

LIETUVOS TEISĖS UNIVERSITETAS

Vitalija Rudzkienė

**STATISTINĖS
TECHNOLOGIJOS
TEISĖJE IR VALDYME**

Praktiniai darbai

Mokomasis leidinys

Vilnius
2003

UDK 004:311(075.8)

Ru44

Recenzavo:

Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakulteto Matematinės statistikos katedros vedėjas prof. habil. dr. **Leonas Saulis**;

Lietuvos teisės universiteto Valstybinio valdymo fakulteto Teisinės informatikos katedros vedėjas prof. dr. **Rimantas Petrauskas**

Lietuvos teisės universiteto Valstybinio valdymo fakulteto Viešojo administravimo katedros 2002 m. balandžio 22 d. posėdžio sprendimu (protokolas Nr. 9) leidinys rekomenduotas spausdinti

Lietuvos teisės universiteto vadovėlių, monografijų, mokslinių, mokomųjų, metodinių bei kitų leidinių aprobavimo spaudai komisija 2002 m. rugsėjo 30 d. posėdyje (protokolas Nr. 4) leidinį patvirtino spausdinti

TURINYS

Įvadas.....	4
1. Teisinė–kriminalinė statistika.....	6
1.1. pavyzdys. Padėties ir sklaidos charakteristikos: Lietuvos apskrityse užregistruotų nusikaltimų skaičius.....	6
1.2 pavyzdys. Užregistruotų nusikaltimų laiko eilučių analizė.....	11
1.3 pavyzdys. Poriniai stebėjimai: asmenų, kaltinamų padarius nusikaltimus, užimtumo struktūros analizė (tiesinė regresija). Uždaviniai.....	17 25
2. Užimtumo ir nedarbo statistika.....	30
2.1 pavyzdys. Gyventojų užimtumo tyrimas: pagrindiniai užimtumo ir nedarbo rodikliai.....	32
2.2 pavyzdys. Užimtųjų pasiskirstymas Lietuvoje pagal išsimokslinimą.....	33
2.3 pavyzdys. Rodiklių priklausomumo tikrinimas.....	35
2.4 pavyzdys. Darbo patirties ir įsidarbinimo priklausomumas.....	38
2.5 pavyzdys. Bedarbių šeiminės padėties analizė..... Uždaviniai.....	40 42
3. Statistinė gyvenimo lygio analizė: visuotinė gerovė ir nelygybė....	45
3.1. Gyventojų ir namų ūkių pajamų ir išlaidų statistika.....	46
3.1 pavyzdys. Decilinis ir kvartilinis diferenciacijos koeficientai, Lorenzo kreivė, Lorenzo ir Džini koeficientai.....	50 55
Uždaviniai.....	55
4. Integruoti visuotinės gerovės rodikliai.....	59
4.1 pavyzdys. Socialinės pažangos indeksas.....	63
4.2 pavyzdys. Besivystančių šalių skurdo indeksas.....	65
4.3 pavyzdys. Išsivysčiusių šalių skurdo indeksas..... Uždaviniai.....	66 66
Literatūra.....	73
Priedai.....	74

IVADAS

Atsisakius populiaraus XIX a. termino *moralinė statistika*, kurios tyrimų objektas buvo nusikaltimai, bausmės, santuokos, skyrybos, gimimai, abortai, venerinės ligos, alkoholizmas, narkomanija ir t. t., ši statistika buvo priskirta socialinės statistikos sričiai. Terminas *socialinė statistika* taip pat neturi griežto ir visuotinai pripažinto apibrėžimo. Socialine sritimi paprastai vadinama veikla, sudaranti sąlygas asmenybės raidai bei veiklai visuomenėje. Sąlyginai socialinę statistiką galima suskirstyti į teisinę–kriminalinę statistiką (nusikaltimai, jų priežastys, asmenys, padarę nusikaltimus, nusikaltimų aukos, teismų statistika ir pan.), demografinę statistiką (gyventojai, gyvenimo trukmė, be tėvų priežiūros augantys vaikai, bedarbiai ir t. t.), sveikatos apsaugos statistiką (gydymo įstaigos, personalas, įvairios ligos, įskaitant narkomaniją, alkoholizmą, venerines ligas ir kt.), politikos bei mokslo statistiką. Demografinė ir sveikatos apsaugos statistika parodo tarsi visuomenės foną, turintį didelės įtakos kriminalinei statistikai.

Šiuolaikinėje visuomenėje informacija, jos analizė ir socialinių–ekonominių reiškinių numatymas tapo vienu iš svarbiausių strateginių uždavinių, leidžiančių kurti veiksmingas valdymo sistemas. Šiuo tikslu sukurtas platus pasaulinis socialinės, ekonominės ir teisinės statistinės informacijos tinklas. Demokratinės visuomenės raida pagrįsta egzistuojančios tikrovės pažinimu, kuris neįmanomas be statistinių bei socialinių metodų. Egzistuojančią tikrovę galima pažinti tik stebint tikrovės faktus bei procesus. Tuo tarpu kiekviena mokslo sritis, tarp jų ir socialiniai mokslai, kuria savo atskirą tikrovę (jurisprudencijoje tai yra įstatymai, teisės aktai, nutarimai, teismų nutartys ir kt.). Ši abstraktaus mąstymo sukurta tikrovė ne visada tiksliai atspindi egzistuojančią tikrovę ir net bando ją pakeisti sukurtomis loginėmis konstrukcijomis. Egzistuojančiam pasauliui pažinti neužtenka loginio mąstymo. Tik empirinių stebėjimų gausa kuria sąlygas atskleisti dėsningumus, kitimo ypatybes ir kitokias žinias, būtinas šiuolaikinės visuomenės santykiams tobulinti. Statistikos metodai, paremti empiriniais stebėjimais, padeda suprasti ir numatyti socialinius ir teisinius reiškinius, analizuoti jų išplitimą ir dinamiką taip gaunant informacijos, reikalingos santykiams visuomenėje tobulinti. Socialinės statistikos uždavinių sprendimas pagrįstas aprašomosios ir matematinės statistikos metodais – grafiniu stebėjimų vaizda-

vimu, dinamikos eilutėmis, duomenų padėties ir sklaidos charakteristikomis, koreliacija, hipotezių tikrinimu, regresine analize. Uždaviniuose, skirtuose pokyčiams socialinėje aplinkoje įvertinti ir prognozuoti, labai svarbi griežta statistinių rezultatų interpretacija.

Pateikiamuose praktiniuose darbuose dažniausiai nagrinėjami tikri duomenys, pasitelkiant pavyzdžius aiškinami įvairių socialinių rodiklių skaičiavimo metodai, statistinių metodų ir modelių taikymas socialiniams reiškiniams analizuoti ir prognozuoti. Uždaviniams spręsti naudojamos šiuolaikinės informacijos technologijos priemonės – kompiuteriai ir taikomoji programinė įranga. Tinkamų priemonių statistinei analizei atlikti yra gana daug. Galima naudoti sistemas EXCEL, STATISTICA, SPSS, STATGRAPHICS bei kitas programinės informacijos technologijos priemones, tinkančias duomenų statistinei analizei atlikti.

Šis mokamasis leidinys skirtas teisės, teisės ir valdymo bei viešojo administravimo specialybių studentams. Jo tikslas – padėti jaunam specialistui suprasti ir numatyti egzistuojančią tikrovę, socialinius ir teisinius reiškinius ir praktiškai panaudoti šias žinias tobulinant rinkos ekonomika pagrįstus demokratinės visuomenės santykius.



1. TEISINĖ–KRIMINALINĖ STATISTIKA

Teisinė statistika nagrinėja įvairių teisės sričių (kriminalinės, civilinės, administracinės, šeimos teisės ir kt.) rodiklius. Ypač plačiai statistikos metodai taikomi kriminologijoje atskleidžiant tiriamų procesų ypatybes, tendencijas ir raidos dėsningumus, nulemtus konkrečių laiko ir erdvės sąlygų. Šio pobūdžio uždaviniuose dažniausiai tiriamas nusikalstamumo – bendro arba pagal nusikaltimų rūšis – kitimas.

Uždaviniuose naudojami Lietuvos statistikos departamento ir Lietuvos nusikalstamumo prevencijos centro duomenys.

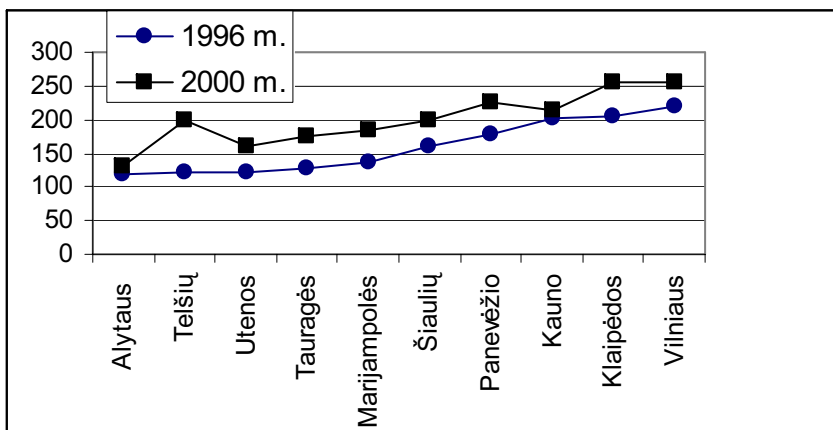
Tipinių uždavinių sprendimą nagrinėsime toliau pateiktuose pavyzdžiuose.

1.1 pavyzdys. Padėties ir sklaidos charakteristikos: Lietuvos apskrityse užregistruotų nusikaltimų skaičius (žr. 1.1 lentelę).

1.1 lentelė. Lietuvos apskrityse užregistruotų nusikaltimų ir gyventojų skaičius

Apskritys	Nusikaltimų sk. 10 000 gyventojų		Gyventojų skaičius	
	1996 m. (x_1)	2000 m. (x_2)	1997 m. pradžioje (f_1)	2001 m. pradžioje (f_2)
Alytaus	119	130	202400	202000
Kauno	202	215	757000	749000
Klaipėdos	206	256	415400	402800
Marijampolės	137	183	198500	198000
Panevėžio	179	227	323300	320100
Šiaulių	160	198	401900	394300
Tauragės	127	175	130200	142400
Telšių	121	199	182600	188700
Utenos	121	160	202400	199400
Vilniaus	219	255	895400	895900

1.1.1 užduotis. Apskaičiuokite kiekvienų metų duomenų padėties ir sklaidos rodiklius: aritmetinį vidurkį, medianą, modą, duomenų aibės plotį, dispersiją, standartinį nuokrypį ir variacijos (kitimo) koeficientą. Skaičiavimų rezultatus pateikite lentelės pavidalu ir apskaičiuokite 2000 metų rodiklių kitimo indeksus lygindami su 1996 metais (žr. 1.3 lentelę). Sudarykite diagramą horizontalioje ašyje atidėdami apskritis, išrikiuotas pagal nusikalstamumo lygį 1996 metais. Vertikaliajoje ašyje atidėkite apskrityse užregistruotų nusikaltimų skaičių 10 000 gyventojų. Taškus sujunkite (žr. 1.1 pav.)



1.1 pav. Lietuvos apskrityse užregistruotų nusikaltimų skaičius 10 000 gyventojų 1996 ir 2000 metais

Skaičiuojant taikomos šios formulės:

- *Negrupuotiems duomenims:*

Mediana yra vidurinė reikšmė. Ją nustatome iš **didėjimo tvarka** surašytų skaičių (variacinės eilutės x_1, x_2, \dots, x_n) išrinkę vidurinį skaičių. Esant lyginiam reikšmių skaičiui dvi vidurinės reikšmės sudedamos ir padalijamos pusiau:

$$Md = \begin{cases} x_{(n+1)/2}, & \text{kai } n \text{ nelyginis,} \\ \frac{x_{n/2} + x_{n/2+1}}{2}, & \text{kai } n \text{ lyginis.} \end{cases} \quad (1.1)$$

Moda – tai duomenų aibėje dažniausiai pasikartojanti reikšmė.

Duomenų aibės plotis $R = x_{\max} - x_{\min}$.

Aritmetinis vidurkis yra visų eilutės elementų suma, padalyta iš jų skaičiaus:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (1.2)$$

Dispersija rodo duomenų sklaidą apie vidurkį:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}. \quad (1.3)$$

- *Grupuotiems duomenims* (t. y. duomenims, suskirstytiems į intervalus):

Paprasčiausias būdas rasti grupuotų duomenų *medianą* būtų visas į intervalą patekusias reikšmes pakeisti vidurinėmis intervalo reikšmėmis ir pritaikyti medianos skaičiavimo formulę (1.1).

Grupuotų duomenų *moda* apytiksliai lygi intervalo, į kurį pateko daugiausia duomenų, vidurinei reikšmei.

Aritmetinis vidurkis

$$\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i}, \quad (1.4)$$

čia f_i – grupės elementų skaičius, m – grupių skaičius.

Dispersija

$$S^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum_{i=1}^m f_i}. \quad (1.5)$$

Standartinis nuokrypis yra kvadratinė šaknis, ištraukta iš dispersijos

$$S = \sqrt{S^2}. \quad (1.6)$$

Variacijos (kitimo) koeficientas naudojamas lyginant skirtingų duomenų aibių sklaidas:

$$CV = \frac{S}{\bar{x}}. \quad (1.7)$$

Indeksas – tai santykinis dydis, gautas palyginus to paties rodiklio (paprasto arba sudėtingo socialinio–ekonominio rodiklio) lygius laike, erdvėje arba su planuojamomis ir numatomomis rodiklio reikšmėmis:

$$i = \frac{p_1}{p_0}. \quad (1.8)$$

Rodiklių skaičiavimas

Norėdami rasti vidutinį užregistruotų nusikaltimų skaičių negalime sudėti kiekvienoje apskrityje užregistruotų nusikaltimų skaičiaus 10 000 gyventojų ir padalyti iš apskričių skaičiaus, nes gyventojų skaičius apskrityse yra skirtingas. Skaičiuodami taikysime grupuotiems duomenims skirtas formules. Tarpiniai skaičiavimo rezultatai pateikti lentelės pavidalu (žr. 1.2 lentelę).

Apskaičiuosime rodiklių reikšmes (žr. 1.3 lentelę):

- Lietuvoje užregistruotas vidutinis nusikaltimų skaičius 10 000 gyventojų (visose apskrityse)
1996 metais:

$$\bar{x}_1 = \sum_{i=1}^{10} x_i f_i / \sum_{i=1}^{10} f_i = 671154200 / 3709100 = 180,9;$$

2000 metais: $\bar{x}_2 = 800209700/3692600 = 216,7$.

1.2 lentelė. Tarpiniai skaičiavimo rezultatai

Grupė	(x ₁)	(x ₂)	(f ₁)	(f ₂)	x ₁ *f ₁	x ₂ *f ₂	(x ₁ - \bar{x}) ² *f ₁	(x ₂ - \bar{x}) ² *f ₂
1	119	130	202400	202000	24085600	26260000	776720891	1518632615
2	202	215	757000	749000	152914000	161035000	335492565	2180694.812
3	206	256	415400	402800	85572400	103116800	260706314	621920987
4	137	183	198500	198000	27194500	36234000	383388075	224950762.7
5	179	227	323300	320100	57870700	72662700	1226818.7	33917849.47
6	160	198	401900	394300	64304000	78071400	176361115	137975754
7	127	175	130200	142400	16535400	24920000	378932247	247692814.9
8	121	199	182600	188700	22094600	37551300	656220912	59159944.44
9	121	160	202400	199400	24490400	31904000	727377396	641191631
10	219	255	895400	895900	196092600	228454500	1296499134	1313754134
Suma			3709100	3692600	671154200	800209700	4992925468	4801377187

- Duomenų aibės plotis

1996 metais: $R_1 = x_{\max} - x_{\min} = 219 - 119 = 100$;

2000 metais: $R_2 = 256 - 130 = 126$.

- Dispersija

1996 metais:

$$S_1^2 = \sum_{i=1}^{10} (x_i - \bar{x})^2 f_i / \sum_{i=1}^{10} f_i = 4992925467 / 3709100 = 1346;$$

2000 metais: $S_2^2 = 4801377187 / 3692600 = 1300$.

- Standartinis nuokrypis:

1996 metais: $S_1 = \sqrt{S^2} = 36,7$;

2000 metais: $S_2 = 36,10$.

- Variacijos koeficientas

1996 metais: $CV_1 = S / \bar{x} = 36,7 / 180,9 = 0,203$, arba 20,3 proc.

2000 metais: $CV_2 = 36,1 / 216,7 = 0,166$, arba 16,6 proc.

1.3 lentelė. Apskaičiuoti padėties ir sklaidos rodikliai

Metai	\bar{x}	R	S ²	S	CV
1996	180,9	100	1346	36,7	0,23
2000	216,7	126	1300	36,1	0,166
$i_{2000/1996}$	1,2	1,26	0,97	0,98	0,82

1.1.2 užduotis. Parašykite trumpą tyrimo analizę pasinaudodami gautais skaičiavimų rezultatais ir diagrama.

Nuo 1996 iki 2000 metų registruotas nusikalstamumas Lietuvoje padidėjo 20 procentų ($i_{\bar{x}_2/\bar{x}_1}=1,2$). Nepaisant didėjančio nusikaltimų skaičiaus, per šį laikotarpį skirtumai tarp apskričių nepadidėjo, netgi priešingai, sprendžiant pagal informatyviausią palygimo rodiklį – variacijos koeficientą – sumažėjo nuo 0,203 iki 0,166, o tai sudaro 82 procentus pradinio lygmens ($i_{CV_2/CV_1}=0,82$). Iš gautų kitimo rodiklių reikšmių (duomenų aibės plotis, t. y. skirtumas tarp maksimalios ir minimalios reikšmės, padidėjo 26 procentais ($i_{R_2/R_1}=1,26$), dispersija beveik nepakito, o variacija sumažėjo 18 procentų) matome, kad atlikdami sudėtingesnius ir daugiau pastangų reikalaujančius skaičiavimus gauname tikslesnius rezultatus ir lengviau suformuluojame pagrįstas išvadas.

Per šį laikotarpį užregistruotų nusikalstamų skaičius apskrityse kito gana tolygiai. Iš bendro kitimo pobūdžio išsiskiria trys apskritys: Telšių, kurioje užregistruotų nusikaltimų padidėjo daugiausia, ir Alytaus bei Kauno, kuriose užregistruotų nusikaltimų skaičius padidėjo nedaug. Aukščiausias nusikalstamumo lygis išliko Klaipėdos ir Vilniaus apskrityse. Galime manyti, kad socialinės–ekonominės sąlygos, sudarančios terpę nusikalstamumui, per šį laikotarpį tarp apskričių pasikeitė nedaug.

1.2 pavyzdys. Užregistruotų nusikaltimų laiko eilučių analizė

Pradiniai duomenys – nepilnamečių padaryti arba jiems dalyvaujant padaryti nusikaltimai (žr. 1.4 lentelę).

1.4 lentelė. **Nepilnamečių arba jiems dalyvaujant padaryti nusikaltimai ir bendras užregistruotų nusikaltimų skaičius**

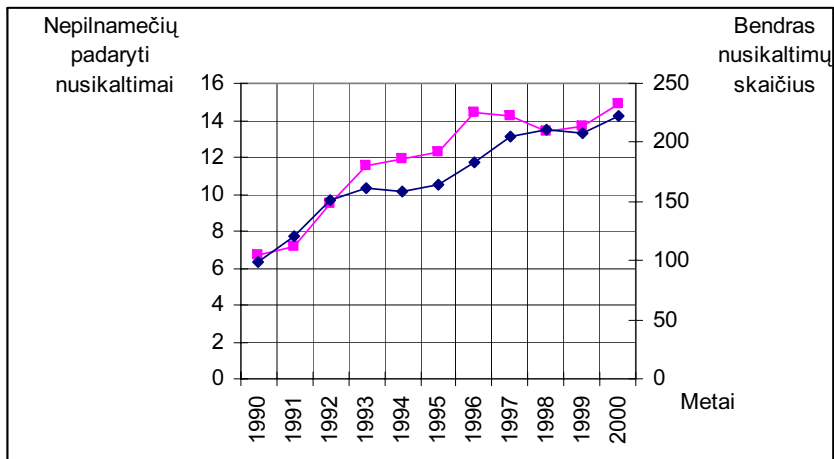
Metai	Išnagrinėta nusikaltimų, padarytų nepilnamečių arba padarytų jiems dalyvaujant (10 000 gyventojų)	Bendras užregistruotų nusikaltimų skaičius (10 000 gyventojų)
1990	6,7	99,2
1991	7,2	120
1992	9,5	150,5
1993	11,5	162
1994	11,9	158
1995	12,3	163,6
1996	14,4	183,3
1997	14,2	204,5
1998	13,4	210,9
1999	13,7	208,3
2000	14,9	222

1.2.1 užduotis. Nusikalstamumo lygio kitimą pavaizduokite grafiškai ir pateikite pradinę analizę. Apskaičiuokite apibendrinančius rodiklius.

Analizę tikslinga pradėti bandymu suvokti bendras kitimo tendencijas. Kiekvienoje laiko eilutėje yra du pagrindiniai elementai: laiko skalė t ir eilutės lygis, paprastai žymimas y raide. Norint atlikti laiko eilutės pradinę analizę būtina nupiešti diagramą. Jei diagramą sudarysime tiesiogiai iš duomenų, pateiktų 1.4 lentelėje, diagrama bus neinformatyvi, nes duomenų masteliai skiriasi eile. Masteliams sulygininti naudojamos dviejų tipų diagramos: 1) su dviem vertikaliomis ašimis sekų reikšmes atidedant atskirose ašyse (1.2 pav.); 2) abiejų sekų pradinį lygį prilyginant 100 ir apskaičiuojant rodiklių kitimo greitį (1.3 pav.).

Kitoje diagramoje (žr. 1.3 pav.) abiejų rodiklių lygis 1990 metais prilygintas 100 ir apskaičiuotas rodiklių didėjimo greitis. Didėjimo greitis apskaičiuojamas dviem būdais: skaičiuojant pirmuoju būdu nagrinėjamas lygis lyginamas su pradiniu lygiu (y_0), o nauja eilutė (santykinis didėjimo greitis) apskaičiuojama taip:

$$S_1 = \frac{y_n}{y_0} \cdot 100. \quad (1.9)$$



1.2 pav. Laiko eilutės diagrama su dviem skirtingomis vertikaliomis ašimis. X ašyje žymimas laikas $t=1,2,\dots,11$.

Čia santykinis didėjimo greitis išreikštas procentais.

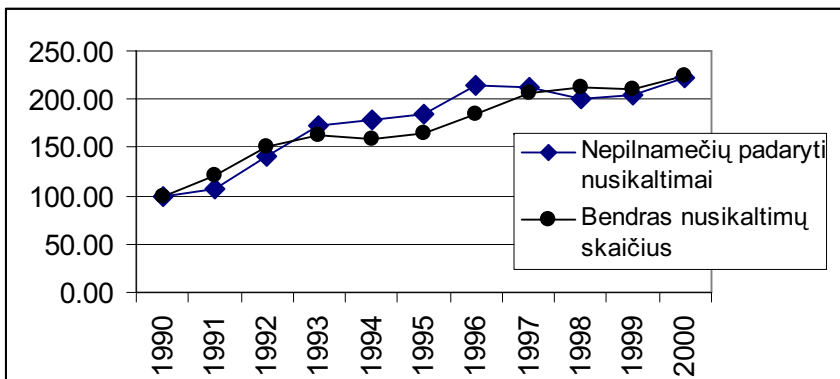
Didėjimo greitį apskaičiuojant antruoju būdu nagrinėjamas lygis lyginamas su prieš jį esančiu kintančiu lygiu ir šiuo atveju didėjimo greitis yra grandininis:

$$S_2 = \frac{y_n}{y_{n-1}} \cdot 100. \quad (1.10)$$

Pastaba: jei šių formulių dešiniųjų pusių nedauginsime iš 100, didėjimo greitis bus išreikštas ne procentais, o kitimo koeficientais.

1.3 pav. diagramoje matome pagal formulę (1.9) apskaičiuotą nusikalstamumo lygio santykinio didėjimo greičio kitimą. Diagramoje galima išskirti kelis nusikalstamumo lygio kitimo laikotarpius. Nusikalstamumo didėjimas, apie kurį sprendžiame iš abiejų rodiklių kitimo, stabilizavosi 1993 metais ir nedaug kito iki 1995 metų. Nuo 1995 metų abu nusikalstamumo rodikliai vėl didėjo, o 1997–1998 metais – mažėjo. Nuo 1998 metų rodikliai vėl šiek tiek didėjo. Pagrindinė diagramos tyrimo išvada yra ta, kad tolesnei analizei atlikti reikia atskirai ištirti 1990–1992, 1993–1995 ir 1996–2000 metų laikotarpius. Šias išvadas

buvo galima padaryti ir nagrinėjant pradinių duomenų lentelę, tačiau iš jos gerokai sudėtingiau išskirti pagrindines kitimo tendencijas.



1.3 pav. Nusikalstamumo lygio santykinio didėjimo greičio kitimas

Kiekybinėms kitimo tendencijoms įvertinti būtina apskaičiuoti šiuos rodiklius:

- vidutinį laiko eilutės lygį;
- kasmetinį vidutinį absoliutų padidėjimą (sumažėjimą);
- vidutinį santykinį didėjimo (mažėjimo) greitį.

Bendros kitimo tendencijos dažnai išryškėja analitiškai išlyginant laiko eilutes, tačiau šiuo atveju to nedarysime, nes kiekvieno atskiro laikotarpio metų nėra daug.

Rodikliams apskaičiuoti taikysime šias formules:

- intervalinių laiko eilučių vidutinis laiko eilutės lygis apskaičiuojamas pagal aritmetinio vidurkio formulę;
- vidutinis metinis padidėjimas (sumažėjimas) rodo, keliais vienetais vidutiniškai pasikeičia reiškinio lygis per metus. Jis apskaičiuojamas pagal formulę:

$$\bar{d} = \frac{y_m - y_1}{m} . \quad (1.11)$$

Šiuo atveju y_m – galutinė laikotarpio reikšmė, y_1 – pradinė reikšmė.

- vidutinio santykinio didėjimo (mažėjimo) greičiui apskaičiuoti taikoma geometrinio vidurkio formulė. Geometrinis vidurkis gaunamas ištraukiant m -tojo laipsnio šaknį iš m skaičių sandaugos $\bar{k} = \sqrt[m]{x^m}$. Pritaikius šią formulę vidutinio santykinio didėjimo (mažėjimo) greičiui skaičiuoti, gaunama:

$$\bar{k} = m \sqrt[m]{\prod_{i=1}^m \frac{y_m}{y_{m-1}}}. \quad (1.12)$$

Reikia pasakyti, kad skaičiuojant geometrinį vidurkį pirmajai laiko periodo reikšmei suteikiamas nulinis numeris, todėl šaknies eilė yra vietu mažesnė nei reikšmių skaičius.

Apskaičiuosime šių rodiklių reikšmes pirmuoju laikotarpiu, t. y. 1990–1992 metais.

- 1) nepilnamečių padaryti nusikaltimai 10 000 gyventojų:

$$\bar{y} = (6,7 + 7,2 + 9,5) / 3 = 23,4 / 3 = 7,8$$

$$\bar{d} = (9,5 - 6,7) / 2 = 1,4$$

$$\bar{k} = 2 \sqrt{\frac{7,2}{6,7} \cdot \frac{9,5}{7,2}} = 1,19.$$

- 2) bendras užregistruotų nusikaltimų skaičius:

$$\bar{y} = (99,2 + 120 + 150,5) / 3 = 369,7 / 3 = 123,23$$

$$\bar{d} = (150,5 - 99,2) / 2 = 25,65$$

$$\bar{k} = 2 \sqrt{\frac{120}{99,2} \cdot \frac{150,5}{120}} = 1,23.$$

Apskaičiuotos šių rodiklių reikšmės atskirais laikotarpiais pateiktos 1.5 lentelėje.

1.5 lentelė. Vidutinis laiko eilutės lygis, vidutinis metinis padidėjimas, vidutinis santykinis didėjimo greitis atskirais laikotarpiais

Laikotarpis	Nepilnamečių padaryti nusikaltimai 10 000 gyv.			Bendras užregistruotų nusikaltimų skaičius 10 000 gyventojų		
	\bar{y}	\bar{d}	\bar{k}	\bar{y}	\bar{d}	\bar{k}
1990–1992 m.	7,80	1,40	1,19	123,23	25,65	1,23
1993–1995 m.	11,9	0,4	1,03	161,2	0,8	1,00
1996–2000 m.	14,12	0,125	1,01	205,8	9,68	1,05

1.2.2 užduotis. Parašykite trumpą gautų tyrimo rezultatų analizę.

Iš gautų rezultatų ir diagramos galime padaryti tokias išvadas: pirmajam laikotarpiui būdingas greitas nusikalstamumo didėjimas: kasmet 23 procentais padidėdavo bendras užregistruotų nusikaltimų skaičius 10 000 gyventojų ($\bar{k}=1,23$). Išnagrinėtas nusikaltimų, padarytų nepilnamečių arba padarytų jiems dalyvaujant, skaičius 10 000 gyventojų kasmet padidėdavo 19 procentų ($\bar{k}=1,19$). Antruoju laikotarpiu ir bendras nusikalstamumo lygis, ir nepilnamečių nusikalstamumas išliko beveik nepakitę (atitinkamai $\bar{k}=1,0$ ir $\bar{k}=1,03$). Trečiuoju laikotarpiu nusikalstamumas vėl didėjo: 5 procentais kasmet didėjo bendras užregistruotų nusikaltimų skaičius, o nepilnamečių nusikalstamumas didėjo ne daug, kasmet vidutiniškai po vieną procentą.

Lyginami diagramos kreives pastebime, kad kitimo tendencijos yra panašios. Nepilnamečių nusikalstamumas, sparčiai didėjęs iki 1996 metų, vėliau nustojo didėti ir vėl pradėjo didėti 2000 metais. Tačiau pagal vienerių metų duomenis sunku spręsti, ar tai naujos kitimo tendencijos pradžia, ar tik trumpalaikis nukrypimas nuo bendros kitimo tendencijos.

Reikia pasakyti, kad nepilnamečių nusikalstamumo kitimo tendencijos panašios į suaugusiųjų nusikalstamumo tendencijas, tačiau silpniau išreikštos. Nagrinėdami vidutinį metinį padidėjimą (\bar{d}), t. y. kiek vienetų vidutiniškai pasikeičia reiškinio lygis per laiko vienetą, matome, kad, lyginant su bendru nusikalstamumu, nepilnamečių nusikalstamumas padidėjo daug mažiau, tačiau tikslesnis ir informatyvesnis yra vi-

duotinis santykinis didėjimo (mažėjimo) greitis (\bar{k}), išreiškiamas procentais arba matavimo koeficientu.

1.3 pavyzdys. Poriniai stebėjimai: asmenų, kaltinamų padarius nusikaltimus, užimtumo struktūros analizė (tiesinė regresija)

Pradiniai duomenys – asmenys, kaltinami padarę nusikaltimus, ir jų užimtumas (žr. 1.6 lentelę).

1.6 lentelė. Asmenys, kaltinami padarę nusikaltimus, ir jų užimtumas

Metai	Nustatyta asmenų, kaltinamų padarius nusikaltimus (x)	Iš jų mokėsi (y1)	Iš jų dirbo (y2)	Nedirbo ir nesimokė (y3)	Kita
1990	12556	1382	6964	2808	1402
1991	13268	1055	6978	3841	1394
1992	18810	1231	8257	7300	2022
1993	20424	1182	6026	10909	2301
1994	21290	952	5068	12854	2416
1995	22969	1341	4266	14182	3180
1996	22269	1724	2815	14066	3664
1997	25542	1979	2283	16146	5134
1998	25373	2455	1765	15920	6998
1999	25160	2761	1595	15900	5210
2000	25046	3156	1304	16653	3933

1.3.1 užduotis. Įvertinkite nagrinėjamų rodiklių koreliacinę priklausomybę.

Šiame pavyzdyje nagrinėsime, kokios priežastys formuoja tiriamo rodiklio lygį ir kokia konkrečios priežasties įtaka. Būtent rodiklio kitimo priklausomybės nuo aplinkos sąlygų tyrimas sudaro koreliacijos teorijos turinį. Tiriant priklausomybes tarp rodiklių išskiriamos dviejų tipų priklausomybės: 1) funkcinės ir 2) koreliacinės.

Funkcinėms (determinuotoms) priklausomybėms būdingas visiškas faktorinio rodiklio (t. y. rodiklio, lemiančio kitų rodiklių kitimą) ir veikiamo rodiklio (t. y. rodiklio, atspindinčio faktorinio rodiklio veikimą) atitikimas. Esant funkicinei priklausomybei kiekvieną faktorinio rodiklio reikšmę atitinka visiškai apibrėžta veikiamo rodiklio reikšmė.

Esant koreliacinei (stochastinei) priklausomybei faktorinis ir veikiamas rodikliai nevysiškai vienas kitą atitinka, atskirų faktorių įtaka išryškėja tik gana didelės apimties duomenų vidutinėse reikšmėse. Paprastai faktorinis rodiklis nėra vienintelė priežastis, lemianti kitų rodiklių kitimą. Kita vertus, faktorinis rodiklis dažnai priklauso nuo įvairių aplinkybių kitimo.

Tiriant koreliacines–regresines rodiklių priklausomybes (apie regresiją skaitykite 3.3 užduotyje) reikia išspręsti šiuos uždavinius:

- atlikti pradinę duomenų analizę ir nustatyti, ar yra priklausomybė tarp rodiklių;
- nustatyti priklausomybės stiprumą;
- sudaryti regresinį modelį;
- įvertinti modelio tinkamumą.

Koreliacinio ryšių buvimui ar nebuvimui nustatyti taikomi įvairūs metodai, iš kurių pats paprasčiausias – grafinis dviejų sekų lyginimas. Tam tikslui faktorinis rodiklis **išdėstomas didėjimo tvarka** ir pagal jį tiriamas, kaip keičiasi kitų rodiklių (ar rodiklio) reikšmės (žr. 1.4 pav.).

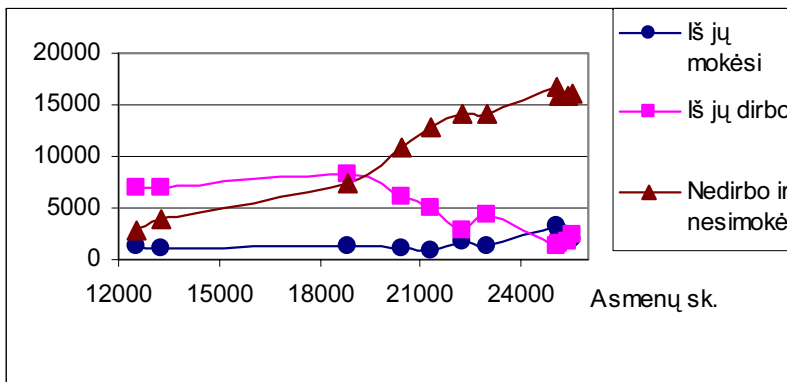
Remdamiesi diagramos (žr. 1.4 pav.) ir 1.6 lentelės duomenimis matome, kad didėjant asmenų, kaltinamų padarius nusikaltimus, skaičiui, tarp jų daugėja asmenų, kurie nedirbo ir nesimokė, ir mažėja dirbančių asmenų. Besimokančių asmenų skaičius kito nedaug.

Tais atvejais, kai didėjant faktorinio rodiklio reikšmėms didėja veikiamų rodiklių reikšmės, koreliacinis ryšys yra tiesioginis. Jei faktoriui didėjant veikiamo rodiklio reikšmės mažėja, ryšys yra atvirkštinis.

Kai diagramoje nubrėžtos linija forma artima tiesei, galima manyti tarp rodiklių egzistuojant tiesinį koreliacinį ryšį. Tuo atveju, kai kitimas netolygus ir linija artima kokiam nors kreivės tipui, koreliacinis ryšys gali būti netiesinis.

Diagramoje (žr. 1.4 pav.) matome, kad asmenų, kaltinamų padarius nusikaltimus, koreliacinis ryšys su asmenimis, kurie nedirbo ir nesimokė, yra tiesioginis, o su dirbusiais asmenimis – atvirkštinis. Besimo-

kančių asmenų skaičius kito mažiau, ir tai leidžia manyti, kad šiuo atveju koreliacinis ryšys yra silpnėsnis.



1.4 pav. Asmenų, kaltinamų padarius nusikaltimus, užimtumas

1.3.2 užduotis. Nustatykite porinių stebėjimų koreliacinio ryšio stiprumo laipsnį.

Koreliacinio ryšio stiprumo rodikliai sudaro galimybę įvertinti veikiamo rodiklio priklausomybę nuo faktorinio rodiklio. Žinant šiuos rodiklius galima išspręsti tokius klausimus:

- lyginant šių rodiklių reikšmes įvairiose situacijose galima spręsti apie skirtumų lygį ir kaip šis lygis kinta atsižvelgiant į konkrečias sąlygas;
- lyginant atskirų rodiklių ryšių stiprumą galima nustatyti veiksnius, kurie konkrečiose sąlygose yra svarbiausi ir stipriausiai veikia kitų rodiklių kitimą.

Vienas iš dažniausiai naudojamų tiesinės sąveikos stiprumo įverčių yra porinis koreliacijos koeficientas (vadinamasis Pirsono (Pearson) koreliacijos koeficientas). Šis koeficientas gali įgyti bet kokią intervalo $[-1;1]$ reikšmę.

Porinis (Pirsono) koreliacijos koeficientas skaičiuojamas taip:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}; \quad (1.13)$$

čia x_i ir y_i – rodiklių reikšmės,

n – išmatuotų reikšmių kiekis,

\bar{x} ir \bar{y} – rodiklių aritmetiniai vidurkiai.

Porinis koreliacijos koeficientas pasižymi tuo, kad:

- koreliacijos koeficientas yra dydis, kintantis intervale nuo -1 iki 1 ;
- kai $r_{xy} = 1$, visi taškai (x_i, y_i) yra tiesėje, kurios krypties koeficientas teigiamas;
- Kai $r_{xy} = -1$, visi taškai (x_i, y_i) yra tiesėje, kurios krypties koeficientas neigiamas;
- $r_{xy} = 0$, kai kintamieji yra tiesiškai nepriklausomi.

Apskaičiuokime empirines koreliacijos koeficiento reikšmes. Aiškumo dėlei apskaičiuotą požymių x ir y koreliacijos koeficientą žymėsime $corr(x, y)$:

$$corr(x, y1) = 0,66;$$

$$corr(x, y2) = -0,85;$$

$$corr(x, y3) = 0,99.$$

Matome, kad didžiausia statistinė priklausomybė yra tarp asmenų, kaltinamų padarius nusikaltimus, skaičiaus ir jų dalies, kuri nedirbo ir nesimokė. Šiuo atveju koreliacijos koeficiento reikšmė lygi $0,99$, o tai rodo stiprų tiesinį ryšį.

Koreliacijos koeficientas tarp bendro asmenų, kaltinamų padarius nusikaltimus, skaičiaus ir dirbančių asmenų skaičiaus yra neigiamas ir lygus $-0,85$. Šiuo atveju ryšys atvirkštinis: asmenų, kaltinamų padarius nusikaltimus, skaičius didėjo, o iš jų dirbusių asmenų dalis mažėjo.

Besimokiusių asmenų skaičiaus ir visų nusikaltimais kaltinamų asmenų skaičiaus koreliacijos koeficiento reikšmė yra mažiausia. Ji teigiama ir lygi $0,66$.

Ar galima iš gautos koreliacijos koeficiento reikšmės spręsti apie koreliacijos stiprumą? Ilgoms imtims (t. y. dideliems n) yra nusistovėjęs tam tikros taisyklės, kokią koreliaciją laikyti stipria (apie tai – 1.7 lentelėje).

1.7 lentelė. Koreliacijos stiprumo interpretacija

Koreliacijos koeficiento reikšmė	Interpretacija
Nuo 0,9 iki 1,0 arba nuo –0,9 iki –1,0	Labai stipri koreliacija
Nuo 0,7 iki 0,9 arba nuo –0,7 iki –0,9	Stipri koreliacija
Nuo 0,5 iki 0,7 arba nuo –0,5 iki –0,7	Vidutinė koreliacija
Nuo 0,3 iki 0,5 arba nuo –0,3 iki –0,5	Silpna koreliacija
Nuo 0,3 iki –0,3	Nereikšminga koreliacija

Šios taisyklėmis tinka pradiniam koreliacijos stiprumui įvertinti esant didelėms imtims (kai n daugiau už 30). Daug tiksliau koreliacinį ryšį įvertinsime patikrinę hipotezę apie koreliacijos koeficiento lygybę nuliui. Kitame tyrimo etape turime įsitikinti, kad koreliacijos koeficientas statistiškai reikšmingai skiriasi nuo nulio.

Suformuluokime statistinę hipotezę:

Nulinė (pagrindinė) hipotezė H_0 teigia, kad koreliacijos koeficientas lygus nuliui.

Alternatyvioji hipotezė teigia, kad koreliacijos koeficientas nelygus nuliui.

Taisyklė, kuria remdamiesi darome išvadą apie hipotezės teisingumą ar klaidingumą, vadinama statistiniu kriterijumi.

Tikrindami taikysime kriterijų

$$T = r \sqrt{\frac{n-2}{1-r^2}}. \quad (1.14)$$

Čia r – apskaičiuota koreliacijos koeficiento reikšmė, n – išmatuotų reikšmių kiekis.

Nustatykite sprendimo priėmimo taisyklę. Parinkime reikšmingumo lygmenį. Tradiciškai reikšmingumo lygmenys yra $\alpha=0,1$, $\alpha=0,05$ ir $\alpha=0,01$. Tarkime, kad reikšmingumo lygmuo $\alpha=0.05$. Hipotezė H_0

atmetama, jei absoliuti T reikšmė viršija dydį $t_{\alpha/2(n-2)}$, kuris yra Stjudento skirstinio su $(n-2)$ laisvės laipsniais $\alpha/2$ lygmens kritinė reikšmė.

Patikrinkime, ar reikšmingas koreliacijos koeficientas
 $corr(x,y1)=0,66$.

$$T = 0.66 \sqrt{\frac{11-2}{1-0.66^2}} = 2,63$$

Statistikos $t_{\alpha/2(n-2)}$ reikšmes surasime padedami statistinės duomenų analizės programos STATISTICA ir pasinaudodami tikimybių pasiskirstymo skaičiuotuvu. Gauname reikšmę $t_{0.0025(9)} = 3,69$. Kadangi apskaičiuota T reikšmė 2,63 yra mažesnė už statistikos $t_{\alpha/2(n-2)}$ reikšmę 3,69, hipotezės atmeti negalime. Koreliacija tarp asmenų, kaltinamų padarius nusikaltimus, ir jų dalies, kuri mokėsi, statistiškai nereikšminga.

Apskaičiuokime kriterijaus T reikšmes kitais dviem atvejais.

$$corr(x,y2)=-0,85; T=-4,84;$$

$$corr(x,y3)=0,99. T=21,05.$$

Abiem atvejais gauta kriterijaus T reikšmė absoliučiu dydžiu viršija reikšmę 3,69, todėl hipotezę H_0 atmetame ir teigiame, kad koreliacija šiais atvejais yra statistiškai reikšminga.

Šio tyrimo pagrindu galime daryti išvadą, kad hipotezė apie asmenų, kaltinamų padarius nusikaltimus, kitimą ir užimtumo struktūrą pasitvirtino. Statistinis asmenų, kaltinamų padarius nusikaltimus, skaičiaus ir jų dalies, kuri nedirbo ir nesimokė, ryšys yra teigiamas, t. y. didėjant asmenų, kaltinamų padarius nusikaltimus, skaičiui kartu didėja ir jų dalis, kuri nedirbo ir nesimokė.

Statistiškai reikšmingas ir bendro asmenų, kaltinamų padarius nusikaltimus, skaičiaus ir dirbančių asmenų skaičiaus koreliacijos koeficientas. Tačiau jis yra neigiamas, todėl galime teigti, kad didėjant asmenų, kaltinamų padarius nusikaltimus, skaičiui, dirbusių asmenų dalis tarp jų mažėja.

1.3.3 užduotis. Sudarykite tiesinės regresijos modelį ir nubrėžkite diagramą.

Metodiniai nurodymai

Apskaičiuoju koreliacijos koeficientą nustatėme priklausomybių bu-

vimą, tačiau negalime pasakyti, kaip keisis, pavyzdžiui, dirbančių asmenų dalis keičiantis asmenų, kaltinamų padarius nusikaltimus, skaičiui, t. y. negalime to prognozuoti. Prognozuojant dažnai naudojami regresijos modeliai. Regresinėje analizėje prognozės yra kiekybinės – regresijos funkcija aprašo tiriamo rodiklio (kintamojo) vidurkio priklausomybę nuo kito rodiklio reikšmių kitimo. Svarbiausias regresinės analizės pranašumas yra tas, kad parenkama kintamuosius siejanti funkcija (sudaromas modelis).

Pirmiausia pasirenkamas regresijos modelio tipas – nusprendžiama, kokia priklausomybė sieja kintamuosius. Mes nagrinėsime *tiesinės* regresijos modelį, t. y. atvejį, kai vienas kintamasis nuo kito priklauso tiesiškai. Šis modelis dažnai taikomas socialiniuose moksluose, nes dažniausiai procesai visuomenėje kinta lėtai ir tiesinis modelis juos prognozuoja gana gerai.

Tiesinė kintamojo Y priklausomybė nuo kintamojo X nusakoma lygtimi

$$y = \alpha + \beta x + \varepsilon. \quad (1.15)$$

Šioje lygtyje α ir β yra nežinomi koeficientai (konstantos), o ε – atsitiktinė paklaida. Kaip atsiranda atsitiktinė paklaida? Ją gali sudaryti matavimo paklaidos arba neįvardytos ir netiriamos priežastys.

Parinkę įvairias koeficientų α ir β reikšmes gausime tiesių šeimą. Kaip apskaičiuoti koeficientus geriausiai tinkančiai tiesei? Tai atliekama taikant matematinį metodą, vadinamą *mažiausių kvadratų metodu*. Šiuo metodu gautos koeficientų α ir β reikšmės, kitaip vadinamos *įverčiais* (jas žymėsime raidėmis a ir b), randamos minimizuojant reiškinį

$$SSE = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2. \quad (1.16)$$

Šią sumą SSE minimuoja šios koeficientų α ir β reikšmės:

$$b = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2}, \quad (1.17)$$

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}. \quad (1.18)$$

Lygtis $\hat{y}(x) = a + bx$, kurios koeficientai apibrėžti (1.17) ir (1.18) lygybėmis, vadinama regresijos tiesės lygtimi, o

$$\varepsilon = y(x) - \hat{y}(x)$$

vadinama liekamąja paklaida, arba liekanomis.

Vertinant modelio tinkamumą atliekama liekamųjų paklaidų analizė ir skaičiuojamas determinacijos koeficientas. Šis koeficientas žymimas R^2 ir interpretuojamas taip:

$$R^2 = 1 - \frac{S_{\varepsilon}^2}{S_y^2} = \frac{\text{Variacijos dalis, kurią paaiškina regresijos modelis}}{\text{Visa variacija}}. \quad (1.19)$$

Čia S_{ε}^2 – liekanų (ε) dispersija, o S_y^2 – stebimų kintamojo (Y) reikšmių dispersija.

Paprastosios tiesinės regresijos atveju determinacijos koeficientas skaičiuojamas kvadratu keliant koreliacijos koeficientą:

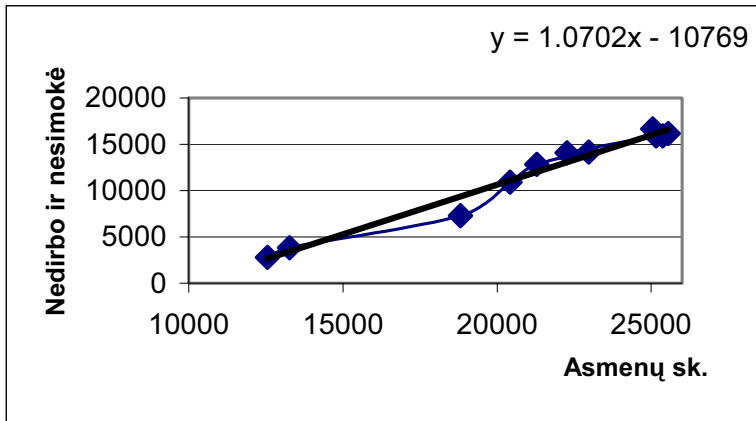
$$R^2 = r_{xy}^2. \quad (1.20)$$

Apskaičiuokime tiesinės regresijos lygties koeficientus tuo atveju, kai x – asmenų, kaltinamų padarius nusikaltimus, skaičius, o y – jų dalis, kuri nedirbo ir nesimokė (žr. 1.6 lentelę). Apskaičiavę gauname tokias reikšmes:

$$a = -10769.1, b = 1,07.$$

Užrašome regresijos lygtį ir nubraižome diagramą (žr. 1.5 pav.):

$$y = -10769,1 + 1,07x.$$



1.5 pav. Regresijos tiesės diagrama

Apskaičiuoto koeficiento b reikšmė didesnė už vieneta, taigi galime teigti, kad nedirbančių ir nesimokančių asmenų dalis tarp kaltinamų padarius nusikaltimus asmenų didėja greičiau nei bendras asmenų skaičius. Ši lygtis leidžia prognozuoti nedirbančių ir nesimokančių asmenų skaičių, pavyzdžiui, kai $x=26000$, $y=-10769,1+1,07*26000=17051$.

Nagrinėjamame pavyzdyje $r^2=0,99*0,99=0,98$ didesnis determinacijos koeficientas reiškia, kad duomenys yra mažai nutolę nuo mažiausių kvadratų metodu gautos tiesės. Jei duomenys idealiai atitinka regresijos tiesės lygtį, t. y. visos y_i reikšmės patenka ant regresijos tiesės, determinacijos koeficientas lygus 1. Jei regresijos tiesės lygtis visiškai netinkama prognozuoti, determinacijos koeficientas artimas nuliui. Praktiškai taikant tiesinę regresinę analizę dažniausiai reikalaujama, kad $r^2 \geq 0,25$. Jei $r^2 < 0,25$, tiesinės regresijos modelio taikymas labai abejotinas.

Uždaviniai

1. 1.8 lentelėje pateikti Jungtinių Tautų duomenys apie įvairiose valstybėse užregistruotų nusikaltimų skaičių 100 000 gyventojų. Nustatykite, kaip pasikeitė valstybių diferenciacija pagal bendrą nusikalsta-

mumo lygį šiuo laikotarpiu. Apskaičiuokite visus galimus rodiklius. Sudarykite diagramas ir suformuluokite išvadas.

1.8 lentelė. **Bendras užregistruotų nusikaltimų skaičius 100 000 gyventojų**

Metai	Baltarusija	Bulgarija	Kipras	Danija	Suomija	Graikija	Vengrija	Japonija	Lietuva
1980	471	506	1333	8282	10062	3063	1219	1178	410
1981	479	507	1365	8313	10596	3179	1260	1244	415
1982	456	526	1277	8512	10881	3424	1308	1291	435
1983	557	580	1496	9015	12040	3996	1421	1292	565
1984	588	552	1569	9756	13416	3562	1478	1324	585
1985	584	541	1481	10391	13390	2933	1567	1331	600
1986	541	624	1939	11091	14208	2954	1737	1302	573
1990	738	772	541	11531	8779	3256	3291	1397	992
1991	792	1877	522	11471	7812	3503	4256	1451	1200
1992	937	2387	539	11907	7792	3678	4332	1469	1505
1993	998	2328	599	12084	7644	3454	3895	1512	1620
1994	1161	2363	596	11878	7650	2909	3795	1491	1580
1995	1276	2465	550	10309	7472	3147	4907	1421	1636
1996	1236	2348	612	10043	7352	3336	4572	1409	1833
1997	1251	2909	523	10081	7273	3591	5066	1507	2045

2. Remdamiesi Lietuvos nusikalstamumo prevencijos centro duomenimis (žr. 1.9 lentelę) ištikite bendro užregistruotų nusikaltimų skaičiaus ir šių nusikaltimų rūšių priklausomybę: a) sunkių nusikaltimų, b) tyčinių nužudymų, c) išžaginimų, d) vagysčių, e) automobilių vagysčių, f) plėšimų, g) turto prievartavimo, h) sukčiavimų, i) chuliganizmo. Sudarykite regresinį modelį, atlikite jo analizę ir paaiškinkite gautus rezultatus.

3. Atlikite užregistruotų nusikaltimų Baltijos valstybėse (žr. 1.10 lentelę) laiko eilučių analizę (pagal 2 pavyzdį) atskirais laiko tarpais remdamiesi Lietuvos nusikalstamumo prevencijos centro 1990–2001 metų duomenis.

1.9 lentelė. **Bendras užregistruotų nusikaltimų skaičius 10 000 gyventojų ir jo struktūrinės dalys**

Metai	Nusikaltimų sk. 10 000 gyv.	Sunkūs nusikaltimai	Tyčiniai nužudymai	Išžaginimai	Vagystės	Automobilių vagystės	Plėšimai	Turto prievartavimas	Sukčiaviškai	Chuliganizmas
1988	58,16		0,43	0,40						2,7
1989	84,16		0,43	0,40						2,7
1990	99,2	10,8	0,60	0,53	23572		0,90		1,2	3,0
1991	120	12,1	0,69	0,51	30486		1,07		1,0	3,2
1992	150,5	15,9	0,81	0,51	41053		1,30	0,12	1,2	3,8
1993	162	21,9	1,29	0,53	42291		1,98	0,63	2,2	3,9
1994	158	25	1,41	0,44	40252		11,33	0,85	2,4	4,6
1995	163,6	35,5	1,35	0,54	41619		7,64	0,61	2,7	6,9
1996	183,3	53,8	1,09	0,45	44600	10,5	9,38	0,65	2,9	8,1
1997	204,5	57,2	1,06	0,45	47193	10,3	10,72	0,64	3,6	8,4
1998	210,9	59,7	0,96	0,45	48213	11,2	9,85	0,52	3,9	8,0
1999	208,3	56,7	0,93	0,61	47865	9,9	9,13	0,62	2,5	7,3
2000	222,5	49,4	1,08	0,49	53225	14,0	11,82	0,52	1,8	7,3
2001	214,5	48,7	1,02	0,48	48899	15,8	11,33	0,53	2,3	6,5

1.10 lentelė. **Užregistruoti nusikaltimai Baltijos valstybėse 10 000 gyventojų**

Metai	Lietuva	Latvija	Estija
1990	99.2	129	150
1991	120	151	206
1992	150.5	235	267
1993	162	204	245
1994	158	161	236
1995	163.6	155	265
1996	183.3	152	240
1997	204.5	147	280
1998	210.9	147.9	314.4
1999	208.3	180.2	356.7
2000	222.5	211.3	419.8
2001	214.5	215	428

4. Iširkite užregistruotų nusikaltimų Baltijos valstybėse priklausomybę (žr. 1.10 lentelę), sudarykite prognozavimo modelį ir paaiškinkite gautus rezultatus.

5. Atlikite asmenų, kaltinamų padarius nusikaltimus (žr. 1.11 lentelę), socialinės–ekonominės struktūros dinamikos analizę. Gavę rezultatus suformuluokite išvadas. Patikrinkite šių darbinių hipotezių teisingumą:

Tarp asmenų, kaltinamų padarius nusikaltimus, kasmet daugėja:

- moterų dalis;
- nepilnamečių dalis;
- neturinčių nuolatinio darbo dalis;
- pakartotinai nusikaltusių asmenų dalis;
- neblaivių asmenų dalis.

1.11 lentelė. Asmenų, kaltinamų padarius nusikaltimus, struktūra

Metai	Visi asmenys, 10 000 gyv.	Moterys, proc.	Nepilnamečiai, proc.	Nedirbo ir nesimokė, proc.	Nusikalto pakartotinai, proc.	Neblaivūs, proc.
1988	34,9		11,9			
1989	34,8		14,6			
1990	33,7	12,5	16,3	22,4	20,2	37,9
1991	35,5	10,1	14,5	28,9	18,9	39,8
1992	50,3	10,2	14,6	38,8	18,5	35,9
1993	54,8	10,3	15,6	53,4	21,3	37,4
1994	57,2	11,4	14,3	60,4	24,0	38,1
1995	61,8	10,4	14,7	61,7	21,7	33,9
1996	60,0	11,4	15,3	63,2	23,2	32,9
1997	68,9	13,9	13	63,2	22,6	29,2
1998	68,5	13,5	13,1	62,7	22,7	27,6
1999	68,0	13,0	13,3	62,7	23,4	25,7
2000	67,7	10,0	14,3	66,5	27,4	28,1
2001	72,2	8,7	13,8	67,2	29,6	30,8

6. Remdamiesi statistikos duomenimis (žr. 1.12 lentelę) apie kai kurių turtingiausių pasaulio valstybių išsivystymo lygį 1991 metais (*New York: Bantam books, 1992*; <http://www.korpios.org>) raskite statiškai reikšmingas priklausomybes požymių, turinčių didžiausią įtaką kriminaliniams rodikliams. Parašykite išvadas.

1.12 lentelė. Socialiniai–ekonominiai kai kurių turtingiausių pasaulio valstybių rodikliai

Valstybė	Nužudy- mai 100 000 gyv.	Išžagini- mai 100 000 gyv.	Kalinių sk. 1000 gyv.	Vidutinės klasės dydis, proc.	Perkamoji galia, JAV dol.	Procentinė nesantuokoje gimusių vaikų dalis
Kanada	5,45	–	–	58,2	19178	12,1
Danija	5,17	11,23	0,7	–	17621	41,9
Vokietija	4,20	8,60	0,8	70,1	19500	9,4
Norvegija	1,99	7,87	–	73	16904	–
Jungtinė Karalystė	1,97	7,26	1,0	58,5	–	19,2
Švedija	1,73	15,70	0,6	79	19729	46,4
Japonija	1,20	1,40	0,4	90	19107	1,0
Suomija	0,70	7,20	–	–	15997	–
JAV	8,40	37,20	4,2	53,7	22204	21,5



2. UŽIMTUMO IR NEDARBO STATISTIKA

Su gyventojų užimtumu susijusių sąvokų turinys šiek tiek skiriasi. Tai priklauso nuo užimtumo tyrimo metodikos. Aptarkime šias sąvokas ir šios statistikos rodiklius.

Bedarbiai:

Oficialioje statistikoje (remiantis darbo biržos duomenimis) bedarbiai yra nedirbantys darbingo amžiaus gyventojai, nesimokantys dienišėse mokymo įstaigose ir užsiregistravę valstybinėje darbo biržoje kaip ieškantys darbo ir pasirengę profesiniam mokymuisi.

Gyventojų užimtumo tyrimuose pagal EUROSTAT reikalavimus bedarbiai yra tiriamo amžiaus gyventojai, kurie tiriamąją savaitę neturėjo darbo, o jį radę buvo pasirengę per artimiausias dvi savaites pradėti dirbti, keturias savaites įvairiais būdais intensyviai ieškojo mokamo darbo. Bedarbių grupei priskiriami ir tie moksleiviai, studentai, namų šeimininkės ir kiti neaktyvūs gyventojai, kurie tiriamąją savaitę aktyviai ieškojo darbo ir buvo pasirengę per artimiausias dvi savaites pradėti dirbti.

Užimtieji gyventojai:

Tiriamo amžiaus gyventojai, kurie tiriamąją savaitę dirbo kokį nors darbą ne mažiau kaip 1 valandą, už kurį gavo darbo užmokestį pinigais arba natūra (maisto produktais ar kitais gaminiais) arba turėjo pelno (pajamų). Užimtieji kaime yra darbingo amžiaus žemdirbiai (turintys iki 3 ha žemės), kurie neturi kito pragyvenimo šaltinio, išskyrus tą žemę, nesvarbu, ar joje užaugintą produkciją panaudoja savo reikmėms, ar dalį parduoda. Smulkūs žemdirbiai – pensininkai užimtųjų kategorijai priskiriami tik tada, kai jų pajamos iš žemės ūkio didesnės nei gaunama pensija. Kiti pensininkai priskiriami neaktyviems gyventojams.

Neaktyvūs gyventojai:

Tiriamo amžiaus gyventojai, nepriskirti užimtųjų ar bedarbių grupei. Jiems priskiriami moksleiviai, studentai, namų šeimininkės, pensininkai, invalidai, nuteistieji, asmenys, praradę viltį rasti darbą.

Darbo jėga (ekonomiškai aktyvūs gyventojai):

Visi užimtieji gyventojai ir bedarbiai.

Pagrindiniai rodikliai yra šie:

- darbo jėgos aktyvumo lygis (gyventojų ekonominio aktyvumo koeficientas)

$$DJAL_1 = \frac{\text{Darbo jėga}}{\text{Darbingo amžiaus ir vyresni gyventojai}} \quad (2.1)$$

arba

$$DJAL_2 = \frac{\text{Darbo jėga}}{\text{Bendras gyventojų skaičius}}; \quad (2.2)$$

- gyventojų užimtumo koeficientas

$$GUK = \frac{\text{Užimtųjų gyventojų skaičius}}{\text{Darbingo amžiaus ir vyresni gyventojai}}; \quad (2.3)$$

- nedarbo lygis (nedarbo koeficientas)

$$NL = \frac{\text{Bedarbių skaičius}}{\text{Darbo jėga}}; \quad (2.4)$$

- bedarbių skaičius, tenkantis tūkstančiui užimtųjų

$$BS_1 = \frac{\text{Bedarbių skaičius}}{\text{Užimtųjų skaičius}} \cdot 1000; \quad (2.5)$$

- vidutinė nedarbo trukmė (bedarbių skaičiaus pasiskirstymas darbo neturėjimo laiko intervaluose)

$$T = \frac{\sum_i T_i N_i}{\sum_i N_i}. \quad (2.6)$$

2.1 pavyzdys. Gyventojų užimtumo tyrimas: pagrindiniai užimtumo ir nedarbo rodikliai

Lietuvos statistikos departamentas, vertindamas gyventojų užimtumą, atliko gyventojų užimtumo tyrimus. 2.1 lentelėje pateikti šio tyrimo metu gauti duomenys.

2.1 lentelė. Gyventojų užimtumas 2001 m. gegužės mėn.

Rodiklis	2001 gegužės mėn.
Bedarbiai, registruoti darbo biržoje, tūkst.	220,3
Bedarbiai, statistikos departamento duomenimis, tūkst.	292,9
Užimtieji gyventojai, iš viso, tūkst.	1481,8
Užimtieji gyventojai 15–64 metų, tūkst.	1451,2
Neaktyvūs gyventojai, tūkst.	1211,1
Darbo jėga, tūkst.	1774,7
Gyventojai metų pradžioje, iš viso, tūkst.	3692,6
15 m. ir vyresni gyventojai, tūkst.	2985,8
15–64 metų gyventojai, tūkst.	2483,1

2.1.1 užduotis. Apskaičiuokite pagrindinius užimtumo ir nedarbo rodiklius.

1. Darbo jėgos aktyvumo lygis

$$\text{a) } DJAL_1 = \frac{1774,7}{2985,8} = 0,594, \text{ arba } 59,4 \text{ procento;}$$

$$\text{b) } DJAL_2 = \frac{1774,7}{3692,6} = 0,481, \text{ arba } 48,1 \text{ procento;}$$

2. Gyventojų užimtumo koeficientas

$$GUK = \frac{1481,8}{2985,8} = 0,496, \text{ arba } 49,6 \text{ procento;}$$

3. Nedarbo lygis (nedarbo koeficientas)

$$NL = \frac{220,3}{1774,7} = 0,124, \text{ arba } 12,4 \text{ procento};$$

4. Bedarbių skaičius, tenkantis tūkstančiui užimtųjų

$$BS_1 = \frac{220,3}{1481,8} \cdot 1000 = 148,67.$$

2.2 pavyzdys. Užimtųjų pasiskirstymas Lietuvoje pagal išsimokslinimą.

Užimtųjų pasiskirstymas Lietuvoje pagal išsimokslinimą 1997 ir 2000 metais apibūdinamas 2.2 lentelėje pateiktais duomenimis.

2.2 lentelė. Užimtųjų gyventojų išsimokslinimas, tūkstančiais

Grupės nr.	Užimtųjų gyventojų išsimokslinimas (tūkst.)	1997	proc.	2000	proc.
	<i>Iš viso</i>	1570,7	100	1518	100
1	Aukštasis	314,3	20,0	316,7	20,9
2	Aukštesnysis	402,1	25,6	379,3	25,0
3	Vidurinis be prof. mok.	375,8	23,9	303,7	20,0
4	Vidurinis su prof. mok.	194,3	12,4	268,7	17,7
5	Devynmetis be prof. mok.	149,1	9,5	116,1	7,6
6	Devynmetis su prof. mok.	62,7	4,0	76,1	5,0
7	Pradinis, neturi pradinio	72,4	4,6	57,4	3,8

2.2.1 užduotis. Apibūdinkite skirtumus išsimokslinimo struktūroje 1997 ir 2000 metais naudodami šiuos rodiklius:

1. Skirtumų indeksą:

$$I_{skirt} = \frac{1}{2} \sum |w_1 - w_0|. \quad (2.7)$$

Čia w_1 ir w_0 – rodiklių dalys (svoriai), kurie kinta nuo 0 iki 1. Indekso I_{skirt} kitimo ribos irgi nuo 0 iki 1. Jei w išreikštas procentais, ši požymio dalis kinta $0 \leq w \leq 100$. Skirtumų indeksas šiuo atveju

$$I_{skirt} = \frac{1}{2} \frac{\sum |w_1 - w_0|}{100}. \quad (2.8)$$

2. Absoliučią struktūrinių pokyčių tiesinį koeficientą:

$$d = \frac{1}{n} \sum |w_1 - w_0|. \quad (2.9)$$

Čia n – grupių skaičius.

3. Absoliučią struktūrinių pokyčių kvadratinio nuokrypio koeficientą:

$$s = \sqrt{\frac{1}{n} \sum (w_1 - w_0)^2}. \quad (2.10)$$

Tarpinio skaičiavimo rezultatai pateikti 2.3 lentelėje. Čia w_0 – tam tikro išsilavinimo lygio užimtų gyventojų skaičius 1997 metais (procentais), w_1 – tas pats 2000 metais (procentais).

2.3 lentelė. Tarpinio skaičiavimo rezultatai

Grupės nr.	w_0 , proc.	w_1 , proc.	$ w_1 - w_0 $	$(w_1 - w_0)^2$
1	20,0	20,85	0,85	0,73
2	25,6	24,99	0,61	0,38
3	23,92	20,0	3,92	15,36
4	12,37	17,7	5,33	28,42
5	9,49	7,64	1,85	3,40
6	3,99	5,0	1,01	1,04
7	4,6	3,78	0,82	0,69
Σ	100	100	14,41	50,01

Skirtumų indeksas $I_{skirt} = \frac{1}{2} \cdot \frac{14,41}{100} = 0,072$. Gauta indekso reikšmė rodo, kad nuo 1997 iki 2000 metų skirtumai užimtųjų išsilavinimo struktūroje pakito nedaug – apie 7 procentus.

Absoliučių struktūrinių pokyčių tiesinis koeficientas $d = \frac{1}{n} \sum |w_1 - w_0| = \frac{1}{7} \cdot 14,41 = 2,05$. Vadinasi, vidutiniškai struktūriniai skirtumai sudaro 2,05 procento.

Absoliučių struktūrinių pokyčių kvadratinio nuokrypio koeficientas $s = \sqrt{\frac{1}{n} \sum (w_1 - w_0)^2} = \sqrt{\frac{1}{7} \cdot 50,01} = 2,67$. Šis rodiklis tiksliausiai parodo dirbančių asmenų išsilavinimo lygio 1997 ir 2000 metais skirtumą. Šis skirtumas sudaro 2,67 procento.

2.3 pavyzdys. Rodiklių priklausomumo tikrinimas

Tiriama, ar Lietuvoje dirbančių vyrų ir moterų atliekamas darbo tipas yra vienodas. Paklausus 1430 atsitiktinai parinktų vyrų ir moterų, kokio tipo darbą jie dirba, buvo gauti rezultatai, pateikti 2.4 lentelėje.

2.4 lentelė. Dirbantys Lietuvos gyventojai

Lytis	Darbdaviai ir sau dirbantys	Samdomieji asmenys	Iš viso
Vyrai	147	562	709
Moterys	89	632	721
Iš viso	236	1194	1430

2.3.1 užduotis. Remdamiesi šiais duomenimis nustatykite, ar vyrai ir moterys, tiek kaip darbdaviai, tiek kaip samdomieji darbuotojai, dirbo vienodai, t. y. ar užimtumo struktūra priklauso nuo lyties?

Metodiniai nurodymai

Šiuo atveju reikia iširti dviejų kokybinių kintamųjų, t. y. užimtumo tipo ir lyties, priklausomybę ir nustatyti, ar šie kintamieji priklausomi, ar nepriklausomi. Tokį statistinį ryšį galima įvertinti specialiais matematiniais metodais, iš kurių pats populiariausias – χ^2 kriterijus.

Suformuluokime hipotezę. Hipotezė H_0 teigia, kad vyrai ir moterys vienodai dirba tiek kaip darbdaviai, tiek kaip samdomieji darbuotojai. Alternatyva H_1 – darbdavių ir samdomųjų darbuotojų tarp vyrų ir moterų yra nevienodai.

Analizuojant statistinius ryšius taikant kriterijų χ^2 pirmiausia apskaičiuojama, kokios turėtų būti lentelės gardelių reikšmės, jeigu tarp kintamųjų nebūtų statistinio ryšio. Šios reikšmės (arba dažnumai) nustatomos pagal formulę:

$$q_{ij} = \frac{f_{i*} f_{*j}}{n}; \quad (2.11)$$

čia q_{ij} – tikėtinas dažnumas gardelėje, esančioje i-tos eilutės ir j-tojo stulpelio sankirtoje,

f_{i*} – i-tos eilutės dažnumų suma, $f_{i*} = f_{i1} + f_{i2}$,

f_{*j} – j-tojo stulpelio dažnumų suma, $f_{*j} = f_{1j} + f_{2j}$,

n – nagrinėjamų elementų skaičius, šiuo atveju $n=1430$.

Šioje formulėje naudojamų simbolių reikšmės paaiškintos 2.5-oje lentelėje.

Kadangi 2.4-os lentelės pagrindinių (paryškintų) duomenų eilutės bei stulpeliai turi po dvi gardeles, tokia lentelė vadinama 2×2 ir jos gardelės žymimos taip (žr. 2.5 lentelę):

2.5 lentelė. Simbolinis 2.4 lentelės reikšmių žymėjimas

Lytis	Darbdaviai ir sau dirbantys	Samdomieji asmenys	Iš viso
Vyrai	f_{11}	f_{12}	f_{1*}
Moterys	f_{21}	f_{22}	f_{2*}
Iš viso	f_{*1}	f_{*2}	

Remdamiesi 2.4-tos lentelės duomenimis apskaičiuojame tikėtinus dažnius:

$$q_{11} = \frac{709 \cdot 236}{1430} \approx 117, \quad q_{12} = \frac{709 \cdot 1194}{1430} \approx 592,$$

$$q_{21} = \frac{721 \cdot 236}{1430} \approx 119, \quad q_{22} = \frac{721 \cdot 1194}{1430} \approx 602.$$

Toliau apskaičiuojame išmatuotų reikšmių (žr. 2.4 lentelę) ir apskaičiuotų tikėtinų reikšmių skirtumus. Žinoma, jei hipotezė teisinga, šie skirtumai neturėtų būti dideli. Taip sudaromas χ^2 kriterijus:

$$\chi^2 = \sum_{i,j=1}^n \frac{(f_{ij} - q_{ij})^2}{q_{ij}}; \quad (2.12)$$

čia f_{ij} – gaudelės, esančios i -tosios eilutės ir j -tojo stulpelio sankirtoje, tikrasis (stebėtas) dažnumas, kurio reikšmės nurodytos 2.4-oje lentelėje.

Įrašę reikšmes, gausime:

$$\chi^2 = \frac{(147 - 117)^2}{117} + \frac{(562 - 592)^2}{592} + \frac{(89 - 119)^2}{119} + \frac{(632 - 602)^2}{602} \approx 18,26$$

Tuomet apskaičiuojami lentelės laisvės laipsniai:

$$df = (s-1)(t-1);$$

čia df – lentelės laisvės laipsnių skaičius,

s – eilučių skaičius,

t – stulpelių skaičius.

Mūsų nagrinėjamą lentelę sudaro dvi eilutės ir du stulpeliai. Taigi

$$df = (2-1)(2-1) = 1.$$

Galiausiai reikia nustatyti, ar ši gauta kriterijaus χ^2 reikšmė yra pakankamai didelė, kad būtų galima teigti, jog tiriami požymiai yra priklausomi. Kintamųjų nepriklausomumo hipotezė atmetama, kai apskaičiuotos statistikos χ^2 reikšmė yra didesnė už χ^2 skirstinio su $(s-1)(t-1)$ laisvės laipsnių α lygmens kritinę reikšmę. Čia α – pasirinktas reikšmingumo lygmuo. Tarkime, pasirinkome reikšmingumo lygmenį $\alpha = 0,05$. Kaip gauti kritinę reikšmę? Ją galima rasti matematinėse len-

telėse, kurios pateikiamos prieduose, arba pasinaudoti statistinių programinių sistemų (STATISTICA, SPSS, STATGRAPHICS ir kt.) tikimybių skaičiavimo programomis (moduliais). Vienaip ar kitaip rastume, kad kritinė reikšmė $\chi_{0,05}^2(1) = 3,84$, ir ji yra daug mažesnė už gautą reikšmę 18,26, todėl hipotezę H_0 turime atmesti.

Nagrinėjamame pavyzdyje gauta kriterijaus reikšmė yra daug didesnė už kriterijaus reikšmę, atitinkančią reikšmingumo lygmenį 0,05 ar 0,005 ($\chi_{0,005}^2(1) = 7.88$). Tai rodo, kad atmetant hipotezę daroma labai nedidelė klaida.

χ^2 kriterijaus taikymas susijęs su kai kuriais apribojimais, taikomais stebėjimų skaičiui n . Šio kriterijaus verčiau netaikyti, kai stebėjimų skaičius n mažesnis už 30 ir bent 75 proc. dažnių lentelės gardelių tikėtini dažniai ne mažesni kaip 5.

Jei tikėtini dažniai mažesni už 5, 2×2 lentelei galime taikyti Fišerio kriterijų (žr. 2.4 pavyzdį).

Jei n labai didelis, daug didesnis už 1000, hipotezė apie požymių nepriklausomumą beveik visuomet atmetama.

2.4 pavyzdys. Darbo patirties ir įsidarbinimo priklausomumas

Per mėnesį į darbo biržą kreipėsi 38 bedarbiai. Rezultatai apie įdarbinimą atsižvelgiant į darbo patirtį pateikiami 2.6 lentelėje.

2.4.1 užduotis. Nustatykite, ar darbo patirtis padeda įsidarbinti. Įvertinkite, kaip skiriasi tikimybė įsidarbinti atsižvelgiant į darbo patirtį.

2.6 lentelė. Darbo patirtis ir galimybė įsidarbinti

Įsidarbino per tris mėnesius	Turi darbo patirtį	Neturi darbo patirties	Iš viso
Ne	2 a	3 b	$a+b=5$
Taip	27 c	5 d	$c+d=32$
Iš viso	$a+c=29$	$b+d=8$	$a+b+c+d=38$

Šiuo atveju, užuot taikius χ^2 kriterijų, verčiau taikyti Fišerio kriterijų, paremtą hipergeometriniu skirstiniu. Taikant šį kriterijų tikimybės p -reikšmės tiksliai apskaičiuojamos, tuo tarpu χ^2 kriterijaus p -reikšmės yra apytikslės. Kadangi Fišerio kriterijus paprastai taikomas tik lentelėms 2×2 su keturiomis gardelėmis, tradiciškai jų reikšmės žymimos raidėmis a, b, c, d .

Paprasciausia Fišerio kriterijaus tikimybės reikšmę apskaičiuoti taip: tikimybė, kad požymiai yra nepriklausomi (hipotezė H_0), skaičiuojama taikant formulę

$$p(a, b, c, d) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Ši formulė taikoma pradinei dažnių lentelei ir visoms galimoms lentelėms, kurios sudaromos iš pradinės lentelės atimant po vieną iš mažiausios gardelės reikšmės (atitinkamai keičiant kitų gardelių reikšmes, kad eilučių ir stulpelių sumos išliktų tos pačios), kol mažiausios gardelės reikšmė tampa lygi nuliui. Gautos tikimybės sudedamos. Jei gauta reikšmė mažesnė už parinktą reikšmingumo lygmenį, hipotezė atmetama.

Iš 2.6 lentelės sudaromos šios papildomos lentelės (žr. 2.7 ir 2.8 lenteles):

2.7 lentelė. Darbo patirtis ir galimybė įsidarbinti, $a=a-1$

Įsidarbino per tris mėnesius	Turi darbo patirtį	Neturi darbo patirties	Iš viso
Ne	1 a	4 b	$a+b=5$
Taip	28 c	4 d	$c+d=32$
Iš viso	$a+c=29$	$b+d=8$	$a+b+c+d=38$

2.8 lentelė. Darbo patirtis ir galimybė įsidarbinti, $a=a-1$

Įsidarbino per tris mėnesius	Turi darbo patirtį	Neturi darbo patirties	Iš viso
Ne	0 a	5 b	$a+b=5$
Taip	29 c	3 d	$c+d=32$
Iš viso	$a+c=29$	$b+d=8$	$a+b+c+d=38$

Apskaičiavę šių trijų lentelių tikimybes gausime:

$$p(2,3,27,5) = \frac{5!32!29!8!}{2!3!27!5!38!} = 0,052;$$

$$p(1,4,28,4)=0,0047;$$

$$p(0,5,29,3)=0,00013.$$

Šių tikimybių suma lygi 0,057. Jei parinktas reikšmingumo lygmuo 0,1, hipotezę reikia atmesti ir teigti, kad įdarbinimas priklauso nuo darbo patirties. Jei parinktas reikšmingumo lygmuo 0.05, hipotezės atmesti negalime, nes gauta tikimybių suma didesnė už šią reikšmę.

2.5 pavyzdys. Bedarbių šeiminės padėties analizė

Darbo birža pateikė šiuos duomenis apie šeiminių bedarbių padėtį (žr. 2.9 lentelę):

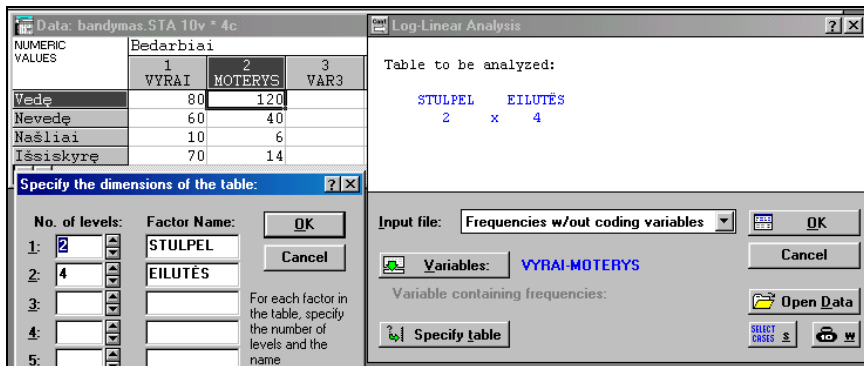
2.9 lentelė. Duomenys apie šeiminių bedarbių padėtį

Šeiminė padėtis	Vyrai	Moterys	Iš viso
Vedę	80	120	200
Nevedę	60	40	100
Našliai, našlės	10	6	16
Išsiskyre	70	14	84
Iš viso	220	180	400

2.5.1 užduotis. Nustatykite, ar yra priklausomybė tarp lyties ir šeiminės padėties.

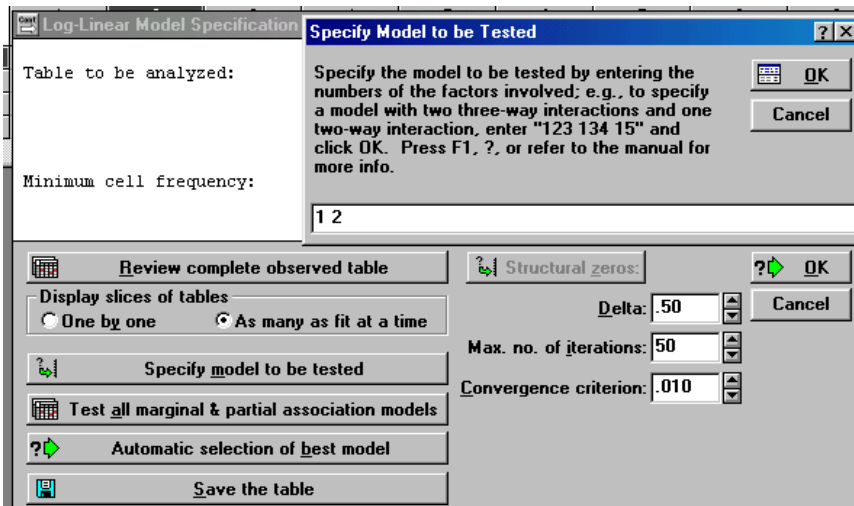
Iki šiol tyrėme lenteles 2×2 , o ši lentelė yra 2×4 . Tirdami galime taikyti χ^2 kriterijų, kuris apskaičiuojamas taip pat kaip 2×2 atveju. Kuo didesnė lentelė, tuo sudėtingiau skaičiuoti, todėl skaičiuojama pasitelkiant kompiuterines programas. Trumpai panagrinėkime, kaip šią užduotį galime atlikti naudodami STATISTICA programą.

Pirmiausia turime parinkti reikalingą modulį. Lentelėms 2×2 skaičiuoti naudojamas **Nonparametrics/Distrib** modulis, o didesnėms nei 2×2 lentelėms – **Log-linear Analysis** modulis. Duomenys įvedami tokiu pat pavidalu kaip 2.9 lentelėje, tik be suminių reikšmių (be šeštos eilutės ir ketvirto stulpelio). 2.1 paveiksle matome tris langus: pradinių duomenų, parametrų įvedimo formos (langas dešinėje) ir stulpelių ir eilučių skaičiaus nustatymo langą (šis langas iškviečiamas mygtuku **Specify table**).



2.1 pav. Pradinių duomenų ir parametrų įvedimo langai

Svarbus yra kitas žingsnis, kuriuo parenkamas skaičiavimo modelis. Šio lango forma iškviečiama mygtuku **Specify model to be tested**. Mūsų atveju šioje formoje įrašome 1 2, nurodydami, kad būtų skaičiuojami efektai tikrinant pirmą ir antrą faktorius, t. y. lyties (stulpeliai) ir šeiminės padėties (eilutės) faktorius (žr. 2.2 pav.).



2.2 pav. Modelio parinkimo langas

Toliau (paspaudę **OK** klavišą) jau gauname skaičiavimo rezultatus. Rezultatų lange pateikiama apskaičiuota statistikos reikšmė, laisvės laipsniai ir pirmosios rūšies klaidos tikimybė (t. y. kokia klaida padaroma atmetant nulinę hipotezę). Šiuo atveju rezultatai tokie:

Pearson Chi-square=46.24887df=3p=0.0000000

Kaip suprasti šiuos rezultatus? **Pearson Chi-square** – tai apskaičiuota χ^2 kriterijaus reikšmė, **df** – laisvės laipsniai, o **p** – kriterijaus reikšmę atitinkantis reikšmingumo lygmuo (α), pateiktas septynių skaitmenų po kablelio tikslumu. Visi šie skaitmenys lygūs nuliui, todėl hipotezę, kad bedarbių lytis ir šeiminė padėtis nesusiję, galime atmesti esant labai mažam reikšmingumo lygmeniui.

Uždaviniai

1. Užimtųjų ir bedarbių gyventojų pasiskirstymas pagal išsilavinimą 2000 metais apibūdinamas tokiais duomenimis (žr. 2.10 lentelę):

2.10 lentelė. Lietuvos gyventojų pasiskirstymas pagal išsilavinimą

Išsilavinimas	Užimtieji, tūkst.	Bedarbiai, tūkst.
Aukštasis	316,7	19,8
Aukštesnysis	379,3	59,9
Vidurinis be prof. mokymo	303,7	63,6
Vidurinis su prof. mokymu	268,7	61,3
Devynmetis be prof. mokymo	116,1	41,2
Devynmetis su prof. mokymu	76,1	24
Pradinis, neturi pradinio	57,4	5,9

Apibūdinkite užimtųjų ir bedarbių skirtumus išsilavinimo struktūroje.

2. Statistikos departamentas pateikė šiuos duomenis apie Lietuvos apskričių gyventojų užimtumą 1997 metais (žr. 2.11 lentelę).

2.11 lentelė. Lietuvos apskričių gyventojų užimtumas

Apskritis	Gyventojų skaičius	Bedarbių skaičius, tūkst.	Nedarbo lygis, proc.
Alytaus	202400	6,6	13,4
Kauno	757000	14,7	14,3
Klaipėdos	415400	9,7	13,3
Marijampolės	198500	5,1	7,2
Panevėžio	323300	9,5	12,6
Šiaulių	401900	12,9	13,3
Tauragės	130200	4,8	20,3
Telšių	182600	4	11,8
Utenos	202400	6,4	15,2
Vilniaus	895400	30,8	16,1

Apskaičiuokite:

- darbo jėgos dydį;
- darbo jėgos aktyvumo lygį;
- kiek 1000 užimtųjų tenka bedarbių;

3. Lentelėje 2.12 pateikti duomenys apie nedarbą Rusijoje 2000 metų pradžioje:

2.12 lentelė. **Duomenys apie nedarbą Rusijoje**

Rodiklis	
Bendras gyventojų skaičius	147 (mln.)
Bendras bedarbių skaičius (tyrimų duomenys)	8,79 (mln.)
Nedarbo lygis	11,7 proc.
Oficialiai užregistruoti bedarbiai	1 263 (tūkst.)
Bedarbio pašalpą gaunantys bedarbiai	1 090 (tūkst.)

Apskaičiuokite:

- darbo jėgos dydį;
- darbo jėgos aktyvumo lygį;
- kiek bedarbių tenka 1000 užimtųjų;
- kaip skiriasi oficialiai užregistruotų bedarbių skaičius nuo bedarbių skaičiaus pagal tyrimų duomenis;
- kiek procentų bedarbių negauna bedarbio pašalpos;
- apibūdinkite ekonomiškai aktyvių gyventojų struktūrą.

4. Gyventojų tyrimo metu apklausus 706 gyventojus buvo gauti šie duomenys (žr. 2.13 lentelę).

2.13 lentelė. **Darbo ieškančių gyventojų pasiskirstymas pagal amžiaus grupes**

Amžiaus grupė	Apklausti gyventojai	Ieško darbo
iki 20 m.	47	9
20–29 m.	130	15
30–39 m.	125	13
40–49 m.	140	17
50–59 m.	182	21
60 m. ir vyresni	82	6

Nubrėžkite duomenų diagramas ir nustatykite, ar darbo ieškančių gyventojų skaičius priklauso nuo amžiaus grupės.

3. STATISTINĖ GYVENIMO LYGIO ANALIZĖ: VISUOTINĖ GEROVĖ IR NELYGYBĖ

Atliekant tarptautinio masto tyrimus apie ekonomikos raidos įtaką socialinei gyventojų padėčiai vartojami įvairūs terminai ir rodikliai. Reikia pripažinti, kad ne visi jie vienareikšmiškai apibrėžti. Netgi tokie visuotinai vartojami terminai kaip *vidurinė klasė* ar *skurdo riba* neturi vieno apibrėžimo. Tokių apibendrinančių rodiklių vertinimas valstybėse su skirtingomis ekonominėmis sistemomis ir skirtingu ekonomikos lygiu yra gana sudėtingas ir komplikluotas.

Terminas *vidurinė klasė* dažnai apibrėžiamas kaip sluoksniškas gyventojų, kurių pajamos patenka į intervalą nuo 33 procentų mažiau visų šalies gyventojų pajamų medianos ir iki 50 procentų daugiau nei mediana.

Nagrinėjant skurdo problemą skiriamas *santykinis* ir *absoliutus* skurdas.

Pagal *santykinio* skurdo kriterijų prie skurstančiųjų priskiriami tie visuomenės nariai, kurių asmeninės pajamos daug mažesnės nei vidutinės. Pavyzdžiui, santykinio skurdo riba dažnai laikomas dydis, lygus 50 (kartais nuo 40 iki 60) procentų asmeninių pajamų ar faktinių vartojimo išlaidų vidurkio arba medianos.

Akivaizdus prieštaravimas šiam apibrėžimui yra tai, kad gyventojų pajamų medianos atskirose valstybėse yra skirtingos. Aukščiausia mediana yra JAV, taigi JAV žemiau skurdo ribos gyvenantys žmonės gali būti turtingesni negu trečiojo pasaulio šalių vidurinės klasės atstovai. Santykinio skurdo skaičiavimas turi prasmę tose valstybėse, kur asmeninės pajamos, palyginti su objektyviais minimaliais normatyvais, yra didelis. Kitais atvejais galime gauti kuriozinius rezultatus, visiškai neatspindinčius tikrosios situacijos.

Absoliutus skurdą ir jo ribą galime apskaičiuoti remdamiesi minimaliais asmeninio vartojimo normatyvais. 1990 m. priimtame Lietuvos Respublikos gyventojų pajamų garantijų įstatyme minimalus gyvenimo lygis (MGL) apibrėžtas kaip „šeimos mėnesinių pajamų suma, tenkanti vienam žmogui per mėnesį ir garantuojanti visiems minimalų socialiai priimtina poreikių tenkinimo lygį, atitinkantį organizmo maisto poreikius pagal fiziologines normas, taip pat minimalius drabužių, avalynės, baldų, ūkinių, sanitarijos ir higienos reikmenų, buto, komunalinių, buitinių

nių, transporto, ryšių, kultūros ir švietimo paslaugų poreikius”. Tai visiškai atitinka ir Visuotinės Žmogaus Teisių Deklaracijos nuostatą, kurioje teigiama, jog „kiekvienas žmogus turi teisę į pakankamą gyvenimo lygį, kuris garantuotų jo ir jo šeimos gerovę, ir ypač jo maistą, drabužius, būstą, medicininę priežiūrą ir būtiną socialinį aptarnavimą” [12].

Lietuvos Vyriausybė oficialiai tvirtina *taikomąjį MGL*. Šis MGL (šiuo metu jo dydis – 125 Lt per mėnesį) nėra pagrįstas kokiais nors objektyviais asmeninio vartojimo normatyvais: jo dydį lemia iš esmės vienintelis kriterijus – konkrečios biudžeto galimybės jau šiandien tenkinti tam tikrus visuomenės požiūriu svarbiausius socialinius poreikius. Šis MGL negali būti kriterijus skurdo ribai pagrįsti.

Lietuvoje sudarinėjamas ir *skaičiuojamasis MGL*. 1999 m. vidutiniškai per mėnesį jis sudarė 187,5 lito. Tačiau ir šis MGL negali būti naudojamas socialinei skurdo ribai nustatyti arba jį naudojant būtina daryti dideles išlygas. Vienas iš esminių šio MGL trūkumų – iškreipta vartojimo (šeimos išlaidų) struktūra, sumažinanti jo dydį: vadovaujantis Pasaulio banko rekomendacijomis, iš skaičiuojamojo MGL net 60 procentų šeimų išlaidų skiriama maistui. Šis MGL daugiau nukreiptas į *fiziologinį* minimumą ir į žemutinę skurdo ribą.

JAV skurdo riba skaičiuojama vidutinės maistui skiriamas išlaidas dauginant iš trijų. Šią metodiką dar 1960 m. pasiūlė Mollie Orshansky. Skurdo riba indeksuojama pagal vartotojo kainų indekso kitimą (įvertinant infliaciją). Šis skurdo ribos vertinimo būdas dažnai kritikuojamas, nes išlaidos būstui ir komunalinėms paslaugoms didėja greičiau nei maistui, taigi išlaidos maistui šeimų biudžete dabar sudaro mažesnę dalį nei 1960-aisiais.

Pasaulio bankas kaip didelio ir vidutinio skurdo standartą siūlo naudoti vieno ir dviejų dolerių pajamas per dieną. Taigi jei šalies vidutinės pajamos pakankamai didelės, tai net skurdžiausiai gyvenančios šeimos yra virš šios skurdo ribos.

3.1. Gyventojų ir namų ūkių pajamų ir išlaidų statistika

Gyventojų pajamų ir išlaidų lygis ir socialinis–ekonominis gyventojų pasiskirstymas yra gyvenimo lygį šalyje apibūdinantys rodikliai, kuriuose glūdi socialinio visuomenės teisingumo esmė. Šiuolaikinėje visuomenėje gyventojų pajamų paskirstymas paklūsta ne tik griežtiems

ekonomikos dėsniams. Jis yra socialinių procesų ir valstybės, verslininkų ir profsajungų sutarčių rezultatas. Esant žemam pajamų lygiui apskritai ir palyginti didelei pajamų diferenciacijai daugėja žmonių, gaunančių labai mažas pajamas. Tokios situacijos padarinys – aukštas bendras skurdo lygis bei vaikus auginančių šeimų skurdas. Tai daro didelę neigiamą įtaką valstybės demografinėi situacijai, naujosios kartos ugdymui ir yra viena iš pagrindinių nusikalstamumo didėjimą skatinančių priežasčių.

Statistiškai tiriant gyventojų pajamas ir išlaidas išskiriamos šios dalys:

- gyventojų pajamų ir išlaidų dydžio, struktūros ir kitimo vertinimas;
- pajamų ir išlaidų pasiskirstymo vertinimas;
- gyventojų pasiskirstymo pagal pajamas ir išlaidas vertinimas;
- gyventojų pajamų modelių sudarymas;
- gyventojų ir namų ūkių skurdo lygio vertinimas.

Gyventojų pajamų ir išlaidų dydis bei struktūra vertinama naudojantis dviem nepriklausomais informacijos šaltiniais:

- a. Gyventojų finansinių deklaracijų pajamų ir išlaidų balansu;
- b. Namų ūkių biudžeto tyrimų rezultatais (arba panašaus tipo tyrimų rezultatais)

Gyventojų *finansinėse deklaracijose* pajamos skirstomos į šias pagrindines rūšis:

- su darbo santykiais susijusios pajamos;
- autorinis atlyginimas;
- pensija;
- dovanotos lėšos (turtas);
- paveldėtos lėšos ir turtas;
- užsienyje ar iš užsienio valstybės gautos lėšos (turtas);
- turto pardavimo pajamos;
- turto nuomos pajamos;
- draudimo išmokos;
- stipendijos;
- kitos pajamos.

Namų ūkių biudžeto tyrimo objektas – namų ūkis, t. y. vienas asmuo ar asmenų grupė, kuri gyvena viename bute (name), turi bendrą

biudžetą ir kartu maitinasi. Namų ūkiai atrenkami atsitiktinių imčių metodu, iš gyventojų registro. Tokia atranka užtikrina vienodas galimybes visų visuomenės sluoksnių atstovams būti atrinktiems.

Atrinkti namų ūkiai tyrime dalyvauja tik vieną mėnesį. Po mėnesio jie keičiami naujais. Tyrime dalyvaujantys namų ūkiai specialiose anketose užrašo visas savo pajamas, išlaidas ne maisto prekėms ir paslaugoms apmokėti. Išlaidas maistui tiriamieji fiksuoja 15 dienų (pusė namų ūkių – pirmą mėnesio pusę, kita pusė – antrą). Suvedant informaciją išlaidos maistui perskaičiuojamos mėnesiui.

Dalis atrinktų namų ūkių tyrimuose nedalyvauja. Pavyzdžiui, 2001 metais namų ūkių biudžeto tyrimui buvo atrinkta 10560 namų ūkių, iš kurių tyrime dalyvavo 8001, t. y. 76 procentai. Penkiuose didžiausiuose miestuose nedalyvavo 36 procentai, kituose miestuose tyrime nedalyvavo 20 procentų, o kaime – 13 procentų atrinktų namų ūkių.

Pagrindinės sąvokos, vartojamos tiriant namų ūkių biudžetą

Namų ūkio tipas nustatomas pagal namų ūkio demografinę padėtį. Išskiriami šie tipai:

Vienišas asmuo;

Vienas suaugęs asmuo su vaikais iki 18 metų;

Sutuoktinių pora su vaikais iki 18 metų;

Kiti namų ūkiai su vaikais iki 18 m. Šiam namų ūkio tipui priklauso namų ūkiai, kuriuos sudaro tėvai su vaikais iki 18 m. ir vyresniais, kelių kartų namų ūkiai su vaikais iki 18 m., seneliai su vaikais iki 18 m. ir pan.

Sutuoktinių pora be vaikų;

Kiti namų ūkiai be vaikų.

Deciliai skaičiuojami į dešimt lygių dalių padalijus eilutę, kurią sudaro tiriami asmenys, išdėstyti didėjimo tvarka pagal vartojimo išlaidų lygį. Pirmąjį dešimtadalį, t. y. pirmąjį decilį, sudaro namų ūkiai, kurių narių išlaidos yra mažiausios, antrąjį – namų ūkiai, kurių išlaidos yra didesnės nei pirmojo decilio tiriamųjų, bet mažesnės nei trečiojo, ir t. t.

Disponuojamos pajamos – tai visos piniginės ir natūrinės pajamos, gautos už darbą, iš ūkininkavimo, verslo, amatų, laisvosios prekybos, laisvosios profesinės veiklos, taip pat pensijos, įvairios pašalpos, stipendijos, pajamos iš turto, renta.

Samdomojo darbo pajamos – tai darbo užmokestis pinigais ir natūra, premijos, įvairios priemokos, dienpinigiai, gauti komandiruotėms, įvairios kompensacijos, kurias moka darbovietė, ir pan.

Nesamdomojo darbo pajamos – tai pajamos iš verslo, amatų, laisvosios profesinės veiklos, žemės ūkio gamybos ir kitokios veiklos. Skaičiuojant nesamdomojo darbo pajamas imamos tik grynosios pajamos, t. y. iš gautų pajamų atėmus einamąsias išlaidas medžiagoms, žaliavoms, energijai, įrankiams, verslo mokesčiams ir pan.

Pajamos iš žemės ūkio – tai piniginės ir natūrinės pajamos, gautos iš žemės ūkio. Natūrinės žemės ūkio pajamos – tai visa žemės ūkio produkcija, kurią namų ūkis pats suvartojo ar atidavė kitiems.

Kitų (ne verslo) veiklų pajamos – tai pajamos, gautos pardavus surinktus grybus, uogas, antrines žaliavas, natūrinės įplaukos iš medžioklės, žūklės, grybavimo.

Renta – pajamos, kurias gauna namų ūkiai nuomodami nekilnojamoji turtą (namą, butą, garažą ir pan.), žemę arba kitoki turtą.

Senatvės pensijos – tai senatvės, valstybinės pensijos bei pensijos, gaunamos iš kitų šalių.

Kitos pašalpos, pensijos – tai invalidumo, šalpos (socialinės) pensijos, našlaičių (maitintojo netekimo) pensijos, ligos pašalpos, stipendijos, įvairios pašalpos, susijusios su motinyste ir kūdikio gimimu, laidojimo pašalpos ir t. t.

Kitos pajamos – tai alimentai, pinigai, gauti pardavus daiktus, laimėjimai loterijoje ar kitokiame žaidime, išmokos, gautos už gyvybės, nekilnojamojo turto, namų įrangos ir transporto priemonių draudimą, iš tėvų, vaikų, giminių ar draugų gauti pinigai, nemokamai gautos paslaugos, maisto produktai, ne maisto prekės ir pan.

Vartojimo išlaidos – tai piniginės ir natūrinės išlaidos, skirtos namų ūkių vartojimo poreikiams tenkinti, t. y. išlaidos maistui, drabužiams, avalynei, būstui, sveikatos priežiūrai, kultūros, poilsio reikmėms.

Gyventojų ir namų ūkių pajamų ir išlaidų pasiskirstymo apibūdinimas

Apibūdinant remiamasi šiais rodikliais:

1. Struktūriniais koeficientais, apskaičiuotais remiantis gyventojų pajamų (išlaidų) skirtumais. Pasaulinėje praktikoje dažnai taikomi deci-

liniai, kvartiliniai ir kvantiliniai diferenciacijos koeficientai, socialinio susisluoksniavimo koeficientas ir kt. Taikant decilinį koeficientą gyventojų skaičius dalijamas į dešimt lygių dalių, taikant kvartilinį koeficientą – į keturias dalis, taikant kvantilinį – į penkias. Grupės suskirstomos tokia eile pagal pajamas: pirmoji – su mažiausiomis pajamomis, paskutinioji – su didžiausiomis pajamomis. Vienas iš dažniausiai naudojamų būdų diferenciacijai vertinti yra sukauptų (kumuliatyvinių) dalių nustatymas eilėje pagal pajamas ir eilėje pagal gyventojų skaičių.

2. Rodikliais, vertinančiais visų gyventojų pasiskirstymą pagal pajamas (išlaidas). Tai Lorenzo (Lorenz) kreivė ir koeficientas, Džini (Gini) koeficientas ir kt.

3. Gyventojų suskirstymu į grupes su panašia perkamąja galia (pagal pragyvenimo minimumą).

3.1 pavyzdys. Decilinis ir kvartilinis diferenciacijos koeficientai, Lorenzo kreivė, Lorenzo ir Džini koeficientai

Lietuvos statistikos departamento duomenimis, 1996 m. vidutinės mėnesinės išlaidos deciliuose buvo tokios (žr. 3.1 lentelę):

3.1 lentelė. Vidutinės mėnesinės išlaidos deciliuose, 1996 m.

Decilio nr.	Vidutinės išlaidos deciliuose, Lt	Gyventojų dalis
1	105,6	0,1
2	160,7	0,1
3	199	0,1
4	235,7	0,1
5	273,5	0,1
6	313,9	0,1
7	364,2	0,1
8	428,9	0,1
9	530,6	0,1
10	920,7	0,1

3.1.1 užduotis. Apskaičiuokite decilinį ir kvartilinį diferenciacijos koeficientus, Lorenzo bei Džini koeficientus ir palyginkite gyventojų išlaidų struktūrą pagal pragyvenimo minimumą.

Metodiniai nurodymai

Decilinis diferenciacijos koeficientas yra kraštinių decilių santykis, parodantis, kiek kartų 10 procentų turtingiausių gyventojų pajamos (išlaidos) viršija 10 procentų skurstančių gyventojų pajamas (išlaidas):

$$K_d = \frac{D_{10}}{D_1} = 920,7/105,6 = 8,72$$

1996 metais vidutinės 10 proc. turtingiausių gyventojų pajamos viršijo 10 proc. skurdžiausiai gyvenančių gyventojų pajamas 8,72 karto.

Kvartiliai variacinę eilutę dalija į keturias maždaug lygias dalis ir žymimi Q_1 , Q_2 , Q_3 . Q_2 sutampa su mediana ir dalija imtį į dvi dalis. Tuomet Q_1 yra apatinės dalies mediana, o Q_3 – viršutinės dalies mediana. Apytikriai medianą Q_1 galime prilyginti trečio decilio vidurkiui, o Q_3 – aštunto decilio vidurkiui. Tada kvartilinis diferenciacijos koeficientas

$$K_{kv} = \frac{Q_3}{Q_1} = 428,9/199 = 2,155.$$

Kvartilinis koeficientas apytikriai atspindi vidurinėsios klasės diferenciaciją.

Šie koeficientai skaičiuojami tik iš dviejų decilių reikšmių, todėl neatspindi visų likusių decilių. Korektiškesni yra rodikliai, apskaičiuoti iš pajamų (išlaidų) pasiskirstymo visuose deciliuose. Dažniausiai naudojami rodikliai yra Lorenzo kreivė, Lorenzo ir Džini koeficientai.

Lorenzo kreivė padeda įvertinti, kaip lygiai pasiskirsto pajamų, turto ar kitokių rodiklių reikšmės intervaluose. Ši kreivė yra sukauptų pajamų dalių (dažnių) grafinis vaizdas. Brėžiant Lorenzo kreivę tiriama seka išrikiuojama didėjimo tvarka $x_1 < x_2 < \dots < x_n$ ir apskaičiuojami

taškai $(h/n, L_h/L_n)$, kur $h=0, 1 \dots n$, $L_0=0$ ir $L_h = \sum_{i=1}^h x_i$, n – intervalų skai-

čius. Jei visos sekos reikšmės yra vienodo dydžio, Lorenzo kreivė yra lygi įstriža linija, vadinama lygybės linija. Brėžiant Lorenzo kreivę veiksmai atliekami šia tvarka:

1. Kiekvienam kintamajam (rodikliui) apskaičiuojami santykiniai dažniai, t. y. sumos dalys.
2. Stebėjimai išdėstomi didėjimo tvarka pagal kintamąjį x .

3. Sudaroma sukauptųjų dažnių lentelė kiekvienam išrikiuotam kintamajam.

4. Nubrėžiama diagrama ir tiesi diagonalinė linija.

Lorenzo kreivė rodo, kurią pajamų dalį gauna tam tikra visuomenės dalis. Įstriža linija rodo idealiai lygų pajamų pasiskirstymą. Kuo toliau Lorenzo kreivė nukrypusi nuo įstrižos linijos, tuo nelygesnis pajamų pasiskirstymas. Ši kreivė gali būti transformuota į Lorenzo arba Džini koeficientą. Lorenzo koeficientas vertina Lorenzo kreivės nuokrypius ir apskaičiuojamas pagal formulę:

$$L = \frac{1}{2} \sum_{i=1}^n |x_i - y_i|. \quad (3.1)$$

Čia x_i ir y_i – santykiniai dažniai, n – intervalų skaičius.

Džini koeficientas paprastai žymimas raide G ir lygus santykiui

$$G = \frac{A}{A+B}. \quad (3.2)$$

Čia A yra sritis tarp diagonalės ir Lorenzo kreivės, o B yra sritis, ribojama Lorenzo kreivės, taigi $A+B$ yra sritis, ribojama diagonalės. Kai pajamos pasiskirstę idealiai lygiai, ir srities A plotas, ir Džini koeficientas lygus nuliui. Kai pasiskirstymas absoliučiai neteisingas, B lygus nuliui, o Džini koeficientas lygus vienetui. Praktiškai, kai Džini koeficiento reikšmė didesnė nei 0.30, galima teigti, kad yra esminių pajamų paskirstymo netolygumų.

Kai dydžiai x išrikiuojami didėjimo tvarka, G gali būti apskaičiuojamas pagal formulę

$$G = \frac{\sum_{i=1}^n (2 \cdot i - n - 1) \cdot x_i}{n^2 \cdot \mu}. \quad (3.3)$$

Čia i – decilio eilės numeris, n – decilių skaičius ($n=10$), μ – vidutinių išlaidų deciliuose aritmetinis vidurkis.

Reikia pastebėti, kad Džini koeficientas neapima visos informacijos apie pajamų pasiskirstymą. Dviems skirtingoms Lorenzo kreivėms apskaičiuotas Džini koeficientas gali būti vienodas. Pagrindinė šio koeficiento savybė yra ta, kad vieno kurio nors gyventojų sluoksnio pajamų

didėjimas didina Džini koeficiento reikšmę, tuo tarpu kitų sluoksnių pajamos nesikeičia.

Tarpiniai rodiklių skaičiavimo rezultatai pateikti 4.2 lentelėje. Brėždami Lorenzo kreivę (žr. 3.1 pav.) horizontalioje ašyje atidėsime sukauptuosius namų ūkių dažnius, o vertikalioje ašyje – sukauptąsias išlaidas $cum F_p$ ir $cum F_t$ (žr. 3.2 lentelę).

3.2 lentelė. Tarpiniai skaičiavimo rezultatai

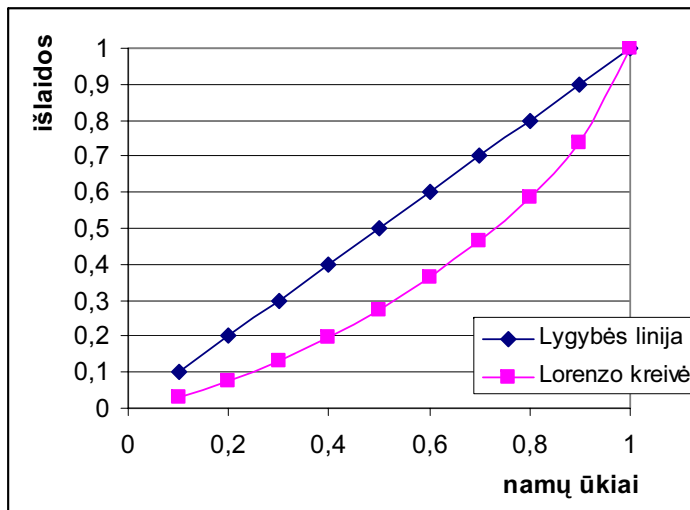
Decilio nr. i	Vidutinės išlaidos deciliuose, Lt x_i	Gyventojų dalis Fp_i	$cum Fp$	$x_i * Fp_i = D_i$	$Ft_i = \frac{D_i}{\sum_{i=1}^n D_i} = \frac{D_i}{353,28}$	$cum Ft$	$(2 \cdot i - n - 1) \cdot x_i$
1	105,6	0,1	0,1	10,56	0,030	0,030	-950,4
2	160,7	0,1	0,2	16,07	0,045	0,075	-1124,9
3	199	0,1	0,3	19,9	0,056	0,131	-995
4	235,7	0,1	0,4	23,57	0,067	0,198	-707,1
5	273,5	0,1	0,5	27,35	0,077	0,275	-273,5
6	313,9	0,1	0,6	31,39	0,089	0,364	313,9
7	364,2	0,1	0,7	36,42	0,103	0,467	1096,6
8	428,9	0,1	0,8	42,89	0,121	0,588	2144,5
9	530,6	0,1	0,9	53,06	0,150	0,738	3714,2
10	920,7	0,1	1,0	92,07	0,261	1,0	8286,3
Suma	3532,8			353,28	1,0		11500,6

Kaip matome 3.1 paveiksle, tikrasis išlaidų pasiskirstymas labai skiriasi nuo idealaus. Didžiausi skirtumai yra kreivės viduryje, t. y. penktame, šeštame ir septintame deciliuose. Šiuose deciliuose ryškiausiai nukrypimai nuo lygaus pajamų pasiskirstymo.

Lorenzo koeficientą apskaičiuojame pagal (3.1) formulę

$$L = \frac{\sum_{i=1}^n |Ft_i - Fp_i|}{2}, \quad (3.4)$$

čia F_p – dalis gyventojų, kurių vidutinės išlaidos pateko į decilio intervalą;



3.1 pav. Lorenzo kreivė ir lygybės linija

F_t – dalis gyventojų, patekusių į decilį, kai išlaidos deciliuose apskaičiuotos pagal formulę

$$Ft_i = \frac{x_i \cdot Fp_i}{\sum(x_i \cdot Fp_i)} \quad (3.5)$$

Įrašę reikšmes iš 3.2 lentelės gauname Lorenzo koeficiento reikšmę

$$L = (|0,030 - 0,1| + |0,045 - 0,1| + |0,056 - 0,1| + \dots + |0,261 - 0,1|) / 2 = 0,235.$$

Lorenzo koeficiento reikšmės, artimos nuliui, rodo, kad gyventojų pajamos deciliuose pasiskirsčiusios tolygiai.

Džini koeficientas skaičiuojamas pagal 3.3 formulę. Nagrinėjamu atveju $n=10$. Apskaičiavę reikšmių x_i aritmetinio vidurkio įvertį gauname reikšmę 353,28. Įrašę reikšmes į formulę gauname

$$G = \frac{11500,6}{10 * 10 * 353,28} = 0,326.$$

Remdamiesi apskaičiuota Džini koeficiento reikšme galime teigti, kad gyventojų pajamų diferenciacijos lygis yra aukštas. 1991 m. Džini koeficientas atskirose valstybėse buvo toks: JAV – 0,326, Švedijoje – 0,205, Norvegijoje – 0,243, Vengrijoje – 0,244, Lotynų Amerikoje – vidutiniškai 0,495.

Analizuojant gyventojų pasiskirstymą pagal pajamas (išlaidas) svarbus rodiklis yra pragyvenimo minimumas. Jei šis rodiklis lygus 125 Lt, matome, kad 10 procentų gyventojų pajamos (išlaidos) yra mažesnės už pragyvenimo minimumą. Šis rodiklis rodo, kad 10 procentų gyventojų gyvybiškai svarbūs poreikiai nėra patenkinti. Skaičiuojant taip pat naudojami dviejų ir trijų pragyvenimo minimumo dydžio rodikliai.

Mažesnes pajamas (išlaidas) nei du pragyvenimo minimumai, t. y. $125 \cdot 2 = 250$ Lt, turi 40 procentų gyventojų ($cum F_p = 0.4$). Mažesnes nei trijų pragyvenimo lygių pajamas (išlaidas), t. y. $125 \cdot 3 = 375$ Lt, turi 70 procentų gyventojų. Šie nesudėtingi skaičiavimo rezultatai pateikti 3.3 lentelėje.

3.3 lentelė. Gyventojų pasiskirstymas pagal pragyvenimo minimumą

Pajamos	Gyventojų dalis ($cum F_p$)
Mažiau nei 1 pragyvenimo minimumas	0,1, arba 10 proc.
Mažiau nei 2 pragyvenimo minimumai	0,4, arba 40 proc.
Mažiau nei 3 pragyvenimo minimumai	0,7, arba 70 proc.
Daugiau nei 3 pragyvenimo minimumai	0,3, arba 30 proc.

Uždaviniai

1. Lietuvos statistikos departamentas pateikia šiuos duomenis (žr. 3.4 lentelę) apie gyventojų vidutines mėnesines išlaidas deciliuose 1997, 1998, 1999 ir 2000 metais.

Nubrėžkite Lorenzo kreivę ir apskaičiuokite kiekvienų metų decilinius diferenciacijos koeficientus, taip pat Lorencio ir Džini koeficientus. Palyginkite gyventojų išlaidų struktūros pagal pragyvenimo minimumą pokyčius. Remdamiesi gautais rezultatais parašykite gyventojų pasiskirstymo pagal pajamas ir išlaidas 1997–2000 metais analizę.

3.4 lentelė. Vidutinės mėnesinės išlaidos deciliuose 1997, 1998, 1999 ir 2000 metais

Deciliai	1997 m.	1998 m.	1999 m.	2000 m.
1	118,3	135,3	133,5	128,4
2	176,9	199,5	202,8	189,6
3	219,2	247,6	248,8	234,9
4	257,3	292,5	290,4	276,8
5	295,7	335,2	333,7	319,2
6	340,1	380,6	382,4	365,6
7	391,4	438,8	440,2	420,9
8	458,5	516,6	512,5	491,5
9	564,1	642,2	631,2	608,1
10	1004,6	1080,3	1077,7	1008,7

2. 3.5 lentelėje pateikti duomenys apie Rusijos gyventojų pasiskirstymą pagal vidutines mėnesines pajamas vienam asmeniui 1992 metais.

Apskaičiuokite decilinius diferenciacijos koeficientus, taip pat Lorenco ir Džini koeficientus. Skaičiuodami naudokite vidurinę vidutinių mėnesinių pajamų intervalo reikšmę. Atlikite gyventojų išlaidų struktūros pagal pragyvenimo minimumą analizę, kai pragyvenimo minimumo reikšmė lygi 980 rublių.

3.5 lentelė. Rusijos gyventojų pasiskirstymas pagal vidutines mėnesines pajamas vienam asmeniui 1992 metais

Vidutinės mėnesinės pajamos, rubliais	Gyventojų dalis, proc.
Iki 1000	4,7
1000,1–2000,0	21,8
2000,1–3000,0	23,0
3000,1–4000,0	17,0
4000,1–5000,0	13,5
5000,1–6000,0	6,5
6000,1–7000,0	4,3
7000,1–8000,0	2,9
8000,1–9000,0	1,9
9000,1–10 000,0	1,3
Daugiau nei 10 000	3,2

3. 3.6 ir 3.7 lentelėse pateiktas Lietuvos gyventojų vidutinių mėnesinių disponuojamų (piniginių ir natūrinių) pajamų ir vidutinių mėnesinių vartojimo išlaidų (piniginių ir natūrinių) pasiskirstymas deciliuose 1999 metais.

Pateiktiems namų ūkio tipams nubrėžkite Lorenzo kreivę ir apskaičiuokite kiekvieno namų ūkio decilinius diferenciacijos koeficientus, taip pat Lorenzo ir Džini koeficientus. Palyginkite gyventojų pajamų ir išlaidų struktūros pagal pragyvenimo minimumą skirtumus. Remdamiesi gautais rezultatais parašykite gyventojų pasiskirstymo pagal namų ūkio tipus analizę.

3.6 lentelė. Lietuvos gyventojų vidutinių mėnesinių disponuojamų (piniginių ir natūrinių) pajamų pasiskirstymas deciliuose vidutiniškai vienam namų ūkio nariui, Lt

Namų ūkio tipas	Deciliai									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Vienišas asmuo	150,6	293,8	341,8	383,7	428,6	481,8	552,5	648,5	820,8	1434,4
Vienas suaugęs asmuo su vaikais iki 18 m.	93,7	146,4	184,5	236,8	268,2	292,8	344,8	409,7	487,4	765,6
Poros su vaikais iki 18 m.	80,9	178,1	231,6	275,5	314,3	361,6	411,8	480,6	594,9	992,0
Kiti namų ūkiai su vaikais iki 18 m.	77,6	168,6	215,5	254,1	291,3	330,5	378,8	433,8	508,0	780,4
Poros be vaikų	149,8	289,3	343,1	368,7	430,9	478,8	546,1	647,1	804,8	1348,2
Kiti namų ūkiai be vaikų	131,3	240,0	292,0	340,6	388,2	436,9	495,4	567,7	693,9	1024,3
Lietuvoje	92,6	200,5	256,6	301,0	346,7	394,4	448,2	521,1	646,9	1072,2

3.7 lentelė. Lietuvos gyventojų vidutinių mėnesinių disponuojamų (piniginių ir natūrinių) išlaidų pasiskirstymas deciliuose vidutiniškai vienam namų ūkio nariui, Lt

Namų ūkio tipas	Deciliai									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Vienišas asmuo	204,2	292,5	352,1	405,9	457,7	517,0	598,8	694,2	855,3	1469,1
Vienas suaugęs asmuo su vaikais iki 18 m.	111,8	163,3	195,4	231,8	264,2	306,1	359,2	425,6	532,0	819,9
Poros su vaikais iki 18 m.	126,6	190,2	233,4	271,7	309,7	353,6	413,2	482,0	589,4	974,2
Kiti namų ūkiai su vaikais iki 18 m.	110,8	167,7	210,7	244,2	276,0	312,7	358,1	412,9	501,5	801,5
Poros be vaikų	196,9	278,7	327,9	371,5	419,4	463,8	521,2	610,0	756,2	1358,3
Kiti namų ūkiai be vaikų	152,1	222,6	267,1	309,5	351,9	397,0	452,7	522,7	634,5	1045,3
Lietuvoje	133,5	202,8	248,8	290,4	333,7	382,4	440,2	512,5	631,2	1077,7

4. INTEGRUOTI VISUOTINĖS GEROVĖS RODIKLIAI

Iki 1970 metų BVP buvo vyraujantis ekonominio progreso ir socialinės gerovės matas. Šiuo laikotarpiu tapo akivaizdu, kad vertinant gyvenimo lygį atskirose valstybėse nepakanka įvertinti tik sukurtą prekių ir paslaugų kiekį. Tai sudarė sąlygas atsirasti dviejų tipų rodikliams, naudojamiems atliekant tarptautinius palyginimus. Vertinant BVP vienam gyventojui buvo pradėta skaičiuoti gyventojų perkamoji galia arba tikrasis BVP *per capita*. Kita vertus, buvo sukurti papildomi matai raidos pažangai vertinti. Vieną iš pirmųjų gerovės vertinimo kriterijų 1979 m. sukūrė profesorius Morris, pavadindamas jį fizinės gyvenimo kokybės rodikliu (*Physical Quality of Life Index- PQLI*). Šis rodiklis vertina galimybę ilgai gyventi, vaikų mirtingumą ir raštingumą. Apskaičiuojant šį rodiklį statistikos standartizuojamos perskaičiuojant į intervalą nuo 0 iki 1. Pavyzdžiui, aukščiausiam vaikų mirtingumui priskiriama reikšmė 0, žemiausiam – 1. Visos kitos reikšmės irgi patenka į šį intervalą. Galiausiai visos šių trijų standartizuotų rodiklių reikšmės su vienodais svoriais sudedamos ir taip gaunamas šalies *PQLI* indeksas. Iš tikrųjų šis indeksas matuoja ne tiek socialinę pažangą, kiek gyventojų pagrindinių poreikių patenkinimą.

Vienas iš svarbiausių šiuo metu naudojamų rodiklių yra socialinės pažangos indeksas (*Human Development Index – HDI*), kuris pirmą kartą pasirodė 1990 m. ataskaitoje apie socialinę pažangą (*Human Development Report – HDR*), pateiktoje JT plėtros programoje (*United Nations Development Programme – UNDP*). Pagal šią programą kasmet pateikiamoje socialinės pažangos ataskaitoje minimi socialiniai rodikliai, priešingai nei Pasaulio banko ataskaitoje apie pasaulio pažangą (*World Development Report – WDR*), kurioje dėmesys skiriamas ekonominiams rodikliams.

Socialinės pažangos indeksas (*HDI*) glaudžiai susijęs su BVP vienam gyventojui, tačiau atskiroms valstybėms jo reikšmės labai skiriasi. Tai paaiškinama valstybių socialinių politikų skirtumais ir skirtingu pajamų pasiskirstymu.

Be šių indeksų, dar vartojami:

- Išsivystymo indeksas, įvertinantis lyčių faktorių (*Gender-related Development Index - GDI*)

- Technologijų išsivystymo indeksas (*Technology Achievement Index – TAI*)
- Gyventojų skurdo indeksai (*Human Poverty Index – HPI₁* ir *HPI₂*)
- Moterų galimybių didėjimo rodikliai ir kt.

Iš šių rodiklių dažniausiai taikomi socialinės pažangos indeksas (*HDI*) ir gyventojų skurdo indeksai (*HPI₁* ir *HPI₂*).

Socialinės pažangos indeksas (HDI) yra sudėtinis indeksas, kuriuo apibendrinami trys pagrindiniai gyventojų gyvenimo kokybės komponentai:

1. Galimybė ilgai gyventi, matuojama tikėtina gyvenimo trukme gimstant (p_1).
2. Informacijos gyventojams prieinamumas (galimybė gauti mokyklinį išsilavinimą ir profesinį parengimą), kuri apibūdina du rodikliai: suaugusių gyventojų raštingumas (p_2) ir besimokančio jaunimo dalis iš bendro jaunimo skaičiaus (p_3)
3. Ekonominės gyventojų galimybės, atsispindinčios bendrajame vidaus produkte (BVP), apskaičiuotame vienam gyventojui (*per capita*). Tai atspindi galimybę pasiekti tokį pajamų lygį, kad gyventojai galėtų apsirūpinti visomis būtinomis gerybėmis. Šis rodiklis paprastai transformuojamas į perkamąją galią, kuri tiksliau išmatuoja gyvenimo standartą, nes įvertina atskirų prekių kainas įvairiose šalyse, pavyzdžiui, kiek mokama už kepalą duonos. Šį rodiklį pažymėkime p_4 .

Vienas iš šių indeksų komponentų yra *gyventojų raštingumas*. Raštingumo turinys įvairiais visuomenės ekonominės raidos etapais nebuvo vienodas. Besivystančiose šalyse raštingais laikomi asmenys, mokantys skaityti, išsivysčiusiose – mokantys ir skaityti, ir rašyti. UNESCO generalinė konferencija 1958 m. rekomendavo gyventojų surašymo metu raštingais laikyti asmenis, mokančius skaityti, suprasti skaitomą tekstą ir galinčius parašyti rašinėlių iš kasdienio gyvenimo.

Kiekvienam šių rodiklių (p_1, p_2, p_3, p_4) sudaromas komponentinis indeksas pagal formulę

$$i_p = \frac{p_{fakt} - p_{\min}}{p_{\max} - p_{\min}}. \quad (4.1)$$

Čia p_{fakt} – faktinė rodiklio reikšmė tam tikroje šalyje;

p_{min} ir p_{max} – atitinkamai minimali ir maksimali kiekvieno rodiklio reikšmė (ribinė reikšmė).

Skaičiuojant p_4 indeksą (tikrasis BVP vienam gyventojui JAV doleriais) pagal perkamosios galios paritetą skaičiuojama šio rodiklio natūrinio logaritmo reikšmė:

$$i_{BVP} = \frac{\ln p_{4\text{fakt}} - \ln p_{4\text{min}}}{\ln p_{4\text{max}} - \ln p_{4\text{min}}}. \quad (4.2)$$

Skaičiuojant komponentinius indeksus naudojamos šios ribinės reikšmės (žr. 4.1 lentelę):

4.1 lentelė. Ribinės rodiklių reikšmės

Rodiklis	p_{min}	p_{max}
Tikėtina gyvenimo trukmė, metais	25	85
Suaugusių gyventojų raštingumo lygis, procentais	0	100
Besimokančio jaunimo dalis, procentais	0	100
Tikrasis BVP <i>per capita</i> pagal perkamosios galios paritetą, JAV dol.	100	40000

Įvertinant šias ribines reikšmes trys komponentiniai indeksai skaičiuojami pagal tokias formules:

- tikėtinos gyvenimo trukmės gimstant indeksas

$$i_{p1} = \frac{p_{1\text{fakt}} - 25}{85 - 25}. \quad (4.3)$$

- suaugusių gyventojų raštingumo indeksas

$$i_{p2} = \frac{p_{2\text{fakt}} - 0}{100 - 0}; \quad (4.4)$$

- besimokančio jaunimo dalies indeksas

$$i_{p3} = \frac{p_{3\text{fakt}} - 0}{100 - 0}. \quad (4.5)$$

- pasiekto išsilavinimo lygio indeksas apibendrina du nagrinėtus indeksus

$$i_{apib} = (2 \cdot i_{p2} + i_{p3}) : 3. \quad (4.6)$$

- tikrojo BVP *per capita* pagal perkamosios galios paritetą indeksas

$$i_{p4} = \frac{\ln p_{4\text{ fakt}} - \ln 100}{\ln 40000 - \ln 100} = \frac{\ln p_{4\text{ fakt}} - 4.605}{10.597 - 4.605} = \frac{\ln p_{4\text{ fakt}} - 4.605}{5.992}. \quad (4.7)$$

- Taigi socialinės pažangos indeksas apibendrina tris komponentinius indeksus

$$HDI = (i_{p1} + i_{apib} + i_{p4}) : 3. \quad (4.8)$$

Išsivystymo indeksas, įvertinantis lyčių faktorių (GDI), įvertina lyčių lygybę trimis socialinės pažangos aspektais. GDI skaičiuojamas pagal tą pačią metodiką kaip ir socialinio išsivystymo indeksas HDI. Kitaip nei HDI, šis indeksas įvertina kiekvienos šalies laimėjimus gyvenimo trukmės, raštingumo ir pajamų aspektais siekiant vyrų ir moterų lygybės. Atliekant tarptautinius palyginimus naudojami abu indeksai, o valstybių išrikiavimas pagal HDI ir GDI rodo lyčių išsivystymo netolygumą.

Gyventojų skurdo indeksas apibūdina gyventojų vargingumo lygį. Jis vertina tris svarbiausias žmogaus gyvenimo sritis: gyvenimo trukmę, žinias ir aprūpinimą ekonominiais ištekliais. Šis indeksas skaičiuojamas pagal metodikas, sukurtas atskirai besivystančioms šalims (HPI_1) ir išsivysčiusioms šalims (HPI_2).

Besivystančių šalių gyventojų skurdo indeksas (HPI_1) skaičiuojamas vertinant šiuos rodiklius:

- p_1 – procentinė dalis gyventojų, kurių tikėtina gyvenimo trukmė neviršys 40 metų;
- p_2 – neraštingų suaugusių gyventojų procentinė dalis;
- p_3 – gyventojų aprūpinimas ekonominiais ištekliais, kurį apibūdina du rodikliai:
 - p_{31} – procentinė gyventojų, kurie neturi galimybės apsirūpinti tinkamos kokybės geriamuoju vandeniu, dalis;
 - p_{32} – procentinė vaikų iki 5 metų amžiaus, kurių svoris daug mažesnis už normalų, dalis.

Bendras gyventojų aprūpinimo ekonominiais ištekliais rodiklis yra šių dviejų smulkesnių rodiklių vidurkis

$$p_3 = \frac{p_{31} + p_{32}}{2}. \quad (4.9)$$

Besivystančių šalių gyventojų skurdo indeksas skaičiuojamas pagal formulę

$$HPI_1 = \sqrt[3]{\frac{1}{3}(p_1^3 + p_2^3 + p_3^3)}. \quad (4.10)$$

Išsivysčiusių šalių gyventojų skurdo indeksas (HPI₂) apibendrina keturių rodiklių teikiamą informaciją:

- p_1 – procentinė gyventojų, kurių tikėtina gyvenimo trukmė neviršys 60 metų, dalis;
- p_2 – procentinė neraštingų suaugusių gyventojų dalis;
- p_3 – gyventojų, gyvenančių žemiau skurdo ribos, procentinė dalis (tai gyventojai, kurių pajamos mažesnės nei 50 proc. visų gyventojų pajamų medianos);
- p_4 – bedarbių, neturinčių darbo ilgą laikotarpį (12 mėn. ir daugiau) procentinė dalis iš bendros darbo jėgos dalies.

Išsivysčiusių šalių gyventojų skurdo indekso reikšmė apskaičiuojama pagal formulę

$$HPI_2 = \sqrt[3]{\frac{1}{4}(p_1^3 + p_2^3 + p_3^3 + p_4^3)}. \quad (4.11)$$

Kaip skaičiuojami šie gyvenimo lygį apibendrinantys rodikliai, parodysime tolesniuose pavyzdžiuose.

4.1 pavyzdys. Socialinės pažangos indeksas

Apskaičiuokime Lietuvos socialinės pažangos indeksą remdamiesi 1996 metų duomenimis (žr. 4.2 lentelę).

Komponentinių indeksų reikšmės:

$$i_{p1} = \frac{70,3 - 25}{85 - 25} = 0,755;$$

$$i_{p2} = \frac{99,8 - 0}{100 - 0} = 0,998;$$

$$i_{p3} = \frac{48 - 0}{100 - 0} = 0,48;$$

$$i_{apib} = [2 \cdot 0,998 + 0,48] : 3 = 0,825;$$

$$i_{p4} = \frac{\ln 1701 - 4,605}{5,992} = 0,473$$

Galutinė indekso reikšmė

$$HDI = \frac{0,755 + 0,825 + 0,473}{3} = 0,683$$

4.2 lentelė. 1996 metų Lietuvos rodikliai

Rodiklis	p _{fakt}
Tikėtina gyvenimo trukmė, metais	p _{vyrų} =64,47; p _{moterų} =76,0; p ₁ =70,3
Suaugusių gyventojų raštingumo lygis, proc.	p ₂ =99,8
Besimokančio jaunimo dalis, proc.	p ₃ =48% (EUROSTAT duomenys)
Tikrasis BVP <i>per capita</i> , JAV dol., palyginamosiomis kainomis	p ₄ =1701

Palygindami apskaičiuokime Danijos socialinės pažangos indeksą remdamiesi 1995 metų duomenimis (žr. 4.3 lentelę).

4.3 lentelė. 1995 m. Danijos duomenys

Rodiklis	p _{fakt}
Tikėtina gyvenimo trukmė, metais	p ₁ =75.3
Suaugusių gyventojų raštingumo lygis, proc.	p ₂ =99
Besimokančio jaunimo dalis, proc.	p ₃ =89
Tikrasis BVP <i>per capita</i> , JAV dol., palyginamosiomis kainomis	p ₄ =21963

Komponentinių indeksų reikšmės:

$$i_{p1} = \frac{75,3 - 25}{85 - 25} = 0,838;$$

$$i_{p2} = \frac{99 - 0}{100 - 0} = 0,99;$$

$$i_{p3} = \frac{89-0}{100-0} = 0,89;$$

$$i_{apib} = (2 \cdot 0,99 + 0,89) : 3 = 0,957;$$

$$i_{p4} = \frac{\ln 21983 - 4,605}{5,992} = 0,900.$$

Galutinė indekso reikšmė

$$HDI = \frac{0,838 + 0,957 + 0,900}{3} = 0,898.$$

Lietuvos ir Danijos socialinės pažangos indeksų santykis lygus $0,898/0,683=1,315$. Tai reiškia, kad Danijos gyventojų galimybės išvystytos 1,315 karto daugiau nei Lietuvos.

4.2 pavyzdys. Besivystančių šalių skurdo indeksas

Apskaičiuokite Kongo gyventojų skurdo indeksą (HPI_I), kai žinomi šie duomenys (žr. 4.4 lentelę):

4.4 lentelė. **Kongo duomenys**

Rodiklis	P_{fakt}
Gyventojai, kurių tikėtina gyvenimo trukmė neviršys 40 metų, proc.	32
Neraštingų suaugusių gyventojų procentinė dalis	25,1
Gyventojai, kurie neturi galimybės apsirūpinti tinkamos kokybės geriamuoju vandeniu, proc.	66
Vaikai iki 5 metų amžiaus, kurių svoris mažesnis už normalų, proc.	24

$$\text{Apskaičiuokime } p_3 = (66 + 24) : 2 = 45;$$

$$\text{Tada } HPI_I = (1/3(32^3 + 25,1^3 + 45^3))^{1/3} = 35,98.$$

Tai rodo, kad 35,98 proc. Kongo gyventojų gyvena skurdžiai ir neturi galimybių gyventi žmoniškomis sąlygomis.

4.3 pavyzdys. Išsivysčiusių šalių skurdo indeksas

Apskaičiuokime Australijos gyventojų skurdo indeksą (HPI_2) remdamiesi šiais duomenimis (žr. 4.5 lentelę):

4.5 lentelė. Australijos duomenys

Rodiklis	P _{fakt}
Gyventojai, kurių tikėtina gyvenimo trukmė neviršys 60 metų	9
Neraštingų suaugusių gyventojų procentinė dalis	17
Gyventojai, kurių pajamos mažesnės nei 50 proc. visų gyventojų pajamų medianos	12.9
Bedarbių, nedirbančių ilgiau nei 1 metai, procentinė dalis iš bendros darbo jėgos	2.6

Australijos gyventojų skurdo indeksas

$$HPI_2 = [1/4(9^3 + 17^3 + 12,9^3 + 2,6^3)]^{1/3} = 12,5.$$

Tai reiškia, kad 12,6 proc. Australijos gyventojų gyvena žemiau skurdo ribos.

Uždaviniai

1. Lentelėje 4.6 pateikti duomenys apibūdina devintojo dešimtmečio vidurio gyvenimo kokybę kai kuriose išsivysčiusiose pasaulio šalyse.

Remdamiesi šiais duomenimis:

1.1. Apskaičiuokite kiekvienos valstybės socialinės pažangos indeksus (komponentinius indeksus ir bendrąjį indeksą HDI).

1.2. Atlikite visų šių šalių duomenų analizę atskirai pagal kiekvieną rodiklį.

1.3. Apskaičiuokite Spirmeno (Spearman) koreliacijos koeficiento reikšmes, apibūdinančias bendrojo socialinės pažangos indekso HDI priklausomybę nuo komponentinių indeksų rangų. Kuris komponentinis indeksas turi didžiausią įtaką indeksui HDI išsivysčiusių šalių grupėje?

4.6 lentelė. Duomenys apie kai kurių išsivysčiusių pasaulio šalių gyvenimo kokybę devintojo dešimtmečio viduryje

Valstybė	Tikėtina gyvenimo trukmė gimstant, metais	Suaugusių gyventojų raštingumo lygis, proc.	Besimokančio jaunimo dalis, proc.	BVP <i>per capita</i> , JAV dol., palyginamosiomis kainomis
Australija	78,2	99,0	79	19632
Belgija	76,9	99,0	86	21548
Austrija	76,7	99,0	87	21322
Ispanija	77,7	97,1	90	14789
Graikija	77,9	96,7	82	11636
Airija	76,4	99,0	88	17590
Italija	78,0	98,1	73	20174
Olandija	77,5	99,0	91	19876
Naujoji Zelandija	76,6	99,0	94	17267
Norvegija	77,6	99,0	92	22427
Kanada	79,1	99,0	100	21916
Islandija	79,2	99,0	83	21064
JAV	76,4	99,0	96	26977
Prancūzija	78,7	99,0	89	21176
Japonija	79,9	99,0	78	21930

Spirmeno koreliacijos koeficientas skaičiuojamas pagal formulę

$$r_S = 1 - \frac{6D^2}{n(n^2 - 1)}. \quad (4.12)$$

Čia $D^2 = \sum_{i=1}^n (r_i - s_i)^2$, r_1, r_2, \dots, r_n – vieno rodiklio rangų seka, o

s_1, s_2, \dots, s_n – kito rodiklio rangų seka; n yra stebėjimų (šiuo atveju valstybių) skaičius.

Spirmeno koreliacijos koeficientas kinta nuo +1 iki –1 (kaip ir porinis (Pearson) koreliacijos koeficientas). Kuo koeficiento reikšmė arčiau nulio, tuo ryšys silpnesnis.

Kai išmatuotų reikšmių kiekis $n \geq 10$, Spirmeno koreliacijos koeficiento skirstinys yra apytikriai Stjudento skirstinys su $n-2$ laisvės laipsniais, todėl hipotezei apie šio koreliacijos koeficiento lygybę nuliui tik-

rinti galime naudoti statistiką (kriterijų) $T = r_S \sqrt{\frac{n-2}{1-r_S^2}}$.

Sprendimo priėmimo taisyklė tokia: parenkame reikšmingumo lygmenį $\alpha=0,05$. Hipotezė H_0 atmetama, jei absoliuti T reikšmė viršija dydį $t_{\alpha/2(n-2)}$, kuris yra Stjudento skirstinio su $(n-2)$ laisvės laipsniais $\alpha/2$ lygmens kritinė reikšmė.

Statistikos $t_{\alpha/2}$ reikšmes galime rasti pasitelkę statistinės duomenų analizės programą STATISTICA, pasinaudodami tikimybių pasiskirstymo skaičiuotuvais.

Tuo atveju, kai $n \leq 10$, hipotezei apie šio koreliacijos koeficiento reikšmingumą tikrinti naudojamos specialios lentelės arba statistinės duomenų analizės programos.

2. 4.7 lentelėje pateikti duomenys, apibūdinantys devintojo dešimtmečio vidurio gyvenimo kokybę besivystančiuose šalyse.

4.7 lentelė. **Duomenys apie kai kurių besivystančių pasaulio šalių gyvenimo kokybę devintojo dešimtmečio viduryje**

Valstybė	Tikėtina gyvenimo trukmė gimstant, metais	Suaugusių gyventojų raštingumo lygis, proc.	Besimokančio jaunimo dalis, proc.	BVP per capita, JAV dol., palyginamosiomis kainomis
Argentina	72,6	96,2	79	8498
Brazilija	66,6	83,3	61	5928
Venesuela	72,3	91,1	67	8090
Singapūras	77,1	91,1	68	22604
Kolumbija	70,3	91,3	69	6347
Tailandas	69,5	93,8	55	7742
Malaizija	71,4	83,5	61	9572
Meksika	72,1	89,6	67	6769
Turkija	68,5	82,3	60	5516
Omanas	70,7	59,0	60	9383
Kuveitas	75,4	78,6	58	23848
Honkongas	79,0	92,2	67	22950
Čilė	75,1	95,2	73	9930
Bahreinas	72,2	85,2	84	16751
Fidži	72,1	91,6	78	6159

Šių duomenų pagrindu:

2.1. Apskaičiuokite kiekvienos valstybės socialinės pažangos indeksus (komponentinius indeksus ir bendrąjį indeksą *HDI*).

2.2. Atlikite visų šių šalių analizę atskirai pagal kiekvieną rodiklį.

2.3. Apskaičiuokite Spirmeno (Spearman) koreliacijos koeficiento reikšmes, apibūdinančias bendrojo socialinės pažangos indekso *HDI* priklausomybę nuo komponentinių indeksų rangų. Kuris komponentinis indeksas turi didžiausią įtaką indeksui *HDI* analizuojamų šalių grupėje?

3. 4.8 lentelėje pateikti duomenys, apibūdinantys gyventojų skurdo lygį besivystančių pasaulio šalių grupėje.

4.8 lentelė. **Duomenys, apibūdinantys gyventojų skurdo lygį besivystančiose pasaulio šalyse, proc.**

Valstybė	Gyventojai, kurių gyvenimo trukmė neviršys 40 metų	Raštingumo lygis	Gyventojai, neaprupinti kokybišku vandeniu	Per mažo svorio vaikai iki 5 m.
Kamerūnas	26	63.4	50	14
Gana	23	64.5	35	27
Nigerija	31	57.5	50	36
Mauritanija	29	37.7	26	23
Madagaskaras	21	45.8	66	34
Gvinėja–Bisau	42	54.9	41	23
Gvinėja	38	35.5	54	26
Mali	36	31.0	34	27
Etiopija	34	35.5	75	48
Burkina–Faso	38	19.2	22	30
Siera Leonė	36	13.6	52	36
Mozambikas	38	40.1	37	27

3.1. Apskaičiuokite besivystančių šalių gyventojų skurdo indeksą. Kurioje šalyje skurdo indekso reikšmė didžiausia, kurioje – mažiausia?

3.2. Tarp gautos skurdo indekso reikšmės ir jo komponentių apskaičiuokite porinius koreliacijos koeficientus ir patikrinkite jų reikšmingumą. Sulyginkite koeficientus. Kuri komponentė turi didžiausią įtaką skurdo lygiui, o kuri – mažiausią?

3.3. Parašykite rezultatų analizę.

4. 4.9 lentelėje pateikti duomenys, apibūdinantys gyventojų skurdo lygį išsivysčiusių pasaulio šalių grupėje.

4.9 lentelė. **Duomenys, apibūdinantys gyventojų skurdo lygį 1995 m. išsivysčiusiose pasaulio šalyse, proc.**

Valstybė	Gyventojai, kurių gyvenimo trukmė neviršys 60 metų	Raštingumo lygis	Gyventojai, gyvenantys žemiau skurdo ribos (pagal ES standartą)	Ilgalaikis nedarbas (proc. nuo visos darbo jėgos)
Airija	9	99	11,1	7,6
Kanada	9	99	11,7	1,3
Prancūzija	11	99	7,5	4,9
Norvegija	9	99	6,6	1,3
JAV	13	99	19,1	0,5
Suomija	11	99	6,2	6,1
Olandija	9	99	6,7	3,2
Japonija	8	99	11,8	0,6
Švedija	8	99	6,7	1,5
Ispanija	10	97,5	10,4	13
Vokietija	12	99	7,5	2
Jungtinė Karalystė	9	99	13,5	3,8
Danija	11	99	5,9	4

4.1. Apskaičiuokite išsivysčiusių šalių gyventojų skurdo indeksą. Kurioje šalyje skurdo indekso reikšmė didžiausia, kurioje – mažiausia?

4.2. Apskaičiuokite gautos skurdo indekso reikšmės ir jo komponentių porinius koreliacijos koeficientus ir patikrinkite jų reikšmingumą. Sulyginkite koeficientus tarpusavyje. Kuri komponentė turi didžiausią įtaką skurdo lygiui, o kuri mažiausią?

4.3. Parašykite rezultatų analizę.

5. 4.10 lentelėje pateikti 1998 m. duomenys, apibūdinantys kai kurių pasaulio šalių lyčių išsivystymą.

4.10 lentelė. Duomenys apie lyčių išsivystymą 1998 m.

Valstybė	Tikėtina gyvenimo trukmė gimstant Moterų	Tikėtina gyvenimo trukmė gimstant Vyrų	Suaugusių gyventojų raštingumo lygis, proc. Moterų	Suaugusių gyventojų raštingumo lygis, proc. Vyrų	Besimokančio jaunimo dalis, proc. Moterų	Besimokančio jaunimo dalis, proc. Vyrų	Tikrasis BVP <i>per capita</i> , JAV dol. Moterų	Realus BVP <i>per capita</i> , JAV dol. Vyrų
Kanada	81,9	76,2	99	99	100	98	17,980	29,294
Norvegija	81,3	75,4	99	99	98	93	22,400	30,356
JAV	80,2	73,5	99	99	97	91	22,565	36,849
Japonija	83,0	76,9	99	99	83	86	14,091	32,794
Suomija	80,8	73,2	99	99	100	95	17,063	24,827
Prancūzija	82,1	74,4	99	99	94	91	16,437	26,156
Vokietija	80,3	74,1	99	99	88	90	15,189	29,476
Ispanija	81,6	74,7	96,5	98,4	96	90	9,636	23,078
Slovėnija	78,9	72,0	89,0	94,2	94	88	10,215	19,538
Čekija	77,7	70,6	99	99	74	73	9,713	15,153
Vengrija	75,1	67,1	99,1	99,4	75	73	7,452	13,267
Lenkija	77,1	68,4	99,7	99,7	79	78	5,821	9,519
Estija	74,7	63,4	99	99	87	82	6,076	9,492
Lietuva	75,7	64,7	99,4	99,6	78	74	5,037	7,998
Turkija	72,0	66,8	75,0	92,8	54	67	4,703	8,104

Remdamiesi šiais duomenimis:

5.1. Apskaičiuokite kiekvienos valstybės socialinės pažangos indeksą *HDI* ir išsivystymo indeksą, įvertinantį lyčių faktorių *GDI*.

5.2. Palyginkite šių šalių išsivystymą atskirai pagal kiekvieną indeksą.

6. 4.11 lentelėje pateikti duomenys apie Lietuvos apskritis. Nustatykite, ar šios požymių mažėjimo tvarka sudarytos apskričių eilės tarpusavyje susijusios. Apskaičiuokite Spirmeno koreliacijos koeficiento reikšmę. Patikrinkite hipotezę, kad šis koeficientas lygus nuliui.

4.11 lentelė. Lietuvos apskritys, išdėstytos rodiklių mažėjimo tvarka

Užregistruota nusikaltimų 10 000 gyv. 2000 m. (mažėjimo tv.)	Gyventojų skaičius 1 kv. km. (mažėjimo tv.)	Nedarbo lygis 2000 m. (mažėjimo tv.)	Automobilių sk. 10 000 gyv. (mažėjimo tv.)
Klaipėda	Kaunas	Šiauliai	Marijampolė
Vilnius	Vilnius	Tauragė	Klaipėda
Panevėžys	Klaipėda	Marijampolė	Tauragė
Kaunas	Šiauliai	Panevėžys	Alytus
Telšiai	Marijampolė	Telšiai	Telšiai
Šiauliai	Telšiai	Alytus	Utena
Marijampolė	Panevėžys	Utena	Panevėžys
Tauragė	Alytus	Klaipėda	Šiauliai
Utena	Tauragė	Kaunas	Kaunas
Alytus	Utena	Vilnius	Vilnius



LITERATŪRA

1. Čekanavičius V., Murauskas G. Statistika ir jos taikymai I. – Vilnius: TEV, 2000.
2. Francis A. Business Mathematics and Statistics. – London: DP Publications, 1993.
3. Практикум по социальной статистике. Под редакцией И. И. Елисеевой. – Москва: Финансы и статистика, 2002.
4. Duncan J. W., Gross A. C. Statistics for the 21st century. – Irwin: The Dun and Breadstreet Co, 1995.
5. Sakalauskas V. Statistika su *Statistica*. – Vilnius: Margi raštai, 1998.
6. Лунеев В. В. Юридическая статистика. – Москва: Юристъ, 1999.
7. Aksomaitis A. Tikimybių teorija ir statistika. – Kaunas: Technologija, 2000.
8. Gudavičiūtė – Bartosevičienė V. Ekonominė statistika. – Kaunas: Technologija, 1997. II d.
9. Rudzkienė V. Kompiuterinė duomenų analizė teisėje ir valdyme. – Vilnius: LTU, 2001.
10. Ефимова М. Р., Петрова Е. В., Румянцев В. Н. Общая теория статистики. – Москва: ИНФРА-М, 1999.
11. Lietuvos statistikos metraštis. – Vilnius: Statistikos departamentas, 2001.
12. Visuotinė Žmogaus Teisių deklaracija. Teisės problemos. – Vilnius: Teisės institutas, 1999. Nr.1–2.
13. Statistikos departamentas prie Lietuvos Respublikos Vyriausybės. URL: <http://www.std.lt/>
14. Nusikalstamumo prevencijos Lietuvoje centras. URL: <http://www.nplc.lt/>



PRIEDAI

1 lentelė. Stjudento skirstinio α lygmens kritinės reikšmės $t_{\alpha}(n)$

n/ α	0,40	0,10	0,05	0,025	0,01	0,005
1	0,325	3,078	6,314	12,706	31,821	63,657
2	0,289	1,886	2,920	4,303	6,965	9,925
3	0,277	1,638	2,353	3,182	4,541	5,841
4	0,271	1,533	2,132	2,776	3,747	4,604
5	0,267	1,476	2,015	2,571	3,365	4,032
6	0,265	1,440	1,943	2,447	3,143	3,707
7	0,263	1,415	1,895	2,365	2,998	3,499
8	0,262	1,397	1,860	2,306	2,896	3,355
9	0,261	1,383	1,833	2,262	2,821	3,250
10	0,260	1,372	1,812	2,228	2,764	3,169
11	0,260	1,363	1,796	2,201	2,718	3,106
12	0,259	1,356	1,782	2,179	2,681	3,055
13	0,259	1,350	1,771	2,160	2,650	3,012
14	0,258	1,345	1,761	2,145	2,624	2,977
15	0,258	1,341	1,753	2,131	2,602	2,947
16	0,258	1,337	1,746	2,120	2,583	2,921
17	0,257	1,333	1,740	2,110	2,567	2,898
18	0,257	1,330	1,734	2,101	2,552	2,878
19	0,257	1,328	1,729	2,093	2,539	2,861
20	0,257	1,325	1,725	2,086	2,528	2,845
21	0,257	1,323	1,721	2,080	2,518	2,831
22	0,256	1,321	1,717	2,074	2,508	2,819
23	0,256	1,319	1,714	2,069	2,500	2,807
24	0,256	1,318	1,711	2,064	2,492	2,797
25	0,256	1,316	1,708	2,060	2,485	2,787
30	0,256	1,310	1,697	2,042	2,457	2,750
40	0,255	1,303	1,684	2,021	2,423	2,704
60	0,254	1,296	1,671	2,000	2,390	2,660
120	0,254	1,289	1,658	1,980	2,358	2,617

2 lentelė. χ^2 skirstinio α lygmens kritinės reikšmės

n/α	a = 0,10	a = 0,05	a = 0,025	a = 0,01	n/α	a = 0,10	a = 0,05	a = 0,025	a = 0,01
1	2,71	3,84	5,02	6,63	21	29,62	32,67	35,48	38,93
2	4,61	5,99	7,38	9,21	22	30,81	33,92	36,78	40,29
3	6,25	7,81	9,35	11,34	23	32,01	35,17	38,08	41,64
5	9,24	11,07	12,83	15,09	24	33,20	36,42	39,36	42,98
6	10,64	12,59	14,45	16,81	25	34,38	37,65	40,65	44,31
7	12,02	14,07	16,01	18,48	26	35,56	38,89	41,92	45,64
8	13,36	15,51	17,53	20,09	27	36,74	40,11	43,19	46,96
9	14,68	16,92	19,02	21,67	28	37,92	41,34	44,46	48,28
10	15,99	18,31	20,48	23,21	29	39,09	42,56	45,72	49,59
12	18,55	21,03	23,34	26,22	30	40,26	43,77	46,98	50,89
13	19,81	22,36	24,74	27,69	40	51,80	55,76	59,34	63,69
14	21,06	23,68	26,12	29,14	50	63,17	67,50	71,42	76,15
15	22,31	25,00	27,49	30,58	60	74,40	79,08	83,30	88,38
16	23,54	26,30	28,85	32,00	70	85,53	90,53	95,02	100,42
18	25,99	28,87	31,53	34,81	80	96,58	101,8	106,63	112,33
19	27,20	30,14	32,85	36,19	90	107,5	113,1	118,14	124,12
20	28,41	31,41	34,17	37,57	100	118,5	124,3	129,56	135,81



doc. dr. Rudzkienė, Vitalija

Ru44 Statistinės technologijos teisėje ir valdyme. Praktiniai darbai: mokomasis leidinys. –Vilnius: Lietuvos teisės universiteto Leidybos centras, 2003. – 76 p.,43 lentelės, 8 paveikslai.

Bibliogr.: p. 73.

ISBN 9955 – 563 – 21 – 4

Štame mokomajame leidinyje nagrinėjami praktiniai socialinės statistikos uždaviniai ir jų taikymo sritys. Tokių uždavinių sprendimo pagrindą sudaro tipiniai aprašomos ir matematinės statistikos metodai. Pateikiami statistinių rezultatų interpretacijos pavyzdžiai padeda suvokti gautų rezultatų naudojimo galimybes vertinant ir numatant socialinėje aplinkoje vykstančius pokyčius. Leidinyje pateikiamos pagrindinės socialinės statistikos sąvokos, uždavinių sprendimo būdai, pavyzdžiai, sprendimų analizė ir kontroliniai uždaviniai.

Leidinys skirtas socialinių mokslų studijų pakraipos pagrindinių ir magistrantūros studijų studentams.

UDK 004:311(075.8)

Vitalija Rudzkienė
STATISTINĖS TECHNOLOGIJOS TEISĖJE IR VALDYME
Praktiniai darbai
Mokomasis leidinys

Redaktorė *Jurgita Marija Bagdonavičienė*
Maketavo *Regina Bernadišienė*
Viršelio dailininkė *Stanislava Narkevičiūtė*

SL 585. 2003 04 14. 3,44 leidyb. apsk. l.

Tiražas 300 egz. Užsakymas .

Išleido Lietuvos teisės universiteto Leidybos centras, Ateities g. 20, 2057 Vilnius.

Tinklapis internete www.ltu.lt

Elektroninis paštas leidyba@ltu.lt

Spausdino AB spaustuvė „Aušra“, Vytauto pr. 23, 3000 Kaunas.

Tinklapis internete www.ausra.lt

Elektroninis paštas ausra@ausra.lt