VILNIAUS UNIVERSITETAS
ŠIAULIŲ AKADEMIJA

MATEMATIKOS MAGISTRO STUDIJŲ PROGRAMA
Didžiųjų duomenų analitikos specializacija

IEVA BELECKAITĖ

Magistro studijų baigiamasis darbas

LIETUVOS GYVENTOJŲ SERGAMUMO PSICHIKOS IR ELGESIO SUTRIKIMAIS XXI A. PRADŽIOJE STATISTINĖ ANALIZĖ

Darbo vadovė: lekt. dr. Karolina Kanišauskienė

Šiauliai, 2022

Studijuojančiojo, teikiančio baigiamąjį darbą, GARANTIJA

WARRANTY of Final Thesis

Vardas, pavardė *Name, Surname*	**Ieva Beleckaitė**
Padalinys *Faculty*	**Šiaulių akademija** ***Šiauliai Academy***
Studijų programa *Study Programme*	**Matematika** ***Mathematics***
Darbo pavadinimas *Thesis topic*	**Lietuvos gyventojų sergamumo psichikos ir elgesio sutrikimais XXI a. pradžioje statistinė analizė** **The Statistical Analysis of the Morbidity of Mental and Behavioural Disorders of Lithuania's Population in the Early 21st Century**
Darbo tipas *Thesis type*	**Baigiamasis darbas** ***Final Thesis***

Garantuoju, kad mano baigiamasis darbas yra parengtas sąžiningai ir savarankiškai, kitų asmenų indėlio į parengtą darbą nėra. Jokių neteisėtų mokėjimų už šį darbą niekam nesu mokėjęs.

I guarantee that my thesis is prepared in good faith and independently, there is no contribution to this work from other individuals. I have not made any illegal payments related to this work.

Šiame darbe tiesiogiai ar netiesiogiai panaudotos kitų šaltinių citatos yra pažymėtos literatūros nuorodose.

Quotes from other sources directly or indirectly used in this thesis, are indicated in literature references.

Aš, Ieva Beleckaitė, pateikdama šį darbą, patvirtinu (*pažymėti*)
I, Ieva Beleckaitė, by submitting this paper confirm (check) ☒

Embargo laikotarpis
Embargo Period

Prašau nustatyti šiam baigiamajam darbui toliau nurodytos trukmės embargo laikotarpį:
I am requesting an embargo of this thesis for the period indicated below:

☐ __________ mėnesių / *months*
(embargo laikotarpis negali viršyti 60 mėn. / *an embargo period shall not exceed 60 months*).

☒ Embargo laikotarpis nereikalingas / *no embargo requested.*

Embargo laikotarpio nustatymo priežastis / *Reason for embargo period*:

TURINYS

ĮVADAS

Temos aktualumas. Žmonių sergamumas skirtingomis ligomis yra nagrinėjamas įvairiai. Psichikos sutrikimas, psichikos liga – tam tikras psichikos (mąstymo, elgsenos, jausmų, nuotaikos) sutrikimas. Žodžiai „psichika“, „psichologas“ ar „psichoterapeutas“ dažnam Lietuvos (ir ne tik) gyventojui sukelia baimę, nepasitikėjimą, nežinomybės jausmą. Šių sutrikimų-ligų atsiradimo priežastys gali būti įvairios: paveldimumo veiksniai, neigiami išgyvenimai (stresas, prievarta), biologiniai faktoriai, trauminis smegenų sužalojimas, sunkios ligos, socialinė atskirtis ar psichoaktyviųjų medžiagų, alkoholio vartojimas.

Kiekvienam iš mūsų vis dažniau tenka susidurti su psichikos ir elgesio sutrikimų susirgimų požymiais: nerimas, neigiamų emocijų padažnėjimas (pvz.: dėl pandemijos šalyje, dėl karo Ukrainoje), vis dažniau patiriamas stresas, panikos priepuoliai, depresija ir kiti nuotaikos sutrikimai. Šie išgyvenimai ar susirgimas psichikos ir elgesio sutrikimais nepriklauso nei nuo intelekto lygio, nei nuo išsilavinimo, nei nuo socialinės padėties, amžiaus, lyties ar kitų veiksnių. Turintieji psichikos ir elgesio sutrikimų žmonės dažniausiai yra diskriminuojami, nedrįsta kreiptis pagalbos į specialistus, todėl ypač svarbu rūpintis tiek savo, tiek artimųjų psichine sveikata.

Magistro darbe nagrinėjamas Lietuvos gyventojų sergamumas psichikos ir elgesio sutrikimais nuo 2006 m. iki 2019 m. pasitelkiant kasmetinius Higienos instituto Sveikatos informacijos centro leidinius „Lietuvos gyventojų sveikata ir sveikatos priežiūros įstaigų veikla“. Remiantis Tarptautine psichikos ir elgesio sutrikimų klasifikacija TLK-10-AM nagrinėjamas ligų klasifikatoriaus efektyvumas renkant ir apdorojant Lietuvos gyventojų medicininius duomenis. TLK-10-AM ligų klasifikacija buvo patvirtinta 1989 m.

Pasitelkiant sudėtingesnius matematinės statistikos metodus Lietuvos žmonių sergamumo psichikos ir elgesio sutrikimais dinamika TLK-10-AM ligų klasifikatoriaus pagalba XXI a. pradžioje tirta nebuvo, todėl šis tyrimas yra ypač aktualus šiomis dienomis. Šią temą pasirinkau todėl, jog mano artimas žmogus sirgo depresija ir todėl nusprendė pasitraukti iš gyvenimo pats.

Tyrimo objektas. Statistiniai ir demografiniai duomenys iš Higienos instituto Sveikatos informacijos centro išleistų periodinių sveikatos statistikos leidinių „Lietuvos gyventojų sveikata ir sveikatos priežiūros įstaigų veikla“ apie sergančiuosius psichikos ir elgesio sutrikimais žmonių skaičius Lietuvoje.

Chronologinės ribos. 2006–2019 metai.

Tyrimo tikslas: Statistiniais metodais ištirti sergančiųjų psichikos ir elgesio sutrikimais skaičiaus dinamiką Lietuvoje nagrinėjamu laikotarpiu.

Tyrimo uždaviniai:

1. Ištirti Lietuvos sergančiųjų psichikos ir elgesio sutrikimais skaičiaus priklausomybę nuo gyventojų skaičiaus panaudojant koreliacinę ir regresinę analizę.
2. Taikant statistines hipotezes apie polinominio skirstinio parametrų reikšmes, nustatyti, ar nagrinėjamų 7 ligų-grupių sergančiųjų skaičiai kardinaliai pasikeičia praėjus tam tikram laikotarpiui, jei taip, nurodyti po kokio laikotarpio (po kelerių metų).
3. Ištirti sergančiųjų skaičiaus pokyčių dinamiką kasmet ir kas ketverius metus, tai aprašyti ir pavaizduoti vizualiai.
4. Taikant serijų kriterijų apie duomenų atsitiktinumą ir nepriklausomumą, nustatyti, kuriems sergančiųjų psichikos ir elgesio sutrikimais duomenims galima taikyti laiko eilučių metodus.
5. Taikant laiko eilučių išlyginimo metodą, nustatyti, kurios duomenų eilutės turi trendą, o kurios ne.
6. Taikant laiko eilučių metodus rasti AR(1) ir AR(2) modelių parametrų įverčius panaudojant trendą ir jo nepanaudojant.

Tyrimo metodai. Taikomi tikimybiniai, matematinės statistikos, lyginamosios analizės, vizualinis-dedukcinis ir laiko eilučių metodai.

Darbo struktūra. Magistro darbą sudaro įvadas, du skyriai, išvados, literatūros sąrašas, santrauka lietuvių ir anglų kalbomis, 8 priedai.

Pirmajame skyriuje pateikiama pagrindinė teorija ir sąvokos, kurios bus naudojamos magistro darbe. Antrajame skyriuje atliekamas tyrimas. Magistro darbo pabaigoje pateikiamas literatūros sąrašas, santrauka lietuvių ir anglų kalbomis, priedai.

1. TEORINĖ DALIS

1.1 PAGRINDINĖS MATEMATINĖS STATISTIKOS SĄVOKOS

Pagrindinės matematinės sąvokos, formulės ir apibrėžimai paimti iš [4] šaltinio.

Apibrėžimas. Imtis – stebėtų rezultatų sąrašas.

Apibrėžimas. Variacinė eilutė – imties reikšmių išdėstymas didėjančia tvarka. T. y. ${x_1}^* \leq {x_2}^* \leq \ldots \leq {x_n}^*$; čia $x_{min} = {x_1}^*$ – mažiausias elementas, o $x_{max} = {x_n}^*$ – didžiausias imties elementas.

Apibrėžimas. Empirinės imties charakteristikos – empirinio skirstinio skaitinės charakteristikos (vidurkis, dispersija).

Apibrėžimas. Vidurkis – tai vidutinė statistinės eilutės reikšmė apskaičiuojama formule:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,;$$

čia n – imties dydis, $x_1,\ x_2, \ldots, x_n$ – statistinės eilutės elementai.

Apibrėžimas. Mediana – vidurinė variacinės eilutės reikšmė, apskaičiuojama formule:

$$M_e = \begin{cases} X_{\left(\frac{n+1}{2}\right)},\ jei\ n = 2l + 1, \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right),\ jei\ n = 2l. \end{cases} \tag{1}$$

Apibrėžimas. Imties dispersija parodo duomenų sklaidą apie vidurkį ir randama pagal formulę:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Apibrėžimas. Imties standartinis nuokrypis yra dažniausiai taikomas sklaidos matas. Jis gaunamas ištraukus kvadratinę šaknį iš dispersijos:

$$S = \sqrt{S^2}.$$

Apibrėžimas. Hipoteze vadinamas kiekvienas nepatvirtintas teiginys apie nežinomą populiacijos požymio skirstinį ar jo parametrus arba kelių populiacijų nepriklausomumą. Paprastai turimas ne tik pats teiginys – nulinė hipotezė H_0, bet ir alternatyvus teiginys – alternatyvioji hipotezė (alternatyva) H_1.

1.2 STATISTINĖS HIPOTEZĖS IR JŲ KLAIDOS TIKIMYBĖS

Teorija pateikta remiantis [6] knyga.

Tarkime, turime binarinį statistinį eksperimentą $\mathcal{E} = (\mathcal{X}, \mathcal{B}, \{\mathrm{P}_0, \mathrm{P}_1\})$ su imtimi $X \in \mathcal{X}$. Iš anksto nėra žinoma, kuris tikrasis imties X skirstinys P_0 ar P_1.

Apibrėžimas. Bet koks teiginys apie tikrąjį imties X skirstinį P^X vadinamas hipoteze.

Kadangi šiuo atveju tėra dvi galimybės, tai tikriname tik dvi hipotezes:

$$H_0 \colon \mathrm{P}^X = \mathrm{P}_0, \quad H_1 \colon \mathrm{P}^X = \mathrm{P}_1.$$

(užrašas $\mathrm{P}^X = \mathrm{P}_i$ reiškia teiginį: tikrasis X skirstinys P^X yra P_i).

Apibrėžimas. Hipotezė, sudaryta iš vienos galimybės, vadinama paprastąja.

Apibrėžimas. Hipotezių H_0 ir H_1 tikrinimo statistiniu kriterijumi vadinama kiekviena statistika $\delta = \delta(X)$, įgyjanti dvi reikšmes 0 ir 1. Kada $\delta(X)$ įgyja reikšmę 0, tada hipotezė H_0 priimama (H_1 atmetama), o kai įgyja reikšmę 1 – H_1 priimama, o H_0 atmetama.

Šitaip apibrėžtas statistinis kriterijus $\delta(X)$ galimas X reikšmes $\mathcal{X}$ padalija į dvi nesikertančias aibes $\mathcal{X}_0 = \mathcal{X} \backslash W$, kur priimama hipotezė H_0, o H_1 atmetama, ir $\mathcal{X}_1 = W$, kur priimama H_1, o H_0 atmetama.

Apibrėžimas. Sritis W, kurioje priimama hipotezė H_1, o H_0 atmetama, vadinama kritine sritimi.

Apibrėžimas. Skaičius α_0 (atitinkamai α_1), vadinamas kriterijaus $\delta = \delta(X)$ 1-osios (atitinkamai 2-osios) rūšies klaidos tikimybe, žymi tikimybę atmesti hipotezę H_0 (atitinkamai H_1), kada ji teisinga.

1.3 POLINOMINIO SKIRSTINIO TAIKYMAS

Teorija pateikta remiantis [7] šaltiniu.

Turime nepriklausomų eksperimentų seką tokią, kad kiekviename bandyme gali įvykti nesutaikomi įvykiai $A_1, \dots, A_k$, kur $k > 2$. Įvykio A_i tikimybė kiekviename bandyme yra lygi $p_i > 0$, $p_1 + p_2 + \cdots + p_k = 1$, ir nepriklauso nuo kitų bandymų rezultatų. Atlikus m bandymų, įvykusių įvykių A_i skaičių žymime X_i, $i = 1, 2, \dots, k$. Tada atsitiktinio vektoriaus $X = (X_1, \dots, X_k)$ tikimybinis skirstinys vadinamas *polinominiu* ir žymimas $X \sim P(m, p_1, \dots, p_k)$. Jį nusako tikimybės:

$$P(X_1 = m_1, X_2 = m_2, \dots, X_k = m_k) = \frac{m!}{m_1! \dots m_k!} p_1^n \dots p_k^m,$$

čia $m \geq m_i \geq 0$ ir $m_1 + \cdots + m_k = m$.

Tarkime, kad turime paprastąją atsitiktinę imtį $X^n = (X_1, X_2, \dots, X_n)$, kur $X_1, X_2, \dots, X_n$ yra nepriklausomi atsitiktiniai elementai, simbolizuojantys apskrities ar kitos didelės teritorijos gyventojus. Bendrą gyventojų skaičių žymime n. Tegu A_j simbolizuoja j-ąją apskritį, kai $j = 1,2, \dots, k$; k – apskričių skaičius. Pažymėkime Y_j gyventojų, patekusių į A_j apskritį, skaičių:

$$Y_j = \sum_{i=1}^{n} \mathbb{I}(X_i \in A_j), j = 1, \dots, k.$$

Vietoje pradinės imties $X^n = (X_1, X_2, \dots, X_n)$, gauname imtį $Y^k = (Y_1, Y_2, \dots, Y_k)$. Atsitiktinis vektorius $Y^k = (Y_1, Y_2, \dots, Y_k)$ turi polinominį skirstinį $Y^k \sim P = (p_1, p_2, \dots, p_k, n)$, t. y.

$$P(Y_1 = n_1, Y_2 = n_2, \dots, Y_k = n_k = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

čia $n_1 + \dots + n_k = n$ ir $p_1 + p_2 + \dots + p_k = 1, p_i, i = 1, \dots, k$ yra tikimybės.

Yra žinoma, kad parametrų p_i maksimalaus tikėtinumo įverčiai yra pavidalo $\frac{Y_i}{n}$.

Tikriname statistinę hipotezę $H_0: p_i = p_{i0}, i = 1, \dots, k,$ turėdami imtį $Y^k = (Y_1, Y_2, \dots, Y_k)$.

Taikome Chi-kvadrato kriterijų. Sudarome Pirsono statistiką:

$$\chi^2 = \sum_{i=1}^{k} \frac{(Y_i - n \cdot p_{i0})^2}{n \cdot p_{i0}} = \sum_{i=1}^{k} \frac{n}{p_{i0}} \cdot (\hat{p}_i - p_{i0})^2,$$

kuri asimptotiškai turi χ^2 skirstinį su $k - 1$ laisvės laipsnių, kai hipotezė H_0 teisinga.

Parinkę reikšmingumo lygmenį α, hipotezei H_0 tikrinti konstruojame tokią kritinę sritį:

jei imties realizacija $y^k = (y_1, y_2, \dots, y_k)$ tokia, kad

$$\chi^2 = \chi^2(y^k) > \chi^2_{1-\alpha}(k - 1),$$

tada hipotezę H_0 atmetame, o jei

$$\chi^2 \leq \chi^2_{1-\alpha}(k - 1),$$

tada hipotezę priimame.

1.4 SERIJŲ KRITERIJUS

Teorija pateikta remiantis [6] šaltiniu.

Atsitiktinė imtis $X^n = (X_1, \dots, X_n)$ pagal prielaidą yra sudaryta iš nepriklausomų atsitiktinių dydžių. Konkrečiame modelyje imties elementai $X_1, \dots, X_n$ gali būti ir net visai ne atsitiktiniai ir priklausomi. Norint įsitikinti, kaip yra iš tikrųjų, tikrinama nulinė hipotezė H_0: imties elementai $X_1, \dots, X_n$ yra atsitiktiniai ir nepriklausomi. Tokiai hipotezei tikrinti taikomas serijų kriterijus su empirine mediana.

Pagal šį kriterijų pirmiausia surašomi imties $X_1, \dots, X_n$ elementai ne mažėjančia tvarka. Pagal (1) formulę randama imties empirinė medianą M_e. Po to M_e reikšmė lyginama su pradinės imties elementais $X_1, \dots, X_n$. Rašomas „+“, jei $X_1 > M_e$, ir „–“, jei $X_1 < M_e$. Jei reikšmės sutampa, nerašoma nieko. Ši procedūra tęsiama, kol patikrinami visi imties elementai. Gautoje „+“-ų ir „–“-ų eilutėje esančių „+“ skaičių žymėsime raide k_1, o „–“ skaičių – k_2. Bendrą visų serijų skaičių žymėsime raide N. $N = N(X^n)$ bus serijų kriterijaus statistika.

Mažas imties serijų skaičius N rodo, kad imties elementai yra giminingi, o didelis imties serijų skaičius N nurodo tarp imties elementų esamą dėsningą kaitą. Abiem šiais atvejais pažeidžiamas imties elementams keliamas atsitiktinumo ir nepriklausomumo reikalavimas. Tuo remiantis hipotezei H_0 tikrinti gali susidaryti viena iš trijų alternatyvų:

$H_1^{(1)}$: imties elementai piklausomi ir neatsitiktiniai,

$H_1^{(2)}$: imties elementai giminingi,

$H_1^{(3)}$: tarp imties elementų yra dėsninga kaita.

Hipotezei H_0 tikrinti, priklausomai nuo alternatyvos, turime kritines sritis:

$W_1 = \left\{X^n : N \leq N_{\frac{\alpha}{2}}(A, n_1, n_2)\ arba\ N \geq N_{\frac{\alpha}{2}}(V, n_1, n_2)\right\}$, kai alternatyva $H_1^{(1)}$,

$W_2 = \{X^n : N \leq N_\alpha(A, n_1, n_2)\}$, kai alternatyva $H_1^{(2)}$,

$W_3 = \{X^n : N \geq N_\alpha(V, n_1, n_2)\}$, kai alternatyva $H_1^{(3)}$,

čia $n_1 = \max(k_1, k_2), n_2 = \min(k_1, k_2)$, $N_\alpha(A, n_1, n_2) = N_\alpha(Apatinė, n_1, n_2)$ ir $N_\alpha(V, n_1, n_2) = N_\alpha(Viršutinė, n_1, n_2)$ – tam tikros kritinės reikšmės, priklausančios nuo kriterijaus reikšmingumo lygmens α ir n_1, n_2 reikšmių, atitinkančių nagrinėjamą imtį.

1.5 KORELIACINĖ ANALIZĖ

1.5.1 EMPIRINIS KORELIACIOS KOEFICIENTAS

Teorija pateikta remiantis [4], [6] šaltiniais.

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.

Apibrėžimas. Dvimatės imties išsisklaidymo diagrama vadinama imties taškų $(X_i, Y_i), i = 1, \dots, n$, atidėtų stačiakampėje koordinačių sistemoje, visuma.

Apibrėžimas. Dvimatės imties empiriniu skirstiniu vadinamas sąlyginis skirstinys diskretaus atsitiktinio vektoriaus (X, Y), įgyjančio reikšmes (X_i, Y_i), $i = 1, \dots, n$ 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 $(X_1, Y_1), \dots, (X_n, Y_n)$ koreliacijos koeficiento

$$\rho = \frac{\boldsymbol{E}(X_i - \boldsymbol{E}X_i)(Y_i - \boldsymbol{E}Y_i)}{\sqrt{DX_i DY_i}}$$

empirinis analogas yra empirinis koreliacijos koeficientas, apskaičiuojamas pagal formulę

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}};$$

čia

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

Koreliacijos koeficiento r realizacija turi tokias savybes:

1. $-1 \leq r \leq 1$. Kuo r reikšmė absoliučiuoju didumu arčiau 1, tuo tiesinė Y priklausomybė nuo X stipresnė.
2. Jeigu $r > 0$, tai didėjant X, didėja ir Y. Jeigu $r < 0$, tai didėjant X, mažėja Y.
3. r neparodo netiesinės priklausomybės.
4. r priklauso nuo duomenų homogeniškumo. Kuo $x_1, x_2, \dots, x_n$ arba $y_1, y, \dots, y_n$ vienodesni, tuo r mažesnis. Jeigu visi x_i vienodi, tai $r = 0$.
5. Kuo didesnė imtis, tuo r yra arčiau nežinomo tikrojo koreliacijos koeficiento ρ.

Nustatytos tam tikros koreliacijos koeficiento interpretacijos:

1 lentelė. *Koreliacijos koeficiento interpretacijos*

ρ reikšmė	**Interpretacija**
nuo -0,2 iki 0,2	Labai silpna teigiama (neigiama) koreliacija
nuo $\pm 0{,}2$ iki $\pm 0{,}5$	Silpna teigiama (neigiama) koreliacija
nuo $\pm 0{,}5$ iki $\pm 0{,}7$	Vidutinio stiprumo teigiama (neigiama) koreliacija
nuo $\pm 0{,}7$ iki $\pm 0{,}9$	Stipri teigiama (neigiama) koreliacija
Nuo $\pm 0{,}9$ iki ± 1	Labai stipri teigiama (neigiama) koreliacija

Koreliacijos koeficientas ρ apibūdina tiesinę priklausomybę tarp dydžių $X_i\ ir\ Y_1$.

Kadangi teorinio (tikrojo) koreliacijos koeficiento ρ tarp $X_i \; ir \; Y_1$ nežinome, tai vietoj jo naudojame jo įvertį – empirinį koreliacijos koeficientą *r*.

1.5.2 HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTĄ

Teorija pateikta remiantis [6] šaltiniu.

Empirinis koreliacijos koeficientas *r* yra atsitiktinis dydis, todėl turėdami jo reikšmę, apskaičiuotą pagal vieną dvimatės imties realizaciją, negalime būti visiškai tikri, kad ji pakankamai artima tikrajai koreliacijos koeficiento ρ reikšmei. Šiuo atveju daug patikimiau yra patikrinti statistinę hipotezę apie galimą ρ reikšmę. Kadangi praktiškai yra labai svarbu žinoti, ar tarp dydžių $X_i \; ir \; Y_1$ yra (tiesinė) priklausomybė ar ne, tai pirmiausia tikrinama hipotezė $H_0 = \rho = 0$.

Tarkime, kad turime dvimatę atsitiktinę imtį $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, kurios elementai (X_i, Y_1) turi dvimatį normalųjį skirstinį

$$N_2\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

su parametrais $a_1\,, a_2\,, \sigma_X^2, \sigma_Y^2, \rho$, čia ρ – koreliacijos koeficientas tarp $X_i \; ir \; Y_1$.

Išskiriami keli atvejai.

1) Tegu imties tūris n yra pakankamai didelis. Tada empirinis koreliacijos koeficientas r turi asimptotiškai normalųjį skirstinį su vidurkiu $\boldsymbol{E}r = \rho$ ir dispersija $\boldsymbol{D}r = \frac{(1-p^2)^2}{n}$.

   Todėl, kai didelis imties tūris *n* ir teisinga hipotezė $H_0\colon \rho = 0$, statistika

$$r\sqrt{n} \sim N(0{,}1).$$

Tuo remiantis pasirenkamas reikšmingumo lygmuo α ir hipotezei $H_0\colon \rho = 0$ tikrinti apibrėžiama kritinė sritis:

$$W = \left\{(X_i, Y_i), i = 1 \dots, n \colon |r|\sqrt{n} > u_{1-\frac{\alpha}{2}}\right\},\ \text{jei } H_o \text{ alternatyva yra } H_1\colon \rho \neq 0.$$

2) Kai imties tūris $n > 10$, o $|r|$ artimas vienetui, tada taikoma R. A. Fišerio transformacija

$$Arthr = \frac{1}{2} ln \frac{1 + r}{1 - r}.$$

Šiuo atveju, jei teisinga hipotezė $H_0\colon \rho = \rho_0$, kur ρ_0 – duotas skaičius, tai

$$Arthr \sim N\left(Arth\rho_0, \frac{1}{n - 3}\right).$$

Todėl

$$Z = \sqrt{n - 3}(Arthr - Arth\rho_0) \sim N(0{,}1),$$

kai H_0 teisinga.

Tuo remiantis, hipotezei $H_0\colon \rho = \rho_0$ tikrinti su reikšmingumo lygmeniu α sudaroma kritinė sritis:

$W = \left\{(X_i, Y_i), i = 1 \dots, n\colon |Z| > u_{1-\frac{\alpha}{2}}\right\}$, jei H_o alternatyva yra $H_1\colon \rho \neq \rho_0$.

1.5.3 TIESINĖ REGRESIJA

Teorija pateikta remiantis [6] šaltiniu.

Sakykime, kad turime du atsitiktinius dydžius X ir Y. Jei apskaičiavę jų koreliacijos koeficientą $\rho = \rho(X, Y)$ gauname, kad jis artimas ± 1, tai reiškia, kad tarp dydžių X ir Y egzistuoja beveik tiesinė priklausomybė, kuri užrašoma per sąlyginį vidurkį

$$\boldsymbol{E}(Y|X = x) = \beta_0 + \beta_1 x. \tag{2}$$

Ši formulė nusako atsitiktinio dydžio Y tiesinės regresijos lygtį X atžvilgiu. Skaičiai β_0 ir β_1 vadinami tiesinės regresijos parametrais.

Paprastai atsitiktinių dydžių X ir Y tikimybiniai skirstiniai tiksliai nežinomi, todėl tiesinės regresijos lygtį $y = \beta_0 + \beta_1 x$ įvertiname pagal (X, Y) stebėjimo rezultatus $(x_1, y_1), \dots, (x_n, y_n)$, kurie teoriškai yra realizacija atsitiktinės dvimatės imties (X_1, Y_1),..., (X_n, Y_n), kur $\boldsymbol{E}(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i, i = 1, \dots, n$. Toliau daroma prielaida, kad

$$(Y_i|X_i = x_i) \sim N(\beta_0 + \beta_1 x_i, \sigma^2),$$

kuri reiškia, kad

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

čia ε_i~N(0, σ^2), $i = 1, \dots, n$ – nepriklausomi atsitiktiniai dydžiai.

Kadangi kiekvienai fiksuotai X_i reikšmei x_i turėjome normalųjį atsitiktinio dydžio išsibarstymą, tai pradinės dvimatės imties nagrinėjimas ekvivalentus nagrinėjimui dvimatės imties (X_1, Y_1),..., (X_n, Y_n), kurioje $X_1, \dots, X_n$ yra neatsitiktiniai dydžiai, o atsitiktiniai dydžiai Y_i užrašomi (2) pavidalu. Kitaip sakant, turime iš anksto žinomų skaičių rinkinį $(X_1, \dots, X_n)$ ir atsitiktinę nepriklausomų dydžių imtį $(Y_1, \dots, Y_n)$, kur

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2),\ i = 1, \dots, n. \tag{3}$$

Ši formulė nusako tiesinės regresijos modelį.

Tiesinės regresinės analizės uždavinys yra toks: naudojantis imtimi $(X_1, Y_1), \dots, (X_n, Y_n)$

1) gauti „geriausius“ regresijos modelio nežinomų parametrų $\beta_0, \beta_1\ ir\ \sigma^2$ taškinius ir intervalinius įvertinius,
2) patikrinti statistines hipotezes apie parametrus $\beta_0\ ir\ \beta_1$,
3) nustatyti, ar modelis gerai suderintas su imties duomenimis (modelio adekvatumo patikrinimas).

Tiesinės regresijos parametrų $\beta_0 \ ir \ \beta_1$ įverčius $\widehat{\beta_0}$ ir $\widehat{\beta_1}$ randame mažiausių kvadratų metodu, minimizuodami dydį

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\beta_0} - \widehat{\beta_1} X_i)^2 \to \min_{\widehat{\beta_0}, \widehat{\beta_1}}.$$

Panaudoję būtiną ekstremumo sąlygą $\frac{\partial S^2}{\partial \widehat{\beta_1}} = 0$ ir $\frac{\partial S^2}{\partial \widehat{\beta_0}} = 0$, gauname

$$\begin{cases} \sum_{i=1}^{n}(Y_i - \widehat{\beta_1} X_i - \widehat{\beta_0})X_i = 0, \\ \sum_{i=1}^{n}(Y_i - \widehat{\beta_1} X_i - \widehat{\beta_0}) = 0. \end{cases}$$

Išsprendę šią sistemą, turime įverčius

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}, \qquad \widehat{\beta_0} = \frac{Q_{X,Y}}{Q_X},$$

kur

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

$$Q_X = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$$Q_{X,Y} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

1.5.4 PROGNOZĖS INTERVALAS

Pagrindinė šio skyrelio teorija paimta iš [4] šaltinio.

Esant fiksuotam x, galima prognozuoti galimų Y reikšmių intervalą. Faktiškai reikia mokėti sudaryti skirtumo $Y - \hat{y}(x)$ prognozės intervalą. Fiksuokime nepriklausomojo kintamojo reikšmę x. Tegul

$$SY = MSE\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}\right);$$

čia $MSE = \frac{SSE}{(n-2)}$ – nežinomos dispersijos σ^2 įvertis, o $SSE = \sum_{i=1}^{n} \hat{e}_i^2$ – liekamųjų paklaidų kvadratų suma, kur $\hat{e}_i$ parodo, kiek stebėtoji y_i reikšmė skiriasi nuo reikšmės, kurią gautume prognozuodami pagal regresijos tiesę; s_x^2 yra empirinė $x_1, \dots, x_n$ dispersija, pagal formulę $(n-1)s_x^2 = \sum_1^n x_i^2 - \frac{(\sum_1^n x_i)^2}{n}$.

Yra žinoma, kad $(Y - \hat{y}(x))/\sqrt{SY}$ turi Stjudento skirstinį su $(n-2)$ laisvės laipsnių. Todėl Q pasikliovimo lygmens $y(x)$ pasikliautinasis intervalas yra

$$\hat{y}(x) \pm t_{\frac{1-Q}{2}}(n-2)\sqrt{SY};$$

čia $t_{\frac{1-Q}{2}}(n-2)$ yra Stjudento kriterijaus su $(n-2)$ laisvės laipsnių $(1-Q)/2$ lygmens kritinė reikšmė.

Reikalaujama, kad x būtų iš duomenų intervalo, t.y. $\min(x_1, \dots, x_n) \leq x \leq \max(x_1, \dots, x_n)$. Dažniausiai skaičiuojamas 95% prognozės intervalas.

Kintamojo Y, esant fiksuotam x, 95% prognozės intervalas yra

$$\left[\hat{y}(x) - \sqrt{SY} t_{0,975}(n-2); \hat{y}(x) + \sqrt{SY} t_{0,975}(n-2)\right];$$

Akivaizdu, kad prognozės intervalo plotis priklauso ne tik nuo imties didumo, bet ir nuo x reikšmės. Kuo x arčiau $\bar{x}$, tuo prognozės intervalas siauresnis.

1.6 TIESINĖ REGRESIJA SU PROGRAMA „SPSS”

Pagrindinė šio poskyrio teorija paimta iš [3] ir [11] šaltinių.

Tikrinant, ar daugiamatė tiesinės regresijos lygtis gali būti naudojama prognozavimui, ar ji yra tikslinga – svarbu atkreipti dėmesį į kai kuriuos rodiklius, kuriuos gauname skaičiuodami tiesinę regresiją programoje „SPSS“.

Pirmiausias rodiklis, į kurį reikia atkreipti dėmesį, yra koreliacijos koeficientai. Visi nepriklausomi kintamieji turi koreliuoti su priklausomu kintamuoju, o tarpusavyje stipriai koreliuoti neturėtų.

Antra, reikia atkreipti dėmesį į „Model Summary“ lentelės rezultatus. Koreguotas determinacijos koeficientas (Adjusted R Square) parodo, kiek % modelis paaiškina pasirinkto kintamojo reikšmių sklaidos apie vidurkį tiesinę regresiją nepriklausomų kintamųjų atžvilgiu. Kuo šis rodiklis didesnis, tuo modelis tinkamesnis. Taip pat ir determinacijos koeficientas (R Square) rodo tą patį. Jeigu imtis gerokai didesnė už nepriklausomus kintamuosius, dažniausiai žiūrima į determinacijos koeficiento reikšmę. Mažiausia reikšmė tinkamam modeliui yra $R^2 \geq 0{,}2$. Taip pat lentelėje yra standartinės regresijos paklaidos (Std. Error of the Estimate) dydis, kuo jis mažesnis, tuo geriau regresijos modelis apibūdina duomenis.

2 lentelė. *„Model Summary“ lentelė*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate

Sekanti lentelė yra “Anova”. Ieškant daugiamatės regresijos tikrinama hipotezė apie regresijos netiesiškumą. Jei lentelėje reikšmingumo lygmuo (Sig.) yra <0,05, hipotezė atmetama ir tai reiškia, kad regresija yra tiesinė – bent vienas koeficientas nelygus nuliui ir regresijos modelis bent jau iš dalies tinka prognozėms.

__3 lentelė.__ „ANOVA" lentelė

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression					
	Residual					
	Total					

Lentelėje „Coefficients“ skiltyje B „Unstandardized Coefficients“ gauname regresijos lygties koeficientus. Jeigu koeficientas reikšmingas, lentelės skiltyje „Sig.“ reikšmė turi būti <0,05, tokiu atveju kintamasis statistiškai reikšmingas ir yra tinkamas prognozavimui. Kitu atveju jis nėra tinkamas. Standartizuotų koeficientų (Standardized Coefficients) reikšmės dydis parodo, kuris nepriklausomas kintamasis yra svarbiausias.

__4 lentelė.__ „Coefficients“ lentelė

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)					

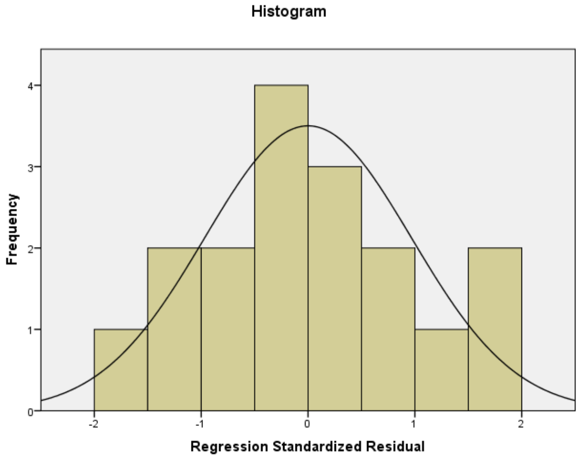
Lentelėje „Residual Statistics“ reiktų atkreipti dėmesį į Kuko matą (Cook‘s Distance), kuris parodo, ar duomenyse yra išskirčių. Priešpaskutinėje eilutėje, antrame stulpelyje rašoma Kuko mato maksimumo reikšmė. Taigi, jei ši reikšmė neviršija vieneto – išskirčių nėra.

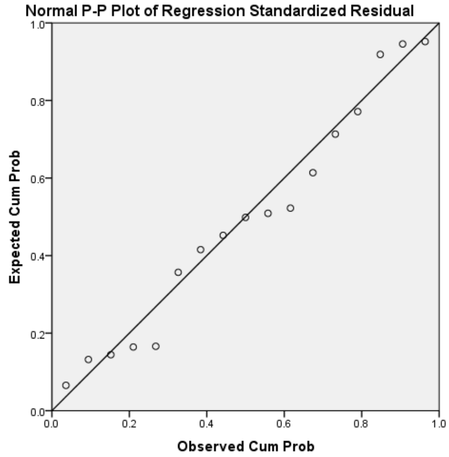
__5 lentelė.__ „Residual Statistics“ lentelė

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value					
…	…	…	…	…	…
Cook's Distance					
Centered Leverage Value					

Informacija apie modelio standartizuotuojų liekamųjų paklaidų normalumą yra pateikiama dviejuose grafikuose „Histogram“ ir „P-P plot“. Histogramos grafike yra nubraižyta ir normalioji kreivė, t.y. tai, į ką turėtų būti panaši histograma, jeigu liekamosios paklaidos būtų normalios. Antrajame grafike atidėti standartizuotojų liekamųjų paklaidų ir normaliojo atsitiktinio kintamojo santykiniai procentinai dažniai. Kuo taškai arčiau nubrėžtos tiesės (idealiu atveju visi taškai yra ant tiesės), tuo duomenys normalesni.





Standartizuotosios liekamosios paklaidos naudojamos patikrinimui, ar Y normaliai pasiskirstęs. Normalumui ištirti naudojamas Šapiro-Vilko (Shapiro-Wilk) testas. Jei lentelėje „Tests of Normality“ „Sig.“ reikšmė $> 0{,}05$ teigiame, kad standartizuotosios liekamosios paklaidos yra normalios. Taip pat, jei pateikiamo Kolmogorovo-Smirnovo testo „Sig.“ reikšmė $> 0{,}05$, teigiame, kad standartizuotosios liekamosios paklaidos yra normalios.

Tokiu pat būdu galima patikrinti, ar duomenys turi normalųjį skirstinį.

6 lentelė. *„Test of Normality“ lentelė*

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual						

1.7 DUOMENŲ NORMALUMAS SU PROGRAMA „SPSS”

Pagrindinė šio poskyrio teorija paimta iš [3] šaltinio.

Patikrinti, ar duomenys yra pasiskirstę normaliai, naudojamas Šapiro-Vilko testas. Jei lentelėje „Tests of Normality“ „Sig.“ Reikšmė $> 0{,}05$, teigiame, kad duomenys pasiskirstę normaliai. Taip pat pateikiamas Kolmogorovo-Smirnovo testas, lygiai taip pat jei „Sig“ reikšmė $> 0{,}05$, teigiame, kad duomenys pasiskirstę normaliai.

7 lentelė. *„Tests of Normality“ lentelė*

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual						

Kolmogorovo-Smirnovo testas gerai nustato nukrypimus nuo normaliojo pasiskirtymo dėsnio, esant nedideliam duomenų skaičiui. Šapiro-Vilko testas yra efektyvesnis (su didesne tikimybe atmeta neteisingą hipotezę) tikrinant normalumą, kai duomenų skaičius nedidelis ($n \leq 50$). Mažoms imtims (tarkime, kai $n < 20$) kriterijų taikymui gali nepakakti statistinės galios atmesti nulinę hipotezę, todėl tokioje situacijoje dažnai gausime klaidinantį rezultatą, kad nuokrypiai nuo normalumo yra nereikšmingi, tačiau tiesinė regresija yra pakankamai atspari nedideliems prielaidų pažeidimams.

1.8 LAIKO EILUTĖS

1.8.1 AUTOREGRESINIAI PROCESAI

Teorija pateikta remiantis [9] ir [10] šaltiniais.

Apibrėžimas. Atsitiktine seka $\{X_t\}$ yra vadinama pirmosios eilės autoregresijos procesu, arba AR(1), jeigu kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + \varepsilon_t,$$

čia dydis a yra konstanta, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Apibrėžimas. Atsitiktine seka $\{X_t\}$ yra vadinama antrosios eilės autoregresijos procesu, arba AR(2), jeigu kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + \varepsilon_t,$$

čia $a_{1,} a_{2,}$ yra konstantos, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Akivaizdu, kad pirmosios ir antrosios eilės autoregresijos procesų natūralus apibendrinimas yra baigtinės eilės procesas.

Apibrėžimas. Atsitiktine $\{X_t\}$ seka yra vadinama p-tos eilės autoregresijos procesu arba AR(p), jei kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + \ldots + a_p \cdot X_{t-p} + \varepsilon_t, \tag{4}$$

čia $a_1, a_2, \ldots, a_p$ yra pastovieji dydžiai, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Padauginę abi lygties (4) puses iš X_{t-r}, turime:

$$X_t \cdot X_{t-r} = a_1 \cdot (X_{t-1} \cdot X_{t-r}) + a_2 \cdot (X_{t-2} \cdot X_{t-r}) + \ldots + a_p \cdot (X_{t-p} \cdot X_{t-r}) + (X_{t-r} \cdot \varepsilon_t),$$

čia $r = 0,1,2, \ldots$.

Iš šios lygybės kartu su (4) gauname:

$$\left\{ \begin{array}{c} \hat{\gamma}(0) = \alpha_1 \cdot \hat{\gamma}(1) + \alpha_2 \cdot \hat{\gamma}(2) + \cdots + \alpha_p \cdot \hat{\gamma}(p) + \sigma^2 \\ \hat{\gamma}(1) = \alpha_1 \cdot \hat{\gamma}(0) + \alpha_2 \cdot \hat{\gamma}(1) + \cdots + \alpha_p \cdot \hat{\gamma}(p-1) \\ \hat{\gamma}(2) = \alpha_1 \cdot \hat{\gamma}(1) + \alpha_2 \cdot \hat{\gamma}(0) + \cdots + \alpha_p \cdot \hat{\gamma}(p-2) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . \\ \hat{\gamma}(p) = \alpha_1 \cdot \hat{\gamma}(p-1) + \alpha_2 \cdot \hat{\gamma}(p-2) + \cdots + \alpha_p \cdot \hat{\gamma}(0) \end{array} \right\} \tag{5}$$

Ši lygtis vadinama Yule-Walker lygtimi ir naudojama AR koeficientų ir koreliacinės funkcijos ryšiui išreikšti.

Apibrėžimas. Atsitiktinio proceso $\{X_t, t \in T\}$, su visais $t \in T$ tenkinančio sąlygą $\mathbf{D}X_t < \infty$, kovariacinė funkcija apibrėžiama lygybe

$$r(s,t) = Cov(X_8, X_t) = \boldsymbol{E}(X_8 - EX_8)(X_t - \boldsymbol{E}X_t), s, t \in T.$$

Apibrėžimas. Seka $\{X_t, t \in \mathbb{Z}\}$ vadinama stacionaria, jeigu:

1) $\forall t \in \mathbb{Z}\ \boldsymbol{E}|X_t|^2 < \infty$,
2) $\forall t \in \mathbb{Z}\ \ \boldsymbol{E}X_t = \boldsymbol{E}X_0$,
3) $r(s,t) = r(s+h, t+h)$ su visais s, t, h iš $\mathbb{Z}$.

Dažnai literatūroje taip apibrėžtas stacionarumas vadinamas stacionarumu plačiąja (arba silpnąja) prasme.

Kadangi stacionariai sekai teisinga $r(s,t) = r(s-t, 0)$ su visais $s, t \in \mathbb{Z}$, tai patogiau kovariacinę funkciją traktuoti kaip vieno argumento funkciją ir rašyti tiesiog $r(s) \equiv r(s, 0)$ visiems $s \in \mathbb{Z}$.

Apibrėžimas. Seka $\{Z_t\}$ vadinama balto triukšmo seka su vidurkiu 0 ir dispersija σ^2, jeigu $\mathbf{E}Z_t = 0$ ir

$$r(h) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

Žymėsime $Z_t \sim BT(0, \sigma^2)$.

AR(1) procesas mažiausių kvadratų metodu. Kovariacinis funkcijos įvertis pasižymi tuo, kad su visais $n \geq 1$ matrica

$$\hat{R}_n = \begin{pmatrix} \hat{r}(0) & \hat{r}(1) & \hat{r}(2) & \cdots & \hat{r}(n-1) \\ \hat{r}(1) & \hat{r}(0) & \hat{r}(1) & \cdots & \hat{r}(n-2) \\ \hat{r}(n-1) & \hat{r}(n-2) & \hat{r}(n-3) & \cdots & \hat{r}(0) \end{pmatrix}$$

yra neneigiamai apibrėžta.

$$\rho(h) = \frac{cov(X_t X_{t+h})}{\sqrt{DX_t DX_{t+h}}}$$

$$\hat{\rho}(h) = \frac{\hat{r}(h)}{\hat{r}(0)}$$

Toliau $\mathbf{E}X_t = 0$

$$\hat{r}(h) = \frac{1}{n} \sum_{k=1}^{n-|h|} X_k X_{k+|h|} \qquad\qquad 0 \leq |h| < n$$

Žinoma, kad $\mathbf{E}\hat{r}(h) = (1 - |h|/n) r(h) \rightarrow r(h), \quad n \rightarrow \infty$

$\hat{r}(h)$ – nepaslinktas $r(h)$ įvertis asimptotiškai $AR(p)$ modelio parametro įverčiui.

$AR(p)$ procesas:

$$X_t = X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t;$$

čia $\varepsilon_t \sim BT(0, \sigma^2)$ – balto triukšmo su $\mathbf{D}\varepsilon_t = \sigma^2$.

Lygtį galima užrašyti taip:

$$X_t = X_{t-1} a + \varepsilon_t;$$

čia $a = (a_1, a_2, \dots, a_p), X_{t-1} = (X_{t-1}, X_{t-2}, \dots, X_{t-p})$.

Gerai žinoma, kad a įvertis mažiausio kvadrato metodu yra:

$$\hat{a}' = \left(\sum_{t=1}^{n} \mathbb{X}_{t-1} \mathbb{X}'_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \mathbb{X}_{t-1} X_t \right)$$

PASTABA. Kai ε_t yra n.v.p $N(0, \sigma^2)$, tai $\hat{a}$ yra ir maksimalaus tikėtinumo įvertis.

Kai $p = 1$, turime $AR(1)$, $X_t = X_{t-1} a + \varepsilon_t$.

Gauname:

$$\hat{a} = \frac{\sum_{t=1}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_{t-1}^2}.$$

Teorema. Jei ε_t – n.v.p su $\mathbf{E}\varepsilon_t = 0, \mathbf{D}\varepsilon_t = \sigma^2$, o $\mathbf{E}\varepsilon_t^4 < \infty (AR(p))$

tai $\sqrt{n}(\hat{a} - a) \xrightarrow{d} N(0, \sigma^2 \Gamma_p{}^{-1})$;

$$\text{čia } \Gamma_p = \mathrm{E}\mathbb{X}\mathbb{X}' = \begin{pmatrix} r(0) & r(1) & r(2) & \cdots & r(p-1) \\ r(1) & r(0) & r(1) & \cdots & r(p-2) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r(p-1) & r(p-2) & r(p-3) & \cdots & r(0) \end{pmatrix}.$$

Kai $p = 1, AR(1)$

$$\sqrt{n}(\hat{a} - a) \xrightarrow{d} N\left(0, \frac{\sigma^2}{r(0)}\right).$$

Kadangi $r(0) = \frac{a^2}{1-a^2}$ buvo $r(0) = \boldsymbol{D}X(0) = \frac{\gamma_\varepsilon^2}{1-a^2}$,

tai:

$$\sqrt{n}(\hat{a} - a) \xrightarrow{d} N\left(0 \frac{\sigma^2}{\frac{\gamma_\varepsilon^2}{1-a^2}}\right) = N(0, 1 - a^2)$$

arba $\hat{a} \sim AN\left(a, \frac{1-a^2}{n}\right)$.

AR(2) procesas mažiausių kvadratų metodu.

Buvo

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t;$$

čia $\varepsilon_t \sim BT(0, \sigma^2)$ – balto triukšmo su $\mathbf{D}\varepsilon_t = \sigma^2$.

Lygtį galima užrašyti tokiu pavidalu:

$$X_t = X_{t-1}a + \varepsilon_t;$$

kur $a = \left(a_1, a_2, \dots, a_p\right), X_{t-1} = (\, X_{t-1}, X_{t-2}, \dots, X_{t-p})$.

Gerai žinoma, kad a įvertis mažiausių kvadratų metodu yra

$$\hat{a} = \left(\sum_{t=1}^{n} X_{t-1}X_{t-1}\right)^{-1}\left(\sum_{t=1}^{n} X_{t-1}X_t\right)$$

$$\widehat{a_1}\sum_{t=1}^{n} X_{t-1}^2 + \widehat{a_2}\sum_{t=1}^{n} X_{t-1}X_{t-2} - \widehat{a_2}\sum_{t=1}^{n} X_{t-2}^2 = 0$$

$$\widehat{a_1}\sum_{t=1}^{n} X_{t-1}X_{t-2} + \widehat{a_2}\sum_{t=1}^{n} X_{t-2}^2 = \sum_{t=1}^{n} X_tX_{t-2}$$

$$\Delta = \begin{vmatrix} \sum_{t=1}^{n} X_{t-1}^2 & \sum_{t=1}^{n} X_{t-1}X_{t-2} \\ \sum_{t=1}^{n} X_{t-1}X_{t-2} & \sum_{t=1}^{n} X_{t-2}^2 \end{vmatrix} = \left(\sum_{t=1}^{n} X_{t-1}^2\right)\left(\sum_{t=1}^{n} X_{t-2}^2\right) - \left(\sum_{t=1}^{n} X_{t-1}X_{t-2}\right)^2$$

$$\widehat{a_1} = \frac{\Delta\widehat{a_1}}{\Delta} = \frac{\begin{vmatrix} \sum_{t=1}^{n} X_tX_{t-1} & \sum_{t=1}^{n} X_{t-1}X_{t-2} \\ \sum_{t=1}^{n} X_tX_{t-2} & \sum_{t=1}^{n} X_{t-2} \end{vmatrix}}{\Delta} = \frac{\begin{vmatrix} \sum_{t=1}^{n} X_tX_{t-1} & \sum_{t=1}^{n} X_{t-1}X_{t-2} \\ \sum_{t=1}^{n} X_tX_{t-2} & \sum_{t=1}^{n} X_{t-2} \end{vmatrix}}{(\sum_{t=1}^{n} X_{t-1}^2)(\sum_{t=1}^{n} X_{t-2}^2) - (\sum_{t=1}^{n} X_{t-1}X_{t-2})^2}$$

$$= \frac{(x_2x_1 + x_3x_2 + \dots + x_{19}x_{18})(x_1^2 + x_2^2 + \dots + x_{17}^2) - (x_3x_1 + x_4x_2 + \dots + x_{19}x_{17})(x_2x_1 + x_3x_2 + \dots + x_{18}x_{17})}{(x_1^2 + x_2^2 + \dots + x_{18}^2)(x_1^2 + x_2^2 + \dots + x_{17}^2) - (x_2x_1 + x_3x_2 + \dots + x_{18}x_{17})^2}$$

$$\widehat{a_2} = \frac{\Delta\widehat{a_2}}{\Delta} = \frac{\begin{vmatrix} \sum_{t=1}^{n} X_{t-1}^2 & \sum_{t=1}^{n} X_tX_{t-1} \\ \sum_{t=1}^{n} X_{t-1}X_{t-2} & \sum_{t=1}^{n} X_tX_{t-2} \end{vmatrix}}{\Delta} = \frac{\begin{vmatrix} \sum_{t=1}^{n} X_{t-1}^2 & \sum_{t=1}^{n} X_tX_{t-1} \\ \sum_{t=1}^{n} X_{t-1}X_{t-2} & \sum_{t=1}^{n} X_tX_{t-2} \end{vmatrix}}{(\sum_{t=1}^{n} X_{t-1}^2)(\sum_{t=1}^{n} X_{t-2}^2) - (\sum_{t=1}^{n} X_{t-1}X_{t-2})^2}$$

$$= \frac{(x_1^2 + x_2^2 + \dots + x_{18}^2)(x_3x_1 + x_4x_2 + \dots + x_{19}x_{17}) - (x_2x_1 + x_3x_2 + \dots + x_{19}x_{18})(x_2x_1 + x_3x_2 + \dots + x_{18}x_{17})}{(x_1^2 + x_2^2 + \dots + x_{18}^2)(x_1^2 + x_2^2 + \dots + x_{17}^2) - (x_2x_1 + x_3x_2 + \dots + x_{18}x_{17})^2}$$

$$X_t = \widehat{a_1} \cdot X_{t-1} + \widehat{a_2} \cdot X_{t-2} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

1.8.2 AUTOREGRESINIŲ PROCESŲ PARAMETRŲ VERTINIMAS

Teorija pateikta remiantis [10] šaltiniu.

Tarkime, kad turime autoregresijos proceso $\{X_t\}$, išreiškiamo (4) lygtimi, stebinius $X_1, X_2, \dots, X_n$ ir žinoma lygties eilė *p*. Reikia gauti nežinomų parametrų $a_1, a_2, \dots, a_p, \sigma^2$.

Žinome, kad

$$\hat{\gamma}(p) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(X_{j+k} - \bar{X}\right) \cdot (X_j - \bar{X}),$$

čia

$$\bar{X} = \frac{1}{n} \cdot \sum_{j=1}^{n} X_j.$$

Iš lygčių sistemos apskaičiuojame įverčius $\hat{\gamma}(0), \hat{\gamma}(1), \dots, \hat{\gamma}(p)$ ir įstatome į (5) sistemą. Gauname lygtį su nežinomaisiais parametrais $a_1, a_2, \dots, a_p$, σ^2. Tuomet surandame $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ įverčius.

1.8.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS

Teorija pateikta remiantis [2] šaltiniu.

Esminį vaidmenį sprendžiant išskyrimo ir įvertinius trendinės $f_{TR}(t)$, sezoninės $\mu(t)$ ir ciklinės $\psi(t)$ sudaromosios išskaidyme yra pradinis analizės etapas kuriame išaiškinamas pats faktas buvimo ar nebuvimo neatsitiktinės (ir priklausančios nuo laiko t) komponentės išskaidyme, iš esmės kalba eina apie statistinės hipotezės $H_0: \boldsymbol{E}X(t) = a = const.$ patikrinimą. Konstruojamas įvertis (aproksimacija) dėl nežinomos integralinės neatsitiktinės dalies $f(t) = \chi(A) \cdot f_{TR}(t) + \chi(B) \cdot \mu(t) + \chi(C) \cdot \psi(t)$, t.y. sprendžiamas sulyginimas (eliminavimas atsitiktinės liekanos $a(t)$) analizuojamos laiko eilutės $X(t)$. Surašome analizuojamos laiko eilutės $X(t_1), \dots, X(t_n)$ narius didėjančia tvarka (variacine eile). Jai apibrėžkime medianą M_e pagal (1) formulę.

Po to sudarome serijas iš pliusų ir minusų kaip tikrindami nepriklausomumą. Atskiru atveju grįžtant prie pradinės eilės $X(t_1), \dots, X(t_n)$ vietoj kiekvieno nario $X(t)$, $t = 1,2, \dots, n$, rašome +, jei $X(t) > M_e$ ir minusą (-), jei $X(t) < M_e$ (laiko eilutės nariai kurie yra lygūs M_e neimami).

Gauta pliusų ir minusų eilutė charakterizuojama bendru serijų skaičiumi $\gamma(n)$ ir pačios ilgiausios serijos ilgiu $\tau(n)$. Čia serija laikoma kartu einanti pliusų arba minusų grupė. Akivaizdu, kad jei nagrinėjame seką $X(t_1), \dots, X(t_n)$, sudarytą iš statistiškai nepriklausomų stebėjimų, atsitiktinai varijuojančių apie tam tikro pastovaus lygio a (t.y. jei teisinga hipotezė $H_0: \boldsymbol{E}X(t) = a = const.$), tai kaita pliusų ir minusų nurodytoje sukonstruotoje jų eilutėje turi būti daugiau ar mažiau atsitiktinė, t.y. ta seka privalo neturėti labai ilgos pliusų arba minusų serijos, einančios iš eilės, ir atitinkamai bendras serijų skaičius $\gamma(n)$ negali būti per daug mažas. Šiame kriterijuje tikslinga nagrinėti statistinę porą $(\gamma(n); \tau(n))$. Todėl mes turime užfiksuoti ir sužinoti atsitiktinių dydžių $(\gamma(n); \tau(n))$ dvimatį skirstinį, galiojantį su sąlyga, kad tikrinama hipotezė $H_0: \boldsymbol{E}X(t) = a = const.$ teisinga. Čia nurodysime apytikslį kriterijų ir jo sukonkstravimui pasinaudosime:

- penkių procentų taško įverčiu $\tau_{0,05}(n)$ pasiskirsčiusio $\tau_{0,05}(n) \approx [1{,}43 \cdot \ln(n+1)]$ (čia $[x]$ žymi “sveiką x dalį”)
- įverčiais iš viršaus ir apačios dėl tikimybių $P\{\gamma(n) < \gamma_{0,95}(n); \tau(n) > \tau_{0,05}(n)\}$.

Tai leidžia suformuluoti taisyklę (tikrinant hipotezę H_0: $\boldsymbol{EX}(t) = a = const.$):
jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}\left(n + 2 - 1{,}96\sqrt{n-1}\right)\right]$$

$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

1.9 LIGŲ KLASIFIKACIJA PAGAL TLK-10-AM LIGŲ KLASIFIKATORIŲ

Pagrindinė šio poskyrio teorija paimta iš [1] šaltinio.

TLK-10-AM – Pasaulio sveikatos organizacijos *Tarptautinė statistinė ligų ir sveikatos sutrikimų klasifikacija, dešimtasis leidimas*, Australijos modifikacija. Jos devintasis leidimas pakeičia aštuntąjį, išleistą 2013 metais. TLK-10-AM devintąjį leidimą parengė Australijos klasifikacijos kūrimo konsorciumas (ACCD). Leidinio rengimo metu ACCD konsultavosi su ACCD TLK Technine grupe (ITG) ir Klasifikavimo klinikine patariamąja grupe (CCAG), kartu su įvairių specialybių klinicistais. Šias grupes sudaro klinikinio kodavimo specialistai iš kiekvienos valstijos ar teritorijos, privačių ligoninių, klinikų specialistai, Australijos sveikatos ir gerovės instituto (AIHW), Sandraugos sveikatos departamento, Australijos sveikatos apsaugos ir kokybės komisijos, Nacionalinio sveikatos informacijos standartų ir statistikos komiteto bei Nepriklausomos ligoninėms kainas nustatančios institucijos (IHPA).

Higienos instituto Sveikatos centro periodiniame leidinyje „Lietuvos gyventojų sveikata ir sveikatos priežiūros įstaigų veikla“ psichikos ir elgesio sutrikimai yra suskirstyti iš viso į 11 grupių ir pogrupių:

8 lentelė. *Psichikos ir elgesio sutrikimų sąrašas*

PSICHIKOS IR ELGESIO SUTRIKIMAI, iš jų:	**5**
demencija ir Alzheimerio liga	5.1
demencija	5.1.1
psichikos ir elgesio sutrikimai dėl alkoholio vartojimo	5.2
psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo	5.3
šizofrenija, šizotipinis ir kliedesiniai sutrikimai	5.4
šizofrenija	5.4.1
nuotaikos (afektiniai) sutrikimai	5.5
depresijos	5.5.1
protinis atsilikimas ir psichologinės raidos sutrikimai, iš jų:	5.6
psichologinės raidos sutrikimai	5.6.1
elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje	5.7

1.10 ŽODŽIŲ TRUMPINIMAI IR ŽYMĖJIMAI

Magistro darbe naudosime pagrindines išskirtas 7 ligų-sutrikimų grupes ir tokius žymėjimus:

1 – Demencija ir Alzheimerio liga (F00-F03),

2 – Psichikos ir elgesio sutrikimai vartojant alkoholį (F10),

3 – Psichikos ir elgesio sutrikimai vartojant kitas psichoaktyviąsias medžiagas (F11-F19),

4 – Šizofrenija, šizotipinis ir kliedesiniai sutrikimai (F20-F29),

5 – Nuotaikos (afektiniai sutrikimai) (F30-F39),

6 – Protinis atsilikimas ir psichologinės raidos sutrikimai (F70-F89),

7 – Elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje (F90-F98),

SPES – sergantieji psichikos ir elgesio sutrikimais.

2. TYRIMAS

2.1 DUOMENYS

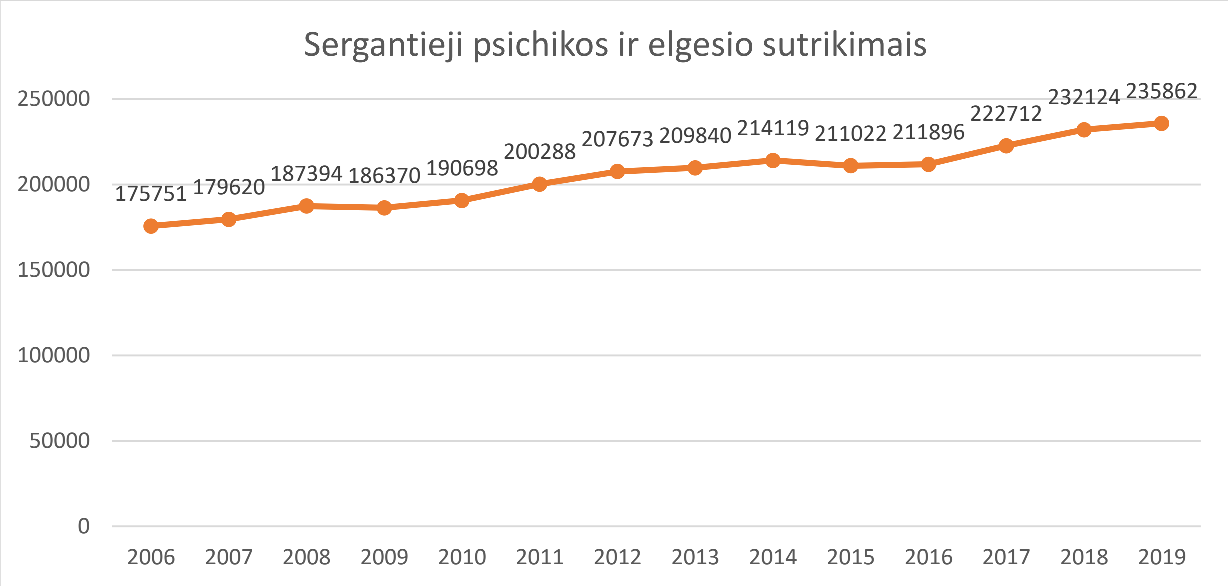
Turimi duomenys apie sergančiųjų psichikos ir elgesio sutrikimais žmonių skaičius Lietuvoje 2006–2019 m. laikotarpiu (1 Priedas) gauti iš Higienos instituto Sveikatos informacijos centro leidinio [5]:

9 lentelė. *SPES skaičiai*

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
1	11340	13915	17068	19609	21233	25156	28655
2	10686	11847	10865	9129	9542	12306	14001
3	714	686	847	979	958	1278	1408
4	17042	17167	17479	17900	18096	18648	18877
5	37818	40308	42593	42963	44534	46821	48594
6	18532	19367	21028	22573	24273	26057	27114
7	9862	9887	10031	10186	10429	10442	11014
SPES	**175751**	**179620**	**187394**	**186370**	**190698**	**200288**	**207673**

	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1	29184	30806	31559	33170	33056	32362	32617
2	18259	19373	19908	20445	19127	18308	18632
3	1722	1844	2153	2132	2269	3001	3278
4	19115	18872	18591	18969	18765	18588	18603
5	47862	47584	45793	46099	46873	46329	46345
6	27281	28655	28827	28773	29677	29501	30445
7	10781	11407	11383	10816	11515	11145	11247
SPES	**209840**	**214119**	**211022**	**211896**	**222712**	**232124**	**235862**

Patogu bendrus SPES skaičius pavaizduoti grafiškai – linijine diagrama:



1 pav. *Sergančiųjų psichikos ir elgesio sutrikimais Lietuvoje dinamika*

Apskaičiuosime šių duomenų empirines charakteristikas (vidurkį, dispersiją, standartinį nuokrypį, medianą).

Pirmiausia sudarome variacinę duomenų eilutę:

175751, 179620, 186370, 187394, 190698, 200288, 207673, 209840, 211022, 211896, 214119, 222712, 232124, 235862,

Apskaičiuojame empirines charakteristikas:

Vidurkis:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{14}\sum_{i=1}^{14} X_i = \frac{1}{14}(175751 + 179620 + \cdots + 235862) = 204669{,}214.$$

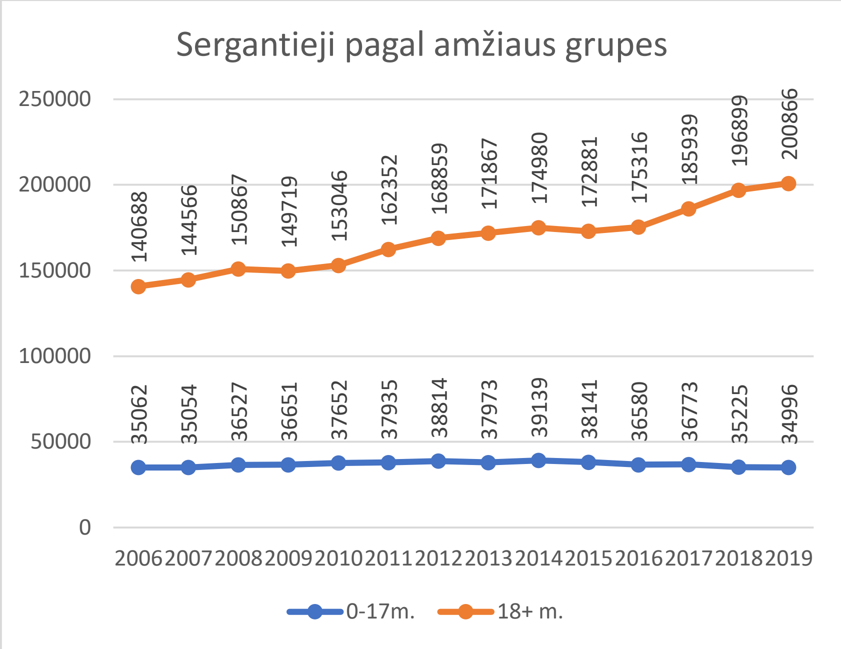
Dispersija:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{14-1}\sum_{i=1}^{14}(X_i - \bar{X})^2 = \frac{1}{14-1}((175751 - 204669{,}214)^2 + (179620 - 204669{,}214)^2 + \cdots + (235862 - 204669{,}214)^2) = 352411551{,}6$$

Standartinis nuokrypis: $S = \sqrt{S^2} = \sqrt{352411551{,}6} = 18772{,}62772$

Mediana: $M_e = \frac{1}{2}\left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right) = \frac{1}{2}\left(X_{(\frac{14}{2})} + X_{(\frac{14}{2}+1)}\right) = \frac{1}{2}\left(X_{(7)} + X_{(8)}\right) = \frac{417513}{2} = 208756{,}5.$

Kadangi turime duomenis sergančiųjų psichikos ir elgesio sutrikimais pagal amžiaus grupes, nuo 0 iki 17 metų imtinai ir nuo 18 metų iki pat gyvenimo galo, patogu šias abi grupes pavaizduoti linijine diagrama (duomenys paimti iš 1 priedo):

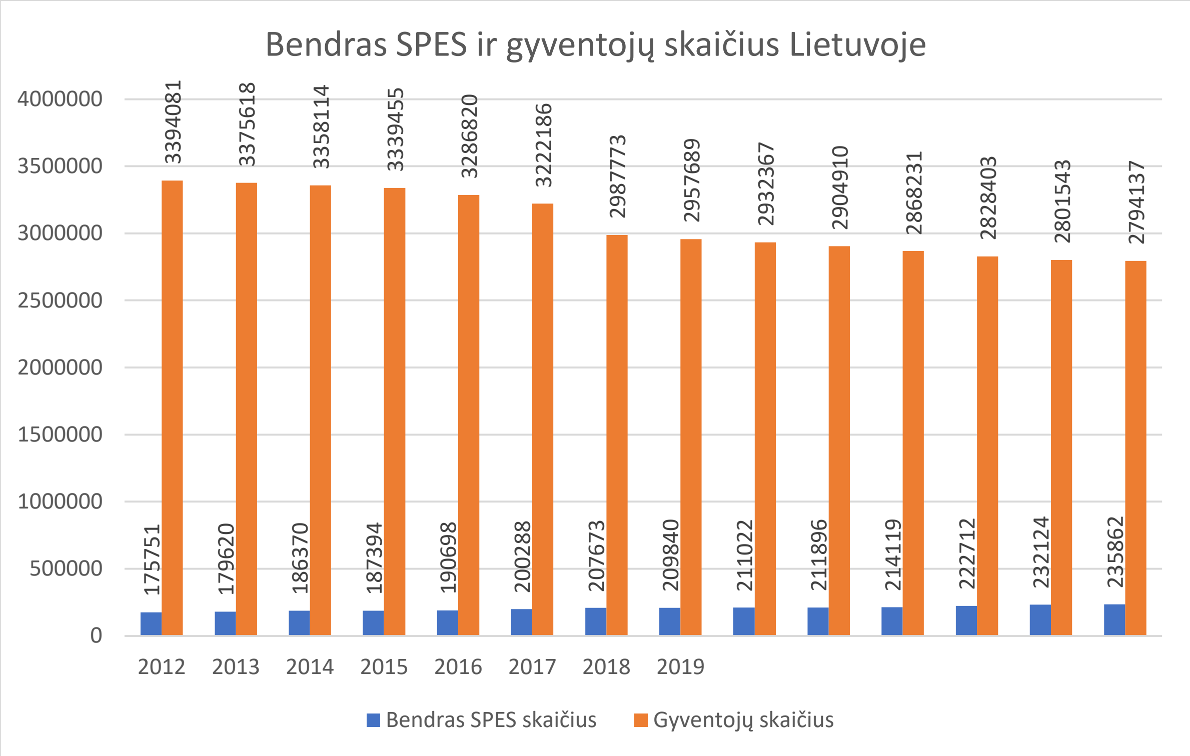


2 pav. *SPES skaičius pagal amžiaus grupes*

Iš šios diagramos galime pastebėti, kad sergančiųjų psichikos ir elgesio sutrikimais vaikų (nuo 0 iki 17 m. imtinai) kreivė yra gana tolydi, sergančiųjų skaičiai svyruoja nežymiai. Iš sergančiųjų psichikos ir elgesio sutrikimais suaugusiųjų (nuo 18m. iki gyvenimo galo) kreivės matome, kad suaugusiųjų skaičiai beveik kiekvienais metais didėja, tad kreivė yra nuolat „auganti“, priešingai nei vaikų iki 17 m. imtinai atveju.

2.2 PRIKLAUSOMYBĖ TARP DUOMENŲ

Ištirsime, ar yra priklausomybė tarp sergančiųjų psichikos ir elgesio sutrikimais ir gyventojų skaičiaus Lietuvoje 2006-2019 metais.



3 pav. *SPES ir gyventojų skaičiai Lietuvoje*

Pasinaudodami 1.5.1 skyrelio teorija, apskaičiuosime empirinį koreliacijos koeficientą tarp sergančiųjų psichikos ir elgesio sutrikimais ir Lietuvos gyventojų skaičiaus.

10 lentelė. *SPES ir gyventojų Lietuvoje duomenys*

	X (Gyventojų skaičius)	**Y (Sergančiųjų skaičius)**
06 m.	3394081	175751
07 m.	3375618	179620
08 m.	3358114	186370
09 m.	3339455	187394
10 m.	3286820	190698
11 m.	3222186	200288
12 m.	2987773	207673
13 m.	2957689	209840
14 m.	2932367	211022
15 m.	2904910	211896
16 m.	2868231	214119
17 m.	2828403	222712
18 m.	2801543	232124
19 m.	2794137	235862

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{14}(3394081 + 3375618 + 3358114 + 3339455 + 3286820 + 3222186 + 2987773 + 2957689 + 2932367 + 2904910 + 2868231 + 2828403 + 2801543 + 2794137) = 3075094{,}786$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{14}(175751 + 179620 + 186370 + 187394 + 190698 + 200288 + 207673 + 209840 + 211022 + 211896 + 214119 + 222712 + 232124 + 235862) = 204669{,}2143$$

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = -55786758404{,}357.$$

$$\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 58094522311.$$

Tada viską surašome į formulę ir gauname

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{-55786758404{,}357}{58094522311} = -0{,}960.$$

r(L(serg.),L(gyv.))= -0,960

Kadangi empirinis koreliacijos koeficientas yra pakankamai didelis, patikrinsime statistinę hipotezę apie galimą ρ reikšmę. Sužinosime, ar tarp 2006-2019 metų sergančiųjų ir gyventojų yra priklausomybė. Imties tūris $n = 14$.

Šiems duomenims koreliacijos koeficientą apskaičiavome anksčiau, todėl pasinaudodami 1.5.2 skyrelio teorija, iškart galime taikyti R. A. Fišerio transformaciją:

$$Arthr = \frac{1}{2} ln \frac{1+r}{1-r}$$

Įrašome duomenis

$$Arthr = \frac{1}{2} ln \frac{1+(-0{,}960)}{1-(-0{,}960)} = -1{,}946$$

Hipotezė $H_0 : \rho = -0{,}9$, tai

$$Arthr \sim N\left(Arth\rho_0, \frac{1}{n-3}\right)$$

$$Arth(-0{,}9) = \frac{1}{2} ln \frac{1+(-0{,}9)}{1-(-0{,}9)} = -1{,}472$$

Todėl

$$Z = \sqrt{n-3}(Arthr - Arth\rho_0) \sim N(0,1)$$

$$Z = \sqrt{14-3} \cdot \left(-1{,}946 - (-1{,}472)\right) = -1{,}572$$

Tuo remdamiesi hipotezei $H_o: \rho = \rho_0$ tikrinti su reikšmingumo lygmeniu $\alpha = 0{,}01$ sudarome kritinę sritį:

$$W_1 = \left\{(X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}}\right\}, \text{ jei } H_o \text{ alternatyva yra } H_1^{(1)}: \rho \neq \rho_0.$$

Kadangi $|Z| = 1{,}572 \leq 2{,}576 = u_{0{,}995}$, tai hipotezė H_o priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = -0{,}9$ teisingumui.

Galime teigti, jog Lietuvos sergančiųjų skaičius yra atvirkščiai proporcingas gyventojų skaičiui, kadangi koreliacijos koeficientas yra neigiamas ir ganėtinai artimas vienetui.

Išvada. Kadangi empirinis koreliacijos koeficientas yra artimas minus vienetui, vadinasi, egzistuoja atvirkštinė tiesinė priklausomybė tarp gyventojų ir sergančiųjų psichikos ir elgesio sutrikimais Lietuvoje. Kadangi turime labai stiprią koreliaciją, galime aprašyti šią priklausomybę tiesine lygtimi (pagal 1.5.3 skyrelio teoriją). Tolimesniame tyrime rasime tiesinės priklausomybės lygtį.

Tiesinę priklausomybę nusako tokia regresijos lygtis:

$$y = \beta_0 + \beta_1 x + \varepsilon, \qquad \varepsilon \sim N(0, \varsigma^2)$$

Rasime tos lygties parametrų įverčius:

$$y = \widehat{\beta_0} + \widehat{\beta_1} x + \varepsilon, \ \varepsilon \sim N(0, \hat{\sigma}^2), \hat{\sigma}^2 = S^2$$

Apibrėžkime kintamuosius:

y – SPES skaičius,

x – gyventojų skaičius.

Tiesinės regresijos radimas:

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 3075094{,}786.$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 204669{,}214.$

$Q_X = \sum_1^n (X_i - \bar{X})^2 = 736676612148{,}357.$

$Q_{X,Y} = \sum_1^n (X_i - \bar{X})(Y_i - \bar{Y}) = -\ 55786758404{,}357.$

$\widehat{\beta_1} = \widehat{\beta_1} = \frac{Q_{X,Y}}{Q_X} = \frac{-55786758404{,}357}{736676612148{,}357} = -0{,}0757.$

$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1} \cdot \bar{X} = 204669{,}214 - (-0{,}076) \cdot 3075094{,}786 = 437539.$

$y = 437538{,}790 - 0{,}076 \cdot x + \varepsilon, \ \varepsilon \sim N(0, \hat{\sigma}^2).$

$S^2 = \frac{1}{14-1}\sum_{i=1}^{14}\left(Y_i - \widehat{\beta_0} - \widehat{\beta_1} \cdot X_i\right)^2 = 1444406{,}189.$

$S = \sqrt{S^2} = 1201{,}835.$

Apskaičiavę gauname tiesinės regresijos lygtį:

$$y = 437539 - 0{,}0757 \cdot x + \varepsilon, \qquad \varepsilon \sim N(0; 1444406{,}189)$$

Kadangi empirinis koreliacijos koeficientas gautas didelis, galima teigti, kad tiesinė priklausomybė tarp šių duomenų yra didelė.

Išvada. Pagal šią tiesinės regresijos lygtį galima prognozuoti, kiek pagal gyventojų skaičių bus sergančiųjų psichikos ir elgesio sutrikimais. Mažėjant gyventojų skaičiui, sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius didėja. Iš tiesinės regresijos lygties matome, kad koeficientas prie x (gyventojų skaičius) yra –0,0757 — kiekvienas gyventojas tiek sumažina sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičių pridėjus konstantą.

Taip pat galime užrašyti tiesinės regresijos lygtį programos SPSS pagalba. Žinome, kad gyventojų skaičiaus ir sergančiųjų gyventojų koreliacijos koeficientas yra didesnis nei vidutinis. Patikrinsime, ar modelis pilnai tinkamas prognozuoti duomenims.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,960[a]	,922	,916	5452,464

a. Predictors: (Constant), Gyventojų skaičius

b. Dependent Variable: Sergančiųjų skaičius

***4 pav.** „Model Summary“ lentelės rezultatai*

Matome, kad „R Square“ yra 0,922, o tai reiškia, kad pagal turimus duomenis 92,2% sergančiųjų psichikos ir elgesio sutrikimais gyventojų pokyčių yra susiję su gyventojų skaičiaus pokyčiais. Regresijos paklaida gauta labai didelė – 5452,464, todėl galima abejoti modelio tinkamumu.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4224597824,8	1	4224597824,8	142,102	<,001[b]
	Residual	356752345,55	12	29729362,129		
	Total	4581350170,4	13			

a. Dependent Variable: Sergančiųjų skaičius

b. Predictors: (Constant), Gyventojų skaičius

***5 pav.** „ANOVA“ lentelės rezultatai*

Iš „ANOVA“ lentelės rezultatų matome, kad modelis iš dalies yra tinkamas prognozavimui, nes „Sig“ reikšmė yra mažesnė už 0,05.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	99,0% Confidence Interval for B Lower Bound	99,0% Confidence Interval for B Upper Bound
1	(Constant)	437538,790	19589,252		22,336	<,001	377702,645	497374,934
	Gyventojų skaičius	-,076	,006	-,960	-11,921	<,001	-,095	-,056

a. Dependent Variable: Sergančiųjų skaičius

6 pav. *„Coefficients" lentelės rezultatai*

Pagal koeficientų lentelę gauta lygtis atrodo taip:

$$y = 437538{,}790 - 0{,}076 \cdot x$$

Ši daugiamatė tiesinės regresijos lygtis yra taikytina, nes atmesta nulinė hipotezė apie koeficientų lygybę nuliui (sig. stulpelio reikšmė koeficientui yra <0,05). Galima daryti išvadą, kad apskaičiuota tiesinės regresijos lygtis tarp gyventojų ir tarp sergančiųjų psichikos ir elgesio sutrikimais gyventojų prognozavimui tinka. Prognozuojant sergančiųjų skaičių gyventojų skaičius yra svarbus ir regresijos modelyje yra reikalingas. Užrašykime modelio parametrų a ir b 99% pasikliautinuosius intervalus: koeficiento a pasikliautinasis intervalas yra (377702,645; 497374,934), koeficiento b pasikliautinasis intervalas yra (–0,095; –0,056).

Taip pat galime patikrinti Kuko mato reikšmę, kuri esant tinkamam modeliui turėtų būti mažesnė už 1:

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	180513,16	225945,47	204669,21	18026,898	14
Std. Predicted Value	-1,340	1,180	,000	1,000	14
Standard Error of Predicted Value	1559,245	2495,965	2039,353	308,002	14
Adjusted Predicted Value	181775,63	223974,77	204588,25	17869,942	14
Residual	-6215,514	9916,524	,000	5238,558	14
Std. Residual	-1,140	1,819	,000	,961	14
Stud. Residual	-1,222	2,007	,007	1,041	14
Deleted Residual	-7140,308	12072,443	80,961	6156,346	14
Stud. Deleted Residual	-1,250	2,357	,037	1,104	14
Mahal. Distance	,135	1,796	,929	,545	14
Cook's Distance	,001	,438	,089	,114	14
Centered Leverage Value	,010	,138	,071	,042	14

a. Dependent Variable: Sergančiųjų skaičius

7 pav. *„Residual Statistics" lentelės rezultatai*

Kuko mato reikšmės atitinka reikalavimus, todėl lieka patikrinti standartizuotų liekanų normalumą.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
	,135	14	,200*	,957	14	,667

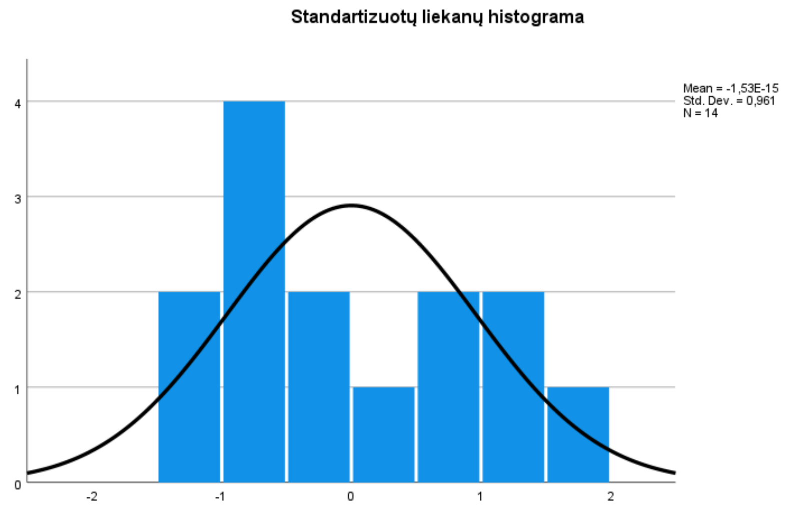
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

8 pav. „Tests of Normality“ lentelės rezultatai

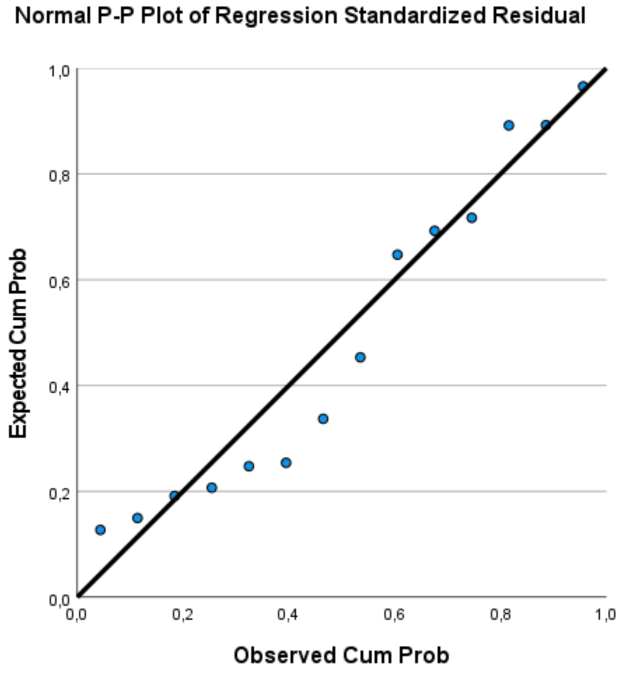
Shapiro-Wilk testo rezultatai rodo, kad standartizuotos liekanos yra pasiskirsčiusios normaliai. Galime teigti, kad tiesinės regresijos modelis, nusakantis sergančiųjų psichikos ir elgesio sutrikimais skaičių pokyčius kintant gyventojų skaičiui, yra tinkamas.

Nubraižysime standartizuotų liekanų histogramą su normaliąja kreive ir standartizuotų liekanų sklaidos diagramą.



9 pav. *Standartizuotų liekanų histograma*

Pagal standartizuotųjų liekanų sklaidos diagramą galime spręsti, ar duomenys yra labai išsisklaidę. Jei regresijos modelis tinkamas, tai standartizuotosios liekanos turėtų būti nelabai didelės ir beveik vienodai išsibarščiusios apie tiesę $y = 0$.



10 pav. *P-P plot grafikas*

Iš grafikų galime matyti, kad visus nagrinėjamų duomenų reikalavimus standartizuotosios liekanos tenkina – tiesinės regresijos modelis tinkamas.

Patikrinkime tiesinės regresijos modelį 2019 metų duomenims. Naudojame duomenis iš 10 lentelės:

$$y = 437538{,}790 - 0{,}076 \cdot x + \varepsilon,$$

$$y = 437538{,}790 - 0{,}076 \cdot 2794137 + \varepsilon,$$

Gauname

$$y = 225184{,}378 + \varepsilon$$

Pagal 1.5.4 skyrelio teoriją apskaičiuokime prognozės intervalą. Esant fiksuotam x, galima prognozuoti galimų y reikšmių intervalą. Apskaičiuokime sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičiaus, esant 2794137 gyventojų (pagal 2019 m. duomenis), 95% prognozės intervalą.

Jau buvome radę

$$\hat{y}(2794137) = 225184{,}378.$$

Stjudento kriterijus su $(n - 2)$ laisvės laipsnių

$$t_{1-\frac{\alpha}{2}}(14-2) = t_{1-\frac{0{,}05}{2}}(14-2) = t_{0{,}975}(12) = 2{,}179;$$

$$\bar{x} = 3075094{,}79;$$

$$(n-1)s_x^2 = (14-1)s_x^2 = 13s_x^2 = 736676612148{,}36;$$

$$SSE = \sum_{i=1}^{n} \hat{e}_i^2 = 366629693{,}993\,;$$

$$MSE = \frac{SSE}{(n-2)} = \frac{366629693{,}993}{12} = 30552474{,}499;$$

$$SY = MSE\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}\right) =$$

$$= 30552474{,}499\left(1 + \frac{1}{14} + \frac{(2794137 - 3075094{,}79)^2}{736676612148{,}36}\right) =$$

$$= 36008590{,}388.$$

Prognozės intervalas apskaičiuojamas pagal formulę:

$$\left[\hat{y}(x) - \sqrt{SY}t_{0,975}(n-2); \hat{y}(x) + \sqrt{SY}t_{0,975}(n-2)\,\right]$$

Gauname

$$\left[225184{,}378 - \sqrt{36008590{,}388} \cdot 2{,}179; 225184{,}378 + \sqrt{36008590{,}388} \cdot 2{,}179\,\right] \rightarrow$$

$$\rightarrow [212108{,}82;\ 238259{,}94]$$

Taigi, su 95% pasikliovimu galime prognozuoti, kad 2019 metais, kai gyventojų skaičius yra 2794137, sergančiųjų psichikos ir elgesio sutrikimais skaičius bus nuo 212108,82 iki 238259,94.

Išvada. 2019 metais sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius Lietuvoje buvo 235862. Įrašius 2019 metų gyventojų skaičių Lietuvoje į gautą lygtį gavome $y = 225184{,}378 + \varepsilon$. Matome, kad su paklaida tiesinės regresijos lygtis yra teisinga. Mūsų prognozuojama reikšmė patenka į prognozės intervalą, vadinasi, sudarytas tiesinės regresijos modelis ne tik teoriškai yra tinkamas, bet ir tinkamai prognozuoja praktikoje.

2.3 SERGANČIŲJŲ SKAIČIAUS POKYČIO TYRIMAS PAGAL SUTRIKIMUS

Pagal Higienos instituto Sveikatos informacijos centro leidinį, turimi duomenys apie sergančiųjų psichikos ir elgesio sutrikimais žmonių skaičius Lietuvoje 2006-2019 m. laikotarpiu, kurie yra išskaidyti į 7 smulkesnes grupes, pagal ligas ar elgesio sutrikimus.

Visi šie psichikos ir elgesio sutrikimai priklauso tai pačiai ligų grupei pagal TLK-10-AM klasifikatorių. TLK-10-AM – Pasaulio sveikatos organizacijos *Tarptautinė statistinė ligų ir sveikatos sutrikimų klasifikacija.*

11 lentelė. *Psichikos ir elgesio sutrikimų sutrumpinti žymėjimai ir kodai pagal TLK-10-AM klasifikatorių*

Sutrumpintas žymėjimas	**Diagnozės pavadinimas**	**Kodas pagal TLK-10-AM**
	Psichikos ir elgesio sutrikimai, iš jų:	F00-F99
1	demencija ir Alzheimerio liga	F00-F03
2	psichikos ir elgesio sutrikimai vartojant alkoholį	F10
3	psichikos ir elgesio sutrikimai vartojant kitas psichoaktyviąsias medžiagas	F11-F19
4	šizofrenija, šizotipinis ir kliedesiniai sutrikimai	F20-F29
5	nuotaikos sutrikimai	F30-F39
6	protinis atsilikimas ir psichologinės raidos sutrikimai	F70-F89
7	elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje	F90-F98

2.3.1 POLINOMINIO SKIRSTINIO TAIKYMAS GRUPIŲ DUOMENIMS

Norint sužinoti, ar sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kitimas kiekvienoje ligos grupėje tiriamuoju laikotarpiu yra statistiškai reikšmingas, taikoma statistinė hipotezė su polinominiu skirstiniu, taip kaip buvo taikoma V. Kanišausko straipsnyje „Vidinės migracijos matematinis modelis" [7]. Polinominio skirstinio matematinį modelį taikysime tiriant gyventojų sergamumo kiekvienoje ligų grupėje dinamiką (žr. 3 priedą).

Pradedame nuo 2006-2010 metų duomenų:

12 lentelė. *2006m. ir 2010m. SPES duomenys*

Ligos pogrupis	**Iš viso 2006m.**	**Procentinė dalis 2006m.**	**Iš viso 2010m.**	**Procentinė dalis 2010m.**
Iš viso	175751		190698	
1	11340	6,45%	21233	11,13%
2	10686	6,08%	9542	5,00%
3	714	0,41%	958	0,50%
4	17042	9,70%	18096	9,49%
5	37818	21,52%	44534	23,35%
6	18532	10,54%	24273	12,73%
7	9862	5,61%	10429	5,47%

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais kiekio procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2006$ į $t_2 = 2010$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,05$. Turimi kiekvienos grupės procentinės dalies kiekiai lyginant su bendru sergančiųjų skaičiumi. Grupių procentinės dalys, padalintos iš 100, laikomos polinominio skirstinio parametrais. Taikant statistinę hipotezę apie polinominio skirstinio parametrus tikrinama

$H_0: p_1 = 0{,}0645, p_2 = 0{,}0608, p_3 = 0{,}0041, \quad p_4 = 0{,}0970, \quad p_5 = 0{,}2152, \quad p_6 = 0{,}1054, p_7 = 0{,}0561,$

kuri sako, kad praėjus ketveriems metams nuo 2006 iki 2010 sergančiųjų skaičius (procentinė dalis) nepakito, t. y. sergančiųjų skaičiaus kitimas nevyko. 2010 m. p_i reikšmės virto $\hat{p}_i$, kai $i = 1,2,\dots,7$. Mūsų atveju 2010 m. procentinės dalys, padalintos iš 100 bus: $p_1 = 0{,}1113, p_2 = 0{,}0500, p_3 = 0{,}0050, p_4 = 0{,}0949, p_5 = 0{,}2335, p_6 = 0{,}1273, p_7 = 0{,}0547$. Naujoje situacijoje $n = n(2010) = 190698$ (visų Lietuvos sergančiųjų skaičius 2010 m.). Jei įverčiai (su atitinkamomis procentinėmis dalimis) nežymiai skirsis nuo ankstesnių metų (2006 m.) polinominio skirstinio parametrų, suformuluotų statistinėje hipotezėje, tada juos atitinkanti Pirsono statistika bus maža, t.y. mažesnė už atitinkamą Chi-kvadrat $\chi^2_{1-\alpha}(k-1)$ kvantilį. Tuo atveju nulinė hipotezė būtų priimama ir būtų galima teigti, kad per nagrinėjamą laikotarpį, nuo 2006 iki 2010 m. Lietuvos žmonių sergamumas psichikos ir elgesio sutrikimais pagal TLK-10-AM klasifikatorių statistiškai nepakito su tikimybe $1-\alpha$. Priešingu atveju — daroma išvada, jog gyventojų sergamumo skaičiai, išskirstyti į smulkesnes grupes, pakito statistiškai reikšmingai ir reikalauja papildomų duomenų bei tyrimų, norint išsiaiškinti bei suprasti tokių pokyčių priežastis. Viską surašome į Pirsono statistiką ir gauname:

$$\chi^2 = 190698\left(\frac{(0{,}1113-0{,}0645)^2}{0{,}0645}+\frac{(0{,}0500-0{,}0608)^2}{0{,}0608}+\frac{(0{,}0050-0{,}0041)^2}{0{,}0041}\right.$$
$$+\frac{(0{,}0949-0{,}0970)^2}{0{,}0970}+\frac{(0{,}2335-0{,}2152)^2}{0{,}2152}+\frac{(0{,}1273-0{,}1054)^2}{0{,}1054}$$
$$\left.+\frac{(0{,}0547-0{,}0561)^2}{0{,}0561}\right) = 80{,}6222$$

Į formulę $\chi^2_{1-\alpha}(k-1)$ įrašome $k = 7$, $\alpha = 0{,}05$ ir gauname $\chi^2_{0{,}95}(6) = 12{,}6$ (reikšmė paimta iš kvantilių lentelės).

$\chi^2 = 80{,}6222 > \chi^2_{0{,}95}(6) = 12{,}6$, todėl hipotezę H_0 apie tai, kad tarp sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kaita nevyksta, atmetame.

Išvada. Sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius nuo 2006 m. iki 2010 m. iš esmės statistiškai pakito, pokyčiai buvo reikšmingi. Gyventojų sergamumas psichikos ir elgesio sutrikimais ir ligomis iš esmės statistiškai pakito per ketverius metus nuo 2006 m. iki 2010 m.

2010-2014 metų atvejis

13 lentelė. *2010m. ir 2014m. SPES duomenys*

Ligų grupės pagal TLK-10-AM	Iš viso 2010m.	Procentinė dalis 2010m.	Iš viso 2014m.	Procentinė dalis 2014m.
Iš viso	190698		214119	
1	21233	11,13%	30806	14,39%
2	9542	5,00%	19373	9,05%
3	958	0,50%	1844	0,86%
4	18096	9,49%	18872	8,81%
5	44534	23,35%	47584	22,22%
6	24273	12,73%	28655	13,38%
7	10429	5,47%	11407	5,33%

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais kiekio procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2010$ į $t_2 = 2014$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,05$. Tikriname statistinę hipotezę:

$H_0: p_1 = 0,1113, p_2 = 0,0500, p_3 = 0,0050, \quad p_4 = 0,0949, \quad p_5 = 0,2335, \quad p_6 = 0,1273, p_7 = 0,0547,$

kuri sako, kad praėjus ketveriems metams nuo 2010 iki 2014 sergančiųjų skaičius (procentinė dalis) nepakito, t. y. sergančiųjų skaičiaus kitimas nevyko. 2014 m. p_i reikšmės virto $\hat{p}_i$, kai $i = 1,2,\ldots,7$. Mūsų atveju 2014 m. procentinės dalys, padalintos iš 100 bus: $p_1 = 0,1439, p_2 = 0,0905, p_3 = 0,0086, p_4 = 0,0881, p_5 = 0,2222, p_6 = 0,1338, p_7 = 0,0533$. Naujoje situacijoje $n = n(2014) = 214119$ (visų Lietuvos sergančiųjų skaičius 2014 m.). Viską surašome į Pirsono statistiką ir gauname:

$$\chi^2 = 214119 \left(\frac{(0,1439 - 0,1113)^2}{0,1113} + \frac{(0,0905 - 0,0500)^2}{0,0500} + \frac{(0,0086 - 0,0050)^2}{0,0050} + \frac{(0,0881 - 0,0949)^2}{0,0949} + \frac{(0,2222 - 0,2335)^2}{0,2335} + \frac{(0,1338 - 0,1273)^2}{0,1273} + \frac{(0,0533 - 0,0547)^2}{0,0547} \right) = 98,8198$$

$\chi^2 = 98,8198 > \chi^2_{0,95}(6) = 12,6$, todėl hipotezę H_0 apie tai, kad tarp sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kaita nevyksta, atmetame.

Išvada. Sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius nuo 2010 m. iki 2014 m. iš esmės statistiškai pakito, pokyčiai buvo reikšmingi. Gyventojų sergamumas psichikos ir elgesio sutrikimais ir ligomis iš esmės statistiškai pakito per penkeris metus nuo 2010 m. iki 2014 m.

2014-2018 metų atvejis

14 lentelė. *2014m. ir 2018m. SPES duomenys*

Ligų grupės pagal TLK-10-AM	Iš viso 2014m.	Procentinė dalis 2014m.	Iš viso 2018m.	Procentinė dalis 2018m.
Iš viso	214119		232124	
1	30806	14,39%	32362	13,94%
2	19373	9,05%	18308	7,89%
3	1844	0,86%	3001	1,29%
4	18872	8,81%	18588	8,01%
5	47584	22,22%	46329	19,96%
6	28655	13,38%	29501	12,71%
7	11407	5,33%	11145	4,80%

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais kiekio procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2014$ į $t_2 = 2018$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,05$. Tikriname statistinę hipotezę:

$H_0 \colon p_1 = 0{,}1439, p_2 = 0{,}0905, p_3 = 0{,}0086, p_4 = 0{,}0881, p_5 = 0{,}2222, p_6 = 0{,}1338, p_7 = 0{,}0533,$

kuri sako, kad praėjus ketveriems metams nuo 2014 iki 2018 sergančiųjų skaičius (procentinė dalis) nepakito, t. y. sergančiųjų skaičiaus kitimas nevyko. 2018 m. p_i reikšmės virto $\hat{p}_i$, kai $i = 1,2,\ldots,7$. Mūsų atveju 2018 m. procentinės dalys, padalintos iš 100 bus: $p_1 = 0{,}1394, p_2 = 0{,}0789, p_3 = 0{,}0129, p_4 = 0{,}0801, p_5 = 0{,}1996, p_6 = 0{,}1271, p_7 = 0{,}0480.$ Naujoje situacijoje $n = n(2018) = 232124$ (visų Lietuvos sergančiųjų skaičius 2018 m.). Viską surašome į Pirsono statistiką ir gauname:

$$\chi^2 = 232124 \left(\frac{(0{,}1394 - 0{,}1439)^2}{0{,}1439} + \frac{(0{,}0789 - 0{,}0905)^2}{0{,}0905} + \frac{(0{,}0129 - 0{,}0086)^2}{0{,}0086} + \frac{(0{,}0801 - 0{,}0881)^2}{0{,}0881} + \frac{(0{,}1996 - 0{,}2222)^2}{0{,}2222} + \frac{(0{,}1271 - 0{,}1338)^2}{0{,}1338} + \frac{(0{,}0480 - 0{,}0533)^2}{0{,}0533} \right) = 17{,}8578$$

$\chi^2 = 17{,}8578 > \chi^2_{0,95}(6) = 12{,}6$, todėl hipotezę H_0 apie tai, kad tarp sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kaita nevyksta, atmetame.

Išvada. Sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius nuo 2014 m. iki 2018 m. iš esmės statistiškai pakito, pokyčiai buvo reikšmingi. Gyventojų sergamumas psichikos ir elgesio sutrikimais ir ligomis iš esmės statistiškai pakito per penkeris metus nuo 2014 m. iki 2018 m.

Analogiškai galima patikrinti pokyčius per vienerius ir dvejus metus. Pirmiausia patikrinsime hipotezės teisingumą praėjus vieneriems metams:

2018-2019 metų atvejis

15 lentelė. *2018m. ir 2019m. SPES duomenys*

Ligų grupės pagal TLK-10-AM	Iš viso 2018m.	Procentinė dalis 2018m.	Iš viso 2019m.	Procentinė dalis 2019m.
Iš viso	232124		235862	
1	32362	13,94%	32617	13,83%
2	18308	7,89%	18632	7,90%
3	3001	1,29%	3278	1,39%
4	18588	8,01%	18603	7,89%
5	46329	19,96%	46345	19,65%
6	29501	12,71%	30445	12,91%
7	11145	4,80%	11247	4,77%

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais kiekio procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2018$ į $t_2 = 2019$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,05$. Tikriname statistinę hipotezę:

$H_0: p_1 = 0{,}1394, p_2 = 0{,}0789, p_3 = 0{,}0129, p_4 = 0{,}0801, p_5 = 0{,}1996, p_6 = 0{,}1271, p_7 = 0{,}0480,$

kuri sako, kad praėjus vieneriems metams nuo 2018 iki 2019 sergančiųjų skaičius (procentinė dalis) nepakito, t. y. sergančiųjų skaičiaus kitimas nevyko. 2019 m. p_i reikšmės virto $\hat{p}_i$, kai $i = 1,2,\dots,7$. Mūsų atveju 2019 m. procentinės dalys, padalintos iš 100 bus: $p_1 = 0{,}1383, p_2 = 0{,}0790, p_3 = 0{,}0139, p_4 = 0{,}0789, p_5 = 0{,}1965, p_6 = 0{,}1291, p_7 = 0{,}0477$. Naujoje situacijoje $n = n(2019) = 235862$ (visų Lietuvos sergančiųjų skaičius 2019 m.). Viską surašome į Pirsono statistiką ir gauname:

$$\chi^2 = 235862\left(\frac{(0{,}1383-0{,}1394)^2}{0{,}1394}+\frac{(0{,}0790-0{,}0789)^2}{0{,}0789}+\frac{(0{,}0139-0{,}0129)^2}{0{,}0129}+\frac{(0{,}0789-0{,}0801)^2}{0{,}0801}+\frac{(0{,}1965-0{,}1996)^2}{0{,}1996}+\frac{(0{,}1291-0{,}1271)^2}{0{,}1271}+\frac{(0{,}0477-0{,}0480)^2}{0{,}0480}\right)=0{,}4282$$

$\chi^2 = 0{,}4282 < \chi^2_{0,95}(6) = 12{,}6$, todėl hipotezę H_0 apie tai, kad tarp sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kaita nevyksta, priimame.

Išvada. Sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius nuo 2018 m. iki 2019 m. iš esmės statistiškai nepakito, pokyčiai nebuvo reikšmingi. Gyventojų sergamumas psichikos ir elgesio sutrikimais ir ligomis iš esmės statistiškai nepakito per metus nuo 2018 m. iki 2019 m.

2017-2019 metų atvejis

16 lentelė. *2017m. ir 2019m. SPES duomenys*

Ligų grupės pagal TLK-10-AM	Iš viso 2017m.	Procentinė dalis 2017m.	Iš viso 2019m.	Procentinė dalis 2019m.
Iš viso	222712		235862	
1	33056	14,84%	32617	13,83%
2	19127	8,59%	18632	7,90%
3	2269	1,02%	3278	1,39%
4	18765	8,43%	18603	7,89%
5	46873	21,05%	46345	19,65%
6	29677	13,33%	30445	12,91%
7	11515	5,17%	11247	4,77%

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais kiekio procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2017$ į $t_2 = 2019$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,05$. Tikriname statistinę hipotezę:

$H_0^{(1)}: p_1 = 0{,}1484, p_2 = 0{,}0859, p_3 = 0{,}0102, p_4 = 0{,}0843, p_5 = 0{,}2105, p_6 = 0{,}1333, p_7 = 0{,}0517,$

kuri sako, kad praėjus dvejiems metams nuo 2017 iki 2019 sergančiųjų skaičius (procentinė dalis) nepakito, t. y. sergančiųjų skaičiaus kitimas nevyko. 2019 m. p_i reikšmės virto $\hat{p}_i$, kai *i* = 1,2,..,7. Mūsų atveju 2019 m. procentinės dalys, padalintos iš 100 bus: $p_1 = 0{,}1383, p_2 = 0{,}0790, p_3 = 0{,}0139, p_4 = 0{,}0789, p_5 = 0{,}1965, p_6 = 0{,}1291, p_7 = 0{,}0477.$

Naujoje situacijoje $n\ = n(2019) = 235862$ (visų Lietuvos sergančiųjų skaičius 2019 m.). Viską surašome į Pirsono statistiką ir gauname:

$$\chi^2 = 235862\left(\frac{(0{,}1383-0{,}1484)^2}{0{,}1484}+\frac{(0{,}0790-0{,}0859)^2}{0{,}0859}+\frac{(0{,}0139-0{,}0102)^2}{0{,}0102}\right.$$
$$+\frac{(0{,}0789-0{,}0843)^2}{0{,}0843}+\frac{(0{,}1965-0{,}2105)^2}{0{,}2105}+\frac{(0{,}1291-0{,}1333)^2}{0{,}1333}$$
$$\left.+\frac{(0{,}0477-0{,}0517)^2}{0{,}0517}\right)=10{,}1662$$

$\chi^2 = 10{,}1662 < \chi^2_{0,95}(6) = 12{,}6$ todėl hipotezę $H_0^{(1)}$ apie tai, kad tarp sergančiųjų psichikos ir elgesio sutrikimais skaičiaus kaita nevyksta, priimame.

Išvada. Sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius nuo 2018 m. iki 2019 m. iš esmės statistiškai nepakito, pokyčiai nebuvo reikšmingi. Gyventojų sergamumas psichikos ir elgesio sutrikimais ir ligomis iš esmės statistiškai nepakito per dvejus metus nuo 2017 m. iki 2019 m.

2.3.2 SERGAMUMO POKYČIŲ LIETUVOJE DINAMIKA

Psichikos ir elgesio sutrikimų ligos pagal TLK-10-AM klasifikatorių yra išskaidytos į dar 7 smulkesnes grupes. Pagal turimus duomenis (žr. 3 priedą) galime matyti sergančiųjų psichikos ir elgesio sutrikimais skaičiaus judėjimą Lietuvoje pagal ligas bei sutrikimus, t.y. kur sergančiųjų daugėja ar mažėja kiekvienais metais, o tam reikalingi duomenys pagal kasmetinius pokyčius nuo 2006m. iki 2019m.

17 lentelė. *Kasmetiniai SPES pokyčiai*

Ligos pogrupis	06-07 m.	07-08 m.	08-09 m.	09-10 m.	10-11 m.	11-12 m.	12-13 m.
1	1,29%	1,36%	1,41%	0,61%	1,43%	1,24%	0,11%
2	0,52%	-0,80%	-0,90%	0,11%	1,14%	0,60%	1,96%
3	-0,02%	0,07%	0,07%	-0,02%	0,14%	0,04%	0,14%
4	-0,14%	-0,23%	0,28%	-0,12%	-0,18%	-0,22%	0,02%
5	0,92%	0,29%	0,32%	0,30%	0,02%	0,02%	-0,59%
6	0,24%	0,44%	0,89%	0,62%	0,28%	0,05%	-0,06%
7	-0,11%	-0,15%	0,11%	0,00%	-0,26%	0,09%	-0,17%

Ligos pogrupis	13-14 m.	14-15 m.	15-16 m.	16-17 m.	17-18 m.	18-19 m.
1	0,48%	0,57%	0,70%	-0,81%	-0,90%	-0,11%
2	0,35%	0,39%	0,21%	-1,06%	-0,70%	0,01%
3	0,04%	0,16%	-0,01%	0,01%	0,27%	0,10%
4	-0,30%	0,00%	0,14%	-0,53%	-0,42%	-0,12%
5	-0,59%	-0,52%	0,05%	-0,71%	-1,09%	-0,31%
6	0,38%	0,28%	-0,08%	-0,25%	-0,62%	0,20%
7	0,19%	0,07%	-0,29%	0,07%	-0,37%	-0,03%

Iš tokios duomenų lentelės sunku ką nors pasakyti, todėl panaudosime vizualinį metodą, sergamumo pokyčius pažymėsime spalvomis. Mus domina sergančiųjų psichikos ir elgesio sutrikimais skaičiaus judėjimas Lietuvoje pagal ligos pogrupius (sutrikimus), t.y. kokių sutrikimų daugėja ar mažėja kiekvienais metais, todėl panaudosime duomenis pagal kasmetinius pokyčius nuo 2006 m. iki 2019 m.

Jei sergamumas tam tikroje ligų grupėje per nurodytą laikotarpį didėjo – žymėsime raudona spalva (raudona spalva parodo sergančiųjų skaičiaus padidėjimą procentais):

$[0; 0{,}09]$

$(0{,}09; 0{,}9]$

$> 0{,}9$

Jei sergamumas tam tikroje ligų grupėje per nurodytą laikotarpį mažėjo – žymėsime žalia spalva (žalia spalva parodo sergančiųjų skaičiaus sumažėjimą procentais):

$[-0{,}09; 0)$

$[-0{,}9; -0{,}09)$

$< -0{,}9$

18 lentelė. *SPES dinamika kas vienerius metus*

Ligos pogrupis	06-07 m.	07-08 m.	08-09 m.	09-10 m.	10-11 m.	11-12 m.	12-13 m.
1							
2							
3							
4							
5							
6							
7							

Ligos pogrupis	13-14 m.	14-15 m.	15-16 m.	16-17 m.	17-18 m.	18-19 m.
1						
2						
3						
4						
5						
6						
7						

Iš pirmo žvilgsnio matome, jog dažniausiai kai kuriose sutrikimų grupėse retai kada išlieka ta pati spalva, o tai reiškia, jog sergančiųjų skaičius nėra pastovus, jis nuolat kinta. Pavyzdžiui, pirmaisiais analizuojamais metais sergančiųjų psichikos ir elgesio sutrikimais dėl alkoholio vartojimo (2) skaičius didėjo, antraisiais ir trečiaisiais mažėjo, nuo ketvirtųjų metų vėl pradėjo didėti ir didėjo iki 2014 metų. Sergančiųjų demencija ir Alzheimerio liga(1) bei sergančiųjų psichikos ir elgesio sutrikimais dėl alkoholio vartojimo (2) duomenų eilutės turi unikalius, atitinkamai 10 metų ir 7 metus trunkančius sergančiųjų skaičiaus didėjimo laikotarpius. Unikalių sergančiųjų mažėjimo atvejų nepastebėta, ilgiausias mažėjimo laikotarpis trukdavo 3 metus (Demencija ir Alzheimerio liga (1); Šizofrenija, šizotipinis ir kliedesiniai sutrikimai(4); Nuotaikos, afektiniai sutrikimai (5); Protinis atsilikimas ir psichologinės raidos sutrikimai (6)).

19 lentelė. *SPES dinamika kas ketverius metus*

Ligos pogrupis	06-10 m.	10-14 m.	14-18 m.
1			
2			
3			
4			
5			
6			
7			

Kai pasirenkame analizuoti ilgiausio laikotarpio (4 metų) pokyčius, galime daryti išvadą, jog šis grafikas yra pozityviausias ir atspindi geriausius rezultatus – jog pastaruoju metu, t.y. 2014-2018 metais, daugelyje (6 iš 7) ligų grupių sergančiųjų psichikos ir elgesio sutrikimais gyventojų skaičius sumažėjo (lyginant 2018 metų duomenis su 2014 metų duomenimis). Norint apibūdinti pastarųjų metų situaciją, idealiausiai tinka antrasis grafikas, tačiau kruopščiai analizuojant duomenis, tiksliausiais galime vadinti pirmojo grafiko duomenis — jame pavaizduoti realūs kasmetiniai sergančiųjų psichikos ir elgesio sutrikimais skaičiaus pokyčiai, lyginant kiekvienų metų sergančiųjų skaičius su prieš tai buvusių metų sergančiųjų skaičiais.

Išvados:

1. Žmonių sergamumas demencija ir Alzheimerio liga (1) padidėjo nuo 2006 m. iki 2014 m. žymiai, po to nuo 2014 m. iki 2018 m. sumažėjo vidutiniškai.
2. Žmonių sergamumas psichikos ir elgesio sutrikimais dėl alkoholio vartojimo (2) stipriai sumažėjo nuo 2006 m. iki 2010 m., po to nuo 2010 m. iki 2014 m. stipriai padidėjo ir nuo 2014 iki 2018 m. vėl stipriai sumažėjo.
3. Žmonių sergamumas psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo (3) vidutiniškai padidėjo nuo 2006 m. iki 2018 m.
4. Žmonių sergamumas šizofrenija, šizotipiniu ir kliedesiniais sutrikimais (4) vidutiniškai sumažėjo nuo 2006 m. iki 2018 m.
5. Žmonių sergamumas nuotaikos, afektiniais sutrikimais (5) stipriai padidėjo nuo 2006 m. iki 2010 m., po to nuo 2010 m. iki 2018 m. stipriai sumažėjo.
6. Žmonių sergamumas protiniu atsilikimu ir psichologinės raidos sutrikimais (6) stipriai padidėjo nuo 2006 m. iki 2010 m., po to nuo 2010 m. iki 2014 m. padidėjo vidutiniškai, o nuo 2014 m. iki 2018 m. sumažėjo vidutiniškai.
7. Žmonių sergamumas elgesio ir emocijų sutrikimais, prasidedančiais vaikystėje ir paauglystėje (7) visu analizuojamu laikotarpiu nuo 2006 m. iki 2018 m. vidutiniškai mažėjo.

2.3.3 LYGINAMOJI ANALIZĖ

TLK-10-AM ligų klasifikatoriaus 5-ame punkte (F00-F99 dalyje), kuriame nurodyti psichikos ir elgesio sutrikimų skaičiai, šie sutrikimai yra suskirstyti į 7 smulkesnes sutrikimų-ligų grupes, kurios iš pirmo žvilgsnio skamba gana panašiai, todėl gali kilti klausimas, ar tikrai tos ligos yra tarpusavyje susijusios ir priklausomos. Tyrimui parinksime kelias ligų grupių poras, kurios pagal savo pavadinimą, atrodo, turėtų būti susijusios ir viena nuo kitos priklausomos. Ištirsime, ar tarp šių ligų grupių yra kažkoks ryšys.

Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo-Psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo

Pirmuoju tyrimu išsiaiškinsime, ar žmonių sergamumas **psichikos ir elgesio sutrikimais dėl alkoholio vartojimo (2)** priklauso nuo sergamumo **psichikos ir elgesio**

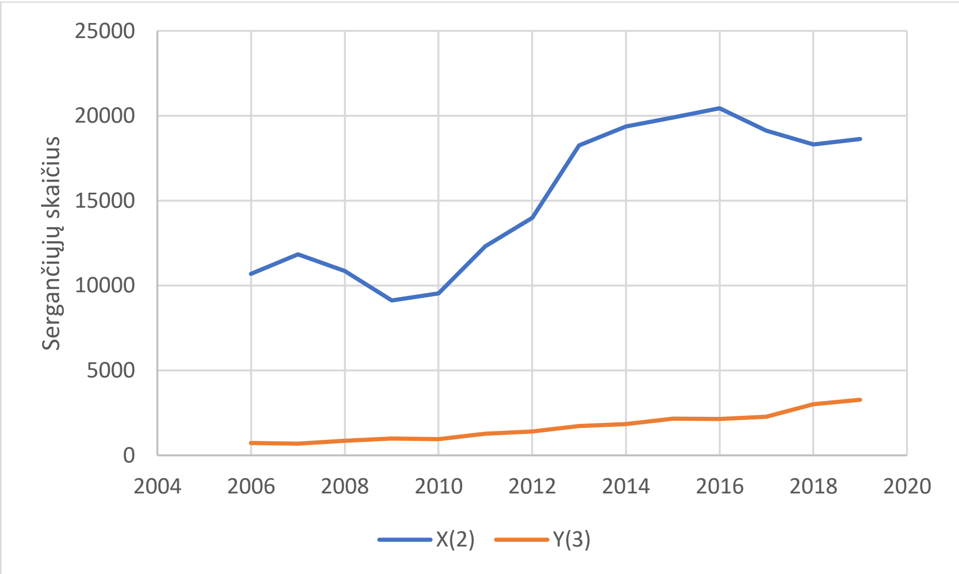
sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo (3). Galbūt šios ligos tarpusavyje nesusijusios (nepriklausomos)?

Iš duomenų lentelės (4 priedas) išrašomi tokie duomenys:

20 lentelė. *(2) ir (3) ligų grupių SPES duomenys*

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
X(2)	10686	11847	10865	9129	9542	12306	14001
Y(3)	714	686	847	979	958	1278	1408

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
X(2)	18259	19373	19908	20445	19127	18308	18632
Y(3)	1722	1844	2153	2132	2269	3001	3278



11 pav. *(2) ir (3) ligų grupių SPES diagrama*

Iš grafiko sunku ką nors pasakyti todėl, kad **(3)** ligos duomenys žymiai mažesni už **(2)** ligos duomenis. Toliau bus taikoma koreliacinė analizė. Tam tikslui ieškosime empirinio koreliacijos koeficiento:

$$r = \frac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}}$$

$$Q_{xy} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$Q_x = \sum_{i=1}^{n} (X_i - \bar{Y})^2 \,, \; Q_y = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

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

__21 lentelė.__ Empirinio koreliacijos koeficiento tarpiniai skaičiavimai

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$
10686	714	-4487,429	-948,071	4254402,816	20137015,184	898839,434
11847	686	-3326,429	-976,071	3246831,888	11065127,041	952715,434
10865	847	-4308,429	-815,071	3511677,031	18562556,755	664341,434
9129	979	-6044,429	-683,071	4128776,459	36535116,755	466586,577
9542	958	-5631,429	-704,071	3964927,959	31712987,755	495716,577
12306	1278	-2867,429	-384,071	1101297,388	8222146,612	147510,862
14001	1408	-1172,429	-254,071	297880,602	1374588,755	64552,291
18259	1722	3085,571	59,929	184913,8878	9520751,041	3591,434
19373	1844	4199,571	181,929	764022,0306	17636400,184	33098,005
19908	2153	4734,571	490,929	2324336,388	22416166,612	241010,862
20445	2132	5271,571	469,929	2477262,031	27789465,327	220832,862
19127	2269	3953,571	606,929	2399535,459	15630727,041	368362,291
18308	3001	3134,571	1338,929	4196967,245	9825538,041	1792729,719
18632	3278	3458,571	1615,929	5588804,388	11961716,327	2611225,148
10686	714	-4487,429	-948,071	4254402,816	20137015,184	898839,434

Q_{xy} =38441635,571*;* Q_x =242390303,429*;* $Q_y = 8961112{,}929$.

Iš čia $r = \frac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}} = \frac{38441635{,}571}{\sqrt{242390303{,}429 \cdot 8961112{,}929}} = 0{,}825$

Išvada. Galima teigti, kad žmonių sergamumas psichikos ir elgesio sutrikimais dėl alkoholio vartojimo priklauso nuo žmonių sergamumo psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo.

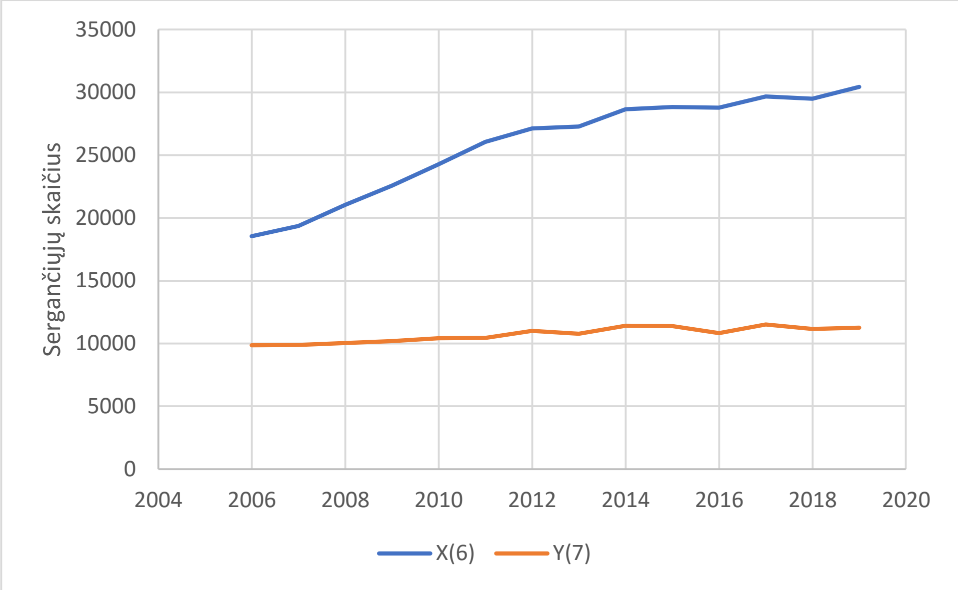
Protinis atsilikimas ir psichologinės raidos sutrikimai-Elgesio ir emocijų sutrikimai prasidedantys vaikystėje ir paauglystėje

Iš duomenų lentelės (4 priedas) išrašomi tokie duomenys:

__22 lentelė.__ (6) ir (7) ligų grupių SPES duomenys

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
X(6)	18532	19367	21028	22573	24273	26057	27114
Y(7)	9862	9887	10031	10186	10429	10442	11014

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
X(6)	27281	28655	28827	28773	29677	29501	30445
Y(7)	10781	11407	11383	10816	11515	11145	11247



12 pav. *(6) ir (7) ligų grupių SPES diagrama*

Iš grafiko sunku ką nors pasakyti todėl, kad **(7)** ligos duomenys žymiai mažesni už **(6)** ligos duomenis. Toliau bus taikoma koreliacinė analizė. Tam tikslui ieškosime empirinio koreliacijos koeficiento:

23 lentelė. *Empirinio koreliacijos koeficiento tarpiniai skaičiavimai*

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
18532	9862	-7332,500	-862,643	6325328,75	53765556,250	744152,699
19367	9887	-6497,500	-837,643	5442584,464	42217506,250	701645,556
21028	10031	-4836,500	-693,643	3354803,679	23391732,250	481140,413
22573	10186	-3291,500	-538,643	1772942,964	10833972,250	290136,128
24273	10429	-1591,500	-295,643	470515,6071	2532872,250	87404,699
26057	10442	192,500	-282,643	-54408,75	37056,250	79886,985
27114	11014	1249,500	289,357	361551,75	1561250,250	83727,556
27281	10781	1416,500	56,357	79829,89286	2006472,250	3176,128
28655	11407	2790,500	682,357	1904117,607	7786890,250	465611,270
28827	11383	2962,500	658,357	1950383,036	8776406,250	433434,128
28773	10816	2908,500	91,357	265712,25	8459372,250	8346,128
29677	11515	3812,500	790,357	3013236,607	14535156,250	624664,413
29501	11145	3636,500	420,357	1528628,75	13224132,250	176700,128
30445	11247	4580,500	522,357	2392656,893	20980980,250	272856,985
18532	9862	-7332,500	-862,643	6325328,75	53765556,250	744152,699

Q_{xy} =28807883,500; Q_x =210109355,500; Q_y = 4452883,214.

Iš čia $r = \frac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}} = \frac{28807883{,}500}{\sqrt{210109355{,}500 \cdot 4452883{,}214}} = 0{,}942$

Išvada. Galima teigti, kad žmonių protinis atsilikimas ir psichologinės raidos sutrikimai priklauso nuo elgesio ir emocijų sutrikimų prasidedančių vaikystėje ir paauglystėje.

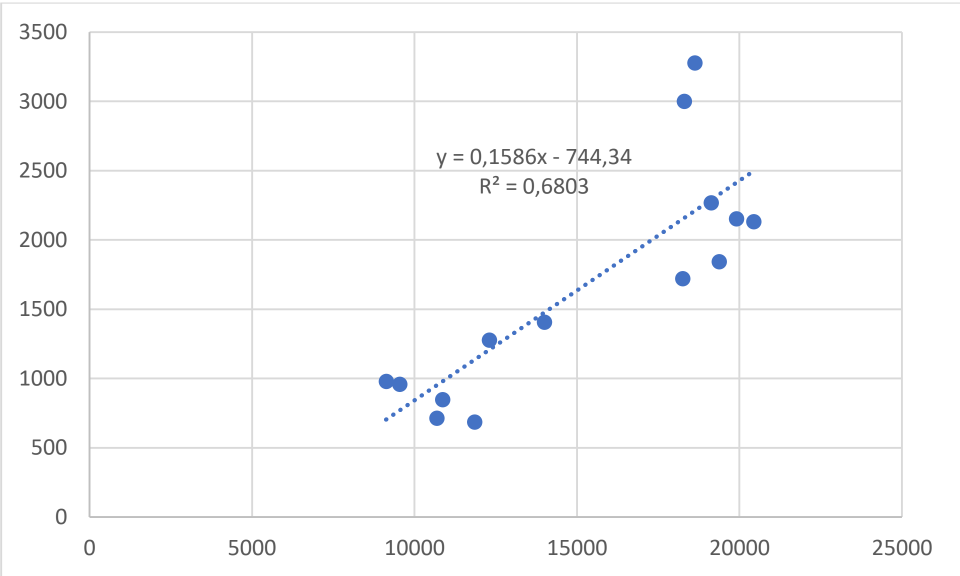
Kadangi empiriniai koreliacijos koeficientai gauti gana dideli, užrašysime tiesinės regresijos lygtis.

Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo (2)-Psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo (3)

Tiesinės regresijos lygtis

$$y = -744{,}34 + 0{,}1586 \cdot x + \varepsilon.$$

Tiesinės regresijos lygtis užrašyta programos Excel pagalba, iš šios lygties matome, kad koeficientas prie x (sergančiųjų psichikos ir elgesio sutrikimais dėl alkoholio vartojimo) yra $0{,}1586$ – kiekvienas sergantysis psichikos ir elgesio sutrikimais dėl alkoholio vartojimo tiek padidina sergančiųjų psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų skaičių pridėjus konstantą.



13 pav. *(2) ir (3) ligų grupių SPES duomenų sklaidos diagrama*

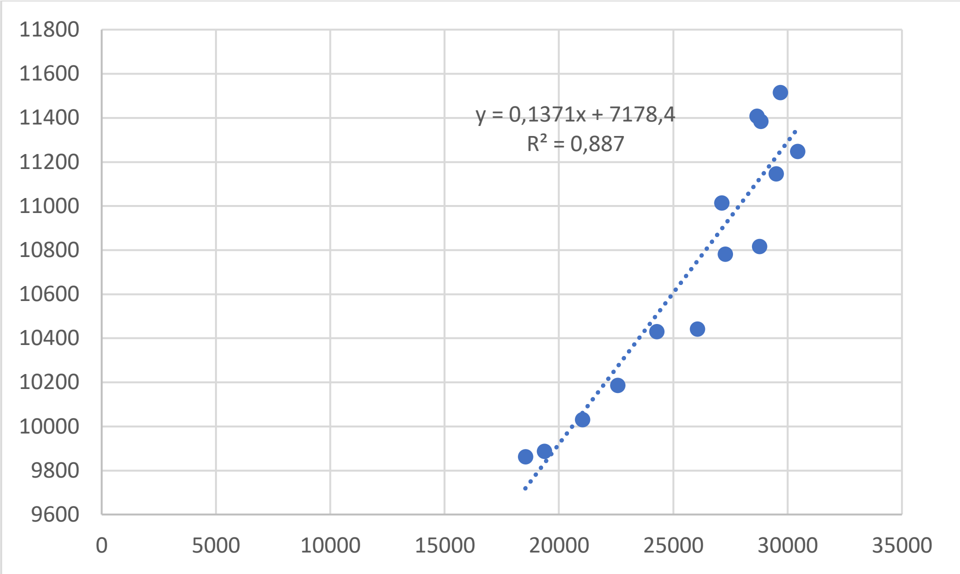
Pagal duomenų sklaidos diagramą matome, jog duomenys gana stipriai išsibarstę aplink tiesę, o koreliacijos koeficientas lygus $0{,}825$ – vidutinio stiprumo koreliacija.

Protinis atsilikimas ir psichologinės raidos sutrikimai (6)-Elgesio ir emocijų sutrikimai prasidedantys vaikystėje ir paauglystėje (7)

Tiesinės regresijos lygtis

$$y = 7178{,}4 + 0{,}1371 \cdot x + \varepsilon.$$

Tiesinės regresijos lygtis užrašyta programos Excel pagalba, iš šios lygties matome, kad koeficientas prie x (sergančiųjų protiniu atsilikimu ir psichologinės raidos sutrikimais) yra $0{,}1371$ – kiekvienas sergantysis protiniu atsilikimu ir psichologinės raidos sutrikimais tiek padidina sergančiųjų elgesio ir emocijų sutrikimais prasidedančiais vaikystėje ir paauglystėje skaičių pridėjus konstantą.



14 pav. *(6) ir (7) ligų grupių SPES duomenų sklaidos diagrama*

Pagal duomenų sklaidos diagramą matome, jog duomenys išsidėstę gana arti regresijos tiesės. Kadangi koreliacijos koeficientas gautas pakankamai didelis ($r = 0{,}942$ – stipri teigiama koreliacija), sklaidos diagramoje lengva įžvelgti ryšius tarp dviejų kintamųjų.

2.3.4 DUOMENŲ NEPRIKLAUSOMUMAS IR ATSITIKTINUMAS

Pagal TLK-10-AM ligų klasifikatorių Lietuvos gyventojų sergamumas psichikos ir elgesio sutrikimais bei ligomis yra pateiktas kiekvienais metais nuo 2006 m. iki 2019 m., tai leidžia ištirti duomenų pokyčius laike ir taikyti laiko eilučių teoriją. Jei duomenys išsidėstę laiko skalėje yra atsitiktiniai ir nepriklausomi, jiems galima taikyti klasikinės statistikos metodus, t. y. braižome histogramas, tikriname hipotezes apie galimą duomenų skirstinį ir t.t. Jei duomenys yra priklausomi, jiems galime taikyti laiko eilučių metodus, todėl tikslinga patikrinti statistinę hipotezę apie duomenų atsitiktinumą ir nepriklausomumą, taikant Serijų kriterijų.

Matematiškai patikrinsime, ar sergančiųjų psichikos ir elgesio sutrikimais Lietuvoje duomenims galime taikyti laiko eilučių teoriją.

Tikriname hipotezę H_0 : duomenys nepriklausomi ir atsitiktiniai; kai $\alpha = 0{,}05$; $\alpha = 0{,}1$.

Duomenų eilutes imame iš 4 priedo:

24 lentelė. *Sergančiųjų demencija ir Alzheimerio liga duomenys*

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	11340	13915	17068	19609	21233	25156	28655
Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	29184	30806	31559	33170	33056	32362	32617

Užrašome duomenis variacine eile:

11340, 13915, 17068, 19609, 21233, 25156, 28655, 29184, 30806, 31559, 33170, 33056, 32362, 32617.

$$n = 14.$$

Surandame medianą : $M_e = \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right) = 28919{,}5$

Lyginame

$$(-------)(+++++++)$$

N = N(serijų skaičius) = 2.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

$N_{0{,}025}(A, 7{,}7) = 3.$

$N_{0{,}025}(V, 7{,}7) = 13.$

Jeigu $N_{0{,}025}(A, 7{,}7) < N < N_{0{,}025}(V, 7{,}7)$, tada H_0 teisinga. Kadangi 3 $\nless$ 2 < 13, H_0 atmetame. Tai reiškia, kad duomenys yra priklausomi ir neatsitiktiniai, todėl laiko eilučių teorija tinka.

2) Kai $\alpha = 0{,}1$

$N_{0{,}05}(A, 7{,}7) = 4.$

$N_{0{,}05}(V, 7{,}7) = 12.$

Kadangi 4 $\nless$ 2 <12 H_0 atmetame. Tai reiškia, kad duomenys yra priklausomi ir neatsitiktiniai, todėl laiko eilučių teorija tinka.

Toliau tą patį procesą analogiškai taikome kiekvienai sergančiųjų duomenų eilutei (5 priedas).

Duomenims, kuriems hipotezė apie duomenų nepriklausomumą ir atsitiktinumą atmetama, priklausomybę galėsime užrašyti laiko eilučių proceso pagalba. Priešingu atveju būtų galima taikyti tik klasikinės matematinės statistikos metodus. Norint lengviau pastebėti rezultatus ir žinoti kurių ligų-sutrikimų duomenų eilutėms galėsime taikyti laiko eilučių modelius, rezultatus surašome į lentelę:

+ – duomenys atsitiktiniai ir nepriklausomi;

— – duomenys yra neatsitiktiniai ir priklausomi.

Gauname tokius rezultatus:

25 lentelė. *Hipotezės apie duomenų nepriklausomumą ir atsitiktinumą tikrinimo rezultatai*

H_0	(1)	(2)	(3)	(4)	(5)	(6)	(7)	∑
α=0,05	—	—	—	+	+	—	+	—
α=0,1	—	—	—	+	—	—	—	—

Išvada. Patikrinus statistinę hipotezę apie duomenų nepriklausomumą ir atsitiktinumą, su skirtingomis α reikšmėmis, galima pastebėti, kad jei reikšmingumo lygmuo mažas (0,05), penkioms duomenų eilutėms galėsime taikyti laiko eilučių metodus. Tačiau jei reikšmingumo lygmenį padidiname iki $\alpha = 0{,}1$, laiko eilučių modelius galime taikyti beveik visoms – 7 iš 8 duomenų eilutėms. Lietuvos žmonių sergamumo psichikos ir elgesio sutrikimais duomenys pagal visas TLK-10-AM ligų klasifikatoriaus F grupės ligas per visą 2006-2019 metų laikotarpį yra priklausomi ir neatsitiktiniai, išskyrus: Šizofrenija, šizotipinis ir kliedesiniai sutrikimai (4).

Nepriklausomi ir atsitiktiniai duomenys gali būti tiriami klasikiniais matematinės statistikos metodais, tad duomenims apie sergančiuosius šizofrenija, šizotipiniu ir kliedesiniais sutrikimais surasime vidurkį, dispersiją, standartinį nuokrypį, medianą, nubraižysime histogramą bei patikrinsime, ar duomenys yra pasiskirstę pagal normalųjį skirstinį.

Turime 4-os ligų grupės duomenų eilutę: 17042, 17167, 17479, 17900, 18096, 18648, 18877, 19115, 18872, 18591, 18969, 18765, 18588, 18603.

Apskaičiuosime šių duomenų empirines charakteristikas (vidurkį, dispersiją, standartinį nuokrypį, medianą).

Pirmiausia sudarome variacinę duomenų eilutę:

17042, 17167, 17479, 17900, 18096, 18588, 18591, 18603, 18648, 18765, 18872, 18877, 18969, 19115

Apskaičiuojame empirines charakteristikas:

Vidurkis:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{14}\sum_{i=1}^{14} X_i = \frac{1}{14}(17042 + 17167 + \cdots + 19115) = 18336{,}57.$$

Dispersija:

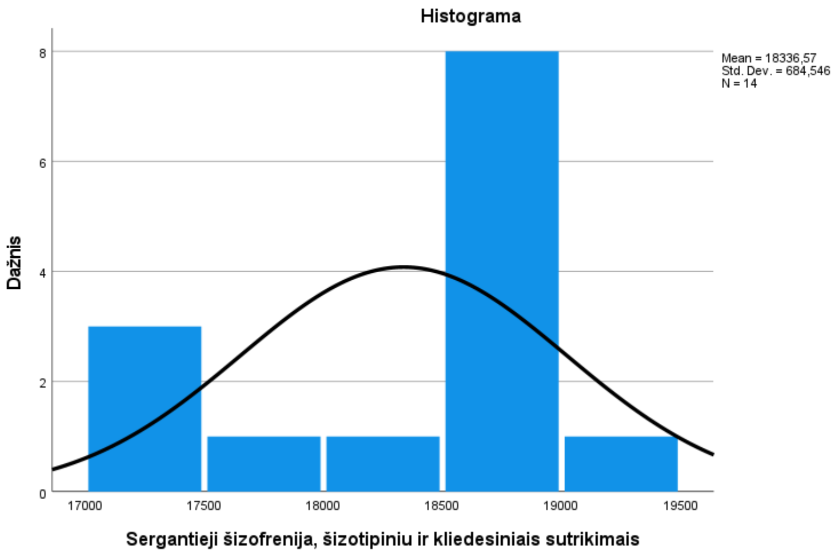
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{14-1}\sum_{i=1}^{14}(X_i - \bar{X})^2 = \frac{1}{14-1}((17042 - 18336{,}57)^2 + (17167 - 18336{,}57)^2 + \cdots + (19115 - 18336{,}57)^2) = 468603{,}648$

Standartinis nuokrypis: $S = \sqrt{S^2} = \sqrt{468603{,}648} = 684{,}546$

Mediana:

$M_e = \frac{1}{2}\left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right) = \frac{1}{2}\left(X_{(\frac{14}{2})} + X_{(\frac{14}{2}+1)}\right) = \frac{1}{2}\left(X_{(7)} + X_{(8)}\right) = \frac{37197}{2} = 18597.$

Programos SPSS pagalba nubraižysime histogramą:



15 pav. *Sergančiųjų šizofrenija, šizotipiniu ir kliedesiniais sutrikimais histograma*

Duomenų normalumo tikrinimui galime nubraižyti histogramą su normaliąja kreive, tačiau vien tik iš histogramos duomenų daryti išvadų apie duomenų normalumą negalime, todėl būtina panaudoti duomenų normalumo testus (Kolmogorov-Smirnov ir Shapiro-Wilk testus). Pasinaudodami SPSS statistikos paketu patikrinsime, ar sergančiųjų šizofrenija, šizotipiniu ir kliedesiniais sutrikimais duomenys yra pasiskirstę pagal normalųjį skirstinį 2006-2019 m. laikotarpiu. Gauname tokius rezultatus:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
	,286	14	,003	,863	14	,034

a. Lilliefors Significance Correction

16 pav. *„Tests of Normality" rezultatų lentelė*

Paprastai, kai yra mažai duomenų (mažiau nei 50), rekomenduojama vadovautis Shapiro-Wilk testo rezultatais. Kadangi lentelėje „Tests of Normality" Shapiro-Wilk testo

„Sig.“ Reikšmė < 0,05, galime teigti, kad sergančiųjų šizofrenija, šizotipiniu ir kliedesiniais sutrikimais gyventojų duomenys nėra pasiskirstę pagal normalųjį skirstinį.

2.4 KOMPLEKSIŠKAS LAIKO EILUČIŲ TAIKYMAS

Turimi duomenys apie sergančiųjų psichikos ir elgesio sutrikimais žmonių skaičius Lietuvoje yra pateikti kasmečiui, 2006-2019 m. laikotarpiu. Kadangi duomenys kasmet keičiasi, t.y. kinta laike, mus domina, ar duomenys priklauso nuo praeitų ir užpraeitų metų duomenų. Pasinaudodami 1.6 poskyrio laiko eilučių teorija, rasime AR(1) ir AR(2) modelius.

2.4.1 AR(1) MODELIO PARAMETRŲ RADIMAS

Surasime Laiko eilučių AR(1) modelius mažiausių kvadratų metodu 1,2,3,5,6,7 grupių ir bendro sergančiųjų skaičiaus duomenims. Taip pat apskaičiuosime σ^2 įverčius.

Pavyzdys. Kaip pavyzdį pateiksiu skaičiavimus su sergančiųjų **Psichikos ir elgesio sutrikimais dėl alkoholio vartojimo (2)** skaičiais (duomenys paimti iš 4 priedo):

26 lentelė. *Sergančiųjų psichikos ir elgesio sutrikimais dėl alkoholio vartojimo duomenys*

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	10686	11847	10865	9129	9542	12306	14001

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	18259	19373	19908	20445	19127	18308	18632

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 10686$, $X_2 = 11847$, $X_3 = 10865$, $X_4 = 9129$, $X_5 = 9542$, $X_6 = 12306$, $X_7 = 14001$, $X_8 = 18259$, $X_9 = 19373$, $X_{10} = 19908$, $X_{11} = 20445$, $X_{12} = 19127$, $X_{13} = 18308$, $X_{14} = 18632$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t (X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{11847 \cdot 10686 + 10865 \cdot 11847 + \cdots + 18308 \cdot 19127 + 18632 \cdot 18308}{10686^2 + 11847^2 + 10865^2 + \cdots + 18308^2 + 18632^2} = 1{,}031$$

σ^2 įvertis:

$$\widehat{\sigma^2} = \frac{S(\hat{a})}{n}$$

Čia, $S(\hat{a}_1) = \sum_{t=2}^{n}(x_t - \hat{a}_1 \cdot x_{t-1})^2 + \left(1 - \hat{a}_1{}^2\right) \cdot x_1^2 = (x_2 - \hat{a}_1 x_1)^2 + \left(x_3 - \widehat{a_1} x_2\right)^2 + \cdots + (x_{19} - \hat{a}_1 x_{18})^2 + (1 - \hat{a}_1{}^2) \cdot x_1^2 = 28202942{,}82$

$$\widehat{\sigma^2} = \frac{S(\hat{a}_1)}{n} = \frac{28202942{,}82}{14} = 1828901{,}29$$

Atlikus tokius veiksmus su duomenimis, kuriems galima taikyti laiko eilučių metodus, gauti tokie rezultatai:

27 lentelė. *AR(1) modelio parametrai*

Duomenų grupė	$\hat{a}_1$	σ^2	**AR(1) modelis**
1	1,049	1828901,29	$X_t = 1{,}049 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
2	1,031	2014495,92	$X_t = 1{,}031 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
3	1,125	20828,86	$X_t = 1{,}125 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
5	1,013	-1071290,35	$X_t = 1{,}013 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
6	1,033	-1102650,65	$X_t = 1{,}033 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
7	1,009	-6365,31	$X_t = 1{,}009 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$
$\sum$	1,023	-85334870,09	$X_t = 1{,}023 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$

Išvada. AR(1) modelio parametras $\hat{a}_1$ rodo priklausomybę tam tikrų metų, nuo prieš tai buvusių, tai gali būti bet kurie metai iš mūsų analizuojamo laikotarpio – nuo 2006 m. iki 2019 m. Iš šios lentelės galima pastebėti, kad šis parametras svyruoja nuo 1,009 iki 1,125; tai rodo didelę priklausomybę tarp duomenų lyginant su prieš tai (prieš metus) buvusiais duomenimis. Jei parametras $\hat{a}$ būtų lygus vienetui, tuomet šių metų duomenys būtų lygūs praeitų metų duomenims – duomenys būtų identiški. Nei vienoje ligų grupėje parametras $\hat{a}$ nėra mažesnis už vienetą, tai rodo, jog sergančiųjų psichikos ir elgesio sutrikimais skaičius metai iš metų po truputį didėja.

Visų AR(1) modelių radimą žr. 6 priede.

2.4.2 AR(2) MODELIO PARAMETRŲ RADIMAS

Surasime Laiko eilučių AR(2) modelio $\hat{a}_1$ ir $\hat{a}_2$ įverčius mažiausių kvadratų metodu 1,2,3,5,6,7 grupių ir bendro sergančiųjų skaičiaus duomenims.

Pavyzdys. Kaip pavyzdį pateiksiu skaičiavimus su sergančiųjų **Psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo (3)** skaičiais (duomenys paimti iš 4 priedo).

28 lentelė. *Sergančiųjų psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo duomenys*

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	714	686	847	979	958	1278	1408

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	1722	1844	2153	2132	2269	3001	3278

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$$X_1 = 714,\ X_2 = 686,\ X_3 = 847,\ X_4 = 979,\ X_5 = 958,\ X_6 = 1278,\ X_7 = 1408,\ X_8 = 1722,\ X_9 = 1844, X_{10} = 2153,\ X_{11} = 2132,\ X_{12} = 2269,\ X_{13} = 3001,\ X_{14} = 3278$$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 0{,}862$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 0{,}306$$

$$X_t = 0{,}862 \cdot X_{t-1} + 0{,}306 \cdot X_{t-2} + \varepsilon_t$$

29 lentelė. AR(2) modelio parametrai

Duomenų grupė	$\hat{a}_1$	$\hat{a}_2$	**AR(2) modelis**
1	1,331	-0,303	$X_t = 1{,}331 \cdot X_{t-1} + (-0{,}303) \cdot X_{t-2} + \varepsilon_t$
2	1,193	-0,175	$X_t = 1{,}193 \cdot X_{t-1} + (-0{,}175) \cdot X_{t-2} + \varepsilon_t$
3	0,862	0,306	$X_t = 0{,}862 \cdot X_{t-1} + 0{,}306 \cdot X_{t-2} + \varepsilon_t$
5	1,072	-0,063	$X_t = 1{,}072 \cdot X_{t-1} + (-0{,}063) \cdot X_{t-2} + \varepsilon_t$
6	1,054	-0,022	$X_t = 1{,}054 \cdot X_{t-1} + (-0{,}022) \cdot X_{t-2} + \varepsilon_t$
7	0,992	0,019	$X_t = 0{,}992 \cdot X_{t-1} + 0{,}019 \cdot X_{t-2} + \varepsilon_t$
$\sum$	1,023	-0,00002	$X_t = 1{,}023 \cdot X_{t-1} + (-0{,}00002) \cdot X_{t-2} + \varepsilon_t$

Išvada. AR(2) modeliuose galima pastebėti, kaip šių metų sergančiųjų duomenys priklauso nuo prieš tai buvusių ir užpernykščių metų sergančiųjų duomenų. Pagal parametą $\hat{a}_1$, kuris svyruoja tarp 0,862 ir 1,331, matoma didelio stiprumo priklausomybė tarp šių metų duomenų ir pernykščių metų duomenų visais atvejais. Lyginant metus su užpernai buvusiais, priklausomybės nėra beveik visose grupėse (kadangi parametras $\hat{a}_2$ visais atvejais mažesnis už nulį, svyruoja nuo -0,303 iki -0,00002), išskyrus 3 ir 7 grupes, kuriose parametras $\hat{a}_2$ yra teigiamas (atitinkamai 0,306 ir 0,019).

Visų AR(2) modelių radimą žr. 7 priede.

2.4.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS

Pagal AR(1) proceso parametrą $\hat{a}_1$, kuris visais atvejais yra didesnis už vienetą, darome išvadą, kad šie procesai nėra asimptotiškai stacionarūs, yra priklausomi nuo laiko, vadinasi, šie procesai turėtų turėti tam tikrą trendą. Nustačius, kad duomenys turi trendą, jį eliminuosime ir rasime AR(1) proceso lygtį.

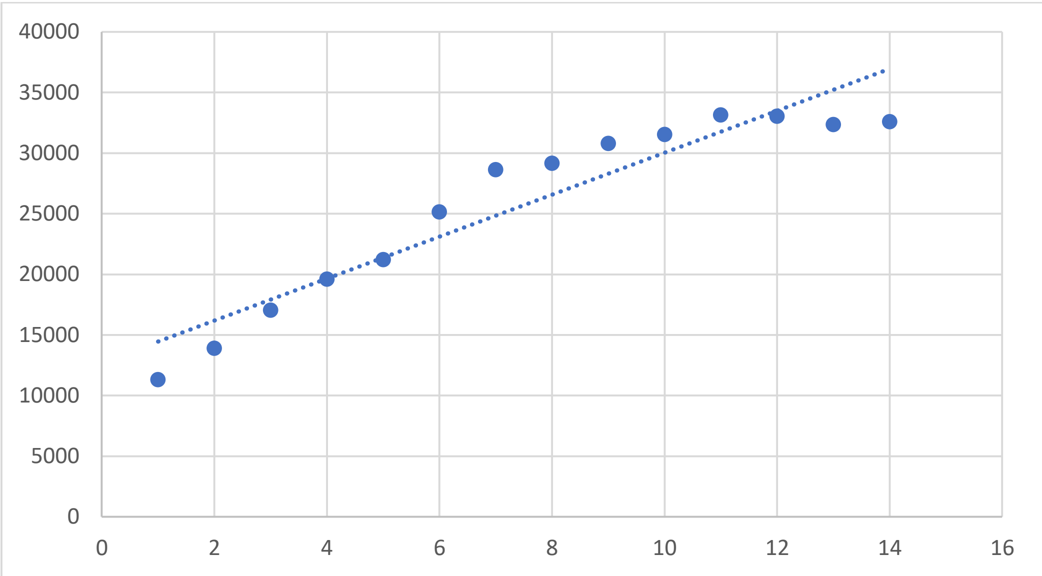
Pavyzdys. Kaip pavyzdį pateiksiu skaičiavimus su sergančiųjų **demencija ir Alzheimerio liga (1)** duomenimis (duomenys paimti iš 4 priedo):

__30 lentelė.__ Sergančiųjų demencija ir Alzheimerio liga duomenys

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	11340	13915	17068	19609	21233	25156	28655

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	29184	30806	31559	33170	33056	32362	32617

Pirmiausia pavaizduosime kaip atrodo sergančiųjų **demencija ir Alzheimerio liga (1)** Lietuvos gyventojų grafikas:



17 pav. *Sergančiųjų demencija ir Alzheimerio liga Lietuvos gyventojų grafikas*

Matome, kad duomenys yra nestipriai išsibarstę aplink tiesę, todėl patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+” ir „-“ eilutes bei radome medianos reikšmę, kur

$$M_e = 28919{,}5$$

$$n = 14$$

$$\gamma(n) = 2 - bendras\ serijų\ skaičius$$

$$\tau(n) = 7 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: \boldsymbol{E}X_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$

$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$2 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14-1})\right] \Rightarrow 2 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

2.3.4 skyrelyje sužinojome, kad laiko eilučių teoriją taip pat galima taikyti 2,3,5,6,7 ir bendro SPES skaičiaus duomenims. Patikrinome, ar šie duomenys turi trendą, gautus rezultatus surašėme į lentelę:

31 lentelė. *Hipotezės apie trendo egzistavimą tikrinimo rezultatai*

Duomenų grupė	$\boldsymbol{H_0: MX_t = a = const}$ **priimta / atmesta**
1	-
2	-
3	-
5	-
6	-
7	-
$SPES$	-

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Pavyzdys. Sergančiųjų **demencija ir Alzheimerio liga (1)** gyventojų duomenys 17 pav. atrodo nestipriai išsibarstę apie tiesę, todėl galime daryti prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1}t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 1730{,}7 \cdot t + 12715$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (1730{,}7 \cdot t + 12715)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

__32 lentelė.__ Duomenys, gauti eliminavus trendą

Metai	X_t	t	$1730{,}7 \cdot t + 12715$	$\boldsymbol{Y_t}$
2006	11340	1	14445,700	**-3105,700**
2007	13915	2	16176,400	**-2261,400**
2008	17068	3	17907,100	**-839,100**
2009	19609	4	19637,800	**-28,800**
2010	21233	5	21368,500	**-135,500**
2011	25156	6	23099,200	**2056,800**
2012	28655	7	24829,900	**3825,100**
2013	29184	8	26560,600	**2623,400**
2014	30806	9	28291,300	**2514,700**
2015	31559	10	30022,000	**1537,000**
2016	33170	11	31752,700	**1417,300**
2017	33056	12	33483,400	**-427,400**
2018	32362	13	35214,100	**-2852,100**
2019	32617	14	36944,800	**-4327,800**

Surasime AR(1) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t$$

ir stebinius:

$$Y_1 = -3105{,}700;\ Y_2 = -2261{,}400;\ Y_3 = -839{,}100;\ Y_4 = -28{,}800;$$
$$Y_5 = -135{,}500; Y_6 = 2056{,}800;\ Y_7 = 3825{,}100; Y_8 = 2623{,}400;\ Y_9 = 2514{,}700;$$
$$Y_{10} = 1537{,}000;\ Y_{11} = 1417{,}300;\ Y_{12} = -427{,}400;\ Y_{13} = -2852{,}100;$$
$$Y_{14} = -4327{,}800.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = -0{,}250.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 5640603{,}305 \\ \hat{\gamma}(1) = 3726523{,}796 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 5640603{,}305 = \hat{a}_1 \cdot 3726523{,}796 + \sigma^2 \\ 3726523{,}796 = \hat{a}_1 \cdot 5640603{,}305 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}661;\ \hat{\sigma}^2 = 3178636{,}234.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}661 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1782{,}873^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 586{,}707 \cdot t + 4310{,}385 + 0{,}661 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1782{,}873^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Žemiau pateikiame tinkamų duomenų grupių trendo lygtis:

33 lentelė. *Trendo lygtys visoms duomenų eilutėms*

Duomenų grupė	**Trendo lygtis**
1	$f_t = 1730{,}7 \cdot t + 12715$
2	$f_t = 890{,}61 \cdot t - 1777181{,}19$
3	$f_t = 192{,}64 \cdot t - 386034$
5	$f_t = 539{,}35 \cdot t - 1040412{,}76$
6	$f_t = 919{,}37 \cdot t - 1824374{,}92$
7	$f_t = 125{,}36 \cdot t - 241572$
SPES	$f_t = 4397{,}39 \cdot t - 8645071{,}75$

Eliminavę trendą gauname AR(1) modelio lygtis:

34 lentelė. *AR(1) proceso lygtys visoms duomenų eilutėms*

Duomenų grupė	**AR(1) proceso lygtis**
1	$Y_t = 0{,}661 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1782{,}873^2)$
2	$Y_t = 0{,}663 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1574{,}694^2)$
3	$Y_t = 0{,}301 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 183{,}441^2)$
5	$Y_t = 0{,}649 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1542{,}719^2)$
6	$Y_t = 0{,}731 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 769{,}993^2)$
7	$Y_t = -0{,}015 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 250{,}317^2)$
SPES	$Y_t = 0{,}385 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 3329{,}914^2)$

Apjungiame AR(1) proceso lygtis su trendo lygtimis ir gauname tokius rezultatus:

35 lentelė. *AR(1) proceso, eliminavus trendą, lygtys visoms duomenų eilutėms*

Duomenų grupė	**AR(1) proceso, eliminavus trendą, lygtis**
1	$X_t = 586{,}707 \cdot t + 4310{,}385 + 0{,}661 \cdot X_{t-1} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 1782{,}873^2)$

2	$X_t = 300{,}136 \cdot t - 598910{,}061 + 0{,}663 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 1574{,}694^2)$
3	$X_t = 134{,}655 \cdot t - 269837{,}766 + 0{,}301 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 183{,}441^2)$
5	$X_t = 189{,}312 \cdot t - 365184{,}879 + 0{,}649 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 1542{,}719^2)$
6	$X_t = 247{,}311 \cdot t - 490756{,}853 + 0{,}731 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 769{,}993^2)$
7	$X_t = 127{,}240 \cdot t - 245195{,}580 - 0{,}015 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 250{,}317^2)$
SPES	$X_t = 2704{,}395 \cdot t - 5316719{,}126 + 0{,}385 \cdot X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0; 3329{,}914^2)$

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = -3105{,}700;\ Y_2 = -2261{,}400;\ Y_3 = -839{,}100;\ Y_4 = -28{,}800;$$
$$Y_5 = -135{,}500; Y_6 = 2056{,}800;\ Y_7 = 3825{,}100; Y_8 = 2623{,}400;\ Y_9 = 2514{,}700;$$
$$Y_{10} = 1537{,}000;\ Y_{11} = 1417{,}300;\ Y_{12} = -427{,}400;\ Y_{13} = -2852{,}100;$$
$$Y_{14} = -4327{,}800.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = -0{,}250$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 5640603{,}305 \\ \hat{\gamma}(1) = 3726523{,}796 \\ \hat{\gamma}(2) = 2548115{,}342 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 5640603{,}305 = \hat{a}_1 \cdot 3726523{,}796 + \hat{a}_2 \cdot 25481115{,}342 + \sigma^2 \\ 3726523{,}796 = \hat{a}_1 \cdot 5640603{,}305 + \hat{a}_2 \cdot 3726523{,}796 \\ 2548115{,}342 = \hat{a}_1 \cdot 3726523{,}796 + \hat{a}_2 \cdot 5640603{,}305 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}643; \hat{a}_2 = 0{,}027; \hat{\sigma}^2 = 3176301{,}420.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}643 \cdot Y_{t-1} + 0{,}027 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1782{,}218^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 571{,}378 \cdot t + 4197{,}764 + 0{,}643 \cdot X_{t-1} + 0{,}027 \cdot X_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1782{,}218).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Žemiau pateikiame tinkamų duomenų grupių AR(2) modelio lygtis:

36 lentelė. *AR(2) proceso lygtys visoms duomenų eilutėms*

Duomenų grupė	**AR(2) proceso lygtis**
1	$Y_t = 0{,}643 \cdot Y_{t-1} + 0{,}027 \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 1782{,}218^2)$
2	$Y_t = 0{,}642 \cdot Y_{t-1} + 0{,}032 \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 1573{,}883^2)$
3	$Y_t = 0{,}244 \cdot Y_{t-1} + 0{,}192 \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 180{,}029^2)$
5	$Y_t = 0{,}691 \cdot Y_{t-1} + (-0{,}064) \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 1539{,}512^2)$
6	$Y_t = 0{,}779 \cdot Y_{t-1} + (-0{,}066) \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 768{,}298^2)$
7	$Y_t = (-0{,}015) \cdot Y_{t-1} + (-0{,}014) \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 250{,}293^2)$
SPES	$Y_t = 0{,}367 \cdot Y_{t-1} + 0{,}045 \cdot Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0; 3326{,}495^2)$

Apjungiame AR(2) proceso lygtis su trendo lygtimis ir gauname tokius rezultatus:

37 lentelė. *AR(2) proceso, eliminavus trendą, lygtys visoms duomenų eilutėms*

Duomenų grupė	**AR(2) proceso, eliminavus trendą, lygtis**
1	$X_t = 571{,}378 \cdot t + 4197{,}764 + 0{,}643 \cdot X_{t-1} + 0{,}027 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 1782{,}218^2)$
2	$X_t = 290{,}524 \cdot t - 579730{,}559 + 0{,}642 \cdot X_{t-1} + 0{,}032 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 1573{,}883^2)$
3	$X_t = 108{,}731 \cdot t - 217888{,}422 + 0{,}244 \cdot X_{t-1} + 0{,}192 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 180{,}029^2)$
5	$X_t = 201{,}465 \cdot t - 388627{,}750 + 0{,}691 \cdot X_{t-1} - 0{,}064 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 1539{,}512^2)$

6	$X_t = 263{,}899 \cdot t - 523673{,}699 + 0{,}779 \cdot X_{t-1} - 0{,}066 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 768{,}298^2)$
7	$X_t = 129{,}002 \cdot t - 248589{,}995 - 0{,}015 \cdot X_{t-1} - 0{,}014 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 250{,}293^2)$
$\boldsymbol{SPES}$	$X_t = 2583{,}671 \cdot t - 5079382{,}118 + 0{,}367 \cdot X_{t-1} + 0{,}045 \cdot X_{t-2} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 3326{,}495^2)$

Laiko eilučių išlyginimo metodo taikymą visoms duomenų eilutėms žr. 8 priede.

IŠVADOS

1. Nagrinėjami sergančiųjų psichikos ir elgesio sutrikimais (SPES) skaičiai koreliuoja su Lietuvos gyventojų skaičiais. 2006 – 2019 m. laikotarpiu sergančiųjų skaičiai stipriai priklauso nuo Lietuvos gyventojų skaičiaus (priimama hipotezė $H_0: \rho = -0{,}9$, su tikimybe $0{,}95$). Ši priklausomybė yra tiesinė ir gali būti užrašyta regresijos lygtimi:

$$y = 437539 - 0{,}0757 \cdot x + \varepsilon, \varepsilon \sim N(0;\ 1444406{,}189),$$

čia y – SPES skaičius, x – gyventojų skaičius.

2019 metų tos lygties prognozės intervalas yra $y \in [212108{,}82;\ 238259{,}94]$. Atkreiptinas dėmesys, kad SPES skaičiaus vidurkis lygus 204669,214, o SPES skaičius 2019 metais lygus 235862.

2. Pritaikius statistinę hipotezę apie polinominio skirstinio parametro reikšmes, nustatyta, kad nagrinėjamų 7 ligų grupių SPES žmonių skaičiai statistiškai reikšmingai pasikeičia tik kas ketverius metus ir tas pokytis dar neišryškėja praėjus vieneriems.

3. Pavaizdavus SPES pokyčių dinamiką spalvų kitimo lentele kasmet ir kas ketverius metus, nustatyta, kad kasmetiniuose pokyčiuose spalvos dažnai keičiasi, tačiau esminės tendencijos geriau pastebimos kas ketverius metus, o vienerių metų pokyčiai nebuvo statistiškai reikšmingi.

4. Taikant serijų kriterijų apie duomenų atsitiktinumą ir nepriklausomumą buvo nustatyta, kad duomenys yra nepriklausomi ir atsitiktiniai tik 4 grupėje (sergančiųjų šizofrenija, šizotipiniu ir kliedesiniais sutrikimais). Visoms kitoms duomenų grupėms (1,2,3,5,6,7) ir bendram SPES skaičiui hipotezė buvo atmesta, vadinasi, joms yra taikytini laiko eilučių metodai.

5. Taikant laiko eilučių išlyginimo metodą, nustatyta, kad visos duomenų eilutės, kurioms yra taikytini laiko eilučių metodai, turi trendą.

6. Pritaikius laiko eilučių metodus tinkamoms duomenų grupėms, buvo rasti AR(1) ir AR(2) modelių parametrų įverčiai, kai iš duomenų eilučių nėra eliminuojamas trendas ir kai trendas yra eliminuojamas.

LITERATŪROS SĄRAŠAS

1. Australijos klasifikacijos kūrimo konsorciumas, *TLK-10-AM Sisteminis ligų sąrašas.* Australija, 2015. Prieiga internetu: http://ebook.vlk.lt/e.vadovas/index.jsp?topic=/lt.webmedia.vlk.drg.icd.ebook.content/html/icd/ivadas.html [žiūrėta 2022-04-26]
2. V. Boguslauskas, *Ekonometrika*. Kaunas, 2008.
3. V. Čekanavičius, *Taikomoji regresinė analizė socialiniuose tyrimuose.* Vilniaus universiteto leidykla, Vilnius, 2014. Prieiga internetu: http://www.statistika.mif.vu.lt/wp-content/uploads/2014/04/regresine-analize.pdf [žiūrėta 2022-04-26]
4. V. Čekanavicius, G. Murauskas, *Statistika ir jos taikymai.* II dalis. Vilnius, 2008.
5. Higienos instituto Sveikatos informacijos centras. *Lietuvos gyventojų sveikata ir sveikatos priežiūros įstaigų veikla.* Vilnius, 2000–2018 m. Prieiga internetu: https://hi.lt/lt/lietuvos-gyventoju-sveikata-ir-sveikatos-prieziuros-istaigu-veikla-2013-m.html [žiūrėta 2022–04–26]
6. V. Kanišauskas. *Tikimybių teorijos ir matematinės statistikos pagrindai.* Šiauliai: Šiaulių universiteto leidykla, 2000.
7. V. Kanišauskas. *Vidinės migracijos matematinis modelis*. Lietuvos matematikos rinkinys, 2016. 57: 7–12.
8. J. Kruopis, *Matematinė statistika.* Vilnius: Mokslo ir enciklopedijų leidykla, 1993.
9. R. Leipus. *Laiko eilučių teorijos įvadas*, mokomoji knyga. Vilniaus universitetas, Vilnius, 1995.
10. R. Leipus, *Finansinės laiko eilutės.* Vilnius, 2010. Prieiga internetu: http://web.vu.lt/mif/a.reklaite/files/2012/09/fin_laik_eil_2.pdf [žiūrėta 2022-04-26]
11. T. Leonavičienė, *SPSS programų paketo taikymas statistiniuose tyrimuose.* Vilnius, 2007.

LIETUVOS GYVENTOJŲ SERGAMUMO PSICHIKOS IR ELGESIO SUTRIKIMAIS XXI A. PRADŽIOJE STATISTINĖ ANALIZĖ

SANTRAUKA

Magistro darbo objektas yra statistiniai ir demografiniai duomenys iš Higienos instituto Sveikatos informacijos centro išleistų periodinių sveikatos statistikos leidinių „Lietuvos gyventojų sveikata ir sveikatos priežiūros įstaigų veikla“ apie sergančiuosius psichikos ir elgesio sutrikimais žmonių skaičius Lietuvoje 2006 – 2019 metais. Darbo tikslas – statistiniais metodais ištirti sergančiųjų psichikos ir elgesio sutrikimais skaičiaus dinamiką Lietuvoje nagrinėjamu laikotarpiu. Tikslui pasiekti keliami uždaviniai: pirmiausia, ištirti Lietuvos sergančiųjų psichikos ir elgesio sutrikimais skaičiaus priklausomybę nuo gyventojų skaičiaus panaudojant koreliacinę ir regresinę analizę, tuomet taikant statistines hipotezes apie polinominio skirstinio parametrų reikšmes, nustatyti, ar nagrinėjamų 7 ligų-grupių sergančiųjų skaičiai kardinaliai pasikeičia praėjus tam tikram laikotarpiui, jei taip, nurodyti po kokio laikotarpio (po kelerių metų). Tęsiant tyrimą pagal sutrikimus, ištirti sergančiųjų skaičiaus pokyčių dinamiką kasmet ir kas ketverius metus, tai aprašyti ir pavaizduoti vizualiai, tuomet taikant serijų kriterijų apie duomenų atsitiktinumą ir nepriklausomumą, nustatyti, kuriems sergančiųjų psichikos ir elgesio sutrikimais duomenims galima taikyti laiko eilučių metodus. Tyrimo pabaigoje taikant laiko eilučių išlyginimo metodą, nustatyti, kurios duomenų eilutės turi trendą, o kurios ne ir taikant laiko eilučių metodus rasti AR(1) ir AR(2) modelių parametrų įverčius panaudojant trendą ir jo nepanaudojant.

Atlikus sergančiųjų psichikos ir elgesio sutrikimais gyventojų ir Lietuvos gyventojų duomenų analizę, buvo nustatyta, kad SPES skaičiai koreliuoja su Lietuvos gyventojų skaičiais. 2006 – 2019 m. laikotarpiu sergančiųjų skaičiai stipriai priklauso nuo Lietuvos gyventojų skaičiaus (priimama hipotezė $H_0\colon \rho = -0{,}9$, su tikimybe $0{,}95$). Ši priklausomybė yra tiesinė ir gali būti užrašyta regresijos lygtimi: $y = 437539 - 0{,}0757 \cdot x + \varepsilon, \varepsilon \sim N(0;\ 1444406{,}189)$, čia y – SPES skaičius, x – gyventojų skaičius. 2019 metų tos lygties prognozės intervalas yra $y \in [212108{,}82;\ 238259{,}94]$. Pritaikius statistinę hipotezę apie polinominio skirstinio parametro reikšmes, nustatyta, kad nagrinėjamų 7 ligų grupių SPES žmonių skaičiai statistiškai reikšmingai pasikeičia tik kas ketverius metus ir tas pokytis dar neišryškėja praėjus vieneriems ar dvejiems metams. Tai atsispindi ir SPES pokyčių dinamikos spalvų kitimo lentelėje. Taikant serijų kriterijų apie duomenų atsitiktinumą ir nepriklausomumą buvo nustatyta, kurioms duomenų grupėms yra taikytini laiko eilučių metodai. Taikant laiko eilučių išlyginimo metodą nustačius, kad visos atrinktos duomenų

eilutės turi trendą, buvo rasti AR(1) ir AR(2) modelių parametrų įverčiai, kai iš duomenų eilučių nėra eliminuojamas trendas ir kai trendas eliminuojamas.

THE STATISTICAL ANALYSIS OF THE MORBIDITY OF MENTAL AND BEHAVIOURAL DISORDERS OF LITHUANIA'S POPULATION IN THE EARLY 21ST CENTURY

SUMMARY

The object of this master's thesis is statistical and demographic data about the number of people suffering from mental and behavioral disorders (hereinafter referred to as ''PMBD") in Lithuania over the period from 2006 to 2019. Data was taken from the periodical health statistics publications ''Health of the Lithuanian Population and Activities of Health care institutions", published by Health Information Centre of the Institute of Hygiene. The aim of the work is to investigate the dynamics of the number of patients with mental and behavioral disorders by statistical methods during the period under consideration in Lithuania. The tasks set to achieve the goal are: first of all, to investigate the dependence of the number of Lithuanian patients with mental and behavioral disorders on the population using correlational and regression analysis, then applying statistical hypotheses about the meanings of the parameters of the polynomial distribution, to determine whether the number of patients of the 7 disease-groups under consideration changes dramatically after a certain period of time, if so, indicated after some period (after a few years). Continuing the study by disorder to investigate the dynamics of changes in the number of patients annually and every four years, to describe and depict it visually and then using a series of criteria on the randomness and independence of the data, to determine which data of patients with mental and behavioral disorders can be used time-series methods. Lastly, using the time series equalization method, determine which data rows have a trend and which do not, using time series methods and to find estimates of parameters for the AR(1) and AR(2) models when using and without using the trend.

After an analysis of the data of patients with mental and behavioral disorders and the population of Lithuania it was established that the PMBD figures correlate with the population figures of Lithuania. During the period from 2006 to 2019 the number of sick people strongly depends on the population of Lithuania (hypothesis $H_0: \rho = -0.9$, with a probability of 0.95 is adopted). This dependence is linear and can be written in the regression equation: $y = 437539 - 0{,}0757 \cdot x + \varepsilon, \varepsilon \sim N(0;\ 1444406{,}189)$, here y – number of PMBD, x – population. The forecast range for this equation for 2019 is $y \in [212108{,}82;\ 238259{,}94]$. After applying the statistical hypothesis on the values of the polynomial distribution parameter it was found that the number of people in the PMBD of the 7 disease groups under consideration changes statistically significantly only every four years and this change does

not yet appear after one or two years. This is also reflected in the color variation table of the dynamics of PMBD changes. The series criterion on data randomness and independence identified which data groups are subject to time series methods. When using the time series equalization method it was determined that all selected data rows have a trend, estimates of parameters for the AR(1) and AR(2) models were found when the trend is not eliminated from the data rows and when the trend is eliminated.

PRIEDAI

1 Priedas

Sergančiųjų psichikos ir elgesio sutrikimais (SPES) skaičius

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
0-17m.	35062	35054	36527	36651	37652	37935	38814
18+ m.	140688	144566	150867	149719	153046	162352	168859
SPES	**175751**	**179620**	**187394**	**186370**	**190698**	**200288**	**207673**

	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
0-17m.	37973	39139	38141	36580	36773	35225	34996
18+ m.	171867	174980	172881	175316	185939	196899	200866
SPES	**209840**	**214119**	**211022**	**211896**	**222712**	**232124**	**235862**

2 Priedas

Sergančiųjų psichikos ir elgesio sutrikimais (SPES) gyventojų pasiskirstymas pagal ligas-sutrikimus

Ligų grupės pagal TLK-10-AM	Iš viso 2006m.	Procentinė dalis 2006m.	Iš viso 2007m.	Procentinė dalis 2007m.	Iš viso 2008m.	Procentinė dalis 2008m.
Iš viso	175751		179620		187394	
1	11340	6,45%	13915	7,75%	17068	9,11%
2	10686	6,08%	11847	6,60%	10865	5,80%
3	714	0,41%	686	0,38%	847	0,45%
4	17042	9,70%	17167	9,56%	17479	9,33%
5	37818	21,52%	40308	22,44%	42593	22,73%
6	18532	10,54%	19367	10,78%	21028	11,22%
7	9862	5,61%	9887	5,50%	10031	5,35%

Ligų grupės pagal TLK-10-AM	Iš viso 2009m.	Procentinė dalis 2009m.	Iš viso 2010m.	Procentinė dalis 2010m.	Iš viso 2011m.	Procentinė dalis 2011m.
Iš viso	186370		190698		200288	
1	19609	10,52%	21233	11,13%	25156	12,56%
2	9129	4,90%	9542	5,00%	12306	6,14%
3	979	0,53%	958	0,50%	1278	0,64%
4	17900	9,60%	18096	9,49%	18648	9,31%
5	42963	23,05%	44534	23,35%	46821	23,38%
6	22573	12,11%	24273	12,73%	26057	13,01%
7	10186	5,47%	10429	5,47%	10442	5,21%

Ligų grupės pagal TLK-10-AM	Iš viso 2012m.	Procentinė dalis 2012m.	Iš viso 2013m.	Procentinė dalis 2013m.	Iš viso 2014m.	Procentinė dalis 2014m.
Iš viso	207673		209840		214119	
1	28655	13,80%	29184	13,91%	30806	14,39%
2	14001	6,74%	18259	8,70%	19373	9,05%
3	1408	0,68%	1722	0,82%	1844	0,86%
4	18877	9,09%	19115	9,11%	18872	8,81%
5	48594	23,40%	47862	22,81%	47584	22,22%
6	27114	13,06%	27281	13,00%	28655	13,38%
7	11014	5,30%	10781	5,14%	11407	5,33%

Ligų grupės pagal TLK-10-AM	Iš viso 2015m.	Procentinė dalis 2015m.	Iš viso 2016m.	Procentinė dalis 2016m.	Iš viso 2017m.	Procentinė dalis 2017m.
Iš viso	211022		211896		222712	
1	31559	14,96%	33170	15,65%	33056	14,84%
2	19908	9,43%	20445	9,65%	19127	8,59%
3	2153	1,02%	2132	1,01%	2269	1,02%
4	18591	8,81%	18969	8,95%	18765	8,43%
5	45793	21,70%	46099	21,76%	46873	21,05%
6	28827	13,66%	28773	13,58%	29677	13,33%
7	11383	5,39%	10816	5,10%	11515	5,17%

Ligų grupės pagal TLK-10-AM	Iš viso 2018m.	Procentinė dalis 2018m.	Iš viso 2019m.	Procentinė dalis 2019m.
Iš viso	232124		235862	
1	32362	13,94%	32617	13,83%
2	18308	7,89%	18632	7,90%
3	3001	1,29%	3278	1,39%
4	18588	8,01%	18603	7,89%
5	46329	19,96%	46345	19,65%
6	29501	12,71%	30445	12,91%
7	11145	4,80%	11247	4,77%

3 Priedas

Procentinis sergančiųjų psichikos ir elgesio sutrikimais (SPES) gyventojų pasiskirstymas pagal ligas-sutrikimus

Ligos pogrupis	2006	2007	2008	2009	2010	2011	2012
1	6,45%	7,75%	9,11%	10,52%	11,13%	12,56%	13,80%
2	6,08%	6,60%	5,80%	4,90%	5,00%	6,14%	6,74%
3	0,41%	0,38%	0,45%	0,53%	0,50%	0,64%	0,68%
4	9,70%	9,56%	9,33%	9,60%	9,49%	9,31%	9,09%
5	21,52%	22,44%	22,73%	23,05%	23,35%	23,38%	23,40%
6	10,54%	10,78%	11,22%	12,11%	12,73%	13,01%	13,06%
7	5,61%	5,50%	5,35%	5,47%	5,47%	5,21%	5,30%
SPES	175751	179620	187394	186370	190698	200288	207673

Ligos pogrupis	2013	2014	2015	2016	2017	2018	2019
1	13,91%	14,39%	14,96%	15,65%	14,84%	13,94%	13,83%
2	8,70%	9,05%	9,43%	9,65%	8,59%	7,89%	7,90%
3	0,82%	0,86%	1,02%	1,01%	1,02%	1,29%	1,39%
4	9,11%	8,81%	8,81%	8,95%	8,43%	8,01%	7,89%
5	22,81%	22,22%	21,70%	21,76%	21,05%	19,96%	19,65%
6	13,00%	13,38%	13,66%	13,58%	13,33%	12,71%	12,91%
7	5,14%	5,33%	5,39%	5,10%	5,17%	4,80%	4,77%
SPES	209840	214119	211022	211896	222712	232124	235862

4 Priedas

Sergančiųjų psichikos ir elgesio sutrikimais (SPES) gyventojų pasiskirstymas pagal ligas-sutrikimus

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
1	11340	13915	17068	19609	21233	25156	28655
2	10686	11847	10865	9129	9542	12306	14001
3	714	686	847	979	958	1278	1408
4	17042	17167	17479	17900	18096	18648	18877
5	37818	40308	42593	42963	44534	46821	48594
6	18532	19367	21028	22573	24273	26057	27114
7	9862	9887	10031	10186	10429	10442	11014
SPES	175751	179620	187394	186370	190698	200288	207673

	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1	29184	30806	31559	33170	33056	32362	32617
2	18259	19373	19908	20445	19127	18308	18632
3	1722	1844	2153	2132	2269	3001	3278
4	19115	18872	18591	18969	18765	18588	18603
5	47862	47584	45793	46099	46873	46329	46345
6	27281	28655	28827	28773	29677	29501	30445
7	10781	11407	11383	10816	11515	11145	11247
SPES	209840	214119	211022	211896	222712	232124	235862

Serijų kriterijaus taikymas

Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo (2)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	10686	11847	10865	9129	9542	12306	14001

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	18259	19373	19908	20445	19127	18308	18632

$$M_e = 16130$$

Lyginame

$$(-------)(+++++++)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

N = N(serijų skaičius) = 2.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3.$

   $N_{0,025}(V, 7{,}7) = 13.$

   $\overline{W} = \{3 \nless 2 < 13\} \Rightarrow H_0$ atmetama.

2) Kai $\alpha = 0{,}1$

$N_{0,05}(A, 7{,}7) = 4.$

$N_{0,05}(V, 7{,}7) = 12.$

   $\overline{W} = \{4 \nless 2 < 12\} \Rightarrow H_0$ atmetama.

Psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo (3)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	714	686	847	979	958	1278	1408

Metai	2013	2014	2015	2016	2017	2018	2019

Sergančiųjų skaičius	1722	1844	2153	2132	2269	3001	3278

$$M_e = 1565$$

Lyginame

$$(- - - - - - -)(+ + + + + + +)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

$N = N($serijų skaičius$) = 2$.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3.$

   $N_{0,025}(V, 7{,}7) = 13.$

   $\overline{W} = \{3 \nless 2 < 13\} \Rightarrow H_0$ atmetama.

2) Kai $\alpha = 0{,}1$

$N_{0,05}(A, 7{,}7) = 4.$

$N_{0,05}(V, 7{,}7) = 12.$

   $\overline{W} = \{4 \nless 2 < 12\} \Rightarrow H_0$ atmetama.

Šizofrenija, šizotipinis ir kliedesiniai sutrikimai (4)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	17042	17167	17479	17900	18096	18648	18877

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	19115	18872	18591	18969	18765	18588	18603

$$M_e = 18597$$

Lyginame

$$(- - - - -)(+ + + +)(-)(+ +)(-)(+)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

$N = N($serijų skaičius$) = 6$.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7$.

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3$.

   $N_{0,025}(V, 7{,}7) = 13$.

   $\overline{W} = \{3 < 6 < 13\} \Rightarrow H_0$ priimama.

2) Kai $\alpha = 0{,}1$

$N_{0,05}(A, 7{,}7) = 4$.

$N_{0,05}(V, 7{,}7) = 12$.

   $\overline{W} = \{4 < 6 < 12\} \Rightarrow H_0$ priimama.

Nuotaikos (afektiniai sutrikimai) (5)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	37818	40308	42593	42963	44534	46821	48594

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	47862	47584	45793	46099	46873	46329	46345

$$M_e = 46214$$

Lyginame

$$(-----)(++++)(--)(+\ +\ +)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

N = *N(*serijų skaičius) = 4.

$k_1 = 7, k_2 = 7$.

$n_1 = 7, n_2 = 7$.

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3$.

   $N_{0,025}(V, 7{,}7) = 13$.

   $\overline{W} = \{3 < 4 < 13\} \Rightarrow H_0$ priimama.

2) Kai $\alpha = 0{,}1$

$N_{0,05}(A, 7{,}7) = 4$.

$N_{0,05}(V, 7{,}7) = 12$.

   $\overline{W} = \{4 \nless 4 < 12\} \Rightarrow H_0$ atmetama.

Protinis atsilikimas ir psichologinės raidos sutrikimai (6)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	18532	19367	21028	22573	24273	26057	27114

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	27281	28655	28827	28773	29677	29501	30445

$$M_e = 27197{,}5$$

Lyginame

$$(-------)(+++++++)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

N = N(serijų skaičius) = 2.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

   $N_{0{,}025}(A, 7{,}7) = 3.$

   $N_{0{,}025}(V, 7{,}7) = 13.$

   $\overline{W} = \{3 \nless 2 < 13\} \Rightarrow H_0$ atmetama.

2) Kai $\alpha = 0{,}1$

$N_{0{,}05}(A, 7{,}7) = 4.$

$N_{0{,}05}(V, 7{,}7) = 12.$

   $\overline{W} = \{4 \nless 2 < 12\} \Rightarrow H_0$ atmetama.

Elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje (7)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	9862	9887	10031	10186	10429	10442	11014

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	10781	11407	11383	10816	11515	11145	11247

$$M_e = 10798{,}5$$

Lyginame

$$(------)(+)(-)(++++++)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

N = N(serijų skaičius) = 4.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3.$

   $N_{0,025}(V, 7{,}7) = 13.$

   $\bar{W} = \{3 < 4 < 13\} \Rightarrow H_0$ priimama.

2) Kai $\alpha = 0{,}1$

$N_{0,05}(A, 7{,}7) = 4.$

$N_{0,05}(V, 7{,}7) = 12.$

   $\bar{W} = \{4 \nless 4 < 12\} \Rightarrow H_0$ atmetama.

Bendras sergančiųjų psichikos ir elgesio sutrikimais (SPES) gyventojų skaičius

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	175751	179620	187394	186370	190698	200288	207673

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	209840	214119	211022	211896	222712	232124	235862

$$M_e = 208756{,}5$$

Lyginame

$$(-------)(+++++++)$$

H_0: duomenys nepriklausomi ir atsitiktiniai.

N = N(serijų skaičius) = 2.

$k_1 = 7, k_2 = 7.$

$n_1 = 7, n_2 = 7.$

1) Kai $\alpha = 0{,}05$

   $N_{0,025}(A, 7{,}7) = 3.$

$N_{0,025}(V, 7,7) = 13.$

$\bar{W} = \{3 \nless 2 < 13\} \Rightarrow H_0$ atmetama.

2) Kai $\alpha = 0,1$

$N_{0,05}(A, 7,7) = 4.$

$N_{0,05}(V, 7,7) = 12.$

$\bar{W} = \{4 \nless 2 < 12\} \Rightarrow H_0$ atmetama.

AR(1) modelių radimas

Demencija ir Alzheimerio liga (1)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	11340	13915	17068	19609	21233	25156	28655

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	29184	30806	31559	33170	33056	32362	32617

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 11340,\ X_2 = 13915,\ X_3 = 17068,\ X_4 = 19609,\ X_5 = 21233,\ X_6 = 25156,\ X_7 = 28655,\ X_8 = 29184,\ X_9 = 30806,\ X_{10} = 31559,\ X_{11} = 33170,\ X_{12} = 33056,\ X_{13} = 32362,\ X_{14} = 32617$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t (X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{13915 \cdot 11340 + 17068 \cdot 13915 + \cdots + 32362 \cdot 33056 + 32617 \cdot 32362}{11340^2 + 13915^2 + 17068^2 + \cdots + 32362^2 + 32617^2} = 1{,}049$$

Psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo (3)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	714	686	847	979	958	1278	1408

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	1722	1844	2153	2132	2269	3001	3278

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 714,\ X_2 = 686,\ X_3 = 847,\ X_4 = 979,\ X_5 = 958,\ X_6 = 1278,\ X_7 = 1408,\ X_8 = 1722,\ X_9 = 1844, X_{10} = 2153,\ X_{11} = 2132,\ X_{12} = 2269,\ X_{13} = 3001,\ X_{14} = 3278$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t(X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{686 \cdot 714 + 847 \cdot 686 + \cdots + 3001 \cdot 2269 + 3278 \cdot 3001}{714^2 + 686^2 + 847^2 + \cdots + 3001^2 + 3278^2} = 1{,}125$$

Nuotaikos (afektiniai sutrikimai) (5)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	37818	40308	42593	42963	44534	46821	48594

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	47862	47584	45793	46099	46873	46329	46345

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 37818,\ X_2 = 40308,\ X_3 = 42593,\ X_4 = 42963,\ X_5 = 44534,\ X_6 = 46821,\ X_7 = 48594,\ X_8 = 47862,\ X_9 = 47584,\ X_{10} = 45793, X_{11} = 46099,\ X_{12} = 46873, X_{13} = 46329,\ X_{14} = 46345$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t(X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{40308 \cdot 37818 + 42593 \cdot 40308 + \cdots + 46329 \cdot 46873 + 46345 \cdot 46329}{37818^2 + 40308^2 + 42593^2 + \cdots + 46329^2 + 46345^2} = 1{,}013$$

Protinis atsilikimas ir psichologinės raidos sutrikimai (6)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	18532	19367	21028	22573	24273	26057	27114

Metai	2013	2014	2015	2016	2017	2018	2019

Sergančiųjų skaičius	27281	28655	28827	28773	29677	29501	30445

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 18532,\ X_2 = 19367,\ X_3 = 21028,\ X_4 = 22573,\ X_5 = 24273,\ X_6 = 26057,\ X_7 = 27114,\ X_8 = 27281,\ X_9 = 28655,\ X_{10} = 28827,\ X_{11} = 28773,\ X_{12} = 29677,\ X_{13} = 29501,\ X_{14} = 30445$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t (X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{19367 \cdot 18532 + 21028 \cdot 19367 + \cdots + 29501 \cdot 29677 + 30445 \cdot 29501}{18532^2 + 19367^2 + 21028^2 + \cdots + 29501^2 + 30445^2} = 1{,}033$$

Elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje (7)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	9862	9887	10031	10186	10429	10442	11014

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	10781	11407	11383	10816	11515	11145	11247

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 9862,\ X_2 = 9887,\ X_3 = 10031,\ X_4 = 10186,\ X_5 = 10429,\ X_6 = 10442,\ X_7 = 11014,\ X_8 = 10781,\ X_9 = 11407,\ X_{10} = 11383,\ X_{11} = 10816,\ X_{12} = 11515,\ X_{13} = 11145,\ X_{14} = 11247$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t (X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{9887 \cdot 9862 + 10031 \cdot 9887 + \cdots + 11145 \cdot 11515 + 11247 \cdot 11145}{9862^2 + 9887^2 + 10031^2 + \cdots + 11145^2 + 11247^2}$$

$$= 1{,}009$$

Bendras sergančiųjų psichikos ir elgesio sutrikimais gyventojų (SPES) skaičius

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	175751	179620	187394	186370	190698	200288	207673

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	209840	214119	211022	211896	222712	232124	235862

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$$X_1 = 175751,\ X_2 = 179620,\ X_3 = 187394,\ X_4 = 186370,\ X_5 = 190698,\ X_6 = 200288,\ X_7 = 207673,\ X_8 = 209840,\ X_9 = 214119, X_{10} = 211022,\ X_{11} = 211896,\ X_{12} = 222712,\ X_{13} = 232124,\ X_{14} = 235862$$

Mažiausių kvadratų metodu $\hat{a}_1$ įvertis yra:

$$\hat{a}_1 = \frac{\sum_{t=2}^{n} X_t(X_{t-1})}{\sum_{t=2}^{n} X_{t-1}^2} = \frac{X_2X_1 + X_3X_2 + \cdots + X_{13}X_{12} + X_{14}X_{13}}{X_1^2 + X_2^2 + X_3^2 + \cdots + X_{13}^2 + X_{14}^2}$$

$$= \frac{179620 \cdot 175751 + 187394 \cdot 179620 + \cdots + 232124 \cdot 222712 + 235862 \cdot 232124}{175751^2 + 179620^2 + 187394^2 + \cdots + 232124^2 + 235862^2}$$

$$= 1{,}023$$

AR(2) modelių radimas

Demencija ir Alzheimerio liga (1)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	11340	13915	17068	19609	21233	25156	28655

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	29184	30806	31559	33170	33056	32362	32617

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$$X_1 = 11340,\ X_2 = 13915,\ X_3 = 17068,\ X_4 = 19609,\ X_5 = 21233,\ X_6 = 25156,\ X_7 = 28655,\ X_8 = 29184,\ X_9 = 30806, X_{10} = 31559,\ X_{11} = 33170,\ X_{12} = 33056,\ X_{13} = 32362,\ X_{14} = 32617$$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 1{,}331$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = -0{,}303$$

$$X_t = 1{,}331 \cdot X_{t-1} + (-0{,}303) \cdot X_{t-2} + \varepsilon_t$$

Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo (2)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	10686	11847	10865	9129	9542	12306	14001

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**

Sergančiųjų skaičius	18259	19373	19908	20445	19127	18308	18632

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 10686,\ X_2 = 11847,\ X_3 = 10865,\ X_4 = 9129,\ X_5 = 9542,\ X_6 = 12306,\ X_7 = 14001,\ X_8 = 18259,\ X_9 = 19373, X_{10} = 19908,\ X_{11} = 20445,\ X_{12} = 19127,\ X_{13} = 18308,\ X_{14} = 18632$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_3x_1 + x_4x_2 + \dots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = 1{,}193$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_3x_1 + x_4x_2 + \dots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = -0{,}175$$

$$X_t = 1{,}193 \cdot X_{t-1} + (-0{,}175) \cdot X_{t-2} + \varepsilon_t$$

Nuotaikos (afektiniai sutrikimai) (5)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	37818	40308	42593	42963	44534	46821	48594

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	47862	47584	45793	46099	46873	46329	46345

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 37818,\ X_2 = 40308,\ X_3 = 42593,\ X_4 = 42963,\ X_5 = 44534,\ X_6 = 46821,\ X_7 = 48594,\ X_8 = 47862,\ X_9 = 47584, X_{10} = 45793,\ X_{11} = 46099,\ X_{12} = 46873,\ X_{13} = 46329,\ X_{14} = 46345$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_3x_1 + x_4x_2 + \dots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = 1{,}072$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_3x_1 + x_4x_2 + \dots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = -0{,}063$$

$$X_t = 1{,}072 \cdot X_{t-1} + (-0{,}063) \cdot X_{t-2} + \varepsilon_t$$

Protinis atsilikimas ir psichologinės raidos sutrikimai (6)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	18532	19367	21028	22573	24273	26057	27114

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	27281	28655	28827	28773	29677	29501	30445

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$$X_1 = 18532,\ X_2 = 19367,\ X_3 = 21028,\ X_4 = 22573,\ X_5 = 24273,\ X_6 = 26057,\ X_7 = 27114,\ X_8 = 27281,\ X_9 = 28655, X_{10} = 28827,\ X_{11} = 28773,\ X_{12} = 29677,\ X_{13} = 29501,\ X_{14} = 30445$$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_3x_1 + x_4x_2 + \dots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = 1{,}054$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_3x_1 + x_4x_2 + \dots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \dots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \dots + x_{13}^2)(x_1^2 + x_2^2 + \dots + x_{12}^2) - (x_2x_1 + x_3x_2 + \dots + x_{13}x_{12})^2} = -0{,}022$$

$$X_t = 1{,}054 \cdot X_{t-1} + (-0{,}022) \cdot X_{t-2} + \varepsilon_t$$

Elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje (7)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	9862	9887	10031	10186	10429	10442	11014

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	10781	11407	11383	10816	11515	11145	11247

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$$X_1 = 9862,\ X_2 = 9887,\ X_3 = 10031,\ X_4 = 10186,\ X_5 = 10429,\ X_6 = 10442,\ X_7 = 11014,\ X_8 = 10781,\ X_9 = 11407, X_{10} = 11383,\ X_{11} = 10816,\ X_{12} = 11515,\ X_{13} = 11145,\ X_{14} = 11247$$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 0{,}992$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 0{,}019$$

$$X_t = 0{,}992 \cdot X_{t-1} + 0{,}019 \cdot X_{t-2} + \varepsilon_t$$

Bendras sergančiųjų psichikos ir elgesio sutrikimais gyventojų (SPES) skaičius

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	175751	179620	187394	186370	190698	200288	207673

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	209840	214119	211022	211896	222712	232124	235862

$$n = 14$$

$$X_t = \hat{a}_1 \cdot X_{t-1} + \hat{a}_2 \cdot X_{t-2} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \widehat{\sigma^2})$$

$X_1 = 175751,\ X_2 = 179620,\ X_3 = 187394,\ X_4 = 186370,\ X_5 = 190698,\ X_6 = 200288,\ X_7 = 207673,\ X_8 = 209840,\ X_9 = 214119, X_{10} = 211022,\ X_{11} = 211896,\ X_{12} = 222712,\ X_{13} = 232124,\ X_{14} = 235862$

$$\hat{a}_1 = \frac{(x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = 1{,}023$$

$$\hat{a}_2 = \frac{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_3x_1 + x_4x_2 + \cdots + x_{14}x_{12}) - (x_2x_1 + x_3x_2 + \cdots + x_{14}x_{13})(x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})}{(x_1^2 + x_2^2 + \cdots + x_{13}^2)(x_1^2 + x_2^2 + \cdots + x_{12}^2) - (x_2x_1 + x_3x_2 + \cdots + x_{13}x_{12})^2} = -0{,}00002$$

$$X_t = 1{,}023 \cdot X_{t-1} + (-0{,}00002) \cdot X_{t-2} + \varepsilon_t$$

Laiko eilučių išlyginimo metodo taikymas duomenims, kuriems taikytina laiko eilučių teorija

Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo (2)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	10686	11847	10865	9129	9542	12306	14001

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	18259	19373	19908	20445	19127	18308	18632

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+" ir „-" eilutes bei radome medianos reikšmę, kur

$$M_e = 16130$$

$$n = 14$$

$$\gamma(n) = 2 - bendras\ serijų\ skaičius$$

$$\tau(n) = 7 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$

$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$2 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14-1})\right] \Rightarrow 2 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1} t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 890{,}61 \cdot t - 1777181{,}19$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (890{,}61 \cdot t - 1777181{,}19)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$890{,}61 \cdot t - 1777181{,}19$	$\boldsymbol{Y_t}$
2006	10686	1	-1776290,580	1786976,580
2007	11847	2	-1775399,970	1787246,970
2008	10865	3	-1774509,360	1785374,360
2009	9129	4	-1773618,750	1782747,750
2010	9542	5	-1772728,140	1782270,140
2011	12306	6	-1771837,530	1784143,530
2012	14001	7	-1770946,920	1784947,920
2013	18259	8	-1770056,310	1788315,310
2014	19373	9	-1769165,700	1788538,700
2015	19908	10	-1768275,090	1788183,090
2016	20445	11	-1767384,480	1787829,480
2017	19127	12	-1766493,870	1785620,870
2018	18308	13	-1765603,260	1783911,260
2019	18632	14	-1764712,650	1783344,650

Surasime AR(1) proceso lygtį sergančiųjų psichikos ir elgesio sutrikimais dėl alkoholio vartojimo duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t,$$

ir stebinius:

$$Y_1 = 1786976{,}580;\ Y_2 = 1787246{,}970;\ Y_3 = 1785374{,}360;\ Y_4 = 1782747{,}750;$$
$$Y_5 = 1782270{,}140; Y_6 = 1784143{,}530;\ Y_7 = 1784947{,}920; Y_8 = 1788315{,}310;$$
$$Y_9 = 1788538{,}700$$
$$Y_{10} = 1788183{,}090;\ Y_{11} = 1787829{,}480;\ Y_{12} = 1785620{,}870;\ Y_{13} = 1783911{,}260;$$
$$Y_{14} = 1783344{,}650.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1785675{,}044.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 4424289{,}179 \\ \hat{\gamma}(1) = 2933188{,}343 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 4424289{,}179 = \hat{a}_1 \cdot 2933188{,}343 + \sigma^2 \\ 2933188{,}343 = \hat{a}_1 \cdot 4424289{,}179 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}663;\ \hat{\sigma}^2 = 2479661{,}802.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}663 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1574{,}694193^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 300{,}136 \cdot t - 598910{,}061 + 0{,}663 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1574{,}694193^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{\mathrm{a}}_1 \cdot Y_{t-1} + \hat{\mathrm{a}}_2 \cdot Y_{t-2} + \varepsilon_t,$$

ir stebinius:

$$Y_1 = 1786976{,}580;\ Y_2 = 1787246{,}970;\ Y_3 = 1785374{,}360;\ Y_4 = 1782747{,}750;$$
$$Y_5 = 1782270{,}140; Y_6 = 1784143{,}530;\ Y_7 = 1784947{,}920; Y_8 = 1788315{,}310;$$
$$Y_9 = 1788538{,}700$$
$$Y_{10} = 1788183{,}090;\ Y_{11} = 1787829{,}480;\ Y_{12} = 1785620{,}870;\ Y_{13} = 1783911{,}260;$$
$$Y_{14} = 1783344{,}650.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1785675{,}044.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 4424289{,}179 \\ \hat{\gamma}(1) = 2933188{,}343 \\ \hat{\gamma}(2) = 2024221{,}825 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 4424289{,}179 = \hat{a}_1 \cdot 2933188{,}343 + \hat{a}_2 \cdot 2024221{,}825 + \sigma^2 \\ 2933188{,}343 = \hat{a}_1 \cdot 4424289{,}179 + \hat{a}_2 \cdot 2933188{,}343 \\ 2024221{,}825 = \hat{a}_1 \cdot 2933188{,}343 + \hat{a}_2 \cdot 4424289{,}179 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}642; \hat{a}_2 = 0{,}032; \hat{\sigma}^2 = 2477106{,}907.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}642 \cdot Y_{t-1} + 0{,}032 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1573{,}883^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 290{,}524 \cdot t - 579730{,}559 + +0{,}642 \cdot X_{t-1} + 0{,}032 \cdot X_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 1573{,}883^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Psichikos ir elgesio sutrikimai dėl kitų psichoaktyviųjų medžiagų vartojimo (3)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	714	686	847	979	958	1278	1408

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	1722	1844	2153	2132	2269	3001	3278

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+" ir „-" eilutes bei radome medianos reikšmę, kur

$$M_e = 1565$$
$$n = 14$$
$$\gamma(n) = 2 - bendras\ serij\text{ų}\ skai\check{c}ius$$
$$\tau(n) = 7 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$
$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$2 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14 - 1})\right] \Rightarrow 2 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1} t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 192{,}64 \cdot t - 386034$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (192{,}64 \cdot t - 386034)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$192{,}64 \cdot t - 386034$	$\boldsymbol{Y_t}$
2006	714	1	-385841,360	386555,360
2007	686	2	-385648,720	386334,720
2008	847	3	-385456,080	386303,080
2009	979	4	-385263,440	386242,440
2010	958	5	-385070,800	386028,800
2011	1278	6	-384878,160	386156,160
2012	1408	7	-384685,520	386093,520
2013	1722	8	-384492,880	386214,880
2014	1844	9	-384300,240	386144,240
2015	2153	10	-384107,600	386260,600
2016	2132	11	-383914,960	386046,960
2017	2269	12	-383722,320	385991,320
2018	3001	13	-383529,680	386530,680
2019	3278	14	-383337,040	386615,040

Surasime AR(1) proceso lygtį sergančiųjų psichikos ir elgesio sutrikimais dėl kitų psichoaktyviųjų medžiagų vartojimo duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t,$$

ir stebinius:

$Y_1 = 386555{,}360;\ Y_2 = 386334{,}720;\ Y_3 = 386303{,}080;\ Y_4 = 386242{,}440;$

$Y_5 = 386028{,}800; Y_6 = 386156{,}160;\ Y_7 = 386093{,}520; Y_8 = 386214{,}880;\ Y_9 = 386144{,}240;$

$$Y_{10} = 386260{,}600;\ Y_{11} = 386046{,}960;\ Y_{12} = 385991{,}320;\ Y_{13} = 386530{,}680;$$
$$Y_{14} = 386615{,}040.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 386251{,}271.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 37014{,}471 \\ \hat{\gamma}(1) = 11158{,}552 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 37014{,}471 = \hat{a}_1 \cdot 11158{,}552 + \sigma^2 \\ 11158{,}552 = \hat{a}_1 \cdot 37014{,}471 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}301;\ \hat{\sigma}^2 = 33650{,}562.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}301 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 183{,}441^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 134{,}655 \cdot t - 269837{,}766 + 0{,}301 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 183{,}441^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 386555{,}360;\ Y_2 = 386334{,}720;\ Y_3 = 386303{,}080;\ Y_4 = 386242{,}440;$$
$$Y_5 = 386028{,}800; Y_6 = 386156{,}160;\ Y_7 = 386093{,}520; Y_8 = 386214{,}880;\ Y_9 = 386144{,}240;$$
$$Y_{10} = 386260{,}600;\ Y_{11} = 386046{,}960;\ Y_{12} = 385991{,}320;\ Y_{13} = 386530{,}680;$$
$$Y_{14} = 386615{,}040.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 386251{,}271.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k} \left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 37014{,}471 \\ \hat{\gamma}(1) = 11158{,}552 \\ \hat{\gamma}(2) = 9824{,}264749 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 37014{,}471 = \hat{a}_1 \cdot 11158{,}552 + \hat{a}_2 \cdot 9824{,}264749 + \sigma^2 \\ 11158{,}552 = \hat{a}_1 \cdot 37014{,}471 + \hat{a}_2 \cdot 11158{,}552 \\ 9824{,}264749 = \hat{a}_1 \cdot 11158{,}552 + \hat{a}_2 \cdot 37014{,}471 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}244; \hat{a}_2 = 0{,}192; \hat{\sigma}^2 = 32410{,}280.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}244 \cdot Y_{t-1} + 0{,}192 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 180{,}029^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 108{,}731 \cdot t - 217888{,}422 + 0{,}244 \cdot X_{t-1} + 0{,}192 \cdot X_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 180{,}029^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Nuotaikos (afektiniai sutrikimai) (5)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	37818	40308	42593	42963	44534	46821	48594

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	47862	47584	45793	46099	46873	46329	46345

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+" ir „-" eilutes bei radome medianos reikšmę, kur

$$M_e = 46214$$
$$n = 14$$
$$\gamma(n) = 4 - bendras\ serijų\ skaičius$$
$$\tau(n) = 5 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$

$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$4 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14-1})\right] \Rightarrow 4 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1} t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 539{,}35 \cdot t - 1040412{,}76$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (539{,}35 \cdot t - 1040412{,}76)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$539{,}35 \cdot t - 1040412{,}76$	$\boldsymbol{Y_t}$
2006	37818	1	-1039873,410	1077691,410
2007	40308	2	-1039334,060	1079642,060
2008	42593	3	-1038794,710	1081387,710
2009	42963	4	-1038255,360	1081218,360
2010	44534	5	-1037716,010	1082250,010
2011	46821	6	-1037176,660	1083997,660
2012	48594	7	-1036637,310	1085231,310
2013	47862	8	-1036097,960	1083959,960
2014	47584	9	-1035558,610	1083142,610
2015	45793	10	-1035019,260	1080812,260
2016	46099	11	-1034479,910	1080578,910
2017	46873	12	-1033940,560	1080813,560
2018	46329	13	-1033401,210	1079730,210
2019	46345	14	-1032861,860	1079206,860

Surasime AR(1) proceso lygtį sergančiųjų nuotaikos (afektiniais sutrikimais) duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 1077691{,}410;\ Y_2 = 1079642{,}060;\ Y_3 = 1081387{,}710;\ Y_4 = 1081218{,}360;$$
$$Y_5 = 1082250{,}010; Y_6 = 1083997{,}660;\ Y_7 = 1085231{,}310; Y_8 = 1083959{,}960;$$
$$Y_9 = 1083142{,}610;$$
$$Y_{10} = 1080812{,}260; Y_{11} = 1080578{,}910;\ Y_{12} = 1080813{,}560;\ Y_{13} = 1079730{,}210;$$
$$Y_{14} = 1079206{,}860.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1081404{,}492$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 4112702{,}195 \\ \hat{\gamma}(1) = 2669486{,}923 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 4112702{,}195 = \hat{a}_1 \cdot 2669486{,}923 + \sigma^2 \\ 2669486{,}923 = \hat{a}_1 \cdot 4112702{,}195 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}649;\ \hat{\sigma}^2 = 2379982{,}418.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}649 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1542{,}719^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 189{,}312 \cdot t - 365184{,}879 + 0{,}649 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1542{,}719^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 1077691{,}410;\ Y_2 = 1079642{,}060;\ Y_3 = 1081387{,}710;\ Y_4 = 1081218{,}360;$$
$$Y_5 = 1082250{,}010; Y_6 = 1083997{,}660;\ Y_7 = 1085231{,}310; Y_8 = 1083959{,}960;$$
$$Y_9 = 1083142{,}610;$$

$$Y_{10} = 1080812{,}260; Y_{11} = 1080578{,}910;\ Y_{12} = 1080813{,}560;\ Y_{13} = 1079730{,}210;$$
$$Y_{14} = 1079206{,}860.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1081404{,}492.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 4112702{,}195 \\ \hat{\gamma}(1) = 2669486{,}923 \\ \hat{\gamma}(2) = 1579335{,}636 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 4112702{,}195 = \hat{a}_1 \cdot 2669486{,}923 + \hat{a}_2 \cdot 1579335{,}636 + \sigma^2 \\ 2669486{,}923 = \hat{a}_1 \cdot 4112702{,}195 + \hat{a}_2 \cdot 2669486{,}923 \\ 1579335{,}636 = \hat{a}_1 \cdot 2669486{,}923 + \hat{a}_2 \cdot 4112702{,}195 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}691; \hat{a}_2 = -0{,}064; \hat{\sigma}^2 = 2370097{,}179.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}691 \cdot Y_{t-1} - 0{,}064 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 1539{,}512^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 201{,}465 \cdot t - 388627{,}750 + 0{,}691 \cdot X_{t-1} - 0{,}064 \cdot X_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 1539{,}512^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Protinis atsilikimas ir psichologinės raidos sutrikimai (6)

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	18532	19367	21028	22573	24273	26057	27114

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	27281	28655	28827	28773	29677	29501	30445

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+” ir „-“ eilutes bei radome medianos reikšmę, kur

$$M_e = 27197,5$$

$$n = 14$$

$$\gamma(n) = 2 - bendras\ serijų\ skaičius$$

$$\tau(n) = 7 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1,96\sqrt{n-1})\right]$$

$$\tau(n) < [1,43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$2 > \left[\frac{1}{2}(14 + 2 - 1,96\sqrt{14-1})\right] \Rightarrow 2 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1}t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 919,37 \cdot t - 1824374,92$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (919,37 \cdot t - 1824374,92)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$919,37 \cdot t - 1824374,92$	$\boldsymbol{Y_t}$
2006	18532	1	-1823455,550	1841987,550
2007	19367	2	-1822536,180	1841903,180
2008	21028	3	-1821616,810	1842644,810
2009	22573	4	-1820697,440	1843270,440
2010	24273	5	-1819778,070	1844051,070
2011	26057	6	-1818858,700	1844915,700

2012	27114	7	-1817939,330	1845053,330
2013	27281	8	-1817019,960	1844300,960
2014	28655	9	-1816100,590	1844755,590
2015	28827	10	-1815181,220	1844008,220
2016	28773	11	-1814261,850	1843034,850
2017	29677	12	-1813342,480	1843019,480
2018	29501	13	-1812423,110	1841924,110
2019	30445	14	-1811503,740	1841948,740

Surasime AR(1) proceso lygtį sergančiųjų protiniu atsilikimu ir psichologinės raidos sutrikimais duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 1841987{,}550;\ Y_2 = 1841903{,}180;\ Y_3 = 1842644{,}810;\ Y_4 = 1843270{,}440;$$
$$Y_5 = 1844051{,}070; Y_6 = 1844915{,}700;\ Y_7 = 1845053{,}330; Y_8 = 1844300{,}960;$$
$$Y_9 = 1844755{,}590;$$
$$Y_{10} = 1844008{,}220;\ Y_{11} = 1843034{,}850;\ Y_{12} = 1843019{,}480;\ Y_{13} = 1841924{,}110;$$
$$Y_{14} = 1841948{,}740.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1843344{,}145$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 1272533{,}303 \\ \hat{\gamma}(1) = 929983{,}301 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 1272533{,}303 = \hat{a}_1 \cdot 929983{,}301 + \sigma^2 \\ 929983{,}301 = \hat{a}_1 \cdot 1272533{,}303 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}731;\ \hat{\sigma}^2 = 592889{,}840.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}731 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 769{,}993^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 247{,}311 \cdot t - 490756{,}853 + 0{,}731 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 769{,}993^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 1841987{,}550;\ Y_2 = 1841903{,}180;\ Y_3 = 1842644{,}810;\ Y_4 = 1843270{,}440;$$
$$Y_5 = 1844051{,}070; Y_6 = 1844915{,}700;\ Y_7 = 1845053{,}330; Y_8 = 1844300{,}960;$$
$$Y_9 = 1844755{,}590;$$
$$Y_{10} = 1844008{,}220;\ Y_{11} = 1843034{,}850;\ Y_{12} = 1843019{,}480;\ Y_{13} = 1841924{,}110;$$
$$Y_{14} = 1841948{,}740.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 1843344{,}145$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 1272533{,}303 \\ \hat{\gamma}(1) = 929983{,}301 \\ \hat{\gamma}(2) = 640316{,}7532 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 1272533{,}303 = \hat{a}_1 \cdot 929983{,}301 + \hat{a}_2 \cdot 640316{,}7532 + \sigma^2 \\ 929983{,}301 = \hat{a}_1 \cdot 1272533{,}303 + \hat{a}_2 \cdot 929983{,}301 \\ 640316{,}7532 = \hat{a}_1 \cdot 929983{,}301 + \hat{a}_2 \cdot 1272533{,}303 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}779;\ \hat{a}_2 = -0{,}066;\ \hat{\sigma}^2 = 590281{,}278.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}779 \cdot Y_{t-1} - 0{,}066 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 768{,}298^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 263{,}899 \cdot t - 523673{,}699 + 0{,}779 \cdot X_{t-1} - 0{,}066 \cdot X_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 768{,}298^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Elgesio ir emocijų sutrikimai, prasidedantys vaikystėje ir paauglystėje (7)

Metai	2006	2007	2008	2009	2010	2011	2012
Sergančiųjų skaičius	9862	9887	10031	10186	10429	10442	11014

Metai	2013	2014	2015	2016	2017	2018	2019
Sergančiųjų skaičius	10781	11407	11383	10816	11515	11145	11247

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+" ir „-" eilutes bei radome medianos reikšmę, kur

$$M_e = 10798{,}5$$

$$n = 14$$

$$\gamma(n) = 4 - bendras\ serij\text{ų}\ skai\check{c}ius$$

$$\tau(n) = 6 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$

$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$4 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14-1})\right] \Rightarrow 4 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1}t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 125{,}36 \cdot t - 241572$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (125{,}36 \cdot t - 241572)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$125{,}36 \cdot t - 241572$	$\boldsymbol{Y_t}$
2006	9862	1	-241446,640	251308,640
2007	9887	2	-241321,280	251208,280
2008	10031	3	-241195,920	251226,920
2009	10186	4	-241070,560	251256,560
2010	10429	5	-240945,200	251374,200
2011	10442	6	-240819,840	251261,840
2012	11014	7	-240694,480	251708,480
2013	10781	8	-240569,120	251350,120
2014	11407	9	-240443,760	251850,760
2015	11383	10	-240318,400	251701,400
2016	10816	11	-240193,040	251009,040
2017	11515	12	-240067,680	251582,680
2018	11145	13	-239942,320	251087,320
2019	11247	14	-239816,960	251063,960

Surasime AR(1) proceso lygtį sergančiųjų elgesio ir emocijų sutrikimais, prasidedančiais vaikystėje ir paauglystėje duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 251308{,}640;\ Y_2 = 251208{,}280;\ Y_3 = 251226{,}920;\ Y_4 = 251256{,}560;$$

$$Y_5 = 251374{,}200; Y_6 = 251261{,}840;\ Y_7 = 251708{,}480; Y_8 = 251350{,}120;\ Y_9 = 251850{,}760;$$

$$Y_{10} = 251701{,}400;\ Y_{11} = 251009{,}040;\ Y_{12} = 251582{,}680;\ Y_{13} = 251087{,}320;$$

$$Y_{14} = 251063{,}960.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 251356{,}443.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 62672{,}531 \\ \hat{\gamma}(1) = -927{,}605 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 62672{,}531 = \hat{a}_1 \cdot (-927{,}605) + \sigma^2 \\ -927{,}605 = \hat{a}_1 \cdot 62672{,}531 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = -0{,}015;\ \hat{\sigma}^2 = 62658{,}802.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = -0{,}015 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 250{,}317^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 127{,}240 \cdot t - 245195{,}580 - 0{,}015 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 250{,}317^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 251308{,}640;\ Y_2 = 251208{,}280;\ Y_3 = 251226{,}920;\ Y_4 = 251256{,}560;$$
$$Y_5 = 251374{,}200; Y_6 = 251261{,}840;\ Y_7 = 251708{,}480; Y_8 = 251350{,}120;\ Y_9 = 251850{,}760;$$
$$Y_{10} = 251701{,}400;\ Y_{11} = 251009{,}040;\ Y_{12} = 251582{,}680;\ Y_{13} = 251087{,}320;$$
$$Y_{14} = 251063{,}960.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 251356{,}443$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k}\left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 62672{,}531 \\ \hat{\gamma}(1) = -927{,}605 \\ \hat{\gamma}(2) = -866{,}169 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 62672{,}531 = \hat{a}_1 \cdot (-927{,}605) + \hat{a}_2 \cdot (-866{,}169) + \sigma^2 \\ -927{,}605 = \hat{a}_1 \cdot 62672{,}531 + \hat{a}_2 \cdot (-927{,}605) \\ -866{,}169 = \hat{a}_1 \cdot (-927{,}605) + \hat{a}_2 \cdot 62672{,}531 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = -0{,}015;\ \hat{a}_2 = -0{,}014;\ \hat{\sigma}^2 = 62646{,}446.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = -0{,}015 \cdot Y_{t-1} - 0{,}014 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 250{,}293^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 129{,}002 \cdot t - 248589{,}995 - 0{,}015 \cdot Y_{t-1} - 0{,}014 \cdot Y_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 250{,}293^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Bendras sergančiųjų psichikos ir elgesio sutrikimais (SPES) skaičius

Metai	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Sergančiųjų skaičius	175751	179620	187394	186370	190698	200288	207673

Metai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Sergančiųjų skaičius	209840	214119	211022	211896	222712	232124	235862

Patikrinsime, ar duomenys, kuriems galime taikyti laiko eilučių metodus, turi X_t trendą. Kadangi tikrinant hipotezę apie duomenų nepriklausomumą ir atsitiktinumą jau sudarėme „+” ir „-“ eilutes bei radome medianos reikšmę, kur

$$M_e = 208756{,}5$$
$$n = 14$$
$$\gamma(n) = 2 - bendras\ serij\text{ų}\ skai\check{c}ius$$
$$\tau(n) = 7 - ilgiausios\ serijos\ ilgis$$

Tikriname hipotezę

$$H_0: MX_t = a = const.$$

jei nors viena iš nelygybių

$$\gamma(n) > \left[\frac{1}{2}(n + 2 - 1{,}96\sqrt{n-1})\right]$$
$$\tau(n) < [1{,}43\ln(n+1)]$$

negalioja, tada hipotezė atmetama su tikimybe α, esančia tarp 0,05 ir 0,0975.

Į pirmąją nelygybę įsirašome:

$$n = 14$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$2 > \left[\frac{1}{2}(14 + 2 - 1{,}96\sqrt{14-1})\right] \Rightarrow 2 > 4$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

Rasime AR(1) ir AR(2) modelių parametrų įverčius, kai trendas eliminuojamas.

Darome prielaidą, kad trendas yra tiesinis:

$$f_t = \widehat{\beta_0} + \widehat{\beta_1} t$$

Programos MS Excel pagalba suskaičiuojame įverčius ir gauname lygtį:

$$f_t = 4397{,}39 \cdot t - 8645071{,}75$$

Eliminuojame trendą pagal formulę:

$$Y_t = X_t - (4397{,}39 \cdot t - 8645071{,}75)$$

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis:

Metai	X_t	t	$4397{,}39 \cdot t - 8645071{,}75$	$\boldsymbol{Y_t}$
2006	175751	1	-8640674,360	8816425,360
2007	179620	2	-8636276,970	8815896,970
2008	187394	3	-8631879,580	8819273,580
2009	186370	4	-8627482,190	8813852,190
2010	190698	5	-8623084,800	8813782,800
2011	200288	6	-8618687,410	8818975,410
2012	207673	7	-8614290,020	8821963,020
2013	209840	8	-8609892,630	8819732,630
2014	214119	9	-8605495,240	8819614,240
2015	211022	10	-8601097,850	8812119,850
2016	211896	11	-8596700,460	8808596,460
2017	222712	12	-8592303,070	8815015,070
2018	232124	13	-8587905,680	8820029,680
2019	235862	14	-8583508,290	8819370,290

Surasime AR(1) proceso lygtį bendram sergančiųjų psichikos ir elgesio sutrikimais (SPES) duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 8816425{,}360;\ Y_2 = 8815896{,}970;\ Y_3 = 8819273{,}580;\ Y_4 = 8813852{,}190;$$

$$Y_5 = 8813782{,}800; Y_6 = 8821963{,}020;\ Y_7 = 8821963{,}020; Y_8 = 8819732{,}630;$$

$$Y_9 = 8819614{,}240;$$

$$Y_{10} = 8812119{,}850;\ Y_{11} = 8808596{,}460;\ Y_{12} = 8815015{,}070;\ Y_{13} = 8820029{,}680;$$

$$Y_{14} = 8819370{,}290.$$

taip pat žinome lygties eilę – 1 (AR(1)).

Įvertinsime koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 8816760{,}539.$

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k} \left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 13012872{,}630 \\ \hat{\gamma}(1) = 5004386{,}597 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 13012872{,}630 = \hat{a}_1 \cdot 5004386{,}597 + \sigma^2 \\ 5004386{,}597 = \hat{a}_1 \cdot 13012872{,}630 \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį, gauname:

$$\hat{a}_1 = 0{,}385;\ \hat{\sigma}^2 = 11088325{,}612.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0{,}385 \cdot Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 3329{,}914^2)$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 2704{,}395 \cdot t - 5316719{,}126 + 0{,}385 \cdot X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 3329{,}914^2).$$

Naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.

Surasime AR(2) proceso lygtį sergančiųjų demencija ir Alzheimerio liga duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi:

$$Y_t = \hat{a}_1 \cdot Y_{t-1} + \hat{a}_2 \cdot Y_{t-2} + \varepsilon_t$$

ir stebinius:

$$Y_1 = 8816425{,}360;\ Y_2 = 8815896{,}970;\ Y_3 = 8819273{,}580;\ Y_4 = 8813852{,}190;$$
$$Y_5 = 8813782{,}800;\ Y_6 = 8821963{,}020;\ Y_7 = 8821963{,}020;\ Y_8 = 8819732{,}630;$$
$$Y_9 = 8819614{,}240;$$
$$Y_{10} = 8812119{,}850;\ Y_{11} = 8808596{,}460;\ Y_{12} = 8815015{,}070;\ Y_{13} = 8820029{,}680;$$
$$Y_{14} = 8819370{,}290.$$

taip pat žinome lygties eilę – 2 (AR(2)).

Įvertinsime koeficientus $\hat{a}_1, \hat{a}_2$ ir $\hat{\sigma}^2$. Stebėjimų skaičius: $n = 14$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 8816760{,}539$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n} \cdot \sum_{j=1}^{n-k} \left(Y_{j+k} - \bar{Y}\right) \cdot (Y_j - \bar{Y})$ su programos MS Excel pagalba, apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 13012872{,}630 \\ \hat{\gamma}(1) = 5004386{,}597 \\ \hat{\gamma}(2) = 2426894{,}054 \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 5640603{,}305 = \hat{a}_1 \cdot 5004386{,}597 + \hat{a}_2 \cdot 2426894{,}054 + \sigma^2 \\ 5004386{,}597 = \hat{a}_1 \cdot 5640603{,}305 + \hat{a}_2 \cdot 5004386{,}597 \\ 2426894{,}054 = \hat{a}_1 \cdot 5004386{,}597 + \hat{a}_2 \cdot 5640603{,}305 \end{cases}$$

Iš antrosios lygties išsireiškiame $\hat{a}_2$ įvertį, įstatome į trečiąją lygtį ir gauname $\hat{a}_1$ bei $\hat{a}_2$ įverčius. Šiuos įverčius įstatome į pirmąją lygtį ir gauname $\hat{\sigma}^2$ įvertį, Excel programos pagalba gauname tokias reiškmes:

$$\hat{a}_1 = 0{,}367; \hat{a}_2 = 0{,}045; \hat{\sigma}^2 = 11065567{,}213.$$

Taigi, gauname AR(2) proceso lygtį:

$$Y_t = 0{,}367 \cdot Y_{t-1} + 0{,}045 \cdot Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0; 3326{,}495^2)$$

Gautą AR(2) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 2583{,}671 \cdot t - 5079382{,}118 + 0{,}367 \cdot Y_{t-1} + 0{,}045 \cdot Y_{t-2} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 3326{,}495^2).$$

Naudodami gautą AR(2) proceso lygtį galime su nedidele paklaida prognozuoti sergančiųjų skaičius.