VILNIAUS PEDAGOGINIS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
ALGEBROS IR STATISTIKOS KATEDRA

Romanas Januškevičius, Olga Januškevičienė

Elementarusis tikimybių ir statistikos kursas

informatikams

2 dalis. STATISTIKOS PRADMENYS

METODINĖ PRIEMONĖ

Leidykla

Vilnius, **2006**

Metodinė priemonė rekomenduota spaudai VPU Matematikos ir informatikos fakulteto Tarybos 2006-05-16 sprendimu (Protokolo Nr.2).

Recenzentai:
prof. hab. dr. Kęstutis Kubilius
doc. dr. Jonas Banys

Metodinės priemonės redaktorius – profesorius hab. dr. Romanas Januškevičius

ISBN 9955-20-089-8 (bendras) elektroninis išteklius
ISBN 9955-20-087-1 (D.1)
ISBN 9955-20-088-X (D.2)

TURINYS

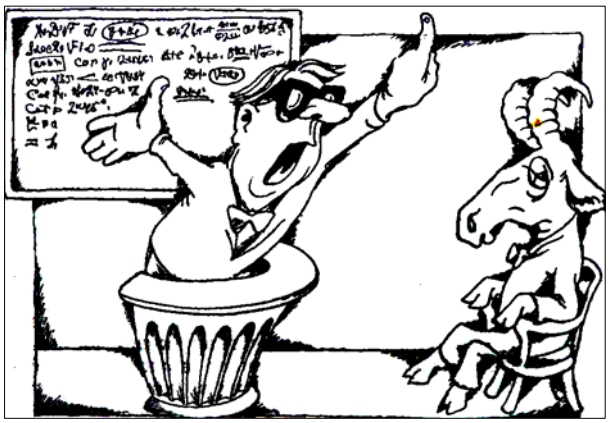
STATISTIKOS PRADMENYS

1. KAS YRA STATISTIKA?

XX-ajame amžiuje mokslininkų, tiriančių stochastinius reiškinius, darbuose galima įžvelgti šias dvi pagrindines tyrimų kryptis:

- Rytuose (lyderis – buvusi Sovietų Sąjunga) pagrindinis dėmesys skiriamas ne praktiniams, o teoriniams uždaviniams, ypač tikimybių teorijai;
- Vakaruose (lyderis – JAV) didžiausias dėmesys skiriamas problemoms, kurias padiktavo praktiniai uždaviniai, ypač statistikai.

Todėl, pradedant nagrinėti statistiką iš praktiko pozicijų, būtų naudinga pažvelgti į kokio nors JAV universiteto statistikos vadovėlį. Čia ir sekančiuose penkiuose nedidelės apimties aprašomojo pobūdžio skyreliuose mes cituojame profesorių Richard A. Johnson ir Gouri K. Bhattacharyya iš Viskonsino universiteto Madisone vadovėlio „Statistika. Pagrindai ir metodai“ trečiąjį leidimą (leidykla John Wiley and Sons).



Net ir šiandien didesnė visuomenės dalis mano, kad statistika yra sinonimas atstumiančioms skaičių eilėms ir begalei grafikų. Toks mūsų statistikos kurso įvaizdis, kurį pavaizdavo dailininkas piešinyje, yra visiškai nepriimtinas

Žodis statistika kilęs iš lotyniško žodžio „status“, kas reiškia „tvirtinti, konstatuoti, būsena“. Labai ilgai statistika buvo tapatinama su duomenų vaizdavimu įvairiomis diagramomis,

atspindinčiomis ekonominę, demografinę ir politinę šalies situaciją. Net ir šiandien didesnė visuomenės dalis mano, kad statistika yra sinonimas atstumiančioms skaičių eilėms ir begalei grafikų. Toks statistikos įvaizdis buvo primestas daugybės vyriausybės ataskaitų, kuriose dominavo skaičiai ir kurių pavadinimuose buvo žodis statistika, pvz. „Žemės ūkio produkcijos statistika“, „Prekybos ir laivininkystės statistika“, „Darbo statistika“.

Tačiau milžiniška XX amžiaus pažanga suteikė statistikai galimybę plėtotis, vystytis ir įgyti šiuolaikinę šios pagrįstos mąstymu disciplinos svarbą. Dabar monotoniškas skaičių ir diagramų vaizdavimas yra tik nereikšmingas statistikos aspektas, o šiuolaikinių statistikos specialistų, užimtų rutininiais lentelių bei diagramų sudarymais, yra labai mažai, jei dar išvis yra.

Tuomet koks gi yra statistikos kaip mokslinės disciplinos vaidmuo, kokie pagrindiniai tikslai? Be duomenų vaizdavimo, statistika užsiima duomenų (informacijos) rinkimu, apdorojimu (interpretacija) bei formuluoja išvadas apie tiriamą objektą. Jos veiklos sritys paprastai apima visus žinių apie faktų radimo, renkant ir apdorojant duomenis, įgijimo procesus. Keli pavyzdžiai gali būti tokie: įvairiausios apklausos šeimos, sociologijos, ekonomikos, sveikatos ir kt. klausimais, žemdirbystės eksperimentai (su naujomis sėklomis, pesticidais, darbo įrankiais), klinikiniai vakcinų tyrimai ir net lietaus iškvietimas (debesų užsėjimas). Statistikos pagrindai ir metodologija praverčia ieškant atsakymų į šiuos klausimus:

- Kiek ir kokio pobūdžio duomenų reikia rinkti?
- Kaip reikia organizuoti (surūšiuoti) ir apdoroti duomenis?
- Kaip reikia analizuoti duomenis ir daryti išvadas?
- Kaip įvertinti išvadų tvirtumą ir išmatuoti jų patikimumą?

Taigi, statistika aprūpina metodologija

- **duomenų rinkimo proceso modeliavimą,**
- **duomenų sumavimą ir apdorojimą,**
- **išvadų bei apibendrinimų darymą**.

Klausimai ir užduotys

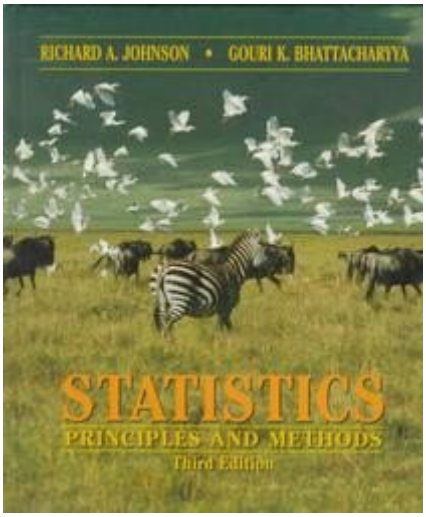
1. Žodis statistika kilęs iš lotyniško žodžio „status“. Ką reiškia šis žodis?
   a) tvirtinti;
   b) konstatuoti;
   c) aprašyti būseną;

d) visus tris variantus a), b) ir c) kartu.

2. Kokius klausimus nagrinėja statistika?

2. STATISTIKA MŪSŲ KASDIENINIAME GYVENIME

Faktų ieškojimas, renkant ir apdorojant duomenis, neapsiriboja profesionaliais tyrimais. Bandydami suprasti valstybės saugumo problemas, nedarbo būseną, futbolo komandų varžovių pasirodymo strategiją, peržiūrime ir interpretuojame skaitinę informaciją ir diagramas. Kasdieniniame gyvenime taip pat dažnai mokomės atlikti faktinės informacijos analizę. Be to, kiekvienas daugiau ar mažiau susipažinęs su statistika per žiniasklaidą.



Įdarbinimas. Kas mėnesį, vykdydamas Nuolatinės Gyventojų Apžvalgos programą, Surašymo Biuras surenka informaciją apie darbo statusą iš maždaug 65000 šeimų. Šeimos keičiamos, bet $\frac{3}{4}$ imties lieka nepakitę du mėnesius iš eilės. Apžvalgos duomenis analizuoja Darbo Statistikos Biuras, kuris kiekvieną mėnesį įvertina nedarbą.

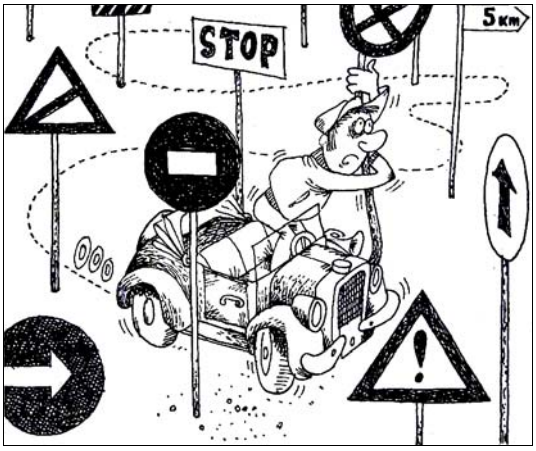
Kainos. Vartotojo kainos indeksas (VKI) matuoja fiksuoto vartotojo krepšio, į kurį įeina daugiau nei 400 prekių ir paslaugų, kainą. Kiekvieną mėnesį iš 85 JAV sostinės rajonų, t.y. iš daugiau kaip 18000 mažmeninės prekybos parduotuvių, gaunamos kainos. Šios kainos kombinuojamos, įskaičiuojant santykinį vartojamų (naudojamų) prekių ir paslaugų kiekį (pagal hipotetinį „miestiečio atlyginimą“, 1967). Nesigilinsime į prekių parinkimo ir skaičiavimo metodus, nes pastarieji yra labai painūs. Tačiau jie yra kruopščiai nagrinėjami, nes yra labai svarbūs šimtams tūkstančių amerikiečių, kurių atlyginimai arba nedarbo pašalpos yra pririštos prie VKI.

Rinkimai. Ši sritis yra labiausiai žinoma. Remiantis pokalbiais su daugiau kaip 1500 suaugusiųjų įvertinama, kiek yra populiarus vienas ar kitas rinkimų kandidatas. Jau keletas mėnesių prieš prezidento rinkimus reguliariai spausdinami tokių apklausų rezultatai. Tokie pranešimai padeda preliminariai nustatyti nugalėtojus ir fiksuoti rinkėjų simpatijų pakitimus.

Mūsų faktinės informacijos šaltiniai varijuoja nuo asmeninės patirties iki žiniasklaidos pranešimų ir straipsnių specializuotuose leidiniuose. Tokių pranešimų vartotojams – gyventojams – reikia kai kurių statistikos žinių, kad jie galėtų teisingai interpretuoti duomenis ir įvertinti išvadas. Statistikos aparatas pateikia kriterijų, padedančių nustatyti, kurios išvados paremtos duomenimis, o kurios – ne. Išvadų patikrinimas daugiausia priklauso nuo statistikos metodų panaudojimo, renkant duomenis. Statistika duoda raktą kiekvienam sistemingam įvertinimui, kaip pagerinti bet kokio tipo procesą, nuo gamybos iki paslaugų.

Kokybės ir našumo gerinimas. Per pastaruosius 30 metų JAV parodė didėjantį konkurencingumą pasaulinėje rinkoje. Tarptautinė kokybės ir našumo gerinimo revoliucija padidino spaudimą į JAV ekonomiką. W. Deming‘o idėjos ir mokymas padėjo Japonijai atnaujinti pramonę 1940 – 1950-aisiais. 1980 – 1990-aisiais Deming‘as pabrėžė, kad, norėdama išgyventi, Amerika privalo mobilizuoti visas darbo jėgas nuolatiniam kokybės gerinimui. Su savo idėjomis jis taip pat kreipėsi į vyriausybę.

Madisono miestas panaudojo kokybės gerinimo projektus policijoje, autobusų remontui ir tvarkaraščių sudarymui. Kiekvienu atveju projekto tikslas – tai geresnės paslaugos už mažesnę kainą. Pirmas žingsnis buvo bendravimas su gyventojais valstybinėse įstaigose – taip buvo renkama informacija apie sritis, kurios reikalauja pagerinimo. Šios statistinės informacijos pagrindu gautas toks rezultatas – strategiškai parinkta vieta naujajai policijos nuovadai, po to padidintas pėsčiųjų patruliuojančių policininkų skaičius (sąveikai su visuomene gerinti).



Ir Lietuvoje būtų naudinga rinkti informaciją apie svarbiausias sritis, kurios reikalauja pagerinimo. Šios statistinės informacijos pagrindu būtų gauti strategiškai svarbūs rezultatai – tada nei Seime, nei visuomenėje nepasitaikytų situacijų, panašių į tą, kurioje atsidūrė šis vairuotojas

Kai jau išrinktas pagerinimo projektas, reikia surinkti duomenis esamai situacijai įvertinti, vėliau dar surinkti duomenų tokiam tikslui – nustatyti kaip pasireiškia pasikeitimo efektas. Šioje stadijoje statistiniai įgūdžiai yra ne tik pageidautini, bet ir būtini kiekvienam dalyviui.

Statistinės žinios kiekvienam pramonės darbuotojui – stovinčiam prie konvejerio, sėdinčiam kontoroje, kontrolieriui ar vadybininkui – yra gyvybingai reikalingos kokybės kontrolei.

Klausimai ir užduotys

1. Kokiose mūsų gyvenimo srityse yra taikomi statistiniai metodai?
2. Kokius klausimus nagrinėja statistika?

3. STATISTIKA PADEDA MOKSLINIAMS TYRIMAMS

Moksliniai tyrimai prasideda nuo sistemingo mokymosi proceso. Mokslininkas iškelia tyrimo tikslą, renka tinkamą faktinę informaciją (duomenis), analizuoja duomenis, padaro išvadas ir nustato tolesnę veiklos kryptį. Pailiustruosime tai keliais pavyzdžiais.

Mokomosios programos. Daugumos sričių mokomosios programos sukurtos tam tikrai žmonių kategorijai:

- studentams,
- pramonės darbuotojams,
- mažumoms,
- žmonėms su fizine negalia,
- protiškai atsilikusiems vaikams ir t.t.

Jos yra nuolat stebimos, testuojamos, įvertinamos ir keičiamos, kad būtų naudingesnės vartotojams. Kad sužinotume apie skirtingų programų suderinimo efektyvumą, reikia būtinai rinkti duomenis apie dirbančių su programa žmonių pasiekimus arba įgūdžių padidėjimą.

Reklama. Visuomenė nuolatos yra atakuojama reklamų, kurios tikina, kad vienos prekės pranašesnės už kitas. Jei tokie pranašumai yra patvirtinti teisingų eksperimentų, tai jie tarnauja vartotojų švietimui. Tačiau ne retenybė, kai klaidinantys pareiškimai yra sukurti remiantis nepakankamais eksperimentais, neteisinga duomenų analize arba net skandalingomis eksperimento rezultatų manipuliacijomis. Vyriausybės agentūros ir vartotojų grupės turi būti pasiruošę įvertinti (patikrinti) laukiamą produkcijos kokybę, pasitelkę adekvačias duomenų rinkimo procedūras ir tinkamus statistinės analizės metodus.

Augalų auginimas. Siekdami padidinti maisto gamybą mokslininkai, dirbantys žemės ūkio srityje, kuria naujus augalų hibridus, sukryžiuodami skirtingas augalų rūšis. Žadamos naujos rūšys turi būti palygintos su geriausiomis jau naudojamomis rūšimis. Jų santykinis produktyvumas įvertinamas auginant keletą naujų rūšių kai kuriuose miestuose. Derliai yra aprašomi ir analizuojami jų akivaizdūs skirtumai. Taip pat turi būti palygintas trąšų poreikis ir atsparumas įvairiems veiksniams.

Pastatų sijos. Medinės sijos, kurios palaiko namų ir viešųjų įstaigų stogus, turi būti labai tvirtos. Dauguma sijų gaminama sujungiant kartu keletą lentų. Mokslininkai, tyrinėję medieną, surinko duomenis, parodančius jog, kietesnės lentos išlaiko didesnį svorį. Šis ryšys gali būti panaudotas būsimų sijų tvirtumui nustatyti.

Taigi, faktinė informacija yra reikšminga bet kokiam tyrimui. Statistikos sritis- **eksperimento planavimas**[1] – gali padėti tyrėjui nustatyti renkamų duomenų pobūdį ir kiekį.

[1] Experimental design.

Po to, kai surinkti duomenys, statistiniai metodai tinka matomoms duomenų ypatybėms susumuoti ir aprašyti. Visi bendrai jie žinomi kaip **aprašomoji statistika**[2]. Šiandien labiau akcentuojamas duomenimis pagrįstos informacijos pateikimas ir naujų žinių, gautų iš tos informacijos, įvertinimas. Tai yra statistikos srities, žinomos kaip **statistinės išvados**[3], metodai.

Turi būti suvokta, kad mokslinis tyrimas yra tipiškas bandymų ir klaidų procesas. Labai retais atvejais fenomenas gali būti pilnai suprastas arba teorija užbaigta remiantis tik vieno atskiro eksperimento rezultatais. Per daug yra tikėtis gauti viską vienu bandymu!

Netgi po pirmos sėkmės su elektros lempute Thomas Edison‘as eksperimentus su daugybe įvairių medžiagų kaitinimo siūlui pagaminti tęsė tol, kol pagaliau rado tinkamiausią to meto sąlygomis. Eksperimento metu gauti duomenys suteikia naujų žinių. Dažnai tos žinios siūlo peržiūrėti egzistuojančią teoriją ir pačios gali pareikalauti tolesnio tyrimo, t.y. daugiau eksperimentų ir duomenų analizių.

Klausimai ir užduotys

1. Koks yra tikimybių teorijos ir statistikos objektas?
   a) galimybių gauti patikimas išvadas (statistinių duomenų pagrindu) tyrimas;
   b) praktinių uždavinių sprendimas statistinių išvadų pagrindu.
2. Suformuluokite, kuo skiriasi terminų statistinė interpretacija nuo jų tikimybinės interpretacijos?

4. DVI PAGRINDINĖS SĄVOKOS – GENERALINĖ AIBĖ (POPULIACIJA) IR IMTIS

Ankstesniuose skyreliuose pristatėme keletą pavyzdžių, kuriuose faktinės informacijos įvertinimas yra būtinas naujų žinių įgijimui. Nors šie pavyzdžiai paimti iš įvairių sričių ir pristatyta tik padrika kurso apimtis ir tikslai, tačiau jau galima išskirti kelis bendrus bruožus.

Pirmiausia, norėdami įgyti naujų žinių, turime surinkti tinkamus duomenis. Antra vertus, neišvengiamas nedidelis duomenų varijavimas, net jei stebėjimai atlikti vienodomis arba labai panašiomis sąlygomis. Pvz., vaistas nuo alergijos gali suteikti ilgalaikį palengvėjimą vieniems,

[2] Descriptive statistics.
[3] Inferential statistics, statistical inference.

laikiną (arba jokio efekto) – kitiems. Taip pat nerealu yra tikėtis, kad pirmakursiai, kurių mokykliniai pažymiai buvo panašūs, panašiai mokysis ir universitete. Gamtoje tiksli tiesinė priklausomybė neveikia.

Trečias požymis yra tas, kad gauti pilną duomenų rinkinį yra arba fiziškai neįmanoma, arba nepraktiška. Kai laboratorinių arba lauko eksperimentų duomenys surinkti, visiškai nesvarbu, kiek bandymų buvo atlikta – visada galima atlikti daugiau. Tyrinėjant viešąją nuomonę arba vartotojų išlaidas, pilna informacija bus surinkta tik tuomet, kai duomenys bus paimti iš kiekvieno individo – neginčijamai monumentali (jei išvis įmanoma) užduotis. Iš tiesų, tarkime, kad jūs norite surinkti išsamią informaciją apie ryšį tarp tam tikro automobilio modelio, važiavusio fiksuotu greičiu ir patekusio į avariją, ir automobiliui padarytos žalos. (Tikriausiai kiekvienas matėte eksperimentus, kai mašina su manekenu visu greičiu atsitrenkia į sieną). Tuomet reikia kiekvieną to modelio automobilį sudaužyti avarijoje! Vadinasi, laiko, galimybių, išteklių apribojimas, o kartais ir destruktyvus eksperimento pobūdis reiškia, kad privalome dirbti su nepilna informacija – duomenimis, kurie iš tikrųjų surinkti eksperimento metu.

Ankstesni aprašymai parodė skirtumą tarp gautos tyrimo procese duomenų aibės ir plačios visų potencialių stebinių, kurie gali būti gauti dominama tema, aibės. Statistikoje pirmoji vadinama **imtimi**[4], o antroji- **generaline aibe** (**statistine populiacija**[5], arba tiesiog **populiacija**). Tolesniam sąvokų aiškinimui pastebėsime, kad duomenys atsiranda iš skirtingų šaltinių, kurie gali būti pacientas, medis, ferma, šeima arba bet kokio kito pobūdžio, priklausomai nuo tiriamo objekto. Kiekvieno duomens šaltinis vadinamas **imties vienetu**[6], arba tiesiog **vienetu**. Tuomet **imtis** (**pavyzdinis duomenų rinkinys**) sudaryta iš tų duomenų, kurių vienetai iš tikrųjų buvo stebėti. Jie sudaro dalį daug didesnio komplekto, apie kurį mes norėtume padaryti išvadas. Duomenų rinkinys, kuris atsirastų, jei visi vienetai didesniame komplekte būtų stebėti, apibrėžiamas kaip **generalinė aibė (populiacija**).

Generalinė aibė (populiacija) – tai duomenų (kiekybinių charakteristikų), atitinkančių pilną vienetų komplektą, apie kurį yra ieškoma informacija, rinkinys.

[4] Sample, sample data set.
[5] Statistical population.
[6] Sampling unit, unit.

Generalinė aibė yra tyrimo „taikinys“, objektas. Mes sužinome apie ją, paimdami iš jos imtį.

Imtis – tai duomenų rinkinys, kuris yra realiai surinktas tyrimo metu.

Pavyzdys.

Vienos muzikinės radijo laidos vedėja paskelbė, kad nori nustatyti, kuris atlikėjas yra populiariausias tarp šio miesto gyventojų. Klausytojai buvo raginami skambinti ir pasakyti jų mėgstamus atlikėjus.

Nustatykite generalinę aibę (populiaciją) ir imtį. Pakomentuokite, kaip iš miesto gyventojų sudaryti imtį, kad ji būtų reprezentatyvi[7] (**reprezentãcija** – *atstovavimas (paprastai geras) kam nors*), t.y. charakterizuojanti visą populiaciją, informatyvi.

SPRENDIMAS

Generalinė aibė (populiacija) sudaryta iš visų miesto gyventojų mėgstamų atlikėjų. Kadangi beveik neįmanoma apklausti visų gyventojų, iškyla būtinybė kalbėti apie imtį.

Vietiniu ryšiu skambinantys gyventojai – labai pigus, bet prastas imties sudarymo būdas. Imtis tokiu atveju būtų sudaryta tik iš visų tų atlikėjų, kuriuos įvardijo tik prisiskambinę į radijo stotį. Aišku, kad dėl tokios imties sudarymo procedūros pati imtis nebus reprezentatyvi, nes radijo stoties klausytojai, skambinantys į radijo redakciją, jau patys savaime sudaro specifinį pogrupį, beje, labai tikėtina, kad su panašiais muzikiniais skoniais. Dar daugiau, radę laiko paskambinti paprastai jaučiasi labiau pasitikintys savo teisumu. Tokių asmenų atsakymai gali išskirti country ar roko muzikos atlikėją, tuo tarpu visų gyventojų (ar radijo stoties klausytojų) muzikinis prioritetas gali būti visiškai kitoks.

Jeigu iš tikrųjų norima nustatyti populiariausią tarp miesto gyventojų atlikėją, reikia veikti kitaip. Vienas iš plačiai naudojamų būdų – apklausa telefonu, kur telefonų numeriai parenkami atsitiktinai. Pvz., skaičiai nuo 0 iki 9 surašyti ant atskirų lapelių ir įdėti į skrybėlę. Po to skiautelės traukiamos viena po kitos. Aišku, kompiuteris gali atlikti tokią atranką daug greičiau ir lengviau. Tokiu būdu atrinkti telefonų numeriai sudaro reprezentatyvesnę (informatyvesnę, vaizdesnę) imtį, negu savaiminė imtis (iš paskambinusiųjų į radijo stotį).

[7] Representative.

Savaiminės imtys, sudarytos iš skambinančių ar rašančių žmonių atsakymų, neatstovauja visuomenei, nėra reprezentatyvios[8]. Jos, pirmiausia, susidaro iš tų asmenų, kurie jaučiasi stiprūs dėl klausimo temos. Dėl to dauguma TV žinių ir pramoginių laidų dabar skelbia, kad jų telefoniniai balsavimai yra nemokslіniai ir atspindi tik tai paskambinusių žmonių nuomones.

Duomenys, surinkti siekiant aiškaus tikslo, labai skiriasi nuo **anekdotinių duomenų**[9]. Dažnas mūsų girdėjo kažką sakant, kad jisai laimėjo pinigų kazino, bet, aišku, dauguma žmonių negali nuolatos laimėti, nes kazino neužsiima pinigų davimo verslu. Žmonės linkę kalbėti gerų dalykų apie save. Analogiškai, net iš keleto vairuotojų galite išgirsti, kad jie buvo neprisisegę saugos diržo ir būtent todėl išsigelbėjo, nes išlėkė iš mašinos avarijos metu. Nors tokios istorijos yra pasakojamos, bet reikia atminti, kad nėra realios galimybės išgirsti istorijas tų vairuotojų, kurie būtų buvę likę gyvi, jei būtų buvę prisisegę saugos diržu. Anekdotinė informacija dažniausiai kartojama todėl, kad ji turi kažkokių stebėtinų bruožų, kurie tačiau dažniausiai neatspindi tokių įvykių koncentracijos realioje būtyje.

Klausimai ir užduotys

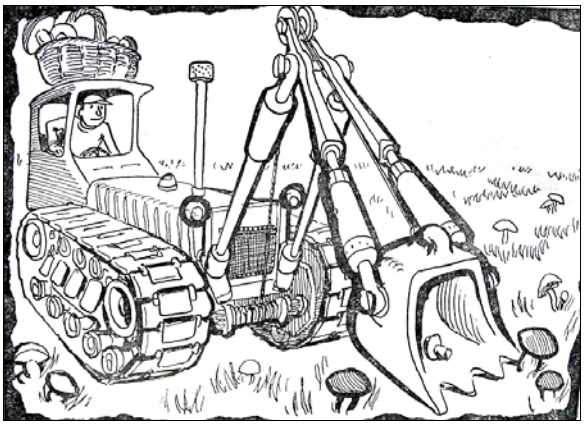
1. Kuo skiriasi generalinės aibės ir statistinės populiacijos sąvokos?
   a) generalinės aibės sąvoka yra bendresnė;
   b) statistinės populiacijos sąvoka taikoma tik praktinių uždavinių sprendime statistinių išvadų pagrindu;
   c) šios sąvokos yra sinonimai.
2. Ką vadiname imtimi?
3. Kuo skiriasi statistiniai duomenys nuo anekdotinių duomenų?
4. Paaiškinkite, kaip jūs suprantate reprezentatyvios imties sąvoką?

5. KRYPTINGAS DUOMENŲ RINKIMAS

Daug klaidingų sprendimų priimama verslo ir kasdieninių darbų srityse dėl nepakankamo galimybių supratimo ir įvertinimo. Suprantama, kad vieno žmogaus pirkingumo (**pirkingỹstė** – *pirkimo pamėgimas, aistra*) įpročiai negali būti apibendrinami visiems gyventojams, arba, analogiškai, vienos pelės reakcija į potencialiai toksinio cheminio mišinio veikimą dar nieko nesako

[8] Not representative.
[9] Anecdotal data.

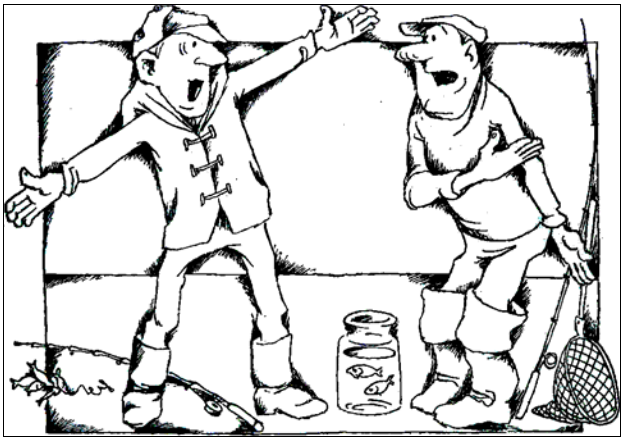
apie didesnės pelių populiacijos reakciją. Tačiau, nekreipdami dėmesio į individo pirkingumo įpročius, mes galime gauti tikslius duomenis apie visų gyventojų pirkingumo įpročius. Tam yra būtina surinkti duomenis iš didesnio žmonių skaičiaus. Analogiškai, daugiau galima sužinoti apie chemikalo toksiškumą, jei juo bus paveikta daug pelių.



Daug klaidingų sprendimų priimama verslo ir kasdieninių darbų srityse dėl nepakankamo galimybių supratimo ir įvertinimo

Pirmas žingsnis, norint atsakyti į rūpimą klausimą, argumentuoti pasirinkto veiksmo reikalingumą arba argumentuotai pagrįsti procesą – tai paprasčiausias sprendimas rinkti duomenis. Kai toks sprendimas priimtas, kitas svarbus žingsnis yra specifinio ir nesavanaudiško tikslo formulavimas (t.y. išreiškimas žodžiais). Jeigu tyrimo objektas yra vėluojantis visuomeninis transportas, reikia atsargiai apibrėžti, ką reiškia sąvoka „vėluoja“. Ar 1, 5, 10 minučių po tvarkaraštyje nurodyto laiko reikia vadinti vėlavimu? Tokias sąvokas, kaip *minkšta*, *patogu* dar sunkiau įvertinti. Vienas iš būdų – paprašyti keleivius norimą sąvoką įvertinti pagal skalę, kur skaičius 1 reiškia „labai nepatogu” ir t.t. iki 5, kuris reiškia „labai patogu“:

1	2	3	4	5
Labai nepatogu		Neutralu		Labai patogu



Tokias sąvokas, kaip *didelis* ar *mažas,* sunku įvertinti!

Aiškiai suformuluotas tikslas padės apsispręsti, kokius duomenis rinkti, ir užtikrins jų tinkamumą. Be aiškiai suformuluoto tikslo ar tinkamai apibrėžtų sąvokų (terminų) daug bergždžių pastangų gali būti sugaišta duomenų, kurie nepadės atsakyti į mus dominantį klausimą, rinkimui.

Pilnas tyrimo siekis gali būti apibendrintas, bet kiekviename žingsnyje reikalingas specifiškai suformuluotas tikslas. Pvz., pirmosios pagalbos paslaugos telefonu (t.y. pacientai gali skambinti pasiteirauti) pasidarė tokios prieinamos, kad kartais atsakymai į pacientų skambučius užima per daug laiko.. Tokie skambučiai turi būti nukreipiami gydytojams arba slaugėms, kuriems gali tekti renkti papildomą informaciją, kad galėtų tinkamai atsakyti į skambučius. Pilnas tyrimo tikslas buvo toks: perprasti teikiamą paslaugą ir ją patobulinti. Kaip geras pirmas žingsnis buvo nuspręsta išsiaiškinti, kiek laiko užtrunka atsakyti į skambutį. Buvo laukiamas laiko varijavimas skirtingiems skambučiams, todėl inicijuoto tyrimo tikslas buvo užfiksuoti šios paslaugos sklaidą, surinkus laiko imtį.

Tikslas: Inspektuoti šią paslaugą, surinkus duomenis apie tai, kiek laiko (min.) užtrunka atsakyti į skambutį.

Savaitės bėgyje buvo sudaryta įeinančių skambučių imtis, skambučio laikas buvo užrašomas kartu su skambinusiojo klausimu. Taip pat buvo fiksuojamas ir laikas (min.), reikalingas atsakymui. Kiekviena iš šių laiko reikšmių (kiek laukė skambinęs asmuo) pavaizduotas tašku 1 diagramoje. Pastebėsime, kad daugiau nei 1/3 skambinusiųjų laukė virš 120 min., kol sulaukė atsakymą. Tai gali būti labai ilgas laiko tarpas, jei lauki informacijos apie karščiuojantį vaiką ar suaugusį su aštriais simptomais. Jei tikslas buvo nustatyti, kuri skambinusiųjų dalis per ilgai laukia atsakymo, reikia

tiksliau apibrėžti sąvoką „per ilgai“ – išreikšti šią sąvoką minutėmis. Tuo pačiu šie duomenys aiškiai parodo, kad paslauga reikalauja patobulinimo, ir kitas žingsnis turi būti padarytas ta kryptimi.

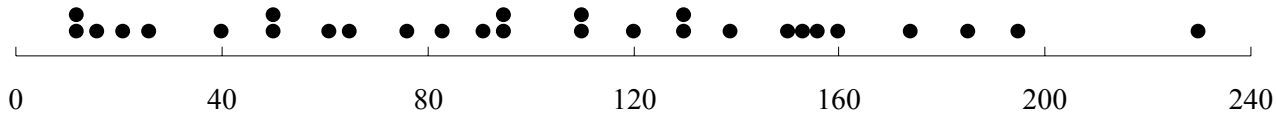


Diagrama 1. Laikas (min) iki atsakymo į skambutį.

Kaip bebūtų, tęsiant paslaugos tobulinimą, reikia susikoncentruoti ties detalėmis. Į tris klausimus

Kada? Kur? Kas?

visuomet turi būti atsakyta prieš pradedant rinkti daugiau duomenų. Dar detaliau, reikia ieškoti tokių duomenų, kurie atsakytų į šiuos klausimus.

Kada atsiranda sunkumai? Ar per pastarąsias valandas, dienas, savaites, mėnesius, ar tai sutapimas su kitais veiksniais.

Kur atsiranda sunkumai? Stengtis surasti silpnas vietas ir nereikalingus užlaikymus (delsimus).

Kas dirbo ir kas kontroliavo? Idėja yra ne priekaištauti, bet suprasti dalyvių vaidmenis bei juos patobulinti.

Labai dažnai praverčia sukonstruoti **priežasties-ir-pasekmės diagramą** (**žuvies kaulo diagramą**)[10]. Pagrindinė centrinė tiesė simbolizuoja problemą arba pasekmę. Žuvies kaulo diagrama pavaizduota 2 diagramoje, kurioje klausimas *Kur*? siejamas su skambučių užlaikymų vietomis. Pagrindinė tiesė centre simbolizuoja problemą: kur yra užlaikymai? Skambučiai priimami registratūroje, bet, kai tos linijos yra užimtos, skambučiai peradresuojamos tiesiai medicinos seselėms (3-as ir 4-as aukštai). Pasviros tiesės simbolizuoja aukštus, o maži horizontalūs brūkšniai- specifines vietas, kuriose gali atsirasti užlaikymas. Galutinė diagrama primena žuvies kaulą. Diagramos aptarimas gali padėti apsispręsti, kokius naujus duomenis reikia rinkti.

[10] Cause-and-effect diagram, fishbone diagram.

Laimei, kvalifikuota tyrėjų komanda jau iš anksto apsvarstė klausimus *Kada*?, *Kur*? ir *Kas*? bei fiksavo ne tik laiką, bet ir datą bei atsiliepusį į skambutį darbuotoją. T.y., jų turimi duomenys paskatino domėtis, ar atsakymo laikas priklauso nuo to, kada ir kur skambutis buvo gautas.

Taip pat kvalifikuota komanda sudarė daugiau tyrimui reikalingų detalių diagramų. Po to jie nustatė kliūtis informacijos sraute, jas pašalino ir paslauga buvo patobulinta. Šiuo pavyzdžiu norėta akcentuoti kryptingo tinkamų duomenų rinkimo idėją.

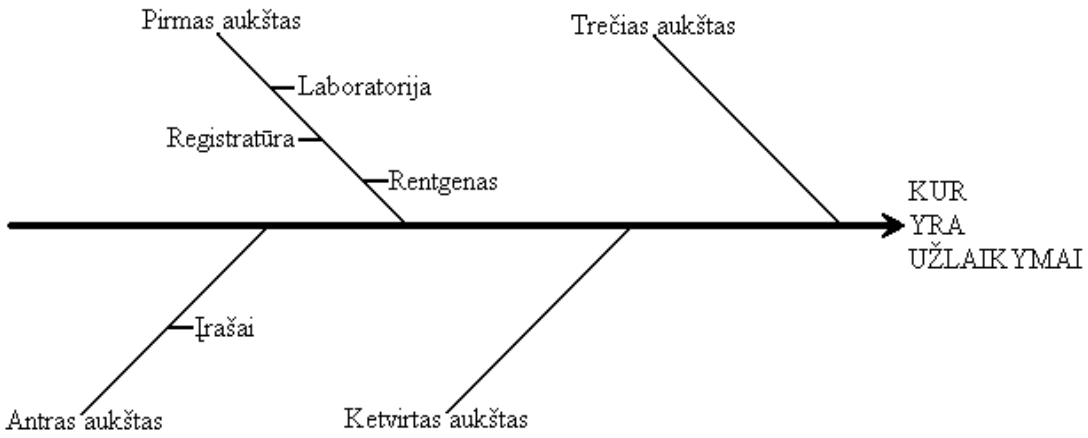


Diagrama 2. Priežasties-ir-pasekmės diagrama.

Klausimai ir užduotys

1. Ar 1, 5, 10 minučių po tvarkaraštyje nurodyto laiko reikia vadinti autobuso vėlavimu?
   a) 1 minutė po tvarkaraštyje nurodyto laiko – tai joks vėlavimas;
   b) 1 ar 5 minutės po tvarkaraštyje nurodyto laiko – tai joks vėlavimas autobusui;
   c) 10 minučių po tvarkaraštyje nurodyto laiko – tai jau vėlavimas;
   d) 1,5, 10 minučių po tvarkaraštyje nurodyto laiko – tai jau vėlavimas.
2. Konkrečiu pavyzdžiu pademonstruokite žuvies kaulo diagramos sudarymo tikslingumą.

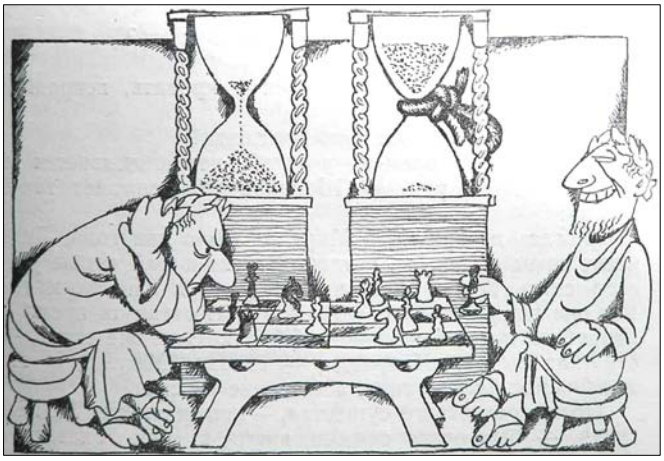
6. STATISTIKOS TIKSLAI

Statistikos objektas aprūpina ją metodologija: kaip daryti išvadas apie visą populiaciją, sudarius ir išanalizavus tik jos imtį. Šie metodai duoda galimybę atlikti tikroviškus apibendrinimus ir įvertinti jų patikimumo laipsnį. Statistikos sąvokos taip pat svarbios planuojant tyrimo eigą, kai reikia nustatyti veiklos pobūdį ir imties dydį.

Pagrindiniai statistikos tikslai yra:

- padaryti **išvadas** apie visą generalinę aibę (populiaciją), išanalizavus imties duomenis, ir įvertinti šių išvadų patikimumą;
- **suplanuoti tyrimo eigą ir imties dydį** taip, kad šio tyrimo dėka atliekamas realaus proceso stebėjimas padėtų pagrindą svarių išvadų darymui.

Toks imties planavimas yra labai svarbus. Geras duomenų rinkimo proceso planavimas leidžia padaryti veiksmingas išvadas, dažnai kartu su tiesiogine analize. Deja, net patys sudėtingiausi duomenų analizavimo metodai negali patys išgauti daug informacijos iš duomenų, gautų blogai suplanuoto eksperimento ar stebėjimo metu.



Negalima išgauti teisingos informacijos iš duomenų, gautų blogai suplanuoto eksperimento metu

Ankstesnis statistikos panaudojimas kompiliuojant ir pasyviai vaizduojant duomenis jau beveik pakeistas šiuolaikiniais metodais, kurių tikslas – pateikti analitinius įrankius, su kuriais duomenys gali būti efektyviai renkami, vienareikšmiškai suprantami ir korektiškai interpretuojami. Statistikos sąvokos ir metodai leidžia daryti svarias išvadas apie visą populiaciją iš jos imties. Statistika jau įsiskverbė į visas žmonių veiklos sritis, kuriose informacijos įvertinimas turi būti pagrįstas duomenų pagrindu gautais įrodymais.

Klausimai ir užduotys

1. Kokia yra statistikos metodologija?
    a) daryti išvadas apie visą imtį, sudarius ir išanalizavus tik jos populiaciją;
    b) daryti išvadas apie visą populiaciją, sudarius ir išanalizavus tik jos imtį;
2. Suformuluokite pagrindinius statistikos tikslus.

7. POPULIACIJOS IR IMTIES SĄVOKOS PAGAL LIETUVOS STANDARTĄ

Pagal Lietuvos standartą,

Populiacija - tai *visų stochastiniame eksperimente nagrinėjamų elementų aibė.*

Ėmimo vienetas - tai *populiacijos elementas.*

Pagal tą patį standartą, **imtis** (arba **atranka)** - tai *vienas ar daugiau ėmimo vienetų, paimtų iš populiacijos, siekiant gauti informaciją apie populiaciją.* **Imties dydis** - tai *ėmimo vienetų skaičius imtyje.*

Šiuolaikiniai statistikos vadovėliai pateikia dar vieną elementariųjų įvykių erdvės Ω apibrėžimą - *imčių erdvė*. Pateiksime šį ir gimininguis apibrėžimus pilnai.

Imčių erdve, *susijusia su stochastiniu eksperimentu, vadinama šio eksperimento visų galimų skirtingų rezultatų (baigčių) aibė.*

Kiekvienas šio eksperimento rezultatas vadinamas **elementariąją baigtimi,** *arba* **elementariuoju (paprastuoju) įvykiu,** arba **imčių erdvės elementu, arba stebiniu.**

Atsitiktiniu įvykiu *vadinama elementariųjų baigčių, turinčių apibrėžtas savybes, aibė.*

Populiacijos elementai į imtį gali būti atrenkami įvairiais būdais. Pakartosime, kad

vienas iš svarbiausių reikalavimų – imties *reprezentatyvumas*. **Imtis reprezentatyvi**, *jei ji teisingai atspindi tiriamo požymio galimų reikšmių populiacijoje proporcijas.*

Būtent imties reprezentatyvumas lemia, ar ištyrus imtį gausime patikimas išvadas apie visą populiaciją. Akivaizdu, kad imties reprezentatyvumas glaudžiai susijęs su imties didumu. Jeigu imtis apima beveik visą populiaciją, tai ji labai reprezentatyvi.

Klausimai ir užduotys

1. Kas lemia, kad ištyrus imtį gausime patikimas išvadas apie visą populiaciją?
   a) statistinio tyrimo atlikėjo susitarimas su užsakovu;
   b) atsitiktinis įvykis;
   c) imties reprezentatyvumas;
   d) populiacijos reprezentatyvumas.

2. Ką vadiname atsitiktiniu įvykiu?
3. Ką vadiname populiacija? imčių erdve?

8. STATISTINIŲ DUOMENŲ APDOROJIMAS

1. *Statistinius duomenis* galima suskirstyti į du pagrindinius tipus:

(A) *kokybiniai duomenys* (arba *duomenys pagal kategorijas* (klases, grupes)),

(B) *kiekybiniai duomenys* (arba *skaitiniai duomenys).*

2. Iš pradžių duomenys paprastai apdorojami kokybiškai, t.y. suskirstomi į *kategorijas (klases, grupes)* pagal požymius. Pavyzdžiui, nustatomos amžiaus grupės (iki 19 metų, 20-24, 25-29, 30-39, 40-49, 50 ir daugiau *ar pan.),* plaukų spalva (blondinas, brunetas *ir pan.),* užimtumas (bedarbis, užimtas darbu), mirtingumo priežastys (infekcinės ir parazitinės ligos, kraujo apytakos sistemos ligos, piktybiniai augliai *ir pan.).*

Štai paprasčiausias kokybinių duomenų tipo pavyzdys.

Pavyzdys. *Patikrinus 20 magistrantų kraujo grupę (I, II, III arba IV), buvo gauti tokie duomenys:*

I II I I II IV I III I I
III I II II III I II II I II

3. Išreiškus rezultatus kiekybinėmis (t.y. skaitinėmis) charakteristikomis pagal kategorijas, duomenys tuo pačiu apdorojami *kiekybiškai*. Kiekvienas stebinys turi būti priskirtas tik vienai iš kelių kategorijų. Tokie duomenys paprastai pateikiami *dažnių lentelės pavidalu*, kurį parodo kiekvienos kategorijos elementų skaičių (*dažnį*). Šio dažnio santykis su visų stebinių skaičiumi vadinamas *santykiniu dažniu*.

Dažnių lentelę pavyzdyje su magistrantų kraujo grupėmis galima sudaryti taip:

Kraujo grupė	Dažnis	Santykinis dažnis
I	*9*	*9/20 = 0.45*
II	*7*	*7/20 - 0.35*
III	*3*	*3/20 = 0.15*
IV	*1*	*1/20 = 0.05*
Viso:	*20*	*1.00*

4. Dabar pacituosime Lietuvos standartą.

Požymis - tai savybė, padedanti identifikuoti, atskirti ar klasifikuoti populiacijos individus.

Stebinys - tai požymio reikšmė, gauta kaip atskiro stebėjimo rezultatas

Klasė, grupė: *(1)* kokybinio požymio atveju - grupė elementų, turinčių tam tikras bendras savybes. Grupės nesikerta ir apima visą populiaciją;

(2) kiekybinio požymio atveju - kiekvienas iš nesikertančių intervalų, į kuriuos padalytas visas kitimo intervalas.

Dažnis - tai tam tikro tipo įvykių skaičius arba į tam tikrą klasę patekusių stebinių skaičius.

Santykinis dažnis - *tai* dažnis, padalytas iš viso bandymų ar stebinių skaičiaus.

Dažnių skirstinys - *tai* požymio reikšmių ir jų dažnių sąsaja. (Skirstinys gali būti grafiškai pateiktas kaip histograma, stulpelinė diagrama, ... arba kaip dažnių lentelė).

Pavyzdys. Susituokusių 1997 metais Lietuvos vyrų dažnių lentelė pagal amžių:

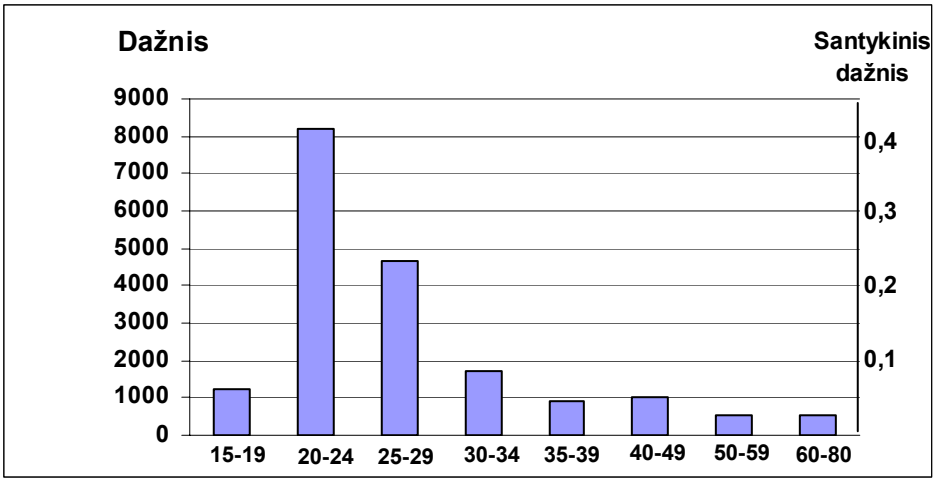
1 lentelė

Kategorija (amžiaus klasė, grupė)	Dažnis (vyrų skaičius amžiaus klasėje)	Santykinis dažnis
15-19	1225	1225/18796 = 0.0652
20-24	8181	8181/18796 = 0.4353
25-29	4672	4672/18796 = 0.2486
30-34	1729	1729/18796 = 0.0920
35-39	928	928/18796 = 0.0494
40-49	995	995/18796 = 0.0529
50-59	512	512/18796 = 0.0272
60-80	554	554/18796 = 0.0294
Viso	18796	1.0000

5. Aptarsime stulpelinės diagramos ir histogramos sąvokas, kurios buvo minimos iliustruojant dažnių skirstinio sąvoką.

Stulpelinė diagrama - tai *kiekybinio požymio dažnių grafinė išraiška, susidedanti iš vienodo pločio stulpelių, kurių ilgiai proporcingi dažniams.*

Atkreipsime dėmesį į svarbią detalę - stulpelių pločiai yra vienodi, nors nagrinėjamos klasės kiekybine prasme gali būti skirtingos. Mūsų pavyzdyje vyrų klasėje nuo 20 iki 24 metų amžiaus skirtumas gali siekti 5 metus, klasėje nuo 40 iki 49 metų šis skirtumas gali siekti 10 metų, o klasėje nuo 60 iki 80 - net 21 metus. Taigi, klasių plotis (ši sąvoka aiškinama žemiau) gali būti skirtingas, tačiau diagramos stulpelių plotis vienodas. Tai naudinga tuo atveju, kai pagrindinis dėmesys koncentruojamas į dažnius. Mūsų pavyzdyje apie susituokusius vyrus stulpelinė diagrama yra tokia:



Susituokusių 1997 m. Lietuvos vyrų stulpelinė diagrama pagal amžių

6. Grįžkime prie paskutinio pavyzdžio dažnių lentelės. Blokai 20-24, 25-29, 30-34 ir pan. yra vadinami *klasių intervalais.* Apatinis klasės intervalo galas vadinamas *klasės apatine riba,* o viršutinis galas - *klasės viršutine riba.* Šiose ribose turi tilpti nagrinėjamos klasės intervalas. Taigi, mūsų pavyzdyje 15, 20, 25, 30, 35, 40, 50, 60 yra klasės apatinės ribos, o 19+1, 24+1, 29+1, 34+1, 39+1, 49+1, 59+1, 80+1 - viršutinės ribos. Kodėl plius l? Jei žmogui yra 24 metai ir 364 dienos, t.y. *beveik* 25 metai, mes jį vis tiek priskiriame klasei 20-24. Tai daroma tam, kad vienas ir tas pats žmogus būtų priskirtas tik vienai klasei.

Tokioms situacijoms Lietuvos standarte yra skirtos dvi pastabos: 1) turi būti patikslinta, kurios iš šių dviejų ribų (apatinės ar viršutinės) priskiriamos klasei; 2) jei įmanoma, klasės ribos neturi sutapti su galima reikšme.

Mūsų pavyzdyje yra patikslinta, kad apatinė riba priklauso klasei, o viršutinė - ne. Be to, klasės viršutinė riba nėra jos galima reikšmė. Taigi, nėra dviprasmybės ir painiavos.

Klasės plotis - tai *klasės viršutinės ribos ir apatinės ribos skirtumas.* Nagrinėjamajame pavyzdyje klasių 15-19, 20-24, 25-29, 30-34, 35-39 pločiai yra lygūs 5, klasių 40-49 ir 50-59 pločiai - 10, o klasės 60-80 plotis yra 21.

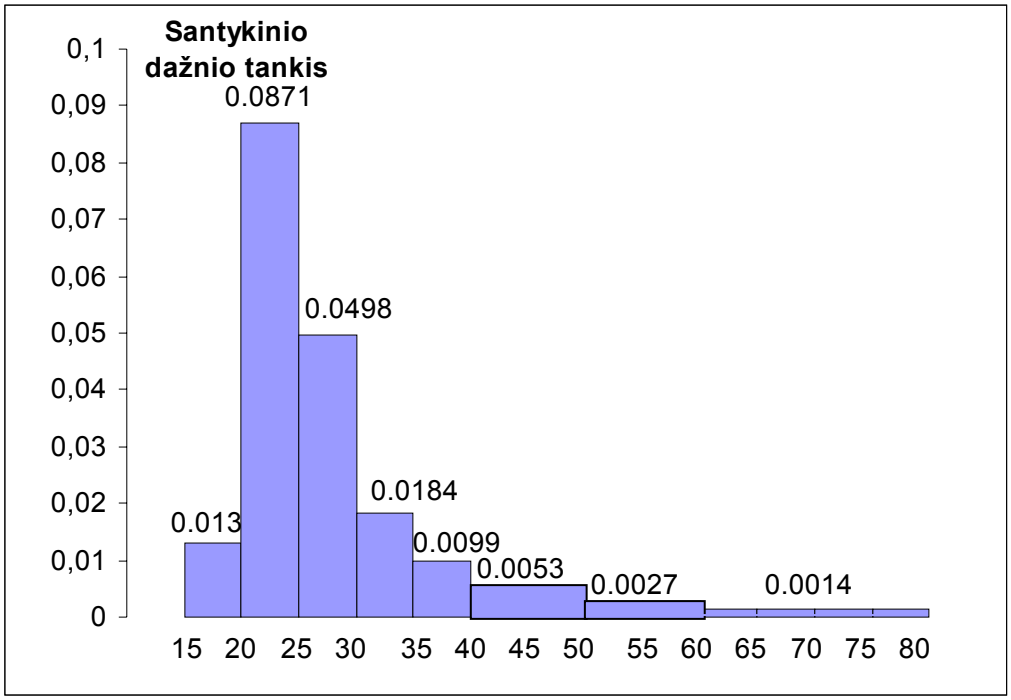
7. Dabar turime visas reikalingas sąvokas histogramos apibrėžimui.

Histograma - tai *kiekybinio požymio dažnių skirstinio grafinė išraiška, susidedanti iš keleto besiribojančių stačiakampių, kurių kiekvieno pagrindas lygus klasės pločiui, o plotas lygus klasės santykiniam dažniui.*

Kadangi visų stačiakampių plotas yra lygus klasių santykinių dažnių sumai, tai gauname tokią taisyklę: **histogramos plotas lygus 1.**

Lietuvos standarte pateikiamas histogramos apibrėžimas skiriasi tik paskutiniaisiais žodžiais: „... o plotas proporcingas klasės dažniui".Taip apibrėžus histogramą, jos plotas nebūtinai lygus l, nors sąvokos esmė išlieka ta pati. Statistikos vadovėliuose yra patogesnis pirmasis apibrėžimas, nes asocijuojasi su tankio sąvokos įvedimu.

Mūsų pavyzdyje apie susituokusius vyrus histograma yra tokia (skaičius virš stačiakampio išreiškia jo aukštį):



Susituokusių 1997 m. Lietuvos vyrų histograma pagal amžių.

Klausimai ir užduotys

1. Ką vadiname stulpeline diagrama?
   a) tai kiekybinio požymio dažnių grafinė išraiška, susidedanti iš stulpelių, kurių ilgiai proporcingi dažniams;
   b) tai kiekybinio požymio dažnių grafinė išraiška, susidedanti iš vienodo pločio stulpelių, kurių ilgiai proporcingi dažniams.;
   c) tai kiekybinio požymio dažnių grafinė išraiška, susidedanti iš vienodo pločio stulpelių, kurių ilgiai proporcingi dažniams, o jų bendras plotas lygus 1.
2. Ką vadiname santykiniu dažniu?
3. Ką vadiname histograma? Kuo ji skiriasi nuo stulpelinės diagramos?

9. ĮVAIRŪS IMTIES APRAŠYMO BŪDAI

Neapdorota imtis dažnai nepatogi tolimesnei analizei. Pirmiausia imties duomenys yra grupuojami, kad galima būtų išskirti imties charakteringus ypatumus.

Imties elementų (stebinių), išdėstytų nemažėjančia tvarka, seka vadinama **variacine eilute.** *Vienodi elementai kartojami.*

Pavyzdžiui, imties 90, 56, 104, 150, 120 variacine eilutė yra : 56, 90, 104, 120, 150.

Variacinės eilutės žymėjimas: $x_{(1)}, x_{(2)}, ..., x_{(n)}$.

Minimalus ir maksimalus imties elementai vadinami **ekstremaliosiomis** *(arba* **kraštinėmis***) reikšmėmis:*

$$x_{min} = x_{(1)}, \quad x_{(max)} = x_{(n)}.$$

Imties pločiu *vadinamas vadinamas imties ekstremaliųjų reikšmių skirtumas R,*

$$R = x_{max} - x_{min}$$

Pavyzdys. Vieno miligramo tikslumu išmatavus detalių nuokrypį nuo nominalo, gauta tokia 15 stebinių imtis:

0, 3, -5, -3, 1, 0, 1, 3, 0, 0, -3, 1, -1, 0, -1.

Šios imties variacinė eilutė yra: ši:

-5, -3, -3, -1, -1, 0, 0, 0, 0, 0, 1, 1, 1, 3, 3.

Lengva pastebėti, kad

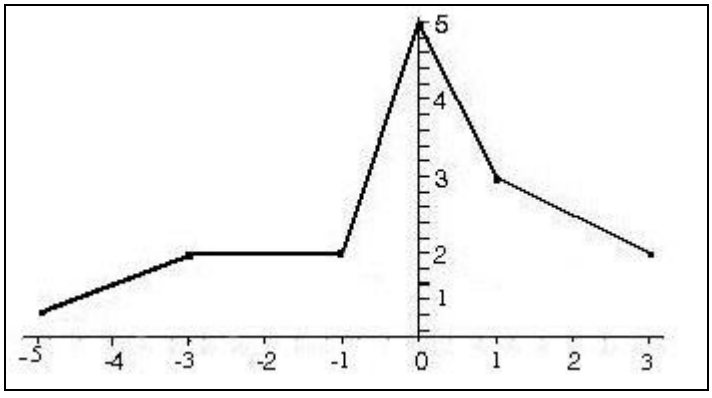
$$x_{min} = -5; \quad x_{max} = 3; \quad R = 3 - (-5) = 8.$$

Skirtingų imties elementų (stebinių), išdėstytų didėjimo tvarka, seka z_1, z_2, ... , z_k su atitinkamais dažniais n_1, ..., n_k vadinama **statistine eilute**.

Statistinė eilutė paprastai užrašoma lentelės pavidalu – ankstesniuose skyreliuose tokias lenteles jau esame sudarę. Nagrinėjamo pavyzdžio atveju statistinė eilutė yra tokia:

z_i	-5	-3	-1	0	1	3
n_i	1	2	2	5	3	2

Statistinę eilutę patogu vaizduoti *linijų diagrama* arba *poligonu*[11], kurioje abscisių ašyje žymimi stebiniai z_i, o ordinačių ašyje – jų dažniai n_i:



[11] polygon – angl.

Populiacijos, o tuo pačiu ir imties elementus vienija tiriamasis požymis. *Matuodami šį požymį, gauname tam tikrą dydį, kuris kinta kartu su imties nariais. Šį dydį ir vadinsime kintamuoju*. Imties duomenų aibė – tai ne kas kita kaip kintamojo reikšmių aibė. Išmatavę visą populiaciją, gautume visas kintamojo reikšmes. Tarkime norime sužinoti, kokia yra VPU studentų tautinė sudėtis. Šiuo atveju VPU studentai – tiriamoji populiacija, tautybė – kintamasis, pasirinktų tyrimui studentų (imties elementų) tautybės – kintamojo reikšmės. *Pagal matuojamo reiškinio prigimtį kintamieji skirstomi į kiekybinius ir kokybinius.*

Kiekybinio kintamojo reikšmė, tai atsakymas, kiek tiriamo požymio turi populiacijos elementas, tuo tarpu kokybiniai kintamieji nusako dydžius, kurių neįmanoma įvertinti skaičiais.

Kiekybinių kintamųjų pavyzdžiai – laikas, aukštis, sesijos pažymių vidurkis ir pan. Kokybinių kintamųjų pavyzdžiai – lytis, rasė, krepšinio komandos pavadinimas ir pan.

Kokybinių ir kiekybinių kintamųjų analizės metodai yra skirtingai, nes, pavyzdžiui, kokybiniai kintamieji negali būti sudedami, dauginami, apskaičiuojami jų vidurkiai.

Statistinėje eilutėje kintamojo x reikšmės gali kartotis. Tarkime, kad statistinėje eilutėje yra k skirtingų reikšmių. Sakykime, kad stebima reikšmė x^j pasikartojo f_j kartų. Tuomet $f_1 + f_2 + ... + f_k = n$, o x^j statistinėje eilutėje sudaro f_j / n dalį visų stebėjimų.

2 *Kintamojo reikšmės dažnis* f_j *- tai skaičius, nusakantis, kiek kartų reikšmė* x^j *pasikartojo statistinėje eilutėje*.

3 *Kintamojo reikšmės santykinis dažnis* f_j / n *- tai skaičius, nusakantis, kurią statistinės eilutės dalį sudaro* x^j .

Yra skaičiuojami tiek kiekybinių, tiek ir kokybinių kintamųjų dažniai, bei santykiniai dažniai.

Jei mūsų stebimas kintamasis įgyja nedaug skirtingų reikšmių, tai duomenis patogu yra surašyti į dažnių ir santykinių dažnių lenteles.

Taip pateiktą informaciją yra daug paprasčiau suvokti, o taip pat pastebėti įvairias duomenų savybes (pvz., dažniausiai pasikartojančią reikšmę, mažiausiąją bei didžiausias reikšmes). Duomenims sisteminti V. Čekanavičius ir G. Murauskas vadovėlyje „Statistika ir jos taikymai“ (I d., TEV, Vilnius, 2000) rekomenduoja naudoti sukauptuosius bei santykinius sukauptuosius dažnius tokios lentelės pavidalu:

2 lentelė

Reikšmė	x_1	x_2	...	x_k
Dažnis	f_1	f_2	...	f_k
Sukauptas dažnis	f_1	f_1+f_2	...	$f_1+f_2+...+f_k=n$
Santykinis dažnis	f_1/n	f_2/n	...	f_k/n
Sukauptas santykinis dažnis	f_1/n	$(f_1+f_2)/n$	...	$(f_1+f_2+...+f_k)/n=1$

Pavyzdys. Tarkime, kad 100 studentų išreiškė požiūrį į studijas universitete : "Labai patinka" – atsakė 20, "patinka" – 30, neturi nuomonės - 30, "nepatinka" - 16 ir labai nepatinka - 4.

Sprendimas. Užpildome lentelę:

3 lentelė

Reikšmė	**Dažnis**	**Sukauptas dažnis**	**Santykinis dažnis**	**Sukauptas santykinis dažnis**
Labai patinka	20	20	0,20	0,20
Patinka	30	50	0,30	0,50
Neturiu nuomonės	30	80	0,30	0,80
Nepatinka	16	96	0,16	0,96
Labai nepatinka	4	100	0,04	1,00

Iš sąlygoje pateiktų duomenų sunku nustatyti, koks yra vyraujantis požiūris vienu ar kitu požiūriu. Tokiu atveju dažniai ir sukauptieji dažniai, pateikti lentele, yra informatyvesni. Pvz. galime teigti, kad 25 studentams patinka arba labai patinka studijos universitete ir net 0,80 (80%) procentai nėra prieš jas nusistatę.

*****

Tačiau gali būti taip, kad mūsų tiriamas kintamasis įgyja labai daug skirtingų reikšmių. Tokiu atveju dažnių lentelė nėra nei labai informatyvi, o tuo labiau patogi – dėl skirtingų duomenų gausos. O ypač dar, kai kurie stebėjimai gali labai mažai tarpusavyje skirtis. Tokius duomenis patogiausia yra grupuoti. Tačiau prieš grupavimą dar reikia nustatyti: 1) grupavimo intervalų skaičių , 2) jų plotį, 3) intervalų kraštinius taškus. Dažniausiai yra pasirenkama nuo 6 iki 20 intervalų.

Intervalų skaičiui parinkti tinka tokia Sturgeso taisyklė:

$$k = 1 + 3{,}222 \cdot \log_{10} n$$

Čia k – intervalų skaičius, n – imties dydis.

Tada visas intervalas $[x_{\min}, x_{\max}]$ žingsniu h yra skaidomas į k mažesnių vienodo ilgio intervalų. Dalinio intervalo ilgį h galima parinkti taikant formulę:

$$h = \frac{x_{\max} - x_{\min}}{k}.$$

Jei iš anksto žinoma, į kiek intervalų (tarkime m), norima suskirstyti, tada

$$h = \frac{x_{\max} - x_{\min}}{m}.$$

Sudarinėjant intervalus yra laikomasi tokių taisyklių:

1 grupavimo intervalų ilgiai yra vienodi;

2 intervalai nesikerta;

3 kiekviena kintamojo reikšmė patenka tik į vieną intervalą.

Gali būti taip, kad mūsų tiriamas kintamasis įgyja labai daug skirtingų reikšmių

Pažymėkime i-ajį grupavimo intervalą $[c_{i-1}, c_i)$. Grupuodami imame atvirus iš dešinės intervalus. Aišku galima imti atvirus ir iš kairės, svarbiausia, kad duomuo patektų tik į vieną intervalą. Sugrupavus duomenis prarandame informacija apie konkrečią kiekvieno duomens reikšmę, todėl norėdami apibūdinti į konkretų intervalą pakliuvusius duomenis yra imamas to intervalo vidurys. *Tuomet visi patekę į intervalą* $[c_{i-1}, c_i)$ *duomenys yra laikomi lygiais* $(c_{i-1} + c_i)/2$, *t.y. intervalo viduriui.* Taip sugrupavus duomenis patogu yra juos surašyti į *intervalinių dažnių lentelę*.

4 lentelė

Intervalas	$[c_1, c_2)$	$[c_2, c_3)$	...	$[c_k, c_{k+1})$
Dažnis	f_1	f_2	...	f_k
Santykinis dažnis	f_1/n	f_2/n	...	f_k/n
Vidurys	$(c_1 + c_2)/2$	$(c_2 + c_3)/2$	...	$(c_k + c_{k+1})/2$

Taškai c_i, $i = 1,2,\ldots,k$ parenkami taip:

$c_1 = x_{\min}$, $c_2 = c_1 + h, \ldots, c_{k+1} = c_k + h \geq x_{\max}$.

Galima už c_1, parinkti ir kitokią reikšmę, mažesnę už $x_{\min}$, tačiau turėtų būti tenkinamos nelygybės

$$x_{\min} - c_1 < \frac{h}{2} \quad \text{ir} \quad c_{k+1} - x_{\max} < \frac{h}{2}.$$

Panagrinėkime tokį pavyzdį.

Pavyzdys. 5 lentelėje pateikti 80 studentų egzamino metu surinkti balai:

5 lentelė

62	73	85	42	68	54	38	27	32	63
68	69	75	59	52	58	36	85	88	72
52	52	63	68	29	73	29	76	29	57
46	43	28	32	9	66	72	68	42	76
38	38	39	28	19	12	78	72	92	82
72	33	92	69	28	39	85	59	68	52
85	59	76	80	72	74	54	48	29	36
10	82	58	88	68	58	46	37	29	35

Sugrupuoti duomenis ir užpildyti intervalinių dažnių lentelę.

Sprendimas. Matome, kad skirtingų reikšmių yra daug, todėl sudarinėti dažnių lentelę yra nepatogu. Tokiu atveju informatyviau - intervalinių dažnių lentelė.

Prieš sudarydami lentelę, turime nuspręsti į kiek intervalų mes suskirstysime turimą duomenų aibę. Mūsų tiriamos imties dydis $n = 80$.

Tuomet galime pasinaudoti formule $k = 1 + 3,222 \cdot \log_{10} 80 \approx 7$. Surandame didžiausią $x_{\max} = 92$ ir mažiausią $x_{\min} = 9$ reikšmes. Apskaičiuojame skirtumą tarp didžiausios ir mažiausios reikšmės $x_{\max} - x_{\min} = 83$, bei dalinio intervalo ilgį $h = \frac{x_{\max} - x_{\min}}{k} = \frac{92-9}{7} \approx 12$.

Dabar jau galime turimus duomenis sugrupuoti į septynis intervalus, kurio kiekvieno ilgis -12.

[9 , 21), [21 , 33), [33 , 45), [45 , 57), [57 , 69), [69 , 81), [81 , 93).

Galėjome pradinį intervalą imti kitokį, pvz. [5 , 17). Svarbiausia, kad visos reikšmės papultų į intervalus. Todėl gali prireikti papildomo intervalo arba padidinti dalinio intervalo ilgį.

Užpildome intervalinių dažnių lentelę:

6 lentelė

Intervalas	Dažnis	Sukauptas dažnis	Santykinis dažnis	Sukauptas santykinis dažnis	Vidurys
[9 , 21)	4	4	0.0500	0.0500	15
[21 , 33)	11	15	0.1375	0.1875	27
[33 , 45)	13	28	0.1625	0.3500	39
[45 , 57)	9	37	0.1125	0.4625	51
[57 , 69)	17	54	0.2125	0.6750	63
[69 , 81)	16	70	0.2000	0.8750	75
[81 , 93)	10	80	0.1250	1.0000	87

Iš taip sugrupuotų duomenų galime pasakyti, kad daugiausia yra studentų, surinkusių nuo 57 iki 69 (17 studentų) ir nuo 69 iki 81 (16 studentų) egzamino balų. Taip pat matome, pvz., kad tik 10 studentų arba 12.50 % visų studentų surinko daugiau nei 81 balą. Štai kodėl dažnai dalinių intervalų ilgiai priklauso nuo to, kokią informaciją mes norime gauti iš sugrupuotų duomenų.

Klausimai ir užduotys

1. Ką vadiname variacine eilute?
   a) Imties elementų, išdėstytų nemažėjančia tvarka, seka;
   b) Imties stebinių, išdėstytų nemažėjančia tvarka, seka;
   c) Imties elementų (stebinių), išdėstytų didėjančia tvarka, seka;
   d) Imties elementų, išdėstytų didėjančia tvarka, seka.
2. Ką vadiname statistine eilute?
3. Ką vadiname poligonu?

10. DUOMENŲ PADĖTIES CHARAKTERISTIKOS

Pagrindinės duomenų padėties charakteristikos yra vidurkis, moda ir mediana, apibūdinančios duomenų „centrą“, bei kvantiliai. Visos charakteristikos, išskyrus modą, skaičiuojamos tik kiekybiniams duomenims.

Vidurkis

Vidurkis – tai taškas, kuris vidutiniškai artimiausias visiems statistinės eilutės nariams. Yra skaičiuojamas tik kiekybinių duomenų vidurkis, su šia charakteristika mes jau turėjome reikalų šiame kurse. Taigi, *imties vidurkis (aritmetinis vidurkis) yra visų statistinės eilutės elementų suma, padalyta iš jų skaičiaus.* Analogiškai apibrėžiamas ir populiacijos vidurkis.

Imties vidurkis: $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$. *Populiacijos vidurkis:* $\mu = \frac{1}{N}\sum_{j=1}^{N} x_j$

Pavyzdys. Dirbančių studentų atlyginimai (Lt per mėn.) yra 500, 600, 600, 750, 850 Lt. Apskaičiuoti studentų vidutinį atlygį (atlyginimų vidurkį).

Sprendimas. Vidutinis dirbančių studentų atlygis (vidurkis) yra:

$$\bar{x} = (500 + 600 + 600 + 750 + 850)/5 = 660 Lt.$$

*****

Grupuotų duomenų vidurkiui skaičiuoti pasirenkame vidurinius intervalų taškus. Tuomet pažymėję *j*-ojo intervalo vidurio tašką $x_j^* = (c_{j-1} + c_j)/2$, o dažnį f_j, turime:

$$\textit{imties vidurkis} \quad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j * f_j = \sum_{j=1}^{n} x_j * \frac{f_j}{n}.$$

Ši formulė parodo, kad aritmetiniam vidurkiui skaičiuoti gali būti panaudoti santykiniai dažniai.

Pavyzdys. Mokslo darbuotojų atlyginimai (Lt per mėn.) yra pateikti 7 lentelėje. Apskaičiuosime atlyginių vidurkį pasinaudodami ką tik išvesta formule.

7 lentelė

Atlyginimas	Darbuotojų skaičius
1400	8
1500	10
1650	14
1700	11
1800	9
2000	5

Sprendimas.

Mokslo darbuotojų atlyginimo vidurkis yra:

$$\bar{x} = (1400 \cdot 8 + 1500 \cdot 10 + 1650 \cdot 14 + 1700 \cdot 11 + 1800 \cdot 9 + 2000 \cdot 5) / 57 \approx 1652{,}63 Lt.$$

*****

Kartais *vidurkis* nėra pati geriausia duomenų centro charakteristika. Tokiu atveju geriau naudoti kitas skaitines charakteristikas.

Panagrinėsime elementarų pavyzdį. Dauguma žmonių turi didesnį, nei vidutinį, kojų skaičių. Iš tikrųjų, tarp 3 milijonų Lietuvos gyventojų yra apie 1000 vienakojų. Taigi, vidutinis vieno gyventojo kojų skaičius yra $\bar{x} =$ (2999000*2+1000*1)/3000000=1,999666... . O kadangi dauguma turi dvi kojas ir 2>1,999666..., tai teiginys "Dauguma žmonių turi didesnį, nei vidutinį, kojų skaičių" yra įrodytas.

Matematiškai viskas teisinga, tačiau būtų sunku įsivaizduoti žmogų, turintį 1,999666... kojas. Tokiu atveju informatyviau būtų naudoti kitą skaitinę charakteristiką – imties modą (Mo).

Moda

Moda – dažniausiai duomenų aibėje pasikartojusi reikšmė.

Pavyzdžiui, duomenų aibės 1; 1; 2; 3; 4; 5 moda *Mo*=1, nes 1 pasikartoja dažniau (2 kartus) negu bet kuri kita šios duomenų aibės reikšmė. Jeigu visos reikšmės statistinėje eilutėje pasikartoja vienodai dažnai, sakoma, kad tokia duomenų aibė neturi modos. Pavyzdžiui, duomenų aibė 2; 2; 3; 3; 8; 8 modos neturi, nes visos reikšmės pasikartoja vienodu dažnumu. *Jeigu kelios gretimos variacinės eilutės reikšmės pasirodo vienodu dažniu ir šis dažnis yra didesnis, negu bet kuris kitas dažnis, tai moda yra šių reikšmių aritmetinis vidurkis.* Pavyzdžiui, duomenų aibės -1;0; 1; 1; 2; 2; 2; 3; 3; 3; 7 moda *Mo* = (2+3)/2 = 2,5.

Gali būti ir daugiau modų. *Jeigu dvi negretimos variacinės eilutės reikšmės pasikartoja vienodu dažniu ir jis didesnis negu bet kurių kitų reikšmių dažnis, tai sakoma, kad egzistuoja dvi modos ir toks dažnių skirstinys vadinamas bimodžiu.* Pavyzdžiui, statistinė eilutė 8; 9; 9; 9; 11; 12; 13; 14; 14; 14; 17 turi dvi modas – 9 ir 14. Jeigu negretimų vienodo dažnio variacinės eilutės narių yra daugiau nei du, modų taip pat yra daugiau.

Grįžtant prie nagrinėto pavyzdžio apie kojas gauname, kad moda Mo=2 , nes dvi kojos tikrai dažniau pasitaiko, nei viena.

Mediana

Tarkime, kad turime variacine eilutę

$$x_1 \le x_2 \le \ldots \le x_n$$

Imties mediana Md yra skaičius, už kurį 50% variacinės eilutės reikšmių yra nedidesnės ir 50% nemažesnės. Tiksliau medianą galime apibrėžti taip:

jeigu n nelyginis, tai mediana yra variacinės eilutės reikšmė, atitinkanti (n+1)/2 poziciją. Jeigu stebėjimų skaičius n lyginis, tai mediana yra variacinės eilutės reikšmių, atitinkančių pozicijas (n/2) ir (n/2)+1, aritmetinis vidurkis.

Taigi,

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

Panagrinėkime tokį pavyzdį:

Pavyzdys. 7 darbininkai per dieną uždirba:

125, 128, 120, 104, 147, 115, 155 Lt.

Surakite šio dienos uždarbių aibės medianą.

Sprendimas. Pirmiausia užrašome variacinę eilutę, t.y. uždarbius išdėstome didėjančia tvarka:

104, 115, 120, 125, 128, 147, 155.

Kadangi darbininkų skaičius yra nelyginis ir imties dydis $n = 7$, tai mediana yra variacinės eilutės reikšmė, atitinkanti (7+1)/2 = 4 poziciją. Gauname, kad $Md = 125$. Matome, kad trys uždarbiai : 104, 115 ir 120 yra mažesni ir trys – 128, 147, 155 – didesni, nei mediana.

Papildykime mūsų sąlygą dar vienu uždarbiu. T.y. tegu aštuoni darbininkai uždirba atitinkamai:

125, 128, 120, 104, 147, 115, 155, 160 Lt.

Suraskime šių uždarbių medianą. Užrašome variacinę eilutę:

104, 115, 120, 125, 128, 147, 155, 160.

Dabar, ieškodami medianos, jau atkreipiame dėmesį į tai, kad darbininkų skaičius yra lyginis ir imties dydis $n = 8$, todėl mediana bus variacinės eilutės reikšmių, atitinkančių pozicijas (8/2) = 4 ir (8/2)+1 = 5, aritmetinis vidurkis. Taigi,

$$Md = \frac{125+128}{2} = 126.5$$

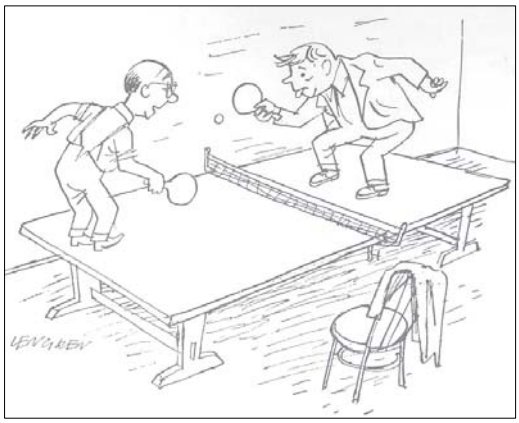
*****

Kaip ir *aritmetinis vidurkis, mediana* charakterizuoja duomenų centrą. Iškyla logiškas klausimas: jei tiek *aritmetinis vidurkis,* tiek *mediana* charakterizuoja duomenų centrą, tai kam reikalingos net dvi centro charakteristikos?

Medianą yra patariama naudoti kaip duomenų centro charakteristiką, kai yra *išskirčių*.

Išskirtis – tai tokia duomenų aibės reikšmė, kuri yra nenatūraliai didesnė arba mažesnė už kitas reikšmes.

Išskirtis – tai reiškinių išskyrimas iš jų natūralaus arba istorinio sąryšio



– Ar jūs esate tikras, kad „šitą" mes žaidžiame taip pat natūraliai, kaip visi?

Panagrinėkime dvi duomenų aibes: 10; 20; 30; 40 ir 10; 20; 30; 100. Abiejų aibių medianos yra lygios – 25, nes n-lyginis, todėl $Md = \frac{x_2 + x_3}{2} = \frac{20+30}{2} = 25$.

Apskaičiuokime abiejų aibių aritmetinius vidurkius. Pirmosios aibės vidurkis yra $\overline{x}_1 = \frac{10+20+30+40}{4} = 25$, o antrosios – $\overline{x}_2 = \frac{10+20+30+100}{4} = 40$.

Aišku, kad antruoju atveju *aritmetinis vidurkis* nėra pati tinkamiausia centro charakteristika, nes 75% visų reikšmių yra mažesnės už vidurkį. O tai nulemia vienintelė didelė reikšmė – *išskirtis* 100.

Ir *vidurkis*, ir *moda*, ir *mediana* yra duomenų centro charakteristikos. Kokią charakteristiką geriau naudoti, priklauso nuo tiriamos duomenų aibės, taip pat nuo tyrimo tikslų. Panagrinėkime tokį pavyzdį.

Pavyzdys. Nedidelėje firmoje dirbančių pareigos ir alga yra nurodyti 8 lentelėje.

8 lentelė

Pareigos	**Alga**
Direktorius	8500
Vadybininkas	6500

Buhalteris	6000
1 vairuotojas	2100
2 vairuotojas	2000
3 vairuotojas	1900
Sargas	700
Valytoja	700

Sprendimas. Vidutinė firmos darbuotojų alga (vidurkis):

$$\bar{x} = \frac{8500 + 6500 + 6000 + 2100 + 2000 + 1900 + 2*700}{8} = 3550.$$

Moda *Mo* = 700 , nes 700 yra dažniausiai firmoje pasikartojanti alga.

Žiūrint algų lentelėje iš apačios į viršų gauname variacine eilutę, o kadangi $n = 8$ (lyginis), tai mediana yra $Md = \frac{x_4 + x_5}{2} = \frac{2000 + 2100}{2} = 2050$.

Taigi, matome, kad vidutinė alga – tai, vaizdžiai kalbant, visų darbuotojų algos, sudėtos į krūvą ir padalintos po lygiai. Matome, kad tik trys firmos darbuotojai (direktorius, vadybininkas ir buhalteris) uždirba daugiau – ir gerokai daugiau, – nei vidurkis. Moda *Mo* = 700 yra alga, kurią gauna daugiausiai firmos darbuotojų (sargas ir valytoja). Mediana *Md* = 2050 yra visų algų, išrikiuotų pagal didumą (sudarius algų variacine eilutę), viduriukas. Todėl atsakymas į klausimą, kuri charakteristika yra tinkamesnė ir kokia prasme tikslesnė priklauso nuo to, ką mes norime ištirti.

Kvantiliai

*Charakteristika, dalijanti variacinę eilutę į q*100 ir (1-q)*100 procentinių dalių, vadinama q-osios (0<q<1) eilės kvantiliu.*

Kvantiliui skaičiuoti V. Čekanavičius, G. Murauskas vadovėlyje „Statistika ir jos taikymai. I“ (Vilnius, TEV, 2000) rekomenduoja naudotis tokiu algoritmu:

1) Iš pradžių surandame indeksą *i*:

$$i = q*n.$$

2) Jeigu *i* nėra sveikasis skaičius, tai imama sveikoji jo dalis [*i*]. Ir tada ieškomas *kvantilis* yra [*i*]+1 *variacinės eilutės* narys, t.y. $x_{([i]+1)}$.

3) Jeigu sudauginus q^*n gauname, kad i yra sveikasis skaičius, tai ieškomasis *kvantilis* yra $(x_{([i])} + x_{([i]+1)})/2$.

Pavyzdys. Tarkime, norime rasti algų lentelės (žr. 8 lentelę) duomenų 15% (q = 0,2) kvantilį.

Sprendimas. Pastebime, kad i = 0,15 * 8 = 1,2. Kadangi i nėra sveikasis skaičius, tai imame jo sveikąją dalį $[i]$ = 1. Gauname, kad ieškomasis *kvantilis* yra variacinės eilutės $[i]$+1 narys, t.y $x_{(1+1)} = x_{(2)} = 700$.

*****

Kvantiliai, dalijantys variacinę eilutę į keturias maždaug lygias dalis, vadinami kvartiliais. Jie žymimi Q_1, Q_2, Q_3.

Kvartilių radimo būdą galime pavaizduoti taip (žr. 2 pav.): Q_2 sutampa su *mediana* ir dalija *imtį* į dvi dalis; tuomet Q_1 yra apatinės dalies *mediana*, o Q_3 yra viršutinės dalies *mediana*.

Taigi Q_1 - 25% *kvantilis*, $Q_2 = Md$ – 50% *kvantilis*, Q_3 - 75% *kvantilis*.

Klausimai ir užduotys

1. Imties vidurkis (aritmetinis vidurkis) yra visų statistinės eilutės elementų suma, padalyta iš jų skaičiaus. Ar imties vidurkis
   a) sutampa su populiacijos vidurkiu?
   b) yra didesnis už populiacijos vidurkį?
   c) yra mažesnis už populiacijos vidurkį?
   d) nebūtinai sutampa su populiacijos vidurkiu?
   e) nieko bendro su populiacijos vidurkiu neturi?
2. Kodėl aritmetinis vidurkis nėra pati tinkamiausia centro charakteristika?
3. Jei tiek aritmetinis vidurkis, tiek mediana charakterizuoja duomenų centrą, tai kam reikalingos net dvi centro charakteristikos?

11. DUOMENŲ SKLAIDOS CHARAKTERISTIKOS

Pati paprasčiausia sklaidos charakteristika yra *duomenų aibės plotis*:

$$r = x_n - x_1 = x_{\max} - x_{\min}$$

Jis parodo intervalo, kuriame išsidėsčiusios (t.y. išsisklaidžiusios) visos reikšmės, ilgį.

Dispersija

Studijuodami tikimybių teorijos skyrių matėme, kad *dispersija* parodo, kaip požymio reikšmės *imtyje* yra pasklidusios (išsibarsčiusios) vidurkio atžvilgiu. *Vidurkis* charakterizuoja duomenų centrą, tačiau nieko nepasako apie tai, kaip dažnai požymio reikšmės *imtyje* yra nutolusios nuo šio duomenų centro, koks jų *susitelkimas* apie vidurkį. Norint įvertinti šią duomenų sklaidą apie vidurkį, reikėtų imti skirtumus, susidarančius *tarp* konkrečios *požymio reikšmės* ir visų reikšmių *vidurkio*. Tačiau realiosios požymių reikšmės nukrypsta nuo *vidurkio į abi puses* ir aritmetinė visų skirtumų suma, kaip matėme, yra lygi nuliui.

Todėl sumuoti pačius skirtumus yra beprasmiška. Vietoje jų yra imami jų *kvadratai*. Tokiu atveju gauname tuos skirtumus tiesiogiai atspindinčius *teigiamus* dydžius, kuriuos jau galima sumuoti. Gautąsias sumas priimta yra dalinti iš vienetu sumažinto imties dydžio. Šis gautasis dydis ir yra vadinamas *imties dispersija:*

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

Pavyzdys. Palyginkime dviejų firmų programuotojų atlyginimus. Pirmoje firmoje dirba keturi programuotojai, per mėnesį uždirbantys atitinkamai 3000; 4000; 4000 ir 7000 Lt. Antroje firmoje dirba taip pat keturi programuotojai, uždirbantys atitinkamai 4500; 4500; 4500; 4500 Lt.

Sprendimas. Abiejų firmų atlyginimų *vidurkis* yra 4500 Lt. Iš tiesų:

$$\bar{x}_1 = \frac{3000 + 4000 + 4000 + 7000}{4} = 4500$$

$$\bar{x}_2 = \frac{4500 + 4500 + 4500 + 4500}{4} = 4500$$

Apskaičiuojame abiejų firmų programuotojų atlyginimų *dispersijas*. Gauname, kad pirmosios ji yra

$$s_1^2 = \frac{1}{3}((3000-4500)^2 + (4000-4500)^2 + (4000-4500)^2 + (7000-4500)^2) = 3000000.$$

Antrosios firmos atlyginimų dispersija:

$$s_2^2 = \frac{1}{3}((4500-4500)^2 + (4500-4500)^2 + (4500-4500)^2 + (4500-4500)^2) = 0$$

Taigi pirmosios firmos atlyginimų *dispersija* (sklaida) didelė, o antrosios maža – jos iš viso nėra.

*****

Standartinis nuokrypis

Dydis, gaunamas ištraukus kvadratinę šaknį iš dispersijos, vadinamas standartiniu nuokrypiu.

Kaip ir dispersija, standartinis nuokrypis parodo vidutinę duomenų sklaidą apie vidurkį.

Kam tada reikia pereiti nuo *dispersijos* prie *standartinio nuokrypio*? Pirmiausia pastebime, kad *standartinis nuokrypis* matuojamas tokiais pačiais vienetais, kaip ir patys duomenys. Prisiminkime mūsų nagrinėtą pavyzdį apie dviejų firmų atlyginimus. Pastebime, kad ir patys duomenys, ir vidurkis, ir standartinis nuokrypis matuojami litais. Tuo tarpu dispersijos matavimo vienetai būtų litai kvadratu. Todėl standartinį nuokrypį lengviau interpretuoti ir lyginti su duomenimis.

Imties standartinis nuokrypis:

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

Taigi mūsų nagrinėtų dviejų firmų *standartiniai nuokrypiai* yra :

$$s_1 = \sqrt{2250000} \approx 474.34\,\text{Lt};\ s_2 = \sqrt{0} = 0\,\text{Lt}.$$

Taigi iš šių duomenų jau galime teigti, kad pirmosios firmos algos yra vidutiniškai 474 Lt pasklidusios (išsivarsčiusios, nutolusios) vidurkio atžvilgiu, o antrosios – visiškai nepasisklidusios vidurkio atžvilgiu. Kitaip sakant, antrosios firmos darbuotojų visos algos yra vienodo dydžio.

Klausimai ir užduotys

1. Nurodykite, kuris iš žemiau išvardytų teiginių yra klaidingas:
   a) Dispersija parodo, kaip požymio reikšmės imtyje yra pasklidusios (išsibarsčiusios) vidurkio atžvilgiu.
   b) Vidurkis charakterizuoja duomenų centrą, tačiau nieko nepasako apie tai, kaip dažnai požymio reikšmės imtyje yra nutolusios nuo šio duomenų centro, koks jų susitelkimas apie vidurkį.
   c) Norint įvertinti šią duomenų sklaidą apie vidurkį, reikėtų imti skirtumus, susidarančius tarp konkrečios požymio reikšmės ir visų reikšmių vidurkio. Tačiau realiosios požymių reikšmės nukrypsta nuo vidurkio į abi puses ir aritmetinė visų skirtumų suma yra lygi nuliui.
   d) Visi trys teiginiai a), b) ir c) yra klaidingi.
2. Kokia paprasčiausia duomenų sklaidos charakteristika?

3. Kodėl naudinga pereiti nuo dispersijos prie standartinio nuokrypio?

12. GRAFINIS STEBĖJIMŲ VAIZDAVIMAS

Grafikas yra vaizdinė priemonė glaustai tiek pradiniams duomenims, tiek ir analizės rezultatams pateikti. Grafiniai elementai lengvai suvokiami, todėl grafikas suteikia daugiau informacijos nei „pliki" skaičiai. Šio skyriaus tikslas nurodyti pagrindinius požymius, charakterizuojančius gerą grafiką. Tipiniai reikalavimai grafikams yra:

1 *Aiškumas - grafikai turi būti suvokiami be papildomų aprašymų.*

2 *Skiriamoji galia - kiekvienas grafiko elementas turi būti lengvai įžiūrimas.*

3 *Kopijuojamumas - grafiko kopija (pvz., nespalvota) turi likti informatyvi.*

Nėra paprasta „išgauti" visą informaciją, esančią duomenyse. Negalima tikėtis pateikti jos visos vienu grafiku. Tenka braižyti daug grafikų, iš kurių kiekvienas padeda atskleisti (aprėpti) vis daugiau informacijos apie duomenis. Tačiau negalima grafiko perkrauti. Grafikų įvairovė yra labai didelė, todėl nėra galimybės visų jų aprašyti. Būdingiausi grafikų tipai yra: histograma, stulpelinė, skritulinė, sklaidos, stačiakampė bei linijų diagramos.

Stulpelinė diagrama, histograma

Stulpelinės diagramos ir *histogramos* apibrėžimus mes detaliai nagrinėjome šio Statistikos pradmenų skyriaus pradžioje (žr. skyrelį *Statistinių duomenų apdorojimas*) – skaitytojui belieka šiuos apibrėžimus pakartoti. Tačiau naują realų pavyzdį panagrinėsime detaliai.

Pavyzdys. 77 studentų surinkti testo rezultatai yra pateikti 9 lentelėje:

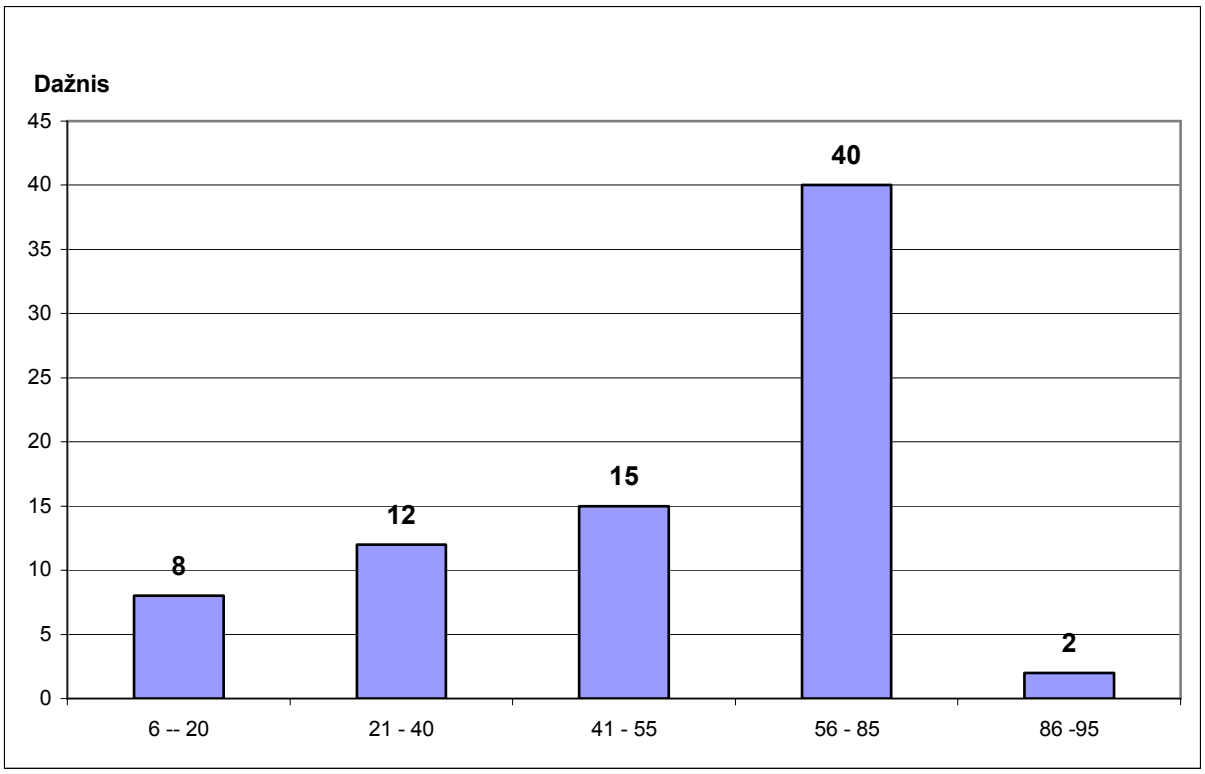
9 lentelė

Kategorija (balų klasė)	**Dažnis (studentų skaičius)**
6 - 20	8
21- 40	12
41 - 55	15
56 - 85	40
86 - 95	2
Viso	77

Sudarykite gautų rezultatų stulpelinę diagramą ir histogramą.

Sprendimas. Mūsų pavyzdyje surinktų balų klasėje nuo 6 iki 20 skirtumas gali siekti 15 balų, klasėje nuo 21 iki 40 šis skirtumas gali siekti 20 balų, klasėje nuo 41 iki 55 - 15 balų, klasėje nuo 56 iki 85 – 30 balai ir klasėje nuo 86 iki 95 – 10 balų.

Taigi, klasių plotis (ši sąvoka irgi buvo detaliai aiškinama) gali būti skirtingas, tačiau diagramos stulpelių plotis vienodas. Tai naudinga tuo atveju, kai pagrindinis dėmesys koncentruojamas į dažnius. Mūsų nagrinėjamame pavyzdyje 77 studentų surinktų testo balų stulpelinė diagrama (žr. 3 pav.) yra tokia:



Klasių intervalais šiuo atveju yra blokai 6-20, 21-40, 41-55, 56-85 ir 86-95. Kadangi apatinis klasės intervalo galas vadinamas *klasės apatine riba,* o viršutinis galas - *klasės viršutinė riba*, tai šiose ribose turi tilpti nagrinėjamos klasės intervalas. Taigi, mūsų pavyzdyje 6-0.5, 21-0.5, 41-0.5, 56-0.5, 86-0.5 yra klasės apatinės ribos, o 20+0.5, 40+0.5, 55+0.5, 85+0.5, 95+0.5 - viršutinės ribos. Kodėl iš apatinės ribos atimam 0.5, o prie viršutinės – pridedam 0.5 ? Tai daroma tam, kad vieno studento surinktų balų skaičius patektų tik į vieną balų klasę.

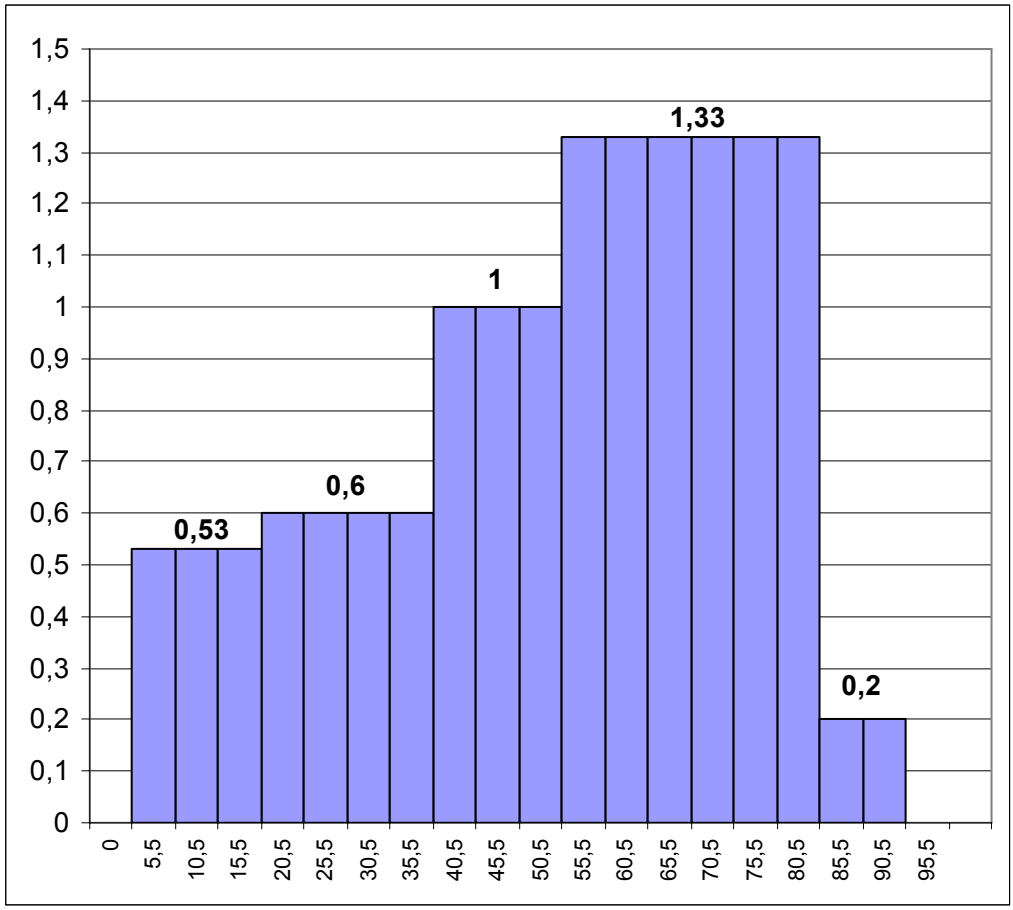
Kadangi histograma - tai kiekybinio požymio dažnių skirstinio grafinė išraiška, susidedanti iš keleto besiribojančių stačiakampių, kurių kiekvieno pagrindas lygus klasės pločiui, o plotas lygus

klasės santykiniam dažniui, tai papildome mūsų turimą 9 lentelę klasės pločiu, klasės santykiniu dažniu ir ribiniais taškais:

10 lentelė

Kategorija (balų klasė)	Dažnis (studentų skaičius)	Klasės plotis	Ribiniai taškai	Klasės santykinis dažnis (dažnis/klasės plotis)
6 - 20	8	15	5.5	8/15 = 0.53
21- 40	12	20	20.5	12/20 = 0.60
41 - 55	15	15	40.5	15/15 = 1.00
56 - 85	40	30	55.5	40/30 = 1.33
86 - 95	2	10	85.5	2/10 = 0.20
			95.5	

Taigi, mūsų pavyzdyje apie studentų surinktus testo balus histograma yra tokia (skaičius virš stačiakampio išreiškia jo aukštį):



Skritulinė diagrama

Skritulinė diagrama naudojama kaip alternatyva *stulpelinei diagramai*. Ji perteikia tą pačią informaciją, tik kita forma. Skritulys atitinka visą populiaciją (100%), o išpjovos - kategorijas, proporcingas jų santykiniam dažniui.

Norint pavaizduoti išsamią *skritulinę diagramą*, reikia žinoti šias taisykles:

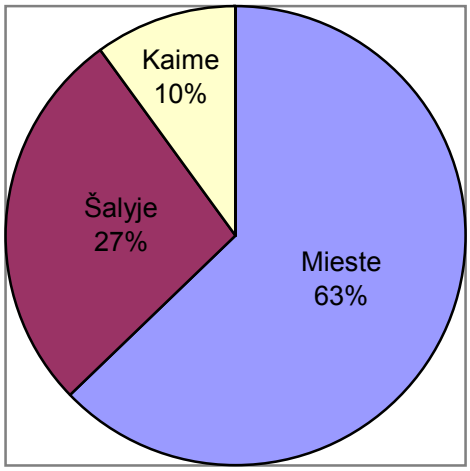
1 Diagramoje išpjovos išdėstomos mažėjimo tvarka pagal laikrodžio rodyklę pradedant 12-aja pozicija.

2 Skritulinė diagrama yra per daug marga ir nevaizdi, jei kategorijų skaičius didesnis už 5 arba mažiausia išpjova yra mažesnė nei 3% viso skritulio.

Panagrinėsime konkretų pavyzdį – skritulinę diagramą, kurioje pavaizduotas vidutinis moksleivių skaičius vienoje mokykloje 1999-2000 metais (žr. atitinkamą pav.) pagal 11 lentelę:

11 lentelė

Mieste	582
Kaime	93
Šalyje	251



Taškinės (sklaidos) ir linijinės diagramos

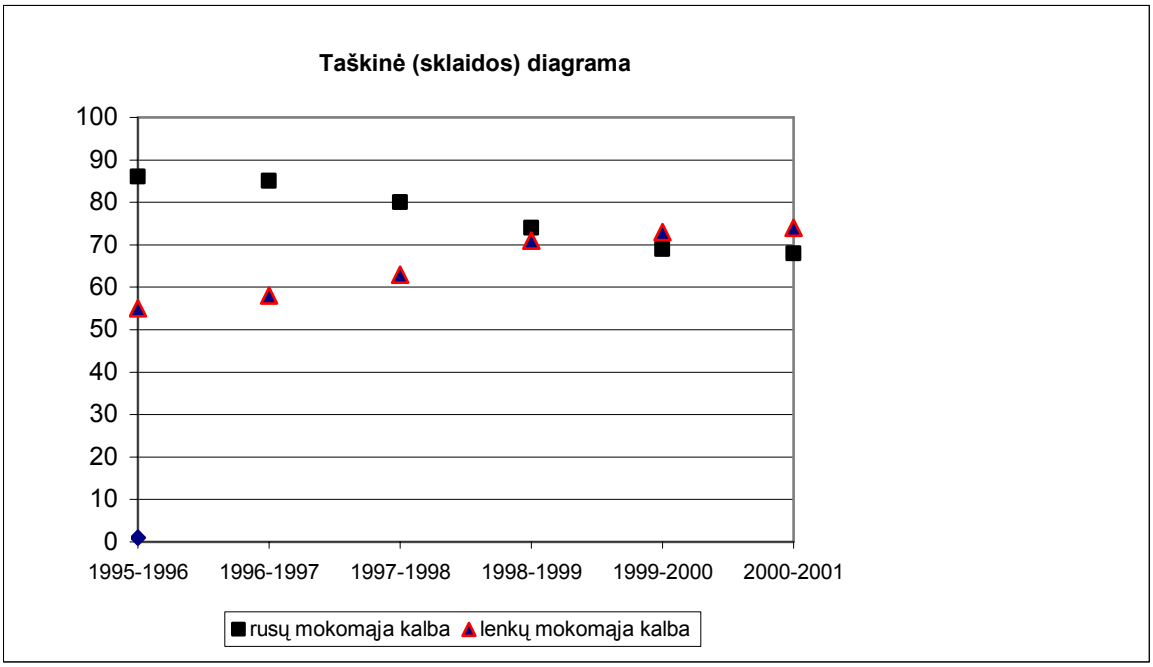
Tai ne tokios išvaizdžios kaip anksčiau minėtosios, bet kartu – už jas dažnai net „skaidresnės“ bei informatyvesnės diagramos, kuriose požymio reikšmių dažniai vaizduojami taškais. Pagal šias diagramas apie dažnių tarpusavio santykius sprendžiama iš taškų aukščių, t.y. jų padėties *y*-ų ašies atžvilgiu. Jeigu taškai yra atskiri, tarpusavyje nesujungti, tai tokios diagramos vadinamos taškinėmis.

Kartais, ypač – kai taškų atsiranda daug, sakysim, vaizduojant kelis požymius vienoje diagramoje, pravartu būna taškus, atitinkančius vaizduojamą požymį, dirbtinai sujungti linijomis. Taip gaunama *linijinė diagrama*, kuri yra parankesnė dėl to, kad linijos tiesiog padeda akims greičiau susekti rūpimus taškus.

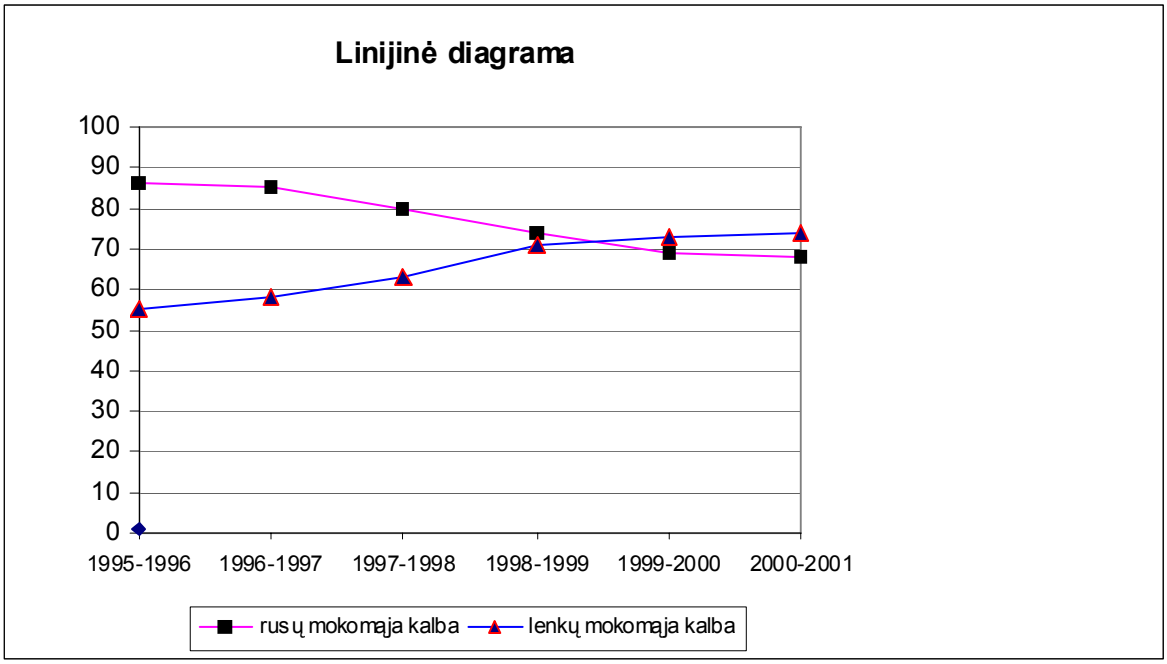
Žemiau galite palyginti dvi diagramas: taškinę ir linijinę (žr. 1 ir 2 pav.), kuriose pavaizduotas mokyklų skaičius Lietuvoje lenkų ir rusų mokomąja kalba.

12 lentelė

Metai	Mokyklų skaičius rusų mokomąja kalba	Mokyklų skaičius lenkų mokomąja kalba
1995-1996	86	55
1996-1997	85	58
1997-1998	80	63
1998-1999	74	71
1999-2000	69	73
2000-2001	68	74



1 pav.



2 pav.

Klausimai ir užduotys

1. Nurodykite, kuris iš žemiau išvardytų teiginių yra klaidingas:
    a) Skritulinė diagrama naudojama kaip alternatyva stulpelinei diagramai. Ji perteikia tą pačią informaciją, tik kita forma. Skritulys atitinka visą populiaciją (100%), o išpjovos - kategorijas, proporcingas jų santykiniam dažniui.
    b) kai taškų atsiranda ypač daug, sakysim, vaizduojant kelis požymius vienoje diagramoje, pravartu būna taškus, atitinkančius vaizduojamą požymį, dirbtinai sujungti linijomis. Taip gaunama linijinė diagrama, kuri yra parankesnė dėl to, kad linijos tiesiog padeda akims greičiau susekti rūpimus taškus.
    c) Sklaidos diagrama nukrypsta nuo vidurkio į abi puses ir aritmetinė visų skirtumų suma yra lygi nuliui.
    d) Visi trys teiginiai a), b) ir c) yra klaidingi.
2. Įrašykite trūkstamus skaičius teiginyje: *Skritulinė diagrama yra per daug marga ir nevaizdi, jei kategorijų skaičius didesnis už ..... arba mažiausia išpjova yra mažesnė nei .....% viso skritulio.*
3. Kokie yra tipiniai reikalavimai grafikams?

13. GRUPUOTOJI STATISTINĖ EILUTĖ IR JOS HISTOGRAMA

Kai turime daug tolydžiojo atsitiktinio dydžio stebėjimų, dažnių lentelė tampa nebeinformatyvi – joje yra labai daug skirtingų reikšmių. Kartu dingsta didžiausias dažnių lentelės privalumas, nes informacija nebekoncentruojama. Be to, kai kurie stebiniai gali labai mažai skirtis tarpusavyje, nes pakankamai tikslaus matavimo atveju praktiškai nebus sutampančių imties elementų, tokio sutapimo tikimybė tolygaus skirstinio atveju lygi nuliui.

Taigi, tokius duomenis reikia grupuoti. Su grupavimo procedūra mes susipažinome skyrelyje *Įvairūs imties aprašymo būdai.* Matėme, kad pirmiausia reikia nustatyti grupavimo intervalų skaičių, jų plotį ir, pagaliau, reikia nustatyti grupavimo intervalų kraštinius taškus.

Grupavimo intervalų ilgiai vienodi, intervalai nesikerta, kiekvienas stebinys patenka tik į vieną intervalą. Kuo grupavimo intervalų skaičius didesnis, tuo mažiau informacijos prarandame. Tačiau pernelyg didindami šį skaičių, rizikuojame dažnių lentelę padaryti neinformatyvia. Imties elementų grupavimas atliekamas taip. Intervalas $[x_{min}, x_{max}]$ dalomas į k lygių pagal ilgį intervalų. Pažymėkime juos $\Delta_1, \Delta_2, ..., \Delta_k$. Jei taškus, skiriančius šiuos *grupavimo intervalus*, pažymėsime $a_0, a_1, ..., a_k$ kur $a_0 = x_{\min}$, $a_k = x_{\max}$, tai $\Delta_1 = [a_0, a_1]$, $\Delta_2 = (a_1, a_2], ..., \Delta_i = (a_{i-1}, a_i], ..., \Delta_k = (a_{k-1}, a_k]$.

Tegu į intervalą Δ_i papuolė n_i elementų. Tada skaičiai $n_1, n_2, ..., n_k$ vadinami *intervaliniais dažniais, o intervalų* $\Delta_1, \Delta_2, ..., \Delta_k$ *seka su atitinkamais intervalinias dažniais* $n_1, n_2, ..., n_k$ *vadinama* **grupuotąja statistine eilute.**

Duomenis sugrupavus, dingsta informacija apie konkrečią kiekvieno duomens (stebinio) reikšmę. Todėl į konkretų intervalą pakliuvusius duomenis apibūdinti imamas *intervalo vidurys* z_i. Iškyla natūralus klausimas – kaip parinkti intervalų skaičių *k?* Šiam tikslui mes naudojome Sturgeso taisyklę. Tačiau galimi ir kiti variantai. Pavyzdžiui, galima rekomenduoti pusiau empirinę (empirinis – *paremtas patyrimu*) formulę

$$k \approx 1,72\sqrt[3]{n};$$

čia n – imties dydis, $30 \le n \le 1000$. Pastebėsime, kad:

$$n = 10 \Rightarrow k = 6;$$
$$n = 100 \Rightarrow k = 8;$$
$$n = 200 \Rightarrow k = 10;$$
$$n = 400 \Rightarrow k = 12;$$
$$n = 1000 \Rightarrow k = 17;$$

Intervalų $\Delta_1, \Delta_2, ..., \Delta_k$ ilgis h apibrėžiamas taip:

$$h = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k}.$$

Vietoj patekusių į intervalų Δ_i elementų grupės nagrinėjamas jų vienas atstovas – intervalo Δ_i vidurio taškas z_i.

Jei grupuotoji statistinė eilutė analizuojama be kompiuterio, tai techniškai ši procedūra atliekama taip.

Kiekvieną imties elementą priskiriame į atitinkamą intervalą Δ_i, darydami atitinkamą atžymą |. Susikaupus ketvertui šių ženklų |||| ir pasirodžius šiame intervale penktajam elementui, jie perbraukiami: 𝍸. Taip užfiksuojami penki stebiniai ir procedūra pradedama iš naujo.

Išnagrinėsime konkretų pavyzdį.

Pavyzdys. Atlikta 100 plieno takumo $\sigma_s (kg/mm^2)$ ribos matavimų. Gauti tokie duomenys:

26.7	37.0	30.5	28.1	27.4	25.4	36.1	33.4	29.7	26.4
34.0	26.6	27.9	35.4	35.3	29.6	29.4	26.4	37.0	28.7
33.7	31.2	34.5	31.8	25.5	30.2	32.7	28.3	**39.9**	33.5
34.1	30.0	35.8	30.7	25.9	31.6	34.4	31.5	31.8	27.4
24.7	33.4	33.1	33.2	30.3	31.6	35.8	32.2	35.2	30.8
37.0	26.3	30.2	31.8	32.5	29.7	28.0	32.4	32.3	31.9
26.7	34.7	33.6	30.7	31.9	30.8	32.3	30.4	33.8	29.5
29.7	31.8	30.1	32.4	30.6	28.1	32.2	30.5	30.8	33.4
28.7	32.6	32.7	32.3	29.8	30.6	37.2	38.4	35.0	33.1
30.6	35.4	25.6	33.5	32.0	31.6	26.1	29.4	36.4	28.0

Apskaičiuoti šios imties empirinį vidurkį ir empirinę dispersiją, nubrėžti histogramą ir poligoną.

SPRENDIMAS.

Atidžiai peržiūrėję visus duomenis nustatome, kad $x_{\min} = 24.7$, $x_{\max} = 39.9$. Tai reiškia, kad

$R = x_{\max} - x_{\min} = 39.9 - 24.7 = 15.2$; $k = 8$; $h = R/k = 15.2/8 = 1.9$.

Intervalą [24.7 ; 39.9] dalome į 8 dalis ir apskaičiuojame dažnius. Gauname tokią lentelę:

Intervalo nr.	Intervalų kraštai		Atžymų skaičius	n_i	Intervalo vidurio taškas
	a_{i-1}	a_i			
1	24.7	26.6	𝍸 𝍸 |	10	1
2	26.6	28.5	𝍸 𝍸	10	2
3	28.5	30.4	𝍸 𝍸 𝍸 |	16	3
4	30.4	32.3	𝍸 𝍸 𝍸 𝍸 ||	27	4

5	32.3	34.2	𝍸 𝍸 𝍸 ||||	19	5
6	34.2	36.1	𝍸 𝍸 |	11	6
7	36.1	38.0	𝍸	5	7
8	38.0	39.9	||	2	8
				100	

Apskaičiuosime šios imties empirinį vidurkį ir empirinę dispersiją. Tuo tikslu pastebime, kad

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (z_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i ({z_i}^2 - 2\bar{x}z_i + \bar{x}^2) = \frac{1}{n-1}\sum_{i=1}^{k} n_i {z_i}^2 - \frac{2\bar{x}}{n-1}\frac{n}{n}\sum_{i=1}^{k} n_i z_i +$$

$$+ \frac{n}{n-1}\bar{x}^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i {z_i}^2 - \frac{2n}{n-1}\bar{x}\bar{x} + \frac{n}{n-1}\bar{x}^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i {z_i}^2 - \frac{n}{n-1}\bar{x}^2. \qquad (1)$$

Sudarome lentelę:

i	n_i	z_i	z_i^2	$n_i z_i$	$n_i z_i^2$
1	10	25.65	657.9	230.85	5921.1
2	10	27.55	759.0	303.05	8349.0
3	16	29.45	867.3	471.20	13876.8
4	27	31.35	982.8	846.45	26535.6
5	19	33.25	1105.6	631.75	21006.4
6	11	35.15	1235.5	386.65	13590.5
7	5	37.05	1372.7	185.25	6863.5
8	2	38.95	1517.1	77.90	3034.2
Σ	100	–	–	3133.1	99177.1

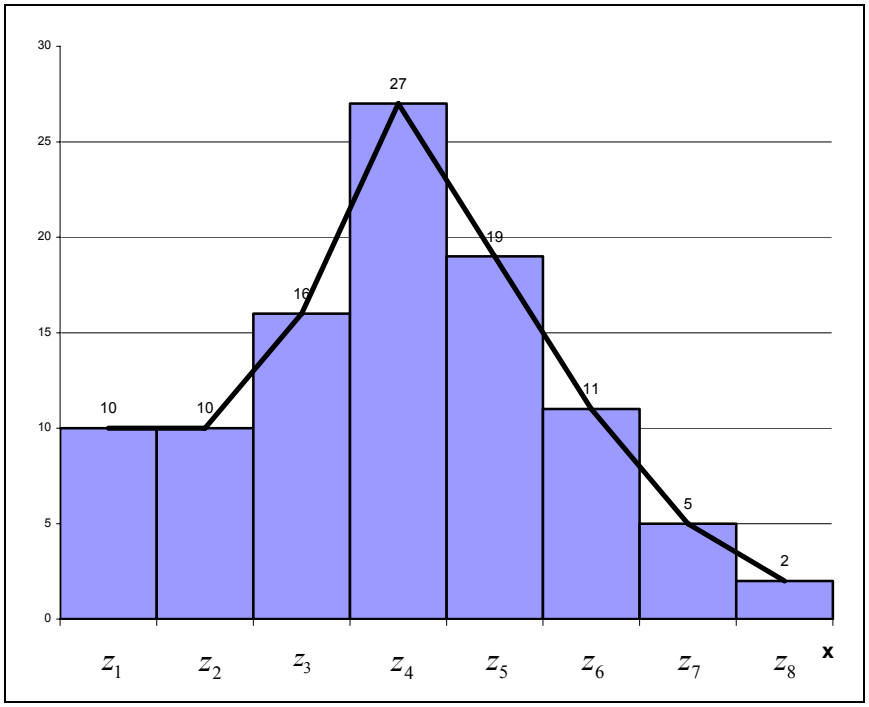
Iš čia nesunkiai gauname, kad $\bar{x}$ = 3133.1/100 = 31.331, o iš formulės (1) gauname, kad

$$s^2 = 99177.1/99 - \frac{n}{n-1}\bar{x}^2 = 99177.1/99 - \frac{100}{99}(31.331)^2 = 10.2418.$$

Grupuotąją statistinę eilutę patogu vaizduoti histograma. Šiuo atveju histogramos apibrėžimą patogu performuluoti taip: **Imties histograma** *vadinama figūra, sudaryta iš stačiakampių su pagrindais* Δ_i *ir aukščiais* $n_i/(nh)$*, kur i=1, 2, ..., k.*

Grupuotąją statistinę eilutę patogu vaizduoti ir *poligonu* – laužte su viršūnėmis taškuose (z_i , $n_i/(nh)$).

Mūsų nagrinėjamo pavyzdžio atveju histogramą ir poligoną galima pavaizduoti taip:



Šios geometrinės charakteristikos leidžia iškelti hipotezę, kad nagrinėjama populiacija turi normalųjį skirstinį, nes laužtės, ribojančios histogramą, ir poligono kontūrai artimi normaliojo tankio grafikui.

Klausimai ir užduotys

1. Nurodykite, kuris iš žemiau išvardytų teiginių yra klaidingas:
    a) Grupavimo intervalų ilgiai vienodi, intervalai nesikerta, kiekvienas stebinys patenka tik į vieną intervalą.
    b) Kuo grupavimo intervalų skaičius didesnis, tuo mažiau informacijos prarandame.
    c) Pernelyg didindami intervalų skaičių, rizikuojame dažnių lentelę padaryti neinformatyvia.
    d) Visi trys teiginiai a), b) ir c) yra klaidingi.
2. Kaip nustatyti grupavimo intervalų skaičių?

LITERATŪRA

1. Aksomaitis. Tikimybių teorija ir statistika. Kaunas, Technologija, 2002, 347 psl.
2. V. Čekanavičius, G. Murauskas. Statistika ir jos taikymai. I, Vilnius, TEV, 2000, 239 psl.
3. K.J. Hastings. Probability and statistics. – New York, Addison-Wesley, 1997, 414 pages
4. R. Januškevičius. Statistikos įvadas. Vilnius, VPU, 2000, 172 psl.
5. Richard A. Johnson, Gouri K. Bhattacharyya. Statistics: Principles and Methods, 5th Edition, John Wiley & Sons, 2005, 736 pages
6. Richard A. Johnson, Gouri K. Bhattacharyya. Statistics: Principles and Methods, third Edition, John Wiley & Sons, 1996, 720 pages
7. J. Kubilius. Tikimybių teorija ir matematinė statistika. Vilnius, Mokslas, 1996, 439 psl.
8. R. Khazanie. Statistics in a world of applications. – New York, HarperCollins, 1996, 887 pages
9. Lietuvių kalbos žodynas, http://www.lkz.lt
10. Lietuvos standartas LST ISO 3534-1. Statistika. I dalis. Tikimybių ir bendrieji statistikos terminai, 1996, 65 psl.
11. Под ред. Ю. Д. Максимова. Вероятностные разделы математики. Санкт-Петербург, «Иван Федоров», 2001, 589 psl.
12. Б. В. Гнеденко. Курс теории вероятностей. М: “УРСС”, 2001, 318 psl.
13. R. Razmas, J. Teišerskis, V. Vitkus. Matematikos uždavinynas XI-XII klasei, 4-asis papild. leid., Kaunas, Šviesa, 1999, 181 psl.

PIEŠINIAI IR NUOTRAUKOS

1. Akademiko A. Kolmogorovo nuotraukos iš http://www.kolmogorov.pms.ru/
2. Dailininkų Г. Бойко, И. Шалито piešiniai iš serijos „Eureka“ knygos Л. Бобров “Фундамент оптимизма”, Москва, Молодая Гвардия, 1976, 208 стр.
3. Dailininkų А. Колли, И. Чураков piešiniai iš serijos „Eureka“ knygos Л. Бобров “По следам сенсаций”, Москва, Молодая Гвардия, 1966, 272 стр.
4. Dailininkų А. Колли, И. Чураков piešiniai iš serijos „Eureka“ knygos В. Португал “Беседы об АСУ”, Москва, Молодая Гвардия, 1977, 208 стр.
5. Dailininkų А. Колли, И. Чураков piešiniai iš serijos „Eureka“ knygos Я. Коломинский “Беседы о тайнах психики”, Москва, Молодая Гвардия, 1976, 208 стр.

6. Dailininko K. Мошкин piešiniai iš serijos „Eureka“ knygos Л. Растригин, П. Граве “Кибернетика как она есть”, Москва, Молодая Гвардия, 1975, 208 стр.
7. Dailininko K. Мошкин piešiniai iš serijos „Eureka“ knygos Р. Петров “Беседы о новой иммунологии”, Москва, Молодая Гвардия, 1976, 224 стр.
8. Dailininko В. Ковынев piešiniai iš serijos „Eureka“ knygos А. Шилейко, Т. Шилейко “Информация или интуиция?”, Москва, Молодая Гвардия, 1983, 208 стр.
9. Ленгрен. 100 юмористических рисунков, RSW “Prasa”, Warszawa, 1957, 130 s.

PADĖKA

Autoriai nuoširdžiai dėkoja šiuos magistrantus (dauguma iš jų šiandien jau yra magistrai), aktyviai talkinusius, rengiant šį leidinį ar atskiras jo dalis:

- Kristiną Aldošiną,
- Virginiją Baranauskaitę,
- Olgą Gluchovskają.
- Remigijų Paulavičių,
- Sigitą Pedzevičienę,
- Dalių Pumputį.

TESTO KLAUSIMŲ ATSAKYMAI

1 skyrelis: 1d
3 skyrelis: 1a
4 skyrelis: 1c
5 skyrelis: 1d
6 skyrelis: 1b
7 skyrelis: 1c
8 skyrelis: 1b
9 skyrelis: 1a, 1b
10 skyrelis: 1d
11 skyrelis: 1d
12 skyrelis: 1c
13 skyrelis: 1d

Romanas Januškevičius, Olga Januškevičienė

ELEMENTARUSIS TIKIMYBIŲ IR STATISTIKOS KURSAS INFORMATIKAMS

2 dalis. Statistikos pradmenys

METODINĖ PRIEMONĖ

Maketavo Romanas Januškevičius

Už redagavimą atsakingi autoriai

Išleido Vilniaus pedagoginis universitetas, Studentų g. 39, LT-08106 Vilnius