



Statistinio raštingumo ugdymas

METODINĖ PRIEMONĖ 9-12 KLASIŲ MOKYTOJAMS

Sandra Aleksienė (*Garliavos Jonučių progimnazija*) ir Alfredas Račkauskas (*VU, MIF*)

2022

Ižanga

Statistika yra apie duomenis. Duomenys yra skaičiai, bet jie nėra „tik skaičiai“. Duomenys yra skaičiai su kontekstu. Pavyzdžiui, skaičius 2 neturi informacijos. Bet, jei išgirstame, kad kažkas loterijoje laimėjo 2 milijonus eurų, su pavydu sveikiname jį. O jei avarijoje žuvo 2 jaunuoliai – apima susirūpinimas. Kontekstas daro skaičius informatyvius. Bet išgauti informaciją iš jų – menas, reikalaujantis daug žinių ir specialiųjų gebėjimų.

Ši metodinė priemonė yra pagalba matematikos mokytojams, kurie rengiasi ugdyti moksleivių statistinį raštingumą pagal atnaujintą Matematikos mokymo programą. Ji rengta atsižvelgiant į programos turinio dalį „Duomenys ir interpretavimas“, skirtą 9, 10 ir 12 klasėms. Dauguma šioje metodinėje priemonėje pateiktų pavyzdžių paimta iš šių šaltinių:

1. David S. Moore and William I. Notz, *Statistics Concepts and Controversies*, 7 edition, H. Freeman and Company, New York, 2009.
2. *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)*. A Framework for Statistics and Data Science Education. Writing Committee: Anna Bargagliotti (co-chair), Christine Franklin (co-chair), Pip Arnold, Sheri Johnson, Leticia Perez, Denise A. Spangler, Rob Gould. *National Council of Teachers of Mathematics, 2020*.

Terminas „Statistika“ turi kelias prasmes.

Statistika - tai mokslas apie mokymąsi iš duomenų, informacijos rinkimą, sisteminimą, analizavimą ir interpretavimą. Ji vysto duomenų rinkimo ir jų analizavimo metodus.

Statistika – tai matematinių metodų ir įrankių rinkinys, leidžiantis atsakyti į svarbius klausimus apie duomenis. Ji skirstomas į dvi kategorijas:

- *Aprašomoji statistika* - siūlo metodus, kaip apibendrinti duomenis, paverčiant neapdorotus stebėjimus prasminga informacija, kurią lengva interpretuoti ir dalytis.
- *Išvadinė statistika* - siūlo metodus, kaip iš imties duomenų padaryti išvadas visai populiacijai.

Statistika - tai informacija apie veiklą ar procesą, išreikšta skaičiais, lentelėmis ar diagramomis.

Dažnai sąvokos „duomenys“ ir „statistika“ vartojamos pakaitomis, tačiau tarp jų yra esminis skirtumas. Duomenys yra atskiros faktinės informacijos dalys, įrašytos ir naudojamos analizei. Tai pirminė, neapdorota informacija, iš kurios kuriama statistika. Statistika yra duomenų analizės, jos aiškinimo ir pateikimo, rezultatai. Kitaip tariant, su duomenimis atlikti tam tikri skaičiavimai/veiksmai, kurie leidžia suprasti, ką tie duomenys reiškia. Statistika dažnai, nors ir nebūtinai, pateikiama lentelės, diagramos ar grafiko pavidalu. Vyriausybės agentūros (pvz., Lietuvos Statistikos departamentas) dažnai praneša apie įvairias statistikas. Tarp įvairių pavyzdžių yra

- *gyvybinė statistika* – tai skaitiniai duomenys apie gimstamumą, santuokas, skyrybas, užkrečiamas ligas, derlius, nelaimingus atsitikimus ir kt.;
- *verslo ir ekonomikos statistika* – tai skaitiniai duomenys apie užimtumą, gamybą, kainas, pardavimus, atleidimus ir kt.;
- *socialinė statistika* – tai skaitiniai duomenys apie būstą, nusikalstamumą, išsilavinimą ir kt.

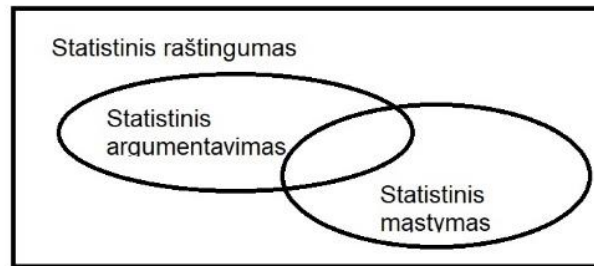
Dažnai šios rūšies statistika vadinama *statistiniais duomenimis*. Statistinis tyrimas analizuoja duomenis apie įvairius objektus (žmones, gyvūnus ar daiktus).

STATISTINIS RAŠTINGUMAS

Pagrindiniai statistinio turinio įtraukimo į matematikos mokymo turinį tikslai paprastai siejami su statistinio mąstymo ir statistinio raštingumo ugdymu.

Statistinis raštingumas - pagrindinės statistikos kalbos bei pagrindinių statistikos idėjų supratimas (pvz., žinojimas, ką reiškia statistiniai terminai ir simboliai bei mokėjimas skaityti statistinę informaciją).

Svarbu pabrėžti, kad vis dar nepasiektas sutarimas dėl vieno statistinio raštingumo apibrėžimo, nes pati koncepcija vis dar tobulinama. Daugelis tyrėjų statistinį raštingumą apibrėžia kaip statistikos kalbos „supratimą ir vartojimą“. Kiti tai sieja su aukštesnio lygio įgūdžių naudojimu (pvz., gebėjimu kritiškai vertinti informaciją ir interpretuoti duomenis pagal kontekstą). Yra ir toks statistinio raštingumo apibrėžimas: „žinoti ką reiškia statistiniai terminai ir simboliai, suprasti pagrindinę kalbą ir pagrindines statistikos idėjas, mokėti interpretuoti statistinius grafikus“. Bet kuriuo atveju, visuotinai sutariama, kad pagrindinės statistinio raštingumo dalys yra statistinis argumentavimas ir statistinis mąstymas.



Statistinis mąstymas – tai mąstymo tipas, kurį statistikai naudoja aiškindamiesi ar sprenddami statistines problemas.

Yra įvairių statistinio mąstymo apibūdinimų. Pavyzdžiui:

- supratimas apie duomenų neapibrėžtumą ir kintamumą bei jų įtaką sprendimų priėmimui; mokslinių metodų naudojimas sprendžiant iššūkius ir problemas;
- supratimas apie duomenų poreikį, duomenų parengimą, kintamumo visur buvimą, jo kiekybinį įvertinimą ir sprendimus.

Statistinis argumentavimas - būdas, kaip žmonės argumentuoja statistinėmis idėjomis ir įprasmina statistinę informaciją. Tai apima argumentus, pagrįstus duomenų rinkiniais, grafiniais vaizdais ir statistinėmis suvestinėmis.

MOTYVACIJA

Statistika tampa vis svarbesne šiuolaikinei visuomenei. Jos vaidmuo kasdieniam gyvenimui, svarba plėtojant kritinį mąstymą ir svarbus vaidmuo kitose disciplinose bei daugelyje profesijų lėmė išaugusį statistinio raštingumo ugdymo aktualumą visų išsilavinimo lygių mokiniams. Visuotinai sutariama, jog tam, kad statistinių samprotavimų esmė taptų mąstymo įpročių dalimi, statistinio raštingumo ugdymas turi prasidėti jau mokantis mokykloje, ir turėtų būti išlaikytas per ilgus metus trunkančią ugdymo ir praktinę patirtį. Dėl to daugelyje šalių statistinis turinys yra įtraukiamas į mokyklų bei kolegines ar universitetines programas arba pertvarkomas jau esamas statistikos ir tikimybių mokymas.

Statistika dažniausiai naudojama kitų disciplinų problemoms spręsti. Iš to gimsta nauji statistikos metodai ar naujos tyrimų kryptys. Pavyzdžiui, Gauss'as mažiausių kvadratų metodą surado analizuodamas duomenis apie asteroidus, Gosste'as t-testą atrado tyrinėdamas alaus daryklos duomenų imtis, Wilcoxon'as neparametrinius testus sukonstravo tyrinėdamas vaistų pramonės duomenis.

Dauguma statistinių argumentų sujungia idėjas apie duomenis ir atsitiktinumą. Tai leidžia daryti išvadas ir interpretuoti statistinius rezultatus. Statistinio argumentavimo pagrindas yra konceptualus svarbių idėjų, tokių kaip pasiskirstymas, centras, sklaida, sąryšis, netikrumas, atsitiktinumas ir atranka, supratimas. Statistiniai argumentai reikalingi daugelio profesijų žmonėms, įskaitant:

- psichologus, kurie tiria, kaip žmonės priima sprendimus apskritai ir sprendimus, susijusius su statistine informacija (dažnai pasitelkdami neteisingą nuojautą ar klaidingą supratimą),
- gydytojus ir kitus medicinos profesijos atstovus, kuriems reikia suprasti ir interpretuoti riziką, skirtingų medicininių rezultatų galimybes bei tyrimų rezultatus,
- žurnalistus ir mokslo propaguotojus, kurie domisi, kaip geriausiai paaiškinti ar kritikuoti žiniasklaidoje pateikiamą statistinę informaciją,
- politikos analitikus, kurie nori analizuoti bei aiškinti apklausų bei rinkimų rezultatus,
- statistikos mokytojus, norinčius ne tik ugdyti įgūdžius ir pateikti koncepcijas, bet ir mokyti argumentuoti kalbant apie duomenis bei tikimybes.

Statistinio raštingumo ugdymas

Nors statistikos specifika yra plačiai pripažinta, ji nėra savarankiška mokyklos mokymo programos tema. Statistika mokoma kaip matematikos dalis, nors ji tokia nėra (dar Fišeris savo knygoje pabrėžė, kad statistika yra taikomosios matematikos dalis). Ir tai kelia nemažus iššūkius matematikos mokytojams. Beje, tokiam matematikos ir statistikos sąryšiui susiformuoti prireikė nemažai laiko. Istorijoje žinomi bandymai statistiką mokyti per kitas disciplinas (biologiją, fiziką, ir kt.), tačiau tai nedavė reikiamo efekto ugdant statistiškai raštingą jaunimą.

Labai dažnai statistiniai metodai laikomi įrankiais, kurie taikomi ribotose situacijose. Pavyzdžiui, pasakyti „sudarykite histogramą, kad ištirtumėte šių duomenų elgseną“ arba „atlikite t testą, kad įvertintumėte, ar šios priemonės statistiškai skiriasi“. Toks požiūris leidžia mokiniams susidaryti labai siaurą statistinio taikymo vaizdą: statistikos įrankiai taikomi atskirai, kaip nurodyta. Mokymas turėtų paskatinti mokinius matyti *visą statistinio tyrimo procesą*. Jie turėtų suvokti, kad jų statistikos tyrimuose nepakanka skaitinio atsakymo, kol šis atsakymas nesusietas su kontekstu, su tyrimo klausimu. Mokiniai taip pat turi būti skatinami susieti statistinį tyrimą su ankstesne patirtimi ir kitu išoriniu kontekstu. Taigi vidurkio ar dispersijos pateikimas turėtų būti laikomas nepakankamu rezultatu. Gauti skaičiai turi būti paaiškinami atitinkamame kontekste.

Statistikos mokymo metodų yra įvairių. Koks bebūtų pasirinktas, jis turi:

1. Ugdyti statistinį mąstymą.
2. Telkti dėmesį į konceptualų statistikos supratimą.
3. Integruoti realius duomenis su kontekstu ir tikslais.
4. Skatinti aktyvų mokymąsi.
5. Analizuoti sąvokas ir duomenis naudojant šiuolaikines technologijas.
6. Mokyti statistikos per statistinį problemos sprendimą.

Paprastai statistikos mokytojai turėtų daug mažiau kliauti standartinėmis pamokomis, daug labiau – šiais alternatyviais studijų metodais:

1. Grupinis problemų sprendimas ir diskusija.
2. Laboratoriniai pratimai.
3. Demonstracijos, pagrįstos klasėje sugeneruotais duomenimis.
4. Pristatymai raštu ir žodžiu.
5. Projektai, grupiniai arba individualūs.

Projektinis metodas yra labai naudingas. Taikant jį bendraamžių grupėse arba individualiai, projektas gali suteikti moksleiviams galimybę aktyviai įsitraukti, bendradarbiauti, planuoti ir atlikti jiems rūpimas užduotis. Projektinis mokymas padeda bendrauti tarpusavyje ir su savo mokytoju bei motyvuoja juos tyrinėti statistines sąvokas ir procesus.

Kompiuterių panaudojimas yra labai svarbus aspektas. Šiandien mokiniai ir mokytojai turi galimybę naudotis virtualiu mokymusi, taip pat gausiais ir įvairiais internetiniais ištekliais.

Geriausios rezultatai ugdant statistinį raštingumą yra pasiekiami *atliekant tikrus statistinių duomenų tyrimus* nuo pirmųjų minčių iki tyrimo planavimo, renkant ir tyrinėjant duomenis, teikiant ataskaitas apie nustatytas duomenų ypatybes. Statistinių duomenų tyrimai taip pat sudaro idealias sąlygas aktyviam mokymuisi, praktinei patirčiai ir statistinių problemų sprendimui.

Konceptuali statistinio raštingumo ugdymo struktūra turi dvi dimensijas:

- Statistinis problemų sprendimo procesas.
- Pasiiekimų lygmenys.

Statistinio problemos sprendimo proceso tikslas - rinkti ir analizuoti duomenis, siekiant atsakyti į statistinio tyrimo klausimus.

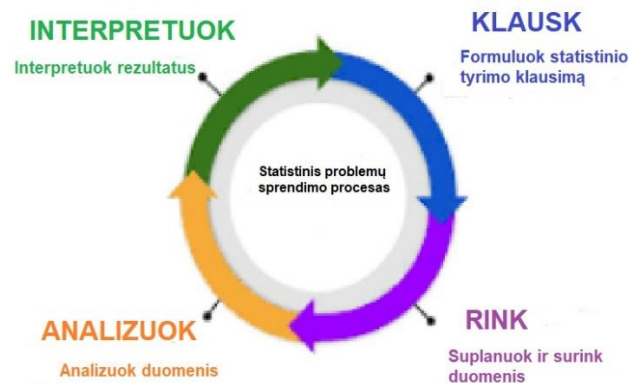
Statistinis problemų sprendimo procesas yra *statistinis tyrimas*, sudarytas iš keturių dalių:

(a) **klausimo iškėlimas/formulavimas** (išsiaiškinama esama problema, suformuluojamas vienas (ar daugiau) klausimų, į kuriuos galima atsakyti naudojant duomenis);

(b) **duomenų rinkimas** (parengiamas planas tinkamiems duomenims rinkti, pagal jį surenkami duomenys)

(c) **duomenų analizavimas** (pasirenkami tinkami grafiniai ar skaitiniai metodai, kuriais pasinaudojama analizuojant duomenis);

(d) **rezultatų interpretavimas** (interpretacijos siejamos su iškeltais klausimais).



KRAI: Klausk - Rink - Analizuok - Interpretuok

Toks statistinių problemų sprendimo procesas prieinamas visuose švietimo lygiuose, nuo pradinio ugdymo iki universitetinių studijų programų. Jis rekomenduojamas nuo tos klasės, kai pradedamas statistinio raštingumo ugdymas.

Mokyklinės matematikos programose, paprastai, statistinio raštingumo lygiai yra trys: A, B ir C.

(A) lygyje mokymasis yra labiau nukreiptas į mokytoją (labai sąlyginai, tikėtina, kad pasiekiamas 1-7(8) klasėse);

(B) lygyje pereinama prie į moksleivį orientuoto darbo (labai sąlyginai, tikėtina, kad pasiekiamas 8-10 klasėse);

(C) lygyje jis tampa labai orientuotas į moksleivį (taikomas labiau išplėstiniais kursams, arba 11-12 klasėms).

Šie trys lygiai pagrįsti statistikos mąstymo raida, o ne amžiumi. Taigi, vidurinėje mokykloje mokinys, kuris neturėjo ankstesnės patirties su statistika, turi pradėti nuo A lygio sąvokų ir veiklos prieš pereinant į B lygį. Jei mokinys nebuvo įgijęs A ir B lygių, nedera pereiti į C lygį. Taigi svarbu įsidėmėti, kad **lygiai pagrįsti patirtimi, o ne amžiumi**.

Visi keturi statistinio tyrimo proceso etapai KRAI naudojami visuose trijuose lygiuose, tačiau visuose lygiuose didėja naudojamų metodų supratimo ir sudėtingumo gylis. Visi trys lygiai trumpai pristatyti B priede.

9 ir I gimnazijos klasė¹

Duomenys ir interpretavimas

Nagrinėjamos taškinės (sklaidos) diagramos, vaizduojančios statistinį ryšį tarp dviejų kintamųjų (stebimų požymių) reikšmių. Mokomasi iš sklaidos diagramos įvertinti šio ryšio buvimą ar nebuvimą, aptariama, kokiais atvejais kalbama apie kintamųjų koreliacinį ryšį. Detaliau aptariama tiesinė koreliacija. Mokomasi užrašyti sklaidos diagramoje pavaizduotos tiesės lygtį $y = kx + b$, koeficiento k reikšmę, interpretuoti šia lygtimi aprašomą duomenų ryšį. Aptariama, kodėl negalime daryti išvados apie tiesinės priklausomybės egzistavimą populiacijoje, jei duomenys imtyje yra neatsitiktiniai ar jų yra per mažai.

Pavyzdžiai, kurie nagrinėjami 9 (ir I-oje gimnazijos) klasėje, mokiniams gali būti jau pažįstami ir gali remtis situacijomis, aptartomis iki 8 klasės. Šiame etape galima (ir reikia) naudoti sąvokas, grafikus ir kitas duomenų santraukas, kurios buvo nagrinėjamos iki 8 klasės. Skaitytojų patogumui, B priede surinktos visos sąvokos, kurios įvardintos programoje iki 8 klasės imtinai.

9 (ir I-oje gimnazijos) klasėje pristatomos naujos sąvokos ir nauji duomenų analizės įrankiai bei pagilinamos jau įgytos žinios. Šio etapo programai įgyvendinti reikės šių sąvokų bei analizės įrankių:

- *statistinis kintamasis ir jų tipai;*
- *sklaidos diagrama;*
- *koreliacijos koeficientas;*

¹ Logiškiau būtų sukeisti 9 ir 10 klases vietomis.

- *tiesinis sąryšis.*

Šiame etape pavyzdžiai ir naujas turinys yra kuriami atliekant statistinį duomenų tyrimą, kuris apima:

- statistinio klausimo formulavimą (apima taip pat kylančių problemų nustatymą ir duomenų rinkimo planavimą);
- duomenų rinkimą, tvarkymą ir tikrinimą;
- duomenų tyrinėjimą;
- rezultatų interpretavimą kontekste.

Vis dar reikalinga mokytojo pagalba. Kadangi viskas prasideda nuo statistinio klausimo, pravartu prisiminti, kad tinkamai suformuluoto statistinio klausimo kriterijai yra šie:

- Statistiniame klausime aiškiai pasakytas dominantis statistinis kintamasis (-ieji).
- Statistinis klausimas aiškiai nurodo populiaciją, kuriai jis skirtas.
- Statistinis klausimas aiškiai nurodo tikslą.
- Iš statistinio klausimo turėtų būti aišku, ar duomenys, padedantys atsakyti į klausimą gali būti renkami (vadinami pirminiais duomenimis), ar duomenys jau yra prieinami (vadinami antriniais duomenimis).
- Statistinis klausimas yra vertas tyrimo, jis įdomus ir/ar turi tikslą, padeda atsakyti į tyrimo klausimą.

Ką dar mokytojui pravartu prisiminti? Pagal aukščiau pateiktą matematikos programos 9 klasės skyrelį „Duomenys ir interpretavimas“, siūlome tokį turinį.

1. Statistinių kintamųjų tipai.
2. Sklaidos diagrama ir jos interpretavimas.
3. Koreliacijos koeficiento samprata.
4. Tiesinis dviejų kintamųjų sąryšis.

STATISTINIŲ KINTAMŲJŲ TIPAI

Kai renkame arba stebime duomenis, tai *ką ketiname* stebėti arba *ką stebime*, vadiname *statistiniu kintamuoju* arba tiesiog *kintamuoju*. Galime galvoti apie statistinį kintamąjį kaip objekto, kuris yra stebimas arba bus stebimas, aprašymą.

(Statistinis) kintamasis. Populiacijos, kartu ir imties elementus vienija tiriamasis požymis. Jį vadiname *statistiniu kintamuoju* arba tiesiog *kintamuoju*. Amžius, lytis, namų ūkio pajamos ir išlaidos, galvos apimtis, transporto priemonės tipas yra kintamųjų pavyzdžiai. Pavyzdžiui, „pajamos“ yra kintamasis, kurio reikšmės gali skirtis tiek skirtingiems namų ūkiams, tiek laikui bėgant.

Duomenys (duomenų rinkinys) - imties elementų tiriamo statistinio kintamojo išmatuotos reikšmės.

Yra įvairių būdų, kaip statistinius kintamuosius galima apibūdinti, atsižvelgiant į tai, kaip jie gali būti matuojami, tiriami ir pateikiami. O duomenų tipai tiesiog sutampa su statistinių kintamųjų tipais. Jų yra įvairių.

Kiekybinį (skaitinį) statistinį kintamąjį aprašo skaičiai atsakantys į klausimą „kiek?“.

Kiekybiniai kintamieji gali būti *tolydieji* arba *diskretieji*. Kiekybinis kintamasis vadinamas *tolydžiuoju*, jei jo galimų reikšmių skirtumas gali būti kiek norima mažas. Tolydaus statistinio kintamojo pavyzdžiai:

- laikas minutėmis pasiekti mokyklą,
- 10 metų berniukų dešinės pėdos ilgis centimetrais,
- oro temperatūra lauke,
- per savaitę iškritęs kritulių kiekis.

Tolydiesiems duomenims reikalingi matavimo vienetai, taigi stebėjimai turi būti registruojami norimais (pasirinktais) vienetais. Tačiau reikia turėti omeny, kad tolydžių kintamųjų stebėjimai niekada nėra tikslūs. Galima interpretuoti, kad stebima reikšmė reprezentuoja galimas reikšmes iš kažkokio skaičių intervalo. Pavyzdžiui, kai sakoma, kad oro temperatūra dieną buvo 15 laipsnių šilumos, galime suprasti, kad jos reikšmės buvo kažkuriame intervale, į kurį patenka 15, pavyzdžiui nuo 13 iki 17 laipsnių. Intervalo didumas yra svarbu, bet tai jau gilesnės statistinės analizės tema.

Kiekybinis kintamasis vadinamas *diskrečiuoju*, jei jo galimų reikšmių skirtumas yra ne mažesnis už tam tikrą minimalų pokytį. Tarp diskrečiųjų kintamųjų bene svarbiausi yra *sveikareikšmiai* arba *skaičiuojantieji*. Jie skaičiuoja daiktų arba žmonių skaičių tam tikru laiku ar tam tikroje vietoje, ar tam tikra proga ir pan. Kiekvienas sveikareikšmio kintamojo stebėjimas yra sveikasis skaičius - skaičiavimo reikšmė. Sveikareikšmio statistinio kintamojo pavyzdžiai:

- vaikų skaičius šeimoje;
- žmonių, atvykstančių į geležinkelio stotį 5 minučių intervalu, skaičius;

- gautų per dieną SMS žinučių skaičius;
- apsilankymų duotoje interneto svetainėje per savaitę skaičius;
- kuriame nors mieste per metus registruotų automobilių skaičius.

Kategoriniai (arba kokybiniai) kintamieji apibūdina duomenų vieneto „kokybę“ arba „charakteristiką“, pvz., „koks tipas“ arba „kuri kategorija“.

Kategoriniai kintamieji paprastai pateikiami ne skaitine reikšme. Jie negali būti sudedami, dauginami ar vidurkinami. Kategorinis statistinis kintamasis turi keletą skirtingų kategorijų. Taigi ir atitinkami stebėjimai suskirstomi į atitinkamas kategorijas. Kategorinio statistinio kintamojo pavyzdžiai:

- naminiai augintiniai,
- mėgstamiausia TV laida,
- mėgstamiausia atostogų veikla,
- mėgstamiausias prekybos centras,
- mėgstamiausia sporto šaka.

Kartais kategorijos yra natūralios, pvz., lytis arba pirmenybė tarp katės ir šuns, o kartais jas reikia pasirinkti ir kruopščiai aprašyti, pavyzdžiui, mėgstama atostogų veikla ar mėgstamas maistas.

SKLAIDOS DIAGRAMA IR JOS INTERPRETAVIMAS

Ką turime omenyje, kai sakome du statistiniai kintamieji yra „susieti“? Šio klausimo supratimas yra labai svarbus norint suprasti statistiką apskritai. Iš esmės sąsaja reiškia, kad vieno kintamojo duomenys atsiranda tam tikru būdu, kuris yra susijęs su kito kintamojo duomenimis. Jei šis apibūdinimas pernelyg techniškas, pagalvokite apie šiuos pavyzdžius:

- Jei atliksite mažmeninės prekybos tinklo rinkos tyrimą, tai pastebėsite, kad vandens ir ledų pardavimai kartu didėja arba mažėja. Didėjant vandens pardavimui, didėja ir ledų pardavimas.
- Žmonės su šviesiais plaukais dažniau turi mėlynas akis nei žmonės su tamsiais plaukais. Ar tikrai?

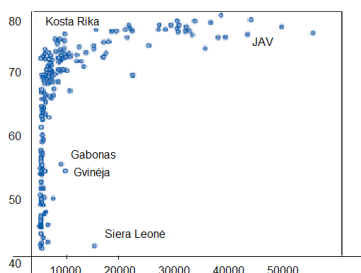
Geriausias būdas pastebėti sąsają tarp dviejų kiekybinių kintamųjų yra pasinaudoti sklaidos diagrama.

Skaidos diagrama rodo ryšį tarp dviejų kiekybinių kintamųjų, išmatuotų tiems patiems individams. Vieno kintamojo reikšmės rodomos horizontalioje ašyje, o kito kintamojo reikšmės – vertikaloje ašyje. Kiekvienas duomenų vienetas rodomas kaip taškas diagramoje, atitinkantis abi individo kintamųjų reikšmes.

Paprastai vienas iš kintamųjų interpretuojamas kaip *priklausomas kintamasis* arba *atsako kintamasis*. Kitas – kaip *aiškinantysis* arba *nepriklausomas kintamasis*. Apibūdindami skaidos diagramą, statistikai paprastai įvardija priklausomą kintamąjį (esantį y ašyje) ir nepriklausomą kintamąjį (esantį x ašyje).

1 Pavyzdys. Sveikata ir turtas

1 skaidos diagramoje yra Pasaulio banko duomenų skaidos diagrama. Individai yra visos pasaulio tautos, apie kurias yra prieinami duomenys. Aiškinantysis kintamasis yra šalies turtingumo matas: bendrasis vidaus produktas (BVP) tenkantis vienam gyventojui. BVP yra bendra šalyje pagamintų prekių ir suteiktų paslaugų vertė, konvertuota į šalies valiutą. Atsako kintamasis yra vidutinė gyvenimo trukmė metais.



1 skaidos diagrama. Gyvenimo trukmės ir šalies bendrojo vidaus produkto tenkančio vienam asmeniui taškinė diagrama. Duomenys paimti iš Pasaulio banko.

Tikėtina, kad turtingesnėse šalyse žmonės gyvena ilgiau. Iš esmės, skaidos diagrama tai ir rodo, tačiau sąsajos turi įdomią formą. Kai BVP didėja, vidutinė gyvenimo trukmė iš pradžių ilgėja labai greitai, bet vėliau greitis lėtėja arba augimas visai sustoja. Žmonės labai turtingose šalyse, pavyzdžiui, JAV, paprastai gyvena ne ilgiau nei žmonės skurdesnėse, bet ne itin skurdžiose šalyse. Palyginkite pvz., Kosta Rika su JAV. Trys Afrikos tautos yra išskirtys. Jų gyvenimo trukmė panaši į kaimynų, tačiau jų BVP yra didesnis. Pusiaujo Gvinėja ir Gabonas išgauna naftą, o Siera Leonė – deimantus. Gali būti, kad pajamos iš naudingųjų iškasenų eksporto daugiausia

atitenka grupei oligarchų ir tai padidina BVP vienam asmeniui, nedarant didelės įtakos nei eilinių piliečių pajamoms, nei gyvenimo trukmei. Tai reiškia, kad BVP vienam asmeniui yra vidurkinis rodiklis, ir mes žinome, kad vidutinės pajamos gali būti daug didesnės nei medianinės pajamos.

2 pavyzdys (mokytojo savarankiškam darbui). Intelektas ir smegenų apimtis

Šimtmečius žmonės intelektą siejo su smegenų dydžiu. Neseniai buvo atliktas tyrimas smegenų apimčiai išmatuoti panaudojant magnetinio rezonanso tomografiją. Šešių asmenų IQ ir smegenų dydis (10 000 pikselių vienetais) yra tokie:

Smegenų dydis	100	90	95	92	88	106
IQ	140	90	100	135	80	103

1 duomenys.

- Ar galime įvardinti, kuris iš kintamųjų yra aiškinantysis?
- Kuris yra atsako kintamasis?
- Sudarykite šių duomenų sklaidos diagramą. Ką pastebite?



Norėdami interpretuoti sklaidos diagramą, galime takyti įprastas duomenų analizės strategijas.

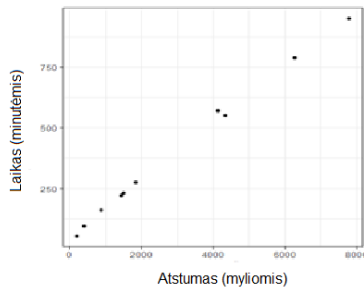
Pirmiausia duomenų diagramoje reikia paieškoti bendros struktūros (bendrumo bruožų) ir ryškių nukrypimų nuo jų. Bendrą sklaidos diagramos struktūrą galime apibūdinti pagal išvelgiamo sąryšio *kryptį, formą ir stiprumą*. Svarbi nukrypimų nuo išvelgto ryšio rūšis yra išskirtys - individualios reikšmės, kurios nepatenka į bendrą sąryšio modelį.

1 sklaidos diagramoje aiškiai matyti sąryšio kryptis: gyvenimo trukmė, apskritai imant, didėja, kai didėja BVP vienam gyventojui. Šiuo atveju sakome, kad stebime teigiamą sąsają (tarp gyvenimo trukmės ir BVP vienam gyventojui).

Du statistiniai kintamieji yra *teigiamai susieti*, kai vieno didesnės nei vidutinės reikšmės būna kartu su didesnėmis nei vidutinėmis kito reikšmėmis, o mažesnės nei vidutinės reikšmės taip pat būna kartu. Sklaidos diagramoje judant iš kairės į dešinę, stebime taškų kylimą aukštyn.

Kitaip tariant, teigiamas ryšys atsiranda tada, kai didelės vieno kintamojo reikšmės yra susietos su didelėmis kito kintamojo reikšmėmis, o mažos su mažomis.

2 sklaidos diagrama rodo dviejų kintamųjų teigiamą sąsają. Duomenys yra „United Airline“ tiesioginių skrydžių iš Čikagos į įvairius miestus į vakarus nuo Čikagos atstumą myliomis ir laikas minutėmis. Matome, kad didėjant atstumui nuo Čikagos, ilgėja ir skrydžio laikas.

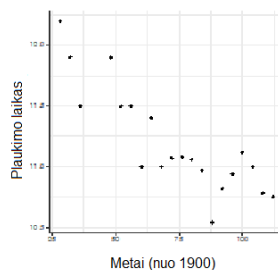


2 sklaidos diagrama. „United Airline“ tiesioginių skrydžių atstumas myliomis ir laikas minutėmis.

Du kintamieji yra *neigiamai susieti*, kai vieno kintamojo didesnės nei vidutinės reikšmės būna kartu su mažesnėmis nei vidutinės kito kintamojo reikšmėmis ir atvirkščiai. Sklaidos diagrama matoma kryptis žemyn iš kairės į dešinę.

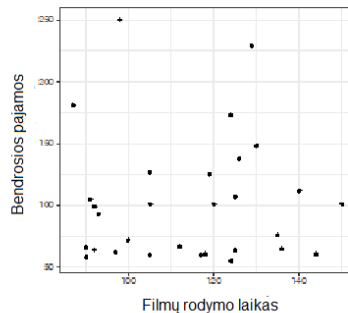
Neigiamas ryšys atsiranda, kai vieno kintamojo reikšmės mažėja tuomet, kai kito kintamojo reikšmės didėja.

3 sklaidos diagrama rodo sklaidos diagramą vaizduojančią neigiamą sąsają. Duomenys yra 100 metrų moterų plaukimo laisvu stiliumi Olimpinėse žaidynėse laikas pasiektas nuo 1912 metų. Metams bėgant, moterų plaukimo 100 m laisvuju stiliumi laikas gerėja.



3 sklaidos diagrama. 100 metrų plaukimo Olimpinėse žaidynėse laiką nuo 1912 metų

4 sklaidos diagrama rodo, kad du kintamieji nėra susiję. Duomenys yra pasirinktos filmų grupės rodymo laikas (minutės) ir bendrosios pajamos.



4 sklaidos diagrama. Filmų grupės rodymo laikas (minutėmis) ir bendrosios pajamos.

Ilgėjant filmo rodymo laikui, pinigų sumą, kurią uždirbo filmas sunku nuspėti – kartais ji buvo didelė, o kartais – gana pastovi.

Sąsajos tarp dviejų statistinių kintamųjų nėra, kai vienas kintamasis didėja ar mažėja, o kito kintamojo reakcijos į tai nėra.

KORELIACIJOS KOEFICIENTO SAMPRATA

Sklaidos diagrama rodo dviejų statistinių kintamųjų sąsajos kryptį, formą ir stiprumą. Kryptį jau aptarėme. Toliau aptarsime formą ir stiprumą.

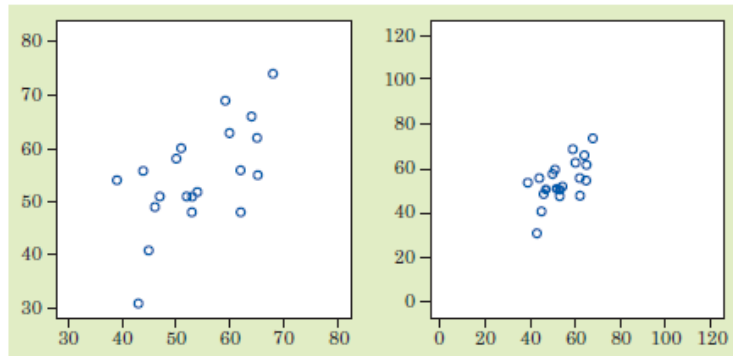
Paprasčiausia sąsajos (ryšio) tarp dviejų statistinių kintamųjų forma yra tiesinė.

Tiesinės sąsajos stebimos, kai sklaidos diagramoje taškai yra išsidėstę apie tiesią liniją.

Šis sąsajos tipas vadinamas tiesiniu ir yra labai svarbus. Mat tiesi linija yra paprastas modelis, kuris yra, be to, gana dažnas.

Tiesinė sąsaja yra stipri, jei sklaidos diagramoje taškai yra arti tiesės, ir silpni, jei jie yra plačiai išsibarstę ties linija.

Tačiau mūsų akys nėra geras sąsajos stiprumo vertintojas. Dvi sklaidos diagramos 5 sklaidos diagramoje vaizduoja tuos pačius duomenis, bet dešinioji diagrama nubrėžta didesniame lauke nei kairioji. Sprendžiant akimis, dešinės pusės diagrama rodo stipresnį tiesinį ryšį. Taigi mūsų akis gali apgauti pakeitus brėžinių skales arba tuščios vietos aplink taškų debesį sklaidos diagramoje kiekį. Todėl duomenų analizės strategiją verta papildyti, panaudojant kokį nors skaitinį įrankį. Koreliacijos koeficientas, paprastai žymimas r , kaip tik yra tai ko reikia.



5 sklaidos diagrama. Koreliacijos koeficientas apibūdina tiesinio ryšio tarp dviejų kiekybinių kintamųjų kryptį ir stiprumą.

Koreliacijos koeficiento skaičiavimas

Turime tokius $n=5$ stebėjimų duomenis apie du kintamuosius x ir y :

x	38	56	59	64	74
y	41	63	70	72	84

1 žingsnis. Raskime x ir y vidurkius ir standartinius nuokrypius. Priminsime, kad

$$\bar{x} = \frac{1}{n} \sum x, \quad s_x^2 = \frac{1}{n-1} \sum (x - \bar{x})^2.$$

Simbolis $\sum a$ reiškia, kad reikia sudėti visas a reikšmes, o n yra tokių reikšmių skaičius.

Pasidarykime vidurkio skaičiavimo lentelę:

x	y
38	41
56	63
59	70
64	72
74	84
$\Sigma x = 291.0$	$\Sigma y = 330.0$

$\bar{x} = \frac{291.0}{5} = 58.2$	$\bar{y} = \frac{330.0}{5} = 66.0$
------------------------------------	------------------------------------

Standartiniais nuokrypiais suskaičiuoti padės ši lentelė:

$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$
-20.2	408.04	-25.0	625.0
-2.2	4.84	-3.0	9.0
0.8	0.64	4.0	16.0
5.8	33.64	6.0	36.0
15.8	249.64	18.0	324.0
	$\Sigma (x - \bar{x})^2 = 696.8$		$\Sigma (y - \bar{y})^2 = 1010$
	$s_x = \sqrt{696.8} = 26.40$		$s_y = \sqrt{1010} = 10.49$

2 žingsnis. Kiekvieną duomenį standartizuokime: atimkime atitinkamą vidurkį ir padalinkime iš standartinio nuokrypio (iš čia ir pavadinimas – standartizavimas). Taigi x standartizuota reikšmė yra $(x - \bar{x})/s_x$, o y standartizuota reikšmė yra $(y - \bar{y})/s_y$. Šito žingsnio tikslas yra panaikinti kintamųjų matavimo vienetus. Pavyzdžiui, padalinę 10 kg iš 2 kg (10 kg / 2 kg = 5) gausime tiesiog 5 be matavimo vienetų. Tuos pačius 5 gautume ir 10 metrų padalinę iš 2 metrų. Taigi matavimo vienetai pranyksta. Pastebėkime, kad $x - \bar{x}$ ir s_x matavimo vienetai yra tokie pat.

x reikšmė	Standartizuota x reikšmė= $(x - \bar{x})/s_x$	y reikšmė	Standartizuota y reikšmė= $(y - \bar{y})/s_y$
38	$(38 - 58.2)/26.40 = -0.76$	41	$(41 - 66.0)/10.49 = -2.38$
56	$(56 - 58.2)/26.40 = -0.08$	63	$(63 - 66.0)/10.49 = -0.28$
59	$(59 - 58.2)/26.40 = 0.03$	70	$(70 - 66.0)/10.49 = 0.38$
64	$(64 - 58.2)/26.40 = 0.22$	72	$(72 - 66.0)/10.49 = 0.57$
74	$(74 - 58.2)/26.40 = 0.60$	84	$(84 - 66.0)/10.49 = 1.72$

3 žingsnis. Koreliacija yra standartizuotų duomenų sandaugų vidurkis. Tik vidurkį randame sandaugų sumą padalindami iš $n-1$, vienetu mažesnio skaičiaus nei duomenų skaičius:

$$\begin{aligned}
r &= \frac{1}{4}[(-0.76)(-2.38) + (-0,08)(-0.28) + (0.03)(0.38) + (0.22)(0.57) \\
&\quad + (0.60)(1.72)] = \frac{1}{4}(1.8088 + 0.0224 + 0.0114 + 0.1254 + 1.032) \\
&= \frac{3}{4} = 0.75
\end{aligned}$$

Rezultatas: $r = 0.75$.

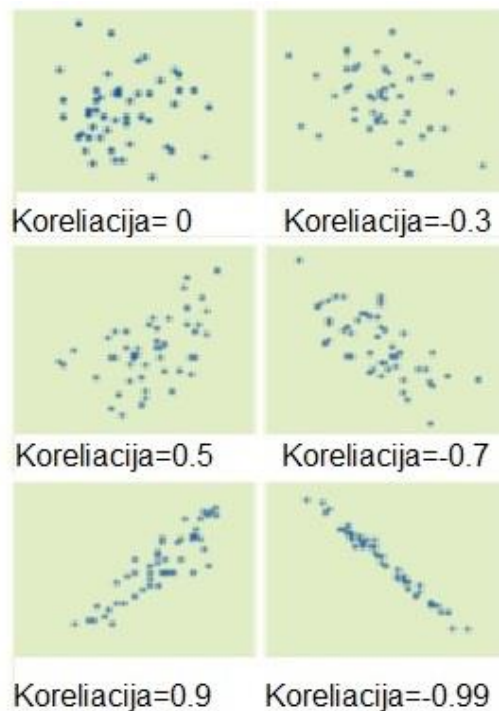
Atliktų skaičiavimų algebrinis trumpinys yra

$$r = \frac{1}{n-1} \sum \frac{x - \bar{x}}{s_x} \frac{y - \bar{y}}{s_y}$$

Koreliacija apibūdina tiesinio ryšio tarp dviejų kiekybinių kintamųjų kryptį ir stiprumą. Svarbiau nei r apskaičiavimas (skaičiavimo užduotis) yra suprasti, kaip koreliacija matuoja dviejų statistinių kintamųjų tiesinį ryšį ir jo stiprumą. Paminėtini keli faktai.

- Teigiamas r rodo teigiamą sąsają tarp kintamųjų, o neigiamas r rodo neigiamą sąsają.
- Koreliacija r visada yra tarp -1 ir 1 . Reikšmės artimos -1 arba 1 rodo, kad sklaidos diagramos taškai yra arti tiesės.
- Kraštutinės reikšmės $r = -1$ ir $r = 1$ atsiranda tik tada, kai taškai sklaidos diagramoje yra tiksliai išilgai tiesios linijos.
- r reikšmės šalia 0 rodo labai silpną tiesinį ryšį. Ryšio stiprumas didėja, kai r tolsta nuo 0 link -1 arba 1 .

Pavyzdyje gauta koreliacijos koeficiento reikšmė $r = 0.994$ rodo stiprą teigiamą tiesinį sąryšį tarp statistinių kintamųjų x ir y . Žemiau pateiktos sklaidos diagramos iliustruoja, kaip r matuoja ir tiesinio ryšio kryptį, ir jo stiprumą. Atkreipkite dėmesį, kad kiekvienoje sklaidos diagramoje r ženklas atitinka nuolydžio kryptį, o r artėjant prie -1 arba 1 , sąsajos modelis artėja prie tiesinio.



6 sklaidos diagrama. Įvairių stiprumų ir krypčių koreliacijos.

- Kadangi r naudoja standartizuotus duomenis, koreliacija nesikeičia, kai keičiame statistinių kintamųjų, vieno kurio ar abiejų matavimo vienetus.
- Koreliacija nepaiso skirtumo tarp aiškinamųjų ir atsakomųjų kintamųjų. Sukeitus vietomis x 'o ir y 'o reikšmes, koreliacijos koeficientas liks toks pat.

Dar kartą pabrėšime, kad koreliacija matuoja *tik tiesinės sąsajos* tarp dviejų kintamųjų stiprumą ir kryptį. Koreliacijos koeficientas nieko nesako apie galimą netiesinį ryšį. Be to, kaip ir vidurkiui bei standartiniam nuokrypiui, koreliacijai didelę įtaką daro išskirtys.

[2 pavyzdžio tęsinys \(mokytojo savarankiškam darbui\). Intelektas ir smegenų apimtis.](#)

Kaip „iš akies“ įvertintumėte koreliacijos koeficientą? O pagal formulę?

www.istics.net/Correlations - šioje svetainėje galima pasipraktikuoti koreliacijos koeficientą vertinti „iš akies“.

Egzistuoja daugybė kintamųjų ryšių ir daugybė būdų juos išmatuoti. Nors koreliacija yra labai paplitusi, atsiminkite jos apribojimus. Koreliacija prasminga tik kiekybiniam kintamiesiems – galime kalbėti apie sąsajas tarp rinkėjų lyties ir

politinės partijos, kuriai jie teikia pirmenybę, bet ne apie koreliaciją tarp šių kintamųjų.

Taip pat atmintkite, kad koreliacija nėra išsamus dviejų kintamųjų duomenų aprašymas, net jei tarp kintamųjų yra tiesinis ryšys. Mat koreliacijos koeficientas naudoja tik duomenų vidurkį ir standartinį nuokrypį. O šie parametrai neatspindi viso duomenų savybių spektro. Be to, koreliacija nieko nepasako apie priežastinį kintamųjų ryšį.

Priežastinis ryšys apibrėžiamas kaip ryšys tarp dviejų kintamųjų, kai vienas kintamasis veikia kitą. Priežastinis ryšys aptinkamas, kai vieno kintamojo reikšmės padidėja arba sumažėja dėl kito kintamojo pasikeitusios reikšmės.

Pavyzdys: jei turite papildomų pamokų per savaitę, jūsų pajamos greičiausiai padidės. Tai reiškia, kad vienas įvykis (papildomos pamokos) turi įtakos kitam įvykiui (pajamoms). Tai yra priežastinis arba priežasties-pasekmės ryšys.

Kita vertus ne visada sąsają tarp dviejų kintamųjų sukelia priežastinis ryšys. Sąsajos gali atsirasti tarp kintamųjų, turinčių priežastinį ryšį, arba tų, kurie neturi priežastinio ryšio. Apie priežastinį ryšį ir tiesinę regresiją kalbama 11 klasės programoje (Dėmesio! Joje yra loginių klaidų).

Tiesinę sąsają tarp statistinių kintamųjų x ir y galime modeliuoti tiese, kurios matematinė išraiška yra

$$y = kx + b$$

Skaičius k vadinamas krypties koeficientu, o b yra laisvasis narys. Yra daug būdų kaip parinkti k ir b . Pavyzdžiui, galima iš turimų duomenų parinkti dvi poras (x, y) ir užrašyti tiesės, einančios per tuos du taškus lygtį. Turėdami k ir b reikšmes, tiesinį modelį $y = kx + b$ galime panaudoti kintamojo y reikšmei numatyti, kai yra žinoma kintamojo x reikšmė.

Tikslas tiesės lygtyje $y = kx + b$ rasti tokius k ir b , kad gauta tiesė geriausiai aprašytų tiesinį statistinių kintamųjų x ir y sąryšį. Mažiausių kvadratų metodo esmė labia paprasta: k ir b , parenkame taip, kad dydis $\sum(y - kx - b)^2$ įgytų mažiausią galimą reikšmę. Čia sumuojama pagal visas turimas statistinių kintamųjų x ir y reikšmes. Galima išvesti šias formules:

$$k = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}, \quad b = \frac{1}{n}(\sum y - k \sum x).$$

Taigi turėdami n kintamųjų x ir y stebėjimų atliekame šiuos veiksmus:

1 veiksmas : kiekvienai porai (x, y) apskaičiuojame x^2 ir xy ;

2 veiksmas : susumuojame visus x, y, x^2 ir xy , gaudami $\sum x, \sum y, \sum x^2, \sum xy$ (kaip ir anksčiau, ženklas $\sum a$ reiškia „sumuoti visas a reikšmes“);

3 veiksmas : apskaičiuojame nuolydį k :

$$k = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} ;$$

4 veiksmas : apskaičiuojame laisvąjį narį b : $b = \frac{1}{n} (\sum y - k \sum x)$;

5 žingsnis : užrašome gautą tiesės lygtį $y = kx + b$.

Pavyzdys. Jonas dirbo ledų pardavėju Palangoje ir tyrė, kaip ledų pardavimus veikia saulėtos valandos. Stebėjęs nuo pirmadienio iki penktadienio surinko tokius duomenis:

Saulėtų valandų skaičius - x	Parduotų ledų skaičius - y
2	4
3	15
5	9
7	10
9	25

1 veiksmas. kiekvienai porai (x, y) apskaičiuojame x^2 ir xy ;

x	y	x^2	xy
2	4	4	8
3	15	9	45
5	9	25	45
7	10	49	70
9	25	81	225

2 veiksmas : randame sumas $\sum x, \sum y, \sum x^2, \sum xy$

x	y	x^2	xy
-----	-----	-------	------

2	4	4	8
3	5	9	45
5	9	25	45
7	10	49	70
9	25	81	225
$\sum x$ = 26	$\sum y = 53$	$\sum x^2 = 168$	$\sum xy$ = 393

3 veiksmas : apskaičiuojame nuolydį k :

$$k = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{5 \times 393 - 26 \times 53}{5 \times 168 - 26^2} = \frac{387}{164} = 2.36$$

4 veiksmas : apskaičiuojame laisvąjį narį b :

$$b = \frac{1}{n} (\sum y - k \sum x) = \frac{1}{5} (53 - 2.36 \times 26) = -1.67$$

5 veiksmas : Užrašome gautą tiesės lygtį $y = kx + a$:

$$y = 2.36x - 1.67.$$

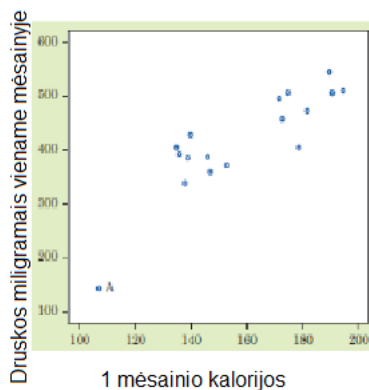
Jonas išgirsta orų prognozę, kuri sako: „rytoj tikimės 8 saulėtų valandų“, todėl jis naudoja aukščiau pateiktą lygtį, kad įvertintų, kiek ledų jis gali tikėtis parduoti. Jonas apskaičiuoja: $y = 2.36 \times 8 - 1.67 = 17,21$ ledų.

APIBENDRINIMAS

Dauguma statistinių tyrimų tiria ryšį tarp dviejų ar daugiau statistinių kintamųjų. Taškinė diagrama yra dviejų kiekybinių kintamųjų sąsajos grafikas. Jei turite aiškinantį ir atsako kintamuosius, atidėkite aiškinantį kintamąjį į sklaidos diagramos x (horizontaliąją) ašį. Nagrinėdami sklaidos diagramą, ieškokite sąryšio krypties, formos ir stiprumo bei galimų nuokrypių. Jei yra aiški kryptis, ar ji teigiama (taškinė diagrama svyruoja aukštyn iš kairės į dešinę) arba neigiamas (grafikas pasviręs žemyn)? Ar forma tiesi ar išlenkta? Ar yra stebėjimų grupių? Ar santykinai stiprūs (taškai sukonzentruoti) ar silpnas (taškai plačiai išsibarstę)? Koreliacija r matuoja tiesinio ryšio tarp dviejų kiekybinių kintamųjų kryptį ir stiprumą. Koreliacija yra skaičius nuo -1 iki 1 . r ženklas rodo, ar sąsaja yra teigiama, ar neigiamas. r reikšmė artėja prie -1 arba 1 , kai sklaidos diagramoje taškai glaudžiau susirenka tiesioje linijoje. Kraštutinės reikšmės -1 ir 1 atsiranda tik tada, kai sklaidos diagrama rodo visiškai tiesią liniją.

KELETAS PRATIMŲ

1 pratimas. *Kalorijos ir druska dešrainiuose.* 8 paveiksle parodytas dviejų statistinių kintamųjų sklaidos diagrama. x ašyje yra vieno mėsinio kalorijų kiekis, y ašyje - natrio kiekis viename mėsainyje. Viso yra 17 stebėjimo rezultatų (duomenų). Apibūdinkite bendrą šių duomenų modelį. Ar taškas pažymėtas A yra išskirtis?



8 sklaidos diagrama

2 pratimas. *Širdies ritmas.* Mokytojas Marius reguliariai nuplaukia 2000 metrų, bandydamas gerinti savo sportinę formą. Štai jo laikai (minutėmis) ir pulso dažnis (tvinksniai per minutę) po 23 plaukimo seansų:

Laikas	34.12	35.72	34.72	34.05	34.13	35.72	36.17	35.57
Pulsas	152	124	140	152	146	128	136	144
Laikas	35.37	35.57	35.43	36.05	34.85	34.70	34.75	33.93
Pulsas	148	144	136	124	148	144	140	156
Laikas	34.60	34.00	34.35	35.62	35.68	35.28	35.97	
Pulsas	136	148	148	132	124	132	139	

- Padarykite sklaidos diagramą. (Kas yra aiškinamasis kintamasis?)
- Ar sąsaja tarp šių kintamųjų yra teigiama ar neigiama? Paašškinkite, kodėl tikite, kad sąsaja yra tokios krypties.
- Apibūdinkite sąsajos formą ir stiprumą.
- Skaičiuotuvu suraskite koreliacijos koeficientą r . Žvelgdami į sklaidos diagramą, paašškinkite, kodėl ši r reikšmė yra pagrįsta.
- Tarkime, kad laikas buvo įrašytas sekundėmis. Pavyzdžiui, laikas 34,12 minutės būtų 2047 sekundės. Kaip pasikeistų r reikšmė?

3 pratimas. *Atspėkite koreliaciją.* Ar tikėtumėte reikšmingos neigiamos koreliacijos, reikšmingos teigiamos koreliacijos ar nedidelės koreliacijos su kiekviena iš šių kintamųjų porų?

- a) Interneto tiekėjo paslaugos kaina ir paslaugos atsisiuntimo greitis,
- b) naujų automobilių arklio galios ir jų degalų sąnaudos,
- c) deimanto dydis ir deimanto kaina,
- d) kolegijų moterų ūgis ir mokyklos matematikos valstybinio egzamino balai.



Darbai su mokiniais

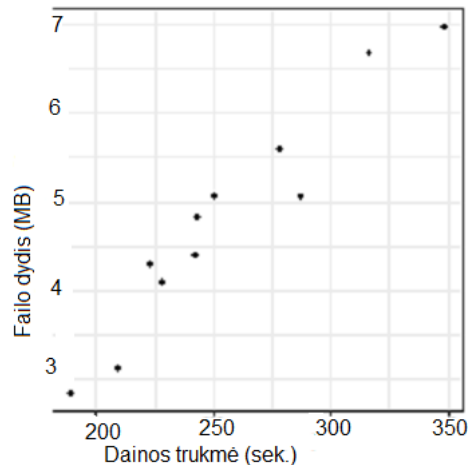
Projektus galima siūlyti mokinių grupelėms po tris ar daugiau mokinių vienoje. Labai svarbu, kad mokiniai suprastu visų keturių statistinio problemų sprendimo etapų svarbą.

1 projektas. *Dainos trukmė ir MP3 failo dydis.*

Nereikėtų nukrypti nuo statistinio duomenų tyrimo eiliškumo KRAI:

1. *Statistinio klausimo formulavimas.* Mokiniais reikia pagelbėti. Pavyzdžiui, paprašykite savo mokinių atsakyti į šiuos klausimus.
 - (a) Ar turite MP3 grotuvą?
 - (b) Ar daug dainų saugoma jūsų mobiliajame telefone?
 - (c) Kiek atminties naudoja jūsų dainų įrenginys?
 - (d) Ar manote, kad yra ryšys tarp dainos trukmės minutėmis ir MP3 failo dydžio? Šis klausimas tinka statistiniam tyrimui. (Galimi ir kitokie klausimai, pavyzdžiui: kiek tikėtina dainų galite įsirašyti įrenginyje, turinčiame 1 GB atminties? Tik šiam klausimui tinkamos statistinės priemonės nukeltos į 10 klasę!)
2. *Duomenų rinkimas, tvarkymas ir tikrinimas.* Suplanuokite, kaip surinksite duomenis. Jei mokytojas turi surinktų ar iš interneto ištrauktų duomenų, galima pradėti darbą su jais. Kiekvienas mokinys gali duomenis susirinkti iš savo mobiliojo telefono ar kompiuterio.
3. *Duomenų tyrinėjimas.* Verta paaiškinti kas yra sklaidos diagrama, kam ir kaip ji naudojama.

- (a) Paprašykite mokinių pateikti duomenis apie kintamuosius (failo dydis ir dainos trukmė). Padėkite mokiniams perkelti duomenis į sklaidos diagramą.



9 sklaidos diagrama. Sklaidos diagrama rodo failų dydžius (MB) skirtingo ilgio dainoms (sekundėmis).

Paašškinkite, kad šioje sklaidos diagramoje dainos trukmė yra nepriklausomas kintamasis arba aiškinantysis kintamasis, o failo dydis – priklausomas kintamasis arba atsako kintamasis. Ko gero teks priminti ir sąvoką „statistinis kintamasis“ bei pakabėti apie jų tipus.

- (b) Paklauskite, kokias mokiniai įžiūri tendencijas šioje sklaidos diagramoje? **Galimas atsakymas:** *Ilgesnė daina (sekundėmis) linkusi užimti daugiau atminties (MB).*
- (c) Kas atsitinka su failo dydžiu, kai daina ilgėja? **Galimas atsakymas:** *Failo apimtis didėja.*

Paašškinkite mokiniams, kad sąsajos tarp statistinių kintamųjų būna teigiamos ir neigiamos, jei jos apskritai yra ir patarkite kaip jas galima atpažinti.

- (a) Kaip apibūdinsite sąsają tarp dainos ilgio ir failo apimties? **Galimas atsakymas:** *sąsaja yra teigiama.*

Paašškinkite, kad duomenų analizavimo vienas iš tikslų yra rasti matematinį modelį, apibūdinantį ryšį tarp dviejų kiekybinių kintamųjų. Galima padiskutuoti apie tiesinius ir netiesinius modelius.

- (b) Koks matematinis modelis tiktų ryšiui tarp dainos failo apimties ir dainos trukmės apibūdinimui. **Galimas atsakymas:** *tiesinis modelis turi prasmę, nes sklaidos diagramoje taškai yra glaudžiai išsidėstę apie tiesią liniją.*

Paaškindite koreliacijos koeficiento prasmę, savybes ir kaip jį suskaičiuoti.

Rezultatų interpretavimas. Grįžkite prie statistinio klausimo. Ar gavote atsakymą? Kokį atsakymą gavote? Paprašykite atliktą tyrimą pristatyti žodžiu. Patarkite, kaip tai galima padaryti.

Atlikę šią užduotį, moksleiviai sužinojo visas sąvokas ir instrumentus, kurie numatyti programoje. T.y.,

- *statistinis kintamasis ir jų tipai;*
- *sklaidos diagrama;*
- *koreliacijos koeficientas;*
- *tiesinis sąryšis.*

Pasiūlykite mokiniams išspręsti dar kelis uždavinius įgytų žinių užtvirtinimui. Beje, tam galima panaudoti ir visus pavyzdžius, kurie skirti mokytojams.

2 projektas. Ryšys tarp žmogaus ūgio ir rankų ilgio.

Galimi statistinio tyrimo klausimai:

- Ar žmogaus ūgis ir rankų ilgis susiję?
- Kiek stiprus yra ryšys tarp rankų ilgio ir ūgio?

Pastaba. Rankų ilgį galima keisti kitu parametru, pavyzdžiui, batų dydžiu, galvos apimtimi ir pan.

3 projektas. Reikia ištirti ar matematika yra raktas į sėkmę?

Tai padėtų padaryti gauti atsakymai į šiuos klausimus.

- (a) Ar yra ryšys tarp matematikos ir kitų dalykų pasiekimų? Galima analizuoti vieną klasę, kelias klases arba visą mokyklą. Sąsajų ieškoti galima tarp matematikos pasiekimų ir kurio nors kito dalyko (užsienio kalbos, fizikos, chemijos ir kt.) arba kitų dalykų pasiekimų vidurkio.
- (b) Kokie kintamieji galėtų paaiškinti ryšį tarp matematikos pasiekimų vidurinėje mokykloje ir sėkmės stojant į universitetus/kolegijas?

Problema gali būti duomenų surinkimas. Tam gali prireikti pagalbos iš mokyklos administracijos.

UŽDUOTYS

1 užduotis. Kiekviename iš šių teiginių yra klaida. Kiekvienu atveju paaiškinkite kas negerai.

- (a) „Yra didelis ryšys tarp universiteto studentų pagrindinio dalyko ir jo ar jos pradinio atlyginimo baigus studijas.“
- (b) „Buvo nustatyta didelė koreliacija ($r = 1,09$) tarp statistikos kurso pirmojo ir antrojo kontrolinių darbų balų.
- (c) „Nustatyta, kad koreliacija tarp mokymosi metų skaičiaus ir pajamų yra $r = 0,53$ metų.“

Galime pasiūlyti pažaisti, bandant atspėti koreliacijos koeficiento reikšmę. <http://guessthecorrelation.com> - žaidimas, kuriame mokiniai spėja koreliaciją ir gali varžytis tiek dėl aukštesnių balų, tiek tarpusavyje.

2 užduotis. Ar tikėtumėtės reikšmingos neigiamos koreliacijos, reikšmingos teigiamos koreliacijos ar nedidelės koreliacijos kiekvienai iš šių kintamųjų porų?

- a) Interneto tiekėjo paslaugos kaina ir paslaugos atsiuntimo greitis.
- b) Naujų automobilių arklio galios ir jų 10 litrų degalų rida kilometrais .
- c) Deimanto dydis ir deimanto kaina.
- d) Mokyklos merginų ūgiai ir baigiamųjų egzaminų rezultatai.

3 užduotis. Šioje lentelėje pateikiami duomenys apie Kolos ir mėšainių kainas 16 beisbolo stadionų:

Stadiono Nr.	Mėšainio kaina	Kolos kaina	Stadiono Nr.	Mėšainio kaina	Kolos kaina	Stadiono Nr.	Mėšainio kaina	Kolos kaina
1	2.75	2.29	8	3.75	3.00	15	2.25	2.25
2	4.00	2.86	9	2.50	3.00	16	3.75	2.86
3	3.75	2.80	10	3.00	2.67	17	3.25	3.00
4	2.50	2.67	11	4.50	2.55	18	2.50	2.29
5	3.00	3.14	12	3.00	3.00	19	1.60	2.00
6	3.75	2.50	13	3.25	2.40	20	3.00	2.20
7	1.47	2.20	14	2.25	2.25	21	2.75	2.57

Išanalizuokite sąsajas tarp mėšainio ir kolos kainų.

4 užduotis. *Bebrai ir vabalai.*

Ekologai mūsų aplinkoje kartais randa gana keistų santykių. Vienas jų tyrimas rodo, kad bebrai naudingi vabalams. Tyrėjai parinko 23 apskritus sklypus, kurių

kiekvienas buvo 4 metrų skersmens, tose vietose, kur bebrai kirto medvilnės medžius.



Kiekviename sklype jie skaičiavo kelmus iš bebrų nukirstų medžių ir vabalų lervų sankaupas.

Kelmai	2	2	1	3	3	4	3	1	2	5	1	3
Lervų sankaupos	10	30	12	24	36	40	43	11	27	56	18	40
Kelmai	2	1	2	2	1	1	4	1	2	1	4	
Lervų sankaupos	25	8	21	14	16	6	54	9	13	14	50	

- (a) Padarykite sklaidos diagramą, kuri parodytų, kiek bebrų padarytų kelmų skaičius turi įtakos vabalų lervų sankaupų skaičiui. Ką rodo diagrama? (Ekologai mano, kad nauji kelmų daigai yra švelnesni ir vabalams labiau patinka.)
- (b) Regresijos linija yra
$$\text{lervų sankaupos} = -1,286 + (11,894 \times \text{kelmai})$$
 Nubrėžkite šią liniją savo sklaidos diagramoje. (Norėdami nubrėžti liniją, naudokite lygtį $y = -1,286 + 11,894x$, ir raskite y imdami $x = 1$ ir $x = 5$. Per surastus du (x, y) taškus nubrėžkite tiesią liniją.)
- (c) Koreliacija tarp šių kintamųjų yra $r = 0,916$. Kiek procentų stebimo vabalų lervų sankaupų skaičiaus kitimo galima paaiškinti tiesinio modelio pagalba?
- (d) Ar, remdamasis savo pastebėjimais (a), (b) ir (c) punktuose, manote, kad kelmų skaičiavimas yra patikimas būdas numatyti vabalų lervų sankaupas? Atsakymą pagrįskite.

5 uždutis. Sudarykite šių duomenų sklaidos diagramą:

x	1	2	3	4	9	10
y	10	3	3	5	9	11

Naudodami skaičiuotuvą parodykite, kad koreliacija yra maždaug 0,5. Kokia duomenų savybė lėmė tokią mažą koreliaciją, nepaisant stiprios tiesinės sąsajos tarp x ir y daugumoje stebėjimų?

10 ir II gimnazijos klasė²

Duomenys ir interpretavimas.

Paaiškinama, kaip imties iš populiacijos sudarymas susijęs su pagrįstų išvadų darymu, ką vadiname duomenų rinkinių kintamumu, duomenų pasiskirstymu, kaip galima apibūdinti ir kiekybiškai interpretuoti duomenų rinkinius. Aptiriamos sąvokos: dispersija, standartinis nuokrypis, skirstinys, normalusis skirstinys, simetriškasis skirstinys, asimetriškasis skirstinys. Nagrinėjant realaus gyvenimo konteksto pavyzdžius, diskutuojama apie duomenų rinkimą ir analizavimą. Svarstoma, kokias išvadas apie duomenis leidžia daryti jų pasiskirstymą aproksimuojančios kreivės forma ar apskaičiuotos duomenų centro (pavyzdžiui, vidurkio) ir sklaidos (pavyzdžiui, standartinio nuokrypio, kvartilų) charakteristikos. Analizuojamas statistinis patikimumas.

Pavyzdžiai, kurie nagrinėjami 10 (ir II-oje gimnazijos) klasėje, mokiniams gali būti jau pažįstami ir gali remtis situacijomis, aptartomis iki 9 klasės. Šiame etape galima (ir reikia) naudoti sąvokas, grafikus ir kitas duomenų santraukas, kurios buvo nagrinėjamos iki 9 klasės.

10 (ir II-oje gimnazijos) klasėje pristatomos naujos sąvokos bei nauji duomenų analizės įrankiai. Šio etapo programai įgyvendinti reikės šių sąvokų bei analizės įrankių:

- *Populiacija ir imtis*

² Šį turinį geriau išdėstyti prieš dviejų statistinių kintamųjų analizę, pvz., 9-oje klasėje. Taigi 9 klasės turinį reikėtų sukeisti vietomis su 10 klase.

- *Duomenų pasiskirstymas*
- *Duomenų pasiskirstymo parametrai*
- *Statistinis patikimumas*

Ką mokytojui pravartu prisiminti? Siūlome tokį turinį.

1. Imtys: geros ir blogos.
2. Duomenų pasiskirstymas: histograma ir medžio diagrama.
3. Duomenų pasiskirstymo parametrai: vidurkis, dispersija, standartinis nuokrypis.
4. Statistinis patikimumas.

IMTYS: GEROS IR BLOGOS

Populiacija - statistinio tyrimo objektų grupė, turinti bent vieną bendrą požymį (parametrą), apie kurį atliekamas statistinis tyrimas. Pvz.: populiacija – Lietuvos gimnazistai; požymis – akių spalva; matematikos žinios, statistinio raštingumo lygis ir pan. Populiacijas gali sudaryti žmonės, gyvūnai, įmonės, pastatai, transporto priemonės, namų ūkiai, renginiai ir pan.

Imtis - populiacijos dalis, naudojama statistiniam tyrimui. Pavyzdžiui, 20 atsitiktinai atrinktų gimnazistų; 1000 apklausoje dalyvaujančių respondentų ir pan.

Taigi parametras (šiuo etape) yra skaičius³, apibūdinantis populiaciją. Parametras yra fiksuotas skaičius, bet tikrosios jo reikšmės nežinome. Statistika yra skaičius, apibūdinantis imtį. Statistikos reikšmė yra žinoma, kai turime imtį, tačiau ji gali keistis nuo vienos imties iki kitos. Vertindami populiacijos parametro reikšmę, dažnai naudojame imties statistiką. Norėdami apibūdinti, kiek patikima yra imtis, klausiamo: „Kas nutiktų, jei paimtume daug imčių iš tos pačios populiacijos?“ Jei beveik visų imčių rezultatai būtų artimi, galėtume pasitikėti turima imtimi, nors, negalėtume būti visiškai tikri, kad rezultatas yra arti teisingo. Tai yra tarsi šaudymas į taikinį, nežinant į kurią vietą reikia pataikyti.

Imtis turi būti pakankamai didelė, kad būtų galima patikimai reprezentuoti visą populiaciją, tačiau ne visada pavyksta tokią sudaryti. Yra įvairių imties sudarymo būdų. Verta juos prisiminti.

³ Sudėtingesni parametrai: skaičių rinkinys, pavyzdžiui, kraujo tyrimo duomenys; kreivės, pavyzdžiui, širdies kardiograma.

Surašymas yra kiekvieno populiacijos vieneto, visų ar visko, tyrimas.

Imties atranka gali būti atsitiktinė arba neatsitiktinė. Atitinkamai imtis vadinama atsitiktine arba neatsitiktine.

Atsitiktinė imtis (atranka) yra tokia, kai kiekvienas populiacijos vienetas turi galimybę būti atrinktas ir šią galimybę galima tiksliai aprašyti.

Atsitiktinė atranka apima paprastąją, sisteminę, sluoksninę, lizdinę.

- *Paprastoji atsitiktinė imtis (atranka)* yra tokia, kai visi imties nariai parenkami atsitiktinai ir turi vienodą galimybę patekti į imtį. Loterija yra paprastos atsitiktinės atrankos pavyzdys, kai skaičiai generuojami iš apibrėžto skaičių diapazono (pvz., nuo 1 iki 60) taip, kad kiekvienas skaičius turi vienodą galimybę būti pasirinktam.

- *Sisteminė atsitiktinė imtis (atranka)* yra tokia, kai pirmasis imties narys parenkamas atsitiktinai, o kiti imties nariai imami tam tikrais intervalais (pvz., kas 4-as vienetas). Pavyzdžiui, daugelis gyventojų

sąrašų būna sudaryti, išrikiavus pavardes abėcėlės tvarka. Sudarant sisteminę imtį pirmasis imties elementas išrenkamas atsitiktinai iš pirmųjų q sąrašo elementų, o likusieji imties elementai vienareikšmiškai nustatomi vis pridėdant po q prie pirmuoju išrinkto elemento numerio iki pat sąrašo pabaigos.

- *Sluoksninė (stratifikuota) atsitiktinė imtis (atranka)* yra tokia, kai nustatomi atitinkami populiacijos pogrupiai ir iš kiekvieno sluoksnio atrenkamos atsitiktinės imtys.

- *Lizdinė imtis (atranka)*, yra tokia, kai visa populiacija suskirstoma į panašias pagal tam tikrą požymį grupes – lizdus (klasterius). Iš visų lizdų aibės paprastosios atsitiktinės imties būdu parenkama dalis lizdų. Į imtį patenka visi atrinktųjų lizdų elementai.

Neatsitiktinėje (ar netikimybinėje) imtyje kai kurie populiacijos vienetai neturi galimybės būti atrinkti arba jų atrankos galimybių negalima nustatyti.

Neatsitiktinė atranka apima patogią atranką, tikslią atranką, atranką pagal kvotą ir savanorių atranką.

- *Patogi atranka* - vienetai parenkami atsižvelgiant į jų prieinamumą.

- *Tikslinė (ekspertinė) atranka* - imtis parenkama atsižvelgiant į tai, kas, tyrėjo nuomone, tinka tyrimui.
- *Kvotinė atranka* - populiacija suskaidoma į dalis pagal kuriuos nors tyrėjų pasirinktus požymius, tarkime, pagal demografinius rodiklius: amžių, lytį, gyvenamąją vietą, tautybę. Imtis sudaroma taip, kad dalis imties elementų, turinčių požymį, būtų lygi daliai visos populiacijos elementų, turinčių tą patį požymį. Pavyzdžiui, numatoma, kad imtį sudarys 86 lietuviai, 7 lenkai, 5 rusai, 1 baltarusis ir 1 ukrainietis.
- *Savanorių atranka* - dalyviai savanoriškai dalyvauja apklausoje (paprastas metodas, naudojamas internetinėse nuomonės apklausoje, kai nekontroliuojama, kiek ar kas balsuoja).
- *Pagavimo, paleidimo ir pakartotinio pagavimo principu pagrįsta imtis*. Tai tokia imtis, kai judantys populiacijos elementai (pavyzdžiui, žuvis, paukščiai) pagaunami, pažymimi, paleidžiami ir pakartotinai gaudomi. Imtį sudaro tiesiog pakartotinai pagauti elementai. Tokioje imtyje skaičiuojami žymėtieji elementai ir, remiantis šios imties duomenimis, vertinami populiacijos parametrai.

Imties sudarymo planas, o kartu ir statistinio tyrimo planas yra *šališkas*, jei jis sistemingai teikia pirmenybę tam tikriems rezultatams. Pavyzdžiui, patogi atranka ir savanorių atranka dažniausiai yra šališkos. Paprastoji atsitiktinė imtis yra vienas iš labiausiai paplitusių imties sudarymo būdų ir pasižymi nešališkumu.

Klasės pavyzdys. Klasėje mokosi 20 mokinių. Jie sėdi keturiose eilėse po penkis mokinius kiekvienoje. Norime sudaryti paprastą atsitiktinę imtį, iš 4 klasės mokinių. Norėdami tai padaryti, iš kiekvienos eilės pasirenkime po vieną mokinį taip. Užrašykime skaičius nuo 1 iki 5 ant vienodų lapelių. Sumaišius lapelius kepurėje, ištraukiame vieną lapelį. Pasirenkamas mokinys, atitinkantis lapelyje užrašytą skaičių, pradeda pirmą eilę iš kairės. Pavyzdžiui, jei pasirinktas skaičius yra 3, pasirenkame trečią mokinį iš kairės pirmoje eilėje. Vėl sumaišome lapelius ir ištraukiame naują skaičių. Atrenkamas mokinys, sėdintis iš kairės antroje eilėje, pagal lapelyje nurodytą skaičių. Pakartojame šį procesą likusioms dviem eilėms. Kiekvienas klasės mokinys turi 1 iš 5 galimybių būti atrinktas. Ar kiekvienas mokinys turi vienodas galimybes būti atrinktas? Ar šis pavyzdys yra paprastos atsitiktinės imties pavyzdys? Paaiškinkite.

Vardų ar skaičių traukimas iš kepurės parodo, ką reiškia suteikti kiekvienam asmeniui ir kiekvienai galimai n asmenų grupei vienodą galimybę būti išrinktam. Tokia yra paprastosios atsitiktinės atrankos idėja. Be abejo, renkant iš 3 milijonų šalies gyventojų, traukti lapelius iš kepurės būtų šiek tiek nepatogu. Praktikoje

naudojame kompiuteriu sugeneruotus atsitiktinius skaičius. Galima pasinaudoti ir internetiniais resursais:

<https://www.calculator.net/random-number-generator.html>

DUOMENŲ PASISKIRSTYMAS: HISTOGRAMA IR MEDŽIO DIAGRAMA

Statistinio kintamojo pasiskirstymas nurodo, kokias reikšmes jis įgyja ir kaip dažnai šias reikšmes įgyja.

Dažniausias kiekybinio kintamojo pasiskirstymo grafikas yra histograma.

Kaip sudaryti histogramą?

1 lentelėje pateiktas 65 metų ir vyresnių gyventojų dalis procentais kiekvienoje iš 50 JAV valstybių. Norėdami sudaryti šio skirstinio histogramą, atlikite tokius veiksmus.

1 veiksmas. Padalinkite duomenų diapazoną į vienodo ilgio intervalus.

Valstija	65 m. ir vyresnių gyventojų dalis, %	Valstija	65 m. ir vyresnių gyventojų dalis, %	Valstija	65 m. ir vyresnių gyventojų dalis, %
Alabama	13,4	Luiziana	12,2	Ohajas	13,4
Alaska	6,8	Meinas	14,6	Oklahoma	13,2
Arizona	12,8	Merilandas	11,6	Oregonas	12,9
Arkanzasas	13,9	Masačusetsas	13,3	Pensilvanija	15,2
Kalifornija	10,8	Mičiganas	12,5	Rodo sala	13,9
Koloradas	10,0	Minesota	12,1	Pietų Karolina	12,8
Konektikutas	13,4	Misisipė	12,4	Pietų Dakota	14,2
Delaveras	13,4	Misūris	13,3	Tenesis	12,7
Florida	16,8	Montana	11,1	Teksasas	9,9
Džordžija	9,8	Nebraska	12,4	Juta	8,8
Havajai	14,0	Nevada	12,9	Vermontas	13,3
Aidaho	11,5	Naujasis Hampšyras	12,4	Virdžinija	11,6
Ilinojaus	11,9	Naujasis Džersis	12,9	Vašingtonas	11,5
Indiana	12,4	Naujasis Meksikas	12,4	Vakarų Virdžinija	15,3
Ajova	14,6	Niujorkas	13,1	Viskonsinas	13,0

Kanzasas	12,9	Šiaurės Karolina	12,2	Vijomingas	12,2
Kentukis	12,4	Šiaurės Dakota	14,6		

Šaltinis: 2008 Statistical Abstract of the United States, pasiekama internete www.census.gov/compendia/statab/.

1 lentelė. 65 m. ir vyresnių gyventojų dalis procentais.

Duomenys svyruoja nuo 6,8 iki 16,7, todėl pasirenkame tokius intervalus:

$6,0 \leq$ procentas virš $65 < 7,0$;

$7,0 \leq$ procentas virš $65 < 8,0$...

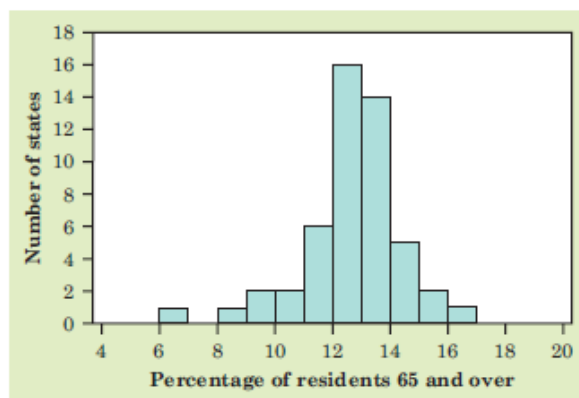
$16,0 \leq$ procentas virš $65 < 17,0$.

Intervalus būtinai parinkite taip, kad kiekvienas duomuo priklausytų tiksliai vienam iš jų. Kitaip tariant, įsitikinkite, kad joks duomuo nepriklauso daugiau nei vienam intervalui ir kiekvienas duomuo yra tam tikroje klasėje. Valstybė, kurioje 6,9% gyventojų yra 65 metų ir vyresni, patenka į pirmą intervalą, tačiau 7,0% patenka jau į antrąjį.

2 veiksmas. Suskaičiuokite valstijų skaičių kiekviename intervale. Štai rezultatas:

Intervalas	Valstijų skaičius	Intervalas	Valstijų skaičius	Intervalas	Valstijų skaičius
[6, 7)	1	[10, 11)	2	[14, 15)	5
[7, 8)	0	[11, 12)	6	[15, 16)	2
[8, 9)	1	[12, 13)	16	[16, 17)	1
[9, 10)	2	[13, 14)	14		

3 veiksmas. Nubraižykite histogramą. Horizontalioje ašyje pažymėkite kintamojo, kurio pasiskirstymą rodote, skalę. Šiame pavyzdyje tai yra „65 metų ir vyresnių gyventojų dalis procentais“. Skalė yra nuo 4 iki 20, nes šis diapazonas apima pasirinktas klases. Vertikalioje ašyje yra skaičiavimų skalė. Kiekviena juosta reiškia klasę. Juostos pagrindas yra atitinkama klasė, o juostos aukštis yra klasės skaičius. Tarp juostų nėra horizontalaus tarpo, nebent klasė yra tuščia, todėl jos juostos aukštis yra nulis.



1 histograma. Kintamojo „65 metų ir vyresnių gyventojų dalis procentais“ histograma

Užduotis. Žemiau pateikiami 2006 m. duomenys apie 15–44 metų amžiaus gyventojus iš 50 JAV valstijų. Tiriamas parametras - 15–44 metų amžiaus gyventojų dalis procentais.

State	Percent	State	Percent
Alabama	41.0	Montana	39.1
Alaska	45.0	Nebraska	41.2
Arizona	42.3	Nevada	43.0
Arkansas	40.8	New Hampshire	41.1
California	44.4	New Jersey	41.3
Colorado	44.1	New Mexico	41.4
Connecticut	40.7	New York	42.4
Delaware	41.4	North Carolina	42.4
Florida	39.7	North Dakota	41.3
Georgia	44.5	Ohio	40.8
Hawaii	41.6	Oklahoma	41.2
Idaho	42.0	Oregon	41.2
Illinois	42.8	Pennsylvania	39.8
Indiana	41.8	Rhode Island	42.2
Iowa	40.3	South Carolina	41.5
Kansas	41.4	South Dakota	40.1
Kentucky	41.6	Tennessee	41.7
Louisiana	41.7	Texas	44.3
Maine	39.2	Utah	45.6
Maryland	42.4	Vermont	40.2
Massachusetts	42.2	Virginia	43.0
Michigan	41.4	Washington	42.6
Minnesota	42.2	West Virginia	39.1
Mississippi	41.5	Wisconsin	41.6
Missouri	41.2	Wyoming	40.8

Padarykite šio parametro reikšmių pasiskirstymo skirstinio histogramą pagal pavyzdyje aprašytus tris žingsnius.

Histogramos sudarymas nėra tikslas savaime. Tikslas – padėti suprasti duomenyse paslėptą informaciją apie populiaciją. Po to, kai histograma padaryta, verta paklausti „ką matome?“ Bendra strategija, žiūrint į grafiką gali būti tokia.

- Ieškokite bendro modelio ir ryškių nukrypimų nuo to modelio.

- Apibūdinkite centrą ir sklaidą: centrą galime apibūdinti vidurio tašku, kai maždaug pusė stebėjimų yra mažesnės reikšmės, o pusė – didesnės. Pasiskirstymo sklaidą galime apibūdinti pateikdami mažiausią ir didžiausią reikšmes.
- Apibūdinkite simetriškumą.

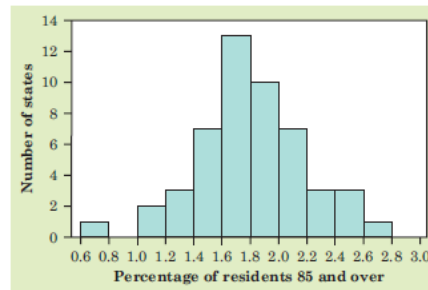
Pasiskirstymas yra simetriškas, jei dešinė ir kairė histogramos pusės yra apytiksliai vienas kito veidrodiniai atvaizdai.

Pasiskirstymas yra iškreiptas į dešinę, jei dešinioji histogramos pusė (kurioje yra pusė stebėjimų su didesnėmis reikšmėmis) yra daug toliau nei kairioji pusė.

Pasiskirstymas yra iškreiptas į kairę, jei kairioji histogramos pusė yra daug toliau nei dešinioji pusė.

Dar kartą pažiūrėkite į 1 histogramą. Histogramos forma: pasiskirstymas turi vieną smailę. Jis yra maždaug simetriškas, tai yra, modelis yra panašus abiejose smailės pusėse. Centras: pasiskirstymo vidurio taškas yra arti vienos smailės, maždaug 13%. Sklaida: skirtumas yra apie 8–17%, jei nepaisysime išskirčių.

Užduotis. 85 metų ir vyresni. 2 histograma yra 85 metų ir vyresnių gyventojų procentas 50 valstybių 2006 m. duomenimis. Apibūdinkite šio pasiskirstymo formą, centrą ir sklaidą. Ar yra kokių nors nukrypimų?



2 histograma. 85 metų ir vyresni

Histogramos nėra vienintelis grafinis pasiskirstymo vaizdas. Mažiams duomenų rinkiniams medžio diagrama yra ne tik sukuriama greičiau, bet dažnai pateikia išsamesnę informaciją.

Medžio diagrama.

1. Atskirkime duomenį į kamieną, kurį sudaro visi, išskyrus paskutinį skaitmenį, ir lapą kuris yra paskutinis skaitmuo. Kamienas gali turėti tiek skaitmenų, kiek reikia, tačiau kiekviename lape yra tik vienas skaitmuo.
2. Įrašykite kamienus į vertikalų stulpelį, kurio viršuje yra mažiausias, o stulpelio dešinėje nubrėžkite vertikalią liniją.
3. Užrašykite kiekvieną lapą eilutėje, esančioje dešinėje nuo jo kamieno bet kuria arba didėjančia tvarka.

Pavyzdys duomenims “ 65 metų ir vyresnių gyventojų dalis procentas ”.

Pirmiausia sudarome kamieną. Pavyzdžiui, Alabamos procentas yra 13.4. Jo kamienas 13, o lapas - 4.

Kitame žingsnyje kiekvienam stebėjimui surašome lapus. Galiausiai, lapus išrikiuojame didėjimo tvarka. Štai rezultatas:

6	6	8	6	8
7	7		7	
8	8	8	8	8
9	9	89	9	89
10	10	80	10	08
11	11	56165	11	15566
12	12	80494251449429872	12	01222444445788999
13	13	49443383142930	13	01233334444899
14	14	06662	14	02666
15	15	23	15	23
16	16	8	16	8

Kintamojo pasiskirstymas parodo, kokias reikšmes kintamasis įgauna ir kaip dažnai jis naudoja kiekvieną reikšmę.

Norėdami parodyti kiekybinio kintamojo reikšmių pasiskirstymą, naudokite histogramą arba medžio diagramą. Pirmenybę teikite medžio diagramai, kai turite nedaug stebėjimų ir histogramoms, kai turite didesnius duomenų rinkinius.

DUOMENŲ PASISKIRSTYMO PARAMETRAI

Mediana ir kvartilai.

Paprastas ir efektyvus būdas apibūdinti duomenų centrą ir sklaidą yra pateikti medianą ir kvartilius.

Mediana - duomenų rinkinio pasiskirstymo mediana yra sutvarkyto duomenų sąrašo centrą nusakanti reikšmė.

Norėdami rasti pasiskirstymo medianą atlikite šiuos veiksmus:

1. Išdėstykite visus stebėjimus pagal dydį, nuo mažiausio iki didžiausio (gausite taip vadinamą variacinę eilutę arba sutvarkytą duomenų sąrašą).
2. Jei stebėjimų skaičius n yra nelyginis, mediana m yra sutvarkyto duomenų sąrašo centre esanti reikšmė. Pavyzdžiui, duomenų rinkinyje $\{1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 26\}$ šeštoji reikšmė turi penkias virš jos ir penkias po ja. Taigi šeštoji reikšmė 3 yra šio duomenų rinkinio mediana.
3. Jei duomenų rinkinyje yra lyginis reikšmių skaičius, mediana randama imant dviejų sutvarkyto duomenų sąrašo centre esančių reikšmių vidurkį. Pavyzdžiui, duomenų rinkinio $\{2, 2, 4, 5, 7, 7\}$ mediana yra $(4+5)/2=4.5$. Ji yra tarp trečio ir ketviro duomenų rinkinio narių.

Kvartiliai yra skaičiai, padalijantys sutvarkytą duomenų rinkinį į keturias dalis, kurių kiekvienoje yra maždaug ketvirtadalis duomenų. Dvidešimt penki procentai duomenų rinkinio patenka į pirmąjį kvartilį (Q_1). Mediana yra antrasis kvartilis (Q_2); 50% duomenų yra prieš medianą. Septyniasdešimt penki procentai duomenų yra prieš trečiąjį kvartilį (Q_3).

Norėdami apskaičiuoti duomenų pasiskirstymo kvartilius, atlikite šiuos veiksmus:

1. Išdėstykite stebėjimus didėjančia tvarka ir sutvarkytame stebėjimų sąrašė suraskite medianą m .
2. Pirmasis kvartilis Q_1 yra stebėjimų, kurių padėtis sutvarkytame sąrašė yra į kairę nuo medianos vietos, mediana. Bendra mediana į stebėjimus į kairę nuo bendros medianos neįtraukiama.
3. Trečiasis kvartilis Q_3 yra stebėjimų, kurių padėtis sutvarkytame sąrašė yra į dešinę nuo medianos vietos, mediana. Bendra mediana neįtraukiama į stebėjimus, kurie laikomi dešinėje nuo bendros medianos.

5	16	19	24	25	25	26	28	33	33	34	34	37	37	40	42	45	45	46	46	49	73
										↑		↑						↑			
										Q_1		m						Q_3			

Penkių skaičių santrauka

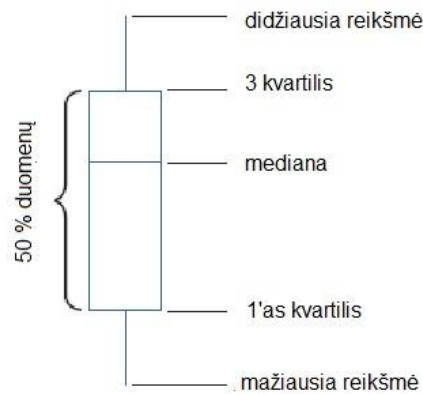
Duomenų pasiskirstymo *penkių skaičių santrauka* susideda iš mažiausios reikšmės, pirmojo kvartilio, medianos, trečiojo kvartilio ir didžiausios reikšmės, surašytos eilės tvarka nuo mažiausio iki didžiausio.

Simboliuose, penkių skaičių santrauka yra

Mažiausia reikšmė Q_1 m Q_3 Didžiausia reikšmė

Stačiakampė diagrama, taip pat žinoma kaip *stačiakampio ir ūsų diagrama*, yra grafinis duomenų rinkinio penkių skaičių suvestinės vaizdas. Iš apatinio kvartilio (Q_1) į viršutinį kvartilį (Q_3) nubrėžiamas stačiakampis langelis; horizontali linija skersai langelio rodo medianą. Nubrėžiami du ūsai, vienas iš apatinio kvartilio iki minimumo ir vienas nuo viršutinio kvartilio iki maksimumo.

Stačiakampes diagramas galima naudoti grafiniam duomenų rinkinių palyginimus arba skirtumams duomenų rinkinio dalyse matuoti.



Vidurkis ir standartinis nuokrypis.

Vidurkis - duomenų rinkinio vidurkis yra duomenų rinkinio aritmetinis vidurkis, kuris gaunamas sudėjus visas reikšmes, ir gautą sumą padaliję iš reikšmių skaičiaus.

Pavyzdžiui, duomenų rinkinio $\{1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 26\}$, vidurkis yra 5; jį rasite padalydami tame rinkinyje esančių reikšmių sumą (55) iš reikšmių skaičiaus (11).

Statistinio kintamojo x imties vidurkio \bar{x} (skaitome „ x su brūkšniu“) algebrinė išraiška yra

$$\bar{x} = \frac{1}{n} \sum x.$$

Priminsime, kad simbolinis užrašas $\sum x$ reiškia, kad reikia susumuoti visas statistinio kintamojo x stebėtas reikšmes.

Standartinis nuokrypis matuoja duomenų rinkinio vidutinį kvadratinį nuokrypį nuo jų vidurkio.

Norėdami rasti n stebėjimų standartinį nuokrypį s_x :

1. Raskite kiekvieno stebėjimo nuokrypio nuo vidurkio kvadratą. Algebrinė šio veiksmo išraiška yra

$$(x - \bar{x})^2$$

2. Apskaičiuokite gautų reikšmių sumą ir padalinkite iš $n - 1$. Algebrinė šio veiksmo išraiška yra

$$\frac{1}{n - 1} \sum (x - \bar{x})^2$$

3. Standartinis nuokrypis s_x yra kvadratinė šaknis iš gautos reikšmės. Algebrinė išraiška yra

$$s_x = \sqrt{\frac{1}{n - 1} \sum (x - \bar{x})^2}$$

Keletas pastabų.

- s_x matuoja statistinio kintamojo x duomenų išsibarstymą (sklaidą) apie vidurkį \bar{x} .
- Naudokite s_x paskirstymo sklaidos apibūdinimui tik tada, kai vidurkį \bar{x} naudojate paskirstymo centrui apibūdinti.
- $s_x = 0$ atsitinka tik tada, kai visi stebėjimai turi tą pačią reikšmę. Taigi standartinis nuokrypis lygus nuliui reiškia, kad iš viso nėra duomenų sklaidos. Kitu atveju $s_x > 0$.
- s_x tuo didesnis, kuo duomenys yra labiau išsisklaidę apie jų vidurkį.

Dar reikia atkreipti dėmesį į tai, kad tiek vidurkį, tiek standartinį nuokrypį iškreipia išsikirtys. Medianą ir kvartilius jos veikia mažiau. Taigi naudokite \bar{x} ir s_x tik pakankamai simetriškiems skirstiniams, kuriuose nėra išskirčių.

STATISTINIS PATIKIMUMAS

Planuojant apklausą reikia siekti kuo mažesnio šališkumo. Be abejo, kuo didesnis imties dydis, tuo geriau. Tai pagerina rezultatų patikimumą. Norėdami pasakyti,

kokio tikslumo yra išvados apie populiaciją, padarytos išanalizavus imtį, turime *suprasti statistinį patikimumą*. Ką tai reiškia?

Jokia atsitiktinai pasirinkta imtis tiksliai neįvertina tikrosios populiacijos parametro reikšmės (nes pastarosios nežinome). Taigi parametro įvertinys turės paklaidą. Todėl reikalingos paklaidos ribos, kad galima būtų pasakyti, kiek tikslus yra gautas įvertinys. Tačiau ir žinant bent apytikres paklaidos ribas, vis vien negalime būti tikri, kad tikroji parametro reikšmė nuo įvertinimo skiriasi ne daugiau nei leidžia paklaidos ribos.

Nors 95 % visų imčių duos rezultatą arti tiesos, 5 % jų gali išeiti už paklaidos ribų. Nežinome tiesos apie populiaciją, todėl nežinome, ar mūsų imtis yra viena iš tų 95 %, kurie pataikė, ar viena iš 5 %, kurios nepataiko. Šiuo atveju sakome, kad galime būti 95 % tikri, kad tiesa yra paklaidos ribose.

Statistinį patikimumą sudaro dvi dalys: *paklaidos riba* ir *pasitikėjimo lygmuo*. Klaidos riba rodo, kaip arti populiacijos parametro yra imties statistika. Pasitikėjimo lygmuo parodo, kiek procentų visų galimų imčių atitinka paklaidos ribą.

Dažniausia pasitikėjimo lygmuo yra 95%. Tai yra, jei pasirinktume daug imčių, tiesa apie populiaciją būtų paklaidos ribose 95% atveju. Grubiai galime įvertinkite 95 % pasiklovimo lygmenį atitinkančią paklaidą. Paprastosios atsitiktinės n dydžio imties atveju ji yra $1/\sqrt{n}$. Kaip rodo ši formulė, svarbu imties dydis, o ne populiacijos dydis.

Pavyzdys. Ar pritariate įstatymo pataisai?

Visų suaugusiųjų dalis, pasisakančių už įstatymo pataisą, yra parametras, apibūdinantis, tarkime, 2 milijonus suaugusiųjų. Pavadinkite tą parametną p . Deja, p skaitinės reikšmės nežinome. Norime ją įvertinti. Tuo tikslu sudaryta atsitiktinė imtis iš 2527 suaugusiųjų. Imties dalis, kuri pritarė konstitucijos pataisai yra statistika, kuria įvertiname parametną p . Pavadinkime ją \hat{p} (skaitome „ p -su kepure“). Tarkime, kad 1289 iš šios 2527 dydžio imties teigė, kad pritaria įstatymo pataisai, todėl šios imties atveju $\hat{p} = \frac{1289}{2527} = 0.51$ (t.y. 51%). Kadangi visi suaugusieji turėjo vienodą galimybę būti tarp pasirinktų 2527, atrodo pagrįsta naudoti statistiką $\hat{p} = 0.51$, kaip nežinomo parametro p įvertinimą. Faktą, kad 51 % imties narių pritarė pakeitimui – žinome, nes jų paklausėme. Tačiau nežinome, koks procentas visų suaugusiųjų pritaria pataisai, bet manome, kad tai bus maždaug 51 proc.

Jei sudarytume antrą atsitiktinę 2527 suaugusiųjų imtį, nauja imtis duotų kitą rezultatą, nes joje būtų taip pat įvairių žmonių. Beveik neabejotina, kad pritariančiųjų skaičius nebus lygiai 1289. Tai yra, statistikos \hat{p} reikšmė skirsis

skirtingoms imtims. Ar gali atsitikti taip, kad viena atsitiktinė imtis rodo kad 51% suaugusiųjų pritaria pataisai, o antrosios atsitiktinės imties rezultatas sako, kad tik 37% pritaria pataisai? Atsitiktinės imtys nėra šališkos, tačiau gauti atsakymai vis tiek gali būti klaidingi, nes atsiranda kintamumas dėl atsitiktinės atrankos. Čia gelbsti du atsitiktinių imčių privalumai. Pirmas privalumas yra tas, kad atsitiktinis pasirinkimas yra nešališkas. Antrasis privalumas: jei iš tos pačios populiacijos paimtume daug tokio paties dydžio atsitiktinių imčių, svyravimai tarp imčių nebūtų dideli.

APIBENDRINIMAS

Imties sudarymo tikslas – gauti informaciją apie populiaciją. Norėdami įvertinti populiacijos parametą, dažniausiai naudojame imties statistiką. Norėdami apibūdinti, imties patikimumą paklauskite: „Kas nutiktų, jei paimtume daug imčių iš tos pačios populiacijos? Jei beveik visų imčių rezultatas būtų artimas tiesai, galime pasitikėti savo vienu pavyzdžiu, nors negalime būti tikri, kad tai yra arti tiesos.

Kintamojo pasiskirstymas parodo, kokias reikšmes kintamasis įgyja ir kaip dažnai. Norėdami parodyti kiekybinio kintamojo pasiskirstymą, naudokite histogramą arba medžio diagramą. Dažniausiai pirmenybę teikiame medžio diagramai, kai turime nedaug stebėjimų, ir histogramai - didesniems duomenų rinkiniams. Grafikuose ieškokite bendro modelio ir nukrypimų nuo to modelio, pvz., išskirčių. Galime apibūdinti histogramą ir medžio diagramą nurodydami jo formą, centrą ir sklaidą. Kai kurie skirstiniai turi paprastas formas, pvz., simetriškas arba iškreiptas, tačiau kiti yra pernelyg netaisyklingi, kad juos būtų galima apibūdinti paprasta forma.

Yra du bendri centro ir sklaidos aprašymai: penkių skaičių santrauka ir vidurkis bei standartinis nuokrypis. Penkių skaičių santrauką sudaro mediana ir du kvartiliai bei mažiausi ir didžiausi stebėjimai, apibūdinantys sklaidą. Stačiakampė diagrama yra penkių skaičių suvestinės grafikas.

DARBUI SU MOKINIAIS

Projektai

Detaliau pristatysime vieną projektą. Jame gali dalyvauti visa klasė.

1 projektas. Ar muzikos klausymas trukdo mokytis?

Pradėkime nuo motyvacijos.

Įprasta, kad besimokydami mokiniai klausosi muzikos. Kai kurie mokiniai teigia, kad muzikos klausymas jiems padeda susikaupti ir nuteikia teigiamai. Kitiems priešingai – trukdo susikaupti. Reikia šią problemą iširti. Pradėti galime nuo klasės. Tuo tikslu moksleiviai gali kurti eksperimentinį tyrimą, skirtą iširti muzikos klausymosi poveikį bendraklasių gebėjimui įsiminti žodžius.

Statistinio klausimo formulavimas.

Statistinis tyrimo klausimas gali būti:

Ar klausydami muzikos mokiniai sugeba įsiminti daugiau žodžių nei nesiklausydami?

Kitas klausimas kuris gali kilti, kiek įtakos atminčiai turi muzikos klausymosi trukmė bei muzikos stilius. Čia vienas kintamasis būtų laikas arba stilius, kitas – įsimenamų žodžių skaičius.

Duomenų surinkimas.

Klasė turi sukurti eksperimentinių duomenų rinkimo strategiją, kad būtų galima atsakyti į tiriamąjį klausimą. Reikės, kiek įmanoma, pašalinti galimus pašalinius kintamumo šaltinius, kurie gali trukdyti interpretuoti rezultatus.

Vienas paprastas eksperimentas, kurį galima atlikti su klase, būtų atsitiktinai padalyti klasę į dvi vienodo dydžio grupes. Taip moksleiviams ugdomas supratimas, kad atsitiktinis priskyrimas yra svarbi eksperimentinio plano dalis, nes ji linkusi mokinių gebėjimus atskirti nuo kitų savybių, kurie gali paveikti atsakymus. Kaip tai padaryti? Tarkime, kad 28 mokiniai atsitiktinai suskirstyti į dvi grupes po 14. Jei įmanoma, klasė turėtų atsitiktinės grupės priskyrimui naudoti technologiją, tačiau taip pat veiktų tokie įrankiai kaip kortų kaladė. Pavyzdžiui, mokytojas gali sumaišyti 28 kortas, 14 raudonų ir 14 juodų kortelių ir išdalinti po vieną kiekvienam mokiniui. Gavę raudoną kortelę dalyvaus muzikos klausytojų grupėje, o gavę juodą – tylos grupėje.

Klasė turėtų sukurti eksperimentinę procedūrą ir susitarti dėl jos detalių. Pavyzdžiui, kiekvienas dalyvis gali turėti dvi minutes 20 žodžių sąrašui išnagrinėti, o po to daryti vienos minutės pauzę, o tada – per dvi minutes užsirašyti kuo daugiau žodžių iš to sąrašo. Viso eksperimento metu grupės dalyviai klausysis tos pačios dainos, o kontrolinė grupė visą laiką tylės. Teisingai įsimenamų žodžių skaičius yra dominantis statistinis kintamasis.

Išties tyrimas apima du kintamuosius: dalyvio atsitiktinai priskirtą grupę, kuri yra kategorinis kintamasis, ir teisingai įsimintų žodžių skaičius, kuris yra kiekybinis kintamasis.

Duomenų analizavimas.

Mokiniai turėtų apskaičiuoti suvestinę statistiką muzikos klausymo grupei ir tylos grupei, nubrėžti histogramas ir medžio diagramas atskirai kiekvienai eksperimentinei grupei. Stačiakampės diagramos suteiks daug informacijos tiriamam klausimui atsakyti.

Jei duomenų yra daug, moksleiviai gali naudoti technologijas. Tik svarbu žinoti, kaip jas naudoti.

Rezultatų interpretavimas. Koks gaunasi atsakymas į iškeltą klausimą?

Mokiniam gali kilti klausimas, ar muzikos tipas buvo svarbus veiksnys rezultatui. Pavyzdžiui, ar žodžių buvimas dainoje prisideda prie skirtumų tarp dviejų grupių? Šiuo atveju galima kartoti eksperimentą su nauja moksleivių grupe, šį kartą naudojant tik instrumentinę muziką.

Moksleiviai turėtų suprasti šio eksperimento išvadų apimtį. Atranka nebuvo atsitiktinė, jie buvo atsitiktinai priskirti tyrimo grupėms. Galima diskutuoti apie rezultatų patikimumą. Tuo tikslu, galima eksperimentą plėtoti kitoje klasėje.

2 projektas. Kiek trunka 1 minutė?

Šį projektą galima siūlyti jau nuo 5 klasės. Duomenims surinkti reikės chronometro. Imtis gali būti iš vienos klasės ar kelių klasių. Galima kelti, pavyzdžiui, tokius statistinius klausimus.

- (a) Koks yra klasės mokinių laiko suvokimas?
- (b) Ar vienos ir kitos klasės mokinių laiko suvokimas yra skirtingas?
- (c) Ar skiriasi mergaičių ir berniukų laiko suvokimas?

3 projektas. Fizinių duomenų statistiniai tyrimai

Šiame tyrime moksleiviai domisi įvairiais kūno matavimo duomenimis (galvos apimtis, rankų ilgis, pėdos ilgis ir kt.) ir asmens gebėjimais atlikti unikalūs su kūnu susijusius veiksmus. Tam prireiks pasirinkti skirtingus kūno parametrus, taip pat registruoti lytį ir amžių bei kelis su kūnu susijusius pratimus.

UŽDUOTYS

Norėdami įtvirtinti įgytas žinias ir panaudotus statistinius instrumentus, galima moksleiviams pasiūlyti išspręsti kelias užduotis. Atkreipiame dėmesį į tai, kad geriau pateikti užduotis su kontekstu. Tik retais atvejais galima pasiūlyti duomenis be konteksto.

1 užduotis. *Automobilių, varomų dujomis, rida.*

JAV vyriausybės nuostatai reikalauja, kad automobilių gamintojai kiekvienam automobilio modeliui nurodytų degalų sąnaudas. Lentelėje pateikiamos 31 2008 metų sedanų degalų sąnaudos važiuojant greitkelyje (mylios už galoną).

Model	mpg	Model	mpg
Acura RL	24	Lexus ES350	27
Bentley Arnage	15	Lexus GS460	24
BMW 535i	26	Lincoln Town Car	23
Buick Lacrosse	24	Maybach 57	16
Cadillac CTS	26	Maybach 62	16
Cadillac STS	26	Mazda 6	26
Chevrolet Malibu	30	Mercedes-Benz E350	24
Chrysler Sebring	27	Mercedes-Benz E550	22
Dodge Avenger	30	Nissan Maxima	25
Honda Accord	31	Pontiac Grand Prix	28
Hyundai Sonata	31	Rolls Royce Phantom	18
Infiniti G35	25	Saturn Aura	30
Infiniti M35	23	Toyota Camry	31
Jaguar S-Type R	22	Volkswagen Passat	26
Kia Optima	31	Volvo S80 AWD	24
Kia Spectra	32		

Source: www.fueleconomy.gov.

Išanalizuokite šių automobilių degalų sąnaudas greitkelyje.

- Sudarykite histogramą. Ką galite pasakyti apie bendrą pasiskirstymo formą?
- Kur yra pasiskirstymo centras?
- Trys iš šių automobilių buvo apmokestinti „degalų perviršio mokesčiu“, nes jų rida yra per maža. Kokie tie trys automobiliai?
- Padarykite šių duomenų medžio diagramą.
- Raskite penkių skaičių suvestinę. Kurie automobiliai patenka į apatinę degalų sąnaudų ketvirtį?
- Medžio diagrama parodo faktą apie bendrą pasiskirstymo formą, kurio negali apibūdinti penkių skaičių santrauka. Ar tai faktas?

2 užduotis. *Rašymo stiliaus statistika.*

Skaitiniais duomenimis galima išskirti skirtingus rašymo tipus, o kartais net atskirus autorius. Štai moksleivių surinkti duomenys apie 1–15 raidžių žodžių procentą, žurnalo „Popular Science“ straipsniuose:

Žodžio ilgis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Procentas	3.6	14.8	18.7	16.0	12.5	8.2	8.1	5.9	4.4	3.6	2.1	0.9	0.6	0.4	0.2

Statistinis kintamas: raidžių skaičius žodyje, žurnalo „Popular Science“ straipsniuose.

- Padarykite šio statistinio kintamojo pasiskirsymo histogramą. Apibūdinkite jo formą, centrą ir išplitimą.
- Raskite penkių skaičių suvestinę.
- Analogišką užduotį atlikite su Lietuviškais leidiniais (tai gali tikti ir projektiniam darbui).

3 užduotis. Kiek kalorijų turi dešrainis?

Žurnalas „Consumer Reports“ pateikė šiuos duomenis apie kalorijų skaičių dešrainiuose kiekviename iš 17 dešrainių rūšių. Štai rezultatai:

173 191 182 190 172 147 146 139 175 136 179 153 107 195 135 140 138

Padarykite dešrainių kalorijų pasiskirstymo medžio diagramą ir trumpai apibūdinkite paskirstymo formą.

- Daugelyje dešrainių rūšių naudojamas jautienos ir kiaulienos mišinys, kuriame, pagal nuostatas, leidžiama iki 15 % paukštienos. Žinoma, kad tik viena rūšis pasinaudojo šia galimybe. Kuris medžio diagramos kamienas ir lapas, jūsų manymu, tai parodo?
- Raskite penkių skaičių suvestinę. Ko apie duomenų pasiskirstymą neparodo skaitinė suvestinė?

4 užduotis.

Norėdami rasti vidurkį ir standartinę nuokrypį, naudokite skaičiuotuvą iš šių dviejų skaičių rinkinių:

- 4 0 1 4 3 6

ii) 5 3 1 3 4 2

- a) Kuris duomenų rinkinys turi didesnę sklaidą?
- b) Pridėkite po 2 prie kiekvieno skaičiaus duomenų rinkinyje (i). Dabar (i) duomenys yra 6 2 3 6 5 8. Naudokite skaičiuotuvą, kad surastumėte vidurkį ir standartinį nuokrypį ir palygintumėte savo atsakymus su (i) duomenų rinkinio atsakymais. Kaip prie kiekvieno skaičiaus pridėjus 2 pasikeis vidurkis?
- c) Kaip tai pakeis standartinį nuokrypį?
- d) Neatlikus skaičiavimo, pasakykite kas nutiktų vidurkiui ir standartiniam nuokrypiui, jei prie kiekvienos (i) pratimo duomenų rinkinio reikšmės pridėtume 10?

5 užduotis. Vidurkio ir standartinio nuokrypio nepakanka.

Vidurkis ir standartinis nuokrypis matuoja centrą ir sklaidą, bet nėra išsamus pasiskirstymo aprašymas. Skirtingų formų duomenų rinkiniai gali turėti tą patį vidurkį ir standartinį nuokrypį. Norėdami parodyti šį faktą, naudokite skaičiuotuvą, kad surastumėte šių dviejų mažų duomenų rinkinių vidurkius ir standartinius nuokrypius. Tada padarykite kiekvieno iš jų medžio diagramą ir pakomentuokite kiekvieno paskirstymo formą

A duomenys	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74
B duomenys	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.50

IV gimnazijos klasė. Bendrasis kursas

Įvadas į taikomąją duomenų analizę. Nagrinėdami straipsnius apie mokslo pasiekimus, statistikos ir technologijų vaidmenį šiuolaikiniame pasaulyje, mokiniai sužino, kad funkcijos gali būti naudojamos ir duomenims apibūdinti, o jei duomenys susiję tiesiniu ryšiu, tai tas ryšys gali būti modeliuojamas tiese (regresijos tiese), o jo stiprumas ir kryptis išreikšti koreliacijos koeficientu. Visas naujas sąvokas mokiniai išsiaiškina, nagrinėdami konkrečius pavyzdžius, o reikiamai skaitinei informacijai gauti pasitelkia skaitmenines technologijas. Mokiniai išsiaiškina, kad statistinės analizės (regresinė analizė yra viena iš jos dalių) tikslas – ištyrus dalį respondentų (imtį), padaryti išvadą apie visą populiaciją. Aptariamos kintamojo, kintamojo matavimo skalių, duomenų sąvokos ir išsiaiškinama, kodėl tik intervaliniams duomenims taikomos vidurio, standartinio nuokrypio, koreliacijos (tiesinės koreliacijos koeficiento) skaičiavimo procedūros. Mokoma(si) praktiškai, naudojantis skaitmeninėmis technologijomis, apskaičiuoti duomenų rinkinio imties vidurkį, standartinį nuokrypį, interpretuoti, kaip jie charakterizuoja imtį. Nagrinėjami pavyzdžiai, kai sprendimui dėl kintamųjų ryšio ir jo stiprumo priimti naudojama koreliacija (pavyzdžiui, laiko ir pažymių, amžiaus ir atlyginimo, IQ ir darbo kompiuteriu). Atkreipiamas dėmesys, kad koreliacija nepaaiškina priežastingumo. Paaiškinama, kad priežastingumui tarp kintamųjų nustatyti taikomas kitas matematinis modelis – tiesinė regresija. Nagrinėjamos paprasčiausios tipinės situacijos, kai gali būti taikoma tiesinė regresija (pavyzdžiui, ar per egzaminą surinktų balų skaičius priklauso nuo socialinio statuso). Išsiaiškinama, kaip priimamas sprendimas, kuris kintamasis vadinamas priklausomu kintamuoju, o kuris – aiškinamuoju (regresoriumi). Naudojantis skaičiuoklės programa, demonstruojama, kaip atrodo grafinis duomenų rinkinio vaizdas („taškų debesėlis“). Nagrinėjama problema – ar įmanoma šiuos duomenis aprašyti modeliu (tiese). Išsiaiškinama, kad svarbiausia šio modelio (tiesės) charakteristika – determinacijos koeficientas (R^2) ir mokomasi, jį žinant (suradus), priimti sprendimą dėl gauto modelio tinkamumo duomenims aprašyti.

Kritiškai peržiūrėdami statistinių duomenų naudojimą viešojoje žiniasklaidoje ir įvairiose ataskaitose, mokiniai mokosi diskutuoti apie tyrimo struktūrą, duomenų rinkimo sąlygas ir būdą, duomenų analizei taikytus metodus, duomenų santraukas ir padarytas išvadas.

Kelios pastabos programos turiniui.

„Paaiškinama, kad priešastingumui tarp kintamųjų nustatyti taikomas kitas matematinis modelis – tiesinė regresija.“ Šis teiginys yra klaidingas. Tiesinė regresija nepaaiškina priešastingumo. Yra daug taip vadinamų „klaidingos regresijos“ pavyzdžių. Klaidinga regresija yra statistinis modelis, rodantis klaidinančius statistinius tiesinio priešastinio ryšio įrodymus. Vienas iš daugybės pavyzdžių:

1. Metinis JAV eksporto indeksas (y) (1960-1990 metų duomenimis) ir Australijos vyrų gyvenimo trukmė (x): nustatytas tiesinis modelis

$$y = -2943. + 45.7974 x, \quad \text{koreliacijos koeficientas } r=0.9570 \text{ ir R kvadratas} \\ = 0.916.$$

2. JAV išlaidos gynybai (y), 1971-1990 metais ir PAR gyventojų skaičius: nustatytas tiesinis modelis

$$y = -368.99 + .0179 X, \quad R \text{ kvadratas} = .940, \quad r = .9694.$$

Intuityviai aišku, kad kartu didėjantys arba kartu mažėjantys kintamieji gali būti aprašyti tiesinės regresijos modeliu, tačiau nei vienas iš jų nebus priešastimi kito.

„Išsiaiškinama, kaip priimamas sprendimas, kuris kintamasis vadinamas priklausomu kintamuoju, o kuris – aiškinamuoju (regresoriumi)“.

Priešastingumui nustatyti dažniausiai naudojamas 2003 metų Nobelio premijos laureato Granger priešastingumo testas. Tačiau tai, mūsų nuomone, yra už mokyklinės programos ribų. Todėl lieka ekspertinis kintamųjų paskirstymas į priklausomus ir aiškinančiuosius.

Iš programos turinio matosi, kad reikalingas tik kiek platesnis supratimas apie regresiją. Nauja sąvoka R-kvadratas. Visa kita reikalinga teorija jau pristatyta kartu su 9 klasės turiniu. Šiame etape galima (ir reikia) naudoti sąvokas, grafikus ir kitas duomenų santraukas, kurios buvo nagrinėjamos 9 ir 10 klasėse.

TIESINĖ REGRESIJA

Verta detaliau prisiminti koreliaciją ir regresiją, bei santykį tarp jų.

Jei sklaidos diagrama rodo tiesinį ryšį tarp dviejų kiekybinių kintamųjų, galime apibendrinti šį bendrą modelį nubrėždami liniją grafike. Regresijos linija apibendrina ryšį tarp dviejų kintamųjų: vienas iš kintamųjų padeda paaiškinti arba nuspėti kitą. Tai yra, regresija apibūdina ryšį tarp aiškinamojo kintamojo (x) ir atsako kintamojo (y). Regresijos linija yra tiesi linija, nusakanti, kaip keičiasi atsako kintamasis y, kai keičiasi aiškinamasis kintamasis x. Regresijos liniją dažniausiai naudojame, norėdami numatyti y reikšmę tam tikrai x reikšmei.

Koreliacija matuoja tiesinio ryšio tarp dviejų statistinių kintamųjų kryptį ir stiprumą. Koreliacija ir regresija yra glaudžiai susijusios. Jei dviejų kintamųjų koreliacija rodo tiesinę jų sąsają, tai regresijos linija pasitarnauja tos sąsajos vizualizavimui. Be to, regresijai reikia pasirinkti aiškinamąjį kintamąjį, o koreliacijai to nereikia. Tiek koreliaciją, tiek regresiją stipriai veikia išskirtys.

Regresijos tiesės naudingumas prognozavimui priklauso nuo sąsajos stiprumo. Koreliacijos kvadratas yra geras rodiklis prognozavimo naudingumui įvertinti.

Koreliacijos kvadratas (r-kvadratas), yra y reikšmių kintamumo dalis, kurią paaiškina regresijos tiesė, nubrėžta mažiausių kvadratų metodu.

Verta prisiminti, kaip brėžiama regresijos tiesė mažiausių kvadratų metodu.

Pavyzdys. Turime tris taškus: (1, 1), (-2,-1) ir (3, 2). Šiems trims taškams nubrėžkime tiesę mažiausių kvadratų metodu.

Duomenų skaičius yra $n = 3$. Sudarykime lentelę:

x	y	xy	x ²
1	1	1	1
-2	-1	2	4
3	2	6	9
$\sum x = 2$	$\sum y = 2$	$\sum xy = 9$	$\sum x^2 = 14$

Raskime krypties koeficientą k:

$$k = (n\sum xy - \sum y \sum x) / (n\sum x^2 - (\sum x)^2)$$

$$k = [(3 \times 9) - (2 \times 2)] / (3 \times 14) - (2)^2$$

$$k = (27 - 4)/(42 - 4)$$

$$k = 23/38$$

Raskime laisvąjį narį b :

$$b = (\sum y - k\sum x)/n$$

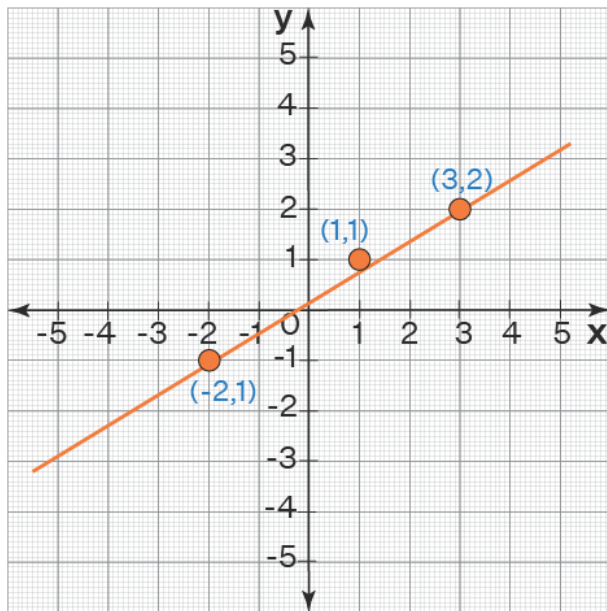
$$b = [2 - (23/38) \times 2]/3$$

$$b = [2 - (23/19)]/3$$

$$b = 15/(3 \times 19)$$

$$b = 5/19$$

Taigi mažiausių kvadratų tiesė yra $y = kx + b = 23/38x + 5/19$, o $r^2 = 0.97$. Taigi regresijos tiesė paaiškina 97% statistinio kintamojo y kintamumo.



Pagrindinės tiesinės regresijos koncepcijos:

Bendroji kvadratų suma (SST), $SST = \sum (y - \bar{y})^2$

regresijos kvadratų suma (SSR), $SSR = \sum (kx + b - \bar{y})^2$

paklaidų kvadratų suma (SSE) $SSE = \sum(y - kx - b)^2$

ir jų ryšys ($SST=SSR+SSE$)

Determinacijos koeficientas R-kvadratas kaip modelio atitikimo matas $R^2 = \frac{SSR}{SST}$ parodo kokią dalį statistinio kintamojo y reikšmių paaiškina tiesinė regresija.

DARBUI SU MOKINIAIS

Projektai

1 projektas. Kodėl mokiniai neišsimiega?

Švietimo administracija ir mokytojai yra susirūpinę dėl pastebimo mokinių miego trūkumo. Mokiniai klasėje atrodo labiau nei bet kada pavargę, sunkiai gali susikaupti ar ilgesnį laiką išlaikyti dėmesį. Dėl to sulaukiama tėvų skundų, kad mokiniams užduodama per daug namų darbų. Tačiau manoma, kad yra keletas galimų veiksnių, kurie prisideda prie miego trūkumo ir tie veiksniai nėra akademiniai. Jie apima mokinių užklausinės veiklos, kurioje aktyviai dalyvauja, skaičių ir jiems skiriamą laiką bei laiką praleistą prie interneto ar elektroninių prietaisų. Nusprendžiama atlikti atrinktų mokyklų mokinių apklausą vienoje ar keliose savivaldybėse ir nustatyti galimus veiksnius, lemiančius vidurinės mokyklos mokinių miego trūkumą.

Tai labai svarbi problema, reikalaujanti rimto mokslininkų dėmesio. Prie jos sprendimo gali prisidėti ir patys moksleiviai, keldami paprastus, su šia problema susijusius klausimus. Pavyzdžiui:

- (a) Kiek būrelių lanko 9-12 klasių mokiniai?
- (b) Kiek laiko užklausinėms veikloms (per savaitę; per mėnesį) skiria 9-12 klasių mokiniai?
- (c) Kiek valandų per dieną po pamokų skiriama internetui?
- (d) Kiek laiko per dieną praleidžiama žaidžiant kompiuterinius žaidimus?

2 projektas. Drabužių dizaineriai turi sukurti drabužių modelius, kurie geriausiai tiktų jų pirkėjams. Kad drabužiai tiktų žmonėms, dizaineris turi suprasti ryšį tarp įvairių žmogaus matmenų. Pavyzdžiui, marškinių dizaineris turi atsižvelgti į žmogaus rankos ilgį ir kaklo apimtį.

Tarkime, klasei buvo pavesta sukurti šabloną mokyklos spektaklio kostiumams. Kostiumams naudojami apsiaustai, besitęsiantys iki pat žemės, nes aktorius pėdos neturi matytis. Apsiaustus reikia sukurti taip, kad rankovė tvirtai priglustų prie rankos, einant nuo alkūnės iki plaštakos. Kaip statistinis tyrimas gali pagelbėti? Bandykime atsakyti į šiuos klausimus.

- Kaip žmogaus dilbio ilgis priklauso nuo jo ūgio?
- Ar dilbio ilgis gali būti naudojamas ūgiui numatyti?

Žemiau pateikti duomenys paimti iš {šaltinis}, tačiau galima organizuoti duomenų surinkimą klasėje arba mokykloje atsitiktinai parinktoje imtyje.

Dilbis (cm)	Ūgis (cm)	Dilbis (cm)	Ūgis (cm)
45.0	180.0	41.0	163.0
44.5	173.2	39.5	155.0
39.5	155.0	43.5	166.0
43.9	168.0	41.0	158.0
47.0	170.0	42.0	165.0
49.1	185.2	45.5	167.0
48.0	181.1	46.0	162.0
47.9	181.9	42.0	161.0
40.6	156.8	46.0	181.0
45.5	171.0	45.6	156.0
46.5	175.5	43.9	172.0
43.0	158.5	44.1	167.0

3 projektas. Kojų ilgis ir žingsnių ilgis.

Padalinkite klasę į grupes po 3 arba 4 mokinius.

Tegu grupės išmatuoja kiekvieno mokinio kojos ilgį nuo išorinio klubo kaulo iki grindų.

Tuo pačiu metu leiskite kiekvienam mokiniui eiti tam tikru maršrutu. Tas pats maršrutas kiekvienam mokiniui. Mokiniai tyliai suskaičiuoja žingsnių skaičių, eidami šiuo maršrutu.

Paprašykite mokinių įvesti duomenis į paprastą duomenų rinkimo formą.

Vardas	Lytis	Dešinės kojos ilgis (cm)	Žingsnių skaičius

Surinkite klasės duomenis ir naudokite sklaidos diagramą koreliacijai ir regresijai iliustruoti.

Tikėtina, kad sąsaja bus silpna ir neigiama.

UŽDUOTYS

1 uždutis. Vietinė ledų parduotuvė seka, kiek ledų jie parduoda ir nori rasti atsakymą į klausimą: kaip dienos ledų pardavimai priklauso nuo tos dienos vidurdienio temperatūros. Štai jų surinkti 12 dienų duomenys:

Vidurdienio temperatūra	Ledų pardavimai (Eurais)
14.2	215
16.4	325
11.9	185
15.2	332
18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408

Išanalizuokite šiuos duomenis ir padėkite parduotuvei atsakyti į jai rūpimą klausimą.

2 uždutis.

Žemiau pateikti duomenys rodo klientų skaičių prie kiekvieno iš šešių restorano stalų ir arbatpinigių, paliktų prie kiekvieno stalo valgio pabaigoje dydį. Išstirkite šiuos duomenis. Raskite tiesinį modelį. Naudokite savo rezultatą ir spėkite arbatpinigių dydį prie stalo, kuriame bus penki klientai.

Pietaujančių skaičius 1 2 3 4 6 8

Arbatpinigiai 3 4 6 7 14 20

3 uždutis.

Pateikti duomenys rodo aštuonių moksleivių išankstinio egzamino balus (iš 20) ir atitinkamo egzamino pažymius (iš 100). Sudarykite tiesinį modelį mažiausių kvadratų metodu. Koks yra numatomas moksleivio, gavusio 17 balų išankstiniame egzamine, galutinio egzamino rezultatas?

Išankstinis egzaminas 20 15 13 18 18 20 14 16

Galutinis egzaminas 92 72 72 95 88 98 65 77

4 uždutis.

Atlikdamas naminių bičių tyrimą, Seeley (2010) pastebėjo, kad bitės skautės šoka, kad praneštų bitėms, grįžusioms į lizdą, apie atstumą iki naujos lizdo vietos. Žemiau esančioje lentelėje parodytas septynių skirtingų bičių šokio trukmė (sekundėmis) ir atstumas iki naujos vietos (metrais).

Šokio trukmė (sec.)	0.40	0.45	0.95	1.30	2.00	3.10	4.10
Atstumas (metrai)	200	250	500	950	1950	3500	4300

Išanalizuokite duomenis. Panaudokite tiesinį modelį ir nuspėkite atstumą iki naujo lizdo, kai bitė šoka 1.5 sekundės.

5 uždutis. Mėnesinės trikotažo gamyklos gamybos sąnaudos ir našumas 4 metų laikotarpiu.

1	46.75	92.64	10	39.84	83.72	20	37.58	30	35.1	40	13.11
2	42.18	88.81	11	39.15	84.54		83.35		77.14		29.45
3	41.86	86.44	12	39.2	85.66	21	36.48	31	33.75	41	9.50
4	43.29	88.8	13	39.52	85.87		82.29		75.47		29.02
5	42.12	86.38	14	38.05	85.23	22	38.25	32	34.29	42	9.74
6	41.78	89.87	15	39.16	87.75		80.92		70.37		19.05

7 41.47 88.53	16 38.59 92.62	23 37.26	33 32.26	43 9.34
8 42.21 91.11	17 36.54 91.56	76.92	66.71	20.36
9 41.03 81.22	18 37.03 84.12	24 38.59	34 30.97	44 7.51
	19 36.6 81.22	78.35	64.37	17.68
		25 40.89	35 28.2	45 8.35
		74.57	56.09	19.23
		26 37.66 71.6	36 24.58	46 6.25
		27 38.79	50.25	14.92
		65.64	37 20.25	47 5.45
		28 38.78	43.65	11.44
		62.09	38 17.09	48 3.79
		29 36.7 61.66	38.01	12.69
			39 14.35	
			31.4	

Kokius statistinius klausimus galima formuluoti?

Atsakykite į iškeltus klausimus.

A PRIEDAS

5-8 METŲ MOKYMOSI PATIRTIS.

5 klasė

Duomenys ir interpretavimas.

Apibrėžiamos sąvokos: **imtis**, imties **vidurkis**. Mokomasi kelti statistinius klausimus apie artimą aplinką, į kuriuos atsakyti galima, surinkus **kokybinius** ir **kiekybinius** duomenis. Aiškinamasi, kokie galėtų būti apklausos ar anketos klausimai; mokomasi numatyti galimų atsakymų reikšmes. Išsiaiškinama, kuomet galima apskaičiuoti imties vidurkį ir kokia gautos skaitinės reikšmės prasmė. Nagrinėjamos, interpretuojamos ir tokios situacijos, kai dažnių lentelėje ar stulpelinėje diagramoje pateikiamas labai didelis duomenų skaičius.

Imtis - populiacijos dalis, naudojama statistiniam tyrimui. Pavyzdžiui, 20 atsitiktinai atrinktų gimnazijos mokinių; 1000 apklausoje dalyvaujančių respondentų ir pan.

Duomenys (duomenų rinkinys) - imties elementų tiriamo statistinio kintamojo išmatuotos reikšmės.

Kiekybiniai duomenys - duomenys apibūdinantys tiriamų objektų kiekybines charakteristikas arba duomenys, surinkti apie kiekybinius (skaitinius) kintamuosius. Jie išreiškiami skaičiais.

Kokybiniai duomenys - duomenys apibūdinantys tiriamų objekto kokybines savybes, charakteristikas arba duomenys, surinkti apie kategorinius (kokybinius) kintamuosius.

Vidurkis - duomenų rinkinio vidurkis yra duomenų rinkinio aritmetinis vidurkis, kuris gaunamas sudėjus visus reikšmės, tada padalijus iš reikšmių skaičiaus. Pavyzdžiui, duomenų rinkinio {1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 26}, vidurkis yra 5; jį rasite padalydami tame rinkinyje esančių reikšmių sumą (55) iš reikšmių skaičiaus (11).

6 klasė

Duomenys ir interpretavimas.

Mokomasi kelti statistinius klausimus, į kuriuos atsakyti galima analizuojant diskrečiuosius duomenis, pateiktus dvigubomis stulpelinėmis diagramomis, linijinėmis diagramomis. Praktikuojamasi išskirti požymį ir numatyti jo reikšmes, rūšiuoti duomenis pagal pasirinktą požymį. Išsiaiškinama, ką vadiname imties **moda**, **mediana**. Mokomasi apskaičiuoti kiekybinių duomenų vidurkį, modą ir medianą iš duomenų (dažnių) lentelės ar stulpelinės diagramos, aptariama, kuo svarbi kiekviena šių charakteristikų, kaip jos viena kitą papildo. Braižant diagramas ir duomenų lenteles, randant skaitines charakteristikas, pasitelkiamos ir skaitmeninės technologijos.

Mediana - duomenų rinkinio mediana yra sutvarkyto duomenų sąrašo centre esanti reikšmė. Tai taip pat ta reikšmė, aukščiau kurios yra tiek reikšmių, kiek ir žemiau jos. Pavyzdžiui, duomenų rinkinyje {1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 26} šeštoji reikšmė turi penkis virš jos ir penkios po ja. Taigi 3 yra mediana. Jei duomenų rinkinyje yra lyginis

reikšmių skaičius, mediana randama imant dviejų eilės sąrašo centre esančių reikšmių vidurkį.

Moda yra dažniausiai pasitaikanti reikšmė duomenų rinkinyje. Pavyzdžiui, duomenų rinkinyje {1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 26} moda yra 4, o jo dažnis yra trys. Duomenų rinkinys gali turėti daugiau nei vieną modą. Duomenų rinkinys neturės modos, jei visos jo reikšmės yra vienodo dažnio.

7 klasė

Duomenys ir interpretavimas.

Aptariamos sąvokos: **populiacija** ir **imtis**, **imties dydis**, **reprezentatyvioji imtis**, **atsitiktinumas**. Paaiškinama, kas yra atsitiktinė imties elementų atranka, kaip galima organizuoti atsitiktinę imties elementų atranką (pavyzdžiui, pasinaudoti generatoriais). Susipažįstama su įvairiais imčių sudarymo būdais: *sisteminė atranka*, *sluoksninė atranka*, *lizdinė atranka*. Aiškinamasi įvairių rūšių duomenų pobūdis, kaip praktikoje gali būti interpretuojamas duomenų rinkinių kintamumas. Nagrinėjant konkrečias situacijas, aptariami imčių sudarymo ir gautų išvadų apie jas pagrįstumo klausimai (pavyzdžiui, mokomasi nuspėti mokykloje vykstančių rinkimų nugalėtoją, remiantis atsitiktinės atrankos tyrimo duomenimis). Mokomasi duomenis pateikti skrituline diagrama ir spręsti uždavinius, kai duomenys pateikiami šios rūšies diagramomis.

Populiacija - statistinio tyrimo objektų grupė, turinti bent vieną bendrą požymį, apie kurį atliekamas statistinis tyrimas. Pvz.: populiacija – Lietuvos mokyklų mokiniai; požymis – lankantys muzikos mokyklą (balsavimo prioritetai). Populiacijas gali sudaryti žmonės, gyvūnai, įmonės, pastatai, transporto priemonės, namų ūkiai, renginiai ir pan.

Imties dydis – imties elementų skaičius.

Atsitiktinė imtis (atranka) yra tokia, kai kiekvienas populiacijos vienetas turi galimybę būti atrinktas ir šią galimybę galima tiksliai aprašyti. Atsitiktinė atranka apima paprastąją, sisteminę, sluoksninę, lizdinę.

- *Paprastoji atsitiktinė imtis* (atranka) yra tokia, kai visi imties nariai parenkami atsitiktinai ir turi vienodą galimybę patekti į imtį. Loterija yra paprastos atsitiktinės atrankos pavyzdys, kai skaičiai generuojami iš apibrėžto skaičių diapazono (pvz., nuo 1 iki 60) taip, kad kiekvienas skaičius turi vienodą galimybę būti pasirinktam.

- *Sisteminė atsitiktinė imtis* (atranka) yra tokia, kai pirmasis imties narys parenkamas atsitiktinai, o kiti imties nariai imami tam tikrais 5 intervalais (pvz., kas 4-as vienetą). Pavyzdžiui, daugelis gyventojų sąrašų būna sudaryti, išrikiavus pavardes abėcėlės tvarka. Sudarant sisteminę imtį pirmasis imties elementas išrenkamas atsitiktinai iš pirmųjų q sąrašo elementų, o likusieji imties elementai vienareikšmiškai nustatomi vis pridėdant po q prie pirmuoju išrinkto elemento numerio iki pat sąrašo pabaigos.
- *Sluoksninė (stratifikuota) atsitiktinė imtis* (atranka) yra tokia, kai nustatomi atitinkami populiacijos pogrupiai ir iš kiekvieno sluoksnio atrenkamos atsitiktinės imtys.
- *Lizdinė imtis* (atranka), yra tokia, kai visa populiacija suskirstoma į panašias pagal tam tikrą požymį grupes – lizdus (klasterius). Iš visų lizdų aibės paprastosios atsitiktinės imties būdu parenkama dalis lizdų. Į imtį pakliūna visi atrinktųjų lizdų elementai.

8 klasė

Duomenys interpretavimas.

Nagrinėjamos situacijos, kai keliami sudėtingesni statistiniai klausimai. Aiškinamasi, kaip surinkti duomenys grupuojami į vienodo ilgio intervalus. Nagrinėjant konkrečius pavyzdžius, aptariamos **histogramos**, **empirinio tankio** sąvokos. Mokomasi duomenis suskirstyti į vienodo ilgio intervalus, taip pat įvertinti, koks galėtų būti į intervalus patekusių duomenų vidurkis. Apibrėžiama **kvartilio** sąvoka. Mokomasi surasti duomenų pirmąjį, antrąjį, trečiąjį kvartilius, grafiškai pavaizduoti duomenų išsibarstymą *stačiakampe diagrama* (su „ūšais“), skaityti ir suprasti tokia diagrama pavaizduotą informaciją. Mokomasi interpretuoti duomenis, kai yra išskirčių (stipriai išsiskiriančių duomenų). Nagrinėjant praktines situacijas, aptariama, kaip apskaičiuojamas **sukaupstasis dažnis**, **sukaupstasis santykinis dažnis**. Aiškinamasi, kaip sukauptojo dažnio ir sukauptojo santykinio dažnio lentelės duomenys pavaizduojami sukauptojo dažnio ar sukauptojo santykinio dažnio diagrama, kaip skaityti ir interpretuoti tokiomis diagramomis pateiktus duomenis.

Duomenų pasiskirstymas apibūdina duomenų rinkinio formą, kai ji rodoma histogramoje. Duomenyse randama įvairių specifinių statistinių skirstinių, tačiau du iš labiausiai paplitusių yra: vienodas pasiskirstymas (vienodo dažnio intervalai) ir normalusis pasiskirstymas (varpelio formos histograma).

Kvartiliai yra skaičiai, padalijantys sutvarkytą duomenų rinkinį į keturias dalis, kurių kiekvienoje yra maždaug ketvirtadalis duomenų. Dvidešimt penki procentai duomenų rinkinio patenka į pirmąjį kvartilį (Q_1). Mediana yra antrasis kvartilis (Q_2); 50% duomenų yra prieš medianą. Septyniasdešimt penki procentai duomenų yra prieš trečiąjį kvartilį (Q_3) (3 ketvirtį).

B priedas

Detaliau žr. ⁴

Statistinis problemų sprendimas A lygyje.

Klausimo iškėlimas:

- mokytojai padeda užduoti klausimus (klausimus į mokinius dominančius kontekstus),
- mokiniai skiria statistinius sprendimus nuo fiksuoto atsakymo.

Duomenų rinkimas: mokiniai

- geba atlikti klasės surašymą,
- supranta natūralų kintamumą,
- supranta eksperimentus su nenuspėjamomis baigtimis.

Duomenų analizavimas: mokiniai

- gali lyginti individą su individu,
- gali lyginti individą su grupe,
- yra susipažinę su dviejų grupių palyginimu,
- supranta pasiskirstymo idėją,
- gali apibūdinti paskirstymą,
- pastebi ryšį tarp dviejų kintamųjų,
- naudojami pasiskirstymo ir sąryšių tyrinėjimo įrankiais, tarp kurių yra:
 - juostinė diagrama,
 - taškinė diagrama,
 - sklaidos diagrama,
 - lentelės (naudojant skaičius),
 - vidurkis, mediana, moda, diapazonas,
 - (Patikslinti, kokiais įrankiais gali naudotis 1-8 klasių mokiniai)

Rezultatų interpretavimas: mokiniai

- daro išvadas apie klasę,

⁴ Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) A Framework for Statistics and Data Science Education. National Council of Teachers of Mathematics and American Statistical Association.

- supranta, kad skirtingi rezultatai gali būti kitoje klasėje ar grupėje,
- supranta išvadų apie klasę platesnio taikymo ribotumą.

Statistinis problemų sprendimas B lygyje.

Klausimo iškėlimas: mokiniai

- pradeda kelti savo klausimus,
- kelia klausimus, susijusius su didesne nei jų klasė grupe, ir pradeda atpažinti skirtumą tarp populiacijos surašymo ir imties.

Duomenų rinkimas: mokiniai

- atlieka dviejų ar daugiau klasių surašymus,
- kuria ir atlieka neatsitiktinių imčių apklausas ir pradeda naudoti atsitiktinę atranką,
- kuria ir atlieka lyginamuosius eksperimentus.

Duomenų analizavimas: mokiniai

- praplečia savo supratimą apie duomenų pasiskirstymą,
- kiekybiškai įvertina kintamumą grupės viduje,
- lygina du ar daugiau pasiskirstymų naudodami grafinius vaizdus ir skaitines santraukas,
- naudoja sudėtingesnius pasiskirstymo apibūdinimo ir palyginimo įrankius, įskaitant:
 - histogramas,
 - IQR (tarpkvartilinį diapazoną) ir MAD (vidutinį absoliutinį nuokrypį),
 - penkių skaičių suvestines ir stačiakampes su ūsais diagramas (boxplots),
- supranta atrankos paklaidas,
- kiekybiškai įvertina dviejų kintamųjų sąsajos stiprumą, kuria paprastus modelius, kaip susieti du skaitinius kintamuosius, ir naudoja išplėstus įrankius sąryšių tyrinėjimui, įskaitant:
 - dviejų kategorinių kintamųjų lenteles,
 - laikinių duomenų grafikus,
 - QCR (Quadrant Count Ratio) kaip sąryšio stiprumo matą,
 - grafikus, dviejų skaitinių kintamųjų ryšių modeliavimui.

Rezultatų interpretavimas: mokiniai

- apibūdina skirtumus tarp dviejų ar daugiau grupių, atsižvelgiant į centrą, išsibarstymą ir profilį,
- supranta, kad imtis gali neatspindėti didesnės populiacijos,
- supranta pagrindines sąryšio matų interpretacijas,
- pradeda atskirti stebėjimo tyrimą nuo suplanuoto eksperimento tyrimo,
- pradeda skirti „sąryšį“ nuo „priežasties ir pasekmės“,
- atpažįsta imties kintamumą suvestinėje statistikoje, pavyzdžiui, imties vidurkio ir imties dydžio.

Statistinis problemų sprendimas C lygyje.

Klausimo iškėlimas:

- mokiniai turėtų gebėti suformuluoti klausimus ir nustatyti, kaip galima rinkti ir analizuoti duomenis, kad būtų galima pateikti atsakymą.

Duomenų rinkimas: mokiniai turėtų

- suprasti, kokia yra geroji atrankinės apklausos praktika,
- suprasti, kokia yra geroji eksperimento atlikimo praktika,
- suprasti, kokia yra gera praktika atliekant stebėjimo tyrimą,
- gali sukurti ir įgyvendinti duomenų rinkimo planą statistiniams tyrimams.

Duomenų analizavimas: mokiniai turėtų

- gebėti nustatyti tinkamus būdus, kaip apibendrinti skaitinius arba kategorinius duomenis naudodami lenteles, grafinius vaizdus ir suvestinę statistiką,
- suprasti, kaip atrankos skirstiniai (sukurti modeliavimo būdu) naudojami apibūdinti imties statistikos kintamumą,
- gebėti atpažinti ryšį tarp dviejų kategorinių kintamųjų,
- sugebėti atpažinti, kada ryšys tarp dviejų skaitinių kintamųjų yra tiesinis, žinoti, kad Pirsono koreliacijos koeficientas yra tiesinio ryšio tarp dviejų skaitinių kintamųjų stiprumo matas, ir suprasti mažiausių kvadratų kriterijų.

Rezultatų interpretavimas: mokiniai turėtų

- suprasti statistinio reikšmingumo reikšmę ir skirtumą tarp statistinio reikšmingumo ir praktinio reikšmingumo,
- suprasti p reikšmių vaidmenį nustatant statistinį reikšmingumą,
- gebėti interpretuoti paklaidą, susijusią su populiacijos charakteristikos įvertinimu.