

## APRAŠOMOJI STATISTIKA

Pagrindinės sąvokos

**Statistika** – keliareikšmė sąvoka. Skirtinos bent jau šios ryškios bei kartu skirtingos reikšmės:

- a) tokia duomenų apie valstybę, jos būklę, gyventojus ir t.t. pateikimo sistema ir iš jos gaunami duomenų srautai, padedantys efektyviau valdyti. Dažniausiai čia įvairiais pavidalais sutelkiama valstybinės svarbos informacija. Tokia ir šio termino genezė (iš it. *stata* - valstybė; *statista* – „valstybininkas“, t.y. tas žmogus, kuris tvarko valstybės reikalus).
- b) toks mokslas ir jo pagrindu atsirandanti metodologija, mokanti kaip racionaliai reikėtų rinkti, sistematinti, analizuoti ir pateikti duomenis. Gali atliepti pačių įvairiausių tyrinėjimo sričių (ekonomikos, biologijos, medicinos, fizikos ir t.t.) poreikiams. Adaptabili ir adaptuotina prie tos srities, kuriai taikoma. Ta prasme atsiranda kaip ir konkrečių mokslo ar taikymo sričių „statistikos“, pvz., ekonominė statistika. Čia pritaipytą ir mūsų kalboje, filologams aktuali *kalbos statistika*.
- c) konkrečiais skaičiavimais nustatytas tiriamo požymio įvertinimas imtyje (plg. pozicinė statistika).

**Populiacija** (kitaip – generalinė visuma, generalinė aibė) – visa objektų, kurie statistiškai tiriami, visuma. Būna baigtinė (tada iš principo galima ištirti visus jos objektus) ir begalinė (nebaigtinė; tada visų jos objektų ištirti apskritai neįmanoma).

**Imtis** (kitaip – praba, atranka) – tyrimui atrinktų ir ištirtų objektų aibė. Paprastai tai būna populiacijos dalis, o jei-ku ištiriama visa baigtinė populiacija, tai imtis ir populiacija sutampa, nebesiskiria.

**Požymiai** (kitaip – kintamieji) – tyrinėtoją dominantys populiacijos bei imties objektų ypatumai, kurių konkrečios reikšmės kinta (įvairuoja).

**Stebėjimas** (arba statistinis eksperimentas) – tiriamųjų požymių (iš)matavimas arba (su)skaičiavimas imtyje ir jų registravimas. Tai – pirminis statistinio pobūdžio informacijos šaltinis, todėl nuo jo priklauso viso tyrimo efektyvumas bei objektyvumas.

Statistikos atmainos (rūšys)

Galima išskirti kelias statistikos, kaip metodologinio pobūdžio mokslo, atmainas arba rūšis. Minėtinos:

- a) **Aprašomoji** (arba deskriptyvinė) statistika. Jos sritis – duomenų sisteminimo ir jų grafinio pateikimo metodai, pritaikyti konkrečiai mokslinių tyrimų sferai. Čia vyrauja faktografija, faktologija ir konstatuojamojo pobūdžio teiginiai. Aktuali kiekvienai statistika besiremiančiai mokslo sričiai, kartu – ir filologijai.
- b) **Indukcinė** (ar – induktyvinė? Kartais sakoma - išvadų) statistika. Jos sfera – metodai, leidžiantys pagal prabos duomenis daryti išvadas apie visą tiriamųjų objektų populiaciją. Šios išvados visada yra tikimybinio pobūdžio ir pasižymi tam tikru tikėtinumo laipsniu. Tikslas – parinkti kuo didesnį tikėtinumo laipsnį (patikimumą) užtikrinančius metodus.
- c) **Lyginamoji** statistika. Didžiumoje knygų ir vadovėlių ji atskirai neišskiriama, bet filologams yra gana aktuali. Sfera – statistinių metodų, leidžiančių palyginti dvi ar kelias objektų, turinčių tuos pačius tiriamuosius požymius, imtis, parinkimas ir taikymas. Tikslas - nustatyti, ar tiriamosios imtys iš esmės yra vienodos ir požymių reikšmės jose įvairuoja (varijuoja) atsitiktinai, dėl atsitiktinės poveikio, ar skiriasi esmingai. Išvados čia būna irgi tikimybinio pobūdžio, tad tikslinga orientuotis į kuo didesnį jų patikimumą (tikėtinumą).

# I. Bendrasis supratimas

## 1. Eksperimentas statistikoje

Žodžiui *eksperimentas* šiame kurse irgi suteikiama kiek kitokia prasmė, negu ta, kuria jis vartojamas kasdieninėje kalboje. Mes *eksperimentą* labiausiai vartosime ta prasme, kuria jis suprantamas tikimybių teorijoje ir matematinėje statistikoje.

**Eksperimentas** Bet koks rūpimų (tiriamųjų) objektų, mūsų atveju – dažniausiai kalbinių, ar jų požymių (su)skaičiavimas arba (iš)matavimas, kurio metu gaunami tolesnio apdoravimo bei aiškinimo reikalaujantys duomenys.

Atskiri matavimai ar tiesiog tiriamųjų objektų bei rūpimų jų požymių užfiksavimai dar vadinami įvairiai: *bandymais*, *stebėjimais* ir pan. Savaime suprantama, jog iš jų kartojimo, didesnės ar mažesnės jų serijos ir „susidaro“ eksperimentas.

Natūralu, kad su „tikrais“, įprastinę to žodžio prasmę atitinkančiais eksperimentais tiesiogiai susiduriama instrumentiniuose kalbos signalo tyrinėjimuose, dažnai ir vadinamuose tiesiog *eksperimentine* fonetika. Tačiau šia specifine prasme ir vienkiių ar kitokių teksto elementų rinkimas bei suskaičiavimas, ir kalbos dalykus liečiančios anketinės apklausos taip pat yra „eksperimentai“.

Paprastai stebėjimų metu dėmesys kreipiamas *ne į visas* tiriamųjų objektų savybes bei ypatybes, o tik į *kai kurias* iš jų, į tas, kurias šiuo konkrečiu atveju norima iširti. Tokių *tiriamųjų* ypatybių gali būti ir tiksliai viena, ir kelios iškart – nelygu ko kuriuo tyrimu siekiama ir kaip projektuojamas tyrimui turintis padėti statistinis eksperimentas. Šios savybės ar ypatybės nebūtinai yra pačios svarbiausios, ryškiausios ar kaip kitaip apskritai dominuojančios tiriamųjų objektų ypatybės: svarbu tik, kad jos būtų tikrai *būdingos* tiriamiesiems objektams ir kad būtų pagrindo iškelti vienokias ar kitokias *hipotezes* apie jų *svarbą* tiriamuoju požūriu. Objektų savybės bei ypatybės, kurias norima iširti ir kurioms išryškinti atliekami statistiniai eksperimentai, dažnai vadinamos *tiriamaisiais požymiais* ar tiesiog *požymiais*.

**Požymis (iai)** Tiriamųjų objektų savybė(s) ar ypatybė(s), į kurią (as) nukreipiamas tyrimo dėmesys ir kurias siekiama užfiksuoti tyrimui skirtais eksperimentais.

Tiriamuoju požymiu iš principo gali tapti bet kuri bet kurio lingvistinio objekto (garso, skiemens, žodžio, sakinio ir t. t.) savybė ar ypatybė, bet praktiškai jais dažniausiai tampa tos iš jų, kurios yra labiausiai „įtartinos“, t. y. kelia tyrinėtoji daugiausiai spėlionių ir prielaidų.

Požymius, kadangi jie gali būti labai įvairūs, galima įvairiai grupuoti, sisteminti ir klasifikuoti. Labai dažnai yra skiriami *kiekybiniai* ir *kokybiniai* požymiai. Kitu gi atžvilgiu požymius priimta skirstyti į *diskrečiuosius* ir *tolydžiuosius*.

Požymiai:

<b>kokybiniai:</b>	Paprastai apibūdinami juos <i>įvardijant</i> , t. y. pavadinant ar specialiais moksliniais terminais, ar kasdieninės kalbos žodžiais. Fiksuojant priimta ženklinti <i>sutartiniais ženklais</i> (emblemomis, etiketėmis, simboliais, kurių funkciją gali atlikti net ir <i>skaitmenys</i> – tik tokiu atveju jie nėra <i>skaičiai</i> , nes nereiškia dydžio, <i>kiekio</i> , ir dažnai nėra loginio pamato su jais atlikti aritmetinius veiksmus). Pvz., žodžių priklausomybė kalbos dalims, sintaksinės žodžių funkcijos (priklausymas sakinio dalims), darybinės kategorijos, kaitybinės formos ir t.t.
<b>kiekybiniai:</b>	Lengvai <i>matuojami</i> , t. y. įvertinami vienokiais ar kitokiais matais ar skaičiavimo vienetais. Matavimo rezultatai paprastai išreiškiami ir užrašomi <i>skaičiais</i> . Pvz., žodžio ilgis garsais (fonemomis), raidėmis ar skiemenimis, sakinio ilgis žodžiais, pastraipų ilgis sakiniais, garso trukmė milisekundėmis, pagrindinio tono aukštis hercais, į anketos klausimą vienaip ar kitaip atsakusių respondentų kiekiai ir pan. Su šiais dydžiais, kiekybiniais požymių įvertinimais, galima atlikti ir aritmetinius veiksmus.
<b>diskretieji:</b>	Tokie, kurių galimų reikšmių skirtumai iš principo negali būti mažesni už tam tikrą „minimumą“. Dažnai tas slenkstis ar žingsnis, minimalus galimų reikšmių skirtumas, būna lygus 1 (pvz., vaikų skaičius šeimoje, žodžio ilgis skiemenimis ar raidėmis, sakinio ilgis žodžiais), tačiau – nebūtinai (sakysim, medicininis termometras fiksuoja 0,1 laipsnio kūno temperatūros pokyčius). Lingvistikoje diskretieji požymiai itin dažni. Kartais diskrečiaisiais laikomi taip pat ir <i>kokybiniai</i> požymiai.

<b>tolydieji:</b>	Jų galimų reikšmių skirtumai iš principo gali būti kiek tik norint maži, todėl jų neįmanoma absoliučiai tiksliai užrašyti baigtiniais skaičiais (pvz., matematinės konstantos $\pi$ , $e$ ), o taip pat neįmanoma suskaičiuoti visų jų potencialių reikšmių, nes jų yra be galo daug, ir tos reikšmės viena į kitą pereina <i>laipsniškai</i> , nenutrūkstamai. Todėl ir juos <i>išmatuoti</i> iš principo teįmanoma <i>tiktai apytiksliai</i> , vienokiu ar kitokiu tikslumu. Matuojant šių požymių reikšmės neišvengiamai <i>diskretizuojamos</i> pagal pasirinktuosius mato vienetų: kuo šie mažesni, tuo ta diskretizacija subtilesnė ir tuo tiksliau atspindima tolydiška požymio prigimtis, tačiau pati jų diskretizacija matuojant išlieka visuomet. Pvz., garso trukmė, intensyvumas, pagrindinio tono ir formančių (harmonikų) dažniai ir pan.
-------------------	---

Kartais dar atskirai yra minimi tarpinę padėtį tarp kiekybinių ir kokybinių užimantys požymiai, vadinami *ranginiais*. Jie turi ir kiekybiniam, ir kokybiniam požymiams būdingų bruožų. Ranginių požymių pavyzdžiai galėtų būti kokiose nors varžybose (sporto rungtynėse, meno kolektyvų apžiūrose, gražuolių konkursuose) užimtos vietos, žinių įvertinimas balais ir pan.

Šiuolaikinėse statistikos knygos *išmatuotos* požymių reikšmės (t.y. matavimo metu gauti rezultatai) itin dažnai vadinami *kintamaisiais*, tuo tarsi specialiai pabrėžiant, kad kiekvieno matavimo atveju galima vis kitokia, kintanti to paties požymio reikšmė. Be to, linkstama akcentuoti ne tiek pačių požymių skirstymą į kiekybinius ir kokybinius, kiek jų reikšmių matavimui taikomų *skalių* pobūdį ar tipologiją (plg., pvz., V. Čekanavičius, G. Murauskas. Statistika ir jos taikymai I, p. 17–20). Skiriamos 4 tipų skalės: pavadinimų arba nominalinė, rangų, intervalų ir santykių.

**Pavadinimų** (nominalinė) skalė leidžia požymio reikšmes tik suklasifikuoti ir įvardinti, bet išvis neperteikia „kiekybinio matmens“. Todėl ji taikoma tik kokybinių požymių reikšmėms apibūdinti.

**Rangų** skalė perteikia tik patį požymio reikšmės padidėjimo ar sumažėjimo viename tiriamajame objekte, lyginant jį su kitu objektu, faktą, bet neperteikia tikslaus jos kiekio įvertinimo. Tinka ranginių kintamųjų reikšmėms apibūdinti, o šia skale išreikštas požymio reikšmes tegalima tik surikiuoti, išdėstyti didėjimo ar mažėjimo tvarka.

**Intervalų** skalė taikoma kiekybiniam požymiams ir leidžia nustatyti, *kiek* daugiau ar mažiau to požymio rasta viename tiriamajame objekte, palyginus su kitu. Tačiau nulis čia parenkamas „laisvai“, todėl nulį atitinkanti požymio reikšmė čia nereiškia, kad šiuo atveju šio požymio iš viso nėra (pvz., oro temperatūra, lygi  $0^{\circ}\text{C}$ , nereiškia temperatūros nebuvimo apskritai). Matuojamas tokia skale kiekybinis požymis gali įgauti tiek teigiamas, tiek ir neigiamas reikšmes.

**Santykių** skalė irgi taikoma kiekybinių požymių reikšmėms įvertinti ir pasižymi tuo, kad nulis joje yra „absolutus“, rodantis matuojamojo požymio nebuvimą. Todėl ja matuojami požymiai tegali įgyti tik teigiamas reikšmes (pvz., žmogaus amžius, sakinio ilgis, garso trukmė).

Eksperimentų metu gauti bei sukaupti (surinkti, užfiksuoti) tiriamųjų požymių įvertinimai lingvistui, kaip ir bet kurios kitos srities tyrinėtoji, yra *pirminiai statistiniai duomenys*. Jie – ne tik pats svarbiausias visokių tolesnių tyrimų bei apibendrinimų pamatas, bet taip pat ir savotiška medžiaga („žaliava“) tolesnei statistinei analizei. Todėl labai svarbu yra lingvistinius eksperimentus planuoti, projektuoti ir atlikti kaip galima korektiškiau: jeigu tik iš principo įmanoma, būtina iširti *pakankamai didelį* (dažnai sakoma – reprezentatyvų, t. y. statistiškai pakankamą) tirtinųjų objektų kiekį. Atrinkti objektus tyrimui reikia *asitiktinai*. Požymius matuoti ar skaičiuoti dera kaip galima atidžiau, reikia vengti galimų jų matavimo ar skaičiavimo riktų, rezultatų užrašymo klaidų ir pan. Atsimintina, kad apskritai *nėra* jokių būdų ar gudrybių, kurios iš *klaidingų* pirminių duomenų leistų gauti *teisingus* rezultatus ar daryti pagrįstas išvadas. Todėl visur, kur tikta įmanoma, ir pirminių duomenų rinkimui reikėtų panaudoti kompiuterį: klaidoms liks kur kas mažiau galimybių.

## 2. Imtis ir populiacija (generalinė aibė)

**Imtis** (kartais pasakoma ir *praba*) yra viena iš svarbiausių statistikos kategorijų. Tuo žodžiu šiaipjau vadinama eksperimentui parinktų (paimtų) ir jo metu *ištirtų objektų visuma*. Savaime suprantama, jog imtys viena nuo kitos pirmiausiai skiriasi pagal tai, kiek objektų kurio eksperimento metu iširta, ir todėl vienas iš svarbesniųjų bet kokią imtį apibūdinančių parametrų yra jos didumas arba, kaip labiau įprasta sakyti, *tūris*.

**Imties tūris** Bendras iširtų objektų kiekis (tuo pačiu – ir atliktų bandymų skaičius). Paprastai žymimas raide  $n$ . Jis kartu atitinka ir eksperimento metu gautų tiriamojo požymio įvertinimų (matavimų) bendrąjį kiekį.

Ta tiriamųjų objektų visuma, iš kurios jie atrenkami (paimami) į imtį, vadinama *generaline aibe* arba – naujesnėse statistikos knygos – *populiacija*. Ji gali būti baigtinė (ribota) arba begalinė. Savaime suprantama, jog baigtinės populiacijos (generalinės aibės) atveju lieka bent jau principinė galimybė iširti *visus* ( $N$ ) jai priklausančius objek-

tus, ir imties sąvoka tokiu atveju kaip ir nebetektų prasmės, nes imtis faktiškai sutaptų su populiacija. Bet jeigu populiacija yra begalinio dydžio, nebaigtinė, tai ištirti visų jos objektų niekaip nebeįmanoma, ir imtis turi būti formuojama (sudaroma) būtinai. Panašiai būna ir tada, kai populiacija yra baigtinė, bet labai skaitlinga, sudaryta iš didžiulės daugybės objektų: ištirti visus jos objektus būtų itin keblu, o dažnai – irgi net neįmanoma, tad parankiau būna verstis pakankamo dydžio, bet kur kas mažiau objektų turinčia imtimi. Imtis iš tiesų yra ne kas kita, kaip savotiškas, dirbtinai susikurtas populiacijos „mini modelis“ ar „mini atvaizdas“, jos eksperimentinis atitikmuo, dirbtinis analogas. Čia ir yra tikroji statistikos stichija: tirti palyginti nedidelę, ribotą imtį, o tyrimo rezultatus tikimybiskai apibendrinti visai populiacijai. Todėl visų svarbiausia yra, kad imtis populiaciją, iš kurios buvo paimta, atspindėtų kaip galima *adekvačiau* ir teisingiau. Paprastai sakoma, kad didesnio tūrio imtis esanti *reprezentatyvesnė*, didėjant imčiai didesnė, artimesnė vienetui darosi *tikimybė*, kad šioje imtyje rastos požymių reikšmės bei jų pasiskirstymas (pasklidimo, sklaidos po objektus ypatumai) iš esmės būtų tokios pačios, jeigu paimtume ir kitus analogiškus populiacijos objektus, šį kartą į imtį nepatekusių. Tačiau svarbu pabrėžti, kad imties reprezentatyvumą apsprendžia ne vien tik jos tūris (didumas), bet labiausiai – jos sudarymo principai. Imties tūris turi būti pakankamas, o praktiškai tai reikštų, kad imtyje turėtų, jeigu tai įmanoma, būti nuo kelių dešimčių iki kelių šimtų objektų. Dar daugiau: pernelyg didelės, daugiataktantinės imtys tampa savaip problemiškos, nes neretai dėl grynai formalių, matematinių priežasčių tuomet gali būti „sureikšminami“ ir iš tikrųjų menki, nežymūs skirtumai.

Jeigu, esant galimybėms, ištiriami *visi* populiacijos ar populiacijų objektai (sakysim, visi rūpimo teksto ar tekstų grupės žodžiai, sakiniai ir pan.), tai, viena vertus, smarkiai susiaurėja daugelio tradicinių statistikos metodų taikymo erdvė (nes nebelieka reikalo apie populiaciją spręsti iš jos „sumažinto“ atvaizdo imtyje), o kita vertus – gaunami nebe tikimybiskai apibendrinti, bet visiškai *tikslius* rezultatai, kuriuos belieka tik korektiškai aprašyti ir pateikti.

Imties parinkimas, konkrečių objektų „patekimas“ arba „nepatekimas“ į ją – atskira problema, nors čia plačiau ir nesvarstoma, tačiau kelianti itin daug tiek „teorinių“, tiek ir praktinių klausimų. Įsidėmėtina, kad matematinės statistikos vadovėliai labai dažnai pabrėžia, jog imtis turinti būti *atsitiktinė*: bet kuris konkretus tiriamosios rūšies objektas turi turėti vienodas galimybes papulti į imtį, o ar bus jis iš tikro atrinktas, ar nebus – turinti lemti vien tiktai *atsitiktinybė* (kartais tai atsitiktinybei užtikrinti iš tiesų pasinaudojama kokiais nors patikimais atsitiktinybės šaltiniais – pvz., urnomis, atsitiktinių skaičių lentelėmis ir pan.). Palyginti retais ir specifiniais atvejais yra galimos ir kitaip, ne atsitiktinumo pagrindu sudarytos (neatsitiktinės) imtys.

Imtin atrinktus objektus itin pravartu sunumeruoti, kad prirėikus visuomet būtų galima vienareikšmiškai atkurti pirminę jų seką (t.y. išrikiuoti juos pagal eilės numerius). Savaime suprantama, jog eilės numeriai turėtų būti fiksuojami kartu su išmatuotomis tiriamojo požymio reikšmėmis.

Specialiai pabrėžtina, kad tendencingos ar falsifikuotos imtys populiacijos „vaizdą“ visuomet iškreipia, perteikia jį neadekvatų ir deformuotą ir todėl jokios mokslinės vertės ar įrodomosios galios neturi: jos tėra tiesiog falsifikatai. Tuo pačiu atskira kalba būtų apie pasitikėjimą statistika (kartais statistika apibūdinama kaip tam tikra [V. Čekanavičiaus ir G. Murausko kn. „Statistika ir jos taikymai“ sako „trečioji“, žr. p. 60–62] melo rūšis; pats esu girdėjęs taip sakant M. Gorbačiova). Mat, svarbu ne vien patys duomenys, bet ir tam tikras jų „kontekstas“: sakysim, ką reiškia „objektyvus“ teiginys „Šiandien yra *dešimt* laipsnių Celsijaus“ - ar šilta, ar šalta?

Sausio 15 d. – šilta, netgi labai šilta

Balandžio 2 ar spalio 23 d. – normalu, nei šilta, nei šalta

Birželio 27 d. – šalta, galbūt net labai šalta

Ištyrus imtin patekusių objektus gaunami tiriamąjį požymį (ar – požymius) apibūdinantys *statistiniai duomenys*.

Išrikiuoti tokia tvarka, kokia buvo atliekami bandymai (matavimai) su imties objektais, jie dar nėra pakankamai vaizdūs ar iškalbūs, todėl paprastai būna tvarkomi ir perdirbinėjami toliau.

### 3. Imties duomenų pirminis sutvarkymas: statistinė ir variacinė eilutės

Duomenys, surašyti tokia tvarka, kokia jie buvo gauti eksperimento metu, sudaro vadinamąją *statistinę* eilutę. Statistinė eilutė, t.y. pirminių duomenų tvarka, atitinkanti imties (ar populiacijos, jeigu buvo ištirta visa populiacija ištiesai) objektų „tvarką“, yra irgi svarbi, ypač – kai tenka specialiai tikrinti, ar imtį (požymio reikšmių seką) galima laikyti esant atsitiktine.

Paprasčiausias koki nors tiriamąjį požymį apibūdinančių statistinių duomenų sutvarkymo atvejis – jų išdėstymas didėjančia tvarka, kur žiūrima jau nebe pačių bandymų eilės (jų *numeriu*), bet požymį įvertinančių reikšmių didumo. Surašyti šitokiu būdu, jie sudaro *variacinę eilutę*, kurios ilgis, savaime aišku, atitinka imties tūrį. Natūralu, kad į variacinę eilutę lengviausia yra „išrikiuoti“ *kiekybinių* požymių įvertinimus. Kokybinių gi požymių variacinės eilutės visuomet yra daugiau ar mažiau *sąlygiškos*: tokie požymiai, kaip minėta, iš esmės tik įvardijami, tad nėra tikro pamato vienas kokybinio požymio laikyti didesnėmis ar mažesnėmis už kitas ir nebelieka kaip nustatyti jų didėjančios tvarkos.



*pasiskirsčiusios* požymio reikšmės, kaip jos pasisklidusios, *kokia dalis* (procentas) kuriai reikšmei imtyje tenka. Vėliau susidursime ir su teoriškai sukonstruotais, t. y. labiau virtualiais, „idealizuotais“ pasiskirstymais, turinčiais „individualius“ vardus.

Kadangi iš esmės *visi* statistinio pobūdžio tyrimai be pasiskirstymų neišsiverčia ir vienaip ar kitaip į juos orientuojasi, pasiskirstymams vaizduoti, jų savybėms nagrinėti ir kt. panašioms dalykams skiriama ypač daug dėmesio. Pasiskirstymus galima vaizduoti įvairiai: ne tik *variacinėmis eilutėmis*, kaip pateiktoji, bet ir įvairiomis *lentelėmis*, *grafikais*, analitiniu būdu (*formulėmis*) ir t. t.

Nors pateiktosiose eilutėse esanti informacija visiškai pilnai apibūdina tirtojo požymio reikšmių pasiskirstymą (t. y. sklaidą, tų reikšmių proporcijas) imtyje, kartais pasiskirstymų lentelės daromos dar išsamesnės: jose būna ir papildomų žinių apie pasiskirstymą. Mūsų pasiskirstymas, pateiktas tokia išplėtota lentele, galėtų atrodyti maždaug taip:

<i>Žodžio ilgis skiemenimis:</i>		1	2	3	4	5
Dažnumas	<i>Absolutus:</i>	12	33	24	4	2
	<i>Santykinis:</i>	0.16	0.44	0.32	0.053	0.027
<i>Pasiskirstymo funkcija:</i>		0.16	0.60	0.92	0.973	1.00

Kaip matome, pačių reikšmių (1, 2, 3, 4, 5), kurias įgyja tiriamasis požymis – žodžio ilgis skiemenimis, – čia atsispindi abi dažnumo atmainos – ir absoliutinis, ir santykinis dažnumas, o paskutiniojoje eilutėje yra pateikiamas dar ir *sukaupstasis* arba *kumuliatyvinis* dažnumas, rodantis, koku mastu požymio reikšmės, palaipsniui besikaupdamos, „užpildo“ imtį, t. y. kokią jos dalį sudaro pirmoji, pirmoji ir antroji kartu paimtos, pirmoji, antroji ir trečioji kartu paimtos ir t. t. Sukaupstasis dažnumas yra nepaprastai svarbus: jeigu jis žinomas (duotas), tai iš jo lengva elementariais aritmetikos veiksmais apskaičiuoti ir santykinį, ir absoliutinį kiekvienos iš reikšmių dažnumą. Todėl pasiskirstymui adekvačiai pavaizduoti pakaktų pateikti tiksliai pirmąją ir paskutinąją, ketvirtąją tos lentelės eilutes, o antrosios ir trečiosios eilučių duomenis yra galima, jei jų prireiktų, išvesti (apskaičiuoti) iš anų ir tikrai *naujos informacijos* jie iš tikrųjų neduoda! Dėl tokių savybių ir dėl ypatingos svarbos sukaupstasis dažnumas paprastai yra vadinamas tiesiog *pasiskirstymo funkcija*. Mes su ja susidursime dar daugelį kartų, ir jos reikės visuomet, kuomet nagrinėsime kokius nors pasiskirstymus. Pabrėžtina, jog empirinių pasiskirstymų funkcija yra ne kas kita, kaip sukaupstasis dažnumas, o teorinių – sukauptoji (kumuliatyvinė) tikimybė. Todėl pasiskirstymo funkcija įgyja reikšmes irgi tik iš intervalo [0; 1].

Tad „minimaliomis priemonėmis“ adekvačiai pavaizduoti pasiskirstymą galima dviejų eilučių (kvazi)lentele: vienoje jos eilutėje surašomos reikšmės, kurias įgyja tiriamasis požymis, o antroje – toms jo reikšmėms atliepančios pasiskirstymo funkcijos reikšmės.

O kaip pasiskirstymą „suformuoti“ tais atvejais, kai požymis yra tolydusis ir imties objektams būdingos jo reikšmės praktiškai nesikartoja? Tokiais atvejais visas variacinėje eilutėje atsispindintis reikšmių ruožas – nuo mažiausios iki didžiausios – suskaidomas į keletą ar keliolika intervalų ir žiūrima, kokie yra į tuos intervalus (padalas) papuolančių požymio reikšmių kiekiai bei dažnumai. Tokie pasiskirstymai, besiremiant pavienėmis diskrečiomis požymio reikšmėmis, bet tolydžiųjų reikšmių intervalais, vadinami intervaliniais pasiskirstymais.

## 2. Imtį (statistinę ar variacinę eilutę) apibūdinantys parametrai (charakteristikos)

Jų yra įvairių, ir visi jie yra skaičiai, apskaičiuojami iš imties ar populiacijos (jei ištiriama visa populiacija) duomenų, tiksliau – iš imtyje (ar populiacijoje) nustatytų požymio reikšmių. Todėl savaime suprantama, kad imtį apibūdinančius parametrus apskaičiuoti galima tik tada, kai tiriamieji požymiai yra *kiekybiniai* ir imtyje randamos jų reikšmės išreiškiamos skaičiais. Patikimiausios yra santykių skale išreikštos kiekybinių požymių reikšmės. Dažnai skiriamos tokios tų parametru grupės ar atmainos:

- duomenų padėtį apibūdinančios charakteristikos (parametrai): *vidurkis*, *moda*, *mediana*, *kvantiliai* (kvartilai ir kitokie kvantiliai)
- duomenų sklaidą apibūdinančios charakteristikos: *dispersija*, *standartinis* (kitaip – vidutinis kvadratinis) *nuokrypis*, *linijinis nuokrypis*, *variacijos žingsnis* (plotis), *variacijos* (kitaip – imties kitimo) *koeficientas*, *kvartilų skirtumas* *IQR*, *kokybinės įvairovės indeksas* ir kt.
- pasiskirstymo formą apibūdinančios charakteristikos: *asimetrijos koeficientas* ir *eksceso koeficientas*. Jos reikalauja supratimo apie normaliąją (Gauso) kreivę.

Galima įsidėmėti kai kurių imtį apibūdinančių parametrų apibrėžimus, įsiminti jų pavadinimus; pvz.:

<b>Variacijos žingsnis</b>	<i>Didžiausios ir mažiausios požymio reikšmių variacinėje eilutėje skirtumas. Ankstesniaime pavyzdyje variacijos žingsnis (pažymėkime jį <math>v</math>) būtų: <math>v = 5 - 1 = 4</math>.</i>
<b>Moda</b>	<i>Dažniausiai variacinėje eilutėje pasitaikanti požymio reikšmė. Mūsų variacinėje eilutėje – tai 2. Priklausomai nuo to, kelios požymio reikšmės imtyje vienodai dažnos, galima skirti unimodalius, bimodalius ir polimodalius pasiskirstymus. Kai dažniausios bimodalaus pasiskirstymo reikšmės eina pagret, jis laikomas unimodaliu pasiskirstymu ir jo moda apskaičiuojama kaip tų reikšmių vidurkis.</i>
<b>Mediana</b>	<i>Viduriniojo (centrinio) variacinės eilutės įrašo, dalinančio ją pusiau, reikšmė, kitaip tariant, ji yra <math>n/2</math> - toji pozicinė statistika. Kadangi mūsų pavyzdžio variacinėje eilutėje iš viso yra 75 įrašai, tai vidurinysis (centrinis) iš jų yra 38-asis: prieš jį eina 37 įrašai, ir tiek pat – po jo. 38-ojo įrašo (reikšmė yra 2, vadinasi, ir tos variacinės eilutės mediana yra 2.</i>

Variacinės eilutės, sudarytos iš *nelyginio* įrašų skaičiaus, visada būna medianos požiūriu „neproblemiškos“, nes turi *vieną centrinį* įrašą, kurio atžvilgiu visi ankstesni ir tolesni įrašai pasidalina po lygiai. To įrašo reikšmė ir yra mediana. Kiek kebliau, jeigu įrašų skaičius variacinėje eilutėje būna *lyginis*: tokios eilutės centre atsiduria *du gretimi įrašai* (mat, kitu atveju likusieji negalėtų pasidalinti po lygiai). Mediana lyginį įrašų kiekį turinčios eilutės atveju randama taip: *sudedamos abiejų „centrinųjų“ įrašų reikšmės ir gauta jų suma dalinama iš dviejų.*

Yra du *patys svarbiausi* kiekybinio požymio reikšmės apibūdinantys parametrai: **vidurkis** ir **standartinis** (kitai – vidutinis kvadratinis) **nuokrypis** (arba artimas jo atitikmuo – **dispersija**). Tad apie juos – atskirai.

### 3. Vidurkiai

Statistikoje yra žinomos minimos kelios vidurkių atmainos: aritmetinis, geometrinis, harmoninis, nupjautasis... Bet pats populiariausias ir labiausiai naudojamas yra pirmasis iš paminėtųjų – aritmetinis vidurkis. Todėl jam – daugiausia dėmesnio.

Derėtų skirti **teorinį** aritmetinį vidurkį (rus. *matematičeskoe ožidanije*, angl. *mean* ar *estimation* liet. kartais irgi pasakoma *matematinė viltis* - aklas vertinys iš rusų k.), apskaičiuojamą pagal atitinkamas formules teoriniams („idealizuotiems“) pasiskirstymų modeliams, ir **empirinį** aritmetinį vidurkį (rus. *arifmetičeskoe srednee*, angl. *average*). Kadangi aritmetinis vidurkis ‘linksniuojamas’ visų dažniausiai, tai apibūdinamas *aritmetinis* paprastai praleidžiamas ir sakoma tiesiog *vidurkis* (savaime suprantama, kad kalbant apie kitokius vidurkius, jų konkrečią atmainą apibūdinantys žodžiai – harmoninis, geometrinis ir pan. – būtinai pridedami). Patikslinimai *teorinis* arba *empirinis* paprastai pridedami irgi tik norint aktualiai pabrėžti, apie kokio tipo vidurkį – teorinį ar empirinį – kalbama. Prisimenant gi imties ir populiacijos priešpriešą, derėtų taip pat skirti *imties vidurkį* (tai tas pat, kas empirinis vidurkis) ir *populiacijos vidurkį*, kuris gali būti arba nustatomas empiriškai (kai ištiriama visa baigtinė populiacija), arba tegalimas prognozuoti teoriškai (kai populiacija nebaigtinė arba kai tiriama tik iš jos atrinkta imtis).

Tyrinėtojų pats aktualiausias vis dėlto lieka imties empirinis vidurkis.

#### **Empirinis vidurkis**

Tai tarsi koks visų imtyje randamų požymio reikšmių, pasiskirsčiusių vienokiu ar kitokiu būdu, *svorio centras*, vidutiniškoji, visos imties mastu imant, požymio reikšmė, „vidutinis“ jos įvertinimas. Gaunamas visų įvertinimų sumą dalijant iš imties tūrio; taigi, jis irgi yra *santykis*, – visos įvertinimų (reikšmių) visumos santykis su ištirtųjų objektų kiekiu.

Formulės empiriniam vidurkiui apskaičiuoti pateikiamos ir gali būti taikomos įvairios; renkantis konkretų jo skaičiavimo būdą (algoritmą) pravartu pasiremti net ir praktiniu protu, nes vidurkis – dydis, dažnai pasitaikantis ir praktiniame gyvenime, kasdieniniuose reikaluose. Būtinai žinoti dalykai yra du: **visų** tiriamojo požymio reikšmių (įvertinimų) **suma** (visuma) ir **bendras bandymų skaičius** (imties tūris). O konkreti formulė labiausiai priklauso nuo to, kokių būdu yra pateikta tiriamojo požymio pasiskirstymą reprezentuojanti statistinė informacija:

<i>Jeigu turime:</i>	<i>Vidurkis skaičiuojamas taip:</i>
Pirminio pavidalo (nesuglausta) variacinę eilutę	Susumuojamos visos variacinėje eilutėje įrašytos požymio reikšmės (įvertinimai) ir gautoji suma dalinama iš imties tūrio ( $n$ ). Taip apskaičiuojamas vidurkis kartais vadinamas paprastuoju (nesvertiniu) vidurkiu: mat, kiekvienos reikšmės (bandymo) „svoris“ čia yra pastovus ir lygus 1.

Suglaustą variacinę eilutę su absoliutiniais dažnumais	Kiekviena požymio reikšmė dauginama iš atitinkamo absoliutinio dažnumo, sandaugos susumuojamos ir gautoji suma dalinama iš imties tūrio. Ankstesniame pavyzdyje būtų: $(1 \times 12 + 2 \times 33 + 3 \times 24 + 4 \times 4 + 5 \times 2) / 75 = 2.347$
Suglaustą variacinę eilutę su santykiniais dažnumais	Kiekviena požymio reikšmė dauginama iš atitinkamo santykinio dažnumo ir sandaugos susumuojamos. Gautoji suma ir yra empirinis vidurkis. Ankstesniame pavyzdyje: $1 \times 0.16 + 2 \times 0.44 + 3 \times 0.32 + 4 \times 0.053 + 5 \times 0.027 = 2.347$

Antruoju ir trečiuoju būdu apskaičiuojami vidurkiai kartais vadinami svertiniais, o dažnumai, iš kurių dauginamos požymių įgyjamos reikšmės, – svoriais.

Skaičiuojant vidurkį su programa *Excel* jokio pirminio duomenų apdorojimo ar grupavimo nereikia, tiesiog pirminiai matavimo rezultatai (pirminiai statistiniai duomenys, „statistinė eilutė“) turi būti *Excel* lakšte surašyti į vieną bloką (paprastai – stulpelio ar eilutės pavidalo, nors tai ir nėra privalu). Vidurkiui apskaičiuoti skirta *Excel* funkcija:

[ = ] AVERAGE(skDuomenųBlokas)

Pridurtina, kad diskrečiųjų požymių vidurkis yra iš esmės *abstraktus* pobūdžio dydis, tinkantis labiausiai įvairiems palyginimams bei sugretinimams, bet realiai dažniausiai nesutampantis *nė su viena* iš imtyje turimų požymio reikšmių. Todėl kartais jis atrodo net kiek nelogiškai. Sakome: žodžių ilgio (skiemėmis) vidurkis yra 2,347 skiemens, šeimoje vidutiniškai yra 2,25 vaiko ir pan., bet juk negali būti žodžių, sudarytų iš 2,347 skiemens, negali kas nors turėti 2,25 vaiko! Tuo atžvilgiu logiškesnė būtų, sakysim, moda.

Aritmetinis vidurkis yra labai jautrus vadinamosioms ekstremalioms – pačioms didžiausioms bei pačioms mažiausioms – požymio reikšmėms: net ir pavienė tokia reikšmė gali vidurkį labai žymiai pakeisti, tiesiai pasakius – iškreipti (taip leistina sakyti tada, kai kyla įtarimas, jog labai didelė arba labai maža požymio reikšmė imtyje yra atsiradusi per klaidą ar neapsižiūrėjimą).

Trumpai apie kelias kitas vidurkio atmainas:

**Nupjautuoju** vadinamas toks vidurkis, kuris skaičiuojamas ne iš visos variacinės eilutės, bet tik iš centrinės jos dalies, gautos atmetus po lygiai mažiausių ir didžiausiųjų požymio reikšmių (pvz., 80% nupjautasis vidurkis skaičiuojamas atmetant 10% mažiausių ir 10% didžiausių imties reikšmių). Dėl to nupjautasis vidurkis yra žymiai atsparesnis ekstremalių reikšmių įtakai ir naudotinas tada, kai šios atrodo esančios nepatikimos. *Excel*’yje jam apskaičiuoti skirta funkcija [=]TRIMMEAN (skDuomenųBlokas; skAtmestinasProcentas). Pagal antrąjį parametą apskaičiuojama, po kelias ekstremalias reikšmes atmesti iš variacinės eilutės pradžios ir pabaigos (jei pagal tą procentą gaunamas nelyginis atmestinių reikšmių kiekis, jis apvalinamas iki artimiausio lyginio, kad variacinės eilutės pradžios ir iš jos galo būtų atmetama po lygiai reikšmių).

**Geometrinis** vidurkis gaunamas iš visų požymio reikšmių sandaugos ištraukus šaknį, kurios laipsnis atitinka imties tūrį (jis paprastai apskaičiuojamas pasinaudojant reikšmių logaritmais). *Excel*’yje jam apskaičiuoti skirta funkcija [=]GEOMEAN(skDuomenųBlokas).

**Kvadratinis** vidurkis gaunamas ištraukus kvadratinę šaknį iš reikšmių kvadratų vidurkio. Su *Excel*’iu jį apskaičiuoti galima skaičiuojamąja išraiška [=] SQRT(AVERAGE(skDuomenųBlokas)); tik duomenų bloke turi būti ne pačios reikšmės, o jų *kvadratai*.

**Harmoninis** vidurkis gaunamas imties tūrį padalijus iš skaičių, atvirkštinių požymio reikšmėms, bendros sumos. *Excel*’yje jam apskaičiuoti skirta funkcija [=]HARMEAN(skDuomenųBlokas).

#### 4. Dispersija ir standartinis nuokrypis

Abu šie „giminingi“ parametrai rodo, kaip konkrečios požymio reikšmės imtyje yra pasklidusios (išsibarsčiusios) vidurkio atžvilgiu. Vidurkis teišryškina tik vieną, dažniausiai abstraktų, išivaizduojamą „centrinį“ požymio reikšmių (įvertinimų) tašką, jų „svorio centrą“, bet nieko nepasako apie tai, koku mastu ir kaip dažnai *realiosios* požymio reikšmės imtyje yra nutolusios nuo šio abstraktaus taško, koks jų *susitelkimo* apie vidurkį *laipsnis*. Norint įvertinti šį pasiskirstymo aspektą, vertėtų imti *skirtumus*, susidarančius tarp konkrečios požymio reikšmės, gaunamos atskiro bandymo metu, ir visų reikšmių vidurkio. Tačiau, kadangi realiosios požymių reikšmės nukrypsta nuo vidurkio *į abi puses*, aritmetinė visų skirtumų suma palaipsniui išsilygintų ir galų gale anuliuotųsi, prilygtų nuliui. Todėl sumuoti pačių skirtumų neišėina, ir vietoje jų yra imami arba jų *absolutiniai dydžiai* (moduliai), arba – dar dažniau – jų *kvadratai*: ir vienos, ir kitos transformacijos (pakeitimo) atveju nuosekliai gaunami tuos skirtumus tiesiogiai atspindintys *teigiami* dydžiai, kuriuos jau galima sumuoti. Gautąsias sumas – tiek modulių, tiek ir kvadratų – priimta yra vėlgi dalinti iš imties tūrio, kitaip tariant, visiems ištirtiesiems objektams „padalinti po lygiai“. Taigi – ir vėl *vidurkis*: arba *skirtumų modulių*, arba *skirtumų kvadratų* vidurkis. Abu jie rodo iš esmės tą patį: *ko-*



kiu laipsniu požymio reikšmės imtyje **vidutiniškai** yra nutolusios nuo savojo vidurkio, tad užtenka apskaičiuoti vieną kurį iš šių dydžių. Paprastai skaičiuojamas tų skirtumų kvadratų vidurkis. Jis vadinamas *dispersija*.

### **Dispersija**

*Skirtumų*, susidarančių tarp reikšmių vidurkio ir konkrečių požymio reikšmių, atitinkančių kiekvieną imties objektą, kvadratų vidurkis. Įprasta žymėti  $s^2$ . Apskaičiuojama taip pat, kaip ir „paprastas“ vidurkis, tik pirma reikia apskaičiuoti atitinkamus skirtumus ir pakelti juos kvadratu.

Vienas niuansas: **jeigu imties tūris (dydis, tiriamųjų objektų skaičius) yra palyginti labai mažas, mažesnis negu 30, apskaičiuojant dispersiją skirtumų kvadratų suma dalinama ne iš „pilno“, bet iš vienetu sumažinto imties tūrio, t. y. ne iš  $n$ , bet iš  $(n-1)$** . Teorinis to reikalavimo pagrindimas čia neapartarinėjamas, tik pridurtina, kad neretai dispersija, gauta kvadratų sumą dalinant iš  $n-1$ , vadinama imties dispersija, o gauta dalinant iš  $n$  – populiacijos dispersija (ją dera skaičiuoti tada, kai populiacija ištiriama visa ištisai).

### **Vidutinis kvadratinis nuokrypis arba standartinis nuokrypis**

Dydis, gaunamas ištraukus kvadratinę šaknį iš dispersijos. Tokiu būdu jis paverčiamas vėl „linijiniu“ (t. y. geometriškai galimu pavaizduoti atitinkamo ilgio *linija*; dispersiją gi tektų vaizduoti *plotu*) dydžiu ir tampa tam tikra prasme bendramatiškas tiek imtyje figūruojančioms požymio reikšmėms, tiek ir pačiam empiriniam vidurkiui. Paprastai žymimas  $s$  (dažniausiai – empirinis) arba graikiška raide *sigma* (ypač – apskaičiuotas teoriškai). Vienas iš pačių populiariausių pasiskirstymo parametrų. Kartais vadinamas tiesiog (nuokrypio) *standartiniu nuokrypiu*.

Vidutinį kvadratinį nuokrypį praktiškai tenka skaičiuoti taip pat dažnai, kaip ir vidurkį. Abu tie dydžiai (vidurkis ir vidutinis kvadratinis nuokrypis), paimti drauge, jau gana išsamiai apibūdina esmines, daugeliu atžvilgių svarbias pasiskirstymų ypatybes, todėl su jais teks susidurti kone „kiekviename žingsnyje“.

Lyginant skirtingus pasiskirstymus, ypač – tuos, kur požymių reikšmės nevienodos ar išmatuotos skirtingais matais (sakysim, norint lyginti žodžių ilgio skiemenimis pasiskirstymą ir sakinių ilgio žodžiais pasiskirstymą) kartais logiškiau yra gretinti ne pačius vidurkius ir vidutinius kvadratinius nuokrypius, bet tų dydžių tarpusavio *santykį*, kuris vadinamas *variacijos koeficientu*. Tai – vidutinio kvadratinio (standartinio) nuokrypio santykis su vidurkiu.

Rečiau praktikuojami empiriniai pasiskirstymo parametrai yra *asimetrijos koeficientas* ir *eksceso koeficientas*, kurie dažnai vadinami tiesiog *asimetrija* ir *ekscesu*. Šių parametrų prasmės ir skaičiavimo kol kas dar nesiaiškinsime, bet vėliau su jais gal ir susidursime.

## **5. Pagrindinės pasiskirstymo parametrų formulės ir funkcijos**

Įsidėmėtini trys dalykai:

1. Tradicinės formulės, pagal kurias galima apskaičiuoti vieną ar kitą empirinio pasiskirstymo parametru, įvairiuose šaltiniuose (matematinės statistikos vadovėliuose ir pan.) gali būti pateikiamos įvairios, nes dažnai *tą pačią* parametro *reikšmę* galima gauti keliais *skirtingais būdais*. Ta prasme ir patys tokie apskaičiavimo būdai, ir juos „reprezentuojančios“ formulės yra lygiagrečios, savaip *sinonimiškos*, ir rinktis galima bet kurią iš jų, paprastai – tą, kuri dėl kokių nors priežasčių pasirodo esanti parankesnė už kitas.

2. Formulėse, net ir perteikiančiose tuos pačius parametrų apskaičiavimo būdus, gali būti panaudoti *skirtingi sutartiniai žymėjimai* (simboliai, kintamieji). Matematinę simbolių funkciją paprastai atlieka lotyniškojo alfabeto raidės, bet skirtinguose šaltiniuose *tos pačios raidės* gali turėti skirtingą prasmę, žymėti *skirtingus dalykus*. Paprastai greta formulių visuomet yra nurodoma (paaiškinama) ir joje pavartotų simbolių (raidžių) prasmė, tad pakanka tik sutelkti į tai dėmesį.

3. Tos pačios formulės, kuriose pavartoti net ir tie patys simboliai, gali turėti *skirtingą grafinę išvaizdą*, kitaip tariant, gali būti *užrašytos skirtingais būdais* (matematikos „normos“ gana dažnai leidžia tą patį reiškinį užrašyti įvairiai).

Čia pateikiamos kelios buvusios pačios reikalingiausios formulės, stengiantis perteikti jų matematinę prasmę. O įvairuojančiai užrašomi jų variantai šiandieną beturi labiau „istorinę“ ar pažintinę prasmę, todėl čia jie nepateikiami; kita vertus, juos irgi būtina pasiaiškinti ir suprasti, kai norima skaityti matematinei statistikai skirtą literatūrą. O praktiniams skaičiavimams svarbiau yra mokėti pasinaudoti atitinkamomis kompiuterinių programų galimybėmis. Todėl lygiagrečiai pateikiamos „tą patį darbą dirbančios“ *Excel* funkcijos.

„Pirminiai“ raidiniai simboliai visose čia pateikiamose formulėse reiškia iš esmės tą patį:

$n$  – bendras bandymų kiekis, t.y. imties tūris  
 $i$  – eilės numeris (1, 2, 3, 4, ...  $k$ )  
 $x_i$  –  $i$ -tosios (pirmosios, antrosios, trečiosios iš eilės ir t.t.) požymio reikšmės „vertė“ (dydis)  
 $m_i$  –  $i$ -tosios požymio reikšmės absoliutinis dažnumas (pasikartojimų imtyje kiekis)  
 $x_{\min}$ ,  $x_{\max}$  – mažiausioji ir didžiausioji požymio reikšmės imtyje

Kaip *Excel* funkcijų (ar skaičiuojamųjų išraiškų) argumentas visur nurodomas duomenų blokas, žymimas tiesiog „skDB“. Tai ištisinė *Excel* lakšto ląstelių zona, kurioje įrašyti pirminiai statistiniai duomenys (tiriamąjį požymio reikšmės).

Parametras	Formulė	<i>Excel</i> funkcija arba išraiška
Variacijos žingsnis (R)	$R = x_{\max} - x_{\min}$	[=] MAX(skDB)-MIN(skDB)
Santykinis dažnumas ( $p$ ; $p^*$ ; $d$ )	$p_i = m_i / n$	Dalybos operacija (ir atitink. operandai)
Vidurkis ( $\bar{x}$ ; $x_{\text{vid}}$ )	$x_{\text{vid}} = (x_1 + x_2 + x_3 + \dots + x_n) / n$	[=] AVERAGE(skDB)
Dispersija ( $s^2$ )	$s^2 = ((x_1 - x_{\text{vid}})^2 + (x_2 - x_{\text{vid}})^2 + \dots + (x_n - x_{\text{vid}})^2) / (n - 1)$ ; arba $s^2 = (x^2)_{\text{vid}} - (x_{\text{vid}})^2$	[=] VAR(skDB)
Standartinis (arba vidutinis kvadratinis) nuokrypis ( $s$ )	$s = \sqrt{s^2}$	[=] STDEV(skDB)
Linijinis nuokrypis ( $l$ )	$l = ((x_1 - x_{\text{vid}}) + (x_2 - x_{\text{vid}}) + \dots + (x_n - x_{\text{vid}})) / n$	[=] AVEDEV(skDB)

Kitų parametų prasmę perteikiančios formulės ir/ar jų apskaičiavimo su *Excel* būdai, kai jų prireiks, bus nurodomi atitinkamose tolesnių konspektų vietose.