Test

		Per para gime
N		10
Normal Parameters ^a	Mean	89.60
	Std. Deviation	6.802
Most Extreme Differences	Absolute	.202
	Positive	.202
	Negative	-.141
Kolmogorov-Smirnov Z		.638
Asymp. Sig. (2-tailed)		.811
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Per para mire
N		10
Normal Parameters ^a	Mean	116.80
	Std. Deviation	4.803
Most Extreme Differences	Absolute	.246
	Positive	.246
	Negative	-.159
Kolmogorov-Smirnov Z		.778
Asymp. Sig. (2-tailed)		.580
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Gyventoju skaicius metu pradžioje
N		10
Normal Parameters ^a	Mean	3271403.90
	Std. Deviation	133247.323
Most Extreme Differences	Absolute	.123
	Positive	.097
	Negative	-.123
Kolmogorov-Smirnov Z		.390
Asymp. Sig. (2-tailed)		.998
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Mirusiu kudikiu skaicius
N		16
Normal Parameters ^a	Mean	245.88
	Std. Deviation	79.902
Most Extreme Differences	Absolute	.167
	Positive	.167
	Negative	-.101
Kolmogorov-Smirnov Z		.668
Asymp. Sig. (2-tailed)		.764
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Negyvagimiai
N		16
Normal Parameters ^a	Mean	180.62
	Std. Deviation	37.727
Most Extreme Differences	Absolute	.256
	Positive	.256
	Negative	-.124
Kolmogorov-Smirnov Z		1.024
Asymp. Sig. (2-tailed)		.245
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Negyvagimiai vyrai
N		16
Normal Parameters ^a	Mean	101.94
	Std. Deviation	23.901
Most Extreme Differences	Absolute	.173
	Positive	.173
	Negative	-.162
Kolmogorov-Smirnov Z		.692
Asymp. Sig. (2-tailed)		.725
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Negyvagimes moterys
N		16
Normal Parameters ^a	Mean	78.69
	Std. Deviation	22.576
Most Extreme Differences	Absolute	.207
	Positive	.207
	Negative	-.123
Kolmogorov-Smirnov Z		.829
Asymp. Sig. (2-tailed)		.498
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Mirusieji
N		16
Normal Parameters ^a	Mean	41923.50
	Std. Deviation	1824.483
Most Extreme Differences	Absolute	.188
	Positive	.188
	Negative	-.098
Kolmogorov-Smirnov Z		.752
Asymp. Sig. (2-tailed)		.624
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		1000 gyventoju tenka mirusiu
N		16
Normal Parameters ^a	Mean	12.256
	Std. Deviation	.8066
Most Extreme Differences	Absolute	.187
	Positive	.187
	Negative	-.125
Kolmogorov-Smirnov Z		.749
Asymp. Sig. (2-tailed)		.630
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Gimusieji
N		16
Normal Parameters ^a	Mean	33934.25
	Std. Deviation	2996.697
Most Extreme Differences	Absolute	.162
	Positive	.162
	Negative	-.109
Kolmogorov-Smirnov Z		.649
Asymp. Sig. (2-tailed)		.793
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		1000 gyventoju tenka gimusiu
N		16
Normal Parameters ^a	Mean	9.919
	Std. Deviation	.9094
Most Extreme Differences	Absolute	.162
	Positive	.160
	Negative	-.162
Kolmogorov-Smirnov Z		.650
Asymp. Sig. (2-tailed)		.792
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		1000 gimusiu tenka mirusiu kudikiu
N		16
Normal Parameters ^a	Mean	7.200
	Std. Deviation	1.9356
Most Extreme Differences	Absolute	.122
	Positive	.122
	Negative	-.122
Kolmogorov-Smirnov Z		.489
Asymp. Sig. (2-tailed)		.971
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Naturali gyventoju kaita
N		16
Normal Parameters ^a	Mean	-7989.25
	Std. Deviation	3737.761
Most Extreme Differences	Absolute	.140
	Positive	.115
	Negative	-.140
Kolmogorov-Smirnov Z		.559
Asymp. Sig. (2-tailed)		.914
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Gydytojas. Sveikatos priežiūros specialistu skaičius
N		16
Normal Parameters ^a	Mean	13859.69
	Std. Deviation	549.214
Most Extreme Differences	Absolute	.157
	Positive	.157
	Negative	-.155
Kolmogorov-Smirnov Z		.626
Asymp. Sig. (2-tailed)		.828
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Gydytojas. Praktikuojanciu sveikatos prieziuros specialistu skaicius
N		5
Normal Parameters ^a	Mean	12357.60
	Std. Deviation	148.313
Most Extreme Differences	Absolute	.230
	Positive	.213
	Negative	-.230
Kolmogorov-Smirnov Z		.515
Asymp. Sig. (2-tailed)		.953
a. Test distribution is Normal.		

--	--

One-Sample Kolmogorov-Smirnov Test

		Gydytojas. Sveikatos prieziuros specialistu skaicius, tenkantis 10000 gyventoju
N		16
Normal Parameters ^a	Mean	40.050
	Std. Deviation	.6261
Most Extreme Differences	Absolute	.157
	Positive	.157
	Negative	-.101
Kolmogorov-Smirnov Z		.627
Asymp. Sig. (2-tailed)		.826
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Gydytojas. Praktikuojanciu sveikatos prieziuros specialistu skaicius, tenkantis 10000 gyventoju
N		5
Normal Parameters ^a	Mean	37.500
	Std. Deviation	.8276
Most Extreme Differences	Absolute	.205
	Positive	.205
	Negative	-.142
Kolmogorov-Smirnov Z		.457
Asymp. Sig. (2-tailed)		.985
a. Test distribution is Normal.		

--	--

One-Sample Kolmogorov-Smirnov Test

		Slaugytojas. Sveikatos prieziuros specialistu skaicius, tenkantis 10000 gyventoju
N		16
Normal Parameters ^a	Mean	6.744
	Std. Deviation	.8099
Most Extreme Differences	Absolute	.180
	Positive	.162
	Negative	-.180
Kolmogorov-Smirnov Z		.718
Asymp. Sig. (2-tailed)		.681
a. Test distribution is Normal.		

--	--

One-Sample Kolmogorov-Smirnov Test

		Slaugytojas. Praktikuojanciu sveikatos prieziuros specialistu skaicius, tenkantis 10000 gyventoju
N		5
Normal Parameters ^a	Mean	7.200
	Std. Deviation	.4950
Most Extreme Differences	Absolute	.190
	Positive	.180
	Negative	-.190
Kolmogorov-Smirnov Z		.426
Asymp. Sig. (2-tailed)		.993
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Lovu skaicius ligoninese
N		16
Normal Parameters ^a	Mean	30812.75
	Std. Deviation	4048.477
Most Extreme Differences	Absolute	.215
	Positive	.215
	Negative	-.161
Kolmogorov-Smirnov Z		.858
Asymp. Sig. (2-tailed)		.453
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		10000 gyventoju tenka lovu ligoninese
N		16
Normal Parameters ^a	Mean	88.794
	Std. Deviation	7.7294
Most Extreme Differences	Absolute	.193
	Positive	.193
	Negative	-.160
Kolmogorov-Smirnov Z		.771
Asymp. Sig. (2-tailed)		.592
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Lovu skaicius ligoninese (be slaugos lovu)
N		5
Normal Parameters ^a	Mean	22729.80
	Std. Deviation	396.697
Most Extreme Differences	Absolute	.124
	Positive	.113
	Negative	-.124
Kolmogorov-Smirnov Z		.278
Asymp. Sig. (2-tailed)		1.000
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		10000 gyventoju tenka lovu ligoninese (be slaugos lovu)
N		5
Normal Parameters ^a	Mean	68.920
	Std. Deviation	.9311
Most Extreme Differences	Absolute	.274
	Positive	.274
	Negative	-.220
Kolmogorov-Smirnov Z		.613
Asymp. Sig. (2-tailed)		.847
a. Test distribution is Normal.		

One-Sample Kolmogorov-Smirnov Test

		Sveikatos prieziuros islaidos
N		4
Normal Parameters ^a	Mean	24323.250
	Std. Deviation	874.1673
Most Extreme Differences	Absolute	.291
	Positive	.196
	Negative	-.291
Kolmogorov-Smirnov Z		.583
Asymp. Sig. (2-tailed)		.886
a. Test distribution is Normal.		

7 priedas. Išlaidų sveikatos priežiūrai (mln. Lt) ir mirusiųjų skaičiaus sąryšio tikrinimas programa SPSS.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.145 ^a	.021	-.468	1405.902

a. Predictors: (Constant), Sveikatos priežiūros išlaidos

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	85286.874	1	85286.874	.043	.855 ^a
Residual	3953119.876	2	1976559.938		
Total	4038406.750	3			

a. Predictors: (Constant), Sveikatos priežiūros išlaidos

b. Dependent Variable: Mirusieji 2008 - 2011 metais

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	37563.796	22596.005		1.662	.238	-59658.966	134786.557
	Sveikatos priežiūros išlaidos	.193	.929	.145	.208	.855	-3.802	4.188

a. Dependent Variable: Mirusieji 2008 - 2011 metais

8 priedas. Mirusiųjų ir lovų skaičiaus ligoninėse skaičiaus sąryšio tikrinimas programa SPSS.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.479 ^a	.229	.174	1657.983

a. Predictors: (Constant), Lovu skaicius ligoninese

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	1.145E7	1	1.145E7	4.164	.061 ^a
Residual	3.848E7	14	2748907.895		
Total	4.993E7	15			

a. Predictors: (Constant), Lovu skaicius ligoninese

b. Dependent Variable: Mirusieji

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	48572.056	3284.426		14.789	.000	41527.663	55616.448
	Lovu skaicius ligoninese	-.216	.106	-.479	-2.041	.061	-.443	.011

a. Dependent Variable: Mirusieji

9 priedas. Mirusiųjų ir gydytojų (sveikatos priežiūros specialistų) skaičiaus sąryšio tikrinimas programa SPSS.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.324 ^a	.105	.041	1786.397

a. Predictors: (Constant), Gydytojas. Sveikatos priežiūros specialistu skaičius

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5254119.711	1	5254119.711	1.646	.220 ^a
	Residual	4.468E7	14	3191212.592		
	Total	4.993E7	15			

a. Predictors: (Constant), Gydytojas. Sveikatos priežiūros specialistu skaičius

b. Dependent Variable: Mirusieji

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	56858.865	11648.317		4.881	.000	31875.711	81842.020
	Gydytojas. Sveikatos priežiūros specialistu skaičius	-1.078	.840	-.324	-1.283	.220	-2.879	.724

a. Dependent Variable: Mirusieji

10 priedas. Mirusiųjų ir gydytojų (sveikatos priežiūros specialistų) skaičiaus bei lovų skaičiaus ligoninėse sąryšio tikrinimas programa SPSS.

Coefficients ^a									
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	5381.836	22635.500		.238	.816			
	Gydytoju skaicius	4.366	2.268	1.314	1.925	.076	-.324	.471	.414
	Lovu ligoninese skaicius	-.778	.308	-1.726	-2.529	.025	-.479	-.574	-.543

a. Dependent Variable: Mirusiųjų skaicius